Programme de la journée thématique du GDR MIA "Parcimonie et applications"

Le 03.05.2018 à l'Institut de Mathématiques de Bordeaux

Conférences plénières

Parcimonie et problèmes inverses

Nelly Pustelnik (CNRS, ENS Lyon)

Résumé : La première partie de cet exposé détaillera les avancées majeures en problèmes inverses qui découlent de la notion de parcimonie et requièrent à la fois des représentations sophistiquées (trames, variation totale non locale,...) et des algorithmes adaptés (algorithmes proximaux). Nous présenterons ensuite une contribution récente utilisant la notion de parcimonie dans le modèle de Mumford-Shah permettant de combiner les étapes de restauration et de détection de contours. Plus précisément, nous proposons une adaptation de l'algorithme PALM pour résoudre une version discrète du problème de Mumford-Shah.

Parcimonie et dictionnaires continus - un point de vue variationnel

Vincent Duval (INRIA, Equipe Mokaplan)

Résumé : De nombreux problèmes inverses en traitement du signal et des images visent à reconstruire des objets qui vivent dans un domaine continu (fréquence d'un signal, position d'un point lumineux, contour d'un objet...). Une pratique courante consiste à introduire une grille discrète pour décrire et manipuler ces objets sur ordinateur, et éventuellement résoudre dans ce cadre un problème variationnel . Cet exposé portera sur des approches variationnelles sans grille, apparues récemment autour de la minimisation L1 (LASSO, basis pursuit). On travaille alors avec un dictionnaire continu, et le problème d'optimisation associé est formulé dans l'espace des mesures (de Radon). Pour les problèmes de déconvolution et de super-résolution, nous verrons que cette formulation continue, a priori moins accessible du point de vue de la théorie et de l'implémentation numérique, permet en fait de lever plusieurs limitations des modèles discrets, que ce soit en terme de garanties théoriques ou de performance algorithmique, et de poser un regard nouveau sur le comportement des problèmes discrets utilisés depuis bien longtemps.

Approche bayésienne de type Approximate Message Passing pour la localisation de source en milieu fluctuant.

Guillaume Beaumont[1], Ronan Fablet[2], Angélique Drémeau[1]

March 2018

[1]Lab-STICC UMR 6285, CNRS, ENSTA Bretagne, Brest, F-29200, France
[2]Lab-STICC UMR 6285, CNRS, IMT-Atlantique, Brest F-29200, France

Abstract

En acoustique sous-marine, la propagation d'une onde peut être grandement perturbée par les fluctuations du milieu de propagation.

Plus spécifiquement, les mesures de phase du champ de pression complexe peuvent alors être fortement perturbées et mettre en échec les algorithmes classiques d'estimation de direction d'arrivée. Ces fluctuations ont cependant largement été étudiées dans la littérature et ont permis de définir des statistiques sur l'impact de cette perturbation sur le signal acoustique reçu.

Dans ces travaux nous proposons une nouvelle approche bayésienne capable de prendre en compte cette perturbation sous la forme d'un a priori sur nos mesures. Ce modèle combiné à un a priori parcimonieux sur le nombre de directions d'arrivées nous permet d'obtenir une méthode hautement-résolue d'estimation des directions d'arrivées. Cette méthode est basée sur un algorithme de type Approximate Message Passing nommé "paSAMP '' (pour phase-aware Swept Approximate Message Passing) et peut être considéré comme une extension de l'algorithme de phase-retrieval "prSAMP" aux a priori informatifs sur le bruit de phase.

Testés sur données simulées reproduisant ces perturbations du milieu, paSAMP s'avère intégrer avec succès ce modèle génératif et offre de meilleures performances en terme de restitution des directions d'arrivée que d'autres approches conventionnelles (e.g. beamforming classique, MUSIC). De plus, il se révèle être plus robuste au bruit additif (type blanc gaussien) que d'autres méthodes variationnelles basées sur ce même modèle (e.g. exploitant une approximation de champ moyen).

