Topic 3 Convex optimization problems

stured? 1. A general optimization problem:

$$\min f_0(x)$$
$$\text{s.t. } f_i(x) \le 0 \quad i=1,\dots,m$$
$$h_i(x) = 0 \quad i=1,\dots,p$$

Terminology:

- $x \in \mathbb{R}^n$: optimization variable
- $f_0 : \mathbb{R}^n \to \mathbb{R}$: objective function (or cost function)
- $f_i : \mathbb{R}^n \to \mathbb{R}$: inequality constraint functions
- $h_i : \mathbb{R}^n \to \mathbb{R}$: equality constraint functions

If there are no constraints, we say the problem is unconstrained.

- $D = \bigcap_{i=1}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$ is the domain of the optimization problem.

- $x \in D$ is a feasible point if it satisfies all the constraints
- If there exists at least one feasible point, the optimization problem is said to be feasible.

  Otherwise it is infeasible (no solution exists).
- Set of all feasible points is called the feasible set or the constraint set.

- Optimal value p^* of the problem is the optimal objective function:

$$p^* = \inf\{f_0(x) \mid f_i(x) \le 0,\ i=1\dots m,\ h_i(x)=0,\ i=1\dots m\}$$

$p^* = \infty$ if the problem is infeasible; if $p^* = -\infty$ then problem is unbounded

∘ <u>Optimal point</u> x^* is a feasible point such that

$$f_0(x^*) = p^*$$

The set of all optimal points is called the optimal set X_{opt}.

- If there exists at least one optimal point, the problem is solvable and the optimal value is attained or achieved.
- If X_{opt} is empty then the problem is not attained/or achieved (for example, when the problem is unbounded)

∘ x is locally optimal if it minimizes f_0 over nearby points in the feasible set.

That is, $\exists\, R > 0$ such that x is optimal for problem

$$\begin{aligned} \min \quad & f(z) \\ \text{s.t.} \quad & f_i(z) \le 0, \quad i = 1 \dots m \\ & h_i(z) = 0, \quad i = 1 \dots p \\ & \|z - x\|_2 \le R \end{aligned}$$

"Globally optimal" is when x solves the original optimization problem. In this course, we use optimal to mean globally optimal.

<u>Examples</u>: i) $f_0(x) = \frac{1}{x}$, $p^* = 0$ but x^* is not achieved.

ii) $f_0(x) = -\log x$, $p^* = -\infty$, problem unbounded below.

iii) $f_0(x) = x \log x$, $p^* = -\frac{1}{e}$, $x^* = \frac{1}{e}$

All above 3 examples have $\operatorname{dom} f_0 = R_{++}$. $(x > 0)$.

iv) $f_0(x) = x^3 - 3x$ has a local optimal point $x^* = 1$.

+) <u>Implicit constraints</u>: Note that even when an optimization problem is unconstrained, it may still has implicit constraints as a result of its domain.

Example: $\min f_0(x) = -\sum_{i=1}^{k} \log(-a_i^T x + b_i)$
is an unconstrained problem with implicit constraints as
$$a_i^T x - b_i \leq 0.$$

+) <u>Feasibility problem</u>: This kind of problems is to determine if a set of constraints is feasible (or consistent).

A feasibility problem can be written as

$$\begin{aligned} &\text{find } x \\ &\text{s.t. } f_i(x) \leq 0, \ i=1\ldots m \\ &\quad\ \ h_i(x) = 0, \ i=1\ldots p \end{aligned} \qquad \text{OR} \qquad \begin{aligned} &\min 0 \\ &\text{s.t. } f_i(x) \leq 0, \ i=1\ldots m \\ &\quad\ \ h_i(x) = 0, \ i=1\ldots p \end{aligned}$$

Lecture 8

2. Convex optimization problems.

$$\begin{aligned} \min \ & f_0(x) \\ \text{s.t. } & f_i(x) \leq 0, \ i=1,\ldots,m \\ & a_i^T x = b_i, \ i=1,\ldots,p \end{aligned}$$

where $f_0, f_1, \ldots, f_m$ are convex function.
the equality constraints are affine

+) If $f_0(x)$ is quasiconvex, the problem is quasiconvex (other $f_i(x)$ are still convex).

+) The feasible set of a convex optimization problem is convex.

$$D = \underbrace{\bigcap_{i=0}^{m} \operatorname{dom} f_i}_{\substack{\text{sublevel sets} \\ \{x \mid f_i(x) \leq 0\}}} \cap \underbrace{\bigcap_{i=1}^{p} \operatorname{dom} h_i}_{\text{hyperplanes } \{x \mid a_i^T x = b_i\}}$$

Example: $\min f_0(x) = -\sum_{i=1}^{k} \log(-a_i^T x + b_i)$
is an unconstrained problem with implicit constraints as

$$a_i^T x - b_i \leq 0 .$$

+) Feasibility problem: This kind of problems is to determine if a set of constraints is feasible (or consistent).

A feasibility problem can be written as

$$\begin{array}{llll} \text{find } x & & & \min\ 0 \\ \text{s.t. } f_i(x) \leq 0,\ i=1\ldots m & & \text{OR} & \text{s.t. } f_i(x) \leq 0,\ i=1\ldots m \\ \phantom{\text{s.t. }} h_i(x) = 0,\ i=1\ldots p & & & \phantom{\text{s.t. }} h_i(x) = 0,\ i=1\ldots p \end{array}$$

Lecture 8

2. Convex optimization problems.

$$\begin{aligned} \min\ & f_0(x) \\ \text{s.t. } & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & a_i^T x = b_i, \quad i = 1, \ldots, p \end{aligned}$$

where $f_0, f_1, \ldots, f_m$ are convex function.
the equality constraints are affine

+) If $f_0(x)$ is quasiconvex, the problem is quasiconvex (other $f_i(x)$ are still convex).

+) The feasible set of a convex optimization problem is convex.

$$D = \underbrace{\bigcap_{i=0}^{m} \operatorname{dom} f_i}_{\substack{\text{sublevel sets} \\ \{x \mid f_i(x) \leq 0\}}} \cap \underbrace{\bigcap_{i=1}^{p} \operatorname{dom} h_i}_{\text{hyperplanes } \{x \mid a_i^T x = b_i\}}$$

Note that this definition of convex problem is strict. A more general definition, or abstract convex problem, is that minimizes a convex function over a convex set. As it turns out, solving such an abstract convex optimization problem requires us to find a description of the set in terms of convex inequalities and linear equality constraints. For this reason, we focus on this restrict form of convex problems.

