VI.2] Lagrange Multipliers = Derivative of the ~~Cost~~ Optimal value

Example:

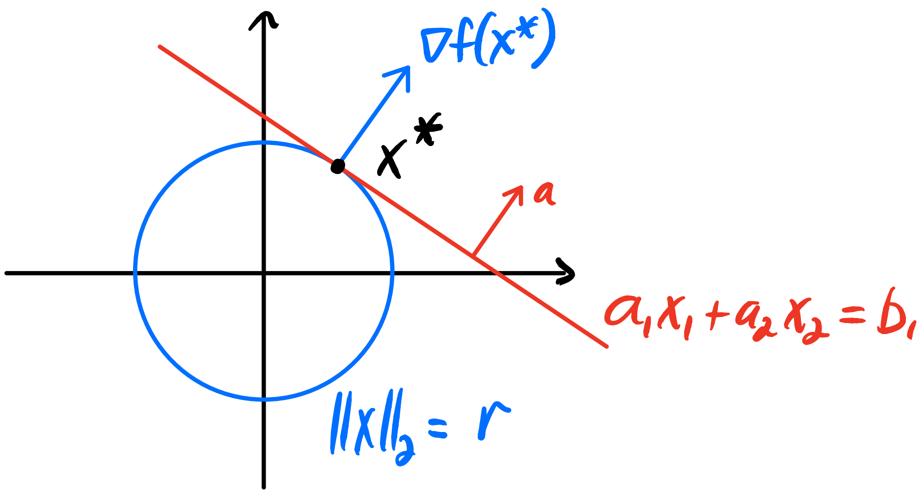
Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, with $m < n$.

$$\boxed{\min \ \tfrac{1}{2}\|x\|_2^2 \quad \text{s.t.} \quad Ax = b, \quad x \in \mathbb{R}^n}$$

$m = 1, \ n = 2$:

$f(x) = \tfrac{1}{2}\|x\|_2^2$

$r = \|x^*\|_2$



Note that $\nabla f(x^*)$ must be a scalar multiple of a:

$$\boxed{\nabla f(x^*) = \lambda a}.$$

The scalar λ is the <u>Lagrange multiplier</u>.

The gradient of the objective function $f(x)=\frac{1}{2}\|x\|_2^2$ is

$$\nabla f(x) = x.$$

Thus, $\nabla f(x^*) = \lambda a$ implies that

$$x^* = \lambda a.$$

Also, $a^T x^* = b_1$, so

$$\lambda = \frac{b_1}{a^T a}.$$

Thus, the optimal sol'n is

$$\boxed{x^* = \frac{b_1}{a^T a} a}$$

and the optimal value is

$$f(x^*) = \frac{1}{2}\left(\frac{b_1}{a^T a}\right)^2 a^T a = \frac{1}{2}\frac{b_1^2}{a^T a}.$$

Notice that

$$\frac{d}{db_1}\left(\frac{1}{2}\frac{b_1^2}{a^T a}\right) = \frac{b_1}{a^T a} = \lambda.$$

Consider the general problem above with n variables and m constraints.

Each constraint has a Lagrange multiplier, so there are m multipliers: $\lambda_1, \dots, \lambda_m \in \mathbb{R}$.
Let $\lambda \in \mathbb{R}^m$ be the vector of Lagrange multipliers.

The <u>Lagrangian</u> is

$$\boxed{L(x, \lambda) = \frac{1}{2}\|x\|_2^2 + \lambda^T(b - Ax)}.$$

Note that

$$\max_{\lambda \in \mathbb{R}^m} L(x, \lambda) = \begin{cases} \frac{1}{2}\|x\|_2^2 & \text{if } Ax = b, \\ +\infty & \text{if } Ax \neq b. \end{cases}$$

Thus, our problem can be written as:

$$\min_x \max_\lambda L(x, \lambda).$$

This is a **two-player game**. Player X goes first and wants to make $L(x,\lambda)$ as small as possible. Player Λ goes second and wants to make $L(x,\lambda)$ as large as possible. Clearly, Player X will want to choose x that satisfies $Ax=b$, since otherwise Player Λ can make $L(x,\lambda)$ arbitrarily large.

