Conic Relaxations for Spase Linear Regression

Hongbo Dong[1]

[1]Assistant Professor, Department of Mathematics,
Washington State University, Pullman, WA

U.S.-Mexico Workshop on Optimization and Applications,
Merida, Mexico, Jan. 5, 2016
Collaborators: Kun Chen and Jeff Linderoth.

Outline

Outline

Perspective relaxation and its projection

A minimax problem and its SDP formulation

Connection with Max-Cut and Goemans-Williamson rounding

Numerical Experiments

Sparse linear regression

$$\min_{\beta \in \mathbb{R}^p} \quad \frac{1}{2}\|X\beta - y\|_2^2 + \lambda\|\beta\|_0 \qquad (\ell_2\text{-}\ell_0)$$

where $\|\beta\|_0 := \#\{i : \beta_i \neq 0\}$.

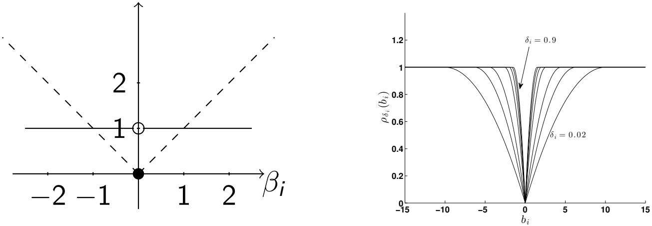
$$\min_{\beta \in \mathbb{R}^p} \quad \frac{1}{2}\left\| \begin{bmatrix} & X & \\ & & \\ & & \end{bmatrix}_{n\times p} \begin{bmatrix} \beta \\ \\ \end{bmatrix}_{p\times 1} - \begin{bmatrix} y \\ \\ \\ \end{bmatrix}_{n\times 1} \right\|_2^2 + \lambda\|\beta\|_0$$

- Each **row** of X and corresponding entry of y is a sample of predictor and response variables;
- Quadratic form $\frac{1}{n}X^TX$ is the empirical covariance matrix of predictor random variables; (if independence among predictor vars is assumed, $\frac{1}{n}X^TX \mapsto$ diagonal, as $n \mapsto +\infty$.)

Large literature on penalty functions

$$\min_{\beta} \quad \frac{1}{2}\|X\beta - y\|_2^2 + \sum_i \rho(\beta_i; \lambda, \delta);$$

where δ is some other parameter that controls concavity, etc.



(a) $\rho(\beta_i; \lambda, \delta) = \lambda|\beta_i|$ (b) Minimax Concave Penalty

Figure: Penalty functions

We are interested in constructions in some *lifted space* and their *projected form*.

Some previous work in optimization community

- Bienstock, D.: Computational study of a family of mixed-integer quadratic programming problems. *Mathematical Programming, Series A* 74(2), 121–140 (1996)
- Bertsimas, D., Shioda, R.: Algorithm for cardinality-constrained quadratic optimization. *Computational Optimization and Applications* 43(1), 1–22 (2009)
- Zheng, X.J., Sun, X.L., and Li, D. Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: a semidefinite program approach, *Informs Journal on Computing*, http://dx.doi.org/10.1287/ijoc.2014.0592, (2014)
- Bertsimas, D., King, A., Mazumder, R.: Best subset selection via a modern optimization lens. submitted to *Annals of Statistics* (2014)
- M. Feng, J. E. Mitchell, J.-S. Pang, X. Shen, and A. Wächter Complementarity Formulations of l0-norm Optimization Problems, Technical Report, Sept. 2013.
- Pilanci, M., Wainwright, M.J., Ghaoui, L.E.: Sparse learning via Boolean relaxations. *Mathematical Programming (Series B)* 151, 63–87 (2015)
- ......

Binary indicator variables and perspective set

$$z_i = \mathbb{I}_{\beta_i \neq 0} := \begin{cases} 1 & \text{if } \beta_i \neq 0, \\ 0 & \text{if } \beta_i = 0. \end{cases}$$

Need big-M to formulate as MIQP in the original variable space.

