# Optimality and Support Projection Algorithm for Sparsity Constrained Minimization

Lili Pan[†‡], Naihua Xiu[†], Shenglong Zhou[†]

† Department of Applied Mathematics, Beijing Jiaotong University
‡ School of Science, Shandong University of Technology

Sept 2014

# Outline

# Introduction

- In this talk, we mainly consider the nonlinear minimization with sparse and nonnegative constraints. By discussing tangent cone and normal cone of sparse constraint, we give the first necessary optimality conditions, $\alpha$-Stability, T-Stability and N-Stability, and the second necessary and sufficient optimality conditions for the nonlinear problem.
- By adopting Armijo-type stepsize rule, we present a gradient support projection algorithmic framework for the problem and establish its full convergence and computational complexity under mild conditions. By doing some numerical experiments, we show the excellent performance of the new algorithm for the least squares without and with noise.

# Introduction

**Model Representation**

- Sparsity and Nonnegativity Constrained Nonlinear Optimization

$$\min \quad f(x), \qquad \text{s.t.} \quad \|x\|_0 \le s, x \ge 0. \tag{1}$$

where $f(x) : \mathbb{R}^N \to \mathbb{R}$ is a continuously differentiable or twice differentiable function, $\|x\|_0$ is the $l_0$-norm of $x$.

- The special case of problem (1)

$$\min \|Ax - b\|^2 \quad s.t. \|x\|_0 \le s, x \ge 0, \tag{2}$$

where $A \in \mathbb{R}^{M\times N}$, $b \in \mathbb{R}^M$, $s < M < N$ and $\|\cdot\|$ is $l_2$-norm.

# Introduction

- We study the first and second order optimality conditions of the following model

$$\min \quad f(x), \qquad \text{s.t.} \quad \|x\|_0 \leq s. \tag{3}$$

Let $S \triangleq \{x \in \mathbb{R}^N | \ \|x\|_0 \leq s\}$.

- Support Projection

$$\mathrm{P}_S(x) = \left\{ \ y \in \mathbb{R}^N | y_i = x_i, i \in I_s(x); \ \ y_i = 0, i \notin I_s(x) \ \right\}.$$

where $I_s(x) := \{j_1, j_2, \cdots, j_s\} \subseteq \{1, 2, \cdots, N\}$ of indices of $x$ with

$$\min_{i \in I_s(x)} |x_i| \geq \max_{i \notin I_s(x)} |x_i|.$$

# Optimality Conditions (I)

**Definition of Bouligand Tangent Cone**

For any nonempty set $\Omega \subseteq \mathbb{R}^N$, its *Bouligand Tangent Cone* $T_\Omega^B(\overline{x})$, and corresponding *Normal Cone* $N_\Omega^B(\overline{x})$ at point $\overline{x} \in \Omega$ are defined as:

$$T_\Omega^B(\overline{x}) := \left\{ d \in \mathbb{R}^N \;\middle|\; \begin{array}{l} \exists \; \{x^k\} \subset \Omega, \lim\limits_{k\to\infty} x^k = \overline{x}, \; \lambda_k \geq 0, k = 1, \\ 2, \cdots, \text{such that } \lim\limits_{k\to\infty} \lambda_k (x^k - \overline{x}) = d \end{array} \right\},$$

$$N_\Omega^B(\overline{x}) := \left\{ d \in \mathbb{R}^N \mid \langle d, z \rangle \leq 0, \; \forall \; z \in T_\Omega^B(\overline{x}) \right\},$$

# Optimality Conditions (I)

## Definition of Clarke Tangent Cone

The *Clarke Tangent Cone* $T^C_\Omega(\overline{x})$ and corresponding *Normal Cone* $N^C_\Omega(\overline{x})$ at point $\overline{x} \in \Omega$ are defined as:

$$T^C_\Omega(\overline{x}) := \left\{ d \in \mathbb{R}^N \;\middle|\; \begin{array}{r} \forall\ \{x^k\} \subset \Omega,\ \forall\ \{\lambda_k\} \subset \mathbb{R}_+ \text{ with } \lim\limits_{k\to\infty} x^k = \overline{x}, \\ \lim_{k\to\infty} \lambda_k = 0, \exists\ \{y^k\} \text{ such that } \lim\limits_{k\to\infty} y^k = d \\ \text{and } x^k + \lambda_k y^k \in \Omega,\ \ k \in \mathbb{N} \end{array} \right\},$$

$$N^C_\Omega(\overline{x}) := \left\{\ d \in \mathbb{R}^N \mid \langle d, z\rangle \le 0,\ \forall\ z \in T^C_\Omega(\overline{x})\ \right\}.$$

