Atomic Decomposition of Mixtures of Translation-Invariant Signals

Gongguo Tang
University of Wisconsin-Madison
Email: gtang5@wisc.edu

Benjamin Recht
University of Wisconsin-Madison
Email: brecht@cs.wisc.edu

***Abstract*—This paper develops a theory for the atomic decomposition of mixtures of translation-invariant signals. Suppose we have a linear combination of shifted copies of a known waveform with unknown shifts and coefficients. These shifts assume continuous values on the real line. We show that one can recover the exact translations and amplitudes by solving a continuous analog of ℓ_1 minimization, provided the translations are well separated. The minimal separation depends on properties of the auto-correlation function of the known waveform. This work lays a foundation for the study of denoising and reconstruction of translation-invariant signals from linear measurements, and finds applications in microscopy imaging, radar, and communication.**

I. Introduction

The theory for decomposition, reconstruction, denoising, and demixing of signals having sparse representations with respect to finite dictionaries finds numerous applications in signal processing and is now considered well-studied. In many applications of practical interest, however, finite dictionaries are insufficient to sparsely represent signals. For example, audios and images typically contain features that could appear at any time, location, and scale; wavelet transform employs atoms of continuously translated and scaled copied of mother functions. Radar and seismic signals are modeled as sparse combinations of shifted and frequency-modulated template waveforms. A distinctive feature of these signals is that they can only be sparsely represented as mixtures of elements from continuously parameterized dictionaries.

Despite the ubiquity of continuous dictionaries, little is known in theory about their powers and limitations in sparse signal processing, especially when it comes to convex relaxation approaches. A common strategy to handle continuous parameterization in practice is discretization [1], [2], [3], [4], [5]. Leaving potential losses in this process aside, even for the discretized system, existing theories can not adequately explain the successes and failures of ℓ_1 minimization due to high coherence of the resulting dictionaries. In this paper, we instead focus on studying the continuous dictionaries directly.

More precisely, we consider the following special mixture of translated waveforms:

$$x(t) = \sum_j c_j \psi(t-\tau_j), t \in \mathbb{R}, \tag{1}$$

with $\psi(t)$ a known waveform, and $\tau_j \in \mathbb{R}$ and $c_j \in \mathbb{R}$ unknown translation parameter and coefficient. Equation (1) is a proper model for signals in radar, LIDAR, ultrasound imaging, and communication channel estimation when the Doppler effect is negligible. Another way to look at the signal (1) is to view it as the convolution of $\psi(t)$ with a mixture of point sources $\sum_j c_j \delta(t-\tau_j)$, where $\delta(t)$ is the Dirac function. In this case, the function $\psi(t)$ can be the point spread function in single-molecule imaging [3] and astronomical imaging [6].

Several fundamental questions about the signal (1) are of interest:

1) Atomic Decomposition: under what conditions the decomposition in (1) is most economical, i.e., has the minimal $\sum_j |c_j|$? Given a set of elementary functions $\mathcal{A} = \{a_j\}$ and a general function x, the problem of finding a decomposition $x = \sum_j c_j a_j$ with the least $\sum_j |c_j|$ is called atomic decomposition [7]. From this point of view, we are looking for conditions to guarantee (1) is the unique atomic decomposition of $x(t)$ corresponding to the atomic set $\{\psi(t-\tau) : \tau \in \mathbb{R}\}$.
2) Reconstruction: when can one recover $x(t)$ given its linear measurements? A particular interesting linear measurement operator is subsampling, in which one observes samples of $x(t)$.
3) Denoising: if $x(t)$ or its linear measurements are contaminated by noise, how well can one perform denoising using, e.g., a generalized LASSO formulation?

Among these questions, atomic decomposition is the most fundamental as it lays theoretical foundations that solutions of other problems can build upon. For finite dictionaries, the study of the noise-free basis pursuit algorithm and the RIP-type conditions serves as the basis for noisy analysis; for matrix completion, the fact that nuclear norm is achieved by singular value decomposition is essential in its analysis [8]. Candès and Fernandez-Granda's result [9] on decomposition of line spectral signals form the foundations for later developments in subsampling and noise analysis [10], [11], [12]. Following a similar roadmap, this paper addresses the atomic decomposition of translation-invariant mixtures.

