Identifiability, regularity conditions and a two-metric adaptive projection for ℓ_1-norm minimization

Hanju Wu under supervision of Dr. Yue Xie

University of Hong Kong

July 3, 2025

Example 1: active sets in bound-constrained optimization

For a c^2-smooth and strongly convex function $f : \mathbb{R}^n \to \mathbb{R}$, the optimality condition of the bound-constrained problem

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{subject to} \quad x^i \geq 0, i = 1, \ldots, n.$$

is given by

$$0 \in \nabla f(x^*) + N_{\mathcal{X}}(x^*),$$

i.e.

$$\nabla f(x^*)^i \geq 0, \quad i = 1, \ldots, n, \quad \text{and} \quad x^{*i} > 0 \Rightarrow \nabla f(x^*)^i = 0.$$

Projected gradient iteration

$$x_{k+1} = P(x_k - \alpha_k \nabla f(x_k)),$$

converge to the optimal solution x^* with

$$\text{dist}(0, \nabla f(x_k) + N_{\mathcal{X}}(x_k)) \to 0$$

Moreover, if

$$0 \in \text{ri}(\nabla f(x^*) + N_{\mathcal{X}}(x^*)),$$

iterates identify an active manifold: eventually

$$x_k \in \mathcal{M} \triangleq \{x \geq 0 : x_i = 0, \forall i \in \mathcal{A}(x^*)\}$$

where $\mathcal{A}(x^*) = \{i : x_i^* = 0\}$.

Example 2: sparsity in ℓ_1-norm minimization

Consider ℓ_1 norm regularization problem, where f is a c^2-smooth and strongly convex function

$$\min f(x) + \gamma \|x\|_1$$

the optimality condition is given by

$$0 \in \nabla f(x^*) + \gamma \partial \|x^*\|_1.$$

Proximal gradient iteration

$$x_{k+1} = \text{prox}_{\gamma\|\cdot\|_1}(x_k - \alpha_k \nabla f(x_k)),$$

converge to the optimal solution x^* with

$$\text{dist}(0, \nabla f(x_k) + \gamma \partial \|x_k\|_1) \to 0.$$

Let $\mathcal{M} \triangleq \{x : x_i = 0, \forall i \in \mathcal{A}(x^*)\}$, where $\mathcal{A}(x^*) = \{i : x_i^* = 0\}$. Then $\mathcal{M}$ is identifiable at x^* if $0 \in \nabla f(x^*) + \gamma \text{ri}\, \partial \|x^*\|_1$.

Example 3: matrix rank in SDP

For a c^2-smooth and strongly convex function f, the optimality condition of the semidefinite programming problem

$$\min_{X \in \mathbb{S}^n} \; f(X) \quad \text{subject to} \quad X \succeq 0.$$

is given by

$$0 \in \nabla f(X^*) + N_{\mathbb{S}^n_+}(X^*).$$

Projected gradient iteration

$$X_{k+1} = P(X_k - \alpha_k \nabla f(X_k)),$$

converge to the optimal solution X^* with

$$\operatorname{dist}(0, \nabla f(X_k) + N_{\mathbb{S}^n_+}(X_k)) \to 0.$$

Moreover, if

$$0 \in \operatorname{ri}(\nabla f(X^*) + N_{\mathbb{S}^n_+}(X^*)),$$

iterates identify an active manifold: eventually

$$X_k \in \mathcal{M} \triangleq \{X : \operatorname{rank}(X) = \operatorname{rank}(X^*), X \succeq 0\},$$

Identifiability

Each example involves an active manifold of solutions and can be identified by diverse algorithms. Which means that

$$\text{dist}(0, \partial f(x_k)) \to 0 \Longrightarrow x_k \in \mathcal{M}, \forall k \quad \text{large enough}$$

Hence high-dimensional nonsmooth optimization

$$\min f(x)$$

reduce locally to low-dimensional smooth optimization on $\mathcal{M}$.

Some basic properties of slope

slope

The slope of a function f at a point x is defined as: if x is not a local minimizer

$$|\nabla f|(x) = \limsup_{y \neq x, y \to x} \frac{f(x) - f(y)}{d(x, y)}.$$

otherwise, $|\nabla f|(x) = 0$.

