AN ELLIPSOID BASED, TWO-STAGE SCREENING TEST FOR BPDN

Liang Dai, Kristiaan Pelckmans

Uppsala University,
Institute of Information Technology,
Division of Systems and Control, Uppsala, Sweden

ABSTRACT

Consider the Basis Pursuit De-Noising (BPDN) estimator for recovery of unknown, sparse parameters. This note presents an ellipsoid-based, two-stage screening test method which aims to reduce a-priori the dimensionality of the resulting optimization problem. The new elements of the proposed method are given by (i) using an efficient ellipsoid approximation scheme in both stages and (ii) making better use of the information which has been calculated during the first stage. A comparative experiment indicates that this procedure can lead to better overall time complexity compared to known screening tests, while *screening away* more irrelevant variables in a preprocessing stage.

1. INTRODUCTION

We will first introduce the Basis Pursuit De-Noising (BPDN) estimator briefly in the following. Let $n \in \mathbb{N}$ denotes the number of observations, and $m \gg n$ denotes the dimensionality of the problem. Given a measurement vector $\mathbf{x} \in \mathbb{R}^n$, m dictionary vectors(atoms) $\mathbf{b}_i \in \mathbb{R}^n$. Let these atoms be organized in a matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$ such that the ith column of $\mathbf{B}$ equals $\mathbf{b}_i$. Assume that there is an (unknown) vector $\mathbf{w}_0 \in \mathbb{R}^m$ (which is assumed to be sparse) and a vector $\mathbf{e} \in \mathbb{R}^n$ (which represents noise), such that

$$\mathbf{x} = \mathbf{B}\mathbf{w}_0 + \mathbf{e}.$$

The task is to recover $\mathbf{w}_0$ from $\mathbf{x}$ and $\mathbf{B}$. A survey of techniques applicable to this task is given in [3, 4].

A reasonable estimate $\mathbf{w} = (w_1, \ldots, w_m)^T$ of $\mathbf{w}_0$ is given by solving the following problem for a given $\lambda > 0$:

$$\min_{\mathbf{w}\in\mathbb{R}^m} \frac{1}{2}\|\mathbf{x} - \mathbf{B}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_0, \tag{1}$$

where $\|\mathbf{w}\|_0 = \sum_{d=1}^m I(w_d \neq 0)$, with the indicator $I(z)$ equals to one iff z holds true, and zero otherwise. Here the parameter $\lambda > 0$ regulates the tradeoff between the data fit and representation complexity.

While (1) is non-smooth, non-convex, and strictly NP-hard [8], one often resorts to solving the convex BPDN which serves as tractable proxy to (1) by solving

$$\bar{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathbb{R}^m} \frac{1}{2}\|\mathbf{x} - \mathbf{B}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1, \tag{2}$$

where the convex L_1-norm is defined as $\|\mathbf{w}\|_1 = \sum_{d=1}^m |w_d|$. Here, the norm $\|\cdot\|_1$ is regarded as the convex envelop to the non-convex $\|\cdot\|_0$. This estimator (2) is sometimes referred to as to the Least Absolute Shrinkage and Selection Operator (LASSO) estimator. As in [1], we also assume that $\|\mathbf{x}\|_2 = 1$ and $\|\mathbf{b}_i\|_2 = 1$ for $i = 1 \ldots m$.

The formulation of BPDN has found many interesting applications and theoretical results, in particular because of the facts that:

- Since the BPDN boils down to a convex optimization problem, it can be solved efficiently with well-known tools as the Interior Point Method (IPM) [5]. This is a general numerical solver for problems of convex optimization. Extensive research on this particular problem resulted in a wide variety of numerical solvers which obtain better practical as well as theoretical performance by exploiting more structure information of the problem. For an up-to-date collection of such methods, please consult[1].
- Theoretical excitement stems from the fact that recoverability(such as the support of $\mathbf{w}_0$,or some 'good' estimations of $\mathbf{w}_0$) of $\mathbf{w}_0$ can be guaranteed under certain conditions of the measurement matrix $\mathbf{B}$ (Restricted Isometry Property, Null Space Property, Spherical Section Property, etc, see [3, 4]) and the sparsity level of $\mathbf{w}_0$. Such guarantees come in different forms as surveyed in [3, 4] and citations.

