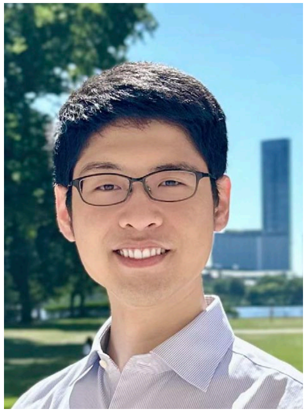
Level-Set Geometry and the Performance of Restarted-PDHG for Conic LP

Zikai Xiong (with Robert Freund)

Workshop celebrating of Don Goldfarb



Zikai Xiong
(MIT OR Center)

Robert Freund
(MIT Sloan)

Don and Eleuthera 1989

Huge-scale optimization is everywhere

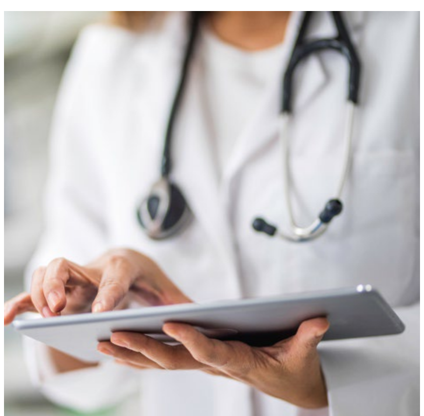
Manufacturing

Machine Learning

Energy

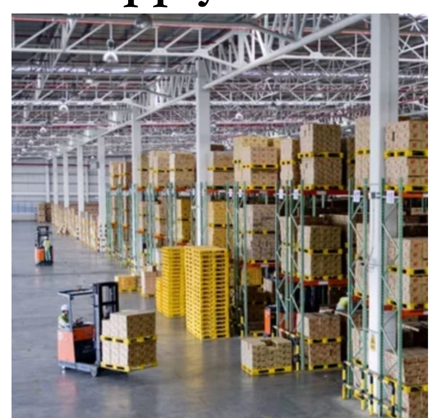
Transportation

Healthcare



Markets and Auctions

Supply Chains



Agriculture

History of Linear Optimization (“LO” or “LP”)

1947

Simplex Method

[George Dantzig, 1947]
75+ years ago

1984

Interior Point Method

[Narendra Karmarkar, 1984]
40 years ago

Simplex and IPMs require expensive matrix factorizations

Consider an LP instance with n decision variables and $\frac{n}{2}$ linear constraints, whose constraint matrix has sparsity = 0.05

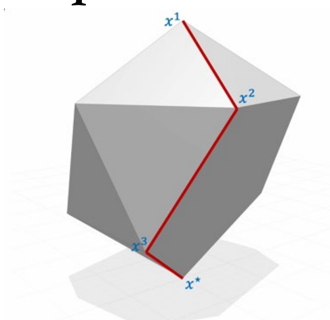
Number of variables (n)	Cost of one IPM iteration

Hence the emergence of FOMs for solving huge (and also not-so-huge) LP instances

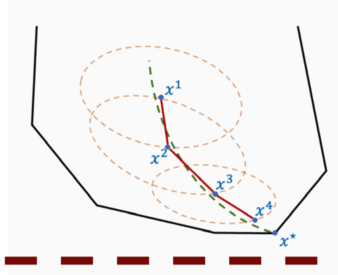
Recent Advances on Huge-Scale LP Solvers in the Industry

Classic methods

Simplex method



Interior-point method



First-order methods

- Primal-Dual Hybrid Gradient ("PDHG", "Chambolle-Pock method")
- Tackles huge-scale problems
- Benefits from modern computational architecture (such as GPU)

Google — 2021

COPT Cardinal Optimizer — State-of-art commercial optimization solver — February, 2024

NVIDIA — March, 2024

HiGHS — An open-source optimization software — March, 2024

FICO — State-of-art commercial optimization solver — April, 2024

GUROBI OPTIMIZATION — October, 2024

We are witnessing a dramatic shift from classic methods to first-order methods

Huge-Scale LP Research

- SCS: Operator splitting/ADMM [O'Donoghue, Chu, Parikh, Boyd, 2016]
- ABIP+: ADMM-based interior-point method
  [Lin, Ma, Ye, Zhang, 2021] & [Deng, et al., 2022]
- Semi-smooth Newton augmented Lagrangian [Li, Sun, Toh, 2020]
- **Primal-Dual Hybrid Gradient (PDHG)** with restarts, applied directly to the primal-dual saddle point problem [Applegate, Hinder, Lu, Lubin, 2023] & [Applegate, et al., 2021] **(2024 Beale-Orchard-Hays Prize)**
- **GPU implementations** of PDHG in Julia and C
  [Lu and Yang, 2023] & [Lu et al., 2023]
- **Guarantees for PDHG for LP** using "Limiting Error Ratios" and LP Sharpness [Xiong and F 2023]
- **Guarantees for PDHG for CLP** – using level-set geometry
  [Xiong and F 2024]

This talk is based on material from three papers

For LP and conic optimization:

- New computational guarantees based on problem (sub)level-set geometry

Xiong, Z., and Freund, R. M. (2024). The Role of Level-Set Geometry on the Performance of PDHG for Conic Linear Optimization.

For LP with unique optima:

- Closed-form iteration bound
- Two-stage performance of PDHG
- "Average-case" polynomial-time complexity guarantee

Xiong, Z. (2024). Accessible Theoretical Complexity of the Restarted Primal-Dual Hybrid Gradient Method for Linear Programs with Unique Optima.

Xiong, Z. (2024). Probabilistic Analysis of Restarted PDHG for Linear Programming Problems (working paper).

Conic Linear Optimization ("CLO" or "CLP")

CLP in standard form

(primal)

$$\boxed{\begin{aligned} \min\ & c^\top x \\ \text{s.t. } & Ax = b \\ & x \in \mathcal{K} \end{aligned}}$$

(dual)

$$\begin{aligned} \max\ & b^\top y \\ \text{s.t. } & c - A^\top y \in \mathcal{K}^* \end{aligned}$$

Decision variables

- $x \in R^n$ (for primal problem)
- $y \in R^m$ (for dual problem)

CLP saddlepoint formulation

$$\boxed{\min_{x \in \mathcal{K}} \max_{y} \ c^\top x - y^\top A x + b^\top y}$$

Primal-Dual Hybrid Gradient Method (PDHG)

Conic Optimiztion in Saddlepoint Form

$$\min_{x \in \mathcal{K}} \max_{y} \; c^\top x + b^\top y - y^\top A$$

PDHG

$$x^{k+1} \leftarrow \mathrm{Proj}_{\mathcal{K}}\left(x^k - \tau\underbrace{\left(c - A^\top y^k\right)}_{\text{Gradient w.r.t. } x^k}\right)$$

$$y^{k+1} \leftarrow y^k + \sigma\underbrace{\left(b - Ax^{k+1}\right)}_{\text{Gradient w.r.t. } y^k} - \sigma\underbrace{A\left(x^{k+1} - x^k\right)}_{\text{Momentum Term}}$$

Primal-Dual Hybrid Gradient for Conic Optimization

PDHG

$$x^{k+1} \leftarrow \mathrm{Proj}_{\mathcal{K}}\left(x^k - \tau\left(c - A^\top y^k\right)\right)$$
$$y^{k+1} \leftarrow y^k + \sigma\left(b - Ax^{k+1}\right) - \sigma A\left(x^{k+1} - x^k\right)$$

- **Inexpensive iterations**:
  Only requires matrix-vector multiplications
- **"Fast" convergence rates**:
  Adaptive restarts based on average iterates yield global linear convergence on LP [Applegate, Hinder, Lu, Lubin, 2023]

We use "PDHG" to denote "PDHG with adaptive restarts"

