ON THE LAGRANGIAN BIDUALITY OF SPARSITY MINIMIZATION PROBLEMS

Dheeraj Singaraju, Roberto Tron†, Ehsan Elhamifar†, Allen Y. Yang*, and S. Shankar Sastry**

*University of California, Berkeley , CA 94720, USA
†Center for Imaging Science, Johns Hopkins University, Baltimore MD 21218, USA

ABSTRACT

We present a novel primal-dual analysis on a class of NP-hard sparsity minimization problems to provide new interpretations for their well known convex relaxations. 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 ℓ_1-minimization; and the bidual of $\ell_{0,1}$-minimization for enforcing group sparsity on structured data is $\ell_{1,\infty}$-minimization problem. Intuitions from the bidual-based relaxation are used to introduce a new family of relaxations for the group sparsity minimization problem.

Index Terms— Lagrangian biduality, group sparsity

1. INTRODUCTION

The ℓ_0-minimization problem aims to find the solution to an underdetermined system of equations $A\boldsymbol{x} = \boldsymbol{b}$, $A \in \mathbb{R}^{m\times n}, \boldsymbol{b} \in \mathbb{R}^m$, where $m \ll n$, by regularizing its solution to be sparse, i.e., having very few non-zero entries, as

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

The problem (P_0) is intended to seek *entry-wise sparsity* in $\boldsymbol{x}$ and is known to be NP-hard in general. This problem has found broad applications in error decoding, image denoising, face recognition and subspace clustering, to name a few.

In recent years, the notion of *group sparsity* has attracted increasing attention. In this case, one assumes that the matrix A has some underlying structure and can be grouped into blocks: $A = \begin{bmatrix}A_1, \cdots, A_K\end{bmatrix}$, where $A_k \in \mathbb{R}^{m\times n_k}$ and $\sum_{k=1}^K n_k = n$. Accordingly, the vector $\boldsymbol{x}$ is split into several blocks as $\boldsymbol{x} = \begin{bmatrix}\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_K^\top\end{bmatrix}^\top$, where $\boldsymbol{x}_k \in \mathbb{R}^{n_k}$. In this case, it is of interest to estimate $\boldsymbol{x}$ with the least number of blocks containing non-zero entries by solving the problem:

$$\begin{aligned}(P_{0,p}): \quad \boldsymbol{x}_{0,p}^* = \operatorname*{argmin}_{\boldsymbol{x}} \sum_{k=1}^K \mathcal{I}(\|\boldsymbol{x}_k\|_p > 0), \text{s.t.} \\ A\boldsymbol{x} \doteq \begin{bmatrix}A_1, \cdots, A_K\end{bmatrix} \begin{bmatrix}\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_K^\top\end{bmatrix}^\top = \boldsymbol{b},\end{aligned} \tag{2}$$

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

Group sparsity lends itself naturally to applications such as face recognition using sparse representation [8], where the columns of A are vectorized images of human faces that can be grouped into blocks of different subjects. Furthermore, the problem of robust face recognition considers an interesting modification: $\boldsymbol{b} = A\boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{e} \in \mathbb{R}^m$ represents sparse error corruption on the observation $\boldsymbol{b}$ [9]. It can be argued that this model can be solved as a group sparsity problem in (2), 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 $\boldsymbol{x} = \boldsymbol{0}$. Hence, one should consider the following *mixed sparsity* minimization problem where $\boldsymbol{x}$ has very few number of non-zero blocks *and* the error $\boldsymbol{e}$ is entry-wise sparse:

$$\begin{aligned}(MP_{0,p}): \{\boldsymbol{x}_{0,p}^*, \boldsymbol{e}_0^*\} = \operatorname*{argmin}_{(\boldsymbol{x},\boldsymbol{e})} \ell_{0,p}(\boldsymbol{x}) + \gamma\|\boldsymbol{e}\|_0, \text{s.t.} \\ \begin{bmatrix}A_1, \cdots, A_K\end{bmatrix} \begin{bmatrix}\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_K^\top\end{bmatrix}^\top + \boldsymbol{e} = \boldsymbol{b},\end{aligned} \tag{3}$$

where $\gamma \geq 0$ is a tradeoff parameter.