Binary indicator variables and perspective set

$$z_i = \mathbb{I}_{\beta_i \neq 0} := \begin{cases} 1 & \text{if } \beta_i \neq 0, \\ 0 & \text{if } \beta_i = 0. \end{cases}$$

Need big-M to formulate as MIQP in the original variable space.
However, with lifted variables $s_i \leftrightarrow \beta_i^2$ we have

$$\textbf{conv} \left\{ (\beta_i, s_i, z_i) \middle| s_i = \beta_i^2, z_i = \mathbb{I}_{\beta_i \neq 0} \right\} = \\ \left\{ (\beta_i, s_i, z_i) \middle| s_i z_i \geq \beta_i^2, s_i \geq 0, 0 \leq z_i \leq 1 \right\}.$$

Binary indicator variables and perspective set

$$z_i = \mathbb{I}_{\beta_i \neq 0} := \begin{cases} 1 & \text{if } \beta_i \neq 0, \\ 0 & \text{if } \beta_i = 0. \end{cases}$$

Need big-M to formulate as MIQP in the original variable space.
However, with lifted variables $s_i \leftrightarrow \beta_i^2$ we have

$$\textbf{conv}\left\{(\beta_i, s_i, z_i)\middle| s_i = \beta_i^2, z_i = \mathbb{I}_{\beta_i \neq 0}\right\} = \left\{(\beta_i, s_i, z_i)\middle| s_i z_i \geq \beta_i^2, s_i \geq 0, 0 \leq z_i \leq 1\right\}.$$

Perspective relaxation by diagonal splitting ($\delta_i \in \mathbb{R}_+^p$ s.t. $X^T X - \textbf{diag}(\delta) \succeq 0$)

$$\min_{b,z} \ \frac{1}{2} b^T (X^T X - \textbf{diag}(\delta)) b - (X^T y)^T b + \frac{1}{2} \sum_i \delta_i s_i + \lambda \sum_i z_i$$

$$s.t., \ \ s_i z_i \geq b_i^2, \ s_i \geq 0, \ 0 \leq z_i \leq 1 \ \ \forall i.$$

Fully solves the ℓ_2-ℓ_0 problem if $X^T X$ were diagonal.

Assumption: $X^T X \succ 0$

- In order for our relaxations later to be meaningful, we assume the quadratic form in our objective function $X^T X$ is positive definite, (e.g. more data points than dimension of β).
- If this is not the case, (e.g. $p > n$), an additional regularization term $\mu \|\beta\|_2^2$ must be added. In statistics, this technique is called "stabilization", and is the basic idea of elastic net.

Perspective relaxation: the equivalent projected form

$$\min_b \ \frac{1}{2}\|Xb - y\|_2^2 + \sum_i \rho_{\delta_i,\lambda}(b_i), \qquad (PR_\delta : reg)$$

where

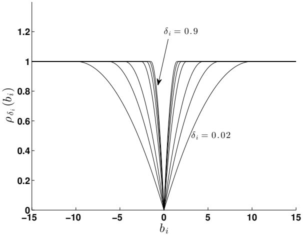
$$\rho_{\delta_i,\lambda}(b_i) = \min_{s_i z_i \geq b_i^2, s_i \geq 0, z_i \in [0,1]} \frac{1}{2}\delta_i \left(s_i - b_i^2\right) + \lambda z_i.$$

Can find explicit form of $\rho_{\delta_i,\lambda}(b_i)$.

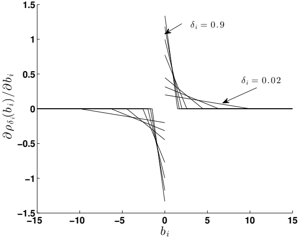
$$\rho_{\delta_i,\lambda}(b_i) = \begin{cases} \sqrt{2\delta_i\lambda}|b_i| - \frac{1}{2}\delta_i b_i^2, & \textit{if } \delta_i b_i^2 \leq 2\lambda; \\ \lambda, & \textit{if } \delta_i b_i^2 > 2\lambda. \end{cases} \qquad \text{(PR:penalty)}$$

Concave in terms of b_i on $[0, +\infty)$, δ_i controls "concavity". Also concave in terms of δ_i for fixed b_i.

