SEMI-SMOOTH SECOND-ORDER METHODS FOR COMPOSITE CONVEX PROGRAMS

Zaiwen Wen

Beijing International Center For Mathematical Research
Peking University

Joint work: Xiantao Xiao, Liwei Zhang (Dalian University of Technology)
Yongfeng Li (PKU)
wenzw@pku.edu.cn

Introduction

- composite convex programs
- operator splitting and fixed-point iteration
- semi-smoothness
- semi-smooth Newton method
- semi-smooth LM method
- numerical result
- conclusion and further research

Optimization problem

Composite convex problem

$$\min_{x \in \mathbb{R}^n} \quad f(x) + h(x),$$

where f and h are real-valued convex functions.

Applications:

- sparse optimization problem in signal or image processing.
- regularized empirical risk minimization problem in machine learning.
- general convex constrained optimization problem.

Operator splitting and fixed-point algorithm

Examples:

- forward-backward splitting(FBS).
- Douglas-Rachford splitting(DRS).
- Peaceman-Rachford splitting(PRS).
- alternating direction method of multipliers(ADMM).

Advantages:

- easy to implement;
- converge fast to a solution with moderate accuracy.

Disadvantages:

- slow tail convergence.

Forward-backward splitting

- Let h be a continuously differentiable function and

$$||\nabla h(x) - \nabla h(y)|| \leq L||x - y||.$$

- The FBS algorithm is known as proximal gradient algorithm

$$x^{k+1} = \text{prox}_{tf}(x^k - t\nabla h(x^k)), k = 0, 1, \cdots .$$

- gradient method when $f = 0$.
- proximal point algorithm when $h = 0$.
- projected gradient method when $f = 1_C(x)$.
- fixed point iteration

$$x^{k+1} = T_{\text{FBS}}(x^k).$$

where

$$T_{\text{FBS}} := \text{prox}_{tf} \circ (I - t\nabla h).$$

Douglas-Rachford splitting

- DRS algorithm

$$
\begin{aligned}
x^{k+1} &= \mathrm{prox}_{th}(z^k), \\
y^{k+1} &= \mathrm{prox}_{tf}(2x^{k+1} - z^k), \\
z^{k+1} &= z^k + y^{k+1} - x^{k+1}.
\end{aligned}
$$

- fixed point iteration

$$
z^{k+1} = T_{\mathrm{DRS}}(z^k),
$$

where

$$
T_{\mathrm{DRS}} := I + \mathrm{prox}_{tf} \circ (2\mathrm{prox}_{th} - I) - \mathrm{prox}_{th}.
$$

Douglas-Rachford splitting

- PRS algorithm

$$z^{k+1} = T_{\mathrm{PRS}}(z^k),$$

where

$$T_{\mathrm{PRS}} := (2\mathrm{prox}_{tf} - I) \circ (2\mathrm{prox}_{th} - I).$$

- relaxed PRS algorithm

$$z^{k+1} = T_{\lambda_k}(z^k)$$

where

$$T_{\lambda_k} := (1 - \lambda_k)I + \lambda_k T_{\mathrm{PRS}}, \quad \lambda_k \in (0, 1].$$

- $\lambda_k = 1/2$ DRS algorithm; $\lambda_k = 1$: PRS algorithm.

Dual Splitting

- For a linear constrained program

$$\begin{aligned} \min_{x_1 \in \mathbb{R}^{n_1}, x_2 \in \mathbb{R}^{n_2}} \quad & f_1(x_1) + f_2(x_2) \\ \text{s.t.} \quad & A_1 x_1 + A_2 x_2 = b, \end{aligned}$$

- the dual problem

$$\min_{w \in \mathbb{R}^m} \quad d_1(w) + d_2(w),$$

where

$$d_1(w) := f_1^*(A_1^T w), \quad d_2(w) := f_2^*(A_2^T w) - b^T w.$$

- The dual FBS algorithm is equal to the alternating minimization (AM) algorithm.
- The dual DRS algorithm is equal to the alternating direction method of multipliers (ADMM) algorithm。

Convergence of the fixed-point algorithms

- Require:
  - f: closed proper and convex;
  - h: convex and continuously differentiable;
  - z^*: a fixed-point of T_{FBS}.

  Consenquences:
  - z^* is an optimal solution.
- Require:
  - f and h: closed proper and convex.
  - z^*: a fixed-point of T_{DRS} or T_{PRS}.

  Consenquences:
  - $\mathrm{prox}_{th}(z^*)$ is an optimal solution .

Sublinear convergence of FBS

Require:

- f: closed proper and convex;
- h: convex and continuously differentiable and $||\nabla h(x) - \nabla h(y)|| \leq L||x - y||$.
- x^*: an optimal solution to problem;
- $x^0 \in \mathbf{dom}(f) \cap \mathbf{dom}(h)$
- $\{x^k\}$: generated by $x^{k+1} = T_{\mathrm{FBS}}(x^k)$ with tuning parameter $t = 1/L$

Consenquences:

- for all $k \geq 1$,

$$f(x^k) + h(x^k) - f(x^*) - h(x^*) \leq \frac{L\|x^0 - x^*\|^2}{2k}.$$

- for all $k \geq 1$, we have $\|T_{\mathrm{FBS}}(x^k) - x^k\|^2 = o(1/k^2)$ and

$$\|T_{\mathrm{FBS}}(x^k) - x^k\|^2 \leq \frac{\|x^0 - x^*\|^2}{k^2}.$$

Error bound condition

- optimal solution set X^*
- test set T
- residual function $F(x)$: F(x) = 0 when x is optimal solution
- (error bound condition) if there exists a constant $\kappa > 0$ such that

$$\text{dist}(x, X^*) \leq \kappa ||F(x)|| \quad \text{for all } x \in T.$$

- (error bound with residual-based test set(EBR))
  $T := \{x \in \mathbb{R}^n | f(x) + h(x) \leq v, \|F(x)\| \leq \varepsilon\}$ for some constant $\varepsilon \geq 0$ and any $v \geq v^* := \min_x f(x) + h(x)$.

Linear convergence of FBS

Require:

- error bound condition (EBR) holds with parameter κ for residual function $F_{\mathrm{FBS}} = T_{\mathrm{FBS}} - I$
- $\{x^k\}$: generated by the fixed-point iteration $x^{k+1} = T_{\mathrm{FBS}}(x^k)$ with $t \leq \beta^{-1}$ for some constant $\beta > 0$
- x^*: the limit point of the sequence $\{x^k\}$

Consenquences:

- there exists an index r such that for all $k \geq 1$,

$$\|x^{r+k} - x^*\|^2 \leq \left(1 - \frac{1}{2\beta\kappa}\right)^k C \cdot (f(x^k) + h(x^k) - f(x^*) - h(x^*)),$$

where $C := \frac{2}{\beta(1-\sqrt{1-(2\beta\gamma)^{-1}})^2}$.

Sublinear convergence of relaxed PRS

Require:

- f and h: closed proper and convex
- $\{\lambda_k \in (0,1)\}$: a sequence of positive numbers
- $\tau_k := \lambda_k(1-\lambda_k)$
- z^*: a fixed point of T_{λ_k}
- $\{z^k\}$: generated by the fixed-point iteration $z^{k+1} = T_{\lambda_k}(z^k)$

Consenquences:

- if $\tau_k > 0$ for all $k \geq 0$,

$$\|T_{\lambda_k}(z^k) - z^k\|^2 \leq \frac{\|z^0 - z^*\|^2}{\sum_{i=0}^{k} \tau_i}.$$

- In particular, if $\tau_k \in (\varepsilon, \infty)$ for all $k \geq 0$ and some $\varepsilon > 0$, then $\|T_{\lambda_k}(z^k) - z^k\| = o(1/(k+1))$.

