Lecture 1

Optimization for F (SF1851)

Amol Sasane sasane@kth.se Room 3711

Exercise classes: Anders Möller
Tore Odland

Course will be given partly in Swedish and partly in English.

Lectures and lecture notes: in English.

Exercise classes: in Swedish.

Exams and take-home assignments: both in English and in Swedish.

Your answers: either in English or in Swedish; you can choose the language you prefer.

Course homepage:

www.math.kth.se/optsyst/grundutbildning/kurser/SF1851

Literature: Complete set of lecture notes

"Optimization for E by Amol Sasane and Krister Svanberg".

Buy from the student office in the Maths Dept.

Contains all the material taught in this course.

Schedule of lectures.

Exercise classes (Exercises are taken from the lecture notes)

Solutions to all the exercises.

My handwritten lecture notes. (outlines of things explained during the lectures)

Some old exams and their solutions.

Assessment:

Written exam : Oct. 21 50 points

+

2 Take-home assignments (*)

If you pass (*), then you can skip Question 1 in the final exam, worth 9 points.

Detailed information about this is on the course homepage.

The first take-home assignment has already been posted on the web.

Deadline for this is: Sept. 17.

Formula sheet for the final exam is also there on the website.

Course is on Optimization.

<u>What is optimization?</u>

<u>Basic object of study</u>: a real-valued function on a set.

$$\underbrace{f}_{\text{objective function}} : \underbrace{\mathcal{F}}_{\text{feasible set}} \longrightarrow \mathbb{R}$$

<u>Optimization problem</u>: Mimimize f.

Find $\hat{x} \in \mathcal{F}$ s.t. for all other $x \in \mathcal{F}$, $f(\hat{x}) \leq f(x)$

(Maximization problems can be converted into equivalent minimization problems by considering $-f$ instead of f.)

Depending on the nature of $\mathcal{F}$, and there are various subdisciplines of optimization; for example:

combinatorial optimization: $\mathcal{F}$ discrete

calculus of variations: $\mathcal{F}$ is a set of functions $\mathcal{F} = C[a,b]$

...

In this course: $\boxed{\mathcal{X} \subset \mathbb{R}^n}$

$$
\begin{array}{ccccc}
f : & \mathcal{X} & \longrightarrow & \mathbb{R} \\
& \cup\!\!\!\!| & & \\
& \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} & \longmapsto & f\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) \\
& \parallel & & \\
& x & \longmapsto & f(x)
\end{array}
$$

So the central problem in the course is:

Given $f : \mathcal{X} \longrightarrow \mathbb{R}$,

$$
(P) : \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathcal{X}. \end{cases}
$$

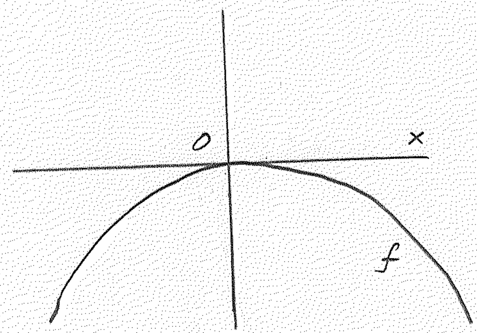
But a solution may not exist:

Examples:

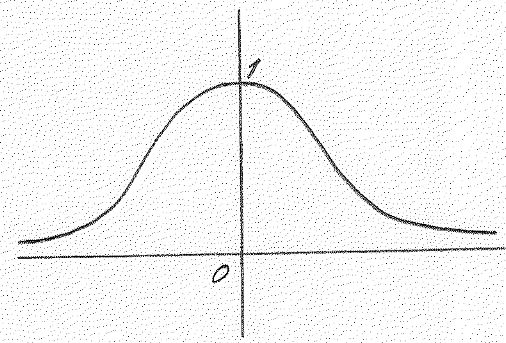
(1) $\begin{cases} \text{minimize} & x \\ \text{s.t.} & 1+x^2 = 0 \end{cases}$

$\mathcal{F} = \emptyset$!

