# RANDOMIZATION AND COMPANION ALGORITHMS IN STOCHASTIC APPROXIMATION WITH SEMIMARTINGALES

Von der Fakultät Mathematik und Physik der Universität Stuttgart
zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung


Vorgelegt von

TIMO PFROMMER

aus Stuttgart


| | |
|---|---|
| **Hauptberichter:** | PD Dr. Jürgen Dippon |
| **Mitberichter:** | Prof. Dr. Ingo Steinwart |
| | |
| **Tag der mündlichen Prüfung:** | 08.02.2018 |


Institut für Stochastik und Anwendungen
der Universität Stuttgart


2018

# Abstract


Methods of stochastic approximations are used for the recursive estimation of parameters of an unknown function which can only be observed with noise. These parameters are for example roots or extreme points. For root-finding the Robbins-Monro algorithm

$$Z_{n+1} = Z_n - a_n Y_n(Z_n) \quad \text{with} \quad Y_n(z) = f(z) - M_n,$$

and $a_n > 0$ is very common. In order to estimate extrema of multivariate functions $f\colon \mathbb{R}^d \to \mathbb{R}$ the Kiefer-Wolfowitz recursion

$$Z_{n+1} = Z_n - a_n Y_n(Z_n)$$

with

$$Y_n(z) = \frac{1}{2c_n} \left\{ \left( f(z + c_n e_i) - M_{n,i}^+ \right) - \left( f(z - c_n e_i) - M_{n,i}^- \right) \right\}_{i \in \{1,\dots,d\}},$$

$a_n > 0$, $c_n > 0$, is widely used. Here only $Y_n(Z_n)$ is observable, but not its individual components. Moreover at each iteration step of the Kiefer-Wolfowitz recursion $2d$ observations are required. In order to reduce the number of observations, randomized stochastic approximation algorithms were introduced, such as

$$Z_{n+1} = Z_n - \frac{a_n}{2c_n} D_n^{-1} \left\{ \left( f(z + c_n D_n) - M_n^+ \right) - \left( f(z - c_n D_n) - M_n^- \right) \right\}$$

with a $d$-dimensional random sequence $D_n$ and $D_n^{-1}$ it's component-wise inverse. These require only two observations per step. Over the years many extensions of such randomized algorithms were developed. Of a special interest is

$$Z_{n+1} = Z_n - \frac{a_n}{c_n} D_n^{-1} \left\{ f(z + c_n D_n) - M_n^+ \right\}$$

which requires actually only one observation per step, using slightly stronger conditions. However one condition, that all these algorithms have in common, is the assumption of $D_n$ being independent and identically distributed (i.i.d.). This makes an extension of such algorithms to a path-continuous setting impossible. In this thesis a unification of the preceding procedures is presented and extended to a semimartingale setting. For that purpose $D_n$ is no longer assumed to be i.i.d. and thereby more sophisticated methods of proof are to be performed.

The first chapter gives a historical introduction to stochastic approximation in general. Moreover semimartingale settings and randomized algorithms are introduced as well as their benefits. After that, a generic randomized semimartingale algorithm is presented. It does not only establish the theory for not yet investigated time-continuous randomized procedures but also contains all known time-discrete special cases as well as non-discovered ones.

Chapter 2 investigates the almost sure convergence of the generic algorithm. Special cases in a semimartingale, time-continuous and time-discrete setting are derived. It turns out that the presented framework also offers the possibility to handle deterministic perturbation functions which yield the same a.s. convergence results as the randomization processes. Particular examples of useful randomization are handled as well. The chapter closes with visualization of simulation results.

In the third chapter almost sure convergence rates are derived. Again the rates of the randomized semimartingale setting are shown to hold true in the special cases that were presented in previous chapters. Finally the different perturbation designs are compared by simulations of the empirical $L^2$-error.

Based on methods for the estimation of roots or extreme points, the second part of this thesis presents a generalized companion stochastic approximation method of the form

$$\Upsilon_n - \Upsilon_{n-1} = -\tilde{a}_n \Upsilon_{n-1} + \tilde{a}_n \widetilde{Y}_n(Z_n), \text{ with } \tilde{a}_n > 0,\ \widetilde{Y}_n(Z_n) = G_n(Z_n) + \frac{k_n}{\tilde{a}_n} M_n(Z_n).$$

Here $Z_n$ is generated by a leading algorithm, like the Robbins-Monro or the Kiefer-Wolfowitz procedure and $\widetilde{Y}_n(Z_n)$ stands for the noisy observation at $Z_n$ of the parameter of interest. The individual components of $\widetilde{Y}_n(Z_n)$, namely the value of the estimator $G_n(Z_n)$, as well as its measuring error $M_n(Z_n)$, are not observable. Companion algorithms can be interpreted as a solution process of a generalized semimartingale stochastic integral equation of the form

$$\Upsilon_t = \Upsilon_0 + \int_0^t \tilde{a}_s \Big( G_s - \Upsilon_{s-} \Big) \mathrm{d}R_s + \int_0^t k_s M(\mathrm{d}s, \Upsilon_{s-}).$$

The asymptotic behaviour of this process $\Upsilon = (\Upsilon_t)_{t \geqslant 0}$ is discussed. The dicussion is based on the work of Mokkadem and Pelletier where two companion-type algorithms for the underlying Kiefer-Wolfowitz algorithm were presented, but only in a time-discrete setting.

The first chapter of the thesis' second part gives a technical motivation for the time-discrete companion algorithms of Mokkadem and Pelletier, that apply to the Kiefer-Wolfowitz algorithm.

In the sixth chapter, a general semimartingale-type companion algorithm is presented. This generalizes the ideas of Mokkadem and Pelletier in two ways. On one hand, it presents a companion to an arbitrary stochastic approximation algorithm, and not only to the Kiefer-Wolfowitz algorithm. For example companion algorithms for the Robbins-Monro procedure but also new companions for the Kiefer-Wolfowitz procedure can be derived. On the other hand, the algorithm is presented in a semimartingale framework. It turns out, that it includes time-discrete as well as time-continuous

settings. We show consistency of the general semimartingale-type companion algorithm and consider special settings in the semimartingale context. Corresponding results in a time-continuous and a time-discrete framework are derived. It unfolds, that the consistency results of Mokkadem and Pelletier are special cases of the time-discrete framework.

The seventh chapter is devoted to the rate of convergence of the general algorithm. We consider special settings of the algorithms of the previous chapter. We point out how the rate of convergence depends on the gain processes $(a_t)_{t\geqslant 0}$, $(c_t)_{t\geqslant 0}$, $(\tilde{a}_t)_{t\geqslant 0}$, $(k_t)_{t\geqslant 0}$ and the smoothness of $f$ in the leading algorithm. Consequently we discuss settings in which the underlying and the companion algorithm cannot simultaneously converge at an optimal rate. Again, time-continuous and time-discrete settings are established as special cases.

Chapter eight establishes asymptotic normality results. For that purpose an almost $L^2$-convergence result of the underlying algorithm is used. In contrast to the Kiefer-Wolfowitz algorithm, there was no such result for the Robbins-Monro algorithm before. After showing this missing almost $L^2$-result, we attend to asymptotic normality under parameter settings given in the previous chapters.

# Zusammenfassung


Die Methoden der stochastischen Approximation werden zur rekursiven Bestimmung von unbekannten Parametern einer Funktion verwendet, die nur mit Rauschen beobachtet werden kann. Diese Parameter sind beispielsweise Null- oder Extremstellen. Zur Ermittlung von Nullstellen hat sich der Robbins-Monro Algorithmus

$$Z_{n+1} = Z_n - a_n Y_n(Z_n) \quad \text{mit} \quad Y_n(z) = f(z) - M_n,$$

und $a_n > 0$ etabliert. Für die Schätzung von Extremstellen ist das Kiefer-Wolfowitz Verfahren

$$Z_{n+1} = Z_n - a_n Y_n(Z_n)$$

mit

$$Y_n(z) = \frac{1}{2c_n} \left\{ \left( f(z + c_n e_i) - M_{n,i}^+ \right) - \left( f(z - c_n e_i) - M_{n,i}^- \right) \right\}_{i \in \{1,\dots,d\}},$$

$a_n > 0$, $c_n > 0$ weit verbreitet. Hierbei ist nur $Y_n(Z_n)$ beobachtbar, aber nicht seine einzelnen Komponenten. Desweiteren werden für jede Iteration des Kiefer-Wolfowitz Verfahrens $2d$ Beobachtungen benötigt. Um diese Anzahl verringern zu können, wurden randomisierte stochastische Approximationsverfahren eingeführt, wie beispielsweise

$$Z_{n+1} = Z_n - \frac{a_n}{2c_n} D_n^{-1} \left\{ \left( f(z + c_n D_n) - M_n^+ \right) - \left( f(z - c_n D_n) - M_n^- \right) \right\}$$

mit einer $d$-dimensionalen Zufallsfolge $D_n$ und $D_n^{-1}$ als deren komponentenweiser Inversen. Diese benötigen lediglich zwei Beobachtungen pro Schritt. Mit den Jahren wurden vielerlei solcher randomisierter Algorithmen entwickelt. Von einem speziellen Interesse ist

$$Z_{n+1} = Z_n - \frac{a_n}{c_n} D_n^{-1} \left\{ f(z + c_n D_n) - M_n^+ \right\}$$

bei welchem nur eine einzige Beobachtung pro Schritt genügt, auch wenn die Voraussetzungen leicht verschärft werden müssen. Eine Bedingung jedoch, die alle diese Verfahren gemeinsam haben, ist die Eigenschaft, dass die $D_n$ unabhänging und identisch verteilt (u.i.v.) sind. Dies macht die Erweiterung solcher Algorithmen in einen pfadstetigen Kontext unmöglich. Diese Thesis stellt die Vereinheitlichung der zuvor präsentierten Verfahren vor, welche zudem in einen Semimartingalrahmen erweitert

werden. Daher wird $D_n$ nicht mehr als u.i.v. vorausgesetzt und weitergehende Methoden sind anzuwenden, um die Beweise durchführen zu können.

Das erste Kapitel gibt eine historische Einführung in die stochastische Approximation. Desweiteren werden Semimartingal- und Randomisierungsverfahren vorgestellt sowie deren Vorzüge. Anschließend wird ein generischer randomisierter Semimartingalalgorithmus präsentiert. Dieser umfasst nicht nur die Theorie der bisher nicht untersuchten zeitstetigen randomisierten Verfahren, sondern auch alle bekannten zeitdiskreten Spezialfälle und auch noch nicht entdeckte.

Kapitel 2 behandelt die fast sichere Konvergenz des generischen Verfahrens. Spezialfälle sowohl in einem zeitstetigen als auch in einem zeitdiskreten Rahmen werden vorgestellt. Es stellt sich heraus, dass der präsentierte Kontext die Möglichkeit bietet, deterministische Störungsfunktionen zu verwenden, welche die selben fast sicheren Konvergenzresultate liefern wie die Randomisierungsprozesse. Zudem werden spezielle Beispiele von nutzbringender Randomisierung behandelt. Das Kapitel schließt mit der Visualisierung von Simulationsresultaten.

Im dritten Kapitel werden fast sichere Konvergenzraten hergeleitet. Auch hier wird gezeigt, dass die Raten des Semimartingalverfahrens sich auf die in den vorigen Kapiteln vorgestellen Spezialfälle übertragen. Am Ende werden den verschiedenen Designs von Störungen mithilfe der Simulation des empirischen $L^2$-Fehlers verglichen.

Begründet auf den Methoden zur Schätzung von Null- oder Extremstellen wird im zweiten Teil der Arbeit ein verallgemeinertes begleitendes stochastisches Approximationsverfahren der Form

$$\Upsilon_n - \Upsilon_{n-1} = -\tilde{a}_n \Upsilon_{n-1} + \tilde{a}_n \widetilde{Y}_n(Z_n), \text{ mit } \tilde{a}_n > 0,\ \widetilde{Y}_n(Z_n) = G_n(Z_n) + \frac{k_n}{\tilde{a}_n} M_n(Z_n)$$

vorgeschlagen. Hierbei wird $Z_n$ erzeugt durch einen zugrundeliegenden Algorithmus, wie dem Robbins-Monro- oder dem Kiefer-Wolfowitz-Verfahren, und $\widetilde{Y}_n(Z_n)$ steht für die verrauschte Beobachtung des zu schätzenden Parameters an der Stelle $Z_n$. Die einzelnen Komponenten von $\widetilde{Y}_n(Z_n)$, also der Wert des Schätzers $G_n(Z_n)$, sowie dessen Messfehler $M_n(Z_n)$, sind nicht getrennt beobachtbar. Begleitende Verfahren können als Lösungsprozess verallgemeinerter semimartingalartiger stochastischer Integralgleichungen der Form

$$\Upsilon_t = \Upsilon_0 + \int_0^t \tilde{a}_s \Big( G_s - \Upsilon_{s-} \Big) \mathrm{d}R_s + \int_0^t k_s M(\mathrm{d}s, \Upsilon_{s-})$$

interpretiert werden. Neben den bereits erwähnten zeitdiskreten Verfahren sind hier auch zeitkontinuierliche Verfahren enthalten. Es wird das asymptotische Verhalten dieses Lösungsprozesses $\Upsilon = (\Upsilon_t)_{t \geqslant 0}$ diskutiert. Diese Überlegung fußt auf der Arbeit von Mokkadem und Pelletier, in der Begleitalgorithmen in einem zeitdiskreten Zusammenhang zum zugrundeliegenden Kiefer-Wolfowitz Verfahren vorgestellt wurden.

Das erste Kapitel des zweiten Teils der Arbeit gibt eine heuristische Begründung für die Form des Begleitalgorithmus von Mokkadem und Pelletier. In diesem wird zusätzlich zum Kiefer-Wolfowitz-Algorithmus, welcher die Minimalstelle behandelt, noch der Funktionswert geschätzt.

Im sechsten Kapitel wird ein verallgemeinerter semimartingalartiger Begleitalgorithmus vorgeschlagen. Dieser erweitert die Ideen von Robbins und Monro in zweierlei Hinsicht. Zum einen stellt er einen Begleitalgorithmus für ein beliebiges stochastisches Approximationsverfahren dar, und nicht lediglich zum Kiefer-Wolfowitz-Algorithmus. Beispielsweise werden ein Algorithmus zum Robbins-Monro-Verfahren und ein weiteres Verfahren für den Kiefer-Wolfowitz-Algorithmus hergeleitet. Zum anderen wird der Begleitalgorithmus in einem Semimartingal-Zusammenhang dargestellt. Entsprechende Ergebnisse im zeitdiskreten und im zeitstetigen Kontext ergeben sich als Spezialfälle. Es stellt sich heraus, dass die Resultate von Mokkadem und Pelletier in den zeitdiskreten Ergebnissen enthalten sind.

Das siebte Kapitel widmet sich der Konvergenzgeschwindigkeit des verallgemeinerten Verfahrens. Wir betrachten spezielle Fassungen der Algorithmen des vorigen Kapitels. Es wird untersucht, wie die Konvergenzgeschwindigkeit von den Schrittweitenprozessen $(a_t)_{t\geqslant 0}$, $(c_t)_{t\geqslant 0}$, $(\tilde{a}_t)_{t\geqslant 0}$, $(k_t)_{t\geqslant 0}$ und von der Glattheit von $f$ im zugrundeliegenden Verfahren abhängt. Somit werden auch Szenarien diskutiert, in denen der führende und der begleitende Algorithmus nicht jeweils mit optimaler Rate konvergieren. Wie zuvor, wird auf zeitdiskrete und zeitstetige Fassungen eingegangen.

In Kapitel acht werden Ergebnisse zur asymptotischen Normalität präsentiert. Hierfür wird ein Resultat zur fast-$L^2$-Konvergenzrate des zugrundeliegenden Verfahrens verwendet. Im Gegensatz zum Kiefer-Wolfowitz-Algorithmus liegt ein solches Resultat für den Robbins-Monro-Algorithmus nicht vor und muss zunächst erarbeitet werden. Daraufhin widmen wir uns der asymptotischen Normalität unter den Parametern des vorigen Kapitels.

# Acknowledgments

Ich danke allen, die mich während des Entstehens dieser Arbeit unterstützt haben.

Zuerst gebührt mein Dank meinem Doktorvater PD Dr. Jürgen Dippon, der mir den Reiz an meinem Promotionsthema sowie der Stochastik im Allgemeinen offenbart hat, für seine Betreuung.

Desweiteren möchte ich Herrn Prof. Dr. Ingo Steinwart dafür danken, meine Arbeit zu begutachten und mich an seinem Institut zu beschäftigen.

Zudem möchte ich Philipp Thomann und Matthias Giegerich für die kritischen Hinweise zur Rohfassung meiner Arbeit danken.

Besonderer Dank gilt meinen Eltern sowie meiner Freundin Katharina, die mir jederzeit den nötigen Rückhalt gegeben haben.

# Contents

# List of Figures

# Abbreviations

| | |
|---|---|
| RM | Robbins-Monro |
| KW | Kiefer-Wolfowitz |
| RDSA | Random Direction Stochastic Approximation |
| SPSA | Simultaneous Perturbation Stochastic Approximation |
| SPSA1 | One-measurement version of SPSA |
| Gen-Rand | Generic randomized algorithm |
| Ker-Rand-1 | One-measurement randomized kernel gradient algorithm |
| Ker-Rand-2 | Two-measurement randomized kernel gradient algorithm |
| c-Ker-Rand-1 | Continuous counterpart of Ker-Rand-1 |
| c-Ker-Rand-2 | Continuous counterpart of Ker-Rand-2 |
| d-Ker-Rand-1 | Recursive counterpart of Ker-Rand-1 |
| d-Ker-Rand-2 | Recursive counterpart of Ker-Rand-2 |
| Gen-Comp | Generic companion algorithm |
| c-Gen-Comp | Itô type counterpart for Gen-Comp |
| d-Gen-Comp | Recursive counterpart for Gen-Comp |
| RM-J | Companion for Robbins-Monro for estimation of Jacobian |
| KW-F-1 | Companion for Kiefer-Wolfowitz for estimation of the function value using one additional function evaluation |
| KW-F-2 | Companion for Kiefer-Wolfowitz for estimation of the function value reusing function evaluations |
| KW-H | Companion for Kiefer-Wolfowitz for estimation of Hessian |
| c-RM | Itô type counterpart for Robbins-Monro |
| c-KW | Itô type counterpart for Kiefer-Wolfowitz |
| c-KW-1D | One-dimensional special case of c-KW |
| c-RM-J | Itô type counterpart for RM-J |
| c-KW-F-1 | Itô type counterpart for KW-F-1 |
| c-KW-F-2 | Itô type counterpart for KW-F-2 |
| c-KW-H | Itô type counterpart for KW-H |
| d-RM | Recursive counterpart for Robbins-Monro |

| | |
|---|---|
| d-KW | Recursive counterpart for Kiefer-Wolfowitz |
| d-RM-J | Recursive counterpart for RM-J |
| d-KW-F-1 | Recursive counterpart for KW-F-1 |
| d-KW-F-2 | Recursive counterpart for KW-F-2 |
| d-KW-H | Recursive counterpart for KW-H |
| càdlàg | Right-continuous with left limits |

# Part I

# RANDOMIZED ALGORITHMS

# 1 Introduction

This part of the thesis deals with optimization of systems with multiple unknown parameters by randomized stochastic approximation in a semimartingale context. Before presenting the results, a historical overview is given.

## 1.1 Historical Introduction

Stochastic approximation has its origin in the 1950s, when Robbins and Monro [33] presented a recursive algorithm for finding the root $z^*$ of an unknown increasing function $f\colon \mathbb{R} \to \mathbb{R}$, where the statistician only has noisy observations of the function values. They suggested a recursion of the form

$$Z_{n+1} - Z_n = -a_n Y_n(Z_n) \quad \text{with} \quad Y_n(z) := f(z) + M_n(z), \tag{1.1}$$

where $M_n(z)$ represents the additional noise of $f$ at $z$. This procedure resembles Newton's method

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

in numerical analysis. At step $n+1$ the user observes the function $f$ at the point $Z_n$, receiving a noisy function value $Y_n(Z_n)$. Unlike Newton's method, for the Robbins-Monro algorithm usually no observation of the gradient $\nabla f$ nor noisy observations of it are assumed to be available. Instead of the gradient one uses a damping sequence $(a_n)$ which is chosen by the experimenter. A typical choice is $a_n := a/n$ with $a > 0$. Robbins and Monro [33] proved that $(Z_n)$ converges in probability to the root $z^*$ of $f$.

On this basis Kiefer and Wolfowitz [20] suggested in 1952 an algorithm for the search of stationary points related to minima. Like Robbins and Monro they did not assume an observable gradient $\nabla f$. Instead it was estimated by $(2c_n)^{-1}\big(f(x+c_n) - f(x-c_n)\big)$ with a sequence $(c_n)$ tending to zero. Hence the recursion is of the form

$$Z_{n+1} - Z_n = -a_n Y_n(Z_n) \quad \text{with} \quad Y_n(z) := \frac{1}{2c_n}\Big(f(z+c_n) - f(z-c_n) + M_n(z)\Big).$$

Common choices to achieve convergence are $a_n := a/n$ and $c_n := cn^{-\gamma}$ with $\gamma \in (0, 1/2)$. Stationary points for maxima can be found by changing the recursion to $Z_{n+1} - Z_n = +a_n Y_n(Z_n)$, but in the following we will stick to minima. Kiefer and

Wolfowitz showed convergence in probability for their algorithm.

In 1954, Blum [3] presented a multidimensional extension of the Robbins-Monro recursion to handle multivariate functions $f\colon \mathbb{R}^d \to \mathbb{R}^d$. Furthermore he suggested a multi-variable Kiefer-Wolfowitz recursion

$$Z_{n+1} - Z_n = -a_n Y_n(Z_n)$$

with

$$Y_n(z) = \frac{1}{2c_n} \{f(z + c_n e_i) - f(z - c_n e_i) + M_{n,i}\}_{i \in \{1,\dots,d\}}, \tag{1.2}$$

where $f\colon \mathbb{R}^d \to \mathbb{R}$, and $M_{n,i}$ comprising the observation noise. Note that in each iteration step in (1.2) $2d$ observations have to be made. Asymptotic normality of both kinds of processes $(Z_n)$, generated by (1.1) and (1.2), was first shown by Sacks [36] in 1957. In 1967 Fabian [15] suggested a modified estimator $Y_n(z)$ for the gradient of $f$ at $z$, if differentiability of $f$ of odd order $p \geqslant 3$ can be assumed. He showed that his estimate, which is based on $d(p-1)$ observations, achieves the a.s. and $L^2$-convergence rate $n^{-(p-1)/(2p)+\epsilon}$, $\forall \epsilon > 0$ and $n^{-(p-1)/(2p)}$, respectively. Therefore it approaches the rate of the Robbins-Monro process, which is $n^{-1/2}$, if $f$ is differentiable of any order. Dippon and Renz [12] constructed an unbiased estimator. It is worth to mention, that Chen [6] showed in 1988 that the rate $n^{-(p-1)/(2p)}$ is optimal.

A ground-breaking innovation regarding the Robbins-Monro method was suggested by Polyak [31] in 1990. Instead of the original algorithm he considered an averaged Robbins-Monro scheme with slowly decaying weights $a_n = an^{-\alpha}$, where $0 < \alpha < 1$. It turned out that except of $f'(z^*) > 0$ no assumption on the usually unknown first derivative of $f$ has to be made. In the classical context the asymptotic variance is given by $(a^2\sigma^2)/(2af'(z^*) - 1)$ such that $a$ should not be chosen too large in order to attain small variance. But the stability condition $f'(z^*) > 1/(2a)$ requires $a$ to be chosen large enough. When using averaged algorithms this dilemma does not arise as no condition relating $a$ and the derivative of $f$ is necessary. Although using a Robbins-Monro algorithm with asymptotic rate less than $n^{-1/2}$, its averaged process achieves rate $n^{-1/2}$ and is optimal regarding the variance. Hence, the averaged scheme has asymptotic and stability benefits over the original algorithm. In 1996 and 1997 Dippon and Renz [11, 12] applied Polyak's ideas to a Kiefer-Wolfowitz algorithm with weighted means. Important surveys of time-discrete stochastic approximation are the books of Ljung et al. [24] and Duflo [14].

Even though Itô published his epoch making contributions on stochastic integral equations in the 1940s, it took until the 1970s when Nevel'son and Has'minskiĭ [29] studied stochastic approximation processes which are generated by the following stochastic integral equations of Itô type. To estimate the root of $f\colon \mathbb{R} \to \mathbb{R}$ in a time-continuous framework with a $d$-dimensional Brownian motion $(W_t)_{t\geqslant 0}$ and diffusion function $\sigma\colon [0,\infty) \times \mathbb{R}^d \to \mathbb{R}^d$,

$$Z_t = Z_0 - \int_0^t a_s f(Z_s)\mathrm{d}s - \int_0^t a_s \sigma_s(Z_s)\mathrm{d}W_s, \tag{c-RM}$$

and to estimate the minimum of $f\colon \mathbb{R} \to \mathbb{R}$, with a 1-dimensional Brownian motion $(W_t)_{t\geqslant 0}$ and diffusion function $\sigma\colon [0,\infty) \times \mathbb{R} \to \mathbb{R}$,

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s}\Big(f(Z_s + c_s) - f(Z_s - c_s)\Big)\mathrm{d}s - \int_0^t \frac{a_s}{2c_s}\sigma_s(Z_s)\mathrm{d}W_s. \qquad \text{(c-KW-1D)}$$

In the same book they treated consistency, rate of convergence and asymptotic normality of the Robbins-Monro algorithm in both, the recursive and the Itô framework. Furthermore they showed consistency of the Itô type Kiefer-Wolfowitz algorithm.

Until the end of the 1980s time-discrete and time-continuous algorithms were treated separately. In the second half of the 20th century, the foundations of stochastic analysis were established. As a generalization of Itô-processes, semimartingales were investigated. Semimartingales offer a self-contained integration theory and can, roughly speaking, be considered as the sum of a process of finite variation on compacts and a local martingale. As a consequence Levy processes are included and thereby time-discrete recursions as well. Then in 1989 Melnikov [27] found a unification of time-discrete and time-continuous algorithms within a semimartingale framework. Afterwards, together with Rodkina [34] and Valkeila [42], consistency as well as asymptotic normality of the Robbins-Monro process and consistency of the Kiefer-Wolfowitz process were shown. The conditions for these results however are very technical and hard to verify. A few years later, Lazrieva, Sharia and Toronjadze [22] suggested the solution of

$$Z_t = Z_0 - \int_0^t H_s(Z_{s-})\mathrm{d}R_s - \int_0^t M(\mathrm{d}s, Z_{s-}),$$

with $Z_{s-}$ the left-continuous modification of $Z_s$, as a general semimartingale version of the Robbins-Monro process. The choices $H_s(Z_{s-}) := a_s f(Z_s)$ and $M(\mathrm{d}s, Z_{s-}) := a_s\sigma \mathrm{d}W_s$ as well as $H_s(Z_{s-}) := a_n f(Z_{n-1})$ and $M(\mathrm{d}s, Z_{s-}) := a_n V_n$ show how the Itô type and the recursive Robbins-Monro algorithms are embedded in the semimartingale framework as special cases. For the proof of consistency they showed a generalized Robbins-Siegmund theorem, which in turn is based on a multiplicative decomposition theorem. Thereby the conditions of their theorem are weaker and less technical then those of Melnikov et al. [42]. Furthermore Lazrieva et al. showed asymptotic normality of the original and the averaged process in a path-continuous semimartingale framework. In 2010 Schnizler [37] studied the related Kiefer-Wolfowitz algorithm in detail. However, to prove asymptotic normality he did not need to assume path-continuity.

One of the main disadvantages of the Kiefer-Wolfowitz algorithm is that it requires $2d$ observations of $f$ in each iteration step. In order to handle this high-dimensional problem, several algorithms have been suggested that need only two evaluations per step. Kushner and Clark [21] suggested a method which estimates the gradient of $f$ at $X_n$ by estimating the directional derivative along a randomly chosen direction of the unit sphere $S^d$. Spall [38] formulated an alternative approach, namely simultaneous perturbation stochastic approximation (SPSA), choosing a distribution $F_{SP}$ on $\mathbb{R}^d$ which is the $d$-fold tensor product of a symmetrical distribution concentrated on $\mathbb{R}^d\backslash\{0\}$. Possible distributions are for example the Bernoulli distribution or the uniform distribution concentrated on the vertices of the cube $[-1,1]^d$. In 2002 Dippon

[10] unified such two-measurement randomized stochastic approximation algorithms by using a randomized kernel gradient estimate

$$Y_n(z) = \frac{1}{2c_n} K(D_n) \left\{ \left( f(z + c_n D_n) - W_{n,1} \right) - \left( f(z - c_n D_n) - W_{n,2} \right) \right\}$$

with a kernel function $K\colon \mathbb{R}^d \to \mathbb{R}^d$ and random vectors $D_n \in \mathbb{R}^d$ satisfying $\mathbb{E}(K(D_n) \otimes D_n) = \mathbb{1}_d$. In 1997 Spall [39] presented SPSA1, an estimator

$$Y_n(z) = \frac{1}{c_n} D_n^{-1} \left\{ \left( f(z + c_n D_n) - W_{n,i} \right) \right\}, \text{ with } D^{-1} := (1/D^{(1)}, \dots, 1/D^{(d)}),$$

that only needs one evaluation per iteration step. Therefore the question arises how one can extend Dippon's ideas to such one-measurement algorithms.

All these randomized procedures are in discrete time. In this thesis an abstract algorithm in a semimartingale framework is suggested, including both, one-measurement and two-measurement algorithms. Semimartingales are the largest class of integrators for which an integral of the form $\int_0^t H_s \mathrm{d}X_s$, with $H$ a locally bounded predictable process, is closed, and hence a powerful calculus is available. Obviously time-discrete as well as Itô type versions of one- and two-measurement algorithms follow as special cases. But also more general settings, where observations can only be taken at random times, are possible. Apart from theoretical interest there is a large field of applications that explicitly use general semimartingale models, and not only time-discrete or time-continuous special cases. Some of them are presented in later sections.

Classical Robbins-Monro algorithms are known to converge with rate $n^{-1/2}$, whereas standard Kiefer-Wolfowitz type algorithms converge with a rate not better than $n^{-1/3}$. Therefore it seems natural to prefer Robbins-Monro in stochastic gradient optimization whenever the gradient is available. But in some circumstances it is preferable to apply a Kiefer-Wolfowitz algorithm instead of a stochastic gradient method. We point out the examples given by Spall [40, Ch. 6.2].

- Calculating the gradient can be too costly in time or computational steps.
- There can be human errors by doing the derivations.
- In complex calculations there is the possibility of software coding errors in the implementation of the algorithm.
- Computer algebra packages may have difficulties with the gradient calculations in high dimensions.
- When applying so-called automatic differentiation methods, one needs huge knowledge of the "inner workings" of the software.

Additionally, sometimes it is sensible to formulate a root-finding problem as a stochastic gradient optimization problem. Typical for the Robbins-Monro algorithm is the assumption that the infimum of $(x - x^*)^T f(x)$ over a compact set not containing $x^*$ must be positive, which directs the process to the direction of the root of $f$. Loosely speaking this means, one has to know, whether the function $f$ is increasing or decreasing in the vicinity of the root. Ruppert [35] was one of the first who treated this problem. Instead of finding the root of $f(x)$ one minimizes $g := \|f\|^2$. Now, for the optimization algorithm the assumption $(x - x^*)^T \nabla g(x)$ is bounded away from 0 on a compact set not containing $x^*$ has to be fulfilled. Here, one only has to know whether $g$ has to be minimized or maximized. As the problem is user-defined this is, in contrast to the mean-reverting assumption, always clear. Due to the fact that randomized stochastic approximation algorithms work very well with only one or two measurements per iteration step, and that the rate of convergence of $n^{-1/2}$ can be approached arbitrarily closely, it represents a very powerful alternative.

## 1.2 General Assumptions

In the following, results on consistency of an abstract semimartingale algorithm of the form (1.3) are given. We specialize this setting to one- and two-measurement randomized kernel gradient estimators (1.5) and (1.6). On this basis, time-continuous (2.11) and (2.12) as well as time-discrete (2.13) and (2.14) special cases are derived. Apart from the proof of consistency for algorithm (2.14), which has been done by Dippon [10] under slightly different assumptions, all results are completely new.

We use the following notations. The tensor $x \otimes y \colon \mathbb{R}^d \to \mathbb{R}^d$ is the linear mapping $\langle y, . \rangle x$ , where $x$ and $y$ are two vectors in $\mathbb{R}^d$. The open ball around $x$ with radius $\epsilon$ is given by $U_\epsilon(x)$. Considering a multi-index $m = (m_1, \ldots, m_d) \in \mathbb{N}_0^d$ the length $m_1 + \ldots + m_d$ is denoted by $|m|$ and $m!$ means $m_1! \cdot \ldots \cdot m_d!$. The $m$-th power of a vector $x \in \mathbb{R}^d$ is defined as $x^m = x_1^{m_1} \cdot \ldots \cdot x_d^{m_d}$, where we assume that $0^0 := 1$. The differential operator $\nabla^m$ with respect to $x$ is defined by $\frac{\partial^{m_1}}{(\partial x_1)^{m_1}} \cdots \frac{\partial^{m_d}}{(\partial x_d)^{m_d}}$. The notation $X_t \simeq Y_t$ means that $X_t$ and $Y_t$ are asymptotically equal, i.e. $\lim_{t\to\infty} X_t/Y_t = 1$. We make the general assumption, that all relations, unless explicitly otherwise specified, shall hold a.s.

## 1.3 General Semimartingale Algorithms

We consider a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ satisfying the usual conditions. This means that $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets of $\mathcal{F}$, and that the filtration $\mathbb{F}$ is right-continuous. On this basis a random variable $Z_0$, a random field $M \in \mathbb{R}^d$ and processes $(a_t)_{t\geqslant 0}$, $(c_t)_{t\geqslant 0}$, $(D_t)_{t\geqslant 0}$, and $(R_t)_{t\geqslant 0}$ are defined. The processes $(a_t)_{t\geqslant 0}$ and $(c_t)_{t\geqslant 0}$ shall be predictable with respect to $\mathbb{F}$, and moreover $(\frac{a_t}{c_t})_{t\geqslant 0}$ has to be locally bounded. Furthermore it is assumed that $(R_t)_{t\geqslant 0}$ is increasing, càdlàg (i.e. right-continuous with left-sided limits), predictable with respect to $\mathbb{F}$, and $R_0 = 0$ as well as $\Delta R_0 = 0$ hold. The process $(D_t)_{t\geqslant 0}$ is assumed to be $\mathbb{R}^d$-valued and predictable with respect to $\mathbb{F}$. By $\mathcal{M}^2_{\mathrm{loc}}(\mathbb{P})$ we denote the set of locally square-integrable martingales with respect to $\mathbb{P}$ and $\mathbb{F}$. The random field $\{M(t, D_{t-}, v) \colon t \geqslant 0, v \in \mathbb{R}^d\}$ is $\mathbb{F}$-adapted for all $v \in \mathbb{R}^d$. Furthermore for every $t \geqslant 0$, $v \in \mathbb{R}^d$ the relations $M(t, D_{t-}, v) \in \mathcal{M}^2_{\mathrm{loc}}(\mathbb{P})$ and $\int_0^t \frac{a_s}{c_s}(M(\mathrm{d}s, D_{s-}, v))_{t\geqslant 0} \in \mathcal{M}^2_{\mathrm{loc}}(\mathbb{P})$ hold.

By $o$ and $\mathcal{O}$ we denote the Landau symbols. Moreover, for a stochastic process $(X_t)_{t\geqslant 0}$, we write $X_t = o_{\mathrm{b}}(r_t)$ if $t$ is increasing to infinity and $(X_t/r_t)$ is bounded a.s. Moreover $R^d$ denotes the purely discontinuous part of the process $R$. By $\Delta R_t$ we define the jump $R_t - R_{t-}$. We note that $\Delta R_t = \mathrm{d}R_t^d$. The covariation and the predictable covariation processes of $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$ are denoted by $\lfloor X, Y \rfloor_t$ and $\lceil X, Y \rceil_t$, respectively. The unit vectors of the Euclidean space $\mathbb{R}^d$ are written as $e_1, \dots, e_d$.

In the subsequent sections, it is shown that the stochastic integral equation [Gen-Rand]

$$Z_t = Z_0 - \int_0^t \frac{a_s}{c_s} F(D_{s-}, c_s, Z_{s-}) \mathrm{d}R_s - \int_0^t \frac{a_s}{c_s} M(\mathrm{d}s, D_{s-}, Z_{s-}) \tag{1.3}$$

can be considered as an abstract formulation of randomized stochastic approximation procedures of various forms. In this stochastic differential equation, the process $D$ represents the incorporated randomization that can be chosen by the statistician. As $D$ is assumed to be predictable, the notation $D_{s-}$ instead of $D_s$ seems superfluous here. However the arising time-discrete special cases are easier to compare with already existing results, when $D$ and $Z$ have the same index in the same iteration step. We note that the function $F$ is a composition of predictable processes and thus predictable as well. Hence

$$\sum_{i=1}^d \left\lceil \int_0^{\cdot} \Big( F(D_{\tau-}, c_\tau, Z_{\tau-}) \mathrm{d}R_\tau \Big)_i \right\rceil_t = \int_0^t \| F(D_{s-}, c_s, Z_{s-}) \|^2 \Delta R_s \mathrm{d}R_s^d \tag{1.4}$$

holds true. Equation (1.3) unifies randomized one-measurement and two-measurement stochastic approximation procedures. In the following this will be investigated in detail. It is worth mentioning that algorithms with $2m$, $m \in \mathbb{N}$, simultaneous measurements are included in the two-measurement framework. Details can be found in the papers of Fabian [15], Dippon and Renz [11], [12] and Dippon [10]. But also three-measurement or other odd-valued measurement algorithms are included in (1.3). Even a deterministic $D_s$ is feasible. A sufficient condition for the existence of these integrals including $F$, is that $F$ is continuous in its arguments. In an analogous way the terms of the second integral in (1.3), which include $M$, are defined. The stochastic integral equation (1.3) is assumed to be well-defined. We are interested in the asymptotic behaviour of the process $Z$. Therefore we assume the existence of a unique strong solution on $[0, \infty)$. Existence and uniqueness of stochastic integral equations are well-investigated in the book of Protter [32].

In this thesis a function $K\colon \mathbb{R}^d \to \mathbb{R}^d$ with $\mathbb{E}(K(D_{s-}) \otimes D_{s-}) = \mathbb{1}_d$ is called a kernel function. Here $\mathbb{1}_d$ denotes a $d$-dimensional diagonal matrix such that all diagonal entries are 1, whereas $\vec{1}$ denotes a $d$-dimensional vector with all entries equal to 1. Assume the existence of $\breve{M}$ such that $K(D_{s-})\breve{M}(\mathrm{d}s, Z_{s-})$ equals $M(\mathrm{d}s, D_{s-}, Z_{s-})$ or $2M(\mathrm{d}s, D_{s-}, Z_{s-})$. As special methods to estimate the minimum of $f$ by employing one-measurement and the two-measurement randomized kernel gradient estimations [Ker-Rand-1] and [Ker-Rand-2] there are stochastic integral equations given by

$$Z_t = Z_0 - \int_0^t \frac{a_s}{c_s} K(D_{s-}) \{ f(Z_{s-} + c_s D_{s-}) \} \mathrm{d}R_s - \int_0^t \frac{a_s}{c_s} K(D_{s-}) \breve{M}(\mathrm{d}s, Z_{s-}) \tag{1.5}$$

and

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s} K(D_{s-})\{f(Z_{s-} + c_s D_{s-}) - f(Z_{s-} - c_s D_{s-})\} \mathrm{d}R_s - \int_0^t \frac{a_s}{2c_s} K(D_{s-}) \breve{M}(\mathrm{d}s, Z_{s-}) \tag{1.6}$$

respectively. Note that in these instances $F$ consists of the function $f\colon \mathbb{R}^d \to \mathbb{R}$ and the kernel function $K$ as follows:

$$F(D_{s-}, c_s, Z_{s-}) = \begin{cases} K(D_{s-})\{f(Z_{s-} + c_s D_{s-})\} & \text{in (1.5)} \\ \frac{1}{2} K(D_{s-})\{f(Z_{s-} + c_s D_{s-}) - f(Z_{s-} - c_s D_{s-})\} & \text{in (1.6).} \end{cases}$$

Moreover the classical Kiefer-Wolfowitz algorithm

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s} \{f(Z_{s-} + c_s e_i) - f(Z_{s-} - c_s e_i)\}_{i \in \{1,\dots,d\}} \mathrm{d}R_s - \int_0^t \frac{a_s}{2c_s} M(\mathrm{d}s, Z_{s-})$$

is included in this framework by setting $F(D_{s-}, c_s, Z_{s-}) = \frac{1}{2}\{f(Z_{s-} + c_s e_i) - f(Z_{s-} - c_s e_i)\}_{i \in \{1,\dots,d\}}$.

Throughout the rest of part one of this thesis the following general conditions shall hold.

**Assumption 1.3.1.**

- *The function $F\colon \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ is differentiable with respect to $c$. Here $\nabla_c^k$ denotes the $k$-fold derivative with respect to $c$.*
- *Let $F$ be* factorizable *at $c = 0$ with respect to $\mathtt{d}$ and $z$ in the sense that there are measurable functions $\tilde{f}_0\colon \mathbb{R}^d \to \mathbb{R}$, $\tilde{f}_1\colon \mathbb{R}^d \to \mathbb{R}^d$, $g_0\colon \mathbb{R}^d \to \mathbb{R}^d$ and $g_1\colon \mathbb{R}^d \to \mathbb{R}^{d \times d}$, such that for all $k \in \{0, 1\}$*

$$\nabla_c^k F(\mathtt{d}, 0, z) = g_k(\mathtt{d}) \tilde{f}_k(z). \tag{1.7}$$

- *The norm of $g_1(.)$ is defined as the Frobenius norm, i.e.*

$$\|g_1(.)\| := \sqrt{\sum_{i=1}^d \sum_{j=1}^d |g_1(.)_{ij}|^2}.$$

- *$F$ is* affine *in the sense of*

$$\nabla_c^k F(\mathtt{d}, c, z) = \nabla_c^k F(\mathtt{d}, 0, z + c\mathtt{d}) \text{ for all } k \in \{0, 1\}, \mathtt{d} \in \mathbb{R}^d \text{ and } z \in \mathbb{R}^d. \tag{1.8}$$

In both, (1.5) and (1.6), we find $g_0(\mathtt{d}) = K(\mathtt{d})$, $g_1(\mathtt{d}) = K(\mathtt{d}) \otimes \mathtt{d}$ and $\tilde{f}_1(z) = \nabla f(z)$. However in (1.5) $\tilde{f}_0(z) = f(z)$ whereas in (1.6) $\tilde{f}_0(z) = 0$ holds true.

# 2 Almost Sure Convergence

In this chapter almost sure convergence is investigated. We begin with a general result on a generic semimartingale algorithm. Later on, kernel-based algorithms are derived, which in turn include interesting continuous-time and discrete-time special cases. Moreover several possible applications, randomization designs and simulations are presented.

## 2.1 A General Semimartingale Algorithm

We state conditions, which are helpful for the investigation of almost sure convergence of $Z$ defined in (1.3).

**Assumption 2.1.1.** *Let Assumption 1.3.1 hold.*

*(A)* *Lipschitz condition for $\tilde{f}_1(.)$: There exists a constant $L$ such that for all $z_1, z_2 \in \mathbb{R}^d$,*

$$\|\tilde{f}_1(z_1) - \tilde{f}_1(z_2)\| \leqslant L\|z_1 - z_2\|.$$

*(B)* *There exists a unique point $z^*$ such that $\nabla_c F(\mathtt{d}, 0, z^*) = 0$ for all $\mathtt{d} \in \mathbb{R}^d$.*

*(C)*

$$\underset{\varepsilon>0}{\forall} \; \underset{C(\varepsilon)>0}{\exists} \; \underset{\{z\in\mathbb{R}^d | \varepsilon \leqslant \|z-z^*\| \leqslant \frac{1}{\varepsilon}\}}{\forall} \left\langle \tilde{f}_1(z), z - z^* \right\rangle \geqslant C(\varepsilon).$$

*(D)* *The processes $(a_t)_{t\geqslant 0}$, $(c_t)_{t\geqslant 0}$ satisfy*

$$a_t, c_t > 0 \qquad\qquad a_t, c_t \downarrow 0$$
$$\int_0^\infty a_s \mathrm{d}R_s = \infty \qquad\qquad \int_0^\infty a_s c_s \mathrm{d}R_s < \infty.$$

*(E)* *If $(R_t)_{t\geqslant 0}$ is not pathwise continuous we furthermore assume that*

$$\int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s^d < \infty.$$

*(F) For all $t \geqslant 0$ let $g_1(D_t)$ have non-negative eigenvalues and $\|g_1(D_t)\|\|D_t\|$ be square integrable such that for any $t \geqslant 0$, $\mathbb{E}(g_1(D_t)) > c\mathbb{1}_d$ with $c > 0$ and*

$$\left|\int_0^\infty a_s\Big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\Big)\mathrm{d}R_s\right| < \infty,$$

$$\left|\int_0^\infty a_s c_s\Big(\|g_1(D_{s-})\|\|D_{s-}\| - \mathbb{E}(\|g_1(D_{s-})\|\|D_{s-}\|)\Big)\mathrm{d}R_s\right| < \infty,$$

$$\left|\int_0^\infty a_s^2 c_s^2\Big(\|g_1(D_{s-})\|^2\|D_{s-}\|^2 - \mathbb{E}(\|g_1(D_{s-})\|^2\|D_{s-}\|^2)\Big)\Delta R_s \mathrm{d}R_s^d\right| < \infty,$$

$$\left|\int_0^\infty a_s^2\Big(\|g_1(D_{s-})\|^2 - \mathbb{E}(\|g_1(D_{s-})\|^2)\Big)\Delta R_s \mathrm{d}R_s^d\right| < \infty.$$

*(G) Assume for all $s \in [0,\infty)$ and all $z \in \mathbb{R}^d$ that $\mathbb{E}(g_0(D_{s-}))\tilde{f}_0(z) = 0$,*

$$\left|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s}{c_s}\langle Z_{s-}, g_0(D_{s-})\tilde{f}_0(Z_{s-})\rangle \mathrm{d}R_s\right| < \infty,$$

*and*

$$\int_0^\infty \frac{a_s^2}{c_s^2}\frac{1}{1+\|Z_{s-}\|^2}\|g_0(D_{s-})\|^2\|\tilde{f}_0(Z_{s-})\|^2\Delta R_s \mathrm{d}R_s^d < \infty.$$

*(H) For every $i \in \{1,\dots,d\}$ and $x \in \mathbb{R}^d$ assume*

$$\int_0^\infty \frac{a_s^2}{c_s^2}\frac{h_s^{ii}(Z_{s-})}{1+\|Z_{s-}\|^2}\mathrm{d}R_s < \infty \text{ with } h_s^{ii}(z) := \frac{\mathrm{d}[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, z))_i]_s}{\mathrm{d}R_s}.$$

**Theorem 2.1.1.** *Let Assumption 2.1.1 be fulfilled. Then the process $(Z_t)_{t\geqslant 0}$, generated by algorithm* (1.3)*, converges almost surely to the point $z^*$ fulfilling $\tilde{f}_1(z^*) = 0$.*

*Remark* 2.1.1. Note that in this theorem, instead of $F$ we could consider an explicitly time-dependent $F_s$. This enables us to minimize more general problems. One way to handle these cases is to modify Assumption 2.1.1 such that $L$ is replaced by $L_s$ and in conditions with $a_s$, the $a_s$-terms are replaced by $a_sL_s$, where $L_s$ is a time-dependent process. However, in the following sections we restrict ourselves to time-independent functions $F$ to avoid technicalities.

*Remark* 2.1.2. As

$$\int_0^t \mathrm{d}[R,R]_s = \int_0^t \mathrm{d}[R,R]_s^c + \sum_{0<s\leqslant t}(\Delta R_s)^2 = \sum_{0<s\leqslant t}(\Delta R_s)^2 = \int_0^t \Delta R_s \mathrm{d}R_s^d$$

holds true, we can write $\Delta R_s \mathrm{d}R_s^d$ instead of $\mathrm{d}[R.,R.]_s$. In the following we use this identity without explicitly mentioning it.

According to (1.4) and the affine condition (1.8),

$$\sum_{i=1}^{d}\Big[\int_0^{\cdot}\Big(F(D_{\tau-},c_\tau,Z_{\tau-})\mathrm{d}R_\tau\Big)_i\Big]_t = \int_0^t \|F(D_{s-},c_s,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d$$
$$= \int_0^t \|F(D_{s-},0,Z_{s-}+c_sD_{s-})\|^2\Delta R_s\mathrm{d}R_s^d$$

holds true.

*Remark* 2.1.3. Condition (G) is fulfilled, for example if $Z$ is time-discrete and $D$ is a i.i.d. process. Cf. the paper of Dippon [10] and the references therein.

*Remark* 2.1.4. In the two-measurement algorithm (1.6), which is investigated later, $\tilde{f}_0(z) = 0$ for any $z \in \mathbb{R}^d$ and therefore $\mathbb{E}\big(g_0(D_{s-})\big)\tilde{f}_0(z) = 0$ holds for any $z \in \mathbb{R}^d$. Hence (G) is trivially fulfilled in such cases.

*Remark* 2.1.5. Assume that $\tilde{f}_0(z)$ is sublinear in the sense that there exist positive constants $C_1, C_2$ such that $C_1(1+\|z\|) \leqslant \tilde{f}_0(z) \leqslant C_2(1+\|z\|)$. Moreover let $F(\mathtt{d},0,z) = g_0(\mathtt{d})\tilde{f}_0(z)$ such that, for all $t \geqslant 0$, $\mathbb{E}(g_0(D_t)) = 0$ and $\mathbb{E}(\|g_0(D_t)\|^2) < \infty$ hold true. Consequently the second condition of (G) can be replaced by

$$\Big|\int_0^\infty \frac{a_s^2}{c_s^2}\Big(\|g_0(D_{s-})\|^2 - \mathbb{E}\big(\|g_0(D_{s-})\|^2\big)\Big)\Delta R_s\mathrm{d}R_s^d\Big| < \infty$$

and

$$\int_0^\infty \frac{a_s^2}{c_s^2}\Delta R_s\mathrm{d}R_s^d < \infty.$$

The first condition in (G) can be replaced by additionally assuming a random time $\tau(\omega) < \infty$ such that

$$\Big|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s}{c_s}\langle Z_{s-}, g_0(D_{s-})\tilde{f}_0(Z_{s-})\rangle\mathrm{d}R_s\Big| \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\Big|\int_{\tau(\omega)}^\infty \frac{a_s}{c_s}g_0(D_{s-})\mathrm{d}R_s\Big|$$

and

$$\Big|\int_0^\infty \frac{a_s}{c_s}g_0(D_{s-})\mathrm{d}R_s\Big| < \infty.$$

*Remark* 2.1.6. Theorem 2.1.1 still holds true if the process $(D_t)_{t\geqslant 0}$ is replaced by a deterministic, periodic function and the expectation values of terms including $D_s$ are substituted by the mean over the period. In this case with a sublinear $\tilde{f}_0(z)$ as in the previous remark, we can employ the bound

$$\Big|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s}{c_s}\langle Z_{s-}, g_0(D_{s-})\tilde{f}_0(Z_{s-})\rangle\mathrm{d}R_s\Big| \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\Big|\int_0^\infty \frac{a_s}{c_s}k_s g_0(D_{s-})\mathrm{d}R_s\Big| \tag{2.1}$$

with a predictable $(k_t)_{t\geqslant 0}$ and $|k_t|$ bounded for all $t \geqslant 0$. Such a $k$ is tolerable due to the fact that we can find deterministic functions $D$ such that $\|\int_0^\infty D_s \mathrm{d}R_s\|^2 = \mathcal{O}(1)$. For many random processes $D$, useful for our purposes, we typically cannot achieve rates better than $\mathbb{E}\|\int_0^t D_{s-} - \mathbb{E}(D_{s-})\mathrm{d}R_s\|^2 = \mathcal{O}(R_t)$. However we also create processes $D$ with rate $\mathbb{E}\|\int_0^t D_{s-} - \mathbb{E}(D_{s-})\mathrm{d}R_s\|^2 = \mathcal{O}(1)$. More details are given in the examples of the following sections.

*Remark* 2.1.7. Assumption (H) is fulfilled if $h^{ii}(x) \leqslant C_s^i(1+\|x\|^2)$ with $\int_0^\infty \frac{a_s^2}{c_s^2} C_s^i \mathrm{d}R_s < \infty$, or if $h_s^{ii}(x) \leqslant \mathcal{C}$ as well as $\int_0^\infty \frac{a_s^2}{c_s^2}\mathrm{d}R_s < \infty$ hold.

*Proof of Theorem 2.1.1.* Without loss of generality let $z^* = 0$. We consider the stochastic integral equation

$$Z_t = Z_0 - \int_0^t \frac{a_s}{c_s} F(D_{s-}, c_s, Z_{s-})\mathrm{d}R_s - \int_0^t \frac{a_s}{c_s} M(\mathrm{d}s, D_{s-}, Z_{s-}).$$

The main idea of this proof is to bound $X_t := \|Z_t\|^2 = \langle Z_t, Z_t\rangle$ by $A_t^1 - A_t^2 + \widetilde{M}$, with predictable, increasing processes $A^1$, $A^2$ and a local martingale $\widetilde{M}$. Lemma A.1.1 applied to $A^1$ yields convergence of $X$. Convergence of $X$ to 0 follows by investigation of $A^2$ and a contradiction to the assertion of the same lemma.

Integration by parts yields

$$\begin{aligned}
\mathrm{d}\langle Z_s, Z_s\rangle = &-2\frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-}, c_s, Z_{s-})\Big\rangle \mathrm{d}R_s - 2\frac{a_s}{c_s}\Big\langle Z_{s-}, M(\mathrm{d}s, D_{s-}, Z_{s-})\Big\rangle \\
&+ \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\Big[\int_0^\cdot (F(D_{\tau-}, c_\tau, Z_{\tau-})\mathrm{d}R_\tau)_i\Big]_s \\
&+ 2\frac{a_s^2}{c_s^2}\sum_{i=1}^d (F(D_{s-}, c_\tau, Z_{s-}))_i \Delta R_s (M(\mathrm{d}s, D_{s-}, Z_{s-}))_i \\
&+ \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\Big[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\Big]_s.
\end{aligned}$$

This can be decomposed as $\int_0^t \mathrm{d}\langle Z_s, Z_s\rangle = \int_0^t \mathrm{d}A_s + \int_0^t \mathrm{d}\widetilde{M}_s$, with

$$\begin{aligned}
A_t - A_0 :=& -2\int_0^t \frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-}, c_s, Z_{s-})\Big\rangle \mathrm{d}R_s + \int_0^t \frac{a_s^2}{c_s^2}\|F(D_{s-}, c_s, Z_{s-})\|^2 \Delta R_s \mathrm{d}R_s^d \\
&+ \int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i]_s \\
\widetilde{M}_t - \widetilde{M}_0 :=& +2\int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d (F(D_{s-}, c_\tau, Z_{s-}))_i \Delta R_s (M(\mathrm{d}s, D_{s-}, Z_{s-}))_i \\
&- 2\int_0^t \frac{a_s}{c_s}\Big\langle Z_{s-}, M(\mathrm{d}s, D_{s-}, Z_{s-})\Big\rangle \\
&+ \int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\left([\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i]_s - \lceil\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\rceil_s\right)
\end{aligned} \tag{2.2}$$

where $(A_t)_{t\geqslant 0} \in \mathcal{V} \cap \mathcal{P}$ as $F$ is predictable and due to the definition of the predictable quadratic variation. The first and second term in the definition of $\widetilde{M}_t$ are in $\mathcal{M}_{\text{loc}}$ as the integrands are predictable and the integrators are local martingales. By definition of the compensator, the third term in the definition of $\widetilde{M}_t$ is in $\mathcal{M}_{\text{loc}}$.

We employ the Lipschitz condition (A) to show that

$$\begin{aligned}
&\Big\|\frac{1}{c_s}\Big(F(D_{s-},c_s,Z_{s-}) - c_s\nabla_c F(D_{s-},0,Z_{s-}) - F(D_{s-},0,Z_{s-})\Big)\Big\| \\
&\quad= \Big\|\frac{1}{c_s}\Big(\int_0^1 c_s\nabla_c F(D_{s-},tc_s,Z_{s-})\mathrm{d}t + F(D_{s-},0,Z_{s-}) - c_s\nabla_c F(D_{s-},0,Z_{s-}) \\
&\qquad - F(D_{s-},0,Z_{s-})\Big)\Big\| \\
&\quad= \Big\|\frac{1}{c_s}\Big(\int_0^1 c_s\nabla_c F(D_{s-},tc_s,Z_{s-})\mathrm{d}t - c_s\nabla_c F(D_{s-},0,Z_{s-})\Big)\Big\| \\
&\quad\leqslant \int_0^1 \Big\|g_1(D_{s-})\big(\tilde{f}_1(Z_{s-} + tc_sD_{s-}) - \tilde{f}_1(Z_{s-})\big)\Big\|\mathrm{d}t \\
&\quad\leqslant \|g_1(D_{s-})\|\|D_{s-}\|c_sL
\end{aligned} \tag{2.3}$$

and analogously

$$\begin{aligned}
&\Big\|\frac{1}{c_s}F(D_{s-},c_s,Z_{s-})\Big\|^2 = \Big\|\frac{1}{c_s}F(D_{s-},c_s,Z_{s-}) - \nabla_c F(D_{s-},0,Z_{s-}) + \nabla_c F(D_{s-},0,Z_{s-})\Big\|^2 \\
&= \Big\|\int_0^1 \nabla_c F(D_{s-},tc_s,Z_{s-}) - \nabla_c F(D_{s-},0,Z_{s-})\mathrm{d}t + \frac{1}{c_s}F(D_{s-},0,Z_{s-}) \\
&\quad + \nabla_c F(D_{s-},0,Z_{s-})\Big\|^2 \\
&\leqslant 3\|\int_0^1 \nabla_c F(D_{s-},tc_s,Z_{s-}) - \nabla_c F(D_{s-},0,Z_{s-})\mathrm{d}t\|^2 + \frac{3}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2 \\
&\quad + 3\|\nabla_c F(D_{s-},0,Z_{s-})\|^2 \\
&\leqslant 3\int_0^1 \|g_1(D_{s-})\big(\tilde{f}_1(Z_{s-} + tc_sD_{s-}) - \tilde{f}_1(Z_{s-})\big)\|^2\mathrm{d}t + \frac{3}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2 \\
&\quad + 3\|\nabla_c F(D_{s-},0,Z_{s-})\|^2 \\
&\leqslant 3\|g_1(D_{s-})\|^2\|D_{s-}\|^2\int_0^1 L^2|tc_s|^2\mathrm{d}t + \frac{3}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2 + 3\|\nabla_c F(D_{s-},0,Z_{s-})\|^2 \\
&\leqslant \|g_1(D_{s-})\|^2\|D_{s-}\|^2c_s^2L^2 + \frac{3}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2 + 3\|\nabla_c F(D_{s-},0,Z_{s-})\|^2
\end{aligned} \tag{2.4}$$

hold. From (2.3) we conclude

$$\begin{aligned}
&-2\frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},c_s,Z_{s-})\Big\rangle \\
&\quad= -2\frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},c_s,Z_{s-}) - c_s\nabla_c F(D_{s-},0,Z_{s-}) + c_s\nabla_c F(D_{s-},0,Z_{s-}) \\
&\qquad - F(D_{s-},0,Z_{s-}) + F(D_{s-},0,Z_{s-})\Big\rangle
\end{aligned}$$

$$
\begin{aligned}
&= -2a_s\Big\langle Z_{s-}, \nabla_c F(D_{s-},0,Z_{s-})\Big\rangle - 2\frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},0,Z_{s-})\Big\rangle \\
&\quad - 2\frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},c_s,Z_{s-}) - c_s\nabla_c F(D_{s-},0,Z_{s-}) - F(D_{s-},0,Z_{s-})\Big\rangle \\
&\leqslant -2a_s\Big\langle Z_{s-}, \nabla_c F(D_{s-},0,Z_{s-})\Big\rangle - 2\frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},0,Z_{s-})\Big\rangle \\
&\quad + 2\|g_1(D_{s-})\|\|D_{s-}\|La_sc_s\|Z_{s-}\|. \qquad (2.5)
\end{aligned}
$$

Polarisation identity, (2.5) and (2.4) yield

$$
\begin{aligned}
&A_t - A_0 \\
&\quad= -2\int_0^t \frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},c_s,Z_{s-})\Big\rangle \mathrm{d}R_s + \int_0^t \frac{a_s^2}{c_s^2}\|F(D_{s-},c_s,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d \\
&\qquad + \int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\Big[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\Big]_s \\
&\quad\leqslant -2\int_0^t a_s\Big\langle Z_{s-}, \nabla_c F(D_{s-},0,Z_{s-})\Big\rangle \mathrm{d}R_s - 2\int_0^t \frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},0,Z_{s-})\Big\rangle \mathrm{d}R_s \\
&\qquad + 2\int_0^t \|g_1(D_{s-})\|\|D_{s-}\|La_sc_s\|Z_{s-}\|\mathrm{d}R_s + \int_0^t a_s^2\|g_1(D_{s-})\|^2\|D_{s-}\|^2c_s^2L^2\Delta R_s\mathrm{d}R_s^d \\
&\qquad + 3\int_0^t \frac{a_s^2}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d + 3\int_0^t a_s^2\|\nabla_c F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d \\
&\qquad + \int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\Big[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\Big]_s \\
&\quad\leqslant 2\int_0^t \|g_1(D_{s-})\|\|D_{s-}\|La_sc_s\|Z_{s-}\|\mathrm{d}R_s + 2\Big|\int_0^t \frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},0,Z_{s-})\Big\rangle \mathrm{d}R_s\Big| \\
&\qquad - 2\int_0^t a_s\Big\langle Z_{s-}, \nabla_c F(D_{s-},0,Z_{s-})\Big\rangle \mathrm{d}R_s \\
&\qquad + \int_0^t a_s^2\|g_1(D_{s-})\|^2\|D_{s-}\|^2c_s^2L^2\Delta R_s\mathrm{d}R_s^d + 3\int_0^t \frac{a_s^2}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d \\
&\qquad + 3\int_0^t a_s^2\|\nabla_c F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d + \int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\Big[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\Big]_s \\
&\quad\leqslant 2\int_0^t \|g_1(D_{s-})\|\|D_{s-}\|La_sc_s\|Z_{s-}\|\mathrm{d}R_s \\
&\qquad + 2\Big|\int_0^t \frac{a_s}{c_s}\Big\langle Z_{s-}, F(D_{s-},0,Z_{s-})\Big\rangle \mathrm{d}R_s\Big| - 2\int_0^t a_s\Big\langle Z_{s-}, g_1(D_{s-})\tilde{f}_1(Z_{s-})\Big\rangle \mathrm{d}R_s \\
&\qquad + \int_0^t a_s^2\|g_1(D_{s-})\|^2\|D_{s-}\|^2c_s^2L^2\Delta R_s\mathrm{d}R_s^d + 3\int_0^t \frac{a_s^2}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d \\
&\qquad + 3\int_0^t a_s^2\|\nabla_c F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d \\
&\qquad + \int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\Big[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\Big]_s.
\end{aligned}
$$

We decompose the right side of the previous inequality into

$$\begin{aligned}
A_t^1 - A_0^1 := & \int_0^t 2a_s\|g_1(D_{s-})\|\|D_{s-}\|\|Z_{s-}\|c_sL\mathrm{d}R_s + 2\Big|\int_0^t \frac{a_s}{c_s}\langle Z_{s-}, F(D_{s-},0,Z_{s-})\rangle \mathrm{d}R_s\Big| \\
& + \int_0^t \|g_1(D_{s-})\|^2\|D_{s-}\|^2 a_s^2c_s^2L^2\Delta R_s\mathrm{d}R_s^d \\
& + 3\int_0^t \frac{a_s^2}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d \\
& + 3\int_0^t a_s^2\|\nabla_c F(D_{s-},c_s,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d \\
& + \int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\Big[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\Big]_s \\
-A_t^2 + A_0^2 := & -2\int_0^t a_s\Big\langle Z_{s-}, g_1(D_{s-})\tilde{f}_1(Z_{s-})\Big\rangle \mathrm{d}R_s,
\end{aligned}$$

such that $A \leqslant A^1 - A^2$. According to conditions (C) and (F), $A^2 \geqslant 0$ holds true. Moreover, $0 \leqslant A^2 \leqslant A^1 - A$, and hence the conditions of Lemma A.1.1 are fulfilled.

In order to make sure that $(Z_t)_{t\geqslant 0}$ converges, we show

$$\left\{\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\mathrm{d}A_s^1 < \infty\right\} = \Omega \text{ a.s.}$$

This is done by the investigation of the following terms. Assumptions (F), (D), (E) and (G) yield

$$\begin{aligned}
& 2\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s\|g_1(D_{s-})\|\|D_{s-}\|\|Z_{s-}\|c_sL\mathrm{d}R_s \\
& \quad \leqslant \mathcal{C}\int_0^\infty \|g_1(D_{s-})\|\|D_{s-}\|a_sc_sL\mathrm{d}R_s \\
& \quad \leqslant \mathcal{C}\Big|\int_0^\infty \big(\|g_1(D_{s-})\|\|D_{s-}\| - \mathbb{E}(\|g_1(D_{s-})\|\|D_{s-}\|)\big)a_sc_sL\mathrm{d}R_s\Big| \\
& \qquad + \mathcal{C}\int_0^\infty \mathbb{E}(\|g_1(D_{s-})\|\|D_{s-}\|)a_sc_sL\mathrm{d}R_s \\
& \quad < \infty, \\
& \int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\|g_1(D_{s-})\|^2\|D_{s-}\|^2a_s^2c_s^2L^2\Delta R_s\mathrm{d}R_s^d \\
& \quad \leqslant \mathcal{C}\Big|\int_0^\infty a_s^2c_s^2L^2\big(\|g_1(D_{s-})\|^2\|D_{s-}\|^2 - \mathbb{E}(\|g_1(D_{s-})\|^2\|D_{s-}\|^2)\big)\Delta R_s\mathrm{d}R_s^d\Big| \\
& \qquad + \mathcal{C}\int_0^\infty a_s^2c_s^2L^2\mathbb{E}(\|g_1(D_{s-})\|^2\|D_{s-}\|^2)\Delta R_s\mathrm{d}R_s^d < \infty
\end{aligned}$$

and

$$
\begin{aligned}
&3\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s^2}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad = 3\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s^2}{c_s^2}\|g_0(D_{s-})\|^2\|\tilde{f}_0(Z_{s-})\|^2\Delta R_s \mathrm{d}R_s < \infty.
\end{aligned}
$$

Furthermore, by assumption (G), it holds true that

$$
\begin{aligned}
&2\Big|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s}{c_s}\langle Z_{s-}, F(D_{s-},0,Z_{s-})\rangle \mathrm{d}R_s\Big| \\
&\qquad = 2\Big|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s}{c_s}\langle Z_{s-}, \big(g_0(D_{s-}) - \mathbb{E}(g_0(D_{s-}))\big)\tilde{f}_0(Z_{s-})\rangle \mathrm{d}R_s\Big| \\
&\qquad\quad + 2\Big|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s}{c_s}\langle Z_{s-}, \mathbb{E}(g_0(D_{s-}))\tilde{f}_0(Z_{s-})\rangle \mathrm{d}R_s\Big| < \infty.
\end{aligned}
$$

Using the Lipschitz continuity condition, the Cauchy-Schwarz inequality and conditions (E) and (F) we obtain

$$
\begin{aligned}
&3\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s^2\|\nabla_c F(D_{s-},0,Z_{s-})\|^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad = 3\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s^2\|\nabla_c F(D_{s-},0,Z_{s-}) - \nabla_c F(D_{s-},0,0)\|^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad = 3\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s^2\|g_1(D_{s-})\big(\tilde{f}_1(Z_{s-}) - \tilde{f}_1(0)\big)\|^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad \leqslant 3\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s^2 L^2\|g_1(D_{s-})\|^2\|Z_{s-}\|^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad \leqslant \mathcal{C}\int_0^\infty a_s^2 L^2\Big(\|g_1(D_{s-})\|^2 - \mathbb{E}(\|g_1(D_{s-})\|^2)\Big)\Delta R_s \mathrm{d}R_s^d \\
&\qquad\quad + \mathcal{C}\int_0^\infty a_s^2 L^2 \mathbb{E}(\|g_1(D_{s-})\|^2)\Delta R_s \mathrm{d}R_s^d \\
&\qquad < \infty.
\end{aligned}
$$

Finally, by condition (H),

$$
\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s^2}{c_s^2}\mathrm{d}\Big[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\Big]_s = \int_0^\infty \frac{a_s^2}{c_s^2}\frac{h_s^{ii}(Z_{s-})}{1+\|Z_{s-}\|^2}\mathrm{d}R_s < \infty
$$

holds true. Consequently it is shown that $(Z_t)_{t\geqslant 0}$ converges. We now show that $Z$ converges to the stationary point of $F$. This is proven by contradiction. It is already shown that

$$
\Omega = \{\|Z_s\|^2 \to\} \cap \{A_\infty^2 < \infty\}.
$$

However, as the smallest eigenvalue of $g_1(D)$ is not strictly greater than zero we cannot employ the previous decomposition $A \leqslant A^1 - A^2$ and especially the term $A^2$

to establish a contradiction. For that purpose assume that there exists a set $N \subset \Omega$ of non-zero probability, such that for all its elements our stochastic integral equation does not converge to the stationary point. The following considerations are done on this set $N$. This means, we assume that $Z$ does not converge to zero. Now we choose another decomposition with an $A^2$ which is more useful. Let

$$\begin{aligned}
A_t^1 - A_0^1 :=& \int_0^t 2a_s\|g_1(D_{s-})\|\|D_{s-}\|\|Z_{s-}\|c_s L \mathrm{d}R_s + 2\Big|\int_0^t \frac{a_s}{c_s}\langle Z_{s-}, F(D_{s-},0,Z_{s-})\rangle \mathrm{d}R_s\Big| \\
&+ 2\Big|\int_0^t a_s\langle Z_{s-}, \nabla_c F(D_{s-},0,Z_{s-}) - \mathbb{E}(g_1(D_{s-}))\tilde{f}_1(Z_{s-})\rangle \mathrm{d}R_s\Big| \\
&+ \int_0^t \|g_1(D_{s-})\|^2\|D_{s-}\|^2 a_s^2 c_s^2 L^2 \Delta R_s \mathrm{d}R_s^d \\
&+ 3\int_0^t \frac{a_s^2}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2 \Delta R_s \mathrm{d}R_s^d \\
&+ 3\int_0^t a_s^2\|\nabla_c F(D_{s-},c_s,Z_{s-})\|^2 \Delta R_s \mathrm{d}R_s^d \\
&+ \int_0^t \frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i]_s \\
-A_t^2 + A_0^2 :=& -2\int_0^t a_s\Big\langle Z_{s-}, \mathbb{E}(g_1(D_{s-}))\tilde{f}_1(Z_{s-})\Big\rangle \mathrm{d}R_s,
\end{aligned}$$

and note again that $A \leqslant A^1 - A^2$ holds true. On the contrary to the previous part of the proof we can now utilize that $Z$ converges. By assumption on $N$ for almost all $\omega \in N$

$$\underset{\epsilon^*>0}{\exists}\ \underset{s_0}{\exists}\ \underset{s\geqslant s_0}{\forall}\ \epsilon^* \leqslant \|Z_s\| \leqslant 1/\epsilon^*.$$

We imply this property to bound the following term:

$$\begin{aligned}
2\Big|&\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s\langle Z_{s-}, \nabla_c F(D_{s-},0,Z_{s-}) - \mathbb{E}(g_1(D_{s-}))\tilde{f}_1(Z_{s-})\rangle \mathrm{d}R_s\Big| \\
&= 2\Big|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s\langle Z_{s-}, g_1(D_{s-})\tilde{f}_1(Z_{s-}) - \mathbb{E}(g_1(D_{s-}))\tilde{f}_1(Z_{s-})\rangle \mathrm{d}R_s\Big| \\
&= 2\Big|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s \sum_{i=1}^d Z_{s-}^{(i)}\Big(\sum_{j=1}^d \Big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\Big)^{(ij)} \tilde{f}_1^{(j)}(Z_{s-})\Big)\mathrm{d}R_s\Big| \\
&\leqslant 4\Big|\int_0^{s_0} \frac{1}{1+\|Z_{s-}\|^2} a_s \sum_{i=1}^d Z_{s-}^{(i)}\Big(\sum_{j=1}^d \Big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\Big)^{(ij)} \tilde{f}_1^{(j)}(Z_{s-})\Big)\mathrm{d}R_s\Big| \\
&\quad + 4\Big|\int_{s_0+}^\infty \frac{1}{1+\|Z_{s-}\|^2} a_s \sum_{i=1}^d Z_{s-}^{(i)}\Big(\sum_{j=1}^d \Big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\Big)^{(ij)} \tilde{f}_1^{(j)}(Z_{s-})\Big)\mathrm{d}R_s\Big| \\
&\leqslant \mathcal{C}(\omega) + \mathcal{C}\Big|\int_{s_0+}^\infty a_s\Big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\Big)\mathrm{d}R_s\Big| < \infty.
\end{aligned}$$

All other terms in the expansion of $\int_0^\infty (1 + \|Z_{s-}\|^2)^{-1}\mathrm{d}A_t^1$ have already been handled before. Now also for the new decomposition it is shown that

$$\Omega = \{\|Z_s\|^2 \to\} \cap \{A_\infty^2 < \infty\}.$$

Now the set $N$ is used to find a contradiction to the fact that $\Omega \subset \{A_\infty^2 < \infty\}$ holds. Note that

$$A_\infty^2 = \int_0^\infty \mathrm{d}A_s^2 + A_0^2 = 2\int_0^\infty a_s \Big\langle Z_{s-}, \mathbb{E}(g_1(D_{s-}))\tilde{f}_1(Z_{s-})\Big\rangle \mathrm{d}R_s + A_0^2.$$

We already know that $Z$ converges for almost all $\omega \in \Omega$, but for all $\omega \in N$ its limit is not 0. Recall that for almost all $\omega \in N$

$$\underset{\epsilon^*>0}{\exists}\ \underset{s_0}{\exists}\ \underset{s \geqslant s_0}{\forall}\ \epsilon^* \leqslant \|Z_s\| \leqslant 1/\epsilon^*.$$

Therefore for almost all $\omega \in N$ it follows that

$$\begin{aligned} A_\infty^2 &= 2\int_0^\infty a_s \Big\langle Z_{s-}, \mathbb{E}(g_1(D_{s-}))\tilde{f}_1(Z_{s-})\Big\rangle \mathrm{d}R_s + A_0^2 \\ &= 2\int_0^{s_0} a_s \Big\langle Z_{s-}, \mathbb{E}(g_1(D_{s-}))\tilde{f}_1(Z_{s-})\Big\rangle \mathrm{d}R_s \\ &\quad + 2\int_{s_0+}^\infty a_s \Big\langle Z_{s-}, \mathbb{E}(g_1(D_{s-}))\tilde{f}_1(Z_{s-})\Big\rangle \mathrm{d}R_s + A_0^2 \\ &\geqslant \mathcal{C}(\omega) + 2C(\epsilon^*)\int_{s_0+}^\infty a_s \mathrm{d}R_s = \infty \end{aligned}$$

which is a contradiction to $\Omega = \{A_\infty^2 < \infty\}$. Consequently such a set $N$ cannot exist, and $Z_t \xrightarrow{t\to\infty} 0$ is proven. Due to the convexity-type condition (C) this stationary point is a minimizer of $\tilde{f}_0$. □

## 2.2 Algorithms Using Kernel-Based Gradient Estimates

Now we state assumptions for algorithms of the one- and two-measurement forms (1.5) and (1.6), respectively. For that purpose we formulate corresponding specialized conditions of Assumption 2.1.1.

