Generalized greedy algorithms.

François-Xavier Dupé & **Sandrine Anthoine**

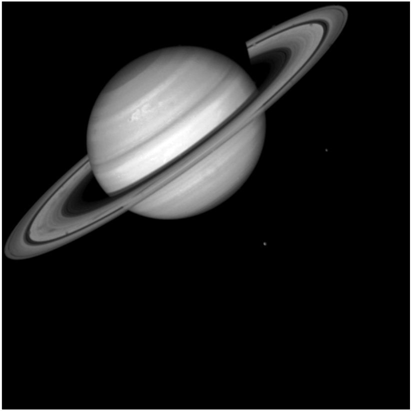
LIF & I2M
Aix-Marseille Université - CNRS - Ecole Centrale Marseille, Marseille
ANR Greta

INSTITUT de MATHÉMATIQUES de MARSEILLE

Séminaire Parisien des Mathématiques Appliquées à l'Imagerie,
05/01/2017

Sparse approximation

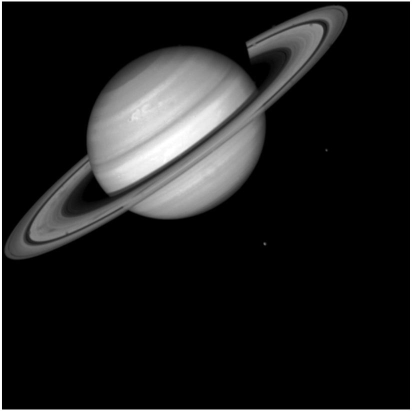
Sparsity is good,



Saturn (original)

Sparse approximation

Sparsity is good,

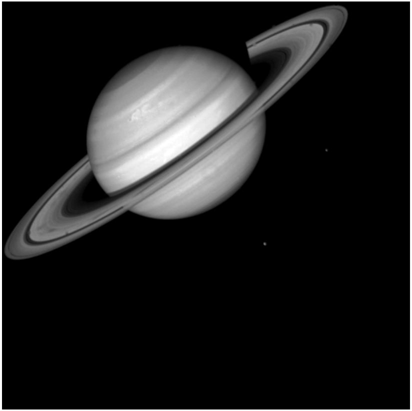


Saturn (original)

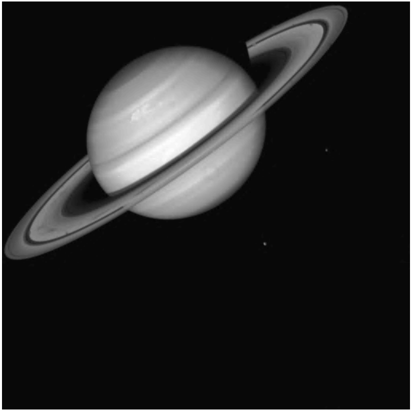
Saturn (coefficients)

Sparse approximation

Sparsity is good,



Saturn (original)



Saturn (2.6% of the coefficients)

Sparse approximation

Classically

Find the *best* k-sparse minimizer of y on the dictionary Φ,

$$\min_{x\in\mathbb{R}^n} \|y - \Phi x\|_2^2 \quad \text{s.t.} \ \|x\|_0 \leqslant k \tag{A}$$

Sparse approximation

Classically

Find the *best* k-sparse minimizer of y on the dictionary Φ,

$$\min_{x \in \mathbb{R}^n} \|y - \Phi x\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leqslant k \tag{A}$$

More generally

Find the *best* k-sparse minimizer,

$$\min_{x \in \mathcal{H}} f(x) \quad \text{s.t.} \quad \|x\|_0 \leqslant k \tag{P}$$

Sparse approximation

Classically

Find the *best* k-sparse minimizer of y on the dictionary Φ,

$$\min_{x \in \mathbb{R}^n} \|y - \Phi x\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leqslant k \tag{A}$$

More generally

Find the *best* k-sparse minimizer,

$$\min_{x \in \mathcal{H}} f(x) \quad \text{s.t.} \quad \|x\|_0 \leqslant k \tag{P}$$

Today

Find a k-sparse zero of an operator $\mathbf{T} : \mathcal{H} \to \mathcal{H}$ (e.g. $\mathbf{T} = \nabla f$),

$$\text{Find } x \in \mathcal{H} \quad \text{s.t} \begin{cases} \|x\|_0 \leqslant k \\ 0 = \mathbf{T}(x) \end{cases} \tag{Q}$$

$\mathcal{H}$: Hilbert space.

Literature

How to solve these problems ?

Literature

How to solve these problems ?

Linear setting (A), at least four families of methods,

- convex relaxation;
- MP, OMP;
- CoSaMP, SP;
- IHT, HTP.

Literature

How to solve these problems ?

Linear setting (A), at least four families of methods,

- convex relaxation;
- MP, OMP;
- CoSaMP, SP;
- IHT, HTP.

How to generalize ? Convergence ?

Literature

Generalization for (P),

- OMP [Zhang 2011],
- GraSP [Bahmani et. al. 2013],
- IHT, HTP [Yuan et. al. 2013];

Literature

Generalization for (P),

- OMP [Zhang 2011],
- GraSP [Bahmani et. al. 2013],
- IHT, HTP [Yuan et. al. 2013];

Today's talk:

Goal solve (P) or/and (Q),

Approach greedy

- theoretically grounded,
- inspired by CoSaMP [Needell, Tropp, 2009] and GraSP [Bahmani et. al., 2013].

Today's topic

1. CoSaMP and its generalizations
2. The Restricted Diagonal Property
3. Three other generalizations
4. Poisson Noise Removal

Today's topic

1. CoSaMP and its generalizations
   - CoSaMP and its guarantees
   - GraSP and its guarantees
   - GCoSaMP and its guarantees
2. The Restricted Diagonal Property
3. Three other generalizations
   - Generalized Subspace Pursuit
   - Generalized Hard Thresholding Pursuit
   - Generalized Iterative Hard Thresholding
4. Poisson Noise Removal
   - Moreau-Yosida regularization
   - Experiments

Optimization problem

Let

- $\mathbf{T} : \mathcal{H} \to \mathcal{H}$ be an operator
- k the expected sparsity.