# Optimality Conditions (I)

**Bouligand Tangent Cone of Sparse Set**

### Theorem

*For any* $\overline{x} \in S$ *and letting* $\Gamma = \mathrm{supp}(\overline{x})$, *the Bouligand tangent cone and corresponding normal cone of* $S$ *at* $\overline{x}$ *are*

$$T_S^B(\overline{x}) = \bigcup_{\Upsilon} \mathrm{span}\{\, \mathrm{e}_i, \quad i \in \Upsilon \supseteq \Gamma, |\Upsilon| \leq s \,\} \tag{4}$$

$$N_S^B(\overline{x}) = \begin{cases} \mathrm{span}\{\, \mathrm{e}_i, \quad i \notin \Gamma \,\}, & \text{if } |\Gamma| = s \\ \{0\}, & \text{if } |\Gamma| < s \end{cases} \tag{5}$$

*where* $\mathrm{e}_i \in \mathbb{R}^N$ *is a vector whose the ith component is one and others are zeros,* $\mathrm{span}\{\mathrm{e}_i, i \in \Gamma\}$ *denotes the subspace of* $\mathbb{R}^N$ *spanned by* $\{\, \mathrm{e}_i, \ i \in \Gamma\}$, *and* $\mathrm{supp}(x) = \{i \in \{1, \cdots, N\} \mid x_i \neq 0\}$.

# Optimality Conditions (I)

**Clarke Tangent Cone of Sparse Set**

## Theorem

*For any* $\overline{x} \in S$ *and letting* $\Gamma = \text{supp}(\overline{x})$, *then the Clarke tangent cone and corresponding normal cone of* $S$ *at* $\overline{x}$ *are*

$$T_S^C(\overline{x}) = \{ d \in \mathbb{R}^N \mid \text{supp}(d) \subseteq \Gamma \} = \text{span}\{ \mathrm{e}_i, \ \ i \in \Gamma \} \tag{6}$$

$$N_S^C(\overline{x}) = \text{span}\{ \mathrm{e}_i, \ \ i \notin \Gamma \}. \tag{7}$$

# Optimality Conditions (I)

**$\alpha$-Stability, $N$-Stability and $T$-Stability**

## Definition

For real number $\alpha > 0$, a vector $x^* \in S$ is called an $\alpha$-stationary point, $N^\sharp$-stationary point and $T^\sharp$-stationary point of (3) if it respectively satisfies the relation

$$\begin{aligned} \alpha - \text{stationary point:} \quad & x^* && \in && P_S\left(x^* - \alpha\nabla f(x^*)\right), && (8)\\ N^\sharp - \text{stationary point:} \quad & 0 && \in && \nabla f(x^*) + N_S^\sharp(x^*), && (9)\\ T^\sharp - \text{stationary point:} \quad & 0 && = && \|\nabla_S^\sharp f(x^*)\|, && (10) \end{aligned}$$

where $\nabla_S^\sharp f(x^*) = \arg\min\{\ \|x + \nabla f(x^*)\| \mid x \in T_S^\sharp(x^*)\ \}$, $\sharp \in \{B, C\}$ stands for the sense of Bouligand tangent cone or Clarke tangent cone.

# Optimality Conditions (I)

**Relationship of the Three Kinds of Stability**

Theorem

*Under the concept of Bouligand tangent cone, for model* (3) *and* $\alpha > 0$, *if the vector* $x^* \in S$ *satisfies* $\|x^*\|_0 = s$, *then*

$$\alpha\text{−stationary point} \Longrightarrow N^B\text{−stationary point} \Longleftrightarrow T^B\text{−stationary point};$$

