Convergence rate analysis of a Dykstra-type projection algorithm

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

pre-SIAM Optimization Workshop
University of Washington
May 2023
(Joint work with Xiaozhou Wang)

Motivating applications

- Nearest correlation matrices: (Higham '02)

$$\underset{x\in S^n}{\text{Minimize}}\ \|x-\bar{v}\|_F^2 \ \text{ subject to } \ x\in\cap_{i=1}^3 C_i,$$

where $\bar{v}\in S^n$, $C_1 = S^n_+$, $C_2=\{x\in S^n:\ L_{ij}\le x_{ij}\le U_{ij}\ \forall i,j\}$ and

$$C_3 := \{x\in S^n:\ x_{ii}=1\ \forall i\}.$$

- System identification: (Liu et al. '20)

$$\underset{x_1,\ldots,x_m\in\mathbb{R}^n}{\text{Minimize}}\ \sum_{i=1}^m \|x_i-\bar{v}_i\|^2 \ \text{ subject to } \ x:=(x_1,\ldots,x_m)\in D_1\cap D_2,$$

where $\bar{v}$ is given, and

$$D_1 := \{x\in\mathbb{R}^{mn}: \|\mathcal{H}_{r_{i+1}}(x_i)\|_*\le k_i, i=1,\ldots,m\},$$
$$D_2 := \{x\in\mathbb{R}^{mn}: \|[\mathcal{H}_{r+1}(x_1)\ \cdots\ \mathcal{H}_{r+1}(x_m)]\|_*\le k\},$$

with $\mathcal{H}_s$ being a linear map that returns a suitable Hankel matrix, $\|\cdot\|_*$ is the nuclear norm, $r_i,\ r\in\mathbb{N}$, and $k_i,\ k>0$.

Best approximation problems

Consider the following best approximation problem:

$$\operatorname*{Minimize}_{x\in\mathbb{R}^n} \ \frac{1}{2}\|x-\bar{v}\|^2 \ \text{ subject to } \ x \in \bigcap_{i=1}^{\ell} A_i^{-1}C_i,$$

where

- $A_i : \mathbb{R}^n \to \mathbb{R}^{m_i}$ is linear and nonzero;
- each $C_i \subseteq \mathbb{R}^{m_i}$ is closed and convex;
- projection onto C_i can be computed more efficiently than projection onto $A_i^{-1}C_i$;
- For technical succintness:

$$\bigcap_{i=1}^{\ell} A_i^{-1}C_i \neq \emptyset.$$

When $A_i = I$

Dykstra's projection algorithm: (Boyle, Dykstra '86, Han '88)

Set $x_\ell^0 = \bar{v}$, $y_1^0 = \cdots = y_\ell^0 = 0$. For each $t \geq 0$, set $x_0^{t+1} = x_\ell^t$ and compute

$$x_i^{t+1} = \mathrm{Proj}_{C_i}(y_i^t + x_{i-1}^{t+1}), \quad y_i^{t+1} = y_i^t + x_{i-1}^{t+1} - x_i^{t+1} \text{ for } i = 1, \ldots, \ell.$$

When $A_i = I$

Dykstra's projection algorithm: (Boyle, Dykstra '86, Han '88)

Set $x_\ell^0 = \bar{v}$, $y_1^0 = \cdots = y_\ell^0 = 0$. For each $t \geq 0$, set $x_0^{t+1} = x_\ell^t$ and compute

$$x_i^{t+1} = \mathrm{Proj}_{C_i}(y_i^t + x_{i-1}^{t+1}), \quad y_i^{t+1} = y_i^t + x_{i-1}^{t+1} - x_i^{t+1} \text{ for } i = 1, \ldots, \ell.$$

Known facts:

- $x_\ell^t =: x^t \to \mathrm{Proj}_{\cap_{i=1}^\ell C_i}(\bar{v})$.
- Reduces to cyclic projection when each C_i is affine. (Gaffke, Mathar '89).
- Equivalent to CGD applied to

$$\underset{y_1,\ldots,y_\ell}{\text{Minimize}} \ \frac{1}{2}\left\|\sum_{i=1}^{\ell} y_i - \bar{v}\right\|^2 - \frac{1}{2}\|\bar{v}\|^2 + \sum_{i=1}^{\ell} \sigma_{C_i}(y_i)$$

starting from $y_1^0 = \cdots = y_\ell^0 = 0$.

