Improved error bounds for quantization based numerical schemes

Gilles Pagès
(joint work with Abass Sagna)

LPMA-Université Pierre et Marie Curie

—

—

Conference in honour of Vlad Bally

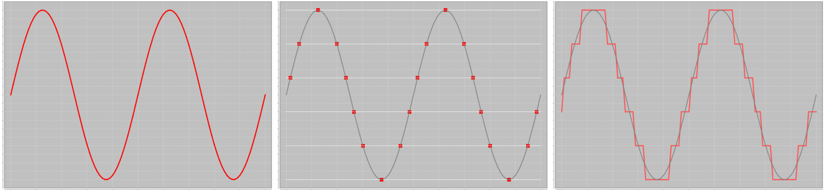
–

Le Mans

9 Octobre 2015

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 its recovery/reconstruction from the discretized one.



- Examples: Pulse-Code-Modulation (PCM), JPEG-Compression
- Extensive Survey about the IEEE-History: *IEEE on Inf. Theory*, 1982, [Gersho-Gray eds]
- Mathematical Foundation of Quantization Theory: S. Graf & H. Luschgy in *Foundation of quantization of probability measures*, LNM 2000.
- P. : Survey on Optimal Vector Quantization and its applications for numerics, *ESAIM Proc. & Surveys*, CEMRACS'13 course, 2015.

At the beginning was rough quantization

At the beginning was rough quantization

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

$$\mathbb{E}|X|^p < +\infty \qquad \text{for some } p \in (0, +\infty).$$

At the beginning was rough quantization

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

$$\mathbb{E}|X|^p < +\infty \qquad \text{for some } p \in (0, +\infty).$$

▷ Let $q : \mathbb{R}^d \to \Gamma \subset \mathbb{R}^d$ be a Borel function, Γ a finite subset of $\mathbb{R}^d$.

At the beginning was rough quantization

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

$$\mathbb{E}|X|^p < +\infty \qquad \text{for some } p \in (0, +\infty).$$

▷ Let $q : \mathbb{R}^d \to \Gamma \subset \mathbb{R}^d$ be a Borel function, Γ a finite subset of $\mathbb{R}^d$.

$$\widehat{X} = q(X)$$

is called a quantization of X. It aims at being a *discretization* of X

▷ Example: if X is $[0,1]$-valued, one may choose

$$q(x) = \frac{\lfloor Nx \rfloor}{N},\ x \in [0,1]$$

or

$$q(x) = \frac{2k-1}{2N},\ \text{ if } \frac{k-1}{N} \le x \le \frac{k}{N},\ x \in [0,1]$$

At the beginning was rough quantization

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

$$\mathbb{E}|X|^p < +\infty \qquad \text{for some } p \in (0, +\infty).$$

▷ Let $q : \mathbb{R}^d \to \Gamma \subset \mathbb{R}^d$ be a Borel function, Γ a finite subset of $\mathbb{R}^d$.

$$\widehat{X} = q(X)$$

is called a quantization of X. It aims at being a *discretization* of X

▷ Example: if X is $[0,1]$-valued, one may choose

$$q(x) = \frac{\lfloor Nx \rfloor}{N}, \; x \in [0,1]$$

or

$$q(x) = \frac{2k-1}{2N}, \;\; \text{if} \;\; \frac{k-1}{N} \le x \le \frac{k}{N}, \; x \in [0,1]$$

▷ L^p-mean quantization error:

$$e_{p,N}(X; q) = \big\| X - q(X) \big\|_p$$

What is it for? Quantization for Cubature

Let $\Gamma = \{x_1, \ldots, x_N\}$. Assume that we have access to the *elementary quantizers* x_i and the weights

$$w_i(q) := \mathbb{P}(\widehat{X} = x_i), \; i = 1, \ldots, N.$$

What is it for? Quantization for Cubature

Let $\Gamma = \{x_1, \dots, x_N\}$. Assume that we have access to the *elementary quantizers* x_i and the weights

$$w_i(q) := \mathbb{P}(\widehat{X} = x_i),\ i = 1, \dots, N.$$

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

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

What is it for? Quantization for Cubature

Let $\Gamma = \{x_1, \ldots, x_N\}$. Assume that we have access to the *elementary quantizers* x_i and the weights

$$w_i(q) := \mathbb{P}(\widehat{X} = x_i),\ i = 1, \ldots, N.$$

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

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

▷ If F is Lipschitz continuous, a first error estimate reads

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

Quantization for Conditional expectation

▷ Applications in Numerical Probability = conditional expectation approximation.

$$\widehat{X} = q_X(X) \qquad \widehat{Y} = q_Y(Y)$$

Quantization for Conditional expectation

▷ Applications in Numerical Probability = conditional expectation approximation.

$$\widehat{X} = q_x(X) \qquad \widehat{Y} = q_Y(Y)$$

Proposition (Pythagoras' Theorem for conditional expectation)

Let $P(y, du) = \mathcal{L}(X \mid Y = y)$ be a regular version of the conditional distribution of X given Y, so that

$$\mathbb{E}(g(X) \mid Y) = Pg(Y) \text{ a.s.}$$

Then

$$\begin{aligned}\big\|\mathbb{E}(g(X) \mid Y) - \mathbb{E}(g(\widehat{X}) \mid \widehat{Y})\big\|_2^2 &\leq [g]_{\mathrm{Lip}}^2 \big\|X - \widehat{X}\big\|_2^2 + \big\|Pg(Y) - Pg(\widehat{Y})\big\|_2^2 \\ &\leq [g]_{\mathrm{Lip}}^2 \big\|X - \widehat{X}\big\|_2^2 + [Pg]_{\mathrm{Lip}}^2 \big\|Y - \widehat{Y}\big\|_2^2.\end{aligned}$$

If P propagates Lipschitz continuity:

$$[Pg]_{\mathrm{Lip}} \leq [P]_{\mathrm{Lip}}[g]_{\mathrm{Lip}}.$$

then quantization produces a control of the error.