We wish to solve

$$\text{Find } x \in \mathcal{H} \text{ such that } \mathbf{T}(x) = 0 \quad \text{and} \quad \|x\|_0 \leqslant k \ . \tag{Q}$$

Optimization problem

Let

- $\mathbf{T} : \mathcal{H} \to \mathcal{H}$ be an operator
- k the expected sparsity.

We wish to solve

$$\text{Find } x \in \mathcal{H} \text{ such that } \mathbf{T}(x) = 0 \quad \text{and} \quad \|x\|_0 \leqslant k \ . \tag{Q}$$

Special case: $\mathbf{T} = \boldsymbol{\nabla} f$

- (Q): find a critical point of f that is k-sparse.
- Related problem: find a minimizer of f among the k-sparse vectors.
- Related algorithms:
  - $\mathbf{T}(x) = \boldsymbol{\nabla}_x(||\Phi x - y||_2^2)$: CoSaMP, SP...
  - $\mathbf{T}(x) = \boldsymbol{\nabla} f(x)$, f convex: GraSP, GOMP...

CoSaMP [Needell, Tropp, 2009]

Goal: $\min_x ||\Phi x - y||_2^2$ s. t. $\|x\|_0 \leqslant k$

Algorithm

Require: y, Φ, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N - 1$,

$g = \Phi^*(\Phi x^t - y)$, *(select new directions)*
$\mathcal{G} = \text{supp}(g_{|2k})$,

$\mathcal{S} = \mathcal{G} \cup \text{supp}(x^t)$, *(set extended support)*

$z = \Phi^\dagger_{\mathcal{S}} y$, *(solve on extended support)*

$\mathcal{T} = \text{supp}(z_{|k})$. *(set support)*

$x^{t+1} = z_{|k}$. *(approximately solve on the support)*

Output : x^N.

CoSaMP [Needell, Tropp, 2009]

Goal: $\min_x ||\Phi x - y||_2^2$ s. t. $\|x\|_0 \leqslant k$

Algorithm

Require: y, Φ, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N-1$,

$g = \Phi^*(\Phi x^t - y)$, *(select new directions)*

$\mathcal{G} = \text{supp}(g_{|2k})$,

$\mathcal{S} = \mathcal{G} \cup \text{supp}(x^t)$, *(set extended support)*

$z = \operatorname*{argmin}_{\{x/\,\text{supp}(x) \subseteq \mathcal{S}\}} ||\Phi x - y||_2^2$, *(solve on extended support)*

$\mathcal{T} = \text{supp}(z_{|k})$. *(set support)*

$x^{t+1} = z_{|k}$. *(approximately solve on the support)*

Output : x^N.

CoSaMP [Needell, Tropp, 2009]

Goal: $\min_x ||\Phi x - y||_2^2$ s. t. $\|x\|_0 \leqslant k$

Algorithm

Require: y, Φ, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N - 1$,

$\mathcal{G} = \mathrm{supp}([\Phi^*(\Phi x^t - y)]_{|2k})$, *(select new directions)*

$\mathcal{S} = \mathcal{G} \cup \mathrm{supp}(x^t)$, *(set extended support)*

$z = \underset{\{x/\,\mathrm{supp}(x) \subseteq \mathcal{S}\}}{\mathrm{argmin}} ||\Phi x - y||_2^2$, *(solve on extended support)*

$\mathcal{T} = \mathrm{supp}(z_{|k})$. *(set support)*

$x^{t+1} = z_{|k}$. *(approximately solve on the support)*

Output : x^N.

CoSaMP guarantees

$$\text{Goal: } \min_x ||\Phi x - y||_2^2 \text{ s. t. } \|x\|_0 \leqslant k$$

Restricted Isometry Property

Φ is has the *Restricted Isometry Property* with constant δ_k if and only if

$$\forall x \text{ s.t. } \mathrm{card}(\mathrm{supp}(x)) \leq k, \qquad (1-\delta_k)\|x\| \leq \|\Phi x\| \leq (1+\delta_k)\|x\|$$

CoSaMP guarantees

Goal: $\min_x ||\Phi x - y||_2^2$ s. t. $\|x\|_0 \leqslant k$

Restricted Isometry Property

Φ is has the *Restricted Isometry Property* with constant δ_k if and only if

$$\forall x \text{ s.t. } \operatorname{card}(\operatorname{supp}(x)) \leq k, \qquad (1-\delta_k)\|x\| \leq \|\Phi x\| \leq (1+\delta_k)\|x\|$$

Consider $y = \Phi u + e$

Theorem (CoSaMP error bound)

If Φ is has the RIP with constant $\delta_{4k} \leqslant 0.1$, then at iteration t, x^t verifies

$$\left\|x^t - u\right\| \leqslant \frac{1}{2^t}\|u\| + 20\nu \ ,$$

with ν the incompressible error:

$$\left\|u - u_{|k}\right\|_2 + \frac{1}{\sqrt{k}}\left\|u - u_{|k}\right\|_1 + \|e\|_2 \ .$$

Gradient Support Pursuit [Bahmani et. al., 2013]

Goal: $\min_x f(x)$ s. t. $\|x\|_0 \leqslant k$

CoSaMP

Require: f, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N-1$,

$\mathcal{G} = \mathrm{supp}([\Phi^*(\Phi x^t - y)]_{|2k})$, *(select new directions)*

$\mathcal{S} = \mathcal{G} \cup \mathrm{supp}(x^t)$, *(set extended support)*

$z = \underset{\{x / \mathrm{supp}(x) \subseteq \mathcal{S}\}}{\mathrm{argmin}} ||\Phi x - y||_2^2$, *(solve on extended support)*

$\mathcal{T} = \mathrm{supp}(z_{|k})$. *(set support)*

$x^{t+1} = z_{|k}$. *(approximately solve on the support)*

Output : x^N.

