Analysis and algorithms for some compressed sensing models based on the ratio of ℓ_1 and ℓ_2 norms

Ting Kei Pong
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hong Kong

One World Optimization Seminar
Nov 2021

(Joint work with Peiran Yu and Liaoyuan Zeng)

Motivating applications

- Basis pursuit:

$$\min_x \; \|x\|_1 \;\text{ subject to }\; Ax = b,$$

where $A \in \mathbb{R}^{m\times n}$ has full row rank, $b \in \mathbb{R}^m \backslash \{0\}$. (hence, $A^{-1}\{b\} \neq \emptyset$)

Motivating applications

- Basis pursuit:

$$\min_x \ \|x\|_1 \ \text{ subject to } \ Ax = b,$$

where $A \in \mathbb{R}^{m\times n}$ has full row rank, $b \in \mathbb{R}^m \backslash \{0\}$. (hence, $A^{-1}\{b\} \neq \emptyset$)

- Basis pursuit with Gaussian noise:

$$\min_x \ \|x\|_1 \ \text{ subject to } \ \|Ax - b\| \leq \sigma,$$

where $A \in \mathbb{R}^{m\times n}$ has full row rank, $b \in \mathbb{R}^m$, $\sigma \in (0, \|b\|)$.

Motivating applications

- Basis pursuit:

$$\min_x \ \|x\|_1 \ \text{ subject to } \ Ax = b,$$

where $A \in \mathbb{R}^{m\times n}$ has full row rank, $b \in \mathbb{R}^m \setminus \{0\}$. (hence, $A^{-1}\{b\} \neq \emptyset$)

- Basis pursuit with Gaussian noise:

$$\min_x \ \|x\|_1 \ \text{ subject to } \ \|Ax - b\| \le \sigma,$$

where $A \in \mathbb{R}^{m\times n}$ has full row rank, $b \in \mathbb{R}^m$, $\sigma \in (0, \|b\|)$.

- Other sparsity inducing objective? Other noise models?

L1 over L2 models

- ℓ_1/ℓ_2 for compressed sensing dates back to (Yin, Esser, Xin '14), and has recently been extensively studied (Rahimi, Wang, Dong, Lou '19), (Wang, Yan, Lou '20), (Wang, Tao, Nagy, Lou '21)...

L1 over L2 models

- ℓ_1/ℓ_2 for compressed sensing dates back to (Yin, Esser, Xin '14), and has recently been extensively studied (Rahimi, Wang, Dong, Lou '19), (Wang, Yan, Lou '20), (Wang, Tao, Nagy, Lou '21)...
- Noiseless model: (Rahimi, Wang, Dong, Lou '19)

$$\min_x \frac{\|x\|_1}{\|x\|} \text{ subject to } Ax = b,$$

where $A \in \mathbb{R}^{m\times n}$ has full row rank and $b \in \mathbb{R}^m \backslash \{0\}$.

L1 over L2 models

- ℓ_1/ℓ_2 for compressed sensing dates back to (Yin, Esser, Xin '14), and has recently been extensively studied (Rahimi, Wang, Dong, Lou '19), (Wang, Yan, Lou '20), (Wang, Tao, Nagy, Lou '21)...
- Noiseless model: (Rahimi, Wang, Dong, Lou '19)

$$\min_x \frac{\|x\|_1}{\|x\|} \text{ subject to } Ax = b,$$

where $A \in \mathbb{R}^{m \times n}$ has full row rank and $b \in \mathbb{R}^m \setminus \{0\}$.

- Noisy model: (Zeng, Yu, P. '21)

$$\min_x \frac{\|x\|_1}{\|x\|} \text{ subject to } q(x) \leq 0,$$

where $q = P_1 - P_2$ with P_1 Lipschitz differentiable and P_2 convex finite-valued, $[q \leq 0] \neq \emptyset$ and $q(0) > 0$.

L1 over L2 models cont.

Three concrete noisy models:

- Gaussian noise:

$$q(x) = \|Ax - b\|^2 - \sigma^2,$$

  where A has full row rank, $b \in \mathbb{R}^m$, $\sigma \in (0, \|b\|)$.

L1 over L2 models cont.

Three concrete noisy models:

- Gaussian noise:

$$q(x) = \|Ax - b\|^2 - \sigma^2,$$

  where A has full row rank, $b \in \mathbb{R}^m$, $\sigma \in (0, \|b\|)$.

- Cauchy noise (Carrilo, Barner, Aysal '10):

$$q(x) = \|Ax - b\|_{LL_2,\gamma} - \sigma,$$

  where A has full row rank, $b \in \mathbb{R}^m$, $\sigma \in (0, \|b\|_{LL_2,\gamma})$, with

$$\|y\|_{LL_2,\gamma} := \sum_{i=1}^{m} \log\left(1 + \frac{y_i^2}{\gamma^2}\right).$$

Note: These q are Lipschitz differentiable.

L1 over L2 models cont.

Three concrete noisy models cont.:

- Electromyographic + Gaussian noise (Carrilo, Barner, Aysal '10), (Liu, P., Takeda '19):

$$q(x) = \mathrm{dist}(Ax - b, S)^2 - \sigma^2,$$

where A has full row rank, $b \in \mathbb{R}^m$, $S = \{z : \|z\|_0 \leq r\}$, and $\sigma \in (0, \mathrm{dist}(b, S))$.

L1 over L2 models cont.

Three concrete noisy models cont.:

- Electromyographic + Gaussian noise (Carrilo, Barner, Aysal '10), (Liu, P., Takeda '19):

$$q(x) = \operatorname{dist}(Ax - b, S)^2 - \sigma^2,$$

  where A has full row rank, $b \in \mathbb{R}^m$, $S = \{z : \|z\|_0 \le r\}$, and $\sigma \in (0, \operatorname{dist}(b, S))$.

