INTRODUCTION

General optim problem: $\min_{x \in S} f(x)$

$f : S \to \mathbb{R}$ ≡ objective function

x = variable, unknown
S = feasible region

Def x^* is a [STRICT] minimizer of f on S if

$$f(x^*) \underset{[<]}{\leq} f(x) \quad \forall x \in S \quad [x \neq x^*]$$

Convention: We only look at minimization problems as

$$\max_{x \in S} f(x) \iff \min_{x \in S} -f(x)$$

Important questions:

1. existence of a sol
2. uniqueness of a sol
3. characterization of a sol

} briefly

4. Numerical approx of sol } most of the class.

S can be a discrete set $\Rightarrow$ discrete optim problem (another class)

Here we shall assume that: for $g : \mathbb{R}^n \to \mathbb{R}^m$ & $h : \mathbb{R}^n \to \mathbb{R}^p$.

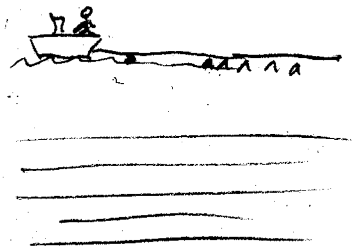
$$S = \{ x \mid \underbrace{g(x) = 0}_{\text{equal constr}}, \underbrace{h(x) \geq 0}_{\text{inequal constr}} \}$$

this is Continuum optimization.

Many real world problems can be formulated (modeled) as a continuum optim problem. Here are some examples.

Examples:

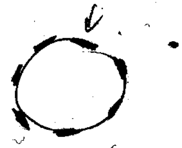
- oil exploration:



find wave velocity that better fits observed measu[rements]

- medical imaging

electrodes



find conductivity inside Ω that better fits electrical measurements at boundary

- economy:

find portfolio that max returns while keeping risk within a certain level.

- design:

find wing cross section that maximizes lift

etc....

Most of the results and algorithms we shall see are inspired by results from 1D: $f:\mathbb{R}\to\mathbb{R}$ s.t.

$$\min_{x\in\mathbb{R}} f(x)$$

- First order necessity

If x^* minimizes f, then $f'(x^*)=0$

- Second order necessity

If x^* minimizes f, then $f''(x^*)\geq 0$.

find candidate for min

- Second order sufficiency:

If x^* is s.t. $f'(x_*)=0$
$f''(x_*)>0$

- then x^* is a strict local min of f:

$$\exists\,\varepsilon>0\quad \forall x\neq x^*, |x-x^*|<\varepsilon \Rightarrow f(x^*)<f(x)$$

Cond on curvature basically means that Taylor exp up to 2nd ord[er]
$f(x^*+h)=f(x^*)+f'(x^*)h$
$+f''(x)h$
looks like parabola

$f(x)=x^4$
example where sufficiency not satisfied

• Existence:
If f is continuous on the closed & bdd interval $[a\ b]$,
then f has at least one minimizer in $[a, b]$.

And most of the algorithms are descendants of Newton's method
for solving $\min\limits_{x \in \mathbb{R}} f(x)$. (1)

Derivation of Newton's method in 1D

Given current approx x_c to x^* minimizer of (1); find "best" update

Taylor's theorem:

$$f(x_c + s) = f(x_c) + f'(x_c)\, s + \frac{1}{2} f''(x_c)\, s^2 + r(s^2)$$

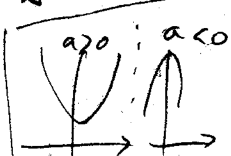
where $\lim\limits_{s \to 0} \frac{r(s^2)}{s^2} = 0$
or in other words
$r(s^2) = o(s^2)$

$$\approx f(x_c) + f'(x_c)\, s + \frac{1}{2} f''(x_c)\, s^2 = m_c(s)$$

Idea: minimize quadratic model $m_c(s)$ instead of f:
to find best update

$$\min_s f(x_c + s) \approx \min_s m_c(s)$$

How to minimize quadratic



$q(x) = \frac{1}{2} a x^2 + bx + c$
$q'(x) = ax + b = 0$
$\Rightarrow x^* = -\frac{b}{a}$
but in order to have min we need $a > 0$
exactly what I meant before.