Rediscovery of Minimax Concave Penalty



(a) Penalty function



(b) Penalty gradient

In [Zhang, 2010], MCP is intuitively constructed by: find a C^1 function on $[0, +\infty)$

- has positive direction derivative at $0+$;
- becomes flat after a threshold;
- minimize the max concavity.
- all δ_i are the same, and this parameter is tuned by some heuristics.

A convex relaxation in [Pilanci, Wainwright & Ghaoui, 2015]

$$\min_{\beta} \frac{1}{2}\|X\beta - y\|_2^2 + \rho\|\beta\|_2^2 + \lambda\|\beta\|_0$$

Using Fenchel conjugacy, [Pilanci et al., 2015] derived convex relaxation:

$$\min_{\beta} \frac{1}{2}\|X\beta - y\|_2^2 + 2\lambda \sum_i H\left(\frac{\sqrt{\rho}\beta_i}{\sqrt{\lambda}}\right)$$

where $H(t)$ is called reverse Huber penalty

$$H(t) = \begin{cases} |t| & if |t| \leq 1 \\ \frac{t^2+1}{2}, & otherwise \end{cases}.$$

Derivation using perspective relaxation

$$\min_{\beta} \frac{1}{2}\|X\beta - y\|_2^2 + \rho\|\beta\|_2^2 + \lambda\|\beta\|_0$$

$$\min_{\beta} \frac{1}{2}\|X\beta - y\|_2^2 + \rho B_{ii} + \lambda z_i, \ s.t. \ B_{ii} z_i \geq \beta_i^2, z_i \in [0,1]$$

$$\min_{\beta} \frac{1}{2}\|X\beta - y\|_2^2 + \min_{z_i \in [0,1]} \rho \frac{\beta_i^2}{z_i} + \lambda z_i$$

where

$$\min_{z_i \in [0,1]} \rho \frac{\beta_i^2}{z_i} + \lambda z_i = \begin{cases} 2\sqrt{\rho\lambda}|\beta_i| & if \sqrt{\frac{\rho}{\lambda}}|\beta_i| \leq 1 \\ \frac{1}{2}\rho\beta_i^2 + \lambda, & otherwise \end{cases} = 2\lambda H(\frac{\sqrt{\rho}\beta_i}{\sqrt{\lambda}})$$

[Pilanci et. al., 2015] also proposed a convex relaxation (an SDP) for ℓ_0 constrained case, which can also be equivalently derived by perspective relaxation in a constrained form.

Parameter selection for general perspective relaxation

Recall the diagonal splitting form,

$$\min_{b,z} \quad \frac{1}{2}b^T(X^TX - \mathbf{diag}(\delta))b - (X^Ty)^Tb + \frac{1}{2}\sum_i \delta_i s_i + \lambda \sum_i z_i$$

$$s.t., \quad s_i z_i \geq b_i^2, \; s_i \geq 0, \; 0 \leq z_i \leq 1 \quad \forall i.$$

δ is a $p \times 1$ vector. How do we choose parameter vector δ given this large degree of freedom?

Parameter selection for general perspective relaxation

Recall the diagonal splitting form,

$$\min_{b,z} \ \frac{1}{2} b^T (X^T X - \mathbf{diag}(\delta)) b - (X^T y)^T b + \frac{1}{2} \sum_i \delta_i s_i + \lambda \sum_i z_i$$

$$s.t., \ \ s_i z_i \geq b_i^2, \ s_i \geq 0, \ 0 \leq z_i \leq 1 \ \ \forall i.$$

δ is a $p \times 1$ vector. How do we choose parameter vector δ given this large degree of freedom?

Intuition: It is a convex relaxation iff $X^T X - \mathbf{diag}(\delta) \succeq 0$.

- Want δ "large" such that $X^T X - \mathbf{diag}(\delta)$ has zero eigenvalues, however such δ is not unique;

Outline

A Minimax formulation

$$\inf_b \frac{1}{2}\|Xb - y\|_2^2 + \sup_{\delta \in \mathbb{R}_+^p} \left\{ \sum_i \rho_{\delta_i,\lambda}(b_i) \middle| X^T X - \mathbf{diag}(\delta) \succeq 0 \right\}. \tag{Inf-Sup}$$

or equivalently

$$\inf_b \sup_{\delta \in \mathbb{R}_+^p} \left\{ \frac{1}{2}\|Xb - y\|_2^2 + \sum_i \rho_{\delta_i,\lambda}(b_i) \middle| X^T X - \mathbf{diag}(\delta) \succeq 0 \right\}. \tag{Inf-Sup}$$

Interpretation:

- Use pointwise supremum of all penalty functions that maintains convexity;
- As sup of convex functions is convex, outer minimization is still a convex problem.

