Nonconvex Sorted ℓ_1 Minimization for Sparse Approximation

Xiaolin Huang (Friedrich-Alexander University Erlangen-Nrnberg)
Lei Shi (Fudan University), and
Ming Yan (Michigan State University, CMSE/Math)

SIAM Conference on Imaging Science
May 25th, 2016

Sparse Approximation

Find $\mathbf{u}$ from

$$\mathbf{Au} = \mathbf{b} \text{ (Noiseless), or } \quad \|\mathbf{Au} - \mathbf{b}\|_2^2 \leq \epsilon \text{ (Noisy)},$$

where $\mathbf{A} \in \mathbf{R}^{m \times n} (m \ll n)$ and $\mathbf{u}$ is sparse.

- ℓ_0-minimization (ideal)

$$\underset{\mathbf{u} \in \mathbf{R}^n}{\text{minimize}} \|\mathbf{u}\|_0, \text{ subject to } \mathbf{Au} = \mathbf{b}$$

$$\underset{\mathbf{u} \in \mathbf{R}^n}{\text{minimize}} \|\mathbf{u}\|_0, \text{ subject to } \|\mathbf{Au} - \mathbf{b}\|_2^2 \leq \epsilon$$

- ℓ_1-minimization (convex relaxation)

$$\underset{\mathbf{u} \in \mathbf{R}^n}{\text{minimize}} \|\mathbf{u}\|_1, \text{ subject to } \mathbf{Au} = \mathbf{b}$$

$$\underset{\mathbf{u} \in \mathbf{R}^n}{\text{minimize}} \|\mathbf{u}\|_1, \text{ subject to } \|\mathbf{Au} - \mathbf{b}\|_2^2 \leq \epsilon$$

$$\min_{\mathbf{u} \in \mathbf{R}^n} \|\mathbf{u}\|_1 + \alpha \|\mathbf{Au} - \mathbf{b}\|_2^2$$

Nonconvex Regularizations

- Separable
  - ℓ_p norm ($0 \le p < 1$): $\|\mathbf{u}\|_p^p$ ($r(u_i) = |u_i|^p$)
  - smoothly clipped absolute deviation (Fan-Li, 2001):

$$r_{\mathsf{SCAD}}(u_i) = \begin{cases} a_1|u_i| & \text{if } |u_i| < a_1, \\ -\frac{a_1|u_i|^2 - 2a_1a_2|u_i| + a_1^3}{2(a_2 - a_1)} & \text{if } a_1 \le |u_i| \le a_2, \\ \frac{a_1a_2 + a_1^2}{2} & \text{if } |u_i| > a_2. \end{cases}$$

  - minimax concave penalty (Zhang, 2010)
  - generalized shrinkage (Chartrand, 2013): no explicit objective function
- Non-separable
  - $\ell_1 - \ell_2$, (Yin-Lou-He-Xin, 2015)
  - $\frac{\ell_1}{\ell_2}$
  - K-support
  - iterative support detection (Wang-Yin, 2010)
  - partial regularization (Lu-Li, 2015)

Algorithms for Nonconvex Problems

- Forward-backward: iterative thresholding
- ADMM for nonconvex problems
- Linearization: One convex problem + one linearization step
- Alternating Minimization: One convex problem + one simple problem

Purpose

- A generalization of several convex and nonconvex regularizations including ℓ_1, K-support, iterative support detection.
- Two algorithms for solving the nonconvex problems: iteratively reweighted ℓ_1 minimization, iterative sorted thresholding
- Showing that stationary points are local minimizers.