When $A_i = I$ cont.

Dykstra's projection algorithm: (Boyle, Dykstra '86, Han '88)

Set $x_\ell^0 = \bar{v}$, $y_1^0 = \cdots = y_\ell^0 = 0$. For each $t \geq 0$, set $x_0^{t+1} = x_\ell^t$ and compute

$$x_i^{t+1} = \mathrm{Proj}_{C_i}(y_i^t + x_{i-1}^{t+1}), \quad y_i^{t+1} = y_i^t + x_{i-1}^{t+1} - x_i^{t+1} \text{ for } i = 1, \ldots, \ell.$$

Known facts cont.:

- Local linear convergence of $\{x_\ell^t\}$ and all $\{y_i^t\}$ when each C_i is polyhedral. (Luo, Tseng '93)
- Convergence rate unknown for general C_i.

When $A_i = I$ cont.

Dykstra's projection algorithm: (Boyle, Dykstra '86, Han '88)

Set $x_\ell^0 = \bar{v}$, $y_1^0 = \cdots = y_\ell^0 = 0$. For each $t \geq 0$, set $x_0^{t+1} = x_\ell^t$ and compute

$$x_i^{t+1} = \mathrm{Proj}_{C_i}(y_i^t + x_{i-1}^{t+1}), \quad y_i^{t+1} = y_i^t + x_{i-1}^{t+1} - x_i^{t+1} \text{ for } i = 1, \ldots, \ell.$$

Known facts cont.:

- Local linear convergence of $\{x_\ell^t\}$ and all $\{y_i^t\}$ when each C_i is polyhedral. (Luo, Tseng '93)
- Convergence rate unknown for general C_i.

Outline:

- Develop Dykstra-type projection algorithm when $A_i \neq I$.
- Identify a class of sets C_i for convergence rate analysis.
- Explicit error bounds and convergence rate, with "tightness" examples.

A Dykstra-type algorithm

Dykstra-type projection algorithm:
Set $y_i^0 = 0 \in \mathbb{R}^{m_i}$ and $\gamma_i := \lambda_{\max}(A_i^T A_i)$ for all i, $x_\ell^0 = \bar{v} \in \mathbb{R}^n$.
For each $t \geq 0$, set $x_0^{t+1} = x_\ell^t$ and $x^t = x_\ell^t$. Compute, for $i = 1, \ldots, \ell$,

$$x_i^{t+1} = (I - \gamma_i^{-1} A_i^T A_i) x_{i-1}^{t+1} + \gamma_i^{-1} A_i^T \mathrm{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}),$$
$$y_i^{t+1} = y_i^t + \gamma_i^{-1} A_i x_{i-1}^{t+1} - \gamma_i^{-1} \mathrm{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}).$$

A Dykstra-type algorithm

Dykstra-type projection algorithm:
Set $y_i^0 = 0 \in \mathbb{R}^{m_i}$ and $\gamma_i := \lambda_{\max}(A_i^T A_i)$ for all i, $x_\ell^0 = \bar{v} \in \mathbb{R}^n$.
For each $t \geq 0$, set $x_0^{t+1} = x_\ell^t$ and $x^t = x_\ell^t$. Compute, for $i = 1, \ldots, \ell$,

$$x_i^{t+1} = (I - \gamma_i^{-1} A_i^T A_i) x_{i-1}^{t+1} + \gamma_i^{-1} A_i^T \text{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}),$$
$$y_i^{t+1} = y_i^t + \gamma_i^{-1} A_i x_{i-1}^{t+1} - \gamma_i^{-1} \text{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}).$$

For notational simplicity, write $\boldsymbol{y}^t := (y_1^t, \ldots, y_\ell^t)$ for all t.

