On the Lagrangian Biduality of Sparsity Minimization Problems

Dheeraj Singaraju
Roberto Tron
Ehsan Elhamifar
Allen Yang
S. Shankar Sastry

Electrical Engineering and Computer Sciences
University of California at Berkeley

Report Documentation Page

1. REPORT DATE **03 JUN 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**
4. TITLE AND SUBTITLE **On the Lagrangian Biduality of Sparsity Minimization Problems**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California, Berkely,Electrical Engineering and Computer Sciences,Berkeley,CA,94720**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Recent results in Compressive Sensing have shown that, under certain conditions the solution to an underdetermined system of linear equations with sparsity-based regularization can be accurately recovered by solving convex relaxations of the original problem. In this work, we present a novel primal-dual analysis on a class of sparsity minimization problems. We show that the Lagrangian bidual (i.e. the Lagrangian dual of the Lagrangian dual) of the sparsity minimization problems can be used to derive interesting convex relaxations: the bidual of the '0- minimization problem is the '1-minimization problem; and the bidual of the '0;1- minimization problem for enforcing group sparsity on structured data is the '1;1- minimization problem. The analysis provides a means to compute per-instance non-trivial lower bounds on the (group) sparsity of the desired solutions. In a realworld application, the bidual relaxation improves the performance of a sparsitybased classification framework applied to robust face recognition.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**12**	

On the Lagrangian Biduality of Sparsity Minimization Problems

Dheeraj Singaraju, Allen Y. Yang, and S. Shankar Sastry[*]
Department of EECS
University of California, Berkeley
Berkeley, CA 94720
{dheeraj,yang,sastry}@eecs.berkeley.edu

Ehsan Elhamifar and Roberto Tron
Center for Imaging Science
Johns Hopkins University
Baltimore, MD 21218
{ehsan,tron}@cis.jhu.edu

Abstract

Recent results in Compressive Sensing have shown that, under certain conditions, the solution to an underdetermined system of linear equations with sparsity-based regularization can be accurately recovered by solving convex relaxations of the original problem. In this work, we present a novel primal-dual analysis on a class of sparsity minimization problems. We show that the Lagrangian bidual (i.e., the Lagrangian dual of the Lagrangian dual) of the sparsity minimization problems can be used to derive interesting convex relaxations: the bidual of the ℓ_0-minimization problem is the ℓ_1-minimization problem; and the bidual of the $\ell_{0,1}$-minimization problem for enforcing group sparsity on structured data is the $\ell_{1,\infty}$-minimization problem. The analysis provides a means to compute per-instance non-trivial lower bounds on the (group) sparsity of the desired solutions. In a real-world application, the bidual relaxation improves the performance of a sparsity-based classification framework applied to robust face recognition.

1 Introduction

The last decade has seen a renewed interest in the problem of solving an underdetermined system of equations $A\boldsymbol{x} = \boldsymbol{b}$, $A \in \mathbb{R}^{m \times n}, \boldsymbol{b} \in \mathbb{R}^m$, where $m << n$, by regularizing its solution to be sparse, i.e., having very few non-zero entries. Specifically, if one aims to find $\boldsymbol{x}$ with the least number of nonzero entries that solves the linear system, the problem is known as ℓ_0-minimization:

$$(P_0): \quad \boldsymbol{x}_0^* = \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{x}\|_0 \quad \text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}. \tag{1}$$

The problem (P_0) is intended to seek *entry-wise sparsity* in $\boldsymbol{x}$ and is known to be NP-hard in general. In Compressive Sensing (CS) literature, it has been shown that the solution to (1) often can be obtained by solving a more tractable linear program, namely, ℓ_1-minimization [4, 8]:

$$(P_1): \quad \boldsymbol{x}_1^* = \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}. \tag{2}$$

---

[*]Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

This unconventional equivalence relation between (P_0) and (P_1) and the more recent numerical solutions [3, 16] to efficiently recover high-dimensional sparse signal have been a very competitive research area in CS. Its broad applications have included sparse error correction [6], compressive imaging [23], image denoising and restoration [11, 17], and face recognition [13, 21], to name a few.

In addition to enforcing entry-wise sparsity in a linear system of equations, the notion of *group sparsity* has attracted increasing attention in recent years [12, 13, 18]. In this case, one assumes that the matrix A has some underlying structure, and can be grouped into blocks: $A = [A_1 \ \cdots \ A_K]$, where $A_k \in \mathbb{R}^{m \times d_k}$ and $\sum_{k=1}^{K} d_k = n$. Accordingly, the vector $\boldsymbol{x}$ is split into several blocks as $\boldsymbol{x}^\top = \begin{bmatrix} \boldsymbol{x}_1^\top & \ldots & \boldsymbol{x}_K^\top \end{bmatrix}$, where $\boldsymbol{x}_k \in \mathbb{R}^{d_k}$. In this case, it is of interest to estimate $\boldsymbol{x}$ with the least number of blocks containing non-zero entries. The group sparsity minimization problem is posed as

$$(P_{0,p}): \quad \boldsymbol{x}_{0,p}^* = \operatorname*{argmin}_{\boldsymbol{x}} \sum_{k=1}^{K} \mathcal{I}(\|\boldsymbol{x}_k\|_p > 0), \quad \text{s.t.} \quad A\boldsymbol{x} \doteq [A_1 \ \cdots \ A_K] \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} = \boldsymbol{b}, \quad (3)$$

where $\mathcal{I}(\cdot) \in \mathbb{R}$ is the indicator function. Since the expression $\sum_{k=1}^{K} \mathcal{I}(\|\boldsymbol{x}_k\|_p > 0)$ can be written as $\| [\|\boldsymbol{x}_1\|_p \ \cdots \ \|\boldsymbol{x}_K\|_p] \|_0$, it is also denoted as $\ell_{0,p}(\boldsymbol{x})$, the $\ell_{0,p}$-norm of $\boldsymbol{x}$.

