Lecture 12

Last time

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

Today

- Finish analysis
- Guarantees for strongly convex
- Accelerated Forward Backward Method.
- More proximal methods
- Alternating Projections

Theorem

For any convex, $L$-smooth $f$ and convex $h$ such that $x^* \in \text{argmin} (f+h)(x)$.

Then, the iterates of $\text{FBM}$ with $\alpha_k = \frac{1}{k}$ satisfies

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

Proof: We start by proving

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

By definition $x_{k+1}$ minimizes

\[
\varphi_k(x) = f(x_k) + \nabla f(x_k)^T (x - x_k) + h(x) + \frac{1}{2} \|x - x_k\|^2
\]

By $\text{HW2} \ p2$:

\[
\Phi_k(x_{k+1}) + \frac{1}{2} \|x^* - x_{k+1}\|^2 \leq \Phi_k(x^*)
\]
Using the characterization of L-smooth convex functions

\[(2) \quad (f + h)(x_{k+1}) \leq \Psi_k(x_{k+1})\]

Using the convexity of

\[(3) \quad \Psi_k(x^*) \leq f(x^*) + h(x^*) + \frac{L}{2} \|x^* - x_k\|^2 \]

Then

\[(f + h)(x_{k+1}) - \min (f + h) \leq \Psi_k(x_{k+1}) - \min (f + h) \]

\[(3) \quad \Psi_k(x^*) - \frac{L}{2} \|x^* - x_k\|^2 - \min (f + h) \leq \frac{L}{2} (\|x^* - x_k\|^2 - \|x^* - x_{k+1}\|^2),\]

which establishes \((\times)\).

\[\square\]

Again convergence should speed up under quadratic growth.

**Theorem** If in addition, we suppose that \(f + h\) is \(\mu\) strongly convex. Then

\[(f + h)(x_{k+1}) - \min (f + h) \leq \frac{1}{2} ((f + h)(x_0) - \min f + h)\]

for \(k \geq \lceil 2L/\mu \rceil\).
Proof: By the previous theorem
\[
(f + h)(x_{k+1}) - \min f + h \leq \frac{L}{2K} \|x_0 - x^*\|^2
\]

Thus we achieve accuracy \(\varepsilon\) after
\[
\log_2 \left( \frac{(f + h)(x_0) - \min f + h}{\varepsilon} \right)
\]
iterations.

Acceleration

We consider the algorithm that starts at \(y_0 = x_0\) and \(\lambda_0 = 0\), and updates
\[
y_{k+1} = \text{prox}_{\alpha h} (x_k - \alpha \nabla f(x_k))
\]
\[
x_{k+1} = y_{k+1} + \left( \frac{\lambda_k - 1}{\lambda_k K_{k+1}} \right) (y_{k+1} - y_k)
\]
\[
\lambda_{k+1} = 1 + \sqrt{1 + 4\lambda_k^2}
\]
This algorithm goes by different names:

- Accelerated/Fast proximal/Projected Gradient Method
- FISTA.

Just as before, it exhibits faster convergence.

**Theorem:** For any convex $f$ with $L$-Lipschitz gradient and convex $h$, the iterates of AFBM with $\alpha = \frac{1}{2}$ satisfy

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

**Proof:** Details are very similar to the proof for AGD (see Beck’s book Theorem 10.34).

More proximal methods

A natural question is what happens when we have

$$\min f(x) + g(x).$$
If \( f, g : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\} \), and none of the two is smooth. Maybe the proximal operator of both \( f \) and \( g \) is easy to compute.

**Examples**

- **Intersection of two sets**
  
  Find \( x \in C_1 \cap C_2 \) where \( C_1, C_2 \) are convex and closed.
  
  Then the two prox are projection.

- **Compressed sensing**
  
  \[
  \min \|x\|_1 \quad \text{s.t.} \quad Ax = b,
  \]

  \[
  \min \|x\|_1 \quad \text{\textcolor{red}{\textbf{\textit{Proj is easy}}}}
  \]

  There are a number of methods to tackle these problems:
  - **Alternating Projections** (Example 1)
  - **Alternating Direction Method of Multipliers (ADMM)**
  - **Primal - Dual Hybrid Gradient (PDHG)** (Example 2)

  In order to understand the ideas behind
ADMM and PDLP we need more convex analysis, so these algorithms will be covered in Nonlinear 2.

