# Invariant Multiscale Statistics for Inverse Problems

Ivan Dokmanić
CSL, ECE, UIUC
Urbana, IL, USA
dokmanic@illinois.edu

Joan Bruna
Courant Institute, NYU
New York, NY, USA
bruna@cims.nyu.edu

Stéphane Mallat
Ecole Normale Superiure
Paris, France
stephane.mallat@ens.fr

Maarten de Hoop
CAAM, Rice University
Houston, TX, USA
mdehoop@rice.edu

***Abstract*—We propose a new approach to linear ill-posed inverse problems. Our algorithm stabilizes the inversion by enforcing a new statistical constraint in a suitable feature space. We use the non-linear multiscale scattering transform—a complex convolutional network which discards the phase and thus exposes strong spectral correlations otherwise hidden beneath the phase fluctuations. We apply the algorithm to super-resolution and tomography with synthetic signals, and show that it outperforms regularized methods and stably recovers the missing spectrum. Further, we discuss the choice of the feature transform as a function of the operator and input statistics, and we prove convergence of the proposed iterative algorithm.**

## I. A New Approach to Inverse Problems

A standard inverse problem in imaging is to estimate $x \in \mathcal{X}$ from measurements $y$ corrupted by noise $b$:

$$y = \Gamma x + b, \tag{1}$$

with $\Gamma$ being a singular operator so that the inversion is ill posed. Examples of $\Gamma$ are lowpass filtering and partial Radon transform.

The usual way to address ill-posedness is to search for a solution which minimizes a regularized cost functional [1], [2]:

$$\widehat{x} = \underset{u \in \mathcal{X}}{\arg\min} \ \tfrac{1}{2}\|y - \Gamma u\|^2 + \lambda h(u), \tag{2}$$

where the regularizer $h$ is convex in $u$. Restriction to convex functionals of this form may be a bottleneck when the problem is highly ill-posed.

We propose a formulation that is not based on a minimization of a regularized cost function,[1] but rather on iterative linear estimation in the space of some feature transform $\Phi$. We show that a particularly good choice of $\Phi$ for many inverse problems is the *scattering transform* [3], [4]. Scattering transform has a structure of a convolutional network with complex wavelets as filters and complex modulus as the non-linearity. Its outputs are made locally invariant to translations by averaging. By eliminating the phase it yields coefficients with strong linear correlations.

For many $\Gamma$, solving the inverse problem can be rephrased as recovering missing spectrum from known spectrum (see Fig. 1), and the role of the regularizer is to stabilize this inversion (e.g. stabilize the high frequencies). In what follows, we describe how we achieve such stabilization in a statistical, data-driven manner.

Let $\mathcal{A} \overset{\text{def}}{=} \{u \ : \ \|\Gamma u - y\| \leqslant \epsilon\}$ be the set of admissible signals (random vectors). The idea is to search for a signal $\widehat{x}$ in $\mathcal{A}$ such that applying any linear estimator to $\Phi\widehat{x}$ cannot improve the feature-space MSE. Assuming $\Phi$ takes values in $\mathbb{R}^d$, this can be written as follows:

$$\widehat{x} \in \mathcal{A}, \tag{A1}$$

$$\mathbb{E}[\Phi\widehat{x}] = \mathbb{E}[\Phi x], \tag{A2}$$

$$\mathbb{E}[(\Phi\widehat{x})_i(\Phi\widehat{x} - \Phi x)] = \mathbf{0}_d, \ \forall i \in \{1, \ldots, d\}\,. \tag{A3}$$

The condition (A3) resembles the usual orthogonality relation, although, importantly, it contains no data term—it simply states the constraint on our estimator. The defined $\widehat{x}$ is not linear in either domain.

More generally, the transform $\Phi$ should be adapted to $\Gamma$ and the input statistics; it could be learned or designed. We show (and it is not difficult to see) that $\Phi$ should not be chosen linear. It is important to note that it need not be invertible.

[1] At least not in the usual sense where it is determined by a single realization.

## II. Computing the Estimator

To compute $\widehat{x}$, we define two operators acting on vectors in the feature space. The first one, denoted $\mathsf{P}_{\Phi\mathcal{A}}$, is a projection onto $\Phi\mathcal{A}$—the image of $\mathcal{A}$ under $\Phi$. For any $u \in \mathcal{X}$, $\mathsf{P}_{\Phi\mathcal{A}}(\Phi u)$ satisfies (A1). Since $\Phi$ is non-linear, $\Phi\mathcal{A}$ is in general not a convex set even when $\Gamma$ is linear. We implement $\mathsf{P}_{\Phi\mathcal{A}}$ by projected gradient descent. The second operator, denoted $\mathsf{L}$, is a linear-minimum-mean-squared-error (LMMSE) estimator of $\Phi x$ given some input data $\Phi u$. For any $u \in \mathcal{X}$, $\mathsf{L}(\Phi u)$ satisfies (A2, A3). To compute $\mathsf{L}$ in practice, we use empirical estimates of the involved first- and second-order statistics.

In general, for some $u$, $\mathsf{P}_{\Phi\mathcal{A}}(\Phi u)$ will not satisfy (A2, A3), and $\mathsf{L}(\Phi u)$ (or its preimage) will not satisfy (A1). Thus, not surprisingly, we propose to iterate the two operators which gives the following algorithm:

$$\Phi x^{(k+1)} = \mathsf{P}_{\Phi\mathcal{A}}\big(\mathsf{L}\big(\Phi x^{(k)}\big)\big). \tag{3}$$

A signal-domain estimate can be computed as any vector in the preimage $x^{(k)} \in \Phi^{-1}\Phi x^{(k)}$. This inversion is again implemented by gradient descent.

