CONSTRAINED OPTIMIZATION

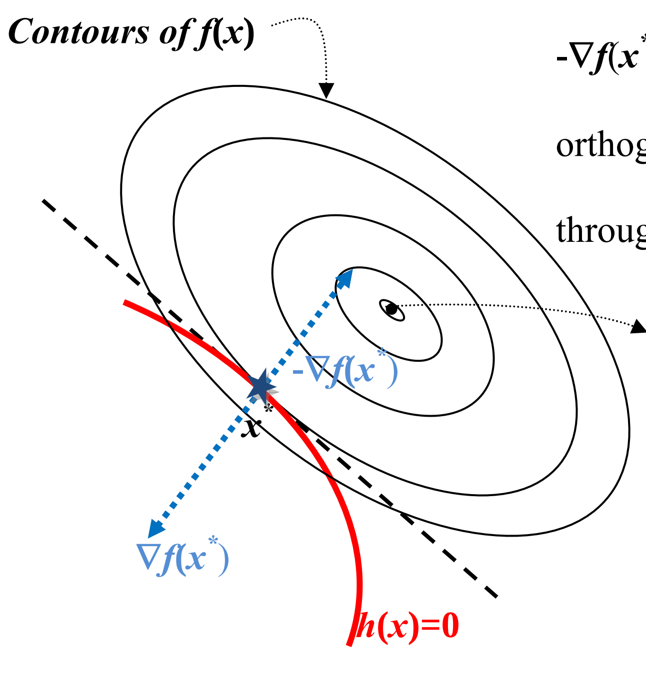
1. EQUALITY CONSTRAINTS

Consider the problem **(P1)**: Minimize $f(\boldsymbol{x})$

st $h_j(\boldsymbol{x}) = 0, \quad j=1,2,\ldots,m$

$\boldsymbol{x} \in \mathbb{R}^n$

Let us first examine the case where m=1 (i.e., a single constraint). Without this constraint the necessary condition for optimality was $\nabla \boldsymbol{f}(\boldsymbol{x})=\mathbf{0}$. With the constraint $h(\boldsymbol{x})=0$, we also require that $\boldsymbol{x}$ lie on the graph of the (nonlinear) equation $h(\boldsymbol{x})=0$.



$-\nabla \boldsymbol{f}(\boldsymbol{x}^*) \equiv$ steepest descent direction, is orthogonal to the tangent of the contour of f through the point $\boldsymbol{x}^*$.

It must also be orthogonal to the tangent of the curve $h(\boldsymbol{x})=0$ at $\boldsymbol{x}^*$!

But we know that $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$ is also orthogonal to the curve $h(\boldsymbol{x})=0$ at $\boldsymbol{x}^*$. Therefore...

$-\nabla \boldsymbol{f}(\boldsymbol{x}^*)$ and $\nabla \boldsymbol{h}(\boldsymbol{x}^*)$ MUST BOTH LIE ALONG THE SAME LINE, i.e., for some $\lambda \in \mathbb{R}$, we must have

$$\boxed{-\nabla \boldsymbol{f}(\boldsymbol{x}^*) = \lambda \nabla \boldsymbol{h}(\boldsymbol{x}^*)}$$

So, if $\boldsymbol{x}^*$ is a minimizer, the necessary condition reduces to

$$\nabla \boldsymbol{f}(\boldsymbol{x}^*) + \lambda \nabla \boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$$

We now proceed to the general case where the above necessary condition should hold for *each* constraint.

DEFINITION: The ***Lagrangian*** function for Problem P1 is defined as

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = f(\boldsymbol{x}) + \Sigma_{j=1\ldots,m}\, \lambda_j\, h_j(\boldsymbol{x})$$

The **KARUSH-KUHN-TUCKER** Conditions

If the point $\boldsymbol{x}^* \in \mathbb{R}^n$ is a minimizer for problem P1, it must satisfy the following necessary condition for some $\lambda^* \in \mathbb{R}^m$:

$$\boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j \boldsymbol{\nabla h_j}(\boldsymbol{x}^*) = 0 \qquad \text{and} \qquad h_j(\boldsymbol{x}^*) = 0 \;\; \forall\, j$$

or equivalently,

$$\nabla_{x,\lambda} \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}.$$

EXAMPLE:

Minimize $\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 3x_2$

st $x_1 + x_2 = 3$ $(\Rightarrow h(\mathbf{x}) = x_1 + x_2 - 3 = 0)$

The Lagrangian is

$L(\mathbf{x},\lambda) = (\frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 3x_2) + \lambda(x_1 + x_2 - 3)$

The **K-K-T** conditions yield

$$\frac{\partial f}{\partial x_1} = x_1 - x_2 + \lambda = 0$$

$$\frac{\partial f}{\partial x_2} = x_2 - x_1 - 3 + \lambda = 0$$

$$\frac{\partial f}{\partial \lambda} = x_1 + x_2 - 3 = 0 \qquad \text{(original constraints)}$$

Solving these (in *this particular* case, linear) equations:

$x_1^* = 0.75$, $x_2^* = 2.25$, $\lambda^* = 1.5$, with $f(\mathbf{x}^*) = -5.625$

The above point thus satisfies the necessary conditions for it to be a minimum.

<u>SUFFICIENT CONDITIONS</u>: IF in addition to the K-K-T necessary conditions, suppose that for arbitrary $\mathbf{z} \in \mathbb{R}^n$ with $\mathbf{z} \neq \mathbf{0}$, the following statement is true:

"$\mathbf{z}^{\mathrm{T}} \nabla \mathbf{h}_j(\mathbf{x}^*) = 0, j=1,2,\ldots,m$ implies that $\mathbf{z}^{\mathrm{T}} \nabla_{\mathbf{x}}^2 \mathbf{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*)\mathbf{z} > 0$"

THEN f has a strict local minimum at $\mathbf{x}^*$, st $h_j(\mathbf{x}^*)=0, j=1,2,\ldots,m$.

For our example,

$$\nabla_x^2 \boldsymbol{L}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \text{ and } \nabla \boldsymbol{h}(\boldsymbol{x}^*) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Consider $\boldsymbol{z} \in \mathbb{R}^2 = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Then $\boldsymbol{z}^T \nabla \boldsymbol{h}_j(\boldsymbol{x}^*) = z_1 + z_2 = 0$ (i.e., $z_2 = -z_1$) implies that

$$\boldsymbol{z}^T \nabla_x^2 \boldsymbol{L}(\boldsymbol{x}^*,\lambda^*)\, \boldsymbol{z} = [z_1 \;\; z_2] \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = (z_1 - z_2)^2 = (2z_1)^2 > 0$$

$\therefore$ $\boldsymbol{x}^*$ is a strict local minimum.

As in unconstrained optimization, in practice, sufficient conditions become quite complicated to verify, and most algorithms only look for points satisfying the necessary conditions.

2. INEQUALITY-CONSTRAINED OPTIMIZATION

Consider the problem (**P2**): Minimize $f(\boldsymbol{x})$

st $g_j(\boldsymbol{x}) \leq 0$, $j=1,2,\ldots,m$

$\boldsymbol{x} \in \mathbb{R}^n$.

Again, let us first consider the case with a single inequality constraint (i.e., m=1).

Q. Can this be converted to an equivalent equality constrained problem?

1. Suppose that (as in LP) we write $g(\boldsymbol{x})\leq 0$ in the form $g(\boldsymbol{x}) + S = 0$, where S is a nonnegative slack variable. Then we have added the constraint $S\geq 0$; effectively, we have not accomplished anything (we just trade one inequality for another)!
2. Instead, let us write $g(\boldsymbol{x}) + S^2 = 0$

   ($S^2\geq 0$ is of course unnecessary…)

The Lagrangian is

$$L(\boldsymbol{x},S,\lambda) = f(\boldsymbol{x}) + \lambda(g(\boldsymbol{x})+S^2)$$

and the **K-K-T** optimality conditions become

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i}(\boldsymbol{x}^*) + \lambda^* \frac{\partial g}{\partial x_i}(\boldsymbol{x}^*) = 0, \qquad i = 1,2,\ldots,n$$

$$\frac{\partial L}{\partial \lambda} = g(\boldsymbol{x}^*) + (S^*)^2 = 0 \qquad \text{(original constraints)}$$

$$\frac{\partial L}{\partial S} = 2\lambda^* S^* = 0 \qquad \text{(COMPLEMENTARY SLACKNESS)}$$

<u>EXAMPLE</u>: Minimize $x_1^2 + ½\, x_2^2 - 8x_1 - 2x_2 - 60$

st $40x_1 + 20x_2 \leq 140 = 0$

Converting to an equality constraint, we have

$$40x_1 + 20x_2 + S^2 – 140 = 0$$

$$L(x_1, x_2, S, \lambda) = (x_1^2 + ½\, x_2^2 - 8x_1 - 2x_2 – 60) + \lambda(40x_1 + 20x_2 + S^2 – 140)$$

The necessary conditions are therefore

$$\frac{\partial L}{\partial x_1} = 2x_1 - 8 + 40\lambda = 0, \quad \frac{\partial L}{\partial x_2} = x_2 - 2 + 20\lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = 40x_1 + 20x_2 + S^2 - 140 = 0 \qquad \text{(original constraints)}$$

$$\frac{\partial L}{\partial S} = 2\lambda S = 0$$

Here we have nonlinear system of equations. Again note that these are only necessary conditions, and there may be one or more points that satisfy these. We seek an optimal solution to the original problem that is among the solution(s) to the necessary conditions.

Rather than attempt a direct solution, let us use the complementary slackness condition to investigate two cases: (i) λ=0, (ii) S=0.

Case (i) $\lambda=0 \Rightarrow \quad 2x_1 - 8 = 0 \Rightarrow x_1 = 4$

$x_2 - 2 = 0 \Rightarrow x_2 = 2$

$40x_1 + 20x_2 + S^2 - 140 = 0 \Rightarrow S^2 = -60$ INFEASIBLE!

Case (ii) $S=0 \Rightarrow \quad 2x_1 - 8 + 40\lambda = 0 \Rightarrow x_1 = 4$

$x_2 - 2 + 20\lambda = 0 \Rightarrow x_2 = 2$

$40x_1 + 20x_2 - 140 = 0 \quad \Rightarrow x_1 = 3,\ x_2 = 1,\ \lambda = 0.05$

$\therefore$ the only solution to the necessary conditions is the point $\boldsymbol{x}$=(3,1). Thus, if the original problem had a solution, then this must be it!

