# LECTURE 24

# LECTURE OUTLINE

- Gradient proximal minimization method
- Nonquadratic proximal algorithms
- Entropy minimization algorithm
- Exponential augmented Lagrangian mehod
- Entropic descent algorithm

**************************************

References:


- Beck, A., and Teboulle, M., 2010. “Gradient-Based Algorithms with Applications to Signal Recovery Problems, in Convex Optimization in Signal Processing and Communications (Y. Eldar and D. Palomar, eds.), Cambridge University Press, pp. 42-88.

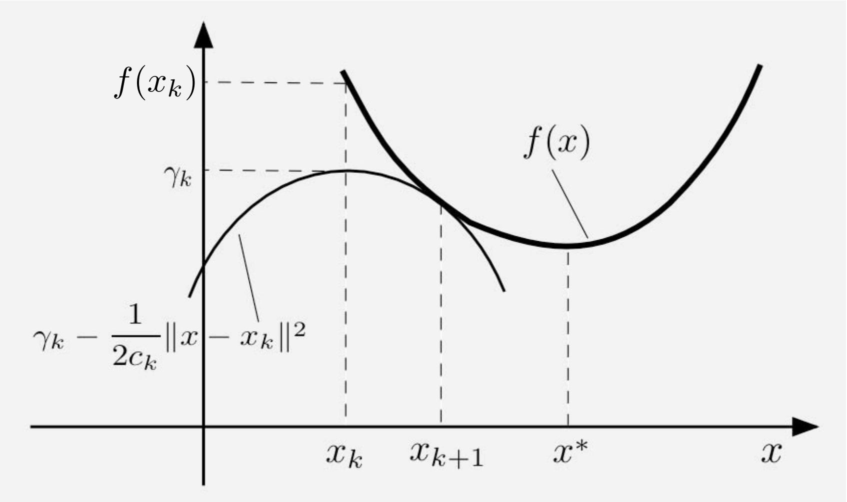
- Beck, A., and Teboulle, M., 2003. “Mirror Descent and Nonlinear Projected Subgradient Methods for Convex Optimization,” Operations Research Letters, Vol. 31, pp. 167-175.

- Bertsekas, D. P., 1999. Nonlinear Programming, Athena Scientific, Belmont, MA.

# PROXIMAL AND GRADIENT PROJECTION

- Proximal algorithm to minimize convex $f$ over closed convex $X$

$$x_{k+1} \in \arg\min_{x \in X} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$




- Let $f$ be differentiable and assume

$$\big\|\nabla f(x) - \nabla f(y)\big\| \leq L\,\|x - y\|, \qquad \forall\ x, y \in X$$

- Define the *linear approximation function at* $x$

$$\ell(y; x) = f(x) + \nabla f(x)'(y - x), \qquad y \in \Re^n$$

- Connection of proximal with gradient projection

$$y = \arg\min_{z \in X} \left\{ \ell(z; x) + \frac{1}{2\alpha} \|z - x\|^2 \right\} = P_X\big(x - \alpha \nabla f(x)\big)$$

# GRADIENT-PROXIMAL METHOD I

- Minimize $f(x)+g(x)$ over $x \in X$, where $X$: closed convex, $f$, $g$: convex, *$f$ is differentiable.*
- **Gradient-proximal method:**

$$x_{k+1} \in \arg\min_{x \in X} \left\{ \ell(x; x_k) + g(x) + \frac{1}{2\alpha} \|x - x_k\|^2 \right\}$$

- Recall key inequality: For all $x, y \in X$

$$f(y) \le \ell(y; x) + \frac{L}{2} \|y - x\|^2$$

- Cost reduction for $\alpha \le 1/L$:

$$\begin{aligned} f(x_{k+1}) + g(x_{k+1}) &\le \ell(x_{k+1}; x_k) + \frac{L}{2} \|x_{k+1} - x_k\|^2 + g(x_{k+1}) \\ &\le \ell(x_{k+1}; x_k) + g(x_{k+1}) + \frac{1}{2\alpha} \|x_{k+1} - x_k\|^2 \\ &\le \ell(x_k; x_k) + g(x_k) \\ &= f(x_k) + g(x_k) \end{aligned}$$

- This is a key insight for the convergence analysis.

# GRADIENT-PROXIMAL METHOD II

- Equivalent definition of gradient-proximal:

$$z_k = x_k - \alpha \nabla f(x_k)$$

$$x_{k+1} \in \arg\min_{x \in X} \left\{ g(x) + \frac{1}{2\alpha} \|x - z_k\|^2 \right\}$$

- Simplifies the implementation of proximal, by using gradient iteration to deal with the case of an inconvenient component $f$

- This is similar to incremental subgradient-proximal method, but the gradient-proximal method does not extend to the case where the cost consists of the sum of multiple components.

