Coordinate Descent Methods for Fractional Minimization

Ganzhao Yuan

Peng Cheng Laboratory, Shenzhen, China

Outline of This Talk

1. Introduction
2. Coordinate Descent Methods
3. Theoretical Analysis
   - Convex-Convex FMP
   - Convex-Concave FMP
4. A Breakpoint Searching Method
5. Experimental Results

Introduction

Introduction

The Fractional Minimization Problem (FMP):

$$\bar{\mathbf{x}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \; F(\mathbf{x}) \triangleq \frac{f(\mathbf{x}) + h(\mathbf{x})}{g(\mathbf{x})} \tag{1}$$

Assumptions

1. $f(\mathbf{x}) + h(\mathbf{x}) \geq 0$ and $g(\mathbf{x}) > 0$ for all $\mathbf{x}$.
2. $f(\cdot)$ 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.

3. Two Cases for $g(\cdot)$:
   1. Convex-Convex FMP: $g(\cdot)$ is cvx, not necessarily differentiable
   2. Convex-Concave FMP: $g(\cdot)$ is concave and differentiable

Introduction

The Fractional Minimization Problem (FMP):

$$\bar{\mathbf{x}} \in \arg\min_{\mathbf{x}\in\mathbb{R}^n} \ F(\mathbf{x}) \triangleq \frac{f(\mathbf{x}) + h(\mathbf{x})}{g(\mathbf{x})}$$

Assumptions

1. $h(\cdot) = \sum_{i=1}^{n} h_i(\mathbf{x}_i)$ is convex and coordinate-wise separable.
2. One of the following subproblem can be solved exactly:

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

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

Remarks. $\arg\min_{\mathbf{x}} \frac{f(\mathbf{x})+h(\mathbf{x})}{g(\mathbf{x})} \Leftrightarrow \arg\max_{\mathbf{x}} \frac{g(\mathbf{x})}{f(\mathbf{x})+h(\mathbf{x})} \Leftrightarrow$
$\arg\max_{\mathbf{x}} \frac{g(\mathbf{x})+\lambda(f(\mathbf{x})+h(\mathbf{x}))}{f(\mathbf{x})+h(\mathbf{x})} \Leftrightarrow \arg\min_{\mathbf{x}} \frac{f(\mathbf{x})+h(\mathbf{x})}{g(\mathbf{x})+\lambda(f(\mathbf{x})+h(\mathbf{x}))}$.

Examples

1. ℓ_p Norm PCA Problem

$$\max_{\mathbf{x}} \ \|\mathbf{Gx}\|_p, \ s.t. \ \|\mathbf{x}\| = 1$$

$$\min_{\mathbf{x}} \ \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{Gx}\|_p^2} \ \Leftrightarrow \ \min_{\mathbf{x}} \ \frac{\|\mathbf{x}\|_2^2+1}{\|\mathbf{Gx}\|_p}$$

2. Sparse Recovery

$$\min_{\mathbf{x}} \ \frac{\frac{1}{2}\|\mathbf{Gx}-\mathbf{y}\|_2^2+\gamma\|\mathbf{x}\|_1}{\gamma\sum_{j=1}^{k}|\mathbf{x}_{[j]}|}, s.t.\|\mathbf{x}\|_\infty \leq \vartheta$$

3. Independent Component Analysis (ICA)

$$\bar{\mathbf{v}} \in \mathbb{R}^n = \arg\max_{\mathbf{v}} \|\mathbf{Gv}\|_4^4, \ s.t. \ \|\mathbf{v}\| = 1$$

$$\bar{\mathbf{x}} = \arg\min_{\mathbf{x}} \ \frac{\mathbf{x}^T\mathbf{x}}{\|\mathbf{Gx}\|_4^2}$$

Note that $\bar{\mathbf{v}} = \frac{\bar{\mathbf{x}}}{\|\bar{\mathbf{x}}\|}$.

Examples

1. Regularized Total Least Squares problem

$$\min_{\mathbf{E},\mathbf{x},\mathbf{r}} \ \|\mathbf{E}\|_F^2 + \|\mathbf{r}\|_2^2, \ \ s.t. \ (\mathbf{A}+\mathbf{E})\mathbf{x} \leq \mathbf{b} - \mathbf{r}$$

It is shown that it is equivalent to the following problem:

$$\min_{\mathbf{x}} \ \frac{\|\max(0, \mathbf{A}\mathbf{x} - \mathbf{b})\|_2^2}{\|\mathbf{x}\|_2^2 + 1}$$

2. Transmit Beamforming

$$\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{x}\|_2^2, \ \ s.t. \ |\mathbf{A}\mathbf{x}| \geq \mathbf{1}$$

It is equivalent to the following problem:

$$\min_{\mathbf{x}} \ \frac{\|\mathbf{x}\|_2^2}{\lambda\|\mathbf{x}\|_2^2 + \min(|\mathbf{A}\mathbf{x}|)^2}$$

Related Fractional Minimization Algorithms (i)

1. Dinkelbach's Parametric Algorithm

   $\min_{\mathbf{x}} f(\mathbf{x}) + h(\mathbf{x}) - \bar{\lambda} g(\mathbf{x})$, where $\bar{\lambda} = \frac{f(\bar{\mathbf{x}})+h(\bar{\mathbf{x}})}{g(\bar{\mathbf{x}})}$.