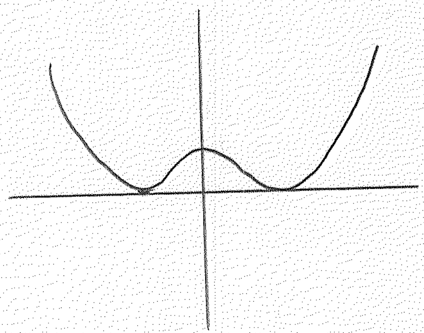
(2) $\begin{cases} \text{minimize} & -x^2 \\ \text{s.t.} & x \in \mathbb{R} \end{cases}$



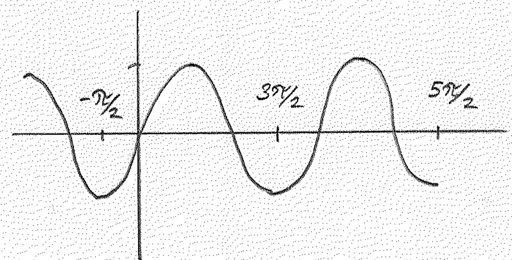
(3) $\begin{cases} \text{minimize} & \frac{1}{1+x^2} \\ \text{s.t.} & x \in \mathbb{R} \end{cases}$

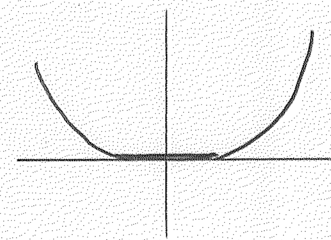


More than one solution may exist:



$\begin{cases} \text{minimize} & \sin x \\ \text{s.t.} & x \in \mathbb{R} \end{cases}$.





If a solution exists, how do we find it?

Depending on $\mathcal{F}$ and f, this course is divided into 3 parts:

I. Linear programming

Objective function f: linear

$\mathcal{F}$: linear ineq. constraints

$$\begin{cases} \text{minimize} & f(x) := c^T x \\ \text{s.t.} & Ax \geqslant b \end{cases}$$

$A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$

$$c^T x = [c_1 \dots c_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = c_1 x_1 + \dots + c_n x_n$$

$$\mathcal{F} = \{ x : \underbrace{Ax \geqslant b}_{\equiv} \}$$

$$\begin{cases} a_{11} x_1 + \dots + a_{1n} x_n \geqslant b_1 \\ \vdots \\ a_{m1} x_1 + \dots + a_{mn} x_n \geqslant b_m \end{cases}$$

II. Quadratic optimization

f: quadratic

$\mathcal{F}$: linear ineq. constr.

$$\begin{cases} \text{minimize} & \frac{1}{2} x^T H x + c^T x + c_0 \\ \text{s.t.} & Ax \geqslant b \end{cases}$$

$H \in \mathbb{R}^{n\times n}$

$c \in \mathbb{R}^n$

$c_0 \in \mathbb{R}$

III. Nonlinear optimization

f: general nonlinear

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & g_1(x) \leq 0 \\ & \vdots \\ & g_m(x) \leq 0 \end{cases}$$

$f, g_1, \dots, g_m$

(differentiable)

Difficulty level increases. $\longrightarrow$

Part I : Linear programming.

$$\begin{cases} \text{minimize} & f(x) \\ \text{s.t.} & x \in \mathcal{F} \end{cases}$$

Objective function f is linear : $f(x) = c_1 x_1 + \ldots + c_n x_n$

$$= \begin{bmatrix} c_1 & \ldots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= c^T x \quad \text{where} \quad c = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}, \; x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Constraints are a bunch of linear inequalities:

$$\mathcal{F} = \{ x \in \mathbb{R}^n : Ax \geq b \}$$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m$$

$$\begin{cases} a_{11} x_1 + \ldots + a_{1n} x_n \geq b_1 \\ \vdots \\ a_{m1} x_1 + \ldots + a_{mn} x_n \geq b_m. \end{cases}$$

<u>Why the name?</u> Linear: obvious

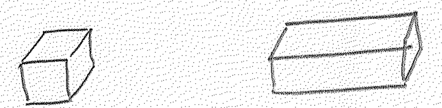
Programming: Historical reason (military program ≡ x)

Algorithm for solving this: <u>G. Dantzig; 1947; Simplex method.</u>

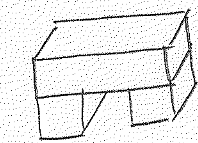
<u>Why study this?</u> Need arises in applications.

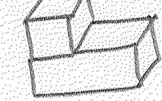
Although linear functions are simple, linear programming problems arise frequently in engineering and economics. We will start with a very simple <u>toy</u> example from production planning.

