FOUNDATIONS OF OPTIMIZATION

Basics

- ***Optimization problems***
  - An optimization problem is

    $$\text{minimise } f(\boldsymbol{x}) \text{ subject to } \boldsymbol{x} \in \mathcal{C}$$

    f is the *objective* (real) $\mathcal{C}$ is the constraint set/feasible set/search space.
  - $\boldsymbol{x}^*$ is an *optimal solution* (*global minimizer*) if and only if

    $$f(\boldsymbol{x}^*) \le f(\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in \mathcal{C}$$
  - Maximizing $f(\boldsymbol{x})$ is equivalent to minimizing $-f(\boldsymbol{x})$.
  - We consider problems in the following form

    $$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & h_i(\boldsymbol{x}) = 0 \qquad \forall\, 1 \le i \le m \\ & g_i(\boldsymbol{x}) \le 0 \qquad \forall\, m \le i \le r \\ & \boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$
  - We consider the following subsets of the problem
    - In *linear programming*, all functions are linear.
    - In convex programming, the f and g are convex, and the h are linear.
  - If $\mathcal{C}$ is the feasible set of a problem, a point $\boldsymbol{x} \in \mathcal{C}$ is a *local minimum* if there exists a neighborhood $N_r(\boldsymbol{x})$ such that $f(\boldsymbol{x}) \le f(\boldsymbol{y})\ \ \forall \boldsymbol{y} \in \mathcal{C} \cap N_r(\boldsymbol{x})$. It is an *unconstrained local minimum* if $f(\boldsymbol{x}) \le f(\boldsymbol{y})\ \ \forall \boldsymbol{y} \in N_r(\boldsymbol{x})$. (Strict equivalents exist).
- ***Topology***
  - An *open ball* around a point $\boldsymbol{x} \in \mathbb{R}^n$ with radius $r > 0$ is the set $N_r(\boldsymbol{x}) = \left\{ \boldsymbol{y} \in \mathbb{R}^n : \left\| \boldsymbol{x} - \boldsymbol{y} \right\| < r \right\}$, where $\left\| \boldsymbol{x} \right\| = \sqrt{\sum x_i^2}$.
  - A point $\boldsymbol{x} \in \mathcal{E} \subset \mathbb{R}^n$ is an *interior point* if there exists an open ball such that $N_r(\boldsymbol{x}) \subset \mathcal{E}$. A set $\mathcal{E} \subset \mathbb{R}^n$ is *open* if $\mathcal{E} = \operatorname{int} \mathcal{E}$.

- ***Definition***: A function is *convex* if $f\left(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2\right) \leq \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2)$. It is *strictly convex* if the inequality is strict for $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$. We say f is convex over $\mathcal{X} = \operatorname{dom} f$ if it is convex when restricted to $\mathcal{X}$. f is (strictly)concave if $-f$ is (strictly) convex.
- ***Definition***: If f is convex with a convex domain $\mathcal{X}$, we define the *extended-value extension* $\tilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ by

$$\tilde{f}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & \text{if } \boldsymbol{x} \in \mathcal{X} \\ \infty & \text{otherwise} \end{cases}$$

  and we let

$$\operatorname{dom} \tilde{f} = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \tilde{f}(\boldsymbol{x}) < \infty \right\}$$

  An extended-value function is convex if
  - Its domain is convex.
  - The standard convexity property holds.
- Given $\mathcal{C} \subset \mathbb{R}^n$, the *indicator function* $I_{\mathcal{C}} : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ as

$$I_{\mathcal{C}}(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \in \mathcal{C} \\ \infty & \text{otherwise} \end{cases}$$

  If $\mathcal{C}$ is a convex set, then $I_{\mathcal{C}}$ is a convex function.
- ***Theorem***: If f is convex over a convex set $\mathcal{C} \subset \mathbb{R}^n$, then every sublevel set $\left\{ \boldsymbol{x} \in \mathcal{C} : f(\boldsymbol{x}) \leq \gamma \right\}$ is a convex subset of $\mathbb{R}^n$. The converse is *not* true (eg: $\log x$ on $(0, \infty)$). However, we define…
- …***Definition***: A extended real valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is *quasiconvex* if, every one of its sublevel sets (ie: for every $\gamma \in \mathbb{R}$) is convex.

