

Distribution-aware Block-sparse Recovery via Convex Optimization

Sajad Daei, Farzan Haddadi, Arash Amini

Abstract—We study the problem of reconstructing a block-sparse signal from compressively sampled measurements. In certain applications, in addition to the inherent block-sparse structure of the signal, some prior information about the block support, i.e. blocks containing non-zero elements, might be available. Although many block-sparse recovery algorithms have been investigated in Bayesian framework, it is still unclear how to incorporate the information about the probability of occurrence into regularization-based block-sparse recovery in an optimal sense. In this work, we bridge between these fields by the aid of a new concept in conic integral geometry. Specifically, we solve a weighted optimization problem when the prior distribution about the block support is available. Moreover, we obtain the unique weights that minimize the expected required number of measurements. Our simulations on both synthetic and real data confirm that these weights considerably decrease the required sample complexity.

Index Terms—Block sparse recovery, Bayesian information, Conic integral geometry, Convex optimization.

I. INTRODUCTION

COMPRESSED Sensing (CS) has emerged in the past decade as a modern technique for recovering a sparse vector $\mathbf{x} \in \mathbb{R}^n$ from compressed measurements (see [1], [2] for more explanations about this field). In this work, we consider signals $\mathbf{x} \in \mathbb{R}^n$ that have a block-sparse structure, namely their non-zero entries appear in blocks. This property has been referred in the literature as block-sparsity. It is common to use the following optimization problem for recovering the signal \mathbf{x} from compressive measurements.

$$P_{1,2}^\eta : \min_{\mathbf{z} \in \mathbb{R}^n} \|\mathbf{z}\|_{1,2} := \sum_{b=1}^q \|\mathbf{z}_{\mathcal{V}_b}\|_2, \quad \text{s.t. } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ represents a fat measurement matrix with $m \ll n$, $\{\mathcal{V}_b\}_{b=1}^q$ are the default disjoint blocks of size $\{k_b\}_{b=1}^q$ that partition the set $\{1, \dots, n\}$, $\mathbf{y} := \mathbf{A}\mathbf{x} + \mathbf{e} \in \mathbb{R}^m$ is the observation vector¹, \mathbf{e} is the noise term which is considered to be i.i.d. Gaussian with variance σ^2 , and η is an upper-bound for $\|\mathbf{e}\|_2$. Most of the earlier literature in block-sparse recovery is focused on the case of single constraint $\|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta$. However, in many applications such as DNA micro-arrays [3], [4], computational neuroscience [5] multi-band signal reconstruction, multiple measurement vector (MMV) problem [6], and the reconstruction of signals in union of subspaces [7],

[8] [9], [10], there exist additional information (or alternatively additional constraints in $P_{1,2}^\eta$) about the signal of interest. In this work, we explore the benefits of having access to extra information about the distribution of the block support (blocks containing non-zero elements) on the required number of measurements. To this end, we propose the optimization problem

$$P_{1,2,w}^\eta : \min_{\mathbf{z} \in \mathbb{R}^n} \|\mathbf{z}\|_{1,2,w} := \sum_{b=1}^q w_b \|\mathbf{z}_{\mathcal{V}_b}\|_2 \\ \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \leq \eta, \\ \mathbf{z} \in \mathcal{M}$$

where the quantities $w_b, b = 1, \dots, q$ are some positive scalars, $\mathbf{w} := [w_1, \dots, w_q]^T$ and \mathcal{M} is some predefined model that restricts the feasible set of the solution. Specifically, we consider two new models for prior information that are of practical interest:

- **Model 1 (Prior distribution)** : We assume that the prior distribution of the block support is available. Under this setting, there are known probabilities associated with each block index $b \in \{1, \dots, q\}$. Namely,

$$\mathbb{P}(b \in \text{bsupp}(\mathbf{x})) = p_b \quad b = 1, \dots, q, \quad (2)$$

where $\text{bsupp}(\cdot)$ returns the block support of a vector.