Note:

- $$\begin{aligned} q(x) &= \min_{z \in S} \|Ax - b - z\|^2 - \sigma^2 \\ &= \underbrace{\|Ax - b\|^2 - \sigma^2}_{P_1(x)} - \underbrace{\max_{z \in S}\{2\langle z, Ax - b\rangle - \|z\|^2\}}_{P_2(x)}. \end{aligned}$$
- $2A^T \mathrm{Proj}_S(Ax - b) \subseteq \partial P_2(x)$.

Previous work

- The ℓ_1/ℓ_2 models are motivated from the spherical section property (SSP) (Vavasis '09, Zhang '13).

Previous work

- The ℓ_1/ℓ_2 models are motivated from the spherical section property (SSP) (Vavasis '09, Zhang '13).
- For noiseless model, under s-nullspace property, any s-sparse solution is a local minimizer. (Rahimi, Wang, Dong, Lou '19)

Previous work

- The ℓ_1/ℓ_2 models are motivated from the spherical section property (SSP) (Vavasis '09, Zhang '13).
- For noiseless model, under s-nullspace property, any s-sparse solution is a local minimizer. (Rahimi, Wang, Dong, Lou '19)
- A Dinkelbach-type algorithm was proposed for the noiseless case with subsequential convergence established: (Wang, Yan, Lou '20)

$$\begin{cases} x^{t+1} = \arg\min_{Ax=b} \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{1}{2}\|x - x^t\|^2, \\ \omega_{t+1} = \|x^{t+1}\|_1 / \|x^{t+1}\|. \end{cases}$$

Previous work

- The ℓ_1/ℓ_2 models are motivated from the spherical section property (SSP) (Vavasis '09, Zhang '13).
- For noiseless model, under s-nullspace property, any s-sparse solution is a local minimizer. (Rahimi, Wang, Dong, Lou '19)
- A Dinkelbach-type algorithm was proposed for the noiseless case with subsequential convergence established: (Wang, Yan, Lou '20)

$$\begin{cases} x^{t+1} = \arg\min_{Ax=b} \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{1}{2}\|x - x^t\|^2, \\ \omega_{t+1} = \|x^{t+1}\|_1/\|x^{t+1}\|. \end{cases}$$

- Does a (globally optimal) solution exist?

Previous work

- The ℓ_1/ℓ_2 models are motivated from the spherical section property (SSP) (Vavasis '09, Zhang '13).
- For noiseless model, under s-nullspace property, any s-sparse solution is a local minimizer. (Rahimi, Wang, Dong, Lou '19)
- A Dinkelbach-type algorithm was proposed for the noiseless case with subsequential convergence established: (Wang, Yan, Lou '20)

$$\begin{cases} x^{t+1} = \underset{Ax=b}{\arg\min} \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{1}{2}\|x - x^t\|^2, \\ \omega_{t+1} = \|x^{t+1}\|_1/\|x^{t+1}\|. \end{cases}$$

- Does a (globally optimal) solution exist?
- What is the rate of convergence of the above algorithm?

Previous work

- The ℓ_1/ℓ_2 models are motivated from the spherical section property (SSP) (Vavasis '09, Zhang '13).
- For noiseless model, under s-nullspace property, any s-sparse solution is a local minimizer. (Rahimi, Wang, Dong, Lou '19)
- A Dinkelbach-type algorithm was proposed for the noiseless case with subsequential convergence established: (Wang, Yan, Lou '20)

$$\begin{cases} x^{t+1} = \underset{Ax=b}{\arg\min} \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{1}{2}\|x - x^t\|^2, \\ \omega_{t+1} = \|x^{t+1}\|_1/\|x^{t+1}\|. \end{cases}$$

- Does a (globally optimal) solution exist?
- What is the rate of convergence of the above algorithm?
- How about algorithm for the noisy case?

Spherical section property

Definition: (Spherical section property) (Vavasis '09, Zhang '13)
Let m, n be two positive integers such that $m < n$. Let V be an $(n-m)$-dimensional subspace of $\mathbb{R}^n$ and s be a positive integer. We say that V has the s-spherical section property (s-SSP) if

$$\inf_{v \in V \setminus \{0\}} \frac{\|v\|_1}{\|v\|} \geq \sqrt{\frac{m}{s}}.$$

Spherical section property

Definition: (Spherical section property) (Vavasis '09, Zhang '13)
Let m, n be two positive integers such that $m < n$. Let V be an $(n-m)$-dimensional subspace of $\mathbb{R}^n$ and s be a positive integer. We say that V has the s-spherical section property (s-SSP) if

$$\inf_{v \in V \setminus \{0\}} \frac{\|v\|_1}{\|v\|} \geq \sqrt{\frac{m}{s}}.$$

Intuition: Any k-sparse vector $u \in \mathbb{R}^n$ satisfies $\|u\|_1 \leq \sqrt{k}\|u\|$.

Spherical section property

Definition: (Spherical section property) (Vavasis '09, Zhang '13)
Let m, n be two positive integers such that $m < n$. Let V be an $(n-m)$-dimensional subspace of $\mathbb{R}^n$ and s be a positive integer. We say that V has the s-spherical section property (s-SSP) if

$$\inf_{v \in V \setminus \{0\}} \frac{\|v\|_1}{\|v\|} \geq \sqrt{\frac{m}{s}}.$$

Intuition: Any k-sparse vector $u \in \mathbb{R}^n$ satisfies $\|u\|_1 \leq \sqrt{k}\|u\|$.

Fact: (Vavasis '09)
If $A \in \mathbb{R}^{m \times n}$ ($m < n$) has i.i.d. standard Gaussian entries, then its nullspace has the s-SSP for $s = c_1(\log(n/m) + 1)$ with probability at least $1 - e^{-c_0(n-m)}$, where $c_0, c_1 > 0$ are independent of m and n.

