Lower-order Regularization for Sparse Optimization with Applications

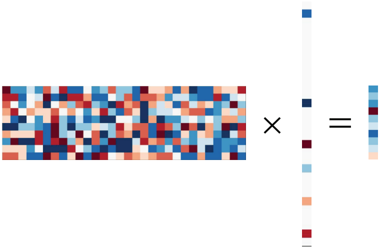
Xiaoqi YANG
Department of Applied Mathematics,
Hong Kong Polytechnic University (mayangxq@polyu.edu.hk)

Joint with
Yaohua HU (SZU), Chong LI (ZJU),
Kaiwen MENG (SWJTU), Jing QIN(SZU)

Outline

In many applications, the underlying data usually can be represented approximately by a linear system

$$Ax = b + \varepsilon.$$



The sparse optimization problem can be modeled as

$$\begin{aligned} \min \quad & \|x\|_0 \\ \text{s.t.} \quad & \|Ax - b\|_2 \le \epsilon. \end{aligned}$$

The ℓ_q regularization model $(0 \le q \le 1)$:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|x\|_q^q, \text{ where } \|x\|_q = \left(\sum_{i=1}^n |x_i|^q \right)^{1/q}.$$

Main questions:

1. How far is the solution of the regularization problem from that of the original sparse optimization problem?
2. How to design the efficient numerical algorithms for the ℓ_1 regularization problem?
3. How to employ the sparse optimization technique to application fields.

The sparse optimization problem can be modeled as

$$\begin{aligned} \min \quad & \|x\|_0 \\ \text{s.t.} \quad & \|Ax - b\|_2 \leq \epsilon. \end{aligned}$$

The ℓ_q regularization model $(0 \leq q \leq 1)$:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|x\|_q^q, \text{ where } \|x\|_q = \left(\sum_{i=1}^n |x_i|^q \right)^{1/q}.$$

Main questions:

1. How far is the solution of the regularization problem from that of the original sparse optimization problem?
2. How to design the efficient numerical algorithms for the ℓ_1 regularization problem?
3. How to employ the sparse optimization technique to application fields.

The sparse optimization problem can be modeled as

$$\begin{aligned} \min \quad & \|x\|_0 \\ \text{s.t.} \quad & \|Ax - b\|_2 \le \epsilon. \end{aligned}$$

The ℓ_q regularization model $(0 \le q \le 1)$:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|x\|_q^q, \text{ where } \|x\|_q = \left(\sum_{i=1}^n |x_i|^q \right)^{1/q}.$$

Main questions:

1. How far is the solution of the regularization problem from that of the original sparse optimization problem?
2. How to design the efficient numerical algorithms for the ℓ_1 regularization problem?
3. How to employ the sparse optimization technique to application fields.

The sparse optimization problem can be modeled as

$$\begin{array}{ll} \min & \|x\|_0 \\ \text{s.t.} & \|Ax-b\|_2 \leq \epsilon. \end{array}$$

The ℓ_q regularization model $(0 \leq q \leq 1)$:

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 + \lambda\|x\|_q^q, \text{ where } \|x\|_q = \left(\sum_{i=1}^n |x_i|^q\right)^{1/q}.$$

Main questions:

1. How far is the solution of the regularization problem from that of the original sparse optimization problem?
2. How to design the efficient numerical algorithms for the ℓ_1 regularization problem?
3. How to employ the sparse optimization technique to application fields.

The sparse optimization problem can be modeled as

$$\begin{array}{ll} \min & \|x\|_0 \\ \text{s.t.} & \|Ax - b\|_2 \leq \epsilon. \end{array}$$

The ℓ_q regularization model $(0 \leq q \leq 1)$:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|x\|_q^q, \text{ where } \|x\|_q = \left(\sum_{i=1}^n |x_i|^q \right)^{1/q}.$$

Main questions:

1. How far is the solution of the regularization problem from that of the original sparse optimization problem?
2. How to design the efficient numerical algorithms for the ℓ_1 regularization problem?
3. How to employ the sparse optimization technique to application fields.

- The ℓ_1 regularization model has attracted much attention and has been accepted as a most useful tool for the sparse optimization problem, which is widely applied in compressive sensing, image science, machine learning, system biology, etc.

- Recovery bound for ℓ_1 regularization:

$$\|x^*(\ell_1) - \bar{x}\|_2^2 = O(\lambda^2 s),$$

under the assumption of restricted isometry property (RIP) or restricted eigenvalue condition (REC).

- The ℓ_1 regularization model has attracted much attention and has been accepted as a most useful tool for the sparse optimization problem, which is widely applied in compressive sensing, image science, machine learning, system biology, etc.

- Recovery bound for ℓ_1 regularization:

$$\|x^*(\ell_1) - \bar{x}\|_2^2 = O(\lambda^2 s),$$

under the assumption of restricted isometry property (RIP) or restricted eigenvalue condition (REC).

Existing Algorithms

ℓ_1 regularization model:

- ℓ_1 Magic [Candes, Romberg and Tao 2006]
- LARs [Efron, Hastie, Johnstone and Tibshirani 2004]
- GPSR and SpaRSA [Figueiredo, Nowak and Wright 2007,2009]
- ISTA [Daubechies, Defrise and De Mol 2004], APG [Nesterov 2013], FISTA [Beck and Teboulle 2009], PGH [Xiao and Zhang 2013]
- ADMM [Yang and Zhang 2011; He and Yuan 2012,2013]

ℓ_q regularization model:

- Reweight scheme [Chartrand 2007,2008; Lai and Yin 2012,2013]
- Smoothing approach [Chen and Ye, 2010,2012]
- IHTA (Half) [Xu et al. 2012]

ℓ_0 regularization model:

- Reweight ℓ_1 [Candes, Wakin and Boyd 2008]
- IHTA (Hard) [Blumensath and Davies 2008,2009]

Existing Algorithms

ℓ_1 regularization model:

- ℓ_1 Magic [Candes, Romberg and Tao 2006]
- LARs [Efron, Hastie, Johnstone and Tibshirani 2004]
- GPSR and SpaRSA [Figueiredo, Nowak and Wright 2007,2009]
- ISTA [Daubechies, Defrise and De Mol 2004], APG [Nesterov 2013], FISTA [Beck and Teboulle 2009], PGH [Xiao and Zhang 2013]
- ADMM [Yang and Zhang 2011; He and Yuan 2012,2013]

ℓ_q regularization model:

- Reweight scheme [Chartrand 2007,2008; Lai and Yin 2012,2013]
- Smoothing approach [Chen and Ye, 2010,2012]
- IHTA (Half) [Xu et al. 2012]

ℓ_0 regularization model:

- Reweight ℓ_1 [Candes, Wakin and Boyd 2008]
- IHTA (Hard) [Blumensath and Davies 2008,2009]

Existing Algorithms

ℓ_1 regularization model:

- ℓ_1 Magic [Candès, Romberg and Tao 2006]
- LARs [Efron, Hastie, Johnstone and Tibshirani 2004]
- GPSR and SpaRSA [Figueiredo, Nowak and Wright 2007,2009]
- ISTA [Daubechies, Defrise and De Mol 2004], APG [Nesterov 2013], FISTA [Beck and Teboulle 2009], PGH [Xiao and Zhang 2013]
- ADMM [Yang and Zhang 2011; He and Yuan 2012,2013]

ℓ_q regularization model:

- Reweight scheme [Chartrand 2007,2008; Lai and Yin 2012,2013]
- Smoothing approach [Chen and Ye, 2010,2012,2014]
- IHTA (Half) [Xu et al. 2012]

ℓ_0 regularization model:

- Reweight ℓ_1 [Candes, Wakin and Boyd 2008]
- IHTA (Hard) [Blumensath and Davies 2008,2009]

Lower Order Regularization

ℓ_q regularization:

- [Chartrand and Staneva 2007,2008]: a weaker RIP is sufficient to guarantee perfect recovery;
- [Xu, Chang, Xu, and Zhang. 2012]: admits a significantly stronger sparsity promoting capability;
- [Qin, Hu, Xu, Yalamanchili, and Wang 2014]: achieves a more reliable solution in biological sense.

