DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Solving optimal stopping problems by empirical dual optimization and penalization

D. Belomestny

Preprint 110

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Solving optimal stopping problems by empirical dual optimization and penalization

D. Belomestny

Preprint 110

Solving optimal stopping problems by empirical dual optimization and penalization

Denis Belomestny

Duisburg-Essen University, Duisburg, Germany

In this article we propose a novel simulation-based algorithm for solving optimal stopping problems in discrete and continuous time. The algorithm involves the optimization of a genuinely penalized dual objective functional over a class of adapted martingales. The convergence analysis of the proposed algorithm reveals that the variance of the resulting estimate can be bounded from above by a multiple of the smallest approximation error within the class of martingales we optimize over, i.e., the better are the approximation properties of the martingale class, the less paths are needed to estimate the value of the underlying optimal stopping problem. The latter rather attractive variance "self-reduction" property not shared with by other existing simulation-based algorithms, is illustrated by a numerical example.

Keywords: optimal stopping, simulation-based algorithms, functional optimization, empirical variance, concentration inequalities, self-normalized processes

1 Introduction

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathrm{P})$ be a standard filtered probability space and let Z_t be a nonnegative adapted process. Consider the following optimal stopping problem:

$$Y^* = \sup_{\tau\in\mathcal{T}[0,T]} \mathrm{E}[Z_\tau], \tag{1}$$

where $\mathcal{T}[0,T]$ is the set of stopping times taking values in $[0,T]$ for some $T>0$. Solving the optimal stopping problem (1) is straightforward in low dimensions. However many problems arising in practice have high dimensions, and these applications have motivated the development of simulation-based algorithms for optimal stopping problems. Unfortunately, simulation-based algorithms often lead to estimates with high variance and consequently have rather slow convergence rates. To exemplify the problem consider first a much simpler situation where the expected value $H = \mathrm{E}[Z_T]$ needs to be approximated. The standard Monte Carlo estimate

$$H_n = \frac{1}{n}\sum_{j=1}^{n} Z_T^{(n)},$$

where $Z_T^{(1)}, \ldots, Z_T^{(n)}$ are i.i.d. copies of the r.v. Z_T, has the variance $\mathrm{Var}[Z_T]/n$. As a result, the convergence rates of the estimate H_n are of oder $\sqrt{\mathrm{Var}[Z_T]/n}$ and the only way to speed up the convergence is to reduce the variance of the r.v. Z_T. This is why variance reduction plays such an important role in the literature on the Monte Carlo methods. A standard approach to variance reduction is the so-called method of control variates where one tries to find a zero-mean random variable ξ that minimizes the variance $\mathrm{Var}[Z_T - \xi]$. Then the unbiased estimate

$$H_{\xi,n} = \frac{1}{n} \sum_{j=1}^{n} (Z_T^{(n)} - \xi^{(n)})$$

of H has the variance $\mathrm{Var}[Z_T - \xi]/n$ which is usually much smaller than $\mathrm{Var}[Z_T]/n$ leading to faster convergence rates. In order to find a r.v. ξ with the desired properties one can consider the optimization problem

$$\inf_{\xi \in \Xi} \mathrm{Var}[Z_T - \xi], \tag{2}$$

where the infimum is taken over a class Ξ of random variables ξ with $\mathrm{E}\,\xi = 0$. Unfortunately, the variance of the r.v. $Z_T - \xi$ can not be usually computed in a closed form and has to be approximated using n replications $Z^{(1)}, \ldots, Z^{(n)}$ of the r.v. Z_T. Ultimately, we are interested in constructing a simulation-based algorithm that uses $Z^{(1)}, \ldots, Z^{(n)}$ as an input and delivers a r.v. ξ_n fulfilling the inequality

$$\mathrm{Var}[Z_T - \xi_n] \leq C \inf_{\xi \in \Xi} \mathrm{Var}\left[Z_T - \xi\right] \tag{3}$$

with high probability for some constant $C > 0$. To achieve this goal, consider the empirical optimization problem

$$\inf_{\xi \in \Xi} V_n\left(Z_T - \xi\right), \tag{4}$$

where for any random variable ζ, $V_n(\zeta)$ stands for the empirical variance of ζ based on n replications of ζ :

$$V_n(\zeta) = \frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} (\zeta^{(i)} - \zeta^{(j)})^2.$$

Using the methods developed in this paper, one can prove that, under some conditions, any solution ξ_n of (4) would fulfill (3) with high probability (depending on n and C). Turning back to the original stopping problem (1), one may wonder if an optimization problem similar to (4) can be formulated. It can be indeed constructed by using the so-called dual representation for the optimal stopping problems (see Rogers (2002) or Haugh and Kogan (2004)) implying that the optimal solution Y^* of (1) can be represented in the form

$$Y^* = \inf_{M \in \mathcal{A}} \mathrm{E}\left[\sup_{t \in [0,T]} (Z_t - M_t)\right], \tag{5}$$

where $\mathcal{A}$ is the class of all adapted martingales. Hence for any given adapted martingale M, $Y(M) = \mathrm{E}\left[\sup_{t\in[0,T]}(Z_t - M_t)\right]$ is an upper bound for Y^* and it is natural to choose M in such a way that $Y(M)$ is minimized. On the other hand, we are interested in martingales M leading to the random variable $\sup_{t\in[0,T]}(Z_t - M_t)$ with a lower variance. By compromising both requirements, one ends up with the optimization problem

$$\inf_{M\in\mathcal{A}} \left\{ \mathrm{E}\left[\sup_{t\in[0,T]}(Z_t - M_t)\right] + \lambda\sqrt{\mathrm{Var}\left[\sup_{t\in[0,T]}(Z_t - M_t)\right]} \right\}, \tag{6}$$

where λ is a non-negative number determining the degree of penalization by the variance. One solution of the optimization problem (6) is well known; it is the martingale part M_t^* (Doob martingale) of the Doob-Meyer decomposition of the supermartingale

$$Y_t^* = \sup_{\tau\in\mathcal{T}[t,T]} \mathrm{E}[Z_\tau|\mathcal{F}_t], \tag{7}$$

since $Y^* = \sup_{t\in[0,T]}(Z_t - M_t^*)$ with probability 1. In fact, there are *infinitely many* martingales M fulfilling $Y^* = \sup_{t\in[0,T]}(Z_t - M_t)$ almost surely (see Schoenmakers (2011)). Simulate now n paths of the process (Z_t, M_t) on $[0,T]$. Fixing a set of martingales $\mathfrak{M} \subset \mathcal{A}$ and replacing the true quantities in (6) by their empirical counterparts, we arrive at the following empirical optimization problem

$$M_n = \operatorname*{arginf}_{M\in\mathfrak{M}} \left(\frac{1}{n}\sum_{j=1}^n Z^{(j)}(M) + \lambda\sqrt{V_n(M)} \right), \quad \lambda > 0, \tag{8}$$

where $Z^{(j)}(M) = \sup_{s\in[0,T]}(Z_s^{(j)} - M_s^{(j)})$ and

$$V_n(M) = \frac{1}{n(n-1)} \sum_{1\le i<j\le n} (Z^{(i)}(M) - Z^{(j)}(M))^2.$$

Let us compare our method based on (8) with the so-called martingale duality approach (see, e.g., Glasserman (2003) or Kohler (2010) for description). The latter approach aims at approximating the Doob martingale M^* and then uses this approximation to compute an upper bound on Y^* by Monte Carlo. There are several methods towards approximating the Doob martingale M^*. The early paper of Andersen and Broadie (2004) used, for example, sub-simulation to approximate M^* and in this way quite tight bounds on both sides for a number of test examples were obtained. The theoretical analysis of this algorithm was carried out in Kohler et al. (2010). The recent work of Desai, Farias and Moallemi (2010) uses optimization and sub-simulation to approximate M^* and Y_t^* simultaneously in an efficient way. Another type of algorithms that does not involve sub-simulation, was suggested in Belomestny et al. (2007), where an approximation for the martingale M^* was constructed using the martingale representation theorem and some approximation of the true price process. Our current approach, however, has several important advantages. First, it delivers "true" upper bound without use of sub-simulation, thus resulting in a non-nested Monte Carlo. Second, it does not exclusively concentrate on finding Doob martingale and takes advantage of the richness of the class $\mathcal{A}^*$ of adapted martingales satisfying

$$Y^* = \sup_{t\in[0,T]} (Z_t - M_t), \quad \text{a.s..} \tag{9}$$

The most useful feature of our algorithm, not shared by other simulation-based approaches, is that the variance of the r.v $Z(M_n) = \sup_{s\in[0,T]}(Z_s - M_{n,s})$ is, with high probability, bounded by a multiple of the r.v.

$$\inf_{M\in\mathfrak{M},\, M'\in\mathcal{A}^*} d(M, M'),$$

where d is a deterministic metric on $\mathcal{A}$. The above property implies that the variance of $Z(M_n)$ can be made arbitrary small by considering classes of martingales $\mathfrak{M}$ with better approximation properties with respect to the solution class $\mathcal{A}^*$. The last but not the least, our approach is applicable to the case of continuous optimal stopping problems, as it does not involve regression at each discretization step as in other approaches based on dynamic programming formulation.

The paper is organized as follows. In Section 2 the main results concerning the properties of the estimate M_n are formulated and discussed. We also present a rather generic class of examples showing that without use of penalization one would get estimates with significantly larger variance. In Section 3 we discuss how to build up a class of martingales with good approximation properties using the so-called martingale representation. The Section 4 contains a numerical example illustrating our approach. Finally, in Section 6 the proofs of the main results together with some auxiliary results are collected. In particular, we derive a novel concentration inequality for the empirical process $V_n(M) - V(M)$ over parameterized classes of martingales.

2 Main results

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathrm{P})$ be a filtered probability space and let (Ψ, ρ) be a metric space. Furthermore, let $\mathcal{M} = \{M(\psi) : \psi \in \Psi\}$ be a family of adapted continuous local martingales defined on $(\Omega, \mathcal{F}, \mathrm{P})$.

Definition 2.1. A quadratic ρ-modulus $\|\mathcal{M}\|_\rho$ of a family $\mathcal{M} = \{M(\psi) : \psi \in \Psi\}$ of continuous local martingales is defined as an $[0, T] \cap \{\infty\}$-valued stochastic process $t \to \|\mathcal{M}\|_{\rho,t}$ given by

$$\|\mathcal{M}\|_{\rho,t} = \sup_{\psi,\phi\in\Psi} \frac{\sqrt{\langle M(\psi) - M(\phi)\rangle_t}}{\rho(\psi, \phi)}, \quad t \in [0, T],$$

where $\langle M\rangle$ stands for the quadratic variation process of the continuous local martingale M.

For a given subspace $\widetilde{\Psi}$ of the metric space (Ψ, ρ) denote by $N(\varepsilon, \widetilde{\Psi}, \rho)$ the smallest number of closed balls, with ρ-radius $\varepsilon > 0$, which cover the set $\widetilde{\Psi}$ and set

$$J(\delta) = \int_0^\delta \sqrt{\log[1 + N(\varepsilon, \widetilde{\Psi}, \rho)]}\, d\varepsilon$$

for all $\delta > 0$. Denote also by $\mathcal{M}^* = \{M(\psi) : \psi \in \Psi^*\}$ a subset of $\mathcal{M}$ containing all martingales M that fulfill (9). In the sequel we shall assume that the family $\mathcal{M}$ is rich enough so that $\mathcal{M}^*$ is not empty. Let us now formulate the main result on the convergence of Y_n.

