Sparse Approximation via Penalty Decomposition Methods

Zhaosong Lu

Simon Fraser University

SIAM Conference on Imaging Science / May 20, 2012

Joint work with Yong Zhang (SFU)

Outline

Some sparse decision problems

Penalty decomposition methods for l_0 minimization

Computational results

Sparse least squares regression

Consider a regression model:

$$y = f(\xi, x) + \eta.$$

Given a data set consisting of n points (ξ^i, y_i), $i = 1, \ldots, n$, the least squares model is:

$$\min_x \sum_{i=1}^{n} (y_i - f(\xi^i, x))^2.$$

The sparse least squares models are:

$$\min_x \sum_{i=1}^{n} (y_i - f(\xi^i, x))^2 + \nu \|x\|_0,$$

$$\min_x \{\sum_{i=1}^{n} (y_i - f(\xi^i, x))^2 : \|x\|_0 \leq r\}.$$

Sparse least squares regression

Consider a regression model:

$$y = f(\xi, x) + \eta.$$

Given a data set consisting of n points (ξ^i, y_i), $i = 1, \ldots, n$, the least squares model is:

$$\min_x \sum_{i=1}^n (y_i - f(\xi^i, x))^2.$$

The sparse least squares models are:

$$\min_x \sum_{i=1}^n (y_i - f(\xi^i, x))^2 + \nu \|x\|_0,$$

$$\min_x \{\sum_{i=1}^n (y_i - f(\xi^i, x))^2 : \|x\|_0 \le r\}.$$

Sparse least squares regression

Consider a regression model:

$$y = f(\xi, x) + \eta.$$

Given a data set consisting of n points (ξ^i, y_i), $i = 1, \ldots, n$, the least squares model is:

$$\min_x \sum_{i=1}^{n} (y_i - f(\xi^i, x))^2.$$

The sparse least squares models are:

$$\min_x \textstyle\sum_{i=1}^{n} (y_i - f(\xi^i, x))^2 + \nu \|x\|_0,$$

$$\min_x \{\textstyle\sum_{i=1}^{n} (y_i - f(\xi^i, x))^2 : \|x\|_0 \leq r\}.$$

Sparse logistic regression

Given scaled samples $\{a^1, \ldots, a^n\}$ and binary outcomes $b_1, \ldots, b_n$, the *average logistic loss* function is defined as

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

where θ is the *logistic loss* function

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

The *sparse logistic regression* models are:

$$\min_{v,w} l_{\text{avg}}(v, w) + \nu \|w\|_0,$$

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

Sparse logistic regression

Given scaled samples $\{a^1, \dots, a^n\}$ and binary outcomes $b_1, \dots, b_n$, the *average logistic loss* function is defined as

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

where θ is the *logistic loss* function

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

The *sparse logistic regression* models are:

$$\min_{v,w} l_{\text{avg}}(v, w) + \nu \|w\|_0,$$

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

Sparse inverse covariance selection

Given a sample covariance matrix $\hat{\Sigma}$ and a set Ω consisting of pairs of known conditionally independent nodes, the sparse inverse covariance selection models are:

$$\max_{X \succeq 0} \left\{ \log\det X - \left\langle \hat{\Sigma}, X \right\rangle - \nu \sum_{(i,j)\in\bar{\Omega}} \|X_{ij}\|_0 : X_{ij} = 0 \quad \forall (i,j) \in \Omega \right\},$$

$$\max_{X \succeq 0} \left\{ \log\det X - \left\langle \hat{\Sigma}, X \right\rangle : X_{ij} = 0 \quad \forall (i,j) \in \Omega, \sum_{(i,j)\in\bar{\Omega}} \|X_{ij}\|_0 \le r \right\},$$

where $\bar{\Omega} = \{(i,j) : (i,j) \notin \Omega,\ i \neq j\}$.

l_0 minimization

Consider

$$\min_{x \in \mathcal{X}} \{ f(x) : g(x) \leq 0,\ h(x) = 0,\ \|x_J\|_0 \leq r \}, \tag{1}$$

$$\min_{x \in \mathcal{X}} \{ f(x) + \nu \|x_J\|_0 : g(x) \leq 0,\ h(x) = 0 \}. \tag{2}$$

Assumption:

- $\mathcal{X}$ is a closed convex set.
- $f : \Re^n \to \Re$, $g : \Re^n \to \Re^m$, $h : \Re^n \to \Re^p$ are continuously differentiable.
- A feasible point $x^{\rm feas}$ is known.

l_0 minimization

Consider

$$\min_{x\in\mathcal{X}}\{f(x) : g(x) \leq 0,\ h(x) = 0,\ \|x_J\|_0 \leq r\}, \quad (1)$$

$$\min_{x\in\mathcal{X}}\{f(x) + \nu\|x_J\|_0 : g(x) \leq 0,\ h(x) = 0\}. \quad (2)$$

Assumption:

- $\mathcal{X}$ is a closed convex set.
- $f : \Re^n \to \Re$, $g : \Re^n \to \Re^m$, $h : \Re^n \to \Re^p$ are continuously differentiable.
- A feasible point $x^{\rm feas}$ is known.

l_0 minimization

Consider

$$\min_{x \in \mathcal{X}} \{f(x) : g(x) \leq 0,\ h(x) = 0,\ \|x_J\|_0 \leq r\}, \tag{1}$$

$$\min_{x \in \mathcal{X}} \{f(x) + \nu \|x_J\|_0 : g(x) \leq 0,\ h(x) = 0\}. \tag{2}$$

