Influence of the dimension on square-norm based quantization

Gudhi, 29 Octobre 2014

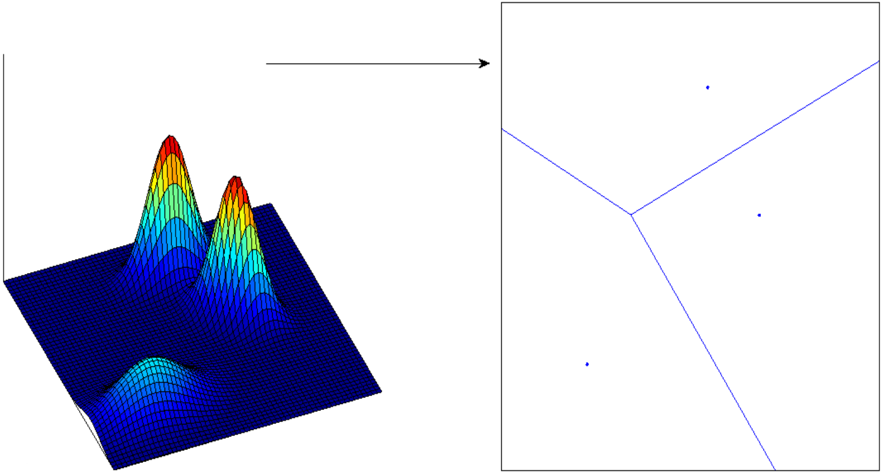
Clément Levrard, INRIA Saclay

1. Introduction
   - Quantization principle
   - Quantizers/Risk
   - ERM strategy
2. Approximation error
   - Finite-dimensional spaces
   - Infinite-dimensional spaces
3. Estimation error
   - Finite-dimensional case
   - Hilbertian case
4. Variable selection

Partie 1

1. Introduction
    - Quantization principle
    - Quantizers/Risk
    - ERM strategy
2. Approximation error
    - Finite-dimensional spaces
    - Infinite-dimensional spaces
3. Estimation error
    - Finite-dimensional case
    - Hilbertian case
4. Variable selection

Principle



Principle

P a distribution over $\mathbb{H}$, $dim(\mathbb{H}) = d \leq +\infty$

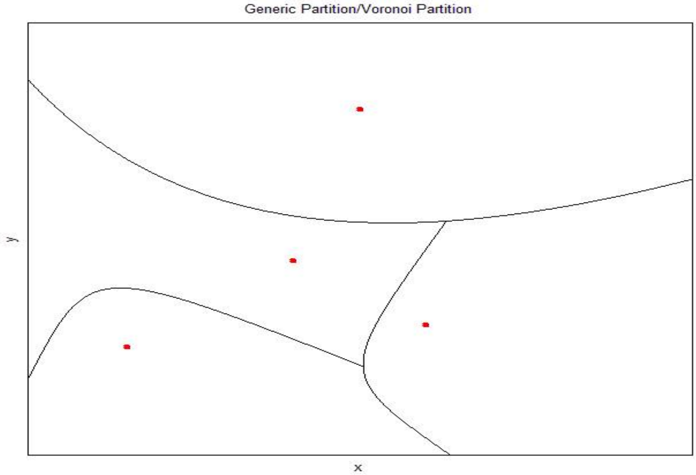
- divide $\mathbb{H}$ into k groups
- without too much loss
- according to a n-sample

Quantizers/Codebooks/Risk

P a distribution over $\mathbb{H}$, Q a k-point quantizer

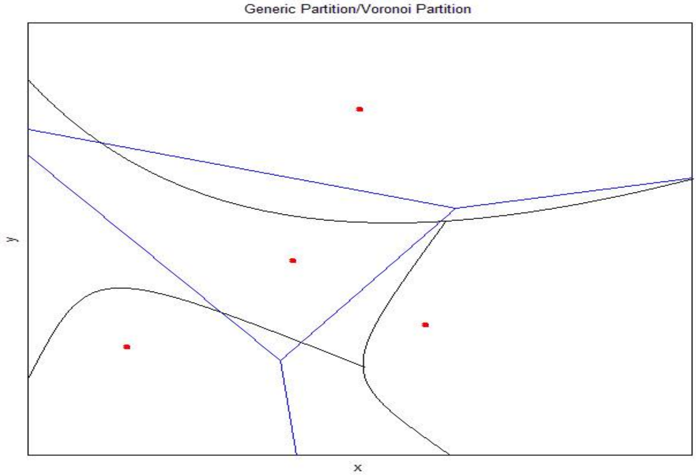
- $(c_1, \ldots, c_k)$ its images (codepoints)
- $R(Q) = P\|x - Q(x)\|^2 = \sum_{i=1}^{k} P\|x - c_i\|^2 \mathbb{1}_{Q^{-1}\{c_i\}}(x)$