Motivation for Restarts for PDHG: "Visualization"

$$\min_x \max_y x \cdot y$$

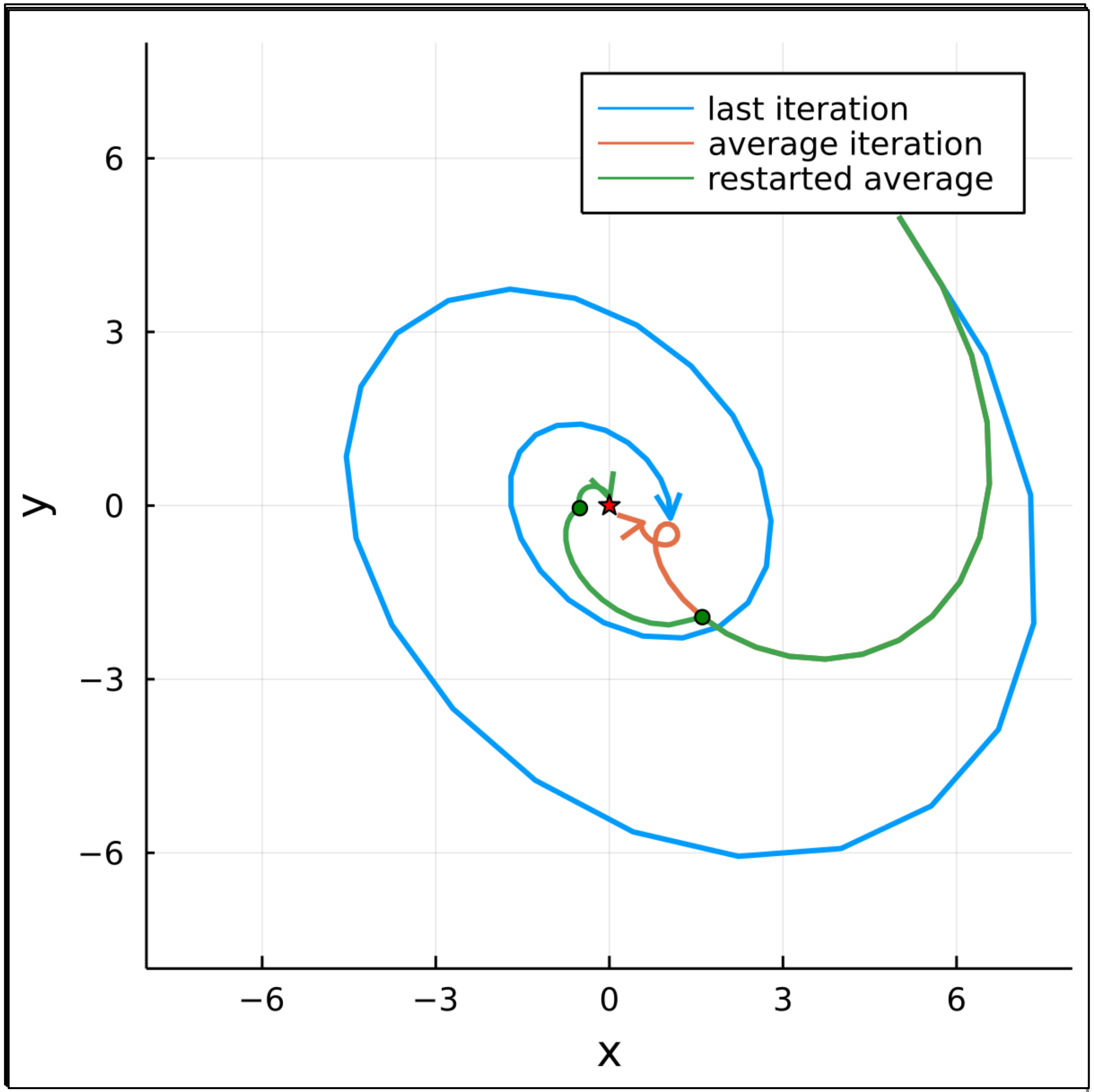
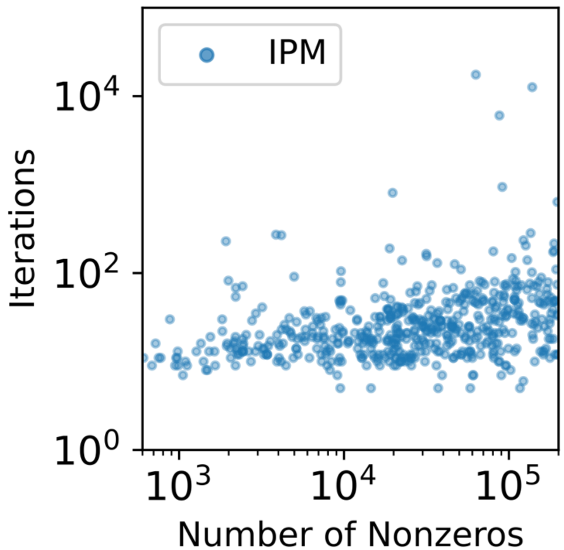


*figure courtesy Haihao Lu

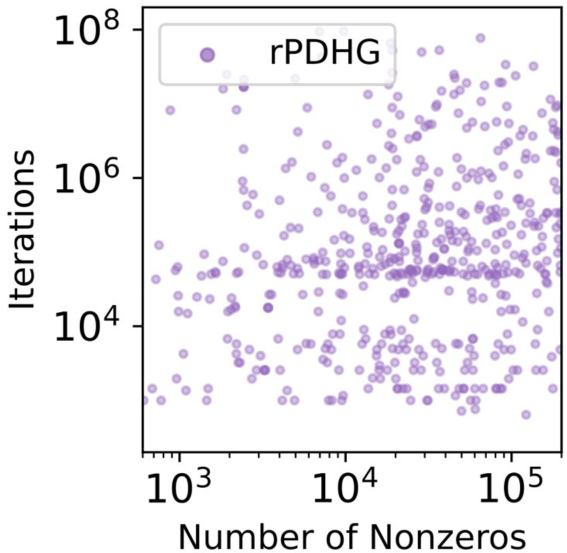
Challenge I: Variability in the Performance of PDHG

- PDHG uses many more iterations than an IPM
  makes sense, it is a first-order method … IPM iterations are hugely expensive while PDHG iterations are very cheap
- Some small problem instances require a very large number of PDHG iterations
  a real challenge for PDHG

IPM Iterations needed for
LP relaxations from MIPLIB 2017



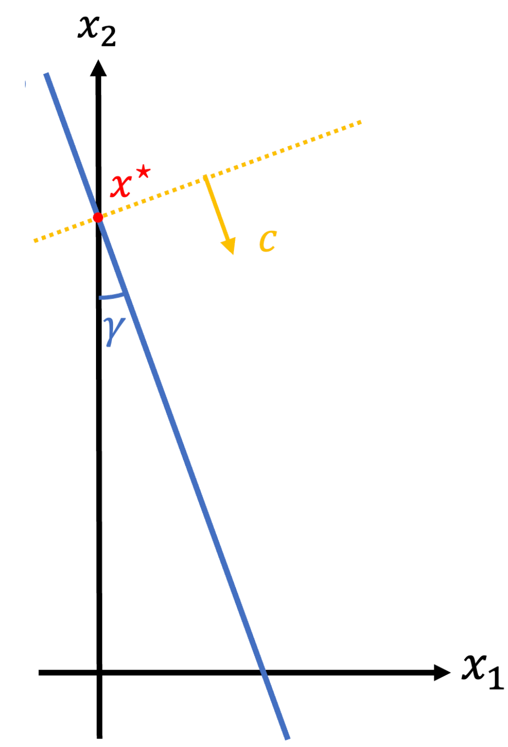
PDHG iterations
LP relaxations from MIPLIB 2017



A seemingly easy LP instance

For $\gamma \in \left(0, \frac{\pi}{2}\right)$ define:

$$P(\gamma): \quad \begin{aligned} \min_{x_1, x_2} \quad & \sin(\gamma)\, x_1 - \cos(\gamma)\, x_2 \\ \text{s.t.} \quad & \sin(\gamma)\, x_1 + \cos(\gamma)\, x_2 = 1 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$



$P(\gamma)$ is always easy for the simplex method and interior-point methods

However, when γ is very small, PDHG requires at least 1,000,000 iterations. What **conditions** of $P(\gamma)$ make it so hard for PDHG?