Nonconvex Sorted ℓ_1 Minimization

$$R_\lambda(\mathbf{u}) = \lambda_1|\mathbf{u}|_{[1]} + \lambda_2|\mathbf{u}|_{[2]} + \cdots + \lambda_n|\mathbf{u}|_{[n]}, \tag{1}$$

where $|\mathbf{u}|_{[1]} \geq |\mathbf{u}|_{[2]} \geq \cdots \geq |\mathbf{u}|_{[n]}$ are the absolute values ranked in decreasing order and $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

- If $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 1$, it is the ℓ_1 norm of $\mathbf{u}$.
- If $\lambda_1 = \cdots = \lambda_K = 0$ and $\lambda_{K+1} = \cdots = \lambda_n = +\infty$ for some K satisfying $0 < K \leq n$, it is the indicator function for $\{\mathbf{u} : \|\mathbf{u}\|_0 \leq K\}$.
- If $\lambda_1 = \cdots = \lambda_K = 0$ and $\lambda_{K+1} = \cdots = \lambda_n = 1$ for some K satisfying $0 < K < n$, it corresponds to the iterative support detection in (Wang-Yin, 2010).
- If $\lambda_1 = \cdots = \lambda_K = w_1$ and $\lambda_{K+1} = \cdots = \lambda_n = w_2$ for $0 < w_1 < w_2 < \infty$, it is the two-level ℓ_1 "norm" in (Huang et al., 2015).
- If $\lambda_i = w_1 + w_2(i-1)$, where $w_1 \geq 0$ and $w_2 > 0$, it is the small magnitude penalized (SMAP) in (Zeng-Figueiredo, 2014).

Nonconvex Relaxation

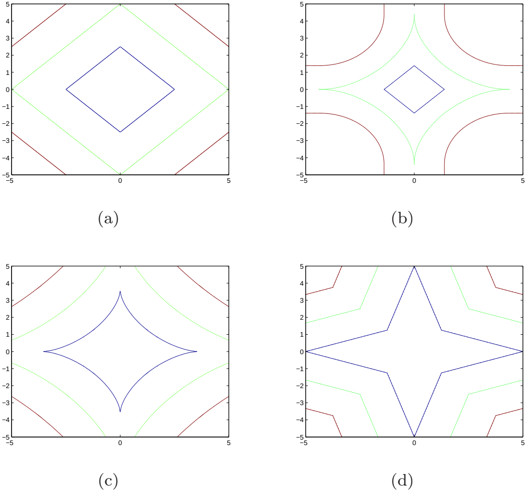


Figure: The contour maps of several penalties: (a) the ℓ_1 norm; (b) the SCAD penalty with $a_1 = 1.1, a_2 = 3.7$; (c) the ℓ_p-norm with $p = 2/3$; (d) the nonconvex sorted ℓ_1 with $\lambda_1 = 1/3$ and $\lambda_2 = 1$.

Properties of Nonconvex Sorted ℓ_1 Norm

$$
\begin{aligned}
F_1(\mathbf{u},\mathbf{P}) &= \sum_{i=1}^{n}(\mathbf{P}\lambda)_i|u_i|,\\
F_2(\mathbf{u},\mathbf{v}) &= \lambda_1\|\mathbf{u}\|_1 + (\lambda_2-\lambda_1)\|\mathbf{u}-\mathbf{v}^1\|_1 + (\lambda_3-\lambda_2)\|\mathbf{u}-\mathbf{v}^1-\mathbf{v}^2\|_1\\
&\quad + \cdots + (\lambda_n-\lambda_{n-1})\|\mathbf{u}-\sum_{j=1}^{n-1}\mathbf{v}^j\|_1,\\
F_3(\mathbf{u},\Lambda) &= \lambda_1\|\mathbf{u}\|_1 + (\lambda_2-\lambda_1)\|\mathbf{u}\odot\Lambda^1\|_1 + (\lambda_3-\lambda_2)\|\mathbf{u}\odot\Lambda^1\odot\Lambda^2\|_1\\
&\quad + \cdots + (\lambda_n-\lambda_{n-1})\|\mathbf{u}\odot\Lambda^1\odot\Lambda^2\odot\cdots\odot\Lambda^{n-1}\|_1.
\end{aligned}
$$

Theorem

Let $\mathcal{P}$ be the set of all permutation matrices.

