Optimization for Machine Learning

Lecture 21: Interior Point Methods – Intro

6.881: MIT

Suvrit Sra
Massachusetts Institute of Technology

11 May, 2021

Background and Motivation

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable
- **Newton method**: $x_{k+1} \leftarrow x_k - [f''(x_k)]^{-1} f'(x_k)$

Background and Motivation

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable
- **Newton method**: $x_{k+1} \leftarrow x_k - [f''(x_k)]^{-1} f'(x_k)$

Explore. Complexity analysis of Newton's method.

Background and Motivation

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable
- **Newton method**: $x_{k+1} \leftarrow x_k - [f''(x_k)]^{-1} f'(x_k)$

Explore. Complexity analysis of Newton's method.

- Newton's method for a ***constrained*** convex problem:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad Ax = b. \end{aligned}$$

Background and Motivation

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable
- **Newton method**: $x_{k+1} \leftarrow x_k - [f''(x_k)]^{-1} f'(x_k)$

Explore. Complexity analysis of Newton's method.

- Newton's method for a ***constrained*** convex problem:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad Ax = b. \end{aligned}$$

- Interior Point Methods build on the Newton method to tackle above convex optimization problem

Exercise: How'd you solve above prob using Newton?

Preliminaries

(handling constraints)

Barrier functions for constraints

$$\min \{ f(x) \mid x \in \mathcal{X} \}$$

Barrier functions for constraints

$$\min \{ f(x) \mid x \in \mathcal{X} \}$$

- $\mathcal{X} \subset \mathbb{R}^n$, closed, convex set with nonempty interior

Barrier functions for constraints

$$\min \{ f(x) \mid x \in \mathcal{X} \}$$

- $\mathcal{X} \subset \mathbb{R}^n$, closed, convex set with nonempty interior
- Equip $\mathcal{X}$ with an **internal penalty** or **barrier** function F

Barrier functions for constraints

$$\min \{f(x) \mid x \in \mathcal{X}\}$$

- $\mathcal{X} \subset \mathbb{R}^n$, closed, convex set with nonempty interior
- Equip $\mathcal{X}$ with an **internal penalty** or **barrier** function F
- F is smooth, strictly convex on $\text{int}(\mathcal{X})$;

Barrier functions for constraints

$$\min \{f(x) \mid x \in \mathcal{X}\}$$

- $\mathcal{X} \subset \mathbb{R}^n$, closed, convex set with nonempty interior
- Equip $\mathcal{X}$ with an **internal penalty** or **barrier** function F
- F is smooth, strictly convex on $\text{int}(\mathcal{X})$; $F(x_k) \to +\infty$ for every sequence $\{x_k\} \subset \text{int}(\mathcal{X})$ that converges to a point $\bar{x} \in \partial X$

Barrier functions for constraints

$$\min \{f(x) \mid x \in \mathcal{X}\}$$

- $\mathcal{X} \subset \mathbb{R}^n$, closed, convex set with nonempty interior
- Equip $\mathcal{X}$ with an **internal penalty** or **barrier** function F
- F is smooth, strictly convex on $\text{int}(\mathcal{X})$; $F(x_k) \to +\infty$ for every sequence $\{x_k\} \subset \text{int}(\mathcal{X})$ that converges to a point $\bar{x} \in \partial X$
- **Barrier family** of objective functions

$$F_t(x) := tf_0(x) + F(x),$$

where $t > 0$ is the **penalty** parameter.

Barrier functions for constraints

$$\min \{f(x) \mid x \in \mathcal{X}\}$$

- $\mathcal{X} \subset \mathbb{R}^n$, closed, convex set with nonempty interior
- Equip $\mathcal{X}$ with an **internal penalty** or **barrier** function F
- F is smooth, strictly convex on $\text{int}(\mathcal{X})$; $F(x_k) \to +\infty$ for every sequence $\{x_k\} \subset \text{int}(\mathcal{X})$ that converges to a point $\bar{x} \in \partial X$
- **Barrier family** of objective functions

  $$F_t(x) := tf_0(x) + F(x),$$

  where $t > 0$ is the **penalty** parameter.
- Say $\mathcal{X}$ is bounded, then every $F_t(x)$ attains its minimum in $\text{int}(\mathcal{X})$; call this $x^*(t)$ (unique since $F(x)$ is strictly convex)

Barrier functions for constraints

$$\min \{f(x) \mid x \in \mathcal{X}\}$$

- $\mathcal{X} \subset \mathbb{R}^n$, closed, convex set with nonempty interior
- Equip $\mathcal{X}$ with an **internal penalty** or **barrier** function F
- F is smooth, strictly convex on $\text{int}(\mathcal{X})$; $F(x_k) \to +\infty$ for every sequence $\{x_k\} \subset \text{int}(\mathcal{X})$ that converges to a point $\bar{x} \in \partial X$
- **Barrier family** of objective functions

$$F_t(x) := tf_0(x) + F(x),$$

  where $t > 0$ is the **penalty** parameter.
- Say $\mathcal{X}$ is bounded, then every $F_t(x)$ attains its minimum in $\text{int}(\mathcal{X})$; call this $x^*(t)$ (unique since $F(x)$ is strictly convex)
- Let **central path** be $\{x^*(t) \mid t \geq 0\}$;

