

INVERSE PROBLEMS WITH INVARIANT MULTISCALE STATISTICS

Ivan Dokmanić¹, Joan Bruna², Stéphane Mallat³, and Maarten de Hoop⁴

¹Institut Langevin, CNRS, ESPCI, PSL, Paris, France

²Courant Institute, NYU, New York, USA

³Ecole Normale Supérieure, PSL, Paris, France

⁴CAAM, Rice University, Houston, USA

ABSTRACT

We propose a new approach to linear ill-posed inverse problems. Our algorithm alternates between enforcing two constraints: the measurements and the statistical correlation structure in some transformed space. We use a non-linear multiscale scattering transform which discards the phase and thus exposes strong spectral correlations otherwise hidden beneath the phase fluctuations. As a result, both constraints may be put into effect by linear projections in their respective spaces. We apply the algorithm to super-resolution and tomography and show that it outperforms ad hoc convex regularizers and stably recovers the missing spectrum.

Index Terms—Inverse problems, tomography, super-resolution, scattering transform, regularization.

1. INTRODUCTION

In this paper we propose a new way to solve ill-posed linear inverse problems. We consider the usual problem statement: estimate $x \in \mathcal{X}$ from noisy measurements y by a singular operator Γ :

$$y = \Gamma x + b, \quad (1)$$

with b being the measurement noise. The estimator $\hat{x} = \hat{x}(y)$ of x is a function of the measurements y .

A standard approach to deal with ill-posedness is to cook up a regularizer $h(u)$ which promotes signals within the desired model and solve

$$\min_{u \in \mathcal{X}} d(y, \Gamma u) + \lambda h(u),$$

where $d(y, \Gamma u)$ is the data fidelity term [1, 2]. Both d and h are typically convex which may be undesirable in severely ill-posed problems (cf. Section 5.3).

We propose a different route. Measurements y reveal certain information about the signal x that can often be interpreted spectrally: a part of the spectrum is known from y , and solving the inverse problem means reconstructing the unknown part from the known part. This is illustrated in Fig. 1 for the two inverse problems we study—super-resolution and tomography. Any signal can be made to satisfy the measurements by projecting it in an appropriate space \mathcal{A} of signals consistent with measurements. For example, in the case of super-resolution, this projection replaces the low frequencies by the measured ones.

For some z , $P_{\mathcal{A}}(z)$ may be a bad estimate of x in the sense that the joint statistics of the known and unknown spectral regions will be wrong. We show that these joint statistics are captured by the *scattering coefficients* [3, 4]—local averages of moduli of complex

This work was supported by the ERC grant InvariantClass 320959. ID was funded by LABEX WIFI (Laboratory of Excellence within the French Program “Investments for the Future”) under references ANR-10-LABX-24 and ANR-10-IDEX-0001-02 PSL* and by Agence Nationale de la Recherche under reference ANR-13-JS09-0001-01. We would like to thank Laurent Daudet for valuable discussions in the early stages of this research.

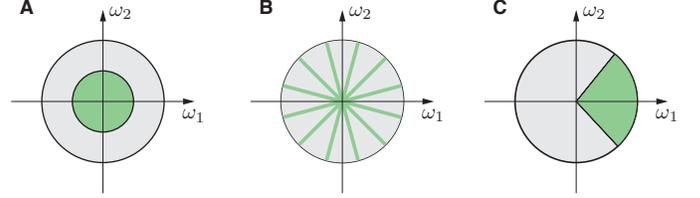


Fig. 1. Known part of the Fourier transform with super-resolution measurements (A); Radon transform uniformly subsampled in angle (B); Radon transform with directions restricted to a cone (C).

wavelet coefficients which encode interactions across scales. Scattering has been used earlier to build probabilistic models for super-resolution [5].