- **Model 2 (Multiple block support estimates)**: We consider L disjoint sets $\mathcal{P}_i \subset \{1, \dots, q\}$ with $i = 1, \dots, L$ that intersect $\text{bsupp}(\mathbf{x})$ with the expected accuracy

$$\alpha_i := \frac{\mathbb{E}|\text{bsupp}(\mathbf{x}) \cap \mathcal{P}_i|}{|\mathcal{P}_i|} \quad i = 1, \dots, L. \quad (3)$$

To each subset $\mathcal{P}_i \subset \{1, \dots, q\}$, we assign a fixed weight λ_i . In fact, it holds that

$$\mathbf{w} = \sum_{i=1}^L \lambda_i \mathbf{1}_{\mathcal{P}_i} = \mathbf{D}\boldsymbol{\lambda} \in \mathbb{R}^q, \quad (4)$$

where $\mathbf{1}_C$ is the indicator function of the set C , $\mathbf{D} := [\mathbf{1}_{\mathcal{P}_1}, \dots, \mathbf{1}_{\mathcal{P}_L}]$, and $\boldsymbol{\lambda} := [\lambda_1, \dots, \lambda_L]^T$.

One of the applications of Model 1 and 2 is in direction of arrival (DOA) estimation. In this application, Model 1 implies that the probability p_b of having a target in an angle indexed by b is known in advance [11]. In Model 2, however, we assume to know the expected number of targets in a range of angles represented by \mathcal{P}_i . Such statistics might be available from previous measurements in a dynamic scenario. Obviously, Model 1 imposes more strict conditions as knowing the probabilities for all angles is not easily achievable. In

S. Daei and F. Haddadi are with the School of Electrical Engineering, Iran University of Science & Technology. A. Amini is with EE department, Sharif University of Technology.

¹Our analysis holds for both real and complex-valued signals and measurements.

contrast, Model 2 could be applicable as the full angular range could be divided into 3 or 4 intervals, for which we can evaluate the average number of included targets.

In this work, we obtain the weights $\mathbf{w}^* \in \mathbb{R}^q$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^L$ that minimize a threshold m_0 describing the expected number of required measurements for Models 1 and 2, respectively. Our approach is to find a suitable upper-bound for m_0 . The bound is not necessarily tight but leads to closed-form expressions for $\mathbf{w} \in \mathbb{R}^q$ and $\boldsymbol{\lambda} \in \mathbb{R}^L$ in Models 1 and 2, respectively.

A. Related works

CS in presence of prior information has been studied in different signal models. While a large part of research (see for example [12]–[17]) deals with deterministic signal models, only a few works (see [18]–[21]) have investigated random signal models with Bayesian information. In the deterministic model, the ground-truth signal has intersected with a few sets which called support estimates. The contributing level of each set to the support is available to the experimenter [12], [13]. This exact situation is investigated in [17]. They propose a non-uniform model for capturing deterministic prior information. The work [19] considers a probabilistic model where there is a continuous shape function describing the probability of contributing each element to the support. The authors obtain an upper-bound for failure probability of weighted ℓ_1 minimization. Their approach is based on calculating the internal and external angles of a weighted cross polytope. With a different approach, [20] has investigated a discrete measure for Bayesian information (a special case of Model 1 with $k = 1$). They relate the weights of weighted ℓ_1 minimization to the discrete probability distribution by minimizing the expected intrinsic volumes of a weighted cone.

B. Contributions

As listed below, we have three main contributions in this work. Besides, our results are also applicable in DOA estimation (see Section I-C) and functional magnetic resonance imaging (fMRI) reconstruction with parallel coils [22].

- 1) *Optimally exploiting the block distribution.* In presence of a block distribution, we obtain the optimal weights in $\mathbb{P}_{1,2,\mathbf{w}}^\eta$. This result can be considered as an extension of [20] to the block-sparse (and joint-sparse²) setting. However, the derivation of the optimal weights in this case is non-trivial and rather challenging. Further, our derivation approach is different from [20].
- 2) *Optimally exploiting multiple estimates.* In presence of multiple block-support estimates, we derive the optimal penalizing coefficients corresponding to each set in weighted ℓ_1 and $\ell_{1,2}$ minimizations.
- 3) *Robustness of optimal weights against inaccurate prior information.* We analytically examine how close one can get to the optimal weights if the prior information (p_b s and α_i s) is inaccurate. This result is important

²In this case, the non-zero blocks have common support.