Provided $f''(x_c) > 0$, the minimizer to quad model is

$$s_c = -\frac{f'(x_c)}{f''(x_c)}$$

thus we take iteration $x_+ = x_c + s_c$ or in algorithmic notation:

choose x_0 initial point
for $k = 0 \ldots$

$$x_{k+1} = x_k \underbrace{- \frac{f'(x_c)}{f''(x_c)}}_{s_c}$$

Convergence of Newton method

Let be smooth and assume x^* is a minim of f satisfying 2nd order sufficieny cdt° (i.e. $f'(x^*)=0$ and $f''(x^*)>0$) then if the start point x_0 is sufficiently close to x^*,

$$|x_{n+1} - x^*| \leq C|x_n - x^*|^2$$

quadratic convergence, fastest one can expect in general.

how to generalize to $\mathbb{R}^n$?
how to deal with local convergence?
what to do if $f'(x)$ or $f''(x)$ have errors or are expensive to compute?

Tools we need for $\mathbb{R}^n$

Gradient $\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$

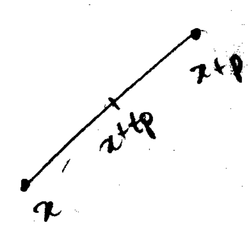
Hessian $\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$ symmetric matrix provided f is twice cont diff'ble (f, ∇f & $\nabla^2 f$ are cont.)

Dot product: $x^T y = \sum_{i=1}^{n} x_i y_i$

Matrix vector mult: $(Ax)_i = \sum_{j=1}^{n} A_{ij} x_j$

Taylor's theorem

Let $f: \mathbb{R}^n \to \mathbb{R}$ be continuously diff'ble then:



$$\exists\, t \in (0,1) \text{ s.t. } f(x+p) = f(x) + \nabla f(x+tp)^T p$$

or equivalently:

$$f(x+p) = f(x) + \nabla f(x)^T p + R_1(p;x) \text{ s.t. } \lim_{p \to 0} \frac{R_1(p;x)}{\|p\|} = 0$$

If f is twice cont. diff'ble then: (I5)

$\exists\, t \in (0,1)$ s.t. $f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p$.

or equivalently:

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x) p + R_2(p;x)$$

with $\lim_{p \to 0} \frac{R_2(p;x)}{\|p\|^2} = 0$ (residual is negligible)

<u>Fundamental theorem of calculus</u>

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)\, p\, dt.$$

You can already imagine that we are going to use <u>quadratic model</u> for deriving Newton's method in $\mathbb{R}^n$.

$$q(x) = c + g^T x + \frac{1}{2} x^T H x \quad = \text{quadratic function in } \mathbb{R}^n.$$

We assume H is symmetric

<u>Gradient computation: (Taylor based)</u>

$$\nabla q(x)^T v = \lim_{t \to 0} \frac{q(x+tv) - q(x)}{t}$$

$$= \lim_{t \to 0} \frac{t g^T v + \frac{1}{2} t\, v^T H x + \frac{1}{2} x^T H v + \frac{t^2}{2} v^T H v}{t}$$

$$= g^T v + \frac{1}{2}(Hx + H^T x)^T v = (g + Hx)^T v.$$

$\Rightarrow \nabla q(x) = g + Hx$ (by identification)

Hessian computation (Taylor based)

$$\frac{1}{2} v^T \nabla^2 q(x) v = \lim_{t \to 0} \frac{q(x+tv) - q(x) - \nabla q(x)^T t v}{t^2}$$

$$= \lim_{t \to 0} \frac{\frac{1}{2} t^2 v^T H v}{t^2} = \frac{1}{2} v^T H v$$

$$\Rightarrow \boxed{\nabla^2 q(x) = H}$$ · by identification (works for symm matrix only)

We need tool to tell "sign" of Hessian ($f''(x) > 0$ ....)