Due to the use of the ℓ_0-norm, the optimization problems in (1)–(3) are all NP-hard in general. Recent works have focused on developing tractable convex relaxations for these problems. The relaxation for entry-wise sparsity replaces the ℓ_0-norm with the ℓ_1-norm [4]. Relaxations for group sparsity replace the $\ell_{0,p}$-norm with the $\ell_{1,p}$-norm, where $\ell_{1,p}(\boldsymbol{x}) \doteq \left\|\begin{bmatrix}\|\boldsymbol{x}_1\|_p & \cdots & \|\boldsymbol{x}_K\|_p\end{bmatrix}\right\|_1 = \sum_{k=1}^K \|\boldsymbol{x}_k\|_p$. These relaxations are also used for the mixed sparsity case [6].

Paper contributions. In this work, we present a new framework for analyzing convex relaxations of the problems in (1)–(3). These relaxations have previously been analyzed from several different viewpoints such as convex polytope theory [5] and submodularity [1]. We present a novel optimization-theoretic interpretation based on Lagrangian duality.

We introduce a new class of equivalent optimization problems for (P_0), $(P_{0,p})$ and $(MP_{0,p})$, and derive their Lagrangian duals. We then consider the Lagrangian dual of the Lagrangian dual to get a new optimization problem called the *Lagrangian bidual* of the primal problem. We show that the Lagrangian biduals are convex relaxations of the original sparsity minimization problems. Importantly, the derived 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.

We note that there are other works on using Lagrangian biduality to derive relaxations of sparsity minimization problems [3, 7]. The work most related to this paper is that of [3], which derives a semi-definite program (SDP) as the bidual for the problem P_0. In contrast, we derive the well-known ℓ_1-minimization problem to be the bidual for the problem P_0, which is a linear program and easier to solve than an SDP.

2. LAGRANGIAN BIDUALITY

In what follows, we will derive the Lagrangian bidual for:

$$\begin{aligned} \boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} \sum_{k=1}^{K} \big[\alpha_k \mathcal{I}(\|\boldsymbol{x}_k\|_p > 0) + \beta_k \|\boldsymbol{x}_k\|_0\big], \\ \text{s.t. } \big[A_1, \cdots, A_K\big] \big[\boldsymbol{x}_1^\top, \cdots, \boldsymbol{x}_K\big]^\top = \boldsymbol{b}, \end{aligned} \tag{4}$$

where $\forall k = 1, \ldots, K : \alpha_k \geq 0$ and $\beta_k \geq 0$. This optimization problem generalizes the entry-wise sparsity and group sparsity cases (also see Section 3). Given any unique, finite solution $\boldsymbol{x}^*$ to (4), 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 (4) does not have a unique solution, it may 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. For such M, we consider the following modified version of (4) where we introduce the box constraint that $\|\boldsymbol{x}\|_\infty \leq M$:

$$\begin{aligned} \boldsymbol{x}^*_{\text{primal}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \sum_{k=1}^{K} \big[\alpha_k \mathcal{I}(\|\boldsymbol{x}_k\|_p > 0) + \beta_k \|\boldsymbol{x}_k\|_0\big], \\ \text{s.t. } A\boldsymbol{x} = \boldsymbol{b} \text{ and } \|\boldsymbol{x}\|_\infty \leq M. \end{aligned} \tag{5}$$

Note that M is chosen such that $\boldsymbol{x}^*_{\text{primal}}$ solves (4).