Theorem 2.2. *Let $\mathfrak{M} = \{M(\psi) : \psi \in \widetilde{\Psi}\}$ be a family of continuous local martingales satisfying $\|\mathfrak{M}\|_{\rho,T} \leq \Theta$ almost surely, for some finite Θ. Let also ψ^* be an element of Ψ^* such that $\rho(\psi, \psi^*) \leq \sigma$ for all $\psi \in \widetilde{\Psi}$ and some $\sigma < \infty$. Set*

$$\mathfrak{C} = \mathfrak{C}(\widetilde{\Psi}) = \int_0^\sigma \varepsilon^{-1} J(\varepsilon) \sqrt{\log[1 + N(\varepsilon, \widetilde{\Psi}, \rho)]}\, d\varepsilon$$

and assume that $\mathfrak{C} < \infty$. Fix some $0 < \delta < 1$ and $\varkappa > 0$ with $J(1)\log(1/\delta) \leq \sqrt{n}$ and define

$$M_n = \operatorname*{arginf}_{M \in \mathfrak{M}} \left(\frac{1}{n} \sum_{j=1}^{n} Z^{(j)}(M) + (\varkappa + \lambda_n(\delta/4)) \sqrt{V_n(M)} \right) \tag{10}$$

with $\lambda_n(\alpha) = 4(2\sqrt{2\log(2/\alpha)} + \mathfrak{C})/\sqrt{n}$ for any $\alpha > 0$. Then it holds for some constant $C > 0$ with probability at least $1 - \delta$

$$0 \leq Y(M_n) - Y^* \quad \leq \quad C(\varkappa + 2\lambda_n(\delta/4)) \inf_{\psi \in \widetilde{\Psi}} \mathcal{R}(\psi, \psi^*), \tag{11}$$

$$\sqrt{V(M_n)} \quad \leq \quad C\left(1 + \frac{2\lambda_n(\delta/4)}{\varkappa}\right) \inf_{\psi \in \widetilde{\Psi}} \mathcal{R}(\psi, \psi^*), \tag{12}$$

where $Y(M) = \mathrm{E}\left[\sup_{s\in[0,T]}(Z_s - M_s)\right]$, $V(M) = \mathrm{Var}\left[\sup_{s\in[0,T]}(Z_s - M_s)\right]$ and

$$\mathcal{R}(\psi, \psi^*) = \rho(\psi, \psi^*)\sqrt{1 \vee |\log(\rho(\psi, \psi^*))|}$$

for any $\psi \in \Psi$.

Discussion Theorem 2.2 shows that the martingale M_n delivered by our algorithm has a nice property that the corresponding approximation error $Y(M_n) - Y^*$ and the square root variance $\sqrt{V(M_n)}$ can be bounded from above with high probability by the quantities proportional to the smallest distance between the classes of martingales $\mathfrak{M}$ and $\mathcal{M}^*$ as measured by ρ. Hence, if the set $\mathfrak{M}$ contains at least one martingale solving (9) we get, as expected, $Y(M_n) = Y^*$ with probability 1. In general, the larger is the class $\mathfrak{M}$, the smaller is the above distance. However, if the class $\mathfrak{M}$ is infinite-dimensional, maximizing the empirical objective functional in (10) over $\mathfrak{M}$ may not be well-defined or even if M_n exists, it is generally difficult to compute. Instead, one can restrict the maximization to a sequence of finite-dimensional approximating spaces $\mathfrak{M}_n = \{M(\psi) : \psi \in \Psi_n\}$ such that $\cup_n \Psi_n$ is dense in Ψ^*. Such a sequence of approximating spaces is usually called a sieve. We are interested in sieves that are compact, non-decreasing ($\mathfrak{M}_n \subset \mathfrak{M}_{n+1} \subset \ldots \subset \mathfrak{M}$) and such that for any $n \in \mathbb{N}$ and some $\psi^* \in \Psi^*$ there exists an element $\pi_n\psi^*$ in Ψ_n satisfying $\rho(\psi^*, \pi_n\psi^*) \to 0$ as $n \to \infty$, where π_n can be regarded as a projection of ψ^* to Ψ_n. For such sieves Theorem 2.2 implies that

$$V(M_n) \xrightarrow{\mathrm{P}} 0, \quad n \to \infty, \tag{13}$$

provided $\mathfrak{C}(\widetilde{\Psi}_n)/\sqrt{n}$ remains bounded as $n \to \infty$. In the next section we discuss how to get the martingale sieves $\mathfrak{M}_n$ in a constructive way. The asymptotic relation (13) implies that the variance of the Monte Carlo estimate

$$Y_m(M_n) = \frac{1}{m} \sum_{j=n+1}^{n+m} Z^{(j)}(M_n),$$

based on a new, independent of $(Z_t^{(1)}, M_t^{(1)}), \ldots, (Z_t^{(n)}, M_t^{(n)})$, set of paths

$$(Z_t^{(n+1)}, M_t^{(n+1)}), \ldots, (Z_t^{(n+m)}, M_t^{(n+m)}), \quad t \in [0,T],$$

has the standard deviation of order $o(1/\sqrt{m})$ as $m, n \to \infty$. Therefore one can speak about fast convergence rates in this situation.

Remark 2.3. If the class Ψ is of Vapnik-Cervonenkis type, i.e.,

$$N(\varepsilon, \widetilde{\Psi}, \rho) \lesssim \varepsilon^{-\beta}, \quad \varepsilon \to 0$$

for some $\beta > 0$, then $\mathfrak{C}$ and $J(\delta)$ are finite for any $\delta > 0$.

Remark 2.4. A natural question is whether the bounds of Theorem 2.2 can be achieved without using the penalization by empirical variance. The answer is, in general, no. To see this, let Z_t be an uniformly integrable submartingale. Then Z_t admits the so-called Doob-Meyer decomposition:

$$Z_t = Z_0 + M_t + A_t,$$

where M_t with $M_0 = 0$ is a uniformly integrable martingale and A_t is an increasing predictable process. Using the optional sampling theorem, we derive

$$Y^* = \sup_{\tau \in \mathcal{T}[0,T]} \mathrm{E}[Z_\tau] = \mathrm{E}[Z_T] = Z_0 + \mathrm{E}[A_T].$$

Define $M_t^* = M_t + \mathrm{E}[A_T|\mathcal{F}_t] - \mathrm{E}[A_T]$, then $Y^* = \sup_{t\in[0,T]}(Z_t - M_t^*)$ with probability 1. Furthermore, the martingale $\widetilde{M} = M$ fulfills

$$Y^* = \mathrm{E}\left[\sup_{t\in[0,T]} (Z_t - \widetilde{M}_t)\right],$$

and if A_T is not deterministic, then $Y^* \neq \sup_{t\in[0,T]}(Z_t - \widetilde{M}_t) = Z_0 + A_T$ with positive probability. Hence, $\widetilde{M}$ solves, along with M^*, the original dual problem (5), but does not have the almost sure property (9). Consider now the empirical optimization problem

$$M_n = \arg \inf_{M\in\left\{M^*, \widetilde{M}\right\}} \left(\frac{1}{n}\sum_{j=1}^{n} Z^{(j)}(M)\right)$$

with $Z(M) = \sup_{t\in[0,T]}(Z_t - M_t)$. It obviously holds

$$\mathrm{P}(M_n = \widetilde{M}) = \mathrm{P}\left(\sum_{j=1}^{n} \xi_j < 0\right) > 0,$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. random variables distributed as $A_T - \mathrm{E}[A_T]$. Therefore

$$V(M_n) = V(\widetilde{M}) = \mathrm{Var}\left[\sup_{t\in[0,T]} (Z_t - \widetilde{M}_t)\right] = \mathrm{Var}[A_T] > 0$$

with positive probability for any natural number n and the bound (12) does not hold any longer.

3 Martingales via martingale representations

Suppose that $Z_t = G_t(X_t)$, where for any $t \in [0,T]$, $G_t : \mathbb{R}^d \to \mathbb{R}$ (dependence on t ???) is a Hölder function and X_t is a d-dimensional Markov process solving the following system of SDE's:

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = x. \tag{14}$$

The coefficient functions $\mu : [0,T] \times \mathbb{R}^d \to \mathbb{R}$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are supposed to be Lipschitz in space and 1/2-Hölder continuous in time, with m denoting the dimension of the Brownian motion $W = (W^1, \ldots, W^m)^\top$ under measure P. It is well known that under the assumption that a martingale M_t is square integrable and is adapted to the filtration generated by W_t, there is a square integrable (row vector valued) process $H_t = (H_t^1, \ldots, H_t^m)$ satisfying

$$M_t = \int_0^t H_s dW_s. \tag{15}$$

It is not hard to see that in our Markovian setting, it holds (????) $H_s = \psi(s, X_s)$ for some ψ satisfying $\int_0^T \mathrm{E}[|\psi(s, X_s)|^2]\, ds < \infty$. As a result,

$$M_t = M_t(\psi) = \int_0^t \psi(s, X_s)dW_s.$$

Thus, the set of adapted square-integrable martingales can be "parameterized" by the set $L_{2,P}([0,T]\times\mathbb{R}^d)$ of square-integrable functions ψ on $[0,T] \times \mathbb{R}^d$ w.r.t the product measure $\lambda \times \mathrm{P}$, where λ is the Lebesgue measure on $\mathbb{R}$. Note that for any ψ from $L_{2,P}([0,T] \times \mathbb{R}^d)$ it holds $\|\psi\|_{2,\mathrm{P}} := \int_0^T \mathrm{E}[|\psi(s, X_s)|^2]\, ds < \infty$. Let $\widetilde{\Psi}^*$ be a set of $\psi \in L_{2,P}([0,T] \times \mathbb{R}^d)$ such that $M_t(\psi)$ solves (9). Choose a family of finite-dimensional linear models of functions, called sieves, with good approximation properties. We consider linear sieves of the form

$$\widetilde{\Psi}_K = \{\beta_1\phi_1 + \ldots + \beta_K\phi_K : \beta_1, \ldots, \beta_K \in \mathbb{R}\}, \tag{16}$$

where $\phi_1, \ldots, \phi_K$ are some given functions from $L_{2,P}([0,T] \times \mathbb{R}^d) \cap C([0,T] \times \mathbb{R}^d)$. Next define a class of adapted square-integrable martingales via

$$\mathfrak{M}_K = \{M_t(\psi) : \psi \in \widetilde{\Psi}_K\}$$

and set

$$M_n = \arg \inf_{M \in \mathfrak{M}_{K_n}} \left(\frac{1}{n} \sum_{j=1}^{n} Z^{(j)}(M) + (\varkappa + \lambda_n)\sqrt{V_n(M)} \right), \tag{17}$$

where $K_n \to \infty$ as $n \to \infty$. As can be easily seen

$$\sqrt{\langle M - M' \rangle_T} \leq \sqrt{T} \sup_{(s,x) \in [0,T] \times \mathbb{R}^d} |\psi(s,x) - \psi'(s,x)| := \sqrt{T} \cdot \rho(\psi, \psi')$$

with $M_t = M_t(\psi)$ and $M_t' = M_t(\psi')$ for any $\psi, \psi' \in L_{2,P}([0,T] \times \mathbb{R}^d) \cap C([0,T] \times \mathbb{R}^d)$. Hence the quadratic ρ-modulus of the family $\mathfrak{M}_K$ is bounded by $\sqrt{T}$ with probability 1. For many linear sieves of the form (16) it holds (see, e.g., Devroye, Györfi and Lugosi (1996))

$$\log[1 + N(\varepsilon, \widetilde{\Psi}_K, \rho)] \lesssim K^{d+1} \log(1/\varepsilon), \quad \varepsilon \to 0$$

and in this situation we have with probability at least $1 - \delta$

$$\sqrt{V(M_n)} = O\left(a_n\right),$$

where $a_n = \inf_{\psi \in \widetilde{\Psi}_{K_n}, \psi^* \in \widetilde{\Psi}^*} \|\psi - \psi^*\|_\infty$, provided $K_n^{d+1}/\sqrt{n} = O(1)$ for $n \to \infty$.