Gradient Support Pursuit [Bahmani et. al., 2013]

Goal: $\min_x f(x)$ s. t. $\|x\|_0 \leqslant k$

Gradient Support Pursuit (GraSP)

Require: f, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N - 1$,

$\mathcal{G} = \text{supp}([\nabla f(x)]_{|2k})$, *(select new directions)*

$\mathcal{S} = \mathcal{G} \cup \text{supp}(x^t)$, *(set extended support)*

$z \in \underset{\{x/\,\text{supp}(x) \subseteq \mathcal{S}\}}{\text{argmin}} f(x)$, *(solve on extended support)*

$\mathcal{T} = \text{supp}(z_{|k})$. *(set support)*

$x^{t+1} = z_{|k}$. *(approximately solve on the support)*

Output : x^N.

GraSP guarantees

$$\text{Goal: } \min_x f(x) \text{ s. t. } \|x\|_0 \leqslant k$$

Stable Restricted Hessian

f has a *Stable Restricted Hessian* with constant μ_k if and only if

$$\frac{A_k(x)}{B_k(x)} \leq \mu_k, \qquad \forall x \text{ s.t. } \operatorname{card}(\operatorname{supp}(x)) \leq k,$$

where

$$A_k(x) = \sup \left\{ \frac{\langle y, H_f(x)y \rangle}{\|y\|_2^2} \mid \quad \operatorname{card}(\operatorname{supp}(x) \cup \operatorname{supp}(y)) \leq k \right\} ,$$

$$B_k(x) = \inf \left\{ \frac{\langle y, H_f(x)y \rangle}{\|y\|_2^2} \mid \quad \operatorname{card}(\operatorname{supp}(x) \cup \operatorname{supp}(y)) \leq k \right\} .$$

GraSP guarantees

Goal: $\min_x f(x)$ s. t. $\|x\|_0 \leqslant k$

Theorem (GraSP error bound)

f has a Stable Restricted Hessian with constant $\mu_{4k} \leqslant \frac{1+\sqrt{3}}{2}$ and there exists $\epsilon > 0$ such that $B_{4k}(u) > \epsilon\ \forall u$, then at iteration t, x^t verifies

$$\left\|x^t - u^\star\right\| \leqslant \frac{1}{2^t}\left\|u^\star\right\| + \frac{C}{\epsilon}\left\|\boldsymbol{\nabla} f(u^\star)_{|3k}\right\| \ .$$

GraSP guarantees

$$\text{Goal: } \min_x f(x) \text{ s. t. } \|x\|_0 \leqslant k$$

Theorem (GraSP error bound)

f has a Stable Restricted Hessian with constant $\mu_{4k} \leqslant \frac{1+\sqrt{3}}{2}$ and there exists $\epsilon > 0$ such that $B_{4k}(u) > \epsilon \ \forall u$, then at iteration t, x^t verifies

$$\left\|x^t - u^\star\right\| \leqslant \frac{1}{2^t}\left\|u^\star\right\| + \frac{C}{\epsilon}\left\|\nabla f(u^\star)_{|3k}\right\| .$$

- Error bound for f only once "differentiable".
- Notion of convexity restricted to k-sparse vectors.

Generalized CoSaMP

Goal: Find $x \in \mathcal{H}$ s. t. $\mathbf{T}(x) = 0$ and $\|x\|_0 \leqslant k$

Gradient Support Pursuit (GraSP)

Require: $\mathbf{T}$, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N - 1$,

$\mathcal{G} = \operatorname{supp}([\nabla f(x)]_{|2k})$, *(select new directions)*

$\mathcal{S} = \mathcal{G} \cup \operatorname{supp}(x^t)$, *(set extended support)*

$z \in \underset{\{x/\operatorname{supp}(x) \subseteq \mathcal{S}\}}{\operatorname{argmin}} f(x)$, *(solve on extended support)*

$\mathcal{T} = \operatorname{supp}(z_{|k})$. *(set support)*

$x^{t+1} = z_{|k}$. *(approximately solve on the support)*

Output : x^N.

Generalized CoSaMP

Goal: Find $x \in \mathcal{H}$ s. t. $\mathbf{T}(x) = 0$ and $\|x\|_0 \leqslant k$

GCoSaMP

Require: $\mathbf{T}$, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N - 1$,

$\mathcal{G} = \operatorname{supp}([\mathbf{T}(x)]_{|2k})$, *(select new directions)*

$\mathcal{S} = \mathcal{G} \cup \operatorname{supp}(x^t)$, *(set extended support)*

$z \quad \text{s. t.} \begin{cases} \operatorname{supp}(z) \subseteq \mathcal{S} \\ \mathbf{T}(z)_{|\mathcal{S}} = 0. \end{cases}$, *(solve on extended support)*

$\mathcal{T} = \operatorname{supp}(z_{|k})$. *(set support)*

$x^{t+1} = z_{|k}$. *(approximately solve on the support)*

Output : x^N.

Uniform Restricted Diagonal Property

$$\mathcal{D}_1 = \left\{ \mathbf{D} : \begin{array}{ccc} \mathcal{H} & \rightarrow & \mathcal{H}, \\ x & \mapsto & \sum_i d_i x_i e_i \end{array} \text{ s.t. } \forall x \, \|\mathbf{D}x\| \geq \|x\| \right\}.$$

Definition (Uniform Restricted Diagonal Property)

$\mathbf{T}$ *is said to have the Uniform Restricted Diagonal Property (URDP) of order* k *if there exists* $\alpha_k > 0$ *and a diagonal operator* $\mathbf{D}_k$ *in* $\mathcal{D}_1$ *such that* $\forall (x, y) \in \mathcal{H}^2$,

$\mathrm{card}(\mathrm{supp}(x) \cup \mathrm{supp}(y)) \leqslant k \Rightarrow$

$$\|\mathbf{T}(x) - \mathbf{T}(y) - \mathbf{D}_k(x - y)\| \leqslant \alpha_k \|x - y\|.$$

Uniform Restricted Diagonal Property

$$\mathcal{D}_1 = \left\{ \mathbf{D} : \begin{array}{ccc} \mathcal{H} & \to & \mathcal{H}, \\ x & \mapsto & \sum_i d_i x_i e_i \end{array} \text{ s.t. } \forall x \, \|\mathbf{D}x\| \geq \|x\| \right\}.$$

Definition (Restricted Diagonal Property)