A Dykstra-type algorithm

Dykstra-type projection algorithm:
Set $y_i^0 = 0 \in \mathbb{R}^{m_i}$ and $\gamma_i := \lambda_{\max}(A_i^T A_i)$ for all i, $x_\ell^0 = \bar{v} \in \mathbb{R}^n$.
For each $t \geq 0$, set $x_0^{t+1} = x_\ell^t$ and $x^t = x_\ell^t$. Compute, for $i = 1, \ldots, \ell$,

$$x_i^{t+1} = (I - \gamma_i^{-1} A_i^T A_i) x_{i-1}^{t+1} + \gamma_i^{-1} A_i^T \mathrm{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}),$$
$$y_i^{t+1} = y_i^t + \gamma_i^{-1} A_i x_{i-1}^{t+1} - \gamma_i^{-1} \mathrm{Proj}_{C_i}(\gamma_i y_i^t + A_i x_{i-1}^{t+1}).$$

For notational simplicity, write $\boldsymbol{y}^t := (y_1^t, \ldots, y_\ell^t)$ for all t.

- The algorithm is equivalent to a proximal CGD applied to

$$\underset{y_1,\ldots,y_\ell}{\text{Minimize}} \;\; d(\boldsymbol{y}) := \frac{1}{2}\left\|\sum_{i=1}^{\ell} A_i^T y_i - \bar{v}\right\|^2 - \frac{1}{2}\|\bar{v}\|^2 + \sum_{i=1}^{\ell} \sigma_{C_i}(y_i)$$

starting from $y_i^0 = 0$ for all i.

A Dykstra-type algorithm cont.

Key facts:

- It holds that

$$-\inf_{\boldsymbol{y}} d(\boldsymbol{y}) = \inf_{x} \left\{ \frac{1}{2}\|x - \bar{v}\|^2 : \; x \in \cap_{i=1}^{\ell} A_i^{-1} C_i \right\}.$$

A Dykstra-type algorithm cont.

Key facts:

- It holds that

$$-\inf_{\boldsymbol{y}} d(\boldsymbol{y}) = \inf_{x} \left\{ \frac{1}{2}\|x - \bar{v}\|^2 : \ x \in \cap_{i=1}^{\ell} A_i^{-1} C_i \right\}.$$

- $\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\| \to 0$ and $\mathrm{dist}(\mathbf{0}, \partial d(\boldsymbol{y}^t)) \to 0$.
- Every accumulation point of $\{\boldsymbol{y}^t\}$ minimizes d.

A Dykstra-type algorithm cont.

Key facts:

- It holds that

$$-\inf_{\boldsymbol{y}} d(\boldsymbol{y}) = \inf_{x} \left\{ \frac{1}{2}\|x - \bar{v}\|^2 : \; x \in \cap_{i=1}^{\ell} A_i^{-1} C_i \right\}.$$

- $\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\| \to 0$ and $\mathrm{dist}(\mathbf{0}, \partial d(\boldsymbol{y}^t)) \to 0$.
- Every accumulation point of $\{\boldsymbol{y}^t\}$ minimizes d.
- If $\boldsymbol{y}^* \in \mathrm{Arg\,min}\, d$, then $\mathrm{Proj}_{\cap_{i=1}^{\ell} A_i^{-1} C_i}(\bar{v}) = \bar{v} - \sum_{i=1}^{\ell} A_i^T y_i^*$.

A Dykstra-type algorithm cont.

Key facts:

- It holds that

$$-\inf_{\boldsymbol{y}} d(\boldsymbol{y}) = \inf_{x} \left\{ \frac{1}{2}\|x - \bar{v}\|^2 : \ x \in \cap_{i=1}^{\ell} A_i^{-1} C_i \right\}.$$

- $\|\boldsymbol{y}^{t+1} - \boldsymbol{y}^t\| \to 0$ and $\operatorname{dist}(\boldsymbol{0}, \partial d(\boldsymbol{y}^t)) \to 0$.
- Every accumulation point of $\{\boldsymbol{y}^t\}$ minimizes d.
- If $\boldsymbol{y}^* \in \operatorname{Arg\,min} d$, then $\operatorname{Proj}_{\cap_{i=1}^{\ell} A_i^{-1} C_i}(\bar{v}) = \bar{v} - \sum_{i=1}^{\ell} A_i^T y_i^*$.
- (Auslender, Cominetti, Crouziex '93) Suppose that $\cap_{i=1}^{\ell} A_i^{-1} \operatorname{ri} C_i \neq \emptyset$. Then $\operatorname{Arg\,min} d = E_1 + E_2 \neq \emptyset$, where E_1 is compact and E_2 is a subspace. Moreover, we have