Challenge II: Loose/unworkable computational guarantees

Existing computational guarantees:

<u>**Theorem**</u> [Applegate, Hinder, Lu, Lubin, 2023] PDHG computes an ε-optimal solution within

$$O\left((\|x^\star\| + \|y^\star\|) \cdot ||A|| \cdot \boldsymbol{H(K)} \cdot \log\left(\frac{\|x^\star\| + \|y^\star\|}{\varepsilon} \right) \right)$$

iterations.

$\boldsymbol{H(K)}$ is the global Hoffman constant of the matrix $\boldsymbol{K}$ of the KKT system

$$\boldsymbol{K} = \begin{pmatrix} A & \\ -A & \\ & A^\top \\ c^\top & -b^\top \end{pmatrix} \qquad \boldsymbol{H(K)} \approx \frac{1}{\min_{J \subseteq \{1,\ldots,2m+n+1\}} \sigma^+_{\min}(K_J)}$$

Likely way too loose, and very hard to compute/validate/analyze.

Challenge II: Loose/unworkable computational guarantees

Existing computational guarantees:

Theorem [Applegate, Hinder, Lu, Lubin, 2023] PDHG computes an ε-optimal solution within

$$O\left((\|x^\star\| + \|y^\star\|) \cdot ||A|| \cdot \boldsymbol{H(K)} \cdot \log\left(\frac{\|x^\star\| + \|y^\star\|}{\varepsilon} \right) \right)$$

iterations.

Key question:

- What conditions of the problem actually drive the performance of PDHG?

Sublevel-set Geometry

Sublevel-set geometry and new performance guarantees for PDHG

Primal-Dual Slack Space

Primal	Dual
$\min_{x} c^\top x$ $\text{s.t. } Ax = b$ $x \in \mathcal{K}$	$\max_{y,s} b^\top y$ $\text{s.t. } A^\top y + s = c$ $s \in \mathcal{K}^*$

The "primal-dual slack-space variable" is w :

$$w \coloneqq (x, s) \text{ are primal/dual feasible slacks}$$

Duality gap: $\mathrm{Gap}(x, s) = c^\top x - b^\top y$

(which is a linear function of x and s)

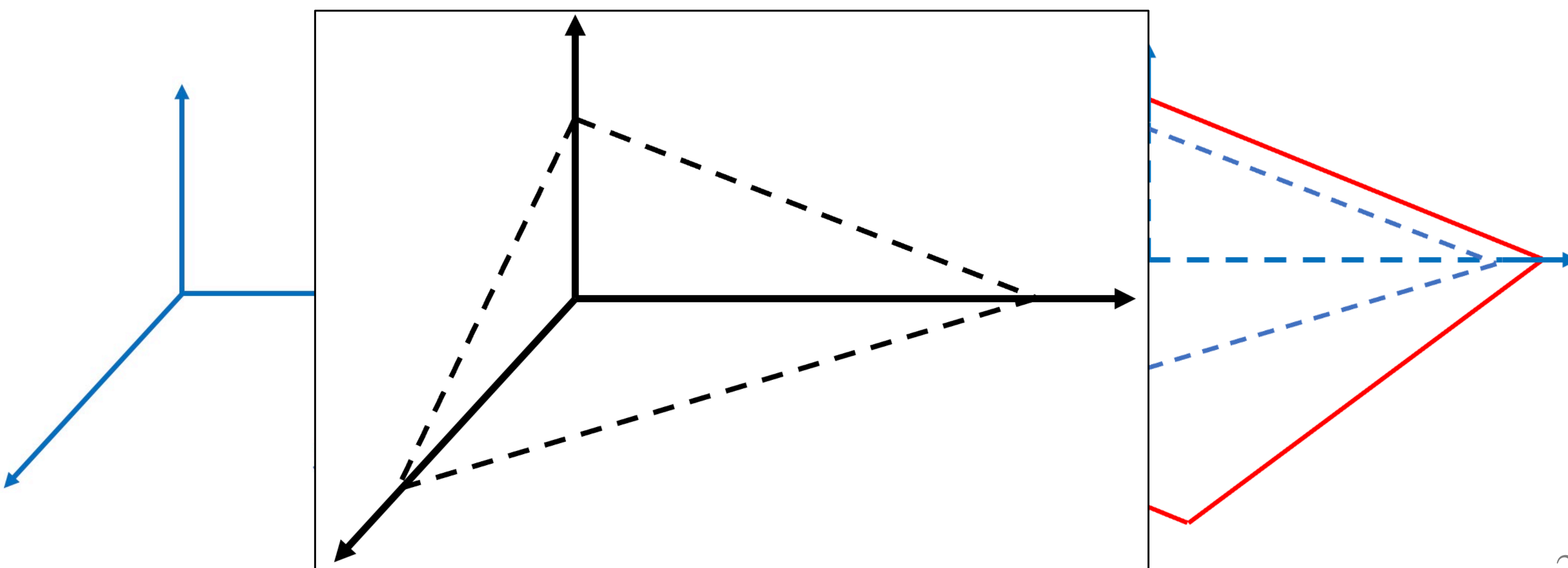
The feasible primal-dual slack-space variables

$$\min_{x} c^\top x \quad \text{s.t.} \quad \boxed{Ax = b}, \quad \boxed{x \in \mathcal{K}}$$

$$\max_{y,s} b^\top y \quad \text{s.t.} \quad \boxed{A^\top y + s = c}, \quad \boxed{s \in \mathcal{K}^*}$$

(x, s) in the primal and dual cone for CLP

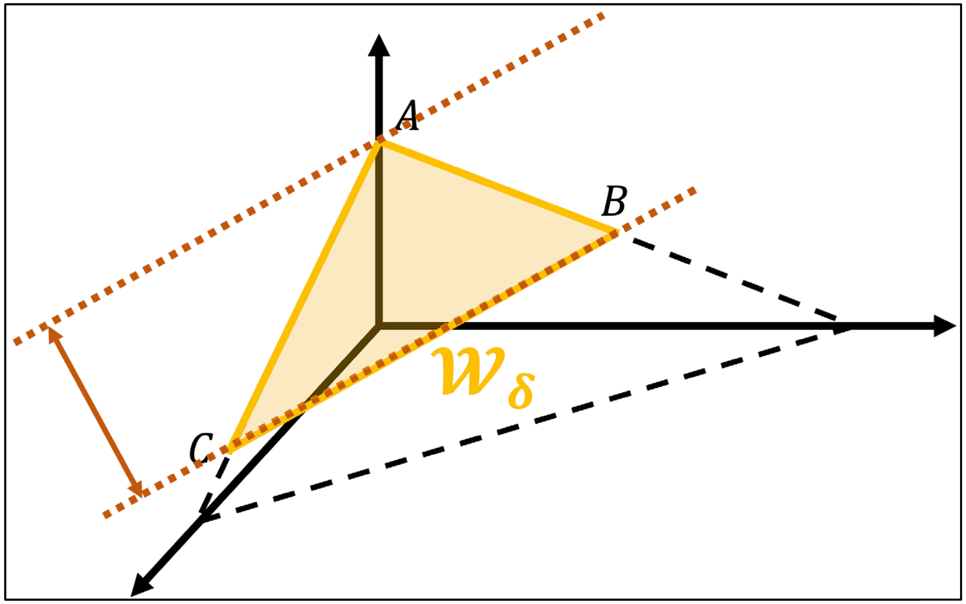
(x, s) lies in an affine subspace



Primal-Dual Slack Sublevel Set

$$\mathcal{W}_\delta := \left\{ w := (x, s) \;\middle|\; \begin{array}{c} w \text{ is primal/dual feasible} \\ \mathbf{Gap}(w) \leq \delta \end{array} \right\}$$