Calculus

- A function $f : \mathcal{X} \to \mathbb{R}$ with $\mathcal{X} \subset \mathbb{R}^n$ is *differentiable* at $\boldsymbol{x} \in \operatorname{int} \mathcal{X}$ if there exists a vector $\nabla f(\boldsymbol{x}) \in \mathbb{R}^n$, known as the *gradient*, such that

$$\lim_{\boldsymbol{d} \to \boldsymbol{0}} \frac{f(\boldsymbol{x} + \boldsymbol{d}) - f(\boldsymbol{x}) - \nabla f(\boldsymbol{x}) \cdot \boldsymbol{d}}{\left\| \boldsymbol{d} \right\|} = 0$$

  And

$$\left(\boldsymbol{d}^\top + \dot{\boldsymbol{u}}(0)^\top \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top\right)\nabla \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$$

But since $\boldsymbol{d} \in \mathcal{V}(\boldsymbol{x}^*)$, we also have that $\boldsymbol{d}^\top \nabla \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$, and so $\dot{\boldsymbol{u}}(0) = \boldsymbol{0}$, and $\dot{\boldsymbol{x}}(0) = \boldsymbol{d}$, as required.

- [It will be useful for later to note that if $\boldsymbol{h}(\cdot)$ is twice continuously differentiable, then so is $\boldsymbol{x}(\cdot)$. Though I'm not quite sure how to prove that result].

  We now have our elusive curve! Let's now prove the theorem.

- $\boxed{\mathcal{V}(\boldsymbol{x}^*) \subset \mathcal{T}(\boldsymbol{x}^*)}$: choose $\boldsymbol{d} \in \mathcal{V}(\boldsymbol{x}^*) \setminus \{\boldsymbol{0}\}$ and let $\boldsymbol{x}(t)$ be the curve discussed above. Take a sequence $t_k \subset (0, \tau), t_k \to 0$, so that $\boldsymbol{x}(t_k) \neq \boldsymbol{x}^*$. Then, by the mean value theorem, there is some $\tilde{t} \in [0, t_k]$ such that

$$\boldsymbol{x}(t_k) - \boldsymbol{x}(0) = \dot{\boldsymbol{x}}(\tilde{t})(t_k - 0)$$
$$\frac{\boldsymbol{x}(t_k) - \boldsymbol{x}^*}{\left\|\boldsymbol{x}(t_k) - \boldsymbol{x}^*\right\|} = \frac{\dot{\boldsymbol{x}}(\tilde{t})}{\left\|\boldsymbol{x}(t_k) - \boldsymbol{x}^*\right\| / t_k}$$

  As $t_k \to 0$ and therefore $\tilde{t} \to 0$, this tends to

$$\frac{\dot{\boldsymbol{x}}(0)}{\left\|\dot{\boldsymbol{x}}(0)\right\|} = \frac{\boldsymbol{d}}{\left\|\boldsymbol{d}\right\|}$$

  So $\boldsymbol{d} \in \mathcal{T}(\boldsymbol{x}^*)$.

- $\boxed{\mathcal{T}(\boldsymbol{x}^*) \subset \mathcal{V}(\boldsymbol{x}^*)}$: consider $\boldsymbol{d} \in \mathcal{T}(\boldsymbol{x}^*) \setminus \{\boldsymbol{0}\}$ and an associated sequence $\{\boldsymbol{x}_k\}$ in the feasible set, as defined in the definition of $\mathcal{T}(\boldsymbol{x}^*)$. By the mean value theorem, there is some $\tilde{\boldsymbol{x}} \in [\boldsymbol{x}_k, \boldsymbol{x}^*]$ such that

$$\boldsymbol{h}(\boldsymbol{x}_k) - \boldsymbol{h}(\boldsymbol{x}^*) = \nabla \boldsymbol{h}(\tilde{\boldsymbol{x}})^\top (\boldsymbol{x}_k - \boldsymbol{x}^*)$$

  But since $\boldsymbol{x}^*$ and every $\boldsymbol{x}_k$ are in the feasible set, $\boldsymbol{h}(\boldsymbol{x}_k) = \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$, and

$$0 = \boldsymbol{h}(\boldsymbol{x}_k) = \boldsymbol{h}(\boldsymbol{x}^*) + \nabla \boldsymbol{h}(\tilde{\boldsymbol{x}}_k)^\top (\boldsymbol{x}_k - \boldsymbol{x}^*) = \nabla \boldsymbol{h}(\tilde{\boldsymbol{x}}_k)^\top (\boldsymbol{x}_k - \boldsymbol{x}^*)$$
$$\nabla \boldsymbol{h}(\tilde{\boldsymbol{x}}_k)^\top \frac{\boldsymbol{x}_k - \boldsymbol{x}^*}{\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|} = 0$$
$$\to \nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{d} = \boldsymbol{0}$$

  And so $\boldsymbol{d} \in \mathcal{V}(\boldsymbol{x}^*)$.

