(Stochastic) Dynamic Programming and Delaunay (dual) quantization

Gilles Pagès

joint work with V. Bally, 03 , B. Wilbertz '11

—

LPMA-Université Pierre et Marie Curie

—

Workshop "Stochastic Optimization and dynamic programming‘"
28 June 2012

From optimal stopping...

▷ Dynamics: Let $(X_t)_{t\in[0,T]}$ be a quasi-left continuous càdlàg dynamics, say

$$dX_t = b(t,X_t)dt + \sigma(t,X_t)dW_t + \kappa(t,X_t)d\zeta_t,\ X_0 = x \in \mathbb{R}^d$$

where are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$,

- $W = (W_t)_{t\in[0,T]}$ is q-dimensional Brownian motion,
- $\zeta = (\zeta_t)_{t\in[0,T]}$ is a martingale Lévy process with $\zeta^c \equiv 0$ and Lévy measure ν on $\mathbb{R}^d \setminus \{0\}$ satisfying $\displaystyle\int_{|z|\geq 1} |z|^p \nu(dz) < +\infty$, $p \in (1,+\infty)$.
- The functions b, σ, κ satisfy

$$b : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d,\ \sigma,\kappa : [0,T]\times\mathbb{R}^d \to \mathbb{M}(d,q,\mathbb{R}) \text{ are continuous,}$$

  Lipschitz in x uniformly in $t \in [0,T]$.

▷ Obstacle/reward process: $(h(t,X_t)_{t\in[0,T]})$ where $h : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ with polynomial growth

$$|h(t,X_t)| \leq C(1+|x|^r), \quad x \in \mathbb{R}^d, r \in (0,p).$$

▷ Optimal stopping problem... We consider the Snell enveloppe

$$Y_t := \mathbb{P}\text{- supess}\left\{\mathbb{E}\Big(h(\tau, X_\tau)\,|\,\mathcal{F}_t^{W,\zeta}\Big),\ \tau\in\mathcal{T}_{[t,T]}^{\mathcal{F}^{W,\zeta}}\right\}\geq h(t,X_t) \tag{1}$$

where $\mathcal{T}_{[t,T]}^{\mathcal{F}^{W,\zeta}} = \{\tau:\Omega\to[t,T],\ \tau\ \mathcal{F}^{W,\zeta}\text{-stopping time}\}$.

The Snell enveloppe represents the honest optimal mean gain when starting to play at time t if the reward is $h(s,X_s)$ when leaving the game at time $s\in[t,T]$. Under the above assumption

$$\tau_t^* = \inf\{s\in[t,T],\ Y_s = h(s,X_s)\} \qquad \text{is an optimal stopping time}$$

i.e.

$$Y_{\tau_t^*} = \mathbb{E}\Big(h(\tau_t^*, X_{\tau_t^*})\,|\,\mathcal{F}_t^{W,\zeta}\Big).$$

...to Variational Inequalities

Assume $\kappa \equiv 0$ (No jump component).

Theorem Under appropriate assumptions and in an appropriate sense

$$Y_t = u(t, X_t)$$

where u satisfies

$$\max\left(\frac{\partial u}{\partial t} + Lu, h - u\right) = 0, \quad u(T, x) = h(T, x)$$

Time discretization

Let

$$t_k^n = k\frac{T}{n},\ k = 0, \dots, n.$$

▷ $(X_{t_k^n})_{0\le k\le n}$ is an $(\mathcal{F}_{t_k^n})_{0\le k\le n}$-Markov chain with transition

$$P_k(x, dy) = \mathbb{P}(X_{t_{k+1}^n} \in dy \,|\, X_{t_k^n} = x).$$

▷ The $(\mathbb{P}, (\mathcal{F}_{t_k^n})_{0\le k\le n})$-Snell envelope (≡ Bermuda options)

$$\widetilde{Y}_{t_k^n} := \mathbb{P}\text{- supess}\left\{\mathbb{E}\Big(h(\tau, X_\tau) \,|\, \mathcal{F}_{t_k^n}^{W,\zeta}\Big),\ \tau \in \mathcal{T}_{t_k^n,T}^n\right\} \ge h(t_k^n, X_{t_k^n}) \tag{2}$$

where $\mathcal{T}_{t_k^n,T}^n = \{\tau : \Omega \to \{t_k^n, \dots, t_n^n = T\},\ \tau\ (\mathcal{F}_{t_\ell^n}^{W,\zeta})_{0\le \ell\le n}\text{-stopping time}\}$.

▷ The Backward Dynamic Programing Principle reads

$$\widetilde{Y}_{t_k^n} = \max\Big(h(t_k^n, X_{t_k^n}), \mathbb{E}\big(\widetilde{Y}_{t_{k+1}^n} \mid \cancel{\mathcal{F}_{t_k^n}}\, X_{t_k^n}\big)\Big), \quad \widetilde{Y}_T = h(T, X_T),$$

so that $\widetilde{Y}_{t_k^n} = \widetilde{u}_k(X_{t_k^n})$, $k = 0, \ldots, n$ satisfying

$$\widetilde{u}_k(x) = \max\Big(h(t_k^n, x), P_k(u_{k+1})(x)\Big), \quad \widetilde{u}_n(x) = h(T, x).$$

▷ The Backward Dynamic Programing Principle reads

$$\widetilde{Y}_{t_k^n} = \max\Big(h(t_k^n, X_{t_k^n}), \mathbb{E}\big(\widetilde{Y}_{t_{k+1}^n} \,|\, \cancel{\mathcal{F}_{t_k^n}}\, X_{t_k^n}\big)\Big), \quad \widetilde{Y}_T = h(T, X_T),$$

so that $\widetilde{Y}_{t_k^n} = \widetilde{u}_k(X_{t_k^n})$, $k = 0, \ldots, n$ satisfying

$$\widetilde{u}_k(x) = \max\Big(h(t_k^n, x), P_k(u_{k+1})(x)\Big), \quad \widetilde{u}_n(x) = h(T, x).$$

▷ Theorem (Bally-P., SPA 2003) (a) If h is Lipschitz in x, uniformly in $t \in [0, T]$,

$$\forall\, p \in (0, r), \quad \Big\| \max_{0 \le k \le n} |Y_{t_k^n} - \widetilde{Y}_{t_k^n}| \Big\|_p \le C_{b,\sigma,\kappa,h,T} \sqrt{\frac{T}{n}}$$

(b) If furthermore h is semi-convex *i.e.* there exists $\delta_h [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ bounded s.t.

$$\exists\, \rho > 0,\ \forall\, x, y \in \mathbb{R}^d,\ h(t, y) - h(t, x) \ge (\delta_h(t, x)|y - x) - \rho|y - x|^2$$

then

$$\Big\| \max_{0 \le k \le n} |Y_{t_k^n} - \widetilde{Y}_{t_k^n}| \Big\|_p \le C_{b,\sigma,\kappa,h,T} \frac{T}{n}.$$

▷ Gaps:

▷ Gaps:

- [hedge? ...]

▷ Gaps:

- [hedge? ...]
- Simulation of the Markov chain $(X_{t_k^n})_{0\le k\le n}$ especially when d or $q \ge 2$

▷ Gaps:

- [hedge? ...]
- Simulation of the Markov chain $(X_{t_k^n})_{0\le k\le n}$ especially when d or $q \ge 2$
- Computation of conditional expectations...

Simulatable Markov chain: Euler scheme

The Euler scheme of (SDE) is defined by

$$\begin{aligned}\bar{X}^n_{t^n_{k+1}} = \bar{X}^n_{t^n_k} + \frac{T}{n} b(t^n_k, \bar{X}^n_{t^n_k}) \quad &+ \quad \sigma(t^n_k, \bar{X}^n_{t^n_k})(W_{t^n_{k+1}} - W_{t^n_k}) \\ &+ \quad \kappa(t^n_k, \bar{X}^n_{t^n_k})(\zeta_{t^n_{k+1}} - \zeta_{t^n_k})\end{aligned}$$

- If $\kappa \equiv 0$, it is always a simulatable Markov chain with transition

$$\bar{P}^{(n)}(x, dy) = \mathbb{P}(\bar{X}^n_{t^n_{k+1}} \in dy \,|\, \bar{X}^n_{t^n_k} = x)$$

and

$$\left\| \max_{k=0,\ldots,n} |X_{t^n_k} - \bar{X}^n_{t^n_k} \right\|_p \leq C_{b,\sigma,T} \sqrt{\frac{T}{n}}$$

▷ The Backward Dynamic Programing Principle reads

$$\bar{Y}_{t_k^n} = \max\Big(h(t_k^n, \bar{X}_{t_k^n}^n), \mathbb{E}\big(\bar{Y}_{t_{k+1}^n} \mid \cancel{\mathcal{F}_{t_k^n}}\, \bar{X}_{t_k^n}^n\big)\Big), \quad Y_T = h(T, \bar{X}_T^n),$$

so that $\bar{Y}_{t_k^n} = \bar{u}_k(\bar{X}_{t_k^n}^n)$, $k = 0, \ldots, n$ satisfying

$$\bar{u}_k(x) = \max\Big(h(t_k^n, x), \bar{P}^{(n)}(u_{k+1})(x)\Big), \quad \bar{u}_n(x) = h(T, x).$$

▷ The Backward Dynamic Programing Principle reads

$$\bar{Y}_{t_k^n} = \max\Big(h(t_k^n, \bar{X}_{t_k^n}^n), \mathbb{E}\big(\bar{Y}_{t_{k+1}^n} \mid \cancel{\mathcal{F}_{t_k^n}}\, \bar{X}_{t_k^n}^n\big)\Big), \quad Y_T = h(T, \bar{X}_T^n),$$

so that $\bar{Y}_{t_k^n} = \bar{u}_k(\bar{X}_{t_k^n}^n)$, $k = 0, \ldots, n$ satisfying

$$\bar{u}_k(x) = \max\Big(h(t_k^n, x), \bar{P}^{(n)}(u_{k+1})(x)\Big), \quad \bar{u}_n(x) = h(T, x).$$

...and corresponds to an a "Bermuda like" optimal stopping problem (with $\bar{X}$ instead of X).

▷ The Backward Dynamic Programing Principle reads

$$\bar{Y}_{t_k^n} = \max\Big(h(t_k^n, \bar{X}_{t_k^n}^n), \mathbb{E}\big(\bar{Y}_{t_{k+1}^n} \,|\, \cancel{\mathcal{F}_{t_k^n}}\, \bar{X}_{t_k^n}^n\big)\Big), \quad Y_T = h(T, \bar{X}_T^n),$$

so that $\bar{Y}_{t_k^n} = \bar{u}_k(\bar{X}_{t_k^n}^n)$, $k = 0, \ldots, n$ satisfying

$$\bar{u}_k(x) = \max\Big(h(t_k^n, x), \bar{P}^{(n)}(u_{k+1})(x)\Big), \quad \bar{u}_n(x) = h(T, x).$$

...and corresponds to an a "Bermuda like" optimal stopping problem (with $\bar{X}$ instead of X).

▷ Theorem (Bally-P., SPA 2003) If h is Lipschitz in x, uniformly in $t \in [0, T]$,

$$\forall\, p \in (0, r), \quad \Big\| \max_{0 \le k \le n} |Y_{t_k^n} - \bar{Y}_{t_k^n}| \Big\|_p \le C_{b,\sigma,\kappa,h,T} \sqrt{\frac{T}{n}}$$

Remark. No loss w.r.t. the Euler scheme itself.

• Otherwise it depends on the simulability of ζ (see Protter-talay, Jacod, Jacod-Protter, etc for convergence rate(s) of the Euler schmes existence of approximate schemes.

• In case of non simulability: design of approximate schemes: Roszincky's "Wienerisation of small jumps", (see Roszincky, Cohen, Rubenthaler, Panloup, etc)...

Abstract "Bermuda" Markov optimal stopping framework

▷ Let $(X_k)_{0\le k\le n}$ be an $\mathbb{R}^d$-valued homogeneous Feller Markov chain defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with transition

$$P(x, dy) = \mathbb{P}(X_{k+1} \in dy \,|\, X_k = x),\ k = 0, \ldots, n-1.$$

Filtration : $\mathcal{F}_k^X = \sigma(X_0, \ldots, X_k)$, $k = 0, \ldots, n$.

MArkov property :

$$\mathbb{E}(f(X_{k+1}) \,|\, \mathcal{F}X_k) = \mathbb{E}(f(X_{k+1}) \,|\, X_k) = \int_{\mathbb{R}^d} f(y)P(x, dy) := Pf(x).$$

Abstract "Bermuda" Markov optimal stopping framework

▷ Let $(X_k)_{0\le k\le n}$ be an $\mathbb{R}^d$-valued homogeneous Feller Markov chain defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with transition

$$P(x, dy) = \mathbb{P}(X_{k+1} \in dy \,|\, X_k = x),\ k = 0, \ldots, n-1.$$

Filtration : $\mathcal{F}_k^X = \sigma(X_0, \ldots, X_k)$, $k = 0, \ldots, n$.

MArkov property :

$$\mathbb{E}(f(X_{k+1}) \,|\, \mathcal{F}X_k) = \mathbb{E}(f(X_{k+1}) \,|\, X_k) = \int_{\mathbb{R}^d} f(y)P(x, dy) := Pf(x).$$

▷ PROBLEM TO BE SOLVED: Compute the premium of a Bermuda option with an *integrable* payoff $\Big(h_k(X_k)\Big)_{0\le k\le n}$ *i.e.*

the right to receive $h_k(X_k)$ *once* between $k = 0$ and $k = n$.

Stopping time $\equiv$ "honnest stopping rule".

$$\tau : \Omega \to \{0, \ldots, n\}, \qquad \{\tau = k\} \in \mathcal{F}_k^X,\ k = 0, \ldots, n.$$

$$V_0 = v_0(X_0) = \text{esssup}\Big\{\mathbb{E}\Big(h_\tau(X_\tau) \,|\, \mathcal{F}_0^X\Big),\ \tau \text{ stopping time}\Big\}$$

...and more generally, its premium at time k,

$$V_k = v_k(X_k) = \text{esssup}_{\tau \in \mathcal{T}_{k,n}} \mathbb{E}\Big(h_\tau(X_\tau) \,|\, \mathcal{F}_k^X\Big),\ k = 0, \ldots, n.$$

where

$$\mathcal{T}_{k,n} := \Big\{\tau : \Omega \to \{k, \ldots, n\}\ \mathcal{F}^X\text{-stopping time}\Big\}.$$

$(V_k)_{0 \le k \le n}$ is called the *Snell envelope* of $(h_k(X_k))_{0 \le k \le n}$.

Backward Dynamic Programing Principle

▷ The $(\mathbb{P}, \mathcal{F}^X)$-*Snell envelope* $(V_k)_{0\le k\le n}$ of the so-called *obstacle process* $(h(X_k))_{0\le k\le n}$ satisfies the *BDPP*

$$V_n = h_n(X_n), \quad V_k = \max\Big(h_k(X_k), \mathbb{E}\big(V_{k+1} \,|\, \cancel{\mathcal{F}_k^X} X_k\big)\Big)$$

or equivalently (in distribution) $V_k = v_k(X_k)$ where

$$v_n = h_n \qquad v_k = \max\Big(h_k, Pv_{k+1}\Big),\ k = 0, \ldots, n-1.$$

▷ Alternative approach (cf. Longstaff-Schwarz, 1993) : the *BDPP* approach for optimal stopping times

$$\tau_k = \min\{\ell \ge k,\ V_k = h_k(X_k)\},\ k = 0, \ldots, n$$

which satisfy

$$\tau_k = k\mathbf{1}_{\{h_k(X_k) > \mathbb{E}(V_{k+1} \,|\, X_k)\}} + \tau_{k+1}\mathbf{1}_{\{h_k(X_k) \le \mathbb{E}(V_{k+1} \,|\, X_k)\}}.$$

and $V_k = \mathbb{E}\big(h_{\tau_k}(X_{\tau_k}) \,|\, X_k\big)$.

Conditional expectation computation

▷ In both cases the point is to compute/estimate

$$\mathbb{E}\big(V_{k+1}\,|\,X_k\big) = \mathbb{E}\big(v_{k+1}(X_{k+1})\,|\,X_k\big) = \mathbb{E}h_{\tau_{k+1}}(X_{\tau_{k+1}})\,|\,X_k),\ k = 0, \ldots, n-1$$

Two approaches have been developed

- Randomization of the *BDPP* (ex: regression methods, Monte Carlo-Malliavin)
- Structural approximation of the Markov dynamics (ex: tree methods)

Markov Dynamics approximation: the paradigm of Quantized BDPP

▷ Two-folded natural idea

Markov Dynamics approximation: the paradigm of Quantized BDPP

▷ TWO-FOLDED NATURAL IDEA

- Step 1 (Markov dynamics Approximation): Approximation of X_k

$$X_k : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}^d \rightsquigarrow \widehat{X}_k : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \Gamma_k := \left\{x_1^k, \ldots, x_{N_k}^k\right\}$$

where

$$\widehat{X}_k = \pi_k(X_k, U_k), \quad k = 0, \ldots, n,$$

where $(U_k)_{0 \le k \le n}$ is an i.i.d. sequence of $U([0,1]^d)$-distributed *exogeneous* (= simulated) r.v.'s.

Note that $(\widehat{X}_k)_{0 \le k \le n}$ is usually NOT a Markov chain.

Markov Dynamics approximation: the paradigm of Quantized BDPP

▷ TWO-FOLDED NATURAL IDEA

- Step 1 (Markov dynamics Approximation): Approximation of X_k

$$X_k : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{R}^d \rightsquigarrow \widehat{X}_k : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \Gamma_k := \{x_1^k, \ldots, x_{N_k}^k\}$$

  where

$$\widehat{X}_k = \pi_k(X_k, U_k), \quad k = 0, \ldots, n,$$

  where $(U_k)_{0 \le k \le n}$ is an i.i.d. sequence of $U([0,1]^d)$-distributed *exogeneous* (= simulated) r.v.'s.

  Note that $(\widehat{X}_k)_{0 \le k \le n}$ is usually NOT a Markov chain.

- Step 2: Force the Markov property in the $BDPP$:

$$\widehat{V}_n = h(\widehat{X}_n), \quad \widehat{V}_k = \max\Big(h_k(\widehat{X}_k), \mathbb{E}\big(\widehat{V}_{k+1} \,|\, \widehat{X}_k\big)\Big)$$

Resulting tree algorithm

▷ THE RESULTING ALGORITHM:

$$\widehat{V}_k = \widehat{v}_k(\widehat{X}_k),\ k = 0, \ldots, n$$

with

$$\forall\, i \in \{1, \ldots, N_k\}, \quad \widehat{v}_k(x_i^k) = \max\Big(h_k(x_i^k), \widehat{P}(\widehat{v}_{k+1})(x_i^k)\Big)$$

where $\widehat{P}$ displays on Borel test functions

$$\begin{aligned} \widehat{P}(f)(x_i^k) &= \sum_{j=1}^{N_{k+1}} \widehat{\pi}_{ij}^{k,k+1} f(x_j^{k+1}) \\ \widehat{\pi}_{ij}^{k,k+1} &= \mathbb{P}\Big(\widehat{X}_{k+1} = x_j^{k+1} \,|\, \widehat{X}_k = x_i^k\Big). \end{aligned}$$

▷ *Markov Dynamics approximation*: The matrices $\widehat{\pi}^{k,k+1}$, $k = 0, \ldots, n$ need to be computed by a massive Monte Carlo simulation.

Quantization tree ($d = 1$, 3 periods)

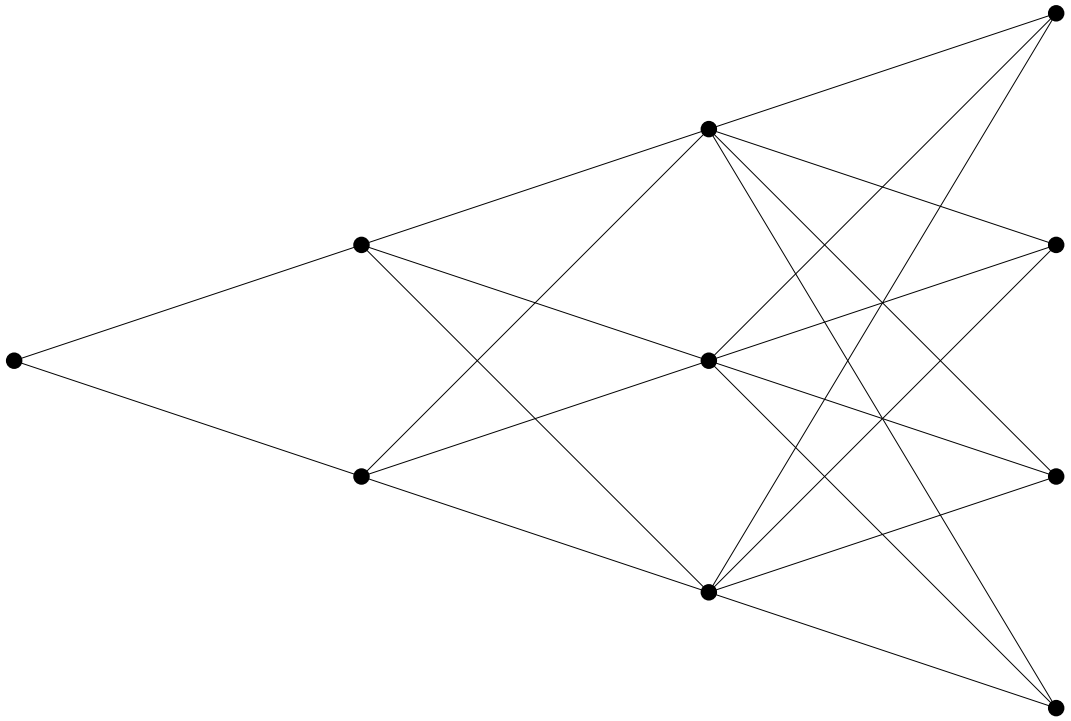


Figure: A typical 1-dimensional quantization tree

- A quantization tree is not re-combining.
- But its size is designed *a priori* (and subject to possible optimization).

Conditional expectation approximation by quantization

▷The natural idea is to use the approximation

$$\mathbb{E}\big(f(X_2)\,|\,X_1\big) \approx \mathbb{E}\big(\hat{f}(\widehat{X}_2)\,|\,\widehat{X}_1\big).$$

since

$$\mathbb{E}\big(\hat{f}(\widehat{X}_2)\,|\,\widehat{X}_1 = x_i^1\big) = \sum_{j=1}^{N_2} \pi_{ij}\hat{f}(x_j^2)$$

is computable.

▷ Can we control in L^p the induces error based on the spatial discretization L^p error ?

$$\Big\|\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(\hat{f}(\widehat{X}_2)|\widehat{X}_1\big)\Big\|_p \leq \Phi\Big(\|X_1 - \widehat{X}_1\|_p, \|f(X_2) - \widehat{f}(\widehat{X}_2)\|_p\Big) \text{ ????}$$

The key is the following one-step estimate

Proposition (Key lemma)

Let $p \in [1, +\infty)$. Assume that $\|X_1\|_p + \|X_2\|_p < +\infty$. Assume $P(x, dy) = P_1(x, dy)$ uniformly propagates Lipschitz functions i.e., for every Lipschitz continuous $f : \mathbb{R}^d \to \mathbb{R}$,

$$[Pf]_{\mathrm{Lip}} \leq [P]_{\mathrm{Lip}} [f]_{\mathrm{Lip}}.$$

(a) If $p = 2$, then

$$\left\| \mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(\hat{f}(\widehat{X}_2)|\widehat{X}_1\big) \right\|_2 \leq \Big([f]_{\mathrm{Lip}}^2 [P]_{\mathrm{Lip}}^2 \|X_1 - \widehat{X}_1\|_2^2 + \|f(X_2) - \hat{f}(\widehat{X}_2)\|_2^2 \Big)^{\frac{1}{2}}$$

(b) If $p \neq 2$

$$\left\| \mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(\hat{f}(\widehat{X}_2)|\widehat{X}_1\big) \right\|_p \leq [f]_{\mathrm{Lip}} [P]_{\mathrm{Lip}} \|X_1 - \widehat{X}_1\|_p + \|f(X_2) - \hat{f}(\widehat{X}_2)\|_p$$

Proof of (a)**.** Keep in mind $\widehat{X}_1 = \pi_1(X_1, U_1)$. For notational convenience we write $\pi_u(x)$ for $\pi_1(x, u)$.