Our objectives:

- The recovery bound for the ℓ_q regularization model.
- Linear convergence of numerical algorithm.
- Application to structured sparse optimization and real applications.

Lower Order Regularization

ℓ_q regularization:

- [Chartrand and Staneva 2007,2008]: a weaker RIP is sufficient to guarantee perfect recovery;
- [Xu, Chang, Xu, and Zhang. 2012]: admits a significantly stronger sparsity promoting capability;
- [Qin, Hu, Xu, Yalamanchili, and Wang 2014]: achieves a more reliable solution in biological sense.

Our objectives:

- The recovery bound for the ℓ_q regularization model.
- Linear convergence of numerical algorithm.
- Application to structured sparse optimization and real applications.

The ℓ_q regularization model $(0 \le q \le 1)$:

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 + \lambda\|x\|_q^q, \text{ where } \|x\|_q = \left(\sum_{i=1}^n |x_i|^q\right)^{1/q}.$$

$\mathcal{I}(x;t)$: the subset of $\{1,\ldots,n\}$ corresponding to the first t largest coordinates in absolute value of x in $\mathcal{I}^c$.

Definition (REC, Bickel, Ritov and Tsybakov 2009)

The restricted eigenvalue condition relative to (s,t) ($\mathrm{REC}(s,t)$) is said to be satisfied if

$$\phi(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_\mathcal{T}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_1 \le \|x_\mathcal{I}\|_1, \mathcal{T} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$

Definition (q-REC)

The q-restricted eigenvalue condition relative to (s,t) (q-$\mathrm{REC}(s,t)$) is said to be satisfied if

$$\phi_q(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_\mathcal{T}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_q \le \|x_\mathcal{I}\|_q, \mathcal{T} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$

The ℓ_q regularization model ($0 \le q \le 1$):

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 + \lambda\|x\|_q^q, \text{ where } \|x\|_q = \left(\sum_{i=1}^n |x_i|^q\right)^{1/q}.$$

$\mathcal{I}(x;t)$: the subset of $\{1,\dots,n\}$ corresponding to the first t largest coordinates in absolute value of x in $\mathcal{I}^c$.

Definition (REC, Bickel, Ritov and Tsybakov 2009)

The restricted eigenvalue condition relative to (s,t) ($\mathrm{REC}(s,t)$) is said to be satisfied if

$$\phi(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_{\mathcal{T}}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_1 \le \|x_{\mathcal{I}}\|_1, \mathcal{T} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$

Definition (q-REC)

The q-restricted eigenvalue condition relative to (s,t) (q-$\mathrm{REC}(s,t)$) is said to be satisfied if

$$\phi_q(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_{\mathcal{T}}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_q \le \|x_{\mathcal{I}}\|_q, \mathcal{T} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$

The ℓ_q regularization model ($0 \le q \le 1$):

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 + \lambda\|x\|_q^q, \text{ where } \|x\|_q = \left(\sum_{i=1}^n |x_i|^q\right)^{1/q}.$$

$\mathcal{I}(x;t)$: the subset of $\{1,\dots,n\}$ corresponding to the first t largest coordinates in absolute value of x in $\mathcal{I}^c$.

Definition (REC, Bickel, Ritov and Tsybakov 2009)

The restricted eigenvalue condition relative to (s,t) ($\mathrm{REC}(s,t)$) is said to be satisfied if

$$\phi(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_\mathcal{T}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_1 \le \|x_\mathcal{I}\|_1, \mathcal{T} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$

Definition (q-REC)

The q-restricted eigenvalue condition relative to (s,t) (q-$\mathrm{REC}(s,t)$) is said to be satisfied if

$$\phi_q(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_\mathcal{T}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_q \le \|x_\mathcal{I}\|_q, \mathcal{T} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$

The ℓ_q regularization model $(0 \le q \le 1)$:

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 + \lambda\|x\|_q^q, \text{ where } \|x\|_q = \Big(\sum_{i=1}^n |x_i|^q\Big)^{1/q}.$$

$\mathcal{I}(x;t)$: the subset of $\{1,\ldots,n\}$ corresponding to the first t largest coordinates in absolute value of x in $\mathcal{I}^c$.

Definition (REC, Bickel, Ritov and Tsybakov 2009)

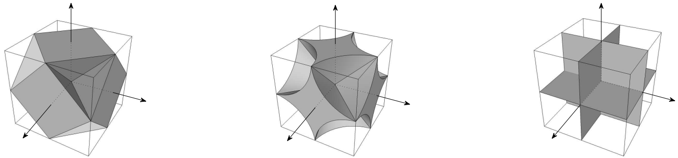
The restricted eigenvalue condition relative to (s,t) ($\mathrm{REC}(s,t)$) is said to be satisfied if

$$\phi(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_{\mathcal{T}}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_1 \le \|x_{\mathcal{I}}\|_1, \mathcal{T} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$

Definition (q-REC)

The q-restricted eigenvalue condition relative to (s,t) (q-$\mathrm{REC}(s,t)$) is said to be satisfied if

$$\phi_q(s,t) := \min\left\{\frac{\|Ax\|_2}{\|x_{\mathcal{T}}\|_2} : |\mathcal{I}| \le s, \|x_{\mathcal{I}^c}\|_q \le \|x_{\mathcal{I}}\|_q, \mathcal{T} = \mathcal{I}(x;t)\cup\mathcal{I}\right\} > 0.$$



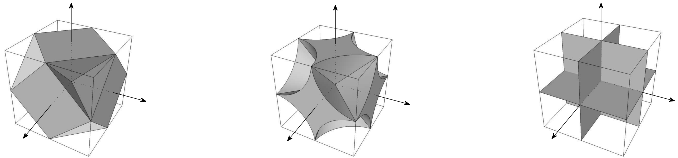
(a) REC (b) 1/2-REC (c) 0-REC

Figure 1: The geometric interpretation of the RECs: the q-REC holds if and only if the null space of A does not intersect the gray region.

Proposition

Let $0 \leq q_1 \leq q_2 \leq 1$. Then

$$q_2\text{-REC}(s,t) \quad \Rightarrow \quad q_1\text{-REC}(s,t).$$



(a) REC (b) 1/2-REC (c) 0-REC

Figure 1: The geometric interpretation of the RECs: the q-REC holds if and only if the null space of A does not intersect the gray region.

Proposition

Let $0 \leq q_1 \leq q_2 \leq 1$. *Then*

$$q_2\text{-REC}(s,t) \quad \Rightarrow \quad q_1\text{-REC}(s,t).$$

Theorem (Oracle Inequality and Global Recovery Bound)

Notations:

- $0 \le q \le 1$, $A\bar{x} = b$, $\mathcal{S} := \mathrm{supp}(\bar{x})$, $s := |\mathrm{supp}(\bar{x})|$;
- x^* *be a global minimum of the* ℓ_q *regularization problem,* K *be the smallest integer such that* $2^{K-1}q \ge 1$.