Solution existence

Theorem 1. (Zeng, Yu, P. '21)
For the noiseless model, suppose that $\ker A$ has the s-spherical section property for some $s > 0$ and there exists $\widetilde{x} \in \mathbb{R}^n$ such that

$$\|\widetilde{x}\|_0 < m/s \quad \text{and} \quad A\widetilde{x} = b.$$

Then the set of optimal solutions is nonempty.

Solution existence

Theorem 1. (Zeng, Yu, P. '21)
For the noiseless model, suppose that $\ker A$ has the s-spherical section property for some $s > 0$ and there exists $\widetilde{x} \in \mathbb{R}^n$ such that

$$\|\widetilde{x}\|_0 < m/s \quad \text{and} \quad A\widetilde{x} = b.$$

Then the set of optimal solutions is nonempty.

Idea:

- Consider $F(x) := \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ and

$$\nu_d^* := \inf\{\|d\|_1 : \ Ad = 0, \|d\| = 1\}.$$

  One can show that every minimizing sequence of F is bounded if and only if $\nu_d^* > \inf F$.

Solution existence

Theorem 1. (Zeng, Yu, P. '21)
For the noiseless model, suppose that $\ker A$ has the s-spherical section property for some $s > 0$ and there exists $\widetilde{x} \in \mathbb{R}^n$ such that

$$\|\widetilde{x}\|_0 < m/s \quad \text{and} \quad A\widetilde{x} = b.$$

Then the set of optimal solutions is nonempty.

Idea:

- Consider $F(x) := \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ and

$$\nu_d^* := \inf\{\|d\|_1 : \ Ad = 0, \|d\| = 1\}.$$

  One can show that every minimizing sequence of F is bounded if and only if $\nu_d^* > \inf F$.

- Notice that

$$\inf F \le \frac{\|\widetilde{x}\|_1}{\|\widetilde{x}\|} \le \sqrt{\|\widetilde{x}\|_0} < \sqrt{\frac{m}{s}} \le \nu_d^*.$$

Solution existence

Theorem 1. (Zeng, Yu, P. '21)
For the noiseless model, suppose that $\ker A$ has the s-spherical section property for some $s > 0$ and there exists $\widetilde{x} \in \mathbb{R}^n$ such that

$$\|\widetilde{x}\|_0 < m/s \quad \text{and} \quad A\widetilde{x} = b.$$

Then the set of optimal solutions is nonempty.

Idea:

- Consider $F(x) := \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ and

$$\nu_d^* := \inf\{\|d\|_1 : \ Ad = 0, \|d\| = 1\}.$$

  One can show that every minimizing sequence of F is bounded if and only if $\nu_d^* > \inf F$.

- Notice that

$$\inf F \le \frac{\|\widetilde{x}\|_1}{\|\widetilde{x}\|} \le \sqrt{\|\widetilde{x}\|_0} < \sqrt{\frac{m}{s}} \le \nu_d^*.$$

Note: Recovery results were proved under suitable s-SSP. (Xu, Narayan, Tran, Webster '21)

KL property & exponent

Definition: (Attouch, Bolte, Redont, Soubeyran '10)
Let h be proper closed and $\alpha \in [0, 1)$.

- h is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \operatorname{dom} \partial h$ if there exist c, ν, $\epsilon > 0$ so that

$$c[h(x) - h(\bar{x})]^{\alpha} \leq \operatorname{dist}(0, \partial h(x))$$

whenever $x \in \operatorname{dom} \partial h$, $\|x - \bar{x}\| \leq \epsilon$ and $h(\bar{x}) < h(x) < h(\bar{x}) + \nu$.

KL property & exponent

Definition: (Attouch, Bolte, Redont, Soubeyran '10)
Let h be proper closed and $\alpha \in [0, 1)$.

- h is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \operatorname{dom} \partial h$ if there exist c, ν, $\epsilon > 0$ so that

$$c[h(x) - h(\bar{x})]^{\alpha} \leq \operatorname{dist}(0, \partial h(x))$$

  whenever $x \in \operatorname{dom} \partial h$, $\|x - \bar{x}\| \leq \epsilon$ and $h(\bar{x}) < h(x) < h(\bar{x}) + \nu$.
- If h has the KL property at every $\bar{x} \in \operatorname{dom} \partial h$ with the same α, then h is said to be a KL function with exponent α.

KL property & exponent

Definition: (Attouch, Bolte, Redont, Soubeyran '10)
Let h be proper closed and $\alpha \in [0, 1)$.

- h is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \mathrm{dom}\,\partial h$ if there exist c, ν, $\epsilon > 0$ so that

$$c[h(x) - h(\bar{x})]^\alpha \le \mathrm{dist}(0, \partial h(x))$$

  whenever $x \in \mathrm{dom}\,\partial h$, $\|x - \bar{x}\| \le \epsilon$ and $h(\bar{x}) < h(x) < h(\bar{x}) + \nu$.
- If h has the KL property at every $\bar{x} \in \mathrm{dom}\,\partial h$ with the same α, then h is said to be a KL function with exponent α.

Examples:

- Proper closed semialgebraic functions are KL functions with exponent $\alpha \in [0, 1)$. (Bolte, Daniilidis, Lewis '07)
- Piecewise linear quadratic (PLQ) functions are KL functions with exponent $\frac{1}{2}$. (Li, P. '18)

KL calculus rules

Consider

$$G(x) := \frac{f(x)}{g(x)} \text{ and } H_u(x) := f(x) - \frac{f(u)}{g(u)} g(x).$$

Theorem 2. (Zeng, Yu, P. '21)
Let f be proper closed with $\inf f \geq 0$, and let g be a nonnegative continuous function that is C^1 on an open set containing $\operatorname{dom} f$ with $\inf_{\operatorname{dom} f} g > 0$. Assume that

- $f = h + \delta_D$ for some locally Lipschitz function h and nonempty closed set D, and h and D are regular at every point in D.

