Barrier Smoothing for Nonsmooth Convex Minimization

Quoc Tran-Dinh
Yen-Huan Li
Volkan Cevher

*Laboratory for Information and Inference Systems-**LIONS***
École Polytechnique Fédérale de Lausanne (EPFL)
{quoc.trandinh, yen-huan.li, volkan.cevher}@epfl.ch

ICASSP 2014
Florence, May 04-09

Outline

- Prologue
- Theory of barrier smoothing
- Numerical examples
- Conclusions and future work

PROLOGUE

A Stylized Example

Basis Pursuit:

$$\max_{y \in \mathbb{R}^n} \{- \|y\|_1 : \Phi y = v, y \in \mathcal{Y}\}$$

$\mathcal{Y}$ a given convex closed set

Equivalent formulation:

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \max_{y \in \mathcal{Y}} \left\{ \left\langle \Phi^T x, y \right\rangle - \|y\|_1 \right\} - \langle v, x \rangle \right\}$$

Formal Problem Formulation

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \max_{y \in \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) \} + \langle c, x \rangle \right\}$$

G a convex function

[Nesterov'2005]

Recall the basis pursuit:

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \max_{y \in \mathcal{Y}} \left\{ \left\langle \Phi^T x, y \right\rangle - \|y\|_1 \right\} - \langle v, x \rangle \right\}$$

Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program., Ser. A*, 2005

Main Idea of Smoothing

$$\min_{x \in \mathbb{R}^n} \left\{ f_\lambda(x) := \max_{y \in \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \lambda h(y) \} + \langle c, x \rangle \right\}$$

λ small enough $\Rightarrow$ $\min f_\lambda \leq (1 + \varepsilon_\lambda) \min f$

$y^*_\lambda(x)$ unique $\Rightarrow$ $\nabla f_\lambda(x) = A^T y^*_\lambda(x) + c$ exists

$$y^*_\lambda(x) = \arg\max_{y \in \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \lambda h(y) \}$$

Smoothing by *Proximity Functions*

$$\min_{x \in \mathbb{R}^n} \left\{ f_\tau(x) := \max_{y \in \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \tau p_{\mathcal{Y}}(y) \} + \langle c, x \rangle \right\}$$

$p_{\mathcal{Y}}$ strongly convex ⇨ ∇f_τ exists and ∇f_τ Lipschitz

$$\|\nabla f_\tau(y) - \nabla f_\tau(x)\| \leq L \|y - x\|$$

⇨ **First-order methods**

Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program., Ser. A*, 2005

Smoothing by *Self-Concordant Barriers*

$$\min_{x\in\mathbb{R}^n}\left\{f_\tau(x) := \max_{y\in\mathcal{Y}}\{\langle Ax, y\rangle - G(y) - \tau p_{\mathcal{Y}}(y)\} + \langle c, x\rangle\right\}$$

$p_{\mathcal{Y}}$ strongly convex ⇨ ∇f_τ exists and ∇f_τ Lipschitz

Smoothing by proximity-functions

Smoothing by self-concordant barriers

$$\min_{x\in\mathbb{R}^n}\left\{f_\sigma(x) := \max_{y\in\operatorname{int}\mathcal{Y}}\{\langle Ax, y\rangle - G(y) - \sigma b_{\mathcal{Y}}(y)\} + \langle c, x\rangle\right\}$$

$b_{\mathcal{Y}}$ self-concordant barrier ⇨ ∇f_σ exists and ∇f_σ Lipschitz-like

Our Contribution

- **The framework of *barrier smoothing***

- **A gradient method with *performance guarantee***

Benefits of Our Contribution

- **The framework of barrier smoothing**
  - Calculating ∇f_σ is *easier*, compared with smoothing by proximity functions.

- **A gradient method with performance guarantee**
  - *Analytic, optimal adaptive step-size*

The Barrier Smoothing Framework

WHY IS CALCULATING THE GRADIENT EASIER?

