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

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \qquad h(x) = 0, \end{aligned}$$

where $f : \Re^n \to \Re$ and $h : \Re^n \to \Re^m$ are continuous, and X is closed.

Quadratic Penalty Approach

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

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

<u>Note:</u> Problem is unchanged – has same local minima

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.

Quadratic Penalty Approach

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?

Convergence mechanisms

1) Take λ close to λ^*.

Let x^*, λ^* 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 λ^*.

This suggest that if we set λ close to λ^* 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 $x^* \in \Re^n$ and $\lambda^* \in \Re^m$ satisfy

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

$$y' \nabla^2_{xx} L(x^*, \lambda^*) y > 0, \quad \forall\ y \neq 0 \text{ with } \nabla h(x^*)' y = 0.$$

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 $L_c(., \lambda^*)$ and hence is a strict local minimum of the Augmented Lagrangian $L_c(., \lambda^*)$ corresponding to λ^*.

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:

$$\nabla_x L_c(x,\lambda) = \nabla f(x) + \nabla h(x)\big(\lambda + ch(x)\big),$$

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

If x*, λ* satisfy the sufficiency conditions of second-order for original problem, we get:

$$\nabla_x L_c(x^*,\lambda^*) = \nabla f(x^*) + \nabla h(x^*)\big(\lambda^* + ch(x^*)\big) = \nabla_x L(x^*,\lambda^*) = 0,$$

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 $y' \nabla^2_{xx} L(x^*, \lambda^*) y > 0, \quad \forall\, y \neq 0 \text{ with } \nabla h(x^*)' y = 0$
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 $\bar{c}$ such that

$$P + cQ : \text{positive definite}, \qquad \forall\, c > \bar{c}.$$

Convergence mechanisms

1) Take λ close to λ*.

2) Take c very large, c ➔ ∞.

   – For large c and any λ

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

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 $x_1 = 1$

$$L(x,\lambda) = \tfrac{1}{2}(x_1^2 + x_2^2) + \lambda(x_1 - 1) \qquad x^* = (1,0) \qquad \lambda^* = -1$$

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

$$x_1(\lambda, c) = \frac{c - \lambda}{c + 1}, \qquad x_2(\lambda, c) = 0$$

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 λ

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

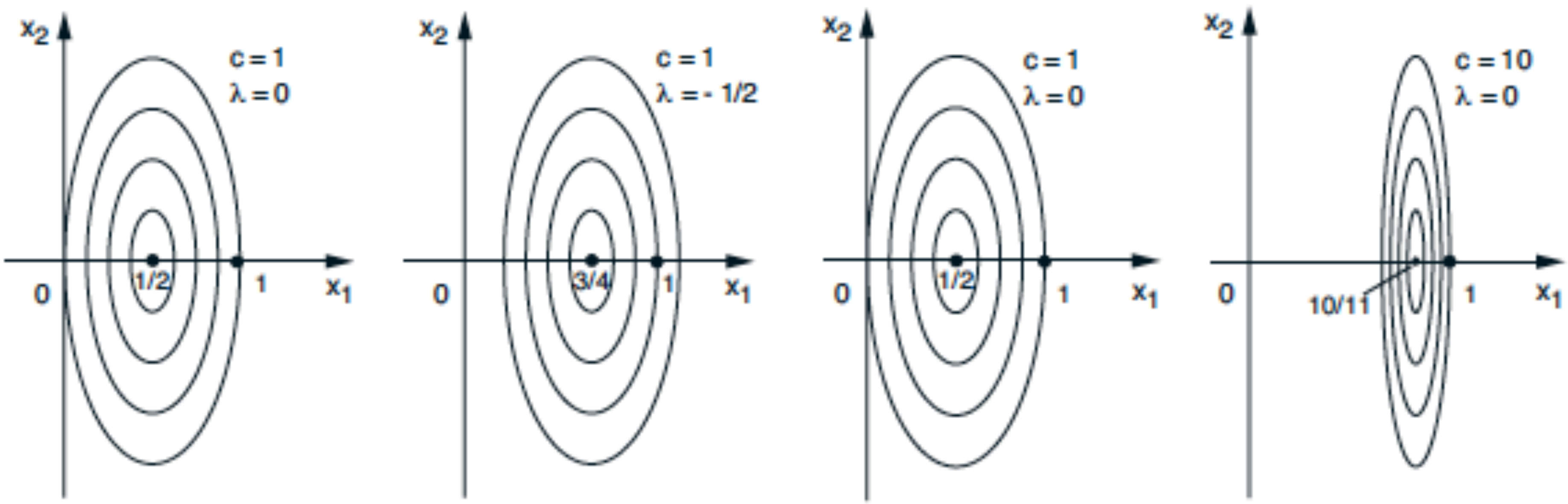
Example

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

subject to $x_1 = 1$

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

$$\lim_{\lambda \to \lambda^*} x_1(\lambda, c) = x_1(-1, c) = 1 = x_1^*, \qquad \lim_{c \to \infty} x_1(\lambda, c) = 1 = x_1^*$$



Quadratic Penalty Approach

How to choose λ and c?

