1. INTRODUCTION

Optimization models

Optimization = find optimal = best = maximum or minimum

General optimization problem: max/min $f(x)$ $\leftarrow$ *objective function*, real-valued
subject to $x = (x_1, x_2, \ldots, x_n) \in S$ $\leftarrow$ *feasible set*, often described by *constraints*

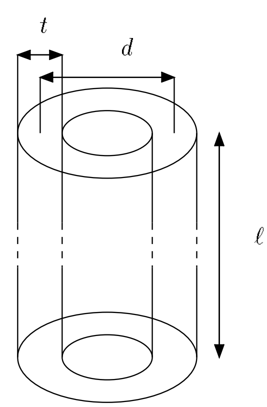
Variables $x_1, x_2, \ldots$ here could be anything.

Types of optimization:

- Discrete, e.g. Travelling Salesman, Shortest Path, ...
- Mixed
- Continuous
  - Linear: special theory, applications to discrete optimization (Math 4/6620).
  - Nonlinear: complicated real-world situations; physical quantities; production and design problems (this class).

Mathematical theory of optimization now important in practical situations because of availability of computers and software to handle realistic problems, with thousands of variables.

Example: A uniform tubular column must handle a compressive load of $P = 25,000$ N (newton). The column is to be made of a material with yield stress $\sigma_y = 5,000$ N/cm^2, modulus of elasticity $E = 8.5 \times 10^6$ N/cm^2, and weight density $\rho = 2.0 \times 10^{-2}$ N/cm^3. The length is to be $\ell = 250$ cm. The mean diameter d must be between 2 cm and 14 cm, and the thickness t between 0.2 cm and 0.8 cm. The induced stress $\sigma_i = P/(\pi d t)$ must not exceed either σ_y or the buckling stress $\sigma_b = \pi^2 E(d^2 + t^2)/(8\ell^2)$. Design the column to minimize its overall cost, which is $c = 0.5W + 2d$, where $W = \pi \ell d t \rho$ is the weight (in N) and d is the mean diameter (in cm).



Steps:

(1) Variables: choose 'design variables', those over which you have control.

d (cm), t (cm). Specifying units is important!

(2) Objective: formulate objective function as function of design variables and determine whether it is to be maximized or minimized.

$$\begin{aligned} \min\ c &= 0.5W + 2d = 0.5\pi \ell d t \rho + 2d \\ &= 0.5\pi(250) d t (2.0 \times 10^{-2}) + 2d \\ &= 2.5\pi d t + 2d. \end{aligned}$$

(3) Constraints: formulate restrictions given in problem as equations or inequalities involving design variables (including perhaps upper and lower bounds on the variables).

Bounds:

$d \geq 2$

$$\begin{aligned} & d \le 14 \\ & t \ge 0.2 \\ & t \le 0.8 \end{aligned}$$

$$\sigma_i \le \sigma_y: \quad \begin{aligned} \frac{P}{\pi dt} &\le 5000 \\ \frac{25000}{\pi dt} &\le 5000 \\ \frac{5}{\pi dt} &\le 1 \end{aligned} \qquad \text{or (since } d \text{ and } t \text{ positive)} \qquad dt \ge 5/\pi.$$

- LHS now polynomial in variables, easy to compute, linear in each variable so behaves nicely with respect to approximation.

$$\sigma_i \le \sigma_b: \quad \begin{aligned} \frac{P}{\pi dt} &\le \frac{\pi^2 E(d^2+t^2)}{8\ell^2} \\ \frac{25000}{\pi dt} &\le \frac{\pi^2(8.5\times 10^6)}{8(250^2)}(d^2+t^2) \\ \frac{10^5}{\pi dt} &\le 68\pi^2(d^2+t^2) \end{aligned} \qquad \text{or (since } d \text{ and } t \text{ positive)} \qquad dt(d^2+t^2) \ge \frac{10^5}{68\pi^3}.$$

- Again perhaps nicer because LHS polynomial. Or maybe original better because both sides lower degree? Experiment!

Message: same constraint can be written in different ways; some may work better than others.

Example: Curve fitting: Find best line $y = a_0 + a_1 x$ through $(1,3)$, $(3,5)$ and $(4,7)$. Not collinear!
Variables: a_0, a_1 (dimension-free, no unit).

Objective: First need to decide how to measure 'best'. Convenient choice: minimize sum of squares of errors - least squares. Other reasonable choices: sum of absolute values of errors, maximum of absolute values of errors. Advantages of least squares:

(1) Differentiable (sum or maximum of absolute values is not).
(2) Can be justified on statistical grounds: if errors normally distributed, corresponds to 'maximum likelihood estimate'.
(3) Easy to solve.