**Alternating Projections**

Assume we want to solve

\[ \min \|x - y\| \text{ st. } x \in \mathcal{C}_1, \ y \in \mathcal{C}_2. \]

The alternating projections method was originally proposed by John von Neumann. It updates as follows

\[ x_{k+1} = P_{\mathcal{C}_1} P_{\mathcal{C}_2}(x_k) \]

**Intuition**

[Diagram of sets \( \mathcal{C}, \mathcal{C}_1, \mathcal{C}_2 \) with alternating projections.]

Another perspective to analyze iterated algorithms based on proximal mappings is through the lens of a fixed-point iteration.
Def: Given an operator $F : \mathbb{R}^d \rightarrow \mathbb{R}^d$, a fixed-point iteration updates

$$x_{k+1} < F(x_k).$$

The goal of this iteration is to find a fixed point $x^* = F(x^*)$.

Proposition: The following two are equivalent

- $x^*$ is a fixed point of $P_{c_1} P_{c_2}$.
- $(x^*, P_{c_1} x^*)$ is a solution of

$$\min_{x \in c_2} \frac{1}{2} \| x - y \|^2,
\quad \min_{y \in c_1} \{ \| x - y \|^2 + \tau_{c_1}(x) + \tau_{c_2}(y) \}$$

Proof: Consider

$$f(x, y) = \frac{1}{2} \| x - y \|^2 + \tau_{c_1}(x) + \tau_{c_2}(y)$$

Then $(x^*, y^*)$ is a solution if

$$[0] \in \left[ x^* - y^* \right] + \left[ \partial \tau_{c_1}(x^*) \right]$$

Thus

$$y^* - x^* \in \partial \tau_{c_1}(x^*) \Leftrightarrow \text{proj}_{c_1}(y^*) = x^*$$
$$x^* - y^* \in \partial \tau_{c_2}(y^*) \Leftrightarrow \text{proj}_{c_2}(x^*) = y^*$$
Fact 1: $R = 2F - I$ is 1-Lipschitz.

Check!

Fact 2: For all $a, b \in \mathbb{R}^d$

$$\|\frac{1}{2}a + \frac{1}{2}b\|^2 = \frac{1}{2}\|a\|^2 + \frac{1}{2}\|b\|^2 - \frac{1}{4}\|a - b\|^2$$

Theorem: The iterates of AP satisfy

$$\frac{1}{T} \sum_{k=0}^{T-1} \| x_k - F(x_k) \|^2 \leq \frac{\| x_0 - x^* \|^2}{T}$$

Proof: Rewrite $F = \frac{R}{2} + \frac{I}{2}$, then

$$\| x_k - x^* \|^2 = \| \frac{1}{2} (x_k - x^*) + \frac{1}{2} (R(x_k) - R(x^*)) \|^2$$

\[= \frac{1}{2} \| x_k - x^* \|^2 + \frac{1}{2} \| R(x_k) - R(x^*) \|^2 - \frac{1}{4} \| x_k - x^* - \frac{R(x_k) - R(x^*)}{\| x_k - x^* \|^2} \|^2 \]

\[\leq \| x_k - x^* \|^2 - \frac{1}{4} \| x_k - R(x_k) \|^2. \]

Reordering and summing up first $T - 1$ ineq

$$\frac{1}{4T} \sum_{k=0}^{T-1} \| x_k - R(x_k) \|^2 \leq \frac{1}{T} \left( \| x_0 - x^* \|^2 - \| x_T - x^* \|^2 \right)$$

$$\| 2(x_k - F(x_k)) \|^2 \leq \frac{1}{T} \| x_0 - x^* \|^2$$
Corollary: The iterates converge to a fixed point $x^*$.  

Proof: Let $S = \{ x | F(x) = x \}$. By the previous Theorem, the $x_k$'s are bounded. Thus, there is some accumulation point $x^*$. By the previous Theorem $\| x_k - F(x_k) \| \to 0$, by continuity $x^* = F(x^*)$. Moreover by $(\ast)$ $x_k \to x^*$.  

More generally one can prove that the convergence depends on transversality.  

$\theta$ - angle between sets.  

Slow convergence due to lack of transversality.