Sparse Recovery Over Continuous Dictionaries -Just Discretize

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

Badri Narayan Bhaskar
University of Wisconsin-Madison
Email: bnbhaskar@wisc.edu

Benjamin Recht
University of California, Berkeley
Email: brecht@berkeley.edu

***Abstract*—In many applications of sparse recovery, the signal has a sparse representation only with respect to a continuously parameterized dictionary. Although atomic norm minimization provides a general framework to handle sparse recovery over continuous dictionaries, the computational aspects largely remain unclear. By establishing various convergence results as the discretization gets finer, we promote discretization as a universal and effective way to approximately solve the atomic norm minimization problem, especially when the dimension of the parameter space is low.**

I. Introduction

Sparse modeling forms a pillar of contemporary signal processing, providing useful techniques for estimation denoising, and detection. The majority of existing work on sparsity focuses on finite dictionaries. However, in applications such as line spectral estimation, ultrasound imaging, radar, MRI, and microscopy imaging, the signals are more appropriately modeled as mixtures of continuously parameterized elementary signals. For example, in radar, the received signal is a mixture of translated and frequency-modulated versions of a known waveform, where the translation and modulation parameters are continuous.

A straightforward strategy to solve sparse recovery with a continuously parameterized dictionary is to discretize the parameter space and reduce the problem to one with a finite dictionary. Indeed, this approach has yielded state-of-the-art results when applied to super-resolution microscopy [1], communication channel sensing, direction-of-arrival (DOA) estimation, high-resolution radar, signal sampling, and remote sensing, to name a few.

However, from a theoretical perspective, signals that have sparse representations in a continuous dictionary might not have sparse representations after discretization [2], raising concerns about the validity of applying sparse recovery algorithms and theories. For sparse frequency estimation, recent work has reformulated the resulted atomic norm minimization as a semidefinite program [3], [4], [5], thus avoiding discretization issues. Similar reformulations to other problems seem rather difficult to obtain. It is our goal in this work to bridge the gap between the superiority of discretization in various practical problems and its theoretical shortcomings, and to promote discretization as a universal recipe to approximately solve atomic norm minimization problems when the parameters indexing the dictionary lie in a space of small dimension.

The paper is organized as follows. In Section II we introduce several atomic norm minimization problems. In Section III, we present our main theorems with proofs. Section IV is devoted to numerical experiments and Section V concludes the paper.

II. Signal Model and Problem Setup

Suppose we have a signal $x^\star \in \mathbb{F}^p$ with $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ represented as a sparse mixture of atoms from the set

$$\mathcal{A} = \left\{a\left(\omega\right) \in \mathbb{F}^p : \omega \in \Omega \subset \mathbb{R}^d\right\},$$

that is

$$x^\star = \textstyle\sum_{j=1}^k c_j^\star a\left(\omega_j^\star\right), c_j^\star \in \mathbb{F}, \omega_j^\star \in \Omega. \tag{1}$$

We consider atoms $a\left(\omega\right) \in \mathbb{F}^p$ that are continuously indexed by the parameter ω in a compact set $\Omega \subset \mathbb{R}^d$. We list a few examples of continuously parameterized atomic sets:

1) spectral estimation, signal sampling, DOA finding:
$$[a(\omega)]_j = e^{ij\omega}, \omega \in [-\pi, \pi]$$
2) delay estimation, imaging:
$$[a(\tau)]_j = g(t_j - \tau), \tau \in [\tau_1, \tau_2]$$
3) radar, GPS:
$$[a(\tau, \nu)]_j = g(t_j - \tau)e^{i2\pi\nu t_j}, \tau \in [\tau_1, \tau_2], \nu \in [\nu_1, \nu_2]$$
4) continuous wavelets:
$$[a(\mu, \sigma)]_j = g(\frac{x_j - \mu}{\sigma}), \mu \in [\mu_1, \mu_2], \sigma \in [\sigma_1, \sigma_2]$$

Here $j \in \{0, \ldots, p-1\}$ and $g(\cdot)$ is a known function.

The atomic norm, defined for any $x \in \mathbb{F}^p$ as

$$\|x\|_{\mathcal{A}} = \inf\left\{\textstyle\sum_j |c_j| : x = \sum_j c_j a\left(\omega_j\right)\right\} \tag{2}$$

was first proposed and analyzed in [6] as a convex proxy to promote sparsity with respect to a general atomic set $\mathcal{A}$. The dual norm of $\|\cdot\|_{\mathcal{A}}$ is $\|z\|_{\mathcal{A}}^* := \sup_{\omega\in\Omega} |\langle z, a\left(\omega\right)\rangle|$.

