Theorem 2.1 If $f \in C^1$, then the followings are equivalent:

1. f is convex
2. $f(y) \geq f(x) + \nabla^T f(x)(y-x)$ for $\forall x, y \in \operatorname{dom} f$, i.e., f lines above the tangent line.

Prof. 1. From the definition for convexity,

$$f(y) - f(x) \geq \frac{f(\lambda x + (1-\lambda)y) - f(x)}{1-\lambda}$$

Letting $\lambda \to 1$, the RHS becomes a direction derivative:

$$f(y) - f(x) \geq \nabla^T f(x)(y-x)$$

2. To show the converse, we let $z = \lambda x + (1-\lambda)y$. By applying the inequality in (2.1) twice, we have

$$f(x) \geq f(z) + \nabla^T f(z)(x-z) \tag{2.1}$$

$$f(y) \geq f(z) + \nabla^T f(z)(y-z) \tag{2.2}$$

Letting Eq.(2.1) times λ add Eq.(2.2) times $(1-\lambda)$, we derive that f is convex.

Theorem 2.2 If $f \in C^2$, then the followings are equivalent:

1. f is convex
2. $\nabla^2 f(x) \succeq 0$ for $\forall x \in \operatorname{dom} f$.

Prof. We rewrite $f(y)$ by applying Taylor expansion:

$$f(y) = f(x) + \nabla^T f(x)(y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(x + t(y-x))(y-x), \tag{2.3}$$

for some $t \in [0,1]$.
1. If f is convex, from Theorem(2.1) and Eq.(2.3), we derive

$$(y-x)^T \nabla^2 f(x + t(y-x))(y-x) \geq 0 \implies \frac{(y-x)^T}{\|y-x\|} \nabla^2 f(x + t(y-x)) \frac{(y-x)}{\|y-x\|} \geq 0$$

Set $d := \frac{|y-x|}{|y-x|}$ and let $y \to x$, we derive

$$d^T \nabla^2 f(x) d \geq 0,$$

which implies $\nabla^2 f(x) \succeq 0$ since d could have an arbitrary direction.
2. To show the converse, due to the semidefiniteness of $\nabla^2 f(x)$, we obtain a new inequality from Eq.(2.3):

$$f(y) \geq f(x) + \nabla^T f(x)(y-x)$$

From Theorem(2.1) we imply f is convex.

Definition 2.4 [Epigraph] The Epigraph of f is given by:

$$\operatorname{Epi}(f) := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid x \in \operatorname{dom} f, t \geq f(x)\} \subseteq \mathbb{R}^{n+1}$$

■

Theorem 2.3 f is convex iff $\operatorname{Epi}(f)$ is convex.

Prof. 1. Suppose f is convex. For any $(x,t), (y,s) \in \operatorname{Epi}(f)$, it suffices to show

$$(\lambda x + (1-\lambda)y, \lambda t + (1-\lambda)s) \in \operatorname{Epi}(f) \iff \lambda t + (1-\lambda)s \geq f(\lambda x + (1-\lambda)y).$$

Put $\Gamma = \operatorname{epi}(f)$. Suppose first that f is convex, and let $(x_1, t_1), \ldots, (x_n, t_n) \in \Gamma$. For any $\lambda_1, \ldots, \lambda_n \in [0,1]$ with $\sum \lambda_i = 1$, the point $(x,t) = \lambda_i \sum (x_i, t_i) = (\sum \lambda_i x_i, \sum \lambda_i t_i)$ has

$$t = \sum \lambda_i t_i \geq \sum \lambda_i f(x_i) \geq f\left(\sum \lambda_i x_i\right) = f(x).$$

Hence $(x,t) \in \Gamma$, and Γ is copvex. The converse is entirely similar.

Choice of Step Length. To get the optimal step length α, differentiating the RHS w.r.t. α leads to

$$\langle \nabla f(x), d\rangle + L\alpha\|d\|^2 = 0 \implies \alpha = -\frac{\langle \nabla f(x), d\rangle}{L\|d\|^2} > 0,$$

which seems a reasonable choice. If d is the steepest descent direction, the step-length becomes a constant:

$$\alpha = \frac{1}{L}.$$

Theorem 2.4 If f is convex in C^1, then $\nabla f(x) = 0$ is the necessary and sufficient condition for global minimum.