Proof of (a)**.** Keep in mind $\widehat{X}_1 = \pi_1(X_1, U_1)$. For notational convenience we write $\pi_u(x)$ for $\pi_1(x, u)$.

STEP 1 Using that U_1 is independent of $(X_1, \widehat{X}_2)$

$$\mathbb{E}\big(f(\widehat{X}_2)|\widehat{X}_1\big) = \int_{\mathbb{R}^{d_0}} \mathbb{E}\big(f(\widehat{X}_2)|\pi_u(X_1)\big)\mathbb{P}_{U_1}(du)$$

so that

$$\begin{aligned}\Big\|\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(f(\widehat{X}_2)|\widehat{X}_1\big)\Big\|_2^2 &= \mathbb{E}\left(\int_{\mathbb{R}^{d_0}} \mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(f(\widehat{X}_2)|\pi_u(X_1)\big)\mathbb{P}_{U_1}(du)\right)^2 \\ &\leq \int_{\mathbb{R}^{d_0}} \mathbb{E}\Big(\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(\hat{f}(\widehat{X}_2)|\pi_u(X_1)\big)\Big)^2 \mathbb{P}_{U_1}(du).\end{aligned}$$

by Jensen's Inequality.

STEP 2

$$\begin{aligned}\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(f(\widehat{X}_2)|\pi_u(X_1)\big) &= \Big[\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\Big(\mathbb{E}\big(f(X_2)|X_1\big)|\pi_u(X_1)\Big)\Big] \\ &\overset{\perp}{+} \Big[\mathbb{E}\big(f(X_2)|\pi_u(X_1)\big) - \mathbb{E}\big(f(\widehat{X}_2)|\pi_u(X_1)\big)\Big]\end{aligned}$$

where we first used that $\sigma(\pi_u(X_1)) \subset \sigma(X_1)$. The orthogonality follow from the very definition of conditional expectation $\mathbb{E}(\,.\,|\sigma(\pi_u(X_1)))$.

Pythagorus Theorem implies

$$\begin{aligned}\big\|\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(f(\widehat{X}_2)|\pi_u(X_1)\big)\big\|_2^2 &= \big\|\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(f(X_2)|\pi_u(X_1)\big)\big\|_2^2 \\ &\quad + \big\|\mathbb{E}\big(f(X_2) - f(\widehat{X}_2)|\pi_u(X_1)\big)\big\|_2^2 \\ &\leq \big\|\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(f(X_2)|\pi_u(X_1)\big)\big\|_2^2 \\ &\quad + \|f(X_2) - \hat{f}(\widehat{X}_2)\|_2^2\end{aligned}$$

by the contraction property. Now, using again that $\sigma(\pi_u(X_1)) \subset \sigma(X_1)$, we get

$$\begin{aligned}\mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\big(f(X_2)|\pi_u(X_1)\big) &= \mathbb{E}\big(f(X_2)|X_1\big) - \mathbb{E}\Big(\mathbb{E}\big(f(X_2)|X_1\big)|\pi_u(X_1)\Big) \\ &= Pf(X_1) - \mathbb{E}\big(Pf(X_1)|\pi_u(X_1)\big)\end{aligned}$$

so that ...

$$\begin{aligned}\left\|\mathbb{E}\big(f(X_2)|X_1\big)-\mathbb{E}\big(f(X_2)|\pi_u(X_1)\big)\right\|_2^2 &= \|Pf(X_1)-\mathbb{E}\big(Pf(X_1)|\pi_u(X_1)\big)\|_2^2\\ &\le \|Pf(X_1)-Pf(\pi_u(X_1))\|_2^2\\ &\le [Pf]_{\mathrm{Lip}}^2\,\|X_1-\pi_u(X_1)\|_2^2\,.\end{aligned}$$

Hence

$$\mathbb{E}\Big(\mathbb{E}\big(f(X_2)|X_1\big)-\mathbb{E}\big(f(\widehat{X}_2)|\pi_u(X_1)\big)\Big)^2 \le [f]_{\mathrm{Lip}}\Big(\|X_2-\widehat{X}_2\|_2^2+[P]_{\mathrm{Lip}}\,\|X_1-\pi_u(X_1)\|_2^2\Big)$$

Integrating with respect to $\mathbb{P}_{U_1}(du)$ (*i.e.* the exogenous innovation) yields

$$\left\|\mathbb{E}\big(f(X_2)|X_1\big)-\mathbb{E}\big(\hat{f}(\widehat{X}_2)|\widehat{X}_1\big)\right\|_2^2 \le \Big(\|f(X_2)-\hat{f}(\widehat{X}_2)\|_2^2+[f]_{\mathrm{Lip}}[P]_{\mathrm{Lip}}\,\|X_1-\pi_u(X_1)\|_2^2\Big)$$

since, by the chain rule for conditional expectation,

$$\|X_1-\widehat{X}_1\|_2^2=\int_{\mathbb{R}^{d_0}}\|X_1-\pi_u(X_1)\|_2^2\mathbb{P}_{U_1}(du). \qquad \square$$

A priori error bounds

Then we have the following general result about the rate of approximation of the Snell envelope $(V_k)_{0\le k\le n}$ by its "quantized" counterpart $(\widehat{V}_k)_{0\le k\le n}$.

Theorem (Bally-P.-Printems '01, P.-Wilbertz '10)

Let $p\in[1,+\infty)$. Assume that all the functions h_k, $k=0,\dots,n$, are Lipschitz continuous. and that the $P_k(x,dy)$ uniformly propagate Lipschitz functions i.e.

$$[P]_{\mathrm{Lip}} := \max_{0\le k\le n-1}[P_k]_{\mathrm{Lip}} < +\infty \text{ and } \max_{0\le k\le n}\left(\|X_k\|_p + \|\widehat{X}_k\|_p\right) < +\infty.$$

(a) If $p=2$, then, for every $k\in\{0,\dots,n\}$,

$$\|V_k-\widehat{V}_k\|_2 \le \sqrt{2}\left(\sum_{\ell=k}^{n}\left(C_{n,\ell}([P]_{\mathrm{Lip}},[h_\ell]_{\mathrm{Lip}})\right)^2\|X_\ell-\widehat{X}_\ell\|_2^2\right)^{\frac{1}{2}}$$

(b) If $p\neq 2$, then for every $k\in\{0,\dots,n\}$,

$$\|V_k-\widehat{V}_k\|_p \le 2\sum_{\ell=k}^{n}C_{n,\ell}([P]_{\mathrm{Lip}},[h_\ell]_{\mathrm{Lip}})\|X_\ell-\widehat{X}_\ell\|_p$$

$$\text{where}\qquad C_{n,k}([P]_{\mathrm{Lip}},[h_.]_{\mathrm{Lip}}) = \max_{k\le\ell\le n}\left([P]_{\mathrm{Lip}}^{\ell-k}[h_\ell]_{\mathrm{Lip}}\right).$$

Proof (not so sketchy)

Proof (not so sketchy)

Proof. STEP 1. *The functions v_k are Lipschitz.*

$$V_k = v_k(X_k), \quad k = 0, \ldots, n,$$

where the functions v_k are Lipschitz continuous satisfying

$$v_n = h_n \quad \text{and} \quad v_k = \max(h_k, Pv_{k+1}),\ k = 0, \ldots, n-1.$$

In particular, for every $k = 0, \ldots, n$ (with the convention $[v_{n+1}]_{\mathrm{Lip}} = 0$),

$$[v_k]_{\mathrm{Lip}} \leq \max\left([h_k]_{\mathrm{Lip}}, [P]_{\mathrm{Lip}}[v_{k+1}]_{\mathrm{Lip}}\right)$$

since $|\sup_{i\in I} a_i - \sup_{i\in I} b_i| \leq \sup_{i\in I} |a_i - b_i|$.
Standard induction yields

$$[v_k]_{\mathrm{Lip}} \leq \max_{k\leq \ell\leq n}\left([P]_{\mathrm{Lip}}^{\ell-k}[h_\ell]_{\mathrm{Lip}}\right),\ k = 0, \ldots, n.$$

Proof (not so sketchy)

Proof. STEP 1. *The functions v_k are Lipschitz.*

$$V_k = v_k(X_k), \quad k = 0, \dots, n,$$

where the functions v_k are Lipschitz continuous satisfying

$$v_n = h_n \quad \text{and} \quad v_k = \max(h_k, Pv_{k+1}),\ k = 0, \dots, n-1.$$

In particular, for every $k = 0, \dots, n$ (with the convention $[v_{n+1}]_{\mathrm{Lip}} = 0$),

$$[v_k]_{\mathrm{Lip}} \le \max\left([h_k]_{\mathrm{Lip}}, [P]_{\mathrm{Lip}}[v_{k+1}]_{\mathrm{Lip}}\right)$$

since $|\sup_{i\in I} a_i - \sup_{i\in I} b_i| \le \sup_{i\in I} |a_i - b_i|$.
Standard induction yields

$$[v_k]_{\mathrm{Lip}} \le \max_{k\le \ell\le n}\left([P]_{\mathrm{Lip}}^{\ell-k}[h_\ell]_{\mathrm{Lip}}\right),\ k = 0, \dots, n.$$

STEP 2.

$$\begin{aligned} |V_k - \widehat{V}_k|^2 &\le \max\left(|h_k(X_k) - h_k(\widehat{X}_k)|^2, |\mathbb{E}(V_{k+1}|X_k) - \mathbb{E}(\widehat{V}_{k+1}|\widehat{X}_k)|^2\right) \\ &\le |h_k(X_k) - h_k(\widehat{X}_k)|^2 + |\mathbb{E}(v_{k+1}(X_{k+1})|X_k) - \mathbb{E}(\widehat{v}_{k+1}(\widehat{X}_{k+1}))|\widehat{X}_k)|^2 \end{aligned}$$

so that by the key lemma

$$\left\|V_k - \widehat{V}_k\right\|_2^2 \le [h_k]_{\mathrm{Lip}}^2\left\|X_k - \widehat{X}_k\right\|_2^2 + [Pv_{k+1}]_{\mathrm{Lip}}\|X_k - \widehat{X}_k\|_2^2 + \left\|V_{k+1} - \widehat{V}_{k+1}\right\|_2^2$$

The result follows from the bounds on $[Pv_{k+1}]_{\mathrm{Lip}}$ and the discrete Gronwall lemma.
□

Applications to diffusions: the Euler scheme (homogeneous)

[Homogeneous case for expository].

- For the above jump diffusion (when $p \geq 2$), the Euler scheme with step $\frac{T}{n}$ satisfies

$$\begin{aligned}
\mathbb{E}\big(\bar{X}^{(n),y}_{\frac{T}{n}} &- \bar{X}^{(n),x}_{\frac{T}{n}}\big)^2 \\
&= \mathbb{E}\Big|y - x + \frac{T}{n}(b(y) - b(x)) + (\sigma(Y) - \sigma(x))W_{\frac{T}{n}} + (\kappa(y) - \kappa(x))\zeta_{\frac{T}{n}}\Big|^2 \\
&= |y - x|^2 + \big(\frac{T}{n}\big)^2|b(y) - b(x)|^2 + \frac{T}{n}|\sigma(y) - \sigma(x)|^2 \\
&\quad +(\kappa(y) - \kappa(x))^2\mathbb{E}\zeta^2_{\frac{T}{n}} \\
&\leq |y - x|^2\Big(1 + 2C_{b,\sigma,\kappa,T}\frac{T}{n}\Big).
\end{aligned}$$

so that

$$|\bar{P}^{(n)}(f)(x) - \bar{P}^{(n)}(f)(y)| \leq [f]_{\mathrm{Lip}}\|\bar{X}^{(n),y}_{\frac{T}{n}} - \bar{X}^{(n),x}_{\frac{T}{n}}\|_1 \leq [f]_{\mathrm{Lip}}\|\bar{X}^{(n),y}_{\frac{T}{n}} - \bar{X}^{(n),x}_{\frac{T}{n}}\|_2$$

and finally

$$[\bar{P}^{(n)}]_{\mathrm{Lip}} \leq \Big(1 + C_{b,\sigma,\kappa,T}\frac{T}{n}\Big).$$

Conclusion:

$$\sup_n \max_{0\leq k\leq n} [\bar{P}^{(n)}]^k_{\mathrm{Lip}} \leq e^{C_{b,\sigma,\kappa,T}T}$$

Proposition (Bally-P.-Printems '03 [BP03], Wilbertz-P., (2010) [PW09])

We consider the optimal stopping problem related to a Brownian diffusions ($\kappa \equiv 0$) with coefficient b and σ and with obstacle function $h(t,x)$, all assumed to be Lipschitz in $x \in \mathbb{R}^d$ uniformly in $t \in [0,T]$.

$$\begin{aligned}\left\| \max_{0\le k\le n} |Y_{t_k^n} - \widehat{\bar{Y}^n_{t_k^n}}| \right\|_2 &\le C\sqrt{\frac{T}{n}} + C_{b,\sigma,h,T}\left(\sum_{k=0}^{n} \left\| \bar{X}^{(n)}_{t_k^n} - \widehat{\bar{X}^{(n)}}_{t_k^n} \right\|_2^2\right)^{\frac{1}{2}} \\ &\le C\left(\sqrt{\frac{T}{n}} + \sqrt{n}\max_{0\le k\le n}\left\| \bar{X}^{(n)}_{t_k^n} - \widehat{\bar{X}^{(n)}}_{t_k^n} \right\|_2\right)\end{aligned}$$

This strongly suggests to investigate methods to reduce/minimize

the quantization error(s)

$$\left\| X - \widehat{X} \right\|_p \ldots \text{especially when } p = 2.$$

Optimal quantization(s)

or

How to optimize the approximation of X by $\widehat{X}$ taking at most N values?

We temporarily turn now to this *static* problem also known as

OPTIMAL (VECTOR) QUANTIZATION...

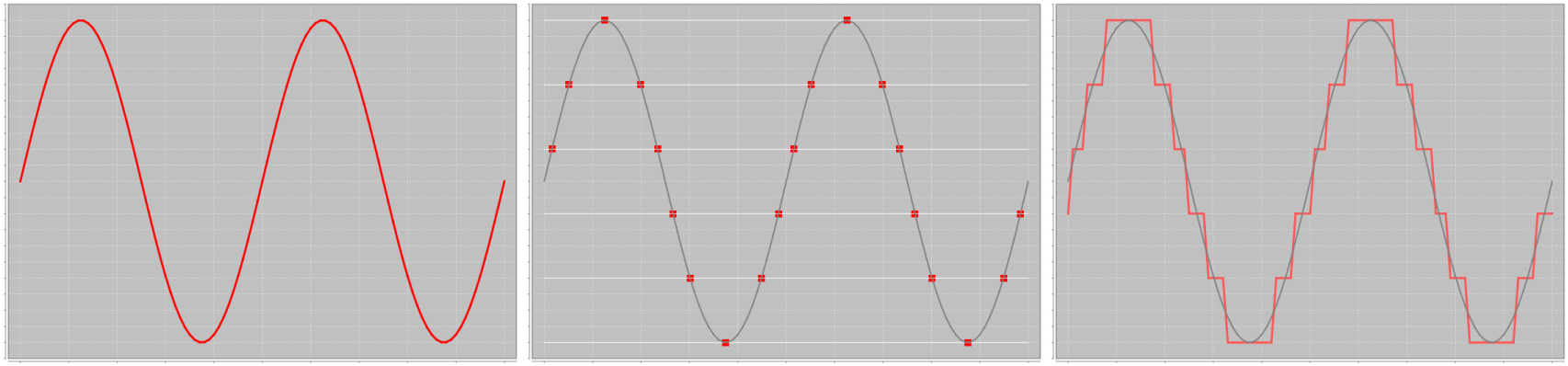
Let $\Gamma \subset \mathbb{R}^d$ be a *grid* with size at most $N \geq 1$.

- $\widehat{X} = \pi(X)$, $\pi : \mathbb{R}^d \to \Gamma$ ($\rightsquigarrow$ Voronoi quantization).
- $\widehat{X} = \pi(X, U)$, $\pi : \mathbb{R}^d \times [0,1] \to \Gamma$, $U \perp\!\!\!\perp X$ ($\rightsquigarrow$ Delaunay (or dual) quantization).

$\boxed{\text{In practice how to optimize the underlying grid } \Gamma?}$

What is Vector Quantization?

- Has its origin in the fields of signal processing in the late 1940's
- Describes the discretization of a random signal and analyses the recovery of the original signal from the discrete one



- Examples: Pulse-Code-Modulation(PCM), JPEG-Compression
- Extensive Survey about the IEEE-History: [GN98]
- Mathematical Foundation of Quantization Theory: [GL00]

Voronoi-Quantization

Voronoi-Quantization

▷ Let $X : (\Omega, \mathcal{S}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}^d, \|\cdot\|)$ be a random vector such that

$$\mathbb{E}\|X\|^p < +\infty \qquad \text{for some } p \in [1, \infty).$$

Voronoi-Quantization

▷ Let $X : (\Omega, \mathcal{S}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}^d, \|\cdot\|)$ be a random vector such that

$$\mathbb{E}\|X\|^p < +\infty \qquad \text{for some } p \in [1, \infty).$$

▷ Given a (finite) "grid" $\Gamma = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^d$, we discretize of the r.v. X using a Nearest Neighbor projection.

- Let $\big(C_i(\Gamma)\big)_{1 \le i \le N}$ be a *Voronoi partition* of $\mathbb{R}^d$ generated by Γ, *i.e.* $\big(C_i(\Gamma)\big)$ is a Borel partition of $\mathbb{R}^d$ satisfying

$$C_i(\Gamma) \subset \Big\{ z \in \mathbb{R}^d : \|z - x_i\| \le \min_{1 \le j \le N} \|z - x_j\| \Big\}.$$

- Let $\pi_\Gamma : \mathbb{R}^d \to \Gamma$ the induced Nearest Neighbor projection,

$$\xi \mapsto \sum_{i=1}^{N} x_i \mathbf{1}_{C_i(\Gamma)}(\xi).$$

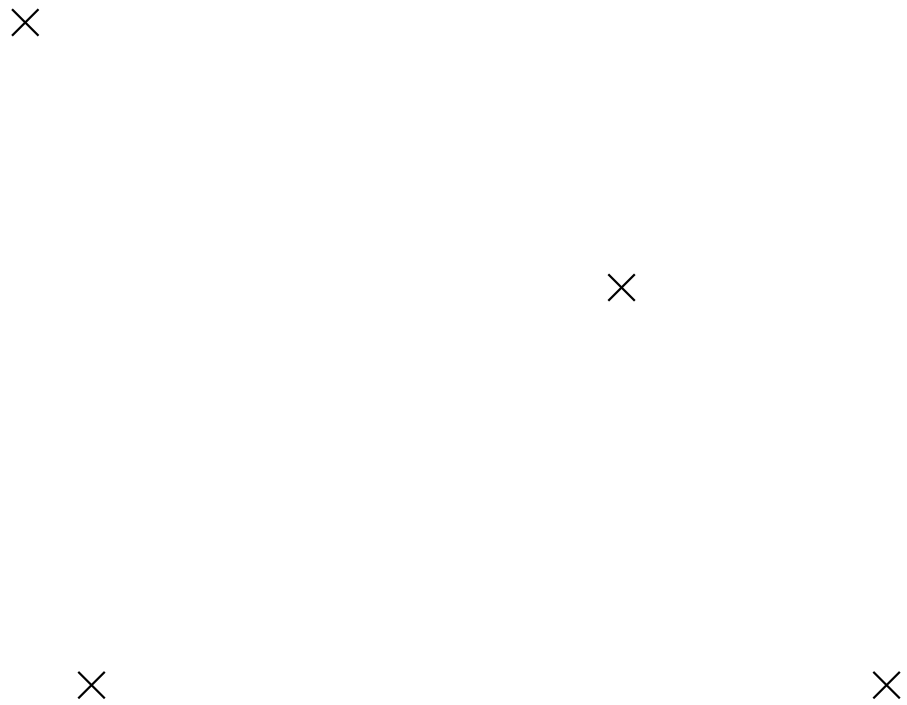
so that

$$\|\xi - \pi_\Gamma(\xi)\| = \mathrm{dist}(\xi, \Gamma)$$

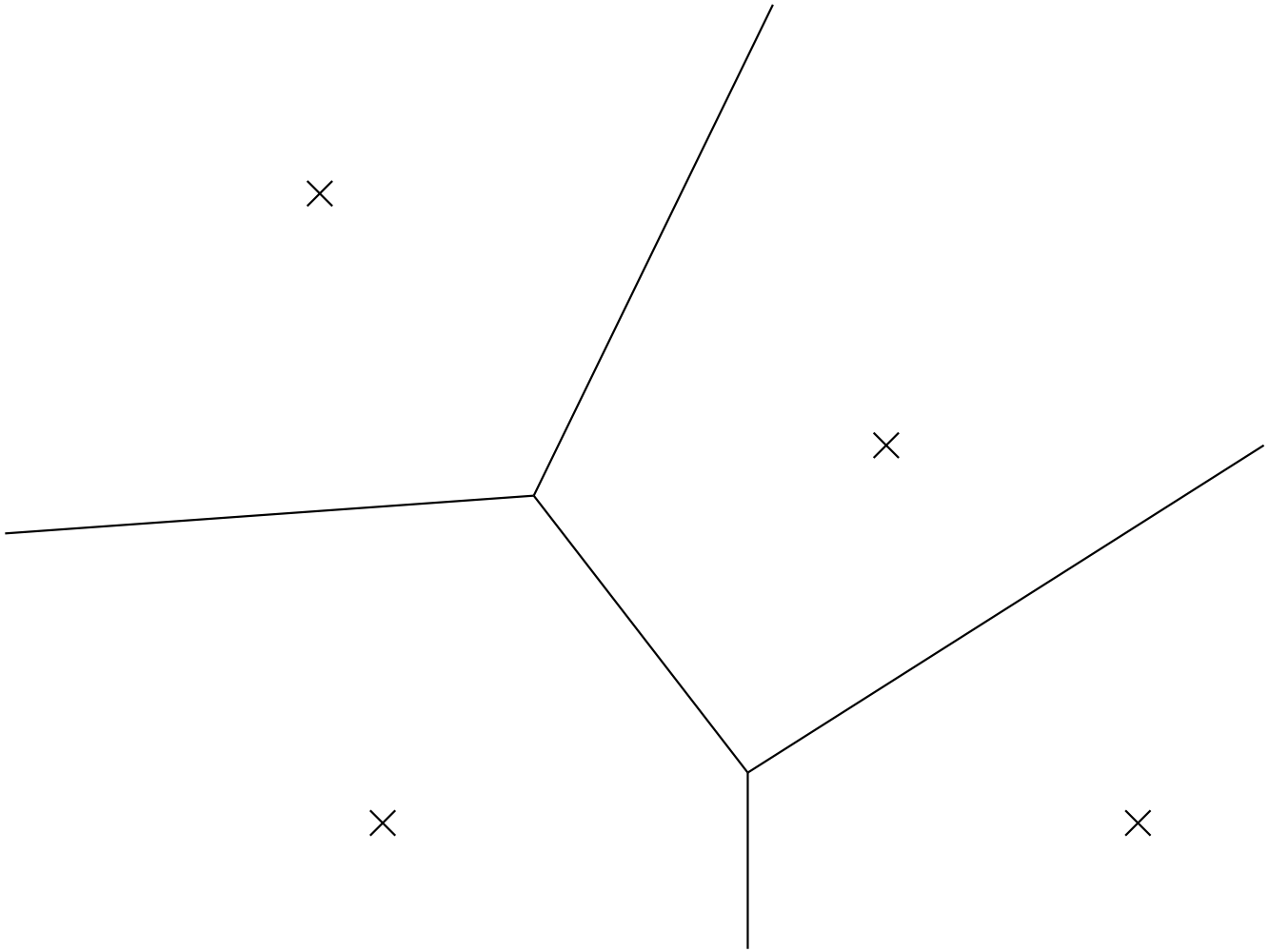
$\Rightarrow$ We define the *Voronoi Quantization* as

$$\widehat{X}^{\Gamma} = \pi_{\Gamma}(X) = \sum_{i=1}^{N} x_i \mathbf{1}_{C_i(\Gamma)}(X).$$