   Since $\bar{\lambda}$ is unknown, iterative procedures are needed.

$$\mathbf{x}^{t+1} = \arg\min_{\mathbf{x}} \; f(\mathbf{x}) + h(\mathbf{x}) - \lambda^t g(\mathbf{x})$$

$$\lambda^t = \frac{f(\mathbf{x}^t) + h(\mathbf{x}^t)}{g(\mathbf{x}^t)}$$

2. Proximal Gradient Algorithm: $g(\mathbf{x}^t)$ is differentiable

$$\mathbf{x}^{t+1} = \arg\min_{\mathbf{x}} f(\mathbf{x}) + h(\mathbf{x}) - \lambda^t \langle \nabla g(\mathbf{x}^t), \mathbf{x} - \mathbf{x}^t \rangle + \frac{1}{2\eta^t} \|\mathbf{x} - \mathbf{x}^t\|_2^2$$

$$\lambda^t = F(\mathbf{x}^t)$$

   Here $\eta^t > 0$ is a stepsize.

Related Fractional Minimization Algorithms (ii)

- Proximal Gradient-Subgradient Algorithm

  We assume that $\nabla f(\cdot)$ is L-Lipschitz continuous:

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

  The algorithm:

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

$$\lambda^t = F(\mathbf{x}^t).$$

- Quadratic Transform Parametric Algorithm: originally designed for solving multiple-ratio FMPs.

$$\min_{\mathbf{x}} \frac{-g(\mathbf{x})}{f(\mathbf{x})+h(\mathbf{x})} \Leftrightarrow \min_{\mathbf{x},\beta} \beta^2(f(\mathbf{x})+h(\mathbf{x})) - 2\beta\sqrt{g(\mathbf{x})}$$

$$\mathbf{x}^{t+1} = \arg\min_{\mathbf{x}} (\beta^t)^2(f(\mathbf{x})+h(\mathbf{x})) - 2\beta^t\sqrt{g(\mathbf{x})}$$

$$\beta^t = \sqrt{g(\mathbf{x}^t)}/(f(\mathbf{x}^t)+h(\mathbf{x}^t))$$

Related Fractional Minimization Algorithms (iii)

- Charnes-Cooper Transform Algorithm.

  Using the transformation

$$\mathbf{y} = \frac{\mathbf{x}}{g(\mathbf{x})}, \; t = \frac{1}{g(\mathbf{x})}$$

  Equivalent Problem:

$$\min_{t,\mathbf{y}} tf(\mathbf{y}/t) + th(\mathbf{y}/t), \; s.t. \; tg(\mathbf{y}/t) = 1$$

  Remarks. (i) $tf(\mathbf{y}/t)$ is convex jointly w.r.t. $\mathbf{y}$ and t if $f(\cdot)$ is convex. (ii) Perspective operation preserves convexity.

- Other fractional optimization algorithms: PGSA with line search, extrapolated PGSA

Related Work on CD Methods

- Convex Problems
  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 problems (Lasso,SVM,NMF,PageRank)
- Nonconvex Problems
  1. Strong optimality guarantees and superior empirical performance
  2. Its popularity continues to grow: ℓ_0 norm minimization, Eigenvalue complementarity problem, K-means clustering, Sparse phase retrieval, Penalized regression, Resource allocation problem, Leading eigenvector computation ...

Theory in Nonconvex Optimization

1. Optimality analysis: Finding 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
   - weakly convex / bounded nonconvex
   - Luo-Tseng error bound assumption
   - Kurdyka-Łojasiewicz inequality
   - sharpness condition

Contributions

1. Two new CD methods based on sequential nonconvex approximation: Fractional CD, Parametric CD
2. Optimality Analysis: *FCD*-point and *PCD*-point are equivalent, they are stronger than the critical/directional point
3. Convergence Analysis:
    1. Convex-Convex: Q-linearly convergence rate under *Luo-Tseng Error bound* condition or *Sharpness* condition
    2. Convex-Concave: Sublinear convergence rate
4. Breakpoint searching methods for computing the proximal operators, Extensive experiments on many applications

Coordinate Descent Methods

Coordinate Descent Methods

The Raw CD Method:

$$\bar{\eta}^t \in \arg\min_{\eta \in \mathbb{R}} \ \frac{f(\mathbf{x}^t + \eta e_i) + h(\mathbf{x}^t + \eta e_i)}{g(\mathbf{x}^t + \eta e_i)}, \ \mathbf{x}^{t+1} = \mathbf{x}^t + \bar{\eta}^t e_i$$

The Fractional CD Method (♠ The Proposed Method):

$$\bar{\eta}^t = \arg\min_{\eta} \frac{\mathbf{Q}_i(\mathbf{x}, \eta) + h(\mathbf{x} + \eta e_i)}{g(\mathbf{x}^t + \eta e_i)}, \ \mathbf{x}^{t+1} = \mathbf{x}^t + \bar{\eta}^t e_i$$

The Parametric CD Method (♠ The Proposed Method):

$$\bar{\eta}^t = \arg\min_{\eta} \mathbf{Q}_i(\mathbf{x}, \eta) + h(\mathbf{x} + \eta e_i) - \lambda^t g(\mathbf{x}^t + \eta e_i), \ \mathbf{x}^{t+1} = \mathbf{x}^t + \bar{\eta}^t e_i$$

The Proximal Gradient-Subgradient CD Method:

$$\bar{\eta}^t = \arg\min_{\eta} \mathbf{Q}_i(\mathbf{x}, \eta) + h(\mathbf{x} + \eta e_i) - \lambda^t \partial_i g(\mathbf{x}^t)\eta, \ \mathbf{x}^{t+1} = \mathbf{x}^t + \bar{\eta}^t e_i$$

Recall that: $f(\mathbf{x} + \eta e_i) \leq \mathbf{Q}_i(\mathbf{x}, \eta) \triangleq f(\mathbf{x}) + \nabla_i f(\mathbf{x})\eta + \frac{\mathbf{c}_i}{2}\eta^2$

if $g(\cdot)$ is cvx: $-g(\mathbf{x}^t + \eta e_i) \leq -g(\mathbf{x}^t) - \langle \partial g(\mathbf{x}^t), (\mathbf{x}^t + \eta e_i) - \mathbf{x}^t \rangle$

CD Methods for Fractional Minimization

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 $i^t \in \{1, ..., n\}$.