Let $\bar{x}$ be such that $0 \in \partial G(\bar{x})$. Then $\bar{x} \in \operatorname{dom} \partial H_{\bar{x}}$. If $H_{\bar{x}}$ satisfies the KL property with exponent $\theta \in [0, 1)$ at $\bar{x}$, then so does G.

KL calculus rules

Consider

$$G(x) := \frac{f(x)}{g(x)} \text{ and } H_u(x) := f(x) - \frac{f(u)}{g(u)} g(x).$$

Theorem 2. (Zeng, Yu, P. '21)
Let f be proper closed with $\inf f \geq 0$, and let g be a nonnegative continuous function that is C^1 on an open set containing $\operatorname{dom} f$ with $\inf_{\operatorname{dom} f} g > 0$. Assume that

- $f = h + \delta_D$ for some locally Lipschitz function h and nonempty closed set D, and h and D are regular at every point in D.

Let $\bar{x}$ be such that $0 \in \partial G(\bar{x})$. Then $\bar{x} \in \operatorname{dom} \partial H_{\bar{x}}$. If $H_{\bar{x}}$ satisfies the KL property with exponent $\theta \in [0, 1)$ at $\bar{x}$, then so does G.

Remark:

- Proper closed convex functions are regular.
- Any closed convex set is regular.

KL calculus rules cont.

Theorem 3. (Zeng, Yu, P. '21)
Let p be a proper closed function, and let $\bar{x} \in \operatorname{dom} p$ be such that $p(\bar{x}) > 0$. Then the following statements hold.

(i) We have $\partial(p^2)(x) = 2p(x)\partial p(x)$ for all x sufficiently close to $\bar{x}$.

(ii) Suppose in addition that $\bar{x} \in \operatorname{dom} \partial(p^2)$ and p^2 satisfies the KL property at $\bar{x}$ with exponent $\theta \in [0, 1)$.
Then p satisfies the KL property at $\bar{x}$ with exponent $\theta \in [0, 1)$.

KL calculus rules cont.

Theorem 3. (Zeng, Yu, P. '21)
Let p be a proper closed function, and let $\bar{x} \in \operatorname{dom} p$ be such that $p(\bar{x}) > 0$. Then the following statements hold.

(i) We have $\partial(p^2)(x) = 2p(x)\partial p(x)$ for all x sufficiently close to $\bar{x}$.

(ii) Suppose in addition that $\bar{x} \in \operatorname{dom} \partial(p^2)$ and p^2 satisfies the KL property at $\bar{x}$ with exponent $\theta \in [0, 1)$.
Then p satisfies the KL property at $\bar{x}$ with exponent $\theta \in [0, 1)$.

Theorem 4. (Zeng, Yu, P. '21)
The function $x \mapsto \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ is a KL function with exponent $\frac{1}{2}$.

Proof idea

KL exponent of $x \mapsto \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

Proof idea

KL exponent of $x \mapsto \|x\|_1^2/\|x\|^2 + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

Proof idea

KL exponent of $x \mapsto \|x\|_1^2 - \frac{\|\bar{x}\|_1^2}{\|\bar{x}\|^2}\|x\|^2 + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1^2/\|x\|^2 + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

Proof idea

PLQ functions are KL functions with exponent $\frac{1}{2}$.

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1^2 - \frac{\|\bar{x}\|_1^2}{\|\bar{x}\|^2}\|x\|^2 + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1^2/\|x\|^2 + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$

Proof idea

PLQ functions are KL functions with exponent $\frac{1}{2}$.

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1^2 - \frac{\|\bar{x}\|_1^2}{\|\bar{x}\|^2}\|x\|^2 + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$ is $\frac{1}{2}$.

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1^2/\|x\|^2 + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$ is $\frac{1}{2}$.

$\Downarrow$

KL exponent of $x \mapsto \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$ at $\bar{x}$ is $\frac{1}{2}$.

Linear convergence

Corollary 1. (Zeng, Yu, P. '21)
Suppose that x^0 satisfy $Ax^0 = b$. Set $\omega_0 := \|x^0\|_1/\|x^0\|$ and update

$$\begin{cases} x^{t+1} & = \underset{Ax=b}{\arg\min} \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{1}{2}\|x - x^t\|^2, \\ \omega_{t+1} & = \|x^{t+1}\|_1/\|x^{t+1}\|. \end{cases}$$

If $\{x^t\}$ is bounded, then it converges locally linearly to a stationary point of the function $F(x) := \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$.

Linear convergence

Corollary 1. (Zeng, Yu, P. '21)
Suppose that x^0 satisfy $Ax^0 = b$. Set $\omega_0 := \|x^0\|_1/\|x^0\|$ and update

$$\begin{cases} x^{t+1} & = \underset{Ax=b}{\arg\min} \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{1}{2}\|x - x^t\|^2, \\ \omega_{t+1} & = \|x^{t+1}\|_1/\|x^{t+1}\|. \end{cases}$$

If $\{x^t\}$ is bounded, then it converges locally linearly to a stationary point of the function $F(x) := \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$.

Idea:

- Since F is semialgebraic, the convergence of $\{x^t\}$ to some $\bar{x}$ follows from a standard line of analysis. (Attouch, Bolte, Svaiter '13) (Bolte, Sabach, Teboulle '14)

Linear convergence

Corollary 1. (Zeng, Yu, P. '21)
Suppose that x^0 satisfy $Ax^0 = b$. Set $\omega_0 := \|x^0\|_1/\|x^0\|$ and update

$$\begin{cases} x^{t+1} & = \operatorname*{arg\,min}_{Ax=b} \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{1}{2}\|x - x^t\|^2, \\ \omega_{t+1} & = \|x^{t+1}\|_1/\|x^{t+1}\|. \end{cases}$$

If $\{x^t\}$ is bounded, then it converges locally linearly to a stationary point of the function $F(x) := \|x\|_1/\|x\| + \delta_{A^{-1}\{b\}}(x)$.