Quantization for Conditional expectation

▷ Sketch of proof As

$$Pg(Y) - \mathbb{E}\big(Pg(Y)\,|\,\widehat{Y}\big) \overset{L^2(\mathbb{P})}{\perp} \sigma(\widehat{Y})$$

so that by Pythagoras' theorem

$$\big\|\mathbb{E}\big(g(X)\,|\,Y\big) - \mathbb{E}\big(g(\widehat{X})|\widehat{Y}\big)\big\|_2^2 = \big\|Pg(Y) - \mathbb{E}\big(Pg(Y)|\widehat{Y}\big)\big\|_2^2 + \big\|\mathbb{E}\big(Pg(X)|\widehat{Y}\big) - \mathbb{E}\big(g(\widehat{X})|\widehat{Y}\big)\big\|_2^2$$

$$\leq \big\|Pg(Y) - Pg(\widehat{Y})\big)\big\|_2^2 + \big\|g(X) - g(\widehat{X})\big\|_2^2.$$

$$\leq [Pg]_{\rm Lip}^2\big\|Y - \widehat{Y}\big)\big\|_2^2 + [g]_{\rm Lip}^2\big\|X - \widehat{X}\big\|_2^2.$$

▷ If $p \neq 2$, a Minkowski like control is preserved

$$\begin{aligned}\big\|\mathbb{E}\big(g(X)\,|\,Y\big) - \mathbb{E}\big(g(\widehat{X})\,|\,\widehat{Y}\big)\big\|_p &\leq [g]_{\rm Lip}\big\|X - \widehat{X}\big\|_p + \big\|Pg(Y) - Pg(\widehat{Y})\big\|_p \\ &\leq [g]_{\rm Lip}\big\|X - \widehat{X}\big\|_p + [Pg]_{\rm Lip}\big\|Y - \widehat{Y}\big\|_p.\end{aligned}$$

A typical result (BSDE)

▷ We consider a "standard" BSDE:

$$Y_t = h(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)ds - \int_t^T Z_s dW_s, \quad t \in [0, T],$$

where the exogenous process $(X_t)_{t\in[0,T]}$ is a diffusion

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \quad x \in \mathbb{R}^d.$$

with b, σ, h Lipschitz continuous in x, f Lipschitz in (x, y, z) uniformly in $t \in [0, T]$...

▷ which is the probabilistic representation of the partially non-linear *PDE*

$$\partial_t u(t, x) + Lu(t, x) + f(t, x, u(t, x), (\partial_x^* u\sigma)(t, x)) = 0 \text{ on } [0, T) \times \mathbb{R}^d, \quad u(T, .) = h$$

with $Lg = (\nabla b|g) + \frac{1}{2}\mathrm{Tr}\big(\sigma^* D^2 g\sigma\big)$.

▷ ...and its time discretization scheme with step $\Delta_n = \frac{T}{n}$ recursively defined by

$$\begin{aligned}
\bar Y_{t_n^n} &= h(\bar X_{t_n^n}),\\
\bar Y_{t_k^n} &= \mathbb{E}(\bar Y_{t_{k+1}^n}|\mathcal{F}_{t_k^n}) + \Delta_n f\big(t_k^n, \bar X_{t_k^n}, \mathbb{E}(\bar Y_{t_{k+1}^n}|\mathcal{F}_{t_k^n}), \bar\zeta_{t_k^n}\big),\\
\bar\zeta_{t_k^n} &= \frac{1}{\Delta_n}\mathbb{E}\big(\bar Y_{t_{k+1}^n}(W_{t_{k+1}^n} - W_{t_k^n})|\mathcal{F}_{t_k}\big) = \frac{1}{\Delta_n}\mathbb{E}\big((\bar Y_{t_{k+1}^n} - \bar Y_{t_k^n})(W_{t_{k+1}^n} - W_{t_k^n})|\mathcal{F}_{t_k}\big)
\end{aligned}$$

where $\bar X$ is the Euler scheme of X defined by

$$\bar X_{t_{k+1}^n} = \bar X_{t_k^n} + b(_k^n, \bar X_{t_k^n})\Delta_n + \sigma(_k^n, \bar X_{t_k^n})(W_{t_{k+1}^n} - W_{t_k^n}).$$

▷ . . . spatially discretized by quantization:

$$\begin{aligned} \widehat{Y}_n &= h(\widehat{X}_n) \\ \widehat{Y}_k &= \widehat{\mathbb{E}}_k(\widehat{Y}_{k+1}) + \Delta_n f_k\big(\widehat{X}_k, \widehat{\mathbb{E}}_k(\widehat{Y}_{k+1}), \widehat{\zeta}_k\big) \\ \text{with} \qquad \widehat{\zeta}_k &= \frac{1}{\Delta_n}\widehat{\mathbb{E}}_k(\widehat{Y}_{k+1}(W_{t^n_{k+1}} - W_{t^n_k})) \end{aligned}$$

where $\widehat{\mathbb{E}}_k = \mathbb{E}(\cdot|\widehat{X}_k)$.

▷ A Quantization tree: $N = N_0 + \cdots + N_n$, N_k = size of layer t^n_k.

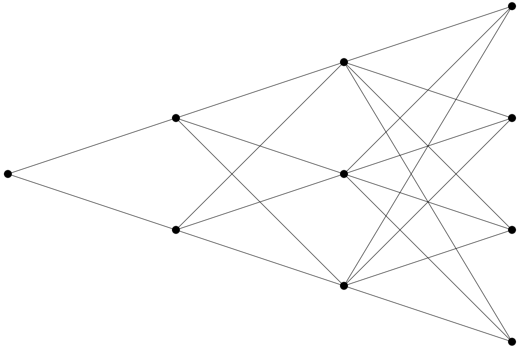


Figure: A typical 1-dimensional quantization tree

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

Theorem (A priori error estimates (Sagna-P., 2014), (P.,Wilbertz, 2012))

Suppose that all the "Lipschitz" assumptions on b, σ, f, h are fulfilled.

(a) "Price": *Then, for every $k = 0, \ldots, n$,*

$$\big\|\bar{Y}_{t_k^n} - \widehat{Y}_k\big\|_2^2 \le [f]_{\rm Lip}^2 \sum_{i=k}^{n} e^{(1+[f]_{\rm Lip})t_i^n} K_i(b,\sigma,T,f,h)\big\|\bar{X}_{t_i^n} - \widehat{X}_{t_i^n}\big\|_2^2 = O\left(\frac{n}{N^{\frac{2}{d}}}\right),$$

(b) "Hedge":

$$\sum_{k=0}^{n-1} \Delta_n \big\|\bar{\zeta}_{t_k^n} - \widehat{\zeta}_k\big\|_2^2 \le \sum_{k=0}^{n-1} e^{(1+[f]_{\rm Lip})t_k^n}\big\|Y_{t_{k+1}^n} - \widehat{Y}_{t_{k+1}^n}\big\|_2^2 + K_k(b,\sigma,T,f,h)\big\|X_{t_k^n} - \widehat{X}_{t_k^n}\big\|_2^2$$

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

Theorem (A priori error estimates (Sagna-P., 2014), (P.,Wilbertz, 2012))

Suppose that all the "Lipschitz" assumptions on b, σ, f, h are fulfilled.