S2 Define

$$\mathcal{J}_i(\mathbf{x}, \eta, \theta) \triangleq f(\mathbf{x}) + \nabla_i f(\mathbf{x})\eta + \frac{\mathbf{c}_i+\theta}{2}\eta^2 + h(\mathbf{x} + \eta e_i)$$

Solve one of the following subproblems *globally*.

- Option I: *Fractional Coordinate Descent* (**FCD**):

$$\bar{\eta}^t = \mathcal{P}_{i^t}(\mathbf{x}^t) \triangleq \arg\min_{\eta} \frac{\mathcal{J}_{i^t}(\mathbf{x}^t, \eta, \theta)}{g(\mathbf{x}^t + \eta e_{i^t})} \quad (2)$$

- Option II: *Parametric Coordinate Descent* (**PCD**):

$$\bar{\eta}^t = \mathcal{P}_{i^t}(\mathbf{x}^t) \triangleq \arg\min_{\eta} \mathcal{J}_{i^t}(\mathbf{x}^t, \eta, \theta) - F(\mathbf{x}^t)g(\mathbf{x}^t + \eta e_{i^t}) \quad (3)$$

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

S4 Increment t by 1

end

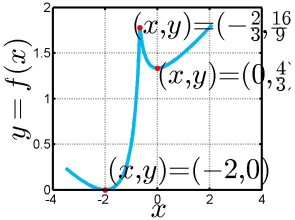
One-dimensional example

We consider the following simple one-dimensional example:

$$\min_{x} F(\mathbf{x}) \triangleq \frac{(x+2)^2}{|3x+2|+1}$$

The following table shows the points satisfying different optimality conditions.

x	$F(x)$	C-point	D-point	FCW-point	PCW-point
-2	0	Yes	Yes	Yes	Yes
$-\frac{2}{3}$	$(\frac{4}{3})^2$	Yes	No	No	No
0	$\frac{4}{3}$	Yes	Yes	No	No



Implementation for the ℓ_1 Norm PCA Problem (i)

$$\max_{\mathbf{x}} g(\mathbf{x}) \triangleq \|\mathbf{A}\mathbf{x}\|_1, \ s.t. \ \|\mathbf{x}\| = 1$$

$$g(\mathbf{x}) \text{ is convex: } -g(\mathbf{x}) \leq -g(\mathbf{x}^t) - \langle \partial g(\mathbf{x}^t), \ \mathbf{x} - \mathbf{x}^t \rangle$$

Power Method / Sub-gradient Method:

$$\mathbf{x}^{t+1} = \arg\min_{\|\mathbf{x}\|=1} -g(\mathbf{x}^t) - \langle \partial g(\mathbf{x}^t), \ \mathbf{x} - \mathbf{x}^t \rangle, \ \partial g(\mathbf{x}^t) = \mathbf{A}^T sign(\mathbf{A}\mathbf{x}^t)$$

$$= \frac{\mathbf{A}^T sign(\mathbf{A}\mathbf{x}^t)}{\|\mathbf{A}^T sign(\mathbf{A}\mathbf{x}^t)\|}$$

Equivalent fractional minimization problems:

$$\min_{\mathbf{x}} \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{A}\mathbf{x}\|_1^2}, \ \text{or} \ \min_{\mathbf{x}} \frac{\|\mathbf{x}\|_2^2 + 1}{\|\mathbf{A}\mathbf{x}\|_1}.$$

Implementation for the ℓ_1 Norm PCA Problem (ii)

We consider **PCD** to solve the $\min_{\mathbf{x}} \frac{\|\mathbf{x}\|_2^2+1}{\|\mathbf{A}\mathbf{x}\|_1}$.

The resulting univariate subproblem is:

$$\min_{\eta} \|\mathbf{x}^t + \eta e_{i^t}\|_2^2 + \frac{\theta}{2}\eta^2 - F(\mathbf{x}^t) g(\mathbf{x}^t + \eta e_{i^t})$$

$$\min_{\eta} \|\mathbf{x}^t + \eta e_{i^t}\|_2^2 + \frac{\theta}{2}\eta^2 - F(\mathbf{x}^t) \left(\|\mathbf{A}(\mathbf{x}^t + \eta e_{i^t})\|_1\right)$$

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

Implementation for the ℓ_1 Norm PCA Problem (iii)

We use breakpoint search method to solve the following problem:

$$\min_{\eta} p(\eta) \triangleq \frac{a}{2}\eta^2 + b\eta - \|\mathbf{g}\eta + \mathbf{d}\|_1 \quad (4)$$

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 $\Theta = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, ..., \boldsymbol{\eta}_{(2m+1)}\}$.

Theoretical Analysis for Convex-Convex FMP

Optimality Definition

Definition (Critical Point)

(Critical Point, or C-Point for short) A solution $\check{\mathbf{x}}$ is called a C-point if: $0 \in \partial F(\check{\mathbf{x}}) \triangleq \nabla f(\check{\mathbf{x}}) + \partial h(\check{\mathbf{x}}) - F(\check{\mathbf{x}}) \cdot \partial g(\check{\mathbf{x}})$.