Self-Concordant Barriers: Definition

b_Ω is a ν-***self-concordant barrier for the set*** Ω if

$$|\phi'''(t)| \leq 2\left[\phi''(t)\right]^{3/2}$$ (self-concordance)

$$|\phi'(t)| \leq \sqrt{\nu}\left[\phi''(t)\right]^{1/2}$$

$$b_\Omega(x) \to \infty \text{ as } x \to \partial\Omega$$

(barrier property)

$$\phi(t) := b_\Omega(x + tv)$$

$$\forall t \in \mathbb{R}, x + tv \in \operatorname{dom} b_\Omega$$

Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithm in Convex Programming*, 1994

Self-Concordant Barriers: Examples

$$\Omega = [\ell, u] \qquad b_\Omega(x) = -\log(x - \ell) - \log(u - x)$$

$$\Omega = \mathbb{R}_+ \qquad b_\Omega(x) = -\log(x)$$

$$\Omega = \mathcal{S}_+ \qquad b_\Omega(X) = -\log(\det X)$$

$$\vdots \qquad \vdots$$

Calculating the Gradient Becomes Easy! (1/3)

b_{Ω} is a ν-***self-concordant barrier for the set*** Ω if...

$$b_{\Omega}(x) \to \infty \text{ as } x \to \partial\Omega$$

$$\min_{x \in \mathbb{R}^n} \left\{ f_{\sigma}(x) := \max_{y \in \mathcal{Y}} \{\langle Ax, y\rangle - G(y) - \sigma b_{\mathcal{Y}}(y)\} + \langle c, x\rangle \right\}$$

$$\Updownarrow$$

$$\min_{x \in \mathbb{R}^n} \left\{ f_{\sigma}(x) := \max_{y \in \operatorname{int} \mathcal{Y}} \{\langle Ax, y\rangle - G(y) - \sigma b_{\mathcal{Y}}(y)\} + \langle c, x\rangle \right\}$$

Calculating the Gradient Becomes Easy! (2/3)

$$\min_{x \in \mathbb{R}^n} \left\{ f_\sigma(x) := \max_{y \in \mathrm{int}\, \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \sigma b_{\mathcal{Y}}(y) \} + \langle c, x \rangle \right\}$$

$$\nabla f_\sigma(x) = A^T y_\sigma^*(x) + c$$

$$y_\sigma^*(x) := \arg \max_{y \in \mathrm{int}\, \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \sigma b_{\mathcal{Y}}(y) \}$$

⇕ G smooth

$$y_\sigma^*(x) \text{ s.t. } Ax - \nabla G(y_\sigma^*(x)) - \sigma \nabla b_{\mathcal{Y}}(y_\sigma^*(x)) = 0$$

Recall: Smoothing by *Proximity Functions*

$$\min_{x \in \mathbb{R}^n} \left\{ f_\tau(x) := \max_{y \in \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \tau p_{\mathcal{Y}}(y) \} + \langle c, x \rangle \right\}$$

$$\nabla f_\tau(x) = A^T y_\tau^*(x) + c$$

$$y_\tau^*(x) := \arg\max_{y \in \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \tau p_{\mathcal{Y}}(y) \}$$

Another constrained convex optimization problem!

Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program., Ser. A*, 2005

Calculating the Gradient Becomes Easy! (3/3)

$$\min_{x\in\mathbb{R}^n}\left\{f_\sigma(x) := \max_{y\in \mathrm{int}\,\mathcal{Y}}\{\langle Ax, y\rangle - G(y) - \sigma b_{\mathcal{Y}}(y)\} + \langle c, x\rangle\right\}$$

$$y^*_\sigma(x) \text{ s.t. } Ax - \nabla G(y^*_\sigma(x)) - \sigma\nabla b_{\mathcal{Y}}(y^*_\sigma(x)) = 0$$

Smoothing by self-concordant barriers

Smoothing by proximity functions

$$\min_{x\in\mathbb{R}^n}\left\{f_\tau(x) := \max_{y\in\mathcal{Y}}\{\langle Ax, y\rangle - G(y) - \tau p_{\mathcal{Y}}(y)\} + \langle c, x\rangle\right\}$$

$$y^*_\tau(x) \text{ s.t. } \langle Ax - \nabla G(y^*_\tau(x)) - \sigma\nabla b_{\mathcal{Y}}(y^*_\tau(x)), y - y^*_\sigma(x)\rangle \leq 0$$

$$\forall y \in \mathcal{Y}$$

A Gradient Method

HOW TO GET THE ANALYTIC, OPTIMAL ADAPTIVE STEP-SIZE?