Barrier functions for constraints

$$\min \{f(x) \mid x \in \mathcal{X}\}$$

- $\mathcal{X} \subset \mathbb{R}^n$, closed, convex set with nonempty interior
- Equip $\mathcal{X}$ with an **internal penalty** or **barrier** function F
- F is smooth, strictly convex on $\text{int}(\mathcal{X})$; $F(x_k) \to +\infty$ for every sequence $\{x_k\} \subset \text{int}(\mathcal{X})$ that converges to a point $\bar{x} \in \partial X$
- **Barrier family** of objective functions

$$F_t(x) := tf_0(x) + F(x),$$

  where $t > 0$ is the **penalty** parameter.
- Say $\mathcal{X}$ is bounded, then every $F_t(x)$ attains its minimum in $\text{int}(\mathcal{X})$; call this $x^*(t)$ (unique since $F(x)$ is strictly convex)
- Let **central path** be $\{x^*(t) \mid t \geq 0\}$; as $t \to \infty$, central path converges to solution of original problem.

Path-following algorithm

1. Suppose $t_k > 0$; some $x_k \in \text{int}(\mathcal{X})$ s.t. x_k "close" to $x^*(t_k)$
2. Repeat until "done":
    1. Replace penalty t_k by a larger value t_{k+1}
    2. Run some method to minimize $F_{t_{k+1}}$ "warm-starting" at x_k until a point x_{k+1} "close" to $x^*(t_{k+1})$ is found
    3. New pair (t_{k+1}, x_{k+1}) is close to the "path"

Path-following algorithm

1. Suppose $t_k > 0$; some $x_k \in \text{int}(\mathcal{X})$ s.t. x_k "close" to $x^*(t_k)$
2. Repeat until "done":
   1. Replace penalty t_k by a larger value t_{k+1}
   2. Run some method to minimize $F_{t_{k+1}}$ "warm-starting" at x_k until a point x_{k+1} "close" to $x^*(t_{k+1})$ is found
   3. New pair (t_{k+1}, x_{k+1}) is close to the "path"

Fairly old idea (e.g., Frisch, 1955)

Path-following algorithm

1. Suppose $t_k > 0$; some $x_k \in \text{int}(\mathcal{X})$ s.t. x_k "close" to $x^*(t_k)$
2. Repeat until "done":
   1. Replace penalty t_k by a larger value t_{k+1}
   2. Run some method to minimize $F_{t_{k+1}}$ "warm-starting" at x_k until a point x_{k+1} "close" to $x^*(t_{k+1})$ is found
   3. New pair (t_{k+1}, x_{k+1}) is close to the "path"

Fairly old idea (e.g., Frisch, 1955)

- Any unconstrained method to solve for x_{k+1}

Path-following algorithm

1. Suppose $t_k > 0$; some $x_k \in \text{int}(\mathcal{X})$ s.t. x_k "close" to $x^*(t_k)$
2. Repeat until "done":
   1. Replace penalty t_k by a larger value t_{k+1}
   2. Run some method to minimize $F_{t_{k+1}}$ "warm-starting" at x_k until a point x_{k+1} "close" to $x^*(t_{k+1})$ is found
   3. New pair (t_{k+1}, x_{k+1}) is close to the "path"

Fairly old idea (e.g., Frisch, 1955)

- Any unconstrained method to solve for x_{k+1}
- What is complexity of such a scheme?

Path-following algorithm

1. Suppose $t_k > 0$; some $x_k \in \text{int}(\mathcal{X})$ s.t. x_k "close" to $x^*(t_k)$
2. Repeat until "done":
   1. Replace penalty t_k by a larger value t_{k+1}
   2. Run some method to minimize $F_{t_{k+1}}$ "warm-starting" at x_k until a point x_{k+1} "close" to $x^*(t_{k+1})$ is found
   3. New pair (t_{k+1}, x_{k+1}) is close to the "path"

Fairly old idea (e.g., Frisch, 1955)

- Any unconstrained method to solve for x_{k+1}
- What is complexity of such a scheme?
- Numerical problems when t_k becomes large; breakdown?
- Standard theory of unconstrained minimization predicts slowdown as t_k becomes larger ...

What barriers?

♠ Renegar (1988) and Gonzaga (1989) introduced improved path-following methods for linear programming

What barriers?

- Renegar (1988) and Gonzaga (1989) introduced improved path-following methods for linear programming
- In particular, for linear-programming with feasible set

$$\mathcal{X} = \left\{ x \mid a_i^T x \leq b_i, 1 \leq i \leq m \right\},$$

What barriers?

