Rate allocation in quantization

Steffen Dereich

Institut für Mathematik
TU Berlin
http://www.math.tu-berlin.de/~dereich/

Workshop on Numerics and Stochastics

August 25, 2008

Outline of the talk

I Introduction

- Quantization
- Applications
- Coarse descriptions of good quantizations: *rate allocation problems*

II Hilbert space-valued Gaussian originals

- Asymptotic formulae
- Kolmogorov's inverse water filling principle

III Diffusions

- Preliminaries - Quantization of the Wiener process
- Asymptotic formulae
- Rate allocation for diffusions
- Constructive quantization

IV Lévy processes

- Asymptotic formulae
- Rate allocation

I Quantization

Let
- $(\mathfrak{X}, \|\cdot\|)$ be a separable Banach space (e.g. $\mathfrak{X} = C[0,1]$)
- X $\mathfrak{X}$-valued random vector (e.g. X sol. to SDE)
- μ distribution of X

Quantization error of rate $r \geq 0$ and order $s > 0$

$$D^{(s)}(r) = \inf\{\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} : \hat{X} \text{ r.v. with } \#\mathrm{range}(\hat{X}) \leq e^r\}$$

I Quantization

Let
- $(\mathfrak{X}, \|\cdot\|)$ be a separable Banach space (e.g. $\mathfrak{X} = C[0,1]$)
- X $\mathfrak{X}$-valued random vector (e.g. X sol. to SDE)
- μ distribution of X

Quantization error of rate $r \geq 0$ and order $s > 0$

$$D^{(s)}(r) = \inf\{\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} : \hat{X} \text{ r.v. with } \#\mathrm{range}(\hat{X}) \leq e^r\}$$

Coding: minimal coding error when coding with a fixed number of elements (studied since the 1940s)

Aim:
- find good *codebook* $\mathcal{C}$ with $\#\mathcal{C} \leq e^r$
- find fast projection $\pi : \mathfrak{X} \to \mathcal{C}$!

Then: Approximate X by $\hat{X} = \pi(X)$.

I Example

Problem: A number of service centers shall be opened in a city !

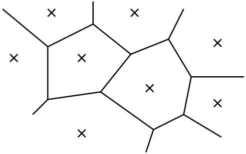
Wants: Potential customers shall be close to service centers.

Aim: Solve

$$\mathbb{E}[\min_{x \in \mathcal{C}} \|X - x\|^s]^{1/s} = \min!$$

where

- minimum is taken over $\mathcal{C} \subset \mathbb{R}^2$ with $\#\mathcal{C} \leq n$ (possible service center locations)
- X denotes the random location of a typical demand.



I Quadrature and quantization

Quadrature formulas

$$D^{(1)}(r) = \inf\left\{ \sup_{f \in \mathrm{Lip}(1)} \left| \int f \, d\mu - \int f \, d\widetilde{\mu} \right| : \#\mathrm{range}(\widetilde{\mu}) \le e^r \right\}$$

is the worst case error for $\mathrm{Lip}(1)$-quadrature
(Kantorovich, Rubinstein '58)

Aim: Construct $\widetilde{\mu}$ (supporting points and weights)

I Quadrature and quantization

Quadrature formulas

$$D^{(1)}(r) = \inf\Big\{ \sup_{f\in\mathrm{Lip}(1)} \Big| \int f\,d\mu - \int f\,d\widetilde{\mu} \Big| : \#\mathrm{range}(\widetilde{\mu}) \le e^r \Big\}$$

is the worst case error for $\mathrm{Lip}(1)$-quadrature
(Kantorovich, Rubinstein '58)

Aim: Construct $\widetilde{\mu}$ (supporting points and weights)

Further applications:

- Variance reduction (Pagès, ...)
- Worst case error analysis of stochastic algorithms (CDMR '08)

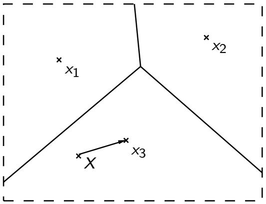
I Variance reduction

Approach: Partition $\mathfrak{X}$ into $\mathfrak{X} = \dot{\bigcup}_j V_j$, choose $\pi : E \to E$ mapping each cell V_j on a single element, and write

$$\mathbb{E}\big[f(X)\big] = \mathbb{E}\big[f(\pi(X))\big] + \mathbb{E}\big[f(X) - f(\pi(X))\big] = \mathrm{I} + \mathrm{II}$$

