Sparsity Recovery by Iterative Orthogonal Projections of Nonlinear Mappings

Alessandro Adamo and Giuliano Grossi

Dipartimento di Scienze dell'Informazione
Università degli Studi di Milano
Via Comelico 39, I-20135 Milan, Italy
Email: adamo@dsi.unimi.it, grossi@dsi.unimi.it

Outline

- Sparse Solutions to Underdetermined Systems
- Sparsity Promoting Mappings
- The Algorithm LiMapS
- Sparsity Minimization
- Numerical Results
- Conclusions and Future Works

Sparse Solutions to Underdetermined Systems

- Let $s = (s_1, \ldots, s_n)$ be a discrete-time signal of length n
- $\Phi = [\phi_1, \ldots, \phi_m]$ be a collection of m basic waveforms or vectors in $\mathbb{R}^n$
- Assuming $m > n$, the dictionary will result in an overcomplete frame leading to infinite solutions of the underdetermined linear system

$$\Phi\alpha = s,$$

- NP-hard combinatorial optimization problem:

$$\min_{\alpha} \| \alpha \|_0 \quad \text{subject to} \quad \Phi\alpha = s, \qquad \text{(P0)}$$

where $\| \alpha \|_0 = |\{k : \alpha_k \neq 0\}|$ denotes the ℓ^0-norm

Sparsity Promoting Mappings

- Let $\mathcal{F} = \{f_\lambda \mid \lambda \in \mathbb{R}^+\}$ family of nonlinear maps
- Let $f_\lambda : \mathbb{R} \to \mathbb{R}$ be a function depending on a real parameter $\lambda > 0$, defined as

$$f_\lambda(x) = x(1 - e^{-\lambda|x|}). \tag{1}$$

- the mapping f_λ is uniformly Lipschitzian with respect to λ with Lipschitz constant $1 + e^{-2}$
- given that $|f_\lambda'(x)| < 1$ on the interval $(-1/\lambda, 1/\lambda)$, the mapping is contractive within that interval with fixed-point at the origin

Sparsity Promoting Mappings: The function f_λ

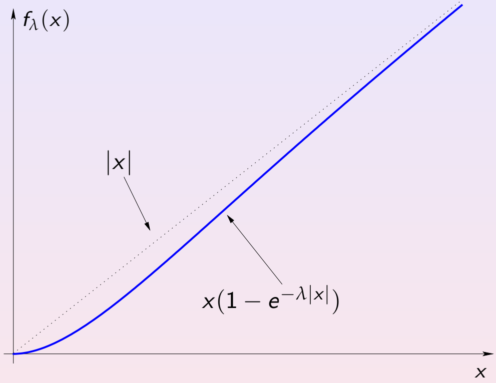

Figure: The graph of function f_λ

Sparsity Promoting Mappings: Derivative of f_λ

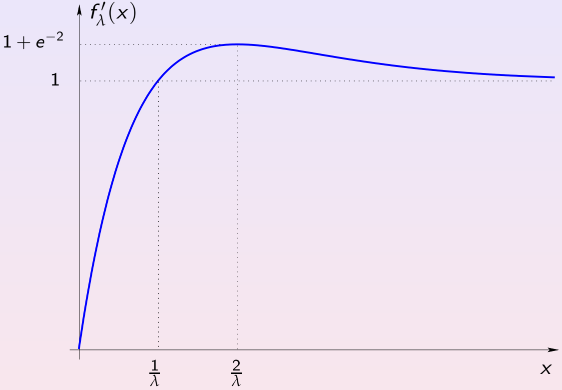

Figure: The graph of the first derivative of f_λ.

Sparsity Promoting Mappings: Extendind the function $f\lambda$

- To deal with high dimensional data, we extend mapping (1) to many dimensions via the elementwise Hadamard product of vectors, obtaining the one-parameter family of nonlinear functions $\mathcal{F} = \{f_\lambda : \mathbb{R}^m \to \mathbb{R}^m \mid \lambda \in \mathbb{R}^+\}$ where each component of $\mathcal{F}$ extends (1) as follows

$$f_\lambda(x) = x \odot (1 - e^{-\lambda|x|}), \tag{2}$$

where $\odot$ denotes the Hadamard product.

Sparsity Promoting Mappings: Orthogonal Projection

- Let an orthogonal projector aimed to map every point falling in the range of (2) into the nearest point in the affine space $\mathcal{A}_{\Phi,s} = \{\alpha \in \mathbb{R}^m : \Phi\alpha = s\}$ (supposed not empty) given by equation $\Phi\alpha = s$.
- denote by $\Phi^\dagger = (\Phi^T\Phi)^{-1}\Phi^T$ the Moore-Penrose pseudo-inverse of a full-rank matrix Φ and by $\nu = \Phi^\dagger s$ the closed-form least-squares solution
- in order to minimize the Euclidean norm of the residual vector $s - \Phi\alpha$, we use orthogonal projections as given by the following mapping:

$$\alpha \mapsto \alpha + \Phi^\dagger(s - \Phi\alpha) = P\alpha + \nu, \tag{3}$$

where $P = I - \Phi^\dagger\Phi$ is the orthogonal projector onto the kernel of Φ.

