Augmented Lagrangian & the Method of Multipliers

Lecturer: Aarti Singh
Co-instructor: Pradeep Ravikumar

Convex Optimization 10-725/36-725
Constrained optimization

So far:

- Projected gradient descent
- Conditional gradient method
- Barrier and Interior Point methods
- Augmented Lagrangian/Method of Multipliers (today)

- Consider the equality constrained problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in X, \quad h(x) = 0,
\end{align*}
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \) and \( h : \mathbb{R}^n \to \mathbb{R}^m \) are continuous, and \( X \) is closed.
Quadratic Penalty Approach

Add a quadratic penalty instead of a barrier. For some $c > 0$

$$\begin{align*}
\text{minimize} & \quad f(x) + \frac{c}{2} \| h(x) \|^2 \\
\text{subject to} & \quad h(x) = 0,
\end{align*}$$

\textbf{Note:} Problem is unchanged – has same local minima

\textbf{Augmented Lagrangian:}

$$L_c(x, \lambda) = f(x) + \lambda^\top h(x) + \frac{c}{2} \| h(x) \|^2$$

- Quadratic penalty makes new objective strongly convex if $c$ is large
- Softer penalty than barrier – iterates no longer confined to be interior points.
Solve unconstrained minimization of Augmented Lagrangian:

\[ x = \arg \min_{x \in X} L_c(x, \lambda) \]

where

\[ L_c(x, \lambda) = f(x) + \lambda^\top h(x) + \frac{c}{2} \| h(x) \|^2 \]

When does this work?
1) Take $\lambda$ close to $\lambda^*$. Let $x^*$, $\lambda^*$ satisfy the sufficiency conditions of second-order for the original problem. We will show that if $c$ is larger than a threshold, then $x^*$ is a strict local minimum of the Augmented Lagrangian $L_c(., \lambda^*)$ corresponding to $\lambda^*$.

This suggest that if we set $\lambda$ close to $\lambda^*$ and do unconstrained minimization of Augmented Lagrangian:

$$x = \arg \min_{x \in X} L_c(x, \lambda)$$

Then we can find $x$ close to $x^*$. 
Second-order sufficiency conditions

Second Order Sufficiency Conditions: Let $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$ satisfy

$$\nabla_x L(x^*, \lambda^*) = 0, \quad \nabla_\lambda L(x^*, \lambda^*) = 0,$$

$$y^T \nabla_{xx}^2 L(x^*, \lambda^*) y > 0, \quad \forall y \neq 0 \text{ with } \nabla h(x^*)'y = 0.$$

Then $x^*$ is a strict local minimum.

We will show that if $c$ is larger than a threshold, then $x^*$ also satisfies these conditions for the Augmented Lagrangian $L_c(\cdot, \lambda^*)$ and hence is a strict local minimum of the Augmented Lagrangian $L_c(\cdot, \lambda^*)$ corresponding to $\lambda^*$. 
Convergence mechanisms

Augmented Lagrangian:

\[ L_c(x, \lambda) = f(x) + \lambda^\top h(x) + \frac{c}{2} \|h(x)\|^2 \]

Gradient and Hessian of Augmented Lagrangian:

\[ \nabla_x L_c(x, \lambda) = \nabla f(x) + \nabla h(x)(\lambda + ch(x)), \]

\[ \nabla_{xx} L_c(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^{m} (\lambda_i + ch_i(x)) \nabla^2 h_i(x) + c \nabla h(x) \nabla h(x)'. \]

If \( x^*, \lambda^* \) satisfy the sufficiency conditions of second-order for original problem, we get:

\[ \nabla_x L_c(x^*, \lambda^*) = \nabla f(x^*) + \nabla h(x^*)(\lambda^* + ch(x^*)) = \nabla_x L(x^*, \lambda^*) = 0, \]
Convergence mechanisms

\[ \nabla^2_{xx} L_c(x^*, \lambda^*) = \nabla^2 f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 h_i(x^*) + c \nabla h(x^*) \nabla h(x^*)' \]

\[ = \nabla^2_{xx} L(x^*, \lambda^*) + c \nabla h(x^*) \nabla h(x^*)'. \]