Note: $\mathcal{W}_0 = \mathcal{W}^\star$



Worst-case complexity of PDHG (under unique optima)

Theorem [Xiong and F 2024]: Suppose $w^\star$ is unique. PDHG computes an $\boldsymbol{\varepsilon}$-optimal solution within

$$\tilde{O}\left(\kappa \cdot \lim_{\delta\to 0} \frac{\boldsymbol{D_\delta}}{\boldsymbol{r_\delta}} \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)$$

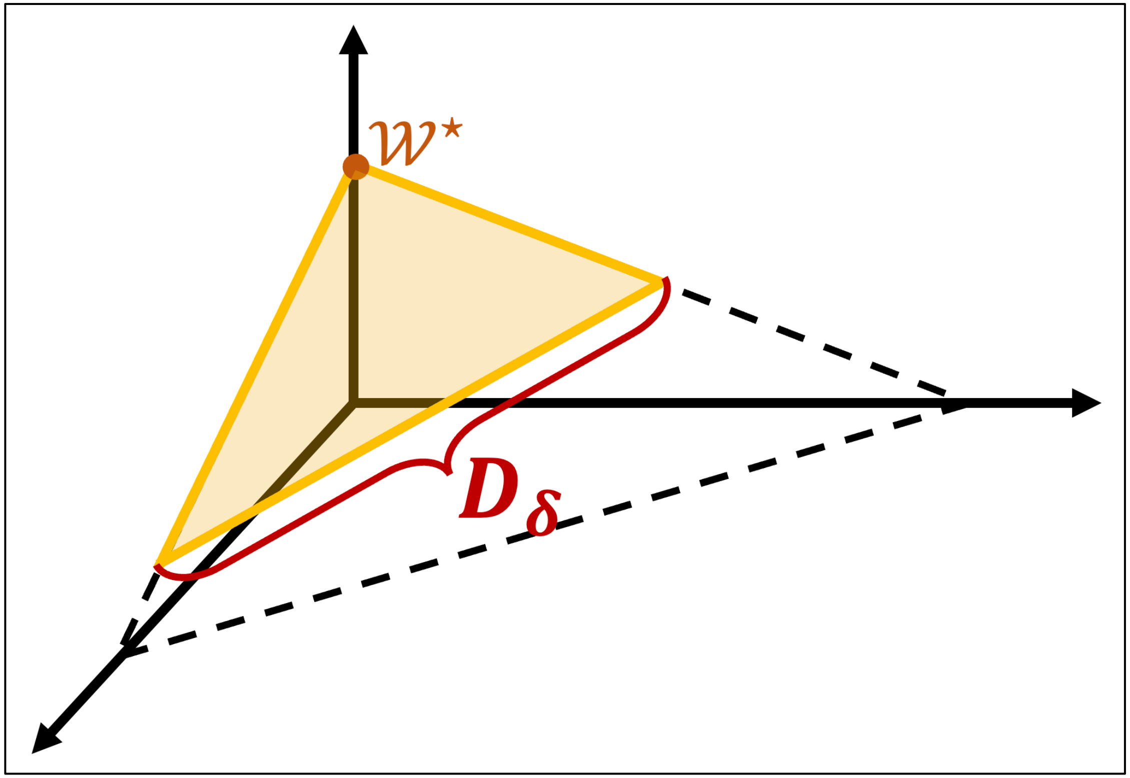
iterations.

Matrix condition number of A:
$\kappa := \sigma_{\max}^{+}(A)/\sigma_{\min}^{+}(A)$

"Sublevel-set geometry"

D_δ: Diameter of δ-sublevel set $\mathcal{W}_\delta$

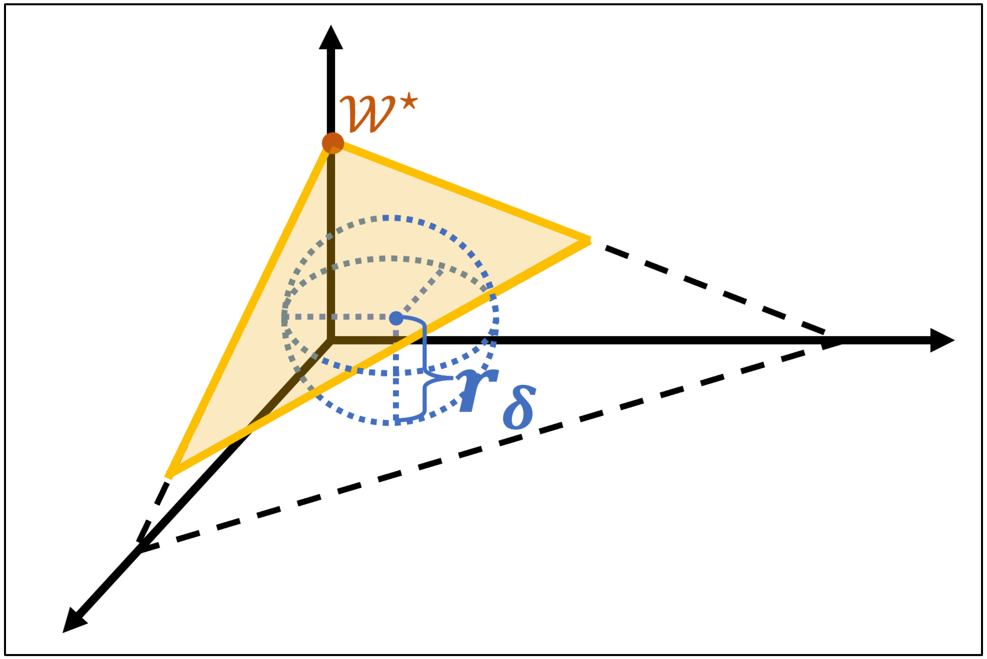
$$\boxed{D_\delta := \max_{\bar{w},\, \widehat{w} \in \mathcal{W}_\delta} \|\bar{w} - \widehat{w}\|}$$



r_δ: "Conic Radius" of $\mathcal{W}_\delta$

$$r_\delta := \max_{r\geq 0,\, w\in\mathcal{W}_\delta} r$$
$$\text{s.t.} \quad B_w(r) \subset \mathcal{K} \times \mathcal{K}^*$$

r_δ is the radius of the maximum ball inscribed in $\mathcal{K} \times \mathcal{K}^*$ and centered at a point in $\mathcal{W}_\delta$



Target: ε-optimal solution

(x, s) is an $\boldsymbol{\varepsilon}$-optimal solution if:

- distance to each type of constraint is no larger than ε , and
- the duality gap is not larger than ε

(x, s) is an $\boldsymbol{\varepsilon}$-optimal solution if:

- $\mathrm{Dist}(x, \{x | Ax = b\}) \leq \varepsilon$
- $\mathrm{Dist}(x, \mathcal{K}) \leq \varepsilon$
- $\mathrm{Dist}(s, \{s | \exists\ y \ \text{ s. t. } A^\top y + s = c\}) \leq \varepsilon$
- $\mathrm{Dist}(s, \mathcal{K}^\star) \leq \varepsilon$
- $c^\top x - b^\top (AA^\top)^{-1} A(c - s) \leq \varepsilon$

Worst-case complexity of PDHG (under unique optima)

Theorem [Xiong and F 2024]: Suppose $w^\star$ is unique. PDHG computes an $\boldsymbol{\varepsilon}$-optimal solution within

$$\tilde{O}\left(\kappa \cdot \lim_{\delta\to 0} \frac{\boldsymbol{D_\delta}}{\boldsymbol{r_\delta}} \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)$$

iterations.

