# Reduced representations of non linear manifolds: from reduced basis to (conditional) vector quantization of measures

Gabriel Turinici

CEREMADE
Université Paris Dauphine - PSL
Paris, France

The Tenth Congress of Romanian Mathematicians

Pitești, June 30 - July 5th, 2023

# Outline

# Motivation: latent dimension in high dimensional objects

• Parametric PDE : for $\omega \in \Omega =$ parameter, the solution $u_\omega(\cdot)$ of a parametric PDE needs to be computed for many values of $\omega$.

• reduced basis idea: when some linearity is present : $(\mathcal{T}_0 + \omega \mathcal{T}_1) u_\omega = 0$, e.g. $\mathcal{T} = -\Delta$, $\mathcal{T}_1 = f$ try to solve the PDE in a small reduced basis made by previously computed solutions $u_{\omega_q}$ i.e., on $W_Q = span\{u_{\omega_q}, q = 1, .., Q\}$. Typically $Q \simeq 10 - 100$.

• offline-online decomposition : if the operators $\mathcal{T}_i$ $(i = 0, 1)$ are precomputed in $W_Q$, then finding the solution $u_\omega$ is extremely fast.

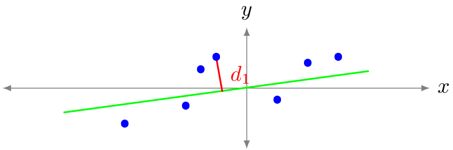
• beyond linear setting : cf. [1] and related works.

# Motivation: latent dimension in high dimensional objects

- Questions : does this work ? how to choose $\omega_q$ ?

- Answers: ok for the 1D linear case, ok for multi-linear. In general ok if the set of solutions $\{u_\omega; \omega \in \Omega\}$ will have a small Kolmogorov n-width (cf [2] and subsequent literature).

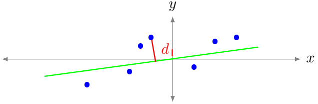
Recall n-width of a set $S$ in a space $H$ measures how well $S$ is approximated by linear space of dimension $n$ :
$d_n(S, H) = \inf_{H_n \subset H, dim(X_n)=n} \sup_{u \in S} \inf_{v_n \in H_n} \|u - v_n\|$.




- Related idea: *proper orthogonal decomposition (POD)* : (some related) eigenvalues decay fast.

# Motivation: latent dimension in high dimensional objects

- Remark : max-norm is used, sensitive to outliers : one point can change the n-width ... not robust.




- alternative form : some average over the solutions $\int F(\omega, u_\omega)\rho(\omega)d\omega$ : $\rho(\omega)$ is a probability density over $\Omega$

quadrature formula $\int F(\omega, u_\omega)\rho(\omega)d\omega \simeq \sum_q w_q F(\omega_q, u_{\omega_q})$.

How to choose representative points $u_{\omega_q}$ ?

- Quantization : procedure similar to clustering and compression


Figure: **Left:** example of clustering. **Middle and right:** compression of the middle image into the right image (credits: Wikipedia)

# Motivation: summarize information from a set of objects

Example 1: alternative to Monte Carlo to approximate integrals over high dimensional spaces : for $\int_\Omega f(\omega)\rho(\omega)d\omega$ it is good to have a sample $\frac{1}{K}\sum_{k=1}^K \delta_{\omega_k}$ close, as measure, to $\rho(\omega)d\omega$ : if $\rho(\omega)d\omega \simeq \frac{1}{K}\sum_{k=1}^K \delta_{\omega_k}$ then $\int_\Omega f(\omega)\rho(\omega)d\omega \simeq \frac{1}{K}\sum_{k=1}^K f(\omega_k)$

- lower dimensional objects : quadrature;
- more exotic objects: $\omega$ (a curve) is a realization of a $W_t$ = Brownian mvt. "cubature". $\int f(t, W_t)dW_t$

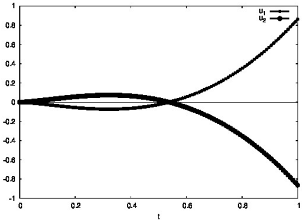



Figure: Example of cubature "points" for a Brownian motion, from [5].

