Dual reweighted ℓ_1-algorithms for sparse optimization problem

Yunbin Zhao (with Kocvara & Luo)
School of Mathematics, University of Birmingham, UK
http://web.mat.bham.ac.uk/Y.Zhao
email: y.zhao.2@bham.ac.uk

Sparse optimization

ℓ_0-minimization:

$$\min\{\|x\|_0 : x \in P\}.$$

For instance,

$$P = \{x : \ Ax = b\},$$

$$P = \{x : \ Ax = b, \ x \geq 0\}$$

or other more general convex sets.

How to solve the ℓ_0-problem?

- ℓ_0-problem is NP-hard (Natarajan 1995).
- Various methods have been developed:
  - Orthogonal matching pursuit
  - Compressive sampling matching pursuit
  - Thresholding-type method
  - Subspace pursuit
  - Continuous approximation (typically, concave minimization)
  - ℓ_1-minimization
  - (Re)weighted ℓ_1-methods

(I) From approximation to weighted ℓ_1-algorithms

Definition. A function $\Phi_\varepsilon(x)$ from R^n_+ to R is said to be a **merit function for sparsity** if

$$\Phi_\varepsilon(x) \to \|x\|_0$$

as $\varepsilon \to 0$.

Examples.

$$\Phi_\varepsilon(s) = \sum_{i=1}^{n} \left(1 - e^{-\frac{s_i}{\varepsilon}}\right), \quad \text{where } s \in R_+^n;$$

$$\Phi_\varepsilon(s) = \sum_{i=1}^{n} \frac{s_i}{s_i + \varepsilon}, \quad \text{where } s_i > -\varepsilon \text{ for all } i = 1, ..., n;$$

$$\Phi_\varepsilon(s) = n - \frac{1}{\log \varepsilon} \left(\sum_{i=1}^{n} \log(s_i + \varepsilon) \right), \quad \text{where } s_i > -\varepsilon \text{ for all } i = 1, ..., n.$$

Where $\varepsilon \in (0, 1)$ is a given parameter.

Properties of these functions

- (i) $\Phi_\varepsilon(s)$ is continuously differentiable in s over an open neighborhood of $\mathbb{R}^n_+$;
- (ii) $\Phi_\varepsilon(s)$ is concave in $s \in \mathbb{R}^n_+$;
- (iii) $\Phi_\varepsilon(s) \to \|s\|_0$ as $\varepsilon \to 0$ for every given $s \in \mathbb{R}^n_+$;
- (iv) $\Phi_\varepsilon(s)$ is increasing w.r.t every component of s.

Concave approximation of ℓ_0-problem

The ℓ_0 problem

$$\min\{\|x\|_0 : \ Ax = b, \ x \geq 0\}$$

is approximated by

$$\min\{\Phi_\varepsilon(x) : Ax = b, \ x \geq 0\}.$$

The first-order approximation of Φ_ε

$$\Phi_\varepsilon(x) = \Phi_\varepsilon(x^k) + \nabla\Phi_\varepsilon(x^k)^T(x - x^k) + o(\|x - x^k\|)$$

leads to a general framework of (re)weighted ℓ_1-minimization

Reweighted ℓ_1-method [Zhao and Li (2012)[1]]:

S1. Choose $\varepsilon_0 \in (0,1)$, and let $x^0 \in R^n_+$ be an initial point.

S2. At the current iterate x^k with $\varepsilon_k > 0$, compute

$$x^{k+1} = \arg\min \left\{ \|W^k x\|_1 : \ Ax = b, \ x \geq 0 \right\},$$

where

$$W^k = \mathrm{diag}(\nabla \Phi_{\varepsilon_k}(x^k)).$$

S3. Set $\varepsilon_{k+1} = \alpha \varepsilon_k$. Replace (x^k, ε_k) with $(x^{k+1}, \varepsilon_{k+1})$ and repeat the step S2.

[1] Y.B. Zhao and D. Li, Reweighted ℓ_1-minimization for sparse solutions of underdetermined linear systems, *SIAM J. Optim.*, 22(2012), no.3, 1065-1088.