*if the vector* $x^* \in S$ *satisfies* $\|x^*\|_0 < s$, *then*

$$\alpha\text{−stationary point} \Longleftrightarrow N^B\text{−stationary point} \Longleftrightarrow T^B\text{−stationary point}$$

$$\Longleftrightarrow \nabla f(x^*) = 0.$$

# Optimality Conditions (I)

## Relationship of the Three Kinds of Stability

| | $\|x^*\|_0 = s$ | | $\|x^*\|_0 < s$ |
|---|---|---|---|
| $\alpha$ – stationary point<br>$x^* \in P_S(x^* - \alpha \nabla f(x^*))$ | $\|(\nabla f(x^*))_i\|$ | $\begin{cases} = 0, & i \in \Gamma \\ \leq \frac{1}{\alpha} M_s(\|x^*\|), & i \notin \Gamma, \end{cases}$ | $\nabla f(x^*) = 0$ |
| $N^B$ – stationary point<br>$-\nabla f(x^*) \in N_S^B(x^*)$ | $(\nabla f(x^*))_i$ | $\begin{cases} = 0, & i \in \Gamma \\ \in \mathbb{R}, & i \notin \Gamma, \end{cases}$ | $\nabla f(x^*) = 0$ |
| $T^B$ – stationary point<br>$\nabla_S^B f(x^*) = 0$ | $(\nabla f(x^*))_i$ | $\begin{cases} = 0, & i \in \Gamma \\ \in \mathbb{R}, & i \notin \Gamma, \end{cases}$ | $\nabla f(x^*) = 0$ |

# Optimality Conditions (I)

**Relationship of the Three Kinds of Stability**

## Theorem

*Under the concept of Clarke tangent cone, we consider the problem* (3). *For* $\alpha > 0$, *if* $x^* \in S$ *then*

$$\alpha\text{-stationary point} \Longrightarrow N^C\text{-stationary point} \Longleftrightarrow T^C\text{-stationary point.}$$

# Optimality Conditions (I)

**Theorem**

*Let function $f(x)$ satisfy Assumption 1, we have if $x^* \in S$ is the optimal solution of (3), then for $0 < \alpha < \frac{1}{L_f}$, $x^*$ is also the $\alpha$-stationary point. On the contrary, let's further assume that $f(x)$ is convex, if $\|x^*\|_0 < s$ and $x^*$ is the $\alpha$-stationary point of (3), then $x^*$ is the optimal solution of (3).*

# Optimality Conditions (I)

**Theorem (Second Order Necessary Optimality)**

*If $x^* \in S$ is the optimal solution of (3) , then for $0 < \alpha < \frac{1}{L_f}$ we have*

$$d^\top \nabla^2 f(x^*) d \geq 0, \quad \forall\ d \in T_S^C(x^*). \tag{11}$$

*where $\nabla^2 f(x^*)$ is the Hessian matrix of $f$ at $x^*$.*

# Optimality Conditions (I)

**Theorem (Second Order Sufficient Optimality)**

*If $x^* \in S$ is an $\alpha$-stationary point of (3) and satisfies*

$$d^\top \nabla^2 f(x^*) d > 0, \quad \forall\, d \in T_S^C(x^*), \tag{12}$$

*then $x^*$ is the strictly locally optimal solution of (3). Moreover, there are $\eta > 0$ and $\delta > 0$, for any $x \in B(x^*, \delta) \cap S$, it holds*

$$f(x) \geq f(x^*) + \eta \|x - x^*\|^2. \tag{13}$$

# Optimality Conditions (II)

**Support projection and Tangent cones for (1)**

- $\mathrm{P}_{S\cap\mathbb{R}^N_+}(x) = \mathrm{P}_S \cdot \mathrm{P}_{\mathbb{R}^N_+}(x)$.

**Theorem**

*For $\overline{x} \in S \cap \mathbb{R}^N_+$, by denoting $\mathbb{R}^N_+(\overline{x}) := \{\, x \in \mathbb{R}^N \mid x_i \geq 0, i \notin \Gamma \,\}$, it has*

$$T^B_{S\cap\mathbb{R}^N_+}(\overline{x}) = T^B_S(\overline{x}) \cap \mathbb{R}^N_+(\overline{x}), \qquad N^B_{S\cap\mathbb{R}^N_+}(\overline{x}) = T^B_S(\overline{x}) \cap (-\mathbb{R}^N_+(\overline{x}))$$

$$T^C_{S\cap\mathbb{R}^N_+}(\overline{x}) = T^C_S(\overline{x}), \qquad N^C_{S\cap\mathbb{R}^N_+}(\overline{x}) = N^C_S(\overline{x}).$$

