A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers

Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright and Bin Yu

Reviewer: Neil Dhingra

M-Estimators

Objective: Recover true θ^* via M-estimator,

$$\hat{\theta} = \arg\min \;\; \mathcal{L}(\theta) \; + \; \lambda_n \mathcal{R}(\theta)$$

M-Estimators

Objective: Recover true θ^* via M-estimator,

$$\hat{\theta} = \arg\min \ \ \mathcal{L}(\theta) \ + \ \lambda_n \mathcal{R}(\theta)$$

Loss function $\mathcal{L}(\theta)$ is *convex* and *differentiable*

M-Estimators

Objective: Recover true θ^* via M-estimator,

$$\hat{\theta} = \arg\min \; \mathcal{L}(\theta) \; + \; \lambda_n \mathcal{R}(\theta)$$

Loss function $\mathcal{L}(\theta)$ is *convex* and *differentiable*

Regularizer $\mathcal{R}(\theta)$ is *decomposable* and a *norm*

M-Estimators

Objective: Recover true θ^* via M-estimator,

$$\hat{\theta} = \arg\min \; \mathcal{L}(\theta) \; + \; \lambda_n \mathcal{R}(\theta)$$

Loss function $\mathcal{L}(\theta)$ is *convex* and *differentiable*

Regularizer $\mathcal{R}(\theta)$ is *decomposable* and a *norm*

Examples

- LASSO: $\mathcal{L} = \|y - X\theta\|_2^2$, $\mathcal{R} = \|\theta\|_1$
- Low-Rank Matrix Approximation: $\mathcal{L} = \|A - \Theta\|_F^2$, $\mathcal{R} = \|\Theta\|_*$ (1)

(1) analytically solvable

Model and Perturbation Subspaces

Partition space of possible θ into

- Model subspace $\mathcal{M} \in \bar{\mathcal{M}}$
- Perturbation subspace $\bar{\mathcal{M}}^{\perp}$

Model and Perturbation Subspaces

Partition space of possible θ into

- Model subspace $\mathcal{M} \in \bar{\mathcal{M}}$
- Perturbation subspace $\bar{\mathcal{M}}^{\perp}$

The true θ^* should lie *mostly* in $\mathcal{M}$

Model and Perturbation Subspaces

Partition space of possible θ into

- Model subspace $\mathcal{M} \in \bar{\mathcal{M}}$
- Perturbation subspace $\bar{\mathcal{M}}^{\perp}$

The true θ^* should lie *mostly* in $\mathcal{M}$

Examples

LASSO - $\mathcal{M}$ is support of a sparse θ^*

Low Rank Matrices - $\mathcal{M}$ is matrices whose singular vectors span same space as those of Θ^*

Decomposability

$\mathcal{R}$ is decomposable with respect to $(\mathcal{M}, \bar{\mathcal{M}}^{\perp})$ if

$$\mathcal{R}(\theta + \gamma) = \mathcal{R}(\theta) + \mathcal{R}(\gamma)$$

$$\begin{aligned} \forall\theta &\in \mathcal{M} \\ \gamma &\in \bar{\mathcal{M}}^{\perp} \end{aligned}$$

Decomposability

$\mathcal{R}$ is decomposable with respect to $(\mathcal{M}, \bar{\mathcal{M}}^{\perp})$ if

$$\mathcal{R}(\theta + \gamma) = \mathcal{R}(\theta) + \mathcal{R}(\gamma)$$

$$\begin{aligned} \forall\theta &\in \mathcal{M} \\ \gamma &\in \bar{\mathcal{M}}^{\perp} \end{aligned}$$