Quantizers/Codebooks/Risk



$$\to R(Q) = P\|x - Q(x)\|^2 = \sum_{i=1}^{k} P\|x - c_i\|^2 \mathbb{1}_{Q^{-1}\{c_i\}}(x)$$

Quantizers/Codebooks/Risk



$$\rightarrow R(Q_{nn}(\mathbf{c})) \leq R(Q)$$

Quantizers/Codebooks/Risk

P a distribution over $\mathbb{H}$, Q a k-point nearest-neighbor quantizer

- Risk/Distortion $R(Q) = R(\mathbf{c}) = P \min_{i=1,\ldots,k} \|x - c_i\|^2$
- $\mathbf{c} = (c_1, \ldots, c_k)$ a codebook
- c_i codepoints
- Remark:

$$R(\mathbf{c}) = \min_{\nu | Supp(\nu) = \mathbf{c}} \mathcal{W}_2^2(P, \nu)$$

ERM Strategy

$X_1, \ldots, X_n$ a n-sample.

- Target: $\mathbf{c}^* = \arg\min R(\mathbf{c})$
- Available: $\frac{1}{n}\sum_{i=1}^{n} \min_{j=1,\ldots,k} \|X_i - c_j\|^2 = P_n \min_{i=1,\ldots,k} \|x - c_i\|^2$

$\rightarrow$ $\hat{\mathbf{c}}_n \in \arg\min P_n \min_{i=1,\ldots,k} \|x - c_i\|^2$

ERM Strategy

$X_1, \ldots, X_n$ a n-sample.

- Target: $\mathbf{c}^* = \arg\min R(\mathbf{c})$
- Available: $\frac{1}{n}\sum_{i=1}^{n} \min_{j=1,\ldots,k} \|X_i - c_j\|^2 = P_n \min_{i=1,\ldots,k} \|x - c_i\|^2$

$\rightarrow$ $\hat{\mathbf{c}}_n \in \arg\min P_n \min_{i=1,\ldots,k} \|x - c_i\|^2$

Approximation/Estimation Error

$$R(\hat{\mathbf{c}}_n) = \underbrace{R(\mathbf{c}^*)}_{Approximation} + \underbrace{R(\hat{\mathbf{c}}_n) - R(\mathbf{c}^*)}_{Estimation}$$

Partie 2

1. Introduction
   - Quantization principle
   - Quantizers/Risk
   - ERM strategy
2. Approximation error
   - Finite-dimensional spaces
   - Infinite-dimensional spaces
3. Estimation error
   - Finite-dimensional case
   - Hilbertian case
4. Variable selection

Finite-dimensional case

Zador Theorem

If $\mathcal{H} = \mathbb{R}^d$, and $P\|x\|^{2+\delta} < \infty$, then

$$R_k^* \underset{k\to\infty}{\sim} Ck^{-\frac{2}{d}}.$$

Finite-dimensional case

Zador Theorem

If $\mathcal{H} = \mathbb{R}^d$, and $P\|x\|^{2+\delta} < \infty$, then

$$R_k^* \underset{k\to\infty}{\sim} Ck^{-\frac{2}{d}}.$$

Canas, Poggio, Rosasco

If $\mathcal{H}$ is a smooth compact Riemmanian d-submanifold embedded in $\mathcal{B}_D(0, M)$, then

$$R_k^* \underset{k\to\infty}{\sim} Ck^{-\frac{2}{d}}.$$

Gaussian measures only

$\mathcal{H}$ a Hilbert space, P a Gaussian process with eigenvalues $\lambda_1 \geq \lambda_2 \ldots$