$\mathbf{T}$ *is said to have the Restricted Diagonal Property (RDP) of order* k *if there exists* $\alpha_k > 0$ *such that for all subsets* $\mathcal{S}$ *of* $\mathbb{N}$ *of cardinal at most* k*, there exists a diagonal operator* $\mathbf{D}_\mathcal{S}$ *in* $\mathcal{D}_1$ *such that* $\forall (x, y) \in \mathcal{H}^2$,

$$\left. \begin{array}{l} \operatorname{supp}(x) \subseteq \mathcal{S} \\ \operatorname{supp}(y) \subseteq \mathcal{S} \end{array} \right\} \Rightarrow$$

$$\|\mathbf{T}(x) - \mathbf{T}(y) - \mathbf{D}_\mathcal{S}(x - y)\| \leqslant \alpha_k \|x - y\|.$$

Generalized CoSaMP

Goal: Find $x \in \mathcal{H}$ s. t. $\mathbf{T}(x) = 0$ and $\|x\|_0 \leqslant k$

Theorem (Generalized CoSaMP error bound)

Denote by $x^\star$ any k-sparse vector and $\alpha^C = \frac{2}{\sqrt{3}} - 1$. If there exists $\rho > 0$ such that $\rho\mathbf{T}$ has the Restricted Diagonal Property of order $4k$ with $\alpha_{4k} \leqslant \alpha^C$ then at iteration t, x^t verifies

$$\left\|x^t - x^\star\right\| \leqslant \frac{1}{2^t}\left\|x^\star\right\| + 12\rho\left\|\mathbf{T}(x^\star)_{|3k}\right\| \ . \quad (1)$$

Today's topic

1. CoSaMP and its generalizations
   - CoSaMP and its guarantees
   - GraSP and its guarantees
   - GCoSaMP and its guarantees
2. The Restricted Diagonal Property
3. Three other generalizations
   - Generalized Subspace Pursuit
   - Generalized Hard Thresholding Pursuit
   - Generalized Iterative Hard Thresholding
4. Poisson Noise Removal
   - Moreau-Yosida regularization
   - Experiments

The Restricted Diagonal Property and RIP

Definition (Restricted Diagonal Property)

$\mathbf{T}$ *has RDP of order* k : $\exists \alpha_k > 0, \forall \mathcal{S} \subseteq \mathbb{N}$, *with* $|\mathcal{S}| \leq k$, $\exists \mathbf{D}_{\mathcal{S}} \in \mathcal{D}_1$ *such that* $\operatorname{supp}(x) \subseteq \mathcal{S}$ & $\operatorname{supp}(y) \subseteq \mathcal{S} \Rightarrow$

$$\|\mathbf{T}(x) - \mathbf{T}(y) - \mathbf{D}_{\mathcal{S}}(x-y)\| \leqslant \alpha_k \|x - y\| .$$

Assume that $\mathbf{T}$ has RDP of order $2k$, then if $\|x\|_0 \leq k$ and $\|y\|_0 \leq k$:

- $\|\mathbf{T}(x) - \mathbf{T}(y)\| \geq (1 - \alpha_{2k}) \|x - y\|$ (injectivity)

The Restricted Diagonal Property and RIP

Definition (Restricted Diagonal Property)

$\mathbf{T}$ *has RDP of order* k : $\exists \alpha_k > 0, \forall \mathcal{S} \subseteq \mathbb{N}$, *with* $|\mathcal{S}| \leq k$, $\exists \mathbf{D}_{\mathcal{S}} \in \mathcal{D}_1$ *such that* $\mathrm{supp}(x) \subseteq \mathcal{S}$ & $\mathrm{supp}(y) \subseteq \mathcal{S} \Rightarrow$

$$\|\mathbf{T}(x) - \mathbf{T}(y) - \mathbf{D}_{\mathcal{S}}(x-y)\| \leqslant \alpha_k \|x-y\| .$$

Assume that $\mathbf{T}$ has RDP of order $2k$, then if $\|x\|_0 \leq k$ and $\|y\|_0 \leq k$:

- $\|\mathbf{T}(x) - \mathbf{T}(y)\| \geq (1 - \alpha_{2k}) \|x - y\|$ (injectivity)
- $(1 - \alpha_{2k}) \|x - y\| \leqslant \|\mathbf{T}(x) - \mathbf{T}(y)\| \leqslant (D + \alpha_{2k}) \|x - y\|$
  $(D = |||\mathbf{D}_{2k}|||$ or $\sup |||\mathbf{D}_{\mathcal{S}}|||$ if exists).

The Restricted Diagonal Property and RIP

Definition (Restricted Diagonal Property)

$\mathbf{T}$ *has RDP of order* k : $\exists \alpha_k > 0, \forall \mathcal{S} \subseteq \mathbb{N}$, *with* $|\mathcal{S}| \leq k$, $\exists \mathbf{D}_\mathcal{S} \in \mathcal{D}_1$ *such that* $\operatorname{supp}(x) \subseteq \mathcal{S}$ & $\operatorname{supp}(y) \subseteq \mathcal{S} \Rightarrow$

$$\|\mathbf{T}(x) - \mathbf{T}(y) - \mathbf{D}_\mathcal{S}(x-y)\| \leqslant \alpha_k \|x-y\| .$$

Assume that $\mathbf{T}$ has RDP of order $2k$, then if $\|x\|_0 \leq k$ and $\|y\|_0 \leq k$:

- $\|\mathbf{T}(x) - \mathbf{T}(y)\| \geq (1-\alpha_{2k}) \|x-y\|$ (injectivity)
- $(1-\alpha_{2k}) \|x-y\| \leqslant \|\mathbf{T}(x) - \mathbf{T}(y)\| \leqslant (D+\alpha_{2k}) \|x-y\|$
  ($D = |||\mathbf{D}_{2k}|||$ or $\sup |||\mathbf{D}_\mathcal{S}|||$ if exists).
- $\mathbf{T}(x) = \Phi^*(\Phi x - z) \quad \Leftrightarrow \quad \Phi$ is RIP.