Properties of slope

- $|\nabla f|(x) \geq 0$ for all x, $|\nabla f|(x) = 0$ if x is a local minimizer.
- If f is convex, then $|\nabla f|(x) = \text{dist}(0, \partial f(x))$.
- If f is differentiable at x, then $|\nabla f|(x) = \|\nabla f(x)\|$.

Therefore, in the language of slope, identifiability can be rewrite as

$$|\nabla f|(x_k) \to 0 \Longrightarrow x_k \in \mathcal{M}, \forall k \quad \text{large enough}$$

Definition of identifiability on (X, d)

Definition

For a function f on a metric space X, consider a point $\bar{x} \in \operatorname{dom} f$, and a set $x \in \mathcal{M} \subset X$. The modulus of identifiability for $\mathcal{M}$ at $\bar{x}$ is

$$\liminf_{x \to_f \bar{x}, x \notin \mathcal{M}} |\nabla f|(x)$$

$\mathcal{M}$ is identifiable at $\bar{x}$ iff the modulus of identifiability is strictly positive.

Proposition

For a function f on a metric space X, consider a critical point $\bar{x} \in \operatorname{dom} f$, and a set $\mathcal{M}$ such that $x \in \mathcal{M} \subset X$ is identifiable at $\bar{x}$. If $x_k \to_f \bar{x}$, then

$$|\nabla f|(x_k) \to 0 \Longrightarrow x_k \in \mathcal{M}, \forall k \quad \textit{large enough}$$

Identifiability in Euclidean space

Definition

Given a closed function $f : \mathbb{R}^n \to (-\infty, +\infty]$, consider a point $\bar{x} \in \operatorname{dom} f$ and a manifold $\mathcal{M}$ containing $\bar{x}$. Then f is partly smooth at $\bar{x}$ for a subgradient $\bar{y} \in \partial f(\bar{x})$ relative to $\mathcal{M}$ if it satisfies the following properties:

Prox-regularity: the function f is prox-regular at $\bar{x}$ for $\bar{y}$.

Restricted smoothness: the restriction $f|_{\mathcal{M}}$ is smooth around $\bar{x}$.

Sharpness: the subspace $\operatorname{par} \hat{\partial} f(\bar{x})$ is just the normal space $N_{\mathcal{M}}(\bar{x})$.

Inner semicontinuity: for all $y \in \partial f(\bar{x})$ near $\bar{y}$ and sequences $x_r \to \bar{x}$ in $\mathcal{M}$, there exist $y_r \in \partial f(x_r)$ converging to y.

Theorem (Identifiability and partial smoothness)

Consider a closed function $f : \mathbb{R}^n \to (-\infty, +\infty]$ with slope zero at a point $\bar{x}$ lying in a manifold $\mathcal{M}$. Suppose that $f|_{\mathcal{M}}$ is smooth around $\bar{x}$. Then $\mathcal{M}$ is identifiable at $\bar{x}$ if and only if f is partly smooth at $\bar{x}$ for zero relative to $\mathcal{M}$ with $0 \in \operatorname{ri} \hat{\partial} f(\bar{x})$.

Uniqueness in Euclidean space

Proposition

If a closed function $f : \mathbb{R}^n \to (-\infty, +\infty]$ *is partly smooth for zero at a point* $\bar{x}$ *relative to manifold* $\mathcal{M}_1$ *and* $\mathcal{M}_2$*, there exists a neighborhood* $B_\epsilon(\bar{x})$ *with* $\epsilon > 0$ *such that*

$$\mathcal{M}_1 \cap B_\epsilon(\bar{x}) \;=\; \mathcal{M}_2 \cap B_\epsilon(\bar{x})$$

Regularity conditions

slope Error-Bound

We say f satisfies the slope Error Bound with constant $\mu > 0$ ($\mu - EB$) around $\bar{x}$ if

$$\mu \operatorname{dist}(x, S) \leq |\nabla f|(x)$$

hold for all x in some neighborhood of $\bar{x}$. Where S is an arbitrary set containing $\bar{x}$.