Overview of the Algorithm

$$\min_{x \in \mathbb{R}^n} \left\{ f_\sigma(x) := \max_{y \in \operatorname{int} \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \sigma b_{\mathcal{Y}}(y) \} + \langle c, x \rangle \right\}$$

1. Find $y_\sigma^*(x^k)$ such that

   $$Ax^k - \nabla G(y_\sigma^*(x^k)) - \sigma \nabla b_{\mathcal{Y}}(y_\sigma^*(x^k)) = 0$$

2. Compute $\nabla f_\sigma(x^k) = A^T y_\sigma^*(x^k) + c$

3. Update $x^{k+1} = x^k - \alpha_k \nabla f_\sigma(x^k)$

Recall: Smoothing by *Proximity Functions*

$$\min_{x \in \mathbb{R}^n} \left\{ f_\tau(x) := \max_{y \in \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \tau p_{\mathcal{Y}}(y) \} + \langle c, x \rangle \right\}$$

∇f_τ exists and ∇f_τ Lipschitz

$$\|\nabla f_\tau(y) - \nabla f_\tau(x)\| \leq L \|y - x\|$$

$$f_\tau(y) \leq f_\tau(x) + \langle \nabla f_\tau(x), y - x \rangle + \frac{1}{2L} \|y - x\|^2 := \mathrm{bound}(y; x)$$

$$x^{k+1} = x^k - \frac{1}{L} \nabla f_\tau(x^k) = \arg\min_y \{\mathrm{bound}(y; x^k)\}$$

Analytic, optimal, non-adaptive!

Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program., Ser. A*, 2005

Smoothing by *Barrier Functions*

$$\min_{x \in \mathbb{R}^n} \left\{ f_\sigma(x) := \max_{y \in \mathrm{int}\, \mathcal{Y}} \{ \langle Ax, y \rangle - G(y) - \sigma b_{\mathcal{Y}}(y) \} + \langle c, x \rangle \right\}$$

∇f_σ exists and ∇f_σ Lipschitz-like

$$\|\nabla f_\sigma(y) - \nabla f_\sigma(x)\| \leq \frac{(c_A)^2 \|y - x\|}{\sigma - c_A \|y - x\|}$$ when $c_A \|y - x\| < \sigma$

$$f_\sigma(y) \leq f_\sigma(x) + \langle \nabla f_\sigma(x), y - x \rangle + \sigma \omega_* \left(\sigma^{-1} c_A \|y - x\| \right) := \mathrm{bound}\,(y; x)$$

$$\omega_*(x) := -x - \ln(1 - x)$$

$$x^{k+1} = x^k - \frac{\sigma}{c_A^k \left(c_A^k + r_k \right)} \nabla f_\tau(x^k) = \arg\min_y \{ \mathrm{bound}(y; x^k) \}$$

Analytic, optimal, adaptive!

$$c_A^k := c_A(x^k), \quad c_A := c_A(x) := \left\| A^T \nabla^2 b_{\mathcal{Y}}(y_\sigma^*(x))^{-1} A \right\|^{1/2}$$

Q. Tran-Dinh, *Sequential convex programming and decomposition approaches for nonlinear optimization*, 2012

Performance Guarantee

With the optimal analytic adaptive step-size:

$$f_\sigma^* = f_\sigma(x_\sigma^*) = \min_{x \in \mathbb{R}^n} \{f_\sigma(x)\}, \quad \overline{c_A} := \sup_{x \in \mathrm{dom}\, f_\sigma} \{c_A(x)\}$$

$$f_\sigma(x^k) - f_\sigma^* \leq \frac{4\overline{c_A}^2 \left\| x^0 - x_\sigma^* \right\|^2}{\sigma k}$$

