Review for the second Midterm

October 30, 2015

The second midterm is focused on **constrained optimization** and related topics. **No least square**, but you are definitely going to see it in the final. Most of the examples are worked during the lectures, and the details of some problems are not given here.

Contents

1 Basic Definitions

Standard form: The *standard form* for constrained optimization is

$$
\begin{aligned}
&\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) \\
&\text{subject to} \\
&\qquad c_i(x) = 0, \quad i \in \mathcal{E}, \\
&\qquad c_i(x) \geq 0, \quad i \in \mathcal{I}.
\end{aligned}
\tag{1}
$$

The reason for introducing the standard form is that the necessary or sufficient conditions for optimality can be **different by a sign**.

Example 1.1. Convert the following optimization into an equivalent standard form.

$$
\begin{aligned}
&\max \quad x_1 + x_2 \\
&\text{subject to} \\
&\qquad x_1^2 + x_2^2 - 2 \leq 0
\end{aligned}
$$

(Hint: there are two places to change: max $\rightarrow$ min for the objective function and $\leq \rightarrow \geq$ for the constraint.) $\qquad\square$

Active set $\mathcal{A}(x)$ at x: The *active set* $\mathcal{A}(x)$ at any feasible point x

$$
\mathcal{A}(x) = \mathcal{E} \bigcup \{i \in \mathcal{I} \mid c_i(x) = 0\}.
$$

Example 1.2. Let the feasible set Ω defined by

$$
\begin{aligned}
\Omega = \{(x_1, x_2, x_3) \mid\ &c_1(x) = x_1 + x_2 + x_3 = 1, \\
&c_2(x) = x_1^2 + x_2^2 + x_3^2 \leq 1, \\
&c_3(x) = x_3 \geq 0.\}
\end{aligned}
$$

Find $\mathcal{E}$, $\mathcal{I}$ and the active set $\mathcal{A}(x)$ for the points

$$
(1,0,0), \quad (0,1,0), \quad (0,0,1), \quad (1/2, 1/2, 0). \qquad\square
$$

For the optimization problem (1), if we know the active set $\mathcal{A}$ at the minimizer x^*, we can solve the equivalent problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)$$

$$\text{subject to} \tag{2}$$

$$c_i(x) = 0, \quad i \in \mathcal{A}.$$

This usually simplifies the problem a lot (that's why we start with equality constraints first), and we can take the inactive constraints just as if they are not there . The real difficult is we don't know the active set at the minimizer unless we solve it.

Lagrange Multiplier: Equality constrained optimization problems are usually solved using Lagrange multipliers. Even for inequality constrained problems, we can solve it in a similar way by assuming different active sets.

Example 1.3 (AM-GM inequality). Let $x_1, x_2, \ldots, x_n$ be n non-negative numbers, show that

$$(x_1 x_2 \ldots x_n)^{1/n} \le \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

This can be formulated as the optimization problem

$$\min \quad \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\text{subject to} \quad x_1 x_2 \ldots x_n = G^n$$

or equivalently

$$\max \quad x_1 x_2 \ldots x_n$$

$$\text{subject to} \quad \frac{x_1 + x_2 + \cdots + x_n}{n} = A.$$

Example 1.4 (Maximum entropy). Assume a physical system has n states $i = 1, 2, \ldots, n$ with energy ϵ_i in i-th states. Let p_i be the probability in i-th state, find the maximum of the probability Let p_i be the probability in i-th state with total energy E, find the probabilities p_i such that maximize the entropy $S = -\sum_i p_i \log p_i$.

Solution: This problem can be formulated as

$$\max_{p_i} \quad S = -\sum_i p_i \log p_i$$

$$\text{subject to} \quad \sum_{i=1}^{n} p_i = 1$$

$$\sum_{i=1}^{n} \epsilon_i p_i = E$$

The Lagrange function is

$$L(p, \lambda, \mu) = -\sum_{i=1}^{n} p_i \log p_i - \lambda \left(\sum_{i=1}^{n} p_i - 1 \right) - \mu \left(\sum_{i=1}^{n} \epsilon_i p_i - E \right).$$

The minimizer p^* satisfies

$$\frac{\partial L}{\partial p_i} = -1 - \log p_i - \lambda - \mu \epsilon_i = 0$$

2

Therefore, p_i can be represented as

$$p_i = e^{-1-\lambda-\mu\epsilon_i}.$$

Substituting it into the two constraints,

$$1 = \sum_{i=1}^{n} p_i = e^{-1-\lambda} \sum_{i=1}^{n} e^{-\mu\epsilon_i}, \qquad E = \sum_{i=1}^{n} \epsilon_i p_i = e^{-1-\lambda} \sum_{i=1}^{n} \epsilon_i e^{-\mu\epsilon}.$$

In general, we can not solve λ and μ from above two equations. However, we can get one standard alone equation for μ by considering the ratio of these two equations, i.e.,

$$E = \frac{\sum_{i=1}^{n} \epsilon_i e^{-\mu\epsilon_i}}{\sum_{i=1}^{n} e^{-\mu\epsilon_i}}.$$

Then μ can be obtained from other ways. In physics, $\mu = 1/kT$ where k is the *Boltzmann constant* and T is the absolute temperature. □

Significance of Lagrange Multipliers: The Lagrange Multipliers measure how the optimal value depends on the constant on the constraint. This also intimately related to the necessary conditions with inequality constraints: the corresponding Lagrange multiplier can only have one sign for a point to be minimizer or maximizer.

Example 1.5. Consider the problem

$$(P_0) \qquad\qquad \begin{aligned} &\text{min} &&f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x^2 \\ &\text{subject to} &&x_1 + x_2 = 2. \end{aligned}$$

The Lagrange function is

$$L(x,\lambda) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - \lambda(x_1 + x_2 - 2)$$

and the minimizer satisfies

$$\nabla_x L(x,\lambda) = \begin{pmatrix} x_1 - \lambda \\ x_2 - \lambda \end{pmatrix} = 0$$

or $x_1^* = x_2^* = \lambda^* = 1$. The minimal value if $f(x^*) = 1$.