Note that convex function does not have a local minimum that is not global minimum

Prof. If $f \in C^1$ is convex, recall the Theorem(2.1) that

$$f(y) \geq f(x) + \nabla^T f(x)(y-x) \tag{2.4}$$

1. If $\nabla f(x) = 0$, then Eq.(2.3) implies $f(y) \geq f(x)$ for $\forall y$.
2. If x is the global minimum, recall the optimality conditon, $\nabla f(x) = 0$.

Definition 4.1 [Q_1 Factor] Restrict the attention to a convergent sequence $\{x^k\}$ with limit x^*. Define an error function $e_k = \|x^k - x^*\| \to 0$. The Q_1 factor of $\{x^k\}$ is given as:

$$Q_1 = \limsup_{k\to\infty} \frac{e_{k+1}}{e_k}$$

■

Convergence Rate Analysis. We apply the Lipschitzness to analysis the rate of convergence first. Setting $h(t) = f(x + t\alpha d)$, we find that

$$\begin{aligned}
f(x+\alpha d) - f(x) &= h(1) - h(0) = \int_0^1 h'(t)\,dt \\
&= \int_0^1 \langle \nabla f(x + t\cdot\alpha d), \alpha d\rangle\,dt \\
&= \int_0^1 [\langle \nabla f(x + t\cdot\alpha d), \alpha d\rangle - \langle \nabla f(x), \alpha d\rangle + \langle \nabla f(x), \alpha d\rangle]\,dt \\
&= \langle \nabla f(x), \alpha d\rangle + \int_0^1 \langle \nabla f(x + t\cdot\alpha d) - \nabla f(x), \alpha d\rangle\,dt \\
&\leq \langle \nabla f(x), \alpha d\rangle + \int_0^1 \|\nabla f(x + t\cdot\alpha d) - \nabla f(x)\|\cdot\|\alpha d\|\,dt \\
&\leq \langle \nabla f(x), \alpha d\rangle + L\int_0^1 t\alpha^2\|d\|^2\,dt \\
&= \underbrace{\langle \nabla f(x)}, \underbrace{\alpha d\rangle} + \frac{L\alpha^2\|d\|^2}{2}
\end{aligned}$$

Here we want to study the performance of e_k. In our case, we compare $\{e_k\}$ with the geometric progression

$$\beta^k, \quad k = 0, 1, \ldots$$

- If there exists $\beta \in (0,1)$ such that

$$Q_1 \leq \beta,$$

  then we can show $e_k \leq q\beta^k$ for some $q > 0$. In this case $\{e_k\}$ is said to be Q-linear convergent.
- If $Q_1 = 0$, then we say $\{e_k\}$ is Q-super-linear convergent
- If $Q_1 = 1$, then we say $\{e_k\}$ is Q-sub-linear convergent

Definition 4.2 [Q_2 Factor]

$$Q_2 = \limsup_{k\to\infty} \frac{e^{k+1}}{(e^k)^2}$$

If $Q_2 = M < +\infty$, i.e., $e^{k+1} = O((e^k)^2)$, then $\{x^k\} \to x^*$ Q-quadratically. ■

To get the optimal solution, it suffices to solve (4.1) for d:

$$d = -(\nabla^2 f(x))^{-1}\nabla f(x),$$

and hence update the solution to be $x \leftarrow x + \alpha d$.

Proposition 4.1 Newton's method gurantees the Q-quadratic convergence.

Prof. Given a nonlinear system $F(x) = 0$, suppose the sequence $\{x^k\}$ is generated by Newton with limit x^* and $F(x^*) = 0$. By Newton's iteration,

$$x^{k+1} = x^k - [F'(x^k)]^{-1}F(x^k), \tag{4.2}$$

which follows that

$$\begin{aligned}
x^{k+1} - x^* &= x^k - x^* - [F'(x^k)]^{-1}\left(F(x^k) - F(x^*)\right) \\
&= [F'(x^k)]^{-1}\left(F(x^*) - F(x^k) - F'(x^k)(x^* - x^k)\right)
\end{aligned} \tag{4.3}$$

