Coordinate Descent Methods for DC Minimization: Optimality Conditions and Global Convergence

Ganzhao Yuan

Peng Cheng Laboratory, China

Outline of This Talk

1. Introduction
2. Coordinate Descent Methods
3. Theoretical Analysis
4. A Breakpoint Searching Method
5. Experimental Results
6. Discussions and Extensions

Introduction

Introduction

The DC minimization problem:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \ \mathcal{F}(\mathbf{x}) \triangleq f(\mathbf{x}) + h(\mathbf{x}) - g(\mathbf{x}) \qquad (1)$$

Assumptions

1. $f(\cdot)$ is convex and continuously differentiable:

$$\forall \mathbf{x}, \eta, \ \ f(\mathbf{x}+\eta e_i) \leq f(\mathbf{x}) + \langle \nabla_i f(\mathbf{x}), \ \eta e_i\rangle + \frac{\mathbf{c}_i}{2}\|\eta e_i\|_2^2$$

   $e_i \in \mathbb{R}^n$ is an indicator vector with one on the i-th entry and zero everywhere else.

2. $h(\cdot) = \sum_{i=1}^n h_i(\mathbf{x}_i)$ is convex and coordinate-wise separable.
3. $g(\cdot)$ is convex and its associated proximal operator can be computed exactly:

$$\min_{\eta\in\mathbb{R}} p(\eta) \triangleq \frac{a}{2}\eta^2 + b\eta + h_i(\mathbf{x}+\eta e_i) - g(\mathbf{x}+\eta e_i),$$

Examples

1. ℓ_p Norm Generalized Eigenvalue Problem

$$
\begin{aligned}
& \max_{\mathbf{x}} \ \|\mathbf{G}\mathbf{x}\|_p, \ \ s.t. \ \mathbf{x}^T\mathbf{Q}\mathbf{x} = 1 \\
\Leftrightarrow \quad & \bar{\mathbf{x}} = \arg\min_{\mathbf{x}} \ \tfrac{\alpha}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \|\mathbf{G}\mathbf{x}\|_p,
\end{aligned}
$$

2. Generalized Linear Regression

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \tfrac{1}{2}\|\sigma(\mathbf{G}\mathbf{x}) - \mathbf{y}\|_F^2
$$

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \ \tfrac{1}{2}\|\sigma(\mathbf{G}\mathbf{x})\|_2^2 - \langle \mathbf{1}, \sigma(\mathrm{diag}(\mathbf{y})\mathbf{G}\mathbf{x})\rangle + \tfrac{1}{2}\|\mathbf{y}\|_2^2
$$

RELU Neural Network: $\sigma(t) = \max(0, t)$

Phase Retrieval: $\sigma(t) = |t|$.

Examples

1. Approximate Sparse Optimization

$$\min_{\mathbf{x}} \ \tfrac{1}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2, \ \ s.t. \ \|\mathbf{x}\|_0 \le s$$

Using the fact that: $\|\mathbf{x}\|_0 \le s \Leftrightarrow \|\mathbf{x}\|_1 = \sum_{i=1}^{s} |\mathbf{x}_{[i]}|$, we have the following equivalent DC problem:

$$\min_{\mathbf{x}} \ \tfrac{1}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2 + \rho(\|\mathbf{x}\|_1 - \sum_{i=1}^{s} |\mathbf{x}_{[i]}|)$$

2. Approximate Binary Optimization

$$\min_{\mathbf{x}\in\{-1,+1\}^n} \ \tfrac{1}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2$$

Using the fact that: $\mathbf{x} \in \{-1,+1\}^n \Leftrightarrow -1 \le \mathbf{x} \le 1, \|\mathbf{x}\|_2^2 = n$, we have the following equivalent DC problem:

$$\min_{\|\mathbf{x}\|_\infty \le 1} \ \tfrac{1}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2 + \rho(\sqrt{n} - \|\mathbf{x}\|)$$

Related Topics of This Paper

1. DC programming
2. Coordinate Descent Methods
3. Iterative Majorization Minimization
4. Provable Nonconvex Algorithms

DC programming

1. An extension of convex maximization over a convex set, closely related to CCCP and alternating minimization
2. The class of DC functions is very broad, considered in global optimization
3. Recent developments focus on local solution methods (proximal bundle DC methods, double bundle DC methods, inertial proximal methods, enhanced proximal methods)
4. Many applications (sparse PCA, variable selection, single source localization, piecewise linear programming)

Coordinate descent methods

1. A popular method for solving large-scale problems
2. Enjoys faster convergence, avoids tricky parameters tuning, allows for easy parallelization
3. Well studied for convex optimization (Lasso, SVM, NMF, PageRank)
4. Extended to solve nonconvex problems (penalized regression, eigenvalue complementarity problem, ℓ_0 norm minimization, resource allocation problem, leading eigenvector computation, sparse phase retrieval)

Iterative Majorization Minimization

1. Lipschitz gradient surrogate
2. proximal gradient surrogate
3. DC programming surrogate
4. variational surrogate
5. saddle point surrogate
6. Jensen surrogate
7. quadratic surrogate
8. Frank-Wolfe surrogate
9. cubic surrogate

Provable Nonconvex Algorithms

1. find stronger stationary points
   - second-order stationary point $\in$ first-order stationary point
   - block-k stationary point $\in$ coordinate-wise stationary point $\in$ Lipschitz stationary point
2. Convergence analysis
   - Kurdyka-Łojasiewicz inequality
   - weakly convex, a regularity condition, a sharpness condition
   - *Luo-Tseng* error bound assumption

Related Work

1. Multi-Stage Convex Relaxation

$$\mathbf{x}^{t+1} = \arg\min_{\mathbf{x}} \; f(\mathbf{x}) + h(\mathbf{x}) - \langle \mathbf{x} - \mathbf{x}^t, \; \mathbf{g}^t \rangle, \; \mathbf{g}^t \in \partial g(\mathbf{x}^t)$$

2. Proximal DC algorithm (PDCA)

$$\mathbf{x}^{t+1} = \arg\min_{\mathbf{x}} \; \mathcal{Q}(\mathbf{x}, \mathbf{x}^t) + h(\mathbf{x}) - \langle \mathbf{x} - \mathbf{x}^t, \; \mathbf{g}^t \rangle$$

$$\mathcal{Q}(\mathbf{x}, \mathbf{x}^t) \triangleq f(\mathbf{x}^t) + \langle \nabla f(\mathbf{x}^t), \; \mathbf{x} - \mathbf{x}^t \rangle + \frac{L}{2} \|\mathbf{x} - \mathbf{x}^t\|_2^2$$

3. Toland's duality method

$$\min_{\mathbf{y}} \bar{g}^*(\mathbf{y}) - f^*(\mathbf{A}^T \mathbf{y}) - h^*(\mathbf{A}^T \mathbf{y})$$

