LECTURE 20

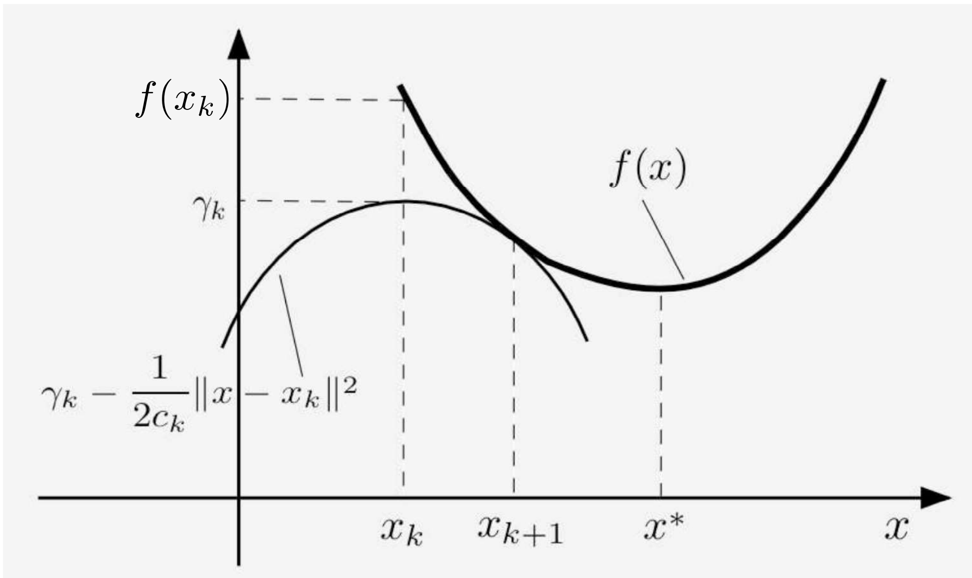
LECTURE OUTLINE

- Review of proximal and augmented Lagrangians
- Alternating direction methods of multipliers (ADMM)
- Applications of ADMM
- Extensions of proximal algorithm

************ References *************

- Bertsekas, D. P., and Tsitsiklis, J. N., 1989. Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, Englewood Cliffs, N. J.
- Eckstein, J., and Bertsekas, D. P., 1992. "On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators," Math. Progr., Vol. 55, pp. 293-318.
- Eckstein, J., 2012. "Augmented Lagrangian and Alternating Direction Methods for Convex Optimization," Rutgers, Univ. Report
- Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Now Publishers Inc.

RECALL PROXIMAL ALGORITHM



- Minimizes closed convex proper f:

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

where x_0 is an arbitrary starting point, and $\{c_k\}$ is a positive parameter sequence.

- We have $f(x_k) \to f^*$. Also $x_k \to$ some minimizer of f, provided one exists.

- When applied with $f = -q$, where q is the dual function of a constrained optimization problem, we obtain the augmented Lagrangian method.

AUGMENTED LAGRANGIAN METHOD

- Consider the convex constrained problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \;\; x \in X, \quad Ax = b \end{aligned}$$

- Primal and dual functions:

$$p(u) = \inf_{\substack{x \in X \\ Ax-b=u}} f(x), \qquad q(\lambda) = \inf_{x \in X} \big\{ f(x) + \lambda'(Ax-b) \big\}$$

- Augmented Lagrangian function:

$$L_c(x,\lambda) = f(x) + \lambda'(Ax-b) + \frac{c}{2}\|Ax-b\|^2$$

- Augmented Lagrangian algorithm: Find

$$x_{k+1} \in \arg\min_{x \in X} L_{c_k}(x, \lambda_k)$$

and then set

$$\lambda_{k+1} = \lambda_k + c_k(Ax_{k+1} - b)$$

A DIFFICULTY WITH AUGM. LAGRANGIANS

- Consider the (Fenchel format) problem

$$\begin{aligned} &\text{minimize} \quad f_1(x) + f_2(z) \\ &\text{subject to} \;\; x \in \Re^n,\ z \in \Re^m, \quad Ax = z, \end{aligned}$$

and its augmented Lagrangian function

$$L_c(x, z, \lambda) = f_1(x) + f_2(z) + \lambda'(Ax - z) + \frac{c}{2}\|Ax - z\|^2.$$

- The problem is separable in x and z, but $\|Ax - z\|^2$ couples x and x inconveniently.
- We may consider minimization by a block coordinate descent method:
  - Minimize $L_c(x, z, \lambda)$ over x, with z and λ held fixed.
  - Minimize $L_c(x, z, \lambda)$ over z, with x and λ held fixed.
  - Repeat many times, then update the multipliers, then repeat again.
- The ADMM does one minimization in x, then one minimization in z, before updating λ.

ADMM

- Start with some λ_0 and $c > 0$:

$$x_{k+1} \in \arg \min_{x \in \Re^n} L_c(x, z_k, \lambda_k),$$

$$z_{k+1} \in \arg \min_{z \in \Re^m} L_c(x_{k+1}, z, \lambda_k),$$

$$\lambda_{k+1} = \lambda_k + c(Ax_{k+1} - z_{k+1}).$$

- The penalty parameter c is kept constant in the ADMM (no compelling reason to change it).