Assumption:

- $\mathcal{X}$ is a closed convex set.
- $f : \Re^n \to \Re$, $g : \Re^n \to \Re^m$, $h : \Re^n \to \Re^p$ are continuously differentiable.
- A feasible point x^{feas} is known.

l_0 minimization

Consider

$$\min_{x \in \mathcal{X}} \{ f(x) : g(x) \le 0,\ h(x) = 0,\ \|x_J\|_0 \le r \}, \qquad (1)$$

$$\min_{x \in \mathcal{X}} \{ f(x) + \nu \|x_J\|_0 : g(x) \le 0,\ h(x) = 0 \}. \qquad (2)$$

Assumption:

- $\mathcal{X}$ is a closed convex set.
- $f : \Re^n \to \Re$, $g : \Re^n \to \Re^m$, $h : \Re^n \to \Re^p$ are continuously differentiable.
- A feasible point x^{feas} is known.

Special l_0 minimization

Proposition

Consider

$$\min\left\{\phi(x) = \sum_{i=1}^{n} \phi_i(x_i) : \|x\|_0 \le r,\ x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n\right\}. \qquad (3)$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \operatorname{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- I^* is the index set corresp. to the r largest values of $\{\phi_i(0) - \phi_i(\tilde{x}_i^*)\}_{i=1}^n$.

Then, x^* is an optimal solution of (3), where

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

Special l_0 minimization

Proposition

Consider

$$\min\left\{\phi(x) = \sum_{i=1}^{n} \phi_i(x_i) : \|x\|_0 \le r,\ x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n\right\}. \quad (3)$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \operatorname{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- I^* is the index set corresp. to the r largest values of $\{\phi_i(0) - \phi_i(\tilde{x}_i^*)\}_{i=1}^n$.

Then, x^* is an optimal solution of (3), where

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

Special l_0 minimization

Proposition

Consider

$$\min\left\{\phi(x) = \sum_{i=1}^{n}\phi_i(x_i) : \|x\|_0 \le r,\ x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n\right\}. \tag{3}$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \mathrm{Arg}\min\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- I^* is the index set corresp. to the r largest values of $\{\phi_i(0) - \phi_i(\tilde{x}_i^*)\}_{i=1}^n$.

Then, x^* is an optimal solution of (3), where

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

Special l_0 minimization

Proposition

Consider

$$\min\left\{\phi(x) = \sum_{i=1}^{n} \phi_i(x_i) : \|x\|_0 \le r,\ x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n\right\}. \quad (3)$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \operatorname{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- I^* is the index set corresp. to the r largest values of $\{\phi_i(0) - \phi_i(\tilde{x}_i^*)\}_{i=1}^n$.

Then, x^* is an optimal solution of (3), where

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

Special l_0 minimization

Proposition

Consider

$$\min\left\{\phi(x) = \sum_{i=1}^{n} \phi_i(x_i) : \|x\|_0 \le r,\ x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n\right\}. \tag{3}$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \operatorname{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- I^* is the index set corresp. to the r largest values of $\{\phi_i(0) - \phi_i(\tilde{x}_i^*)\}_{i=1}^n$.

Then, x^* is an optimal solution of (3), where

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

Special l_0 minimization

Proposition

Consider

$$\min\left\{\nu\|x\|_0 + \sum_{i=1}^{n}\phi_i(x_i) : \; x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n\right\}. \qquad (4)$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \mathrm{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- $v_i^* = \phi_i(0) - \nu - \phi_i(\tilde{x}_i^*)$ for $i = 1, \ldots, n$.

Then, x^* is an optimal solution of (4), where

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

Special l_0 minimization

Proposition

Consider

$$\min\left\{\nu\|x\|_0 + \sum_{i=1}^{n}\phi_i(x_i) : \ x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n\right\}. \qquad (4)$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \mathrm{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- $v_i^* = \phi_i(0) - \nu - \phi_i(\tilde{x}_i^*)$ for $i = 1, \ldots, n$.

Then, x^* is an optimal solution of (4), where

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

Special l_0 minimization

Proposition

Consider

$$\min \left\{ \nu \|x\|_0 + \sum_{i=1}^{n} \phi_i(x_i) : \quad x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \right\}. \tag{4}$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \operatorname{Arg\,min}\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- $v_i^* = \phi_i(0) - \nu - \phi_i(\tilde{x}_i^*)$ for $i = 1, \ldots, n$.

Then, x^* is an optimal solution of (4), where

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

Special l_0 minimization

Proposition

Consider

$$\min \left\{ \nu \|x\|_0 + \sum_{i=1}^{n} \phi_i(x_i) : \; x \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \right\}. \tag{4}$$

Suppose that

- $0 \in \mathcal{X}_i$ for all i;
- $\tilde{x}_i^* \in \mathrm{Arg}\min\{\phi_i(x_i) : x_i \in \mathcal{X}_i\}$;
- $v_i^* = \phi_i(0) - \nu - \phi_i(\tilde{x}_i^*)$ for $i = 1, \ldots, n$.

Then, x^* is an optimal solution of (4), where

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

Special l_0 minimization

Proposition

Consider

$$\min\left\{\nu\|x\|_0 + \sum_{i=1}^{n}\phi_i(x_i):\ \ x\in\mathcal{X}_1\times\cdots\times\mathcal{X}_n\right\}. \tag{4}$$

Suppose that

- $0\in\mathcal{X}_i$ for all i;
- $\tilde{x}_i^*\in\operatorname{Arg\,min}\{\phi_i(x_i): x_i\in\mathcal{X}_i\}$;
- $v_i^* = \phi_i(0)-\nu-\phi_i(\tilde{x}_i^*)$ for $i=1,\ldots,n$.

