Lecture 9

- HW4 posted!

Necessary + Sufficient Condition for linear inequality constraints
min $f(x)$
st. $Ax \ge b$

- Gradient Descent + Steepest Descent + (Line Search)
- Converges on a simple quadratic!

Necessary Conditions for
min $f(x)$
s.t. $Ax \ge b$ ①

This time

$$\left.\begin{array}{l} g(x^*) = A^T \lambda^* \\ \lambda^* \ge 0 \\ \lambda^{*T}(Ax - b) = 0 \\ Ax^* \ge b \end{array}\right] ②$$

Last time

$g(x^*) = \bar{A}^T \bar{\lambda}^*$
$\bar{\lambda}^* \ge 0$
$\bar{A}x^* \ge \bar{b}^*$
$\bar{A}, \bar{b}$ based on the active set!

What the key difference!
There is no active set!
· uses full A, λ^*

But we add
$(\lambda^*)^T(Ax - b) = 0$
<=>
$\sum \lambda_i (a_i^T x - b) = 0$
$\lambda_i \ge 0$.

This is called a COMPLEMENTARITY Condition

Since $a_i^T x - b \ge 0$ (Const.)
$\lambda_i \ge 0$

If $a_i^T x - b_i > 0$
=> non-active Constraint
=> $\lambda_i = 0$.
Also, if $\lambda_i > 0$
Then $a_i^T x - b_i = 0$
=> active constraint.

To go from Last time to ②.
Let $\lambda_i = 0$ for all inactive constraints.
$A = \begin{bmatrix} \bar{A} \\ A_I \end{bmatrix}$

$\bar{A}^T \bar{\lambda}^* = g(x^*)$
↑
based on act. set

$$\begin{bmatrix} \bar{A}^T & A_I^T \end{bmatrix} \underbrace{\begin{bmatrix} \bar{\lambda}^* \\ 0 \end{bmatrix}}_{\lambda^*} = g(x^*)$$

So we get
$A^T \lambda^* = g(x^*)$
in ②.

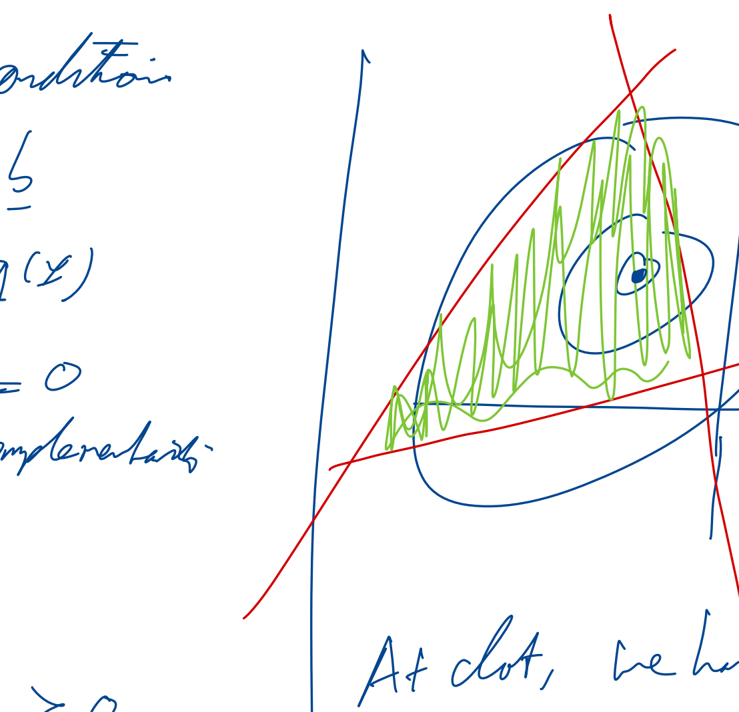
A problem for sufficient condition is to read off the active set.
We can have $\lambda_i = 0$ for an active Constraint! ☹
$a_i^T x - b_i = 0$ AND
$\lambda_i = 0$.

For suff. conditions, we assume Strict Complementarity.
$\lambda_i = 0$ xor $a_i^T x - b_i = 0$
i.e. only 1 is true! <=>
active set <=> $\lambda_i > 0$.

Sufficient Condition
$Ax \ge b$
$A^T \lambda = g(x)$
$\lambda^T (Ax - b) = 0$
w/ Strict Complementarity
$\lambda \ge 0$