# Motivation: summarize information from a set of objects

Example 2: summarize a distribution with $K$ points, e.g. 2D Gaussian.

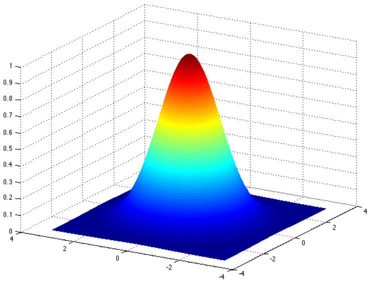


Figure: 2D Gaussian (credits: Wikipedia)

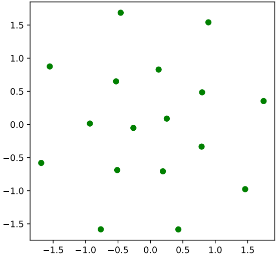


Figure: Example of compression with $K = 17$ points of a 2D Gaussian using special statistical distances (cf. [7]).
Presence of a three layers point structure: inner 2, middle 7, outer 8 (from [6]).

# Motivation: summarize information from a set of objects

Example 3: summarize a large database of objects (e.g. MNIST, FMNIST, CIFAR10, ...)

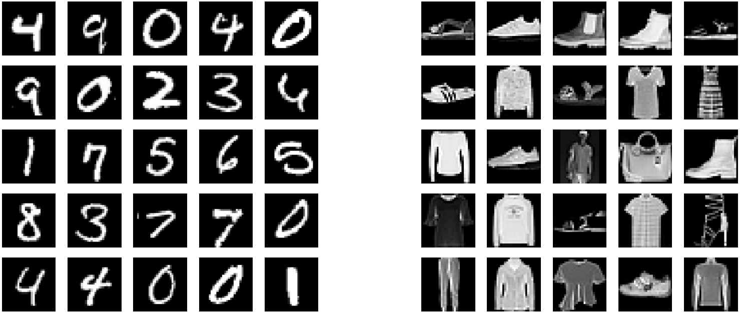


Figure: Left: MNIST samples (25 out of 60'000). Right: Fashion MNIST samples (25 out of 60'000), from [7]

# Outline

# Quantization : tools and algorithms: idea

- Goal: quantize a finite Borel measure $\mu$ (usually a probability measure).
- Idea: minimize $dist(\frac{1}{K}\sum_{k=1}^{K}\delta_{x_k},\mu)^2$.
- Naive idea : $dist(\delta_x,\delta_y)^2 = \|x-y\|^2$ then by linearity (Hilbert)

$$dist(\mu,\nu)^2 = (cst)\int \|x-y\|^2(\nu-\mu)(dx)(\nu-\mu)(dy) = (\mathbb{E}_\mu - \mathbb{E}_\nu)^2 \quad (1)$$

$\mu,\nu$ are measures; only average object is reproduced ($dist(\mu, E_\mu)=0$ !) ...

Denote $h(x,y) = d(\delta_x,\delta_y)^2$ ; often $h(x,y) = h(|x-y|)$ (translation and rotation invariant kernel) $h(x)$ = important function to choose.

Statistical distance: $$d^2(\mu,\nu) = -\frac{1}{2}\int h(x,y)(\mu-\nu)(dx)(\mu-\nu)(dy) \quad (2)$$

Example $h = |\cdot|$ : $$d^2\left(\frac{1}{K}\sum_{k=1}^{K}\delta_{x_k},\mu\right) = \left(c(h,\mu) - \frac{1}{2K^2}\sum_{k\neq l}^{K}|x_k - x_l| + \frac{1}{K}\sum_{k=1}^{K} g_\mu(x_k)\right) \quad (3)$$

DAUPHINE UNIVERSITÉ PARIS | CEREMADE

# Energy distance: historical parenthesis

• Historical idea: the "energy distance" builds on the Newton's potential energy concept, cf Szekely 2002.