Assumptions:

- q-$\mathrm{REC}(s, s)$ *holds.*

Conclusions:

- *oracle inequality:*

$$\|Ax^* - A\bar{x}\|_2^2 + \lambda \|x^*_{\mathcal{S}^c}\|_q^q \le \lambda^{\frac{2}{2-q}} s^{(1-2^{-K})\frac{2}{2-q}} / \phi_q^{\frac{2q}{2-q}}(s, s) = O(\lambda^{\frac{2}{2-q}} s),$$

- *global recovery bound:*

$$\|x^* - \bar{x}\|_2^2 \le 2\lambda^{\frac{2}{2-q}} s^{\frac{q-2}{q} + (1-2^{-K})\frac{4}{q(2-q)}} / \phi_q^{\frac{4}{2-q}}(s, s) = O(\lambda^{\frac{2}{2-q}} s).$$

Theorem (Oracle Inequality and Global Recovery Bound)

Notations:

- $0 \le q \le 1$, $A\bar{x} = b$, $\mathcal{S} := \mathrm{supp}(\bar{x})$, $s := |\mathrm{supp}(\bar{x})|$;
- x^* *be a global minimum of the* ℓ_q *regularization problem,* K *be the smallest integer such that* $2^{K-1}q \ge 1$.

Assumptions:

- q-$\mathrm{REC}(s, s)$ *holds.*

Conclusions:

- *oracle inequality:*

$$\|Ax^* - A\bar{x}\|_2^2 + \lambda \|x^*_{\mathcal{S}^c}\|_q^q \le \lambda^{\frac{2}{2-q}} s^{(1-2^{-K})\frac{2}{2-q}} / \phi_q^{\frac{2q}{2-q}}(s, s) = O(\lambda^{\frac{2}{2-q}} s),$$

- *global recovery bound:*

$$\|x^* - \bar{x}\|_2^2 \le 2\lambda^{\frac{2}{2-q}} s^{\frac{q-2}{q} + (1-2^{-K})\frac{4}{q(2-q)}} / \phi_q^{\frac{4}{2-q}}(s, s) = O(\lambda^{\frac{2}{2-q}} s).$$

Example

$$A = \begin{pmatrix} 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

This A satisfies the 1/2-REC(1,1), but not REC(1,1).

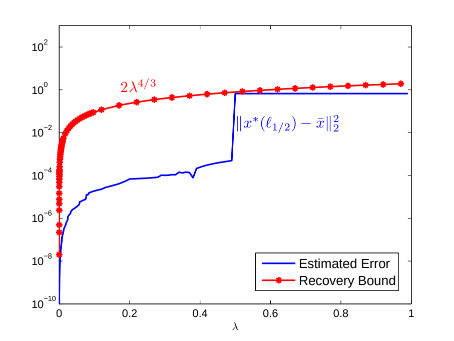


Figure 2: The illustration of the recovery bound and estimated error.

Let $\bar{x} = (\bar{x}_{\mathcal{S}}, 0)$.

- to construct a smooth path by applying an implicit theorem to the function $H : \mathbb{R}^{s+1} \to \mathbb{R}^s$:

$$H(z, \lambda) = 2A_{\mathcal{S}}^{\top}(A_{\mathcal{S}} z - b) + \lambda q \begin{pmatrix} |z_1|^{q-1}\text{sign}(z_1) \\ \vdots \\ |z_s|^{q-1}\text{sign}(z_s) \end{pmatrix}.$$

- to apply following first-order growth condition:

$$\|A_{\mathcal{S}}^c y\|_2^2 + 2\langle A_{\mathcal{S}}\xi(\lambda) - b, A_{\mathcal{S}}^c y\rangle - 2\epsilon_0\|y\|_2^2 + \lambda\|y\|_q^q \geq \epsilon\|y\|_2 \quad \forall y \in \mathbf{B}(0, \delta).$$

to show that the path is a local optimal one.

- to verify

$$\|x^*(\lambda) - \bar{x}\|_2^2 = \|\xi(\lambda) - \bar{x}_{\mathcal{S}}\|_2^2 \leq \lambda^2 q^2 s \|(A_{\mathcal{S}}^{\top} A_{\mathcal{S}})^{-1}\|^2 \max_{\bar{x}_i \neq 0} \left(|\bar{x}_i|^{2(q-1)}\right).$$

via a first-order condition.

Let $\bar{x} = (\bar{x}_{\mathcal{S}}, 0)$.

- to construct a smooth path by applying an implicit theorem to the function $H : \mathbb{R}^{s+1} \to \mathbb{R}^s$:

$$H(z,\lambda) = 2A_{\mathcal{S}}^{\top}(A_{\mathcal{S}}z - b) + \lambda q \begin{pmatrix} |z_1|^{q-1}\mathrm{sign}(z_1) \\ \vdots \\ |z_s|^{q-1}\mathrm{sign}(z_s) \end{pmatrix}.$$

- to apply following first-order growth condition:

$$\|A_{\mathcal{S}}^c y\|_2^2 + 2\langle A_{\mathcal{S}}\xi(\lambda) - b, A_{\mathcal{S}}^c y\rangle - 2\epsilon_0\|y\|_2^2 + \lambda\|y\|_q^q \geq \epsilon\|y\|_2 \quad \forall y \in \mathbf{B}(0,\delta).$$

  to show that the path is a local optimal one.

- to verify

$$\|x^*(\lambda) - \bar{x}\|_2^2 = \|\xi(\lambda) - \bar{x}_{\mathcal{S}}\|_2^2 \leq \lambda^2 q^2 s \|(A_{\mathcal{S}}^{\top}A_{\mathcal{S}})^{-1}\|^2 \max_{\bar{x}_i \neq 0}\left(|\bar{x}_i|^{2(q-1)}\right).$$

  via a first-order condition.

Let $\bar{x} = (\bar{x}_{\mathcal{S}}, 0)$.

- to construct a smooth path by applying an implicit theorem to the function $H : \mathbb{R}^{s+1} \to \mathbb{R}^s$:

$$H(z,\lambda) = 2A_{\mathcal{S}}^{\top}(A_{\mathcal{S}}z - b) + \lambda q \begin{pmatrix} |z_1|^{q-1}\mathrm{sign}(z_1) \\ \vdots \\ |z_s|^{q-1}\mathrm{sign}(z_s) \end{pmatrix}.$$

- to apply following first-order growth condition:

$$\|A_{\mathcal{S}}^c y\|_2^2 + 2\langle A_{\mathcal{S}}\xi(\lambda) - b, A_{\mathcal{S}}^c y\rangle - 2\epsilon_0\|y\|_2^2 + \lambda\|y\|_q^q \geq \epsilon\|y\|_2 \quad \forall y \in \mathbf{B}(0,\delta).$$

  to show that the path is a local optimal one.

- to verify

$$\|x^*(\lambda) - \bar{x}\|_2^2 = \|\xi(\lambda) - \bar{x}_{\mathcal{S}}\|_2^2 \leq \lambda^2 q^2 s \|(A_{\mathcal{S}}^{\top}A_{\mathcal{S}})^{-1}\|^2 \max_{\bar{x}_i \neq 0}\left(|\bar{x}_i|^{2(q-1)}\right).$$

  via a first-order condition.