We consider recovering $x^\star$ using atomic norm minimization from both noise-free and noisy linear measurements. In the noise-free case, given $x^\star$ in (1) we observe $y = \Phi x^\star$, where $\Phi \in \mathbb{F}^{n\times p}$ is a sensing matrix. We recover $x^\star$ by solving the Basis Pursuit type atomic norm minimization:

$$\underset{x\in\mathbb{F}^p}{\text{minimize}} \ \|x\|_{\mathcal{A}} \ \text{subject to} \ y = \Phi x. \tag{3}$$

The dual problem of (3) is

$$\underset{z\in\mathbb{F}^n}{\text{maximize}} \ \langle z, y\rangle_{\mathbb{R}} \quad \text{subject to} \quad \|\Phi^* z\|_{\mathcal{A}}^* \leq 1. \tag{4}$$

or equivalently, a semi-infinite program (SIP) that has a finite-dimensional decision variable and infinitely many constraints:

$$\underset{z\in\mathbb{F}^n}{\text{maximize}}\,\langle z,y\rangle_{\mathbb{R}}\ \text{ subject to }\ |\langle\Phi^* z,a(\omega)\rangle|\le 1,\forall\omega\in\Omega. \quad (5)$$

We have used $\langle x,y\rangle = y^*x$ to denote the inner product and $\langle x,y\rangle_{\mathbb{R}} = \text{Real}(\langle x,y\rangle$ takes the real part.

In the noisy case, we observe $y = \Phi x^\star + w$ with $w\in\mathbb{F}^n$ the additive noise. We use a LASSO-type formulation

$$\underset{x\in\mathbb{F}^p}{\text{minimize}}\ \ \frac{1}{2}\left\|y-\Phi x\right\|_2^2+\tau\left\|x\right\|_{\mathcal{A}} \quad (6)$$

to remove noise. The dual of (6) is

$$\underset{z\in\mathbb{F}^n}{\text{maximize}}\ \ \frac{1}{2}(\|y\|_2^2-\|y-\tau z\|_2^2)\ \text{ subject to }\ \|\Phi^*z\|_{\mathcal{A}}^*\le 1 \quad (7)$$

which is equivalent to another SIP

$$\begin{aligned}&\underset{z\in\mathbb{F}^n}{\text{maximize}}\ \ \frac{1}{2}(\|y\|_2^2-\|y-\tau z\|_2^2)\\ &\text{subject to}\ \ |\langle\Phi^*z,a(\omega)\rangle|\le 1,\forall\omega\in\Omega.\end{aligned} \quad (8)$$

We will see that viewing the dual problems as SIPs provides insight into the discretization approach for solving them.

III. Convergence Results for Discretization

In this section, we present several results on the discretization approach for atomic norm minimization problems. We start with the following general convex SIP

$$\begin{aligned}&\underset{z\in\mathbb{F}^n}{\text{maximize}}\ \ f(z)\\ &\text{subject to}\ \ |\langle\Phi^*z,a(\omega)\rangle|\le 1,\forall\omega\in\Omega.\end{aligned} \quad (9)$$

As special cases, $f(z)=\frac{1}{2}(\|y\|_2^2-\|y-\tau z\|_2^2)$ corresponds to the LASSO dual (8) and $f(z)=\langle z,y\rangle_{\mathbb{R}}$ corresponds to the Basis Pursuit dual (5). The SIP (9) satisfies the Slater's condition [7] since $|\langle\Phi^*0,a(\omega)\rangle|=0<1$.

Throughout the paper, we make the following assumption:

A1: The parameter set $\Omega\subset\mathbb{R}^d$ is compact, and $a(\omega):\Omega\to\mathbb{F}^p$ is continuously differentiable.

This assumption is valid the examples mentioned in Section I. Under Assumption A1, we define a dual of the SIP (5) whose variable is a (signed or complex) measure μ over Ω:

$$\underset{\mu}{\text{minimize}}\ \ \|\mu\|_{\text{TV}}\ \text{ subject to }\ \ y=\int_\Omega\Phi a(\omega)\,d\mu(\omega)\,, \quad (10)$$

where $\|\mu\|_{\text{TV}} := |\mu|(\Omega)$ is the total variation norm of μ. Similarly, the dual of (8) is

$$\underset{\mu}{\text{minimize}}\ \ \frac{1}{2}\|y-\int_\Omega\Phi a(\omega)d\mu(\omega)\|_2^2+\tau\left\|\mu\right\|_{\text{TV}}. \quad (11)$$

Interested readers are encouraged to refer to [7] for a formal and rigorous treatment of the duality theory for SIPs.

