A polynomial time interior point method for problems with nonconvex constraints

Oliver Hinder, Yinyu Ye

Department of Management Science and Engineering
Stanford University

June 28, 2018

The problem

- Consider problem

$$\min_{x \in \mathbb{R}^d} f(x) \text{ s.t. } a(x) \leq 0$$

where $f : \mathbb{R}^d \to \mathbb{R}$ and $a : \mathbb{R}^d \to \mathbb{R}^m$ have Lipschitz first and second derivatives.

The problem

- Consider problem

$$\min_{x \in \mathbb{R}^d} f(x) \text{ s.t. } a(x) \leq 0$$

  where $f : \mathbb{R}^d \to \mathbb{R}$ and $a : \mathbb{R}^d \to \mathbb{R}^m$ have Lipschitz first and second derivatives.

- Captures a huge variety of practical problems ... drinking water network, electricity network, trajectory optimization, engineering design, etc.

The problem

- Consider problem

$$\min_{x \in \mathbb{R}^d} f(x) \text{ s.t. } a(x) \leq 0$$

  where $f : \mathbb{R}^d \to \mathbb{R}$ and $a : \mathbb{R}^d \to \mathbb{R}^m$ have Lipschitz first and second derivatives.
- Captures a huge variety of practical problems ... drinking water network, electricity network, trajectory optimization, engineering design, etc.
- In worst-case finding global optima takes an exponential amount of time.

The problem

- Consider problem

$$\min_{x \in \mathbb{R}^d} f(x) \text{ s.t. } a(x) \leq 0$$

  where $f : \mathbb{R}^d \to \mathbb{R}$ and $a : \mathbb{R}^d \to \mathbb{R}^m$ have Lipschitz first and second derivatives.
- Captures a huge variety of practical problems ... drinking water network, electricity network, trajectory optimization, engineering design, etc.
- In worst-case finding global optima takes an exponential amount of time.
- Instead we want to find an 'approximate local optima', more precisely a Fritz John point.

μ-approximate Fritz John point

$a(x) < \mathbf{0}$ **(Primal feasibility)**

$\left\|\nabla_x f(x) - y^T \nabla a(x)\right\|_2 \leq \mu \sqrt{\|y\|_1 + 1} \quad y > \mathbf{0}$ **(Dual feasibility)**

$\frac{y_i a_i(x)}{\mu} \in [1/2, 3/2] \quad \forall i \in \{1, \ldots, m\}$ **(Complementarity)**

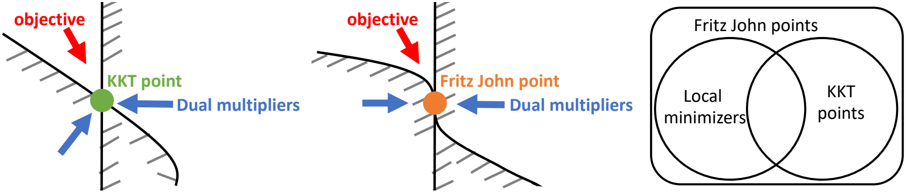
μ-approximate Fritz John point

$$a(x) < \mathbf{0} \qquad \textbf{(Primal feasibility)}$$

$$\left\|\nabla_x f(x) - y^T \nabla a(x)\right\|_2 \le \mu\sqrt{\|y\|_1 + 1} \quad y > \mathbf{0} \qquad \textbf{(Dual feasibility)}$$

$$\frac{y_i a_i(x)}{\mu} \in [1/2, 3/2] \quad \forall i \in \{1, \dots, m\} \qquad \textbf{(Complementarity)}$$

- Fritz John is a necessary condition for local optimality
- MFCQ constraint qualification, i.e., dual multipliers are bounded then Fritz John = KKT point



How do we solve this problem?

How do we solve this problem?

- One popular approach, move constraints into a log barrier:

$$\psi_\mu(x) := f(x) - \mu \log(-a(x)).$$

and apply (modified) Newton's method to find an approximate local optima.

How do we solve this problem?

- One popular approach, move constraints into a log barrier:

$$\psi_\mu(x) := f(x) - \mu \log(-a(x)).$$

  and apply (modified) Newton's method to find an approximate local optima.
- Used for real systems, e.g., drinking water networks, electricity networks, trajectory control, etc.

