Week 1

Note Title 3/29/2021

An optimization problem is about maximizing/minimizing a quantity that varies with a set of variables.

This quantity to be optimized is usually call the objective function.

The set of variables may consist of finite or infinite degrees of freedom. Moreover, these degrees of freedom may be subject to constraints.

The objective function, dependent on as many variables as there are in the problem, is always scalar-valued. (We can only talk about 'bigger' or 'smaller' for a single quantity.)

Aside: there is something called multi-objective optimization problem, in which more than one quantities seek to be optimized (e.g. we may want to be maximize both wealth and happiness). In this case, an 'optimal solution' requires a kind of `optimal tradeoff' among the different objectives. We shall not discuss this type of optimization problem in this course.

Example 1

Among all the cylinders with the same surface area, find the one with the largest capacity.

Solution: The shape of a cylinder is determined by two degrees of freedom, its height (h) and radius (r)

Surface area : $A(h,r) = 2\pi r^2 + 2\pi r h = 2\pi r(r+h)$

Volume : $V(h,r) = \pi r^2 h$

objective function

constraint

Optimization problem : $\max\limits_{h,r} \pi r^2 h$ s.t. $\underbrace{2\pi r(r+h)}_{\text{constraint}} = A_0$

Some fixed value

# of d.o.f. : 2

# of constraint : 1

In this case, we can easily convert the optimizaton problem to another one with only 1 degree of freedom, and 0 constraint:

$$2\pi r(r+h) = A_0 \Rightarrow h = \frac{A_0}{2\pi r} - r$$

$$\text{Volume} = \pi r^2 h = \pi r^2\left(\frac{A_0}{2\pi r} - r\right) = \frac{A_0}{2} r - \pi r^3$$

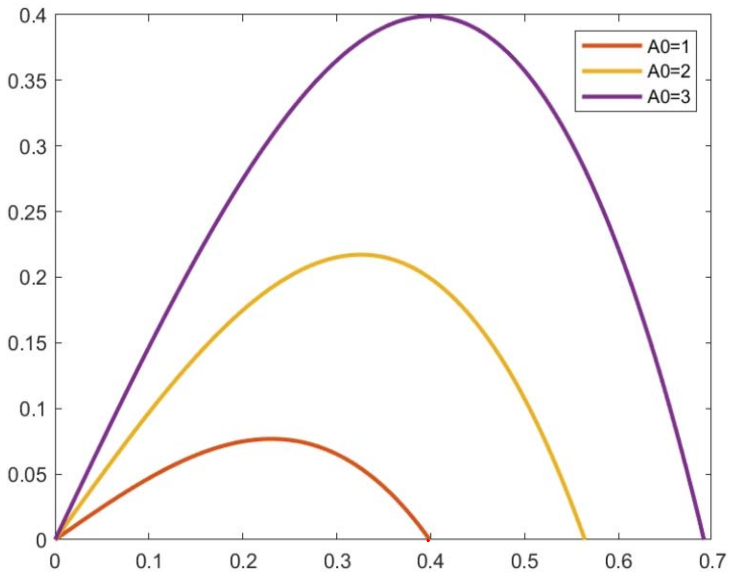
The problem becomes: $\boxed{\max_{r} \frac{A_0}{2} r - \pi r^3}$

A fine detail: Now that the variable 'h' is gone, and the 'r' variable is essentially "free" (or "unconstrained"), we may still want to add that r should be constrained to be positive, i.e. $r \geq 0$. Moreover, since $h = \frac{A_0}{2\pi r} - r \geq 0$, this means

$$\frac{A_0}{2\pi r} \geq r, \text{ or } r^2 \leq \frac{A_0}{2\pi}, \text{ or } r \leq \sqrt{\frac{A_0}{2\pi}}.$$

So, maybe it is better to cast the problem as

$$\max_{0 \le r \le \sqrt{\frac{A_0}{2\pi}}} \quad \frac{A_0}{2} r - \pi r^3$$



Does it matter to be careful about the bounds of the variables? Usually, it does. But not in this case.

It happens that the objective function is negative when $r < 0$ or $r > \sqrt{A_0/2\pi}$, so the bounds have no effects on where the maximizer is.

