Note: the treatment I give of the KKT conditions roughly follows that in Chapter 10 (sections 1-3) of Lauritzen's book Undergraduate Convexity. But, I really only focus on the convex case, which is simpler. Most treatments of the KKT conditions, and Lagrange multipliers, begin with the case of general functions and general (or convex) constraints, and then observe that the convex case is special. Treatments of this form appear in Chapter 5 of BV, and Chapter 12 of LW. The treatment in LW is probably closer to the one I give here.

In the next class, we will introduce Lagrange multipliers and duality.

I am going to skip most proofs this week so that I can instead cover more examples.

A general constrained convex optimization problem is

$\min f(x)$ such that $g_i(x) \le 0$, for $1 \le i \le k$, $x \in \mathbb{R}^n$,

and $f, g_1, \ldots, g_k$ are convex functions.

Our goal for today is to learn how to certify a solution. That is, the KKT (Karush-Kuhn-Tucker) conditions.

We covered the case of $k=0$ (unconstrained)

$x_* \in \arg\min_x f(x)$ iff $(\nabla f)(x_*) = \bar{0}$.

For continuous, differentiable f, the following condition is very helpful.

<u>Lemma 1</u> Let C be a closed convex set (e.g. $\{x : g_i(x) \le 0, \forall i\}$), and let f be differentiable on C.

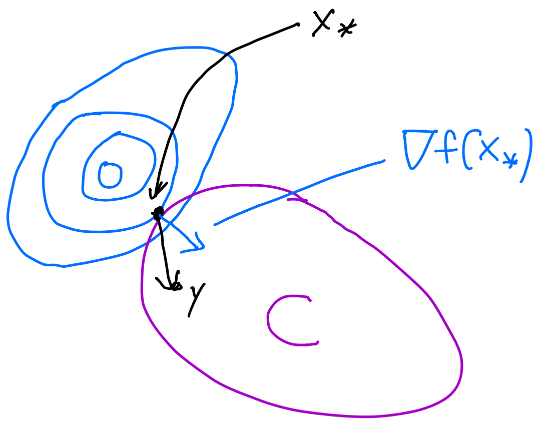
Then $x_* \in \arg\min_{x \in C} f(x)$ (1)

implies $\nabla f(x_*)^T (y - x_*) \ge 0 \quad \forall y \in C.$ (2)

If f is convex, then also (2) $\Rightarrow$ (1).

Note if x_* is a global minimizer, $\nabla f(x_*) = 0$ in which case (2) must hold.

<u>Picture</u> level sets of f



<u>proof</u> (1) => (2). Prove not(2) => not(1).

If $\exists\, y \in C$ s.t. $\nabla f(x_*)^T (y - x_*) < 0$,
then moving in direction $y - x_*$ decreases f,
but stays inside C.

For small δ, $f(x_* + \delta) \approx f(x_*) + \nabla f(x_*)^T \delta$.

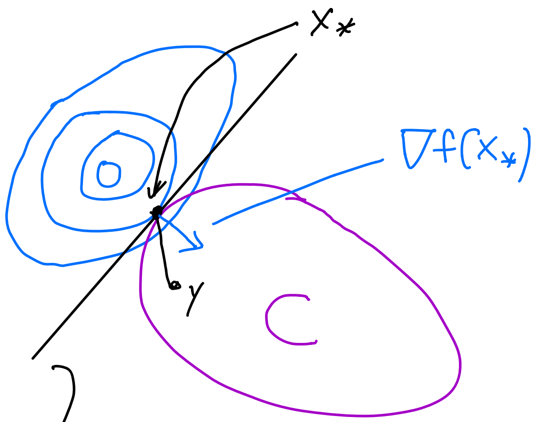
Setting $\delta = \varepsilon (y - x_*)$ for $\varepsilon > 0$ but small gives
$$f(x_* + \delta) < f(x_*)$$

As C is convex, $x_* + \delta = (1-\varepsilon) x_* + \varepsilon y \in C$,
contradicting (1)

(2) and f convex => (1)

Recall convexity => $f(y) \geq f(x_*) + \nabla f(x_*)^T (y - x)$

(2) => $y \in C$ => $\nabla f(x_*)^T (y - x_*) \geq 0$, so $f(y) \geq f(x_*)$, $\forall y \in C$.