In our exposition, we will refer to ℓ_p-minimization and $\ell_{p,q}$-minimization as ℓ_p-min and $\ell_{p,q}$-min, respectively

Primal problem. We now frame an equivalent optimization problem for (5). Let $\boldsymbol{z} \in \{0,1\}^n$ be an entry-based sparsity indicator vector for $\boldsymbol{x}$, i.e., $z_i = 0$ if $x_i = 0$ and $z_i = 1$ otherwise. We introduce a group-based sparsity indicator vector $\boldsymbol{g} \in \{0,1\}^K$, whose k^{th} entry g_k denotes whether the k^{th} 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 i^{th} entry of $\boldsymbol{x}$ belongs to the j^{th} block and $\Pi_{i,j} = 0$ otherwise. We separate 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, (5) can be reformulated as

$$\begin{aligned} &\{\boldsymbol{x}_+^*, \boldsymbol{x}_-^*, \boldsymbol{z}^*, \boldsymbol{g}^*\} = \underset{\{\boldsymbol{x}_+, \boldsymbol{x}_-, \boldsymbol{z}, \boldsymbol{g}\}}{\operatorname{argmin}} \big[\boldsymbol{\alpha}^\top \boldsymbol{g} + \boldsymbol{\beta}^\top \boldsymbol{z}\big], \text{ s.t.} \\ &\text{(a) } \boldsymbol{x}_+ \geq 0, \text{ (b) } \boldsymbol{x}_- \geq 0, \text{ (c) } \boldsymbol{g} \in \{0,1\}^K, \text{(d) } \boldsymbol{z} \in \{0,1\}^n \\ &\text{(e) } A(\boldsymbol{x}_+ - \boldsymbol{x}_-) = \boldsymbol{b}, \text{ (f) } \Pi \boldsymbol{g} \geq \frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-), \text{ and} \\ &\text{(g) } \boldsymbol{z} \geq \frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-), \text{ where} \end{aligned} \tag{6}$$

$\boldsymbol{\alpha} = \big[\alpha_1, \cdots, \alpha_k\big]^\top \in \mathbb{R}^k$ and $\boldsymbol{\beta} = [\cdots \underbrace{\beta_k, \cdots, \beta_k}_{n_k \text{times}} \cdots]^\top \in \mathbb{R}^n$.

Constraints (a)–(d) 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 group and entry-wise sparsity of $\boldsymbol{x}$.

Lagrangial dual. The Lagrangian function for (6) is given as

$$\begin{aligned} &L(\boldsymbol{x}_+, \boldsymbol{x}_-, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3, \boldsymbol{\lambda}_4, \boldsymbol{\lambda}_5) = \\ &\boldsymbol{\alpha}^\top \boldsymbol{g} + \boldsymbol{\beta}^\top \boldsymbol{z} - \boldsymbol{\lambda}_1^\top \boldsymbol{x}_+ - \boldsymbol{\lambda}_2^\top \boldsymbol{x}_- + \boldsymbol{\lambda}_3^\top (\boldsymbol{b} - A\boldsymbol{x}_+ + A\boldsymbol{x}_-) \\ &+ \boldsymbol{\lambda}_4^\top (\frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-) - \Pi \boldsymbol{g}) + \boldsymbol{\lambda}_5^\top (\frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-) - \boldsymbol{z}), \end{aligned} \tag{7}$$

where $\boldsymbol{\lambda}_1 \geq \boldsymbol{0}$, $\boldsymbol{\lambda}_2 \geq \boldsymbol{0}$, $\boldsymbol{\lambda}_4 \geq \boldsymbol{0}$, and $\boldsymbol{\lambda}_5 \geq \boldsymbol{0}$. We now minimize $L(\cdot)$ with respect to $\boldsymbol{x}_+$, $\boldsymbol{x}_-$, $\boldsymbol{g}$ and $\boldsymbol{z}$ to obtain the Lagrangian dual function [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. Hence, 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 g_k only takes values 0 or 1, its optimal value $\hat{g}_k$ that minimizes $L(\cdot)$ is given as

$$\hat{g}_k = \begin{cases} 0 & \text{if } \alpha_k - (\Pi^\top \boldsymbol{\lambda}_4)_k > 0, \text{ and} \\ 1 & \text{otherwise.} \end{cases} \tag{8}$$