(*a*) "Price": *Then, for every* $k = 0, \dots, n$,

$$\big\|\bar{Y}_{t_k^n} - \widehat{Y}_k\big\|_2^2 \le [f]_{\rm Lip}^2 \sum_{i=k}^{n} e^{(1+[f]_{\rm Lip})t_i^n} K_i(b,\sigma,T,f,h)\big\|\bar{X}_{t_i^n} - \widehat{X}_{t_i^n}\big\|_2^2 = O\left(\frac{n}{N^{\frac{2}{d}}}\right),$$

(*b*) "Hedge":

$$\sum_{k=0}^{n-1} \Delta_n\big\|\bar{\zeta}_{t_k^n} - \widehat{\zeta}_k\big\|_2^2 \le \sum_{k=0}^{n-1} e^{(1+[f]_{\rm Lip})t_k^n}\big\|Y_{t_{k+1}^n} - \widehat{Y}_{t_{k+1}^n}\big\|_2^2 + K_k(b,\sigma,T,f,h)\big\|X_{t_k^n} - \widehat{X}_{t_k^n}\big\|_2^2$$

(*c*) "RBSDE": *The same error bounds hold with Reflected BSDE (so far without Z in f) by replacing h by $h_k = h(t_k^n, .)$ where $h(t, X_t)$ is the obstacle process in the resulting quantized scheme.*

What is new (compared to Bally-P. 2003 for *reflected BSDE*)?

- $+$: Z in f for quantization error bounds.
- $+$: The square everywhere

How to reduce the (quadratic) quantization error $\|X - \widehat{X}\|_2$?

or to be more precise

Given a (finite) grid $\Gamma = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^d$, how to solve

$$\inf_{q:\mathbb{R}^d \to \Gamma} \|X - q(X)\|_2 \ ?$$

Voronoi Quantization

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

$$\mathbb{E}|X|^p < +\infty \qquad \text{for some } p \in (0, \infty).$$

Voronoi Quantization

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

$$\mathbb{E}|X|^p < +\infty \qquad \text{for some } p \in (0, \infty).$$

▷ Given a (finite) "grid" $\Gamma = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^d$ and $q : \mathbb{R}^d \to \Gamma$

$$\|X - \underbrace{q(X)}_{\in\Gamma}\|_p \geq \|\mathrm{dist}(X, \Gamma)\|_p.$$

This suggests to discretize of the random vector X using a Nearest Neighbor projection as a quantization function q.

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

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

- Let $q := \mathrm{Proj}_\Gamma : \mathbb{R}^d \to \Gamma$ be the induced Nearest Neighbor projection,

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

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

▷ We define the *Voronoi Quantization* of the random vector X as

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

▷ This is a purely geometric optimization only depending on the norm.

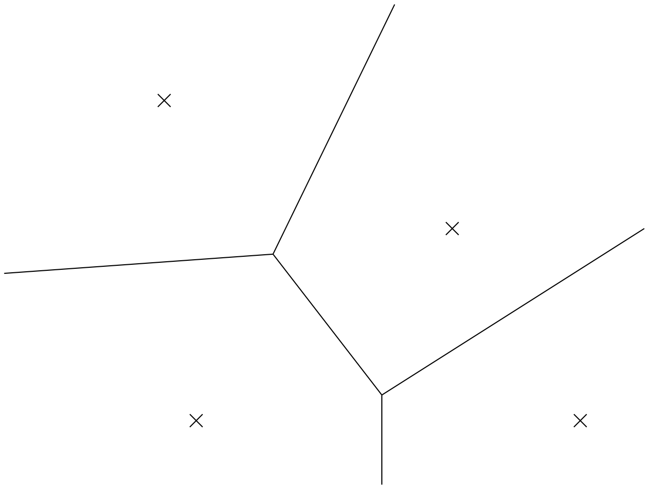
▷ The *L^p-mean quantization error induced by a grid* Γ ($p \in (0, +\infty)$) induced by a grid $\Gamma \subset \mathbb{R}^d$ with size $|\Gamma| \leq N$, $N \in \mathbb{N}$

Definition (*L^p-mean quantization error*)

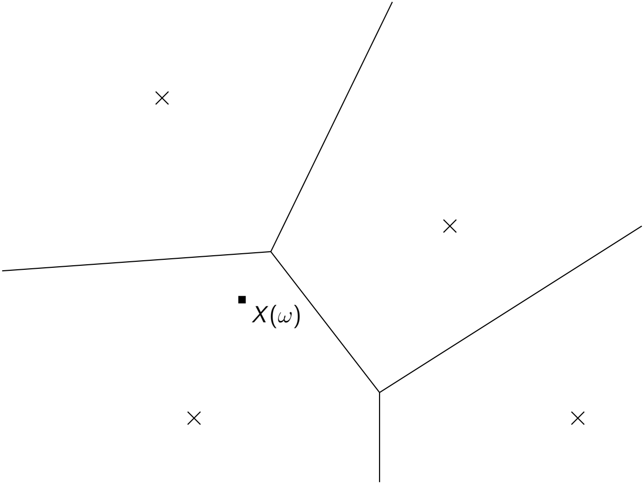
$$e_p(X;\Gamma) = \|X - \widehat{X}^{\Gamma}\|_p = \|\mathrm{dist}(X,\Gamma)\|_p = \Big\| \min_{x \in \Gamma} |X - x| \Big\|_p.$$

Voronoi Quantization

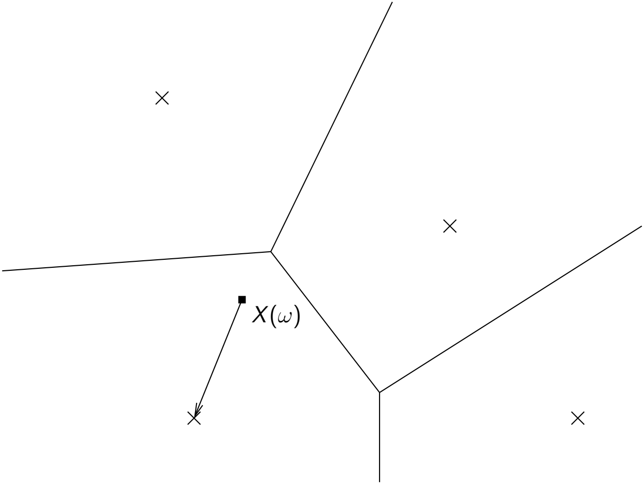
Voronoi Quantization



Voronoi Quantization



Voronoi Quantization



Optimal L^p-mean quantization problem

▷ Second idea to minimize the quantization error

Optimally fit the grid Γ to (the distribution of $\mathbb{P}_X$) of X for a given "complexity".

▷ It amounts to solve the optimal L^p-mean quantization problem at level N, $N \geq 1$.