To see why we work with amplitudes and discard the phases, consider estimating a complex-valued random variable $X = |X|e^{j\phi_X}$ from $Y = |Y|e^{j\phi_Y}$. Assume that phases and magnitudes are independent, ϕ_X is distributed uniformly on a circle, and phases are perfectly correlated: $\phi_Y = -\phi_X$. One may expect that this correlation is beneficial. However, the quality of linear estimation depends on achieving significant linear correlations between coordinates of the transform, as measured by the correlation coefficients [6]:¹

$$\text{corr}(X_i, X_j) = \frac{\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]}{\sqrt{\mathbb{V}(X_i)\mathbb{V}(X_j)}}. \quad (2)$$

A quick computation shows that $\mathbb{E}[XY^*] = 0$, so that correlation is zero and at best we can restore the mean (0 in this case). In general, phases inevitably reduce linear correlations: an instance of a phenomenon known in statistics as regression to the mean.

By eliminating the phase, a scattering transform yields strongly correlated coefficients which can thus be estimated *linearly* from the scattering coefficients of the measured signal. This estimator defines another projection in the space of square integrable random variables. Signals obtained by this second projection will generally not belong to \mathcal{A} . We then propose a natural idea: iterate the two projections—a linear one onto \mathcal{A} and a non-linear one that adjusts the transform statistics—until we obtain a signal in \mathcal{A} with the right statistics.

We demonstrate the method in two inverse problems: super-resolution and tomography, showing that our approach stably estimates the missing spectral information. The paper is organized as follows: Section 2 reviews scattering transforms, Section 3 presents our main approach and Section 4 gives the algorithmic details. Finally, Section 5 presents numerical experiments.

2. SCATTERING TRANSFORM

A scattering representation is a vector of signal descriptors computed as moduli of iterated wavelet coefficients, which are made locally

¹Even in the Gaussian case, absence of correlation in the sense of (2) limits our estimation potential to that of restoring the mean.

invariant to translations. By eliminating the phase, it creates a set of highly correlated coefficients so that the missing spectrum can be linearly estimated from known coefficients at different scales.

Let $x(u)$ for $u \in \mathbb{Z}^d$ be a stationary process, with $\mathbb{E}[|x|^2] < \infty$. A wavelet transform analyzes multiscale variability of x through convolutions with dilated wavelet band-pass filters. For image processing, the wavelets ψ_k may be constructed by rotating a single wavelet ψ through an angle $k\pi/K$: $\psi_k(u) = \psi(r_k u)$. They are dilated diadically by factors of 2^j , $\psi_{j,k}(u) = 2^{-jn} \psi_k(2^{-j}u)$, up to the maximum scale 2^J . The resulting wavelet transform of x is

$$\mathcal{W}_J x = \{x \star \phi_J(u), x \star \psi_{j,k}(u)\}_{j \leq J, 1 \leq k \leq K}, \quad (3)$$

where $\phi_J(u) = 2^{-dJ} \phi(2^{-J}u)$ is a scaled averaging filter, with $\int \phi(u) du = 1$.

If x is stationary then $x \star \psi_{j,k}(u)$ is also stationary. We consider complex-valued wavelets ψ whose Fourier transform has a support concentrated on one-half of the Fourier domain—for any $\omega \in \mathbb{R}^d$ either $\hat{\psi}(\omega) \approx 0$ or $\hat{\psi}(-\omega) \approx 0$. We define K wavelets $\psi_k(u)$, which are regular functions with a fast decay and a zero average $\int \psi_k(u) du = 0$. One can verify that the wavelet transform is invertible with a stable inverse if the wavelet Fourier transform satisfies the following Littlewood-Paley condition at all frequencies ω :

$$1 - \epsilon \leq |\phi_J(\omega)|^2 + \frac{1}{2} \sum_{j,k} (|\hat{\psi}_{j,k}(\omega)|^2 + |\hat{\psi}_{j,k}(-\omega)|^2) \leq 1. \quad (4)$$

First-order scattering coefficients are obtained by averaging the envelope of the complex wavelet coefficients:

$$\Phi x(u, j, k) = |x \star \psi_{j,k}| \star \phi_J(u). \quad (5)$$

Because of the averaging, these coefficients are locally invariant to translations at a scale 2^J . However, the averaging erases information, and to recover it we must recover the variability of the wavelet transform envelope $|x \star \psi_{j,k}(u)|$ for each j, k .

