Linear Convergence of Bregman First-Order Methods: applications to Poisson Inverse Problems

Jonathan Chirinos-Rodríguez, Christian Daniele, Cédric Févotte and Emmanuel Soubies

July 9, 2026

EUROPT 26, Linz

Co-funded by the European Union (ERC, MALIN, 10117133). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the EU nor the granting authority can be held responsible for them.

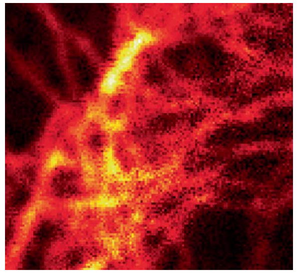
Poisson Inverse Problems

$$\boxed{y = \text{Poisson}(Ax + b)}$$

$A \in \mathbb{R}^{m \times n}_{\geq 0}$
$x \in \mathbb{R}^{n}_{\geq 0}, y \in \mathbb{R}^{m}_{\geq 0}, b \geq 0$

Motivation: Imaging with low photon count (e.g. Microscopy) →

[Bertero and Boccacci '98]



$$\boxed{\operatorname*{argmin}_{x \geq 0} \mathrm{KL}(y, Ax + b) + R(x)}$$

R convex

$$\mathrm{KL}(y, Ax + b) = \sum_{m=1}^{M} y_m \log\left(\frac{y_m}{[Ax]_m + b}\right) + [Ax]_m + b - y_m$$

- Convex but **not strongly**
- Coercive
- $L-$smooth if $b > 0$

How to effective handle KL minimization?

UniGe | MaLGa

KL minimization: Proximal Gradient Descent

$$\boxed{\underset{x \geq 0}{\operatorname{argmin}}\, \mathsf{KL}(y, Ax + b) + R(x) =: J(x)}$$

$$x^{k+1} = \underset{x \in \mathbb{R}^n_{\geq 0}}{\operatorname{argmin}} \langle x - x^k, \nabla(\mathsf{KL}(y, A \cdot + b)\rangle + R(x) + \frac{1}{\tau}\|x - x^k\|_2^2$$

[Harmany, Marcia, and Willett '12]

If $b > 0 \implies L \propto \frac{1}{b^2}$

$$\boxed{\text{If } \tau \in (0, 2/L) \implies J(x^k) - \bar{J} = O(1/k)}$$

[Nesterov '13]

[Nesterov '13] $\mu-$strong convexity allows for achieving $O(\rho^k)$, with $\rho := 1 - \mu/L$ rates but **not satisfied by KL**

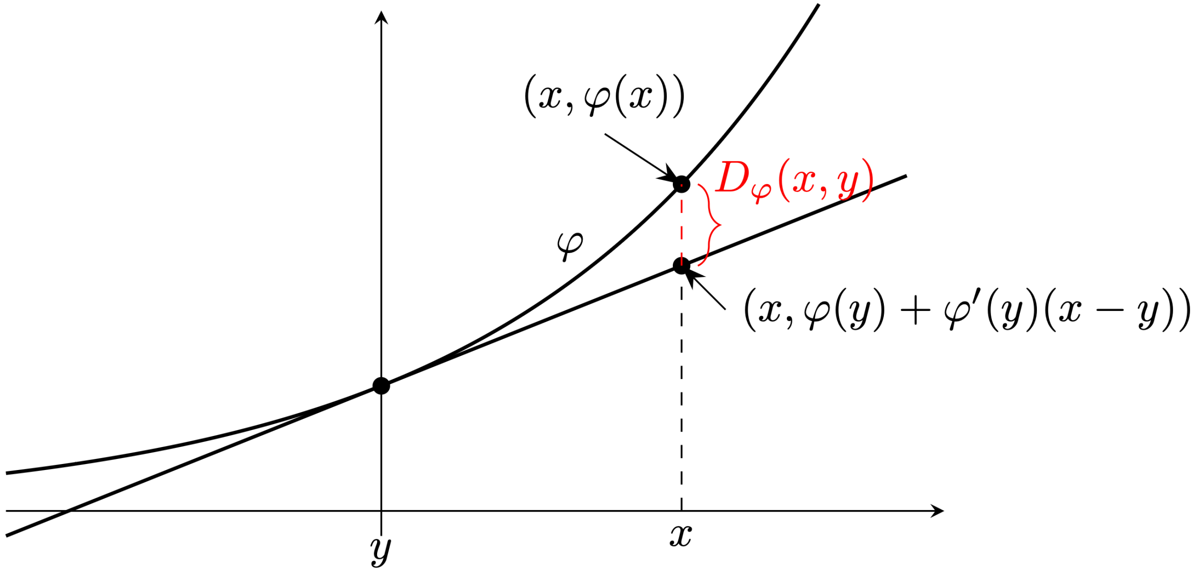
KL minimisation: Bregman Proximal Gradient Descent

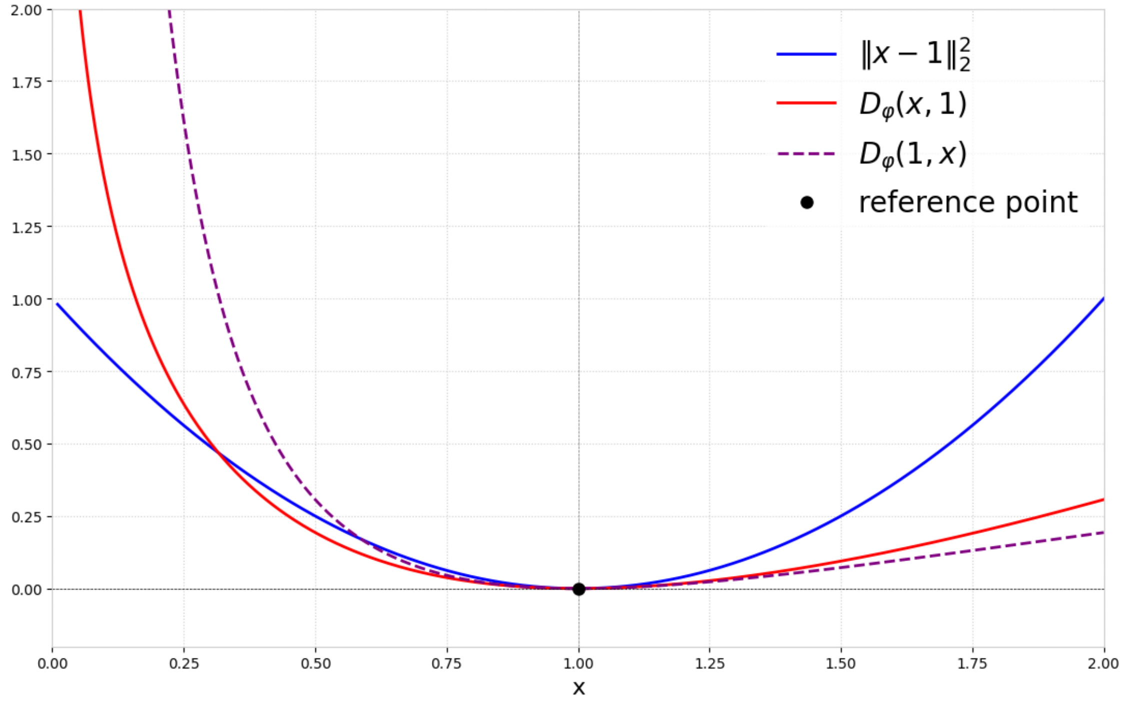
$$\varphi = \frac{1}{2}\|x\|_2^2 \implies$$
$$D_\varphi(x,y) = \frac{1}{2}\|x-y\|_2^2$$

$$\boxed{\underset{x \geq 0}{\operatorname{argmin}}\, \mathsf{KL}(y, Ax+b) + R(x) =: J(x)}$$

Given a Legendre function φ:

$$x^{k+1} = \underset{x \in \mathbb{R}^n_{\geq 0}}{\operatorname{argmin}}\, \langle x - x^k, \nabla(\mathsf{KL}(y, A\cdot + b))\rangle + R(x) + \frac{1}{\tau} D_\varphi(x, x^k)$$





KL minimisation: Bregman Proximal Gradient Descent

$$x^{k+1} = \underset{x \in \mathbb{R}^n_{\geq 0}}{\operatorname{argmin}} \langle x - x^k, \nabla \underbrace{(\mathsf{KL}(y, A \cdot + b))}_{F(x)} \rangle + R(x) + \frac{1}{\tau} D_\varphi(x, x^k)$$