ℓ_1 and nuclear norms are decomposable

Combinations of these norms also decomposable

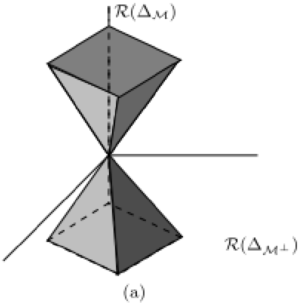
Lemma 1

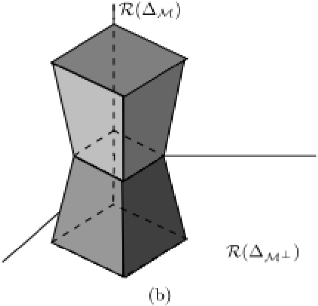
If $\lambda_n \geq \mathcal{R}^*(\nabla \mathcal{L}(\theta^*))$, the estimation error will lie in a certain region

$$\hat{\theta} - \theta^* \in \mathbb{C}(\mathcal{M}, \bar{\mathcal{M}}^{\perp}; \theta^*)$$

If $\theta^* \in \mathcal{M}$

$$\mathbb{C}(\mathcal{M}, \bar{\mathcal{M}}^{\perp}; \theta^*) := \left\{ \Delta \,\middle|\, \mathcal{R}(\Delta_{\bar{\mathcal{M}}^{\perp}}) \leq 3\mathcal{R}(\Delta_{\bar{\mathcal{M}}}) \right\}$$





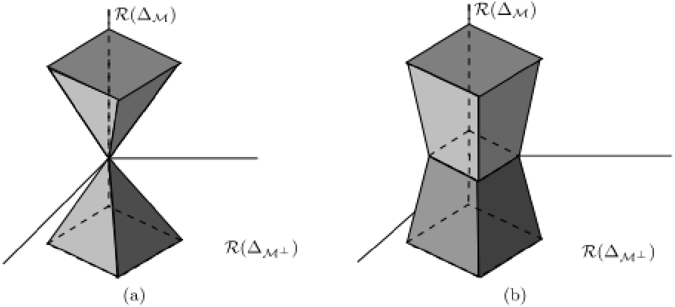
Lemma 1

If $\lambda_n \geq \mathcal{R}^*(\nabla \mathcal{L}(\theta^*))$, the estimation error will lie in a certain region

$$\hat{\theta} - \theta^* \;\in\; \mathbb{C}(\mathcal{M}, \bar{\mathcal{M}}^\perp; \theta^*)$$

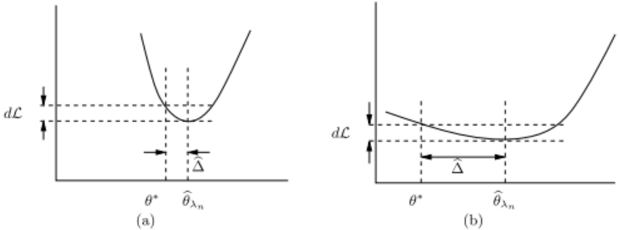
If $\theta^* \notin \mathcal{M}$

$$\mathbb{C}(\mathcal{M}, \bar{\mathcal{M}}^\perp; \theta^*) \;:=\; \left\{ \Delta \;\middle|\; \mathcal{R}(\Delta_{\bar{\mathcal{M}}^\perp}) \;\leq\; 3\mathcal{R}(\Delta_{\bar{\mathcal{M}}}) + 4\mathcal{R}(\theta^*_{\mathcal{M}^\perp}) \right\}$$



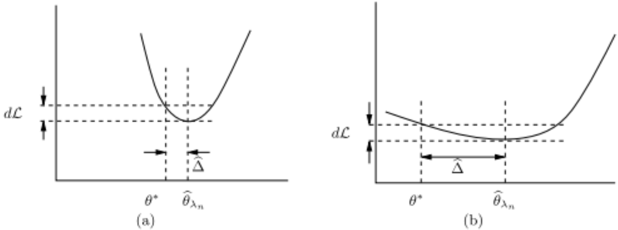
Restricted Strong Convexity

Strength of $\mathcal{L}$ convexity important



Restricted Strong Convexity

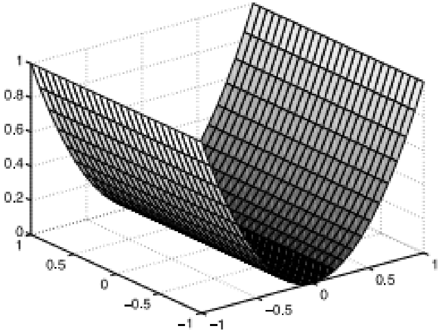
Strength of $\mathcal{L}$ convexity important