- Renegar (1988) and Gonzaga (1989) introduced improved path-following methods for linear programming
- In particular, for linear-programming with feasible set

$$\mathcal{X} = \left\{ x \mid a_i^T x \leq b_i, 1 \leq i \leq m \right\},$$

they used the **logarithmic barrier**

$$F(x) := -\sum_i \log(b_i - a_i^T x).$$

What barriers?

- Renegar (1988) and Gonzaga (1989) introduced improved path-following methods for linear programming
- In particular, for linear-programming with feasible set

$$\mathcal{X} = \left\{x \mid a_i^T x \leq b_i, 1 \leq i \leq m\right\},$$

they used the **logarithmic barrier**

$$F(x) := -\sum\nolimits_i \log(b_i - a_i^T x).$$

- And with this $F(x)$, they showed path-following based on Newton's method can be made **polynomial time**.

Breakthrough result, though *ad-hoc* analysis of NM

What barriers?

- Renegar (1988) and Gonzaga (1989) introduced improved path-following methods for linear programming
- In particular, for linear-programming with feasible set

$$\mathcal{X} = \left\{ x \mid a_i^T x \leq b_i, 1 \leq i \leq m \right\},$$

they used the **logarithmic barrier**

$$F(x) := -\sum\nolimits_i \log(b_i - a_i^T x).$$

- And with this $F(x)$, they showed path-following based on Newton's method can be made **polynomial time**.

Breakthrough result, though *ad-hoc* analysis of NM

Shortly thereafter, Nesterov realized what intrinsic properties of the log-barrier made it work!

Newton method – affine invariance

- Consider $f(x)$ and $\phi(y) = f(Ay)$, where A is invertible

Newton method – affine invariance

- Consider $f(x)$ and $\phi(y) = f(Ay)$, where A is invertible

Lemma Let $\{x_k\}$ be generated by Newton method for f:

$$x_{k+1} = x_k - [f''(x_k)]^{-1} f'(x_k) \quad k \geq 0.$$

Let $\{y_k\}$ be seq. generated by NM for ϕ:

$$y_{k+1} = y_k - [\phi''(y_k)]^{-1} \phi'(y_k),$$

with $Ay_0 = x_0$. Then, $Ay_k = x_k$ for all $k \geq 0$.

Newton method – affine invariance

- Consider $f(x)$ and $\phi(y) = f(Ay)$, where A is invertible

Lemma Let $\{x_k\}$ be generated by Newton method for f:

$$x_{k+1} = x_k - [f''(x_k)]^{-1} f'(x_k) \quad k \geq 0.$$

Let $\{y_k\}$ be seq. generated by NM for ϕ:

$$y_{k+1} = y_k - [\phi''(y_k)]^{-1} \phi'(y_k),$$

with $Ay_0 = x_0$. Then, $Ay_k = x_k$ for all $k \geq 0$.

Newton method remains same—strong contrast to gradient method! **Stopping condition:**

$$\langle [f''(x_k)]^{-1} f'(x_k), f'(x_k) \rangle < \epsilon$$

independent of choice of basis A!

Newton method – local convergence

Assumptions

- **Lipschitz Hessian**: $\|\nabla^2 f(x) - \nabla^2 f(y)\| \le M\|x - y\|$
- **Local strong convexity**: There exists a local minimum x^* with

$$\nabla^2 f(x^*) \succeq \mu I, \qquad \mu > 0.$$

- **Locality**: Starting point x_0 "close enough" to x^*

Theorem. Suppose x_0 satisfies

$$\|x_0 - x^*\| < r := \frac{2\mu}{3M}.$$

Then, $\|x_k - x^*\| < r, \forall k$ and the NM converges quadratically

$$\|x_{k+1} - x^*\| \le \frac{M\|x_k - x^*\|^2}{2(\mu - M\|x_k - x^*\|)}$$

Newton method – local convergence

What's wrong / missing?

Newton method – local convergence

What's wrong / missing?

- Convergence analysis depends on μ, and M
- These quantities are **not basis independent!**

Newton method – local convergence

What's wrong / missing?

- Convergence analysis depends on μ, and M
- These quantities are **not basis independent!**
- Mismatch between geometry and analysis

What's missing

☞ Key condition used was $\|f''(x) - f''(y)\| \leq M\|x - y\|$

What's missing

- ☞ Key condition used was $\|f''(x) - f''(y)\| \le M\|x - y\|$
- ☞ Third derivative in direction $u \in \mathbb{R}^n$ is

$$f'''(x)[u] = \lim_{\alpha \to 0} \frac{f''(x + \alpha u) - f''(x)}{\alpha}$$

What's missing

- ☞ Key condition used was $\|f''(x) - f''(y)\| \leq M\|x - y\|$
- ☞ Third derivative in direction $u \in \mathbb{R}^n$ is

$$f'''(x)[u] = \lim_{\alpha \to 0} \frac{f''(x + \alpha u) - f''(x)}{\alpha}$$

- ☞ Lipschitz Hessian equivalent (prove!) to

$$\|f'''(x)[u]\| \leq M\|u\|$$

What's missing

- ☞ Key condition used was $\|f''(x) - f''(y)\| \leq M\|x - y\|$
- ☞ Third derivative in direction $u \in \mathbb{R}^n$ is