To find the maximizer, we just have to locate the unique stationary point of the objective function : i.e. to solve

$$\frac{d}{dr}\left(\frac{A_0}{2}r - \pi r^3\right) = 0 \quad \Leftrightarrow \quad \boxed{r = \sqrt{\frac{A_0}{6\pi}}} \leftarrow \text{maximizer, call it } r^*$$

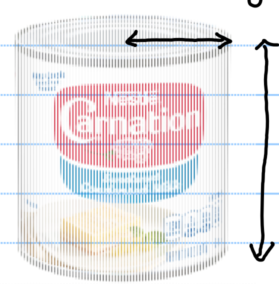
$$\frac{A_0}{2} - 3\pi r^2$$

The corresponding height of the cylinder is :

$$h^* = \frac{A_0}{2\pi r^*} - r^* = \frac{A_0}{2\pi\sqrt{\frac{A_0}{6\pi}}} - \sqrt{\frac{A_0}{6\pi}} = \cdots = 2\sqrt{\frac{A_0}{6\pi}} = 2r^*$$

We have the elegant conclusion that a volume maximizing cylinder satisfies :

$$\boxed{\frac{h^*}{r^*} = 2}$$



Example 1 is relatively easy because it is essentially a single-variable problem. So all we need is univariate calculus.

The second example is (literally!) infinitely harder, as it involves an infinite number of degrees of freedom. I want you to see this problem, understand what the problem is about and appreciates its difficulty. But it is out of the scope of this course to solve this problem fully.

Example 2: Among all the planar shapes with a fixed perimeter, find the one with the largest area.

Notice that the set of all "planar shapes with a fixed perimeter" is extremely vast. In the HW, we will explore a very special case of this problem in which we restrict ourselves only to the set of "all triangles with a fixed parameter". As you will see, a natural way to formulate the triangular case gives an optimization problem with 2 degress of freedom and 3 (linear) constraints.

Now, you only need to imagine what happens if the problem allows for all possible shapes, this will in particular include all triangles (3-gons), quadrilaterials (4-gons), pentagons (5-gons), and in fact all n-gons, for any n=3,4,5,6,.....

And this hasn't accounted for shapes with curved boundaries, like circles and ellipses. It is quite clear that, in some appropriate sense, the problem has an infinite number of degrees of freedom.

This kind of optimization problems has a different name: "Calculus of Variations".

Example 3: The Professor's Dairy Business Problem (Section 1.1 of textbook)

Each week, a professor makes butter and icecream from milk, and sell them at a local store.

Milk constraints:
His cows produce 22 gallons of milk each week.
Each kilogram of butter requires 2 gallons of milk to produce.
Each gallon if ice-cream requires 3 gallons of milk to produce.

Freezer constraint:
His freezer can hold (practically) unlimited amounts of butter, but at most 6 gallons of icecream.

Time constraint:
Professor has only 6 hours per week.
One hour of work needed to produce either 4 gallons of icre-cream or one kilogram of butter.

He can always sell everything he produces, at a price that ensures a profit of $5 per gallon of icecream and $4 per kilogram of butter.

Problem: How much ice-cream and butter should be produce to maximize his profit?

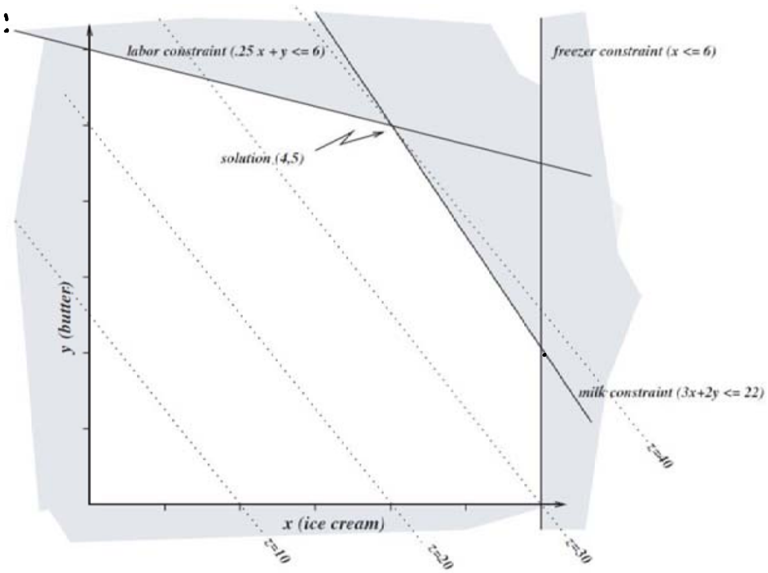
x = # of gallons of ice-cream produced
y = # of kilograms of butter produced

assumed to be any non-negative real values

so, the problem becomes: $\max_{x,y} \underbrace{5x+4y}$ s.t. $\begin{cases} 3x+2y \le 22 \\ x \le 6 \\ \frac{1}{4}x + y \le 6 \\ x, y \ge 0 \end{cases}$

Call this objective function z (a linear function of two variables)

Graphical solution:



The maximizer

$\boxed{(x,y) = (4,5)}$

Example 3 has three important properties:

- It variables are <u>continuous</u> variables.
- All constraints and bounds involve <u>linear</u> functions of variables.
- The objective function is also a <u>linear</u> function of the variables.