Note that $F(x^*) = F(x^k) + F'(x^k)(x^* - x^k) + O(\|x^k - x^*\|^2)$, which implies

$$\|x^{k+1} - x^*\| \leq \|[F'(x^k)]^{-1}\| O(\|x^k - x^*\|^2) = O(\|x^k - x^*\|^2).$$

$$x^{k+1} = x^k - \alpha^k\nabla f(x^k) = (I - \alpha^k Q)x^k,$$

$$\begin{aligned}
\|x^{k+1}\|^2 &= (x^k)^T(I - \alpha^k Q)^2 x^r \leq \lambda_1[(I - \alpha^k Q)^2]\|x^r\|^2 \\
&= \|x^k\|^2 \cdot \max\{(1-\alpha^k m)^2, (1-\alpha^k M)^2\},
\end{aligned}$$

The idea for feasible direction method is that we generate a sequence of $\{x^k\} \subseteq X$ such that $f(x^{r+1}) \leq f(x^r)$, or equivalently, find $\hat{x}^r \in X$ such that the $\hat{x}^r - x^r$ is the descent direction:

$$\langle \nabla f(x^r), (\hat{x}^r - x^r)\rangle \leq 0 \tag{7.7a}$$

and therefore

$$x^{r+1} = x^r + \alpha_r(\hat{x}^r - x^r), \quad \alpha_r \in (0,1] \tag{7.7b}$$

The feasibility of x^{r+1} is guaranteed since x^{r+1} is a convex combination of feasible points x^r and $\hat{x}^r$. Here the problem remains to find $\hat{x}^r$

Projection Gradient Method. One way of finding $\hat{x}^r$ is to compute $x^r - s_r\nabla f(x^r)$ and project it back into X, as $\hat{x}^r$:

$$\hat{x}^r = [x^r - s_r\nabla f(x^r)]^+, \quad s_r > 0\,.$$

Conditional Gradient. Another way is to linearize the objective function and solve for $\hat{x}^r$:

$$\hat{x}^r \approx \arg\min_{x\in X} f(x) \implies \hat{x}^r = \arg\min_{x\in X} f(x^r) + \langle \nabla f(x^r), (x - x^r)\rangle.$$

$$\frac{\partial f^T g}{\partial x} = \frac{\partial f}{x}g + \frac{\partial g}{x}f$$

which implies

$$\frac{\|x^{k+1}\|}{\|x^k\|} \leq \max\{|1-\alpha^k m|, |1-\alpha^k M|\}$$

Choosing $\alpha^k = \frac{2}{M+m}$ s.t. $\max\{|1-\alpha^k m|, |1-\alpha^k M|\}$ is maximized, we have

$$\frac{\partial x^T A x}{\partial x} = (A + A^T)x$$

$$\frac{\|x^{k+1}\|}{\|x^k\|} \leq \frac{M-m}{M+m}.$$

$$\frac{\partial\|y - Ax\|_2^2}{\partial x} = 2\frac{\partial y - Ax}{\partial x}(y - Ax) = -2A^T(y - Ax)$$

GD with fixed stepsize.

$$\underbrace{\|x^{k+1} - x^*\|^2}_{\text{non-increasing}} = \|x^k - x^*\|^2 - 2\alpha\langle g^k, x^k - x^*\rangle + \alpha^2\|g^k\|^2$$

when $0 < \alpha < \frac{2}{L}$.

$$\|\nabla f(x^k)\| = o\left(\frac{1}{k}\right)$$

$f(x^k)$, $\|\nabla f(x^k)\| \downarrow$

convergence rate be q-quadratic if $\nabla^2 L(x,\lambda)$ Lipschitz continuous nonsingular,

$$\|\nabla^2 L(x,\lambda) - \nabla^2 L(y,\mu)\|_F^2 \leq k\|(x,\lambda) - (y,\mu)\|$$

MVT:

$$r(x) - r(x^*) = (x - x^*)\cdot\int_0^1 J(x^* + t(x - x^*))\,dt \quad ①$$

$$\begin{aligned}
x^+ - x^* &= (J^TJ)^{-1}J^T \cdot ① \\
&\leq (J^TJ)^{-1}J^T \cdot \int_0^1 L(1-t)\,dt \cdot \|x - x^*\| \\
&= \frac{L}{2}[J^TJ]^{-1}J^T\|x - x^*\|^2
\end{aligned}$$

