We denote $f(\boldsymbol{x}) = -x_1x_2x_3$ and $h(\boldsymbol{x}) = x_1x_2+x_2x_3+x_3x_1-A/2$. We have $\nabla f(\boldsymbol{x}) = -[x_2x_3, x_1x_3, x_1x_2]^\top$ and $\nabla h(\boldsymbol{x}) = [x_2+x_3, x_1+x_3, x_1+x_2]^\top$. Note that all feasible points are regular in this case. By the Lagrange condition, the dimensions of the box with maximum volume satisfies

$$\begin{aligned} x_2x_3 - \lambda(x_2 + x_3) &= 0 \\ x_1x_3 - \lambda(x_1 + x_3) &= 0 \\ x_1x_2 - \lambda(x_1 + x_2) &= 0 \\ x_1x_2 + x_2x_3 + x_3x_1 &= \frac{A}{2}, \end{aligned}$$

where $\lambda \in \mathbb{R}$.

We now solve these equations. First, we show that that x_1, x_2, x_3, and λ are all nonzero. Suppose that $x_1 = 0$. By the constraints, we have $x_2x_3 = A/2$. However, the second and third equations in the Lagrange condition yield $\lambda x_2 = \lambda x_3 = 0$, which together with the first equation implies that $x_2x_3 = 0$. This contradicts the constraints. A similar argument applies to x_2 and x_3.

Next, suppose that $\lambda = 0$. Then, the sum of the three Lagrange equations gives $x_2x_3 + x_1x_3 + x_1x_2 = 0$, which contradicts the constraints.

We now solve for x_1, x_2, and x_3 in the Lagrange equations. First, multiply the first equation by x_1 and the second by x_2, and subtract one from the other. We arrive at $x_3\lambda(x_1 - x_2) = 0$. Because neither x_3 nor λ can be zero (by part b), we conclude that $x_1 = x_2$. We similarly deduce that $x_2 = x_3$. From the constraint equation, we obtain $x_1 = x_2 = x_3 = \sqrt{A/6}$.

Notice that we have ignored the constraints that x_1, x_2, and x_3 are positive so that we can solve the problem using Lagrange's theorem. However, there is only one solution to the Lagrange equations, and the solution is positive. Therefore, if a solution exists for the problem with positivity constraints on the variables x_1, x_2, and x_3, then this solution must necessarily be equal to the solution above obtained by ignoring the positivity constraints. ∎

Next we provide an example with a quadratic objective function and a quadratic constraint.

Example 20.7 Consider the problem of extremizing the objective function

$$f(\boldsymbol{x}) = x_1^2 + x_2^2$$

on the ellipse

$$\{[x_1, x_2]^\top : h(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1 = 0\}.$$

We have

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= [2x_1, 2x_2]^\top, \\ \nabla h(\boldsymbol{x}) &= [2x_1, 4x_2]^\top. \end{aligned}$$

Thus,

$$D_x l(\boldsymbol{x}, \lambda) = D_x[f(\boldsymbol{x}) + \lambda h(\boldsymbol{x})] = [2x_1 + 2\lambda x_1, 2x_2 + 4\lambda x_2]$$

and

$$D_\lambda l(\boldsymbol{x}, \lambda) = h(\boldsymbol{x}) = x_1^2 + 2x_2^2 - 1.$$

Setting $D_x l(\boldsymbol{x}, \lambda) = \mathbf{0}^\top$ and $D_\lambda l(\boldsymbol{x}, \lambda) = 0$, we obtain three equations in three unknowns

$$\begin{aligned} 2x_1 + 2\lambda x_1 &= 0, \\ 2x_2 + 4\lambda x_2 &= 0, \\ x_1^2 + 2x_2^2 &= 1. \end{aligned}$$

All feasible points in this problem are regular. From the first of the equations above, we get either $x_1 = 0$ or $\lambda = -1$. For the case where $x_1 = 0$, the second and third equations imply that $\lambda = -1/2$ and $x_2 = \pm 1/\sqrt{2}$. For the case where $\lambda = -1$, the second and third equations imply that $x_1 = \pm 1$ and $x_2 = 0$. Thus, the points that satisfy the Lagrange condition for extrema are

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \boldsymbol{x}^{(2)} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \end{bmatrix}, \quad \boldsymbol{x}^{(3)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{x}^{(4)} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Because

$$f(\boldsymbol{x}^{(1)}) = f(\boldsymbol{x}^{(2)}) = \frac{1}{2}$$

and

$$f(\boldsymbol{x}^{(3)}) = f(\boldsymbol{x}^{(4)}) = 1$$

we conclude that if there are minimizers, then they are located at $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$, and if there are maximizers, then they are located at $\boldsymbol{x}^{(3)}$ and $\boldsymbol{x}^{(4)}$. It turns out that, indeed, $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ are minimizers and $\boldsymbol{x}^{(3)}$ and $\boldsymbol{x}^{(4)}$ are maximizers. This problem can be solved graphically, as illustrated in Figure 20.14. ∎

In the example above, both the objective function f and the constraint function h are quadratic functions. In the next example we take a closer look at a class of problems where both the objective function f and the constraint h are quadratic functions of n variables.