The paper is organized as follows. In Section II we set up the signal model and state the main problem more precisely. In Section III, we present the main theorem and outline its proof. Section IV is devoted to numerical experiments and Section V concludes the paper.

II. Signal Model and Problem Setup

In the signal model (1), we assume that the waveform $\psi(\cdot)$ is smooth and has unit energy, i.e., $\int_{\mathbb{R}} |\psi(t)|^2 dt = 1$. Examples of $\psi(\cdot)$ include

- Gaussian function: $\psi(t) = \frac{1}{\pi^{1/4}\sqrt{\sigma}} \exp\left(-\frac{t^2}{2\sigma^2}\right), t \in \mathbb{R}$
- Cauchy function: $\psi(t) = \sqrt{\frac{2}{\pi\sigma}} \frac{1}{1+(t/\sigma)^2}, t \in \mathbb{R}$

- Sinc function: $\psi(t) = \sqrt{\frac{1}{\pi\sigma}}\frac{\sin(t/\sigma)}{t/\sigma}, t \in \mathbb{R}$.
- Airy function: $\psi(t) = C(\sigma)\left(\frac{J_1(t/\sigma)}{t/2}\right)^2, t \in \mathbb{R}$

where $J_1(\cdot)$ is the Bessel functions of the first kind and $C(\sigma)$ is a constant to ensure unit energy.

Let $C_0(\mathbb{R})$ be the space of continuous functions on $\mathbb{R}$ that vanish at infinity. The smoothness and unit energy assumption guarantee that both $\psi(t)$ and $x(t)$ are in $C_0(\mathbb{R})$. For any $x(t) \in C_0(\mathbb{R})$, we define its atomic norm [7] corresponding to $\mathcal{A} = \{\pm\psi(t-\tau) \in C_0(\mathbb{R}) : \tau \in \mathbb{R}\}$ as

$$\|x\|_{\mathcal{A}} := \inf\Big\{\sum_j |c_j| : x(t) = \sum_j c_j\psi(t-\tau_j)\Big\}. \quad (2)$$

The definition of the atomic norm involves an infinite analog of ℓ_1 minimization. A decomposition that achieves the atomic norm of $x(t)$ is called an atomic decomposition.

An alternative way to write model (1) is the following:

$$x(s) = \int_{\mathbb{R}} \psi(s-t)\mu(dt), s \in \mathbb{R}, \quad (3)$$

where

$$\mu = \sum_j c_j\delta(t-\tau_j) \quad (4)$$

with $\delta(t)$ the Dirac measure. Here the support $\{\tau_j\}$ of the Borel measure μ does not have to be finite, as long as $\sum_j |c_j| < \infty$.

Denote by $\mathcal{B}(\mathbb{R})$ the set of regular Borel measures over $\mathbb{R}$ with bounded total variation. Our goal in this paper is to answer the question: among all possible decompositions of the form (3) for $\mu \in \mathcal{B}(\mathbb{R})$, when the measure in (4) represents the unique atomic decomposition of $x(t)$?

III. Main Theorem

In this section, we present the main theorem and outline its proof. Due to space limit, we omit proofs for all intermediate lemmas.

We heuristically argue conditions the signal should satisfy to guarantee unique atomic decomposition. First of all, if two translations $\tau_1 \approx \tau_2$, then due to continuity, $\psi(t-\tau_1)$ and $\psi(t-\tau_2)$ are very similar to each other, making the decomposition not economical. Therefore, the conditions involve the separation among translations:

$$\Delta := \min_{\tau_i \neq \tau_j} |\tau_i - \tau_j| \quad (5)$$

Suppose now the minimal separation Δ is large. If $\psi(t-\tau_1)$ has high correlation with another shifted copy, say, $\psi(t-\tau_2)$, then the decomposition is still not economical. Our conditions also involve the auto-correlation function of of $\psi(\cdot)$:

$$K(t) = \int_{\mathbb{R}} \psi(s-t)\psi(s)ds\,. \quad (6)$$

Denote the derivatives of $K(t)$ as $K_n(t) := \frac{d^n}{dt^n}K(t)$ and $\kappa_{2n} := |K_{2n}(0)| = \int_{\mathbb{R}} |\psi^{(n)}(s)|^2 ds$.