4. Subgradient descent method

$$\mathbf{x}^{t+1} = \mathcal{P}(\mathbf{x}^t - \eta^t \mathbf{g}^t)$$

Contributions

1. A new coordinate descent method based on sequential nonconvex approximation
2. Coordinate-wise optimality condition is always stronger than the critical/directional point condition
3. Linear convergence rate
4. A breakpoint searching method for computing the proximal operator
5. Extensive experiments on some statistical learning tasks
6. Several important discussions and extensions

Coordinate Descent Methods

Coordinate Descent Methods

The Coordinate Descent Methods:

$$\bar{\eta}^t = \arg\min_{\eta \in \mathbb{R}} \; f(\mathbf{x}^t + \eta e_{i^t}) + h(\mathbf{x}^t + \eta e_{i^t}) - g(\mathbf{x}^t + \eta e_{i^t})$$

$$\mathbf{x}^{t+1} = \mathbf{x}^t + \bar{\eta}^t e_{i^t}$$

Choosing the Majorization Function

$$f(\mathbf{x}^t + \eta e_{i^t}) \leq \mathcal{S}_{i^t}(\mathbf{x}^t, \eta) \triangleq f(\mathbf{x}^t) + \langle \nabla f(\mathbf{x}^t), \; \eta e_{i^t} \rangle + \frac{\mathbf{c}_{i^t}}{2}\eta^2,$$

$$-g(\mathbf{x}^t + \eta e_{i^t}) \leq \mathcal{R}_{i^t}(\mathbf{x}^t, \eta) \triangleq -g(\mathbf{x}^t) - \langle \partial g(\mathbf{x}^t), \; (\mathbf{x}^t + \eta e_{i^t}) - \mathbf{x}^t \rangle.$$

The two CD methods:

$$NonConvex: \quad \bar{\eta}^t = \arg\min_{\eta} \; \mathcal{S}_{i^t}(\mathbf{x}^t, \eta) + h_{i^t}(\mathbf{x}^t + \eta e_{i^t}) - g(\mathbf{x}^t + \eta e_{i^t})$$

$$Convex: \quad \bar{\eta}^t = \arg\min_{\eta} \; \mathcal{S}_{i^t}(\mathbf{x}^t, \eta) + h_{i^t}(\mathbf{x}^t + \eta e_i) + \mathcal{R}_{i^t}(\mathbf{x}^t, \eta)$$

Coordinate Descent Methods

Input: an initial feasible solution $\mathbf{x}^0$, $\theta > 0$. Set $t = 0$.

while *not converge* **do**

S1 Use some strategy to find a coordinate $i^t \in \{1, ..., n\}$ for the t-th iteration.

S2 Solve the following nonconvex or convex subproblem.

• Option I: SNCA strategy.

$$\bar{\eta}^t = \arg\min_{\eta} \ \mathcal{S}_{i^t}(\mathbf{x}^t, \eta) + h_{i^t}(\mathbf{x}^t + \eta e_{i^t}) - g(\mathbf{x}^t + \eta e_{i^t}) + \frac{\theta}{2}\|\eta e_{i^t}\|_2^2$$

• Option II: SCA strategy.

$$\bar{\eta}^t = \arg\min_{\eta} \ \mathcal{S}_{i^t}(\mathbf{x}^t, \eta) + h_{i^t}(\mathbf{x}^t + \eta e_i) + \mathcal{R}_{i^t}(\mathbf{x}^t, \eta) + \frac{\theta}{2}\|\eta e_{i^t}\|_2^2$$

S3 $\mathbf{x}^{t+1} = \mathbf{x}^t + \bar{\eta}^t \cdot e_{i^t} \quad (\Leftrightarrow \mathbf{x}_{i^t}^{t+1} = \mathbf{x}_{i^t}^t + \bar{\eta}^t)$

S4 Increment t by 1

end

Algorithm 1: Coordinate Descent Methods for Minimizing DC functions using **SNCA** or **SCA** strategy.

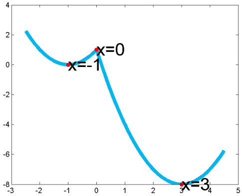
Remarks

1. A proximal term is used $\Rightarrow$ sufficient descent condition
2. The subproblem is equivalent to solving the following nonconvex problem which has a bilinear structure:

$$\min_{\eta, \mathbf{y}} \; \mathcal{S}_{i^t}(\mathbf{x}^t, \eta) + \frac{\theta}{2}\eta^2 + h(\mathbf{x}^t + \eta e_{i^t}) - \langle \mathbf{y}, \mathbf{x}^t + \eta e_{i^t} \rangle + g^*(\mathbf{y})$$

3. One can apply CD to the primal/dual
4. CD fails for *nonseparable nonsmooth convex* functions.
   Example: $\min_{x,y} x^2 + y^2 + 2|x - y|$.
   $(x^0, y^0) = (1, 1)$. It gets stuck at $(x^\infty, y^\infty) = (1, 1)$.
5. CD converges for *nonseparable nonsmooth concave* functions.
   Example: $\min_{x,y} x^2 + y^2 - 2|x - y|$.
   $(x^0, y^0) = (1, 1)$. It stops at $(x^\infty, y^\infty) = (-1, 1)$ or $(1, -1)$.

Remarks



1. **CD-SNCA** is more accurate than **CD-SCA**.

   Example: $\min_x (x-1)^2 - 4|x|$. Three critical points $\{-1, 0, 3\}$

   CD-SCA only finds one of the critical points

   CD-SNCA finds the global optimal solution $x = 3$

   This is achieved by using a breakpoint searching algorithm

Theoretical Analysis

Theoretical Analysis

Assumption (globally ρ-bounded nonconvexity)

$\ddot{g}(\mathbf{x}) \triangleq -g(\mathbf{x})$ *is ρ-bounded nonconvex that:*

$$\ddot{g}(\mathbf{x}) \leq \ddot{g}(\mathbf{y}) + \langle \mathbf{x} - \mathbf{y},\ \partial \ddot{g}(\mathbf{x}) \rangle + \frac{\rho}{2}\|\mathbf{x} - \mathbf{y}\|_2^2,\ \forall \mathbf{x}, \mathbf{y}.$$

Optimality Definition

Definition (Critical Point)

A solution $\check{\mathbf{x}}$ is called a critical point if the following holds:

$$0 \in \nabla f(\check{\mathbf{x}}) + \partial h(\check{\mathbf{x}}) - \partial g(\check{\mathbf{x}})$$

Definition (Directional Point)

A solution $\grave{\mathbf{x}}$ is called a directional point if the following holds:

$$\mathcal{F}'(\grave{\mathbf{x}}; \mathbf{y} - \grave{\mathbf{x}}) \triangleq \lim_{t \downarrow 0} \frac{\mathcal{F}(\grave{\mathbf{x}} + t(\mathbf{y} - \grave{\mathbf{x}})) - \mathcal{F}(\grave{\mathbf{x}})}{t} \geq 0, \ \forall \mathbf{y}$$

with $\mathbf{y} \in \mathrm{dom}(\mathcal{F}) \triangleq \{\mathbf{x} : |\mathcal{F}(\mathbf{x})| < +\infty\}$.

