Penalty Decomposition Methods for Rank and l_0-Norm Minimization

Zhaosong Lu

Simon Fraser University

Joint work with Yong Zhang (Simon Fraser)

Outline of Talk

- Rank and l_0-norm minimization problems
- Technical preliminaries
- PD methods for rank minimization
- PD methods for l_0-norm minimization
- Numerical results

Rank minimization

Rank minimization:

$$\min_X \{f(X) : \ \mathrm{rank}(X) \le r, \ X \in \mathcal{X} \cap \Omega\},$$
$$\min_X \{f(X) + \nu \, \mathrm{rank}(X) : X \in \mathcal{X} \cap \Omega\}$$

for some r, $\nu \ge 0$, where $\mathcal{X}$ is a closed convex set, Ω is a closed unitarily invariant set in $\Re^{m \times n}$, and $f : \Re^{m \times n} \to \Re$ is a continuously differentiable function.

Applications:

combinatorial optimization; nonconvex QP; image recovery; nearest low-rank correlation matrix and etc.

l_0-norm minimization

l_0-norm minimization:

$$\min_x \{f(x) : \|x_J\|_0 \le r,\ x \in \mathcal{X}\},$$
$$\min_x \{f(x) + \nu\|x_J\|_0 : x \in \mathcal{X}\}$$

for some integer $r \ge 0$ and $\nu \ge 0$ controlling the sparsity of the solution, where $\mathcal{X}$ is a closed convex set in $\Re^n$, $f : \Re^n \to \Re$ is a continuously differentiable function, and $\|x_J\|_0$ denotes the cardinality of the subvector formed by the entries of x indexed by J.

Applications:

compressed sensing; sparse logistic regression; sparse inverse covariance selection and etc.

Technical preliminaries

Proposition: Let $\|\cdot\|$ be a unitarily invariant norm on $\Re^{m\times n}$, and let $F : \Re^{m\times n} \to \Re$ be a unitarily invariant function. Suppose that $\mathcal{X} \subseteq \Re^{m\times n}$ is a unitarily invariant set. Let $A \in \Re^{m\times n}$ be given, $q = \min(m, n)$, and let ϕ be a non-decreasing function on $[0, \infty)$. Suppose that $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem

$$\begin{array}{ll} \min & F(X) + \phi(\|X - A\|) \\ \text{s.t.} & X \in \mathcal{X}, \end{array}$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\begin{array}{ll} \min & F(\mathscr{D}(x)) + \phi(\|\mathscr{D}(x) - \Sigma(A)\|) \\ \text{s.t.} & \mathscr{D}(x) \in \mathcal{X}. \end{array}$$

Technical preliminaries (cont'd)

Corollary 1: Let $\nu \geq 0$ and $A \in \Re^{m\times n}$ be given, and let $q = \min(m, n)$. Suppose that $\mathcal{X} \subseteq \Re^{m\times n}$ is a unitarily invariant set, and $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem

$$\min\{\nu \operatorname{rank}(X) + \frac{1}{2}\|X - A\|_F^2 : \ X \in \mathcal{X}\},$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min\{\nu\|x\|_0 + \frac{1}{2}\|x - \sigma(A)\|_2^2 : \ \mathscr{D}(x) \in \mathcal{X}\}.$$

Technical preliminaries (cont'd)

Corollary 2: Let $r \geq 0$ and $A \in \Re^{m \times n}$ be given, and let $q = \min(m, n)$. Suppose that $\mathcal{X} \subseteq \Re^{m \times n}$ is a unitarily invariant set, and $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem

$$\min\{\|X - A\|_F : \ \mathrm{rank}(X) \leq r, \ X \in \mathcal{X}\},$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min\{\|x - \sigma(A)\|_2 : \ \|x\|_0 \leq r, \ \mathscr{D}(x) \in \mathcal{X}\}.$$

Technical preliminaries (cont'd)

Corollary 3: Let $\nu \geq 0$ and $A \in \Re^{m \times n}$ be given, and let $q = \min(m, n)$. Suppose that $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem

$$\min \nu \|X\|_* + \frac{1}{2}\|X - A\|_F^2,$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min \nu \|x\|_1 + \frac{1}{2}\|x - \sigma(A)\|_2^2.$$

Technical preliminaries (cont'd)

Corollary 4: Let $r \geq 0$ and $A \in \Re^{m \times n}$ be given, and let $q = \min(m, n)$. Suppose that $U\Sigma(A)V^T$ is the singular value decomposition of A. Then, $X^* = U\mathscr{D}(x^*)V^T$ is an optimal solution of the problem

$$\min\{\|X - A\|_F : \ \|X\|_* \leq r\},$$

where $x^* \in \Re^q$ is an optimal solution of the problem

$$\min\{\|x - \sigma(A)\|_2 : \ \|x\|_1 \leq r\}.$$

Technical preliminaries (cont'd)

Proposition: Let $\mathcal{X}_i \subseteq \Re$ and $\phi_i : \Re \to \Re$ for $i = 1, \ldots, n$ be given. Suppose that r is a positive integer and $0 \in \mathcal{X}_i$ for all i. Consider the following l_0-norm minimization problem:

$$\min \left\{ \phi(x) = \sum_{i=1}^{n} \phi_i(x_i) : \|x\|_0 \leq r,\ x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \right\}. \qquad (1)$$

Let $\tilde{x}_i^* \in \operatorname{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$ and $I^* \subseteq \{1, \ldots, n\}$ be the index set corresponding to r largest values of $\{v_i^*\}_{i=1}^n$, where $v_i^* = \phi_i(0) - \phi_i(\tilde{x}_i^*)$ for $i = 1, \ldots, n$. Then, x^* is an optimal solution of problem (1), where x^* is defined as follows:

$$x_i^* = \begin{cases} \tilde{x}_i^* & \text{if } i \in I^*; \\ 0 & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, n.$$

Technical preliminaries (cont'd)

Proposition: Let $\mathcal{X}_i \subseteq \Re$ and $\phi_i : \Re \to \Re$ for $i = 1, \ldots, n$ be given. Suppose that $\nu \geq 0$ and $0 \in \mathcal{X}_i$ for all i. Consider the following l_0-norm minimization problem:

$$\min \left\{ \nu \|x\|_0 + \sum_{i=1}^{n} \phi_i(x_i) : \quad x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \right\}. \tag{2}$$

Let $\tilde{x}_i^* \in \operatorname{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$ and $v_i^* = \phi_i(0) - \nu - \phi_i(\tilde{x}_i^*)$ for $i = 1, \ldots, n$. Then, x^* is an optimal solution of problem (2), where x^* is defined as follows:

$$x_i^* = \begin{cases} \tilde{x}_i^* & \text{if } v_i^* \geq 0; \\ 0 & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, n.$$

PD methods for rank minimization

Consider

$$\min_X \{ f(X) : \ \mathrm{rank}(X) \leq r, \ X \in \mathcal{X} \cap \Omega \}, \tag{3}$$

$$\min_X \{ f(X) + \nu \, \mathrm{rank}(X) : X \in \mathcal{X} \cap \Omega \} \tag{4}$$

for some r, $\nu \geq 0$, where $\mathcal{X}$ is a closed convex set, Ω is a closed unitarily invariant set in $\Re^{m \times n}$, and $f : \Re^{m \times n} \to \Re$ is a continuously differentiable function.