Enforcing group sparsity exploits the problem's underlying structure and can improve the solution's interpretability. For example, in a sparsity-based classification (SBC) framework applied to face recognition, the columns of A are vectorized training images of human faces that can be naturally grouped into blocks corresponding to different subject classes, $\boldsymbol{b}$ is a vectorized query image, and the entries in $\boldsymbol{x}$ represent the coefficients of linear combination of all the training images for reconstructing $\boldsymbol{b}$. Group sparsity lends itself naturally to this problem since it is desirable to use images of the smallest number of subject classes to reconstruct and subsequently classify a query image.

Furthermore, the problem of robust face recognition has considered an interesting modification known as the *cross-and-bouquet* (CAB) model: $\boldsymbol{b} = A\boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{e} \in \mathbb{R}^m$ represents possible sparse error corruption on the observation $\boldsymbol{b}$ [22]. It can be argued that the CAB model can be solved as a group sparsity problem in (3), where the coefficients of $\boldsymbol{e}$ would be the $(K+1)^{\text{th}}$ group. However, this problem has a trivial solution for $\boldsymbol{e} = \boldsymbol{b}$ and $\boldsymbol{x} = \boldsymbol{0}$, which would have the smallest possible group sparsity. Hence, it is necessary to further regularize the entry-wise sparsity in $\boldsymbol{e}$.

To this effect, one considers a mixture of the previous two cases, where one aims to enforce entry-wise sparsity as well as group sparsity such that $\boldsymbol{x}$ has very few number of non-zero blocks *and* the reconstruction error $\boldsymbol{e}$ is also sparse. The *mixed sparsity* minimization problem can be posed as

$$(MP_{0,p}): \quad \{\boldsymbol{x}_{0,p}^*, \boldsymbol{e}_0^*\} = \operatorname*{argmin}_{(\boldsymbol{x}, \boldsymbol{e})} \ell_{0,p}(\boldsymbol{x}) + \gamma\|\boldsymbol{e}\|_0, \quad \text{s.t.} \quad [A_1 \ \cdots \ A_K] \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} = \boldsymbol{b} + \boldsymbol{e}, \quad (4)$$

where $\gamma \geq 0$ controls the tradeoff between the entry-wise sparsity and group sparsity.

Due to the use of the counting norm, the optimization problems in (3) and (4) are also NP-hard in general. Hence, several recent works have focused on developing tractable convex relaxations for these problems. In the case of group sparsity, the relaxation involves replacing the $\ell_{0,p}$-norm with the $\ell_{1,p}$-norm, where $\ell_{1,p}(\boldsymbol{x}) \doteq \left\| [\|\boldsymbol{x}_1\|_p \ \cdots \ \|\boldsymbol{x}_K\|_p] \right\|_1 = \sum_{k=1}^{K} \|\boldsymbol{x}_k\|_p$. These relaxations are also used for the mixed sparsity case [13].

In this work, we are interested in deriving and analyzing convex relaxations for general sparsity minimization problems. In the entry-wise case, the main theoretical understanding of the link between the original NP-hard problem in (1) and its convex relaxation has been given by the simple fact that the ℓ_1-norm is a convex surrogate of the ℓ_0-norm. However, in the group sparsity case, a similar relaxation produces a family of convex surrogates, i.e., $\ell_{1,p}(\boldsymbol{x})$, whose value depends on p. This raises the question whether there is a preferable value of p for the relaxation of the group sparsity minimization problem? In fact, we consider the following more important question:

Is there a unified framework for deriving convex relaxations of general sparsity recovery problems?

2

1.1 Paper contributions

We present a new optimization-theoretic framework based on Lagrangian duality for deriving convex relaxations of sparsity minimization problems. Specifically, we introduce a new class of equivalent optimization problems for (P_0), $(P_{0,p})$ and $(MP_{0,p})$, and derive the Lagrangian duals of the original NP-hard problems. We then consider the Lagrangian dual of the Lagrangian dual to get a new optimization problem that we term as the *Lagrangian bidual* of the primal problem. We show that the Lagrangian biduals are convex relaxations of the original sparsity minimization problems. Importantly, we show that the Lagrangian biduals for the (P_0) and $(P_{0,p})$ problems correspond to minimizing the ℓ_1-norm and the $\ell_{1,\infty}$-norm, respectively.

Since the Lagrangian duals for (P_0), $(P_{0,p})$ and $(MP_{0,p})$ are linear programs, there is no duality gap between the Lagrangian duals and the corresponding Lagrangian biduals. Therefore, the bidual based convex relaxations can be interpreted as maximizing the Lagrangian duals of the original sparsity minimization problems. This provides new interpretations for the relaxations of sparsity minimization problems. Moreover, since the Lagrangian dual of a minimization problem provides a lower bound for the optimal value of the primal problem, we show that the optimal objective value of the convex relaxation provides a non-trivial lower bound on the sparsity of the true solution to the primal problem.

2 Lagrangian biduals for sparsity minimization problems

In what follows, we will derive the Lagrangian bidual for the mixed sparsity minimization problem, which generalizes the entry-wise sparsity and group sparsity cases (also see Section 3). Specifically, we will derive the Lagrangian bidual for the following optimization problem:

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\text{argmin}} \sum_{k=1}^{K} \left[\alpha_k \mathcal{I}(\|\boldsymbol{x}_k\|_p > 0) + \beta_k \|\boldsymbol{x}_k\|_0 \right], \text{ s.t. } \begin{bmatrix} A_1 & \cdots & A_K \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_K \end{bmatrix} = \boldsymbol{b}, \quad (5)$$

where $\forall k = 1, \ldots, K : \alpha_k \geq 0$ and $\beta_k \geq 0$. Given any unique, finite solution $\boldsymbol{x}^*$ to (5), there exists a constant $M > 0$ such that the absolute values of the entries of $\boldsymbol{x}^*$ are less than M, namely, $\|\boldsymbol{x}^*\|_\infty \leq M$. Note that if (5) does not have a unique solution, it might not be possible to choose a finite-valued M that upper bounds all the solutions. In this case, a finite-valued M may be viewed as a regularization term for the desired solution. To this effect, we consider the following modified version of (5) where we introduce the box constraint that $\|\boldsymbol{x}\|_\infty \leq M$:

$$\boldsymbol{x}^*_{\text{primal}} = \underset{\boldsymbol{x}}{\text{argmin}} \sum_{k=1}^{K} \left[\alpha_k \mathcal{I}(\|\boldsymbol{x}_k\|_p > 0) + \beta_k \|\boldsymbol{x}_k\|_0 \right], \text{ s.t. } A\boldsymbol{x} = \boldsymbol{b} \text{ and } \|\boldsymbol{x}\|_\infty \leq M, \quad (6)$$

where M is chosen as described above to ensure that the optimal values of (6) and (5) are the same.

Primal problem. We will now frame an equivalent optimization problem for (6), for which we introduce some new notation. Let $\boldsymbol{z} \in \{0,1\}^n$ be an entry-based sparsity indicator for $\boldsymbol{x}$, namely, $z_i = 0$ if $x_i = 0$ and $z_i = 1$ otherwise. We also introduce a group-based sparsity indicator vector $\boldsymbol{g} \in \{0,1\}^K$, whose k^{th} entry g_k denotes whether the kth block $\boldsymbol{x}_k$ contains non-zero entries or not, namely, $g_k = 0$ if $\boldsymbol{x}_k = \boldsymbol{0}$ and $g_k = 1$ otherwise. To express this constraint, we introduce a matrix $\Pi \in \{0,1\}^{n \times K}$, such that $\Pi_{i,j} = 1$ if the ith entry of $\boldsymbol{x}$ belongs to the jth block and $\Pi_{i,j} = 0$ otherwise. Finally, we denote the positive component and negative component of $\boldsymbol{x}$ as $\boldsymbol{x}_+ \geq 0$ and $\boldsymbol{x}_- \geq 0$, respectively, such that $\boldsymbol{x} = \boldsymbol{x}_+ - \boldsymbol{x}_-$.

Given these definitions, we see that (6) can be reformulated as

$$\{\boldsymbol{x}^*_+, \boldsymbol{x}^*_-, \boldsymbol{z}^*, \boldsymbol{g}^*\} = \underset{\{\boldsymbol{x}_+, \boldsymbol{x}_-, \boldsymbol{z}, \boldsymbol{g}\}}{\text{argmin}} \left[\boldsymbol{\alpha}^\top \boldsymbol{g} + \boldsymbol{\beta}^\top \boldsymbol{z} \right], \text{ s.t. (a) } \boldsymbol{x}_+ \geq 0, \text{ (b) } \boldsymbol{x}_- \geq 0, \text{ (c) } \boldsymbol{g} \in \{0,1\}^K,$$

$$\text{(d) } \boldsymbol{z} \in \{0,1\}^n \text{ (e) } A(\boldsymbol{x}_+ - \boldsymbol{x}_-) = \boldsymbol{b}, \text{ (f) } \Pi \boldsymbol{g} \geq \frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-), \text{ and (g) } \boldsymbol{z} \geq \frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-), \quad (7)$$

where $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_k \end{bmatrix}^\top \in \mathbb{R}^k$ and $\boldsymbol{\beta} = \begin{bmatrix} \cdots & \underbrace{\beta_k \cdots \beta_k}_{d_k \text{ times}} & \cdots \end{bmatrix}^\top \in \mathbb{R}^n$.

3

Constraints (a)–(d) are used to enforce the aforementioned conditions on the values of the solution. While constraint (e) enforces the condition that the original system of linear equations is satisfied, the constraints (f) and (g) ensure that the group sparsity indicator $\boldsymbol{g}$ and the entry-wise sparsity indicator $\boldsymbol{z}$ are consistent with the entries of $\boldsymbol{x}$.

Lagrangial dual. The Lagrangian function for (7) is given as

$$L(\boldsymbol{x}_+, \boldsymbol{x}_-, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4, \boldsymbol{\lambda}_5) = \boldsymbol{\alpha}^\top \boldsymbol{g} + \boldsymbol{\beta}^\top \boldsymbol{z} - \boldsymbol{\lambda}_1^\top \boldsymbol{x}_+ - \boldsymbol{\lambda}_2^\top \boldsymbol{x}_- + \boldsymbol{\lambda}_3^\top (\boldsymbol{b} - A\boldsymbol{x}_+ + A\boldsymbol{x}_-)$$
$$+ \boldsymbol{\lambda}_4^\top (\frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-) - \Pi \boldsymbol{g}) + \boldsymbol{\lambda}_5^\top (\frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-) - \boldsymbol{z}),$$

(8)

where $\boldsymbol{\lambda}_1 \geq \mathbf{0}$, $\boldsymbol{\lambda}_2 \geq \mathbf{0}$, $\boldsymbol{\lambda}_4 \geq \mathbf{0}$, and $\boldsymbol{\lambda}_5 \geq \mathbf{0}$. In order to obtain the Lagrangian dual function, we need to minimize $L(\cdot)$ with respect to $\boldsymbol{x}_+$, $\boldsymbol{x}_-$, $\boldsymbol{g}$ and $\boldsymbol{z}$ [2]. Notice that if the coefficients of $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$, i.e., $\frac{1}{M}(\boldsymbol{\lambda}_4 + \boldsymbol{\lambda}_5) - A^\top \boldsymbol{\lambda}_3 - \boldsymbol{\lambda}_1$ and $\frac{1}{M}(\boldsymbol{\lambda}_4 + \boldsymbol{\lambda}_5) + A^\top \boldsymbol{\lambda}_3 - \boldsymbol{\lambda}_2$ are non-zero, the minimization of $L(\cdot)$ with respect to $\boldsymbol{x}_+$ and $\boldsymbol{x}_-$ is unbounded below. To this effect, the constraints that these coefficients are equal to 0 form constraints on the dual variables. Next, consider the minimization of $L(\cdot)$ with respect to $\boldsymbol{g}$. Since each entry g_k only takes values 0 or 1, its optimal value $\hat{g}_k$ that minimizes $L(\cdot)$ is given as

$$\hat{g}_k = \begin{cases} 0 & \text{if } \alpha_k - (\Pi^\top \boldsymbol{\lambda}_4)_k > 0, \text{ and} \\ 1 & \text{otherwise.} \end{cases}$$