Example (i) Choose

$$\Phi_\varepsilon(x) = \sum_{i=1}^{n} \log(x_i + \varepsilon), \quad x \in R^n_+$$

At $x \in R^n_+$, the gradient is given by

$$\nabla \Phi_\varepsilon(x) = \left(\frac{1}{x_1 + \varepsilon}, \ldots, \frac{1}{x_n + \varepsilon} \right)^T \in R^n_{++},$$

yielding the well-known Candés, Walkin and Boyd (CWB)'s method (2008) with

$$w_i = \frac{1}{|x_i| + \varepsilon}, \; i = 1, \ldots, n$$

Example (ii) For a given $p \in (0, 1)$, Choose

$$\Phi_\varepsilon(x) = \frac{1}{p} \sum_{i=1}^{n} (x_i + \varepsilon)^p, \quad x \in R_+^n$$

At $x \in R_+^n$, the gradient is

$$\nabla \Phi_\varepsilon(x) = \left(\frac{1}{(x_1 + \varepsilon)^{1-p}}, ..., \frac{1}{(x_n + \varepsilon)^{1-p}} \right)^T \in R_{++}^n.$$

yielding the weight

$$w_i = \frac{1}{(|x_i| + \varepsilon)^{1-p}}$$

which was proposed by Foucart and Lai (2009), etc.

Example (iii) Let $p \in (0, 1)$. Choose

$$\Phi_\varepsilon(x) = \sum_{i=1}^{n} \log\left(x_i + \varepsilon + (x_i + \varepsilon)^p\right), \quad x \in R^n_+.$$

This merit function yields a reweighted ℓ_1-algorithm with

$$w_i = [\nabla \Phi_\varepsilon(|x|)]_i = \frac{p + (|x_i| + \varepsilon)^{1-p}}{(|x_i| + \varepsilon)^{1-p}\left[|x_i| + \varepsilon + (|x_i| + \varepsilon)^p\right]},$$

which was proposed by Zhao and Li [2012].

(II) Dual reweighted ℓ_1-method: A new design principle

Consider the weighted ℓ_1-minimization:

$$\min_x \{\|Wx\|_1 : \; Ax = b, \; x \geq 0\} \tag{1}$$

where $W = \mathrm{diag}(w)$, $w \in R^n_+$, is given.

Optimal weight.

$w \in R^n_+$ is said to be an optimal weight if the solution of problem (1) is the sparsest point of

$$P = \{x : \; Ax = b, \; x \geq 0\}$$

Question: *How to find the optimal or nearly optimal weight for weighted ℓ_1-method?*

Does an optimal weight exists?

Theorem [Zhao, Kočvara and Luo (2015, 2017)]

Let x^* be a sparsest point in $P = \{x : Ax = b, x \geq 0\}$. Then for any weight $w \in R^n_+$ satisfying

$$w_{J_0(x^*)} > A^T_{J_0(x^*)} A_{supp(x^*)} (A^T_{supp(x^*)} A_{supp(x^*)})^{-1} w_{supp(x^*)}, \qquad (2)$$

x^* is the unique solution to the weighted ℓ_1-problem

$$\min\{\|Wx\|_1 : \ Ax = b, \ x \geq 0\},$$

where $W = diag(w)$ and

$$supp(x^*) = \{i : x^*_i \neq 0\}, J_0(x^*) = \{i : x^*_i = 0\}$$

Dual problem of weighted ℓ_1-minimization

Given $w \in R^n_+$, the dual of the problem

$$\min_x \{\|Wx\|_1 : \; x \in P\} \tag{3}$$

is given as

$$\max_{(y,s)} \left\{ b^T y : \; A^T y + s = w, \; s \geq 0 \right\}. \tag{4}$$

Let $x(w)$ denote an optimal solution of (3), and let $(y(w), s(w))$ the optimal solution of (4).