Definition (*Optimal L^p-mean quantization error at level N*)

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

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

▷ One shows the more general optimality result

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

i.e. no possible alternative than quantization

Theorem (Kieffer, Cuesta-Albertos, P., Graf-Luschgy, 1988 $\longrightarrow$ 2000)

(a) EXISTENCE AN OPTIMAL QUANTIZER: *For every level $N \geq 1$, there exists (at least) one L^p-optimal quantization grid $\Gamma^{N,p}$ at level N.*

(b) *If $p = 2$,* STATIONARITY PROPERTY/SELF-CONSISTENCY: $\mathbb{E}\Big(X \,|\, \widehat{X}^{\Gamma^{N,2}}\Big) = \widehat{X}^{\Gamma^{N,2}}$.

Theorem (Kieffer, Cuesta-Albertos, P., Graf-Luschgy, 1988 → 2000)

(a) EXISTENCE AN OPTIMAL QUANTIZER: *For every level $N \geq 1$, there exists (at least) one L^p-optimal quantization grid $\Gamma^{N,p}$ at level N.*

(b) *If $p = 2$,* STATIONARITY PROPERTY/SELF-CONSISTENCY: $\mathbb{E}\Big(X \,|\, \widehat{X}^{\Gamma^{N,2}}\Big) = \widehat{X}^{\Gamma^{N,2}}$.

▷ One checks that

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

Let $(z_N)_{N\geq 1}$ be an everywhere dense sequence in $\mathbb{R}^d$

$$e_{p,N}(X) \leq e_p\big(X, \{z_1, \ldots, z_N\}\big) \downarrow 0 \quad N \to +\infty.$$

by the Lebesgue dominated convergence theorem.

▷ But. . . at which rate ?

Rates of Optimal Quantization

Theorem (Zador's Theorem)

(a) SHARP ASYMPTOTIC *(Zador, Kieffer, Bucklew & Wise, Graf & Luschgy (2000)):*

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 *(Pierce, Luschgy-P. (2006)):*

Let $p' > p$. *There exists* $C_{p,p',d} \in (0,+\infty)$ *such that, for every* $\mathbb{R}^d$*-valued* X *r.v.*

$$\forall\, N \geq 1, \quad e_{p,N}(X) \leq C_{p,p',d}\, \sigma_{p'}(X).\, N^{-\frac{1}{d}}.$$

Remarks. • $\sigma_{p'}(X) := \inf_{a\in\mathbb{R}^d} \|X-a\|_{p'} \leq +\infty$ is the $L^{p'}$-(pseudo-)standard deviation.

• $N^{\frac{1}{d}}$ is known as the *curse of dimensionality*.

Numerical computation of quantizers

▷ (Nearly) optimal grids can be computed by optimization algorithms :

- When $d = 1$ (or even $d = 2$): deterministic Newton-Raphson like methods
- In higher dimension ($d \geq 2$ or 3), stochastic optimization methods based on a stochastic gradient approach:
  - Competitive Learning Vector Quantization algorithm (*CLVQ*),
  - (Fixed point) randomized Lloyd's I procedure.

  both based on Monte Carlo simulations.

In fact, Optmial quantization appears as

Compressed Monte Carlo method

On the numerical computation optimal quantizers ($p = 2$)

Computing optimal grids: the quadratic distorsion ($p = 2$)

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

On the numerical computation optimal quantizers ($p = 2$)

Computing optimal grids: the quadratic distorsion ($p = 2$)

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

$$\min_{x \in (\mathbb{R}^d)^N} D_N(x) \Longleftrightarrow \min_{|\Gamma| \le N} e_2(X; \Gamma)$$

On the numerical computation optimal quantizers ($p = 2$)

Computing optimal grids: the quadratic distorsion ($p = 2$)

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

$$\min_{x\in(\mathbb{R}^d)^N} D_N(x) \Longleftrightarrow \min_{|\Gamma|\le N} e_2(X;\Gamma)$$

▷ If $\|\cdot\|$ is the canonical Euclidean norm on $\mathbb{R}^d$ and x has distincts components

$$\nabla D_N(x) = \frac{1}{2}\left(\mathbb{E}\Big[(x_i - X)\mathbf{1}_{\{X\in C_i(x)\}}\Big]\right)_{1\le i\le N}$$

On the numerical computation optimal quantizers ($p = 2$)

Computing optimal grids: the quadratic distorsion ($p = 2$)

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

$$\min_{x\in(\mathbb{R}^d)^N} D_N(x) \iff \min_{|\Gamma|\le N} e_2(X;\Gamma)$$

▷ If $\|\cdot\|$ is the canonical Euclidean norm on $\mathbb{R}^d$ and x has distincts components

$$\nabla D_N(x) = \frac{1}{2}\left(\mathbb{E}\Big[(x_i - X)\mathbf{1}_{\{X\in C_i(x)\}}\Big]\right)_{1\le i\le N}$$

▷ The grid $\Gamma^{*,N} = \{x_1^*, \dots x_N^*\}$ is L^2-optimal iff $x^{*,N} \in \mathrm{argmin} D_N$.

Hence

$$\Gamma^{*,N} \text{ is } L^2\text{-optimal} \iff \nabla D_N\big(x^{*,N}\big) = 0.$$

On the numerical computation optimal quantizers ($p = 2$)

Computing optimal grids: the quadratic distorsion ($p = 2$)

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

$$\min_{x \in (\mathbb{R}^d)^N} D_N(x) \iff \min_{|\Gamma| \le N} e_2(X; \Gamma)$$

▷ If $\|\cdot\|$ is the canonical Euclidean norm on $\mathbb{R}^d$ and x has distincts components

$$\nabla D_N(x) = \frac{1}{2}\left(\mathbb{E}\Big[(x_i - X)\mathbf{1}_{\{X \in C_i(x)\}}\Big]\right)_{1 \le i \le N}$$

▷ The grid $\Gamma^{*,N} = \{x_1^*, \ldots x_N^*\}$ is L^2-optimal iff $x^{*,N} \in \operatorname{argmin} D_N$.

Hence

$$\Gamma^{*,N} \text{ is } L^2\text{-optimal } \iff \nabla D_N\big(x^{*,N}\big) = 0.$$

▷ Connection critical point and stationarity:

$$\nabla D_N(x) = 0 \quad \iff \quad x_i = \frac{\mathbb{E}\Big(X\mathbf{1}_{\{X \in C_i(x)\}}\Big)}{\mathbb{P}\big(X \in C_i(x)\big)}, \; i = 1, \ldots, N$$

$$\iff \quad \widehat{X}^x = \mathbb{E}\big(X \,|\, \widehat{X}^x\big)$$

$d = 1$ $\mu = \mathbb{P}_X$, $C_i(x_1, \ldots, x_N) = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, $i = 1, \ldots, N$,.

$$x_i = \frac{\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \xi \mathbb{P}_X(d\xi)}{\mu([x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]}, i = 1, \ldots, N$$

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

$d \geq 2$

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: $i^* \in \text{argmin}_i |x_i^n - \xi_n|$