We use some bounding functions to control the decay speeds of the auto-correlation function and its derivatives:

Assumption 1: For $\ell = 0, 1, 2, 3$, there exist a constant ξ_0 and upper bound functions $\{B_\ell(t), |t| \geq \xi_0\}$ such that

- $|K_\ell(t)| \leq B_\ell(t), |t| \geq \xi_0$,
- $B_\ell(t)$ are symmetric, nonincreasing, and integrable over $[\xi_0, \infty)$,
- $B_\ell(\xi - t) + B_\ell(\xi + t)$ is non-decreasing for $t \in \left[0, \frac{\xi}{2}\right]$ for any $\xi > 2\xi_0$.

For smooth $B_\ell(t)$, the last part of the assumption is true if $B_\ell''(t) \geq 0$ for $t \geq \xi_0$.

The decay speed of $K_\ell(t)$ is summarized by the quantity

$$\bar{B}_\ell := \sup_{\Delta > 2\xi_0}\left[\Delta B_\ell\left(\frac{\Delta}{2}\right) + 2\int_{\frac{\Delta}{2}}^{\infty} B_\ell(t)\,dt\right]. \quad (7)$$

We have separated the bound $\bar{B}_\ell$ into two parts, one summarizing the behavior near zero, and one bound the tail. Decomposing our bound into these two components tends to yield better constants than if we were to try to upper bound the integral from 0 to infinity.

Assumption 2: The separation between the shifts is sufficiently large:

$$\Delta \geq \frac{1}{c}\max\left\{2\xi_0, \bar{B}_0, \frac{\bar{B}_1}{\sqrt{\kappa_2}}, \frac{\bar{B}_2}{\kappa_2}, \frac{\bar{B}_3}{\sqrt{\kappa_2\kappa_4}}\right\}, \quad (8)$$

where $c > 0$ is a small constant.

Now we are ready to present the main theorem of the paper:

Theorem 1: Under Assumptions 1 and 2, if in addition the constant c satisfies

$$c < \min\left\{\frac{1}{11}, \frac{1}{7\sqrt{\kappa_4/\kappa_2^2}}, \frac{1 - B_0\left(\sqrt{\kappa_2/\kappa_4}\right)}{5}\right\} \quad (9)$$

then the decomposition in (1) is the unique atomic decomposition of $x(t)$.

Several remarks follow:

Remark 1 (Scaling Invariance): If for some $\psi(t)$ and Δ the conditions of Theorem 1 are satisfied, then a scaled version $\psi_\sigma(t) = \frac{1}{\sqrt{\sigma}}\psi(\frac{t}{\sigma})$ and $\Delta_\sigma = \Delta\sigma$ also satisfy these conditions. To see this, we note that the auto-correlation function of ψ_σ and its derivatives are $\frac{1}{\sigma^\ell}K_\ell(\frac{t}{\sigma})$ with $\kappa_{2\ell}^\sigma = \frac{1}{\sigma^{2\ell}}\kappa_{2\ell}$. Therefore, we could take $\frac{1}{\sigma^\ell}B_\ell(\frac{t}{\sigma}), |t| \geq \xi_0\sigma$ as the new bounding functions, which implies $\bar{B}_\ell^\sigma = \frac{1}{\sigma^{\ell-1}}\bar{B}_\ell$. The right hand side of (8) is then scaled by σ. The right hand side of (9), however, will remain the same. So if we scale the waveform $\psi(t)$, the minimal separation should also be scaled by the same quantity.

Remark 2: The quantity $1/\sqrt{\kappa_2}$ is a natural measure of the width of the energy density $\psi^2(t)$. To see this, we observe that $\kappa_2 = \int_{\mathbb{R}} |\psi'(t)|^2 dt = \int_{\mathbb{R}} \omega^2|\hat{\psi}(\omega)|^2 d\omega$ where $\hat{\psi}$ is the Fourier transform of ψ. Since Parseval's identity implies $\int_{\mathbb{R}} |\hat{\psi}(\omega)|^2 d\omega = 1$, we could view κ_2 as the variance of energy density $|\hat{\psi}(\omega)|^2$. Similarly, the quantity κ_4/κ_2^2 is the kurtosis of the energy density $|\hat{\psi}(\omega)|^2$.

