Lift and Unmask: Sparse Models by Lifting the Atomic Norm

Pablo Rademacher
Institute for Mathematical and Computational Engineering
Pontificia Universidad Católica de Chile
Santiago, Chile
pablo.rademacher@uc.cl

Carlos A. Sing Long
Institute for Mathematical and Computational Engineering
Pontificia Universidad Católica de Chile
Santiago, Chile
casinglo@uc.cl

***Abstract*—Sparse signal recovery has become one of the preferred methods to recover signals from a set of incomplete linear measurements. This is due both to its appealing computational properties, as it involves solving a convex optimization problem, and its rigorous justification by the theory of Compressed Sensing. When the underlying signal is not sparse, but it is instead a sparse combination of elementary building blocks called *atoms*, the signal can be recovered by minimizing the *atomic norm*, i.e., the gauge associated to the convex hull of the *atomic set*. Although this approach has been successfully used in several applications, there is an implicit geometric constraint in this approach: only the atoms that are *exposed points* of the convex hull will be selected to represent the solution to atomic norm minimization. This can be an issue when the representation of the underlying signal is sparse when using all the atoms, but dense when using exposed ones. In this work, we propose an approach based on *lifting* that allows us to promote representations using atoms that are not exposed. Our method is based on convex optimization, preserving many of the computational benefits of atomic norm minimization. We present phase diagrams derived from a suitable signal model showing the benefits of using our approach.**

***Index Terms*—Signal reconstruction, signal recovery, Compressed Sensing, sparse signal recovery, atomic norm, atomic norm minimization.**

I. Introduction

In the last decades, ℓ_1-norm recovery has become the favored approach for solving discrete signal recovery problems when the underlying signal is *sparse*, i.e., when the number of its non-zero components is small [1], [2]. Let $\boldsymbol{\Phi} \in \mathbb{R}^{m\times d}$ and suppose that we want to recover an unknown signal $\boldsymbol{x}_0 \in \mathbb{R}^d$ from $m \ll d$ linear measurements $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}_0$. Then ℓ_1-norm recovery consists in solving

$$\min_{\boldsymbol{x}\in\mathbb{R}^d} \quad \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad \Phi\boldsymbol{x} = \boldsymbol{y}. \tag{1}$$

The theory of Compressed Sensing [2], [3] provides conditions under which $\boldsymbol{x}_0$ is the *unique* optimal solution to this problem. If the signal $\boldsymbol{x}_0$ is s-sparse, i.e., at most s of its components are non-zero, and $\boldsymbol{\Phi}$ is a Gaussian matrix with $\Phi_{i,j} \overset{\text{iid}}{\sim} N(0, 1/\sqrt{m})$ then only $m \sim s\log(d/s)$ measurements are enough to ensure that $\boldsymbol{x}_0$ is the unique solution to (1) with high probability [4].

P. R. and C. SL. were partially funded by grants ANID – FONDECYT 1211643. C. SL. was partially funded by the National Center for Artificial Intelligence CENIA FB210017, Basal ANID.

The underlying assumption in ℓ_1-norm recovery is that the underlying signal is sparse. However, in some applications [5]–[7], the underlying signal is not necessarily sparse, but instead a sparse combination of elementary building blocks called *atoms*. *Atomic norm recovery* is the natural extension of ℓ_1-norm recovery to this case [8]. The *atomic norm* induced by a finite collection of atoms $\mathcal{A} := \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ is

$$\rho_{\mathcal{A}}(\boldsymbol{x}) = \inf\left\{\sum\nolimits_{i=1}^n c_i : \ \boldsymbol{x} = \sum\nolimits_{i=1}^n c_i\boldsymbol{a}_i,\ c_i \geq 0\right\}. \tag{2}$$

It follows that $\mathbf{dom}(\rho_{\mathcal{A}}()) = \mathbf{cone}(\mathcal{A})$ and that $\rho_{\mathcal{A}}(\boldsymbol{x}) = \infty$ when $\boldsymbol{x}$ is not a conic combination of atoms. Atomic norm recovery consists on solving

$$\min\nolimits_{\boldsymbol{x}\in\mathbb{R}^d} \quad \rho_{\mathcal{A}}(\boldsymbol{x}) \quad \text{s.t.} \quad \Phi\boldsymbol{x} = \boldsymbol{y}. \tag{3}$$