Max-min problem

$$\sup_{\delta\in\mathbb{R}^p_+} \inf_b \left\{ \frac{1}{2}\|Xb-y\|_2^2 + \sum_i \rho_{\delta_i,\lambda}(b_i) \middle| X^TX - \mathbf{diag}(\delta) \succeq 0 \right\}. \tag{Sup-Inf}$$

Interpretation:

- Inner minimization problem is always a convex relaxation for (ℓ_2-ℓ_0);
- How to choose the parameter vector δ to maximize the lower bound?

Max-min problem

$$\sup_{\delta \in \mathbb{R}^p_+} \inf_b \left\{ \frac{1}{2}\|Xb - y\|_2^2 + \sum_i \rho_{\delta_i,\lambda}(b_i) \middle| X^TX - \mathbf{diag}(\delta) \succeq 0 \right\}. \tag{Sup-Inf}$$

Interpretation:

- Inner minimization problem is always a convex relaxation for (ℓ_2-ℓ_0);
- How to choose the parameter vector δ to maximize the lower bound?

In literature of perspective relaxation, e.g. [Billionnet and Elloumi, 2007] or [Zheng, Sun and Li, 2014], this "tightest lower bound" can be computed using a semidefinite relaxation.

- We show it is also the case here, an SDP relaxation computes a saddle point for (Inf-Sup) and (Sup-Inf);
- All "sup" and "inf" are *attained* given $X^TX \succ 0$.

Existence of saddle point

Let $C = \left\{\delta \in \mathbb{R}_+^p \middle| X^T X - \mathbf{diag}(\delta) \succeq 0\right\}$, $D := \mathbb{R}^p$

$$K(\delta, b) := \begin{cases} \frac{1}{2}\|Xb - y\|_2^2 + \sum_i \rho_{\delta_i,\lambda}(b_i), & \forall \delta \in C \\ -\infty, & \forall \delta \notin C \end{cases}. \qquad (1)$$

Theorem (Theorem 37.6 in *Convex Analysis*, Rockafellar)

Let $K(\delta, b)$ be a closed proper concave-convex function with effective domain $C \times D$. If both of the following conditions,

1. *The convex functions $K(\delta, \cdot)$ for $\delta \in \mathbf{ri}(C)$ have no common direction of recession;*
2. *The convex functions $-K(\cdot, b)$ for $b \in \mathbf{ri}(D)$ have no common direction of recession;*

are satisfied, then K has a saddle-point in $C \times D$. In other words, there exists $(\delta^*, b^*) \in C \times D$, such that

$$\inf_{b \in D} \sup_{\delta \in C} K(\delta, b) = \sup_{\delta \in C} \inf_{b \in D} K(\delta, b) = K(\delta^*, b^*).$$

SDP relaxation - primal and dual form

$$\begin{aligned} \min_{b,z,B} \quad & \frac{1}{2}\left\langle X^T X, B\right\rangle - y^T X b + \lambda \sum_i z_i, \\ s.t. \; & \begin{pmatrix} 1 & b^T \\ b & B \end{pmatrix} \succeq 0, \begin{pmatrix} z_i & b_i \\ b_i & B_{ii} \end{pmatrix} \succeq 0, \quad \forall i. \end{aligned} \tag{SDP}$$

$$\begin{aligned} \max_{\epsilon\in\mathbb{R},\delta,t\in\mathbb{R}^p} \quad & -\frac{1}{2}\epsilon \\ s.t. \quad & \begin{pmatrix} \epsilon & -y^T X - t^T \\ -X^T y - t & X^T X - \mathbf{diag}(\delta) \end{pmatrix} \succeq 0, \\ & \begin{pmatrix} 2\lambda & t_i \\ t_i & \delta_i \end{pmatrix} \succeq 0, \forall i, \end{aligned} \tag{DSDP}$$

Strong duality holds by strict feasibility of (SDP).

