Lecture 11

Last time

- ▷ Review of smooth optimization
- ▷ Motivating Problems
- ▷ Proximal operator

Today

- ▷ Forward - Backward method
- ▷ Examples
- ▷ Constraints via proximal operator
- ▷ Analysis

Forward - Backward Method.

When we have a sum $\underset{\text{smooth}}{f} + \underset{\text{convex}}{h}$. we have

a natural approximation

$$\Psi_k(x) = \underbrace{f(x^k) + \langle \nabla f(x^k), x - x^k \rangle}_{\text{linear approximation}} + \underbrace{h(x)}_{\text{perfect approx.}}$$

Then, at each iteration we update

$$x_{k+1} \leftarrow \underset{x}{\operatorname{argmin}} \Big\{ \underbrace{h(x)}_{\text{convex}} + \underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2\alpha_k} \|x - x_k\|^2}_{\text{smooth part}} \Big\}$$

By Lemma ★

$$\frac{1}{\alpha_k}\left(x_k - \alpha_k \nabla f(x_k) - x_{k+1}\right) \in \partial h(x)$$

By Proposition ♡, this is equivalent to

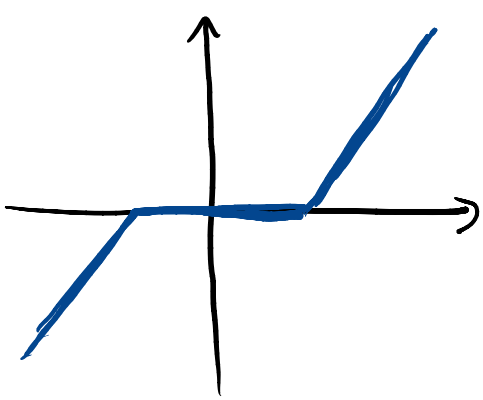
$$x_{k+1} = \underbrace{\text{prox}_{\alpha_k h}}_{\text{Backward step}} \big(\underbrace{x_k - \alpha_k f(x_k)}_{\text{Forward step}} \big).$$

Thus, this method works well for convex functions for which we can compute proximal operators efficiently.

Examples

- The ℓ_1 norm $\|\cdot\|_1$ (HW 3)

$$[\text{prox}_{\alpha \|\cdot\|_1}]_i = \begin{cases} x_i + \alpha & x_i < -\alpha \\ 0 & -\alpha \le x_i \le \alpha \\ x_i - \alpha & x_i > \alpha. \end{cases}$$



← Known as hard thresholding.

- The nuclear norm $\|\cdot\|_*$.

U diag $(\sigma(X))$ V

SVD decomposition of

$$\text{prox}_{\alpha\|\cdot\|_*}(X) = U \,\text{diag}\big(\text{prox}_{\alpha\|\cdot\|_1}(\sigma(X))\big) V^T$$

Constraints via the proximal operator

Suppose we want to minimize

$$\min_{\substack{x\in S \\ \uparrow \\ \text{convex closed}}} \underset{\nwarrow \text{ smooth.}}{f(x)}$$

We can capture these problems using the extended reals

$$\min f(x) + \imath_S(x), \qquad \imath_S(x) = \begin{cases} 0 & x\in S, \\ \infty & x\notin S. \end{cases}$$

which matches the template we are considering (smooth + convex).

Lemma: $\mathrm{prox}_{\alpha \imath_S}(x) = \mathrm{proj}_S(x)$.

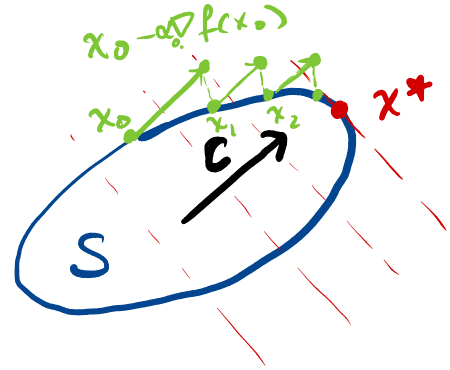
Proof:
$$\begin{aligned}\mathrm{prox}_{\alpha \imath_S}(x) &= \mathrm{argmin}\left\{ \imath_S(y) + \frac{1}{2\alpha}\|y-x\|^2 \right\} \\ &= \underset{x\in S}{\mathrm{argmin}}\left\{ \|y-x\|^2 \right\} \\ &= \mathrm{proj}_S(x). \qquad \square\end{aligned}$$