Performance Guarantee

With the optimal analytic adaptive/nonadaptive step-size:

$$f_\sigma^* = f_\sigma(x_\sigma^*) = \min_{x \in \mathbb{R}^n} \{f_\sigma(x)\}, \quad \overline{c_A} := \sup_{x \in \mathrm{dom}\, f_\sigma} \{c_A(x)\}$$

$$f_\sigma(x^k) - f_\sigma^* \le \frac{4\overline{c_A}^2 \left\| x^0 - x_\sigma^* \right\|^2}{\sigma k}$$

Smoothing by self-concordant barriers

Smoothing by proximity functions

$$f_\tau^* = f_\tau(x_\tau^*) = \min_{x \in \mathbb{R}^n} \{f_\tau(x)\}, \quad \|\nabla f_\tau(y) - \nabla f_\tau(x)| \le L \|y - x\|$$

$$f_\tau(x^k) - f_\tau^* \le \frac{2L \left\| x^0 - x_\sigma^* \right\|^2}{k + 4}$$

Y. Nesterov, *Introductory Lectures on Convex Optimization*, 2004

NUMERICAL RESULTS

Examples 1/2: Basis Pursuit

Original problem:

$$\max_{y \in \mathbb{R}^n} \{- \|y\|_1 : \Phi y = b, y \in \mathcal{Y}\}$$

$$\mathcal{Y} := [-1/2, 1/2]^n$$

Equivalent dual formulation:

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \max_{y \in \mathcal{Y}} \{ \langle \Phi^T x, y \rangle - \|y\|_1 \} - \langle b, x \rangle \right\}$$

Examples 1/2: Basis Pursuit

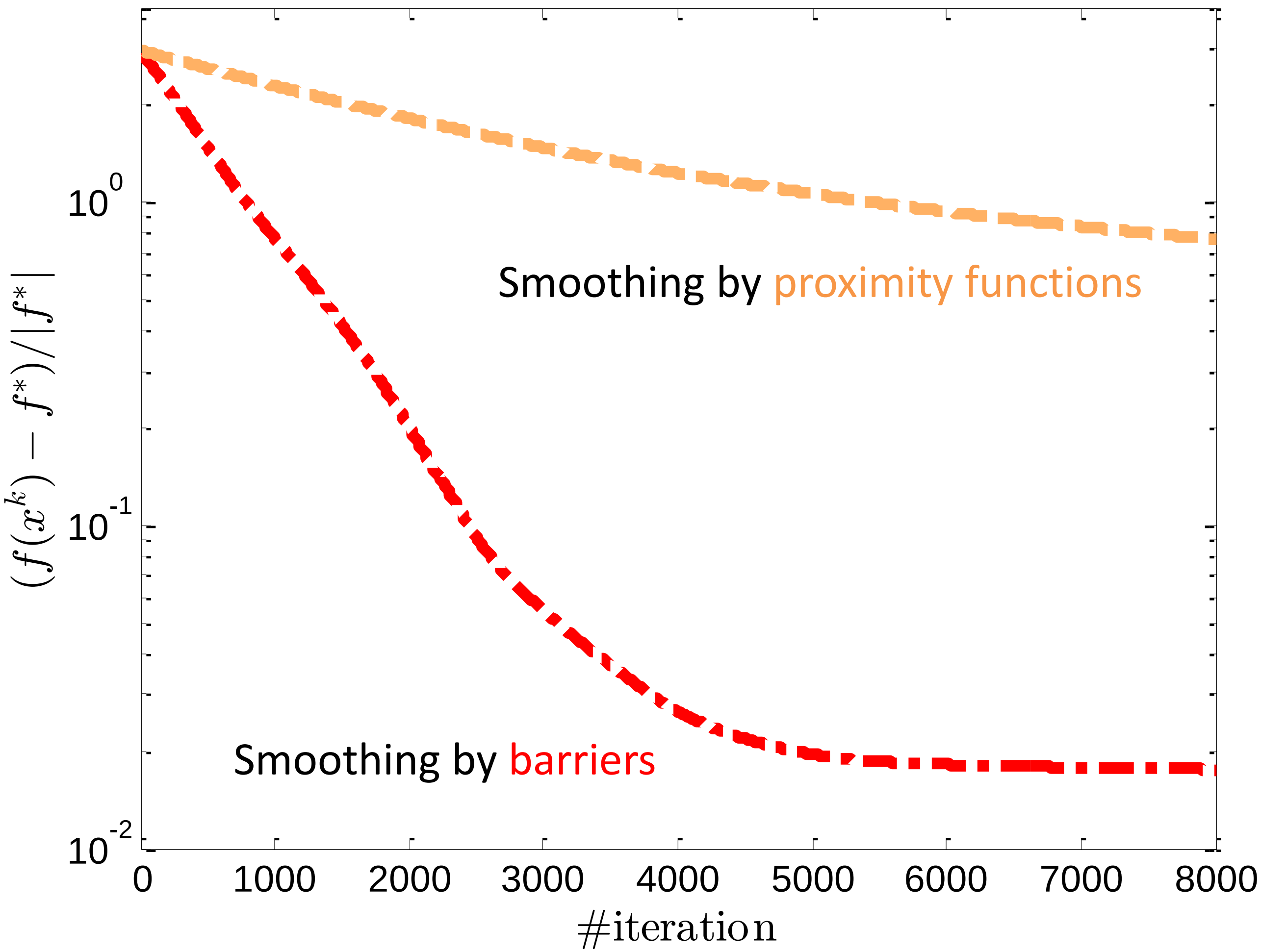
$$f^* := \min_{x \in \mathbb{R}^n} \left\{ f(x) := \max_{y \in \mathcal{Y}} \{ \langle \Phi^T x, y \rangle - \|y\|_1 \} - \langle b, x \rangle \right\}$$