(9)

A similar expression can be computed for the minimization with respect to $\boldsymbol{z}$. As a consequence, the Lagrangian dual problem can be derived as

$$\{\boldsymbol{\lambda}_i^*\}_{i=1}^5 = \underset{\{\boldsymbol{\lambda}_i\}_{i=1}^5}{\operatorname{argmax}} \big[\boldsymbol{\lambda}_3^\top \boldsymbol{b} + \mathbf{1}^\top \min\{\mathbf{0}, \boldsymbol{\alpha} - \Pi^\top \boldsymbol{\lambda}_4\} + \mathbf{1}^\top \min\{\mathbf{0}, \boldsymbol{\beta} - \boldsymbol{\lambda}_5\} \big], \text{s.t.}$$

$$(a) \ \forall i = 1, 2, 4, 5 : \boldsymbol{\lambda}_i \geq \mathbf{0}, \ (b) \ \frac{1}{M}(\boldsymbol{\lambda}_4 + \boldsymbol{\lambda}_5) - A^\top \boldsymbol{\lambda}_3 - \boldsymbol{\lambda}_1 = \mathbf{0} \qquad (10)$$

$$\text{and (c)} \ \frac{1}{M}(\boldsymbol{\lambda}_4 + \boldsymbol{\lambda}_5) + A^\top \boldsymbol{\lambda}_3 - \boldsymbol{\lambda}_2 = \mathbf{0}.$$

This can be further simplified by rewriting it as the following linear program:

$$\{\boldsymbol{\lambda}_i^*\}_{i=3}^7 = \underset{\{\boldsymbol{\lambda}_i\}_{i=3}^7}{\operatorname{argmax}} \big[\boldsymbol{\lambda}_3^\top \boldsymbol{b} + \mathbf{1}^\top \boldsymbol{\lambda}_6 + \mathbf{1}^\top \boldsymbol{\lambda}_7 \big], \text{s.t. (a) } \boldsymbol{\lambda}_4 \geq \mathbf{0}, \text{ (b) } \boldsymbol{\lambda}_5 \geq \mathbf{0}, \text{ (c) } \boldsymbol{\lambda}_6 \leq \mathbf{0}, \text{ (d) } \boldsymbol{\lambda}_7 \leq \mathbf{0},$$

(11)

$$\text{(e) } \boldsymbol{\lambda}_6 \leq \boldsymbol{\alpha} - \Pi^\top \boldsymbol{\lambda}_4, \text{ (f) } \boldsymbol{\lambda}_7 \leq \boldsymbol{\beta} - \boldsymbol{\lambda}_5 \text{ and (g) } -\frac{1}{M}(\boldsymbol{\lambda}_4 + \boldsymbol{\lambda}_5) \leq A^\top \boldsymbol{\lambda}_3 \leq \frac{1}{M}(\boldsymbol{\lambda}_4 + \boldsymbol{\lambda}_5).$$

Notice that we have made two changes in going from (10) to (11). First, we have replaced constraints (b) and (c) in (10) with the constraint (g) in (11) and eliminated $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ from (11). Second, we have introduced variables $\boldsymbol{\lambda}_6$ and $\boldsymbol{\lambda}_7$ to encode the "min" operator in the objective function of (10).

Lagrangian bidual. We will now consider the Lagrangian dual of (11), which will be referred to as the *Lagrangian bidual* of (7). It can be verified that the Lagrangian dual of (11) is given as

$$\{\boldsymbol{x}_+^*, \boldsymbol{x}_-^*, \boldsymbol{z}^*, \boldsymbol{g}^*\} = \underset{\{\boldsymbol{x}_+, \boldsymbol{x}_-, \boldsymbol{z}, \boldsymbol{g}\}}{\operatorname{argmin}} \boldsymbol{\alpha}^\top \boldsymbol{g} + \boldsymbol{\beta}^\top \boldsymbol{z} \quad \text{s.t. (a) } \boldsymbol{x}_+ \geq \mathbf{0}, \text{ (b) } \boldsymbol{x}_- \geq \mathbf{0}, \text{ (c) } \boldsymbol{g} \in [0,1]^K,$$

(12)

$$\text{(d) } \boldsymbol{z} \in [0,1]^n \text{ (e) } A(\boldsymbol{x}_+ - \boldsymbol{x}_-) = \boldsymbol{b}, \text{ (f) } \Pi \boldsymbol{g} \geq \frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-) \text{ and (g) } \boldsymbol{z} \geq \frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-).$$

Notice that in going from (7) to (12), the discrete valued variables $\boldsymbol{z}$ and $\boldsymbol{g}$ have been relaxed to take real values between 0 and 1. Given that $\boldsymbol{z} \leq \mathbf{1}$ and noting that $\boldsymbol{x}$ can be represented as $\boldsymbol{x} = \boldsymbol{x}_+ - \boldsymbol{x}_-$, we can conclude from constraint (g) in (12) that the solution $\boldsymbol{x}^*$ satisfies $\|\boldsymbol{x}^*\|_\infty \leq M$. Moreover, given that $\boldsymbol{g}$ and $\boldsymbol{z}$ are relaxed to take real values, we see that the optimal values for g_k^* and z_i^* are $\frac{1}{M}\|\boldsymbol{x}_k^*\|_\infty$ and $\frac{1}{M}|x_i^*|$, respectively. Hence, we can eliminate constraints (f) and (g) by replacing $\boldsymbol{z}$ and $\boldsymbol{g}$ by these optimal values. It can then be verified that solving (12) is equivalent to solving the problem:

$$\boldsymbol{x}_{\text{bidual}}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{M} \sum_{k=1}^K \big[\alpha_k \|\boldsymbol{x}_k\|_\infty + \beta_k \|\boldsymbol{x}_k\|_1 \big] \quad \text{s.t. (a) } A\boldsymbol{x} = \boldsymbol{b} \text{ and (b) } \|\boldsymbol{x}\|_\infty \leq M. \quad (13)$$

This is the Lagrangian bidual for (7).