Definition (Directional Point)

(Directional Point, or D-Point for short) A solution $\grave{\mathbf{x}}$ is called a D-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 (Fractional Coordinate-Wise Point)

(Fractional Coordinate-Wise Point, or *FCW*-Point for short) Given a constant $\theta \geq 0$. Define $\mathcal{K}_i(\mathbf{x}, \eta) \triangleq \frac{\mathcal{J}_i(\mathbf{x},\eta,\theta)}{g(\mathbf{x}+\eta e_i)}$. A solution $\ddot{\mathbf{x}}$ is called a *FCW*-point if: $\mathcal{K}_i(\ddot{\mathbf{x}}, 0) = \min_{\boldsymbol{\eta}_i} \mathcal{K}_i(\ddot{\mathbf{x}}, \boldsymbol{\eta}_i)$, $\forall i = 1, ..., n$.

Definition (Parametric Coordinate-Wise Point)

(Parametric Coordinate-Wise Point, or *PCW*-Point for short) Given a constant $\theta \geq 0$. Define $\mathcal{M}_i(\mathbf{x}, \eta) \triangleq \mathcal{J}_i(\mathbf{x}, \eta, \theta) - F(\mathbf{x})g(\mathbf{x} + \eta e_i)$. A solution $\dot{\mathbf{x}}$ is called a *PCW*-point if: $\mathcal{M}_i(\dot{\mathbf{x}}, 0) = \min_{\boldsymbol{\eta}_i} \mathcal{M}_i(\dot{\mathbf{x}}, \boldsymbol{\eta}_i)$, $\forall i = 1, ..., n$.

Theoretical Analysis

Assumption

(**Boundedness of the Denominator**) *There exists a constant* $\bar{g} > 0$ *such that* $\forall \mathbf{x} \in \{\mathbf{z} \mid F(\mathbf{z}) \leq F(\mathbf{x}^0)\},\ g(\mathbf{x}) \leq \bar{g}$.

Definition

(**Globally/Locally ρ-Bounded Non-Convexity**) A function $\tilde{g}(\mathbf{x}) = -g(\mathbf{x})$ is globally ρ-bounded non-convex if it holds that $\tilde{g}(\mathbf{x}) \leq \tilde{g}(\mathbf{y}) + \langle \mathbf{x} - \mathbf{y},\ \partial \tilde{g}(\mathbf{x}) \rangle + \frac{\rho}{2} \|\mathbf{x} - \mathbf{y}\|_2^2$ for all $\mathbf{x}$ and $\mathbf{y}$ with a constant $\rho < +\infty$. In particular, $\tilde{g}(\mathbf{x})$ is locally ρ-bounded non-convex if $\mathbf{x}$ is restricted to some point $\mathbf{x}$ with $\mathbf{x} = \check{\mathbf{x}}$.

Optimality Hierarchy (i)

Lemma (Properties of *FCW*-point and *PCW*-point.)

For any FCW-point $\ddot{\mathbf{x}}$ *and any PCW-point* $\dot{\mathbf{x}}$*, assume that* $\tilde{g}(\mathbf{x}) = -g(\mathbf{x})$ *is locally* ρ*-bounded non-convex at the point* $\ddot{\mathbf{x}}$ *(or* $\dot{\mathbf{x}}$*) with* $\rho < +\infty$*. We define* $\mathcal{C}(\mathbf{x}, \boldsymbol{\eta}) \triangleq \frac{1}{2}\|\boldsymbol{\eta}\|_{\mathbf{c}+\theta}^2 + \frac{\rho}{2}\|\boldsymbol{\eta}\|_2^2 F(\mathbf{x})$*. We have:*

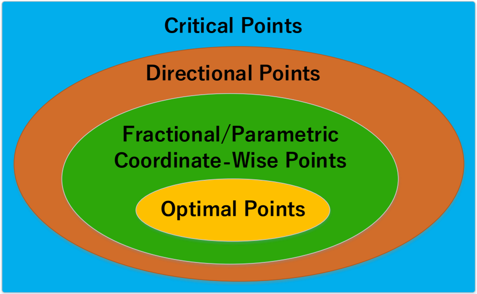
(i) $\forall \boldsymbol{\eta},\ F(\ddot{\mathbf{x}}) - F(\ddot{\mathbf{x}} + \boldsymbol{\eta}) \leq \frac{\mathcal{C}(\ddot{\mathbf{x}}, \boldsymbol{\eta})}{g(\ddot{\mathbf{x}} + \boldsymbol{\eta})}$.

(ii) $\forall \boldsymbol{\eta},\ F(\dot{\mathbf{x}}) - F(\dot{\mathbf{x}} + \boldsymbol{\eta}) \leq \frac{\mathcal{C}(\dot{\mathbf{x}}, \boldsymbol{\eta})}{g(\dot{\mathbf{x}} + \boldsymbol{\eta})}$.

Optimality Hierarchy (ii)

We use $\check{\mathbf{x}}$, $\grave{\mathbf{x}}$, $\dot{\mathbf{x}}$, $\ddot{\mathbf{x}}$, and $\bar{\mathbf{x}}$ to denote a C-point, a D-point, a FCW-point, a PCW-point, and an optimal point, respectively. Based on the the assumption made in the previous lemma. The following relation holds:

$$\{\bar{\mathbf{x}}\} \overset{\textbf{(a)}}{\subseteq} \{\ddot{\mathbf{x}}\} \overset{\textbf{(b)}}{\Leftrightarrow} \{\dot{\mathbf{x}}\} \overset{\textbf{(c)}}{\subseteq} \{\grave{\mathbf{x}}\} \overset{\textbf{(d)}}{\subseteq} \{\check{\mathbf{x}}\}$$



Optimality Hierarchy (iii)

How is *FCW*/*PCW*-point compared with local minimum point?