How do we solve this problem?

- One popular approach, move constraints into a log barrier:

$$\psi_\mu(x) := f(x) - \mu \log(-a(x)).$$

  and apply (modified) Newton's method to find an approximate local optima.
- Used for real systems, e.g., drinking water networks, electricity networks, trajectory control, etc.
- Examples of practical codes using this method include IPOPT, KNITRO, LOQO, One-Phase (my code), etc.

How do we solve this problem?

- One popular approach, move constraints into a log barrier:

$$\psi_\mu(x) := f(x) - \mu \log(-a(x)).$$

  and apply (modified) Newton's method to find an approximate local optima.
- Used for real systems, e.g., drinking water networks, electricity networks, trajectory control, etc.
- Examples of practical codes using this method include IPOPT, KNITRO, LOQO, One-Phase (my code), etc.
- Local superlinear convergence is known.

How do we solve this problem?

- One popular approach, move constraints into a log barrier:

$$\psi_\mu(x) := f(x) - \mu \log(-a(x)).$$

  and apply (modified) Newton's method to find an approximate local optima.
- Used for real systems, e.g., drinking water networks, electricity networks, trajectory control, etc.
- Examples of practical codes using this method include IPOPT, KNITRO, LOQO, One-Phase (my code), etc.
- Local superlinear convergence is known.
- Global convergence: most results show convergence in limit but provide no explicit runtime bound. Implicitly runtime bound is exponential in $1/\mu$.

How do we solve this problem?

- One popular approach, move constraints into a log barrier:

$$\psi_\mu(x) := f(x) - \mu \log(-a(x)).$$

  and apply (modified) Newton's method to find an approximate local optima.
- Used for real systems, e.g., drinking water networks, electricity networks, trajectory control, etc.
- Examples of practical codes using this method include IPOPT, KNITRO, LOQO, One-Phase (my code), etc.
- Local superlinear convergence is known.
- Global convergence: most results show convergence in limit but provide no explicit runtime bound. Implicitly runtime bound is exponential in $1/\mu$.
- Goal: runtime with polynomial in $1/\mu$ to find μ-approximate Fritz John point.

Brief history of IPM theory

Brief history of IPM theory

Linear programming

Birth of interior point methods
[Karmarkar, 1984]

$$O(\sqrt{m}\log(1/\epsilon))$$

Brief history of IPM theory

Linear programming

Birth of interior point methods
[Karmarkar, 1984]

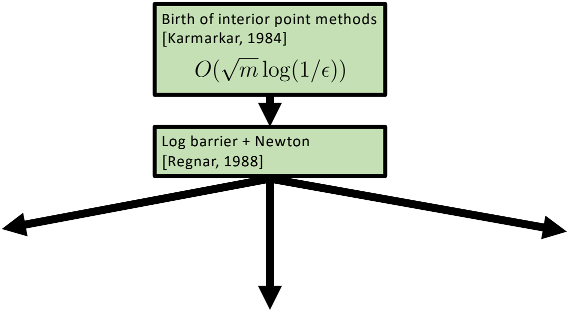
$$O(\sqrt{m}\log(1/\epsilon))$$

↓

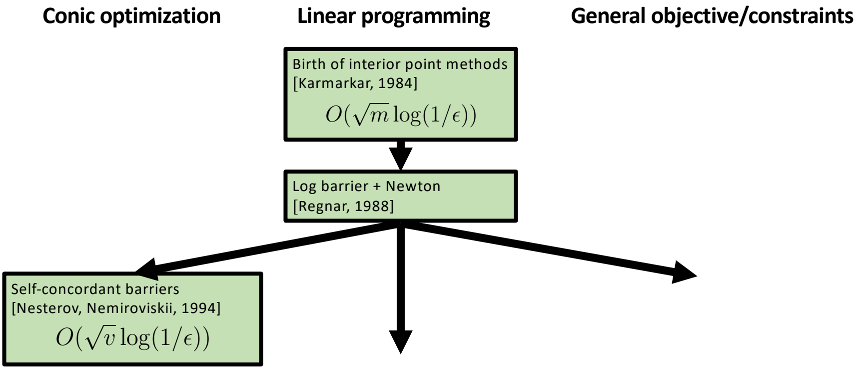
Log barrier + Newton
[Regnar, 1988]