Uniform Restricted Diagonal Property: characterization

Theorem

$\beta\mathbf{T}$ *is URDP of order k for* $\mathbf{D}$, *with* $\alpha_k < 1$ *and* $\beta > 0$ $\Leftrightarrow$
$\exists(m, L)$ *such that* $0 < m$ *and* $0 \leq |||D|||^2 - \frac{m^2}{L^2} < 1$ *and*

$$|\operatorname{supp}(x) \cup \operatorname{supp}(y)| \leq k \Rightarrow \begin{cases} \|\mathbf{T}(x) - \mathbf{T}(y)\| \leq L\,\|x - y\| \\ \langle \mathbf{T}(x) - \mathbf{T}(y), D(x - y)\rangle \geq m\,\|x - y\|^2 . \end{cases}$$

Uniform Restricted Diagonal Property: characterization

Theorem

$\beta\mathbf{T}$ *is URDP of order k for* $\mathbf{D}$, *with* $\alpha_k < 1$ *and* $\beta > 0$ $\Leftrightarrow$
$\exists(m, L)$ *such that* $0 < m$ *and* $0 \leq |||D|||^2 - \dfrac{m^2}{L^2} < 1$ *and*

$$|\operatorname{supp}(x) \cup \operatorname{supp}(y)| \leq k \Rightarrow \begin{cases} \|\mathbf{T}(x) - \mathbf{T}(y)\| \leq L\,\|x - y\| \\ \langle \mathbf{T}(x) - \mathbf{T}(y), D(x - y)\rangle \geq m\,\|x - y\|^2 \,. \end{cases}$$

- L-Lipschitz property on sparse elements.
- $\mathbf{D} = \mathbf{I}$: “monotone operator”.

Uniform Restricted Diagonal Property: characterization

Theorem

$\beta\nabla f$ is URDP of order k for $\mathbf{D}$, with $\alpha_k < 1$ and $\beta > 0$ $\Leftrightarrow$
$\exists(m, L)$ such that $0 < m$ and $0 \leq |||D|||^2 - \dfrac{m^2}{L^2} < 1$ and

$$|\operatorname{supp}(x) \cup \operatorname{supp}(y)| \leq k \Rightarrow \begin{cases} \|\nabla f(x) - \nabla f(y)\| \leq L\,\|x - y\| \\ \langle \nabla f(x) - \nabla f(y), D(x - y) \rangle \geq m\,\|x - y\|^2 . \end{cases}$$

- L-Lipschitz property on sparse elements.
- $\mathbf{D} = \mathbf{I}$: “monotone operator”.

If $\mathbf{T} = \nabla f$

- L-Lipschitz property: *Restricted Strong Smoothness.*
- if $\mathbf{D} = \mathbf{I}$: *Restricted Strong Convexity*
  $\hookrightarrow$ recovers the conditions of [Bahmani et. al., 2013].

Today's topic

1. CoSaMP and its generalizations
   - CoSaMP and its guarantees
   - GraSP and its guarantees
   - GCoSaMP and its guarantees
2. The Restricted Diagonal Property
3. Three other generalizations
   - Generalized Subspace Pursuit
   - Generalized Hard Thresholding Pursuit
   - Generalized Iterative Hard Thresholding
4. Poisson Noise Removal
   - Moreau-Yosida regularization
   - Experiments

Generalized Subspace Pursuit

GCoSaMP

Require: $\mathbf{T}$, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N-1$,

$\mathcal{G} = \text{supp}([\mathbf{T}(x)]_{|2k})$,

$\mathcal{S} = \mathcal{G} \cup \text{supp}(x^t)$,

$$z \text{ s. t. } \begin{cases} \text{supp}(z) \subseteq \mathcal{S} \\ \mathbf{T}(z)_{|\mathcal{S}} = 0. \end{cases} ,$$

$\mathcal{T} = \text{supp}(z_{|k})$.

$x^{t+1} = z_{|k}$.

Output : x^N.

GSP

Require: $\mathbf{T}$, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N-1$,

$\mathcal{G} = \text{supp}([\mathbf{T}(x)]_{|k})$,

$\mathcal{S} = \mathcal{G} \cup \text{supp}(x^t)$,

$$z \text{ s. t. } \begin{cases} \text{supp}(z) \subseteq \mathcal{S} \\ \mathbf{T}(z)_{|\mathcal{S}} = 0. \end{cases} ,$$

$\mathcal{T} = \text{supp}(z_{|k})$.

$$x^{t+1} \text{ s. t. } \begin{cases} \text{supp}(x^{t+1}) \subseteq \mathcal{T} \\ \mathbf{T}(x^{t+1})_{|\mathcal{S}} = 0. \end{cases} .$$

Output : x^N.

Generalized Subspace Pursuit

Theorem (Generalized CoSaMP error bound)

If $\rho\mathbf{T}$ has the Restricted Diagonal Property of order $4k$ with $\alpha_{4k} \leqslant \alpha^C = \frac{2}{\sqrt{3}} - 1$ then at iteration t of GCoSaMP, x^t verifies

$$\left\|x^t - x^\star\right\| \leqslant \frac{1}{2^t}\left\|x^\star\right\| + 12\rho\left\|\mathbf{T}(x^\star)_{|3k}\right\| \ .$$

Theorem (Generalized Subspace Pursuit error bound)

If $\rho\mathbf{T}$ has the Restricted Diagonal Property of order $3k$ with $\alpha_{3k} \leqslant \alpha^S$ then at iteration t of GSP, x^t verifies

$$\left\|x^t - x^\star\right\| \leqslant \frac{1}{2^t}\left\|x^\star\right\| + 12\rho\left\|\mathbf{T}(x^\star)_{|2k}\right\| \ .$$

$x^\star$ is k-sparse. α^S is the real root of $x^3 + x^2 + 7x - 1$.

Generalized Hard Thresholding Pursuit

GCoSaMP

Require: $\mathbf{T}$, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N-1$,

$\mathcal{G} = \operatorname{supp}([\mathbf{T}(x)]_{|2k})$,

$\mathcal{S} = \mathcal{G} \cup \operatorname{supp}(x^t)$,

$$z \text{ s. t. } \begin{cases} \operatorname{supp}(z) \subseteq \mathcal{S} \\ \mathbf{T}(z)_{|\mathcal{S}} = 0. \end{cases} ,$$

$\mathcal{T} = \operatorname{supp}(z_{|k})$.