# Optimality Conditions (II)

- **$\alpha$-stationary point of (1) is defined as:**

$$x^* \in P_{S\cap\mathbb{R}^N_+}\left(x^* - \alpha\nabla f(x^*)\right). \tag{14}$$

**Theorem**

*For any $\alpha > 0$, $x^* \in S \cap \mathbb{R}^N_+$ is $\alpha$-stationary point of (1) if and only if*

$$\nabla_i f(x^*) \begin{cases} = 0, & \text{if } \ i \in \mathrm{supp}(x^*), \\ \in [-\frac{1}{\alpha} M_s(x^*), +\infty), & \text{if } \ i \notin \mathrm{supp}(x^*), \end{cases} \tag{15}$$

# Optimality Conditions (II)

**Relationship of the Three Kinds of Stability for model** $(1)$

### Theorem

*For the model* $(1)$ *and any* $\alpha > 0$.

***A****) Under the concept of Bouligand tangent cone, if* $\|x^*\|_0 = s, x^* \geq 0$*, then*

$$\alpha-\text{stationary point} \Longrightarrow N^B-\text{stationary point} \Longleftrightarrow T^B-\text{stationary point.}$$

***B****) Under the concept of Clarke tangent cone, if* $\|x^*\|_0 \leq s, x^* \geq 0$*, then*

$$\alpha-\text{stationary point} \Longrightarrow N^C-\text{stationary point} \Longleftrightarrow T^C-\text{stationary point.}$$

# Optimality Conditions (II)

- Assumption 1. The gradient of the objective function $f(x)$ is Lipschitz with constant $L_f$ over $\mathbb{R}^N$:

$$\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\|, \quad \forall\ x, y \in \mathbb{R}^N. \tag{16}$$

# $\alpha$-stationary point of $(1)$

## Theorem (Second Order Optimality for model (1))

*If* $x^* \in S \cap \mathbb{R}^N_+$ *is the optimal solution of* $(1)$, *then for* $0 < \alpha < \frac{1}{L_f}$, $x^*$ *is also the* $\alpha$*-stationary point of* $(1)$, *and moreover,*

$$d^\top \nabla^2 f(x^*) d \geq 0, \quad \forall\, d \in T^C_S(x^*). \tag{17}$$

*On the contrary, if* $x^* \in S \cap \mathbb{R}^N_+$ *is an* $\alpha$*-stationary point of* $(1)$ *and*

$$d^\top \nabla^2 f(x^*) d > 0, \quad \forall\, d \in T^C_S(x^*), \tag{18}$$

*then* $x^*$ *is the strictly locally optimal solution of* $(1)$. *Moreover, there is a* $\gamma > 0$ *and* $\delta > 0$, *when any* $x \in B(x^*, \delta) \cap S \cap \mathbb{R}^N_+$, *it holds*

$$f(x) \geq f(x^*) + \gamma \|x - x^*\|^2. \tag{19}$$

# Gradient Support Projection Algorithm for (1)

**Step 0** Initialize $x^0 = 0$, $\Gamma^0 = \mathrm{supp}(\mathrm{P}_{S\cap\mathbb{R}^N_+}(\nabla f(x^0)))$, $0 < \alpha_0 < \frac{1}{L_f}$, $0 < \sigma \le \frac{1}{4L_f}$, $0 < \beta < 1, \epsilon > 0$. Set $k \Leftarrow 0$;

**Step 1** Compute $\tilde{x}^{k+1} = \mathrm{P}_{S\cap\mathbb{R}^N_+}\left(x^k - \alpha_0 \nabla f(x^k)\right)$;

**Step 2** If $\mathrm{supp}(\tilde{x}^{k+1}) = \Gamma^k$, then $x^{k+1} = \tilde{x}^{k+1}, \Gamma^{k+1} = \mathrm{supp}(x^{k+1})$;
Else $x^{k+1} = \mathrm{P}_{S\cap\mathbb{R}^N_+}\left(x^k - \alpha_k \nabla f(x^k)\right)$, $\Gamma^{k+1} = \mathrm{supp}(x^{k+1})$,
where $\alpha_k = \alpha_0 \beta^{m_k}$ and $m_k$ is the smallest positive integer $m$ such that

$$f(x^k(\alpha_0\beta^m)) \le f(x^k) - \frac{\sigma}{2}\frac{\|x^k(\alpha_0\beta^m) - x^k\|^2}{(\alpha_0\beta^m)^2},$$

here $x^k(\alpha) = \mathrm{P}_{S\cap\mathbb{R}^N_+}(x^k - \alpha\nabla f(x^k))$;