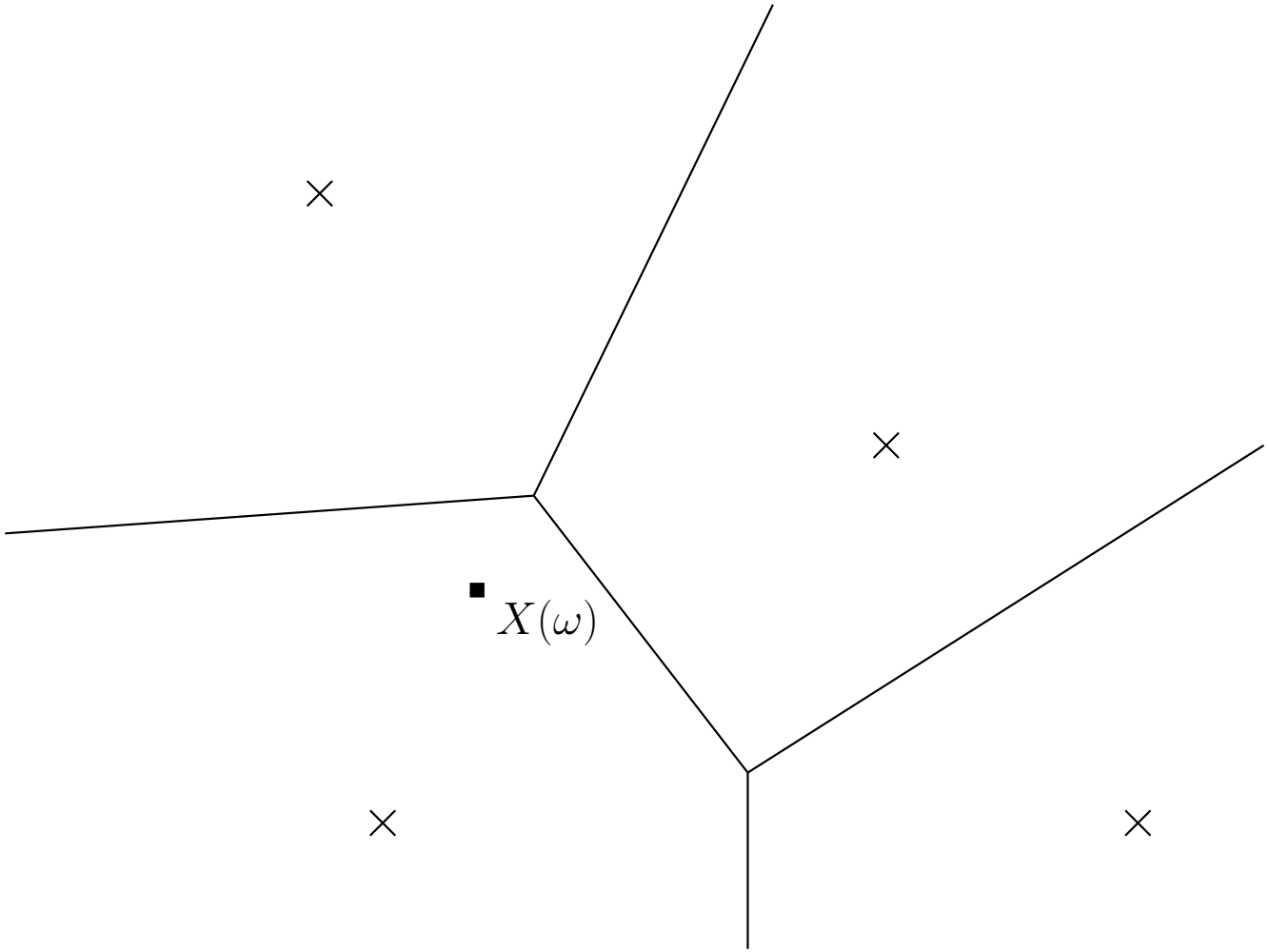
Voronoi-Quantization



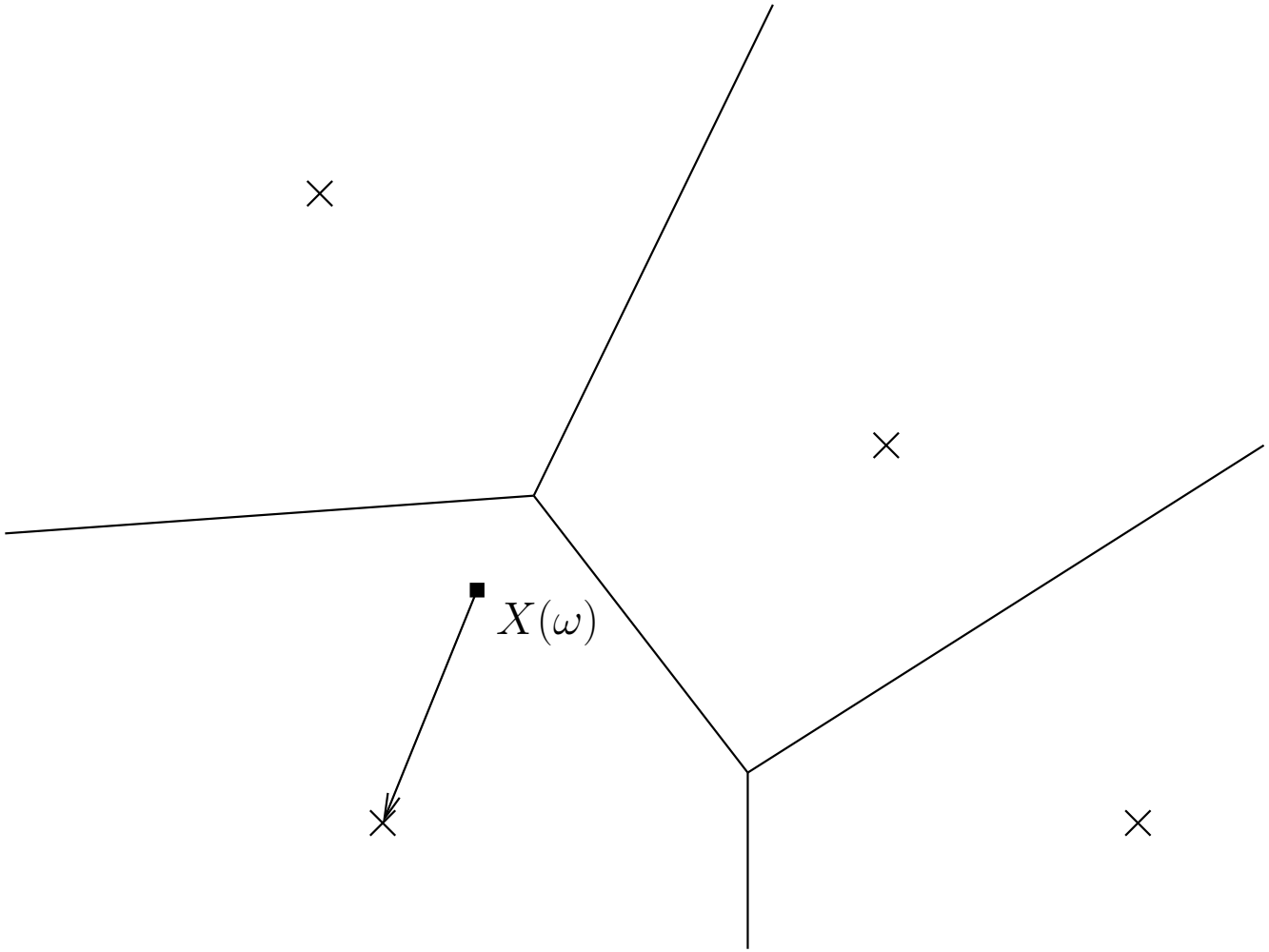
Voronoi-Quantization



Voronoi-Quantization



Voronoi-Quantization



▷ The companion functional approximation operator is

$$\boxed{\mathrm{F}(\widehat{X}^{\Gamma}) = (F \circ \pi_{\Gamma})(X).}$$

It maps F in a *stepwise constant* (on Voronoi partitions...) functions.

▷ The companion functional approximation operator is

$$\boxed{\mathrm{F}(\widehat{X}^{\Gamma}) = (F \circ \pi_{\Gamma})(X).}$$

It maps F in a *stepwise constant* (on Voronoi partitions...) functions.

▷ If F is Lipschitz continuous

$$\left|\mathbb{E}F(X) - \mathbb{E}F(\widehat{X}^{\Gamma})\right| \leq [F]_{\mathrm{Lip}} \left\|X - \widehat{X}^{\Gamma}\right\|_1 = \|\mathrm{dist}(X,\Gamma)\|_1$$

and, since $\xi \mapsto \mathrm{dist}(\xi,\Gamma)$ is 1-Lipschitz, one has

$$\sup_{[F]_{\mathrm{Lip}} \leq 1} \left|\mathbb{E}F(X) - \mathbb{E}F(\widehat{X}^{\Gamma})\right| = \left\|X - \widehat{X}^{\Gamma}\right\|_1 = \left\|\mathrm{dist}(X,\Gamma)\right\|_1.$$

(Wasserstein distance between $\mathcal{L}(X)$ and the set of Γ-supported distributions).

L^p-mean quantization error

▷ The L^p-*mean quantization error* induced by a grid $\Gamma \subset \mathbb{R}^d$ with size $|\Gamma| \leq N$, $N \in \mathbb{N}$

$$e_p(X;\Gamma) = \|\mathrm{dist}(X,\Gamma)\|_p = \left\| \min_{x\in\Gamma} \|X - x\| \right\|_p. \tag{3}$$

L^p-mean quantization error

▷ The L^p-*mean quantization error* induced by a grid $\Gamma \subset \mathbb{R}^d$ with size $|\Gamma| \leq N$, $N \in \mathbb{N}$

$$e_p(X;\Gamma) = \|\mathrm{dist}(X,\Gamma)\|_p = \left\| \min_{x\in\Gamma} \|X - x\| \right\|_p. \tag{3}$$

▷ The optimal L^p-mean quantization problem consists in minimizing (3) over all grids of size $|\Gamma| \leq N$.

We define the L^p-*optimal mean quantization error* of level N as

$$e_{p,N}(X) := \inf\left\{ \left\| \min_{x\in\Gamma} \|X - x\| \right\|_p : \Gamma \subset \mathbb{R}^d,\ |\Gamma| \leq N \right\}.$$

Voronoi-Quantization

One shows the more general optimality result

$$e_{p,N}(X) = \inf\big\{\|X - \Xi\|_p : \Xi \in L^p(\mathbb{R}^d),\ |\Xi(\Omega)| \le N\big\}.$$

Voronoi-Quantization

One shows the more general optimality result

$$e_{p,N}(X) = \inf\big\{\|X - \Xi\|_p : \Xi \in L^p(\mathbb{R}^d),\ |\Xi(\Omega)| \le N\big\}.$$

$\Rightarrow$ Voronoi Quantization $\widehat{X}^\Gamma$ provides an optimal L^p-mean discretization of X (as soon as Γ is an optimal quantization grid for $X\ldots$).

Voronoi-Quantization

One shows the more general optimality result

$$e_{p,N}(X) = \inf\big\{\|X - \Xi\|_p : \Xi \in L^p(\mathbb{R}^d),\ |\Xi(\Omega)| \le N\big\}.$$

$\Rightarrow$ Voronoi Quantization $\widehat{X}^\Gamma$ provides an optimal L^p-mean discretization of X (as soon as Γ is an optimal quantization grid for $X\ldots$).

Voronoi-Quantization

One shows the more general optimality result

$$e_{p,N}(X) = \inf\big\{\|X - \Xi\|_p : \Xi \in L^p(\mathbb{R}^d),\ |\Xi(\Omega)| \le N\big\}.$$

$\Rightarrow$ Voronoi Quantization $\widehat{X}^{\Gamma}$ provides an optimal L^p-mean discretization of X (as soon as Γ is an optimal quantization grid for $X\ldots$).

$\Rightarrow$ The Nearest Neighbor projection is the coding rule, which yields the smallest L^p-mean approximation error for X.

Theorem (Kiefer,..., Cuesta-Albertos, P. (1997))

(a) For every level $N \ge 1$, there exists (at least) an L^p-optimal quantization grid $\Gamma^{N,*}$ at level N.

(b) If $p = 2$, $\mathbb{E}\Big(X \,|\, \widehat{X}^{\Gamma^{N,*}}\Big) = \widehat{X}^{\Gamma^{N,*}}$ (stationarity/self-consistency).

Rates of Optimal Quantization

▷ It is easy to check that (everywhere dense sequence...)

$$e_{p,N}(X) \to 0 \quad \text{as} \quad N \to \infty.$$

At which rate ?

Rates of Optimal Quantization

▷ It is easy to check that (everywhere dense sequence...)

$$e_{p,N}(X) \to 0 \quad \text{as} \quad N \to \infty.$$

At which rate ?

Theorem (Zador's Theorem)

(*a*) Sharp asymptotic *(Zador, Kiefer, Bucklew & Wise, Graf & Luschgy, cf. [GL00]): Let* $X \in L^{p+}(\mathbb{R}^d)$ *with distribution* $\mathbb{P}_X = \varphi.\lambda^d \overset{\perp}{+} \nu$.
Then

$$\lim_{N\to\infty} N^{\frac{1}{d}} \cdot e_{p,N}(X) = Q_{p,\|\cdot\|} \cdot \left(\int_{\mathbb{R}^d} \|\varphi\|^{d/(d+p)}\, d\lambda^d\right)^{(d+p)/d}$$

where $Q_{p,\|\cdot\|} = \inf_N N^{\frac{1}{d}} \cdot e_{p,N}\big(U([0,1]^d)\big)$.

(*b*) Non-asymptotic *(Luschgy-P. (2007), cf. []): Let* $p' > p$. *There exists* $C_{p,p',d}$ *such that, for every r.v.* $\mathbb{R}^d$*-valued* X

$$\forall\, N \geq 1, \quad e_{p,N}(X) \leq C_{p,p',d} \inf_{a\in\mathbb{R}^d} \|X - a\|_{p'} \cdot N^{-\frac{1}{d}}.$$

A Side Note on Numerical Computation of Quantizers

Computing optimal grids

Consider for $D_N : (\mathbb{R}^d)^N \to \mathbb{R}$ the optimization problem (here $p = 2$)

$$D_N(x) := \mathbb{E} \min_{1 \le i \le N} \|X - x_i\|^2 \ \to \min_{x \in (\mathbb{R}^d)^N} .$$

A Side Note on Numerical Computation of Quantizers

Computing optimal grids

Consider for $D_N : (\mathbb{R}^d)^N \to \mathbb{R}$ the optimization problem (here $p = 2$)

$$D_N(x) := \mathbb{E} \min_{1 \le i \le N} \|X - x_i\|^2 \ \to \min_{x \in (\mathbb{R}^d)^N} .$$

As soon as $\|\cdot\|$ is a.s. smooth $\Rightarrow D_N$ is differentiable.

A Side Note on Numerical Computation of Quantizers

Computing optimal grids

Consider for $D_N : (\mathbb{R}^d)^N \to \mathbb{R}$ the optimization problem (here $p = 2$)

$$D_N(x) := \mathbb{E} \min_{1 \le i \le N} \|X - x_i\|^2 \ \to \min_{x \in (\mathbb{R}^d)^N} .$$

As soon as $\|\cdot\|$ is a.s. smooth $\Rightarrow D_N$ is differentiable.

$d = 1$:

$$D_N(x) = \sum_{i=1}^{N} \int_{x_{i-1/2}}^{x_{i+1/2}} |\xi - x_i|^2 d\mathbb{P}^X(\xi)$$

$\Rightarrow$ Evaluation of Voronoi-Cells, Gradient and Hessian is simple $\rightsquigarrow$ Newton-Raphson

A Side Note on Numerical Computation of Quantizers ($p = 2$)

$d \geq 2$: 1. Stochastic Gradient Method: CLVQ

- Simulate $\xi_1, \xi_2, \ldots$ independent copies of X
- Generate step sequence $\gamma_1, \gamma_2, \ldots$
  Usually: step $\gamma_n = \frac{A}{B+n} \searrow 0$ or $\gamma_n = \eta \approx 0$
- Grid updating $n \mapsto n+1$:
  Competition: select winner index: $i^* \in \operatorname{argmin}_i |x_i^n - \xi_n|$
  Learning: $\begin{cases} x_{i^*}^{n+1} := x_{i^*}^n + \gamma_n (x_{i^*}^n - \xi_n) \\ x_j^{n+1} := x_j^n, \qquad \text{for } j \neq i^*. \end{cases}$

A Side Note on Numerical Computation of Quantizers ($p = 2$)

$d \geq 2$:

1. Stochastic Gradient Method: CLVQ
   - Simulate $\xi_1, \xi_2, \ldots$ independent copies of X
   - Generate step sequence $\gamma_1, \gamma_2, \ldots$
     Usually: step $\gamma_n = \frac{A}{B+n} \searrow 0$ or $\gamma_n = \eta \approx 0$
   - Grid updating $n \mapsto n+1$:
     Competition: select winner index: $i^* \in \operatorname{argmin}_i |x_i^n - \xi_n|$
     Learning: $\begin{cases} x_{i^*}^{n+1} := x_{i^*}^n + \gamma_n (x_{i^*}^n - \xi_n) \\ x_j^{n+1} := x_j^n, \qquad \text{for } j \neq i^*. \end{cases}$
2. Lloyd's algorithm as a randomized fix-point method.
   - Initial grid $\Gamma^{(0)} = \{x_1^0, \ldots, x_N^0\}$
   - Usual step : $\widehat{X}^{\Gamma^{(n+1)}} = \mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma^{(n)}}\big)$ i.e. $x_k^{(n+1)} = \mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma^{(n)}} = x_k^{(n)}\big)$
   - so that $\|X - \widehat{X}^{\Gamma^{(n+1)}}\|_2 \leq \|X - \widehat{X}^{\Gamma^{(n)}}\|_2$

A Side Note on Numerical Computation of Quantizers ($p = 2$)

$d \geq 2$:

1. Stochastic Gradient Method: CLVQ
   - Simulate $\xi_1, \xi_2, \ldots$ independent copies of X
   - Generate step sequence $\gamma_1, \gamma_2, \ldots$
     Usually: step $\gamma_n = \frac{A}{B+n} \searrow 0$ or $\gamma_n = \eta \approx 0$
   - Grid updating $n \mapsto n+1$:
     Competition: select winner index: $i^* \in \operatorname{argmin}_i |x_i^n - \xi_n|$
     Learning: $\begin{cases} x_{i^*}^{n+1} := x_{i^*}^n + \gamma_n (x_{i^*}^n - \xi_n) \\ x_j^{n+1} := x_j^n, \qquad \text{for } j \neq i^*. \end{cases}$
2. Lloyd's algorithm as a randomized fix-point method.
   - Initial grid $\Gamma^{(0)} = \{x_1^0, \ldots, x_N^0\}$
   - Usual step : $\widehat{X}^{\Gamma^{(n+1)}} = \mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma^{(n)}}\big)$ i.e. $x_k^{(n+1)} = \mathbb{E}\big(X \,|\, \widehat{X}^{\Gamma^{(n)}} = x_k^{(n)}\big)$
   - so that $\|X - \widehat{X}^{\Gamma^{(n+1)}}\|_2 \leq \|X - \widehat{X}^{\Gamma^{(n)}}\|_2$
3. "Batch" approach [...]

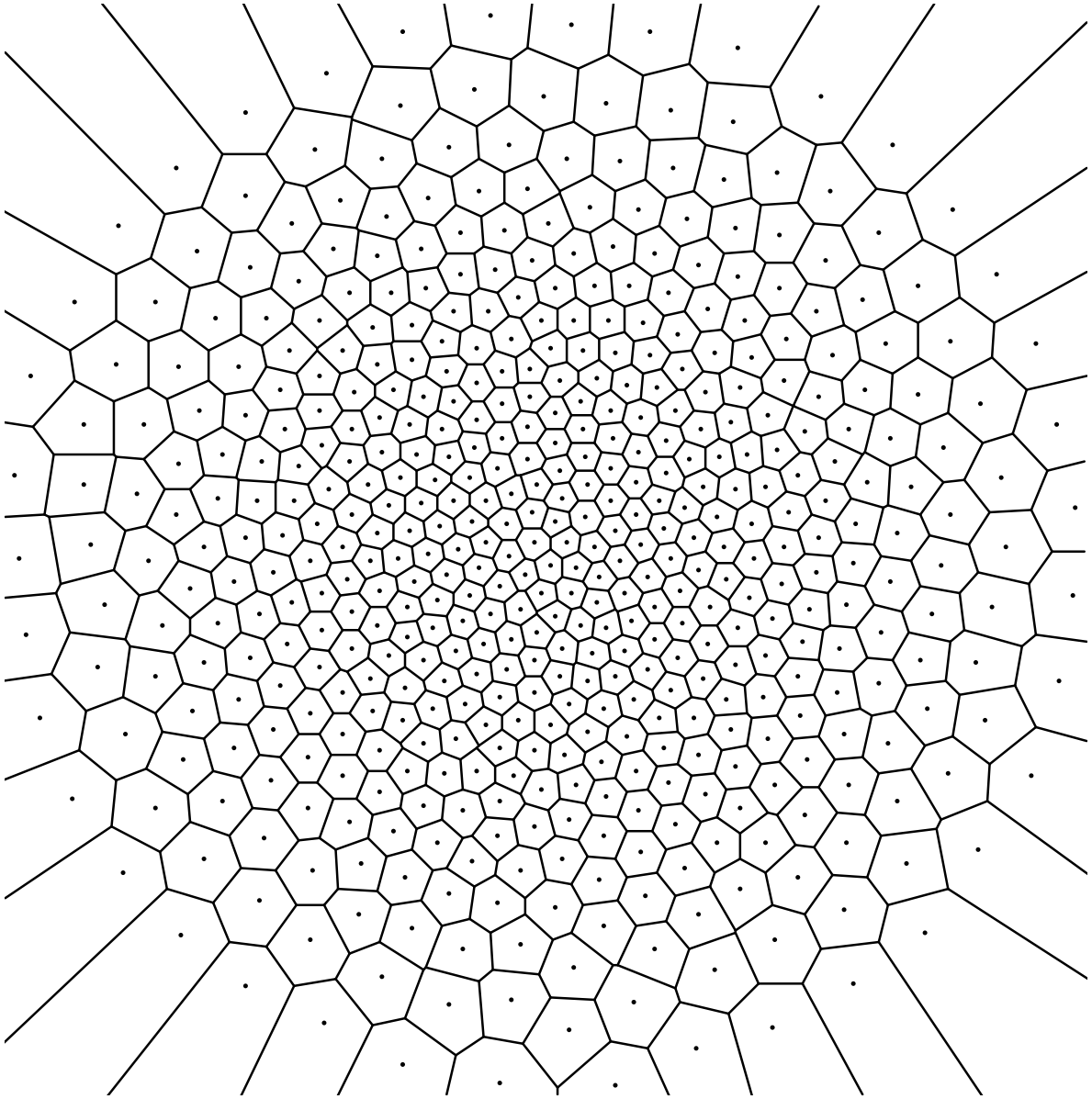


Figure: A Quantizer for $\mathcal{N}(0, I_2)$ of size $N = 500$ in $(R^2, \|\cdot\|_2)$.