According to the reference [1], the Lagrangian dual to problem (2) is given as follows:

$$\bar{\theta} = \arg\max_{\theta\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}\|_2^2 - \frac{\lambda^2}{2}\left\|\theta - \frac{\mathbf{x}}{\lambda}\right\|_2^2 \quad \text{s.t. } |\theta^T\mathbf{b}_i| \leq 1, \forall i = 1, 2, \ldots, m. \tag{3}$$

The optimal solutions $\bar{\mathbf{w}}$ to problem (2) and $\bar{\theta}$ to problem (3) are connected through eq. (4) and eq. (5). We refer the readers to the reference[1] for details.

$$\mathbf{x} = \sum_{i=1}^m \bar{w_i}\mathbf{b}_i + \lambda(\bar{\theta}), \tag{4}$$

and

$$\bar{\theta}^T\mathbf{b}_i \in \begin{cases} \text{sign}(\bar{w_i}) & \text{iff } \bar{w_i} \neq 0 \\ [-1, 1] & \text{iff } \bar{w_i} = 0. \end{cases} \tag{5}$$

Define the halfspace $H(\mathbf{y})$ for $\mathbf{y} \in \mathbb{R}^n$ as

$$H(\mathbf{y}) = \left\{\mathbf{z} : \mathbf{z}^T\mathbf{y} \leq 1\right\} \subset \mathbb{R}^n.$$

Let $L(\mathbf{y})$ be the corresponding hyperplane

$$L(\mathbf{y}) = \left\{\mathbf{z} : \mathbf{z}^T\mathbf{y} = 1\right\} \subset \mathbb{R}^n.$$

The reasoning behind the construction of a screening test goes as follows. From eq. (3), (4) and (5), we can see that if $\bar{\theta}$ is not on

[1] http://ugcs.caltech.edu/~srbecker/wiki/Category:Solvers

$L(\mathbf{b_i})$ nor $L(-\mathbf{b_i})$, then $\bar{w}_i$ will be zero. This is a crucial observation for screening tests as pointed out in [1, 2, 6]. The idea is then to make a set $Q \subset \mathbb{R}^n$ which contains $\bar{\theta}$, and check for $i = 1, \ldots, m$, whether $L(\mathbf{b}_i)$ or $L(-\mathbf{b_i})$ intersects Q or not. If for i, no intersection takes place, one can conclude that $\bar{w}_i$ is zero, and it doesn't need to be included in later stages of the optimization problem. That is, this corresponding dictionary $\mathbf{b}_i$ is *screened away* in the subsequent optimization problem.

The aim of this paper is to reduce m before actually solving (2). That is, we aim to filter out (or *screen out*) as many different columns of $\mathbf{B}$ as possible, before performing the convex optimization problem (2) completely. Such preprocessing stage could then lead to subsequent less time and memory intensive optimization procedures since m could be reduced severely. An important point is that such *screening* stage should not be too computationally involved to perform.

Some test methods have been devised already in [1, 2, 6] with different levels of effectiveness. This note introduces a two-stage ellipsoid based screening test which further improves the screening performance. This means that in total, the computational cost including the cost for the screening test and the cost for the subsequent optimization will be reduced. Our strategy is composed of two stages, which in general are:

1. Approximate the basic potential region Q for $\bar{\theta}$ with an ellipsoid, and then perform the 'intersection test'. If neither $L(\mathbf{b}_i)$ nor $L(-\mathbf{b}_i)$ intersect with this ellipsoid, then the corresponding $\bar{\mathbf{w}}_i$ is set to zero. This is similar to the tests as performed in the traditional screening tests.
2. In the second stage, a new approximation of the potential region of $\bar{\theta}$ based on the information which is obtained earlier (we only choose one halfspace which shrink the volume most, details are in section 3). Then another round screening test is obtained based on this updated ellipsoid. Note that, this stage, the test only is carried out on those atoms with haven't been determined to be screened out in the first stage.

Our method is motivated as follows: (1) the update rule of the ellipsoid approximation is simple; (2) while performing the 'intersection test' in the first round, information can also be used for obtaining a tighter approximation of the potential region of $\bar{\theta}$; (3) the 'intersection test' in every round also requires low time cost.

We will use the following notational conventions throughout. A lower-case letter denotes a scalar, a boldface lowercase a vector and a boldface capital denotes a matrix. This paper is organized as follows. Section II describes the ellipsoid related results, including the update rule, and some related geometrical results. Section III describes our algorithm in detail. Section IV gives experimental results indicating the efficacy of the method, and compares to existing approaches. Section IV concludes this paper and points towards interesting open avenues for further research.

2. ELLIPSOID RELATED RESULTS

In this section, we will give the ellipsoid update rule and some related results. These results will be used in the forming of our proposed algorithm in the following sections.

