Mathematical methods for Image Processing

François Malgouyres

Institut de Mathématiques de Toulouse, France

invitation by
Jidesh P., NITK Surathkal

funding
Global Initiative on Academic Network

Oct. 23–27

Plan

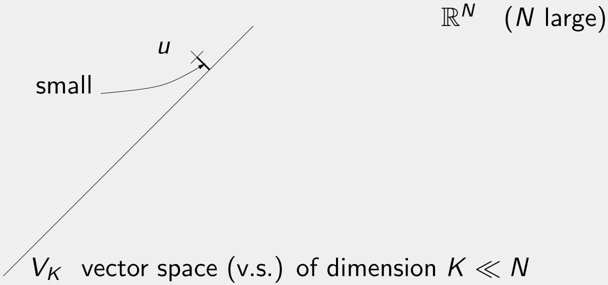
1. Sparsity prior : ℓ^0 case
2. Sparsity prior : ℓ^1 case

Dimensionality reduction

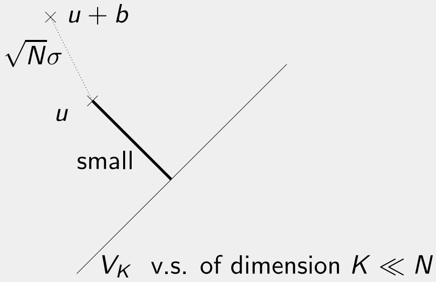
$\mathbb{R}^N \quad (N \text{ large})$

$u \quad \times$

Dimensionality reduction for approximation

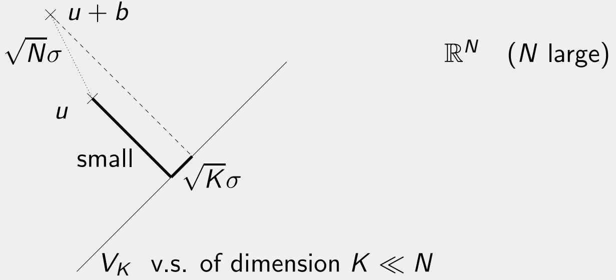


Dimensionality reduction for denoising



$\mathbb{R}^N$ (N large)

Dimensionality reduction for denoising



Dimensionality reduction

Applications:

- Approximation
- Denoising, compressed sensing/inverse problems (later in the talk)
- Feature extraction

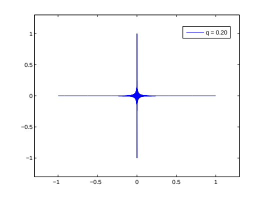
Questions:

- Can we do it for images?
- How to construct V_K in a stable and realistic manner (from a numerical standpoint)?

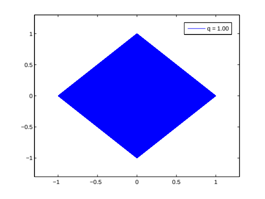
Images are smooth

Let B be an orthonormal basis (e.g. Fourier, ondelettes...)
We assume B is such that for any image u and x such that $u = Bx$.

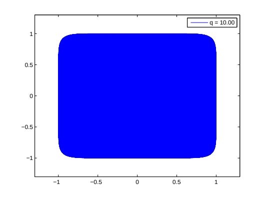
$$\|x\|_p \text{ is small for } p \text{ small.}$$



$p = 0.2$



$p = 1$



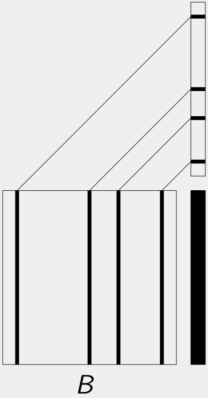
$p = 10$

Assymptotically : Assumption on Sobolev smoothness for Fourier, Besov smoothness for Wavelets.

Images are smooth

We denote S the index of the K largest entries in x and

$$\tilde{x}_k = \begin{cases} x_k & \text{, if } k \in S, \\ 0 & \text{, otherwise.} \end{cases}$$



$$B\tilde{x} \sim u \qquad B\tilde{x} \in V_K = \mathrm{Vect}\{B^k, k \in S\}$$

Non-linear approximation

Theorem

We have for $0 \leq p \leq 2$,

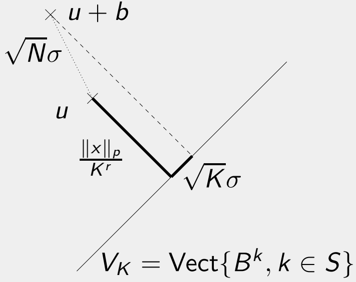
$$\|u - B\tilde{x}\|_2 = \|x - \tilde{x}\|_2 \leq \frac{\|x\|_p}{K^r},$$

for $r = \frac{1}{p} - \frac{1}{2}$.