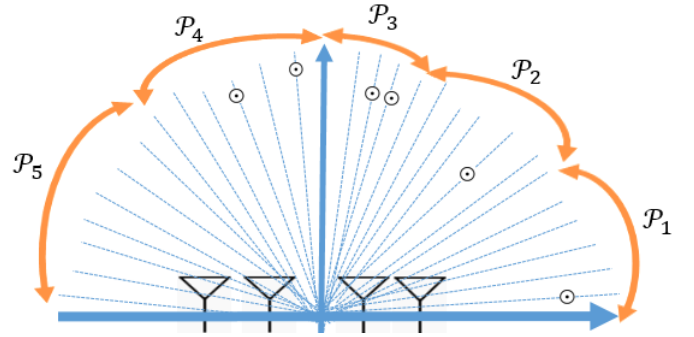


Fig. 1. Schematic diagram of DOA estimation of far-field sources. The angular half-space is divided into $q = 30$ angular clusters $\{\mathcal{V}_b\}_{b=1}^q$ with equal length. The associated parameters corresponding to Model 2 are $L = 5$, $\alpha_1 = \frac{1}{6}$, $\alpha_2 = \frac{1}{6}$, $\alpha_3 = \frac{1}{2}$, $\alpha_4 = \frac{1}{3}$ and $\alpha_5 = 0$.

in practical scenarios, as we only have access to the approximations of p_b s and α_i s.

C. Application (Broadband DOA Estimation)

Suppose that s far-field broadband signals $\{x_i(t, f)\}_{i=1}^s$ in the frequency range $f \in [f_L, f_H]$ incident on an q -element uniform linear array (ULA). The received signal in m sensors at time t and l -th frequency bin can be expressed as:

$$\mathbf{y}(t, f_l) = \sum_{i=1}^s \mathbf{a}_i(f_l) x_i(t, f_l) + \mathbf{e}(t, f_l) \in \mathbb{C}^m, \quad (5)$$

where $\mathbf{a}_i(f_l) := [1, \dots, e^{-j2\pi f_l \frac{(q-1)d}{c} \sin(\theta_i)}]^T$ is the steering vector, c is the propagation velocity, d is the inter-sensor spacing and $\mathbf{e}(t, f_l)$ is a Gaussian noise term with variance σ^2 . In practice, one has to take several snapshots $\{\mathbf{y}(t, f_l)\}_{t=1}^k$. This temporal redundancy is crucial in practice since the array size is limited due to physical constraints [23], [24]. Consequently, one may write

$$\begin{aligned} \mathbf{Y}(f_l) &:= [\mathbf{y}(1, f_l), \dots, \mathbf{y}(k, f_l)] = \\ &\mathbf{A}(f_l) \mathbf{X}(f_l) + \mathbf{E}(f_l) \in \mathbb{C}^{m \times k}, \end{aligned} \quad (6)$$

where $\mathbf{X}(f_l) = [\mathbf{x}(1, f_l), \dots, \mathbf{x}(k, f_l)] \in \mathbb{C}^{q \times k}$, $\mathbf{A}(f_l) = [\mathbf{a}_1(f_l), \dots, \mathbf{a}_q(f_l)] \in \mathbb{C}^{m \times q}$ and $\mathbf{E}(f_l) \in \mathbb{C}^{m \times k}$ is defined similar to $\mathbf{Y}(f_l)$. If the sources is time-invariant over the period of snapshotting, then for all $t = 1, \dots, k$ the non-zero dominant peaks in $\mathbf{x}(t, f_l)$ occur at the same locations corresponding to the ground-truth DOAs. Hence, DOA estimation can be cast as recovering a joint sparse signal $\mathbf{X}(f_l)$ from $\mathbf{Y}(f_l)$. In addition, it is realistic for a radar engineer to know the probability of appearing the ground-truth DOAs in some angular bands [11] (see Figure 1 for a schematic model of this scenario).

