Lecture 32

- Theorem 3.2 from Nocedal & Wright
(convergence of line search with strong wolfe)

- A practical Newton iteration when the Hessian is indefinite
- Bounded Perturbation
- Modified Cholesky

Logistics:
- Report 1 due Monday
- We will flip class starting Monday.
- Weekly lectures (recorded with 2x 75 min vs. 3x 50 min)
- Q&A during class periods (can be shorter than 50 mins)
- Still recorded, etc.
- Need to come to class at least 1x / week to submit question.

Back to Line Search (Nocedal & Wright)

Assume $f(x)$

Assume $g(x)$ is available

Alg: while $\|g_k\| \geq$ tol
find p_k in direction.
find α_k to satisfy
Strong Wolfe
set $x_{k+1} = x_k + \alpha_k p_k$

If $\boxed{\text{Lim}_{k\to\infty} \frac{\cos(\theta_k)}{\|p_k\|} = 0}$ then

$\|p_k\| \to 0$

via Backtrack LS.

Using Strong Wolfe.
Theorem 3.2 in N&W.
- Lots of assumptions
- f bounded below
- f differentiable
- ∇f Lipschitz

Then $\sum_{k=0}^{\infty} \cos^2(\theta_k)\|p_k\|^2 < \infty$

$\cos(\theta_k) = \frac{-g_k^T p_k}{\|g_k\|\|p_k\|}$

So if $\cos^2(\theta_k) \geq \rho$
then $\|p_k\| \to 0$.

Called Zutendijk condition.

Newton for Opt

"Simple idea"

$p_k = -H_k^{-1} g_k$

Why?

$f(x_k + p) \approx M_k(p) =$
$f_k + p^T g_k + \frac{1}{2} p^T H_k p$

Quiz: What is
$\min_p M_k(p)$

(or argmin of you prefer)

If $H_k \succ 0$, then all fine.
Convex.
So $\frac{\partial M_k}{\partial p} = g + Hp = 0$
$p = -H^{-1} g$

If H_k is indef.
$M_k(p) \to -\infty$

If $H_k \succeq 0$, then there is
$M_k(p)$ Solve $(H_k + \lambda I)p = -H^{-1} g$
λ is a Lagrange multiplier to enforce $\|p\| \leq \Delta$

In Newton:
if $H_k \succ 0$, the
good!
But... What if not?
1. Make it work!
2. Express a step
Stay Backtrack
(if not the Search
$\alpha = 1$ good)
3. Change
step that dir.
4. use grad dir.
5. pick a different model!

If H_k is not pos def.
then we can pick
$C_k \succ 0$
$B_k = H_k + C_k$
is pos def.
then use $p = -B_k^{-1} g$
as a search dir.

How to find C_k?
if $B_k = H_k + \tau I$
if $\tau \geq |\lambda_{min}(H_k)|$
then $B_k \succ 0$.
1. A is a sym def
$A = PP^T$ for
some P
2. Cholesky factorization will
give you P!

? We can
do Bisection
to find τ

Idea 2

Use Chol Cholesky
find U
A is not sym

$A = \begin{bmatrix} a & g^T \\ g & B \end{bmatrix}$

Then a Cholesky factor
looks like:

$= \begin{bmatrix} 1 & 0 \\ l & R \end{bmatrix} \begin{bmatrix} \gamma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & l^T \\ 0 & R^T \end{bmatrix}$

$a = \gamma$, $l = g/\gamma$

$RDR^T = B - gl^T$

If A is sym
then $\gamma \geq 0$
and all $D_{ii} \geq 0$.

If $\gamma \leq 0$
use $\gamma = 0$
as given by Fix!

Then
Cholesky of A fails.
$A + E_k$

Theory

If $B_k = H_k + E_k$

and $\|B_k\|_2 \|B_k^{-1}\|_2 \geq k$, $\square$
then $\frac{-g_k^T p_k}{\|g_k\|\|p_k\|} \geq \frac{1}{k}$

$\cos(p_k, g_k) \geq \frac{1}{k}$

$\cos(\theta_k, g_k)$

$\frac{g_k^T p_k}{\|g_k\|\|p_k\|} = \frac{1}{\kappa(M_k)}$ with $M_k \succ 0$
$M_k p_k = -g_k$

$\frac{1}{\kappa(M)} = \frac{\lambda_{min}}{\lambda_{max}}$

$\frac{g^T p}{\|g\|\|p\|} = \frac{g^T M g}{g^T g \cdot \|p\|} = \frac{\|Mg\|}{\|g\|} \geq \frac{\lambda_{min} \cdot \|Mg\|}{\|Mg\|} \geq \frac{\lambda_{min}}{\lambda_{max}}$

$\frac{\|Bx\|}{\|x\|} \geq \frac{\|M^{-1}Bx\|}{\|x\|} \geq \frac{1}{\lambda_{max}} \Rightarrow$ $\boxed{k}$