$H = \{y : \nabla f(x_*)^T y = \nabla f(x_*)^T x_*\}$

is a hyperplane that separates C from $\{y : f(y) \le f(x_*)\}$

We will now do some examples.

---

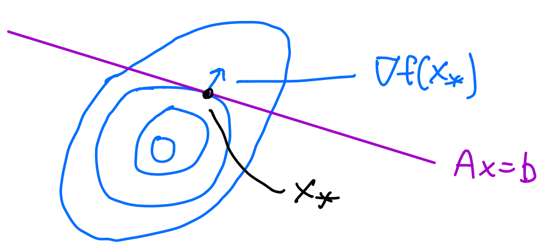
1. For a convex function f, consider the problem

$\min f(x)$ s.t. $Ax = b$, where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^k$, $k < n$

A is k-by-n.

(can think of as $Ax \le b$ and $Ax \le -b$)

$Ax = b$ defines an $(n-k)$ dimensional hyperplane, so think of it like this



At the solution x_*, we have $\nabla f(x_*)^T(y - x_*) \geq 0$ for all y s.t. $Ay = b$.

As $A(2x_* - y) = 2b - b = b$,

$$\nabla f(x_*)^T(2x_* - y - x_*) \geq 0$$

$$\iff \nabla f(x_*)^T(x_* - y) \geq 0$$

$$\iff \nabla f(x_*)^T(y - x_*) = 0, \quad \forall y \text{ s.t. } Ay = b$$

This is $A(y - x_*) = 0 \Rightarrow \nabla f(x_*)^T(y - x_*) = 0$

So, for all z s.t. $Az = 0$, $\nabla f(x_*)^T z = 0$

$$\iff \nabla f(x_*) \in \text{row-span}(A)$$

(if $g \notin$ row-span(A), $\exists z$ s.t. $Az = 0$ but $g^T z \neq 0$)

That is, $\exists \nu$ ("nu") $\in \mathbb{R}^k$ s.t. $\nabla f(x_*) = A^T \nu$

x_* is the solution iff $Ax_* = b$

<u>and</u> $\exists \nu_*$ s.t. $\nabla f(x_*) + A^T \nu_* = 0$

$$(\nu_* = -\nu)$$

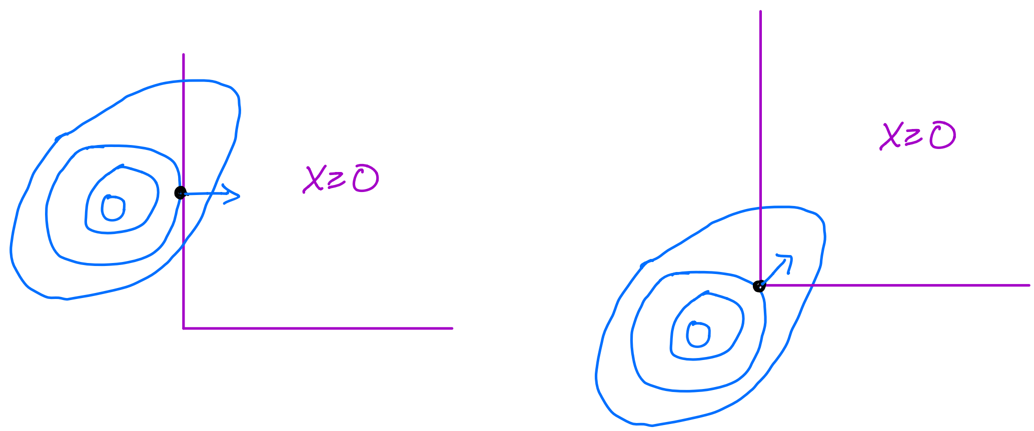
I'm introducing this odd notation to be consistent with what comes later.