**Assumption 2.2.1.** *Let conditions (D) and (E) from Assumption 2.1.1 hold.*

*(kA) The gradient of $f\colon \mathbb{R}^d \to \mathbb{R}$ satisfies the Lipschitz condition*

$$\underset{x,y\in\mathbb{R}^d}{\forall}\ \|\nabla f(x) - \nabla f(y)\| \leqslant L\|x-y\|,$$

*with a constant $L$.*

*(kB) There exists a $z^*$ such that $\nabla f(z^*) = 0$.*

(kC) *The gradient at $z^*$ satisfies the following condition:*

$$\underset{\varepsilon>0}{\forall}\ \underset{C(\varepsilon)>0}{\exists}\ \underset{\{z\in\mathbb{R}^d|\varepsilon\leqslant\|z-z^*\|\leqslant\frac{1}{\varepsilon}\}}{\forall} \left\langle\nabla f(z), z-z^*\right\rangle \geqslant C(\varepsilon)$$

(kF) *For all $s\in[0,\infty)$ let*

$$\mathbb{E}(K(D_{s-})\otimes D_{s-}) = \mathbb{1}_d, \quad \mathbb{E}(\|D_{s-}\|^4\|K(D_{s-})\|^2) < \infty \text{ and } \mathbb{E}\left(\|K(D_{s-})\|^2\right) < \infty.$$

*Moreover*

$$\Big|\int_0^\infty a_s\Big(K(D_{s-})\otimes D_{s-} - \mathbb{E}(K(D_{s-})\otimes D_{s-})\Big)\mathrm{d}R_s\Big| < \infty$$
$$\Big|\int_0^\infty a_sc_s\Big(\|K(D_{s-})\|\|D_{s-}\|^2 - \mathbb{E}(\|K(D_{s-})\|\|D_{s-}\|^2)\Big)\mathrm{d}R_s\Big| < \infty$$
$$\Big|\int_0^\infty a_s^2c_s^2\Big(\|D_{s-}\|^4\|K(D_{s-})\|^2 - \mathbb{E}(\|D_{s-}\|^4\|K(D_{s-})\|^2)\Big)\Delta R_s\mathrm{d}R_s^d\Big| < \infty$$
$$\Big|\int_0^\infty a_s^2\Big(\|D_{s-}\|^2\|K(D_{s-})\|^2 - \mathbb{E}(\|D_{s-}\|^2\|K(D_{s-})\|^2)\Big)\Delta R_s\mathrm{d}R_s^d\Big| < \infty.$$

(kG) *For the one-measurement algorithm* (1.5) *we assume there exists a random time $\tau(\omega)<\infty$ such that*

$$\begin{aligned}\Big|\int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s}{c_s}\big\langle Z_{s-}, K(D_{s-})f(Z_{s-})\big\rangle\mathrm{d}R_s\Big|&\\ \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\Big|\int_{\tau(\epsilon)}^\infty \frac{a_s}{c_s}\sum_{i=1}^d K(D_{s-})^{(i)}\mathrm{d}R_s\Big|&,\end{aligned}$$

$$\Big|\int_0^\infty \frac{a_s}{c_s}\sum_{i=1}^d\big(K(D_{s-})^{(i)} - \mathbb{E}(K(D_{s-}))^{(i)}\big)\mathrm{d}R_s\Big| < \infty,$$

*for all $s\geqslant 0$, $\mathbb{E}(K(D_{s-}))=0$,*

$$\underset{C_1,C_2\geqslant 0}{\exists}\ \underset{z\in\mathbb{R}^d}{\forall} C_1(1+\|z\|) \leqslant f(z) \leqslant C_2(1+\|z\|),$$

$$\Big|\int_0^\infty \frac{a_s^2}{c_s^2}\Big(\|K(D_{s-})\|^2 - \mathbb{E}(\|K(D_{s-})\|^2)\Big)\Delta R_s\mathrm{d}R_s^d\Big| < \infty,$$

*and*

$$\int_0^\infty \frac{a_s^2}{c_s^2}\Delta R_s\mathrm{d}R_s^d < \infty.$$

*(kH)* *For every* $i \in \{1, \dots, d\}$ *and* $z \in \mathbb{R}^d$

$$\int_0^\infty \frac{a_s^2}{c_s^2} \frac{h_s^{ii}(Z_{s-})}{1 + \|Z_{s-}\|^2} \mathrm{d}R_s < \infty \text{ with } h_s^{ii}(z) := \frac{\mathrm{d}[\int_0^\cdot (K(D_{\tau-})\breve{M}(\mathrm{d}\tau, z))_i]_s}{\mathrm{d}R_s},$$

*where* $\breve{M}$ *was introduced on page 8.*

**Corollary 2.2.1.** *Let Assumption 2.2.1 be fulfilled. Then the process* $(Z_t)_{t \geqslant 0}$, *generated by the one- or two-measurement algorithms* (1.5) *or* (1.6) *converges almost surely to the minimizing point of* $f$.

*Proof of Corollary 2.2.1.* We show that the assertion is a consequence of Theorem 2.1.1. Set

$$\begin{aligned} &F(D_{s-}, c_s, Z_{s-}) \\ &\quad = \begin{cases} K(D_{s-}) f(Z_{s-} + c_s D_{s-}) & \text{for algorithm (1.5)} \\ \frac{1}{2} K(D_{s-})\{f(Z_{s-} + c_s D_{s-}) - f(Z_{s-} - c_s D_{s-})\} & \text{for algorithm (1.6).} \end{cases} \end{aligned}$$

We obtain

$$\begin{aligned} &\nabla_c F(D_{s-}, c_s, Z_{s-}) \\ &\quad = \begin{cases} \big(K(D_{s-}) \otimes D_{s-}\big) \nabla f(Z_{s-} + c_s D_{s-}) & \text{for alg. (1.5)} \\ \frac{1}{2}\big(K(D_{s-}) \otimes D_{s-}\big)\{\nabla f(Z_{s-} + c_s D_{s-}) + \nabla f(Z_{s-} - c_s D_{s-})\} & \text{for alg. (1.6).} \end{cases} \end{aligned}$$

We show that Assumption 2.2.1 implies conditions (A), (B), (C), (F), (G) and (H) in Assumption 2.1.1. Note that according to the factorizing and the affine condition

$$\nabla_c F(\mathtt{d}, c, z) = \nabla_c F(\mathtt{d}, 0, z + c\mathtt{d}) = g_1(\mathtt{d}) \tilde{f}_1(z + c\mathtt{d}) = (K(\mathtt{d}) \otimes \mathtt{d}) \nabla f(z + c\mathtt{d})$$

holds true for any $z \in \mathbb{R}^d$ in both algorithms (1.5) and (1.6). Consequently $\tilde{f}_1(z) = \nabla f(z)$ for any $z \in \mathbb{R}^d$. Therefore condition (kA), (kB) and (kC) imply (A), (B) and (C), respectively. Choosing $g_1(\mathtt{d}) := \big(K(\mathtt{d}) \otimes \mathtt{d}\big)$ yields (F) from (kF). With $g_0(\mathtt{d}) := K(\mathtt{d})$ the first part of condition (G) follows by

$$F(D_{s-}, 0, x) = g_0(D_{s-}) \tilde{f}_0(x) = \begin{cases} K(D_{s-}) f(x) & \text{for algorithm (1.5)} \\ 0 & \text{for algorithm (1.6),} \end{cases}$$

such that it is trivially fulfilled for (1.6). For (1.5) we get

$$\begin{aligned} &\int_0^\infty \frac{1}{1 + \|Z_{s-}\|^2} \frac{a_s}{c_s} \big\langle Z_{s-}, g_0(D_{s-}) \tilde{f}_0(Z_{s-}) \big\rangle \mathrm{d}R_s \\ &\qquad = \int_0^\infty \frac{1}{1 + \|Z_{s-}\|^2} \frac{a_s}{c_s} \big\langle Z_{s-}, K(D_{s-}) f(Z_{s-}) \big\rangle \mathrm{d}R_s \\ &\qquad \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega) \Big| \int_0^\infty \frac{a_s}{c_s} \sum_{i=1}^d K(D_{s-})^{(i)} \mathrm{d}R_s \Big| \end{aligned}$$

$$
\begin{aligned}
&\leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\Big|\int_0^\infty \frac{a_s}{c_s}\sum_{i=1}^d \big(K(D_{s-})^{(i)} - \mathbb{E}(K(D_{s-})^{(i)})\big)\mathrm{d}R_s\Big| \\
&< \infty.
\end{aligned}
$$

To show the second assumption of (G) we note that

$$
\|F(D_{s-},0,x)\|^2 = \begin{cases} \|K(D_{s-})\|^2\|f(x)\|^2 & \text{for algorithm (1.5)} \\ 0 & \text{for algorithm (1.6)} \end{cases}
$$

and apply a Taylor expansion and the sublinearity condition form (kG) to prove

$$
\|f(z)\|^2 \leqslant \mathcal{C}(1+\|z\|)^2 \leqslant \mathcal{C}(1+\|z\|^2).
$$

Then

$$
\begin{aligned}
\int_0^\infty & \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s^2}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2\Delta R_s \mathrm{d}R_s^d \\
&= \int_0^\infty \frac{1}{1+\|Z_{s-}\|^2}\frac{a_s^2}{c_s^2}\|K(D_{s-})\|^2\|f(Z_{s-})\|^2\Delta R_s \mathrm{d}R_s^d \\
&\leqslant \mathcal{C}\int_0^\infty \frac{a_s^2}{c_s^2}\|K(D_{s-})\|^2\Delta R_s \mathrm{d}R_s^d \\
&\leqslant \mathcal{C}\Big|\int_0^\infty \frac{a_s^2}{c_s^2}\Big(\|K(D_{s-})\|^2 - \mathbb{E}(\|K(D_{s-})\|^2)\Big)\Delta R_s \mathrm{d}R_s^d\Big| \\
&\quad + \mathcal{C}\int_0^\infty \frac{a_s^2}{c_s^2}\mathbb{E}(\|K(D_{s-})\|^2)\Big)\Delta R_s \mathrm{d}R_s^d < \infty.
\end{aligned}
$$

Finally choosing $M(\mathrm{d}s, D_{s-}, x) := K(D_{s-})\check{M}(\mathrm{d}s, x)$ assumption (kH) yields (H). $\square$

*Remark* 2.2.1. It is worth mentioning that consistency of the classical Kiefer-Wolfowitz algorithm follows from Theorem 2.1.1 by setting $g_0(D_{s-}) = \vec{1}$, $g_1(D_{s-}) = \mathbb{1}_d$ and

$$
\tilde{f}_k(Z_{s-}) = \begin{cases} 0 & \text{if } k \in \{0,2\} \\ \nabla^k f(Z_{s-}) & \text{if } k = 1. \end{cases}
$$

### 2.2.1 An Application in Analog Computing

In current machine learning applications deep learning architectures are performing extremely well [16]. However computation increases as dimensionality of the input space increases. In such designs the computation is performed concurrently, CPUs however run sequentially. To overcome this fact, GPUs are widely used. But these are power hungry as well. The following examples show how analog computing offers an alternative.

Analog VLSI (very large scale integration) implementations are working in a slow but massively parallel fashion. Moreover they are tolerant with respect to inaccuracies, while digital computers only accept two states and ignore values in the middle (to achieve noise immunity).

Off-chip learning is effective as long as training is performed in the loop. However, I/O bandwidth limitations make systems with a large number of weight parameters impractical. On-chip learning on the other hand provide autonomous, self-contained systems which are able to adapt continuously in the environment they are operating in. A detailed description can be found in a paper of Cauwenberghs [5].

Backpropagation, definitely the most common way to train neural networks, can have vanishing or exploding gradients when training classical recurrent neural networks. Hence it is better to either use an advanced recurrent neural network architecture, such as LSTM (long short-term memory), or to optimize with another algorithm. In the following, we see how randomized stochastic approximation can be used for such problems.

Cauwenberghs [4] considered a recurrent neural network with continuous-time dynamics:

$$\tau \frac{\mathrm{d}}{\mathrm{d}t} x_{i,t} = -x_{i,t} + \sum_{j=1}^{6} w_{ij}^{(n)} \sigma(x_{j,t}(t) - \theta_j^{(n)}) + \epsilon_{i,t} \ , \quad t \in (t_n, t_{n+1}) \ , \ i \in \{1, \ldots, 6\},$$

with unknown states and thresholds $w_{ij}$ and $\theta_j$, observable neuron state variables $x_t$ and external inputs $\epsilon_t$, the sigmoid activation function $\sigma$ and a fixed time constant $\tau$ governing the dynamics. Hence 36 weights $w_{ij}$ and 6 thresholds $\theta_j$ are to be estimated. The usage of external input is not essential in this model. Hence we could choose $\epsilon$ to be zero. However one can use $\epsilon$ as a teacher forcing signal $\epsilon_{i,t} := \lambda\gamma(x_{i,t}^{\mathrm{tar}} - x_{i,t})$ with a constant $\lambda$, a symmetrical and monotonically increasing function $\gamma \colon \mathbb{R} \to \mathbb{R}$ and $x_t^{\mathrm{tar}}$ the target output signal. Let $Z$ be a 42-dimensional vector, representing the weights and thresholds. We are looking for a value $z^*$ such that $\{x_t(z^*) \colon t \in [0, \infty)\}$ tracks $\{x_t^{\mathrm{tar}}(z^*) \colon t \in [0, \infty)\}$ the best. However at each iteration step $n$ we only have information on $x_t$ at a single point $t$ in time and not for the complete time interval. Due to this lack of information, we choose a loss function $\mathcal{L}_n(Z_n)$ which can be considered as a noisy observation of the path of $\|x_s^{\mathrm{tar}} - x_s(Z_n)\|^2_{s \in [0, t_{n+1})}$. Then the loss function to be minimized in a discrete-time setting is $\mathcal{L}_n(Z_n) := \|x_{t_{n+1}}^{\mathrm{tar}} - x_{t_{n+1}}(Z_n)\|^2$. This leads to the learning algorithm

$$Z_{n+1} = Z_n - D_n \frac{a_n}{2c_n} \Big( \mathcal{L}_n(Z_n + c_n D_n) - \mathcal{L}_n(Z_n - c_n D_n) \Big).$$

Cauwenberghs successfully implemented this model on an analog VLSI and used it for tracking a circular target trajectory.

In Maeda's and Wakamura's paper [26] recurrent neural networks with a simultaneous perturbation learning scheme are considered. In contrast to ordinary correlation rules, this method can be applied to analog learning and the learning of oscillatory solutions of recurrent neural networks. They considered the implementation of a Hopfield neural network with a field-programmable gate array (FPGA). With examples of Hopfield neural networks for analog and for oscillatory targets they showed the feasibility of such a learning scheme. Moreover it can be used for trajectory learning. Usually backpropagation through time (BPTT) is employed to propagate an error quantity through time from the current state to a state which is several time steps in the past. Such a procedure is complicated as it takes a long time to compute the modifying quantities that correspond to all weights. Moreover its realization on a hardware system is hard as well. This is where simultaneous perturbation has its benefits. It is easy to implement and the modification of all weights can be done without the need of a complicated error propagation through time. Unlike Hebbian learning, it can also be applied to analog problems. Furthermore it is not necessary to work with an energy function.

In contrast to digital computation, analog neural circuits require a smaller number of elements and less power consumption. Moreover parallel information processing is possible and therefore high-speed operation can be expected. Additional information can be found in a paper of Maeda and Kusuhashi [25].

Time-continuous stochastic approximation methods fit perfectly for such models, as no artificial discretization has to be done. Then the optimization problem to be solved is

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s} D_{s-} \big\{ \mathcal{L}_s(Z_{s-} + c_s D_{s-}) - \mathcal{L}_s(Z_{s-} - c_s D_{s-}) \big\} \mathrm{d}R_s.$$

Some recent analog computing applications can be worth considering. Sarpeshkar et al. [9] showed how the behaviour of genetic circuits can be modelled by analog circuits. They used analog electronic circuits to model interactions between proteins and DNA in a cell with a remarkable accuracy. When treated as an analog device, one single transistor has an infinite number of possible conductivities. However when it is treated as a binary switch, there are only two possible states. Hence in such a case, one would need a large number of transistors to model a large number of concentrations.

In another paper [8] the opposite thing was done: Bacterial cells have been transformed into living calculators.

Recently Sarpeshkar et al. [1] suggested a compiler which enables faster programming of analog devices instead of programming by hand. Differential equations can be translated into current flows and voltages. The laws of physics yield that the voltages and currents across an analog circuit will balance out. When the variables in a set of differential equations are encoded by those voltages and currents, then varying one will also vary the others. This is in contrast to the inner workings of a digital circuit. There time has to be split into a huge amount of intervals and the equations have to be solved in each of these intervals. Moreover a transistor in such a circuit can only represent one of two possible values instead of a continuous range of values.

Hence analog computers seem to be a good application of time-continuous randomized stochastic approximation algorithms. This motivates the deduction of time-continuous special cases of our original randomized stochastic approximation algorithms 1.5 and 1.6, which will be done in subsection 2.2.3.

### 2.2.2 An Application in Model-Free Control

In 1998 Spall and Cristion [41] considered the problem of developing controllers for a nonlinear, stochastic system whose equations are unknown. Consider a general dynamic process $X$ which is typically involving nonlinear dynamics and stochastic effects. Instead of $X$, only a sequence of discrete-time measurements of a process $Y$, which is a noisy observation of a function of $X$, is observable. In tracking applications $Y_{k+1}$ is compared with a target value $T_{k+1}$. On the basis of this information the goal is to choose a sequence of corresponding controls $(u(Z_k))$ with $u(Z_k) := u(Z_k; Y_k, T_{k+1})$ to optimize a function of future system measurements $Y_{k+1}$ via $Z_k$ in such a way that these measurements approach the target values $T_{k+1}$ in a certain sense. The structure of $u$ is fixed but might be unknown and $Z$ represents the parameters to be optimized. Assume the system dynamics and the measurements are given by

$$\begin{aligned} X_{k+1} &= X_k + f(Z; X_k) + M_k \\ Y_k &= h(X_k) + N_k, \end{aligned}$$

with $k = 0, 1, 2, \ldots$ where $f(Z; X_k) := f(u(Z, Y_k, T_{k+1}); X_k)$ and $h$ are typically unknown nonlinear functions and $M_k$ as well as $N_k$ are noise terms. An optimal $Z$ can be found by minimizing a loss function $\mathcal{L}_k$ related to the next measurement $Y_{k+1}$, comparing $Y_{k+1}$ with a target value $T_{k+1}$. A common choice of $\mathcal{L}_k$ is the regularized least squares loss function

$$\mathcal{L}_k(Z) = (Y_{k+1} - T_{k+1})^T A (Y_{k+1} - T_{k+1}) + u(Z)^T B u(Z) \tag{2.6}$$

with positive semidefinite matrices $A$ and $B$ representing the weight put on large deviations from the target and the cost of large values of $u(Z_k)$. The parameter $Z_k$ only affects $\mathcal{L}_k$ via $u$. Note that, except of the dependence on $Z$, the structure of $u$ is left open. We can consider it as a direct approximator given no analytical structure of the measurements. Spall and Cristion also considered more general forms of $u$ where previous measurements and controls are available as well. Besides that, they presented a model where $u$ is a neural network and $Z$ are the weights to be optimized. We seek for an optimal $Z^*$ minimizing the expectation value of $\mathcal{L}_k$. Namely

$$\frac{\partial \mathbb{E}(\mathcal{L}_k)}{\partial Z_k} = \frac{\partial u}{\partial Z_k} \frac{\partial \mathbb{E}(\mathcal{L}_k)}{\partial u} = 0 \text{ at } Z_k = Z^*$$

must hold. Note, that sometimes in control literature, the expectation value is replaced by a conditional expectation given the previous measurements and control. But it turns out, that under standard assumptions yielding the interchange of derivative and integral, both are minimized by the same $Z^*$. Namely if we denote the conditional expectation by $\mathbb{E}^*$, then $\partial \mathbb{E}^*(\mathcal{L}_k)/\partial Z_t$ implies $\mathbb{E}(\partial \mathbb{E}^*(\mathcal{L}_k)/\partial Z_t) = \partial \mathbb{E}(\mathbb{E}^*(\mathcal{L}_k))/\partial Z_t = \partial \mathbb{E}(\mathcal{L}_k)/\partial Z_t = 0$. Since $f$ and $h$ are not completely known we are unable to compute the term $\partial \mathbb{E}(\mathcal{L}_k)/\partial Z_k$ which includes $\partial h/\partial X_{k+1}$ and $\partial f/\partial u_k$. For the unregularized least squares loss function this is illustrated by

$$\frac{\partial (Y_{k+1} - T_{k+1})^2}{\partial Z_k} = 2(Y_{k+1} - T_{k+1}) \frac{\partial h}{\partial X_{k+1}} \frac{\partial f}{\partial u_k} \frac{\partial u_k}{\partial Z_k}.$$

Consequently a Robbins-Monro type algorithm cannot be applied as the derivative is unavailable. As the number of parameters could be hundreds, Spall and Cristion preferred SPSA1 and SPSA instead of the standard Kiefer-Wolfowitz algorithm. Thus their algorithms are of the form

$$Z_k = Z_{k-1} - a_k g(Z_{k-1})$$

where the $l$-th component of $g(\hat{Z}_{k-1})$, $l = \{1, \ldots, p\}$ is given by

$$g(Z_{k-1}) = \begin{cases} \frac{\mathcal{L}_k^{(+)}}{c_k D_{kl}} & \text{in the one-measurement, and} \\ \frac{\mathcal{L}_k^{(+)} - \mathcal{L}_k^{(-)}}{2c_k D_{kl}} & \text{in the two-measurement form.} \end{cases}$$

The estimates $\mathcal{L}_k^{(\pm)} := \mathcal{L}_k(Z_{k-1} \pm c_k D_k)$ make use of the observed $Y_{k+1}^{(\pm)}$ and $u^{(\pm)}(Z)$. In the tracking loss function from the beginning this means

$$\mathcal{L}_k(Z) = (Y_{k+1}^{(\pm)} - T_{k+1})^T A (Y_{k+1}^{(\pm)} - T_{k+1}) + (u^{(\pm)}(Z))^T B u^{(\pm)}(Z).$$

Assume now, that system dynamics and measurements are not time-discrete but have the following semimartingale form

$$X_t = X_0 + \int_0^t f(Z; X_{s-}) \mathrm{d}R_s + \int_0^t M(\mathrm{d}s, X_{s-}) \tag{2.7}$$

$$Y_t = h(X_t) + N_t. \tag{2.8}$$

The functions $f$ and $h$ are unknown nonlinear functions governing the system dynamics and the measurement process, respectively. Note that only $(Y_t)_{t \geqslant 0}$ is observable. Consider a randomized semimartingale stochastic approximation algorithm given by

$$Z_t = Z_0 - \int_0^t \frac{a_s}{c_s} K(D_{s-}) \big\{ \mathcal{L}_s(Z_{s-} + c_s D_{s-}) \big\} \mathrm{d}R_s \tag{2.9}$$

or

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s} K(D_{s-}) \big\{ \mathcal{L}_s(Z_{s-} + c_s D_{s-}) - \mathcal{L}_s(Z_{s-} - c_s D_{s-}) \big\} \mathrm{d}R_s, \tag{2.10}$$

with an appropriate loss function $\mathcal{L}_s$. Given an initial guess $Z_0$ of the parameter $Z$, it can be shown that under certain regularity conditions the solutions of these stochastic differential equations approach the optimal $Z^*$.

### 2.2.3 Continuous-Time Algorithms

The preceding results have interesting special cases. We begin with the Itô type stochastic integral equations [c-Ker-Rand-1] and [c-Ker-Rand-2] given by

$$Z_t = Z_0 - \int_0^t \frac{a_s}{c_s} K(D_s) \big\{ f(Z_s + c_s D_s) \big\} \mathrm{d}s - \int_0^t \frac{a_s}{c_s} K(D_s) \sum_{j=1}^d \sigma_s^j(Z_s) \mathrm{d}W_s^j \tag{2.11}$$

and

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s} K(D_s)\big\{f(Z_s + c_s D_s) - f(Z_s - c_s D_s)\big\}\mathrm{d}s$$
$$- \int_0^t \frac{a_s}{2c_s} K(D_s) \sum_{j=1}^{d} \sigma_s^j(Z_s)\mathrm{d}W_s^j \tag{2.12}$$

with $\sigma^j \colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ and independent standard Brownian motions $(W_t)_{t\geqslant 0}$, defined on the standard basis.

Consider the following assumptions.

**Assumption 2.2.2.** *Let (kA)–(kC) from Assumption 2.2.1 hold.*

*(kD') Let $(a_t)_{t\geqslant 0}$, $(c_t)_{t\geqslant 0}$ be continuous processes satisfying*

$$a_t, c_t > 0 \qquad a_t, c_t \downarrow 0 \qquad \int_0^\infty a_s \mathrm{d}s = \infty \qquad \int_0^\infty a_s c_s \mathrm{d}s < \infty.$$

*(kF') For all $t \geqslant 0$ let the $\mathbb{R}^d$-valued, continuous random process $(D_t)_{t\geqslant 0}$ and the measurable function $K \colon \mathbb{R}^d \to \mathbb{R}^d$ fulfil $\mathbb{E}(\|K(D_t)\|\|D_t\|^2) < \infty$ and $\mathbb{E}(K(D_t) \otimes D_t) = \mathbb{1}_d$. Moreover assume*

$$\Big|\int_0^\infty a_s \Big(K(D_s) \otimes D_s - \mathbb{E}(K(D_s) \otimes D_s)\Big)\mathrm{d}s\Big| < \infty$$
$$\Big|\int_0^\infty a_s c_s \Big(\|K(D_s)\|\|D_s\|^2 - \mathbb{E}(\|K(D_s)\|\|D_s\|^2)\Big)\mathrm{d}s\Big| < \infty.$$

*(kG') For the one-measurement algorithm* (2.11) *we assume: There exists a random time $\tau(\omega) < \infty$ such that*

$$\Big|\int_0^\infty \frac{1}{1+\|Z_s\|^2} \frac{a_s}{c_s} \big\langle Z_s, K(D_s) f(Z_s)\big\rangle \mathrm{d}s\Big| \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\Big|\int_{\tau(\omega)}^\infty \frac{a_s}{c_s} \sum_{i=1}^{d} K(D_s)^{(i)} \mathrm{d}s\Big|,$$
$$\Big|\int_0^\infty \frac{a_s}{c_s} \sum_{i=1}^{d} \big(K(D_s)^{(i)} - \mathbb{E}(K(D_s))^{(i)}\big)\mathrm{d}s\Big| < \infty,$$

*for all $s \geqslant 0$, $\mathbb{E}(K(D_s)) = 0$.*

*(kH') For every $j \in \{1, \cdots, d\}$*

$$\int_0^\infty \frac{a_s^2}{c_s^2} \frac{\sigma_s^j(Z_s)^2}{1+\|Z_s\|^2} \mathrm{d}s < \infty$$

*holds.*

*Remark* 2.2.2. Note that the conditions in $(kG)$ concerning the jump part $\Delta R$ are superfluous in condition $(kG')$. For the same reason there is no condition $(kE')$.

**Corollary 2.2.2.** *Let Assumption 2.2.2 hold. Then the strong solution* $(Z_t)_{t\geqslant 0}$ *of the Itô type algorithms* (2.11) *or* (2.12) *converges to the minimizing point of the function* $f$.

*Proof.* We put the result down to Corollary 2.2.1. Setting $R_s := s$ assures assumptions (D) and (kG). Assumptions (E) and (F) follow immediately from the continuity of the processes $Z$ and $D$, respectively. Setting $\breve{M}(\mathrm{d}s, x) = \sum_{j=1}^{d} \sigma_s^j(x)\mathrm{d}W_s^j$ we find

$$\begin{aligned}
&\left[\int_0^{\cdot} K(D_s)\breve{M}(\mathrm{d}s, x_s)\right]_t \\
&\quad = \left[\int_0^{\cdot} K(D_s)\sum_{j=1}^{d}\sigma_s^j(x)\mathrm{d}W_s^j\right]_t = \int_0^{\cdot} \|K(D_s)\|^2 \sum_{j,k=1}^{d} (\sigma_s^j(x)\sigma_s^k(x)[\mathrm{d}W_s^j, \mathrm{d}W_s^k]_t \\
&\quad = \int_0^{\cdot} \|K(D_s)\|^2 \sum_{j=1}^{d} (\sigma_s^j(x)^2 \mathrm{d}s
\end{aligned}$$

hence choosing $h_s^{ii}(x) := \left(\|K(D_s)\|^2 \sum_{j=1}^{d}(\sigma_s^j(x)\right)_i^2$ yields (kH) from (kH'). □

**Example 2.2.1.** We establish two randomized Itô type stochastic approximation algorithms. An overview of time-discrete examples can be found for instance in the paper of Dippon [10]. By definition, the directional derivative is

$$\lim_{h\to 0} \frac{f_s(x+hv) - f_s(x)}{h}, \qquad v \in \mathbb{R}^d.$$

Hence,

$$\frac{a_t}{c_t}\{f(Z_t + c_t D_t)\} \approx a_t \nabla f(Z_t) D_t + \frac{a_t}{c_t} f(Z_t)$$

and

$$\frac{a_t}{2c_t}\{f(Z_t + c_t D_t) - f(Z_t - c_t D_t)\} \approx a_t \nabla f(Z_t) D_t$$

hold true. Assuming $\mathbb{E}(D_t \otimes D_t) = \frac{1}{d}\mathbb{1}_d$, this motivates an Itô type version of the one- and two-measurement random direction stochastic approximation

$$Z_t = Z_0 - \int_0^t \frac{a_s}{c_s} D_s\{f_s(Z_s + c_s D_s)\}\mathrm{d}s - \int_0^t \frac{a_s}{c_s} D_s \mathrm{d}W_s$$

and

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s} D_s\{f_s(Z_s + c_s D_s) - f_s(Z_s - c_s D_s)\}\mathrm{d}s - \int_0^t \frac{a_s}{2c_s} D_s \mathrm{d}W_s$$

where for the first one we additionally assume $\mathbb{E}D_t = 0$.

A possible choice for $D$ for $d$-dimensional problems is the Brownian motion on the $(d-1)$-sphere $S$ with an initial value generated by a uniform distribution on the sphere. A Brownian motion on the unit sphere $S$ of $\mathbb{R}^d$, $d \geqslant 3$, can be constructed by applying the function $\phi\colon \mathbb{R}^d \backslash \{0\} \to S, x \mapsto x\|x\|^{-1}$ to the $d$-dimensional Brownian motion $B = (B_1, \dots, B_d)$. This yields a stochastic integral $Y = (Y_1, \dots, Y_d) = \phi(B)$, which according to Itô's formula is given by

$$\mathrm{d}Y_i = \frac{\|B\| - B_i^2}{\|B\|^3}\mathrm{d}B_i - \sum_{j \neq i} \frac{B_j - B_i}{\|B\|^3}\mathrm{d}B_j - \frac{n-1}{2}\frac{B_i}{\|B\|^3}\mathrm{d}t \quad \text{with } i, j = 1, \dots, d.$$

Now the process $Z$, defined by the time change

$$Z_t(\omega) := Y_{\alpha(t,\omega)}(\omega) \quad \text{with } \alpha(t,\omega) := \beta(t,\omega)^{-1}, \beta(t,\omega) := \int_0^t \frac{1}{\|B\|^2}\mathrm{d}s,$$

is a Brownian motion on the unit sphere $S$. Details can be found in the book of Øksendal [30].

We now turn to the special case $d = 2$, and verify the conditions of Assumption 2.2.2 that involve the process $D$. Here, with a 1-dimensional Brownian motion $B$, $Y = e^{iB} = (\cos(B), \sin(B))$ is a Brownian motion on the unit circle. With a random variable $U \sim \mathrm{Unif}[0, 2\pi)$ independent of the standard Brownian motion $B$ we find

$$\begin{aligned}\mathbb{E}(D_t) = \mathbb{E}\begin{pmatrix}\cos(B_t + U) \\ \sin(B_t + U)\end{pmatrix} &= \mathbb{E}\begin{pmatrix}\cos(B_t)\cos(U) - \sin(B_t)\sin(U) \\ \sin(B_t)\cos(U) + \cos(B_t)\sin(U)\end{pmatrix} \\ &= \begin{pmatrix}\mathbb{E}(\cos(B_t)) \cdot 0 - \mathbb{E}(\sin(B_t)) \cdot 0 \\ \mathbb{E}(\sin(B_t)) \cdot 0 + \mathbb{E}(\cos(B_t)) \cdot 0\end{pmatrix} = \begin{pmatrix}0 \\ 0\end{pmatrix}.\end{aligned}$$

Now consider

$$\begin{aligned}&\mathbb{E}\left(\begin{pmatrix}\cos(B_t + U) \\ \sin(B_t + U)\end{pmatrix} \otimes \begin{pmatrix}\cos(B_t + U) \\ \sin(B_t + U)\end{pmatrix}\right) \\ &\quad = \mathbb{E}\begin{pmatrix}\cos(B_t + U)^2 & \sin(B_t + U)\cos(B_t + U) \\ \sin(B_t + U)\cos(B_t + U) & \sin(B_t + U)^2\end{pmatrix}\end{aligned}$$

and calculate the matrix entries

$$\begin{aligned}\mathbb{E}\big(\cos(B_t + U)^2\big) &= \mathbb{E}\Big(\big(\cos(B_t)\cos(U) - \sin(B_t)\sin(U)\big)^2\Big) \\ &= \mathbb{E}\Big(\cos(B_t)^2\cos(U)^2 - 2\cos(B_t)\sin(B_t)\cos(U)\sin(U) \\ &\qquad + \sin(B_t)^2\sin(U)^2\Big) \\ &= \mathbb{E}(\cos(B_t)^2)\mathbb{E}(\cos(U)^2) - 2 \cdot 0 \cdot 0 + \mathbb{E}(\sin(B_t)^2)\mathbb{E}(\sin(U)^2) \\ &= \frac{1}{2}\mathbb{E}\big(\cos(B_t)^2 + \sin(B_t)^2\big) = \frac{1}{2},\end{aligned}$$

$$\begin{aligned}&\mathbb{E}\big(\sin(B_t + U)\cos(B_t + U)\big) \\ &\quad = \mathbb{E}\Big(\big(\sin(B_t)\cos(U) + \cos(B_t)\sin(U)\big) \cdot \big(\cos(B_t)\cos(U) - \sin(B_t)\sin(U)\big)\Big)\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\Big(\sin(B_t)\cos(B_t)\Big)\mathbb{E}\Big(\cos(U)^2 - \sin(U)^2\Big) \\
&\quad + \mathbb{E}\Big(\sin(U)\cos(U)\Big)\mathbb{E}\Big(\cos(B_t)^2 - \sin(B_t)^2\Big) = 0
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\big(\sin(B_t + U)^2\big) &= \mathbb{E}\Big(\big(\sin(B_t\cos(U) + \cos(B_t)\sin(U)\big)^2\Big) \\
&= \mathbb{E}\Big(\sin(B_t)^2\cos(U)^2 + 2\sin(B_t)\cos(B_t)\cos(U)\sin(U) \\
&\quad + \cos(B_t)^2\sin(U)^2\Big) \\
&= \mathbb{E}(\sin(B_t)^2)\mathbb{E}(\cos(U)^2) + 2\cdot 0\cdot 0 + \mathbb{E}(\cos(B_t)^2)\mathbb{E}(\sin(U)^2) \\
&= \frac{1}{2}\mathbb{E}\big(\sin(B_t)^2 + \cos(B_t)^2\big) = \frac{1}{2}.
\end{aligned}
$$

Consequently

$$
\mathbb{E}\left(\begin{pmatrix}\cos(B_t+U)\\ \sin(B_t+U)\end{pmatrix} \otimes \begin{pmatrix}\cos(B_t+U)\\ \sin(B_t+U)\end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}
$$

and thereby the choice $K(D_t) = 2D_t$ makes sense to achieve $\mathbb{E}(K(D_t)) = \mathbb{E}(2D_t) = 0$ and $\mathbb{E}(K(D_t) \otimes D_t) = \mathbb{1}_d$ for any $t \geqslant 0$.

Next we turn to assumptions $(kG')$ and consider

$$
\Big|\int_0^\infty \frac{a_s}{c_s}\Big(K(D_s) - \mathbb{E}(K(D_s))\Big)\mathrm{d}s\Big| < \infty.
$$

For that purpose employ Lemma A.1.4. It is sufficient to show that the rate-condition of the second moment is fulfilled. This is shown component-wise. As $U$ is not time-dependent, and we are only interested in an asymptotic result, it is sufficient to consider $\cos(B_t)$ and $\sin(B_t)$ instead of $\cos(D_t)$ and $\sin(D_t)$, respectively. In order to show

$$
\mathbb{E}\Big|\int_0^t \cos(B_s)\mathrm{d}s\Big|^2 = \mathcal{O}(t),
$$

Itô's formula and Itô's isometry yield

$$
\begin{aligned}
\mathbb{E}\Big|\int_0^t \cos(B_s)\mathrm{d}s\Big|^2 &= 2\mathbb{E}\Big|-\cos(B_t) + \cos(B_0) - \int_0^t \sin(B_s)\mathrm{d}B_s\Big|^2 \\
&\leqslant 4\mathbb{E}\Big|-\cos(B_t) + \cos(B_0)|^2 + 4\mathbb{E}|\int_0^t \sin(B_s)\mathrm{d}B_s\Big|^2 \\
&\leqslant \mathcal{C} + 4\mathbb{E}\Big|\int_0^t \sin(B_s)\mathrm{d}B_s\Big|^2 \leqslant \mathcal{C} + \mathcal{C}\int_0^t \mathbb{E}(\sin(B_s)^2)\mathrm{d}s = \mathcal{O}(t).
\end{aligned}
$$

In the same way $\mathbb{E}|\int_0^t \sin(B_s)\mathrm{d}s|^2 = \mathcal{O}(t)$ holds.

Now we show

$$\mathbb{E}\Big|\int_0^t \Big(K(D_s)\otimes D_s - \mathbb{E}(K(D_s)\otimes D_s)\Big)^{(jk)} \mathrm{d}s\Big|^2 = \mathcal{O}(t)$$

for each $k, j \in \{1, \dots, d\}$ to verify $\int_0^\infty a_s\Big(K(D_s)\otimes D_s - \mathbb{E}(K(D_s)\otimes D_s)\Big)\mathrm{d}s < \infty$ from condition $(kF')$. We calculate the corresponding matrix entries. Again it is sufficient to consider $\cos(B_t)$ and $\sin(B_t)$ instead of $\cos(D_t)$ and $\sin(D_t)$, respectively. Itô's formula and Itô's isometry yield

$$\begin{aligned}
&\mathbb{E}|\int_0^t (2\cos(B_s)-1)\mathrm{d}s|^2 \\
&\quad = \mathbb{E}|\int_0^t \cos(2B_s)\mathrm{d}s|^2 \leqslant \mathcal{C}\mathbb{E}|\cos(2B_t)^2| + \mathcal{C} + \mathcal{C}\mathbb{E}|\int_0^t \sin(2B_s)\mathrm{d}B_s|^2 \\
&\quad \leqslant \mathcal{C} + \mathcal{C}\int_0^t \mathbb{E}|\sin(2B_s)|^2\mathrm{d}s = \mathcal{O}(t).
\end{aligned}$$

Analogously we find

$$\begin{aligned}
&\mathbb{E}|\int_0^t (2\sin(B_s)-1)\mathrm{d}s|^2 \\
&\quad = \mathbb{E}|\int_0^t -\cos(2B_s)\mathrm{d}s|^2 \leqslant \mathcal{C}\mathbb{E}|\cos(2B_t)^2| + \mathcal{C} + \mathcal{C}\mathbb{E}|\int_0^t \sin(2B_s)\mathrm{d}B_s|^2 \\
&\quad \leqslant \mathcal{C} + \mathcal{C}\int_0^t \mathbb{E}|\sin(2B_s)|^2\mathrm{d}s = \mathcal{O}(t)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}|\int_0^t 2\sin(B_s)\cos(B_s)\mathrm{d}s|^2 \\
&\quad \leqslant \mathcal{C}\mathbb{E}|-\frac{B_t}{2} - \frac{1}{4}\sin(2B_t) + \frac{B_0}{2} + \frac{1}{4}\sin(2B_0)|^2 + \mathcal{C}\mathbb{E}|\int_0^t \cos(B_s)^2\mathrm{d}B_s|^2 \\
&\quad \leqslant \mathcal{C}\mathbb{E}(B_t^2) + \mathcal{C}\mathbb{E}|\sin(2B_t)^2| + \mathcal{C} + \mathcal{C}\int_0^t \mathbb{E}\cos(B_s)^4\mathrm{d}s = \mathcal{O}(t).
\end{aligned}$$

As $\|K(D_s)\| = 2$ and $\|D_s\| = 1$ it holds

$$\|K(D_s)\|\|D_s\|^2 - \mathbb{E}(\|K(D_s)\|\|D_s\|^2) = 2 - \mathbb{E}(2) = 0.$$

This shows the remaining condition of $(kF')$. For a one-dimensional stochastic approximation algorithm it is worth mentioning that the 0-sphere is not connected and therefore not useful. But for that case a counterpart can be found just by setting $D_t = \cos(B_t + U)$.

**Example 2.2.2.** For another example we return to the $d$-dimensional case and let

$$D_t^{-1} := (1/D_t^{(1)}, \dots, 1/D_t^{(d)})^T.$$

Then the Itô type simultaneous perturbation stochastic approximation (SPSA) algorithms (2.11) and (2.12) would be given by

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s} D_s^{-1}\{f(Z_s + c_s D_s) - f(Z_s - c_s D_s)\}\mathrm{d}s - \int_0^t \frac{a_s}{2c_s} D_s^{-1}\mathrm{d}W_s$$

and

$$Z_t = Z_0 - \int_0^t \frac{a_s}{c_s} D_s^{-1}\{f(Z_s + c_s D_s)\}\mathrm{d}s - \int_0^t \frac{a_s}{c_s} D_s^{-1}\mathrm{d}W_s.$$

However in a time-continuous framework it must be possible that $D_t$ equals zero for some $t$. It is hard to find distributions fulfilling the corresponding conditions on $D_t$ and $K(D_t)$.

**Example 2.2.3.** For $D$ consider the deterministic, $2\pi$-periodic function

$$D_t = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \quad \text{with } t \in [0, \infty).$$

We calculate the counterparts for the expectation values in (kF') and (kG'):

$$\frac{1}{2\pi}\int_0^{2\pi} D_t \mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi} \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \mathrm{d}t = 0$$

and

$$\frac{1}{2\pi}\int_0^{2\pi} D_t \otimes D_t \mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi} \begin{pmatrix} \cos(t)^2 & \sin(t)\cos(t) \\ \sin(t)\cos(t) & \sin(t)^2 \end{pmatrix} \mathrm{d}t = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Moreover, for $t > 0$,

$$\frac{1}{t}\| \int_0^t D_s \mathrm{d}s \|^2 = \mathcal{O}(1)$$

and

$$\frac{1}{t}\| \int_0^t (D_s \otimes D_s - \mathbb{1}_2)\mathrm{d}s \|^2 = \mathcal{O}(1)$$

hold true. Obviously the bounds are also of order $\mathcal{O}(t)$, such that we can apply Lemma A.1.4 analogously to the previous examples, such that the remaining conditions of (kF') and (kG') are verified. However having bounds of order $\mathcal{O}(1)$ also gives us the possibility to apply Corollary A.1.1 and thereby make less strict conditions for one-measurement algorithms.

**Example 2.2.4.** This example is similar to the previous one. However now the path of $D$ consists of sub-paths which we sample without replacement. Let the process $D$

go along one of the paths

$$t \mapsto \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \quad \text{or} \quad t \mapsto \begin{pmatrix} \cos(-t) \\ \sin(-t) \end{pmatrix}$$

with $t \geqslant 0$ for $m \geqslant 1$ times each. More precisely, let the process $D$ start at $D_0 = (1,0)^T$. At every time $t = k\pi$, $k \in \mathbb{N}_0$ it is $D_t = ((-1)^k, 0)^T$ where we sample without replacement, from initially $m$ upper arcs and $m$ lower arcs, which direction to choose next. Obviously the probability which path to choose next is dependent on the previously chosen paths. Every $2m$ steps no paths are left to choose from and the complete sample experiment will be repeated. In this case

$$\mathbb{E}\| \int_0^t D_t \mathrm{d}t \|^2 = \mathcal{O}(1)$$

and

$$\mathbb{E}\| \int_0^t (D_t \otimes D_t - \mathbb{1}_2) \mathrm{d}t \|^2 = \mathcal{O}(1)$$

hold, although $D$ is not deterministic. This is an example for a dependent perturbation where Corollary A.1.1 is applicable to verify the integral conditions in (kF') and (kG').

### 2.2.4 Discrete-Time Algorithms

We carry on with the time-discrete recursive algorithms [d-Ker-Rand-1] and [d-Ker-Rand-2] given by

$$Z_n = Z_{n-1} - \frac{a_n}{c_n} K(D_{n-1}) \left( f(Z_{n-1} + c_n D_{n-1}) + V_n \right) \tag{2.13}$$

and

$$Z_n = Z_{n-1} - \frac{a_n}{2c_n} K(D_{n-1}) \left( \left\{ f(Z_{n-1} + c_n D_{n-1}) - f(Z_{n-1} - c_n D_{n-1}) \right\} + V_n \right). \tag{2.14}$$

Consider the following assumptions.

**Assumption 2.2.3.** *Let (kA)–(kC) from Assumption 2.2.1 hold.*

*(kD") Let $(a_n)$, $(c_n)$ be sequences satisfying*

$$a_n, c_n > 0 \qquad a_n, c_n \downarrow 0 \qquad \sum_{n=1}^{\infty} a_n = \infty \qquad \sum_{n=1}^{\infty} a_n c_n < \infty.$$

*(kE") Assume*

$$\sum_{i=1}^{\infty} a_n^2 < \infty.$$

(kF”) *For all* $n \geqslant 0$ *let the* $\mathbb{R}^d$*-valued random process* $(D_n)_{n\geqslant 0}$ *and the measurable function* $K\colon \mathbb{R}^d \to \mathbb{R}^d$ *fulfil*

$$\mathbb{E}(K(D_n)\otimes D_n) = \mathbb{1}_d, \quad \mathbb{E}(\|D_n\|^4\|K(D_n)\|^2) < \infty \quad \textit{and} \quad \mathbb{E}\left(\|K(D_n)\|^2\right) < \infty.$$

*Moreover assume*

$$\Big|\sum_{n=1}^{\infty} a_n\Big(K(D_{n-1})\otimes D_{n-1} - \mathbb{E}(K(D_{n-1})\otimes D_{n-1})\Big)\Big| < \infty,$$

$$\Big|\sum_{n=1}^{\infty} a_n c_n\Big(\|K(D_{n-1})\|\|D_{n-1}\|^2 - \mathbb{E}(\|K(D_{n-1})\|\|D_{n-1}\|^2)\Big)\Big| < \infty,$$

$$\Big|\sum_{n=1}^{\infty} a_n^2 c_n^2\Big(\|D_{n-1}\|^4\|K(D_{n-1})\|^2 - \mathbb{E}(\|D_{n-1}\|^4\|K(D_{n-1})\|^2)\Big)\Big| < \infty,$$

$$\Big|\sum_{n=1}^{\infty} a_n^2\Big(\|D_{n-1}\|^2\|K(D_{n-1})\|^2 - \mathbb{E}(\|D_{n-1}\|^2\|K(D_{n-1})\|^2)\Big)\Big| < \infty.$$

(kG”) *For the one-measurement algorithm* (2.13) *we assume: There exists a stopping time* $\tau(\omega) < \infty$ *such that*

$$\Big|\sum_{n=1}^{\infty} \frac{1}{1+\|Z_{n-1}\|^2}\frac{a_n}{c_n}\Big\langle Z_{n-1}, K(D_{n-1})f(Z_{n-1})\Big\rangle\Big|$$

$$\leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\Big|\sum_{n=\tau(\omega)}^{\infty}\frac{a_n}{c_n}\sum_{i=1}^{d} K(D_{n-1})^{(i)}\Big|,$$

$$\Big|\sum_{n=1}^{\infty}\frac{a_n}{c_n}\sum_{i=1}^{d}\big(K(D_{n-1})^{(i)} - \mathbb{E}(K(D_{n-1}))^{(i)}\big)\Big| < \infty,$$

*for all* $n \geqslant 1$, $\mathbb{E}(K(D_{n-1})) = 0$,

$$\underset{C_1,C_2\geqslant 0}{\exists}\ \underset{z\in\mathbb{R}^d}{\forall}\ C_1(1+\|z\|) \leqslant f(z) \leqslant C_2(1+\|z\|),$$

$$\Big|\sum_{n=1}^{\infty}\frac{a_n^2}{c_n^2}\Big(\|K(D_{n-1})\|^2 - \mathbb{E}(\|K(D_{n-1})\|^2)\Big)\Big| < \infty,$$

*and*

$$\sum_{n=1}^{\infty}\frac{a_n^2}{c_n^2} < \infty.$$

(kH”) *Assume*

$$\sup_{n\in\mathbb{N}} \mathrm{E}\|K(D_{n-1})V_n\|^2 \mid \mathcal{G}_n) < \infty \quad \textit{and} \quad \mathrm{E}(K(D_{n-1})V_n \mid \mathcal{G}_n) = 0 \ \textit{a.s.}$$

*where* $\mathcal{G}_n := \sigma(Z_1, \dots, Z_{n-1}, D_1, \dots, D_{n-1})$.

**Corollary 2.2.3.** *Let Assumption 2.2.3 hold. Then the strong solution* $(Z_n)$ *of the time-discrete recursive algorithms* (2.13) *or* (2.14), *respectively converges to the minimizing point of the function* $f$.

*Proof.* It is sufficient to check the assumptions of Corollary 2.2.1. We only investigate recursion (2.13). The case for (2.14) follows analogously. Set $F(D_{s-}, c_s, Z_{s-}) := K(D_{s-}) f(Z_{s-} + c_s D_{s-})$. We extend the sequence $(V_n)$ to a time-continuous process $(\widetilde{V}_t)_{t \geqslant 0}$ defined by

$$\widetilde{V}_t := \begin{cases} V_1 & \text{for } t = 0 \\ V_n & \text{for } n-1 < t \leqslant n \text{ with } n \in \mathbb{N}. \end{cases}$$

Furthermore we define

$$M(\mathrm{d}s, D_{s-}, x) := K(D_{s-}) \widetilde{V}_s \mathrm{d}R_s \quad \text{and} \quad R_s := \max_{\substack{n \in \mathbb{N} \\ n \leqslant s}} \{n\} = \lfloor s \rfloor \text{ for } s \geqslant 0.$$

These definitions are used for

$$\begin{aligned} \int_0^t M(\mathrm{d}s, D_{s-}, x) &= \int_0^t K(D_{s-}) \widetilde{V}_s \mathrm{d}R_s = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} K(D_{n-1}) \widetilde{V}_n (\Delta R_n) = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} K(D_{n-1}) \widetilde{V}_n \\ &= \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} K(D_{n-1}) V_n =: H_t. \end{aligned}$$

We now show that $\int_0^t M(\mathrm{d}s, D_{s-}, x)$ is a martingale with respect to $\widetilde{\mathcal{F}}_t := \mathcal{F}_{R_t}$, $t \geqslant 0$.

$$\begin{aligned} \mathbb{E}(H_t \mid \mathcal{F}_s) &= \mathbb{E}(H_t \mid \mathcal{F}_{\lfloor s \rfloor}) = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} \mathbb{E}(K(D_{n-1}) V_n \mid \mathcal{F}_{\lfloor s \rfloor}) \\ &= \sum_{\substack{n \leqslant \lfloor s \rfloor \\ n \in \mathbb{N}}} \mathbb{E}(K(D_{n-1}) V_n \mid \mathcal{F}_{\lfloor s \rfloor}) + \sum_{\substack{\lfloor s \rfloor < n \leqslant t \\ n \in \mathbb{N}}} \mathbb{E}(K(D_{n-1}) V_n \mid \mathcal{F}_{\lfloor s \rfloor}) \\ &= \sum_{\substack{n \leqslant \lfloor s \rfloor \\ n \in \mathbb{N}}} K(D_{n-1}) V_n + 0 = \sum_{\substack{n \leqslant \lfloor s \rfloor \\ n \in \mathbb{N}}} K(D_{n-1}) V_n = H_s \end{aligned}$$

We get for $n \in \mathbb{N}$

$$\begin{aligned} Z_n - Z_0 &= -\sum_{j=1}^n \frac{a_j}{c_j} K(D_{j-1}) \{f(Z_{j-1} + c_j D_{j-1})\} (\Delta R_j) - \sum_{j=1}^n \frac{a_j}{c_j} K(D_{j-1}) V_j (\Delta R_j) \\ &= -\sum_{j=1}^n \frac{a_j}{c_j} K(D_{j-1}) \{f(Z_{j-1} + c_j D_{j-1})\} - \sum_{j=1}^n \frac{a_j}{c_j} K(D_{j-1}) V_j. \end{aligned}$$

Writing recursion (2.13) in a telescoping series yields

$$Z_n - Z_0 = \sum_{j=1}^{n}(Z_j - Z_{j-1}) = -\sum_{j=1}^{n} \frac{a_j}{c_j} K(D_{j-1})\{f(Z_{j-1} + c_j D_{j-1})\} - \sum_{j=1}^{n} \frac{a_j}{c_j} K(D_{j-1}) V_j.$$

As a consequence it is sufficient to verify the assumptions of the theorem of the semimartingale case. Assumptions (D) and (E) follow by

$$\int_0^\infty a_s \mathrm{d}R_s = \sum_{j=1}^{\infty} a_j(\Delta R_j) = \sum_{j=1}^{\infty} a_j = \infty,$$
$$\int_0^\infty a_s c_s \mathrm{d}R_s = \sum_{j=1}^{\infty} a_j c_j(\Delta R_j) = \sum_{j=1}^{\infty} a_j c_j < \infty,$$

and

$$\int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s = \sum_{j=1}^{\infty} a_j^2(\Delta R_j)^2 = \sum_{j=1}^{\infty} a_j^2 < \infty.$$

Noting that $\mathrm{d}R_s = \mathrm{d}\lfloor s \rfloor$, assumptions (kF) and (kG) follow obviously from (kF”) and (kG”), respectively. Finally we deduce (kH) from (kH”). Knowing that

$$\begin{aligned}[\int_0^\cdot M(\mathrm{d}s, D_{s-}, x)]_t = [\int_0^\cdot K(D_{s-})\widetilde{V}_s \mathrm{d}R_s]_t &= \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} \mathbb{E}\left(K(D_{n-1})^2 V_n^2 (\Delta R_n)^2 \mid \mathcal{F}_{n-1}\right) \\ &= \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} \mathbb{E}\left(K(D_{n-1})^2 V_n^2 \mid \mathcal{F}_{n-1}\right)\end{aligned}$$

it is sufficient to show

$$\int_0^\infty \frac{a_s^2}{c_s^2} \frac{h_s^{ii}(Z_{s-})}{1 + \|Z_{s-}\|^2} \mathrm{d}R_s \leqslant \sum_{n \in \mathbb{N}} \frac{a_n^2}{c_n^2} \mathbb{E}\left(K(D_{n-1})^2 V_n^2 \mid \mathcal{F}_{n-1}\right) < \infty.$$

As the sum consists of positive terms only, we can apply the monotone convergence theorem:

$$\mathbb{E} \sum_{n \in \mathbb{N}} \frac{a_n^2}{c_n^2} \mathbb{E}\left(K(D_{n-1})^2 V_n^2 \mid \mathcal{F}_{n-1}\right) = \sum_{n \in \mathbb{N}} \frac{a_n^2}{c_n^2} \mathbb{E}\,|K(D_{n-1}) V_n|^2$$

Finally, Hölder’s inequality yields

$$\sum_{n \in \mathbb{N}} \frac{a_n^2}{c_n^2} \mathbb{E}\,|K(D_{n-1}) V_n|^2 \leqslant \left(\sup_{n \in \mathbb{N}} \mathbb{E}\|K(D_{n-1}) V_n\|^2\right) \sum_{n \in \mathbb{N}} \frac{a_n^2}{c_n^2} < \infty.$$

$\square$

**Example 2.2.5.** Imagine an irreducible, stationary Markovian chain $(D_n)_{n \geqslant 1}$ that moves along the vertices of the hypercube $[-1, 1]^d$ and has a symmetric transition matrix. For instance, consider two kinds of settings.

- $D$ moves to vertices to which it is connected via a common edge or stays at the same vertex. All these events shall have the same probability. All other vertices are not accessible in one step.
- $D$ must leave the current vertex. It moves to vertices to which it is connected via a common edge with same probability. All other vertices are not accessible in one step.

The verifications of both settings are done in the same way. More specifically we will consider irreducible, symmetric, doubly stochastic transition matrices.

Choose $K(D_n) := D_n$. Then for any $i \in \{1, \ldots, n\}$ it holds

$$\mathbb{E}\big(K(D_i) \otimes D_i\big)^{jk} = \begin{cases} \mathbb{E}(D_i^j D_i^k) = \mathbb{E}((D_i^j)^2) = 1 & \text{if } j = k \\ \mathbb{E}(D_i^j D_i^k) = \frac{1}{4}\big(1 - 1 - 1 + (-1)^2\big) = 0 & \text{if } j \neq k. \end{cases}$$

Note that in this case Random Direction Stochastic Approximation (RDSA)

$$Z_{n+1} = Z_n - \frac{a_n}{2c_n} D_n \left\{\big(f(Z_n + c_n D_n) - W_{n,1}\big) - \big(f(Z_n - c_n D_n) - W_{n,2}\big)\right\}$$

and Simultaneous Perturbation Stochastic Approximation (SPSA)

$$Z_{n+1} = Z_n - \frac{a_n}{2c_n} D_n^{-1} \left\{\big(f(Z_n + c_n D_n) - W_{n,1}\big) - \big(f(Z_n - c_n D_n) - W_{n,2}\big)\right\}$$

are identical. In order to achieve moment conditions like $\mathbb{E}(K(D_n)) = 0$ we choose the initial value $D_0$ uniformly distributed on $\{-1, 1\}^d$. Consequently $\mathbb{E}(D_0) = 0 \in \mathbb{R}^d$. Due to stationarity, $\mathbb{E}(K(D_n)) = \mathbb{E}(D_n)$ equals zero for any $n$.

Now we show that $\sum_{n=1}^{\infty} \frac{a_n}{c_n}\Big(K(D_n) - \mathbb{E}(K(D_n))\Big) < \infty$ in $(kG'')$ is fulfilled. For that purpose verify the conditions of Lemma A.1.4. The only non-obvious one is

$$\mathbb{E}\|\sum_{i=1}^{n} D_i\|^2 = \mathcal{O}(n).$$

First note that

$$\mathbb{E}\langle X, Y\rangle = \sum_x \sum_y \langle x, y\rangle \mathbb{P}(X = x, Y = y) = \sum_x \sum_y \langle x, y\rangle \mathbb{P}(Y = y \mid X = x)\mathbb{P}(X = x).$$

Let $\delta$ and $\delta'$ denote elements of the set of all vertices of our hypercube. In our setting, symmetry yields $\sum_\delta \delta = 0$. Bringing these ideas together yields

$$\mathbb{E}\langle D_i, D_j\rangle = \sum_\delta \sum_{\delta'} \langle \delta, \delta'\rangle \mathbb{P}(D_i = \delta' \mid D_j = \delta)\mathbb{P}(D_j = \delta).$$

Due to stationarity, $\mathbb{P}(D_j = \delta) = \frac{1}{2^d}$ for all $j$. Note that the $2^d$-dimensional transition matrix from $D_i$ to $D_{i+1}$, which we denote by $M$ in the following, is an irreducible, symmetric, doubly stochastic matrix. For every stochastic matrix, 1 is an eigenvalue.

According to the Perron-Frobenius theorem for stochastic matrices [19, Thm. A.2.4], all eigenvalues have absolute value $\leqslant 1$. For an irreducible, non-negative matrix A, the spectral radius $\rho(A)$ is a positive, simple eigenvalue of $A$. Hence the eigenvalue 1 has algebraic (and thereby geometric) multiplicity 1. As $M$ is a stochastic matrix, the eigenvector related to eigenvalue 1 is $v_1 := (1, 1, \ldots, 1)^T = \vec{1}$. In order to calculate $M^n$, we diagonalize $M$ by $\Lambda := T^T M T$. It is of the form

$$\Lambda = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_{2^d} \end{pmatrix}.$$

The remaining eigenvalues $\lambda_2, \ldots, \lambda_d$ are either $-1$ or have absolute value less than 1. Note that there is no eigenvalue $-1$ if $M$ is primitive. This however does not apply to all possible matrices in our example. The matrix

$$\begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

for example has eigenvalues 1, $-1$ and 0.

Now the $n$-th power of $\Lambda$ is

$$\Lambda^n = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \lambda_2^n & \ldots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_{2^d}^n \end{pmatrix}.$$

Calculating $M^n = T \Lambda^n T^T$ with

$$T = \begin{pmatrix} \frac{v_1}{\|v_1\|} & \frac{v_2}{\|v_2\|} & \cdots & \frac{v_{2^d}}{\|v_{2^d}\|} \end{pmatrix} = \begin{pmatrix} \frac{\vec{1}}{\sqrt{2^d}} & \frac{v_2}{\|v_2\|} & \cdots & \frac{v_{2^d}}{\|v_{2^d}\|} \end{pmatrix}$$

yields the desired matrix. Due to the special form of the eigenvector $v_1$ all entries of $M^n$ consist, amongst others, of a summand with eigenvalue 1 divided by the squared length of its eigenvector, namely $1 \cdot \frac{1}{2^d}$. This very same summand appears in every matrix entry. All other summands consist of the other eigenvalues and are not necessarily the same for each matrix entry as their corresponding eigenvector is not a multiple of $\vec{1}$. Hence we use different arguments for them in the next step.

Now calculate $\mathbb{E}\langle D_0, D_n \rangle$ by multiplying each component of $M^n$ with its corresponding $\langle \delta, \delta' \rangle$ and the probability $\mathbb{P}(D_0 = \delta)$ and then summing up all entries. As a result we get

$$\mathbb{E}\langle D_0, D_n \rangle \leqslant \frac{1}{2^d} \sum_{\delta} \sum_{\delta'} \frac{1}{2^d} \langle \delta, \delta' \rangle \cdot 1 + \mathcal{C} \sum_{k=2}^{2^d} \lambda_k^n.$$

The first term on the right side represents the impact of the eigenvalue 1 and second term stands for the remaining ones, those structure is not so handy, as their corresponding eigenvectors are less trivial. Due to symmetry

$$\sum_{\delta}\sum_{\delta'}\langle\delta,\delta'\rangle = \sum_{\delta}\langle\delta,\sum_{\delta'}\delta'\rangle = \sum_{\delta}\langle\delta,0\rangle = 0,$$

hence the first term vanishes. Now compute

$$\sum_{i=0}^{n-1}\mathbb{E}\langle D_0, D_i\rangle \leqslant \sum_{i=0}^{n-1}\mathcal{C}\sum_{k=2}^{2^d}\lambda_k^i.$$

There are two possibilities for $\lambda_k$. Either it is $-1$ or $|\lambda_k| < 1$. As $\sum_{i=0}^{\infty}(-1)^i < \infty$ and the geometric series converges, we can find a bound $\mathcal{C}$ which is independent of $n$ such that $\sum_{i=0}^{n-1}\mathbb{E}\langle D_0, D_n\rangle < \mathcal{C}$. Finally

$$\mathbb{E}\|\sum_{i=0}^{n-1}D_i\|^2 = \sum_{i=0}^{n-1}\mathbb{E}\sum_{j=0}^{n-1}\langle D_i, D_j\rangle \leqslant \mathcal{C}\sum_{i=0}^{n-1}1 = \mathcal{O}(n).$$

Now check $\sum_{n=1}^{\infty}a_n\Big(K(D_{n-1})\otimes D_{n-1} - \mathbb{E}(K(D_{n-1})\otimes D_{n-1})\Big) < \infty$ in condition $(kF'')$. By a component-wise investigation of the entries we show

$$\mathbb{E}\|\sum_{i=0}^{n-1}\big(K(D_i)\otimes D_i\big) - \mathbb{E}\big(K(D_i)\otimes D_i\big)\|^2 = \mathcal{O}(n).$$

Note that all principal diagonal entries of $\big(K(D_i)\otimes D_i\big) - \mathbb{E}\big(K(D_i)\otimes D_i\big)$ are zero. All other entries, namely

$$\Big(\big(K(D_i)\otimes D_i\big) - \mathbb{E}\big(K(D_i)\otimes D_i\big)\Big)^{j,k} = \Big(K(D_i)\otimes D_i\Big)^{j,k}$$

for $j \neq k$, behave in the same way as $\mathbb{E}\|\sum_{i=1}^{n}D_{i-1}\|^2$ which was just proven to be $\mathcal{O}(n)$. Namely we consider

$$\begin{aligned}&\mathbb{E}\Big\langle (D_i\otimes D_i) - \mathbb{E}(D_i\otimes D_i), (D_j\otimes D_j) - \mathbb{E}(D_j\otimes D_j)\Big\rangle\\ &= \sum_{\delta}\sum_{\delta'}\Bigg(\Big\langle\Big((\delta\otimes\delta) - \mathbb{E}(D_i\otimes D_i)\Big), \Big((\delta'\otimes\delta') - \mathbb{E}(D_j\otimes D_j)\Big)\Big\rangle\\ &\cdot\mathbb{P}(D_i = \delta' \mid D_j = \delta)\mathbb{P}(D_j = \delta)\Bigg),\end{aligned}$$

where for matrices $X$ and $Y$, $\langle X, Y\rangle$ denotes the Frobenius inner product, which means that the elements of the matrices shall be multiplied element-wise and then summed up. Then, similarly as before, $\sum_{\delta}\Big((\delta\otimes\delta) - \mathbb{E}(D_i\otimes D_i)\Big) = \sum_{\delta}\Big((\delta\otimes\delta) - \mathbb{1}_d\Big) = 0$. Just consider the non-diagonal entries $\sum_{\delta}\delta^{(i)}\cdot\delta^{(j)}$ with $i \neq j$. Then summing over the $2^d$ vertices $\delta$ yields

$$\sum_{\delta} \delta^{(i)} \cdot \delta^{(j)} = \left( \frac{2^d}{2} \cdot 1 + \frac{2^d}{2} \cdot (-1) \right) = 0.$$

Consequently we can cope with the simple, maximum eigenvalue 1 of the transition matrix. All other eigenvalues are either $-1$ of have absolute value less than 1 and can be handled as before. As $\|K(D_n)\| = \|D_n\| = \sqrt{d}$, it holds

$$\|K(D_n)\|^2 - \mathbb{E}\big(\|K(D_n)\|^2\big) = d - \mathbb{E}\big(d\big) = 0.$$

Thus condition $\sum_{n=1}^{\infty} \frac{a_n^2}{c_n^2} \Big( \|K(D_{n-1})\|^2 - \mathbb{E}\big(\|K(D_{n-1})\|^2\big) \Big) < \infty$ in $(kG'')$ holds. All assumptions of $(kF'')$ including the norm of $K(D_n)$ or $D_n$ follow with the same argument.

**Example 2.2.6.** Let $D$ follow the deterministic, periodic sequence of the 3-dimensional vectors

$$\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\-1\end{pmatrix}, \begin{pmatrix}-1\\1\\-1\end{pmatrix}, \begin{pmatrix}-1\\-1\\-1\end{pmatrix}, \begin{pmatrix}-1\\-1\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\1\end{pmatrix}.$$

The mean of these vectors is zero. Moreover $\sum_{k=1}^{8} D_k \otimes D_k = \mathbb{1}_3$. As the sequence of vectors is deterministic and periodic,

$$\mathbb{E}\|\sum_{i=1}^{n} D_i\|^2 = \|\sum_{i=1}^{n} D_i\|^2 = \mathcal{O}(1)$$

and

$$\mathbb{E}\|D_i \otimes D_i\|^2 = \|D_i \otimes D_i\|^2 = \mathcal{O}(1)$$

hold true. This is sufficient to apply Lemma A.1.4, for which we only need the rate $\mathcal{O}(n)$. But we could even apply Corollary A.1.1 and thereby make less strict conditions for one-measurement algorithms.

*Remark* 2.2.3. It is worth to mention that deterministically perturbated algorithms where already handled by Bhatnagar [2]. However much stricter assumptions were made there and only time-discrete algorithms were considered.

**Example 2.2.7.** Consider $D$ to follow the vertices of a $d$-dimensional cube that are sampled without replacement. After $2^d$ steps, when no vertices are left to choose, we start again sampling from all vertices. Here

$$\mathbb{E}\|\sum_{i=1}^{n} D_i\|^2 = \mathcal{O}(1)$$

and

$$\mathbb{E}\|D_i \otimes D_i\|^2 = \mathcal{O}(1)$$

hold. Hence this is an application for Corollary A.1.1 with a non-deterministic $D$.

A part of the following result has already been shown by Dippon [10] under slightly different assumptions. There only the two-measurement algorithm (2.14) was inquired. The rest of the corollary is new.

**Corollary 2.2.4.** *Let all conditions of Corollary 2.2.3 hold and furthermore assume that $(D_n)_{n\geqslant 1}$ is not predictable but a sequence of i.i.d. random variables. Then the strong solution $(Z_n)$ of the time-discrete recursive algorithms* (2.13) *or* (2.14)*, respectively converges to the minimizing point of function $f$.*

*Proof.* The proof is a direct consequence of Corollary 2.2.3, noting that most conditions of assumption $(kF'')$ are fulfilled due to the Khintchine-Kolmogorov convergence theorem.

Alternatively one can follow Dippon's approach [10] to consider $\langle Z_n, Z_n\rangle$ and follow similar steps as in the proof of Theorem 2.1.1. Decompose it into

$$\begin{aligned}\|Z_n\|^2 = {} & \Big(\mathbb{E}(\|Z_n\|^2 \mid Z_0,\dots,Z_n,D_0,\dots,D_{n-1})\Big)\\ & + \Big(\|Z_n\|^2 - \mathbb{E}(\|Z_n\|^2 \mid Z_0,\dots,Z_n,D_0,\dots,D_{n-1})\Big).\end{aligned}$$

The first summand represents the predictable part and the second one the martingale part. The bounds

$$\begin{aligned}\Big|\sum_{n=1}^{\infty} a_n\Big(K(D_{n-1})\otimes D_{n-1} - \mathbb{E}(K(D_{n-1})\otimes D_{n-1})\Big)\Big| &< \infty\\ \Big|\sum_{n=1}^{\infty} a_n c_n\Big(\|K(D_{n-1})\|\|D_{n-1}\|^2 - \mathbb{E}(\|K(D_{n-1})\|\|D_{n-1}\|^2)\Big)\Big| &< \infty\\ \Big|\sum_{n=1}^{\infty} a_n^2 c_n^2\Big(\|D_{n-1}\|^4\|K(D_{n-1})\|^2 - \mathbb{E}(\|D_{n-1}\|^4\|K(D_{n-1})\|^2)\Big)\Big| &< \infty\\ \Big|\sum_{n=1}^{\infty} a_n^2\Big(\|D_{n-1}\|^2\|K(D_{n-1})\|^2 - \mathbb{E}(\|D_{n-1}\|^2\|K(D_{n-1})\|^2)\Big)\Big| &< \infty,\end{aligned}$$

from condition (kF") are trivially fulfilled. For example

$$\begin{aligned}&\Big|\sum_{n=1}^{\infty}\mathbb{E}\Big(a_n\Big(K(D_{n-1})\otimes D_{n-1} - \mathbb{E}(K(D_{n-1})\otimes D_{n-1})\Big)\Big|Z_0,\dots,Z_{n-1},D_0,\dots,D_{n-2}\Big)\Big|\\ &\quad= \Big|\sum_{n=1}^{\infty} a_n\Big(\mathbb{E}(K(D_{n-1})\otimes D_{n-1} \mid Z_0,\dots,Z_{n-1},D_0,\dots,D_{n-2})\\ &\qquad - \mathbb{E}(K(D_{n-1})\otimes D_{n-1})\Big)\Big|\end{aligned}$$

$$= \Big|\sum_{n=1}^{\infty} a_n \Big(\mathbb{E}(K(D_{n-1}) \otimes D_{n-1}) - \mathbb{E}(K(D_{n-1}) \otimes D_{n-1})\Big)\Big|$$
$$= 0.$$

For the same reason the terms

$$\Big|\sum_{n=1}^{\infty} \frac{1}{1+\|Z_{n-1}\|^2} \frac{a_n}{c_n} \big\langle Z_{n-1}, K(D_{n-1}) f(Z_{n-1})\big\rangle\Big|,$$
$$\Big|\sum_{n=1}^{\infty} \frac{a_n}{c_n} \sum_{i=1}^{d} \big(K(D_{n-1})^{(i)} - \mathbb{E}(K(D_{n-1}))^{(i)}\big)\Big|,$$
$$\Big|\sum_{n=1}^{\infty} \frac{a_n^2}{c_n^2} \Big(\|K(D_{n-1})\|^2 - \mathbb{E}(\|K(D_{n-1})\|^2)\Big)\Big|,$$

in (kG”) can be handled, as the summands turn to zero after applying the conditional expectation to them. Note that this alternative approach is not applicable for dependent perturbations $(D_n)_{n\geqslant 1}$.

□

*Remark* 2.2.4. In a continuous-time framework we do not have such a process $(D_t)_{t\geqslant 0}$ which is path continuous, $D_{t_1}$ and $D_{t_2}$ are i.i.d. for $t_1 \neq t_2$, and satisfies the moment conditions like $\mathbb{E}(K(D_t)) = 0$ and $\mathbb{E}(K(D_t) \otimes D_t) = \mathbb{1}_d$ in order to achieve consistency of $(Z_t)_{t\geqslant 0}$.

A comprehensive description of time-discrete examples with i.i.d. randomization can be found in [10, Chapter 5].

### 2.2.5 An Application in Wing Design Optimization

This application deals with the design of a wing shape such that the lift $L$ to drag $D$ ratio (L/D-ratio) is maximized with the wing weight as a constraint. The equations which show the relation of $L$ and $D$ with the quantities to be optimized, are very complex. In the following paragraph, the underlying equations and the relation to the parameters to be optimized, are presented. A detailed description of this example can be found in a paper of Xing and Damodaran from 2002 [44].

We begin with the description of the numerator. Here $L$ is defined as $L = C_L q S$, with $q = \frac{1}{2}\rho V^2$ the dynamic pressure, $\rho$ the density of the air, $V$ the flight speed, $C_L = C_{L\alpha}\alpha$ the lift coefficient, with $\alpha$ the angle of attack and $C_{L\alpha} = 2\pi A_R/\big(2 + \sqrt{4 + (A_R\beta/\eta)^2(1 + \tan^2\lambda/\beta^2))}\big)$ the lift curve slope. In the lift curve slope expression, $A_R = b^2/S$ is the wing aspect ratio, $b$ the wing span, $\lambda$ the wing sweep angle, $\eta$ the airfoil efficiency factor, $\beta = 1 - M^2$ the compressibility factor, and $M$ the Mach number. The total drag is defined by $D = C_{\mathrm{D}} q S$, with total drag coefficient $C_{\mathrm{D}} = C_{\mathrm{Di}} + C_{\mathrm{D0}}$, induced drag coefficient $C_{\mathrm{Di}} = C_L^2/(\pi A_R e)$ and zero-lift drag coefficient $C_{\mathrm{D0}} = C_f F Q$. Here $e = 4.61(1 - 0.045 A_R^{0.68})(\cos\lambda)^{0.15} - 3.1$ is the wing planform efficiency factor, $C_f = 0.455/(\log_{10}\mathrm{Re})^{2.58}(1 + 0.144M^2)^{0.65}$ the surface skin-friction coefficient, which in turn is a function of the Reynolds number Re,

$F = \big(1 + (0.6/(x/c)_m)(t/c) + 100(t/c)^4\big)(1.34M^{0.18}(\cos\lambda)^{0.28})$, $t/c$ the airfoil thickness-to-chord ratio, $(x/c)_m$ the chord-wise location of the maximum thickness-to-chord ratio, and $Q$ a factor which stands for interference effects on drag. The weight of the wing is $W_{\text{wing}} = 0.0106(W_{\text{dg}}N_z)^{0.5}S^{0.622}A_R^{0.75}(t/c)^{-0.4}(\cos\lambda)^{-1}$, with design gross weight $W_{\text{dg}}$ and ultimate load factor $N_z$.

The variables to be optimized are angle of attack $\alpha$, wing span $b$, mean aerodynamic chord $c$, sweep angle $\lambda$ and wing weight $W_{\text{wing}}$. Additionally the following constraints are made:

$$\begin{aligned} 1.0\text{deg} \leqslant \alpha \leqslant 10.0\text{deg} \qquad & \qquad 10.0 \leqslant b \leqslant 50.0 \\ 3.5 \leqslant c \leqslant 10.0 \qquad & \qquad 0.0\text{deg} \leqslant \lambda \leqslant 35.0\text{deg} \\ 0.5 \leqslant A_R \leqslant 15.0 \qquad & \qquad W_{\text{wing}} \leqslant 2473(lb) \end{aligned}$$

Hence

$$f(Z) = \frac{D(Z)}{L(Z)} + \sum_{j\in J} \max(0, g_j(Z))^2$$

with $Z = (\alpha, b, c, \lambda, W_{\text{wing}})^T$ and $J$ the number of conditions, is to be minimized with $Z$ representing the five design variables and the design constraints $g_j(Z) \leqslant 0$ formulated as inequality constraints.

Xing and Damodaran simulated this optimization problem with simultaneous perturbation stochastic approximation (SPSA), simultaneous annealing (SA) and a genetic algorithm (GA). It turned out that SPSA reached the stopping criteria after 383 iterations, where GA took more than 13000 and SA more than 9000 iterations. Moreover SPSA is easier to implement.

It is worth mentioning that the same authors also investigated other aerodynamic shape design optimization problems with the SPSA method in 2005 [45].