- *FCW*-point

$$\forall \boldsymbol{\eta},\ F(\ddot{\mathbf{x}}) \leq F(\ddot{\mathbf{x}} + \boldsymbol{\eta}) + \frac{\frac{1}{2}\|\boldsymbol{\eta}\|_{\mathbf{z}}^2}{g(\ddot{\mathbf{x}} + \boldsymbol{\eta})},\ \mathbf{z} = \mathbf{c} + \theta\mathbf{1} + \rho F(\ddot{\mathbf{x}})\mathbf{1}$$

- local minimum point

$$\forall \boldsymbol{\eta},\ F(\ddot{\mathbf{x}}) \leq F(\ddot{\mathbf{x}} + \boldsymbol{\eta}), \|\boldsymbol{\eta}\| \leq \epsilon$$

  where $\epsilon > 0$ is sufficiently small.

Conclusion: Neither condition is stronger than the other.

Global Convergence

Definition

Given $\epsilon > 0$, the solution $\mathbf{x}$ is said to be an ϵ-approximate *FCD*/*PCD* point if it holds that:

$$\frac{1}{n}\sum_{i=1}^{n} |\mathcal{P}_i(\mathbf{x})|^2 \triangleq Z(\mathbf{x}) \leq \epsilon.$$

Theorem (Global Convergence)

(a) *(***Sufficient Decrease Condition***)*
$F(\mathbf{x}^{t+1}) - F(\mathbf{x}^t) \leq -\frac{\theta}{2g(\mathbf{x}^{t+1})}\|\mathbf{x}^{t+1} - \mathbf{x}^t\|_2^2$.
(b) *FCD/PCD find an ϵ-approximate FCD-point/PCD-point in at most T iterations in expectation, where*

$$T \leq \lceil \frac{n\bar{g}[F(\mathbf{x}^0) - F(\bar{\mathbf{x}})]}{\theta\epsilon} \rceil = \mathcal{O}(\epsilon^{-1})$$

Convergence Rate

Assumption

(Luo-Tseng Error Bound) We define a residual function as $\mathcal{R}(\mathbf{x}) \triangleq \frac{1}{n}\sum_{i=1}^{n}|\mathcal{P}_i(\mathbf{x})|$, *where* $\mathcal{P}_i(\mathbf{x})$ *is defined in (2) (or (3)). For any* $\zeta \geq \min_{\mathbf{x}} F(\mathbf{x})$, *there exist scalars* $\delta > 0$ *and* $\varrho > 0$ *such that:*

$$dist(\mathbf{x}, \mathcal{X}) \leq \delta \cdot \mathcal{R}(\mathbf{x}), \ \ whenever\ F(\mathbf{x}) \leq \zeta,\ \mathcal{R}(\mathbf{x}) \leq \varrho. \tag{5}$$

Here, $dist(\mathbf{x}, \mathcal{X}) = \inf_{\mathbf{z}\in\mathcal{X}} \|\mathbf{z} - \mathbf{x}\|$, $\mathcal{X}$ *is the set of the FCW-point (or the PCW-point).*

Convergence Rate

Necessary first-order optimality conditions:

$$\alpha^t \partial_{i^t} g(\mathbf{x}^t + \bar{\eta}^t e_{i^t}) \in \partial \mathcal{J}_{i^t}(\mathbf{x}^t, \bar{\eta}^t, \theta), \;\; \alpha^t \triangleq \frac{\mathcal{J}_{i^t}(\mathbf{x}^t, \bar{\eta}^t, \theta)}{g(\mathbf{x}^{t+1})}.$$

$$F(\mathbf{x}^t) \cdot \partial_{i^t} g(\mathbf{x}^t + \bar{\eta}^t e_{i^t}) \in \partial \mathcal{J}_{i^t}(\mathbf{x}^t, \bar{\eta}^t, \theta). \tag{6}$$

Lemma

(**Property of FCD**) *The value of the parameter* α^t *defined in (6) is sandwiched as*

$F(\mathbf{x}^{t+1}) \leq \alpha^t \leq F(\mathbf{x}^{t+1}) + \sigma(F(\mathbf{x}^t) - F(\mathbf{x}^{t+1})) \leq \sigma F(\mathbf{x}^0)$ *with* $\sigma \triangleq \frac{\max(\mathbf{c}) + \theta}{\theta}$.

Convergence Rate

Theorem

(Convergence Rate of **FCD***). For any FCW-point* $\ddot{\mathbf{x}}$*, we define* $q^t \triangleq F(\mathbf{x}^t) - F(\ddot{\mathbf{x}})$, $r^t \triangleq \frac{1}{2}\|\mathbf{x}^t - \ddot{\mathbf{x}}\|_{\bar{\mathbf{c}}}^2$, $\bar{\mathbf{c}} \triangleq \mathbf{c} + \theta$*. Assume that* $\tilde{g}(\mathbf{x}) = -g(\mathbf{x})$ *is globally* ρ*-bounded non-convex, and* $F(\cdot)$ *satisfies Assumption 2. We define:* $\varpi \triangleq \frac{\max(\bar{\mathbf{c}})}{\min(\bar{\mathbf{c}})} \cdot \frac{\rho}{\theta} \cdot F(\mathbf{x}^0)$*. We have the following inequality:*

$(1-\varpi)\mathbb{E}_{i^t}[r^{t+1}] + \frac{g(\bar{\mathbf{x}})}{n} q^{t+1} \leq (1-\varpi) r^t + \frac{\varpi}{n} r^t$*. When the proximal parameter* θ *is sufficiently large such that* $\varpi \leq 1$*, we obtain:* $q^{t+1} \leq (\frac{\kappa_1}{\kappa_1 + \kappa_0})^{t+1} q^0$*, where* $\kappa_0 \triangleq \frac{g(\bar{\mathbf{x}})}{\bar{g}}$ *and* $\kappa_1 \triangleq (n+1)\max(\bar{\mathbf{c}})\delta^2/\theta$*.*