$$f'''(x)[u] = \lim_{\alpha \to 0} \frac{f''(x + \alpha u) - f''(x)}{\alpha}$$

- ☞ Lipschitz Hessian equivalent (prove!) to

$$\|f'''(x)[u]\| \leq M\|u\|$$

- ☞ Thus, at $x \in \operatorname{dom} f$, and any $u, v \in \mathbb{R}^n$ we have

$$\langle f'''(x)[u]v, v\rangle \leq M\|u\| \|v\|^2$$

What's missing

☞ Using $x \leftarrow Ay$, $u' \leftarrow Au$, $v' \leftarrow Av$, $\phi(y) = f(Ay)$

$$\langle f'''(x)[u]v,\, v\rangle = \langle \phi'''(x)[u']v',\, v'\rangle$$

What's missing

☞ Using $x \leftarrow Ay, u' \leftarrow Au, v' \leftarrow Av, \phi(y) = f(Ay)$

$$\langle f'''(x)[u]v, v\rangle = \langle \phi'''(x)[u']v', v'\rangle$$

☞ Thus, in the inequality $\langle f'''(x)[u]v, v\rangle \leq M\|u\|\,\|v\|^2$, lhs is affine invariant, but rhs is not

What's missing

☞ Using $x \leftarrow Ay$, $u' \leftarrow Au$, $v' \leftarrow Av$, $\phi(y) = f(Ay)$

$$\langle f'''(x)[u]v,\, v\rangle = \langle \phi'''(x)[u']v',\, v'\rangle$$

☞ Thus, in the inequality $\langle f'''(x)[u]v,\, v\rangle \le M\|u\|\,\|v\|^2$, lhs is affine invariant, but rhs is not

☞ What can be a quick fix? Observation, use local norms on rhs

$$\|u\|_{f''(x)} := \langle f''(x)u,\, u\rangle^{1/2} = \sqrt{u^T f''(x) u}$$

What's missing

☞ Using $x \leftarrow Ay$, $u' \leftarrow Au$, $v' \leftarrow Av$, $\phi(y) = f(Ay)$

$$\langle f'''(x)[u]v,\, v\rangle = \langle \phi'''(x)[u']v',\, v'\rangle$$

☞ Thus, in the inequality $\langle f'''(x)[u]v,\, v\rangle \leq M\|u\|\,\|v\|^2$, lhs is affine invariant, but rhs is not

☞ What can be a quick fix? Observation, use local norms on rhs

$$\|u\|_{f''(x)} := \langle f''(x)u,\, u\rangle^{1/2} = \sqrt{u^T f''(x) u}$$

Then, we immediately have

$$\|A^{-1}u\|_{f''(Ax)} = \|u\|_{f''(x)}$$

What's missing

☞ Using $x \leftarrow Ay$, $u' \leftarrow Au$, $v' \leftarrow Av$, $\phi(y) = f(Ay)$

$$\langle f'''(x)[u]v,\, v\rangle = \langle \phi'''(x)[u']v',\, v'\rangle$$

☞ Thus, in the inequality $\langle f'''(x)[u]v,\, v\rangle \leq M\|u\|\,\|v\|^2$, lhs is affine invariant, but rhs is not

☞ What can be a quick fix? Observation, use local norms on rhs

$$\|u\|_{f''(x)} := \langle f''(x)u,\, u\rangle^{1/2} = \sqrt{u^T f''(x) u}$$

Then, we immediately have

$$\|A^{-1}u\|_{f''(Ax)} = \|u\|_{f''(x)}$$

☞ This brings us to the idea of **self-concordance**

Self-concordant functions

- Let $f \in C^3(\operatorname{dom} f)$ be a closed, convex with **open** domain

Self-concordant functions

- Let $f \in C^3(\mathrm{dom} f)$ be a closed, convex with **open** domain
- Fix $x \in \mathrm{dom} f$ and a direction vector $u \in \mathbb{R}^n$

Self-concordant functions

- Let $f \in C^3(\mathrm{dom} f)$ be a closed, convex with **open** domain
- Fix $x \in \mathrm{dom} f$ and a direction vector $u \in \mathbb{R}^n$
- Denote restriction to line $\phi(x;t) := f(x+tu)$

Derivatives

$$
\begin{aligned}
Df(x)[u] &= \phi'(x;t) = \langle f'(x),\, u\rangle \\
D^2f(x)[u,u] &= \phi''(x;t) = \langle f''(x)u,\, u\rangle = \|u\|^2_{f''(x)} \\
D^3f(x)[u,u,u] &= \phi'''(x;t) = \langle D^3f(x)[u]u,\, u\rangle
\end{aligned}
$$

Self-concordant functions

- Let $f \in C^3(\operatorname{dom} f)$ be a closed, convex with **open** domain
- Fix $x \in \operatorname{dom} f$ and a direction vector $u \in \mathbb{R}^n$
- Denote restriction to line $\phi(x;t) := f(x+tu)$