Theorem (Local Recovery Bound)

Notations:

- $0 < q < 1$, $A\bar{x} = b$, $\mathcal{S} := \mathrm{supp}(\bar{x})$.

Assumptions:

- *The columns of $A_{\mathcal{S}}$ are linearly independent.*

Conclusion:

- *there exist $\kappa > 0$ and a path of local minima of the l_q regularization problem, $x^*(\lambda)$, such that, for $\lambda < \kappa$,*

$$\|x^*(\lambda) - \bar{x}\|_2^2 \le \lambda^2 q^2 s \|(A_{\mathcal{S}}^\top A_{\mathcal{S}})^{-1}\|^2 \max_{\bar{x}_i \neq 0} \left(|\bar{x}_i|^{2(q-1)} \right) = O(\lambda^2 s).$$

Theorem (Local Recovery Bound)

Notations:

- $0 < q < 1$, $A\bar{x} = b$, $\mathcal{S} := \mathrm{supp}(\bar{x})$.

Assumptions:

- *The columns of $A_{\mathcal{S}}$ are linearly independent.*

Conclusion:

- *there exist $\kappa > 0$ and a path of local minima of the l_q regularization problem, $x^*(\lambda)$, such that, for $\lambda < \kappa$,*

$$\|x^*(\lambda) - \bar{x}\|_2^2 \leq \lambda^2 q^2 s \|(A_{\mathcal{S}}^\top A_{\mathcal{S}})^{-1}\|^2 \max_{\bar{x}_i \neq 0} \left(|\bar{x}_i|^{2(q-1)} \right) = O(\lambda^2 s).$$

The nonsmooth composite optimization problem

$$\min_{x\in\mathbb{R}^n} F(x) := f(x) + \phi(x),$$

Proximal gradient algorithm (PGA):

$$\begin{aligned} z^k &= x^k - v\nabla f(x^k), \\ x^{k+1} &\in \operatorname{Arg}\min_{x\in\mathbb{R}^n} \{\phi(x) + \frac{1}{2v}\|x - z^k\|_2^2\}. \end{aligned}$$

Convex composite optimization: ISTA, APG, FISTA, PGH.

Nonconvex composite optimization:

- Kurdyka-Łojasewicz (KL) theory [Bolte, Sabach and Teboulle 2013]
- majorization-minimization (MM) scheme [Mairal 2013]
- coordinate gradient descent (CGD) method [Tseng and Yun 2009]
- successive upper-bound minimization (SUM) approach [Razaviyayn, Hong and Luo 2013]

The nonsmooth composite optimization problem

$$\min_{x\in\mathbb{R}^n} F(x) := f(x) + \phi(x),$$

Proximal gradient algorithm (PGA):

$$\begin{aligned} z^k &= x^k - v\nabla f(x^k), \\ x^{k+1} &\in \operatorname{Arg}\min_{x\in\mathbb{R}^n}\{\phi(x) + \frac{1}{2v}\|x - z^k\|_2^2\}. \end{aligned}$$

Convex composite optimization: ISTA, APG, FISTA, PGH.
Nonconvex composite optimization:

- Kurdyka-Łojasewicz (KL) theory [Bolte, Sabach and Teboulle 2013]
- majorization-minimization (MM) scheme [Mairal 2013]
- coordinate gradient descent (CGD) method [Tseng and Yun 2009]
- successive upper-bound minimization (SUM) approach [Razaviyayn, Hong and Luo 2013]

The ℓ_q regularization problem

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2+\lambda\|x\|_q^q.$$

Theorem (Global Convergence of PGA)

Let $\{x^k\}$ be a sequence generated by the PGA with $v<\frac{1}{2}\|A\|_2^{-2}$. Then the following statements hold:

(i) *if $q=1$, then $\{x^k\}$ converges to a global minimizer of the ℓ_1 regularization problem,*

(ii) *if $q=0$, then $\{x^k\}$ converges to a local minimizer of the ℓ_0 regularization problem,*

(iii) *if $0<q<1$, then $\{x^k\}$ converges to a critical point of the ℓ_q regularization problem.*

The linear convergence of PGA for solving the ℓ_1 regularization:

[Hale, Yin and Zhang 2008] under one of the assumptions:

- $A|_J$ is injective; or
- Strict complementarity condition (SCC): $\mathrm{supp}(x^*) = J$,

where

$$J := \{k \in \mathbb{N} : |(A^\top (Ax^* - b))_k| = \frac{\lambda}{2}\}.$$

[Bredies and Lorenz 2008] for infinite-dimensional Hilbert spaces.

Lemma (second-order sufficient condition and second-order growth condition)

Notations: $x^* \in R^n \setminus \{0\}$; $I = supp(x^*)$.

Conclusion: the following statements are equivalent:

- x^* is a local minimum of the l_q regularization problem;
- the following first- and second-order conditions hold:

$$2A_I^\top (A_I x_I^* - b) + \lambda q \left((|x_i^*|^{q-1} sign(x_i^*))_{i \in I} \right) = 0,$$

$$2A_I^\top A_I + \lambda q(q-1) diag \left(|x_i^*|^{q-2} \right)_{i \in I} \succ 0; \quad (1)$$

- the second-order growth condition holds at x^*:

$$F(x) \geq F(x^*) + \varepsilon \|x - x^*\|_2^2 \quad \text{for any } x \in B(x^*, \delta).$$

Remarks: Second-order necessary condition (1) (with $\succeq$) is obtained in Chen et al (2010).

Second order growth condition is established for a convex composite problem $g(M(x))$ in Bonnans and Shapiro (2000).

Lemma (second-order sufficient condition and second-order growth condition)

Notations: $x^* \in R^n \setminus \{0\}$; $I = supp(x^*)$.

Conclusion: *the following statements are equivalent:*

- x^* *is a local minimum of the* l_q *regularization problem;*
- *the following first- and second-order conditions hold:*

$$2A_I^\top(A_I x_I^* - b) + \lambda q \left((|x_i^*|^{q-1} sign(x_i^*))_{i\in I} \right) = 0,$$

$$2A_I^\top A_I + \lambda q(q-1) diag\left(|x_i^*|^{q-2} \right)_{i\in I} \succ 0; \quad (1)$$

- *the second-order growth condition holds at* x^**:*

$$F(x) \geq F(x^*) + \varepsilon \|x - x^*\|_2^2 \quad \text{for any } x \in B(x^*, \delta).$$

Remarks: Second-order necessary condition (1) (with $\succeq$) is obtained in Chen et al (2010).

Second order growth condition is established for a convex composite problem $g(M(x))$ in Bonnans and Shapiro (2000).

Lemma (second-order sufficient condition and second-order growth condition)

Notations: $x^* \in R^n \setminus \{0\}$; $I = supp(x^*)$.

Conclusion: *the following statements are equivalent:*

- x^* *is a local minimum of the* l_q *regularization problem;*
- *the following first- and second-order conditions hold:*

$$2A_I^\top(A_I x_I^* - b) + \lambda q \left((|x_i^*|^{q-1} sign(x_i^*))_{i\in I} \right) = 0,$$

$$2A_I^\top A_I + \lambda q(q-1) diag\left((|x_i^*|^{q-2})_{i\in I} \right) \succ 0; \quad (1)$$

- *the second-order growth condition holds at* x^*:

$$F(x) \geq F(x^*) + \varepsilon \|x - x^*\|_2^2 \quad \text{for any } x \in B(x^*, \delta).$$

Remarks: Second-order necessary condition (1) (with $\succeq$) is obtained in Chen et al (2010).