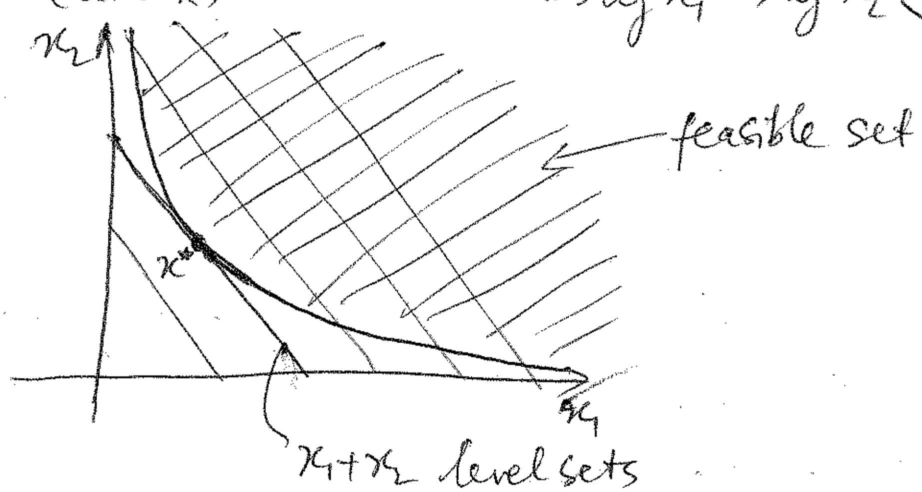
+) <u>Example</u>:

$$
\begin{aligned}
\min\ & x_1 + x_2 \\
\text{s.t.}\ & -x_1 \le 0 \\
& -x_2 \le 0 \\
& 1 - x_1 x_2 \le 0.
\end{aligned}
$$

This problem is not convex in the standard form since $f_3(x) = 1 - x_1 x_2$ is not convex on $\mathbb{R}^2_+$.
But this problem can be cast as a standard convex problem:

$$
\begin{aligned}
\min\ & x_1 + x_2 \\
\text{s.t.}\ & -x_1 \le 0 \\
& -x_2 \le 0 \\
& 1 - \sqrt{x_1 x_2} \le 1 \quad \text{(convex)}
\end{aligned}
\qquad \text{OR} \qquad
\begin{aligned}
\min\ & x_1 + x_2 \\
\text{s.t.}\ & -x_1 \le 0 \\
& -x_2 \le 0 \\
& -\log x_1 - \log x_2 \le 0.
\end{aligned}
$$

$x^* = (1,1)$
$p^* = 2$.



+) A convex optimization problem can also be written as

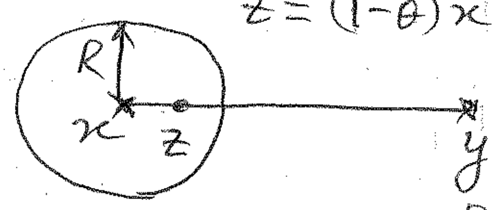
$$
\begin{aligned}
\min\ & f_0(x) \\
\text{s.t.}\ & f_i(x) \le 0 \qquad i = 1 \dots m \\
& Ax = b \qquad , A \in \mathbb{R}^{p\times n},\ b \in \mathbb{R}^p.
\end{aligned}
$$

+) <u>Local and global optimality:</u>

For a convex optimization problem, a fundamental property is that any locally optimal point is also globally optimal.

This property is easy to see:

$$z = (1-\theta)x + \theta y$$



By contradiction, suppose there exists a point y outside the local circle of radius R centered at x, such that $f_0(y) < f_0(x)$.

e.g. $\theta = \dfrac{R}{2\|y-x\|_2}$

Take a small step from x towards y: $z = (1-\theta)x + \theta y$ such that z is still inside the circle.

Then $\quad f_0(z) \leq (1-\theta) f_0(x) + \theta f_0(y) < f_0(x)$

$\rightarrow$ contradictory to the assumption that x is locally optimal.

+) <u>Optimality criterion for differentiable f_0:</u>

Recall the first-order condition of convex functions:

$$f(y) \geq f(x) + \nabla^T f(x)(y-x) \quad \forall x, y \in \operatorname{dom} f$$

Then for x^* to be optimal, it must satisfies

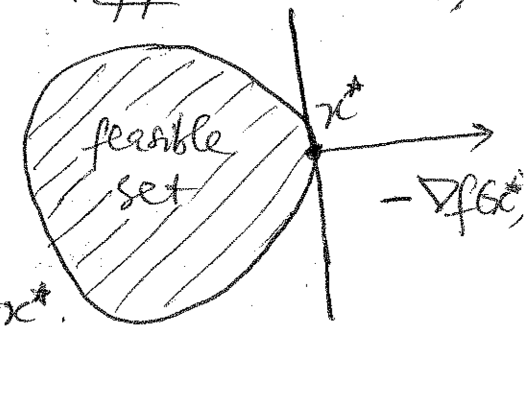
$$\nabla^T f(x^*)(y - x^*) \geq 0 \quad \forall y \text{ feasible.}$$

The converse is also true (i.e. this is an <u>iff</u> statement)

This condition means that:

either $\nabla f(x^*) = 0$

or $-\nabla f(x^*)$ defines a supporting hyperplane to the feasible set at x^*.



Special cases:

- Unconstrained problems: The optimality criterion reduces to the well-known condition

x^* optimal $\overset{\text{iff}}{\Longleftrightarrow} \nabla f_0(x^*) = 0$.

This is because for unconstrained problems, the feasible set is the domain of $f_0(x)$.

For differentiable $f_0(x)$ then its domain is an open set. Therefore all points surrounding x^* and sufficiently close are feasible.

Thus let $y = x^* - t\nabla f_0(x^*)$, for $t > 0$ and small enough then $y \in \mathrm{dom}\, f_0$.

$$\rightarrow 0 \leq \nabla f_0(x^*)^T (y - x^*) = -t\|\nabla f_0(x^*)\|_2^2$$

$$\rightarrow \nabla f_0(x^*) = 0.$$

Example: $\min f_0(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$, $a_i \in \mathbb{R}^n$.