<u>Problem.</u> We own a furniture company that manufactures two types of furniture: tables and chairs.

These are made from two types of parts: 

small parts big parts.

To make a table, we need 1 big and 2 small parts: 

To make a chair, we need 1 big and 1 small part: 

A table sells for SEK 400, and a chair for SEK 300.

Suppose we have 200 big parts and 300 small parts.

Question. How many tables and chairs should we make so as to maximize the profit?

Suppose $T =$ no. of tables

$C =$ no. of chairs.

Then profit $= 400 \cdot T + 300 \cdot C$

We want to maximize this.

But there are constraints!

For making T tables, we need T big and $2T$ small parts.

For making C chairs, we need C big and $2C$ small parts.

Totally we need $T+C$ big parts, and this must be ≤ 200.

Totally we need $2T+C$ small parts, and this must be ≤ 300.

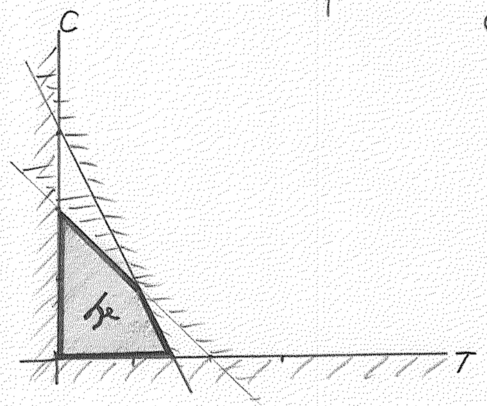
Also $T \geq 0$, $C \geq 0$.

So we arrive at:
$$\left\{\begin{array}{ll} \text{maximize} & 400T + 300C \\ \text{s.t.} & T + C \leq 200 \\ & 2T + C \leq 300 \\ & T \geq 0 \\ & C \geq 0 \end{array}\right\}$$
, a linear programming problem.

What does the feasible set look like?

$$\mathcal{F} = \left\{ (T, C) \;\middle|\; \begin{array}{rl} T + C \leq 200 & \text{half plane} \\ 2T + C \leq 300 & \text{half plane} \\ T \geq 0 & \text{half plane} \\ C \geq 0 & \text{half plane} \end{array} \right\}$$

intersection of half planes.

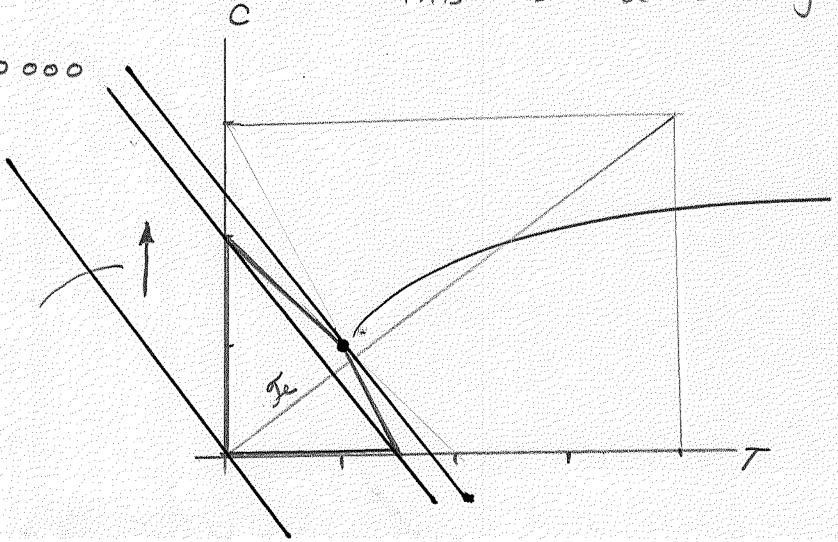


Graphical solution: Fix a profit (P, say) and look at all (T, C) giving this profit, i.e., (T, C) satisfying $400T + 300C = P$.

This is a straight line, perpindular to the line joining $(0,0)$ and $(400,300)$:

P = 60 000

P = 0



As P ↑, line moves upwards.