Assumption: Problems (3) and (4) are feasible, and moreover, at least a feasible solution, denoted by X^{feas}, is known.

PD methods for rank minimization

$$\min_X \{ f(X) : \ \mathrm{rank}(X) \le r, \ X \in \mathcal{X} \cap \Omega \}$$

$$\Updownarrow$$

$$\min_{X,Y} \{ f(X) : \ X - Y = 0, \ X \in \mathcal{X}, \ Y \in \mathcal{Y} \}, \tag{5}$$

where

$$\mathcal{Y} := \{ Y \in \Omega | \ \mathrm{rank}(Y) \le r \}.$$

Given $\varrho > 0$, define:

$$\begin{aligned} Q_\varrho(X,Y) &:= f(X) + \frac{\varrho}{2} \|X - Y\|_F^2, \\ \widetilde{Q}_\varrho(X,U,V) &:= Q_\varrho(X, UV) \quad \forall X \in \Re^{m \times n}, U \in \Re^{m \times r}, V \in \Re^{r \times n}. \end{aligned}$$

PD method for (5) (asymmetric matrices):

Let $\{\epsilon_k\}$ be a positive decreasing sequence. Let $\varrho_0 > 0$, $\sigma > 1$ be given. Choose an arbitrary $Y_0^0 \in \mathcal{Y}$ and a constant $\Upsilon \geq \max\{f(X^{\text{feas}}), \min_{X\in\mathcal{X}} Q_{\varrho_0}(X, Y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$ and apply the BCD method to find an approximate solution $(X^k, Y^k) \in \mathcal{X} \times \mathcal{Y}$ for the penalty subproblem:

$$\min\{Q_{\varrho_k}(X, Y) : \ X \in \mathcal{X}, \ Y \in \mathcal{Y}\} \tag{6}$$

by performing steps 1a)-1d):

1a) Solve $X_{l+1}^k \in \operatorname{Arg}\min_{X\in\mathcal{X}} Q_{\varrho_k}(X, Y_l^k)$.

1b) Solve $Y_{l+1}^k \in \operatorname{Arg}\min_{Y\in\mathcal{Y}} Q_{\varrho_k}(X_{l+1}^k, Y)$.

1c) Set $(X^k, Y^k) := (X_{l+1}^k, Y_{l+1}^k)$. If (X^k, Y^k) satisfies

$$\operatorname{dist}\left(-\nabla_X Q_{\varrho_k}(X^k, Y^k), \mathcal{N}_{\mathcal{X}}(X^k)\right) \quad \leq \quad \epsilon_k, \tag{7}$$

$$\|\nabla_U \widetilde{Q}_{\varrho_k}(X^k, U^k, V^k) + Z_Y^k (V^k)^T\|_F \quad \leq \quad \epsilon_k, \tag{8}$$

$$\|\nabla_V \widetilde{Q}_{\varrho_k}(X^k, U^k, V^k) + (U^k)^T Z_Y^k\|_F \quad \leq \quad \epsilon_k \tag{9}$$

for some $Z_Y^k \in \mathcal{N}_\Omega(Y^k)$, $U^k \in \Re^{m\times r}$, $V^k \in \Re^{r\times n}$ such that

$$(U^k)^T U^k = I, \quad Y^k = U^k V^k, \tag{10}$$

then go to step 2).

1d) Set $l \leftarrow l+1$ and go to step 1a).

2) Set $\varrho_{k+1} := \sigma \varrho_k$.

3) If $\min_{X\in\mathcal{X}} Q_{\varrho_{k+1}}(X, Y^k) > \Upsilon$, set $Y_0^{k+1} := X^{\text{feas}}$. Otherwise, set $Y_0^{k+1} := Y^k$.

4) Set $k \leftarrow k+1$ and go to step 1).

end

Theorem (outer iterations): Assume that $\epsilon_k \to 0$. Let $\{(X^k, Y^k)\}$ be generated by the above PD method, and $\{(U^k, V^k, Z_Y^k)\}$ be the associated seq satisfying (7)-(10). Suppose that the level set $\mathcal{X}_\Upsilon := \{X \in \mathcal{X} | f(X) \leq \Upsilon\}$ is compact. Then:

(a) $\{(X^k, Y^k, U^k, V^k)\}$ is bounded;

(b) Suppose that $\{(X^k, Y^k, U^k, V^k)\}_{k\in K}$ converges to (X^*, Y^*, U^*, V^*). Then, (X^*, Y^*) is a feasible point of problem (5). Moreover, if

$$\left\{ \begin{pmatrix} d_X - d_U V^* - U^* d_V \\ d_U V^* + U^* d_V - d_Y \end{pmatrix} : \begin{array}{l} d_X \in \mathcal{T}_{\mathcal{X}}(X^*), d_U \in \Re^{m\times r}, \\ d_V \in \Re^{r\times n}, d_Y \in \mathcal{T}_\Omega(X^*) \end{array} \right\} = \Re^{m\times n} \times \Re^{m\times n}$$

holds, then $\{(Z_X^k, Z_Y^k)\}_{k\in K}$ is bounded, where $Z_X^k := \varrho_k(X^k - Y^k)$, and each cluster point (Z_X^*, Z_Y^*) of $\{(Z_X^k, Z_Y^k)\}_{k\in K}$ together with (X^*, U^*, V^*) satisfies

$$\begin{aligned} -\nabla f(X^*) - Z_X^* &\in \mathcal{N}_{\mathcal{X}}(X^*), \\ (Z_X^* - Z_Y^*)(V^*)^T &= 0, \\ (U^*)^T(Z_X^* - Z_Y^*) &= 0, \\ X^* - U^*V^* = 0, \quad Z_Y^* &\in \mathcal{N}_\Omega(X^*). \end{aligned}$$