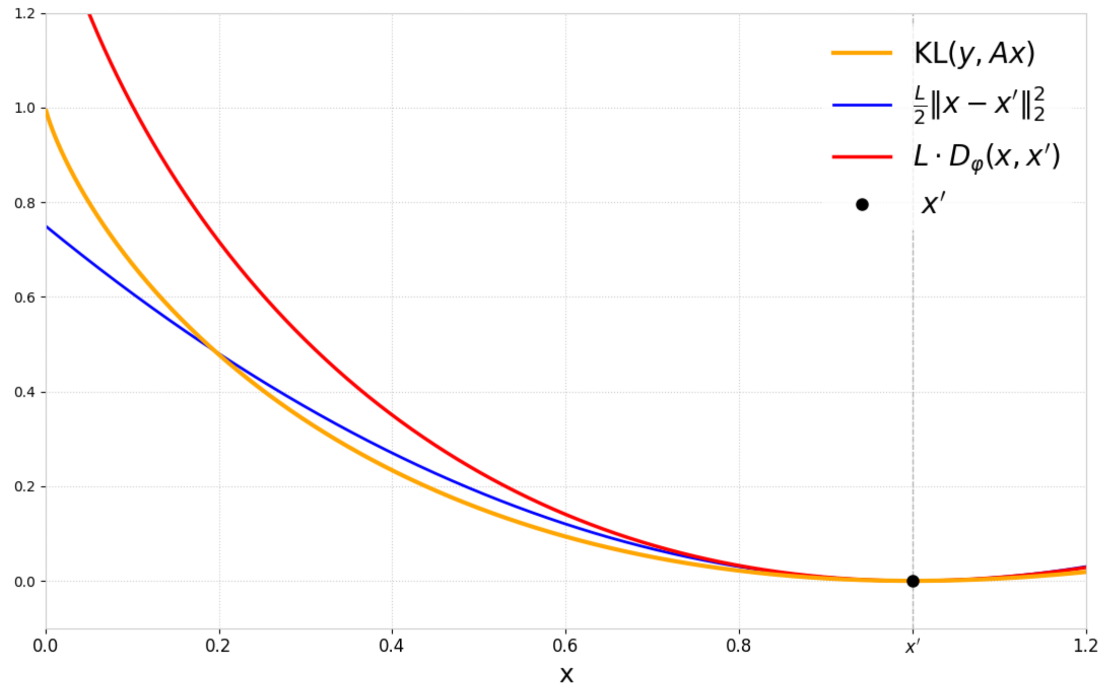
Convergence under **relative smoothness**

$$F(x) \leq F(x') + \langle \nabla F(x'), x - x' \rangle + L D_\varphi(x, x')$$

$$\varphi = -\log(\,\cdot\,) \implies L \geq \|y\|_1$$

[Bauschke, Bolte, Teboulle '17]

$$\boxed{\tau \in (0, 1/L) \implies J(x^k) - \bar{J} = O(1/k)}$$



UniGe | MaLGa

KL minimisation: Bregman Proximal Gradient Descent

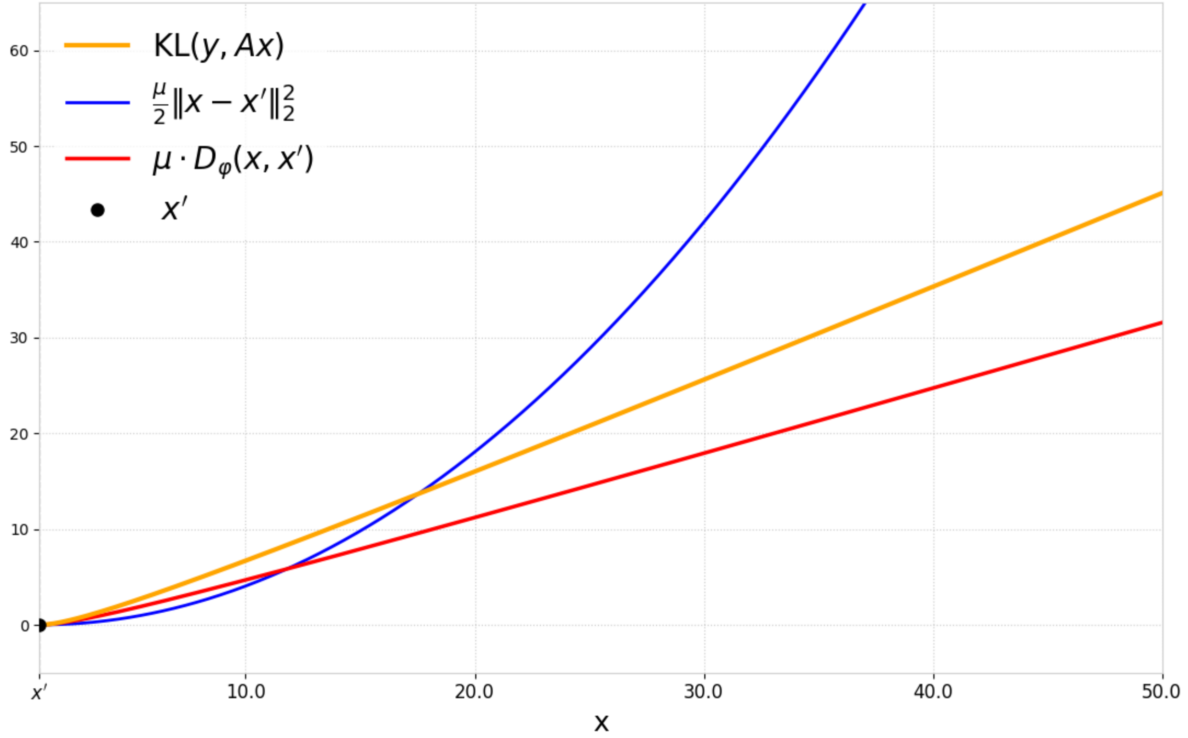
$$x^{k+1} = \underset{x \in \mathbb{R}^n_{\geq 0}}{\operatorname{argmin}} \langle x - x^k, \nabla \underbrace{(\mathrm{KL}(y, A \cdot + b))}_{F(x)} \rangle + R(x) + \frac{1}{\tau} D_\varphi(x, x^k)$$

Linear rate under **relative strong convexity (RSC)**

$$F(x) \geq F(x') + \langle \nabla F(x'), x - x' \rangle + \mu D_\varphi(x, x')$$

$$\boxed{\text{RSC} \implies J(x^k) - \bar{J} = O(\rho^k),\ \rho = 1 - \mu/L}$$

Lu, Freund, Nesteorov '18, Teboulle '18]



RSC does not hold if $n \geq 2$ for $\varphi = -\log(\cdot)$

It is possible to relax RSC and get linear rate?

Is (relative) global strong convexity necessary?

Restricted strong convexity (ReSC) [Zhang, Cheng '15]

$$\boxed{\langle \nabla F(x) - \nabla F(\bar{x}_{\ell_2}), x - \bar{x}_{\ell_2} \rangle \geq \mu \|x - \bar{x}_{\ell_2}\|_2^2}$$

$\bar{x}_{\ell_2} = \operatorname*{argmin}_{z \in S} \|x - z\|_2^2$, with $S = \operatorname{argmin}(J)$, with $J = F + R$

Upon **ReSC**, it is possible to retrieve linear convergence rate [Necoara, Nesterov, Glineur '19]

Restricted relative strong convexity (RRSC)

$$\boxed{\langle \nabla F(x) - \nabla F(\bar{x}_{\ell_2}), x - \bar{x}_{\ell_2} \rangle \geq \mu \langle \nabla \varphi(x) - \nabla \varphi(\bar{x}_{\ell_2}), x - \bar{x}_{\ell_2} \rangle}$$

If $\varphi = 1/2\| \cdot \|_2^2$, then **RRSC = ReSC**

Linear rate for BPGM?