$\mathrm{dom}\, f_0 = \{x \mid Ax < b\}$ is an open set.

Thus the optimality conditions are:

$$Ax < b \; ; \quad \nabla f_0(x) = \sum_{i=1}^{m} \frac{1}{b_i - a_i^T x} a_i = 0$$

Then we solve this set of equations and inequalities to find x^*.

- Problems with equality constraints only:

$$\min f_0(x)$$
$$\text{s.t. } Ax = b \quad , A \in \mathbb{R}^{p\times n}$$

The optimality condition is

$$\nabla^T f_0(x^*)(y - x^*) \geq 0 \text{ for all } y : Ay = b.$$

Since x, y are feasible $\rightarrow y = x + w$ for $w \in N(A)$, the null space of A.

Then the optimality conditions can be written as

$$\nabla f_0(x)^T w \geq 0 \quad \forall w \in \mathcal{N}(A)$$

$$\Longrightarrow \nabla f_0(x)^T w = 0 \quad \forall w \in \mathcal{N}(A) \text{ (since } \mathcal{N}(A) \text{ is}$$

a subspace, which contains w, $-w$)

Thus $\nabla f_0(x) \perp \mathcal{N}(A)$ or $\nabla f_0(x) \in \mathcal{N}(A)^{\perp} = \mathcal{R}(A^T)$.

That is, $\nabla f_0(x)$ can be written as

$$\nabla f_0(x) = A^T u \text{ for some } u \in \mathbb{R}^p.$$

In other words, x is optimal iff there exists a $v \in \mathbb{R}^p$ such that

$$\nabla f_0(x) + A^T v = 0$$
$$Ax = b.$$

This is the classical Lagrange multiplier (more in the duality topic).

+) Equivalent convex problems:

Two optimization problems are equivalent if from a solution of one, a solution of the other is readily found, and vice versa.

Next we describe some general transformation that yield equivalent problems.

The first two are general transformations. The others are transformation that also preserve convexity.

(i) Change of variable:

Suppose $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a one-to-one function with image covering the problem domain D. Let $x = \phi(z)$ and define

$$\tilde{f}_i(z) = f_i(\phi(z)) \quad i = 0, \dots, m$$
$$\tilde{h}_i(z) = h_i(\phi(z)) \quad i = 1, \dots, p$$

$$\nabla f_0(x)^T w \geq 0 \quad \forall w \in N(A)$$

$$\Longrightarrow \nabla f_0(x)^T w = 0 \quad \forall w \in N(A)$$ (since $N(A)$ is a subspace, which contains $w, -w$)

Thus $\nabla f_0(x) \perp N(A)$ or $\nabla f_0(x) \in N(A)^\perp = R(A^T)$.

That is, $\nabla f_0(x)$ can be written as

$$\nabla f_0(x) = A^T u \text{ for some } u \in \mathbb{R}^p.$$

In other words, x is optimal iff there exists a $v \in \mathbb{R}^p$ such that

$$\nabla f_0(x) + A^T v = 0$$
$$Ax = b.$$

This is the classical Lagrange multiplier (more in the duality topic).

Lecture 9:

+) Equivalent convex problems:

Two optimization problems are equivalent if from a solution of one, a solution of the other is readily found, and vice versa.

Next we describe some general transformation that yield equivalent problems.
The first two are general transformations. The others are transformation that also preserve convexity.

(i) Change of variable:

Suppose $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a one-to-one function with image covering the problem domain D. Let $x = \phi(z)$ and define

$$\tilde{f}_i(z) = f_i(\phi(z)) \qquad i = 0, \dots, m$$
$$\tilde{h}_i(z) = h_i(\phi(z)) \qquad i = 1, \dots, p$$

then the following problem is equivalent to the original prob:

$$\begin{aligned} \min \quad & \tilde{f}_0(z) \\ \text{s.t.} \quad & \tilde{f}_i(z) \le 0 \quad i = 1 \dots m \\ & \tilde{h}_i(z) = 0 \quad i = 1 \dots p. \end{aligned}$$

(ii) Transformation of objective and constraint functions:

Suppose that

$\varphi_0 : \mathbb{R} \to \mathbb{R}$ is monotone increasing

$\varphi_1 \dots \varphi_m : \mathbb{R} \to \mathbb{R}$ satisfy $\varphi_i(u) \le 0$ iff $u \le 0$

$\varphi_{m+1} \dots \varphi_{m+p} : \mathbb{R} \to \mathbb{R}$ satisfy $\varphi_i(u) = 0$ iff $u = 0$

then we can define

$$\tilde{f}_i(x) = \varphi_i(f_i(x)) \qquad i = 0 \dots m$$

$$\tilde{h}_i(x) = \varphi_{m+i}(h_i(x)) \quad i = 1 \dots p$$

and obtain an equivalent problem

$$\begin{aligned} \min \quad & \tilde{f}_0(x) \\ \text{s.t.} \quad & \tilde{f}_i(x) \le 0 \quad i = 1 \dots m \\ & \tilde{h}_i(x) = 0 \quad i = 1 \dots p \end{aligned}$$

Example: $\min \ \|Ax - b\|_2 \quad (= f_0(x))$

is equivalent to

$$\min \ (Ax-b)^T(Ax-b) \qquad (= \tilde{f}_0(x))$$

Note that $f_0(x)$ is not differentiable at $Ax - b = 0$, but $\tilde{f}_0(x)$ is differentiable everywhere.

(iii) Eliminating equality constraints:

A convex problem in standard form

$$\min f_0(x)$$
$$\text{s.t. } f_i(x) \le 0, \quad i=1\dots m$$
$$Ax = b, \quad A \in \mathbb{R}^{p\times n}$$

is equivalent to

$$\min_z f_0(Fz + x_0)$$
$$\text{s.t. } f_i(Fz+x_0) \le 0, \quad i = 1\dots m$$

where F is a matrix whose range is the null space of A, that is

$$Ax = b \iff x = Fz + x_0 \text{ for some } z,$$
and x_0 s.t. $Ax_0 = b$.

Note that $AF = 0$.

This transformation preserves convexity since composition of a convex function with an affine function preserves convexity.