PD methods for rank minimization (cont'd)

Remark: The above cluster point $(X^*, U^*, V^*, Z_X^*, Z_Y^*)$ satisfies the KKT conditions of the following reformulation of (5) (or, equivalently, (3)):

$$\min_{X,U,V} \{ f(X) : \ X - UV = 0, \ UV \in \Omega, \ X \in \mathcal{X}, \ U \in \Re^{m \times r}, \ V \in \Re^{r \times n} \}.$$

Theorem (inner iterations): Suppose that the following condition:

$$\left\{ d_U \bar{V} + \bar{U} d_V - d_Y : d_U \in \Re^{m \times r}, d_V \in \Re^{r \times n}, d_Y \in \mathcal{T}_\Omega(\bar{Y}) \right\} = \Re^{m \times n}$$

holds for any $\bar{U} \in \Re^{m \times r}$, $\bar{V} \in \Re^{r \times n}$ such that $\bar{U}^T \bar{U} = I$ and $\bar{Y} = \bar{U}\bar{V} \in \Omega$. The approximate solution $(X^k, Y^k) \in \mathcal{X} \times \mathcal{Y}$ for problem (6) satisfying (7)-(10) can be found by the BCD method described in steps 1a)-1d) within a finite number of iterations.

PD methods for rank minimization (cont'd)

Penalty decomposition method for (4):

Let $\varrho_0 > 0$, $\sigma > 1$ be given. Choose an arbitrary $Y_0^0 \in \Omega$ and a constant Υ such that $\Upsilon \geq \max\{f(X^{\rm feas}) + \nu\,{\rm rank}(X^{\rm feas}), \min_{X\in\mathcal{X}} P_{\varrho_0}(X, Y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$ and apply the BCD method to find an approximate solution $(X^k, Y^k) \in \mathcal{X} \times \Omega$ for the penalty subproblem

$$\min\{P_{\varrho_k}(X, Y) :\ X \in \mathcal{X},\ Y \in \Omega\}$$

by performing steps 1a)-1c):

1a) Solve $X_{l+1}^k \in \operatorname{Arg}\min\limits_{X\in\mathcal{X}} P_{\varrho_k}(X, Y_l^k)$.

1b) Solve $Y_{l+1}^k \in \operatorname{Arg}\min\limits_{Y\in\Omega} P_{\varrho_k}(X_{l+1}^k, Y)$.

1c) Set $l \leftarrow l + 1$ and go to step 1a).

2) Set $\varrho_{k+1} := \sigma\varrho_k$.

3) If $\min\limits_{X\in\mathcal{X}} P_{\varrho_{k+1}}(X, Y^k) > \Upsilon$, set $Y_0^{k+1} := X^{\rm feas}$. Otherwise, set $Y_0^{k+1} := Y^k$.

4) Set $k \leftarrow k + 1$ and go to step 1).

PD methods for l_0-norm minimization

Consider the l_0-norm minimization problems:

$$\min_x \{f(x) : \|x_J\|_0 \leq r,\ x \in \mathcal{X}\}, \tag{11}$$

$$\min_x \{f(x) + \nu\|x_J\|_0 : x \in \mathcal{X}\} \tag{12}$$

for some integer $r \geq 0$ and $\nu \geq 0$ controlling the sparsity of the solution, where $\mathcal{X}$ is a closed convex set in $\Re^n$, $f : \Re^n \to \Re$ is a continuously differentiable function, and $\|x_J\|_0$ denotes the cardinality of the subvector formed by the entries of x indexed by J.

Assumption: Problems (11) and (12) are feasible, and moreover, at least a feasible solution, denoted by x^{feas}, is known.

PD methods for l_0-norm minimization

For simplicity, assume $J = \{1, 2, \ldots, n\}$. Define:

$$\mathcal{X}_M = \{\mathscr{D}(x) : x \in \mathcal{X}\}, \qquad f_M(X) = f(\mathscr{D}^*(X)) \quad \forall X \in \mathcal{D}^n.$$

Observe:

$$\min_x \{f(x) : \ \|x\|_0 \le r, \ x \in \mathcal{X}\}$$
$$\Updownarrow$$
$$\min_X \{f_M(X) : \operatorname{rank}(X) \le r, \ X \in \mathcal{X}_M\}, \tag{13}$$

which can be suitably solved by the above PD method.

Define:

$$\begin{aligned}
\mathcal{Y}_M &:= \{Y \in \mathcal{S}^n | \operatorname{rank}(Y) \le r\}, \\
Q_\varrho(X, Y) &:= f_M(X) + \frac{\varrho}{2}\|X - Y\|_F^2, \\
\widetilde{Q}_\varrho(X, U, D) &:= Q_\varrho(X, UDU^T) \quad \forall X \in \mathcal{D}^n, U \in \Re^{n \times r}, D \in \mathcal{D}^r.
\end{aligned}$$

Penalty decomposition method for (13):

Let $\{\epsilon_k\}$ be a positive decreasing sequence. Let $\varrho_0 > 0$, $\sigma > 1$ be given. Choose an arbitrary $Y_0^0 \in \mathcal{Y}_M$ and a constant $\Upsilon \geq \max\{f(x^{\text{feas}}), \min_{X \in \mathcal{X}_M} Q_{\varrho_0}(X, Y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$ and apply the BCD method to find an approximate solution $(X^k, Y^k) \in \mathcal{X}_M \times \mathcal{Y}_M$ for the penalty subproblem:

$$\min\{Q_{\varrho_k}(X, Y) : \ X \in \mathcal{X}_M, \ Y \in \mathcal{Y}_M\}$$

by performing steps 1a)-1d):

1a) Solve $X_{l+1}^k \in \text{Arg} \min_{X \in \mathcal{X}_M} Q_{\varrho_k}(X, Y_l^k)$.

1b) Solve $Y_{l+1}^k \in \text{Arg} \min_{Y \in \mathcal{Y}_M} Q_{\varrho_k}(X_{l+1}^k, Y)$.