Idea:

- Since F is semialgebraic, the convergence of $\{x^t\}$ to some $\bar{x}$ follows from a standard line of analysis. (Attouch, Bolte, Svaiter '13) (Bolte, Sabach, Teboulle '14)
- The role of KL exponent:

$$F(x^{t+1}) - F(\bar{x}) \le C_1[\operatorname{dist}(0, \partial F(x^{t+1}))]^2$$
$$\le C_2\|x^{t+1} - x^t\|^2 \le C_3[F(x^t) - F(x^{t+1})].$$

Translation to sequential convergence is standard.

Algorithm for noisy model

The noisy model:

$$\min_x \frac{\|x\|_1}{\|x\|} \quad \text{subject to} \quad q(x) \le 0,$$

where

- $q = P_1 - P_2$ with $[q \le 0] \neq \emptyset$ and $q(0) > 0$.
- P_1 is Lipschitz differentiable and $P_2 : \mathbb{R}^n \to \mathbb{R}$ is convex.

We also assume the generalized MFCQ holds at every feasible x, i.e.,

If $q(x) = 0$, then $\nabla P_1(x) \notin \partial P_2(x)$.

Algorithm for noisy model

The noisy model:

$$\min_x \frac{\|x\|_1}{\|x\|} \text{ subject to } q(x) \le 0,$$

where

- $q = P_1 - P_2$ with $[q \le 0] \ne \emptyset$ and $q(0) > 0$.
- P_1 is Lipschitz differentiable and $P_2 : \mathbb{R}^n \to \mathbb{R}$ is convex.

We also assume the generalized MFCQ holds at every feasible x, i.e.,

If $q(x) = 0$, then $\nabla P_1(x) \notin \partial P_2(x)$.

Remark: The generalized MFCQ holds for our 3 choices of q.

Algorithm for noisy model

The noisy model:

$$\min_x \frac{\|x\|_1}{\|x\|} \text{ subject to } q(x) \le 0,$$

where

- $q = P_1 - P_2$ with $[q \le 0] \neq \emptyset$ and $q(0) > 0$.
- P_1 is Lipschitz differentiable and $P_2 : \mathbb{R}^n \to \mathbb{R}$ is convex.

We also assume the generalized MFCQ holds at every feasible x, i.e.,

If $q(x) = 0$, then $\nabla P_1(x) \notin \partial P_2(x)$.

Remark: The generalized MFCQ holds for our 3 choices of q.

Algorithmic ideas:

- Augmented Lagrangian?

Algorithm for noisy model

The noisy model:

$$\min_{x} \frac{\|x\|_1}{\|x\|} \text{ subject to } q(x) \le 0,$$

where

- $q = P_1 - P_2$ with $[q \le 0] \neq \emptyset$ and $q(0) > 0$.
- P_1 is Lipschitz differentiable and $P_2 : \mathbb{R}^n \to \mathbb{R}$ is convex.

We also assume the generalized MFCQ holds at every feasible x, i.e.,

If $q(x) = 0$, then $\nabla P_1(x) \notin \partial P_2(x)$.

Remark: The generalized MFCQ holds for our 3 choices of q.

Algorithmic ideas:

- Augmented Lagrangian?
- Moving balls approximation...

Moving balls approximation

Moving balls approximation algorithm (Auslender, Shefi, Teboulle '10) was designed for

$$\min_{x}\; f(x) \;\text{ subject to } g_i(x) \leq 0 \;\; \forall i = 1, \ldots, m.$$

Moving balls approximation

Moving balls approximation algorithm (Auslender, Shefi, Teboulle '10) was designed for

$$\min_x \; f(x) \;\text{ subject to } g_i(x) \le 0 \;\; \forall i = 1, \ldots, m.$$

Key update: At an x^t satisfying $\max_{1\le i\le m} g_i(x^t) \le 0$, compute

$$\begin{aligned} x^{t+1} = \arg\min_x \quad & f(x^t) + \langle \nabla f(x^t), x - x^t \rangle + \frac{L_f}{2}\|x - x^t\|^2 \\ \text{s.t.} \quad & g_i(x^t) + \langle \nabla g_i(x^t), x - x^t \rangle + \frac{L_{g_i}}{2}\|x - x^t\|^2 \le 0 \;\; \forall i. \end{aligned}$$

Moving balls approximation

Moving balls approximation algorithm (Auslender, Shefi, Teboulle '10) was designed for

$$\min_x \ f(x) \ \text{ subject to } g_i(x) \le 0 \ \ \forall i = 1, \ldots, m.$$

Key update: At an x^t satisfying $\max_{1\le i\le m} g_i(x^t) \le 0$, compute

$$\begin{aligned} x^{t+1} = \arg\min_x \quad & f(x^t) + \langle \nabla f(x^t), x - x^t\rangle + \tfrac{L_f}{2}\|x - x^t\|^2 \\ \text{s.t.} \quad & g_i(x^t) + \langle \nabla g_i(x^t), x - x^t\rangle + \tfrac{L_{g_i}}{2}\|x - x^t\|^2 \le 0 \ \ \forall i. \end{aligned}$$

- The above algorithm is well defined and any accumulation point of $\{x^t\}$ is stationary. (Auslender, Shefi, Teboulle '10)
- Convergence of $\{x^t\}$ under convexity (Auslender, Shefi, Teboulle '10) or semialgebraicity (Bolte, Pauwels '16) is known.
- Variants with line-search scheme have been proposed (Lu '12) (Bolte, Chen, Pauwels '20).