2.1. Ellipsoid Update Rule

Given a halfspace represented as

$$H_h(\mathbf{x}_p, \mathbf{g}) = \left\{ \mathbf{z} \in \mathbb{R}^n : \mathbf{g}^T(\mathbf{z} - \mathbf{x}_p) + h \leq 0 \right\},$$

and the corresponding hyperplane as

$$L_h(\mathbf{x}_p, \mathbf{g}) = \left\{ \mathbf{z} \in \mathbb{R}^n : \mathbf{g}^T(\mathbf{z} - \mathbf{x}_p) + h = 0 \right\},$$

where $h \geq 0, \mathbf{g}, \mathbf{x}_p \in \mathbb{R}^n$ are given, and an ellipsoid

$$E(\mathbf{x}_p, \mathbf{P}_p) = \{\mathbf{z} \in \mathbb{R}^n : (\mathbf{z} - \mathbf{x}_p)^T \mathbf{P}_p^{-1} (\mathbf{z} - \mathbf{x}_p) \leq 1\},$$

where $\mathbf{P}_p \in \mathbb{R}^{n \times n}$ and $\mathbf{P}_p \succeq 0$. Then the ellipsoid with the minimum volume which contains the intersection of $H_h(\mathbf{x}_p, \mathbf{g})$ and $E(\mathbf{P}_p, \mathbf{x}_p)$ could be represented as

$$E(\mathbf{x}_u, \mathbf{P}_u) = \left\{ \mathbf{z} \in \mathbb{R}^n : (\mathbf{z} - \mathbf{x}_u)^T \mathbf{P}_u^{-1} (\mathbf{z} - \mathbf{x}_u) \leq 1 \right\}.$$

Here we define

$$\left\{ \begin{array}{l} \mathbf{x}_u = \mathbf{x}_p - \frac{1+\alpha n}{n+1} \mathbf{P}_p \bar{\mathbf{g}} \\ \mathbf{P}_u = \frac{n^2(1-\alpha^2)}{n^2-1} \left(\mathbf{P}_p - \frac{2(1+\alpha n)}{(n+1)(\alpha+1)} \mathbf{P}_p \bar{\mathbf{g}} \bar{\mathbf{g}}^T \mathbf{P}_p \right), \end{array} \right. \tag{6}$$

where $\bar{\mathbf{g}} = \frac{\mathbf{g}}{\sqrt{\mathbf{g}^T \mathbf{P}_p \mathbf{g}}}$ and $\alpha = \frac{h}{\sqrt{\mathbf{g}^T \mathbf{P}_p \mathbf{g}}}$.

The derivation of this update rule is given in [5]. It has found main application for bounding convex sets as in the membership set method [7] as commonly used in system identification. It also plays an important historic role in finding polynomial time solver for solving Linear Programming (LP) problems [5]. In the following, we will give a rule which decides if a hyperplane intersects with an ellipsoid or not.

2.2. Intersection test

Lemma 1 *Given $h > 0, \mathbf{x}_p \in \mathbb{R}^n, \mathbf{g} \in \mathbb{R}^n, \mathbf{P}_p \in \mathbb{R}^{n \times n}$. Define*

$$\alpha = \frac{h}{\sqrt{\mathbf{g}^T \mathbf{P}_p \mathbf{g}}}.$$

If $|\alpha| > 1$, then the intersection of the hyperplane $L_h(\mathbf{x}_p, \mathbf{g})$ with the ellipsoid $E(\mathbf{x}_p, \mathbf{P}_p)$ is empty.

Proof 1 *From a geometric viewpoint, this lemma follows by the following reasoning. Since*

$$\{\mathbf{z} : \mathbf{g}^T(\mathbf{z} - \mathbf{x}_p) + h = 0\} \cap \{\mathbf{z} : (\mathbf{z} - \mathbf{x}_p)^T \mathbf{P}_p^{-1} (\mathbf{z} - \mathbf{x}_p) \leq 1\} = \varnothing, \tag{7}$$

holds if and only if

$$\{\mathbf{z} : \mathbf{g}^T \mathbf{P}_p^{\frac{1}{2}} \mathbf{z} + h = 0\} \cap \{\mathbf{z} : \mathbf{z}^T \mathbf{z} \leq 1\} = \varnothing. \tag{8}$$

Notice that the distance from 0 to the hyperplane given as $\{\mathbf{z} : \mathbf{g}^T \mathbf{P}_p^{\frac{1}{2}} \mathbf{z} + h = 0\}$ is equal to

$$\frac{|h|}{\sqrt{\mathbf{g}^T \mathbf{P}_p \mathbf{g}}}.$$

Hence it follows that if $|\alpha| > 1$, the intersection will be empty. This concludes the proof.