Strict complementarity [Goldman & Tucker (1956)]

$x(w)$ and $s(w)$ are complementary in the sense that

$$x(w)^T s(w) = 0, \ x(w) \geq 0, \ s(w) \geq 0$$

which implies that

$$\|x(w)\|_0 + \|s(w)\|_0 \leq n, \ \forall \ w \geq 0.$$

The equality can be achieved when $x(w)$ and $s(w)$ are strictly complementary, i.e.,

$$x(w)^T s(w) = 0, \ x(w) \geq 0, \ s(w) \geq 0, \ x(w) + s(w) > 0.$$

The strict complementarity of $s(w)$ and $x(w)$ implies that

$$\|x(w)\|_0 + \|s(w)\|_0 = n.$$

Thus an increase in $\|s(w)\|_0$ leads to a decrease of $\|x(w)\|_0$.

Corollary [Zhao, Kočvara and Luo (2015, 2017)].

If $s(w^*)$ is the densest vector among all possible choices of $w \in R^n_+$, i.e.,

$$s(w^*) = \arg\max\{\|s(w)\|_0 : w \in R^n_+\}. \tag{5}$$

then the corresponding $x(w^*)$ must be the sparsest point of $P = \{x : Ax = b, x \geq 0\}$. Here w^* must be the optimal weight.

Finding optimal weight via bilevel optimization

Theorem [Zhao, Kočvara and Luo (2015, 2017)] *Let (w^*, y^*, s^*) be an optimal solution to the bilevel optimization*

$$
\begin{aligned}
\max_{(w,y,s)} \quad & \|s\|_0 \\
s.t. \quad & b^T y = \gamma, \\
& s = w - A^T y, \\
& w \geq 0, \\
& \gamma = \min_x \{\|Wx\|_1 : x \in P\},
\end{aligned}
$$

where $W = \mathrm{diag}(w)$. Then with $w = w^*$, any optimal solution to the weighted ℓ_1-problem is the sparsest point of P.

New sparsity-seeking principle

- *Seeking sparsity in primal space can be achieved by seeking the maximal density of the associated dual variable in dual space.*
- Finding the densest point in dual space yields the optimal weight for weighted ℓ_1-problem.
- Find optimal weight w^* and associated densest vector $s(w^*)$ can be achieved by tackling a structured bilevel optimization

This connection leads to an entirely new principle for sparse data reconstruction. The ground base of this new principle is the fundamental complementarity theory for linear optimization.

Relaxation of bilevel optimization

- First replacing the objective by $\Phi_\varepsilon(s)$, we get

$$\begin{aligned} \max_{(w,s)} \quad & \Phi_\varepsilon(s) \\ \text{s.t.} \quad & b^T y = \gamma, \\ & s = w - A^T y, \\ & w \geq 0, \\ & \gamma = \min_x \{\|Wx\|_1 : x \in P\}, \end{aligned}$$

- By weak duality of linear programming problem, the condition

$$b^T y = \gamma = \min\{\|Wx\|_1 : x \in P\}$$

can be relaxed to

$$b^T y \leq \gamma.$$

- Note that the scaling of weight does not affect the solution of weighted ℓ_1-minimization. We confine w to a bounded convex set B in order to make the model well-defined, for instance,

$$B = \{w \geq 0 : \ (x^0)^T w \leq M, \ w \leq M^* e\}$$

where x^0 is the solution of the initial weighted ℓ_1-minimization problem.

- So after such processes, we obtain the convex optimization

$$\begin{aligned} \max \quad & \alpha \Phi_\varepsilon(s) + b^T y \\ \text{s.t.} \quad & s = w - A^T y, \ b^T y \leq 1, \ w \in B, \end{aligned}$$

where α is a given parameter.

Algorithms [Zhao and Luo (2017)]

Algorithm Let α, $\varepsilon \in (0,1)$ and $w^0 \in \mathbb{R}^n_{++}$ be a given vector.