2. Consider $\min f(x)$ s.t. $x \in \mathbb{R}^n_+ \overset{\text{def}}{=} \{x : x \in \mathbb{R}^n \text{ and } x \geq 0\}$

That is, we have constraints $x(i) \geq 0$, $\forall i$

By lemma 1, x_* is optimal iff $x_* \geq 0$ and $\forall y \geq 0$, $\nabla f(x_*)^T (y - x_*) \geq 0$.

For convenience, I will write $\nabla = \nabla f(x_*)$.



$$\left.\begin{array}{l} \bar{0} \in \mathbb{R}^n_+ \Rightarrow \nabla^T(-x_*) \geq 0 \Rightarrow \nabla^T x_* \leq 0 \\ 2x_* \in \mathbb{R}^n_+ \Rightarrow \nabla^T(2x_* - x_*) \geq 0 \Rightarrow \nabla^T x_* \geq 0 \end{array}\right\} \nabla^T x_* = 0 \quad (1)$$

So, $\forall y \in \mathbb{R}^n_+$, $\nabla^T(y - x_*) \geq 0 \Rightarrow \nabla^T y \geq 0$, $\forall y \in \mathbb{R}^n_+$

$$\Rightarrow \nabla \geq 0 \quad (\text{if } \nabla(i) < 0,\ \nabla^T e_i < 0,\ \text{※}) \qquad (2)$$

Conversely, (1) and (2) $\Rightarrow \forall y \in \mathbb{R}^n_+$,

$$\nabla f(x_*)^T (y - x_*) = \nabla f(x_*)^T y \geq 0$$

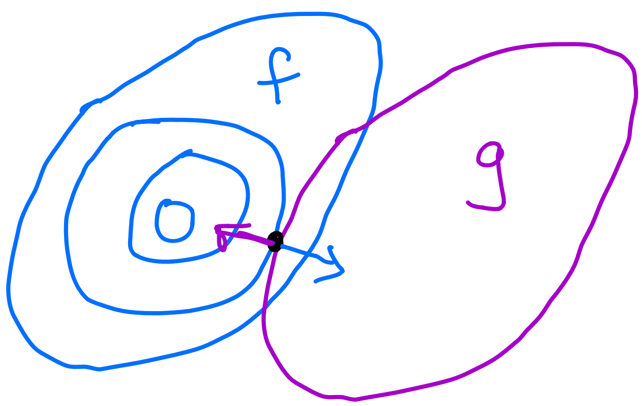
So, $x_* \in \mathbb{R}^n_+$ is optimal iff

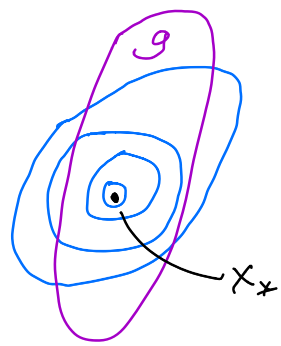
$$\nabla f(x_*) \geq 0 \quad \text{and} \quad \nabla f(x_*)^T x_* = 0$$

Complementary slackness: $x_*(i) \cdot \nabla f(x_*)(i) = 0 \ , \ \forall i$

---

3. $\min f(x)$ s.t. $g(x) \leq 0$,
   where f and g are differentiable and convex.





x_* is optimal iff $\qquad g(x_*) < 0$ and $\nabla f(x_*) = \bar{0}$

or $\quad g(x_*) = 0$ and $\nabla f(x_*) = -\lambda \nabla g(x_*)$ for some $\lambda > 0$.

First, understand condition $\nabla f(x_*) = -\lambda \nabla g(x_*)$, $\lambda > 0$

Use tight case for Cauchy-Schwartz: $|u^T v| \leq \|u\| \cdot \|v\|$
with equality only when $u = \lambda v$, some λ.