Then, x^* is an optimal solution of (4), where

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

Necessary conditions for general l_0 min (1)

Theorem

Assume that x^* is a local minimizer of (1). Let $J^* \subseteq J$ be an index set with $|J^*| = r$ such that $x_j^* = 0$ for all $j \in \bar{J}^* := J \setminus J^*$. Suppose that the Robinson condition

$$\left\{ \begin{bmatrix} g'(x^*)d - v \\ h'(x^*)d \\ (I_{\bar{J}^*})^T d \end{bmatrix} : \begin{array}{l} d \in \mathcal{T}_{\mathcal{X}}(x^*), v \in \Re^m, \\ v_i \le 0, i \in \mathcal{A}(x^*) \end{array} \right\} = \Re^m \times \Re^p \times \Re^{|J|-r} \tag{5}$$

holds. Then, there exists (λ^*, μ^*, z^*) together with x^* satisfying KKT conditions

$$\begin{gathered} -\nabla f(x^*) - \nabla g(x^*)\lambda^* - \nabla h(x^*)\mu^* - z^* \in \mathcal{N}_{\mathcal{X}}(x^*), \\ \lambda_i^* \ge 0, \ \lambda_i^* g_i(x^*) = 0, \ i = 1, \ldots, m; \quad z_j^* = 0, \ j \in \bar{J} \cup J^*. \end{gathered} \tag{6}$$

where $\bar{J} := \{1, \ldots, n\} \setminus J$.

Necessary conditions for general l_0 min (1)

Theorem

Assume that x^* is a local minimizer of (1). Let $J^* \subseteq J$ be an index set with $|J^*| = r$ such that $x^*_j = 0$ for all $j \in \bar{J}^* := J \setminus J^*$. Suppose that the Robinson condition

$$\left\{ \begin{bmatrix} g'(x^*)d - v \\ h'(x^*)d \\ (I_{\bar{J}^*})^T d \end{bmatrix} : \begin{array}{l} d \in \mathcal{T}_{\mathcal{X}}(x^*), v \in \Re^m, \\ v_i \le 0, i \in \mathcal{A}(x^*) \end{array} \right\} = \Re^m \times \Re^p \times \Re^{|J|-r} \tag{5}$$

holds. Then, there exists (λ^*, μ^*, z^*) together with x^* satisfying KKT conditions

$$\begin{array}{c} -\nabla f(x^*) - \nabla g(x^*)\lambda^* - \nabla h(x^*)\mu^* - z^* \in \mathcal{N}_{\mathcal{X}}(x^*), \\ \lambda^*_i \ge 0, \ \lambda^*_i g_i(x^*) = 0, \ i = 1, \ldots, m; \quad z^*_j = 0, \ j \in \bar{J} \cup J^*. \end{array} \tag{6}$$

where $\bar{J} := \{1, \ldots, n\} \setminus J$.

Necessary conditions for general l_0 min (1)

Theorem

Assume that x^* is a local minimizer of (1). Let $J^* \subseteq J$ be an index set with $|J^*| = r$ such that $x_j^* = 0$ for all $j \in \bar{J}^* := J \setminus J^*$. Suppose that the Robinson condition

$$\left\{ \begin{bmatrix} g'(x^*)d - v \\ h'(x^*)d \\ (I_{\bar{J}^*})^T d \end{bmatrix} : \begin{array}{l} d \in \mathcal{T}_{\mathcal{X}}(x^*), v \in \Re^m, \\ v_i \leq 0, i \in \mathcal{A}(x^*) \end{array} \right\} = \Re^m \times \Re^p \times \Re^{|J|-r} \tag{5}$$

holds. Then, there exists (λ^*, μ^*, z^*) together with x^* satisfying KKT conditions

$$\begin{array}{c} -\nabla f(x^*) - \nabla g(x^*)\lambda^* - \nabla h(x^*)\mu^* - z^* \in \mathcal{N}_{\mathcal{X}}(x^*), \\ \lambda_i^* \geq 0, \; \lambda_i^* g_i(x^*) = 0, \; i = 1, \ldots, m; \quad z_j^* = 0, \; j \in \bar{J} \cup J^*. \end{array} \tag{6}$$

where $\bar{J} := \{1, \ldots, n\} \setminus J$.

Necessary conditions for general l_0 min (2)

Theorem

Assume that x^* is a local minimizer of (2). Let $J^* = \{j \in J : x_j^* \neq 0\}$ and $\bar{J}^* = J \setminus J^*$. suppose that the following Robinson condition

$$\left\{ \begin{bmatrix} g'(x^*)d - v \\ h'(x^*)d \\ (I_{\bar{J}^*})^T d \end{bmatrix} : \begin{array}{l} d \in \mathcal{T}_{\mathcal{X}}(x^*), v \in \Re^m, \\ v_i \leq 0, i \in \mathcal{A}(x^*) \end{array} \right\} = \Re^m \times \Re^p \times \Re^{|\bar{J}^*|} \tag{7}$$

holds. Then, there exists (λ^*, μ^*, z^*) together with x^* satisfying KKT conditions (6).