We summarize the relationships of the optimizations for the noise-free recovery problem in the following diagram:

$$\begin{array}{ccc}(3) & \xleftrightarrow{\text{duality}} & (4)\\ {\scriptstyle\text{reformulation}}\updownarrow & & \updownarrow{\scriptstyle\text{reformulation}}\\ (10) & \xleftrightarrow{\text{duality}} & (5)\end{array} \quad (12)$$

The optimizations for the noisy recovery problem have a similar relationship. The diagram (12) suggests that, due to strong duality of (3) and (4), strong duality should hold for (10) and (5). Indeed, we have the following proposition that justifies this statement:

Proposition 1: [7, Theorem 2.3, Theorem 3.2] Under Assumption A1, for the primal-dual pairs (10) and (5), (and (11) and (8)):

1) strong duality holds;
2) there exists at least one optimal solution $\hat{\mu}$ to (10) (and (11)) such that $\hat{\mu}=\sum_{j=1}^{n'}\hat{c}_j\delta(\omega-\hat{\omega}_j)$ with $n'\le n$.

A. Discretization

Given a finite set $\Omega_m=\{\omega_1,\dots,\omega_m\}\subset\Omega$, the associated finite optimization

$$\begin{aligned}&\underset{z\in\mathbb{F}^n}{\text{maximize}}\ \ f(z)\\ &\text{subject to}\ \ |\langle\Phi^*z,a(\omega_i)\rangle|\le 1, i=1,\dots,m\end{aligned} \quad (13)$$

is an approximation to (9) under Assumption A1. The optimal value of (13) is larger than that of (9) as the feasible set of the former is larger than that of the latter.

Define $A^m=[a(\omega_1),\cdots,a(\omega_m)]\in\mathbb{F}^{p\times m}$ and $c^m\in\mathbb{F}^m$. The dual problem of (13) is

$$\underset{c^m\in\mathbb{F}^m}{\text{minimize}}\ \|c^m\|_1\ \text{ subject to }\ y=\Phi A^m c^m \quad (14)$$

when $f(z)=\langle z,y\rangle$, or

$$\underset{c^m\in\mathbb{F}^m}{\text{minimize}}\ \frac{1}{2}\|y-\Phi A^m c^m\|_2^2+\tau\|c^m\|_1 \quad (15)$$

when $f(z)=\frac{1}{2}\left(\|y\|_2^2-\|y-\tau z\|_2^2\right)$. We identify c^m with a discrete measure $\mu_m=\sum_{j=1}^m c_j^m\delta(w-\omega_j)$ which is supported on Ω_m. Therefore, the effect of discretization on (10) and (11) is replacing a general measure μ supported on Ω with one supported only on Ω_m.

We use the Hausdorff distance between Ω and Ω_m to quantify the fineness of the discretization:

$$\rho(\Omega_m)=\sup_{\omega\in\Omega}\inf_{\omega'\in\Omega_m}d(\omega,\omega') \quad (16)$$

where $d(\cdot,\cdot)$ is the metric on Ω. We will establish results on two notions of convergence when $\rho(\Omega_m)\to 0$:

1) convergence of the solutions of the discretized dual problems (13) to those of the original problems (9);
2) convergence in distribution of the solutions of the discretized primal problems (14) and (15) to those of the original problems (10) and (11), respectively.

For the LASSO formulation, we also establish results on the speed of convergence.

B. Convergence of the Duals

In this section we present results on the convergence of the solutions and optimal values of the duals. Denote by $S_{\text{d}}(\Omega)$, $F_{\text{d}}(\Omega)$, and $v_{\text{d}}(\Omega)$ the solution set, the feasible set, and the optimal value of the dual problem (9), respectively. $S_{\text{d}}(\Omega_m)$, $F_{\text{d}}(\Omega_m)$, and $v_{\text{d}}(\Omega_m)$ are corresponding quantities for (13).

To guarantee that the feasible set $F_{\rm d}(\Omega)$ and its finer enough discretization $F_{\rm d}(\Omega_m)$ are bounded, we further assume

A2: There exist parameters $\omega_1, \ldots, \omega_n \in \Omega$ such that $\{\Phi a(\omega_1), \ldots, \Phi a(\omega_n)\}$ are linearly independent.

The following theorem shows that, as the discretization gets finer, the solutions and the optimal values of the discretized optimization (13) converge to the those of optimization (9).

Theorem 1: Under Assumptions A1 and A2, if $\rho(\Omega_m) \to 0$, then each sequence of solutions $\hat{z}_m \in S_{\rm d}(\Omega_m)$ satisfies

$$d(\hat{z}_m, S_{\rm d}(\Omega)) \to 0 \quad \text{and} \quad v_{\rm d}(\Omega_m) \to v_{\rm d}(\Omega) \, \text{as} \, m \to \infty$$

where $d(z, S) := \inf_{w \in S} \|z - w\|_2$.