Optimality Definition

Definition (Coordinate-Wise Stationary Point)

We let

$$\mathcal{M}_i(\mathbf{x}, \eta) \triangleq \frac{\mathbf{c}_i + \theta}{2}\eta^2 + \nabla f(\mathbf{x})_i \eta + h(\mathbf{x} + \eta e_i) - g(\mathbf{x} + \eta e_i)$$

for a given constant $\theta \geq 0$. A solution $\ddot{\mathbf{x}}$ is called a coordinate-wise stationary point if the following holds:

$$0 \in \arg\min_{\eta} \; \mathcal{M}_i(\ddot{\mathbf{x}}, \eta), \forall i = 1, ..., n.$$

Optimality Hierarchy

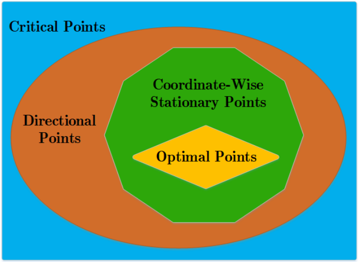
Theorem (Optimality Hierarchy between the Optimality Conditions.)

Assume that $-g(\cdot)$ is globally ρ-bounded nonconvex.

*(**a**) It holds that* $\forall \mathbf{d},\ \mathcal{F}(\ddot{\mathbf{x}}) \leq \mathcal{F}(\ddot{\mathbf{x}} + \mathbf{d}) + \frac{1}{2}\|\mathbf{d}\|^2_{(\mathbf{c}+\theta+\rho)}$.

*(**b**) The following relation holds:*

$$\{\bar{\mathbf{x}}\} \overset{\textbf{(b-i)}}{\subseteq} \{\ddot{\mathbf{x}}\} \overset{\textbf{(b-ii)}}{\subseteq} \{\grave{\mathbf{x}}\} \overset{\textbf{(b-iii)}}{\subseteq} \{\check{\mathbf{x}}\}$$



Global Convergence

Theorem (Global Convergence)

(a) *For* **CD-SNCA**, *we have:* $\mathcal{F}(\mathbf{x}^{t+1}) - \mathcal{F}(\mathbf{x}^t) \leq -\frac{\theta}{2}\|\mathbf{x}^{t+1} - \mathbf{x}^t\|^2$. *Algorithm 1 finds an* ϵ*-approximate* **coordinate-wise stationary point** *of Problem (1) in at most* T *iterations in the sense of expectation, where* $T \leq \lceil \frac{2n(\mathcal{F}(\mathbf{x}^0) - \mathcal{F}(\bar{\mathbf{x}}))}{\theta\epsilon} \rceil = O(\epsilon^{-1})$.

(b) *For* **CD-SCA**, *we have:* $\mathcal{F}(\mathbf{x}^{t+1}) - \mathcal{F}(\mathbf{x}^t) \leq -\frac{\beta}{2}\|\mathbf{x}^{t+1} - \mathbf{x}^t\|^2$ *with* $\beta \triangleq \min(\mathbf{c}) + 2\theta$. *Algorithm 1 finds an* ϵ*-approximate* **critical point** *of Problem (1) in at most* T *iterations in the sense of expectation, where* $T \leq \lceil \frac{2n(\mathcal{F}(\mathbf{x}^0) - \mathcal{F}(\bar{\mathbf{x}}))}{\beta\epsilon} \rceil = O(\epsilon^{-1})$.

Convergence Rate

Assumption

(Luo-Tseng Error Bound) We define a residual function as $\mathcal{R}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} |dist(0, \bar{\mathcal{M}}_i(\mathbf{x}))|$ *or* $\mathcal{R}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} |dist(0, \bar{\mathcal{P}}_i(\mathbf{x}))|$*, where* $\bar{\mathcal{M}}_i(\mathbf{x})$ *and* $\bar{\mathcal{P}}_i(\mathbf{x})$ *are respectively defined in* **CD-SNCA** *and* **CD-SCA***. For any* $\varsigma \geq \min_{\mathbf{x}} F(\mathbf{x})$*, there exist scalars* $\delta > 0$ *and* $\varrho > 0$ *such that:*

$$\forall \mathbf{x},\ \ dist(\mathbf{x}, \mathcal{X}) \leq \delta \mathcal{R}(\mathbf{x}),\ \ whenever\ F(\mathbf{x}) \leq \varsigma, \mathcal{R}(\mathbf{x}) \leq \varrho.$$

Here, $\mathcal{X}$ *is the set of stationary points satisfying* $\mathcal{R}(\mathbf{x}) = 0$.

Convergence Rate

$$\ddot{q}^t \triangleq F(\mathbf{x}^t) - F(\ddot{\mathbf{x}}), \ddot{r}^t \triangleq \frac{1}{2}\|\mathbf{x}^t - \ddot{\mathbf{x}}\|_{\bar{\mathbf{c}}}^2$$

$$\bar{\mathbf{c}} \triangleq \mathbf{c} + \theta, \bar{\rho} = \frac{\rho}{\min(\bar{\mathbf{c}})},\ \gamma \triangleq 1 + \frac{\rho}{\theta}, \varpi \triangleq 1 - \bar{\rho}$$

Theorem (Convergence Rate for CD-SNCA)

Assume that $z(\mathbf{x}) \triangleq -g(\mathbf{x})$ *is globally* ρ*-bounded non-convex.*

(a) *We have* $\varpi \mathbb{E}_{i^t}[\ddot{r}^{t+1}] + \gamma \mathbb{E}_{i^t}[\ddot{q}^{t+1}] \leq \varpi \ddot{r}^t + \gamma \ddot{q}^t + \frac{\bar{\rho}}{n}\ddot{r}^t - \frac{\ddot{q}^t}{n}$.

(b) *If* θ *is sufficiently large such that* $\varpi \geq 0$, $\mathcal{M}_{i^t}(\mathbf{x}^t, \eta)$ *in (2) is convex w.r.t.* η *for all* t.

(c) $\ddot{q}^{t+1} \leq (\frac{\kappa_1 - \frac{1}{n}}{\kappa_1})^{t+1} \ddot{q}^0$, *where* $\kappa_0 \triangleq \max(\bar{\mathbf{c}})\frac{\delta^2}{\theta}$ *and* $\kappa_1 \triangleq n\kappa_0(\varpi + \frac{\bar{\rho}}{n}) + \gamma$.