Convergence Rate

Theorem

(Convergence Rate of **PCD***). For any PCW-point $\dot{\mathbf{x}}$, we define $q^t \triangleq F(\mathbf{x}^t) - F(\dot{\mathbf{x}})$, $r^t \triangleq \frac{1}{2}\|\mathbf{x}^t - \dot{\mathbf{x}}\|_{\bar{\mathbf{c}}}^2$, $\bar{\mathbf{c}} \triangleq \mathbf{c} + \theta$. Assume that $\tilde{g}(\mathbf{x}) = -g(\mathbf{x})$ is globally ρ-bounded non-convex, and $F(\cdot)$ satisfies Assumption 2. We define: $\varpi \triangleq \frac{\rho}{\min(\bar{\mathbf{c}})} F(\mathbf{x}^0)$. We have the following inequality:*

$$\mathbb{E}_{i^t}[(1-\varpi) r^{t+1}] + \frac{\bar{g}}{n} q^{t+1} \leq (1-\varpi) r^t + \frac{\varpi}{n} r^t - \frac{g(\bar{\mathbf{x}})}{n} q^t + \frac{\bar{g}}{n} q^t.$$

When the proximal parameter θ is sufficiently large such that $\varpi \leq 1$, we obtain: $q^{t+1} \leq (\frac{\kappa_1 + 1 - \kappa_0}{\kappa_1 + 1})^{t+1} q^0$, where $\kappa_0 \triangleq \frac{g(\bar{\mathbf{x}})}{\bar{g}}$ and $\kappa_1 \triangleq (n+1)\max(\bar{\mathbf{c}})\delta^2/\theta$.

Convergence Rate

Remarks

1. Algorithm 1 converges to the *FCW*-point (or the *PCW*-point) with a Q-linear convergence rate.
2. We compare the convergence rate of **FCD** and **PCD** which depend on κ_0 and κ_1:
$(\frac{\kappa_1+1-\kappa_0}{\kappa_1+1}) - (\frac{\kappa_1}{\kappa_1+\kappa_0}) = \frac{1}{(\kappa_1+\kappa_0)(\kappa_1+1)}[\kappa_1(\kappa_1+\kappa_0) + (\kappa_1+\kappa_0) - \kappa_0(\kappa_1+\kappa_0) - \kappa_1(\kappa_1+1)] = \frac{\kappa_0(1-\kappa_0)}{(\kappa_1+\kappa_0)(\kappa_1+1)} \geq 0$.
Thus, **FCD** is faster than **PCD**.
3. The condition $\varpi < 1$ essentially implies that the θ is chosen to be large enough that the univariate subproblem is convex.

Theoretical Analysis for Convex-Concave FMP

Theoretical Analysis

Proposition

(i) $F(\cdot)$ *is quasiconvex that:* $F(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \leq \max(F(\mathbf{x}), F(\mathbf{y})), \forall \alpha \in [0,1], \mathbf{x}, \mathbf{y}$. *(ii) Any critical point of Problem (2) is a global minimum.*

Theorem (Convergence Rate.)

For any global optimal solution $\bar{\mathbf{x}}$*-point of Problem (2), we define* $q^t \triangleq F(\mathbf{x}^t) - F(\bar{\mathbf{x}}),\ r^t \triangleq \frac{1}{2}\|\mathbf{x}^t - \bar{\mathbf{x}}\|_{\bar{\mathbf{c}}}^2,\ \bar{\mathbf{c}} \triangleq \mathbf{c} + \theta$.

For **FCD**, *we have:* $\mathbb{E}_{\xi^{t-1}}[q^t] \leq \frac{n(\bar{g}\sigma q^0 + r^0)}{g(\bar{\mathbf{x}})t}$, *where* σ *is defined in Lemma 9.*

For **PCD**, *we have:* $\mathbb{E}_{\xi^{t-1}}[q^t] \leq \frac{n(\bar{g} q^0 + r^0)}{g(\bar{\mathbf{x}})(t+1)}$.

A Breakpoint Searching Method

A Breakpoint Searching Method

Two steps:

1. identifies all the possible critical points / breakpoints Θ
2. picks the solution from Θ that leads to the lowest value as the optimal solution.

Examples:

1. ℓ_∞ Norm PCA Problem: $F(\mathbf{x}) \triangleq \frac{\|\mathbf{x}\|_2^2+1}{\|\mathbf{A}\mathbf{x}\|_\infty}$.
2. Sparse Recovery Problem: $F(\mathbf{x}) \triangleq \frac{\frac{1}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2+\gamma\|\mathbf{x}\|_1}{\gamma\sum_{j=1}^{k}|\mathbf{x}_{[j]}|}$.
3. ICA Problem: $F(\mathbf{x}) \triangleq \frac{\mathbf{x}^T\mathbf{x}}{\|\mathbf{G}\mathbf{x}\|_4^2}$.
4. RTLS Problem: $F(\mathbf{x}) \triangleq \min_{\mathbf{x}} \frac{\|\max(0,\mathbf{A}\mathbf{x}-\mathbf{b})\|_2^2}{\|\mathbf{x}\|_2^2+1}$.
5. Transmit Beamforming Problem: $\frac{\|\mathbf{x}\|_2^2}{\lambda\|\mathbf{x}\|_2^2+\min(|\mathbf{A}\mathbf{x}|)^2}$

Example 1. ℓ_∞ Norm PCA Problem: $F(\mathbf{x}) \triangleq \frac{\|\mathbf{x}\|_2^2+1}{\|\mathbf{A}\mathbf{x}\|_\infty}$

We consider Parametric CD to solve the ℓ_∞ Norm PCA Problem.

