**Introduction**

General optimization problem: \( \text{min } f(x) \)
\[ x \in S \]
\( f : S \to \mathbb{R} \) = objective function
\( x \) = variable, unknown
\( S \) = feasible region

**Def:** \( x^* \) is a strict minimizer of \( f \) on \( S \) if
\[ f(x^*) < f(x) \quad \forall x \in S \quad [x \neq x^*] \]

**Convention:** We only look at minimization problems as
\[ \max f(x) \quad \Leftrightarrow \quad \min -f(x) \]
\[ x \in S \]

**Important questions:**
1. existence of a sol
2. uniqueness of a sol
3. characterization of a sol
4. Numerical approx of sol

**S** can be a discrete set = discrete optimization problem (another class)
Here we shall assume that: for \( g : R^n \to \mathbb{R}^m \) & \( R : \mathbb{R}^n \to \mathbb{R}^p \)
\[ S = \{ x : \begin{array}{l} g(x) = 0, \quad f(x) > 0 \end{array} \text{ and } \begin{array}{l} \text{equal constr.} \quad \text{inequal constr.} \end{array} \} \]

this is continuum optimization.

Many real-world problems can be formulated (modeled) as
a continuum optimization problem. Here are some examples:
Examples:

- off exploration
- medical imaging
- economy
- design

find mean velocity that better fits observed measurements
find conductivity inside S2 that better fits electrical measurements at boundary
find portfolio that max returns while keeping risk within a certain level
find wing cross section that maximizes lift

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

\[
\min_{x \in \mathbb{R}} f(x)
\]

- First order necessary
  If $x^*$ minimizes $f$, then $f'(x^*) = 0$

- Second order necessary
  If $x^*$ minimizes $f$, then $f''(x^*) > 0$

- Second order sufficiency:
  If $x^* \in \mathbb{R}$ s.t. $f(x^*) = 0$

  $f''(x^*) > 0$

  then $x^*$ is a strict local minimizer of $f$

  $\epsilon > 0 \quad \forall x^* \quad |x - x^*| < \epsilon \Rightarrow f(x^*) < f(x)$

Cand on curvature basically means that Taylor approx to 2nd order:

$f(x + h) = f(x^*) + f'(x^*)h + \frac{1}{2}f''(x^*)h^2$

Looks like parabola

$f(x) = x^4$ example when sufficiency not valid.
**Existence:**
If $f$ is continuous on the closed interval $[a,b]$, then $f$ has at least one minimizer on $[a,b]$.

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

**Derivation of Newton's method on $\mathbb{R}$**

Given current approx $x_c$ to $x^*$ minimizer of $f(x)$, find best update.

**Taylor's theorem:**

\[
f(x_c + \delta) = f(x_c) + f'(x_c) \delta + \frac{1}{2} f''(x_c) \delta^2 + o(\delta^2)
\]

where $\lim_{\delta \to 0} \frac{o(\delta^2)}{\delta^2} = 0$.

Or in other words:

\[
r(\delta^2) = o(\delta^2)
\]

\[
\therefore f(x_c) + f'(x_c) \delta + \frac{1}{2} f''(x_c) \delta^2 = m(\delta)
\]

**Idea:** minimize quadratic model $m_c(\delta)$ instead of $f$.

\[
\min_{\delta} f(x_c + \delta) \approx \min_{\delta} m_c(\delta)
\]

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

\[
\delta_c = -\frac{f'(x_c)}{f''(x_c)}
\]

Thus we take iteration $x_{k+1} = x_k + \delta_c$ or in algorithmic notation:

choose $x_0$ initial point

\[
\begin{align*}
\text{for } k = 0 \ldots \delta_c \quad x_{k+1} &= x_k - \frac{f'(x_c)}{f''(x_c)}
\end{align*}
\]
Convergence of Newton method

Let $f$ be smooth and assume $x^*$ is a minimizer of $f$ satisfying the 2nd order sufficient condition (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 we 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 needed for $\mathbb{R}^n$

