LECTURE 20

LECTURE OUTLINE

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

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


RECALL PROXIMAL ALGORITHM

- Minimizes closed convex proper $f$:

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

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

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in X, \quad Ax = b
\end{align*}
\]

- Primal and dual functions:

\[
\begin{align*}
p(u) &= \inf_{x \in X, Ax - b = u} f(x), \quad q(\lambda) = \inf_{x \in X} \{ f(x) + \lambda'(Ax - b) \}
\end{align*}
\]

- 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

\[
\begin{align*}
\text{minimize} \quad & f_1(x) + f_2(z) \\
\text{subject to} \quad & x \in \mathbb{R}^n, \quad z \in \mathbb{R}^m, \quad Ax = z,
\end{align*}
\]

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 \(\lambda\) held fixed.
  
  – Minimize \(L_c(x, z, \lambda)\) over \(z\), with \(x\) and \(\lambda\) held fixed.
  
  – Repeat many times, then update the multipliers, then repeat again.

• The ADMM does \textbf{one} minimization in \(x\), then \textbf{one} minimization in \(z\), before updating \(\lambda\).
ADMM

- Start with some $\lambda_0$ and $c > 0$:

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

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

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

- 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 $\lambda$ are needed).
FAVORABLY STRUCTURED PROBLEMS I

• Additive cost problems:

\[
\begin{align*}
\text{minimize} \quad & \sum_{i=1}^{m} f_i(x) \\
\text{subject to} \quad & x \in \bigcap_{i=1}^{m} X_i,
\end{align*}
\]

where \( f_i : \mathbb{R}^n \mapsto \mathbb{R} \) 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 \( \mathbb{R}^n \), find a point in \( \cap_{i=1}^{m} X_i \).

• Problems involving \( \ell_1 \) norms: A key fact is that proximal works well with \( \ell_1 \). For any \( \alpha > 0 \) and \( w = (w^1, \ldots, w^m) \in \mathbb{R}^m \),

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

is easily computed by the shrinkage operation:

\[
S^i(\alpha, w) = \begin{cases} 
  w^i - \alpha & \text{if } w^i > \alpha, \\
  0 & \text{if } |w^i| \leq \alpha, \\
  w^i + \alpha & \text{if } w^i < -\alpha,
\end{cases} \quad i = 1, \ldots, m.
\]
FAVORABLY STRUCTURED PROBLEMS II

• Basis pursuit:

\[
\begin{align*}
\text{minimize} \quad & \|x\|_1 \\
\text{subject to} \quad & Cx = b,
\end{align*}
\]

where \(\| \cdot \|_1\) is the \(\ell_1\) norm in \(\mathbb{R}^n\), \(C\) is a given \(m \times n\) matrix and \(b\) is a vector in \(\mathbb{R}^m\). Use \(f_1 = \text{indicator fn of } \{x \mid Cx = b\}\), and \(f_2(z) = \|z\|_1\).

• \(\ell_1\) Regularization:

\[
\begin{align*}
\text{minimize} \quad & f(x) + \gamma \|x\|_1 \\
\text{subject to} \quad & x \in \mathbb{R}^n,
\end{align*}
\]

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

• Least Absolute Deviations Problem:

\[
\begin{align*}
\text{minimize} \quad & \|Cx - b\|_1 \\
\text{subject to} \quad & x \in \mathbb{R}^n,
\end{align*}
\]

where \(C\) is an \(m \times n\) matrix, and \(b \in \mathbb{R}^m\) is a given vector. Use \(f_1 = 0\), and \(f_2(z) = \|z\|_1\).
PROXIMAL AS FIXED POINT ALGORITHM

- Back to the proximal algorithm for minimizing closed convex $f : \mathbb{R}^n \mapsto (-\infty, \infty]$.
- Proximal operator corresponding to $c$ and $f$:
  \[
P_{c,f}(z) = \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2c} \|x - z\|^2 \right\}, \quad z \in \mathbb{R}^n
  \]
- 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:
  \[
  \bar{z} = P_{c,f}(z) \quad \text{iff} \quad \bar{z} = z - cv \quad \text{for some} \quad v \in \partial f(\bar{z})
  \]
- Important mapping $N_{c,f}(z) = 2P_{c,f}(z) - z$
The mapping $N_{c,f} : \mathbb{R}^n \mapsto \mathbb{R}^n$ given by

$$N_{c,f}(z) = 2P_{c,f}(z) - z, \quad z \in \mathbb{R}^n,$$

is nonexpansive:

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

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).