- Step 0. Solve $\min\{\|W^0x\|_1 : x \in P\}$ to obtain a minimizer x^0. Set $\gamma^0 = \|W^0x^0\|_1$ and choose M and M^* such that $1 \leq M \leq M^*$ and $M\|w^0\|_\infty/\gamma^0 \leq M^*$.
- Step 1. Solve the convex problem to get a solution $(\widetilde{w}, \widetilde{y}, \widetilde{s})$

$$
\begin{array}{ll}
\max_{(w,y,s)} & b^Ty + \alpha\Phi_\varepsilon(s) \\
s.t. & s = w - A^Ty \geq 0,\ b^Ty \leq 1, \\
 & (x^0)^Tw \leq M,\ 0 \leq w \leq M^*e.
\end{array}
$$

- Step 2. Solve the following problem to obtain a point $\widetilde{x}$

$$
\min\{\|\widetilde{W}x\|_1 : x \in P\}.
$$

A key difference compared to traditional framework

- A key difference between the Algorithm and existing weighted ℓ_1-methods lies in the principle for tackling ℓ_0-minimization.
- Existing weighted ℓ_1-algorithms working in primal space were derived from minimizing nonlinear merit functions for sparsity via linearization, which results in the weight given by the gradient of merit functions at the current iterate.
- The weight $\widehat{w}$ in the dual algorithm can be viewed as an approximate optimal weight generated by convex optimization in dual space.

Theoretical performance under strictly regularity

Definition 4.6. P is said to be *strictly regular* if there is a sparsest point $x \in P$ such that

$$\left\| A^T_{\overline{J_+(x)}} A_{J_+(x)} (A^T_{J_+(x)} A_{J_+(x)})^{-1} \right\|_\infty < 1.$$

Numerical Results

- The experiments have been carried out by realizing the entries of $A \in R^{m \times n}$ and the nonzero entries of the sparse vector x^* from the standard normal distribution.
- For each realized pair $(A, x^* \geq 0)$, we set $b = Ax^*$ and then apply the algorithms to

$$P = \{x : Ax = b, x \geq 0\}$$

to test their performance on reconstructing x^*.
- We take $w^0 = e$ as the initial weight. With this w^0 and a prescribed number $M \geq 1$, we set

$$M^* = M(\max(1, 1/\gamma^0) + 1) \tag{6}$$

which satisfies that $1 \leq M \leq M^*$ and $M^* \geq M\|w^0\|_\infty/\gamma^0$.

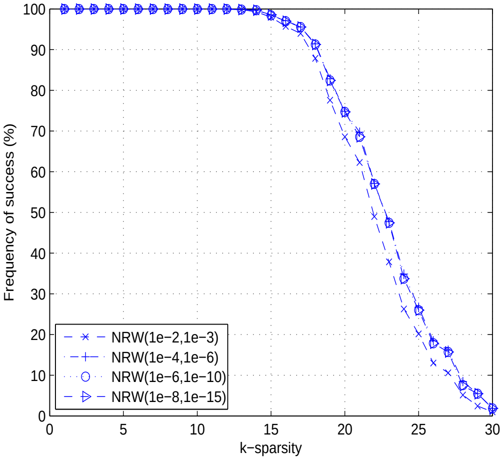

Figure: Comparison of the performance of the NRW algorithm with different (α, ε). The experiments were carried out on polyhedral sets with $A \in R^{40 \times 200}$, and 300 attempts were made for each sparsity level $k = 1, 2, ..., 25$.

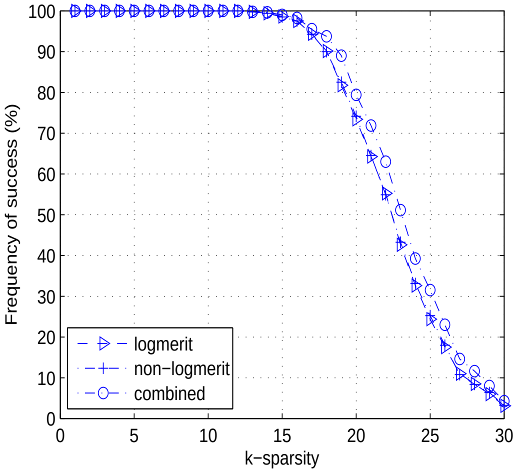

Figure: Performance comparison of the NRW algorithm with different choices of merit functions.