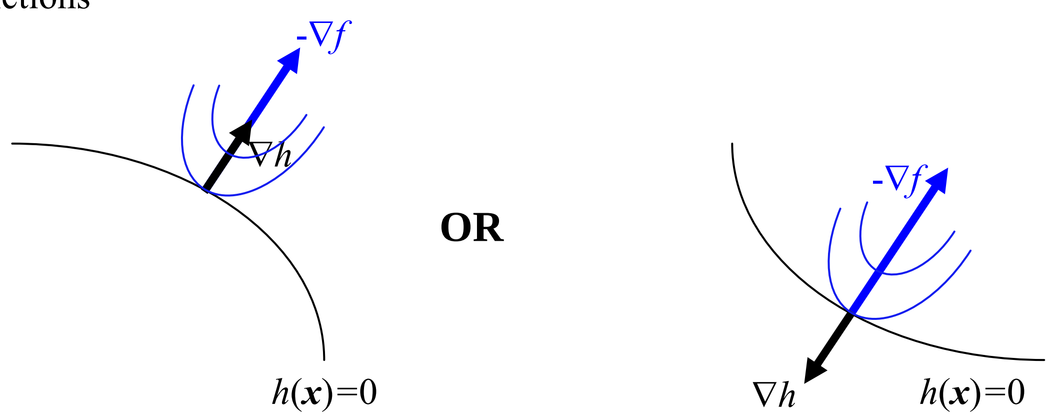
Geometric Interpretation

Consider Minimize $f(\boldsymbol{x})$, st $g(\boldsymbol{x})\leq 0$.

Just as in equality constrained optimization, here too we have one of the necessary conditions as

$$-\nabla \boldsymbol{f}(\boldsymbol{x}^*) = \lambda.\nabla \boldsymbol{g}(\boldsymbol{x}^*) \text{ for some value of } \lambda$$

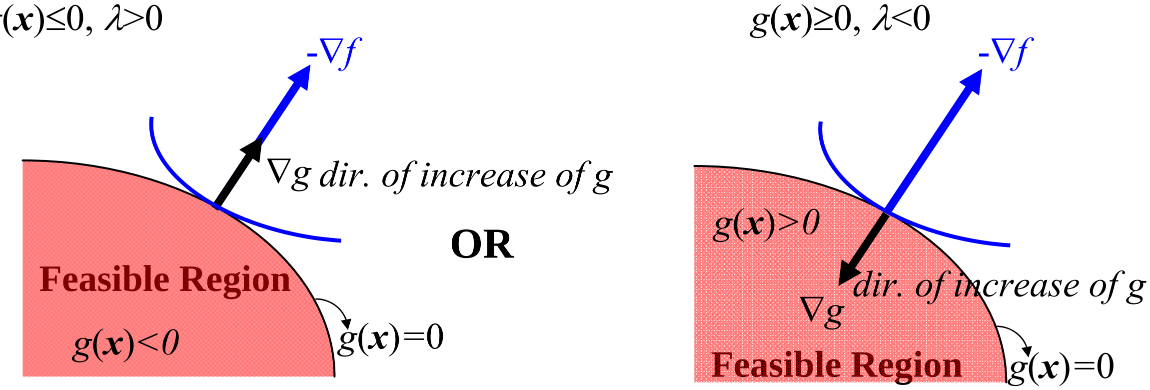
With equality constraints, recall that λ was just some real number with no sign restrictions



($\lambda>0$) ($\lambda<0$)

With inequality constraints we can "predict" the correct sign for λ

$g(\boldsymbol{x})\leq 0$, $\lambda>0$ $g(\boldsymbol{x})\geq 0$, $\lambda<0$



<u>Thus for minimization, $g(\boldsymbol{x})\leq 0 \Rightarrow \lambda\geq 0$; and $g(\boldsymbol{x})\geq 0 \Rightarrow \lambda\leq 0$.</u>

In practice, we dispense with the slack variable S and state the NECESSARY CONDITIONS in a standard form that does not use the slack variable.

Let $L(\boldsymbol{x},\lambda) = f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})$

Then the necessary conditions are:

(1) $$\frac{\partial L}{\partial x_i}(\boldsymbol{x}^*,\lambda^*) = \frac{\partial f}{\partial x_i}(\boldsymbol{x}^*) + \lambda^* \frac{\partial g}{\partial x_i}(\boldsymbol{x}^*) = 0, \qquad \forall i$$

i.e., $\nabla_x \boldsymbol{L}(\boldsymbol{x}^*,\lambda^*) = \nabla \boldsymbol{f}(\boldsymbol{x}^*) + \lambda^* \nabla \boldsymbol{g}(\boldsymbol{x}^*) = 0$

(2) $g(\boldsymbol{x}^*) \leq 0$

(3) $\lambda^* \cdot g(\boldsymbol{x}^*) = 0 \qquad \lambda^* \geq 0$

Condition 2 yields the original constraints.

Condition 3 is complementary slackness: $(g(\boldsymbol{x}^*)<0 \Rightarrow \lambda^* = 0; \quad \lambda^* >0 \Rightarrow g(\boldsymbol{x}^*)=0)$.

For the general case with m inequality constraints (i.e., Problem **P2**) the **K-K-T** necessary conditions are that if $\boldsymbol{x}^*$ is a minimizer then $\exists\ \lambda_j^* \geq 0$ for j=1,2, ... ,m such that for $L(\boldsymbol{x},\lambda)=f(\boldsymbol{x}) + \Sigma_{j=1\ldots m} \lambda_j\, g_j(\boldsymbol{x})$

(1) $$\boldsymbol{\nabla L_x}(\boldsymbol{x}^*,\boldsymbol{\lambda}^*) = \boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \boldsymbol{\nabla g_j}(\boldsymbol{x}^*) = 0 \qquad (n \text{ equations})$$

(2) $g_j(\boldsymbol{x}^*) \leq 0, \quad j=1,2,\ldots, m$

(3) $\{g_j(\boldsymbol{x}^*)\} \cdot \lambda_j^* = 0, \quad j=1,2,\ldots, m$

<u>NOTE</u>: If a constraint were of the form $g(\boldsymbol{x})\geq 0$, we would require that $\lambda_j^* \leq 0$.

Alternatively, we could redefine $g_j(\boldsymbol{x})\geq 0$ as $-g(\boldsymbol{x})\leq 0$ and retain $\lambda_j^* \geq 0$!

We now look at the general NLP problem (**P3**)

Minimize $f(\boldsymbol{x})$

st $g_j(\boldsymbol{x}) \leq 0,\ \ j=1,2,\ldots,m,$

$g_j(\boldsymbol{x}) \geq 0,\ \ j=m+1,\ldots,p,$

$g_j(\boldsymbol{x}) = 0,\ \ j=p+1,\ldots,q,$

$\boldsymbol{x}\in\mathbb{R}^n.$

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{m}\lambda_j g_j(\boldsymbol{x}) - \sum_{j=m+1}^{p}\lambda_j g_j(\boldsymbol{x}) + \sum_{j=p+1}^{q}\mu_j g_j(\boldsymbol{x})$$

<u>Karush-Kuhn-Tucker Necessary Conditions</u>

If $\boldsymbol{x}^*$ minimizes $f(\boldsymbol{x})$ while satisfying the above constraints, then *provided certain regularity conditions are met*, there exist vectors $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ such that

(1) $$\boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m}\lambda_j^*\boldsymbol{\nabla g_j}(\boldsymbol{x}^*) - \sum_{j=m+1}^{p}\lambda_j^*\boldsymbol{\nabla g_j}(\boldsymbol{x}^*) + \sum_{j=p+1}^{q}\mu_j^*\boldsymbol{\nabla g_j}(\boldsymbol{x}^*) = \boldsymbol{0}$$

(2) All original constraints are satisfied

(3) $\lambda_j^*\cdot\{g_j(\boldsymbol{x}^*)\} = 0 \qquad j=1,2,\ldots,p$

(4) $\lambda_j^* \geq 0 \qquad j=1,2,\ldots,p$

(5) μ_j^* are unrestricted in sign for $j=p+1,\ldots, q$.

The regularity conditions referred to above are also known as ***constraint qualifications***. We will address these later…

<u>NON-NEGATIVITY</u>: Many engineering problems require the decision variables to be non-negative. While each non-negativity requirement could be treated as a constraint, we don't have to...

Consider the following problem (for simplicity, assume only ≤ constraints….):

Minimize $f(\boldsymbol{x})$

st $\quad g_j(\boldsymbol{x}) \leq 0; \qquad j=1,2,\ldots,m$

$\quad q_i(\boldsymbol{x}) = x_i \geq 0; \qquad i=1,2,\ldots,n.$

Let us define Lagrange multipliers $\rho_1, \rho_2,\ldots, \rho_n$ corresponding to the non-negativity constraints. Then in the **K-K-T** conditions we have via complementary slackness

$\rho_i^* x_i^* = 0; \qquad i=1,2,\ldots,n$

and Condition (1) becomes

$$\boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \boldsymbol{\nabla g_j}(\boldsymbol{x}^*) - \sum_{i=1}^{n} \rho_i^* \boldsymbol{\nabla q_i}(\boldsymbol{x}^*) = \boldsymbol{0}$$

$$\Rightarrow \quad \boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \boldsymbol{\nabla g_j}(\boldsymbol{x}^*) - \rho_1^* \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \end{bmatrix} - \rho_2^* \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 0 \end{bmatrix} - \ldots - \rho_{n-1}^* \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 1 \\ 0 \end{bmatrix} - \rho_n^* \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{0}$$

$\Rightarrow \quad \boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \boldsymbol{\nabla g_j}(\boldsymbol{x}^*) = \boldsymbol{\rho}^*,$ where $\boldsymbol{\rho}^* \in \mathbb{R}^n = [\rho_1^* \;\; \rho_2^* \;\ldots\; \rho_n^*]^T$

Note that we want $\boldsymbol{\rho}^* \geq \boldsymbol{0}$

Thus the necessary conditions become

(1) $\boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \boldsymbol{\nabla g_j}(\boldsymbol{x}^*) \geq \mathbf{0}$ (n conditions)

(2) All original constraints are satisfied

(3) $\lambda_j^* \cdot \{g_j(\boldsymbol{x}^*)\} = 0$ $j=1,2,\ldots,m$

(4) $\lambda_j^* \geq 0$ $j=1,2,\ldots,m$

(5) $\rho_i^* x_i^* = 0 \;\forall\, i$, i.e., $\sum_{i=1}^{n} x_i^* \rho_i^* = (\boldsymbol{x}^*)^{\mathrm{T}} \cdot \left[\boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \boldsymbol{\nabla g_j}(\boldsymbol{x}^*) \right] = 0$

NOTE THAT NO EXPLICIT MULTIPLIERS ARE USED FOR THE NONNEGATIVITY CONSTRAINTS!

The Karush-Kuhn-Tucker conditions are the most commonly used ones to check for optimality. However, they are actually *not valid under all conditions*.

There do exist a somewhat weaker set of necessary conditions (known as the Fritz John conditions) that are valid at all times; however in many instances they do not provide the same information as the Karush-Kuhn-Tucker conditions.