Quantization for Cubature

Assume that we have access to the Voronoi-Cell weights

$$w_i(\Gamma) := \mathbb{P}(X \in C_i(\Gamma)),\ i = 1, \ldots, N.$$

Quantization for Cubature

Assume that we have access to the Voronoi-Cell weights

$$w_i(\Gamma) := \mathbb{P}(X \in C_i(\Gamma)),\ i = 1, \ldots, N.$$

$\Longrightarrow$ The computation of $\mathbb{E}F(\widehat{X}^\Gamma)$ for some Lipschitz continuous $F : \mathbb{R}^d \to \mathbb{R}$ becomes straightforward:

$$\mathbb{E}\, F(\widehat{X}^\Gamma) = \sum_{i=1}^{N} w_i(\Gamma) F(x_i).$$

Quantization for Cubature

Assume that we have access to the Voronoi-Cell weights

$$w_i(\Gamma) := \mathbb{P}(X \in C_i(\Gamma)),\ i = 1, \ldots, N.$$

$\Longrightarrow$ The computation of $\mathbb{E}F(\widehat{X}^\Gamma)$ for some Lipschitz continuous $F : \mathbb{R}^d \to \mathbb{R}$ becomes straightforward:

$$\mathbb{E}\, F(\widehat{X}^\Gamma) = \sum_{i=1}^{N} w_i(\Gamma) F(x_i).$$

▷ As a first error estimate, we already know that

$$|\mathbb{E}F(X) - \mathbb{E}\, F(\widehat{X}^\Gamma)| \leq [F]_{\mathrm{Lip}}\, \mathbb{E}\|X - \widehat{X}^\Gamma\|.$$

Further Error Estimates

Moreover

$$\inf\Big\{\sup_{[F]_{\mathrm{Lip}}\le 1}|\mathbb{E}F(X)-\mathbb{E}\,F(Y)|,\ Y(\Omega)\subset\Gamma\Big\}$$
$$=\sup_{[F]_{\mathrm{Lip}}\le 1}|\mathbb{E}F(X)-\mathbb{E}\,F(\widehat{X}^{\Gamma})|=\mathbb{E}\|X-\widehat{X}^{\Gamma}\|$$

i.e. Quantization is optimal for the class of Lipschitz functions.

Further Error Estimates

Moreover

$$\inf\Big\{\sup_{[F]_{\mathrm{Lip}}\le 1}|\mathbb{E}F(X)-\mathbb{E}\,F(Y)|,\ Y(\Omega)\subset\Gamma\Big\}$$
$$=\sup_{[F]_{\mathrm{Lip}}\le 1}|\mathbb{E}F(X)-\mathbb{E}\,F(\widehat{X}^{\Gamma})|=\mathbb{E}\|X-\widehat{X}^{\Gamma}\|$$

i.e. Quantization is optimal for the class of Lipschitz functions.

Second order rate

▷ If $F\in C^1_{\mathrm{Lip}}$ and the grid Γ is *stationary*, *i.e.*

$$\widehat{X}^{\Gamma}=\mathbb{E}(X|\widehat{X}^{\Gamma}),$$

then a Taylor expansion yields

$$\begin{aligned}|\mathbb{E}\,F(X)-\mathbb{E}\,F(\widehat{X}^{\Gamma})| &= |\mathbb{E}\,F(X)-\mathbb{E}\,F(\widehat{X}^{\Gamma})-\mathbb{E}\,DF(\widehat{X}^{\Gamma}).(X-\widehat{X}^{\Gamma})|\\ &\le [DF]_{\mathrm{Lip}}\cdot\mathbb{E}\|X-\widehat{X}^{\Gamma}\|^2.\end{aligned}$$

▷ Furthermore, if F is convex, then Jensen's inequality implies for stationary Γ

$$\mathbb{E}\, F(\widehat{X}^{\Gamma}) \leq \mathbb{E}\, F(X).$$

Further Applications

Applications for optimal quantization grids

Further Applications

Applications for optimal quantization grids

Further Applications

Applications for optimal quantization grids

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's (Bally-P.-Printems '01, '03 et '05, Illand '11)

Further Applications

Applications for optimal quantization grids

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's (Bally-P.-Printems '01, '03 et '05, Illand '11)
- δ-Hedging for American options (ibid. '05)

Further Applications

Applications for optimal quantization grids

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's (Bally-P.-Printems '01, '03 et '05, Illand '11)
- δ-Hedging for American options (ibid. '05)
- Optimal Stochastic Control problems (P.-Pham-Printems 06'), Pricing of Swing options (Bouthemy-Bardou-P.'09)... on massively parallel architecture (GPU, Bronstein-P.-Wilbertz, '10)

Further Applications

Applications for optimal quantization grids

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's (Bally-P.-Printems '01, '03 et '05, Illand '11)
- δ-Hedging for American options (ibid. '05)
- Optimal Stochastic Control problems (P.-Pham-Printems 06'), Pricing of Swing options (Bouthemy-Bardou-P.'09)... on massively parallel architecture (GPU, Bronstein-P.-Wilbertz, '10)
- Non-linear filtering and stochastic volatility models (P.-Pham-Pprintems '05, Pham-Sellami-Runggaldier'06, Sellami '09 &'10, Callegaro-Sagna '10)

Further Applications

Applications for optimal quantization grids

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's (Bally-P.-Printems '01, '03 et '05, Illand '11)
- δ-Hedging for American options (ibid. '05)
- Optimal Stochastic Control problems (P.-Pham-Printems 06'), Pricing of Swing options (Bouthemy-Bardou-P.'09)... on massively parallel architecture (GPU, Bronstein-P.-Wilbertz, '10)
- Non-linear filtering and stochastic volatility models (P.-Pham-Pprintems '05, Pham-Sellami-Runggaldier'06, Sellami '09 &'10, Callegaro-Sagna '10)
- Discretization of SPDE's (stochastic Zakaï & McKean-Vlasov equations) [Gobet-P.-Pham-Printems '07]

Further Applications

Applications for optimal quantization grids

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's (Bally-P.-Printems '01, '03 et '05, Illand '11)
- δ-Hedging for American options (ibid. '05)
- Optimal Stochastic Control problems (P.-Pham-Printems 06'), Pricing of Swing options (Bouthemy-Bardou-P.'09)... on massively parallel architecture (GPU, Bronstein-P.-Wilbertz, '10)
- Non-linear filtering and stochastic volatility models (P.-Pham-Pprintems '05, Pham-Sellami-Runggaldier'06, Sellami '09 &'10, Callegaro-Sagna '10)
- Discretization of SPDE's (stochastic Zakaï & McKean-Vlasov equations) [Gobet-P.-Pham-Printems '07]
- Quantization based Universal Stratification (variance reduction) [Corlay-P. '10]

Further Applications

Applications for optimal quantization grids

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's (Bally-P.-Printems '01, '03 et '05, Illand '11)
- δ-Hedging for American options (ibid. '05)
- Optimal Stochastic Control problems (P.-Pham-Printems 06'), Pricing of Swing options (Bouthemy-Bardou-P.'09)... on massively parallel architecture (GPU, Bronstein-P.-Wilbertz, '10)
- Non-linear filtering and stochastic volatility models (P.-Pham-Pprintems '05, Pham-Sellami-Runggaldier'06, Sellami '09 &'10, Callegaro-Sagna '10)
- Discretization of SPDE's (stochastic Zakaï & McKean-Vlasov equations) [Gobet-P.-Pham-Printems '07]
- Quantization based Universal Stratification (variance reduction) [Corlay-P. '10]
- CVaR-based dynamical risk hedging [Bardou-Frikha-P., '10), etc.

First conclusions on Voronoi quantization

- Voronoi quantization is optimal for "Lipschitz approximation"
- Paradox: it does not preserve regularity
- Second order (stationarity) : (almost) only optimal grids $\Rightarrow$ lack of flexibility
- Download free pre-computed grids of $\mathcal{N}(0; I_d)$ distributions at the URL

  `www.quantize.maths-fi.com`

  for $d = 1, \ldots, 10$ and $N = 1, \ldots 10^4$.
- and many others items related to optimal quantization.

Starting with dual Quantization ($d = 1$ and $d \geq 2$)

P.-Wilbertz '09 JCF and '12 SINUM & *Num. Meth. in Fin.*, Springer

Starting with dual Quantization ($d = 1$ and $d \geq 2$)
P.-Wilbertz '09 JCF and '12 SINUM & *Num. Meth. in Fin.*, Springer

Idea

▷ No longer maps $X(\omega)$ to its nearest neighbor, but splits up the projection randomly between the "surrounding" neighbors of $X(\omega)$.

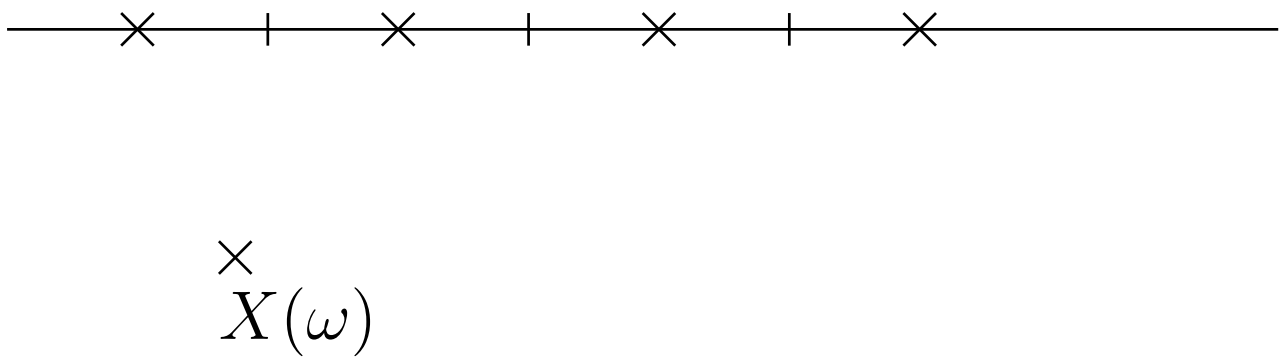
▷ Let "$\Gamma = \{\times, \ldots, \times\}$" in the figure below.

Starting with dual Quantization ($d = 1$ and $d \geq 2$)
P.-Wilbertz '09 JCF and '12 SINUM & *Num. Meth. in Fin.*, Springer

Idea

▷ No longer maps $X(\omega)$ to its nearest neighbor, but splits up the projection randomly between the "surrounding" neighbors of $X(\omega)$.

▷ Let "$\Gamma = \{\times, \ldots, \times\}$" in the figure below.

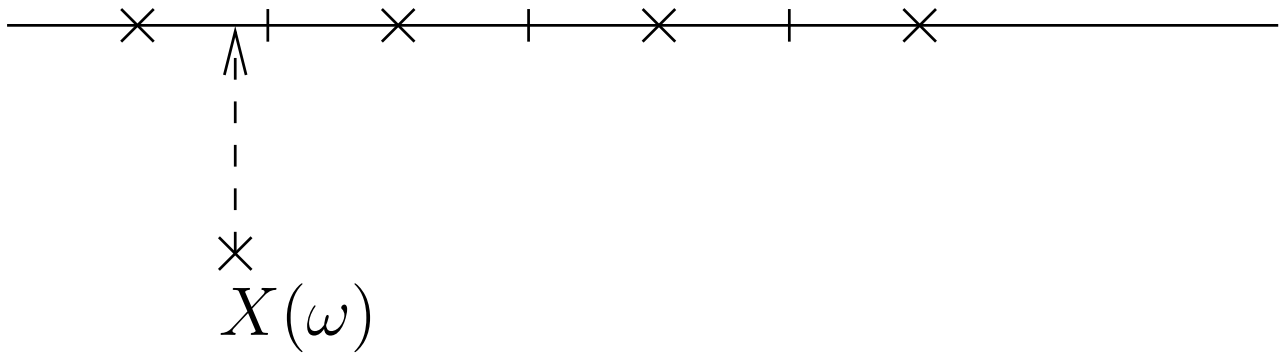


Starting with dual Quantization ($d = 1$ and $d \geq 2$)
P.-Wilbertz '09 JCF and '12 SINUM & *Num. Meth. in Fin.*, Springer

Idea

▷ No longer maps $X(\omega)$ to its nearest neighbor, but splits up the projection randomly between the "surrounding" neighbors of $X(\omega)$.

▷ Let "$\Gamma = \{\times, \ldots, \times\}$" in the figure below.

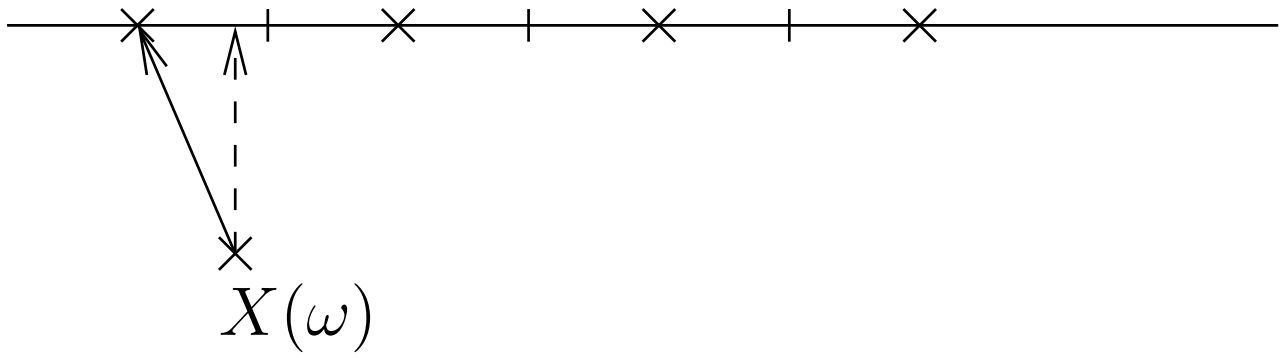


Starting with dual Quantization ($d = 1$ and $d \geq 2$)
P.-Wilbertz '09 JCF and '12 SINUM & *Num. Meth. in Fin.*, Springer

Idea

▷ No longer maps $X(\omega)$ to its nearest neighbor, but splits up the projection randomly between the "surrounding" neighbors of $X(\omega)$.

▷ Let "$\Gamma = \{\times, \ldots, \times\}$" in the figure below.

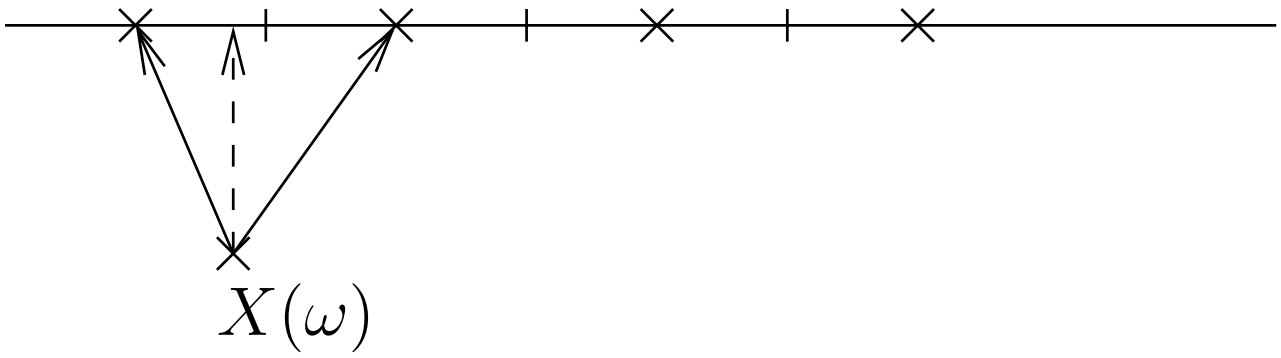


Starting with dual Quantization ($d = 1$ and $d \geq 2$)
P.-Wilbertz '09 JCF and '12 SINUM & *Num. Meth. in Fin.*, Springer

Idea

▷ No longer maps $X(\omega)$ to its nearest neighbor, but splits up the projection randomly between the "surrounding" neighbors of $X(\omega)$.

▷ Let "$\Gamma = \{\times, \ldots, \times\}$" in the figure below.

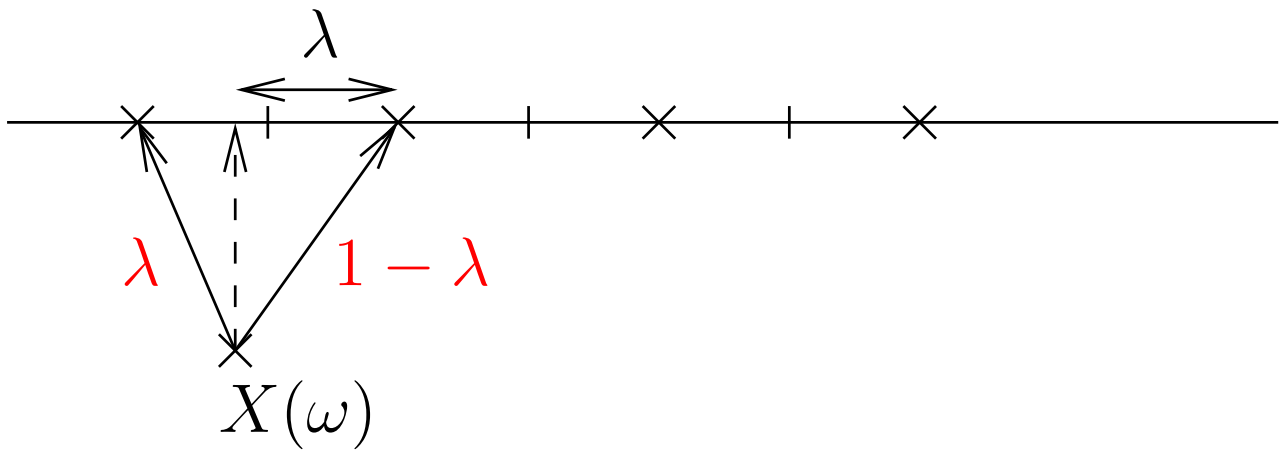


Starting with dual Quantization ($d = 1$ and $d \geq 2$)
P.-Wilbertz '09 JCF and '12 SINUM & *Num. Meth. in Fin.*, Springer

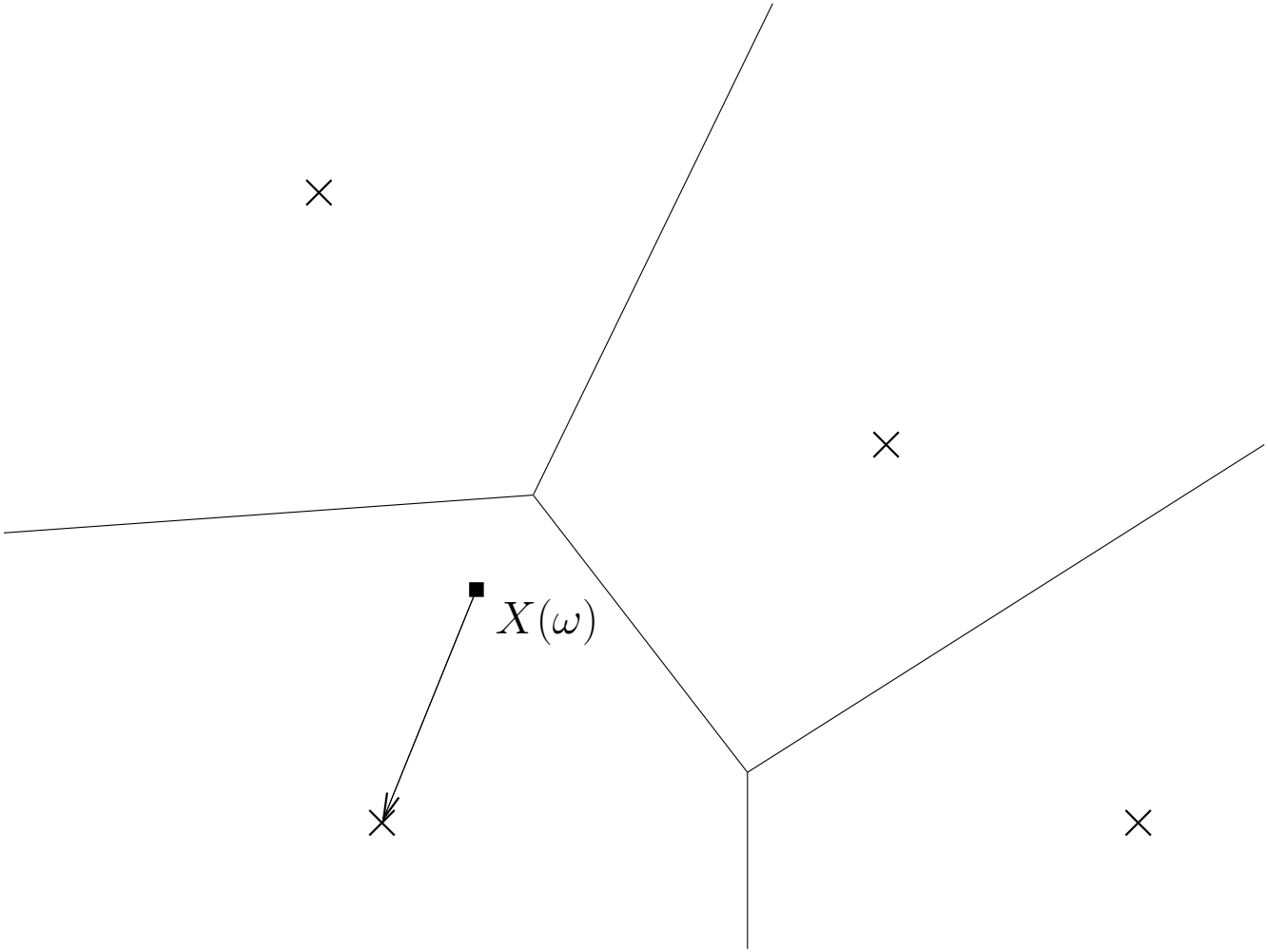
Idea

▷ No longer maps $X(\omega)$ to its nearest neighbor, but splits up the projection randomly between the "surrounding" neighbors of $X(\omega)$.

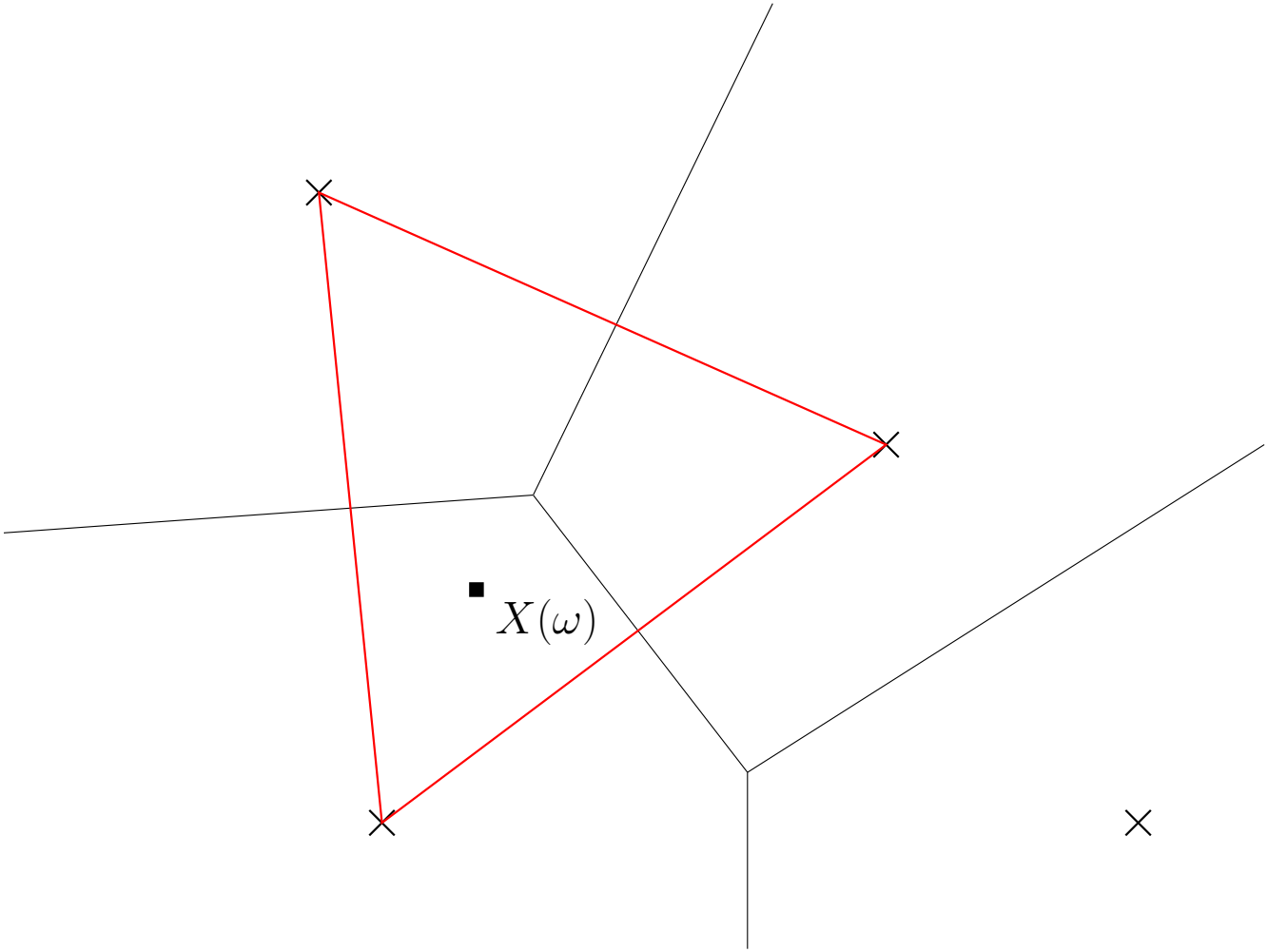
▷ Let "$\Gamma = \{\times, \ldots, \times\}$" in the figure below.