Convergence Rate

$$\check{q}^t \triangleq F(\mathbf{x}^t) - F(\check{\mathbf{x}}), \check{r}^t \triangleq \frac{1}{2}\|\mathbf{x}^t - \check{\mathbf{x}}\|_{\bar{\mathbf{c}}}^2, \ \bar{\mathbf{c}} \triangleq \mathbf{c} + \theta, \bar{\rho} = \frac{\rho}{\min(\bar{\mathbf{c}})}.$$

Theorem (Convergence Rate for CD-SCA)

Assume that $z(\mathbf{x}) \triangleq -g(\mathbf{x})$ is globally ρ-bounded non-convex.

(a) *We have* $\mathbb{E}_{i^t}[\check{r}^{t+1}] + \mathbb{E}[\check{q}^{t+1}] \leq \check{r}^t + \frac{\bar{\rho}}{n}\check{r}^t - \frac{1}{n}\check{q}^t + \check{q}^t$.

(b) *It holds that:* $\check{q}^{t+1} \leq (\frac{\kappa_2 - \frac{1}{n}}{\kappa_2})^{t+1}\check{q}^0$, *where* $\kappa_0 \triangleq \max(\bar{\mathbf{c}})\frac{\delta^2}{\theta}$ *and* $\kappa_2 = n\kappa_0(1 + \frac{\bar{\rho}}{n}) + 1$.

Conclusions:

- Q-linearly convergence rate for **CD-SNCA** and **CD-SCA**
- When n is large and we choose $0 \leq \varpi < 1$, **CD-SNCA** is much faster than **CD-SCA**.

A Breakpoint Searching Method

A Breakpoint Searching Method

Two steps:

1. identifies all the possible critical points / breakpoints Θ for $\min_{\eta \in \mathbb{R}} p(\eta)$
2. picks the solution that leads to the lowest value as the optimal solution.

Examples:

1. $g(\mathbf{y}) \triangleq \|\mathbf{Ay}\|_\infty$ and $h_i(\cdot) \triangleq 0$
2. $g(\mathbf{y}) \triangleq \|\mathbf{Ay}\|_2$ and $h_i(\cdot) \triangleq 0$
3. $g(\mathbf{y}) \triangleq \sum_{i=1}^{s} |\mathbf{y}_{[i]}|$ and $h_i(\mathbf{y}) \triangleq |\mathbf{y}_i|$
4. $g(\mathbf{y}) \triangleq \|\mathbf{Ay}\|_1$ and $h_i(\cdot) \triangleq 0$
5. $g(\mathbf{y}) \triangleq \|\max(0, \mathbf{Ay})\|_1$ and $h_i(\cdot) \triangleq 0$

Example 1: $g(\mathbf{y}) \triangleq \|\mathbf{Ay}\|_\infty$ and $h_i(\cdot) \triangleq 0$

Consider the problem:

$$\begin{aligned} &\min_\eta \tfrac{a}{2}\eta^2 + b\eta - \|\mathbf{A}(\mathbf{x} + \eta e_i)\|_\infty \\ \Leftrightarrow\quad &\min_\eta \frac{a}{2}\eta^2 + b\eta - \|\mathbf{g}\eta + \mathbf{d}\|_\infty \\ \Leftrightarrow\quad &\min_\eta p(\eta) \triangleq \frac{a}{2}\eta^2 + b\eta + \max_{i=1}^{2m}(\bar{\mathbf{g}}_i\eta + \bar{\mathbf{d}}_i) \end{aligned}$$

with $\bar{\mathbf{g}} = [\mathbf{g}_1, \mathbf{g}_2, ..., \mathbf{g}_m, -\mathbf{g}_1, -\mathbf{g}_2, ..., -\mathbf{g}_m]$ and $\bar{\mathbf{d}} = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_m, -\mathbf{d}_1, -\mathbf{d}_2, ..., -\mathbf{d}_m]$.

Letting $0 \in \partial p(\cdot)$, we have: $a\eta + b + \bar{\mathbf{g}}_i = 0$ with $i = 1, 2, ..., (2m)$. We have $\boldsymbol{\eta} = (-b - \bar{\mathbf{g}})/a$.

This problem contains $2m$ breakpoints $\Theta = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, ..., \boldsymbol{\eta}_{2m}\}$.

Example 2: $g(\mathbf{y}) \triangleq \|\mathbf{A}\mathbf{y}\|_2$ and $h_i(\cdot) \triangleq 0$

Consider the problem:

$$\min_{\eta} \tfrac{a}{2}\eta^2 + b\eta - \|\mathbf{A}(\mathbf{x} + \eta e_i)\|_p \Leftrightarrow \min_{\eta} p(\eta) \triangleq \frac{a}{2}\eta^2 + b\eta - \|\mathbf{g}\eta + \mathbf{d}\|_p$$

We have

$$0 \in \partial p(\eta) = a\eta + b + \|\mathbf{g}\eta - \mathbf{d}\|_p^{1-p} \langle \mathbf{g}, \operatorname{sign}(\mathbf{g}\eta + \mathbf{d}) \odot |\mathbf{g}\eta + \mathbf{d}|^{p-1} \rangle.$$

We only focus on $p = 2$. We obtain:

$$\begin{aligned} 0 = -a\eta - b = \frac{\langle \mathbf{g}, \mathbf{g}\eta + \mathbf{d} \rangle}{\|\mathbf{g}\eta - \mathbf{d}\|} \quad &\Leftrightarrow \quad \|\mathbf{g}\eta - \mathbf{d}\|(-a\eta - b) = \langle \mathbf{g}, \mathbf{g}\eta + \mathbf{d} \rangle \\ &\Leftrightarrow \quad \|\mathbf{g}\eta - \mathbf{d}\|_2^2 (a\eta + b)^2 = (\langle \mathbf{g}, \mathbf{g}\eta + \mathbf{d} \rangle)^2 \end{aligned}$$

Solving this quartic equation we obtain all of its real roots $\{\eta_1, \eta_2, ..., \eta_c\}$ with $1 \le c \le 4$.

This problem at most contains 4 breakpoints $\Theta = \{\eta_1, \eta_2, ..., \eta_c\}$.

Example 3: $g(\mathbf{y}) \triangleq \sum_{i=1}^{s} |\mathbf{y}_{[i]}|$ and $h_i(\mathbf{y}) \triangleq |\mathbf{y}_i|$

Consider the problem:

$$\min_{\eta} \ \frac{a}{2}\eta^2 + b\eta + |\mathbf{x}_i + \eta| - \sum_{i=1}^{s} |(\mathbf{x} + \eta e_i)_{[i]}|$$

Since the variable η only affects the value of $\mathbf{x}_i$, we consider two cases for $\mathbf{x}_i + \eta$.