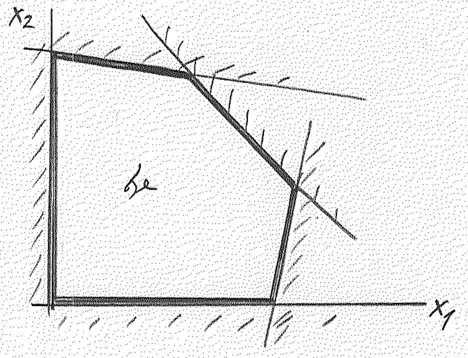
$$\left\{ \begin{array}{l} T + C = 200 \\ 2T + C = 300 \end{array} \right\} \Rightarrow T = C = 100$$

$$f_{max} = 400 \cdot 100 + 300 \cdot 100 = 70\,000 \text{ SEK}$$

General case in $\mathbb{R}^2$: similar as above.

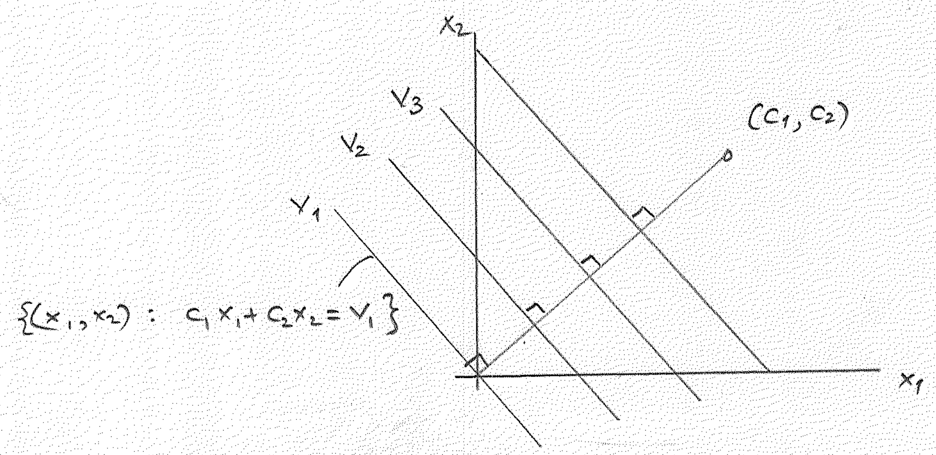
Constraints $\quad a_{i1}x_1 + a_{i2}x_2 \geqslant b_i \quad (i \in I)$ determines a half plane.

So $\mathcal{F}$ = intersection of half planes = convex polygonal region.

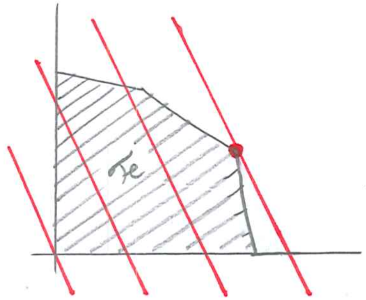


Level sets of the function to be optimized are straight lines perpendicular to c.

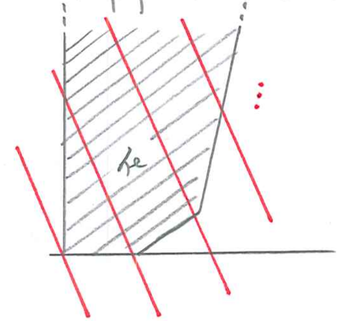
$$f(x_1, x_2) = c^T x = c_1 x_1 + c_2 x_2 = \underset{\text{fixed.}}{v}$$



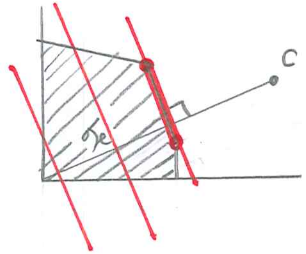
So by a reasoning similar to the above, we see that the function is maximized or minimized at a <u>corner point</u> of the convex polygonal region.



Of course it may happen that there is no extremizer at all:



It may also happen that there are infinitely many extremizers (when c is $\perp$ to one of the sides of $\mathscr{R}$):



<u>Observation:</u> If there exists an extremizer, then there exists an extremizer which is a <u>corner point of $\mathscr{R}$</u>.

<u>Conclusion:</u> It suffices to search the corner points of $\mathscr{R}$.

But how do we work in $\mathbb{R}^n$ $(n > 3)$?