Necessary conditions for general l_0 min (2)

Theorem

Assume that x^* is a local minimizer of (2). Let $J^* = \{j \in J : x_j^* \neq 0\}$ and $\bar{J}^* = J \setminus J^*$. suppose that the following Robinson condition

$$\left\{ \begin{bmatrix} g'(x^*)d - v \\ h'(x^*)d \\ (I_{\bar{J}^*})^T d \end{bmatrix} : \begin{array}{l} d \in \mathcal{T}_{\mathcal{X}}(x^*), v \in \Re^m, \\ v_i \leq 0, i \in \mathcal{A}(x^*) \end{array} \right\} = \Re^m \times \Re^p \times \Re^{|\bar{J}^*|} \tag{7}$$

holds. Then, there exists (λ^*, μ^*, z^*) together with x^* satisfying KKT conditions (6).

Necessary conditions for general l_0 min (2)

Theorem

Assume that x^* is a local minimizer of (2). Let $J^* = \{j \in J : x_j^* \neq 0\}$ and $\bar{J}^* = J \setminus J^*$. suppose that the following Robinson condition

$$\left\{ \begin{bmatrix} g'(x^*)d - v \\ h'(x^*)d \\ (I_{\bar{J}^*})^T d \end{bmatrix} : \begin{array}{l} d \in \mathcal{T}_{\mathcal{X}}(x^*), v \in \Re^m, \\ v_i \leq 0, i \in \mathcal{A}(x^*) \end{array} \right\} = \Re^m \times \Re^p \times \Re^{|\bar{J}^*|} \tag{7}$$

holds. Then, there exists (λ^*, μ^*, z^*) together with x^* satisfying KKT conditions (6).

Sufficient conditions for general l_0 min

Theorem

Assume that h's are affine functions, and f and g's are convex functions. Let x^* be a feasible point of (1), and let $\mathcal{J}^* = \{J^* \subseteq J : |J^*| = r, x_j^* = 0, \forall j \in J \setminus J^*\}$. Suppose that for any $J^* \in \mathcal{J}^*$, there exists some (λ^*, μ^*, z^*) such that KKT conditions (6) *hold. Then, x^* is a local minimizer of (1).*

Theorem

Assume that h's are affine functions, and f and g's are convex functions. Let x^* be a feasible point of (2), and let $J^* = \{j \in J : x_j^* \neq 0\}$. Suppose that for such J^*, there exists some (λ^*, μ^*, z^*) such that KKT conditions (6) *hold. Then, x^* is a local minimizer of (2).*

Sufficient conditions for general l_0 min

Theorem

Assume that h's are affine functions, and f and g's are convex functions. Let x^* be a feasible point of (1), and let $\mathcal{J}^* = \{J^* \subseteq J : |J^*| = r, x_j^* = 0, \forall j \in J \setminus J^*\}$. Suppose that for any $J^* \in \mathcal{J}^*$, there exists some (λ^*, μ^*, z^*) such that KKT conditions (6) *hold. Then, x^* is a local minimizer of (1).*

Theorem

Assume that h's are affine functions, and f and g's are convex functions. Let x^* be a feasible point of (2), and let $J^* = \{j \in J : x_j^* \neq 0\}$. Suppose that for such J^*, there exists some (λ^*, μ^*, z^*) such that KKT conditions (6) *hold. Then, x^* is a local minimizer of (2).*

PD method for general l_0 min (1)

Observe

$$\min_{x\in\mathcal{X}}\{f(x) : g(x) \le 0,\ h(x) = 0,\ \|x_J\|_0 \le r\}$$

$$\Updownarrow$$

$$\min_{x\in\mathcal{X},y\in\mathcal{Y}}\{f(x) :\ g(x) \le 0,\ h(x) = 0,\ x_J - y = 0\}, \qquad (8)$$

where

$$\mathcal{Y} = \{y \in \Re^{|J|} : \|y\|_0 \le r\}.$$

Define

$$q_\varrho(x, y) = f(x) + \frac{\varrho}{2}(\|[g(x)]^+\|^2 + \|h(x)\|^2 + \|x_J - y\|^2) \quad \forall x, y.$$

PD method for general l_0 min (1)

Observe

$$\min_{x\in\mathcal{X}}\{f(x) : g(x) \le 0,\ h(x) = 0,\ \|x_J\|_0 \le r\}$$
$$\Updownarrow$$
$$\min_{x\in\mathcal{X},y\in\mathcal{Y}}\{f(x) :\ g(x) \le 0,\ h(x) = 0,\ x_J - y = 0\}, \qquad (8)$$

where

$$\mathcal{Y} = \{y \in \Re^{|J|} : \|y\|_0 \le r\}.$$

Define

$$q_\varrho(x, y) = f(x) + \frac{\varrho}{2}(\|[g(x)]^+\|^2 + \|h(x)\|^2 + \|x_J - y\|^2) \quad \forall x, y.$$

PD method for (1)

Choose $\varrho_0 > 0$, $\sigma > 1$, $y_0^0 \in \mathcal{Y}$,
$\Upsilon \geq \max\{f(x^{\text{feas}}), \min_{x\in\mathcal{X}} q_{\varrho_0}(x, y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$. Find an approx. solution (x^k, y^k) to

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

by performing:

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

$$\|\mathcal{P}_{\mathcal{X}}(x^k - \nabla_x q_{\varrho_k}(x^k, y^k)) - x^k\| \leq \epsilon_k,$$

then go to step 2).