Brief history of IPM theory

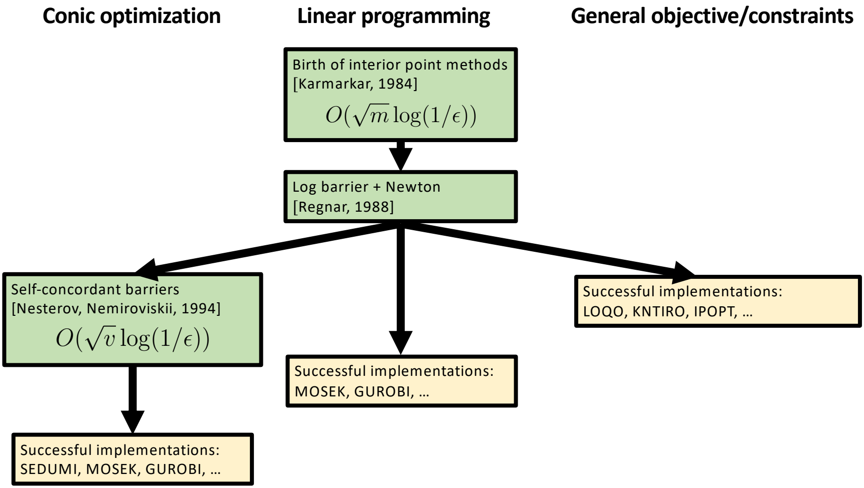
Linear programming



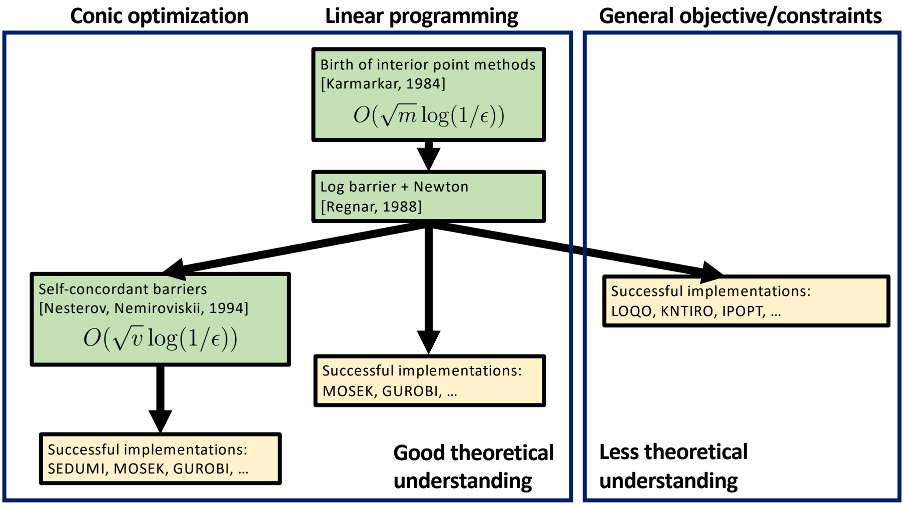
Brief history of IPM theory



Brief history of IPM theory



Brief history of IPM theory



Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.

Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.

Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.
  - Cubic regularization $O(\epsilon^{-3/2})$ [Nesterov, Polyak, 2006].

Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.
  - Cubic regularization $O(\epsilon^{-3/2})$ [Nesterov, Polyak, 2006].
  - Best achievable runtimes (of any method) with ∇f and $\nabla^2 f$ Lipschitz respectively [Carmon, Duchi, Hinder, Sidford, 2017].

Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.
  - Cubic regularization $O(\epsilon^{-3/2})$ [Nesterov, Polyak, 2006].
  - Best achievable runtimes (of any method) with ∇f and $\nabla^2 f$ Lipschitz respectively [Carmon, Duchi, Hinder, Sidford, 2017].
- Nonconvex objective, linear inequality constraints. Affine scaling IPM [Ye, (1998), Bian, Chen, and Ye (2015), Haeser, Liu, and Ye (2017)]. Has $O(\epsilon^{-3/2})$ runtime to find KKT points.

Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.
  - Cubic regularization $O(\epsilon^{-3/2})$ [Nesterov, Polyak, 2006].
  - Best achievable runtimes (of any method) with ∇f and $\nabla^2 f$ Lipschitz respectively [Carmon, Duchi, Hinder, Sidford, 2017].
- Nonconvex objective, linear inequality constraints. Affine scaling IPM [Ye, (1998), Bian, Chen, and Ye (2015), Haeser, Liu, and Ye (2017)]. Has $O(\epsilon^{-3/2})$ runtime to find KKT points.
- Nonconvex objective and constraints.

Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.
  - Cubic regularization $O(\epsilon^{-3/2})$ [Nesterov, Polyak, 2006].
  - Best achievable runtimes (of any method) with ∇f and $\nabla^2 f$ Lipschitz respectively [Carmon, Duchi, Hinder, Sidford, 2017].
- Nonconvex objective, linear inequality constraints. Affine scaling IPM [Ye, (1998), Bian, Chen, and Ye (2015), Haeser, Liu, and Ye (2017)]. Has $O(\epsilon^{-3/2})$ runtime to find KKT points.
- Nonconvex objective and constraints.
  - Apply cubic regularization to quadratic penalty function [Birgin et al, 2016]. Has $O(\epsilon^{-2})$ runtime.

Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.
  - Cubic regularization $O(\epsilon^{-3/2})$ [Nesterov, Polyak, 2006].
  - Best achievable runtimes (of any method) with ∇f and $\nabla^2 f$ Lipschitz respectively [Carmon, Duchi, Hinder, Sidford, 2017].
- Nonconvex objective, linear inequality constraints. Affine scaling IPM [Ye, (1998), Bian, Chen, and Ye (2015), Haeser, Liu, and Ye (2017)]. Has $O(\epsilon^{-3/2})$ runtime to find KKT points.
- Nonconvex objective and constraints.
  - Apply cubic regularization to quadratic penalty function [Birgin et al, 2016]. Has $O(\epsilon^{-2})$ runtime.
  - Sequential linear programming [Cartis et al, 2015] has $O(\epsilon^{-2})$ runtime.

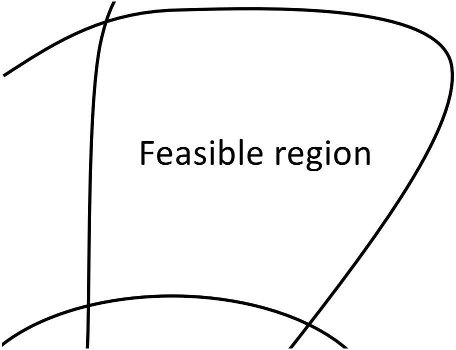
Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.
  - Cubic regularization $O(\epsilon^{-3/2})$ [Nesterov, Polyak, 2006].
  - Best achievable runtimes (of any method) with ∇f and $\nabla^2 f$ Lipschitz respectively [Carmon, Duchi, Hinder, Sidford, 2017].
- Nonconvex objective, linear inequality constraints. Affine scaling IPM [Ye, (1998), Bian, Chen, and Ye (2015), Haeser, Liu, and Ye (2017)]. Has $O(\epsilon^{-3/2})$ runtime to find KKT points.
- Nonconvex objective and constraints.
  - Apply cubic regularization to quadratic penalty function [Birgin et al, 2016]. Has $O(\epsilon^{-2})$ runtime.
  - Sequential linear programming [Cartis et al, 2015] has $O(\epsilon^{-2})$ runtime.
  - Both these methods plug and play directly with unconstrained optimization methods, e.g., they need their penalty functions to have Lipschitz derivatives.

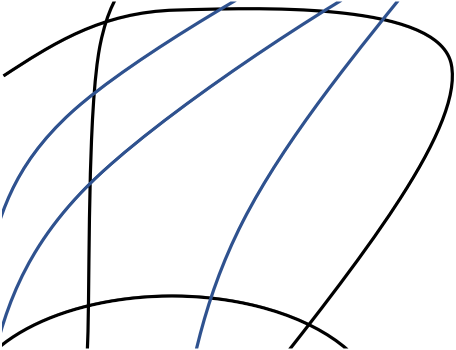
Literature review nonconvex optimization