Example:

$$\begin{cases} \text{minimize} & x_1 + x_2 - 2x_3 + 7x_4 \\ \text{s.t.} & x_1 + 2x_2 - 8x_3 \leq 1 \\ & 2x_1 - x_2 + 3x_4 \leq 3 \\ & x_1 \geq 0 \\ & x_3 \geq 0 \end{cases}$$

We cannot draw pictures now.

It turns out again that one can show that:

If there is an extremizer of the linear programming problem,

then there is an extremizer which is a corner point of $\mathcal{S}e$.

So it suffices to search among the corner points of $\mathcal{S}e$

But what exactly do we mean by a corner point of $\mathcal{S}e$?

And how do we determine them?

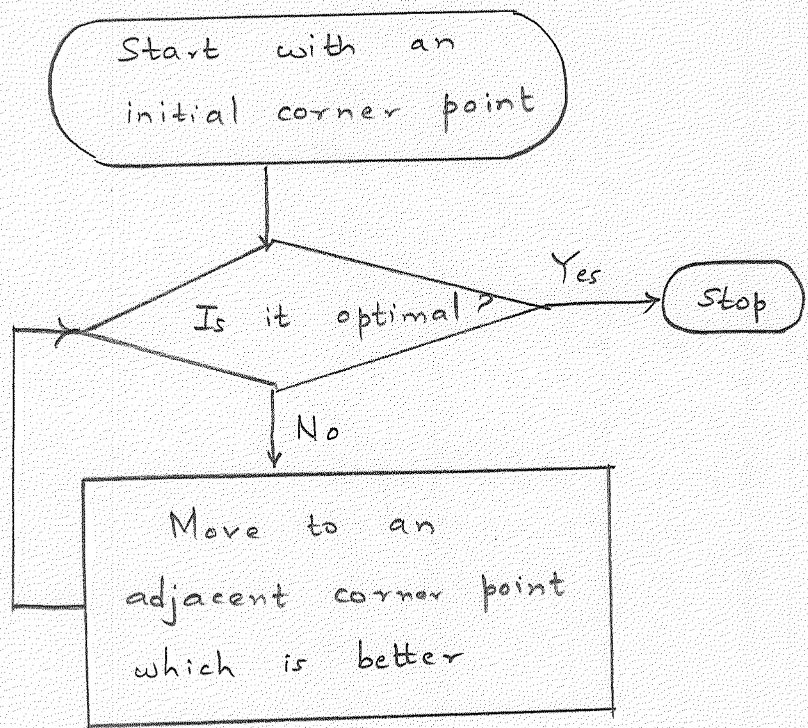
Answer: Linear algebra.

An important issue is that even with a small number of inputs (m, n small), number of corner points can be very large.

So calculating <u>all</u> corner points is not an option.

Instead, there is an efficient way of searching among corner points:

this is the <u>simplex method</u>.



But first we will learn about the <u>Standard Form</u> of linear programming problems.

<u>General linear programming problem</u>

$$\text{minimize} \quad c^T x$$
$$\text{s.t.} \quad Ax \le b$$

e.g. can't subtract two inequalities

$$\Rightarrow$$

<u>Linear programming problem in Standard Form</u>

$$\left\{ \begin{array}{ll} \text{minimize} & \tilde{c}^T \tilde{x} \\ \text{s.t.} & \tilde{A} \tilde{x} = \tilde{b} \\ & \tilde{x} \geq 0 \end{array} \right.$$

linear algebraic manipulations possible

simple

Standard Form : – special way of writing problems;

– every linear programming problem can be written as an equivalent problem in standard form (we will see this soon);

– then we will learn the simplex method for solving <u>linear programming problems in the standard form</u>

Standard form of linear programming problems

$$\begin{cases} \text{minimize} & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{cases}$$

$f(x) = c^T x$

$\mathcal{F}_e = \{ x \in \mathbb{R}^n \mid Ax = b, \ x \geq 0 \}$

$c \in \mathbb{R}^n, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m \qquad x \text{ (variable)} \in \mathbb{R}^n.$

Written out:

$$\begin{cases} \text{minimize} & c_1 x_1 + \dots + c_n x_n \\ \text{s.t.} & a_{11} x_1 + \dots + a_{1n} x_n = b_1 \\ & \vdots \\ & a_{mn} x_1 + \dots + a_{mn} x_n = b_m \\ & x_1 \geq 0 \\ & \vdots \\ & x_n \geq 0 \end{cases}$$

$\tilde{A}\tilde{x} \geq \tilde{b} \quad \Rightarrow \quad \begin{matrix} Ax = b \\ x \geq 0 \end{matrix}$

hard to manipulate/solve

$Ax = b$ — easy to manipulate/solve

$x \geq 0$ — simple inequality constraints

equivalent: One problem has a solution iff the other has a solution, and $\tilde{x} \leftrightarrow \hat{x}$.