Write $f(x) = a_0 + a_1 x$. In applications, think of f as model function, known values = measurements with errors.

Residual (error = measured versus model value)

at $x = 1$: $3 - f(1) = 3 - (a_0 + a_1)$,
at $x = 3$: $5 - f(3) = 5 - (a_0 + 3a_1)$,
at $x = 4$: $7 - f(4) = 7 - (a_0 + 4a_1)$.

So objective is

$$\min((3-(a_0+a_1))^2 + (5-(a_0+3a_1))^2 + (7-(a_0+4a_1))^2$$

Constraints: None. Problem is *unconstrained*.

Linear least squares problems easily solved via linear algebra, in theory. (See problems.) For large problems must address practical issues of numerical linear algebra.

Standard nonlinear optimization problem:

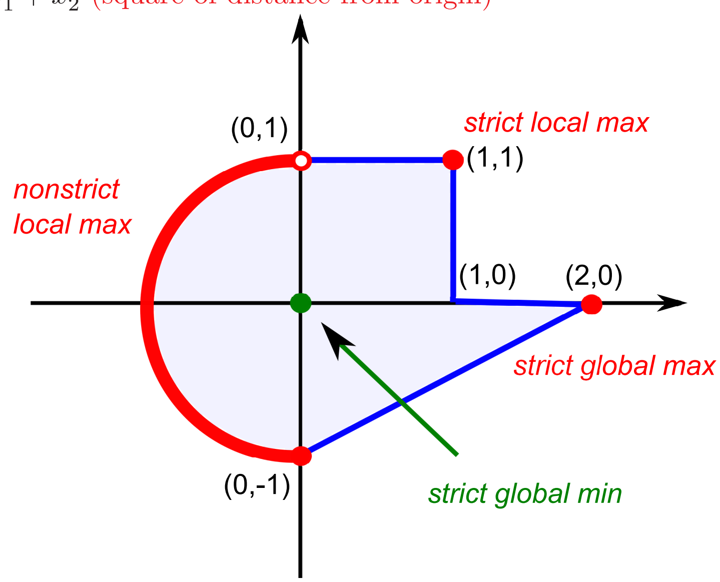
$$\begin{array}{rlll} \min & f(x) & x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n & \\ \text{such that} & g_i(x) = 0, & i \in \mathcal{E} & \textit{equality constraints}, \\ & g_i(x) \ge 0, & i \in \mathcal{I} & \textit{inequality constraints}. \end{array}$$

Constraints define *feasible set*.
– If max, negate objective function.
– Move all expressions in constraints to LHS.
– If $\leq$, negate.
– All constraints assumed to have *nonstrict* inequalities so that feasible region is *closed* if g_i's are continuous. Avoids problems where desired point on boundary of set but not feasible.

Conditions for optimality (need some terminology first)

Suppose f is a real-valued function on a set $S \subseteq \mathbf{R}^n$. Then $x^* \in S$ is a
global minimizer of f on S if $f(x) \geq f(x^*)$ $\forall\, x \in S$;
strict global minimizer of f on S if $f(x) > f(x^*)$ $\forall\, x \in S$ with $x \neq x^*$;
local minimizer of f on S if $\exists\, \varepsilon > 0$ so that $f(x) \geq f(x^*)$ $\forall\, x \in S$ with $\|x - x^*\| < \varepsilon$;
strict local minimizer of f on S if $\exists\, \varepsilon > 0$ so that $f(x) > f(x^*)$ $\forall\, x \in S$ with $\|x - x^*\| < \varepsilon$ and $x \neq x^*$;
Similar definitions for *maximizers*: reverse inequalities for f values.

Example: $f(x) = x_1^2 + x_2^2$ (square of distance from origin)



local/global minimum value of f = value of f at local/global minimizer. Minimizer is $x \in \mathbf{R}^n$, minimum value is $f(x) \in \mathbf{R}$.

How do we find minimizers? For continuous function of one variable on closed interval, need to look at (a) critical points ($f' = 0$), and (b) boundary points (ends). Worry about (b) later; for now, what about minimizers in interior of feasible set?

$B(x_0, \varepsilon) = \{x \in \mathbf{R}^n \mid \|x - x_0\| < \varepsilon\}$ – *open ball of radius ε around x_0*.
interior point x_0 of S: $\exists\, \varepsilon > 0$ such that $B(x_0, \varepsilon) \subseteq S$ (hence $x_0 \in S$) (positive distance to $\overline{S} = \mathbf{R}^n - S$).
boundary point x_0 of S: $\forall \varepsilon > 0$, $B(x_0, \varepsilon)$ intersects both S and $\overline{S} = \mathbf{R}^n - S$ (may have $x_0 \in S$ or $x_0 \notin S$) (arbitrarily close to both S and $\overline{S}$).
closed set S contains all its boundary points.
gradient of f is $\nabla f = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}]^{\mathrm{T}}$.
NOTE: All vectors are column vectors!