In many cases, however, it is better (more efficient) to retain the equality constraints, especially when x is of large dimension, then equality constraints can provide sparsity or some other useful structure.

(iv) <u>Introducing equality constraints</u>:

$$\min f_0(A_0 x + b_0) \qquad A_i \in \mathbb{R}^{k_i \times n}$$
$$\text{s.t. } f_i(A_i x + b_i) \le 0 \quad i = 1\dots m$$

is equivalent to

$$\min_{x, y_i} f_0(y_0)$$
$$\text{s.t. } f_i(y_i) \le 0 \quad i = 1,\dots, m$$
$$y_i = A_i x + b_i \quad i = 0, 1, \dots m$$

independent objective & inequality constraints.

This problem has $k_0 + \dots + k_m$ new variables and $k_0 + \dots + k_m$ new equality constraints.

(v) Introducing Slack variables:

Observe that $f_i(x) \leq 0$ iff there is an $s_i \geq 0$ such that
$$f_i(x) + s_i = 0.$$

Using this transformation, we obtain an equivalent problem
$$\begin{aligned} \min\ & f_0(x) \\ \text{s.t.}\ & s_i \geq 0, \quad i=1\ldots m \\ & f_i(x) + s_i = 0, \quad i=1\ldots m \\ & h_i(x) = 0, \quad i=1\ldots p \end{aligned}$$

where the variables of this new problem are x, s_i ($i=1\ldots m$) (or $s \in \mathbb{R}^m$). This new problem has $n+m$ variables and $m+p$ equality constraints.

(vi) Minimizing (optimizing) over some variables:

Observe that
$$\inf_{x,y} f(x,y) = \inf_x \tilde{f}(x)$$
where
$$\tilde{f}(x) = \inf_y f(x,y)$$

Thus we can partition a big problem into smaller problem(s)
$$\begin{aligned} \min\ & f_0(x_1, x_2) \\ \text{s.t.}\ & \left.\begin{aligned} & f_i(x_1) \leq 0, \quad i=1\ldots m_1 \\ & \tilde{f}_i(x_2) \leq 0, \quad i=1\ldots m_2 \end{aligned}\right\} \text{note these are independent constraints} \end{aligned}$$

Since the constraints are independent (one set depends on x_1 only whereas the other set depends on x_2 only) or separable, we can first minimize over **one** set of variables first.

Note here $x \in \mathbb{R}^n$, $x = (x_1, x_2)$ where $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$, $n_1 + n_2 = n$.

Define $\tilde{f}_0(x_1) = \inf\{ f_0(x_1, z) \mid \tilde{f}_i(z) \leq 0, \ i=1 \dots m_2 \}$.
then the original problem is equivalent to

$$\begin{aligned} &\min \ \tilde{f}_0(x_1) \\ &\text{s.t. } f_i(x_1) \leq 0, \ i=1 \dots m_1 \end{aligned}$$

<u>Example:</u>

$$\begin{aligned} &\min \ x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2 \\ &\text{s.t. } f_i(x_1) \leq 0, \quad i=1, \dots, m \end{aligned}$$

We can analytically minimize the objective over x_2:

$$\begin{aligned} &\inf_{x_2} \left(x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2 \right) \\ &= x_1^T (P_{11} - P_{12} P_{22}^{-1} P_{12}^T) x_1 \qquad (\text{assuming } P_{22} \succ 0) \end{aligned}$$

Thus the problem is equivalent to

$$\begin{aligned} &\min \ x_1^T (P_{11} - P_{12} P_{22}^{-1} P_{12}^T) x_1 \\ &\text{s.t. } f_i(x_1) \leq 0, \quad i=1, \dots, m \end{aligned}$$

(vii) <u>Epigraph form:</u> The original problem is equivalent to

$$\begin{aligned} &\min \ t \\ &\text{s.t. } f_0(x) - t \leq 0 \\ &\qquad f_i(x) \leq 0, \ i=1 \dots m \\ &\qquad a_i^T x = b_i, \ i=1 \dots p \end{aligned}$$

This new problem is the epigraph form and is convex, since the objective is linear and constraints are convex.

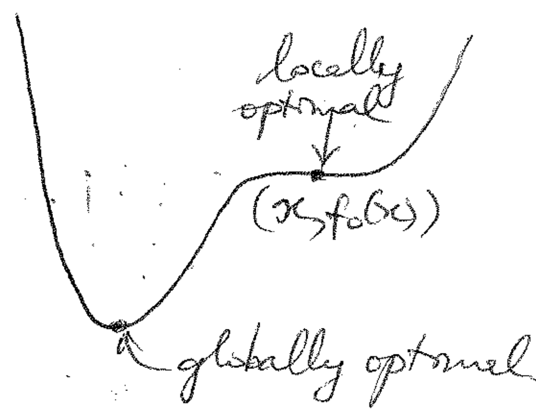
This epigraph form can be useful for theoretical analysis since it transforms any convex problem into a convex problem with linear objective.

+) <u>Quasiconvex optimization problems:</u>

$$\begin{aligned} \min\ & f_0(x) \\ \text{s.t.}\ & f_i(x) \leq 0 \quad , i = 1, \dots, m \\ & Ax = b \quad , A \in \mathbb{R}^{p \times n} \end{aligned}$$

where $f_0(x)$ is quasiconvex
$f_i(x)$ is convex $i = 1 \dots m$.

Quasiconvex problems can have locally optimal points that are not globally optimal.



Quasiconvex optimization can be solved through a series of convex optimizations.

∘ Sublevel sets are convex: We can find convex representation of sublevelsets of $f_0(x)$.

If $f_0(x)$ is quasiconvex $\rightarrow \exists$ a family of convex function indexed by t, $\phi_t : \mathbb{R}^n \rightarrow \mathbb{R}$, such that

$$f_0(x) \leq t \iff \phi_t(x) \leq 0$$

That is, the t-sublevel set of $f_0(x)$ is the 0-sublevel set of $\phi_t(x)$.

A trivial example of such an representation is the indicator function
$$\phi_t(x) = \begin{cases} 0 & \text{if } f(x) \leq t \\ \infty & \text{otherwise} \end{cases}$$
But we can often find "nicer" functions that are differentiable

<u>Example</u>: $f_0(x) = \dfrac{p(x)}{q(x)}$ where $p(x)$ is convex, $q(x)$ concave then $f_0(x)$ is quasiconvex (assuming $q(x) > 0$, $p(x) \geq 0$).