SDP solves the minimax pair

Theorem

Assume $X^T X \succ 0$, a saddle point for the minimax pair (Inf-Sup) and (Sup-Inf) can be obtained by solving (SDP) and (DSDP). Let (b^*, z^*, B^*) and $(\epsilon^*, \delta^*, t^*)$ be optimal solutions to (SDP) and (DSDP) respectively, then (δ^*, b^*) is a saddle point for (Inf-Sup) and (Sup-Inf).

- Goal:
$$\max_{\delta \in C} K(\delta, b^*) = \zeta_{SDP} = \min_{b \in \mathbb{R}^p} K(\delta^*, b);$$
- It suffices to show $\max_\delta \zeta_{PR(\delta)} = \zeta_{SDP} = \zeta_{PR(\delta^*)}$, both cases of $\leq$ are proved by analyzing the relaxations.

Outline

Perspective relaxation and its projection

A minimax problem and its SDP formulation

Connection with Max-Cut and Goemans-Williamson rounding

Numerical Experiments

Two level formulation of (ℓ_2-ℓ_0)

(ℓ_2-ℓ_0) is equivalent to

$$\min_{u\in\mathbb{R}^p} \min_{z\in\{0,1\}^p} \frac{1}{2}\|X\mathbf{diag}(u)z - y\|_2^2 + \lambda \sum_i z_i.$$

Reformulation

$$\min_{u\in\mathbb{R}^p} \min_{z\in\{0,1\}^p} \frac{1}{2}\left\langle Q(u), \begin{bmatrix} 1 & z^T \\ z & zz^T \end{bmatrix} \right\rangle,$$

where

$$Q(u) = \begin{bmatrix} y^T y & -y^T X\mathbf{diag}(u) \\ -\mathbf{diag}(u)X^T y & \mathbf{diag}(u)X^T X\mathbf{diag}(u) + 2\lambda I \end{bmatrix}$$

The inner problem is a quadratic program with binary variables (BQP).

Replacing the inner BQP with its SDP relaxation

$$\min_{u\in\mathbb{R}^p} \min_{z,Z} \frac{1}{2} \left\langle Q(u), \begin{bmatrix} 1 & z^T \\ z & Z \end{bmatrix} \right\rangle, \tag{2LvlSDP}$$
$$s.t. \begin{bmatrix} 1 & z^T \\ z & Z \end{bmatrix} \succeq 0, Z_{ii} = z_i, \forall i.$$

It turns out, that this two level problem is equivalent to our relaxation (SDP). Given (b^*, B^*, z^*) to be an optimal solution to (SDP), we can recover an optimal solution to

$$u_i^* = \begin{cases} \frac{B_{ii}^*}{b_i^*} & \text{if } b_i^* \neq 0 \\ 1 & \text{if } b_i^* = 0 \end{cases}, Z_{ij}^* = \begin{cases} \frac{B_{ij}^* b_i^* b_j^*}{B_{ii}^* B_{jj}^*}, & \text{if } B_{ii} B_{jj} \neq 0, \\ 0 & \text{if } B_{ii} B_{jj} = 0 \end{cases}.$$

Then can apply the Goemans-Williamson rounding to (z^*, Z^*) to generate a binary vector $\hat{z}^{GW}$. An approximate solution to (ℓ_2-ℓ_0) is then constructed as

$$\beta_i := u_i^* \hat{z}^{GW}.$$

Outline

Perspective relaxation and its projection

A minimax problem and its SDP formulation

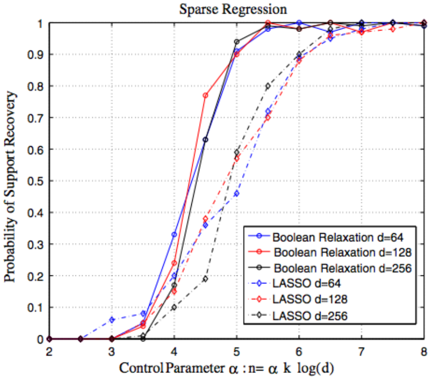
Connection with Max-Cut and Goemans-Williamson rounding