FRITZ JOHN CONDITIONS

These weaker necessary conditions for the problem

Minimize $f(\boldsymbol{x})$

st $\quad g_j(\boldsymbol{x}) \leq 0, \quad j=1,2,\ldots,m$

$\boldsymbol{x} \in \mathbb{R}^n.$

are based on the so-called ***weak*** Lagrangian

$$\tilde{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \lambda_0 f(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{x})$$

and may be stated as follows:

If $\boldsymbol{x}^*$ is a minimizer then $\exists\ \boldsymbol{\lambda^*} \in \mathbb{R}^{m+1}$ such that

(1) $$\boldsymbol{\nabla_x \tilde{L}}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \lambda_0^* \boldsymbol{\nabla f}(\boldsymbol{x}^*) + \sum_{j=1}^{m} \lambda_j^* \boldsymbol{\nabla g_j}(\boldsymbol{x}^*) = \boldsymbol{0}$$

(2) $\quad g_j(\boldsymbol{x}^*) \leq 0, \qquad j=1,2,\ldots,m$

(3) $\quad \lambda_j^* \cdot \{ g_j(\boldsymbol{x}^*)\} = 0, \qquad j=1,2,\ldots,m$

(4) $\quad \boldsymbol{\lambda}^* \geq \boldsymbol{0}$ and $\boldsymbol{\lambda}^* \neq \boldsymbol{0}.$

While the Fritz John conditions are *always* necessary for $\boldsymbol{x}^*$ to be a solution, the Karush-Kuhn-Tucker conditions are necessary <u>*provided*</u> certain conditions known as CONSTRAINT QUALIFICATIONS are met. In effect:

$$\boxed{\textbf{Local Optimum}} \Rightarrow \boxed{\textbf{Fritz John}} \xrightarrow[\textbf{\textit{Qualification}}]{\textbf{\textit{Constraint}}} \boxed{\textbf{Karush-Kuhn-Tucker}}$$

An example of a constraint qualification is:

The set $\mathbb{S} = \{\nabla \boldsymbol{g}_j(\boldsymbol{x}^*) \mid j \in J\}$ is *linearly independent*, where $J = \{j | g_j(\boldsymbol{x}^*)=0\}$.

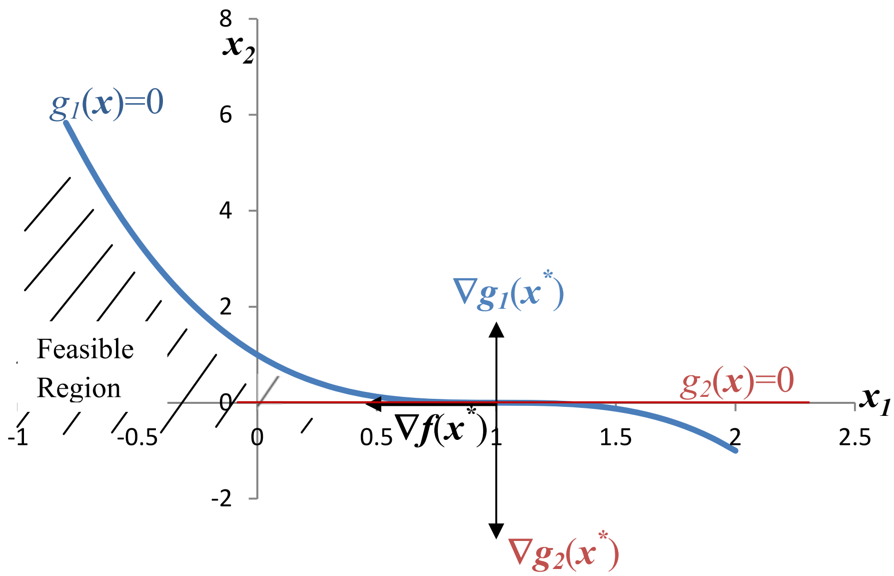
The above C.Q. states that the gradient vectors of all the constraints that are binding (satisfied as strict equalities) at $\boldsymbol{x}^*$ should be linearly independent.

In the Fritz John conditions, suppose that $\lambda_0^*= 0$. Then the conditions are satisfied for *any* function f at the point $\boldsymbol{x}^*$, regardless of whether or not that function has a minimum at $\boldsymbol{x}^*$!

This is the main weakness of the Fritz John conditions. If $\lambda_0^*= 0$, then they make no use of the objective and they are of no practical use in locating an optimal point.

Constraint qualifications essentially are constraints that *ensure* that $\lambda_0^*>0$, so that if we redefine λ_j^* by $(\lambda_j^*/\lambda_0^*)$ for $j=0,1,...,m$, we can reduce the Fritz John conditions to the Karush-Kuhn-Tucker conditions ($\lambda_0^* =1$ now and may thus be ignored).

EXAMPLE: Minimize $-x_1$, st $g_1(\boldsymbol{x})=x_2-(1-x_1)^3\leq 0$; $g_2(\boldsymbol{x})= -x_2\leq 0$.



The optimum solution is at $\boldsymbol{x}^* = \begin{bmatrix}1\\0\end{bmatrix}$, where we also have

$$\nabla f(\boldsymbol{x}^*) = \begin{bmatrix}-1\\0\end{bmatrix}; \qquad \nabla g_1(\boldsymbol{x}^*) = \begin{bmatrix}0\\1\end{bmatrix}; \qquad \nabla g_2(\boldsymbol{x}^*) = \begin{bmatrix}0\\-1\end{bmatrix}$$

Note that the constraint qualification is not met since both constraints are active at $\boldsymbol{x}^*$, but the gradient vectors of these constraints are not linearly independent.

FRITZ JOHN CONDITIONS: $\lambda_0\begin{bmatrix}-1\\0\end{bmatrix} + \lambda_1\begin{bmatrix}0\\1\end{bmatrix} + \lambda_2\begin{bmatrix}0\\-1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$

Satisfied for $\lambda_0=0$, and $\lambda_1= \lambda_2=$ arbitrary $\alpha > 0$.

KARUSH-KUHN-TUCKER CONDITIONS: $-\nabla f(\boldsymbol{x}^*)= \begin{bmatrix}1\\0\end{bmatrix} = \lambda_1\begin{bmatrix}0\\1\end{bmatrix} + \lambda_2\begin{bmatrix}0\\-1\end{bmatrix}$

$\Rightarrow \lambda_1(0) + \lambda_2(0) = 1$ and $\lambda_1(1) + \lambda_2(-1) = 0$

Inconsistent! (cannot be satisfied).

CONVEX PROGRAMS

The following NLP (Program **P4**) is referred to as a convex program:

Minimize $f(\boldsymbol{x})$

st $\quad g_j(\boldsymbol{x}) \leq 0, \quad j=1,2,\ldots,m,$

$g_j(\boldsymbol{x}) \geq 0, \quad j=m+1,\ldots,p,$

$h_j(\boldsymbol{x}) = 0, \quad j=1,2,\ldots,q,$

$\boldsymbol{x} \in \mathbb{R}^n.$

where

f is a proper convex function on $\mathbb{R}^n$

g_j are proper convex functions on $\mathbb{R}^n$, $j=1,\ldots,m$

g_j are proper concave functions on $\mathbb{R}^n$, $j=m+1,\ldots,p$

h_j are linear functions of the form

$$h_j(\boldsymbol{x}) = \sum_{k=1}^{n} a_{jk} x_k - b_j$$

(Note: A *proper* function has at least one point in its domain that is $<+\infty$ (i.e., its *effective* domain is non-empty), and it is also $>-\infty$ everywhere in its domain.)

Thus a convex minimization program has a convex objective, and the set of feasible solutions is a convex set.

EXERCISE: Show that the set of feasible points determined by the constraints of program **P4** is convex.

Q. Why are convex programs important?

A. Because they are 'well-behaved' programs with nice results!

Slater's Condition: For convex programs, if $\exists\ \boldsymbol{x}' \in \mathbb{R}^n \ni$

$g_j(\boldsymbol{x}') < 0$ for $j=1,2,\ldots,m$, and

$g_j(\boldsymbol{x}') > 0$ for $j=m+1,\ldots,p$,

then the constraint qualification holds at $\boldsymbol{x}'$ (the relative interior is non-empty)

Furthermore, for a convex program where the constraint qualification holds, the **K-K-T** necessary conditions are also sufficient.

Many algorithms have been designed specifically for convex programming problems. A special case is (posynomial) geometric programming, where the constrained local optima also turn out to be global optima.

LAGRANGIAN DUALITY

Just like LP, NLP also has a duality theory. For the problem

$$\text{Min } f(\boldsymbol{x}), \quad \text{st} \quad g_j(\boldsymbol{x})\le 0,\ j=1,2,\ldots,m; \quad h_j(\boldsymbol{x})=0,\ j=1,2,\ldots,p; \quad \boldsymbol{x}\in\mathbb{S}$$

this is based upon the Lagrangian function $L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) = f(\boldsymbol{x}) + \Sigma_j\lambda_j g_j(\boldsymbol{x}) + \Sigma_j\mu_j h_j(\boldsymbol{x})$.

Primal	**Dual**
MINIMAX	**MAXIMIN**
Minimize $\bar{L}(\boldsymbol{x})$, st $\boldsymbol{x}\in\mathbb{S}$	Maximize $\hat{L}(\boldsymbol{\lambda},\boldsymbol{\mu})$, st $\boldsymbol{\lambda}\ge\boldsymbol{0}$
where $\bar{L}(\boldsymbol{x}) = \text{Max}_{\boldsymbol{\lambda}\ge\boldsymbol{0},\boldsymbol{\mu}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})$	where $\hat{L}(\boldsymbol{\lambda},\boldsymbol{\mu}) = \text{Min}_{\boldsymbol{x}\in\mathrm{S}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})$
$= \begin{cases} f(\boldsymbol{x}) \text{ if } g_j(\boldsymbol{x})\le 0\ \forall j \text{ and } h_j(\boldsymbol{x})=0\ \forall \text{ j} \\ +\infty \text{ if } g_j(\boldsymbol{x})>0 \text{ or } h_j(\boldsymbol{x})\ne 0 \text{ for some } j\end{cases}$	In particular, under conditions of convexity & differentiability this reduces to
and so	Max $L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})$
$\underset{\boldsymbol{x}\in\mathrm{S}}{\text{Min}}\ \bar{L}(\boldsymbol{x}) = \underset{\boldsymbol{x}\in\mathrm{S}}{\text{Min}}\left\{\underset{\boldsymbol{\lambda}\ge\boldsymbol{0},\boldsymbol{\mu}}{\text{Max}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})\right\}$	st $\boldsymbol{\nabla_x L(x,\lambda,\mu)=0}$
$= \underset{\boldsymbol{x}\in\mathrm{S}}{\text{Min}}\{f(\boldsymbol{x}) \mid g_j(\boldsymbol{x})\le 0, h_j(\boldsymbol{x})=0\ \forall\ j\}$	$\boldsymbol{\lambda}\ge\boldsymbol{0}$, $\boldsymbol{x}\in\mathbb{S}$
This is our original (Primal) problem!	