Now consider the perturbed problem

$$(P_\delta) \qquad\qquad \begin{aligned} &\text{min} &&f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x^2 \\ &\text{subject to} &&x_1 + x_2 = 2 + \delta. \end{aligned}$$

We want to know how the optimal value $f(x_\delta^*)$ depends on δ ($|\delta| \ll 1$). We can find that $x_\delta^* = (1 + \delta/2, 1 + \delta/2)$ and

$$f(x_\delta^*) = (1 + \delta/2)^2 = 1 + \delta + \delta^2/4 = f(x^*) + \lambda^*\delta + O(\delta^2).$$

This relation can also be expressed as

$$\left.\frac{d}{d\delta} f(x_\delta^*)\right|_{\delta=0} = \lambda^* = 1. \qquad\qquad □$$

The last relation holds for optimization with more constraints, and the Lagrange multipliers tell how sensitively the optimal value depends on the corresponding constraint. For inactive constraint, the corresponding Lagrange multiplier is zero. We can change that constraint a little, without leading to any change in the optimizer and optimal value (that's another reason we call it *inactive*).

General solution of linear equations: Let A be a $n \times n$ matrix, then the solution of the equation $Ax = b$ for $b \in \mathbb{R}^n$ depends the properties of A. If A is non-singular $(\det(A) \neq 0$ or A has full rank), then there is a unique solution, given by $x = A^{-1}b$. Otherwise, there may be no solution or there are infinity many solutions of the form $x = \bar{x} + Zv$, with $A\bar{x} = b$ and $AZ = 0$.

If A is a $m \times n$ matrix with $(m < n)$, similarly for the under-determined equation $Ax = b$ there is no solution or infinity many solutions of the form $x = \bar{x} + Zv$, with $A\bar{x} = b$ and $AZ = 0$.

Example 1.6. Find the general solution of the equation

$$x_1 + x_2 + x_3 = 1.$$

The general solution is $\bar{x} = (1, 0, 0)$. The columns of Z are in the null space of the constraint, or all the points $x = (x_1, x_2, x_3)$ such that $x_1 + x_2 + x_3 = 0$.

There is one equation but three unknowns, implying Z has two columns (the general solution has two degree of freedom). One way is that we can write the solution in term of those free variables, and then change them into those free variables. More precisely, taking x_2 and x_3 as the free variables,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_2 - x_3 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

The columns of Z are exactly the vectors on the right hand side.

$$Z = \begin{pmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

$\square$

Example 1.7. Find the general solution of the equation

$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 + x_2 + 2x_3 = 2. \end{cases}$$

The particular solution $\bar{x}$ can be found by inspection or by assuming some of the variables to be zero. For this example, you CAN NOT assume $\bar{x}_3$ to be zero, otherwise there is no solution. Assuming $\bar{x}_2 = 0$, we can get the unique solution $\bar{x} = (0, 0, 1)$. There are three variables and two equations with independent rows, implying that there is one degree of freedom for the solution of the null space

$$\begin{cases} x_1 + x_2 + x_3 = 0 \\ x_1 + x_2 + 2x_3 = 0. \end{cases} \tag{3}$$

For the same reason, we can not take x_3 to be the free parameter. Taking x_2 be the free parameter, we need to solve x_1 and x_3 from the system

$$\begin{cases} x_1 + x_3 = -x_2 \\ x_1 + 2x_3 = -x_2. \end{cases}$$

The solution is given by $x_1 = -x_2$ and $x_3 = 0$. Therefore the solution for the homogeneous equation (3) is given by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_2 \\ 0 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Therefore $Z = (-1, 1, 0)^t$.

$\square$

Example 1.8. Find the general solution of the following linear system of equations

$$x_1 + x_2 + x_3 = 1,$$
$$x_2 - 2x_3 = -1,$$
$$x_1 + 3x_3 = 2.$$

□

2 Optimality conditions

2.1 Linear Equality constraints:

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{subject to} \quad Ax = b,$$

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

There are two basic ways to solve it:

Lagrange multiplier: Introducing the Lagrange function

$$L(x, \lambda) = f(x) - \lambda^t (Ax - b), \qquad \lambda \in \mathbb{R}^m.$$

The minimizer x^* satisfies the equation $\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - A^t \lambda^*$. This condition together with the constraint $Ax^* = b$, gives $m + n$ equations for $m + n$ unknowns, x^* and λ^*.

Reduction of variables: Write the general solution $x = \bar{x} + Zv$ and define $\phi(v) = f(\bar{x} + Zv)$. Then the original problem is equivalent to the **unconstrained problem**:

$$\min_{v \in \mathbb{R}^r} \quad \phi(v)$$

or

$$0 = \varphi(v^*) = Z^t \nabla f(\bar{x} + Zv^*).$$

Remember that if the rows in A are linear dependent, there may be no solution (inconsistent) or the size of v is larger than $n - m$, see Example 1.8.

Example 2.1. Show that for $\phi(v) = f(\bar{z} + Zv)$, we have the gradient and Hessian matrix of ϕ in terms of that of f.

$$\nabla \phi(v) = Z^t \nabla f(\bar{x} + Zv), \qquad \nabla^2 \phi(v) = Z^t \nabla^2 f(\bar{x} + Zv) Z.$$

□

Example 2.2. Show that when A has full row rank, the two first order optimality conditions $Z^t \nabla f(x^*) = 0$ and $\nabla f(x^*) = A^t \lambda^*$ are equivalent.

If $\nabla f(x^*) = A^t \lambda^*$, then

$$Z^t \nabla f(x^*) = Z^t A^t \lambda^* = (AZ)^t \lambda^* = 0$$

since columns of Z are in the null space of A.