Since \( y' \nabla^2_{xx} L(x^*, \lambda^*) y > 0, \quad \forall \ y \neq 0 \) with \( \nabla h(x^*)' y = 0 \) from sufficiency condition, we have for large enough \( c \)

\[ y' \nabla^2_{xx} L_c(x^*, \lambda^*) y > 0, \quad \forall \ y \neq 0 \]

using the following lemma:

**Lemma:** Let \( P \) and \( Q \) be two symmetric matrices. Assume that \( Q \geq 0 \) and \( P > 0 \) on the nullspace of \( Q \), i.e., \( x'Px > 0 \) for all \( x \neq 0 \) with \( x'Qx = 0 \). Then there exists a scalar \( \overline{c} \) such that

\[ P + cQ : \text{positive definite}, \quad \forall \ c > \overline{c}. \]
Convergence mechanisms

1) Take $\lambda$ close to $\lambda^*$. 

2) Take $c$ very large, $c \to \infty$.

- For large $c$ and any $\lambda$

$$L_c(\cdot, \lambda) \approx \begin{cases} f(x) & \text{if } x \in X \text{ and } h(x) = 0 \\ \infty & \text{otherwise} \end{cases}$$

If $c$ is very large, then solution of unconstrained Augmented Lagrangian $x$ is nearly feasible.
Example

minimize \[ f(x) = \frac{1}{2}(x_1^2 + x_2^2) \]
subject to \( x_1 = 1 \)

\[ L(x, \lambda) = \frac{1}{2}(x_1^2 + x_2^2) + \lambda(x_1 - 1) \quad x^* = (1, 0) \quad \lambda^* = -1 \]

\[ L_c(x, \lambda) = \frac{1}{2}(x_1^2 + x_2^2) + \lambda(x_1 - 1) + \frac{c}{2}(x_1 - 1)^2 \]

\[ x_1(\lambda, c) = \frac{c - \lambda}{c + 1}, \quad x_2(\lambda, c) = 0 \]

We also have for all \( c > 0 \)

\[ \lim_{\lambda \to \lambda^*} x_1(\lambda, c) = x_1(-1, c) = 1 = x_1^* \]

We also have for all \( \lambda \)

\[ \lim_{c \to \infty} x_1(\lambda, c) = 1 = x_1^* \]
Example

minimize \( f(x) = \frac{1}{2}(x_1^2 + x_2^2) \)

subject to \( x_1 = 1 \)

\( x^* = (1, 0) \quad \lambda^* = -1 \)

\[
\lim_{\lambda \to \lambda^*} x_1(\lambda, c) = x_1(-1, c) = 1 = x_1^*, \quad \lim_{c \to \infty} x_1(\lambda, c) = 1 = x_1^*
\]
Quadratic Penalty Approach

How to choose $\lambda$ and $c$?

Solve sequence of unconstrained minimization of Augmented Lagrangian:

$$x^k = \arg \min_{x \in X} L_{ck}(x, \lambda^k)$$

where

$$L_{ck}(x, \lambda^k) \equiv f(x) + \lambda^k h(x) + \frac{c^k}{2} \|h(x)\|^2$$
Proposition: Assume that $f$ and $h$ are continuous functions, that $X$ is a closed set, and that the constraint set \( \{ x \in X \mid h(x) = 0 \} \) is nonempty. For $k = 0, 1, \ldots$, let $x^k$ be a global minimum of the problem

\[
\begin{align*}
\text{minimize} & \quad L_{c^k}(x, \lambda^k) \\
\text{subject to} & \quad x \in X,
\end{align*}
\]

where \( \{ \lambda^k \} \) is bounded, $0 < c^k < c^{k+1}$ for all $k$, and $c^k \to \infty$. Then every limit point of the sequence \( \{ x^k \} \) is a global minimum of the original problem.