Semi-smooth Newton-type method

- Solving the system
$$F(z) = 0,$$
where $F(z) = T(z) - z$ and $T(z)$ is a fixed-point mapping.
- Fixed-point algorithms suffer from slow tail convergence and is not suitable for high accuracy applications.
- $F(z)$ fails to be differentiable in many interesting applications.
- F(z): (strongly) semi-smooth and monotone.
- semi-smooth Newton's type method
  - Semi-smooth Newton method
  - Semi-smooth Levenberg-Marquardt method

Semi-smoothness

- $F : \mathcal{O} \to \mathbb{R}^m$ be locally Lipschitz continuous.
- The B-subdifferential of F at x is defined by

$$\partial_B F(x) := \left\{ \lim_{k\to\infty} F'(x^k) | x^k \in D_F, x^k \to x \right\}.$$

  The set

$$\partial F(x) = \mathrm{co}(\partial_B F(x))$$

  is called Clarke's generalized Jacobian
- We say that F is semismooth at $x \in \mathcal{O}$ if
  - F is directionally differentiable at x;
  - for any $d \in \mathcal{O}$ and $J \in \partial F(x+d)$,

$$\|F(x+d) - F(x) - J(d)\| = o(\|d\|) \quad \text{as } d \to 0.$$

- F is said to be strongly semi-smooth at $x \in \mathcal{O}$ if F is semi-smooth and for any $d \in \mathcal{O}$ and $J \in \partial F(x+d)$,

$$\|F(x+d) - F(x) - J(d)\| = O(\|d\|^2) \quad \text{as } d \to 0.$$

Semi-smoothness

- (Strongly) semi-smoothness is closed under scalar multiplication, summation and composition.
- A vector-valued function is (strongly) semi-smooth if and only if each of its component functions is (strongly) semi-smooth.
- Examples:
  - semi-smooth
    - the smooth functions
    - all convex functions (thus norm)
    - the piecewise differentiable functions
  - strongly semi-smooth
    - Differentiable functions with Lipschitz gradients
    - For every $p \in [1, \infty]$, the norm $\|\cdot\|_p$
    - Piecewise affine functions

Semi-smoothness of proximal mappings

- The semi-smoothness of proximal mapping does not hold generally.
- Examples:
  - Sparcity inducing norm $||\cdot||_1$.
  - Piecewise $\mathcal{C}^k$ functions.
  - Metric projections over a polyhedral set or symmetric cones.
  - Piecewise linear-quadratic functions.
  - Convex functions.

Semi-smooth Newton method

- $J_k \in \partial_B F(z^k)$: positively semidefinite.
- regularized Newton's method

$$(J_k + \mu_k I)d = -F_k,$$

where $F_k = F(z^k)$, $\mu_k = \lambda_k \|F_k\|$ and $\lambda_k > 0$ is a regularization parameter.

- solve the linear system inexactly.

$$r_k := (J_k + \mu_k I)d^k + F_k.$$

- seek to step d^k by solving the system approximately such that

$$\|r_k\| \leq \tau \min\{1, \lambda_k \|F_k\| \cdot \|d^k\|\},$$

where $0 < \tau < 1$ is some positive constant.

Semi-smooth Newton method

- $d^k = 0$, then x_k is the optimal solution.
- A trial point
$$u^k = z^k + d^k.$$
- d_k is small enough,
$$\left\langle F(u^k), z^k - u^k \right\rangle = -\left\langle F(u^k), d^k \right\rangle > 0.$$
- By monotonicity of F, for any optimal solution z^*
$$\left\langle F(u^k), z^* - u^k \right\rangle \leq 0.$$
- Therefore the hyperplane
$$H_k := \{z \in \mathbb{R}^n | \left\langle F(u^k), z - u^k \right\rangle = 0\}$$
strictly separates z^k from the solution set Z^*.

Semi-smooth Newton method

- Define a ratio

$$\rho_k = \frac{-\left\langle F(u^k), d^k \right\rangle}{\|d^k\|^2}.$$

- If ρ_k is big enough,

$$z^{k+1} = z^k - \frac{\left\langle F(u^k), z^k - u^k \right\rangle}{\|F(u^k)\|^2} F(u^k),$$

  which is the projection onto the hyperplane H_k.
- If ρ_k is too small, $z^{k+1} = z^k$ and increase the parameter.

Semi-smooth Newton method

- Select some parameters $0 < \eta_1 \leq \eta_2 < 1$ and $1 < \gamma_1 \leq \gamma_2$. $\underline{\lambda} > 0$ is a small positive constant.
- Update the point

$$z^{k+1} = \begin{cases} z^k - \frac{\langle F(u^k), z^k - u^k \rangle}{\|F(u^k)\|^2} F(u^k), & \text{if } \rho_k \geq \eta_1, \\ z^k, & \text{otherwise.} \end{cases}$$

- Update the regularization prameter

$$\lambda_{k+1} \in \begin{cases} (\underline{\lambda}, \lambda_k), & \text{if } \rho_k \geq \eta_2, \\ [\lambda_k, \gamma_1 \lambda_k], & \text{if } \eta_1 \leq \rho_k < \eta_2, \\ (\gamma_1 \lambda_k, \gamma_2 \lambda_k], & \text{otherwise,.} \end{cases}$$

- For any $z^* \in Z^*$ and any successful iteration

$$\|z^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 - \|z^{k+1} - z^k\|^2.$$

Globel covergence of semi-smooth Newton method

Assumption:

- Suppose that there exists a constant $c_1 > 0$ such that $\|J_k\| \leq c_1$ for any $k \geq 0$ and any $J_k \in \partial_B F(z^k)$.

glabal convergence:

- the successful index set

$$\mathcal{S} := \{k \geq 0 : \rho_k \geq \eta_1\}.$$

- Suppose that $\mathcal{S}$ is finite. Then $z^k = z^*$ for all sufficiently large k and $F(z^*) = 0$.
- Suppose that $\mathcal{S}$ is infinite. Then $\{z^k\}$ converges to some point $\bar{z}$ such that $F(\bar{z}) = 0$.

Semi-smooth Levenberg-Marquardt method

- Define a merit function

$$\phi(z) := \frac{1}{2}\|F(z)\|^2.$$

- Choose an element $J_k \in \partial_B F(z^k)$ and denote

$$F_k = F(z^k), \quad \phi_k = \phi(z^k), \quad g_k = J_k^T F_k.$$

- Consider a quadratic model

$$m_k(s) := \frac{1}{2}\|F_k + J_k s\|^2.$$

- The step s^k is obtained by solving

$$\min_s \; m_k(s) + \frac{1}{2}\mu_k \|F_k\| \cdot \|s\|^2,$$

Semi-smooth Levenberg-Marquardt method

- Solve a linear system

$$[J_k^T J_k + \mu_k \|F_k\| I]s = -J_k^T F_k,$$

where $\mu_k > 0$ is the regularization parameter.

- Solving approximately such that

$$\|r_k\| \leq \tau \min\{\bar{\eta}, \|F_k\|^2\},$$

where

$$r_k := (J_k^T J_k + \lambda_k \|F_k\| I)s^k + g_k.$$

Semi-smooth Levenberg-Marquardt method

- Define the ratio

$$\rho_k = \frac{\phi(z^k) - \phi(z^k + s^k)}{m_k(0) - m_k(s^k)}.$$

  which compare the actual reduction and predicted reduction.

- If ρ_k is big enough, the quadratic approximation is proper.

$$z^{k+1} = z^k + s^k.$$

- If ρ_k is too small, the quadratic approximation is bad.

$$z^{k+1} = z^k$$

  Increase the regularization parameter.