1c) Set $(X^k, Y^k) := (X_{l+1}^k, Y_{l+1}^k)$. If (X^k, Y^k) satisfies

$$\begin{aligned} \text{dist}\left(-\nabla_X Q_{\varrho_k}(X^k, Y^k), \mathcal{N}_{\mathcal{X}_M}(X^k)\right) &\leq \epsilon_k, \\ \|\nabla_U \widetilde{Q}_{\varrho_k}(X^k, U^k, D^k)\|_F &\leq \epsilon_k, \\ \|\nabla_D \widetilde{Q}_{\varrho_k}(X^k, U^k, D^k)\|_F &\leq \epsilon_k \end{aligned} \tag{14}$$

for some $U^k \in \Re^{n\times r}$, $D^k \in \mathcal{D}^r$ such that

$$(U^k)^T U^k = I, \quad Y^k = U^k D^k (U^k)^T, \tag{15}$$

then go to step 2).

1d) Set $l \leftarrow l + 1$ and go to step 1a).

2) Set $\varrho_{k+1} := \sigma \varrho_k$.

3) If $\min_{X \in \mathcal{X}_M} Q_{\varrho_{k+1}}(X, Y^k) > \Upsilon$, set $Y_0^{k+1} := \mathscr{D}(x^{\text{feas}})$. Otherwise, set $Y_0^{k+1} := Y^k$.

4) Set $k \leftarrow k + 1$ and go to step 1).

end

Theorem: Assume that $\epsilon_k \to 0$. Let $\{(X^k, Y^k, U^k, D^k)\}$ be generated by the above PD method satisfying (14) and (15). Suppose that $\mathcal{X}_\Upsilon := \{X \in \mathcal{X}_M | f_M(X) \leq \Upsilon\}$ is compact. Then:

(a) $\{(X^k, Y^k, U^k, D^k)\}$ is bounded;

(b) Suppose that $\{(X^k, Y^k, U^k, D^k)\}_{k\in K}$ converges to (X^*, Y^*, U^*, D^*). Then, $X^* = Y^*$ and X^* is a feasible point of problem (13). Moreover, if the following condition

$$\left\{ d_X - d_U D^* (U^*)^T - U^* d_D (U^*)^T - U^* D^* d_U^T : \begin{array}{l} d_X \in \mathcal{T}_{\mathcal{X}_M}(X^*), \\ d_U \in \Re^{n\times r}, d_D \in \mathcal{D}^r \end{array} \right\} \supseteq \mathcal{D}^n$$

holds, then $\{Z^k\}_{k\in K}$ is bounded, where $Z^k := \varrho_k(X^k - Y^k)$, and each cluster point Z^* of $\{Z^k\}_{k\in K}$ together with (X^*, U^*, D^*) satisfies

$$\begin{aligned} -\nabla f_M(X^*) - Z^* &\in \mathcal{N}_{\mathcal{X}_M}(X^*), \\ Z^* U^* D^* &= 0, \\ \widetilde{\mathscr{D}}\left((U^*)^T Z^* U^*\right) &= 0, \\ X^* - U^* D^* (U^*)^T &= 0. \end{aligned}$$

PD methods for l_0-norm minimization

Remark: The above cluster point (X^*, U^*, D^*, Z^*) satisfies the KKT conditions of the following reformulation of (13) (or, equivalently, (11)).

$$\min_{X,U,D} \{ f_M(X) : \ X - UDU^T = 0, \ X \in \mathcal{X}_M, \ U \in \Re^{n \times r}, \ D \in \mathcal{D}^r \}.$$

Goal: Transfer the above PD method into the one involving vector operations only.

Define:

$$\mathcal{Y} = \{ y \in \Re^n : \|y\|_0 \leq r \}, \quad q_\varrho(x, y) = f(x) + \frac{\varrho}{2} \|x - y\|_2^2 \quad \forall x, y \in \Re^n.$$

Penalty decomposition method for (11):

Let $\{\epsilon_k\}$ be a positive decreasing sequence. Let $\varrho_0 > 0$, $\sigma > 1$ be given. Choose an arbitrary $y_0^0 \in \mathcal{Y}$ and a constant $\Upsilon \geq \max\{f(x^{\text{feas}}), \min_{x\in\mathcal{X}} q_{\varrho_0}(x, y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$ and apply the BCD method to find an approximate solution $(x^k, y^k) \in \mathcal{X} \times \mathcal{Y}$ for the penalty subproblem

$$\min\{q_{\varrho_k}(x, y) : \ x \in \mathcal{X}, \ y \in \mathcal{Y}\}$$

by performing steps 1a)-1d):

1a) Solve $x_{l+1}^k \in \operatorname{Arg}\min_{x\in\mathcal{X}} q_{\varrho_k}(x, y_l^k)$.

1b) Solve $y_{l+1}^k \in \operatorname{Arg}\min_{y\in\mathcal{Y}} q_{\varrho_k}(x_{l+1}^k, y)$.

1c) Set $(x^k, y^k) := (x_{l+1}^k, y_{l+1}^k)$. If (x^k, y^k) satisfies

$$\operatorname{dist}\left(-\nabla_x q_{\varrho_k}(x^k, y^k), \mathcal{N}_{\mathcal{X}}(x^k)\right) \quad \leq \quad \epsilon_k,$$

then go to step 2).

1d) Set $l \leftarrow l + 1$ and go to step 1a).

2) Set $\varrho_{k+1} := \sigma \varrho_k$.

3) If $\min_{x \in \mathcal{X}} q_{\varrho_{k+1}}(x, y^k) > \Upsilon$, set $y_0^{k+1} := x^{\text{feas}}$. Otherwise, set $y_0^{k+1} := y^k$.

4) Set $k \leftarrow k + 1$ and go to step 1).

end

Theorem: Assume that $\epsilon_k \to 0$. Let $\{(x^k, y^k)\}$ be generated by the above PD method and $J_k = \{j_1^k, \ldots, j_r^k\}$ a set of r distinct indices such that $(y^k)_j = 0$ for all $j \notin J_k$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} | f(x) \leq \Upsilon\}$ is compact. Then:

(a) $\{(x^k, y^k)\}$ is bounded;

(b) Suppose (x^*, y^*) is a cluster point of $\{(x^k, y^k)\}$. Then, $x^* = y^*$ and x^* is a feasible point of problem (11). Moreover, there exists a subsequence K such that $\{(x^k, y^k)\}_{k \in K} \to (x^*, y^*)$ and $J_k = J^*$ for some index set J^* when $k \in K$ is sufficiently large. Furthermore, if

$$\{d_x + d_d : d_x \in \mathcal{T}_\mathcal{X}(x^*),\ d_d \in \Re^n,\ (d_d)_j = 0\ \forall j \notin J^*\} = \Re^n$$

holds, then $\{z^k := \varrho_k(x^k - y^k)\}_{k \in K}$ is bounded and each cluster point z^* of $\{z^k\}_{k \in K}$ together with x^* satisfies

$$-\nabla f(x^*) - z^* \in \mathcal{N}_\mathcal{X}(x^*), \qquad z_j^* = 0 \quad \forall j \in J^*. \tag{16}$$

Remark: The optimality condition (16) is generally stronger than one natural optimality condition for (11).

