Lecture 14

Last time
- Back to IPM
- A "complete" IPM
- Answering our questions

Today
- Finish proof
- New topic
- Proximal point method

Proof of proposition (♡):

We prove 3 inequalities:

(1) $(c^Tx - c^Tx^*)(1 - \|x - x^*_\eta\|_x) \leq c^Tx^*_\eta - c^Tx^*$

(2) $\|x - x^*_\eta\|_x \leq 4\|n_\eta(x)\|_x$

(3) $c^Tx^*_\eta - c^Tx \leq \frac{m}{\eta}$.

Clearly the combination of these three establishes the proposition

Proof of (1): We start by adding and sustracting

$$c^Tx - c^Tx^* = (c^Tx - c^Tx^*_\eta) + (c^Tx^*_\eta - c^Tx^*) \quad (zz)$$

We focus on bounding the first term:

$$c^T x - c^T x^*_\eta = c^T (x - x^*_\eta)$$
$$\le \langle H(x)^{-1} c, x - x^*_\eta \rangle_x$$
Cauchy Schwarz
$$\le \| H(x)^{-1} c \|_x \, \| x - x^*_\eta \|_x$$

This is already close to what we want. Let's bound $\| H(x)^{-1} c \|_x$.

Claim (Exercise) $x - H(x)^{-1} c / \| H(x)^{-1} c \|_x \in P$.

Therefore

$$c^T x^* \le c^T x - c^T H(x)^{-1} c / \| H(x)^{-1} c \|_x$$
$$\Updownarrow$$
$$\| H(x)^{-1} c \|_x = \frac{c^T H(x)^{-1} c}{\| H(x)^{-1} c \|} \le c^T x - c x^*$$

Combining this with (22) gives (1).

Proof of (2): Suppose that we take h s.t. $\| h \|_x \le 1/6$. Expanding $f(x+h)$ and using

The mean value theorem gives

$$f_\eta(x+h) = f_\eta(x) + \langle h, \nabla f_\eta(x)\rangle + \frac{1}{2} h^T \nabla^2 f_\eta(\theta)\, h$$

for some $\theta \in (x, x+h)$. By Cauchy Schwarz

$$\begin{aligned}\langle h, \nabla f_\eta(x)\rangle &= \langle h, n_\eta(x)\rangle_x \\ &\geq -\|h\|_x \, \|n_\eta(x)\|_x.\end{aligned}$$

Further, since f is SC

$$\begin{aligned}h^T \nabla^2 f_\eta(\theta)\, h &= h^T H(\theta)\, h \\ &\geq \frac{1}{2} h^T H(x)\, h \\ &= \frac{1}{2} \|h\|_x^2.\end{aligned}$$

SC with $\delta = \frac{1}{6}$

Altogether, for all h s.t. $\|h\|_x \leq \frac{1}{6}$,

$$f_\eta(x+h) \geq f_\eta(x) - \|h\|_x \|n_\eta(x)\|_x + \frac{1}{4}\|h\|_x^2.$$

Consider now the funky sphere $S_r = \{ y \mid \|y\|_x = r \}$ with $r = 4\|n_\eta(x)\|_x$. Note that for

$y \in S_r$ we have

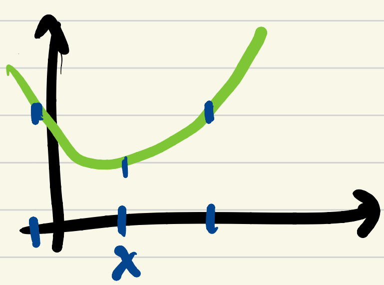
$$\|y\|_x = 4\|n_\eta(x)\|_x \le \frac{4}{24} = \frac{1}{6}$$

and so (08) applies:

$$f_\eta(x+y) \ge f_\eta(x) - \|y\|_x \|n_\eta(x)\|_x + \frac{1}{4}\|y\|_x^2$$
$$\ge f_\eta(x).$$

Claim (Exercise): Since the f_η is strictly convex and $f_\eta(x+y) \ge f(x)$ $\forall y \in S_r$, then $\|x - x^*_\eta\|_x \le r$.