A similar expression can be computed for the minimization with respect to $\boldsymbol{z}$. Using these results, the Lagrangian dual problem can be derived as the following linear program (LP):

$$\begin{aligned} &\{\boldsymbol{y}_i^*\}_{i=1}^5 = \underset{\{\boldsymbol{y}_i\}_{i=1}^5}{\operatorname{argmax}} \big[\boldsymbol{y}_1^\top \boldsymbol{b} + \boldsymbol{1}^\top \boldsymbol{y}_4 + \boldsymbol{1}^\top \boldsymbol{y}_5\big], \text{s.t. (a) } \boldsymbol{y}_2 \geq \boldsymbol{0}, \\ &\text{(b) } \boldsymbol{y}_3 \geq \boldsymbol{0}, \text{ (c) } \boldsymbol{y}_4 \leq 0, \text{ (d) } \boldsymbol{y}_5 \leq 0 \text{ (e) } \boldsymbol{y}_4 \leq \boldsymbol{\alpha} - \Pi^\top \boldsymbol{y}_2, \\ &\text{(f) } \boldsymbol{y}_5 \leq \boldsymbol{\beta} - \boldsymbol{y}_3, \text{ and (g) } |A^\top \boldsymbol{y}_1| \leq \frac{1}{M}(\boldsymbol{y}_2 + \boldsymbol{y}_3). \end{aligned} \tag{9}$$

Lagrangian bidual. We can similarly derive the Lagrangian dual of (9), i.e., the *Lagrangian bidual* of (6), as:

$$\begin{aligned} &\{\boldsymbol{x}_+^*, \boldsymbol{x}_-^*, \boldsymbol{z}^*, \boldsymbol{g}^*\} = \underset{\{\boldsymbol{x}_+, \boldsymbol{x}_-, \boldsymbol{z}, \boldsymbol{g}\}}{\operatorname{argmin}} \ \boldsymbol{\alpha}^\top \boldsymbol{g} + \boldsymbol{\beta}^\top \boldsymbol{z} \quad \text{s.t.} \\ &\text{(a) } \boldsymbol{x}_+ \geq 0, \text{ (b) } \boldsymbol{x}_- \geq 0, \text{ (c) } \boldsymbol{g} \in [0,1]^K, \text{(d) } \boldsymbol{z} \in [0,1]^n, \\ &\text{(e) } A(\boldsymbol{x}_+ - \boldsymbol{x}_-) = \boldsymbol{b}, \text{ (f) } \Pi \boldsymbol{g} \geq \frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-) \\ &\text{and (g) } \boldsymbol{z} \geq \frac{1}{M}(\boldsymbol{x}_+ + \boldsymbol{x}_-). \end{aligned} \tag{10}$$

Comparing (6) and (10), the discrete valued variables $\boldsymbol{z}$ and $\boldsymbol{g}$ in (6) have been relaxed in (10) to take real values between 0 and 1. Since $\boldsymbol{z} \leq \boldsymbol{1}$ and $\boldsymbol{x} = \boldsymbol{x}_+ - \boldsymbol{x}_-$, constraint

(g) in (10) implies 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, 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 (10) is equivalent to solving the problem:

$$\boldsymbol{x}^*_{\text{bidual}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{M}\sum_{k=1}^{K}\left[\alpha_k\|\boldsymbol{x}_k\|_\infty + \beta_k\|\boldsymbol{x}_k\|_1\right], \quad \text{s.t. (a) } A\boldsymbol{x} = \boldsymbol{b} \text{ and (b) } \|\boldsymbol{x}\|_\infty \leq M. \tag{11}$$

This is the Lagrangian bidual for (6).

3. THEORETICAL RESULTS FROM BIDUALITY

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

Theorem 1. *The optimal value of the bidual in* (11) *is a lower bound on the optimal value of the primal problem in* (6).