$x^{t+1} = z_{|k}$.

Output : x^N.

GHTP

Require: $\mathbf{T}$, k, η .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N-1$,

$\mathcal{G} = \operatorname{supp}([\mathbf{T}(x)]_{|k})$,

$\mathcal{S} = \mathcal{G} \cup \operatorname{supp}(x^t)$,

$$z = \left[(\mathbf{I} - \eta\mathbf{T})(x^t)\right]_{|\mathcal{S}} ,$$

$\mathcal{T} = \operatorname{supp}(z_{|k})$.

$$x^{t+1} \text{ s. t. } \begin{cases} \operatorname{supp}(x^{t+1}) \subseteq \mathcal{T} \\ \mathbf{T}(x^{t+1})_{|\mathcal{S}} = 0. \end{cases} .$$

Output : x^N.

Generalized Hard Thresholding Pursuit

Theorem (Generalized CoSaMP error bound)

If $\rho\mathbf{T}$ *has the Restricted Diagonal Property of order* $4k$ *with* $\alpha_{4k} \leqslant \alpha^C = \frac{2}{\sqrt{3}} - 1$ *then at iteration* t *of GCoSaMP,* x^t *verifies*

$$\left\|x^t - x^\star\right\| \leqslant \frac{1}{2^t}\left\|x^\star\right\| + 12\rho\left\|\mathbf{T}(x^\star)_{|3k}\right\| \ .$$

Theorem (Generalized HTP error bound)

If $\mathbf{T}$ *has the Uniform Restricted Diagonal Property of order* $2k$ *with* $\mathbf{D}_{2k} = \mathbf{I}$, $\alpha_{2k} \leqslant \alpha^H$ *and* $\frac{3}{4} < \eta < \frac{5}{4}$, *then at iteration* t *of GHTP,* x^t *verifies*

$$\left\|x^t - x^\star\right\| \leqslant \frac{1}{2^t}\left\|x^\star\right\| + 2\frac{(1+2\eta)(1-\alpha_{2k})+4}{(1-\alpha_{2k})^2}\left\|\mathbf{T}(x^\star)_{|2k}\right\| \ . \qquad (2)$$

$x^\star$ is k-sparse. $\alpha^H = 7 - 2\sqrt{11}$.

Generalized Iterative Hard Thresholding

GCoSaMP

Require: $\mathbf{T}$, k .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N-1$,

$\mathcal{G} = \mathrm{supp}([\mathbf{T}(x)]_{|2k})$,

$\mathcal{S} = \mathcal{G} \cup \mathrm{supp}(x^t)$,

z s. t. $\begin{cases} \mathrm{supp}(z) \subseteq \mathcal{S} \\ \mathbf{T}(z)_{|\mathcal{S}} = 0. \end{cases}$,

$\mathcal{T} = \mathrm{supp}(z_{|k})$.

$x^{t+1} = z_{|k}$.

Output : x^N.

GIHT

Require: $\mathbf{T}$, k, η .

Initialization: $x^0 = 0$.

For $t = 0$ **to** $N-1$,

$z = (\mathbf{I} - \eta\mathbf{T})(x^t)$,

$\mathcal{T} = \mathrm{supp}(z_{|k})$.

$x^{t+1} = z_{|k}$.

Output : x^N.

Generalized Iterative Hard Thresholding

Theorem (Generalized CoSaMP error bound)

If $\rho\mathbf{T}$ *has the Restricted Diagonal Property of order* $4k$ *with* $\alpha_{4k} \leqslant \alpha^C = \frac{2}{\sqrt{3}} - 1$ *then at iteration* t *of GCoSaMP,* x^t *verifies*

$$\left\|x^t - x^\star\right\| \leqslant \frac{1}{2^t}\left\|x^\star\right\| + 12\rho\left\|\mathbf{T}(x^\star)_{|3k}\right\| .$$

Theorem (Generalized IHT error bound)

If $\mathbf{T}$ *has the Uniform Restricted Diagonal Property of order* $2k$ *with* $\mathbf{D}_{2k} = \mathbf{I}$, $\frac{3}{4} < \eta < \frac{5}{4}$ *and* $\alpha_{2k} \leqslant \alpha^\eta$ *then at iteration* t *of GIHT,* x^t *verifies*

$$\left\|x^t - x^\star\right\| \leqslant \frac{1}{2^t}\left\|x^\star\right\| + 4\eta\left\|\mathbf{T}(x^\star)_{|3k}\right\| .$$

$x^\star$ is k-sparse. $\alpha^\eta = \frac{1-4|\eta-1|}{4(1+|\eta-1|)}$.

About these error bounds

For all four algorithms

- Guaranteed convergence to unique solution if it exists, at an exponential rate.
- Incompressible error of the form $\left\|T(x)_{|\cdot k}\right\|$.

About these error bounds

For all four algorithms

- Guaranteed convergence to unique solution if it exists, at an exponential rate.
- Incompressible error of the form $\left\|T(x)_{|\cdot k}\right\|$.

For GCoSaMP and GSP

- Invariance to scaling of the algorithm, no parameters to set.
- Guarantee in the RDP case (no “monotonicity” of $\mathbf{T}$ or “convexity” of f required).

About these error bounds

For all four algorithms

- Guaranteed convergence to unique solution if it exists, at an exponential rate.
- Incompressible error of the form $\left\|T(x)_{|\cdot k}\right\|$.

For GCoSaMP and GSP

- Invariance to scaling of the algorithm, no parameters to set.
- Guarantee in the RDP case (no "monotonicity" of $\mathbf{T}$ or "convexity" of f required).

For GHTP and GIHT

- Requires careful setting of the step η w.r.t to $\mathbf{T}$.
- Guarantee only in the URDP case with $\mathbf{D} = \mathbf{I}$ ("monotonicity" of $\mathbf{T}$ or "convexity" of f required).