- **Gradient**: $Df(x) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$

- **Hessian**: $D^2f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$

- **Symmetric Matrix**: $D^2f(x)$ is symmetric since $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$

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

- **Matrix-Vector Multiplication**: $(A x)^T = \sum_{j=1}^{n} A_{ij} x_j$

- **Taylor's Theorem**

Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be continuously differentiable then:

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

leading to equivalence:

$$
f(x + p) \approx f(x) + Df(x)^T p + R_f(p; x) \text{ s.t. } \lim_{p \to 0} \frac{R_f(p; x)}{\|p\|} = 0
$$
If $f$ is twice and differentiable then:

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

or equivalently:

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

with

$$ \lim_{p \to 0} \frac{R_2(p; x)}{p^T p} = 0 $$

(residual is negligible)

Fundamental Theorem of Calculus

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

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

$$ q(x) = c + g^T x + \frac{1}{2} x^T H x $$

is a quadratic function in $\mathbb{R}^n$.

We assume $H$ is symmetric.

**Gradient computation (Taylor based)**

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

$$ = \lim_{t \to 0} \frac{tg^T \nu + \frac{1}{2} t \nu^T H \nu + \frac{1}{2} \nu^T H \nu}{t} $$

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

$$ = \nabla q(x) = g + H x \quad \text{(by identification)}$$
Hessian computation (Taylor, feed)

\[
\frac{1}{2} \mathbf{v}^T \mathbf{D}^2 q(x) \mathbf{v} = \lim_{t \to 0} \frac{q(x + t\mathbf{v}) - q(x) - \nabla q(x)^T t \mathbf{v}}{t^2} = \lim_{t \to 0} \frac{1}{2} \frac{t^2 \mathbf{v}^T \mathbf{H} \mathbf{v}}{t^2} = \frac{1}{2} \mathbf{v}^T \mathbf{H} \mathbf{v}
\]

\[\Rightarrow \nabla^2 q(x) = \mathbf{H}\] by identification (works for vector matrix only)

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

**Def**: Hessian is said to be symmetric positive definite if:

\[\forall \mathbf{v} \in \mathbb{R}^n \quad \mathbf{v}^T \mathbf{H} \mathbf{v} > 0 \quad \mathbf{v} \neq 0\]

**H** symmetric positive semidefinite if:

\[\forall \mathbf{v} \in \mathbb{R}^n \quad \mathbf{v}^T \mathbf{H} \mathbf{v} \geq 0\]

[Note: symmetric def.: for inequalities]

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

- \( \mathbf{H} \) pos def \( \Rightarrow \) Bawl

- \( \mathbf{H} \) pos semidef \( \Rightarrow \) degenerate bowl: \( q(x)\)
(\lambda, \mathbf{u}) \text{ is an eigenpair of } A \text{ if } \mathbf{u} \neq \mathbf{0} \text{ and } A\mathbf{u} = \lambda \mathbf{u}

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

- \( A \) is real: \( \mathbf{u}^T A \mathbf{u} = A \mathbf{u}^T \mathbf{u} \)

If \((\lambda, \mathbf{u}) (\mu, \mathbf{v})\) are eigenpairs of \( A \) with \( \mathbf{f} \neq \mathbf{v} \), \( \mathbf{u}^T \mathbf{v} = 0 \)

Rayleigh quotient:

\[
\begin{align*}
\lambda_{\min}(A) &\leq \frac{\mathbf{u}^T A \mathbf{u}}{\mathbf{u}^T \mathbf{u}} \leq \lambda_{\max}(A) \quad \text{for } \mathbf{u} \neq \mathbf{0} \\
\lambda_{\max}(A) &= \max_{\mathbf{u} \neq \mathbf{0}} \frac{\mathbf{u}^T A \mathbf{u}}{\mathbf{u}^T \mathbf{u}} \\
\lambda_{\min}(A) &= \min_{\mathbf{u} \neq \mathbf{0}} \frac{\mathbf{u}^T A \mathbf{u}}{\mathbf{u}^T \mathbf{u}}
\end{align*}
\]

In quadratics, eigenvectors give axes of hyperboloid.