- $\mathrm{I} = \sum_x \mathbb{P}(\pi(X) = x)\, f(x)$
- II is approximated by Monte Carlo; note that

$$\mathrm{var}(f(X) - f(\pi(X))) \le \mathrm{Lip}(f)^2\, \mathbb{E}\big\|X - \pi(X)\big\|^2$$



I Optimal point density for $\mathfrak{X} = \mathbb{R}^d$

In general, it is hard to construct (close to) optimal quantizations, e.g., by competitive learning vector quantization algorithm (Bouton, Pagès '97).

Paradigm: Find an intermediate convex optimization problem that coarsely describes good quantizations!

I Optimal point density for $\mathfrak{X} = \mathbb{R}^d$

In general, it is hard to construct (close to) optimal quantizations, e.g., by competitive learning vector quantization algorithm (Bouton, Pagès '97).

Paradigm: Find an intermediate convex optimization problem that coarsely describes good quantizations!

Example:
- $\mathfrak{X} = (\mathbb{R}^d, \|\cdot\|)$
- μ possesses a density h
- $\mathbb{E}\|X\|^{s+\varepsilon} < \infty$ for some $\varepsilon > 0$

Then the empirical measures $\nu_n = \frac{1}{n}\sum_{x\in\mathcal{C}(n)} \delta_x$ for optimal codebooks $\mathcal{C}(n)$ of size n converge to a measure ν that has density

$$\xi = \frac{1}{Z} h^{d/(d+s)}.$$

It is the solution to

$$\int_{\mathbb{R}^d} \xi(x)^{-s/d}\, h(x)\, dx = \min$$

under the constraint $\int \xi(x)\, dx = 1$ (Bucklew '84, Graf, Luschgy '00).

II Gaussian signals in Hilbert spaces

Let
- $\mathfrak{X}$ be a Hilbert space
- μ be a centered Gaussian measure

Use *Karhunen-Loève expansion*

$$X = \sum_j \sqrt{\lambda_j} X_j e_j$$

where
- (λ_j) $\mathbb{R}_+$-valued sequence of eigenvalues
- (e_j) orthonormal system of eigenvectors
- (X_j) i.i.d. standard normals

II Gaussian signals in Hilbert spaces

Let • $\mathfrak{X}$ be a Hilbert space
• μ be a centered Gaussian measure

Use *Karhunen-Loève expansion*

$$X = \sum_j \sqrt{\lambda_j} X_j e_j$$

where
- (λ_j) $\mathbb{R}_+$-valued sequence of eigenvalues
- (e_j) orthonormal system of eigenvectors
- (X_j) i.i.d. standard normals

Example: (Wienerprocess in $L^2[0,1]$)

$$\lambda_j = \frac{1}{(j-1/2)^2\pi^2} \quad \text{and} \quad e_j(t) = \sqrt{2}\,\sin((j-1/2)\pi t)$$

II Shannon's distortion rate function

Quantization error

$$D^{(s)}(r) = \inf\{\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} : \hat{X} \text{ r.v. with } \#\mathrm{range}(\hat{X}) \leq e^r\}$$

II Shannon's distortion rate function

Shannon's distortion rate function

$$D^{(s)}_{\mathrm{Shannon}}(r) = \inf\left\{\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} \, : \, \hat{X} \text{ r.v. with } I(X;\hat{X}) \le r\right\},$$

where

$$I(X;\hat{X}) = H(\mathbb{P}_{X,\hat{X}} \| \mathbb{P}_X \otimes \mathbb{P}_{\hat{X}}) \qquad \text{(mutual information)}$$

II Shannon's distortion rate function

Shannon's distortion rate function

$$D^{(s)}_{\mathrm{Shannon}}(r) = \inf\{\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} \,:\, \hat{X} \text{ r.v. with } I(X;\hat{X}) \le r\},$$

where

$$I(X;\hat{X}) = H(\mathbb{P}_{X,\hat{X}} \| \mathbb{P}_X \otimes \mathbb{P}_{\hat{X}}) \qquad \text{(mutual information)}$$