Subproblem needs iterative solver except for $m = 1$.

$\text{MBA}_{\ell_1/\ell_2}$: The algorithm

Algorithm 1: $\text{MBA}_{\ell_1/\ell_2}$

Step 0. Choose x^0 with $q(x^0) \le 0$, $\alpha > 0$ and $0 < l_{\min} < l_{\max}$. Set $\omega_0 = \|x^0\|_1/\|x^0\|$ and $t = 0$.

Step 1. Choose $l_t^0 \in [l_{\min}, l_{\max}]$ arbitrarily and set $l_t = l_t^0$. Choose $\zeta^t \in \partial P_2(x^t)$.

(1a) Solve the subproblem

$$\begin{aligned} \widetilde{x} = \mathop{\arg\min}_{x \in \mathbb{R}^n} \quad & \|x\|_1 - \frac{\omega_t}{\|x^t\|}\langle x, x^t\rangle + \frac{\alpha}{2}\|x - x^t\|^2 \\ \text{s.t.} \quad & q(x^t) + \langle \nabla P_1(x^t) - \zeta^t, x - x^t\rangle + \frac{l_t}{2}\|x - x^t\|^2 \le 0. \end{aligned}$$

(1b) If $q(\widetilde{x}) \le 0$, go to **Step 2**. Else, update $l_t \leftarrow 2l_t$ and go to (1a).

Step 2. Set $x^{t+1} = \widetilde{x}$ and compute $\omega_{t+1} = \|x^{t+1}\|_1/\|x^{t+1}\|$. Set $\bar{l}_t := l_t$. Update $t \leftarrow t + 1$ and go to **Step 1.**

$\text{MBA}_{\ell_1/\ell_2}$: Subsequential convergence

Theorem 5. (Zeng, Yu, P. '21)

(i) $\text{MBA}_{\ell_1/\ell_2}$ is well defined.

(ii) The Slater condition holds for each subproblem.

(iii) Let $\{x^t\}$ be the sequence generated by $\text{MBA}_{\ell_1/\ell_2}$ and suppose that $\{x^t\}$ is bounded. Then $\lim_{t\to\infty} \|x^{t+1} - x^t\| = 0$, and any accumulation point x^* is a Clarke critical point, in the sense that

$$0 \in \partial \frac{\|x^*\|_1}{\|x^*\|} + \bar{\lambda}\nabla P_1(x^*) - \bar{\lambda}\partial P_2(x^*)$$

for some $\bar{\lambda} \geq 0$ satisfying $\bar{\lambda} q(x^*) = 0$.

$\text{MBA}_{\ell_1/\ell_2}$: Subsequential convergence

Theorem 5. (Zeng, Yu, P. '21)

(i) $\text{MBA}_{\ell_1/\ell_2}$ is well defined.

(ii) The Slater condition holds for each subproblem.

(iii) Let $\{x^t\}$ be the sequence generated by $\text{MBA}_{\ell_1/\ell_2}$ and suppose that $\{x^t\}$ is bounded. Then $\lim_{t\to\infty} \|x^{t+1} - x^t\| = 0$, and any accumulation point x^* is a Clarke critical point, in the sense that

$$0 \in \partial \frac{\|x^*\|_1}{\|x^*\|} + \bar{\lambda} \nabla P_1(x^*) - \bar{\lambda} \partial P_2(x^*)$$

for some $\bar{\lambda} \geq 0$ satisfying $\bar{\lambda} q(x^*) = 0$.

If q is also regular at x^*, then x^* is stationary in the sense that

$$0 \in \partial \left[\frac{\|\cdot\|_1}{\|\cdot\|} + \delta_{[q \leq 0]} \right] (x^*).$$

Global convergence

Define

$$\widetilde{F}(x, y, \zeta, w) := \frac{\|x\|_1}{\|x\|} + \delta_{[\widetilde{q} \leq 0]}(x, y, \zeta, w) + \delta_{\|\cdot\| \geq \rho}(x),$$

with

$$\widetilde{q}(x, y, \zeta, w) := P_1(y) + \langle \nabla P_1(y), x - y \rangle + P_2^*(\zeta) - \langle \zeta, x \rangle + \frac{w}{2}\|x - y\|^2,$$

where $\rho > 0$ is such that $[q \leq 0] \subseteq \{x : \|x\| > \rho\}$.

Global convergence

Define

$$\widetilde{F}(x,y,\zeta,w) := \frac{\|x\|_1}{\|x\|} + \delta_{[\widetilde{q}\le 0]}(x,y,\zeta,w) + \delta_{\|\cdot\|\ge\rho}(x),$$

with

$$\widetilde{q}(x,y,\zeta,w) := P_1(y) + \langle \nabla P_1(y), x-y\rangle + P_2^*(\zeta) - \langle \zeta, x\rangle + \frac{w}{2}\|x-y\|^2,$$

where $\rho > 0$ is such that $[q \le 0] \subseteq \{x : \ \|x\| > \rho\}$.

Theorem 6. (Zeng, Yu, P. '21)
Assume in addition that P_1 is C^2. Let $\{x^t\}$ be generated by $\mathrm{MBA}_{\ell_1/\ell_2}$ and assume that $\{x^t\}$ is bounded.
If $\widetilde{F}$ is a KL function, then $\{x^t\}$ converges to a Clarke critical point x^*: This x^* is a stationary point if q is in addition regular at x^*.