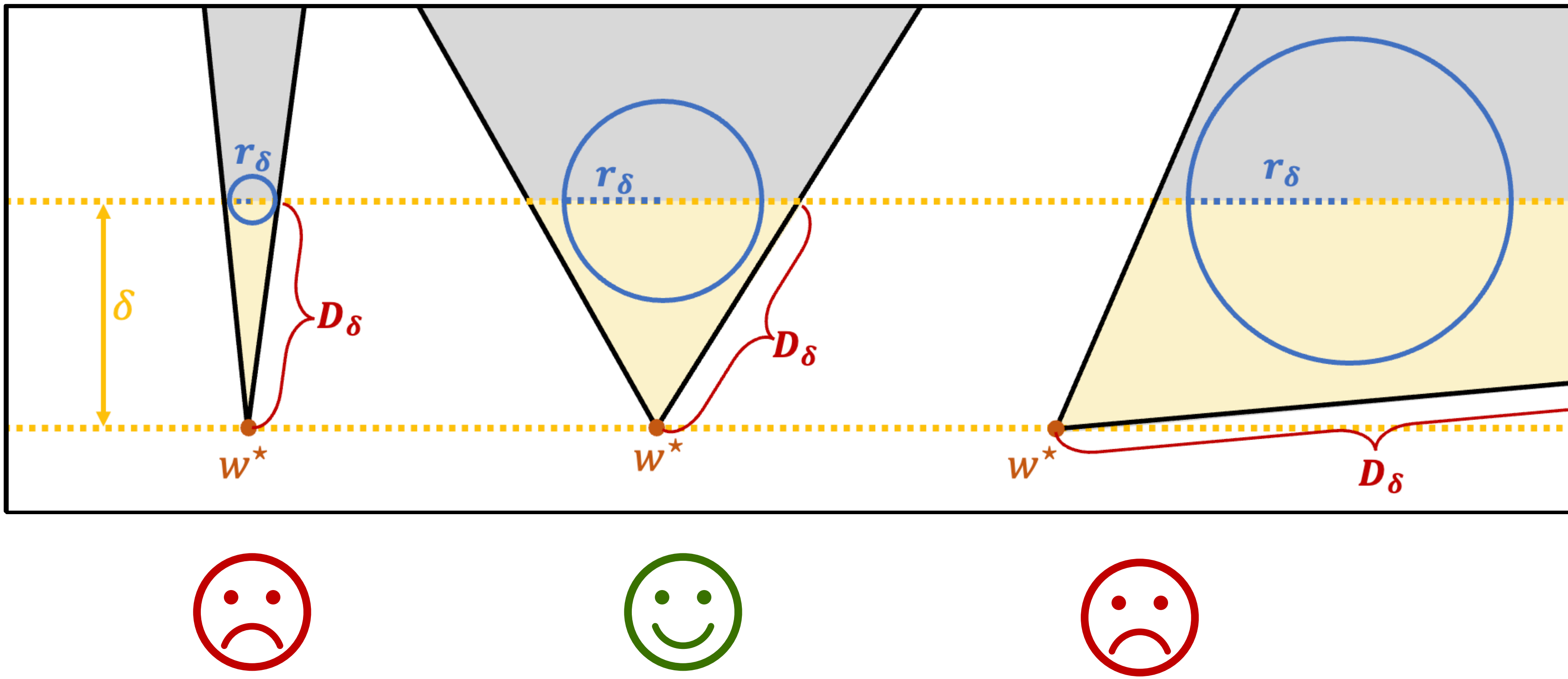
Matrix condition number of A:

$$\kappa := \sigma_{\max}^{+}(A)/\sigma_{\min}^{+}(A)$$

"Sublevel-set geometry"

Local Geometry of and $\lim_{\delta \to 0} \frac{D_\delta}{r_\delta}$ in the case of LP

When $w^\star$ is unique and δ is sufficiently small, $\mathcal{W}_\delta$ is a slice of a pointed cone at $w^\star$.



Very small r_δ
Intermediate D_δ

Intermediate r_δ
Intermediate D_δ

Intermediate r_δ
Very large D_δ

Worst-case complexity of PDHG (under unique optima)

Matrix condition number of A: $\kappa = \sigma_{\max}^{+}(A)/\sigma_{\min}^{+}(A)$

Theorem [Xiong and F 2024]: Suppose $w^\star$ is unique. PDHG computes an $\boldsymbol{\varepsilon}$-optimal solution within

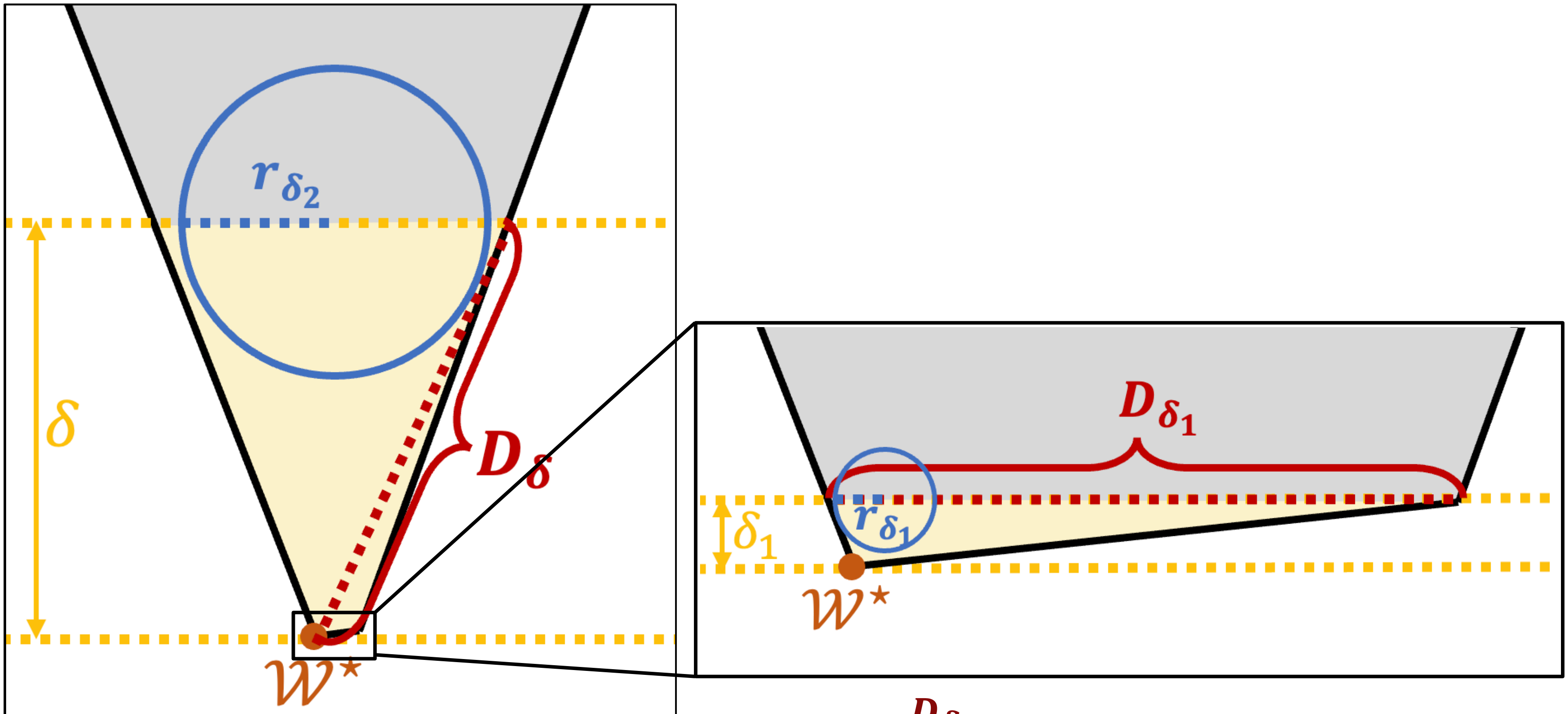
$$\tilde{O}\left(\kappa \cdot \lim_{\delta \to 0} \frac{\boldsymbol{D_\delta}}{\boldsymbol{r_\delta}} \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)$$

iterations.