**Step 3** If $\|x^{k+1} - x^k\| \le \epsilon$, stop; Otherwise $k \Leftarrow k+1$, go to **Step 1**.

# Gradient Support Projection Algorithm for (1)

- **Lemma**
  *Let Assumption 1. hold and $\{x^k\}$ be the iterative point in Step 2 in GSPA. Then*

$$f(x^k(\alpha)) \leq \begin{cases} f(x^k) - \frac{1}{2}(\frac{1}{\alpha} - L_f)\|x^k(\alpha) - x^k\|^2, \alpha \in \left(0, \frac{1}{L_f}\right) \\ f(x^k) - \frac{\sigma}{2}\frac{\|x^k(\alpha)-x^k\|^2}{\alpha^2}, \alpha \in \left[\frac{1-\sqrt{1-4\sigma L_f}}{2L_f}, \frac{1+\sqrt{1-4\sigma L_f}}{2L_f}\right]. \end{cases}$$

# Gradient Support Projection Algorithm for (1)

**Theorem**

*Let Assumption 1 hold and the sequence $\{x^k\}$ be generated by GSPA, we have*

$(i)$ $\lim\limits_{k\to\infty} \frac{\|x^{k+1}-x^k\|}{\alpha_k} = 0$;

$(ii)$ *any accumulation point of* $\{x^k\}$ *is the* $\alpha$*-stationary point of* (3);

$(iii)$ $\lim_{k\to\infty} \|\nabla^C_{S\cap\mathbb{R}^N_+} f(x^k)\| = 0$.

# Gradient Supp-Projection Algorithm for (2)

Let $r(x) = \frac{1}{2}\|Ax - b\|^2$, we consider the problem (2).

**Step 0** Initialize $x^0 = 0$, $\Gamma^0 = \mathrm{supp}(\mathrm{P}_{S\cap\mathbb{R}^N_+}(A^T b))$, $0 < \sigma \le \frac{1}{4L_r}$, $0 < \beta < 1, \epsilon > 0$. Set $k \Leftarrow 0$;

**Step 1** Compute $\tilde{x}^{k+1} = \mathrm{P}_{S\cap\mathbb{R}^N_+}\left(x^k - \alpha_0^k \nabla r(x^k)\right)$;

$$\alpha_0^k = \frac{\|A^T_{\Gamma^k}(b - Ax^k)\|^2}{\|A_{\Gamma^k}A^T_{\Gamma^k}(b - Ax^k)\|^2}.$$

**Step 2** If $\mathrm{supp}(\tilde{x}^{k+1}) = \Gamma^k$, then $x^{k+1} = \tilde{x}^{k+1}, \Gamma^{k+1} = \mathrm{supp}(x^{k+1})$;
Else $x^{k+1} = \mathrm{P}_{S\cap\mathbb{R}^N_+}\left(x^k - \alpha_k \nabla r(x^k)\right)$, $\Gamma^{k+1} = \mathrm{supp}(x^{k+1})$,

where $\alpha_k = \alpha_0^k \beta^{m_k}$ and $m_k$ is the smallest positive integer $m$ such that

$$r(x^k(\alpha_0^k\beta^m)) \le r(x^k) - \frac{\sigma}{2}\frac{\|x^k(\alpha_0^k\beta^m) - x^k\|^2}{(\alpha_0^k\beta^m)^2},$$

here $x^k(\alpha) = \mathrm{P}_{S\cap\mathbb{R}^N_+}(x^k - \alpha\nabla r(x^k))$;

**Step 3** If $\|x^{k+1} - x^k\| \le \epsilon$, stop; Otherwise $k \Leftarrow k + 1$, go to Step 1.

# Gradient Supp-Projection Algorithm for (2)

- Assumption 2. Matrix $A$ is $s$-regular if any $s$ of its columns are linearly independent, namely,

$$d^\top A^\top A d > 0, \quad \forall \; \|d\|_0 \leq s.$$

# Gradient Supp-Projection Algorithm for (2)

- Theorem
  *Let the sequence $\{x^k\}$ be generated by GSPA, then $\{x^k\}$ converges to a local minimizer of (2) if $A$ is s-regular.*

