# Constrained Optimization

- Lagrange Multiplier
- KKT conditions
  - projection onto simplex
  - e.g. sampling for depth sensing
- Duality
  - Lagrange Dual function
  - Dual problem
  - duality gap
  - strong duality
  - Conjugate function
- Method of Multiplier
  - Quadratic penalty
  - Multiplier update : Pictorial illustration
  - Augmented Lagrangian method
  - ADMM
  - ADMM examples

# Lagrange Multiplier

We will consider general constrained optimization

$$\min \ f(x)$$
$$\text{s.t.} \ \ h(x) \geq 0$$

The standard procedure is to consider

$$\mathcal{L}(x, \lambda) = f(x) - \lambda h(x),$$

where $\lambda$ is called the Lagrange-multiplier. If there are multiple constraints, e.g. $h_i(x) \geq 0$ for $i = 1, 2, \ldots, n$, then $\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^{n} \lambda_i h_i(x)$. For $(x^*, \lambda^*)$ to be the solution, we need

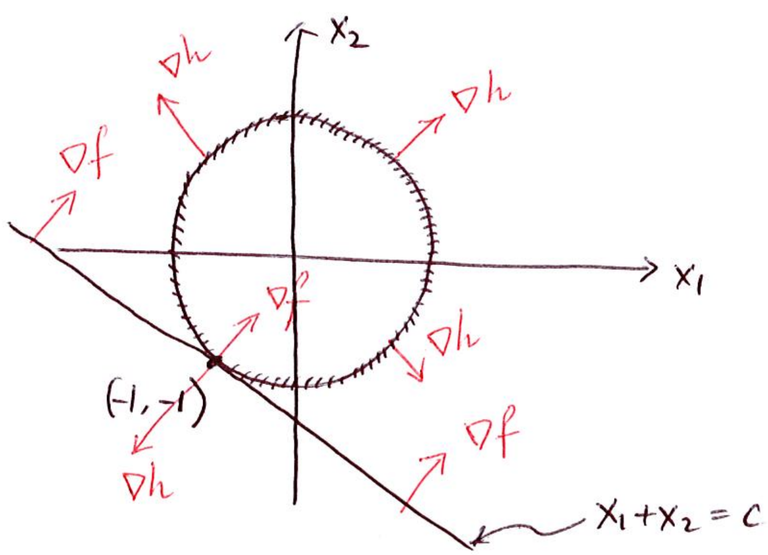
$$\boxed{\nabla \mathcal{L}(x^*, \lambda^*) = 0}, \quad \text{where } \nabla \text{ is taken wrt } x \text{ and } \lambda.$$

---

Case Study 1 : $\begin{cases} \min & x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 = 2 \end{cases}$

The solution $(x_1, x_2)$ must live on the circle defined by $h(x) = 0$.

$$f(x) = x_1 + x_2$$
$$h(x) = x_1^2 + x_2^2 - 2.$$



At optimal point $(x_1^*, x_2^*) = (-1, -1)$, we can show that

$$\nabla f(x^*) = \begin{bmatrix} \left.\frac{\partial f}{\partial x_1}\right|_{x_1^*} \\ \left.\frac{\partial f}{\partial x_2}\right|_{x_2^*} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\nabla h(x^*) = \begin{bmatrix} \left.\frac{\partial h}{\partial x_1}\right|_{x_1^*} \\ \left.\frac{\partial h}{\partial x_2}\right|_{x_2^*} \end{bmatrix} = \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

That means, ~~∇f~~ $\nabla f(x^*) = \frac{-1}{2} \nabla h(x^*)$,

or

$$\nabla f(x^*) = \lambda \nabla h(x^*).$$

this says: $\boxed{\nabla f \text{ is parallel to } \nabla h.}$

$$\Longrightarrow \underbrace{\nabla f(x^*) - \lambda \nabla h(x^*)}_{= \nabla \mathcal{L}(x^*)} = 0$$

In general, if we have an equality constraint, we can show that $\nabla f$ must be in parallel to $\nabla h$ at optimal point.