In the following, we will characterize how much of the volume will be shrunken by the update.

2.3. Shrinkage of the Volume

Lemma 2 *Define*

$$\alpha = \frac{h}{\sqrt{\mathbf{g}^T \mathbf{P}_p \mathbf{g}}}.$$

If $0 \leq \alpha \leq 1$, *then one has that after the ellipsoid update as depicted in eq. (6), the volumes are shrunken as:*

$$\frac{\mathrm{vol}\,(E(\mathbf{x}_p, \mathbf{P}_p))}{\mathrm{vol}\,(E(\mathbf{x}_u, \mathbf{P}_u))} = \frac{n^n}{(1+n)(n^2-1)^{\frac{n-1}{2}}}(1-\alpha)(1-\alpha^2)^{\frac{n-1}{2}}.$$

Proof 2 *We have that*

$$\begin{aligned}
\frac{\mathrm{vol}^2\,(E(\mathbf{x}_p, \mathbf{P}_p))}{\mathrm{vol}^2\,(E(\mathbf{x}_u, \mathbf{P}_u))} &= \frac{|\mathbf{P}_u|}{|\mathbf{P}_p|} \qquad (9)\\
&= \frac{\left|\frac{n^2(1-\alpha^2)}{n^2-1}\left(\mathbf{P}_p - \frac{2(1+\alpha n)}{(n+1)(\alpha+1)}\mathbf{P}_p \bar{g}\bar{g}^T \mathbf{P}_p\right)\right|}{|\mathbf{P}_p|}\\
&= \frac{\left|\frac{n^2(1-\alpha^2)}{n^2-1}\mathbf{P}_p^{\frac{1}{2}}\left(I - \frac{2(1+\alpha n)}{(n+1)(\alpha+1)}\mathbf{P}_p^{\frac{1}{2}}\bar{\mathbf{g}}\bar{\mathbf{g}}^T\mathbf{P}_p^{\frac{1}{2}}\right)\mathbf{P}_p^{\frac{1}{2}}\right|}{|\mathbf{P}_p|}\\
&= \left(\frac{n^2(1-\alpha^2)}{n^2-1}\right)^n \left|I - \frac{2(1+\alpha n)}{(n+1)(\alpha+1)}\mathbf{P}_p^{\frac{1}{2}}\bar{\mathbf{g}}\bar{\mathbf{g}}^T\mathbf{P}_p^{\frac{1}{2}}\right|\\
&= \left(\frac{n^2(1-\alpha^2)}{n^2-1}\right)^n \left|1 - \frac{2(1+\alpha n)}{(n+1)(\alpha+1)}\bar{\mathbf{g}}^T\mathbf{P}_p\bar{\mathbf{g}}\right|\\
&= \left(\frac{n^2(1-\alpha^2)}{n^2-1}\right)^n \left|1 - \frac{2(1+\alpha n)}{(n+1)(\alpha+1)}\right|\\
&= \frac{n^{2n}}{(1+n)^2(n^2-1)^{n-1}}(1-\alpha)^2(1-\alpha^2)^{n-1},
\end{aligned}$$

as desired.

Remark 1 *From both lemmas, we see that* α *plays a remarkable role. This factor not only let us decide whether the hyperplane will intersect with the ellipsoid or not, but also can help to characterize how much the volume of the updated ellipsoid will be shrunken. Especially, from Lemma 2, we can see that the larger* α *is, the more the volume of the updated ellipsoid will shrink.*

3. ALGORITHM

Using the same notations as in [1, 2], we define $\lambda_{\max} = \max_i |\mathbf{x}^T \mathbf{b}_i|$. The vector $\mathbf{b}_*$ is defined so as to satisfy $\lambda_{\max} = \mathbf{x}^T \mathbf{b}_*$. It can be verified that $\mathbf{x}/\lambda_{\max}$ is a feasible solution to the dual (3). In order to avoid the trivial case, we assume that $\lambda < \lambda_{\max}$ as in [1, 2]. Define the region $R_1 \subset \mathbb{R}^n$ as

$$R_1 = \{\theta : \mathbf{b}_*^T \theta \leq 1\} \bigcap \left\{\theta : \|\theta - \mathbf{x}/\lambda\|_2 \leq \sqrt{1/\lambda - 1/\lambda_{\max}}\right\}.$$