### 2.2.6 A Neural Network Application

Consider a neural network with $d$ weights to be optimized. Typically such a problem is solved by gradient descent. If not all sample data points are accessible, a stochastic gradient descent procedure is a typical choice. Now we go one step further and assume no knowledge of how the weights are connected. Thus, the exact function representing the neural network is not accessible. The reason could be that the network structure is too complicated or simply unknown. This means we cannot compute the gradient of our loss function directly and Kiefer-Wolfowitz type algorithms come into play. As mentioned in the introduction these algorithms require $2d$ observations per iteration step. Especially in high-dimensional online optimization problems the system might change faster than the corresponding weights can be estimated. For this purpose randomized stochastic approximation algorithms are a good choice. With the ideas in this thesis it is even possible to optimize problems where the randomization has some dependency restrictions.

## 2.3 Simulations

This section occupies with the comparison of Kiefer-Wolfowitz, one- and two-measurement algorithms. The simulations should give some recommendations when to use which procedure. Besides the almost sure convergence rate and the asymptotic $L^2$-error, which are simulated in the following sections, there is also the quality criterion pre-convergence. This is important as in real-world applications the number of iterations cannot be infinite. Hence it makes sense to also compare the procedures in the first few iteration steps.

The following plots consider stochastic approximation of the minimum of the functions $\mathbb{R}^3 \to \mathbb{R}\colon (x_1, x_2, x_3)^T \mapsto \sum_{i=1}^3 x_i^2$ for the two-measurement algorithm and a linearly continued variant

$$\mathbb{R}^3 \to \mathbb{R}\colon (x_1, x_2, x_3)^T \mapsto \sum_{i=1}^3 y_i \text{ with } y_i = \begin{cases} x_i^2 & \text{if } |x_i| \leqslant 1, \\ |x_i| & \text{else,} \end{cases}$$

$i \in \{1, 2, 3\}$, for the one-measurement algorithm, where the latter grows linearly outside the unit cube. The step sequences are chosen as $a_n = 2/(20+n)$, $c_n = 1/n^{1/6}$. We denote our stochastic approximation process by $Z = (Z_n)_{n\geqslant 0} = (Z_n^{(1)}, Z_n^{(2)}, Z_n^{(3)})^T_{n\geqslant 0}$. The starting value is $Z_0 = (-5, -5, -5)^T$. For each algorithm 10000 observations are made. The observation noise is Bernoulli distributed with values $\pm 1$. In order to keep the algorithms comparable, we do not update at each iteration step. For example, the classical Kiefer-Wolfowitz algorithm keeps the same values in each component for the first $2d = 6$ iterations. Then an update at every component is done, which is followed by freezing the values for another six steps. As the name implies, the two-measurement algorithms need two evaluations per iteration step. Hence we update it every second step. One-measurement algorithms are renewed in each step.

### 2.3.1 Comparison of One- and Two-measurement Algorithms and Kiefer-Wolfowitz

We begin with the simulation (Figure 2.1) of the classical Kiefer-Wolfowitz procedure

$$Z_{n+1} = Z_n - \frac{a_n}{2c_n} \{f(Z_n + c_n e_i) - f(Z_n - c_n e_i) + M_{n,i}\}_{i\in\{1,\dots,d\}}$$

in $\mathbb{R}^3$. Note that it needs $2d = 6$ function evaluations per update.

Next we turn to the two-measurement (RDSA) in Figure 2.2

$$Z_{n+1} = Z_n - \frac{a_n}{2c_n} D_n \left\{\left(f(Z_n + c_n D_n) - W_{n,1}\right) - \left(f(Z_n - c_n D_n) - W_{n,2}\right)\right\}$$

and the one-measurement random direction stochastic approximation algorithm (RDSA1) in Figure 2.3

$$Z_{n+1} = Z_n - \frac{a_n}{c_n} D_n \left\{\left(f(Z_n + c_n D_n) - W_n\right)\right\}.$$

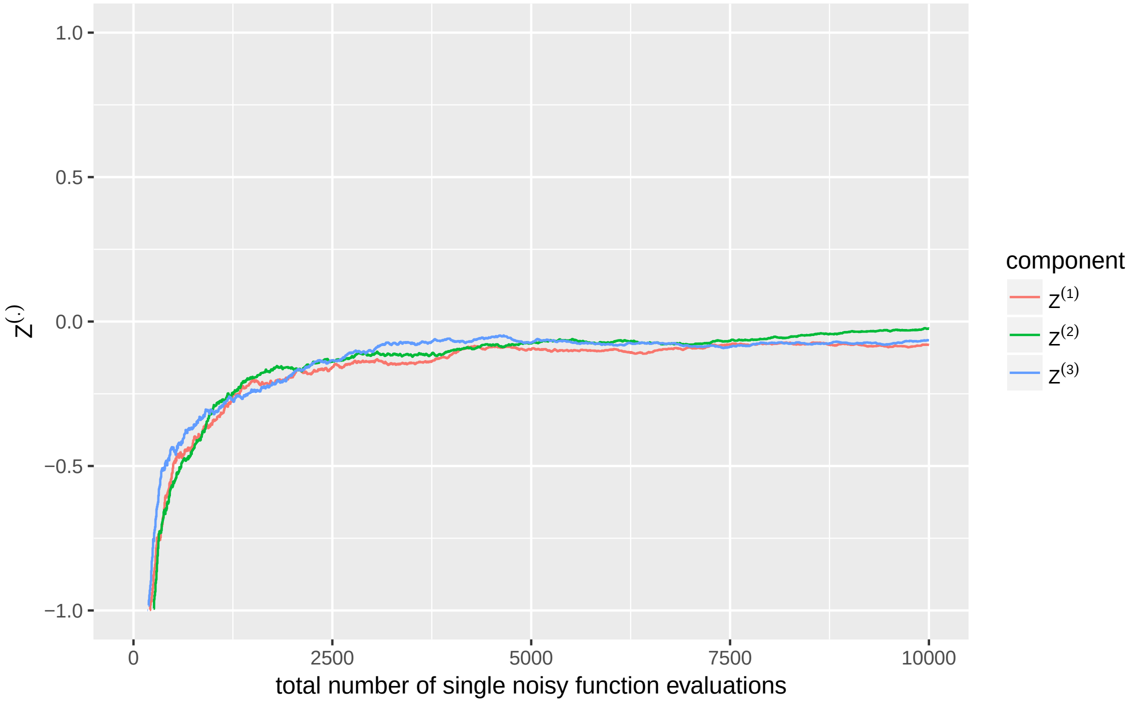


**Figure 2.1.** The three components of a simulated path of a regular Kiefer-Wolfowitz procedure in $\mathbb{R}^3$

These use an i.i.d. noise process. A stronger fluctuation of the one-measurement algorithm is observable. The lack of a second, symmetrical, evaluation which could give more stability yields a worse behaviour.

Dependent (Markovian) noise algorithms use a noise process on the vertices of a cube, with a random vertex as starting value. The process changes the sign in exactly one dimension $d \in \{1, 2, 3\}$. The probability for each dimension is $1/3$. The two-measurement version (RDSA) is presented in Figure 2.4. It's one-measurement counterpart (RDSA1) on the other hand (Figure 2.5) does not indicate to eventually converge when simulating 10000 evaluations. However using an alternative gain $a_n = 2/(200 + n)$, yields another impression (Figure 2.6). Nevertheless these plots show a drawback of one-measurement stochastic approximation procedures.

Another example for dependent perturbation is sampling without replacement from the directions

$$(1,1,1)^T, (-1,-1,-1)^T, (-1,1,1)^T, (1,-1,1)^T, (1,1,-1)^T, (-1,-1,1)^T,$$
$$(-1,1,-1)^T, (1,-1,-1)^T.$$

When all directions are chosen, the replacement experiment will be repeated. The one- and two-measurement simulations are given in Figures 2.7 and 2.8, respectively.

Deterministically perturbated algorithms start at a random starting value, but follow a deterministic rule: Starting with an arbitrary value one first changes the sign of the first dimension then the second, and so on. The period is of length $2d$. If we started the simulation with $D_0 = (1, 1, 1)^T$, $D$ in (RDSA) and (RDSA1) periodically

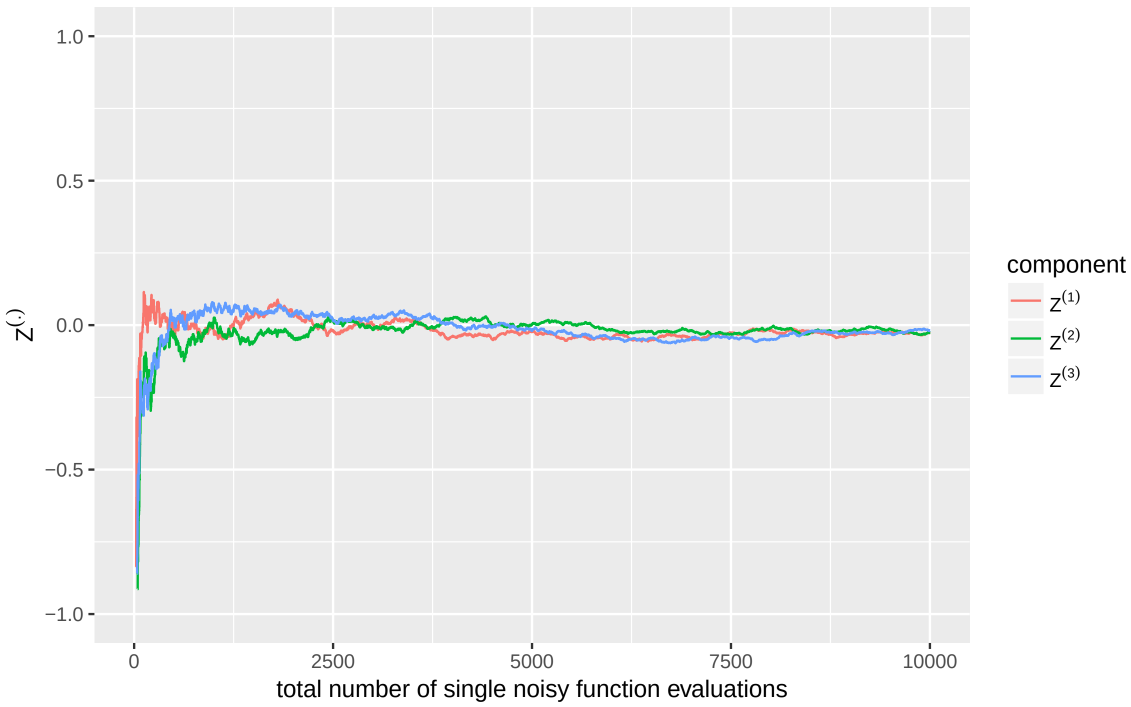


**Figure 2.2.** The three components of a simulated path of a two-measurement procedure with i.i.d. perturbation in $\mathbb{R}^3$

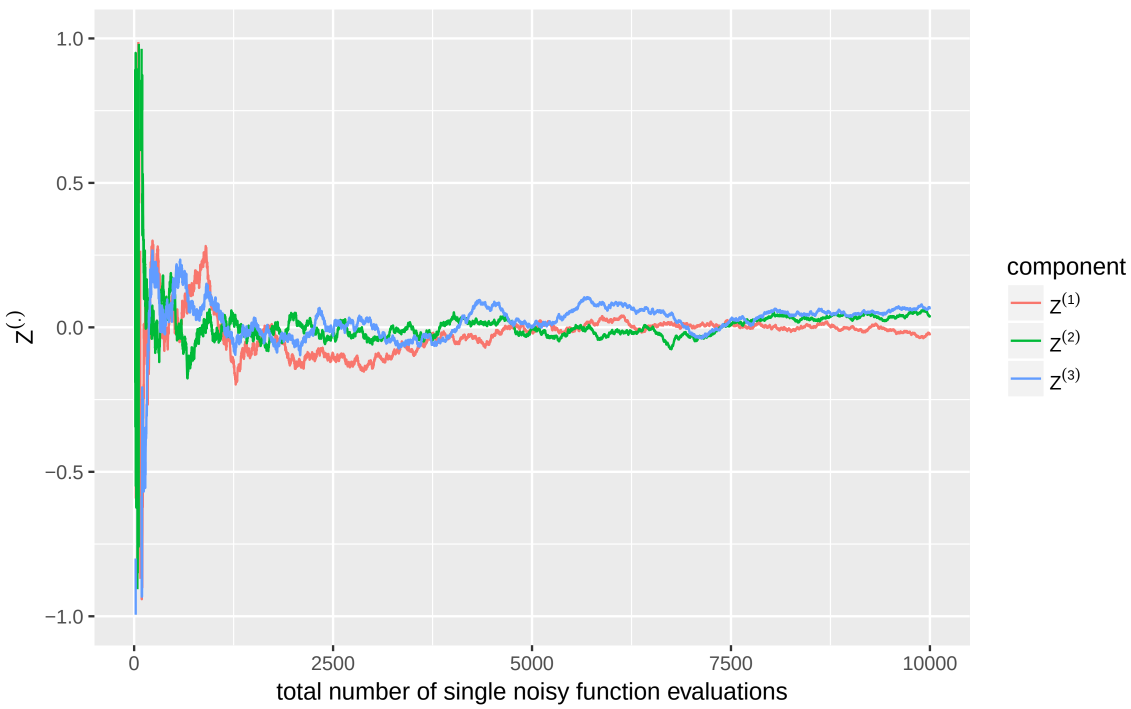


**Figure 2.3.** The three components of a simulated path of a one-measurement procedure with i.i.d. perturbation in $\mathbb{R}^3$

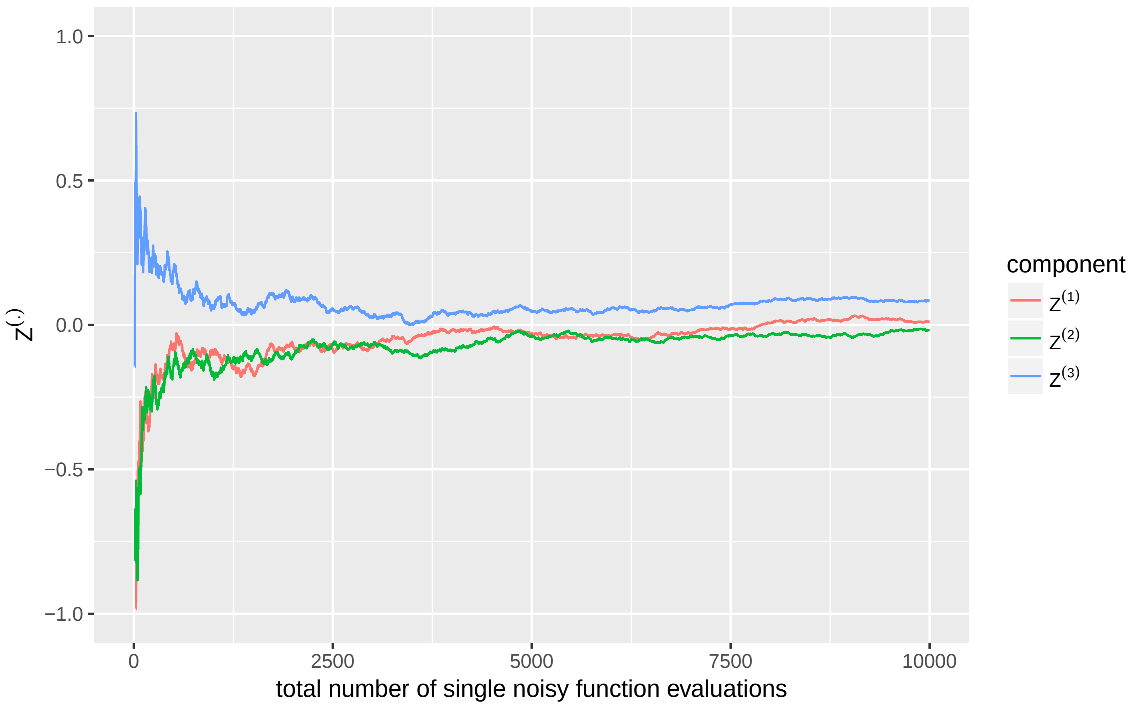


**Figure 2.4.** The three components of a simulated path of a two-measurement procedure with dependent perturbation

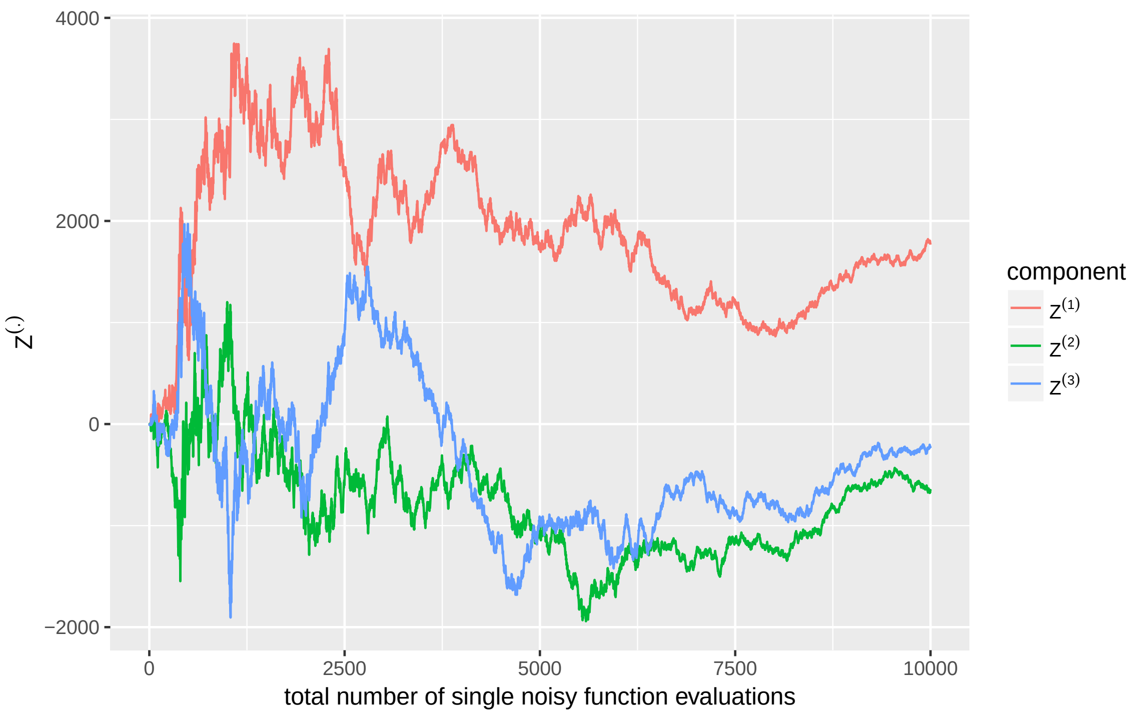


**Figure 2.5.** The three components of a simulated path of a one-measurement procedure with dependent perturbation

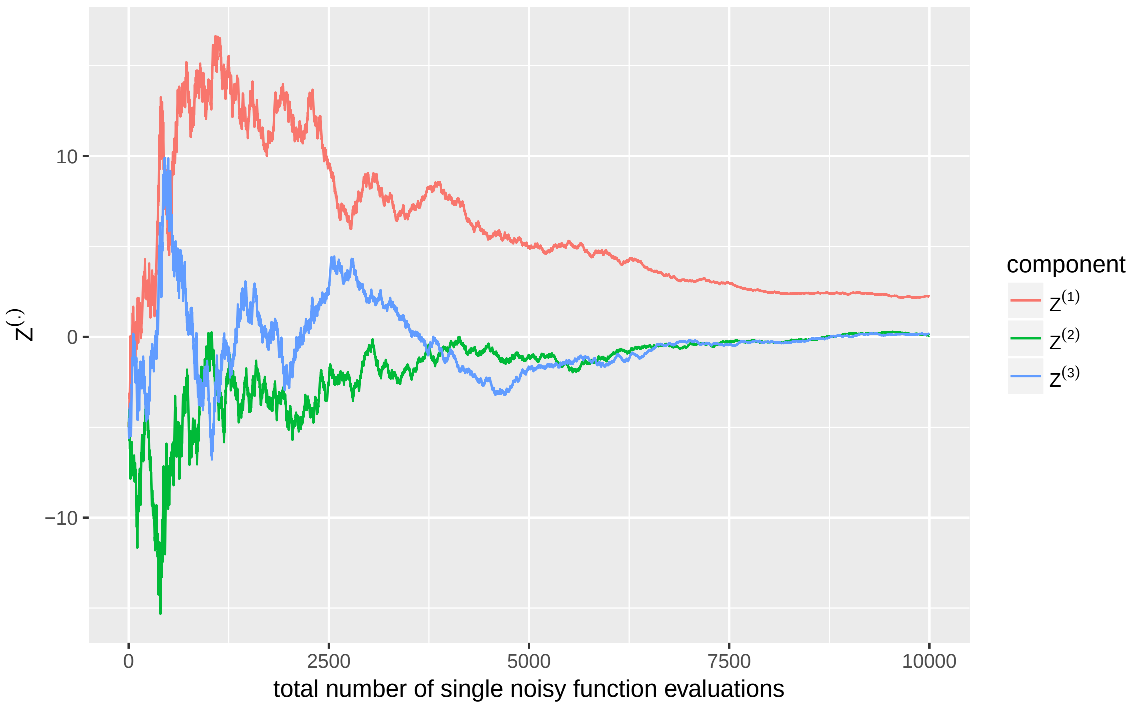


**Figure 2.6.** The three components of a simulated path of a one-measurement procedure with dependent perturbation and $a_n = 2/(200 + n)$

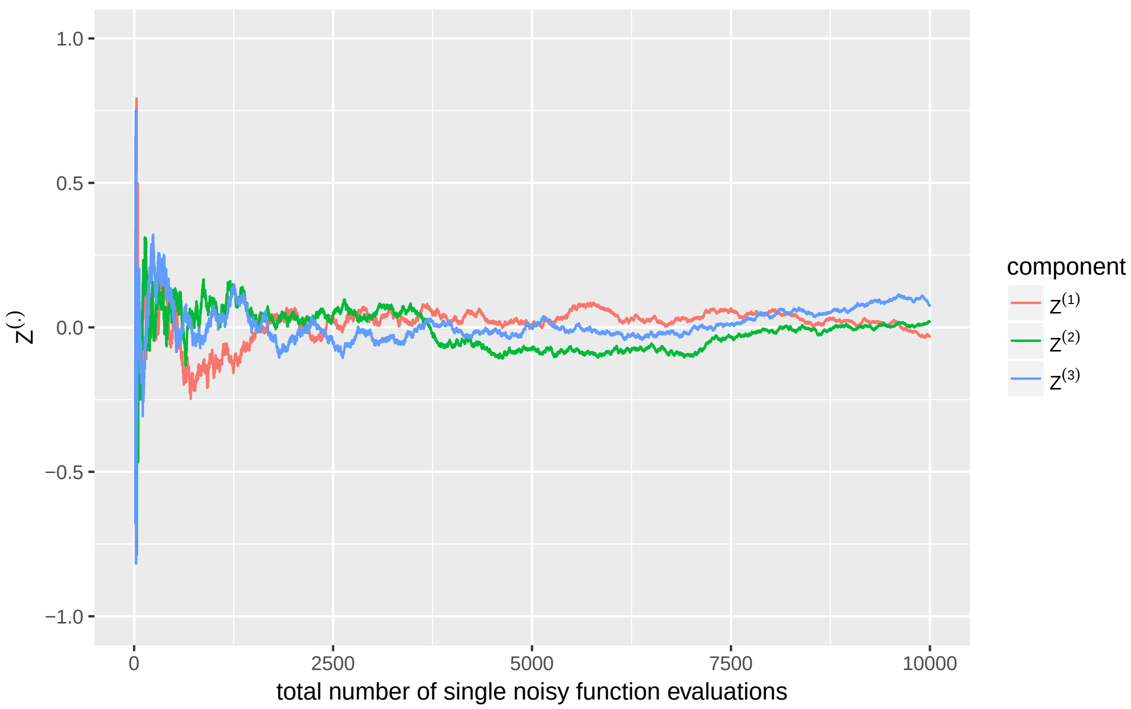


**Figure 2.7.** The three components of a simulated path of a one-measurement procedure with dependent perturbation without replacement

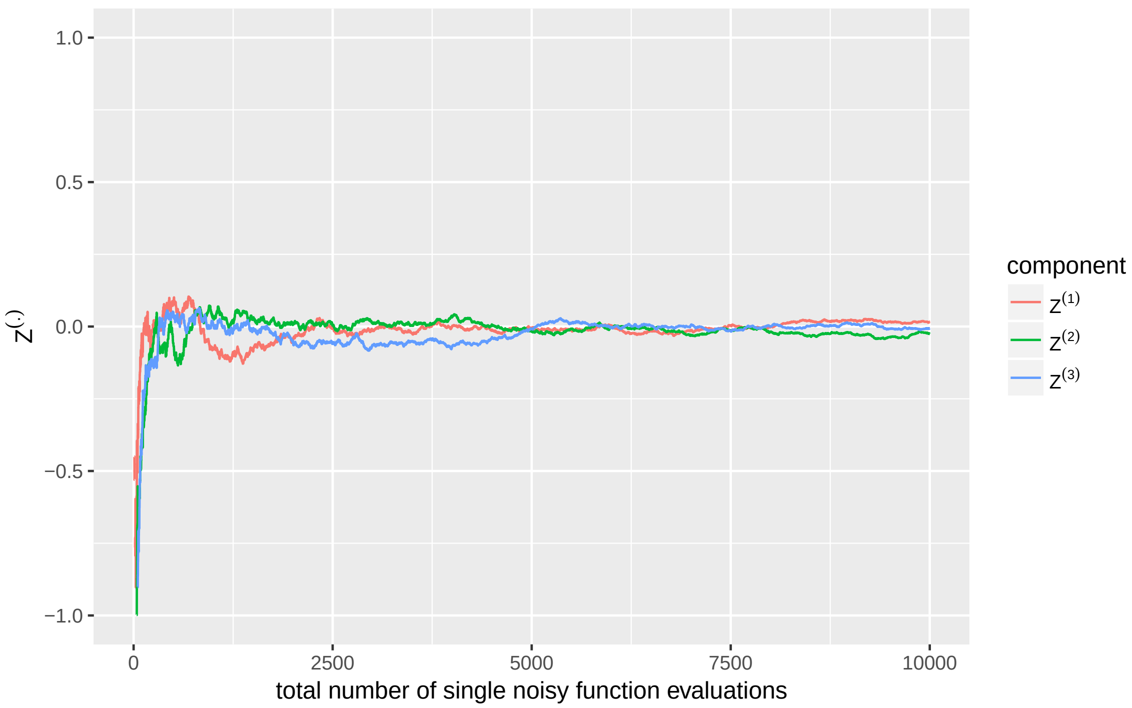


**Figure 2.8.** The three components of a simulated path of a two-measurement procedure with dependent perturbation without replacement

had the values

$$(1,1,1)^T, (-1,1,1)^T, (-1,-1,1)^T, (-1,-1,-1)^T, (1,-1,-1)^T, (1,1,-1)^T.$$

The corresponding simulations are given in Figures 2.9 and 2.10.

Now we present another procedure with deterministic directions which has a close similarity to the classical Kiefer-Wolfowitz algorithm and shall hence be called pseudo-Kiefer-Wolfowitz algorithm. It's simulation is shown in Figures 2.11 and 2.12. The direction is 6-periodic with $D$ in (RDSA) and (RDSA1) periodically having the values

$$(1,0,0)^T, (0,1,0)^T, (0,0,1)^T, (-1,0,0)^T, (0,-1,0)^T, (0,0,-1)^T.$$

That is to say $D_1 = (1,0,0)^T$, $D_2 = (0,1,0)^T$, until $D_7 = D_1$ and so on. In contrast to the regular Kiefer-Wolfowitz algorithm, each evaluation requires not $2d$ function evaluations but only one. Although this procedure looks similar to the classical Kiefer-Wolfowitz algorithm, it is actually nothing else but the deterministic procedure we simulated before with $D$ concentrated on the vertices of a rotated hypercube. For this reason we omit its simulation in the following sections.

The comparisons of one- and two-measurement algorithms show that the first ones are very sensitive about poorly chosen initial values $Z_0$. If additionally the random perturbations are showing into an unfavourable direction, the iterates $Z_n$ move away from the solution even further. Hence it is useful to investigate one-measurement algorithms when they already start in the point which they actually shall converge to. This is done in the following subsection. As an intermediate result, it seems

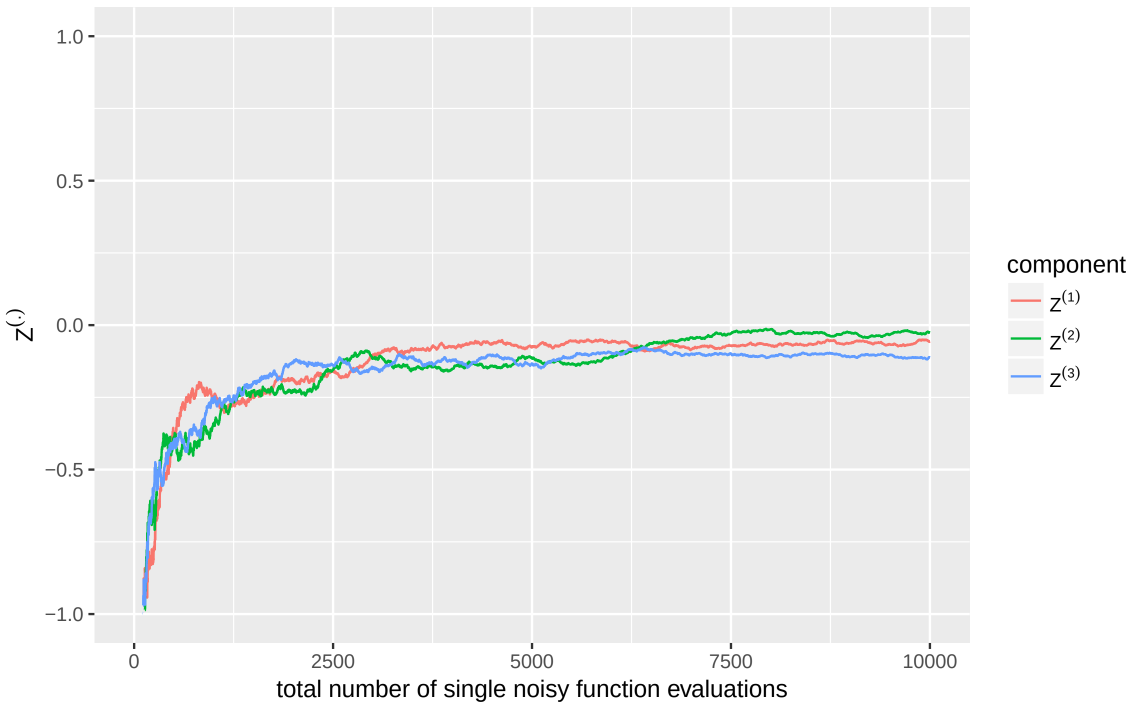


**Figure 2.9.** The three components of a simulated path of a two-measurement procedure with deterministic perturbation

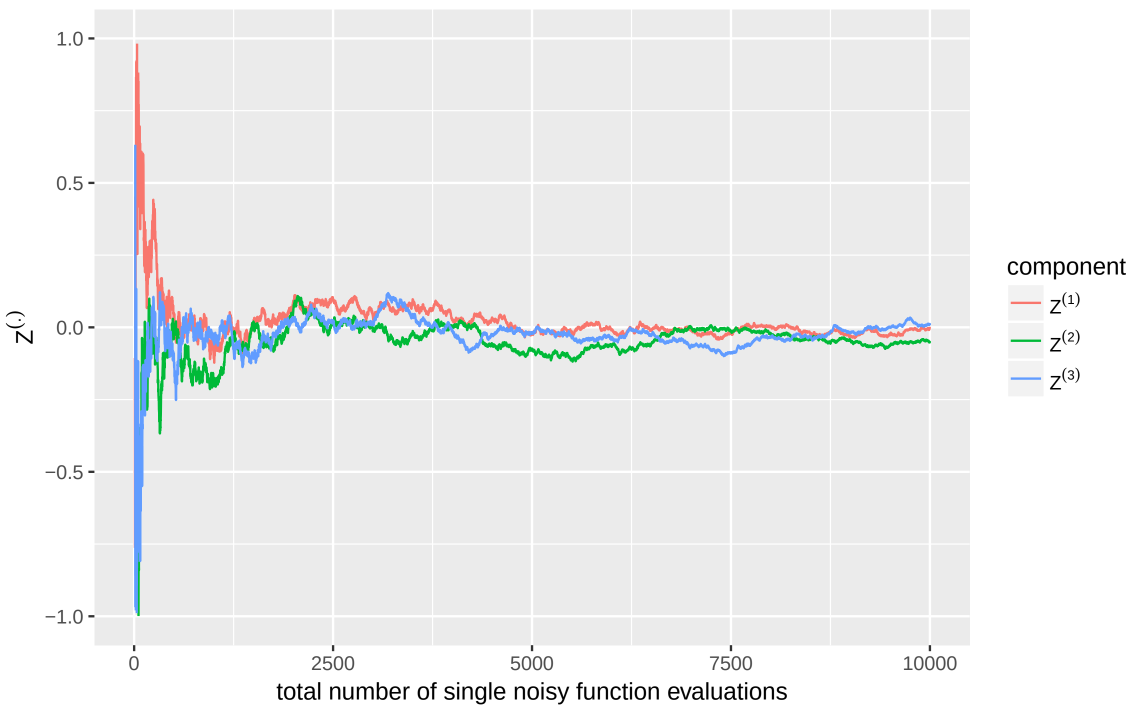


**Figure 2.10.** The three components of a simulated path of a one-measurement procedure with deterministic perturbation


**Figure 2.11.** The three components of a simulated path of a Pseudo Kiefer-Wolfowitz procedure (one-measurement)

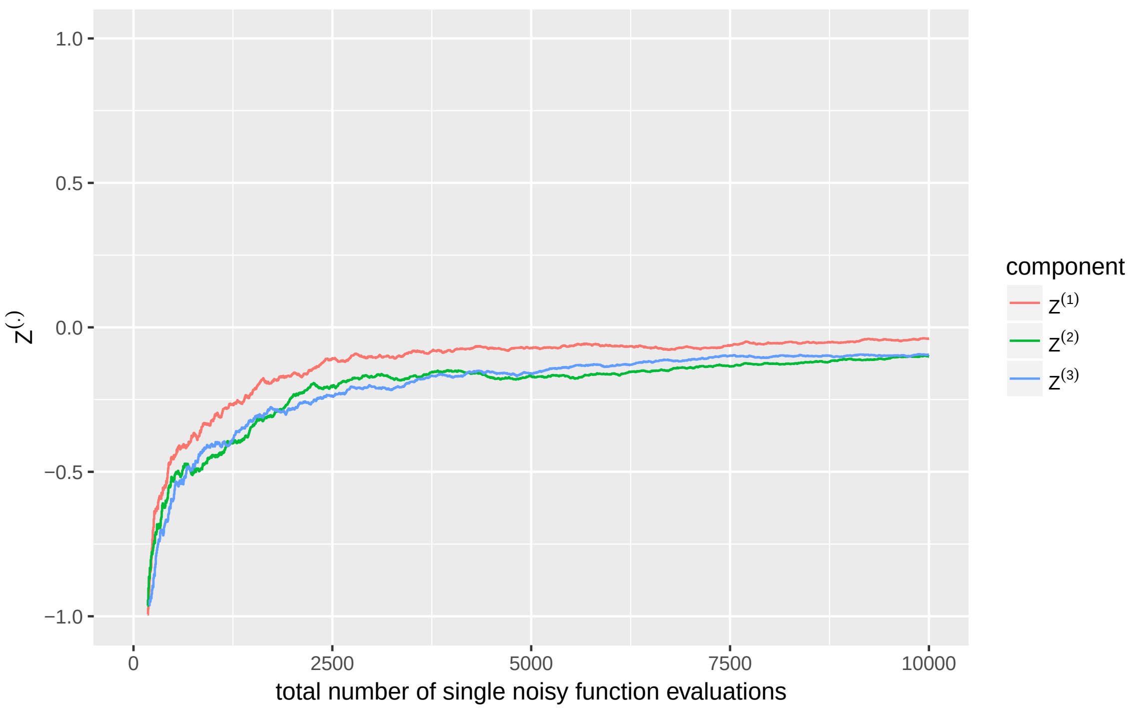


**Figure 2.12.** The three components of a simulated path of a Pseudo Kiefer-Wolfowitz procedure (two-measurement)

that two-measurement algorithms are more robust at the first few iteration steps, where one-measurement algorithms seem to have no benefit beyond requiring fewer evaluations per iteration step.

### 2.3.2 Comparison of One-measurement Algorithms Starting at the Extremum of $f$

Now we compare one-measurement algorithms with each other. In contrast to the previous simulations we start at $Z_0 = (0,0,0)^T$, i.e. the extremum we are actually searching for, and regard only 1000 evaluations. All other settings remain as before. We begin with i.i.d. (Figure 2.13) and Markovian (Figure 2.13) perturbation settings for which $\mathbb{E}\|\sum_{k=0}^{n-1} D_k\|^2 = \mathcal{O}(n)$ holds true.

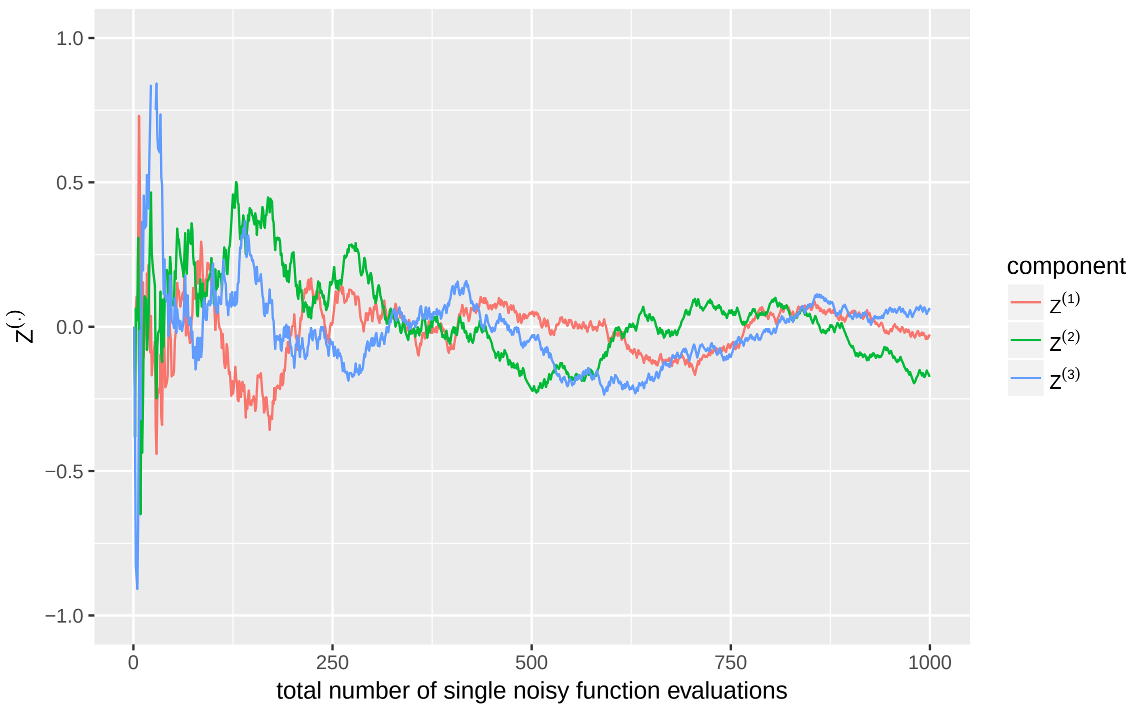


**Figure 2.13.** The three components of a simulated path of a one-measurement procedure with i.i.d. perturbation

Next we turn to algorithms for which $\mathbb{E}\|\sum_{k=0}^{n-1} D_k\|^2 = \mathcal{O}(1)$ holds. This is the case for sampling without replacement (Figure 2.15) and deterministic (Figure 2.16) perturbation. It is observable that the latter two algorithms behave better at the first evaluation steps. This becomes particularly obvious if one looks the paths up to step 250.

The simulation of one single path gives a first impression on its behaviour though it is not very representative for the whole process. For that reason in the next subsection its empirical $L^2$-error is investigated.

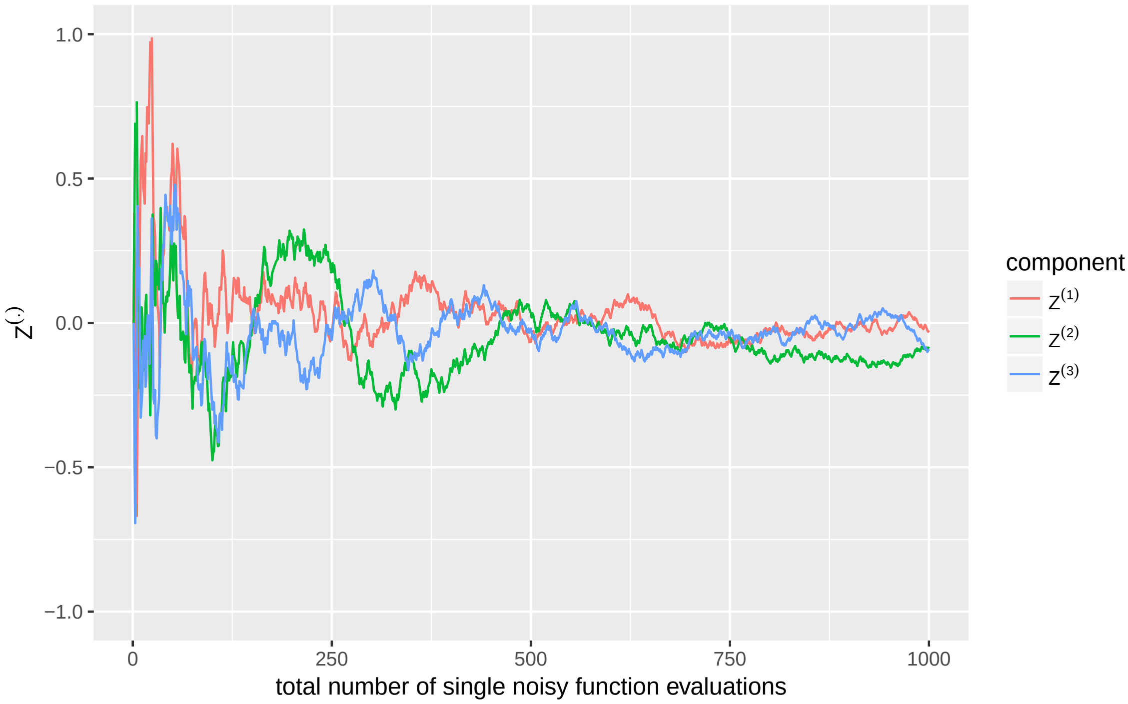


**Figure 2.14.** The three components of a simulated path of a one-measurement procedure with Markovian perturbation


**Figure 2.15.** The three components of a simulated path of a one-measurement procedure with dependent (sample without replacement) perturbation

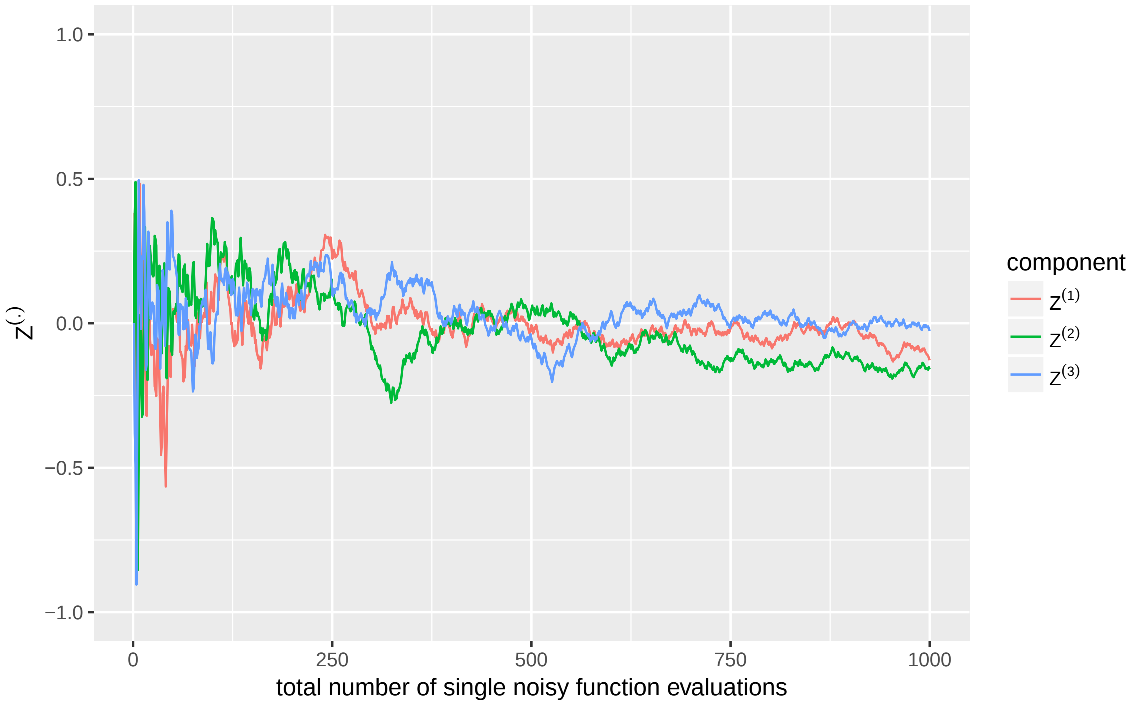


**Figure 2.16.** The three components of a simulated path of a one-measurement procedure with deterministic perturbation

### 2.3.3 $\mathbf{L^2}$-Convergence of Algorithms Starting at the Extremum of $f$

In order to provide a fair comparison of the individual algorithms, we estimate the $L^2$-error empirically. The parameters are $a_n = 2/(70+n)$, $c_n = 1/n^{1/6}$, and starting value $Z_0 = (0,0,0)^T$. For each algorithm $N = 1000$ paths with $n = 10000$ single observations, which equals the number of evaluation steps in one-measurement algorithms, were performed. We begin with the one-measurement procedures. The empirical $L^2$-errors of these paths are given in Figure 2.17.

This plot upholds the assumption that algorithms with $\mathbb{E}\|\sum_{k=0}^{n-1} D_k\|^2 = \mathcal{O}(1)$ have a better behaviour at the first evaluation steps than algorithms for which only $\mathbb{E}\|\sum_{k=0}^{n-1} D_k\|^2 = \mathcal{O}(n)$ only holds true.

Next (Figure 2.18) we apply the same setting to Kiefer-Wolfowitz and the two-measurement procedures. Note that similar to the previous simulations we freeze the iterations for two or six evaluations. In two-measurement procedures the rate of $\mathbb{E}\|\sum_{k=0}^{n-1} D_k\|^2$ seems to have no effect on the pre-asymptotic behaviour. This is not very surprising, as due to symmetry our theorems and corollaries (cf. Corollary 2.2.3) on these algorithms did not need a condition including $\mathbb{E}\|\sum_{k=0}^{n-1} D_k\|^2$ or $\mathbb{E}(D_k)$. Note that the pre-asymptotic behaviour of multi-measurement algorithms is generally better than that of one-measurement. Although at very few steps the Kiefer-Wolfowitz procedure has lower error, this changes after about 1000 evaluations. This is due to the fact that the process is frozen three times as long as the the other simulations in this plot. If we only compared with the total number of iterations without regarding the evaluations per step, this would look completely different.

In the following chapter we occupy with the almost sure convergence rate of random-

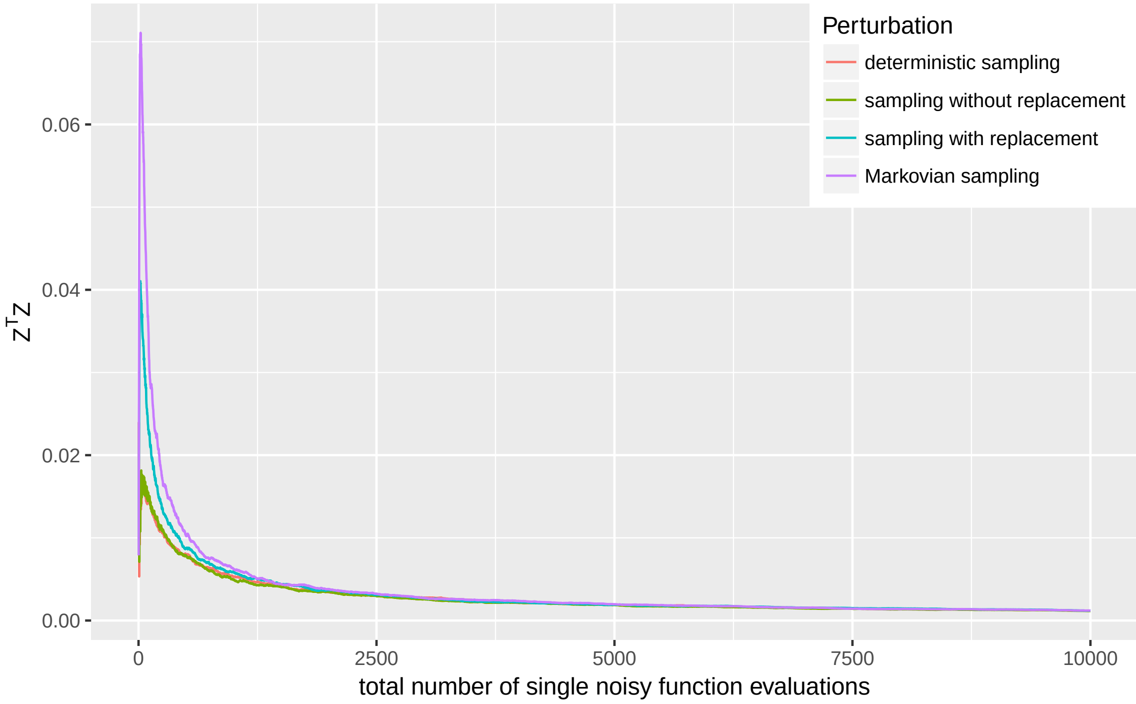


**Figure 2.17.** Empirical $L^2$-error of $Z_n$ generated by one-measurement procedures

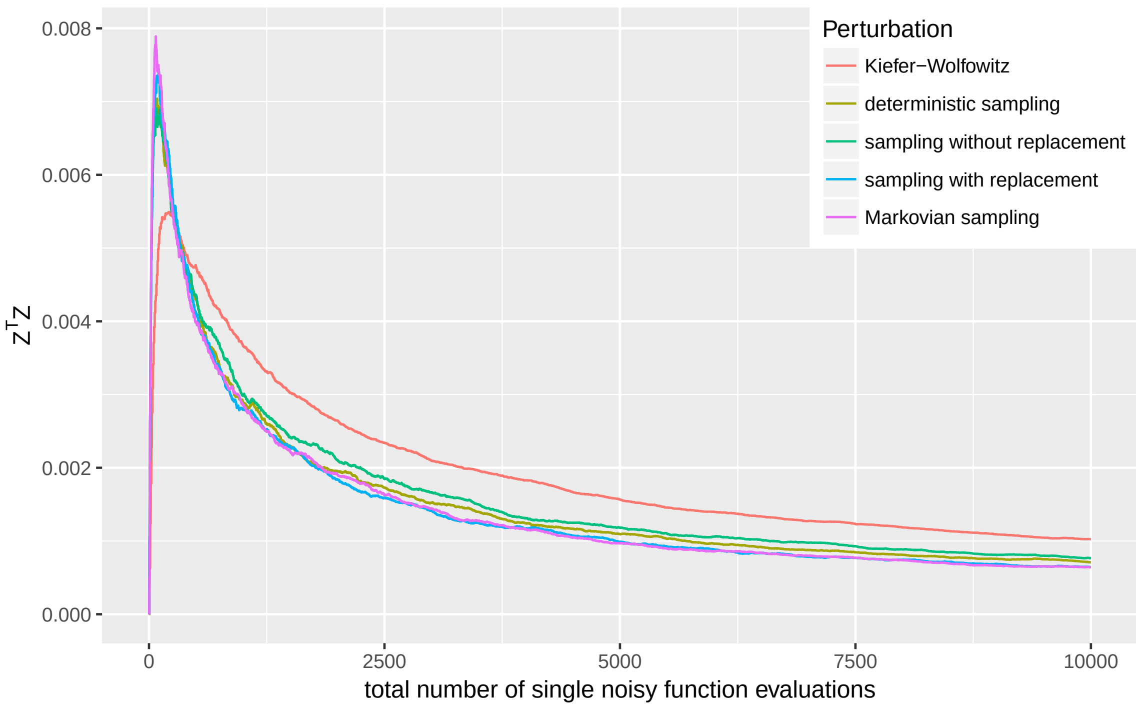


**Figure 2.18.** Empirical $L^2$-error of $Z_n$ generated by Kiefer-Wolfowitz and two-measurement procedures

ized algorithms. The associated simulations yield more insights into which procedure to prefer in which situation. Apart form already visible pre-asymptotic differences the plots of the $L^2$-converge rate can point out differences for large numbers of iterations. Moreover it should be noted that the asymptotic $L^2$-error is also dependent on the form of $f$. In order to exclude the possibility that the simulated behaviour changes for a different $f$ we would need more advanced results like asymptotic normality.

# 3 Almost Sure Convergence Rate

After the verification of consistency, the question arises how fast the process $(Z_t)_{t\geqslant 0}$ defined in (1.3) converges towards the minimizing point $z^*$ of $\tilde{f}_0$. Again we present a general framework for semimartingales and deduce special cases.

## 3.1 A General Semimartingale Algorithm

We define $\gamma_t(\delta) := \mathcal{E}_t(\delta \int_0^{\cdot} a_s \mathrm{d}R_s)$, where $\mathcal{E}_t(.)$ is the stochastic exponential, and investigate how $\delta$ can be chosen such that

$$\gamma_t(\delta)\|Z_t - z^*\| \to 0 \text{ a.s.}$$

can be assured. Note that the stochastic exponential is the solution of $Z_t = 1 + \int_0^t Z_{s-}\mathrm{d}X_s,\ X_0 = 0$ which is given by $\mathcal{E}_t(X) := \exp\left(X_t - \frac{1}{2}[X,X]_t\right)\prod_{0<s\leqslant t}(1 + \Delta X_s)\exp\left(-\Delta X_s + \frac{1}{2}(\Delta X_s)^2\right)$. In that context we need additional assumptions that are given below.

For the sake of simpler proofs let $D_{s-}$ as well as all functions which are only dependent on $D_{s-}$ be bounded. The function $F\colon \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ is called *$p$-smooth* at $(c^*, z^*)$ if for all $\mathtt{d} \in \mathbb{R}^d$

$$\left\|\nabla_c F(\mathtt{d}, c, z) - \sum_{\substack{n_1\colon |n_1|\leqslant m_1 \\ n_2\colon n_2 \leqslant m_2 \\ m_1+m_2\leqslant \lfloor p\rfloor -1}} \frac{1}{n_1! n_2!}\nabla_z^{n_1}\nabla_c^{n_2}\nabla_c F(\mathtt{d}, c^*, z^*)(z-z^*)^{n_1}(c-c^*)^{n_2}\right\|$$
$$= o\Big(\|z - z^*\|^{m_1+\epsilon_1}\Big) + o\Big(|c - c^*|^{m_2+\epsilon_2}\Big)$$

with $\epsilon_1, \epsilon_2 \in [0,1)$ and $m_1+\epsilon_1+m_2+\epsilon_2 = p-1$ holds, and $\mathbb{E}\|g_i(D_s)\|^2 < \infty$ for $i \in \{1,2\}$ and any $s \in [0,\infty)$. Note that this definition employs the multi-index notation which was already defined in Section 1.2.

Now we extend the general Assumption 1.3.1 to $p$-smooth functions $F$.

**Assumption 3.1.1.**

- *$F$ is* factorizable *at $c = 0$ with respect to $\mathtt{d}$ and $z$ in the sense that there are measurable functions $\tilde{f}_k\colon \mathbb{R}^d \to \mathbb{R}^{d^k}$, $g_k\colon \mathbb{R}^d \to \mathbb{R}^{d^{k+1}}$ such that*

$$\nabla_z^l \nabla_c^k F(\mathtt{d}, 0, z) = g_k(\mathtt{d})\nabla_z^l \tilde{f}_k(z) \quad \textit{for } l \in \{0,1\} \textit{ and } l+k \in \{0,\ldots,\lfloor p\rfloor\}. \tag{3.1}$$

- *$F$ is* affine *in the sense of*

$$\nabla_c^k F(\mathtt{d}, c, z) = \nabla_c^k F(\mathtt{d}, 0, z + c\mathtt{d}) \text{ for } k \in \{0, \dots, \lfloor p \rfloor\}. \tag{3.2}$$

As mentioned before, $D_{s-}$ as well as all $g_k(D_{s-})$ with $k \in \{0, \dots, \lfloor p \rfloor\}$ are bounded. Define $\nabla_z \tilde{f}_1(z^*) =: H_{z^*}$. In applications later, we observe that $H_{z^*}$ coincides with the Hessian of a function $f$ at $z^*$. This condition will be naturally fulfilled in all our applications. Its largest and smallest eigenvalues are denoted by $\lambda_{\max}$ or $\lambda_{\min}$, respectively.

We formulate the following assumptions.

**Assumption 3.1.2.** *Let Assumption 3.1.1 and conditions (A)–(G) in Assumption 2.1.1 hold. Assume $F\colon \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ p-smooth with $p \geqslant 2$ at $(c^*, z^*)$.*

*($\widetilde{D}$)*

$$\int_0^\infty \gamma_{s-}(\delta) a_s c_s^{p-1} \mathrm{d}R_s < \infty$$

*($\widetilde{F}$)*

$$\Big| \int_0^\infty \gamma_{s-}(\delta) a_s \sum_{m=0}^{\lfloor p \rfloor - 2} \Big( g_{m+2}(D_{s-}) - \mathbb{E}\big(g_{m+2}(D_{s-})\big) \Big) \tilde{f}_{m+2}(z^*) c_s^{m+1} \mathrm{d}R_s \Big| < \infty$$

*($\widetilde{G}$)*

$$\Big| \int_0^\infty \frac{1}{1 + \gamma_{s-}^2(\delta) \|Z_{s-}\|^2} \gamma_{s-}^2(\delta) \frac{a_s}{c_s} \big\langle Z_{s-}, F(D_{s-}, 0, Z_{s-}) \big\rangle \mathrm{d}R_s \Big| < \infty$$

*and let for every $i \in \{1, \dots, d\}$*

$$\int_0^\infty \frac{a_s^2}{c_s^2} \frac{\gamma_{s-}^2(\delta) \|F(D_{s-}, 0, Z_{s-})\|^2}{1 + \gamma_{s-}^2(\delta) \|Z_{s-}\|^2} \Delta R_s \mathrm{d}R_s^d < \infty.$$

*If $p \geqslant 3$ assume $\mathbb{E}\left(g_{m-1}(D_{s-})\right) \tilde{f}_{m-1}(z^*) = 0$ for any $m \in \{3, \dots, \lfloor p \rfloor\}$.*

*($\widetilde{H}$) For every $i \in \{1, \dots, d\}$ and all $z \in \mathbb{R}^d$ let*

$$\int_0^\infty \frac{a_s^2}{c_s^2} \frac{\gamma_{s-}^2(\delta) h_s^{ii}(Z_{s-})}{1 + \gamma_{s-}^2(\delta) \|Z_{s-}\|^2} \mathrm{d}R_s < \infty \quad \text{where} \quad h_s^{ii}(z) := \frac{\mathrm{d}[\int_0^{\cdot} (M(\mathrm{d}t, D_{t-}, z))_i]_s}{\mathrm{d}R_s}.$$

The following theorem and its associated corollaries are new.

**Theorem 3.1.1.** *Let Assumption 3.1.2 hold. Then for all $\delta \in [0, \lambda_{min})$ the solution $Z$ of algorithm* (1.3) *satisfies*

$$\gamma_t(\delta) \|Z_t - z^*\| \overset{t \to \infty}{\longrightarrow} 0 \ a.s.$$

*Proof.* Without loss of generality, let $z^* = 0$. The idea of the proof is similar to that of Theorem 2.1.1. We investigate $\langle \gamma_t(\delta)Z_t, \gamma_t(\delta)Z_t\rangle$ instead of $\|Z_t\|^2$. With integration by parts and Taylor expansions, we find a decomposition that is handled with Lemma A.1.1. Integration by parts yields

$$\begin{aligned}\gamma_t^2(\delta) = \gamma_t(\delta)\gamma_t(\delta) &= \mathcal{E}_t\left(2\delta\int_0^\cdot a_s \mathrm{d}R_s + \int_0^\cdot \delta^2 a_s^2 \mathrm{d}[R,R]_s\right)\\ &= \mathcal{E}_t\left(2\delta\int_0^\cdot a_s \mathrm{d}R_s + \delta^2\int_0^\cdot a_s^2 \Delta R_s \mathrm{d}R_s^d\right)\end{aligned}$$

as well as

$$\mathrm{d}\gamma_s^2(\delta) = \gamma_{s-}^2(\delta)\left(2\delta a_s \mathrm{d}R_s + \delta^2 a_s^2 \Delta R_s \mathrm{d}R_s\right).$$

Using integration by parts as well as Lemma A.1.6 results in

$$\begin{aligned}&\gamma_t^2(\delta)\langle Z_t, Z_t\rangle - \gamma_0^2(\delta)\langle Z_0, Z_0\rangle\\
&\quad= \int_0^t \gamma_{s-}^2(\delta)\mathrm{d}\langle Z_s, Z_s\rangle + \int_0^t \langle Z_{s-}, Z_{s-}\rangle \mathrm{d}\gamma_s^2(\delta) + \int_0^t \mathrm{d}[\gamma^2(\delta), \langle Z, Z\rangle]_s\\
&\quad= \int_0^t \gamma_{s-}^2(\delta)\mathrm{d}\langle Z_s, Z_s\rangle + \int_0^t \langle Z_{s-}, Z_{s-}\rangle \mathrm{d}\gamma_s^2(\delta) + \int_0^t \Delta\gamma_s^2(\delta)\mathrm{d}\langle Z, Z\rangle_s\\
&\quad= \int_0^t \gamma_{s-}^2(\delta)\mathrm{d}\langle Z_s, Z_s\rangle + \int_0^t \langle Z_{s-}, Z_{s-}\rangle \mathrm{d}\gamma_s^2(\delta) + \int_0^t \left(\gamma_s^2(\delta) - \gamma_{s-}^2(\delta)\right)\mathrm{d}\langle Z, Z\rangle_s\\
&\quad= \int_0^t \gamma_s^2(\delta)\mathrm{d}\langle Z_s, Z_s\rangle + \int_0^t \langle Z_{s-}, Z_{s-}\rangle \mathrm{d}\gamma_s^2(\delta)\\
&\quad= -2\int_0^t \gamma_s^2(\delta)\frac{a_s}{c_s}\langle Z_{s-}, F(D_{s-}, c_s, Z_{s-})\rangle \mathrm{d}R_s\\
&\qquad+ \int_0^t \gamma_s^2(\delta)\frac{a_s^2}{c_s^2}\|F(D_{s-}, c_s, Z_{s-})\|^2 \Delta R_s \mathrm{d}R_s^d\\
&\qquad+ \int_0^t \gamma_s^2(\delta)\frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}[\int_0^\cdot (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i]_s + 2\delta\int_0^t a_s\gamma_{s-}^2(\delta)\|Z_{s-}\|^2 \mathrm{d}R_s\\
&\qquad+ \delta^2\int_0^t \gamma_{s-}^2(\delta)\|Z_{s-}\|^2 a_s^2 \Delta R_s \mathrm{d}R_s^d + \int_0^t \gamma_s^2(\delta)\mathrm{d}\widetilde{M}_s\end{aligned}$$

where $\mathrm{d}\widetilde{M}_s$ is given in (2.2). With the same arguments as in the proof of consistency $\int_0^t \gamma_s^2(\delta)\mathrm{d}\widetilde{M}_s \in \mathcal{M}_{\mathrm{loc}}$ follows. Now we have

$$\begin{aligned}\frac{1}{c_s}F(D_{s-}, c_s, Z_{s-}) &= \nabla_z\nabla_c F(D_{s-}, 0, 0)Z_{s-}\\ &\quad+ \underbrace{\nabla_c F(D_{s-}, 0, Z_{s-}) - \nabla_z\nabla_c F(D_{s-}, 0, 0)Z_{s-}}_{=:B_s} + C_s\\ &= g_1(D_{s-})\nabla_z \tilde{f}_1(0)Z_{s-} + B_s + C_s\end{aligned}$$

where $H_0 = \nabla_z \tilde{f}_1(0)$, $C_s := \frac{1}{c_s}F(D_{s-}, c_s, Z_{s-}) - \nabla_c F(D_{s-}, 0, Z_{s-})$.

We conclude

$$\frac{1}{c_s^2}\|F(D_{s-},c_s,Z_{s-})\|^2 \leqslant 3\|g_1(D_{s-})\|^2\lambda_{\max}^2\|Z_{s-}\|^2 + 3\|B_s\|^2 + 3\|C_s\|^2$$

and

$$\begin{aligned}
&-\frac{1}{c_s}\langle F(D_{s-},c_s,Z_{s-}),Z_{s-}\rangle \\
&\quad\leqslant -\langle Z_{s-}, g_1(D_{s-})\nabla_z\tilde{f}_1(0)Z_{s-}\rangle + \|B_s\|\|Z_{s-}\| - \langle C_s, Z_{s-}\rangle \\
&\quad= -\lambda_{\min}\Big\langle Z_{s-}, \big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\big)Z_{s-}\Big\rangle \\
&\qquad - \lambda_{\min}\|Z_{s-}\|^2 + \|B_s\|\|Z_{s-}\| - \langle C_s, Z_{s-}\rangle.
\end{aligned}$$

A Taylor expansion at $(0,0)$ yields the following asymptotic behaviour of $B_s$:

$$\begin{aligned}
\|B_s\| &= \Big\|\nabla_c F(D_{s-},0,Z_{s-}) - \nabla_z\nabla_c F(D_{s-},0,0)Z_{s-}\Big\| \\
&= \Big\|\nabla_c F(D_{s-},0,0) + \nabla_z\nabla_c F(D_{s-},0,0)Z_{s-} - \nabla_z\nabla_c F(D_{s-},0,0)Z_{s-} \\
&\quad + g_1(D_{s-})o(\|Z_{s-}\|)\Big\| \\
&= o(\|Z_{s-}\|).
\end{aligned}$$

Furthermore $\|B_s\|^2 = o(\|Z_{s-}\|^2)$. This holds true for any smoothness order $p \geqslant 2$ of $F$.

The investigation of $C_s$ depends on the smoothness of $F$ at $(c^*,z^*)$. In the case where $F$ is $p$-smooth with $p \in [2,3)$, it is already known from the proof of consistency, namely equations (2.3) and (2.4), that

$$-C_s \leqslant \vec{1}\|g_1(D_{s-})\|\|D_{s-}\|Lc_s - \frac{1}{c_s}F(D_{s-},0,Z_{s-}),$$

with $\vec{1} := (1,\dots,1)^T$ and

$$\|C_s\|^2 \leqslant \frac{2}{3}d\|g_1(D_{s-})\|^2\|D_{s-}\|^2L^2c_s^2 + \frac{2}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2.$$

If $F$ is $p$-smooth with $p \geqslant 3$, there exists a $\theta \in [0,1]$ such that

$$\begin{aligned}
C_s &= \frac{1}{c_s}\Big(F(D_{s-},c_s,Z_{s-}) - c_s\nabla_c F(D_{s-},0,Z_{s-})\Big) \\
&= \frac{1}{c_s}\Big(F(D_{s-},0,Z_{s-}) + c_s\nabla_c F(D_{s-},0,Z_{s-}) + \frac{1}{2}c_s^2\nabla_c^2 F(D_{s-},\theta c_s,Z_{s-}) \\
&\quad - c_s\nabla_c F(D_{s-},0,Z_{s-})\Big) \\
&= \frac{1}{c_s}\Big(F(D_{s-},0,Z_{s-}) + \frac{1}{2}c_s^2\nabla_c^2 F(D_{s-},\theta c_s,Z_{s-})\Big) \\
&= \frac{1}{c_s}F(D_{s-},0,Z_{s-}) + \frac{1}{2}c_s\nabla_c^2 F(D_{s-},\theta c_s,Z_{s-})
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{c_s} F(D_{s-},0,Z_{s-}) + \frac{1}{2} c_s \Big( \nabla_c^2 F(D_{s-},0,0) + \nabla_z \nabla_c^2 F(D_{s-},0,0) Z_{s-} + o(\|Z_{s-}\|) \\
&\quad + \sum_{m=1}^{\lfloor p \rfloor -2} \frac{1}{m!} \nabla_c^{m+2} F(D_{s-},0,0) \theta^m c_s^m + o(\theta^{p-2} c_s^{p-2}) \Big) \\
&= \frac{1}{c_s} F(D_{s-},0,Z_{s-}) + \frac{1}{2} c_s \Big( g_2(D_{s-}) \Big( \tilde{f}_2(0) + \nabla_z \tilde{f}_2(0) Z_{s-} + o(\|Z_{s-}\|) \Big) \\
&\quad + \sum_{m=1}^{\lfloor p \rfloor -2} \frac{1}{m!} g_{m+2}(D_{s-}) \tilde{f}_{m+2}(0) \theta^m c_s^m + g_{\lfloor p \rfloor -2}(D_{s-}) o(\theta^{p-2} c_s^{p-2}) \Big) \\
&= \frac{1}{c_s} F(D_{s-},0,Z_{s-}) + g_2(D_{s-}) \mathcal{O}(c_s \|Z_{s-}\|) \\
&\quad + \sum_{m=0}^{\lfloor p \rfloor -2} g_{m+2}(D_{s-}) \tilde{f}_{m+2}(0) \mathcal{O}(c_s^{m+1}) + o(c_s^{p-1})
\end{aligned}
$$

and

$$
\begin{aligned}
\|C_s\|^2 &\leqslant \frac{2}{c_s^2} \|F(D_{s-},0,Z_{s-})\|^2 + c_s^2 \|\nabla_c^2 F(D_{s-},\theta c_s, Z_{s-})\|^2 \\
&\leqslant \frac{2}{c_s^2} \|F(D_{s-},0,Z_{s-})\|^2 \\
&\quad + 5 c_s^2 \Big( \|\nabla_c^2 F(D_{s-},0,0)\|^2 + \|\nabla_z \nabla_c^2 F(D_{s-},0,0) Z_{s-}\|^2 + o(\|Z_{s-}\|^2) \\
&\quad + \Big\| \sum_{m=1}^{\lfloor p \rfloor -2} \frac{1}{m!} \nabla_c^{m+2} F(D_{s-},0,0) \theta^m c_s^m \Big\|^2 + o(\theta^{p-2} c_s^{p-2}) \Big) \\
&= \frac{2}{c_s^2} \|F(D_{s-},0,Z_{s-})\|^2 + \mathcal{O}(c_s^2) + \mathcal{O}(c_s^2 \|Z_{s-}\|^2) + o(c_s^2 \|Z_{s-}\|^2) + \mathcal{O}(c_s^4) \\
&= \frac{2}{c_s^2} \|F(D_{s-},0,Z_{s-})\|^2 + \mathcal{O}(c_s^2) + \mathcal{O}(c_s^2 \|Z_{s-}\|^2).
\end{aligned}
$$

Note that for two-measurement algorithms like (1.6) it actually holds true that

$$
\|C_s\|^2 = \frac{2}{c_s^2} \|F(D_{s-},0,Z_{s-})\|^2 + \mathcal{O}(c_s^4) + \mathcal{O}(c_s^2 \|Z_{s-}\|^2).
$$

For one-measurement algorithms this would only be guaranteed by an additional assumption like $\tilde{f}_2(0) = 0$ and thereby $\|\nabla_c^2 F(D_{s-},0,0)\|^2 = 0$. We find

$$
C_s = \begin{cases} \frac{1}{c_s} F(D_{s-},0,Z_{s-}) + \mathcal{O}(c_s) & \text{if } F \text{ is } p\text{-smooth with } p \in [2,3) \\ \frac{1}{c_s} F(D_{s-},0,Z_{s-}) + \mathcal{O}(c_s \|Z_{s-}\|) & \\ \quad + \sum_{m=0}^{\lfloor p \rfloor -2} g_{m+2}(D_{s-}) \tilde{f}_{m+2}(0) \mathcal{O}(c_s^{m+1}) & \\ \quad + o(c_s^{p-1}) & \text{if } F \text{ is } p\text{-smooth with } p \geqslant 3. \end{cases}
$$

We seek for predictable processes $(A_t^1)_{t \geqslant 0}$ and $(A_t^2)_{t \geqslant 0}$ of finite variation with

$$
\gamma_t^2(\delta) \langle Z_t, Z_t \rangle - \gamma_0^2(\delta) \langle Z_0, Z_0 \rangle \leqslant A_t^1 - A_t^2 + N_t,
$$

where $N_t \in \mathcal{M}^2_{\text{loc}}$. Using the asymptotic behaviour of $B_s$ and $C_s$, we find that the following decomposition makes sense. Choose

$$A_t^2 := 2\int_0^t \gamma_s^2(\delta) a_s \Big( (\delta - \lambda_{\min})\|Z_{s-}\|^2 + \|B_s\|\|Z_{s-}\| \Big)^- \mathrm{d}R_s$$

$$\begin{aligned}
A_t^1 := {} & 2\int_0^t \gamma_s^2(\delta) a_s \Big( (\delta - \lambda_{\min})\|Z_{s-}\|^2 + \|B_s\|\|Z_{s-}\| \Big)^+ \mathrm{d}R_s \\
& + 2\left|\lambda_{\min} \int_0^t \gamma_s^2(\delta) a_s \big\langle Z_{s-}, \big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\big) Z_{s-} \big\rangle \mathrm{d}R_s \right| \\
& + 3\int_0^t \gamma_s^2(\delta) a_s^2 \|B_s\|^2 \Delta R_s \mathrm{d}R_s^d + (3\|g_1(D_{s-})\|^2 \lambda_{\max}^2 + \delta^2) \int_0^t \gamma_s^2(\delta) a_s^2 \|Z_{s-}\|^2 \Delta R_s \mathrm{d}R_s^d \\
& + 2\left| \int_0^t \gamma_s^2(\delta) a_s \langle C_s, Z_{s-}\rangle \mathrm{d}R_s \right| + 3\int_0^t \gamma_s^2(\delta) a_s^2 \|C_s\|^2 \Delta R_s \mathrm{d}R_s^d \\
& + \int_0^t \gamma_s^2(\delta) \frac{a_s^2}{c_s^2} \sum_{i=1}^d \mathrm{d}\Big[\int_0^{\cdot} (M(\mathrm{d}\tau, D_{\tau-}, Z_{\tau-}))_i\Big]_s
\end{aligned}$$

and

$$N_t := \int_0^t \gamma_t^2(\delta) \mathrm{d}\widetilde{M}_s,$$

with $\mathrm{d}\widetilde{M}_s$ as given in (2.2).