Numerical simulations I

- Solve
$$\min_x \frac{\|x\|_1}{\|x\|} \text{ subject to } \|Ax - b\|_{LL_2,\gamma} \le \sigma.$$
- Consider random instances: generate an $m \times n$ matrix A, a k-sparse vector $\tilde{x}$, a Cauchy noise vector $\hat{n}$ (s.d. 0.01) and set $b = A\tilde{x} + \hat{n}$. Set $\gamma = 0.02$ and $\sigma = 1.2\|\hat{n}\|_{LL_2,\gamma}$.
- Initialize at an approximate solution of
$$\min_x \|x\|_1 \text{ subject to } \|Ax - b\|_{LL_2,\gamma} \le \sigma,$$
obtained via SCP_{ls} initialized at $A^\dagger b$.
- Terminate when $\|x^t - x^{t-1}\| \le tol \cdot \max\{1, \|x^t\|\}$.
- $(n, m, k) = i \cdot (2560, 720, 80)$.

Numerical simulations I

Table: $tol = 10^{-6}$ for $\mathrm{SCP}_{\mathrm{ls}}$ and $\mathrm{MBA}_{\ell_1/\ell_2}$

i	CPU		$\frac{\|x-\tilde{x}\|}{\max\{1,\|\tilde{x}\|\}}$		$\|Ax-b\|_{LL_2,\gamma}-\sigma$	
	$\mathrm{SCP}_{\mathrm{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$	$\mathrm{SCP}_{\mathrm{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$	$\mathrm{SCP}_{\mathrm{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$
2	10.0	0.6 (11.1)	1.3e-01	6.5e-02	-2e-07	-8e-08
4	52.4	2.0 (57.5)	1.3e-01	6.6e-02	-6e-07	-2e-07
6	87.3	4.1 (100.9)	1.3e-01	6.6e-02	-9e-07	-2e-07
8	281.6	7.0 (312.1)	1.3e-01	6.5e-02	-1e-06	-3e-07
10	285.5	11.4 (339.5)	1.3e-01	6.5e-02	-2e-06	-4e-07

Numerical simulations I

Table: *tol* $= 10^{-6}$ for $\mathrm{SCP_{ls}}$ and $\mathrm{MBA}_{\ell_1/\ell_2}$

i	CPU		$\frac{\|x-\tilde{x}\|}{\max\{1,\|\tilde{x}\|\}}$		$\|Ax-b\|_{LL_2,\gamma}-\sigma$	
	$\mathrm{SCP_{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$	$\mathrm{SCP_{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$	$\mathrm{SCP_{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$
2	10.0	0.6 (11.1)	1.3e-01	6.5e-02	-2e-07	-8e-08
4	52.4	2.0 (57.5)	1.3e-01	6.6e-02	-6e-07	-2e-07
6	87.3	4.1 (100.9)	1.3e-01	6.6e-02	-9e-07	-2e-07
8	281.6	7.0 (312.1)	1.3e-01	6.5e-02	-1e-06	-3e-07
10	285.5	11.4 (339.5)	1.3e-01	6.5e-02	-2e-06	-4e-07

Table: *tol* $= 10^{-3}$ for $\mathrm{SCP_{ls}}$ and *tol* $= 10^{-6}$ for $\mathrm{MBA}_{\ell_1/\ell_2}$

i	CPU		$\frac{\|x-\tilde{x}\|}{\max\{1,\|\tilde{x}\|\}}$		$\|Ax-b\|_{LL_2,\gamma}-\sigma$	
	$\mathrm{SCP_{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$	$\mathrm{SCP_{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$	$\mathrm{SCP_{ls}}$	$\mathrm{MBA}_{\ell_1/\ell_2}$
2	3.0	50.8 (54.3)	1.8e+00	1.6e+00	-3e+01	-6e-05
4	11.8	457.6 (472.5)	4.3e+00	4.2e+00	-1e+02	-5e-04
6	30.5	4.9 (44.9)	2.1e-01	6.6e-02	-9e-01	-2e-07
8	37.7	78.5 (139.2)	9.7e+00	9.6e+00	-6e+01	-9e-03
10	71.9	3164.0 (3277.6)	2.1e+00	1.7e+00	-1e+02	-2e-04

Numerical simulations II

- Solve
$$\min_x \frac{\|x\|_1}{\|x\|} \text{ subject to } \|Ax - b\|^2 \le \sigma^2.$$
- Generate $A = [a_1, \cdots, a_n] \in \mathbb{R}^{m \times n}$ with
$$a_j = \frac{1}{\sqrt{m}} \cos\left(\frac{2\pi w j}{F}\right), \quad j = 1, \ldots, m,$$
where w has i.i.d. entries uniformly chosen in $[0, 1]$.
- Badly scaled instances: Generate $\tilde{x} \in \mathbb{R}^n$ in MATLAB by:
```
I = randperm(n); J = I(1:k); tx = zeros(n,1);
tx(J) = sign(randn(k,1)).*10.^(D*rand(k,1));
```
- Set $b = A\tilde{x} + \hat{n}$, where $\hat{n} \sim N(0, 0.01^2 I)$, and set $\sigma = 1.2\|\hat{n}\|$.
- Initialize at an approximate solution computed by SPGL1, backtrack to feasibility if necessary.
- Terminate when $\|x^t - x^{t-1}\| \le 10^{-8} \cdot \max\{1, \|x^t\|\}$.