$$d(\boldsymbol{y}^t) \to \inf d, \qquad \operatorname{dist}(\boldsymbol{y}^t, \operatorname{Arg\,min} d) \to 0,$$

  and $d(\boldsymbol{y} + \boldsymbol{u}) = d(\boldsymbol{y})$ for all $\boldsymbol{y} \in \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_\ell}$ and $\boldsymbol{u} \in E_2$.

$C^{1,\alpha}$-cone reducibility

Let $\mathbb{X}$ and $\mathbb{Y}$ be two finite dimensional Hilbert spaces.

Definition: Let $\alpha \in (0,1]$. A closed set $\Omega \subseteq \mathbb{X}$ is said to be $C^{1,\alpha}$-cone reducible at $\hat{x} \in \Omega$ if $\exists\ \rho > 0$, a closed convex pointed cone $K \subseteq \mathbb{Y}$ and a mapping $\Xi : \mathbb{X} \to \mathbb{Y}$ satisfies $\Xi(\hat{x}) = 0$ and is $C^{1,\alpha}$ in $B(\hat{x}, \rho)$ with $D\Xi(\hat{x})$ being surjective, and moreover

$$\Omega \cap B(\hat{x}, \rho) = \{x : \Xi(x) \in K\} \cap B(\hat{x}, \rho).$$

We say that Ω is $C^{1,\alpha}$-cone reducible if it is so at each $x \in \Omega$.

$C^{1,\alpha}$-cone reducibility

Let $\mathbb{X}$ and $\mathbb{Y}$ be two finite dimensional Hilbert spaces.

Definition: Let $\alpha \in (0,1]$. A closed set $\Omega \subseteq \mathbb{X}$ is said to be $C^{1,\alpha}$-cone reducible at $\hat{x} \in \Omega$ if $\exists\ \rho > 0$, a closed convex pointed cone $K \subseteq \mathbb{Y}$ and a mapping $\Xi : \mathbb{X} \to \mathbb{Y}$ satisfies $\Xi(\hat{x}) = 0$ and is $C^{1,\alpha}$ in $B(\hat{x}, \rho)$ with $D\Xi(\hat{x})$ being surjective, and moreover

$$\Omega \cap B(\hat{x}, \rho) = \{x : \Xi(x) \in K\} \cap B(\hat{x}, \rho).$$

We say that Ω is $C^{1,\alpha}$-cone reducible if it is so at each $x \in \Omega$.

Examples:

- A set C is $C^{1,1}$-cone reducible at any $x \in \operatorname{int} C$: just take $\mathbb{Y} = \{0\}$.

$C^{1,\alpha}$-cone reducibility

Let $\mathbb{X}$ and $\mathbb{Y}$ be two finite dimensional Hilbert spaces.

Definition: Let $\alpha \in (0,1]$. A closed set $\Omega \subseteq \mathbb{X}$ is said to be $C^{1,\alpha}$-cone reducible at $\hat{x} \in \Omega$ if $\exists\ \rho > 0$, a closed convex pointed cone $K \subseteq \mathbb{Y}$ and a mapping $\Xi : \mathbb{X} \to \mathbb{Y}$ satisfies $\Xi(\hat{x}) = 0$ and is $C^{1,\alpha}$ in $B(\hat{x}, \rho)$ with $D\Xi(\hat{x})$ being surjective, and moreover

$$\Omega \cap B(\hat{x}, \rho) = \{x : \Xi(x) \in K\} \cap B(\hat{x}, \rho).$$

We say that Ω is $C^{1,\alpha}$-cone reducible if it is so at each $x \in \Omega$.

Examples:

- A set C is $C^{1,1}$-cone reducible at any $x \in \operatorname{int} C$: just take $\mathbb{Y} = \{0\}$. Note that $C^{1,1}$-cone reducibility $\Longrightarrow$ $C^{1,\alpha}$-cone reducibility for any $\alpha \in (0,1]$.