Second-order scattering coefficients provide information about this variability by filtering the envelope with a new set of wavelets $\psi_{j',k'}$ and averaging the amplitude of the resulting coefficients with ϕ_J :

$$\Phi x(u, j, k, j', k') = ||x \star \psi_{j,k}| \star \psi_{j',k'}| \star \phi_J(u).$$

They measure the average multiscale time variations of $|x \star \psi_{j,k}(u)|$. Note that in general, the scattering transform is not invertible.

If $x(u)$ is a stationary process then $|x \star \psi_{j,k}| \star \phi_J(u)$ and $||x \star \psi_{j,k}| \star \psi_{j',k'}| \star \phi_J(u)$ are also stationary, but also slowly varying in space. They are strongly correlated with correlation coefficients which depend upon the properties of x . In the following we explain how learning these correlation coefficients enables us to estimate the missing spectral information.

3. OUR APPROACH IN A NUTSHELL

As mentioned in the introduction, we work with two types of information about x : measurements y and the correlation structure of our representation Φx . Satisfying the measurement constraints is achieved with a linear projection in the original signal domain, whereas adjusting the correlation structure is achieved with a linear operator in the transformed domain Φ . We assume knowing the first and second moments of $(\Phi x, \Phi \Gamma^+ y)$.

In general, the two criteria will not have the same minimizers, therefore the estimator resulting from the best linear predictor in the scattering domain will not satisfy (1). More precisely, since the involved

distributions are not Gaussian and Φ is not invertible, the best linear estimator (LMMSE)² will in general not satisfy measurements (1) even with $b = 0$. That is, letting \mathbf{R} be the estimator, the set

$$\mathcal{L} \stackrel{\text{def}}{=} \{z : \Phi z = \mathbf{R} \Phi \Gamma^+ y \wedge \Gamma z = y\}$$

could be empty, and in principle it will. To resolve this issue, we define the estimator $\hat{x}(y)$ by the following three relations (strictly speaking, \hat{x} is a set):

$$\|\Gamma \hat{x} - y\| \leq \epsilon, \quad (\text{A1})$$

$$\mathbb{E}[\Phi \hat{x}] = \mathbb{E}[\Phi x], \quad (\text{A2})$$

$$\mathbb{E}[(\Phi \hat{x})_i (\Phi \hat{x} - \Phi x)] = \mathbf{0}, \quad \forall i. \quad (\text{A3})$$

Let us explain these conditions. The first one is a natural requirement that our estimate satisfy the measurement constraint (1) up to the noise ball, where ϵ^2 is proportional to variance of the noise b . For convenience, we denote by \mathcal{A} the set of vectors satisfying the measurement constraint,

$$\mathcal{A} \stackrel{\text{def}}{=} \{u : \|\Gamma u - y\| \leq \epsilon\}.$$

The second and the third condition are related to linearity of estimation. This is not to say that the defined $\Phi \hat{x}$ is linear in $\Phi \Gamma^+ y$ —it is somewhat more subtle: we ask that our Φ -domain estimate $\Phi \hat{x}$ optimally exploits the statistical information about x in the sense that no linear (affine) estimator Q applied to $\Phi \hat{x}$ can extract additional information about Φx that would make $\mathbb{E}\|\Phi x - Q \Phi \hat{x}\|^2$ smaller than $\mathbb{E}\|\Phi x - \Phi \hat{x}\|^2$. This is equivalent to asking that the error vector $\Phi(\hat{x}) - \Phi(x)$ have zero mean and be uncorrelated with every coordinate of the estimate $\Phi \hat{x}$.

The goal is thus to find an estimator \hat{x} which satisfies conditions (A1,A2,A3). In the next section we propose a new algorithm to compute such an estimator.

4. ALGORITHM

4.1. Projection on the Transform-Domain Statistics

As argued above, the LMMSE estimator of Φx given $\Phi \Gamma^+ y$ will in general not be consistent with the measurements. This means that the relations (A1,A2,A3) do not specify the LMMSE estimator. However, we propose to use the LMMSE estimator as a building block of an iterative algorithm that will lead to \hat{x} satisfying the three conditions.

