Average performance of the sparsest approximation using a general dictionary

Mila Nikolova

CMLA (ENS Cachan and CNRS) PRES UniverSud, France

http://www.cmla.ens-cachan.fr/fileadmin/Membres/nikolova/

nikolova@cmla.ens-cachan.fr

Joint work with Franois Malgouyres

LAGA/L2TI, Université Paris 13, France

http://www.math.univ-paris13.fr/ malgouy/

malgouy@math.univ-paris13.fr

Mathematical and Algorithmical Challenges for Modeling and Analyzing Modern Data Sets

21-25 April 2008 HKBU - Hong Kong

Our Problem

Data $\boldsymbol{d} \in \mathrm{R}^N$ assumed uniform on $\mathcal{L}_{f_d}(\theta) = \{v \in \mathrm{R}^N, f_d(v) \leq \theta\}$

Dictionary $(\boldsymbol{\psi_i})_{\boldsymbol{i \in I}}$, $\{\psi_i : i \in I\} = \mathrm{R}^N$ (e.g. any frame or basis)

The most economical way to represent $\boldsymbol{d} \approx \sum \boldsymbol{\lambda_i \psi_i}$ = solving

$$(\mathcal{P}_d) : \quad \begin{cases} \min_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_0 \\ \text{under the constraint : } \quad \left\| \sum_{\boldsymbol{i \in I}} \boldsymbol{\lambda_i \psi_i} - \boldsymbol{d} \right\| \leq \boldsymbol{\tau} \end{cases}$$

$\|\boldsymbol{\lambda}\|_0 \overset{\text{def}}{=} \#\{\boldsymbol{i \in I : \lambda_i \neq 0}\}$ Non-linear approximation

$\|.\|$ and f_d — norms to choose by the user

θ and τ — parameters to choose by the user

Measure the obtaining of an (exactly) K-sparse solution

$\equiv$ $\mathrm{val}(\mathcal{P}_d) \leq \boldsymbol{K}$ = length (code ($\boldsymbol{d}$))

$\mathrm{val}(\mathcal{P}_d) = \|\lambda\|_0$ for a solution λ of $(\mathcal{P}_d)$ and for $K = 0, \ldots, N$, $\tau > 0$.

Related problems

Minimum description length principle of Rissanen 1967

$$\|\lambda\|_0 = \sum_{i \in I} \varphi(\lambda_i) \quad \text{where} \quad \varphi(t) = \begin{cases} 0 & \text{if} \quad t = 0 \\ 1 & \text{if} \quad t \neq 0 \end{cases}$$

Maximum a posteriori (MAP) energies with Markov-random field prior (signals, images)

$$(\boldsymbol{\mathcal{E}_d}) \qquad \boldsymbol{\mathcal{E}(\lambda) = \|\Psi\lambda - d\|_2^2 + \beta \sum_{i \sim j} \varphi(\lambda_i - \lambda_j)}, \quad \text{where } i \sim j \text{ neighbors}$$

- λ = image composed of labels $\Rightarrow$ Potts model *[Geman^2 1984, Besag 1986,...]*
- $\lambda \in \mathrm{R}^N$—since *[Leclerc 1989]*
- Theoretical results on $\boldsymbol{\arg\min \mathcal{E}(\lambda)}$ in *[Nikolova 2005]*
- Fast optimization scheme *[Robini at al 2008]*
- $(\mathcal{P}_d)$ and $(\mathcal{E}_d)$ related but no equivalent in general

Hard thresholding to denoise wavelet coefficients *[Donoho & Johnstone 1992]*

$$\text{minimize } \boldsymbol{\|\lambda_i - g_i\|_2^2 + \beta\phi(\lambda_i), \ \forall i \in I}$$

where noisy coefficients $\boldsymbol{g_i = \langle \psi_i^*, d\rangle, \forall i \in I}$

- Amounts to a special case of $(\mathcal{E}_d)$

Nonlinear approximation theory (*K best term approximation.*) *(Review—see [DeVore 1998])*

Performance measured by $\inf_{J \subset I, \alpha \in \mathrm{R}^J} \left\{ \|d - \sum_{i \in J} \alpha_i \psi_i\| \ : \ \text{subjsct to} \ \dim(\mathrm{span}(\psi_i)_{i \in J} = K \right\}$

Results depend on the regularity of data d but independent of their distribution. Performance evaluation often depend on pessimistic constants.

Our main focus: **Compression of large data sets (e.g. images)**

Build a coder such that $\mathbb{P}\left(\textit{length} \left(\textit{code} \left(\boldsymbol{d} \right)\right)\right) \begin{cases} \textit{large} \ \text{ if } \boldsymbol{K} \text{ is small} \\ \textit{small} \ \text{ if } \boldsymbol{K} \text{ is large} \end{cases}$

- *E.g., JPEG and JPEG200 use basis*
- *Much more can be obtained if $(\psi_i)_{i \in I}$ is redundant (e.g. dictionaries)*

Alternative formulations

$$\min_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_{\boldsymbol{p}} \quad \textit{subject to} \quad \left\| \sum_{\boldsymbol{i} \in \boldsymbol{I}} \boldsymbol{\lambda_i} \boldsymbol{\psi_i} - \boldsymbol{d} \right\| \le \boldsymbol{\tau}, \quad \boldsymbol{p} \in [\mathbf{0}, \mathbf{1}], \quad \boldsymbol{\tau} \ge \mathbf{0}$$

e.g. Compressed sensing: $p = 1$ and $\tau = 0$ (=no noise) + numerous restrictions