# Gradient Supp-Projection Algorithm for (2)

**Theorem**

*If Assumption 2 holds for matrix $A$, then the local solutions of problem (2) exist and are finite. Moreover, if $A$ and $b$ satisfies*

$$\|\Pi_{\Gamma_i} b\| \neq \|\Pi_{\Gamma_j} b\| \quad \text{with } \Gamma_i \neq \Gamma_j,\ |\Gamma_i| \leq s, |\Gamma_j| \leq s \tag{20}$$

*where $\|\Pi_{\Gamma_i} b\| = b^T A_{\Gamma_i}(A_{\Gamma_i}^T A_{\Gamma_i})^{-1} A_{\Gamma_i}^T b$. then problem (2) has a unique solution.*

# Numerical Experiments

- **Greedy methods**
  - MP — Matching pursuit[MZ]
  - OMP — Orthogonal MP[DM]
  - CoSaMP — Compressive sampling matching pursuit [NT]
  - SP — Subspace pursuit[DM]
  - NIHT — Iterative hard thresholding algorithm [B]
  - · · ·

[MZ] S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Trans. Signal Process., 41, pp. 3397-3415, 1993.

[NT] D. Needell and J.A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal., 26, pp.301-32,2009.

[DM] W. Dai and O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inform. Theory, 55, pp.2230-2249, 2009.

[B] T Blumensath, Normalized iterative hard thresholding: Guaranteed stability and performance , Selected Topics in Signal Processing, IEEE Journal of, vol. 4. no. 2, pp. 298-309, 2010..

# Numerical Experiments

## Exact recovery

GSPA and NIHT for (2) with sparsity and nonnegativity

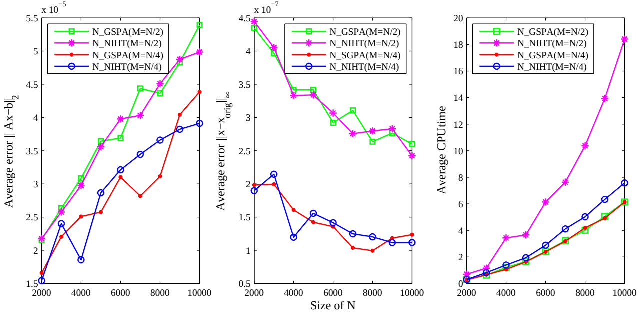
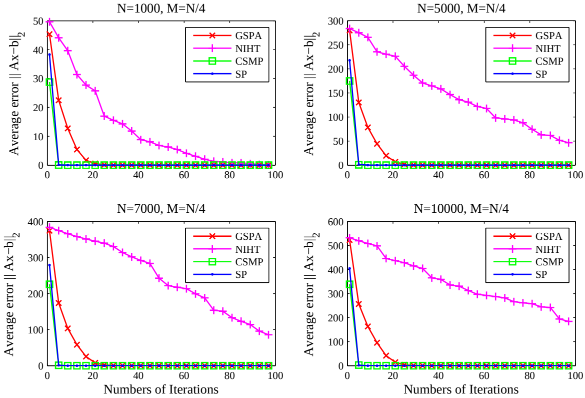



Figure: Average results yielded by *Non_NIHT* and *Non_GSPA*.

# Numerical Experiments

## Exact recovery

*GSPA*, *NIHT*, *CoSaMP*(short for CSMP) and *SP* for (2) with sparsity



# Numerical Experiments

**Exact recovery:** *GSPA*, *NIHT*, *CoSaMP* and *SP* for (2) with sparsity

Table: The average CPU time over 40 simulations with $k = 5\%N$.

| $N$ | $M$ | $GSPA$ | $NIHT$ | $CSMP$ | $SP$ |
|---|---|---|---|---|---|
| $N = 1000$ | $M = N/4$ | 0.0689 | 0.2583 | 0.1492 | 0.0961 |
| | $M = N/2$ | 0.0677 | 0.2459 | 0.1687 | 0.1307 |
| $N = 3000$ | $M = N/4$ | 0.5385 | 3.3210 | 1.9171 | 1.1197 |
| | $M = N/2$ | 0.5756 | 2.6228 | 1.8754 | 1.3627 |
| $N = 5000$ | $M = N/4$ | 1.5583 | 11.246 | 8.0507 | 4.5900 |
| | $M = N/2$ | 1.5114 | 8.0690 | 7.7457 | 5.0981 |
| $N = 7000$ | $M = N/4$ | 3.0050 | 20.761 | 19.698 | 10.729 |
| | $M = N/2$ | 2.9543 | 16.389 | 19.336 | 12.613 |
| $N = 10000$ | $M = N/4$ | 6.3880 | 52.257 | 51.680 | 27.864 |
| | $M = N/2$ | 5.9462 | 38.256 | 53.707 | 30.924 |