{*Done! Take a deep breath!*}

- ***Theorem – necessary condition***: If $\boldsymbol{x}^*$ is a local minimum that is a regular point, then there is no descent direction that is also a first order feasible variation:

$$\nabla f(\boldsymbol{x}^*) \cdot \boldsymbol{d} = 0 \qquad \forall\ \boldsymbol{d} \in \mathcal{V}(\boldsymbol{x}^*)$$

- Find the set of non-regular points
- Choose the point with the lowest objective.

For example, consider the problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^3} \tfrac{1}{2}(x_1^2 + x_2^2 + x_3^2) \text{ s.t. } x_1 + x_2 + x_3 \le -3$$

The objective and constraints are continuously differentiable, and minima exist (by coerciveness). The Lagrangian is

$$\mathcal{L}\left(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}\right) = \tfrac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \mu(x_1 + x_2 + x_3 + 3)$$

The first-order conditions are

$$\begin{array}{ccc} \nabla_x \mathcal{L}\left(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) = \boldsymbol{0} & \Rightarrow & x_1^* + \mu^* = x_2^* + \mu^* = x_3^* + \mu^* = 0 \\ \boldsymbol{g}(\boldsymbol{x}^*) \le \boldsymbol{0} & \Rightarrow & x_1^* + x_2^* + x_3^* \le -3 \\ \mu_j^* g_j(\boldsymbol{x}^*) = 0 & \Rightarrow & \mu^*(x_1^* + x_2^* + x_3^* + 3) = 0 \end{array}$$

The solution is $\boldsymbol{x}^* = (-1,-1,-1)$ and $\mu^* = 1$ which satisfies $\mu \ge 0$. Furthermore, all points are regular, so this is the global minimum.

- ***Theorem (KKT Sufficient Conditions)***: Assume that f, $\boldsymbol{h}$ and $\boldsymbol{g}$ are twice continuously differentiable, and that $\boldsymbol{x}^* \in \mathbb{R}^n, \boldsymbol{\lambda}^* \in \mathbb{R}^m, \boldsymbol{\mu}^* \in \mathbb{R}^r$ satisfy

$$\begin{array}{ccc} \nabla_x \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \boldsymbol{0} & \boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0} & \boldsymbol{g}(\boldsymbol{x}^*) \le \boldsymbol{0} \\ \boldsymbol{\mu}^* \ge \boldsymbol{0} & & \mu_j^* = 0 \ \forall j \notin \mathcal{A}(\boldsymbol{x}^*) \\ \boldsymbol{d}^\top \nabla_{xx} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\boldsymbol{d} > 0 & & \forall \boldsymbol{d} \in \mathcal{V}^{\text{EQ}}(\boldsymbol{x}^*) \setminus \{\boldsymbol{0}\} \end{array}$$

Assume also that

$$\mu_j^* > 0 \qquad \forall j \in \mathcal{A}(\boldsymbol{x}^*)$$

Then $\boldsymbol{x}^*$ is a strict local minimum.

Proof: We follow the equality case. Suppose $\boldsymbol{x}^*$ is not a strict local minimum. Then the exists $\{\boldsymbol{x}_k\} \subset \mathbb{R}^n, \boldsymbol{h}(\boldsymbol{x}_k) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}_k) \le \boldsymbol{0}, \boldsymbol{x}_k \ne \boldsymbol{x}^*, \boldsymbol{x}_k \to \boldsymbol{x}^*$ with $f(\boldsymbol{x}_k) \le f(\boldsymbol{x}^*)$. We define

$$\boldsymbol{d}_k = \frac{\boldsymbol{x}_k - \boldsymbol{x}^*}{\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|} \qquad \delta_k = \left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|$$

Without loss of generality, assume $\boldsymbol{d}_k \to \boldsymbol{d}$. Using the same mean-value-theorem argument as in the sufficient conditions proof for Lagrange multipliers, we find that $\nabla \boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{d} = \boldsymbol{0}$.