The reduced univariate subproblem is

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

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. Sparse Recovery: $F(\mathbf{x}) \triangleq \frac{\frac{1}{2}\|\mathbf{G}\mathbf{x}-\mathbf{y}\|_2^2+\gamma\|\mathbf{x}\|_1}{\gamma\sum_{j=1}^{k}|\mathbf{x}_{[j]}|}$

We consider Parametric CD to solve this problem. The reduced univariate subproblem is

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

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 3. ICA Problem: $F(\mathbf{x}) \triangleq \frac{\mathbf{x}^T\mathbf{x}}{\|\mathbf{G}\mathbf{x}\|_4^2}$

We consider Fractional CD to solve the ICA Problem.

The reduced univariate subproblem is $\min_\eta \frac{\|\mathbf{x}^t\|_2^2 + 2\mathbf{x}_i\eta + \frac{2+\theta}{2}\eta^2}{\sqrt{\|\mathbf{G}(\mathbf{x}^t + \eta e_{i^t})\|_4^4}}$

$$\min_\eta p(\eta) \triangleq \frac{a_2\eta^2 + a_1\eta + a_0}{\sqrt{b_4\eta^4 + b_3\eta^3 + b_2\eta^2 + b_1\eta + b_0}}$$

Setting the gradient of $p(\cdot)$ to zero yields: $2a_2\eta + a_1 = p(\eta)\frac{1}{2}(b_4\eta^4 + b_3\eta^3 + b_2\eta^2 + b_1\eta + b_0)^{-\frac{1}{2}} \cdot (4b_4\eta^3 + 3b_3\eta^2 + 2b_2\eta + b_1)$.

It reduces to a quartic equation which can be solved analytically by Lodovico Ferrari's method: $c_4\eta^4 + c_3\eta^3 + c_2\eta^2 + c_1\eta + c_0 = 0$.

This problem contains 4 breakpoints $\Theta = \{\bar{\eta}_1, \bar{\eta}_2, \bar{\eta}_3, \bar{\eta}_4\}$.

Example 4. RTLS: $F(\mathbf{x}) \triangleq \min_{\mathbf{x}} \frac{\|\max(0,\mathbf{Ax}-\mathbf{b})\|_2^2}{\|\mathbf{x}\|_2^2+1}$

We consider Parametric CD to solve the RTLS Problem.

The reduced univariate subproblem is

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

Letting $p = 2$, 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:

$$0 = -a\eta - b = \frac{\langle \mathbf{g}, \mathbf{g}\eta + \mathbf{d}\rangle}{\|\mathbf{g}\eta - \mathbf{d}\|} \Leftrightarrow \|\mathbf{g}\eta - \mathbf{d}\|(-a\eta - b) = \langle \mathbf{g}, \mathbf{g}\eta + \mathbf{d}\rangle$$

$$\Leftrightarrow \|\mathbf{g}\eta - \mathbf{d}\|_2^2(a\eta + b)^2 = (\langle \mathbf{g}, \mathbf{g}\eta + \mathbf{d}\rangle)^2$$

Solving this quartic equation we obtain all of its real roots $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, ..., \boldsymbol{\eta}_c\}$ with $1 \le c \le 4$.

This problem at most contains 4 breakpoints $\Theta = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, ..., \boldsymbol{\eta}_c\}$.

Example 4. Beamforming Problem: $\frac{\|\mathbf{x}\|_2^2}{\lambda\|\mathbf{x}\|_2^2+\min(|\mathbf{A}\mathbf{x}|)^2}$

We consider Parametric CD to solve the beamforming Problem.

The reduced univariate subproblem is

$$
\begin{aligned}
& \min_\eta \tfrac{a}{2}\eta^2 + b\eta - \lambda \min(|\mathbf{A}(\mathbf{x}+\eta e_i)|)^2 \\
\Leftrightarrow\quad & \min_\eta p(\eta) \triangleq \tfrac{a}{2}\eta^2 + b\eta - \tfrac{1}{2}\min(\mathbf{g}\eta + \mathbf{d})^2 \\
\Leftrightarrow\quad & \min_\eta p(\eta) \triangleq \tfrac{a}{2}\eta^2 + b\eta - \tfrac{1}{2}\min_{i=1}^{m}[(\mathbf{g}_i\eta + \mathbf{d}_i)^2]
\end{aligned}
$$

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

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

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 $h(\cdot)$ does not bring much difficulty for solving the subproblem since it is separable.

Experimental Results

Experiments for Sparse Recovery and ICA Problem

We consider four publicly available real-world data sets: 'e2006tfidf', 'news20', 'sector', and 'TDT2' for the sensing/channel matrix $\mathbf{G} \in \mathbb{R}^{m \times n}$

The size of $\mathbf{G} \in \mathbb{R}^{m \times n}$ are are chosen from the following set $(m, n) \in \{(1000, 1024), (1000, 2048), (1024, 1000), (2048, 1000)\}$.

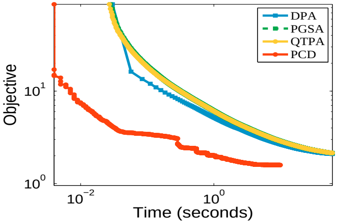
To generate the original k-sparse signal $\bar{\mathbf{x}}$ for the sparse recovery problem, we randomly select a support set S of size 100 and set $\bar{\mathbf{x}}_{\{1,\ldots,n\}\setminus S} = \mathbf{0}$, $\bar{\mathbf{x}}_S = \text{randn}(|S|, 1)$. We generate the observation vector via $\mathbf{y} = \mathbf{G}\bar{\mathbf{x}} + 0.1\|\mathbf{G}\bar{\mathbf{x}}\| \cdot \text{randn}(m, 1)$.