$C^{1,\alpha}$-cone reducibility

Let $\mathbb{X}$ and $\mathbb{Y}$ be two finite dimensional Hilbert spaces.

Definition: Let $\alpha \in (0, 1]$. A closed set $\Omega \subseteq \mathbb{X}$ is said to be $C^{1,\alpha}$-cone reducible at $\hat{x} \in \Omega$ if $\exists\ \rho > 0$, a closed convex pointed cone $K \subseteq \mathbb{Y}$ and a mapping $\Xi : \mathbb{X} \to \mathbb{Y}$ satisfies $\Xi(\hat{x}) = 0$ and is $C^{1,\alpha}$ in $B(\hat{x}, \rho)$ with $D\Xi(\hat{x})$ being surjective, and moreover

$$\Omega \cap B(\hat{x}, \rho) = \{x : \Xi(x) \in K\} \cap B(\hat{x}, \rho).$$

We say that Ω is $C^{1,\alpha}$-cone reducible if it is so at each $x \in \Omega$.

Examples:

- A set C is $C^{1,1}$-cone reducible at any $x \in \operatorname{int} C$: just take $\mathbb{Y} = \{0\}$. Note that $C^{1,1}$-cone reducibility $\Longrightarrow$ $C^{1,\alpha}$-cone reducibility for any $\alpha \in (0, 1]$.
- A closed convex pointed cone is $C^{1,1}$-cone reducible at the origin.

$C^{1,\alpha}$-cone reducibility cont.

Examples cont.:

- Let C be closed and convex, and B be the unit ball, then for any $\epsilon > 0$, the set $C + \epsilon B$ is $C^{1,1}$-cone reducible.

$C^{1,\alpha}$-cone reducibility cont.

Examples cont.:

- Let C be closed and convex, and B be the unit ball, then for any $\epsilon > 0$, the set $C + \epsilon B$ is $C^{1,1}$-cone reducible. Indeed,

$$C + \epsilon B = \{x : \text{dist}(x, C)^2 - \epsilon^2 \leq 0\};$$

  for boundary points, take $K = \mathbb{R}_-$ and $\Xi(x) := \text{dist}(x, C)^2 - \epsilon^2$.

$C^{1,\alpha}$-cone reducibility cont.

Examples cont.:

- Let C be closed and convex, and B be the unit ball, then for any $\epsilon > 0$, the set $C + \epsilon B$ is $C^{1,1}$-cone reducible. Indeed,

$$C + \epsilon B = \{x : \operatorname{dist}(x, C)^2 - \epsilon^2 \leq 0\};$$

  for boundary points, take $K = \mathbb{R}_-$ and $\Xi(x) := \operatorname{dist}(x, C)^2 - \epsilon^2$.
- Let $p \in (1, \infty)$ and let $C = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_p \leq r\}$. Then C is $C^{1,\alpha}$-cone reducible with $\alpha = \min\{1, p - 1\}$.

$C^{1,\alpha}$-cone reducibility cont.

Examples cont.:

- Let C be closed and convex, and B be the unit ball, then for any $\epsilon > 0$, the set $C + \epsilon B$ is $C^{1,1}$-cone reducible. Indeed,

$$C + \epsilon B = \{x : \operatorname{dist}(x, C)^2 - \epsilon^2 \leq 0\};$$

  for boundary points, take $K = \mathbb{R}_-$ and $\Xi(x) := \operatorname{dist}(x, C)^2 - \epsilon^2$.

- Let $p \in (1, \infty)$ and let $C = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_p \leq r\}$. Then C is $C^{1,\alpha}$-cone reducible with $\alpha = \min\{1, p-1\}$. Indeed, if we let $G(x, r) := \|x\|_p - r$, then $\nabla G(x, r)$ is

$$\|x\|_p^{1-p} \cdot \begin{bmatrix} \operatorname{sgn}(x_1)|x_1|^{p-1} & \cdots & \operatorname{sgn}(x_n)|x_n|^{p-1} & -\|x\|_p^{p-1} \end{bmatrix}^T,$$

  which is nonzero at any nonzero boundary points, and is $\min\{1, p-1\}$- Hölder continuous.

$C^{1,\alpha}$-cone reducibility cont.