$$\begin{aligned}(\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)\cdot \boldsymbol{x} &\geq (\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)\cdot \hat{\boldsymbol{x}}_k \\ &= (\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)\cdot(\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k) + (\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)\cdot \boldsymbol{x}_k \\ &\geq (\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)\cdot \boldsymbol{x}_k\end{aligned}$$

More succinctly

$$(\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)\cdot \boldsymbol{x} \geq (\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)\cdot \boldsymbol{x}_k = \text{constant} \qquad \forall \boldsymbol{x} \in \bar{\mathcal{C}}$$

In other words, $\bar{\mathcal{C}}$ lies on one side of each of those hyperplanes.

But now, set

$$\boldsymbol{\mu}_k = \frac{\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k}{\left\|\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k\right\|}$$

Then the equation above can be written as

$$\boldsymbol{\mu}_k \cdot \boldsymbol{x} \geq \boldsymbol{\mu}_k \cdot \boldsymbol{x}_k \qquad \forall k, x \in \mathcal{C}$$

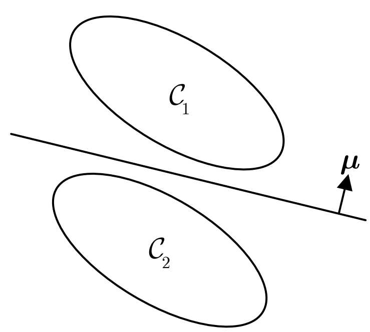
Since $\left\|\boldsymbol{\mu}_k\right\| = 1$, the sequence $\{\boldsymbol{\mu}_k\}$ is bounded has a non-zero subsequential limit $\boldsymbol{\mu}$. Letting $k \to \infty$, we get $\boldsymbol{\mu}_k \to \boldsymbol{\mu}$ and $\boldsymbol{x}_k \to \bar{\boldsymbol{x}}$ so

$$\boldsymbol{\mu}\cdot \boldsymbol{x} \geq \boldsymbol{\mu}\cdot \bar{\boldsymbol{x}} \qquad \forall k, x \in \bar{\mathcal{C}}$$

As required.

- ***Theorem (separating hyperplane)***: Let $\mathcal{C}_1, \mathcal{C}_2 \in \mathbb{R}^n$ be two disjoint non-empty convex sets. Then there exists a hyperplane that separates them; ie: a vector $\boldsymbol{\mu} \in \mathbb{R}^n, \boldsymbol{\mu} \neq 0$ and a scalar $b \in \mathbb{R}$ with

$$\boldsymbol{\mu}\cdot \boldsymbol{x}_1 \leq b \leq \boldsymbol{\mu}\cdot \boldsymbol{x}_2 \qquad \forall \boldsymbol{x}_1 \in \mathcal{C}_1, \boldsymbol{x}_2 \in \mathcal{C}_2$$



Proof: Consider the convex set

$$\mathcal{D} = \mathcal{C}_1 - \mathcal{C}_2 = \left\{x_1 - x_2 : x_1 \in \mathcal{C}_1, x_2 \in \mathcal{C}_2\right\}$$

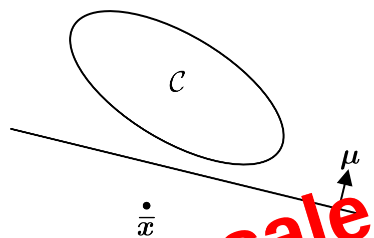
Since the two sets are disjoint, $\boldsymbol{0} \notin \mathcal{D}$. Thus, by the supporting hyperplane theorem, there exists a vector $\boldsymbol{\mu} \neq \boldsymbol{0}$ with

$$\boldsymbol{0} \leq \boldsymbol{\mu}^\top \left(\boldsymbol{x}_1 - \boldsymbol{x}_2\right) \qquad \forall \boldsymbol{x}_1 \in \mathcal{C}_1, \boldsymbol{x}_2 \in \mathcal{C}_2$$

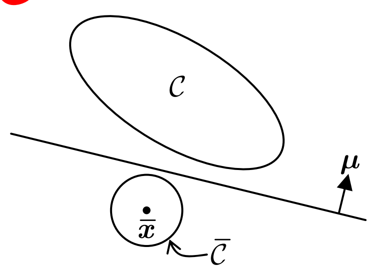
Setting $b = \sup_{\boldsymbol{x}_2 \in \mathcal{C}_2} \boldsymbol{\mu}^\top \boldsymbol{x}_2$, we obtain the desired result.