3 Theoretical results from the biduality framework

In this section, we first describe some properties of the biduality framework in general. We will then focus on some important results for the special cases of entry-wise sparsity and group sparsity.

Theorem 1. *The optimal value of the Lagrangian bidual in* (13) *is a lower bound on the optimal value of the NP-hard primal problem in* (7).

Proof. Since there is no duality gap between a linear program and its Lagrangian dual [2], the optimal values of the Lagrangian dual in (11) and the Lagrangian bidual in (13) are the same. Moreover, we know that the optimal value of a primal minimization problem is always bounded below by the optimal value of its Lagrangian dual [2]. We hence have the required result. □

Remark 1. *Since the original primal problem in* (7) *is NP-hard, we note that the duality gap between the primal and its dual in* (11) *is non-zero in general. Moreover, we notice that as we increase M (i.e., a more conservative estimate), the optimal value of the primal is unchanged, but the optimal value of the bidual in* (13) *decreases. Hence, the duality gap increases as M increases.*

M in (6) should preferably be equal to $\|x^*_{\text{primal}}\|_\infty$, which may not be possible to estimate accurately in practice. Therefore, it is of interest to analyze the effect of taking a very conservative estimate of M, i.e., choosing a large value for M. In what follows, we show that taking a conservative estimate of M is equivalent to dropping the box constraint in the bidual.

For this purpose, consider the following modification of the bidual:

$$x^*_{\text{bidual-conservative}} = \underset{x}{\arg\min} \sum_{k=1}^K \left[\alpha_k \|x_k\|_\infty + \beta_k \|x_k\|_1 \right] \quad \text{s.t.} \quad Ax = b, \tag{14}$$

where we have essentially dropped the box constraint (b) in (13). It is easy to verify that $\forall M \geq \max\{\|x^*_{\text{primal}}\|_\infty, \|x^*_{\text{bidual-conservative}}\|_\infty\}$, we have that $x^*_{\text{bidual}} = x^*_{\text{bidual-conservative}}$. Therefore, we see that taking a conservative value of M is equivalent to solving the modified bidual in (14).

3.1 Results for entry-wise sparsity minimization

Notice that by substituting $\alpha_1 = \cdots = \alpha_K = 0$ and $\beta_1 = \cdots = \beta_K = 1$, the optimization problem in (5) reduces to the entry-wise sparsity minimization problem in (1). Hence, the Lagrangian bidual to the M-regularized entry-wise sparsity problem (P_0) is:

$$x^*_{\text{entry-wise-bidual}} = \underset{x}{\arg\min} \frac{1}{M} \|x\|_1 \quad \text{s.t. (a) } Ax = b \text{ and (b) } \|x\|_\infty \leq M. \tag{15}$$

More importantly, we can also conclude from (14) that solving the Lagrangian bidual to the entry-wise sparsity problem with a conservative estimate of M is equivalent to solving the problem:

$$x^*_{\text{entry-wise-bidual-conservative}} = \underset{x}{\arg\min} \|x\|_1 \quad \text{s.t.} \quad Ax = b, \tag{16}$$

which is precisely the well-known ℓ_1-norm relaxation for (P_0). Our framework therefore provides a new interpretation for this relaxation:

Remark 2. *The ℓ_1-norm minimization problem in* (16) *is the Lagrangian bidual of the ℓ_0-norm minimization problem in* (1), *and solving* (16) *is equivalent to maximizing the dual of* (1).

We further note that we can now use the solution of (15) to derive a non-trivial lower bound for the primal objective function which is precisely the sparsity of the desired solution. More specifically, we can use Theorem 1 to conclude the following result:

Corollary 1. *Let x^*_0 be the solution to* (1). *We have that $\forall M \geq \|x^*_0\|_\infty$, the sparsity of x^*_0, i.e., $\|x^*_0\|_0$ is bounded below by $\frac{1}{M}\|x^*_{\text{entry-wise-bidual}}\|_1$.*

Due to the non-zero duality gap in the primal entry-wise sparsity minimization problem, the above lower bound provided by Corollary 1 is not tight in general.

3.2 Results for group sparsity minimization

Notice that by substituting $\alpha_1 = \cdots = \alpha_K = 1$ and $\beta_1 = \cdots = \beta_K = 0$, the optimization problem in (5) reduces to the group sparsity minimization problem in (3). Hence, the Lagrangian bidual of the group sparsity problem is:

$$\boldsymbol{x}^*_{\text{group-bidual}} = \operatorname*{argmin}_{\boldsymbol{x}} \frac{1}{M} \sum_{k=1}^{K} \|\boldsymbol{x}_k\|_\infty \quad \text{s.t.} \quad \text{(a) } A\boldsymbol{x} = \boldsymbol{b} \text{ and (b) } \|\boldsymbol{x}\|_\infty \leq M. \tag{17}$$

As in the case of entry-wise sparsity above, solving the bidual to the group sparsity problem with a conservative estimate of M is equivalent to solving:

$$\boldsymbol{x}^*_{\text{group-bidual-conservative}} = \operatorname*{argmin}_{\boldsymbol{x}} \sum_{k=1}^{K} \|\boldsymbol{x}_k\|_\infty \quad \text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}, \tag{18}$$

which is the convex $\ell_{1,\infty}$-norm relaxation of the $\ell_{0,p}$-min problem (3). In other words, the biduality framework selects the $\ell_{1,\infty}$-norm out of the entire family of $\ell_{1,p}$-norms as the convex surrogate of the $\ell_{0,p}$-norm.