- Allows a constant stepsize (under the restriction $\alpha \leq 1/L$). This does not extend to incremental methods.

- Like all gradient and subgradient methods, convergence can be slow.

- There are special cases where the method can be fruitfully applied (see the reference by Beck and Teboulle).

# GENERALIZED PROXIMAL ALGORITHM

- Introduce a general regularization term $D_k$:

$$x_{k+1} \in \arg\min_{x \in X} \big\{ f(x) + D_k(x, x_k) \big\}$$

- **Example:** Bregman distance function

$$D_k(x, y) = \frac{1}{c_k}\Big(\phi(x) - \phi(y) - \nabla\phi(y)'(x - y)\Big),$$

where $\phi : \Re^n \mapsto (-\infty, \infty]$ is a convex function, differentiable within an open set containing $\text{dom}(f)$, and $c_k$ is a positive penalty parameter.

- All the ideas for applications and connections of the quadratic form of the proximal algorithm extend to the nonquadratic case (although the analysis may not be trivial). In particular we have:
  - A dual proximal algorithm (based on Fenchel duality)
  - Equivalence with (nonquadratic) augmented Lagrangean method
  - Combinations with polyhedral approximations (bundle-type methods)
  - Incremental subgradient-proximal methods
  - Nonlinear gradient projection algorithms

# ENTROPY MINIMIZATION ALGORITHM

- A special case involving entropy regularization:

$$x_{k+1} \in \arg\min_{x \in X} \left\{ f(x) + \frac{1}{c_k} \sum_{i=1}^{n} x^i \left( \ln\left(\frac{x^i}{x_k^i}\right) - 1 \right) \right\}$$

where $x_0$ and all subsequent $x_k$ have positive components

- We use Fenchel duality to obtain a dual form of this minimization
- Note: The logarithmic function

$$p(x) = \begin{cases} x(\ln x - 1) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ \infty & \text{if } x < 0, \end{cases}$$

and the exponential function

$$p^\star(y) = e^y$$

are a conjugate pair.

- The dual problem is

$$y_{k+1} \in \arg\min_{y \in \Re^n} \left\{ f^\star(y) + \frac{1}{c_k} \sum_{i=1}^{n} x_k^i e^{c_k y^i} \right\}$$

# EXPONENTIAL AUGMENTED LAGRANGIAN

- The dual proximal iteration is

$$x_{k+1}^i = x_k^i e^{c_k y_{k+1}^i}, \qquad i = 1, \ldots, n$$

where $y_{k+1}$ is obtained from the dual proximal:

$$y_{k+1} \in \arg \min_{y \in \Re^n} \left\{ f^\star(y) + \frac{1}{c_k} \sum_{i=1}^{n} x_k^i e^{c_k y^i} \right\}$$

- A special case for the convex problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g_1(x) \le 0, \ldots, g_r(x) \le 0, \quad x \in X \end{aligned}$$

is the **exponential augmented Lagrangean method**

- Consists of unconstrained minimizations

$$x_k \in \arg \min_{x \in X} \left\{ f(x) + \frac{1}{c_k} \sum_{j=1}^{r} \mu_k^j e^{c_k g_j(x)} \right\},$$

followed by the multiplier iterations

$$\mu_{k+1}^j = \mu_k^j e^{c_k^j g_j(x_k)}, \qquad j = 1, \ldots, r$$

# NONLINEAR PROJECTION ALGORITHM

- Subgradient projection with general regularization term $D_k$:

$$x_{k+1} \in \arg\min_{x\in X} \Big\{ f(x_k) + \tilde\nabla f(x_k)'(x - x_k) + D_k(x, x_k) \Big\}$$

where $\tilde\nabla f(x_k)$ is a subgradient of $f$ at $x_k$. Also called **mirror descent** method.

- Linearization of $f$ simplifies the minimization
- The use of nonquadratic linearization is useful in problems with special structure
- **Entropic descent method:** Minimize $f(x)$ over the unit simplex $X = \big\{ x \geq 0 \mid \sum_{i=1}^n x^i = 1 \big\}$.
- Method:

$$x_{k+1} \in \arg\min_{x\in X} \sum_{i=1}^n x^i \left( g_k^i + \frac{1}{\alpha_k} \ln\left(\frac{x^i}{x_k^i}\right)\right)$$

where $g_k^i$ are the components of $\tilde\nabla f(x_k)$.

- This minimization can be done in closed form:

$$x_{k+1}^i = \frac{x_k^i e^{-\alpha_k g_k^i}}{\sum_{j=1}^n x_k^j e^{-\alpha_k g_k^j}}, \qquad i = 1, \ldots, n$$

6.253 Convex Analysis and Optimization
Spring 2012