A case of exact recovery using OMP with continuous dictionaries

Clément Elvira[(1)], Rémi Gribonval[(1)], Cédric Herzet[(1)] et Charles Soussen[(2)]
[(1)] Univ Rennes, Inria, CNRS, IRISA
mails : prenom.nom@inria.fr
[(2)] L2S - Laboratoire des signaux et systèmes
mail : charles.soussen@centralesupelec.fr

We present new theoretical results on sparse recovery guarantees for a greedy algorithm, orthogonal matching pursuit (OMP), in the context of continuous dictionaries. Consider a sparse linear combination of atoms from a dictionary parameterized by some real parameters, for example, a combination of shifted versions of a basic waveform as in the context of sparse spike deconvolution. Currently, performance guarantees for greedy algorithms are typically carried out in the discrete setting associated to a grid of atom parameters, and based on, e.g., the coherence of the considered discretized dictionary [1]. However, such analyses fail to be conclusive for grid-based approaches when the discretization step tends to zero, as the coherence goes to one. Instead, our analysis is directly conducted in the continuous setting. For atoms parametrized by a real parameter that are elements of the (infinite-dimensional) Hilbert space $L_2(\mathbb{R})$ of square integrable real functions, and such that the inner product between two atoms is the exponential of the negative absolute difference of the corresponding parameters, we show in the noise-free setting that OMP exactly recovers the atom parameters as well as their amplitudes, regardless of the number of distinct atoms. We exhibit a convolutive dictionary of exponentially decaying pulses for which the atoms have an analytic definition while their pairwise inner products have the prescribed form. The established guarantees rely on a proof technique which is the continuous equivalent of Tropp's Exact Recovery Condition (ERC) [1]. The proof exploits specific properties of the positive definite kernel between atom parameters defined by the inner product between the corresponding atoms. Future work will aim at characterizing the class of kernels for which such an analysis holds –in particular for higher dimensional parameters– and the compatibility of the guarantees with dimension reduction techniques such as sketching, which would pave the way to provably good greedy algorithms for compressive statistical learning [2]. In light of the existing links between Tropp's ERC and recovery guarantees for ℓ_1 minimization [3], an interesting question is whether these guarantees extend to sparse spike recovery with total variation norm minimization [4, 5].

References

[1] J. A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–-2242, 2004.

[2] R. Gribonval, G. Blanchard, N. Keriven, Y. Traonmilin. Compressive Statistical Learning with Random Feature Moments. *preprint arxiv*, 2017.

[3] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030—1051, 2006.

[4] V. Duval and G. Peyré. Exact Support Recovery for Sparse Spikes Deconvolution. *Foundations of Computational Mathematics*, 15(2):1315–1355, 2015.

[5] E. J. Candès and C. Fernandez-Granda. Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6), 906–-956, 2014.

Non-negative orthogonal greedy algorithms for sparse reconstruction

Thanh T. Nguyen, Jérôme Idier, Charles Soussen, El-Hadi Djermoune

March 13, 2018

In this communication, we address sparse approximation under non-negativity constraints, formulated as the ℓ_0 minimization problem:

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - H\boldsymbol{x}\|_2^2 \qquad \text{subject to} \qquad \left\{ \ \|\boldsymbol{x}\|_0 \leq K \ , \ \boldsymbol{x} \geq \boldsymbol{0} \ \right\}$$

We introduce a family of greedy algorithms, so-called Non-Negative Orthogonal Greedy (NNOG) algorithms, having the following structure:

1. Selection of a dictionary atom.
2. Sparse coefficient update by solving a Non-Negative Least Squares (NNLS) subproblem.
3. Support compression.

This family covers, among others, Non-Negative Orthogonal Matching Pursuit (NNOMP) [1] and the Non-Negative Orthogonal Least Squares versions (NNOLS, SNNOLS) introduced in [2]. The main structural difference between NNOG and the standard orthogonal greedy algorithms is the inclusion of the support compression step. Indeed, the sparse representation coefficients obtained when solving the NNLS subproblems are likely to vanish when non-negativity constraints are active (that is, when $x_i = 0$). This motivated us to propose the support compression step, which removes the zero-valued coefficients from the support. The support compression step was omitted in earlier works [1, 2] and enables us to maintain consistency between the support and iterates, and also to yield residual vectors that are orthogonal to the span of the selected atoms.