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

PD method for (1)

Choose $\varrho_0 > 0$, $\sigma > 1$, $y_0^0 \in \mathcal{Y}$,
$\Upsilon \geq \max\{f(x^{\text{feas}}), \min_{x\in\mathcal{X}} q_{\varrho_0}(x, y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$. Find an approx. solution (x^k, y^k) to

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

by performing:

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

$$\|\mathcal{P}_{\mathcal{X}}(x^k - \nabla_x q_{\varrho_k}(x^k, y^k)) - x^k\| \leq \epsilon_k,$$

then go to step 2).

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

PD method for (1)

Choose $\varrho_0 > 0$, $\sigma > 1$, $y_0^0 \in \mathcal{Y}$,
$\Upsilon \geq \max\{f(x^{\text{feas}}), \min_{x\in\mathcal{X}} q_{\varrho_0}(x, y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$. Find an approx. solution (x^k, y^k) to

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

by performing:

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

$$\|\mathcal{P}_{\mathcal{X}}(x^k - \nabla_x q_{\varrho_k}(x^k, y^k)) - x^k\| \leq \epsilon_k,$$

then go to step 2).

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

PD method for (1)

Choose $\varrho_0 > 0$, $\sigma > 1$, $y_0^0 \in \mathcal{Y}$,
$\Upsilon \geq \max\{f(x^{\text{feas}}), \min_{x\in\mathcal{X}} q_{\varrho_0}(x, y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$. Find an approx. solution (x^k, y^k) to

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

by performing:

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

$$\|\mathcal{P}_{\mathcal{X}}(x^k - \nabla_x q_{\varrho_k}(x^k, y^k)) - x^k\| \leq \epsilon_k,$$

then go to step 2).

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

PD method for (1)

Choose $\varrho_0 > 0$, $\sigma > 1$, $y_0^0 \in \mathcal{Y}$,
$\Upsilon \geq \max\{f(x^{\text{feas}}), \min_{x \in \mathcal{X}} q_{\varrho_0}(x, y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$. Find an approx. solution (x^k, y^k) to

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

by performing:

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

$$\|\mathcal{P}_{\mathcal{X}}(x^k - \nabla_x q_{\varrho_k}(x^k, y^k)) - x^k\| \leq \epsilon_k,$$

then go to step 2).

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

PD method for (1)

Choose $\varrho_0 > 0$, $\sigma > 1$, $y_0^0 \in \mathcal{Y}$,
$\Upsilon \geq \max\{f(x^{\text{feas}}), \min_{x \in \mathcal{X}} q_{\varrho_0}(x, y_0^0)\}$. Set $k = 0$.

1) Set $l = 0$. Find an approx. solution (x^k, y^k) to

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

by performing:

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

$$\|\mathcal{P}_{\mathcal{X}}(x^k - \nabla_x q_{\varrho_k}(x^k, y^k)) - x^k\| \leq \epsilon_k,$$

then go to step 2).

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

PD method for (1)

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

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

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

end

PD method for (1)

2) Set $\varrho_{k+1} := \sigma \varrho_k$.
3) If $\min_{x \in \mathcal{X}} q_{\varrho_{k+1}}(x, y^k) > \Upsilon$, set $y_0^{k+1} := x^{\mathrm{feas}}$. Otherwise, set $y_0^{k+1} := y^k$.
4) Set $k \leftarrow k + 1$ and go to step 1).

end

PD method for (1)

2) Set $\varrho_{k+1} := \sigma \varrho_k$.
3) If $\min_{x \in \mathcal{X}} q_{\varrho_{k+1}}(x, y^k) > \Upsilon$, set $y_0^{k+1} := x^{\rm feas}$. Otherwise, set $y_0^{k+1} := y^k$.
4) Set $k \leftarrow k + 1$ and go to step 1).

end

Convergence theorem for (1)

Assume that $\epsilon_k \to 0$. Let $I_k = \{i_1^k, \ldots, i_r^k\}$ be a set of r distinct indices in $\{1, \ldots, |J|\}$ such that $(y^k)_i = 0$ for any $i \notin I_k$, and let $J_k = \{J(i) : i \in I_k\}$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \leq \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of (1). Moreover, there exists a subsequence K such that $\{(x^k, y^k)\}_{k \in K} \to (x^*, y^*)$, $I_k = I^*$ and $J_k = J^* := \{J(i) : i \in I^*\}$ for some index set $I^* \subseteq \{1, \ldots, |J|\}$ when $k \in K$ is sufficiently large.

(c) Let x^*, K and J^* be defined above, and let $\bar{J}^* = J \setminus J^*$. Suppose that the Robinson condition (5) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ is bounded, where

$$\lambda^k = \varrho_k [g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k (x_J^k - y^k).$$

Convergence theorem for (1)

Assume that $\epsilon_k \to 0$. Let $I_k = \{i_1^k, \ldots, i_r^k\}$ be a set of r distinct indices in $\{1, \ldots, |J|\}$ such that $(y^k)_i = 0$ for any $i \notin I_k$, and let $J_k = \{J(i) : i \in I_k\}$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \le \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of (1). Moreover, there exists a subsequence K such that $\{(x^k, y^k)\}_{k \in K} \to (x^*, y^*)$, $I_k = I^*$ and $J_k = J^* := \{J(i) : i \in I^*\}$ for some index set $I^* \subseteq \{1, \ldots, |J|\}$ when $k \in K$ is sufficiently large.