If $Z^t \nabla f(x^*) = 0$, but $\nabla f(x^*) \neq A^t \lambda$ for any $\lambda \in \mathbb{R}^m$. Let λ^* be the solution of the least square problem

$$\min_{\lambda \in \mathbb{R}^m} \|\nabla f(x^*) - A^t \lambda\|_2$$

and define $r = \nabla f(x^*) - A^t \lambda^* \neq 0$. Since λ^* satisfies the normal equation

$$A(\nabla f(x^*) - A^t \lambda^*) = 0$$

or $Ar = 0$. This implies that r in the column space of Z and we should have $r^t \nabla f(X^*) = 0$ (since $Z^t \nabla f(x^*) = 0$). On the other hand, from $r^t A^t \lambda^* = (Ar)^t \lambda^* = 0$

$$r^t \nabla f(x^*) = r^t(\nabla f(x^*) - A^t \lambda^*) = \|r\|_2^2 > 0$$

contradicting the previous statement. (This also tells you how to find a direction in the feasible region to decrease the function value, if the optimality condition is violated). □

Theorem 2.1 (Necessary Conditions). *If x^* is a local minimizer, then*

$$Z^t \nabla f(x^*) = 0 \text{ and } Z^t \nabla^2 f(x^*)Z \text{ is positive semidefinite.}$$

Theorem 2.2 (Sufficient Conditions). *If x^* satisfies $Ax^* = b$, $Z^t \nabla f(x^*) = 0$ and $Z^t \nabla^2 f(x^*)Z$ is positive definite, then x^* is a strict local minimizer.*

These conditions can be written as for all $p \in \mathcal{N}(A)$

$$p^t \nabla f(x^*) = 0, \qquad p^t \nabla^2 f(x^*)p \geq 0.$$

If the constraints are redundant, you have to do some preprocessing, to make sure the rows in the constraints are linear independent. See the following example.

Example 2.3. Find the minimizers of the following problem

$$
\begin{aligned}
& \min && f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2 \\
(P) \quad & \text{subject to} && x_1 + x_2 + x_3 = 1 \\
& && x_2 - 2x_3 = -1 \\
& && x_1 + \textcolor{red}{3x_3} = 2.
\end{aligned}
$$

and the second order sufficient is satisfied at the minimizer.

Solution: The constraint is redundant and consistent (the sum of the second and the third equation is the first one). We can solve the following equivalent problem

$$
\begin{aligned}
& \min && f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2 \\
(P') \quad & \text{subject to} && x_1 + x_2 + x_3 = 1 \\
& && x_2 - 2x_3 = -1.
\end{aligned}
$$

The Lagrange function is

$$L(x, \lambda) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \frac{1}{2}x_3^2 - \lambda_1(x_1 + x_2 + x_3 - 1) - \lambda_2(x_2 - 2x_3 + 1)$$

The minimizer satisfies

$$\nabla_x L(x^*, \lambda^*) = \begin{pmatrix} x_1^* - \lambda_1^* \\ x_2^* - \lambda_1^* - \lambda_2^* \\ x_3^* - \lambda_1^* + 2\lambda_2^* \end{pmatrix} = 0$$

or $x_1^* = \lambda_1^*, x_2^* = \lambda_1^* + \lambda_2^*, x_3^* = \lambda_1^* - 2\lambda_2^*$. Substituting it into the constraint, we have

$$1 = x_1^* + x_2^* + x_3^* = 3\lambda_1^* - \lambda_2^*$$
$$-1 = x_2^* - 2x_3^* = -\lambda_1^* + 5\lambda_2^* \tag{4}$$

The solution of this system is given by $\lambda_1^* = 2/7$ and $\lambda_2^* = -1/7$. Finally the minimizer is $x^* = (2/7, 1/7, 4/7)$.

The necessary conditions and sufficient condition should also be applied to the reduced system. Write the solution $\bar{x} = x^* + Zv$, we should have the equivalent first order condition $Z^t \nabla f(x^*) = 0$. If the columns of Z are linear independent, then $Z^t \nabla^2 f(x^*) Z$ must be positive definite because $\nabla^2 f(x)$ is obviously positive definite (It is possible that $Z^t \nabla^2 f(x^*) Z$ without $\nabla^2 f(x^*)$ being positive definite. See the next example).

If you using Lagrange Multiplier for the original problem (P), there is one parameter family of solution. The minimizer is unique but the Lagrange multiplier is not. □

Example 2.4. Find the minimizer of the following problem

$$\begin{aligned} \min & \quad f(x) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 \\ \text{subject to} & \quad x_1 - x_2 + 2x_3 = 2. \end{aligned}$$

and show that the second order sufficient condition is satisfied at this point. Is $\bar{x} = (2, 0, 0)$ is a minimizer? If not, find a direction such that f decrease from this point.

Solution: The Lagrange function is

$$L(x, \lambda) = x_1^2 - 2x_1 + x_2^2 - x_3^2 + 4x_3 - \lambda(x_1 - x_2 + 2x_3 - 2).$$

The minimizer satisfy the condition

$$\nabla_x L(x^*, \lambda^*) = \begin{pmatrix} 2x_1^* - 2 - \lambda^* \\ 2x_2^* + \lambda^* \\ -2x_3^* + 4 - 2\lambda^* \end{pmatrix} = 0$$

or $x^* = (1 + \lambda^*/2, -\lambda^*/2, 2 - \lambda^*)$. Substituting it into the constraint

$$2 = x_1^* - x_2^* + 2x_3^* = 5 - \lambda^*$$

or $\lambda^* = 3$ and $x^* = (5/2, -3/2, -1)$. The general solution can be written as $x = x^* + Zv$ where

$$Z = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

It is easy to see that

$$\nabla f(x^*) = \begin{pmatrix} 3 \\ -3 \\ 6 \end{pmatrix}, \qquad \nabla^2 f(x^*) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Since

$$Z^t \nabla f(x^*) = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ -3 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

the first order condition is satisfies. Since

$$Z^t \nabla^2 f(x^*) Z = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 2 & 0 \\ 4 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix}$$

is positive definite, the second order sufficient condition is satisfied too; x^* is a local minimizer.