We require our estimate to be uncorrelated with the estimation error in the Φ -domain. Thus, given some Φ -domain data $Z = f(y)$ which we interpret as an intermediate estimate (say $Z = \Phi \Gamma^+ y$), we want to produce a random vector $\hat{X} = \hat{X}(Z)$ such that

$$\mathbb{E}[\hat{X}] = \mathbb{E}[X] \quad (6)$$

$$\text{and } \mathbb{E}[\hat{X}_j (\hat{X} - X)] = 0, \quad \forall j \in \{1, \dots, n\}. \quad (7)$$

Note that there is no typo in having \hat{X} in both terms of the orthogonality relation (7). This is because our goal is not to specify how \hat{X} is related to this particular Z . Rather, we are stating a property of the estimator \hat{X} . Showing how to obtain from Z a vector \hat{X} is then only a mechanism to satisfy this property which we refer to as projection.

Regardless of whether or not Z and X are jointly Gaussian, a random vector satisfying the above conditions can always be found by linear regression over coordinates of Z which optimizes the MSE. In other words, if $\mathbf{G} \in \mathbb{R}^{n \times n}$, $\mathbf{h} \in \mathbb{R}^n$ are (uniquely) defined as

²Although it is common to call this estimator linear, it is in fact affine.

$$\mathbf{G}, \mathbf{h} = \arg \min_{\mathbf{G}, \mathbf{h}} \mathbb{E} \|X - (\hat{\mathbf{G}}Z + \hat{\mathbf{h}})\|^2, \quad (8)$$

then $\mathbf{G}Z + \mathbf{h}$ satisfies (6) and (7). Conveniently, \mathbf{G} and \mathbf{h} only depend on the first and second moments of (X, Z) and we can find them in closed form:

$$\mathbf{G} = \mathbf{K}_{XZ} \mathbf{K}_{ZZ}^{-1}, \quad \mathbf{h} = \mathbb{E}[X] - \mathbf{G} \mathbb{E}[Z]. \quad (9)$$

4.2. Projection on the Measurements

Given $z \in \mathbb{R}^d$ we define a second projector $z' = P_{\mathcal{A}}z$ where $\Phi z'$ is such that $z' \in \mathcal{A}$ and $\mathbb{E} \|\Phi z' - \Phi z\|^2$ is minimized. We can obtain such a projector on a per-realization basis as follows,

$$z' \in P_{\mathcal{A}}z \stackrel{\text{def}}{=} \arg \min_{u \in \mathcal{A}} \|\Phi u - \Phi z\|^2. \quad (10)$$

Since Φ is non-linear this projection is non-convex. We can associate to $P_{\mathcal{A}}$ a projector in the Φ -domain. Given $Z = \Phi z$, we have

$$Z' \in P_{\Phi(\mathcal{A})}Z \stackrel{\text{def}}{=} \left\{ \Phi v : v \in \arg \min_{u \in \mathcal{A}} \|\Phi u - Z\|^2 \right\}. \quad (11)$$

4.3. Algorithm

We now state the alternating projection algorithm. Let $z^{(0)} = 0$. Then for any $k \geq 0$ we define

$$\Phi z^{(k+1)} \in P_{\Phi(\mathcal{A})} R \Phi z^{(k)}.$$

At each iteration, the linear regression operator R applied to $\Phi z^{(k)}$ is computing the closest $Z^{(k+1)}$ such that $\Phi x - Z^{(k+1)}$ is uncorrelated with $\Phi z^{(k)}$. Note that to compute $P_{\Phi(\mathcal{A})}$, we must first compute $P_{\mathcal{A}}$, so the algorithm also maintains an estimate in the signal domain, $z^{(k)}$. The operator $P_{\Phi(\mathcal{A})}$ is a non-linear operator which is approximated with a projected gradient descent algorithm. The alternating projection can be explicitly written as

$$\Phi z^{(k+1)} \in P_{\Phi(\mathcal{A})} \left[\mathbf{G}^{(k)} \Phi z^{(k)} + \mathbf{h}^{(k)} \right].$$

The algorithm is stopped after r iterations and the associated (non-unique) $z^{(k)}$ is taken as an approximation of \hat{x} . It is important to point out that since in every iteration we compute a new estimator, we need not only the first- and second-order statistics of $(\Phi x, \Phi \Gamma^+ y)$, but also of $(\Phi x, \Phi z^{(k)})$ for all iterations k .