Proof. Since there is no duality gap between a linear program and its Lagrangian dual [2], the optimal values of the Lagrangian dual in (9) and the Lagrangian bidual in (11) are the same. Moreover, 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 result. □

Remark 1. *Since the original primal problem in* (6) *is NP-hard, the duality gap between the primal and its dual in* (9) *is non-zero in general. Moreover, as we increase M (i.e., a more conservative estimate), the duality gap increases.*

M in (5) should preferably be equal to $\|\boldsymbol{x}^*_{\text{primal}}\|_\infty$, 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. Consider the following modification of the bidual:

$$\boldsymbol{x}^\dagger_{\text{bidual}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \sum_{k=1}^{K}\left[\alpha_k\|\boldsymbol{x}_k\|_\infty + \beta_k\|\boldsymbol{x}_k\|_1\right] \text{ s.t. } A\boldsymbol{x} = \boldsymbol{b}, \tag{12}$$

where we have dropped the box constraint (b) in (11). Notice that $\forall M \geq \max\{\|\boldsymbol{x}^*_{\text{primal}}\|_\infty, \|\boldsymbol{x}^\dagger_{\text{bidual}}\|_\infty\}$, we have that $\boldsymbol{x}^*_{\text{bidual}} = \boldsymbol{x}^\dagger_{\text{bidual}}$. Hence, *using a conservatively large value of M is equivalent to solving the modified bidual in* (12).

3.1. Results for entry-wise sparsity minimization

When $\alpha_1 = \cdots = \alpha_K = 0$ and $\beta_1 = \cdots = \beta_K = 1$, the problem in (4) reduces to the entry-wise sparsity minimization problem in (1). Importantly, we conclude from (12) that solving the Lagrangian bidual with a conservative estimate of M is equivalent to solving the problem:

$$\boldsymbol{x}^\dagger_{\text{entry-wise-bidual}} = \underset{\boldsymbol{x}}{\operatorname{argmin}}\|\boldsymbol{x}\|_1 \quad \text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}, \tag{13}$$

which is the well-known ℓ_1-norm relaxation for (P_0) [4]. Hence, we provide a new interpretation for this relaxation: *The ℓ_1-min problem in* (13) *is the Lagrangian bidual of the ℓ_0-min problem in* (1).

Our bidual-based relaxation is different from the SDP bidual-based relaxation derived by [3], due to the different (albeit equivalent) reformulations of the primal problem.

3.2. Results for group sparsity minimization

When $\forall k = 1, \ldots, K, \alpha_k = 1$ and $\beta_k = 0$, (4) reduces to the group sparsity minimization problem in (2). Hence, we conclude from (12) that solving the bidual to the group sparsity problem with a conservative estimate of M is equivalent to:

$$\boldsymbol{x}^\dagger_{\text{group-bidual}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \sum_{k=1}^{K}\|\boldsymbol{x}_k\|_\infty \quad \text{s.t.} \quad A\boldsymbol{x} = \boldsymbol{b}, \tag{14}$$

which is the convex $\ell_{1,\infty}$-norm relaxation of the $\ell_{0,p}$-min problem (2). In addition to the submodularity-based interpretation [1], we now have: *the $\ell_{1,\infty}$-min problem in* (14) *is the Lagrangian bidual of the $\ell_{0,p}$-min problem in* (2).

Interestingly, we get the $\ell_{1,\infty}$-norm for the bidual due to the constraint (f) in (6) which implies that $g_k \geq \frac{\|\boldsymbol{x}_k\|_\infty}{M}$. Similar relaxations can be derived for other p-norms for finite-valued p, by replacing this with $g_k \geq \frac{\|\boldsymbol{x}_k\|_p}{Mn_k}$ where n_k denotes the number of entries in $\boldsymbol{x}_k$, i.e., the number of columns in A_k. Now, if we relax $\boldsymbol{g}$ to take values in $[0,1]$ as observed in our derived bidual, we get the following relaxations for (6):

$$\boldsymbol{x}^*_{1,\text{p}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \sum_{k=1}^{K}\frac{\|\boldsymbol{x}_k\|_p}{Mn_k} \text{ s.t. } A\boldsymbol{x} = \boldsymbol{b} \text{ and } \|\boldsymbol{x}\|_\infty \leq M. \tag{15}$$

Compare the above to the following family of relaxations:

$$\boldsymbol{x}^\dagger_{1,\text{p}} = \underset{\boldsymbol{x}}{\operatorname{argmin}}\, \ell_{1,p}(\boldsymbol{x}) \text{ s.t. } A\boldsymbol{x} = \boldsymbol{b} \text{ and } \|\boldsymbol{x}\|_\infty \leq M. \tag{16}$$

For a conservatively large M, the solutions of (16) are precisely the solutions to the $\ell_{1,p}$ relaxations discussed in [6]. In contrast to (15), the relaxations in (16) ignore scaling the cost function by the block lengths. This may result in biases due to difference in sizes of training data for each class.