We are now prepared to prove

$$\int_0^\infty \frac{1}{1 + \gamma_{s-}^2(\delta)\langle Z_{s-}, Z_{s-}\rangle} \mathrm{d}A_s^1 < \infty. \tag{3.3}$$

A quick calculation yields

$$\begin{aligned}
\gamma_t(\delta) &= \exp(\delta \tilde{a}_t \Delta R_t)\Big(\exp\Big(\delta \int_0^{t-} \tilde{a}_s \mathrm{d}R_s\Big) \prod_{0<s<t} (1 + \delta \tilde{a}_s \Delta R_s) \exp(-\delta \tilde{a}_s \Delta R_s)\Big) \\
&\quad \cdot (1 + \delta \tilde{a}_t \Delta R_t) \exp(-\delta \tilde{a}_t \Delta R_t) \\
&= \gamma_{t-}(\delta)(1 + \delta \tilde{a}_t \Delta R_t)
\end{aligned} \tag{3.4}$$

which is a useful representation for the investigation of $\frac{1}{1+\gamma_{s-}^2\langle Z_{s-}, Z_{s-}\rangle}$. This, together with the assumptions $\int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s^d < \infty$ and $\int_0^\infty a_s \mathrm{d}R_s = \infty$, implies

$$\frac{\gamma_t(\delta)}{\gamma_{t-}(\delta)} = (1 + \delta a_t \Delta R_t) = (1 + o_{\mathrm{b}}(1)) = \mathcal{C}(\omega).$$

Let us now expand (3.3) with the definition of $A^1$. The first term in this expansion is handled by showing that

$$\underset{s_0(\omega)}{\exists} \ \underset{s \geqslant s_0(\omega)}{\forall} \Big( (\delta - \lambda_{\min})\|Z_{s-}\|^2 + \|B_s\|\|Z_{s-}\| \Big)^+ = 0.$$

By the assumption that $\lambda_{\min} > \delta$ and $\|B_s\| = o(\|Z_{s-}\|)$ for increasing $s$, there exists such an $s_0$ because the term in the brackets is negative for an $s_0$ large enough. On the other hand

$$\int_0^{s_0(\omega)} \frac{1}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\Big((\delta-\lambda_{\min})\|Z_{s-}\|^2 + \|B_s\|\|Z_{s-}\|\Big)^+ \gamma_s^2(\delta)a_s\mathrm{d}R_s \leqslant \mathcal{C}(\omega)\int_0^{s_0(\omega)} \gamma_s^2(\delta)a_s\mathrm{d}R_s < \infty.$$

From Theorem 2.1.1 we know that $(Z_t)_{t\geqslant 0}$ converges. Moreover there are no explosion times as $(Z_t)_{t\geqslant 0}$ is a strong solution of the stochastic integral equation (1.3). Consequently $\sup_t \|Z_t\| \leqslant \mathcal{C}(\omega) < \infty$. Therefore there exists a stopping time $\tau(\omega)$ such that the second term in the expansion of (3.3) can be handled as follows:

$$\begin{aligned}
2\lambda_{\min}\Big|\int_0^\infty & \frac{\gamma_s^2(\delta)a_s}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\big\langle Z_{s-}, \big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\big)Z_{s-}\big\rangle \mathrm{d}R_s\Big| \\
&\leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\Big|\int_{\tau(\omega)}^\infty a_s\Big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\Big)\mathrm{d}R_s\Big| \\
&\leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\Big|\int_0^\infty a_s\Big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\Big)\mathrm{d}R_s\Big| \\
&< \infty.
\end{aligned}$$

The fourth term in the expanded (3.3) is bounded by

$$\mathcal{C}\int_0^\infty \left(\frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)}\right)^2 \frac{\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} a_s^2\Delta R_s \mathrm{d}R_s^d \leqslant \mathcal{C}(\omega)\int_0^\infty a_s^2\Delta R_s\mathrm{d}R_s^d < \infty$$

and the seventh term by

$$\begin{aligned}
\mathcal{C}\int_0^\infty \left(\frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)}\right)^2 & \frac{\gamma_{s-}^2(\delta)}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\frac{a_s^2}{c_s^2}\sum_{i=1}^d \mathrm{d}\Big[\int_0^\cdot (M(\mathrm{d}t, D_{t-}, Z_{t-}))_i\Big]_s \\
&\leqslant \int_0^\infty \frac{\gamma_{s-}^2(\delta)h_s^{ii}(Z_{s-})}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\frac{a_s^2}{c_s^2}\mathrm{d}R_s < \infty.
\end{aligned}$$

The remaining terms dependent on $B_s$ or $C_s$, respectively. We make use of the fact, that they are bounded almost surely and we know their asymptotic properties.

Investigation of the third term leads to

$$\begin{aligned}
\int_0^\infty & \frac{\gamma_s^2(\delta)\|B_s\|^2}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}a_s^2\Delta R_s\mathrm{d}R_s^d \\
&\leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\int_{\tau(\omega)}^\infty \left(\frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)}\right)^2 \frac{\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}a_s^2\Delta R_s\mathrm{d}R_s^d \\
&\leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\int_0^\infty a_s^2\Delta R_s\mathrm{d}R_s^d < \infty.
\end{aligned}$$

For the fifth summand we distinguish two cases. If $F$ is $p$-smooth at $(0,0)$ with

$p \in [2,3)$, it holds

$$
\begin{aligned}
\Big| \int_0^\infty & \left( \frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)} \right)^2 \frac{\gamma_{s-}^2(\delta)}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} \langle C_s, Z_{s-} \rangle a_s \mathrm{d}R_s \Big| \\
&\leqslant \mathcal{C}(\omega) \int_0^\infty \gamma_{s-}(\delta) a_s c_s \mathrm{d}R_s \\
&\quad + \mathcal{C}(\omega) \Big| \int_0^\infty \gamma_{s-}^2(\delta) \frac{a_s}{c_s} \frac{\langle Z_{s-}, F(D_{s-},0,Z_{s-})\rangle}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} \mathrm{d}R_s \Big| < \infty,
\end{aligned}
$$

and, if $F$ is $p$-smooth at $(0,0)$ with $p \geqslant 3$, then

$$
\begin{aligned}
&\Big| \int_0^\infty \frac{\gamma_s^2(\delta)\langle C_s, Z_{s-}\rangle}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} a_s \mathrm{d}R_s \Big| \\
&\quad \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega) \Big| \int_{\tau(\omega)}^\infty \gamma_{s-}^2(\delta) \frac{a_s}{c_s} \frac{\langle Z_{s-}, F(D_{s-},0,Z_{s-})\rangle}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} \mathrm{d}R_s \Big| \\
&\qquad + \mathcal{C}(\omega) \int_{\tau(\omega)}^\infty \left( \frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)} \right)^2 \\
&\qquad \cdot \frac{\Big( c_s\|Z_{s-}\| + \sum_{m=0}^{\lfloor p \rfloor -2} g_{m+2}(D_{s-}) \tilde{f}_{m+2}(0) c_s^{m+1} + c_s^{p-1} \Big) \gamma_{s-}^2(\delta)\|Z_{s-}\|}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} a_s \mathrm{d}R_s \\
&\quad \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega) \Big| \int_{\tau(\omega)}^\infty \gamma_{s-}^2(\delta) \frac{a_s}{c_s} \frac{\langle Z_{s-}, F(D_{s-},0,Z_{s-})\rangle}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} \mathrm{d}R_s \Big| \\
&\qquad + \mathcal{C}(\omega) \int_0^\infty a_s c_s \mathrm{d}R_s + \mathcal{C}(\omega) \Big| \int_0^\infty \gamma_{s-}(\delta) a_s \sum_{m=0}^{\lfloor p \rfloor -2} g_{m+2}(D_{s-}) \tilde{f}_{m+2}(0) c_s^{m+1} \mathrm{d}R_s \Big| \\
&\qquad + \mathcal{C}(\omega) \int_0^\infty \gamma_{s-}(\delta) a_s c_s^{p-1} \mathrm{d}R_s \\
&\quad \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega) \Big| \int_{\tau(\omega)}^\infty \gamma_{s-}^2(\delta) \frac{a_s}{c_s} \frac{\langle Z_{s-}, F(D_{s-},0,Z_{s-})\rangle}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} \mathrm{d}R_s \Big| \\
&\qquad + \mathcal{C}(\omega) \int_0^\infty a_s c_s \mathrm{d}R_s + \mathcal{C}(\omega) \Big| \int_0^\infty \gamma_{s-}(\delta) a_s \sum_{m=0}^{\lfloor p \rfloor -2} \mathbb{E}\big(g_{m+2}(D_{s-})\big) \tilde{f}_{m+2}(0) c_s^{m+1} \mathrm{d}R_s \Big| \\
&\qquad + \mathcal{C}(\omega) \Big| \int_0^\infty \gamma_{s-}(\delta) a_s \sum_{m=0}^{\lfloor p \rfloor -2} \Big( g_{m+2}(D_{s-}) - \mathbb{E}\big(g_{m+2}(D_{s-})\big) \tilde{f}_{m+2}(0) \Big) c_s^{m+1} \mathrm{d}R_s \Big| \\
&\qquad + \mathcal{C}(\omega) \int_0^\infty \gamma_{s-}(\delta) a_s c_s^{p-1} \mathrm{d}R_s \\
&\quad = \mathcal{C}(\omega) + \mathcal{C}(\omega) \Big| \int_{\tau(\omega)}^\infty \gamma_{s-}^2(\delta) \frac{a_s}{c_s} \frac{\langle Z_{s-}, F(D_{s-},0,Z_{s-})\rangle}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} \mathrm{d}R_s \Big| + \mathcal{C}(\omega) \int_0^\infty a_s c_s \mathrm{d}R_s \\
&\qquad + \mathcal{C}(\omega) \Big| \int_0^\infty \gamma_{s-}(\delta) a_s \sum_{m=0}^{\lfloor p \rfloor -2} \Big( g_{m+2}(D_{s-}) - \mathbb{E}\big(g_{m+2}(D_{s-})\big) \tilde{f}_{m+2}(0) \Big) c_s^{m+1} \mathrm{d}R_s \Big| \\
&\qquad + \mathcal{C}(\omega) \int_0^\infty \gamma_{s-}(\delta) a_s c_s^{p-1} \mathrm{d}R_s < \infty.
\end{aligned}
$$

Finally we have two cases for the sixth term as well. If $F$ is $p$-smooth at $(0,0)$ with $p \in [2,3)$, we use that $\int_0^\infty \gamma_{s-}(\delta) a_s c_s \mathrm{d}R_s < \infty$ together with $\int_0^\infty a_s \mathrm{d}R_s = \infty$ implies $\gamma_{s-}(\delta)c_s \to 0$ and hence

$$\begin{aligned}
&\int_0^\infty \frac{\gamma_s^2(\delta)\|C_s\|^2 a_s^2}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\Delta R_s \mathrm{d}R_s \\
&\quad \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\int_{\tau(\omega)}^\infty \frac{(\gamma_{s-}(\delta)c_s)^2 a_s^2}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\Delta R_s \mathrm{d}R_s^d \\
&\quad\quad + \mathcal{C}(\omega)\int_{\tau(\omega)}^\infty \frac{\gamma_{s-}^2(\delta) a_s^2 L^2 \|Z_{s-}\|^2}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\Delta R_s \mathrm{d}R_s^d \\
&\quad\quad + \mathcal{C}(\omega)\int_0^\infty \frac{a_s^2}{c_s^2}\frac{\gamma_{s-}^2(\delta)}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\|F(D_{s-},0,Z_{s-})\|^2 \Delta R_s \mathrm{d}R_s^d \\
&\quad \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s^d \\
&\quad\quad + \mathcal{C}(\omega)\int_0^\infty \frac{a_s^2}{c_s^2}\frac{\gamma_{s-}^2(\delta)}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\|F(D_{s-},0,Z_{s-})\|^2 \Delta R_s \mathrm{d}R_s^d < \infty.
\end{aligned}$$

If $F$ is $p$-smooth at $(0,0)$ with $p \geqslant 3$, we make use of the facts that

$$\|C_s\|^2 \leqslant \mathcal{C}(\omega)\Big(c_s^2\|Z_{s-}\|^2 + c_s^2 + \frac{1}{c_s^2}\|F(D_{s-},0,Z_{s-})\|^2\Big)$$

holds for an $s$ larger than $\tau(\omega)$ and that $\gamma_s c_s \to 0$ implies $\gamma_s^2 c_s^2 \to 0$, to conclude

$$\begin{aligned}
&\int_0^\infty \left(\frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)}\right)^2 \frac{\gamma_{s-}^2(\delta)\|C_s\|^2}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2} a_s^2 \Delta R_s \mathrm{d}R_s^d \\
&\quad \leqslant \mathcal{C}(\omega) + \mathcal{C}(\omega)\int_{\tau(\omega)}^\infty \Big(c_s^2 + \gamma_{s-}^2(\delta)c_s^2\Big) a_s^2 \Delta R_s \mathrm{d}R_s^d \\
&\quad\quad + \mathcal{C}(\omega)\int_0^\infty \frac{a_s^2}{c_s^2}\frac{\gamma_{s-}^2(\delta)}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\|F(D_{s-},0,Z_{s-})\|^2 \Delta R_s \mathrm{d}R_s^d \\
&\quad < \infty.
\end{aligned}$$

Hence $\gamma_t^2(\delta)\|Z_t\|^2$ converges by Lemma A.1.1.

Following the arguments of [37] in the same way an investigation of $A_t^2$ yields that our algorithm converges to $z^* = 0$, the minimizing point of $\tilde{f}_0$ .

According to Lemma A.1.1 we know that

$$\Omega = \{\gamma_s^2(\delta)\|Z_s\|^2 \to\} \cap \{A_\infty^2 < \infty\}.$$

Assume now that there exists a set $N$ of non-zero probability on which $\gamma_t^2(\delta)\|Z_t\|^2$ does not convergence to 0. A contradiction to $\Omega \subset \{A_\infty^2 < \infty\}$ completes the proof. Note that

$$A_\infty^2 = \int_0^\infty \mathrm{d}A_s^2 + A_0^2 \geqslant 2\int_0^\infty \gamma_s^2(\delta) a_s \Big((\delta - \lambda_{\min})\|Z_{s-}\|^2 + \|B_s\|\|Z_{s-}\|\Big)^- \mathrm{d}R_s + A_0^2$$

and $\|B_s\| = o(\|Z_{s-}\|)$. Consequently there exists an $s_0^1(\omega) < \infty$ such that for all $s \geqslant s_0^1(\omega)$ the relation $\|B_s\| \leqslant \frac{1}{2}(\lambda_{\min} - \delta)\|Z_{s-}\|$ holds true. As $\lambda_{\min} > \delta$ we conclude for all $s \geqslant s_0^1(\omega)$ that

$$\Big((\delta - \lambda_{\min})\|Z_{s-}\|^2 + \|B_s\|\|Z_{s-}\|\Big)^- \geqslant \Big((\delta - \lambda_{\min})\|Z_{s-}\|^2 + \frac{1}{2}(\lambda_{\min} - \delta)\|Z_{s-}\|^2\Big)^- = \frac{1}{2}(\lambda_{\min} - \delta)\|Z_{s-}\|^2.$$

It is already known that the process $\gamma^2(\delta)\|Z\|^2$ converges for almost all $\omega \in \Omega$. But according to our assumption it does not converge to 0 for all $\omega \in N$. This implies that for almost all $\omega \in N$,

$$\underset{\epsilon^*>0}{\exists}\ \underset{s_0^2}{\exists}\ \underset{t\geqslant s_0^2}{\forall}\ \epsilon^* \leqslant \gamma_{t-}^2(\delta)\|Z_{t-}\|^2.$$

Bringing these ideas together with $s_0 := \max\{s_0^1, s_0^2\}$ yields

$$\begin{aligned} A_\infty^2 &\geqslant 2\int_0^\infty \Big((\delta - \lambda_{\min})\|Z_{s-}\|^2 + \|B_s\|\|Z_{s-}\|\Big)^- a_s\gamma_s^2(\delta)\mathrm{d}R_s + A_0^2 \\ &\geqslant (\lambda_{\min} - \delta)\int_{s_0}^\infty a_s\gamma_{s-}^2(\delta)\|Z_{s-}\|^2\mathrm{d}R_s \geqslant (\lambda_{\min} - \delta)\epsilon^*\int_{s_0}^\infty a_s\mathrm{d}R_s \\ &\geqslant \underbrace{(\lambda_{\min} - \delta)\epsilon^*}_{>0}\Bigg(\underbrace{\int_0^\infty a_s\mathrm{d}R_s}_{=\infty} - \underbrace{\int_0^{s_0} a_s\mathrm{d}R_s}_{\leqslant(R_{s_0}-R_0)<\infty}\Bigg) = \infty \end{aligned}$$

for almost all $\omega \in N$, which leads to the desired contradiction and completes the proof. □

## 3.2 Algorithms Using Kernel-Based Gradient Estimates

From Theorem 3.1.1 we can deduce many interesting special cases. We begin with the most general ones for semimartingales and proceed with time-continuous and time-discrete settings.

We have results for the one- and two-measurement algorithms (1.5) and (1.6). The concept of $p$-smoothness is simpler in these cases. The function $f$ fulfils the $p$-smooth condition at $z^*$ if

$$\Big\|\nabla f(z) - \sum_{m\leqslant\lfloor p\rfloor-1}\frac{1}{m!}\nabla^m\nabla f(z^*)(z - z^*)^m\Big\| = o\Big(\|z - z^*\|^{p-1}\Big).$$

We furthermore assume, that $D$ and $K(D)$ are bounded and therefore

$$\mathbb{E}(\|K(D_s)\|^2\|D_s\|^{2(p-1)}) < \infty$$

holds true.

The following assumptions are important.

**Assumption 3.2.1.** *Let conditions (kA)–(kC), (kF), (kG), (D) and (E) from Assumption 2.2.1 and condition $(\tilde{D})$ from Assumption 3.1.2 hold.*

$(\widetilde{kF})$ *If $p \geqslant 3$ assume*

$$\Big|\int_0^\infty \gamma_{s-}(\delta) a_s \sum_{m=0}^{\lfloor p\rfloor -2}\Big(\big(K(D_{s-})D_{s-}^{m+2}\big) - \mathbb{E}\big(\big(K(D_{s-})D_{s-}^{m+2}\big)\big)\Big) c_s^{m+1}\mathrm{d}R_s\Big| < \infty$$

*as well as $\mathbb{E}\left(K(D_s)D_s^{m-1}\right) = 0$ for all $m$ with $|m| \in \{3,\dots,\lfloor p\rfloor\}$ for algorithm* (1.5) *and*

$$\Big|\int_0^\infty \gamma_{s-}(\delta) a_s \sum_{k=1}^{\frac{\lfloor p\rfloor}{2}-1}\Big(\big(K(D_{s-})D_{s-}^{2k+1}\big) - \mathbb{E}\big(\big(K(D_{s-})D_{s-}^{2k+1}\big)\big)\Big) c_s^{2k}\mathrm{d}R_s\Big| < \infty$$

*as well as $\mathbb{E}\left(K(D_s)D_s^{m-1}\right) = 0$ for all $m$ with odd length $|m| \in \{3,\dots,\lfloor p\rfloor\}$ for algorithm* (1.6).

$(\widetilde{kG})$ *If $p \geqslant 3$ we furthermore assume $\mathbb{E}\left(K(D_s)D_s^{m-1}\right) = 0$ for all $m$ with $|m| \in \{3,\dots,\lfloor p\rfloor\}$ for algorithm* (1.5).

$(\widetilde{kH})$ *Let for every $i \in \{1,\dots,d\}$ and all $z \in \mathbb{R}^d$*

$$\int_0^\infty \frac{a_s^2}{c_s^2}\frac{\gamma_{s-}^2(\delta) h_s^{ii}(Z_{s-})}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\mathrm{d}R_s < \infty \quad \text{where} \quad h_s^{ii}(z) := \frac{\mathrm{d}[\int_0^\cdot (K(D_s)\breve{M}(\mathrm{d}t,z))_i]_s}{\mathrm{d}R_s},$$

*where $\breve{M}$ was introduced on page 8.*

*Remark* 3.2.1. Random variables fulfilling the moment conditions of $(\widetilde{kG})$ can be generated with the help of a Vandermonde matrix. Details can be found in the paper of Dippon [10, section 5.2].

**Corollary 3.2.1.** *Let Assumption 3.2.1 hold, and assume a continuous Hessian $H$ around $z^*$. Assume that $f$ is $p$-smooth at $z^*$ with $p \geqslant 2$. Then for both, the one-measurement* (1.5) *and the two-measurement* (1.6) *algorithm, for all $0 \leqslant \delta < \lambda_{min}$*

$$\gamma_t(\delta)\|Z_t - z^*\| \xrightarrow{t\to\infty} 0 \text{ a.s.}$$

*holds.*

*Proof.* We trace the result back to Theorem 3.1.1. Conditions (D), (E) and $(\widetilde{D})$ are also assumed there. In the proof of Corollary 2.2.1 we already showed that (A), (B), (C), (F) and (G) follow from (kA), (kB), (kC), (kF) and (kG). It remains to show that $(\widetilde{kF})$, $(\widetilde{kG})$ and $(\widetilde{kH})$ yield $(\widetilde{F})$, $(\widetilde{G})$ and $(\widetilde{H})$, respectively. It holds true that

$$\tilde{f}_k(z) = \begin{cases} \nabla^k f(z) & \text{if } k \text{ is odd} \\ 0 & \text{if } k \text{ is even.} \end{cases}$$

Condition $(\widetilde{F})$ follows from $(\widetilde{kF})$ by $g_k(D_{s-}) = K(D_{s-})D_{s-}^k$ and $\mathbb{E}(g_1(D_{s-}) = \mathbb{E}(K(D_{s-}) \otimes D_{s-}) = \mathbb{1}_d$. Hence

$$\begin{aligned}&\left|\int_0^\infty \gamma_{s-}(\delta)a_s\Big(g_1(D_{s-}) - \mathbb{E}(g_1(D_{s-}))\Big)\mathrm{d}R_s\right|\\&\qquad = \left|\int_0^\infty \gamma_{s-}(\delta)a_s\Big((K(D_{s-})\otimes D_{s-}) - \mathbb{1}_d\Big)\mathrm{d}R_s\right| < \infty.\end{aligned}$$

Moreover for (1.5)

$$\begin{aligned}&\Big|\int_0^\infty \gamma_{s-}(\delta)a_s\sum_{m=0}^{\lfloor p\rfloor-2}\Big(g_{m+2}(D_{s-}) - \mathbb{E}\big(g_{m+2}(D_{s-})\big)\Big)\tilde{f}_{m+2}(z^*)c_s^{m+1}\mathrm{d}R_s\Big|\\&= \Big|\int_0^\infty \gamma_{s-}(\delta)a_s\sum_{m=0}^{\lfloor p\rfloor-2}\Big((K(D_{s-})D_{s-}^{m+2}) - \mathbb{E}\big((K(D_{s-})D_{s-}^{m+2})\big)\Big)\nabla^{m+2}f(Z_{s-})c_s^{m+1}\mathrm{d}R_s\Big|\\&< \infty\end{aligned}$$

and for (1.6), as $\tilde{f}_k(z) = 0$ if $k$ is odd,

$$\begin{aligned}&\Big|\int_0^\infty \gamma_{s-}(\delta)a_s\sum_{m=1}^{\lfloor p\rfloor-2}\Big(g_{m+2}(D_{s-}) - \mathbb{E}\big(g_{m+2}(D_{s-})\big)\Big)\tilde{f}_{m+2}(z^*)c_s^{m+1}\mathrm{d}R_s\Big|\\&= \Big|\int_0^\infty \gamma_{s-}(\delta)a_s\sum_{l=1}^{\frac{\lfloor p\rfloor}{2}-1}\Big((K(D_{s-})D_{s-}^{2l+1}) - \mathbb{E}\big((K(D_{s-})D_{s-}^{2l+1})\big)\Big)\nabla^{2l+1}f(Z_{s-})c_s^{2l}\mathrm{d}R_s\Big|\\&< \infty.\end{aligned}$$

Now we turn to $(\widetilde{G})$. For the one-measurement algorithm (1.5), the first condition of $(\widetilde{G})$ holds by assumption and $\mathbb{E}(g_0(D_{s-})) = \mathbb{E}(K(D_{s-})) = 0$ by (kG). The two-measurement (1.6) case is verified as $\tilde{f}_0(z) = 0$. Turning to the last part of $(\widetilde{G})$ yields

$$\begin{aligned}&\int_0^\infty \frac{a_s^2}{c_s^2}\frac{\gamma_s^2(\delta)}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\|F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d\\&\quad= \int_0^\infty\left(\frac{\gamma_s^2(\delta)}{\gamma_{s-}^2(\delta)}\right)\frac{a_s^2}{c_s^2}\frac{\gamma_{s-}^2(\delta)}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\|F(D_{s-},0,Z_{s-})\|^2\Delta R_s\mathrm{d}R_s^d\\&\quad= \begin{cases}0 & \text{for algorithm (1.5)}\\ \mathcal{C}(\omega)\displaystyle\int_0^\infty\frac{a_s^2}{c_s^2}\frac{\gamma_{s-}^2(\delta)\|f(Z_{s-})\|^2}{1+\gamma_{s-}^2(\delta)\|Z_{s-}\|^2}\Delta R_s\mathrm{d}R_s^d & \text{for algorithm (1.6)}\end{cases}\\&\quad< \infty.\end{aligned}$$

Choosing $M(\mathrm{d}s, D_{s-}, x) := K(D_{s-})\widetilde{M}(\mathrm{d}s, x)$ yields $(\widetilde{H})$ from $(\widetilde{kH})$. □

### 3.2.1 Continuous-Time Algorithms

Likewise consistency, the almost-sure convergence rate includes an interesting special case for the Itô setting.

**Assumption 3.2.2.** *Let $f$ be $p$-smooth at $z^*$ with $p \geqslant 2$. Assumption 2.2.2 holds with (kD') replaced by $a_t := a(1+t)^{-1}$, $a > 0$, as well as $c_t := c(1+t)^{-\frac{1}{2p}}$, $c > 0$, and (kH') replaced by $\sum_{j=1}^{d} \sigma_s^{ij}(x) \leqslant C(1 + \|x\|)$ for all $i \in \{1, \ldots, d\}$.*

$(\widetilde{kF'})$ *If $p \geqslant 3$ and algorithm (2.11) is used, assume*

$$\left|\int_0^\infty s^\delta a_s \sum_{m=0}^{\lfloor p\rfloor-2} \Big((K(D_s)D_s^{m+2}) - \mathbb{E}\big((K(D_s)D_s^{m+2})\big)\Big) c_s^{m+1} \mathrm{d}s\right| < \infty$$

*as well as $\mathbb{E}\left(K(D_s)D_s^{m-1}\right) = 0$ for all $m$ with $|m| \in \{3, \ldots, \lfloor p\rfloor\}$.*
*If $p \geqslant 3$ and algorithm (2.12) is used, assume*

$$\left|\int_0^\infty s^\delta a_s \sum_{m=1}^{\frac{\lfloor p\rfloor}{2}-1} \Big((K(D_s)D_s^{2m+1}) - \mathbb{E}\big((K(D_s)D_s^{2m+1})\big)\Big) c_s^{2m} \mathrm{d}s\right| < \infty$$

*as well as $\mathbb{E}\left(K(D_s)D_s^{m-1}\right) = 0$ for all $m$ with odd length $|m| \in \{3, \ldots, \lfloor p\rfloor\}$.*

$(\widetilde{kG'})$ *For the one-measurement algorithm (2.11) let*

$$\left|\int_0^\infty \frac{a_s}{c_s} s^{2\delta}(\delta) \frac{\langle Z_s, K(D_s) f(Z_s)\rangle}{1 + s^{2\delta}\|Z_s\|^2} \mathrm{d}s\right| < \infty.$$

*If $p \geqslant 3$ we furthermore assume $\mathbb{E}\left(K(D_s)D_s^{m-1}\right) = 0$ for all $m$ with $|m| \in \{3, \ldots, \lfloor p\rfloor\}$ for algorithm (2.11).*

**Corollary 3.2.2.** *For the Itô type stochastic integral equations (2.11) and (2.12) let Assumption 3.2.2 hold. For $p$-smooth $f$ at $z^*$ with $p \geqslant 3$, the Hessian $H$ of $f$ shall exist and be continuous around $z^*$, and $\lambda_{\min} > \frac{p-1}{2pa}$.*
*Then*

$$(1+t)^\delta \|Z_t - z^*\| \to 0 \ a.s.$$

*for all $\delta \in (0, \frac{p-1}{2p})$.*

*Proof.* It is sufficient to show that the conditions of Corollary 3.2.1 are fulfilled. In both corollaries, (kA), (kB) and (kC) are assumed. Due to the continuity of the Itô type stochastic integral equation, $(\widetilde{kG})$ and (E) are trivially fulfilled. Hence it remains to verify that conditions (kG), (D)–(F), $(\tilde{D})$, $(\widetilde{kF})$ and $(\widetilde{kH})$ hold. As well as in the proof of consistency we choose $R_s := s$ and $\widetilde{M}(\mathrm{d}s, x) := \sum_{j=1}^{d} \sigma_s^j(x) \mathrm{d}W^j(s)$. The path-continuity of $(R_t)_{t \geqslant 0}$ transfers to $(\gamma_t(\delta))_{t \geqslant 0}$.

$$\gamma_t(\delta) = \mathcal{E}_t\left(\delta \int_0^\cdot a_s \mathrm{d}R_s\right) = \mathcal{E}_t\left(\delta a \int_0^\cdot \frac{1}{1+s} \mathrm{d}s\right) = \exp(\delta a \ln(1+t)) = (1+t)^{a\delta}$$

Condition $(\widetilde{kF})$ follows due to $R_s = s$. Moreover assume a sufficiently small $\epsilon > 0$. According to our conditions, $0 \leqslant \delta < \lambda_{\min}$ holds. We can show $(\widetilde{D})$ by choosing $\delta \in \left(0, \frac{p-1}{2pa}\right)$, which yields

$$\int_0^\infty \gamma_{s-}(\delta) a_s c_s \mathrm{d}R_s = ac\int_0^\infty (1+s)^{a\delta-1-\frac{p-1}{2p}}\mathrm{d}s \leqslant ac\int_0^\infty (1+s)^{\frac{p-1}{2p}-\epsilon-1-\frac{p-1}{2p}}\mathrm{d}s$$
$$= ac\int_0^\infty (1+s)^{-1-\epsilon}\mathrm{d}s < \infty.$$

Assumption $(\widetilde{kH})$ follows by

$$\int_0^\infty \frac{\gamma_{s-}^2(\delta) h_s^{ii}(Z_{s-})}{1+\gamma_{s-}(\delta)\|Z_{s-}\|^2}\frac{a_s^2}{c_s^2}\mathrm{d}R_s$$
$$= \sum_{j=1}^d \int_0^\infty \frac{\gamma_s^2(\delta)\sigma_s^j(Z_s)^2}{1+\gamma_s(\delta)\|Z_s\|^2}\frac{a_s^2}{c_s^2}\mathrm{d}s \leqslant \mathcal{C}\int_0^\infty \gamma_s^2(\delta)\underbrace{\frac{(1+\|Z_s\|)^2}{1+\gamma_s(\delta)\|Z_s\|^2}}_{\leqslant 1}\frac{a_s^2}{c_s^2}\mathrm{d}s$$
$$\leqslant \mathcal{C}\int_0^\infty \gamma_s^2(\delta)\frac{a_s^2}{c_s^2}\mathrm{d}s = \mathcal{C}\frac{a^2}{c^2}\int_0^\infty (1+s)^{2a\delta-2+\frac{p-1}{p}}\mathrm{d}s$$
$$= \mathcal{C}\int_0^\infty (1+s)^{\frac{p-1}{p}-\epsilon-2-\frac{p-1}{p}}\mathrm{d}s \leqslant \mathcal{C}\int_0^\infty (1+s)^{-1-\epsilon}\mathrm{d}s < \infty.$$

Condition (F) is a direct consequence of (kF'). The choice of $(a_t)_{t\geqslant 0}$ and $(c_t)_{t\geqslant 0}$ yields (D) form (kD'). In an analogous way, (kG') implies (kG).

□

### 3.2.2 Discrete-Time Algorithms

We proceed with a time-discrete special case. In the following results we obtain the same rates of convergence as in the previous subsection.

**Assumption 3.2.3.** *Let $f$ be $p$-smooth at $z^*$ with $p \geqslant 2$. Assumption 2.2.3 holds for algorithms* (2.13) *and* (2.14), *with (kD") and (kE") replaced by $a_n := an^{-1}$, $a > 0$, and $c_n := cn^{-\frac{1}{2p}}$, $c > 0$.*

$(\widetilde{kF''})$ *If $p \geqslant 3$ and algorithm* (2.13) *is used, assume*

$$\left|\sum_{k=1}^\infty k^\delta a_k \sum_{m=0}^{\lfloor p\rfloor-2}\Big((K(D_{k-1})D_{k-1}^{m+2}) - \mathbb{E}\big((K(D_{k-1})D_{k-1}^{m+2})\big)\Big)c_k^{m+1}\right| < \infty$$

*as well as $\mathbb{E}\left(K(D_k)D_k^{m-1}\right) = 0$ for all $m$ with $|m| \in \{3,\dots,\lfloor p\rfloor\}$.*

*If $p \geqslant 3$ and algorithm* (2.14) *is used, assume*

$$\left|\sum_{k=1}^\infty k^\delta a_k \sum_{m=1}^{\frac{\lfloor p\rfloor}{2}-1}\Big((K(D_{k-1})D_{k-1}^{2m+1}) - \mathbb{E}\big((K(D_{k-1})D_{k-1}^{2m+1})\big)\Big)c_k^{2m}\right| < \infty$$

*as well as* $\mathbb{E}\left(K(D_k)D_k^{m-1}\right) = 0$ *for all* $m$ *with odd length* $|m| \in \{3, \dots, \lfloor p \rfloor\}$*.*

$(\widetilde{kG''})$ *For the one-measurement algorithm* (2.13) *let*

$$\left|\sum_{k=1}^{\infty} \frac{a_k}{c_k} k^{2\delta} \frac{\langle Z_{k-1}, K(D_{k-1}) f(Z_{k-1})\rangle}{1 + k^{2\delta}\|Z_{k-1}\|^2}\right| < \infty$$

*and*

$$\sum_{k=1}^{\infty} \frac{a_k^2}{c_k^2} k^{2\delta} \frac{\|f(Z_{k-1})\|^2}{1 + k^{2\delta}\|Z_{k-1}\|^2} < \infty.$$

*If* $p \geqslant 3$ *we furthermore assume* $\mathbb{E}\left(K(D_k)D_k^{m-1}\right) = 0$ *for all* $m$ *with* $|m| \in \{3, \dots, \lfloor p \rfloor\}$ *for algorithm* (2.13)*.*

**Corollary 3.2.3.** *Let Assumption 3.2.3 hold. If* $f$ *is* $p$*-smooth at* $z^*$*, the Hessian* $H$ *of* $f$ *exists and is continuous around* $z^*$*, and* $\lambda_{\min} > \frac{p-1}{2pa}$*, then*

$$n^{\delta}\|Z_n - z^*\| \to 0 \ a.s.$$

*for all* $\delta \in (0, \frac{p-1}{2p})$*.*

*Proof.* Likewise the previous proof, it is sufficient to verify the conditions of Corollary 3.2.1. Hence we show that conditions (kG), (D)–(F), $(\tilde{D})$, $(\widetilde{kF})$, $(\widetilde{kG})$ and $(\widetilde{kH})$ hold. Conditions (kA), (kB) and (kC) are assumed in both corollaries. We choose the same notations as in the proof of consistency of the recursion. Together with $R_s := \lfloor s \rfloor$ and $a_s := as^{-1}$ a Taylor expansion, and using that $\ln(1+x) \approx x$ for $x \ll 1$ holds, yield

$$\begin{aligned}
\gamma_t(\delta) &= \mathcal{E}_t\left(a\delta \int_0^{\cdot} \frac{1}{s} \mathrm{d}R_s\right) = \prod_{i=1}^{\lfloor t \rfloor}\left(1 + \frac{a\delta}{i}\right) = \exp\left(\sum_{i=1}^{\lfloor t \rfloor} \ln(1 + \frac{a\delta}{i})\right) \\
&=: C_{\lfloor t \rfloor} \exp\left(a\delta \sum_{i=1}^{\lfloor t \rfloor} \frac{1}{i}\right)
\end{aligned}$$

where $C_{\lfloor t \rfloor} \to C_\infty \in (0, \infty)$ as $t \to \infty$. Investigating this exponential term yields

$$\exp\left(a\delta \sum_{i=1}^{\lfloor t \rfloor} \frac{1}{i}\right) \geqslant \exp\left(a\delta \int_1^{\lfloor t \rfloor} \frac{1}{x} \mathrm{d}x\right) = \lfloor t \rfloor^{a\delta}$$

as well as

$$\exp\left(a\delta \sum_{i=1}^{\lfloor t \rfloor} \frac{1}{i}\right) = \exp\left(a\delta + a\delta \sum_{i=2}^{\lfloor t \rfloor} \frac{1}{i}\right) \leqslant \exp\left(a\delta + a\delta \int_1^{\lfloor t \rfloor} \frac{1}{x} \mathrm{d}x\right) = \exp(a\delta)\lfloor t \rfloor^{a\delta}$$

where the inequalities in the previous term follow by the integral test for convergence. These results enable us to verify the assumptions of the semimartingale case with

$\gamma_{t-}(\delta)$ replaced by $\lfloor t \rfloor^{a\delta}$. Assume a sufficiently small $\epsilon > 0$. With the conditions on $\delta$, the bound

$$\int_0^\infty \lfloor s \rfloor^{a\delta} a_s c_s \mathrm{d}R_s \leqslant ac \sum_{n\in\mathbb{N}} n^{a\delta - 1 - \frac{p-1}{2p}} = \mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{p-1}{2p} - \epsilon - 1 - \frac{p-1}{2p}} = \mathcal{C} \sum_{n\in\mathbb{N}} n^{-1-\epsilon} < \infty$$

yields $(\widetilde{D})$. Condition $(\widetilde{kF})$ follows due to $R_s = s$. In order to show $(\widetilde{kH})$, we use the positivity of the processes $\lfloor s \rfloor^{2a\delta} \frac{a_s^2}{c_s^2} h_s^{ii}(Z_{s-})$ and $\Delta R_s$ as well as the monotone convergence theorem. Positivity enables us to investigate the expectation only. The latter, together with Hölder's inequality, yields

$$\begin{aligned} \mathbb{E} \int_0^\infty \lfloor s \rfloor^{2a\delta} \frac{a_s^2}{c_s^2} h_s^{ii}(Z_{s-}) \mathrm{d}R_s &\leqslant \mathcal{C} \mathbb{E} \sum_{n\in\mathbb{N}} \mathbb{E}(K(D_n)^2 V_n^2 \mid \mathcal{F}_{n-1}) n^{2a\delta - 2 - \frac{p-1}{2p}} \\ &\leqslant \mathcal{C} \left( \sup_{n\in\mathbb{N}} \mathbb{E} \|K(D_s)^2\| V_n^2 \right) \sum_{n\in\mathbb{N}} n^{\frac{p-1}{p} - \epsilon - 2 - \frac{p-1}{p}} \\ &\leqslant \mathcal{C} \sum_{n\in\mathbb{N}} n^{-1-\epsilon} < \infty. \end{aligned}$$

Validity of condition $(\widetilde{kG})$ follows analogously. The choice of $(a_n)$ and $(c_n)$ yields (D) and (E) via (kD") and (kE"), respectively. Assumptions (F) and (kG) follow directly form (kF") and (kG"), respectively. □

## 3.3 Simulations

After the almost sure convergence rates were derived, we deal with simulations to find out how processes behave when they are maximally weighted such that convergence is still achieved. Moreover it is an interesting question, whether they have the same empirical $L^2$-error and which have a better pre-asymptotic behaviour. It should also be noted that all processes with a perturbated direction can be modified such that it reaches a rate close to $n^{-1/2}$. Details can be found in Section 5.1 in Dippon's paper [10]. Therefore these procedures are superior to the Kiefer-Wolfowitz algorithm which reaches a maximum almost sure convergence rate close to $n^{-1/3}$. These special settings of $D$ are not simulated in this thesis.

### 3.3.1 $\mathbf{L^2}$-Convergence Rate of One-measurement Algorithms Starting at the Extremum of $f$

As in Subsection 2.3.3, we choose $a_n = 2/(70+n)$, $c_n = 1/n^{1/6}$, and starting value $Z_0 = (0,0,0)^T$. For each algorithm $N = 1000$ paths with $n = 10000$ single observations were simulated. The estimated $L^2$-errors of $Z_n$ are given in the following plots. In Corollary 3.2.3 we found out that $n^\delta Z_n$ converges with $\delta \in [0, 1/3)$. In Figure 3.1 $Z_n$ is multiplied by $n^{1/3}$. Convergence of $n^{1/3} Z_n$ in any sense is not proven in this thesis, under the same assumptions as above. However it is well-known that the classical Kiefer-Wolfowitz procedure converges with $L^2$-rate $n^{1/3}$ under similar conditions to the assumptions of Corollary 3.2.3.

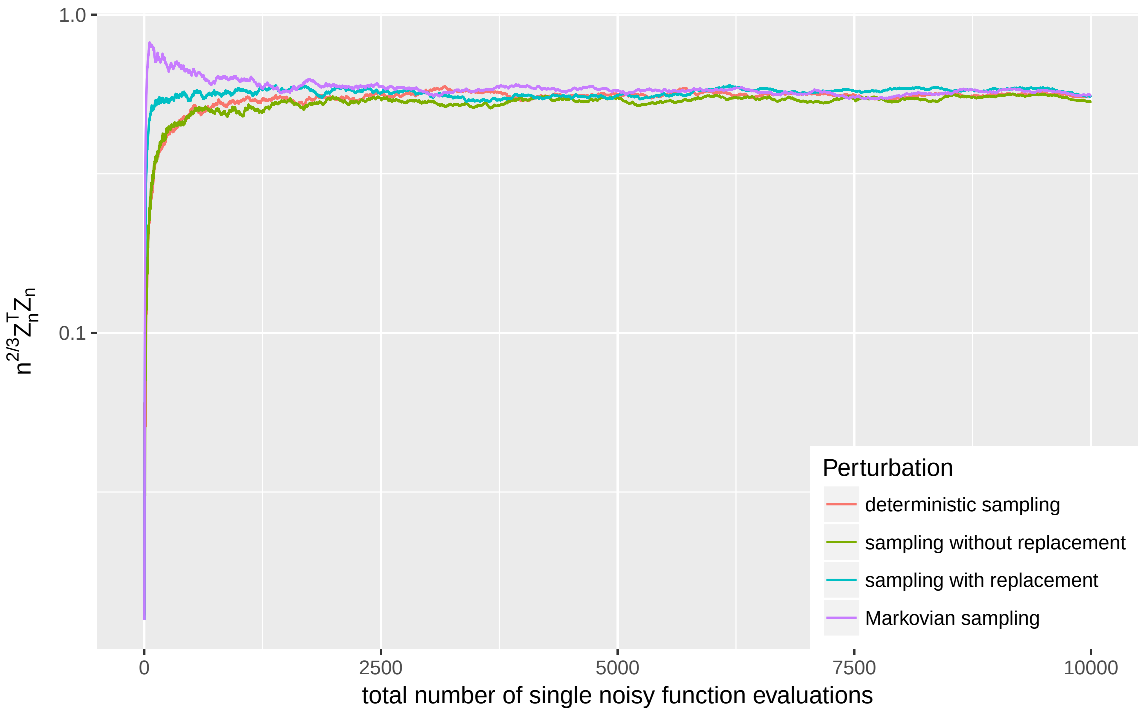


**Figure 3.1.** Empirical $L^2$-error of $n^{1/3}Z_n$ generated by one-measurement procedures

Next, in Figure 3.2, we investigate $n^{1/4}Z_n$ with all other settings as before. In addition to the almost sure convergence rate, this plot also raises the conjecture that the $L^2$-convergence rate is near $n^{1/3}$. Note that all procedures seem to have the same limit. A reason for that could be that the distributions of all chosen perturbations $D$ are too similar in our simulations. Different constructions of $D$ might yield other results.

### 3.3.2 $\mathbf{L^2}$-Convergence Rate of Two-measurement Algorithms Starting at the Extremum of $f$

After the investigation of one-measurement algorithms we devote ourselves to two-measurement procedures. In order to make the comparison more complete we also include the classical Kiefer-Wolfowitz algorithms which even takes $2d$ evaluations per iteration step. These simulations can be found in Figure 3.3. Note that the scale of the vertical axis depends on the number $n$ of iterations, where the horizontal is scaled with the number $v(n)$ of evaluations. In order to make this plot, $\|Z_n\|^2$ is multiplied by $(v(n)/6)^{2/3}$ for the Kiefer-Wolfowitz procedure, as in this case we have 6 evaluations per iteration step, hence $v(n) = 6n$ holds true. Analogously multiply $\|Z_n\|^2$ by $(v(n)/2)^{2/3}$ in the two-measurement algorithms.

The Kiefer-Wolfowitz algorithm has the lowest $L^2$-error for large $n$. All two-measurement algorithms show a very similar pre-asymptotic behaviour. The asymptotic $L^2$-error seem to coincide for Markovian perturbation and sampling with replacement. These are the procedures for which $\|\sum_{i=1}^n D_i - \mathbb{E}(D_i)\|^2 = \mathcal{O}(n)$ holds. For the procedures with $\|\sum_{i=1}^n D_i - \mathbb{E}(D_i)\|^2 = \mathcal{O}(1)$, namely deterministic and sampling

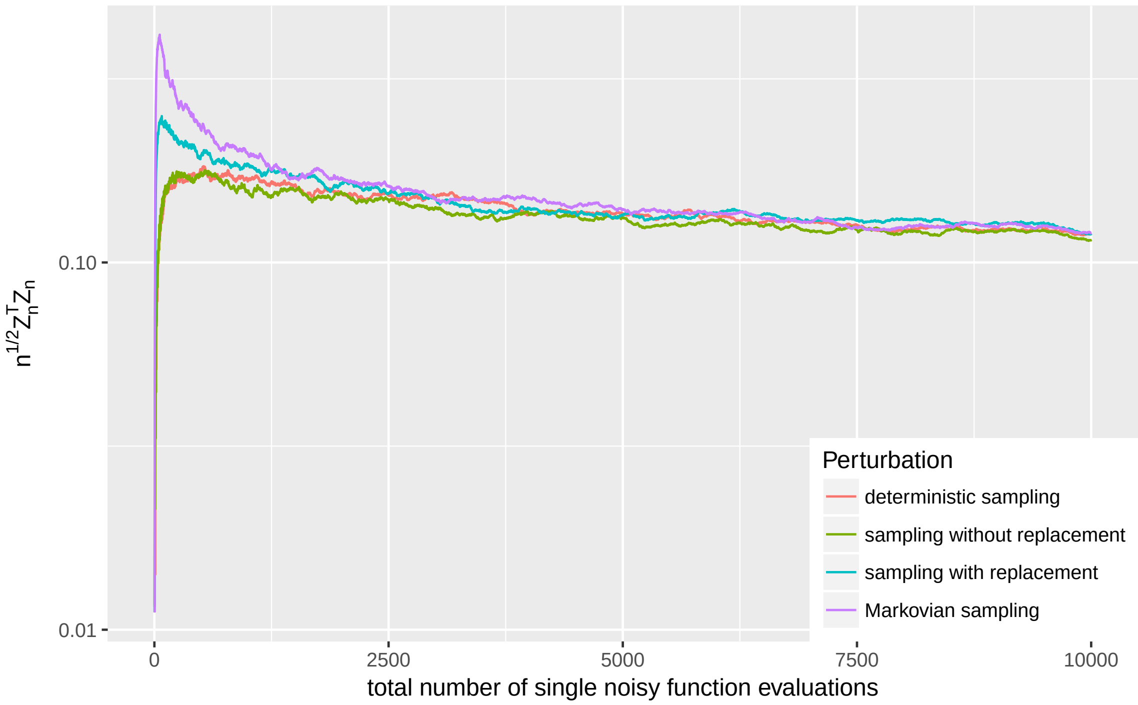


**Figure 3.2.** Empirical $L^2$-error of $n^{1/4}Z_n$ generated by one-measurement procedures

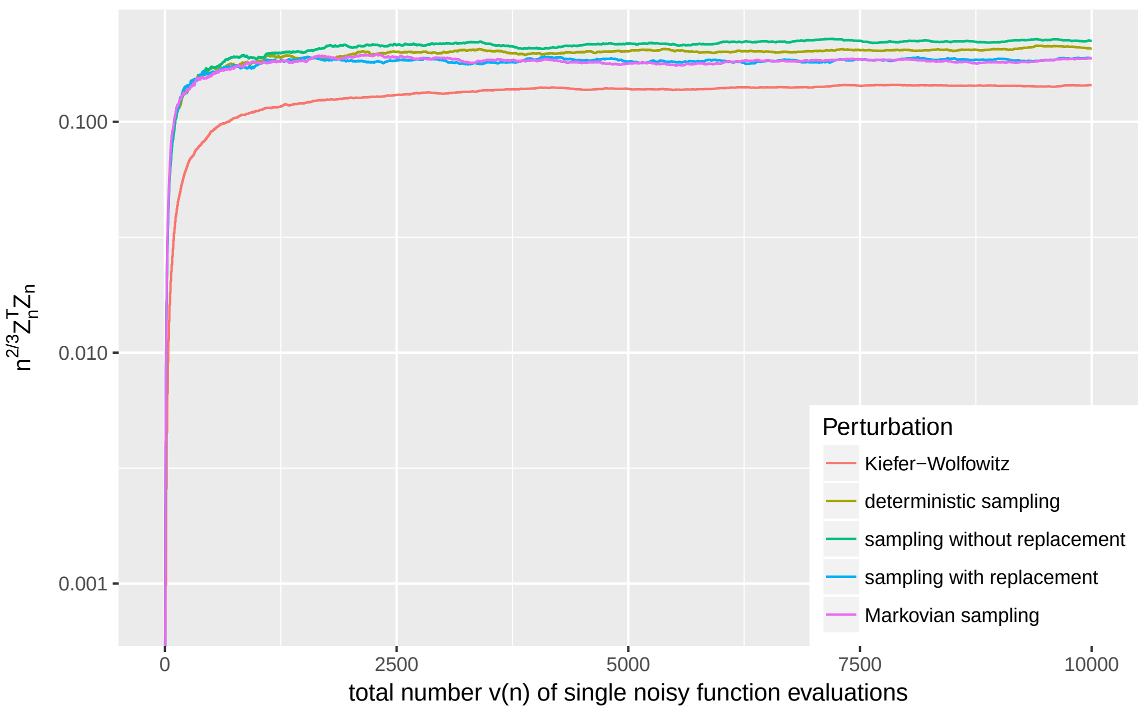


**Figure 3.3.** Empirical $L^2$-error of $n^{1/3}Z_n$ generated by Kiefer-Wolfowitz and two-measurement procedures

without replacement, the asymptotic $L^2$-error seems to be a little bit larger. In one-measurement settings the results are different. We observed a different pre-asymptotic behaviour but a similar asymptotic $L^2$-error.

The reader might have noticed, that in Figure 3.3 the error of Kiefer-Wolfowitz is lower although in Figure 2.18 it was higher. This is due to the fact that the process is weighted by the number of iterations $n$ and not by the number of single function observations. If instead we weighted with the latter, this would result in Figure 3.4. Hence in this respect Kiefer-Wolfowitz seem to behave weaker. However, this

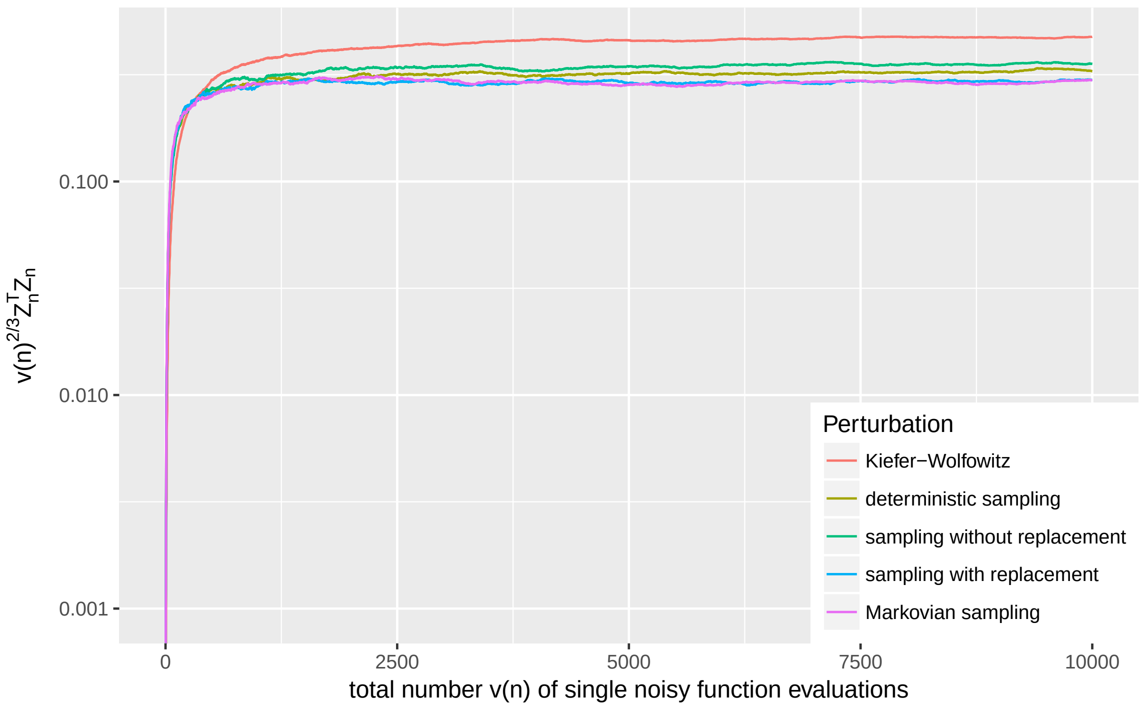


**Figure 3.4.** Empirical $L^2$-error of $v(n)^{1/3}Z_n$ generated by Kiefer-Wolfowitz and two-measurement procedures

effect is more pronounced in high dimensional problems. This would be extreme if we simulated not only a 3-dimensional optimization problem but had hundreds of parameters to be adjusted.

In all simulations one-measurement algorithms perform worse than two-measurement algorithms or the Kiefer-Wolfowitz procedure. Moreover they have a similar behaviour. However concerning the pre-asymptotic behaviour, we distinguished two subclasses. The methods with $\|\sum_{i=1}^n D_i - \mathbb{E}(D_i)\|^2 = \mathcal{O}(1)$ behave better than the methods with only $\|\sum_{i=1}^n D_i - \mathbb{E}(D_i)\|^2 = \mathcal{O}(n)$. Such a difference does not occur in two-measurement settings. This behaviour seems to be independent of the function $f$, however we have no rigorous proof for this claim. It is worth mentioning that methods with $\|\sum_{i=1}^n D_i - \mathbb{E}(D_i)\|^2 = \mathcal{O}(n)$ could even fulfil $\|\sum_{i=1}^n D_i - \mathbb{E}(D_i)\|^2 = \mathcal{O}(n^\alpha)$ with $\alpha \in (0, 1)$. A more detailed investigation on this asymptotic behaviour as well as the study of the $L^2$-error of $Z_n$ with respect to $\alpha$ could be interesting.

# 4 Concluding Remarks

In this part of the thesis, many known as well as new results were derived from a generic, semimartingale-type, randomization algorithm. Consistency and almost sure convergence rate results transfer directly to these special cases. This general formulation gives the opportunity to build new algorithms, just by employing this framework. One key message is that multiple-measurement procedures are more robust than one-measurement procedures, especially at the first iteration steps. If one can afford a second observation per step, one should prefer to do so. But that's not expensive compared by the savings in high-dimensional problems where we need $2d$ observations in the regular Kiefer-Wolfowitz case. If randomization is to be avoided, then the presented deterministic perturbations have the same dimension-reducing effect. Moreover a variety of randomization as well as deterministic perturbation designs were introduced. Due to the general assumptions, a whole bunch of new methods can easily be constructed.

An open issue is the derivation of asymptotic normality results. However, with the almost sure results at hand together with the methods presented in the second part of this thesis, the main difficulties appear to be tractable.

# Part II

# COMPANION ALGORITHMS

# 5 Introduction

This part of the thesis is on companion algorithms. These always refer to a leading algorithm as for instance the Robbins-Monro or Kiefer-Wolfowitz procedure. An introduction to these and stochastic approximation in general can be found at the beginning of the previous part.

## 5.1 Previous Work

In 2006 Mokkadem and Pelletier [28] suggested an algorithm to estimate the minimizer of $f\colon \mathbb{R}^d \to \mathbb{R}$ and its minimum simultaneously. As a companion to the Kiefer-Wolfowitz algorithm (1.2),

$$Z_{n+1} - Z_n = -a_n Y_n(Z_n)$$

with

$$Y_n(z) = \frac{1}{2c_n} \left\{ f(z + c_n e_i) - f(z - c_n e_i) + M_{n,i} \right\}_{i \in \{1,\dots,d\}},$$

which estimates the minimizing value $z^*$ of $f$, they suggested the recursion

$$\Upsilon_{n+1} = (1 - \tilde{a}_n)\Upsilon_n + \tilde{a}_n \widetilde{Y}_n(Z_n) \tag{5.1}$$

to estimate the minimum $f(z^*)$, where $\widetilde{Y}_n(z)$ is a noisy estimator of the function value $f(z)$ and the sequence $(\tilde{a}_n)$ tends towards zero. The basic idea of procedure (5.1) is to calculate a weighted mean of $\widetilde{Y}_n(Z_n)$. In the following the algorithm generating $Z_n$ will be denoted as leading algorithm. Here the term $\widetilde{Y}_n(Z_n)$ in (5.1) should not be confused with $Y_n(Z_n)$ in the leading Kiefer-Wolfowitz algorithm. It is worth to mention that a related learning rule is given in the book of Ljung et al. [24, Ch. 4].

Now we show that the explicit representation for (5.1) is a weighted mean. For that purpose choose the sequence $(b_n)$ with $b_i > 0$ for all $i \in \mathbb{N}$ such that $\tilde{a}_n = b_n / \sum_{i=1}^n b_i$ holds. Consider the weighted mean

$$\overline{\Upsilon}_{n+1} := \frac{\sum_{i=1}^n b_i \widetilde{Y}_i(Z_i)}{\sum_{i=1}^n b_i}$$

then

$$\begin{aligned}
&\overline{\Upsilon}_{n+1} - \overline{\Upsilon}_n \\
&= \frac{\sum_{i=1}^{n} b_i \tilde{Y}_i(Z_i)}{\sum_{i=1}^{n} b_i} - \frac{\sum_{i=1}^{n-1} b_i \tilde{Y}_i(Z_i)}{\sum_{i=1}^{n-1} b_i} \\
&= \frac{\Big(\sum_{i=1}^{n-1} b_i\Big)\Big(\sum_{i=1}^{n} b_i \tilde{Y}_i(Z_i)\Big) - \Big(\sum_{i=1}^{n} b_i\Big)\Big(\sum_{i=1}^{n-1} b_i \tilde{Y}_i(Z_i)\Big)}{(\sum_{i=1}^{n} b_i)(\sum_{i=1}^{n-1} b_i)} \\
&= \frac{\Big(\sum_{i=1}^{n-1} b_i\Big)\Big(\sum_{i=1}^{n-1} b_i \tilde{Y}_i(Z_i) + b_n \tilde{Y}_n(Z_n)\Big) - \Big(\sum_{i=1}^{n-1} b_i + b_n\Big)\Big(\sum_{i=1}^{n-1} b_i \tilde{Y}_i(Z_i)\Big)}{(\sum_{i=1}^{n} b_i)(\sum_{i=1}^{n-1} b_i)} \\
&= \frac{\Big(\sum_{i=1}^{n-1} b_i\Big) b_n \tilde{Y}_n(Z_n) - b_n\Big(\sum_{i=1}^{n-1} b_i \tilde{Y}_i(Z_i)\Big)}{(\sum_{i=1}^{n} b_i)(\sum_{i=1}^{n-1} b_i)} \\
&= \frac{b_n \tilde{Y}_n(Z_n)}{(\sum_{i=1}^{n} b_i)} - \frac{b_n\Big(\sum_{i=1}^{n-1} b_i \tilde{Y}_i(Z_i)\Big)}{(\sum_{i=1}^{n} b_i)(\sum_{i=1}^{n-1} b_i)} = \tilde{a}_n \tilde{Y}_n(Z_n) - \tilde{a}_n \frac{\sum_{i=1}^{n-1} b_i \tilde{Y}_i(Z_i)}{\sum_{i=1}^{n-1} b_i} \\
&= \tilde{a}_n \tilde{Y}_n(Z_n) - \tilde{a}_n \overline{\Upsilon}_n.
\end{aligned}$$

As a result $\overline{\Upsilon}$ fulfills recursion (5.1). Consequently we can consider (5.1) as a weighted mean

$$\Upsilon_{n+1} = \frac{\sum_{i=1}^{n} b_i \widetilde{Y}_i(Z_i)}{\sum_{i=1}^{n} b_i}$$

of noisy function observations $\widetilde{Y}_n(Z_n)$. For example let $\widetilde{Y}_n(Z_n) = f(Z_n) + M_n(Z_n)$. Then

$$\Upsilon_{n+1} = \frac{\sum_{i=1}^{n} b_i \widetilde{Y}_i(Z_i)}{\sum_{i=1}^{n} b_i} = \frac{\sum_{i=1}^{n} b_i f(Z_i)}{\sum_{i=1}^{n} b_i} + \frac{\sum_{i=1}^{n} b_i M_i(Z_i)}{\sum_{i=1}^{n} b_i}. \tag{5.2}$$

According to Toeplitz's lemma (Lemma A.1.2) the first term in (5.2) converges to $f(z^*)$ for $Z_t \to z^*$. A law of large numbers can be applied to achieve the second term asymptotically vanishing.

Mokkadem and Pelletier [28] basically used two different methods to estimate the function value $f(z^*)$. The first one reuses the function evaluations taken by the Kiefer-Wolfowitz algorithm setting

$$\widetilde{Y}_n(Z_n) = \frac{1}{2|\mathcal{S}|} \sum_{i \in \mathcal{S}} Y_{n,i}(Z_n),$$

where

$$Y_{n,i}(Z_n) = \Big(f(Z_n + c_n e_i) + f(Z_n - c_n e_i) + M_{n,i}(Z_n)\Big)$$

with $M_{n,i}$ representing the observation noise, $\mathcal{S}$ a nonempty subset of $\{1, \ldots, d\}$, and

$|\mathcal{S}|$ denoting the cardinality of $\mathcal{S}$. Note that we do not need observations in all dimensions to make the algorithm work. Unfortunately it turns out that it is impossible to choose the sequences $(a_n)$, $(c_n)$ and $(\tilde{a}_n)$ such that $(Z_n)$ given in (1.2) and $(\Upsilon_n)$ given in (5.1) converge simultaneously with optimal rates (c.f. Corollaries 7.4.3 and 7.4.6). As another algorithm they investigated the average of $|\mathcal{S}|$ function evaluations

$$\widetilde{Y}_n(Z_n) = \frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} Y_{n,i}(Z_n), \text{ where } Y_{n,i}(Z_n) = \Big(f(Z_n) + M_{n,i}(Z_n)\Big). \tag{5.3}$$

When using (5.3) in (5.1), $\Upsilon_n$ is not explicitly dependent on $(c_n)$, but only implicitly via $Z_n$. Hence it is possible to choose $(a_n)$, $(c_n)$ and $(\tilde{a}_n)$ such that optimal convergence rates can be achieved simultaneously. However, $|\mathcal{S}|$ additional function evaluations per iteration step are required.

It is worth mentioning, that a slight modification of the second estimator (5.3) was presented in [28] as well. Instead of $Z_n$, a weighted mean $\bar{Z}_n := \frac{1}{\sum_{i=1}^n c_i^2}\sum_{i=1}^n c_i^2 Z_i$ was inserted, which results in replacing $\widetilde{Y}_n(Z_n)$ by

$$\bar{Y}_n(\bar{Z}_n) = \frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} Y_{n,i}(\bar{Z}_n), \text{ with } Y_{n,i}(\bar{Z}_n) = \Big(f(\bar{Z}_n) + M_{n,i}(\bar{Z}_n)\Big).$$

As the convergence rate of $(Z_n)$ transfers to $(\bar{Z}_n)$, they found out that it does not improve the convergence rate of $(\Upsilon_n)$ given in (5.1), albeit there are good reasons to prefer $(\bar{Z}_n)$ instead of $(Z_n)$ in the leading algorithms (e.g. the dilemma of asymptotics and stability mentioned earlier in Section 1.1). This is based on the fact that only the rates and not the bias or variance of the asymptotic distribution of the leading algorithm is employed to investigate the companion algorithms.

In this thesis the ideas of Mokkadem and Pelletier are generalized to the semimartingale framework. Weaker assumptions on the smoothness of the function $f$ are used. A new, generic algorithm is presented, without prescribing how to construct $\widetilde{Y}$. It is formulated in such a general way that it can easily be applied to further algorithms. Besides generalizations of the original companion algorithms of Mokkadem and Pelletier, two new algorithms are suggested. The first one estimates the Jacobian of $f$ at the root as a companion algorithm to the Robbins-Monro algorithm. The second one estimates the Hessian of $f$ at the minimizer on the basis of a leading Kiefer-Wolfowitz algorithm. Time-discrete as well as time-continuous versions follow as special cases.

## 5.2 General Assumptions

We now state assumptions that shall hold true for the rest of the thesis. We consider a stochastic basis $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ satisfying the usual conditions. This means that $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets of $\mathcal{F}$, and that the filtration $\mathbb{F}$ is right-continuous. On this basis an $\mathcal{F}_0$-measurable random variable $\Upsilon_0$ and a random field $(M(t,\upsilon))_{t\geqslant 0}$ are given, with $\upsilon \in \mathbb{R}$, $\upsilon \in \mathbb{R}^d$ or $\upsilon \in \mathbb{R}^{d\times d}$, depending on the algorithm. Define processes $(a_t)_{t\geqslant 0}$, $(c_t)_{t\geqslant 0}$, $(\tilde{a}_t)_{t\geqslant 0}$ and $(k_t)_{t\geqslant 0}$ that are predictable with respect to $\mathbb{F}$, and $(k_t)_{t\geqslant 0}$ is locally bounded. Furthermore it is assumed that $(R_t)_{t\geqslant 0}$ is increasing, càdlàg (i.e. right-continuous with left-sided limits), predictable with respect to $\mathbb{F}$, $R_0 = 0$ as well as $\Delta R_0 = 0$. By $\mathcal{M}^2_{\text{loc}}(\mathbb{P})$ we denote the set of locally square-integrable martingales with respect to $\mathbb{P}$ and $\mathbb{F}$. The random field $(M(t,\upsilon))_{t\geqslant 0}$ is $\mathbb{F}$-adapted, the relation $(M(t,\upsilon))_{t\geqslant 0} \in \mathcal{M}^2_{\text{loc}}(\mathbb{P})$ holds for every $\upsilon \in \mathbb{R}$, $\upsilon \in \mathbb{R}^d$ or $\upsilon \in \mathbb{R}^{d\times d}$, and $(\int_0^t k_s M(\mathrm{d}s, \Upsilon_{s-}))_{t\geqslant 0} \in \mathcal{M}^2_{\text{loc}}(\mathbb{P})$ is satisfied.

By $o$ and $\mathcal{O}$ we denote the Landau symbols. Moreover, for a stochastic process $(X_t)_{t\geqslant 0}$, we write $X_t = o_{\text{b}}(r_t)$ if $r_t$ is increasing to infinity and $(X_t/r_t)$ is bounded a.s. If $\lim_{t\to\infty} a_t/b_t = 1$, $a_t$ and $b_t$ are called asymptotically equal, which is denoted by $a_t \simeq b_t$. Moreover $R^d$ denotes the purely discontinuous part of $R$. By $\Delta R_t$ we define the jump $R_t - R_{t-}$. We note that $\Delta R_t = \mathrm{d}R^d_t$. The covariation and the predictable covariation processes of $(X_t)_{t\geqslant 0}$ and $(Y_t)_{t\geqslant 0}$ are denoted by $([X,Y]_t)_{t\geqslant 0}$ and $(\lceil X,Y\rceil_t)_{t\geqslant 0}$, respectively. The unit vectors of the Euclidean space $\mathbb{R}^d$ are written as $e_1,\ldots,e_d$. By $\mathcal{C}$ (and $\mathcal{C}(\omega)$) we denote a non-negative, real, generic constant (which also depends on $\omega \in \Omega$). If not stated otherwise, the statements concerning random variables and stochastic processes are to be interpreted in the almost surely sense.

## 5.3 The General Semimartingale Framework

The generic companion algorithm [Gen-Comp]

$$\Upsilon_t = \Upsilon_0 + \int_0^t \tilde{a}_s\Big(G_s - \Upsilon_{s-}\Big)\mathrm{d}R_s + \int_0^t k_s M(\mathrm{d}s, \Upsilon_{s-}), \tag{5.4}$$

is run to estimate $\upsilon^*$ consistently by $\Upsilon_t$. Essentially, in (5.4) we assume that the process $(\Upsilon_t)_{t\geqslant 0}$ can be decomposed in a finite variation part and a local martingale, and that there exists an intermediate process $(G_t)_{t\geqslant 0}$ which in general cannot be observed directly but approximate the quantity of interest $\upsilon^*$ sufficiently fast. Under appropriate conditions this will force $\Upsilon_t$ to converge to $\upsilon^*$ almost surely. In the algorithms of Mokkadem and Pelletier, introduced in Section 5.1, $G_t$ is given as $(2|\mathcal{S}|)^{-1}\sum_{i\in\mathcal{S}} f(Z_{t-}+c_te_i) + f(Z_{t-}+c_te_i)$ or $|\mathcal{S}|^{-1}\sum_{i\in\mathcal{S}} f(Z_{t-})$, $Z_t$ is generated by the Kiefer-Wolfowitz algorithm, and $\upsilon^* = f(z^*)$. Later on we deal with Robbins-Monro and Kiefer-Wolfowitz type leading algorithms although this general framework is not restricted to those two. The observation noise of $G_t - \Upsilon_{t-}$ is absorbed by $M$. Moreover, in this thesis we make the general assumption that for (5.4) there exists a unique strong solution $\Upsilon$ on $[0,\infty)$. The starting point $\Upsilon_0$ is a random variable or a fixed point. In practice the statistician chooses $\Upsilon_0$ either deterministically or in a random fashion. The processes $(\tilde{a}_t)_{t\geqslant 0}$ and $(k_t)_{t\geqslant 0}$ have a damping effect. It will turn out, that they must be chosen to be positive and monotonously decreasing to zero. The rate of convergence of $(\tilde{a}_t)_{t\geqslant 0}$ and $(k_t)_{t\geqslant 0}$ to zero is important as well. If the rate is chosen too slowly, $\Upsilon_t$ in (5.4) will not converge, whereas choosing it too high doesn't ensure the convergence to $\upsilon^*$ anymore. Typically one chooses $R_t := \lfloor t\rfloor$ or $R_t := t$. In the first case new observations are only taken at times $t \in \mathbb{N}$, whereas in the second case there is a continuous update of data. The semimartingale framework of (5.4), however enables to chose $(R_t)_{t\geqslant 0}$ as a stochastic process. For example it is possible to model a situation in which new updates of data can only be taken at random times.

# 6 Almost Sure Convergence of Companion Algorithms

In this chapter, consistency of companion algorithms is investigated. After proving consistency of the generic algorithm, companion algorithms for the Kiefer-Wolfowitz algorithm and the Robbins-Monro algorithm are investigated. These can be $\mathbb{R}$, $\mathbb{R}^d$-, or $\mathbb{R}^{d\times d}$-valued. For the sake of clarity assume the companion process $(\Upsilon_t)_{t\geqslant 0}$ to be $\mathbb{R}$-valued. A generalization to $\mathbb{R}^d$- and $\mathbb{R}^{d\times d}$-valued processes is straightforward.

## 6.1 Consistency of the Generic Algorithm

Usually the first and most important question concerning an estimator is if it is convergent. Typically if it is not, further investigation is redundant. Moreover many results, as for instance on the rate of convergence, assume consistency. In order to show strong consistency we state the following conditions.

**Assumption 6.1.1.**

*(A)* $(G_t)_{t\geqslant 0}$ *is an adapted left-continuous process with* $G_t \xrightarrow{t\to\infty} \upsilon^*$ $\mathbb{P}$*-a.s.*

*(B)* *Let* $(\tilde{a}_t)_{t\geqslant 0}$ *satisfy*

$$\tilde{a}_t > 0, \quad \tilde{a}_t \downarrow 0 \quad \text{and} \quad \int_0^\infty \tilde{a}_s \mathrm{d}R_s = \infty.$$

*(C)* *Assume*

$$\int_0^\infty \tilde{a}_s |G_s - \upsilon^*| \mathrm{d}R_s < \infty.$$

*(D)* *For all* $y \in \mathbb{R}$ *assume there exists a process* $(k_t)_{t\geqslant 0}$ *with*

$$\int_0^\infty k_s^2 \frac{h_s(\Upsilon_{s-})}{1+\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \quad \text{where} \quad h_s(y) := \frac{\mathrm{d}[\int_0^\cdot M(\mathrm{d}t, y)]_s}{\mathrm{d}R_s}.$$

*(E)* *If the process* $\Upsilon$ *is not purely continuous, assume*

$$\int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d < \infty.$$

**Theorem 6.1.1.** *If Assumption 6.1.1 is satisfied, then for* (5.4), $\Upsilon_t \to \upsilon^*$ $\mathbb{P}$*-a.s. as* $t \to \infty$.

*Remark* 6.1.1. Condition $(C)$ ensures that the leading algorithm converges sufficiently fast to $z^*$.

*Remark* 6.1.2. Sufficient conditions for assumption $(D)$ are $h_s(y) \leqslant K_s(1 + |y|^2)$ together with $\int_0^\infty k_s^2 K_s \mathrm{d}R_s < \infty$. Even simpler it is to assume $h_s(y) \leqslant \mathcal{C} < \infty$ together with $\int_0^\infty k_s^2 \mathrm{d}R_s < \infty$.

*Remark* 6.1.3. Condition $(E)$ guarantees that the damped jumps $a_s \Delta R_s$ tend to zero.

*Remark* 6.1.4. Note that under appropriate conditions the rates of convergence of $(Z_t)_{t\geqslant 0}$ transfer to that of some weighted average process $(\bar{Z}_t)_{t\geqslant 0}$, c.f. Schnizler [37, Theorem 4.1]. Consequently the result of Theorem 6.1.1 holds true if we replace $(Z_t)_{t\geqslant 0}$ by $(\bar{Z}_t)_{t\geqslant 0}$ in algorithm (5.4).

*Proof of Theorem 6.1.1.* We consider the stochastic integral equation (5.4). Without loss of generality let $\upsilon^* = 0$.

The idea of the proof is to bound $X := \Upsilon^2$ by $A_t^1 - A_t^2 + \tilde{M}$, with predictable, increasing processes $A^1$, $A^2$ and a local martingale $\widetilde{M}$.

In a first step the Robbins-Siegmund lemma (Lemma A.1.1 in the appendix) applied to $A^1$ yields $\mathbb{P}$-a.s. convergence of $X$. Applying the same lemma also to $A^2$ yields $\mathbb{P}$-a.s. convergence of $X$ to 0. In that second part, the punchline is different from consistency proofs of classical stochastic approximation algorithms like Robbins-Monro or Kiefer-Wolfowitz. Application of the Robbins-Siegmund lemma for the consistency proof of companion algorithms has not been performed before. Mokkadem and Pelletier used a different method for which they needed unlike stronger assumptions and traced it back to a consistency theorem on Robbins-Monro algorithms.

Application of integration by parts [32, II.6.Cor. 2] yields

$$\mathrm{d}\Upsilon_s^2 = 2\Upsilon_{s-}\mathrm{d}\Upsilon_s + \mathrm{d}[\Upsilon]_s,$$

where

$$\Upsilon_{s-}\mathrm{d}\Upsilon_s = \tilde{a}_s \Upsilon_{s-} G_s \mathrm{d}R_s - \tilde{a}_s \Upsilon_{s-}^2 \mathrm{d}R_s + k_s \Upsilon_{s-} M(\mathrm{d}s, \Upsilon_{s-})$$

and

$$\begin{aligned}
\mathrm{d}[\Upsilon]_s &= \tilde{a}_s^2 (G_s - \Upsilon_{s-})^2 \Delta R_s \mathrm{d}R_s^d + 2\tilde{a}_s k_s (G_s - \Upsilon_{s-}) \Delta R_s M(\mathrm{d}s, \Upsilon_{s-}) \\
&\quad + k_s^2 \mathrm{d}[\int_0^\cdot M(\mathrm{d}s, \Upsilon_{s-})]_s \\
&= \tilde{a}_s^2 G_s^2 \Delta R_s \mathrm{d}R_s^d - 2\tilde{a}_s^2 G_s \Upsilon_{s-} \Delta R_s \mathrm{d}R_s^d + \tilde{a}_s^2 \Upsilon_{s-}^2 \Delta R_s \mathrm{d}R_s^d \\
&\quad + 2\tilde{a}_s k_s G_s \Delta R_s M(\mathrm{d}s, \Upsilon_{s-}) - 2\tilde{a}_s k_s \Upsilon_{s-} \Delta R_s M(\mathrm{d}s, \Upsilon_{s-}) \\
&\quad + k_s^2 \mathrm{d}[\int_0^\cdot M(\mathrm{d}s, \Upsilon_{s-})]_s.
\end{aligned}$$

Therefore we have

$$\mathrm{d}\Upsilon_s^2 = 2\tilde{a}_s \Upsilon_{s-} G_s \mathrm{d}R_s - 2\tilde{a}_s \Upsilon_{s-}^2 \mathrm{d}R_s + 2k_s \Upsilon_{s-} M(\mathrm{d}s, \Upsilon_{s-})$$

$$
\begin{aligned}
&+ \tilde{a}_s^2 G_s^2 \Delta R_s \mathrm{d}R_s^d - 2\tilde{a}_s^2 G_s \Upsilon_{s-} \Delta R_s \mathrm{d}R_s^d + \tilde{a}_s^2 \Upsilon_{s-}^2 \Delta R_s \mathrm{d}R_s^d \\
&+ 2\tilde{a}_s k_s G_s \Delta R_s M(\mathrm{d}s, \Upsilon_{s-}) - 2\tilde{a}_s k_s \Upsilon_{s-} \Delta R_s M(\mathrm{d}s, \Upsilon_{s-}) \\
&+ k_s^2 \mathrm{d}[\int_0^{\cdot} M(\mathrm{d}s, \Upsilon_{s-})]_s. 
\end{aligned} \tag{6.1}
$$

In order to apply the Robbins-Siegmund lemma, we define

$$
\begin{aligned}
\mathrm{d}A_t^1 &:= 2\tilde{a}_s|\Upsilon_{s-}G_s|\mathrm{d}R_s + 2\tilde{a}_s^2|G_s\Upsilon_{s-}|\Delta R_s \mathrm{d}R_s^d + \tilde{a}_s^2 G_s^2 \Delta R_s \mathrm{d}R_s^d + \tilde{a}_s^2 \Upsilon_{s-}^2 \Delta R_s \mathrm{d}R_s^d \\
&\quad + k_s^2 \mathrm{d}\left[\int_0^{\cdot} M(\mathrm{d}s, \Upsilon_{s-})\right]_s \\
-\mathrm{d}A_t^2 &:= -2\tilde{a}_s \Upsilon_{s-}^2 \mathrm{d}R_s \\
\mathrm{d}\tilde{M}_t &:= +2k_s \Upsilon_{s-} M(\mathrm{d}s, \Upsilon_{s-}) + 2\tilde{a}_s k_s G_s \Delta R_s M(\mathrm{d}s, \Upsilon_{s-}) \\
&\quad - 2\tilde{a}_s k_s \Upsilon_{s-} \Delta R_s M(\mathrm{d}s, \Upsilon_{s-}) \\
&\quad + k_s^2 \mathrm{d}\left(\left[\int_0^{\cdot} M(\mathrm{d}s, \Upsilon_{s-})\right]_s - \left[\int_0^{\cdot} M(\mathrm{d}s, \Upsilon_{s-})\right]_s\right).
\end{aligned}
$$

If we can show $\int_0^\infty \frac{1}{1+\Upsilon_{s-}^2} \mathrm{d}A_s^1 < \infty$, the Robbins-Siegmund lemma yields that $(\Upsilon_t)_{t \geqslant 0}$ converges and $\int_0^\infty \mathrm{d}A_s^2 < \infty$.