Just like standard orthogonal greedy algorithms in the unconstrained case, NNOG algorithms can be interpreted as descent algorithms aiming to minimize the approximation error $\|\boldsymbol{y} - H\boldsymbol{x}\|_2^2$. On the implementation viewpoint, the main difference with the unconstrained case is the need for solving NNLS subproblems at each iteration instead of unconstrained least squares subproblems. Unfortunately, NNLS subproblems do not have a closed-form solution. We thus propose to make use of the active-set implementation of Lawson and Hanson [3] with a warm start initialization, corresponding to the current NNOG iterate. Since the active-set algorithm has a greedy structure, the resulting NNOG implementations appear to be recursive and fully greedy. Numerical tests on sparse deconvolution problems confirm the efficiency of the proposed implementations. The NNOG algorithms are also compared in term of tradeoff computational time *vs* sparse support reconstruction accuracy on both simulated and real data.

References

[1] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equation", *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4813–4820, Nov. 2008.

[2] M. Yaghoobi and M. E. Davies, "Fast non-negative orthogonal least squares", in *Actes de la 23e conf. EUSIPCO*, Nice, France, Aug. 2015, pp. 479–483.

[3] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, pp. 149–199, Prentice-Hall, Saddle River, NJ, Society for Industrial and Applied Mathematics (SIAM) edition, 1974.

Thanh T. NGUYEN: Centre de Recherche en Automatique de Nancy, UMR 7039, 54506 Vandœuvre-lès-Nancy. Email: *thi-thanh.nguyen@univ-lorraine.fr*

Jérôme IDIER: Laboratoire des Sciences du Numérique de Nantes, UMR 6004, 44321 Nantes. Email: *jerome.idier@ls2n.fr*

Charles SOUSSEN: Laboratoire des Signaux et Systèmes, UMR 8506, 91192 Gif-sur-Yvette. Email: *charles.soussen@centralesupelec.fr*

El-Hadi DJERMOUNE: Centre de Recherche en Automatique de Nancy, UMR 7039, 54506 Vandoeuvre-les-Nancy. Email: *el-hadi.djermoune@univ-lorraine.fr*

Structured sparsity in inverse problems and support recovery with mirror-stratifiable functions

Guillaume GARRIGOS
École Normale Supérieure

Jalal FADILI
ENSICAEN, CNRS

Jérôme MALICK
CNRS, LJK

Gabriel PEYRÉ
École Normale Supérieure, CNRS

Lorenzo ROSASCO
Università degli Studi di Genova

Silvia VILLA
Politecnico di Milano

Résumé. We consider inverse problems in separable Hilbert spaces where the prior on the data is an assumption of structured sparsity. We look at a class of regularizers for which minimization algorithms identify in finite time the extended support of the original data. This is a direct consequence of a more general identification theorem, involving the mirror stratifiability of the regularizer, a notion developped in [1], and based on duality arguments. As a by-product, we derive improved rates of convergence for the minimization algorithms, like a new linear rate result for the soft-thresholding algorithm in $\ell^2(\mathbb{N})$ with no assumptions. We discuss as well whether stochastic algorithms can (or cannot) enjoy this identification property.

We then provide necessary and sufficient conditions for norm regularizers to be mirror stratifiable, and show its tight relationship with the geometry of the corresponding unit ball. We apply this characterization result to show that norm regularizers inducing group sparsity with overlap are not mirror-stratifiable. We then discuss how to adapt the notion of mirror-stratifiability to treat these regularizers.

Mots-clefs : Sparse inverse problems, support recovery, group sparsity, optimization algorithms

Références

[1] J. Fadili, J. Malick, and G. Peyré. *Sensitivity Analysis for Mirror-Stratifiable Convex Functions*. Preprint on arXiv:1707.03194, 2017.

[2] G. Garrigos, L. Rosasco and S. Villa. *Thresholding gradient methods in Hilbert spaces: support identification and linear convergence*. Preprint on arXiv:1712.00357, 2017.

[3] G. Garrigos, L. Rosasco and S. Villa. *Sparse Multiple Kernel Learning: Support Identification via Mirror Stratifiability* . Preprint on arXiv:1803.00783, 2018.

[4] J. Fadili, G. Garrigos, J. Malick, and G. Peyré. *Model Consistency for Learning with Mirror-Stratifiable Regularizers*. submitted, 2018.