WEAK DUALITY THEOREM: This states that

$\hat{L}(\boldsymbol{\lambda},\boldsymbol{\mu}) \le L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) \le \bar{L}(\boldsymbol{x})$ for all $\boldsymbol{\lambda}\ge\boldsymbol{0}$ and $\boldsymbol{x}\in\mathbb{S}$

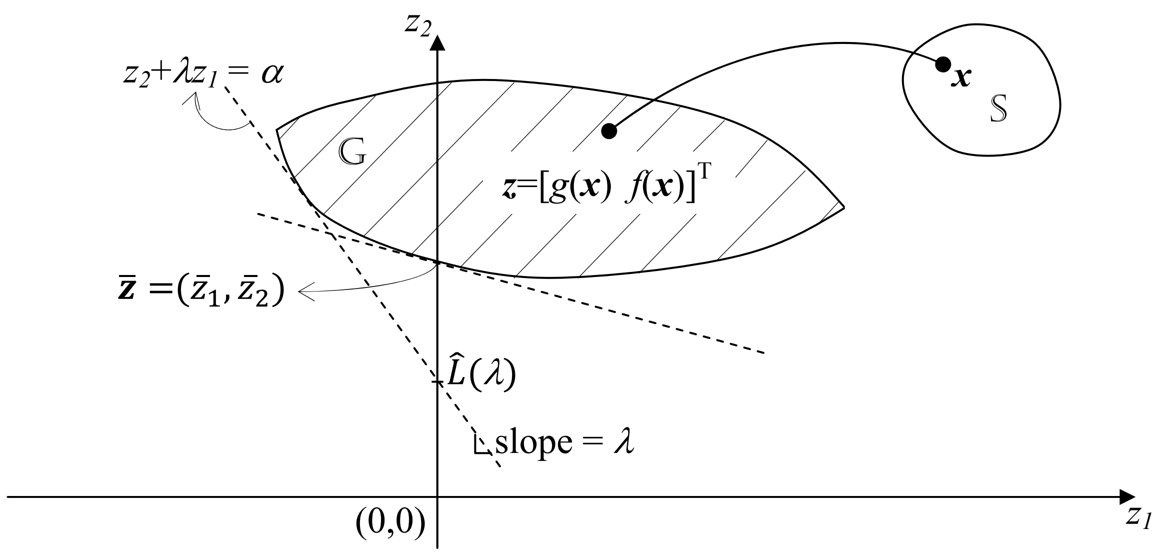
$\text{Max}_{\lambda,\mu}\ \{\text{Min}_x\ L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})\} \le \text{Min}_x\ \{\text{Max}_{\lambda,\mu}\ L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu})\}$

The quantity $[\bar{L}(\boldsymbol{x}) - \hat{L}(\boldsymbol{\lambda},\boldsymbol{\mu})]$ is referred to as the "duality gap."

GEOMETRIC INTERPRETATION OF DUALITY

Consider the problem: Minimize $f(\boldsymbol{x})$, st $g(\boldsymbol{x}) \leq 0$, $\boldsymbol{x} \in \mathbb{S}$.

Let $z_1 = g(\boldsymbol{x})$ and $z_2 = f(\boldsymbol{x})$



$\mathbb{G}$ is the image of the set $\mathbb{S}$ under the map $z_1 = g(\boldsymbol{x})$, $z_2 = f(\boldsymbol{x})$. Thus the original problem is equivalent to

Minimize z_2 , st $z_1 \leq 0$, $\boldsymbol{z} = (z_1, z_2) \in \mathbb{G}$

which yields the optimum solution $\bar{\boldsymbol{z}}$ shown above.

For $\lambda \geq 0$ given, $\hat{L}(\lambda) = \text{Min}_{\boldsymbol{x} \in \mathrm{S}} L(\boldsymbol{x}, \lambda) = \text{Min}_{\boldsymbol{x} \in \mathrm{S}} \{f(\boldsymbol{x}) + \lambda g(\boldsymbol{x})\}$

$$= \text{Min}_{\boldsymbol{z} \in \mathrm{G}} \, z_2 + \lambda z_1 = \alpha$$

Thus the dual problem is to find the slope λ of a supporting (tangential) line (or hyperplane) for which the intercept with the z_2 axis is maximized.

EXAMPLE 1: Minimize $4x_1^2 + 2x_1x_2 + x_2^2$

st $3x_1 + x_2 \geq 6$; $x_1, x_2 \geq 0$

Let $\mathbb{S} = \{\boldsymbol{x} \mid x_1 \geq 0, x_2 \geq 0\}$. The Lagrangian is

$$L(\boldsymbol{x},\lambda) = 4x_1^2 + 2x_1x_2 + x_2^2 + \lambda\cdot(-3x_1 - x_2 + 6)$$

The PRIMAL problem is

$$\underset{\boldsymbol{x}\in S}{\text{Minimize}}\left\{\underset{\boldsymbol{\lambda}\geq\boldsymbol{0}}{\text{Maximize}}\, L(\boldsymbol{x},\boldsymbol{\lambda})\right\} \qquad (= \underset{\boldsymbol{x}\in S}{\text{Min}}[\bar{L}(\boldsymbol{x})])$$

i.e., Min $4x_1^2 + 2x_1x_2 + x_2^2$

st $3x_1 + x_2 \geq 6$; $x_1, x_2 \geq 0$

The corresponding DUAL problem is

$$\underset{\boldsymbol{\lambda}\geq 0}{\text{Maximize}}\left\{\underset{\boldsymbol{x}\in S}{\text{Minimize}}\, L(\boldsymbol{x},\boldsymbol{\lambda})\right\} \qquad (= \underset{\boldsymbol{\lambda}\geq 0}{\text{Max}}\left[\hat{L}(\boldsymbol{\lambda})\right])$$

Consider the dual objective

$$\hat{L}(\lambda) = \underset{\boldsymbol{x}\in S}{\text{Minimize}}\{L(\boldsymbol{x},\lambda) = 4x_1^2 + 2x_1x_2 + x_2^2 + \lambda(-3x_1 - x_2 - 6)\}$$

For this particular example, $L(\boldsymbol{x},\lambda)$ (the objective function) is convex; therefore to get the minimum the necessary and sufficient conditions are that

$$\frac{\partial L}{\partial x_1}(\boldsymbol{x},\lambda) = 8x_1 + 2x_2 - 3\lambda = 0$$

$$\frac{\partial L}{\partial x_2}(\boldsymbol{x},\lambda) = 2x_1 + 2x_2 - \lambda = 0 \qquad \text{i.e.,} \qquad x_1^* = \frac{\lambda}{3},\ x_2^* = \frac{\lambda}{6}$$

Note that if $\lambda \geq 0$, this satisfies $\boldsymbol{x}\in\mathbb{S}$.

Thus we have $\hat{L}(\lambda) = L[(\lambda/3),(\lambda/6),\lambda] = 6\lambda - (7/12)\lambda^2$

The dual problem thus reduces to

Maximize $6\lambda - (7/12)\lambda^2$

st $\lambda \geq 0$

The objective function for this subproblem turns out to be concave (CHECK and VERIFY for yourself):

$$\therefore \frac{\partial \hat{L}}{\partial \lambda} = 6 - \frac{7}{12}(2\lambda) = 0 \implies \lambda = \frac{36}{7} \ (> 0)$$

Since 36/7 is not < 0, it solves the problem, i.e.,

DUAL SOLUTION: $\lambda^* = \frac{36}{7}, \hat{L}(\lambda^*) = \frac{108}{7}$

PRIMAL SOLUTION: $x_1^* = \frac{\lambda^*}{3} = \frac{12}{7} > 0; \quad x_2^* = \frac{\lambda^*}{6} = \frac{6}{7} > 0;$

and $3x_1^* + x_2^* = (36/7) + (6/7) = 6$ ---- feasible.

Complementary slackness is satisfied since $\lambda^*(6\text{-}3x_1^* + x_2^*) = 0$

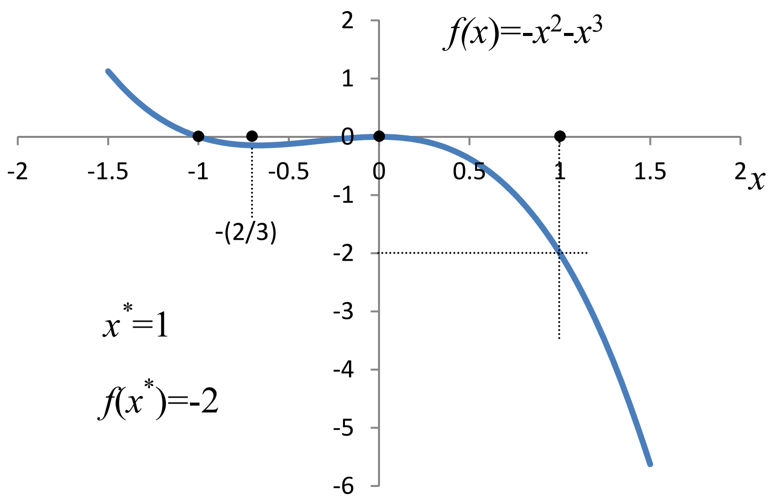
$\therefore$ (12/7, 6/7) is optimal for the Primal

Furthermore

$$4\left(\frac{12}{7}\right)^2 + 2\left(\frac{12}{7}\right)\left(\frac{6}{7}\right) + \left(\frac{6}{7}\right)^2 = \left(\frac{108}{7}\right) \implies \textbf{NO DUALITY GAP!}$$

<u>EXAMPLE 2</u>: Min $(-x^2 - x^3)$

(P) st $x^2 \leq 1$



The Lagrangian for this problem is $L(x,\lambda) = -x^2 - x^3 + \lambda(x^2-1)$

Consider the PRIMAL problem; optimality conditions are:

$\nabla_x \boldsymbol{L}(\boldsymbol{x}, \lambda) = 0 \quad \Rightarrow \quad -2x - 3x^2 + 2\lambda x = 0$

$\lambda(x^2-1)=0, \qquad \lambda \geq 0; \qquad x^2 \leq 1.$

(a) $\lambda=0 \quad \Rightarrow \quad x = (-2/3)$ OR $x = 0;$

(b) $\lambda>0 \quad \Rightarrow \quad x^2-1= 0 \quad \Rightarrow \quad x = \pm 1$

$x=+1 \quad \Rightarrow \quad \lambda=2.5 \qquad\qquad x=-1 \quad \Rightarrow \quad \lambda= -0.5$

Therefore the solutions are

(x,λ) =	(-2/3,0)	or	(0,0)	or	(1,2.5)
$f(x)$ =	-4/27	or	0	or	-2

Thus the primal optimum is $x^* = 1$, $f(x^*) = -2$.