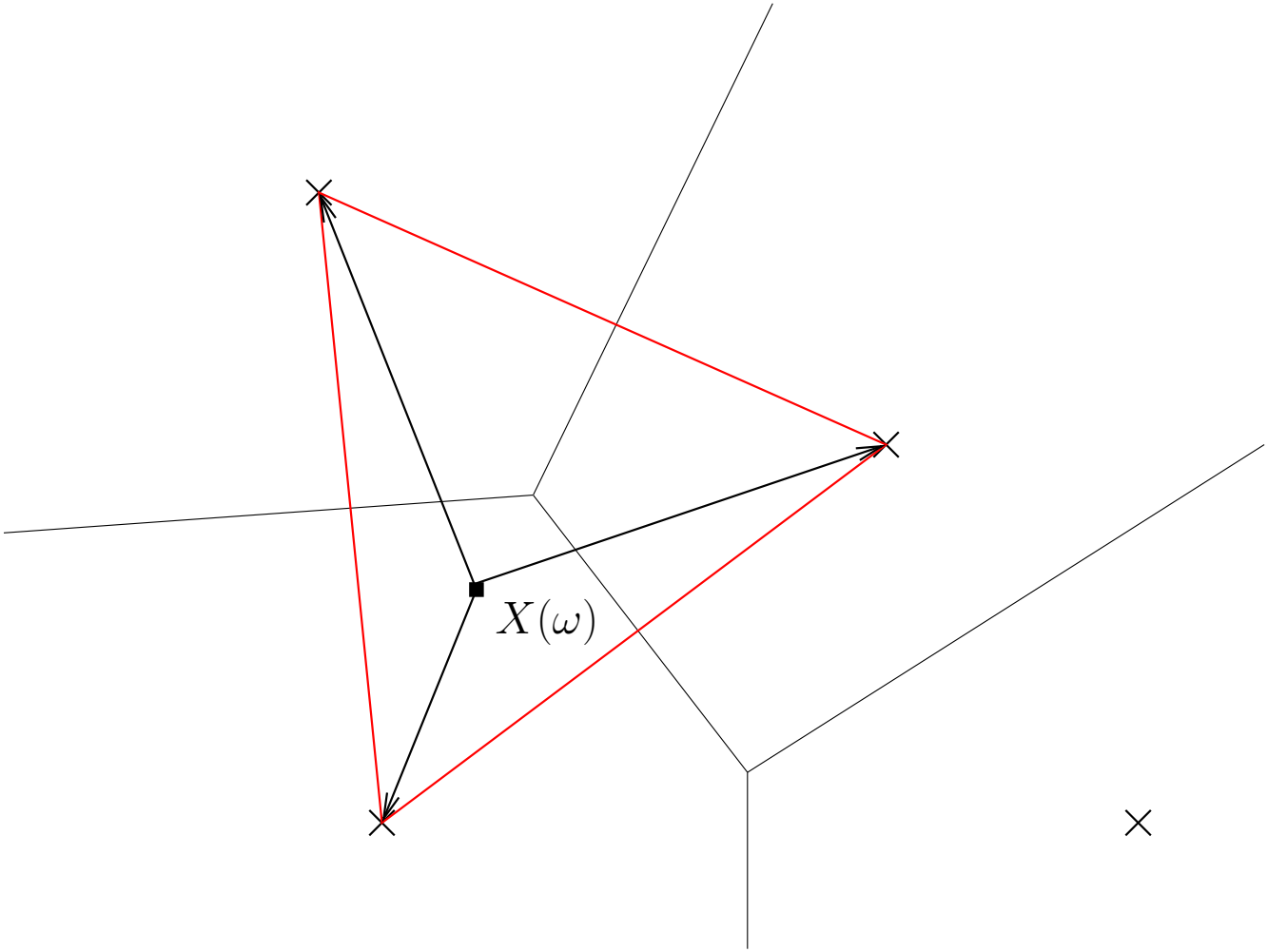
Starting with dual Quantization: $d = 2$



Starting with dual Quantization: $d = 2$



Starting with dual Quantization: $d = 2$



Random splitting operator

▷ Let $\tau = \{x_1, \ldots, x_{d+1}\} \subset \mathbb{R}^d$ be a d-simplex in $\mathbb{R}^d$,

i.e. $x_1, \ldots, x_{d+1}$ are *affinely independent.*

Random splitting operator

▷ Let $\tau = \{x_1, \ldots, x_{d+1}\} \subset \mathbb{R}^d$ be a d-simplex in $\mathbb{R}^d$,

i.e. $x_1, \ldots, x_{d+1}$ are *affinely independent.*

▷ Let $\lambda(\xi)$ be the barycentric coordinates of $\xi \in \mathrm{conv}(\tau)$.

Definiton of the τ-splitting operator

$$\begin{array}{rcl} \mathcal{J}_\tau^U : \mathrm{conv}(\tau) & \longrightarrow & \tau \\ \xi & \longmapsto & \displaystyle\sum_{i=1}^{d+1} x_i \mathbf{1}_{\left\{\sum_{j=1}^{i-1} \lambda_j(\xi) \le U < \sum_{j=1}^{i} \lambda_j(\xi)\right\}} \end{array}$$

where $U \sim \mathcal{U}([0,1])$ is defined on an exogeneous space $(\Omega_0, \mathcal{S}_0, \mathbb{P}_0)$.

Random splitting operator

▷ Let $\tau = \{x_1, \ldots, x_{d+1}\} \subset \mathbb{R}^d$ be a d-simplex in $\mathbb{R}^d$,

i.e. $x_1, \ldots, x_{d+1}$ are *affinely independent.*

▷ Let $\lambda(\xi)$ be the barycentric coordinates of $\xi \in \mathrm{conv}(\tau)$.

Definiton of the τ-splitting operator

$$\begin{array}{rcl} \mathcal{J}_\tau^U : \mathrm{conv}(\tau) & \longrightarrow & \tau \\ \xi & \longmapsto & \displaystyle\sum_{i=1}^{d+1} x_i \mathbf{1}_{\left\{\sum_{j=1}^{i-1} \lambda_j(\xi) \le U < \sum_{j=1}^{i} \lambda_j(\xi)\right\}} \end{array}$$

where $U \sim \mathcal{U}([0,1])$ is defined on an exogeneous space $(\Omega_0, \mathcal{S}_0, \mathbb{P}_0)$.

▷ This τ-splitting operator always satisfies a *mean preserving* property:

$$\mathbb{E}_0\big(\mathcal{J}_\tau^U(\xi)\big) = \sum_{i=1}^{d+1} \lambda_i(\xi) \cdot x_i = \xi, \qquad \forall \xi \in \mathrm{conv}(\tau). \tag{4}$$

Functional approximation operator

▷ The τ-splitting operator is in fact a probabilistic representation of the classical *interpolation operator*

$$\mathbb{J}_\tau(F) \equiv \xi \longmapsto \mathbb{E}_0\big(F(\mathcal{J}_\tau^U(\xi))\big) = \sum_{i=1}^{d+1} \lambda_i(\xi) \cdot F(x_i), \qquad \forall \xi \in \mathrm{conv}(\tau). \tag{5}$$

P1. $\mathbb{J}_\tau(F)$ is affine on $\mathrm{conv}(\tau)$.

P2. If F is convex $\mathbb{J}_\tau(F) \geq F$.

(Not so) naive extension to triangulations: Cubature I

▷ The notion of τ-splitting operator can be extended to any given *triangulation* $\mathcal{T}_\Gamma$ of a grid $\Gamma = \{x_1, \ldots, x_N\}$, so that (4) and (5) hold for any $\xi \in \mathrm{conv}(\Gamma)$ for $\mathcal{J}_{\mathcal{T}_\Gamma}$ and $\mathbb{J}_{\mathcal{T}_\Gamma}$.

Such an operator $\mathbb{J}_{\mathcal{T}_\Gamma}$ also satisfies

P'1. $\mathbb{J}_{\mathcal{T}_\Gamma}(F)$ is continuous, piecewise affine on $\mathrm{conv}(\Gamma)$.

P'2. If F is convex $\mathbb{J}_{\mathcal{T}_\Gamma}$ is convex on $\mathrm{conv}(\Gamma)$ and $\mathbb{J}_\tau(F) \geq F$.

P3. Random splitting operators preserve the *convex order* on distributions, namely

$$\Big(\forall\, F : \mathrm{conv}(\Gamma) \overset{convex}{\longrightarrow} \mathbb{R},\ \mathbb{E}F(X) \leq \mathbb{E}F(Y)\Big)$$
$$\Longrightarrow \quad \Big(\forall\, F : \mathrm{conv}(\Gamma) \overset{convex}{\longrightarrow} \mathbb{R},\ \mathbb{E}F\big(\mathcal{J}_{\mathcal{T}_\Gamma}(X)\big) \leq \mathbb{E}F\big(\mathbb{J}_{\mathcal{T}_\Gamma}(Y)\big)\Big)$$

▷ Induced cubature formulas. Let $F : \mathbb{R}^d \to \mathbb{R}$.

$$\begin{aligned}
\mathbb{E}\, F(\mathcal{J}_{\mathcal{T}_\Gamma}(X)) &= \mathbb{E}\left(\mathbb{J}_{\mathcal{T}_\Gamma}(F)(X)\right) \\
&= \sum_{\tau \in \mathcal{T}_\Gamma} \sum_{a \in \tau} \mathbb{E}(\lambda_a(X)) F(a) = \sum_{a \in \Gamma} \Big(\sum_{\tau \in \mathcal{T}_\Gamma, a \in \tau} \mathbb{E}(\lambda_a(X)) \Big) F(a) \\
&= \sum_{a \in \Gamma} w_a F(a).
\end{aligned}$$

Intrinsic (dual) stationarity

Motivated by this observation...

▷ DEFINITION. Let Γ be a grid of $\mathbb{R}^d$. An application

$$\mathcal{J}_\Gamma : \Omega_0 \times \mathbb{R}^d \to \Gamma$$

is *intrinsic stationary* if

$$\forall \xi \in \mathrm{conv}(\Gamma), \qquad \mathbb{E}_0\big(\mathcal{J}_\Gamma(\xi)\big) = \xi.$$

Intrinsic (dual) stationarity

Motivated by this observation...

▷ DEFINITION. Let Γ be a grid of $\mathbb{R}^d$. An application

$$\mathcal{J}_\Gamma : \Omega_0 \times \mathbb{R}^d \to \Gamma$$

is *intrinsic stationary* if

$$\forall \xi \in \mathrm{conv}(\Gamma), \qquad \mathbb{E}_0\big(\mathcal{J}_\Gamma(\xi)\big) = \xi.$$

▷ The following proposition is an easy consequence of Fubini's Theorem

Proposition

$\mathcal{J}_\Gamma$ is intrinsic stationary, if and only if it satisfies the dual stationarity condition

$$\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(\mathcal{J}_\Gamma(X)|X\big) = X$$

for any r.v. $X : (\Omega, \mathcal{S}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}^d)$ *with* $\mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}(\Gamma)$.

Intrinsic (dual) stationarity

Motivated by this observation...

▷ DEFINITION. Let Γ be a grid of $\mathbb{R}^d$. An application

$$\mathcal{J}_\Gamma : \Omega_0 \times \mathbb{R}^d \to \Gamma$$

is *intrinsic stationary* if

$$\forall \xi \in \mathrm{conv}(\Gamma), \qquad \mathbb{E}_0\big(\mathcal{J}_\Gamma(\xi)\big) = \xi.$$

▷ The following proposition is an easy consequence of Fubini's Theorem

Proposition

$\mathcal{J}_\Gamma$ is intrinsic stationary, if and only if it satisfies the dual stationarity condition

$$\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(\mathcal{J}_\Gamma(X)|X\big) = X$$

for any r.v. $X : (\Omega, \mathcal{S}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}^d)$ *with* $\mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}(\Gamma)$.

▷ QUESTION. Are all intrinsic stationary operators random splitting operators defined on a triangulation? (YES...)

Intrinsic (dual) stationarity

Motivated by this observation...

▷ DEFINITION. Let Γ be a grid of $\mathbb{R}^d$. An application

$$\mathcal{J}_\Gamma : \Omega_0 \times \mathbb{R}^d \to \Gamma$$

is *intrinsic stationary* if

$$\forall \xi \in \mathrm{conv}(\Gamma), \qquad \mathbb{E}_0\big(\mathcal{J}_\Gamma(\xi)\big) = \xi.$$

▷ The following proposition is an easy consequence of Fubini's Theorem

Proposition

$\mathcal{J}_\Gamma$ is intrinsic stationary, if and only if it satisfies the dual stationarity condition

$$\mathbb{E}_{\mathbb{P}\otimes\mathbb{P}_0}\big(\mathcal{J}_\Gamma(X)|X\big) = X$$

for any r.v. $X : (\Omega, \mathcal{S}, \mathbb{P}) \to (\mathbb{R}^d, \mathcal{B}^d)$ with $\mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}(\Gamma)$.

▷ QUESTION. Are all intrinsic stationary operators random splitting operators defined on a triangulation? (YES...)

NEW! This (dual) kind of stationarity is very robust, since *it holds by construction for any r.v. X with support in Γ.*

Dual stationarity II

▷ Like with "regular" quantization, intrinsic stationary operators also yields an improved second order bound.

Dual stationarity II

▷ Like with "regular" quantization, intrinsic stationary operators also yields an improved second order bound.

Proposition

(a) Let $F \in C^1_{Lip}(\Gamma)$, $\Gamma \subset \mathbb{R}^d$ *and* $\mathcal{J}_\Gamma$ *be intrinsic stationary. Then it holds for any r.v.* $X \in L^2(\mathbb{P})$ *with* $\mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}(\Gamma)$,

$$\begin{aligned} |\mathbb{E}\, F(X) - \mathbb{E}\, F(\mathcal{J}_\Gamma(X))| &= |\mathbb{E}\, F(\mathcal{J}_\Gamma(X)) - \mathbb{E}\, F(X) - \mathbb{E}\big(DF(X)(\mathcal{J}_\Gamma(X) - X)\big)| \\ &\leq [F']_{Lip} \cdot \mathbb{E}\|X - \mathcal{J}_\Gamma(X)\|^2. \end{aligned}$$

Dual stationarity II

▷ Like with "regular" quantization, intrinsic stationary operators also yields an improved second order bound.

Proposition

(a) Let $F \in C^1_{Lip}(\Gamma)$, $\Gamma \subset \mathbb{R}^d$ *and* $\mathcal{J}_\Gamma$ *be intrinsic stationary. Then it holds for any r.v.* $X \in L^2(\mathbb{P})$ *with* $\mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}(\Gamma)$,

$$\begin{aligned} |\mathbb{E}\, F(X) - \mathbb{E}\, F(\mathcal{J}_\Gamma(X))| &= |\mathbb{E}\, F(\mathcal{J}_\Gamma(X)) - \mathbb{E}\, F(X) - \mathbb{E}\big(DF(X)(\mathcal{J}_\Gamma(X) - X)\big)| \\ &\leq [F']_{Lip} \cdot \mathbb{E}\|X - \mathcal{J}_\Gamma(X)\|^2. \end{aligned}$$

(b) If $F : \Gamma \to \mathbb{R}^d$ *is convex, then Jensen's inequality implies*

$$\mathbb{E}\, F(\mathcal{J}_\Gamma(X)) \geq \mathbb{E}\, F(X)$$

Dual stationarity II

▷ Like with "regular" quantization, intrinsic stationary operators also yields an improved second order bound.

Proposition

(a) Let $F \in C^1_{Lip}(\Gamma)$, $\Gamma \subset \mathbb{R}^d$ *and* $\mathcal{J}_\Gamma$ *be intrinsic stationary. Then it holds for any r.v.* $X \in L^2(\mathbb{P})$ *with* $\mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}(\Gamma)$,

$$\begin{aligned} |\mathbb{E}\, F(X) - \mathbb{E}\, F(\mathcal{J}_\Gamma(X))| &= |\mathbb{E}\, F(\mathcal{J}_\Gamma(X)) - \mathbb{E}\, F(X) - \mathbb{E}\big(DF(X)(\mathcal{J}_\Gamma(X) - X)\big)| \\ &\leq [F']_{Lip} \cdot \mathbb{E}\|X - \mathcal{J}_\Gamma(X)\|^2. \end{aligned}$$

(b) If $F : \Gamma \to \mathbb{R}^d$ *is convex, then Jensen's inequality implies*

$$\mathbb{E}\, F(\mathcal{J}_\Gamma(X)) \geq \mathbb{E}\, F(X)$$

(c) $\mathcal{J}_\Gamma$ *preserves convex order on random vectors.*

This property also follows from the *dual stationarity*.

Remark. If $\mathcal{J}_\Gamma$ is a random splitting operator, it follows from the stronger fact that

$$F(\mathcal{J}_\Gamma(X)) = \mathbb{J}_\Gamma(F)(X) \geq F(X).$$

Question

Let Γ be a grid of size $N \in \mathbb{N}$. What is the best approximation, which can be achieved by an intrinsic stationary operator $\mathcal{J}_\Gamma$?

Question

Let Γ be a grid of size $N \in \mathbb{N}$. What is the best approximation, which can be achieved by an intrinsic stationary operator $\mathcal{J}_\Gamma$?

PROBLEM: The grid Γ gives raise to (finitely) many possible triangulations.

Question

Let Γ be a grid of size $N \in \mathbb{N}$. What is the best approximation, which can be achieved by an intrinsic stationary operator $\mathcal{J}_\Gamma$?

PROBLEM: The grid Γ gives raise to (finitely) many possible triangulations.

▷ We aim at selecting the triangulation with the lowest p-inertia *i.e.*

$$\forall\, \xi \in \mathrm{conv}(\Gamma), \qquad F_p^p(\xi;\Gamma) = \min_{\substack{\lambda \in \mathbb{R}^N \\ \text{s.t. } \begin{bmatrix} x_1 & \dots & x_N \\ 1 & \dots & 1 \end{bmatrix}\lambda = \begin{bmatrix} \xi \\ 1 \end{bmatrix},\ \lambda \geq 0}} \left(\sum_{i=1}^N \lambda_i \left\| \xi - x_i \right\|^p \right)^{\frac{1}{p}}$$

Question

Let Γ be a grid of size $N \in \mathbb{N}$. What is the best approximation, which can be achieved by an intrinsic stationary operator $\mathcal{J}_\Gamma$?

PROBLEM: The grid Γ gives raise to (finitely) many possible triangulations.

▷ We aim at selecting the triangulation with the lowest p-inertia *i.e.*

$$\forall\, \xi \in \mathrm{conv}(\Gamma), \qquad F_p^p(\xi;\Gamma) = \min_{\substack{\lambda \in \mathbb{R}^N \\ \text{s.t. } \begin{bmatrix} x_1 & \dots & x_N \\ 1 & \dots & 1 \end{bmatrix}\lambda = \begin{bmatrix} \xi \\ 1 \end{bmatrix},\ \lambda \geq 0}} \left(\sum_{i=1}^N \lambda_i \left\| \xi - x_i \right\|^p \right)^{\frac{1}{p}}$$

Hence, for every $\xi \in \mathrm{conv}(\Gamma)$ we choose the best "triangle" in Γ which contains ξ.

Question

Let Γ be a grid of size $N \in \mathbb{N}$. What is the best approximation, which can be achieved by an intrinsic stationary operator $\mathcal{J}_\Gamma$?

PROBLEM: The grid Γ gives raise to (finitely) many possible triangulations.

▷ We aim at selecting the triangulation with the lowest p-inertia *i.e.*

$$\forall\, \xi \in \mathrm{conv}(\Gamma), \qquad F_p^p(\xi;\Gamma) = \min_{\substack{\lambda \in \mathbb{R}^N \\ \text{s.t. } \begin{bmatrix} x_1 & \dots & x_N \\ 1 & \dots & 1 \end{bmatrix}\lambda = \begin{bmatrix} \xi \\ 1 \end{bmatrix},\ \lambda \geq 0}} \left(\sum_{i=1}^N \lambda_i \, \|\xi - x_i\|^p \right)^{\frac{1}{p}}$$

Hence, for every $\xi \in \mathrm{conv}(\Gamma)$ we choose the best "triangle" in Γ which contains ξ.

▷ The optimal L^p dual quantization error is then defined as

$$d_{p,N}(X) = \inf\big\{ \|F_p(X;\Gamma)\|_p,\ \Gamma \subset \mathbb{R}^d, |\Gamma| \leq N \big\}.$$

Optimality regions for $F^p(\xi;\Gamma)$

▷ To design a splitting operator matching $F^p(\xi;\Gamma)$, we need to determine optimality regions, counterparts of the Voronoi regions for regular quantization.

Optimality regions for $F^p(\xi;\Gamma)$

▷ To design a splitting operator matching $F^p(\xi;\Gamma)$, we need to determine optimality regions, counterparts of the Voronoi regions for regular quantization.

▷ $(\lambda_i)_{1\le i\le N} \mapsto \min_{\substack{\lambda\in\mathbb{R}^N \\ \text{s.t. } \begin{bmatrix} x_1 & \dots & x_N \\ 1 & \dots & 1 \end{bmatrix}\lambda=\begin{bmatrix} \xi \\ 1 \end{bmatrix},\, \lambda\ge 0}} \left(\sum_{i=1}^N \lambda_i \|\xi - x_i\|^p\right)^{\frac{1}{p}}$ atteins a minimum (at least) at an extremal N-tuple $\lambda^*(\xi)$ of the convex constraint set.

Optimality regions for $F^p(\xi;\Gamma)$

▷ To design a splitting operator matching $F^p(\xi;\Gamma)$, we need to determine optimality regions, counterparts of the Voronoi regions for regular quantization.

▷ $(\lambda_i)_{1\le i\le N} \mapsto \min_{\substack{\lambda\in\mathbb{R}^N \\ \text{s.t. } \begin{bmatrix} x_1 & \dots & x_N \\ 1 & \dots & 1 \end{bmatrix}\lambda=\begin{bmatrix}\xi\\1\end{bmatrix},\ \lambda\ge 0}} \left(\sum_{i=1}^N \lambda_i \|\xi - x_i\|^p\right)^{\frac{1}{p}}$ atteins a minimum (at least) at an extremal N-tuple $\lambda^*(\xi)$ of the convex constraint set.

▷ Therefore, $I^*(\xi) := \{i : \lambda_i^*(\xi) > 0\}$ defines an affinely independent family $(x_i)_{i\in I^*(\xi)}$ which can be completed into a Γ-valued affine basis.

Optimality regions for $F^p(\xi;\Gamma)$

▷ To design a splitting operator matching $F^p(\xi;\Gamma)$, we need to determine optimality regions, counterparts of the Voronoi regions for regular quantization.

▷ $(\lambda_i)_{1\le i\le N} \mapsto \min_{\substack{\lambda\in\mathbb{R}^N \\ \text{s.t. } \begin{bmatrix} x_1 & \dots & x_N \\ 1 & \dots & 1 \end{bmatrix}\lambda=\begin{bmatrix}\xi\\1\end{bmatrix},\ \lambda\ge 0}} \left(\sum_{i=1}^N \lambda_i \|\xi - x_i\|^p\right)^{\frac{1}{p}}$ atteins a minimum (at least) at an extremal N-tuple $\lambda^*(\xi)$ of the convex constraint set.