We apply the theorem to the Gaussian function. We can analytically compute the auto-correlation function and get $\kappa_2 = 1/(2\sigma^2), \kappa_4 = 3/(4\sigma^4)$. Assumption 1 is fulfilled if

we pick $\xi_0 = \sqrt{2\sqrt{10}+10}\sigma \approx 4.05\sigma$ and $B_\ell(t) = |K_\ell(t)|$. We compute $\bar{B}_\ell$ numerically and obtain (8):

$$\Delta \geq \frac{1}{c}\max\{8.09, 0.16, 0.44, 1.12, 1.36\}\sigma = \frac{8.09}{c}\sigma. \quad (10)$$

The condition (9) on c becomes

$$c < \min\left\{\frac{1}{11}, \frac{1}{7\sqrt{3}}, \frac{1-e^{-\frac{1}{6}}}{5}\right\} \approx 0.031. \quad (11)$$

So if the separation $\Delta \geq 264\sigma$ for Gaussian mixtures, then (1) is the unique atomic decomposition. Obviously, this is a terribly conservative bound. However, note that this bound is independent of the number of terms in the original signal and is derived using very few properties of the Gaussian distribution. Numerical simulations in Section IV show that the minimal separation needs to be marginally larger than σ.

A. Duality and Optimality Condition

Define a duality of $C_0(\mathbb{R})$ and $\mathcal{B}(\mathbb{R})$ as

$$\langle f, \mu\rangle := \int_{\mathbb{R}} f(t)\mu(dt), \forall f \in C_0(\mathbb{R}), \mu \in \mathcal{B}(\mathbb{R}). \quad (12)$$

Then the dual problem of atomic decomposition is

$$\begin{aligned} &\underset{\nu\in\mathcal{B}(\mathbb{R})}{\text{maximize}} \quad \langle x, \nu\rangle \\ &\text{subject to} \quad \sup_t |\langle \psi(\cdot - t), \nu\rangle| \leq 1 \end{aligned} \quad (13)$$

Strong duality implies the following proposition:

Proposition 1: Assume $\{\psi(t-\tau_j)\}$ are linearly independent. The decomposition (1) is the unique atomic decomposition of $x(t)$ if there exists a dual certificate measure $\nu \in \mathcal{B}(\mathbb{R})$ such that the corresponding dual function

$$Q(t) := \int_{\mathbb{R}} \psi(s-t)\nu(ds) = \langle \psi(\cdot - t), \nu\rangle \quad (14)$$

satisfies

$$Q(\tau_j) = \text{sign}(c_j), \forall j \quad (15)$$

$$|Q(t)| < 1, t \notin \{t_j\}. \quad (16)$$

In Figure 1, we plot a dual function obtained by solving (a discretized version of) the dual problem (13).

To prove Theorem (1), it suffices to construct a dual function of the form (14) that satisfies (15) and (16).

B. Minimum Energy Construction of Dual Function

In order to construct a dual function in Proposition 1, we require that $Q(t)$ satisfies

$$Q(\tau_j) = \text{sign}(c_j) \text{ and } Q'(\tau_j) = 0, \forall j. \quad (17)$$

Since $Q(t)$ has many more degrees of freedom than the number of constraints, we choose a measure $\nu(ds)$ of the form $q(s)ds$ for some continuous function $q(s)$. Write the conditions in terms of the function q:

$$\int_{\mathbb{R}} \psi(s-\tau_j)q(s)ds = \text{sign}(c_j), \forall j \quad (18)$$

$$\int_{\mathbb{R}} \psi'(s-\tau_j)q(s)ds = 0, \forall j \quad (19)$$

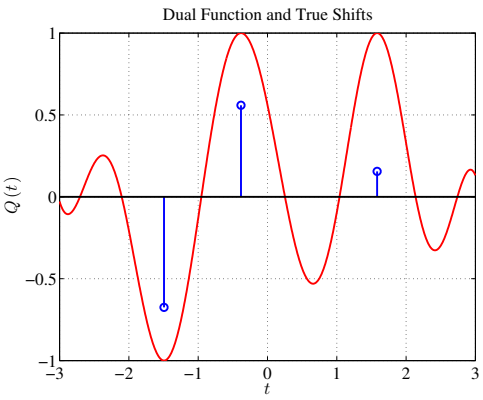

Fig. 1. Dual function and the true shifts.