Then $f(x) \leq t \iff p(x) - tq(x) \leq 0$
and we can take $\phi_t(x) = p(x) - tq(x)$ for $t \geq 0$.
For each t, $\phi_t(x)$ is convex in x.

- Quasiconvex optimization via convex feasibility problems.

Let p^* be the optimal value of the quasiconvex problem.
Consider the feasibility problem, for a fixed t,:

$$
\begin{aligned}
&\text{find } x \\
&\text{s.t. } \phi_t(x) \leq 0 \\
&\quad\;\; f_i(x) \leq 0, \; i = 1 \ldots m \\
&\quad\;\; Ax = b
\end{aligned}
$$

If this problem is feasible then $p^* \leq t$, otherwise $p^* \geq t$.
Thus we can use a bisection method for quasiconvex optimization: First obtain a lower bound l and an upper bound u on the objective function $f_0(x)$.

given $l \leq p^*$, $u \geq p^*$ and tolerence $\varepsilon > 0$
repeat
1. $t = \frac{1}{2}(l + u)$
2. Solve the (convex) feasibility problem
3. If problem is feasible then $u := t$
   else $l := t$

until $u - l \leq \varepsilon$.

This algorithm requires exactly $\lceil \log_2 (\frac{u-l}{\varepsilon}) \rceil$ iterations.

<u>Example</u>: $\min (a^T x + b)/(c^T x + d)$ is quasiconvex optimization.

$$
\begin{aligned}
\text{s.t. } & Ax = b \\
& Fx \leq g \\
& c^T x + d > 0
\end{aligned}
$$

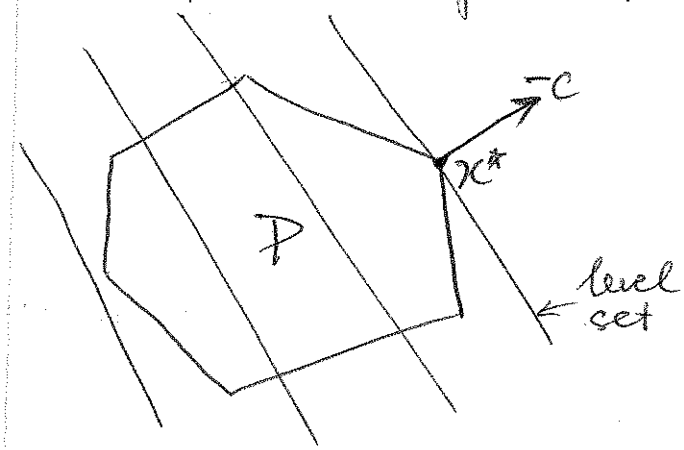
Lecture 10:

3. Linear program (LP)

$$\begin{aligned} \min \quad & c^T x + d \leftarrow \text{often omitted} \\ \text{s.t.} \quad & Gx \preceq h \\ & Ax = b \end{aligned}$$

All objective & constraints are affine → linear program (LP)

The feasible set of an LP is a polyhedron P.



Level sets $\{x \mid c^T x = \alpha\}$ are hyperplanes.

The optimal point x^* is as far as possible in P along the $-c$ direction

+) Standard form LP: An LP can be converted into a standard form as

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \succeq 0. \end{aligned}$$

(only equality constraints and $x \succeq 0$).

Another form of LP is an inequality form LP:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \preceq b \end{aligned}$$

Note that any general LP can be transformed into the standard form above by introducing slack variables for the inequality constraints and writing $x = x^+ - x^-$ where $x^+, x^- \succeq 0$. Then the problem become the standard form in variables (x^+, x^-, s):

$$\begin{aligned} \min \quad & c^T x^+ - c^T x^- \\ \text{s.t.} \quad & Gx^+ - Gx^- + s = h \\ & Ax^+ - Ax^- = b \\ & x^+ \succeq 0, \; x^- \succeq 0, \; s \succeq 0. \end{aligned}$$

Example: (i) Chebyshev center of a polyhedron

Find the largest Euclidean ball that lies inside a polyhedron

polyhedron $P = \{x \in \mathbb{R}^n \mid a_i^T x \le b_i \ , \ i = 1 \ldots m\}$

ball $B = \{x_c + u \mid \|u\|_2 \le r\}$

→ we want to find x_c and the largest r such that $B \subseteq P$.

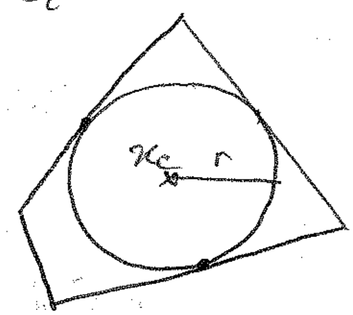
To formulate the problem: For a point in B to satisfy a constraint of P then

$$\forall \ \|u\|_2 \le r \implies a_i^T (x_c + u) \le b_i$$

Since $\sup\{a_i^T u \mid \|u\|_2 \le r\} = r\|a_i\|_2$

then we must have

$$a_i^T x_c + r\|a_i\|_2 \le b_i .$$



Thus we can formulate an LP as

$$\max \ r$$
$$\text{s.t.} \ a_i^T x_c + \|a_i\|_2 r \le b_i$$

Note the optimization variables are x_c and r.

(ii) Piece-wise linear minimization:

$$\min \ \max_i \{a_i^T x + b\}$$

is equivalent to an LP

$$\min \ t$$
$$\text{s.t.} \ a_i^T x + b_i \le t \quad , i = 1 \ldots m.$$

The variable of this LP is (t, x).

+) Linear fractional program:

This is a quasiconvex problem that can be converted into an LP.

linear fractional prog.

LP

(quasiconvex problem, can also be solved via feasibility bisection)

$$\min \frac{c^T x + d}{e^T x + f} \quad \text{s.t.} \quad Gx \preceq h, \quad Ax = b \quad \text{(variable } x) \iff \min \; c^T y + dz \quad \text{s.t.} \quad Gy \preceq hz, \; Ay = bz, \; e^T y + fz = 1, \; z \geq 0 \quad \text{(variable } y, z)$$

These two problems are equivalent since we can map

$$y = \frac{x}{e^T x + f}, \quad z = \frac{1}{e^T x + f} \quad (\text{or } y = xz)$$

+) Generalized linear fractional programming:

$$f_0(x) = \max_{i=1\ldots r} \frac{c_i^T x + d_i}{e_i^T x + f_i}$$

Example: Von Neumann model for a growing economy:

$x \in \mathbb{R}^n$: current activity level in n sectors.