Derivatives

$$Df(x)[u] = \phi'(x;t) = \langle f'(x), u\rangle$$
$$D^2 f(x)[u,u] = \phi''(x;t) = \langle f''(x)u, u\rangle = \|u\|^2_{f''(x)}$$
$$D^3 f(x)[u,u,u] = \phi'''(x;t) = \langle D^3 f(x)[u]u, u\rangle$$

Note: Third derivative: symmetric trilinear operator, so it operates on $[u_1, u_2, u_3]$ to yield a trilinear symmetric form.

$$D^p f(x)[u_1, \ldots, u_p] = \frac{\partial^p}{\partial t_1 \cdots \partial t_p}\bigg|_{t_1 = \cdots = t_p = 0} f(x + t_1 u_1 + \cdots + t_p u_p)$$

Self-concordant functions and barriers

Def. (**Self-concordant**). Let $\mathcal{X}$ be a closed convex set. A function $f : \text{int}(\mathcal{X}) \to \mathbb{R}$ called **self-concordant** (SC) on $\mathcal{X}$ if

☞ $f \in C^3(\mathcal{X})$ with $f(x_k) \to +\infty$ if $x_k \to \bar{x} \in \partial\mathcal{X}$

☞ f satisfies the **SC inequality**

$$|D^3 f(x)[u,u,u]| \leq 2\left(D^2 f(x)[u,u]\right)^{3/2}, \quad \forall x \in \text{int}(\mathcal{X}), u \in \mathbb{R}^n$$

Self-concordant functions and barriers

Def. (**Self-concordant**). Let $\mathcal{X}$ be a closed convex set. A function $f : \mathrm{int}(\mathcal{X}) \to \mathbb{R}$ called **self-concordant** (SC) on $\mathcal{X}$ if

- ☞ $f \in C^3(\mathcal{X})$ with $f(x_k) \to +\infty$ if $x_k \to \bar{x} \in \partial\mathcal{X}$
- ☞ f satisfies the **SC inequality**

$$|D^3 f(x)[u,u,u]| \leq 2\left(D^2 f(x)[u,u]\right)^{3/2}, \quad \forall x \in \mathrm{int}(\mathcal{X}), u \in \mathbb{R}^n$$

Def. Given a real $\vartheta \geq 1$, F is called a **ϑ-self-concordant barrier** (SCB) for $\mathcal{X}$ if F is SC and

$$|DF(x)[u]| \leq \vartheta^{1/2}\left(D^2 f(x)[u,u]\right)^{1/2}, \quad \forall x \in \mathrm{int}(\mathcal{X}), u \in \mathbb{R}^n.$$

Self-concordant functions and barriers

Def. (**Self-concordant**). Let $\mathcal{X}$ be a closed convex set. A function $f : \text{int}(\mathcal{X}) \to \mathbb{R}$ called **self-concordant** (SC) on $\mathcal{X}$ if

- ☞ $f \in C^3(\mathcal{X})$ with $f(x_k) \to +\infty$ if $x_k \to \bar{x} \in \partial\mathcal{X}$
- ☞ f satisfies the **SC inequality**

$$|D^3 f(x)[u,u,u]| \leq 2\left(D^2 f(x)[u,u]\right)^{3/2}, \quad \forall x \in \text{int}(\mathcal{X}), u \in \mathbb{R}^n$$

Def. Given a real $\vartheta \geq 1$, F is called a **ϑ-self-concordant barrier** (SCB) for $\mathcal{X}$ if F is SC and

$$|DF(x)[u]| \leq \vartheta^{1/2}\left(D^2 f(x)[u,u]\right)^{1/2}, \quad \forall x \in \text{int}(\mathcal{X}), u \in \mathbb{R}^n.$$

- Exponents 3/2 and 1/2 crucial—ensure both sides have same degree of homogeneity in u (for affine invariance)
- Factor 2 can be scaled by scaling f; equiv. to $D^2 f$ Lipschitz with constant 2 in norm $\|\cdot\|_{f''(x)}$

Examples of SC functions

Example. $f(x) = -\log x : \mathbb{R}_{++} \to \mathbb{R}$ is a 1-SCB for $\mathbb{R}_+$

Proof: $f''(x) = x^{-2}, f'''(x) = -2x^{-3}$; directly verifies.

Examples of SC functions

Example. $f(x) = -\log x : \mathbb{R}_{++} \to \mathbb{R}$ is a 1-SCB for $\mathbb{R}_+$

Proof: $f''(x) = x^{-2}, f'''(x) = -2x^{-3}$; directly verifies.

- Linear functions are SC; $f'''(x) = 0$
- Convex quadratic functions; $f'''(x) = 0$
- Log-barrier for $\phi(x) = a + \langle b, x \rangle - \frac{1}{2}x^T Ax$; $f(x) = -\log \phi(x)$
  Show: $|D^3 f(x)[u, u, u]| = |2\omega_1^3 + 3\omega_1\omega_2|$, where $\omega_1 = Df(x)[u]$, $\omega_2 = \frac{1}{\phi(x)} u^T Au$; also show that $D^2 f(x)[u, u] = \omega_1^2 + \omega_2$.