Notation. Throughout, scalars are denoted by lowercase letters, vectors by lowercase boldface letters, and matrices by uppercase boldface letters. The i th element of a vector \mathbf{x} is shown either by $x(i)$ or x_i . \mathcal{C}° denotes the polar of a cone \mathcal{C} . We show sets (e.g. \mathcal{B}) by calligraphic uppercase letters. We show the set $\{1, \dots, n\}$ by $[n]$. $\bar{\mathcal{B}}$ is used to represent the complement $[n] \setminus \mathcal{B}$ of a set $\mathcal{B} \subset [n]$.

II. MAIN RESULTS

In the following propositions, we obtain closed-form solutions for optimal weights in case that \mathbf{x} satisfies Model 1 and 2. The proofs are provided in Appendix B.

Proposition 1. *Let $\mathbf{x} \in \mathbb{C}^n$ satisfy Model 1 with parameter $\mathbf{p} = [p_1, \dots, p_q]^T$. Then, the optimal weights $\mathbf{w}^* = [w_1^*, \dots, w_q^*]^T$ in $\mathcal{P}_{1,2,\mathbf{w}}$ are obtained by solving the following equations simultaneously:*

$$\frac{p_b}{1-p_b} w_b^* = \frac{1}{2^{\frac{k_b}{2}-1} \Gamma(\frac{k_b}{2})} \int_{w_b^*}^{\infty} (u - w_b^*) u^{k_b-1} e^{-\frac{u}{2}} du, \quad (7)$$

$$b = 1, \dots, q.$$

Remark 1. (Prior work) The special case $k_b = 1$ (\mathbf{x} is sparse instead of block-sparse) reduces (7) to the weighted ℓ_1 minimization which is studied in [20]. Therefore, Proposition 1 generalizes the results of [20] to the block-sparse case. However, our approach to reach this generalized result is different from and somewhat simpler than [20].

Proposition 2. *Let $\mathbf{x} \in \mathbb{C}^n$ be decomposed into q blocks \mathcal{V}_b of equal size k . Assume that there exist L independent estimates $\{\mathcal{P}_i\}_{i=1}^L$ of $\text{bsupp}(\mathbf{x})$ with parameter $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_L]^T$. Then, the optimal weights $\boldsymbol{\lambda}^* = [\lambda_1^*, \dots, \lambda_L^*]^T$ in Model 2 are obtained by solving*

$$\frac{\alpha_i}{1-\alpha_i} \lambda_i^* = \frac{1}{2^{\frac{k}{2}-1} \Gamma(\frac{k}{2})} \int_{\lambda_i^*}^{\infty} (u - \lambda_i^*) u^{k-1} e^{-\frac{u}{2}} du, \quad (8)$$

$$i = 1, \dots, L.$$

Remark 2. The optimal weights w_b^* 's and λ_i^* 's in Propositions 1 and 2 are respectively obtained by minimizing an upper-bound of the expected number of required measurements in problems $\mathcal{P}_{1,2,\mathbf{w}}$ and $\mathcal{P}_{1,2,D\boldsymbol{\lambda}}$. We numerically observe that the upper-bound is tight for non-uniform distributions of \mathbf{x} , but the exact identification of such distributions is beyond the scope of this work. Further, the system of equations in (7) and (8) are solved using the `fzero` function in MATLAB.

In applications, it is of important practical value to know how the inaccuracies of p_b and α_i in Model 1 and 2, respectively, affect the optimal weights. The following theorem is about this challenge.

Theorem 1. *Assume that \mathbf{p} and \mathbf{p}' be the true and approximate estimate of $\text{bsupp}(\mathbf{x})$, respectively. Let \mathbf{w} and \mathbf{w}' be the optimal weights corresponding to \mathbf{p} , and \mathbf{p}' . Then, there exists a constant $c(k_b, p_b)$ such that*

$$|w_b - w'_b| \leq c(k_b, p_b) |p_b - p'_b| \quad b = 1, \dots, q$$

$$c(k_b, p_b) := \frac{\left(\sqrt{2} h(p_b) \left(\Gamma(\frac{k_b}{2}) - \gamma(\frac{k_b}{2}, \frac{h(p_b)^2}{2}) \right) + 2\gamma(\frac{k_b}{2}, \frac{h(p_b)^2}{2}) \right)^2}{2\sqrt{2} \Gamma(\frac{k_b}{2}) \gamma(\frac{k_b}{2}, \frac{h(p_b)^2}{2})} \quad (9)$$

where $\gamma(a, z) := \int_z^{\infty} u^{a-1} e^{-u} du$ is incomplete gamma function and $h(\cdot) : p_b \rightarrow w_b$ is the nonlinear function in (7).