4 Numerical example

Consider a class of processes Z_t defined as

$$Z_t = \int_0^t f(s, W_s)\, dW_s + \int_0^t g(s, W_s)\, ds, \quad t \geq 0,$$

where $(W_t)_{t \geq 0}$ is the standard Brownian motion and f, g are two functions satisfying

$$\int_0^T \mathrm{E}\,|f(s, W_s)|^2\, ds < \infty, \qquad \int_0^T \mathrm{E}\,|g(s, W_s)|^2\, ds < \infty.$$

If $g \geq 0$, then the process Z_t is uniformly integrable submartingale and the value of the optimal stopping problem (1) is, due to the optional sampling theorem, given by

$$Y^* = \mathrm{E}\left[\int_0^T g(s, W_s)\, ds\right].$$

As a result we obtain an interesting class of test examples, where the exact value of the optimal stopping problem (1) can be computed. As an example let us study the case where $T = 1$,

$f(s,x) = \sin^4(x)$ and $g(s,x) = x^2$ leading to $Y^* = 1/2$. Consider a set of functions on $[0,T] \times \mathbb{R}$:

$$(\phi_1(t,x), \ldots, \phi_7(t,x)) = \{1, x, tx, \sin(x), \cos(x), \sin(2x), \cos(2x)\}$$

and define a sieve $\widetilde{\Psi}$ via

$$\widetilde{\Psi} = \{\beta_1\phi_1 + \ldots + \beta_7\phi_7 : \beta_1, \ldots, \beta_7 \in \mathbb{R}\}.$$

Now simulate n paths of the Brownian motion W_t on $[0,T]$ and consider two optimization problems

$$\psi_n = \operatorname*{arginf}_{\psi \in \widetilde{\Psi}} \left\{ \frac{1}{n} \sum_{j=1}^{n} Z^{(j)}(\psi) \right\} \tag{18}$$

and

$$\psi_{n,\lambda} = \operatorname*{arginf}_{\psi \in \widetilde{\Psi}} \left\{ \frac{1}{n} \sum_{j=1}^{n} Z^{(j)}(\psi) + \frac{\lambda}{n(n-1)} \sum_{1 \le i < j \le n} (Z^{(i)}(\psi) - Z^{(j)}(\psi))^2 \right\} \tag{19}$$

with

$$Z^{(j)}(\psi) = \sup_{t \in [0,T]} \left[Z_t^{(j)} - \int_0^t \psi(s, W_s^{(j)})\, dW_s^{(j)} \right], \tag{20}$$

$$Z_t^{(j)} = \int_0^t f(s, W_s^{(j)})\, dW_s^{(j)} + \int_0^t g(s, W_s^{(j)})\, ds$$

and some $\lambda > 0$. Note that the objective functionals in (18) and (19) are not smooth. In order to avoid computational problems related to the non-smoothness of $Z(M(\psi))$, we smooth it and consider instead Z the functional

$$Z_p(\psi) = p^{-1} \log \left(\int_0^T \exp(p(Z_s - M_s(\psi)))\, ds \right), \tag{21}$$

where $M_t(\psi) = \int_0^t \psi(s, W_s)\, dW_s$. An alternative expression for $Z_p(\psi)$ is

$$Z_p(\psi) = Z(M(\psi)) + p^{-1} \log \left(\int_0^T \exp(p(Z_s - M_s(\psi) - Z(M(\psi))))\, ds \right). \tag{22}$$

It follows from the representation (22) that

$$0 \le Z_p(\psi) - Z(M(\psi)) \le p^{-1} \log T.$$

Hence $Z_p(\psi) \to Z(M(\psi))$ as $p \to \infty$. The advantage of using $Z_p(\psi)$ instead of $Z(M(\psi))$ is that the standard gradient based optimization routines can be used to compute ψ_n and $\psi_{n,\lambda}$. In our numerical implementation we compute 1000 realizations of ψ_n and $\psi_{n,\lambda}$ by simulating 1000 independent sets of $n = 300$ discretised paths of the Brownian motion W_t on $[0,T]$ with the discretization step 0.0015. The integral (21) with $p = 25$ is then discretised using the same discretization step. For each realization of ψ_n and $\psi_{n,\lambda}$ we compute the corresponding mean values and standard deviations of the r.v. $Z(\psi_n)$ and $Z(\psi_{n,\lambda})$. The histograms of the obtained standard deviations along with estimated densities for the choice $\lambda = 2$ are reported in Figure 1. By averaging the mean values of $Z(\psi_n)$ and $Z(\psi_{n,2})$ over all 1000 replications, we get $\mathrm{E}[Z(\psi_n)] \approx$

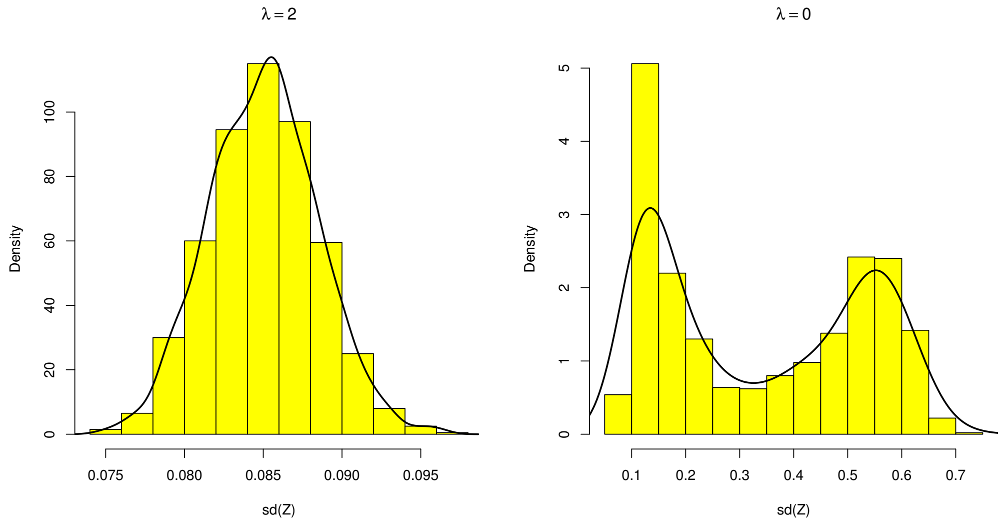

Figure 1: The histograms of the standard deviations of the r.v. $Z(\psi_{n,2})$ (left) and $Z(\psi_n)$ (right) based on 1000 realizations of the solutions $\psi_{n,2}$ and ψ_n.

0.4864316 and $\mathrm{E}[Z(\psi_{n,2})] \approx 0.5044231$ (recall that $Y^* = 0.5$), respectively. As one can see, the penalization via empirical variance leads not only to a significant reduction in varaince, but also to tighter probabilistic bounds for the variance of r.v. $Z(\psi_{n,\lambda})$.

5 Conclusion

This paper proposes a novel approach to variance reduction in simulation-based algorithms for optimal stopping problems. The approach consists in penalizing the empirical dual objective functional by its empirical variance. Our theoretical analysis shows that the variance of a value function estimate can be significantly reduced by a proper choice of the approximating sets of martingales. The approach of the paper is not restricted to the case of optimal stopping problems, it can also be used in optimal control problems and in general Markov decision problems. From the methodological point of view, the probabilistic tools developed in the paper can be used to analyze convergence of various types of empirical optimization problems arising in computational stochastics and finance.

6 Proofs of main results

6.1 Proof of Theorem 2.2

Let us first sketch the main steps of the proof. Our main interest lies in estimating the quantities $Y(M_n) - Y^*$ and $V(M_n)$. In order to obtain these estimates we need a kind of uniform (over $M \in \mathfrak{M}$) concentration inequality for the empirical process

$$\mathcal{E}_n(M) = \frac{1}{n}\sum_{j=1}^{n}(Z^{(j)}(M) - \mathrm{E}[Z(M)]) = \frac{1}{n}\sum_{j=1}^{n} Z^{(j)}(M) - Y(M)$$

that gives probabilistic bounds for $\sqrt{n}\cdot\mathcal{E}_n(M)$ in terms of the empirical variance $V_n(M)$. Indeed, such an inequality would allow us to get an upper bound for the quantity $Y(M_n) + \varkappa\sqrt{V_n(M_n)}$ with $\varkappa > 0$ in terms of $\mathcal{Q}_n(M_n)$, where

$$\mathcal{Q}_n(M) = \frac{1}{n}\sum_{j=1}^{n} Z^{(j)}(M) + (\varkappa + \lambda_n(\delta/2))\sqrt{V_n(M)}.$$

Unfortunately, the usual concentration inequalities could not be used here, as they would provide us with the bounds in terms of the true variance $V(M)$ and not in terms of the empirical one $V_n(M)$. However, there is another less known type of concentration inequalities for self-normalized empirical processes (see Bercu, Gassiat and Rio (2002)) and this is exactly what we need. We extend the above inequalities to the case of general family of random variables. As a next step, in order to derive a bound for $V(M_n)$, we need a kind of uniform concentration inequality for the empirical process $\Delta_n(\psi) = (V(M(\psi)) - V_n(M(\psi)))$ that holds uniformly over the set $\widetilde{\Psi}$ and gives probabilistic bounds for $\sqrt{n}\cdot\Delta_n(\psi)$ in terms of $\rho(\psi,\psi^*)$ for any fixed $\psi^* \in \Psi^*$. The latter type of inequalities can not be derived from the well known concentration inequalities for selfbounding random variables (see, e.g., Devroye and Lugosi (2008)), since variance $V(M)$ is a highly non-linear function of M and the random variable $Z(M)$ is usually not bounded. The corresponding concentration inequality making use of the local subgaussianity of $V(M)$, is presented in Section 6 and can be interesting in its own right. Finally, using the inequality $\mathcal{Q}_n(M_n) \le \mathcal{Q}_n(M)$, that holds for any $M \in \mathfrak{M}$, we will arrive at (11) and (12).

Part 1: The following proposition allows us to derive uniform bounds for the empirical process $\sqrt{n}\cdot\mathcal{E}_n(M)$ in terms of the empirical variance $V_n(M)$.

Proposition 6.1. *Let $\mathfrak{X}$ be a family of centered and normalized random variables on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with finite bracketing number in $L_2(\mathrm{P})$ such that*

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{X\in\mathfrak{X}} \max\left|\sqrt{n}\cdot\mathbb{E}_n[X]\right|\right] \le \mathfrak{C} < \infty$$

for some positive constant $\mathfrak{C} = \mathfrak{C}(\mathfrak{X})$, where

$$\mathbb{E}_n[X] = \frac{1}{n}\sum_{j=1}^{n} X^{(j)}$$

and $X^{(1)}, \ldots, X^{(n)}$ *are i.i.d. copies of the element* $X \in \mathfrak{X}$*. Define*

$$W_n(X) = \frac{\mathbb{E}_n[X]}{\sqrt{V_n(X)}}$$

with

$$V_n(X) = \frac{1}{n}\sum_{j=1}^{n}(X^{(j)})^2.$$

Then for any $\kappa > 0$ *and* $\alpha > \sqrt{2}$*, one can find some positive* θ *and* n_0 *depending on* $\mathfrak{X}$*,* α *and* κ *such that, for* $n \geq n_0$ *and for any* $x \in [0, \theta\sqrt{n}]$

$$\mathrm{P}\left(\sup_{X\in\mathfrak{X}} |\sqrt{n} \cdot W_n(X)| \geq (x + \alpha\mathfrak{C})\right) \leq 2\exp\left(-\frac{x^2}{4\alpha^2(1+\kappa)}\right).$$

For the case of non-centered and non-normalized random variables X, one can derive from Proposition 6.1 the following corollary.