Such problems are called linear programming problems, or just linear programs (LP).

In general, a LP <u>in the standard form</u> an optimization problem of the following form:

$$\underbrace{\min_{x_1,\dots,x_n}}_{\substack{\text{variables}\\ \text{or}\\ \text{degrees of freedom}}} \quad Z = \underbrace{p_1 x_1 + \dots + p_n x_n}_{\text{objective function}} \quad \text{s.t.} \quad \left.\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &\geq b_1 \\ &\vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &\geq b_m \\ x_1, \dots, x_n &\geq 0 \end{aligned}\right\} \text{constraints}$$

n variables / d.o.f.

$m+n$ constraints

In matrix notation, a LP in standard form can be compactly restated as :

$$\min_{x \in \mathbb{R}^n} p^T x \quad \text{s.t. } Ax \geq b, \; x \geq 0$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad p = \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}, \quad A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

The dairy business problem can be expressed in the standard form, with

$$A = -\begin{bmatrix} 1 & 0 \\ 1/4 & 1 \\ 3 & 2 \end{bmatrix}, \quad b = -\begin{bmatrix} 6 \\ 6 \\ 22 \end{bmatrix}, \quad p = -\begin{bmatrix} 5 \\ 4 \end{bmatrix}.$$

A linear program (LP) in the <u>canonical form</u> is an optimization problem of the form:

$$\min \quad z = p^T x \quad \text{s.t. } Ax = b, \; x \geq 0$$

We can always convert a LP in the standard form into one in the canonical form by introducing <u>slack variables</u>.

Ex: the dairy problem can be reformulated as

$$\begin{array}{lrcrcrcrcrcl}
\min\limits_{x \in \mathbb{R}^5} & z = -5x_1 - 4x_2 & & & & & & & & & & \\
\text{s.t.} & x_1 & & & + & x_3 & & & & & = & 6 \\
& \frac{1}{4}x_1 & + & x_2 & & & + & x_4 & & & = & 6 \\
& 3x_1 & + & 2x_2 & & & & & + & x_5 & = & 22 \\
& & & & & & & & & x_1, x_2, x_3, x_4, x_5 & \geq & 0
\end{array}$$

This is a canonical form LP with

$$p = \begin{bmatrix} -5 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1/4 & 1 & 0 & 1 & 0 \\ 3 & 2 & 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 6 \\ 6 \\ 22 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}.$$

There are many optimization problems that are NOT linear programs. We shall see that LP have such special structures that, when combined with techiques in linear algebra, make them relatively easy to solve.

We shall devolop the celebrated simplex method for solving LP. But before we get into it, we shall explore a few interesting applications in which LP arises.

Applications

① The diet problem

n possible foods, indexed by $j=1,2,\cdots,n$
m nutritional categories, indexed by $i=1,\cdots,m$

x_j = amount of food j to be included in the diet, measured in the number of servings.
p_j = cost of one serving of food j
b_i = minimum daily requirement of nutrient i
A_{ij} = amount of nutrient i contained in one serving of food j

If one seeks the diet with the lowest cost that achieves all the nutritional requirements, then she is faced with the following LP:

$$\min_{x_1,\cdots,x_n} p_1x_1+\cdots+p_nx_n \quad \text{s.t.} \quad \begin{aligned} A_{11}x_1+\cdots+A_{1n}x_n &\geq b_1 \\ &\vdots \\ A_{m1}x_1+\cdots+A_{mn}x_n &\geq b_m \\ x_1,\cdots,x_n &\geq 0. \end{aligned}$$

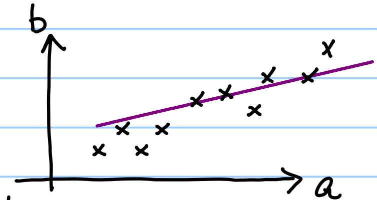
It is exactly a LP in the standard form.

Note: m, n can be quite big, hence the challenge!
(The graphical approach only works for $n=2$.)

(2) Linear regression (This is a special case of the problem considered in Sec 1.3.2. The math is hardly different.)