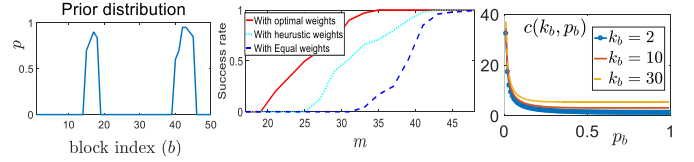


Fig. 2. The middle image compares the success rate of $\mathcal{P}_{1,2,\mathbf{w}^*}^0$ with that in $\mathcal{P}_{1,2}^0$ when there exists a prior distribution $\mathbf{p} \in \mathbb{R}^q$ (depicted in the left image) about the block indexes $\{1, \dots, 50\}$. The optimal weight \mathbf{w}^* is obtained by (7). The right image represents $c(k_b, p_b)$ in Theorem 11 for different k_b 's.

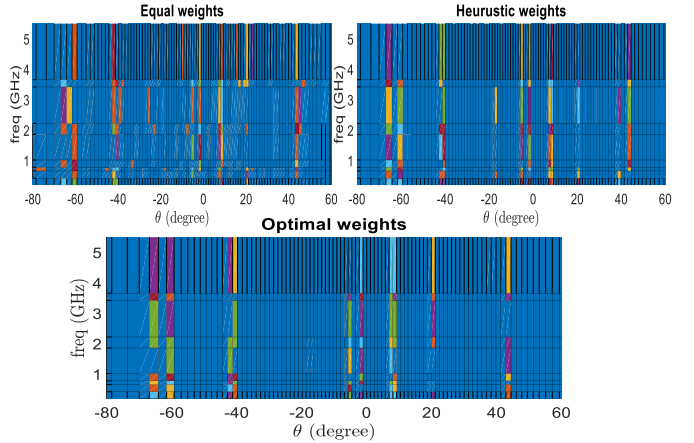


Fig. 3. Broadband DOA estimation for 10 sources (located at the angles $-66.9^\circ, -61.64^\circ, -42.84^\circ, -41.3^\circ, -5.74^\circ, -2.3^\circ, 6.89^\circ, 8.05^\circ, 19.88^\circ, 42.84^\circ$) using an $m = 15$ sensor ULA. The associated parameters are $k = 10, q = 100, d = 5, c = 3 \times 10^8$, and $\sigma = 1$. The expected accuracy corresponding to three sets $\{\mathcal{P}_i\}_{i=1}^3$ are $\alpha_1 = \frac{4}{5}, \alpha_2 = \frac{2}{3}$, and $\alpha_3 = 0$. The top left, top right and bottom images correspond to the recovery using $\mathcal{P}_{1,2,D\boldsymbol{\lambda}}^0$, when equal, heuristic and optimal weights $\boldsymbol{\lambda}$ (calculated using (8)) are applied, respectively.

From the above theorem and the right image of Figure 2, one can infer that the method of obtaining w_b^* and λ_i^* is robust to slight changes of p_b and α_i as long as they are greater than approximately $\frac{1}{10}$.

III. SIMULATIONS

In the first experiment, we construct a random block-sparse $\mathbf{x} \in \mathbb{R}^{250}$, whose building blocks \mathcal{V}_b have equal size $k = 5$. The probability of activating each block is taken from the vector \mathbf{p} depicted in the left image of Figure 2. Then, this signal is observed through a measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times 250}$, of which the elements are drawn from i.i.d. standard normal distribution. We obtain the optimal weights \mathbf{w}^* corresponding to \mathbf{p} by solving (7). We also examine the heuristic weights $w_b = \frac{1}{p_b + \epsilon}$. In the middle image of Figure 2, we plot the success rate as a function of m . For each m , we average over 100 realizations of \mathbf{A} and \mathbf{x} . As expected, $\mathcal{P}_{1,2,\mathbf{w}}$ requires fewer measurements for successful recovery with optimal weights \mathbf{w}^* rather than the other two options (heuristic or equal weights). In turn, heuristic weights are also superior to the equal weights.