▷ Therefore, $I^*(\xi) := \{i : \lambda_i^*(\xi) > 0\}$ defines an affinely independent family $(x_i)_{i\in I^*(\xi)}$ which can be completed into a Γ-valued affine basis.

Let

$$I \in \mathcal{I}(\Gamma) = \big\{J \subset \{1,\dots,N\} : |J| = d+1,\ \mathrm{rk}(A_J) = d+1\big\}.$$

Set

$$D_I(\Gamma) = \big\{\xi \in \mathbb{R}^d : \exists I^*(\xi) \subset I\big\},$$

or equivalently, in term of linear programming,

$$D_I(\Gamma) = \left\{ \xi \in \mathbb{R}^d : \lambda^I = A_I^{-1} \left[\begin{smallmatrix} \xi \\ 1 \end{smallmatrix} \right] \geq 0 \text{ and } \sum_{i \in I} \lambda_i^I \|\xi - x_i\|^p = F^p(\xi; \Gamma) \right\},$$

where A_I denotes the submatrix of $\left[\begin{smallmatrix} x_1 & \cdots & x_N \\ 1 & \dots & 1 \end{smallmatrix} \right]$ whose columns are given by I and

Quadratic Euclidean case

Quadratic Euclidean case

▷ In the case $\|\cdot\| = |\cdot|_2$ and $p = 2$,

optimality regions are the Delaunay "triangles" in Γ,

i.e. the spheres spanned by such optimal d-simplex contain no further point in its interior.

Quadratic Euclidean case

▷ In the case $\|\cdot\| = |\cdot|_2$ and $p = 2$,

optimality regions are the Delaunay "triangles" in Γ,

i.e. the spheres spanned by such optimal d-simplex contain no further point in its interior.

▷ The following theorem is an extention of Rajan's Theorem ([Raj91]).

Theorem (Euclidean case (Rajan '91))

Let $\|\cdot\| = |\cdot|_2$, $p = 2$, *and* $\Gamma = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ *with* $\mathrm{aff.dim}(\Gamma) = d$.
(a) *If* $I \in \mathcal{I}(\Gamma)$ *defines a Delaunay triangle (or d-simplex), then* $\lambda^I = A_I^{-1} \left[\begin{smallmatrix} \xi \\ 1 \end{smallmatrix}\right]$ *provides a solution to* $F^p(\xi; \Gamma)$ *for every* $\xi \in \mathrm{conv}\{x_j : j \in I\}$ *i.e.*

$$D_I(\Gamma) = \mathrm{conv}\{x_j : j \in I\}.$$

Quadratic Euclidean case

▷ In the case $\|\cdot\| = |\cdot|_2$ and $p = 2$,

optimality regions are the Delaunay "triangles" in Γ,

i.e. the spheres spanned by such optimal d-simplex contain no further point in its interior.

▷ The following theorem is an extention of Rajan's Theorem ([Raj91]).

Theorem (Euclidean case (Rajan '91))

Let $\|\cdot\| = |\cdot|_2$, $p = 2$, *and* $\Gamma = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$ *with* $\mathrm{aff.dim}(\Gamma) = d$.
(a) If $I \in \mathcal{I}(\Gamma)$ *defines a Delaunay triangle (or d-simplex), then* $\lambda^I = A_I^{-1} \left[\begin{smallmatrix}\xi\\1\end{smallmatrix}\right]$ *provides a solution to* $F^p(\xi;\Gamma)$ *for every* $\xi \in \mathrm{conv}\{x_j : j \in I\}$ *i.e.*

$$D_I(\Gamma) = \mathrm{conv}\{x_j : j \in I\}.$$

(b) Conversly, if $I \in \mathcal{I}(\Gamma)$ *satisfies* $\mathring{D}_I(\Gamma) \neq \emptyset$, *then the triangle (or d-simplex) defined by* I *has the Delaunay property for* Γ.

Optimal dual quantization operator

For a $\Gamma = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$ with $\mathrm{aff.\,dim}(\Gamma) = d$,

Optimal dual quantization operator

For a $\Gamma = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$ with $\mathrm{aff.\,dim}(\Gamma) = d$,

- choose a Borel partition $(C_I(\Gamma))_{I \in \mathcal{I}(\Gamma)}$ of $\mathrm{conv}(\Gamma)$ such that

$$C_I(\Gamma) \subset D_I(\Gamma),$$

Optimal dual quantization operator

For a $\Gamma = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$ with $\mathrm{aff.\,dim}(\Gamma) = d$,

- choose a Borel partition $(C_I(\Gamma))_{I \in \mathcal{I}(\Gamma)}$ of $\mathrm{conv}(\Gamma)$ such that

$$C_I(\Gamma) \subset D_I(\Gamma),$$

- let $U \sim \mathcal{U}[0,1]$ on $(\Omega_0, \mathcal{S}_0, \mathbb{P}_0)$.

Optimal dual quantization operator

For a $\Gamma = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$ with $\mathrm{aff.dim}(\Gamma) = d$,

- choose a Borel partition $(C_I(\Gamma))_{I \in \mathcal{I}(\Gamma)}$ of $\mathrm{conv}(\Gamma)$ such that

$$C_I(\Gamma) \subset D_I(\Gamma),$$

- let $U \sim \mathcal{U}[0,1]$ on $(\Omega_0, \mathcal{S}_0, \mathbb{P}_0)$.

The *optimal dual quantization operator* $\mathcal{J}_\Gamma^*$ is defined as

$$\mathcal{J}_\Gamma^*(\xi) = \sum_{I \in \mathcal{I}(\Gamma)} \left[\sum_{\ell=1}^{k_I} x_{i_\ell} \cdot \mathbf{1}_{\left\{ \sum_{j=1}^{\ell-1} \lambda_{i_j}^I(\xi) \le U < \sum_{j=1}^{\ell} \lambda_{i_j}^I(\xi) \right\}} \right] \mathbf{1}_{C_I(\Gamma)}(\xi).$$

where $I = \{i_1, \ldots, i_{k_I}\}$.

Optimal dual quantization operator

For a $\Gamma = \{x_1, \ldots, x_N\} \subset \mathbb{R}^d$ with $\mathrm{aff.dim}(\Gamma) = d$,

- choose a Borel partition $(C_I(\Gamma))_{I \in \mathcal{I}(\Gamma)}$ of $\mathrm{conv}(\Gamma)$ such that

$$C_I(\Gamma) \subset D_I(\Gamma),$$

- let $U \sim \mathcal{U}[0,1]$ on $(\Omega_0, \mathcal{S}_0, \mathbb{P}_0)$.

The *optimal dual quantization operator* $\mathcal{J}_\Gamma^*$ is defined as

$$\mathcal{J}_\Gamma^*(\xi) = \sum_{I \in \mathcal{I}(\Gamma)} \left[\sum_{\ell=1}^{k_I} x_{i_\ell} \cdot \mathbf{1}_{\left\{ \sum_{j=1}^{\ell-1} \lambda_{i_j}^I(\xi) \le U < \sum_{j=1}^{\ell} \lambda_{i_j}^I(\xi) \right\}} \right] \mathbf{1}_{C_I(\Gamma)}(\xi).$$

where $I = \{i_1, \ldots, i_{k_I}\}$.

One easily checks that this operator is intrinsic stationary.

Equivalence of optimal dual quantization

Equivalence of optimal dual quantization

The operator $\mathcal{J}_\Gamma^*$ then leads to the following characterizations of the optimal dual quantization error:

Equivalence of optimal dual quantization

The operator $\mathcal{J}_\Gamma^*$ then leads to the following characterizations of the optimal dual quantization error:

Theorem ([PW10a])

Let $X \in L^p(\mathbb{P})$ and $N \in \mathbb{N}$. Then

$$d_{p,N}(X) =$$

Equivalence of optimal dual quantization

The operator $\mathcal{J}_\Gamma^*$ then leads to the following characterizations of the optimal dual quantization error:

Theorem ([PW10a])

Let $X \in L^p(\mathbb{P})$ and $N \in \mathbb{N}$. Then

$$d_{p,N}(X) = \inf\Big\{\big(\mathbb{E}\|X - \mathcal{J}_\Gamma(X)\|^p\big)^{\frac{1}{p}} : \mathcal{J}_\Gamma : \Omega_0 \times \mathbb{R}^d \to \Gamma \text{ is intrinsic stationary}, \\ \operatorname{supp}(\mathbb{P}_X) \subset \operatorname{conv}(\Gamma),\ |\Gamma| \le N\Big\}$$

Equivalence of optimal dual quantization

The operator $\mathcal{J}_\Gamma^*$ then leads to the following characterizations of the optimal dual quantization error:

Theorem ([PW10a])

Let $X \in L^p(\mathbb{P})$ and $N \in \mathbb{N}$. Then

$$\begin{aligned} d_{p,N}(X) = \inf\Big\{ \big(\mathbb{E}\|X - \mathcal{J}_\Gamma(X)\|^p\big)^{\frac{1}{p}} : \mathcal{J}_\Gamma : \Omega_0 \times \mathbb{R}^d \to \Gamma \text{ is intrinsic stationary}, \\ \operatorname{supp}(\mathbb{P}_X) \subset \operatorname{conv}(\Gamma),\ |\Gamma| \le N \Big\} \\ = \inf\big\{ \mathbb{E}\|X - \widehat{Y}\|^p : \widehat{Y} \text{ is a r.v. on } (\Omega \times \Omega_0, \mathcal{S} \otimes \mathcal{S}_0, \mathbb{P} \otimes \mathbb{P}_0), \\ |\widehat{Y}(\Omega \times \Omega_0)| \le N,\ \mathbb{E}(\widehat{Y}|X) = X \Big\}. \end{aligned}$$

Extension to unbounded support

Extension to unbounded support

- It is not possible to obtain intrinsic stationarity for $\xi \notin \mathrm{conv}(\Gamma)$

Extension to unbounded support

▷ It is not possible to obtain intrinsic stationarity for $\xi \notin \mathrm{conv}(\Gamma)$

▷ We extend $\mathcal{J}_\Gamma(X)$ outside $\mathrm{conv}(\Gamma)$ by using a Nearest Neighbor projection (which only preserves stationarity inside $\mathrm{conv}(\Gamma)$).

Extension to unbounded support

▷ It is not possible to obtain intrinsic stationarity for $\xi \notin \mathrm{conv}(\Gamma)$

▷ We extend $\mathcal{J}_\Gamma(X)$ outside $\mathrm{conv}(\Gamma)$ by using a Nearest Neighbor projection (which only preserves stationarity inside $\mathrm{conv}(\Gamma)$).

▷ We therefore drop the requirement $\mathrm{supp}(\mathbb{P}_X) \subset \mathrm{conv}(\Gamma)$ in above theorem and set

$$\bar{d}_{p,N}(X) = \inf\Big\{ \big(\mathbb{E}\|X - \mathcal{J}_\Gamma(X)\|^p\big)^{\frac{1}{p}} : \mathcal{J}_\Gamma \text{ is intrinsic stationary}, |\Gamma| \leq N \Big\}.$$

Existence of optimal dual quantizers

Existence of optimal dual quantizers

Theorem ([PW10a])

Let $p > 1$.

(a) Assume that $\mathbb{P}_X$ has a compact support. Then, for every $N \geq d+1$ there exists at least one optimal dual quantizer at level N (i.e. the dual quantization problem $d_N^p(X)$ attains its infimum).
Moreover, $d_N^p(X)$ strictly decreases to 0 as $N \to \infty$, if not vanishing.

Existence of optimal dual quantizers

Theorem ([PW10a])

Let $p > 1$.

(a) Assume that $\mathbb{P}_X$ has a compact support. Then, for every $N \geq d+1$ there exists at least one optimal dual quantizer at level N (i.e. the dual quantization problem $d_N^p(X)$ attains its infimum).
Moreover, $d_N^p(X)$ strictly decreases to 0 as $N \to \infty$, if not vanishing.

(b) Assume that the distribution $\mathbb{P}_X$ is strongly continuous. Then the same holds for $\bar{d}_N^p(X)$ for every $N \geq 1$.

Asymptotic behavior

Theorem (Sharp rate [PW10b])

(a) *Let* $X \in L^{p+}(\mathbb{R}^d)$ *with distribution* $\mathbb{P}_X = \varphi.\lambda^d \overset{\perp}{+} \nu$.

Asymptotic behavior

Theorem (Sharp rate [PW10b])

(a) Let $X \in L^{p+}(\mathbb{R}^d)$ with distribution $\mathbb{P}_X = \varphi.\lambda^d \overset{\perp}{+} \nu$. Then

$$\lim_{N\to\infty} N^{1/d} \cdot \bar{d}_{p,N}(X) = Q_{d,p,\|\cdot\|} \cdot \left(\int_{\mathbb{R}^d} \varphi^{d/(d+p)}\, d\lambda^d \right)^{\frac{d+p}{d}}$$

where $Q_{d,p,\|\cdot\|} = \inf_N N^{1/d} \cdot d_N^p\big(U([0,1]^d)\big)$.

Asymptotic behavior

Theorem (Sharp rate [PW10b])

(a) Let $X \in L^{p+}(\mathbb{R}^d)$ with distribution $\mathbb{P}_X = \varphi.\lambda^d \overset{\perp}{+} \nu$. Then

$$\lim_{N\to\infty} N^{1/d} \cdot \bar{d}_{p,N}(X) = Q_{d,p,\|\cdot\|} \cdot \left(\int_{\mathbb{R}^d} \varphi^{d/(d+p)}\, d\lambda^d \right)^{\frac{d+p}{d}}$$

where $Q_{d,p,\|\cdot\|} = \inf_N N^{1/d} \cdot d_N^p\big(U([0,1]^d)\big)$.

(b) If $d = 1$, $Q_{d,p,\|\cdot\|} = \frac{2^{1+1/p}}{(p+2)^{1/p}} \lim_{N\to\infty} N^{1/d} \cdot e_{p,N}\big(U([0,1])\big)$. *If $d \geq 2$,* ???

Asymptotic behavior

Theorem (Sharp rate [PW10b])

(a) *Let* $X \in L^{p+}(\mathbb{R}^d)$ *with distribution* $\mathbb{P}_X = \varphi.\lambda^d \overset{\perp}{+} \nu$. *Then*

$$\lim_{N\to\infty} N^{1/d} \cdot \bar{d}_{p,N}(X) = Q_{d,p,\|\cdot\|} \cdot \left(\int_{\mathbb{R}^d} \varphi^{d/(d+p)}\, d\lambda^d \right)^{\frac{d+p}{d}}$$

where $Q_{d,p,\|\cdot\|} = \inf_N N^{1/d} \cdot d_N^p\big(U([0,1]^d)\big)$.

(b) *If* $d = 1$, $Q_{d,p,\|\cdot\|} = \frac{2^{1+1/p}}{(p+2)^{1/p}} \lim_{N\to\infty} N^{1/d} \cdot e_{p,N}\big(U([0,1])\big)$. *If* $d \geq 2$, ???

(c) *Let* $p' > p$. *There exists* $C^{dual}_{p,p',d}$ *and* $N^{dual}_{p,p',d}$ *such that, for every r.v.* $\mathbb{R}^d$*-valued* X

$$\forall\, N \geq N^{dual}_{p,p',d}, \quad \bar{d}_{p,N}(X) \leq C^{dual}_{p,p',d} \inf_{a\in\mathbb{R}^d} \|X - a\|_{p'} \cdot N^{-\frac{1}{d}}.$$

(d) *The same holds for compactly supported r.v. for the mean quantization error* $d_{n,p}(X)$ *with the same asymptotic constant* $Q_{d,p,\|\cdot\|}$.

Asymptotic behavior

Sketch of the proof

Asymptotic behavior

Sketch of the proof

- Prove existence of the limit for $\mathcal{U}([0,1]^d)$

Asymptotic behavior

Sketch of the proof

- Prove existence of the limit for $\mathcal{U}([0,1]^d)$
- Derive upper and lower bounds for piecewise constant densities (with compact support) on hypercubes

Asymptotic behavior

Sketch of the proof

- Prove existence of the limit for $\mathcal{U}([0,1]^d)$
- Derive upper and lower bounds for piecewise constant densities (with compact support) on hypercubes
- Use Differentiation of measure to cover the general case (still compact support)

Asymptotic behavior

Sketch of the proof

- Prove existence of the limit for $\mathcal{U}([0,1]^d)$
- Derive upper and lower bounds for piecewise constant densities (with compact support) on hypercubes
- Use Differentiation of measure to cover the general case (still compact support)
- Random dual quantization argument (so-called extended Pierce Lemma) to get the unbounded case.

Numerical computations: the grids

▷ Differentiability as a function of Γ for every $\xi \in \mathrm{conv}(\Gamma)$,

$$\Gamma = (x_1, \ldots, x_N) \longmapsto F_p^p(\xi, \Gamma)$$

is differentiables except at a grid Γ_0 except for a λ^d-negligible set of values of ξ, namely $\cup_{I \in \mathcal{I}(\Gamma)} \partial D_I(\Gamma)$.

▷ Hence, if X has an absolutely continuous distribution,

$$\Gamma \longmapsto \mathbb{E}(F_p^p(X, \Gamma)) \quad \text{is differentiable}$$

with gradient at Γ given by

$$\mathbb{E}\Big(\frac{\partial F_p^p}{\partial \xi}(X, \Gamma)\Big)$$

▷ Provides a stochasic gradient descent procedure (counterpart of $CLVQ$) or a couterpart of randomized Lloyd's procedure..

