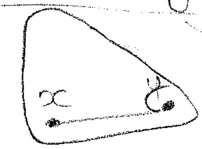
02/13/18 Constrained optimization.
Convexity, Lagrangian, duality.

Def A set S is convex if
$x, y \in S \Rightarrow \underbrace{\theta x + (1-\theta) y}_{,} \in S.$ $\forall \theta \in [0,1]$

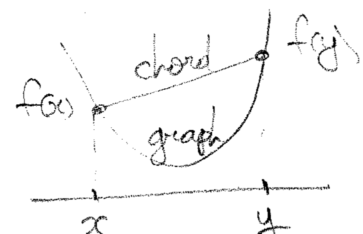
called convex combination



Def $f: \mathbb{R}^n \to \mathbb{R}$ is convex if
$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$

(p. 67 in Boyd-Vandenberghe)

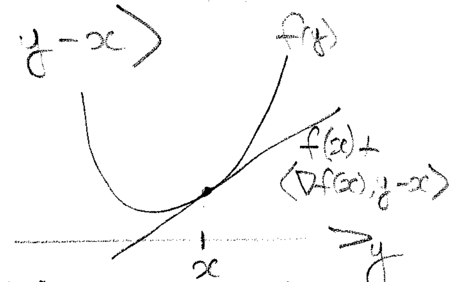
$\forall x, y$, $\theta \in [0,1]$



Prop. A differentiable $f: \mathbb{R}^n \to \mathbb{R}$ is convex iff

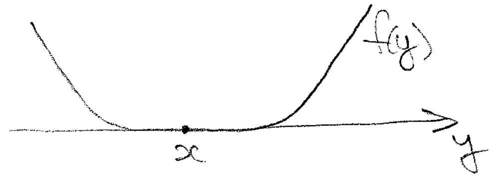
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$$

$\forall x, y$



$\to$ tangent line is a global underestimator

Prop. x is a minimizer of a differentiable, convex f iff $\nabla f(x) = 0$.



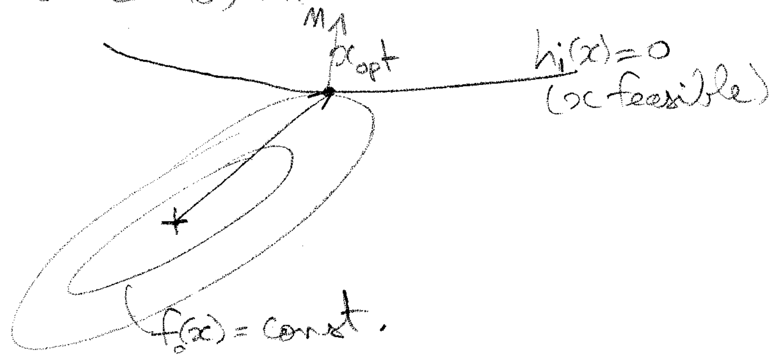
(= global minimizer).

Constrained optimization

$\min \|x\|$ s.t. $Ax = b$

$\rightarrow \min f_0(x)$ s.t. $h_i(x) = 0$ ex. $h_i(x) = a_i^T x - b_i$

assume $f_0, h_i \in C^1$



at $x = x_{opt}$, normal to $f_0(x) = f_0(x_{opt})$

is also normal to all $h_i(x) = 0$

$\nabla f_0 \in \text{Span}\{\nabla h_i\}$.

$$\begin{cases} \nabla f_0 + \sum v_i \nabla h_i = 0 & (1) \\ h_i = 0 & (2) \end{cases}$$

$$\mathcal{L}(x, v) = f_0(x) + \sum v_i h_i(x)$$

$$\nabla_x \mathcal{L} = 0 \iff (1)$$
$$\frac{\partial \mathcal{L}}{\partial v_i} = 0 \iff (2).$$

(1) may determine v as well.
It is a linear system for them
(watch for singular matrix).

$\mathcal{L}$ is the Lagrangian

ν are Lagrange multipliers

(1), (2) are the Karush-Kuhn-Tucker (KKT) conditions, a.k.a. first-order optimality conditions

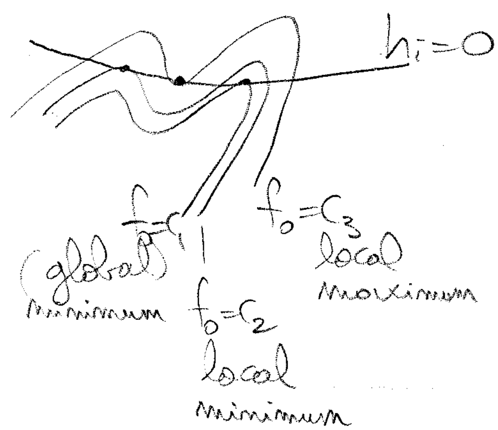
ex. ULS. $\mathcal{L}(x,\nu) = \frac{1}{2}\|x\|^2 - \nu^T(Ax-b)$