Theorem (First Order Necessary Condition): Suppose x^* is a local minimizer of f on $S \subseteq \mathbf{R}^n$ and x^* is an interior point of S. If all partial derivatives of f exist at x^*, then $\nabla f(x^*) = 0$. In other words, all partial derivatives are 0 at x^*.
Proof: For each i consider $f_i(x_i) = f(x_1^*, \ldots, x_{i-1}^*, x_i, x_{i+1}^*, \ldots, x_n^*)$. Since x^* minimizes f, x_i^* must minimize f_i, so by first-year calculus $0 = f_i'(x_i^*) = \frac{\partial f}{\partial x_i}(x^*)$.

Reformulation: Any local minimizer of f on S must occur (i) where $\nabla f = 0$, or (ii) where ∇f does not exist, or (iii) at a boundary point.
Stationary point is where $\nabla f = 0$; *critical point* is where $\nabla f = 0$ or ∇f does not exist.

Given a stationary point, how can we tell if it's really a minimizer? Need something like second deriv. test for funtions of one variable. To get idea, look at multivariable version of Taylor series.

Taylor series for $f(x) = f(x_1, x_2, \ldots, x_n)$: if f has continuous 3rd derivatives,

$$f(x) = f(x_0) + (x - x_0)^{\mathrm{T}} \nabla f(x_0) + \frac{1}{2}(x - x_0)^{\mathrm{T}} \nabla^2 f(x_0)(x - x_0) + O(\|x - x_0\|^3).$$

– x_0, x are column vectors;

– $\nabla^2 f$ = *Hessian matrix* = $\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$; symmetric if f has continuous 2nd derivatives.

So we need to look at expressions like (something)$^{\mathrm{T}}$ $\nabla^2 f$ (something). Some relevant definitions:

A real symmetric $n \times n$ matrix A is defined to be:
positive definite if $v^{\mathrm{T}} A v > 0 \;\forall\, v \in \mathbf{R}^n$ with $v \neq 0$; all eigenvalues > 0
positive semidefinite if $v^{\mathrm{T}} A v \geq 0 \;\forall\, v \in \mathbf{R}^n$; all eigenvalues ≥ 0
negative definite if $v^{\mathrm{T}} A v < 0 \;\forall\, v \in \mathbf{R}^n$ with $v \neq 0$; all eigenvalues < 0
negative semidefinite if $v^{\mathrm{T}} A v \leq 0 \;\forall\, v \in \mathbf{R}^n$; all eigenvalues ≤ 0
indefinite if $\exists\; v_1, v_2 \in \mathbf{R}^n$ with $v_1^{\mathrm{T}} A v_1 > 0$ and $v_2^{\mathrm{T}} A v_2 < 0$. some eigenvalue > 0, some eigenvalue < 0

A real symmetric matrix A (a) has all eigenvalues real, and (b) is *diagonalizable*: $A = U^{-1} D U$ where D = diagonal matrix with eigenvalues of A down diagonal, U = *orthogonal* matrix ($U^{-1} = U^{\mathrm{T}}$). Above properties can be represented in terms of eigenvalues.
Also third way to think of positive definiteness using determinants of submatrics; good for computation; see 1E.

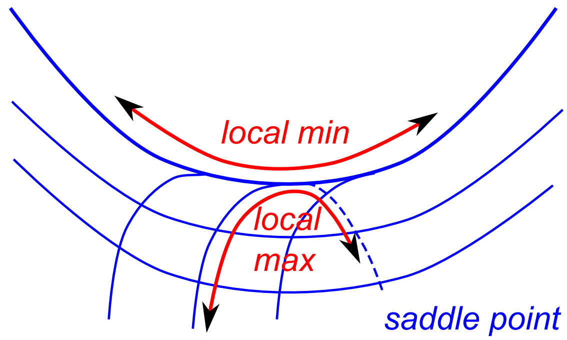
Second Order Conditions: Suppose that x^* is an interior point of S at which $\nabla f(x^*) = 0$, and so that f has continuous second derivatives in an open ball $B(x^*, \varepsilon) \subseteq S$ for some $\varepsilon > 0$.

(i) **Necessary condition for minimizer:** If x^* is a local minimizer of f on S, then $\nabla^2 f(x^*)$ is positive semidefinite.