PL

$$(f(x) - f^\star) \leq \frac{1}{2\mu} |\nabla f|^2(x)$$

Quadratic Growth

$$\frac{\mu}{2} \operatorname{dist}^2(x, S) \leq f(x) - f^\star$$

relationship between $(\mathbb{R}^n, d)$ and $(\mathcal{M}, d_{Rie})$

[Lewis and Tian, 2024][Theorem 5.12, Proposition 8.6] states that

$$\text{KL on} \quad (\mathbb{R}^n, d) \Longleftrightarrow \quad \text{KL on} \quad (\mathcal{M}, d_{Rie})$$

However, there is a critical flaw in the proof $\Longrightarrow$. Actually, follow the idea of [Lewis and Tian, 2024], we can also prove that

$$\text{slope EB on} \quad (\mathbb{R}^n, d) \Longleftarrow \quad \text{slope EB on} \quad (\mathcal{M}, d_{Rie})$$

Theorem (Quadratic growth)

On a complete metric space, suppose that a closed function f has slope zero at a point $\bar{x}$, and consider any identifiable set $\mathcal{M}$ at $\bar{x}$. Then f has quadratic growth around $\bar{x}$ if and only if it has quadratic growth around $\bar{x}$ relative to $\mathcal{M}$. Indeed, the two growth rates are identical:

$$\liminf_{\substack{x\to\bar{x}\\ x\neq\bar{x}}} \frac{f(x)-f(\bar{x})}{d(x,\bar{x})^2} \;=\; \liminf_{\substack{x\to\bar{x},\ x\in\mathcal{M}\\ x\neq\bar{x}}} \frac{f(x)-f(\bar{x})}{d(x,\bar{x})^2}.$$

Relationship between regularity conditions

Now suppose f is convex.

relationship on $(\mathbb{R}^n, d)$

$$\text{PL (KL)} \iff \text{slope EB} \iff \text{Quadratic Growth}$$

subdifferential EB

We say f satisfies the subdifferential Error Bound with constant $\mu > 0$ $(\mu - EB)$ around $\bar{x}$ if

$$\mu \operatorname{dist}(x, S) \leq \operatorname{dist}(0, \partial f(x))$$

hold for all x in some neighborhood of $\bar{x}$. Where S is an arbitrary set containing $\bar{x}$.

Now suppose $f|_{\mathcal{M}}$ is C^2-smooth around $\bar{x}$, $\mathcal{M}$ is a C^2-smooth identifiable manifold with Riemannian metric d_{Rie}, then [Rebjock and Boumal, 2024] states that

relationship on $(\mathcal{M}, d_{Rie})$

$$\text{PL (KL)} \iff \text{slope EB} \iff \text{Quadratic Growth}$$

Note that these conditions are limited to the manifold $\mathcal{M}$, and

$$|\nabla f|(x) = \|\operatorname{grad}_{\mathcal{M}} f(x)\|$$

Suppose that f is convex, $f|_{\mathcal{M}}$ is C^2-smooth around $\bar{x}$, $\mathcal{M}$ is a C^2-smooth identifiable manifold with Riemannian metric d_{Rie}, then

conclusion

$$\begin{array}{ccccc} (\mathbb{R}^n, d) & \text{PL (KL)} & \Longleftarrow & \text{EB} & \Longleftarrow & \text{QG} \\ & \uparrow & & \uparrow & & \\ (\mathcal{M}, d_{Rie}) & \text{PL (KL)} & \Longleftarrow & \text{EB} & \Longleftarrow & \text{QG} \end{array}$$

open question

What is the most general condition can make the following implication hold with $\mathcal{M}$ is a C^2-smooth identifiable manifold with Riemannian metric d_{Rie}?

$$\text{PL (KL) on} \quad (X, d) \Longrightarrow \quad \text{PL (KL) on} \quad (\mathcal{M}, d_{Rie})$$

In which cases EB holds?

convex composite optimization

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

where f is a smooth convex function and h is a proper closed convex function.

[Drusvyatskiy and Lewis, 2018][Theorem 3.3, Corollary 3.6] states that proximal EB is equivalent to subdifferential EB.

proximal-EB

For an optimal point $x^* \in X^*$, there exists a neighborhood $B(x^*, \delta)$ such that for all $x \in B(x^*, \delta)$,

$$\kappa r(x) \geq \operatorname{dist}(x, X^*)$$

where $r(x)$ is a residual function $\|x - \operatorname{prox}_h(x - \nabla f(x))\|$ and $\kappa > 0$ is a constant.