Corollary 6.2. *Let* $\mathfrak{X}$ *be a family class of random variables on a common probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ *with finite bracketing number in* $L_2(\mathrm{P})$ *such that*

$$\sup_{X\in\mathfrak{X}} \mathrm{E}\,|X|^2 < \infty$$

and

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{X\in\mathfrak{X}} \left|\sqrt{n} \cdot \mathbb{E}_n[X - \mathrm{E}\,X]\right|\right] \leq \mathfrak{C} < \infty$$

for some positive constant $\mathfrak{C} = \mathfrak{C}(\mathfrak{X})$*. Define*

$$W_n(X) = \frac{\mathbb{E}_n[X] - \mathrm{E}[X]}{\sqrt{V_n(X)}}$$

with

$$V_n(X) = \frac{1}{n}\sum_{j=1}^{n}(X^{(j)} - \mathbb{E}_n[X])^2.$$

Then for any $\kappa > 0$ *and* $\alpha > \sqrt{2}$*, one can find some positive* θ *and* n_0 *depending on* $\mathfrak{X}$*,* α *and* κ *such that, for* $n \geq n_0$ *and for any* $x \in [0, \theta\sqrt{n}]$

$$\mathrm{P}\left(\sup_{X\in\mathfrak{X}} |\sqrt{n} \cdot W_n(X)| \geq \frac{\sqrt{2}(x + \alpha\mathfrak{C})}{1 - \sqrt{2}(x + \alpha\mathfrak{C})/n}\right) \leq 2\exp\left(-\frac{x^2}{4\alpha^2(1+\kappa)}\right),$$

provided $\sqrt{2}(x + \alpha\mathfrak{C}) < n$*. As a result, by fixing some* $\delta > 0$ *with* $\log(1/\delta) \leq \sqrt{n}$ *and taking* $x = 2\alpha\sqrt{(1+\kappa)\log(4/\delta)}$ *we get with probability at least* $1 - \delta$

$$\sup_{X\in\mathfrak{X}} |\sqrt{n} \cdot W_n(X)| \geq 2\sqrt{2}\alpha \cdot (2\sqrt{(1+\kappa)\log(2/\delta)} + \mathfrak{C})$$

for all $n > n_0$.

Part 2: Next we need the concentration inequality for the empirical process $\sqrt{n} \cdot (V_n(M) - V(M))$. The following proposition is proved in Section 7.1.

Proposition 6.3. *Let $\mathfrak{M} = \{M(\psi) : \psi \in \widetilde{\Psi}\}$ be a family of continuous local martingales, where $\widetilde{\Psi}$ is a subspace of the metric space (Ψ, ρ). Suppose that $\|\mathfrak{M}\|_{\rho,T} \leq \Theta$ a.s. for some finite Θ and*

$$J = \int_0^1 \sqrt{\log[1 + N(\varepsilon, \widetilde{\Psi}, \rho)]}\, d\varepsilon < \infty.$$

Denote $\Delta_n(\psi) = V_n(M(\psi)) - V(M(\psi))$ for any $\psi \in \Psi$, then for any fixed $\psi^* \in \mathcal{M}^*$ such that $\sup_{\psi \in \widetilde{\Psi}} \rho(\psi, \psi^*) < \infty$ it holds

$$\mathrm{P}\left(\sup_{\psi \in \widetilde{\Psi}} \left| \frac{\sqrt{n} \cdot \Delta_n(\psi)}{\mathcal{R}^2(\psi, \psi^*)} \right| > U \right) \leq \exp\left(-\frac{D \cdot U}{J} \right)$$

for any $U > 0$ and some constant $D > 0$ depending on Θ, where

$$\mathcal{R}(\psi, \psi') = \rho(\psi, \psi') \sqrt{1 \vee |\log(\rho(\psi, \psi'))|}$$

for any $\psi, \psi' \in \Psi$.

Part 3: Now we can begin with the proof of Theorem 2.2. By Corollary 6.2 it holds for any $\psi \in \widetilde{\Psi}$ with probability at least $1 - \delta/2$

$$\begin{aligned} Y(M_n) + \varkappa \sqrt{V_n(M_n)} &\leq \frac{1}{n} \sum_{j=1}^n Z^{(j)}(M_n) + (\varkappa + \lambda_n(\delta/4)) \sqrt{V_n(M_n)} \\ &\leq \frac{1}{n} \sum_{j=1}^n Z^{(j)}(M(\psi)) + (\varkappa + \lambda_n(\delta/4)) \sqrt{V_n(M(\psi))} \\ &\leq Y(M(\psi)) + (\varkappa + 2\lambda_n(\delta/4)) \sqrt{V_n(M(\psi))}. \end{aligned}$$

Proposition 6.3 implies that with probability at least $1 - \delta/4$

$$\begin{aligned} V_n(M(\psi)) &\leq V(M(\psi)) + JD^{-1} \log(4/\delta) \frac{\mathcal{R}^2(\psi, \psi^*)}{\sqrt{n}} \\ &\leq V(M(\psi)) + C\, \mathcal{R}^2(\psi, \psi^*) \end{aligned}$$

for some universal constant C, provided $J \log(1/\delta) \leq \sqrt{n}$. Hence, using the elementary inequality $\sqrt{a + b} \leq \sqrt{a} + \sqrt{b}$, we get

$$Y(M_n) + \varkappa \sqrt{V_n(M_n)} \leq Y(M(\psi)) + (\varkappa + 2\lambda_n(\delta/4)) \left[\sqrt{V(M(\psi))} + \sqrt{C}\, \mathcal{R}(\psi, \psi^*) \right]$$

with probability at least $1 - 3\delta/4$. By Burkholder-Davis-Gundy's inequality

$$Y(M(\psi)) - Y^* \leq \Theta\, \rho(\psi, \psi^*)$$

and $V(M(\psi)) \leq \Theta^2 \rho^2(\psi, \psi^*)$ for any $\psi \in \widetilde{\Psi}$. Therefore

$$Y(M_n) - Y^* \leq 2\sqrt{C}\,\Theta\,(1 + \varkappa + 2\lambda_n(\delta/4))\mathfrak{R}(\psi, \psi^*)$$

and

$$\sqrt{V_n(M_n)} \leq 2\sqrt{C}\,\Theta\,\varkappa^{-1}(1 + \varkappa + 2\lambda_n(\delta/4))\mathfrak{R}(\psi, \psi^*).$$

Using again Proposition 6.3, we get with probability at least $1 - \delta$

$$\sqrt{V(M_n)} \leq \sqrt{V_n(M_n)} + \sqrt{C}\mathfrak{R}(\psi, \psi^*) \leq 3\sqrt{C}\,\Theta\,\varkappa^{-1}(1 + \varkappa + 2\lambda_n(\delta/4))\mathfrak{R}(\psi, \psi^*).$$

Part 4: To finish the proof of Theorem 2.2, it suffices to prove the following proposition.

Proposition 6.4. *Let $\widetilde{\Psi}$ be a subspace of the metric space (Ψ, ρ) such that $\rho(\psi, \psi^*) \leq \sigma$ for some $\psi^* \in \Psi^*$, all $\psi \in \widetilde{\Psi}$ and some $\sigma > 0$. Define $\mathfrak{M} = \{M(\psi) : \psi \in \widetilde{\Psi}\}$ and set*

$$\mathfrak{C} = \int_0^\sigma \varepsilon^{-1} J(\varepsilon) \sqrt{\log[1 + N(\varepsilon, \widetilde{\Psi}, \rho)]}\, d\varepsilon.$$

If $\|\mathfrak{M}\|_{\rho,T} \leq \Theta$ a.s. and $\mathfrak{C} < \infty$, then there is a constant A depending on Θ, such that

$$\limsup_{n\to\infty} \mathrm{E}\left[\sup_{M \in \mathfrak{M}} |\mathbb{G}_n[Z(M)]|\right] \leq A\mathfrak{C}$$

with

$$\mathbb{G}_n[Z(M)] = \frac{1}{\sqrt{n}} \sum_{j=1}^n (Z^{(j)}(M) - \mathrm{E}[Z(M)]).$$

Proof. We follow the proof of Lemma 19.34 in van der Vaart (1998) with some straightforward modifications. It holds

$$\mathbb{G}_n(M(\psi)) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \Big(Z^{(j)}(M(\psi)) - \mathrm{E}[Z^{(j)}(M(\psi))]\Big) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \Big(Z^{(j)}(M(\psi)) - Z^{(j)}(M(\psi^*))\Big)$$
$$+ \frac{1}{\sqrt{n}} \sum_{j=1}^n \Big(\mathrm{E}[Z^{(j)}(M(\psi^*))] - \mathrm{E}[Z^{(j)}(M(\psi))]\Big),$$

since $\mathrm{Var}[Z(M(\psi^*))] = 0$. Setting

$$K_T = \sup_{\psi \in \widetilde{\Psi}} \sup_{t \in [0,T]} |M_t(\psi) - M_t(\psi^*)|,$$

we derive

$$|Z(M(\psi)) - Z(M(\psi^*))| \leq K_T, \quad |\mathrm{E}[Z(M(\psi))] - \mathrm{E}[Z(M(\psi^*))]| \leq \mathrm{E}[K_T].$$

As a result,

$$\mathrm{E}\sup_{\psi\in\widetilde{\Psi}}\left|\mathbb{G}_n(M(\psi))1\{K_T > a(\sigma)\sqrt{n}\}\right| \leq \sqrt{n}\cdot\mathrm{E}\left[(K_T+\mathrm{E}[K_T])1\{K_T > a(\sigma)\sqrt{n}\}\right]$$
$$\leq 2\sqrt{n}\cdot\mathrm{E}\left[K_T\cdot 1\{K_T > a(\sigma)\sqrt{n}\}\right],$$

where we set

$$a(\sigma) = \frac{J(\sigma)}{\sqrt{\log[1+N(\sigma,\widetilde{\Psi},\rho)]}}.$$

Under the condition $\|\mathfrak{M}\|_{\rho,T} \leq \Theta$ a.s. one can prove (see, e.g, van Zanten (2005)), that

$$\mathrm{P}\left(\sup_{\substack{\psi,\phi\in\widetilde{\Psi},\\ \rho(\psi,\phi)\leq\delta}}\sup_{t\in[0,T]}|M_t(\psi)-M_t(\phi)| > x\right) \leq 2e^{-x^2/CJ^2(\delta)} \tag{23}$$

for all $x > 0$, where C is a universal constant depending only on Θ. The inequality (23) implies

$$\mathrm{E}\left[K_T\cdot 1\{K_T > a(\sigma)\sqrt{n}\}\right] \leq 2a(\sigma)\sqrt{n}e^{-na^2(\sigma)/CJ^2(\sigma)} + 2\int_{a(\sigma)\sqrt{n}}^{\infty} e^{-x^2/CJ^2(\sigma)}\,dx.$$

Fix an integer q_0 such that $\sigma \leq 2^{-q_0} \leq 2\sigma$. For each natural number $q > q_0$, there exists a nested sequence of partitions $\widetilde{\Psi} = \cup_{i=1}^{N_q}\widetilde{\Psi}_{qi}$ of $\widetilde{\Psi}$ into N_q disjoint subsets such that $\rho(\psi,\phi) \leq 2^{-q}$ for any $\psi,\phi\in\widetilde{\Psi}_{qi}$ and $N_q \leq N(2^{-q+1},\widetilde{\Psi},\rho)$. Denote

$$\Delta_{qi} = \sup_{\psi,\phi\in\widetilde{\Psi}_{qi}}\sup_{t\in[0,T]}|M_t(\psi)-M_t(\phi)|,$$

then (23) implies

$$\mathrm{E}[\Delta_{qi}^2] \leq 2\int_0^\infty xe^{-x^2/CJ^2(2^{-q})}\,dx = CJ^2(2^{-q}).$$