Second order growth condition is established for a convex composite problem $g(M(x))$ in Bonnans and Shapiro (2000).

Lemma (second-order sufficient condition and second-order growth condition)

Notations: $x^* \in R^n \setminus \{0\}$; $I = supp(x^*)$.
Conclusion: *the following statements are equivalent:*

- x^* *is a local minimum of the* l_q *regularization problem;*
- *the following first- and second-order conditions hold:*

$$2A_I^\top(A_I x_I^* - b) + \lambda q \left((|x_i^*|^{q-1} sign(x_i^*))_{i\in I} \right) = 0,$$

$$2A_I^\top A_I + \lambda q(q-1) diag\left(|x_i^*|^{q-2} \right)_{i \in I} \succ 0; \qquad (1)$$

- *the second-order growth condition holds at* x^**:*

$$F(x) \geq F(x^*) + \varepsilon \|x - x^*\|_2^2 \quad \text{for any } x \in B(x^*, \delta).$$

Remarks: Second-order necessary condition (1) (with $\succeq$) is obtained in Chen et al (2010).
Second order growth condition is established for a convex composite problem $g(M(x))$ in Bonnans and Shapiro (2000).

Lemma (second-order sufficient condition and second-order growth condition)

Notations: $x^* \in R^n \setminus \{0\}$; $I = supp(x^*)$.

Conclusion: *the following statements are equivalent:*

- x^* *is a local minimum of the* l_q *regularization problem;*
- *the following first- and second-order conditions hold:*

$$2A_I^\top(A_I x_I^* - b) + \lambda q \left((|x_i^*|^{q-1} sign(x_i^*))_{i \in I} \right) = 0,$$

$$2A_I^\top A_I + \lambda q(q-1) diag\left(|x_i^*|^{q-2} \right)_{i \in I} \succ 0; \tag{1}$$

- *the second-order growth condition holds at* x^**:*

$$F(x) \geq F(x^*) + \varepsilon \|x - x^*\|_2^2 \quad \text{for any } x \in B(x^*, \delta).$$

Remarks: Second-order necessary condition (1) (with $\succeq$) is obtained in Chen et al (2010).

Second order growth condition is established for a convex composite problem $g(M(x))$ in Bonnans and Shapiro (2000).

Lemma (second-order sufficient condition and second-order growth condition)

Notations: $x^* \in R^n \setminus \{0\}$; $I = supp(x^*)$.

Conclusion: *the following statements are equivalent:*

- x^* *is a local minimum of the* l_q *regularization problem;*
- *the following first- and second-order conditions hold:*

$$2A_I^\top(A_I x_I^* - b) + \lambda q \left((|x_i^*|^{q-1} sign(x_i^*))_{i \in I} \right) = 0,$$

$$2A_I^\top A_I + \lambda q(q-1) diag\left(|x_i^*|^{q-2} \right)_{i \in I} \succ 0; \quad (1)$$

- *the second-order growth condition holds at* x^**:*

$$F(x) \geq F(x^*) + \varepsilon \|x - x^*\|_2^2 \quad \text{for any } x \in B(x^*, \delta).$$

Remarks: Second-order necessary condition (1) (with $\succeq$) is obtained in Chen et al (2010).
Second order growth condition is established for a convex composite problem $g(M(x))$ in Bonnans and Shapiro (2000).

Theorem (Linear Convergence of PGA for ℓ_q regularization)

Notations:

- $0 < q < 1$;
- $F(x) := \|Ax - b\|_2^2 + \lambda \|x\|_q^q$;
- $\{x^k\}$ *be a sequence generated by the PGA (converging to* x^**).*

Assumptions:

- x^* *is a local minimum of* F.

Conclusion:

- *Linear convergence: Then there exist* $C > 0$ *and* $\eta \in (0,1)$ *such that*

$$F(x^k) - F(x^*) \le C\eta^k \quad \text{and} \quad \|x^k - x^*\|_2 \le C\eta^k, \quad \textit{for any } k \in \mathbb{N}.$$

Theorem (Linear Convergence of PGA for ℓ_q regularization)

Notations:

- $0 < q < 1$;
- $F(x) := \|Ax - b\|_2^2 + \lambda \|x\|_q^q$;
- $\{x^k\}$ *be a sequence generated by the PGA (converging to* x^**).*

Assumptions:

- x^* *is a local minimum of* F.

Conclusion:

- *Linear convergence: Then there exist* $C > 0$ *and* $\eta \in (0,1)$ *such that*

$$F(x^k) - F(x^*) \le C\eta^k \quad \text{and} \quad \|x^k - x^*\|_2 \le C\eta^k, \quad \text{for any } k \in \mathbb{N}.$$

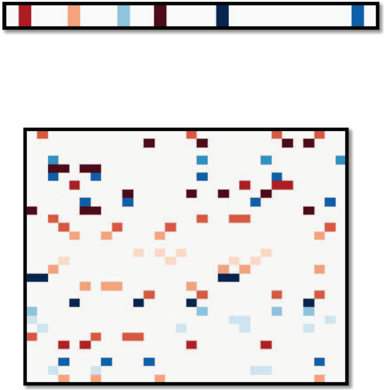
- Recently, enhancing the recoverability due to the special structures has become an active topic in the sparse optimization.
- Group sparse structure: the solution has a natural grouping of its components, and the components within each group are likely to be either all zeros or all nonzeros. The grouping information is usually pre-defined based on prior knowledge of specific problems.
- The group Lasso [Yuan and Lin 2006]:

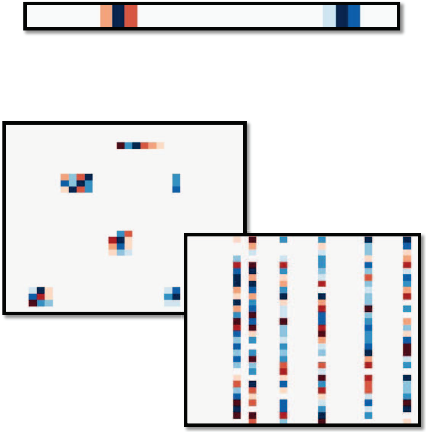
$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2^2 + \lambda\|x\|_{2,1},$$

where $\|x\|_{2,1} = \sum_{i=1}^r \|x_{\mathcal{G}_i}\|$. The group Lasso has been applied in multifactor analysis-of-variance, multi-task learning, dynamic MRI and gene finding.

Sparsity VS Group Sparsity

Sparsity VS Group Sparsity





- Recently, enhancing the recoverability due to the special structures has become an active topic in the sparse optimization.
- Group sparse structure: the solution has a natural grouping of its components, and the components within each group are likely to be either all zeros or all nonzeros. The grouping information is usually pre-defined based on prior knowledge of specific problems.
- The group Lasso [Yuan and Lin 2006]:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|x\|_{2,1},$$

where $\|x\|_{2,1} = \sum_{i=1}^{r} \|x_{\mathcal{G}_i}\|$.

- The group Lasso has been applied in multifactor analysis-of-variance, multi-task learning, dynamic MRI and gene finding.