Sparsity Promoting Mappings: The Iterative System

- A map $T_\lambda : \mathbb{R}^m \to \mathcal{A}_{\Phi,s}$, obtained by combining nonlinear mappings falling in the family $\mathcal{F}$ and the orthogonal projector (3), is given by

$$T_\lambda(\alpha) = Pf_\lambda(\alpha) + \nu. \tag{4}$$

- $$\| T_\lambda(\alpha) - T_\lambda(\beta) \| \leq \sqrt{m}\left(1 + e^{-2}\right) \| \alpha - \beta \| .$$

- fixed a positive sequence $\{\lambda_n\}_{n\geq 0}$, the iterates are inductively defined as

$$\begin{cases} \alpha_0 = \alpha \in \mathbb{R}^m \\ \alpha_{n+1} = T_{\lambda_n}(\alpha_n) \end{cases} . \tag{5}$$

Sparsity Promoting Mappings: The Iterative System

- Let $\{\alpha_n\}$ be the sequence in $\mathcal{A}_{\Phi,s}$ generated by the operator $T_\lambda(\cdot)$ in (5). Then it can be verify that

$$\alpha_n = P\alpha - P\left[\sum_{k=0}^{n-1} \alpha_k \odot e^{-\lambda_k|\alpha_k|}\right] + \nu \ .$$

- the sequence converges when

$$P\alpha + \nu = (I - \Phi^\dagger\Phi)\alpha + \Phi^\dagger s = \alpha,$$

or equivalently, when

$$P\left[\sum_{k=0}^{\infty} \alpha_k \odot e^{-\lambda_k|\alpha_k|}\right] = 0$$

Algorithm 1 LiMapS

Require:
- a dictionary $\Phi \in \mathbb{R}^{n\times m}$
- its pseudo-inverse $\Phi^\dagger$
- a signal $s \in \mathbb{R}^n$
- a sequence $\{\lambda_t\}_{t\geq 0}$

```
t ← 0
α ← ν
while [cond] do
    λ ← λ_t                      <sparsity ratio update>
    β ← f_λ(α)                   <increase sparsity>
    α ← β − Φ†(Φβ − s)           <orthogonal projection>
    t ← t + 1                    <step update>
end while
```

Ensure: a fixed-point $\alpha = P\alpha + \nu$

The Algorithm LiMapS: Stopping Criteria and The Choice of the Parameter λ

- The ability to find the sparse solution is given by whise choices wich be adopted for the sequence $\{\lambda_t\}_{t\geq 0}$

$$\lambda_t = \gamma\lambda_{t-1} = \theta\gamma^t \qquad \text{for } t \geq 1,$$

where $\lambda_0 = \theta$ and γ are positive and fixed constants.

- Possible choices of stopping criteria may include to bound the difference between two successive iterates, that is, until $\| \alpha_n - \alpha_{n-1} \| \geq \varepsilon$, or the discrepancy between the value $\left\| P \left[\sum_{k=0}^{\infty} \alpha_k \odot e^{-\lambda_k |\alpha_k|} \right] \right\|$ and zero

- Sparsity measure "g_λ-norm"

$$g_\lambda(|\alpha|) = \sum_{i=1}^{m}\left(1 - e^{-\lambda|\alpha_i|}\right) = m - \sum_{i=1}^{m} e^{-\lambda|\alpha_i|}, \qquad (6)$$

whith $\lambda > 0$.

- its importance derives from the fact that g-norm behaves like the ℓ^p-norm for $p = 1/\lambda < 1$ and becomes $\| \alpha \|_0$ as λ tends to $+\infty$, as stated in the following properties:

$$g_\lambda(|\alpha|) \approx \| \alpha \|_p^p \quad (\text{with } p = 1/\lambda < 1)$$

$$\lim_{\lambda\to+\infty} g_\lambda(|\alpha|) = \| \alpha \|_0 \quad (\text{asymptotically}).$$

Sparsity Minimization

- It would seem very natural to attempt to solve a new regularization problem of the form:

$$\min_{\alpha} \|\alpha\|_g \quad \text{subject to} \quad \Phi\alpha = s \tag{P1}$$

by applying the mapping (4) and by letting λ tend to infinity.