Quasi-Newton equation

$$B^{k+1}q^k = p^k \qquad p^k = x^{k+1} - x^k, \quad q^k = \nabla f(x^{k+1}) - \nabla f(x^k)$$

$$U = V^T = [Bs \;\; y]$$

$$D = \operatorname{diag}(-1/s^TBs, \; 1/y^Ts)$$

Constraint:

$$\langle \nabla f(x^*), (x - x^*)\rangle \geq 0 \quad \forall x \in X.$$

GN: r_i's are Lipschitz continuasl differentiable

$J(x)$ non-singular near x^*

$$\begin{aligned}
x^+ - x^* &= x - x^* - [J^TJ]^{-1}J^Tr(x) \\
&= [J^TJ]^{-1}J^T\left[J(x - x^*) - (r(x) - r(x^*))\right] \\
&= [J^TJ]^{-1}J^T\int_0^1 J(x) - J(x^* + t(x - x^*))\,dt\,(x - x^*)
\end{aligned}$$

$$\|J(x) - J(x^* + t(x - x^*))\| \leq L|1-t|\,\|x - x^*\|$$

若 E convex

$\forall (x_i, y_i) \in Ef$ for $y_i = f(x_i)$

$f(tx_1 + (1-t)x_2) \leq \ldots$

the quadratic (local) rate of convergence is guaranteed if the following conditions hold:

1. there exists x^* such that $F(x^*) = 0$
2. $[F'(x^*)]^{-1}$ exists
3. F is Lipschitz continuous near x^*.

Theorem 5.4 — Convergence Rate for invariant step-size. Given a convex function $f \in C^2$ with Lipschitz gradient, i.e.,

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|^2,$$

the iteration $x^{k+1} = x^k - \frac{1}{L}\nabla f(x^k)$ with local minimum point x^* gives sub-linear convergence rate, i.e.,

$$f(x^k) - f(x^*) \le \frac{L\|x^0 - x^*\|}{k+1}$$

Proposition 5.1 A convex function $f \in C^2$ with Lipschitz gradient constant L has a bounded Hessian matrix:

$$\nabla^2 f(x) \preceq LI$$

Proof for proposition(5.1). Otherwise $\exists x_0, v$ such that

$$\|\nabla^2 f(x_0)v\| > L\|v\|.$$

Thus we apply taylor expansion near x_0 for $x = x_0 + v$:

$$\nabla f(x) = \nabla f(x_0) + \nabla^2 f(x_0)(x - x_0) + o(1)(x - x_0)$$

It follows that

$$\|\nabla f(x) - \nabla f(x_0)\| = \|\nabla^2 f(x_0)(x - x_0) + o(1)(x - x_0)\| \le L\|x - x_0\|$$

Thus for sufficiently small v, we have $\|\nabla f(x) - \nabla f(x_0)\| \le L\|x - x_0\|$, which is a contradiction. ■

Step 1: Apply Lipschitz condition. We do the taylor expansion of x^k for the point x^{k+1}:

$$\begin{aligned} f(x^{k+1}) &= f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{1}{2}(x^{k+1} - x^k)^T\nabla^2 f(x^k + \tau(x^{k+1-x^k}))(x^{k+1} - x^k \\ &\le f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 \\ &= f(x^k) - \frac{1}{2L}\|\nabla f(x^k)\|^2 \end{aligned}$$

implying that

$$\|\nabla f(x^k)\|^2 \le 2L[f(x^k) - f(x^{k+1})]$$

and therefore

$$\sum_{k=0}^{r}\|\nabla f(x^k)\|^2 \le \sum_{k=0}^{\infty}\|\nabla f(x^k)\|^2 \le 2L[f(x^0) - f(x^*)] \le L^2\|x^0 - x^*\|^2 \quad (5.5)$$

Step 2: Applying Convexity of f. By the convexity of f and the bound on its gradient, we can estimate the total error $\sum_{k=0}^{r}(f(x^k) - f(x^*))$:

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|x^k - \frac{1}{L}\nabla f(x^k) - x^*\|^2 \\ &= \|x^k - x^*\|^2 - \frac{2}{L}\langle \nabla f(x^k), x^k - x^* \rangle + \frac{1}{L^2}\|\nabla f(x^k)\|^2 \\ &\le \|x^k - x^*\|^2 - \frac{2}{L}(f(x^k) - f(x^*)) + \frac{1}{L^2}\|\nabla f(x^k)\|^2 \end{aligned}$$

which implies

$$\begin{aligned} \sum_{k=0}^{r} f(x^k) - f(x^*) &\le \frac{L}{2}\sum_{k=0}^{r}\left[\|x^k - x^*\| - \|x^{k+1} - x^*\|^2\right] + \frac{1}{2L}\sum_{k=0}^{r}\|\nabla f(x^k)\|^2 \\ &\le \frac{L}{2}\left[\|x^0 - x^*\|^2 - \|x^{r+1} - x^*\|^2\right] + \frac{L}{2}\|x^0 - x^*\|^2 \\ &\le L\|x^0 - x^*\|^2 \end{aligned}$$

Step 3: applying monotonicity of $f(x^k) - f(x^*)$. By the monotonicity of $f(x^k) - f(x^*)$,

$$\begin{aligned} f(x^r) - f(x^*) &\le \frac{1}{r+1}\sum_{k=0}^{r} f(x^k) - f(x^*) \\ &\le \frac{L\|x^0 - x^*\|^2}{r+1} \end{aligned}$$

The standard form of linear programming is

$$\begin{aligned} \min \quad & c^T x \\ \text{such that} \quad & Ax = b \\ & x \ge 0 \end{aligned}$$

Define the Lagrange function

$$\begin{aligned} L(x, y, z) &= c^T x - y^T(Ax - b) - z^T x \\ &= (c - A^T y - z)^T x + b^T y \end{aligned}$$

From this Lagrange function we define a dual function

$$Q(y, z) = \inf_x L(x, y, z) = \begin{cases} b^T y, & \text{if } c - A^T y - z = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

For any feasible x and $z \ge 0$, we always have

$$Q(y, x) \le c^T x - z^T x \le c^T x$$

Theorem 7.1 The LP optimality condition is given by:

1. Primal-Feasibility:
$$Ax = b, x \ge 0$$
2. Dual-Feasibility:
$$A^T y + z = c, z \ge 0$$
3. Complementarity/Strong-Duality:
$$x \circ z = 0$$

$$[z]^+ = x^* = \arg\min_{x \in X}\|x - z\|$$

Given the optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{such that} \quad & x \in X \text{ is convex} \end{aligned}$$

Recall that The necessary optimality condition for (7.2) is

$$\langle \nabla f(x^*), (x - x^*) \rangle \ge 0 \quad \forall x \in X$$

Proposition 7.4 — First order projection property. Given convex set X, the necessary and sufficient condition for the local minimum x^* for problem (7.3) is:

$$\langle x - [z]^+, z - [z]^+ \rangle \le 0, \quad \forall x \in X \tag{7.4}$$

Proof. Define $f(x) := \frac{1}{2}\|x - z\|_2^2$, thus $\nabla f(x) = x - z$, with the necessary condition of local minimum x^* as:

$$\langle \nabla f(x^*), (x - x^*) \rangle \ge 0, \quad \forall x \in X \implies \langle x - [z]^+, z - [z]^+ \rangle \le 0.$$

The sufficency of this condition is due to the convexity of problem (7.3). ■

Armijo rule: Let $\sigma \in (0, \frac{1}{2})$. Start with s and continue with $\beta s,\ \beta^2 s, \ldots,$ until $\beta^m s$ falls within the set of α with

$$f(x^r) - f(x^r + \alpha d^r) \ge -\sigma\alpha\nabla f(x^r)'d^r.$$

Claim: if $d^r = -D^r\nabla f(x^r) \ne 0$ with $D^r \succ 0$, then m is finite.

Proposition 7.5 — Non-expansive.

$$\|[z_1]^+ - [z_2]^+\| \le \|z_1 - z_2\|$$

with X to be a convex set.