Let $x$ be a feasible point and $d$ be a feasible direction. Consider $x+d$. In order for $x+d$ remain feasible, we need

$$0 = h(x+d) \cong \underbrace{h(x)}_{=0} + \nabla h(x)^T d$$

So that implies

$$\boxed{\nabla h(x)^T d = 0} \qquad \text{———— (1)}$$

In addition, if $d$ is a good search direction, then we want

$$0 > f(x+d) - f(x) \cong \nabla f(x)^T d,$$

which implies,

$$\boxed{\nabla f(x)^T d < 0} \qquad \text{———— (2)}$$

Therefore, if we have a point $x^*$, then $x^*$ is optimal when we cannot find a search direction $d$ s.t.

$$\begin{cases} \nabla h(x^*)^T d = 0 & \leftarrow d \text{ can make } x+d \text{ feasible} \\ \nabla f(x^*)^T d < 0 & \leftarrow d \text{ can reduce objective} \end{cases}$$

So such $d$ cannot exists if

$\nabla f(x^*)$ is parallel to $\nabla h(x^*)$.

---

<u>Case Study 2</u>

$$\begin{cases} \min & x_1 + x_2 \\ \text{s.t.} & 2 - x_1^2 - x_2^2 \geq 0 \end{cases}$$

The constraint now becomes $h(x) \geq 0$.

Note:

$$0 \leq h(x+d) \cong \underbrace{h(x)}_{\geq 0} + \nabla h(x)^T d$$

$\geq 0$ if $x$ interior.
$= 0$ if $x$ boundary

$$\Rightarrow \boxed{\nabla h(x)^T d \geq -h(x)}$$

If $x$ is an interior point, then define

$$d = -h(x) \frac{\nabla f(x)}{\|\nabla f(x)\|}$$

Claim (i) $\nabla f(x)^T d < 0$ : $\nabla f(x)^T \left[ \frac{\nabla f(x)}{\|\nabla f(x)\|} (-h(x)) \right]$

$$= -\underbrace{h(x)}_{\geq 0 \text{ for interior.}} \|\nabla f(x)\| < 0$$

(ii) $\nabla h(x)^T d \geq -h(x)$ :

$\frac{\nabla f(x)}{\|\nabla f(x)\|}$ = unit step search direction. $h(x)$ = step size.

Therefore, since $x$ is interior, any unit direction will work. The step size just needs to be small enough; e.g. $h(x)$.

So, $d$ exists if we can have

(i) $\nabla h(x)^T d \geq -h(x)$

(ii) $\nabla f(x)^T d < 0$.

This cannot happen if $\nabla f(x) = 0$.

Now $\nabla f(x) = \lambda \nabla h(x)$. If $\nabla f(x) = 0$ ~~and~~ then either $\lambda = 0$ or $\nabla h(x) = 0$ or both. This happens when $h(x) > 0$. So we have

$$\text{if } h(x) > 0, \text{ then } \lambda = 0, \left[\text{or } \nabla h(x) = 0\right]$$

if $x$ is on boundry so that $h(x) = 0$. Then it becomes an equality constraint case. So we have

$$\begin{cases} \nabla h(x)^T d \geqq 0 \leftarrow \text{closed half space} \\ \nabla f(x)^T d < 0 \leftarrow \text{open half space} \end{cases}$$

They cannot intersect if $\nabla f$ and $\nabla h$ are pointing in the same direction. So

$$\nabla f(x) = \lambda \nabla h(x) \text{ and must have } \lambda \geq 0.$$

So

$$\text{if } h(x) = 0, \text{ then } \lambda \geq 0.$$

if $x^*$ is an interior point, then

$$\nabla f(x^*) = 0 .$$

This implies $\lambda^* = 0$.

if $x^*$ is on boundary, then

$$\nabla f(x^*) = \lambda^* \nabla h(x^*) . \quad \text{and } \lambda^* > 0 .$$

So in general, we have

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \quad \text{for } \lambda \geq 0,$$

and $\lambda^* h(x^*) = 0$. ← complementary slackness.

⇓

if $h(x^*) > 0$, then $\lambda^* = 0$.

if $\lambda^* > 0$, then $h(x^*) = 0$.

Can have both:

$\lambda^* = 0$ & $h(x^*) = 0$.

---

## Karush-Kuhn-Tucker Condition

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g_i(x) \geq 0 \\ & h_j(x) = 0 \end{cases}$$

The KKT condition is the first-order necessary condition:

Let $\mathcal{L}(x; \mu, \lambda) = f(x) - \sum_i \mu_i g_i(x) - \sum_j \lambda_j h_j(x)$.