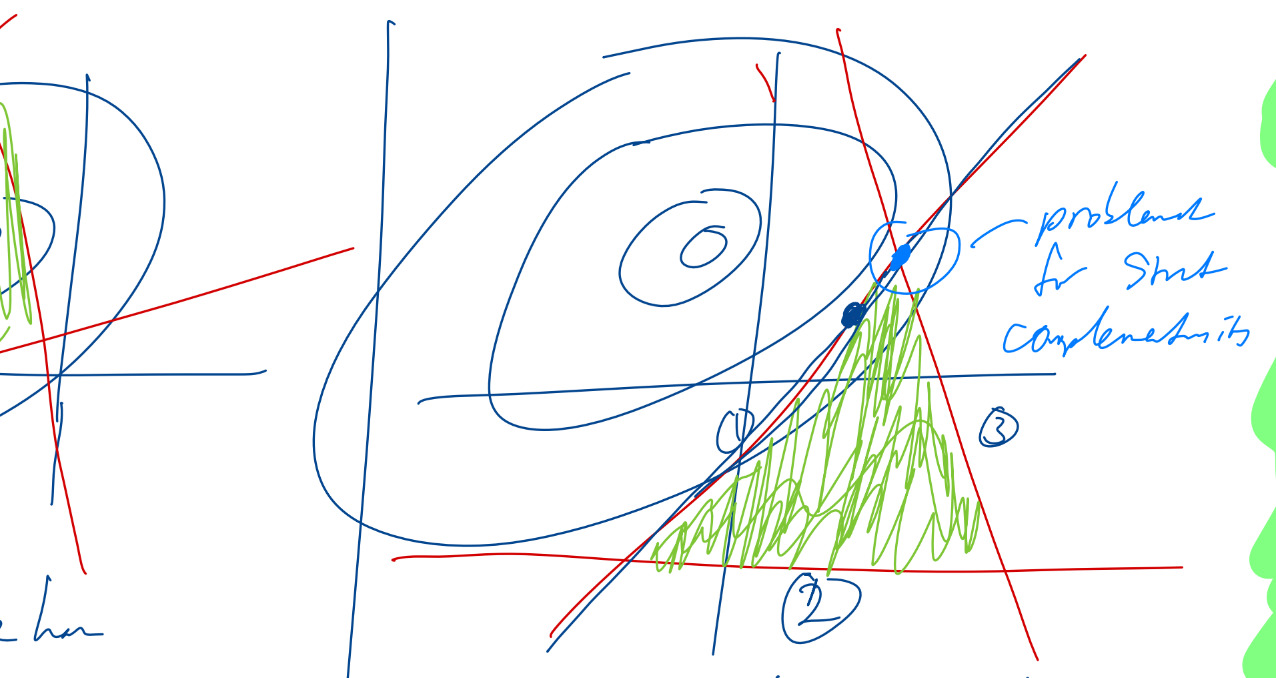
$\bar{Z}^T H(x) \bar{Z} \succ 0$
$\bar{Z}$ is the null space of the active set.

All of this is LOCAL!



At dist, we have
$Ax \ge b$,
$A^T \lambda = g(x) = 0$

$\bar{Z} = I$ => $H(x) \succ 0$.



At dist, we have $Ax \ge b$
We have $\lambda_1 > 0$ for active Constraint 1,
$A^T \lambda = g(x)$ $\lambda \ge 0$ too!
$\lambda^T(Ax - b) = 0$ w/ Strict Complementarity
Hessian along ① is pos. def.

Gradient Descent or Steepest Descent!
It's good! Great!
↳ It's "fast"
↳ low-ish memory
↳ good theory
↳ guarantees.
↳ No Hessian.
↳ Lots of Variants
· SGD
· Adam, } for separable functions
etc.

Today:
It's slow,
and Even "optimized"
it's still slow.

On a quadratic objective.

Scenario:
min $f(x)$
$f: \mathbb{R}^n \to \mathbb{R}$.
$x_k \to x_{k+1}$

Idea:
$f(x_k + p) = f(x_k) + p^T g(x_k) + O(\|p\|^2)$

Let's say $\|p\|$ is small!
Then $p = -g(x)$
Makes $f(x_k + p)$ go down as much as possible!

Alg: Let x_0 be arbitrary

$$x_{k+1} = x_k + \alpha_k \underbrace{(-g(x_k))}_{\text{neg. gradient}}$$

Step length!

New Notation
$g_k = g(x_k)$

Loop
$x_{k+1} = x_k - \alpha_k g_k$

Need to pick α_k!
- Constant
- Some "hyper parameter scheme"
  Small → big
  big → small
- Based on f (if known)
- As far as possible
  $\alpha_k = \operatorname{argmin} f(x_k - \alpha g_k)$
- norm of g_k

The Problem to Analyze
min $\frac{1}{2} x^T Q x$
$Q = Q^T$, $Q \succ 0$
(Quadratic, 1 soln.)

We will show
$f(x_{k+1}) < f(x_k)(1 - c)$
for some $0 < c < 1$
=> $f(x_k) \to 0$
=> We solve this at (at least) a linear rate!

Approach

1. $\alpha_k = \dfrac{\|g_k\|^2}{g_k^T Q g_k}$
gives the Exact line search step. length!

$\alpha_k = \operatorname{argmin} f(x_k - \alpha g_k)$
(for $f(x_k) = \frac{1}{2} x_k^T Q x_k$)
§3 on handout.

2. Then we show

$$f(x_{k+1}) = f(x_k) - \underbrace{\frac{1}{2} \frac{\|g_k\|^4}{g_k^T Q g_k}}_{\text{looks like algebra Convergence ☹}} \rightarrow$$

We use/show
$f(x_k) = \frac{1}{2} g(x)^T Q^{-1} g(x)$
or $f(x_k) = \frac{1}{2} g_k^T Q^{-1} g_k$
(b/c) $g_k = Q x_k$.

$$f(x_{k+1}) = f(x_k) \cdot \left[1 - \frac{\|g_k\|^4}{g_k^T Q g_k \cdot g_k^T Q^{-1} g_k} \right]$$

3 We want to show

$$c < \frac{\|g_k\|^4}{g_k^T Q g_k \cdot g_k^T Q^{-1} g_k} < d < 1$$