Examples cont.:

- Let C be closed and convex, and B be the unit ball, then for any $\epsilon > 0$, the set $C + \epsilon B$ is $C^{1,1}$-cone reducible. Indeed,

$$C + \epsilon B = \{x : \text{ dist}(x, C)^2 - \epsilon^2 \leq 0\};$$

  for boundary points, take $K = \mathbb{R}_-$ and $\Xi(x) := \text{dist}(x, C)^2 - \epsilon^2$.

- Let $p \in (1, \infty)$ and let $C = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : \ \|x\|_p \leq r\}$. Then C is $C^{1,\alpha}$-cone reducible with $\alpha = \min\{1, p-1\}$. Indeed, if we let $G(x, r) := \|x\|_p - r$, then $\nabla G(x, r)$ is

$$\|x\|_p^{1-p} \cdot \begin{bmatrix} \text{sgn}(x_1)|x_1|^{p-1} & \cdots & \text{sgn}(x_n)|x_n|^{p-1} & -\|x\|_p^{p-1} \end{bmatrix}^T,$$

  which is nonzero at any nonzero boundary points, and is $\min\{1, p-1\}$- Hölder continuous.

- Let $p \in (1, \infty)$ and let $C = \{x \in \mathbb{R}^n : \ \|x\|_p \leq 1\}$. Then C is $C^{1,\alpha}$-cone reducible with $\alpha = \min\{1, p-1\}$.

Error bound

Theorem 1. (Wang, P. '22)
For the best approximation problem

$$\operatorname*{Minimize}_{x\in\mathbb{R}^n}\ 0.5\|x-\bar{v}\|^2 \ \text{ subject to }\ x\in\cap_{i=1}^{\ell}A_i^{-1}C_i,$$

suppose that

(i) Each C_i is $C^{1,\alpha}$-cone reducible with $\alpha\in(0,1]$, closed & convex;

(ii) $\cap_{i=1}^{\ell}A_i^{-1}\operatorname{ri}C_i\neq\emptyset$;

(iii) $0\in x^*-\bar{v}+\operatorname{ri}\partial(\sum_{i=1}^{\ell}\delta_{A_i^{-1}C_i})(x^*)$, where $x^*=\operatorname{Proj}_{\cap_{i=1}^{\ell}A_i^{-1}C_i}(\bar{v})$.

Then there exist $\epsilon>0$ and $c>0$ such that

$$\operatorname{dist}(\boldsymbol{y},\operatorname{Arg\,min}d)\leq c\,(d(\boldsymbol{y})-\inf d)^{1-\frac{1}{1+\alpha}}$$

whenever $\boldsymbol{y}$ satisfies $\operatorname{dist}(\boldsymbol{y},\operatorname{Arg\,min}d)\leq\epsilon$ & $\inf d\leq d(\boldsymbol{y})\leq\inf d+\epsilon$.

Example: tightness of exponent

Example: Let $\bar{v} = (2, 0)$ and consider

$$\underset{x\in\mathbb{R}^2}{\text{Minimize}} \ \ \|x - \bar{v}\|^2/2 \ \text{ subject to } \ A_1 x \in C_1,$$

where $A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $C_1 = \{x \in \mathbb{R}^2 : \|x\|_p \le 1\}$ and $p \in (1, 2]$.

Example: tightness of exponent

Example: Let $\bar{v} = (2, 0)$ and consider

$$\underset{x \in \mathbb{R}^2}{\text{Minimize}} \ \|x - \bar{v}\|^2/2 \ \text{ subject to } \ A_1 x \in C_1,$$

where $A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $C_1 = \{x \in \mathbb{R}^2 : \|x\|_p \leq 1\}$ and $p \in (1, 2]$.
The feasible region is $[-1, 1] \times \mathbb{R}$, hence the solution is $x^* = (1, 0)$.