$$
\begin{aligned}
R_\lambda(\mathbf{u}) = \min_{\mathbf{P}\in\mathcal{P}} F_1(\mathbf{u},\mathbf{P}) &= \min_{\{\mathbf{v}^j\}_{j=1}^{n-1}:\|\mathbf{v}^j\|_0\le 1} F_2(\mathbf{u},\mathbf{v})\\
&= \min_{\{\Lambda^j\}_{j=1}^{n-1}:\Lambda_i^j\in\{0,1\},\sum_i\Lambda_i^j\ge n-1} F_3(\mathbf{u},\Lambda)\\
&= \min_{\{\Lambda^j\}_{j=1}^{n-1}:\Lambda_i^j\in[0,1],\sum_i\Lambda_i^j\ge n-1} F_3(\mathbf{u},\Lambda).
\end{aligned}
$$

Two Lemmas for Convergence

$$\underset{\mathbf{u}}{\text{minimize}} \;\; E(\mathbf{u}) := \;\; R_\lambda(\mathbf{u}) + L(\mathbf{u}), \tag{2}$$

$L(\mathbf{u}) = \iota_{\{\mathbf{u}:\mathbf{Au}=\mathbf{b}\}}(\mathbf{u})$ ($L(\mathbf{u}) = \alpha\|\mathbf{Au}-\mathbf{b}\|_2^2$) for the noiseles (noisy) case.

$$\underset{\mathbf{u},\mathbf{P}}{\text{minimize}} \; E_1(\mathbf{u},\mathbf{P}) := \;\; F_1(\mathbf{u},\mathbf{P}) + L(\mathbf{u}),$$

$$\underset{\mathbf{u},\Lambda}{\text{minimize}} \; E_3(\mathbf{u},\Lambda) := \;\; F_3(\mathbf{u},\Lambda) + L(\mathbf{u}),$$

Lemma

If $\mathbf{u}^*$ is a local minimizer of $E(\mathbf{u})$, then for any $\mathbf{P}^*$ minimizing $F_1(\mathbf{u}^*,\mathbf{P})$, $(\mathbf{u}^*,\Lambda^*)$ with Λ^* being constructed from $\mathbf{P}^*$ is a local minimizer of $E_3(\mathbf{u},\Lambda)$.

Lemma

Given fixed $\mathbf{u}^*$, if for all $\bar{\mathbf{P}} \in \mathcal{P}$ minimizing $E_1(\mathbf{u}^*,\mathbf{P})$, we also have $\mathbf{u}^*$ minimizing $E_1(\mathbf{u},\bar{\mathbf{P}})$, then $\mathbf{u}^*$ is a local minimizer of $E(\mathbf{u})$.

Iteratively Reweighted ℓ_1 Minimization

Algorithm 1 Iteratively Reweighted ℓ_1 Minimization

Initialize λ, $\mathbf{u}^0$

for $l = 0, 1, \cdots$ **do**

Update $\mathbf{P}^l = \arg\min_{\mathbf{P}} F_1(\mathbf{u}^l, \mathbf{P})$ with an optimal $\mathbf{P}$ such that $\mathbf{P}^l$ is different from $\{\mathbf{P}^0, \mathbf{P}^1, \cdots, \mathbf{P}^{l-1}\}$. If there is no optimal $\mathbf{P}$ satisfying this condition, break.

Update $\mathbf{u}^{l+1} = \arg\min_{\mathbf{u}} E_1(\mathbf{u}, \mathbf{P}^l)$.

end for

Theorem

Algorithm 1 will converge in finite steps, and the output $\mathbf{u}^*$ is a local minimizer of $E(\mathbf{u})$. In addition, $(\mathbf{u}^*, \Lambda^*)$ is a local minimizer of $E_3(\mathbf{u}, \Lambda)$.