Outline of our work

$$(\mathcal{P}_d) : \qquad \min_{\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\|_0 \qquad \text{subject to :} \qquad \left\| \sum_{i \in I} \boldsymbol{\lambda}_i \psi_i - \boldsymbol{d} \right\| \leq \boldsymbol{\tau}$$

1. The basic concept
2. Measure of cylinder-like bounded subsets
3. Measure of the intersection of such sets
4. Sets of data yielding K-sparse solutions
5. Sets of data yielding exactly K-sparse solutions
6. Statistical meaning
7. Example - Euclidian norms and orthogonal basis
8. Open questions

all our results hold $\forall \|.\|, \quad \forall \quad f_d, \quad \forall (\psi_i)_{i \in I}$

1. The Basic Idea

For any $J \subset I$ define $\mathcal{T}_J = \textbf{span}\left\{(\psi_j)_{j \in J}\right\}$

The **θ-level set of f**: $\mathrm{R}^N \to \mathrm{R}$, for any $\theta \in \mathrm{R}$ is $\mathcal{L}_f(\theta) = \{u \in \mathbf{R}^N, f(u) \leq \theta\}.$

Theorem.

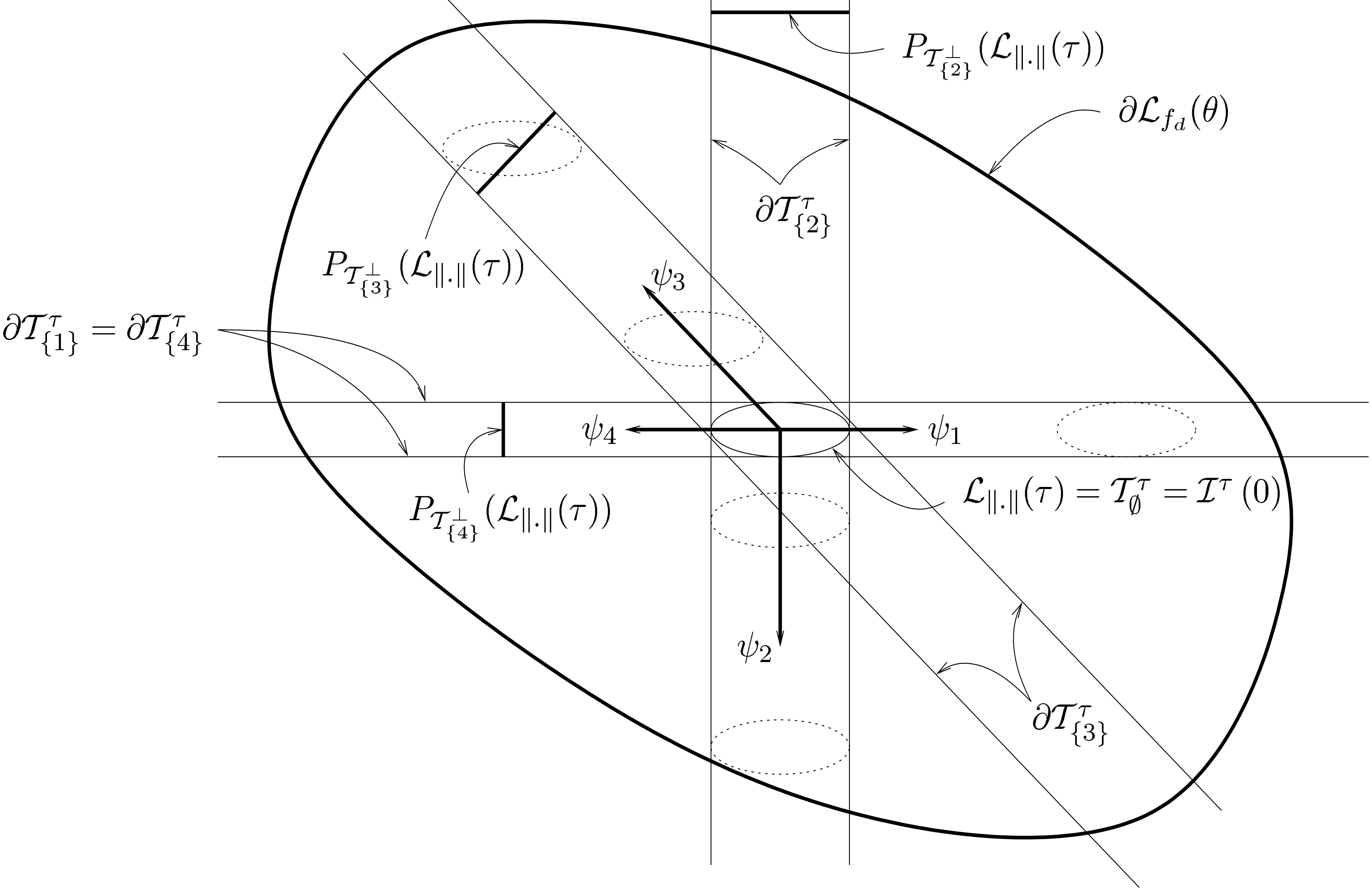
$$\mathcal{I}^\tau(K) \stackrel{\text{def}}{=} \left\{d \in \mathbf{R}^N, \mathrm{val}(\mathcal{P}_d) \leq K\right\} = \bigcup \left\{\mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau) \quad : \quad \dim \mathcal{T}_J = K\right\}$$

all data $\Rightarrow$ solutions with $\leq K$ coefficients $\neq \mathbf{0}$ (i.e. K-sparse)

$$\mathcal{D}^\tau(K) \stackrel{\text{def}}{=} \left\{d \in \mathbf{R}^N, \mathrm{val}(\mathcal{P}_d) = K\right\} = \mathcal{I}^\tau(K) - \mathcal{I}^\tau(K-1)$$

all data yielding exactly K-sparse solutions

We derive tight Upper/Lower bounds for $\mathcal{I}^\tau(K)$ and $\mathcal{D}^\tau(K)$



Derivations:

measure $\bigcup \left\{ \text{all } \underline{\text{different}}\ \mathcal{T}_J + \mathcal{L}_{\|.\|}(\tau) \right\}$ and subtract their multiple intersections

In practice, it is enough we subtract only intersections of order two.