     Learning: $\begin{cases} x_{i^*}^{n+1} := x_{i^*}^n + \gamma_n(x_{i^*}^n - \xi_n) \\ x_j^{n+1} := x_j^n, \qquad \text{for } j \neq i^*. \end{cases}$

$d = 1$ $\mu = \mathbb{P}_X,\ C_i(x_1,\dots,x_N) = [x_{i-\frac12}, x_{i+\frac12}],\ i = 1,\dots,N,.$

$$x_i = \frac{\int_{x_{i-\frac12}}^{x_{i+\frac12}} \xi \mathbb{P}_X(d\xi)}{\mu([x_{i-\frac12}, x_{i+\frac12}])},\ i = 1,\dots,N$$

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

$d \geq 2$

1. Stochastic Gradient Method: *CLVQ*
   - Simulate $\xi_1, \xi_2, \dots$ independent copies of X
   - Generate step sequence $\gamma_1, \gamma_2, \dots$
     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: $i^* \in \mathrm{argmin}_i |x_i^n - \xi_n|$

     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 fixed-point procedure.
   - Initial grid $\Gamma^{(0)} = \{x_1^0, \dots, x_N^0\}$
   - Grid updating $n \mapsto n+1$:

   (i) $x_k^{(n+1)} = \mathbb{E}(X \mid \widehat{X}^{\Gamma^{(n)}} = x_k^{(n)})$

   (ii) $\Gamma^{(n+1)} = \{x_k^{(n+1)},\ k = 1:N\}$ and $\widehat{X}^{\Gamma^{(n+1)}} = \mathrm{Proj}_{\Gamma^{(n+1)}}(X)$,
   - so that $\|X - \widehat{X}^{\Gamma^{(n+1)}}\|_2 \leq \|X - \widehat{X}^{\Gamma^{(n)}}\|_2$

$d = 1$ $\mu = \mathbb{P}_X$, $C_i(x_1,\ldots,x_N) = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, $i = 1,\ldots,N$,.

$$x_i = \frac{\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \xi \mathbb{P}_X(d\xi)}{\mu([x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}])}, i = 1,\ldots,N$$

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

$d \geq 2$

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: $i^* \in \text{argmin}_i |x_i^n - \xi_n|$

     Learning: $\begin{cases} x_{i^*}^{n+1} := x_{i^*}^n + \gamma_n(x_{i^*}^n - \xi_n) \\ x_j^{n+1} := x_j^n, \quad \text{for } j \neq i^*. \end{cases}$

2. LLOYD's algorithm as a randomized fixed-point procedure.
   - Initial grid $\Gamma^{(0)} = \{x_1^0, \ldots, x_N^0\}$
   - Grid updating $n \mapsto n+1$:

   (i) $x_k^{(n+1)} = \mathbb{E}(X \mid \widehat{X}^{\Gamma^{(n)}} = x_k^{(n)})$

   (ii) $\Gamma^{(n+1)} = \{x_k^{(n+1)}, \ k = 1:N\}$ and $\widehat{X}^{\Gamma^{(n+1)}} = \text{Proj}_{\Gamma^{(n+1)}}(X)$,

   - so that $\|X - \widehat{X}^{\Gamma^{(n+1)}}\|_2 \leq \|X - \widehat{X}^{\Gamma^{(n)}}\|_2$

3. "Batch" approach [. . .]

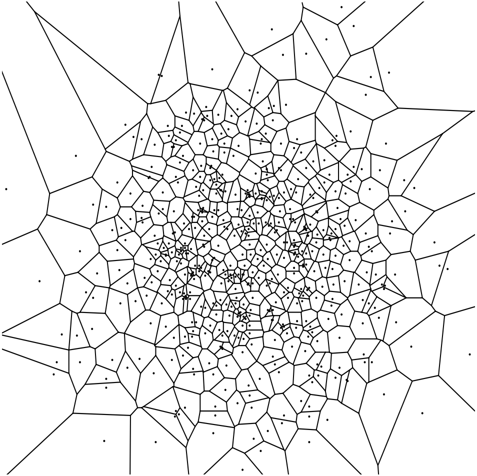


Figure: Two N-quantizers related to $\mathcal{N}(0; I_2)$ of size $N = 500\ldots$

(with J. Printems)

Before. . .

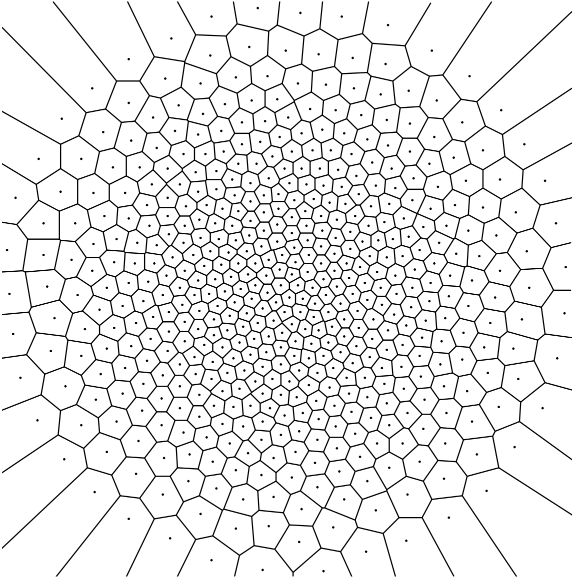


Figure: ... After

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

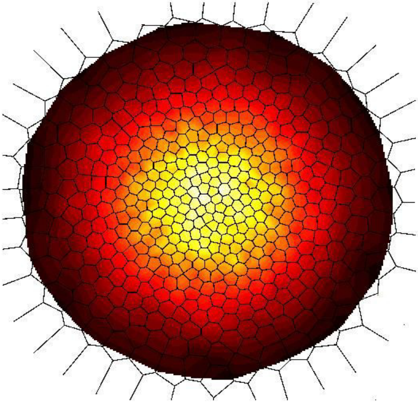


Figure: An N-quantization of $X \sim \mathcal{N}(0; I_2)$ with coloured weights: $\mathbb{P}\big(X \in C_a(\Gamma)\big)$, $a \in \Gamma$. (with J.Printems)

▷ Weights: $\mathbb{P}\big(X \in C_a(\Gamma)\big) \approx C^{st}\,\dfrac{f_X^{\frac{d}{d+2}}(a)}{N^{\frac{1}{d}}}$ (when N is large).

▷ Local inertia: $a \longmapsto \mathbb{E}|X-a|^2 \mathbf{1}_{\{X \in C_a(\Gamma)\}} \approx \dfrac{e_n(\Gamma, X)}{N}$ (for fixed N).

As a result for Gaussian vectors...

▷ Instant search for the unique optimal quantizer using a Newton-Raphson descent on $\mathbb{R}^N$ ... with an arbitrary accuracy.