Let us look at the dual (**D**) and see if there is any duality gap. The dual is:

(**D**) Maximize $\hat{L}(\lambda)$; st $\lambda \geq 0$

where

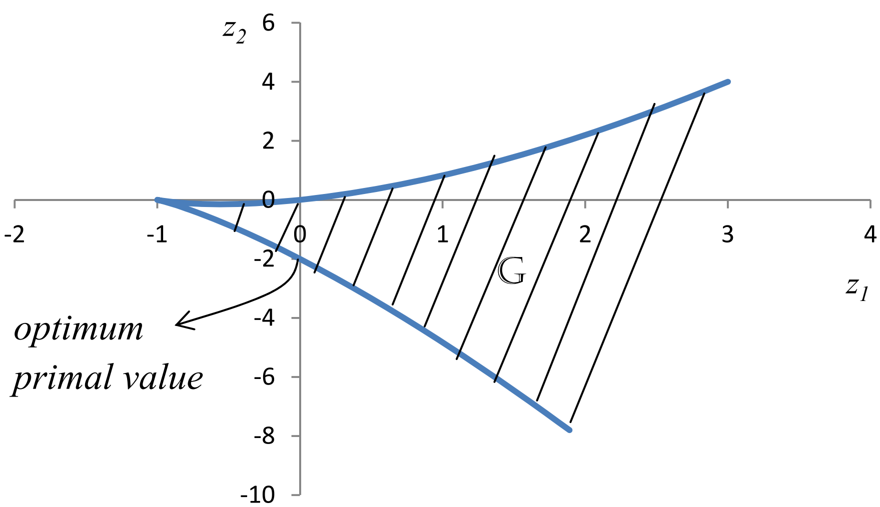
$$\hat{L}(\lambda) = \underset{x}{\text{Min}}\, L(x,\lambda) = \underset{x}{\text{Min}}\{-x^2 - x^3 + \lambda(x^2 - 1)\}$$

$= -\infty$ for all values of λ!

$$\therefore \quad \text{Max}_{\lambda \geq 0}\, \hat{L}(\lambda) = -\infty$$

Thus the primal objective = -2; dual objective = -∞. There is a **DUALITY GAP**!

Looking at it geometrically, let $z_1 = g(x) = x^2 - 1$, and $z_2 = f(x) = -x^2 - x^3$.



The supporting hyperplane is the plane (line) through (-1,0) that intersects the z_2-axis at -∞.

Saddlepoint Sufficiency

Recall that from Weak Duality $\hat{L}(\boldsymbol{\lambda}) \le L(\boldsymbol{x},\boldsymbol{\lambda}) \le \bar{L}(\boldsymbol{x})$.

<u>Definition</u>: The point $(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}})$ is a ***saddlepoint*** of $L(\boldsymbol{x},\boldsymbol{\lambda})$ if

1. $L(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}}) \le L(\boldsymbol{x},\bar{\boldsymbol{\lambda}}) \forall \boldsymbol{x} \in S$, i.e., $L(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}}) = Min_{x \in S} L(\boldsymbol{x},\bar{\boldsymbol{\lambda}}) = \hat{L}(\bar{\boldsymbol{\lambda}})$, and
2. $L(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}}) \ge L(\bar{\boldsymbol{x}},\boldsymbol{\lambda}) \forall \boldsymbol{\lambda} \ge \boldsymbol{0}$, i.e., $L(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}}) = Max_{\lambda \ge 0} L(\bar{\boldsymbol{x}},\boldsymbol{\lambda}) = \bar{L}(\bar{\boldsymbol{x}})$

In other words, if $(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}})$ is a ***saddlepoint*** of $L(\boldsymbol{x},\boldsymbol{\lambda})$ then $\bar{L}(\bar{\boldsymbol{x}}) = L(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}}) = \hat{L}(\bar{\boldsymbol{\lambda}})$,

i.e., primal objective = dual objective at the point and the duality gap = 0.

Now consider the following Primal problem:

Min $f(\boldsymbol{x})$, st $g_j(\boldsymbol{x}) \le 0$, $\quad j=1,2,\ldots,m;\ \boldsymbol{x} \in \boldsymbol{S}$.

where $f(\boldsymbol{x})$ and $g_j(\boldsymbol{x})$ are all convex functions, and $\boldsymbol{S}$ is a convex set.

Saddlepoint Sufficiency Condition: If $\bar{\boldsymbol{x}} \in S$ and $\bar{\boldsymbol{\lambda}} \ge 0$, then $(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}})$ is a saddlepoint of $L(\boldsymbol{x},\boldsymbol{\lambda})$ if, and only if,

- $\bar{\boldsymbol{x}}$ minimizes $L(\boldsymbol{x},\bar{\boldsymbol{\lambda}}) = f(\boldsymbol{x}) + \bar{\boldsymbol{\lambda}}^T \boldsymbol{g}(\boldsymbol{x})$ over $\boldsymbol{S}$.
- $g_j(\boldsymbol{x}) \le 0$ for each $j=1,2,\ldots,m$
- $\bar{\boldsymbol{\lambda}}_j g_j(\bar{\boldsymbol{x}}) = 0$, which implies that $f(\bar{\boldsymbol{x}}) = L(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}})$

If $(\bar{\boldsymbol{x}},\bar{\boldsymbol{\lambda}})$ is a ***saddlepoint*** of $L(\boldsymbol{x},\boldsymbol{\lambda})$ then $\bar{\boldsymbol{x}}$ solves the Primal problem above and $\bar{\boldsymbol{\lambda}}$ solves the Dual problem: Maximize $\hat{L}(\boldsymbol{\lambda})$ st $\boldsymbol{\lambda} \ge \boldsymbol{0}$, where $\hat{L}(\boldsymbol{\lambda}) = Min_{x \in S} L(\boldsymbol{x},\boldsymbol{\lambda})$

Strong Duality Theorem

Consider the primal problem: Find

$$\boxed{\begin{aligned} &\Phi = \inf f(\boldsymbol{x}) \\ &\text{st} \quad g_j(\boldsymbol{x}) \le 0, j = 1, 2, \ldots, m_1 \\ &\qquad h_j(\boldsymbol{x}) = 0, j = 1, 2, \ldots, m_2 \\ &\qquad\quad \boldsymbol{x} \in \boldsymbol{S} \end{aligned}}$$

where $\boldsymbol{S} \subseteq \mathbb{R}^n$ is nonempty and convex, $f(\boldsymbol{x})$ and $g_j(\boldsymbol{x})$ are convex & $h_j(\boldsymbol{x})$ are linear.

<u>Note</u>: *infimum* may be replaced by *minimum* if the minimum is attained at some $\boldsymbol{x}$

Define the dual problem: Find

$$\boxed{\begin{aligned} &\Psi = \sup \hat{L}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &st \qquad \boldsymbol{\lambda} \ge \boldsymbol{0} \end{aligned}}$$

where $\hat{L}(\boldsymbol{\lambda},\boldsymbol{\mu}) = inf_{x \in S} \left\{ f(\boldsymbol{x}) + \boldsymbol{\lambda}^T g(\boldsymbol{x}) + \boldsymbol{\mu}^T h(\boldsymbol{x}) \right\}$.

<u>**Strong Duality Theorem**</u>**:** Assume that the following *Constraint Qualification*

holds:

$$\boxed{\begin{aligned} &\text{There exists } \hat{\boldsymbol{x}} \text{ such that} \\ &g_j(\hat{\boldsymbol{x}}) < 0, j = 1,2,\ldots,m_1 \\ &h_j(\hat{\boldsymbol{x}}) = 0, j = 1,2,\ldots,m_2 \\ &\& \, 0 \in \operatorname{int} \boldsymbol{h}(\boldsymbol{S}) \end{aligned}}$$

where $\boldsymbol{h}(S) = \boldsymbol{h}(\boldsymbol{x})$: $\boldsymbol{x} \in \boldsymbol{S}$.

Then $\boxed{\Phi=\Psi}$, i.e. there is no duality gap. Furthermore, if $\Phi > -\infty$, then

- $\Psi = \hat{L}(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ for some $\boldsymbol{\lambda}^* \ge \mathbf{0}$
- If $\boldsymbol{x}^*$ solves the primal, it satisfies complementary slackness: $\lambda_j^* g_j(\boldsymbol{x}^*) = 0 \; \forall \, j$.

QUADRATIC PROGRAMMING

Quadratic Programs are convex minimization programs that have a convex (quadratic) objective and a convex constraint set formed by linear constraints. The **QP** primal problem is:

(**P**) Minimize $\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}Q\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$

st $\quad \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \qquad (\cong \boldsymbol{b}\text{-}\boldsymbol{A}\boldsymbol{x} \leq \mathbf{0})$

(any nonnegativity restrictions are included in $\boldsymbol{A}\boldsymbol{x}\geq\boldsymbol{b}$)

where

$\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{Q}$ is a $(n\times n)$, real, symmetric, positive definite matrix

$\boldsymbol{c} \in\mathbb{R}^n$, $\boldsymbol{A}$ is a $(m\times n)$ real matrix, $\boldsymbol{b}\in\mathbb{R}^m$.

Thus we have a convex objective & linear constraints.

The Lagrangian function is

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$$

The dual **QP** problem is

$$\operatorname*{Max}_{\boldsymbol{\lambda}\geq\mathbf{0}}\left\{\operatorname*{Min}_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\lambda})\right\}, \quad \text{i.e.,} \quad \operatorname*{Max}_{\boldsymbol{\lambda}\geq\mathbf{0}} L(\boldsymbol{x},\boldsymbol{\lambda})\,, \text{st } \boldsymbol{\nabla_x L}(\boldsymbol{x},\boldsymbol{\lambda}) = \mathbf{0}$$

This yields

(**D**) Maximize $L(\boldsymbol{x},\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$

st $\quad \nabla_x \boldsymbol{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{0}; \qquad \boldsymbol{\lambda}\geq\mathbf{0}.$

The dual constraints imply that $\boldsymbol{x}^{\mathrm{T}}[\boldsymbol{Qx}+ \boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}] = \boldsymbol{x}^{\mathrm{T}}[0]$

i.e., $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Qx}+ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{c} - \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{\lambda}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Qx}+ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{c} - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x} = 0$ $\qquad (\otimes)$

The objective function can be rearranged as

$\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{b} - \frac{1}{2}\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Qx} + (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Qx} + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{c} - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x}) = \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{b} - \frac{1}{2}\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Qx}$ $\qquad$ from $(\otimes)$

Thus the dual reduces to

Maximize $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{b} - \frac{1}{2}\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Qx}$

st $\qquad \boldsymbol{Qx} + \boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{0}$; $\boldsymbol{\lambda}\geq\boldsymbol{0}$.

If $\boldsymbol{Q}=[\boldsymbol{0}]$, then $\qquad$ (**P**) is Min $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$, st $\boldsymbol{Ax}\geq\boldsymbol{b}$

(**D**) is Max $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{b}$, st $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{c}$

the regular LP primal-dual pair!

For $\boldsymbol{Q} \neq [\boldsymbol{0}]$

(**P**) has *n* variables, *m* linear inequality constraints

(**D**) has (*m*+*n*) variables, *n* equality constraints, and *m* nonnegativity constraints

If $\boldsymbol{Q}$ is nonsingular (i.e., $\boldsymbol{Q}^{-1}$ exists) then from the dual constraints,

$$\boldsymbol{Qx}^* + \boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{0} \;\Rightarrow\; \boldsymbol{x}^* + \boldsymbol{Q}^{-1}\boldsymbol{c} - \boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{0}$$

i.e., $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}[\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{c}]$; thus we may eliminate $\boldsymbol{x}$ altogether from the dual problem!