$$\mathcal{Y} := [-1/2, 1/2]^n$$

Smoothing by barrier $\quad b_{\mathcal{Y}}(y) = -\sum_{i=1}^{m} [\log(y_i - \ell_i) + \log(u_i - y_i)]$

Smoothing by proximity functions $\quad p_{\mathcal{Y}}(y) = \frac{1}{2} \|y\|^2$

Examples 1/2: Basis Pursuit



Examples 2/2: Quadratically Constrained Quadratic Programming (QCQP)

Original problem:

$$\min_{y \in \mathbb{R}^m} \left\{ \langle y, Qy \rangle + \langle b, y \rangle : \langle By, y \rangle \leq 1, A^T y + c = 0 \right\}$$

$$Q \geq 0, B > 0, B = B^T$$

Equivalent dual formulation:

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) := \max_{y \in \mathbb{R}^m : \langle By, y \rangle \leq 1} \left\{ \langle Ax - b, y \rangle - \frac{1}{2} \langle Qy, y \rangle \right\} + \langle c, x \rangle \right\}$$

Q singular $\Longrightarrow$ f nonsmooth

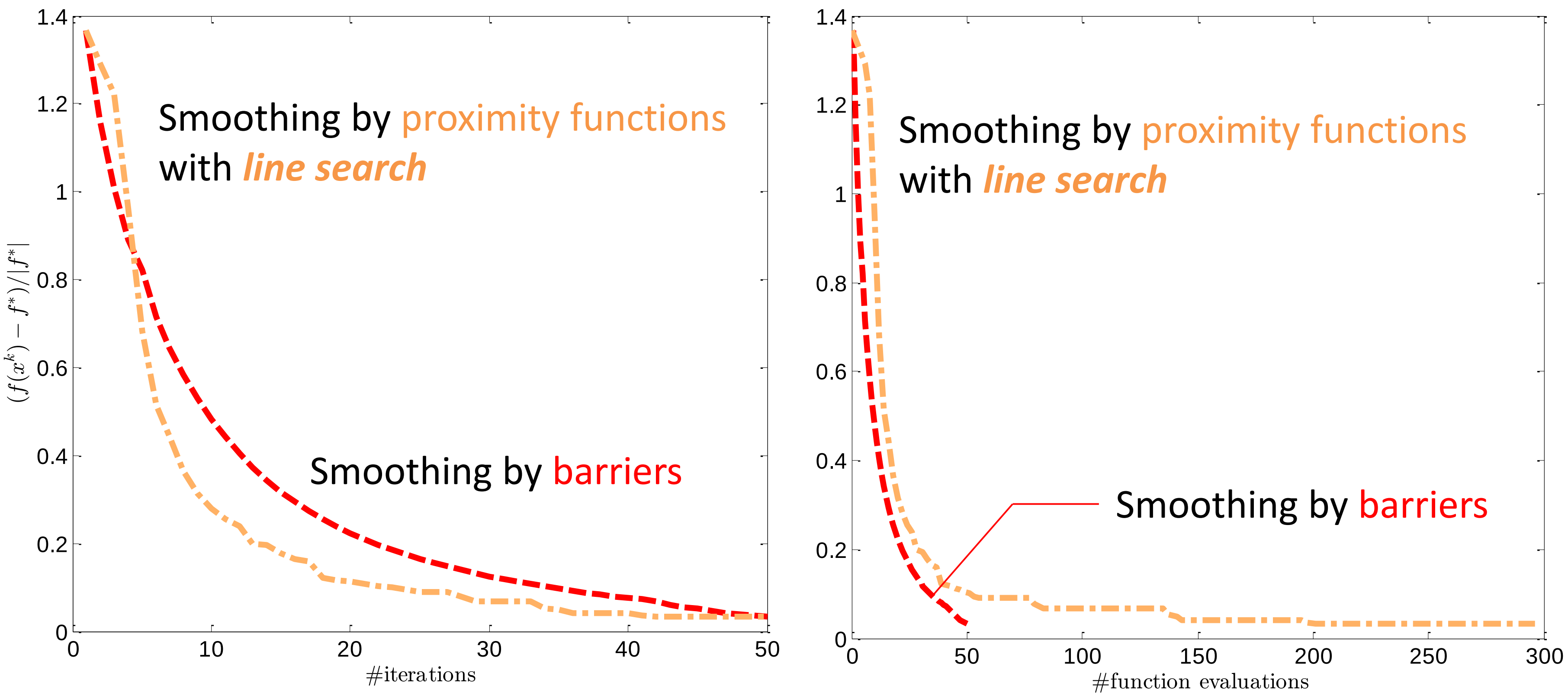
Examples 2/2: Quadratically Constrained Quadratic Programming (QCQP)

$$f^* := \min_{x \in \mathbb{R}^n} \left\{ f(x) := \max_{\langle By, y \rangle \leq 1} \left\{ \langle Ax - b, y \rangle - \frac{1}{2} \langle Qy, y \rangle \right\} + \langle c, x \rangle \right\}$$

Smoothing by barrier $\quad b_{\mathcal{Y}}(y) = -\log\left(1 - \langle y, By \rangle\right)$

Smoothing by proximity functions $\quad p_{\mathcal{Y}}(y) = \frac{1}{2} \langle y, By \rangle$

