Super-resolution via superset selection and pruning

Laurent Demanet
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: laurent@math.mit.edu

Deanna Needell
Department of Mathematics
Claremont McKenna College
Claremont, CA 91711
Email: dneedell@cmc.edu

Nam Nguyen
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
Email: namnguyen@math.mit.edu

***Abstract*—We present a pursuit-like algorithm that we call the "superset method" for recovery of sparse vectors from consecutive Fourier measurements in the super-resolution regime. The algorithm has a subspace identification step that hinges on the translation invariance of the Fourier transform, followed by a removal step to estimate the solution's support. The superset method is always successful in the noiseless regime (unlike ℓ_1 minimization) and generalizes to higher dimensions (unlike the matrix pencil method). Relative robustness to noise is demonstrated numerically.**

Acknowledgments. LD acknowledges funding from the Air Force Office of Scientific Research, the National Science Foundation, and the Alfred P. Sloan Foundation. LD is grateful to Jean-Francois Mercier and George Papanicolaou for early discussions on super-resolution.

I. Introduction

We consider the problem of recovering a sparse vector $x_0 \in \mathbb{R}^n$, or an approximation thereof, from $m \leq n$ contiguous Fourier measurements

$$y = Ax_0 + e, \tag{1}$$

where A is the partial, short and wide Fourier matrix $A_{jk} = e^{2\pi ijk/n}$, $0 \leq j < m$, $-n/2 \leq k < n/2$, n even, and, say, $e \sim N(0, \sigma^2 I_m)$.

When recovery is successful in this scenario of contiguous measurements, we may speak of super-resolution: the spacing between neighboring nonzero components in x_0 can be much smaller than the Rayleigh limit n/m suggested by Shannon-Nyquist theory. But in contrast to the compressed sensing scenario, where the m values of j are drawn at random from $\{0, \ldots, n-1\}$, super-resolution can be arbitrarily ill-posed. Open questions concern not only recovery bounds, but the very algorithms needed to define good estimators.

Various techniques have been proposed in the literature to tackle super-resolution, such as MUSIC [11], Prony's method / finite rate of innovation [8] [1] [12], the matrix pencil method [9], ℓ_1 minimization [7] [5] [3] [2], and greedy pursuits [6].

Prony and matrix pencil methods are based on eigenvalue computations: they work well with exact measurements, but their performance is poorly understood in the presence of noise, and they are not obviously set up in higher dimensions. As for ℓ_1 minimization, there is good evidence that k-sparse *nonnegative* signals can be recovered from only $2k+1$ noiseless Fourier coefficients by imposing the positivity constraint with or without ℓ_1 minimization, see [4] [7] and [5]. The work of [3] extends this result to the continuous setting by using total variation minimization. Recently, Candès and Fernandez-Granda showed that the solution to an ℓ_1-minimization problem with a $\|A^*(y - Ax)\|_1$ misfit will be close to the true signal, assuming that locations of any two consecutive nonzero coefficients are separated by at least four times the super-resolution factor n/m [2]. Such optimization ideas have the advantage of being easily generalizable to higher dimensions. On the flip side, ℓ_1 minimization super-resolution is known to fail on sparse signals with nearby components that alternate signs.

In this paper, we discuss a simple algorithm for solving (1) based on

- subspace identification as in the matrix pencil method, but without the subsequent eigenvalue computation; and
- a removal procedure for tightening the active set, remindful of a step in certain greedy pursuits.

This algorithm can outperform the well-known matrix pencil method, as we show in the numerical section, and it is generalizable to higher dimensions. It is a one-pass procedure that does not suffer from slow convergence in situations of high coherence. We also show that the algorithm provides perfect recovery for the (not combinatorially hard in the Fourier case) noiseless ℓ_0

problem

$$\min_x |\operatorname{supp} x| \qquad \text{s.t.} \qquad Ax = y. \tag{2}$$

II. Noiseless subspace identification

For completeness we start by recalling the classical uniqueness result for (2).

Lemma 1. *Let $x_0 \in \mathbb{R}^n$ with support T such that $m \geq 2|T|$, and let $y = Ax_0$. Then the unique minimizer of (2) is x_0.*

We make use of the following notations. Denote $\operatorname{supp} x_0$ by T, and write A_T for the restriction of A to its columns in T. Let T^c for the complement of T. Let a_k for the k-th column of A. The superscript L is used to denote a restriction of a matrix to its first L rows, as in A_T^L.