Today's topic

1. CoSaMP and its generalizations
   - CoSaMP and its guarantees
   - GraSP and its guarantees
   - GCoSaMP and its guarantees
2. The Restricted Diagonal Property
3. Three other generalizations
   - Generalized Subspace Pursuit
   - Generalized Hard Thresholding Pursuit
   - Generalized Iterative Hard Thresholding
4. Poisson Noise Removal
   - Moreau-Yosida regularization
   - Experiments

Forward model

The observed data y is a Poisson noise corrupted version of x,

$$y \sim \mathcal{P}(x) \ .$$

Sparsity assumption

$$x = \mathbf{\Phi}\alpha \quad \text{with} \quad \|\alpha\|_0 \leqslant k \quad \text{and} \quad k << n.$$

with

- y the observation (e.g. a $\sqrt{n} \times \sqrt{n}$ image),
- x the true image,
- $\mathbf{\Phi} \in \mathbb{R}^{n \times d}$ a dictionary (redundant if $d > n$),
- α coefficients to be found ($x = \mathbf{\Phi}\alpha$).

Regularized Sparse Poisson denoising

Writing $F_y(x) = -\log P(y|x)$, one naturally seeks to

Minimize the neg-log-likelihood under a sparsity constraint

$$\min_{\alpha \in \mathbb{R}^d} F_y(\boldsymbol{\Phi}(\alpha)) \quad \text{s. t. } \|\alpha\|_0 \leqslant k \ .$$

$F_y : x \in \mathbb{R}^n \mapsto \sum_{i=1}^{n} f_y^i(x[i])$, with

$$f_y^i(\xi) = \begin{cases} -y[i]\log(\xi) + \xi & \text{if } y[i] > 0 \text{ and } \xi > 0, \\ \xi & \text{if } y[i] = 0 \text{ and } \xi \geqslant 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Regularized Sparse Poisson denoising

Writing $F_y(x) = -\log P(y|x)$, one naturally seeks to

Minimize the neg-log-likelihood under a sparsity constraint

$$\min_{\alpha \in \mathbb{R}^d} F_y(\mathbf{\Phi}(\alpha)) \quad \text{s. t. } \|\alpha\|_0 \leqslant k \ .$$

$F_y : x \in \mathbb{R}^n \mapsto \sum_{i=1}^{n} f_y^i(x[i])$, with

$$f_y^i(\xi) = \begin{cases} -y[i]\log(\xi) + \xi & \text{if } y[i] > 0 \text{ and } \xi > 0, \\ \xi & \text{if } y[i] = 0 \text{ and } \xi \geqslant 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Issues :

- no full domain, no tolerance if a pixel is removed (set to 0),
- non-Lipschitz gradient.

Regularized Sparse Poisson denoising

Instead, we propose

Min. a regularized neg-log-likelihood under a sparsity constraint

$$\min_{\alpha \in \mathbb{R}^d} \mathcal{M}_{\lambda, F_y}(\mathbf{\Phi}(\alpha)) \quad \text{s. t.} \ \ \|\alpha\|_0 \leqslant k \ .$$

which is equivalent to

Using GCoSaMP/GSP with

$$\mathbf{T} = \frac{1}{\nu\lambda} \mathbf{\Phi}^* \circ (\mathbf{I} - \mathrm{prox}_{\nu\lambda F_y}) \circ \mathbf{\Phi}$$

- $\mathcal{M}_{\lambda,f}$ is the Moreau-Yosida regularization of f with parameter λ.
- prox_f is the proximal operator.
- ν is the frame bound for $\mathbf{\Phi}$.

Experiments

We compare using GSP, GCoSaMP, GHTP or GIHT to solve

$$\min_{\alpha \in \mathbb{R}^d} \mathcal{M}_{\lambda, F_y \circ \mathbf{\Phi}}(\alpha) \quad \text{s. t. } \|\alpha\|_0 \leqslant k \ ,$$

Experiments

We compare using GSP, GCoSaMP, GHTP or GIHT to solve

$$\min_{\alpha \in \mathbb{R}^d} \mathcal{M}_{\lambda, F_y \circ \mathbf{\Phi}}(\alpha) \quad \text{s. t. } \|\alpha\|_0 \leqslant k \ ,$$

to

- Subspace Pursuit (SP):[Dai and Milenkovic, 2009]

$$\min_{\alpha \in \mathbb{R}^d} \|\mathbf{\Phi}(\alpha) - y\|_2^2 \quad \text{s. t. } \|\alpha\|_0 \leqslant k \ ,$$

- l_1-relaxation: using Forward-Backward-Forward primal-dual algorithm [Combettes et. al. 2012] to solve

$$\min_{\alpha \in \mathbb{R}^d} F_y \circ \mathbf{\Phi}(\alpha) + \gamma \|\alpha\|_1 \ ,$$

- SAFIR: adaptation of BM3D [Boulanger et al., 2010].
- MSVST: variance-stabilizing method [Zhang et al., 2008].

Comparing greedy l_0 and l_1 results

(a) Original

(b) Noisy

(c) GSP

(d) GHTP

(e) GCoSaMP

Figure : *Cameraman*, maximal intensity 30, undecimated wavelet transform.

Comparing greedy l_0 and l_1 results

(c) GSP

(d) GHTP

(e) GCoSaMP

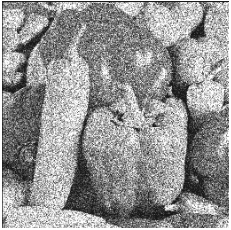
(f) GIHT

(g) ℓ_1 method

Figure : *Cameraman*, maximal intensity 30, undecimated wavelet transform.

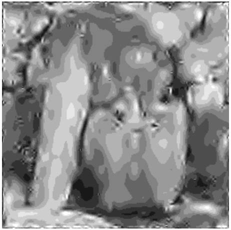
Influence of λ in $\mathcal{M}_{\lambda,\cdot}$.

Original



Noisy

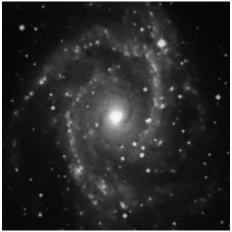
$\lambda = 10$, MAE=0.81



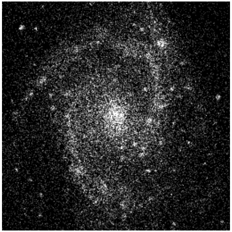
$\lambda = 0.1$, MAE=0.74

Maximal Intensity: 10, k: 1500, $\boldsymbol{\Phi}$: cycle-spinning wavelet transform.