We can see that R_1 is a a region where $\bar{\theta}$ will locate in. This region has been referred to as a 'dome' in [2] (an intersection of a halfspace and a ball). As discussed before, if for $i \in \{1, \ldots, m\}$ neither of the hyperplanes $L(\mathbf{b}_i)$ or $L(-\mathbf{b}_i)$ intersects with R_1, then $\bar{\mathbf{w}}_i$ has to equal zero. In the references, the authors bound R_1 with different balls (different center and radius), which led them to convenient yet effective test as the 'SAFE/ST1','ST2','ST3' test [1, 6] or the 'dome' test[2]. The 'dome' test is considered to be the most effective one in the sense of its effectiveness (the number of irrelevant atoms screened out) and low computation cost. Hence, in the experiment part we will mainly compare the proposed screening test with the 'dome' test. As stated briefly in the previous part, the proposed test will consist of two stages. The formal and precise descriptions are given as follows.

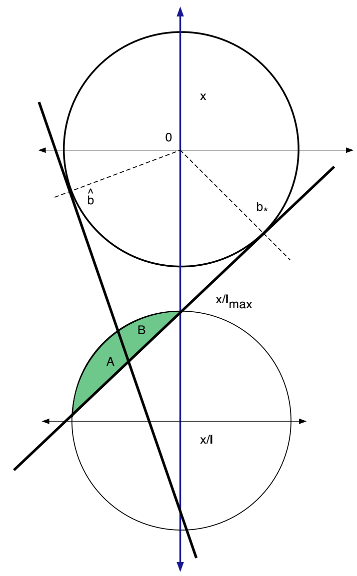

Fig. 1. *A schematic explanation of the idea behind the 2-stage, ellipsoid-based screening test when* $n = 2$. *The unit circle* S^{n-1} *at the top indicates the unit sphere including the vectors* $\{\mathbf{b}_i, -\mathbf{b}_i\}_{i=1}^m$ *and* $\mathbf{x}$. *The circle at the bottom indicates the set of vectors* θ *with distance to* $\frac{\mathbf{x}}{\lambda}$ *equal to* $\|\frac{\mathbf{x}}{\lambda} - \frac{\mathbf{x}}{\lambda_{\max}}\|_2$. *Hence, the optimum* $\bar{\theta}$ *of problem (3) will be inside this circle. The solid lines indicate the hyperplanes* $L(\mathbf{b}^*)$ *and* $L(\hat{\mathbf{b}})$ *as explained in Subsections 3.1 and 3.2. The first stage of the test computes an ellipsoid estimation of the dome* $R_1 = A \bigcup B$ *which contains* $\bar{\theta}$. *Then a first round of ellipsoid based screening is applied, and many of the irrelevant dictionary atoms will be screened away. As a byproduct, those calculations give the halfspace* $H(\hat{\mathbf{b}})$ *which shrinks the volume of the ellipsoid estimation the most. So, the potential region of* $\bar{\theta}$ *will be shrunken from* R_1 *to region* B. *In the second stage of the test, screening is applied to the remaining dictionary atoms using the updated ellipsoid.*

3.1. Stage 1

1. Compute the minimum volume ellipsoid containing R_1. This calculation is a direct consequence of the update rule described in section 2, in which $\mathbf{x}_p = \frac{\mathbf{x}}{\lambda}$, $\mathbf{g} = \mathbf{b}_*$, $h = \frac{\lambda_{\max}}{\lambda} - 1$, and $\mathbf{P}_p = (\frac{1}{\lambda} - \frac{1}{\lambda_{\max}})\mathbf{I}_n$. Denote the updated ellipsoid as

$$E_1(\mathbf{x}_1, \mathbf{P}_1) = \{\mathbf{z} : (\mathbf{z} - \mathbf{x}_1)^T \mathbf{P}_1^{-1}(\mathbf{z} - \mathbf{x}_1) \leq 1\},$$

where

$$\begin{cases} \mathbf{x}_1 = \mathbf{x}_p - \frac{1+\alpha n}{n+1}\mathbf{P}_p\bar{\mathbf{g}} \\ \mathbf{P}_1 = \frac{n^2(1-\alpha^2)}{n^2-1}\left(\mathbf{P}_p - \frac{2(1+\alpha n)}{(n+1)(\alpha+1)}\mathbf{P}_p\bar{\mathbf{g}}\bar{\mathbf{g}}^T\mathbf{P}_p\right), \end{cases} \qquad (10)$$

in which $\bar{\mathbf{g}} = \frac{\mathbf{g}}{\sqrt{\mathbf{g}^T\mathbf{P}_p\mathbf{g}}}$ and $\alpha = \frac{h}{\sqrt{\mathbf{g}^T\mathbf{P}_p\mathbf{g}}}$.