- Unconstrained. Goal: find a point with $\|\nabla f(x)\|_2 \leq \epsilon$.
  - Gradient descent $O(\epsilon^{-2})$.
  - Cubic regularization $O(\epsilon^{-3/2})$ [Nesterov, Polyak, 2006].
  - Best achievable runtimes (of any method) with ∇f and $\nabla^2 f$ Lipschitz respectively [Carmon, Duchi, Hinder, Sidford, 2017].
- Nonconvex objective, linear inequality constraints. Affine scaling IPM [Ye, (1998), Bian, Chen, and Ye (2015), Haeser, Liu, and Ye (2017)]. Has $O(\epsilon^{-3/2})$ runtime to find KKT points.
- Nonconvex objective and constraints.
  - Apply cubic regularization to quadratic penalty function [Birgin et al, 2016]. Has $O(\epsilon^{-2})$ runtime.
  - Sequential linear programming [Cartis et al, 2015] has $O(\epsilon^{-2})$ runtime.
  - Both these methods plug and play directly with unconstrained optimization methods, e.g., they need their penalty functions to have Lipschitz derivatives.
  - Log barrier does not have Lipschitz derivatives!

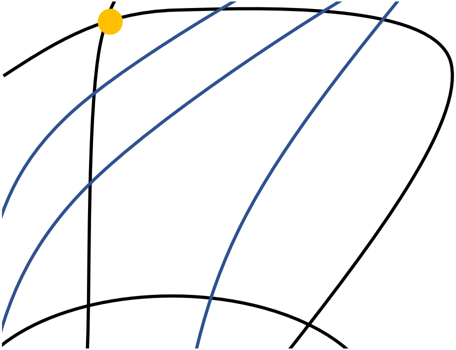
Our IPM for nonconvex optimization



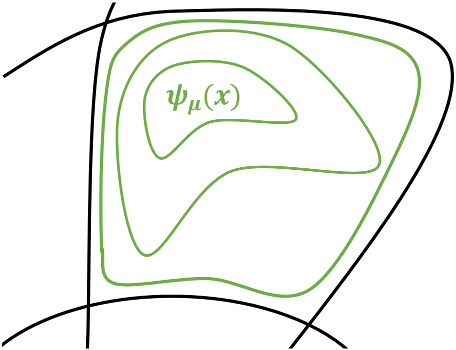
Our IPM for nonconvex optimization



Our IPM for nonconvex optimization



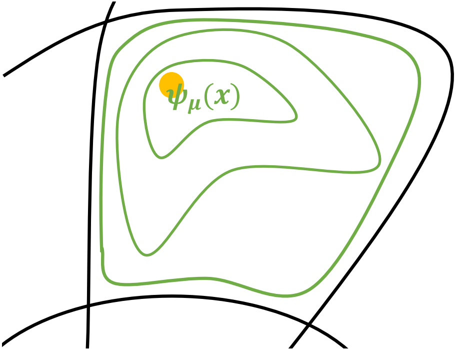
Our IPM for nonconvex optimization



Reminder:

$$\psi_\mu(x) := f(x) - \mu \log(-a(x)).$$

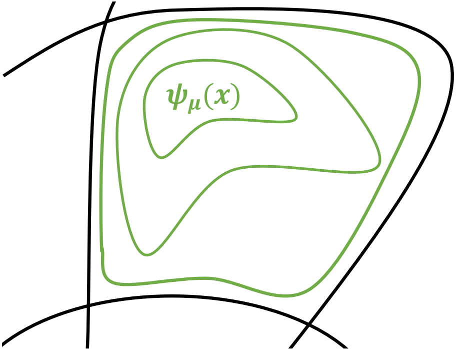
Our IPM for nonconvex optimization



Reminder:

$$\psi_\mu(x) := f(x) - \mu \log(-a(x)).$$

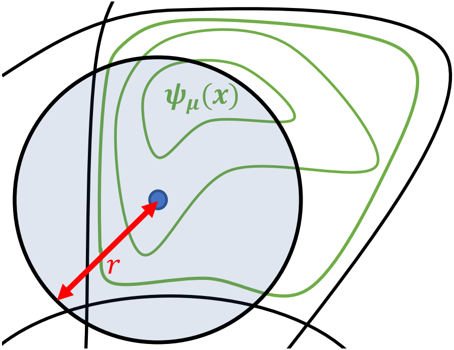
Our IPM for nonconvex optimization



Each iteration of our IPM we:

1. Pick trust region size $r \approx \mu^{3/4}(1 + \|y\|_1)^{-1/2}$.
2. Compute direction d_x by solving trust region problem.
3. Pick step size $\alpha \in (0, 1]$ to ensure $x_{new} = x + \alpha d_x$ satisfies $a(x) < 0$.
4. Is new point Fritz John point? If no then return to step one.