If we let $\boldsymbol{A} \in \mathbb{R}^{d\times n}$ be the matrix with the atoms as its columns then the above is equivalent to solve

$$\min_{\boldsymbol{c}\in\mathbb{R}^n_+} \quad \sum\nolimits_{i=1}^n c_i \quad \text{s.t.} \quad \Phi\boldsymbol{A}\boldsymbol{c} = \boldsymbol{y}. \tag{4}$$

If $\boldsymbol{c}^\star$ is a solution to the above, then $\boldsymbol{x}^\star := \boldsymbol{A}\boldsymbol{c}^\star$ is an optimal solution to (3). It follows that ℓ_1-norm recovery is a special case of atomic norm recovery when $\mathcal{A} = \{\pm\boldsymbol{e}_1, \ldots, \pm\boldsymbol{e}_d\}$. This approach has been successfully applied to super-resolution [9], [10] and signal processing [11], [12], and presents excellent recovery guarantees [13].

There is an implicit geometric constraint in (3) that has not received much attention in the literature, namely, that not all atoms in $\mathcal{A}$ are used to represent the optimal solution $\boldsymbol{x}^\star$ to (3). If we let $\mathcal{A}^* = \mathcal{A} \cup \{\boldsymbol{0}_d\}$ and denote as $C_{\mathcal{A}} = \mathbf{conv}(\mathcal{A}^*)$ its convex hull, then it is apparent that $\rho_{\mathcal{A}}()$ coincides with the gauge of $C_{\mathcal{A}}$ [14]. Hence, the optimal solution $\boldsymbol{x}^\star$ will *always* lie in the boundary of $\rho_{\mathcal{A}}(\boldsymbol{x}^\star)C_{\mathcal{A}}$. As a consequence, when $\boldsymbol{x}_0$ is a sparse linear combination of atoms $\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_s} \in \mathbf{int}(C_{\mathcal{A}})$, atomic norm recovery will attempt to represent $\boldsymbol{x}_0$ as a sparse linear combination of the exposed points $E_{\mathcal{A}}$ of $C_{\mathcal{A}}$. We call these *exposed atoms* and *masked atoms* all the remaining ones. Since by the Krein-Milman theorem $\mathbf{conv}(\mathcal{A}^*) = \mathbf{conv}(E_{\mathcal{A}^*})$, the atoms lying on the interior of $C_{\mathcal{A}}$ may as well never have been included in the signal model! This may have practical implications as a representation of $\boldsymbol{x}_0$ using exposed atoms may require a large number of them compared to using masked atoms.

A. Previous work and contributions

A popular approach to mitigate the effect of masked atoms is to construct an atomic set $\bar{\mathcal{A}}$ of *normalized atoms* [15] to then solve (3). In this case $\bar{\mathcal{A}}$ is a subset of the unit sphere in $\mathbb{R}^d$ and all the atoms become exposed. However, normalizing atoms removes all the information about their magnitude, which may be itself important in applications. To our knowledge, the impact of this approach on the performance of the method has not been compared to alternatives.

In this work, we propose an alternative approach that does not involve normalizing and is based on a *lifting* to a higher-dimensional space. All the atoms become exposed in the lifted space and thus may be selected to represent the optimal solution to atomic norm recovery. We show how this *lifted atomic set* can be used to improve the reconstruction performance for a simple signal model with exposed and masked atoms over using normalized atoms.

II. Preliminaries

The canonical inner product between two vectors $\boldsymbol{x}, \boldsymbol{y}$ is denoted as $\boldsymbol{x}^\top \boldsymbol{y}$. The vector in $\mathbb{R}^d$ with all components equal to one is denoted as $\mathbf{1}_d$ and the zero vector as $\mathbf{0}_d$. The restriction of a function $f : \mathbb{R}^d \to \mathbb{R}$ to a set $S \subset \mathbb{R}^d$ is denoted as $f|_S$. We assume that the atoms in $\mathcal{A}$ are non-zero.