(ii) **Sufficient condition for minimizer:** If $\nabla^2 f(x^*)$ is positive definite, then x^* is a strict local minimizer of f on S.

Note: necessary and sufficient conditions are different.

(iii) **Sufficient condition for saddle point:** If $\nabla^2 f(x^*)$ is indefinite then x^* is a *saddle point* of f: there is a direction in which f has a strict local minimizer at x^*, and a direction in which f has a strict local maximizer at x^*. Some people use much weaker definition of saddle point: stationary but not local maximizer or minimizer.



(Figure above is a 3-dimensional graph of a function of 2 variables.)

– Counterparts of (i) and (ii) for maximizers: replace ‘minimizer’ by ‘maximizer’ and ‘positive’ by ‘negative’.

– Converses of (i), (ii) and (iii) are not true.

Example: $f(x, y) = 2x^2 + 6xy + 5y^2 - 2x - 8y + 10$ on all $\mathbf{R}^2$.

Critical point: $\nabla f(x, y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 4x + 6y - 2 \\ 6x + 10y - 8 \end{bmatrix} = 0$

which gives $\begin{array}{rcrcl} 4x &+& 6y &=& 2 \quad (1) \\ 6y &+& 10y &=& 8 \quad (2) \end{array}$ and $(2) - \frac{3}{2}(1)$ gives $y = 5$ and then we get $x = -7$.

So there is one critical point, at $(-7, 5)$.

Hessian $A = \nabla^2 f(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 10 \end{bmatrix}$

Eigenvalues:

$$\begin{aligned} \det(\lambda I - A) &= \begin{vmatrix} \lambda - 4 & -6 \\ -6 & \lambda - 10 \end{vmatrix} \\ &= (\lambda - 4)(\lambda - 1) - (-6)^2 = \lambda^2 - 14\lambda + 40 - 36 \\ &= \lambda^2 - 14\lambda + 4 = \lambda^2 - 14\lambda + 49 - 45 = (\lambda - 7)^2 - 45 \\ &= (\lambda - 7 + \sqrt{45})(\lambda - 7 - \sqrt{45}) = 0 \end{aligned}$$

and hence $\lambda = 7 \pm \sqrt{45}$, both > 0, so $\nabla^2 f(-7, 5)$ is positive definite, $(-7, 5)$ is strict local minimizer.

In fact, from properties of quadratics, actually strict global minimizer.

Idea of definiteness generalizes test for functions of two variables: condition for local minimizer is $f_{xx} f_{yy} - f_{xy}^2 > 0$ and $f_{xx} > 0$; really checking positive definiteness using determinants of submatrices.

Example: $f(x, y) = x^2 + 6xy + 9y^2 + 4x + 12y - 3$:
$\nabla f = 0$ everywhere along the line $x + 3y = -2$, e.g. at $(1, -1)$;
$\nabla^2 f = \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$ everywhere, eigenvalues $\lambda = 0$, 20, so positive semidefinite.

So cannot say for sure what happens at $(1, -1)$ or other points on line. Can say it’s not a local maximizer since $\nabla^2 f$ not negative semidefinite. In fact all points on line are nonstrict global minimizers.

Local versus global optimality

Usually want *global* minimizer of f on feasible set S. May not exist; if exists, may be hard to find.

Theory

If f continuous, S is closed bounded set then a global minimizer (and global maximizer) of f exists (on boundary or where ∇f is 0 or doesn't exist).

If $S = \mathbf{R}^n$ and f is continuous and *coercive* ($\lim_{\|x\|\to\infty} f(x) = +\infty$) then a global minimizer of f exists, at a critical point. Just apply previous point with S being set where f is smaller than something.

There are conditions for global minimizers involving *convexity* (more details soon), related to properties of $\nabla^2 f$. For example, if $S = \mathbf{R}^n$ and $\nabla^2 f$ is positive semidefinite (definite) everywhere then any stationary point ($\nabla f = 0$) is a (strict) global minimizer.

Methods

Chop S into smaller regions where f can be analysed. Use ideas like 'bisection' and 'branch and bound', using 'interval arithmetic' and local minimizers to get bounds. Sometimes can guarantee we really have global minimizer.

Seach methods that allow non-descent steps, have probabilistic aspect: e.g., taboo search, simulated annealling.

Use population of points spread out over S: genetic algorithms, particle swarm, ant colony.

... and others; see 'Handbook of Global Optimization', volumes 1 and 2.

In this course focus on finding *local* minimizers.

– global opt. methods often rely on local opt. methods;

– in particular situation often have theory (convexity) or practical reaons for expecting local minimizer to also be global minimizer.