We now bound the term $\int_0^\infty \frac{1}{1+\Upsilon_{s-}^2} \mathrm{d}A_s^1$. Assumption $(C)$ yields that

$$
\int_0^\infty \tilde{a}_s \frac{|\Upsilon_{s-}G_s|}{1+\Upsilon_{s-}^2} \mathrm{d}R_s \leqslant \frac{1}{2} \int_0^\infty \tilde{a}_s |G_s| \mathrm{d}R_s < \infty.
$$

For the second and the third term it holds by assumptions $(A)$ and $(E)$

$$
\int_0^\infty \tilde{a}_s^2 \frac{|\Upsilon_{s-}G_s|}{1+\Upsilon_{s-}^2} \Delta R_s \mathrm{d}R_s^d \leqslant \mathcal{C}(\omega) \int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d < \infty
$$

and

$$
\int_0^\infty \tilde{a}_s^2 \frac{1}{1+\Upsilon_{s-}^2} G_s^2 \Delta R_s \mathrm{d}R_s^d \leqslant \mathcal{C}(\omega) \int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d < \infty,
$$

respectively. With condition $(E)$ the fourth term is bounded by

$$
\int_0^\infty \tilde{a}_s^2 \frac{\Upsilon_{s-}^2}{1+\Upsilon_{s-}^2} \Delta R_s \mathrm{d}R_s^d \leqslant \int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d < \infty
$$

and the last term, according to assumption $(D)$, is bounded by

$$
\int_0^\infty k_s^2 \frac{1}{1+\Upsilon_{s-}^2} \mathrm{d}[\int_0^{\cdot} M(\mathrm{d}s, \Upsilon_{s-})]_s = \int_0^\infty k_s^2 \frac{1}{1+\Upsilon_{s-}^2} h_s(\Upsilon_{s-}) \mathrm{d}R_s < \infty.
$$

As a result, we conclude

$$\int_0^\infty \frac{1}{1+\Upsilon_{s-}^2} \mathrm{d}A_s^1 < \infty.$$

Therefore the Robbins-Siegmund lemma (Lemma A.1.1) yields that $(\Upsilon_t^2)_{t\geqslant 0}$ converges and $A_\infty^2 < \infty$ a.s.

The convergence of $\Upsilon$ to 0 is shown by contradiction. For that purpose assume a set $N$ of non-zero probability on which the solution of the stochastic integral equation does not converge to zero. We will deduce a contradiction to

$$\Omega = \{A_\infty^2 < \infty\}.$$

As proven before, $\Upsilon$ converges for almost all $\omega \in \Omega$, but by assumption for all $\omega \in N$ the process does not converge to 0. Hence it follows for all $\omega \in N$

$$\underset{\epsilon^*>0}{\exists} \quad \underset{s_0}{\exists} \quad \underset{s\geqslant s_0}{\forall} \quad \epsilon^* \leqslant \Upsilon_s^2 \leqslant 1/\epsilon^*.$$

In $A_t^2$ the term

$$\int_0^t \tilde{a}_s \Upsilon_{s-}^2 \mathrm{d}R_s$$

is non-negative. Consequently, with condition (B),

$$\begin{aligned} A_\infty^2 &\geqslant \int_0^\infty \tilde{a}_s \Upsilon_{s-}^2 \mathrm{d}R_s = \int_0^{s_0} \tilde{a}_s \Upsilon_{s-}^2 \mathrm{d}R_s + \int_{s_0+}^\infty \tilde{a}_s \Upsilon_{s-}^2 \mathrm{d}R_s \\ &\geqslant \mathcal{C} + \epsilon^* \int_{s_0+}^\infty \tilde{a}_s \mathrm{d}R_s = \infty. \end{aligned}$$

This is a contradiction to what we have shown before. Consequently the set $N$ cannot exist. We conclude $\Upsilon_t^2 \to 0$ and thereby $\Upsilon_t \to 0$ a.s. for $t \to \infty$. □

## 6.2 Consistency of Special Algorithms

We consider two types of leading algorithms, namely the Robbins-Monro algorithm [RM]

$$Z_t = Z_0 - \int_0^t a_s f(Z_{s-}) \mathrm{d}R_s - \int_0^t a_s M(\mathrm{d}s, Z_{s-}) \tag{6.2}$$

and the Kiefer-Wolfowitz algorithm [KW]

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s} \Big\{ f(Z_{s-} + c_s e_i) - f(Z_{s-} - c_s e_i) \Big\}_{i\in\{1,\dots,d\}} \mathrm{d}R_s - \int_0^t \frac{a_s}{2c_s} M(\mathrm{d}s, Z_{s-}). \tag{6.3}$$

In both cases $Z_0$ is assumed to be $\mathcal{F}_0$-measurable and $z^*$ to exist as the limit of the unique solution of $(Z_t)_{t\geqslant 0}$ in $[0,\infty)$ and denotes the parameter that is approximated by the corresponding leading algorithm. These parameters are the root of $f$ estimated by the Robbins-Monro algorithm and the minimizer of $f$ estimated by the Kiefer-Wolfowitz algorithm, respectively.

Let us consider semimartingale versions of two novel algorithms. To estimate the Jacobian $\upsilon^* := J_{z^*}$ at the root of $f\colon \mathbb{R}^d \to \mathbb{R}^d$ on basis of a leading Robbins-Monro algorithm use [RM-J]

$$\Upsilon_t = \Upsilon_0 + \int_0^t \tilde{a}_s\left(\left\{\frac{1}{c_s}\Big(f(Z_{s-}+c_s) - f(Z_{s-})\Big)\right\} - \Upsilon_{s-}\right)\mathrm{d}R_s + \int_0^t \frac{\tilde{a}_s}{c_s} M(\mathrm{d}s, \Upsilon_{s-}). \tag{6.4}$$

As an estimator of the Hessian $\upsilon^* := H_{z^*}$ of $f\colon \mathbb{R}^d \to \mathbb{R}$ at a minimizer based on [KW] consider [KW-H]

$$\begin{aligned}\Upsilon_t = \Upsilon_0 &+ \int_0^t \tilde{a}_s\left(\left\{\frac{1}{c_s^2}\Big(f(Z_{s-}+c_s) + f(Z_{s-}-c_s) - 2f(Z_{s-})\Big)\right\} - \Upsilon_{s-}\right)\mathrm{d}R_s \\ &+ \int_0^t \frac{\tilde{a}_s}{c_s^2} M(\mathrm{d}s, \Upsilon_{s-}).\end{aligned} \tag{6.5}$$

Besides that, semimartingale versions of the algorithms to estimate $\upsilon^* := f(z^*)$ on basis of a leading [KW] that were presented by Mokkadem and Pelletier are surveyed: Algorithm [KW-F-2]

$$\begin{aligned}\Upsilon_t = \Upsilon_0 &+ \int_0^t \tilde{a}_s\left(\left\{\frac{1}{2|\mathcal{S}|}\sum_{i\in\mathcal{S}} f(Z_{s-}+c_s e_i) + f(Z_{s-}-c_s e_i)\right\} - \Upsilon_{s-}\right)\mathrm{d}R_s \\ &+ \int_0^t \tilde{a}_s M(\mathrm{d}s, \Upsilon_{s-})\end{aligned} \tag{6.6}$$

which recycles the observations made in the leading algorithm (6.3), and [KW-F-1]

$$\Upsilon_t = \Upsilon_0 + \int_0^t \tilde{a}_s\Big(f(Z_{s-}) - \Upsilon_{s-}\Big)\mathrm{d}R_s + \int_0^t \tilde{a}_s M(\mathrm{d}s, \Upsilon_{s-}), \tag{6.7}$$

requiring an additional observation. To keep notations simple, all proofs in this thesis referring to these two algorithms (6.4) and (6.5) are only covered in the one-dimensional case, albeit an extension to the multi-dimensional setting is straightforward.

**Assumption 6.2.1.** *Let Assumption 6.1.1 hold and replace conditions (A), (C) and (D) by the following ones.*

*(Asp)* *Let*

$$f\colon \mathbb{R}^d \to \mathbb{R}^d \text{ with } J \text{ be Lipschitz continuous in [RM-J] (6.4),}$$

$$
\begin{aligned}
&f\colon \mathbb{R}^d \to \mathbb{R} \textit{ with } H \textit{ be Lipschitz continuous in [KW-H] } (6.5),\\
&f\colon \mathbb{R}^d \to \mathbb{R} \textit{ with } \nabla f \textit{ be Lipschitz continuous in [KW-F-2] } (6.6),\\
&f\colon \mathbb{R}^d \to \mathbb{R} \textit{ with } \nabla f \textit{ be Lipschitz continuous in [KW-F-1] } (6.7).
\end{aligned}
$$

*(Csp)* *There exists a left-continuous adapted process* $(r_t)_{t\geqslant 0}$ *with* $r_t \to 0$ $\mathbb{P}$*-a.s. for* $t \to \infty$ *such that* $\|Z_t - z^*\| = \mathcal{O}(r_t)$ $\mathbb{P}$*-a.s. and*

$$
\begin{aligned}
&\int_0^\infty \tilde{a}_s r_s \mathrm{d}R_s < \infty \quad \textit{and} \quad \int_0^\infty \tilde{a}_s c_s \mathrm{d}R_s < \infty \quad \textit{for [RM-J] } (6.4),\\
&\int_0^\infty \tilde{a}_s r_s \mathrm{d}R_s < \infty \quad \textit{and} \quad \int_0^\infty \tilde{a}_s c_s \mathrm{d}R_s < \infty \quad \textit{for [KW-H] } (6.5),\\
&\int_0^\infty \tilde{a}_s r_s^2 \mathrm{d}R_s < \infty \quad \textit{and} \quad \int_0^\infty \tilde{a}_s c_s^2 \mathrm{d}R_s < \infty \quad \textit{for [KW-F-2] } (6.6),\\
&\int_0^\infty \tilde{a}_s r_s^2 \mathrm{d}R_s < \infty \quad \textit{for [KW-F-1] } (6.7).
\end{aligned}
$$

*(Dsp)* *Let*

$$
\begin{aligned}
&\int_0^\infty \left(\frac{\tilde{a}_s}{c_s}\right)^2 \frac{h_s(\Upsilon_{s-})}{1+\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \quad \textit{for [RM-J] } (6.4),\\
&\int_0^\infty \left(\frac{\tilde{a}_s}{c_s^2}\right)^2 \frac{h_s(\Upsilon_{s-})}{1+\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \quad \textit{for [KW-H] } (6.5),\\
&\int_0^\infty \tilde{a}_s^2 \frac{h_s(\Upsilon_{s-})}{1+\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \quad \textit{for [KW-F-2] } (6.6) \textit{ and [KW-F-1] } (6.7).
\end{aligned}
$$

**Theorem 6.2.1.** *Let Assumption 6.2.1 hold. Then the companion algorithm, given as the solution of* (6.4), (6.5), (6.6) *or* (6.7) *is consistent, i.e.* $\Upsilon_t \to \upsilon^*$ $\mathbb{P}$*-a.s. as* $t \to \infty$.

*Proof.* We trace the result back to Theorem 6.1.1. Algorithm (5.4) can be rewritten as

$$
\Upsilon_t = \Upsilon_0 + \int_0^t \tilde{a}_s\Big(G_s - \upsilon^* - (\Upsilon_{s-} - \upsilon^*)\Big)\mathrm{d}R_s + \int_0^t k_s M(\mathrm{d}s, \Upsilon_{s-}).
$$

Depending on the considered companion algorithm, we get $G_s$ from (6.4), (6.5), (6.6) or (6.7) and perform a Taylor expansion. The convergence rates that are achieved in the terms of the expansion enable us to deduce the validity of assumptions $(A)$ and $(C)$ from $(Asp)$ and $(Csp)$.

Considering algorithm [KW-F-2] as given in (6.6), a Taylor expansion yields

$$
\begin{aligned}
&G_s - \upsilon^*\\
&\quad\leqslant \left|\frac{1}{2|\mathcal{S}|}\sum_{i\in\mathcal{S}}\Big(f(Z_{s-}+c_s e_i) + f(Z_{s-}-c_s e_i)\Big) - f(z^*)\right|\\
&\quad\leqslant \frac{1}{2|\mathcal{S}|}\left|\sum_{i\in\mathcal{S}}\Big(\int_0^1 \Big\langle c_s e_i, \nabla f(Z_{s-}+tc_s e_i) - \nabla f(Z_{s-}-tc_s e_i)\Big\rangle \mathrm{d}t\Big)\right|
\end{aligned}
$$

$$
\begin{aligned}
&+ |f(Z_{s-}) - f(z^*)| \\
&\leqslant \frac{1}{2|\mathcal{S}|}\left|\sum_{i\in\mathcal{S}}\left(\int_0^1 \Big\langle c_s e_i, \nabla f(Z_{s-} + tc_s e_i) - \nabla f(Z_{s-}) \right.\right. \\
&\quad \left.\left. + \nabla f(Z_{s-}) - \nabla f(Z_{s-} - tc_s e_i)\Big\rangle \mathrm{d}t\right)\right| \\
&\quad + |f(Z_{s-}) - f(z^*)| \\
&\leqslant \mathcal{C}\sum_{i\in\mathcal{S}}\left(\int_0^1 |c_s e_i|\Big(L|tc_s e_i| + L|tc_s e_i|\Big)\mathrm{d}t\right) + |f(Z_{s-}) - f(z^*)| \\
&= \mathcal{O}(c_s^2) + |f(Z_{s-}) - f(z^*)|\,.
\end{aligned}
$$

Furthermore

$$
\begin{aligned}
f(Z_{s-}) - f(z^*) &\leqslant \int_0^1 \|Z_{s-} - z^*\|\|\nabla f(z^* + t(Z_{s-} - z^*))\|\mathrm{d}t \\
&= \int_0^1 \|Z_{s-} - z^*\|\|\nabla f(z^* + t(Z_{s-} - z^*)) - \nabla f(z^*)\|\mathrm{d}t \\
&\leqslant \mathcal{C}\int_0^1 \|Z_{s-} - z^*\|\|t(Z_{s-} - z^*)\|\mathrm{d}t = \mathcal{O}(\|Z_{s-} - z^*\|^2)
\end{aligned}
$$

holds, because $\nabla f(z^*)$ is equal to zero. Consequently

$$
G_s - \upsilon^* = \mathcal{O}(\|Z_{s-} - z^*\|^2) + \mathcal{O}(c_s^2) = \mathcal{O}(r_s^2) + \mathcal{O}(c_s^2). \tag{6.8}
$$

An analogous calculation for algorithm (6.7) results to

$$
G_s - \upsilon^* = \mathcal{O}(r_s^2). \tag{6.9}
$$

Investigating algorithm (6.4) yields

$$
\begin{aligned}
\left|\frac{1}{c_s}\Big(f(Z_{s-} + c_s) - f(Z_{s-})\Big) - J_{z^*}\right| &= \left|\int_0^1 J_{Z_{s-}+tc_s}\mathrm{d}t - J_{z^*}\right| = \left|\int_0^1 J_{Z_{s-}+tc_s} - J_{z^*}\mathrm{d}t\right| \\
&= \left|\int_0^1 J_{Z_{s-}+tc_s} - J_{Z_{s-}} + J_{Z_{s-}} - J_{z^*}\mathrm{d}t\right| \\
&\leqslant \left|\int_0^1 L|Z_{s-} + tc_s - Z_{s-}| + L|Z_{s-} - z^*|\mathrm{d}t\right| \\
&= \mathcal{O}(c_s) + \mathcal{O}(r_s).
\end{aligned}
$$

Finally, for algorithm (6.5)

$$
\begin{aligned}
&\left|\frac{1}{c_s^2}\Big(f(Z_{s-} + c_s) + f(Z_{s-} - c_s) - 2f(Z_{s-})\Big) - H_{z^*}\right| \\
&\quad = \left|\int_0^1 \frac{t}{c_s}\Big(J_{Z_{s-}+tc_s} + J_{Z_{s-}-tc_s}\Big)\mathrm{d}t - H_{z^*}\right|
\end{aligned}
$$

$$
\begin{aligned}
&= \left| \int_0^1 \int_{-1}^1 t H_{Z_{s-}+tuc_s} \mathrm{d}u\mathrm{d}t - H_{z^*} \right| = \left| \int_0^1 \int_{-1}^1 t\Big(H_{Z_{s-}+tuc_s} - H_{z^*}\Big) \mathrm{d}u\mathrm{d}t \right| \\
&= \left| \int_0^1 \int_{-1}^1 t\Big(H_{Z_{s-}+tuc_s} - H_{Z_{s-}} + H_{Z_{s-}} - H_{z^*}\Big) \mathrm{d}u\mathrm{d}t \right| \\
&= \left| \int_0^1 \int_{-1}^1 |t|L|tuc_s| + L|Z_{s-} - z^*|\Big) \mathrm{d}u\mathrm{d}t \right| = \mathcal{O}(c_s) + \mathcal{O}(r_s),
\end{aligned}
$$

holds where $H_z$ denotes the Hessian of $f$ at $z$. As a result we have

$$
G_s - \upsilon^* = \Big\{ \mathcal{O}(c_s) + \mathcal{O}(r_s) \quad \text{for [RM-J] (6.4) and [KW-H] (6.5).} \tag{6.10}
$$

Consequently we deduced the validity of assumption $(A)$ from $(Asp)$. Now (6.8), (6.9) and (6.10) are used such that

$$
\begin{aligned}
\int_0^\infty \tilde{a}_s G_s \mathrm{d}R_s \leqslant \mathcal{C}(\omega) + &\begin{cases}
\mathcal{C}(\omega) \int_{\tau(\omega)}^\infty \tilde{a}_s (r_s^2 + c_s^2) \mathrm{d}R_s & \text{for [KW-F-2] (6.6)} \\
\mathcal{C}(\omega) \int_{\tau(\omega)}^\infty \tilde{a}_s r_s^2 \mathrm{d}R_s & \text{for [KW-F-1] (6.7)} \\
\mathcal{C}(\omega) \int_{\tau(\omega)}^\infty \tilde{a}_s (c_s + r_s) \mathrm{d}R_s & \text{for [RM-J] (6.4)} \\
\mathcal{C}(\omega) \int_{\tau(\omega)}^\infty \tilde{a}_s (c_s + r_s) \mathrm{d}R_s & \text{for [KW-H] (6.5)}
\end{cases} \\
&< \infty
\end{aligned}
$$

directly yields $(C)$ from $(Csp)$. Condition $(D)$ follows from $(Dsp)$ by replacing $k_s^2$ by $(\tilde{a}_s/c_s)^2$, $(\tilde{a}_s/c_s^2)^2$ or $\tilde{a}_s^2$, respectively. □

## 6.3 Itô-Type and Recursive Stochastic Approximation Algorithms

Consider the Itô type, continuous generic companion algorithm [c-Gen-Comp]

$$
\Upsilon_t = \Upsilon_0 + \int_0^t \tilde{a}_s \Big(G_s - \Upsilon_s\Big) \mathrm{d}s + \int_0^t k_s \sigma_s(\Upsilon_s) \mathrm{d}W_s, \tag{6.11}
$$

under the following assumptions.

**Assumption 6.3.1.**

*(cA) There exists an adapted continuous process* $(G_t)_{t \geqslant 0}$ *with* $G_t \xrightarrow{t\to\infty} \upsilon^*$ $\mathbb{P}$*-a.s.*

*(cB)* $(\tilde{a}_t)_{t \geqslant 0}$ *is continuous with*

$$
\tilde{a}_t > 0, \quad \tilde{a}_t \downarrow 0 \quad \text{and} \quad \int_0^\infty \tilde{a}_s \mathrm{d}s = \infty.
$$

*(cC)*

$$\int_0^\infty \tilde{a}_s|G_s - \upsilon^*|\mathrm{d}s < \infty$$

*(cD)*

$$\int_0^\infty k_s^2 \frac{\sigma_s^2(\Upsilon_s)}{1+\Upsilon_s^2}\mathrm{d}s < \infty.$$

**Corollary 6.3.1.** *Let Assumption 6.3.1 hold. Then the solution process $(\Upsilon_t)_{t\geqslant 0}$ of the Itô type stochastic integral equation [c-Gen-Comp] given in* (6.11) *converges almost surely to $\upsilon^*$.*

*Proof.* Without loss of generality let $\upsilon^* = 0$. Setting $R_s := s$ and $M(\mathrm{d}s, y) := \sigma_s(y)\mathrm{d}W_s$ we get the corresponding Itô type stochastic integral equation from the semimartingale stochastic integral equation. Moreover $(cA)$ yields $(A)$. Condition $(B)$ is directly deduced from $(cB)$. Assumption $(cC)$ implies $(C)$. Continuity of $(R_s)_{s\geqslant 0}$ yields assumption $(E)$. Assumption $(D)$ follows from $(cD)$ by

$$\lceil \int_0^\cdot M(\mathrm{d}s, y)\rceil_t = \lceil \int_0^\cdot \sigma_s(y)\mathrm{d}W_s\rceil_t = \int_0^t \sigma_s^2(y)\mathrm{d}s$$

and $h_s(y) = \sigma_s^2(y)$. Consequently all conditions of Theorem 6.1.1 are verified and the corollary is proven. □

The following algorithms are the Itô type stochastic integral equations of the Robbins-Monro algorithm [c-RM]

$$Z_t = Z_0 - \int_0^t a_s f(Z_s)\mathrm{d}s - \int_0^t a_s\sigma_s(Z_s)\mathrm{d}W_s \tag{6.12}$$

with diffusion function $\sigma\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ and a $d$-dimensional standard Brownian motion $W$, and Kiefer-Wolfowitz algorithm [c-KW]

$$Z_t = Z_0 - \int_0^t \frac{a_s}{2c_s}\Big\{f(Z_s + c_s e_i) - f(Z_s - c_s e_i)\mathrm{d}s + \sum_{j=1}^d \sigma_s^{ij}(Z_s)\mathrm{d}W_s^j\Big\}_{i\in\{1,\dots,d\}} \tag{6.13}$$

with diffusion function $\sigma^{ij}\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ and $d$ independent 1-dimensional standard Brownian motions $W^j$. Detailled analyses can be found in Lazrieva et al. [22] and Schnizler [37].

In an analogous way to the semimartingale case we may deduce time-continuous algorithms [c-RM-J], [c-KW-H], [c-KW-F-2] and [c-KW-F-1] as special cases.

**Assumption 6.3.2.** *Let Assumption 6.3.1 hold, with (cA), (cC) and (cD) replaced by the following conditions.*

*(cAsp)* *Let*

$f\colon \mathbb{R}^d \to \mathbb{R}^d$ *with $J$ be Lipschitz continuous in [c-RM-J],*

*$f\colon \mathbb{R}^d \to \mathbb{R}$ with $H$ be Lipschitz continuous in [c-KW-H],*

*$f\colon \mathbb{R}^d \to \mathbb{R}$ with $\nabla f$ be Lipschitz continuous in [c-KW-F-2] and [c-KW-F-1].*

*($cCsp$)* *There exists a continuous adapted process $(r_t)_{t\geqslant 0}$ with $r_t \to 0$ for $t \to \infty$ such that $\|Z_t - z^*\| = \mathcal{O}(r_t)$ $\mathbb{P}$-a.s. as well as*

$$\begin{aligned}
&\int_0^\infty \tilde{a}_s r_s \mathrm{d}s < \infty \quad \text{and} \quad \int_0^\infty \tilde{a}_s c_s \mathrm{d}s < \infty \quad \text{for [c-RM-J]},\\
&\int_0^\infty \tilde{a}_s r_s \mathrm{d}s < \infty \quad \text{and} \quad \int_0^\infty \tilde{a}_s c_s \mathrm{d}s < \infty \quad \text{for [c-KW-H]},\\
&\int_0^\infty \tilde{a}_s r_s^2 \mathrm{d}s < \infty \quad \text{and} \quad \int_0^\infty \tilde{a}_s c_s^2 \mathrm{d}s < \infty \quad \text{for [c-KW-F-2]},\\
&\int_0^\infty \tilde{a}_s r_s^2 \mathrm{d}s < \infty \quad \text{for [c-KW-F-1]}.
\end{aligned}$$

*For [c-KW-H] and [c-KW-F-2], $(c_s)_{s\geqslant 0}$ is the non-negative process from the leading Kiefer-Wolfowitz algorithm.*

*($cDsp$)* *Let*

$$\begin{aligned}
&\int_0^\infty \left(\frac{\tilde{a}_s}{c_s}\right)^2 \frac{\sigma_s^2(\Upsilon_s)}{1+\Upsilon_s^2} \mathrm{d}s < \infty \quad \text{for [c-RM-J]},\\
&\int_0^\infty \left(\frac{\tilde{a}_s}{c_s^2}\right)^2 \frac{\sigma_s^2(\Upsilon_s)}{1+\Upsilon_s^2} \mathrm{d}s < \infty \quad \text{for [c-KW-H]},\\
&\int_0^\infty \tilde{a}_s^2 \frac{\sigma_s^2(\Upsilon_s)}{1+\Upsilon_s^2} \mathrm{d}s < \infty \quad \text{for [c-KW-F-2] and [c-KW-F-1]}.
\end{aligned}$$

**Corollary 6.3.2.** *Under Assumption 6.3.2 the solutions of the Itô type stochastic integral equations [c-RM-J], [c-KW-H], [c-KW-F-2] and [c-KW-F-1] converge almost surely to $\upsilon^*$.*

*Proof.* Without loss of generality let $\upsilon^* = 0$. Setting $R_s := s$ and $M(\mathrm{d}s, y) := \sigma_s(y)\mathrm{d}W_s$ we get the corresponding Itô type stochastic integral equation from the semimartingale stochastic integral equation. This also implies ($Asp$) and ($Csp$) from ($cAsp$) and ($cCsp$). Following the proof of Corollary 6.3.1 yields ($B$), ($D$) and ($E$) such that all conditions of Theorem 6.2.1 are fulfilled. □

Now we consider a generic time-discrete companion algorithm [d-Gen-Comp]

$$\Upsilon_n - \Upsilon_{n-1} = \tilde{a}_n\Big(G_n - \Upsilon_{n-1}\Big) + k_n V_n, \tag{6.14}$$

under the following assumptions.

**Assumption 6.3.3.**

*($dA$)* *There exists a sequence $(G_n)$ with $G_n \xrightarrow{n\to\infty} \upsilon^*$ $\mathbb{P}$-a.s.*

*(dB)* *The sequence* $(\tilde{a}_n)$ *satisfies*

$$\tilde{a}_n > 0, \quad \tilde{a}_n \downarrow 0 \quad \textit{and} \quad \sum_{n=1}^{\infty} \tilde{a}_n = \infty.$$

*(dC)*

$$\sum_{n=1}^{\infty} \tilde{a}_n |G_n - \upsilon^*| < \infty$$

*(dD)* *Let* $\sum_{n=1}^{\infty} k_n^2 < \infty$ *and*

$$\sup_{n \in \mathbb{N}} \mathbb{E}\|V_n\|^2 < \infty \quad \textit{and} \quad \mathbb{E}(V_n \mid \mathcal{F}_{n-1}) = 0$$

*where* $\mathcal{F}_n := \mathcal{F}_n(\Upsilon_1, V_1, \dots, \Upsilon_n, V_n)$.

*(dE)*

$$\sum_{n=1}^{\infty} \tilde{a}_n^2 < \infty$$

**Corollary 6.3.3.** *Under Assumption 6.3.3 the solution process* $(\Upsilon_n)$ *of the recursive algorithm [d-Gen-Comp] given in* (6.14) *converges almost surely to* $\upsilon^*$.

*Proof.* We define $R_s := \max\limits_{n \in \mathbb{N}, n \leqslant s} (n) = \lfloor s \rfloor$, $s \geqslant 0$ and $M(\mathrm{d}s, y) := \tilde{V}_s \mathrm{d}R_s$. where

$$\tilde{V}_t := \begin{cases} V_1 & , t = 0 \\ V_n & , n-1 < t \leqslant n \ , n \in \mathbb{N} \end{cases}$$

is a time-continuous extension of $V_n$. We write

$$\int_0^t M(\mathrm{d}s, y) = \int_0^t \tilde{V}_s \mathrm{d}R_s = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} \tilde{V}_n (\Delta R_n) = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} \tilde{V}_n = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} V_n =: H_t.$$

With $\tilde{\mathcal{F}}_t := \mathcal{F}_{R_t}$ we find

$$\begin{aligned} \mathbb{E}\left(H_t \mid \tilde{\mathcal{F}}_s\right) &= \mathbb{E}\left(H_t \mid \mathcal{F}_{\lfloor s \rfloor}\right) = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} \mathbb{E}\left(V_n \mid \mathcal{F}_{\lfloor s \rfloor}\right) \\ &= \sum_{\substack{n \leqslant \lfloor s \rfloor \\ n \in \mathbb{N}}} \mathbb{E}\left(V_n \mid \mathcal{F}_{\lfloor s \rfloor}\right) + \sum_{\substack{\lfloor s \rfloor < n \leqslant t \\ n \in \mathbb{N}}} \mathbb{E}\left(V_n \mid \mathcal{F}_{\lfloor s \rfloor}\right) \\ &= \sum_{\substack{n \leqslant s \\ n \in \mathbb{N}}} V_n + 0 = \sum_{\substack{n \leqslant s \\ n \in \mathbb{N}}} V_n = H_s. \end{aligned}$$

Consequently $(H_t)$ is a martingale with respect to $\tilde{\mathcal{F}}_t := \mathcal{F}_{R_t}$, $t \geqslant 0$.

Now we are prepared to derive recursion [d-Gen-Comp] given in (6.14) from the stochastic integral equation [Gen-Comp] (5.4). We find

$$\begin{aligned}\Upsilon_n - \Upsilon_0 &= \int_0^n \tilde{a}_s \left(G_s - \Upsilon_{s-}\right) \mathrm{d}R_s + \int_0^n k_s M(\mathrm{d}s, \Upsilon_{s-}) \\ &= \sum_{j=1}^n \tilde{a}_j \left(G_j - \Upsilon_{j-1}\right)(\Delta R_j) + \sum_{j=1}^n k_j V_j \left(\Delta R_j\right) \\ &= \sum_{j=1}^n \tilde{a}_j \left(G_j - \Upsilon_{j-1}\right) + \sum_{j=1}^n k_j V_j.\end{aligned}$$

It suffices to check conditions $(B)$, $(C)$, $(D)$ and $(E)$ from Assumption 6.1.1. In order to show $(D)$ we write

$$[\int_0^\cdot M(\mathrm{d}s, y)]_t = [\int_0^\cdot \tilde{V}_s \mathrm{d}R_s]_t = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} \mathbb{E}\left(V_n^2 (\Delta R_s)^2 \mid \mathcal{F}_{n-1}\right) = \sum_{\substack{n \leqslant t \\ n \in \mathbb{N}}} \mathbb{E}\left(V_n^2 \mid \mathcal{F}_{n-1}\right).$$

The monotone convergence theorem and Hölder's inequality yield

$$\int_0^t k_s^2 \frac{h_s(\Upsilon_{s-})}{1+\Upsilon_{s-}^2} \mathrm{d}R_s \leqslant \mathbb{E} \sum_{n \in \mathbb{N}} k_n^2 \mathbb{E}\left(V_n^2 \mid \mathcal{F}_{n-1}\right) = \sum_{n \in \mathbb{N}} k_n^2 \mathbb{E} V_n^2 \leqslant \left(\sup_{n \in \mathbb{N}} \mathbb{E} V_n^2\right) \sum_{n \in \mathbb{N}} k_n^2 < \infty.$$

Vality of Assumption $(B)$ and $(C)$ follows from

$$\int_0^\infty \tilde{a}_s G_s \mathrm{d}R_s = \sum_{j=1}^\infty \tilde{a}_j G_j (\Delta R_j) = \sum_{j=1}^\infty \tilde{a}_j G_j < \infty.$$

Condition $(E)$ follows obviously from $(dE)$ by

$$\int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d = \sum_{n=1}^\infty \tilde{a}_n^2 < \infty.$$

Consequently all conditions of Theorem 6.1.1 are verified. □

We also investigate time-discrete special cases of (6.2) and (6.3):

$$Z_n - Z_{n-1} = -a_n\{f(Z_{n-1}) + V_n\} \tag{6.15}$$

$$Z_n - Z_{n-1} = -\frac{a_n}{2c_n}\{f(Z_{n-1} + c_n e_i) - f(Z_{n-1} - c_n e_i) + V_n^i\}_{i \in \{1,\dots,d\}} \tag{6.16}$$

which we denote by [d-RM] and [d-KW], respectively.

There are obvious time-discrete variants of [RM-J], [KW-H], [KW-F-2] and [KW-F-1] which we denote as [d-RM-J], [d-KW-H], [d-KW-F-2] and [d-KW-F-1]. For these special cases we formulate the following assumption.

**Assumption 6.3.4.**
*Let Assumption 6.3.3 hold with (dA) and (dC) replaced by the following conditions.*

*(dAsp)* *Let*

$f\colon \mathbb{R}^d \to \mathbb{R}^d$ *with* $J$ *be Lipschitz continuous in [d-RM-J],*
$f\colon \mathbb{R}^d \to \mathbb{R}$ *with* $H$ *be Lipschitz continuous in [d-KW-H],*
$f\colon \mathbb{R}^d \to \mathbb{R}$ *with* $\nabla f$ *be Lipschitz continuous in [d-KW-F-2] and [d-KW-F-1].*

*(dCsp)* *There exists an adapted process* $(r_n)_{n\in\mathbb{N}}$ *with* $r_n \to 0$ *for* $n \to \infty$ *such that* $\|Z_n - z^*\| = \mathcal{O}(r_n)$ $\mathbb{P}$*-a.s. as well as*

$$\begin{array}{llll} \sum_{n=1}^{\infty} \tilde{a}_n r_n < \infty & \text{and} & \sum_{n=1}^{\infty} \tilde{a}_n c_n < \infty & \text{for [d-RM-J],} \\ \sum_{n=1}^{\infty} \tilde{a}_n r_n < \infty & \text{and} & \sum_{n=1}^{\infty} \tilde{a}_n c_n < \infty & \text{for [d-KW-H],} \\ \sum_{n=1}^{\infty} \tilde{a}_n r_n^2 < \infty & \text{and} & \sum_{n=1}^{\infty} \tilde{a}_n c_n^2 < \infty & \text{for [d-KW-F-2], and} \\ \sum_{n=1}^{\infty} \tilde{a}_n r_n^2 < \infty & & & \text{for [d-KW-F-1].} \end{array}$$

*For [d-KW-H] and [d-KW-F-2],* $(c_n)$ *is the non-negative process from the leading Kiefer-Wolfowitz algorithm.*

*(dDsp)* *Moreover let (dD) hold with* $k_s$ *replaced by* $(\tilde{a}_s/c_s)$, $(\tilde{a}_s/c_s^2)$, $\tilde{a}_s$ *or* $\tilde{a}_s$, *for algorithm [d-RM-J], [d-KW-H], [d-KW-F-2] or [d-KW-F-1], respectively.*

**Corollary 6.3.4.** *Under Assumption 6.3.4 the iterates of the recursive algorithms [d-RM-J], [d-KW-H], [d-KW-F-2] and [d-KW-F-1] converge almost surely to* $\upsilon^*$.

*Remark* 6.3.1. Almost sure convergence of [d-KW-F-2] and [d-KW-F-1] has already been shown in [28] under the assumption of a three times differentiable $f$ at $z^*$. Here $f$ is assumed to have a Lipschitz continuous gradient at $z^*$ only. However it is fair to say that a rate $r_t$ for leading [KW] can only be achieved if $f$ is at least two times differentiable.

*Proof.* We follow the steps of Corollary 6.3.3. We define $R_s := \max_{n\in\mathbb{N}, n\leqslant s}(n) = \lfloor s \rfloor$, $s \geqslant 0$ and $M(\mathrm{d}s, y) := \tilde{V}_s \mathrm{d}R_s$, where

$$\tilde{V}_t := \begin{cases} V_1 & , t = 0 \\ V_n & , n-1 < t \leqslant n \ , n \in \mathbb{N} \end{cases}$$

is a time-continuous extension of $V_n$. We write

$$\int_0^t M(\mathrm{d}s, y) = \int_0^t \tilde{V}_s \mathrm{d}R_s = \sum_{\substack{n\leqslant t \\ n\in\mathbb{N}}} \tilde{V}_n(\Delta R_n) = \sum_{\substack{n\leqslant t \\ n\in\mathbb{N}}} \tilde{V}_n = \sum_{\substack{n\leqslant t \\ n\in\mathbb{N}}} V_n =: H_t.$$

With $\tilde{\mathcal{F}}_t := \mathcal{F}_{R_t}$ we find

$$\begin{aligned}\mathbb{E}\left(H_t \mid \tilde{\mathcal{F}}_s\right) &= \mathbb{E}\left(H_t \mid \mathcal{F}_{\lfloor s\rfloor}\right) = \sum_{\substack{n\leqslant t\\ n\in\mathbb{N}}} \mathbb{E}\left(V_n \mid \mathcal{F}_{\lfloor s\rfloor}\right)\\ &= \sum_{\substack{n\leqslant \lfloor s\rfloor\\ n\in\mathbb{N}}} \mathbb{E}\left(V_n \mid \mathcal{F}_{\lfloor s\rfloor}\right) + \sum_{\substack{\lfloor s\rfloor< n\leqslant t\\ n\in\mathbb{N}}} \mathbb{E}\left(V_n \mid \mathcal{F}_{\lfloor s\rfloor}\right)\\ &= \sum_{\substack{n\leqslant s\\ n\in\mathbb{N}}} V_n + 0 = \sum_{\substack{n\leqslant s\\ n\in\mathbb{N}}} V_n = H_s.\end{aligned}$$

Consequently $(H_t)$ is a martingale with respect to $\tilde{\mathcal{F}}_t := \mathcal{F}_{R_t}$, $t \geqslant 0$.

Now we derive recursion [d-KW-F-1] from the stochastic integral equation [KW-F-1] (6.7):

$$\begin{aligned}\Upsilon_n - \Upsilon_0 &= \int_0^n \tilde{a}_s \left(f(Z_{s-}) - \Upsilon_{s-}\right) \mathrm{d}R_s + \int_0^n k_s M(\mathrm{d}s, \Upsilon_{s-})\\ &= \sum_{j=1}^n \tilde{a}_j \left(f(Z_{j-1}) - \Upsilon_{j-1}\right)(\Delta R_j) + \sum_{j=1}^n k_j V_j\,(\Delta R_j)\\ &= \sum_{j=1}^n \tilde{a}_j \left(f(Z_{j-1}) - \Upsilon_{j-1}\right) + \sum_{j=1}^n k_j V_j.\end{aligned}$$

Algorithms [d-RM-J], [d-KW-H] and [d-KW-F-2] follow analogously. Clearly $(dAsp)$ implies $(Asp)$. Conditions $(B)$, $(Dsp)$ and $(E)$ follow from $(dB)$, $(dDsp)$ and $(dE)$ in the same way as shown in the proof of Corollary 6.3.3. Vality of Condition $(Csp)$ follows from

$$\int_0^\infty \tilde{a}_s(r_s^2 + c_s^2)\mathrm{d}R_s = \sum_{j=1}^\infty \tilde{a}_j(r_j^2 + c_j^2)(\Delta R_j) = \sum_{j=1}^\infty \tilde{a}_j(r_j^2 + c_j^2) < \infty,$$

$$\int_0^\infty \tilde{a}_s r_s^2\mathrm{d}R_s = \sum_{j=1}^\infty \tilde{a}_j r_j^2(\Delta R_j) = \sum_{j=1}^\infty \tilde{a}_j r_j^2 < \infty,$$

$$\int_0^\infty \tilde{a}_s(c_s + r_s)\mathrm{d}R_s = \sum_{j=1}^\infty \tilde{a}_j(c_j + r_j)(\Delta R_j) = \sum_{j=1}^\infty \tilde{a}_j(c_j + r_j) < \infty,$$

and

$$\int_0^\infty \tilde{a}_s(c_s + r_s)\mathrm{d}R_s = \sum_{j=1}^\infty \tilde{a}_j(c_j + r_j)(\Delta R_j) = \sum_{j=1}^\infty \tilde{a}_j(c_j + r_j) < \infty,$$

respective to [d-KW-F-1], [d-KW-F-2], [d-RM-J] and [d-KW-H]. Consequently all conditions of Theorem 6.2.1 are verified and the corollary is proven. □

## 6.4 Simulations

The following plots show the leading and companion algorithm together in one figure. Although a random initial value would work as well, we manually choose fixed starting values for the processes to keep the plots clear. All companion processes are chosen to have an initial value $\Upsilon_0 = 0$. The one- or two-dimensional leading algorithms all start at 5 or $(5, 5)$, respectively. In all simulations we set $a_n = n^{-1}$ and $\tilde{a}_n = n^{-1}$. As observation noise standard normal distributed random variables are chosen.

We begin in Figure 6.1 with the investigation of the companion algorithm [RM-J] which estimates the first derivative at the root of a function, which in turn is estimated via the Robbin-Monro procedure.

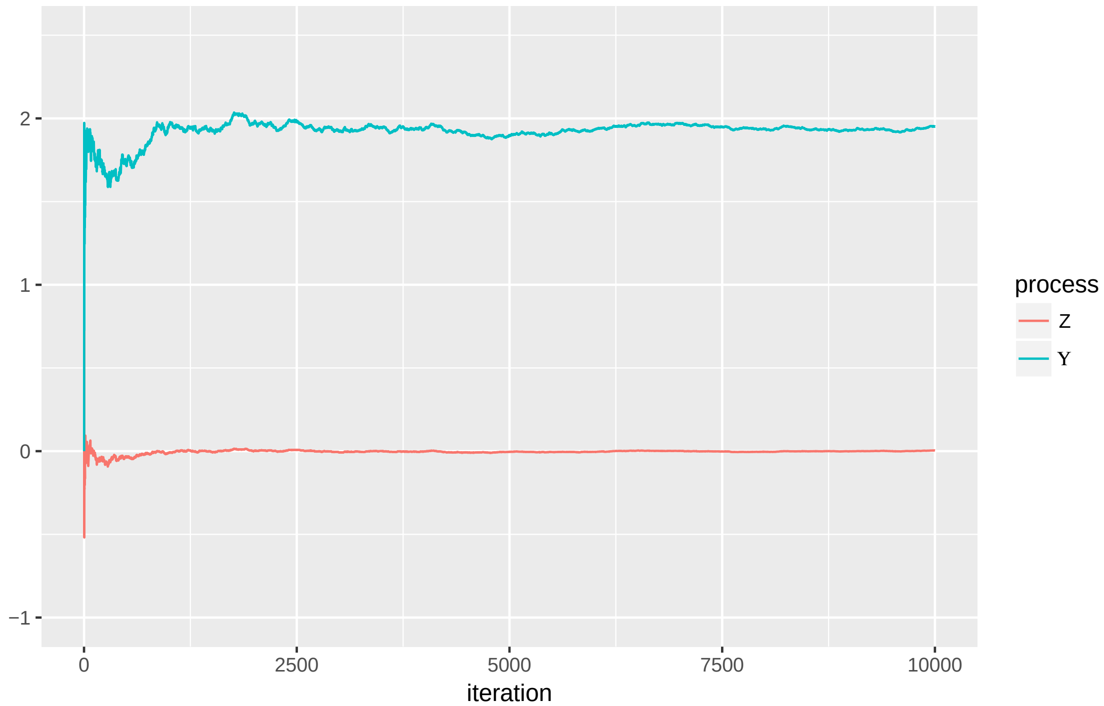


**Figure 6.1.** Paths of Robbins-Monro process $Z$ and companion [RM-J] process $\Upsilon$ related to the function $z \mapsto z + \sin(z)$ with $c_n = 2n^{-1/4}$

In Figure 6.2 the paths of Kiefer-Wolfowitz and [KW-H] are shown. The latter estimates the second derivative at the minimum.

Finally we focus on [KW-F-2] (Figure 6.3) and [KW-F-1] (Figure 6.4) which both estimate the function value at the point of the location of the minimum of $\mathbb{R}^2 \to \mathbb{R}$: $(z_1, z_2)^T \mapsto z_1^2 + z_2^2 + 1$ which in turn is estimated by a leading Kiefer-Wolfowitz algorithm.

It is remarkable that the [KW-F-1] and the [KW-F-2] algorithms are very robust against the observation noise. There is hardly any difference between both companion paths. Another detail to notice is that the paths of [RM-J] and [KW-H] are approaching their respective point of interest very early in contrast to [KW-F-1] or [KW-F-2]. This can be explained easily as both, the first derivative of $z \mapsto z + \sin(z)$ for [RM-J] as well as the second derivative of $z \mapsto z^2 + \cos(z)$ for [KW-H] are bounded for any $z$.

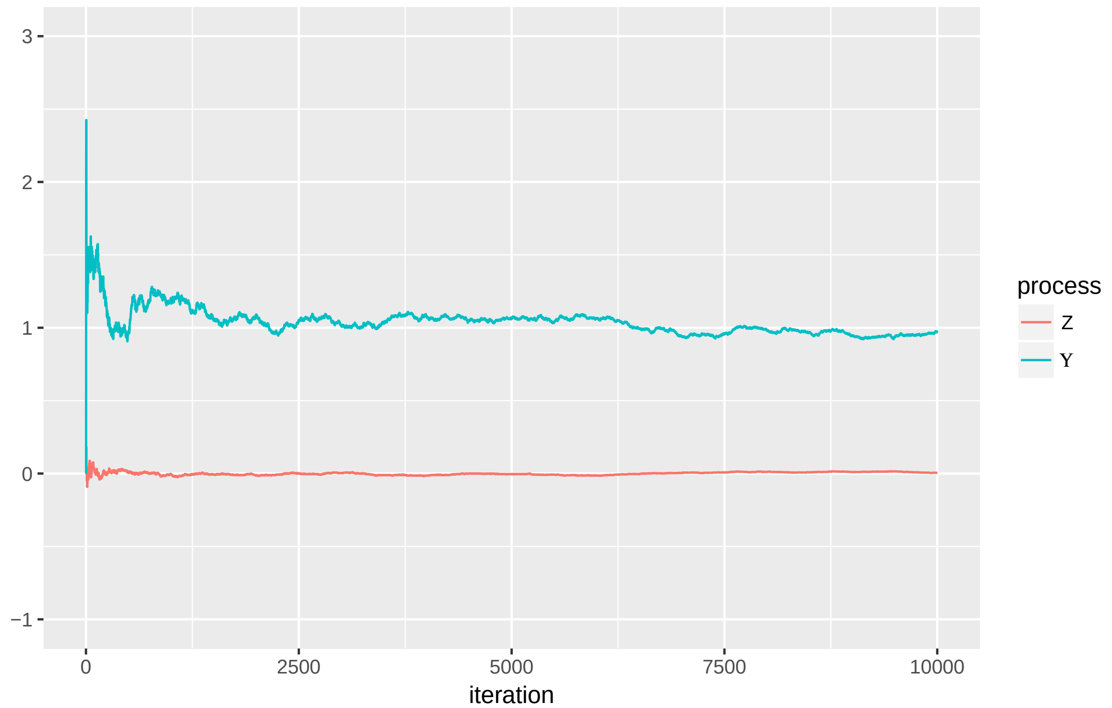


**Figure 6.2.** Paths of Kiefer-Wolfowitz process $Z$ and companion [KW-H] process $\Upsilon$ related to the function $z \mapsto z^2 + \cos(z)$ with $c_n = 2n^{-1/6}$

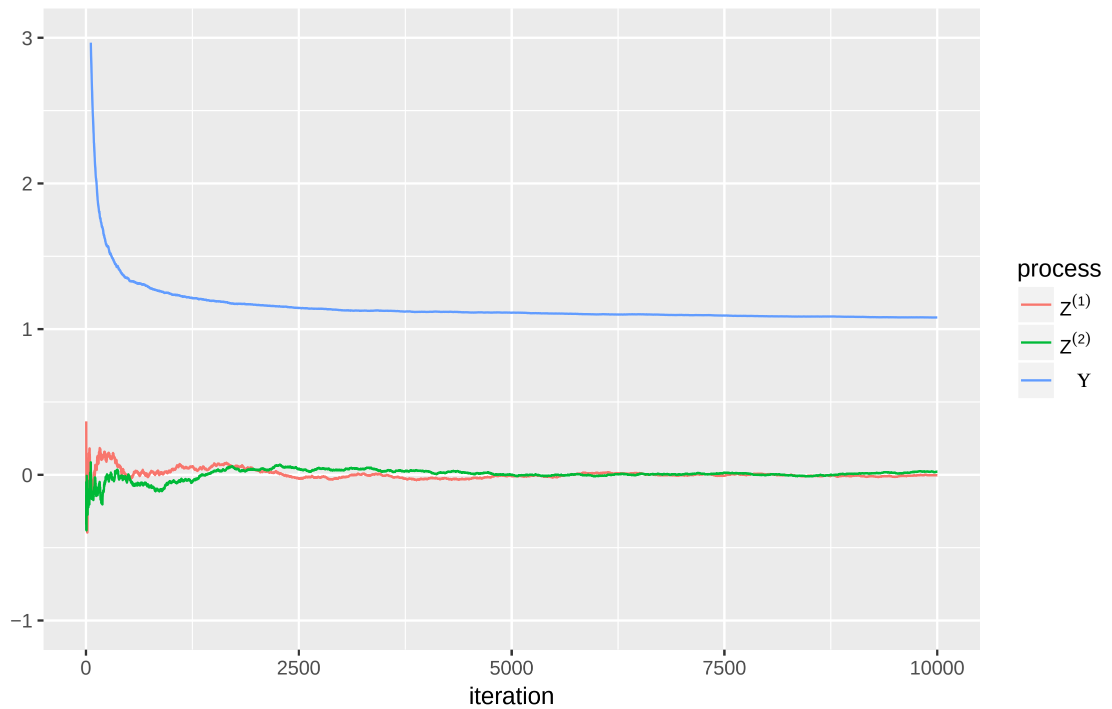


**Figure 6.3.** Paths of Kiefer-Wolfowitz process $Z = (Z^{(1)}, Z^{(2)})^T$ and companion [KW-F-2] process $\Upsilon$ related to the function $(z_1, z_2)^T \mapsto z_1^2 + z_2^2 + 1$ with $c_n = n^{-1/6}$

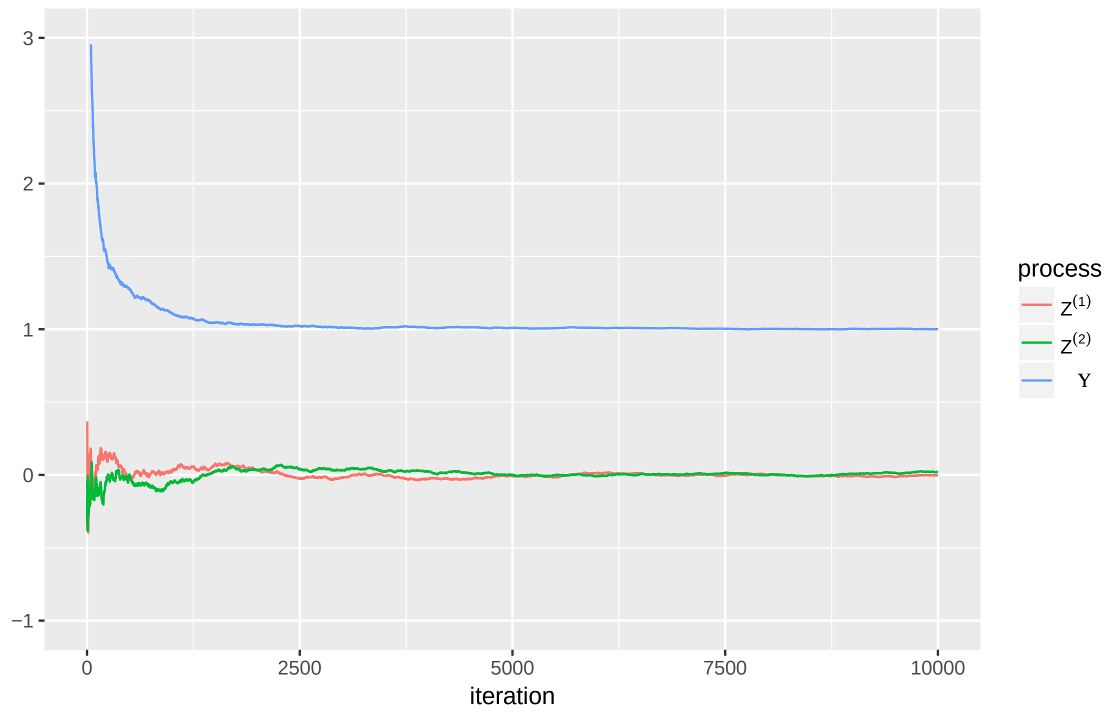


**Figure 6.4.** Paths of Kiefer-Wolfowitz process $Z = (Z^{(1)}, Z^{(2)})^T$ and companion [KW-F-1] process $\Upsilon$ related to the function $(z_1, z_2)^T \mapsto z_1^2 + z_2^2 + 1$ with $c_n = n^{-1/6}$

This comes from the assumption, that the function in [RM] and the first derivative in [KW] must be Lipschitz. The function value in [KW-F-1] or [KW-F-2] however is highly dependent on the value $z$ where it is located. One way to handle this disadvantage could be to start the companion algorithm not at the same time as its leading algorithm. Then it would not be misled by poorly chosen initial values. Moreover it is observable that [KW-F-1] decreases a little bit faster than [KW-F-2]. This is due to the fact that the latter averages four function evaluations per iteration step. Hence the effect of negative noise, which pushes the curve down, is very unlikely. The relatively smooth paths of [KW-F-1] or [KW-F-2] can be explained by a higher almost sure rate of convergence which is almost $n^{-1/2}$ and $n^{-1/3}$, respectively, whereas [RM-J] and [KW-H] have a rate close to $n^{-1/4}$ and $n^{-1/6}$, respectively. A detailed derivation of almost sure convergence rates of companion algorithms is given in Chapter 7. Especially almost surely convergence rates of companion processes with parameters chosen in the same way as for the simulated paths presented above are presented in Section 7.4.1. In Chapter 8 asymptotic normality of the companion processes is investigated. Under the settings of current simulations, Section 8.5 yields that [KW-F-1] and [KW-F-2] converge with rate $n^{-1/2}$ and $n^{-1/3}$, respectively. Moreover they have asymptotic bias 0 and 3/2, respectively. [RM-J] converges unbiasedly with rate $n^{-1/4}$. Finally [KW-H] converges with rate $n^{-1/6}$ and asymptotic bias 0.

# 7 Almost Sure Convergence Rate of Companion Algorithms

Once consistency is ensured, the question arises how fast the process $(\Upsilon_t)_{t\geqslant 0}$ converges. Later, in order to establish asymptotic normality, we need $\Upsilon$ to converge at an a.s. rate.

## 7.1 Semimartingale Companion Algorithms

For fixed $\delta \geqslant 0$, we define $\gamma_t(\delta) := \mathcal{E}_t(\delta \int_0^{\cdot} \tilde{a}_s \mathrm{d}R_s)$, $t \geqslant 0$, where $\mathcal{E}_t(.)$ is the stochastic exponential, and investigate the set of $\delta$ such that

$$\gamma_t(\delta)\|\Upsilon_t - \upsilon^*\| \to 0 \text{ a.s.}$$

Note that the stochastic exponential of a semimartingale $X$ is the solution of $\Upsilon_t = 1 + \int_0^t \Upsilon_{s-} \mathrm{d}X_s$, $X_0 = 0$, which is given by $\mathcal{E}_t(X) := \exp\left(X_t - \frac{1}{2}[X,X]_t\right) \prod_{0<s\leqslant t}(1 + \Delta X_s) \exp\left(-\Delta X_s + \frac{1}{2}(\Delta X_s)^2\right)$, c.f. Protter [32].

## 7.2 Almost Sure Convergence Rate of the General Algorithms

We consider the following conditions.

**Assumption 7.2.1.** *In addition to (A), (B) and (E) from Assumption 6.1.1 let the following conditions hold true.*

$(\tilde{C})$

$$\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s|G_s - \upsilon^*|\mathrm{d}R_s < \infty$$

$(\tilde{D})$ *For all* $y \in \mathbb{R}$

$$\int_0^\infty k_s^2 \frac{\gamma_{s-}^2(\delta) h_s(\Upsilon_{s-})}{1 + \gamma_{s-}^2(\delta)\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \text{ where } h_s(y) := \frac{\mathrm{d}[\int_0^{\cdot} M(\mathrm{d}s, y)]_s}{\mathrm{d}R_s}.$$

**Theorem 7.2.1.** *Let Assumption 7.2.1 hold. Then for all $0 \leqslant \delta < 1$, the solution of the companion algorithm [Gen-Comp] given in (5.4) satisfies*

$$\gamma_t(\delta)|\Upsilon_t - \upsilon^*| \xrightarrow{t\to\infty} 0 \ \mathbb{P}\text{-}a.s.$$

*Proof.* The proof is similar to that of Theorem 6.1.1. Let $\upsilon^* = 0$. We investigate $(\gamma_t^2(\delta)\Upsilon_t^2)_{t\geqslant 0}$ instead of $(\Upsilon_t^2)_{t\geqslant 0}$. We apply the Robbins-Siegmund lemma (Lemma A.1.1 in the appendix) to a decomposition of $(\gamma_t^2(\delta)\Upsilon_t^2)_{t\geqslant 0}$. Integration by parts yields

$$\begin{aligned}
\gamma_t^2(\delta) = \gamma_t(\delta)\gamma_t(\delta) &= \mathcal{E}_t\left(2\delta\int_0^\cdot \tilde{a}_s \mathrm{d}R_s + \delta^2\int_0^\cdot \tilde{a}_s^2 \mathrm{d}[R_\cdot, R_\cdot]_s\right)\\
&= \mathcal{E}_t\left(2\delta\int_0^\cdot \tilde{a}_s \mathrm{d}R_s + \delta^2\int_0^\cdot \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d\right)
\end{aligned}$$

as well as

$$\mathrm{d}\gamma_s^2(\delta) = \gamma_{s-}^2(\delta)\left(2\delta a_s \mathrm{d}R_s + \delta^2 a_s^2 \Delta R_s \mathrm{d}R_s^d\right).$$

Integration by parts, Lemma A.1.6 and (6.1) yield

$$\begin{aligned}
\gamma_t^2(\delta)\Upsilon_t^2 - \gamma_0^2(\delta)\Upsilon_0^2 &= \int_0^t \gamma_{s-}^2(\delta)\mathrm{d}\Upsilon_s + \int_0^t \Upsilon_{s-}^2 \mathrm{d}\gamma_s^2(\delta) + \int_0^t \mathrm{d}[\gamma_\cdot^2(\delta), \Upsilon_\cdot^2]_s\\
&= \int_0^t \gamma_{s-}^2(\delta)\mathrm{d}\Upsilon_s + \int_0^t \Upsilon_{s-}^2 \mathrm{d}\gamma_s^2(\delta) + \int_0^t \Delta\gamma_s^2(\delta)\mathrm{d}\Upsilon_s^2\\
&= \int_0^t \gamma_{s-}^2(\delta)\mathrm{d}\Upsilon_s + \int_0^t \Upsilon_{s-}^2 \mathrm{d}\gamma_s^2(\delta) + \int_0^t (\gamma_s^2(\delta) - \gamma_{s-}^2(\delta))\mathrm{d}\Upsilon_s^2\\
&= \int_0^t \Upsilon_{s-}^2 \mathrm{d}\gamma_s^2(\delta) + \int_0^t \gamma_s^2(\delta)\mathrm{d}\Upsilon_s^2\\
&= 2\int_0^t \gamma_s^2(\delta)\tilde{a}_s\Upsilon_{s-}G_s\mathrm{d}R_s - 2\int_0^t \gamma_s^2(\delta)\tilde{a}_s G_s\Upsilon_{s-}\Delta R_s \mathrm{d}R_s^d\\
&\quad - 2\int_0^t \gamma_s^2(\delta)\tilde{a}_s\Upsilon_{s-}^2\mathrm{d}R_s + \int_0^t \gamma_s^2(\delta)\tilde{a}_s^2 G_s^2\Delta R_s \mathrm{d}R_s^d\\
&\quad + \int_0^t \gamma_s^2(\delta)\tilde{a}_s^2\Upsilon_{s-}^2\Delta R_s\mathrm{d}R_s^d + \int_0^t \gamma_s^2(\delta)k_s^2[M(\mathrm{d}\tau, \Upsilon_{\tau-})]_s\\
&\quad + 2\delta\int_0^t \tilde{a}_s\gamma_{s-}^2(\delta)\Upsilon_{s-}^2\mathrm{d}R_s + \delta^2\int_0^t \gamma_{s-}^2(\delta)\Upsilon_{s-}^2\tilde{a}_s^2\Delta R_s\mathrm{d}R_s^d\\
&\quad + \int_0^t \mathrm{d}\tilde{M}_s
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{d}\tilde{M}_s := &+2\gamma_s^2(\delta)k_s\Upsilon_{s-}M(\mathrm{d}s, \Upsilon_{s-}) + 2\gamma_s^2(\delta)\tilde{a}_s k_s G_s\Delta R_s M(\mathrm{d}s, \Upsilon_{s-})\\
&- 2\gamma_s^2(\delta)\tilde{a}_s k_s\Upsilon_{s-}\Delta R_s M(\mathrm{d}s, \Upsilon_{s-})
\end{aligned}$$

$$+\gamma_s^2(\delta)k_s^2\left([\int_0^{\cdot} M(\mathrm{d}\tau,\Upsilon_{\tau-})]_s - \lceil\int_0^{\cdot} M(\mathrm{d}\tau,\Upsilon_{\tau-})\rceil_s\right).$$

The first, second and third term in the definition of $\widetilde{M}_t$ are in $\mathcal{M}_{\mathrm{loc}}$ as the integrands are predictable and the integrators are local martingales. By definition of the compensator, the fourth term in the definition of $\widetilde{M}_t$ is in $\mathcal{M}_{\mathrm{loc}}$. In order to apply the Robbins-Siegmund lemma (Lemma A.1.1) we bound $\gamma_t^2(\delta)\Upsilon_t^2 \leqslant A_t^1 - A_t^2 + \tilde{M}_t$ with

$$\begin{aligned} A_t^1 := {} & 2\int_0^t \gamma_s^2(\delta)\tilde{a}_s|\Upsilon_{s-}G_s|\mathrm{d}R_s + 2\int_0^t \gamma_s^2(\delta)\tilde{a}_s|G_s\Upsilon_{s-}|\Delta R_s \mathrm{d}R_s^d \\ & + \int_0^t \gamma_s^2(\delta)\tilde{a}_s^2 G_s^2 \Delta R_s \mathrm{d}R_s^d + \int_0^t \gamma_s^2(\delta)\tilde{a}_s^2\Upsilon_{s-}^2\Delta R_s\mathrm{d}R_s^d \\ & + \int_0^t \gamma_s^2(\delta)k_s^2 \mathrm{d}\lceil M(\mathrm{d}s,\Upsilon_{s-})\rceil_s + \delta^2\int_0^t \gamma_{s-}^2(\delta)\Upsilon_{s-}^2\tilde{a}_s^2\Delta R_s \mathrm{d}R_s^d \qquad (7.1) \\ -A_t^2 := {} & (2-2\delta)\int_0^t \tilde{a}_s\gamma_{s-}^2(\delta)\Upsilon_{s-}^2\mathrm{d}R_s. \qquad (7.2) \end{aligned}$$

Now we assess $\int_0^t \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\mathrm{d}A_t^1$. A quick calculation (c.f. (3.4)) yields

$$\gamma_t(\delta) = \gamma_{t-}(\delta)(1+\delta a_t \Delta R_t)$$

which is a useful representation of $\gamma_t(\delta)$ to investigate

$$\int_0^\infty \frac{1}{1+\gamma_{s-}^2\Upsilon_{s-}^2}\mathrm{d}A_s^1. \qquad (7.3)$$

This, together with the assumptions $\int_0^\infty \tilde{a}_s^2\Delta R_s \mathrm{d}R_s^d < \infty$ and $\int_0^\infty \tilde{a}_s \mathrm{d}R_s = \infty$, stated in $(B)$ and $(E)$, respectively, implies

$$\frac{\gamma_t(\delta)}{\gamma_{t-}(\delta)} = (1+\delta\tilde{a}_t\Delta R_t) = (1+o_{\mathrm{b}}(1)) \leqslant \mathcal{C}(\omega).$$

Now we replace the integrand in (7.3) by (7.1). The first term in that substitution can be bounded by

$$\begin{aligned} & \int_0^\infty \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_s^2(\delta)\tilde{a}_s|\Upsilon_{s-}G_s|\mathrm{d}R_s \\ & \qquad \leqslant \int_0^\infty \left(\frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)}\right)^2 \frac{\gamma_{s-}(\delta)|\Upsilon_{s-}|}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_{s-}(\delta)\tilde{a}_s|G_s|\mathrm{d}R_s \\ & \qquad \leqslant \mathcal{C}(\omega)\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s|G_s|\mathrm{d}R_s < \infty \end{aligned}$$

where the last inequality holds by condition $(\tilde{C})$. In the same way we can bound the purely discontinuous term

$$\begin{aligned}
&\int_0^\infty \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_s^2(\delta)\tilde{a}_s|G_s\Upsilon_{s-}|\Delta R_s \mathrm{d}R_s^d \\
&\qquad\leqslant \int_0^\infty \left(\frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)}\right)^2 \frac{\gamma_{s-}(\delta)|\Upsilon_{s-}|}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_{s-}(\delta)\tilde{a}_s|G_s|\Delta R_s \mathrm{d}R_s^d \\
&\qquad\leqslant \mathcal{C}(\omega)\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s|G_s|\mathrm{d}R_s < \infty.
\end{aligned}$$

The third term is handled with condition $(E)$ as follows:

$$\begin{aligned}
&\int_0^\infty \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_s^2(\delta)\tilde{a}_s^2 G_s^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad= \int_0^\infty \left(\frac{\gamma_s(\delta)}{\gamma_{s-}(\delta)}\right)^2 \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_{s-}^2(\delta)\tilde{a}_s^2 G_s^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad\leqslant \mathcal{C}(\omega)\int_0^\infty \gamma_{s-}^2(\delta)\tilde{a}_s^2 G_s^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad\leqslant \mathcal{C}(\omega)\int_0^\infty \tilde{a}_s^2\Delta R_s \mathrm{d}R_s^d \\
&\qquad< \infty.
\end{aligned}$$

The second to last inequality holds for the following reason. By assumption $(B)$

$$\int_0^\infty \tilde{a}_s \mathrm{d}R_s = \infty$$

holds true. Additionally by condition $(\tilde{C})$,

$$\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s|G_s|\mathrm{d}R_s < \infty.$$

This implies $\gamma_{s-}(\delta)|G_s| \to 0$ and hence $\gamma_{s-}^2(\delta)G_s^2 \to 0$. Concerning the fourth term in the expansion of (7.3) we once more apply $(E)$ to get

$$\int_0^\infty \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_s^2(\delta)\tilde{a}_s^2\Upsilon_{s-}^2\Delta R_s \mathrm{d}R_s^d \leqslant \mathcal{C}(\omega)\int_0^\infty \tilde{a}_s^2\Delta R_s \mathrm{d}R_s^d < \infty.$$

This bound is used to handle the sixth term as well. Finally the fifth term can be handled with $(\tilde{D})$ as follows:

$$\begin{aligned}
&\int_0^\infty \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_s^2(\delta)k_s^2 \mathrm{d}[M(\mathrm{d}s,\Upsilon_{s-})]_s \\
&\qquad\leqslant \mathcal{C}(\omega)\int_0^\infty \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\gamma_{s-}^2(\delta)k_s^2 h_s(\Upsilon_{s-})\mathrm{d}R_s < \infty.
\end{aligned}$$

As a result we conclude that $\displaystyle\int_0^\infty \frac{1}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}\mathrm{d}A_s^1 < \infty$. Therefore, according to

Lemma A.1.1, $\gamma_t^2(\delta)\Upsilon_t^2$ converges. Furthermore the same lemma yields that $A_\infty^2 < \infty$.

The convergence of $(\gamma_t(\delta)\Upsilon_t)_{t\geqslant 0}$ to 0 is shown by contradiction. For that purpose assume a set $N$ of non-zero probability on which the solution of the stochastic integral equation does not converge to zero. We will deduce a contradiction to

$$\Omega = \{A_\infty^2 < \infty\}.$$

As proven before, $(\gamma_t(\delta)\Upsilon_t)_{t\geqslant 0}$ converges for almost all $\omega \in \Omega$, but by assumption for all $\omega \in N$ the process does not converge to 0. Hence it follows for all $\omega \in N$

$$\underset{\epsilon^*>0}{\exists} \quad \underset{s_0}{\exists} \quad \underset{s\geqslant s_0}{\forall} \quad \epsilon^* \leqslant \Upsilon_s^2 \leqslant 1/\epsilon^*.$$

As $0 \leqslant \delta < 1$ holds, $A_\infty^2$, namely $(2-2\delta)\int_0^\infty \tilde{a}_s\gamma_{s-}^2(\delta)\Upsilon_{s-}^2 \mathrm{d}R_s$, is non-negative. Consequently, with condition (B),

$$\begin{aligned}
A_\infty^2 &= (2-2\delta)\int_0^\infty \gamma_{s-}^2(\delta)\tilde{a}_s\Upsilon_{s-}^2 \mathrm{d}R_s \\
&\geqslant \mathcal{C}\int_0^\infty \gamma_{s-}^2(\delta)\tilde{a}_s\Upsilon_{s-}^2 \mathrm{d}R_s = \mathcal{C}\int_0^{s_0} \gamma_{s-}^2(\delta)\tilde{a}_s\Upsilon_{s-}^2 \mathrm{d}R_s + \mathcal{C}\int_{s_0+}^\infty \gamma_{s-}^2(\delta)\tilde{a}_s\Upsilon_{s-}^2 \mathrm{d}R_s \\
&\geqslant \mathcal{C} + \epsilon^* \int_{s_0+}^\infty \tilde{a}_s \mathrm{d}R_s = \infty.
\end{aligned}$$

This is a contradiction to what we have shown before. Consequently the set $N$ cannot exist. We conclude $\gamma_t^2(\delta)\Upsilon_t^2 \to 0$ and thus $\gamma_t(\delta)\Upsilon_t \to 0$ a.s. as $t \to \infty$. □

## 7.3 Almost Sure Convergence Rate of Special Algorithms

In order to examine algorithms [RM-J], [KW-H], [KW-F-2] and [KW-F-1] given in (6.4)–(6.7), it makes sense to replace assumption $(\tilde{C})$ by the following one.

**Assumption 7.3.1.** *Let Assumption 6.2.1 hold.*

$(\tilde{C}sp)$ *Dependent on the leading algorithm replace condition (Csp) by*

$$\begin{aligned}
&\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s(c_s + r_s)\mathrm{d}R_s < \infty && \textit{for [RM-J]} \text{ (6.4)},\\
&\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s(c_s + r_s)\mathrm{d}R_s < \infty && \textit{for [KW-H]} \text{ (6.5)},\\
&\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s(r_s^2 + c_s^2)\mathrm{d}R_s < \infty && \textit{for [KW-F-2]} \text{ (6.6)},\\
&\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s r_s^2 \mathrm{d}R_s < \infty && \textit{for [KW-F-1]} \text{ (6.7)}.
\end{aligned}$$

*Moreover assume*

$(\tilde{D}sp)$ *For all $y \in \mathbb{R}$*

$$\int_0^\infty \left(\frac{\tilde{a}_s}{c_s}\right)^2 \frac{\gamma_{s-}^2(\delta)h_s(\Upsilon_{s-})}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \qquad \textit{for [RM-J]}\ (6.4),$$
$$\int_0^\infty \left(\frac{\tilde{a}_s}{c_s^2}\right)^2 \frac{\gamma_{s-}^2(\delta)h_s(\Upsilon_{s-})}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \qquad \textit{for [KW-H]}\ (6.5),$$
$$\int_0^\infty \tilde{a}_s^2 \frac{\gamma_{s-}^2(\delta)h_s(\Upsilon_{s-})}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \qquad \textit{for [KW-F-2]}\ (6.6),$$
$$\int_0^\infty \tilde{a}_s^2 \frac{\gamma_{s-}^2(\delta)h_s(\Upsilon_{s-})}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2} \mathrm{d}R_s < \infty \qquad \textit{for [KW-F-1]}\ (6.7),$$

*where* $h_s(y) := \dfrac{\mathrm{d}[\int_0^\cdot M(\mathrm{d}s, y)]_s}{\mathrm{d}R_s}$ .

**Theorem 7.3.1.** *Let Assumption 7.3.1 hold. Then for the solutions of the companion algorithms [RM-J]* (6.4)*, [KW-H]* (6.5)*, [KW-F-2]* (6.6) *and [KW-F-1]* (6.7)

$$\underset{0\leqslant\delta<1}{\forall} \quad \gamma_t(\delta)|\Upsilon_t - \upsilon^*| \xrightarrow{t\to\infty} 0 \ \textit{a.s.}$$

*Proof.* From the proof of almost sure convergence, we already know how $(Asp)$ is employed to show

$$G_s = \begin{cases} \mathcal{O}(c_s) + \mathcal{O}(r_s) & \text{for [RM-J] (6.4)} \\ \mathcal{O}(c_s) + \mathcal{O}(r_s) & \text{for [KW-H] (6.5)} \\ \mathcal{O}(r_s^2) + \mathcal{O}(c_s^2) & \text{for [KW-F-2] (6.6)} \\ \mathcal{O}(r_s^2) & \text{for [KW-F-1] (6.7).} \end{cases}$$

Consequently $(Asp)$ and $(\tilde{C}sp)$ imply $(A)$ as well as $(\tilde{C})$. Hence this theorem follows directly from Theorem 7.2.1. □

## 7.4 Itô Type and Recursive Stochastic Approximation Algorithms

In this section almost sure convergence rates for Itô type and recursive stochastic approximation algorithms are explored. Like in Chapter 6 generic as well as special types of companion algorithms are investigated. Before we go into the details, we first have to understand the rates of the underlying algorithms. This is the purpose of the following subsection.

### 7.4.1 Rates of the Underlying Algorithms

It is shown by Lazrieva et al. [22] that if $a_s = a(1+R_s)^{-1}$ amongst other conditions in the Robbins-Monro algorithm (6.2) then

$$\underset{\rho\in[0,\frac{1}{2})}{\forall} \quad (1+R_t)^\rho \|Z_t - z^*\| \to 0 \text{ a.s.}$$

Moreover, setting $a_s = a(1+R_s)^{-1}$ and $c_s = c(1+R_s)^{-\gamma}$ as well as assuming further conditions in the the Kiefer-Wolfowitz algorithm (6.3), Schnizler [37] shows

$$\underset{\rho\in[0,\tilde{\rho})}{\forall} (1+R_t)^{\rho}\|Z_t - z^*\| \to 0 \text{ a.s.},$$

where

$$\tilde{\rho} := \min\left\{\gamma(p-1), \frac{1}{2} - \gamma\right\},$$

if $f$ is $p$-times continuously differentiable at $z^*$ with $p \in \{2,3\}$. The optimal $\tilde{\rho}$ is achieved for $\gamma = 1/(2p)$, if $f$ is $p$-times continuously differentiable at $z^*$ with $p \in \{2,3\}$. Then we obtain that $\tilde{\rho}$ is of the form

$$\tilde{\rho} = \frac{p-1}{2p}.$$

The papers of Fabian [15], Dippon and Renz [12] and Dippon [10] deal with modified Kiefer-Wolfowitz algorithms in order to achieve higher rates of convergence. But usually one requires more observations or randomization of the estimator. These algorithms are not handled here.