[Zhou and So, 2017] states that proximal EB holds for a class of structured convex optimization problems, including

- f is strongly convex, ∇f is Lipschitz continuous, h is arbitrary proper closed convex function.
- $f(x) = g(Ax)$, where g is a proper convex function satisfies
  - g is continuously differentiable on $\mathrm{dom}(g)$.
  - g is strongly convex, ∇g is Lipschitz continuous on any compact subset of $\mathrm{dom}(g)$.

  and A is a linear operator. h is group LASSO regularizer.
- $f(x) = g(Ax) + \langle c, x \rangle$, where g is a proper convex function are as above. h has a polyhedral epigraph.

Existed second-order methods

successive quadratic approximation (SQA)

$$\min_{x\in\mathbb{R}^n} q_k(x) := f(x^k) + \nabla f(x^k)^T(x - x^k) + \frac{1}{2}(x - x^k)^T H_k(x - x^k) + h(x),$$

where H_k is an appropriate approximation of the Hessian of f.

Unfortunately, this subproblem has no closed-form solution as H_k is non-diagonal, so one needs to use an iterative solver for this sub-problem and the running time to reach the accuracy requirement can be very large. In this cases, its superlinear convergence thus gives little practical advantage in running time compare with first-order method.

Bound-Constrained Problem

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{subject to} \quad x^i \geq 0, i = 1, \ldots, n. \tag{1}$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is bounded below by f_{low} on the feasible region.

Two-Metric Projection [Bertsekas, 1982]

$$x_{k+1} := P(x_k - \alpha_k D_k \nabla f(x_k)),$$

- $P(z)$ is the projection onto the feasible region, i.e.

$$[P(z)]^i = \max\{z^i, 0\},$$

- $D_k \in \mathbb{R}^{n \times n}$, positive definite matrix
- To ensure descent in f, require $D_k[i,j] = 0$, $\forall i, j \in I_k^+$, $j \neq i$.

$$D_k = \underbrace{\left(\begin{array}{c|ccc} \bar{D}_k & 0 & \cdots & 0 \\ \hline 0 & d^{r_k+1} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & d^n \end{array}\right)}_{I_k^+}$$

$I_k^+ \triangleq \{i \mid 0 \leq x_k^i \leq \epsilon_k, \nabla_i f(x_k) > 0\}$.

Convergence properties

Main result

- Global convergence under suitable line search.
- $$\left[\bar{D}_k^{-1}\right]_{ij} = \frac{\partial^2 f(x_k)}{\partial x^i \partial x^j} \quad \forall i,j \notin I_k^+.$$

  Strict Complementarity + Local Strong Convexity $\Rightarrow$ Identification of Active Set ($I_k^+ = \mathcal{A}(x^*)$) + Quadratic Convergence rate.

- Eigenvalues of D_k should be uniformly bounded.
- Newton's equation should be solved exactly.

These requirements are rarely met in practice.

Bound-constrained formulation of ℓ_1-norm minimization

ℓ_1-norm regularization

$$\min_{x\in\mathbb{R}^n} \psi(x) = f(x) + h(x), \quad h(x) = \gamma\|x\|_1 \tag{2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a C^2-smooth and convex function, and $\gamma > 0$ is a regularization parameter.

Let $x = x^+ - x^-$ in (2), where $x^+ = \max(0, x)$ and $x^- = -\min(0, x)$. Then (2) can be reformulated as the following constrained problem

Bound-constrained formulation

$$\begin{aligned} \min_{x^+,x^-\in\mathbb{R}^n} \quad & f\left(x^+ - x^-\right) + \gamma \textstyle\sum_i \left[x_i^+ + x_i^-\right] \\ \text{s.t.} \quad & x^+ \geq 0, \quad x^- \geq 0. \end{aligned} \tag{3}$$

Why Two-metric projection can not be used directly?