(c) Let x^*, K and J^* be defined above, and let $\bar{J}^* = J \setminus J^*$. Suppose that the Robinson condition (5) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ is bounded, where

$$\lambda^k = \varrho_k [g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k (x_J^k - y^k).$$

Convergence theorem for (1)

Assume that $\epsilon_k \to 0$. Let $I_k = \{i_1^k, \ldots, i_r^k\}$ be a set of r distinct indices in $\{1, \ldots, |J|\}$ such that $(y^k)_i = 0$ for any $i \notin I_k$, and let $J_k = \{J(i) : i \in I_k\}$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \leq \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of (1). Moreover, there exists a subsequence K such that $\{(x^k, y^k)\}_{k \in K} \to (x^*, y^*)$, $I_k = I^*$ and $J_k = J^* := \{J(i) : i \in I^*\}$ for some index set $I^* \subseteq \{1, \ldots, |J|\}$ when $k \in K$ is sufficiently large.

(c) Let x^*, K and J^* be defined above, and let $\bar{J}^* = J \setminus J^*$. Suppose that the Robinson condition (5) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ is bounded, where

$$\lambda^k = \varrho_k [g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k (x_{\bar{J}}^k - y^k).$$

Convergence theorem for (1)

Assume that $\epsilon_k \to 0$. Let $I_k = \{i_1^k, \dots, i_r^k\}$ be a set of r distinct indices in $\{1, \dots, |J|\}$ such that $(y^k)_i = 0$ for any $i \notin I_k$, and let $J_k = \{J(i) : i \in I_k\}$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \leq \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of (1). Moreover, there exists a subsequence K such that $\{(x^k, y^k)\}_{k \in K} \to (x^*, y^*)$, $I_k = I^*$ and $J_k = J^* := \{J(i) : i \in I^*\}$ for some index set $I^* \subseteq \{1, \dots, |J|\}$ when $k \in K$ is sufficiently large.

(c) Let x^*, K and J^* be defined above, and let $\bar{J}^* = J \setminus J^*$. Suppose that the Robinson condition (5) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ is bounded, where

$$\lambda^k = \varrho_k [g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k (x_{\bar{J}}^k - y^k).$$

Convergence theorem for (1) (cont'd)

Moreover, each limit point $(\lambda^*, \mu^*, \varpi^*)$ of $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ together with x^* satisfies the KKT conditions (6) with $z_j^* = \varpi_i^*$ for all $j = J(i) \in \bar{J}^*$.

(d) Further, if $\|x_{\bar{J}}^*\|_0 = r$, h's are affine functions, and f and g's are convex functions, then x^* is a local minimizer of (1).

Convergence theorem for (1) (cont'd)

Moreover, each limit point $(\lambda^*, \mu^*, \varpi^*)$ of $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ together with x^* satisfies the KKT conditions (6) with $z_j^* = \varpi_i^*$ for all $j = J(i) \in \bar{J}^*$.

(d) Further, if $\|x_J^*\|_0 = r$, h's are affine functions, and f and g's are convex functions, then x^* is a local minimizer of (1).

Convergence theorem for (2)

Assume that $\epsilon_k \to 0$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \le \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of problem (2).

(c) Let (x^*, y^*) be defined above. Suppose that $\{(x^k, y^k)\}_{k\in K} \to (x^*, y^*)$ for some subsequence K. Let $J^* = \{j \in J : x_j^* \neq 0\}$, $\bar{J}^* = J \setminus J^*$. Assume that the Robinson condition (7) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ is bounded, where

$$\lambda^k = \varrho_k [g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k (x_J^k - y^k).$$

Moreover, each limit point $(\lambda^*, \mu^*, \varpi^*)$ of $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ together with x^* satisfies the KKT conditions (6) with $z_j^* = \varpi_i^*$ for all $j = J(i) \in \bar{J}^*$.

(d) Further, if h's are affine functions, and f and g's are convex functions, then x^* is a local minimizer of (2).

Convergence theorem for (2)

Assume that $\epsilon_k \to 0$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \leq \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of problem (2).

(c) Let (x^*, y^*) be defined above. Suppose that $\{(x^k, y^k)\}_{k \in K} \to (x^*, y^*)$ for some subsequence K. Let $J^* = \{j \in J : x_j^* \neq 0\}$, $\bar{J}^* = J \setminus J^*$. Assume that the Robinson condition (7) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ is bounded, where

$$\lambda^k = \varrho_k[g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k(x_J^k - y^k).$$

Moreover, each limit point $(\lambda^*, \mu^*, \varpi^*)$ of $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ together with x^* satisfies the KKT conditions (6) with $z_j^* = \varpi_i^*$ for all $j = J(i) \in \bar{J}^*$.

(d) Further, if h's are affine functions, and f and g's are convex functions, then x^* is a local minimizer of (2).

Convergence theorem for (2)

Assume that $\epsilon_k \to 0$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \le \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of problem (2).

(c) Let (x^*, y^*) be defined above. Suppose that $\{(x^k, y^k)\}_{k\in K} \to (x^*, y^*)$ for some subsequence K. Let $J^* = \{j \in J : x_j^* \neq 0\}$, $\bar{J}^* = J \setminus J^*$. Assume that the Robinson condition (7) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ is bounded, where

$$\lambda^k = \varrho_k[g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k(x_J^k - y^k).$$

Moreover, each limit point $(\lambda^*, \mu^*, \varpi^*)$ of $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ together with x^* satisfies the KKT conditions (6) with $z_j^* = \varpi_i^*$ for all $j = J(i) \in \bar{J}^*$.

(d) Further, if h's are affine functions, and f and g's are convex functions, then x^* is a local minimizer of (2).