Semi-smooth Levenberg-Marquardt method

- Set the parameter $0 < \eta_1 \leq \eta_2 < 1$ and $1 < \gamma_1 \leq \gamma_2$,
- Update the iterate point

$$z^{k+1} = \begin{cases} z^k + s^k & \text{if } \rho_k \geq \eta_1, \\ z^k & \text{otherwise.} \end{cases}$$

- Update the regularized parameter μ_k

$$\mu_{k+1} \in \begin{cases} (0, \mu_k) & \text{if } \rho_k > \eta_2, \\ [\mu_k, \gamma_1 \mu_k] & \text{if } \eta_1 \leq \rho_k \leq \eta_2, \\ (\gamma_1 \mu_k, \gamma_2 \mu_k] & \text{otherwise.} \end{cases}$$

Convergence of semi-smooth LM method

Assumptions:

- The level set $L(z^0) := \{z \in \mathbb{R}^n : \phi(z) \leq \phi(z^0)\}$ is bounded.
- $F(z)$ is strongly semi-smooth over $L(z^0)$.

Convergence:

- The sequence $\{z^k\}$ generated by Semi-smoothness LM method satisfies

$$\lim_{k\to\infty} \|g_k\| = 0.$$

- F is BD-regular at $z \in \mathbb{R}^n$: every element in $\partial_B F(z)$ is nonsingular.
- If F is BD-regular at a solution z^* to $F(z) = 0$. Assume that the sequence $\{z^k\}$ lies in the neighborhood of z^*, then

$$\|z^{k+1} - z^*\| = O(\|z^k - z^*\|^2).$$

Globalization of the algorithm

Drawback:

- The semi-smooth LM algorithm converges to a local but not global optimal solution.
- Newton-type methods only have a local quadratic convergence rate and are more expensive than first order method in one iteration.

Goal:

- Develop some globalized technique to escape the local solutions.
- Reduce the complexity.

Globalization of the algorithm

Solution:

- Run the fixed-point iteration until $F(z) \leq \varepsilon_0$.
- Then use the LM method until $F(z) \leq \varepsilon$

Convergence:

- Global convergence and quadratic convergence rate.
- The complexity
  - fixed-point iterations: $O(1/\varepsilon_0^2)$
  - LM iterations: $O(\log(1/\varepsilon))$.

Applications to the FBS Method

- The fixed-point mapping

$$F(x) = \text{prox}_{tf}(x - t\nabla h(x)) - x.$$

- The generalized Jacobian matrix of $F(x)$ is

$$J(x) = M(x)(I - t\partial^2 h(x)) - I,$$

where $M(x) \in \partial \text{prox}_{tf}(x - t\nabla h(x))$ and $\partial^2 h(x)$ is the generalized Hessian matrix of $h(x)$.

LASSO Regression

- The Lasso regression problem

$$\min \frac{1}{2}\|Ax-b\|_2^2 \text{ s.t. } \|x\|_1 \leq \lambda,$$

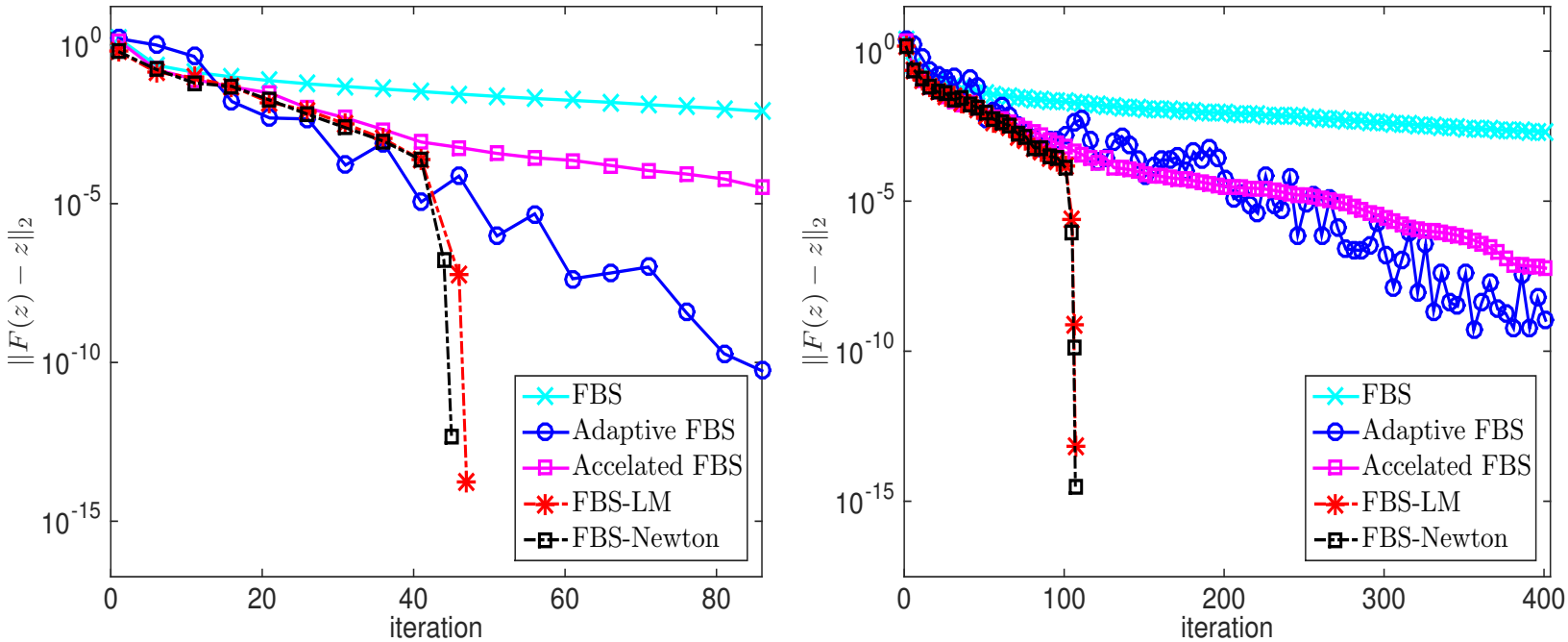
where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $\lambda \geq 0$ are given.

- $h(x) = \frac{1}{2}\|Ax-b\|_2^2$ and $f(x) = 1_\Omega(x)$, where $\Omega = \{x \mid \|x\|_1 \leq \lambda\}$.
- For a given $z \in \mathbb{R}^n$, let $|z_{[1]}| \geq |z_{[2]}| \geq \ldots \geq |z_{[n]}|$, the Jacobian matrix $M(z)$

$$M(z)_{ij} = \begin{cases} 1 & \text{if } \alpha < 0, j = i \\ 1 - \alpha \text{sign}(z_i)\text{sign}(z_j)/p, & \text{if } |z_i| \geq \alpha \text{ and } \alpha > 0, j = [1], \ldots, [p]. \end{cases}$$

where α be the largest value of $\left(\sum_{i=1}^k |z_{[i]}| - \lambda\right)/k$, $k = 1, \ldots, n$, and p be the corresponding k of α.

LASSO Regression



(a) $k = 50$ (b) $k = 150$

Figure: residual history of LASSO on $n = 1000$, $m = 500$ and $\mu = 0.9\|x\|_1$

Logistic Regression

- Sparse logistic regression problem

$$\min \ \mu\|x\|_1 + h(x),$$

where $\sum_{i=1}^{m} \log(e^{A_i x} + 1) - b_i^T A_i x$.