▷ For $\mathcal{N}(0;1)$ and $N = 1, \ldots, 500$, tabulation within 10^{-14} accuracy of optimal N-quantizers and textcolorbluecompanion parameters:

$$\alpha^{(N)} = (\alpha_1^{(N)}, \ldots, \alpha_N^{(N)})$$

and

$$\mathbb{P}(X \in C_i(\alpha^{(N)})),\ i = 1, \ldots N, \qquad \text{and} \qquad \|X - \widehat{X}^{\alpha^{(N)}}\|_2.$$

▷ For $d = 1$ up to 10? Also available for Gaussian $\mathcal{N}(0, I_d)$ $(1 \leq N \leq 4\,000)$.

Download at our WEBSITE :

www.quantize.maths-fi.com

Further Error Estimates

Proposition (First order)

If $\Gamma^{N,*}$ is L^1-optimal at level $N \geq 1$

$$e_{1,N}(X) = \mathbb{E}\|X - \widehat{X}^{\Gamma^{N,*}}\| = \sup_{[F]_{Lip}\leq 1} |\mathbb{E}F(X) - \mathbb{E}\,F(\widehat{X}^{\Gamma^{N,*}})|$$

$$\inf\Big\{\sup_{[F]_{Lip}\leq 1} |\mathbb{E}F(X) - \mathbb{E}\,F(Y)|,\ \mathrm{card}(Y(\Omega)) \leq N\Big\}$$

$= L^1$*-Wasserstein distance between* $\mathcal{L}(X)$ *and the set* $\mathcal{P}_N$.

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

Further Error Estimates

Proposition (First order)

If $\Gamma^{N,*}$ is L^1-optimal at level $N \geq 1$

$$e_{1,N}(X) = \mathbb{E}\|X - \widehat{X}^{\Gamma^{N,*}}\| = \sup_{[F]_{Lip}\leq 1} |\mathbb{E}F(X) - \mathbb{E}\,F(\widehat{X}^{\Gamma^{N,*}})|$$

$$\inf\Big\{ \sup_{[F]_{Lip}\leq 1} |\mathbb{E}F(X) - \mathbb{E}\,F(Y)|,\ \mathrm{card}(Y(\Omega)) \leq N \Big\}$$

$$= L^1\text{-Wasserstein distance between } \mathcal{L}(X) \text{ and the set } \mathcal{P}_N.$$

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

Proposition (Second order)

If $F \in C^1_{Lip}$ and the grid Γ is stationary (e.g. because it is L^2-optimal), i.e.

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

then a Taylor expansion yields

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

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

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

Further Applications

What are these applications using optimal quantization grids?

Further Applications

What are these applications using optimal quantization grids?

Further Applications

What are these applications using optimal quantization grids?

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

Further Applications

What are these applications using optimal quantization grids?

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

Further Applications

What are these applications using optimal quantization grids?

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's [Bally-P.-Printems '01, '03 and '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], Control of PDMP [Dufour-de Sapporta 2013].

Further Applications

What are these applications using optimal quantization grids?

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's [Bally-P.-Printems '01, '03 and '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], Control of PDMP [Dufour-de Sapporta 2013].
- Non-linear filtering and stochastic, volatility models [P.-Pham-Printems '05, Pham-Sellami-Runggaldier'06, Sellami '09 &'10, Callegaro-Sagna '10].

Further Applications

What are these applications using optimal quantization grids?

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's [Bally-P.-Printems '01, '03 and '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], Control of PDMP [Dufour-de Sapporta 2013].
- Non-linear filtering and stochastic, volatility models [P.-Pham-Printems '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

What are these applications using optimal quantization grids?

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's [Bally-P.-Printems '01, '03 and '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], Control of PDMP [Dufour-de Sapporta 2013].
- Non-linear filtering and stochastic, volatility models [P.-Pham-Printems '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

What are these applications using optimal quantization grids?

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's [Bally-P.-Printems '01, '03 and '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], Control of PDMP [Dufour-de Sapporta 2013].
- Non-linear filtering and stochastic, volatility models [P.-Pham-Printems '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.

Further Applications

What are these applications using optimal quantization grids?

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's [Bally-P.-Printems '01, '03 and '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], Control of PDMP [Dufour-de Sapporta 2013].
- Non-linear filtering and stochastic, volatility models [P.-Pham-Printems '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.
- Quadratic *BSDE* schemes by Markovian Quantization [Chassagneux-Richou'14].

Further Applications

What are these applications using optimal quantization grids?

- Obstacle Problems: Valuation of Bermuda and American options, Reflected BSDE's [Bally-P.-Printems '01, '03 and '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], Control of PDMP [Dufour-de Sapporta 2013].
- Non-linear filtering and stochastic, volatility models [P.-Pham-Printems '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.
- Quadratic *BSDE* schemes by Markovian Quantization [Chassagneux-Richou'14].
- New error bounds for *BSDE* schemes by quadratic optimal quantization [P.-Sagna '15]

A new result : distortion mismatch/ L^s-rate optimality, $s > p$

▷ Let $\Gamma_N^{(p)}$, $N \geq 1$, be a sequence L^p-optimal grids.

What about $e_s(X, \Gamma_N^p)$ (L^s-mean quantization error) when $X \in L^s_{\mathbb{R}^d}(\mathbb{P})$ for $s > p$?

Theorem (L^p-L^s-distortion mismatch, Graf-Luschgy-P. 2005, Luschgy-P. 2015)

(a) *Let $X \in L^p_{\mathbb{R}^d}(\mathbb{P})$ and let $(\Gamma_N^{(p)})_{N\geq 1}$ be an L^p-optimal sequence for grids. Let $s \in (p, p+d)$. If*

$$X \in L^{\frac{sd}{d+p-s}+\delta}(\mathbb{P}),\ \delta > 0,$$

(note that $\frac{sd}{d+p-s} > s$ and $\lim_{s\to p+d} \frac{sd}{d+p-s} = +\infty$), then

$$\overline{\lim_N}\, N^{\frac{1}{d}} e_s(\Gamma_N^{(p)}, X) < +\infty.$$

(b) *If $\mathbb{P}_X = f(|x|).\lambda_d(d\xi)$ (radial density) then $\delta = 0$ is admissible.*

(c) *If $\mathbb{E}\,|X|^{\frac{sd}{d+p-s}} = +\infty$, then $\underline{\lim}_N\, N^{\frac{1}{d}} e_s(\Gamma_N^{(p)}, X) = +\infty$.*

▷ Possible perspectives: error bounds for quantization based numerical schemes for *BSDE* with a quadratic Z term ?

▷ So far, an application to quantized non-linear filtering.