$$\nabla_x \mathcal{L} = x^T - \nu^T A = 0 \Rightarrow x = A^T\nu$$
$$\nabla_\nu \mathcal{L} = Ax - b = 0 \Rightarrow Ax = b$$

then $AA^T\nu = b$, $\nu = (AA^T)^{-1} b$
$x = A^T(AA^T)^{-1} b$

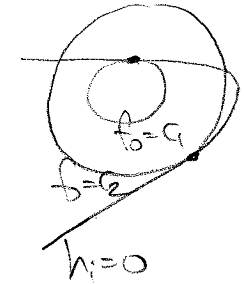
the KKT conditions are necessary
(x_{opt} minimizer $\Rightarrow$ (1), (2)
argue by contradiction)
but they are not in general sufficient
((1), (2) $\overset{?}{\Rightarrow}$ x_{opt} minimizer).

ex 1. (level sets of f_0 nonconvex)



$c_1 < c_2 < c_3$

ex 2.



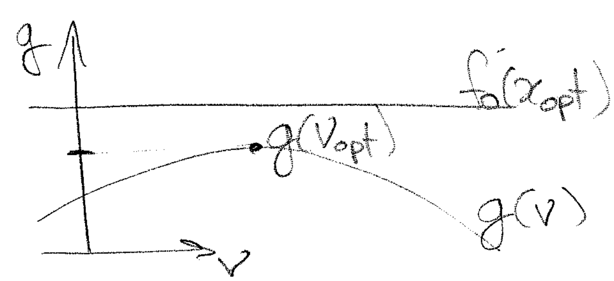
$\boxed{Q.}$ When is KKT sufficient?

Would like to recognize (1) of KKT as a condition of minimization of $\mathcal{L}$ over x.

Def. $g(\nu) = \min_x \mathcal{L}(x, \nu)$

Note $g(\nu) \le \mathcal{L}(\tilde{x}, \nu)$ & pick $\tilde{x}$ feasible : $h_i(\tilde{x}) = 0$
$= f_0(\tilde{x})$
true for all feasible $\tilde{x}$
$\Rightarrow g(\nu) \le f_0(x_{opt})$

true for all ν
$\Rightarrow$
$\max g(\nu) =$
$\boxed{g(\nu_{opt}) \le f_0(x_{opt})}$



so $g(\nu_{opt})$ provides a floor on the optimal value of the objective $f_0(x)$.

$\max_\nu g(\nu)$ is the dual problem
ν dual variables, ν_{opt} dual optimal
$g(\nu)$ is the Lagrangian dual function
$\min_x f_0(x)$: $h_i(x) = 0$ is the primal problem
x : primal variables, x_{opt} primal optimal

$f_0(x_{opt}) - g(\nu_{opt})$: <u>duality gap</u>
$f_0(x_{opt}) \geq g(\nu_{opt})$: weak <u>duality</u>
$f_0(x_{opt}) = g(\nu_{opt})$. strong duality
(holds / obtains)

<u>thm.</u> Assume: $\begin{cases} f_0(x) \text{ is differentiable, convex} \\ h_i(x) = a_i^T x - b_i \text{ is affine} \end{cases}$

(p. 244 in Boyd-Vandenberghe)

Then KKT are sufficient
($(\tilde{x}, \tilde{\nu})$ obey (1), (2) $\Rightarrow$ they are optimal)
and strong duality holds.

<u>Pf.</u> Let $(\tilde{x}, \tilde{\nu})$ obey (1), (2)
(2) $\Rightarrow$ $\tilde{x}$ is feasible.
Consider $\mathcal{L}(x, \tilde{\nu}) = f_0(x) + \sum_i \tilde{\nu}_i h_i(x)$
It is also a convex & differentiable function
(1) means $\nabla_x \mathcal{L}(x, \tilde{\nu}) = 0$ when $x = \tilde{x}$
$\Rightarrow$ $\tilde{x}$ is a minimizer of $\mathcal{L}(x, \tilde{\nu})$.

$$g(\tilde{\nu}) = \min_x \mathcal{L}(x, \tilde{\nu})$$
$$= \mathcal{L}(\tilde{x}, \tilde{\nu}) \quad \text{because } \tilde{x} \text{ is a minimizer}$$
$$= f_0(\tilde{x}) \quad \text{because } \tilde{x} \text{ is feasible}$$

$\Rightarrow$ zero duality gap
Lower bound is attained for f_0
$\Rightarrow$ $\tilde{x}$ is primal optimal
Upper bound is attained for g
$\Rightarrow$ $\tilde{\nu}$ is dual optimal.

$\square$

<u>Rmk:</u> We say $g(\tilde{\nu}) = f_0(\tilde{x})$ is a <u>certificate</u> of <u>optimality</u> of $\tilde{x}$ and $\tilde{\nu}$.

<u>Rmk:</u> $g(\nu)$ is the min of affine functions
$\Rightarrow$ it is con<u>cave</u>