Linear convergence upon RRSC

$$x^{k+1} = \underset{x \in \mathbb{R}^n_{\geq 0}}{\operatorname{argmin}} \langle x - x^k, \nabla F(x^k) \rangle + R(x) + \frac{1}{\tau} D_\varphi(x, x^k)$$

Proposition (Informal) [Chirinos-Rodriguez, D., Févotte, Soubies '26]

If the following holds:

1) F is $\mu-$**restricted strongly convex** relative to φ

2) F is L-**smooth** relative to φ

Then, given $J = F + R$, if $\tau \in (0,1/L)$ holds:

$$J(x^{k+1}) - \bar{J} \leq \frac{1}{\tau} \left(\frac{1}{1 + c(\varphi)\tau\mu} \right)^k D_\varphi(\bar{x}^0_\varphi, x^0),$$

where $\bar{x}^0_\varphi = \underset{z \in S}{\operatorname{argmin}} D_\varphi(x^0, z)$ and $c(\varphi) \in (0,1]$, **measures the symmetry** of D_φ

$$\tau = 1/2L, \rho = \frac{1}{1 + c(\varphi)\frac{\mu}{2L}}$$

What is a good choice of φ?

The smoothed-Burg's entropy: relative smoothness

$$\varphi_\xi(x) = \sum_{i=1}^{n} -\log(x_n + \xi),\ \xi \geq 0$$

Lemma [Chirinos-Rodriguez, D., Févotte, Soubies '26]

KL$(y, A \cdot +b)$ is $L-$smooth relative to φ for any L such that:

$$L \geq \sum_{m=1}^{m} w_m^2 y_m$$

with $w_m^2 = \max\{1, [A\boldsymbol{\xi}]_m / b\}$

The classic Bolte's result $L \geq \|y\|_1$ holds if $\xi \leq \min\limits_{m, a_{mn} \neq 0} \dfrac{b}{N a_{mn}}$, where $a_{mn} = A_{m,n}$

The result holds for $b > 0$

UniGe | MaLGa

The smoothed-Burg's entropy: RRSC, when it may hold?

$$\varphi_\xi(x) = \sum_{i=1}^{n} -\log(x_n + \xi),\ \xi \geq 0$$

Lemma 1 (Necessary condition) [Chirinos-Rodriguez, D., Févotte, Soubies '26]

Suppose that KL$(y, A \cdot + b)$ satisfies $\mu(\xi)-$restricted strong convexity relative to φ_ξ. Then $\xi > 0$

Remark:

If $\xi = 0$, then φ_0, i.e the **classic Burg's entropy, does not satisfy RRSC**

In general, if $\xi > 0$ the property may hold depending on the property of Ker(A) and $S = \text{argmin}(J)$

The smoothed-Burg's entropy: RRSC, where it holds?

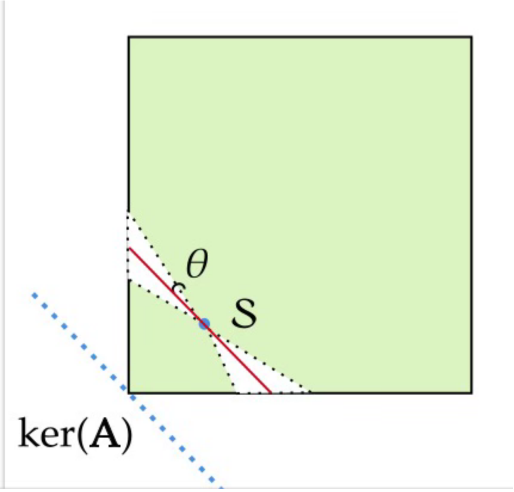
$$\varphi_\xi(x) = \sum_{i=1}^{n} -\log(x_n + \xi),\ \xi \geq 0$$

Lemma 1 (Sufficient condition) [Chirinos-Rodriguez, D., Févotte, Soubies '26]

If $R = 0$ or $\mathrm{Ker}(A) = \{0\}$, then $\xi > 0 \implies F = \mathrm{KL}(y, A\cdot + b)$ is $\mu(\xi)-$restricted strongly convex relative to φ_ξ

Counterexample:

$\mathrm{Ker}(A) \neq \{0\}$ and $R \neq 0$ such that $S = \mathrm{argmin}(\mathrm{KL}(y, A\cdot + b) + R) = \{x^*\}$



$$\langle \nabla F(x) - \nabla F(\bar{x}_{\ell_2}), x - \bar{x}_{\ell_2} \rangle \geq \mu \langle \nabla \varphi(x) - \nabla \varphi(\bar{x}_{\ell_2}), x - \bar{x}_{\ell_2} \rangle$$

=

$$\underbrace{\langle \nabla \mathrm{KL}(y, Ax + b) - \nabla \mathrm{KL}(y, A\bar{x}_{\ell_2} + b), A(x - \bar{x}_{\ell_2}) \rangle}_{=0 \text{ on the red line}}$$

Numerical experiments

$$y = \text{Poisson}(y_0)/Q, \quad y_0 = QAx^* + b$$

$A \in \mathbb{R}_{\geq 0}^{M \times N}$ random matrix, x^* ground truth

$$x^{k+1} = \underset{x \in C}{\operatorname{argmin}} \langle x - x^k, \nabla(\text{KL}(y, A \cdot + b\rangle + R(x) + \frac{1}{\tau} D_\varphi(x, x^k) \qquad C = \mathbb{R}_{\geq 0}^n$$

Setup:

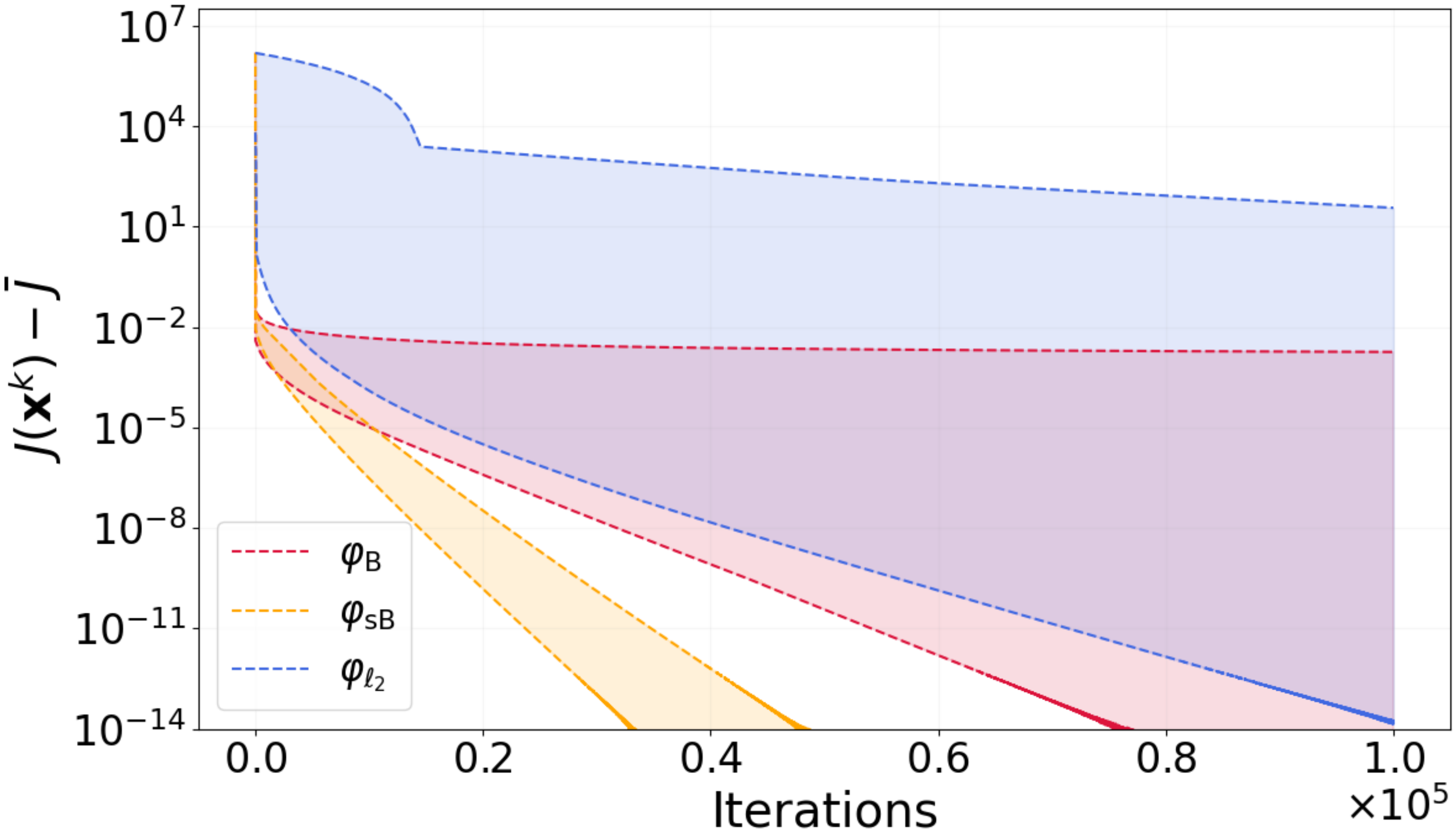
Three geometry: φ_B (Burg's entropy), φ_{sB} (Smoothed Burg's entropy), φ_{ℓ_2} (Euclidean Geometry)