The "superset method" hinges on a special property that the partial Fourier matrix A does not share with arbitrary dictionaries: each column a_k is *translation-invariant* in the sense that any restriction of a_k to $s \leq m$ consecutive elements gives rise to the same sequence, up to an overall scalar. In other words, exponentials are eigenfunctions of the translation operator. This structure is important. There is an opportunity cost in ignoring it and treating (1) as a generic compressed sensing problem.

A way to leverage translation invariance is to recognize that it gives access to the *subspace* spanned by the atoms a_k for $k \in T$, such that $y = \sum_{k\in T}(x_0)_k a_k$. Algorithmically, one picks a number $1 < L < m$ and juxtaposes translated copies of (restrictions of) y into the Hankel matrix $Y = \text{Hankel}(y)$, defined as

$$Y = \begin{pmatrix} y_0 & y_1 & \cdots & y_{m-L-1} \\ y_1 & y_2 & \cdots & y_{m-L} \\ \vdots & \vdots & \vdots & \vdots \\ y_{L-1} & y_L & \cdots & y_m \end{pmatrix}.$$

The range of Y is the subspace we seek.

Lemma 2. *If $L \geq |T|$, then the rank of Y is $|T|$, and*

$$\textit{Ran}\, Y = \textit{Ran}\, A_T^L.$$

The lemma suggests a simple recovery procedure in the noiseless case: loop over all the candidate atoms a_k for $-n/2 \leq k < n/2$ and select those for which the angle

$$\angle(a_k^L, \operatorname{Ran} Y) = 0. \tag{3}$$

Once the set T is identified, the solution is obtained by solving the determined system

$$A_T x_T = y, \qquad x_{T^c} = 0. \tag{4}$$

This procedure (unsurprisingly) provides a solution to the noise-free ℓ_0 sparse recovery problem (2).

Theorem 3. *Let $x_0 \in \mathbb{R}^n$ with support T such that $m > 2|T|$, and let $y = Ax_0$. Consider x defined by (3) and (4), where the Hankel matrix Y is built with $|T| + 1 \leq L \leq m - |T| - 1$. Then $x = x_0$.*

The proofs of lemma 2 and theorem 3 hinge on the fact that A has full spark.

The idea of subspace identification is at the heart of a different method, the matrix pencil, which seeks the rank-reducing numbers z of the pencil

$$\overline{Y} - z\underline{Y},$$

where $\overline{Y}$ is Y with its first row removed, and $\underline{Y}$ is Y with its last row removed. These numbers z are computed as the generalized eigenvalues of the couple $(\underline{Y}^*\overline{Y}, \underline{Y}^*\underline{Y})$. z can also be found via solving the eigenvalues of the matrix $\underline{Y}^\dagger\overline{Y}$. When $|T| \leq L \leq m - |T|$, the collection of these generalized eigenvalues includes $e^{2\pi ijk/n}$ for $k \in T$, as well as $m - L - |T|$ zeros. There exist variants that consider a Toeplitz matrix instead of a Hankel matrix, with slightly better numerical stability properties. When $L = |T|$, the matrix pencil method reduces to Prony's method, a numerically inferior choice that should be avoided in practice if possible.

III. Noisy subspace identification

The problem becomes more difficult when the observations are contaminated by noise. In this situation $\operatorname{Ran} A_T^L \neq \operatorname{Ran} Y$, though in low-noise situations we may still be able to recover T from the indices of the smallest angles $\angle(a_k^L, \operatorname{Ran} Y)$.