At $\bar{x} = (2, 0, 0)$, since $\nabla f(\bar{x}) = (2, 0, 4)^t$ CAN NOT be written as a linear combination of columns of Z (no solution to $\nabla f(\bar{x}) = A^t \lambda$ for any λ). From the discussion in Example 2.2, we can find λ to the problem

$$A(\nabla f(\bar{x}) - A^t \lambda) = A \nabla f(\bar{x}) - AA^t \lambda = 0,$$

where $A = (1, -1, 2)$. Since

$$AA^t = 6, \qquad A \nabla f(\bar{x}) = 10,$$

we get $\lambda = 5/3$. The decreasing direction at $\bar{x}$ is then given by

$$d = A^t \lambda - \nabla f(\bar{x}) = \begin{pmatrix} 5/3 \\ -5/3 \\ 10/3 \end{pmatrix} - \begin{pmatrix} 2 \\ 0 \\ 4 \end{pmatrix} = \begin{pmatrix} -1/3 \\ -5/3 \\ -2/3 \end{pmatrix}.$$

To see this is indeed the decreasing direction, define

$$\varphi(t) = f(\bar{x} + td) = f(2 + t/3, -5t/3, 2t/3) = \frac{22}{9} t^2 - 2t.$$

The fact that $\varphi'(0) = -2 < 0$ implies that as x moves from $\bar{x}$ in the direction d, the function value decreases.

Another way to find the decreasing direction comes from the general solution $x = \bar{x} + Zv$ and the reduction to $\phi(v) = f(\bar{x} + Zv)$. Since

$$\nabla_v \phi(0) = Z^t \nabla f(\bar{x}) = -d' \neq 0$$

We can choose $v = d'$, which is also a decreasing direction. $\qquad\qquad \Box$

2.2 Linear Inequality constraints

For inequality constraints (linear or nonlinear), the most import thing is the **sign of the Lagrange Multiplier**, which tells you how the function value changes if the constant in the constraint changes.
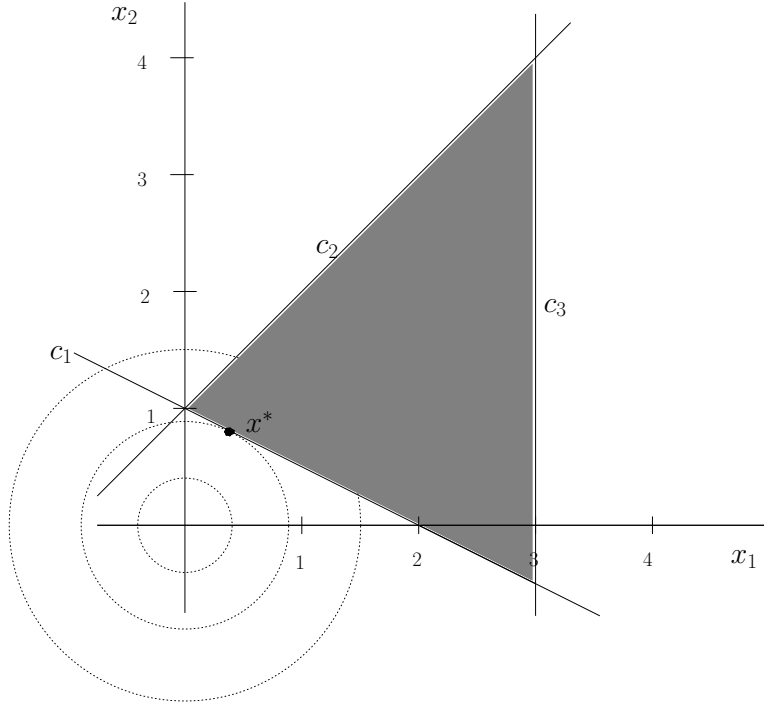
Example 2.5. Consider the following problem

$$\min \qquad f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$$

$$\text{subject to} \quad x_1 + 2x_2 \geq 2 \qquad\qquad\qquad\qquad (c_1)$$

$$x_1 - x_2 \geq -1 \qquad\qquad\qquad\qquad (c_2)$$

$$- x_1 \geq -3. \qquad\qquad\qquad\qquad (c_3)$$

(a) What's the *active set*, according to the graph?

(b) Solve the problem subject inequality constraint(s) in the active set, using Lagrange Multiplier.

(c) If x is close to x^*, but still in the feasible region, how does the objective function change?

(d) How about the Lagrange Multipliers with the rest *inactive set*?

`Solution`: (a) $\mathcal{A}(x^*) = \{1\}$

(b) We are solving the problem

$$\min \qquad f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2$$

$$\text{subject to} \qquad x_1 + 2x_2 = 2.$$

The Lagrange function $L(x, \lambda) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - \lambda(x_1 + 2x_2 - 2)$. The solution is governed by

$$\nabla_x L(x^*, \lambda^*) = \begin{pmatrix} x_1^* - \lambda^* \\ x_2^* - 2\lambda^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

or $x_1^* = \lambda, x_2^* = 2\lambda^*$. Substituting them back into the constraint, we have

$$2 = x_1^* + 2x_2^* = 5\lambda^*$$

9

or $\lambda^* = 2/5$ and $x^* = (2/5, 4/5)$.

(c) For x close to the x^*, we have from Taylor expansion

$$f(x) = f(x^*) + \nabla f(x^*)^t (x - x^*) + \cdots$$

Especially, when the only active constraint c_1 is changed to $x_1 + 2x_2 \geq 2 + \delta$ (for δ small) then we have the new optimal function value is $f(x_\delta^*) = f(x^*) + \lambda^* \delta + \cdots$.

(d) The Lagrange Multipliers for the inactive constraints are always zero, i.e., $\lambda_2^* = \lambda_3^* = 0$. Therefore, in the equation

$$0 = \nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_i \lambda_i^* \nabla c_i(x^*),$$

it does not matter whether the indices i chosen from all the constraints $\mathcal{E} \bigcap \mathcal{I}$ or just the active constraints $\mathcal{A}(x^*)$. For simplicity, we often choose the latter.

$\square$

Example 2.6. The same previous problem, what if we erroneously guessed only the third constraint $-x_1 \geq -3$ was active?

(i) Find the minimizer $\tilde{x}^*$ under this (wrong) active constraint.

(ii) Find a feasible direction at $\tilde{x}^*$ along which the objective function is decreasing.

Solution: (i) If c_3 is the only active constraint, the optimization problem is equivalent to

$$\begin{aligned}
\min \qquad & f(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \\
\text{subject to} \qquad & -x_1 = -3.
\end{aligned}$$

The Lagrange function is $L(x, \lambda) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - \lambda(3 - x_1)$. The solution is given by

$$\nabla_x L(x^*, \lambda^*) = \begin{pmatrix} x_1^* + \lambda^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

or $x^* = (3, 0)$ and $\lambda^* = -3$. Since the Lagrange Multiplier is negative, this point should not be a local minimizer of the original problem.