Finally, we use Theorem 1 to show that the solution obtained by minimizing the $\ell_{1,\infty}$-norm provides a lower bound for the group sparsity.

Corollary 2. *Let $\boldsymbol{x}^*_{0,p}$ be the solution to (3). For any $M \geq \|\boldsymbol{x}^*_{0,p}\|_\infty$, the group sparsity of $\boldsymbol{x}^*_{0,p}$, i.e., $\ell_{0,p}(\boldsymbol{x}^*_{0,p})$, is bounded below by $\frac{1}{M}\ell_{1,\infty}(\boldsymbol{x}^*_{\text{group-bidual}})$.*

The $\ell_{1,\infty}$-norm seems to be an interesting choice for computing the lower bound of the group sparsity, as compared to other $\ell_{1,p}$-norms for finite $p < \infty$. For example, consider the case when $p = 1$, where the $\ell_{1,p}$-norm is equivalent to the ℓ_1-norm. Assume that A consists of a single block with several columns so that the maximum number of non-zero blocks is 1. Denote the solution to the ℓ_1-minimization problem as $\boldsymbol{x}^*_1$. It is possible to construct examples (also see Figure 1) where $\frac{1}{M}\ell_{1,1}(\boldsymbol{x}^*_1) = \frac{1}{M}\ell_1(\boldsymbol{x}^*_1) > 1$. Hence, it is unclear in general if the solutions obtained by minimizing $\ell_{1,p}$-norms for finite-valued $p < \infty$ can help provide lower bounds for the group sparsity.

4 Experiments

We now present experiments to evaluate the bidual framework for minimizing entry-wise sparsity and mixed sparsity. We present experiments on synthetic data to show that our framework can be used to compute non-trivial lower bounds for the entry-wise sparsity minimization problem. We then consider the face recognition problem where we compare the performance of the bidual-based $\ell_{1,\infty}$-norm relaxation with that of the $\ell_{1,2}$-norm relaxation for mixed sparsity minimization.

We use boxplots to provide a concise representation of our results' statistics. The top and bottom edge of a boxplot for a set of values indicates the maximum and minimum of the values. The bottom and top extents of the box indicate the 25 and 75 percentile mark. The red mark in the box indicates the median and the red crosses outside the boxes indicate potential outliers.

Entry-wise sparsity. We now explore the practical implications of Corollary 1 through synthetic experiments. We randomly generate entries of $A \in \mathbb{R}^{128 \times 256}$ and $\boldsymbol{x}_0 \in \mathbb{R}^{256}$ from a Gaussian distribution with unit variance. The sparsity of $\boldsymbol{x}_0$ is varied from 1 to 64 in steps of 3. We solve (15) with $\boldsymbol{b} = A\boldsymbol{x}_0$ using $M = M_0, 2M_0$ and $5M_0$, where $M_0 = \|\boldsymbol{x}_0\|_\infty$. We use Corollary 1 to compute lower bounds on the true sparsity, i.e., $\|\boldsymbol{x}_0\|_0$. We repeat this experiment 1000 times for each sparsity level and Figure 1 shows the boxplots for the bounds computed from these experiments.

We first analyze the lower bounds computed when $M = M_0$, in Figure 1(a). As explained in Section 3.1, the bounds are not expected to be tight due to the duality gap. Notice that for extremely sparse solutions, the maximum of the computed bounds is close to the true sparsity but this diverges as the sparsity of $\boldsymbol{x}_0$ reduces. The median value of the bounds is much looser and we see that the median also diverges as the sparsity of $\boldsymbol{x}_0$ reduces. Furthermore, the computed lower bounds seem to grow linearly as a function of the true sparsity. Similar trends are observed for $M = 2M_0$ and $5M_0$ in Figures 1(b) and 1(c), respectively. As expected from the discussion in Section 3.1, the bounds become very loose as M increases.
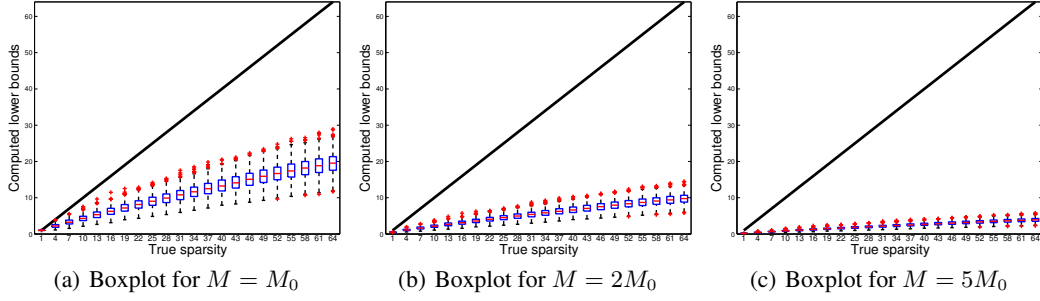
| (a) Boxplot for $M = M_0$ | (b) Boxplot for $M = 2M_0$ | (c) Boxplot for $M = 5M_0$ |

Figure 1: Results for computing the lower bounds on the true (black lines) entry-wise sparsity $\|\boldsymbol{x}_0\|_0$ obtained over 1000 trials. The bounds are computed by solving (15) and using Corollary 1 with $M = M_0, 2M_0$ and $5M_0$, where $M_0 = \|\boldsymbol{x}_0\|_\infty$. Notice that as expected from the discussion in Section 3.1, the bounds are not tight due to the duality gap and become looser as M increases.