- Strong convergence properties: $\{\lambda_k\}$ converges to optimal dual solution, and if $A'A$ is invertible, $\{x_k, z_k\}$ also converge to optimal primal solution.

- Big advantages:
  - x and z are decoupled in the minimization of $L_c(x, z, \lambda)$.
  - Very convenient for problems with special structures.
  - Has gained a lot of popularity for signal processing and machine learning problems.

- Not necessarily faster than augmented Lagrangian methods (many more iterations in λ are needed).

FAVORABLY STRUCTURED PROBLEMS I

- Additive cost problems:

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(x) \\ &\text{subject to} \ \ x \in \cap_{i=1}^{m} X_i, \end{aligned}$$

where $f_i : \Re^n \mapsto \Re$ are convex functions and X_i are closed, convex sets.

- Feasibility problem: Given m closed convex sets $X_1, X_2, \ldots, X_m$ in $\Re^n$, find a point in $\cap_{i=1}^{m} X_i$.

- Problems involving ℓ_1 norms: A key fact is that proximal works well with ℓ_1. For any $\alpha > 0$ and $w = (w^1, \ldots, w^m) \in \Re^m$,

$$S(\alpha, w) \in \arg \min_{z \in \Re^m} \left\{ \|z\|_1 + \frac{1}{2\alpha} \|z - w\|^2 \right\},$$

is easily computed by the shrinkage operation:

$$S^i(\alpha, w) = \begin{cases} w^i - \alpha & \text{if } w^i > \alpha, \\ 0 & \text{if } |w^i| \le \alpha, \\ w^i + \alpha & \text{if } w^i < -\alpha, \end{cases} \qquad i = 1, \ldots, m.$$

FAVORABLY STRUCTURED PROBLEMS II

- Basis pursuit:

$$\text{minimize} \quad \|x\|_1$$
$$\text{subject to} \quad Cx = b,$$

where $\|\cdot\|_1$ is the ℓ_1 norm in $\Re^n$, C is a given $m \times n$ matrix and b is a vector in $\Re^m$. Use $f_1 =$ indicator fn of $\{x \mid Cx = b\}$, and $f_2(z) = \|z\|_1$.

- ℓ_1 Regularization:

$$\text{minimize} \quad f(x) + \gamma\|x\|_1$$
$$\text{subject to} \quad x \in \Re^n,$$

where $f : \Re^n \mapsto (-\infty, \infty]$ is a closed proper convex function and γ is a positive scalar. Use $f_1 = f$, and $f_2(z) = \gamma\|z\|_1$.

- Least Absolute Deviations Problem:

$$\text{minimize} \quad \|Cx - b\|_1$$
$$\text{subject to} \quad x \in \Re^n,$$

where C is an $m \times n$ matrix, and $b \in \Re^m$ is a given vector. Use $f_1 = 0$, and $f_2(z) = \|z\|_1$.

SEPARABLE PROBLEMS I

- Consider a convex separable problem of the form

$$\text{minimize} \quad \sum_{i=1}^{m} f_i(x^i)$$

$$\text{subject to} \quad \sum_{i=1}^{m} A_i x^i = b, \qquad x^i \in X_i, \quad i = 1, \ldots, m,$$

- A plausible idea is the ADMM-like iteration

$$x_{k+1}^i \in \arg\min_{x^i \in X_i} L_c(x_{k+1}^1, \ldots, x_{k+1}^{i-1}, x^i, x_k^{i+1}, \ldots, x_k^m, \lambda_k),$$

$$\lambda_{k+1} = \lambda_k + c\left(\sum_{i=1}^{m} A_i x_{k+1}^i - b\right)$$

- For $m = 1$ it becomes the augmented Lagrangian method, for $m = 2$ it becomes the ADMM, and for $m > 2$ it maintains the attractive variable decoupling property of ADMM

- Unfortunately, it may not work for $m > 2$ (it does work but under restrictive assumptions)

- We will derive a similar but reliable version (a special case of ADMM for $m = 2$, from Bertsekas and Tsitsiklis 1989, Section 3.4).

SEPARABLE PROBLEMS II

- We reformulate the convex separable problem so it can be addressed by ADMM

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(x^i) \\ &\text{subject to} \quad A_i x^i = z^i, \quad x^i \in X_i, \qquad i = 1, \ldots, m, \\ &\qquad\qquad\quad \sum_{i=1}^{m} z^i = b, \end{aligned}$$

- The ADMM is given by

$$x_{k+1}^i \in \arg \min_{x^i \in X_i} \left\{ f_i(x^i) + (A_i x^i - z_k^i)' p_k^i + \frac{c}{2} \|A_i x^i - z_k^i\|^2 \right\},$$

$$z_{k+1} \in \arg \min_{\sum_{i=1}^{m} z^i = b} \left\{ \sum_{i=1}^{m} (A_i x_{k+1}^i - z^i)' p_k^i + \frac{c}{2} \|A_i x_{k+1}^i - z^i\|$$

$$p_{k+1}^i = p_k^i + c(A_i x_{k+1}^i - z_{k+1}^i),$$

where p_k^i is the multiplier of $A_i x^i = z^i$.