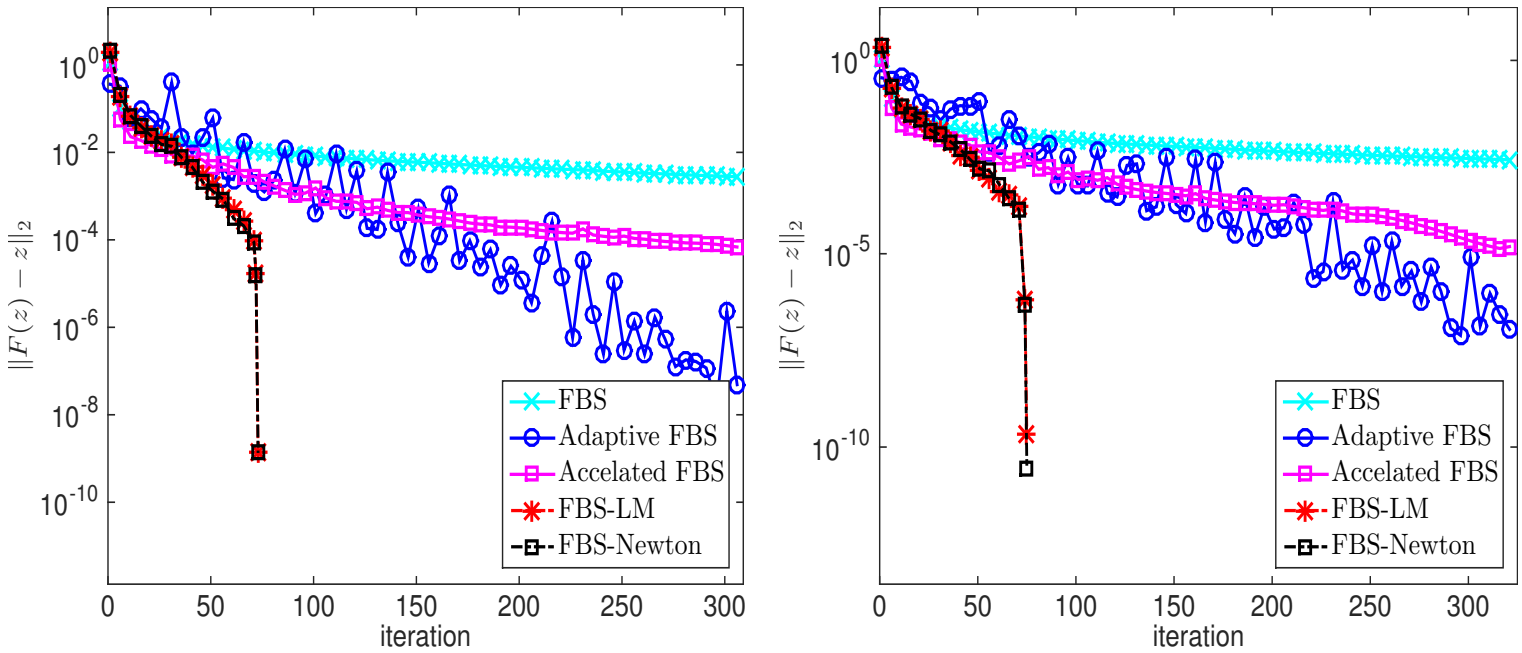
- The proximal mapping corresponding to $f(x) = \mu\|x\|_1$

$$\left(\text{prox}_{tf}(z)\right)_i = \text{sign}(z_i)\max(|z_i| - \mu t, 0).$$

- the Jacobian matrix $M(z)$ is diagonal matrix whose diagonal entries are

$$(M(z))_{ii} = \begin{cases} 1, & \text{if } |z_i| > \mu t, \\ 0, & \text{otherwise.} \end{cases}$$

Logistic Regression



(a) $k = 200$ (b) $k = 600$

Figure: residual history of the logistic regression problem on $n = 2000$, $m = 1000$ and $\mu = 1$

General Quadratic Programming

- The general quadratic programming

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}x^T Qx + c^T x, \text{ s.t. } Ax \leq b,$$

where $Q \in R^{n\times n}$ is symmetric positive definite, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

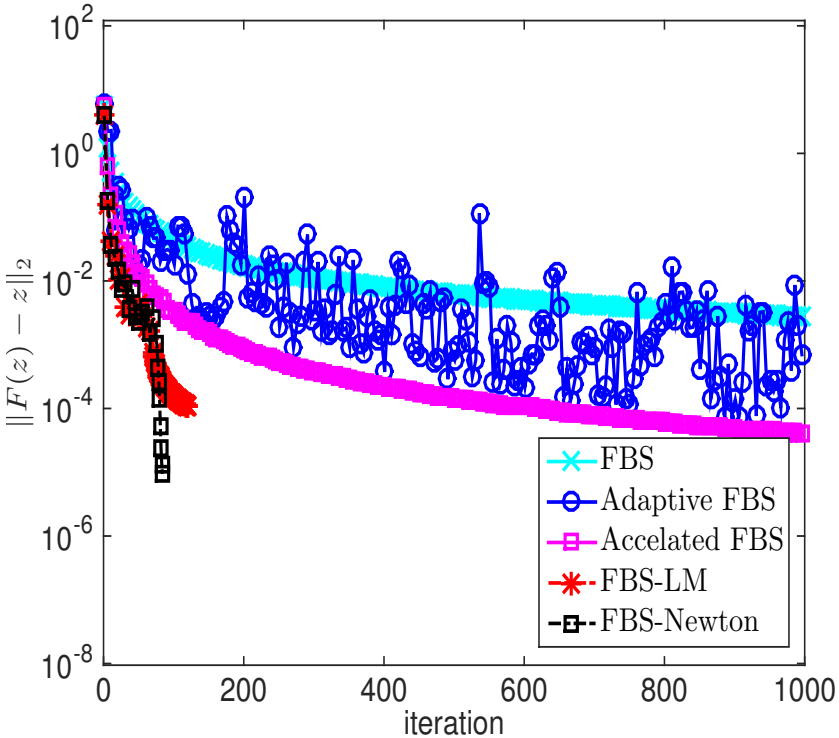
- The dual problem is

$$\max_{y\geq 0} \min_{x\in\mathbb{R}^n} \frac{1}{2}x^T Qx + c^T x + y^T(Ax - b),$$

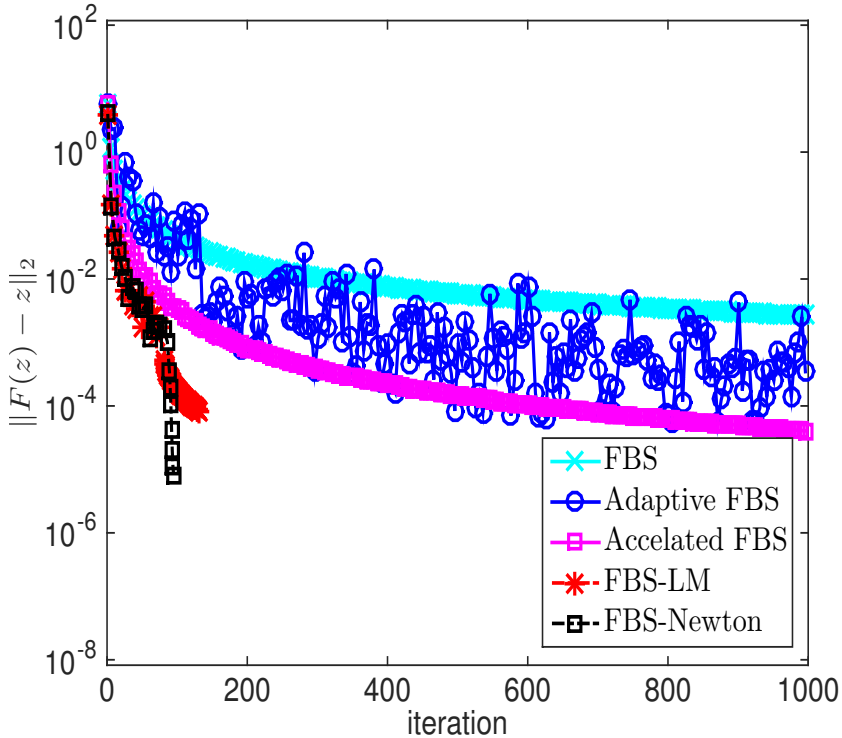
which is equivalent to

$$\min_{y\geq 0} \frac{1}{2}y^T(AQ^{-1}A^T)y + (AQ^{-1}c + b)^T y.$$

General Quadratic Programming



(a) LISWET1



(b) LISWET2

Figure: residual history of quadratic programming

l1-regularized optimization problems

- Consider the ℓ_1-regularized optimization problem of the form

$$\min \ \mu\|x\|_1 + h(x),$$

where $h(x) = \frac{1}{2}\|Ax - b\|_2^2$.