Suppose we have a set of observations

$$\underset{\in \mathbb{R}}{a_i}, \underset{\in \mathbb{R}}{b_i} \qquad i=1,\cdots,m$$



We <u>postulate</u> that, in general, 'b' is approximately a linear function of 'a', i.e. $\exists\, x, \gamma$ s.t.

$$b_i \approx x a_i + \gamma, \quad i=1,\cdots,m$$

Note: If '$\approx$' is replaced by '$=$', then

$$\underbrace{\begin{bmatrix} a_1 & 1 \\ a_2 & 1 \\ \vdots & \\ a_m & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ \gamma \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{b}$$

This is a linear system of m (typically $\gg 2$) equations and 2 unknowns. $\longrightarrow$ mostly likely no solutions

Sometimes, there is truely an underlying linear relation, but that the measurement b_i is corrupted by 'noise', i.e.

$$b_i = \overset{?}{x} a_i + \overset{?}{\gamma} + \overset{?}{\text{"measurement error"}}$$

known

We would like to "learn" this linear model (determined by x and γ) from the noisy data a_i, b_i, $i=1,\dots,m$.

There is a very standard, and widely used, method for this problem, called the "least-square method". It is based on solving:

$$\boxed{\text{Find } x, \gamma \text{ such that } \sum_{i=1}^{m} (x a_i + \gamma - b_i)^2 \text{ is minimized.}}$$

This least square problem has a very elegant solution, note that:

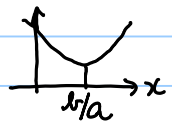
$Q(x)$

$$= \frac{1}{2}\sum_{i=1}^{m}(x a_i + \gamma - b_i)^2 = \frac{1}{2}\|Ax - b\|_2^2 = \frac{1}{2}(x^T A^T - b^T)(Ax - b)$$

a multivariate quadratic polynomial

$$= \frac{1}{2}x^T A^T A x - x^T A^T b + \frac{1}{2} b^T b$$

Just like $q(x) = \frac{1}{2}\overset{\overset{0}{\wedge}}{a}x^2 - bx + c$ has $x = b/a$ as its minimizer.



$Q(x)$ has $x = \underbrace{(A^TA)^{-1}(A^Tb)}_{\text{called the least square solution}}$ as its minimizer. (Proof omitted, but it's not difficult.)

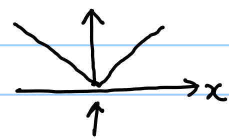
Note: $\nabla Q(x) = A^TAx - A^Tb$

A drawback of the least-square approach is that if the measurement errors in some of the observations b_i are large (these are called "outliers"), then 'square' would further amplifies these errors, causing the learning / fitting to perform poorly.

Remedy: use $|\cdot|$ instead of $(\ )^2$. Solve:

$$\boxed{\text{Find } x, \gamma \text{ such that } \sum_{i=1}^{m}|x a_i + \gamma - b_i| \text{ is minimized.}}$$

This least L^1 error optimization problem is harder to solve than the least L^2 problem, part of the problem is that the $|x|$ function is not differentiable at $x=0$.



Interestingly, the least L^1 error problem can be recast as a LP!

$$\min_{x,\gamma} \underbrace{\sum_{i=1}^{m} |xa_i + \gamma - b_i|}_{\text{not a linear function in } x,\gamma}$$, as it stands, is not a LP.

But we can rewrite it as follows: (Tricky – do you see why?)

free $\rightarrow x, \gamma,$

$\geq 0 \rightarrow y_1, \dots, y_m$

$$\boxed{\min_{\substack{x,\gamma,\\ y_1,\dots,y_m}} \; y_1 + \dots + y_m \quad \text{s.t.} \quad \begin{array}{c} -y_i \leq xa_i + \gamma - b_i \leq y_i \\ y_1, \dots, y_m \geq 0 \end{array}} \quad \star$$

This is a LP, but in neither the standard form nor the canonical form. Below I show how to further rewrite it into the standard form.

Let's rewrite the linear inequalities to comply with the standard form:

$$-y_i \le x a_i + \gamma - b_i \quad \text{and} \quad x a_i + \gamma - b_i \le y_i$$

$$\Updownarrow \qquad\qquad\qquad\qquad \Updownarrow$$

$$a_i x + \gamma + y_i \ge b_i \qquad\qquad a_i x + \gamma - y_i \le b_i$$

$$\Leftrightarrow -a_i x - \gamma + y_i \ge -b_i$$

In the standard form, all the variables are constrained to be non-negetive.
This is the case for $y_1, \dots, y_m$, but <u>not</u> x, γ.