Then the Forward-Backward method reduces to Projected Gradient Descent

$$x_{k+1} \leftarrow \mathrm{proj}_S(x_k - \alpha_k \nabla f(x_k)).$$

Intuition

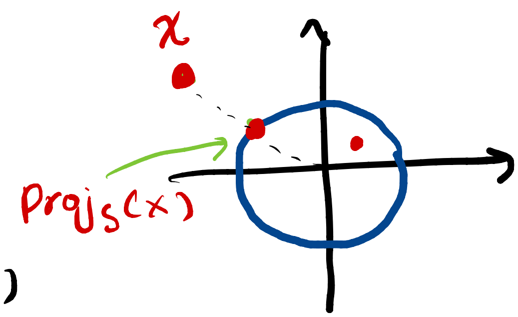
$$\min_{x \in S} -c^T x$$



Examples

ℓ_2 - norm ball

$$S = \{ x \mid \|x\|_2 \leq 1 \}$$

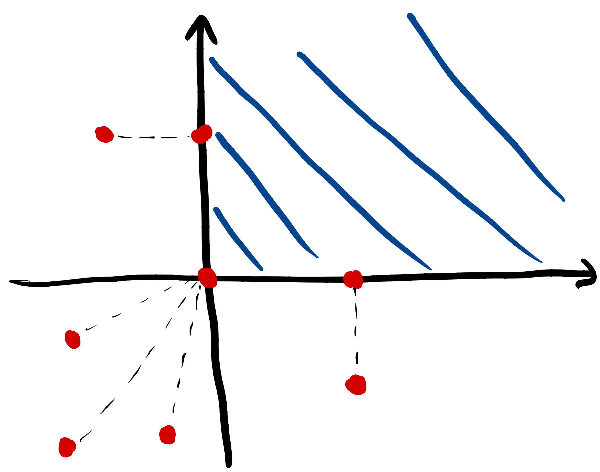


$$\text{proj}_S(x) = \begin{cases} x & x \in S, \\ \dfrac{x}{\|x\|_2} & \text{otherwise.} \end{cases}$$

Nonnegative orthant

$$S = \{ x \mid x_i \geq 0 \;\; \forall i \}$$

$$\text{proj}_S(x)_i = \max\{x_i, 0\}$$



Grading Polyhedral

$$S = \{ (H, M, F) \mid Ax \leq b \}$$

see syllabus.

$$\text{proj}_S(x) = \underset{\text{s.t. } y \in S}{\operatorname{argmin}} \; \|y - x\|^2$$

Quadratic programming problem.

Analysis of FBM

We define the Gradient mapping

$$G_\alpha(x) = \frac{1}{\alpha}\left(x - \underbrace{\text{prox}_{\alpha h}(x - \nabla f(x))}_{= x^+}\right)$$

By definition

$$\frac{1}{\alpha}(x - \alpha \nabla f(x) - x_+) \in \partial h(x_+)$$

Then

$$G_{\alpha_k}(x) \in \nabla f(x) + \partial h(x_+)$$

Thus, when $G_\alpha(x) = 0 \Rightarrow x = x_+$ and

$$-\nabla f(x) \in \partial h(x)$$ ← First order optimality condition.

Thus we use $\|G_\alpha(x)\|$ as a measure of optimality.

Lemma (Descent 2.0): Assume f is L-smooth. Then, for all $x \in \mathbb{R}^d$,

$$(f+h)(x^+) \le (f+h)(x) - \left(\alpha - \frac{L\alpha^2}{2}\right) \|G_\alpha(x)\|^2.$$