In the second experiment, we test DOA estimation using broadband signals (see Subsection I-C and Figure 1). The angular half-space $[-90^\circ, 90^\circ)$ is divided into $q = 100$ angular grids. For each frequency bin $f_l \in [0, 5]$ GHz, we take $k = 10$

snapshots. We assume that there exist three ($L = 3$) sets \mathcal{P}_i with expected accuracies $\alpha_1 = \frac{4}{5}$, $\alpha_2 = \frac{2}{3}$, and $\alpha_3 = 0$. Also, we use $m = 15$ sensors for recovering $s = 10$ ground-truth sources (located at the angles -66.9° , -61.64° , -42.84° , -41.3° , -5.74° , -2.3° , 6.89° , 8.05° , 19.88° , 42.84°) and implement the optimization problem $\mathsf{P}_{1,2,\mathcal{D}\lambda}^\eta$ when λ is chosen optimally (i.e. λ_i^* 's are obtained using (8)), heuristically (i.e. $\lambda_i = \frac{1}{\alpha_i + \epsilon}$) and equally. As it turns out from Figure 3, while $\mathsf{P}_{1,2,\mathcal{D}\lambda}^\eta$ with equal and heuristic weights detects [respectively many and a few] non-existing sources with spurious DOAs, $\mathsf{P}_{1,2,\mathcal{D}\lambda}^\eta$ locates the ground-truth sources correctly. This in turn suggests that our optimal weighting strategy considerably decreases the required number of sensors.

APPENDIX

A. Preliminaries

In this section, we introduce two concepts from conic integral geometry that are used in our analysis.

Descent cone: Let \mathbf{x} be a vector with a special low-dimensional structure (e.g. block-sparsity). Assume that f is a convex function that promotes this structure. Then, the set of decent directions forms a convex set defined by:

$$\mathcal{D}(f, \mathbf{x}) = \bigcup_{t \geq 0} \{z \in \mathbb{C}^n : f(\mathbf{x} + tz) \leq f(\mathbf{x})\}. \quad (10)$$

Statistical dimension: Statistical dimension is intuitively a measure for the size of a cone. It is shown in [25] that statistical dimension of the above decent cone defined by

$$\delta(\mathcal{D}(f, \mathbf{x})) := \mathbb{E}_{\mathbf{g}} \text{dist}^2(\mathbf{g}, \mathcal{D}(f, \mathbf{x})^\circ), \quad (11)$$

specifies the required number of measurements (i.e. m) that

$$\mathsf{P}_f : \min_{z \in \mathbb{C}^n} f(z) \quad \text{s.t.} \quad \mathbf{y}_{m \times 1} = \mathbf{A}z \quad (12)$$

needs for perfect recovery. Here, $\mathbf{g} \in \mathbb{R}^n$ is a vector with i.i.d. standard normal distribution. We define the expected number of measurements needed for $\mathsf{P}_{1,2,w}$ as

$$\bar{m} := \mathbb{E}_{\mathbf{x}} \delta(\mathcal{D}(\|\cdot\|_{1,2,w}, \mathbf{x})). \quad (13)$$

Then, we call the weights that minimize \bar{m} optimal in sense of expected sample complexity.

B. Proof of Propositions 1 and 2

Proof. Since $\delta(\mathcal{D}(\|\cdot\|_{1,2,w}, \mathbf{x}))$ is upper-bounded by an expression that only depends on $\text{bsupp}(\mathbf{x})$ and \mathbf{w} (see Lemma 1), we show the corresponding upper-bound by $\delta(\mathcal{D}(\mathcal{B}, \mathbf{w}))$. It holds that:

$$\bar{m} \leq \mathbb{E}_{\mathbf{x}} \delta(\mathcal{D}(\text{bsupp}(\mathbf{x}), \mathbf{w})) := \mathbb{E}_{\mathbf{x}} \delta(\mathcal{D}(\mathcal{B}, \mathbf{w})) \quad (14)$$

We proceed by using a closed-form expression for $\delta(\mathcal{D}(\mathcal{B}, \mathbf{w}))$ a special case of which is obtained in [17, Lemma 2].