Application to non-linear filtering

- Signal process $(X_k)_{k\geq 0}$ is an $\mathbb{R}^d$-valued Markov chain.
- The observation process $(Y_k)_{k\geq 0}$ is a sequence of $\mathbb{R}^q$-valued random vectors such that

$$(X_k, Y_k)_{k\geq 0} \text{ is a Markov chain.}$$

- The conditional distribution

$$\mathcal{L}(Y_k \,|\, X_{k-1}, Y_{k-1}, X_k) = g_k(X_{k-1}, Y_{k-1}, X_k, y)\lambda_q(dy)$$

- Aim : compute

$$\Pi_{y_{0:n},n}(dx) = \mathbb{P}(X_k \in dx \,|\, Y_1 = y_1, \cdots, Y_n = y_n)$$

- Kallianpur-Streibel formula: set $y = y_{0:n} = (y_0, \ldots, y_n)$ a vector of observations

$$\Pi_{y,n}(dx) = \Pi_{y,n} f = \frac{\pi_{y,n} f}{\pi_{y,n}\mathbf{1}}$$

with the normalized filter $\pi_{y_{0,n},n}$ defined by

$$\pi_{y_{0:n},n} f = \mathbb{E}(f(X_n) L_{y_{0:n},n}) \quad \text{with} \quad L_{y_{0:n},n} = \prod_{k=1}^{n} g_k(X_{k-1}, y_{k-1}, X_k, y_k),$$

solution to both a forward and a backward inductionsbased on the kernels

$$H_{y,k} h(x) = \mathbb{E}(h(X_k) g_k(x, y_{k-1}, X_k, y_k) | X_{k-1} = x), \quad H_{y,0} f(x) = \mathbb{E}(f(X_0)),$$

- Forward: Start from

$$\pi_{y,0} = H_{y,0}$$

and define by a forward induction

$$\pi_{y,k} f = \pi_{y,k-1} H_{y,k} f, \qquad k = 1, \ldots, n.$$

- Backward: We define by a backward induction

$$\begin{aligned} u_{y,n}(f)(x) &= f(x), \\ u_{y,k-1}(f) &= H_{y,k} u_{y,k}(f), \quad k = 0, \ldots, n. \end{aligned}$$

so that

$$\pi_{y,n} f = u_{y,-1}(f)$$

This formulation is useful in order to establish the quantization error bound.

Quantized Kallianpur-Streibel formula (P.-Pham (2005))

- Quantization of the kernel:

$$H_{y_{0:n},k}f(x) \longrightarrow \widehat{H}_{y_{0:n},k}f(x) = \mathbb{E}(f(\widehat{X}_k)g_k(x, y_{k-1}, \widehat{X}_k, y_k)|\widehat{X}_{k-1} = x)$$

- Forward quantized dynamics (I):

$$\widehat{\pi}_{y,k}f = \widehat{\pi}_{y,k-1}\widehat{H}_{y,k}f, \qquad k = 1, \ldots, n.$$

- Forward quantized dynamics (II):

$$\widehat{\Pi}_y(dx) = \widehat{\Pi}_{y,n}f = \frac{\widehat{\pi}_{y,n}f}{\pi_{y_{0:n},n}\mathbf{1}}$$

(finitely supported unnormalized filter satisfies formally the same recursions)

- Weight computation: If $\widehat{X}_n = \widehat{X}_n^{\Gamma_n}$, $\Gamma_n = \{x_1^1, \ldots, x_{N_n}^n\}$ then

$$\widehat{\Pi}_{y,n}(dx) = \sum_{i=1}^{N_n} \widehat{\Pi}_{y,n}^i \delta_{x_i^n} \quad \text{with } \widehat{\Pi}_{y,n}^i = \widehat{\Pi}_{y,n}\big(\mathbf{1}_{C_i(\Gamma_n)}\big).$$

From *Lip* to θ-Liploc assumptions

- Standard $\mathcal{H}_{\rm Lip}$assumption for the conditional densities $g_k(.,y,.,y')$: bounded by K_g and Lipschitz continuity.

$$|g_k(x,y,x',y') - g_k(\widehat{x},y,\widehat{x}',y')| \quad \le \quad [g_k]_{\rm Lip}(y,y')\big(|x-\widehat{x}|+|x'-\widehat{x}'|\big).$$

- The kernels $P_k(x,d\xi)=\mathbb{P}(X_k\in d\xi\,|\,X_{k-1}=x)$ propagate Lipschitz continuity with coefficient $[P_k]_{\rm Lip}s$ such that

$$\max_{k=1,\ldots,n}[P_k]_{\rm Lip}<+\infty$$

Aim: Switch to a θ-local Lipschitz assumption ($\theta:\mathbb{R}^d\to\mathbb{R}_+$, $\uparrow+\infty$ as $|x|\uparrow+\infty$).

$$|h(x,x')-h(\hat{x},\hat{x}')|\le[h]_{\rm loc}\big(|x-\widehat{x}|+|x'-\widehat{x}'|\big)\big(1+\theta(x)+\theta(x')+\theta(\hat{x})+\theta(\hat{x}')\big)$$

- New $(\mathcal{H}^{\theta}_{\rm Liploc})$ assumption: the functions g_k are still bounded by K_g and θ-local Lipschitz continuous

$$|g_k(x,y,x',y')-g_k(\widehat{x},y,\widehat{x}',y')|\le[g_k]_{\rm loc}(y,y')\big(|x-\widehat{x}|+|x'-\widehat{x}'|\big)\big(1+\theta(x)+\theta(x')+\theta(\hat{x})+\theta(\hat{x}')\big)$$

- The kernels $P_k(x,d\xi)=\mathbb{P}(X_k\in d\xi\,|\,X_{k-1}=x)$ propagate θ-local Lipschitz continuity with coefficient $[P_k]_{\rm loc}<+\infty$.
- The kernels $P_k(x,d\xi)$ propagate θ-control: $\max_{0\le k\le n-1}P_k(\theta)(x)\le C\big(1+\theta(x)\big)$.

Typical example: $X_k=\bar{X}^n_{t^n_k}$ (Euler scheme with step $\Delta_n=\frac{T}{n}$), $\theta(\xi)=|\xi|^{\alpha}$, $\alpha>0$.