- The proximal mapping corresponding to $f(x) = \mu\|x\|_1$

$$\left(\text{prox}_{tf}(z)\right)_i = \text{sign}(z_i)\max(|z_i| - \mu t, 0).$$

- the Jacobian matrix $M(z)$ is diagonal matrix whose diagonal entries are

$$(M(z))_{ii} = \begin{cases} 1, & \text{if } |z_i| > \mu t, \\ 0, & \text{otherwise.} \end{cases}$$

l1-regularized optimization problems

Table: Total numbers of A- and A^T- calls N_A and CPU time (in seconds) averaged over 10 independent runs with dynamic range 20 dB and 40dB

method	$\epsilon : 10^{-2}$		$\epsilon : 10^{-4}$		$\epsilon : 10^{-8}$		$\epsilon : 10^{-10}$	
	time	N_A	time	N_A	time	N_A	time	N_A
SNF	1.12	86.6	2.91	232.2	3.75	296	3.76	307
SNF(aCG)	1.16	89	2.86	226.2	3.97	317.4	4.08	332.4
SSM-Newton	1.36	109	1.97	159.4	2.96	239.2	2.97	251.4
FPC-AS	83.2	5956	5.26	379.4	8.99	653.2	9.66	731.2
SpaRSA	0.794	74.6	1.47	141.5	2.74	267.3	3.24	323.9

method	$\epsilon : 10^{-2}$		$\epsilon : 10^{-4}$		$\epsilon : 10^{-8}$		$\epsilon : 10^{-10}$	
	time	N_A	time	N_A	time	N_A	time	N_A
SNF	2.16	162.6	5.08	393.8	6.41	509.6	6.81	531.6
SNF(aCG)	2.17	165	4.86	385.2	6.52	517	7.24	567
SSM-Newton	2.43	192.6	3.78	303	5.02	412.6	5.41	437.8
FPC-AS	5.24	370.6	24.6	1716.6	9.1	685	11.5	847.2
SpaRSA	3.95	374.5	5.43	515.7	7.29	706.6	8.05	778.2

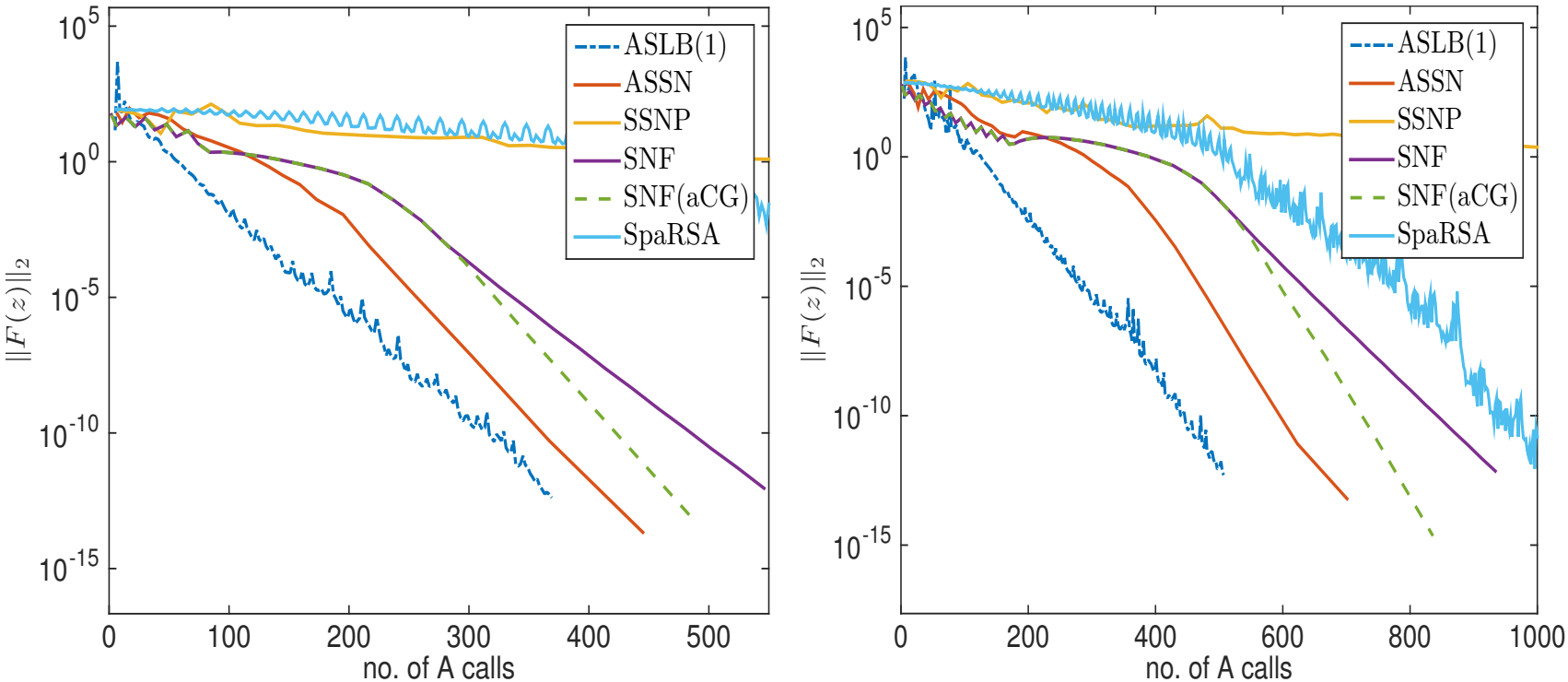
l1-regularized optimization problems

Table: Total numbers of A- and A^T- calls N_A and CPU time (in seconds) averaged over 10 independent runs with dynamic range 60 dB and 80dB

method	$\epsilon : 10^{-2}$		$\epsilon : 10^{-4}$		$\epsilon : 10^{-8}$		$\epsilon : 10^{-10}$	
	time	N_A	time	N_A	time	N_A	time	N_A
SNF	4.16	324	7.21	561.2	10.1	799.6	10.7	832.6
SNF(aCG)	4.16	324	7.3	561.2	10.2	802.4	10.9	871.2
SSM-Newton	3.21	257.4	5.4	433.4	7.44	603.6	8.21	663
FPC-AS	192	13776.8	9.58	658.8	12.8	935	16.4	1184.2
SpaRSA	24.2	2399.3	27.9	2743.3	31	3011.9	31.6	3111.5

method	$\epsilon : 10^{-2}$		$\epsilon : 10^{-4}$		$\epsilon : 10^{-8}$		$\epsilon : 10^{-10}$	
	time	N_A	time	N_A	time	N_A	time	N_A
SNF	7.44	582.6	7.48	589.4	12.4	980.8	12.9	1022.6
SNF(aCG)	8.07	635.2	8.11	635.2	12.1	953	12.8	1030.8
SSM-Newton	6.26	486	6.69	508.8	9.72	758.4	10.6	830
FPC-AS	5.15	368.4	6.41	442.4	11.2	829.4	15.6	1146.2
SpaRSA	160	15660.5	171	16781.3	174	17135.5	174	17260.9