# Numerical Experiments

## Recovery with Noise

*GSPA* and *NIHT* for (2) with sparsity

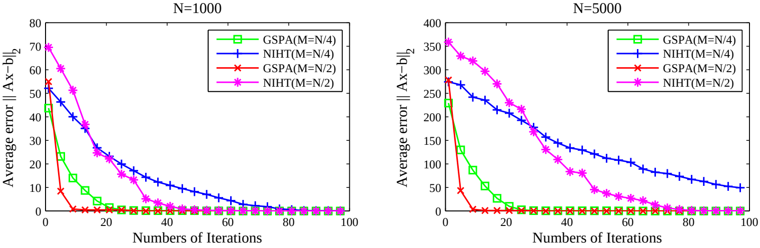



Figure: Average error $\|Ax - b\|_2$ for each iteration with $k = 5\%N$ over 40 simulations with noise.

# Numerical Experiments

## Recovery with Noise

*GSPA* and *CoSaMP* for (2) with sparsity

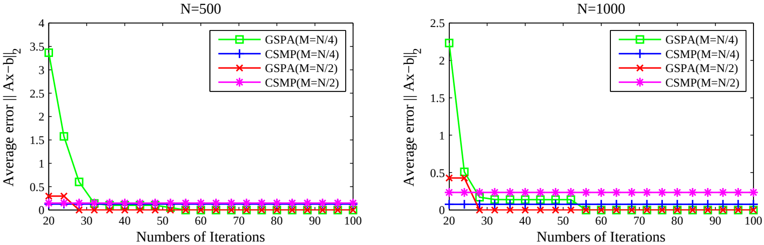



Figure: Average error $\|Ax - b\|_2$ for each iteration with $k = 5\%N$ over 40 simulations with noise.

# Numerical Experiments

## Recovery with Noise

*GSPA* and *SP* for (2) with sparsity

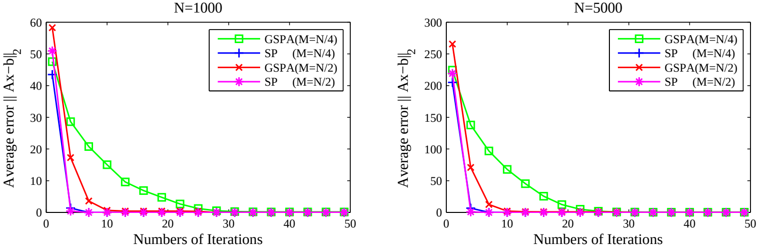



Figure: Average error $\|Ax - b\|_2$ for each iteration with $k = 5\%N$ over 40 simulations with noise.

# Numerical Experiments

## Recovery with Noise

*GSPA*, *NIHT*, *CoSaMP* and *SP* for (2) with sparsity

Table: The average CPU time over 40 simulations with $M = N/4, s = 5\%N$ and noise.

| | $N$ | *GSPA* | *NIHT* | *CSMP* | *SP* |
|---|---|---|---|---|---|
| | 1000 | 0.0812 | 0.3226 | 116.87 | 0.1859 |
| | 3000 | 0.5797 | 3.9317 | 1416.1 | 1.1631 |
| CPU time | 5000 | 1.6221 | 9.6857 | – – | 4.9076 |
| | 7000 | 3.2252 | 25.306 | – – | 11.556 |
| | 10000 | 6.6369 | 38.440 | – – | 28.429 |

# Summary

- **Contributions** We have established the first and second order optimality conditions for problem (1) and (3), proposed a gradient support projection algorithm for (3), and shown that the new algorithm has elegant convergence and exceptional performance.
- **Future Work** In the future, we will further consider conjugate gradient or quasi-Newton direction in stead of negative gradient direction to improve convergence speed. On the other hand, we will think to develop this algorithm for optimization problems with sparsity and other complex constraints.

# *Thank you !*