III. Lifting the atomic norm

To *unmask* the masked atoms, we propose to *lift* the atomic set to a higher-dimensional space. For our purposes, it suffices to map the atomic set to $\mathbb{R}^{d+1}$. Let $\lambda : \mathbb{R}^d \to \mathbb{R}_+$ be the *lifting coordinate* and define the *lifting map* $\Lambda : \mathbb{R}^d \to \mathbb{R}^{d+1}$ as

$$\Lambda(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{x} & \lambda(\boldsymbol{x}) \end{bmatrix}^\top.$$

We let $\widehat{\mathcal{A}} = \Lambda(\mathcal{A}^*)$ be the *lifted atomic set* and we denote $C_{\widehat{\mathcal{A}}} = \mathbf{conv}(\widehat{\mathcal{A}}^*)$. We have the following proposition. We omit its proof for brevity.

Proposition 1. *If $\lambda|_{C_\mathcal{A}} : C_\mathcal{A} \to \mathbb{R}_+$ is concave then every element in $\widehat{\mathcal{A}}$ is in the boundary of $C_{\widehat{\mathcal{A}}}$. If it is strictly concave, then every element in $\widehat{\mathcal{A}}$ is an extreme point of $C_{\widehat{\mathcal{A}}}$.*

If $\widehat{\boldsymbol{a}} \in \widehat{\mathcal{A}}$ is a *lifted atom* then $\widehat{a}_{d+1} = \lambda(\boldsymbol{a})$ for some $\boldsymbol{a} \in \mathcal{A}$. We call $\widehat{a}_{d+1}$ the *lifted coordinate*. From now on we write $\widehat{\mathcal{A}} = \{\widehat{\boldsymbol{a}}_1, \dots, \widehat{\boldsymbol{a}}_{n+1}\}$ where $\widehat{\boldsymbol{a}}_{n+1} = \Lambda(\mathbf{0}_d)$. The *lifted atomic norm* is

$$\rho_{\widehat{\mathcal{A}}}(\widehat{\boldsymbol{x}}) := \inf \left\{ \sum\nolimits_{i=1}^{n+1} c_i : \ \widehat{\boldsymbol{x}} = \sum\nolimits_{i=1}^{n+1} c_i \widehat{\boldsymbol{a}}_i,\ c_i \geq 0 \right\}. \quad (5)$$

The magnitude of the lifted coordinate should reflect the *degree of masking* of the atom, e.g., $\lambda(\boldsymbol{a}) = 0$ if $\boldsymbol{a}$ is an exposed atom in $C_\mathcal{A}$ and $\lambda(\mathbf{0}_d) = 1$. We consider lifting coordinates of the form

$$\lambda(\boldsymbol{x}) = \zeta(\rho_\mathcal{A}(\boldsymbol{x}))$$

where $\zeta : [0,1] \to [0,1]$ is strictly concave with $\zeta(0) = 1$ and $\zeta(1) = 0$. From now on, we let $\boldsymbol{\lambda} \in \mathbb{R}^n$ be the vector of lifted coordinates for all the atoms in $\mathcal{A}$ and we let

$$\widehat{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \mathbf{0}_d \\ \boldsymbol{\lambda} & 1 \end{bmatrix}$$

be the matrix associated to the lifted atomic set. It will be useful to decompose any $\widehat{\boldsymbol{c}} \in \mathbb{R}^{n+1}$ from now on as

$$\widehat{\boldsymbol{c}} = \begin{bmatrix} \boldsymbol{c} & c_{n+1} \end{bmatrix}^\top.$$

Although Proposition 1 does not ensure that all the lifted atoms will be exposed for this choice, they are for the signal model we introduce in Section V-B.