• Gravitational potential energy of an object at distance $R$ from another behaves like $-1/R$; but when an object is close to a surface (e.g., on earth of radius $R$ at height $H << R$) this behaves like $+H$.

• work required to lift at height $H$ is $cst \cdot H$. Modern concept: gravity battery that stores potential gravitational energy.

• other vision: viewer from other galaxy at distance $\xi$, to summarize the solar system (centered at 0) : potential energy with respect to a planet at $x$ : $-\frac{1}{\|\xi\|} - \left(-\frac{1}{\|\xi - x\|}\right) = O\left(\frac{\langle x, \xi \rangle}{\|\xi\|}\right)$.

# Statistical distances: conditionally negative kernels

Question: $h$ defines a proper distance ? Which is better ?

**Definition (conditionally negative definite)**

A kernel $h(\cdot,\cdot)$ is said to be conditionally negative definite if for any $I \in \mathbb{N}$, $p_1, ..., p_I$ with $\sum p_i = 0$ and any $x_1, ..., x_I$: $\sum_{i,j} p_i p_j h(x_i, x_j) \leq 0$.

this would correspond to $dist(\sum_{i,p_i>0} |p_i|\delta_{x_i}, \sum_{i,p_i<0} |p_i|\delta_{x_i})^2 \geq 0$

**Theorem ("Gini difference" Gini 1912; "energy distance" Szekelly 1985, 2002; "maximum mean discrepancy" Gretton 2007, Radon-Sobolev G.T. 2021 [7])**

*The kernel $h(x) = |x|$ is conditionally negative definite.*

Rq: many other kernels are known to be conditionally negative definite: Gaussian, etc.

# Compatibile kernels : sliced Sobolev idea

• Idea: measures are dual of continous functions,

• use Sobolev embeddings and dual norm

• find a Sobolev space $H^s$ containing continous functions .... regularity vs dimension problem. !!

• in 1D good candidate: $\dot{H}^1$: functions with $L^2$ gradient, quotient $f \sim f + cst$, norm $\|\nabla f\|^2$

• second idea : slicing: project on lines $\theta \# \mu =$ projection on line $\theta$ ; Radon the: if all projections are the same measures are the same.

# Statistical distances: conditionally negative kernels

## Proof (GT 2021 version).

Radon transform of the dual of the homogeneous Sobolev space $\dot{H}^1$: take all directions on the unit sphere, project, measure in $\dot{H}^{-1}$, sum up: $d(\mu,\nu)^2 = \frac{1}{\text{area}(\mathbb{S})} \int_{\mathbb{S}} \|\theta_{\#}\mu - \theta_{\#}\nu\|^2_{\dot{H}^{-1}} d\theta$. Obviously positive, non-degenerate by properties of the Radon transform. □

When $d(\delta_x, \delta_y)^2 = |x-y|$, one minimizes terms involving $|\cdot|$ (not $|\cdot|^2$) : gradient descent methods experience instabilities as the differential is $\frac{x}{|x|^2}$.

## Theorem (Schoenberg 1938 [4], Micchelli 1984 [3], GT 2021 [6])

*For any $a \geq 0$, $\alpha \in ]0,1[$, the kernels $h(x) = (a^2 + |x|^2)^\alpha - a^\alpha$ and $h(x) = \frac{\|x\|^2}{(a+|x|^2)^\alpha}$ are conditionally negative definite and can be expressed explicitly as a Gaussian mixture. In particular this is true for $\sqrt{a^2+x^2} - a$.*

Rq: the proof extends to a larger family of kernels.

# Quantization tools and algorithms: in practice

Implementation : minimize $X = (x_1, ..., x_K) \mapsto d^2\left(\frac{1}{K}\sum_{k=1}^{K}\delta_{x_k}, \mu\right)$

• deterministic optimization techniques when $x \mapsto \mathbb{E}_{y\sim\mu}h(x-y)$ has a closed form (e.g. normal mixture)

• ML / stochastic optimization algorithms (e.g. SGD, Adam, momentum, ...) when the database is large: compute a noisy gradient using batches/sampling from the database. BLUE estimators (i.e., minimal variance GT 23) ...