- \( A \) symmetric positive definite: \( \lambda_1(A) > 0 \)
- \( A \) positive semidefinite: \( \lambda_0(A) \geq 0 \).
Norms:

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

\[ \|x\|_1 = \sum_{i=1}^{n} |x_i| = \ell_1 \text{ norm} \]

\[ \|x\|_\infty = \max_{i=1...n} |x_i| \quad \text{Infinity norm} \]

Matrix norms:

\[ \|A\|_F = \left( \sum_{i,j=1}^{n} a_{ij}^2 \right)^{1/2} \quad \text{(simplified: view matrix as a vector)} \]

Induced matrix norm:

\[ \|A\|_2 = \max_{\|u\|_2 = 1} \|Au\|_2 \]

for \( \|\cdot\|_2 \) norm:

\[ \|A\|_2 = \max_{\|u\|_2 = 1} \|Au\|_2 \] (See A.1 for other induced norm)

For A symmetric:

\[ \|A\|_2^2 = \max_{\|u\|_2 = 1} \|Au\|_2^2 = \max_{\|u\|_2 = 1} \frac{u^T A^2 u}{u^T u} = \lambda_{\max}(A^2) \]

\[ \Rightarrow \|A\|_2 = \sqrt{\lambda_{\max}(A^2)} \]

Inequalities:

Cauchy Schwarz: \( \|Au\|_1 \leq \|A\|_1 \|u\|_1 \)

\[ \|Au\|_1 \leq \|A\|_1 \|u\|_1 \]

\( \|Au\|_2 \leq \|A\|_2 \|u\|_2 \) for induced matrix norm.

Triangle inequality:

\[ \|u + v\|_1 \leq \|u\|_1 + \|v\|_1 \]

\[ \|u + v\|_2 \leq \|u\|_2 + \|v\|_2 \]
§2 Unconstrained optimization

(1) \[ \min_{x \in F} f(x) \quad \text{where} \quad F \subset \mathbb{R}^n \text{ is a feasible set} \]

and \( f: \mathbb{R}^n \to \mathbb{R} \) is a smooth function

Def: \( x^\ast \) is called a global [strict] minimum (solution) of (1) if

\[ f(x^\ast) \leq f(x) \quad \forall x \in F \quad [\leq] \]

or \( f(x^\ast) < f(x) \quad \forall x \in F \quad [<] \)

\( x^\ast \) is called a local [strict] minimum (sol) of (1) if

\[ \exists \varepsilon > 0 \quad f(x^\ast) \leq f(x) \quad \forall x \in B_\varepsilon(x^\ast) \cap F \quad [\leq] \]

Existence of a minimizer

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

Note: in \( \mathbb{R}^n \) "compact set \( \Rightarrow \) closed and bounded.

However \( \mathbb{R}^n \) is unbounded.

Thm: Let \( F = \mathbb{R}^n \). Let \( f: \mathbb{R}^n \to \mathbb{R} \) be continuous and have the "infinity property":

\[ \|x\| \to \infty \quad \Rightarrow \quad f(x) \to +\infty \]

then \( \exists x^\ast \text{ s.t. } f(x^\ast) = \inf_{x \in \mathbb{R}^n} f(x) \quad (\text{makes linear } \{ x \mid f(x) < f(x_0) \text{ compact})} \]

In many cases the objective function is of the form:

\[ f(x) = g(x) + \alpha \|x\|^2 \]

with \( g(x) \) bounded below.

Thus existence of solution is guaranteed (sometimes term is added to guarantee existence of sol).
Characterization of local minima

Taylor's theorem

Let \( f: \mathbb{R}^n \to \mathbb{R} \) be continuously differentiable then:

\[ \exists t \in (0,1) \text{ st. } f(x+t) = f(x) + \nabla f(x) \cdot t \]

If \( f \) is twice continuously differentiable then:

\[ \nabla f(x+t) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tv) \cdot v \, dt \]

and

\[ \exists t \in (0,1) \text{ s.t. } f(x+t) = f(x) + \nabla f(x) \cdot t + \frac{1}{2} t^2 \nabla^2 f(x) \cdot t \]

Necessary optimality condition:

Thm (First order necessary optimality condition)

If \( x^* \) is a minimizer and \( f \) is continuously differentiable around \( x^* \) then \( \nabla f(x^*) = 0 \).