Examples of SC functions

Example. $f(x) = -\log x : \mathbb{R}_{++} \to \mathbb{R}$ is a 1-SCB for $\mathbb{R}_+$

Proof: $f''(x) = x^{-2}, f'''(x) = -2x^{-3}$; directly verifies.

- Linear functions are SC; $f'''(x) = 0$
- Convex quadratic functions; $f'''(x) = 0$
- Log-barrier for $\phi(x) = a + \langle b, x\rangle - \frac{1}{2}x^T Ax$; $f(x) = -\log\phi(x)$
  Show: $|D^3 f(x)[u,u,u]| = |2\omega_1^3 + 3\omega_1\omega_2|$, where $\omega_1 = Df(x)[u]$, $\omega_2 = \frac{1}{\phi(x)}u^T Au$; also show that $D^2 f(x)[u,u] = \omega_1^2 + \omega_2$.

Lemma Function f SC iff for any $x \in \text{int}(\mathcal{X})$, $u_1, u_2, u_3 \in \mathbb{R}^n$

$$|D^3 f(x)[u_1, u_2, u_3]| \le 2\|u_1\|_{f''(x)}\|u_2\|_{f''(x)}\|u_3\|_{f''(x)}$$

Proof: Essentially generalized Cauchy-Schwarz (challenge!).

Optimization using SC

Key quantities

- Let $f(x)$ be SC, and that $f''(x) \succ 0$ within $\operatorname{dom} f$
- *not asking* for usual L-smoothness, strong cvx
- simplified notation for the local norms at x

$$
\begin{aligned}
\|u\|_x &:= \langle f''(x)u,\, u\rangle^{1/2} \\
\|v\|_x^* &= \langle [f''(x)]^{-1}v,\, v\rangle^{1/2}
\end{aligned}
$$

Key quantities

- Let $f(x)$ be SC, and that $f''(x) \succ 0$ within $\operatorname{dom} f$
- ***not asking*** for usual L-smoothness, strong cvx
- simplified notation for the local norms at x

$$\begin{aligned} \|u\|_x &:= \langle f''(x)u,\, u\rangle^{1/2} \\ \|v\|_x^* &= \langle [f''(x)]^{-1}v,\, v\rangle^{1/2} \end{aligned}$$

- Let us use these to state three crucial observations

Three key facts (locally structure)

☞ At any point $x \in \operatorname{dom} f = \operatorname{int}(\mathcal{X})$, there is an ***ellipsoid***

$$W(x) := \{y \in \mathbb{R}^n \mid \|y - x\|_x \leq 1\} \subset \operatorname{dom} f.$$

Three key facts (locally structure)

- At any point $x \in \operatorname{dom} f = \operatorname{int}(\mathcal{X})$, there is an ***ellipsoid***

$$W(x) := \{y \in \mathbb{R}^n \mid \|y - x\|_x \leq 1\} \subset \operatorname{dom} f.$$

- Within this Dinkin ellipsoid, f is almost quadratic

$$r := \|u\|_x < 1 \implies$$

$$(1-r)^2 f''(x) \preceq f''(x+u) \preceq \frac{1}{(1-r)^2} f''(x)$$

Three key facts (locally structure)

☞ At any point $x \in \mathrm{dom} f = \mathrm{int}(\mathcal{X})$, there is an ***ellipsoid***

$$W(x) := \{y \in \mathbb{R}^n \mid \|y - x\|_x \leq 1\} \subset \mathrm{dom} f.$$

☞ Within this Dinkin ellipsoid, f is almost quadratic

$$r := \|u\|_x < 1 \implies$$

$$(1-r)^2 f''(x) \preceq f''(x+u) \preceq \frac{1}{(1-r)^2} f''(x)$$

☞ Moreover, linear upper and lower bounds on f:

$$f(x) + \langle f'(x), u\rangle + \rho(-r) \leq f(x+u) \leq f(x) + \langle f'(x), u\rangle + \rho(r),$$

where $\rho(r) := -\log(1-r) - s = s^2/2 + s^3/3 + \cdots$

Proof: See Chap. 4 of Nesterov (2004).

Setting up Newton's Method: Newton Decrement

Newton decrement

$$\lambda_f(x) := \langle [f''(x)]^{-1} f'(x), f'(x) \rangle^{1/2}.$$

Observe: $\lambda_f(x) = \|f'(x)\|_x^*$ (local, dual-norm of gradient).

$$\lambda_f(x) = \max_u \left\{ Df(x)[u] \mid D^2 f(x)[u,u] \le 1 \right\}$$

- $\lambda_f(x)$ vanishes at (unique, if any) minimizer x_f^* of f on $\operatorname{dom} f$

Setting up Newton's Method: Newton Decrement

Newton decrement

$$\lambda_f(x) := \langle [f''(x)]^{-1} f'(x), f'(x) \rangle^{1/2}.$$

Observe: $\lambda_f(x) = \|f'(x)\|_x^*$ (local, dual-norm of gradient).

$$\lambda_f(x) = \max_u \left\{ Df(x)[u] \mid D^2 f(x)[u, u] \leq 1 \right\}$$

- $\lambda_f(x)$ vanishes at (unique, if any) minimizer x_f^* of f on $\operatorname{dom} f$

Theorem. If $\lambda_f(x) < 1$ for some $x \in \operatorname{dom} f$. Then, $\min f(x)$ s.t., $x \in \operatorname{dom} f$, has a unique optimal solution.