Given an ϵ_k, let I_k^1 and I_k^2 are the "I_k^- part" corresponding to x_k^+ and x_k^- respectively. Then we have

$$I_k^1 = \{i \mid 0 \leq (x_k^+)^i \leq \epsilon_k, g_k^i + \gamma \leq 0\} \cup \{i \mid (x_k^+)^i > \epsilon_k\}$$

$$I_k^2 = \{i \mid 0 \leq (x_k^-)^i \leq \epsilon_k, -g_k^i + \gamma \leq 0\} \cup \{i \mid (x_k^-)^i > \epsilon_k\}$$

- $I_k^1 \cap I_k^2 \neq \emptyset$ if $(x_k^+)^i > \epsilon_k, (x_k^-)^i \leq \epsilon_k, -g_k^i + \gamma \leq 0$ or $(x_k^-)^i > \epsilon_k, (x_k^+)^i \leq \epsilon_k, g_k^i + \gamma \leq 0$.
- $\bar{D}_k$ is singular if $I_k^1 \cap I_k^2 \neq \emptyset$.
- Newton's equation is unsolvable even if $\nabla^2 f \succ 0$.
- Inexact apporximation $\nabla^2 f(x_k) + \mu_k I$ will lead to numerical instability when μ_k is small.

Two-metric Adaptive Projection

- Slightly modify the definition of I_k^1 and I_k^2 and reduce the problem size by aggregating x_k^+ and x_k^- to avoid the intersection.

$$I_k^{-+} \triangleq \left\{ i : \begin{array}{l} x_k^i > \epsilon_k \text{ or} \\ 0 \le x_k^i \le \epsilon_k, g_k^i \le -\gamma \end{array} \right\} \subseteq I_k^1,$$

$$I_k^{--} \triangleq \left\{ i : \begin{array}{l} x_k^i < -\epsilon_k \text{ or} \\ -\epsilon_k \le x_k^i \le 0, g_k^i \ge \gamma \end{array} \right\} \subseteq I_k^2,$$

$$I_k^- \triangleq I_k^{-+} \cup I_k^{--}, I_k^{-+} \cap I_k^{--} = \emptyset.$$

- The Newton's equation doesn't need to be solved exactly i.e.

$$(H_k + \mu_k I)[p_k]_{I_k^-} = [g_k + \omega_{k,\epsilon}]_{I_k^-} + r_k$$

- Reserve **identification of active set** ($I_k^+ = \mathcal{A}(x^*)$) and **superlinear convergence** under a weaker condition than local strong convexity.

Two-metric Adaptive Projection

$$x_{k+1} := \mathcal{P}_{k,\epsilon}\left(x_k - t_k p_k\right), \quad \forall k \geq 0,$$

where $t_k > 0$ is the stepsize determinded by line search. The step p_k is defined as below:

$$[p_k]_{I_k^+} \triangleq [g_k + \omega_{k,\epsilon}]_{I_k^+}$$

and $[p_k]_{I_k^-}$ satisfies

$$(H_k + \mu_k I)[p_k]_{I_k^-} = [g_k + \omega_{k,\epsilon}]_{I_k^-} + r_k \tag{4}$$

where H_k is a symmetric positive semi-definite matrix. μ_k is a positive scalar such that

$$\mu_k = c \left\| \begin{bmatrix} [x_k - \mathrm{Prox}_h(x_k - g_k)]_{I_k^+} \\ [g_k + \omega_k]_{I_k^-} \end{bmatrix} \right\|^{\delta}, \delta \in (0,1),$$

and $r_k \in \mathbb{R}^{|I_k^-|}$ is a residual that satisfies the following condition for a fixed $\tau \in (0,1)$:

$$\|r_k\| \leq \tau \min\{\mu_k \|[p_k]_{I_k^-}\|, \|[g_k + \omega_k]_{I_k^-}\|\}$$

Two-metric Adaptive Projection

$\mathcal{P}_{k,\epsilon}$ and $\omega_{k,\epsilon}$ are associated to x_k, ϵ and defined as below (recall that S_t denotes the soft thresholding operator):

$$\mathcal{P}^i_{k,\epsilon}(v) = \begin{cases} \max\left\{v^i, 0\right\} & \text{if} \quad i \in I_k^{-+} \\ \min\left\{v^i, 0\right\} & \text{if} \quad i \in I_k^{--} \\ S_{t_k\gamma}(v^i) & \text{if} \quad i \in I_k^{+} \end{cases}$$

$$\omega^i_{k,\epsilon} = \begin{cases} \gamma & \text{if} \quad i \in I_k^{-+} \\ -\gamma & \text{if} \quad i \in I_k^{--} \\ 0 & \text{if} \quad i \in I_k^{+} \end{cases}.$$

Convergence properties

Theorem (equivalence between EB on (X, d) and $(\mathcal{M}, d_{Rie})$)

When strict complementarity holds at an optimal point x^*, EB on (X, d) implies EB on $(\mathcal{M}, d_{Rie})$. Where $d_{Rie} = d$ in this case.