Now, since (15) is obtained by relaxing the feasible values of $\boldsymbol{g}$ in (6), its optimal value $\sum_{k=1}^{K}\frac{\|\boldsymbol{x}^*_{1,p,k}\|_p}{Mn_k}$ is lower than that of (6), where $\boldsymbol{x}^*_{1,p,k} \in \mathbb{R}^{n_k}$ denotes the k^{th} block of $\boldsymbol{x}^*_{1,p}$. Hence, this value is a lower bound for the group sparsity. We will show that *the bound for group sparsity is the tightest for $p = \infty$, i.e., with the bidual relaxation.*

Notice that $\boldsymbol{x}^*_{1,\infty}$ is the solution to the bidual (11) for the group sparsity case. Theorem 1 states that for any $M \geq \|\boldsymbol{x}^*_{0,p}\|_\infty$, a lower bound for $\ell_{0,p}(\boldsymbol{x}^*_{0,p})$ is given by $\frac{1}{M}\ell_{1,\infty}(\boldsymbol{x}^*_{1,\infty})$. We will show that the optimal value of (15) for finite p, is not greater than $\frac{1}{M}\ell_{1,\infty}(\boldsymbol{x}^*_{1,\infty})$. By the optimality of $\boldsymbol{x}^*_{1,p}$, we have $\sum_{k=1}^{K}\frac{\|\boldsymbol{x}^*_{1,p}\|_p}{Mn_k} \leq \sum_{k=1}^{K}\frac{\|\boldsymbol{x}^*_{1,\infty}\|_p}{Mn_k}$. We also know that $\frac{\|\boldsymbol{x}^*_{1,\infty,k}\|_p}{n_k} \leq \|\boldsymbol{x}^*_{1,\infty,k}\|_\infty$. Hence, we have $\sum_{k=1}^{K}\frac{\|\boldsymbol{x}^*_{1,p}\|_p}{Mn_k} \leq \sum_{k=1}^{K}\frac{\|\boldsymbol{x}^*_{1,\infty}\|_\infty}{M} = \frac{\ell_{1,\infty}(\boldsymbol{x}^*_{1,\infty})}{M}$.

4. EXPERIMENTS

Now, we present experiments on synthetically generated data to show that the performance of the bidual relaxation depends on the distribution of the entries of $\boldsymbol{x}$. More comprehensive experiments on synthetically generated data as well as face recognition can be found in our technical report at `http://arxiv.org/abs/1201.3674`.

We set the dimension of the ambient space to $n = 500$ and vary the number of blocks in A from $K = 60$ to 100 in steps of 10. The size of each block in A is fixed as $n_k = 20$. Hence, $m = Kn_k$ varies from 1200 to 2000 in steps of 200. We ensure that no two blocks of A are linearly dependent.

The number K_{nz} of non-zero blocks in $\boldsymbol{x}$ is varied from 1 to 12. We consider two different test cases, where the entries of each block in $\boldsymbol{x}$ are drawn from two different distributions. The coefficients are drawn from the normal distribution $\mathcal{N}(0, 1)$ in the first case and from $\mathcal{N}(20, 1)$ in the second case. The fraction of non-zero entries in $\boldsymbol{e}$ is set as $\frac{K_{nz}}{K}$. In both cases, the entries of $\boldsymbol{e}$ are drawn from $\mathcal{N}(0, 1)$.