5. NUMERICAL RESULTS

5.1. Figures of Merit

The proposed algorithm iteratively reduces the MSE in the Φ domain. However, this does not translate into improved PSNR in the pixel domain. In fact, in super-resolution, a simple low-pass projection will generally yield better PSNR than our reconstructions. One of the reasons for this is that we reconstruct inexact phase at high frequencies. This results in a slight misalignment of singularities such as points and edges, which in turn leads to large ℓ^2 errors. Lowpassing hedges its bets by blurring singularities thus reducing the ℓ^2 error, although the final result clearly *looks* wrong.

There is a need for a different translation-invariant measure which goes beyond Gaussian statistics, and which is independent of our specific Φ . To this end, we propose to use higher-order moments, and in particular kurtosis.

Kurtosis of a random variable is the standardized fourth central moment. For a random vector $\mathbf{x} = [x_1, \dots, x_N]$, one may define a multivariate (co)kurtosis tensor by analogy to the covariance matrix:

$$\text{kurt}[x]_{ijkl} = \frac{\mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l)]}{\sqrt{\mathbb{V}[x_i]\mathbb{V}[x_j]\mathbb{V}[x_k]\mathbb{V}[x_l]}}. \quad (12)$$

Note that for shift-invariant processes we can fix one index.

To circumvent the challenges of estimating and comparing large tensors, we further use a result by Mardia [7] who showed that $\beta \stackrel{\text{def}}{=} \mathbb{E} \left\{ [(x - \mu)^\top \Sigma^{-1} (x - \mu)]^2 \right\}$ is a good measure of multivariate kurtosis. Given a random sample x_1, \dots, x_n , we can estimate β by the empirical mean using estimated covariance $\hat{\Sigma}$. Mardia demonstrated that for $x \in \mathbb{R}^p$ we have $\mathbb{E}[\hat{\beta}] = p(p+2) \frac{n-1}{n+1}$.

Non-normality is an important aspect of natural signals. For an illustration in texture classification see [3]. We will use the excess sample kurtosis $\hat{\beta} - p(p+2)$ as a measure of how well the statistics of the desired signals are reproduced by the various methods.

5.2. Operators and Processes

We study two different problems: super-resolution and limited-angle tomography (Radon transform). Their action in Fourier space is illustrated in Fig. 1. In super-resolution, the operator is given as a decimator—a composition of lowpass filtering H_{LP} and downsampling S , $\Gamma_{\text{SR}} = S \circ H_{\text{LP}}$, and the corresponding projection is given as (assuming y are measurements and H_{LP} is self-adjoint)

$$P_{\text{SR}}z = H_{\text{LP}} \circ S^* y + (I - H_{\text{LP}})z, \quad (13)$$

where S^* is upsampling.

In tomography, the operator is given as the Radon transform $[\Gamma_{\text{R}}x]_{\ell} = \int_{\ell} x(s) ds$, for a set of lines $\ell \in \Lambda$. Unlike in the case of super-resolution, here we do not use a linear right inverse. A better result can be obtained if we impose positivity and max norm:

$$\Gamma_{\text{R}}^+ y \stackrel{\text{def}}{=} \arg \min_{z: 0 \leq z(u) \leq 1} \|y - \Gamma_{\text{R}}z\|_2^2, \quad (14)$$

Note that this is a right inverse only on signals in the constraint space.

We use two types of signals: realizations of the Ising spin-glass model [8] and Cox point processes [9]. At the right temperature, Ising realizations behave like a mixture of simple shapes and textures while Cox processes are useful for comparisons with sparse regularizations.

5.3. Comparison With ℓ^1 Minimization: Cox Point Processes

We first attempt to demonstrate why convex methods that work in the pixel domain and do not use a phase-removal mechanism fail in some simple cases. Fig. 2C shows a reconstruction of a Cox point process. Measurements are obtained by Gaussian filtering and downsampling by a factor of 4 along each axis. Signal sparsity is set so that combined with the relatively high coherence of the forward operator the ℓ^1 minimization does not have a unique solution [10].