Main result

Strict Complementarity + Error Bound $\Rightarrow$ Identification of Active Set $(I_k^+ = \mathcal{A}(x^*))$ + Superlinear convergence $(1 + \delta)$.

Numerical experiments

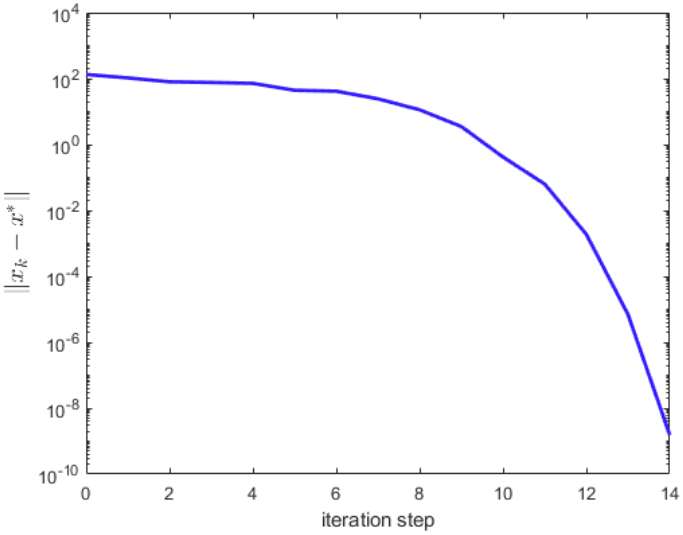
logistic regression

$$\min_{x\in\mathbb{R}^n} \frac{1}{m}\sum_{i=1}^{m} \log\left(1+\exp\left(-b_i\cdot a_i^T x\right)\right)+\gamma\|x\|_1, \gamma=\frac{1}{m}$$

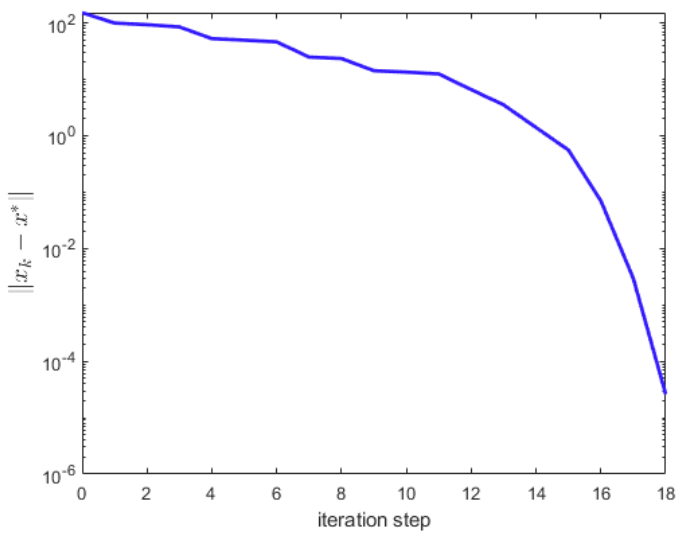
- Second-order method: IRPN [Yue et al., 2019] (SQA-type method), newGLMNET
- First-order method: SpaRSA, FISTA

datasets		IRPN	TMAP
rcv1_train	outer iter.	6	-
	inner iter.	492	13
	time	1.46	0.17
news20	outer iter.	9	-
	inner iter.	1197	18
	time	40.42	1.64
real-sim	outer iter.	13	-
	inner iter.	547	21
	time	4.30	1.11