(ii) The feasible decreasing direction is just $d = \nabla c_3(x^*) = (-1, 0)$. When the active constraint is changed to $-x_1 \geq -3 + \delta$ for δ small enough, it is still active and gives the optimal function value $f(x^*) + \lambda^* \delta + \cdots < f(x^*)$. This is obvious from explicit calculation of this problem.

$\square$

Theorem 2.3 (Necessary condition for linear ineq constr). *If x^* is a local minimizer of f over the set $\{x : Ax \geq b\}$, then for some vector λ^* of Lagrange multipliers,*

- $Ax^* \geq b$ (x^* is feasible)

- $\lambda^{*t}(Ax^* - b) = 0, \lambda^* \geq 0$ (complementarity)

- $\nabla f(x^*) = A^T \lambda^*$ (or $Z^t \nabla f(x^*) = 0$)   (first order condition)

- $Z^t \nabla^2 f(x^*) Z$ is nonnegative definite.   (second order condition)

Remark. Here columns of Z are a basis of the null spaces of the active constraints of $Ax \geq b$, i.e., the submatrix $\hat{A}$, such that $\hat{A}x^* = \hat{b}$.

Remark. We actually have to require that **rows of A are linear independent** (Linear Independence Constraint Qualification induced later), but it is easy to fixed there. We either get redundant constraints or inconsistent constraints (infeasible).

Example 2.7. Show that $x^* = (2/5, 4/5)$ is a strict local minimizer in Example 2.5.

Solution: It is easy to see that x^* is feasible and the complementarity condition $\lambda^{*t}(Ax - b) = 0$ is satisfied and $\lambda^* = (2/5, 0, 0) \geq 0$. The condition $\nabla f(x^*) = A^t \lambda^*$ holds too. The only thing we have to check is the second order sufficient condition. Columns of Z are a basis of the null space of the active constraint, i.e., all vectors $d = (d_1, d_2)$ such that $d_1 + 2d_2 = 0$. Therefore Z has just one column, and can be chosen as $Z = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$. Since $\nabla^2 f(x^*) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, we have

$$Z^t \nabla^2 f(x^*) Z = \begin{pmatrix} 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = 5 > 0,$$

and x^* is a strict minimizer.

$\square$

Example 2.8 (Strict complementarity)**.** Consider the problem

$$\begin{aligned} \min \quad & f(x) = x_1^3 + x_2^2 \\ \text{subject to} \quad & -1 \leq x_1 \leq 0. \end{aligned}$$

Show that $x^* = (0, 0)$ is NOT a local minimizer, but it satisfies all the necessary conditions **except strict complementarity**.

$\square$

Theorem 2.4 (Sufficient Condition 1)**.** *If x^* satisfies all the necessary condition and additionally*

- *Strict complementarity holds*

- *$Z^t \nabla^2 f(x^*) Z$ is positive definite,*

then x^* is a strict local minimizer for the problem

$$\min f(x) \quad \text{subject to} \quad Ax \geq b.$$

If the strict complementary does not hold, one sufficient condition is to test the reduced Hessian $Z^* \nabla^2 f(x^*) Z$, where Z DOES NOT have to be in the null space of those non-strict complementarity constraints.

Theorem 2.5 (Sufficient Condition 2)**.** *Let $\hat{A}_+$ be the submatrix of $\hat{A}$ corresponding to the non-degenerate active constraints at x^* (those constraints whose Lagrange Multiplier are positive). Let Z_+ be a basis matrix for the null space of $\hat{A}_+$. If x^* satisfies*

- *$Z_+^t \nabla^2 f(x^*) Z_+$ is positive definite,*

then x^* is a strict local minimizer for the problem

$$\min f(x) \quad \text{subject to} \quad Ax \geq b.$$

Example 2.9. Redo Example 2.8. Show that the sufficient condition in Theorem 2.5 does not hold at $x^* = (0, 0)$.

Solution: Since the only active constraint $-x_1 \geq 0$ does not satisfy the strict complementarity, there is no constraint on Z or $Z = I$, the identity matrix in $\mathbb{R}^2$. Since

$$Z\nabla^2 f(x^*)Z = \nabla^2 f(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

is only positive semidefinite (not positive definite), the sufficient condition does not hold.

$\square$

If there are mixed equality and inequality constraint, we only need to check the sign of λ_i for $i \in \mathcal{I}$ but not for $i \in \mathcal{E}$.

Example 2.10. Solve the following problem

$$
\begin{aligned}
\min \quad & f(x) = x_1^3 - x_2^3 - 2x_1^2 - x_1 + x_2 \\
\text{subject to} \quad & -x_1 - 2x_2 = -2, \\
& x_1 \geq 0, \\
& x_2 \geq 0.
\end{aligned}
$$

2.3 Nonlinear Constraints

With the presence of nonlinear constraints, there are two more things should be taken care of:

(a) Linear Independence Constraint Qualification

(b) Second order condition is changed to $Z^t \nabla L(x^*, \lambda^*)Z$ instead of $Z^t \nabla^2 f(x^*)Z$.

Linear Independence constraint qualification (LICQ) holds at a point x if the set of active constraint gradients $\{\nabla c_i(x), i \in \mathcal{A}(x)\}$ is linearly independent.

This requires some understanding of the description of the feasible region Ω, given by the constraints. In general, the same feasible can be described in different ways. In general, we want c_i to be smooth and $\nabla c_i(x) \neq 0$.

Example 2.11. For the same region $\Omega = [0, \infty)$, which of the following is a good algebraic description of Ω?

(a) $\{x \mid x^3 \geq 0\}$

(b) $\{x \mid e^x \geq 1\}$

(c) $\{x \mid x^{1/3} \geq 0\}$

$\square$

To understand the constraint qualification, we have to introduce the tangent cone $T_\Omega(x^*)$ and the linearized feasible direction $\mathcal{F}(x^*)$.