Let $I^* = \{j : x_j^* \neq 0\}$. Suppose that x^* is a local minimum of (11). Then x^* is clearly a local minimum of

$$\min_{x \in \mathcal{X}} \{f(x) : x_j = 0 \; \forall j \notin I^*\}.$$

Assume that the constraint qualification

$$\{d_x + d_d : d_x \in \mathcal{T}_{\mathcal{X}}(x^*), (d_d)_j = 0 \; \forall j \notin I^*\} = \Re^n$$

holds at x^*. Then there exists $z^* \in \Re^n$ such that

$$-\nabla f(x^*) - z^* \in \mathcal{N}_{\mathcal{X}}(x^*), \qquad z_j^* = 0 \quad \forall j \in I^*. \tag{17}$$

Clearly, when $I^* \neq J^*$, (16) is generally stronger than (17). For example, when $r = n$ and $J = \{1, \ldots, n\}$, problem (11) reduces to

$$\min_x \{f(x) : x \in \mathcal{X}\}$$

and (16) becomes $-\nabla f(x^*) \in \mathcal{N}_{\mathcal{X}}(x^*)$. But (17) clearly cannot when $I^* \neq \{1, \ldots, n\}$.

Matrix completion

Test I: recover a low-rank matrix $M \in \Re^{m \times n}$ based on a subset of entries of M.

$$\begin{aligned} \min_{X \in \Re^{m \times n}} \quad & \operatorname{rank}(X) \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \ (i,j) \in \Theta, \end{aligned}$$

where Θ is a subset of index pairs (i,j).

Instances: randomly generate 50 copies of M and Θ with $m = n = 40$, $p = 800$ for each $r = 1, \ldots, 10$ by the same procedure as described by Ma, Goldfarb and Chen (2008).

Matrix completion

Initialization:

- $X^{\text{feas}} \in m \times n$ such that $X_{ij}^{\text{feas}} = M_{ij}$ for all $(i,j) \in \Theta$ and $X_{ij}^{\text{feas}} = 0$ for all $(i,j) \notin \Theta$.
- $Y_0^0 = X^{\text{feas}}$.
- $\varrho_0 = 0.1$ and $\sigma = \sqrt{10}$.

Inner termination criterion:

$$\frac{|Q_{\varrho_k}(X_l^k, Y_l^k) - Q_{\varrho_k}(X_{l-1}^k, Y_{l-1}^k)|}{\max(|Q_{\varrho_k}(X_l^k, Y_l^k)|, 1)} \leq 10^{-7}.$$

Outer termination criterion:

$$\max_{ij} |X_{ij}^k - Y_{ij}^k| \leq 10^{-5}.$$

Matrix completion

Relative error:

$$\text{rel\_err} := \frac{\|X^* - M\|_F}{\|M\|_F},$$

where X^* is an approximate recovery for M.

As in Recht, Fazel, and Parrilo (2007) and Candés and Recht (2008), we say a matrix M is successfully recovered by X^* if the relative error is less than 10^{-3}.

Remark: The recoverabily of M depends on two ratios:

- $SR = p/(mn)$.
- $FR = r(m + n - r)/p$.

Matrix completion

Table 1: Numerical results for $m = 40$, $n = 40$ and $p = 800$

Problems		FPCA			LMaFit			PD		
Rank	FR	NS	rel_err	Time	NS	rel_err	Time	NS	rel_err	Time
1	0.0988	50	7.20e−6	2.4	50	1.12e−4	0.003	50	1.17e−5	0.1
2	0.1950	50	1.34e−5	2.4	47	1.62e−4	0.004	50	1.57e−5	1.2
3	0.2888	50	2.20e−5	2.6	31	2.14e−4	0.006	50	1.90e−5	2.4
4	0.3800	50	2.87e−5	2.6	11	2.87e−4	0.009	50	2.34e−5	2.8
5	0.4688	50	3.52e−5	2.7	0	−−	−−	50	3.27e−5	3.5
6	0.5550	50	3.98e−5	2.9	0	−−	−−	50	3.77e−5	5.4
7	0.6388	50	6.13e−5	3.0	0	−−	−−	48	5.64e−5	6.1
8	0.7200	48	1.22e−4	3.2	0	−−	−−	47	9.21e−5	11.5
9	0.7987	47	1.50e−4	3.5	0	−−	−−	40	1.31e−4	19.7
10	0.8750	33	6.24e−4	4.6	0	−−	−−	25	2.07e−4	48.5

Matrix completion

Test II: recover a high-rank matrix $M \in \Re^{n \times n}$, whose most of singular values are nearly zero, by a low-rank matrix based on a subset of entries $\{M_{ij}\}_{(i,j) \in \Theta}$.

Instances: randomly generate 50 instances for each sample ratio varying from 0.5 to 0.9 with the singular values given by $\sigma_i = i^{-4}$ for all i.

Matrix completion

Table 2: Numerical results for $n = 40$ and $\sigma_i = i^{-4}$

		FPCA		LMaFit		PD	
SR	Rank	Rank	rel_err	Rank	rel_err	Rank	rel_err
0.5	5	2.0	1.45e−2	1.0	6.65e−2	5.0	9.91e−4
0.6	5	2.0	1.42e−2	1.0	6.51e−2	5.0	9.56e−4
0.7	5	2.0	1.34e−2	1.0	6.46e−2	5.0	9.37e−4
0.8	5	2.0	1.31e−2	1.0	6.41e−2	5.0	9.25e−4
0.9	5	2.1	1.24e−2	1.0	6.39e−2	5.0	9.01e−4

Matrix completion

Test III (grayscale image inpainting problem): fill the missing pixel values of the image at given pixel locations.