Full rank case: $N = M = 500$, Not full rank case $M = 400, N = 500$

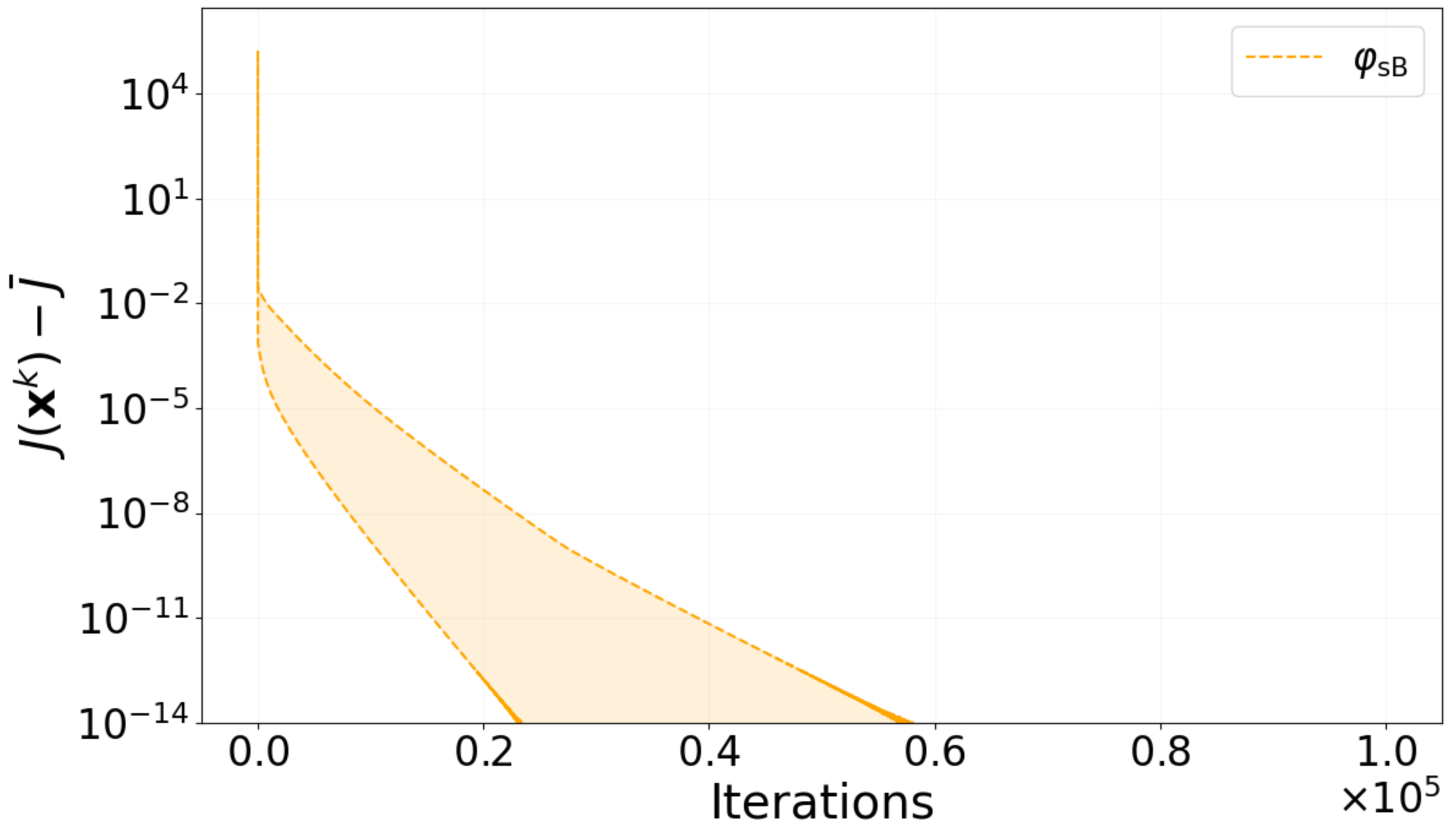
Unregularized setting: $R = 0$, Regularized setting: $R = \|\cdot\|_1$