Preliminary results

$\delta_1 > 0$ and $\delta_2 > 0$ defined by: $f_d(w) \leq \delta_1 \|w\|_2$ & $\|w\|_2 \leq \delta_2 \|w\|$, $\forall w \in \mathbf{R}^N$

$\overline{\Delta} \stackrel{\text{def}}{=} \delta_1 \delta_2$ These constants are independent of the subspace chosen

For any $J \subset I$ and $\tau > 0$: $\begin{cases} \mathcal{T}_J = \text{span}\left\{(\psi_j)_{j \in J}\right\} & \text{vector subspace} \\ \mathcal{T}_J^\tau = \mathcal{T}_J + P_{\mathcal{T}_J^\perp}\left(\mathcal{L}_{\|.\|}(\tau)\right) & \text{cylinder-like subset} \end{cases}$

$K \stackrel{\text{def}}{=} \dim(\mathcal{T}_J)$ $P =$ projection

$(\psi_i)_{i \in I}$ redundant $\Rightarrow$ $\mathcal{T}_J$ can be "any" vector subspace

$$h(u) \stackrel{\text{def}}{=} \inf \left\{ t \geq 0 : \frac{u}{t} \in P_{\mathcal{T}_J^\perp}\left(\mathcal{L}_{\|.\|}(1)\right) \right\}$$

Lemma. *h is a norm on ${\mathcal{T}_J}^\perp$ and $\mathcal{L}_h(\tau) = P_{\mathcal{T}_J^\perp}\left(\mathcal{L}_{\|.\|}(\tau)\right)$*

2. Measure of bounded cylinder-like sets

Proposition.

$\forall J \subset I, \ \ \forall \|.\|, \ \ \forall \tau > 0, \ \ \ \mathcal{T}_J$ is closed and measurable

$\forall f_d$ (norm) $\ \exists \delta_J \in [0, \overline{\Delta}]$ such that if $\theta > \delta_J \tau$ then

$$C_J \tau^{N-K} (\theta - \delta_J \tau)^K \ \leq \ \mathbb{L}^N\big(\mathcal{T}_J^\tau \cap \mathcal{L}_{f_d}(\theta)\big) \ \leq \ C_J \tau^{N-K} (\theta + \delta_J \tau)^K$$

where $\ C_J = \mathbb{L}^{N-K}\big(P_{\mathcal{T}_J^\perp}\,(\mathcal{L}_{\|.\|}(1))\,\big)\ \mathbb{L}^K\big(\mathcal{T}_J \cap \mathcal{L}_{f_d}(1)\big) \in (0, +\infty)$

Sketch of the proof.

$$\begin{aligned} \varphi_0 : \mathcal{L}_h(1) \times \big(V \cap \mathcal{L}_{f_d}(1)\big) &\to \ \mathrm{R}^N \\ (u, v) &\to \ \tau u + (\theta - \delta\tau) v \\ \varphi_0\Big(\mathcal{L}_h(1) \times \big(V \cap \mathcal{L}_{f_d}(1)\big)\Big) &= \ \ B_0 \end{aligned}$$

$$\begin{aligned} \varphi_1 : \mathcal{L}_h(1) \times \big(V \cap \mathcal{L}_{f_d}(1)\big) &\to \ \mathrm{R}^N \\ (u, v) &\to \ \tau u + (\theta + \delta\tau) v \\ \varphi_1\Big(\mathcal{L}_h(1) \times \big(V \cap \mathcal{L}_{f_d}(1)\big)\Big) &= \ \ B_1 \end{aligned}$$

$$B_0 \subset \Big(\mathcal{T}_J^\tau \cap \mathcal{L}_{f_d}(\theta)\Big) \subset B_1$$

φ_0 and φ_1 are Lipschitz homeomorphisms $\ \Rightarrow\ $ we can calculate the Lebesgue measure of B_0 and B_1

$$D\varphi_0 = \begin{bmatrix} \tau I_{N-K} & 0 \\ 0 & (\theta - \delta\tau) I_K \end{bmatrix} \quad \text{and} \quad D\varphi_1 = \begin{bmatrix} \tau I_{N-K} & 0 \\ 0 & (\theta + \delta\tau) I_K \end{bmatrix}$$

Their Jacobians: $[\![\varphi_0]\!] = \det\left(D\varphi_0\right) = \tau^{N-K}(\theta - \delta\tau)^K$ and $[\![\varphi_1]\!] = \det\left(D\varphi_1\right) = \tau^{N-K}(\theta + \delta\tau)^K$.