(i) $\mathbf{x}_i + \eta$ belongs to the top-s subset. It reduces to $\min_\eta \ \frac{a}{2}\eta^2 + b\eta$. It has 1 breakpoint: $\{-\frac{b}{a}\}$.

(ii) $\mathbf{x}_i + \eta$ does not belong to the top-s subset. It reduces to $\min_\eta \ \frac{a}{2}\eta^2 + bt + |\mathbf{x}_i + \eta|$. It has 3 breakpoints $\{-\mathbf{x}_i, \frac{-1-b}{a}, \frac{1-b}{a}\}$.

This problem contains 4 breakpoints $\Theta = \{-\frac{b}{a}, -\mathbf{x}_i, \frac{-1-b}{a}, \frac{1-b}{a}\}$.

Example 4: $g(\mathbf{y}) \triangleq \|\mathbf{A}\mathbf{y}\|_1$ and $h_i(\cdot) \triangleq 0$

Consider the problem:

$$\min_{\eta} \tfrac{a}{2}\eta^2 + b\eta - \|\mathbf{A}(\mathbf{x} + \eta e_i)\|_1 \Leftrightarrow \min_{\eta} p(\eta) \triangleq \frac{a}{2}\eta^2 + b\eta - \|\mathbf{g}\eta + \mathbf{d}\|_1$$

Letting $0 \in \partial p(\eta)$, we have:

$0 \in a\eta + b - \langle \text{sign}(\eta\mathbf{g} + \mathbf{d}), \mathbf{g}\rangle = a\eta + b - \langle \text{sign}(\eta + \mathbf{d} \div |\mathbf{g}|), |\mathbf{g}|\rangle$.

We define $\mathbf{z} \triangleq \{+\frac{\mathbf{d}_1}{\mathbf{g}_1}, -\frac{\mathbf{d}_1}{\mathbf{g}_1}, ..., +\frac{\mathbf{d}_m}{\mathbf{g}_m}, -\frac{\mathbf{d}_m}{\mathbf{g}_m}\} \in \mathbb{R}^{2m\times 1}$, and $\mathbf{z}_1 \leq \mathbf{z}_2 \leq ... \leq \mathbf{z}_{2m}$. The domain $p(\eta)$ can be divided into $2m+1$ intervals: $(-\infty, \mathbf{z}_1)$, $(\mathbf{z}_1, \mathbf{z}_2)$,..., and $(\mathbf{z}_{2m}, +\infty)$. There are $2m+1$ breakpoints $\boldsymbol{\eta} \in \mathbb{R}^{(2m+1)\times 1}$. In each interval, the sign of $(\eta + \mathbf{d} \div |\mathbf{g}|)$ can be determined. Thus, the i-th breakpoints for the i-th interval is: $\boldsymbol{\eta}_i = (\langle \text{sign}(\eta + \mathbf{d} \div |\mathbf{g}|), \mathbf{g}\rangle - b)/a$. It contains $2m+1$ breakpoints $\Theta = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, ..., \boldsymbol{\eta}_{(2m+1)}\}$.

Example 5: $g(\mathbf{y}) \triangleq \|\max(0, \mathbf{A}\mathbf{y})\|_1$ and $h_i(\cdot) \triangleq 0$

Consider the problem:

$$\min_{\eta} \tfrac{a}{2}\eta^2 + b\eta - \|\max(0, \mathbf{A}(\mathbf{x} + \eta e_i))\|_1$$

Using the fact that $\max(0, a) = \frac{1}{2}(a + |a|)$, we have the following equivalent problem:

$$\min_{\eta} \tfrac{a}{2}\eta^2 + b\eta - \frac{1}{2}\langle \mathbf{1}, \mathbf{A}e_i \rangle \eta - \frac{1}{2}\|\mathbf{A}(\mathbf{x} + \eta e_i)\|_1$$

Therefore, the proximal operator of $g(\mathbf{x}) = \|\max(0, \mathbf{A}\mathbf{x})\|_1$ can be transformed to the proximal operator of $g(\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|_1$.

A Breakpoint Searching Method

When the breakpoint set Θ is found, we pick the solution that leads to the lowest value as the global optimal solution $\bar{\eta}$:

$$\bar{\eta} = \arg\min_{\eta} p(\eta), \;\; s.t. \; \eta \in \Theta.$$

The function $h_i(\cdot)$ does not bring much difficulty for solving the subproblem.

Experimental Results

Experimental Results

We consider the following four types of data sets for the sensing/channel matrix $\mathbf{G} \in \mathbb{R}^{m \times n}$

1. 'randn-m-n': $\mathbf{G} = \text{randn}(m, n)$.
2. 'e2006-m-n': $\mathbf{G} = \mathcal{X}$.
3. 'randn-m-n-C': $\mathbf{G} = \mathcal{N}(\text{randn}(m, n))$.
4. 'e2006-m-n-C': $\mathbf{G} = \mathcal{N}(\mathcal{X})$.

$\text{randn}(m, n)$ is a Gaussian random matrix of size $m \times n$. $\mathcal{X}$ is sampled from the data set 'e2006'. $\mathcal{N}(\mathbf{G})$ is defined as: $[\mathcal{N}(\mathbf{G})]_I = 100 \cdot \mathbf{G}_I, [\mathcal{N}(\mathbf{G})]_{\bar{I}} = \mathbf{G}_{\bar{I}}$, where I is a random subset of $\{1, ..., mn\}$, $\bar{I} = \{1, ..., mn\} \setminus I$, and $|I| = 0.1 \cdot mn$.

ℓ_p Norm Generalized Eigenvalue Problem

We consider the following problem:

$$\min_{\mathbf{x}} \ \frac{\alpha}{2}\|\mathbf{x}\|_2^2 - \|\mathbf{G}\mathbf{x}\|_1$$

Compared methods

1. Multi-Stage Convex Relaxation (MSCR)
2. Toland's dual method (T-DUAL)
3. Subgradient method (SubGrad)
4. **CD-SCA**: $\mathbf{x}_{i^t}^{t+1} = \mathbf{x}_{i^t}^t + \arg\min_\eta \frac{\mathbf{c}_i+\theta}{2}\eta^2 + (\nabla_{i^t} f(\mathbf{x}^t) - \mathbf{g}_{i^t}^t)\eta$
5. **CD-SNCA**:
   $\mathbf{x}_{i^t}^{t+1} = \mathbf{x}_{i^t}^t + \arg\min_\eta \frac{\mathbf{c}_i+\theta}{2}\eta^2 + \nabla_{i^t} f(\mathbf{x}^t)\eta - \|\mathbf{G}(\mathbf{x} + \eta e_i)\|_1$