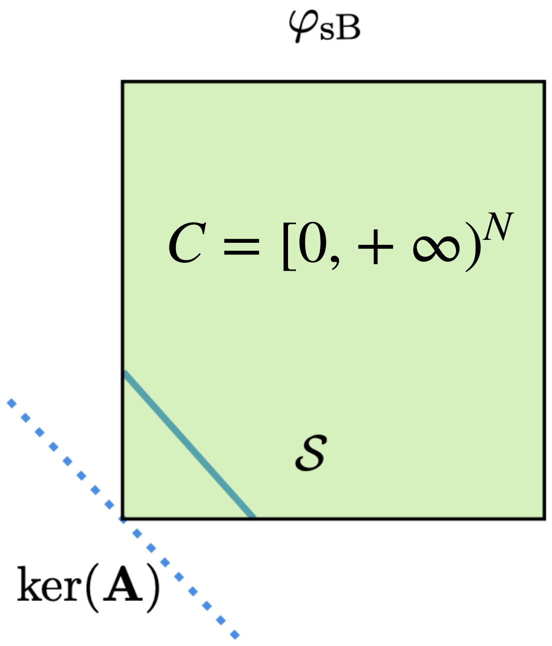
Not full rank Unregularized setting

30 different $x^0 \sim \mathcal{U}((0,K))$

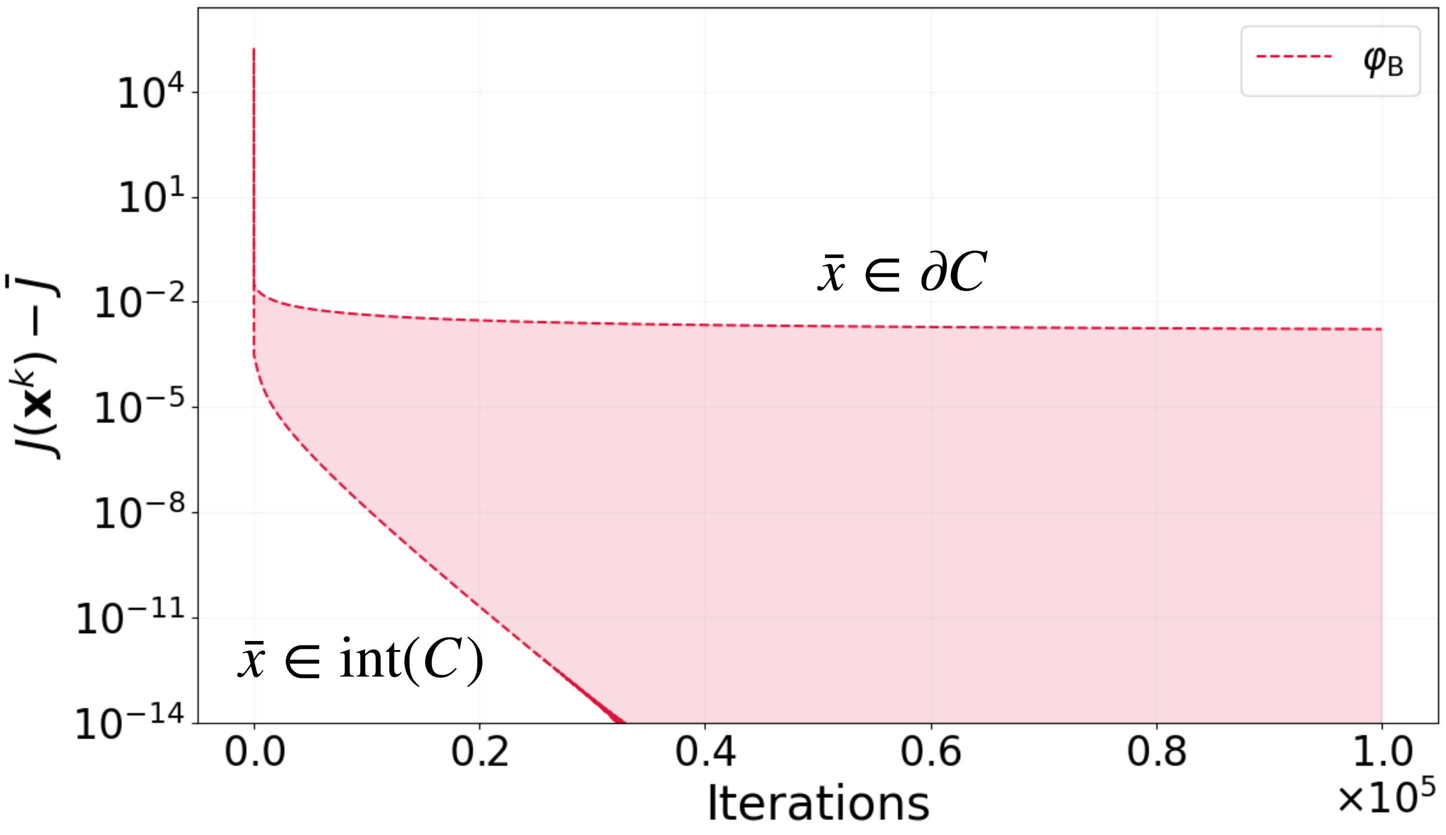


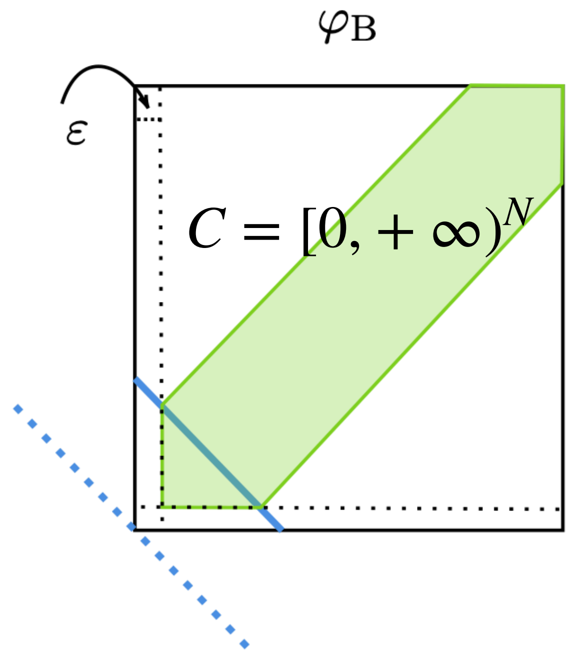
Smoothed Burg's entropy



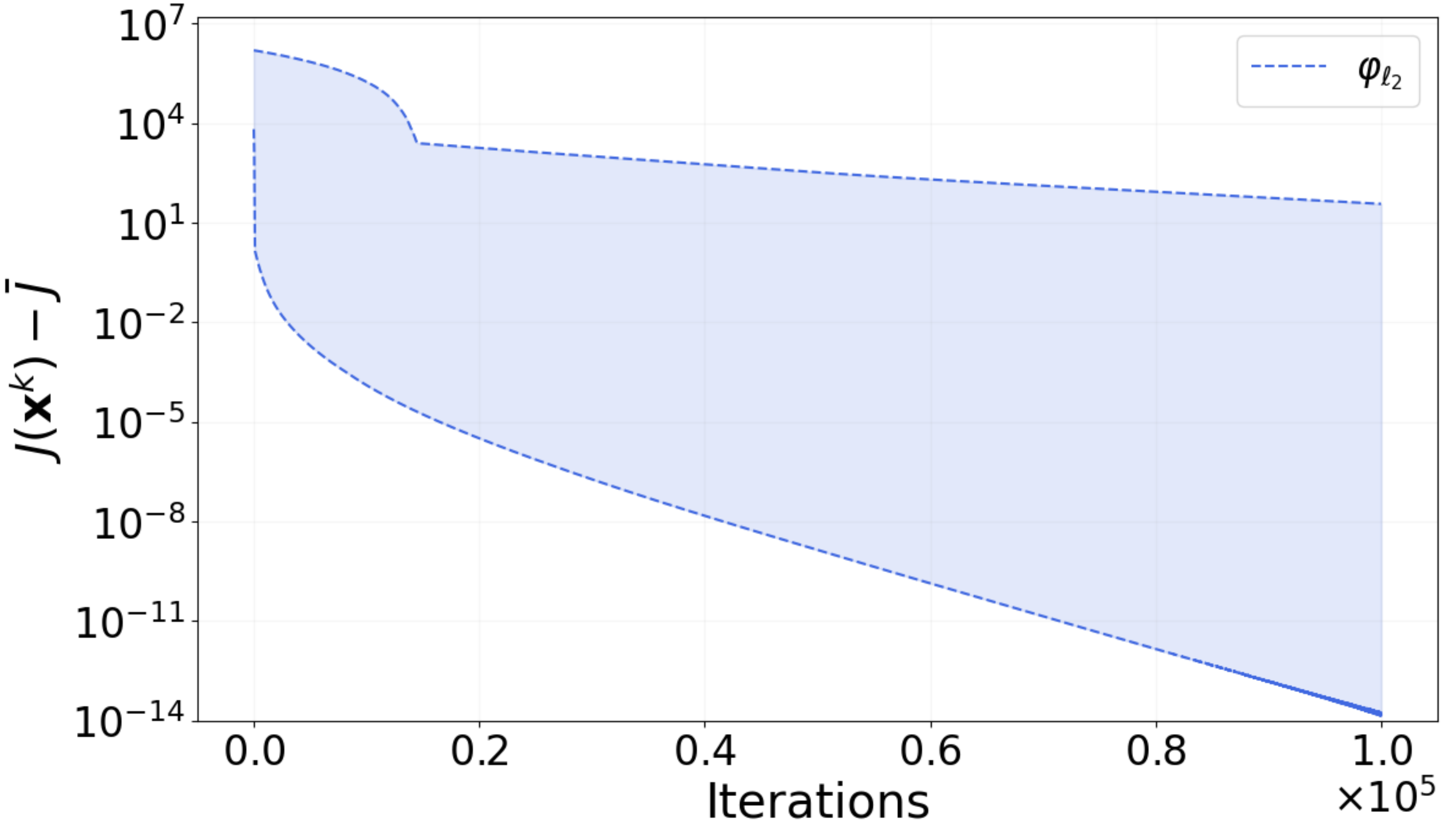


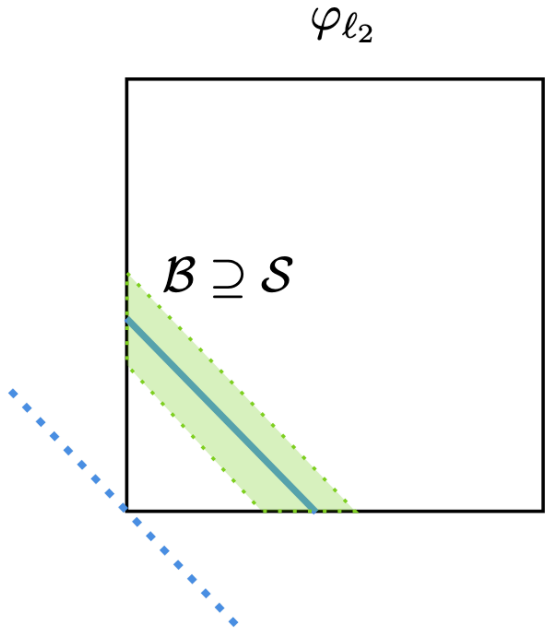
Burg's entropy



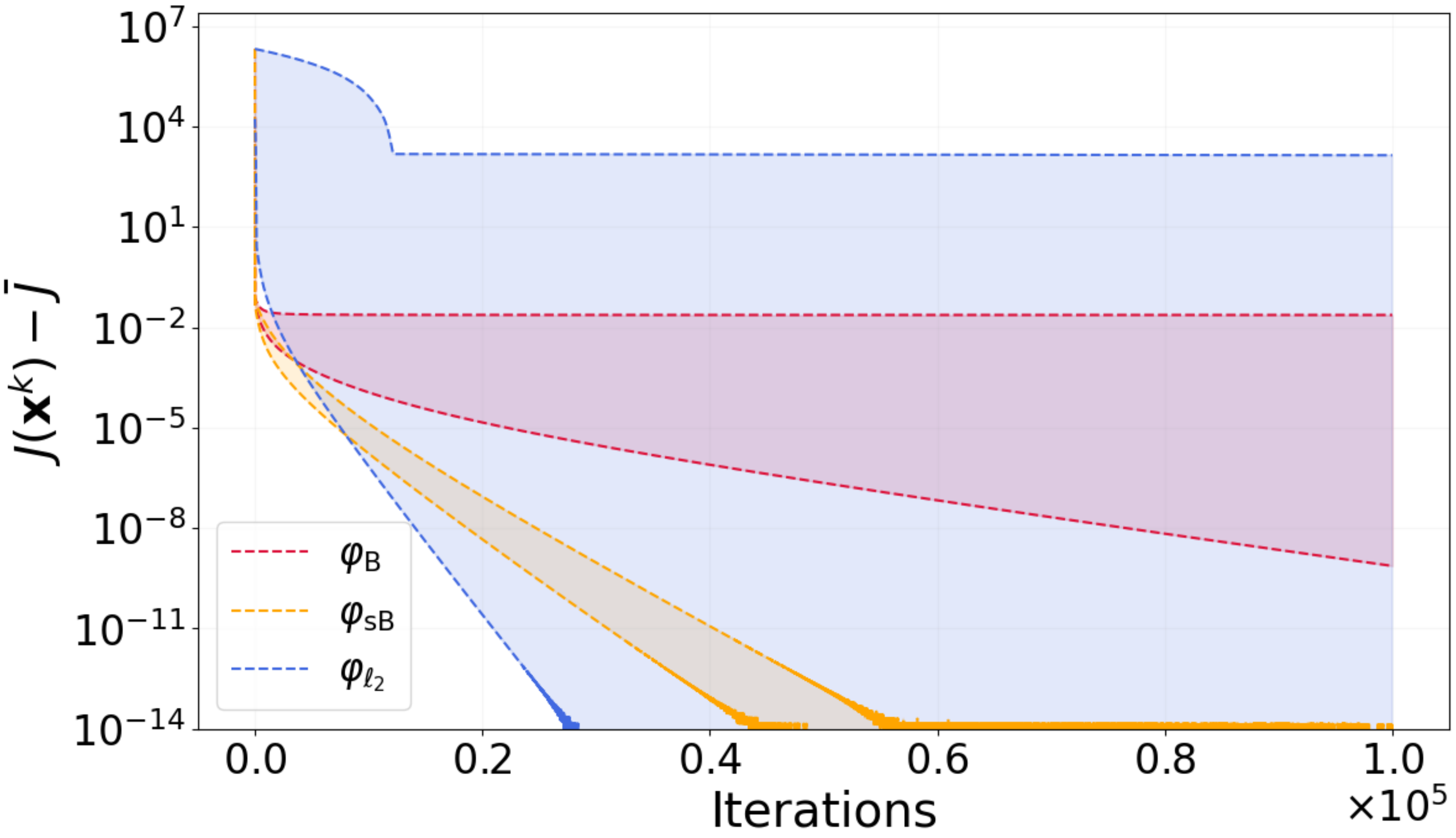


Euclidean Geometry

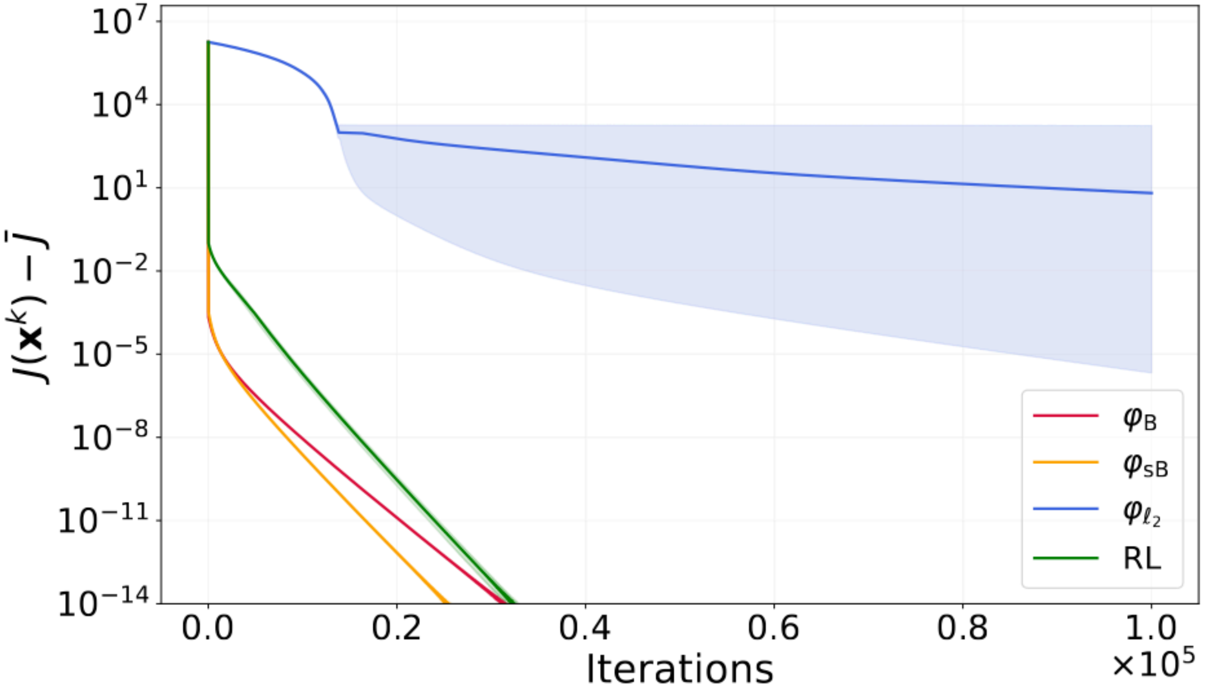




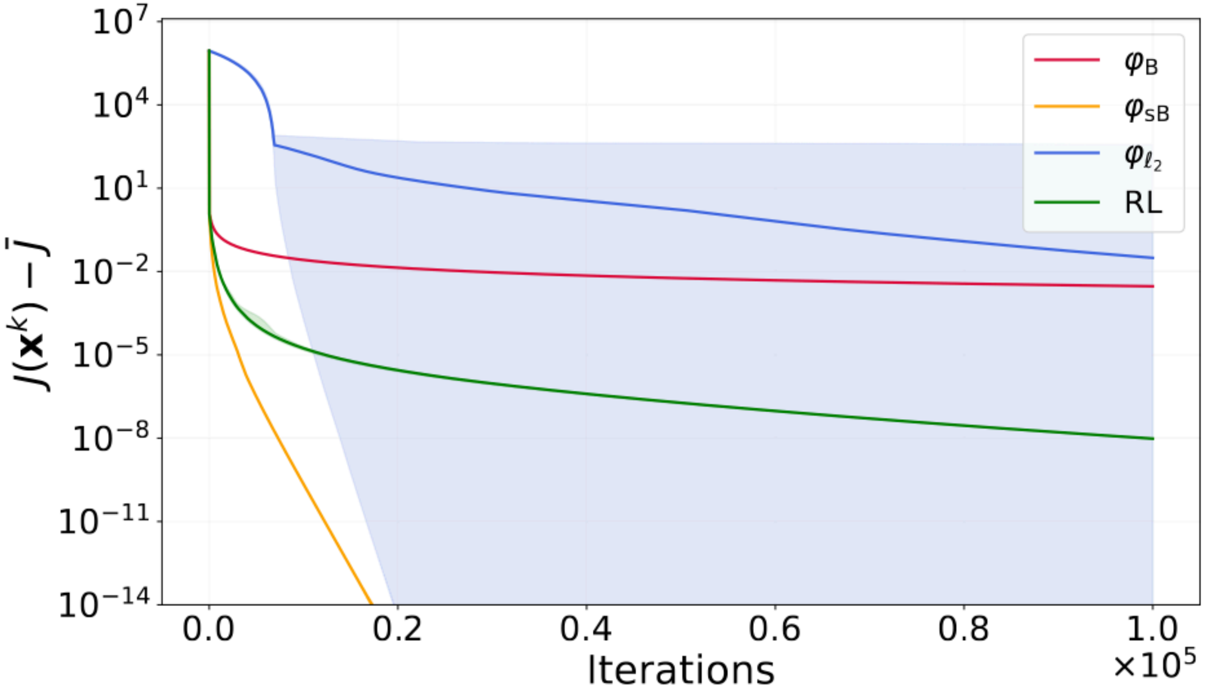
Full rank Regularized setting



Comparison with Richardson-Lucy (RL)



(A) $\bar{\mathbf{x}} \in \mathrm{int}(\mathcal{C})$



(B) $\bar{\mathbf{x}} \in \partial\mathcal{C}$

Conclusions & outlook

Linear convergence rate upon RRSC for Bregman Proximal Gradient Methods in convex setting

Tailored choice of φ for Kullback-Leibler functional

Experiments on simulated data showed the effectiveness of the proposed φ

$\longrightarrow$ What about convergence of iterates?