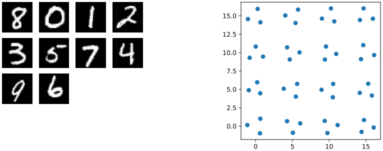



Figure: Measure compression results, from [6].
**Left :** MNIST compression with $K = 10$ samples: we computed the compression then took closest from the database.Note that algorithm chooses by itself to represent the each figure exactly once. **Right :** 16 2D Gaussians on a grid compressed with $K = 16 * 3$ points.

# Outline

1. Motivation
2. Quantization : tools and algorithms
3. Theoretical questions
   - Existence
   - Conditional quantization
4. Application to a transport equation
   - Quantization of the transport equation

# Theoretical questions

• Existence : does the minimization of $X = (x_1, ..., x_Q) \mapsto d^2\left(\frac{1}{Q}\sum_{q=1}^{Q}\delta_{x_q}, \mu\right)$ has a solution ?

• problem with the intuition : for $h = \|x\|$ : the measure $1/a\delta_{a^2} + (a-1)/a\delta_0$ is at distance 1 from $\delta_0$ even as $a \to \infty$ !! The neighborhood of the origin does not contain only bounded support measures ...

• similar problem : $v : \Omega \to \mathcal{H}$ continuous, is $\inf_{\omega \in \Omega^Q} \|u - \sum_q \alpha_q v(\omega_q)\|^2$ attained (note $\Omega$ not compact)

• $\mathcal{H}$ is a RKHS (reproducing kernel Hilbert space) associated to $h(x,y) = d^2(\delta_x, \delta_y)$

# Theoretical questions

**Theorem (GT 22[6])**

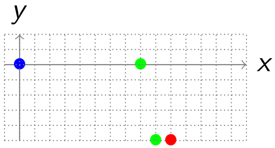
*Consider Huber-energy kernels* $h(x,y) = (a^2 + \|x-y\|^2)^{r/2} - a^r$ *(*$a \geq 0$*,* $r \in ]0,1[$*) or the Gaussian kernel* $g(x,y) = 1 - e^{-\|x-y\|^2/2}$ *and* $\mu$ *a measure with finite moments of order* $r$*. For fixed* $\alpha_q \geq 0$ *with* $\sum \alpha_q = 1$ *the infimum* $\inf_{x \in \mathbb{R}^Q} d(\sum_q \alpha_q \delta_{x_q}, \mu)^2$ *is attained (*$d$ *is the distance associated to kernel* $h$*).*

Rq: two regimes $\lim_{\|x-y\| \to \infty} d(\delta_x, \delta_y)^2 \to \infty$ is finite or infinite.

• particular case 1D for the "energy" kernel, uniform weights : solutions are $q + 1/2/Q$ quantiles, $q = 0, ..., Q-1$.

# Theoretical questions: positivity of the quantization

Question: suppose $\mu \geq 0$; when optimizing both weights $p_k$ and support points $x_k$ variables) is the compressed measure positive too ?




projection of $(0,0)$ to $(8,0),(9,-5),(10,-5)$ has optimal weight negative for red point

## Proposition (..., GT 2021 in some cases)

*Let $\mu$ be a probability law on a convex domain with finite first order moment and $K \in \mathbb{N}^\star$. If*

$$\sum_{k=1}^{K} p_k^* \delta_{x_k^*} \in argmin_{p_k, x_k, \sum_{k=1}^{K} p_k = 1} \left[ d\left( \sum_{k=1}^{K} p_k \delta_{x_k}, \mu \right)^2 \right] \tag{4}$$

*then* $\sum_{k=1}^{K} p_k^* \delta_{x_k^*} \geq 0$ *i.e.,* $p_k^* \geq 0, \ \forall k$.

# Theoretical questions: conditional quantization

• Question: how does the quantization depends on target measure $\mu$ ? Suppose $\mu$ depends on parameter $u$ e.g. $\mu(u) = \mathcal{N}(u, 1)$ (1D normal of mean $u$, variance 1). $Q$ = fixed, compression for given $u$ =ok. What about continuity w/r to u ?