See [Evans & Gariepi 92]:

$$\mathbb{L}^N\left(B_i\right) = \int_{u \in \mathcal{L}_h(1) \subset \mathcal{T}_J{}^\perp} \int_{v \in \mathcal{T}_J \cap \mathcal{L}_{f_d}(1)} [\![\varphi_i]\!] dv du$$

$$\mathbb{L}^N\left(B_0\right) = C_J \tau^{N-K}(\theta - \delta\tau)^K \quad \text{and} \quad \mathbb{L}^N\left(B_1\right) = C_J \tau^{N-K}(\theta + \delta\tau)^K$$

$$C_J = \int_{\mathcal{L}_h(1) \subset \mathcal{T}_J{}^\perp} du \int_{\mathcal{T}_J \cap \mathcal{L}_{f_d}(1)} dv \qquad \in (0, \infty)$$

Using that $\mathbb{L}^N\left(B_0\right) \leq \mathbb{L}^N\left(\mathcal{T}_J^\tau \cap \mathcal{L}_{f_d}(\theta)\right) \leq \mathbb{L}^N\left(B_1\right)$ yields the result.

Lemma. $\boxed{\mathcal{T}_J^\tau = \mathcal{T}_J + \mathcal{L}_{||.||}(\tau)}$ (easier to use!)

3. Measure of the intersection of such sets

The intersection of two cylinder-like bounded subsets is small and becomes negligible as τ/θ decreases. The theorem corroborates this intuition.

Proposition.

$$J_1 \subset I, \quad J_2 \subset I \quad : \quad \mathcal{T}_{J_1} \neq \mathcal{T}_{J_2}, \quad \dim(\mathcal{T}_{J_1}) = \dim(\mathcal{T}_{J_2}) \stackrel{\text{def}}{=} K$$

For $\tau > 0,\ \theta > 0,\ \mathcal{T}_{J_1}^{\tau} \cap \mathcal{T}_{J_2}^{\tau} \cap \mathcal{L}_{f_d}(\theta)$ is closed and measurable.

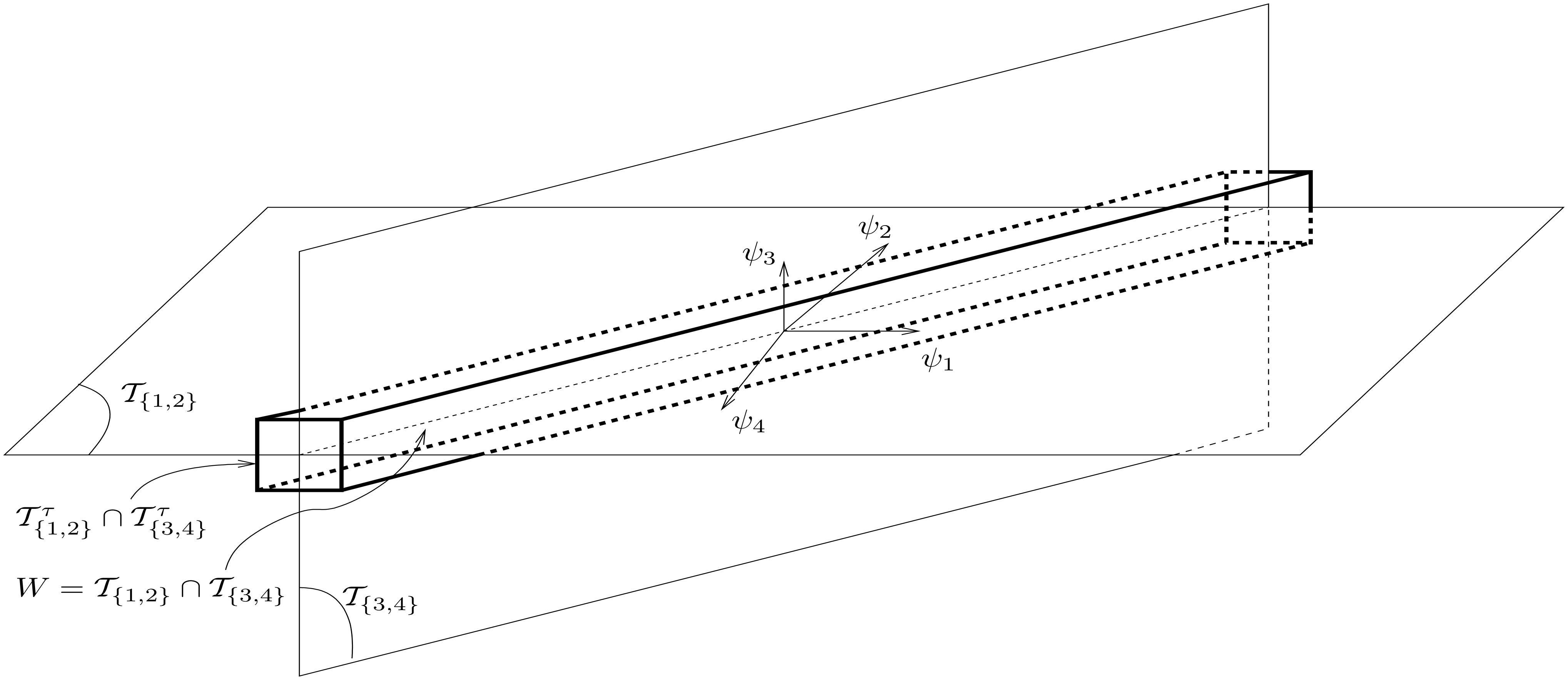
$\exists\ \delta_{J_1,J_2} \in [0, 3\overline{\Delta}]$: if $\theta > \delta_{J_1,J_2}\tau \quad \Rightarrow$