IV. Recovery with the lifted atomic norm

The lifted atomic norm is defined on $\mathbb{R}^{d+1}$ and it cannot be directly used in (3) to replace the atomic norm. Roughly speaking, this is because there is an additional degree of freedom that needs to be accounted for. For instance, if we write

$$\widehat{\boldsymbol{\Phi}} = \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{0}_m \end{bmatrix} \quad \text{and} \quad \widehat{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{x} & \widehat{x}_{d+1} \end{bmatrix}^\top$$

then the optimal solution $\widehat{\boldsymbol{x}}^\star$ to

$$\min_{\widehat{\boldsymbol{x}} \in \mathbb{R}^{d+1}} \quad \rho_{\widehat{\mathcal{A}}}(\widehat{\boldsymbol{x}}) \quad \text{s.t.} \quad \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{x}} = \boldsymbol{y}. \quad (6)$$

is such that $\widehat{x}^\star_{d+1} = 0$ and $\boldsymbol{x}^\star$ is an optimal solution to (3). This follows from the fact that the constraint

$$H_D = \{\widehat{\boldsymbol{x}} \in \mathbb{R}^{d+1} : \ \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{x}} = \boldsymbol{y}\} \quad (7)$$

is a *vertical affine space* in $\mathbb{R}^{d+1}$. For this reason, we propose one strategy to use the lifted atomic norm for signal recovery.

A. Recovery by signal lifting

By fixing the lifting coordinate of the optimization variable to a value $\alpha \geq 0$ we can solve

$$\min_{\widehat{\boldsymbol{x}} \in \mathbb{R}^{d+1}} \quad \rho_{\widehat{\mathcal{A}}}(\widehat{\boldsymbol{x}}) \quad \text{s.t.} \quad \Phi\boldsymbol{x} = \boldsymbol{y},\ \widehat{x}_{d+1} = \alpha. \quad (8)$$

This is equivalent to minimizing the lifted atomic norm on the intersection between the affine space H_D in (7) and the *horizontal slice*

$$H_\alpha := \{\widehat{\boldsymbol{x}} \in \mathbb{R}^{d+1} : \ \widehat{x}_{d+1} = \alpha\}.$$

In practice, this strategy can be implemented by combining (5) with (8) to solve

$$\min_{\boldsymbol{c} \in \mathbb{R}^n_+} \quad (\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c} \quad \text{s.t.} \quad \Phi\boldsymbol{A}\boldsymbol{c} = \boldsymbol{y},\ \boldsymbol{\lambda}^\top \boldsymbol{c} \leq \alpha. \quad (9)$$

If $\boldsymbol{c}^\star$ is an optimal solution to the above, then $\boldsymbol{x}^\star = \boldsymbol{A}\boldsymbol{c}^\star$ is an optimal solution to (8). See Appendix A for details.

The following proposition shows that under mild conditions, by selecting a suitable value of α, we can indeed promote representations of the optimal solution with masked atoms.

Proposition 2. *Let $S \subset \{1, \dots, n\}$ be such that $\boldsymbol{a}_i$ is masked for every $i \in S$. Let $\boldsymbol{x}_0 = \boldsymbol{A}\boldsymbol{c}_0$ with $\mathbf{supp}(\boldsymbol{c}_0) = S$ and let*

$$\lambda_+ = \max\nolimits_{i \in S} \ \lambda(\boldsymbol{a}_i) \quad \textit{and} \quad \lambda_- = \min\nolimits_{i \in S} \ \lambda(\boldsymbol{a}_i)$$

If $\boldsymbol{c}^*$ is an optimal solution to (4)*, and both*

$$\mathbf{1}_n^\top \boldsymbol{c}_0 \leq \frac{\alpha}{1 - \lambda_-} \quad \textit{and} \quad \mathbf{1}_n^\top \boldsymbol{c}_0 < \frac{1}{\lambda_+} \mathbf{1}_n^\top \boldsymbol{c}^*$$

hold, then both $\boldsymbol{c}_0$ and $\boldsymbol{c}^*$ are feasible for (9) *but $\boldsymbol{c}_0$ attains a strictly lower objective than $\boldsymbol{c}^*$.*