Numerical simulations II

Table: Random tests on badly scaled CS problems with Gaussian noise

k	F	D	CPU		$\frac{\|x-\tilde{x}\|}{\max\{1,\|\tilde{x}\|\}}$		$\|Ax-b\|^2-\sigma^2$	
			SPGL1	MBA$_{\ell_1/\ell_2}$	SPGL1	MBA$_{\ell_1/\ell_2}$	SPGL1	MBA$_{\ell_1/\ell_2}$
8	5	2	0.07	0.13 (0.20)	3.2e-02	2.3e-03	-4e-05	-1e-13
8	5	3	0.06	0.14 (0.20)	3.2e-03	6.8e-04	-4e-05	-2e-11
8	15	2	0.08	3.92 (4.01)	4.7e-01	1.5e-01	-9e-05	-7e-13
8	15	3	0.11	31.46 (31.58)	3.8e-01	5.3e-02	2e-02	-5e-11
12	5	2	0.06	2.26 (2.32)	1.4e-01	3.6e-02	-3e-04	-8e-13
12	5	3	0.08	4.05 (4.14)	6.0e-02	3.8e-03	1e-04	-7e-11
12	15	2	0.09	8.32 (8.41)	5.2e-01	2.0e-01	-1e-04	-1e-12
12	15	3	0.11	403.80 (403.91)	5.2e-01	1.5e+00	6e-02	-3e-10

Conclusion and future work

Conclusion:

- We established convergence rate of a Dinkelbach type algorithm for noiseless compressed sensing based on ℓ_1/ℓ_2 minimization via new KL calculus rules (for fractional objectives).
- We proposed and analyzed convergence of $\mathrm{MBA}_{\ell_1/\ell_2}$ for ℓ_1/ℓ_2 minimization subject to measurement noise.

Conclusion and future work

Conclusion:

- We established convergence rate of a Dinkelbach type algorithm for noiseless compressed sensing based on ℓ_1/ℓ_2 minimization via new KL calculus rules (for fractional objectives).
- We proposed and analyzed convergence of $\text{MBA}_{\ell_1/\ell_2}$ for ℓ_1/ℓ_2 minimization subject to measurement noise.

Future work:

- Other fractional objectives? (Boţ, Dao, Li '21)
- KL-type analysis for inexactly solved subproblems.

References:

- L. Zeng, P. Yu, and T. K. Pong. Analysis and algorithms for some compressed sensing models based on L1/L2 minimization. *SIAM J. Optim.*, 31 (2021), pp. 1576–1603.

Thanks for coming! ☺

References

H. Attouch, J. Bolte, P. Redont, and A. Soubeyran.
Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality.
Math. Oper. Res., 35 (2010), pp. 438–457.

H. Attouch, J. Bolte, and B. F. Svaiter.
Convergence of descent methods for semialgebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods.
Math. Program., 137 (2013), pp. 91–129.

A. Auslender, R. Shefi, and M. Teboulle.
A moving balls approximation method for a class of smooth constrained minimization problems.
SIAM J. Optim., 20 (2010), pp. 3232–3259.

References

J. Bolte, Z. Chen, and E. Pauwels.
The multiproximal linearization method for convex composite problems.
Math. Program., 186 (2020), pp. 1–36, doi:10.1007/s10107-019-01382-3.

J. Bolte, A. Daniilidis, and A. Lewis.
The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamic systems.
SIAM. J. Optim., 17 (2007), pp. 1205–1223.

J. Bolte and E. Pauwels.
Majorization-minimization procedures and convergence of SQP methods for semi-algebraic and tame programs.
Math. Oper. Res., 41 (2016), pp. 442–465.

References

J. Bolte, S. Sabach, and M. Teboulle.
Proximal alternating linearized minimization for nonconvex and nonsmooth problems.
Math. Program., 146 (2014), pp. 459–494.

R. I. Boţ, M. N. Dao, and G. Li.
Inertial proximal block coordinate method for a class of nonsmooth and nonconvex sum-of-ratios optimization problems.
Preprint (2020). Available at https://arxiv.org/pdf/2011.09782.pdf.

R. E. Carrillo, K. E. Barner, and T. C. Aysal.
Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise.
IEEE J. Sel. Topics Signal Process., 4 (2010), pp. 392–408.

References

G. Li and T. K. Pong.
Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods.
Found. Comput. Math., 18 (2018), pp. 1199–1232.

T. Liu, T. K. Pong, and A. Takeda.
A refined convergence analysis of pDCA$_e$ with applications to simultaneous sparse recovery and outlier detection.
Comput. Optim. Appl., 73 (2019), pp. 69–100.

Z. Lu,
Sequential convex programming methods for a class of structured nonlinear programming.
Preprint (2012). Available at https://arxiv.org/abs/1210.3039.

References

Y. Rahimi, C. Wang, H. Dong, and Y. Lou.
A scale invariant approach for sparse signal recovery.
SIAM J. Sci. Comput., 41 (2019), pp. A3649–A3672.

S. A. Vavasis.
Derivation of Compressive Sensing Theorems from the Spherical Section Property.
University of Waterloo, 2009.

C. Wang, M. Tao, J. Nagy, and Y. Lou.
Limited-angle CT reconstruction via the L1/L2 minimization.
SIAM J. Imaging Sci., 14 (2021), pp. 749–777.

C. Wang, M. Yan, and Y. Lou.
Accelerated schemes for the L1/L2 minimization.
IEEE Trans. Signal Process., 68 (2020), pp. 2660–2669.

References

Y. Xu, A. Narayan, H. Tran, and C. G. Webster.
Analysis of the ratio of ℓ_1 and ℓ_2 norms in compressed sensing.
Appl. Comput. Harmon. Anal., 55 (2021), pp. 486–511.

P. Yin, E. Esser, and J. Xin.
Ratio and difference of ℓ_1 and ℓ_2 norms and sparse representation with coherent dictionaries.
Commun. Inf. Syst., 14 (2014), pp. 87–109.

L. Zeng, P. Yu, and T. K. Pong.
Analysis and algorithms for some compressed sensing models based on L1/L2 minimization.
SIAM J. Optim., 31 (2021), pp. 1576–1603.

Y. Zhang.
Theory of compressive sensing via L1-minimization: A non-RIP analysis and extensions.
J. Oper. Res. Soc. China, 1 (2013), pp. 79–105.