- Assumes we can do exact minimization of the unconstrained Augmented Lagrangian
Inexact minimization

Proposition: Assume that \( X = \mathbb{R}^n \), and \( f \) and \( h \) are continuously differentiable. For \( k = 0, 1, \ldots \), let \( x^k \) satisfy

\[
\| \nabla_x L_{c^k}(x^k, \lambda^k) \| \leq \varepsilon^k,
\]

where \( \{\lambda^k\} \) is bounded, and \( \{\varepsilon^k\} \) and \( \{c^k\} \) satisfy

\[
0 < c^k < c^{k+1}, \quad \forall \ k, \quad c^k \to \infty, \quad 0 \leq \varepsilon^k, \quad \forall \ k, \quad \varepsilon^k \to 0.
\]

Assume \( x^k \to x^* \), where \( x^* \) is such that \( \nabla h(x^*) \) has rank \( m \). Then

\[
\lambda^k + c^k h(x^k) \to \lambda^*
\]

where \( \lambda^* \) is a vector satisfying, together with \( x^* \), the first order necessary conditions

\[
\nabla f(x^*) + \nabla h(x^*)\lambda^* = 0, \quad h(x^*) = 0.
\]
Practical issues

- Ill-conditioning: The condition number of the Hessian \( \nabla^2_{xx} L_{ck}(x^k, \lambda^k) \) tends to increase with \( c^k \).

Example:

\[
\begin{align*}
\text{minimize } f(x) &= \frac{1}{2} (x_1^2 + x_2^2) \\
\text{subject to } x_1 &= 1.
\end{align*}
\]

\[
\nabla^2_{xx} L_c(x, \lambda) = \begin{pmatrix}
1 + c & 0 \\
0 & 1
\end{pmatrix}.
\]

- To overcome ill-conditioning:
  - Use Newton-like method (and double precision).
  - Use good starting points.
  - Increase \( c^k \) at a moderate rate (if \( c^k \) is increased at a fast rate, \( \{x^k\} \) converges faster, but the likelihood of ill-conditioning is greater).
Method of Multipliers

Solve sequence of unconstrained minimization of Augmented Lagrangian:

\[ x^k = \arg \min_{x \in X} L_{c^k}(x, \lambda^k) \]

where

\[ L_{c^k}(x, \lambda^k) \equiv f(x) + \lambda^k h(x) + \frac{c^k}{2} \|h(x)\|^2 \]

and using the following multiplier update:

\[ \lambda^{k+1} = \lambda^k + c^k h(x^k) \]

- Note: Under some reasonable assumptions this works even if \( \{c^k\} \) is not increased to \( \infty \).
Method of Multipliers

Example: \[ \text{minimize } f(x) = \frac{1}{2}(x_1^2 + x_2^2) \]
subject to \( x_1 = 1. \)
\[ x^* = (1, 0) \quad \lambda^* = -1 \]

Method of Multipliers:
\[ x^k = \arg \min_{x \in \mathbb{R}^n} L_{c^k}(x, \lambda^k) = \left( \frac{c^k - \lambda^k}{c^k + 1}, 0 \right) \]
\[ \lambda^{k+1} = \lambda^k + c^k \left( \frac{c^k - \lambda^k}{c^k + 1} - 1 \right) \]
\[ \lambda^{k+1} - \lambda^* = \frac{\lambda^k - \lambda^*}{c^k + 1} \]

From this formula, it can be seen that

(a) \( \lambda^k \rightarrow \lambda^* = -1 \) and \( x^k \rightarrow x^* = (1, 0) \) for every nondecreasing sequence \( \{c^k\} \) [since the scalar \( 1/(c^k + 1) \) multiplying \( \lambda^k - \lambda^* \) in the above formula is always less than one].