Matrix condition number

Local geometric condition

Another example

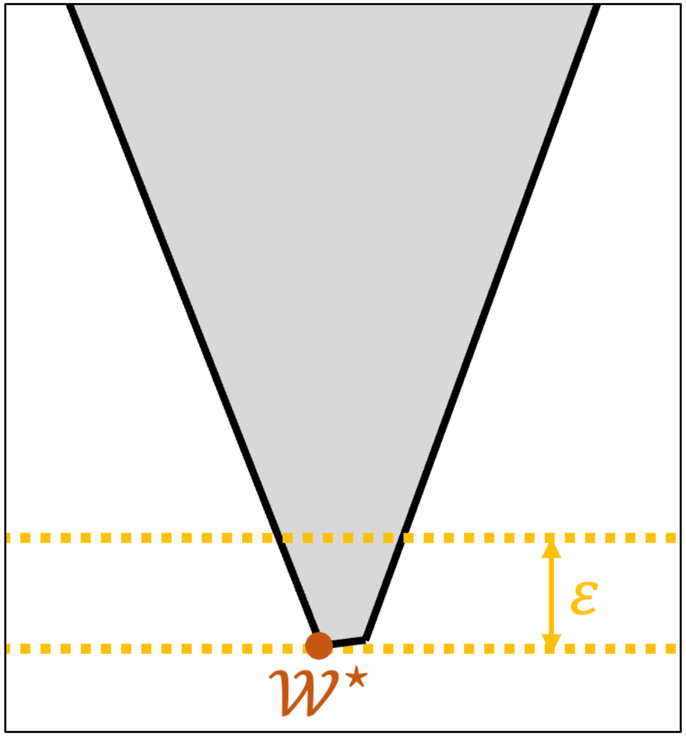


For $\delta_2 > \delta_1$, $\frac{D_{\delta_2}}{r_{\delta_2}}$ becomes smaller/better

$\frac{D_{\delta_1}}{r_{\delta_1}}$ is very large/bad (due to the small r_{δ_1})

Is $\lim_{\delta \to 0} \frac{D_\delta}{r_\delta}$ the only geometric condition?

Suppose we want an ε-optimal solution:



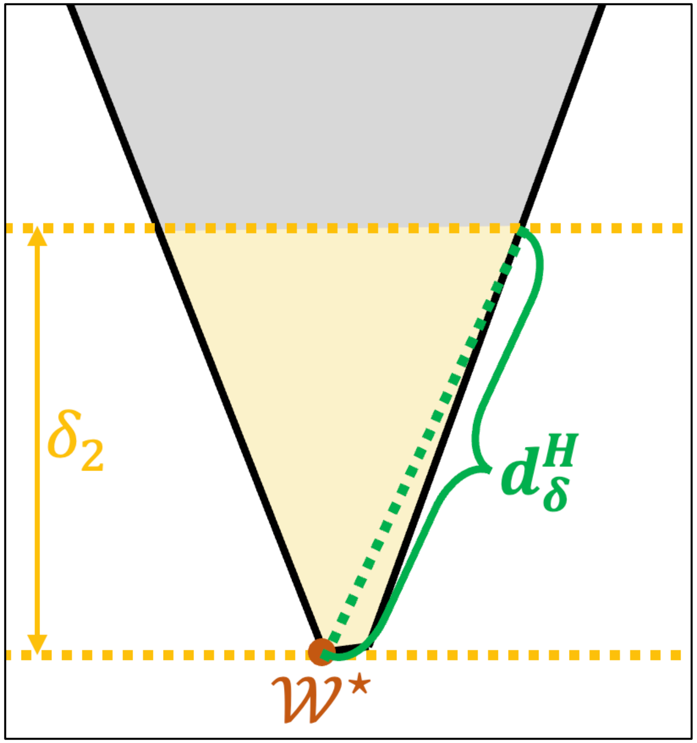
Intuition: The very-local bad geometry should not have a significant impact when the iterates of the algorithm have not yet reached the local neighborhood.

Is $\lim_{\delta \to 0} \frac{D_\delta}{r_\delta}$ the only geometric condition?

We will need a third geometric measure designed to capture "being close to $\mathcal{W}^\star$"

$$d_\delta^H := \max_{w \in \mathcal{W}_\delta} \mathrm{Dist}(w, \mathcal{W}^\star)$$

Hausdorff distance from $\mathcal{W}_\delta$ to $\mathcal{W}^\star$



Our General Conic Optimization Computational Guarantee

Theorem [Xiong and F 2024]: The number of PDHG iterations required to compute an $\boldsymbol{\varepsilon}$-optimal solution is upper bounded by:

$$\tilde{O}\left(\kappa \cdot \max\left\{\frac{\boldsymbol{D_\delta}}{\boldsymbol{r_\delta}} \cdot \ln\left(\frac{1}{\varepsilon}\right), \frac{\boldsymbol{d_\delta^H}}{\varepsilon}(1 + \mathrm{Dist}(0, \mathcal{W}^\star))\right\}\right)$$

for each $\delta > 0$.

How good the geometry of $\mathcal{W}_\delta$ is | How close $\mathcal{W}_\delta$ is to $\mathcal{W}^\star$

Implication: If there is a δ-sublevel set that

(i) has good geometry and (ii) is close to the optimal solution set,

then PDHG may converge faster.

Remark: This result holds for LP with multiple optima, and for general conic optimization too.

$\boldsymbol{D_\delta}$: Diameter of … $\boldsymbol{r_\delta}$: Conic radius of … $\boldsymbol{d_\delta^H}$: Hausdorff distance …

Our General Conic Optimization Computational Guarantee

Theorem [Xiong and F 2024]: The number of PDHG iterations required to compute an $\boldsymbol{\varepsilon}$-optimal solution is upper bounded by:

$$\tilde{O}\left(\kappa \cdot \max\left\{\frac{\boldsymbol{D_\delta}}{\boldsymbol{r_\delta}} \cdot \ln\left(\frac{1}{\varepsilon}\right), \quad \frac{\boldsymbol{d_\delta^H}}{\varepsilon}(1 + \mathrm{Dist}(0, \mathcal{W}^\star))\right\}\right)$$

for each $\delta > 0$.

Small when $\mathcal{W}_\delta$ has good geometry

Small when $\mathcal{W}_\delta$ is close to $\mathcal{W}^\star$

Q: Can we improve $\frac{\boldsymbol{D_\delta}}{\boldsymbol{r_\delta}}$ and $\boldsymbol{d_\delta^H}$?

A: Yes, by using Hessian Rescaling

$\boldsymbol{D_\delta}$: Diameter of … $\boldsymbol{r_\delta}$: Conic radius of … $\boldsymbol{d_\delta^H}$: Hausdorff distance …

Consider the generic case of LP: LPs with unique optima

Matrix condition number of A: $\kappa = \sigma_{\max}^{+}(A)/\sigma_{\min}^{+}(A)$

Theorem [Xiong and F 2024]: Suppose $w^\star$ is unique. PDHG computes an $\boldsymbol{\varepsilon}$-optimal solution within

$$\tilde{O}\left(\kappa \cdot \lim_{\delta \to 0} \frac{\boldsymbol{D_\delta}}{\boldsymbol{r_\delta}} \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)$$

iterations.