In theory, we would like to have *per-instance certificates-of-optimality* of the computed solution, where the lower bound is equal to the true sparsity $\|\boldsymbol{x}_0\|_0$. Nonetheless, we note that this ability to compute a per-instance non-trivial lower bound on the sparsity of the desired solution is an important step forward with respect to the previous approaches that require pre-computing optimality conditions for equivalence of solutions to the ℓ_0-norm and ℓ_1-norm minimization problems.

We have performed a similar experiment for the group sparsity case, and observed that the bidual framework is able to provide non-trivial lower bounds for the group sparsity also.

Mixed sparsity. We now evaluate the results of mixed sparsity minimization for the sparsity-based face recognition problem, where the columns of A represent training images from the K face classes: $A_1, \cdots, A_K$ and $\boldsymbol{b} \in \mathbb{R}^m$ represents a query image. We assume that a subset of pixel values in the query image may be corrupted or disguised. Hence, the error in the image space is modeled by a sparse error term $\boldsymbol{e}$: $\boldsymbol{b} = \boldsymbol{b}_0 + \boldsymbol{e}$, where $\boldsymbol{b}_0$ is the uncorrupted image. A linear representation of the query image forms the following linear system of equations:

$$\boldsymbol{b} = A\boldsymbol{x} + \boldsymbol{e} = \begin{bmatrix} A_1 & \cdots & A_K & I \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1^\top & \cdots & \boldsymbol{x}_K^\top & \boldsymbol{e}^\top \end{bmatrix}^\top, \tag{19}$$

where I is the $m \times m$ identity matrix. The goal of sparsity-based classification (SBC) is to minimize the group sparsity in $\boldsymbol{x}$ and the sparsity of $\boldsymbol{e}$ such that the dominant non-zero coefficients in $\boldsymbol{x}$ reveal the membership of the ground-truth observation $\boldsymbol{b}_0 = \boldsymbol{b} - \boldsymbol{e}$ [13, 21]. In our experiments, we solve for $\boldsymbol{x}$ and $\boldsymbol{e}$ by solving the following optimization problem:

$$\{\boldsymbol{x}_{1,\mathrm{p}}^*, \boldsymbol{e}_1^*\} = \operatorname*{argmin}_{\{\boldsymbol{x},\boldsymbol{e}\}} \sum_{k=1}^K \|\boldsymbol{x}_k\|_p + \gamma \|\boldsymbol{e}\|_1 \quad \text{s.t.} \quad A\boldsymbol{x} + \boldsymbol{e} = \boldsymbol{b}. \tag{20}$$

Notice that for $p = \infty$, this reduces to solving a special case of the problem in (14), i.e., the bidual relaxation of the mixed sparsity problem with a conservative estimate of M. In our experiments, we set $\gamma = 0.01$ and compare the solutions to (20) obtained using $p = 2$ and $p = \infty$.

We evaluate the algorithms on a subset of the AR dataset [1] which has manually aligned frontal face images of size 83×60 for 50 male and 50 female subjects, i.e., $K = 100$ and $m = 4980$. Each individual contributes 7 un-occluded training images, 7 un-occluded testing images and 12 occluded testing images. Hence, we have 700 training images and 1900 testing images. To compute the number of non-zero blocks in the coefficient $\boldsymbol{x}$ estimated for a testing image, we find the number of blocks whose energy $\ell_2(\boldsymbol{x}_k)$ is greater than a specified threshold.

The results of our experiments are presented in Figure 2. The solution obtained with $p = 2$ gives better group sparsity of $\boldsymbol{x}$. However, a sparser error $\boldsymbol{e}$ is estimated with $p = \infty$. The number of non-zero entities in a solution to (20), i.e., the number of non-zero blocks plus the number of non-zero error entries, is lower for the solution obtained using $p = \infty$ rather than that obtained using $p = 2$. However, the primal mixed-sparsity objective value $\ell_{0,p}(\boldsymbol{x}) + \gamma \|\boldsymbol{e}\|_0$ (see (4)) is lower for the solution obtained using $p = 2$.

(a) Group sparsity for $p = 2$ (b) Group sparsity for $p = \infty$ (c) Difference in group sparsities

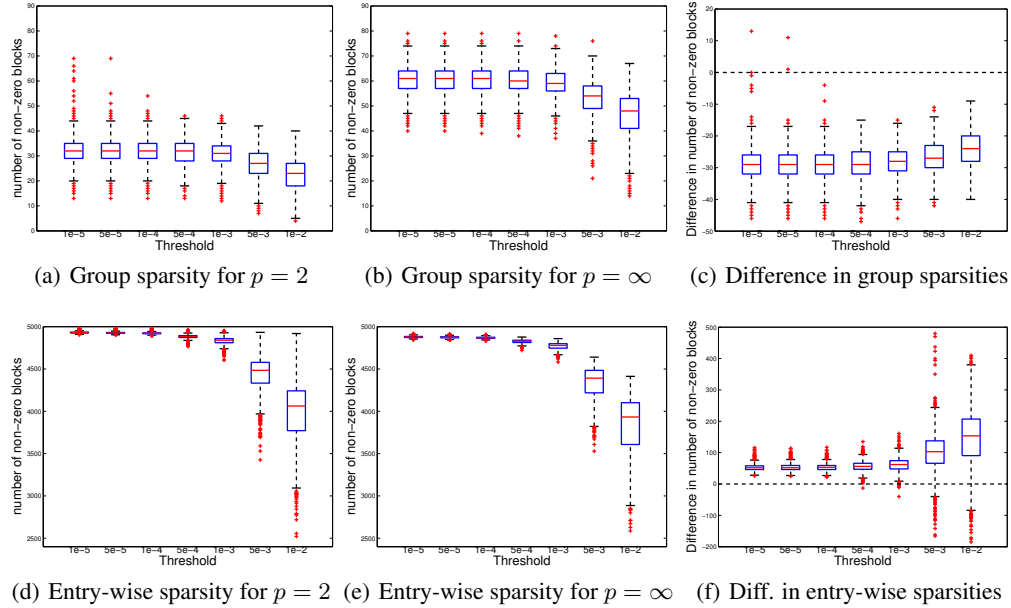(d) Entry-wise sparsity for $p = 2$ (e) Entry-wise sparsity for $p = \infty$ (f) Diff. in entry-wise sparsities