[5] G. Garrigos, L. Rosasco and S. Villa. *Characterization of mirror-stratifiability for sparsity inducing norms*. Work in progress.

Stable recovery of the factors from a deep matrix product

François Malgouyres
Institut de Mathématiques de Toulouse ; UMR5219
Université de Toulouse ; CNRS
UPS IMT, F-31062 Toulouse Cedex 9, France
http://www.math.univ-toulouse.fr/∼fmalgouy/

Joseph Landsberg
Department of Mathematics
Mailstop 3368
Texas A& M University
College Station, TX 77843-3368
Email: jml@math.tamu.edu

***Abstract*—We study a deep matrix factorization problem. It takes as input the matrix X obtained by multiplying K matrices (called factors) and aims at recovering the factors. When $K = 1$, this is the usual compressed sensing framework; $K = 2$: Examples of applications are dictionary learning, blind deconvolution, self-calibration; $K \geq 3$: can be applied to many fast transforms (such as the FFT). In particular, we apply the theorems to deep convolutional network.**

Using a Lifting, we provide : a necessary and sufficient conditions for the identifiability of the factors (up to a scale indeterminacy); - an analogue of the Null-Space-Property, called the Deep-Null-Space-Property which is necessary and sufficient to guarantee the stable recovery of the factors.

The long article corresponding to this work is available in [1].

I. Introduction

Let $K \in \mathbb{N}^*$, $m_1 \dots m_{K+1} \in \mathbb{N}$, write $m_1 = m$, $m_{K+1} = n$. We impose the factors to be structured matrices defined by a (typically small) number S of unknown parameters. More precisely, for $k = 1 \dots K$, let

$$\begin{array}{rcl} M_k : \mathbb{R}^S & \longrightarrow & \mathbb{R}^{m_k \times m_{k+1}}, \\ h & \longmapsto & M_k(h) \end{array}$$

be a linear map.

We assume that we know the matrix $X \in \mathbb{R}^{m \times n}$ which is provided by

$$X = M_1(\mathbf{h}_1) \cdots M_K(\mathbf{h}_K) + e, \tag{1}$$

for an unknown error term e and parameters $\mathbf{h} = (\mathbf{h}_k)_{1 \leq k \leq K} \in \mathcal{M}^{\overline{L}} \subset \mathbb{R}^{S \times K}$ for some $\overline{L}$, where we assume that we know a collection of models $\mathcal{M} = (\mathcal{M}^L)_{L \in \mathbb{N}}$ such that, for every L, $\mathcal{M}^L \subset \mathbb{R}^{S \times K}$.

This work investigates models/constraints imposed on (1) for which we can (up to obvious scale rearrangement) identify or stably recover the parameters $\mathbf{h}$ from X. A preliminary version of this work is presented in [2].

Set $\mathbb{N}_K = \{1, \dots, K\}$ and

$$\mathbb{R}_*^{S \times K} = \{\mathbf{h} \in \mathbb{R}^{S \times K}, \forall k \in \mathbb{N}_K, \|\mathbf{h}_k\| \neq 0\}.$$

Define an equivalence relation in $\mathbb{R}_*^{S \times K}$: for any $\mathbf{h}$, $\mathbf{g} \in \mathbb{R}^{S \times K}$, $\mathbf{h} \sim \mathbf{g}$ if and only if there exists $(\lambda_k)_{k \in \mathbb{N}_K} \in \mathbb{R}^K$ such that

$$\prod_{k=1}^{K} \lambda_k = 1 \qquad \text{and} \qquad \forall k \in \mathbb{N}_K, \mathbf{h}_k = \lambda_k \mathbf{g}_k.$$

Denote the equivalence class of $\mathbf{h} \in \mathbb{R}_*^{S \times K}$ by $[\mathbf{h}]$. We consider a metric denoted d_p on $\mathbb{R}_*^{S \times K} / \sim$. It is based on the l^p norm.