Also, C_1 is $C^{1,p-1}$-cone reducible, $A_1^{-1}\text{ri}\, C_1 \neq \emptyset$ and

$$\begin{bmatrix} 1 & 0 \end{bmatrix}^T = \bar{v} - x^* \in \left\{ \begin{bmatrix} t & 0 \end{bmatrix}^T : t > 0 \right\} = \text{ri}\, \mathcal{N}_{A_1^{-1} C_1}(x^*).$$

Then $1 - \frac{1}{1+\alpha} = 1 - \frac{1}{p} =: \frac{1}{q}$, and **Theorem 1** shows that

$$\text{dist}(y_1, \text{Arg}\min d) \leq c\,(d(y_1) - \inf d)^{\frac{1}{q}}$$

whenever y_1 satisfies $\text{dist}(y_1, \text{Arg}\min d) \leq \epsilon$ & $\inf d \leq d(y_1) \leq \inf d + \epsilon$.

Example: tightness of exponent cont.

Example cont.: Now, note that

$$d(y_1) = (1/2)\|A_1 y_1 - \bar{v}\|^2 - (1/2)\|\bar{v}\|^2 + \|y_1\|_q.$$

Moreover, from duality,

$$\begin{bmatrix} 1 & 0 \end{bmatrix}^T = \bar{v} - x^* = A_1 \hat{y}_1$$

whenever $\hat{y}_1 \in \operatorname{Arg\,min} d$. Thus, the 1st coordinate of $\hat{y}_1$ is 1. Moreover, the 2nd coordinate of $\hat{y}_1$ is 0. Consequently,

$$\operatorname{Arg\,min} d = \{(1, 0)\}.$$

Now, let $y_1^\epsilon = (1, \epsilon)$ for $\epsilon \downarrow 0$. Then we obtain as $\epsilon \downarrow 0$ that

$$d(y_1^\epsilon) - d(\hat{y}_1) = (1 + \epsilon^q)^{\frac{1}{q}} - 1 = \Theta(\epsilon^q) \text{ and } \operatorname{dist}(y_1^\epsilon, \operatorname{Arg\,min} d) = \epsilon,$$

Consequently, $\operatorname{dist}(y_1^\epsilon, \operatorname{Arg\,min} d) = \Theta([d(y_1^\epsilon) - \inf d]^{\frac{1}{q}})$ as $\epsilon \downarrow 0$.

Convergence rate

Theorem 2. (Wang, P. '22)
For the best approximation problem, suppose that

(i) Each C_i is $C^{1,\alpha}$-cone reducible with $\alpha \in (0,1]$, closed & convex;

(ii) $\cap_{i=1}^{\ell} A_i^{-1} \operatorname{ri} C_i \neq \emptyset$;

(iii) $0 \in x^* - \bar{v} + \operatorname{ri} \partial(\sum_{i=1}^{\ell} \delta_{A_i^{-1} C_i})(x^*)$, where $x^* = \operatorname{Proj}_{\cap_{i=1}^{\ell} A_i^{-1} C_i}(\bar{v})$.

Let $\{x^t\}$ and $\{\boldsymbol{y}^t\}$ be generated by the Dykstra-type algorithm. Then there exist $c > 0$ and $\bar{t} \in \mathbb{N}$ such that

- If $\alpha = 1$, then $\exists \boldsymbol{y}^* \in \operatorname{Arg\,min} d$, $\tau \in (0,1)$ such that for any $t \geq \bar{t}$,

$$\|x^t - x^*\| \leq c\tau^t, \quad \|\boldsymbol{y}^t - \boldsymbol{y}^*\| \leq c\tau^t.$$

- If $\alpha \in (0,1)$, then for any $t \geq \bar{t}$,

$$\|x^t - x^*\| \leq ct^{-\alpha^2/2}, \quad d(\boldsymbol{y}^t) - \inf d \leq ct^{-\alpha}, \quad \operatorname{dist}(\boldsymbol{y}^t, \operatorname{Arg\,min} d) \leq ct^{-\alpha^2/2}.$$

Conclusion and future work

Conclusion:

- Explicit convergence rate of a Dykstra-type method is deduced for $C^{1,\alpha}$-cone reducible sets, $\alpha \in (0, 1]$.
- An example was constructed to illustrate the tightness of an essential error bound condition.

Reference:

- X. Wang and T. K. Pong.
  Convergence rate analysis of a Dykstra-type projection algorithm.
  Preprint. Available at https://arxiv.org/abs/2301.03026.

Thanks for coming!