The vector d is a **tangent** (or **tangent vector**) to Ω at a point x if there are a feasible sequence $\{z_k\}$ approaching x and a sequence of positive scalars $\{t_k\}$ with $t_k \to 0$ such that

$$\lim_{k \to \infty} \frac{z_k - x}{t_k} = d$$

The set of all tangents to Ω at x^* is called the **tangent cone** and is denoted by $T_\Omega(x^*)$.

The set of **linearized feasible direction** $\mathcal{F}(x)$ at a feasible point x

$$\mathcal{F}(x) = \left\{ d \mid \begin{array}{l} d^t \nabla c_i(x) = 0, \ \text{for all } i \in \mathcal{E} \\ d^t \nabla c_i(x) \geq 0, \ \text{for all } i \in \mathcal{A}(x) \bigcap \mathcal{I} \end{array} \right\}$$

In general, among all the different ways, we should write our constraints c_i such that $\mathcal{F}(x^*)$ (defined purely from the algebraic expressions c_is) agrees with $T_\Omega(x^*)$ (related only to the unique geometry of Ω).

Example 2.12. Find the tangent cone $T_\Omega(x^*)$ and the linearized feasible direction $\mathcal{F}(x^*)$ for the feasible regions defined below at x^*. Are they the same for both case?

- $\{x \mid x_1^2 + x_2^2 - 2 = 0\}, x^* = (-\sqrt{2}, 0)$
- $\{x \mid x_1^2 + x_2^2 - 2 \leq 0\}, x^* = (-\sqrt{2}, 0)$

$\square$

Example 2.13. Show that the tangent cone $T_\Omega(x^*)$ and the linearized feasible direction $\mathcal{F}(x^*)$ are different at $x^* = (0,0)$ for the feasible region defined by

$$\Omega = \{x \mid 1 - x_1^2 - (x_2 - 1)^2 \geq 0, \quad -x_2 \geq 0\}.$$

$\square$

The first order necessary conditions for the general constraint optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & c_i(x) = 0, \quad i \in \mathcal{E}, \\ & c_i(x) \geq 0, \quad i \in \mathcal{I}. \end{array}$$

are exact the same as that with linear constraints, except with the addition LICQ condition.

Theorem 2.6 (First-order necessary (KKT) conditions). *Suppose x^* is a local solution, f and c_i are continuously differentiable and the **LICQ** holds at x^*. Then there exist a Lagrange multiplier λ^*, $i \in \mathcal{E} \bigcup \mathcal{I}$, such that*

(1) $c_i(x^*) = 0$, for all $i \in \mathcal{E}$ (Feasible condition for equality constraints)

(2) $c_i(x^*) \geq 0$, for all $i \in \mathcal{I}$ (Feasible condition for inequality constraints)

(3) $\lambda_i^* \geq 0$, for all $i \in \mathcal{I}$

(4) $\lambda_i^* c_i(x^*) = 0$, for all $i \in \mathcal{E} \bigcup \mathcal{I}$ (Complementarity)

(5) $\nabla_x L(x^*, \lambda^*) = 0$

To find the second order conditions, we need the critical cone

$$\mathcal{C}(x^*, \lambda^*) = \{w \in \mathcal{F}(x^*) \mid \nabla c_i(x^*)^t w = 0, \ \forall \ i \in \mathcal{A}(x^*) \bigcap \mathcal{I} \text{ with } \lambda_i^* > 0\}$$

$$= \{w \mid \nabla c_i(x^*)^t w = 0 \text{ for } i \in \mathcal{E} \text{ or } i \in \mathcal{A}(x^*) \bigcap \mathcal{I} \text{ with } \lambda_i^* > 0\}.$$

Theorem 2.7 (Second-order necessary conditions).

$$w^T \nabla_{xx} L(x^*, \lambda^*) w \geq 0, \qquad \forall\, w \in \mathcal{C}(x^*, \lambda^*).$$

Theorem 2.8 (Second-order sufficient conditions).

$$w^T \nabla_{xx} L(x^*, \lambda^*) w > 0, \qquad \forall\, w \in \mathcal{C}(x^*, \lambda^*),\ w \neq 0.$$

Example 2.14. Find the solution of the following problem

$$\begin{aligned} \min \qquad & f(x) = x_1 + x_2 \\ \text{subject to} \qquad & c_1(x) = x_1^2 + x_2^2 - 2 = 0. \end{aligned}$$

`Solution`: The Lagrange function is

$$L(x, \lambda) = x_1 + x_2 - \lambda(x_1^2 + x_2^2 - 2)$$

and the solution satisfies

$$\nabla_x L(x^*, \lambda^*) = \begin{pmatrix} 1 - 2\lambda^* x_1^* \\ 1 - 2\lambda^* x_2^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

or $x^* = \left(\dfrac{1}{2\lambda}, \dfrac{1}{2\lambda} \right)$. Substituting it into the constraint

$$2 = x_1^{*2} + x_2^{*2} = \frac{1}{2\lambda^{*2}}$$

or $\lambda^* = \pm\dfrac{1}{2}$. We can use the second order optimality condition to choose the right λ^* for a minimizer. Since $\nabla c_1(x^*) = (2x_1^*, 2x_2^*)^t = (1/\lambda^*, 1/\lambda^*)$, for any vector $w = (d_1, d_2)^t$ in the critical cone $\mathcal{C}(x^*, \lambda^*)$, we have $\nabla c_1(x^*)^t d = 0$ or $d_1 + d_2 = 0$ or $w = (d, -d)$ for $d \in \mathbb{R}$. Moreover, since $\nabla_x^2 L(x^*, \lambda^*) = \begin{pmatrix} -2\lambda^* & 0 \\ 0 & -2\lambda^* \end{pmatrix}$, x^* is a minimizer if and only if $w^t \nabla_x^2 L(x^*, \lambda^*) w = -4\lambda^* d^2 > 0$ for $d \neq 0$. Therefore, we must choose $\lambda^* = -1/2$ and $x^* = (-1, -1)$. $\qquad\square$

Example 2.15. Find the solution of the following problem

$$\begin{aligned} \min \qquad & f(x) = x_1 + x_2 \\ \text{subject to} \qquad & c_1(x) = 2 - x_1^2 - x_2^2 - 2 \geq 0. \end{aligned}$$

In this example, λ^* is determined from the first order necessary condition for minimizers (which is compatible with the second order condition). In the previous example, the sign of λ^* does not matter because the constraint is an equality.