2. Test for any $i = 1, \ldots, m$ whether $L(\mathbf{b}_i)$ or $L(-\mathbf{b}_i)$ intersect with $E_1(\mathbf{x}_1, \mathbf{P}_1)$ or not. If both do not intersect, then set

$\bar{\mathbf{w}}_i = 0$. Formally, calculate

$$\alpha_i^+ = \frac{\mathbf{b}_i^T \mathbf{x_1} - 1}{\sqrt{\mathbf{b}_i^T P_1 \mathbf{b}_i}},$$

and

$$\alpha_i^- = \frac{-\mathbf{b}_i^T \mathbf{x_1} - 1}{\sqrt{\mathbf{b}_i^T P_1 \mathbf{b}_i}}.$$

If $|\alpha_i^+| > 1$ and $|\alpha_i^-| > 1$ hold together, or equivalently if

$$\sqrt{\mathbf{b}_i^T \mathbf{P}_1 \mathbf{b}_i} < \min\{|\mathbf{b}_i^T \mathbf{x_1} + 1|, |\mathbf{b}_i^T \mathbf{x_1} - 1|\}, \qquad (11)$$

then set $\bar{\mathbf{w}}_i = 0$.

Remark 2 *Eq. (11) is a direct application of Lemma 1 in Section 2.*

Stage 1 gives an ellipsoid approximation to the 'dome' area R_1. As we have seen in the previous section, the proposed screening test is also convenient to compute as described above. An interesting fact is that, while we are doing the screening, if eq. (11) does not hold, one has the fact that the corresponding halfspace intersects with ellipsoid $E_1(\mathbf{x}_1, \mathbf{P}_1)$. Since α_i^- and α_i^- have been calculated in hand, we see from Lemma 2 that α_i^-, α_i^+ actually also indicate the volume shrinkage of the ellipsoid approximation of the intersection. The lager they are, the more the volume will shrink. This motivates the next stage which causes no extra significant computational overhead.

3.2. Stage 2

1. Choose the the maximum value $\hat{\alpha}$ from $\{\alpha_i^+, \alpha_i^-\}_{i=1}^m$ which satisfies $0 < \hat{\alpha} < 1$. Denote the corresponding halfspace as $H(\hat{\mathbf{b}})$ and the hyperplane as $L(\hat{\mathbf{b}})$. In the ellipsoid update rules as in eq. (6), let $\mathbf{g} = \hat{\mathbf{b}}$ and $h = \hat{\mathbf{b}}^T \mathbf{x}_1 - 1$ in order to compute the updated ellipsoid

$$E_2(\mathbf{x}_2, \mathbf{P}_2) = \{\mathbf{z} : (\mathbf{z} - \mathbf{x}_2)^T \mathbf{P}_2^{-1} (\mathbf{z} - \mathbf{x}_2) \le 1\},$$

where

$$\begin{cases} \mathbf{x}_2 = \mathbf{x}_1 - \frac{1+\alpha n}{n+1} \mathbf{P}_1 \bar{\mathbf{g}} \\ \mathbf{P}_2 = \frac{n^2(1-\alpha^2)}{n^2-1} \left(\mathbf{P}_1 - \frac{2(1+\alpha n)}{(n+1)(\alpha+1)} \mathbf{P}_1 \bar{\mathbf{g}} \bar{\mathbf{g}}^T \mathbf{P}_1 \right) \end{cases}$$

and in which $\bar{\mathbf{g}} = \frac{\mathbf{g}}{\sqrt{\mathbf{g}^T \mathbf{P}_1 \mathbf{g}}}$ and $\alpha = \frac{h}{\sqrt{\mathbf{g}^T \mathbf{P}_1 \mathbf{g}}}$.

2. For $i = 1, \ldots, m$, if $\bar{w}_i$ is not screened away yet during the first stage, test wether

$$\sqrt{\mathbf{b}_i^T \mathbf{P}_2 \mathbf{b}_i} < \min\{|\mathbf{b}_i^T \mathbf{x_2} + 1|, |\mathbf{b}_i^T \mathbf{x_2} - 1|\}. \qquad (12)$$

If this holds, then set $\bar{\mathbf{w}}_i = 0$.

