Math 408 Review

① Modeling problems as NLPs: examples from §1, examples done in class.

② Fundamentals: (§2.1)
- functions in n variables: graphs, level sets,
- equations and inequalities: - drawing feasible regions
- matrices, operations with matrices - $+$, $\cdot$, inverse, transpose scalar mult^{n}.
- vectors - dot products, norm, properties of norm (pp. 7)

③ Vocabulary of NLPs (§2.2)
standard form, cost function, constraints, feasible region, feasible solution, optima, local & global optima, unconstrained NLP, linear program, quadratic program

④ Solving NLPs in 1 and 2 variables (§2.3, §2.4)
(Look at all problems worked out in the notes and in the exercises)

⑤ Local & Global Optima
- continuity of functions (Defn in terms of converging sequences)
- differentiability of functions
- gradient $\nabla f(x)$
- Hessian $Hf(x)$

Chain rule
$$\frac{d}{dt} f(x_1(t), \ldots, x_n(t)) = \frac{\partial f}{\partial x_1}\frac{d x_1(t)}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{d x_n(t)}{dt}$$

Lemma 3.1.1 $f: \mathbb{R}^n \to \mathbb{R}$, $x^*, \omega \in \mathbb{R}^n$. Let
$$\phi(t) := f(x^* + t\omega) \quad t \in \mathbb{R}$$
If $\frac{\partial f}{\partial x_i}$ conts on $\mathbb{R}^n$ $\forall i$, then $\phi'(t) = \nabla f(x^* + t\omega)^T \omega$ $\forall t \in \mathbb{R}$

If $\frac{\partial f}{\partial x_i \partial x_j}$ conts on $\mathbb{R}^n$ $\forall i,j$ then $\phi''(t) = \omega^T Hf(x^* + t\omega)\omega$ $\forall t \in \mathbb{R}$

Lemma 3.1.2 1^{st} & 2^{nd} order Taylor formula (proof uses Lemma 3.1.1)

Assume 1^{st} & 2^{nd} partials of
$$f(x) = f(x^*) + \nabla f(y)^T (x - x^*) \quad y \in [x, x^*]$$
$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T Hf(z)(x - x^*) \quad z \in [x, x^*]$$

- closed set (if a sequence $\{x^{(k)}\}$ in the set has a limit pt, the limit pt is also in the set)

Lemma 3.1.3 $g_1, \ldots, g_\ell, h_1, \ldots, h_m$ conts $\Rightarrow D = \{x \in \mathbb{R}^n : g_i(x) \le 0, h_i(x) = 0\}$ is closed.

- bounded set

(4) Existence of Global Optima : Thm 3.2.1 If f is continuous everywhere defined on a non-empty, closed, bounded set D in $\mathbb{R}^n$ then $\exists\, x^* \in D$ s.t $f(x^*) \le f(x) \quad \forall\, x \in D$.

(Bolzano-Weierstrass: $D \ne \emptyset$, closed, bounded in $\mathbb{R}^n$ & $\{x^{(k)}\} \subset D$ then $\exists$ a subsequence that converges in D)

Examples that show you cannot drop any of the assumptions.

Corollary 3.2.1 f, g_i, h_i conts. If (1) $D \ne \emptyset$ and (2) either D is bounded or f is coercive then NLP has a global opt.

How to check whether f is coercive.

(Why is Cor 3.2.1 a corollary of Thm 3.2.1?)

(5) Local Optima

pd and psd matrices and all the checks for them.

Thm 3.3.2 (Unconstrained NLP: $\min f(x)$)

a) x^* local opt $\Rightarrow \nabla f(x^*) = 0$ and $Hf(x^*)$ psd

b) $\nabla f(x^*) = 0$ and $Hf(x^*)$ pd $\Rightarrow x^*$ local opt

(Proof uses Taylor's formula)

Algorithm: To find local optima in the unconstrained case. Solve for $\nabla f(x) = 0$. For each sol^n y check $Hf(y)$. If $Hf(y)$ pd then y local opt. If $Hf(y)$ not psd then y not local opt. If $Hf(y)$ is psd but not pd need further checks.