Choose for each $q \geq q_0$ a fixed element ψ_{qi} from each partioning set $\widetilde{\Psi}_{qi}$, and set

$$\Pi_q\left[Z(M(\psi))\right] = Z(\psi_{qi}), \quad \Delta_q\left[Z(M(\psi))\right] = \Delta_{qi}, \quad \text{if } \psi\in\widetilde{\Psi}_{qi}.$$

Then $\Pi_q\left[Z(M(\psi))\right]$ and $\Delta_q\left[Z(M(\psi))\right]$ run through a set of N_q functions if ψ runs through $\widetilde{\Psi}$. Define for each fixed n and $q \geq q_0$ numbers and indicator functions

$$\begin{aligned}
a_q &= J(2^{-q})/\sqrt{\log[1+N_{q+1}]},\\
A_{q-1}[Z(M(\psi))] &= \mathbf{1}\{\Delta_{q_0}[Z(M(\psi))] \leq \sqrt{n}a_{q_0},\ldots,\Delta_{q-1}[Z(M(\psi))] \leq \sqrt{n}a_{q-1}\}\\
B_q[Z(M(\psi))] &= \mathbf{1}\{\Delta_{q_0}[Z(M(\psi))] \leq \sqrt{n}a_{q_0},\ldots,\Delta_{q-1}[Z(M(\psi))] \leq \sqrt{n}a_{q-1},\Delta_q[Z(M(\psi))] > \sqrt{n}a_q\}.
\end{aligned}$$

Now decompose

$$
\begin{aligned}
Z(M(\psi)) - \Pi_{q_0}[Z(M(\psi))] \;&=\; \sum_{q=q_0+1}^{\infty} (Z(M(\psi)) - \Pi_q[Z(M(\psi))])B_q[Z(M(\psi))] \\
&\quad + \sum_{q=q_0+1}^{\infty} (\Pi_q[Z(M(\psi))] - \Pi_{q-1}[Z(M(\psi))])A_{q-1}[Z(M(\psi))].
\end{aligned}
$$

We observe that either all of the $B_q[Z(M(\psi))]$ are zero in which case the $A_{q-1}[Z(M(\psi))]$ are 1. Alternatively, $B_{q_1}[Z(M(\psi))] = 1$ for some $q_1 > q_0$ (and zero for all other q), in which case $A_q[Z(M(\psi))] = 1$ for $q < q_1$ and $A_q[Z(M(\psi))] = 0$ for $q \geq q_1$. Our construction of partitions and choice of q_0 also ensure that

$$
a(\sigma) = \frac{J(\sigma)}{\sqrt{\log[1 + N(\sigma, \widetilde{\Psi}, \rho)]}} \leq \frac{J(2^{-q_0})}{\sqrt{\log[1 + N(2^{-q_0-1}, \widetilde{\Psi}, \rho)]}} \leq a_{q_0},
$$

whence $A_{q_0}[Z(M(\psi))] = 1$. Next we apply the empirical process $\mathbb{G}_n$ to both series on the r.h.s. of separately, take absolute values, and next take suprema over $\psi \in \widetilde{\Psi}$. Because the partitions are nested, $\Delta_q[Z(M(\psi))]B_q[Z(M(\psi))] \leq \Delta_{q-1}[Z(M(\psi))]B_q[Z(M(\psi))] \leq \sqrt{n}a_{q-1}$. The last inequality holds if $B_q[Z(M(\psi))] = 0$ and also if $B_q[Z(M(\psi))] = 1$ by definition. Furthermore, as $B_q[Z(M(\psi))]$ is indicator of the event $\Delta_q[Z(M(\psi))] > \sqrt{n}a_q$, it follows

$$
\sqrt{n}a_q \cdot \mathrm{E}\,[\Delta_q[Z(M(\psi))]B_q[Z(M(\psi))]] \leq \mathrm{E}[(\Delta_q[Z(M(\psi))])^2 B_q[Z(M(\psi))]] \leq J^2(2^{-q})
$$

by the choice of $\Delta_q[Z(M(\psi))]$. Because $|\mathbb{G}_n[Z(M(\psi))]| \leq \mathbb{G}_n[Z'] + 2\sqrt{n}\cdot\mathrm{E}[Z']$ if $|Z(M(\psi))| \leq Z'$, we obtain by triangle inequality and Lemma 8.1

$$
\begin{aligned}
\mathrm{E}\left\| \sum_{q=q_0+1}^{\infty} \mathbb{G}_n (Z(M(\psi)) - \Pi_q[Z(M(\psi))])B_q[Z(M(\psi))] \right\|_{\widetilde{\Psi}} \;&\leq\; \sum_{q=q_0+1} \mathrm{E}\,\|\mathbb{G}_n \Delta_q[Z(M(\psi))]B_q[Z(M(\psi))]\|_{\widetilde{\Psi}} \\
&\quad +2\sqrt{n} \sum_{q=q_0+1}^{\infty} \|\,\mathrm{E}\{\Delta_q[Z(M(\psi))] \\
&\quad \times B_q[Z(M(\psi))]\}\|_{\widetilde{\Psi}} \\
&\lesssim\; \sum_{q=q_0+1}^{\infty} \Big[a_{q-1} \log[1 + N_q] \\
&\quad + CJ(2^{-q})\sqrt{\log[1 + N_q]} + \frac{J^2(2^{-q})}{a_q}\Big].
\end{aligned}
$$

In view of the definition of a_q, the series on the right can be bounded by a multiple of the series $\sum_{q=q_0+1}^{\infty} J(2^{-q})\sqrt{\log[1 + N_q]}$. To establish a similar bound for the second part of equation (24), note that there are at most N_q differences $\Pi_q[Z(M(\psi))] - \Pi_{q-1}[Z(M(\psi))]$ and at most N_{q-1} indicator functions $A_{q-1}[Z(M(\psi))]$. Because the partitions are nested, $(\Pi_q[Z(M(\psi))] - \Pi_{q-1}[Z(M(\psi))])A_{q-1}[Z(M(\psi))]$ is bounded by $\Delta_{q-1}[Z(M(\psi))]A_{q-1}[Z(M(\psi))] \leq \sqrt{n}a_{q-1}$. Moreover, $\mathrm{E}[\Pi_q[Z(M(\psi))] - \Pi_{q-1}[Z(M(\psi))]]^2 \le CJ^2(2^{-q})$. Hence

$$\left\| \sum_{q_0+1}^{\infty} \mathbb{G}_n(\Pi_q[Z(M(\psi))] - \Pi_{q-1}[Z(M(\psi))])A_{q-1}[Z(M(\psi))] \right\|_{\widetilde{\Psi}} \le \sum_{q=q_0+1}^{\infty} \Big[a_{q-1}\log(1+N_q) + CJ(2^{-q})\sqrt{\log[1+N_q]}\Big].$$

Again this is bounded above by a multiple of the series $\sum_{q=q_0+1}^{\infty} J(2^{-q})\sqrt{\log[1+N_q]}$. To conclude the proof it suffices to consider the terms $\Pi_{q_0}[Z(M(\psi))]$. Because $|\Pi_{q_0}[Z(M(\psi))]| \le K_T \le a(\delta)\sqrt{n} \le a_{q_0}\sqrt{n}$ and

$$\mathrm{E}(\Pi_{q_0}[Z(M(\psi))])^2 \le \mathrm{E}\left[\sup_{t\in[0,T]} (M_t(\psi_{q_0 i}) - M_t(\psi^*))\right]^2 \le 4\Theta^2\sigma^2$$

by Burkholder-Davis-Gundy's inequality, we have

$$\mathrm{E}\,\|\mathbb{G}_n\Pi_{q_0}[Z(M(\psi))]\|_{\widetilde{\Psi}} \lesssim a_{q_0}\log[1+N_{q_0}] + \sigma\sqrt{\log[1+N_{q_0}]}.$$

By the choice of q_0, this is bounded by a multiple of the first few items of the series

$$\sum_{q=q_0+1}^{\infty} J(2^{-q})\sqrt{\log[1+N_q]}.$$

□

7 Proofs

7.1 Proof of Proposition 6.1

The proof can be routinely carried out along with lines of Bercu, Gassiat and Rio (2002).

7.2 Proof of Proposition 6.3

In order to prove Proposition 6.3 we need the following lemma.

Lemma 7.1. *Denote*

$$\mathcal{Q}(\psi,\psi') = \rho(\psi,\psi')\sqrt{\log\log(\rho^2(\psi,\psi') \vee e^2)}$$

for any $\psi,\psi' \in \Psi$. There is $\varepsilon > 0$ such that for any $\psi,\psi' \in \Psi$ and $\psi^* \in \Psi^*$ it holds

$$\mathrm{E}\left\{\exp\left(\theta\left[\frac{\sqrt{n}\cdot(\Delta_n(\psi) - \Delta_n(\psi'))}{\mathcal{Q}(\psi,\psi^*)\cdot\mathcal{Q}(\psi,\psi')}\right]\right) - 1\right\} \le C\theta^2$$

for some constant $C > 0$, provided $|\theta| \le \varepsilon$.

Proof. Without loss of generality, we may, and do, assume that $\Theta = 1$. Fix a martingale $M^* = M(\psi^*) \in \mathcal{M}^*$. Since $Z(M^*) = \mathrm{E}[Z(M^*)]$ almost surely, we have for arbitrary $M = M(\psi), M' =$

$M(\psi') \in \mathcal{M}$

$$
\begin{aligned}
V_n(M) - V_n(M') &= \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \left(\widetilde{Z}^{(i)}(M) - \widetilde{Z}^{(j)}(M)\right)^2 \\
&\quad - \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \left(\widetilde{Z}^{(i)}(M') - \widetilde{Z}^{(j)}(M')\right)^2 \\
&= \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \left(\widetilde{Z}^{(i)}(M) - \widetilde{Z}^{(i)}(M') - \widetilde{Z}^{(j)}(M) + \widetilde{Z}^{(j)}(M')\right) \\
&\quad \times \left(\widetilde{Z}^{(i)}(M) - \widetilde{Z}^{(j)}(M) + \widetilde{Z}^{(i)}(M') - \widetilde{Z}^{(j)}(M')\right) \\
&= \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \left(Z^{(i)}(M) - Z^{(i)}(M') - Z^{(j)}(M) + Z^{(j)}(M')\right) \\
&\quad \times \left(\widetilde{Z}^{(i)}(M) - \widetilde{Z}^{(j)}(M) + \widetilde{Z}^{(i)}(M') - \widetilde{Z}^{(j)}(M')\right)
\end{aligned}
$$

with $\widetilde{Z}^{(i)} = Z^{(i)}(M) - Z^{(i)}(M^*)$, $i = 1, \ldots, n$. Set

$$
\xi_i = Z^{(i)}(M) - Z^{(i)}(M'), \quad \zeta_i = \widetilde{Z}^{(i)}(M) + \widetilde{Z}^{(i)}(M'), \quad i = 1, \ldots, n,
$$

then

$$
\begin{aligned}
V_n(M) - V_n(M') &= \frac{1}{n(n-1)} \sum_{1\le i<j\le n} (\xi_i - \xi_j)(\zeta_i - \zeta_j) \\
&= \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \xi_i\zeta_i - \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \xi_i\zeta_j \\
&\quad - \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \xi_j\zeta_i + \frac{1}{n(n-1)} \sum_{1\le i<j\le n} \xi_j\zeta_j \\
&= \frac{2}{n} \sum_{i=1}^{n} \xi_i\zeta_i - \frac{1}{n(n-1)} \sum_{i\ne j} \xi_i\zeta_j.
\end{aligned}
$$