Def H symm is said to be symmetric pos def if

$\forall v \in \mathbb{R}^n$, $v \neq 0$: $v^T H v > 0$

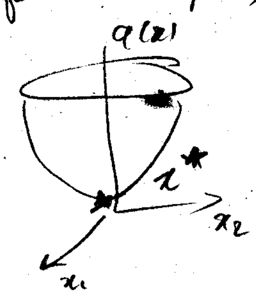
H symmetric positive semidef if:

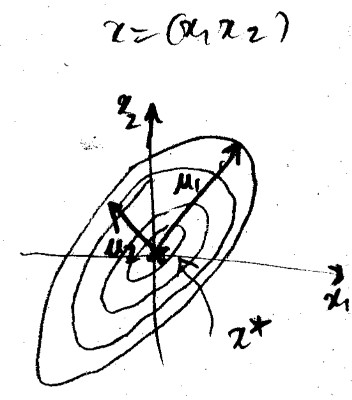
$\forall v \in \mathbb{R}^n$ $v^T H v \geq 0$

(negative [semi] def: flip inequalities)

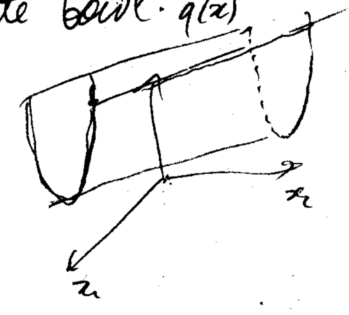
Implications for quadratic: ($\mathbb{R}^2$ for example)

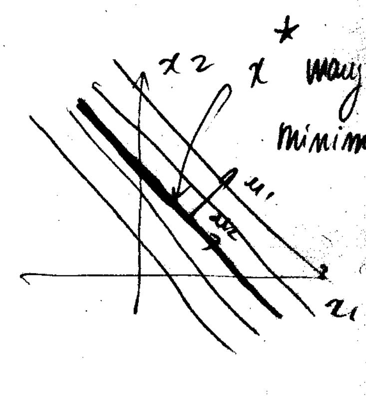
<u>H pos def</u> $\Rightarrow$ Bowl



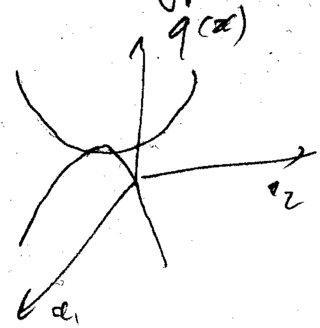


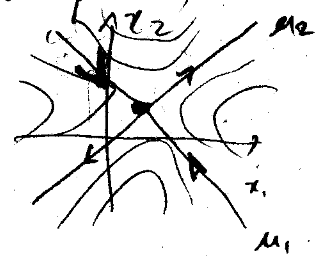
<u>H pos semidef</u> $\Rightarrow$ degenerate bowl





<u>H indefinite</u>: hyperboloid (saddle point, mountain pass)





feature in $\mathbb{R}^n$ that we don't have in $\mathbb{R}$

(λ, u) is an <u>eigenpair</u> of A if $u \neq 0$ and $Au = \lambda u$

eigenvalue eigenvector

If $A \in \mathbb{R}^{n \times n}$ and symmetric:

λ is real: $u^T . Au = \lambda u^T u$

if (λ, u) (μ, v) are eigenpairs of A with $\lambda \neq \mu$ $u^T v = 0$

<u>Rayleigh quotient:</u>

$$\lambda_{min}(A) \leq \frac{u^T A u}{u^T u} \leq \lambda_{max}(A) \quad \text{for } u \neq 0$$

$$\lambda_{max}(A) = \max_{u \neq 0} \frac{u^T A u}{u^T u}$$

$$\lambda_{min}(A) = \min_{u \neq 0} \frac{u^T A u}{u^T u}$$

In quadratics eigenvectors give
- axis of hyperboloid
- axis of ellipsoid

A symm pos def : $\lambda_i(A) > 0$

A — pos semidef : $\lambda_i(A) \geq 0$.