Proof: Assumption A2 implies that

$$\sigma := \inf_z \frac{\left\|[a(\omega_1), \ldots, a(\omega_n)]^* \Phi^* z\right\|_\infty}{\|z\|_2} > 0.$$

For a fine enough grid Ω_m, due to the continuity of $a(\omega)$, we could pick $\{\omega_i'\} \subset \Omega_m$ such that $[a(\omega_1'), \ldots, a(\omega_n')]$ is very close to $[a(\omega_1), \ldots, a(\omega_n)]$ and $\inf_z \frac{\left\|[a(\omega_1'),\ldots,a(\omega_n')]^* \Phi^* z\right\|_\infty}{\|z\|_2} > \sigma/2$. Therefore, $F_{\rm d}(\Omega_m) \subset \left\{z : \|z\|_2 \leq \frac{2}{\sigma}\right\}$ is bounded.

As a consequence of this and the continuity of the constraint function $|\langle \Phi^* z, a(\omega)\rangle|$ for fixed ω, the feasible set $F_{\rm d}(\Omega_m)$ is compact. Therefore, the continuity of the objective $f(\cdot)$ implies that $S_{\rm d}(\Omega_m)$ is nonempty and $v_{\rm d}(\Omega_m)$ is finite. The same is true for $S_{\rm d}(\Omega)$ and $v_{\rm d}(\Omega)$.

Reasoning by contradiction, we suppose there exists a sequence $\{\hat{z}_m\}$ that does not satisfy $d(\hat{z}_m, S_{\rm d}(\Omega)) \to 0$. Then there must be a subsequence $\{\hat{z}_{m_l}\}_{l=1}^\infty$ such that

$$d(\hat{z}_{m_l}, S_{\rm d}(\Omega)) \geq \epsilon > 0 \qquad \forall l$$

Since $\{\hat{z}_{m_l}\} \subset \left\{z : \|z\|_2 \leq \frac{2}{\sigma}\right\}$, we choose a convergent sequence. WLOG, we assume $\{\hat{z}_{m_l}\}$ is convergent and $\hat{z}_{m_l} \to \hat{z}$.

We show $\hat{z} \in F_{\rm d}(\Omega)$ and achieves $v_{\rm d}(\Omega)$. Fix $\omega \in \Omega$, since $\rho(\Omega_m) \to 0$, there exists a sequence $\tilde{\omega}_m \in \Omega_m$ such that $\tilde{\omega}_m \to \omega$. Then we have $|\langle \Phi^* \hat{z}_{m_l}, \tilde{\omega}_{m_l}\rangle| \leq 1$ due to the feasibility of $\hat{z}_{m_l}$. Taking limit yields $|\langle \Phi^* \hat{z}, \omega\rangle| \leq 1$, proving the feasibility of $\hat{z}$. Since $f(\hat{z}_{m_l}) \geq v_d(\Omega)$ due to $F_{\rm d}(\Omega_{m_l}) \supset F_{\rm d}(\Omega)$, the continuity of $f(\cdot)$ entails

$$f(\hat{z}) \geq v_{\rm d}(\Omega), \tag{17}$$

implying $f(\hat{z}) = v_{\rm d}(\Omega)$. We have shown that $\hat{z} \in S_{\rm d}(\Omega)$, contradicting $\epsilon \leq d(\hat{z}_{m_l}, S_{\rm d}(\Omega)) \leq d(\hat{z}_{m_l}, \hat{z}) \to 0$. ∎

C. Convergence of the Primals

In this section, we study the convergence of the solutions $\hat{\mu}_m = \sum_{j=1}^m c_j^m \delta(\omega - \omega_j)$ of the discretized problems (14) (and (15), resp.) to $\hat{\mu}$, the solution of (10) (and (11), resp.). Particularly, we want to establish convergence in distribution.

Denote by $S_{\rm p}(\Omega)$ the set of optimal measures to (3), and $v_{\rm p}(\Omega)$ the optimal value. Define $S_{\rm p}(\Omega_m)$ and $v_{\rm p}(\Omega_m)$ similarly. We have the following theorem:

Theorem 2: Under Assumptions A1 and A2, we have for each sequence of solutions $\hat{\mu}_m \in S_{\rm p}(\Omega_m)$, there exists at least one convergent subsequence, and every convergent subsequence $\{\hat{\mu}_{m_l}\}$ converges in distribution to some $\hat{\mu} \in S_{\rm p}(\Omega)$, and $v_{\rm p}(\Omega_m) \to v_{\rm p}(\Omega)$ as $\rho(\Omega_m) \to 0$.