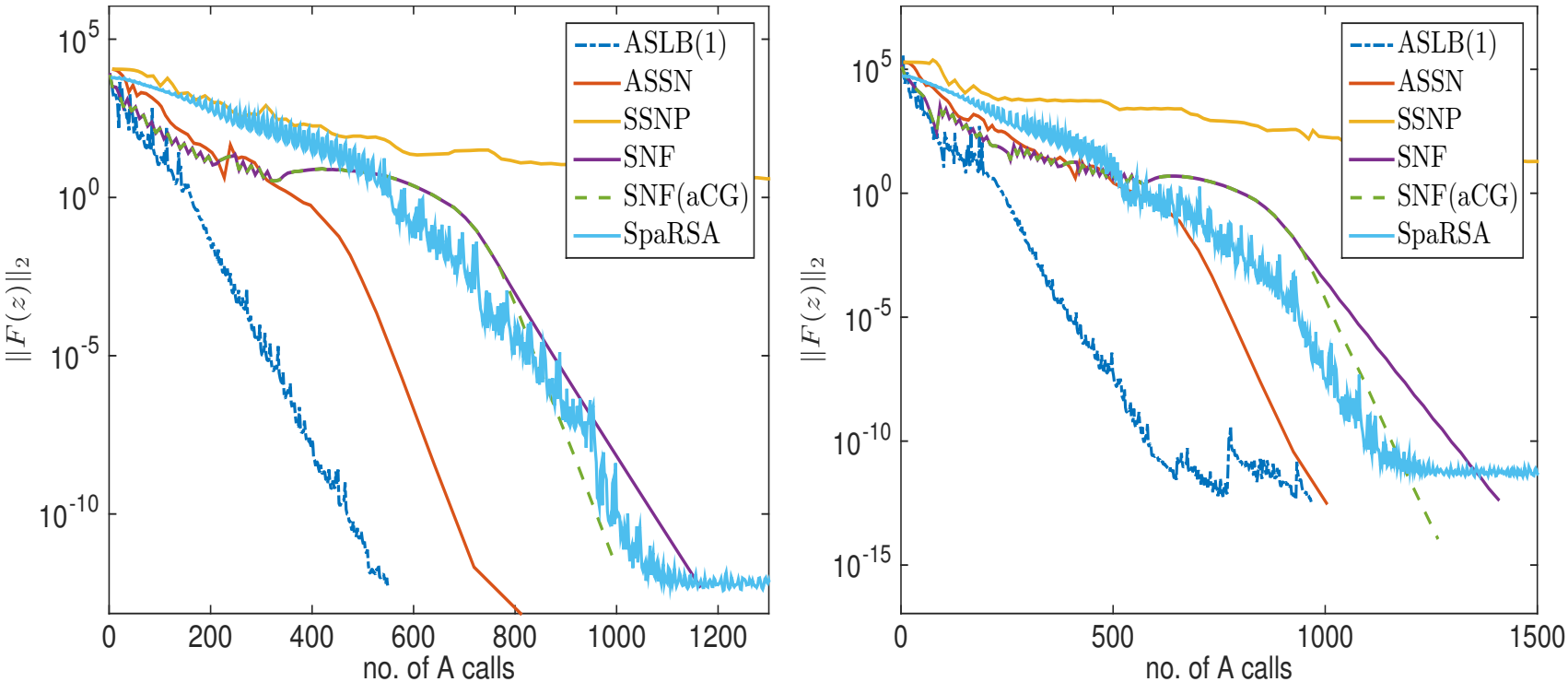
l1-regularized optimization problems



(a) 20dB (b) 40dB

Figure: residual history with respect to total numbers of A- and A^T- calls N_A

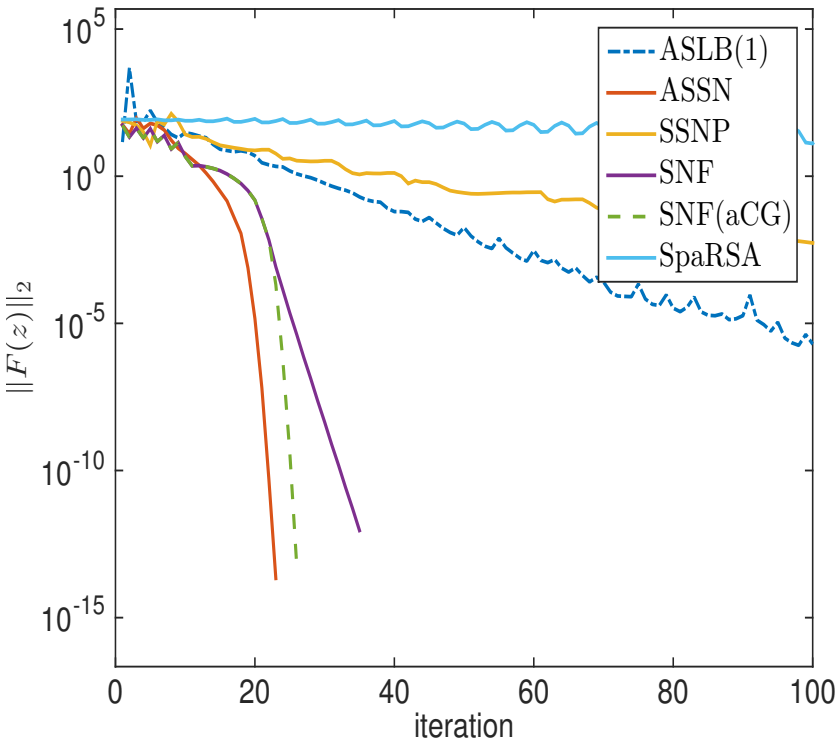
l1-regularized optimization problems



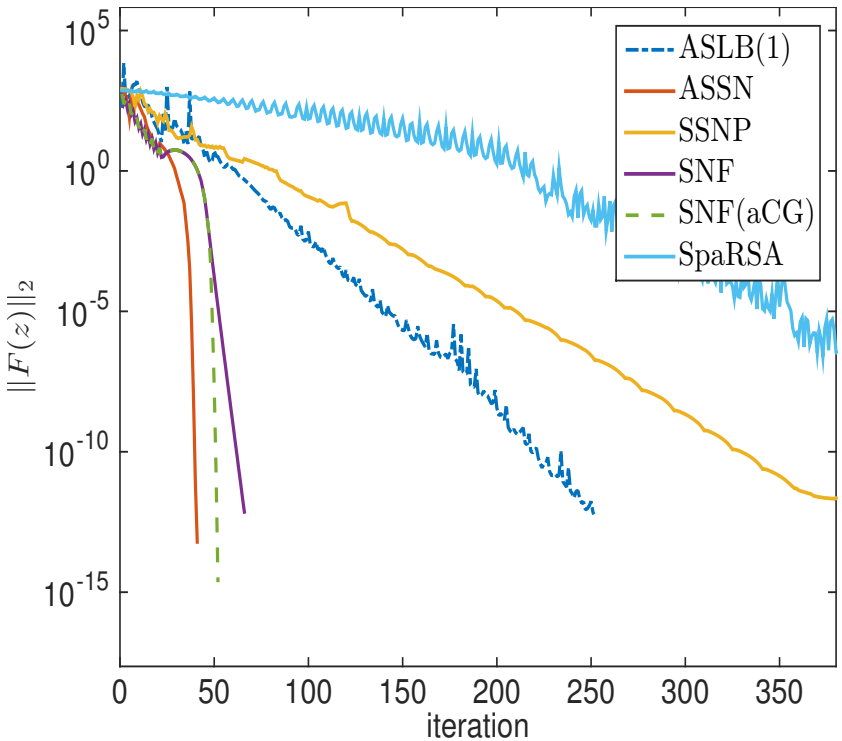
(a) 60dB (b) 80dB

Figure: residual history with respect to total numbers of A- and A^T- calls N_A

l1-regularized optimization problems



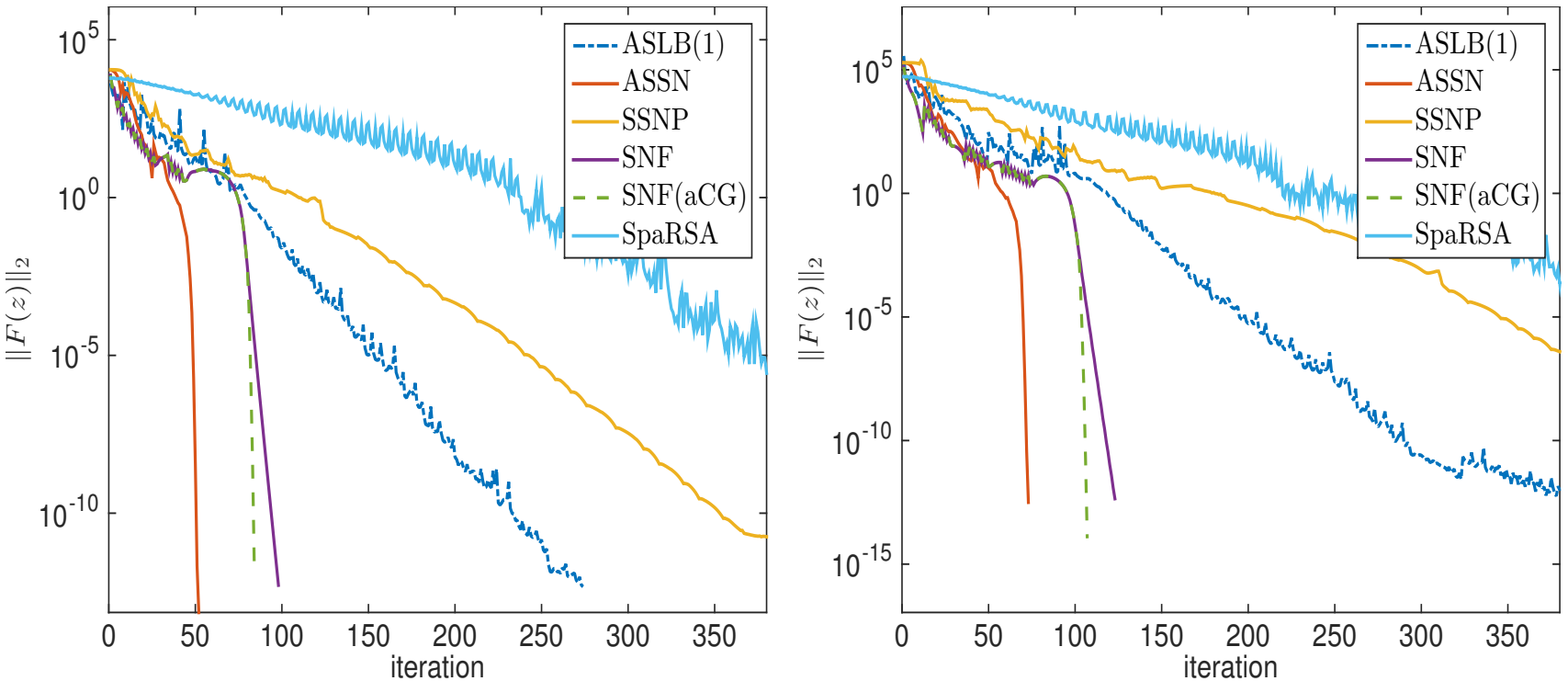
(a) 20dB



(b) 40dB

Figure: residual history with respect to total numbers of iteration

l1-regularized optimization problems



(a) 60dB (b) 80dB

Figure: residual history with respect to total numbers of iteration

Applications to the DRS Method

- Optimization problems

$$\min\ f(x),\ \text{s.t. } Ax = b,$$

where $A \in \mathbb{R}^{m\times n}$ is of full row rank and $b \in \mathbb{R}^m$.

- $h(x) = 1_\Omega(x)$, where $\Omega = \{x \mid Ax = b\}$.
- The proximal mapping with respect to $h(x)$ is