4. ILLUSTRATIVE EXAMPLES

This part describes an example which indicate the efficacy of the proposed screening test, and compares result with earlier proposed screening tests. In our example, we chose the dictionary atoms $\{\mathbf{b}_i\}_{i=1}^m$ sampled randomly from a unit normally distributed random variable, and then normalize each of them to one. we generate the normalized vector $\mathbf{x}$ in the same way. In this example, we let $n = 10$ and $m = 200$. Estimation problems of this size are typical in applications of BPDN, while the relative low-dimensional nature will already indicate the benefit of the proposed technique. The displayed figures are obtained by averaging out results over 50 randomizations of the experiment. In Fig. 2, results of different screening tests are given:

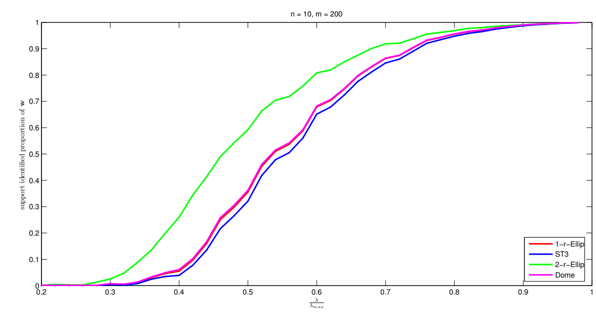

Fig. 2. *Performance of the different screening test methods, including the 'Dome test', the 'ST3 test', the '1-stage ellipsoid test', '2-stage ellipsoid test'. The x-axis represents $\frac{\lambda}{\lambda_{\max}}$, the y-axis represents the proportion of the number of the screened out zero elements in $\bar{\mathbf{w}}$. This result illustrates a significant benefit of the proposed 2-stage screening test for appropriate range of $\lambda/\lambda_{\max}$.*

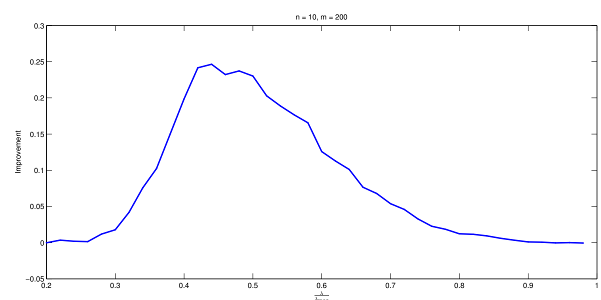

Fig. 3. *Improvement of the performance of the '2-stage ellipsoid test' and the 'Dome test'. The x-axis represents $\frac{\lambda}{\lambda_{\max}}$, the y-axis represents the ratio of proportions of the number the screened out zero elements in $\bar{\mathbf{w}}$. Larger values indicate less remaining dictionary elements after the '2-stage, ellipsoid test'. This plot indicates that the present test can have a significant gain in terms of number of screened out dictionary elements for an appropriate range of $\frac{\lambda}{\lambda_{\max}}$.*

1. the ST3 method [1], in Fig. 2, with the tag 'ST3';
2. the 1-stage ellipsoid method as derived in Subsection 3.1 (only performing the first stage), in Fig. 2, with the tag '1-r-Ellip';
3. the Dome test method [2], in Fig. 2, with the tag 'Dome';
4. the 2-stage ellipsoid method as proposed in Section 3 (including both stages), in Fig. 2, with the tag '2-r-Ellip'.

From these figures, we observe the following:

1. When the ratio $\frac{\lambda}{\lambda_{\max}}$ is relatively small (in this example less than 0.3), then all the screening test methods are relatively ineffective. But in other words, this phenomena is reasonable. If $\lambda_{\max}$ is fixed, when $\frac{\lambda}{\lambda_{\max}}$ is small (which means that λ is small), then the dome R_1 will become very large, and all hyperplanes $\{L(\mathbf{b}_i), L(-\mathbf{b}_i)\}_i$ could be expected to intersect the dome with more chance. In the extremal case that $\lambda \to 0$, any screening test would do poor, meaning that such tests cannot be applied straightforwardly to the noiseless Basis Pursuit