Hence

$$
\begin{aligned}
V_n(M) - V(M) - (V_n(M') - V(M')) &= \frac{2}{n} \sum_{i=1}^{n} (\xi_i\zeta_i - \mathrm{E}[\xi_i\zeta_i]) \\
&\quad - \frac{1}{n(n-1)} \sum_{i\ne j} (\xi_i\zeta_j - \mathrm{E}[\xi_i\zeta_j]).
\end{aligned} \tag{24}
$$

Note that $\xi_1\zeta_1), \ldots, (\xi_n, \zeta_n$ is a family of i.i.d. random two-dimensional vectors such that

$$
|\xi_i| \le \sup_{t\in[0,T]} |M_t^{(i)} - M_t'^{(i)}|, \quad i = 1, \ldots, n,
$$

and

$$|\zeta_i| \le 2 \sup_{t\in[0,T]} |M_t^{(i)} - M_t^{*(i)}|, \quad i = 1,\ldots,n.$$

Lemma 8.3 implies that for any $x > 0$

$$\mathrm{P}\left(\frac{|\xi_i|}{\sqrt{\langle M^{(i)} - M'^{(i)}\rangle_T \log\log(\langle M^{(i)} - M'^{(i)}\rangle_T \vee e^2)}} \ge x\right) \le C(\alpha)e^{-\alpha x^2}$$

and

$$\mathrm{P}\left(\frac{|\zeta_i|}{\sqrt{\langle M^{(i)} - M^{*(i)}\rangle_T \log\log(\langle M^{(i)} - M^{*(i)}\rangle_T \vee e^2)}} \ge x\right) \le C(\alpha)e^{-\alpha x^2/4}.$$

As a result,

$$\mathrm{P}\left(\frac{|\xi_i|}{\rho(\psi,\psi')\sqrt{\log\log(\rho^2(\psi,\psi')\vee e^2)}} \ge x\right) \le C(\alpha)e^{-\alpha x^2}$$

and

$$\mathrm{P}\left(\frac{|\zeta_i|}{\rho(\psi,\psi^*)\sqrt{\log\log(\rho^2(\psi,\psi^*)\vee e^2)}} \ge x\right) \le C(\alpha)e^{-\alpha x^2/4}$$

for $i = 1,\ldots,n$. Using the representation (24), we get

$$\begin{aligned}\frac{\sqrt{n}(\Delta_n(\psi) - \Delta_n(\psi'))}{\mathcal{R}(\psi,\psi^*)\cdot\mathcal{R}(\psi,\psi')} &= \frac{2}{\sqrt{n}}\sum_{i=1}^n(\widetilde{\xi}_i\widetilde{\zeta}_i - \mathrm{E}[\widetilde{\xi}_i\widetilde{\zeta}_i]) - \frac{1}{\sqrt{n}(n-1)}\sum_{i<j}(\widetilde{\xi}_i\widetilde{\zeta}_j - \mathrm{E}[\widetilde{\xi}_i\widetilde{\zeta}_j]) \\ &\quad - \frac{1}{\sqrt{n}(n-1)}\sum_{j<i}(\widetilde{\xi}_i\widetilde{\zeta}_j - \mathrm{E}[\widetilde{\xi}_i\widetilde{\zeta}_j]) = T_{1,n} + T_{2,n} + T_{3,n},\end{aligned}$$

where the "normalized" random variables

$$\begin{aligned}\widetilde{\xi}_i &= \frac{\xi_i}{\rho(\psi,\psi')\sqrt{\log\log(\rho^2(\psi,\psi')\vee e^2)}}, \\ \widetilde{\zeta}_i &= \frac{\zeta_i}{\rho(\psi,\psi^*)\sqrt{\log\log(\rho^2(\psi,\psi^*)\vee e^2)}}\end{aligned}$$

satisfy

$$\mathrm{P}\left(|\widetilde{\zeta}_i| \vee |\widetilde{\xi}_i| \ge x\right) \le C(\alpha)e^{-\alpha x^2/4}, \quad i = 1,\ldots,n. \tag{25}$$

The inequalities (25) immediately imply

$$\mathrm{P}(|\widetilde{\xi}_i\widetilde{\zeta}_i| > x) \le \mathrm{P}(|\widetilde{\xi}_i|^2 + |\widetilde{\zeta}_i|^2 > 2x) \quad \le \mathrm{P}(|\widetilde{\xi}_i| > \sqrt{2x}) + \mathrm{P}(|\widetilde{\zeta}_i| > \sqrt{2x}) \le 2C(\alpha)\exp(-\alpha x/2).$$

Consider first the term $T_{1,n}$. For any $\theta \in \mathbb{R}$ we have

$$\mathrm{E}[\exp(\theta T_{1,n})] = \prod_{i=1}^{n} \mathrm{E}[\exp(\theta(\widetilde{\xi}_i\widetilde{\zeta}_i - \mathrm{E}[\widetilde{\xi}_i\widetilde{\zeta}_i])/\sqrt{n})]. \tag{26}$$

Since the random variables $\widetilde{\xi}_i\widetilde{\zeta}_i - \mathrm{E}[\widetilde{\xi}_i\widetilde{\zeta}_i]$, $i = 1, \ldots, n$, possess finite moments of any order and have zero mean, it holds

$$\log \mathrm{E}\left[\exp(\varepsilon(\widetilde{\xi}_i\widetilde{\zeta}_i - \mathrm{E}[\widetilde{\xi}_i\widetilde{\zeta}_i])\right] = \frac{1}{2}\sigma^2\varepsilon^2 + o(\varepsilon^2), \quad i = 1, \ldots, n,$$

as $\varepsilon \to 0$, where $\sigma^2 = \mathrm{E}\left(\widetilde{\xi}_i\widetilde{\zeta}_i - \mathrm{E}[\widetilde{\xi}_i\widetilde{\zeta}_i]\right)^2$. Hence the inequality

$$\mathrm{E}\left[\exp(\varepsilon(\widetilde{\xi}_i\widetilde{\zeta}_i - \mathrm{E}[\widetilde{\xi}_i\widetilde{\zeta}_i])\right] \leq e^{C_1\varepsilon^2} \tag{27}$$

holds for sufficiently small ε and any $C_1 > \sigma^2/2$. Combining (26) with (27), we get for all $n \in \mathbb{N}$ and sufficiently small $\theta > 0$

$$\mathrm{E}[\exp(\theta T_{1,n}) - 1] \leq e^{C_1\theta^2} - 1 \leq C_2\theta^2.$$

Turn now to the terms $T_{2,n}$ and $T_{3,n}$. We need the following proposition to estimate $T_{2,n}$ and $T_{3,n}$.

Proposition 7.2. *Let $(X_1, Y_1) \ldots, (X_n, Y_n)$ be a sequence of i.i.d. centered random vectors in $\mathbb{R}^2$ such that $\mathrm{E}\,|X_i|^p < \infty$ and $\mathrm{E}\,|Y_i|^p < \infty$ for all $i = 1, \ldots, n$, and some $p \geq 2$. Then*

$$\begin{aligned}\mathrm{E}\left|\sum_{1\leq i<j\leq n} X_iY_j\right|^p \leq\ & C^p \max\Bigg\{\sum_{1\leq i<j\leq n} \mathrm{E}\,|X_i|^p\,\mathrm{E}\,|Y_j|^p, \sum_{i=1}^{n-1}\mathrm{E}\,|X_i|^p\left(\sum_{j=i+1}^{n}\mathrm{E}\,|Y_j|^2\right)^{p/2}, \\ & \sum_{j=2}^{n}\mathrm{E}\,|Y_j|^p\left(\sum_{i=1}^{j-1}\mathrm{E}\,|X_i|^2\right)^{p/2}, \left(\sum_{1\leq i<j\leq n}\mathrm{E}\,|X_i|^2\,\mathrm{E}\,|Y_j|^2\right)^{p/2}\Bigg\}\end{aligned} \tag{28}$$

for some constant $C > 0$ not depending on p.

Proof. Denote $Q_n = \sum_{1\leq i<j\leq n} X_iY_j$ and

$$V_j = \sum_{i=1}^{j-1} X_iY_j, \quad j = 2, \ldots, n.$$

It is clear that $T_{2,n} = \sum_{j=2}^{n} V_j$ and $(V_j,\, j = 2, \ldots, n)$ is a forward martingale-difference sequence (see Appendix for definition) with respect to σ-algebras $\mathcal{F}_j = \sigma((X_1, Y_1), \ldots, (X_j, Y_j))$, $j =$

$2,\dots,n$. By the martingale Rosenthal inequality (see Proposition 8.2 in Appendix)

$$\mathrm{E}[|Q_n|^p] \le B(p/\log p) \max\left\{ \sum_{j=2}^{n} \mathrm{E}\,|V_j|^p, \mathrm{E}\left(\sum_{j=2}^{n} \mathrm{E}\left[V_j^2 \middle| \mathcal{F}_{j-1}\right] \right)^{p/2} \right\}$$

and

$$\mathrm{E}\,|V_j|^p \le B(p/\log p) \cdot \mathrm{E}\,|Y_j|^p \max\left\{ \sum_{i=1}^{j-1} \mathrm{E}\,|X_i|^p, \left(\sum_{i=1}^{j-1} \mathrm{E}\,|X_i|^2 \right)^{p/2} \right\} \tag{29}$$

for all $j = 2,\dots,n$. Then

$$\begin{aligned}
&\mathrm{E}\left(\sum_{j=2}^{n} \mathrm{E}\left[V_j^2 \middle| \mathcal{F}_{j-1}\right] \right)^{p/2} = \\
&\qquad \mathrm{E}\left(\sum_{1\le i<j\le n} |X_i|^2\, \mathrm{E}\,|Y_j|^2 + 2\sum_{j=3}^{n} \sum_{1\le k<l\le j-1} X_k X_l\; \mathrm{E}\,|Y_j|^2 \right)^{p/2} \\
&\qquad \le 2^{p/2-1}\, \mathrm{E}\left(\sum_{1\le i<j\le n} |X_i|^2\, \mathrm{E}\,|Y_j|^2 \right)^{p/2} + 2^{p-1}\, \mathrm{E}\left| \sum_{1\le k<l\le n-1} X_k X_l \sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2 \right|^{p/2}.
\end{aligned}$$

By the Rosenthal inequality

$$\begin{aligned}
&\mathrm{E}\left(\sum_{1\le i<j\le n} |X_i|^2\, \mathrm{E}\,|Y_j|^2 \right)^{p/2} = \mathrm{E}\left(\sum_{i=1}^{n-1} |X_i|^2 \sum_{j=i+1}^{n} \mathrm{E}\,|Y_j|^2 \right)^{p/2} \le \\
&B(p/2)\log^{-1}(p/2) \max\left\{ \sum_{i=1}^{n-1} \mathrm{E}\,|X_i|^p \left[\sum_{j=i+1}^{n} \mathrm{E}\,|Y_j|^2 \right]^{p/2}, \left(\sum_{1\le i<j\le n} \mathrm{E}\,|X_i|^2\, \mathrm{E}\,|Y_j|^2 \right)^{p/2} \right\}.
\end{aligned} \tag{30}$$

Using the Jensen inequality, we get for $2 \le p < 4$

$$\mathrm{E}\left| \sum_{1\le k<l\le n-1} X_k X_l \sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2 \right|^{p/2} \le \left(\sum_{1\le k<l\le n-1} \mathrm{E}\,|X_k X_l|^2 \left(\sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2 \right)^2 \right)^{p/4}.$$