$x^+ \in \mathbb{R}^n$: next period activity level

$Bx \in \mathbb{R}^m$: consumption of activity level x

$Ax \in \mathbb{R}^m$: produce of activity level x.

Then want to maximize the growth rate x_i^+/x_i subject to the constraint that good consumed in the next period cannot exceed good produced in the current period

$$\max \; \min_{i=1\ldots n} \frac{x_i^+}{x_i} \quad \text{s.t.} \quad Bx^+ \preceq Ax, \quad x^+ \geq 0.$$

$\rightarrow$ (generalized) linear fractional program

4. Quadratic optimization problem (QP)

$$\min \; \frac{1}{2} x^T P x + q^T x + r \quad \text{s.t.} \quad Gx \preceq h, \quad Ax = b$$

$P \in S_+^n$ (PSD)

$G \in \mathbb{R}^{m \times n}$, $h \in \mathbb{R}^m$.

$A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$.

In QP, we minimize a convex quadratic function over a polyhedron.

Example: Least-square or regression

$$\min \|Ax - b\|^2$$
$$\text{s.t. } l_i \leq x_i \leq u_i, \quad i = 1 \dots n$$

Without constraints, least-square problems have closed form solutions as $x = A^{+}b$ where A^{+} is the pseudoinverse

With even linear constraints as above, however, no closed form analytical solutions exist. But we can build efficient algorithm to solve the problem.

+) Quadratically constrained quadratic program (QCQP):

This is a special case of QP:

$$\min \frac{1}{2} x^T P x + q^T x + r$$
$$\text{s.t. } \frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1 \dots m$$
$$Ax = b$$

The quadratic constraints define ellipsoids feasible sets if $P_i \in S_{++}^n$ (positive definite).

+) Second-order cone programming (SOCP)

$$\min f^T x$$
$$\text{s.t. } \|A_i x + b_i\|_2 \leqslant c_i^T x + d_i, \quad i = 1 \dots m$$
$$Fx = g$$

where $A_i \in \mathbb{R}^{n_i \times m}$, $F \in \mathbb{R}^{p \times n}$.

Here the inequality constraints are called cone constraints.

if $c_i = 0 \rightarrow$ SOCP becomes a QCQP
$A_i = 0 \rightarrow$ SOCP becomes an LP.

Example: Robust linear programming is an SOCP.

$$\min c^T x$$
$$\text{s.t. } a_i^T x \le b_i \quad i=1,\dots,m$$

where there can be uncertainty in a_i, b_i and c.

- Deterministic approach:

Assume $a_i \in \mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \le 1\}$
that is, a_i lies in an "uncertainty" ellipsoid.

Then $a_i^T x \le b_i \quad \forall a_i \in \mathcal{E}_i$ can be expressed as

$$b_i \ge \sup\{a_i^T x \mid a_i \in \mathcal{E}_i\}$$
$$= \bar{a}_i^T x + \sup\{u^T P_i^T x \mid \|u\|_2 \le 1\}$$
$$= a_i^T x + \|P_i^T x\|_2 .$$

Thus the robust LP problem can be expressed as

$$\min c^T x$$
$$\text{s.t. } a_i^T x + \|P_i^T x\|_2 \le b_i$$

- Stochastic approach:

In a statistical framework, suppose that a_i are independent random vectors that are Gaussian with mean $\bar{a}_i$ and covariance Σ_i.

Formulate a robust optimization problem as

$$\min c^T x$$
$$\text{s.t. } \operatorname{prob}(a_i^T x \le b_i) \ge \eta \quad , i=1\dots m$$

That is, we require that each constraint holds with a probability exceeding η $(\eta > 0.5)$.

For Gaussian random a_i, this probability can be computed as follows.

Let $u_i = a_i^T x$ then u is a Gaussian random variable with mean $\bar{u}_i = \bar{a}_i^T x$ and variance $\sigma_i^2 = x^T \Sigma_i x$.

Now $\Pr(u_i \le b_i) = \Pr\left(\frac{u_i - \bar{u}_i}{\sigma_i} < \frac{b_i - \bar{u}_i}{\sigma_i}\right)$

$$= \frac{1}{2\pi} \int_{-\infty}^{z_i} e^{-t^2/2} dt = \Phi(z_i), \quad z_i = \frac{b_i - \bar{u}_i}{\sigma_i}$$

(t is a standard Gaussian $N(0,1)$, $\Phi(z)$ is its known CDF)

Thus the constraints are equivalent to

$$\Phi\left(\frac{b_i - \bar{u}_i}{\sigma_i}\right) \ge \eta \quad , i = 1 \dots m$$

$$\Leftrightarrow \frac{b_i - \bar{u}_i}{\sigma_i} \ge \Phi^{-1}(\eta) \Leftrightarrow \bar{u}_i + \Phi^{-1}(\eta)\sigma_i \le b_i \; , i = 1 \dots m$$

$$\Leftrightarrow \bar{a}_i^T x + \Phi^{-1}(\eta) \cdot (x^T \Sigma_i x)^{1/2} \le b_i$$

Then the robust LP problem becomes

$$\min c^T x$$
$$\text{s.t. } \bar{a}_i^T x + \Phi^{-1}(\eta) \cdot \|\Sigma_i^{1/2} x\|_2 \le b_i$$

which is a SOCP problem.

Lecture 11:

5. Geometric programming:

This type of problems is not convex in their natural form but can be transformed into convex problems.

+) Monomial and posynomials:

Monomial $f: \mathbb{R}^n \to \mathbb{R}$, $\operatorname{dom} f = R^n_{++}$:

$$f(x) = c x_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \quad , c > 0.$$

$a_i \in \mathbb{R}$ can be any number (+ve or -ve), but $c > 0$.

Note that $f(x) > 0$ since $\operatorname{dom} f = R^n_{++}$.