$\ell_{p,q}$ Regularization Method

The $\ell_{p,q}$ regularization model ($p \geq 1$, $0 < q \leq 1$):

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda \|x\|_{p,q}^q,$$

where $x := (x_{\mathcal{G}_1}^\top, \cdots, x_{\mathcal{G}_r}^\top)^\top$ and

$$\|x\|_{p,q} = \left(\sum_{i=1}^{r} \|x_{\mathcal{G}_i}\|_p^q \right)^{1/q}.$$

- $\|x\|_{p,p} = \|x\|_p$,
- when $\max |\mathcal{G}_i| = 1$, $\|x\|_{p,q} = \|x\|_q$.

Theorem (Oracle Result and Recovery Bound)

Notations:

- $0 \le q \le 1 \le p \le 2$;
- $A\bar{x} = b$, $\mathcal{S} := \{i \in \{1, \ldots, r\} : \bar{x}_{\mathcal{G}_i} \neq 0\}$, $S := |\mathcal{S}|$;
- x^* *be a global minimum of the* $\ell_{p,q}$ *regularization problem,* K *be the smallest integer such that* $2^{K-1}p \ge 1$.

Assumptions:

- (p, q)-$\mathrm{GREC}(S, S)$ *holds.*

Conclusion:

- *oracle inequality:*

$$\|Ax^* - A\bar{x}\|_2^2 + \lambda \|x^*_{\mathcal{G}_{\mathcal{S}^c}}\|_{p,q}^q \le \lambda^{\frac{2}{2-q}} S^{(1-2^{-K})\frac{2}{2-q}} / \phi_{p,q}^{\frac{2q}{2-q}}(S, S) = O(\lambda^{\frac{2}{2-q}} S),$$

- *recovery bound:*

$$\|x^* - \bar{x}\|_2^2 \le 2\lambda^{\frac{2}{2-q}} S^{\frac{q-2}{q} + (1-2^{-K})\frac{4}{q(2-q)}} / \phi_{p,q}^{\frac{4}{2-q}}(S, S) = O(\lambda^{\frac{2}{2-q}} S).$$

Theorem (Oracle Result and Recovery Bound)

Notations:

- $0 \le q \le 1 \le p \le 2$;
- $A\bar{x} = b$, $\mathcal{S} := \left\{ i \in \{1, \ldots, r\} : \bar{x}_{\mathcal{G}_i} \neq 0 \right\}$, $S := |\mathcal{S}|$;
- x^* be a global minimum of the $\ell_{p,q}$ regularization problem, K be the smallest integer such that $2^{K-1} p \ge 1$.

Assumptions:

- (p, q)-$\mathrm{GREC}(S, S)$ holds.

Conclusion:

- *oracle inequality*:

$$\|Ax^* - A\bar{x}\|_2^2 + \lambda \|x^*_{\mathcal{G}_{\mathcal{S}^c}}\|_{p,q}^q \le \lambda^{\frac{2}{2-q}} S^{(1-2^{-K})\frac{2}{2-q}} / \phi_{p,q}^{\frac{2q}{2-q}}(S, S) = O(\lambda^{\frac{2}{2-q}} S),$$

- *recovery bound*:

$$\|x^* - \bar{x}\|_2^2 \le 2\lambda^{\frac{2}{2-q}} S^{\frac{q-2}{q} + (1-2^{-K})\frac{4}{q(2-q)}} / \phi_{p,q}^{\frac{4}{2-q}}(S, S) = O(\lambda^{\frac{2}{2-q}} S).$$

Theorem (Linear Convergence of PGA for $\ell_{1,q}$ regularization)

Notations:

- $0 < q < 1$;
- $F(x) := \|Ax - b\|_2^2 + \lambda \|x\|_{1,q}^q$;
- $\{x^k\}$ *be a sequence generated by the PGA (converging to* x^**).*

Assumptions:

- x^* *is a local minimum of* F*;*
- *any nonzero group of* x^* *is active.*

Conclusion:

- *Linear convergence: Then there exist* $C > 0$ *and* $\eta \in (0, 1)$ *such that*

$$F(x^k) - F(x^*) \leq C\eta^k \quad \text{and} \quad \|x^k - x^*\|_2 \leq C\eta^k, \quad \textit{for any } k \in \mathbb{N}.$$

Theorem (Linear Convergence of PGA for $\ell_{1,q}$ regularization)

Notations:

- $0 < q < 1$;
- $F(x) := \|Ax - b\|_2^2 + \lambda \|x\|_{1,q}^q$;
- $\{x^k\}$ *be a sequence generated by the PGA (converging to* x^**).*

Assumptions:

- x^* *is a local minimum of* F;
- *any nonzero group of* x^* *is active.*

Conclusion:

- *Linear convergence: Then there exist* $C > 0$ *and* $\eta \in (0, 1)$ *such that*

$$F(x^k) - F(x^*) \le C\eta^k \quad \text{and} \quad \|x^k - x^*\|_2 \le C\eta^k, \quad \textit{for any } k \in \mathbb{N}.$$

Lemma (Analytical formulae of proximal optimization subproblems.)

Let $z \in \mathbb{R}^l$, $v > 0$ *and the proximal regularization* $R_{p,q}(x) := \lambda\|x\|_p^q + \frac{1}{2v}\|x - z\|_2^2$. *Then the proximal operator*

$$P_{p,q}(z) \in \arg\min_{x \in \mathbb{R}^l}\{R_{p,q}(x)\} = \arg\min_{x \in \mathbb{R}^l}\{\lambda\|x\|_p^q + \frac{1}{2v}\|x - z\|_2^2\}$$

has the following analytical formula:

(i) *if* $p = 2$ *and* $q = 1$, *then*

$$P_{2,1}(z) = \begin{cases} z - \frac{v\lambda}{\|z\|_2} z, & \|z\|_2 > v\lambda, \\ 0, & \text{otherwise}, \end{cases}$$

(ii) *if* $p \geq 1$ *and* $q = 0$, *then*

$$P_{p,0}(z) = \begin{cases} z, & \|z\|_2 > \sqrt{2v\lambda}, \\ 0 \text{ or } z, & \|z\|_2 = \sqrt{2v\lambda}, \\ 0, & \|z\|_2 < \sqrt{2v\lambda}, \end{cases}$$

Lemma (Con't)

(iii) *if $p = 2$ and $q = 1/2$, then*

$$P_{2,1/2}(z) = \begin{cases} \frac{16\|z\|_2^{3/2}\cos^3(\frac{\pi}{3}-\frac{\psi(z)}{3})}{3\sqrt{3}v\lambda+16\|z\|_2^{3/2}\cos^3(\frac{\pi}{3}-\frac{\psi(z)}{3})}z, & \|z\|_2 > \frac{3}{2}(v\lambda)^{2/3}, \\ 0 \text{ or } \frac{16\|z\|_2^{3/2}\cos^3(\frac{\pi}{3}-\frac{\psi(z)}{3})}{3\sqrt{3}v\lambda+16\|z\|_2^{3/2}\cos^3(\frac{\pi}{3}-\frac{\psi(z)}{3})}z, & \|z\|_2 = \frac{3}{2}(v\lambda)^{2/3}, \\ 0, & \|z\|_2 < \frac{3}{2}(v\lambda)^{2/3}, \end{cases}$$