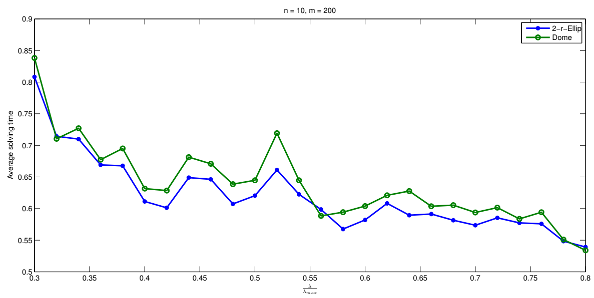

Fig. 4. *Comparison of the average time cost when 'screening with the Dome test, and solving the reduced BPDN', and 'screening with the proposed 2-stage, ellipsoid test, and solving the reduced BPDN', The x-axis represents $\frac{\lambda}{\lambda_{\max}}$, the y-axis represents the time needed to solve a corresponding complete problem. This result indicates that the improved screening capacity of the 2-stage ellipsoid screening test does not result in computational overheads, and may well lead to computational speedups.*

(BP) case.

2. When $\frac{\lambda}{\lambda_{\max}}$ is between 0.35 to 0.8, we see that the 'Dome test' and the '1-stage ellipsoid test' perform quite similar (their performance curve nearly overlapping) to each other, but do slightly better than 'ST3 test'. However, the 2-stage ellipsoid test outperforms those as more irrelevant variables are screened away, the quantitative improvement can be seen from Fig. 3. This phenomena means that, in this case, it's better to use the 2-stage ellipsoid method to do screening. Here, we need to notice that the time cost for the '2-stage-ellipsoid test' is also quite low. Fig. 4, displays the time cost for solving the same BPDN problem with 'Dome test' for screening and '2-stage ellipsoid test' for screening. After screening, we solve the reduced dimension BPDN problem with 'cvx' [5], using the internal Sedumi solver. We can see the time-saving of adopting the '2-stage-ellipsoid test' method for screening.

3. When $\frac{\lambda}{\lambda_{\max}}$ is larger than 0.8, it appears that all the methods give very similar performance. This is due to the fact that in this case the dome area is relatively small, and most of the hyperplanes $\{L(\mathbf{b}_i), L(-\mathbf{b}_i)\}_i$ will not intersect this area (which means that most coefficients of the solution are zero for such λ).

Again, note that by construction the screening tests are conservative, that is, they cannot screen variables away which would be nonzero in the final solution. Or, no performance can be lost, the screening stage can only be beneficial since the resulting optimization problem has smaller dimensionality.

5. CONCLUSION

This note presented an improvement of a screening test method for the BPDN problem. The motivation is based on an ellipsoid approximation of the potential region for $\bar{\theta}$, while the involved quantities are found to be useful in the second stage. This second stage leads to improved screening capabilities, only requring quantities which were computed in the first stage anyway. Through simulations it is found that such screening test is most effective when the ratio $\frac{\lambda}{\lambda_{\max}}$ is moderate. The comparative experiment shows that, the proposed '2-stage ellipsoid' method results in both effectiveness (more irrelevant dictionary atoms are screened away) and efficiency (the time cost for solving the whole optimization problem is reduced) improvement over the state-of-art screening test method.

However, the following questions remain open: (1) Starting with a feasible point $\frac{\mathbf{x}}{\lambda_{\max}}$, the present approach uses an initial potential region for $\bar{\theta}$ which is the 'dome' region R_1. Can we find a better starting feasible point in order to make the initial 'dome' region more accurate? (2) Can we find a way to generalize the method (including the 'SAFE/ST1', 'ST2', 'ST3', 'Dome test', and the proposed method) to the the case where $1/\lambda \to \infty$ as in Basis Pursuit (BP)?

6. REFERENCES

[1] Z. J. Xiang, H. Xu, P. J. Ramadge, Learning sparse representations of high dimensional data on large scale dictionaries, NIPS 2011.

[2] Z. J. Xiang, P. J. Ramadge, Fast lasso screening tests based on correlations, ICASSP 2012.

[3] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer, 2010.

[4] M.A. Davenport, M.F. Duarte, Y.C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing: Theory and Applications, Cambridge University Press, 2011.

[5] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[6] L.E. Ghaoui, V. Viallon, T. Rabbani, Safe feature elimination in sparse supervised learning, Arxiv preprint arXiv:1009.3515, 2010.

[7] R. Kosut, M. Lau, S. Boyd, Set-membership identification of systems with parametric and nonparametric uncertainty, IEEE Transactions on Automatic Control, 37(7):929-941, July 1992.

[8] D. Ge, X. Jiang, Y. Ye. A Note on Complexity of Lp Minimization, Mathematical Programming, Volume 129, Number 2, 285-299, 2010.