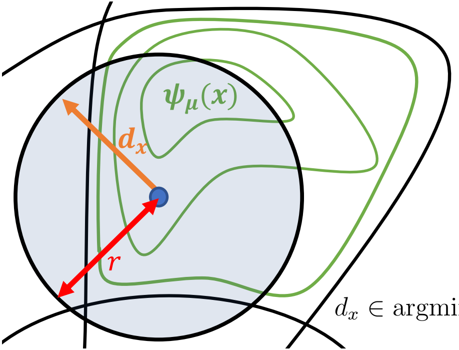
Our IPM for nonconvex optimization



Each iteration of our IPM we:

1. **Pick trust region size r** $\approx \mu^{3/4}(1 + \|\mathbf{y}\|_1)^{-1/2}$**.**
2. Compute direction d_x by solving trust region problem.
3. Pick step size $\alpha \in (0, 1]$ to ensure $x_{new} = x + \alpha d_x$ satisfies $a(x) < 0$.
4. Is new point Fritz John point? If no then return to step one.

Our IPM for nonconvex optimization

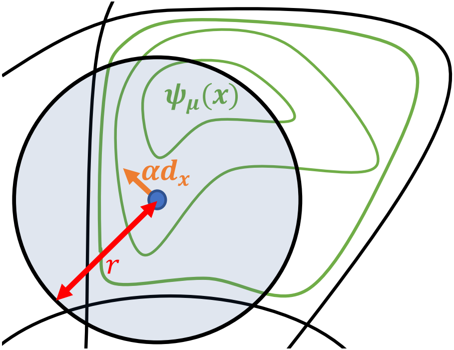


$$d_x \in \operatorname{argmin}_{u:\|u\|_2 \le r} \frac{1}{2} u^T \nabla^2 \psi_\mu(x) u + u^T \nabla \psi_\mu(x)$$

Each iteration of our IPM we:

1. Pick trust region size $r \approx \mu^{3/4}(1 + \|y\|_1)^{-1/2}$.
2. **Compute direction d_x by solving trust region problem.**
3. Pick step size $\alpha \in (0, 1]$ to ensure $x_{new} = x + \alpha d_x$ satisfies $a(x) < 0$.
4. Is new point Fritz John point? If no then return to step one.

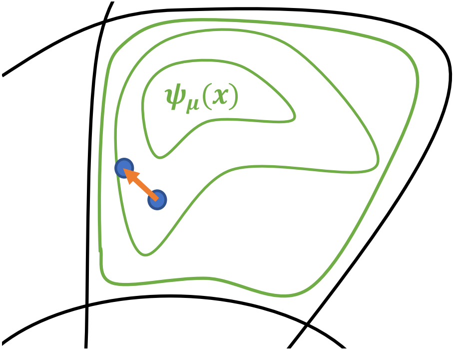
Our IPM for nonconvex optimization



Each iteration of our IPM we:

1. Pick trust region size $r \approx \mu^{3/4}(1 + \|y\|_1)^{-1/2}$.
2. Compute direction d_x by solving trust region problem.
3. **Pick step size $\alpha \in (0, 1]$ to ensure $\mathbf{x}_{\text{new}} = \mathbf{x} + \alpha \mathbf{d}_\mathbf{x}$ satisfies $\mathbf{a}(\mathbf{x}) < \mathbf{0}$.**
4. Is new point Fritz John point? If no then return to step one.