Examples:

- If $p = 0.01$, $r = 99.5$
- If $p = 0.1$, $r = 9.5$
- If $p = 1$, $r = 0.5$

Dimensionality reduction of smooth images



$\mathbb{R}^N$ (N large)

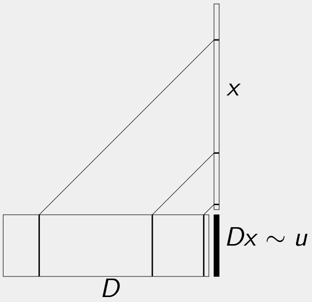
Dimensionality reduction: Manipulating vector spaces

- **Implicit representation** (Often called Analysis, co-sparsity prior): We have an analysis linear operator
$$A : \mathbb{R}^N \longrightarrow \mathbb{R}^P$$
where typically $P \gg N$ and build a set of equation $\mathcal{S} \subset \{1, \ldots, P\}$ such that
$$V_K = \{w \in \mathbb{R}^N | \forall i \in \mathcal{S}, (Aw)_i = 0\}$$
(Ex: A computes finite differences as in Total Variation prior)
- **Explicit representation** (Often called Synthesis, sparsity prior): We have a synthesis linear operator
$$D : \mathbb{R}^P \longrightarrow \mathbb{R}^N$$
where typically $P \gg N$ and build support $\mathcal{S} \subset \{1, \ldots, P\}$ such that
$$V_K = \{Dx | x \in \mathbb{R}^P \text{ and } \mathrm{Supp}\,(x) \subset \mathcal{S}\}$$
(Ex: D is a wavelet transform)

Redundancy in the Synthesis prior

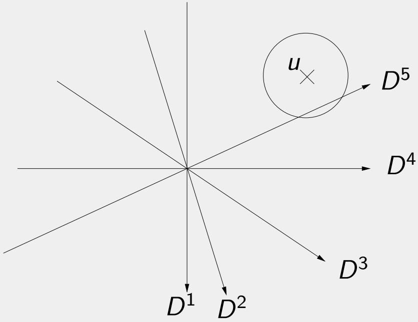
For $P > N$ and $\|x\|_0 = |\{i, x_i \neq 0\}|$

$$(P_0)\left\{\begin{array}{l}\min_{x\in\mathbb{R}^P} \|x\|_0 \\ \text{s.t.} \|Dx-u\|_2 \leq \tau\end{array}\right. \quad \textit{or} \quad (P_0')\left\{\begin{array}{l}\min_{x\in\mathbb{R}^P} \|Dx-u\|_2 \\ \text{s.t. } \|x\|_0 \leq K\end{array}\right.$$



Redundancy: the picture in $\mathbb{R}^N$

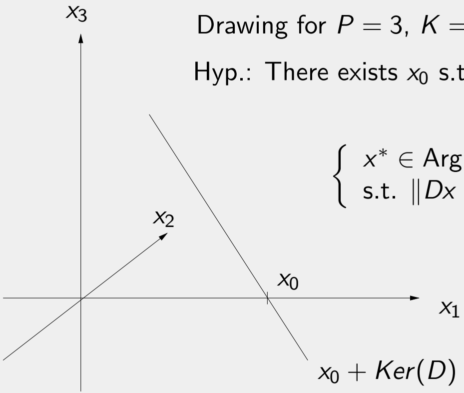
Drawing for $N = 2$, $K = 1$



$$\begin{cases} \min_{x\in\mathbb{R}^P} \|x\|_0 \\ \text{s.t. } \|Dx - u\| \leq \tau \end{cases}$$

Number of v.s. of dimension K:
$C_P^K = \frac{P(P-1)\dots(P-K+1)}{K!}$

Redundancy: the picture in $\mathbb{R}^P$



Drawing for $P = 3$, $K = 1$, $\tau = 0$

Hyp.: There exists x_0 s.t. $\|x_0\|_0 \leq K$ and $u = Dx_0$

$$\begin{cases} x^* \in \text{Argmin}_{x \in \mathbb{R}^P} \|x\|_0 \\ \text{s.t. } \|Dx - u\| = 0 \end{cases}$$