Experimental Results

	MSCR	PDCA	T-DUAL	CD-SCA	CD-SNCA
randn-256-1024	-1.329 ± 0.038	-1.329 ± 0.038	-1.329 ± 0.038	-1.426 ± 0.056	**-1.447 ± 0.053**
randn-256-2048	-1.132 ± 0.021	-1.132 ± 0.021	-1.132 ± 0.021	-1.192 ± 0.019	**-1.202 ± 0.016**
randn-1024-256	-5.751 ± 0.163	-5.751 ± 0.163	-5.664 ± 0.173	-5.755 ± 0.108	**-5.817 ± 0.129**
randn-2048-256	-9.364 ± 0.183	-9.364 ± 0.183	-9.161 ± 0.101	-9.405 ± 0.182	**-9.408 ± 0.164**
e2006-256-1024	-28.031 ± 37.894	-28.031 ± 37.894	-27.996 ± 37.912	-27.880 ± 37.980	**-28.167 ± 37.826**
e2006-256-2048	-22.282 ± 24.007	-22.282 ± 24.007	-22.282 ± 24.007	-22.113 ± 23.941	**-22.448 ± 23.908**
e2006-1024-256	-43.516 ± 77.232	-43.516 ± 77.232	-43.364 ± 77.265	-43.283 ± 77.297	**-44.269 ± 76.977**
e2006-2048-256	-44.705 ± 47.806	-44.705 ± 47.806	-44.705 ± 47.806	-44.633 ± 47.789	**-45.176 ± 47.493**
randn-256-1024-C	-1.332 ± 0.019	-1.332 ± 0.019	-1.332 ± 0.019	-1.417 ± 0.027	**-1.444 ± 0.029**
randn-256-2048-C	-1.161 ± 0.024	-1.161 ± 0.024	-1.161 ± 0.024	-1.212 ± 0.022	**-1.219 ± 0.023**
randn-1024-256-C	-5.650 ± 0.141	-5.650 ± 0.141	-5.591 ± 0.145	-5.716 ± 0.159	**-5.808 ± 0.134**
randn-2048-256-C	-9.236 ± 0.125	-9.236 ± 0.125	-9.067 ± 0.137	-9.243 ± 0.145	**-9.377 ± 0.233**
e2006-256-1024-C	-4.841 ± 6.410	-4.841 ± 6.410	-4.840 ± 6.410	-4.837 ± 6.411	**-5.027 ± 6.363**
e2006-256-2048-C	-4.297 ± 2.825	-4.297 ± 2.825	-4.297 ± 2.823	-4.259 ± 2.827	**-4.394 ± 2.814**
e2006-1024-256-C	-6.469 ± 3.663	-6.469 ± 3.663	-6.469 ± 3.663	-6.470 ± 3.663	**-6.881 ± 3.987**
e2006-2048-256-C	-31.291 ± 60.597	-31.291 ± 60.597	-31.291 ± 60.597	-31.284 ± 60.599	**-32.026 ± 60.393**

Comparisons of objective values of all the methods for solving the ℓ_1 norm PCA problem.

Conclusions: **CD-SNCA** consistently gives the best performance.

Approximate Sparse Optimization

We consider the following problem:

$$\frac{1}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2+\rho\sum_{i=1}^{s}|\mathbf{x}_{[i]}^t|$$

Compared methods

1. Multi-Stage Convex Relaxation (MSCR)
2. Proximal DC algorithm (PDCA)
3. Subgradient method (SubGrad)
4. **CD-SCA**:
   $\mathbf{x}_{i^t}^{t+1}=\mathbf{x}_{i^t}^{t}+\arg\min_{\eta} 0.5(\mathbf{c}_{i^t}+\theta)\eta^2+\rho|\mathbf{x}_{i^t}^t+\eta|+[\nabla f(\mathbf{x}^t)-\mathbf{g}^t]_{i^t}\cdot\eta$
5. **CD-SNCA**: $\mathbf{x}_{i^t}^{t+1}=\mathbf{x}_{i^t}^{t}+\arg\min_{\eta}\frac{\mathbf{c}_i+\theta}{2}\eta^2+\nabla_{i^t}f(\mathbf{x}^t)\eta+\rho|\mathbf{x}_{i^t}^t+\eta|-\rho\sum_{i=1}^{s}|(\mathbf{x}^t+\eta e_i)_{[i]}|$

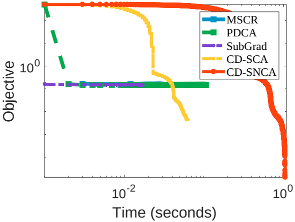
Experimental Results

	MSCR	PDCA	SubGrad	CD-SCA	CD-SNCA
randn-256-1024	0.090 ± 0.017	0.090 ± 0.016	0.775 ± 0.040	0.092 ± 0.018	**0.034 ± 0.004**
randn-256-2048	0.052 ± 0.009	0.052 ± 0.010	1.485 ± 0.030	0.061 ± 0.012	**0.027 ± 0.002**
randn-1024-256	1.887 ± 0.353	1.884 ± 0.352	2.215 ± 0.379	1.881 ± 0.337	**1.681 ± 0.346**
randn-2048-256	3.795 ± 0.518	3.794 ± 0.518	4.127 ± 0.525	3.772 ± 0.522	**3.578 ± 0.484**
e2006-256-1024	0.217 ± 0.553	0.217 ± 0.553	0.597 ± 0.391	0.218 ± 0.556	**0.087 ± 0.212**
e2006-256-2048	0.050 ± 0.068	0.050 ± 0.068	0.837 ± 0.209	0.050 ± 0.068	**0.025 ± 0.032**
e2006-1024-256	3.078 ± 2.928	3.078 ± 2.928	3.112 ± 2.844	3.097 ± 2.960	**2.697 ± 2.545**
e2006-2048-256	1.799 ± 1.453	1.799 ± 1.453	1.918 ± 1.518	1.805 ± 1.456	**1.688 ± 1.398**
randn-256-1024-C	0.086 ± 0.012	0.087 ± 0.012	0.775 ± 0.038	0.083 ± 0.011	**0.033 ± 0.002**
randn-256-2048-C	0.043 ± 0.006	0.044 ± 0.006	1.472 ± 0.027	0.051 ± 0.009	**0.026 ± 0.001**
randn-1024-256-C	1.997 ± 0.250	1.998 ± 0.250	2.351 ± 0.297	1.979 ± 0.265	**1.781 ± 0.244**
randn-2048-256-C	3.618 ± 0.681	3.617 ± 0.682	3.965 ± 0.717	3.619 ± 0.679	**3.420 ± 0.673**
e2006-256-1024-C	0.031 ± 0.031	0.031 ± 0.031	0.339 ± 0.073	0.030 ± 0.028	**0.015 ± 0.014**
e2006-256-2048-C	0.217 ± 0.575	0.217 ± 0.575	0.596 ± 0.418	0.215 ± 0.568	**0.071 ± 0.176**
e2006-1024-256-C	3.789 ± 4.206	3.798 ± 4.213	3.955 ± 4.363	3.851 ± 4.339	**3.398 ± 3.855**
e2006-2048-256-C	4.480 ± 6.916	4.482 ± 6.918	4.710 ± 7.292	4.461 ± 6.844	**4.200 ± 6.608**

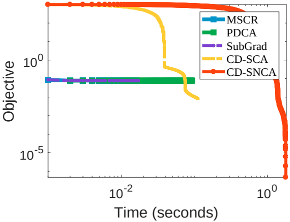
Comparisons of objective values of all the methods for solving the approximate sparse optimization problem.