Convergence theorem for (2)

Assume that $\epsilon_k \to 0$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \leq \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of problem (2).

(c) Let (x^*, y^*) be defined above. Suppose that $\{(x^k, y^k)\}_{k\in K} \to (x^*, y^*)$ for some subsequence K. Let $J^* = \{j \in J : x_j^* \neq 0\}$, $\bar{J}^* = J \setminus J^*$. Assume that the Robinson condition (7) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ is bounded, where

$$\lambda^k = \varrho_k[g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k(x_J^k - y^k).$$

Moreover, each limit point $(\lambda^*, \mu^*, \varpi^*)$ of $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ together with x^* satisfies the KKT conditions (6) with $z_j^* = \varpi_i^*$ for all $j = J(i) \in \bar{J}^*$.

(d) Further, if h's are affine functions, and f and g's are convex functions, then x^* is a local minimizer of (2).

Convergence theorem for (2)

Assume that $\epsilon_k \to 0$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \leq \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of problem (2).

(c) Let (x^*, y^*) be defined above. Suppose that $\{(x^k, y^k)\}_{k \in K} \to (x^*, y^*)$ for some subsequence K. Let $J^* = \{j \in J : x_j^* \neq 0\}$, $\bar{J}^* = J \setminus J^*$. Assume that the Robinson condition (7) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ is bounded, where

$$\lambda^k = \varrho_k [g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k (x_J^k - y^k).$$

Moreover, each limit point $(\lambda^*, \mu^*, \varpi^*)$ of $\{(\lambda^k, \mu^k, \varpi^k)\}_{k \in K}$ together with x^* satisfies the KKT conditions (6) with $z_j^* = \varpi_i^*$ for all $j = J(i) \in \bar{J}^*$.

(d) Further, if h's are affine functions, and f and g's are convex functions, then x^* is a local minimizer of (2).

Convergence theorem for (2)

Assume that $\epsilon_k \to 0$. Suppose that $\mathcal{X}_\Upsilon := \{x \in \mathcal{X} : f(x) \le \Upsilon\}$ is compact. Then, the following statements hold:

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

(b) Suppose (x^*, y^*) is a limit point of $\{(x^k, y^k)\}$. Then, x^* is a feasible point of problem (2).

(c) Let (x^*, y^*) be defined above. Suppose that $\{(x^k, y^k)\}_{k\in K} \to (x^*, y^*)$ for some subsequence K. Let $J^* = \{j \in J : x_j^* \neq 0\}$, $\bar{J}^* = J \setminus J^*$. Assume that the Robinson condition (7) holds at x^* for such $\bar{J}^*$. Then, $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ is bounded, where

$$\lambda^k = \varrho_k [g(x^k)]^+, \quad \mu^k = \varrho_k h(x^k), \quad \varpi^k = \varrho_k (x_J^k - y^k).$$

Moreover, each limit point $(\lambda^*, \mu^*, \varpi^*)$ of $\{(\lambda^k, \mu^k, \varpi^k)\}_{k\in K}$ together with x^* satisfies the KKT conditions (6) with $z_j^* = \varpi_i^*$ for all $j = J(i) \in \bar{J}^*$.

(d) Further, if h's are affine functions, and f and g's are convex functions, then x^* is a local minimizer of (2).

Sparse logistic regression (real data)

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

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

- compare the solution quality of l_0 and l_1 models

Table: Computational results on three real data sets

Data	p	n			SLEP			PD	
			$\lambda/\lambda_{\max}$	r	l_{avg}	Error (%)	Time	l_{avg}	Error (%)	Time
Colon	2000	62	0.5	7	0.4398	17.74	0.2	0.4126	12.9	9.1
			0.1	22	0.1326	1.61	0.5	0.0150	0	6.0
			0.05	25	0.0664	0	0.6	0.0108	0	5.0
			0.01	28	0.0134	0	1.3	0.0057	0	5.4
Iono	34	351	0.5	3	0.4804	17.38	0.1	0.3466	13.39	0.7
			0.1	11	0.3062	11.40	0.1	0.2490	9.12	1.0
			0.05	14	0.2505	9.12	0.1	0.2002	8.26	1.1
			0.01	24	0.1846	6.55	0.4	0.1710	5.98	1.7
Ad	1430	2359	0.5	3	0.2915	12.04	2.3	0.2578	7.21	31.9
			0.1	36	0.1399	4.11	14.2	0.1110	4.11	56.0
			0.05	67	0.1042	2.92	21.6	0.0681	2.92	74.1
			0.01	197	0.0475	1.10	153.0	0.0249	1.10	77.4

Sparse logistic regression (random data)

Table: Computational results on random data sets

Size $n \times p$	$\lambda/\lambda_{\max}$	r	SLEP l_{avg}	SLEP Error (%)	SLEP Time	PD l_{avg}	PD Error (%)	PD Time
1000×2000	0.9	17.0	0.6411	9.76	0.4	0.2145	8.49	9.9
	0.7	52.9	0.5090	3.96	1.0	0.0588	2.66	20.0
	0.5	96.6	0.3838	2.23	1.7	0.0060	0.02	34.9
	0.3	138.7	0.2611	1.22	2.1	0.0022	0	25.5
	0.1	192.0	0.1228	0.31	2.0	0.0013	0	16.0
2000×1000	0.9	11.0	0.6441	11.46	0.4	0.2763	10.67	15.2
	0.7	42.8	0.5083	3.63	1.1	0.0376	1.49	38.9
	0.5	78.0	0.3776	1.65	2.0	0.0032	0	34.4
	0.3	115.5	0.2490	0.6	2.6	0.0015	0	25.3
	0.1	160.8	0.1056	0.03	3.1	0.0010	0	15.8
1000×1000	0.9	11.7	0.6417	11.00	0.1	0.2444	9.67	2.3
	0.7	37.2	0.5086	3.95	0.2	0.0572	2.46	5.8
	0.5	67.6	0.3805	2.15	0.3	0.0060	0.01	6.2
	0.3	100.1	0.2544	0.81	0.4	0.0016	0	4.6
	0.1	137.9	0.1124	0.12	0.5	0.0011	0	3.3