Norms:

$\|x\|_2^2 = x^T x = \sum_{i=1}^{n} x_i^2$ = Euclidean norm (default)

$\|x\|_1 = \sum_{i=1}^{n} |x_i|$ = l_1 norm

$\|x\|_\infty = \max_{i=1..n} |x_i|$ infinity norm

$\boxed{\begin{array}{l} \|x\| = 0 \Leftrightarrow x = 0 \\ \|\alpha x\| = |\alpha| \, \|x\| \end{array}}$

Matrix norms:

$\|A\|_F^2 = \sum_{i,j=1}^{n} a_{ij}^2$ (simplest: view matrix as a vector.

Induced matrix norm:

$$\|A\| = \max_{u \neq 0} \frac{\|Au\|}{\|u\|}$$

for $\|\cdot\|_2$ norms: $\|A\|_2 = \max_{u \neq 0} \frac{\|Au\|}{\|u\|}$ (See A.1 for other induced norms)

For A symmetric: $\|A\|_2^2 = \max_{u \neq 0} \frac{\|Au\|^2}{\|u\|^2} = \max_{u \neq 0} \frac{u^T A^2 u}{u^T u} = \lambda_{max}(A^2)$

$= \lambda_{max}(A)^2$

$$\Rightarrow \|A\|_2 = |\lambda_{max}(A)|$$

Inequalities:

Cauchy Schwartz $|u^T v| \leq \|u\| \, \|v\|$

$\|Au\| \leq \|A\| \, \|u\|$

op-induced matrix norm.

Triangle ineq

$\|u + v\| \leq \|u\| + \|v\|$

MATH 5750/6880 Optimization

§2 Unconstrained optimization

(1) $\min_{x \in \mathcal{F}} f(x)$ where $\mathcal{F} \subset \mathbb{R}^n$ = feasible set and $f: \mathbb{R}^n \to \mathbb{R}$ is a smooth function

UNCONSTRAINED: $\mathcal{F} = \mathbb{R}^n$

Def:

- x^* is called a global [STRICT] minimum (solution) of (1) if

$$f(x^*) \leq f(x) \quad \forall x \in \mathcal{F} \; [x \neq x^*]$$
$$[<]$$

- x^* is called a local [STRICT] minimum (sol) of (1) if

$$\exists \varepsilon > 0 \quad f(x_*) \leq f(x) \quad \forall x \in B_\varepsilon(x^*) \cap \mathcal{F}$$
$$[<]$$

Existence of a minimizer

Thm: A continuous function over a compact set attains its minimum.

Note: in $\mathbb{R}^n$ "compact set $\Leftrightarrow$ closed and bounded.

However $\mathbb{R}^n$ is unbounded!

Thm: Let $\mathcal{F} = \mathbb{R}^n$, Let $f: \mathbb{R}^n \to \mathbb{R}$ be continuous and have the "infinity property":

$$\|x\| \to \infty \Rightarrow f(x) \to +\infty.$$

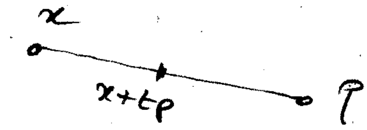
then $\exists x_*$ s.t. $f(x^*) = \inf f(x)$ (makes level sets $\{x \mid f(x) < f(x_0)\}$ compact)

In many case the objective function is of the form

$$f(x) = g(x) + \alpha \|x\|_2^2, \text{ and } g(x) \text{ bounded below}$$

thus existence of solution is guaranteed (sometimes term is added to guarantee existence of sol)

Charaterization of local minima



Taylor's theorem

Let $f: \mathbb{R}^n \to \mathbb{R}$ be continuously diff'ble (means ∇f cont.) then:

$$\exists\, t \in (0,1) \text{ s.t. } f(x+p) = f(x) + \nabla f(x+tp)^T p$$

If f is twice cont. diff'ble ($\nabla^2 f$ cont) then:

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)\, p \, dt$$

and

$$\exists\, t \in (0,1) \text{ s.t.}$$

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp)\, p.$$

Recall: $\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{i,j=1 \cdots n}$ = symm matrix if f is smooth.