Proof. Recall the first order property on z_1, z_2, we obtain

$$\begin{cases} \langle z_1 - [z_1]^+, x - [z_1]^+ \rangle \le 0, \forall x \in X \\ \langle z_2 - [z_2]^+, x - [z_2]^+ \rangle \le 0, \forall x \in X \end{cases} \implies \begin{cases} \langle z_1 - [z_1]^+, [z_2]^+ - [z_1]^+ \rangle \le 0, \forall x \in X \\ \langle z_2 - [z_2]^+, [z_1]^+ - [z_2]^+ \rangle \le 0, \forall x \in X \end{cases}$$

Adding the inequalities above, we derive

$$\langle z_1 - z_2 + [z_2]^+ - [z_1]^+, [z_2]^+ - [z_1]^+ \rangle \le 0 \implies \langle [z_2]^+ - [z_1]^+, [z_2]^+ - [z_1]^+ \rangle \le \langle z_2 - z_1, [z_2]^+ - [z_1]^+ \rangle$$

Applying Cauchy Scharwz inequality, we obtain:

$$\|[z_2]^+ - [z_1]^+\|_2^2 \le \langle z_2 - z_1, [z_2]^+ - [z_1]^+ \rangle \le \|z_2 - z_1\|\|[z_2]^+ - [z_1]^+\|$$

Or equivalently, $\|[z_2]^+ - [z_1]^+\| \le \|z_2 - z_1\|$. ■

Proposition 8.4 — KKT conditions. Consider the standard optimization problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{such that} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned} \tag{8.7}$$

with the dual problem

$$\begin{aligned} \max \quad & g(\lambda, \gamma) \\ & \lambda_i \ge 0, \quad i = 1, \ldots, m \end{aligned} \tag{8.8}$$

1. The necessary condition for primal and dual optimal points x^*, (λ, γ^*) is
   - Primal Feasibility: $f_i(x) \le 0, \quad i = 1, \ldots, m;\ h_i(x) = 0, \quad i = 1, \ldots, p$
   - Dual Feasibility: $\lambda \ge 0$
   - Complementary Slacknes: $\lambda_i f_i(x^*) = 0, \quad i = 1, \ldots, m$
   - Stationarity of Lagrange: $\nabla_x L(x^*, \lambda^*, \gamma^*) = 0$
2. When the primal problem is convex, the KKT conditions above are both necessary and sufficient optimality conditions.

- The first order necessary optimality condition is

$$\nabla_x L(x^*, \lambda^*) = 0, \Longleftrightarrow \nabla f(x^*) + \sum_{i=1}^{m}\lambda_i^*\nabla h_i(x^*) = 0,$$

where λ^* is uniquely determined.

- The second order necessary optimality condition is (pre-assume $f, g \in C^2$):

$$y^T\left[\nabla_{xx}^2 L(x^*, \lambda^*)\right]y \ge 0, \quad \forall y \text{ s.t. } \underline{\langle \nabla h(x^*), y \rangle = 0},$$

or equivalently,

$$y^T\left[\nabla^2 f(x^*) + \sum_{i=1}^{m}\lambda_i^*\nabla^2 h_i(x^*)\right]y \ge 0, \quad \forall y \text{ s.t. } \langle \nabla h(x^*), y \rangle = 0,$$

i.e., the Hessian matrix $L_{xx}(x^*, \lambda^*)$ is PSD over the null space of Jacobian matrix of $h(x^*)$.

second order sufficent optimality conditon is

$$\begin{cases} \nabla_x L(x^*, \lambda^*) = 0, & \nabla_\lambda L(x^*, \lambda^*) = 0 \\ y^T\nabla_{xx}^2 L(x^*, \lambda^*)y > 0, & \forall y \ne 0 \text{ with } \langle \nabla h(x^*), y \rangle = 0 \end{cases}$$

Equality constraint:

min $f(x)$ $h_i(x) = 0$ $i = 1, \ldots, m$

$\nabla_x L(x^*, \lambda^*) = 0$ $\nabla_\lambda L(x^*, \lambda^*) = 0$

$y^T\nabla_{xx}^2 L(x^*, \lambda^*)y \ge 0$

$\forall y$ $\langle \nabla h(x^*), y \rangle = 0$

Suff: $y^T\nabla_{xx}^2 L(x^*, \lambda^*)y > 0$

$\forall y \ne 0$

Inequality:

nec: $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$

$\mu_j^* = 0$ for non-active

$\mu_j^* \ge 0$ $\forall j$

KKT Dual:

max $g(\lambda, \gamma)$

$\lambda_i \ge 0$.