$x^* \in (x_0 + Ker(D)) \cap \left(\cup_{|\mathcal{S}| \leq K} V_{\mathcal{S}}\right)$
But when $\dim(Ker(D)) = P - N$ and $K \ll N$, the intersection of

- $x_0 + Ker(D)$
- and any finite collection of vector space of dimension K

is unlikely in $\mathbb{R}^P$

Compressed sensing, noiseless ℓ^0 case

Theorem

If

- *There exists* x_0 *such that* $\|x_0\|_0 \leq K$ *and* $u = Dx_0$
- D *is such that, for any* $x \neq 0$ *such that* $\|x\|_0 \leq 2K$, $Dx \neq 0$

then x_0 *is the unique solution of* (P_0), *when* $\tau = 0$.

Proof: Otherwise, there exists x with $\|x\|_0 \leq \|x_0\|_0 \leq K$ and such that $Dx = Dx_0$.
Therefore $\|x - x_0\|_0 \leq 2K$ and $D(x - x_0) = 0$. Contradiction. $\square$

Thanks to the sparsity hypothesis, we can invert D !
(But as such it might be unstable and (P_0) is NP-Hard.)

Compressed sensing, noisy ℓ^0 case

We assume the columns D are of norm 1.

Definition (Restricted Isometry Property)

We say that D satisfies the "K-Restricted Isometry Property" (K-RIP)), where $K \in \{1, \dots, N\}$, if there exists $\delta_K \in [0, 1)$ such that

$$(1 - \delta_K)\|x\|^2 \leq \|D_K x\|^2 \leq (1 + \delta_K)\|x\|^2 \qquad , \forall x \in \mathbb{R}^K$$

whatever the matrix D_K extracted from D, of size $N \times K$.

We can equivalently say

$$(1 - \delta_K)\|x\|^2 \leq \|Dx\|^2 \leq (1 + \delta_K)\|x\|^2 \qquad , \forall x \in \mathbb{R}^N \text{ s.t. } \|x\|_0 \leq K$$

(Geometrically, in the parameter space, the (RIP) forces the singular vector of D corresponding to small singular values to be almost orthogonal to all K-sparse vectors.)

Compressed sensing, noisy ℓ^0 case

Remind

$$(P_0) : \begin{cases} x^* \in \text{Argmin} \|x\|_0 \\ x \text{ s. t. } \|Dx - u\| \leq \tau. \end{cases}$$

Theorem

Assume there exists $K \leq N$ and x_0 such that $\|x_0\|_0 \leq K$ and if we consider the data

$$u = Dx_0 + b.$$

If D satisfies the $2K$-RIP with constant δ_{2K}, then

$$\|x_0 - x^*\| \leq \frac{2}{\sqrt{1-\delta_{2K}}}\tau$$

where x^* is any solution of (P_0), with $\|b\| \leq \tau$.

(b contains some approximation error and some noise)

Compressed sensing, noisy ℓ^0 case, Thm proof

Notice first that since $\|b\| \leq \tau$, x_0 satisfies the constraint $\|Dx_0 - u\| \leq \tau$. Therefore

$$\|x^*\|_0 \leq \|x_0\|_0 \leq K$$

and

$$\|x_0 - x^*\|_0 \leq 2K.$$

Therefore

$$\begin{array}{rcl}
(1-\delta_{2K})\|x_0 - x^*\|^2 & \leq & \|D(x_0 - x^*)\|^2 \\
 & = & \|(Dx_0 - u) - (Dx^* - u)\|^2 \\
 & \leq & (\|Dx_0 - u\| + \|Dx^* - u\|)^2 \\
 & \leq & (2\tau)^2
\end{array}$$

$\square$

Rq : The Upper-RIP ($\|Dx\|^2 \leq (1+\delta_K)\|x\|^2$) is not used.

ℓ^0 minimization: A NP-Hard problem

Theorem

All the variants

$$\begin{cases} x^* \in \text{Argmin} \, \|x\|_0 \\ x \; s. \; t. \; \|Dx - u\| \leq \tau. \end{cases}$$

$$\begin{cases} x^* \in \text{Argmin} \, \|Dx - u\|^2 \\ x \; s. \; t. \; \|x\|_0 \leq K. \end{cases}$$

$$x^* \in \text{Argmin} \, \|x\|_0 + \lambda \|Dx - u\|^2$$

are NP-Hard problems.