We point out settings of $(a_s)_{s\geqslant 0}$, $(c_s)_{s\geqslant 0}$ and $(\tilde{a}_s)_{s\geqslant 0}$ for which the companion algorithms [c-KW-F-2] or [c-KW-H] and their leading algorithm [KW] don't converge simultaneously with optimal rate. Analogously such a trade-off can be achieved for [RM-J] and its leading algorithm [RM].

## 7.4.2 Itô Type Stochastic Approximation Algorithms

Now we turn to a generic Itô type result for the almost sure convergence rate of companion algorithms.

**Corollary 7.4.1.** *Consider the Itô type companion algorithm [c-Gen-Comp]* (6.11). *Let Assumption 6.3.1 and $\sigma_s(y) \leqslant C(1+|y|)$ hold. Set $a_s = a(1+s)^{-1}$, $a > 0$, and $k_s = k(1+s)^{-\kappa}$, $k > 0$. Assume $\int_0^\infty \gamma_s(\delta)\tilde{a}_s|G_s - \upsilon^*|\mathrm{d}s < \infty$ $\mathbb{P}$-a.s. Then almost surely*

$$(1+t)^{\delta}|\Upsilon_t - \upsilon^*| \xrightarrow{t\to\infty} 0$$

*for all $\delta \in [0, \kappa - 1/2)$.*

*Proof.* The corollary is traced back to Theorem 7.2.1. As in the proof of Corollary 6.3.2 we choose $R_s := s$ and $M(\mathrm{d}s, y) = \sigma_s(y)\mathrm{d}W(s)$. Continuity of $(R_t)_{t\geqslant 0}$ implies continuity of $(\gamma_t(\delta))_{t\geqslant 0}$. Hence

$$\gamma_t(\delta) = \mathcal{E}_t\left(\delta\int_0^{\cdot} \tilde{a}_s \mathrm{d}R_s\right) = \exp\left(\delta\int_0^t (1+s)^{-1}\mathrm{d}s\right) = \exp\left(\delta\ln(1+t)\right) = (1+t)^{\delta}.$$

Condition $(\tilde{D})$ follows by

$$\begin{aligned}\int_0^\infty \frac{\gamma_{s-}^2(\delta)h_s(\Upsilon_{s-})}{1+\gamma_{s-}^2(\delta)\Upsilon_{s-}^2}k_s^2\mathrm{d}R_s &= \int_0^\infty \frac{\gamma_s^2(\delta)\sigma_s^2(\Upsilon_s)}{1+\gamma_s^2(\delta)\Upsilon_s^2}k_s^2\mathrm{d}s \leqslant \mathcal{C}\int_0^\infty \frac{\gamma_s^2(\delta)(1+|\Upsilon_s|)^2}{1+\gamma_s^2(\delta)\Upsilon_s^2}k_s^2\mathrm{d}s \\ &\leqslant \mathcal{C}(\omega)\int_0^\infty \gamma_s^2(\delta)k_s^2\mathrm{d}s = \mathcal{C}(\omega)\int_0^\infty (1+s)^{2\delta-2\kappa}\mathrm{d}s \\ &< \infty.\end{aligned}$$

Due to continuity, $(E)$ holds and condition $(\tilde{C})$ follows by

$$\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s|G_s|\mathrm{d}R_s = \int_0^\infty \gamma_s(\delta)\tilde{a}_s|G_s|\mathrm{d}s < \infty.$$

Analogously $(B)$ follows from $(cB)$. Consequently all conditions of Theorem 7.2.1 are fulfilled. $\square$

In the following corollary $(a_s)_{s\geqslant 0}$, $(c_s)_{s\geqslant 0}$ and $(\tilde{a}_s)_{s\geqslant 0}$ are chosen such that the companion algorithm converges with optimal rate $\delta \in [0, \frac{1}{2})$. This does not necessarily mean, that the leading algorithm converges optimally as well. Moreover, for algorithms [c-RM-J] and [c-KW-H] there is no possible choice of $(a_s)_{s\geqslant 0}$, $(c_s)_{s\geqslant 0}$ and $(\tilde{a}_s)_{s\geqslant 0}$ such that they converge with optimal rate $\delta \in [0, \frac{1}{2})$ and consequently these algorithms are not mentioned there, but handled in a later corollary.

**Corollary 7.4.2.** *Consider the Itô type stochastic integral equations [c-KW-F-2] and [c-KW-F-1]. Let Assumption 6.3.1 and* $\sigma_s(y) \leqslant C(1+|y|)$ *hold. Set* $a > 0$, $a_s = a(1+s)^{-1}$, *and* $\tilde{a}_s = \tilde{a}(1+s)^{-1}$ *with* $\tilde{a} > 0$. *In case of* $f$ *being* $p$*-times continuously differentiable at* $z^*$, *consider the following cases.*

| *Companion Algorithm* | $p$ | $c_s$ | $\delta$ |
|---|---|---|---|
| *[c-KW-F-2]* | 2 | $c(1+s)^{-\frac{1}{4}}$ | $[0, \frac{1}{2})$ |
| *[c-KW-F-2]* | 3 | $c(1+s)^{-\frac{1}{4}}$ | $[0, \frac{1}{2})$ |
| *[c-KW-F-1]* | 2 | $c(1+s)^{-\frac{1}{4}}$ | $[0, \frac{1}{2})$ |
| *[c-KW-F-1]* | 3 | $c(1+s)^{-\frac{1}{6}}$ | $[0, \frac{1}{2})$ |

*Then almost surely* $(1+t)^\delta|\Upsilon_t - v^*| \xrightarrow{t\to\infty} 0$.

In this corollary we had settings for which the leading algorithm [c-KW] and the companion algorithm [c-KW-F-2] don't converge simultaneously each with optimal rate. Moreover we did not have settings in which [c-RM-J] and [c-KW-H] converged at an optimal rate simultaneously with their respective leading algorithm. These cases are handled in the following. Here we choose $(a_s)_{s\geqslant 0}$, $(c_s)_{s\geqslant 0}$ and $(\tilde{a}_s)_{s\geqslant 0}$ such that the leading algorithm converges with optimal rate and give the resulting rate for the companion algorithm.

**Corollary 7.4.3.** *Consider the Itô type stochastic integral equations [c-KW-F-2], [c-RM-J] and [c-KW-H]. Assume that the leading algorithm converges with optimal rate. Let Assumption 6.3.1 and $\sigma_s(y) \leqslant C(1+|y|)$ hold. Set $a_s = a(1+s)^{-1}$, $a > 0$, and $\tilde{a}_s = \tilde{a}(1+s)^{-1}$, $\tilde{a} > 0$. In case of $f$ being p-times continuously differentiable at $z^*$, assume one of the following cases.*

| *Companion Algorithm* | $p$ | $c_s$ | $\delta$ |
|---|---|---|---|
| *[c-KW-F-2]* | 3 | $c(1+s)^{-\frac{1}{6}}$ | $[0, \frac{1}{3})$ |
| *[c-RM-J]* | 1 | $c(1+s)^{-\frac{1}{4}}$ | $[0, \frac{1}{4})$ |
| *[c-KW-H]* | 3 | $c(1+s)^{-\frac{1}{6}}$ | $[0, \frac{1}{6})$ |

*Then almost surely* $(1+t)^{\delta}|\Upsilon_t - v^*| \xrightarrow{t\to\infty} 0$.

*Proof of Corollaries 7.4.2 and 7.4.3.* The corollary is traced back to Theorem 7.3.1. As in the proof of Corollary 7.4.1 we choose $R_s := s$ and $M(\mathrm{d}s, y) = \sigma_s(y)\mathrm{d}W(s)$. In the same proof we already showed

$$\gamma_t(\delta) = (1+t)^{\delta}$$

and how to deduce $(B)$ and $(\tilde{D})$ from $(cB)$ and $(cD)$, respectively. Continuity of the paths directly yields $(E)$. The rest of the proof deals with the verification of the conditions of $(\tilde{C}sp)$.

Assume a sufficiently small $\epsilon > 0$. For the companion algorithms to estimate the minimum we have

$$\begin{aligned}\int_0^{\infty} \gamma_{s-}(\delta)\tilde{a}_s r_s^2 \mathrm{d}R_s &= \mathcal{C}\int_0^{\infty}(1+s)^{\delta-1-2\rho}\mathrm{d}s\\ &\leqslant \begin{cases}\mathcal{C}\int_0^{\infty}(1+s)^{\frac{1}{2}-1-\frac{1}{2}-\epsilon}\mathrm{d}s & \text{in [c-KW-F-2] for } f\in C^2,\ \delta\in[0,\frac{1}{2})\\ \mathcal{C}\int_0^{\infty}(1+s)^{\frac{1}{2}-1-\frac{2}{3}-\epsilon}\mathrm{d}s & \text{in [c-KW-F-2] for } f\in C^3,\ \delta\in[0,\frac{1}{2})\\ \mathcal{C}\int_0^{\infty}(1+s)^{\frac{1}{2}-1-\frac{1}{2}-\epsilon}\mathrm{d}s & \text{in [c-KW-F-1] for } f\in C^2,\ \delta\in[0,\frac{1}{2})\\ \mathcal{C}\int_0^{\infty}(1+s)^{\frac{1}{2}-1-\frac{2}{3}-\epsilon}\mathrm{d}s & \text{in [c-KW-F-1] for } f\in C^3,\ \delta\in[0,\frac{1}{2})\\ \mathcal{C}\int_0^{\infty}(1+s)^{\frac{1}{3}-1-\frac{2}{3}-\epsilon}\mathrm{d}s & \text{in [c-KW-F-2] for } f\in C^3,\ \delta\in[0,\frac{1}{3})\end{cases}\\ &< \infty\end{aligned}$$

and

$$\begin{aligned}\int_0^{\infty} \gamma_{s-}(\delta)\tilde{a}_s c_s^2 \mathrm{d}R_s &= c^2\int_0^{\infty}(1+s)^{\delta-1-2\gamma}\mathrm{d}s\\ &\leqslant \begin{cases}\mathcal{C}\int_0^{\infty}(1+s)^{\frac{1}{2}-1-\frac{1}{2}-\epsilon}\mathrm{d}s & \text{in [c-KW-F-2] for } f\in C^2,\ \delta\in[0,\frac{1}{2})\\ \mathcal{C}\int_0^{\infty}(1+s)^{\frac{1}{2}-1-\frac{1}{2}-\epsilon}\mathrm{d}s & \text{in [c-KW-F-2] for } f\in C^3,\ \delta\in[0,\frac{1}{2})\\ \mathcal{C}\int_0^{\infty}(1+s)^{\frac{1}{3}-1-\frac{2}{3}-\epsilon}\mathrm{d}s & \text{in [c-KW-F-2] for } f\in C^3,\ \delta\in[0,\frac{1}{3})\end{cases}\end{aligned}$$

$$< \infty.$$

For the companion algorithm [c-RM-J] we find the bound

$$\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s c_s \mathrm{d}R_s = \int_0^\infty (1+s)^{\delta-1-\gamma}\mathrm{d}s \leqslant \mathcal{C}\int_0^\infty (1+s)^{-1-\epsilon}\mathrm{d}s < \infty,$$

and for [c-KW-H]

$$\begin{aligned}\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s c_s \mathrm{d}R_s &= \mathcal{C}\int_0^\infty (1+s)^{\delta-1-\gamma}\mathrm{d}s \\ &\leqslant \mathcal{C}\int_0^\infty (1+s)^{\frac{1}{6}-1-\frac{1}{6}-\epsilon}\mathrm{d}s \\ &< \infty.\end{aligned}$$

Finally

$$\begin{aligned}\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s r_s \mathrm{d}R_s &= \mathcal{C}\int_0^\infty (1+s)^{\delta-1-\rho}\mathrm{d}s \\ &\leqslant \begin{cases}\mathcal{C}\int_0^\infty (1+s)^{\frac{1}{2}-1-\frac{1}{2}-\epsilon}\mathrm{d}s & \text{in [c-RM-J]} \\ \mathcal{C}\int_0^\infty (1+s)^{\frac{1}{6}-1-\frac{1}{6}-\epsilon}\mathrm{d}s & \text{in [c-KW-H]}\end{cases} \\ &< \infty.\end{aligned}$$

□

### 7.4.3 Recursive Stochastic Approximation Algorithms

We also have analogous results for the time-discrete setting. In the following corollaries we achieve the same rates of convergence as in the previous subsection.

**Corollary 7.4.4.** *Consider the companion algorithm [d-Gen-Comp]* (6.14)*. Let Assumption 6.3.3 hold. Set $a_n = an^{-1}$, $a > 0$, and $\tilde{a}_n = n^{-1}$, $\tilde{a} > 0$. Assume $\sum_{n=1}^\infty \gamma_n(\delta)\tilde{a}_n|G_n - \upsilon^*| < \infty$ $\mathbb{P}$-a.s. Then almost surely $n^\delta|\Upsilon_n - \upsilon^*| \xrightarrow{n\to\infty} 0$ for all $\delta \in [0, 1/2)$.*

*Proof.* This corollary is ascribed to Theorem 7.2.1. Define $\tilde{V}_s$ and $M(\mathrm{d}s, y)$ as in the proof of Corollary 6.3.4. Choose $R_s := \lfloor s \rfloor$, $\tilde{a}_s := s^{-1}$, $a_s := as^{-1}$ and $c_s := cs^{-\gamma}$. The definition of $\mathcal{E}_t$, $\exp(\ln(x)) = x$ and a Taylor expansion yield

$$\begin{aligned}\gamma_t(\delta) = \mathcal{E}_t\left(\delta\int_0^\cdot \frac{1}{s}\mathrm{d}R_s\right) &= \prod_{i=1}^{\lfloor t\rfloor}\left(1+\frac{\delta}{i}\right) = \exp\left(\sum_{i=1}^{\lfloor t\rfloor}\ln\left(1+\frac{\delta}{i}\right)\right) \\ &= C_{\lfloor t\rfloor}\exp\left(\delta\sum_{i=1}^{\lfloor t\rfloor}\frac{1}{i}\right)\end{aligned}$$

with a term $C_{\lfloor t\rfloor}$ such that $C_{\lfloor t\rfloor} \to C_\infty$, with $C_\infty \in (0,\infty)$, for $t \to \infty$. Since

$$\exp\left(\delta \sum_{i=1}^{\lfloor t\rfloor} \frac{1}{i}\right) \geqslant \exp\left(\delta \int_1^{\lfloor t\rfloor} \frac{1}{x}\mathrm{d}x\right) = \lfloor t\rfloor^\delta$$

and

$$\exp\left(\delta \sum_{i=1}^{\lfloor t\rfloor} \frac{1}{i}\right) = \exp\left(\delta + \delta \sum_{i=2}^{\lfloor t\rfloor} \frac{1}{i}\right) \leqslant \exp\left(\delta + \delta \int_{i=1}^{\lfloor t\rfloor} \frac{1}{i}\mathrm{d}x\right) = \exp(\delta)\lfloor t\rfloor^\delta,$$

we can replace $\gamma_{t-}(\delta)$ by $\lfloor t\rfloor^\delta$ in Assumption 7.2.1. In order to show assumption $(\tilde{D})$, we recall that $\lfloor s\rfloor^{2\delta} k_s^2 h_s(\Upsilon_{s-})$ is positive for all $s$. Therefore it is sufficient to show $\mathbb{E}\int_0^\infty h_s(\Upsilon_{s-})\lfloor s\rfloor^{2\delta} k_s^2 \mathrm{d}R_s < \infty$. As in the previous proof, assume a sufficiently small $\epsilon > 0$. Condition $(\tilde{D})$ is verified by

$$\begin{aligned}\mathbb{E}\int_0^\infty h_s(\Upsilon_{s-})\lfloor s\rfloor^{2\delta} k_s^2 \mathrm{d}R_s &= \mathbb{E}\sum_{n\in\mathbb{N}} \mathbb{E}\Big(V_n^2 \mid \mathcal{F}_{n-1}\Big) n^{2\delta} n^{-2\kappa} \leqslant \mathcal{C}\Big(\sup_{n\in\mathbb{N}} \mathbb{E}V_n^2\Big) \sum_{n\in\mathbb{N}} n^{1-\epsilon-2}\\ &\leqslant \mathcal{C}\sum_{n\in\mathbb{N}} n^{-1-\epsilon} < \infty,\end{aligned}$$

where we made use of the monotone convergence theorem and the fact that $\delta$ is smaller than $\kappa - \frac{1}{2}$. Moreover

$$\int_0^\infty \gamma_{s-}(\delta)\tilde{a}_s|G_s|\mathrm{d}R_s \leqslant \mathcal{C}\sum_{n=1}^\infty \gamma_n(\delta)\tilde{a}_n|G_n| < \infty$$

yields $(\tilde{C})$. □

Coming to special algorithms [d-RM-J], [d-KW-H], [d-KW-F-2] and [d-KW-F-1] again, we begin with settings where the companion algorithms converge optimally.

**Corollary 7.4.5.** *Consider the algorithms [d-KW-F-2] and [d-KW-F-1]. Let Assumption 6.3.3 hold. Set* $a_n = an^{-1}$, $a > 0$, *and* $\tilde{a}_n = \tilde{a}n^{-1}$, $\tilde{a} > 0$. *In case of* $f$ *being* $p$*-times continuously differentiable at* $z^*$, *assume one of the following cases.*

| *Companion Algorithm* | $p$ | $c_n$ | $\delta$ |
|---|---|---|---|
| *[d-KW-F-2]* | 2 | $cn^{-\frac{1}{4}}$ | $[0, \frac{1}{2})$ |
| *[d-KW-F-2]* | 3 | $cn^{-\frac{1}{4}}$ | $[0, \frac{1}{2})$ |
| *[d-KW-F-1]* | 2 | $cn^{-\frac{1}{4}}$ | $[0, \frac{1}{2})$ |
| *[d-KW-F-1]* | 3 | $cn^{-\frac{1}{6}}$ | $[0, \frac{1}{2})$ |

*Then almost surely* $n^\delta|\Upsilon_n - \upsilon^*| \xrightarrow{n\to\infty} 0$.

We now turn to the settings where only the leading algorithms converge with optimal rate.

**Corollary 7.4.6.** *Consider the algorithms [d-KW-F-2], [d-RM-J] and [d-KW-H]. Assume that the leading algorithm converges with optimal rate. Let Assumption 6.3.3 hold. Set $a_n = an^{-1}$, $a > 0$, and $\tilde{a}_n = \tilde{a}n^{-1}$, $\tilde{a} > 0$. In case of $f$ being p-times continuously differentiable at $z^*$, assume one of the following cases.*

| *Companion Algorithm* | $p$ | $c_n$ | $\delta$ |
|---|---|---|---|
| *[d-KW-F-2]* | 3 | $cn^{-\frac{1}{6}}$ | $[0, \frac{1}{3})$ |
| *[d-RM-J]* | 1 | $cn^{-\frac{1}{4}}$ | $[0, \frac{1}{4})$ |
| *[d-KW-H]* | 3 | $cn^{-\frac{1}{6}}$ | $[0, \frac{1}{6})$ |

*Then almost surely* $n^\delta|\Upsilon_n - \upsilon^*| \xrightarrow{n\to\infty} 0$.

*Proof of Corollaries 7.4.5 and 7.4.6.* As before these corollaries are also ascribed to Theorem 7.3.1. Define $\tilde{V}_s$ and $M(\mathrm{d}s, y)$ as in the proof of Corollary 6.3.4. Choose $R_s := \lfloor s \rfloor$, $\tilde{a}_s := s^{-1}$, $a_s := as^{-1}$ and $c_s := cs^{-\gamma}$. In the proof of Corollary 7.4.4 it is already shown that we can replace $\gamma_{t-}(\delta)$ by $\lfloor t \rfloor^\delta$ in Assumption 7.2.1. Moreover it is shown there that $(B)$ and $(\tilde{D})$ follow from $(dB)$ and $(D)$, respectively. We complete the proof with the verification of $(\tilde{C}sp)$. For the two companion algorithms estimating the function value of $f$ at $z^*$ we have

$$
\begin{aligned}
\int_0^\infty \lfloor s \rfloor^\delta \tilde{a}_s r_s^2 \mathrm{d}R_s &\leqslant \mathcal{C} \sum_{n\in\mathbb{N}} n^{\delta-1-2\rho} \\
&\leqslant \begin{cases}
\mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{2}-1-\frac{1}{2}-\epsilon} & \text{for [d-KW-F-2] if } f \in C^2,\ \delta \in [0, \frac{1}{2}) \\
\mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{2}-1-\frac{2}{3}-\epsilon} & \text{for [d-KW-F-2] if } f \in C^3,\ \delta \in [0, \frac{1}{2}) \\
\mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{2}-1-\frac{1}{2}-\epsilon} & \text{for [d-KW-F-1] if } f \in C^2,\ \delta \in [0, \frac{1}{2}) \\
\mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{2}-1-\frac{2}{3}-\epsilon} & \text{for [d-KW-F-1] if } f \in C^3,\ \delta \in [0, \frac{1}{2}) \\
\mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{3}-1-\frac{2}{3}-\epsilon} & \text{for [d-KW-F-2] if } f \in C^3,\ \delta \in [0, \frac{1}{3})
\end{cases} \\
&< \infty
\end{aligned}
$$

and

$$
\begin{aligned}
\int_0^\infty \lfloor s \rfloor^\delta \tilde{a}_s c_s^2 \mathrm{d}R_s &\leqslant c^2 \sum_{n\in\mathbb{N}} n^{\delta-1-2\gamma} \\
&\leqslant \begin{cases}
\mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{2}-1-\frac{1}{2}-\epsilon} & \text{for [d-KW-F-2] if } f \in C^2,\ \delta \in [0, \frac{1}{2}) \\
\mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{2}-1-\frac{1}{2}-\epsilon} & \text{for [d-KW-F-2] if } f \in C^3,\ \delta \in [0, \frac{1}{2}) \\
\mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{3}-1-\frac{1}{3}-\epsilon} & \text{for [d-KW-F-2] if } f \in C^3,\ \delta \in [0, \frac{1}{3})
\end{cases} \\
&< \infty.
\end{aligned}
$$

For algorithm [d-RM-J], we find

$$\int_0^\infty \lfloor s \rfloor^\delta \tilde{a}_s c_s \mathrm{d}R_s = \sum_{n\in\mathbb{N}} n^{\delta-1-\gamma} \leqslant \mathcal{C} \sum_{n\in\mathbb{N}} n^{-1-\epsilon} < \infty,$$

and for [d-KW-H]

$$\int_0^\infty \lfloor s \rceil^\delta \tilde{a}_s c_s \mathrm{d}R_s = \mathcal{C} \sum_{n\in\mathbb{N}} n^{\delta-1-\gamma} \leqslant \mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{6}-1-\frac{1}{6}-\epsilon} < \infty.$$

Finally

$$\begin{aligned}\int_0^\infty \lfloor s \rceil^\delta \tilde{a}_s r_s \mathrm{d}R_s &= \mathcal{C} \sum_{n\in\mathbb{N}} n^{\delta-1-\rho} \\ &\leqslant \begin{cases} \mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{2}-1-\frac{1}{2}-\epsilon} & \text{for [d-RM-J]} \\ \mathcal{C} \sum_{n\in\mathbb{N}} n^{\frac{1}{6}-1-\frac{1}{6}-\epsilon} & \text{for [d-KW-H]} \end{cases} \\ &< \infty\end{aligned}$$

completes the proof. □

# 8 Asymptotic Normality of Companion Algorithms

In this section the asymptotic distribution of companion processes is identified. Knowledge of this distribution can be used to find optimal design parameters $a$, $c$, $\tilde{a}$ and $k$. From now on, we assume that the process $(R_t)_{t\geqslant 0}$ and especially the processes $(a_t)_{t\geqslant 0}$, $(c_t)_{t\geqslant 0}$, $(\tilde{a}_t)_{t\geqslant 0}$ and $(k_t)_{t\geqslant 0}$ are deterministic of the form

$$a_t = \frac{a}{(1+R_{t-})^{\alpha}}, \quad c_t = \frac{c}{(1+R_{t-})^{\gamma}}, \quad \tilde{a}_t = \frac{\tilde{a}}{(1+R_{t-})^{\tilde{\alpha}}} \quad \text{and} \quad k_t = \frac{k}{(1+R_{t-})^{\kappa}}$$

with $a, c, \tilde{a}, k > 0$ and $0 < \alpha, \gamma, \tilde{\alpha}, \kappa \leqslant 1$.

## 8.1 Almost $\mathbf{L^2}$-Convergence Rate

In order to show asymptotic normality of the companion algorithms (6.5), (6.6) and (6.7), we make use of Theorem A.1.2 in the appendix on the almost $L^2$-convergence rate of the Kiefer-Wolfowitz process [37, Theorem 3.1.]. A process $(Z_t)_{t\geqslant 0}$ is said to converge almost in $L^2$, if for any $\epsilon > 0$, there is an event $A_\epsilon$ of probability $\geqslant 1-\epsilon$, such that $(Z_t \mathbb{1}_{A_\epsilon})_{t\geqslant 0}$ converges in $L^2$. For the companion algorithm (6.4), which refers to the Robbins-Monro process, we need a result on the almost $L^2$-convergence rate given in Theorem 8.1.1 below.

Also useful is the following lemma, which can be found in [37] as Lemma 3.1. It can be employed to handle the impact of the leading algorithms on the companion algorithms.

**Lemma 8.1.1.** *Let $Z$ be a strong solution of the stochastic integral equation* (6.2) *or* (6.3) *on* $[0,\infty)$. *If there is a strictly positive, monotone increasing process $(S_t)_{t\geqslant 0}$ that satisfies $S_t \uparrow \infty$ and $S_t\|Z_t\| \to 0$, then, for all $\epsilon, \delta > 0$, a deterministic time $T(\epsilon,\delta)$ exists with*

$$\mathbb{P}\left[\sup_{t\geqslant T(\epsilon,\delta)} \|Z_t\| > \delta\right] < \epsilon.$$

*Remark* 8.1.1. Choosing $\gamma := 1/4$ (or $\gamma := 1/6$) for two (or three) times differentiable $f$, Theorem A.1.2 yields that there is a $T$ such that, for all $t \geqslant T$, the leading Kiefer-Wolfowitz algorithm converges with rate $-1/4$ (or $-1/3$, respectively) in the almost

$L^2$ sense. The following assumptions refer to the Robbins-Monro algorithm given in (6.2).

**Assumption 8.1.1.**

(RM $A$) *$f\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz-continuous.*

(RM $B$) *There exists a $z^*$ with $f(z^*) = 0$.*

(RM $D$) *The weight process $(a_s)_{s\geqslant 0}$ satisfies*

$$a_s > 0 \qquad a_s \downarrow 0 \qquad \int_0^\infty a_s \mathrm{d}R_s = \infty.$$

(RM $E$) *For every $i \in \{1, \ldots, d\}$ and $x \in \mathbb{R}^d$, we have*

$$\int_0^\infty a_s^2 \frac{h_s^{ii}(Z_{s-})}{1 + \|Z_{s-}\|^2} \mathrm{d}R_s < \infty, \text{ where } h_s^{ij}(z) := \frac{\mathrm{d}[\int_0^\cdot M_i(\mathrm{d}t, z), \int_0^\cdot M_j(\mathrm{d}t, z)]_s}{\mathrm{d}R_s}.$$

(RM $E^*$) *Assume that $f$ is continuously differentiable around $z^*$,*

$$\underset{i,j\in\{1,\ldots,d\}}{\forall} \ \underset{0<C<\infty}{\forall} \ \underset{0<K<\infty}{\exists} \ \underset{z\in\mathbb{R}^d}{\forall} \ \|z\| \leqslant C \Rightarrow \sup_{t\in[0,\infty)} |h_t^{ij}(z)| \leqslant K,$$

$$\int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s^d < \infty \text{ and} \int_0^\infty a_s^2 \mathrm{d}R_s < \infty.$$

The following theorem on the almost $L^2$-convergence rate is useful to investigate companion algorithms with a leading Robbins-Monro algorithm.

**Theorem 8.1.1** (Almost $L^2$-convergence rate of the Robbins-Monro process)**.** *Consider the Robbins-Monro process $Z$ given in [RM] (6.2). Assume a positive, deterministic, monotonously increasing function $(S_t)_{t\geqslant 0}$ with $S_t \uparrow \infty$ and $S_t\|Z_t\| \to 0$ a.s. Assumption 8.1.1 shall hold true. Let $J_{z^*}$ be the Jacobian of $f$ at $z^*$, and $\beta_{min}$ as well as $\beta_{max}$ its minimum and maximum eigenvalue, respectively. Moreover let $\alpha \in (1/2, 1]$. If $\alpha < 1$, assume $\beta_{\min} > 0$; if $\alpha = 1$ assume $\beta_{\min} > 1/(2a)$. Then, for all $\epsilon > 0$, there exists a process $(Y_t)_{t\geqslant 0}$ such that*

$$\mathbb{P}\left[\underset{t\geqslant 0}{\forall} Y_t = Z_t\right] \geqslant 1 - \epsilon \tag{8.1}$$

*and*

$$\mathbb{E}\|Y_t - z^*\|^2 = \mathcal{O}((1 + R_t)^{1-2\alpha}). \tag{8.2}$$

*Proof.* The proof is similar to that of Theorem 3.1 in [37] (Theorem A.1.2 in the appendix) referring to the Kiefer-Wolfowitz algorithm. Without loss of generality we may assume that $z^* = 0$. We construct a process $(Y_t)_{t\geqslant 0}$ with property (8.1). Next we calculate $\|Y_t\|^2$ and show that its local martingale part is even a martingale. Finally we establish a convergence rate for $\mathbb{E}\|Y_t\|^2$.

**Construction of $Y$.** As $f$ is continuous and differentiable at $z^* = 0$ we get

$$f(x) = J_0 x + f(x) - J_0 x =: J_0 x + V(x)$$

where $V(x) = o(\|x\|)$. Choose a $p^* \in [2\alpha - 1, 1]$. We distinguish the cases $\alpha < 1$ and $\alpha = 1$. If $\alpha = 1$ we have $2a\beta_{\min} > 1 \geqslant p^* > 0$. Set

$$\kappa := \begin{cases} \dfrac{2a\beta_{\min} - p^*}{4a} & \text{if } \alpha = 1 \\ \dfrac{\beta_{\min}}{3} & \text{if } \alpha < 1. \end{cases}$$

Note that this constant is strictly positive. The fact that $V(x) = o(\|x\|)$ yields

$$\underset{\rho>0}{\forall}\ \underset{\delta_1>0}{\exists}\ \underset{\|x\|<\delta_1}{\forall}\ \|V(x)\| \leqslant \rho\|x\|.$$

As $\kappa$ is strictly positive, it is acceptable to choose $\rho = \kappa$. With $\epsilon > 0$ and choosing $\delta := \min\{\delta_1, 1\}$, Lemma 8.1.1 guarantees the existence of a deterministic time $T(\epsilon, \delta) < \infty$ with

$$\mathbb{P}\left(\left[\sup_{t \geqslant T(\epsilon,\delta)} \|Z_t\| \leqslant \delta\right]\right) \geqslant 1 - \epsilon.$$

Recall that $a_s = a(1 + R_{s-})^{-\alpha}$. Conditions (RM $D$) and (RM $E^*$) justify the implication

$$\int_0^\infty a_s \mathrm{d}R_s = \infty \ \wedge \ \int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s = \int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s^d < \infty \ \Rightarrow \ \frac{\Delta R_s}{(1 + R_{s-})^\alpha} \xrightarrow{s\to\infty} 0.$$

Together with the assumption that $(R_s)_{s\geqslant 0}$ is deterministic this yields for $\alpha = 1$ that

$$\underset{s_0}{\exists}\ \underset{s\geqslant s_0}{\forall}\ a_s \Delta R_s \leqslant \frac{2a\beta_{\min} - p^*}{2a(\beta_{\max} + \kappa)^2} \tag{8.3}$$

and for $\alpha < 1$

$$\underset{s_1}{\exists}\ \underset{s\geqslant s_1}{\forall} \left( a_s \Delta R_s \leqslant \frac{\beta_{\min}}{(\beta_{\max} + \kappa)^2} \quad \wedge \quad (1 + R_t)^{\alpha - 1} \leqslant \frac{a\beta_{\min}}{3p^*} \right). \tag{8.4}$$

Note that the times $T(\epsilon, \delta)$, $s_0$ and $s_1$ are all deterministic. For that reason

$$T := \begin{cases} \max\{T(\epsilon, \delta), s_0\} & \text{if } \alpha = 1 \\ \max\{T(\epsilon, \delta), s_1\} & \text{if } \alpha < 1 \end{cases}$$

is deterministic as well, and hence especially has the properties of a stopping time. Consequently, as $Z$ is an adapted càdlàg process, according to [32, Theorem 4],

$$D := \inf\{t > T \mid \|Z_t\| > \delta\}$$

defines another, proper, stopping time. Now we are prepared to define $(Y_t)_{t\geqslant 0}$ by

$$\begin{aligned}
Y_t :=\ & Z_t^D \\
=& Z_t \mathbb{1}_{[0,T)}(t) + Z_T^D \mathbb{1}_{[T,\infty)}(t)\mathbb{1}_{[T\neq D]} + \int_{T+}^{t\wedge D} \mathrm{d}Z_s \\
=&\ Z_t \mathbb{1}_{[0,T)}(t) + Z_T \mathbb{1}_{[T,\infty)}(t)\mathbb{1}_{[T\neq D]} \\
&\quad - \int_{T+}^{t} a_s f(Z_{s-})\mathbb{1}_{(T,D]}(s)\mathrm{d}R_s - \int_{T+}^{t} a_s \mathbb{1}_{(T,D]}(s) M(\mathrm{d}s, Z_{s-}).
\end{aligned}$$

Investigating the process $(Y_t)_{t\geqslant 0}$ on the set $[D=\infty]$ yields

$$Y_t \mathbb{1}_{[D=\infty]} = Z_t \mathbb{1}_{[0,T)}(t) + \mathbb{1}_{[T,\infty)}(t)\Big(Z_T + \int_{T+}^{t} \mathrm{d}Z_s\Big) = Z_t \mathbb{1}_{[0,T)}(t) + \mathbb{1}_{[T,\infty)}(t) Z_t = Z_t.$$

Consequently $(Y_t)_{t\geqslant 0}$ may differ from $(Z_t)_{t\geqslant 0}$ only on the set $[D<\infty]$. From $T \geqslant T(\epsilon,\delta)$ we conclude $\sup_{t>T}\|Z_t\| \leqslant \sup_{t>T(\epsilon,\delta)}\|Z_t\|$ and therefore $\mathbb{P}\Big(\Big[\sup_{t>T}\|Z_t\| \leqslant \delta\Big]\Big) \geqslant \mathbb{P}\Big(\Big[\sup_{t>T(\epsilon,\delta)}\|Z_t\| \leqslant \delta\Big]\Big)$. So according to

$$\begin{aligned}
\mathbb{P}\Big(\Big[\underset{t\geqslant 0}{\forall}\, Z_t = Y_t\Big]\Big) &= \mathbb{P}([D<\infty]) + \mathbb{P}([D=\infty]) \geqslant \mathbb{P}([D=\infty]) = \mathbb{P}\Big(\Big[\sup_{t\geqslant T}\|Z_t\| \leqslant \delta\Big]\Big) \\
&\geqslant \mathbb{P}\Big(\Big[\sup_{t>T(\epsilon,\delta)}\|Z_t\| \leqslant \delta\Big]\Big) \geqslant 1-\epsilon
\end{aligned}$$

the $\mathbb{P}$-measure of such a set $[D<\infty]$ is at most $\epsilon$.

From now on, we assume that $t>T$ holds. A straightforward calculation yields

$$\begin{aligned}
\|Y_t\|^2 &= \|Y_T\|^2 + \int_{T+}^{t} \mathrm{d}\|Y_s\|^2 \\
&= \|Z_T^D \mathbb{1}_{[T\neq D]}\|^2 + M_t^* + \sum_{i=1}^{d}\Big(\int_{T+}^{t} a_s^2 h_s^{ii}(Z_{s-})\mathbb{1}_{(T,D]}(s)\mathrm{d}R_s \\
&\quad - 2\int_{T+}^{t} a_s f(Z_{s-})_i Z_{s-}^i \mathbb{1}_{(T,D]}(s)\mathrm{d}R_s + \int_{T+}^{t} a_s^2 f(Z_{s-})_i^2 \mathbb{1}_{(T,D]}(s)\Delta R_s \mathrm{d}R_s^d\Big),
\end{aligned} \tag{8.5}$$

with a local martingale given by

$$\begin{aligned}
M_t^* = \sum_{i=1}^{d}\Bigg(&\int_{T+}^{t} a_s^2 f(Z_{s-})_i \Delta R_s \mathbb{1}_{(T,D]}(s) M_i^d(\mathrm{d}s, Z_{s-}) - \int_{T+}^{t} a_s Z_{s-}^i \mathbb{1}_{(T,D]}(s) M_i(\mathrm{d}s, Z_{s-}) \\
&+ \int_{T+}^{t} a_s^2 \mathbb{1}_{(T,D]}(s)\Big([M_i(\mathrm{d}r, Z_{r-})]_s - \lceil M_i(\mathrm{d}r, Z_{r-})\rceil_s\Big)\Bigg).
\end{aligned} \tag{8.6}$$

We furthermore have

$$\mathbb{1}_{(T,D]}(t) Y_t = Z_T^D \mathbb{1}_{(T,D]}(t)\mathbb{1}_{[T\neq D]} + \mathbb{1}_{(T,D]}(t)\int_{T+}^{t\wedge D} \mathrm{d}Z_s$$

$$
\begin{aligned}
&= Z_T^D \mathbb{1}_{(T,D]}(t) + \mathbb{1}_{(T,D]}(t) \int_{T+}^{t\wedge D} \mathrm{d}Z_s = \mathbb{1}_{(T,D]}(t)\Big(Z_T^D + \int_{T+}^{t} \mathrm{d}Z_s^D\Big) \\
&= \mathbb{1}_{(T,D]}(t) Z_t^D = \mathbb{1}_{(T,D]}(t) Z_t
\end{aligned}
$$

as well as

$$
\mathbb{1}_{(T,D]}(t)\|Y_{t-}\| = \mathbb{1}_{(T,D]}(t)\|Z_{t-}\| \leqslant \delta \mathbb{1}_{(T,D]}(t). \tag{8.7}
$$

In order to bound the terms $\sum_{i=1}^d Z_{s-}^i f(Z_{s-})_i$ and $\sum_{i=1}^d f(Z_{s-})_i^2$ we establish some inequalities. For a time $s$ with $s > T$ we have

$$
f(Z_{s-})_i = (J_0 Z_{s-})^i + V^i(Z_{s-})
$$

which yields

$$
\begin{aligned}
-\sum_{i=1}^d f(Z_{s-})_i Z_{s-}^i &= -\sum_{i=1}^d \Big((J_0 Z_{s-})^i + V^i(Z_{s-})\Big) Z_{s-}^i \\
&\leqslant -\beta_{\min}\|Z_{s-}\|^2 + \|V(Z_{s-})\|\|Z_{s-}\| \\
&\leqslant -\beta_{\min}\|Z_{s-}\|^2 + \kappa\|Z_{s-}\|^2
\end{aligned} \tag{8.8}
$$

and

$$
\sum_{i=1}^d f(Z_{s-})_i^2 = \|f(Z_{s-})\|^2 \leqslant \Big(\|J_0 Z_{s-}\| + \underbrace{\|V(Z_{s-})\|}_{\leqslant \kappa\|Z_{s-}\|}\Big)^2 \leqslant (\beta_{\max} + \kappa)^2 \|Z_{s-}\|^2. \tag{8.9}
$$

**Martingale property of $M^*$.** Next we show that the expectation value of the local martingale $(M_t^*)_{t\geqslant T}$ is zero. For that purpose we show that the local martingale is even a martingale. This can be done by the fact [32, Ch. II.6, Corollary 3] that a local martingale $M_t$ with $\mathbb{E}M_t^2 < \infty$ for all $t \geqslant 0$ is also a martingale if and only if $\mathbb{E}[M]_t < \infty$ for all $t$. As according to [23, p.60, Problem 7] for $M \in \mathcal{M}_{\mathrm{loc}}^2$ with $M_0 = 0$ it holds $\mathbb{E}[M]_t = \mathbb{E}[M]_t$ for all $t$, it is sufficient to show $\mathbb{E}[M]_t < \infty$. Condition (RM $E^*$) will be employed repeatedly. Finally the expectation of a martingale starting at zero is zero.

With (8.7) and condition (RM $E^*$) the second term in (8.6) is handled in the following way:

$$
\begin{aligned}
&\sup_{t>T} \mathbb{E}\Bigg[\sum_{i=1}^d \int_{T+}^{\cdot} a_s Z_{s-}^i \mathbb{1}_{(T,D](s)} M_i(\mathrm{d}s, Z_{s-})\Bigg]_t \\
&\qquad = \sup_{t>T} \mathbb{E} \sum_{i,j=1}^d \int_{T+}^t a_s^2 Z_{s-}^i Z_{s-}^j \mathbb{1}_{(T,D]}(s) [M_i(\mathrm{d}r, Z_{r-}), M_j(\mathrm{d}r, Z_{r-})]_s \\
&\qquad \leqslant \sum_{i,j=1}^d \sup_{t>T} \mathbb{E} \int_{T+}^t a_s^2 Z_{s-}^i Z_{s-}^j \mathbb{1}_{(T,D]}(s) h_s^{ij}(Z_{s-}) \mathrm{d}R_s \leqslant \mathcal{C} \int_0^\infty a_s^2 \mathrm{d}R_s < \infty.
\end{aligned}
$$

Before we investigate the first term, note

$$\begin{aligned}a_s^2 f(Z_{s-})_i \Delta R_s M_i(\mathrm{d}s, Z_{s-}) &= a_s^2 f(Z_{s-})_i \Delta R_s M_i^d(\mathrm{d}s, Z_{s-}) + a_s^2 f(Z_{s-})_i \Delta R_s M_i^c(\mathrm{d}s, Z_{s-})\\ &= a_s^2 f(Z_{s-})_i \Delta R_s M_i^d(\mathrm{d}s, Z_{s-}).\end{aligned}$$

Now use this identity, (8.7), (8.9) and condition (RM $E^*$) to show

$$\begin{aligned}&\sup_{t>T} \mathbb{E}\left[\sum_{i=1}^d \int_{T+}^{\cdot} a_s^2 f(Z_{s-})_i \mathbb{1}_{(T,D]}(s) \Delta R_s M_i(\mathrm{d}s, Z_{s-})\right]_t\\ &\qquad = \sup_{t>T} \mathbb{E} \sum_{i,j=1}^d \int_{T+}^t a_s^4 f(Z_{s-})_i f(Z_{s-})_j (\Delta R_s)^2 h_s^{ij}(Z_{s-}) \mathbb{1}_{(T,D]}(s) \mathrm{d}R_s\\ &\qquad = \mathcal{C} \sup_{t>T} \mathbb{E} \int_{T+}^t a_s^4 \Big(\sum_{l=1}^d f(Z_{s-})_l\Big)^2 \mathbb{1}_{(T,D]}(s) (\Delta R_s)^2 \mathrm{d}R_s\\ &\qquad \leqslant \mathcal{C} \int_0^\infty a_s^4 (\Delta R_s)^2 \mathrm{d}R_s = \mathcal{C} \int_0^\infty a_s^2 (a_s \Delta R_s)^2 \mathrm{d}R_s\\ &\qquad = \mathcal{C} \int_0^\infty a_s^2 o_{\mathrm{b}}(1) \mathrm{d}R_s < \mathcal{C} \int_0^\infty a_s^2 \mathrm{d}R_s < \infty.\end{aligned}$$

In order to investigate the last term in (8.6) we use the following fact. If $M$ is a locally square integrable martingale starting at zero, we have $\mathbb{E}[M]_t = \mathbb{E}\lceil M \rceil_t$ for all $t$ [23, p.60, Problem 7]. Hence

$$\begin{aligned}&\mathbb{E} \int_{T+}^t a_r^2 \mathbb{1}_{(T,D]}(r) \Big([M_i(\mathrm{d}l, Z_{l-})]_r - \lceil M_i(\mathrm{d}l, Z_{l-}) \rceil_r\Big)\\ &\quad = \mathbb{E}\Big[\underbrace{\int_{T+}^{\cdot} a_r \mathbb{1}_{(T,D]}(r) M_i(\mathrm{d}r, Z_{r-})}_{\in \mathcal{M}^2_{\mathrm{loc}}}\Big]_t - \mathbb{E}\Big\lceil \underbrace{\int_{T+}^{\cdot} a_r \mathbb{1}_{(T,D]}(r) M_i(\mathrm{d}r, Z_{r-})}_{\in \mathcal{M}^2_{\mathrm{loc}}}\Big\rceil_t\\ &\quad = 0.\end{aligned}$$

Consequently $(M_t^*)_{t \geqslant T}$ is a proper martingale.

**Convergence rate of $\mathbb{E}\|Y_t\|^2$.** Now we investigate

$$\begin{aligned}&\mathrm{d}\Big((1+R_t)^{p^*} \mathbb{E}\|Y_t\|^2\Big)\\ &\quad = (1+R_{t-})^{p^*} \mathrm{d}\mathbb{E}\|Y_t\|^2 + \mathbb{E}\|Y_{t-}\|^2 \mathrm{d}(1+R_t)^{p^*} + \mathrm{d}[(1+R_\cdot)^{p^*}, \mathbb{E}\|Y_\cdot\|^2]_t. \qquad (8.10)\end{aligned}$$

With (8.5) the expectation of $\|Y_t\|^2$ can be rewritten as

$$\begin{aligned}\mathbb{E}\|Y_t\|^2 = \mathbb{E}\|Z_T^D \mathbb{1}_{[T \neq D]}\|^2 &+ \int_{T+}^t a_s^2 \mathbb{E}\Big(\sum_{i=1}^d f(Z_{s-})_i^2 \mathbb{1}_{(T,D]}(s)\Big) \Delta R_s \mathrm{d}R_s^d\\ &+ \int_{T+}^t a_s^2 \mathbb{E}\Big(\sum_{i=1}^d h_s^{ii}(Z_{s-}) \mathbb{1}_{(T,D]}(s)\Big) \mathrm{d}R_s\end{aligned}$$

$$- 2\int_{T+}^{t} a_s \mathbb{E}\Big(\sum_{i=1}^{d} f(Z_{s-})_i Z_{s-}^i \mathbb{1}_{(T,D]}(s)\Big) \mathrm{d}R_s.$$

Therefore the first term on the right hand side in (8.10) can be bounded with the help of (8.8) and (8.9):

$$\begin{aligned}
&\int_{T+}^{t} (1+R_{s-})^{p^*} \mathrm{d}\mathbb{E}\|Y_s\|^2 \\
&\quad \leqslant \int_{T+}^{t} (1+R_{s-})^{p^*} (\beta_{\max}+\kappa)^2 \frac{a\Delta R_s}{(1+R_{s-})^{\alpha}} a(1+R_{s-})^{-\alpha} \mathbb{E}\|Z_{s-}\|^2 \mathrm{d}R_s \\
&\qquad + \mathcal{C}\int_{T+}^{t} (1+R_{s-})^{p^*-2\alpha} \mathrm{d}R_s \\
&\qquad + 2\int_{T+}^{t} (1+R_{s-})^{p^*} (-\beta_{\min}+\kappa) a(1+R_{s-})^{-\alpha} \mathbb{E}\|Z_{s-}\|^2 \mathrm{d}R_s \\
&\quad \leqslant a\int_{T+}^{t} \Big(-2\beta_{\min} + (\beta_{\max}+\kappa)^2 \frac{a\Delta R_s}{(1+R_{s-})^{\alpha}} + 2\kappa\Big)(1+R_{s-})^{p^*-\alpha} \mathbb{E}\|Z_{s-}\|^2 \mathrm{d}R_s \\
&\qquad + \mathcal{C}\int_{T+}^{t} (1+R_{s-})^{p^*-2\alpha} \mathrm{d}R_s \\
&\quad \leqslant a\int_{T+}^{t} \Big(-2\beta_{\min} + v_s + 2\kappa\Big)(1+R_{s-})^{p^*-\alpha} \mathbb{E}\|Z_{s-}\|^2 \mathrm{d}R_s \\
&\qquad + \mathcal{C}\int_{T+}^{t} (1+R_{s-})^{p^*+\rho-1} \mathrm{d}R_s. 
\end{aligned} \tag{8.11}$$

Here we define $v_s := (\beta_{\max}+\kappa)^2 \frac{a\Delta R_s}{(1+R_{s-})^{\alpha}}$ and $\rho := 1-2\alpha$. By Itô's formula and a Taylor expansion of $f\colon x \mapsto (1+x)^{p^*}$ around $R_{s-}$ with a $\vartheta_s \in [0,1]$ we get

$$\begin{aligned}
&(1+R_t)^{p^*} - 1 \\
&\quad = \int_{0+}^{t} p^*(1+R_{s-})^{p^*-1} \mathrm{d}R_s \\
&\qquad + \sum_{0<s\leqslant t} \Big\{(1+R_s)^{p^*} - (1+R_{s-})^{p^*} - p^*(1+R_{s-})^{p^*-1}\Delta R_s\Big\} \\
&\quad = \int_{0+}^{t} p^*(1+R_{s-})^{p^*-1} \mathrm{d}R_s + \sum_{0<s\leqslant t} \Big\{ \underbrace{\frac{1}{2}p^*}_{\geqslant 0} \underbrace{(p^*-1)}_{<0} \underbrace{(1+R_{s-}+\vartheta_s\Delta R_s)^{p^*-2}}_{\geqslant 0} \Big\} \\
&\quad \leqslant \int_{0+}^{t} p^*(1+R_{s-})^{p^*-1} \mathrm{d}R_s.
\end{aligned}$$

Hence for the second term on the right hand side in (8.10) we have

$$\int_{T+}^{t} \mathbb{E}\|Y_{s-}\|^2 \mathrm{d}(1+R_s)^{p^*} \leqslant p^* \int_{T+}^{t} (1+R_{s-})^{p^*-1} \mathbb{E}\|Y_{s-}\|^2 \mathrm{d}R_s.$$

By the mean value theorem there is a $\vartheta_t \in [0,1]$ such that

$$
\begin{aligned}
0 &\leqslant \Delta(1+R_t)^{p^*} = (1+R_t)^{p^*} - (1+R_{t-})^{p^*} = p^*(1+R_{t-}+\vartheta_t\Delta R_t)^{p^*-1}\Delta R_t \\
&\leqslant p^*(1+R_{t-})^{p^*-1}\Delta R_t.
\end{aligned}
$$

Hence the last term in (8.10) can be bounded in a similar way as (8.11):

$$
\begin{aligned}
&\int_{T+}^{t} \mathrm{d}\Big[(1+R_\cdot)^{p^*}, \mathbb{E}\|Y_\cdot\|^2\Big]_s \\
&\quad = \int_{T+}^{t} \Delta(1+R_s)^{p^*} \mathrm{d}\mathbb{E}\|Y_s\|^2 \\
&\quad \leqslant a \int_{T+}^{t} \Big(-2\beta_{\min} + v_s + 2\kappa\Big)(1+R_{s-})^{-\alpha}\mathbb{E}\|Z_{s-}\|^2 \Delta(1+R_s)^{p^*}\mathrm{d}R_s \\
&\qquad + \mathcal{C}\int_{T+}^{t} (1+R_{s-})^{\rho-1}\Delta(1+R_s)^{p^*}\mathrm{d}R_s \\
&\quad \overset{(\star)}{\leqslant} \mathcal{C}\int_{T+}^{t} (1+R_{s-})^{p^*-1+\rho-1}\Delta R_s \mathrm{d}R_s,
\end{aligned}
$$

where $(\star)$ is discussed below. All terms in the last inequality, especially $v_s$, are purely deterministic. A combination of all bounds for the term on the right hand side of (8.10) yields

$$
\begin{aligned}
&\int_{T+}^{t} \mathrm{d}\Big((1+R_s)^{p^*}\mathbb{E}\|Y_s\|^2\Big) \\
&\quad \leqslant a \int_{T+}^{t} \Big(-2\beta_{\min} + v_s + 2\kappa + \frac{p^*}{a}(1+R_{s-})^{\alpha-1}\Big)(1+R_{s-})^{p^*-\alpha}\mathbb{E}\|Z_{s-}\|^2\mathrm{d}R_s \\
&\qquad + \mathcal{C}\int_{T+}^{t} (1+R_{s-})^{p^*+\rho-1}\mathrm{d}R_s \\
&\quad \overset{(\star)}{\leqslant} \mathcal{C}\int_{T+}^{t} (1+R_{s-})^{p^*+\rho-1}\mathrm{d}R_s
\end{aligned}
\tag{8.12}
$$

where we made use of $\frac{\Delta R_s}{(1+R_s)^\alpha} = o_{\mathrm{b}}(1)$ and $(\star)$. Now we discuss $(\star)$. In the case $\alpha = 1$ (8.3) yields

$$
\begin{aligned}
&-2\beta_{\min} + (\beta_{\max}+\kappa)^2\frac{a\Delta R_s}{(1+R_{s-})} + 2\kappa + \frac{p^*}{a} \\
&\qquad \leqslant -2\beta_{\min} + (\beta_{\max}+\kappa)^2\frac{2a\beta_{\min}-p^*}{2a(\beta_{\max}+\kappa)^2} + 2\frac{2a\beta_{\min}-p^*}{4a} + \frac{p^*}{a} = 0
\end{aligned}
$$

and in the case $\alpha < 1$ with (8.4) we get

$$
\begin{aligned}
&-2\beta_{\min} + (\beta_{\max}+\kappa)^2\frac{a\Delta R_s}{(1+R_{s-})^\alpha} + 2\kappa + \frac{p^*}{a}(1+R_{s-})^{\alpha-1} \\
&\qquad \leqslant -2\beta_{\min} + (\beta_{\max}+\kappa)^2\frac{\beta_{\min}}{(\beta_{\max}+\kappa)^2} + 2\frac{\beta_{\min}}{3} + \frac{p^*}{a}\frac{a\beta_{\min}}{3p^*} = 0.
\end{aligned}
$$

Application of Itô's formula to the right hand side of (8.12) yields

$$\begin{aligned}\int_{T+}^{t} \mathrm{d}\Big((1+R_s)^{p^*}\mathbb{E}\|Y_s\|^2\Big) &\leqslant \mathcal{C}\int_{T+}^{t}(1+R_{s-})^{p^*+\rho-1}\mathrm{d}R_s\\ &= \mathcal{C}\int_{T+}^{t}\mathrm{d}(1+R_s)^{p^*+\rho} + \mathcal{C}\int_{T+}^{t}(1+R_{s-})^{p^*+\rho-2}\Delta R_s\mathrm{d}R_s.\end{aligned}$$

Condition (RM $E^*$) implies $\int_0^\infty (1+R_{s-})^{-2\alpha}\Delta R_s \mathrm{d}R_s < \infty$ and thus by Kronecker's lemma

$$\frac{\int_{T+}^{t}(1+R_{s-})^{p^*+\rho-2}\Delta R_s\mathrm{d}R_s}{(1+R_t)^{p^*+\rho+2\alpha-2}} \to 0 \text{ as } t\to\infty,$$

which yields a convergence rate for $\int_{T+}^{t}(1+R_{s-})^{p^*+\rho-2}\Delta R_s\mathrm{d}R_s$. Hence we have

$$(1+R_t)^{p^*}\mathbb{E}\|Y_t\|^2 \leqslant (1+R_T)^{p^*}\mathbb{E}\|Y_T\|^2 + \mathcal{C}(1+R_t)^{p^*+\rho} + o(1)(1+R_t)^{p^*+\rho+2\alpha-2}$$

and

$$\mathbb{E}\|Y_t\|^2 \leqslant \mathcal{C}(1+R_t)^{-p^*} + \mathcal{C}(1+R_t)^{\rho} + o(1)(1+R_t)^{\rho+2\alpha-2}.$$

Our assumptions and the choice of $p^*$ guarantee $\alpha \leqslant 1$ as well as $-p^* \leqslant \rho$ and $\rho \geqslant \rho + 2\alpha - 2$. As a result

$$\mathbb{E}\|Y_t\|^2 = \mathcal{O}\Big((1+R_t)^{\rho}\Big) = \mathcal{O}\Big((1+R_t)^{1-2\alpha}\Big)$$

holds and the theorem is proven. □

## 8.2 Explicit Solution of a Stochastic Integral Equation

The following lemma, a representation for the solution of [Gen-Comp] (5.4), is employed to show asymptotic normality. As a general assumption for the generic companion algorithm we already assumed the existence of a unique solution of [Gen-Comp] (5.4) on $[0,\infty)$. Now we construct an explicit solution.

**Lemma 8.2.1.** *Let $G$ be a left-continuous adapted process with $G_t \xrightarrow{t\to\infty} \upsilon^*$ $\mathbb{P}$-a.s. Choose $k>0$, $k_s := k(1+R_{s-})^{-\kappa}$, $\tilde{a}>0$ and $\tilde{a}_s := \tilde{a}(1+R_{s-})^{-\tilde{\alpha}}$ such that $\int_0^\infty \tilde{a}_s \mathrm{d}R_s = \infty$ and $\int_0^\infty \tilde{a}_s^2\Delta R_s \mathrm{d}R_s^d < \infty$ hold. Then the companion stochastic integral equation*

$$\Upsilon_t = \Upsilon_0 + \int_0^t \tilde{a}_s\left(G_s - \Upsilon_{s-}\right)\mathrm{d}R_s + \int_0^t k_s M(\mathrm{d}s, \Upsilon_{s-}) \tag{8.13}$$

*is solved by*

$$\Upsilon_t = \phi_t\left(\Upsilon_0 + k\int_0^t (1+R_{s-})^{-\kappa}\phi_s^{-1}M(\mathrm{d}s,\upsilon^*) + \int_0^t \phi_s^{-1}\mathrm{d}\tilde{R}_s\right) \tag{8.14}$$

*where*

$$\tilde{R}_s := \tilde{a}\int_0^t G_s(1+R_{s-})^{-\tilde{\alpha}}\mathrm{d}R_s + \sum_{s\leqslant t}\Upsilon_{s-}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}$$
$$+\, k\int_0^t (1+R_{s-})^{-\kappa}\left(M(\mathrm{d}s,\Upsilon_{s-}) - M(\mathrm{d}s, \upsilon^*)\right)$$

*and*

$$\phi_t := \mathcal{E}_t\left(-\int_0^{\cdot} \bar{a}^s(1+R_{s-})^{-\tilde{\alpha}}\mathrm{d}R_s\right)$$

*with*

$$\bar{a}^s := \tilde{a}\mathbb{1}_{\{\tilde{a}\Delta R_s\neq(1+R_{s-})^{\tilde{\alpha}}\}}.$$

*If $\tilde{\alpha}=1$ and $\sum_{0\leqslant t}\mathbb{1}_{\{\tilde{a}\Delta R_t=1+R_{t-}\}}<\infty$, the function $(\phi_t)_{t\geqslant 0}$ can be represented as*

$$\phi_t = (1+R_t)^{-\tilde{a}}\prod\nolimits_{0,t}$$

*with*

$$\prod\nolimits_{0,t} := \prod_{0<s\leqslant t}\left(\left(1-\frac{\tilde{a}\Delta R_s}{1+R_{s-}}+\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}\right)\left(1+\frac{\Delta R_s}{1+R_{s-}}\right)^{\tilde{a}}\right) \tag{8.15}$$

*where $\prod_{0,t}$ converges pointwise to a real number $\prod_\infty$ as $t\to\infty$.*
*If $\tilde{\alpha}<1$ and $\sum_{0\leqslant t}\mathbb{1}_{\{\tilde{a}\Delta R_t=(1+R_{t-})^{\tilde{\alpha}}\}}<\infty$, the function $(\phi_t)_{t\geqslant 0}$ can be represented as*

$$\phi_t = \exp\Big(-\frac{\tilde{a}}{1-\tilde{\alpha}}(1+R_t)^{1-\tilde{\alpha}}\Big)\prod\nolimits_{0,t}$$

*with*

$$\prod\nolimits_{0,t} :=$$
$$e^{\frac{\tilde{a}}{1-\tilde{\alpha}}}\prod_{0<s\leqslant t}\left(\Big(1-\frac{\tilde{a}\Delta R_s}{(1+R_{s-})^{\tilde{\alpha}}}+\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}\Big)\exp\Big(\frac{\tilde{a}}{1-\tilde{\alpha}}\Delta(1+R_s)^{1-\tilde{\alpha}}\Big)\right) \tag{8.16}$$

*where $\prod_{0,t}$ converges pointwise to a real number $\prod_\infty$ as $t\to\infty$.*
*In both cases, $\tilde{\alpha}=1$ and $\tilde{\alpha}<1$,*

$$\prod\nolimits_{s,t} := \frac{\prod_{0,t}}{\prod_{0,s}} = \big(1+o_{\mathrm{b}}(1)\big) \quad s\to t \text{ pointwise for all } t\geqslant 0 \tag{8.17}$$

*holds true.*

*Proof.* Insert (8.14) on the right and left side of (8.13). Equation (8.13) is equivalent to

$$
\begin{aligned}
\mathrm{d}\Upsilon_t &= \tilde{a}_t\left(G_t - \Upsilon_{t-}\right)\mathrm{d}R_t + k_t M(\mathrm{d}t, \Upsilon_{t-}) \\
&= \tilde{a}(1+R_{t-})^{-\tilde{\alpha}} G_t \mathrm{d}R_t - \tilde{a}\mathbb{1}_{\{\tilde{a}\Delta R_s \neq (1+R_{s-})^{\tilde{\alpha}}\}}(1+R_{t-})^{-\tilde{\alpha}}\Upsilon_{t-}\mathrm{d}R_t \\
&\quad - \tilde{a}\mathbb{1}_{\{\tilde{a}\Delta R_s = (1+R_{s-})^{\tilde{\alpha}}\}}(1+R_{t-})^{-\tilde{\alpha}}\Upsilon_{t-}\mathrm{d}R_t + k(1+R_{t-})^{-\kappa}M(\mathrm{d}t, \upsilon^*) \\
&\quad + k(1+R_{t-})^{-\kappa}\big(M(\mathrm{d}t, \Upsilon_{t-}) - M(\mathrm{d}t, \upsilon^*)\big) \\
&= -\bar{a}^t(1+R_{t-})^{-\tilde{\alpha}}\Upsilon_{t-}\mathrm{d}R_t + k(1+R_{t-})^{-\kappa}M(\mathrm{d}t, \upsilon^*) + \mathrm{d}\tilde{R}_t.
\end{aligned} \tag{8.18}
$$

From (8.14) we get

$$
\begin{aligned}
\mathrm{d}\left(\phi_t^{-1}\Upsilon_t\right) &= k(1+R_{t-})^{-\kappa}\phi_t^{-1}M(\mathrm{d}t, \upsilon^*) + \phi_t^{-1}\mathrm{d}\tilde{R}_t \\
&= \phi_t^{-1}\left(k(1+R_{t-})^{-\kappa}M(\mathrm{d}t, \upsilon^*) + \mathrm{d}\tilde{R}_t\right).
\end{aligned} \tag{8.19}
$$

According to Lemma A.1.6, $\mathrm{d}\left[\phi, \phi^{-1}\Upsilon\right]_t = \Delta\phi_t \mathrm{d}\left(\phi_t^{-1}\Upsilon_t\right)$ holds true. With this fact and equation (8.19) we find

$$
\begin{aligned}
\mathrm{d}\Upsilon_t &= \mathrm{d}\left(\phi_t\phi_t^{-1}\Upsilon_t\right) = \phi_{t-}\mathrm{d}\left(\phi_t^{-1}\Upsilon_t\right) + \left(\phi_{t-}^{-1}\Upsilon_{t-}\right)\mathrm{d}\phi_t + \mathrm{d}\left[\phi, \phi^{-1}\Upsilon\right]_t \\
&= \phi_t\mathrm{d}\left(\phi_t^{-1}\Upsilon_t\right) - \Delta\phi_t\mathrm{d}\left(\phi_t^{-1}\Upsilon_t\right) + \left(\phi_{t-}^{-1}\Upsilon_{t-}\right)\mathrm{d}\phi_t + \Delta\phi_t\mathrm{d}\left(\phi_t^{-1}\Upsilon_t\right) \\
&= \phi_t\mathrm{d}\left(\phi_t^{-1}\Upsilon_t\right) + \left(\phi_{t-}^{-1}\Upsilon_{t-}\right)\mathrm{d}\phi_t \\
&\overset{(8.19)}{=} k\left(1+R_{t-}\right)^{-\kappa}M(\mathrm{d}t, \upsilon^*) + \mathrm{d}\tilde{R}_t + \phi_{t-}^{-1}\Upsilon_{t-}\mathrm{d}\phi_t.
\end{aligned}
$$

By (8.18)

$$
\begin{aligned}
\mathrm{d}\Upsilon_t &= -\bar{a}^t\Upsilon_{t-}\left(1+R_{t-}\right)^{-\tilde{\alpha}}\mathrm{d}R_t + k\left(1+R_{t-}\right)^{-\kappa}M(\mathrm{d}t, \upsilon^*) + \mathrm{d}\tilde{R}_t \\
&= -\Upsilon_{t-}\left(-\phi_{t-}^{-1}\underbrace{\left(-\phi_{t-}\bar{a}^t\left(1+R_{t-}\right)^{-\tilde{\alpha}}\mathrm{d}R_t\right)}_{=\mathrm{d}\phi_t}\right) + k\left(1+R_{t-}\right)^{-\kappa}M(\mathrm{d}t, \upsilon^*) + \mathrm{d}\tilde{R}_t \\
&= \Upsilon_{t-}\phi_{t-}^{-1}\mathrm{d}\phi_t + k\left(1+R_{t-}\right)^{-\kappa}M(\mathrm{d}t, \upsilon^*) + \mathrm{d}\tilde{R}_t.
\end{aligned}
$$

As a result both sides of (8.14) are equal.

In order to show an alternative representation of $\phi$ we first investigate the argument of the stochastic exponential in the definition of $\phi$:

$$
\begin{aligned}
&-\int_0^t \frac{\bar{a}^s}{(1+R_{s-})^{\tilde{\alpha}}}\mathrm{d}R_s \\
&\quad = -\tilde{a}\int_0^t \left(1+R_{s-}\right)^{-\tilde{\alpha}}\mathbb{1}_{\{\tilde{a}\Delta R_s \neq (1+R_{s-})^{\tilde{\alpha}}\}}\mathrm{d}R_s \\
&\quad = -\tilde{a}\int_0^t \left(1+R_{s-}\right)^{-\tilde{\alpha}}\mathrm{d}R_s + \tilde{a}\int_0^t \left(1+R_{s-}\right)^{-\tilde{\alpha}}\mathbb{1}_{\{\tilde{a}\Delta R_s = (1+R_{s-})^{\tilde{\alpha}}\}}\mathrm{d}R_s \\
&\quad = -\tilde{a}\int_0^t \left(1+R_{s-}\right)^{-\tilde{\alpha}}\mathrm{d}R_s + \sum_{0<s\leqslant t}\frac{\tilde{a}\Delta R_s}{(1+R_{s-})^{\tilde{\alpha}}}\mathbb{1}_{\{\tilde{a}\Delta R_s = (1+R_{s-})^{\tilde{\alpha}}\}}
\end{aligned}
$$

$$= -\tilde{a}\int_0^t (1+R_{s-})^{-\tilde{\alpha}}\,\mathrm{d}R_s + \sum_{0<s\leqslant t} \mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}. \tag{8.20}$$

For the case $\tilde{\alpha} = 1$ we apply Itô's formula with $f'(x) = (1+x)^{-1}$:

$$\begin{aligned}
&-\tilde{a}\int_0^t (1+R_{s-})^{-1}\mathrm{d}R_s \\
&\quad= -\tilde{a}\ln(1+R_t) + \tilde{a}\sum_{0<s\leqslant t}\left\{\ln(1+R_s) - \ln(1+R_{s-}) - \frac{\Delta R_s}{1+R_{s-}}\right\} \\
&\quad= -\tilde{a}\ln(1+R_t) + \tilde{a}\sum_{0<s\leqslant t}\left\{\ln\left(\frac{1+R_s}{1+R_{s-}}\right) - \frac{\Delta R_s}{1+R_{s-}}\right\} \\
&\quad= -\tilde{a}\ln(1+R_t) + \tilde{a}\sum_{0<s\leqslant t}\left\{\ln\left(1+\frac{\Delta R_s}{1+R_{s-}}\right) - \frac{\Delta R_s}{1+R_{s-}}\right\}.
\end{aligned}$$

Equation (8.20) yields

$$\begin{aligned}
-\Delta\int_0^t \frac{\bar{a}^s}{1+R_{s-}}\mathrm{d}R_s &= -\tilde{a}\Delta\int_0^t (1+R_{s-})^{-1}\mathrm{d}R_s + \Delta\sum_{0<s\leqslant t}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}} \\
&= -\frac{\tilde{a}\Delta R_t}{1+R_{t-}} + \mathbb{1}_{\{\tilde{a}\Delta R_t=(1+R_{t-})\}}.
\end{aligned}$$

Since $\mathcal{E}_t(X) = e^{X_t - X_0 - \frac{1}{2}[X,X]_t^c}\prod_{0<s\leqslant t}(1+\Delta X_s)e^{-\Delta X_s}$ we obtain

$$\begin{aligned}
\phi_t &= \mathcal{E}_t\left(-\int_0^{\cdot}\frac{\bar{a}^s}{(1+R_{s-})}\mathrm{d}R_s\right) \\
&= \exp\Bigg(-\tilde{a}\ln(1+R_t) + \tilde{a}\sum_{0<s\leqslant t}\left\{\ln\left(1+\frac{\Delta R_s}{1+R_{s-}}\right) - \frac{\Delta R_s}{1+R_{s-}}\right\} \\
&\quad + \sum_{0<s\leqslant t}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}\Bigg) \\
&\quad \cdot \prod_{0<s\leqslant t}\left(1 - \frac{\tilde{a}\Delta R_s}{1+R_{s-}} + \mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}\right)\exp\left(\frac{\tilde{a}\Delta R_s}{1+R_{s-}} - \mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}\right) \\
&= (1+R_t)^{-\tilde{a}}\underbrace{\prod_{0<s\leqslant t}\left(\left(1 - \frac{\tilde{a}\Delta R_s}{1+R_{s-}} + \mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}\right)\left(1+\frac{\Delta R_s}{1+R_{s-}}\right)^{\tilde{a}}\right)}_{=\prod_{0,t}} \\
&= (1+R_t)^{-\tilde{a}}\prod\nolimits_{0,t}.
\end{aligned}$$

To show convergence of $\prod_{0,t}$ we separate it into two factors $\prod_{0,\tau}$ and $\prod_{\tau,t}$ using a stopping time $\tau$. Choose $\tau$ in such a way that the second factor can be approximated by a Taylor expansion. Then convergence of $\prod_{\tau,t}$ and boundedness of $\prod_{0,\tau}$ yield convergence of $\prod_{0,t}$. Consider the following Taylor expansions:

$$\begin{aligned}
\ln(1-ax) &= -ax - \frac{a^2x^2}{2} + o(x^2) \quad &&\text{for } -1 < ax < 1 \\
a\ln(1+x) &= ax - \frac{ax^2}{2} + o(x^2) &&\text{for } -1 < x < 1
\end{aligned}$$

We define a (deterministic) time

$$\tau^1 := \min\left\{t \in \mathbb{R}_+ : \frac{\Delta R_s}{1+R_{s-}} < \frac{1}{\tilde{a}+1} \text{ for all } s > t\right\}.$$

As assumptions $\int_0^\infty \tilde{a}_s \mathrm{d}R_s = \infty$ and $\int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d < \infty$ imply $\frac{\Delta R_s}{1+R_{s-}} \xrightarrow{s\to\infty} 0$, $\tau^1 < \infty$ holds true. With the two Taylor expansions above we obtain

$$\begin{aligned}
&\ln\left(1 - \frac{\tilde{a}\Delta R_s}{1+R_{s-}}\right) + \tilde{a}\ln\left(1 + \frac{\Delta R_s}{1+R_{s-}}\right) \\
&\qquad\qquad = \left(-\frac{\tilde{a}}{2} - \frac{\tilde{a}^2}{2} + \rho_s\right)\left(\frac{\Delta R_s}{1+R_{s-}}\right)^2 \text{ where } \rho_s \xrightarrow{s\to\infty} 0.
\end{aligned}$$

This motivates the choice of the time

$$\tau := \min\left\{t \geqslant \tau^1 : |\rho_s| < 1 \text{ for all } s > t\right\}$$

which determines the decomposition $\prod_{0,t} = \prod_{0,\tau} \cdot \prod_{\tau,t}$. The convergence $\rho_s \xrightarrow{s\to\infty} 0$ yields the finiteness of $\tau < \infty$.

Now we analyse the logarithm of $\prod_{\tau,t}$. Note that $\prod_{\tau,t}$ consists of positive factors only. As $\tau \geqslant \tau^1$ we have $\frac{\Delta R_s}{1+R_{s-}} < \frac{1}{\tilde{a}+1}$ for all $s > \tau$. Together with $\tilde{a}\Delta R_s \geqslant 0$ this means $\tilde{a}\Delta R_s \neq (1+R_{s-})$ for all $s > \tau$. Bringing these ideas together yields

$$\begin{aligned}
\ln\left(\prod\nolimits_{\tau,t}\right) &= \sum_{\tau<s\leqslant t} \ln\left(\left(1 - \frac{\tilde{a}\Delta R_s}{1+R_{s-}} + \mathbb{1}_{\{\tilde{a}\Delta R_s = (1+R_{s-})\}}\right)\left(1 + \frac{\Delta R_s}{1+R_{s-}}\right)^{\tilde{a}}\right) \\
&= \sum_{\tau<s\leqslant t} \ln\left(\left(1 - \frac{\tilde{a}\Delta R_s}{1+R_{s-}}\right)\left(1 + \frac{\Delta R_s}{1+R_{s-}}\right)^{\tilde{a}}\right) \\
&= \sum_{\tau<s\leqslant t} \left(\ln\left(1 - \frac{\tilde{a}\Delta R_s}{1+R_{s-}}\right) + \tilde{a}\ln\left(1 + \frac{\Delta R_s}{1+R_{s-}}\right)\right) \\
&= \sum_{\tau<s\leqslant t} \left(-\frac{\tilde{a}}{2} - \frac{\tilde{a}^2}{2} + \rho_s\right)\left(\frac{\Delta R_s}{1+R_{s-}}\right)^2,
\end{aligned}$$

where the second equation follows from $\frac{\Delta R_s}{1+R_{s-}} < \frac{1}{\tilde{a}+1}$, and the last one from the Taylor expansion above.

Now $\rho_s \to 0$, $|\rho_s| < 1$ and

$$\sum_{0<s\leqslant\infty} \left(\frac{\Delta R_s}{1+R_{s-}}\right)^2 = \frac{1}{\tilde{a}^2} \sum_{0<s\leqslant\infty} \tilde{a}_s^2 (\Delta R_s)^2 = \frac{1}{\tilde{a}^2} \int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d < \infty$$

imply convergence of $\ln(\prod_{\tau,t})$ and $\prod_{\tau,t}$, respectively.

Finally we prove boundedness of $|\prod_{0,\tau}|$. We just showed that

$$\prod\nolimits_{0,\tau} = \prod_{0<s\leqslant\tau}\left(1 - \frac{\tilde{a}\Delta R_s}{1+R_{s-}} + \mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}\right)\left(\prod_{0<s\leqslant\tau}\left(1+\frac{\Delta R_s}{1+R_{s-}}\right)\right)^{\tilde{a}}$$

holds. The second factor is bounded as we have

$$\ln\left(\prod_{0<s\leqslant\tau}\left(1+\frac{\Delta R_s}{1+R_{s-}}\right)\right) = \sum_{0<s\leqslant\tau}\ln\left(1+\frac{\Delta R_s}{1+R_{s-}}\right) \leqslant \sum_{0<s\leqslant\tau}\frac{\Delta R_s}{1+R_{s-}} \leqslant \sum_{0<s\leqslant\tau}\Delta R_s$$
$$\leqslant R_\tau^d < \infty.$$

In the first factor we can neglect the term $\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}$ as the set

$$\{s \in \mathbb{R}_+ : \tilde{a}\Delta R_s = (1+R_{s-})\}$$

is finite.