$\longrightarrow$ Test on Microscopy inverse problems

Thanks for your attention!

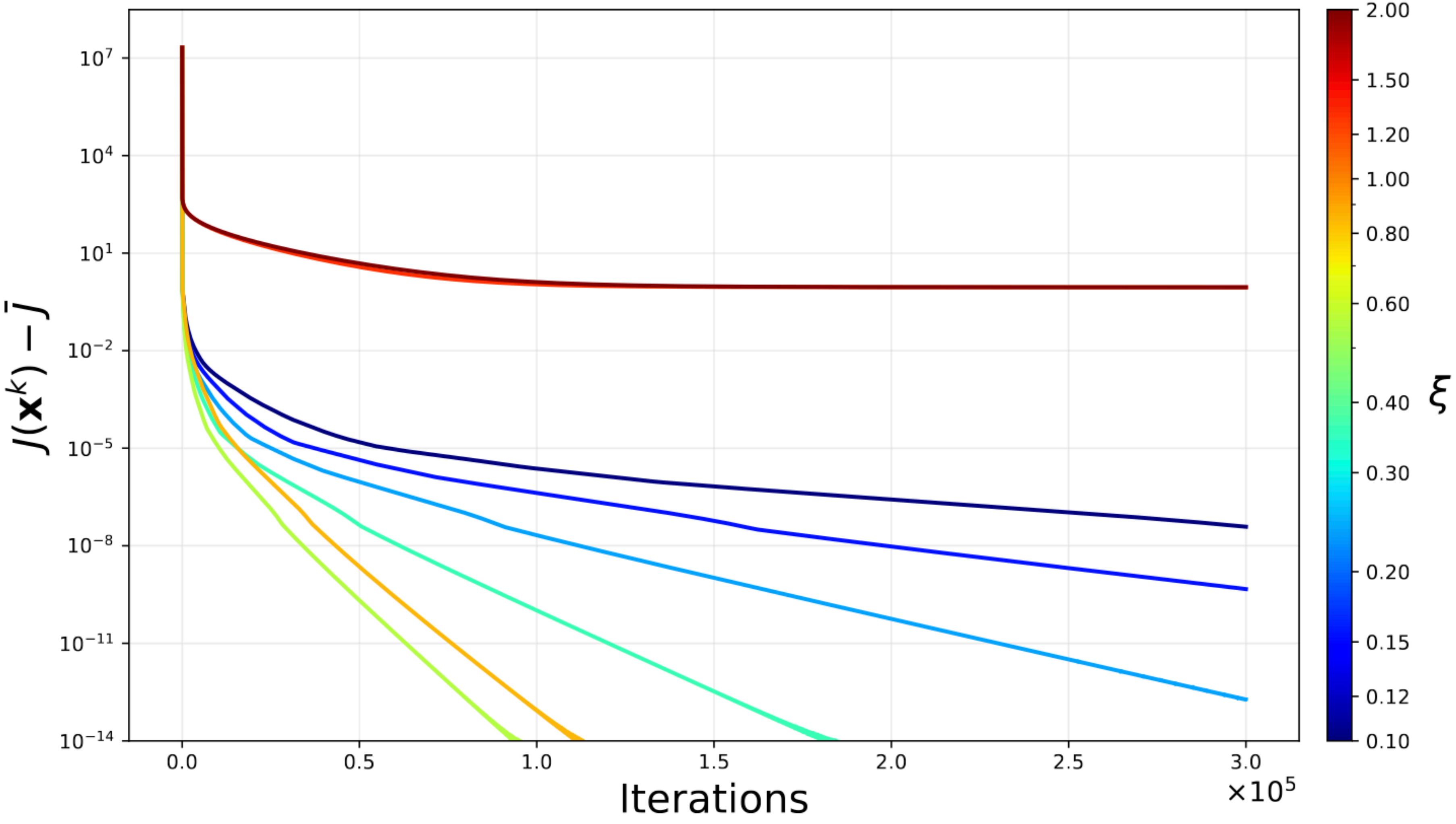
christian.daniele@edu.unige.it

UniGe

MaLGa

ARXIV PREPRINT!

Tuning of ξ



10^{7}
10^{4}
10^{1}
10^{-2}
10^{-5}
10^{-8}
10^{-11}
10^{-14}
$J(\mathbf{x}^k) - \bar{J}$
0.0
0.5
1.0
1.5
2.0
2.5
3.0
$\times 10^5$
Iterations
2.00
1.50
1.20
1.00
0.80
0.60
0.40
0.30
0.20
0.15
0.12
0.10
ξ

RRSC

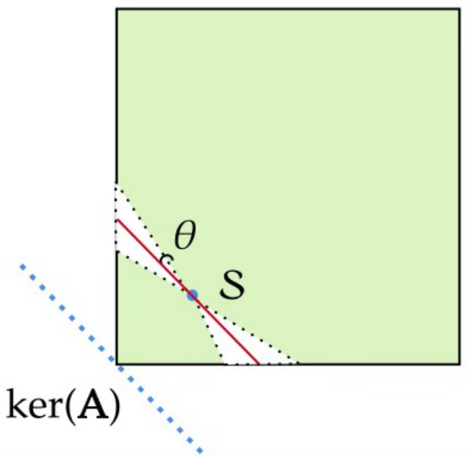
$$\mathcal{S}_{\ker}^{\theta} := \{\mathbf{x} \in \mathcal{C} \setminus \mathcal{S} \mid \angle(\mathbf{x} - \bar{\mathbf{x}}_{\ell_2}, \ker(\mathbf{A})) \leq \theta\}\,.$$

Proposition 3.4 (Restricted relative strong convexity). *Let F be the KL divergence defined in (21). Then, we have that*

- *For any $\theta \in (0, \pi/2)$, F is restricted strongly convex relative to φ_{sB} over $\mathcal{C} \setminus \mathcal{S}_{\ker}^{\theta}$.*
- *For any $\varepsilon > 0$ and any $\theta \in (0, \pi/2)$, F is restricted strongly convex relative to φ_{B} over*

(23) $$\mathcal{K}_{\varepsilon} := \left\{\mathbf{x} \in \mathbb{R}^N_{\geq \varepsilon} \setminus \mathcal{S}_{\ker}^{\theta} \mid \bar{\mathbf{x}}_{\ell_2} \in \mathbb{R}^N_{\geq \varepsilon}\right\}.$$

- *For any $\theta \in (0, \pi/2)$ and any compact set $\mathcal{B} \subseteq \mathcal{C}$ containing $\mathcal{S}$, F is restricted strongly convex (relative to φ_{ℓ_2}) over $\mathcal{B} \setminus \mathcal{S}_{\ker}^{\theta}$.*