First, recall that given any two matrices $\boldsymbol{U}$ and $\boldsymbol{V}$, ***(i)*** $[\boldsymbol{UV}]^{\mathrm{T}}=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}$, ***(ii)*** $[\boldsymbol{U}^{\mathrm{T}}]^{\mathrm{T}} = \boldsymbol{U}$ and ***(iii)*** $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{V}=\boldsymbol{V}^{\mathrm{T}}\boldsymbol{U}$ (assuming compatibility). Also, ***(iv)*** $\boldsymbol{Q}$, $\boldsymbol{Q}^{-1}$ are symmetric and equal to their transposes.

Substituting this value of $\boldsymbol{x}$ into the dual objective function and rearranging:

$$\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{b} - \tfrac{1}{2}\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{b} - \tfrac{1}{2}\,[\boldsymbol{Q}^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{c})]^{\mathrm{T}} \cdot \boldsymbol{Q} \cdot [\boldsymbol{Q}^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{c})]$$

$$= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \tfrac{1}{2}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{c})^{\mathrm{T}} \cdot \boldsymbol{Q}^{-1} \cdot \boldsymbol{Q} \cdot \boldsymbol{Q}^{-1} \cdot (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{c}) \qquad \text{(i) and (iv)}$$

$$= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \tfrac{1}{2}\,[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda})^{\mathrm{T}} - \boldsymbol{c}^{\mathrm{T}}] \cdot \boldsymbol{Q}^{-1} \cdot (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{c})$$

$$= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \tfrac{1}{2}\,[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}^{-1} - \boldsymbol{c}^{\mathrm{T}}\,\boldsymbol{Q}^{-1}] \cdot (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{c}) \qquad \text{(i) and (ii)}$$

$$= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \tfrac{1}{2}\,[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} - \boldsymbol{c}^{\mathrm{T}}\,\boldsymbol{Q}^{-1}\,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{c}]$$

$$= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \tfrac{1}{2}\,[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} - \boldsymbol{c}^{\mathrm{T}}\,\boldsymbol{Q}^{-1}\,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{c})^{\mathrm{T}}\boldsymbol{\lambda}] \qquad \text{(iii)}$$

$$= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \tfrac{1}{2}\,[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} - \boldsymbol{c}^{\mathrm{T}}\,\boldsymbol{Q}^{-1}\,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} - \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{Q}^{-1})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}] \qquad \text{(i)}$$

$$= \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} - \tfrac{1}{2}\,[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} - 2\boldsymbol{c}^{\mathrm{T}}\,\boldsymbol{Q}^{-1}\,\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}] \qquad \text{(iv)}$$

$$= \underbrace{[\boldsymbol{b}^{\mathrm{T}} + \boldsymbol{c}^{\mathrm{T}}\,\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}]}_{}\boldsymbol{\lambda} - \tfrac{1}{2}\,[\boldsymbol{\lambda}^{\mathrm{T}}\underbrace{\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}}_{}\boldsymbol{\lambda}] - \tfrac{1}{2}\,[\boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c}]$$

$$= \boldsymbol{e}^{\mathrm{T}}\boldsymbol{\lambda} - \tfrac{1}{2}\,[\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\lambda}] - \tfrac{1}{2}\,[\boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c}]$$

where

$$\boldsymbol{e} = \boldsymbol{b} + \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{c} \qquad \text{and} \qquad \boldsymbol{D} = -\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}}$$

so that the dual is (<u>*as long as **Q** is nonsingular*</u>)

$$\underset{\boldsymbol{\lambda}\geq\boldsymbol{0}}{\text{Maximize}}\left\{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\lambda} + \frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\lambda} - \frac{1}{2}\boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c}\right\}$$

EXAMPLE: Minimize $\frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 - 2x_1 - 2x_2$

st $0 \le x_1 \le 1 \Rightarrow \begin{cases} -x_1 \ge -1 \\ x_1 \ge 0 \end{cases}$; $0 \le x_2 \le 1 \Rightarrow \begin{cases} -x_2 \ge -1 \\ x_2 \ge 0 \end{cases}$

i.e., Minimize $\frac{1}{2}\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} -2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$\boldsymbol{x}^{\mathrm{T}}$ $\boldsymbol{Q}$ $\boldsymbol{x}$ $\boldsymbol{c}^{\mathrm{T}}$ $\boldsymbol{x}$

st $\begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \ge \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$

$\boldsymbol{A}$ $\boldsymbol{x}$ $\boldsymbol{b}$

In order to define the dual we have $\boldsymbol{Q}^{-1}$ $(=\boldsymbol{Q}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\Rightarrow \boldsymbol{e} = \boldsymbol{b} + \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{c} = \begin{bmatrix} -1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -2 \\ -2 \end{bmatrix} = = \begin{bmatrix} 1 \\ 1 \\ -2 \\ -2 \end{bmatrix}$$

$$\boldsymbol{D} = -\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^{\mathrm{T}} = -\begin{bmatrix} -1 & 0 \\ 0 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Therefore the dual is

$$\underset{\boldsymbol{\lambda} \ge \boldsymbol{0}}{\text{Maximize}} \left\{ \boldsymbol{e}^{\mathrm{T}}\boldsymbol{\lambda} + \frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\lambda} - \frac{1}{2}\boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c} \right\}$$

with the $\boldsymbol{e}$ and $\boldsymbol{D}$ as given above. The optimal primal solution $\boldsymbol{x}^* = \boldsymbol{Q}^{-1}[\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}^* - \boldsymbol{c}]$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\left\{ \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} - \begin{bmatrix} -2 \\ -2 \end{bmatrix} \right\}$$

AN ALGORITHM FOR QUADRATIC PROGRAMMING PROBLEMS: The Method of HILDETH AND D'ESPO

This is a "cyclic co-ordinate search" method applied to the **QP** dual:

$$\underset{\boldsymbol{\lambda}\geq\mathbf{0}}{\text{Maximize}}\ \hat{L}(\boldsymbol{\lambda}) = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{\lambda} + \frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\lambda}$$

STEP 1: Start with an initial $\boldsymbol{\lambda}$ (e.g. $\boldsymbol{\lambda}=\mathbf{0}$)

STEP 2: For j=1,2,...,m do the following

Search for the maximum in the direction parallel to the λ_j-axis by <u>fixing</u> λ_k for $k\neq j$ and solving $\frac{\partial\hat{L}}{\partial\lambda_j}(\boldsymbol{\lambda}) = 0$ for λ_j. If this $\lambda_j<0$, then fix $\lambda_j=0$.

STEP 3: If $\boldsymbol{\lambda}$ from this iteration the same as the $\boldsymbol{\lambda}$ from the previous iteration STOP else go to STEP 2.

NOTE: In order to obtain $\frac{\partial\hat{L}}{\partial\lambda_j}(\boldsymbol{\lambda})$ we need

$$\nabla_{\boldsymbol{\lambda}}\hat{L}(\boldsymbol{\lambda}) = \boldsymbol{D}\boldsymbol{\lambda} + \boldsymbol{e}$$

Consider the previous example where the dual was defined as:

$$\underset{\boldsymbol{\lambda}\geq\mathbf{0}}{\text{Maximize}}\left\{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\lambda} + \frac{1}{2}\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\lambda} - \frac{1}{2}\boldsymbol{c}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{c}\right\}$$

with

$$\boldsymbol{c} = \begin{bmatrix}-2\\-2\end{bmatrix};\ \boldsymbol{Q} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix};\ \boldsymbol{e} = \begin{bmatrix}1\\1\\-2\\-2\end{bmatrix};\ \boldsymbol{D} = \begin{bmatrix}-1 & 0 & 1 & 0\\0 & -1 & 0 & 1\\1 & 0 & -1 & 0\\0 & 1 & 0 & -1\end{bmatrix}$$

Substituting the appropriate values of $\boldsymbol{D}$, $\boldsymbol{c}$, $\boldsymbol{Q}$ and $\boldsymbol{e}$ yields

$$\underset{\boldsymbol{\lambda}\geq\boldsymbol{0}}{\text{Maximize}}[1 \quad 1\ -2\ -2]\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\end{bmatrix}+\frac{1}{2}[\lambda_1\ \ \lambda_2\ \ \lambda_3\ \ \lambda_4]\begin{bmatrix}-1&0&1&0\\0&-1&0&1\\1&0&-1&0\\0&1&0&-1\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\end{bmatrix}$$

$$-\frac{1}{2}[-2\ \ -2]\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}-2\\-2\end{bmatrix}$$

$$\underset{\boldsymbol{\lambda}\geq\boldsymbol{0}}{\text{Maximize}}(\lambda_1+\lambda_2-2\lambda_3-2\lambda_4)-\frac{1}{2}(\lambda_1^2+\lambda_2^2+\lambda_3^2+\lambda_4^2+2\lambda_1\lambda_3+2\lambda_2\lambda_4)-4$$

The **K-K-T** conditions for this yield

$\lambda_1(1-\lambda_1+\lambda_3)=0$

$\lambda_2(1-\lambda_2+\lambda_4)=0$ Need to solve this system to obtain the optimal $\boldsymbol{\lambda}$

$\lambda_3(-2-\lambda_3+\lambda_1)=0$

$\lambda_4(-2-\lambda_4+\lambda_2)=0$

Instead, let us try the method of Hildeth and D'Espo…

$$\boldsymbol{\nabla\hat{L}}=\boldsymbol{D\lambda}+\boldsymbol{e}=\begin{bmatrix}-1&0&1&0\\0&-1&0&1\\1&0&-1&0\\0&1&0&-1\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\end{bmatrix}+\begin{bmatrix}1\\1\\-2\\-2\end{bmatrix}$$

$$=\begin{bmatrix}-\lambda_1+\lambda_3+1\\-\lambda_2+\lambda_4+1\\\lambda_1-\lambda_3-2\\\lambda_2-\lambda_4-2\end{bmatrix}\begin{matrix}\rightarrow & \partial\hat{L}/\partial\lambda_1\\\rightarrow & \partial\hat{L}/\partial\lambda_2\\\rightarrow & \partial\hat{L}/\partial\lambda_3\\\rightarrow & \partial\hat{L}/\partial\lambda_4\end{matrix}$$

CYCLE-1 Let $\boldsymbol{\lambda} = [0\ \ 0\ \ 0\ \ 0]^T$

Solve $\frac{\partial \hat{L}}{\partial \lambda_1} = 0$ (with $\lambda_2 = \lambda_3 = \lambda_4 = 0$) to obtain λ_1

$\Rightarrow$ $-\lambda_1+\lambda_3+1=0$ $\Rightarrow$ $-\lambda_1+0+1=0$ $\Rightarrow$ $\lambda_1=1$