$$\text{prox}_{th}(x) = \mathcal{P}_\Omega(x) = (I - \mathcal{P}_{A^T})x + (A^T(AA^T)^{-1})b,$$

where $\mathcal{P}_{A^T} = A^T(AA^T)^{-1}A$.

Numerical results

- The DRS fixed-point mapping reduces to

$$F(z) = \text{prox}_{tf}((2D - I)z + 2\beta) - Dz - \beta,$$

where

$$D = I - \mathcal{P}_{A^T} \quad \text{and} \quad \beta = (A^T(AA^T)^{-1})b.$$

- The generalized Jacobian matrix of $F(z)$ is in the form of

$$J(z) = M(z)(2D - I) - D = \Psi(z) - \Phi(z)\mathcal{P}_{A^T},$$

where $M(z) \in \partial\text{prox}_{tf}((2D - I)z + 2\beta)$, $\Psi(z) = M(z) - I$ and $\Phi(z) = 2M(z) - I$.

Basis Pursuit

- The ℓ_1 minimization problem:

$$\min_{x\in\mathbb{R}^n} \quad \|x\|_1, \text{ s.t.} \quad Ax = b.$$

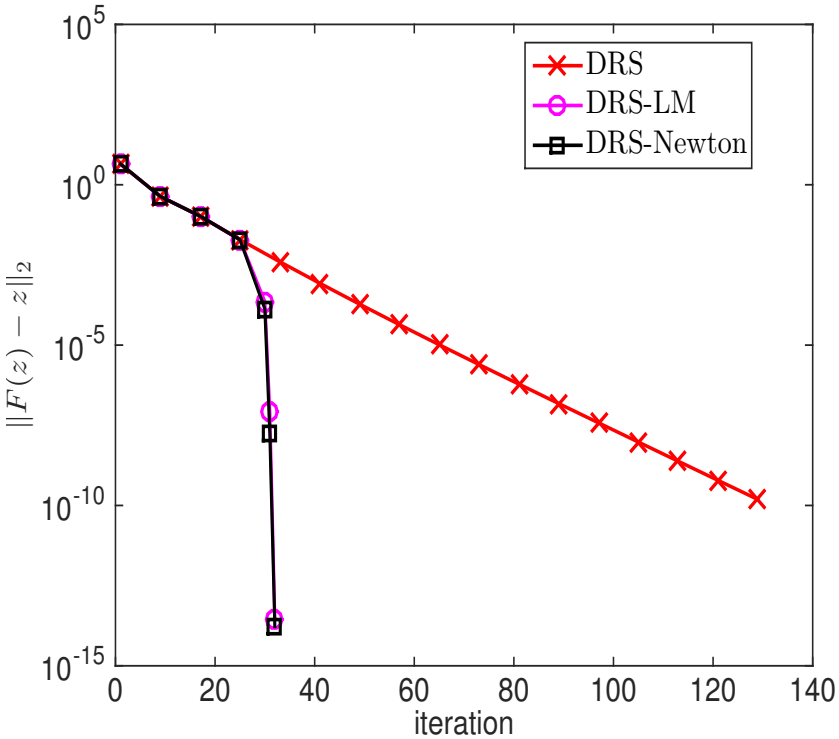
- The generalized Jacobian matrix $M(z)$ is a diagonal matrix with diagonal entries

$$M_{ii}(z)\begin{cases} = 1, & |((2D-I)z+2\beta)_i| > t, \\ = 0, & |((2D-I)z+2\beta)_i| < t, \\ \in [0,1], & |((2D-I)z+2\beta)_i| = t. \end{cases}$$