(a) original image

(b) rank 40 image

(c) 50% masked original image

(d) recovered image by PD

Matrix completion

Test III (grayscale image inpainting problem): fill the missing pixel values of the image at given pixel locations.

(e) 50% masked rank 40 image

(f) recovered image by PD

(g) 6.3% masked rank 40 image

(h) recovered image by PD

Nearest low-rank correlation

Nearest low-rank correlation problem:

$$\begin{aligned} \min_{X \in S^n} \quad & \tfrac{1}{2}\|X - C\|_F^2 \\ \text{s.t.} \quad & \operatorname{diag}(X) = e, \\ & \operatorname{rank}(X) \le r, \; X \succeq 0 \end{aligned}$$

for some correlation matrix $C \in S_+^n$ and some integer $r \in [1, n]$

Nearest low-rank correlation

Table 3: Comparison of Major and PD

Problem	Major			PD		
	Iter	Obj	Time	Iter	Obj	Time
P1n100r5	243	15.0	1.4	976	15.0	8.2
P1n100r10	1698	1.9	12.6	596	1.9	7.8
P1n100r15	5581	0.5	52.5	403	0.5	7.6
P1n100r20	10667	0.2	128.4	208	0.2	5.2
P1n100r25	17430	0.1	259.9	116	0.1	2.4
P1n500r5	488	3107.0	22.9	2514	3107.2	80.7
P1n500r10	836	748.2	51.5	1220	748.2	48.4
P1n500r15	1690	270.2	137.0	804	270.2	37.3
P1n500r20	3106	123.4	329.1	581	123.4	31.5
P1n500r25	5444	65.5	722.0	480	65.5	29.4

Nearest low-rank correlation

Table 4: Comparison of Major and PD

Problem	Major			PD		
	Iter	Obj	Time	Iter	Obj	Time
P2n100r5	528	852.3	3.1	963	852.3	7.7
P2n100r10	180	356.6	1.4	573	356.6	7.1
P2n100r15	122	196.4	1.3	421	196.4	7.0
P2n100r20	105	121.0	1.2	318	121.0	6.5
P2n100r25	102	79.4	1.6	231	79.4	5.9
P2n500r5	2126	24248.5	97.8	3465	24248.5	112.3
P2n500r10	3264	11749.5	199.6	1965	11749.5	76.6
P2n500r15	5061	7584.4	409.9	1492	7584.4	70.4
P2n500r20	4990	5503.2	532.0	1216	5503.2	67.2
P2n500r25	2995	4256.0	404.1	1022	4256.0	69.2

Nearest low-rank correlation

Table 5: Comparison of Major and PD

Problem	Major			PD		
	Iter	Obj	Time	Iter	Obj	Time
P3n100r5	279	29.4	1.6	937	29.4	7.0
P3n100r10	1232	4.4	9.2	601	4.4	7.1
P3n100r15	3685	1.3	34.6	505	1.3	8.6
P3n100r20	7903	0.6	95.2	317	0.6	7.3
P3n100r25	13472	0.3	201.5	176	0.3	5.2
P3n500r5	2541	2869.3	116.4	2739	2869.4	90.4
P3n500r10	2357	981.8	144.2	1410	981.8	55.4
P3n500r15	2989	446.9	241.9	923	446.9	41.6
P3n500r20	4086	234.7	438.4	662	234.7	33.0
P3n500r25	5923	135.9	788.3	504	135.9	29.5

Nearest low-rank correlation

Table 6: Comparison of Major and PD

Problem	Major			PD		
	Iter	Obj	Time	Iter	Obj	Time
P4n100r5	387	9.5	2.2	1573	9.5	12.9
P4n100r10	2062	1.2	15.1	1977	1.2	25.7
P4n100r15	6545	0.4	61.5	1480	0.4	27.8
P4n100r20	13439	0.2	157.5	1089	0.2	27.6
P4n100r25	23470	0.1	350.9	875	0.1	28.8
P4n500r5	160	845.7	8.1	2389	846.2	79.1
P4n500r10	971	109.6	60.8	2044	109.7	80.6
P4n500r15	2674	32.8	215.4	1608	32.9	72.6
P4n500r20	5249	13.9	572.0	1322	14.0	69.0
P4n500r25	8248	7.2	1107.8	1203	7.2	74.7

Sparse logistic regression

Given n samples z^i's with p features, and n binary outcomes b_i's, let $a^i = b_i z^i$ for $i = 1, \ldots, n$.

The *average logistic loss* function is:

$$l_{\text{avg}}(v, w) := \sum_{i=1}^{n} \theta(w^T a^i + v b_i)/n$$

for $v \in \Re$ and $w \in \Re^p$, where θ is the *logistic loss* function

$$\theta(t) := \log(1 + \exp(-t)).$$

The sparse logistic regression problem:

$$\min_{v,w} \{ l_{\text{avg}}(v, w) : \|w\|_0 \leq r \},$$

where integer $r \in [1, p]$ is for controlling the sparsity of the solution.

The l_1-norm regularization problem:

$$\min_{v,w} l_{\text{avg}}(v, w) + \lambda \|w\|_1,$$

where $\lambda \geq 0$ is a regularization parameter.

Sparse logistic regression

Given any model variables (v, w) and a sample vector $z \in \Re^p$, the outcome predicted by (v, w) for z is given by

$$\phi(z) = \mathrm{sgn}(w^T z + v),$$

where

$$\mathrm{sgn}(t) = \begin{cases} +1 & \text{if } t > 0, \\ -1 & \text{otherwise.} \end{cases}$$

The error rate of (v, w) for predicting the outcomes $b_1, \ldots, b_n$:

$$\text{Error} := \left\{ \sum_{i=1}^{n} \|\phi(z^i) - b_i\|_0 / n \right\} \times 100\%.$$

Goal: compare the quality of the solutions of similar sparsity obtained by our PD (l_0) and IPM method (l_1) (Kim et al. (2007)).