- A key fact is that all p_k^i, $i = 1, \ldots, m$, can be shown to be equal to a single vector λ_k, the multiplier of the constraint $\sum_{i=1}^{m} z^i = b$.

- This simplifies the algorithm.

PROXIMAL AS FIXED POINT ALGORITHM I

- Back to the proximal algorithm for minimizing closed convex $f : \Re^n \mapsto (-\infty, \infty]$.
- Proximal operator corresponding to c and f:

$$P_{c,f}(z) = \arg \min_{x \in \Re^n} \left\{ f(x) + \frac{1}{2c} \|x - z\|^2 \right\}, \qquad z \in \Re^n$$

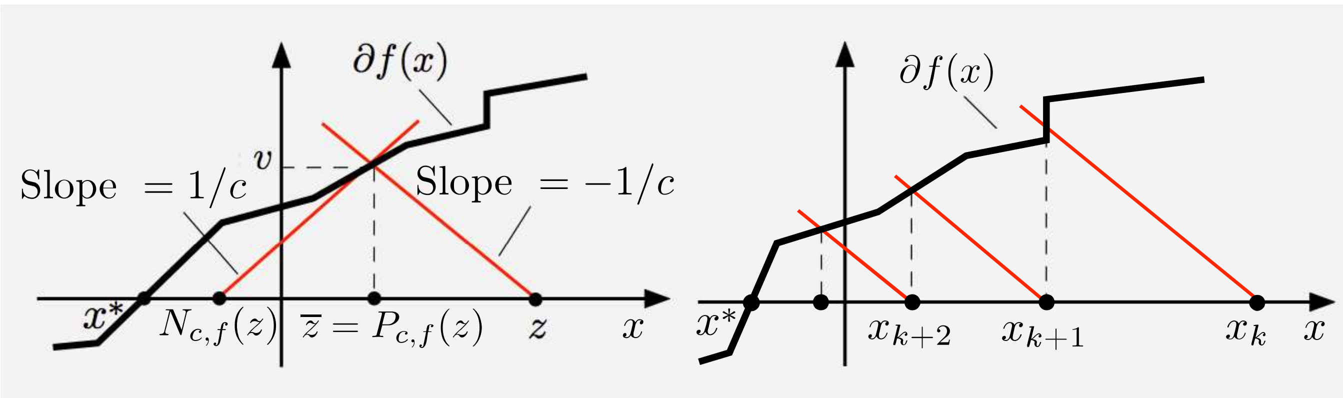
- The set of fixed points of $P_{c,f}$ coincides with the set of minima of f, and the proximal algorithm, written as

$$x_{k+1} = P_{c_k,f}(x_k),$$

may be viewed as a fixed point iteration.

- Decomposition:

$$\overline{z} = P_{c,f}(z) \qquad \text{iff} \qquad \overline{z} = z - cv \text{ for some } v \in \partial f(\overline{z})$$



- Important mapping $N_{c,f}(z) = 2P_{c,f}(z) - z$

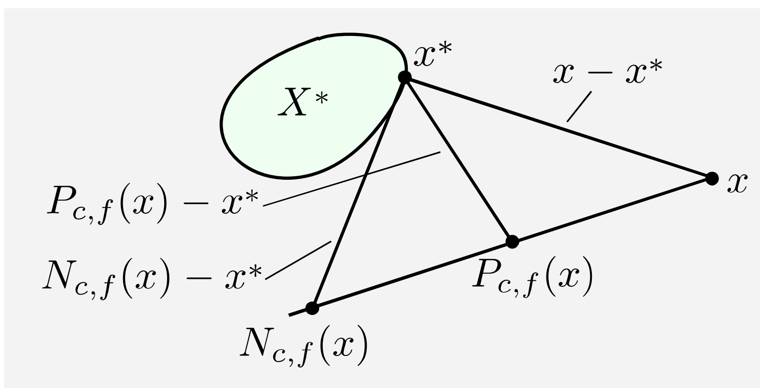
PROXIMAL AS FIXED POINT ALGORITHM II

- The mapping $N_{c,f} : \Re^n \mapsto \Re^n$ given by

$$N_{c,f}(z) = 2P_{c,f}(z) - z, \qquad z \in \Re^n,$$

is nonexpansive:

$$\left\| N_{c,f}(z_1) - N_{c,f}(z_2) \right\| \leq \|z_1 - z_2\|, \quad \forall\, z_1, z_2 \in \Re^n.$$



- The interpolated iteration

$$x_{k+1} = (1 - \alpha_k)x_k + \alpha_k N_{c,f}(x_k),$$

where $\alpha_k \in [\epsilon, 1-\epsilon]$ for some scalar $\epsilon > 0$, converges to a fixed point of $N_{c,f}$, provided $N_{c,f}$ has at least one fixed point.

- Extrapolation is more favorable

- ADMM and proximal belong to the same family of fixed point algorithms for finding a zero of a multivalued monotone operator (see refs).