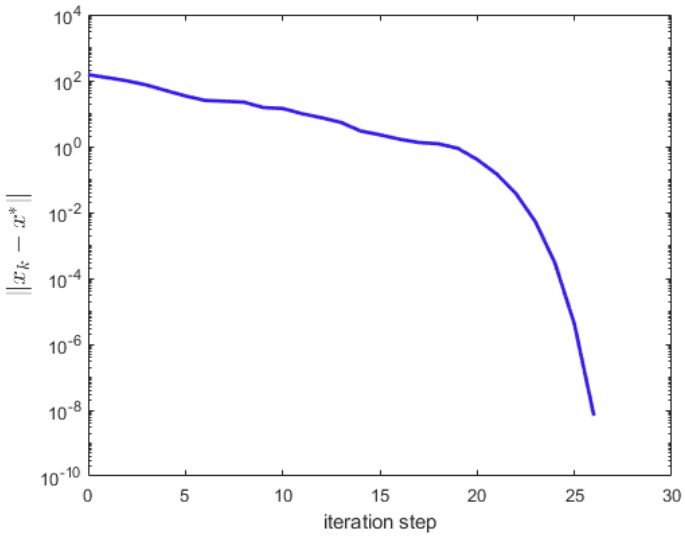
TMAP $\gg$ SQA-type method and newGLMNET (3-10 times associate to the sparsity) $>$ SpaRSA and FISTA.



rcv1



news20



real-sim

Numerical experiments

LASSO (large-scale reconstruction problem)

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \gamma \|x\|_1.$$

- Second-order method: ASSN [Xiao et al., 2018] (Semi-smooth Newton method).
- First-order method: SpaRSA, FPC_AS, ADMM

dataset size

$n = 512^2 = 262144$, $m = n/8 = 32768$, sparsity $= 2.5\%$.

Table: N_A denotes the total number of calls to A and A^T and the CPU time (in seconds) is averaged over 10 independent runs.

Dynamic range		ASSN	TMAP
20dB	N_A	298.2	300.6
	time	1.31	1.30
40dB	N_A	459.2	408.9
	time	2.51	2.32
60dB	N_A	635.4	624.9
	time	2.29	2.23
80dB	N_A	858.2	791.5
	time	2.99	2.74

TMAP $\approx$ Semi-smooth Newton method $\gg$ SpaRSA, FPC_AS and ADMM.

Summary

	Our algorithm	**Bertsekas algorithm**
Problem class	ℓ_1-norm minimization	Bound-constrained
Regularity condition	Error Bound	Local Strong Convexity
Newton's equation	Inexact	Exact
Global convergence	$\surd$	$\surd$
Local convergence rate	Superlinear $(1+\delta)$	Quadratic

Competitive against the state-of-the-art algorithms for large-scale ℓ_1-norm minimization (LASSO, logistic regression).

Future work

- Use BFGS to approximate the Hessian matrix in the Two-metric Adaptive Projection algorithm.
- Extend the algorithm to more general optimization problems.
- Identifiability requires the nondegenracy/strict complementarity, in which case can this condition satisfied?
- In which cases the regularity conditions on (X, d) imply the regularity conditions on $(\mathcal{M}, d_{Rie})$? ($\mathcal{M}$ is identifiable manifolds)

References I


Bertsekas, D. P. (1982).
Projected newton methods for optimization problems with simple constraints.
SIAM Journal on control and Optimization, 20(2):221–246.


Drusvyatskiy, D. and Lewis, A. S. (2018).
Error bounds, quadratic growth, and linear convergence of proximal methods.
Mathematics of Operations Research, 43(3):919–948.


Lewis, A. and Tian, T. (2024).
Identifiability, the kl property in metric spaces, and subgradient curves.
Foundations of Computational Mathematics, pages 1–38.

References II


Rebjock, Q. and Boumal, N. (2024).
Fast convergence to non-isolated minima: four equivalent conditions for c 2 functions.
Mathematical Programming, pages 1–49.


Xiao, X., Li, Y., Wen, Z., and Zhang, L. (2018).
A regularized semi-smooth newton method with projection steps for composite convex programs.
Journal of Scientific Computing, 76:364–389.


Yue, M.-C., Zhou, Z., and So, A. M.-C. (2019).
A family of inexact sqa methods for non-smooth convex minimization with provable convergence guarantees based on the luo–tseng error bound property.
Mathematical Programming, 174(1):327–358.

References III


Zhou, Z. and So, A. M.-C. (2017).
A unified approach to error bounds for structured convex optimization problems.
Mathematical Programming, 165:689–728.