Conclusions: **CD-SNCA** consistently gives the best performance.

Computational Efficiency



(a) randn-256-1024



(b) randn-256-2048

Conclusions:

CD-SNCA generally takes a little more time to converge.

CD-SNCA generally achieves higher accuracy.

Discussions and Extensions: Equivalent Reformulations for the ℓ_p Norm Generalized Eigenvalue Problem

Equivalent Reformulations

We consider the following problems with $\mathbf{Q} \succ \mathbf{0}$:

$$\min_{\mathbf{x}} \ \mathcal{F}_1(\mathbf{x}) \triangleq \tfrac{\alpha}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \|\mathbf{A}\mathbf{x}\|_p \qquad (2)$$

$$\min_{\mathbf{x}} \ \mathcal{F}_2(\mathbf{x}) \triangleq -\|\mathbf{A}\mathbf{x}\|_p, \ \ s.t. \ \mathbf{x}^T\mathbf{Q}\mathbf{x} \leq 1 \qquad (3)$$

$$\min_{\mathbf{x}} \ \mathcal{F}_3(\mathbf{x}) \triangleq \tfrac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}, \ \ s.t. \ \|\mathbf{A}\mathbf{x}\|_p \geq 1 \qquad (4)$$

We have the following results.

(a) If $\bar{\mathbf{x}}$ is an optimal solution to (2), then $\pm\bar{\mathbf{x}}(\bar{\mathbf{x}}^T\mathbf{Q}\bar{\mathbf{x}})^{-\frac{1}{2}}$ and $\frac{\pm\bar{\mathbf{x}}}{\|\mathbf{A}\bar{\mathbf{x}}\|_p}$ are respectively optimal solutions to (3) and (4).

(b) If $\bar{\mathbf{y}}$ is an optimal solution to (3), then $\frac{\pm\|\mathbf{A}\bar{\mathbf{y}}\|_p\cdot\bar{\mathbf{y}}}{\alpha\bar{\mathbf{y}}^T\mathbf{Q}\bar{\mathbf{y}}}$ and $\frac{\pm\bar{\mathbf{y}}}{\|\mathbf{A}\bar{\mathbf{y}}\|_p}$ are respectively optimal solutions to (2) and (4).

(c) If $\bar{\mathbf{z}}$ is an optimal solution to (4), then $\frac{\pm\bar{\mathbf{z}}\|\mathbf{A}\bar{\mathbf{z}}\|_p}{\alpha\bar{\mathbf{z}}^T\mathbf{Q}\bar{\mathbf{z}}}$ and $\pm\bar{\mathbf{z}}(\bar{\mathbf{z}}^T\mathbf{Q}\bar{\mathbf{z}})^{-\frac{1}{2}}$ are respectively optimal solutions to (2) and (3).

Discussions and Extensions: A Local Analysis for the PCA Problem

A Local Analysis for the PCA Problem

The PCA problem:

$$\max_{\mathbf{v}} \ \mathbf{v}^T\mathbf{C}\mathbf{v}, \ s.t. \ \|\mathbf{v}\| = 1$$

where $\mathbf{C} \succcurlyeq \mathbf{0}$ is given.

Equivalent problem:

$$\min_{\mathbf{x}} \ \mathcal{F}(\mathbf{x}) = \frac{\alpha}{2}\|\mathbf{x}\|_2^2 - \sqrt{\mathbf{x}^T\mathbf{C}\mathbf{x}}. \qquad (5)$$

for any given constant $\alpha > 0$.

We assume

$\mathbf{C} = \sum_{i=1}^n \boldsymbol{\lambda}_i \mathbf{u}_i \mathbf{u}_i^T = \mathbf{U}^T \text{diag}(\boldsymbol{\lambda})\mathbf{U}, \ \ \boldsymbol{\lambda}_1 \geq \boldsymbol{\lambda}_2 \geq ... \geq \boldsymbol{\lambda}_n \geq 0.$

A Local Analysis for the PCA Problem

Theorem

We have the following results:

(a) *The set of critical points of Problem (5) are* $\{\{\mathbf{0}\} \cup \{\pm\frac{\sqrt{\lambda_k}}{\alpha}\mathbf{u}_k : k = 1, ..., n\}\}$.

(b) *The PCA Problem in (5) has at most two local minima* $\{\pm\frac{\sqrt{\lambda_1}}{\alpha}\mathbf{u}_1\}$ *which are the global optima with* $\mathcal{F}(\bar{\mathbf{x}}) = -\frac{\lambda_1}{2\alpha}$.

A Local Analysis for the PCA Problem

Theorem

We define $\delta \triangleq 1 - \frac{\lambda_2}{\lambda_1}, \xi \triangleq \frac{\lambda_1}{6}\left(-1 - \frac{3}{\sqrt{\lambda_1}} + \sqrt{(1 + \frac{3}{\sqrt{\lambda_1}})^2 + \frac{12}{\lambda_1}\delta}\right)$. *Assume that* $0 < \delta < 1$. *When* $\mathbf{x}$ *is sufficiently close to the global optimal solution* $\bar{\mathbf{x}}$ *such that* $\|\mathbf{x} - \bar{\mathbf{x}}\| \leq \varpi$ *with* $\varpi < \bar{\varpi} \triangleq \min\{\sqrt{\lambda_1}\mathcal{K}(\frac{\lambda_2}{\lambda_1}), \xi\}$, *we have:*

(a) $\sqrt{\lambda_1} - \varpi \leq \|\mathbf{x}\| \leq \sqrt{\lambda_1} + \varpi$.

(b) $\lambda_1 - \varpi\sqrt{\lambda_1} \leq \|\mathbf{x}\|_{\mathbf{C}} \leq \lambda_1 + \varpi\sqrt{\lambda_1}$.

(c) $\lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \rho \mathbf{I} \succeq \mathbf{x}\mathbf{x}^T \succeq \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T - \rho \mathbf{I}$ *with* $\rho \triangleq 3\varpi^2 + 2\varpi\sqrt{\lambda_1}$.

(d) $\tau \mathbf{I} \succeq \nabla^2 \mathcal{F}(\mathbf{x}) \succeq \sigma \mathbf{I}$ *with* $\sigma \triangleq 1 - \frac{\lambda_2}{\lambda_1} - \varpi(1 + \frac{3}{\sqrt{\lambda_1}}) - \frac{3\varpi^2}{\lambda_1} > 0$ *and* $\tau \triangleq 1 + \frac{\lambda_1^2(\sqrt{\lambda_1} + \varpi)^2}{(\lambda_1 - \varpi\sqrt{\lambda_1})^3}$.