Convexity of the cost function $E(u) \stackrel{\text{def}}{=} d(y, \Gamma u) + \lambda h(u)$ can be a bottleneck in severely ill-posed problems as it makes the estimate vulnerable to regression to the mean. If u_1 and u_2 are two minimizers with $E(u_1) = E(u_2)$, then $\tilde{u} = t u_1 + (1-t) u_2$ is also a minimizer of E . But if u_2 is a small translation or deformation of u_1 , \tilde{u} will be a blurred version of u_1 with corrupted high-frequency information. This can be seen clearly for $h(\cdot) = \|\cdot\|_{\ell^1}$ in Fig. 2C.

The reconstruction using one iteration of the proposed algorithm looks better. Even though it is not equal to the original realization, it has the correct spatial statistics (it *looks* right), and it reproduces

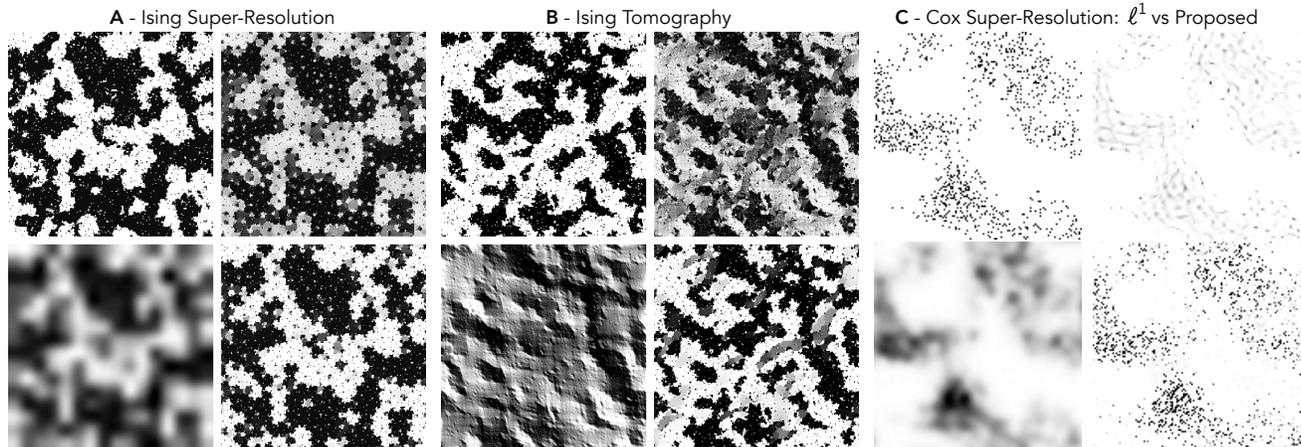


Fig. 2. Results of computer experiments. **A:** upper-left—the original Ising realization of size 256×256 , at temperature $T = 0.3$; lower-left—lowpass version (projection of the original on \mathcal{A}) obtained by decimating by a factor of 16 along each axis; upper-right—reconstruction after the first iteration; lower-right—reconstruction after 20 iterations, projected on \mathcal{A} . **B:** same as **A** but with half-angle Radon measurements. **C:** upper-left—realization of a Cox random process (non-homogeneous Poisson); lower-left—lowpass version; upper-right—reconstruction by ℓ^1 minimization; lower-right—one iteration of the proposed algorithm.

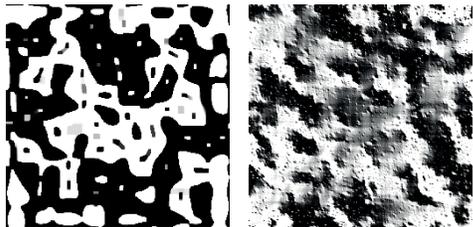


Fig. 3. Reconstructions of Ising measurements from Fig. 2 by TV-norm minimization. Left: super-resolution; right: tomography.