We say that a tensor $T \in \mathbb{R}^{S^K}$ is of *rank* 1 if and only if there exists a collection of vectors $\mathbf{h} \in \mathbb{R}^{S \times K}$ such that T is the outer product of the vectors $\mathbf{h}_k$, for $k \in \mathbb{N}_K$, that is, for any $\mathbf{i} \in \mathbb{N}_S^K$,

$$T_{\mathbf{i}} = \mathbf{h}_{1,i_1} \dots \mathbf{h}_{K,i_K}.$$

The set of all the tensors of rank 1 is denoted by Σ_1.

Moreover, we parametrize $\Sigma_1 \subset \mathbb{R}^{S^K}$ by the Segre embedding

$$\begin{array}{rcl} P : \mathbb{R}^{S \times K} & \longrightarrow & \Sigma_1 \subset \mathbb{R}^{S^K} \\ \mathbf{h} & \longmapsto & (\mathbf{h}_{1,i_1} \mathbf{h}_{2,i_2} \dots \mathbf{h}_{K,i_K})_{\mathbf{i} \in \mathbb{N}_S^K} \end{array}$$

Following [3], [4], [5], [6], [7], [8] where problems such that $K = 2$ are studied, we can *lift* the problem and show that the map

$$(\mathbf{h}_1, \dots, \mathbf{h}_K) \longmapsto M_1(\mathbf{h}_1) M_2(\mathbf{h}_2) \dots M_K(\mathbf{h}_K),$$

uniquely determines a linear map

$$\mathcal{A} : \mathbb{R}^{S^K} \longrightarrow \mathbb{R}^{m \times n},$$

such that for all $\mathbf{h} \in \mathbb{R}^{S \times K}$

$$M_1(\mathbf{h}_1) M_2(\mathbf{h}_2) \dots M_K(\mathbf{h}_K) = \mathcal{A} P(\mathbf{h}).$$

When $\|e\| = 0$, we can prove that every element of $\mathbf{h} \in \mathcal{M}$ is identifiable (i.e. the elements of $[\mathbf{h}]$ are the only solutions of (1)) if and only if for any L and $L' \in \mathbb{N}$

$$\mathrm{Ker}\,(\mathcal{A}) \cap \left(P(\mathcal{M}^L) - P(\mathcal{M}^{L'})\right) = \{0\}.$$

When $\|e\| \leq \delta$, we further assume that we have a way to find L^* and $\mathbf{h}^* \in \mathcal{M}^{L^*}$ such that, for some parameter $\eta > 0$,

$$\|\mathcal{A} P(\mathbf{h}^*) - X\|^2 \leq \eta. \tag{2}$$

Definition 1. Deep-Null Space Property

Let $\gamma > 0$, we say that $\mathrm{Ker}\,(\mathcal{A})$ *satisfies the* deep-Null Space Property (deep-NSP) *with respect to the model collection $\mathcal{M}$ with constant γ if there exists $\varepsilon > 0$ such that for any L and $L' \in \mathbb{N}$, any $T \in P(\mathcal{M}^L) - P(\mathcal{M}^{L'})$ satisfying $\|\mathcal{A}T\| \leq \varepsilon$ and any $T' \in \mathrm{Ker}\,(\mathcal{A})$, we have*

$$\|T\| \leq \gamma \|T - T'\|.$$

Theorem 1. Sufficient condition for stable recovery

Assume $\mathrm{Ker}\,(\mathcal{A})$ *satisfies the deep-NSP with respect to the collection of models $\mathcal{M}$ and with the constant $\gamma > 0$. For any $\mathbf{h}^*$ as in* (2) *with η and δ sufficiently small, we have*

$$\|P(\mathbf{h}^*) - P(\overline{\mathbf{h}})\| \leq \frac{\gamma}{\sigma_{min}} (\delta + \eta),$$

where σ_{min} is the smallest non-zero singular value of $\mathcal{A}$. Moreover, if $\overline{\mathbf{h}} \in \mathbb{R}_*^{S \times K}$

$$d_p([\mathbf{h}^*], [\overline{\mathbf{h}}]) \leq \frac{7 (KS)^{\frac{1}{p}} \gamma}{\sigma_{min}} \min \left(\|P(\overline{\mathbf{h}})\|_\infty^{\frac{1}{K}-1}, \|P(\mathbf{h}^*)\|_\infty^{\frac{1}{K}-1} \right) (\delta + \eta).$$

We also prove that the deep-NSP condition is **necessary** for the stable recovery of the factors. We detail how these results can be applied to obtain sharp conditions for the stable recovery of deep convolutional network as depicted on Figure 1.