(b) The convergence rate becomes faster as \( c^k \) becomes larger; in fact \( \{|\lambda^k - \lambda^*|\} \) converges superlinearly if \( c^k \rightarrow \infty \).
Method of Multipliers

Example:  

\[ \text{minimize } f(x) = \frac{1}{2}(-x_1^2 + x_2^2) \]

subject to  

\[ x_1 = 1. \]

\[ x^* = (1, 0) \quad \lambda^* = 1 \]

Method of Multipliers:

\[ x^k = \arg \min_{x \in \mathbb{R}^n} L_{c^k}(x, \lambda^k) = \left( \frac{c^k - \lambda^k}{c^k - 1}, 0 \right) \]

provided \( c^k > 1 \) (otherwise the min does not exist)

\[ \lambda^{k+1} = \lambda^k + c^k \left( \frac{c^k - \lambda^k}{c^k - 1} - 1 \right) \]

\[ \lambda^{k+1} - \lambda^* = -\frac{\lambda^k - \lambda^*}{c^k - 1} \]

- We see that:
  - No need to increase \( c^k \) to \( \infty \) for convergence; doing so results in faster convergence rate.
  - To obtain convergence, \( c^k \) must eventually exceed the threshold 2.
Practical issues

- Key issue is how to select \( \{c^k\} \).
  - \( c^k \) should eventually become larger than the "threshold" of the given problem.
  - \( c^0 \) should not be so large as to cause ill-conditioning at the 1st minimization.
  - \( c^k \) should not be increased so fast that too much ill-conditioning is forced upon the unconstrained minimization too early.
  - \( c^k \) should not be increased so slowly that the multiplier iteration has poor convergence rate.

- A good practical scheme is to choose a moderate value \( c^0 \), and use \( c^{k+1} = \beta c^k \), where \( \beta \) is a scalar with \( \beta > 1 \) (typically \( \beta \in [5, 10] \) if a Newton-like method is used).
Consider the problem

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad h_1(x) = 0, \ldots, h_m(x) = 0, \\
& \quad g_1(x) \leq 0, \ldots, g_r(x) \leq 0.
\end{align*}
\]

- Convert inequality constraint \( g_j(x) \leq 0 \) to equality constraint \( g_j(x) + z_j^2 = 0 \).
- The penalty method solves problems of the form

\[
\min_{x, z} \bar{L}_c(x, z, \lambda, \mu) = f(x) + \lambda^T h(x) + \frac{c}{2} \|h(x)\|^2
\]

\[
+ \sum_{j=1}^{r} \left\{ \mu_j (g_j(x) + z_j^2) + \frac{c}{2} |g_j(x) + z_j^2|^2 \right\},
\]

for various values of \( \mu \) and \( c \).
Inequality constraints

• First minimize $\bar{L}_c(x, z, \lambda, \mu)$ with respect to $z$,

$$L_c(x, \lambda, \mu) = \min_z \bar{L}_c(x, z, \lambda, \mu) = f(x) + \lambda' h(x) + \frac{c}{2} \|h(x)\|^2$$

$$+ \sum_{j=1}^{r} \min_{z_j} \left\{ \mu_j \left( g_j(x) + z_j^2 \right) + \frac{c}{2} \left| g_j(x) + z_j^2 \right|^2 \right\}$$

and then minimize $L_c(x, \lambda, \mu)$ with respect to $x$.

• Can show this reduces to:

$$L_c(x, \lambda, \mu) = f(x) + \lambda' h(x) + \frac{c}{2} \|h(x)\|^2$$

$$+ \frac{1}{2c} \sum_{j=1}^{r} \left\{ \left( \max\{0, \mu_j + c g_j(x)\} \right)^2 - \mu_j^2 \right\}$$

• Under similar assumptions as before,

$$\left\{ \lambda_i^k + c^k h_i(x^k) \right\} \rightarrow \lambda_i^* \quad \max\{0, \mu_j^k + c^k g_j(x^k)\} \rightarrow \mu_j^*$$