Proof: We prove the theorem for (10) and its discretization (14). The noisy case can be proved with minor modifications. Strong duality established in Proposition 1 implies $v_{\rm d}(\Omega) = v_{\rm p}(\Omega)$. Each discretized problem is an equality constrained linear program, so strong duality also holds, implying $v_{\rm d}(\Omega_m) = v_{\rm p}(\Omega_m)$. Therefore, Theorem 1 entails $v_{\rm p}(\Omega_m) \to v_{\rm p}(\Omega)$ as $\rho(\Omega_m) \to 0$, or equivalently, $\|\hat{\mu}_m\|_{\rm TV} = \|\sum_{j=1}^m c_j^m \delta(w - \omega_j)\|_{\rm TV} = \|c^m\|_1 \to v_{\rm p}(\Omega)$.

The convergence implies that the sequence $\{\hat{\mu}_m\}$ is bounded. According to Banach-Alaoglu theorem [8, Theorem 5.18, pp. 169], the closed ball of the dual space of a normed vector space is compact in the weak* topology (the topology corresponds to convergence in distribution). Therefore, $\{\hat{\mu}_m\}$ has at least one convergent subsequence in weak* topology.

For any subsequence $\{\hat{\mu}_{m_l}\}$ that converges in distribution, i.e., in the weak* topology, to some $\hat{\mu}$. We need to prove $\hat{\mu} \in S_{\rm p}(\Omega)$. The feasibility of $\hat{\mu}$, or $\Phi \int_\Omega a(\omega)\, d\hat{\mu}(\omega) = y$, is simply a consequence of taking limit in $\Phi \int_\Omega a(\omega)\, d\hat{\mu}_{m_l}(\omega) = y$ (Recall that each component of $a(\omega)$ is continuous).

We next show the optimality of $\hat{\mu}$. The Reize Representation Theorem [8, Corollary 7.18, pp. 223] states that the TV norm in the measure space $\mathcal{M}(\Omega)$ is the operator norm of the linear functional $h \mapsto \int_\Omega h d\mu$ and hence can be written as

$$\|\mu\|_{\rm TV} = \sup_h \left\{ \left| \int_\Omega h d\mu \right| h \in C(\Omega, \mathbb{F}), \|h\|_\infty \leq 1 \right\},$$

where $C(\Omega, \mathbb{F})$ is the set of continuous functions from the compact set $\Omega \subset \mathbb{F}^d$ to $\mathbb{F}$. Consequently, given $\varepsilon > 0$, there exists a continuous function $h \in C(\Omega, \mathbb{F})$ with $\|h\|_\infty \leq 1$ such that $\left|\int_\Omega h d\hat{\mu}\right| \leq \|\hat{\mu}\|_{\rm TV} \leq \left|\int_\Omega h d\hat{\mu}\right| + \varepsilon$. Convergence of $\hat{\mu}_{m_l}$ to $\hat{\mu}$ in distribution implies that $\left|\int_\Omega h d\hat{\mu}_{m_l}\right| \to \left|\int_\Omega h d\hat{\mu}\right|$. Therefore, for large enough l,

$$\begin{aligned} \|\hat{\mu}\|_{\rm TV} &\leq \left|\int_\Omega h d\hat{\mu}\right| + \varepsilon \leq \left|\int_\Omega h d\hat{\mu}_{m_l}\right| + 2\varepsilon \\ &\leq \|\hat{\mu}_{m_l}\|_{\rm TV} + 2\varepsilon = v_{\rm p}(\Omega_{m_l}) + 2\varepsilon \end{aligned}$$

Taking limit in l yields $\|\hat{\mu}\|_{\rm TV} \leq v_{\rm p}(\Omega) + 2\varepsilon$, implying $\|\hat{\mu}\|_{\rm TV} \leq v_{\rm p}(\Omega)$. Hence, $\hat{\mu} \in S_{\rm p}(\Omega)$. ∎

When $S_{\rm p}(\Omega)$ is a singleton, any sequence of $\hat{\mu}_m$ converge to $\hat{\mu}$ in distribution. Additionally, if $\hat{\mu}$ is supported on $\{\hat{\omega}_1, \ldots, \hat{\omega}_k\}$: $\hat{\mu} = \sum_{j=1}^k c_j \delta(\omega - \hat{\omega}_j)$, then we could approximately identify the support from discretized solutions:

Corollary 1: Given ε balls $B_j := B(\hat{\omega}_j, \varepsilon) := \{\omega : d(\omega, \hat{\omega}_j) < \varepsilon\}$ around $\hat{\omega}_j$ with ε small enough, we have $\hat{\mu}_m(B_j) \to c_j$, $|\hat{\mu}_m|(B_j) \to |c_j|$ and $|\mu_m|((\cup_j B_j)^c) \to 0$.