$$\mathbb{L}^N\big(\mathcal{T}_{J_1}^{\tau} \cap \mathcal{T}_{J_2}^{\tau} \cap \mathcal{L}_{f_d}(\theta)\big) \leq Q_{J_1,J_2}\tau^{N-k}(\theta + \delta_{J_1,J_2}\tau)^k, \qquad k = \dim\left(\mathcal{T}_{J_1} \cap \mathcal{T}_{J_2}\right)$$

$$Q_{J_1,J_2} \stackrel{\text{def}}{=} \mathbb{L}^{N-k}\big(W^{\perp} \cap \mathcal{L}_{\|.\|_2}(2\delta_2)\big)\ \mathbb{L}^k\big(W \cap \mathcal{L}_{f_d}(1)\big) \quad \text{for} \quad W \stackrel{\text{def}}{=} \mathcal{T}_{J_1} \cap \mathcal{T}_{J_2}$$

Notice: Q_{J_1,J_2} depends only on $\|.\|$, f_d, $(\psi_i)_{i\in J_1}$, $(\psi_i)_{i\in J_2}$

A tighter bound is found in the proof



1

4. Sets of data yielding K-sparse solutions

Remind: $\boldsymbol{\mathcal{I}^\tau(K)} \stackrel{\text{def}}{=} \left\{\boldsymbol{d} \in \mathbf{R}^{\boldsymbol{N}}, \mathbf{val}(\boldsymbol{\mathcal{P}_d}) \leq \boldsymbol{K}\right\}$

(data leading to a K-sparse solution)

Define: $\boldsymbol{G_K} \stackrel{\textbf{def}}{=} \{\boldsymbol{J} \subset \boldsymbol{I} : \mathbf{dim}(\boldsymbol{\mathcal{T}_J}) \leq \boldsymbol{K}\}$

Proposition. (Surprising)

$$\forall \boldsymbol{K} \in \{\mathbf{0}, \dots, \boldsymbol{N}\}, \quad \forall \|.\|, \quad \forall \boldsymbol{\tau} > \mathbf{0} \quad \Rightarrow \quad \boldsymbol{\mathcal{I}^\tau(K)} = \bigcup_{\boldsymbol{J} \in \boldsymbol{G_K}} \boldsymbol{\mathcal{T}_J} + \boldsymbol{\mathcal{L}}_{\|.\|}(\boldsymbol{\tau})$$

Notations to get a non-redundant listing of $\left\{\boldsymbol{\mathcal{T}_J} : \boldsymbol{J} \in \boldsymbol{I}\right\}$

$\forall \boldsymbol{K} = \mathbf{0}, \dots, \boldsymbol{N}$, define $\boldsymbol{\mathcal{J}(K)}$ by:

$$\begin{cases} \boldsymbol{\mathcal{J}(K)} \subset \{\boldsymbol{J} \subset \boldsymbol{I} : \mathbf{dim}(\boldsymbol{\mathcal{T}_J}) = \boldsymbol{K}\} \\ \boldsymbol{J_1}, \boldsymbol{J_2} \in \boldsymbol{\mathcal{J}(K)} \ \text{ and } \ \boldsymbol{J_1} \neq \boldsymbol{J_2} \quad \Longrightarrow \quad \boldsymbol{\mathcal{T}_{J_1}} \neq \boldsymbol{\mathcal{T}_{J_2}} \\ \boldsymbol{\mathcal{J}(K)} \ \textbf{is maximal:} \text{ if } \boldsymbol{J_1} \subset \boldsymbol{I} \text{ and } \mathbf{dim}(\boldsymbol{\mathcal{T}_{J_1}}) = \boldsymbol{K} \ \text{ then } \ \exists \boldsymbol{J} \in \boldsymbol{\mathcal{J}(K)} \ : \ \boldsymbol{\mathcal{T}_J} = \boldsymbol{\mathcal{T}_{J_1}} \end{cases}$$

Notice: $\mathcal{J}(0) = \{\emptyset\}, \ \#\mathcal{J}(N) = 1, \ G_K \supset \bigcup_{k=0}^{K} \mathcal{J}(k), \ \{\mathcal{T}_J : J \in G_K\} = \{\mathcal{T}_J : J \in \mathcal{J}(k), \ k = 0, \dots, K\}$

Theorem. $$\boxed{\mathcal{I}^{\tau}(K) = \bigcup_{J \in \mathcal{J}(K)} \mathcal{T}_J^{\tau}} \quad \Rightarrow \quad \mathcal{I}^{\tau}(K)\text{—closed and measurable}$$

Notations—continued

$\forall \; K = 0, \ldots, N$

$$\hat{\delta}_K = \max_{J \in \mathcal{J}(K)} \delta_J \in [0, \overline{\Delta}] \qquad \overline{C}_K = \sum_{J \in \mathcal{J}(K)} C_J$$

e.g. $\overline{C}_0 = \mathbb{L}^N(\mathcal{L}_{\|.\|}(1)), \;\; \overline{C}_N = \mathbb{L}^N(\mathcal{L}_{f_d}(1))$

$$\mathcal{H}(K,k) = \Big\{ (J_1, J_2) \in \mathcal{J}(K)^2 \; : \; \dim(\mathcal{T}_{J_1} \cap \mathcal{T}_{J_2}) = k \Big\}, \quad 0 \le k \le K - 1$$