Numerical Experiments

Exact recovery rate - PWG vs. lasso

An experiment in [Pilanci, Wainwright & El Ghaoui, 2015]: generate X with i.i.d N(0,1) entries, $y = X\beta_{true} + \epsilon$, where ϵ has i.i.d. $N(0, \gamma^2)$ entries. Solve convex relaxations of

$$\min_{\beta} \quad \frac{1}{2n}\|X\beta - y\|_2^2 + \rho\|\beta\|_2^2 \quad s.t. \quad \|\beta\|_0 \leq k.$$



Directly searching for dual certificate of exact recovery

Consider a variant of (SDP)

$$\begin{aligned} \min_{b\in\mathbb{R}^p, B\in\mathcal{S}^p} \quad & \frac{1}{2}\left\langle \begin{bmatrix} y^Ty & -y^TX \\ -X^Ty & \rho I_p + X^TX \end{bmatrix}, \begin{bmatrix} 1 & b^T \\ b & B \end{bmatrix} \right\rangle \\ s.t. \quad & \begin{bmatrix} 1 & b^T \\ b & B \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} z_i & b_i \\ b_i & B_{ii} \end{bmatrix} \succeq 0, \forall i, \quad \sum_{i=1}^{p} z_i \le k. \end{aligned} \qquad (SDP_{cons})$$

Use the KKT conditions to derive a bisection search to search for a dual certificate that a solution (b^*, B^*, z^*) where z^* is a binary vector corresponding to the correct support of β_{true}. If such a certificate found, then b^* solves (ℓ_2-ℓ_0).

Bisection search for dual certificate

Theorem

Let $S \subseteq \{1, \ldots, n\}$, $|S| = k$ and z^* be a binary vector such that $z_i^* = 1, \forall i \in S$ and $z_i^* = 0, \forall i \notin S$. Further let b^* be the optimal solution of the ridge regression in the restricted subspace, i.e.,

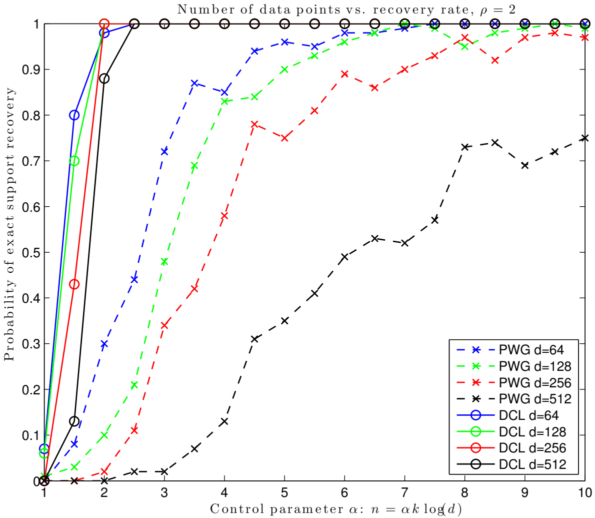
$$b^* \in \arg\min_{\beta \in \mathbb{R}^p} \left\{ \|X\beta - y\|_2^2 + \rho\|\beta\|_2^2 \;\middle|\; \beta_j = 0, \forall j \notin S \right\}$$

Then $(b^*, b^* {b^*}^T, z^*)$ is optimal to (SDP_{cons}) if and only if there is $\mu > 0$ such that

$$f(\mu) := \lambda_{\max}\left\{ \begin{bmatrix} D_S(\mu) & 0 \\ 0 & D_{\bar{S}}(\mu) \end{bmatrix} - X^T X - \rho I_p \right\} \leq 0, \qquad (2)$$