Necessary optimality cdt:

Thm (First order necessary optimality cdt)
If x^* is a local minimizer and f is continuously diff'ble on a nbd of x^* then $\nabla f(x^*) = 0$.

Proof: Suppose for contrad. $\nabla f(x^*) \neq 0$.

Let $p = -\nabla f(x^*)$

then $p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0$

By continuity of $\nabla f(x)$, $\exists\, T$ s.t. $\forall\, t \in [0,T]$:

$$p^T \nabla f(x^* + tp) < 0$$

Now by Taylor's theorem:

$\forall \bar{t} \in [0,T], \exists\, t \in (0,\bar{t})$ s.t.

$$f(x^* + \bar{t}p) = f(x^*) + \bar{t}\, \underbrace{\nabla f(x^* + tp)^T p}_{<0}$$

$$\Rightarrow f(x^* + \bar{t}p) < f(x^*) \rightsquigarrow \text{contradiction}$$

<u>Thm</u> (Second order necessary optimality cdt°)

If x^* is a local minimizer of f and $D^2 f$ exists and is continuous in a neighborhood of x^* then:

$$\nabla f(x^*) = 0 \quad \text{and} \quad \nabla^2 f(x^*) \text{ is pos semidef.}$$

<u>proof</u>: Taylors theorem. See book.

Recall $A \in \mathbb{R}^{n\times n}$ is symm pos semidef $\Leftrightarrow \forall p \in \mathbb{R}^n \quad p^T A p \geq 0$

$$\Leftrightarrow \lambda_{min}(A) \geq 0.$$

pos def same but strict ineq.

<u>Thm</u> (Second-order sufficient optim cdt°)

Suppose • $D^2 f$ is continuous in a nbd of x^*

• $\nabla f(x^*) = 0$

• $\nabla^2 f(x^*)$ is pos definite

$\Rightarrow x^*$ is a strict local minimizer of f. <u>proof</u>: see book, also based on Taylor's theorem.

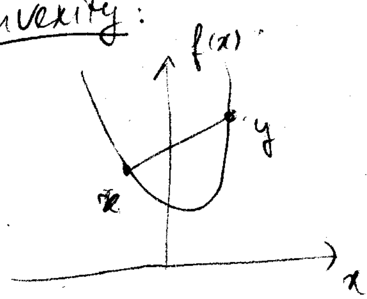
<u>Note</u>: There is a gap between nec and sufficient cdt° for optimality:

I.e. not all minimizers satisfy sufficient cdt°.

<u>Simplest example</u>: $f(x) = x^4 \quad D^2 f(x) = 12x^2$

$$D^2 f(0) = 0$$

<u>Convexity</u>:



f is convex iff $\forall x, y \in \mathbb{R}^n \; \forall \lambda > 0$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

<u>Theorem</u>: When f is convex: x^* local minimizer $\Rightarrow x^*$ global minimizer