Examples 2/2: Quadratically Constrained Quadratic Programming (QCQP)



Observation

Better empirical convergence behavior!

Recall the analytic optimal adaptive step-size

$$x^{k+1} = x^k - \frac{\sigma}{c_A^k\left(c_A^k + r_k\right)} \nabla f_\tau(x^k) = \arg\min_x \{\text{bound}(x)\}$$

$$c_A^k := c_A(x^k) := \left\| A^T \nabla^2 b_{\mathcal{Y}}(y_\sigma^*(x^k))^{-1} A \right\|^{1/2}$$

CONCLUSIONS & FUTURE WORK

Comparison

[Nesterov'2005]

	Barrier function	Proximity function
Convergence Behavior	$O\left(\frac{1}{k}\right)$	$O\left(\frac{1}{k}\right)$
Complexity-per-iteration	Solving a nonlinear equation	Solving a constrained convex minimization problem

Y. Nesterov, "Smooth minimization of non-smooth functions," *Math. Program., Ser. A*, 2005

Our Contribution

- **The framework of *barrier smoothing***

- **A gradient method with *performance guarantee***

Benefits of Our Contribution

- **The framework of barrier smoothing**
  - Calculating ∇f_σ is *easier*, compared with smoothing by proximity functions.

- **A gradient method with performance guarantee**
  - *Analytic, optimal adaptive step-size*
  - *Better empirical convergence behavior*

Future Work

- **Accelerated gradient method**
- **Nonsmooth Composite Minimization**

THANKS FOR YOUR ATTENTION!