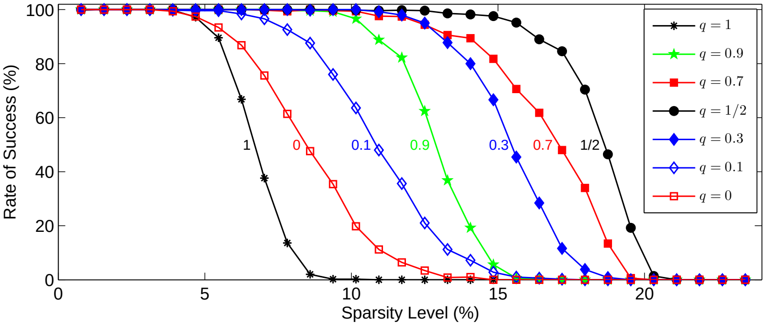
with $\psi(z) = \arccos\left(\frac{v\lambda}{4}(\frac{3}{\|z\|_2})^{3/2}\right)$,

(iv) *if $p = 1$ and $q = 1/2$, then*

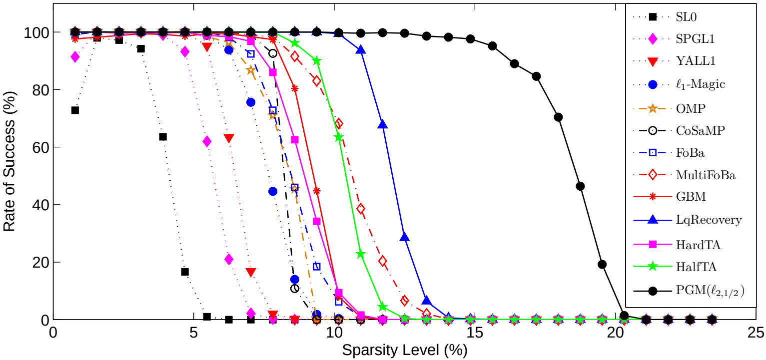
$$P_{1,1/2}(z) = \begin{cases} \tilde{z}, & R_{1,1/2}(\tilde{z}) < R_{1,1/2}(0), \\ 0 \text{ or } \tilde{z}, & R_{1,1/2}(\tilde{z}) = R_{1,1/2}(0), \\ 0, & R_{1,1/2}(\tilde{z}) > R_{1,1/2}(0), \end{cases}$$

where $\tilde{z} = z - \frac{\sqrt{3}v\lambda sgn(z)}{4\sqrt{\|z\|_1}\cos(\frac{\pi}{3}-\frac{\xi(z)}{3})}$, *and* $\xi(z) = \arccos\left(\frac{v\lambda l}{4}(\frac{3}{\|z\|_1})^{3/2}\right)$.

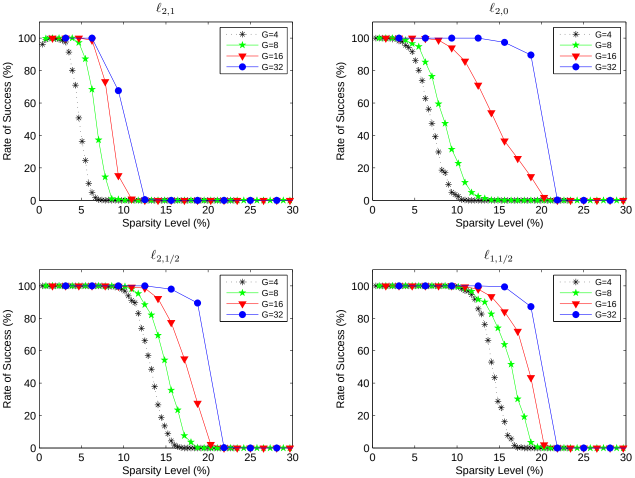
Variation of PGA-GSO when varying the regularization order q



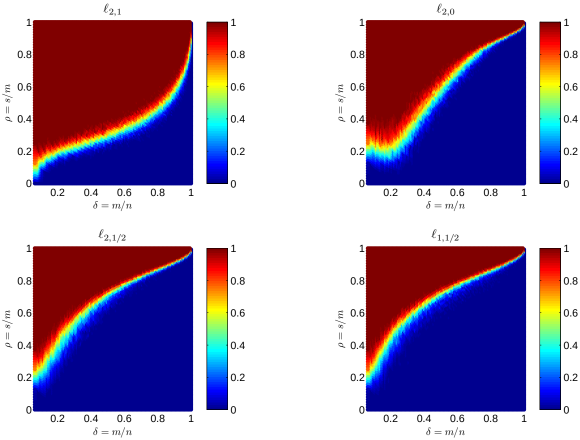
Comparison of PGA-GSO with the state-of-arts algorithms

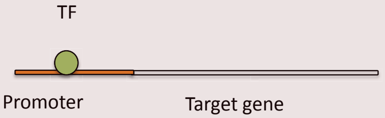


Sensitivity analysis on group size

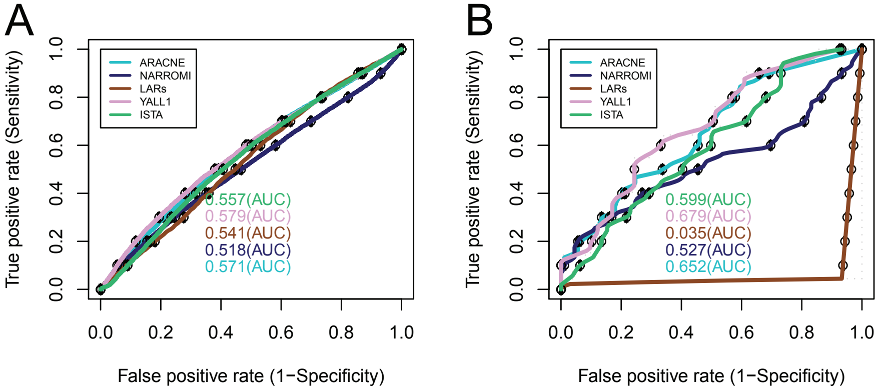


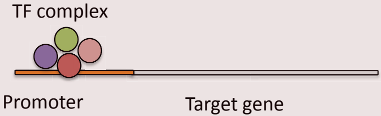
Phase diagram study of $\ell_{p,q}$ regularization



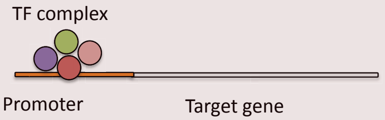


TF regulates a target gene by co-binding on its promoter

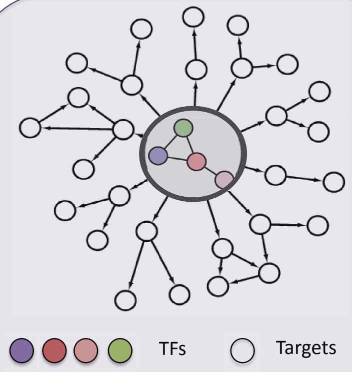

A
ARACNE
NARROMI
LARs
YALL1
ISTA
0.557(AUC)
0.579(AUC)
0.541(AUC)
0.518(AUC)
0.571(AUC)
True positive rate (Sensitivity)
False positive rate (1−Specificity)
B
ARACNE
NARROMI
LARs
YALL1
ISTA
0.599(AUC)
0.679(AUC)
0.035(AUC)
0.527(AUC)
0.652(AUC)
True positive rate (Sensitivity)
False positive rate (1−Specificity)



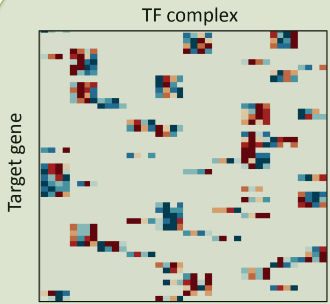
TF complex regulates a target gene by co-binding on its promoter