Sparse logistic regression

Table 7: Computational results on three real data sets

Data	Features p	Samples n	$\lambda/\lambda_{\max}$	r	IPM l_{avg}	IPM Error (%)	PD l_{avg}	PD Error (%)
Colon	2000	62	0.5	7	0.4398	17.74	0.3588	17.74
			0.1	22	0.1326	1.61	0.0003	0
			0.05	25	0.0664	0	0.0003	0
			0.01	28	0.0134	0	0.0003	0
Ionosphere	34	351	0.5	3	0.4804	17.38	0.3389	12.25
			0.1	11	0.3062	11.40	0.2393	9.69
			0.05	14	0.2505	9.12	0.2055	8.83
			0.01	24	0.1846	6.55	0.1707	6.27
Advertisements	1430	2359	0.5	3	0.2915	12.04	0.2242	6.06
			0.1	36	0.1399	4.11	0.1068	4.24
			0.05	67	0.1042	2.92	0.0613	2.25
			0.01	197	0.0475	1.10	0.0238	0.76

Sparse logistic regression

Table 8: Average computational time on six random problems

Size	Time				
$n \times p$	$r = 0.1p$	$r = 0.3p$	$r = 0.5p$	$r = 0.7p$	$r = 0.9p$
100×1000	1.2	0.5	0.2	0.2	0.2
500×5000	25.4	12.1	7.6	5.8	4.4
1000×100	3.8	1.3	1.0	0.7	0.6
5000×500	73.0	25.5	18.8	16.7	13.1
1000×1000	21.3	6.7	4.1	3.2	3.0
5000×5000	403.5	161.4	117.9	88.8	79.7

Sparse inverse covariance selection

Sparse inverse covariance selection:

$$\begin{aligned} \max_{X \succeq 0} \quad & \log\det X - \langle \Sigma, X \rangle \\ \text{s.t.} \quad & \sum_{(i,j)\in\bar{\Omega}} \|X_{ij}\|_0 \le r, \\ & X_{ij} = 0 \quad \forall (i,j) \in \Omega, \end{aligned} \tag{18}$$

where $\bar{\Omega} = \{(i,j) : (i,j) \notin \Omega,\ i \neq j\}$, and $r \in [1, |\bar{\Omega}|]$ is some integer for controlling the sparsity of the solution.

The l_1-norm regularization:

$$\begin{aligned} \max_{X \succeq 0} \quad & \log\det X - \langle \Sigma, X \rangle - \sum_{(i,j)\in\bar{\Omega}} \rho_{ij} |X_{ij}| \\ \text{s.t.} \quad & X_{ij} = 0 \quad \forall (i,j) \in \Omega, \end{aligned} \tag{19}$$

where $\{\rho_{ij}\}_{(i,j)\in\bar{\Omega}}$ is a set of regularization parameters.

Sparse inverse covariance selection

Test I (random data): compare the solution quality of the l_0 problem and its l_1 regularization.

Data: randomly generate Σ and $\bar{\Omega}$ by the same manner as in d'Aspremont et al. (2007) and L. (2008).

Normalized entropy loss:

$$\text{Loss} := \frac{1}{p}(\langle \mathbf{\Sigma^t}, X \rangle - \log\det(\mathbf{\Sigma^t} X) - p).$$

- apply the PPA (Wang, Sun and Toh (2009)) to solve the l_1 regularization problem with $\rho_{\bar{\Omega}} = 0.01$, 0.05, 0.1 and 0.5 and obtain solution X^*;
- set $r = \sum_{(i,j)\in\bar{\Omega}} \|X^*_{ij}\|_0$ for the l_0-norm problem so that the solution by PD method is at least as sparse as X^*.

Sparse inverse covariance selection

Table 9: Computational results for $\delta = 10\%$

Problem				PPA			PD		
p	$\lvert\Omega\rvert$	$\rho_{\bar{\Omega}}$	r	Likelihood	Loss	Time	Likelihood	Loss	Time
500	56724	0.01	183876	−950.88	0.0982	31.9	−936.45	0.0694	7.3
		0.05	108418	−978.45	0.1534	34.6	−949.92	0.0963	11.5
		0.1	45018	−999.89	0.1962	40.8	−978.61	0.1537	16.1
		0.5	8602	−1020.56	0.2376	55.6	−1015.98	0.2284	58.7
1000	226702	0.01	745470	−2247.14	0.0883	147.8	−2220.47	0.0616	55.8
		0.05	459524	−2301.07	0.1422	151.9	−2245.66	0.0868	67.2
		0.1	186602	−2344.03	0.1852	155.4	−2301.11	0.1423	94.1
		0.5	42904	−2370.86	0.2120	247.3	−2358.41	0.1996	278.1
1500	509978	0.01	1686128	−3647.71	0.0941	372.2	−3607.23	0.0672	237.7
		0.05	1072854	−3731.47	0.1500	289.0	−3646.32	0.0932	249.2
		0.1	438146	−3799.02	0.1950	303.5	−3731.17	0.1498	311.8
		0.5	93578	−3832.95	0.2176	646.1	−3819.74	0.2088	718.9
2000	905240	0.01	3012206	−5177.80	0.0868	786.5	−5126.09	0.0609	484.4
		0.05	1974238	−5285.89	0.1408	565.2	−5171.87	0.0838	623.0
		0.1	822714	−5375.21	0.1855	667.1	−5282.37	0.1391	827.1
		0.5	188954	−5412.77	0.2043	1266.1	−5397.87	0.1968	1341.3

Sparse inverse covariance selection

Table 10: Computational results for $\delta = 50\%$

Problem				PPA			PD		
p	$\|\Omega\|$	$\rho_{\bar{\Omega}}$	r	Likelihood	Loss	Time	Likelihood	Loss	Time
500	37738	0.01	202226	−947.33	0.0829	34.0	−937.09	0.0624	11.2
		0.05	119536	−978.51	0.1453	31.4	−953.26	0.0948	12.9
		0.1	50118	−1001.23	0.1907	35.3	−981.95	0.1521	20.1
		0.5	16456	−1022.34	0.2329	58.3	−1005.92	0.2000	32.1
1000	152512	0.01	816070	−2225.74	0.0780	147.3	−2207.32	0.0596	76.4
		0.05	501248	−2288.28	0.1405	138.2	−2227.53	0.0875	91.0
		0.1	203686	−2335.81	0.1881	127.1	−2292.49	0.1447	93.1
		0.5	63646	−2362.87	0.2151	295.7	−2340.06	0.1923	168.0
1500	340656	0.01	1851266	−3649.78	0.0725	366.2	−3623.70	0.0551	217.2
		0.05	1178778	−3742.41	0.1342	267.2	−3660.38	0.0795	283.0
		0.1	475146	−3815.09	0.1827	308.5	−3745.00	0.1359	326.0
		0.5	76206	−3852.35	0.2075	816.7	−3845.43	0.2029	646.4
2000	605990	0.01	3301648	−5149.12	0.0729	823.2	−5116.42	0.0566	557.0
		0.05	2161490	−5272.67	0.1347	539.9	−5160.40	0.0786	595.1
		0.1	893410	−5371.26	0.1840	646.1	−5269.88	0.1333	620.3
		0.5	226194	−5409.58	0.2031	1204.3	−5382.19	0.1895	884.3