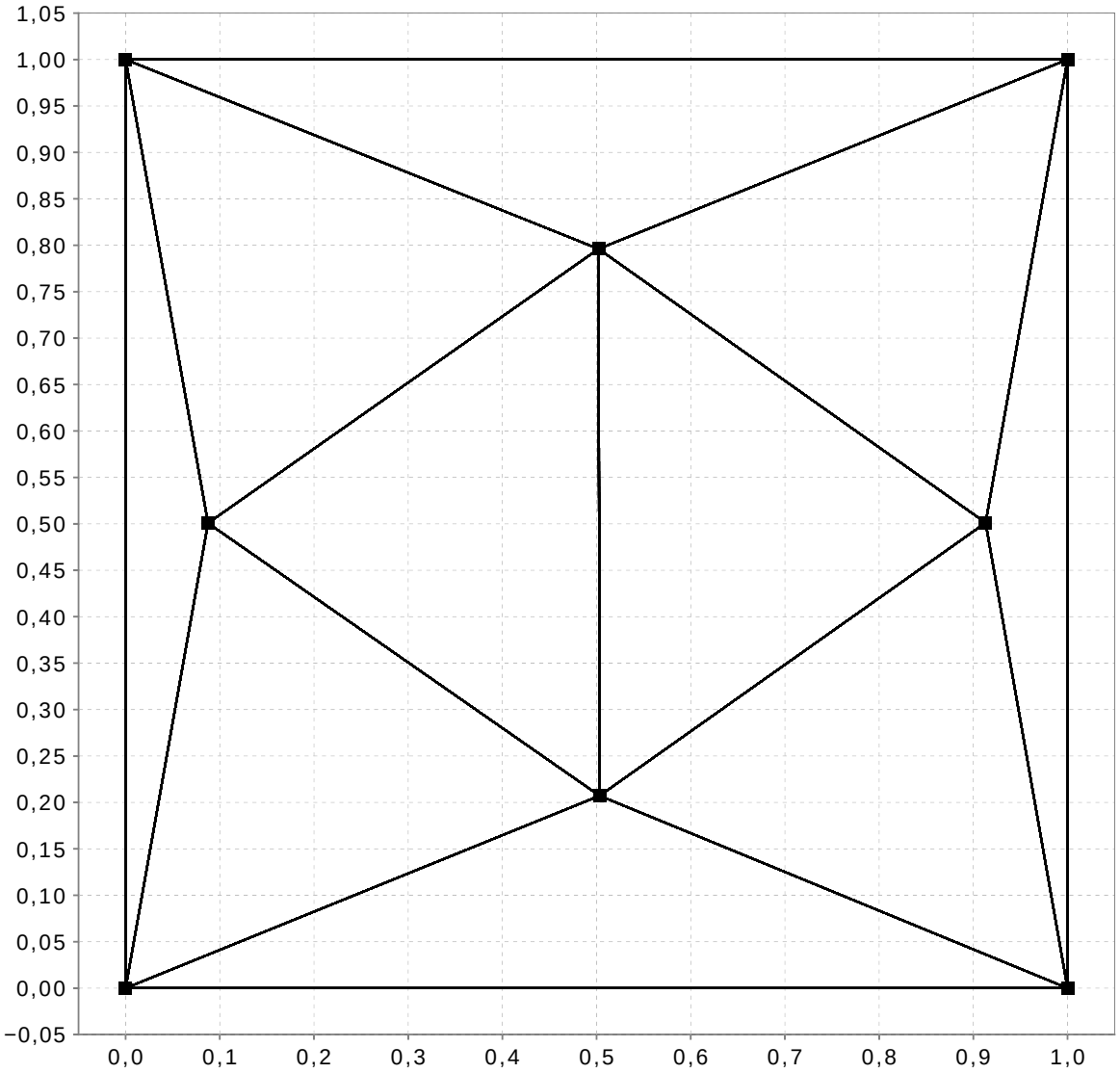


Figure: Dual Quantization for $\mathcal{U}([0,1]^2)$ and $n = 8$

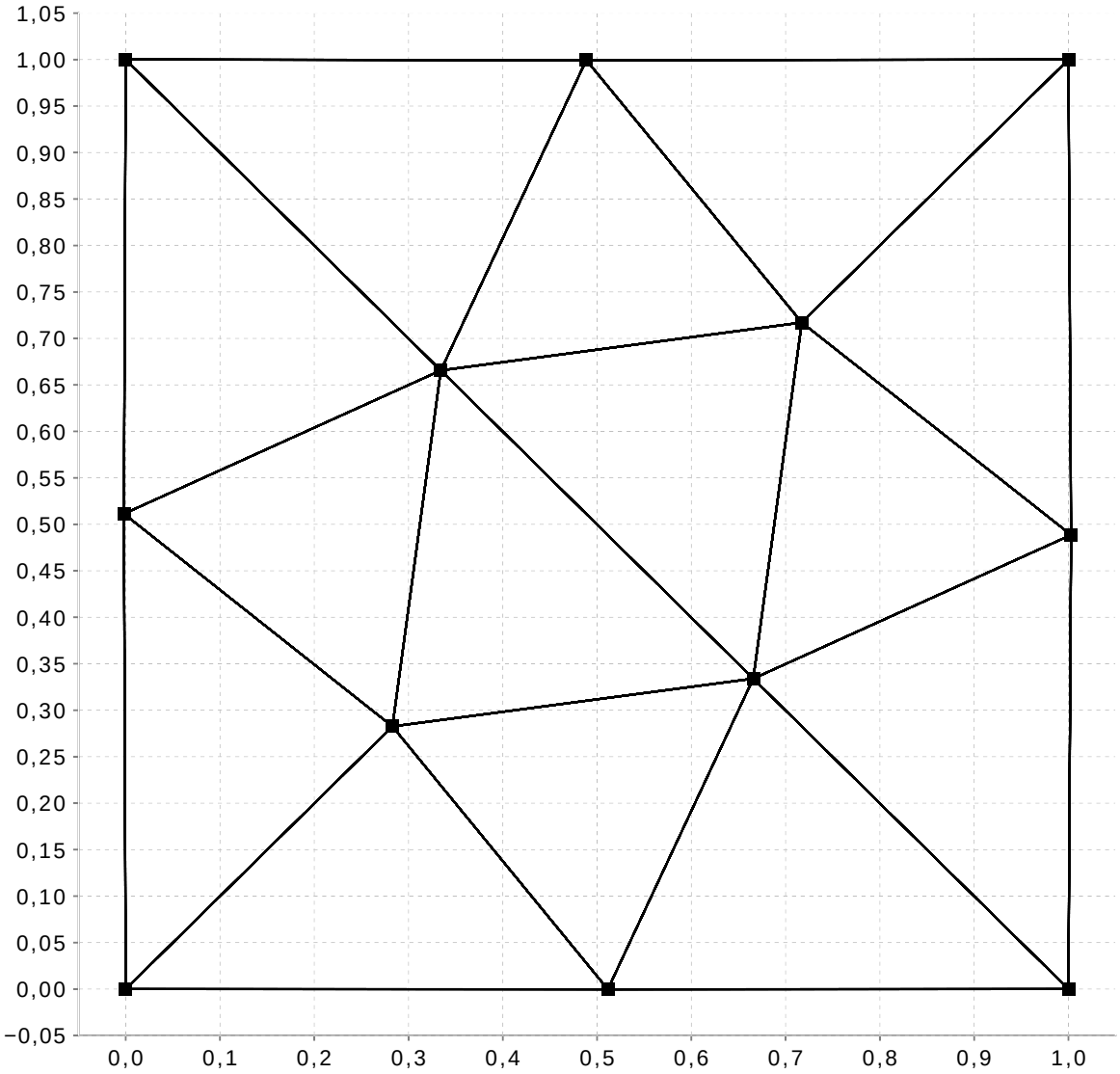


Figure: Dual Quantization for $\mathcal{U}([0,1]^2)$ and $n = 12$

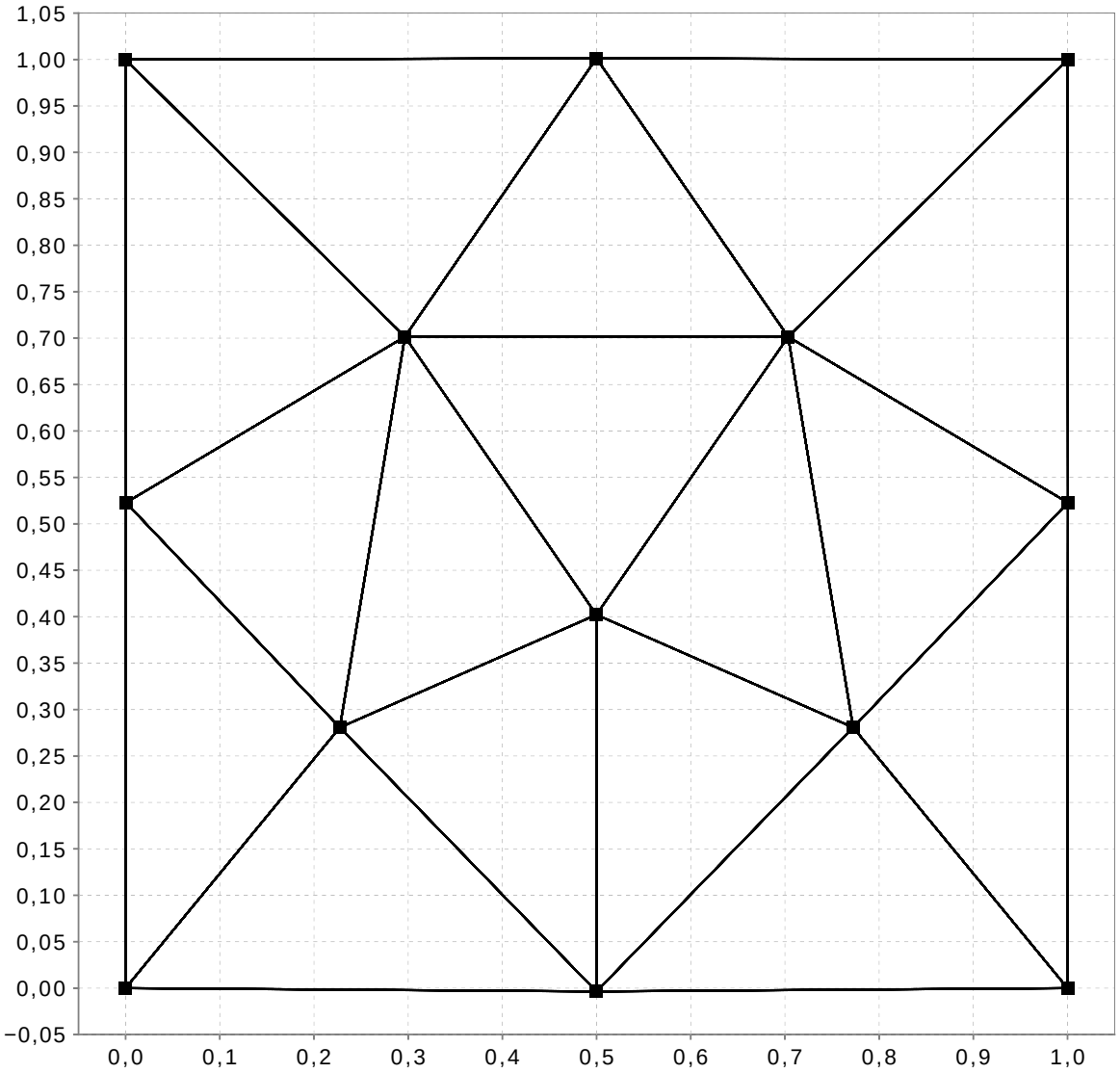


Figure: Dual Quantization for $\mathcal{U}([0,1]^2)$ and $n = 13$

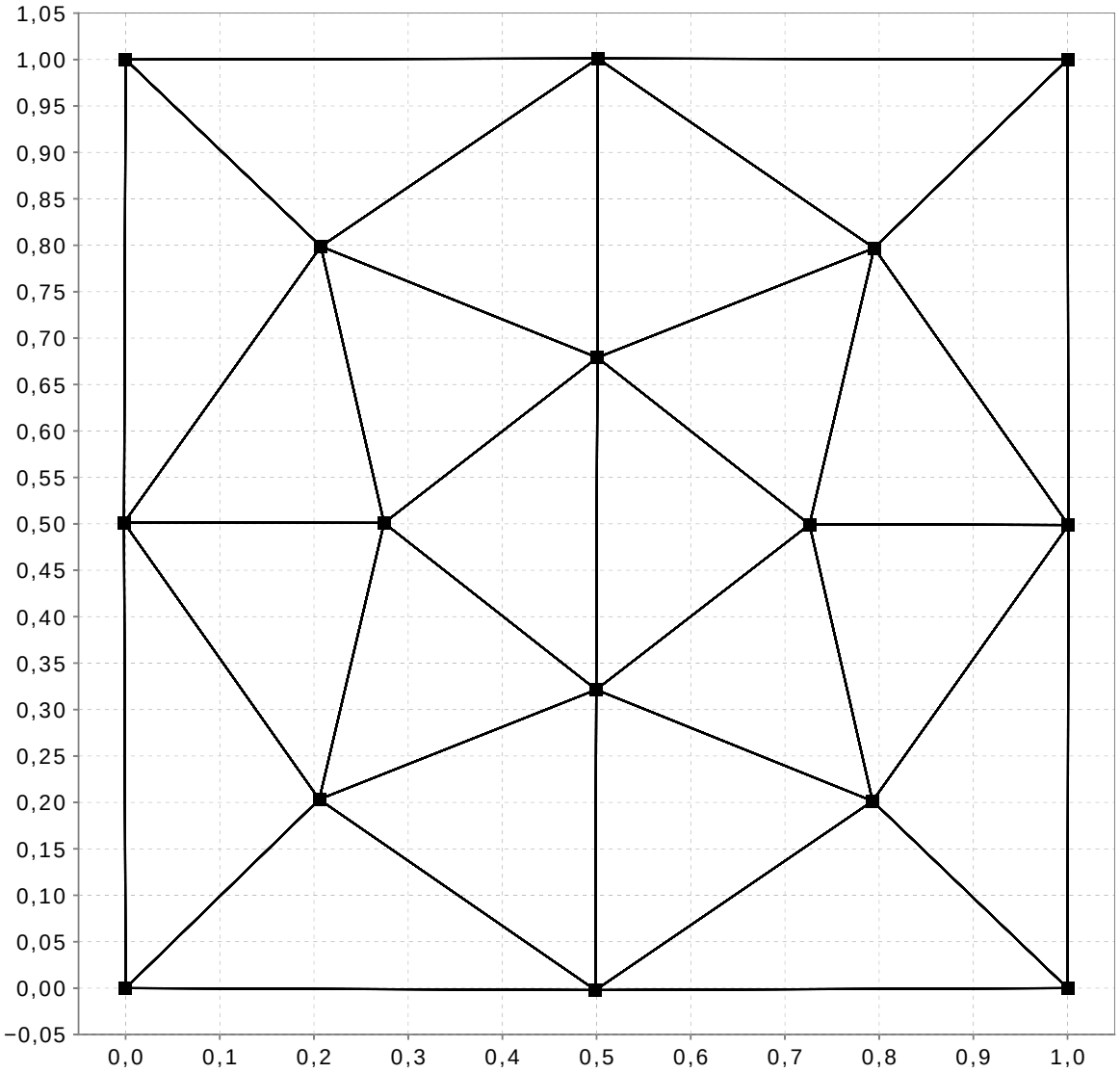


Figure: Dual Quantization for $\mathcal{U}([0,1]^2)$ and $n = 16$

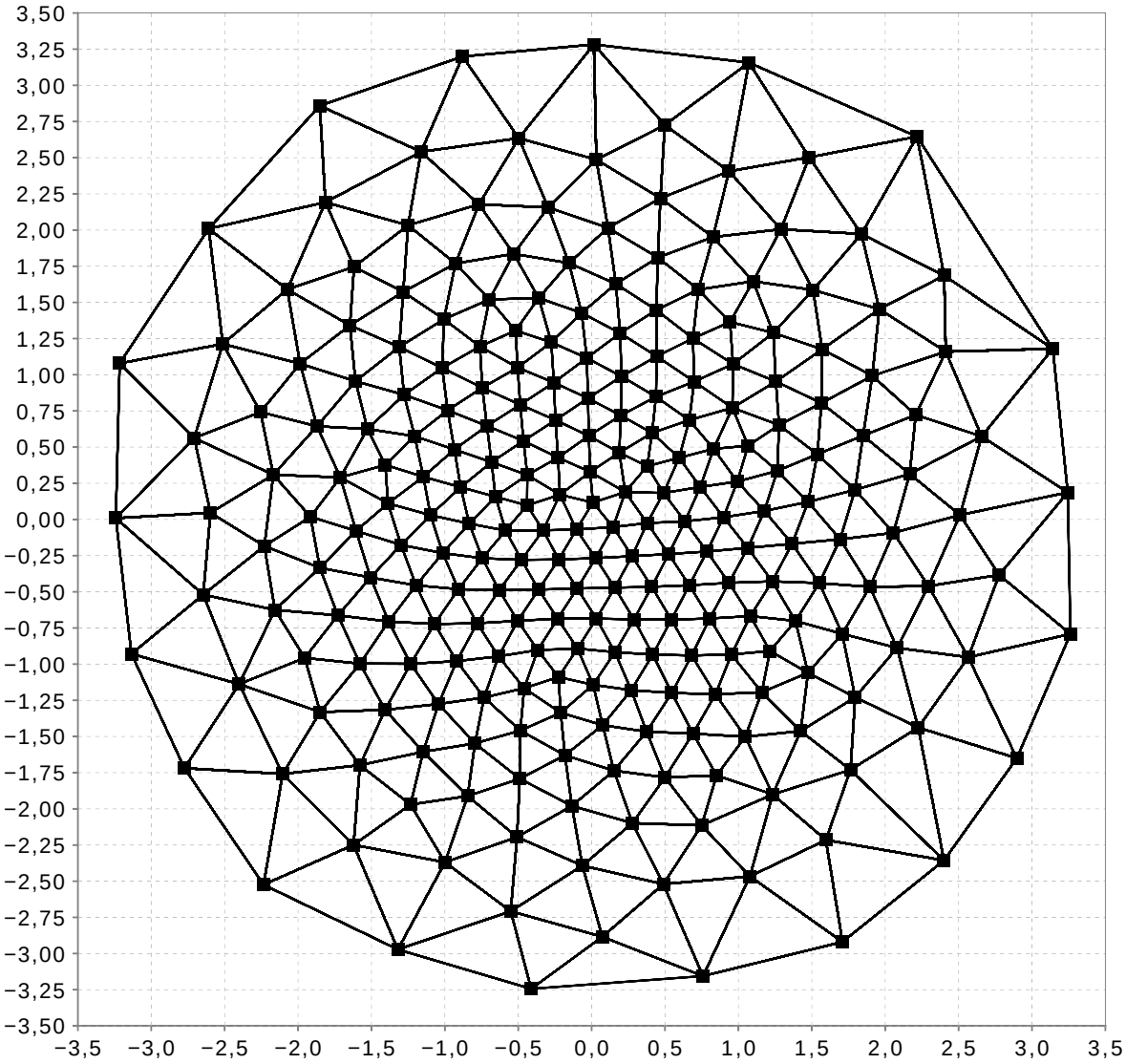


Figure: Dual Quantization for $\mathcal{N}(0, I_2)$ and $N = 250$

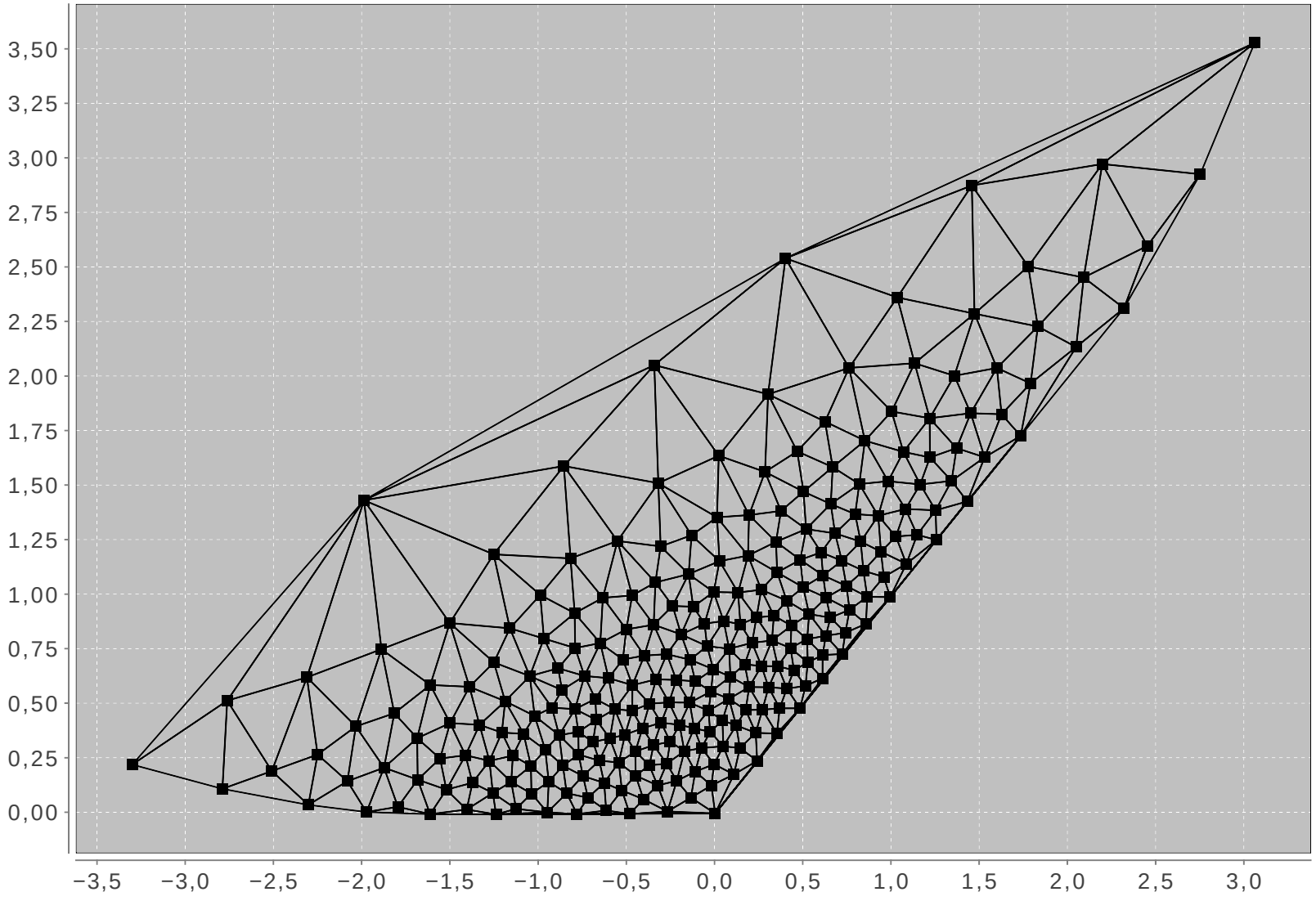


Figure: Joint Dual Quantization of the BM and its supremum, $N = 250$

Numerical computations: the weights

▷ STATIC WEIGHT COMPUTATION. Let $\Gamma = \{x_1, \ldots, x_N\}$. How to compute

$$p_i = \mathbb{P}(\mathcal{J}_\Gamma^U(X) = x_i) \ ?$$

By a (possibly massively parallel) Monte Carlo simulation

$$p_i = \lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} \mathbb{P}\Big(\mathcal{J}_\Gamma^{U_m}(X^{(m)}) = x_i | X^{(m)}\Big)$$

where $(X^{(m)}, U^{(m)})$, $m \geq 1$ are independent copies of $X \perp\!\!\!\perp U$ and

$$\mathbb{E}\Big(\mathcal{J}_\Gamma^{U_m}(X^{(m)}) | X^{(m)}\Big) \quad = \quad \text{barycentric coordinate of } X^{(m)} \text{ in “its” simplex} \ni x_i$$

Bermuda options

In the same way we use the Backward Dynamic Principle for the valuation of Bermuda options:

BDP for Bermuda options

$$\widehat{V}_n = \varphi_{t_n}(\widehat{X}_n)$$
$$\widehat{V}_k = \max\Big\{\varphi_{t_k}(\widehat{X}_k);\, \mathbb{E}\big(\widehat{V}_{k+1}\big|\widehat{X}_k\big)\Big\},\ 0 \le k \le n-1,$$

so that $\widehat{V}_0$ yields an approximation to the Bermuda option premium

$$V_0 = \text{esssup}\{\mathbb{E}\,\varphi(X_\tau) : \tau \text{ is } \{t_0, \ldots, t_n\}\text{-valued stopping time}\}.$$

Error bounds

Theorem (P.-Wilbertz 2010)

$$V_k = v_k(X_k) \quad \textit{and} \quad \widehat{V}_k =: \widehat{v}_k(\widehat{X}_k),\ k = 0, \ldots, n$$

and

$$\|v_k(X_k) - \widehat{v}_k(\widehat{X}_k)\|_p \leq \kappa_{p,p'} \sum_{\ell=k}^{n} C_{n,\ell}([v]_{Lip}, [P]_{Lip}) \sigma_{p'}(X_k) N_k^{-\frac{1}{d}}$$

where $\sigma_{p'}(X_k) = \min_{a \in \mathbb{R}^d} \|X_k - a\|_{p'}$ *is the* $L^{p'}$*-median of* X_k, $p' > p$.

▷ Optimization of the quantization tree structure for a given budget N

$$\min_{N_0 + \cdots N_n \leq N} \sum_{k=0}^{n} C_{n,k}([v]_{Lip}, [P]_{Lip}) \sigma_{p'}(X_k) N_k^{-\frac{1}{d}}$$

$$\Longrightarrow \qquad N_k = \frac{\cdots}{\cdots},\ k = 0, \ldots, n.$$

Numerical experiments I

Example

$2d$-asset (uncorrelated) Black-Scholes model with maturity $T = 1$, 11 exercise dates: $k/10$, $k = 0, \ldots, 10$, and

$$s_0^i = 40^{\frac{2}{d}},\ i = 1, \ldots, k,\ s_0^i = 40^{\frac{2}{d}},\ i = k+1, \ldots, d,\ r = 0.05,$$
$$\sigma_i = 0.2,\ i = 1, \ldots, d,\ \delta_i = 0.05, i = 1, \ldots, k,\ \delta_i = 0.0,\ i = k+1, \ldots, d.$$

for a geometric exchange put option

$$\varphi(S_t^1, S_t^2) = \Big(\prod_{i=1}^{k} S_t^i - \prod_{i=k+1}^{d} S_t^i\Big)_+.$$

This can be reduced for any d to a 2-dim exchange option.
Hence reference values are available using a Boyle-Evnine-Gibbs tree with 10 000 time steps.

Printems's paradigm: log-log plots for true rates

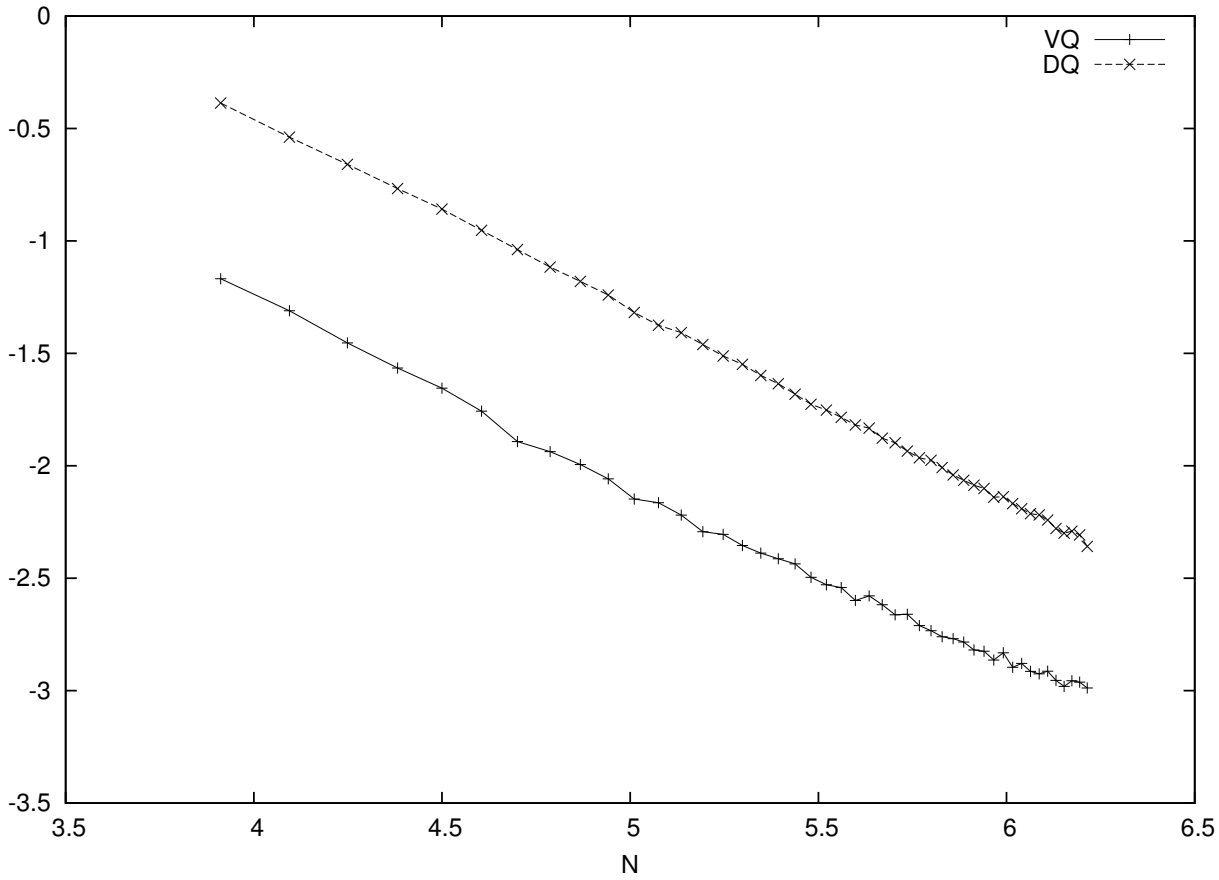


Figure: Log-Log plot of both quantization methods in dimension 2

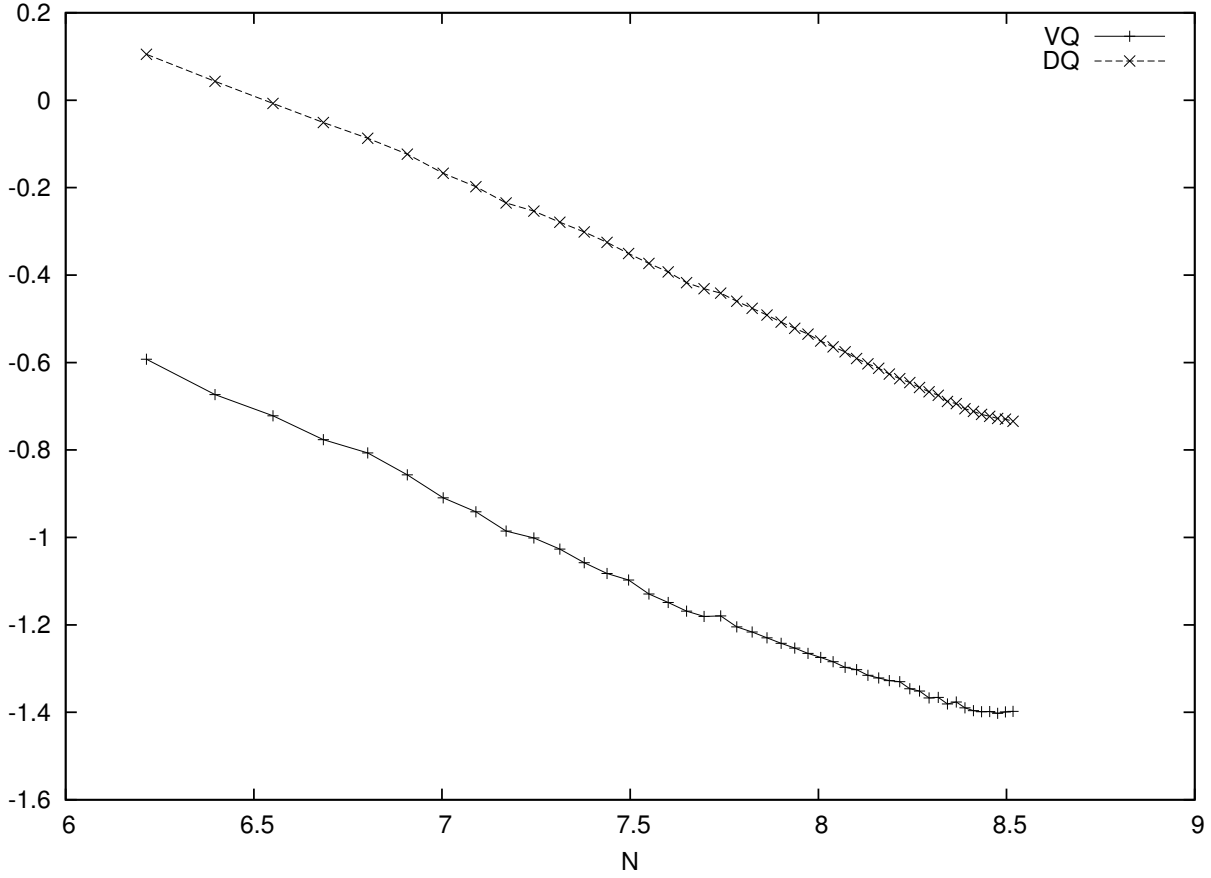


Figure: Log-Log plot of both quantization method in dimension 4

	$2d = 2$	$2d = 4$
Voronoi Quantization	0.73	0.36
Dual Quantization	0.86	0.38

Table: Rates of convergence for the exchange option.