Proposition 4. *Let $y = y_0 + e$ with $e \sim N(0, \sigma^2 I_m)$, and form the corresponding $L \times (m-L)$ matrices Y and Y_0 as previously. Denote the singular values of Y_0 by $s_{n,0}$. Let $r = \textit{rank}(Y_0)$. Then there exists positive c_1, C_1 and c, such that with probability at least $1 - c_1 m^{-C_1}$,*

$$|\sin\angle(a_k^L, \textit{Ran}\, Y_0) - \sin\angle(a_k^L, \textit{Ran}\, Y)| \leq c\,\varepsilon_1, \tag{5}$$

where

$$\varepsilon_1^2 = \sigma \frac{\sqrt{rm\log m}}{s_{1,0} - s_{2,0}}. \tag{6}$$

In order to write bounds involving the singular values s_i of Y instead of those of Y_0, one may use Weyl's inequality $|s_i - s_{i,0}| \leq \|\text{Hankel}(e)\|$, which can in turn be controlled as $O(\sigma\sqrt{m\log m})$ with high probability.

The subspace identification step now gathers all the values of k compatible with the (null) hypothesis that

$\angle(a_k^L, \operatorname{Ran} Y_0) = 0$, i.e., those which for some adequate $c > 0$ obey

$$\sin \angle(a_k^L, \operatorname{Ran} Y) \le c\,\varepsilon_1.$$

The resulting set Ω of indices is only expected to be a *superset* of the true support T, with high probability.

A second step is now needed to prune Ω in order to extract T. For this purpose, a loop over k is set up where we test the membership of y in $\operatorname{Ran} A_{\Omega\setminus k}$, the range of A_Ω with the k-th column removed. We are now considering a new set of angles where the roles of y and A are reversed: in a noiseless situation, $k \in T$ if and only if

$$\angle(y, \operatorname{Ran} A_{\Omega\setminus k}) \neq 0.$$

When noise is present, we first filter out the noise off Ω by projecting y onto the range of A_Ω, then estimate $k \in T$ only when the angle is above a certain threshold. It is easier to work directly with projections Π:

$$\|\Pi_\Omega y - \Pi_{\Omega\setminus k} y\| = \sin \angle(\Pi_\Omega y, \operatorname{Ran} A_{\Omega\setminus k})\ \|\Pi_\Omega y\|.$$

The effect of noise on the left-hand side is as follows.

Proposition 5. *Let $y = y_0 + e$ with $e \sim N(0, \sigma^2 I_m)$. Let $\Pi_\Omega y$ be the projection of y onto Ran A_Ω, and let $\Delta\Pi = \Pi_\Omega - \Pi_{\Omega\setminus k}$. Then there exists $c > 0$ such that, with high probability,*

$$|\,\|\Delta\Pi y\| - \|\Delta\Pi y_0\|\,| \le c\,\varepsilon_2,$$

with $\varepsilon_2 = \sigma$.

Algorithm 1 for the superset method implements the removal step in an iterative fashion, one atom at a time.

IV. Experimental Results

In the first simulation, we fix $n = 1000$ and $m = 120$ and construct an n-dimensional signal x_0 whose nonzero components are well separated by at least $4n/m$, a distance equivalent to four times the super-resolution factor n/m. The spike magnitudes are independently set to $\pm 1/\sqrt{29}$ with probability $1/2$. The noise vector e is drawn from $N(0, \sigma^2 I_m)$ with $\sigma = 10^{-3}$. We fix the thresholds ε_1 via (6) with $c = 1$ and $\varepsilon_2 = 10\sigma$. Throughout our simulations, we set $L = \lfloor m/3 \rfloor$. As can be seen from Fig. 1, top row, the recovered signal from the superset method is reasonable, with $\|\widehat{x} - x_0\|_2 = 0.075$, while the reconstruction via ℓ_1-minimization tends to exhibit incorrect clusters around the true spikes.

Our next simulation considers a more challenging signal model with a strongly coherent matrix A. For example, with $n = 1000$ and $m = 120$, the coherence of the matrix A with normalized columns a_i is $\mu = \max_{i\neq j} |\langle a_i, a_j \rangle| = 0.9765$. The signal in this