(See Davies-Mallat-Avellaneda, "Adaptive Greedy Approximation", Constructive approximation, 13:57-98, 1997.)

ℓ^0 minimization: the proximal gradient algorithm

Theorem

For any $x \in \mathbb{R}^P$ and any $t \geq 0$

$$\text{prox}^t_{\|.\|_0}(x)_i = \begin{cases} x_i & \text{, if } |x_i| \geq \frac{\sqrt{2}}{\sqrt{t}} \\ 0 & \text{, otherwise} \end{cases}$$

It is a hard-thresholding.

Proof: By definition

$$\text{prox}^t_{\|.\|_0}(x) = \text{Argmin}_{x' \in \mathbb{R}^P} \frac{t}{2}\|x - x'\|_2^2 + \|x'\|_0$$

This problem is separable and we can prove that, for all $i = 1..P$,

$$\text{prox}^t_{\|.\|_0}(x)_i = \text{Argmin}_{y \in \mathbb{R}} \frac{t}{2}(x_i - y)^2 + \mathbf{1}_{y \neq 0}(y).$$

We have

$$\frac{t}{2}(x_i - y)^2 + \mathbf{1}_{y \neq 0}(y) = \begin{cases} 1 & \text{, if } y = x_i \neq 0 \\ \frac{t}{2}x_i^2 & \text{, if } y = 0 \end{cases}$$

Choosing the smallest objective value leads to the result. □

ℓ^0 minimization: the proximal gradient algorithm

Theorem

For any starting point and any stepsize $0 \leq t < \frac{1}{L}$, where L is the Lipschitz constant of the gradient

$$x \longmapsto 2\lambda D^*(Dx - u)$$

(i.e. $L = 2\lambda\sigma_{max}(D)^2$ where $\sigma_{max}(D)$ is the largest singular value of D), the proximal point algorithm (also called "Iterative Hard thresholding")

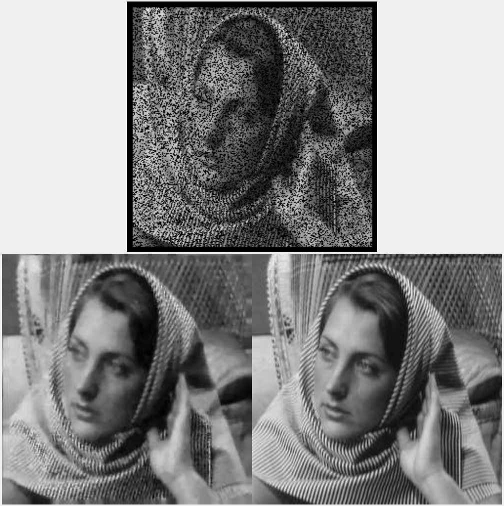
$$x^{k+1} = \text{prox}^t_{\|.\|_0}(x^k \ - \ t\ 2\lambda D^*(Dx^k - u))$$

converges to a **critical point** *of*

$$\|x\|_0 + \lambda\|Dx - u\|^2$$

It is a straightforward consequence of Bolte-Sabach-Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems", Math. Prog. serie A, 2013.

Application of compressed sensing: Inpainting



Top: Image with missing pixels; Bottom: restored and ideal image

Application of compressed sensing: Zoom

Un-zoomed; zoomed (x4) by two methods

Application of compressed sensing: Morphological component analysis

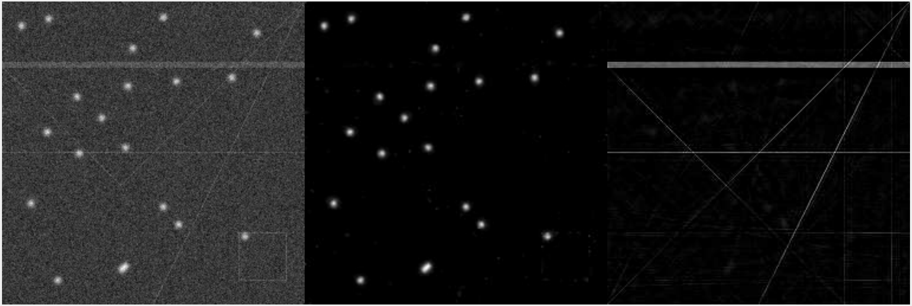


Figure: Initial image, wavelet part, curvelet part (Courtasy of J.L. Starck)

Plan

ℓ^1 minimization

We replace the **nonconvex, not continuous, NP-hard** problem by a **convex, non-differentiable** problem

$$(P_1) : \begin{cases} x^* \in \text{Argmin} \|x\|_1 \\ x \text{ s. t. } \|Dx - u\| \leq \tau. \end{cases}$$

where $\tau \geq 0$ and for the ℓ^1 norm defined by

$$\|x\|_1 = \sum_{i=1}^{P} |x_i|.$$

Or a variant using a Lagrangian

$$x^* \in \text{Argmin} \|x\|_1 + \lambda \|Dx - u\|^2$$

for $\lambda \geq 0$.
(They describe the same solution path, when λ and τ vary.)