$$k_K \stackrel{\text{def}}{=} \max\{0, \; 2K - N\}$$

Observe: $\quad k < k_K \quad \Rightarrow \quad \mathcal{H}(K,k) = \emptyset$

Over-estimations: $\quad \#\mathcal{J}(K) \le \dfrac{\#I!}{K!(\#I - K)!}$ and $\#\mathcal{H}(K,k) \le \#\mathcal{J}(K)\#\mathcal{J}(K-1), \; \forall k < K$

$$\hat{\delta}'_{K,k} = \max\left\{0, \max_{(J_1,J_2)\in\mathcal{H}(K,k)} \delta_{J_1,J_2}\right\} \in [0, 3\overline{\Delta}] \qquad \overline{Q}_{K,k} = \sum_{(J_1,J_2)\in\mathcal{H}(K,k)} Q_{J_1,J_2}$$

$$\Delta_K = \left\{\begin{array}{ll} \hat{\delta}_0 & \text{if } K = 0 \\ \max\left\{\Delta_{K-1}, \hat{\delta}_K, \max_{k_K\le k\le K-1} \hat{\delta}'_{K,k}\right\} & \text{if } 0 < K < N \\ \max\left\{\Delta_{N-1}, \hat{\delta}_N\right\} & \text{if } K = N \end{array}\right\} \in [0,\ 3\overline{\Delta}]$$

All these constants depend only on $(\psi_i)_{i\in I}$, $\|.\|$, f_d, K, k

Theorem.

$\forall K = 0, \ldots, N$, $\forall \tau > 0$ and $\theta \geq \tau\Delta_K \quad \Rightarrow$

$$\overline{C}_K \tau^{N-K} (\theta - \hat{\delta}_K\tau)^K - \theta^N \varepsilon_0(K,\tau,\theta) \leq \mathbb{L}^N\left(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\right) \leq \overline{C}_K \tau^{N-K} (\theta + \hat{\delta}_K\tau)^K$$

$$\varepsilon_0(K,\tau,\theta) = \sum_{k=k_K}^{K-1} \overline{Q}_{K,k} \left(\frac{\tau}{\theta}\right)^{N-k} \left(1 + \hat{\delta}'_{K,k}\frac{\tau}{\theta}\right)^k, \ 0 < K < N, \qquad \varepsilon_0 = 0 \text{ else}$$

5. Sets of data yielding exactly K-sparse

Remind: $$\mathcal{D}^\tau(K) = \left\{ d \in \mathbf{R}^N : \mathrm{val}(\mathcal{P}_d) = K \right\}$$

(data leading to exactly K-sparse solutions)

Notice: $\mathcal{D}^\tau(K) = \mathcal{I}^\tau(K) \setminus \mathcal{I}^\tau(K-1), \;\; 0 \le K \le N, \;\; \mathcal{I}^\tau(-1) \overset{\mathrm{def}}{=} \emptyset$ then

$$\mathbb{L}^N\big(\mathcal{D}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\big) = \mathbb{L}^N\big(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\big) - \mathbb{L}^N\big(\mathcal{I}^\tau(K-1) \cap \mathcal{L}_{f_d}(\theta)\big)$$

Theorem.

$$\theta \ge \tau \max(\Delta_K, \Delta_{K-1}) > 0 \quad \Rightarrow$$

$$\overline{C}_K \tau^{N-K}(\theta - \hat{\delta}_K \tau)^K - \theta^N \varepsilon_0'(K,\tau,\theta) \le \mathbb{L}^N\big(\mathcal{D}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\big) \le \overline{C}_K \tau^{N-K}(\theta + \hat{\delta}_K \tau)^K + \theta^N \varepsilon_1(K,\tau,\theta)$$

$$\varepsilon_0'(K,\tau,\theta) = \varepsilon_0(K,\tau,\theta) + \overline{C}_{K-1}\left(\tfrac{\tau}{\theta}\right)^{N-(K-1)}\left(1 + \hat{\delta}_{K-1}\tfrac{\tau}{\theta}\right)^{K-1}$$

$$\varepsilon_1(K,\tau,\theta) = \varepsilon_0(K-1,\tau,\theta) - \overline{C}_{K-1}\left(\tfrac{\tau}{\theta}\right)^{N-(K-1)}\left(1 - \hat{\delta}_{K-1}\tfrac{\tau}{\theta}\right)^{K-1}$$

6. Statistical meaning

Theorem.

$\forall K = 0, \ldots N, \quad \theta \geq \tau \Delta_K > 0$ if d uniform on $\mathcal{L}_{f_d}(\theta) \quad \Rightarrow$

$$\frac{\overline{C}_K}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \left(\frac{\tau}{\theta}\right)^{N-K} \left(1 - \hat{\delta}_K \frac{\tau}{\theta}\right)^K - \frac{\varepsilon_0(K, \tau, \theta)}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \leq \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K\right)$$

$$\leq \frac{\overline{C}_K}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \left(\frac{\tau}{\theta}\right)^{N-K} \left(1 + \hat{\delta}_K \frac{\tau}{\theta}\right)^K$$

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K\right) = \frac{\overline{C}_K}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \left(\frac{\tau}{\theta}\right)^{N-K} + o\left(\left(\frac{\tau}{\theta}\right)^{N-K}\right) \quad \text{as } \frac{\tau}{\theta} \to 0$$

Theorem.