Graf, Luschgy, Pagés

If $\lambda_j \approx \phi(j)$, where

$$\frac{\phi(tx)}{\phi(x)} \underset{x \to \infty}{\longrightarrow} t^{-b}, b > 1,$$

then

$$R_k^* \approx \left(\Psi(\log(k))\right)^{-1/2},$$

where

$$\Psi(x) = \frac{1}{x\phi(x)}.$$

Gaussian measures only

Graf, Luschgy, Pagés

If $\lambda_j \approx \phi(j)$, where

$$\frac{\phi(tx)}{\phi(x)} \underset{x\to\infty}{\longrightarrow} t^{-b}, b > 1,$$

then

$$R_k^* \approx \left(\Psi(\log(k))\right)^{-1/2},$$

where

$$\Psi(x) = \frac{1}{x\phi(x)}.$$

Example: If $\lambda_j = j^{-b}, b > 1$, then

$$R_k^* \approx (\log(k))^{\frac{1-b}{2}}.$$

Partie 3

1. Introduction
   - Quantization principle
   - Quantizers/Risk
   - ERM strategy
2. Approximation error
   - Finite-dimensional spaces
   - Infinite-dimensional spaces
3. Estimation error
   - Finite-dimensional case
   - Hilbertian case
4. Variable selection

Combinatorial bound

Small Lemma: If $Supp(P) \subset \mathcal{B}_d(0, M)$, then

$$R(\hat{\mathbf{c}}_n) - R_k^* \leq 2M\mathcal{W}_2(P, P_n) + \frac{kM^2}{n}$$

Dereich, Scheutzow, Schottstedt

If $d \geq 4$,

$$\mathbb{E}\mathcal{W}_2(P, P_n) \lesssim Mn^{-\frac{1}{d}}.$$

Combinatorial bound

Chaining of $\mathcal{B}_d(0, M)$ + Dudley's entropy bound

$$\Downarrow$$

Linder

If $Supp(P) \subset \mathcal{B}_d(0, M)$, then

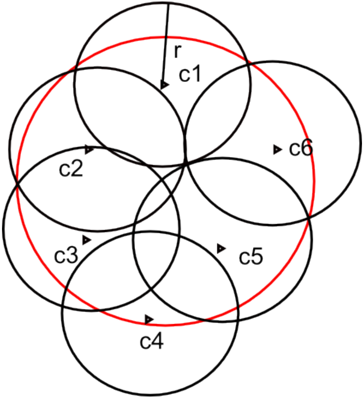
$$\mathbb{E}\left(R(\hat{\mathbf{c}}_n) - R_k^*\right) \lesssim M^2 \sqrt{\frac{kd}{n}}$$

Combinatorial bound

Sketch of proof

Goal:

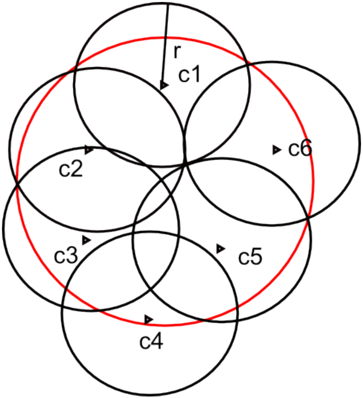
$$R_n := \mathbb{E} \sup_{\mathbf{c} \in \mathcal{B}_d(0,M)^k} (P - P_n) \min_{i=1,\dots,k} \|x - c_i\|^2$$



$\mathcal{C}_r$ r-covering

$$|\mathcal{C}_r| \lesssim r^{-kd}$$

Combinatorial bound



$$R_n \leq \mathbb{E} \sup_{\mathbf{c} \in r-\text{net}} (P - P_n) \min_{i=1,\dots,k} \|x - c_i\|^2 + CM^2 r$$

Combinatorial bound

$$R_n \leq \mathbb{E} \sup_{\mathbf{c} \in r-\text{net}} (P - P_n) \min_{i=1,\dots,k} \|x - c_i\|^2 + CM^2 r$$

Individual concentration:

$$\mathbb{P}\left((P - P_n) \min_{i=1,\dots,k} \|x - c_i\|^2 \geq t \right) \leq e^{-C\frac{nt^2}{M^2}}$$

Sub-Gaussian maximal inequality:

$$\mathbb{E} \sup_{\mathbf{c} \in r-\text{net}} (P - P_n) \min_{i=1,\dots,k} \|x - c_i\|^2 \lesssim M^2 \sqrt{\frac{\log(|\mathcal{C}_r|)}{n}} \approx M^2 \sqrt{\frac{kd \log(1/r)}{n}}$$

$\rightarrow r \approx \sqrt{\frac{kd}{n}}$ gives almost the desired result.