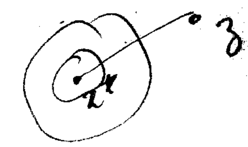
<u>Proof</u>: By contradiction

If x^* = local min, but x^* is not a global min then

$\exists z$ s.t. $f(z) < f(x^*)$

Consider line segment joining x^* to z: *important*

$$x = \lambda z + (1-\lambda)x^*, \quad \lambda \in (0,1]$$

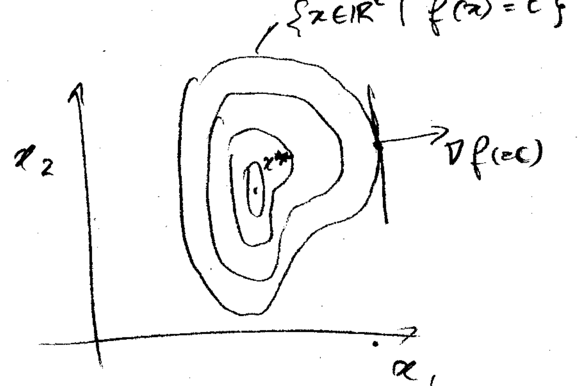


Since f is convex:

$$f(x) \le \lambda f(z) + (1-\lambda) f(x^*) < f(x^*)$$

Thus any ball containing x^* contains a point in line segment, with $f(x) < f(x^*)$

thus x^* is not a local min.



∇f points to increasing vals of f.

Def Descent Direction

d is a descent direction provided

$$d^T \nabla f(x) < 0$$

(angle with gradient is $> \frac{\pi}{2}$)

in particular $d = -\nabla f(x)$ is a descent direction iff $\nabla f(x) \neq 0$.

since $\quad -\nabla f(x)^T \nabla f(x) = -\|\nabla f(x)\|^2 < 0$

Quadratic optimization problems

$$f(x) = g^T x + \frac{1}{2} x^T H x \qquad \text{(const can be added, but does not change min)}$$

$$\nabla f(x) = g + Hx$$

$$\nabla^2 f(x) = H \qquad (H = \text{symm})$$

1st order nec. cond: x^* min $\Rightarrow$ $Hx^* = -g$

If $-g \notin \text{Range}(H)$ then $\nabla f(x) \neq 0$ and so there is no min/max

To ensure we have a minimizer:

then: x^* solves $\min f(x) \iff g + Hx^* = 0$ and H is pos semidef

H pos def $\Rightarrow$ there is a unique sol.

Newton's method $\min_{x \in \mathbb{R}^n} f(x)$ $f: \mathbb{R}^n \to \mathbb{R}$

(P44)

Given approx x_c = (c as in "current") of x^*:

$$f(x_c+s) = f(x_c) + \nabla f(x_c)^T s + \frac{1}{2} s^T \nabla^2 f(x) s + \underbrace{R_2(x_c; s)}_{\text{small residual}}$$

$$\approx f(x_c) + \nabla f(x_c)^T s + \frac{1}{2} s^T \nabla^2 f(x) s = m_c(s) \quad o(\|s\|^2)$$

$= \underset{\text{QUADRATIC}}{\hat{}}$ model at x_c of f.

Minimize model instead of f:

$$\min_s f(x_c+s) \approx \min_s m_c(s)$$

If s_c solves $\boxed{\nabla^2 f(x_c) s_c = -\nabla f(x_c)}$ and

$\nabla^2 f(x_c)$ is pos semidef then s_c is a min of m_c.

take: $x_+ = x_c + s_c$.

If $\nabla^2 f(x_c)$ is pos def then $s_c = -(\nabla^2 f(x_c))^{-1} \nabla f(x_c)$ is a descent direction as:

$$\nabla f(x_c)^T s_c = -\overbrace{\nabla f(x_c)^T \nabla^2 f(x_c) \nabla f(x_c)}^{>0} < 0.$$

Newton's method:

for $k = 1 \ldots$

$s_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

$x_{k+1} = x_k + s_k$

A func f is L iff:

$\|f(x) - f(y)\| \leq L \|x -$

Thm (Convergence of Newton's method)

Let $f: \mathbb{R}^n \to \mathbb{R}$ be twice cont. diff'ble and let $\nabla^2 f$ be Lipschitz cont.