$$\forall K = 1, \ldots, N, \quad \theta \geq \tau \max(\Delta_K, \Delta_{K-1}) > 0, \quad \text{and} \quad d \text{ uniform on } \mathcal{L}_{f_d}(\theta) \quad \Rightarrow$$

$$\frac{\overline{C}_K}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \left(\frac{\tau}{\theta}\right)^{N-K} \left(1 - \hat{\delta}_K \frac{\tau}{\theta}\right)^K - \varepsilon^-(K, \tau, \theta) \leq \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) = K\right)$$
$$\leq \frac{\overline{C}_K}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \left(\frac{\tau}{\theta}\right)^{N-K} \left(1 + \hat{\delta}_K \frac{\tau}{\theta}\right)^K + \varepsilon^+(K, \tau, \theta)$$

$$\varepsilon^-(K, \tau, \theta) = \frac{\varepsilon_0'(K, \tau, \theta)}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \quad \text{and} \quad \varepsilon^+(K, \tau, \theta) = \frac{\varepsilon_1(K, \tau, \theta)}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))}$$

$$\mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) = K\right) = \frac{\overline{C}_K}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \left(\frac{\tau}{\theta}\right)^{N-K} + o\left(\left(\frac{\tau}{\theta}\right)^{N-K}\right) \quad \text{as} \quad \frac{\tau}{\theta} \to 0$$

$$\mathbb{E}\left(\mathrm{val}(\mathcal{P}_d)\right) = \sum_{K=1}^{N} K\ \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) = K\right) = N - \sum_{K=0}^{N-1} \mathbb{P}\left(\mathrm{val}(\mathcal{P}_d) \leq K\right)$$

Theorem.

If $\boldsymbol{\theta} \geq \boldsymbol{\tau} \max_{0 \leq K \leq N} \boldsymbol{\Delta}_K > 0$ and $\boldsymbol{d}$ **uniform on** $\boldsymbol{\mathcal{L}}_{f_d}(\boldsymbol{\theta}) \Rightarrow$

$$N - \sum_{K=0}^{N-1} \frac{\overline{C}_K}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \left(\frac{\tau}{\theta}\right)^{N-K} \left(1 + \hat{\delta}_K \frac{\tau}{\theta}\right)^{K} \leq \mathbb{E}\left(\mathrm{val}(\mathcal{P}_d)\right)$$

$$\leq N - \sum_{K=0}^{N-1} \frac{\overline{C}_K}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \left(\frac{\tau}{\theta}\right)^{N-K} \left(1 - \hat{\delta}_K \frac{\tau}{\theta}\right)^{K} - \frac{\varepsilon_0(K, \tau, \theta)}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))}$$

$$\mathbb{E}\left(\mathrm{val}(\mathcal{P}_d)\right) = N - \frac{\overline{C}_{N-1}}{\mathbb{L}^N(\mathcal{L}_{f_d}(1))} \frac{\tau}{\theta} + o\left(\frac{\tau}{\theta}\right) \quad \text{as} \quad \frac{\tau}{\theta} \rightarrow 0$$

7. Example: Euclidian norms and orthogonal bases

Theorem.

$$\forall K = 0, \ldots, N, \quad \text{if} \;\; \theta \geq \sqrt{2}\tau > 0 \quad \Rightarrow$$

$$\mathcal{C}_N^K \mu(K)\, \tau^{N-K} \theta^K \left(\sqrt{1 - \left(\frac{\tau}{\theta}\right)^2}\right)^K - \theta^N \varepsilon_0(K, \tau, \theta)$$

$$\leq \mathbb{L}^N\big(\mathcal{I}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\big) \leq \; \mathcal{C}_N^K \mu(K) \tau^{N-K} \theta^K$$

$$\underline{\varepsilon_0}(K, \tau, \theta) \leq \varepsilon_0(K, \tau, \theta) \leq \overline{\varepsilon_0}(K, \tau, \theta)$$

$$\underline{\varepsilon_0}(K, \tau, \theta) = \sum_{k=k_K}^{K-1} \#(\mathcal{H}(K, k))\mu(k) \left(\frac{\tau}{\theta}\right)^{N-k} \left(\sqrt{1 - \left(\frac{\tau}{\theta}\right)^2}\right)^k$$

$$\overline{\varepsilon_0}(K, \tau, \theta) = \sum_{k=k_K}^{K-1} \#(\mathcal{H}(K, k))\mu(k) \left(\frac{\tau}{\theta}\right)^{N-k} 2^{\frac{N-k}{2}}$$

$$\mathcal{C}_N^K \frac{N!}{K!(N-K)!} \quad \mu(K) = \frac{4\pi^{\frac{N}{2}}}{K(N-K)\Gamma\left(\frac{N-K}{2}\right)\Gamma\left(\frac{K}{2}\right)} \quad \#\mathcal{H}(K, k) = \mathcal{C}_N^k \mathcal{C}_{N-k}^{K-k} \mathcal{C}_{N-K}^{K-k}$$

Theorem.

$$\forall K = 0,\dots,N, \quad \text{if} \;\; \theta \geq \sqrt{2}\tau > 0 \qquad \Rightarrow$$

$$\mathcal{C}_N^K \mu(K)\, \tau^{N-K}\theta^K \left(\sqrt{1-\left(\tfrac{\tau}{\theta}\right)^2}\right)^K - \theta^N \varepsilon_0'(K,\tau,\theta) \leq \mathbb{L}^N\big(\mathcal{D}^\tau(K) \cap \mathcal{L}_{f_d}(\theta)\big) \leq$$
$$\leq \mathcal{C}_N^K \mu(K)\tau^{N-K}\theta^K + \theta^N \varepsilon_1'(K,\tau,\theta)$$