Proof of Proposition 2. It follows from our previous arguments that $\boldsymbol{c}^*$ only selects atoms in the boundary, i.e., $\boldsymbol{a}_i$ is exposed for $i \in \mathbf{supp}(\boldsymbol{c}^*)$. As $\lambda(\boldsymbol{a}) = 0$ if $\boldsymbol{a}$ is an exposed atom, it follows that $\boldsymbol{\lambda}^\top \boldsymbol{c}^* = 0$ whence $\boldsymbol{c}^*$ is feasible for (9). From the hypothesis, we have that

$$\boldsymbol{\lambda}^\top \boldsymbol{c}_0 \leq \lambda_+ \mathbf{1}_n \boldsymbol{c}_0 \leq \alpha$$

whence $\boldsymbol{c}_0$ is also feasible for (9). Finally,

$$(\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c}_0 \leq (1 - \lambda_-)\mathbf{1}_n^\top \boldsymbol{c}_0 < \mathbf{1}_n^\top \boldsymbol{c}^* = (\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c}^*$$

from where the claim follows. □

V. Numerical Experiment

A. Atomic set

We first generate an atomic set using the following model. Let $\mathcal{A}_e = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$, let $\ell \in \mathbb{N}$ be a number of layers, and let $\rho_1, \ldots, \rho_\ell \in (0, 1]$ be sparsity parameters. We generate the sets $\mathcal{A}'_1, \ldots, \mathcal{A}'_\ell$ as follows. For every $k \in \{1, \ldots, \ell\}$ we generate d random vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_d$ with support of size $s_k = [\rho_k d]$ selected at random. The non-zero entries are drawn independently from an $\mathsf{Exp}(1)$ distribution. The elements of $\mathcal{A}'_k$ are then

$$\boldsymbol{a}_i = \frac{1}{\sqrt{s_k}} \left(1 - \frac{k}{\ell + 1}\right) \frac{1}{\|\boldsymbol{z}_i\|_2} \boldsymbol{z}_i, \qquad i \in \{1, \ldots, d\}.$$

This normalization ensures that $\mathcal{A}'_k$ is a collection of masked atoms, and that the degree of masking increases with k. The atomic set is then $\mathcal{A} = \mathcal{A}_e \cup \mathcal{A}'_1 \cup \ldots \cup \mathcal{A}'_\ell$. By construction, the atomic norm *is* the ℓ_1-norm restricted to the positive orthant.

B. Atomic signal model

We generate signals using the following *atomic signal model*. Fix parameters $0 \leq \nu_- < \nu_+ \leq 1$. Define

$$I(\nu_-, \nu_+) = \{i \in \{1, \ldots, n\} : \ \nu_- \leq \rho_{\mathcal{A}}(\boldsymbol{a}_i) \leq \nu_+\}$$

and let $n^* = |I(\nu_-, \nu_+)|$. For a given sparsity parameter $\rho \in (0, 1]$, we select a random subset $S \subset I(\nu_-, \nu_+)$ with $|S| = [\rho n^*]$ and we draw a vector $\boldsymbol{c}_0$ with support S such that its non-zero entries are drawn independently from an $\mathsf{Exp}(1)$ distribution. We then let $\boldsymbol{x}_0 = \boldsymbol{A}\boldsymbol{c}_0$ and let $(\boldsymbol{x}_0, \boldsymbol{c}_0)$ be a sample from this model. This ensures that we are always taking a combination of a fraction of ρ atoms with a fixed degree of masking. The generated signals belong to the positive orthant.

C. Phase diagram

To evaluate the performance of different recovery methods using the atomic signal model, we estimate their phase diagram[1]. We choose $d = 256$, $\ell = 2$, $\rho_1 = 0.1$ and $\rho_2 = 0.2$ so that the atoms reduce their sparsity as their degree of masking increases. Observe that they cluster in 3 layers (Fig. 1a). We select $\nu_- = 0.4$ and $\nu_+ = 0.6$. For the atomic set we use in our experiments, we have that $n = 768$, 512 atoms are masked, and $n^* = 248$. We use $\zeta(s) = \sqrt{1 - s^2}$ to compute the lifting coordinate.