Figure 2: Comparison of mixed sparsity of the solutions to (20) for $p = 2$ and $p = \infty$. We present boxplots for group sparsity of $\boldsymbol{x}$ and entry-wise sparsity of $\boldsymbol{e}$. The differences are calculated as (# non-zero blocks/elements for $p = 2$) - (# non-zero blocks/elements for $p = \infty$). We see that for $p = 2$ we get better group sparsity of $\boldsymbol{x}$, but we get a more sparse error $\boldsymbol{e}$ when we use $p = \infty$.

	$p = 2$		$p = \infty$	
	#(correct results)	%(correct results)	#(correct results)	%(correct results)
un-occluded	655	93.57%	663	94.71%
occluded	643	53.58%	691	57.58%
total	1298	68.32%	1324	69.68%

Table 1: Classification results on the AR dataset using the solutions obtained by minimizing mixed sparsity. The test set consists of 700 un-occluded images and 1200 occluded images.

We now compare the classification results obtained with the solutions $\boldsymbol{x}$ computed in our experiments. For classification, we consider the non-zero blocks in $\boldsymbol{x}$ and then assign the query image to the block, i.e., subject class, for which it gives the least ℓ_2 residual $\|\boldsymbol{b} - A_k \boldsymbol{x}_k\|_2$. The results are presented in Table 1. Notice that the classification results obtained with $p = \infty$ (the bidual relaxation) are better than those obtained using $p = 2$. Since the classification of un-occluded images is already very good using $p = 2$, classification with $p = \infty$ gives only a minor improvement in this case. However, a more tangible improvement is noticed in the classification of the occluded images. Therefore the classification with $p = \infty$ is in general better than that obtained with $p = 2$, which is considered the state-of-the-art for sparsity-based classification [13].

5 Discussion

We have presented a novel analysis of several sparsity minimization problems which allows us to interpret several convex relaxations of the original NP-hard primal problems as being equivalent to maximizing their Lagrangian duals. The pivotal point of this analysis is the formulation of mixed-integer programs which are equivalent to the original primal problems. While we have derived the biduals for only a few sparsity minimization problems, the same techniques can also be used to easily derive convex relaxations for other sparsity minimization problems [7].

An interesting result of our biduality framework is the ability to compute a per-instance certificate of optimality by providing a lower bound for the primal objective function. This is in contrast to most

previous research which aims to characterize either the subset of solutions or the set of conditions for perfect sparsity recovery using the convex relaxations [5, 6, 8–10, 14, 15, 20]. In most cases, the conditions are either weak or hard to verify. More importantly, these conditions needed to be pre-computed as opposed to verifying the correctness of a solution at run-time. In lieu of this, we hope that our proposed framework will prove an important step towards per-instance verification of the solutions. Specifically, it is of interest in the future to explore tighter relaxations for the verification of the solutions.

Acknowledgments

References

[1] A. M. Martinez, and, R. Benavente, "The AR face database." CVC Technical Report #24, 1998.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[3] A. Bruckstein, D. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, 2009.

[4] E. Candès. Compressive sampling. In *Proceedings of International Congress of Mathematics*, 2006.

[5] E. Candés. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci., Paris, Series I*, 346: 589–592, 2008.

[6] E. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory,* 51(12):4203–4215, 2005.

[7] V. Cevher, M. F. Duarte, C. Hegde, R. G.. Baraniuk. Sparse signal recovery using Markov random fields. In *NIPS* 257–264, 2008.

[8] D. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *PNAS*, 100(5):2197–2202, 2003.

[9] D. Donoho. For most large underdetermined systems of linear equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution. *Communications on Pure and Applied Mathematics*. 2006.

[10] D. Donoho, and M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. on Information Theory,* 52(1): 6–18, 2006.

[11] M. Protter, M. Elad. Image sequence denoising via sparse and redundant representations. *IEEE Transactions on Image Processing* 18(1): 27–35, 2009.

[12] Y. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Transactions on Information Theory,* 55(11):5302–5316, 2009.

[13] E. Elhamifar and R. Vidal. Robust classification using structured sparse representation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[14] A. Fletcher, S. Rangan, and V. Goyal. Necessary and sufficient conditions on sparsity pattern recovery. *IEEE Transactions on Information Theory,* 55(12): 5758-5772 (2009)

[15] A. Iouditski, F. K. Karzan, and A. Nemirovski. Verifiable conditions of ℓ_1-recovery of sparse signals with sign restrictions. *ArXiv e-prints*, 2009.

[16] I. Loris. On the performance of algorithms for the minimization of ℓ_1-penalized functionals. *Inverse Problems*, 25:1–16, 2009.

[17] J. Mairal, M. Elad and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing* 17(1): 53-69, 2008.

[18] M. Stojnic, F. Parvaresh, B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Transactions on Signal Processing* 57(8): 3075–3085, 2009.

[19] J. Tropp. Greed is Good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory* 50(10): 2231–2242, 2004.

[20] G. Reeves and M. Gastpar. Efficient sparsity pattern recovery. In *Proc. 30th Symp. on Information Theory Benelux*, 2009.

[21] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[22] J. Wright and Y. Ma. Dense Error Correction via ℓ_1-Minimization. *IEEE Transactions on Information Theory*, 56(7), 2010.

[23] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly and R. Baraniuk. Compressive imaging for video representation and coding. In *Picture Coding Symposium*, 2006.