We solve for $\boldsymbol{x}$ and $\boldsymbol{e}$ by solving the following problem:

$$\min_{\{\boldsymbol{x},\boldsymbol{e}\}} \sum_{k=1}^{K} \|\boldsymbol{x}_k\|_p + \gamma\|\boldsymbol{e}\|_1, \text{ s.t. } A\boldsymbol{x} + \boldsymbol{e} = \boldsymbol{b}. \qquad (17)$$

In our tests, we set $\gamma = 1$ and compare the solutions obtained using $p = 1, 2$ and ∞. MOSEK (`www.mosek.com`) is used as the numerical solver. For $p = \infty$, this reduces to the bidual relaxation (12) of the mixed sparsity problem (5) with a conservative estimate of M. The case $p = 2$ corresponds to a standard relaxation for minimizing group sparsity [6]. The case $p = 1$ reduces to minimizing entry-wise sparsity of $\boldsymbol{x}$.

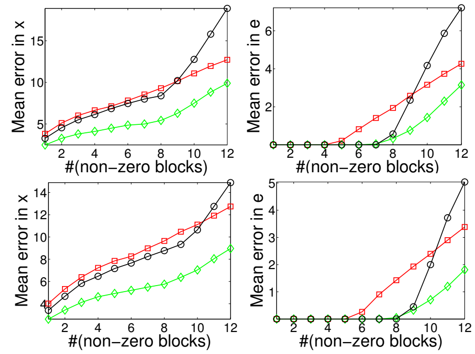
A representative subset of the results is shown in Figure 1 and detailed results are given in our technical report. We see from Figure 1(a) that the $\ell_{1,2}$-based relaxation performs better in the first test case. On the other hand, Figure 1(b) shows that the $\ell_{1,\infty}$-based relaxation performs better in the second case when the entries of $\boldsymbol{x}$ are clustered around a non-zero value. The performance differences are linked to the norms used in the relaxation and the distribution of the entries in $\boldsymbol{x}$.

5. DISCUSSION

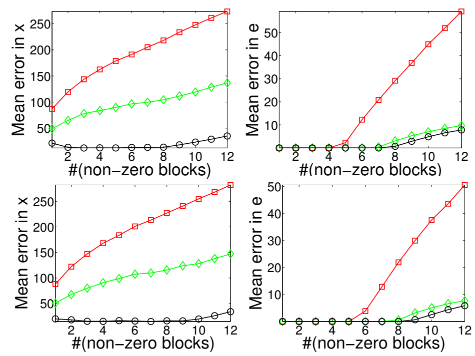
We have presented a novel analysis to interpret convex relaxations of sparsity minimization problems as their Lagrangian biduals. The pivotal point of this analysis is the formulation of mixed-integer programs which are equivalent to the original primal problems. Biduality is only one of the many choices for generating relaxations and the performance of a particular relaxation ultimately depends on the distribution of the data being estimated.

6. ACKNOWLEDGEMENTS

This research was supported by ARO MURI W911NF-06-1-0076, ARL MAST-CTA W911NF-08-2-0004, NSF CNS-0931805, NSF CNS-0941463 and NSF grant 0834470. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute for Government purposes notwithstanding any copyright notation herein.



(a) First case: entries of $\boldsymbol{x}$ drawn from $\mathcal{N}(0, 1)$



(b) Second case: entries of $\boldsymbol{x}$ drawn from $\mathcal{N}(20, 1)$

Fig. 1. Representative results for the two test cases. The top and bottom rows in each subfigure show results for $K = 60$ and $K = 100$, respectively. The color coding for the results is ℓ_1-red squares, $\ell_{1,2}$-green diamonds and $\ell_{1,\infty}$-black circles.

7. REFERENCES

[1] F. Bach. Structured Sparsity-Inducing Norms through Submodular Functions. *NIPS*, 2010.

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

[3] S. Chrétien. An alternating ℓ_1 approach to the compressed sensing problem. IEEE Signal Processing Letters, 17(2), 2010.

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

[5] D. Donoho. Neighborly Polytopes and Sparse Solutions of Underdetermined Linear Equations. Technical Report, 2005.

[6] E. Elhamifar and R. Vidal. Robust classification using structured sparse representation. In *CVPR*, 2011.

[7] R. Luss and M. Teboulle. Convex Approximations to Sparse PCA via Lagrangian Duality. Operations Research Letters, 39(1), 2011.

[8] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *PAMI*, 31(2), 2009.

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