Optimality condition for differentiable f_0.

if f_0 is differentiable, f_0 is convex

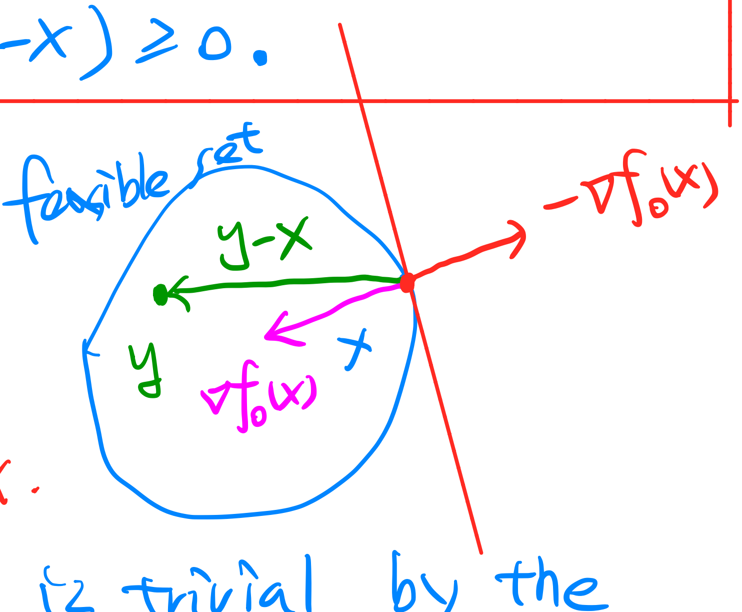
$\forall x, y \in \mathrm{dom}\, f_0$,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T (y-x)$$

(first-order condition).

$\boxed{\begin{array}{l} x \text{ is optimal if and only if} \\ x \text{ is feasible. and for all feasible } y \\ \nabla f_0^T(x)(y-x) \geq 0. \end{array}}$

If $\nabla f_0(x) \neq 0$, it means that $-\nabla f_0(x)$ defines a supporting hyperplane to the feasible set at x.



Proof. Sufficiency is trivial by the first order condition, (Necessity) Suppose x is optimal and $\exists\, y$ s.t. y feasible but $\nabla f_0(x)^T (y-x) < 0$.

Consider $z(t) = ty + (1-t)x$. $t \in [0,1]$.

$z(t)$ is feasible since x, y feasible.

Claim: for small t, $f_0(z(t)) < f_0(x)$

$$\frac{d}{dt} f_0(z(t))\Big|_{t=0} = \nabla f_0(x)^T \cdot \frac{d\, z(t)}{dt}\Big|_{t=0}.$$

$$z(0) = x \qquad = \nabla f_0(x)^T (y-x) < 0.$$

(by assumption).

$$\frac{d}{dt} f_0(z(t))\Big|_{t=0} < 0$$

($f_0(z(t))$ is decreasing at $t=0$).

Unconstrained problem with differentiable f_0.

x is optimal if and only if

$$x \in \mathrm{dom}\, f_0, \quad \nabla f_0(x) = 0.$$

Proof. Sufficiency is trivial by the first order condition.

(Necessity). Suppose x optimal.

$\forall$ feasible y. $\nabla f_0^T(x)(y-x) \geq 0$.

Since f_0 is differentiable, $\mathrm{dom}\, f_0$ is open.

so all y sufficiently close to x are feasible.

(no constraints).

Consider $\boxed{y = x - t\nabla f_0(x)}$, where $t \in \mathbb{R}$.

For small and positive t, y is feasible.

$$0 \leq \nabla f_0(x)^T(y-x) = \underbrace{-t}_{<0}\|\nabla f_0(x)\|_2^2$$

$\Rightarrow \nabla f_0(x) = 0.$

Example. $f_0(x) = \frac{1}{2}x^T P x + q^T x + r$. $P \in S^n$.

$0 = \nabla f_0(x) = Px + q \iff x$ is a minimizer

① If $q \notin \mathcal{R}(P)$, no solution.

$f_0(0) = r$. 0 is feasible.

In this case, $f_0(x)$ is unbounded below.

Consider $x = a \cdot q$. $a \in \mathbb{R}$. $q^T I q$

For $a = -\lambda_{max}(P)^{-1}$

$$f_0(x) = \frac{\lambda_{max}(P)^{-1}}{2} \boxed{q^T \left(\frac{P}{\lambda_{max}(P)}\right) q} - \frac{1}{\lambda_{max}(P)} \boxed{\|q\|_2^2} + r$$

$\preceq I$.

$$P' \triangleq \boxed{\frac{P}{\lambda_{max}(P)}} \preceq I$$

has largest eigenvalue 1.

② If $P \succ 0$, then there is a unique minimizer $x^* = -P^{-1}q$.

③ If P is singular, but $q \in \mathcal{R}(P)$, then the set of optimal points is

$-P^\dagger q + \mathcal{N}(P)$. $P(-P^\dagger q + v) = -q$, $v \in \mathcal{N}(P)$

where $P^\dagger$ denotes the pseudo-inverse of P.

$$P \geq 0 \qquad P = \sum_i \lambda_i v_i v_i^T$$

$$v_i^T v_j = \delta_{ij}.$$

$$P^{\dagger} = \sum_i \frac{1}{\lambda_i} v_i v_i^T.$$

$$P^{\dagger} P = \sum_j \frac{1}{\lambda_j} \lambda_i v_j \underbrace{v_j^T v_i}_{\delta_{ij}} v_i^T.$$

$$= \sum_i v_i v_i^T. \left(= \mathbb{1}. \text{ if } P \text{ is non singular.} \right)$$

If P is singular.

$\{v_i\}$ does not form a basis for $\mathbb{R}^n$.

say $\{v_i\} \sqcup \{w_j\}$ is a basis.

$$P^{\dagger}P = \sum_i v_i v_i^T + \sum_j 0 \cdot w_j w_j^T.$$

$$\mathbb{1} = \sum_i v_i v_i^T + \sum_j w_j w_j^T.$$

$$P^{\dagger}P \leq \mathbb{1}.$$

$$q \in R(P) \Rightarrow q = \sum_i a_i v_i$$

$$P^{\dagger}P q = \sum_i v_i v_i^T \sum_j a_j v_j = q.$$