- ***Theorem (strictly separating hyperplane)***: Let $\mathcal{C} \subset \mathbb{R}^n$ be a *closed* convex set and $\bar{\boldsymbol{x}} \notin \mathcal{C}$ a point. Then there exists a hyperplane that *strictly* separates the point and the set. In other words, $\exists \boldsymbol{\mu} \in \mathbb{R}^n \setminus \{0\}, b \in \mathbb{R}$ such that

$$\boldsymbol{\mu} \cdot \bar{\boldsymbol{x}} < b < \inf_{\boldsymbol{x} \in \mathcal{C}} \boldsymbol{\mu} \cdot \boldsymbol{x}$$



Proof: Define $r = \min_{\boldsymbol{x} \in \mathcal{C}} \left\|\boldsymbol{x} - \bar{\boldsymbol{x}}\right\|$. This will be > 0 because $\mathcal{C}$ is closed, and $\bar{\boldsymbol{x}} \notin \mathcal{C}$. Now, let $\bar{\mathcal{C}} = \{\boldsymbol{x} \in \mathbb{R}^n : \left\|\boldsymbol{x} - \bar{\boldsymbol{x}}\right\| \leq r/2\}$. Clearly, $\mathcal{C}$ and $\bar{\mathcal{C}}$ are disjoint, so we can apply the separating hyperplane theorem. Diagramatically:



- ***Corollary***: If $\mathcal{C} \subsetneq \mathbb{R}^n$ is a closed convex set, then it is the intersection of all closed halfspaces that contain it.

Proof: Let $\mathcal{H}$ be the collection of all closed halfspaces containing $\mathcal{C}$, and let $\bar{H} = \bigcap_{H \in \mathcal{H}} H$.

Since $\mathcal{C} \neq \mathbb{R}^n$, the strictly separating hyperplane theorem implies that $\mathcal{H}$ is non-empty (since there is a point $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{x} \notin \mathcal{C}$) and clearly, $C \subset \bar{H}$.

Now, suppose there exists an $\boldsymbol{x} \in \bar{H}$ with $\boldsymbol{x} \neq \mathcal{C}$, then

$\inf_{\boldsymbol{x}\in\Omega} L(\boldsymbol{x},\boldsymbol{\mu})$ considers the highest intercept for planes *that contain the whole of $\mathcal{S}$ in one of their halfpsaces.*

Proof:

1. The hyperplane is the set $(\boldsymbol{z}, w)$ satisfying

$$\boldsymbol{\mu}\cdot\boldsymbol{z} + w = \boldsymbol{\mu}\cdot\boldsymbol{g}(\boldsymbol{x}) + f(\boldsymbol{x})$$

   Clearly, if $\boldsymbol{z} = \boldsymbol{0}$, we must have $w = L(\boldsymbol{x},\boldsymbol{\mu})$.

2. The hyperplane with normal $(\boldsymbol{\mu},1)$ that intercepts the axis at level c is the set $(\boldsymbol{z}, w)$ with

$$\boldsymbol{\mu}\cdot\boldsymbol{z} + w = c$$

   If $\mathcal{S}$ lies in the positive halfspace, then

$$L(\boldsymbol{x},\boldsymbol{\mu}) = \boldsymbol{\mu}\cdot\boldsymbol{g}(\boldsymbol{x}) + f(\boldsymbol{x}) \geq c \qquad \forall \boldsymbol{x}\in\Omega$$

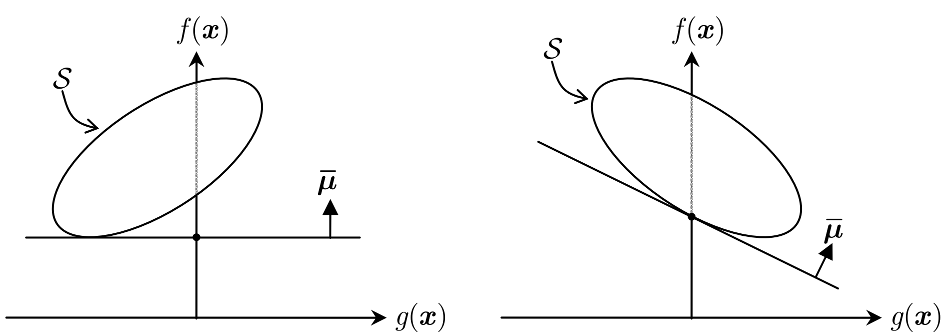
   Thus, the maximum intercept is $\inf_{\boldsymbol{x}\in\Omega} L(\boldsymbol{x},\boldsymbol{\mu})$.