In Fig 1b we see samples for $(\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c}_0$ when $\boldsymbol{c}_0$ is drawn from the atomic signal model whereas in Fig. 1c we show the corresponding samples of $(\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c}_1^*$ where $\boldsymbol{c}_1^*$ a minimum ℓ_1-norm representation of $\boldsymbol{x}_0$. Comparing both figures suggests that $\boldsymbol{c}_0$ often achieves a lower objective in (9) than $\boldsymbol{c}_1^*$.

We select the sparsity parameter ρ and the sampling parameter δ from 10 equispaced values in $[0, 1]$. For each value of the pair (ρ, δ) we draw a sample $\boldsymbol{x}_0$ from the atomic signal model and we generate the measurements $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}_0$ using a random matrix as described in Section I. A recovery method is applied to obtain an estimate $\boldsymbol{x}$ of $\boldsymbol{x}_0$. We repeat the process 20 times and we assume that the recovery is exact when $\|\boldsymbol{x} - \boldsymbol{x}_0\|_2 \leq 10^{-2}\|\boldsymbol{x}_0\|_2$, i.e., if the recovery error is less that 1%.

D. Discussion

In Fig. 1d we show for reference the phase diagram for ℓ_1-norm recovery using a sparse signal model, i.e., the atomic signal model when $\mathcal{A} = \mathcal{A}_e$ and $\nu_- = \nu_+ = 1$. In Fig. 1e we show the phase diagram for atomic norm recovery using the atomic signal model. The performance degrades substantially, as only the extremely sparse signals can be recovered with a substantial number of measurements. Remark that this also reflects the performance of ℓ_1-norm recovery for this model. In Fig. 1f we show the phase diagram for atomic norm minimization with normalized atoms. Observe that in this case the method is unable to succeed as it completely destroys the information encoded in the norms of the masked atoms. Finally, the phase diagram for lifted atomic norm recovery for $\alpha = 50$, $\alpha = 100$ and $\alpha = 150$ shows a substantial improvement. Although it is not able to reproduce the diagram in Fig. 1d, it substantially improves over Figs. 1e and 1f. Finally, it shows how the parameter α impacts the phase diagram.

VI. Conclusion

In this work we addressed an implicit geometric constraint in atomic norm minimization, namely, that only exposed atoms are selected to represent the solution. By lifting the atomic set, we were able to propose a method based on convex optimization that is able to selected masked atoms to represent the optimal solution. The computational burden compared to atomic norm minimization seems to be minimal, while our numerical results show that the performance of our method can improve over atomic norm minimization in the atomic signal model that we used.

As future research directions, we will develop alternative strategies to use the lifted atomic norm for signal recovery. In light of our numerical results, we will also focus on developing a general atomic signal model to derive guarantees for exact signal recovery.

[1]The code to reproduce the results can be found in the GitHub repository `csl-lab/liftedAtomicNormRecovery`