Trick: $x = \underset{\overset{\vee\!\!/}{0}}{x^+} - \underset{\overset{\vee\!\!/}{0}}{x^-}$, $\gamma = \underset{\overset{\vee\!\!/}{0}}{\gamma^+} - \underset{\overset{\vee\!\!/}{0}}{\gamma^-}$ (Every real value can be written as the difference of 2 non-negative values. E.g. $5 = 5 - 0$, $-5 = 0 - 5$)

And ✡ is rewritten as:

$$\boxed{\begin{aligned} \min \; y_1 + \dots + y_m \;\; \text{s.t.} \;\; & \left.\begin{aligned} a_i x^+ - a_i x^- + \gamma^+ - \gamma^- + y_i &\ge b_i, \\ -a_i x^+ + a_i x^- - \gamma^+ + \gamma^- + y_i &\ge -b_i, \quad i = 1, \dots, m \end{aligned}\right\} 2m \\ & \left. y_1, \dots, y_m, x^+, x^-, \gamma^+, \gamma^- \ge 0 \quad \right\} m+4 \end{aligned}}$$

$m+4$ variables

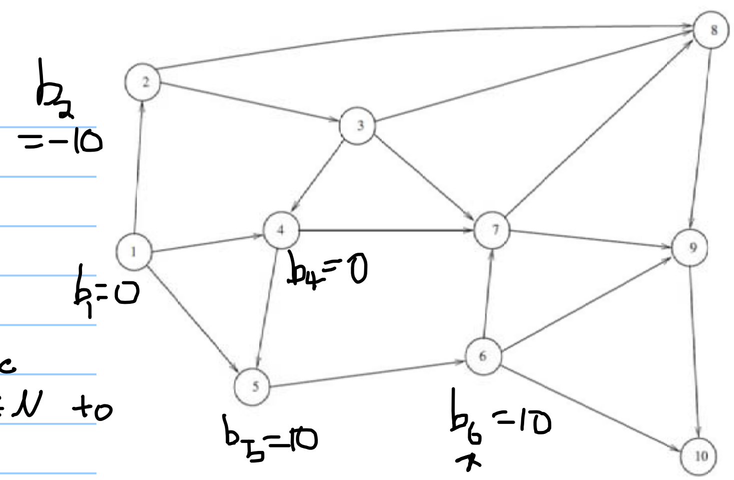
$2m + (m+4)$ constraints

In matrix notation: $\min_{x} p^T x$ s.t $Ax \geq b$, $x \geq 0$ where

$$x = \begin{bmatrix} x^+ \\ x^- \\ \gamma^+ \\ \gamma^- \\ y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad p = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad A = \left[\begin{array}{cccc|cccc} a_1 & -a_1 & 1 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \ddots & & \\ a_m & -a_m & 1 & -1 & 0 & & & 1 \\ \hline -a_1 & a_1 & -1 & 1 & 1 & & 0 & \\ \vdots & & \vdots & \vdots & & \ddots & & \\ -a_m & a_m & -1 & 1 & 0 & & & 1 \end{array}\right], \quad b = \left[\begin{array}{c} b_1 \\ \vdots \\ b_m \\ \hline -b_1 \\ \vdots \\ -b_m \end{array}\right]$$

$(m+4) \times 1$ $\qquad$ $(m+4) \times 1$ $\qquad$ $(2m) \times (m+4)$ $\qquad$ $2m \times 1$

③ minimum cost network flow



$\mathcal{N}$ = set of all nodes

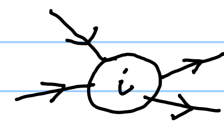
$\mathcal{A}$ = set of all arcs

($(i,j) \in \mathcal{A}$ means there is an arc connecting an origin node $i \in \mathcal{N}$ to a destination node $j \in \mathcal{N}$.)

b_i = amount of product produced or consumed at node i

$b_i > 0$ means node i supplies b_i units of product

$b_i < 0$ means node i demands $-b_i$ units of product



c_{ij} = cost of moving 1 unit of product

The min-cost network flow problem is

$$\min_x \; z = \sum_{(i,j)\in\mathcal{A}} c_{ij} x_{ij} \quad \text{s.t.} \quad \sum_{j:(i,j)\in\mathcal{A}} x_{ij} - \sum_{j:(j,i)\in\mathcal{A}} x_{ji} = b_i \quad \forall i \in \mathcal{N}$$

$$l_{ij} \le x_{ij} \le u_{ij} \quad \forall (i,j) \in \mathcal{A}$$

Note : $\sum_{i \in N} b_i = 0$, otherwise the problem is infeasible.

Typically there is a source node $s \in N$, and a sink node $t \in N$

$$b_s > 0 \ , \ b_t = -b_s < 0 \ , \text{ and } b_i = 0, \ \forall i \neq s, t.$$