Sparse inverse covariance selection (real data)

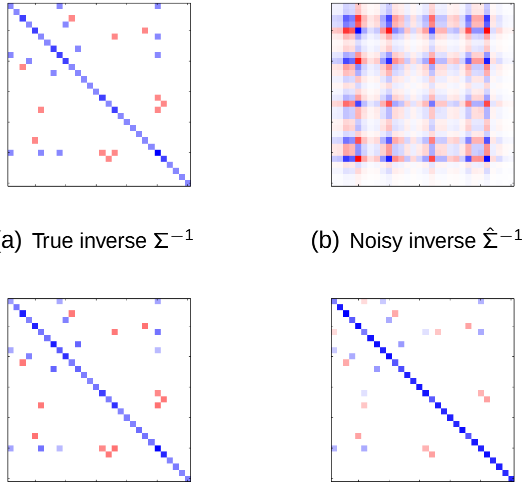
$$\max_{X \succeq 0} \left\{ \log\det X - \left\langle \hat{\Sigma}, X \right\rangle : X_{ij} = 0 \quad \forall (i,j) \in \Omega, \sum_{(i,j)\in\bar{\Omega}} \|X_{ij}\|_0 \le r \right\},$$

$$\max_{X \succeq 0} \left\{ \log\det X - \left\langle \hat{\Sigma}, X \right\rangle - \rho \sum_{(i,j)\in\bar{\Omega}} |X_{ij}| : X_{ij} = 0 \quad \forall (i,j) \in \Omega \right\}.$$

Table: Computational results on two real data sets

Genes	Samples			PPA			PD		
p	n	ρ	r	Likelihood	Loss	Time	Likelihood	Loss	Time
587	148	0.01	144294	790.12	23.24	101.5	1035.24	22.79	38.0
		0.05	67474	174.86	24.35	85.2	716.97	23.27	31.5
		0.10	38504	−47.03	24.73	66.7	389.65	23.85	26.1
		0.50	4440	−561.38	25.52	33.2	−260.32	24.91	24.8
		0.70	940	−642.05	25.63	26.9	−511.70	25.30	22.0
		0.90	146	−684.59	25.70	22.0	−598.05	25.51	14.9
1255	72	0.01	249216	3229.75	28.25	705.7	3555.38	28.12	177.1
		0.05	169144	1308.38	29.85	491.1	2996.95	28.45	189.2
		0.10	107180	505.02	30.53	501.4	2531.62	28.82	202.8
		0.50	37914	−931.59	31.65	345.9	797.23	30.16	256.6
		0.70	4764	−1367.22	31.84	125.7	−1012.48	31.48	271.6
		0.90	24	−1465.70	31.90	110.6	−1301.99	31.68	187.8

Sparse inverse covariance selection (synthetic data)



(a) True inverse Σ^{-1} (b) Noisy inverse $\hat{\Sigma}^{-1}$

(c) Approximate solution of PD (d) Approximate solution of PPA

Summary:

- Study optimality conditions for l_0 minimization
- Propose PD methods for l_0 minimization
- Converge to a KKT point for general l_0 minimization
- Converge to a local minimizer for “convex” l_0 minimization

Reference:

Sparse Approximation via Penalty Decomposition Methods
(with Yong Zhang).

Thank you!

Summary:

- Study optimality conditions for l_0 minimization
- Propose PD methods for l_0 minimization
- Converge to a KKT point for general l_0 minimization
- Converge to a local minimizer for "convex" l_0 minimization

Reference:

Sparse Approximation via Penalty Decomposition Methods (with Yong Zhang).

Thank you!

Summary:

- Study optimality conditions for l_0 minimization
- Propose PD methods for l_0 minimization
- Converge to a KKT point for general l_0 minimization
- Converge to a local minimizer for “convex” l_0 minimization

Reference:

Sparse Approximation via Penalty Decomposition Methods (with Yong Zhang).

Thank you!

Summary:

- Study optimality conditions for l_0 minimization
- Propose PD methods for l_0 minimization
- Converge to a KKT point for general l_0 minimization
- Converge to a local minimizer for “convex” l_0 minimization

Reference:

Sparse Approximation via Penalty Decomposition Methods
(with Yong Zhang).

Thank you!

Summary:

- Study optimality conditions for l_0 minimization
- Propose PD methods for l_0 minimization
- Converge to a KKT point for general l_0 minimization
- Converge to a local minimizer for "convex" l_0 minimization

Reference:
Sparse Approximation via Penalty Decomposition Methods
(with Yong Zhang).

Thank you!

Summary:

- Study optimality conditions for l_0 minimization
- Propose PD methods for l_0 minimization
- Converge to a KKT point for general l_0 minimization
- Converge to a local minimizer for “convex” l_0 minimization

Reference:

Sparse Approximation via Penalty Decomposition Methods
(with Yong Zhang).

Thank you!