Solve sequence of unconstrained minimization of Augmented Lagrangian:

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

where

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

Basic convergence result

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

$$\begin{aligned} &\text{minimize} \quad L_{c^k}(x, \lambda^k) \\ &\text{subject to} \quad x \in X, \end{aligned}$$

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 $X = \Re^n$, and f and h are continuously differentiable. For $k = 0, 1, \ldots$, let x^k satisfy

$$\|\nabla_x L_{c^k}(x^k, \lambda^k)\| \leq \epsilon^k,$$

where $\{\lambda^k\}$ is bounded, and $\{\epsilon^k\}$ and $\{c^k\}$ satisfy

$$0 < c^k < c^{k+1}, \quad \forall\, k, \qquad c^k \to \infty, \qquad 0 \leq \epsilon^k, \quad \forall\, k, \qquad \epsilon^k \to 0.$$

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 λ^* is a vector satisfying, together with x^*, the first order necessary conditions

$$\nabla f(x^*) + \nabla h(x^*)\lambda^* = 0, \qquad h(x^*) = 0.$$

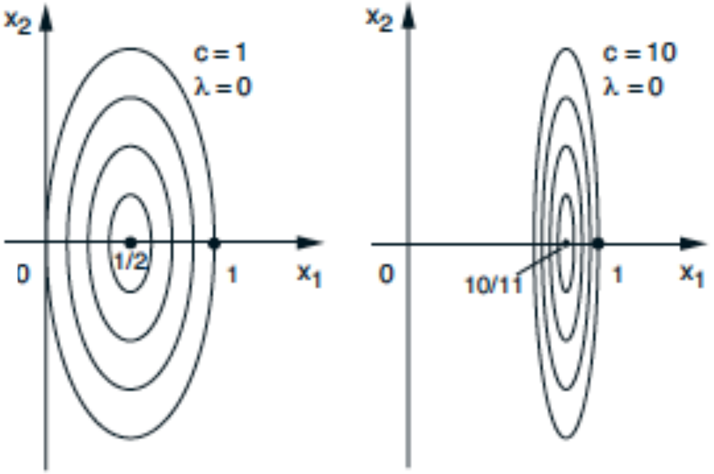
Practical issues

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

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

subject to $x_1 = 1$.

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



- 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

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

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 ∞.

Method of Multipliers

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

subject to $x_1 = 1.$

Convex problem

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

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

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

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

From this formula, it can be seen that

(a) $\lambda^k \to \lambda^* = -1$ and $x^k \to x^* = (1, 0)$ 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 $c^k \to \infty$.

Method of Multipliers

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

subject to $x_1 = 1.$

Non-convex problem

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

Method of Multipliers: $$x^k = \arg\min_{x \in \Re^n} L_{c^k}(x, \lambda^k) = \left(\frac{c^k - \lambda^k}{c^k - 1}, 0\right)$$

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 ∞ 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 β is a scalar with $\beta > 1$ (typically $\beta \in [5, 10]$ if a Newton-like method is used).

Inequality constraints

Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h_1(x) = 0, \ldots, h_m(x) = 0, \\ & g_1(x) \le 0, \ldots, g_r(x) \le 0. \end{array}$$

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

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

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

for various values of μ and c.

Inequality constraints

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

$$L_c(x, \lambda, \mu) = \min_z \bar{L}_c(x, z, \lambda, \mu) = f(x) + \lambda' h(x) + \frac{c}{2}\|h(x)\|^2$$
$$+ \sum_{j=1}^{r} \min_{z_j} \left\{ \mu_j \left(g_j(x) + z_j^2 \right) + \frac{c}{2} |g_j(x) + z_j^2|^2 \right\}$$

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,

$$\{\lambda_i^k + c^k h_i(x^k)\} \to \lambda_i^* \qquad \max\{0, \mu_j^k + c^k g_j(x^k)\} \to \mu_j^*$$