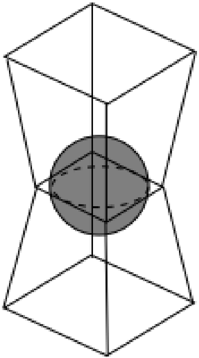
Only need strong convexity WITHIN $\mathbb{C}$

$$\mathcal{L}(\theta^* + \Delta) - \Big(\mathcal{L}(\theta^*) + \langle \nabla \mathcal{L}(\theta^*), \Delta \rangle\Big) \geq \mathcal{K}_{\mathcal{L}} \, \|\Delta\|^2 - \tau_{\mathcal{L}}^2 \, (\theta^*)$$

$$\forall \Delta \in \mathbb{C}(\mathcal{M}, \bar{\mathcal{M}}^{\perp}; \theta^*)$$



(a)



(b)

Subspace Compatibility Constant

$$\Psi(\mathcal{M}) \; := \; \sup_{u \in \mathcal{M} \setminus \{0\}} \frac{\mathcal{R}(u)}{\|u\|}$$

Subspace Compatibility Constant

$$\Psi(\mathcal{M}) := \sup_{u \in \mathcal{M} \setminus \{0\}} \frac{\mathcal{R}(u)}{\|u\|}$$

Example:

$$\begin{aligned} u &\in \mathbb{R}^p \\ \mathcal{R}(u) &= \|u\|_1 \\ \|u\| &= \|u\|_2 \\ \Psi(\mathcal{M}) &= \sqrt{p} \end{aligned}$$

Theorem 1

When $\lambda_n \geq 2\mathcal{R}^*\big(\nabla\mathcal{L}(\theta^*)\big)$,

$$\|\hat{\theta}_{\lambda_n} - \theta^*\|^2 \;\leq\; 9\,\frac{\lambda_n^2}{\mathcal{K}_{\mathcal{L}}^2}\,\Psi^2(\bar{\mathcal{M}}) \;+\; \frac{\lambda_n}{\mathcal{K}_{\mathcal{L}}}\Big(2\,\tau_{\mathcal{L}}^2(\,\theta^*\,) \;+\; 4\,\mathcal{R}(\,\theta^*_{\bar{\mathcal{M}}^\perp}\,)\Big)$$

Depends on subspace, strength of convexity, and regularization parameter

Theorem 1

When $\lambda_n \geq 2\mathcal{R}^*\big(\nabla\mathcal{L}(\theta^*)\big)$,

$$\|\hat{\theta}_{\lambda_n} - \theta^*\|^2 \;\leq\; 9\,\frac{\lambda_n^2}{\mathcal{K}_{\mathcal{L}}^2}\,\Psi^2(\bar{\mathcal{M}}) \;+\; \frac{\lambda_n}{\mathcal{K}_{\mathcal{L}}}\Big(2\,\tau_{\mathcal{L}}^2(\theta^*) + 4\,\mathcal{R}(\theta^*_{\bar{\mathcal{M}}^\perp})\Big)$$

Depends on subspace, strength of convexity, and regularization parameter

Corollary 1: $\theta^* \in \mathcal{M}$

$$\|\hat{\theta}_{\lambda_n} - \theta^*\| \;\leq\; 3\,\frac{\lambda_n}{\mathcal{K}_{\mathcal{L}}}\,\Psi(\bar{\mathcal{M}})$$

$$\mathcal{R}(\hat{\theta}_{\lambda_n} - \theta^*) \;\leq\; 12\,\frac{\lambda_n}{\mathcal{K}_{\mathcal{L}}}\,\Psi^2(\bar{\mathcal{M}})$$

Conclusions

Decomposable Regularizer

- error vector lies in a cone or a star

Conclusions

Decomposable Regularizer

- error vector lies in a cone or a star

Restricted Strong Convexity yields error bound

- strength of restricted convexity
- chosen subspace
  - how well it captures θ^*
  - subspace compatibility constant

Questions?