Properties: • $D^{(s)}_{\mathrm{Shannon}}(r) \le D^{(s)}(r)$

II Shannon's distortion rate function

Shannon's distortion rate function

$$D^{(s)}_{\mathrm{Shannon}}(r) = \inf\{\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} : \hat{X} \text{ r.v. with } I(X;\hat{X}) \leq r\},$$

where

$$I(X;\hat{X}) = H(\mathbb{P}_{X,\hat{X}} \| \mathbb{P}_X \otimes \mathbb{P}_{\hat{X}}) \qquad \text{(mutual information)}$$

Properties: • $D^{(s)}_{\mathrm{Shannon}}(r) \leq D^{(s)}(r)$

• For a r.v. $X = (X_1, \dots, X_n)$ with independent entries

$$\inf\Big\{\mathbb{E}\Big[\sum_{j=1}^{n} \rho_j(X_j, \hat{X}_j)\Big] : I(X;\hat{X}) \leq r\Big\} = \inf\Big\{\sum_{j=1}^{n} D_j(r_j) : r_1 + \dots + r_n \leq r\Big\},$$

where

$$D_j(r_j) := \inf\{\mathbb{E}[\rho_j(X_j, \hat{X}_j)] : I(X_j;\hat{X}_j) \leq r_j\}.$$

II Shannon's distortion rate function

Shannon's distortion rate function

$$D^{(s)}_{\mathrm{Shannon}}(r) = \inf\{\mathbb{E}[\|X - \hat{X}\|^s]^{1/s} \,:\, \hat{X} \text{ r.v. with } I(X;\hat{X}) \le r\},$$

where

$$I(X;\hat{X}) = H(\mathbb{P}_{X,\hat{X}} \| \mathbb{P}_X \otimes \mathbb{P}_{\hat{X}}) \qquad \text{(mutual information)}$$

Properties: • $D^{(s)}_{\mathrm{Shannon}}(r) \le D^{(s)}(r)$

• For a r.v. $X = (X_1, \dots, X_n)$ with independent entries

$$\inf\Big\{\mathbb{E}\Big[\sum_{j=1}^n \rho_j(X_j, \hat{X}_j)\Big] \,:\, I(X;\hat{X}) \le r\Big\} \overset{(\le)}{=} \inf\Big\{\sum_{j=1}^n D_j(r_j) \,:\, r_1 + \dots + r_n \le r\Big\},$$

where

$$D_j(r_j) := \inf\{\mathbb{E}[\rho_j(X_j, \hat{X}_j)] \,:\, I(X_j;\hat{X}_j) \le r_j\}.$$

II Kolmogorov's inverse water filling principle

Thm:

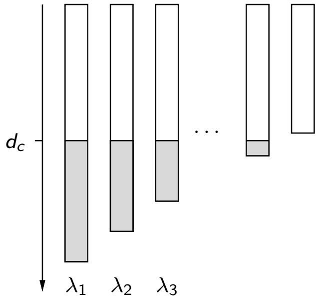
$$\left(D^{(2)}_{\text{Shannon}}(r)\right)^2 = \min \sum_j \lambda_j e^{-2r_j}$$

with constraint $\sum_j r_j = r$.

Sol. Choose r_j as

$$\lambda_j \wedge d_c = \lambda_j e^{-2r_j}$$

for appropriate $d_c > 0$.



II Asymptotic formulae

Thm (D '03): If

$$\lim_{n\to\infty} \frac{\log\log(1/\lambda_n)}{n} = 0,$$

then for all $s > 0$,

$$D^{(s)}(r) \sim D^{(2)}_{\mathrm{Shannon}}(r), \qquad r \to \infty.$$

- Asymptotics do not depend on s
- Approximation error concentrated around $D^{(2)}_{\mathrm{Shannon}}(r)$
- Asymptotic-equipartition-property (Dembo, Kontoyiannis '02)

II Applications

Weakly optimal scheme

- choose r_j according to RAP
- choose r_j-quantizations μ_j for $\mathcal{N}(0, \lambda_j)$ (weakly optimal)

$\rightsquigarrow$ **approximation** $\prod_j \mu_j$

Ref. Luschgy, Pagès '04, D '03

II Applications

Weakly optimal scheme

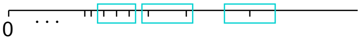
- choose r_j according to RAP
- choose r_j-quantizations μ_j for $\mathcal{N}(0, \lambda_j)$ (weakly optimal)