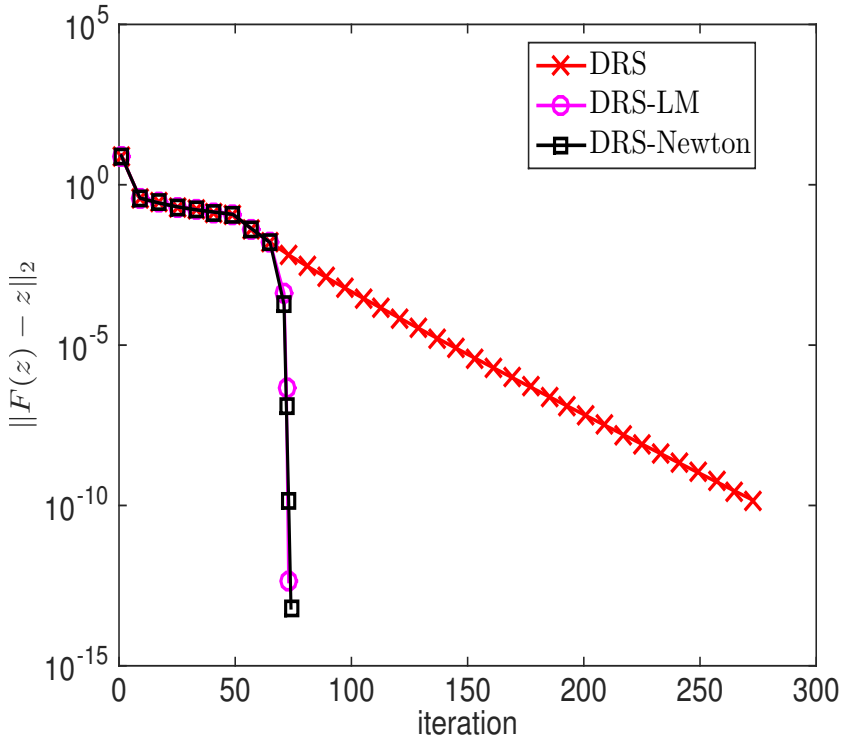
- Choose $M(z)$ such that $M_{ii}(z) = 1$ when $|((2D-I)z+2\beta)_i| = t$.
- $\Phi(z)$ and $\Psi(z)$ are diagonal matrices whose diagonal entries are

$$\begin{cases} \Psi_{ii}(z) = 0, & \Phi_{ii}(z) = 1, & |((2D-I)z+2\beta)_i| \geq t, \\ \Psi_{ii}(z) = -1, & \Phi_{ii}(z) = -1, & |((2D-I)z+2\beta)_i| < t. \end{cases}$$

Basis Pursuit



(a) $k = 50$



(b) $k = 150$

Figure: residual history of the ℓ_1-minimization problem on $n = 1000$ and $m = 500$

Linear Programming

- The classic linear programming problem

$$\min_{x\in\mathbb{R}^n} \quad c^T x, \ \text{s.t.} \quad Ax = b, \ x \geq 0.$$

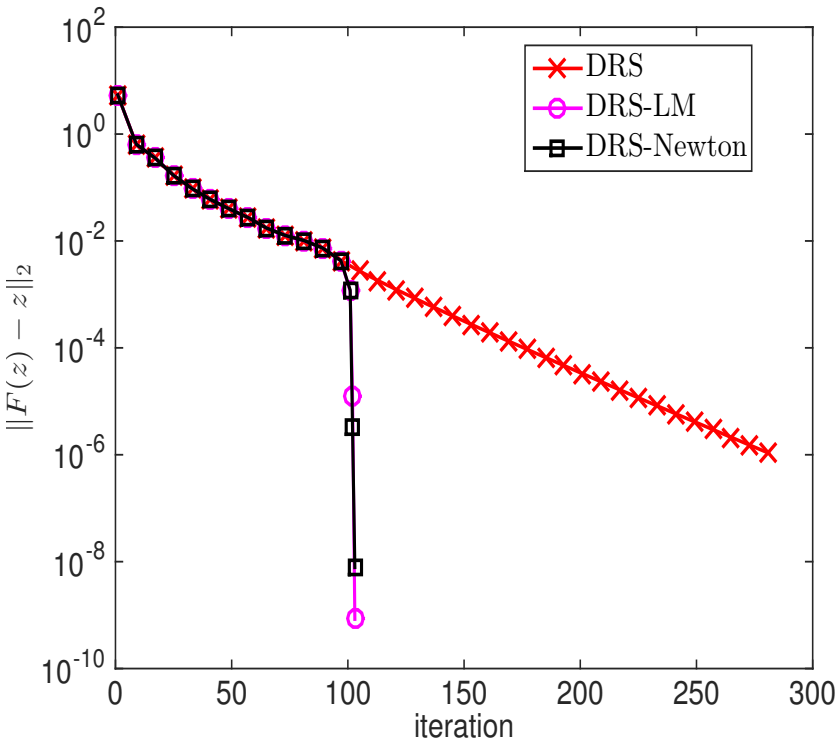
- Let $f(x) = c^T x + 1_K(x)$ where $K := \{x \mid x \geq 0\}$.
- Every element of the generalized Jacobian $\partial \mathcal{P}_K$ at $(2D - I)z + \beta$ is a diagonal matrix with diagonal entries

$$M_{ii}(z) \begin{cases} = 1, & ((2D - I)z + \beta)_i > 0, \\ = 0, & ((2D - I)z + \beta)_i < 0, \\ \in [0, 1], & ((2D - I)z + \beta)_i = 0. \end{cases}$$

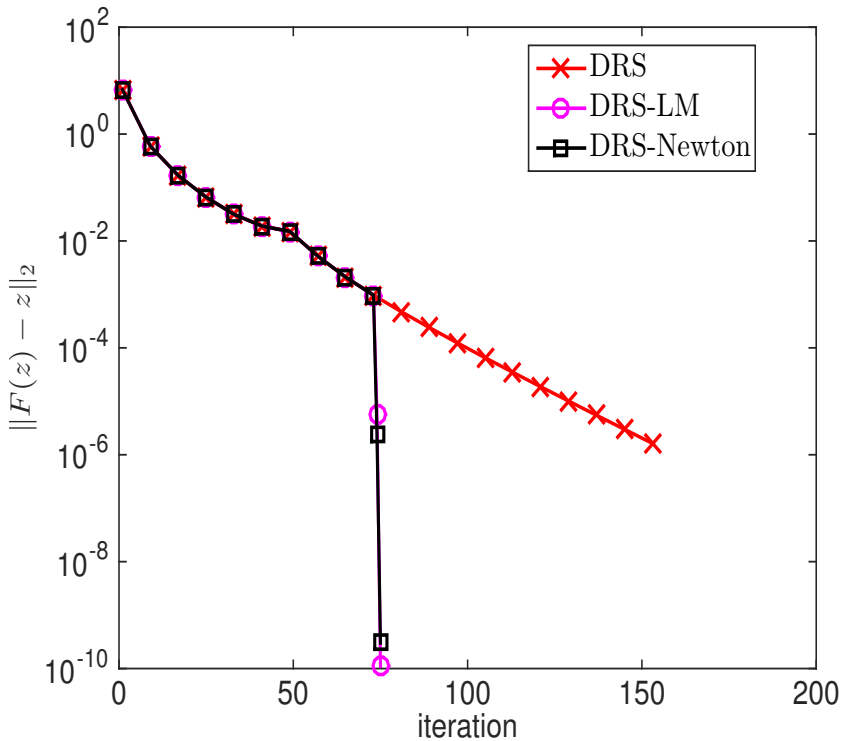
- Choose $M(z)$ such that $M_{ii}(z) = 1$ when $((2D - I)z + \beta)_i = 0$.
- we have

$$\begin{cases} \Psi_{ii}(z) = 0, \quad \Phi_{ii}(z) = 1, & ((2D - I)z + \beta)_i \geq 0, \\ \Psi_{ii}(z) = -1, \quad \Phi_{ii}(z) = -1, & ((2D - I)z + \beta)_i < 0. \end{cases}$$

Linear Programming



(a) $m = 300$



(b) $m = 400$

Figure: residual history of the LP problem on $n = 1000$

Conclusion and further reasearch

Conclusion

1. Second-order Convergence rate.
2. The proposed Newton method is theoretically guaranteed to converge to a global solution.
3. The proposed LM method converges quadratically as long as the initial point is close to the solution set enough.

Further reasearch

1. Second order algorithm on matrix optimization including semidefinite programming and low-rank optimization.
2. The error bound condition corresponding to DRS fixed-point residual function.
3. The linear convergence of DRS method under error bound condition.
4. The convergence rate of the proposed semi-smooth Newton method