Among all possible functions $q(\cdot)$ satisfying (18) and (19), we choose the one with minimal energy $\int_{\mathbb{R}} |q(s)|^2 ds$ and hope this will ensure (16) is satisfied.

Define (possibly infinite) matrices $[D_n]_{i,j} = K_n(\tau_i - \tau_j)$. Then the minimal energy solution is

$$q(s) = \sum_j \alpha_j \psi(s-\tau_j) + \sum_j \beta_j \psi'(s-\tau_j) \quad (20)$$

with

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix} = \begin{bmatrix}D_0 & D_1^T\\ D_1 & D_2\end{bmatrix}^{-1}\begin{bmatrix}\text{sign}(c)\\ 0\end{bmatrix} \quad (21)$$

As a consequence, we get the dual function

$$Q(t) = \sum_j \alpha_j K_0(t-\tau_j) + \sum_j \beta_j K_1(t-\tau_j) \quad (22)$$

We have assumed that the 2×2 block matrix in (21) is invertible.

C. Proof Outline of Theorem 1

In this section, we outline the rest argument showing that the constructed $Q(t)$ indeed satisfies the conditions (15) and (16). The argument consists of three main steps:

1) showing the matrix $\begin{bmatrix}D_0 & D_1^T\\ D_1 & D_2\end{bmatrix}$ is invertible
2) showing $|Q(t)| < 1$ in regions near each τ_j by estimating $Q''(t)$
3) showing $|Q(t)| < 1$ in regions far away from any τ_j.

One key ingredient of the proof is the following lemma that estimates sums of the form $\sum_j |K_n(t-\tau_j)|, n = 0, 1, 2, 3$. We observe that if $K_n(\cdot)$ decays fast and the τ_js are well separated, then these sums are well controlled.

Lemma 1: Assume $\tau_1 = 0$ and $\tau_+ = \min\{\tau_j : \tau_j > 0\}$. Then for all $t \in \left[0, \frac{\tau_+}{2}\right]$, we have

$$\sum_{j:\tau_j\neq 0} |K_\ell(t-\tau_j)| \leq \frac{1}{\Delta}\bar{B}_\ell.$$

We now use Lemma 1 to guarantee the invertibility of the matrix $\begin{bmatrix}D_0 & D_1^T\\ D_1 & D2\end{bmatrix}$ and estimate the coefficients α and β:

Lemma 2: Under Assumptions 1 and 2, if $0 < c < \frac{5-\sqrt{17}}{4}$, then the matrix $\begin{bmatrix} D_0 & D_1^T \\ D_1 & D2 \end{bmatrix}$ is invertible, and the coefficient vectors α, β obeying

$$\|\alpha\|_\infty \leq \frac{1-c}{1-3c+c^2}, \|\beta\|_\infty \leq \frac{c}{1-3c+c^2}\frac{1}{\sqrt{\kappa_2}}$$

Furthermore, if $\text{sign}\,(c_1) = 1$, then

$$\alpha_1 \geq \frac{1-5c+2c^2}{1-3c+c^2}.$$

In constructing the dual function, we require that $Q'(\tau_j) = 0$ to guarantee that $|Q(\tau_j)|$ achieves maximum at τ_j. As a consequence, the second order Taylor expansion of $Q(t)$ in a neighborhood $\mathcal{N}(\tau_j)$ of τ_j is

$$Q(t) = Q(\tau_j) + \frac{1}{2}Q''(\xi_t)(t-\tau_j)^2, \forall t \in \mathcal{N}(\tau_j) \qquad (23)$$

where $\xi_t \in \mathcal{N}(\tau_j)$. If we could show $Q''(t) < 0, t \in \mathcal{N}(\tau_j)$ ($Q''(t) > 0, t \in \mathcal{N}(\tau_j)$, resp.) for τ_j with $Q(\tau_j) = 1$ ($Q(\tau_j) = -1$, reps.), then $|Q(t)| < 1$ in the neighborhood $\mathcal{N}(\tau_j)$. The following lemma uses this idea to ensure $|Q(t)| < 1$ in the neighborhood $\mathcal{N}(\tau_j) := \{t \in \mathbb{R} : |t-\tau_j| \leq \sqrt{\kappa_2/\kappa_4}\}$.