• $\mu = \mu(u)$, each measure is 1D valued; to simplify we take them as probability laws.

## Lemma (regularity w/r to target)

*Suppose $u \mapsto \mu(u)$ is regular enough (...) and the measure $\mu(u)$ is non-atomic $\forall u$; then:*

- *the minimization problem $\frac{1}{Q}\sum_{q=1}^{Q} \delta_{x_q} \mapsto d^2\left(\frac{1}{Q}\sum_{q=1}^{Q} \delta_{x_q}, \mu(u)\right)$ admits a unique solution $C(u) = \frac{1}{Q}\sum_{q=1}^{K} \delta_{x_q}$ (as a probability law);*
- *the mapping $u \mapsto C(u)$ is regular with respect to $u$.*

# Theoretical questions: conditional quantization

• In general : multi-dimensional, conditional quantization $\mu(dy|X = x)$

Problems :

• the quantization is not necessarily unique for some $u$ (e.g. symmetries of $\mu$);

• Can prove the existence of a measurable selection (GT 2023).

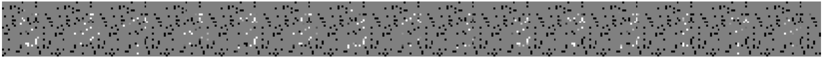


Figure: Numerical results : MNIST conditional reconstruction: the input

# Conditional quantization : numerical results

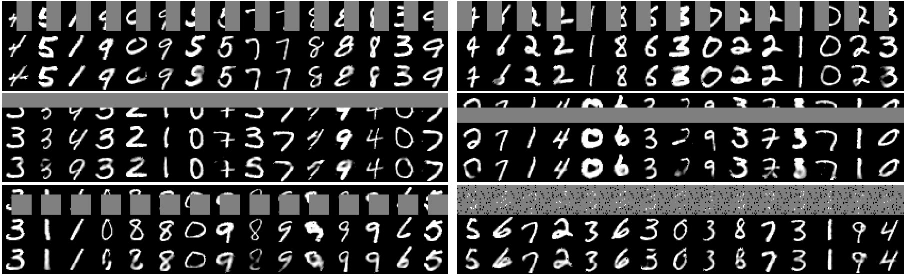


Figure: Numerical results : MNIST conditional reconstruction; line 1 = input $x$, line 2 = truth, line 3 = output $y$

# Outline

# Application to a transport equation, joint work with L. Laguzet

- Transport equation :

$$\frac{1}{c}\partial_t u(t,x,\omega)+\omega\cdot\nabla u(t,x,\omega)+\sigma_a(t,x)u(t,x,\omega)=\sigma_s(t,x)\left[\langle u\rangle(t,x)-u(t,x,\omega)\right]. \quad (5)$$

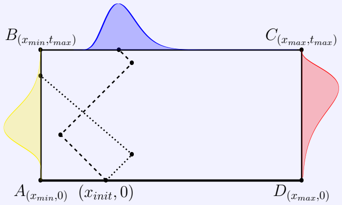
- $\nabla=\nabla_x$, $t$ = time, $x$ = position, $\omega\in\mathcal{S}^d$ = angle of propagation, $\langle u\rangle(t,x)=\fint_{\mathcal{S}^d}u(t,x,\omega)d\omega$; absorption opacity $\sigma_a=\sigma_a(t,x)$, scattering opacity $\sigma_s=\sigma_s(t,x)$ known functions, can be taken piecewise constant.
- very computationally intensive, $\sigma$ spans several orders of magnitude, large total time.

- "diffusion limit" $(\sigma_a, 1/\sigma_s = O(\epsilon))$: $u=\langle u\rangle-\frac{1}{\sigma_a+\sigma_s}\omega\cdot\nabla\langle u\rangle+O(\epsilon)$ where $\frac{1}{c}\partial_t\langle u\rangle(t,x)=\nabla\cdot\left[\frac{1}{3(\sigma_a(t,x)+\sigma_s(t,x))}\nabla\langle u\rangle(t,x)\right]-\sigma_a(t,x)\langle u\rangle(t,x)$.

