On Resolution of ℓ_1-Norm Minimization via a Two-metric Adaptive Projection Method

Hanju Wu, Yue Xie

University of Hong Kong

April 18, 2025

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 $D_k[i,j] = 0$, $\forall i, j \in I_k^+$, $j \neq i$.

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

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

Convergence properties

Main result

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

  Strict Complementarity + Local Strong Convexity ⇒ 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

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

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

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

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 determined 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

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 $\|x - \operatorname{prox}_h(x - g)\|$ and $\kappa > 0$ is a constant.

Hold for a class of structured convex optimization problems.

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)Yue, Zhou, and So] (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.

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)Xiao, Li, Wen, and Zhang] (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).

References

BERTSEKAS D P.
Projected newton methods for optimization problems with simple constraints[J].
SIAM Journal on control and Optimization, 1982, 20(2): 221-246.

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

XIAO X, LI Y, WEN Z, et al.
A regularized semi-smooth newton method with projection steps for composite convex programs[J].
Journal of Scientific Computing, 2018, 76: 364-389.