Example 2.16. Consider the problem

$$\begin{aligned} \min \qquad & f(x) = x_1 \\ \text{subject to} \qquad & (x_1 + 1)^2 + x_2^2 \geq 1 \\ & x_1^2 + x_2^2 \leq 2. \end{aligned}$$

(a) Use graphic method to find the global minimizer and verify the optimality conditions are satisfied there.

(b) Is $(0,0)$ a local minimizer? If not, which optimality condition is violated?

Example 2.17 (Counterexample for LICQ). Consider the following problem

$$
\begin{aligned}
\min \quad & f(x) = 3x_1 + 4x_2 \\
\text{subject to} \quad & (x_1 + 1)^2 + x_2^2 = 1 \\
& (x_1 - 1)^2 + x_2^2 = 1
\end{aligned}
$$

(i) Find the feasible region and the minimizer.

(ii) Can you find the Lagrange Multiplier λ^*?

3 Duality

Lagrange Multipliers play an important role in constrained optimization. Here are some facts:

(a) Complementarity: For inactive constraints $c_i(x^*) \geq 0$, the corresponding Lagrange Multiplier $\lambda_i^* = 0$.

(b) If the constraint $c_i(x) \geq 0$ (or $c_i(x) = 0$) is perturbed to $c_i(x) \geq \delta$ (or $c_i(x) = \delta$) then the Lagrange function evaluated at the optimal x_δ^* and λ_δ^* has the relation

$$
\frac{d}{d\delta} L_\delta(x_\delta^*, \lambda_\delta^*) \bigg|_{\delta=0} = \lambda_i^*
$$

or

$$
L_\delta(x_\delta^*, \lambda_\delta^*) = L(x^*, \lambda^*) + \lambda_i^* \delta + O(\delta^2)
$$

(c) The inequality constraint c_i is *strongly active* (or *Strict Complementarity*) if $i \in \mathcal{A}(x^*)$ and $\lambda_i^* > 0$. It is *weakly active* if $i \in \mathcal{A}(x^*)$ and $\lambda_i^* = 0$.

The Lagrange Multipliers for constraints can be regarded as *dual* variable for an associated problem, the so-called *dual problem*.

3.1 Min-Max Duality

Most of the duality-related problems can be formulated by min-max and max-min.

Example 3.1 (Two-person Zero-sum Game represented as matrix). . Let the payoff that A chooses strategy A_i and B chooses strategy B_j be a_{ij}, represented as

	B chooses B1	B chooses B2	B chooses B3
A chooses A1	+3	-2	+2
A chooses A2	-1	0	+4
A chooses A3	-4	-3	+1

The goal of A is to maximize the payoff (and B to minimize A's payoff). But the strategy depends on who plays first. If A choose strategy A_i, then B chooses the minimal payoff, i.e $\min_j a_{ij}$ and we have that the payoff A gets for choosing A_i is

$$
\min_j a_{1j} = -2, \qquad \min_j a_{2j} = -1, \qquad \min_j a_{3j} = -4.
$$

Then A must choose A_2 to maximize it (and B chooses B_1).

On the other hand, if B plays first and choose B_j, then A chooses the maximum payoff $\max_{ij} a_{ij}$, or

$$\max_i \ a_{i1} = 3, \qquad \max_i \ a_{i2} = 0, \qquad \max_i \ a_{i2} = 4$$

respectively. Then B's optimal choice is B_2 (and A chooses A_2). In this case, we have

$$\max_i \min_j \ a_{ij} = -1 < \min_j \max_i \ a_{ij} = 0.$$

Similarly, for general function $f(x, y)$, we always have the **weak duality**:

$$\max_{y \in Y} \min_{x \in X} \ f(x, y) \le \min_{x \in X} \max_{y \in Y} \ f(x, y).$$

Under certain conditions, we have equality sign:

Theorem 3.1 (Strong duality). *The condition*

$$\max_{y \in Y} \min_{x \in X} \ f(x, y) = \min_{x \in X} \max_{y \in Y} \ f(x, y)$$

holds if and only if there exists a pair (x^*, y^*) that satisfies the saddle-point condition

$$f(x^*, y) \le f(x^*, y^*) \le f(x, y^*)$$

for all $x \in X$ and $y \in y$.

3.2 Lagrangian Duality

The Lagrangian duality for the problem

$$(P) \qquad\qquad \min_{x \in \mathbb{R}^n} f(x) \qquad \text{subject to } c(x) \ge 0,$$

uses the previous min-max techniques to the Lagrange function $L(x, \lambda) = f(x) - \lambda^t c(x)$. The key idea is to **get rid of any constraints on** x, by introducing the Lagrange multipliers λ, such that those constraints are transferred on λ. This can be accomplished by the following fact

$$\max_{\lambda \ge 0} \ -\lambda^t c(x) = \begin{cases} 0, & \text{if } c(x) \ge 0, \\ \infty, & \text{otherwise.} \end{cases}$$

Therefore, if the above problem (P) is feasible, then it is equivalent to the "unconstrained problem" $\min_x L^*(x)$ where

$$L^*(x) = \max_{\lambda \ge 0} L(x, \lambda) = \begin{cases} f(x), & \text{if } g(x) \ge 0, \\ \infty, & \text{otherwise.} \end{cases}$$

The function L^* coincides with f on the feasible region, but becomes $+\infty$ outside it.