Sparse Recovery Problem

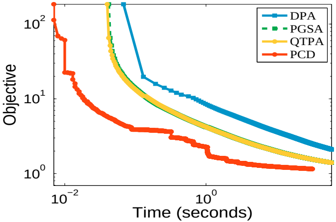
	DPA	PGSA	QTPA	PCD
e2006-1000-1024	1.874 ± 0.315	1.929 ± 0.278	1.923 ± 0.279	**1.530 ± 0.184**
e2006-1000-2048	1.640 ± 0.118	1.663 ± 0.172	1.660 ± 0.177	**1.312 ± 0.061**
e2006-1024-1000	2.610 ± 0.796	2.362 ± 0.533	2.362 ± 0.530	**1.882 ± 0.418**
e2006-2048-1000	5.623 ± 4.005	6.576 ± 4.966	6.593 ± 4.989	**3.068 ± 1.282**
news20-1000-1024	1.750 ± 0.247	1.403 ± 0.128	1.402 ± 0.130	**1.168 ± 0.023**
news20-1000-2048	2.043 ± 0.429	1.424 ± 0.181	1.426 ± 0.180	**1.207 ± 0.065**
news20-1024-1000	1.856 ± 0.353	1.488 ± 0.317	1.487 ± 0.318	**1.195 ± 0.045**
news20-2048-1000	4.997 ± 0.269	2.664 ± 0.604	2.559 ± 0.745	**1.394 ± 0.115**
sector-1000-1024	1.864 ± 0.162	1.337 ± 0.105	1.337 ± 0.104	**1.160 ± 0.016**
sector-1000-2048	1.780 ± 0.040	1.293 ± 0.033	1.293 ± 0.026	**1.148 ± 0.010**
sector-1024-1000	2.039 ± 0.016	1.485 ± 0.194	1.486 ± 0.195	**1.193 ± 0.015**
sector-2048-1000	5.041 ± 1.714	2.477 ± 1.048	2.475 ± 1.046	**1.409 ± 0.108**
TDT2-1000-1024	1.778 ± 0.303	1.646 ± 0.035	1.644 ± 0.032	**1.215 ± 0.047**
TDT2-1000-2048	1.710 ± 0.045	1.398 ± 0.029	1.398 ± 0.028	**1.127 ± 0.016**
TDT2-1024-1000	1.984 ± 0.284	1.555 ± 0.058	1.552 ± 0.050	**1.206 ± 0.067**
TDT2-2048-1000	4.696 ± 1.980	3.846 ± 0.901	3.789 ± 0.800	**1.338 ± 0.038**

Table: Comparisons of objective values for solving the spare recovery problem.

Sparse Recovery Problem



(a) e2006-1000-2048



(b) e2006-2048-1000

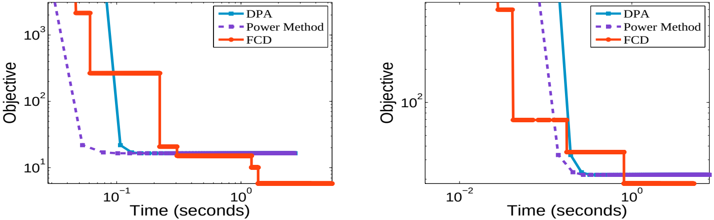
Figure: The convergence curve for solving the sparse recovery problem.

ICA Problem

	PGSA	Power Method	FCD
e2006-1000-1024	12.254 ± 14.922	12.254 ± 14.922	**6.686 ± 4.956**
e2006-1000-2048	16.896 ± 14.521	16.896 ± 14.521	**9.436 ± 6.359**
e2006-1024-1000	5.923 ± 4.485	5.923 ± 4.485	**4.948 ± 2.631**
e2006-2048-1000	16.846 ± 13.916	16.846 ± 13.916	**11.360 ± 8.225**
news20-1000-1024	112.805 ± 58.995	112.805 ± 58.995	**78.183 ± 22.830**
news20-1000-2048	125.440 ± 43.203	125.440 ± 43.203	**120.046 ± 41.353**
news20-1024-1000	99.211 ± 35.338	99.211 ± 35.338	**80.244 ± 22.771**
news20-2048-1000	138.909 ± 49.626	138.909 ± 49.626	**108.080 ± 37.811**
sector-1000-1024	60.813 ± 24.018	60.813 ± 24.018	**50.551 ± 18.675**
sector-1000-2048	139.459 ± 51.094	139.459 ± 51.094	**96.301 ± 42.115**
sector-1024-1000	83.176 ± 38.697	83.176 ± 38.697	**48.559 ± 19.163**
sector-2048-1000	104.654 ± 63.318	104.654 ± 63.318	**78.110 ± 28.532**
TDT2-1000-1024	27.167 ± 12.705	27.167 ± 12.705	**22.308 ± 8.171**
TDT2-1000-2048	27.480 ± 15.468	27.480 ± 15.468	**23.225 ± 12.614**
TDT2-1024-1000	32.334 ± 18.178	32.334 ± 18.178	**21.143 ± 12.143**
TDT2-2048-1000	44.659 ± 19.775	44.659 ± 19.775	**36.517 ± 12.689**

Table: Comparisons of objective values for solving the ICA problem.

ICA Problem



(a) e2006-1000-2048 (b) e2006-2048-1000

Figure: The convergence curve for solving the ICA problem.

Thank You!