$\rightsquigarrow$ **approximation** $\prod_j \mu_j$

Strongly optimal scheme

- represent $\mathbb{N} = \dot{\bigcup} I_k$ such that
  - $\# I_k \to \infty$ as $k \to \infty$
  - $\max_{j \in I_k} \lambda_j / \min_{j \in I_k} \lambda_j \to 1$

  (subband decomposition)
- choose $\sum_{j \in I_k} r_j$-quantizations ν_k for $\prod_{j \in I_k} \mathcal{N}(0, \lambda_j)$ (strongly optimal)



$\rightsquigarrow$ **approximation** $\prod_k \nu_k$

Ref. Luschgy, Pagès '04, D '03

III Diffusions

Ass.: X solution to

$$X_t = \int_0^t b(X_u, u)\, du + \int_0^t \sigma(X_u, u)\, dW_u,$$

where $b, \sigma : \mathbb{R} \times [0, \infty) \to \mathbb{R}$ satisfy for fixed $C > 0$ and $\beta \in (0, 1]$

$$b(x,t) \leq C[|x| + 1] \text{ and}$$
$$|\sigma(x,t) - \sigma(x',t')| \leq C\left[|x - x'|^\beta + |x - x'| + |t - t'|^\beta\right].$$

Moreover, $\mathfrak{X} = L^p[0,1]$ for $p \in [1, \infty)$ or
$\mathfrak{X} = C[0,1]$ ($\rightsquigarrow$ $p = \infty$)

III Preliminaries

When X is Brownian motion one has:

Thm (D, Scheutzow '06): $\exists \kappa_p \in \mathbb{R}_+$ such that for all $s > 0$

$$\lim_{r \to \infty} \sqrt{r}\, D^{(s)}(r) = \kappa_p$$

- Asymptotics do not depend on s
- Approximation error concentrated around $\kappa_p / \sqrt{r}$
- Similar result for fractional Brownian motion

III Quantization of diffusions

Thm (D '08): For $p \in [1, \infty)$,

$$\lim_{r \to \infty} \sqrt{r}\, D^{(s)}(r) = \kappa_p \, \mathbb{E}\Big[\Big(\int_0^1 |\sigma_u|^{\frac{2p}{p+2}} \, du \Big)^{\frac{p+2}{2p} s} \Big]^{1/s} = \kappa_p \, \Big\| \big\| \sigma_\cdot \big\|_{L^{\frac{2p}{p+2}}[0,1]} \Big\|_{L^s(\mathbb{P})}$$

III Quantization of diffusions

Thm (D '08): For $p \in [1, \infty)$,

$$\lim_{r \to \infty} \sqrt{r}\, D^{(s)}(r) = \kappa_p \, \mathbb{E}\Big[\Big(\int_0^1 |\sigma_u|^{\frac{2p}{p+2}}\, du\Big)^{\frac{p+2}{2p}s}\Big]^{1/s} = \kappa_p \Big\| \big\|\sigma.\big\|_{L^{\frac{2p}{p+2}}[0,1]} \Big\|_{L^s(\mathbb{P})}$$

and, for $p = \infty$,

$$\lim_{r \to \infty} \sqrt{r}\, D^{(s)}(r) = \kappa_\infty \Big\| \big\|\sigma.\big\|_{L^2[0,1]} \Big\|_{L^s(\mathbb{P})}$$

III Quantization of diffusions

Thm (D '08): For $p \in [1, \infty)$,

$$\lim_{r\to\infty} \sqrt{r}\, D^{(s)}(r) = \kappa_p\, \mathbb{E}\Big[\Big(\int_0^1 |\sigma_u|^{\frac{2p}{p+2}}\, du\Big)^{\frac{p+2}{2p}s}\Big]^{1/s} = \kappa_p \Big\| \big\|\sigma_\cdot\big\|_{L^{\frac{2p}{p+2}}[0,1]} \Big\|_{L^s(\mathbb{P})}$$

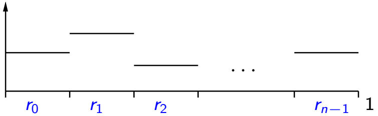
and, for $p = \infty$,

$$\lim_{r\to\infty} \sqrt{r}\, D^{(s)}(r) = \kappa_\infty \Big\| \big\|\sigma_\cdot\big\|_{L^2[0,1]} \Big\|_{L^s(\mathbb{P})}$$

Heuristics: Suppose that $(\sigma_t)_{t\in[0,1]}$ is deterministic and piecewise constant on each interval $I_j = [j/n, (j+1)/n)$. Then for a good approximation $\hat{X}$, typically,

$$\int_0^1 |X_t - \hat{X}_t|^p\, dt \approx \kappa_p^p\, \frac{1}{n} \sum_{j=0}^{n-1} \frac{|\sigma_{j/n}|^p}{(nr_j)^{p/2}}$$

where r_j is the rate assigned for the approximation of the piece I_j.