Proof: By Taylor Approximation

(☺) $$f(x^+) \le f(x) + \nabla f(x)^T (x^+ - x) + \frac{L}{2}\|x - x^+\|^2.$$

Moreover $\frac{1}{\alpha}(x - \alpha \nabla f(x) - x^+) \in \partial h(x^+)$

(♡) $$\Rightarrow \quad h(x) \ge h(x^+) + \frac{1}{\alpha}(x - \alpha \nabla f(x) - x^+)^T (x - x^+)$$

$$= h(x^+) - \nabla f(x)^T (x - x^+) + \frac{1}{\alpha}\|x - x^+\|^2.$$

Then, taking (☺) + (♡)

$$(f+h)(x^+) \le f(x) + h(x) - \left(\frac{1}{\alpha} - \frac{L}{2}\right)\|x - x^+\|^2$$

$$= f(x) + h(x) - \left(\alpha - \frac{\alpha^2 L}{2}\right)\|G_\alpha(x)\|^2.$$

□

Thus, picking $\alpha = \frac{1}{L}$ gives

$$(f+h)(x^+) \le (f+h)(x) - \frac{1}{2L}\|G_{1/L}(x)\|^2.$$

Linesearch procedures work exactly the same as before. If you want the details see Chapter 10 of Amir Beck's "First-Order Methods in Optimization."

Theorem: For any f with L-Lipschitz gradient and convex h. The iterates of FBM with stepsize $\alpha_k = 1/L$ satisfy

$$\frac{1}{T}\sum_{k=0}^{T-1} \| G_{1/L}(x_k) \|^2 \leq \frac{2L\,((f+h)(x_0) - \min f+h)}{T}.$$

$\dashv$

Intuition

There is an iterate that is approximate stationary

$$\min_{k \leq T-1} \| G_{1/L}(x_k) \| = O\left(\frac{1}{\sqrt{T}}\right).$$

Proof: By DL 2.0

$$\| G(x_k) \|^2 \leq 2L\,((f+h)(x_k) - (f+h)(x_{k+1}))$$

Summing up to $T-1$ yields

$$\sum_{k=0}^{T-1} \| G(x_k) \|^2 \leq 2L\,((f+h)(x_0) - (f+h)(x_T))$$
$$\leq 2L\,((f+h)(x_0) - \min f+h),$$

divide by T to get the result $\square$

Theorem For any convex, L-smooth f and convex h such that $x^* \in \operatorname{argmin}(f+h)(x)$.
Then, the iterates of FBM with $\alpha_k = 1/L$ satisfies

$$(f+h)(x_{k+1}) - \min(f+h) \leq \frac{L\| x_0 - x^* \|^2}{2K}.$$

Proof: We start by proving

$$(\ast\ast) \quad 0 \leq (f+h)(x_{k+1}) - \min(f+h) \leq \frac{L}{2}\left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2\right)$$

By definition x_{k+1} minimizes

$$\varphi_k(x) = \underbrace{f(x_k) + \nabla f(x_k)^T(x - x_k) + h(x)}_{\text{0-strongly convex}} + \underbrace{\frac{L}{2}\|x - x_k\|^2}_{\text{L-strongly convex.}}$$

By Hw 2 P2:

(1) $$\varphi_k(x_{k+1}) + \frac{L}{2}\| x^* - x_{k+1}\|^2 \leq \varphi_k(x^*)$$

Using the characterization of L-smooth convex functions

(2) $$(f + h)(x_{k+1}) \leq \varphi_k(x_{k+1})$$

Using the convexity f

(3) $$\varphi_k(x^*) \leq \underbrace{f(x^*) + h(x^*)}_{\min(f+h)} + \frac{L}{2}\| x^* - x_k\|^2$$

Then

$$\begin{aligned}(f+h)(x_{k+1}) - \min(f+h) &\overset{(2)}{\leq} \varphi_k(x_{k+1}) - \min(f+h) \\ &\overset{(1)}{\leq} \varphi_k(x^*) - \frac{L}{2}\|x^* - x_{k+1}\|^2 \\ &\qquad - \min(f+h) \\ &\overset{(3)}{\leq} \frac{L}{2}\left(\|x^* - x_k\|^2 - \|x^* - x_{k+1}\|^2\right),\end{aligned}$$

which establishes (**).

Summing up and dividing by T gives

$$\frac{1}{T}\sum_{k=0}^{T-1}\left[(f+h)(x_{k+1}) - \min(f+h)\right] \leq \frac{L}{2T}\left(\|x_0 - x^*\|^2 - \|x_T - x^*\|^2\right)$$

$$\leq \frac{L}{2T} \|x_0 - x^*\|^2.$$

OL 2.0 ensures that the minimum function gap is achieved at $K = T-1$

$$\Rightarrow \quad f+h(x_T) - \min(f+h) \leq \frac{L\|x_0 - x^*\|^2}{2T}. \qquad \square$$