It is not hard to show that

$$\min_x \max_\lambda L(x,\lambda) \geq \max_\lambda \min_x L(x,\lambda).$$

Thus, switching the order of the players will give us a lower bound on the optimal value of the original problem.

Let's first look at the <u>unconstrained</u> problem

$$\min_x \; L(x, \lambda) = \frac{1}{2}\|x\|_2^2 + \lambda^T(b - Ax).$$

First we find the <u>critical points</u>, so we solve $\nabla_x L(x, \lambda) = 0$ for x:

$$\nabla_x L(x, \lambda) = x - A^T\lambda = 0$$

Thus, the only critical point is

$$\bar{x} = A^T\lambda.$$

Plugging this into $L(x, \lambda)$, we have

$$L(\bar{x}, \lambda) = \frac{1}{2}\|A^T\lambda\|_2^2 + \lambda^T(b - AA^T\lambda)$$

$$= b^T\lambda - \frac{1}{2}\|A^T\lambda\|_2^2.$$

This is what Player Λ wants to maximize.

Let $g(\lambda) = b^T\lambda - \frac{1}{2}\|A^T\lambda\|_2^2$. Then

$$\nabla g(\lambda) = b - AA^T\lambda,$$

so $\nabla g(\lambda) = 0$ <u>iff</u>

$$AA^T\lambda = b.$$

Note that AA^T is nonsingular <u>iff</u> the rows of A are linearly independent. In this case, the only critical point is

$$\bar{\lambda} = (AA^T)^{-1}b$$

and we have

$$L(\bar{x}, \bar{\lambda}) = b^T(AA^T)^{-1}b - \frac{1}{2}\|A^T(AA^T)^{-1}b\|_2^2.$$

This simplifies to

$$L(\bar{x}, \bar{\lambda}) = \frac{1}{2} b^T(AA^T)^{-1}b.$$

This is the value of

$$\max_{\lambda} \min_{x} L(x,\lambda).$$

Therefore,

$$L(\bar{x},\bar{\lambda}) \leq \min_{x} \max_{\lambda} L(x,\lambda),$$

which implies that

$$\frac{1}{2} b^T (AA^T)^{-1} b \leq \frac{1}{2}\|x\|_2^2, \quad \text{if } Ax = b.$$

Notice that

$$AA^T\bar{\lambda} = b \quad \text{and} \quad \bar{x} = A^T\bar{\lambda}$$

implies that $A\bar{x} = b$. Moreover,

$$\frac{1}{2}\|\bar{x}\|_2^2 = \frac{1}{2}\|A^T(AA^T)^{-1}b\|_2^2 = \frac{1}{2} b^T (AA^T)^{-1} b \quad !!!$$

Thus, $\bar{x}$ achieves the lower bound, so is optimal!

Therefore, the **optimal sol'n** of the optimization problem

$$\boxed{\min \ \tfrac{1}{2}\|x\|_2^2 \ \text{ s.t. } \ Ax = b, \ x \in \mathbb{R}^n}$$

is

$$\boxed{x^* = \bar{x} = A^T\bar{\lambda} = A^T(AA^T)^{-1}b}$$

and the **optimal value** is

$$\boxed{v^* = \tfrac{1}{2} b^T(AA^T)^{-1}b}.$$

Furthermore, the **optimal Lagrange multiplier** is

$$\boxed{\lambda^* = \bar{\lambda} = (AA^T)^{-1}b}.$$

[Recall the $n=2$, $m=1$ case:

$$x^* = \frac{b_1}{a^Ta}a, \quad \lambda^* = \frac{b_1}{a^Ta}, \quad v^* = \frac{1}{2}\frac{b_1^2}{a^Ta}.$$
]

Moreover, $\lambda^* = (AA^T)^{-1}b$ is the gradient of the <u>optimal value function</u>

$$\boxed{v(b) = \frac{1}{2} b^T (AA^T)^{-1} b}.$$

Indeed, $\nabla v(b) = (AA^T)^{-1} b$, so

$$\boxed{\nabla v(b) = \lambda^*}.$$

Thus, Lagrange multipliers = rate of change of the optimal value with respect to changes in the right-hand-side of the constraints.