Theorem

Let $s \in (1, 1+\frac{d}{2})$ *and* $\theta(x) = |x|^\alpha$, $\alpha \in (0, \frac{1}{\frac{1}{s-1}-\frac{2}{d}})$.
Assume (X_k) *and* (g_k) *satisfy* $(\mathcal{H}^{\theta}_{\rm Liploc})$ *(in particular* (X_k) *propagates* θ*-Lipschitz continuity) and assume* $X_k \in L^{\frac{2ds}{d+2-2s}}$, $k = 0, \ldots, n$. *Then*

$$|\Pi_{y,n}f - \widehat{\Pi}_{y,n}f|^2 \le \frac{2(K_g^n)^2}{\phi_n^2(y) \vee \widehat{\phi}_n^2(y)} \sum_{k=0}^{n} B_k^n(f,y) \times \underbrace{\|X_k - \widehat{X}_k\|_{2s}^2}_{\asymp \|X_k-\widehat{X}_k\|_2^2 \le c_k N_k^{-\frac{2}{d}} \text{ (Mismatch!!)}} \tag{1}$$

with

$$\phi_n(y) = \pi_{y,n}\mathbf{1} \quad \textit{and} \quad \widehat{\phi}_n(y) = \widehat{\pi}_{y,n}\mathbf{1},$$

$$B_k^n(f,y) := 2[P]_{\rm loc}^{2(n-k)}[f]_{\rm loc}^2 + 2\|f\|_\infty^2 R_{n,k} + \|f\|_\infty R_{n,k}^2,$$

where

$$R_{n,k} = \frac{8^{\frac{s}{s-1}} M_s^n}{K_g^2}\Big[[g_{k+1}]_{\rm loc}^2 + [g_k]_{\rm loc}^2 + \Big(\sum_{m=1}^{n-k}[P]_{\rm loc}^{m-1}(1+[P]_{\rm loc})[g_{k+m}]_{\rm loc}\Big)^2\Big],$$

and

$$M_s^n := 2 \max_{k=0,\ldots,n}\big(\mathbb{E}\big(\theta(X_k)^{\frac{2s}{s-1}}\big) + \mathbb{E}\big(\theta(\widehat{X}_k)^{\frac{2s}{s-1}}\big)\big).$$

Extensions

- Greedy quantization (Luschgy-P., JAT, 2014): sequence $(a_N)_{N\geq 1}$ such that

$$\{a_1,\ldots,a_N\},\ N\geq 1,\ \text{ is } L^p\text{-rate optimal}$$

to spare RAM.

$$a_{N+1}=\mathrm{argmin}_{\xi\in\mathbb{R}^d}\, e_p(\{a_1,\ldots,a_N\}\cup\{\xi\},X)$$

Numerical schemes can be successfully implemented with this quantization.

- Fast recursive quantization (in progress) in medium dimension (Sagna-P. , 2014)

Numerical illustrations

- Risk-neutral price under historical probability (B&S model, Euler scheme)

$$dY_t = \left(rY_t + \frac{\mu - r}{\sigma} Z\right) dt + Z_t dW_t$$

with

$$Y_T = h(X_T) = (X_T - K)_+.$$

▷ Model parameters: $r = 0.1$; $T = 0.1$; $\sigma = 0.25$; $S_0 = K = 100$.

▷ Quantization tree calibration: $7.5\,10^5$ MC and $NbLloyd = 1$.

▷ Reference $\mathrm{call}_{BS}(K, T) = 3.66$, $Z_0 = 14.148$. If $\mu \in \{0.05, 0.1, 0.15, 0.2\}$,

- $n = 10$ and $N_k = \bar{N} = 20$: Q-price $= 3.65$, $\widehat{Z}_0 = 14.06$.
- $n = 10$ and $N_k = \bar{N} = 40$, Q-price $= 3.66$, $\widehat{Z}_0 = 14.08$.

▷ Computation time :

– 5 seconds for one contract.

– Additional contracts for free (more than $10^5/s$).

▷ Romberg extrapolation price $= 2 * Q$-price(N_2)-Q-price(N_1) does improve the price (and the "hedge").

Numerical illustrations

- Bid-ask spreads on interest rates :

$$dY_t = \left(rY_t + \frac{\mu - r}{\sigma} Z_t + (R - r)\min\left(Y_t - \frac{Z_t}{\sigma}, 0\right)\right) dt + Z_t dW_t$$

with

$$Y_T = h(X_T) = (X_T - K_1)_+ - 2(X_T - K_2)_+, \quad K_1 = 95,\ K_2 = 105.$$

$$\mu = 0.05, r = 0.01,\ \sigma = 0.2,\ T = 0.25,\ R = 0.06$$

▷ Reference price $= 2.978$, $\widehat{Z}_0 = 0.553$.

▷ Quantized prices:

- $n = 10$ and $N_k = \bar{N} = 20$: Q-price $= 2.96$, $\widehat{Z}_0 = 0.515$.
- $n = 10$ and $N_k = \bar{N} = 40$, Q-price $= 2.97$, $\widehat{Z}_0 = 0.531$.

▷ Romberg extrapolation price $= 2 * Q$-price(N_2)-Q-price$(N_1) \approx 2.98$ and Romberg $\widehat{Z}_0 \approx 0.547$.
Comparable results though slightly less precise, due to the non linearity in Z compensated by the difference of convex functions. . . .

Multidimensional example

▷ Due to J.-F. Chassagneux: W d-dimensional B.M.

$$dX_t = dW_t, \qquad -dY_t = f(t, Y_t, Z_t)dt - Z_t \cdot dW_t$$

with $f(t,y,z) = (z_1 + \ldots + z_d)\big(y - \frac{2+d}{2d}\big)$. ▷ Solution :

$$Y_t = \frac{e_t}{1+e_t}, \qquad Z_t = \frac{e_t}{(1+e_t)^2} \text{ with } e_t = \exp(x_1 + \ldots + x_d + t).$$

We set $t = 0.5$, $d = 2, 3$, so that $Y_0 = 0.5$ and $Z_0^i = 0.24$, for every $i = 1, \ldots, d$.

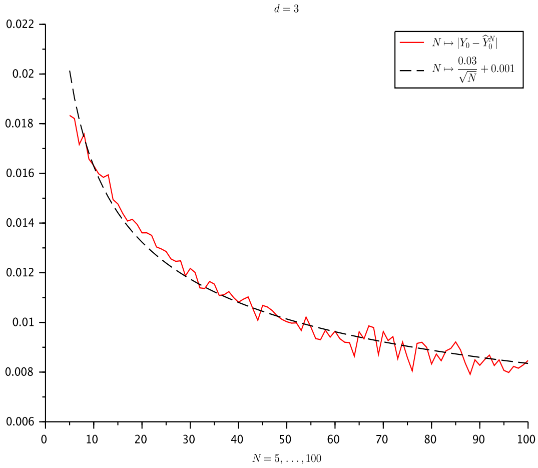


Figure: Convergence rate of the quantization error for the multidimensional example). Abscissa axis: the size $N = 5, \dots, 100$ of the quantization. Ordinate axis: The error $|Y_0 - \widehat{Y}_0^N|$ and the graph $N \mapsto \hat{a}/N + \hat{b}$, where $\hat{a}$ and $\hat{b}$ are the regression coefficients. $d = 3$.

Bon, Vlad on s'y remet quand?

Faut profiter tant qu'on est jeunes !!