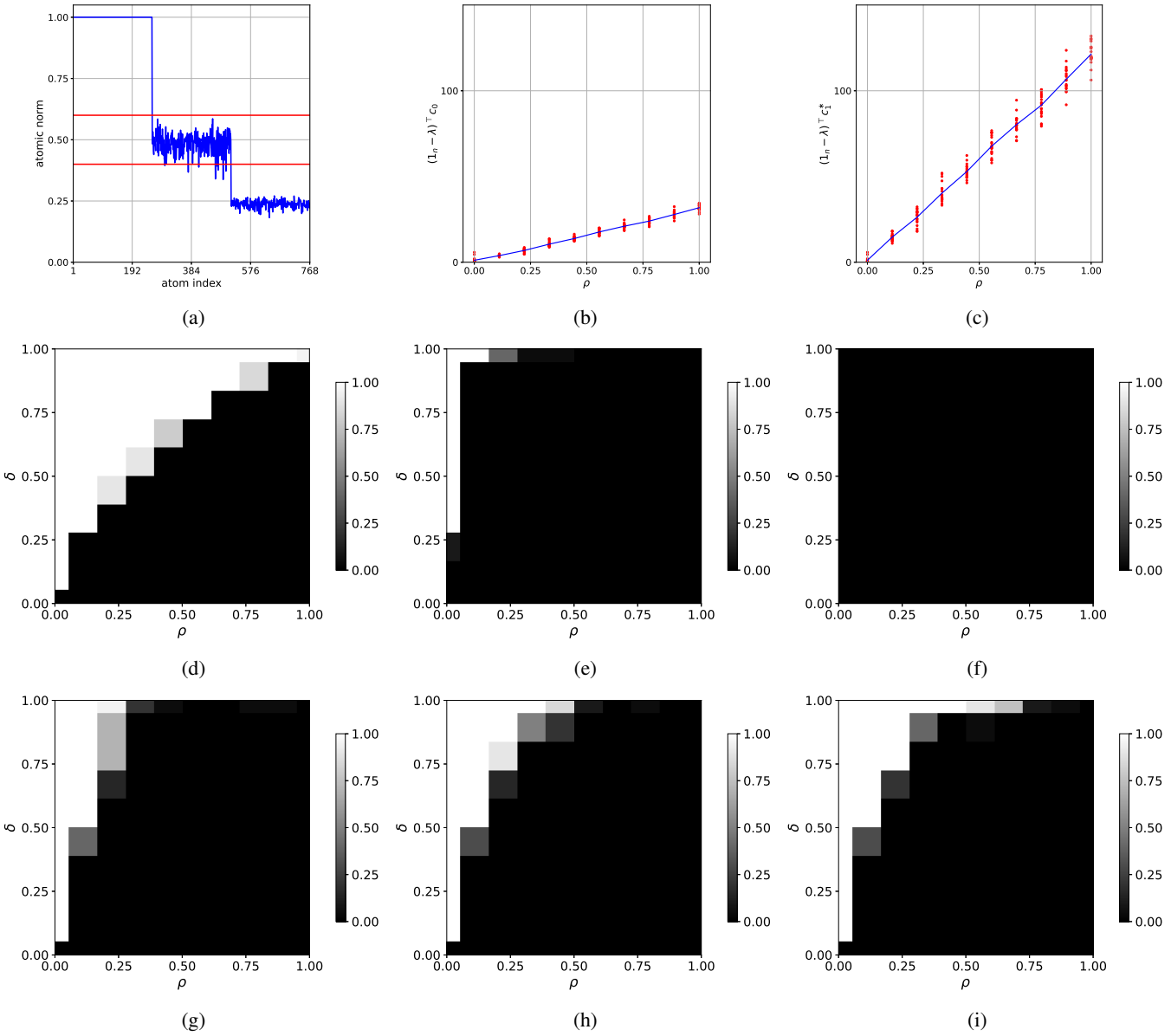

Fig. 1: (a) Atomic norms of the atoms. (b) Samples of $(\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c}_0$ for $\boldsymbol{c}_0$ drawn from the atomic signal model. (c) Samples of $(\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c}_0$ associated to the minimum ℓ_1-norm representation of $\boldsymbol{x}_0$. (d) Phase diagram using a sparse signal model for ℓ_1-norm recovery. (e-i) Phase diagram using the atomic signal model for: (e) Atomic norm recovery; (f) Normalized atomic norm recovery; (g, h, i) Lifted atomic norm with (g) $\alpha = 50$, (h) $\alpha = 100$ and (i) $\alpha = 150$.