Euclidean bound

Comparison Theorem + Cauchy-Schwarz

$$\Downarrow$$

Biau

If $Supp(P) \subset \mathcal{B}_d(0, M)$, then

$$\mathbb{E}\left(R(\hat{\mathbf{c}}_n) - R_k^*\right) \lesssim \frac{kM^2}{\sqrt{n}}$$

Euclidean bound

Sketch of Proof

Almost same goal

$$R_n := \mathbb{E} \sup_{\mathbf{c} \in \mathcal{B}_d(0,M)^k} (P - P_n) \min_{i=1,\dots,k} -2\langle x, c_i \rangle + \|c_i\|^2$$

Symmetrization

$$R_n \lesssim R_{n,g} := \mathbb{E}_{X,g} \sup_{\mathbf{c} \in \mathcal{B}_d(0,M)^k} \frac{1}{n} \sum_{j=1}^{n} g_j \left(\min_{i=1,\dots,k} -2\langle X_j, c_i \rangle + \|c_i\|^2 \right)$$

Comparison

$$R_{n,g} \lesssim k \mathbb{E}_{X,g} \sup_{c \in \mathcal{B}_d(0,M)} \frac{1}{n} \sum_{j=1}^{n} g_j (-2\langle X_j, c \rangle + \|c\|^2)$$

Euclidean bound

$$R_{n,g} \lesssim k\mathbb{E}_{X,g} \sup_{c \in \mathcal{B}_d(0,M)} \frac{1}{n} \sum_{j=1}^{n} g_j(-2\langle X_j, c\rangle + \|c\|^2)$$

Cauchy-Schwarz

$$\begin{aligned} R_{n,g} &\lesssim kM\mathbb{E}_{X,g} \left\| \frac{1}{n} \sum_{j=1}^{n} g_j X_j \right\| + \frac{kM^2}{n} \mathbb{E}_g \sum_{j=1}^{n} |g_j| \\ &\lesssim \frac{kM^2}{\sqrt{n}} \end{aligned}$$

$\rightarrow$ easily extends to Hilbertian case

Dimension-free bounds

Biau

If $Supp(P) \subset \mathcal{B}(0, M)$, then

$$\mathbb{E}\left(R(\hat{\mathbf{c}}_n) - R_k^*\right) \lesssim \frac{kM^2}{\sqrt{n}}$$

Dimension-free bounds

Biau

If $Supp(P) \subset \mathcal{B}(0, M)$, then

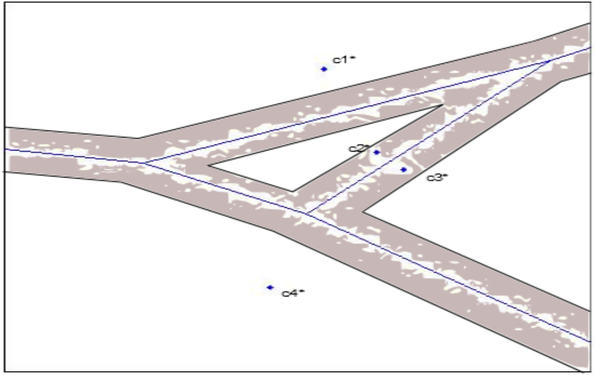
$$\mathbb{E}\left(R(\hat{\mathbf{c}}_n) - R_k^*\right) \lesssim \frac{kM^2}{\sqrt{n}}$$

L

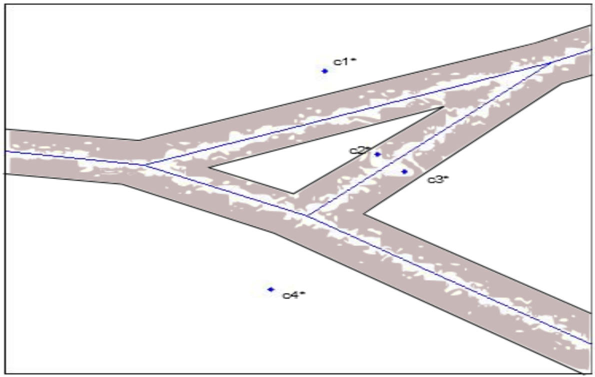
If $Supp(P) \subset \mathcal{B}(0, M)$ + Margin condition, then

$$\mathbb{E}\left(R(\hat{\mathbf{c}}_n) - R_k^*\right) \lesssim C(k, P)\frac{kM^2}{n}$$