Lemma 3: Under Assumptions 1 and 2, if in addition

$$c < \min\left\{\frac{1}{11}, \frac{1}{7\sqrt{\kappa_4/\kappa_2^2}}\right\} \qquad (24)$$

then $|Q(t)| < 1$ for $t \in \mathcal{N}(\tau_j)$.

We bound $|Q(t)|$ directly over the region $\mathbb{R}/\bigcup_j \mathcal{N}(\tau_j)$:

Lemma 4: Under Assumptions 1 and 2, if in addition

$$c < \frac{1 - B_0\left(\sqrt{\kappa_2/\kappa_4}\right)}{5} \qquad (25)$$

then $|Q(t)| < 1$ in $\mathbb{R}/\bigcup_j \mathcal{N}(\tau_j)$.

Combining Lemmas 2, 3, 4, we have proved Theorem 1.

IV. Numerical Experiments

We conducted a numerical experiment to support the theory. We use finite observations and finite discretization of the parameter space and generate the unknown translations from the discretized points, so there is no "off-the-grid" issue. Note that no existing theory can explain the performance of ℓ_1 minimization even for this finite case.

We examine the transition behavior of the rate of success against the separation Δ for the Gaussian function with $\sigma = 1$. We generated the signal (1) using $k = 4$ equispaced translates from $[-5, 5]$ divided into 200 points. We varied the minimal separation $\Delta \in \{0.1 : 0.05 : 2\}$ of the k translates. For each Δ, 100 instances of the signal were generated. Each instance consists of unknown equispaced shifts produced centering around the origin with small random perturbation, and unknown coefficients with random signs and magnitudes of the form $1+N(0,1)^2$. We collected 100 equispaced samples of the signal from $[-5, 5]$. We then solved an ℓ_1 minimization for each signal instance in CVX, thresholded the solution using a cut-off value 10^{-4}, and declared success if CVX successfully returned the correct number of translates at correct locations.

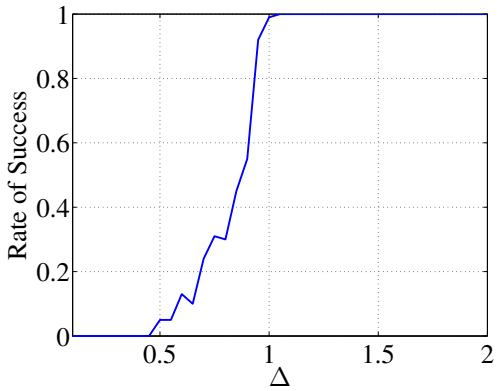

Fig. 2. Rate of success for atomic decomposition as a function of separation.

The rate of success as function of Δ is shown in Figure 2. We see the transition region is only a bit over 1.

V. Conclusions

By explicitly constructing a dual certificate, we derived a sufficient condition for a mixture of translation-invariant signals to be its atomic decomposition. The condition involves the separation between the translation parameters, and the decay speeds of the auto-correlation function of the basic waveform and its derivatives. Numerically experiments were performed to support the theory. In future work, we will analyze the performance of atomic norm minimization for translation-invariant signals with finite samples in noise.

References

[1] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[2] A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, June 2009.

[3] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," *Nature Methods*, vol. 9, no. 7, pp. 721–723, 2012.

[4] H. Rauhut, "Random sampling of sparse trigonometric polynomials," *Applied and Comput. Hamon. Analy.*, vol. 22, no. 1, pp. 16–42, Jan. 2007.

[5] C. Ekanadham, D. Tranchina, and P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, Oct. 2011.

[6] J. Anderson and I. King, "Toward high-precision astrometry with WFPC2. I. deriving an accurate point spread function," *Publ. Astronom. Soc. Pacific*, vol. 112, no. 776, pp. pp. 1360–1382, 2000.

[7] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.

[8] Emmanuel J Candès and Benjamin Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.

[9] E. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, pp. n/a–n/a, 2013.

[10] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *arXiv e-prints*, July 2012.

[11] G. Tang, B. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *ArXiv e-prints*, Mar. 2013.

[12] E. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *arXiv e-prints*, Nov. 2012.