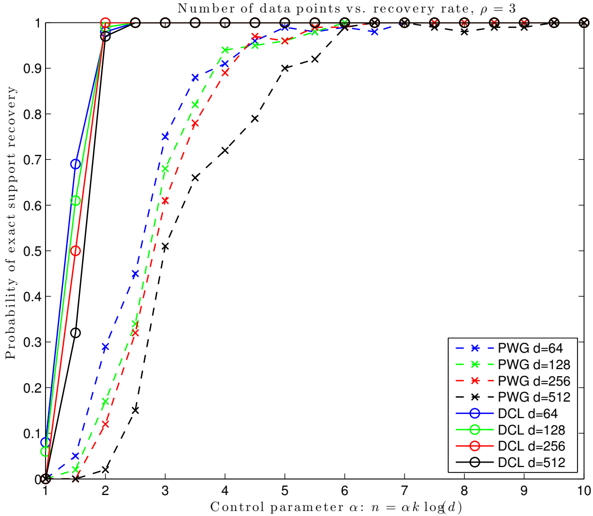
where $D_S(\mu)$ is diagonal with entries $\mu\rho^2 v_i^{-2}, i = 1, \ldots, |S|$, and similarly $D_{\bar{S}}(\lambda)$ is diagonal with entries $\mu^{-1} v_i^2, i = |S| + 1, \ldots, p$, and $v_i = X_i^T \left(\rho I + X_S X_S^T\right)^{-1} y, \forall i$

SDP v.s. PWG



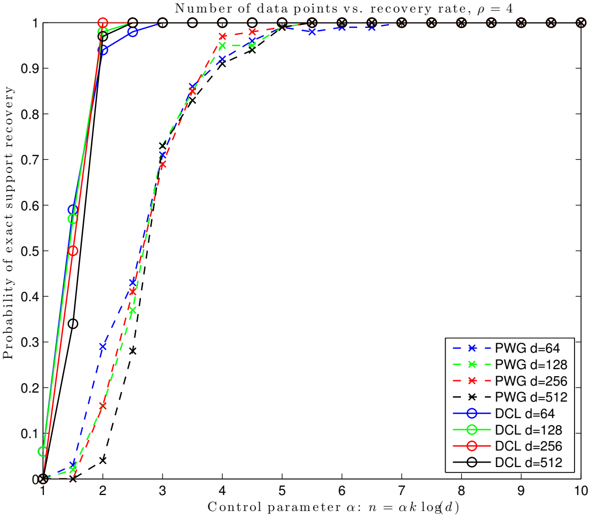
Number of data points vs. recovery rate, $\rho = 2$
Probability of exact support recovery
Control parameter α: $n = \alpha k \log(d)$
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
2
3
4
5
6
7
8
9
10
PWG d=64
PWG d=128
PWG d=256
PWG d=512
DCL d=64
DCL d=128
DCL d=256
DCL d=512

SDP v.s. PWG

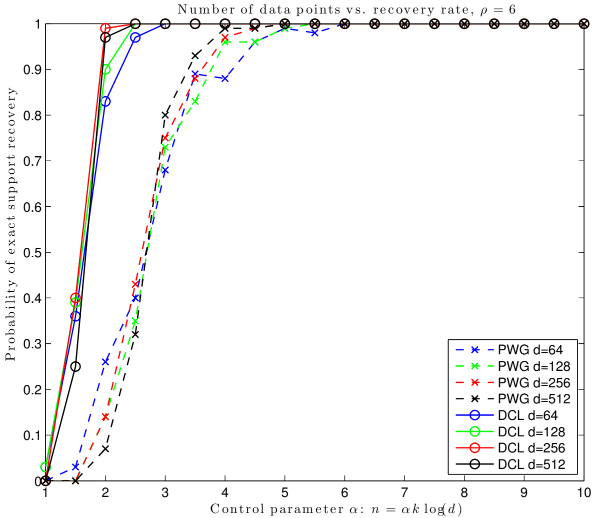


Number of data points vs. recovery rate, $\rho = 3$
Probability of exact support recovery
Control parameter α: $n = \alpha k \log(d)$
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
2
3
4
5
6
7
8
9
10
PWG d=64
PWG d=128
PWG d=256
PWG d=512
DCL d=64
DCL d=128
DCL d=256
DCL d=512