Let $u_i = a_i^T x$ then u is a Gaussian random variable with mean $\bar{u}_i = \bar{a}_i^T x$ and variance $\sigma_i^2 = x^T \Sigma_i x$.

Now $\Pr(u_i \le b_i) = \Pr\left(\frac{u_i - \bar{u}_i}{\sigma_i} < \frac{b_i - \bar{u}_i}{\sigma_i}\right)$

$$= \frac{1}{2\pi} \int_{-\infty}^{z_i} e^{-t^2/2} dt = \Phi(z_i), \quad z_i = \frac{b_i - \bar{u}_i}{\sigma_i}$$

(t is a standard Gaussian $N(0,1)$, $\Phi(z)$ is its known CDF.)

Thus the constraints are equivalent to

$$\Phi\left(\frac{b_i - \bar{u}_i}{\sigma_i}\right) \ge \eta \quad , i = 1 \dots m$$

$$\Leftrightarrow \frac{b_i - \bar{u}_i}{\sigma_i} \ge \Phi^{-1}(\eta) \Leftrightarrow \bar{u}_i + \Phi^{-1}(\eta)\sigma_i \le b_i, \; i = 1 \dots$$

$$\Leftrightarrow \bar{a}_i^T x + \Phi^{-1}(\eta) \cdot (x^T \Sigma_i x)^{1/2} \le b_i$$

Then the robust LP problem becomes

$$\min c^T x$$
$$\text{s.t. } \bar{a}_i^T x + \Phi^{-1}(\eta) \cdot \|\Sigma_i^{1/2} x\|_2 \le b_i$$

which is a SOCP problem.

Lecture 11:

5. Geometric programming:

This type of problems is not convex in their natural form but can be transformed into convex problems.

+) Monomial and posynomials:

Monomial $f: \mathbb{R}^n \to \mathbb{R}$, $\operatorname{dom} f = R^n_{++}$:

$$f(x) = c x_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \quad , c > 0.$$

$a_i \in \mathbb{R}$ can be any number (+ve or +ve), but $c > 0$.

Note that $f(x) > 0$ since $\operatorname{dom} f = R^n_{++}$.

Posynomial : $f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} x_2^{a_{2k}} \dots x_n^{a_{nk}}, \ c_k > 0$

= Sum of monomials.

Monomials are closed under multiplication & division.
Posynomials are closed under addition, multiplication and non-negative scaling.

+) Geometric programming (GP).

$$\begin{aligned} \min \ & f_0(x) \\ \text{s.t.} \ & f_i(x) \le 1, \ i = 1 \dots m \\ & h_i(x) = 1, \ i = 1 \dots p \end{aligned}$$

where $f_i(x)$ are posynomials
$h_i(x)$ are monomials

Here domain of the problem is $D = R^n_{++}$ so the constraint $x > 0$ is implicit.

<u>Example:</u>

$$\begin{aligned} \max \ & x/y \\ \text{s.t.} \ & 2 \le x \le 3 \\ & x^2 + 3y/z \le \sqrt{y} \\ & x/y = z^2 \end{aligned} \iff \begin{aligned} \min \ & x^{-1}y \\ \text{s.t.} \ & 2x^{-1} \le 1, \ \tfrac{1}{3}x \le 1 \\ & x^2 y^{-1/2} + 3y^{1/2}z^{-1} \le 1 \\ & x y^{-1} z^{-2} = 1 \end{aligned}$$

+) GP in convex form.

Define $y_i = \log x_i \iff x_i = e^{y_i}$.

Then
$$\begin{aligned} f(x) &= c\, x_1^{a_1} x_2^{a_2} \dots x_n^{a_n} \\ &= c\,(e^{y_1})^{a_1} (e^{y_2})^{a_2} \dots (e^{y_n})^{a_n} \\ &= e^{a^T y + b}, \quad b = \log_e c. \end{aligned}$$

$\Rightarrow \log f(x) = a^T y + b \quad , \; y = \log x \, , \; b = \log c.$

Similarly for posynomials

$$f(x) = \sum_{k=1}^{K} c_k \, x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}$$

$$\rightarrow \log f(x) = \log \left(\sum_{k=1}^{K} e^{a_k^T x + b_k} \right) \qquad b_k = \log c_k$$

Thus a GP can be transformed into a convex problem as

$$\min \log \left(\sum_{k=1}^{K} \exp \left(a_{0k}^T x + b_{0k} \right) \right)$$

$$\text{s.t. } \log \left(\sum \exp (a_{ik}^T x + b_{ik}) \right) \leq 0 \, , \; i = 1 \dots m$$

$$Gy + d = 0$$

This form is also refered to as GP in convex form (versus GP in posynomial form).

If all posynomials in a GP have only one term then the GP can be transformed into an LP.

<u>Example</u>: (i) Perron - Frobenius eigenvalue $\lambda_{pf}(A)$

$A \in \mathbb{R}^{n \times n}$ is element-wise positive (that is $A_{ij} \geq 0 \; \forall i,j$)

Then $\lambda_{pf}(A) = \max_i |\lambda_i(A)|$ exists.

This PF eigenvalue determines the rate of growth or decay of A^k as $k \to \infty$, often used in control.

Given that the PF eigenvalue can be characterized as

$$\lambda_{pf} = \inf \{ \lambda \mid Av \preceq \lambda v \text{ for some } v \succ 0 \}$$

Then a problem of minimizing λ_{pf} of $A(x)$ where the entries of $A(x)$ are posynomials is a GP

$$\min \lambda$$

$$\text{s.t. } \sum_{j=1}^{n} A_{ij} v_j / (\lambda v_i) \leq 1 \quad , \; i = 1 \dots n.$$

6) <u>Semidefinite programming</u> (SDP)

$$\min \; c^T x$$
$$\text{s.t. } x_1 F_1 + x_2 F_2 + \dots + x_n F_n + F_0 \preceq 0$$
$$Ax = b$$

where $F_i \in S^k$ and $A \in \mathbb{R}^{p \times n}$.

In a SDP, the inequality constraint is a linear matrix inequality (LMI).
If the matrices F_i are diagonal then the SDP reduces to an LP.