How does one convert a general LP problem to one in standard form?

(1) Convert inequalities into equalities with slack variables.

$$a_{i1}x_1 + \dots + a_{in}x_n \leq b_i \iff \begin{cases} a_{i1}x_1 + \dots + a_{in}x_n + y_i = b_i \\ \text{and} \quad \underbrace{y_i}_{\text{'slack' variable}} \geq 0 \end{cases}$$

(2) Write free variables as a difference of two nonnegative ones.

$$x_i \text{ free} \longleftrightarrow \begin{cases} x_i = u_i - v_i, \\ u_i \geq 0 \\ v_i \geq 0 \end{cases}$$

Slack variables.

$$(P): \quad \begin{cases} \text{minimize} & c_1 x_1 + \dots + c_n x_n \\ \text{s.t.} & a_{11} x_1 + \dots + a_{1n} x_n \leq b_1 \\ & \vdots \\ & a_{m1} x_1 + \dots + a_{mn} x_n \leq b_m \\ & x_1 \geq 0, \dots, x_n \geq 0 \end{cases}$$

(P) is equivalent to:

$$(P'): \quad \begin{cases} \text{minimize} & c_1 x_1 + \dots + c_n x_n + 0 y_1 + \dots + 0 y_m \\ \text{s.t.} & a_{11} x_1 + \dots + a_{1n} x_n + y_1 = b_1 \\ & \vdots \\ & a_{m1} x_1 + \dots + a_{mn} x_n + y_m = b_m \\ & x_1 \geq 0, \dots, x_n \geq 0 \\ & y_1 \geq 0, \dots, y_m \geq 0 \end{cases}$$

Old problem:

$$\begin{cases} \text{minimize} & c^T x \\ \text{s.t.} & A x \leq b \\ & x \geq 0 \end{cases}$$

New problem:

$$\begin{cases} \text{minimize} & \tilde{c}^T \tilde{x} \\ \text{s.t.} & \tilde{A} \tilde{x} = \tilde{b} \\ & \tilde{x} \geq 0. \end{cases} \qquad \tilde{c} = \begin{bmatrix} c \\ 0 \end{bmatrix} \quad \tilde{A} = [A \;\vdots\; I_m] \quad \tilde{b} = b. \qquad \tilde{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

$$x \in \mathcal{F} \iff \tilde{x} := \begin{bmatrix} x \\ y := b - Ax \end{bmatrix} \in \tilde{\mathcal{F}}.$$

<u>Example</u>.

$$\left\{\begin{array}{ll} \text{maximize} & 400\,T + 300\,C \\ \text{s.t.} & T + C \leq 200 \\ & 2T + C \leq 300 \\ & T \geq 0 \\ & C \geq 0 \end{array}\right.$$

$$T = x_1 \qquad C = x_2$$

$$\left\{\begin{array}{ll} \text{minimize} & -400\,x_1 + 300\,x_2 \\ \text{s.t.} & x_1 + x_2 + y_1 = 200 \\ & 2x_1 + x_2 + y_2 = 300 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \\ & y_1 \geq 0 \\ & y_2 \geq 0 \end{array}\right\} \quad \text{is of the form} \quad \left\{\begin{array}{ll} \text{minimize} & c^T x \\ \text{s.t} & Ax = b \\ & x \geq 0 \end{array}\right\},$$

where

$$c = \begin{bmatrix} -400 \\ -300 \\ 0 \\ 0 \end{bmatrix}, \quad A = \left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{array}\right], \quad b = \begin{bmatrix} 200 \\ 300 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix}.$$

<u>Free variables.</u> A problem like $(P): \begin{cases} \text{minimize} & x_1 + 3x_2 \\ \text{s.t} & x_1 + 2x_2 = 4 \\ & x_2 \geq 0 \end{cases}$

is not in standard form, since x_1 is not required to be ≥ 0.

x_1 can take any real value.

Every real number can be written as a difference of nonnegative numbers.