Thm 3.4.1 (KKT Theorem) (constrained case) Assume 1^{st} & 2^{nd} partials of $f, g_1, \dots, g_\ell, h_1, \dots, h_m$ are defined & conts on $\mathbb{R}^n$. For any feasible sol^{n} $x \in \mathbb{R}^n$, let $I(x) = \{ i \in \{1, \dots, \ell\} : g_i(x) = 0 \}$

(a) If x^* local opt s.t $\{ \nabla g_i(x^*) \;\; i \in I(x^*), \;\; \nabla h_i(x^*) \;\; i = 1, \dots m \}$ are linearly independent then $\exists$ #s $\lambda_1^*, \dots, \lambda_\ell^*, \mu_1^*, \dots, \mu_m^*$ s.t

$$(KKT) \quad \nabla f(x^*) + \sum_{i=1}^{\ell} \lambda_i^* \nabla g_i(x^*) + \sum_{i=1}^{m} \mu_i^* \nabla h_i(x^*) = 0$$

$$\lambda_i^* \geq 0 \qquad \lambda_i^* g_i(x^*) = 0 \quad \forall \; i = 1, \dots, \ell$$

$$* \quad \text{and} \quad \omega^T M \omega \geq 0 \quad \forall \; \omega \text{ s.t } \begin{cases} \nabla g_i(x^*)^T \omega = 0 & i \in I(x^*) \\ \nabla h_i(x^*)^T \omega = 0 & i = 1, \dots, m \end{cases}$$

where $M = Hf(x^*) + \sum \lambda_i^* H g_i(x^*) + \sum \mu_i^* H h_i(x^*)$

(b) If x^* feasible & $\exists$ λ_i^*, μ_i^* satisfying (KKT) and M pd then x^* is a local opt.

Algorithm: Solve for KKT cond^{n} & feasibility. For each sol^{n} $x^*, \lambda_1^*, \dots, \lambda_\ell^*, \mu_1^*, \dots, \mu_m^*$ check if $M \succ 0$.
If yes then x^* local opt. If M not pd but LI cond^{n} of gradients hold then check if $*$ holds.
If no, then x^* not local opt. If yes then need more checks.

(6) Global Optima

- convex functions
- Different properties of convex functions
- Thm 3.5.1
  (a) If $\frac{\partial f}{\partial x_i}$ conts on $\mathbb{R}^n$, f convex $\iff$ $f(x) + \nabla f(x)^T (y - x) \leq f(y)$ $\forall \; x, y \in \mathbb{R}^n$
  (b) If $\frac{\partial^2 f}{\partial x_i \partial x_j}$ conts on $\mathbb{R}^n$, f convex $\iff$ $Hf(x)$ psd $\forall \; x \in \mathbb{R}^n$

Thm 3.5.2 f, g_i convex with 1st partials conts on $\mathbb{R}^n$
h_i's are affine $\Rightarrow$ for any feasible sol$^{\underline{n}}$ $x^* \in \mathbb{R}^n$
that satisfies KKT for some λ_i^*, μ_i^*, x^* is global opt.

(Includes all of LPs)

Chapter 4: How to find feasible solutions that satisfy KKT

4.1 Unconstrained problems

Idea Generate $x^{(0)}, x^{(1)}, x^{(2)} \ldots$ s.t ① $f(x^{(k+1)}) < f(x^{(k)})$

② If $x^* = \lim_{k\to\infty} x^{(k)}$ then $\nabla f(x^*) = 0$

$x^{(k+1)} = x^{(k)} + t_k d^{(k)}$ Let $\phi_k(t) = f(x^{(k)} + t d^{(k)})$

Choose t_k: ① Take t_k to be global minimizer of $\phi_k(t)$ (1-var problem)

② Armijo's rule: Assume $\phi_k'(0) < 0$ Choose t_k to be largest
$t \in \{1, 1/2, 1/4, 1/8, \ldots\}$ s.t $\dfrac{\phi_k(t) - \phi_k(0)}{t} \leq .1\, \phi_k'(0)$

Choosing $d^{(k)}$ Always need $\nabla f(x^{(k)})^T d^{(k)} < 0$

① Cauchy's steepest descent: $d^{(k)} := -\nabla f(x^{(k)})$

② modified Newton direction: Choose $P^{(k)} \succ 0$

$$d^{(k)} = -(P^{(k)})^{-1} \nabla f(x^{(k)})$$

eg
- $P^{(k)} = I_n$
- $P^{(k)} = Hf(x^{(k)})$ if $Hf(x^{(k)}) \succ 0$
- $P^{(k)} = \begin{bmatrix} \ddots & & 0 \\ & \max\{\frac{\partial^2 f}{\partial x_i^2}(x^{(k)}), 1\} & \\ 0 & & \ddots \end{bmatrix}$ if $Hf(x^{(k)}) \not\succ 0$ & n large
- $P^{(k)} = Hf(x^{(k)}) + \gamma I_n$ if $Hf(x^{(k)}) \not\succ 0$ & n not large
where $\gamma = \max\{0, 1 - \text{min eigenval of } Hf(x^{(k)})\}$

③ Quasi-newton

Proof that you achieve your initial goals with these choices of t_k and $d^{(k)}$

4.2 Methods for the constrained problems ① Penalty
② Lagrange ③ Feasible descent