TF complex regulates a target gene by co-binding on its promoter

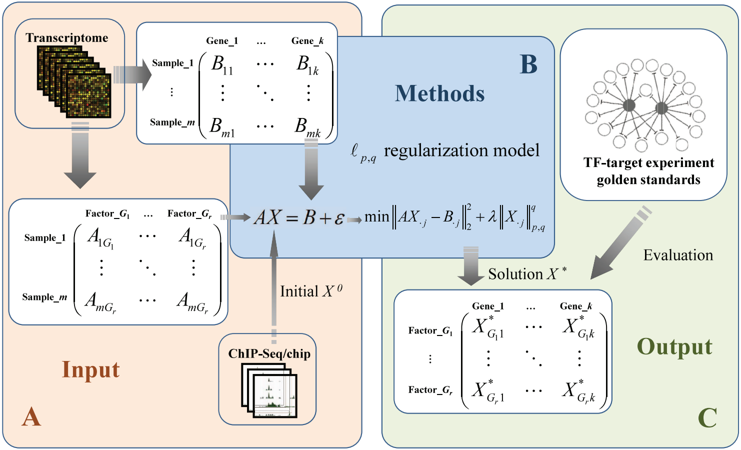


Gene regulatory network controlled by TF complex

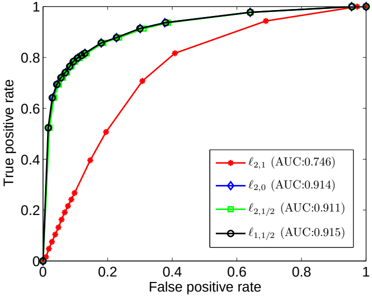


Group structure of a gene regulatory network

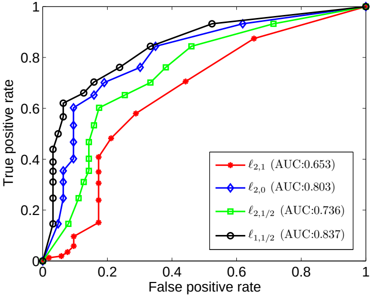
Workflow of gene regulatory network inference



ROC curves and AUCs of PGA-GSO on mESC gene regulatory network inference

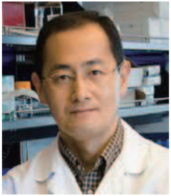


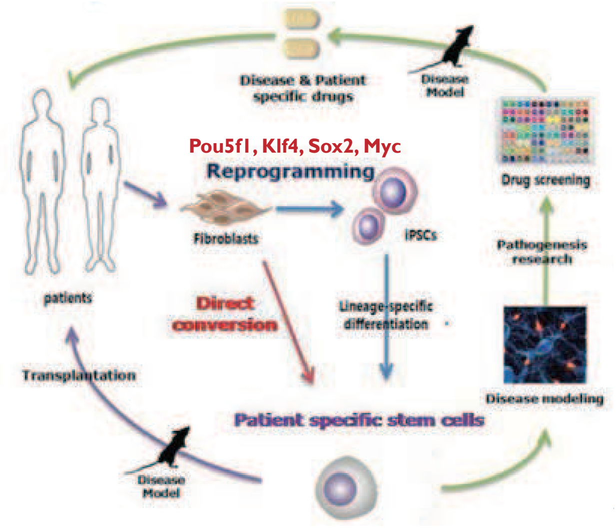
(a) Evaluation with high-throughput golden standard.



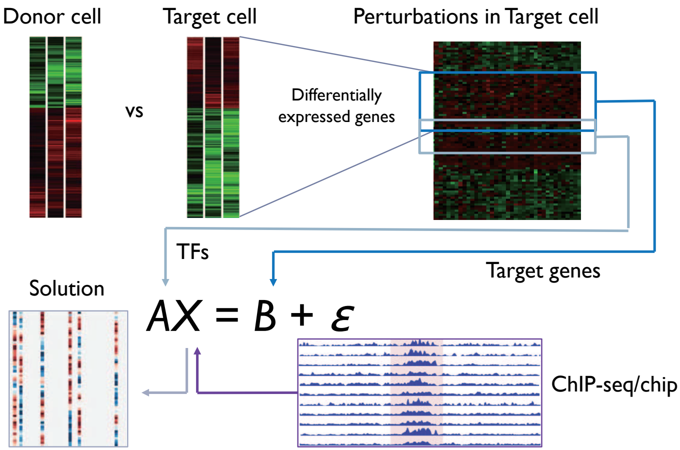
(b) Evaluation with literature-based low-throughput golden standard.

Master Regulator Inference (Group L_q)





Master Regulator Inference (Group L_q)



TF	Pubmed	Mogrify	CellNet	pTFBS_SuperEnh	ChIP_SuperEnh	HubTF	TopTF	gLASSO	GSO	GSO_SuperEnh
Pou5f1	306	3	2		13	5	6		3	2
Sox2	285	2			14	14	15		7	3
Klf4	229				15	15	14		10	8
Myc	226				16	17	16		9	15
Nanog	129	1	3		19	8	8		8	6
Esrrb	2					2	2		2	5
Sall4	2	16	14						17	10
Lin28	2									
Tet1	9									
Prdm14	4	17								
Jarid2	2									
Zic3	1	8								
Glis	1									
Total Score		79	44	0	28	65	65	0	91	98
Total Weighted Score		449	275	0	210	369	365	0	534	571

- **Yaohua Hu**, Chong Li, Kaiwen Meng, Jing Qin and Xiaoqi Yang. Group sparse optimization via $\ell_{p,q}$ regularization. **Journal of Machine Learning Research**, 18:1–52, 2017.
- Jinhua Wang, **Yaohua Hu**, Chong Li and Jen-Chih Yao. Linear convergence of CQ algorithms and applications in gene regulatory network inference. **Inverse Problems**, 33(5): 055017, 2017.
- Lufang Zhang, **Yaohua Hu**, C. Li and Jen-Chih Yao. A new linear convergence result for iterative soft thresholding algorithm. **Optimization**, accepted for publication.
- Jing Qin, **Yaohua Hu**, Jen-Chih Yao, Yiming Qin, Ka Hou Chu, Junwen Wang. Group Sparse Optimization: AnIntegrative OMICs Method to Predict Master Transcription Factors for Cell Fate Conversion, **Nucleic Acids Research**, in revision.
- Jing Qin, **Yaohua Hu**, Feng Xu, Harry Yalamanchili, and Junwen Wang. Inferring gene regulatory networks by integrating ChIP-seq/chip and transcriptome data via LASSO-type regularization methods. **Methods**, 67(3):294–303, 2014.

- A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- P. J. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics*, 37:1705–1732, 2009.
- J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, pages 1–36, 2013.
- E. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51:4203–4215, 2005.
- R. Chartrand and V. Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24:1–14, 2008.
- X. Chen, F. Xu, and Y. Ye. Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. *SIAM Journal on Scientific Computing*, 32(5):2832–2852, 2010.
- D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(8):1289–1306, 2006.
- E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1-minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3): 1107–1130, 2008.
- Z. Xu, X. Chang, F. Xu, and H. Zhang. $L_{1/2}$ regularization: A thresholding representation theory and a fast solver. *IEEE Transactions on Neural Networks and Learning Systems*, 23:1013–1027, 2012.
- M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of The Royal Statistical Society, series B*, 68:49–67, 2006.

Thank You for Your Attention.