3. We need to show that $f^* = \inf_{\boldsymbol{x}\in\Omega} L(\boldsymbol{x},\boldsymbol{\mu})$. It is obvious from part (2) that this is true if and only if among all hyperplanes with normal $(\boldsymbol{\mu}^*,1)$, the highest intersection with the vertical axis is at f^*.

- ***Theorem***: Let $\boldsymbol{\mu}$ be a geometric multiplier. Then $\boldsymbol{x}^*$ is a global minimum if and only if $\boldsymbol{x}^*$ is feasible and

$$\boldsymbol{x}^* \in \operatorname{argmin}_{\boldsymbol{x}\in\Omega} L(\boldsymbol{x},\boldsymbol{\mu}^*) \qquad \mu_j^* g_j(\boldsymbol{x}^*) = 0 \quad \forall 1 \leq j \leq r$$

Geometrically, the first statement is that $\boldsymbol{x}^*$ is, indeed, the value of $\boldsymbol{x}$ that minimizes L at that value of $\boldsymbol{\mu}$ (and recall that $f^* = \inf_{\boldsymbol{x}\in\Omega} L(\boldsymbol{x},\boldsymbol{\mu}^*)$), and the second is that *either* $\boldsymbol{x}^*$ is on the *boundary* of the feasible set (in which case we can "improve no further") or that the geometric multiplier is horizontal (in which case the minimum is attained on the interior of the feasible set).



(We do not need to check for regularity, since the constraints are linear).

Now, since $\mathcal{L}(\boldsymbol{x},\boldsymbol{\mu})$ is convex and $\nabla f(\boldsymbol{x}^*) + A\boldsymbol{\mu}^* = \boldsymbol{0}$, we have that

$$\boldsymbol{x}^* \in \operatorname{argmin}_{\boldsymbol{x}\in\mathbb{R}^n} \mathcal{L}(\boldsymbol{x},\boldsymbol{\mu}^*)$$

As such

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ f(\boldsymbol{x}) + \boldsymbol{\mu}^{*\top}(A\boldsymbol{x} - \boldsymbol{b}) \right\} = q(\boldsymbol{\mu}^*)$$

By weak duality, however, we have that

$$q(\boldsymbol{\mu}^*) \le q^* \le f^* \le f(\boldsymbol{x}^*)$$

However, by the previous statement, $f(\boldsymbol{x}^*) = q(\boldsymbol{\mu}^*)$, equality holds throughout, and so $f^* = q^*$.

- ***Extension to equality constraints***: The above trivially extends to linear equality constraints. More generally, consider the problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \text{ s.t. } A\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}$$

With Lagrangian

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top(A\boldsymbol{x} - \boldsymbol{b}) + \boldsymbol{\mu}^\top \boldsymbol{g}(\boldsymbol{x})$$

(With $\boldsymbol{\mu} \ge \boldsymbol{0}$ and $\boldsymbol{\lambda}$ unrestricted).

Then, provided that

  - There exists an optimal solution $\boldsymbol{x}^*$
  - f and $\boldsymbol{g}$ are continuously differentiable
  - There exists multipliers $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ satisfying the KKT conditions (ie: some sort of regularity).
  - f and $\boldsymbol{g}$ are convex over $\mathbb{R}^n$

Then there is no duality gap and geometric multipliers exists.

(Note that we are effectively requiring inequality constraints to be convex and equality constraints the be linear. One way to look at this requirement is as a requirement that $\mathcal{L}$ be convex. Indeed, since $\boldsymbol{\mu}$ is positive, $\boldsymbol{\mu}^\top \boldsymbol{g}(\boldsymbol{x})$ is convex for all $\boldsymbol{g}$. Since $\boldsymbol{\lambda}$ can take any value, however, it needs to multiply a linear function to retain convexity).

- ***Theorem (Slater's Condition)***: Consider the problem