Then

(1) $\nabla_x \mathcal{L}(x^*, \mu^*, \lambda^*) = 0$ (stationarity)

(2) $g_i(x^*) \geq 0$, $h_j(x^*) = 0$ (primal feasibility)

(3) $\mu_i^* \geq 0$ (dual feasibility)

(4) $\mu_i^* g_i(x^*) = 0$ (complementary slackness)

## Example

$$\min_x \frac{1}{2}\|x-b\|^2$$
$$\text{s.t.} \quad x \geq 0 \;, \; x^T 1 = 1 .$$

The Lagrangian is

$$\mathcal{L}(x, \lambda, \gamma) = \frac{1}{2}\|x-b\|^2 - \lambda^T x - \gamma(1 - x^T 1)$$

The stationarity condition implies

$$\frac{\partial}{\partial x}\mathcal{L} = 0 = x - b - \lambda + \gamma$$
$$\Longrightarrow \quad \boxed{x_i = b_i + \lambda_i - \gamma .} \qquad (1)$$

Primal Feasibility requires $x_i \geq 0$ and $\sum_i x_i = 1$.

Dual Feasibility requires $\lambda_i \geq 0$.

Complementary Slackness: $\lambda_i x_i = 0$.

Consider (1):

if $\lambda_i = 0$, then $x_i = b_i - \gamma$.

By complementary slackness, ~~this~~ $\lambda_i = 0$ implies $x_i > 0$. So we have $b_i - \gamma > 0$.

if $\lambda_i > 0$, then $x_i = 0$ by complementary slackness.

So we have $x_i = b_i + \lambda_i - \gamma = 0$

$$\Longrightarrow \quad b_i + \lambda_i = \gamma$$
$$\Longrightarrow \quad b_i < \gamma \qquad \text{because } \lambda_i > 0 .$$

Combining these two cases, we have

(i) if $b_i > \gamma$, then $x_i = b_i - \gamma$

(ii) if $b_i < \gamma$, then $x_i = 0$

(iii) if $b_i = \gamma$, then $x_i = \lambda_i \Longrightarrow x_i = 0$

This can be compactly written as

$$x_i = \max(b_i - \gamma, 0).$$

It remains to determine $\gamma$: By primal feasibility again, we have

$$\sum_{i=1}^{n} x_i = 1 \implies \sum_{i=1}^{n} \max(b_i - \gamma, 0) = 1.$$

So $\gamma$ can be determined by finding the root of the function $g(\gamma) = \sum_{i=1}^{n} \max(b_i - \gamma, 0) - 1$.

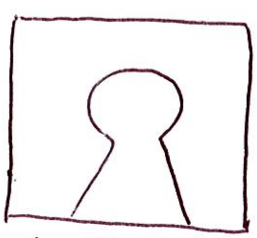
<u>Remark</u>: if the problem is $\min \frac{1}{2}\|Ax - b\|^2$ s.t. $x \geq 0$, $x^T 1 = 1$,

then the above one-shot solution is not applicable because stationarity condition implies

$$A^T A x = A^T b + \lambda - \gamma.$$

coupling of $x_i$ and $x_j$.

<u>Application</u>: Sampling for depth recovery



depth map is typically piece-wise constant

Let $b_i$ = $i^{th}$ pixel's gradient magnitude.

Goal: have more samples along the gradient, and less on flat areas. Also, want to target a fixed number of samples.

Ideal average gradient

$$\mu = \frac{1}{N} \sum_{j=1}^{N} b_j.$$

Randomly selected:

$$Y = \frac{1}{N} \sum_{j=1}^{N} \frac{b_j}{p_j} I_j \qquad \mathbb{P}(I_j = 1) = p_j$$

$$\mathrm{Var}(Y) = \mathbb{E}\left[(Y-\mu)^2\right]$$
$$= \frac{1}{N}\sum_{j=1}^{N} \frac{b_j^2}{p_j^2}\,\mathrm{Var}(I_j)$$
$$= \frac{1}{N}\sum_{j=1}^{N} b_j^2 \left(\frac{1-p_j}{p_j}\right)$$

Optimization:

$$\min_{\{p_j\}} \frac{1}{N}\sum_{j=1}^{N} b_j \left(\frac{1-p_j}{p_j}\right)$$
$$\text{s.t.} \quad \sum_{j=1}^{N} p_j = \xi, \quad 0 \le p_j \le 1.$$

Solution: $p_j = \max(\tau b_j, 1)$

where $\tau$ solves

$$g(\tau) = \sum_{j=1}^{N} \max(\tau b_j, 1) - \xi N.$$