The dual problem can be obtained by exchanging the order of min and max. Define the **dual objective function** q as

$$q(\lambda) \stackrel{\text{def}}{=} \inf_x L(x, \lambda)$$

The **dual problem** is:

$$\max_{\lambda \in \mathbb{R}^m} q(\lambda) \qquad \text{subject to } \lambda \ge 0.$$

Example 3.2. Find the dual problem for

$$\min_{x \in \mathbb{R}^2} \frac{1}{2}(x_1^2 + x_2^2) \qquad \text{subject to } x_1 - 1 \geq 0.$$

`Solution`: We have $L(x, \lambda) = \frac{1}{2}(x_1^2 + x_2^2) - \lambda(x_1 - 1)$ and

$$q(\lambda) = \min_x L(x, \lambda)$$

For fixed λ, the minimizer is given by

$$\nabla_x L(x, \lambda) = \begin{pmatrix} x_1 - \lambda \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Therefore, $x_1^* = \lambda, x_2^* = 0$ and

$$q(\lambda) = L((\lambda, 0), \lambda) = \lambda - \frac{1}{2}\lambda^2.$$

The dual problem is

$$\max_{\lambda \geq 0} q(\lambda) = \max_{\lambda \geq 0} \ \lambda - \frac{1}{2}\lambda^2.$$

The maximizer is $\lambda^* = 1 > 0$. It is easy to check that in this case we have

$$\min_x \max_{\lambda \geq 0} L(x, \lambda) = \max_{\lambda \geq 0} \min_x L(x, \lambda) = 1/2. \qquad \square$$

The dual problem can be different, depending on the type of constraints.

Example 3.3. Find the dual problem and solve it for the following optimization:

(a) $\min_x \dfrac{1}{2}(x_1^2 + x_2^2) \qquad$ subject to $\quad x_1 + x_2 \geq 2$

(b) $\min_x \dfrac{1}{2}(x_1^2 + x_2^2) \qquad$ subject to $\quad x_1 + x_2 \leq 2$

(c) $\min_x \dfrac{1}{2}(x_1^2 + x_2^2) \qquad$ subject to $\quad x_1 + x_2 = 2$

$\square$

Sometimes, for a given λ, the solution x for the minimization problem $\min_x L(x, \lambda$ can not be written explicitly. In this case, we can write x implicitly as $\nabla_x L(x, \lambda) = 0$.

Example 3.4 (Wolfe Duality). Consider the problem

$$\min f(x) \qquad \text{subject to} \ \ c(x) \geq 0.$$

The dual problem is

$$\max_{\lambda \geq 0} \min_x L(x, \lambda)$$

can be written as

$$\max_{\lambda \geq 0} \ L(x, \lambda) \qquad \text{subject to} \ \ \nabla_x L(x, \lambda) = 0.$$

Because of the equality, we can "maximize" w.r.t x in the objective function L, which gives the **Wolfe duality**

$$
\begin{aligned}
\max_{x,\lambda} \quad & L(x,\lambda) \\
\text{subject to} \quad & \nabla_x L(x,\lambda) = 0 \\
& \lambda \geq 0,
\end{aligned}
$$

under the condition that f is convex. □

Example 3.5. Consider the problem

$$
\min \ f(x) = e^x, \qquad \text{subject to } 1 - x^2 \geq 0.
$$

Formulate it as in Wolfe duality. For this one, you can simplify your constraint and get the exact solution. □

For the problem

$$
\min f(x) \qquad \text{subject to} \ \ c(x) \geq 0,
$$

and the dual problem

$$
\max_{\lambda \geq 0} q(\lambda)
$$

we still have the *weak duality*

$$
q(\lambda) \leq f(x)
$$

for any $\lambda \geq 0$ and x feasible ($c(x) \geq 0$). If fact, we have

$$
\max_{\lambda \geq 0} q(\lambda) \leq \min_{c(x) \geq 0} f(x).
$$

However, the equality may not hold in this case, especially when one problem is infeasible or unbounded and we say that there is a **duality gap**.

Example 3.6 (Duality gap). Consider the problem

$$
\begin{aligned}
\min \quad & f(x) = -x^2 \\
\text{subject to} \quad & x = 1 \\
& x \in X = \{x : 0 \leq x \leq 2\}.
\end{aligned}
$$

What's the optimal solution? Formulate the duality problem as (slightly different because we don't introduce Lagrange Multipliers for the constraint $x \in X$)

$$
\min_{x \in X} \max_{\lambda} \ -x^2 - \lambda(x-1), \quad \max_{\lambda} \min_{x \in X} \ -x^2 - \lambda(x-1).
$$

Find the solution. □

Example 3.7 (Infinity Duality Gap). Consider the problem

$$
\begin{aligned}
\min \quad & f(x) = -x^2 \\
\text{subject to} \quad & 0 \leq x \leq 1.
\end{aligned}
$$

Solve this problem and the dual problem. □

Example 3.8 (Duality of a Linear program)**.** Consider the linear program

$$
\begin{aligned}
\min \qquad & f(x) = c^t x \\
\text{subject to} \qquad & Ax = b \\
& x \geq 0.
\end{aligned}
$$

Solution: Introduce y for the constraint $Ax = b$ and λ for $x \geq 0$, then

$$
L(x, y, \lambda) = c^t x - y^t(Ax - b) - \lambda^t x.
$$

The dual problem is

$$
\begin{aligned}
\max \qquad & L(x, y, \lambda) = c^t x - y^t(Ax - b) - \lambda^t x \\
\text{subject to} \qquad & c - A^t y - \lambda = 0 \\
& \lambda \geq 0.
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
\max \qquad & L(x, y, \lambda) = b^t y \\
\text{subject to} \qquad & A^t y \leq c.
\end{aligned}
$$

$\square$

Example 3.9 (Duality of a quadratic program)**.** Consider the quadratic program

$$
\begin{aligned}
\min \qquad & f(x) = \frac{1}{2} x^t Q x + c^t x \\
\text{subject to} \qquad & Ax \geq b.
\end{aligned}
$$

where Q is a positive definite matrix.
    Solution:

$$
\begin{aligned}
\max_{x,\lambda} \qquad & \frac{1}{2} x^t Q x + c^t x - \lambda^t(Ax - b) \\
\text{subject to} \qquad & Qx + c - A^t \lambda = 0 \\
& \lambda \geq 0.
\end{aligned}
$$

We can eliminate x from the constraint and get the equivalent problem

$$
\begin{aligned}
\max_{x,\lambda} \qquad & -\frac{1}{2} \lambda^t (AQ^{-1}A^t)\lambda + (AQ^{-1}c + b)^t \lambda - \frac{1}{2} c^t Q^{-1} c \\
\text{subject to} \qquad & \lambda \geq 0.
\end{aligned}
$$

$\square$