The remaining term $\prod_{0<s\leqslant\tau}\left(1-\frac{\tilde{a}\Delta R_s}{1+R_{s-}}\right)$ is split into two factors. The first factor consists of all factors with norm smaller than one, such that its norm is bounded by one, too. Consequently we are done, when we can show that the remaining factor has only a finite number of factors. The inequation

$$\left|1-\frac{\tilde{a}\Delta R_s}{1+R_{s-}}\right| \geqslant 1 \text{ implies } \Delta R_s \geqslant \frac{2}{\tilde{a}}.$$

The finiteness of the set

$$\left\{s \in \mathbb{R}_+ : s \leqslant \tau \wedge \Delta R_s \geqslant \frac{2}{\tilde{a}}\right\}$$

in turn is assured by $R_\tau^d < \infty$. This proves the convergence of $\prod_{0,t}$ to a limit denoted by $\prod_\infty$.

For the case $\tilde{\alpha} < 1$ we reuse the representation (8.20) as well as

$$-\tilde{a}\int_0^t (1+R_{s-})^{-\tilde{\alpha}}\mathrm{d}R_s$$
$$= -\frac{\tilde{a}}{1-\tilde{\alpha}}\left((1+R_t)^{1-\tilde{\alpha}} - 1 - \sum_{s\leqslant t}\left(\Delta(1+R_s)^{1-\tilde{\alpha}} - (1-\tilde{\alpha})(1+R_{s-})^{-\tilde{\alpha}}\Delta R_s\right)\right)$$

and

$$-\Delta\int_0^t \frac{\bar{a}^s}{(1+R_{s-})^{\tilde{\alpha}}}\mathrm{d}R_s = -\tilde{a}\Delta\int_0^t (1+R_{s-})^{-\tilde{\alpha}}\mathrm{d}R_s + \Delta\sum_{0<s\leqslant t}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{-\tilde{\alpha}}\}}$$
$$= -\frac{\tilde{a}\Delta R_t}{(1+R_{t-})^{\tilde{\alpha}}\}} + \mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{-\tilde{\alpha}}\}}$$

in order to calculate the stochastic exponential $\mathcal{E}_t(X) = e^{X_t - X_0 - \frac{1}{2}[X,X]_t^c}\prod_{s\leqslant t}(1+$

$\Delta X_s)e^{-\Delta X_s}$. This leads to the representation

$$\begin{aligned}
\phi_t &= \mathcal{E}_t\left(-\int_0^\cdot \frac{\bar{a}^s}{(1+R_{s-})^{\tilde{\alpha}}}\mathrm{d}R_s\right)\\
&= \exp\Bigg(-\frac{\tilde{a}}{1-\tilde{\alpha}}\bigg((1+R_t)^{1-\tilde{\alpha}}-1\\
&\quad -\sum_{0<s\leqslant t}(\Delta(1+R_s)^{1-\tilde{\alpha}}-(1-\tilde{\alpha})(1+R_{s-})^{-\tilde{\alpha}}\Delta R_s)\bigg)+\sum_{0<s\leqslant t}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}\Bigg)\\
&\quad\cdot\prod_{0<s\leqslant t}\left(1-\frac{\tilde{a}\Delta R_s}{(1+R_{s-})^{\tilde{\alpha}}}+\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}\right)\\
&\quad\cdot\exp\left(\frac{\tilde{a}\Delta R_s}{(1+R_{s-})^{\tilde{\alpha}}}-\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}\right)\\
&= \exp\left(-\frac{\tilde{a}}{1-\tilde{\alpha}}(1+R_t)^{1-\tilde{\alpha}}\right)\exp\left(\frac{\tilde{a}}{1-\tilde{\alpha}}\left(1+\sum_{0<s\leqslant t}\Delta(1+R_s)^{1-\tilde{\alpha}}\right)\right)\\
&\quad\cdot\prod_{0<s\leqslant t}\left(1-\frac{\tilde{a}\Delta R_s}{(1+R_{s-})^{\tilde{\alpha}}}+\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}\right)\\
&= \exp\left(-\frac{\tilde{a}}{1-\tilde{\alpha}}(1+R_t)^{1-\tilde{\alpha}}\right)\prod\nolimits_{0,t}
\end{aligned}$$

with

$$\prod\nolimits_{0,t} = e^{\frac{\tilde{a}}{1-\tilde{\alpha}}}\prod_{0<s\leqslant t}\left(\left(1-\frac{\tilde{a}\Delta R_s}{(1+R_{s-})^{\tilde{\alpha}}}+\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}\right)\exp\left(\frac{\tilde{a}}{1-\tilde{\alpha}}\Delta(1+R_s)^{1-\tilde{\alpha}}\right)\right).$$

Finally we show that $\prod_{0,t}$ converges as $t\to\infty$. Firstly we observe

$$\int_0^\infty \tilde{a}_s\mathrm{d}R_s=\infty \quad\text{and}\quad \int_0^\infty \tilde{a}_s^2\Delta R_s\mathrm{d}R_s^d<\infty \quad\text{imply}\quad \tilde{a}_s\Delta R_s=\frac{\tilde{a}\Delta R_s}{(1+R_s)^{\tilde{\alpha}}}\xrightarrow{s\to\infty}0.$$

Next, we make use of the well-known Taylor expansion $\ln(1-x)=-\sum_{n=1}^\infty\frac{x^n}{n}$, if $|x|<1$. As $\left(\frac{\tilde{a}\Delta R_t}{(1+R_t)^{\tilde{\alpha}}}\right)\geqslant 0$, this yields

$$\ln\left(1-\frac{\tilde{a}\Delta R_t}{(1+R_{t-})^{\tilde{\alpha}}}\right)=-\sum_{n=1}^\infty\frac{1}{n}\cdot\left(\frac{\tilde{a}\Delta R_t}{(1+R_{t-})^{\tilde{\alpha}}}\right)^n\leqslant-\frac{\tilde{a}\Delta R_t}{(1+R_{t-})^{\tilde{\alpha}}}-\frac{1}{2}\left(\frac{\tilde{a}\Delta R_t}{(1+R_{t-})^{\tilde{\alpha}}}\right)^2$$

for $t$ sufficiently large. Consequently, with the previous inequation and a Taylor expansion around $R_{s-}$, for $s$ sufficiently large

$$\begin{aligned}
&\ln\prod_{s\leqslant t}\left(\exp\left(\frac{\tilde{a}}{1-\tilde{\alpha}}\Delta(1+R_s)^{1-\tilde{\alpha}}\right)\left(1-\frac{\tilde{a}\Delta R_s}{(1+R_{s-})^{\tilde{\alpha}}}\right)\right)\\
&\qquad=\sum_{s\leqslant t}\left(\frac{\tilde{a}}{1-\tilde{\alpha}}\Delta(1+R_s)^{1-\tilde{\alpha}}+\ln\left(1-\frac{\tilde{a}\Delta R_s}{(1+R_{s-})^{\tilde{\alpha}}}\right)\right)
\end{aligned}$$

$$\leqslant \sum_{s\leqslant t}\left(\frac{\tilde a}{1-\tilde\alpha}\Delta(1+R_s)^{1-\tilde\alpha} - \frac{\tilde a\Delta R_s}{(1+R_{s-})^{\tilde\alpha}} - \frac{1}{2}\left(\frac{\tilde a\Delta R_s}{(1+R_{s-})^{\tilde\alpha}}\right)^2\right)$$
$$\leqslant \sum_{s\leqslant t}\left(\rho_s - \frac{\tilde a^2}{2}\right)\left(\frac{\Delta R_s}{(1+R_{s-})^{\tilde\alpha}}\right)^2,$$

with $\rho_s \to 0$, converges. The rest of the proof follows in an analogous way to the case $\tilde\alpha = 1$.

Finally from $|\prod_{0,t}| \leqslant \mathcal{C} < \infty$ and $\prod_{0,t} \to \prod_\infty$ follows

$$\prod_{s,t} := \frac{\prod_{0,t}}{\prod_{0,s}} = \big(1 + o_{\mathrm{b}}(1)\big) \quad s \to t \text{ pointwise.}$$

□

## 8.3 General Distribution Results

Having results on the almost $L^2$-convergence rate of the leading algorithms, we are prepared to formulate a theorem on asymptotic normality of the companion algorithms.

**Assumption 8.3.1.** *Assume* $\sum_{0\leqslant s}\mathbb{1}_{\{\tilde a\Delta R_s=(1+R_{s-})^{\tilde\alpha}\}} < \infty$, $\int_0^\infty \tilde a_s \mathrm{d}R_s = \infty$, $\int_0^\infty \tilde a_s^2 \Delta R_s \mathrm{d}R_s^d < \infty$,

$$h(v^*) := \lim_{\substack{s\to\infty\\ y\to v^*}} h_s(y) = \lim_{\substack{s\to\infty\\ y\to v^*}} h_s(y, v^*)$$

*where*

$$h_s(y_1,y_2) := \frac{\mathrm{d}[\int_0^\cdot M(\mathrm{d}t,y_1), \int_0^\cdot M(\mathrm{d}t,y_2)]_s}{\mathrm{d}R_s}, \quad h_s(y) := h_s(y,y),$$

*and, for all* $\epsilon \in (0,1]$, *the Lindeberg type condition*

$$\frac{\int_0^t \frac{(1+R_s)^{2\tilde a}}{(1+R_{s-})^{2\kappa}} \int_{\mathfrak{G}^\epsilon_{s,t}} x^T x \nu^{M^*}(\mathrm{d}s,\mathrm{d}x)}{(1+R_t)^{2(\tilde a-\beta)}} \xrightarrow{\mathbb{P}} 0 \quad \textit{as } t\to\infty \quad \textit{if } \tilde\alpha = 1$$
$$\frac{\int_0^t (1+R_{s-})^{-2\kappa}\phi_s^{-2} \int_{\mathfrak{G}^\epsilon_{s,t}} x^T x \nu^{M^*}(\mathrm{d}s,\mathrm{d}x)}{(1+R_t)^{-2\beta}\phi_t^{-2}} \xrightarrow{\mathbb{P}} 0 \quad \textit{as } t\to\infty \quad \textit{if } \tilde\alpha < 1,$$

*where* $M_t^* := \int_0^t M(\mathrm{d}s, v^*)$,

$$\mathfrak{G}^\epsilon_{s,t} := \begin{cases} \left\{x \in \mathbb{R}^d \,\middle|\, \|x\| > \epsilon \dfrac{(1+R_t)^{\tilde a-\beta}}{(1+R_{st-})^{-\kappa}(1+R_{st})^{\tilde a}}\right\} & \textit{if } \tilde\alpha = 1 \\ \left\{x \in \mathbb{R}^d \,\middle|\, \|x\| > \epsilon \dfrac{(1+R_t)^{-\beta}}{(1+R_{st-})^{-\kappa}}\phi_t^{-1}\phi_{st}\right\} & \textit{if } \tilde\alpha < 1, \end{cases}$$

*and $\nu^{M^*}$ is the compensator of the jump-measure $\mu^{M^*}$ of the local martingale $M^*$. If $\tilde{\alpha} = 1$ assume $\tilde{a} > 1/2$. If $\tilde{\alpha} < 1$ assume $\tilde{a} > 0$.*

*The process $Z$ converges to $z^*$ in the almost $L^2$ sense with rate $(1 + R_{s-})^{-p}$. There are constants $l, \iota, \nu, \chi > 0$ such that $G_s$ can be decomposed as*

$$G_s = l_s + \mathcal{O}(\|Z_{s-} - z^*\|^{\chi}) + o(m_s),$$

*with $l_s := l(1 + R_{s-})^{-\iota}$ and $m_s := (1 + R_{s-})^{-\eta}$. The constant $\beta$ satisfies $\beta \leqslant \frac{p}{2}\chi$, $\beta \leqslant \iota$ and $\beta \leqslant \eta$.*

*Remark* 8.3.1. In preceding representation of $G_s$, the term $l_s$ causes a bias term in asymptotic normality results, whereas $m_s$ vanishes asymptotically. The condition $\beta \leqslant \frac{p}{2}\chi$ guarantees, that the almost $L^2$-convergence rate of $Z$ is fast enough to still achieve convergence of the disturbed process. For a leading Robbins-Monro process with differentiable $f$ typically $p = 1/2$ holds, whereas in a Kiefer-Wolfowitz process only $p = 1/4$ or $p = 1/3$ for $f \in C^2$ or $f \in C^3$, respectively, can be achieved.

Now we are ready to state an asymptotic normality theorem about a generic companion algorithm.

**Theorem 8.3.1.** *Let Assumption 8.3.1 hold. Then*

$$\begin{aligned}
&(1 + R_t)^{\beta}(\Upsilon_t - \upsilon^*) \xrightarrow{\mathcal{D}} \mathrm{N}(\mu, \Sigma) && \text{as } t \to \infty \text{ if } \beta = \frac{2\kappa - 1}{2}\\
&(1 + R_t)^{\beta}(\Upsilon_t - \upsilon^*) \xrightarrow{\mathbb{P}} \mu && \text{as } t \to \infty \text{ if } \beta < \frac{2\kappa - 1}{2}.
\end{aligned}$$

*Bias $\mu$ and variance $\Sigma$ are given by*

$$\mu = \begin{cases} \dfrac{\tilde{a} l}{\tilde{a} - \iota} & \text{if } \tilde{\alpha} = 1 \text{ and } \beta = \iota \\ l & \text{if } \tilde{\alpha} < 1 \text{ and } \beta = \iota \\ 0 & \text{if } \tilde{\alpha} \leqslant 1 \text{ and } \beta < \iota \end{cases}$$

*and*

$$\Sigma = \begin{cases} \dfrac{k^2}{2(\tilde{a} - \kappa) + 1} h(\upsilon^*) & \text{if } \beta = \frac{2\kappa - 1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{k^2}{2\tilde{a}} h(\upsilon^*) & \text{if } \beta = \frac{2\kappa - 1}{2} \text{ and } \tilde{\alpha} < 1. \end{cases}$$

The previous theorem employs a condition of Lindeberg type, which uses the jump measure of the compensator explicitly. The following corollary demands conditions (S1) and (S2) that are an alternative, which is easier to interpret, to the Lindeberg type condition.

**Corollary 8.3.1.** *If we replace condition*

$$\frac{\int_0^t \frac{(1+R_s)^{2\tilde{a}}}{(1+R_{s-})^{2\kappa}} \int_{\mathfrak{G}^{\epsilon}_{s,t}} x^T x \nu^{M^*}(\mathrm{d}s, \mathrm{d}x)}{(1 + R_t)^{2(\tilde{a} - \beta)}} \xrightarrow{\mathbb{P}} 0 \quad \text{as } t \to \infty \quad \text{if } \tilde{\alpha} = 1$$

$$\frac{\int_0^t (1+R_{s-})^{-2\kappa}\phi_s^{-2}\int_{\mathfrak{G}^\epsilon_{s,t}} x^T x \nu^{M^*}(\mathrm{d}s,\mathrm{d}x)}{(1+R_t)^{-2\beta}\phi_t^{-2}} \xrightarrow{\mathbb{P}} 0 \quad \textit{as } t\to\infty \quad \textit{if } \tilde{\alpha}<1,$$

*in Assumption 8.3.1 by the following conditions*

*(S1)*

$$\frac{\mathbb{E}\sup_{0\leqslant s\leqslant t}(1+R_{s-})^{2(\tilde{a}-\kappa)}(\Delta M_s^*)^2}{(1+R_t)^{2(\tilde{a}-\kappa)+1}} \xrightarrow{t\to\infty} 0 \qquad \textit{if } \tilde{\alpha}=1$$

$$\frac{\mathbb{E}\sup_{0\leqslant s\leqslant t}(1+R_{s-})^{-2\kappa}\phi_{s-}^{-2}(\Delta M_s^*)^2}{(1+R_t)^{-2\kappa+1}\phi_t^{-2}} \xrightarrow{t\to\infty} 0 \qquad \textit{if } \tilde{\alpha}<1$$

*(S2)*

$$\frac{\mathbb{E}\sum_{s\leqslant t}(\Delta M_s^*)^4(1+R_{s-})^{4(\tilde{a}-\kappa)}}{(1+R_t)^{4(\tilde{a}-\kappa)+2}} \xrightarrow{t\to\infty} 0 \qquad \textit{if } \tilde{\alpha}=1$$

$$\frac{\mathbb{E}\sum_{s\leqslant t}(\Delta M_s^*)^4(1+R_{s-})^{-4\kappa}\phi_{s-}^{-4}}{(1+R_t)^{-4\kappa+2}\phi_t^{-4}} \xrightarrow{t\to\infty} 0 \qquad \textit{if } \tilde{\alpha}<1$$

*then the conclusion of Theorem 8.3.1 still holds true.*

*Proof of Theorem 8.3.1.* Without loss of generality we assume $\upsilon^*=0$. Now we analyse $(1+R_t)^\beta\Upsilon_t$, with $\Upsilon$ given in (8.14). Due to Lemma 8.2.1 and Slutsky's theorem it is sufficient to show

(I)

$$(1+R_t)^\beta\phi_t\Upsilon_0 \xrightarrow{\mathbb{P}} 0$$

(II)

$$(1+R_t)^\beta\phi_t\int_0^t \phi_s^{-1}\mathrm{d}\tilde{R}_s \xrightarrow{\mathbb{P}} \mu$$

(III)

$$(1+R_t)^\beta\phi_t k\int_0^t (1+R_{s-})^{-\kappa}\phi_s^{-1}M(\mathrm{d}s,0) \xrightarrow{\mathcal{D}} \mathrm{N}(0,\Sigma)$$

with $\phi$ and $\tilde{R}$ defined in Lemma 8.2.1.

**Verification of (I).** In all settings $\beta\leqslant 1/2$ holds true. If $\tilde{\alpha}<1$

$$(1+R_t)^\beta\exp\Big(-\frac{\tilde{a}}{1-\tilde{\alpha}}(1+R_t)^{1-\tilde{\alpha}}\Big)\underbrace{\prod\nolimits_{0,t}\Upsilon_0}_{|.|\leqslant C} \xrightarrow{\mathbb{P}} 0$$

follows directly. When $\tilde{\alpha} = 1$ we furthermore assumed $\tilde{a} > 1/2$. Therefore $(1 + R_t)^{\beta}(1 + R_t)^{-\tilde{a}} \to 0$ yields $(1 + R_t)^{\beta}\phi_t\Upsilon_0 \xrightarrow{\mathbb{P}} 0$, as $\beta - \tilde{a} < 0$.

**Verification of (II).** According to Lemma 8.2.1

$$\phi_t = \begin{cases} \exp\Big(-\frac{\tilde{a}}{1-\tilde{\alpha}}(1 + R_t)^{1-\tilde{\alpha}}\Big)\prod_{0,t} & \text{if } \tilde{\alpha} < 1 \\ (1 + R_t)^{-\tilde{a}}\prod_{0,t} & \text{if } \tilde{\alpha} = 1 \end{cases}$$

holds true with $\prod_{0,t}$ given in (8.15) or (8.16). We use the abbreviation

$$\bar{\phi}_t := \exp\Big(-\frac{\tilde{a}}{1-\tilde{\alpha}}(1 + R_t)^{1-\tilde{\alpha}}\Big) \quad \text{for } \tilde{\alpha} < 1.$$

The properties $\left|\prod_{0,t}\right| \leqslant \mathcal{C} < \infty$ and $\prod_{0,t} \to \prod_{\infty}$ imply $\frac{\prod_{0,t}}{\prod_{0,s}} = \prod_{s,t} = (1 + o_{\mathrm{b}}(1))$ for $s \to t$. According to the dominated convergence theorem, instead of

$$(1 + R_t)^{\beta}\phi_t\int_0^t \phi_s^{-1}\mathrm{d}\tilde{R}_s \xrightarrow{\mathbb{P}} \mu$$

it is sufficient to show

$$\begin{cases} (1 + R_t)^{\beta-\tilde{a}}\int_0^t(1 + R_s)^{\tilde{a}}\mathrm{d}\tilde{R}_s \xrightarrow{\mathbb{P}} \mu & \text{if } \tilde{\alpha} = 1 \\ (1 + R_t)^{\beta}\bar{\phi}_t\int_0^t \bar{\phi}_s^{-1}\mathrm{d}\tilde{R}_s \xrightarrow{\mathbb{P}} \mu & \text{if } \tilde{\alpha} < 1. \end{cases}$$

Therefore our problem reduces to show

$$(1 + R_t)^{\beta}\phi_t\int_0^t \phi_s^{-1}\mathrm{d}\tilde{R}_s \simeq H_t^1 + H_t^2 + H_t^3 \xrightarrow{\mathbb{P}} \mu$$

where

$$H_t^1 = \begin{cases} (1 + R_t)^{\beta-\tilde{a}}\displaystyle\int_0^t \tilde{a}G_s(1 + R_{s-})^{\tilde{a}-1}\mathrm{d}R_s & \text{if } \tilde{\alpha} = 1 \\ (1 + R_t)^{\beta}\bar{\phi}_t\displaystyle\int_0^t \tilde{a}G_s(1 + R_{s-})^{-\tilde{\alpha}}\bar{\phi}_{s-}^{-1}\mathrm{d}R_s & \text{if } \tilde{\alpha} < 1 \end{cases}$$

$$H_t^2 = \begin{cases} (1 + R_t)^{\beta-\tilde{a}}\displaystyle\sum_{0<s\leqslant t}(1 + R_{s-})^{\tilde{a}}\Upsilon_{s-}\mathbb{1}_{\{\tilde{a}\Delta R_s = (1+R_{s-})\}} & \text{if } \tilde{\alpha} = 1 \\ (1 + R_t)^{\beta}\bar{\phi}_t\displaystyle\sum_{0<s\leqslant t}\bar{\phi}_{s-}^{-1}\Upsilon_{s-}\mathbb{1}_{\{\tilde{a}\Delta R_s = (1+R_{s-})^{\tilde{\alpha}}\}} & \text{if } \tilde{\alpha} < 1 \end{cases}$$

$$H_t^3 = \begin{cases} (1 + R_t)^{\beta-\tilde{a}}\displaystyle\int_0^t k(1 + R_{s-})^{\tilde{a}-\kappa}\left(M(\mathrm{d}s, \Upsilon_{s-}) - M(\mathrm{d}s, 0)\right) & \text{if } \tilde{\alpha} = 1 \\ (1 + R_t)^{\beta}\bar{\phi}_t\displaystyle\int_0^t \bar{\phi}_{s-}^{-1}k(1 + R_{s-})^{-\kappa}\left(M(\mathrm{d}s, \Upsilon_{s-}) - M(\mathrm{d}s, 0)\right) & \text{if } \tilde{\alpha} < 1. \end{cases}$$

The asymptotic equality of $(1 + R_{s-})^{\tilde{a}}$ and $(1 + R_s)^{\tilde{a}}$ follows as $\int_0^{\infty}\tilde{a}_s\mathrm{d}R_s = \infty$ and

$\int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s = \int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d < \infty$ imply

$$\frac{\Delta R_s}{1 + R_{s-}} \to 0. \tag{8.21}$$

Actually we have

$$(1+R_s)^{\tilde{a}} = \left(\frac{1+R_s}{1+R_{s-}}\right)^{\tilde{a}} (1+R_{s-})^{\tilde{a}} = \underbrace{\left(1 + \frac{\Delta R_s}{1+R_{s-}}\right)^{\tilde{a}}}_{\xrightarrow{s\to\infty} 1} (1+R_{s-})^{\tilde{a}}.$$

Analogous arguments yield the asymptotic equality of $\bar{\phi}_{s-}$ and $\bar{\phi}_s$. Consequently we show:

$$\text{(II.1)} \quad H_t^1 \xrightarrow{\mathbb{P}} \mu \qquad\qquad \text{(II.2)} \quad H_t^2 \xrightarrow{\mathbb{P}} 0 \qquad\qquad \text{(II.3)} \quad H_t^3 \xrightarrow{\mathbb{P}} 0$$

**Verification of (II.1).** With the notation

$$G_s = l_s + \mathcal{O}(\|Z_{s-}\|^\chi) + o(m_s) =: l_s + V_s$$

we can split (II.1) into the tasks
(II.1.1)

$$(1+R_t)^{\beta-\tilde{a}} \int_0^t \tilde{a} V_s (1+R_{s-})^{\tilde{a}-1} \mathrm{d}R_s \xrightarrow{\mathbb{P}} 0 \quad \text{if } \tilde{\alpha} = 1$$
$$(1+R_t)^{\beta} \bar{\phi}_t \int_0^t \tilde{a} V_s (1+R_{s-})^{-\tilde{\alpha}} \bar{\phi}_{s-}^{-1} \mathrm{d}R_s \xrightarrow{\mathbb{P}} 0 \quad \text{if } \tilde{\alpha} < 1,$$

as well as
(II.1.2)

$$(1+R_t)^{\beta-\tilde{a}} \int_0^t \tilde{a} l_s (1+R_{s-})^{\tilde{a}-1} \mathrm{d}R_s \xrightarrow{\mathbb{P}} \begin{cases} \dfrac{\tilde{a} l}{\tilde{a} - \iota} & \text{if } \beta = \iota \\ 0 & \text{if } \beta < \iota \end{cases}$$

for $\tilde{\alpha} = 1$ and

$$(1+R_t)^{\beta} \bar{\phi}_t \int_0^t \tilde{a} l_s (1+R_{s-})^{-\tilde{\alpha}} \bar{\phi}_{s-}^{-1} \mathrm{d}R_s \xrightarrow{\mathbb{P}} \begin{cases} l & \text{if } \beta = \iota \\ 0 & \text{if } \beta < \iota \end{cases}$$

for $\tilde{\alpha} < 1$.

In the following steps we use the abbreviation

$$Q_{s,t} := \begin{cases} \dfrac{(1+R_{s-})^{\tilde{a}-1}}{(1+R_t)^{\tilde{a}-\beta}} & \text{if } \tilde{\alpha} = 1 \\ \dfrac{(1+R_{s-})^{-\tilde{\alpha}} \bar{\phi}_{s-}^{-1}}{(1+R_t)^{-\beta} \bar{\phi}_t^{-1}} & \text{if } \tilde{\alpha} < 1. \end{cases}$$

**Verification of (II.1.1).** Now we are prepared to verify

$$\underset{\epsilon_1>0}{\forall}\ \underset{\epsilon_2>0}{\forall}\ \underset{t_0}{\exists}\ \underset{t\geqslant t_0}{\forall}\ \mathbb{P}\left[\left|\int_0^t Q_{s,t}V_s\mathrm{d}R_s\right| > \epsilon_1\right] \leqslant \epsilon_2$$

with arbitrary but fixed $\epsilon_1, \epsilon_2 > 0$.

According to the almost $L^2$-convergence condition in Assumption 8.3.1 there exist a process $Y$ and a deterministic time $T_1 < \infty$ such that for all $t \geqslant T_1$ we have

$$\Big(\mathbb{E}\|Y_t\|\Big)^2 \leqslant \mathbb{E}\|Y_t\|^2 \leqslant K(1+R_t)^{-p} \quad \text{and} \quad \mathbb{P}\left(\left[\underset{t\geqslant 0}{\forall}\ Y_t = Z_t\right]\right) \geqslant 1 - \frac{\epsilon_1}{8}.$$

By definition of $V_s$ it holds

$$V_s = \mathcal{O}(\|Z_{s-}\|^\chi) + o(m_s).$$

Hence we have

$$\underset{\rho>0}{\exists}\ \underset{\|Z_{s-}\|,\, m_s\leqslant\rho}{\forall} |V_s| \leqslant A\|Z_{s-}\|^\chi + Bm_s$$

with appropriate choices for the constants $A$ and $B$.

As the conditions of Lemma 8.1.1 hold, there exists a deterministic $T_2 < \infty$ such that

$$\mathbb{P}\left(\left[\sup_{t\geqslant T_2}\|Z_t\| \leqslant \rho\right]\right) \geqslant 1 - \frac{\epsilon_2}{4}.$$

Furthermore there exists a deterministic $T_3 < \infty$ such that for all $t > T_3$ it holds $m < \rho$. We define $T := \max\{T_1, T_2, T_3\}$. Consequently it remains to show

$$\begin{aligned}
&\mathbb{P}\left(\left[\left|\int_0^T Q_{s,t}V_s\mathrm{d}R_s + \int_{T+}^t Q_{s,t}V_s\mathrm{d}R_s\right| > \epsilon_1\right]\right)\\
&\leqslant \mathbb{P}\left(\left[\left|\int_0^T Q_{s,t}V_s\mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{2}\right] \cup \left[\left|\int_{T+}^t Q_{s,t}V_s\mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{2}\right]\right)\\
&\leqslant \underbrace{\mathbb{P}\left(\left[\left|\int_0^T Q_{s,t}V_s\mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{2}\right]\right)}_{\leqslant \frac{\epsilon_2}{2}} + \underbrace{\mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t}V_s\mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{2}\right]\right)}_{\leqslant \frac{\epsilon_2}{2}}.
\end{aligned} \tag{8.22}$$

We begin with the verification of the bound of the first summand in the last line of (8.22). As $(Z_t)_{t\geqslant 0}$ is a strong solution of the corresponding stochastic integral equation on $[0,\infty)$, no explosion times exist and $Z_t \to z^*$ a.s. Consequently for all $t > 0$ $\|Z_t\| \leqslant C(\omega)$ holds. Furthermore $m$ is bounded. Combining these arguments yields that there exists a constant $C(\omega)$ such that

$$|V_t| \leqslant C(\omega) \quad \text{for all } t \leqslant T.$$

Therefore

$$\left|\int_0^T Q_{s,t} V_s \mathrm{d}R_s\right| \leqslant \begin{cases} \underbrace{C(\omega)}_{<\infty} \underbrace{(1+R_t)^{\beta-\tilde{a}}}_{\to 0} \underbrace{\int_0^T (1+R_{s-})^{\tilde{a}-1} \mathrm{d}R_s}_{<\infty} & \text{if } \tilde{\alpha} = 1 \\ \underbrace{C(\omega)}_{<\infty} \underbrace{(1+R_t)^{\beta} \bar{\phi}_t}_{\to 0} \underbrace{\int_0^T (1+R_{s-})^{-\tilde{\alpha}} \bar{\phi}_{s-}^{-1} \mathrm{d}R_s}_{<\infty} & \text{if } \tilde{\alpha} < 1. \end{cases}$$

This yields almost sure convergence and therefore convergence in probability. Using this fact we find

$$\underset{t_0^1}{\exists} \underset{t \geqslant t_0^1}{\forall} \mathbb{P}\left[\left|\int_0^T Q_{s,t} V_s \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{2}\right] \leqslant \frac{\epsilon_2}{2}.$$

Therefore the first inequation in (8.22) is proven. We now turn to the second bound in the last equation of (8.22):

$$\begin{aligned} &\mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} V_s \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{2}\right]\right) \\ &\leqslant \mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} V_s \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{2}\right] \cap \left[\sup_{t \geqslant T} \|Z_t\| < \rho\right]\right) + \mathbb{P}\left(\left[\sup_{t \geqslant T} \|Z_t\| \geqslant \rho\right]\right) \\ &\leqslant \mathbb{P}\left(\left[\sup_{t \geqslant T} \|Z_t\| \geqslant \rho\right]\right) + \mathbb{P}\left(\left[\left|\int_{T+}^t Q_s \Big(A\|Z_{s-}\|^{\chi} + B m_s\Big) \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{2}\right]\right) \\ &\leqslant \frac{\epsilon_2}{4} + \mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} A\|Z_{s-}\|^{\chi} \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{4}\right]\right) + \mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} B m_s \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{4}\right]\right) \\ &= \frac{\epsilon_2}{4} + \mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} \|Z_{s-}\|^{\chi} \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{4A}\right]\right) + \mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} B m_s \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{4}\right]\right) \end{aligned} \tag{8.23}$$

We start with a bound for the second to last term in (8.23). We have

$$\begin{aligned} &\mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} \|Z_{s-}\|^{\chi} \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{4A}\right]\right) \\ &\leqslant \mathbb{P}\left(\left[\underset{t \geqslant 0}{\forall} Z_t = Y_t\right]^c\right) + \mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} \|Z_{s-}\|^{\chi} \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{4A}\right] \cap \left[\underset{t \geqslant 0}{\forall} Z_t = Y_t\right]\right) \\ &\leqslant \underbrace{\mathbb{P}\left(\left[\underset{t \geqslant 0}{\forall} Z_t = Y_t\right]^c\right)}_{\leqslant \frac{\epsilon_2}{8}} + \underbrace{\mathbb{P}\left(\left[\left|\int_{T+}^t Q_{s,t} \|Y_{s-}\|^{\chi} \mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{4A}\right]\right)}_{\leqslant \frac{\epsilon_2}{8}} \end{aligned}$$

where the first bound is a direct consequence of the assumed almost $L^2$-convergence rate. The second inequation is shown next. The bound

$$\mathbb{E}\|Y_t\|^2 \leqslant K(1+R_t)^{-p}$$

implies

$$\mathbb{E}\|Y_t\| \leqslant K^{\frac{1}{2}}(1+R_t)^{-\frac{p}{2}} \qquad \text{with } \chi > 0.$$

Using Markov's inequality and the latter result yields

$$\begin{aligned}
&\mathbb{P}\left[\left|\int_{T+}^{t} Q_{s,t}\|Y_{s-}\|^{\chi}\mathrm{d}R_s\right| \geqslant \frac{\epsilon_1}{4A}\right] \\
&\quad\leqslant \frac{4A}{\epsilon_1}\int_{T+}^{t} Q_{s,t} \underbrace{\mathbb{E}\|Y_{s-}\|^{\chi}}_{\leqslant K^{\frac{\chi}{2}}(1+R_{s-})^{-\frac{p}{2}\chi}} \mathrm{d}R_s \\
&\quad\leqslant \begin{cases} \dfrac{4A}{\epsilon_1}K^{\frac{\chi}{2}}(1+R_t)^{\beta-\tilde{a}}\displaystyle\int_0^t (1+R_{s-})^{\tilde{a}-1-\frac{p}{2}\chi}\mathrm{d}R_s & \text{if } \tilde{\alpha}=1 \\ \dfrac{4A}{\epsilon_1}K^{\frac{\chi}{2}}(1+R_t)^{\beta}\bar{\phi}_t\displaystyle\int_0^t \bar{\phi}_{s-}^{-1}(1+R_{s-})^{-\tilde{\alpha}-\frac{p}{2}\chi}\mathrm{d}R_s & \text{if } \tilde{\alpha}<1. \end{cases}
\end{aligned}$$

Application of Itô's formula and a Taylor expansion of $f\colon x \mapsto (1+x)^{\tilde{a}-\frac{p}{2}\chi}$ around $R_{s-}$ with a $\vartheta_s \in (0,1)$ yields

$$\begin{aligned}
&\int_0^t (1+R_{s-})^{\tilde{a}-1-\frac{p}{2}\chi}\mathrm{d}R_s \\
&\quad= \frac{(1+R_t)^{\tilde{a}-\frac{p}{2}\chi}-1}{\tilde{a}-\frac{p}{2}\chi} - \sum_{0<s\leqslant t}\left\{\frac{\Delta(1+R_s)^{\tilde{a}-\frac{p}{2}\chi}}{\tilde{a}-\frac{p}{2}\chi} - (1+R_{s-})^{\tilde{a}-1-\frac{p}{2}\chi}\Delta R_s\right\} \\
&\quad= \frac{(1+R_t)^{\tilde{a}-\frac{p}{2}\chi}-1}{\tilde{a}-\frac{p}{2}\chi} \\
&\qquad - \frac{1}{\tilde{a}-\frac{p}{2}\chi}\sum_{0<s\leqslant t}\frac{1}{2}\Big(\tilde{a}-\frac{p}{2}\chi\Big)\Big(\tilde{a}-1-\frac{p}{2}\chi\Big)(1+R_{s-}+\vartheta_s\Delta R_s)^{\tilde{a}-\frac{p}{2}\chi-2}(\Delta R_s)^2 \\
&\quad= \frac{(1+R_t)^{\tilde{a}-\frac{p}{2}\chi}-1}{\tilde{a}-\frac{p}{2}\chi} - \frac{1}{2}\Big(\tilde{a}-1-\frac{p}{2}\chi\Big)\sum_{0<s\leqslant t}(1+R_{s-}+\vartheta_s\Delta R_s)^{\tilde{a}-\frac{p}{2}\chi-2}(\Delta R_s)^2 \\
&\quad= \frac{(1+R_t)^{\tilde{a}-\frac{p}{2}\chi}-1}{\tilde{a}-\frac{p}{2}\chi} \\
&\qquad - \frac{1}{2}\Big(\tilde{a}-1-\frac{p}{2}\chi\Big)\sum_{0<s\leqslant t}(1+R_{s-}+\vartheta_s\Delta R_s)^{\tilde{a}}(1+R_{s-}+\vartheta_s\Delta R_s)^{-\frac{p}{2}\chi-2}(\Delta R_s)^2 \\
&\quad\leqslant \frac{(1+R_t)^{\tilde{a}-\frac{p}{2}\chi}-1}{\tilde{a}-\frac{p}{2}\chi} + \mathcal{C}\sum_{0<s\leqslant t}(1+R_s)^{\tilde{a}}(1+R_{s-})^{-2-\frac{p}{2}\chi}(\Delta R_s)^2 \\
&\quad\leqslant \frac{(1+R_t)^{\tilde{a}-\frac{p}{2}\chi}-1}{\tilde{a}-\frac{p}{2}\chi} + \mathcal{C}\sum_{0<s\leqslant t}\Bigg(\underbrace{\frac{1+R_s}{1+R_{s-}}}_{=1+\frac{\Delta R_s}{1+R_{s-}}=1+o_b(1)\text{ by (8.21)}}\Bigg)^{\tilde{a}}(1+R_{s-})^{\tilde{a}-2-\frac{p}{2}\chi}(\Delta R_s)^2.
\end{aligned}$$

Therefore, for $\tilde{\alpha} = 1$, choosing $A := \frac{\epsilon_2}{64}\epsilon_1(\tilde{a} - \frac{p}{2}\chi)K^{-\frac{\chi}{2}}$,

$$\begin{aligned}
&\frac{4AK^{\frac{\chi}{2}}}{\epsilon_1}\int_0^t \frac{(1+R_{s-})^{\tilde{a}-1-\frac{p}{2}\chi}}{(1+R_t)^{\tilde{a}-\beta}}\mathrm{d}R_s \\
&\qquad \leqslant \frac{4AK^{\frac{\chi}{2}}}{\epsilon_1(\tilde{a}-\frac{p}{2}\chi)}\underbrace{\left(1+R_t\right)^{\beta-\frac{p}{2}\chi}}_{=1 \text{ if } \beta=\frac{p}{2}\chi} + \mathcal{C}\underbrace{\sum_{0<s\leqslant t}\frac{(1+R_{s-})^{\tilde{a}-2-\frac{p}{2}\chi}}{(1+R_t)^{\tilde{a}-\beta}}(\Delta R_s)^2}_{\to 0} \\
&\qquad \to \begin{cases} \frac{\epsilon_2}{16} & \text{if } \beta = \frac{p}{2}\chi \\ 0 & \text{if } \beta < \frac{p}{2}\chi \end{cases}
\end{aligned}$$

holds. The convergence of the sum to zero follows by Kronecker's lemma as $\beta \leqslant \frac{p}{2}\chi$ implies

$$\sum_{0<s}\frac{(1+R_{s-})^{\tilde{a}-2-\frac{p}{2}\chi}}{(1+R_s)^{\tilde{a}-\beta}}(\Delta R_s)^2 \leqslant \mathcal{C}\int_0^\infty \tilde{a}_s^2 \Delta R_s \mathrm{d}R_s^d < \infty.$$

In an analogous way, for $\tilde{\alpha} < 1$, we can show that

$$\frac{4AK^{\frac{\chi}{2}}}{\epsilon_1}\int_0^t \frac{\bar{\phi}_{s-}^{-1}(1+R_{s-})^{-\tilde{a}-\frac{p}{2}\chi}}{\bar{\phi}_t^{-1}(1+R_t)^{\tilde{a}-\beta}}\mathrm{d}R_s \to \begin{cases} \frac{\epsilon_2}{16} & \text{if } \beta = \frac{p}{2}\chi \\ 0 & \text{if } \beta < \frac{p}{2}\chi. \end{cases}$$

Now, for $\tilde{\alpha} = 1$, we show that the last term in (8.23) equals zero. For $t$ large enough we have, using Itô's formula in an analogous way as above,

$$\begin{aligned}
&(1+R_t)^{\beta-\tilde{a}}\int_{T+}^t (1+R_{s-})^{\tilde{a}-1-\eta}\mathrm{d}R_s \\
&\qquad \leqslant \frac{1}{\tilde{a}-\eta}\underbrace{(1+R_t)^{\beta-\eta}}_{=1 \text{ if } \beta=\eta} + \mathcal{C}\underbrace{\sum_{0<s\leqslant t}(\Delta R_s)^2\frac{(1+R_{s-})^{\tilde{a}-2-\eta}}{(1+R_t)^{\tilde{a}-\beta}}}_{\xrightarrow{t\to\infty} 0 \text{ as above}} \\
&\qquad < \frac{2}{\tilde{a}-\eta} \qquad \text{if } \beta \leqslant \eta
\end{aligned}$$

and therefore, for sufficiently large $t$,

$$\mathbb{P}\left(\left[(1+R_t)^{\beta-\tilde{a}}\int_{T+}^t (1+R_{s-})^{\tilde{a}-1-\eta}\mathrm{d}R_s \geqslant \frac{2}{\tilde{a}-\eta}\right]\right) = 0 \text{ holds.}$$

In a similar way, for $\tilde{\alpha} < 1$ and $t$ sufficiently large, it is easy to show

$$\mathbb{P}\left(\left[(1+R_t)^{\beta}\bar{\phi}_t\int_{T+}^t \bar{\phi}_{s-}^{-1}(1+R_{s-})^{-\tilde{a}-\eta}\mathrm{d}R_s \geqslant \frac{2}{\tilde{a}}\right]\right) = 0, \text{ if } \beta \leqslant \eta.$$

**Verification of (II.1.2).** We have, for $\tilde{\alpha} = 1$,

$$\begin{aligned}(1+R_t)^{\beta-\tilde{a}} \int_0^t \tilde{a}l(1+R_{s-})^{-\iota}(1+R_{s-})^{\tilde{a}-1}\mathrm{d}R_s \\ &= \tilde{a}l(1+R_t)^{\beta-\tilde{a}} \int_0^t (1+R_{s-})^{\tilde{a}-1-\iota}\mathrm{d}R_s \\ &\simeq \frac{\tilde{a}l}{\tilde{a}-\iota}(1+R_t)^{\beta-\tilde{a}}(1+R_t)^{\tilde{a}-\iota} \\ &\xrightarrow{\mathbb{P}} \begin{cases} \dfrac{\tilde{a}l}{\tilde{a}-\iota} & \text{if } \beta = \iota \\ 0 & \text{if } \beta < \iota \end{cases}\end{aligned}$$

and analogously, for $\tilde{\alpha} < 1$,

$$(1+R_t)^{\beta}\bar{\phi}_t \int_0^t \tilde{a}l(1+R_{s-})^{-\iota}(1+R_{s-})^{-\tilde{\alpha}}\bar{\phi}_{s-}^{-1}\mathrm{d}R_s \xrightarrow{\mathbb{P}} \begin{cases} l & \text{if } \beta = \iota \\ 0 & \text{if } \beta < \iota. \end{cases}$$

**Verification of (II.2).** We have assumed that $\sum_{s\leqslant t} \mathbb{1}_{\{\tilde{a}\Delta R_s = (1+R_{s-})^{\tilde{\alpha}}\}} < \infty$. Moreover we know that $(\Upsilon_t)_{t\geqslant 0}$ converges and that there are no explosion times. Therefore, if $\tilde{\alpha} = 1$,

$$\begin{aligned}\Big|(1+R_t)^{\beta-\tilde{a}} \sum_{0<s\leqslant t} (1+R_{s-})^{\tilde{a}}\Upsilon_{s-}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}\Big| \\ &\leqslant \mathcal{C}(\omega)\underbrace{(1+R_t)^{\beta-\tilde{a}}}_{\xrightarrow{t\to\infty}0}\underbrace{\sum_{s\leqslant t}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}}}_{<\infty} \xrightarrow{t\to\infty} 0.\end{aligned}$$

If $\tilde{\alpha} < 1$, one can show that

$$\begin{aligned}\Big|(1+R_t)^{\beta}\bar{\phi}_t \sum_{0<s\leqslant t} \bar{\phi}_{s-}^{-1}\Upsilon_{s-}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}}\Big| \\ &\leqslant \mathcal{C}(\omega)(1+R_t)^{\beta}\bar{\phi}_t\sum_{s\leqslant t}\mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})^{\tilde{\alpha}}\}} \xrightarrow{t\to\infty} 0.\end{aligned}$$

**Verification of (II.3).** Now we prove

$$(1+R_t)^{\beta-\tilde{a}} \int_0^t k(1+R_{s-})^{\tilde{a}-\kappa}\left(M(\mathrm{d}s,\Upsilon_{s-}) - M(\mathrm{d}s,0)\right) \xrightarrow{\mathbb{P}} 0 \quad \text{if } \tilde{\alpha} = 1.$$

We apply the Lenglart-Rebolledo inequality, which can be found as Theorem A.1.4 in the appendix. For that purpose choose

$$X_t := \int_0^t \frac{M(\mathrm{d}s,\Upsilon_{s-}) - M(\mathrm{d}s,0)}{(1+R_{s-})^{\kappa-\tilde{a}}} \qquad \text{and} \qquad Y_t := \lceil X \rceil_t.$$

Here $Y_t$ is the predictable compensator of $X_t^2$ and therefore $X_t^2 - Y_t \in \mathcal{M}_{\mathrm{loc}}$. For every stopping time $\tau$ we know according to Theorem A.1.5 that $\mathbb{E}X_\tau^2 = \mathbb{E}Y_\tau$. With arbitrary $\epsilon_1$, $\epsilon_2$ and fixed $t$ large enough, the Lenglart-Rebolledo inequality yields

$$\begin{aligned}
&\mathbb{P}\Big(\Big[\Big|(1+R_t)^{\beta-\tilde{a}}\int_0^t (1+R_s)^{\tilde{a}-\kappa}\left(M(\mathrm{d}s,\Upsilon_{s-})-M(\mathrm{d}s,0)\right)\Big|>\epsilon_1\Big]\Big)\\
&\quad=\mathbb{P}\Big(\Big[\Big|(1+R_t)^{\beta-\tilde{a}}X_t\Big|>\epsilon_1\Big]\Big)=\mathbb{P}\Big(\Big[X_t^2>\epsilon_1^2(1+R_t)^{2(\tilde{a}-\beta)}\Big]\Big)\\
&\quad\leqslant\mathbb{P}\Big(\Big[\sup_{s\leqslant t}X_s^2>\epsilon_1^2(1+R_t)^{2(\tilde{a}-\beta)}\Big]\Big)\leqslant\frac{\mathbb{E}(Y_t\wedge b)}{\epsilon_1^2(1+R_t)^{2(\tilde{a}-\beta)}}+\mathbb{P}\big([Y_t\geqslant b]\big)\\
&\quad\leqslant\frac{b}{\epsilon_1^2(1+R_t)^{2(\tilde{a}-\beta)}}+\mathbb{P}\big([Y_t\geqslant b]\big)
\end{aligned}\tag{8.24}$$

for any $b>0$. As $(1+R_t)^{2(\beta-\tilde{\alpha})}\xrightarrow{t\to\infty}\infty$, the first term in line (8.24) tends to zero. Now with Toeplitz's lemma, $h_s(\Upsilon_{s-},\Upsilon_{s-})-2h_s(\Upsilon_{s-},0)+h_s(0,0)\xrightarrow{s\to\infty}0$ implies

$$\begin{aligned}
&(1+R_t)^{-2(\beta-\tilde{\alpha})}Y_t\\
&\quad=(1+R_t)^{-2(\beta-\tilde{\alpha})}\lceil X\rceil_t=\frac{\displaystyle\int_0^t\frac{\mathrm{d}\big[\int_0^{\cdot}(M(\mathrm{d}\tau,\Upsilon_{\tau-})-M(\mathrm{d}\tau,0))\big]_s}{(1+R_{s-})^{2(\kappa-\tilde{a})}}}{(1+R_t)^{2(\tilde{a}-\beta)}}\\
&\quad=\frac{\displaystyle\int_0^t\frac{h_s(\Upsilon_{s-},\Upsilon_{s-})-2h_s(\Upsilon_{s-},0)+h_s(0,0)}{(1+R_{s-})^{2(\kappa-\tilde{a})}}\mathrm{d}R_s}{(1+R_t)^{2(\tilde{a}-\beta)}}\xrightarrow{t\to\infty}0.
\end{aligned}$$

Consequently

$$\mathbb{P}\big([Y_t\geqslant b]\big)=\mathbb{P}\Bigg(\Bigg[\underbrace{\frac{Y_t}{b}}_{\to 0}\geqslant 1\Bigg]\Bigg)\to 0\text{ as }t\to\infty,$$

hence the second term in line (8.24) tends to zero. Analogously, in the case $\tilde{\alpha}<1$, choosing

$$X_t:=\int_0^t\frac{M(\mathrm{d}s,\Upsilon_{s-})-M(\mathrm{d}s,0)}{\bar{\phi}_{s-}(1+R_{s-})^{\kappa}}\quad\text{and}\quad Y_t:=\lceil X\rceil_t$$

it holds true that

$$\mathbb{P}\Big(\Big[\Big|(1+R_t)^{\beta}\bar{\phi}_t\int_0^t(1+R_s)^{-\kappa}\bar{\phi}_{s-}^{-1}\left(M(\mathrm{d}s,\Upsilon_{s-})-M(\mathrm{d}s,0)\right)\Big|>\epsilon_1\Big]\Big)=0.$$

**Verification of (III).** We show

$$(1+R_t)^{\beta}\phi_t k\int_0^t(1+R_{s-})^{-\kappa}\phi_s^{-1}M(\mathrm{d}s,0)\xrightarrow{\mathcal{D}}\mathrm{N}(0,\Sigma).$$

It is sufficient to show

$$(1+R_{t_n})^{\beta}\phi_{t_n}k\int_0^{t_n}(1+R_{s-})^{-\kappa}\phi_s^{-1}M(\mathrm{d}s,0)\xrightarrow{\mathcal{D}}\mathrm{N}(0,\Sigma),$$

for an arbitrary sequence $t_n \uparrow \infty$. Therefore we investigate the sequence of locally square integrable martingales

$$M_s^n := (1+R_{t_n})^\beta \phi_{t_n} k \int_0^{st_n} (1+R_{r-})^{-\kappa} \phi_r^{-1} M(\mathrm{d}r, 0)$$

for some fixed $s > 0$. Note that $M_s^n \in \mathcal{M}^2_{\mathrm{loc}}$ as $(R)_{t \geqslant 0}$ and $(\phi)_{t \geqslant 0}$ are deterministic and $\int_0^t a_s M(\mathrm{d}s, \Upsilon_{s-}) \in \mathcal{M}^2_{\mathrm{loc}}$ by general assumption in the introduction. Applying the central limit theorem Theorem A.1.6 in the appendix, we show

$$M_1^n \xrightarrow{\mathcal{D}} M \text{ where } M \sim \mathrm{N}(0, \Sigma).$$

For that purpose we verify the Lindeberg-type and variance-type conditions in Theorem A.1.6 with $X^n := M_s^n \in \mathcal{M}^2_{\mathrm{loc}}$.

In the next steps we choose $S = \{1\}$ and $M_t^* := \int_0^t M(\mathrm{d}s, 0)$.

**Verification of the Lindeberg-type condition.** We have

$$\underset{n \in \mathbb{N}}{\forall} \ \underset{s \geqslant 0}{\forall} \ \Delta M_s^n = k \frac{(1+R_{t_n})^\beta}{(1+R_{st_n-})^\kappa} \phi_{t_n} \phi_{st_n}^{-1} \Delta M_{st_n}^*.$$

Therefore

$$\begin{aligned} |\Delta M_s^n|^2 &= k^2 \frac{(1+R_{t_n})^{2\beta}}{(1+R_{st_n-})^{2\kappa}} |\phi_{t_n} \phi_{st_n}^{-1} \Delta M_{st_n}^*|^2 \leqslant k^2 \frac{(1+R_{t_n})^{2\beta}}{(1+R_{st_n-})^{2\kappa}} |\phi_{t_n} \phi_{st_n}^{-1}|^2 |\Delta M_{st_n}^*|^2 \\ &\leqslant \begin{cases} \mathcal{C} \dfrac{(1+R_{st_n-})^{-2\kappa}(1+R_{st_n})^{2\tilde{a}}}{(1+R_{t_n})^{-2\beta+2\tilde{a}}} (\Delta M_{st_n}^*)^2 & \text{if } \tilde{\alpha} = 1 \\ \mathcal{C} \dfrac{(1+R_{t_n})^{2\beta}}{(1+R_{st_n-})^{2\kappa}} \bar{\phi}_{t_n}^2 \bar{\phi}_{st_n}^{-2} (\Delta M_{st_n}^*)^2 & \text{if } \tilde{\alpha} < 1. \end{cases} \end{aligned}$$

and thus

$$\begin{aligned} [|\Delta M_s^n| > \delta] &= \left[ \left| k \frac{(1+R_{t_n})^\beta}{(1+R_{st_n-})^\kappa} \phi_{t_n} \phi_{st_n}^{-1} \Delta M_{st_n}^* \right| > \delta \right] \\ &= \left[ |\phi_{t_n} \phi_{st_n}^{-1} \Delta M_{st_n}^*| > \frac{\delta}{k} \frac{(1+R_{t_n})^{-\beta}}{(1+R_{st_n-})^{-\kappa}} \right] \\ &\subset \left[ |\phi_{t_n} \phi_{st_n}^{-1}| |\Delta M_{st_n}^*| > \frac{\delta}{k} \frac{(1+R_{t_n})^{-\beta}}{(1+R_{st_n-})^{-\kappa}} \right] \\ &\subset \begin{cases} \left[ |\Delta M_{st_n}^*| > \mathcal{C} \dfrac{(1+R_{t_n})^{\tilde{a}-\beta}}{(1+R_{st_n-})^{-\kappa}(1+R_{st_n})^{\tilde{a}}} \right] & \text{if } \tilde{\alpha} = 1 \\ \left[ |\Delta M_{st_n}^*| > \mathcal{C} \dfrac{(1+R_{t_n})^{-\beta}}{(1+R_{st_n-})^{-\kappa}} \phi_{t_n}^{-1} \phi_{st_n} \right] & \text{if } \tilde{\alpha} < 1. \end{cases} \end{aligned}$$

Instead of $[|\Delta M_s^n| > \delta]$ for all $\delta \in ]0,1]$ we investigate the set

$$\mathfrak{G}_{s,t_n}^{\epsilon} := \begin{cases} \left\{ x \in \mathbb{R}^d \colon |x| > \epsilon \dfrac{(1+R_{t_n})^{\tilde{a}-\beta}}{(1+R_{st_n-})^{-\kappa}(1+R_{st_n})^{\tilde{a}}} \right\} & \text{if } \tilde{\alpha} = 1 \\ \left\{ x \in \mathbb{R}^d \colon |x| > \epsilon \dfrac{(1+R_{t_n})^{-\beta}}{(1+R_{st_n-})^{-\kappa}} \phi_{t_n}^{-1}\phi_{st_n} \right\} & \text{if } \tilde{\alpha} < 1 \end{cases}$$

for all $\epsilon \in ]0,1]$. With the inequalities above, as well as $S = \{1\}$ and $t = 1$ we get by Assumption 8.3.1

$$\begin{aligned} &\underset{\omega\in\Omega}{\forall} \; x^2 \mathbb{1}_{[|x|>\delta]} * \nu_t^{M^n} \\ &= x^2 \mathbb{1}_{[|x|>\delta]} * \nu_1^{M^n} = \int_0^1 \int_{\mathbb{R}^d} x^2 \mathbb{1}_{[|x|>\delta]} \nu^{M^n}(\mathrm{d}s, \mathrm{d}x) \\ &\leqslant \begin{cases} \mathcal{C} \displaystyle\int_0^{t_n} \int_{\mathfrak{G}_{s,t_n}^{\epsilon}} \frac{(1+R_{s-})^{-2\kappa}(1+R_s)^{2\tilde{a}}}{(1+R_{t_n})^{2(\tilde{a}-\beta)}} x^2 \nu^{M*}(\mathrm{d}s,\mathrm{d}x) & \text{if } \tilde{\alpha} = 1 \\ \mathcal{C} \displaystyle\int_0^{t_n} \int_{\mathfrak{G}_{s,t_n}^{\epsilon}} \frac{(1+R_{s-})^{-2\kappa}\phi_s^{-2}}{(1+R_{t_n})^{-2\beta}\phi_{t_n}^{-2}} x^2 \nu^{M*}(\mathrm{d}s,\mathrm{d}x) & \text{if } \tilde{\alpha} < 1 \end{cases} \\ &\leqslant \begin{cases} \mathcal{C} \dfrac{\int_0^{t_n} \frac{(1+R_s)^{2\tilde{a}}}{(1+R_{s-})^{2\kappa}} \int_{\mathfrak{G}_{s,t_n}^{\epsilon}} x^2 \nu^{M*}(\mathrm{d}s,\mathrm{d}x)}{(1+R_{t_n})^{2(\tilde{a}-\beta)}} & \text{if } \tilde{\alpha} = 1 \\ \mathcal{C} \dfrac{\int_0^{t_n} (1+R_{s-})^{-2\kappa}\phi_s^{-2} \int_{\mathfrak{G}_{s,t_n}^{\epsilon}} x^2 \nu^{M*}(\mathrm{d}s,\mathrm{d}x)}{(1+R_{t_n})^{-2\beta}\phi_{t_n}^{-2}} & \text{if } \tilde{\alpha} < 1 \end{cases} \\ &\xrightarrow{\mathbb{P}} 0 \text{ as } t_n \to \infty. \end{aligned}$$

**Verification of the variance-type condition.** We show $\lceil M^n \rceil_1 \xrightarrow{\mathbb{P}} \Sigma$. With

$$\begin{aligned} \lceil \phi_{t_n}\phi_r^{-1}M(\mathrm{d}r,0), \phi_{t_n}\phi_r^{-1}M(\mathrm{d}r,0) \rceil_s &= \phi_{t_n}\phi_s^{-1} \cdot \phi_{t_n}\phi_s^{-1} \underbrace{\lceil M(\mathrm{d}s,0), M(\mathrm{d}s,0) \rceil}_{=h_s(0)\mathrm{d}R_s} \\ &= \begin{cases} \dfrac{(\prod_{0,t_n})^2}{(1+R_{t_n})^{2\tilde{a}}} \dfrac{(1+R_s)^{2\tilde{a}}}{(\prod_{0,s})^2} h_s(0)\mathrm{d}R_s & \text{if } \tilde{\alpha} = 1 \\ \dfrac{\bar{\phi}_{t_n}^2(\prod_{0,t_n})^2}{\bar{\phi}_s^2(\prod_{0,s})^2} h_s(0)\mathrm{d}R_s & \text{if } \tilde{\alpha} < 1 \end{cases} \\ &= \begin{cases} \dfrac{(1+R_s)^{2\tilde{a}}}{(1+R_{t_n})^{2\tilde{a}}} \left(\prod_{s,t_n}\right)^2 h_s(0)\mathrm{d}R_s & \text{if } \tilde{\alpha} = 1 \\ \dfrac{\bar{\phi}_{t_n}^2}{\bar{\phi}_s^2} \left(\prod_{s,t_n}\right)^2 h_s(0)\mathrm{d}R_s & \text{if } \tilde{\alpha} < 1 \end{cases} \end{aligned}$$

and the boundedness of $\prod_{s,t_n}$, the dominated convergence theorem yields for $\tilde{\alpha} = 1$ that

$$\lceil M^n, M^n \rceil_1 = k^2 \int_0^{t_n} \frac{(1+R_{s-})^{-2\kappa}}{(1+R_{t_n})^{-2\beta}} \mathrm{d}\lceil \phi_{t_n}\phi_r^{-1}M(\mathrm{d}r,0), \phi_{t_n}\phi_r^{-1}M(\mathrm{d}r,0) \rceil_s$$

$$
\begin{aligned}
&= k^2 h_s(0) \frac{1}{(1+R_{t_n})^{2(\tilde{a}-\beta)}} \int_0^{t_n} (1+o_{\mathrm{b}}(1))(1+R_{s-})^{2(\tilde{a}-\kappa)} \mathrm{d}R_s \\
&\simeq k^2 h_s(0) \frac{\int_0^{t_n} (1+R_{s-})^{2(\tilde{a}-\kappa)} \mathrm{d}R_s}{(1+R_{t_n})^{2(\tilde{a}-\beta)}} \\
&\xrightarrow{n\to\infty} \begin{cases} k^2 h(0) \dfrac{1}{2(\tilde{a}-\kappa)+1} & \text{if } \beta = \frac{2\kappa-1}{2} \\ 0 & \text{if } \beta < \frac{2\kappa-1}{2} \end{cases}
\end{aligned}
$$

and for $\tilde{\alpha} < 1$

$$
\begin{aligned}
&[M^n, M^n]_1 \\
&= k^2 \int_0^{t_n} \frac{(1+R_{s-})^{-2\kappa}}{(1+R_{t_n})^{-2\beta}} \mathrm{d}[\phi_{t_n}\phi_r^{-1} M(\mathrm{d}r,0), \phi_{t_n}\phi_r^{-1} M(\mathrm{d}r,0)]_s \\
&= k^2 h_s(0) \bar{\phi}_{t_n}^2 (1+R_{t_n})^{2\beta} \int_0^{t_n} (1+o_{\mathrm{b}}(1))(1+R_{s-})^{-2\kappa} \bar{\phi}_{s-}^{-2} \mathrm{d}R_s \\
&\simeq k^2 h_s(0) \bar{\phi}_{t_n}^2 (1+R_{t_n})^{2\beta} \int_0^{t_n} (1+R_{s-})^{-2\kappa} \bar{\phi}_{s-}^{-2} \mathrm{d}R_s \\
&= k^2 h_s(0) \frac{\int_0^{t_n} (1+R_{s-})^{-2\kappa} \bar{\phi}_{s-}^{-2} \mathrm{d}R_s}{\int_0^{t_n} \bar{\phi}_{s-}^{-2} \mathrm{d}(1+R_s)^{-2\beta} + \int_0^{t_n} (1+R_{s-})^{-2\beta} \mathrm{d}\bar{\phi}_s^{-2} + \int_0^{t_n} \mathrm{d}[\bar{\phi}_\cdot^{-2}, (1+R_\cdot)^{-2\beta}]_s} \\
&= k^2 h_s(0) \int_0^{t_n} (1+R_{s-})^{-2\kappa} \bar{\phi}_{s-}^{-2} \mathrm{d}R_s \Big/ \Big( \int_0^{t_n} -2\beta (1+R_{s-})^{-2\beta-1} \bar{\phi}_{s-}^{-2} \mathrm{d}R_s \\
&\quad + \int_0^{t_n} \bar{\phi}_{s-}^{-2} 2\tilde{a} (1+R_{s-})^{-2\beta-\tilde{\alpha}} \mathrm{d}R_s + \int_0^{t_n} \mathrm{d}[\bar{\phi}_\cdot^{-2}, (1+R_\cdot)^{-2\beta}]_s \Big) \\
&\simeq k^2 h_s(0) \frac{\int_0^{t_n} (1+R_{s-})^{-2\kappa} \bar{\phi}_{s-}^{-2} \mathrm{d}R_s}{\int_0^{t_n} \bar{\phi}_{s-}^{-2} 2\tilde{a} (1+R_{s-})^{-2\beta-\tilde{\alpha}} \mathrm{d}R_s + \int_0^{t_n} \mathrm{d}[\bar{\phi}_\cdot^{-2}, (1+R_\cdot)^{-2\beta}]_s} \\
&= k^2 h_s(0) \frac{\int_0^{t_n} (1+R_{s-})^{-2\kappa} \bar{\phi}_{s-}^{-2} \mathrm{d}R_s}{\int_0^{t_n} \bar{\phi}_{s-}^{-2} 2\tilde{a} (1+R_{s-})^{-2\beta-\tilde{\alpha}} \mathrm{d}R_s - \int_0^{t_n} \bar{\phi}_{s-}^{-2} 4\tilde{a}\beta (1+R_{s-})^{-4\beta-\tilde{\alpha}-1} \Delta R_s \mathrm{d}R_s^d} \\
&\simeq k^2 h_s(0) \frac{\int_0^{t_n} (1+R_{s-})^{-2\kappa} \bar{\phi}_{s-}^{-2} \mathrm{d}R_s}{\int_0^{t_n} \bar{\phi}_{s-}^{-2} 2\tilde{a} (1+R_{s-})^{-2\beta-\tilde{\alpha}} \mathrm{d}R_s} \\
&\xrightarrow{n\to\infty} \begin{cases} \dfrac{k^2}{2\tilde{a}} h(0) & \text{if } \beta = \frac{2\kappa-\tilde{\alpha}}{2} \\ 0 & \text{if } \beta < \frac{2\kappa-\tilde{\alpha}}{2}, \end{cases}
\end{aligned}
$$

respectively. Note that the last convergence follows according to Toeplitz' lemma A.1.2. □

*Proof of Lemma 8.3.1.* We only have to change part **(III)** in the proof of Theorem 8.3.1. For that purpose we use the theorem of Crimaldi and Pratelli (Theorem A.1.7 in the appendix). Choose

$$
\hat{a}_t := k(1+R_t)^\beta \phi_t \qquad \hat{M}_t := \int_0^t (1+R_{s-})^{-\kappa} \phi_s^{-1} M(\mathrm{d}s, 0)
$$

and check conditions (a), (b) and (c) of their theorem. We only consider the case $\tilde{\alpha} = 1$, as the proof for $\tilde{\alpha} < 1$ is similar.

**Verification of (a).**

With the definitions of $\hat{a}_t$, $\phi_t$ and the condition $\tilde{a} > \frac{1}{2} > \beta$, we have

$$\hat{a}_t^2 \leqslant \mathcal{C}(1+R_t)^{2\beta}\phi_t^2 \leqslant \mathcal{C}(1+R_t)^{2\beta}(1+R_t)^{-2\tilde{a}} \to 0$$

and thereby $\hat{a}_t \to 0$.

**Verification of (b).**

We show that the squared process converges to zero:

$$\begin{aligned}\Big(\mathbb{E}\sup_{0\leqslant s\leqslant t}|\hat{a}_t\Delta\hat{M}_s|\Big)^2 &\leqslant \mathbb{E}\sup_{0\leqslant s\leqslant t}|\hat{a}_t\Delta\hat{M}_s|^2\\
&\leqslant \mathbb{E}\sup_{0\leqslant s\leqslant t}\left(\Big(\frac{1+R_s}{1+R_t}\Big)^{\tilde{a}}\frac{\prod_{0,t}}{\prod_{0,s}}\frac{(1+R_t)^{\beta}}{(1+R_{s-})^{\kappa}}|\Delta M_s|\right)^2\\
&\leqslant \mathbb{E}\sup_{0\leqslant s\leqslant t}\left(\underbrace{\prod\nolimits_{s,t}\Big(1+\frac{\Delta R_s}{1+R_{s-}}\Big)^{\tilde{a}}}_{\leqslant\mathcal{C}}\frac{(1+R_t)^{\beta-\tilde{a}}}{(1+R_{s-})^{\kappa-\tilde{a}}}|\Delta M_s|\right)^2\\
&\leqslant \mathcal{C}\frac{\mathbb{E}\sup_{0\leqslant s\leqslant t}(1+R_{s-})^{2(\tilde{a}-\kappa)}|\Delta M_s|^2}{(1+R_t)^{2(\tilde{a}-\beta)}} \xrightarrow{t\to\infty} 0.\end{aligned}$$

**Verification of (c).**

We show

$$\textbf{(i)}\quad \hat{a}_t^2\lceil\hat{M}\rceil_t \xrightarrow{\mathbb{P}} \Sigma \qquad \textbf{(ii)}\quad \hat{a}_t^2\big([\hat{M}]_t - \lceil\hat{M}\rceil_t\big) \xrightarrow{\mathbb{P}} 0.$$

Part (a) yields $\hat{a}_t \to 0$.

**Verification of (i).**

Consider $\lceil\hat{M}\rceil_t$:

$$\begin{aligned}\lceil\hat{M}\rceil_t &= \int_0^t (1+R_{s-})^{-2\kappa}\lceil(\phi^{-1}M(\mathrm{d}r,0)),(\phi^{-1}M(\mathrm{d}r,0))\rceil_s\\
&= \int_0^t (1+R_{s-})^{2(\tilde{a}-\kappa)}h_s(0)\mathrm{d}R_s.\end{aligned}$$

We obtain

$$\begin{aligned}k^2(1+R_t)^{2\beta}(\phi_t)^2\lceil\hat{M}\rceil_t &= k^2\frac{\int_0^t(\prod_{s,t})^2(1+R_{s-})^{2(\tilde{a}-\kappa)}h_s(0)\mathrm{d}R_s}{(1+R_t)^{2(\tilde{a}-\kappa)}}\\
&= k^2h(0)\underbrace{\frac{\int_0^t(1+o_{\mathrm{b}}(1))(1+R_{s-})^{2(\tilde{a}-\kappa)}\mathrm{d}R_s}{(1+R_t)^{2(\tilde{a}-\kappa)}}}_{\to\frac{1}{2(\tilde{a}-\kappa)+1}}.\end{aligned}$$

As a result we have

$$\hat{a}_t^2 \lceil \hat{M} \rceil_t \xrightarrow{\mathbb{P}} \Sigma \quad \text{with } \Sigma := \frac{k^2 h(0)}{2(\tilde{a}-\kappa)+1}.$$

**Verification of (ii).**
Following the arguments of the previous calculation yields

$$\hat{a}_t^2 \big( [\hat{M}]_t - \lceil \hat{M} \rceil_t \big) = \frac{\int_0^t (1+o_{\mathrm{b}}(1))(1+R_{s-})^{2(\tilde{a}-\kappa)} \mathrm{d}(\overbrace{[M^*]_s - \lceil M^* \rceil_s}^{=:L_s})}{(1+R_t)^{2(\tilde{a}-\kappa)}}.$$

As the term $o_{\mathrm{b}}(1)$ in this expression does not effect the asymptotic result, it is sufficient to show

$$(1+R_t)^{2(\beta-\tilde{a})} \int_0^t (1+R_{s-})^{2(\tilde{a}-\kappa)} \mathrm{d}L_s \xrightarrow{\mathbb{P}} 0. \tag{8.25}$$

By definition of the compensator, $L \in \mathcal{M}_{\mathrm{loc}}$ holds. Note that (8.25) holds if we can show

$$(1+R_t)^{2(\beta-\tilde{a})} \mathbb{E} \sup_{s \leqslant t} \left| \int_0^s (1+R_{r-})^{2(\tilde{a}-\kappa)} \mathrm{d}L_r \right| \to 0.$$

According to Davis' inequality (Theorem A.1.3 in the appendix),

$$\begin{aligned}
&(1+R_t)^{2(\beta-\tilde{a})} \mathbb{E} \sup_{s \leqslant t} \left| \int_0^s (1+R_{r-})^{2(\tilde{a}-\kappa)} \mathrm{d}L_r \right| \\
&\qquad \leqslant \mathcal{C} (1+R_t)^{2(\beta-\tilde{a})} \mathbb{E} \sqrt{\int_0^t (1+R_{s-})^{4(\tilde{a}-\kappa)} \mathrm{d}[L_s]} \\
&\qquad \leqslant \mathcal{C} \frac{\mathbb{E} \sqrt{\sum_{s \leqslant t} (\Delta M_s^*)^4 (1+R_{s-})^{4(\tilde{a}-\kappa)}}}{(1+R_t)^{2(\tilde{a}-\kappa)}} \\
&\qquad\quad + \mathcal{C} \frac{\mathbb{E} \sqrt{\int_0^t (1+R_{s-})^{4(\tilde{a}-\kappa)} h_s(0)^2 \Delta R_s \mathrm{d}R_s}}{(1+R_t)^{2(\tilde{a}-\kappa)}},
\end{aligned} \tag{8.26}$$

where the last inequality follows with $\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b}$, $a, b \geqslant 0$, and

$$\begin{aligned}
[L]_t &= \sum_{s \leqslant t} (\Delta L_s)^2 = \sum_{s \leqslant t} \big( \Delta [M^*]_s - \Delta \lceil M^* \rceil_s \big)^2 \\
&= \sum_{s \leqslant t} \Big( (\Delta M_s^*)^2 - \Delta \int_0^s h_r(0) \mathrm{d}R_r \Big)^2 = \sum_{s \leqslant t} \big( (\Delta M_s^*)^2 - h_s(0) \Delta R_s \big)^2 \\
&= 2 \sum_{s \leqslant t} \big( (\Delta M_s^*)^4 + (h_s(0) \Delta R_s)^2 \big) = 2 \sum_{s \leqslant t} (\Delta M_s^*)^4 + 2 \int_0^t h_s(0)^2 \Delta R_s \mathrm{d}R_s.
\end{aligned}$$

Convergence of the first term in the right side of (8.26) is shown via Jensen's inequality:

$$\left(\frac{\mathbb{E}\sqrt{\sum_{s\leqslant t}(\Delta M_s^*)^4(1+R_{s-})^{4(\tilde{a}-\kappa)}}}{(1+R_t)^{2(\tilde{a}-\beta)}}\right)^2 \leqslant \frac{\mathbb{E}\sum_{s\leqslant t}(\Delta M_s^*)^4(1+R_{s-})^{4(\tilde{a}-\kappa)}}{(1+R_t)^{4(\tilde{a}-\beta)}} \xrightarrow{t\to\infty} 0.$$

In order to verify convergence of the second term in the right side of (8.26), we apply Kronecker's lemma (Lemma A.1.3 in the appendix) to

$$\int_0^t \frac{(1+R_{s-})^{4(\tilde{a}-\kappa)}h_s(0)^2\Delta R_s}{(1+R_{s-})^{4(\tilde{a}-\beta)}}\mathrm{d}R_s \leqslant \mathcal{C}\int_0^t \frac{\Delta R_s}{(1+R_{s-})^{4(\kappa-\beta)}}\mathrm{d}R_s < \infty.$$

□

## 8.4 Special Distribution Results

In this section we derive asymptotic normality results for the special companion algorithms [RM-J], [KW-H], [KW-F-2], [KW-F-1] in a semimartingale context.

**Theorem 8.4.1.** *For the leading algorithm [RM]* (6.2) *or [KW]* (6.3) *let the assumptions of Theorem A.1.2 or Theorem 8.1.1, respectively, hold true. Let Assumption 8.3.1 hold, where the condition on the components of $G_s$ is replaced as follows. Dependent on the algorithm let*

$$\begin{array}{llll} \beta \leqslant \gamma & \textit{and} & \kappa = \tilde{\alpha} - \gamma & \textit{for [RM-J]}\ (6.4) \\ \beta \leqslant \gamma & \textit{and} & \kappa = \tilde{\alpha} - 2\gamma & \textit{for [KW-H]}\ (6.5) \\ \beta \leqslant 2\gamma & \textit{and} & \kappa = \tilde{\alpha} & \textit{for [KW-F-2]}\ (6.6) \\ & & \kappa = \tilde{\alpha} & \textit{for [c-KW-F-1]}\ (6.7). \end{array}$$

*Then*

$$\begin{array}{ll} (1+R_t)^\beta(\Upsilon_t - \upsilon^*) \xrightarrow{\mathcal{D}} \mathrm{N}(\mu,\Sigma) & \textit{if } \beta = \dfrac{2\kappa-1}{2} \\ (1+R_t)^\beta(\Upsilon_t - \upsilon^*) \xrightarrow{\mathbb{P}} \mu & \textit{if } \beta < \dfrac{2\kappa-1}{2}. \end{array}$$

*The parameters $\mu$ and $\Sigma$ are given in the following way. In all settings of algorithms [KW-H]* (6.5) *and [KW-F-1]* (6.7), *as well as [KW-F-2]* (6.6), *if $f \in C^2$, $\mu = 0$ holds true. For algorithm [KW-F-2]* (6.6), *if even $f \in C^3$, the bias is*

$$\mu = \begin{cases} \dfrac{\tilde{a}c^2}{2|\mathcal{S}|(\tilde{a}-2\gamma)}\displaystyle\sum_{i\in\mathcal{S}}\frac{\partial^2}{\partial x_i^2}f(z^*) & \textit{if } \tilde{\alpha} = 1 \textit{ and } \beta = 2\gamma \\ \dfrac{c^2}{2|\mathcal{S}|}\displaystyle\sum_{i\in\mathcal{S}}\frac{\partial^2}{\partial x_i^2}f(z^*) & \textit{if } \tilde{\alpha} < 1 \textit{ and } \beta = 2\gamma \\ 0 & \textit{if } \tilde{\alpha} \leqslant 1 \textit{ and } \beta < 2\gamma. \end{cases}$$

*For [RM-J]* (6.4) *the bias is*

$$\mu = \begin{cases} \dfrac{\tilde{a}cf''(z^*)}{\tilde{a}-\gamma} & \text{if } \tilde{\alpha} = 1 \text{ and } \beta = \gamma \\ cf''(z^*) & \text{if } \tilde{\alpha} < 1 \text{ and } \beta = \gamma \\ 0 & \text{if } \tilde{\alpha} \leqslant 1 \text{ and } \beta < \gamma. \end{cases}$$

*Furthermore for [RM-J] the variance is*

$$\Sigma = \begin{cases} \dfrac{(\tilde{a}/c)^2}{2(\tilde{a}-\tilde{\alpha}+\gamma)+1} h(\upsilon^*) & \text{if } \beta = \frac{2(\tilde{\alpha}-\gamma)-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2c} h(\upsilon^*) & \text{if } \beta = \frac{2(\tilde{\alpha}-\gamma)-1}{2} \text{ and } \tilde{\alpha} < 1, \end{cases}$$

*for [KW-H]*

$$\Sigma = \begin{cases} \dfrac{(\tilde{a}/c^2)^2}{2(\tilde{a}-\tilde{\alpha}+2\gamma)+1} h(\upsilon^*) & \text{if } \beta = \frac{2(\tilde{\alpha}-2\gamma)-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2c^2} h(\upsilon^*) & \text{if } \beta = \frac{2(\tilde{\alpha}-2\gamma)-1}{2} \text{ and } \tilde{\alpha} < 1, \end{cases}$$

*and for [KW-F-1] as well as [KW-F-2]*

$$\Sigma = \begin{cases} \dfrac{\tilde{a}^2}{2\tilde{a}-1} h(\upsilon^*) & \text{if } \beta = \frac{2\tilde{\alpha}-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2} h(\upsilon^*) & \text{if } \beta = \frac{2\tilde{\alpha}-1}{2} \text{ and } \tilde{\alpha} < 1. \end{cases}$$

*Remark* 8.4.1. At first sight one might be confused why [KW-F-2] always has a bias term $\mu = 0$ for $f \in C^2$, but if $f \in C^3$ there are settings such that $\mu \neq 0$. This is due to the fact that by Assumption 8.3.1 $\beta$ must be chosen small enough such that $\beta \leqslant \frac{p}{2}\chi$ and hence we have $\beta < 2\gamma$ for $f \in C^2$.