Matrix condition number	Local geometric condition

[Xiong, 2024]:

- This bound has a closed-form expression
- PDHG has local fast convergence in a small neighborhood of $w^\star$
- This bound is $\tilde{O}\left(n^{2.5} \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)$ with high probability

Iterations Bounds in Closed Form of the Optimal Basis/Solution

A formula for $\lim_{\delta\to 0} \frac{D_\delta}{r_\delta}$ using the optimal basis/solution

Strict Complementary Slackness (under unique optima)**:**

Let $\mathcal{B} := \mathrm{supp}(x^\star)$ and $\mathcal{N} := \mathrm{supp}(s^\star)$. Then $(\mathcal{B}, \mathcal{N})$ is a partition of $\{1,2,\ldots,n\}$.

We use "$\approx$" to denote being equivalent up to an absolute constant (2).

<u>Lemma</u> [Xiong 2024]

Under unique optima, let $B := A_{\mathcal{B}}$, $N := A_{\mathcal{N}}$. Then

$$\lim_{\delta\to 0}\frac{D_\delta}{r_\delta} \approx (\|x^\star\|_1 + \|s^\star\|_1)\cdot \max\left\{\max_{i\in[m]}\frac{\|(B^{-1}N)_{i,:}\|+1}{x^\star_{\mathcal{B}(i)}},\ \max_{j\in[n-m]}\frac{\|(B^{-1}N)_{:,j}\|+1}{s^\star_{\mathcal{N}(j)}}\right\}.$$

ℓ_1-norm of the optimal solution

$$= \max_i\left(\frac{1 + \text{norm of the } i\text{–th row of } (B^{-1}N)}{i\text{–th component of } x^\star_{\mathcal{B}}}\right)$$

$$= \max_j\left(\frac{1 + \text{norm of the j–th column of } (B^{-1}N)}{\text{j–th component of } s^\star_{\mathcal{N}}}\right)$$

$B^{-1}A = (I, B^{-1}N)$ is the simplex method tableau

A formula for $\lim_{\delta\to 0} \frac{D_\delta}{r_\delta}$ using the optimal basis/solution

Strict Complementary Slackness:

Let $\mathcal{B} \coloneqq \mathrm{supp}(x^\star)$ and $\mathcal{N} \coloneqq \mathrm{supp}(s^\star)$. Then $(\mathcal{B}, \mathcal{N})$ is a partition of $\{1,2,\dots,n\}$.

We use "≈" to denote being equivalent up to an absolute constant (2).

<u>Lemma</u> [Xiong 2024]

Under unique optima, let $B \coloneqq A_{\mathcal{B}}$, $N \coloneqq A_{\mathcal{N}}$. Then

$$\lim_{\delta\to 0} \frac{D_\delta}{r_\delta} \approx (\|x^\star\|_1 + \|s^\star\|_1) \cdot \max\left\{ \max_{i\in[m]} \frac{\|(B^{-1}N)_{i,:}\| + 1}{x^\star_{\mathcal{B}(i)}}, \max_{j\in[n-m]} \frac{\|(B^{-1}N)_{:,j}\| + 1}{s^\star_{\mathcal{N}(j)}} \right\}.$$

Furthermore, $\lim_{\delta\to 0} \frac{D_\delta}{r_\delta}$ has a simple upper bound:

$$\lim_{\delta\to 0} \frac{D_\delta}{r_\delta} \le 2 \cdot \frac{\|x^\star\|_1 + \|s^\star\|_1}{\min_{i\in[n]} x_i^\star + s_i^\star} \cdot \|B^{-1}A\|$$

Ratio of ℓ_1-norm to the smallest nonzero

Norm of the simplex tableau at the optimal solution

An iteration bound using the optimal basis/solution

Theorem [Xiong 2024]: PDHG computes an ε-optimal solution within:

$$\tilde{O}\left(\kappa \cdot \frac{\|x^\star\|_1 + \|s^\star\|_1}{\min_{i\in[n]} x_i^\star + s_i^\star} \cdot \|B^{-1}A\| \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)$$

iterations.

The smallest nonzero $\left(\min_{i\in[n]} x_i^\star + s_i^\star\right)$ plays a key role in other methods as well:

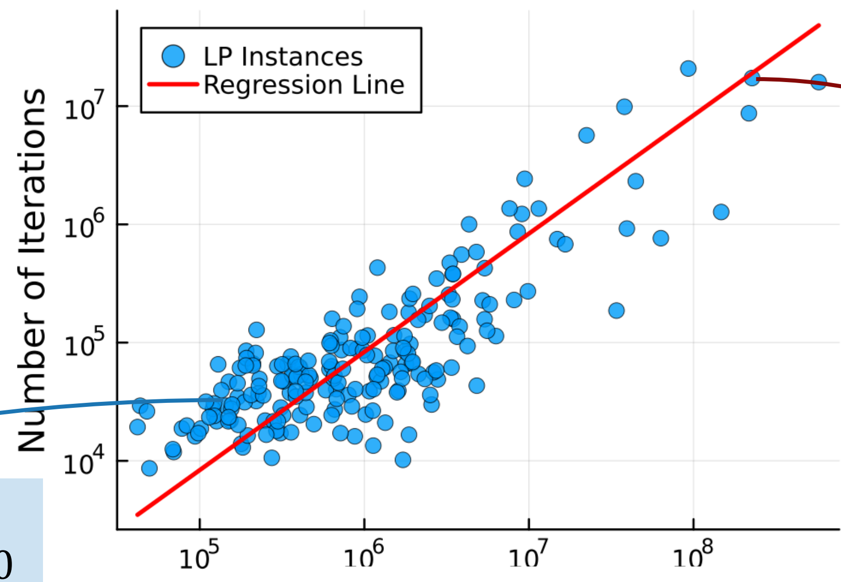
[Güler and Ye, 1993] [Ye, 1992] [Mehrotra and Ye, 1993] [Ye, 2011] …

The product of $\left(\frac{\|x^\star\|_1+\|s^\star\|_1}{\min_{i\in[n]} x_i^\star+s_i^\star}\right)$ and a norm of $B^{-1}A$ also appears in IPM complexity:

[Potra, 1994][Anstreicher, Ji, Potra and Ye, 1999] …

Validation Experiments on random LP instances

Iterations Required to solve LP instance to $\varepsilon = 10^{-8}$



Value of $\kappa \cdot \lim_{\delta \to 0} \frac{D_\delta}{r_\delta}$

200 random LP instances with $\mathrm{m} = 60$, and $\mathrm{n} = 120$

Regression line:
$\log_{10}(\text{PDHG Iteration}) = \log_{10} \kappa\Phi - 1.078$