Algorithm 1 Superset selection and pruning

input: Partial Fourier matrix $A \in C^{m\times n}$, $y = Ax_0 + e$, parameter L, thresholds ε_1 and ε_2.
initialization: $Y = \text{Hankel}(y) \in \mathbb{C}^{L\times(m-L)}$
support identification
  decompose: $\widetilde{Q}\widetilde{R} = Y\widetilde{E}$, $\widetilde{Q} \in \mathbb{C}^{L\times r}$
  project: $a_k \leftarrow A_{\{k\}}$ (for all k)
  $\gamma_k \leftarrow \left\| a_k - \widetilde{Q}\widetilde{Q}^* a_k \right\| / \|a_k\|$
  $\Omega = \{k : \gamma_k \le \varepsilon_1\}$
while true **do**
  decompose: $QR = A_\Omega E$, $Q \in \mathbb{C}^{m\times|\Omega|}$
  remove: $\forall k \in \Omega$: $Q_{(k)}R_{(k)} = A_{\Omega\setminus k}E_{(k)}$
  $\delta_k \leftarrow \|(Q_{(k)}Q_{(k)}^* - QQ^*)y\|_2$
  $k_0 \leftarrow \operatorname{argmin}_k \delta_k$
  if $\delta_{k_0} < \varepsilon_2$, $\Omega \leftarrow \Omega\setminus k_0$
  else break
end while
output: $\widehat{x} = \operatorname{argmin}_x \|y - A_\Omega x\|$

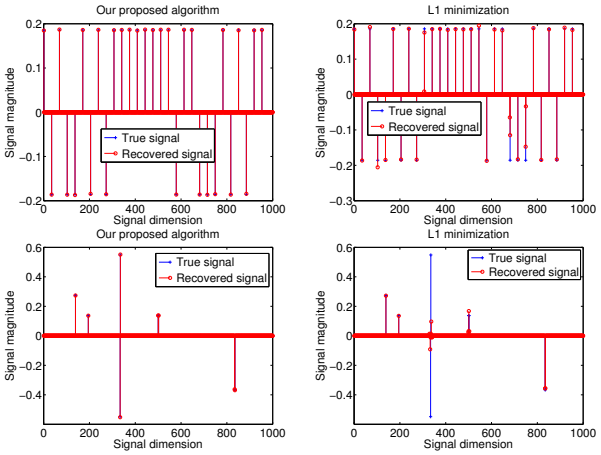

Fig. 1. Original (blue) and recovered (red) signals. Left column: the superset method. Right column: ℓ_1-minimization. Top row: a signal with well-separated spikes. Bottom row: spike spacing below the Rayleigh length.

simulation is shown in Fig. 1, bottom row. It consists of five spike clusters: each of the first two clusters consists of a single spike, and each of the last four clusters contains two neighboring spikes. The signs of these neighboring spikes either agree or differ. We set m, σ and ε_2 as in the previous simulation, and we let the constant c in the equation (6) of ε_1 equal to 5. Recovery via the superset method is accurate, while ℓ_1 minimization fails at least with clusters of opposite-sign spikes.

In the next simulation, we consider a signal of size $n = 1000$ which contains two nearby spikes at locations $[100, 101]$ and has magnitudes $1/\sqrt{2}$ and

$-1/\sqrt{2}$. We empirically investigate the algorithm's ability to recover the signal from varying measurements $m = \{10, 20, ..., 220\}$ and noise levels $log_{10}\sigma = \{-3.5, -3.4, ..., -2\}$. For each pair (m, σ), we report the frequency of success over 100 random realizations of e. The greyscale goes from white (100 successes) to black (100 failures). A trial is declared successful if the recovered $\widehat{x}$ satisfies $\|\widehat{x} - x_0\|_2 / \|x_0\|_2 < 10^{-3}$. The horizontal axis indicates the noise level σ in log scale, and the vertical axis indicates $\log_{10}(1-\mu)$ where μ is the coherence as earlier.

We note that the coherence is inversely proportional to the amount of measurements m and proportional to the super-resolution factor n/m: increasing m (decreasing the super-resolution factor) will reduce the coherence μ. On the vertical axis, smaller values imply higher coherence, or equivalently smaller amount of measurements. As shown in Fig. 2, for reasonably small noise, the algorithm is able to recover the signal exactly even the coherence is nearly 1.