<u>Claim</u> For nonzero vectors u and v,

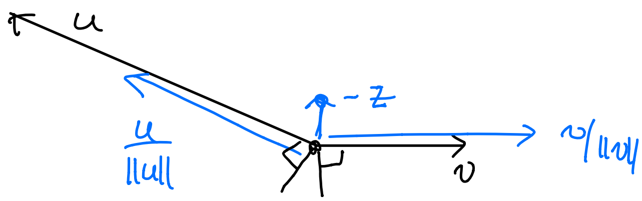
$u = -\lambda v$ for some $\lambda > 0$

<u>iff</u> there does not exist z s.t. $u^T z < 0$ and $v^T z < 0$

<u>proof</u> if $u = -\lambda v$, then $\text{sign}(u^T z) = -\text{sign}(v^T z)$

otherwise,

consider $z = -\left(\frac{u}{\|u\|} + \frac{v}{\|v\|} \right)$



$$u^T \left(\frac{u}{\|u\|} + \frac{v}{\|v\|} \right) = \|u\| + \frac{u^T v}{\|v\|} \geq \|u\| - \|u\|$$

by Cauchy-Schwartz.

And, $\frac{u^T v}{\|v\|} = \|u\|$ only if u is a multiple of v.

<u>proof of optimality condition</u>

If $\exists$ a direction z s.t. $\nabla f(x_*)^T z < 0$

<u>and</u> $\nabla g(x_*)^T z < 0$

Then for small $\varepsilon > 0$, $f(x_* + \varepsilon z) < f(x_*)$

and $g(x_* + \varepsilon z) < g(x_*) \leq 0$,

contradicting the optimality of x_*.

That is, x_* optimal $\Rightarrow \nexists$ such z

$\Rightarrow \nabla f(x_*) = -\lambda \nabla g(x_*)$

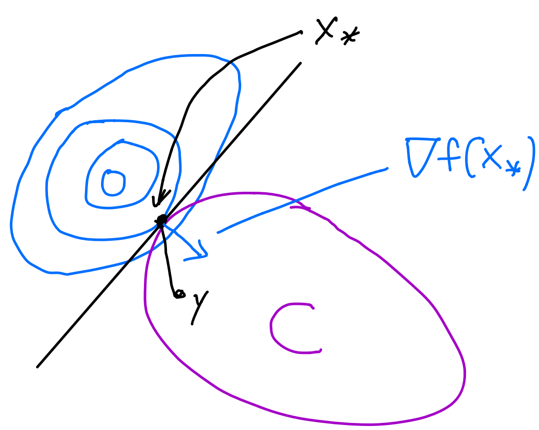
Conversely, if $\nabla f(x_*) = -\lambda \nabla g(x_*)$ and $g(x_*) = 0$

consider hyperplane $\mathcal{H} = \{y : \nabla^T y = \nabla^T x_*\}$

convexity of f and g tells us f is bigger on

one side and g is bigger on the other.

We write this as $\nabla f(x_*) + \lambda g(x_*) = 0, \quad \lambda \geq 0$



The KKT conditions for (see Lauritzen Def. 10.4)

$$\min f(x) \quad \text{s.t. } g_1(x) \le 0, \dots, g_k(x) \le 0$$

for $x_* \in \mathbb{R}^n$ and $\lambda_* \in \mathbb{R}^k_+$

are

1. $g_i(x_*) \le 0$ for $1 \le i \le k$
2. $\lambda_*(i)\, g_i(x) = 0$ for all i
3. $\nabla f(x_*) + \sum_i \lambda_*(i) \nabla g_i(x_*) = \bar{0}$

KKT conditions certify optimality (sufficient conditions)

Theorem 1 If f and $g_1, \dots, g_k$ are convex and x_* and λ_* satisfy (1), (2), and (3), then x_* is an optimal solution.

In many nice problems, these certificates always exist (necessary)

Theorem 2 If f and $g_1, \dots, g_k$ are convex and $\exists$ strictly feasible x_0 : $g_i(x_0) < 0 \ \forall i$ (Slater's condition) then $\exists\, x_*$ and λ_* that satisfy (1), (2), and (3).

Theorem 3 If each $g_i(x)$ is a linear function,

$$g_i(x) = a_i^T x + b_i$$

then $\exists\, x_*$ and λ_* that satisfy (1), (2), and (3).

We will skip the proofs.