Proof:

Suppose for contradiction, \( \nabla f(x^*) \neq 0 \).

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

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

By continuity of \( \nabla f(x) \), \( \exists T \text{ s.t. } \forall t \in (0,T) : \)

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

Now by Taylor's theorem:

\[ \forall t \in [0,T], \exists \epsilon \in (0,1) \text{ s.t. } f(x^*+\epsilon p) = f(x^*) + \epsilon \nabla f(x^*+tp) \cdot p \]

\[ \leq f(x^*+tp) < f(x^*) \implies \text{contradiction} \]
Thus (Second order necessary optimality conditions)
If $x^*$ is a local minimizer of $f$ and $\nabla^2 f$ exists and is continuous in a neighborhood of $x^*$, then:
$$\nabla f(x^*) = 0 \quad \text{and} \quad \nabla^2 f(x^*) \text{ is pos. semidef.}$$
Proof: Taylor's theorem, see book.

Recall $A \in \mathbb{R}^{n \times n}$, symmetric or pos. semidef. $\iff \forall \alpha \in \mathbb{R}^n \quad \alpha^T A \alpha \geq 0 \iff \lambda_{\min}(A) > 0.$

Then (Second-order sufficient optimality conditions)
Suppose $\nabla^2 f$ is continuous in a neighborhood of $x^*$.
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*)$ is pos. definite
$\Rightarrow x^*$ is a strict local minimizer of $f$.
Proof: see book, also based on Taylor's theorem.

Note: There is a gap between necessary and sufficient conditions for optimality.
I.e. not all minimizers satisfy sufficient conditions.

Simplest example: $f(x) = x^4, \quad \nabla^2 f(x) = 12x^2, \quad \nabla^2 f(0) = 0$

Convexity:
$f$ is convex iff $\forall x, y \in \mathbb{R}^n \forall \lambda \geq 0, \quad f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$

Theorem: When $f$ is convex: $x^*$ local minimizer $\Rightarrow x^*$ global minimizer.
Proof: By contradiction.
If $x^* = \text{local min.}$, but $x^*$ is not a global min then
$$\exists \exists x \quad f(z) < f(x^*)$$
Consider line segment joining $x^*$ to $z$: 

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

Since $f$ is convex:

$$f(z) \leq tf(x) + (1-t) f(x^*) \leq f(x^*)$$

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

thus $x^*$ is not a local min.

**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

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

**Quadratic optimization problems**

$$f(x) = \frac{1}{2} x^T H x$$

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

$$\nabla^2 f(x) = H$$

(NH = symmetric)

For order nec. cond: $x^*$ min = $\nabla^2 f(x^*) = -g$

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

To numerically have a minimizer:

then:

To solve $\min f(x) \iff g + H x^* = 0$ and $H$ is positive definite

$H$ pos def $\implies$ there is a unique sol.
Newton's method \[ \min_{x \in \mathbb{R}^n} f(x), \quad f : \mathbb{R}^n \to \mathbb{R} \]

Given approximate \( x_c \) of \( x^* \):
\[
f(x_c + s) = f(x_c) + \nabla f(x_c)^T s + \frac{1}{2} s^T \nabla^2 f(x_c) s + R_c(x_c, s)
\]

where \( R_c \) is small residual
\[
\approx f(x_c) + \nabla f(x_c)^T s + \frac{1}{2} s^T \nabla^2 f(x_c) s = m_c(s) + o(||s||^2)
\]

\( m_c(s) \) = model of \( f_c \) of \( f \)

Minimize model instead of \( f \):
\[
\min_{s} f_c(x_c + s) \approx \min_{s} m_c(s)
\]

If gradient
\[
\nabla^2 f(x_c) s_c = -\nabla f(x_c)
\]
and
\[
\nabla^2 f(x_c) \text{ is pos semi-def then } s_c \text{ is a min of } m_c.
\]

\[ \text{take } \quad x_{c+1} = x_c + s_c \]