Newton Method: Guaranteed Descent

1. Select $x_0 \in \operatorname{dom} f$
2. For $k \geq 0$: $x_{k+1} = x_k - \frac{1}{1+\lambda_f(x_k)}[f''(x_k)]^{-1}f'(x_k)$

Newton Method: Guaranteed Descent

1. Select $x_0 \in \operatorname{dom} f$
2. For $k \geq 0$: $x_{k+1} = x_k - \frac{1}{1+\lambda_f(x_k)}[f''(x_k)]^{-1} f'(x_k)$

Theorem. The iterates of the damped NM satisfy ($k \geq 0$):

$$f(x_{k+1}) \leq f(x_k) - \rho(-\lambda_f(x_k)).$$

Newton Method: Guaranteed Descent

1. Select $x_0 \in \operatorname{dom} f$
2. For $k \geq 0$: $x_{k+1} = x_k - \frac{1}{1+\lambda_f(x_k)}[f''(x_k)]^{-1} f'(x_k)$

Theorem. The iterates of the damped NM satisfy ($k \geq 0$):

$$f(x_{k+1}) \leq f(x_k) - \rho(-\lambda_f(x_k)).$$

Proof: Denote $\lambda = \lambda_f(x_k)$. Also, set $\omega(t) := \rho(-t)$.
Then, $\|x_{k+1} - x_k\|_x = \frac{\lambda}{1+\lambda} = \omega'(\lambda)$. Thus, using one of the key facts

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k \rangle + \omega^*(\|x_{k+1} - x_k\|_x) \\ &= f(x_k) - \frac{\lambda^2}{1+\lambda} + \omega^*(\omega'(\lambda)) \\ &= f(x_k) - \lambda\omega'(\lambda) + \omega^*(\omega'(\lambda)) = f(x_k) - \omega(\lambda). \end{aligned}$$

▶ At each step, $f(x)$ decreases by at least $\omega(\lambda)$

Using Damped Newton

- Globally convegent; iteration complexity can be derived.
- Though, better to start with DN (when $\lambda_f(x_k) \geq \beta$, $\beta \in (0, 0.3819\ldots)$), where $f(x_{k+1}) \leq f(x_k) - \omega(\beta)$, which runs for

$$N \approx \frac{1}{\omega(\beta)[f(x_0) - f(x_f^*)]} \quad \text{iterations.}$$

- After that $\lambda_f(x_k) \leq \beta$, and we apply standard NM which converges quadratically.

SC Barriers

Minimization using SC Barriers

- class of ϑ-SCB smaller than general SC.

Minimization using SC Barriers

- class of ϑ-SCB smaller than general SC.

Standard convex problem

$$\min \quad c^T x \quad x \in \mathcal{X},$$

where $\mathcal{X}$ is a compact set for which $\operatorname{dom} F \equiv \mathcal{X}$.

Minimization using SC Barriers

- class of ϑ-SCB smaller than general SC.

Standard convex problem

$$\min \quad c^T x \quad x \in \mathcal{X},$$

where $\mathcal{X}$ is a compact set for which $\operatorname{dom} F \equiv \mathcal{X}$.

- Recall path-following scheme

$$x^*(t) = \operatorname*{argmin}_{x \in \operatorname{dom} F} \quad tc^T x + F(x), \quad t \geq 0.$$

- Any point of the central path (set $\{x^*(t)\}$) satisfies

$$tc + F'(x^*(t)) = 0.$$

Minimization using SC Barriers

- class of ϑ-SCB smaller than general SC.

Standard convex problem

$$\min \quad c^T x \quad x \in \mathcal{X},$$

where $\mathcal{X}$ is a compact set for which $\operatorname{dom} F \equiv \mathcal{X}$.

- Recall path-following scheme

$$x^*(t) = \operatorname*{argmin}_{x \in \operatorname{dom} F} \quad tc^T x + F(x), \quad t \geq 0.$$

- Any point of the central path (set $\{x^*(t)\}$) satisfies

$$tc + F'(x^*(t)) = 0.$$

- Aim is to iteratively find points close to central path

Minimization using SCBs

Approximate solution: A point x for which

$$\lambda_{F_t}(x) := \|F_t'(x)\|_x^* = \|tc + F'(x)\|_x^* \leq \beta,$$

where β is the "centering parameter"

Minimization using SCBs

Approximate solution: A point x for which

$$\lambda_{F_t}(x) := \|F_t'(x)\|_x^* = \|tc + F'(x)\|_x^* \leq \beta,$$

where β is the "centering parameter"