$$\Delta(K-1,\tau,\theta) + \underline{\varepsilon_0}(K,\tau,\theta) \leq \varepsilon_0'(K,\tau,\theta) \leq \Delta(K-1,\tau,\theta) + \overline{\varepsilon_0}(K,\tau,\theta)$$

$$-\Delta(K-1,\tau,\theta)\left(\sqrt{1-\left(\tfrac{\tau}{\theta}\right)^2}\right)^{K-1} + \underline{\varepsilon_0}(K-1,\tau,\theta) \leq \varepsilon_1'(K,\tau,\theta)$$
$$\leq -\Delta(K-1,\tau,\theta)\left(\sqrt{1-\left(\tfrac{\tau}{\theta}\right)^2}\right)^{K-1} + \overline{\varepsilon_0}(K-1,\tau,\theta)$$

$$\Delta(K-1,\tau,\theta) = \mathcal{C}_N^{K-1}\mu(K-1)\left(\tfrac{\tau}{\theta}\right)^{N-(K-1)}$$

Theorem.

$$\forall \boldsymbol{K} = \mathbf{1}, \ldots, \boldsymbol{N}, \quad \text{if} \quad \boldsymbol{\theta} \geq \sqrt{\mathbf{2}}\boldsymbol{\tau} > \mathbf{0} \quad \text{and} \quad \boldsymbol{d} \quad \textbf{uniform on} \quad \boldsymbol{\mathcal{L}_{f_d}(\theta)} \quad \Rightarrow$$

$$\boldsymbol{\mathcal{C}_N^K \nu(K)} \left(\frac{\boldsymbol{\tau}}{\boldsymbol{\theta}}\right)^{\boldsymbol{N-K}} \left(\sqrt{1-\left(\frac{\tau}{\theta}\right)^2}\right)^K - \frac{\varepsilon_0(K,\tau,\theta)}{\alpha(N)} \leq \mathbb{P}\left(\mathbf{val}(\boldsymbol{\mathcal{P}_d}) \leq \boldsymbol{K}\right) \leq \boldsymbol{\mathcal{C}_N^K \nu(K)} \left(\frac{\boldsymbol{\tau}}{\boldsymbol{\theta}}\right)^{\boldsymbol{N-K}}$$

$$\boldsymbol{\mathcal{C}_N^K \nu(K)} \left(\frac{\boldsymbol{\tau}}{\boldsymbol{\theta}}\right)^{\boldsymbol{N-K}} \left(\sqrt{1-\left(\frac{\tau}{\theta}\right)^2}\right)^K - \frac{\varepsilon_0'(K,\tau,\theta)}{\alpha(N)} \leq \mathbb{P}\left(\mathbf{val}(\boldsymbol{\mathcal{P}_d}) = \boldsymbol{K}\right)$$

$$\leq \boldsymbol{\mathcal{C}_N^K \nu(K)} \left(\tfrac{\boldsymbol{\tau}}{\boldsymbol{\theta}}\right)^{\boldsymbol{N-K}} + \frac{\varepsilon_1'(K,\tau,\theta)}{\alpha(N)}$$

$$\boldsymbol{\nu(K)} = \frac{\boldsymbol{\alpha(K)\alpha(N-K)}}{\boldsymbol{\alpha(N)}}, \quad \boldsymbol{\alpha(n)} = \mathbb{L}^{\boldsymbol{n}}\left(\boldsymbol{\mathcal{L}}_{\|.\|_2}(\mathbf{1})\right) = \frac{\boldsymbol{\pi}^{\boldsymbol{n}/\mathbf{2}}}{\boldsymbol{\Gamma(n/2+1)}}$$

---

8. Conclusion and perspectives

- Practical (lower/upper) bounds for the obtention of K-sparse or sparser solutions:
  - Measure$\Big(\boldsymbol{\mathcal{I}}^{\tau}(\boldsymbol{K}) \cap \boldsymbol{\mathcal{L}}_{\boldsymbol{f_d}}(\boldsymbol{\theta})\Big)$ and Measure$\Big(\boldsymbol{\mathcal{D}}^{\tau}\ (\boldsymbol{K}) \cap \boldsymbol{\mathcal{L}}_{\boldsymbol{f_d}}(\boldsymbol{\theta})\Big)$ $\stackrel{\frac{\tau}{\theta}\to 0}{\longrightarrow}$ $\overline{\boldsymbol{C}}_K \theta^N \left(\frac{\tau}{\theta}\right)^{N-K}$
  - $\mathbb{P}\left(\mathbf{val}(\boldsymbol{\mathcal{P}_d}) \leq \boldsymbol{K}\right)$ and $\mathbb{P}\left(\mathbf{val}(\boldsymbol{\mathcal{P}_d}) = \boldsymbol{K}\right)$
- Evaluations for $\mathbb{E}\left(\mathrm{val}(\mathcal{P}_d)\right)$ (Follows $\overline{\boldsymbol{C}}_{N-1}$as $\frac{\tau}{\theta} \to 0$)
- All bounds can be calculated (numerically) for any particular choice for $(\psi_i)_{i\in I}$, $\|.\|$ and f_d, and τ and θ
- To get the sparsest approximation—tune $\|.\|$, f_d and $(\psi_i)i \in I$ to maximize $\overline{\boldsymbol{C}}_K$
- Upper $-$ Lower bound $\sim (\frac{\tau}{\theta})^{N-K+1}$ (small enough, can be improved)
- More general distribution for the data should be envisaged
- More precise results in the case of orthogonal bases

Papers available on: `http://www.cmla.ens-cachan.fr/fileadmin/Membres/nikolova/`