ℓ^1 minimization: Proximal gradient algorithm

The objective function

$$\mathcal{E}(x) = \lambda \|Dx - u\|^2 + \|x\|_1$$

has the form $E + R$

- Where $E : W \longrightarrow \mathbb{R}$ is convexe, differentiable, with a Lipschitz gradient[1] of constant $L > 0$
- Where R is lower semi-continuous, proper, convex and coercive.

Beside the fact that E might not be coercive (which turn out not to be an issue), this permit to guarantee that the iterates of the proximal gradient algorithm converge (See lecture on "non-smooth optimization").

[1]

$$\forall w, w' \in W, \qquad \|\nabla E(w') - \nabla E(w)\| \leq L \|w' - w\|$$

ℓ^1 minimization: Proximal gradient algorithm

Theorem (Iterative Soft Thresholding convergence)

The iterates

$$x^{k+1} = \text{prox}^t_{\|.\|_1}(x^k \ - \ t\, 2\lambda D^*(Dx^k - u))$$

converge, for any $t < \frac{1}{L}$ *, where L is the Lipschitz constant of the gradient*

$$x \longmapsto 2\lambda D^*(Dx - u)$$

(i.e. $L = 2\lambda\sigma_{max}(D)^2$ *where* $\sigma_{max}(D)$ *is the largest singular value of D).*
Moreover,$(\mathcal{E}(x^k))_{k\in\mathbb{N}}$ *is non-increasing and, for any minimizer* w^* *of* $\mathcal{E}$*,*

$$\mathcal{E}(x^k) - \mathcal{E}(x^*) \leq \frac{L}{2k}\|x^0 - x^*\|_2.$$

We remind that

$$\text{prox}^L_{\|.\|_1}(x')_i = \begin{cases} x'_i - \frac{1}{L} & \text{, if } x'_i > \frac{1}{L}. \\ 0 & \text{, if } -\frac{1}{L} \leq x'_i \leq \frac{1}{L}, \\ x'_i + \frac{1}{L} & \text{, if } x'_i < -\frac{1}{L}, \end{cases}$$

ℓ^1 minimization

We remind

$$(P_1) : \begin{cases} x^* \in \text{Argmin}\, \|x\|_1 \\ x \text{ s. t. } \|Dx - u\| \leq \tau. \end{cases}$$

Theorem (Compressed sensing: ℓ^1 case)

Let $K \leq N$ and x_0 be such that $\|x_0\|_0 \leq K$. We consider the datum

$$u = Dx_0 + b.$$

If D satisfies the $4K$-RIP and the constants δ_{3K} and δ_{4K} are such that $\delta_{3K} + 3\delta_{4K} < 2$, then

$$\|x_0 - x^*\| \leq C_K \tau$$

where x^* is the solution of (P_1), any τ such that $\|b\| \leq \tau$ and

$$C_K = \frac{4}{\sqrt{3 - 3\delta_{4K}} - \sqrt{1 + \delta_{3K}}}.$$

(b contains some approximation error and some noise)

Compressed sensing ℓ^1: Geometric intuition

Lemma

We have

$$\|D(x^* - x_0)\| \leq 2\tau$$

$$\|x^*\|_1 \leq \|x_0\|_1$$

Proof:

$$\begin{aligned} \|D(x^* - x_0)\| &= \|(Dx^* - u) - (Dx_0 - u)\| \\ &\leq \|Dx^* - u\| + \|Dx_0 - u\| \\ &\leq 2\tau \end{aligned}$$

The second inequality is a straightforward consequence of $\|b\| \leq \tau$. □

The rest of the proof quantifies and argues that the ℓ^1 ball is narrow. (See Candes-Romberg-Tao, "Stable Signal recovery from incomplete and inaccurate measurements", Comm. Pure Appl. Math. 2006.)