SDP v.s. PWG

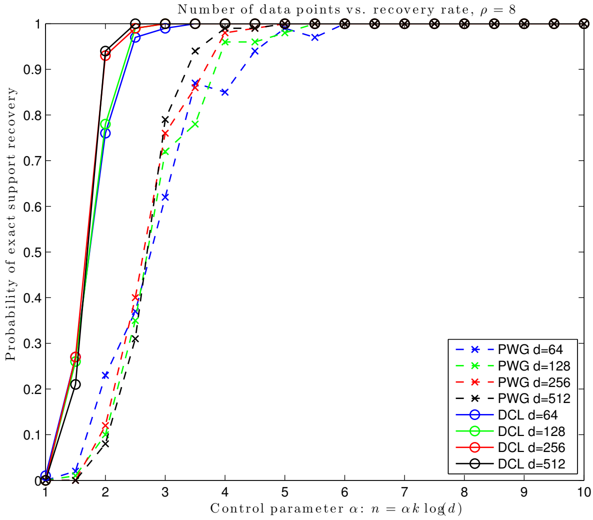

Number of data points vs. recovery rate, $\rho = 4$
Probability of exact support recovery
Control parameter α: $n = \alpha k \log(d)$
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
2
3
4
5
6
7
8
9
10
PWG d=64
PWG d=128
PWG d=256
PWG d=512
DCL d=64
DCL d=128
DCL d=256
DCL d=512

SDP v.s. PWG



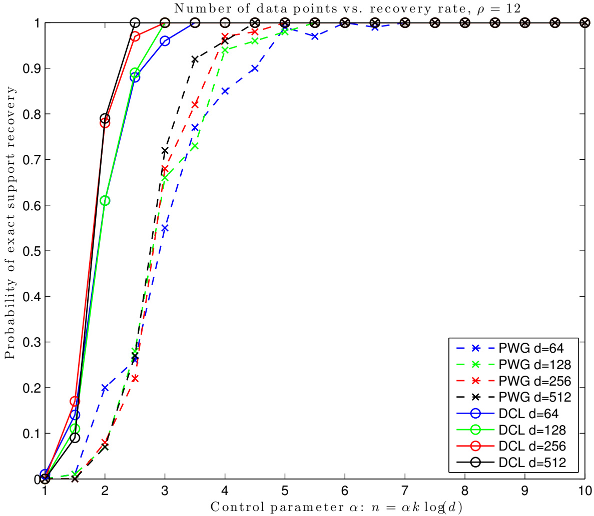
Number of data points vs. recovery rate, $\rho = 6$
Probability of exact support recovery
Control parameter α: $n = \alpha k \log(d)$
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
2
3
4
5
6
7
8
9
10
PWG d=64
PWG d=128
PWG d=256
PWG d=512
DCL d=64
DCL d=128
DCL d=256
DCL d=512

SDP v.s. PWG



Number of data points vs. recovery rate, $\rho = 8$
Probability of exact support recovery
Control parameter α: $n = \alpha k \log(d)$
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
2
3
4
5
6
7
8
9
10
PWG d=64
PWG d=128
PWG d=256
PWG d=512
DCL d=64
DCL d=128
DCL d=256
DCL d=512

SDP v.s. PWG



Number of data points vs. recovery rate, $\rho = 12$
Probability of exact support recovery
Control parameter α: $n = \alpha k \log(d)$
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1
2
3
4
5
6
7
8
9
10
PWG d=64
PWG d=128
PWG d=256
PWG d=512
DCL d=64
DCL d=128
DCL d=256
DCL d=512

Summary and Some interesting questions

- Constructions in the lifted space corresponding to some concave regularizers;
- An SDP relaxation (with moderate complexity) appears to be attractive in recovering sparse signals.

Interesting questions:

- More elegant way to handle $n < p$, i.e., when X^TX has non-trivial null space. (Patching perspective relaxation and ℓ_1 norm in different subspaces?)
- Low rank approximation to (SDP)

$$\min_{b,R} \quad \|Xb - y\|_2^2 + \langle X^TX, RR^T \rangle + \lambda \sum_j \frac{b_j^2}{b_j^2 + \ell_j^T \ell_j},$$

where R is $p \times r$ with r carefully chosen, ℓ_j is the j-th row of matrix R, $j = 1, ..., p$. For Max-Cut SDPs, [Burer, Monteiro & Zhang, 2000], [Grippo, Palagi, Piacentini, Piccialli & Rinaldi, 2010].