Iterative Sorted Thresholding

Algorithm 2 Iteratively Sorted Thresholding

Initialize $\mathbf{u}^0$

for $l = 0, 1, \cdots$ **do**

  Find $\mathbf{u}^{l+1} = \arg\min_{\mathbf{u}} \beta R_\lambda(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - (\mathbf{u}^l - \beta \nabla L(\mathbf{u}^l))\|^2$.

end for

Lemma

The proximal operator of R_λ can be evaluated as

$$\textit{prox}_{R_\lambda}(\mathbf{x}) = \max(|\mathbf{x}| - \mathbf{P}\lambda, \mathbf{0}) \odot \operatorname{sign}(\mathbf{x}),$$

for any $\mathbf{P} \in \mathcal{P}$ such that $R_\lambda(\mathbf{x}) = \sum_{i=1}^{n} (\mathbf{P}\lambda)_i x_i$. Here $\max$ *and* sign *are both component-wise.*

- The proximal operator can be multi-valued: $\lambda = (0, 1)$ and $\mathbf{x} = (1, 1)$, then both $(1, 0)$ and $(0, 1)$ are optimal.
- The proximal operator is expansive: $\lambda = (0.5, 1)$, $\mathbf{x}_1 = (2, 1)$ and $\mathbf{x}_2 = (1, 2)$, then $\text{prox}_{R_\lambda}(\mathbf{x}_1) = (1.5, 0)$ and $\text{prox}_{R_\lambda}(\mathbf{x}_2) = (0, 1.5)$.

Iteratively Sorted Thresholding (cont'd)

Lemma

If there exists $\mathbf{u}^*$ *such that*

$$\mathbf{u}^* \in \mathit{prox}_{\beta R_\lambda}(\mathbf{u}^* - \beta \nabla L(\mathbf{u}^*)),$$

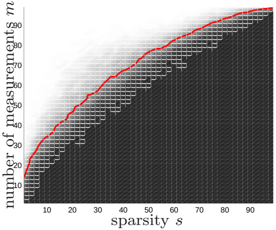
then $\mathbf{u}^*$ *is a local minimizer of* $E(\mathbf{u})$*.*

- $\|\nabla L(\mathbf{u}) - \nabla L(\mathbf{v})\| \leq L_L \|\mathbf{u} - \mathbf{v}\|$

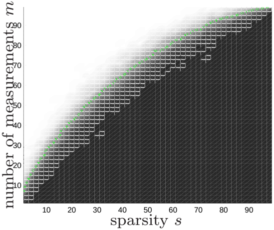
Theorem

If $\beta < 1/L_L$*, the iterative sorted thresholding converges to a local optimum of* (2)*.*

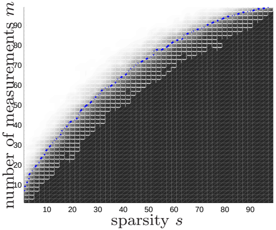
Numerical Experiments



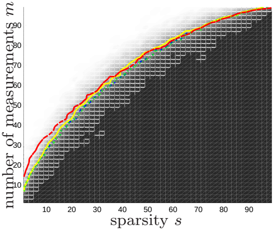
(a)



(b)



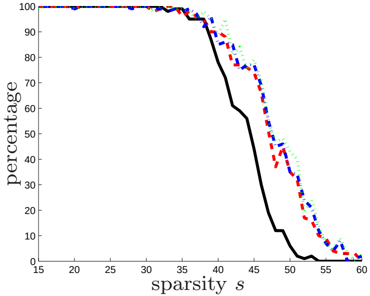
(c)



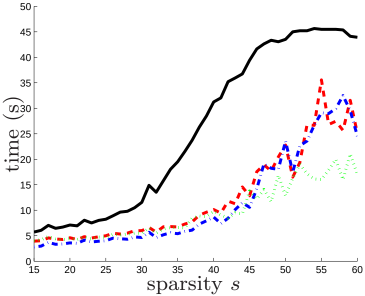
(d)

(a) ISD; (b) 2level; (c) mlevel; (d) IRL1.

Numerical Experiments (cont'd)



(a)



(b)

ISD (red dashed line), 2level (blue dot-dashed line), mlevel (green dotted line), and IRL1 (black solid line):

Conclusion

- nonconvex sorted ℓ_1 generalizes many regularizations
- two algorithms for solving sparse recovery problem with nonconvex sorted ℓ_1
- other algorithms are also available, e.g., ADMM.