$\therefore$ $\boldsymbol{\lambda} = [1\ \ 0\ \ 0\ \ 0]^T$

Solve $\frac{\partial \hat{L}}{\partial \lambda_2} = 0$ (with $\lambda_1 = 1, \lambda_3 = \lambda_4 = 0$) to obtain λ_2

$\Rightarrow$ $-\lambda_2+\lambda_4+1=0$ $\Rightarrow$ $-\lambda_2+0+1=0$ $\Rightarrow$ $\lambda_2=1$

$\therefore$ $\boldsymbol{\lambda} = [1\ \ 1\ \ 0\ \ 0]^T$

Solve $\frac{\partial \hat{L}}{\partial \lambda_3} = 0$ (with $\lambda_1 = \lambda_2 = 1, \lambda_4 = 0$) to obtain λ_3

$\Rightarrow$ $\lambda_1-\lambda_3-2=0$ $\Rightarrow$ $1-\lambda_3-2=0$ $\Rightarrow$ $\lambda_3=-1$ $\therefore$ Fix $\lambda_3=0$

$\therefore$ $\boldsymbol{\lambda} = [1\ \ 1\ \ 0\ \ 0]^T$

Solve $\frac{\partial \hat{L}}{\partial \lambda_4} = 0$ (with $\lambda_1 = \lambda_2 = 1, \lambda_3 = 0$) to obtain λ_4

$\Rightarrow$ $\lambda_2-\lambda_4-2=0$ $\Rightarrow$ $1-\lambda_4-2=0$ $\Rightarrow$ $\lambda_4=-1$ $\therefore$ Fix $\lambda_4=0$

$\therefore$ $\boldsymbol{\lambda} = [1\ \ 1\ \ 0\ \ 0]^T$

END OF CYCLE 1

Another cycle is required since $\boldsymbol{\lambda}$ has changed from $[0\ \ 0\ \ 0\ \ 0]^T$ to $[1\ \ 1\ \ 0\ \ 0]^T$.

1. $-\lambda_1+\lambda_3+1=0$; with $\lambda_2=1,\ \lambda_3=0,\ \lambda_4=0$ $\Rightarrow\ \lambda_1=1$

   $\Rightarrow\ \boldsymbol{\lambda} = [1\ \ 1\ \ 0\ \ 0]^T$

2. $-\lambda_2+\lambda_4+1=0$; with $\lambda_1=1,\ \lambda_3=0,\ \lambda_4=0$ $\Rightarrow\ \lambda_2=1$

   $\Rightarrow\ \boldsymbol{\lambda} = [1\ \ 1\ \ 0\ \ 0]^T$

3. $\lambda_1-\lambda_3-2=0$; with $\lambda_1=1,\ \lambda_2=1,\ \lambda_4=10$ $\Rightarrow\ \lambda_3=-1$

   $\Rightarrow\ \boldsymbol{\lambda} = [1\ \ 1\ \ 0\ \ 0]^T$ ($\lambda_3<0$, so we set it to 0)

4. $\lambda_2-\lambda_4-2=0$; with $\lambda_1=1,\ \lambda_2=1,\ \lambda_3=1$ $\Rightarrow\ \lambda_4=-1$

   $\Rightarrow\ \boldsymbol{\lambda} = [1\ \ 1\ \ 0\ \ 0]^T$ ($\lambda_4<0$, so we set it to 0)

END OF CYCLE 2

Since the value of $\boldsymbol{\lambda}$ has not changed from the previous cycle the algorithm has converged.

Thus $\lambda_1^* = \lambda_2^* = 1$; $\lambda_3^* = \lambda_4^* = 0$. In order to now obtain the optimal primal vector $\boldsymbol{x}^*$

$$\boldsymbol{x}^* = \boldsymbol{Q}^{-1}[\boldsymbol{A}^T\boldsymbol{\lambda}^*-\boldsymbol{c}] = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\left\{\begin{bmatrix}-1 & 0 & 1 & 0\\ 0 & -1 & 0 & 1\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\\ \lambda_4\end{bmatrix} - \begin{bmatrix}-2\\ -2\end{bmatrix}\right\} = \begin{bmatrix}1\\ 1\end{bmatrix}$$

Dual Objective = Primal Objective = -3.

<u>EXERCISE</u>: Verify that **K-K-T** conditions for the dual are satisfied at $\boldsymbol{x}^*$.

WOLFE'S ALGORITHM FOR QP

$$\text{Minimize } \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \quad \text{st} \quad \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0}.$$

KARUSH-**K**UHN-**T**UCKER CONDITIONS

$\nabla_x \boldsymbol{L}(\boldsymbol{x},\boldsymbol{\lambda}) = \boldsymbol{0}$ $\Rightarrow$ $\boldsymbol{Qx} + \boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{0}$ $\Rightarrow$ $\boldsymbol{Qx} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = -\boldsymbol{c}$

Original Constraints $\Rightarrow$ $\boldsymbol{Ax} \geq \boldsymbol{b}$ $\Rightarrow$ $\boldsymbol{Ax} - \boldsymbol{y} = \boldsymbol{b}$ ($\boldsymbol{y}$≡surplus)

Complementary Slackness $\Rightarrow$ $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$

Nonnegativity $\Rightarrow$ $\boldsymbol{x} \geq \boldsymbol{0}$, $\boldsymbol{y} \geq \boldsymbol{0}$, $\boldsymbol{\lambda} \geq \boldsymbol{0}$.

Thus, we need to solve complementary slackness, nonnegativity for $\boldsymbol{\lambda}$, $\boldsymbol{y}$ and $\boldsymbol{x}$, along with

$$\left[\begin{array}{c|c|c} \boldsymbol{Q} & -\boldsymbol{A}^{\mathrm{T}} & \boldsymbol{0} \\ \hline \boldsymbol{A} & \boldsymbol{0} & -\boldsymbol{I} \end{array}\right] \left[\begin{array}{c} \boldsymbol{x} \\ \hline \boldsymbol{\lambda} \\ \hline \boldsymbol{y} \end{array}\right] = \left[\begin{array}{c} -\boldsymbol{c} \\ \hline \boldsymbol{b} \end{array}\right]$$

This is solved in a manner similartothePHASE-1 procedure of the simplex method, while also using a "restricted basis entry" (RBE) rule:

> **“If y_i is already in the basis at a level>0, do not consider λ_i as a candidate for entry into the basis, and vice versa. If y_i (λ_i) is in the basis already at a 0 level, λ_i (y_i) may enter only if y_i (λ_i) remains at the zero level.”**

Let w_j represent the artificial variable for the j^{th} equality. Then, for the objective we minimize the sum of the artificial variables ($\sum_j w_j$). Thus the problem is to

Minimize $W = \sum_j w_j$

st
$$\left[\begin{array}{c:c:c} \boldsymbol{Q} & -\boldsymbol{A}^{\mathrm{T}} & \boldsymbol{0} \\ \hdashline \boldsymbol{A} & \boldsymbol{0} & -\boldsymbol{I} \end{array}\right] \left[\begin{array}{c} \boldsymbol{x} \\ \hdashline \boldsymbol{\lambda} \\ \hdashline \boldsymbol{y} \end{array}\right] + \left[\boldsymbol{w}\right] = \left[\begin{array}{c} -\boldsymbol{c} \\ \hdashline \boldsymbol{b} \end{array}\right]$$

while using the RBE rule.

If the optimum objective value of the LP is zero, we have an optimum solution to the original problem; else the latter has no solution.

====================================

EXAMPLE: Minimize $\frac{1}{2} x_1^2 + x_2^2 - x_1 x_2 - x_1 - x_2$

st $x_1 + x_2 \leq 3$ (i.e., $-x_1 - x_2 \geq -3$)

$2x_1 + 3x_2 \geq 6$

$x_1, x_2 \geq 1.$

i.e.,
$$\text{Min } \underbrace{\frac{1}{2}[x_1 \;\; x_2]}_{\boldsymbol{x}^{\mathrm{T}}} \underbrace{\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}}_{\boldsymbol{Q}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\boldsymbol{x}} + \underbrace{[-1 \quad -1]}_{\boldsymbol{c}^{\mathrm{T}}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\boldsymbol{x}}$$

$$\text{st } \underbrace{\begin{bmatrix} -1 & -1 \\ 2 & 3 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\boldsymbol{x}} \geq \underbrace{\begin{bmatrix} -3 \\ 6 \\ 1 \\ 1 \end{bmatrix}}_{\boldsymbol{b}}, \quad \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\boldsymbol{x}} \geq \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\boldsymbol{0}}$$

Then it follows that

$$\left[\begin{array}{c|c|c} \boldsymbol{Q} & -\boldsymbol{A}^{\mathrm{T}} & \boldsymbol{0} \\ \hline \boldsymbol{A} & \boldsymbol{0} & -\boldsymbol{I} \end{array}\right] = \left[\begin{array}{cc|cccc|cccc} 1 & -1 & 1 & -2 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & 1 & -3 & 0 & -1 & 0 & 0 & 0 & 0 \\ \hline -1 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{array}\right]$$

Therefore the simplex tableau for Wolfe's method is:

x_1	x_2	λ_1	λ_2	λ_3	λ_4	y_1	y_2	y_3	y_4	w_1	w_2	w_3	w_4	w_5	w_6	W	RHS
0	0	0	0	0	0	0	0	0	0	−1	−1	−1	−1	−1	−1	1	0
1	−1	1	−2	−1	0	0	0	0	0	1	0	0	0	0	0	0	1
−1	2	1	−3	0	−1	0	0	0	0	0	1	0	0	0	0	0	1
−1	−1	0	0	0	0	−1	0	0	0	0	0	1	0	0	0	0	−3
2	3	0	0	0	0	0	−1	0	0	0	0	0	1	0	0	0	6
1	0	0	0	0	0	0	0	−1	0	0	0	0	0	1	0	0	1
0	1	0	0	0	0	0	0	0	−1	0	0	0	0	0	1	0	1

The simplex method requires that the RHS be non-negative, so that both sides of equality in the 3rd constraint should be multiplied by -1.Thus we have

x_1	x_2	λ_1	λ_2	λ_3	λ_4	y_1	y_2	y_3	y_4	w_1	w_2	w_3	w_4	w_5	w_6	W	RHS
0	0	0	0	0	0	0	0	0	0	−1	−1	−1	−1	−1	−1	1	0
1	−1	1	−2	−1	0	0	0	0	0	1	0	0	0	0	0	0	1
−1	2	1	−3	0	−1	0	0	0	0	0	1	0	0	0	0	0	1
1	1	0	0	0	0	1	0	0	0	0	0	−1	0	0	0	0	3
2	3	0	0	0	0	0	−1	0	0	0	0	0	1	0	0	0	6
1	0	0	0	0	0	0	0	−1	0	0	0	0	0	1	0	0	1
0	1	0	0	0	0	0	0	0	−1	0	0	0	0	0	1	0	1