Proof: Pick ε small enough such that B_js have empty intersections. Since $\hat{\mu}(\partial B_j) = 0$, B_j are continuity sets of $\hat{\mu}$. Hence, $\hat{\mu}_m(B_j) \to \hat{\mu}(B_j) = c_j$.

Note that $\hat{\mu}_m$ is a discrete measure. Suppose the restriction of $\hat{\mu}_m$ on B_j has form $\hat{\mu}_m\big|_{B_j} = \sum_l c_l^{m,j} \delta(\omega - \omega_l^{m,j})$, then $\sum_l c_l^{m,j} = \hat{\mu}(B_j) \to c_j$. Clearly, $\left|\hat{\mu}_m\big|_{B_j}\right|(B_j) = \sum_l \left|c_l^{m,l}\right| \geq \left|\sum_l c_l^{m,j}\right| \to |c_j|$, implying $\lim_{m\to\infty} |\hat{\mu}_m|(B_j) \geq |c_j|$.

We argue that $\lim_{m\to\infty} |\hat{\mu}_m|(B_j) = |c_j|$ for all j, and $\lim_{m\to\infty} |\hat{\mu}_m|((\cup_j B_j)^c) = 0$. If otherwise, namely, any of these equalities is strict greater than, then we would get the following contradiction

$$\begin{aligned}\lim_{m\to\infty} \|\hat{\mu}_m\|_{\mathrm{TV}} &= \lim_{m\to\infty} (\sum_{j=1}^{k} |\hat{\mu}_m|(B_j) + |\hat{\mu}_m|((\cup_j B_j)^c)) \\ &= \sum_{j=1}^{k} \lim_{m\to\infty} |\hat{\mu}_m|(B_j) + \lim_{m\to\infty} |\hat{\mu}_m|((\cup_j B_j)^c) \\ &> \sum_{j=1}^{k} |\hat{\mu}|(B_j) + |\hat{\mu}|((\cup_j B_j)^c) = \|\hat{\mu}\|_{\mathrm{TV}}.\end{aligned}$$

■

Corollary 1 suggests that for fine enough discretization, the support of the discretized solutions cluster around the support of the original solution. In practice, we find clustering coupled with approximating one active parameter in each cluster using an average weighted by recovered coefficients give an exceptional heuristic for support recovery.

A second consequence of Theorem 2 is the convergence of $\hat{x}_m := A^m c^m = \int_\Omega a(\omega)\hat{\mu}_m(d\omega)$ to $\hat{x} = \int_\Omega a(\omega)\hat{\mu}(d\omega)$, which is an optimal solution to (3) (or (6)).

Corollary 2: Under Assumptions A1 and A2, as $\rho(\Omega_m) \to 0$, for each sequence $\{\hat{x}_m = A^m c^m\}$, there exists at least one convergent subsequence; every convergent subsequence $\{\hat{x}_{m_l}\}$ converges to some $\hat{x} = \int_\Omega a(\omega)\hat{\mu}(d\omega)$ with $\hat{\mu} \in S_{\mathrm{p}}(\Omega)$.

D. Rate of Convergence

In this section, we present a rate of convergence result for the lasso dual (7). Since the objective function of the dual problem (7) is strictly concave, the optimal solution $\hat{z}$ is unique. Similarly, the discretized dual problems (13) also have unique solutions $\hat{z}_m$. Theorem 1 implies that $\hat{z}_m \to \hat{z}$.

As shown in [9] for general SIPs, the speed of convergence for discretization approach depends crucially on how one discretizes the boundary of the parameter space Ω. We need the following regularity assumption:

A3: Ω is prescribed by a set of inequalities:

$$\Omega = \{\omega \in \mathbb{R}^d : v_i(\omega) \le 0, i \in I\} \tag{18}$$

where I is a finite index set and each function $v_i(\cdot)$ has continuous second order partial derivatives. Furthermore, for any $\omega \in \Omega$, the set of vectors with $I(\omega) := \{i : v_i(\omega) = 0\} \neq \emptyset$

$$\{\frac{\partial v_i(\omega)}{\partial \omega}, i \in I(\omega)\}, \tag{19}$$

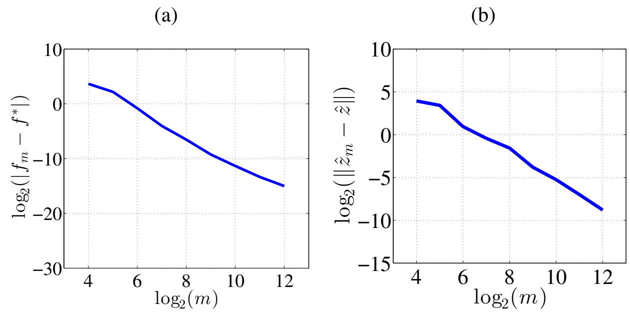
are linearly independent.