- for constant $\sigma_a,\sigma_s$ analytic solutions (in the diffusion limit) exist. What about intermediary regimes ??

# Application to a transport equation

• toy problem $\sigma_a = 0$, $\sigma_s = cst$ on $[x_{min}, x_{max}]$, $\omega = \pm 1$. Fundamental solution: given by a particular method.




Particle starts at $x_{init}$, has collisions after $Exp(\sigma)$ times; can escape through any of the domain's frontiers: time is up or because reaching border of the domain. Escape probability law $\mathcal{E}(\sigma, x_{init}/(x_{max} - x_{min}), t_{max})$ with support on the border (cross the exit angle parameter).

• Idea: use particular methods, precompute the quantization of the escape distribution on a grid of possible values for $t_{max}$, $x_0$, $\sigma_s$.

# Quantization of the transport equation

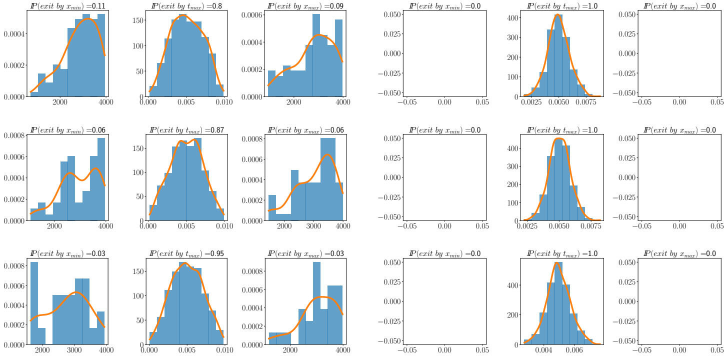



Figure: Escape statistics from a time-space domain $[x_{min}, x_{max}] \times [0, t_{max}]$. columns 1 to 3 : $\sigma = 0.75$ (first line), $\sigma = 1$ (second line) and $\sigma = 1.25$ (third line); the plots in columns 4 to 6 : $\sigma = 7.5$ (first line) $\sigma = 10$ (second line) and $\sigma = 12.5$ (third line).

# Conclusions and future work

Further questions:

- existence results for more general kernels and variable weights
- positivity under more general conditions
- conditional quantization : faster algo

# References I

[1] Yvon Maday, Olga Mula, and Gabriel Turinici. "Convergence analysis of the generalized empirical interpolation method". In: *SIAM Journal on Numerical Analysis* 54.3 (2016), pp. 1713–1731.

[2] Yvon Maday, Anthony T. Patera, and Gabriel Turinici. "A Priori Convergence Theory for Reduced-Basis Approximations of Single-Parameter Elliptic Partial Differential Equations". In: *Journal of Scientific Computing* 17.1 (Dec. 2002), pp. 437–446. ISSN: 1573-7691. DOI: 10.1023/A:1015145924517. URL: https://doi.org/10.1023/A:1015145924517.

[3] Charles A Micchelli. "Interpolation of scattered data: distance matrices and conditionally positive definite functions". In: *Approximation theory and spline functions*. Springer, 1984, pp. 143–145.

# References II


[4] Isaac J Schoenberg. “Metric spaces and completely monotone functions”. In: *Annals of Mathematics* (1938), pp. 811–841.

[5] Gabriel Turinici. “Cubature on C1 Space”. In: *Control and Optimization with PDE Constraints*. Springer, 2013, pp. 159–172.

[6] Gabriel Turinici. *Huber-energy measure quantization*. 2022. DOI: 10.48550/ARXIV.2212.08162. URL: https://arxiv.org/abs/2212.08162.

[7] Gabriel Turinici. “Radon–Sobolev Variational Auto-Encoders”. In: *Neural Networks* 141 (2021), pp. 294–305. ISSN: 0893-6080. DOI: 10.1016/j.neunet.2021.04.018.