Moreover

$$\left(\sum_{1\le k<l\le n-1} \mathrm{E}\,|X_k X_l|^2 \left(\sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2 \right)^2 \right)^{p/4} \le \left(\sum_{1\le i<j\le n} \mathrm{E}\,|X_i|^2\, \mathrm{E}\,|Y_j|^2 \right)^{p/2}. \tag{31}$$

Combining (29), (30) and (31), we arrive at the inequality (28). Thus Lemma 7.2 is proved for all $2 \le p < 4$. Suppose now that the inequality (28) holds for $p \le m-1$ with some $m > 4$. Let us prove it for $p = m$. It follows from the previous steps, that we only need to obtain an upper bound for the term

$$\mathrm{E}\left|\sum_{1\le k<l\le n-1} X_k X_l \sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2\right|^{m/2}.$$

Our induction hypothesis gives

$$\begin{aligned}
\mathrm{E}\left|\sum_{1\le k<l\le n-1} X_k X_l \sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2\right|^{m/2} \le\ & C^{m/2} \max\left\{ \sum_{1\le k<l\le n-1} \mathrm{E}\left|X_k X_l \sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2\right|^{m/2}, \right.\\
& \sum_{k=1}^{n-2} \mathrm{E}\,|X_k|^{m/2} \left(\sum_{l=k+1}^{n-1} |X_l|^2 \sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2\right)^{m/4}, \\
& \sum_{l=2}^{n-1} |X_l|^{m/2} \left(\sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2\right)^{m/2} \left(\sum_{k=1}^{l-1} \mathrm{E}\,|X_k|^2\right)^{m/4}, \\
& \left.\left(\sum_{1\le k<l\le n-1} \mathrm{E}\left[|X_k|^2 |X_l|^2\right] \left(\sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2\right)^2\right)^{m/4} \right\}.
\end{aligned} \tag{32}$$

Let us consider, for example, the first term in the above maximum. Using the inequality

$$\left(\mathrm{E}\sum_{k=1}^{n} |U_k|^p\right)^2 \le \max\left\{ \sum_{k=1}^{n} \mathrm{E}\,|U_k|^{2p}, \left(\sum_{k=1}^{n} \mathrm{E}\,|U_k|\right)^{2p} \right\} \tag{33}$$

that holds for any $p > 1$ and any sequence of independent r.v. $U_1, \ldots, U_n$ with $\mathrm{E}\,|U_k|^{2p} < \infty$, we get

$$\begin{aligned}
\sum_{1\le k<l\le n-1} \mathrm{E}\left|X_k X_l \sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2\right|^{m/2} &= \sum_{1\le k<l\le n-1} \mathrm{E}\,|X_k X_l|^{m/2} \left(\sum_{j=l+1}^{n} \mathrm{E}\,|Y_j|^2\right)^{m/2} \\
&\le \left[\sum_{i=1}^{n-1} \mathrm{E}\,|X_i|^{m/2} \left(\sum_{j=i+1}^{n} \mathrm{E}\,|Y_j|^2\right)^{m/4}\right]^2
\end{aligned}$$

and

$$\left[\sum_{i=1}^{n-1} \mathrm{E}\,|X_i|^{m/2} \left(\sum_{j=i+1}^{n} \mathrm{E}\,|Y_j|^2\right)^{m/4}\right]^2 \leq \max\left\{\sum_{i=1}^{n-1} \mathrm{E}\,|X_i|^m \left(\sum_{j=i+1}^{n} \mathrm{E}\,|Y_j|^2\right)^{m/2}, \left(\sum_{1\le i<j\le n} \mathrm{E}\,|X_i|^2\,\mathrm{E}\,|Y_j|^2\right)^{m/2}\right\}.$$

To see that the inequality (33) holds, just note that the function

$$h(t) = \log\left[\sum_{k=1}^{n} \mathrm{E}\,|U_k|^t\right]$$

is convex in the domain $t > 1$. Due to convexity of $h(t)$, we have

$$\left(\sum_{k=1}^{n} \mathrm{E}\,|U_k|^p\right)^{2p-1} \leq \left(\sum_{k=1}^{n} \mathrm{E}\,|U_k|^{2p}\right)^{p-1} \left(\sum_{k=1}^{n} \mathrm{E}\,|U_k|\right)^{p}$$

for any $p > 1$. Hence

$$\begin{aligned}\left(\sum_{k=1}^{n} \mathrm{E}\,U_k^p\right)^2 &\leq \left(\sum_{k=1}^{n} \mathrm{E}\,|U_k|^{2p}\right)^{\frac{2(p-1)}{2p-1}} \left(\sum_{k=1}^{n} \mathrm{E}\,|U_k|\right)^{\frac{2p}{2p-1}} \\ &\leq \max\left\{\sum_{k=1}^{n} \mathrm{E}\,|U_k|^{2p}, \left(\sum_{k=1}^{n} \mathrm{E}\,|U_k|\right)^{2p}\right\}.\end{aligned}$$

Other terms on the right hand side of (32) can be handled in a similar way. □

Let us proceed with estimating the term $T_{2,n}$. Without loss of generality we may assume that $\mathrm{E}[\widetilde{\xi}] = \mathrm{E}[\widetilde{\zeta}] = 0$. Note that for any natural $p > 0$

$$\begin{aligned}\mathrm{E}[|\widetilde{\xi}|^p] &\leq 2p\,C(\alpha)\int_0^\infty x^{p-1}\exp(-\alpha x^2)dx = \frac{2p\,C(\alpha)}{(2\alpha)^{p/2}}\int_0^\infty y^{p-1}\exp(-y^2/2)dy \\ &\leq \frac{p\sqrt{2\pi}C(\alpha)}{(2\alpha)^{p/2}}\,\mathrm{E}[|Z|^p],\end{aligned}$$

where $Z \sim N(0,1)$. Similarly

$$\mathrm{E}[|\widetilde{\zeta}|^p] \leq \frac{2^{p/2}p\sqrt{2\pi}C(\alpha)}{\alpha^{p/2}}\,\mathrm{E}[|Z|^p].$$

As a result, we get from Proposition 7.2

$$\mathrm{E}[|T_{2,n}|^p] \leq C^p \max\left\{n^{1-p/2}(n-1)^{1-p}\,\mathrm{E}[|Z|^{2p}], n^{-p/2}(n-1)^{1-p/2}\,\mathrm{E}[|Z|^p], (n-1)^{-p/2}\right\}$$

for some constant $C > 0$ and any $p > 1$. Hence for any $\theta \in \mathbb{R}$

$$
\begin{aligned}
\text{(34)} \qquad \mathrm{E}[\exp(\theta T_{2,n}) - 1] &= \sum_{k=2}^{\infty} \frac{\theta^k}{k!} \mathrm{E}[T_{2,n}^k] \\
&\leq \sum_{k=2}^{\infty} \frac{|\theta|^k}{k!} \frac{B_1^k}{(n-1)^{k/2}} \mathrm{E}[Z^{2k}] \\
&= \mathrm{E}[\exp(B_1|\theta|Z^2/\sqrt{n-1})] - 1 - B_1|\theta| \,\mathrm{E}[Z^2]/\sqrt{n-1} \\
&= \frac{1}{\sqrt{1 - 2B_1|\theta|/\sqrt{n-1}}} - 1 - B_1|\theta|/\sqrt{n-1} \\
&\leq B_2\theta^2,
\end{aligned}
$$

provided $B_1|\theta|/\sqrt{n-1} < 1/2$, where B_1 and B_2 are two constants not depending on k and n. Analogously to (34), one can prove that

$$\mathrm{E}[\exp(\theta T_{3,n}) - 1] \leq B_3\theta^2$$

for sufficiently small $|\theta|$. Hence by the Cauchy-Schwarz inequality

$$\mathrm{E}\left[e^{\theta(T_{1,n}+T_{2,n}+T_{3,n})} - 1\right] \leq \left[\mathrm{E}\, e^{2\theta T_{1,n}}\right]^{1/2} \left[\mathrm{E}\, e^{4\theta T_{2,n}}\right]^{1/4} \left[\mathrm{E}\, e^{4\theta T_{3,n}}\right]^{1/4} - 1 \leq B_4\theta^2$$

for some constant $B_4 > 0$. Lemma 7.1 is proved. □

Let us proceed with the proof of Proposition 6.3. Let $\left\{\widetilde{\Psi}^m\right\}_{m\in\mathbb{N}}$ be a sequence of finite subsets of $\widetilde{\Psi}$ such that $\widetilde{\Psi}^m \uparrow \widetilde{\Psi}$ as $m \to \infty$. Introduce the disjoint sets

$$H_p = \{\psi \in \widetilde{\Psi} : 2^{-p-1} < \rho(\psi, \psi^*) \leq 2^{-p}\}$$

for any $p \in \mathbb{Z}$. Without loss of generality we may assume that H_p are empty for $p < 0$. For every $m \in \mathbb{N}$, denote by $q(m,p)$ the smallest integer such that $q(m,p) > p$ and that each of closed balls with centers in $\widetilde{\Psi}^m \cap H_p$ and ρ-radius $2 \cdot 2^{-q(m,p)}$ contains exactly one point in $\widetilde{\Psi}^m \cap H_p$. Then it is clear that $\mathrm{Card}(\widetilde{\Psi}^m \cap H_p) \leq N(2^{-q(m,p)}, \widetilde{\Psi} \cap H_p, \rho)$. Next let us introduce some mappings $\pi_r^{m,p} : \widetilde{\Psi}^m \cap H_p \to \widetilde{\Psi}_r^{m,p}$, $p \leq r \leq q(m,p)$, defined by

$$\pi_r^{m,p} = \lambda_r^{m,p} \circ \lambda_{r+1}^{m,p} \circ \ldots \circ \lambda_{q(m,p)}^{m,p},$$

where the sets $\widetilde{\Psi}_r^{m,p} \subset \widetilde{\Psi}^m \cap H_p$ and the mappings $\lambda_r^{m,p} : \widetilde{\Psi}^m \cap H_p \to \widetilde{\Psi}_r^{m,p}$ are specified in the following way. For $p \leq r < q(m,p)$, choose $\widetilde{\Psi}_r^{m,p}$ and define $\lambda_r^{m,p}$ such that they satisfy the following two conditions: $\mathrm{Card}(\widetilde{\Psi}_r^{m,p}) \leq N(2^{-r}, \widetilde{\Psi} \cap H_p, \rho)$ and $\rho(\psi, \lambda_r^{m,p}(\psi)) \leq 2 \cdot 2^{-r}$ for every $\psi \in \widetilde{\Psi}^m \cap H_p$. For $r = q(m,p)$, put $\widetilde{\Psi}_{q(m,p)}^{m,p} = \widetilde{\Psi}^m \cap H_p$ and denote by $\lambda_{q(m,p)}^{m,p}$ the identical mapping on $\widetilde{\Psi}^m \cap H_p$. In terms of the mappings $\pi_r^{m,p}$ which have been introduced, we consider the chaining given as follows: for every $n \in \mathbb{N}$ and $\psi \in \widetilde{\Psi} \cap H_p$

$$|\Delta_n(\psi)| \leq \sum_{r=p+1}^{q(m,p)} |\Delta_n(\pi_r^{m,p}(\psi)) - \Delta_n(\pi_{r-1}^{m,p}(\psi))| + |\Delta_n(\pi_p^{m,p}(\psi))|.$$