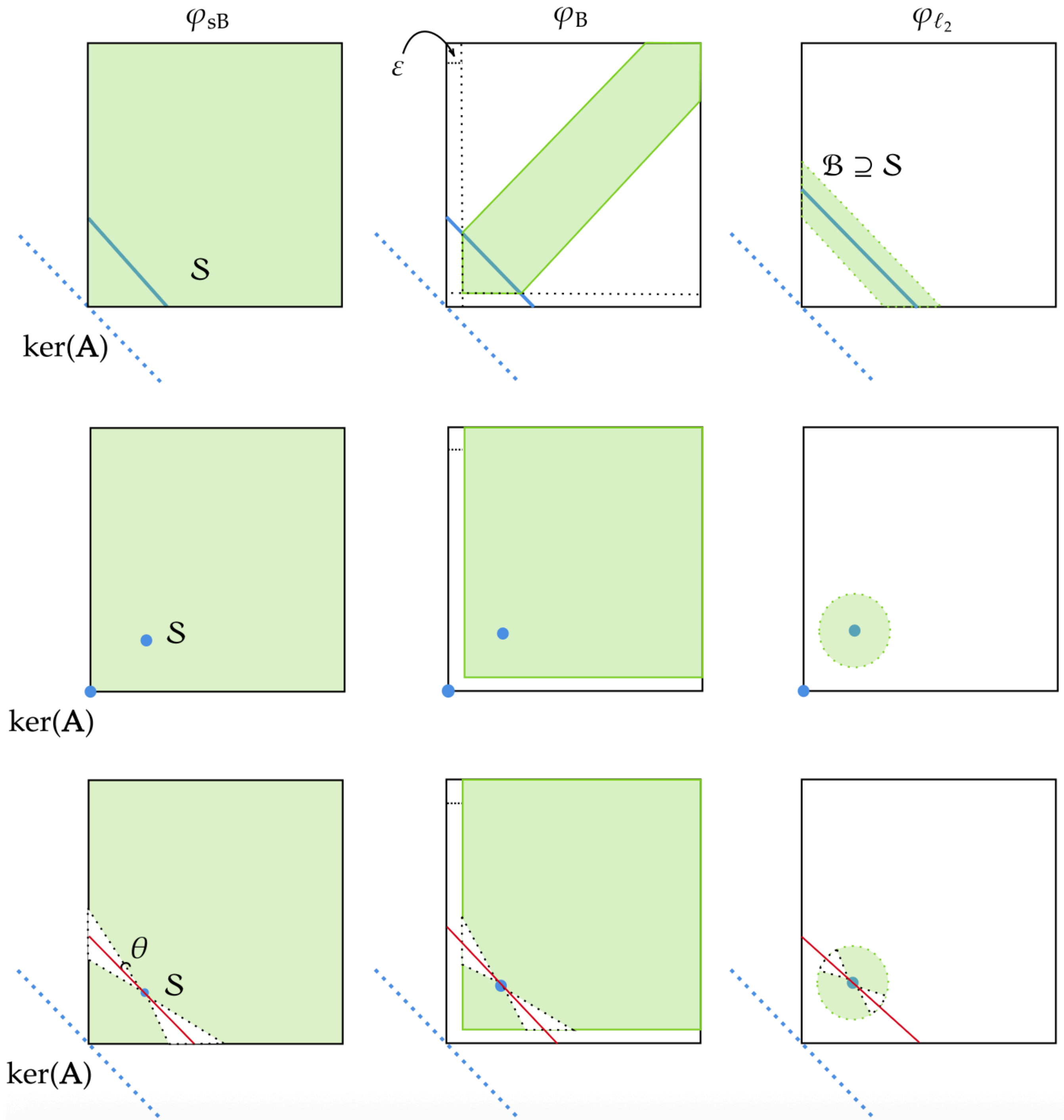
Restricted symmetry coefficient

Definition 2.2 (Restricted symmetry coefficient)**.** Let $\mathcal{D} \subseteq \operatorname{int}(\operatorname{dom}\varphi)$. Then, the (restricted) symmetry coefficient $\alpha_{\mathcal{D}}(\varphi)$ is given by

$$\alpha_{\mathcal{D}}(\varphi) := \inf\left\{\frac{D_\varphi(\mathbf{x},\mathbf{y})}{D_\varphi(\mathbf{y},\mathbf{x})} \mid \mathbf{x} \in \operatorname{int}(\operatorname{dom}\varphi), \mathbf{y} \in \mathcal{D}, \mathbf{x} \neq \mathbf{y}\right\} \in [0,1]. \tag{6}$$

Proposition 3.5 (Restricted symmetry coefficient)**.** *Let $\mathcal{S}$ be the solution set of (1). We have that*

- $\alpha_{\mathcal{S}}(\varphi_{\mathrm{sB}}) > 0$ *over* $\mathcal{C}$,
- $\alpha_{\mathcal{S}\cap\mathbb{R}^N_{\geq\varepsilon}}(\varphi_{\mathrm{B}}) > 0$ *over* $\mathbb{R}^N_{\geq\varepsilon}$,
- $\alpha(\varphi_{\ell_2}) = 1$ *over* $\mathbb{R}^N$.



φ_{sB}
φ_B
φ_{ℓ_2}
ε
$\mathcal{S}$
$\mathcal{B} \supseteq \mathcal{S}$
ker(**A**)
$\mathcal{S}$
ker(**A**)
θ
$\mathcal{S}$
ker(**A**)

Backtracking

Corollary 2.6 (Linear convergence of BPGM with backtracking). *Let Assumption 2.1 be satisfied, and consider the sequence* $(\mathbf{x}^k)_{k\in\mathbb{N}}$ *generated by* (BPGM) *with step-sizes* τ^k *obtained via the following backtracking strategy: for* $\beta \in (0,1)$, *and* $\tau^0 > 0$, *we set, for all* $k \in \mathbb{N}$, $\tau^k := \tau^{k-1} \cdot \beta^j$ *for the smallest* $j = 0, 1, \ldots$ *such that* $\mathbf{x}^{k+1} \in \mathcal{C} \cap \operatorname{int}(\operatorname{dom}\varphi)$ *and*

$$F(\mathbf{x}^{k+1}) \leq F(\mathbf{x}^k) + \langle \nabla F(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k \rangle + \frac{1}{\tau^k} D_\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k) \tag{19}$$

holds true.

Then, the sequence $(\mathbf{x}^k)_{k\in\mathbb{N}}$ *obeys the two statements of Theorem 2.4 by replacing* τ *with* $\bar{\tau} = \beta/L$.

Coercivity of KL

Proof. To prove the result we only need to show that the composite function J is coercive on $\mathbb{R}^N_{\geq 0}$. Indeed, given that J is lower semi-continuous over the closed convex set $\mathbb{R}^N_{\geq 0}$, we get by [8, Proposition 11.15] that coercivity ensures that $\mathcal{S}$ is nonempty and compact. To that end, we will show that, if $\|\mathbf{x}\|_1 \to +\infty$, then $J(\mathbf{x}) \to +\infty$. Recall that, by assumption, the matrix $\mathbf{A} \in \mathbb{R}^{M\times N}_{\geq 0}$, with entries $a_{m,n}$, $m = 1, \dots, M$, $n = 1, \dots, N$, has no column which is entirely zero. This means that, for every $n \in \{1, \dots, N\}$, there exists $m \in \{1, \dots, M\}$ such that $a_{m,n} > 0$. Then, defining the constant $c := \min_n \|\mathbf{a}_n\|_1$, we have that

$$\|\mathbf{A}\mathbf{x}\|_1 = \sum_{m=1}^{M} \left| \sum_{n=1}^{N} a_{m,n} x_n \right| = \sum_{n=1}^{N} x_n \sum_{m=1}^{M} a_{m,n} \geq c\|\mathbf{x}\|_1,$$

for all $\mathbf{x} \in \mathcal{C}$. Consequently, if $\|\mathbf{x}\|_1 \to +\infty$, then indeed $\|\mathbf{A}\mathbf{x}\|_1 \to +\infty$. Now, since $\mathrm{KL}(\mathbf{y}, \cdot + \mathbf{b})$ is coercive, we derive that

$$\lim_{\|\mathbf{x}\|_1 \to +\infty} \mathrm{KL}(\mathbf{y}, \mathbf{A}\mathbf{x} + \mathbf{b}) = \lim_{\mathbf{z} \to +\infty} \mathrm{KL}(\mathbf{y}, \mathbf{z} + \mathbf{b}) = +\infty,$$

showing that $\mathrm{KL}(\mathbf{y}, \mathbf{A} \cdot + \mathbf{b})$ is coercive. Finally, combining this with the fact that R is bounded from below proves that J is coercive on $\mathbb{R}^N_{\geq 0}$ and completes the proof. $\square$