$\sum_j r_j = r$ (rate constraint)

$\int_{I_j} |X_t - \hat{X}_t|^p\, dt \approx \kappa_p^p\, \frac{|\sigma_{j/n}|^p}{(nr_j)^{p/2}}$

III Quantization of diffusions

Thm (D '08): For $p \in [1, \infty)$,

$$\lim_{r\to\infty} \sqrt{r}\, D^{(s)}(r) = \kappa_p\, \mathbb{E}\Big[\Big(\int_0^1 |\sigma_u|^{\frac{2p}{p+2}}\, du\Big)^{\frac{p+2}{2p}s}\Big]^{1/s} = \kappa_p \Big\| \big\|\sigma_\cdot\big\|_{L^{\frac{2p}{p+2}}[0,1]} \Big\|_{L^s(\mathbb{P})}$$

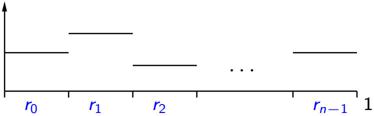
and, for $p = \infty$,

$$\lim_{r\to\infty} \sqrt{r}\, D^{(s)}(r) = \kappa_\infty \Big\| \big\|\sigma_\cdot\big\|_{L^2[0,1]} \Big\|_{L^s(\mathbb{P})}$$

Heuristics: Suppose that $(\sigma_t)_{t\in[0,1]}$ is deterministic and piecewise constant on each interval $I_j = [j/n, (j+1)/n)$. Then for a good approximation $\hat{X}$, typically,

$$\int_0^1 |X_t - \hat{X}_t|^p\, dt \approx \kappa_p^p \frac{1}{n} \sum_{j=0}^{n-1} \frac{|\sigma_{j/n}|^p}{(nr_j)^{p/2}} = \kappa_p^p \int_0^1 \frac{|\sigma_t|^p}{(\bar{r}_t)^{p/2}}\, dt$$

where r_j is the rate assigned for the approximation of the piece I_j.



$\sum_j r_j = r$ (rate constraint)

$\int_{I_j} |X_t - \hat{X}_t|^p\, dt \approx \kappa_p^p \frac{|\sigma_{j/n}|^p}{(nr_j)^{p/2}}$

III Rate allocation problem

Minimize

$$\int_0^1 \frac{|\sigma_t|^p}{(\bar{r}_t)^{p/2}}\, dt$$

over all non-negative $(\bar{r}_t)$ with $\int_0^1 \bar{r}_t\, dt = r$.

III Rate allocation problem

Minimize

$$\int_0^1 \frac{|\sigma_t|^p}{(\bar{r}_t)^{p/2}}\, dt$$

over all non-negative $(\bar{r}_t)$ with $\int_0^1 \bar{r}_t\, dt = r$.

Solution:

$$\bar{r}_t = \frac{1}{Z}\, |\sigma_t|^{\frac{2p}{p+2}}\, r,$$

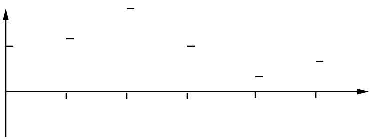
where $Z = \int_0^1 |\sigma_u|^{\frac{2p}{p+2}}\, du$.