APPENDIX

A. *Equivalence between* (8) *and* (9)

From (6) and (8) we can solve

$$\min_{\widehat{\boldsymbol{c}}\in\mathbb{R}_+^{n+1}} \quad \mathbf{1}_{n+1}^\top \widehat{\boldsymbol{c}} \quad \text{s.t.} \quad \Phi \boldsymbol{A}\boldsymbol{c} = \boldsymbol{y},\ \boldsymbol{\lambda}^\top \boldsymbol{c} + c_{n+1} = \alpha. \quad (10)$$

From the second constraint we have $c_{n+1} = \alpha - \boldsymbol{\lambda}^\top \boldsymbol{c}$ whence

$$\mathbf{1}_{n+1}^\top \widehat{\boldsymbol{c}} = \mathbf{1}_n \boldsymbol{c} + c_{n+1} = (\mathbf{1}_n - \boldsymbol{\lambda}^\top)\boldsymbol{c} + \alpha.$$

We relax the constraint $c_{n+1} = \alpha - \boldsymbol{\lambda}^\top \boldsymbol{c}$ to $\alpha - \boldsymbol{\lambda}^\top \boldsymbol{c} \geq 0$. This yields (9). Now, let $\widehat{\boldsymbol{c}}^\star$ be a solution to (10). Then $\boldsymbol{c}^\star$ is feasible for (9). If there exists $\boldsymbol{c}$ attaining a lower objective in (9) then

$$\widehat{\boldsymbol{c}} = \begin{bmatrix} \boldsymbol{c} & \alpha - \boldsymbol{\lambda}^\top \boldsymbol{c} \end{bmatrix}^\top$$

is feasible for (10) and

$$\begin{aligned} \mathbf{1}_{n+1} \widehat{\boldsymbol{c}}^\star &= \mathbf{1}_n^\top \boldsymbol{c} + (\alpha - \boldsymbol{\lambda}^\top \boldsymbol{c}) = (\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c} + \alpha \\ &< (\mathbf{1}_n - \boldsymbol{\lambda})^\top \boldsymbol{c}^\star + \alpha = \mathbf{1}_n^\top \boldsymbol{c}^\star + c_{n+1}^\star = \mathbf{1}_{n+1}^\top \widehat{\boldsymbol{c}}^\star \end{aligned}$$

contradicting the optimality of $\widehat{\boldsymbol{c}}^\star$. Consequently $\boldsymbol{c}^\star$ is optimal for (9). A similar argument proves the converse, showing the equivalence. We omit the details for brevity.

References

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, Jan. 1, 1996. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

[2] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006. DOI: 10.1109/TIT.2005.862083.

[3] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006. DOI: 10.1109/TIT.2006.871582.

[4] J. A. Tropp, "Convex recovery of a structured signal from independent random linear measurements," in *Sampling Theory, a Renaissance*, G. E. Pfander, Ed., Series Title: Applied and Numerical Harmonic Analysis, Cham: Springer International Publishing, 2015, pp. 67–101. DOI: 10.1007/978-3-319-19749-4_2.

[5] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996. DOI: 10.1038/381607a0.

[6] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Sep. 1, 1994. DOI: 10.1093/biomet/81.3.425.

[7] S. Jagabathula and D. Shah, "Inferring rankings using constrained sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7288–7306, Nov. 2011. DOI: 10.1109/TIT.2011.2165827.

[8] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, Dec. 2012. DOI: 10.1007/s10208-012-9135-7.

[9] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, Jun. 2014. DOI: 10.1002/cpa.21455.

[10] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5145–5157, Oct. 1, 2016. DOI: 10.1109/TSP.2016.2576422.

[11] P. Shah, B. N. Bhaskar, G. Tang, and B. Recht, "Linear system identification via atomic norm regularization," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, Maui, HI, USA: IEEE, Dec. 2012, pp. 6265–6270. DOI: 10.1109/CDC.2012.6426006.

[12] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013. DOI: 10.1109/TSP.2013.2273443.

[13] M. März, C. Boyer, J. Kahn, and P. Weiss, "Sampling rates for $$\ell ^1$$-synthesis," *Foundations of Computational Mathematics*, vol. 23, no. 6, pp. 2089–2150, Dec. 2023. DOI: 10.1007/s10208-022-09580-w.

[14] R. T. Rockafellar, *Convex Analysis* (Princeton Landmarks in Mathematics and Physics). Princeton: Princeton University Press, 2015, 470 pp.

[15] M. Aharon, M. Elad, and A. Bruckstein, "$Rm k$-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006. DOI: 10.1109/TSP.2006.881199.