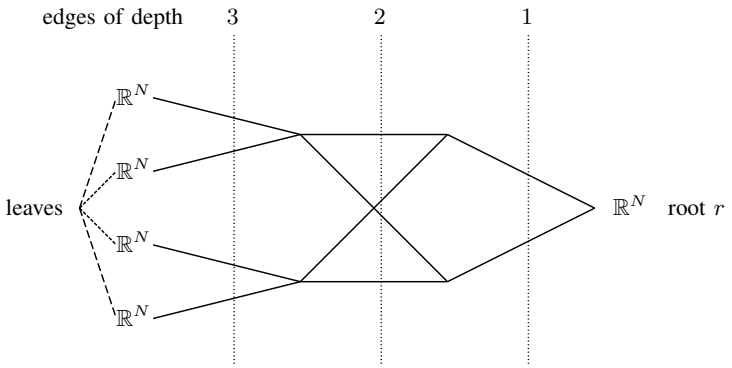

Fig. 1. Example of the considered convolutional network. To every edge is attached a convolution kernel. The network does not involve non-linearities or sampling.

REFERENCES

[1] F. Malgouyres and J. Landsberg, "Stable recovery of the factors from a deep matrix product and application to convolutional network," *preprint hal-01494267*, 2017.

[2] ——, "On the identifiability and stable recovery of deep/multi-layer structured matrix factorization," in *IEEE, Info. Theory Workshop*, Sept. 2016.

[3] A. Ahmed, B. Recht, and J. Romberg, "Blind deconvolution using convex programming," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1711–1732, 2014.

[4] S. Choudhary and U. Mitra, "Identifiability scaling laws in bilinear inverse problems," *arXiv preprint arXiv:1402.2637*, 2014.

[5] X. Li, S. Ling, T. Strohmer, and K. Wei, "Rapid, robust, and reliable blind deconvolution via nonconvex optimization," *CoRR*, vol. abs/1606.04933, 2016. [Online]. Available: http://arxiv.org/abs/1606.04933

[6] S. Bahmani and J. Romberg, "Lifting for blind deconvolution in random mask imaging: Identifiability and convex relaxation," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2203–2238, 2015.

[7] S. Ling and T. Strohmer, "Blind deconvolution meets blind demixing: Algorithms and performance bounds," *arXiv preprint arXiv:1512.07730*, 2015.

[8] ——, "Self-calibration and biconvex compressive sensing," *Inverse Problems*, vol. 31, no. 11, p. 115002, 2015.

Multichannel Cosparse Declipping: Structure Helps

Clément Gaultier, Nancy Bertin, Rémi Gribonval
Univ Rennes, Inria, CNRS, IRISA, France

Abstract

In this work, we investigate the performance of structured cosparse regularizations in jointly restoring 8-channels saturated audio recordings. Beyond the cosparse (also known as sparse analysis) model, results show that exploiting structures across channels is beneficial compared to simple use of channel-by-channel independent cosparse prior.

Motivated by the success of the sparse analysis version of the SPADE algorithm [1] for declipping, as well as the potential of structured sparsity (especially [2] that was able to capture some typical time-frequency patterns), we postulate that coupling these two concepts could be beneficial to audio restoration. The main model characteristics for this work derive from the relation between the time-domain signal of interest $\mathbf{X}$ and its frequency representation $\mathbf{Z}$ as well as properties of $\mathbf{Z}$. The underlying hypothesis behind the structure (group sparsity) in the frequency representation $\mathbf{Z}$ is that non zero coefficients are roughly distributed equivalently from one channel to another. This work compares the performance of structured cosparse models to state-of-the-art regular cosparse method on the multichannel audio declipping problem.