Sparse inverse covariance selection

Table 11: Computational results for $\delta = 100\%$

Problem				PPA			PD		
p	$\lvert\Omega\rvert$	$\rho_{\bar{\Omega}}$	r	Likelihood	Loss	Time	Likelihood	Loss	Time
500	0	0.01	238232	−930.00	0.0445	35.9	−914.62	0.0138	1.7
		0.05	140056	−973.57	0.1317	30.8	−940.49	0.0655	4.2
		0.1	57064	−1000.78	0.1861	36.2	−978.64	0.1418	7.4
		0.5	20968	−1022.41	0.2294	49.1	−1002.64	0.1898	12.6
1000	0	0.01	963400	−2188.06	0.0406	154.3	−2155.70	0.0083	8.0
		0.05	590780	−2276.69	0.1292	140.1	−2210.76	0.0633	21.0
		0.1	231424	−2335.09	0.1876	119.1	−2285.20	0.1378	28.4
		0.5	65682	−2363.94	0.2165	240.7	−2340.02	0.1926	49.5
1500	0	0.01	2181060	−3585.21	0.0365	369.3	−3538.11	0.0051	20.7
		0.05	1385872	−3716.91	0.1243	252.9	−3615.58	0.0568	55.3
		0.1	551150	−3806.07	0.1837	289.1	−3723.29	0.1286	67.8
		0.5	144160	−3840.95	0.2070	670.1	−3811.37	0.1873	110.7
2000	0	0.01	3892952	−5075.44	0.0341	748.9	−5014.78	0.0037	41.8
		0.05	2543142	−5248.42	0.1206	515.2	−5112.48	0.0526	102.1
		0.1	1027584	−5367.86	0.1803	609.7	−5249.32	0.1210	142.9
		0.5	196448	−5410.37	0.2015	1399.0	−5390.21	0.1914	231.3

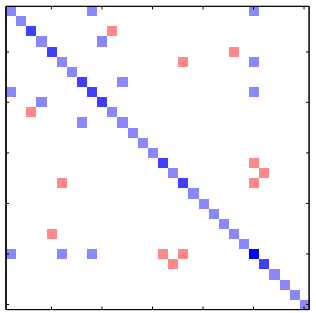
Sparse inverse covariance selection

Test II (random data): compare the sparsity recovery ability of the l_0 problem and its l_1 regularization.

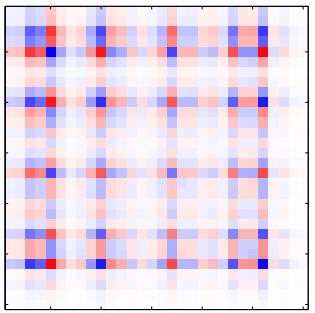
Data:

- set $p = 30$ and randomly generate true and sample covariance matrices $\mathbf{\Sigma^t}$ and Σ by the same manner as in d'Aspremont (2007);
- set $\Omega = \{(i,j) : (\mathbf{\Sigma^t})^{-1}_{ij} = 0,\ |i-j| \geq 15\}$;
- set $\rho_{ij} = \rho_{\bar{\Omega}}$ for all $(i,j) \in \bar{\Omega}$, where $\rho_{\bar{\Omega}}$ is the smallest value such that the total number of nonzero off-diagonal entries of the approximate solution obtained by the PPA when applied to the l_1 problem equals $\sum_{(i,j)\in\bar{\Omega}} \|(\mathbf{\Sigma^t})^{-1}_{ij}\|_0$;
- set $r = \sum_{(i,j)\in\bar{\Omega}} \|(\mathbf{\Sigma^t})^{-1}_{ij}\|_0$ for the l_0 problem.

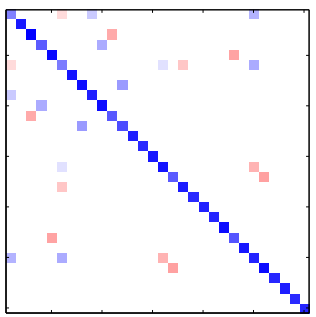
Sparse inverse covariance selection



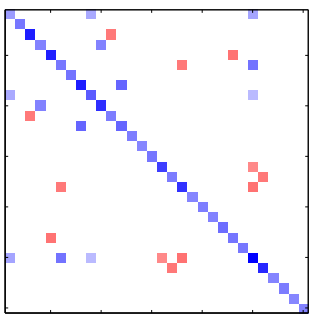
(a) Original inverse $(\mathbf{\Sigma^t})^{-1}$



(b) Noisy inverse Σ^{-1}



(c) Approx solution of (19)



(d) Approx solution of (18)

Sparse inverse covariance selection

Test III (real data): compare the solution quality of the l_0 problem and its l_1 regularization.

Data:

- pre-process two gene expression data by the same procedure as described by Li and Toh (2010) to obtain Σ;
- set $\Omega = \emptyset$ and $\rho_{ij} = \rho_{\bar{\Omega}}$ for some $\rho_{\bar{\Omega}} > 0$;
- choose r so that the solution given by the PD method when applied to the l_0 problem is at least as sparse as the one obtained by the PPA when applied to the l_1 problem.

Sparse inverse covariance selection

Table 12: Computational results on two real data sets

Data	Genes	Samples			PPA			PD		
	p	n	$\rho_{\bar{\Omega}}$	r	Likelihood	Loss	Time	Likelihood	Loss	Time
Lymph	587	148	0.01	144881	790.12	23.23	97.7	1034.99	22.95	44.2
			0.05	68061	174.86	24.35	80.9	724.16	23.27	38.2
			0.1	39091	−47.03	24.74	63.3	395.92	23.85	30.7
			0.5	5027	−561.37	25.52	27.8	−237.81	24.88	47.6
Leukemia	1255	72	0.01	250471	3229.75	0.678	681.6	4306.82	0.627	208.6
			0.05	170399	1308.38	0.769	483.5	3003.61	0.689	233.3
			0.1	108435	505.02	0.810	487.2	2541.73	0.710	251.7
			0.5	39169	−931.59	0.873	343.4	841.71	0.785	328.9