A Local Analysis for the PCA Problem

Theorem (Convergence Rate of CD-SNCA for the PCA Problem)

. We assume that the random-coordinate selection rule is used. Assume that $\|\mathbf{x}^t - \bar{\mathbf{x}}\| \leq \bar{\varpi}$ that $\mathcal{F}(\cdot)$ is σ-strongly convex and τ-smooth. Here the parameters $\bar{\varpi}, \sigma$ and τ are define in Theorem 9. We define $r_t^2 \triangleq \frac{(1+\sigma)\tau}{2}\|\mathbf{x} - \bar{\mathbf{x}}\|_2^2$ and $\beta \triangleq \frac{2\sigma}{1+\sigma}$. We have:

$$\mathbb{E}[r_t^2] \leq (1 - \frac{\beta}{n})^{t+1} \left(r_0^2 + \mathcal{F}(\mathbf{x}^0) - \mathcal{F}(\bar{\mathbf{x}})\right)$$

Note that the theorem above does not rely on the weak convexity condition or the sharpness condition of $\mathcal{F}(\cdot)$.

Discussions and Extensions: Examples for Optimality Hierarchy between the Optimality Conditions

The First Running Example

- . We consider the following problem:

$$\min_{\mathbf{x}} \ \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \langle \mathbf{x}, \mathbf{p} \rangle - \|\mathbf{A}\mathbf{x}\|_1$$

with using the following parameters:

$$\mathbf{Q} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}, \ \mathbf{p} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} 1 & -1 & 1 \\ 3 & 1 & 0 \\ 4 & 2 & -1 \end{pmatrix}.$$

The First Running Example

y	x	Function Value	Critical Point	CWS Point
$[1;1;1]$	$[1.75;0;-1]$	-6.625	**Yes**	No
$[1;1;[-1,1]]$	NA	NA	No	No
$[1;1;-1]$	$[-0.25;-2;-1]$	-8.125	No	No
$[1;[-1,1];1]$	NA	NA	No	No
$[1;[-1,1];[-1,1]]$	NA	NA	No	No
$[1;[-1,1];-1]$	NA	NA	No	No
$[1;-1;1]$	$[0.25;-2;-3]$	-4.1250	No	No
$[1;-1;[-1,1]]$	$[-0.3333;0.2667;-0.1333]$	-1.9956	No	No
$[1;-1;-1]$	$[-1.75;-4;-3]$	-16.1250	No	No
$[[-1,1];1;1]$	NA	NA	No	No
$[[-1,1];1;[-1,1]]$	NA	NA	No	No
$[[-1,1]1;-1]$	$[0;-2;-2]$	-6.0000	No	No
$[[-1,1];[-1,1];1]$	NA	NA	No	No
$[[-1,1];[-1,1];[-1,1]]$	$[0;0;0]$	0	**Yes**	No
$[[-1,1];[-1,1];-1]$	$[0;0;0]$	0	**Yes**	No
$[[-1,1];-1;1]$	NA	NA	No	No
$[[-1,1];-1;[-1,1]]$	$[0;0;0]$	0	**Yes**	No
$[[-1,1];-1;-1]$	$[0;0;0]$	0	**Yes**	No
$[-1;1;1]$	$[1.25;0;-3]$	-7.6250	**Yes**	No
$[-1;1;[-1,1]]$	NA	NA	No	No
$[-1;1;-1]$	$[-0.75;-2;-3]$	-12.1250	No	No
$[-1;[-1,1];1]$	NA	NA	No	No
$[-1;[-1,1];[-1,1]]$	$[0;0;0]$	0	**Yes**	No
$[-1;[-1,1];-1]$	$[0;0;0]$	0	**Yes**	No
$[-1;-1;1]$	$[-0.25;-2;-5]$	-6.6250	No	No
$[-1;-1;[-1,1]]$	$[0;0;0]$	0	**Yes**	No
$[-1;-1;-1]$	$[-2.25;-4;-5]$	-18.625	**Yes**	**Yes**

Table: Solutions satisfying optimality conditions.

The Second Running Example

- **The Second Running Example**. We consider the following example:

$$\min_{\mathbf{x}} \ \frac{1}{2}\mathbf{x}^T\mathbf{x} - \|\mathbf{A}\mathbf{x}\|_2$$

with using the following parameter:

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 0 & 2 \\ 3 & 1 & 0 \\ 4 & 2 & -1 \end{pmatrix}.$$

The Second Running Example

$(\lambda_i,\ \mathbf{u}_i)$	$\mathbf{x}$	Function Value	Critical Point	CWS Point
$(0.5468,\ [-0.2934, 0.8139, 0.5015])$	$\pm[-0.2169, 0.6019, 0.3709]$	-5.7418	**Yes**	No
$(7.8324,\ [0.1733, -0.4707, 0.8651])$	$\pm[0.4850, -1.3172, 2.4212]$	-82.2404	**Yes**	No
$(33.6207,\ [-0.9402, -0.3407, 0.0030])$	$\pm[-5.4514, -1.9755, 0.0172]$	-353.0178	**Yes**	**Yes**
	$[0, 0, 0]$	0	**Yes**	No

Table: Solutions satisfying optimality conditions.

The Third Running Example

- **The Third Running Example**. We consider the following example:

$$\min_{\mathbf{x}} \ \frac{1}{2}\mathbf{x}^T\mathbf{x} - \|\mathbf{A}\mathbf{x}\|_\infty$$

with using the following parameter:

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 0 & 2 \\ 3 & 1 & 0 \\ 4 & 2 & -1 \end{pmatrix}.$$

The Third Running Example

y	x	Function Value	Critical Point	CWS Point
$[1; 0; 0; 0]$	$[1; -1; 1]$	-2.5000	**Yes**	No
$[0; 1; 0; 0]$	$[2; 0; 2]$	-4.0000	**Yes**	No
$[0; 0; 1; 0]$	$[3; 1; 0]$	-9.0000	**Yes**	No
$[0; 0; 0; 1]$	$[4; 2; -1]$	-10.5000	**Yes**	**Yes**
$[-1; 0; 0; 0]$	$[-1; 1; -1]$	-2.5000	**Yes**	No
$[0; -1; 0; 0]$	$[-2; 0; -2]$	-4.0000	**Yes**	No
$[0; 0; -1; 0]$	$[-3; -1; 0]$	-9.0000	**Yes**	No
$[0; 0; 0; -1]$	$[-4; -2; 1]$	-10.5000	**Yes**	**Yes**

Table: Solutions satisfying optimality conditions.

Thank You!