If $x^* \in \mathbb{R}^n$ is a local min at which second order suff optimality cdt are satisfied then: $\exists \varepsilon > 0$ $\forall x_0 \in B_\varepsilon(x_*)$

Newton's method w/ start pt x_0 converges $x_k \to x^*$.

q-quadratically i.e.:

(quotient of errors)

$$\|x_{k+1} - x_* \|_2 \leq C \| x_k - x_* \|^2$$

Proof (by induction).

$$x_{k+1} - x^* = x_k + s_k - x^* = x_k - x^* - (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

$$= \nabla^2 f(x_k)^{-1} \Big[\underbrace{\nabla f(x^*)}_{0} - \nabla f(x_k) + \nabla^2 f(x_k)(x_k - x^*) \Big]$$

By Taylor's theorem:

$$\nabla f(x^*) - \nabla f(x_k) = \int_0^1 \nabla^2 f(x_k + t(x^* - x_k))(x^* - x_k)\, dt$$

Thus:

$$\| \nabla f(x^*) - \nabla f(x_k) + \nabla^2 f(x_k)(x_k - x^*) \|$$

$$= \Big\| \int_0^1 \big[\nabla^2 f(x_k + t(x^* - x_k)) - \nabla^2 f(x_k) \big] (x^* - x_k)\, dt \Big\|$$

$$\leq \int_0^1 \| \quad \| \, \| x^* - x_k \| \, dt$$

$$\overset{\text{L-cont}}{\leq} \int_0^1 L t \| x^* - x_k \|^2 dt = \frac{L}{2} \| x^* - x_k \|^2$$

Thus: $\| x_{k+1} - x^* \| \leq \| \nabla^2 f(x_k)^{-1} \| \frac{L}{2} \| x_k - x^* \|^2$

Now $\exists\, \varepsilon > 0$ s.t. $\forall x \in B_\varepsilon(x^*)$: $\| \nabla^2 f(x)^{-1} \| \leq 2 \| \nabla^2 f(x^*)^{-1} \|$

(proof comes next)

therefore: $\| x_{k+1} - x^* \| \leq \underbrace{\| \nabla^2 f(x^*) \| \, L}_{c} \| x_k - x^* \|^2$

$\Rightarrow$ q- quadratic LOCAL converge

Lemma Under same assumptions as previous theorem: ⑦

$\exists \varepsilon > 0 \ \forall x \in B_\varepsilon(x^*) \ \ \|D^2 f(x)^{-1}\| \leq 2 \|D^2 f(x^*)^{-1}\|$

proof:

$$D^2 f(x) = D^2 f(x^*) + D^2 f(x) - D^2 f(x^*)$$

$$v^T D^2 f(x) v = v^T D^2 f(x^*) v + v^T (D^2 f(x) - D^2 f(x^*)) v$$

$$\geq v^T D^2 f(x^*) v - | \quad \checkmark \quad |$$

$$\geq v^T D^2 f(x^*) v - \underbrace{\| D^2 f(x) - D^2 f(x^*) \|_2}_{\leq L \|x - x^*\|} \|v\|_2^2$$

Denote by $\lambda_{min}(x)$ = smallest eigenvalue of $D^2 f(x)$ then:

$$\lambda_{min}(x) = \min_{v \neq 0} \frac{v^T D^2 f(x) v}{v^T v} \geq \underbrace{\min_{v \neq 0} \frac{v^T D^2 f(x^*) v}{v^T v}}_{\lambda_{min}(x^*)} - L \|x - x^*\|$$

take $\|x - x^*\| \leq \dfrac{\lambda_{min}(x^*)}{2L}$ then.

$$\lambda_{min}(x) \geq \lambda_{min}(x^*) - \frac{\lambda_{min}(x^*)}{2} = \frac{\lambda_{min}(x^*)}{2}$$

thus:

$$\underset{2 \| D^2 f(x^*)^{-1} \|}{\overset{\|}{2 \lambda_{min}(x^*)^{-1}}} \geq \underset{\| D^2 f(x)^{-1} \|}{\overset{\|}{\lambda_{min}(x)^{-1}}}$$

QED