1. Fairly obvious linear dependence on $\kappa \cdot \lim_{\delta \to 0} \frac{D_\delta}{r_\delta}$

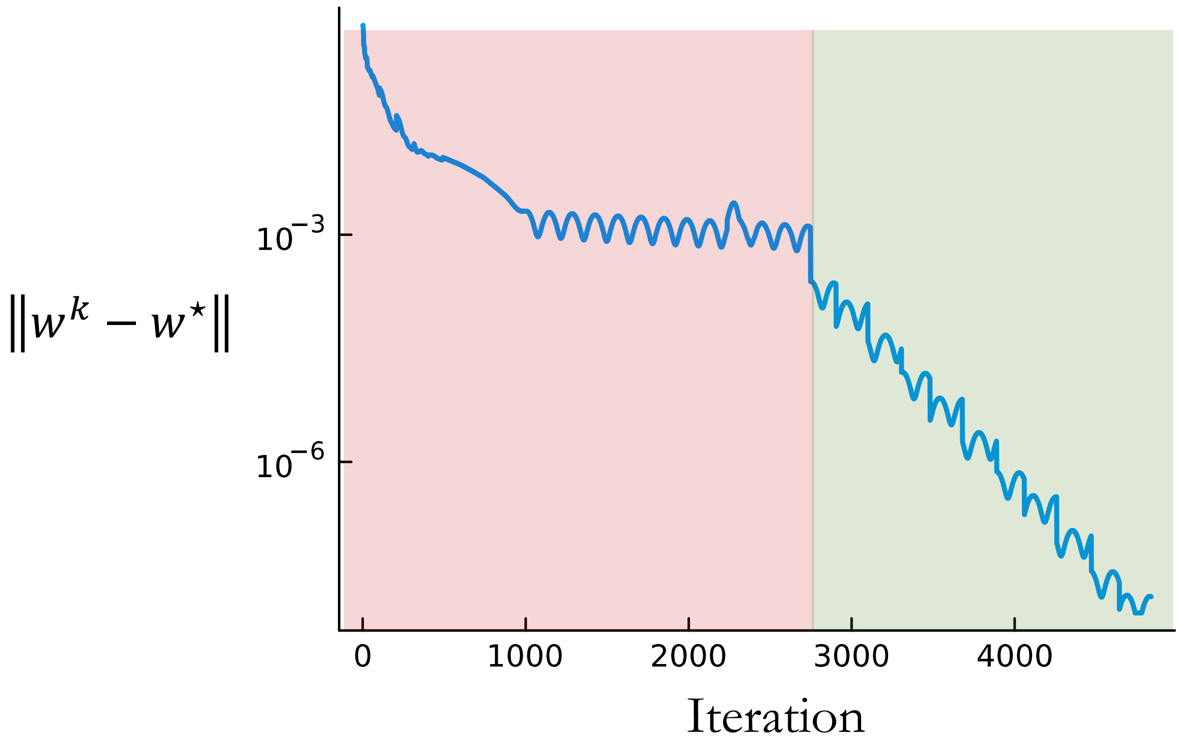
2. Although $\kappa \cdot \lim_{\delta \to 0} \frac{D_\delta}{r_\delta}$ can be extremely large in practice (or not), the random LP instances typically have smaller values of $\kappa \cdot \lim_{\delta \to 0} \frac{D_\delta}{r_\delta}$. More on this later…

Regarding the two-stage performance of PDHG on LP instances with unique optima

Two-Stage Performance of PDHG

Two-stage performance of PDHG on general LP [Lu and Yang 2023]

Typical convergence performance of PDHG



Stage I:
Finite-time basis identification

Stage II:
Fast local convergence

Two-Stage Performance of PDHG (under unique optima)

Theorem (Stage I: Finite-time basis identification) [Xiong 2024]:

Let $w^k = (x^k, s^k)$ denote the k-th iteration solution. The solution x^k identifies the optimal basis $\mathcal{B}$ (i.e. $\mathrm{supp}(x^k) = \mathcal{B}$) for all $k \geq T_{\text{basis}}$, where :

$$T_{\text{basis}} := \tilde{O}\left(\kappa \cdot \lim_{\delta \to 0} \frac{\boldsymbol{D_\delta}}{\boldsymbol{r_\delta}}\right)$$

Theorem (Stage II: Fast local convergence)[Xiong 2024]:

Once the optimal basis has been determined, PDHG computes an $\boldsymbol{\varepsilon}$-optimal solution within an additional

$$T_{\text{local}} := \tilde{O}\left(\|B^{-1}\|\|A\| \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)$$

iterations.

$\|B^{-1}\|\|A\|$: Only matrix condition numbers.

Stage II converges faster because it is not affected by the sublevel set geometry

Can we explain the practical performance of PDHG using high-probability complexity analysis?

Todd's Classic Random LP Model

Definition (Random LP Model):

Select $\mathcal{B} \subset [n]$, and $|\mathcal{B}| = m$ and the solution $(\hat{x}, \hat{s})$ is distributed as follows:

$$\hat{x}_{\mathcal{B}} \sim |\mathcal{N}(0,1)|^m, \qquad \hat{x}_{\mathcal{N}} = 0,$$
$$\hat{s}_{\mathcal{B}} = 0, \qquad \hat{s}_{\mathcal{N}} \sim |\mathcal{N}(0,1)|^{n-m},$$

The random LP of the above optimal solution is distributed as follows:

$$A \sim \mathcal{N}(0,1)^{m \times n}, \qquad b = A\hat{x}, \qquad c = \hat{s}.$$

From a classic random LP model of [Todd, 1991].
Variants studied for IPM [Ye, 1994], [Anstreicher et al., 1999]
We use unit variance for simplicity of result.

This LP model has unique optima with probability = 1

Two-Stage Performance of PDHG

Theorem (Stage I: Finite-time basis identification) [Xiong 2024]:
Let T_{basis} denote the number of PDHG iterations to identify the optimal basis. Then it holds for any $\delta \in \left(\frac{1}{2^{c_0 n}}, 1\right)$ that

$$\mathbb{P}\left[T_{\text{basis}} \leq \tilde{O}\left(\frac{n^{2.5}}{\delta}\right)\right] \geq 1 - \delta.$$

Theorem (Stage II: Fast local convergence) [Xiong 2024]:
After T_{basis} iterations, let T_{local} denote the number of additional PDHG iterations to compute an $\boldsymbol{\varepsilon}$-optimal solution. Then it holds for any $\delta \in (0,1)$ that

$$\mathbb{P}\left[T_{\text{local}} \leq \tilde{O}\left(\frac{n}{\delta} \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)\right] \geq 1 - \delta.$$

Faster local linear convergence indicated by the probabilistic analysis.

Theorem (Stage I: Finite-time basis identification) [Xiong 2024]:
Let T_{basis} denote the number of PDHG iterations to identify the optimal basis. Then it holds for any $\delta \in \left(\frac{1}{2^{c_0 n}}, 1\right)$ that

$$\mathbb{P}\left[T_{\text{basis}} \leq \tilde{O}\left(\frac{n^{2.5}}{\delta}\right)\right] \geq 1 - \delta.$$

Theorem (Stage II: Fast local convergence) [Xiong 2024]:
After T_{basis} iterations, let T_{local} denote the number of additional PDHG iterations to compute an $\boldsymbol{\varepsilon}$-optimal solution. Then it holds for any $\delta \in (0,1)$ that

$$\mathbb{P}\left[T_{\text{local}} \leq \tilde{O}\left(\frac{n}{\delta} \cdot \ln\left(\frac{1}{\varepsilon}\right)\right)\right] \geq 1 - \delta.$$

This provides a possible explanation for why PDHG works well in practice (polynomial-time in most cases), even though $\kappa \cdot \lim_{\delta \to 0} \frac{D_\delta}{r_\delta}$ may take extreme values in the worst case.

But PDHG bound has a heavier tail compared with the two classic methods:

- PDHG has polynomial high-probability complexity
- IPM has polynomial average-case complexity. [Anstreicher, et al., 1999]

Summary/Remarks

- The convergence rate of PDHG for conic optimization is related to the geometry of primal-dual sublevel sets measured with $\boldsymbol{D_\delta}$, $\boldsymbol{r_\delta}$, $\boldsymbol{d_\delta^H}$
- For LP instances with unique optima, the iteration bound has a closed-form expression
- For LP instances with unique optima, PDHG has faster local convergence after identifying the optimal basis
- PDHG is polynomial-time with high probability

Remark:

- These results relied only on PDHG's average iterate convergence and non-expansiveness properties. Similar results might also hold for other FOMs, in particular ADMM, EGM, …

Thank you!