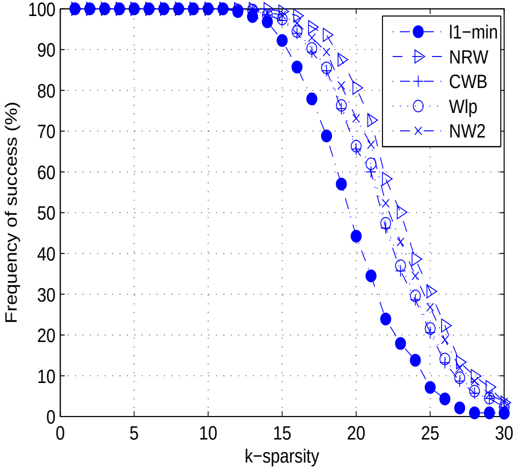

Figure: Comparison of algorithms ($\mu = 10^{-2}$ is taken in CBW, Wlp, and NW2 algorithms).

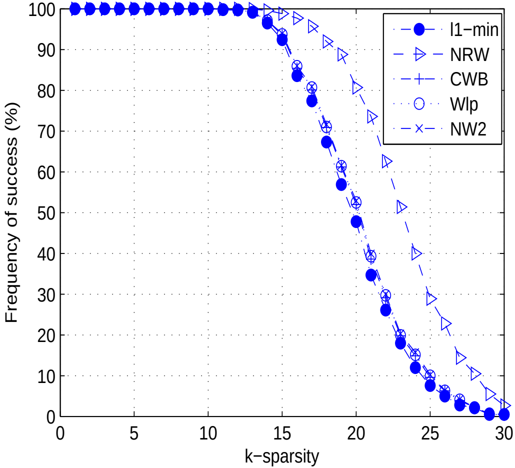

Figure: Comparison of algorithms ($\mu = 10^{-3}$ is taken in CBW, Wlp, and NW2 algorithms).

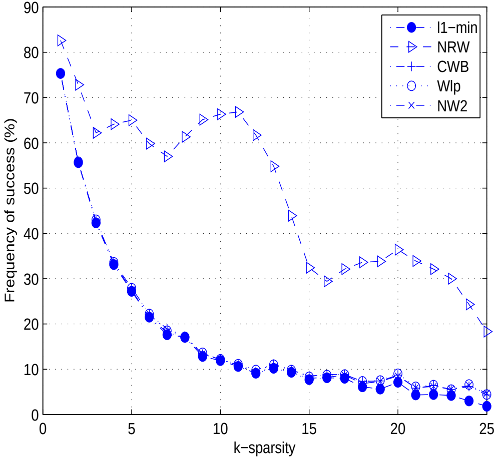

Figure: Comparison of algorithms when P admits multiple sparsest points.

Conclusions

- Seeking the sparsest solutions of linear systems can be achieved by finding the densest slack variable $s \in R^n_+$ of the dual problem of the weighted ℓ_1-problem with all possible choices of $w \in R^n_+$.
- ℓ_0-minimization can be transformed to ℓ_0-maximization with certain bilevel constraints.
- A relaxation of this bilevel optimization leads to a new reweighted ℓ_1-algorithm, going beyond the framework of existing sparsity-seeking methods.
- The simulations indicate that the proposed algorithms outperform ℓ_1-minimization and are comparable to some existing reweighted ℓ_1-algorithms.

Main References

- E.J. Candès, M.B. Wakin and S.P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization,
  J. Fourier Anal. Appl., 14 (2008), pp. 877-905.
- Y.B. Zhao and D. Li, Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems,
  SIAM Journal on Optimization, 22 (2012), 1065–1088.
- Y.-B. Zhao and M. Kocvara, A New Computational Method for the Sparsest Solutions to Systems of Linear Equations,
  SIAM Journal on Optimization, 25 (2015), no.2, pp. 1110–1134
- Y.B. Zhao and Z.Q. Luo, Constructing new reweighted ℓ_1-algorithms for sparsest points of polyhedral sets,
  Mathematics of Operations Research, 42 (2017), No.1, pp. 57–76.)