Therefore, (08) applies:



$$0 \ge \frac{1}{\|x - x^*_\eta\|_x}\left(f_\eta(x^*_\eta) - f_\eta(x)\right)$$
$$\ge \|n_\eta(x)\|_x + \frac{1}{4}\|x - x^*_\eta\|_x,$$

concluding the proof of (2).

Proof of (3): By the optimality of x^*_η

$$0 = \nabla f_\eta(x^*_\eta) = \eta c + g(x^*_\eta).$$

Thus, $-\eta c = g(x^*_\eta)$ and

$$\langle c, x^*_\eta - x^* \rangle = -\langle c, x^* - x^*_\eta \rangle$$

$$= \frac{1}{\eta} \langle g(x^*_\eta), x^* - x^*_\eta \rangle$$

It remains to show $\langle g(x^*_\eta), x^* - x^*_\eta \rangle \leq m$, using our formula for g:

$$\langle g(x^*_\eta), x^* - x^*_\eta \rangle = \sum_{i=1}^{m} \frac{\langle a_i, x^* - x^*_\eta \rangle}{s_i(x^*_\eta)}$$

$$= \sum_{i=1}^{m} \frac{s_i(x^*_\eta) - s_i(x^*)}{s_i(x^*_\eta)}$$

$s_i(x^*_\eta) > 0$

$$= m - \sum_{i=1}^{m} \frac{s_i(x^*)}{s_i(x^*_\eta)}$$

$s_i(x^*) \geq 0$

$$\leq m.$$

This completes the proof of the proposition.

Closing remarks about IPM

We only cover a special type of IPM known as a "Primal" IPM. Commercial solvers use more sophisticated "primal-dual" versions. They often also use heuristics, e.g., Mehrotra predictor-corrector method, that makes them run much faster.

IPM were a subject of intense research in the '80s and '90s.

If you are interested I recommend Steve Wright's "Primal-dual Interior Point methods" book.

New topic

So far we have covered two

methods that have been classically used to solve LPs.

IPM can also be used to solve SDPs

Importantly, both of these methods rely on a very expensive operation: matrix inversion.

Recall Simplex had to invert $A_B \in \mathbb{R}^{m \times m}$ at each iteration, while IPM inverts $\nabla^2 f_\eta(x_k) \in \mathbb{R}^{n \times n}$ at each iteration of Newton's.

Warning: Exactly inverting these matrices require $O(m^3)$ and $O(n^3)$ memory, respectively! $\dashv$

Next, we drive into several methods that turn out to require less memory per iteration, albeit at the cost of regeiring more of them.

Proximal point method and fixed points

Instead of developing problem/algorithmic specific theory, we will build a general framework based on the fixed point iteration.

As a first example suppose we wanted to solve

$$\min_{x \in E} f(x) \qquad f: E \to \mathbb{R} \cup \{+\infty\}$$

with f proper, closed and convex.

A simple (and often not even implementable) method to solve this problem

Proximal point method

Input: $x_0 \in E$

▷ Loop $k \geq 0$:

$$x_{k+1} \leftarrow \underbrace{\operatorname{argmin} \; f(x) + \frac{1}{2\alpha} \|x - x_k\|_2^2}_{\mathrm{prox}_{\alpha f}(x_k).}$$

(I am really glad we are back to two-line algorithms)

A simple application of sub-differential calculus yields

$$x = \mathrm{prox}_{\alpha f}(z) \iff (z - x) \in \partial f(z).$$

This will prove useful later on, but now how can we know that the prox is well defined?

Theorem (:): Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex, closed function and $z \in E$ arbitrary. Then, we have that

$$\inf_{x \in E} \left\{ f(x) + \frac{1}{2}\|x+z\|^2 \right\} + \inf_{y \in E} \left\{ f^*(y) + \frac{1}{2}\|y - z\|^2 \right\} = \|z\|^2$$

Moreover, the optimal solutions of both problems are attained by unique x^*, y^*.

Additionally, they are characterized as the solution of the feasibility problem

$$z = x + y$$
$$y \in \partial f(x).$$

Proof: Exercise. $\square$

Remark. This generalizes the Pythagorean theorem. Simply take $f = \iota_L$ with $L \subseteq E$ a subspace, then $f^* = \iota_{L^\perp}$ and

$$x^* = P_L z \quad \text{and} \quad y^* = P_{L^\perp} z.$$

$\dashv$