<u>Example:</u> (i) LP as SDP:

$$\begin{array}{ccc} \min c^T x & & \min c^T x \\ \text{s.t. } Ax \leq b & \Leftrightarrow & \text{s.t. } \operatorname{diag}(Ax-b) \preceq 0 \\ \text{(LP)} & & \text{(SDP)} \end{array}$$

(ii) Matrix norm minimization:
Let $A(x) = A_0 + x_1 A_1 + \dots + x_n A_n$, $A_i \in \mathbb{R}^{p \times q}$.
Consider unconstrained problem

$$\min \; \|A(x)\|_2$$

Here $\|A(x)\|_2 = \max_i |\sigma_i(A)|$ where σ_i are singular values of A.
We use the fact

$$\|A\|_2 \leq s \Leftrightarrow A^T A \preceq s^2 I \quad (s \geq 0)$$

then the problem becomes

$$\min \; s$$
$$\text{s.t. } A(x)^T A(x) \preceq s^2 I \qquad \text{(variables } x, s)$$

Now using

$$A^T A \preceq t^2 I \Leftrightarrow \begin{bmatrix} tI & A \\ A^T & tI \end{bmatrix} \succeq 0$$

we obtain the SDP

$$\min \; t$$
$$\text{s.t. } \begin{bmatrix} tI & A(x) \\ A(x)^T & tI \end{bmatrix} \succeq 0. \qquad \text{(variables } x, t)$$

7) Some other examples of convex optimization problems:

+) Statistical estimation: Covariance estimation for Gaussian variab

Suppose $y \in \mathbb{R}^n$ is a random vector, $y \in \mathcal{N}(0, R)$

$$R = E\, y y^T, \quad R \in S^n_{++} \quad \text{is the covariance}$$

We want to estimate this covariance from N independent sampl $\{y_1, \dots, y_N\}$.

$$f_R(y) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(R)^{1/2}} e^{-\frac{1}{2} y^T R^{-1} y}.$$

The objective is to maximize the log-likelihood function, under some prior knowledge about R.

$$l(R) = \log f_R(y_1, y_2, \dots, y_N)$$

$$= -N\frac{n}{2} - \frac{N}{2}\log\det(R) + \left(-\frac{1}{2}\sum_{k=1}^{N} y_k^T R^{-1} y_k\right)$$

Let $\quad Y = \frac{1}{N}\sum_{k=1}^{N} y_k^T y_k$

$$\rightarrow l(R) = -\frac{N}{2}n - \frac{N}{2}\log\det R - \frac{N}{2}\operatorname{tr}(R^{-1}Y)$$

Denote $S = R^{-1}$ then $l(R)$ is concave in S.

Suppose that the prior knowledge is that there is a conditi number constraint:

$$\frac{\lambda_{max}(R)}{\lambda_{min}(R)} \leq K$$

Then we can formulate a convex problem

$$\max \ \log\det S - \operatorname{tr}(SY)$$

$$\text{s.t.} \quad uI \preccurlyeq S \preccurlyeq K.uI$$

where the variables are $\quad S \in S^n_{++}$ and $u > 0, u \in \mathbb{R}$

f) Geometric problems: Extremal volume ellipsoids.

(i) Find the minimum volume ellipsoid that contains a set C:

$$\mathcal{E} = \{v \mid \|Av+b\|_2 \leq 1 \ \forall v \in C\}.$$

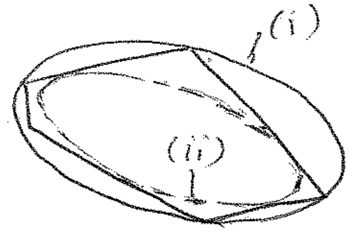
Assuming that $A \in S^n_{++}$ which is invertiable, then the volume of the ellipsoid is proportional to $\det A^{-1}$.

Then the minimum volume can be found through an opt. prob.

$$\begin{aligned} \min \ & \log\det A^{-1} \\ \text{s.t.} \ & \sup_{v\in C} \|Av+b\|_2 \leq 1 \end{aligned}$$

This problem is convex, but evaluating the constraint can be hard (for general C).

For a finite set $C = \{x_1, \dots, x_m\}$.



$$\begin{aligned} \min \ & \log\det A^{-1} \\ \text{s.t.} \ & \|Ax_i+b\|_2 \leq 1, \quad i=1\dots m \end{aligned}$$

→ this problem solves for the Löwner-John ellipsoid for the polyhedron $\mathrm{conv}\{x_1, \dots, x_m\}$. (Covering this polyhedron)

(ii) Find the maximum volume ellipsoid contained in a set:

For polyhedron

$$C = \{x \mid a_i^T x \leq b_i, \ i=1\dots m\}$$

want to find the max ellipsoid

$$\mathcal{E} = \{Bu+d \mid \|u\|_2 \leq 1\}$$ that is inside set C.

Constraint: $\sup_{\|u\|_2\leq 1} a_i^T(Bu+d) \leq b_i \iff \|Ba_i\|_2 + a_i^T d \leq b_i$

Thus we can formulate

$$\begin{aligned} \min \ & \log\det B^{-1} \\ \text{s.t.} \ & \|Ba_i\|_2 + a_i^T d \leq b_i. \end{aligned}$$

+) <u>Generalized posynomials and GP:</u>

$f_i : \mathbb{R}^n \to \mathbb{R}$ posynomials, $i = 1 \dots k$.

$\phi : \mathbb{R}^k \to \mathbb{R}$ polynomial with nonnegative coefficient, but can have fractional exponents.

Then $h(x) = \phi(f_1(x), \dots, f_k(x))$

is a generalized posynomial.

<u>Example:</u> $\phi_1(z_1, z_2) = 3z_1^2 z_2 + 2z_1 + 3z_2^3$

$\rightarrow h = 3f_1^2 f_2 + 2f_1 + f_2^3$ is a posynomial

$\phi_2(z_1, z_2) = 2z_1^{0.3} z_2^{1.2} + z_1 z_2^{0.5} + 2$

$\rightarrow h = 2f_1^{0.3} f_2^{1.2} + f_1 f_2^{0.5} + 2$ is a generalized posynom

Generalized GP:

$$\begin{aligned} \min \ & h_0(x) \\ \text{s.t. } & h_i(x) \le 1 \quad i = 1 \dots m \\ & g_i(x) = 1 \quad i = 1 \dots p \end{aligned}$$

$g_i(x)$: monomials, $h_i(x)$: generalized posynomial

This problem can be transformed into a GP by introducing new variables.