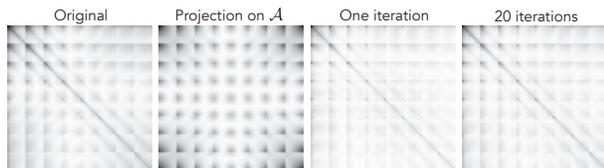


Fig. 4. Translation invariant cokurtosis tensors of size 64×64 computed on patches of size 8×8 and averaged along the first dimension (tomography).

measurements. It is thus a valid solution within the proposed framework. Indeed, Table 1 shows that although the ℓ^1 -minimal solution fares better in terms of the MSE, in terms of higher-order moments it performs much worse than the scattering-based reconstruction.

5.4. Ising Super-Resolution

Fig. 2A shows super-resolution of Ising realizations. The operator is a decimation by a factor of 16 along each axis, thus the data loss is 256-fold. The following can be noted: using a linear reconstruction (lower-left) gives a bad result. Already the first iteration of the proposed algorithm gives a much better result, and iterating further brings out correct structural details as indicated in the lower right. In Table 1 we see that the excess kurtosis corroborates our ocular observations, although the MSE might suggest otherwise. In the light of the severity of data loss we find these results remarkable.

For a comparison, in Fig. 3 we provide a reconstruction regularized by the total variation (TV): $\arg \min_{z: 0 \leq z(u) \leq 1} \|y - \Gamma_{\text{SR}} z\|_2^2 + \lambda \|z\|_{\text{TV}}$. The value $\lambda = 5 \times 10^{-6}$ was tuned by hand.

Experiment	MSE	Ex. Kurtosis
Ising Original	0	1760
Ising SR Projection	0.91e4	85
Ising SR Iteration 1	1.32e4	1282
Ising SR Final	1.35e4	1290
Ising Radon Projection	6.91e3	2285
Ising Radon Iteration 1	6.55e3	1705
Ising Radon Final	7.54e3	1787
Cox Original	0	605
Cox SR Projection	1814	267
Cox ℓ^1	1500	3164
Cox One Iteration	1910	514

Table 1. Figures of merit computed on various sets of images.

5.5. Ising Tomography

In Fig. 2B we present reconstructions of Ising models from tomographic (Radon) measurements. Γ_{R} performs a half angle Radon transform with angles between 0° and 89° in steps of 1° . We see that a simple reconstruction subject to the positivity constraint (14) does not give a very good result. First iteration of our algorithm gives a nicer image, and a considerable improvement comes from iterating.

The excess kurtosis given in Table 1 corroborates what our eyes are telling us. We also show the cokurtosis tensors in Fig. 4. It is clear that the final result best matches the cokurtosis of the original signal.

Finally, a TV-norm regularized reconstruction is shown in Fig. 3 ($\lambda = 1 \times 10^{-6}$). Since in the tomographic experiments the data loss is less severe, this reconstruction looks better than in super-resolution.

6. CONCLUSION

We proposed a new algorithm for solving linear inverse problems which exploits the correlation structure of the data in a non-linearly transformed domain. The modulus non-linearity reveals spectral correlations otherwise hidden by phase fluctuations. Promising initial results on super-resolution and tomography show that we indeed recover the correct missing spectral information in hard problems. Future work includes optimizing the transform, more general classes of operators, efficient implementations, and proofs of convergence.

7. 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] 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.
- [4] S. Mallat, “Group Invariant Scattering,” *Commun. Pur. Appl. Math.*, vol. 65, no. 10, pp. 1331–1398, Oct. 2012.
- [5] J. Bruna, P. Sprechmann, and Y. LeCun, “Super-Resolution with Deep Convolutional Sufficient Statistics,” Nov. 2015.
- [6] K. Pearson, “Contributions to the Mathematical Theory of Evolution. III. Regression, Heredity, and Panmixia.,” *Proc. R. Soc. Lond.*, vol. 59, no. 353-358, pp. 69–71, Jan. 1895.
- [7] K. V. Mardia, “Measures of multivariate skewness and kurtosis with applications,” *Biometrika*, vol. 57, no. 3, pp. 519–530, Dec. 1970.
- [8] 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.
- [9] D. R. Cox and V. Isham, *Point Processes*, CRC Press, July 1980.
- [10] D. L. Donoho, “Compressed Sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.