For any subset $J \subset I$, define the boundary set $S_J = \{\omega : v_i(\omega) = 0, i \in J\}$. The next assumption dictates that boundaries of Ω of all dimensions are gridded equally fine:

A4: If $S_J \neq \emptyset$, the Hausdorff distance between S_J and the grid set Ω_m restricted to S_J is less than $\rho(\Omega_m)$:

$$\max_{\omega\in S_J} \min_{\omega' \in \Omega_m \bigcap S_J} d(\omega, \omega') \le \rho(\Omega_m). \tag{20}$$

Fig. 1: Convergence of optimal values and solutions



We now present the rate of convergence results, which are direct applications of a theorem by Still:

Theorem 3: Let $\hat{z}, \hat{z}_m$ be the optimal dual and primal solutions for the original denoising problem (7) and its discretized version (13), respectively. Suppose Assumptions A1 and A2 hold and the atoms $a(\omega)$ have continuous second-order parietal derivatives. Then we have:

$$v_{\mathrm{d}}(\Omega) - v_{\mathrm{d}}(\Omega_m) = O(\rho_m^{1/2}), \text{and } \|\hat{z}_m - \hat{z}\| = O\left(\rho_m^{1/2}\right).$$

If in addition Assumptions A3 and A4 are true, then we have accelerated rate of convergence:

$$v_{\mathrm{d}}(\Omega) - v_{\mathrm{d}}(\Omega_m) = O(\rho_m) \text{ and } \|\hat{z}_m - \hat{z}\| = O(\rho_m).$$

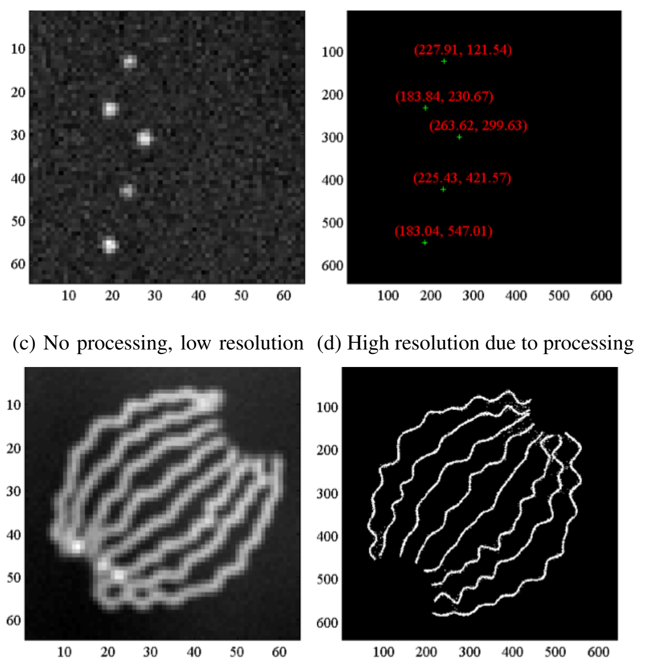
IV. Numerical Experiments

We performed two numerical experiments to validate our theoretical developments. The first experiments is on line spectral estimation, which is the only known example that has atoms with continuous parameterization in a low-dimensional space *and* the atomic norm can be reformulated as a semidefinite program. Hence, the exact solutions to (3), (4), (6), (7) can be found efficiently to quantify the convergence speed.

In the first experiment, we take $\Phi = I$, $p = 64$ and generate the signal $x^\star$ that involves three active frequencies. The noise is complex Gaussian with mean zeros and standard deviation 0.1. We vary the level of discretization m in $\{2^4, 2^5, \ldots, 2^{12}\}$, and solve the finite optimization (13) using cvx for each m to get $\hat{z}_m$ and $v_{\mathrm{d}}(\Omega_m)$. The semidefinite formulation of (7) was solved also using cvx to get $\hat{z}$ and $v_{\mathrm{d}}(\Omega)$. In Figure 1a and 1b, we plot the differences $|v_{\mathrm{d}}(\Omega_m) - v_{\mathrm{d}}(\Omega)|$ and $\|\hat{z}_m - \hat{z}\|_2$ as functions of m in logarithmic scale. The linear decreasing is consistent with our analysis.