Richardson-Romberg extrapolation

▷ We proceed a (heuristic) Richardson-Romberg extrapolation on the (guessed) error expansion.

$$\mathbb{E}\, F(X) \approx \mathbb{E}\, F(\widehat{X}) + \kappa\, N^{-\alpha}$$

▷ We extrapolate the unknown κ using two different grids sizes N_1 and N_2. As a result, we obtain in the above setting for

$$\hat{P}_0^{\mathrm{Rom}} = \hat{P}_0^{N_1} + \frac{\hat{P}_0^{N_1} - \hat{P}_0^{N_2}}{N_2^{-\alpha} - N_1^{-\alpha}} N_1^{-\alpha}$$

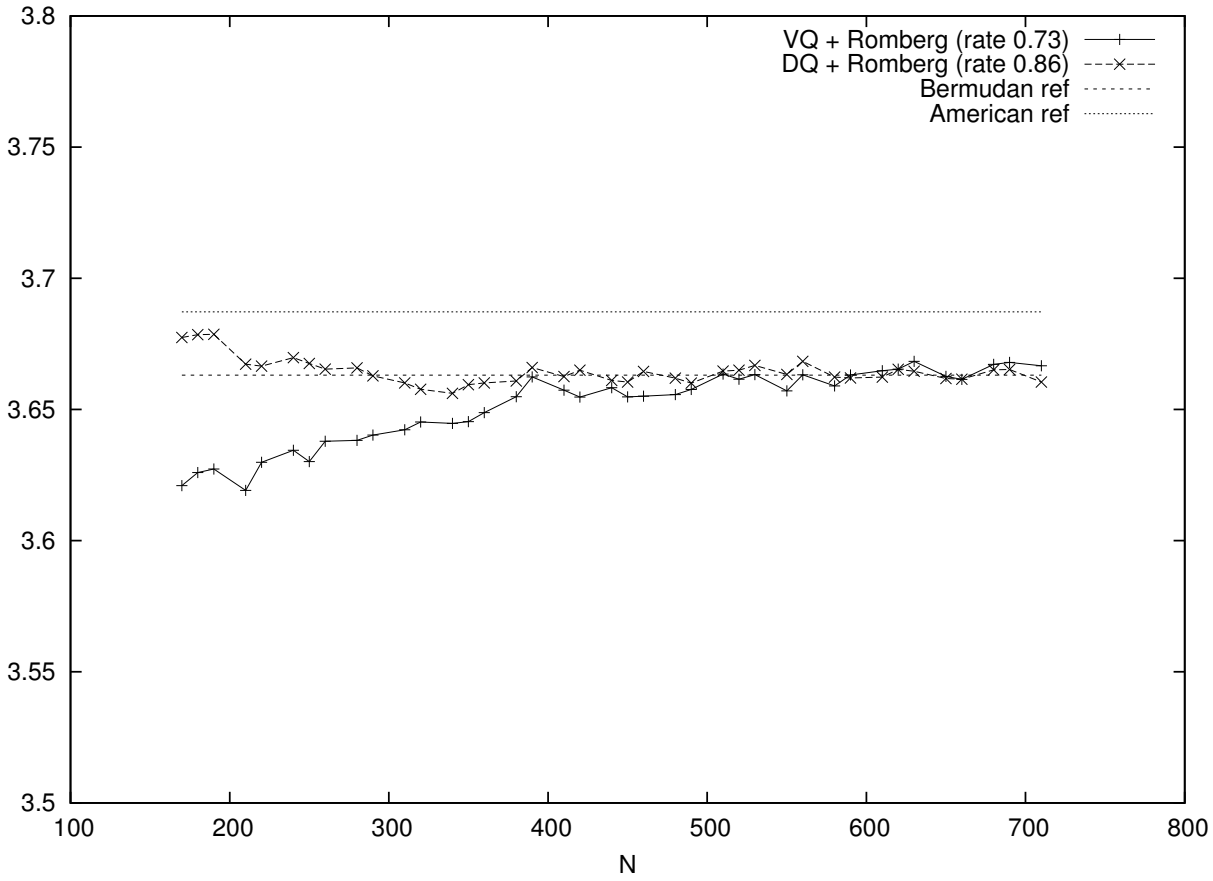


Figure: Convergence of the extrapolated quantization methods for the geometric exchange option in dimension 2

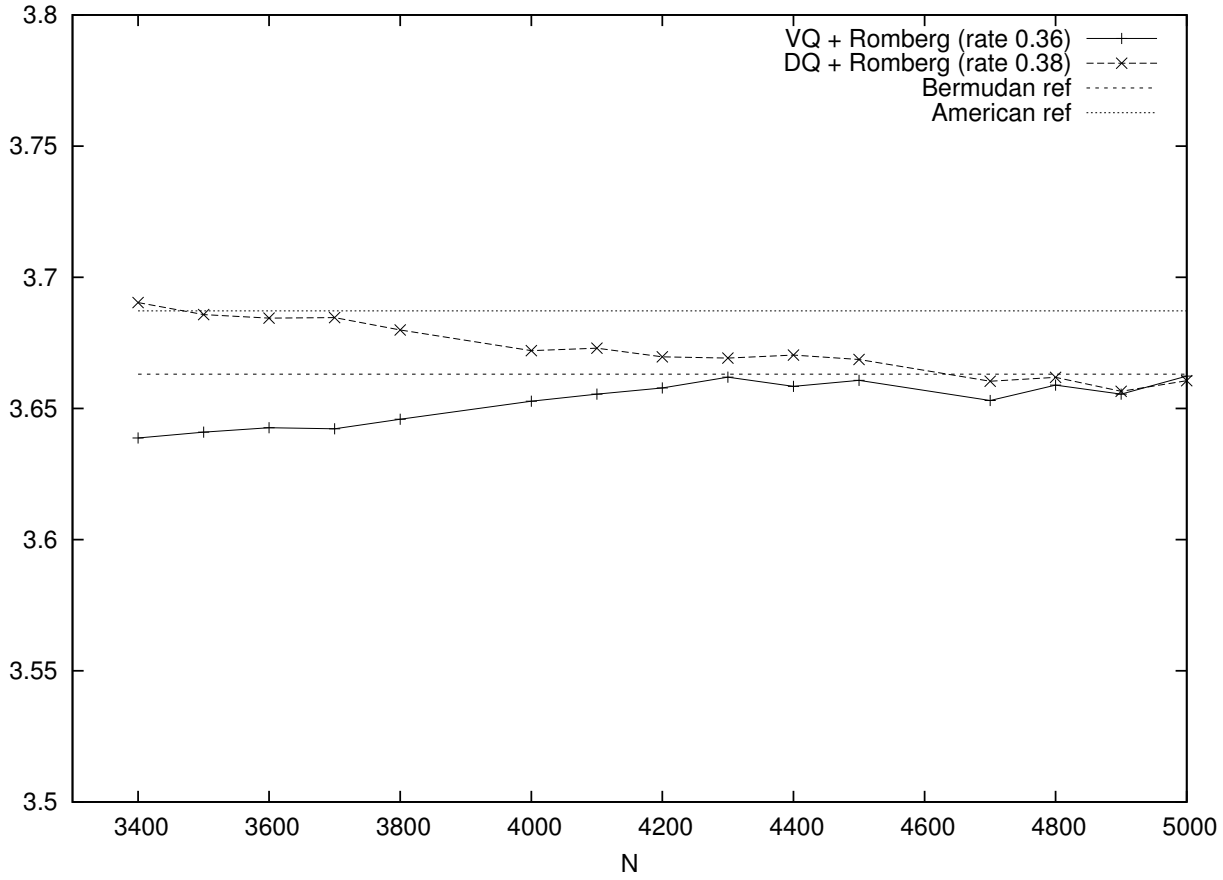


Figure: Convergence of the extrapolated quantization methods for the geometric exchange option in dimension 4

Suggestion: Adopt the mid-price $0.5 \times (Price_{VQ} + Price_{DQ})$ computed on an optimal Voronoi quantization tree.

Bermuda: Numerical experiments II

Example

2-asset Black-Scholes model with

$$s_0^1 = s_0^2 = 40,\ r = 0.05,\ \sigma_1 = 0.2,\ \sigma_2 = 0.3,\ \rho = 0.5,\ K = 40,$$

for a put on the min, i.e. payoff

$$\varphi(S_t^1, S_t^2) = (K - \min\{S_t^1, S_t^2\})^+.$$

Bermuda: Numerical experiments II

Example

2-asset Black-Scholes model with

$$s_0^1 = s_0^2 = 40,\ r = 0.05,\ \sigma_1 = 0.2,\ \sigma_2 = 0.3,\ \rho = 0.5,\ K = 40,$$

for a put on the min, i.e. payoff

$$\varphi(S_t^1, S_t^2) = (K - \min\{S_t^1, S_t^2\})^+.$$

As underlying Markov process we have chosen a 2-dimensional Brownian Motion with correlation ρ.

Bermuda: Numerical experiments II

Example

2-asset Black-Scholes model with

$$s_0^1 = s_0^2 = 40,\ r = 0.05,\ \sigma_1 = 0.2,\ \sigma_2 = 0.3,\ \rho = 0.5,\ K = 40,$$

for a put on the min, i.e. payoff

$$\varphi(S_t^1, S_t^2) = (K - \min\{S_t^1, S_t^2\})^+.$$

As underlying Markov process we have chosen a 2-dimensional Brownian Motion with correlation ρ.

Reference values still computed using a Boyle-Evnine-Gibbs tree with 10.000 timesteps.

Martingale Adjustment

▷ If the structure process $(X_k)_{0\le k\le n}$is a martingale $(\dots)$ and $X_0 = x_0$, the attached quantization trees loose this property.

▷ One idea is to restore the martingality by moving the grids Γ_k:

– Define by a backward induction $\widetilde{\Gamma}_n = \Gamma_n$ and for every $k = 0, \dots, n-1$,

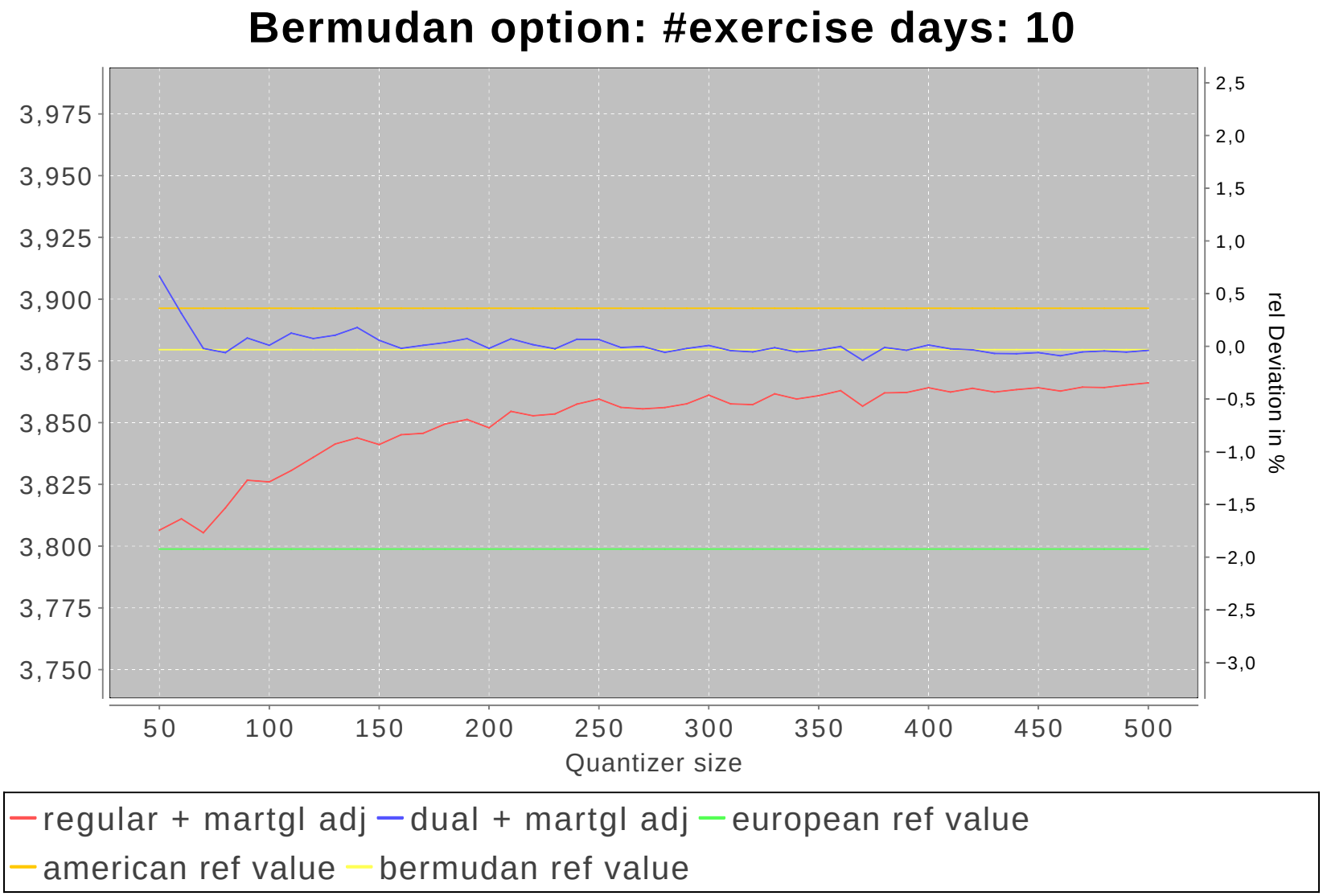
$$\widetilde{\Gamma}_k = \left\{\widetilde{x}_1^k, \dots, \widetilde{x}_{N_k}^k\right\} \quad \text{where} \quad \widetilde{x}_i^k = \sum_{j=1}^{N_{k+1}} p_{ij}^k \widetilde{x}_j^{k+1},\ i = 1, \dots, N_k.$$

– Re-center the grids by setting

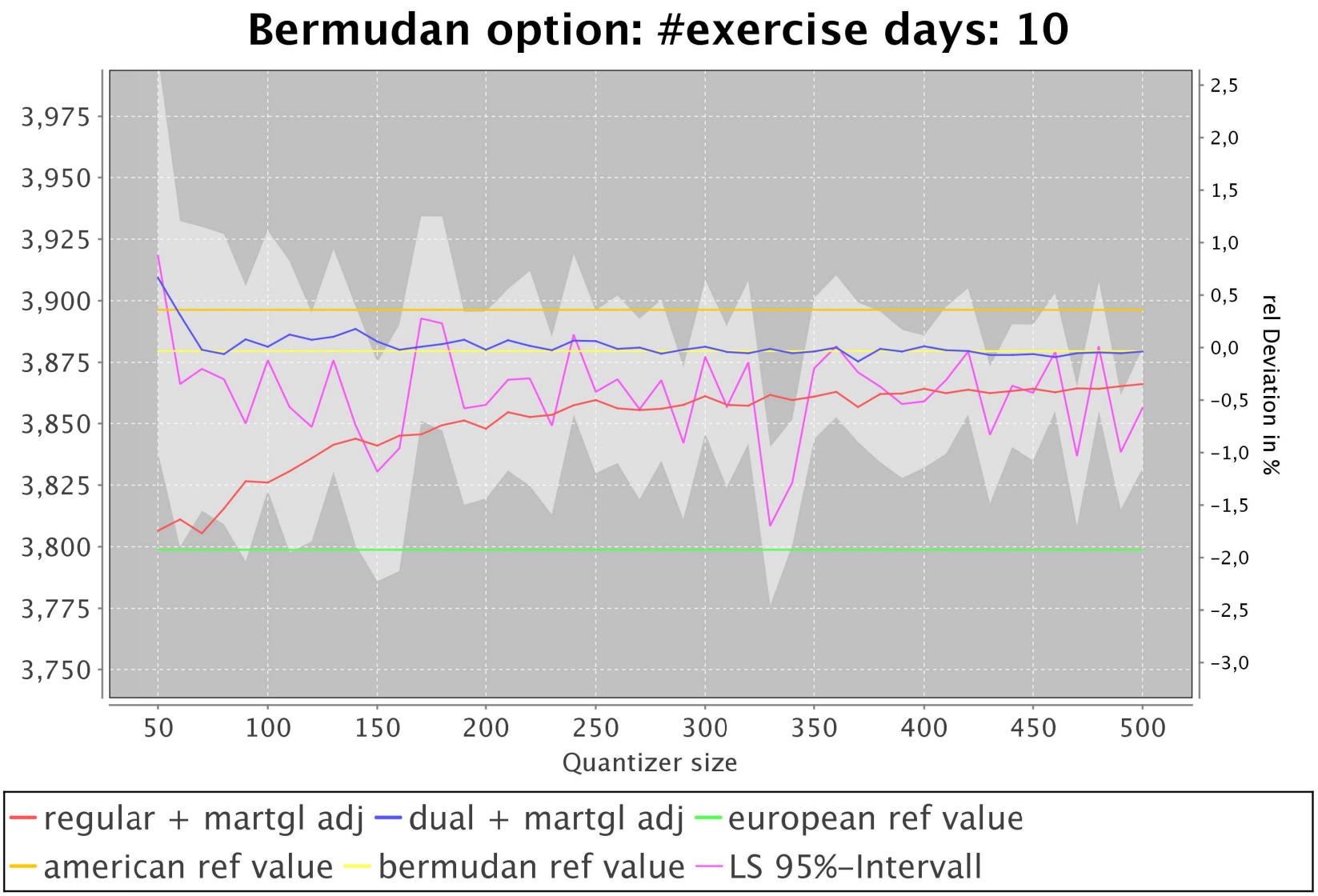
$$\Gamma_k^{mart} = \widetilde{\Gamma}_k + x_0 - \widetilde{x}_0.$$

▷ The resulting quantization tree $(\Gamma_k^{mart}, \mathbf{p}^k)_{0\le k\le n}$ has the distribution of a martingale starting at x_0 at time 0. (It is observed that the translation $x_0 - \widetilde{x}_0$ is negligible in practice).

Martingale Adjustment: numerical experiments...



including Longstaff-Schwartz by Premia



Conclusion / Outlook

Conclusion / Summary

Conclusion / Outlook

Conclusion / Summary

- Interesting and challenging extention of regular Quantization

Conclusion / Outlook

Conclusion / Summary

- Interesting and challenging extention of regular Quantization
- Provides a stationarity, which holds independently of the choice of the quantization grid

Conclusion / Outlook

Conclusion / Summary

- Interesting and challenging extention of regular Quantization
- Provides a stationarity, which holds independently of the choice of the quantization grid
- Represented in the Euclidean case by the dual concept of Voronoi tesselations: the Delaunay triangulation

Conclusion / Outlook

Conclusion / Summary

- Interesting and challenging extention of regular Quantization
- Provides a stationarity, which holds independently of the choice of the quantization grid
- Represented in the Euclidean case by the dual concept of Voronoi tesselations: the Delaunay triangulation
- Yields very promising results in first numerical applications

Conclusion / Outlook

Conclusion / Summary

- Interesting and challenging extention of regular Quantization
- Provides a stationarity, which holds independently of the choice of the quantization grid
- Represented in the Euclidean case by the dual concept of Voronoi tesselations: the Delaunay triangulation
- Yields very promising results in first numerical applications
- Further applications in optimal grid generation, adaptive grid refinements possible

Conclusion / Outlook

Conclusion / Summary

- Interesting and challenging extention of regular Quantization
- Provides a stationarity, which holds independently of the choice of the quantization grid
- Represented in the Euclidean case by the dual concept of Voronoi tesselations: the Delaunay triangulation
- Yields very promising results in first numerical applications
- Further applications in optimal grid generation, adaptive grid refinements possible
- Application to 3-factor models, etc.

References I

V. Bally and G. Pagès.
Error analysis of the optimal quantization algorithm for obstacle problems.
Stochastic Processes and their Applications, 106:1–40(40), July 2003.

S. Graf and H. Luschgy.
Foundations of Quantization for Probability Distributions.
Lecture Notes in Mathematics $n^0$1730. Springer, Berlin, 2000.

R.M. Gray and D.L. Neuhoff.
Quantization.
IEEE Trans. Inform., 44:2325–2383, 1998.

G. Pagès and B. Wilbertz.
Optimal delaunay and voronoi quantization schemes for pricing american style options.
LPMA-1425, to appear in *Numerical methods in Finance* (R. Carmona, P. Hu, P. Del Moral, .N. Oudjane eds.), Springer, 2012, 2009.

G. Pagès and B. Wilbertz.
Intrinsic stationarity for vector quantization: foundations of dual quantization.
Technical report, LPMA-1393, 2010.

References II

G. Pagès and B. Wilbertz.
Sharp rate for the dual quantization problem.
Technical report, LPMA-1402, 2010.

V. T. Rajan.
Optimality of the delaunay triangulation in R^d.
In *SCG '91: Proceedings of the seventh annual symposium on Computational geometry*, pages 357–363, New York, NY, USA, 1991. ACM.