Theorem. For any $t > 0$, we have

$$c^T x^*(t) - c^T x^* \leq \tfrac{\vartheta}{t}.$$

If a point x is an approximate solution (close to $x^*(t)$), then

$$c^T x - c^T x^* \leq \frac{1}{t}\left(\vartheta + \frac{\beta(\beta + \sqrt{\vartheta})}{1-\beta}\right).$$

Path-following algorithm

1. Set $t_0 = 0$. Choose accuracy $\epsilon > 0$ and $x_0 \in \operatorname{dom} F$ such that

$$\|F'(x_0)\|^*_{x_0} \leq \beta$$

Path-following algorithm

1. Set $t_0 = 0$. Choose accuracy $\epsilon > 0$ and $x_0 \in \operatorname{dom} F$ such that

$$\|F'(x_0)\|_{x_0}^* \leq \beta$$

2. At k-th iteration, set

$$t_{k+1} = t_k + \frac{\gamma}{\|c\|_{x_k}^*}, \quad \gamma = \frac{\sqrt{\beta}}{1-\sqrt{\beta}} - \beta,$$

$$x_{k+1} = x_k - [F''(x_k)]^{-1}(t_{k+1}c + F'(x_k))$$

Path-following algorithm

1. Set $t_0 = 0$. Choose accuracy $\epsilon > 0$ and $x_0 \in \operatorname{dom} F$ such that
$$\|F'(x_0)\|^*_{x_0} \leq \beta$$
2. At k-th iteration, set
$$t_{k+1} = t_k + \frac{\gamma}{\|c\|^*_{x_k}}, \quad \gamma = \frac{\sqrt{\beta}}{1-\sqrt{\beta}} - \beta,$$
$$x_{k+1} = x_k - [F''(x_k)]^{-1}(t_{k+1}c + F'(x_k))$$
3. Stop the process if
$$\epsilon t_k \geq \vartheta + \frac{\beta(\beta+\sqrt{\vartheta})}{1-\beta}$$

Path-following algorithm

1. Set $t_0 = 0$. Choose accuracy $\epsilon > 0$ and $x_0 \in \operatorname{dom} F$ such that

$$\|F'(x_0)\|^*_{x_0} \leq \beta$$

2. At k-th iteration, set

$$t_{k+1} = t_k + \frac{\gamma}{\|c\|^*_{x_k}}, \quad \gamma = \frac{\sqrt{\beta}}{1-\sqrt{\beta}} - \beta,$$

$$x_{k+1} = x_k - [F''(x_k)]^{-1}(t_{k+1}c + F'(x_k))$$

3. Stop the process if

$$\epsilon t_k \geq \vartheta + \frac{\beta(\beta + \sqrt{\vartheta})}{1-\beta}$$

Theorem. Above scheme yields $c^T x_N - c^T x^* \leq \epsilon$ after no more than N steps, where

$$N \leq O\left(\sqrt{\vartheta} \log \frac{\vartheta \|c\|^*_{x^*}}{\epsilon}\right).$$

SC Barriers

Recall, F is ϑ-SCB if $F''(x) \succeq \frac{1}{\vartheta} F'(x) F'(x)^T$.

Exercise: Verify that $f(x) = \langle a, x \rangle + b$ with $\mathrm{dom}\, f = \mathbb{R}^n$ is not an SCB. Similarly, convex quadratics are also not SCBs.

Exercise: Let $\phi(x) = b + \langle a, x \rangle - \frac{1}{2} x^T A x$ be concave. Verify that $F(x) = -\log \phi(x)$ with $\mathrm{dom}\, F = \{x \in \mathbb{R}^n \mid \phi(x) > 0\}$ is 1-SCB.

Exercise: $-\log\det X$ barrier for PSD cone

Theorem. If F_1, F_2 are ϑ_i-SCB, then $F = F_1 + F_2$ is ϑ-SCB for $\mathrm{dom}\, F = \mathrm{dom}\, F_1 \cap \mathrm{dom}\, F_2$ with $\vartheta = \vartheta_1 + \vartheta_2$.

Impt: The param ϑ is invariant to affine transformations.

Can apply IPM only if we have SCBs for constraints / epigraphs of costs

Much More

We've just scratched the surface!

Much More

We've just scratched the surface!

- Much more to interior point methods.
- See references and also read: Ch. 9,10,11 of BV for high-level overview.

Read: Universal barriers, entropic barriers, SCB and differential geometry, SCB and optimal transport, …

Much More

We've just scratched the surface!

- Much more to interior point methods.
- See references and also read: Ch. 9,10,11 of BV for high-level overview.

Read: Universal barriers, entropic barriers, SCB and differential geometry, SCB and optimal transport, …

References

♡ Nemirovski, Todd. *Interior-point methods for optimization.* (2008)

♡ Nesterov. *Introductory lectures on convex optimization* (2004).

♡ Nesterov, Nemirovski. *Interior-Point Polynomial Algorithms in Convex Programming* (1994).