Adjusting the objective row via, Row 0 ← Row 0 + $\sum_{j,j\neq 3}$(Row j) yields the final starting tableau below; now we can proceed with the Simplex method (with RBE)

x_1	x_2	λ_1	λ_2	λ_3	λ_4	y_1	y_2	y_3	y_4	w_1	w_2	w_3	w_4	w_5	w_6	W	RHS
4	6	2	−5	−1	−1	0	−1	−1	−1	0	0	0	0	0	0	0	11
1	−1	1	−2	−1	0	0	0	0	0	1	0	0	0	0	0	0	1
−1	2	1	−3	0	−1	0	0	0	0	0	1	0	0	0	0	0	1
1	1	0	0	0	0	1	0	0	0	0	0	−1	0	0	0	0	3
2	3	0	0	0	0	0	−1	0	0	0	0	0	1	0	0	0	6
1	0	0	0	0	0	0	0	−1	0	0	0	0	0	1	0	0	1
0	1	0	0	0	0	0	0	0	−1	0	0	0	0	0	1	0	1

BEALE'S ALGORITHM FOR QP

$$\text{Minimize } \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, \quad \text{st } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \geq \boldsymbol{0}.$$

Let B≡Basic; N≡Nonbasic, so that $\boldsymbol{A}^{\boldsymbol{B}}\boldsymbol{x}_{\boldsymbol{B}} + \boldsymbol{A}^{N}\boldsymbol{x}_{N} = \boldsymbol{b}$

$\boldsymbol{x}^{\mathbf{0}} = \left[\boldsymbol{x}_{\boldsymbol{B}}^{\mathbf{0}} \mid \boldsymbol{x}_{\boldsymbol{N}}^{\mathbf{0}}\right]^{\mathrm{T}} =$ initial basic feasible solution, and

$\boldsymbol{A} = [\boldsymbol{A}^{\boldsymbol{B}} \mid \boldsymbol{A}^{N}]$ where $\boldsymbol{A}^{\boldsymbol{B}}$ has columns corresponding to the basic variables $\boldsymbol{x}_{\boldsymbol{B}}$ and is nonsingular. If $\boldsymbol{x}_N = \mathbf{0}$, then $\boldsymbol{x}_{\boldsymbol{B}} = [\boldsymbol{A}^{\boldsymbol{B}}]^{-1}\boldsymbol{b}$.

In general, $\boldsymbol{x}_{\boldsymbol{B}} = [\boldsymbol{A}^{\boldsymbol{B}}]^{-1}\boldsymbol{b} - [\boldsymbol{A}^{\boldsymbol{B}}]^{-1}\boldsymbol{A}^{N}\boldsymbol{x}_N$ for $\boldsymbol{x}$ that satisfies $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$.

Let us also partition $\boldsymbol{c}$ and $\boldsymbol{Q}$ as

$$\boldsymbol{c} = [\boldsymbol{c}_{\boldsymbol{B}} \mid \boldsymbol{c}_{\boldsymbol{N}}]^{\mathrm{T}} \text{ and } \boldsymbol{Q} = \left[\begin{array}{c|c} \boldsymbol{Q}_{\boldsymbol{B}}^{\boldsymbol{B}} & \boldsymbol{Q}_{\boldsymbol{B}}^{\boldsymbol{N}} \\ \hline \boldsymbol{Q}_{\boldsymbol{N}}^{\boldsymbol{B}} & \boldsymbol{Q}_{\boldsymbol{N}}^{\boldsymbol{N}} \end{array}\right] \text{ where } \boldsymbol{Q}_{\boldsymbol{B}}^{\boldsymbol{N}} = [\boldsymbol{Q}_{\boldsymbol{N}}^{\boldsymbol{B}}]^{\mathrm{T}}$$

EXAMPLE: Min $\frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 + x_3 + 2x_1x_2 + 2x_1x_3 + x_2x_3 + 2x_1 + x_2 + x_3$

$$= \frac{1}{2}[x_1 \;\; x_2 \;\; x_3] \underbrace{\begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \end{bmatrix}}_{\boldsymbol{Q}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{[2 \;\; 1 \;\; 1]}_{\boldsymbol{c}^{\mathrm{T}}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

st

$$\left.\begin{array}{l} x_1 + x_3 = 8 \\ x_2 + x_3 = 10 \end{array}\right\} \Rightarrow \underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}}_{\boldsymbol{A}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 8 \\ 10 \end{bmatrix}}_{\boldsymbol{b}}, \quad \boldsymbol{x} \geq \mathbf{0}$$

An initial BFS might be x_1=8, x_2=10, x_3=0, with the objective equal to 268 so that

$$\boldsymbol{x^0} = [\boldsymbol{x_B^0} \mid \boldsymbol{x_N^0}]^{\mathrm{T}} = [x_1^0 \;\; x_2^0 \mid x_3^0]^{\mathrm{T}} = [8 \;\; 10 \mid 0]^{\mathrm{T}}$$

$$\boldsymbol{A} = [\boldsymbol{A^B} \mid \boldsymbol{A^N}] = \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right]; \boldsymbol{c} = [\boldsymbol{c_B} \mid \boldsymbol{c_N}]^{\mathrm{T}} = [2 \;\; 1 \mid 1]^{\mathrm{T}}$$

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q_B^B} & \boldsymbol{Q_B^N} \\ \boldsymbol{Q_N^B} & \boldsymbol{Q_N^N} \end{bmatrix} = \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 2 & 1 & 1 \\ \hline 2 & 1 & 2 \end{array}\right]$$

Note that $[\boldsymbol{A^B}]^{-1}\boldsymbol{b} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 8 \\ 10 \end{bmatrix} = \begin{bmatrix} 8 \\ 10 \end{bmatrix} = \boldsymbol{x_B}$

$$[\boldsymbol{A^B}]^{-1}\boldsymbol{A^N}\boldsymbol{x_N} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} [0] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The objective function is $f(\boldsymbol{x}) = ½\, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Qx} + \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}$

$$= \frac{\mathbf{1}}{\mathbf{2}}[\boldsymbol{x_B^{\mathrm{T}}} \mid \boldsymbol{x_N^{\mathrm{T}}}]\left[\begin{array}{c|c} \boldsymbol{Q_B^B} & \boldsymbol{Q_B^N} \\ \hline \boldsymbol{Q_N^B} & \boldsymbol{Q_N^N} \end{array}\right]\left[\begin{array}{c} \boldsymbol{x_B} \\ \hline \boldsymbol{x_N} \end{array}\right] + [\boldsymbol{c_B^{\mathrm{T}}} \mid \boldsymbol{c_N^{\mathrm{T}}}]\left[\begin{array}{c} \boldsymbol{x_B} \\ \hline \boldsymbol{x_N} \end{array}\right]$$

$$= \frac{\mathbf{1}}{\mathbf{2}}[\boldsymbol{x_B^{\mathrm{T}}} \mid \boldsymbol{x_N^{\mathrm{T}}}]\left[\begin{array}{c} \boldsymbol{Q_B^B x_B} + \boldsymbol{Q_B^N x_N} \\ \hline \boldsymbol{Q_N^B x_B} + \boldsymbol{Q_N^N x_N} \end{array}\right] + \boldsymbol{c_B x_B} + \boldsymbol{c_N x_N}$$

$$= \frac{\mathbf{1}}{\mathbf{2}}[\boldsymbol{x_B^{\mathrm{T}} Q_B^B x_B} + \boldsymbol{x_B^{\mathrm{T}} Q_B^N x_N} + \boldsymbol{x_N^{\mathrm{T}} Q_N^B x_B} + \boldsymbol{x_N^{\mathrm{T}} Q_N^N x_N}] + \boldsymbol{c_B x_B} + \boldsymbol{c_N x_N}$$

Upon substituting $\boldsymbol{x_B} = [\boldsymbol{A}^{\mathrm{B}}]^{-1}\boldsymbol{b}$ - $[\boldsymbol{A}^{\mathrm{B}}]^{-1}\boldsymbol{A}^{\mathrm{N}}\boldsymbol{x}_N$ we obtain

$$f(\boldsymbol{x}) = \left\{\boldsymbol{c_B}(\boldsymbol{A^B})^{-1}\boldsymbol{b} + \frac{1}{2}[(\boldsymbol{A^B})^{-1}\boldsymbol{b}]^{\mathrm{T}}\boldsymbol{Q_B^B}(\boldsymbol{A^B})^{-1}\boldsymbol{b}\right\} +$$

$$\{[(\boldsymbol{A^B})^{-1}\boldsymbol{b}]^{\mathrm{T}}\boldsymbol{Q_B^N} - [(\boldsymbol{A^B})^{-1}\boldsymbol{b}]^{\mathrm{T}}\boldsymbol{Q_B^B}(\boldsymbol{A^B})^{-1}\boldsymbol{A^N} + \boldsymbol{c_N} - \boldsymbol{c_B^{\mathrm{T}}}(\boldsymbol{A^B})^{-1}\boldsymbol{A^N}\}\boldsymbol{x_N} +$$

$$\frac{1}{2}\boldsymbol{x_N^{\mathrm{T}}}\{\boldsymbol{Q_N^N} - \boldsymbol{Q_N^B}(\boldsymbol{A^B})^{-1}\boldsymbol{A^N} + [(\boldsymbol{A^B})^{-1}\boldsymbol{A^N}]^{\mathrm{T}}\boldsymbol{Q_B^B}(\boldsymbol{A^B})^{-1}\boldsymbol{A^N}\}\boldsymbol{x_N}$$

Thus $f(\boldsymbol{x}) = Z_0 + \boldsymbol{p}^{\mathrm{T}}\boldsymbol{x_N} + \frac{1}{2}\boldsymbol{x_N^{\mathrm{T}}}R\boldsymbol{x_N}$

Computing the partial derivatives, we have

$$\frac{\partial f}{\partial x_i} = p_i + \sum_{k\epsilon \boldsymbol{N}} r_{ik}x_k \text{ for } i\epsilon N$$

$$= p_i \text{ at } \boldsymbol{x_N} = \boldsymbol{0}.$$

Choose the most negative p_i, and increase the corresponding value of x_i until

(1) One of the basic variables becomes zero (cf. the Simplex method...), OR

(2) $\frac{\partial f}{\partial x_i} = p_i + r_{ii}x_i = 0, \qquad \text{i.e.}, x_i = \frac{-p_i}{r_{ii}}$

(this results in a nonbasic but feasible solution)

For our example,

$$\boldsymbol{p} = [8\ \ 10]\begin{bmatrix}2\\1\end{bmatrix} - [8\ \ 10]\begin{bmatrix}1 & 2\\2 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} + 1 - [2\ \ 1]\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = [-30]$$

$$\therefore \frac{\partial f}{\partial x_3} = -30$$

Increase x_3 from 0 until one of the basic variables goes to zero, i.e., by 8 units when x_1=0, x_2=2, x_3 =8 (don't need R now since $\boldsymbol{x_N} = \boldsymbol{0}$). The new objective function value is 92 (compared to 268 at the previous iteration).

At this point $\boldsymbol{x_B} = [x_2\ \ x_3 \mid x_1] = [2\ \ 8 \mid 0]$ etc., etc., etc…