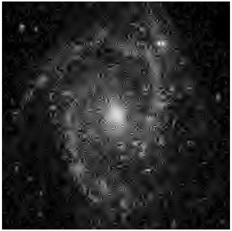
Results



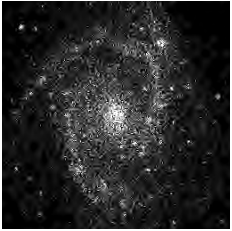
(a) Original



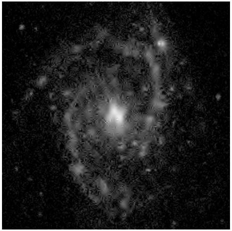
(b) Noisy



(c) GSP



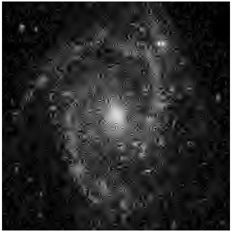
(d) SP



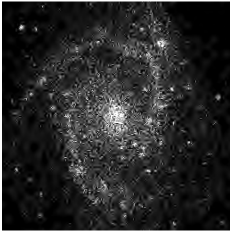
(e) ℓ_1-relaxation

Figure : NGC 2997 Galaxy image.

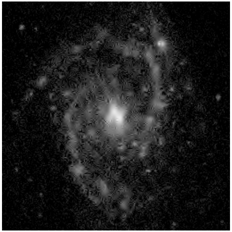
Results



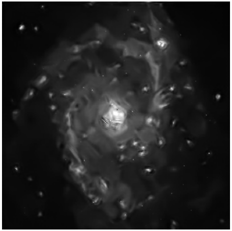
(c) GSP



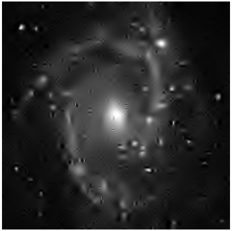
(d) SP



(e) ℓ_1-relaxation



(f) SAFIR



(g) MSVST

Figure : NGC 2997 Galaxy image.

Numerical results

	Sparse Cameraman		Galaxy	
	MAE	SSIM	MAE	SSIM
Noisy	1.57	0.252	0.63	0.19
GSP	**0.252**	**0.87**	0.17	0.71
SP	0.55	0.63	0.28	0.55
SAFIR	0.256	**0.86**	0.15	**0.84**
MSVST	**0.251**	0.84	**0.12**	**0.83**
ℓ_1-relaxation	0.64	0.73	0.252	0.50

Table : Comparison of denoising methods on a sparse version of Cameraman ($k/n = 0.15$) and the NGC 2997 Galaxy.

Take away messages

We have

- presented a general greedy optimization algorithms, with theoretical guarantees;
- applied it to a Moreau-Yosida regularization of Poisson likelihood;
- shown encouraging results for Poisson denoising.

Perspectives includes

- application in other known settings such as dictionary learning, additional regularization…
- gain understanding on the behavior of GCoSaMP, GSP and co.

Thanks for your attention.

Any questions ?

Moreau-Yosida regularization

Definition (Lemaréchal et al, 1997)

Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *be a proper, convex and lower semi-continuous function. The Moreau-Yosida regularization of* f *with parameter* λ *is:*

$$\mathcal{M}_{\lambda,f} : \mathbb{R}^d \to \mathbb{R},$$
$$s \mapsto \inf_{x \in \mathbb{R}^d} \left[\tfrac{1}{2\lambda}||s - x||^2 + f(x)\right]$$

Moreau-Yosida regularization

Definition (Lemaréchal et al, 1997)

Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semi-continuous function. The Moreau-Yosida regularization of f with parameter λ is:

$$\mathcal{M}_{\lambda,f} : \mathbb{R}^d \to \mathbb{R}, \quad s \mapsto \inf_{x \in \mathbb{R}^d} \left[\tfrac{1}{2\lambda} ||s - x||^2 + f(x) \right]$$

Remarks:

- The gradient of $\mathcal{M}_{\lambda,f}$ is linked with proximal operator.
- λ regulates the similarity between f and $\mathcal{M}_{\lambda,f}$.

Gradient for the Moreau-Yosida regularization of Poisson likelihood

Proposition (Combettes and Pesquet, 2007)

Let $\mathbf{\Phi}$ *is a tight frame (i.e.* $\exists\ \nu > 0$*, such that* $\mathbf{\Phi} \circ \mathbf{\Phi}^* = \nu \mathbf{I}$*), then the gradient of the Moreau-Yosida regularization of* $F_y \circ \mathbf{\Phi}$ *is:*

$$\nabla \mathcal{M}_{\lambda, F_y \circ \mathbf{\Phi}}(x) = \frac{1}{\nu\lambda} \mathbf{\Phi}^* \circ (\mathbf{I} - \mathrm{prox}_{\nu\lambda F_y}) \circ \mathbf{\Phi}$$

with

$$\mathrm{prox}_{\nu\lambda F_y}(x)[i] = \frac{x[i] - \nu\lambda + \sqrt{|x[i] - \nu\lambda|^2 + 4\nu\lambda y[i]}}{2}$$

Optimization problem

$$\min_{\alpha \in \mathbb{R}^d} \mathcal{M}_{\lambda, F_y \circ \mathbf{\Phi}}(\alpha) \quad \text{s. t. } \|\alpha\|_0 \leqslant k \quad (\text{P}),$$

with,

- $\mathbf{\Phi}$ the dictionary,
- λ the regularization parameter,
- k sought sparsity.

Optimization problem

$$\min_{\alpha \in \mathbb{R}^d} \mathcal{M}_{\lambda, F_y \circ \mathbf{\Phi}}(\alpha) \quad \text{s. t. } \|\alpha\|_0 \leqslant k \quad \text{(P)},$$

with,

- $\mathbf{\Phi}$ the dictionary,
- λ the regularization parameter,
- k sought sparsity.

The bias between the solution of (P) and the non-regularized problem, under mild condition, is $\mathcal{O}(\sqrt{\lambda})$. [Mahey et Tao, 1993]