Lemma 1. *The statistical dimension of descent cone of any vector $\mathbf{x}_{n \times 1}$ with $\text{bsupp}(\mathbf{x}) := \mathcal{B}$ satisfies:*

$$\begin{aligned} \delta(\mathcal{D}(\|\cdot\|_{1,2,w}, \mathbf{x})) &\leq \delta(\mathcal{D}(\mathcal{B}, \mathbf{w})) := \\ \inf_{t \geq 0} &\left(\sum_{b \in \mathcal{B}} (k_b + (tw_b)^2) + \frac{\sum_{b \in \mathcal{B}} \phi_B(tw_b, k_b)}{2^{\frac{k_b}{2}-1} \Gamma(\frac{k_b}{2})} \right), \\ \text{with } \phi_B(z, k) &:= \int_z^\infty (u-z)^2 u^{k-1} \exp(-\frac{u^2}{2}) du. \end{aligned}$$

Moreover, the minimum is achieved at a unique $t \geq 0$. The inequality is in fact equality in the asymptotic case ($q \rightarrow \infty$) [17, Proposition 3].

Thus, by the aid of Lemma 1, it holds that

$$\begin{aligned} \bar{m} &\stackrel{(I)}{\leq} \mathbb{E} \inf_{t \geq 0} \sum_{b=1}^q \left((k_b + (tw_b)^2) 1_{b \in \mathcal{B}} + \frac{\phi_B(tw_b, k_b) 1_{b \in \mathcal{B}}}{2^{\frac{k_b}{2}-1} \Gamma(\frac{k_b}{2})} \right) \stackrel{(II)}{\leq} \\ \inf_{t \geq 0} &\sum_{b=1}^q \left((k_b + (tw_b)^2) \underbrace{\mathbb{E}[1_{b \in \mathcal{B}}]}_{p_b} + \frac{\phi_B(tw_b, k_b) \overbrace{\mathbb{E}[1_{b \in \mathcal{B}}]}^{1-p_b}}{2^{\frac{k_b}{2}-1} \Gamma(\frac{k_b}{2})} \right) \end{aligned} \quad (15)$$

where in the inequality (I), the notation $1_{\mathcal{E}}$ signifies the indicator function of an event \mathcal{E} . The inequality (II) results from the Jensen inequality for concave functions. Therefore, we have:

$$\bar{m} \leq \inf_{t \geq 0} \sum_{b=1}^q \left[p_b (k_b + t^2 w_b^2) + \frac{(1-p_b)}{2^{\frac{k_b}{2}-1} \Gamma(\frac{k_b}{2})} \phi_B(tw_b, k_b) \right]. \quad (16)$$

Moreover, by partitioning the set $[q]$ into the sets \mathcal{P}_i s, one may write:

$$\begin{aligned} \bar{m} &\leq \inf_{t \geq 0} \sum_{i=1}^L \sum_{b \in \mathcal{P}_i} \left((k_b + (tw_b)^2) p_b + \frac{(1-p_b) \phi_B(tw_b, k_b)}{2^{\frac{k_b}{2}-1} \Gamma(\frac{k_b}{2})} \right) = \\ &\sum_{b=1}^q k_b p_b + \inf_{t \geq 0} \sum_{i=1}^L \sum_{b \in \mathcal{P}_i} \left[t^2 \lambda_i^2 p_b + \frac{(1-p_b)}{2^{\frac{k_b}{2}-1} \Gamma(\frac{k_b}{2})} \phi_B(t\lambda_i, k_b) \right] \end{aligned} \quad (17)$$

where we used the relation (4) in the last equality. By further assumption $k_b = k \forall b$, it holds that:

$$\bar{m} \leq \sum_{b=1}^q k_b p_b + \inf_{t \geq 0} \sum_{i=1}^L \left[t^2 \lambda_i^2 \sum_{b \in \mathcal{P}_i} p_b + \frac{\phi_B(t\lambda_i, k) \sum_{b \in \mathcal{P}_i} (1-p_b)}{2^{\frac{k}{2}-1} \Gamma(\frac{k}{2})} \right]. \quad (18)$$