- Similar results obtained for strong approximation by piecewise linear functions (Müller-Gronbach '96)

III Constructive quantization

Approach:

- 1st step: quantize Brownian motion on a coarse time grid and compute an approximation via Milstein scheme
  $\rightsquigarrow$ each path leads to an approximation $(\hat{\sigma}_t)$

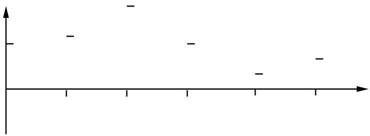


Ref: Müller-Gronbach, Ritter (work in progress)

III Constructive quantization

Approach:

- 1st step: quantize Brownian motion on a coarse time grid and compute an approximation via Milstein scheme
  $\rightsquigarrow$ each path leads to an approximation $(\hat{\sigma}_t)$
- 2nd step: each coarse approximation is refined by inserting bridges according to the rates induced by $(\hat{\sigma}_t)$

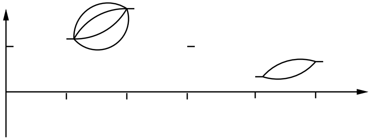


Ref: Müller-Gronbach, Ritter (work in progress)

III Constructive quantization

Approach:

- 1st step: quantize Brownian motion on a coarse time grid and compute an approximation via Milstein scheme
$\rightsquigarrow$ each path leads to an approximation $(\hat{\sigma}_t)$
- 2nd step: each coarse approximation is refined by inserting bridges according to the rates induced by $(\hat{\sigma}_t)$



Ref: Müller-Gronbach, Ritter (work in progress)

IV Lévy processes

Now:

- X is a (ν, σ^2, b)-Lévy process, i.e. a Lévy process with

$$\mathbb{E}e^{i\theta X_1} = \exp\Big\{-\frac{\sigma^2}{2}\theta^2 + ib\theta + \int_{\mathbb{R}} \big(e^{iux} - 1 - \mathbb{1}_{\{|x|\leq 1\}} iux\big)\,\nu(dx)\Big\}$$

- $\mathfrak{X} = L^p[0,1]$ with $p \in [1, \infty)$

IV Lévy processes

Now:

- X is a (ν, σ^2, b)-Lévy process, i.e. a Lévy process with

$$\mathbb{E}e^{i\theta X_1} = \exp\Big\{-\frac{\sigma^2}{2}\theta^2 + ib\theta + \int_{\mathbb{R}} \big(e^{iux} - 1 - \mathbb{1}_{\{|x|\le 1\}} iux\big)\, \nu(dx)\Big\}$$

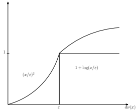
- $\mathfrak{X} = L^p[0,1]$ with $p \in [1,\infty)$

Thm (Aurzada, D '08): Under additional assumptions $\exists$ constants $c_1 = c(p,\nu)$ and c_2, such that for sufficiently small $\varepsilon > 0$ and $s > 0$,

$$\tfrac{1}{c_2}\,\varepsilon \le D^{(1)}(c_1\, F(\varepsilon)) \quad \text{and} \quad D^{(s)}(\tfrac{1}{c_1}\, F(\varepsilon)) \le c_2\, \varepsilon,$$

where

$$F(\varepsilon) = \frac{\sigma^2}{\varepsilon^2} + \int_{\mathbb{R}} \Big[\Big(\frac{x^2}{\varepsilon^2} \wedge 1\Big) + \log_+ \frac{|x|}{\varepsilon}\Big]\, \nu(dx).$$



IV Rate allocation

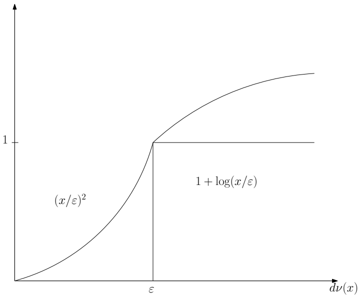
Approximate evolutions induced by *large* and *small* jumps (together with Brownian component) separately.

- **Small jumps.** Approximate exit times of the compensated process out of ε-intervals.

  $\rightsquigarrow$ allocated rate $\frac{\sigma^2}{\varepsilon^2} + \int_{[-\varepsilon,\varepsilon]} \frac{x^2}{\varepsilon^2}\, \nu(dx)$

- **Large jumps.** Approximate individual jumps.

  $\rightsquigarrow$ allocated rate $\int_{[-\varepsilon,\varepsilon]^c} \left(1 + \log \frac{|x|}{\varepsilon}\right) \nu(dx)$



V Final remarks

- Quantization useful in the analysis of quadrature problems
  - Variance reduction
  - Lower bounds for stochastic algorithms
- Coarse descriptions of good quantizations available for a number of random objects
  - Finite dimensional X under Orlicz norm dist. (D, Vormoor (Prep.))
  - Gaussian X in Hilbert space
  - Diffusions
  - Lévy processes
- Asymptotically optimal constructive quantization partially understood