If \( \nabla^2 f(x_c) \) is pos def then \( s_c = -\left(\nabla^2 f(x_c)\right)^{-1} \nabla f(x_c) \) is a
descent direction as:
\[
\nabla f(x_c)^T s_c = -\nabla f(x_c)^T \nabla^2 f(x_c) \nabla f(x_c) s_c < 0
\]

\[ \text{Newton's method } \]

\[
\begin{align*}
    s_k &= -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k) \\
    x_{k+1} &= x_k + s_k
\end{align*}
\]

Then (Convergence of Newton's method)
If \( f : \mathbb{R}^n \to \mathbb{R} \) be twice and differentiable and let \( \nabla^2 f \) be Lipschitz cont.
If \( x^* \in \mathbb{R}^n \) is a local min at which second order suff optimality cond are
satisfied then \( \exists \delta > 0 \) \( \forall x \in B_\varepsilon(x^*) \)

Newton's method will start pt \( x_0 \) converges \( x_k \to x^* \)

\( q \)-quadratically
\[ \| x_{k+1} - x^* \|_2 \leq C \| x_k - x^* \|^2. \]

**Proof (by induction).**

\[
x_{k+1} - x^* = x_k + \Delta x_k - x^* = x_k - x^* - (\nabla^2 f(x_k))^{-1} \Delta \nabla f(x_k)
\]

\[
= \nabla^2 f(x_k)^{-1} \left[ \nabla^2 f(x^*) \Delta \nabla f(x_k) + \nabla^2 f(x_k) (x_k - x^*) \right]
\]

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:

\[
\| \nabla f(x^*) - \nabla f(x_k) \|_2 \leq \int_0^1 \| \nabla^2 f(x_k + t(x^* - x_k)) \|_2 (x^* - x_k) \, dt
\]

\[
\leq \int_0^1 \| x^* - x_k \|_2 \, dt
\]

\[
\leq \int_0^1 \left[ L + \| x^* - x_k \|_2 \right] \, dt = \frac{1}{2} \| x^* - x_k \|^2.
\]

Thus:

\[
\| x_{k+1} - x^* \| \leq \| \nabla^2 f(x_k)^{-1} \| \frac{L}{2} \| x_k - x^* \|^2.
\]

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

(proof comes next)

therefore:

\[
\| x_{k+1} - x^* \| \leq \| \nabla^2 f(x_k)^{-1} \| L \| x_k - x^* \|^2
\]

\[
\leq \delta \implies \text{quadratic LOCAL convergence}
\]
Lemma. Under some assumptions as previous theorem:

\[ \forall x \in B_r(x^*) \quad \| V^2 f(x) \|^2 \leq L \| V^2 f(x^*) \| \]

Proof:

\[ V^2 f(x) = V^2 f(x^*) + V^2 f(x) - V^2 f(x^*) \]

\[ V^T V^2 f(x) V = V^T V^2 f(x^*) V + V^T (V^2 f(x) - V^2 f(x^*)) \]

\[ \geq V^T V^2 f(x^*) V - \frac{2 L}{\lambda_{\min}(x^*)} \]

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

\[ \lambda_{\min}(x) = \min_{V \neq 0} \frac{V^T V^2 f(x) V}{V^T V} \geq \min_{V \neq 0} \frac{V^T V^2 f(x^*) V}{V^T V} - \frac{2 L}{\lambda_{\min}(x^*)} \]

Take \( \| x - x^* \| \leq \lambda_{\min}(x^*) \) then:

\[ \lambda_{\min}(x) \geq \lambda_{\min}(x^*) - \frac{2 L}{\lambda_{\min}(x^*)} = \frac{\lambda_{\min}(x^*)}{2} \]

Thus:

\[ 2 \lambda_{\min}(x^*) \geq \lambda_{\min}(x) \]

\[ 2 \| V^2 f(x^*) \| \leq \| V^2 f(x) \| \]

\[ \boxed{\text{DED}} \]