$(3 = 3 - 0 \quad -3 = 0 - 3 = 1 - 4 = \dots)$

We write $x_1 = u_1 - v_1$ with $u_1 \geq 0$, $v_1 \geq 0$ being new variables

Substitute $u_1 - v_1$ instead of x_1 everywhere.

Then. $(P'): \begin{cases} \text{minimize} & u_1 - v_1 + 3x_2 \\ \text{s.t} & u_1 - v_1 + 2x_2 = 4 \\ & u_1 \geq 0 \\ & v_1 \geq 0 \\ & x_2 \geq 0 \end{cases}$

linearity of objective function is preserved

linearity in the equality constraints is preserved

<u>all</u> variables are nonnegative.

Problem (P') is in standard form.

Moreover, (P), (P′) are equivalent:

(i) If $(\hat{x}_1, \hat{x}_2)$ is optimal for (P),

then $(\hat{u}_1, \hat{v}_1, \hat{x}_2)$ is optimal for (P′),

where $\hat{u}_1, \hat{v}_1$ are any real numbers s.t. $\hat{u}_1 \geq 0$, $\hat{v}_1 \geq 0$ and $\hat{x}_1 = \hat{u}_1 - \hat{v}_1$.

(For e.g., we may take $\hat{u}_1 = \hat{x}_1$, $\hat{v}_1 = 0$ if $\hat{x}_1 \geq 0$ and
$\hat{u}_1 = 0$, $\hat{v}_1 = -\hat{x}_1$ if $x_1 < 0$.)

(ii) If $(\hat{u}_1, \hat{v}_1, \hat{x}_2)$ is optimal for (P′),

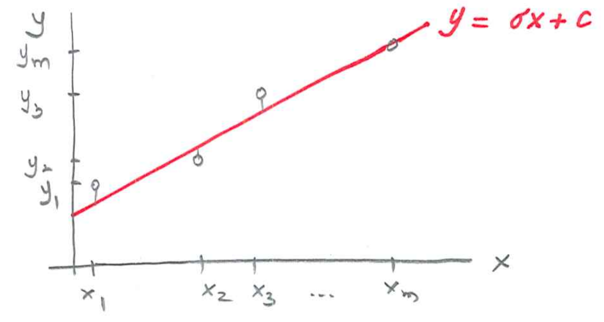
then $(\hat{u}_1 - \hat{v}_1, \hat{x}_2)$ is optimal for (P).

So if (P′) has a solution, we can get a solution for (P)

and if (P′) doesn't have a solution, neither does (P).

Modelling : line fitting as an LP problem.

$(x_1, y_1), \ldots, (x_m, y_m)$: observation points.



E.g. x_i's : blood pressures of patients.

y_i's : corresponding drug dosages.

Problem : Fit a line,

i.e., find σ, c so that the line $y = \sigma x + c$ "best" fits the data

"error" is minimized

Measure of error? **E.g** $|\sigma x_1 + c - y_1| + \ldots + |\sigma x_m + c - y_m|$.

So we arrive at:

$$(P):\begin{cases}\text{minimize} & |\sigma x_1 + c - y_1| + \dots + |\sigma x_m + c - y_m| \\ \text{subject to} & \sigma, c \in \mathbb{R}\end{cases}$$

(P) is <u>not</u> a linear programming problem, since the objective function is not linear.

But we can convert it into an equivalent LP problem:

$$(P'):\begin{cases}\text{minimize} & v_1 + \dots + v_m \\ \text{subject to} & \left.\begin{array}{l} v_i \geq \sigma x_i + c - y_i \\ v_i \geq -(\sigma x_i + c - y_i) \end{array}\right\} \text{ for } i = 1, \dots, m.\end{cases}$$

(P), (P′) are equivalent:

(i) If $(\hat{v}_1, \dots, \hat{v}_m, \hat{\sigma}, \hat{c})$ is optimal for (P′),
then $(\hat{\sigma}, \hat{c})$ is optimal for (P).

(ii) If $(\hat{\sigma}, \hat{c})$ is optimal for (P),
then $(\hat{v}_1, \dots, \hat{v}_m, \hat{\sigma}, \hat{c})$ is optimal for (P), where $\hat{v}_i := |\hat{\sigma} x_i + \hat{c} - y_i|$, $i = 1, \dots, m$.

We will see detailed justification of similar problems in the coming Exercise Class. Useful for Take-Home Assigment I.