Since $\rho(\pi_r^{m,p}(\psi), \pi_{r-1}^{m,p}(\psi))/\rho(\psi, \psi^*) \leq 2^{-r+p+1}$ and $\rho(\pi_r^{m,p}(\psi), \psi^*)/\rho(\psi, \psi^*) \leq 2$ on $\widetilde{\Psi}^m \cap H_p$, it follows from Lemma 7.1 and Lemma 8.2 in Kosorok (2008) that

$$\begin{aligned}
&\mathrm{E}\left[\exp\left(\theta \sup_{\psi \in \widetilde{\Psi}^m \cap H_p} \left\{\frac{\sqrt{n}|\Delta_n(\pi_r^{m,p}(\psi)) - \Delta_n(\pi_{r-1}^{m,p}(\psi))|}{\mathcal{R}^2(\psi, \psi^*)}\right\}\right) - 1\right] \\
&\qquad \leq \mathrm{E}\left[\exp\left(\theta \sup_{\psi \in \widetilde{\Psi}^m \cap H_p} \left\{\frac{\mathcal{Q}(\pi_{r-1}^{m,p}(\psi), \pi_r^{m,p}(\psi))\mathcal{Q}(\pi_r^{m,p}(\psi), \psi^*)}{\mathcal{R}^2(\psi, \psi^*)}\right.\right.\right. \\
&\qquad\qquad \left.\left.\left.\times \frac{\sqrt{n}|\Delta_n(\pi_r^{m,p}(\psi)) - \Delta_n(\pi_{r-1}^{m,p}(\psi))|}{\mathcal{Q}(\pi_{r-1}^{m,p}(\psi), \pi_r^{m,p}(\psi))\mathcal{Q}(\pi_r^{m,p}(\psi), \psi^*)}\right\}\right) - 1\right] \\
&\qquad\qquad \leq K p^{-2} 4^{-r+p+1} \log(1 + N(2^{-r}, \widetilde{\Psi} \cap H_p, \rho))
\end{aligned}$$

for all $|\theta| \leq \varepsilon$, some $\delta > 0$ and some constant $K > 0$. Moreover note that $N(2^{-r}, \widetilde{\Psi} \cap H_p, \rho) \leq N(2^{-r+p+1}, \widetilde{\Psi}, \rho)$. Next

$$\mathrm{E}\left[\exp\left(\theta \sup_{\psi \in \widetilde{\Psi}^m \cap H_p} \left\{\frac{|\sqrt{n} \cdot \Delta_n(\pi_p^{m,p}(\psi))|}{\mathcal{R}^2(\psi, \psi^*)}\right\}\right) - 1\right] \lesssim p^{-2} \log(1 + N(2^{1+p}, \widetilde{\Psi}, \rho)).$$

Finally, we get for any $P > 0$

$$\begin{aligned}
\mathrm{E}\left[\exp\left(\theta \sup_{\psi \in \widetilde{\Psi}^m \cap (H_1 \cup \ldots \cup H_P)} \frac{|\sqrt{n} \cdot \Delta_n(\psi)|}{\mathcal{R}^2(\psi, \psi^*)}\right) - 1\right] &\lesssim \sum_{p=1}^{P} p^{-2} \sum_{r=p+1}^{q(m,p)} 4^{-r+p+1} \log(1 + N(2^{-r+p+1}, \widetilde{\Psi}, \rho)) \\
&\lesssim \sum_{p=1}^{P} p^{-2} \int_0^1 \log(1 + N(\sqrt{\varepsilon}, \widetilde{\Psi}, \rho))\, d\varepsilon \\
&\lesssim \int_0^1 \sqrt{\log(1 + N(\varepsilon, \widetilde{\Psi}, \rho))}\, d\varepsilon.
\end{aligned}$$

The proof of Proposition 6.3 is accomplished by letting $m \to \infty$ and $P \to \infty$.

8 Appendix

The following lemma is a straightforward generalization of Lemma 19.33 in van der Vaart (1998).

Lemma 8.1. *Let $\mathcal{X}$ be a finite collection of bounded real valued random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$, then*

$$\mathrm{E}\, \|\mathbb{G}_n[X]\|_{\mathcal{X}} \lesssim \frac{\sup_{X \in \mathcal{X}} |X|}{\sqrt{n}} \log(1 + |\mathcal{X}|) + \max_{X \in \mathcal{X}} \sqrt{\mathrm{E}\,[|X|^2]} \sqrt{\log(1 + |\mathcal{X}|)},$$

where $\mathbb{G}_n[X] = \frac{1}{n}\sum_{j=1}^{n}(X^{(j)} - \mathrm{E}[X])$ *and* $X^{(1)}, \ldots, X^{(n)}$ *are i.i.d. copies of* X.

Given a sequence of σ-algebras $(\mathcal{F}_n)$, $n \geq 1$ on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$ we call a sequence of integrable r.v. Y_n to be a forward martingale-difference sequence w.r.t. $(\mathcal{F}_n)$ if

- $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$
- Y_n is $\mathcal{F}_n$-measurable
- $\mathrm{E}[Y_n|\mathcal{F}_{n-1}] = 0$ a.s. for any $n \geq 1$

The following proposition can be found in Hitczenko (1990).

Proposition 8.2. *Let* (X_k) *be a forward martingale-difference sequence relative to* $\mathcal{F}_k$ *such that* $\mathrm{E}\,|X_k|^p < \infty$ *for some* $p \geq 2$ *and* $k = 1, \ldots, n$, *then*

$$\mathrm{E}\left|\sum_{k=1}^{n} X_k\right|^p \leq B(k\log^{-1}k)\max\left\{\sum_{k=1}^{n}\mathrm{E}\,|X_k|^p, \mathrm{E}\left[\sum_{k=1}^{n}\mathrm{E}[X_k^2|\mathcal{F}_{k-1}]\right]^{p/2}\right\}$$

for some constant B *not depending on* k.

The next inequality can be found in De la Pena, Klass and Lai (2004).

Lemma 8.3. *For any continuous local martingale* $(M_t)_{t\in[0,T]}$ *with* $M_0 = 0$

$$\mathrm{P}\left(\frac{\sup_{0\leq t\leq T}|M_t|}{\sqrt{\langle M\rangle_T \log\log(\langle M\rangle_T \vee e^2)}} \geq x\right) \leq C(\alpha)e^{-\alpha x^2},$$

where α *is a real number in* $(0, 1/2)$ *and* $C(\alpha)$ *is a positive constant.*

Acknowledgements I would like to thank John Schoenmakers and Mikhail Urusov for remarks and helpful discussions.

References

L. Andersen and M. Broadie (2004). A primal-dual simulation algorithm for pricing multidimensional American options. *Management Sciences*, **50**(9), 1222-1234.

D. Belomestny, Ch. Bender and J. Schoenmakers (2007). True upper bounds for Bermudan products via non-nested Monte Carlo. *Math. Financ.*, **19**(1), 53–71.

B. Bercu, E. Gassiat and E. Rio (2002). Concentration inequalities, large and moderate deviations for self-normalized empirical processes. *Ann. Probab.*, **30**(4), 1576–1604.

V. De la Pena, M. Klass and T. Lai (2004). Self-normalized processes: Exponential inequalities, moments bounds and iterated logarithm law. *Ann. Probab.*, **32**, 1902–1933.

V. Desai, V. Farias and C. Moallemi (2010). Pathwise optimization for optimal stopping problems. Preprint.

L. Devroye and G. Lugosi (2008). Local tail bounds for functions of independent random variables. *Annals of Probability*, **36**(1), 143–159.

L. Devroye, L. Györfi and G. Lugosi (1996). A probabilistic theory of pattern recognition. Application of Mathematics (New York), **31**, Springer.

P. Glasserman (2003). Monte Carlo Methods in Financial Engineering. Springer.

M. Haugh, L. Kogan (2004). Pricing American options: a duality approach. *Operations Research*, **52**(2), 258–270.

P. Hitczenko (1990). Best constants in martingale version of Rosenthal's inequality. *Annals of Probability*, **18**(4), 1656–1668.

M. Kohler, A. Krzyzak und H. Walk (2010). Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps. *Statistics and Decision*, **26**, 275–288.

M. Kohler (2010). A Review on Regression-based Monte Carlo Methods for Pricing American Options. In: L. Devroye, B. Karasözen, M. Kohler und R. Korn (Eds.) Recent Developments in Applied Probability and Statistics - Dedicated to the Memory of Jürgen Lehn. Springer.

M. Kosorok (2008). Introduction to Empirical Processes and Semiparametric Inference. Springer: New York.

L.C.G. Rogers (2002). Monte Carlo valuation of American options. *Mathematical Finance*, **12**, 271-286.

J. Schoenmakers (2011). Optimal dual martingales, their analysis and application to new algorithms for Bermudan products. SSRN paper.

A. van der Vaart (1998). Asymptotic Statistics. Cambridge University Press, Cambridge.

H. van Zanten (2005). On the rate of convergence of the maximum likelihood estimator in Brownian semimartingale models, *Bernoulli*, **11**(4), 643–664.

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multi-level Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.

[92] G. Kutyniok. Data Separation by Sparse Representations. Preprint 92, DFG-SPP 1324, May 2011.

[93] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing. Preprint 93, DFG-SPP 1324, May 2011.

[94] H.-C. Kreusler and H. Yserentant. The Mixed Regularity of Electronic Wave Functions in Fractional Order and Weighted Sobolev Spaces. Preprint 94, DFG-SPP 1324, June 2011.

[95] E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. Preprint 95, DFG-SPP 1324, June 2011.

[96] S. Kunis and I. Melzer. On the Butterfly Sparse Fourier Transform. Preprint 96, DFG-SPP 1324, June 2011.

[97] T. Rohwedder. The Continuous Coupled Cluster Formulation for the Electronic Schrödinger Equation. Preprint 97, DFG-SPP 1324, June 2011.

[98] T. Rohwedder and R. Schneider. Error Estimates for the Coupled Cluster Method. Preprint 98, DFG-SPP 1324, June 2011.

[99] P. A. Cioica and S. Dahlke. Spatial Besov Regularity for Semilinear Stochastic Partial Differential Equations on Bounded Lipschitz Domains. Preprint 99, DFG-SPP 1324, July 2011.

[100] L. Grasedyck and W. Hackbusch. An Introduction to Hierarchical (H-) Rank and TT-Rank of Tensors with Examples. Preprint 100, DFG-SPP 1324, August 2011.

[101] N. Chegini, S. Dahlke, U. Friedrich, and R. Stevenson. Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates. Preprint 101, DFG-SPP 1324, September 2011.

[102] S. Dahlke, P. Oswald, and T. Raasch. A Note on Quarkonial Systems and Multilevel Partition of Unity Methods. Preprint 102, DFG-SPP 1324, September 2011.

[103] A. Uschmajew. Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation. Preprint 103, DFG-SPP 1324, September 2011.

[104] S. Kvaal. Multiconfigurational time-dependent Hartree method for describing particle loss due to absorbing boundary conditions. Preprint 104, DFG-SPP 1324, September 2011.

[105] M. Guillemard and A. Iske. On Groupoid C*-Algebras, Persistent Homology and Time-Frequency Analysis. Preprint 105, DFG-SPP 1324, September 2011.

[106] A. Hinrichs, E. Novak, and H. Woźniakowski. Discontinuous information in the worst case and randomized settings. Preprint 106, DFG-SPP 1324, September 2011.

[107] M. Espig, W. Hackbusch, A. Litvinenko, H. Matthies, and E. Zander. Efficient Analysis of High Dimensional Data in Tensor Formats. Preprint 107, DFG-SPP 1324, September 2011.

[108] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization Problems in Contracted Tensor Networks. Preprint 108, DFG-SPP 1324, October 2011.

[109] S. Dereich, T. Müller-Gronbach, and K. Ritter. On the Complexity of Computing Quadrature Formulas for SDEs. Preprint 109, DFG-SPP 1324, October 2011.

[110] D. Belomestny. Solving optimal stopping problems by empirical dual optimization and penalization. Preprint 110, DFG-SPP 1324, November 2011.