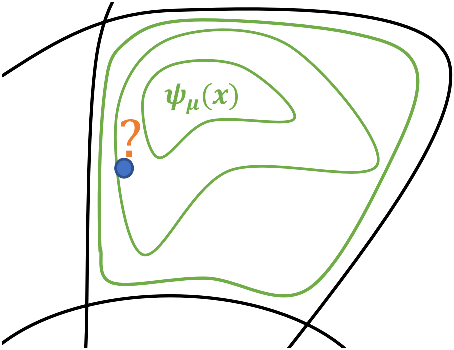
Our IPM for nonconvex optimization



Each iteration of our IPM we:

1. Pick trust region size $r \approx \mu^{3/4}(1 + \|y\|_1)^{-1/2}$.
2. Compute direction d_x by solving trust region problem.
3. **Pick step size $\alpha \in (0, 1]$ to ensure $\mathbf{x}_{new} = \mathbf{x} + \alpha \mathbf{d}_x$ satisfies $\mathbf{a}(\mathbf{x}) < \mathbf{0}$.**
4. Is new point Fritz John point? If no then return to step one.

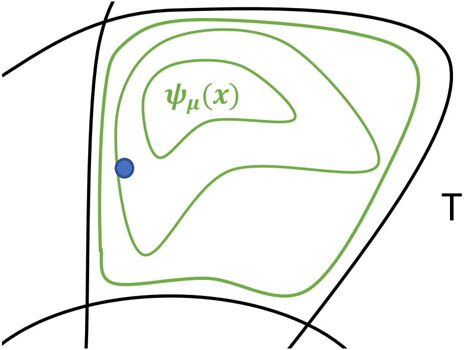
Our IPM for nonconvex optimization



Each iteration of our IPM we:

1. Pick trust region size $r \approx \mu^{3/4}(1 + \|y\|_1)^{-1/2}$.
2. Compute direction d_x by solving trust region problem.
3. Pick step size $\alpha \in (0, 1]$ to ensure $x_{\text{new}} = x + \alpha d_x$ satisfies $a(x) < 0$.
4. **Is new point Fritz John point? If no then return to step one.**

Our IPM for nonconvex optimization



Theoretical properties?

Each iteration of our IPM we:

1. Pick trust region size $r \approx \mu^{3/4}(1 + \|y\|_1)^{-1/2}$.
2. Compute direction d_x by solving trust region problem.
3. Pick step size $\alpha \in (0, 1]$ to ensure $x_{\text{new}} = x + \alpha d_x$ satisfies $a(x) < 0$.
4. Is new point Fritz John point? If no then return to step one.

Nonconvex result

Theorem

Assume:

- *Strictly feasible initial point* $x^{(0)}$.
- ∇f, ∇a, $\nabla^2 f$, $\nabla^2 a$ *are Lipschitz.*

Then our IPM takes as most

$$\mathcal{O}\left(\left(\psi_\mu(x^{(0)}) - \inf_z \psi_\mu(z)\right)\mu^{-7/4}\right)$$

trust region steps to terminate with a μ*-approximate second-order Fritz John point* (x^+, y^+).

Nonconvex result

Theorem

Assume:

- *Strictly feasible initial point* $x^{(0)}$.
- ∇f, ∇a, $\nabla^2 f$, $\nabla^2 a$ *are Lipschitz.*

Then our IPM takes as most

$$\mathcal{O}\left(\left(\psi_\mu(x^{(0)}) - \inf_z \psi_\mu(z)\right) \mu^{-7/4}\right)$$

trust region steps to terminate with a μ*-approximate second-order Fritz John point* (x^+, y^+).

algorithm	# iteration	primitive	evaluates
Birgin et al., 2016	$\mathcal{O}(\mu^{-3})$	vector operation	∇
Cartis et al., 2011	$\mathcal{O}(\mu^{-2})$	linear program	∇
Birgin et al., 2016	$\mathcal{O}(\mu^{-2})$	KKT of NCQP	∇, ∇^2
IPM (this paper)	$\mathcal{O}(\mu^{-7/4})$	**linear system**	∇, ∇^2

Convex result

Modify our IPM have sequence of decreasing μ, i.e., $\mu^{(j)}$ with $\mu^{(j)} \to 0$.

Theorem

Assume:

- *Strictly feasible initial point $x^{(0)}$ and feasible region is bounded.*
- *f, ∇f, ∇a, $\nabla^2 f$, $\nabla^2 a$ are Lipschitz.*
- *Slaters condition holds.*

Then our IPM starting takes at most

$$\tilde{\mathcal{O}}\left(m^{1/3}\epsilon^{-2/3}\right)$$

trust region steps to terminate with a ϵ-optimal solution.

Questions?

One-phase results