We prove that for many $\Phi$ the above iteration converges to a fixed point satisfying (A1)-(A3). The analysis shows that (3) is a new kind of non-convex alternating projection algorithm with unusual geometry.

## III. Experiments

In Fig. 2 we show preliminary results for synthetic signals in super-resolution and tomography. As input we use Ising spin-glass realizations [5] since they look like a mixture of shapes and texture, and Cox point processes [6] for comparison with sparsity-promoting regularizations. Point processes are sparse in the identity basis and thus present an ideal signal for $\ell_1$ norm minimization.

Fig. 2 shows that our approach produces results with significantly better statistics than convex regularizations. Since we are not optimizing the signal-domain $\ell_2$ error, convex approaches give lower MSE due to slight misalignment of the many singularities with our method. However, convex results clearly have wrong spatial statistics, especially at high frequencies. To demonstrate this quantitatively, we compare the multivariate kurtosis [7] on 8 by 8 windows between the original image and the various reconstructions, and show that it is reproduced much more accurately by our method. This generalizes to other higher-order spatial moments.In the super-resolution experiment the downsampling is by a factor of 16 along each side and the results should be interpreted in light of this rather severe (256-fold) data loss.

Fig. 3 gives reconstructions of a point process with a single iteration of (3). Measurements were obtained by Gaussian filtering and downsampling by a factor of 4 along each axis. The number of non-zeros is set so that the $\ell_1$ minimization does not have a unique solution [8] which leads to artifacts, whereas the proposed method recovers a signal with correct spatial statistics. A major part of ongoing work is an analysis of the reconstruction methods in terms of optimal transport metrics.

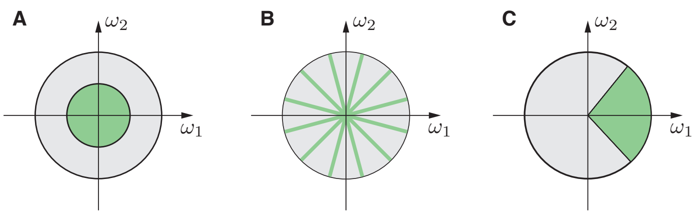


Fig. 1. Known part of the Fourier transform with super-resolution measurements (A); Radon transform uniformly subsampled in angle (B); Radon transform with directions restricted to a cone (C).

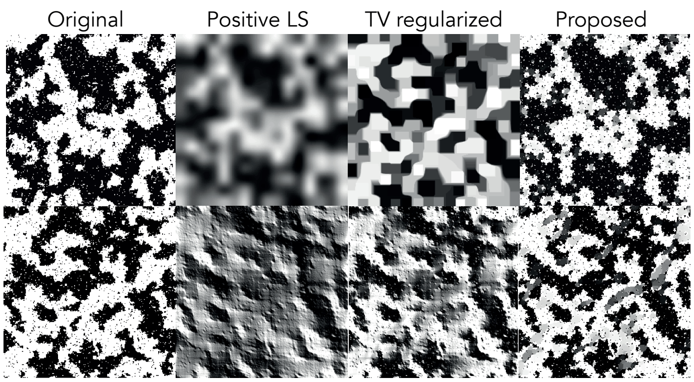


Fig. 2. 20 iterations of the proposed method compared with (positive) least-squares inversions and TV-norm regularized reconstructions for super-resolution (top) and tomography (bottom). The multiplier $\lambda$ (cf (2)) for the TV regularizer was picked by hand to give best MSE. MSE top row (smaller is better): 0, 0.860e4, 0.797e4, 1.291e4; excess kurtosis top row (closer to 1760 is better): 1760, 85, 7861, 1290; MSE bottom row: 0, 6.57e3, 3.90e3, 7.19e3; excess kurtosis bottom row: 1760, 2285, 3640, 1787.

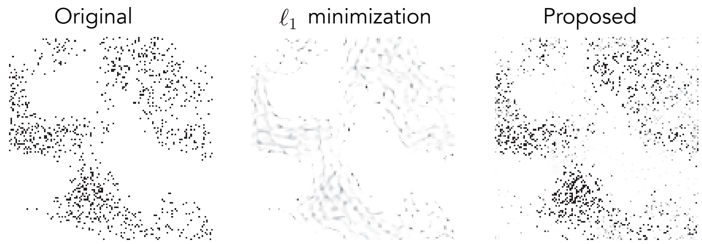


Fig. 3. Comparison with sparsity-promoting regularization for a point process in super-resolution (single iteration).

## References


[1] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Springer Science & Business Media, Mar. 2000.

[2] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pur. Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[3] S. Mallat, "Group Invariant Scattering," *Commun. Pur. Appl. Math.*, vol. 65, no. 10, pp. 1331–1398, Oct. 2012.

[4] J. Bruna and S. Mallat, "Invariant Scattering Convolution Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[5] K. Binder and A. P. Young, "Spin glasses: Experimental facts, theoretical concepts, and open questions," *Rev. Mod. Phys.*, vol. 58, no. 4, pp. 801–976, Oct. 1986.

[6] D. R. Cox and V. Isham, *Point Processes*, CRC Press, July 1980.

[7] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, Dec. 1970.

[8] D. L. Donoho, "Compressed Sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.