Sparsity Minimization

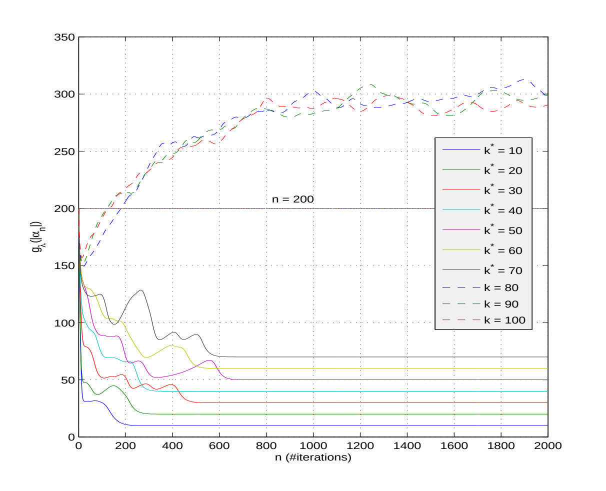

Figure: Plotting of the g_λ-norm functional versus iterations of the **while** loop of Algorithm 1. $n = 200$, $m = 800$

Numerical Results

- We evaluate the performances of the algorithms measuring relative error and computation time.
- as errors we consider the Signal-to-Noise-Ratio (SNR) and the Sum of Squares Error (SSE) of an approximate solution found α with respect the optimum α^*, defined as usual:

$$\text{SNR} = 20 \log_{10} \frac{\|\alpha\|}{\|\alpha - \alpha^*\|}, \quad \text{SSE} = \|s - \Phi\hat{\alpha}\|^2;$$

Numerical Results: Signal To Noise Ratio

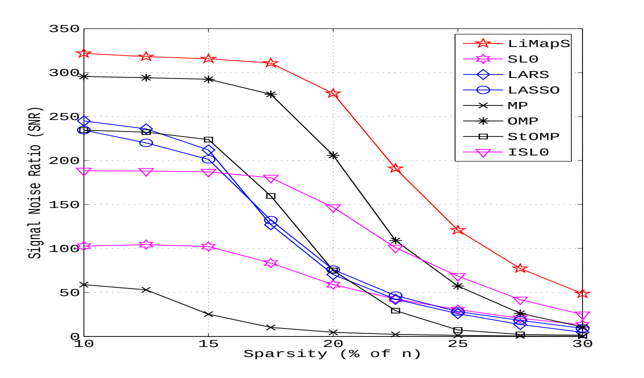

Figure: $n = 200$, $m = [300, 1400]$, 1000 trials

Numerical Results: Times

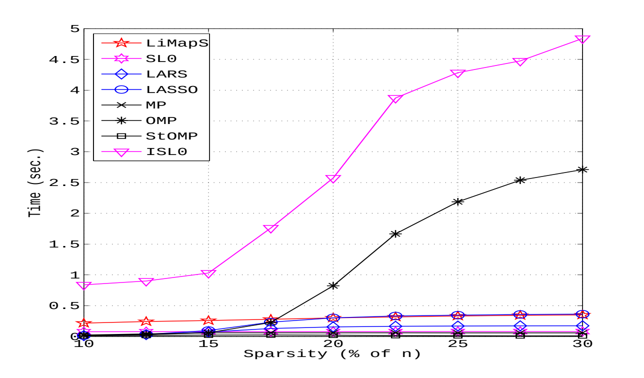

Figure: $n = 200$, $m = [300, 1400]$, 1000 trials

Numerical Results

	μ_{SSE}	σ_{SSE}	μ_{SNR}	σ_{SNR}	μ_{time}	σ_{time}
LIMAPS	**1.5e-24**	1.3e-24	**249.8**	114.9	0.79	0.23
SL0	4.8e-24	7.6e-25	24.7	38.3	**0.15**	0.01
ISL0	4.3e-16	3.7e-15	82.7	89.5	9.10	12.70
LASSO	1.3e+02	1.1e+03	8.3	2.0	1.79	0.20
LARS	2.3e-10	2.3e-09	6.3	2.2	0.79	0.07
MP	2.4e+04	4.3e+03	1.9	0.7	0.18	0.01
OMP	2.4e+00	2.8e-01	1.6	5.5	11.4	0.79
StOMP	3.8e+05	1.4e+05	2.4	0.7	0.02	0.01

Table: Instances of dimensions $m = 800$, $n = 400$ and $k = 200$, 1000 trials.

Conclusions and Future Works

- a critical aspect of the proposed framework deserves a deeper analysis in order to make the search strategy more efficient, reducing the number of iterations of the main cycle and avoiding to get stuck into local minima as well.
- such technique is promising because it exhibits very good performances (high SNR) also in case of very low sparsity (near $n/2$), values for which many others fail.

Thanks ...