Margin condition



Margin condition

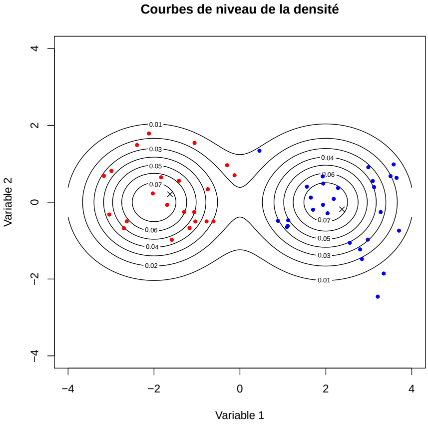


$\rightarrow$ "equivalent" to $R(\mathbf{c}) - R(\mathbf{c}^*) \approx \|\mathbf{c} - \mathbf{c}^*\|^2$

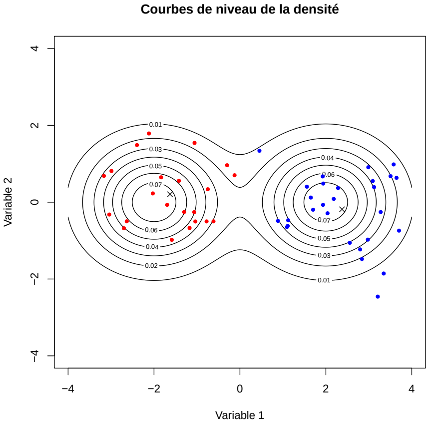
Partie 4

1. Introduction
   - Quantization principle
   - Quantizers/Risk
   - ERM strategy
2. Approximation error
   - Finite-dimensional spaces
   - Infinite-dimensional spaces
3. Estimation error
   - Finite-dimensional case
   - Hilbertian case
4. Variable selection

Motivation


Courbes de niveau de la densité
Variable 1
Variable 2
0.01
0.02
0.03
0.04
0.05
0.06
0.07

Motivation

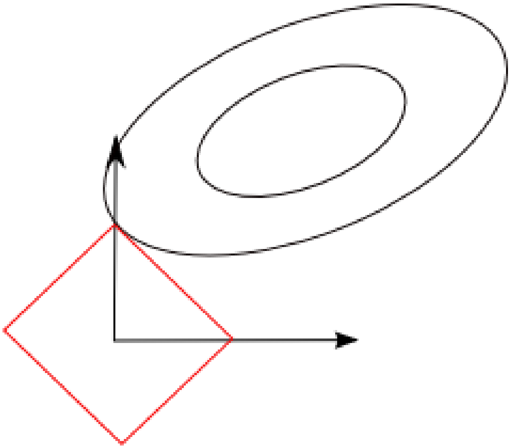


$\rightarrow$ $\hat{\mathbf{c}}_n^{(2)} := (\hat{c}_1^{(2)}, \dots, \hat{c}_k^{(2)}) \neq 0$

$\rightarrow$ sparsify $\hat{\mathbf{c}}_n$?

Variable selection procedure

$$\hat{\mathbf{c}}_{n,\lambda} = \arg\min_{\mathbf{c}} P_n \min_{i=1,\ldots,k} \|x - c_i\|^2 + \lambda \sum_{p=1}^{d} \sqrt{c_1^{(p)2} + \ldots + c_k^{(p)2}}$$



Variable selection procedure

$$\hat{\mathbf{c}}_{n,\lambda} = \arg\min_{\mathbf{c}} P_n \min_{i=1,\ldots,k} \|x - c_i\|^2 + \lambda \sum_{p=1}^{d} \sqrt{c_1^{(p)2} + \ldots + c_k^{(p)2}}$$

$\rightarrow$ Theoretical results, close to

$$\mathbf{c}^*_\lambda = \arg\min_{\mathbf{c}} 3R(\mathbf{c}) + C(P,k)\lambda^2 \|\mathbf{c}\|_0,$$

if

$$\lambda \gtrsim M_\infty \sqrt{\frac{k\log(kd)}{n}}$$

Discarded variables

$$\mathbf{c}_\lambda^* = \arg\min_{\mathbf{c}} 3R(\mathbf{c}) + C(P,k)\lambda^2\|\mathbf{c}\|_0.$$