*Proof of Theorem 8.4.1.* All we have to do is to check that our assumptions also fulfill the assumption on the decomposition of the term $G_s$ in Theorem 8.3.1. We first take a closer look at the $G_s$-terms of companion algorithms [KW-F-2] (6.6) and [KW-F-1] (6.7). A simple calculation concerning algorithm [KW-F-1] (6.7) yields

$$\begin{aligned} G_s - \upsilon^* &= \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} f(Z_{s-}) - f(z^*) \\ &= \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} f(z^*) + \nabla f(z^*)(Z_{s-} - z^*) + \mathcal{O}(\|Z_{s-} - z^*\|^2) - f(z^*) \\ &= \mathcal{O}\left(\|Z_{s-} - z^*\|^2\right), \end{aligned}$$

such that choosing $l = 0$ and $\chi = 2$ yields the result. A Taylor expansion for the case of algorithm [KW-F-2] (6.6) yields

$$G_s - \upsilon^* = \frac{1}{2|\mathcal{S}|} \sum_{i \in \mathcal{S}} (f(Z_{s-} + c_s e_i) + f(Z_{s-} - c_s e_i)) - f(z^*)$$

$$
\begin{aligned}
&= \frac{1}{2|\mathcal{S}|} \sum_{i\in\mathcal{S}} \int_0^1 c_s \big(\frac{\partial}{\partial z_i} f(Z_{s-} + tc_s) - \frac{\partial}{\partial z_i} f(Z_{s-} - tc_s)\big) \mathrm{d}t + f(Z_{s-}) - f(z^*) \\
&= \frac{1}{2|\mathcal{S}|} \sum_{i\in\mathcal{S}} \int_0^1 c_s \big(\frac{\partial^2}{\partial z_i^2} f(z^*) 2tc_s + o(\|Z_{s-} - z^*\|) + o(c_s)\big) \mathrm{d}t + f(Z_{s-}) - f(z^*) \\
&= \frac{1}{2|\mathcal{S}|} \sum_{i\in\mathcal{S}} \big(c_s^2 \frac{\partial^2}{\partial z_i^2} f(z^*) + o(c_s\|Z_{s-} - z^*\|) + o(c_s^2)\big) + f(Z_{s-}) - f(z^*) \\
&= \frac{1}{2|\mathcal{S}|} \sum_{i\in\mathcal{S}} \frac{\partial^2}{\partial z_i^2} f(z^*) c_s^2 + \mathcal{O}\left(\|Z_{s-} - z^*\|^2\right) + o\left(c_s^2\right).
\end{aligned}
$$

It doesn't matter whether $f$ is two or three times differentiable, as we cannot get rid of the term $\frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} \frac{\partial^2}{\partial z_i^2} f(z^*) c_s^2$. Consequently the decomposition is given by

$$
l := \frac{c^2}{2|\mathcal{S}|} \sum_{i\in\mathcal{S}} \frac{\partial^2}{\partial x_i^2} f(z^*), \quad \iota := 2\gamma, \quad \chi := 2, \quad \eta := 2\gamma.
$$

For algorithm [RM-J] (6.4) we get

$$
\begin{aligned}
G_s - \upsilon^* &= \frac{1}{c_s}\Big(f(Z_{s-} + c_s) - f(Z_{s-}\Big) - J_{z^*} = \int_0^1 \nabla f(Z_{s-} + c_s) \mathrm{d}t - \int_0^1 J_{z^*} \mathrm{d}t \\
&= \int_0^1 (J_{Z_{s-}+tc_s} - J_{z^*}) \mathrm{d}t \\
&= \int_0^1 (J_{z^*} + f''(z^*)(Z_{s-} - z^* + tc_s) + o(Z_{s-} - z^* + tc_s) - J_{z^*}) \mathrm{d}t \\
&= f''(z^*) c_s + \mathcal{O}(r_s) + o(c_s).
\end{aligned}
$$

Hence we decompose as

$$
l := f''(z^*), \quad \iota := \gamma, \quad \chi := 1, \quad \eta := \gamma.
$$

Finally, for [KW-H] (6.5)

$$
\begin{aligned}
G_s - \upsilon^* &= \frac{1}{c_s^2}\Big(f(Z_{s-} + c_s) + f(Z_{s-} - c_s) - 2f(Z_{s-})\Big) - H_{z^*} \\
&= \frac{1}{c_s^2}\Big(\int_0^1 c_s \nabla f(Z_{s-} + tc_s) \mathrm{d}t + f(Z_{s-} - c_s) - f(Z_{s-})\Big) - H_{z^*} \\
&= \frac{1}{c_s}\Big(\int_0^1 \nabla f(Z_{s-} + tc_s) - \nabla f(Z_{s-} - tc_s) \mathrm{d}t\Big) - H_{z^*} \\
&= \int_0^1 \int_{-1}^1 t \nabla^2 f(Z_{s-} + tuc_s) \mathrm{d}u \mathrm{d}t - H_{z^*} \\
&= \int_0^1 \int_{-1}^1 t\Big(\nabla^2 f(z^*) + \nabla^3 f(z^*)(Z_{s-} - z^* + tuc_s) \\
&\quad + o(Z_{s-} - z^* + tuc_s)\Big) \mathrm{d}u \mathrm{d}t - H_{z^*}
\end{aligned}
$$

$$
\begin{aligned}
&= \int_0^1 2t\big(\nabla^2 f(z^*) + \nabla^3 f(z^*)(Z_{s-} - z^*) + o(Z_{s-} - z^*)\big) \\
&\quad + \big[\frac{u^2}{2} t^2 c_s(\nabla^3 f(z^*) + o(1))\big]_{-1}^1 \mathrm{d}t - H_{z^*} \\
&= \int_0^1 2t\big(\nabla^2 f(z^*) + \nabla^3 f(z^*)(Z_{s-} - z^*) + o(Z_{s-} - z^*)\big)\mathrm{d}t - H_{z^*} \\
&= \big[t^2\big(\nabla^2 f(z^*) + \nabla^3 f(z^*)(Z_{s-} - z^*) + o(|Z_{s-} - z^*|)\big)\big]_0^1 - H_{z^*} \\
&= \mathcal{O}(|Z_{s-} - z^*|).
\end{aligned}
$$

Similarly as before, choose

$$
l := 0, \quad \chi := 1.
$$

□

## 8.5 Itô Type and Recursive Stochastic Approximation Algorithms

We now turn to Itô type and time-discrete algorithms which are special cases of the semimartingale model.

**Corollary 8.5.1.** *Consider the generic Itô type algorithm [c-Gen-Comp]* (6.11) *with $f \in C^p$. Let conditions (A) and (cD) from Assumption 6.1.1 and 6.3.1, respectively as well as*

$$
\sigma_s(y) \leqslant C(1 + |y|) \quad \text{and} \quad \lim_{\substack{s\to\infty \\ \Upsilon\to\upsilon^*}} \sigma_s(\Upsilon) = \sigma(\upsilon^*)
$$

*hold. If $\tilde{\alpha} = 1$ assume $\tilde{a} > 1/2$. If $\tilde{\alpha} < 1$ assume $\tilde{a} > 0$.*

*The leading process $Z$ shall converge to $z^*$ in the almost $L^2$ sense with rate $(1+s)^{-p}$. Assume $l, \iota, \nu, \chi > 0$ such that $G_s$ can be decomposed as*

$$
G_s = l_s + \mathcal{O}(\|Z_s - z^*\|^\chi) + o(m_s),
$$

*with $l_s := l(1 + s)^{-\iota}$ and $m_s := (1 + s)^{-\eta}$. Moreover choose $\beta$ with $\beta \leqslant \frac{p}{2}\chi$, $\beta \leqslant \iota$ and $\beta \leqslant \eta$.*

*Then*

$$
\begin{aligned}
&(1 + t)^\beta(\Upsilon_t - \upsilon^*) \xrightarrow{\mathcal{D}} \mathrm{N}(\mu, \Sigma) \qquad \text{if } \beta = \frac{2\kappa - 1}{2} \\
&(1 + t)^\beta(\Upsilon_t - \upsilon^*) \xrightarrow{\mathbb{P}} \mu \qquad\qquad\quad \text{if } \beta < \frac{2\kappa - 1}{2}.
\end{aligned}
$$

*Bias and variance are given by*

$$
\mu = \begin{cases} \dfrac{\tilde{a}l}{\tilde{a} - \iota} & \text{if } \tilde{\alpha} = 1 \text{ and } \beta = \iota \\ l & \text{if } \tilde{\alpha} < 1 \text{ and } \beta = \iota \\ 0 & \text{if } \tilde{\alpha} \leqslant 1 \text{ and } \beta < \iota. \end{cases}
$$

*and*

$$\Sigma = \begin{cases} \dfrac{\tilde{a}^2}{2\tilde{a}-1}\sigma(\upsilon^*)^2 & \text{if } \beta = \frac{2\kappa-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2}\sigma(\upsilon^*)^2 & \text{if } \beta = \frac{2\kappa-1}{2} \text{ and } \tilde{\alpha} < 1. \end{cases}$$

Before turning to the proof, we present a corollary to Theorem 8.4.1.

**Corollary 8.5.2.** *Consider the algorithms [c-RM-J], [c-KW-H], [c-KW-F-2] and [c-KW-F-1] with $f \in C^p$. Let conditions (A), (cD) from Assumption 6.1.1 and 6.3.1, respectively,*

$$\sigma_s(y) \leqslant C(1+|y|) \quad \text{and} \quad \lim_{\substack{s\to\infty \\ \Upsilon\to\upsilon^*}} \sigma_s(\Upsilon) = \sigma(\upsilon^*)$$

*hold. If $\tilde{\alpha} = 1$, assume $\tilde{a} > 1/2$. If $\tilde{\alpha} < 1$, assume $\tilde{a} > 0$. The leading process $Z$ shall converge to $z^*$ in the almost $L^2$ sense with rate $(1+s)^{-p}$. Depending on the algorithm, assume*

$$\begin{array}{lllll} \beta \leqslant \gamma, & \beta \leqslant \dfrac{p}{2} & \text{and} & \kappa = \tilde{\alpha} - \gamma & \text{for [c-RM-J]} \\ \beta \leqslant \gamma, & \beta \leqslant \dfrac{p}{2} & \text{and} & \kappa = \tilde{\alpha} - 2\gamma & \text{for [c-KW-H]} \\ \beta \leqslant 2\gamma, & \beta \leqslant p & \text{and} & \kappa = \tilde{\alpha} & \text{for [c-KW-F-2]} \\ & \beta \leqslant p & \text{and} & \kappa = \tilde{\alpha} & \text{for [c-KW-F-1].} \end{array}$$

*Then*

$$\begin{aligned} (1+t)^\beta(\Upsilon_t - \upsilon^*) &\xrightarrow{\mathcal{D}} \mathrm{N}(\mu,\Sigma) && \text{if } \beta = \frac{2\kappa-1}{2} \\ (1+t)^\beta(\Upsilon_t - \upsilon^*) &\xrightarrow{\mathbb{P}} \mu && \text{if } \beta < \frac{2\kappa-1}{2}. \end{aligned}$$

*Here, the bias $\mu$ and the variance $\Sigma$ are defined in the following way.*

*In the case that $f \in C^2$ in algorithm [c-KW-F-2], and for all settings of algorithms [c-KW-F-1] and [c-KW-H], $\mu = 0$ holds true. For algorithm [c-KW-F-2], if even $f \in C^3$,*

$$\mu = \begin{cases} \dfrac{\tilde{a}c^2}{2|\mathcal{S}|(\tilde{a}-2\gamma)} \displaystyle\sum_{i\in\mathcal{S}} \frac{\partial^2}{\partial x_i^2} f(z^*) & \text{if } \tilde{\alpha} = 1 \text{ and } \beta = 2\gamma \\ \dfrac{c^2}{2|\mathcal{S}|} \displaystyle\sum_{i\in\mathcal{S}} \frac{\partial^2}{\partial x_i^2} f(z^*) & \text{if } \tilde{\alpha} < 1 \text{ and } \beta = 2\gamma \\ 0 & \text{if } \tilde{\alpha} \leqslant 1 \text{ and } \beta < 2\gamma. \end{cases}$$

*For [c-RM-J] (6.4) the bias is*

$$\mu = \begin{cases} \dfrac{\tilde{a}cf''(z^*)}{\tilde{a}-\gamma} & \text{if } \tilde{\alpha} = 1 \text{ and } \beta = \gamma \\ cf''(z^*) & \text{if } \tilde{\alpha} < 1 \text{ and } \beta = \gamma \\ 0 & \text{if } \tilde{\alpha} \leqslant 1 \text{ and } \beta < \gamma. \end{cases}$$

*Furthermore for [RM-J] the variance is*

$$\Sigma = \begin{cases} \dfrac{(\tilde{a}/c)^2}{2(\tilde{a}-\tilde{\alpha}+\gamma)+1}\sigma(\upsilon^*)^2 & \text{if } \beta = \frac{2(\tilde{\alpha}-\gamma)-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2c}\sigma(\upsilon^*)^2 & \text{if } \beta = \frac{2(\tilde{\alpha}-\gamma)-1}{2} \text{ and } \tilde{\alpha} < 1, \end{cases}$$

*for [KW-H]*

$$\Sigma = \begin{cases} \dfrac{(\tilde{a}/c^2)^2}{2(\tilde{a}-\tilde{\alpha}+2\gamma)+1}\sigma(\upsilon^*)^2 & \text{if } \beta = \frac{2(\tilde{\alpha}-2\gamma)-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2c^2}\sigma(\upsilon^*)^2 & \text{if } \beta = \frac{2(\tilde{\alpha}-2\gamma)-1}{2} \text{ and } \tilde{\alpha} < 1, \end{cases}$$

*and for [KW-F-1] as well as [KW-F-2]*

$$\Sigma = \begin{cases} \dfrac{\tilde{a}^2}{2\tilde{a}-1}\sigma(\upsilon^*)^2 & \text{if } \beta = \frac{2\tilde{\alpha}-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2}\sigma(\upsilon^*)^2 & \text{if } \beta = \frac{2\tilde{\alpha}-1}{2} \text{ and } \tilde{\alpha} < 1. \end{cases}$$

*Proof of Corollaries 8.5.1 and 8.5.2.* We verify the assumptions of Theorems 8.3.1 and 8.4.1. Due to our assumptions, the conditions of Corollaries 7.4.2 and 7.4.3 are also valid. This guarantees the existence of a function $S_t \uparrow \infty$ with $S_t|\Upsilon_t| \to 0$ a.s. Definitions and notations of the proof of Corollaries 7.4.2 and 7.4.3 are reused. Note that the decomposition of $G_s$ or its special form for non-generic algorithms, respectively, was already given there. The Lindeberg-type condition can be shown using the continuity of the Brownian motion, since

$$\begin{aligned} \mu^W([0,t]\times\Gamma) &= \sum_{0<s\leqslant t} \mathbb{1}_{\{\Delta W_s \in \Gamma\}} = \sum_{0<s\leqslant t} \mathbb{1}_{\{0\in\Gamma\}} \\ &= \sum_{0<s\leqslant t} 0 = 0 \quad \text{for all } \Gamma \in \mathcal{B}_d(\mathbb{R}\backslash\{0\}),\ t \in \mathbb{R}_+, \end{aligned}$$

yields $\nu^W([0,t]\times\Gamma) = 0$ for all $\Gamma \in \mathcal{B}_d(\mathbb{R}\backslash\{0\})$ and $t \in \mathbb{R}_+$.

The condition

$$\sum_{0\leqslant s} \mathbb{1}_{\{\tilde{a}\Delta R_s = (1+R_{s-})\}} < \infty$$

follows from $\Delta R_s = 0$ since $R_s := s$. Computing

$$\left[\int_0^{\cdot} M(\mathrm{d}s, y)\right]_t = \int_0^t \sigma_s^2(y)\mathrm{d}[W_\cdot]_s = \int_0^t \sigma_s^2(y)\mathrm{d}s$$

yields $h_s(y) = \sigma_s^2(y)$ and therefore

$$\lim_{\substack{s\to\infty \\ y\to\upsilon^*}} h_s(y) = \lim_{\substack{s\to\infty \\ y\to\upsilon^*}} \sigma_s^2(y) = \sigma^2(\upsilon^*) = h(\upsilon^*).$$

Furthermore, we conclude

$$|h_s(y)| = |\sigma_s^2(y)| \leqslant dC^2(1+|y|)^2 \leqslant \mathcal{C},$$

as $\Upsilon$ converges a.s. □

**Assumption 8.5.1.** *Let conditions (A) and (dD) from Assumption 6.1.1 and 6.3.3 hold.*

*(dD2)* $\sup_n \mathbb{E}\left(V_n^2 \mid \mathcal{F}_{n-1}\right) < \infty$ *$\mathbb{P}$-a.s.,*

*(dD3)* $\mathbb{E}\left(V_n^2 \mid \mathcal{F}_{n-1}\right) \xrightarrow{n\to\infty} \tilde{h}$ *$\mathbb{P}$-a.s., and*

*(L)* $\sup_n \mathbb{E}\left(V_n^{2+\delta}\right) < \infty$ *$\mathbb{P}$-a.s. for all* $\delta > 0$

Now we turn to a result on time-discrete generic algorithms.

**Corollary 8.5.3.** *Consider the algorithms [d-RM-J], [d-KW-H], [d-KW-F-2] and [d-KW-F-1] with $f \in C^p$. Let Assumption 8.5.1 hold. If $\tilde{\alpha} = 1$ choose $\tilde{a} > 1/2$. If $\tilde{\alpha} < 1$ choose $\tilde{a} > 0$. The leading process $Z$ shall converge to $z^*$ in the almost $L^2$ sense with rate $n^{-p}$. Assume the existence of constants $l, \iota, \nu, \chi > 0$ such that $G_n$ can be decomposed as*

$$G_n = l_n + \mathcal{O}(\|Z_n - z^*\|^{\chi}) + o(m_n),$$

*with $l_n := ln^{-\iota}$ and $m_n := n^{-\eta}$. Moreover assume a $\beta$ with $\beta \leqslant \frac{p}{2}\chi$, $\beta \leqslant \iota$ and $\beta \leqslant \eta$. Then*

$$\begin{aligned} &n^{\beta}(\Upsilon_n - \upsilon^*) \xrightarrow{\mathcal{D}} \mathrm{N}(\mu, \Sigma) && \text{if } \beta = \frac{2\kappa - 1}{2} \\ &n^{\beta}(\Upsilon_n - \upsilon^*) \xrightarrow{\mathbb{P}} \mu && \text{if } \beta < \frac{2\kappa - 1}{2}. \end{aligned}$$

*The bias is*

$$\mu = \begin{cases} \dfrac{\tilde{a}l}{\tilde{a} - \iota} & \text{if } \tilde{\alpha} = 1 \text{ and } \beta = \iota \\ l & \text{if } \tilde{\alpha} < 1 \text{ and } \beta = \iota \\ 0 & \text{if } \tilde{\alpha} \leqslant 1 \text{ and } \beta < \iota. \end{cases}$$

*and the variance*

$$\Sigma = \begin{cases} \dfrac{\tilde{a}^2}{2\tilde{a}-1}h & \text{if } \beta = \frac{2\kappa-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2}h & \text{if } \beta = \frac{2\kappa-1}{2} \text{ and } \tilde{\alpha} < 1. \end{cases}$$

Before presenting the proof we formulate another corollary of Theorem 8.4.1. Results for algorithms [d-KW-F-2] and [d-KW-F-1] for $f \in C^3$ have already been shown by Mokkadem and Pelletier [28].

**Corollary 8.5.4.** *Consider the algorithms [d-RM-J], [d-KW-H], [d-KW-F-2] and [d-KW-F-1] with $f \in C^p$. Let Assumption 8.5.1 hold. If $\tilde{\alpha} = 1$, assume $\tilde{a} > 1/2$. If $\tilde{\alpha} < 1$, assume $\tilde{a} > 0$. The leading process $Z$ shall converge to $z^*$ in the almost $L^2$ sense with rate $n^{-p}$. Depending on the algorithm, assume*

$$\begin{array}{llll} \beta \leqslant \gamma, & \beta \leqslant \dfrac{p}{2} & \text{and} \quad \kappa = \tilde{\alpha} - \gamma & \text{for [d-RM-J]} \\ \beta \leqslant \gamma, & \beta \leqslant \dfrac{p}{2} & \text{and} \quad \kappa = \tilde{\alpha} - 2\gamma & \text{for [d-KW-H]} \\ \beta \leqslant 2\gamma, & \beta \leqslant p & \text{and} \quad \kappa = \tilde{\alpha} & \text{for [d-KW-F-2]} \\ & \beta \leqslant p & \text{and} \quad \kappa = \tilde{\alpha} & \text{for [d-KW-F-1].} \end{array}$$

*Then*

$$\begin{array}{ll} n^\beta(\Upsilon_n - \upsilon^*) \xrightarrow{\mathcal{D}} \mathrm{N}(\mu, \Sigma) & \text{if } \beta = \dfrac{2\kappa-1}{2} \\ n^\beta(\Upsilon_n - \upsilon^*) \xrightarrow{\mathbb{P}} \mu & \text{if } \beta < \dfrac{2\kappa-1}{2}. \end{array}$$

*Here, the parameters $\mu$ and $\Sigma$ are defined in the following way. In the case that $f \in C^2$ in algorithm [d-KW-F-2], and for all settings of algorithms [d-KW-F-1] and [d-KW-H], $\mu = 0$ holds true. For algorithm [d-KW-F-2], if even $f \in C^3$,*

$$\mu = \begin{cases} \dfrac{\tilde{a}c^2}{2|\mathcal{S}|(\tilde{a}-2\gamma)} \displaystyle\sum_{i\in\mathcal{S}} \frac{\partial^2}{\partial x_i^2} f(z^*) & \text{if } \tilde{\alpha} = 1 \text{ and } \beta = 2\gamma \\ \dfrac{c^2}{2|\mathcal{S}|} \displaystyle\sum_{i\in\mathcal{S}} \frac{\partial^2}{\partial x_i^2} f(z^*) & \text{if } \tilde{\alpha} < 1 \text{ and } \beta = 2\gamma \\ 0 & \text{if } \tilde{\alpha} \leqslant 1 \text{ and } \beta < 2\gamma. \end{cases}$$

*For [c-RM-J]* (6.4) *the bias is*

$$\mu = \begin{cases} \dfrac{\tilde{a}cf''(z^*)}{\tilde{a}-\gamma} & \text{if } \tilde{\alpha} = 1 \text{ and } \beta = \gamma \\ cf''(z^*) & \text{if } \tilde{\alpha} < 1 \text{ and } \beta = \gamma \\ 0 & \text{if } \tilde{\alpha} \leqslant 1 \text{ and } \beta < \gamma. \end{cases}$$

*Furthermore for [RM-J] the variance is*

$$\Sigma = \begin{cases} \dfrac{(\tilde{a}/c)^2}{2(\tilde{a}-\tilde{\alpha}+\gamma)+1} h & \text{if } \beta = \frac{2(\tilde{\alpha}-\gamma)-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2c} h & \text{if } \beta = \frac{2(\tilde{\alpha}-\gamma)-1}{2} \text{ and } \tilde{\alpha} < 1, \end{cases}$$

*for [KW-H]*

$$\Sigma = \begin{cases} \dfrac{(\tilde{a}/c^2)^2}{2(\tilde{a}-\tilde{\alpha}+2\gamma)+1} h & \text{if } \beta = \frac{2(\tilde{\alpha}-2\gamma)-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2c^2} h & \text{if } \beta = \frac{2(\tilde{\alpha}-2\gamma)-1}{2} \text{ and } \tilde{\alpha} < 1, \end{cases}$$

*and for [KW-F-1] as well as [KW-F-2]*

$$\Sigma = \begin{cases} \dfrac{\tilde{a}^2}{2\tilde{a}-1} h & \text{if } \beta = \frac{2\tilde{\alpha}-1}{2} \text{ and } \tilde{\alpha} = 1 \\ \dfrac{\tilde{a}}{2} h & \text{if } \beta = \frac{2\tilde{\alpha}-1}{2} \text{ and } \tilde{\alpha} < 1. \end{cases}$$

*Proof of Corollaries 8.5.3 and 8.5.4.* We verify the assumptions of Theorem 8.4.1. Due to our assumptions Corollaries 7.4.5 and 7.4.6 apply. This guarantees the existence of an increasing process $S_t \uparrow \infty$ such that $S_t|\Upsilon_t| \to 0$ a.s. The definitions and notations of the proof of Corollaries 7.4.5 and 7.4.6 are reused. These corollaries also yield the decomposition and alternative representations of $G_n$.

Calculating the quadratic variation

$$\left[\int_0^{\cdot} M(\mathrm{d}s, y)\right]_t = \left[\int_0^{\cdot} \tilde{V}_s \mathrm{d}R_s\right]_t = \sum_{\substack{n\leqslant t\\ n\in\mathbb{N}}} V_n^2(\Delta R_n)^2 = \sum_{\substack{n\leqslant t\\ n\in\mathbb{N}}} V_n^2$$

yields the predictable quadratic variation

$$\left[\int_0^{\cdot} M(\mathrm{d}s, y)\right]_t = \sum_{\substack{n\leqslant t\\ n\in\mathbb{N}}} \mathbb{E}\left(\tilde{V}_n^2 \mid \mathcal{F}_{n-1}\right) = \sum_{\substack{n\leqslant t\\ n\in\mathbb{N}}} \mathbb{E}\left(V_n^2 \mid \mathcal{F}_{n-1}\right).$$

As $h_n(y_1, y_2) = \mathbb{E}\left(V_n^2 \mid \mathcal{F}_{n-1}\right)$, convergence and boundedness of $h_n$ are assured by

$$h_n(y_1, y_2) = \mathbb{E}\left(V_n^2 \mid \mathcal{F}_{n-1}\right) \xrightarrow{n\to\infty} h \text{ a.s.}$$

and

$$\sup_n h_n = \sup_n \mathbb{E}\left(V_n^2 \mid \mathcal{F}_{n-1}\right) < \infty.$$

Since $h_n(y_1, y_2)$ and $h_n(y)$ are independent of $y_1$ and $y_2$, we get

$$\mathbb{E}\left(h_s(\Upsilon_{s-}) - 2h_s(\Upsilon_{s-}, 0) + h_s(0)\right) = \mathbb{E}\left(h_n - 2h_n + h_n\right) = 0.$$

As $\tilde{a}$ is not necessarily integer-valued, we conclude

$$\sum_{0\leqslant s} \mathbb{1}_{\{\tilde{a}\Delta R_s=(1+R_{s-})\}} = \sum_{0\leqslant s} \mathbb{1}_{\{\tilde{a}=1+\lfloor s\rfloor\}} \leqslant 1 < \infty.$$

In central limit theorems for triangular schemes of random variables it is well-known, that the Lindeberg type condition is implied by the Lyapunov type condition. In order to show the Lindeberg type condition in our context, we use assumption $(L)$ which corresponds to a Lyapunov type condition. Dependent on the choice of $\tilde{\alpha}$ we have two kinds of Lyapunov type conditions. If $\tilde{\alpha}=1$ consider

$$\begin{aligned}
&\frac{\int_0^t \frac{(1+R_s)^{2\tilde{a}}}{(1+R_{s-})^{2\kappa}} \int_{\mathfrak{G}^{\epsilon}_{s,t}} x^2 \nu^{M^*}(\mathrm{d}s,\mathrm{d}x)}{(1+R_t)^{2(\tilde{a}-\beta)}} \\
&\quad\leqslant \frac{\int_0^t \frac{(1+R_s)^{2\tilde{a}}}{(1+R_{s-})^{2\kappa}} \int_{\mathbb{R}} x^2 \nu^{M^*}(\mathrm{d}s,\mathrm{d}x)}{(1+R_t)^{2(\tilde{a}-\beta)}} \leqslant \frac{\int_0^t \frac{(1+R_s)^{2\tilde{a}}}{(1+R_{s-})^{2\kappa}} \int_{\mathbb{R}} x^2 N_s(\omega,\mathrm{d}x)\mathrm{d}C_s}{(1+R_t)^{2(\tilde{a}-\beta)}} \\
&\quad\leqslant \mathcal{C}\frac{\int_0^t \frac{(1+R_s)^{2\tilde{a}}}{(1+R_{s-})^{2\kappa}} \int_{\mathbb{R}} x^2 N_s(\omega,\mathrm{d}x)\mathrm{d}R_s}{(1+R_t)^{2(\tilde{a}-\beta)}} \leqslant \mathcal{C}\frac{\sum_{i=1}^{\lfloor t\rfloor} \frac{(1+i)^{2\tilde{a}}}{i^{2\kappa}} \int_{\mathbb{R}} x^2 N_i(\omega,\mathrm{d}x)}{(1+\lfloor t\rfloor)^{2(\tilde{a}-\beta)}} \\
&\quad\leqslant \mathcal{C}\frac{\sum_{i=1}^{\lfloor t\rfloor} \frac{i^{2\tilde{a}}}{i^{2\kappa}} \int_{\mathbb{R}} x^2 \mathbb{P}_{(V_i|\mathcal{F}_{i-1})}(\mathrm{d}x)}{\lfloor t\rfloor^{2(\tilde{a}-\beta)}} \leqslant \mathcal{C}\frac{\sum_{i=1}^{\lfloor t\rfloor} i^{2\tilde{a}-2\kappa}\mathbb{E}(V_i^2 \mid \mathcal{F}_{i-1})}{\lfloor t\rfloor^{2(\tilde{a}-\beta)}} \\
&\quad\leqslant \mathcal{C}\frac{\sum_{i=1}^{n} i^{2\tilde{a}-2\kappa}\mathbb{E}(V_i^2 \mid \mathcal{F}_{i-1})}{n^{2(\tilde{a}-\beta)}}
\end{aligned} \tag{8.27}$$

for $t=n\in\mathbb{N}$. We made use of the relations

$$\nu^{M^*}(\omega,\mathrm{d}t,\mathrm{d}x) = N_t(\omega,\mathrm{d}x)\mathrm{d}C_t, \qquad \text{where } C_t = [M^*_{\cdot}]_t,$$

with $M^* = \int_0^t M(\mathrm{d}s, v^*)$ as given in Theorem 8.3.1 and $N_t(\omega,A) = \mathbb{P}\big(V_t\in A \mid \mathcal{F}_{t-1}\big)$. More details on the latter identities can be found in the book of Jacod and Shiryayev [18, Chapter II] or a paper of Wang [43].

In order to show that (8.27) tends to zero in probability, we consider its expectation value and apply Jensen's inequality, Hölder's inequality and Kronecker's lemma:

$$\begin{aligned}
&\left(\mathbb{E}\frac{\sum_{i=1}^n i^{2\tilde{a}-2\kappa}\mathbb{E}\big(V_i^2 \mid \mathcal{F}_{i-1}\big)}{n^{2(\tilde{a}-\beta)}}\right)^{1+\frac{\delta}{2}} \\
&\quad\leqslant \mathbb{E}\frac{\sum_{i=1}^n i^{(2\tilde{a}-2\kappa)(1+\frac{\delta}{2})}\mathbb{E}\big(V_i^{2+\delta} \mid \mathcal{F}_{i-1}\big)}{n^{2(\tilde{a}-\beta)(1+\frac{\delta}{2})}} \leqslant \frac{\sum_{i=1}^n i^{(2\tilde{a}-2\kappa)(1+\frac{\delta}{2})}\mathbb{E}\big(V_i^{2+\delta}\big)}{n^{2(\tilde{a}-\beta)(1+\frac{\delta}{2})}} \\
&\quad\leqslant \left(\sup_j \mathbb{E}V_j^{2+\delta}\right)\frac{\sum_{i=1}^n i^{(2\tilde{a}-2\kappa)(1+\frac{\delta}{2})}}{n^{2(\tilde{a}-\beta)(1+\frac{\delta}{2})}} \leqslant \mathcal{C}\frac{\sum_{i=1}^n i^{(2\tilde{a}-2\kappa)(1+\frac{\delta}{2})}}{n^{2(\tilde{a}-\beta)(1+\frac{\delta}{2})}} \\
&\quad\xrightarrow{n\to\infty} 0.
\end{aligned}$$

In the last step Kronecker's lemma was applied in the following way:

$$\sum_{i=1}^{n} \frac{i^{(2\tilde{a}-2\kappa)(1+\frac{\delta}{2})}}{i^{2(\tilde{a}-\beta)(1+\frac{\delta}{2})}} = \sum_{i=1}^{n} i^{-(2\kappa-2\beta)(1+\frac{\delta}{2})} \leqslant \sum_{i=1}^{n} i^{-(1+\frac{\delta}{2})} < \infty \Rightarrow \lim_{n\to\infty} \frac{\sum_{i=1}^{n} i^{(2\tilde{a}-2\kappa)(1+\frac{\delta}{2})}}{n^{2(\tilde{a}-\beta)(1+\frac{\delta}{2})}} = 0$$

If $\tilde{\alpha} < 1$ we have to show the Lyapunov type condition

$$\frac{\int_0^t (1+R_{s-})^{-2\kappa} \phi_s^{-2} \int_{\mathfrak{G}_{s,t}^{\epsilon}} x^2 \ \nu^{M*}(\mathrm{d}s, \mathrm{d}x)}{(1+R_t)^{-2\beta} \phi_t^2} \xrightarrow{\mathbb{P}} 0$$

which is handled by following the preceding steps one by one. $\square$

# 9 Concluding Remarks

There are several good reasons to average leading algorithms. If for example in the unaveraged Kiefer-Wolfowitz algorithm (6.3) we choose $a_s = a(1 + R_s)^{-1}$, the additional assumption $a > \frac{1-2\gamma}{2\lambda_{\min}}$ has to be made. Actually we need assumptions concerning the smallest eigenvalue of the Hessian although even the regression function $f$ itself is unknown. Similar problems arise for the Robbins-Monro algorithm. This is one of the main disadvantages of unaveraged algorithms, which does not arise in averaged algorithms.

As mentioned in the introduction, the results concerning the companion algorithms keep valid if in (6.4), (6.5), (6.6) or (6.7) we replace the leading process $(Z_t)_{t \geqslant 0}$ by its averaged process $(\bar{Z}_t)_{t \geqslant 0}$. The same is true for the time-discrete and time-continuous special cases. Consequently replacing $Z$ by its averaged process $\bar{Z}$ does not seem to improve the asymptotic properties of the companion algorithms alone.

If we choose $\tilde{a}_s = \tilde{a}(1 + R_s)^{-1}$ in the companion algorithms, we also have to make a stricter assumption $\tilde{a} > 1/2$ on the constant and not only $\tilde{a} > 0$. But as its form is pretty simple, namely $\tilde{a} > 1/2$, and therefore independent of further knowledge of $f$, there is no disadvantage in having an unknown $f$. Actually it is easy to see, that an optimal choice is $\tilde{a} := 1$, as it reduces the variance in the asymptotic normality results. In an averaged companion algorithm choose $\tilde{a}_s = \tilde{a}(1 + R_s)^{-1+\epsilon}$ for arbitrary $\tilde{a} > 0$. It is expected that averaging yields optimal rates also with $\tilde{a} \in (0, 1/2]$.

This thesis only handled simple estimators for the parameters of interest as special cases. In 1997 Dippon and Renz [12] presented a modification of the gradient estimate in the Kiefer-Wolfowitz algorithm. It is an extension of ideas from Fabian [15]. With a $p$-times differentiable $f$ at $z^*$ they can obtain $n^{\frac{1}{2}(1-1/p)}(Z_n - z^*) \xrightarrow{\mathcal{D}} \mathrm{N}(\mu, \Sigma)$. The main idea is to achieve better rates by increasing the number of observations. It is possible to improve the almost sure rate of convergence and the asymptotic distribution properties of the companion algorithms by using more advanced estimators. For example estimators using more observations or a random design as described in the first part of this thesis. With the template of the generic algorithms this thesis should provide all required tools.

# A Appendix

By $\mathcal{P}$ we denote the space of predictable processes. Furthermore let $\mathcal{V}$ (and $\mathcal{V}^+$) be the (increasing) processes of finite variation. With $\mathcal{M}_{\text{loc}}$ we denote the set of local martingales. The following lemma is essential for almost-sure convergence results.

**Lemma A.1.1** (Generalized Robbins-Siegmund)**.** *Let $X \geqslant 0$ be a special semimartingale with $X = X_0 + A + M$ where $A \in \mathcal{V} \cap \mathcal{P}$ and $M \in \mathcal{M}_{\text{loc}}$. Furthermore let*

$$A \leqslant A^1 - A^2 \text{ with } A^1, A^2 \in \mathcal{V}^+ \cap \mathcal{P} \text{ and } A^1 - A \in \mathcal{V}^+$$

*be fulfilled. Then*

$$\left\{\int_0^\infty \frac{1}{1 + X_{s-}} \mathrm{d}A_s^1 < \infty\right\} \subseteq \{X \rightarrow\} \cap \{A_\infty^2 < \infty\} \text{ a.s.}$$

*Proof.* The lemma was originally stated in [22]. A detailed proof can be found in [37]. □

*Remark* A.1.1. Note that the notation $\{X \rightarrow\}$ denotes the set of all $\omega \in \Omega$ for which $X$ converges to a not further specified value. This is not necessarily the value we want to show $X$ converges to.

**Lemma A.1.2** (Generalized Toeplitz-Lemma)**.** *Let $X$ be a semimartingale and $L \in \mathcal{V}^+ \cap \mathcal{P}$ with $L_0 = 0$ then*

$$\left\{L_\infty = \infty\right\} \cap \left\{X_t \rightarrow x\right\} \subseteq \left\{(1 + L_t)^{-1} \int_0^t X_{s-} \mathrm{d}L_s \rightarrow x\right\} a.s.$$

*Proof.* As no proof of the result could be found in literature, it is given here. For all $\epsilon > 0$ there exists a $T_\epsilon$ such that $|X_{t-} - x| \leqslant \epsilon$ on the set $\{X \rightarrow\} \cap \{t > T_\epsilon\}$. Furthermore we can find a $T_1 > T_\epsilon$ such that $\sup\limits_{s \in [0, t \wedge T_\epsilon]} |X_{s-} - x| \frac{L_{t \wedge T_\epsilon}}{1 + L_{T_1}} < \epsilon$. Obviously

$$x(L_t - L_0) = \int_0^t x \mathrm{d}L_s$$

holds and therefore

$$x = \lim_{t \to \infty} \frac{x(L_t - L_0 + L_0 + 1)}{1 + L_t} = \lim_{t \to \infty} \frac{x(L_t - L_0)}{1 + L_t} + \lim_{t \to \infty} \frac{x(L_0 + 1)}{1 + L_t} = \lim_{t \to \infty} \frac{x(L_t - L_0)}{1 + L_t}$$

$$= \lim_{t\to\infty} \frac{1}{1+L_t} \int_0^t x \mathrm{d}L_s.$$

Then, for $t > T_1$,

$$\begin{aligned}
\left| \frac{1}{1+L_t} \int_0^t (X_{s-} - x) \mathrm{d}L_s \right| &\leqslant \frac{1}{1+L_t} \int_0^t |X_{s-} - x| \mathrm{d}L_s \\
&= \frac{1}{1+L_t} \int_0^{t\wedge T_\epsilon} |X_{s-} - x| \mathrm{d}L_s + \frac{1}{1+L_t} \int_{t\wedge T_\epsilon}^t |X_{s-} - x| \mathrm{d}L_s \\
&\leqslant \sup_{s\in[0,t\wedge T_\epsilon]} |X_{s-} - x| \frac{L_{t\wedge T_\epsilon} - L_0}{1 + L_{T_1}} + \mathbb{1}_{\{T_\epsilon < t\}} \cdot \epsilon \\
&< 2\epsilon.
\end{aligned}$$

□

**Lemma A.1.3** (Generalized Kronecker-Lemma)**.** *Let $X$ be a semimartingale and $L \in \mathcal{V}^+ \cap \mathcal{P}$ with $L_0 = 0$ then*

$$\left\{ L_\infty = \infty \right\} \cap \left\{ (1+L_t)^{-1} \int_0^t X_{s-} \mathrm{d}L_s \to \right\} \subseteq \left\{ \frac{X_t}{L_t} \to 0 \right\} a.s.$$

*Proof.* Kronecker's lemma follows directly from Toeplitz' lemma. Alternatively a direct proof can be found in the book of Liptser and Shiryayev [23, Lemma II.5.3]. □

**Theorem A.1.1** (Minkowski's inequality for integrals)**.** *Let $(S_1, \mu_1)$, $(S_2, \mu_2)$ be measure spaces and $f\colon S_1 \times S_2 \to \mathbb{R}$ a measurable function. Then*

$$\left( \int_{S_2} \left| \int_{S_1} f(x,y) \mathrm{d}\mu_1(x) \right|^p \mathrm{d}\mu_2(y) \right)^{\frac{1}{p}} \leqslant \int_{S_1} \left( \int_{S_2} \left| f(x,y) \right|^p \mathrm{d}\mu_2(y) \right)^{\frac{1}{p}} \mathrm{d}\mu_1(x)$$

*for $1 < p < \infty$.*

*Proof.* The proof can be found in the book of Hardy, Littlewood and Pólya [17, Chapter 6.13]. □

The following lemma is inspired by the ideas of Dippon and Walk [13].

**Lemma A.1.4.** *Let $R_t$ be deterministic with $R_0 = 0$, $R_t \geqslant 0$ monotonously increasing with*

$$\int_0^\infty (1+R_s)^{-1-\epsilon} \mathrm{d}R_s < \infty \quad \textit{for all } \epsilon > 0.$$

*Moreover let $(V_t)_{t\geqslant 0}$ be $\mathbb{R}^d$-valued, adapted with $\mathbb{E}\|V_t\|^2 < \infty$ and $\mathbb{E}\| \int_0^t (V_{s-} - \mathbb{E}(V_{s-})) \mathrm{d}R_s \|^2 = \mathcal{O}(R_t)$. Furthermore let $b_s = b((1+R_s)^{-\beta})$ as well as $b > 0$ and $\beta > 1/2$. Then*

$$\left\| \int_0^\infty b_s (V_{s-} - \mathbb{E}(V_{s-})) \mathrm{d}R_s \right\| < \infty \ a.s.$$

*Proof.* Integration by parts yields

$$
\begin{aligned}
&\int_0^t b_{s-}(V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s \\
&\qquad = b_t \int_0^t (V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s - \int_0^t \Big(\int_0^s (V_{u-} - \mathbb{E}(V_{u-}))\mathrm{d}R_u\Big)\mathrm{d}b_s \\
&\qquad\quad - \int_0^t \mathrm{d}\Big[b_\cdot, \int_0^\cdot (V_{\tau-} - \mathbb{E}(V_{\tau-}))\mathrm{d}R_\tau\Big]_s.
\end{aligned}
$$

Hence with Markov's inequality and Minkowski's inequality for integrals,

$$
\begin{aligned}
&\mathbb{E}\Big\| \int_0^\infty b_{s-}(V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s\Big\|^2 \\
&\quad \leqslant \Big(3b_\infty^2 \mathbb{E}\Big(\int_0^\infty (V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s\Big)^2 \\
&\qquad + 3\Big(\Big(\mathbb{E}\Big\|\int_0^\infty \Big(\int_0^s (V_{u-} - \mathbb{E}(V_{u-}))\mathrm{d}R_u\Big)\mathrm{d}b_s\Big\|^2\Big)^{\frac{1}{2}}\Big)^2 \\
&\qquad + 3\mathbb{E}\Big\|\int_0^\infty \mathrm{d}\Big[b_\cdot, \int_0^\cdot (V_{\tau-} - \mathbb{E}(V_{\tau-}))\mathrm{d}R_\tau\Big]_s\Big\|^2\Big) \\
&\quad \leqslant \Big(3b_\infty^2 \mathbb{E}\| \int_0^\infty (V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s\|^2 + 3\Big(\int_0^\infty \Big(\mathbb{E}\|\int_0^s (V_{u-} - \mathbb{E}(V_{u-}))\mathrm{d}R_u\|^2\Big)^{\frac{1}{2}}|\mathrm{d}b_s|\Big)^2 \\
&\qquad + 3\mathbb{E}\Big\|\big[b_\cdot, \int_0^\cdot (V_{\tau-} - \mathbb{E}(V_{\tau-}))\mathrm{d}R_\tau\big]_\infty\Big\|^2\Big) \\
&\quad \leqslant \Big(3b_\infty^2 \mathbb{E}\| \int_0^\infty (V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s\|^2 + 3\Big(\int_0^\infty \Big(\mathbb{E}\|\int_0^s (V_{u-} - \mathbb{E}(V_{u-}))\mathrm{d}R_u\|^2\Big)^{\frac{1}{2}}|\mathrm{d}b_s|\Big)^2 \\
&\qquad + 3\mathbb{E}\Big\|\big[\int_0^\cdot b\beta(1+R_\tau)^{-\beta-1}\mathrm{d}R_\tau, \int_0^\cdot (V_{\tau-} - \mathbb{E}(V_{\tau-}))\mathrm{d}R_\tau\big]_\infty\Big\|^2\Big) \\
&\quad \leqslant \Big(3b_\infty^2 R_\infty + 3\mathcal{C}\Big(\int_0^\infty \sqrt{R_s}|\mathrm{d}b_s|\Big)^2 \\
&\qquad + 3\mathbb{E}\Big\|\int_0^\infty b\beta(1+R_s)^{-\beta-1}(V_{s-} - \mathbb{E}(V_{s-}))\Delta R_s \mathrm{d}R_s^d\Big\|^2\Big) \\
&\quad \leqslant \Big(3b_\infty^2 R_\infty + 3\beta^2 b^2\Big(\int_0^\infty (1+R_s)^{-\beta-\frac{1}{2}}\mathrm{d}R_s\Big)^2\Big) \\
&\qquad + 3\int_0^\infty b^2\beta^2(1+R_s)^{-2\beta-2}\mathbb{E}\|(V_{s-} - \mathbb{E}(V_{s-}))\|^2(\Delta R_s)^2\mathrm{d}R_s^d\Big) \\
&\quad \leqslant \Big(3b_\infty^2 R_\infty + 3\beta^2 b^2\Big(\int_0^\infty (1+R_s)^{-\beta-\frac{1}{2}}\mathrm{d}R_s\Big)^2\Big) \\
&\qquad + \mathcal{C}\int_0^\infty (1+R_s)^{-2\beta-2}(\Delta R_s)^2\mathrm{d}R_s^d\Big) \\
&\quad \leqslant \Big(3b_\infty^2 R_\infty + 3\beta^2 b^2\Big(\int_0^\infty (1+R_s)^{-\beta-\frac{1}{2}}\mathrm{d}R_s\Big)^2\Big) \\
&\qquad + \mathcal{C}\int_0^\infty (1+R_s)^{-2\beta}\mathrm{d}R_s\Big)
\end{aligned}
$$

$< \mathcal{C} < \infty.$

$\square$

**Corollary A.1.1.** *Let the conditions of Lemma A.1.4 hold with* $\mathbb{E}\|\int_0^t (V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s\|^2 = \mathcal{O}(1)$. *Moreover assume an $\mathbb{R}$-valued process $(k_t)_{t\geqslant 0}$ to be predictable, for all $t \geqslant 0$ bounded from above and below, such that $b_s := k_s(1+R_{s-})^{-\beta}$, $\beta > 1/2$. Then*

$$\left\|\int_0^\infty b_s(V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s\right\| < \infty \text{ a.s.}$$

*Proof.* From the proof of Lemma A.1.4 it follows

$$\begin{aligned}
\mathbb{E}&\left\|\int_0^\infty b_{s-}(V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s\right\|^2 \\
&\leqslant \Bigg(3b_\infty^2\mathbb{E}\|\int_0^\infty (V_{s-} - \mathbb{E}(V_{s-}))\mathrm{d}R_s\|^2 + 3\Big(\int_0^\infty \Big(\mathbb{E}\|\int_0^s (V_{u-} - \mathbb{E}(V_{u-}))\mathrm{d}R_u\|^2\Big)^{\frac{1}{2}}|\mathrm{d}b_s|\Big)^2 \\
&\quad + 3\mathbb{E}\left\|\Big[b_\cdot, \int_0^\cdot (V_{\tau-} - \mathbb{E}(V_{\tau-}))\mathrm{d}R_\tau\Big]_\infty\right\|^2\Bigg). \qquad (A.1)
\end{aligned}$$

The first term on the right hand side of (A.1) clearly tends to zero. The second one is bounded by

$$\begin{aligned}
&\Big(\int_0^\infty \Big(\mathbb{E}\|\int_0^s (V_{u-} - \mathbb{E}(V_{u-}))\mathrm{d}R_u\|^2\Big)^{\frac{1}{2}}|\mathrm{d}b_s|\Big)^2 \\
&\qquad \leqslant \sup_{s\in[0,\infty]}\Big(\mathbb{E}\|\int_0^s (V_{u-} - \mathbb{E}(V_{u-}))\mathrm{d}R_u\|^2\Big)^{1/2}\int_0^\infty |\mathrm{d}b_s| < \infty.
\end{aligned}$$

Finally

$$\begin{aligned}
\mathbb{E}&\left\|\Big[b_\cdot, \int_0^\cdot (V_{\tau-} - \mathbb{E}(V_{\tau-}))\mathrm{d}R_\tau\Big]_\infty\right\|^2 \\
&\leqslant \int_0^\infty k_s^2(1+R_{s-})^{-2\beta-2}\mathbb{E}\|V_{s-} - \mathbb{E}(V_{s-})\|^2(\Delta R_s)^2\mathrm{d}R_s^d \\
&\leqslant \sup_{s\in[0,\infty]}\|k_s^2\mathbb{E}\|V_{s-} - \mathbb{E}(V_{s-})\|^2\|\int_0^\infty (1+R_s)^{-2\beta-2}(\Delta R_s)^2\mathrm{d}R_s^d \\
&\leqslant \mathcal{C}\int_0^\infty (1+R_s)^{-2\beta}\mathrm{d}R_s < \infty
\end{aligned}$$

completes the proof. $\square$

**Lemma A.1.5.** *Let $a_s := a(1+R_{s-})^{-\alpha}$ with $R_t \in \mathcal{V}^+ \cap \mathcal{P}$, $R_0 = 0$ and $\alpha \in (0,1]$. Then $\int_0^\infty a_s^2\Delta R_s\mathrm{d}R_s < \infty$ implies*

$$\int_0^\infty (1+R_{s-})^{-1-\epsilon}\mathrm{d}R_s < \infty \text{ for any } \epsilon > 0.$$

*Proof.* Itô's formula yields

$$\begin{aligned}
&\int_0^t (1+R_{s-})^{-1-\epsilon} \mathrm{d}R_s \\
&\quad = -\frac{1}{\epsilon}\left((1+R_t)^{-\epsilon} - (1+R_0)^{-\epsilon}\right) \\
&\quad\quad - \sum_{0<s\leqslant t} \left\{ -\frac{1}{\epsilon}\Big((1+R_s)^{-\epsilon} - (1+R_{s-})^{-\epsilon}\Big) - (1+R_{s-})^{-\epsilon-1}(R_s - R_{s-}) \right\}.
\end{aligned}$$

The first term tends to $\frac{1}{\epsilon}(1+R_0)^{-\epsilon}$. A Taylor expansion of the second term with $\zeta_s \in [0,1]$ and $t \to \infty$ yields

$$\begin{aligned}
&\sum_{0<s\leqslant\infty} \left\{ \frac{1}{\epsilon}\Big((1+R_s)^{-\epsilon} - (1+R_{s-})^{-\epsilon}\Big) + (1+R_{s-})^{-\epsilon-1}(R_s - R_{s-}) \right\} \\
&\qquad = \frac{1}{2}(\epsilon+1) \sum_{0<s\leqslant\infty} \frac{(\Delta R_s)^2}{(1+R_{s-}+\zeta_s \Delta R_s)^{2+\epsilon}} \\
&\qquad \leqslant \frac{1}{2}(\epsilon+1) \sum_{0<s\leqslant\infty} \left(\frac{\Delta R_s}{1+R_{s-}}\right)^2 \leqslant \frac{1}{2}(\epsilon+1) \sum_{0<s\leqslant\infty} \left(\frac{\Delta R_s}{(1+R_{s-})^{\alpha}}\right)^2 \\
&\qquad = \frac{1}{2}(\epsilon+1)\frac{1}{a^2} \int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s^d < \infty.
\end{aligned}$$

□

**Lemma A.1.6.** *Let $X$ be a semimartingale and $Y$ be a predictable process of finite variation. Then*

$$[X,Y]_t = \int_0^t \Delta Y_s \mathrm{d}X_s.$$

*Proof.* The proof is given in Jacod et al. [18, Proposition I.4.49]. □

The following conditions are needed to show an almost $L^2$-convergence rate of the Robbins-Monro algorithm. (C.f. Theorem 8.1.1.)

**Assumption A.1.1.**

- *$f\colon \mathbb{R}^d \to \mathbb{R}$ has a Lipschitz-continuous gradient.*
- *There exists an $z^*$ with $\nabla f(z^*) = 0$.*
- *The processes $(a_s)_{s\geqslant 0}$ and $(c_s)_{s\geqslant 0}$, which the statistician has to choose, are left-continuous and satisfy*

$$a_s, c_s > 0 \qquad\qquad a_s, c_s \downarrow 0$$

$$\int_0^\infty a_s \mathrm{d}R_s = \infty \qquad\qquad \int_0^\infty a_s c_s \mathrm{d}R_s < \infty.$$

- *For every $i \in \{1,\dots,d\}$ and $x \in \mathbb{R}^d$, we have*

$$\int_0^\infty \frac{a_s^2}{c_s^2}\frac{h_s^{ii}(Z_{s-})}{1+\|Z_{s-}\|^2}\mathrm{d}R_s < \infty, \text{ where } h_s^{ii}(z) := \frac{\mathrm{d}[\int_0^\cdot M_i(\mathrm{d}t,z)]_s}{\mathrm{d}R_s}.$$

- *If the process $(R_s)_{s\geqslant 0}$ is not continuous, then the following condition should also hold:*

$$\int_0^\infty a_s^2 \Delta R_s \mathrm{d}R_s^d < \infty.$$

**Theorem A.1.2** (Almost $L^2$-convergence rate of the Kiefer-Wolfowitz algorithm)**.** *We assume the existence of a positive, deterministic, monotonously increasing process $(S_t)_{t\geqslant 0}$ with $S_t \uparrow \infty$ and $S_t\|Z_t\| \to 0$ a.s. for $Z$ defined in the Kiefer-Wolfowitz algorithm* (6.3)*. Let Assumption A.1.1 be valid. Assume that $f$ is two or three times differentiable at $z^*$ with a continuous Hessian around $z^*$ and*

$$\underset{i,j\in\{1,\dots,d\}}{\forall}\ \underset{0<S<\infty}{\forall}\ \underset{0<K<\infty}{\exists}\ \|x\| \leqslant S \Rightarrow \sup_{t\in[0,\infty)} |h_t^{ij}(x)| \leqslant K \text{ a.s.}$$

*as well as*

$$\int_0^\infty \frac{a_s^2}{c_s^2}\mathrm{d}R_s < \infty \text{ a.s.}$$

*In the case $\alpha < 1$, we assume that the Hessian of $f$ is positive definite at $z^*$, and in the special, yet important, case $\alpha = 1$ we further stipulate $\lambda_{\min} > \frac{1-2\gamma}{2a}$, where $\lambda_{\min}$ denotes the minimum of the eigenvalues of the Hessian. Then, for all $\epsilon > 0$, there exists a process $(Y_t)_{t\geqslant 0}$ such that*

$$\mathbb{P}\left[\underset{t\geqslant 0}{\forall}\ Y_t = Z_t\right] \geqslant 1-\epsilon$$

*and*

$$\mathbb{E}\|Y_t - z^*\|^2 = \mathcal{O}((1+R_t)^{\tilde\beta})$$

*with*

$$\tilde\beta := \max\{1-\alpha-2(p-1)\gamma, 1-2\alpha+2\gamma\} \text{ , if } f \text{ is } p \text{ times differentiable at } z^*$$

*where $p \in \{2,3\}$.*

*Proof.* This theorem has been shown by Schnizler [37, Theorem 3.1]. □

**Theorem A.1.3** (Davis' inequality)**.** *Let $T$ be a stopping time and $M \in \mathcal{M}_{\mathrm{loc}}$ with $M_0 = 0$. Then there exist constants $c$ and $C$ that are independent of $T$ such that*

$$c\mathbb{E}[M,M]_T^{1/2} \leqslant \mathbb{E}\sup_{s\leqslant T}|M_s| \leqslant C\mathbb{E}[M,M]_T^{1/2}.$$

*Proof.* The proof can be found in the book of Liptser and Shiryayev [23, Ch. I.5]. □

**Theorem A.1.4** (Special case of the Lenglart-Rebolledo inequality)**.** *Let $X$ and $Y$ non-negative, $\mathbb{F}$-adapted and càdlàg (right continuous with left limits) processes, and $X_0 = Y_0 = 0$, $Y \in \mathcal{V}^+$. Let $Y$ dominate the process $X$ in the sense that for each stopping time $\tau$*

$$\mathbb{E}X_\tau \leqslant \mathbb{E}Y_\tau.$$

*If, in addition the process $Y$ is predictable, then for each stopping time $T$ with $\mathbb{P}(T < \infty) = 1$ a.s. and all numbers $a > 0$, $b > 0$*

$$\mathbb{P}\left(\sup_{t \leqslant T} X_t \geqslant a\right) \leqslant \frac{1}{a}\mathbb{E}\big[Y_T \wedge b\big] + \mathbb{P}(Y_T \geqslant b).$$

*Proof.* The proof can be found in [23, Theorem 3, p. 66]. □

**Theorem A.1.5.** *Let $A$ be a process of finite variation with $A_0 = 0$ with locally integrable total variation. Then there exists one and only one predictable process $\tilde{A}$ of finite variation with $A_0 = 0$ with locally integrable total variation, such that $A - \tilde{A} \in \mathcal{M}_{\mathrm{loc}}$ or equivalently $\mathbb{E}A_\tau = \mathbb{E}\tilde{A}_\tau$ for any stopping time $\tau$.*

*Proof.* The proof can be found in [23, Theorem I.6.3]. □

*Remark* A.1.2. A process $\tilde{A}$ from the previous theorem is also called compensator of $A$.

**Definition A.1.1.** Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a stochastic basis with $\mathbb{F} = (\mathcal{F}_t)_{t \geqslant 0}$ and $\mathcal{P}(\mathbb{F})$. The predictable $\sigma$-field is the $\sigma$-field on $\Omega \times \mathbb{R}_+$ that is generated by all right-continuous processes (considered as mappings on $\Omega \times \mathbb{R}_+$). By $(E, \mathcal{E})$ we denote a Lusin space, i.e. $E$ is a Borel subspace of a compact metric space, and $\mathcal{E}$ the corresponding Borel $\sigma$-algebra. Furthermore we use the notations $\tilde{\Omega} := \Omega \times \mathbb{R}_+ \times E$, $\tilde{\mathcal{F}} := \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$ and $\tilde{\mathcal{P}} := \mathcal{P} \otimes \mathcal{E}$.

As a random measure on $\mathbb{R}_+ \times E$ we define the family $\mu = \{\mu(\omega; \mathrm{d}t, \mathrm{d}x) \mid \omega \in \Omega\}$ of non-negative measures $\mu(\omega; .)$ on $(\mathbb{R}_+ \times E, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E})$ where $\omega \in \Omega$, such that $\mu(\omega; \{0\} \times E) = 0$.

Let $X = X(\omega, t, x)$ be a non-negative $\tilde{\mathcal{F}}$-measurable function. Then for $\omega \in \Omega$ and $t \in \mathbb{R}_+$ we can define the Lebesgue integral $(X * \mu)_t = \int_{[0,t] \times E} X(\omega, s, x)\mu(\omega; \mathrm{d}s, \mathrm{d}x)$.

A random measure $\mu$ is called predictable, if the process $X * \mu$ is predictable for every predictable function $X$.

Let $X$ be an adapted càdlàg $\mathbb{R}^d$-valued process. Then, according to [18, Proposition II.1.16], we can define an integer-valued random measure, called jump measure, on $\mathbb{R}_+ \times \mathbb{R}^d$ by setting

$$\mu(\omega; \mathrm{d}t, \mathrm{d}x) = \sum_{s > 0} \mathbb{1}_{\{\Delta X_s(\omega) \neq 0\}} \epsilon_{(s, \Delta X_s(\omega))}(\mathrm{d}t, \mathrm{d}x),$$

where $\epsilon_a$ denotes the Dirac measure at point $a$.

For every measure $\mu$ and probability measure $\mathbb{P}$ we can define the Doléans measure $M_\mu^{\mathbb{P}}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}})$, by

$$M_\mu^{\mathbb{P}}(\mathrm{d}\omega, \mathrm{d}t, \mathrm{d}x) := \mathbb{P}(\mathrm{d}\omega)\mu(\omega; \mathrm{d}t, \mathrm{d}x).$$

For every non-negative $\tilde{\mathcal{F}}$-measurable function $X(\omega, t, x)$ define

$$M_\mu^{\mathbb{P}}(X) := \int_{\tilde{\Omega}} X(\omega, t, x) M_\mu^{\mathbb{P}}(\mathrm{d}\omega, \mathrm{d}t, \mathrm{d}x) = \mathbb{E}(X * \mu)_\infty.$$

Furthermore we define that $\mu$ belongs to $\tilde{\mathcal{V}}_{\mathcal{P}}^+$ if $M_\mu^{\mathbb{P}}(\mathbb{1}_{\tilde{\Omega}}) < \infty$, $\tilde{\Omega}_n \in \tilde{\mathcal{P}}$ and $\tilde{\Omega}_n \uparrow \tilde{\Omega}$. A predictable random measure $\nu$ is called compensator of a random measure $\tilde{\mathcal{V}}_{\mathcal{P}}^+$ if for any non-negative $\tilde{\mathcal{P}}$-measurable function $X = X(\omega, t, x)$ we have $M_\mu^{\mathbb{P}}(X) = M_\nu^{\mathbb{P}}(X)$. According to [23, Theorem I.3.2.1], each random measure $\mu \in \tilde{\mathcal{V}}_{\mathcal{P}}^+$ possesses the unique ($\mathbb{P}$-a.s.) compensator $\nu$.

Let $X = (X_t, \mathcal{F}_t)$ and $X^n = (X_t^n, \mathcal{F}_t^n)$, $n \geqslant 1$, be semimartingales, $S$ a non-empty subset of $\mathbb{R}_+$. Then the expression $X^n \xrightarrow{\mathrm{d}_f(S)} X$ denotes the weak converge of a sequence of distributions of vectors $(X_{t_1}^n, \dots, X_{t_m}^n)$, $n \geqslant 1$ to the distribution of a vector $(X_{t_1}, \dots, X_{t_m})$ for each finite subset $\{t_1, \dots, t_m\} \in S$, while the expression $X^n \xrightarrow{\mathrm{d}_f(S)} X$ ($\mathcal{G}$-stable) means that

$$\lim_n \mathbb{E}\xi h(X_{t_1}^n, \dots, X_{t_m}^n) = \mathbb{E}\xi h(X_{t_1}, \dots, X_{t_m})$$

holds for every bounded function $h(x_1, \dots, x_m)$ that is continuous in all variables $(x_1, \dots, x_m)$ and for each bounded $\mathcal{G}$-measurable random variable $\xi$, where $\mathcal{G}$ is a sub-$\sigma$-algebra of $\mathcal{F}$. If $S$ consists of a single point $S = \{t_1\}$, then instead of $\xrightarrow{\mathrm{d}_f(S)}$ we write $\xrightarrow{\mathcal{D}}$, which means convergence of random variables in distribution.

**Theorem A.1.6** (Central Limit Theorem)**.** *Let $X^n = (X_t^n, \mathcal{F}_t^n) \in \mathcal{M}_{\mathrm{loc}}^2$, $X_0^n = 0$, $n \geqslant 1$, $S$ a nonempty subset of $\mathbb{R}_+$, $\mathcal{G} \subseteq \bigcap_{n \geqslant 1} \mathcal{F}_0^n$ and conditions*

*(I)* $x^2 \mathbb{1}_{[|x| > \delta]} * \nu_t^n \xrightarrow{\mathbb{P}} 0$*, for all $\delta \in (0, 1]$ and all $t \in S$ (Lindeberg-type condition),*

*(II)* $\lceil X^n \rceil_t \xrightarrow{\mathbb{P}} \lceil X \rceil_t$*, for all $t \in S$ (variance-type condition),*

*hold. Then*

$$X^n \xrightarrow{\mathrm{d}_f(S)} X \quad (\mathcal{G}\text{-stable}).$$

*Proof.* The proof can be found in [23, Theorem II.5.5.4]. □

**Theorem A.1.7.** *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space with a filtration $\mathcal{F} = (\mathcal{F}_t)_{t \geqslant 0}$ satisfying the usual conditions. Moreover let $\hat{M} = (\hat{M}_t)_{t \geqslant 0}$ be a càdlàg $d$-dimensional martingale and $(\hat{a}_t)_{t \geqslant 0}$ a family of invertible $d \times d$ matrices. Let the following conditions hold, as $t \to \infty$:*

*(a)* $\|\hat{a}_t\| \to 0$

*(b)* $\mathbb{E}\left(\sup_{0\leqslant s\leqslant t}\|\hat{a}_t\Delta\hat{M}_s\|\right)\to 0$

*(c)* $\hat{a}_t[\hat{M}_\cdot,\hat{M}_\cdot]_t\hat{a}_t^T\xrightarrow{\mathbb{P}}\Sigma$ .

*Then the random vector* $\hat{a}_t\hat{M}_t$ *converges* $\mathcal{A}$*-stable to the Gaussian distribution* $\mathrm{N}(0,\Sigma)$ *as* $t\to\infty$*.*

*Proof.* Theorem and proof are given in a paper of Crimaldi and Pratelli [7, Theorem 2.2]. □

# Bibliography


[1] S. Achour, R. Sarpeshkar, and M. C. Rinard. Configuration synthesis for programmable analog devices with Arco. In *Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation*, pages 177–193. ACM, 2016.

[2] S. Bhatnagar, H. Prasad, and L. Prashanth. *Stochastic Recursive Algorithms for Optimization: Simultaneous Perturbation Methods*. Lecture Notes in Control and Information Sciences. Springer London, 2012.

[3] J. R. Blum. Multidimensional stochastic approximation methods. *Ann. Math. Statistics*, 25:737–744, 1954.

[4] G. Cauwenberghs. An analog VLSI recurrent neural network learning a continuous-time trajectory. *IEEE Transactions on Neural Networks*, 7:346–361, 1996.

[5] G. Cauwenberghs. Analog VLSI stochastic perturbative learning architectures. *Analog Integrated Circuits and Signal Processing*, 13(1):195–209, 1997.

[6] H. Chen. Lower rate of convergence for locating a maximum of a function. *The Annals of Statistics*, 16(3):1330–1334, 1988.

[7] I. Crimaldi and L. Pratelli. Convergence results for multivariate martingales. *Stochastic Processes and their Applications*, 115(4):571 – 577, 2005.

[8] R. Daniel, J. R. Rubens, R. Sarpeshkar, and T. K. Lu. Synthetic analog computation in living cells. *Nature*, 497(7451):619–623, 2013.

[9] R. Daniel, S. S. Woo, L. Turicchia, and R. Sarpeshkar. Analog transistor models of bacterial genetic circuits. In *Biomedical Circuits and Systems Conference (BioCAS), 2011 IEEE*, pages 333–336. IEEE, 2011.

[10] J. Dippon. Accelerated randomized stochastic optimization. *The Annals of Statistics*, 31(4):1260–1281, 2003.

[11] J. Dippon and J. Renz. Weighted means of processes in stochastic approximation. *Math. Methods Statist.*, 5(1):32–60, 1996.

[12] J. Dippon and J. Renz. Weighted means in stochastic approximation of minima. *SIAM J. Control Optim.*, 35(5):1811–1827, 1997.

[13] J. Dippon and H. Walk. The averaged Robbins-Monro method for linear problems in a banach space. *Journal of Theoretical Probability*, 19(1):166–189, 2006.

[14] M. Duflo. *Random Iterative Models*. Applications of Mathematics : Stochastic Modelling and Applied Probability. Springer, 1997.

[15] V. Fabian. Stochastic approximation of minima with improved asymptotic speed. *The Annals of Mathematical Statistics*, 38(1):191–200, 1967.

[16] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[17] G. Hardy, J. Littlewood, and G. Pólya. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, 1952.

[18] J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 2000.

[19] S. Karlin and H. Taylor. *A First Course in Stochastic Processes*. Elsevier Science, 2012.

[20] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952.

[21] H. J. Kushner and D. S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer, New York; Heidelberg, 1978.

[22] N. Lazrieva, T. Sharia, and T. Toronjadze. Semimartingale stochastic approximation procedure and recursive estimation. *J. Math. Sci. (N. Y.)*, 153(3):211–261, 2008. Martingale theory and its application.

[23] R. Liptser and A. Shiryayev. *Theory of Martingales*. Kluwer Academic Publisher. Dordrecht, Boston, London, 1986.

[24] L. Ljung, G. Pflug, and H. Walk. *Stochastic Approximation and Optimization of Random Systems*. DMV Seminar. Springer Verlag NY, 1992.

[25] Y. Maeda and T. Kusuhashi. An analog neural network system with learning capability using simultaneous perturbation. *IEICE TRANSACTIONS on Information and Systems*, 82(12):1627–1633, 1999.

[26] Y. Maeda and M. Wakamura. Simultaneous perturbation learning rule for recurrent neural networks and its FPGA implementation. *IEEE Transactions on Neural Networks*, 16(6):1664–1672, 2005.

[27] A. V. Melnikov. Stochastic approximation procedures for semimartingales. In *Statistics and Control of Random Processes (Russian) (Preila, 1987)*, pages 147–156. "Nauka", Moscow, 1989.

[28] A. Mokkadem and M. Pelletier. A companion for the Kiefer-Wolfowitz-Blum stochastic approximation algorithm. *The Annals of Statistics*, 35(4):1749–1772, 2007.

[29] M. B. Nevel'son and R. Z. Has'minskiĭ. *Stochastic Approximation and Recursive Estimation*. Am. Math. Soc., Providence, Rhode Isl., 1973.

[30] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Hochschultext / Universitext. Springer, 2003.

[31] B. T. Polyak. A new method of stochastic approximation type. *Avtomat. i Telemekh.*, (7):98–107, 1990.

[32] P. E. Protter. *Stochastic Integration and Differential Equations: Version 2.1*. Springer, Berlin; Heidelberg, 2. ed., corr. 3. print. edition, 2005.

[33] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951.

[34] A. E. Rodkina and A. V. Melnikov. Martingale approach to the procedures of stochastic approximation. *Front. Pure Appl. Probab.*, 1:165–182, 1993.

[35] D. Ruppert. A Newton-Raphson version of the multivariate Robbins-Monro procedure. *The Annals of Statistics*, 13(1):236–245, 03 1985.

[36] J. Sacks. Asymptotic distribution of stochastic approximation procedures. *The Annals of Mathematical Statistics*, 29(2):373–405, 1958.

[37] P. Schnizler. *Kiefer-Wolfowitz Type Stochastic Approximation with Semimartingales*. PhD thesis, Universität Stuttgart, 2010.

[38] J. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Automat. Control*, 37(3):332–341, 1992.

[39] J. C. Spall. A one-measurement form of simultaneous perturbation stochastic approximation. *Automatica*, 33(1):109–112, 1997.

[40] J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Wiley, Hoboken, NJ, 2003.

[41] J. C. Spall, S. Member, and J. A. Cristion. Model-free control of nonlinear stochastic systems with discrete-time measurements. *IEEE Transactions on Automatic Control*, 43:1198–1210, 1998.

[42] E. Valkeila and A. V. Melnikov. Martingale models of stochastic spproximation and their convergence. *Theory of Probability & Its Applications*, 44(2):333–360, 2000.

[43] J.-G. Wang. The asymptotic behavior of locally square integrable martingales. *The Annals of Probability*, 23(2):552–585, 1995.

[44] X. Xing and M. Damodaran. Assessment of simultaneous perturbation stochastic approximation method for wing design optimization. *Journal of Aircraft*, 39(2):379–381, 2002.

[45] X. Q. Xing and M. Damodaran. Application of simultaneous perturbation stochastic approximation method for aerodynamic shape design optimization. *AIAA journal*, 43(2):284–294, 2005.

# Nomenclature

| | |
|---|---|
| $D = (D_t)_{t\geqslant 0}$ | A randomization process or a deterministic disturbance function |
| $W = (W_t)_{t\geqslant 0}$ | Brownian motion |
| $\mathcal{M}_{\mathrm{loc}}$ | local martingales |
| $\mathcal{M}^2_{\mathrm{loc}}$ | locally square integrable martingales |
| $\breve{M}$ | a locally square integrable martingale representing the observation noise in [Ker-Rand-1] and [Ker-Rand-2] |
| $\langle \cdot, \cdot \rangle$ | inner product of the Euclidean space $\mathbb{R}^d$ |
| $[X]_t$ | quadratic variation of the process $X$ |
| $\\| \cdot \\|$ | norm of the Euclidean space $\mathbb{R}^d$ |
| $[X.Y]_t$ | covariation of the processes $X$ and $Y$ |
| $\lceil X \rceil_t$ | predictable quadratic variation of the process $X$ |
| $\lceil X, Y \rceil_t$ | predictable covariation of the processes $X$ and $Y$ |
| $M^c$ | purely continuous part of the local martingale $M$ (note $M^c \perp M^d$) |
| $M^d$ | purely discontinuous part of the local martingale $M$ (note $M^c \perp M^d$) |
| $R^c$ | continuous part of the process $R$ |
| $R^d_t$ | sum of all jumps of the process $R = (R_s)_{s\geqslant 0}$ up to time $t$ |
| $\Delta X_t$ | jump height of the process $X = (X_s)_{s\geqslant 0}$ at time $t$ |
| $\{X \to\}$ | the set of all events such that $X_\infty$ exists and is a finite random variable |
| $X_{t-}$ | left continuous version of $X_t$, whereas $X$ is a process |
| $\nabla f(x)$ | gradient of the function $f$ at $x$ |
| $\mathcal{E}(M)$ | stochastic exponential of the process $M$ |
| $o(\cdot)$ | Landau symbol |
| $\mathcal{O}(\cdot)$ | Landau symbol |
| $\mathbb{E}$ | expectation value |
| $\mathbb{P}$ | probability measure |
| $Z^T_t$ | process $Z = (Z_t)_{t\geqslant 0}$ stopped at time $T$ |

| | |
|---|---|
| $\lfloor \cdot \rfloor$ | Gauss bracket |
| max | maximum |
| min | minimum |
| sup | supremum |
| inf | infimum |
| $\xrightarrow{\mathbb{P}}$ | convergence in probability |
| $\xrightarrow{\mathcal{D}}$ | convergence in distribution |
| $\simeq$ | asymptotically equal |
| $\mathcal{P}$ | set of predictable processes |
| $\mathcal{V}$ | set of real-valued processes that are càdlàg, adapted, starting at zero with paths of finite variation on compacts |
| $\mathcal{V}^+$ | set of processes belonging to $\mathcal{V}$ with non-decreasing paths |

# Declaration

Hiermit erkläre ich, dass ich die vorliegende Arbeit

*Randomization and Companion Algorithms in Stochastic Approximation with Semimartingales*,

selbstständig und nur unter Verwendung der angegebenen Hilfmittel und Literatur verfasst habe.

Stuttgart, 9. Juni 2018

Timo Pfrommer