Cosparse regularized approaches to inverse problems can be cast as an optimization problem, where the cost function to minimize is a sum of a data-fidelity term and a regularization term enforcing sparsity. To address the multichannel declipping issue, we define an iterative ADMM procedure [3] in which, the estimate is alternatively projected on the structured cosparsity constraint and the data-fidelity (declipping) constraint. A well-chosen *sparsifying operator* is applied at each iteration and acts as a proxy for the cosparse regularization term. The choice of this proxy is the key difference between regular cosparse and structured cosparse algorithms. In the regular cosparse case, the well-known *hard-thresholding* operator is applied at each iteration on the current estimate $\mathbf{A}\hat{\mathbf{X}}$. In the structured cosparse case, as we wish to promote some particular structures across channels and frequency, hard-thresholding is replaced by the *Group Empirical Wiener shrinkage*. The latter was successfully used in [2] for signal decomposition. For the data-fidelity projection, we derive a component-wise closed form solution from the magnitude declipping constraints sets. This procedure, involving alternatively a sparsifying step and data-fidelity projection is applied on frames. The $\hat{\mathbf{X}}$ declipped estimates are used to rebuild the full length estimated signal by means of overlap-and-add synthesis.

The joint use of cosparse and structured sparsity models is particularly efficient on music and speech data from [4] recorded with a 8-channels compact microphones antenna. We show that our method numerically outperforms state-of-the-art simple cosparse A-SPADE algorithm [1] by 1 dB to more than 3 dB while retaining very limited runtime overcost.

References

[1] S. Kitić, N. Bertin, and R. Gribonval, "Sparsity and cosparsity for audio declipping: a flexible non-convex approach," in *Latent Variable Analysis and Signal Separation (LVA/ICA)*. Liberec, Czech Republic: Springer, 2015, pp. 243–250.

[2] C. Févotte and M. Kowalski, "Hybrid sparse and low-rank time-frequency signal decomposition," in *23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 464–468.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[4] N. Bertin, E. Camberlein, R. Lebarbenchon, E. Vincent, S. Sivasankaran, I. Illina, and F. Bimbot, "VoiceHome-2, an extended corpus for multichannel speech processing in real homes," *To appear in Speech Communication*, 2017.

A Dual Certificates Analysis of Compressive Off-the-Grid Recovery

Clarice Poon Nicolas Keriven Gabriel Peyré

Many problems in machine learning and imaging can be framed as an infinite dimensional Lasso problem to estimate a sparse measure $\mu_0 \in \mathcal{M}(\mathcal{X})$. This includes for instance regression using a continuously parameterized dictionary, mixture model estimation and super-resolution of images.

To make the problem tractable, one typically sketches the observations (often called compressive-sensing in imaging) using randomized projections:

$$\forall\, k = 1, \dots, m, \quad y_k \overset{\text{def.}}{=} \langle \varphi_{\omega_k}, \mu_0 \rangle + \varepsilon_k \quad \text{where} \quad \langle \varphi, \mu \rangle \overset{\text{def.}}{=} \int_{\mathcal{X}} \varphi(x) \mathrm{d}\mu(x) \in \mathbb{C},$$

where $\varepsilon_k \in \mathbb{C}$ accounts for noise or modelling errors, $(\omega_1, \dots, \omega_m)$ are identically and independently distributed according to some probability distribution $\Lambda(\omega)$ on $\omega \in \Omega$, and $\varphi_\omega : \mathcal{X} \to \mathbb{C}$ is a continuous function. The studied Lasso problem is:

$$\min_{\mu \in \mathcal{M}(\mathcal{X})} \frac{1}{2m} \sum_{k=1}^{m} |\langle \varphi_{\omega_k}, \mu \rangle - y_k|^2 + \lambda |\mu|(\mathcal{X}).$$

where $|\mu|(\mathcal{X})$ is the total variation of μ.

In this work, we provide a comprehensive treatment of the recovery performances of this class of approaches. We show that sufficient conditions can be expressed on the expectation of the covariance kernel $K(x, x') = \mathbb{E}_\omega \varphi_\omega(x) \overline{\varphi_\omega(x')}$ and a minimal separation of Diracs, thus extending classical proofs in a generic, possibly multi-dimensional setting. We give two examples, the Fejér kernel, which corresponds to discrete Fourier sampling on the (multi-dimensional) torus, and the gaussian kernel, which can be seen as continuous Fourier sampling with gaussian frequencies.