Sufficient condition

If

$$\sigma_p^2 - R_p^* < C(P,k)\lambda^2/3,$$

then

$$\mathbf{c}_\lambda^{*(p)} = (0,\ldots,0).$$

Experimental results

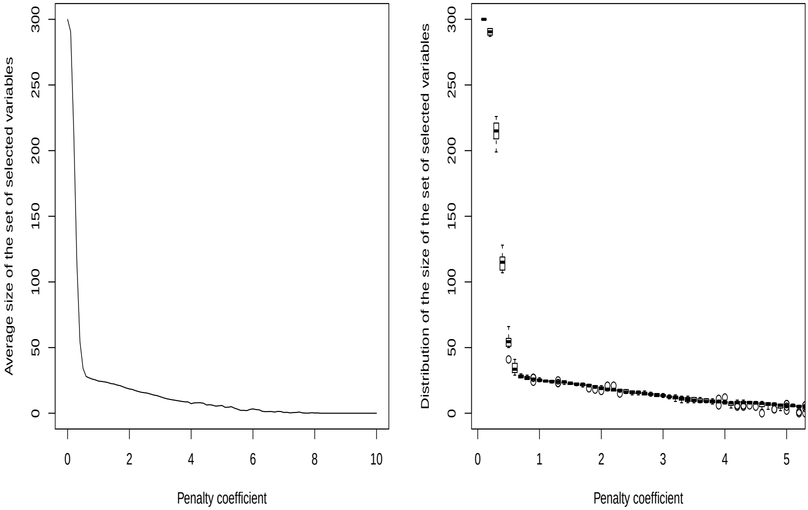
Model

$$\begin{cases} \mu_1 &= (3, 2.9, 2.8, \ldots, 0.1, \underbrace{0, \ldots, 0}_{270}), \\ \mu_2 &= 0, \\ \mu_3 &= -\mu_1, \\ \mu_4 &= (3, -2.9, 2.8, \ldots, -0.1, \underbrace{0, \ldots, 0}_{270}), \end{cases}$$

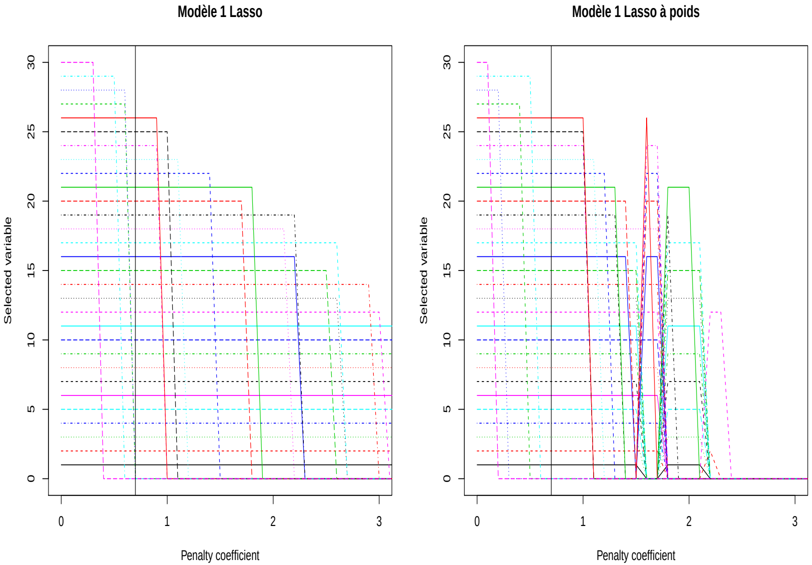
Gaussian mixture

$$P = 0.3\mathcal{N}(\mu_1, I_{300}) + 0.2\mathcal{N}(\mu_2, I_{300}) + 0.2\mathcal{N}(\mu_3, I_{300}) + 0.3\mathcal{N}(\mu_4, I_{300})$$

Experimental results



Experimental results



Conclusions

- Different impacts of dimension in quantization issues
  - → Approximation: $k^{-2/d}$, d "natural" dimension
  - → Estimation: dimension plays a role through the M^2 term only
  - → Variable selection: a "statistical" way to reduce dimension

- Some blanks left
  - → Approximation: Almost no results in Hilbert spaces
  - → Estimation: Dimension-free lower bounds are missing
  - → Variable selection: among others, a way to calibrate λ