For reference, we also compare the superset method with the matrix pencil method as set up in [10]. The noise is filtered out by preparing low-rank approximations of $\underline{Y}$ and $\overline{Y}$ where only the singular values above $c\sigma\sqrt{L \log L}$ are kept, for some heuristically optimized constant c. Two more signals are considered: (1) a 3-spase signal consisting of three neighboring spikes, each of magnitude $1/\sqrt{3}$ with alternating signs, and (2) a 4-sparse signal with neighboring spikes of alternating signs and equal magnitude $1/2$. Fig. 2 is a good illustration of the contrasting numerical behaviors of the two methods: the matrix pencil is often the better method in the special case of a signal with 2 spikes, but loses ground to the superset method in various cases of progressively less sparse signals. Understanding the performance of the matrix pencil would require formulating a lower bound on the (typically extremely small) S-th eigenvalues of Y_0 where S is the sparsity of y_0.

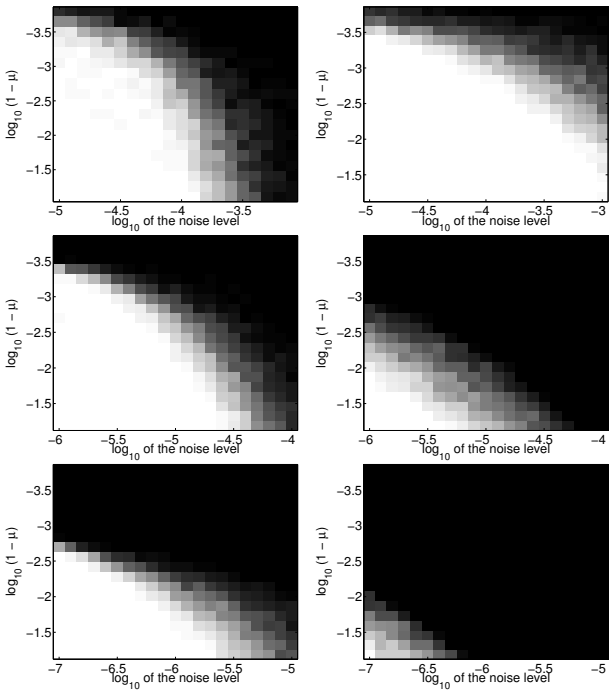

Fig. 2. Probability of recovery, from 1 (white) to 0 (black) for the superset method (left column) and the matrix pencil method (right column). Top row: 2-sparse signal. Middle row: 3-sparse signal. Bottom row: 4-sparse signal. The plots show recovery as a function of the noise level (x-axis, $\log_{10}\sigma$) and the coherence (y-axis, $\log_{10}(1-\mu)$).

V. Conclusion

Empirical evidence is presented for the potential of the superset method as a viable computational method for super-resolution. Further theoretical justifications will be presented elsewhere.

References

[1] V.M. Adamjan, D.Z. Arov, and MG Krein. Analytic properties of schmidt pairs for a hankel operator and the generalized schur-takagi problem. *Sb. Math.*, 15(1):31–73, 1971.

[2] E. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Commun. Pure Appl. Math.* To appear.

[3] Y.D. Castro and F. Gamboa. Exact reconstruction using beurling minimal extrapolation. *Preprint*, 2012.

[4] D.L. Donoho, I.M. Johnstone, J.C. Hoch, and A.S. Stern. Maximum entropy and the nearly black object. *J. Roy. Stat. Soc. B Met.*, pages 41–81, 1992.

[5] D.L. Donoho and J. Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. In *Proc. Nation. Acad. Scien.*, page 94469451, 2005.

[6] A. Fannjiang and W. Liao. Coherence-pattern guided compressive sensing with unresolved grids. *IAM J. Imaging Sci.*, 5:179–202, 2012.

[7] J.J. Fuchs. Sparsity and uniqueness for some specific underdetermined linear systems. In *Proc. of IEEE ICASSP*, page 729732, Philadelphia, PA, USA, 2005. IEEE.

[8] U. Grenander and G. Szegő. *Toeplitz forms and their applications*. U. California Press, Berkeley, 1958.

[9] Y. Hua and T.K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. 38(5):814–824, 1990.

[10] Y. Hua and T.K. Sarkar. On svd for estimating generalized eigenvalues of singular matrix pencil in noise. *IEEE T. Signal Proces.*, 39(4):892–900, 1991.

[11] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. Atten. Prop.*, 34(3):276–280, Apr. 1986.

[12] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE T. Signal Proces.*, 50(6):1417–1428, 2002.