Then, our main contribution is two-fold:

- we prove that, up to log factors, a number of sketches proportional to the sparsity is enough to obtain approximate stability (localization of the atoms) with robustness to noise. The proof introduces a new infinite-dimensional variant of the classical golfing scheme.
- using more classical proof techniques generalized to our framework, we show that a number of sketches quadratic in the sparsity is enough to obtain *exact* support stability (the number of recovered atoms matches that of the measure of interest) in the small noise regime. This number can still be linear, under the somewhat unrealistic assumption that the signs of the sought after Diracs are random.

We finish by giving a few examples, including sketched Gaussian Mixture Model estimation. Unlike non-convex log-likelihood approaches, the infinite-dimensional Lasso may recover exactly the number of components in the mixture when the exact support stability result applies.

When Safe Screening and Structured Dictionaries Join Forces

Cássio Fraga Dantas, Rémi Gribonval
Univ Rennes, Inria, CNRS, IRISA, France

Abstract

We propose a way to combine two existing acceleration techniques for the Lasso problem: safe screening tests, which simplify the problem by eliminating useless dictionary atoms; and the use of structured dictionaries which are faster to operate with. A structured approximation of the true dictionary is used at the initial stage of the optimization, and we show how to define screening tests which are still safe despite the approximation error. In particular, we extend a state-of-the-art screening test, the GAP SAFE sphere test, to this new setting. The practical interest of the proposed methodology is demonstrated by considerable reductions in simulation time.

The l_1-regularized least-squares, referred to as Lasso, is a is a ubiquitous tool for variable selection in the context of underdetermined linear regression problems. Due to its convex cost function, fast solvers with strong theoretical guarantees are available. Nevertheless, for large scale problems such methods may become computationally prohibitive and, for this reason, accelerating techniques are still an intense research topic.

In this work, we demonstrate how to combine two of such techniques:

1) *Structured dictionaries* provide faster matrix-vector multiplications, which dominate the cost of typical iterative optimization algorithms for the Lasso, such as the iterative soft-thresholding algorithms (e.g. ISTA and FISTA).
2) *Safe screening tests* first proposed in [1] allow to safely eliminate inactive dictionary atoms (those associated to zero entries in the solution vector) before having complete knowledge of the solution, with minor computational overhead. In this work, we show how to obtain screening tests that manipulate an approximate (and faster) version of the true dictionary, but remain safe with respect to the original problem, i.e. to the atoms of the true dictionary. In [2], we extended a particular screening test called Dynamic Spherical Test [3] to this new setting, and in [4] we extended a more complex test called GAP SAFE [5] which is the current state-of-the-art in terms of screening capabilities.

The overall idea is the following: starting the iterative Lasso optimization by manipulating the fast approximate dictionary to take advantage of its reduced multiplication cost, and at some point – for instance, when a considerable portion of the atoms have been screened out – switching back to the original dictionary which is now much less costly to operate with and ensures convergence to the solution of the original problem.

Simulation results on synthetic data proves the effectiveness of the proposed technique. Significant reductions in execution time are obtained in comparison to screening rules alone, which, in turn, already represent a considerable improvement with respect to not screening at all. The proposed framework could also be extended to other sparsity-inducing inverse problems such as the Group-Lasso or the regularized logistic regression. Additional experiments with real datasets are a short-term perspective as well as handling multiple approximations of the dictionary with different error levels and complexity gains.

REFERENCES

[1] L. El Ghaoui, V. Viallon, and T. Rabbani, "Safe feature elimination in sparse supervised learning," *UC Berkeley, Tech. Rep*, 2010.
[2] C. F. Dantas and R. Gribonval, "Dynamic Screening with Approximate Dictionaries," in *XXVIème colloque GRETSI*, Juan-les-Pins, France, 2017. [Online]. Available: https://hal.inria.fr/hal-01598021
[3] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, "Dynamic screening: Accelerating first-order algorithms for the lasso and group-lasso," *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5121–5132, 2015.
[4] C. F. Dantas and R. Gribonval, "Faster and still safe: combining screening techniques and structured dictionaries to accelerate the Lasso," in *ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, AB, Canada, 2018. [Online]. Available: https://hal.inria.fr/hal-01706392
[5] O. Fercoq, A. Gramfort, and J. Salmon, "Mind the duality gap: safer rules for the lasso," in *Proc. ICML 2015*, July 2015.