In the second experiment, we apply the discretization approach to single molecule imaging. In a typical setup of single molecule imaging, a sub-cellular structure of size tens of nano meters is dyed with fluorophores before being imaged by a microscope. If all hundreds of thousands fluorophores are activated simultaneously, then due to the diffraction of light, the smallest details that can be resolved are limited to 200-300nm. The STORM approach divided the imaging process into tens of thousands cycles and in each cycle only a very small random portion of fluorophores are activated and imaged. Each frame is thus composed of superpositions of translated point spread functions of the microscope. A typical

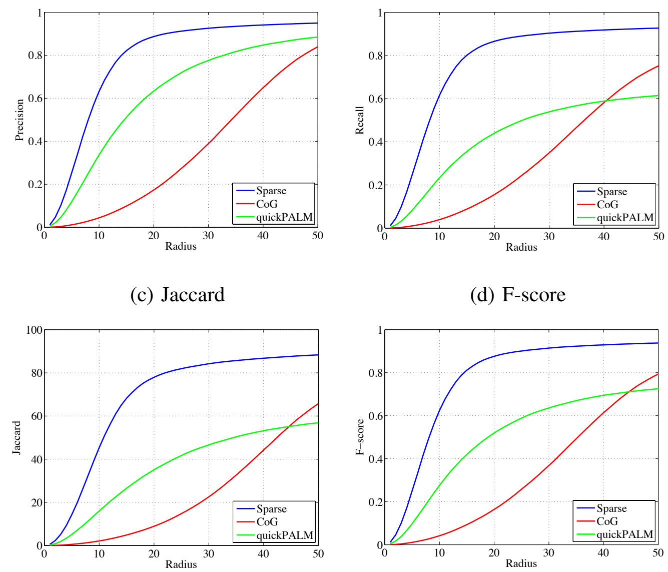
Fig. 2: Single molecule imaging using discretization and Lasso



(a) A typical frame (b) Identified positions

(c) No processing, low resolution (d) High resolution due to processing

Fig. 3: Performance comparison



(a) Precision (b) Recall

(c) Jaccard (d) F-score

frame is shown in Figure 2a. The precise locations of the fluorophores are then estimated from each frame, and stitched together to get a final image of the sub-cellular structure.

Mathematically, each frame is of the form

$$I(k,l) = \sum_j c_j \mathrm{PSF}(k - x_j, l - y_j) \tag{21}$$

where the integral coordinates (k,l) indexes pixels, and the continuous coordinates (x_i, y_j) give the position of the jth fluorophore. The atoms $[a(x,y)]_{k,l} = \mathrm{PSF}(k-x, l-y)$ are indexed by the position parameters. In the "Bundles of tubulins" simulated data provided by the Single-Molecule Localization Microscopy grand challenge organized by ISBI 2013 [10], each frame is of size 64×64 with pixel size 100nm$\times$100nm, while the target resolution is 10nm$\times$10nm. We therefore discretize the field of view into 640$\times$640 pixels, and solve the resulting ℓ_1 minimization problem (15). This approach was first proposed in [1].

Inspired by Corollary 1, we cluster c^m and use a weighted average in each cluster to get an estimate of the position. The final result for a single frame is shown in Figure 2b. Figure 2c shows the image obtained by adding all frames together without processing, while Figure 2d shows the result of adding the processed frames. We observe that processing using sparse recovery significantly improves the resolution.

We also compared the performance of the sparse recovery approach with a naive Center of Gaussian (CoG) approach, and a state-of-the-art quickPALM approach. The plots measure performance using four different metrics: precision, recall, Jaccard, and F-score. More details of these metrics can be found [10]. We see that the sparse approach significantly outperforms the other two methods in all metrics.

V. Conclusions

In this paper we argued that in many applications the sparse recovery problems involve continuously parameterized dictionaries. The atomic norm associated with the dictionary is a powerful convex proxy to promote sparsity. If the parameter space has a small dimension, discretization provides a natural and efficient approach to approximately solve the atomic norm minimization problems. We showed that in both noise-free and noisy cases, the discretized solutions converge to the atomic norm solutions under various criteria, thus providing a solid foundation for applying discretization.

References

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

[2] Y. Chi, L. L. Scharf, A Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[3] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Thy*, vol. 59, no. 11, pp. 7465–7490, 2013.

[4] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, 2013.

[5] 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.

[6] 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.

[7] A. Shapiro, "Semi-infinite programming, duality, discretization and optimality conditions," *Optimization: A Journal of Mathematical Programming and Operations Research*, vol. 58, no. 2, pp. 133–161, 2009.

[8] G. B. Folland, *Real analysis: modern techniques and their applications*, Pure and applied mathematics. Wiley, 1999.

[9] G. Still, "Discretization in semi-infinite programming: the rate of convergence," *Mathematical programming*, vol. 91, no. 1, pp. 53–69, 2001.

[10] EFPL Biomedical Imaging Group, "Single-Molecule Localization Microscopy," http://bigwww.epfl.ch/smlm/, [Online; accessed 20-Nov-2013].