Since $\alpha_i := \frac{\mathbb{E}|\text{bsupp}(\mathbf{x}) \cap \mathcal{P}_i|}{|\mathcal{P}_i|} = \frac{\sum_{b \in \mathcal{P}_i} p_b}{|\mathcal{P}_i|}$ and $1 - \alpha_i = \frac{\mathbb{E}|\text{bsupp}(\mathbf{x}) \cap \mathcal{P}_i^c|}{|\mathcal{P}_i|} = \frac{\sum_{b \in \mathcal{P}_i^c} (1-p_b)}{|\mathcal{P}_i|}$, we have:

$$\bar{m} \leq \sum_{b=1}^q k_b p_b + \inf_{t \geq 0} \sum_{i=1}^L \left[t^2 \lambda_i^2 \alpha_i + \frac{1 - \alpha_i}{2^{\frac{k}{2}-1} \Gamma(\frac{k}{2})} \phi_B(t\lambda_i, k) \right]. \quad (19)$$

As multiplication with a positive scalar t keeps the optimal choices of w_b and λ_i in $\mathsf{P}_{1,2,w^*}$ and $\mathsf{P}_{1,2,\mathcal{D}\lambda^*}$ untouched, by minimizing the expressions in brackets in (16) and (19) with respect to w_b and λ_i (the second derivatives of these expressions are always positive with respect to w_b and λ_i and hence they are strictly convex functions), one can get to the expressions (7) and (8). ■

REFERENCES

- [1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

- [2] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution," *Communications on pure and applied mathematics*, vol. 59, no. 6, pp. 797–829, 2006.
- [3] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed dna microarrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 3, pp. 275–285, 2008.
- [4] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *arXiv preprint arXiv:0804.0041*, 2008.
- [5] T. Euler and T. Baden, "Computational neuroscience: Species-specific motion detectors," *Nature*, 2016.
- [6] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4692–4702, 2008.
- [7] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.
- [8] Y. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE transactions on signal processing*, vol. 56, no. 6, pp. 2334–2345, 2008.
- [9] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 993–1009, 2009.
- [10] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375–391, 2010.
- [11] K. V. Mishra, M. Cho, A. Kruger, and W. Xu, "Spectral super-resolution with prior knowledge," *IEEE transactions on signal processing*, vol. 63, no. 20, pp. 5342–5357, 2015.
- [12] D. Needell, R. Saab, and T. Woolf, "Weighted-minimization for sparse recovery under arbitrary prior information," *Information and Inference: A Journal of the IMA*, vol. 6, no. 3, pp. 284–309, 2017.
- [13] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 1985–2001, 2011.
- [14] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [15] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [16] S. Oymak, M. A. Khajehnejad, and B. Hassibi, "Recovery threshold for optimal weight ℓ_1 minimization," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 2032–2036, IEEE, 2012.
- [17] S. Daei, F. Haddadi, and A. Amini, "Exploiting prior information in block sparse signals," *arXiv preprint arXiv:1804.08444*, 2018.
- [18] W. Xu, *Compressive sensing for sparse approximations: constructions, algorithms, and analysis*. PhD thesis, California Institute of Technology, 2010.
- [19] S. Misra and P. A. Parrilo, "Weighted ℓ_1 -minimization for generalized non-uniform sparse model," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4424–4439, 2015.
- [20] M. Díaz, M. Junca, F. Rincón, and M. Velasco, "Compressed sensing of data with a known distribution," *Applied and Computational Harmonic Analysis*, 2017.
- [21] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse bayesian learning for recovery of block-sparse signals," *IEEE Transactions on Signal Processing*, vol. 63, no. 2, pp. 360–372, 2015.
- [22] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "k-t focuss: a general compressed sensing framework for high resolution dynamic mri," *Magnetic resonance in medicine*, vol. 61, no. 1, pp. 103–116, 2009.
- [23] 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, 2016.
- [24] M. M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed ℓ_0 norm approximation," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, 2010.
- [25] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 224–294, 2014.