Example 20.8 Consider the following problem:

$$\text{maximize} \quad \frac{\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x}},$$

where $\boldsymbol{Q} = \boldsymbol{Q}^\top \geq 0$ and $\boldsymbol{P} = \boldsymbol{P}^\top > 0$. Note that if a point $\boldsymbol{x} = [x_1, \ldots, x_n]^\top$ is a solution to the problem, then so is any nonzero scalar multiple of it,

$$t\boldsymbol{x} = [tx_1, \ldots, tx_n]^\top, \quad t \neq 0.$$

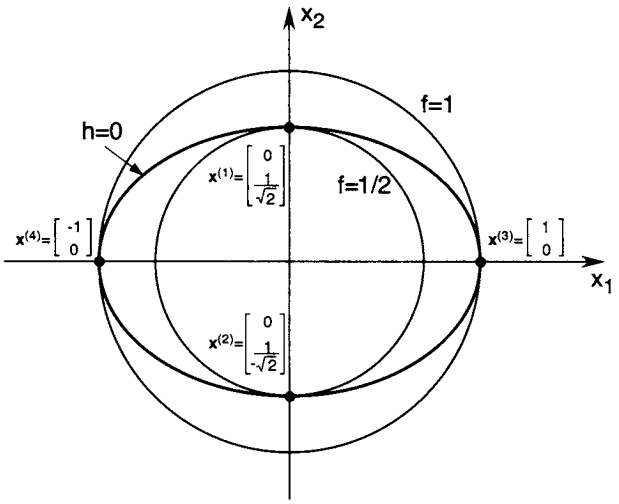


Figure 20.14 Graphical solution of the problem in Example 20.7.

Indeed,

$$\frac{(t\boldsymbol{x})^\top \boldsymbol{Q}(t\boldsymbol{x})}{(t\boldsymbol{x})^\top \boldsymbol{P}(t\boldsymbol{x})} = \frac{t^2 \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}{t^2 \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}} = \frac{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}}.$$

Therefore, to avoid the multiplicity of solutions, we further impose the constraint

$$\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 1.$$

The optimization problem becomes

$$\begin{aligned} &\text{maximize} \quad \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ &\text{subject to} \quad \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 1. \end{aligned}$$

Let us write

$$\begin{aligned} f(\boldsymbol{x}) &= \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}, \\ h(\boldsymbol{x}) &= 1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}. \end{aligned}$$

Any feasible point for this problem is regular (see Exercise 20.13). We now apply Lagrange's method. We first form the Lagrangian function

$$l(\boldsymbol{x}, \lambda) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \lambda(1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}).$$

Applying the Lagrange condition yields

$$\begin{aligned} D_x l(\boldsymbol{x}, \lambda) &= 2\boldsymbol{x}^\top \boldsymbol{Q} - 2\lambda \boldsymbol{x}^\top \boldsymbol{P} = \boldsymbol{0}^\top, \\ D_\lambda l(\boldsymbol{x}, \lambda) &= 1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 0. \end{aligned}$$

The first of the equations above can be represented as

$$\boldsymbol{Q}\boldsymbol{x} - \lambda \boldsymbol{P}\boldsymbol{x} = \boldsymbol{0}$$

or

$$(\lambda \boldsymbol{P} - \boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}.$$

This representation is possible because $\boldsymbol{P} = \boldsymbol{P}^\top$ and $\boldsymbol{Q} = \boldsymbol{Q}^\top$. By assumption $\boldsymbol{P} > 0$, hence $\boldsymbol{P}^{-1}$ exists. Premultiplying $(\lambda \boldsymbol{P} - \boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}$ by $\boldsymbol{P}^{-1}$, we obtain

$$(\lambda \boldsymbol{I}_n - \boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}$$

or, equivalently,

$$\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{x} = \lambda \boldsymbol{x}.$$

Therefore, the solution, if it exists, is an eigenvector of $\boldsymbol{P}^{-1}\boldsymbol{Q}$, and the Lagrange multiplier is the corresponding eigenvalue. As usual, let $\boldsymbol{x}^*$ and λ^* be the optimal solution. Because $\boldsymbol{x}^{*\top}\boldsymbol{P}\boldsymbol{x}^* = 1$ and $\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{x}^* = \lambda^*\boldsymbol{x}^*$, we have

$$\lambda^* = \boldsymbol{x}^{*\top}\boldsymbol{Q}\boldsymbol{x}^*.$$

Hence, λ^* is the maximum of the objective function, and therefore is, in fact, the maximal eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$. It is also called the maximal *generalized eigenvalue.* ∎

In the problems above, we are able to find points that are candidates for extremizers of the given objective function subject to equality constraints. These critical points are the only candidates because they are the only points that satisfy the Lagrange condition. To classify such critical points as minimizers, maximizers, or neither, we need a stronger condition—possibly a necessary and sufficient condition. In the next section we discuss a second-order necessary condition and a second-order sufficient condition for minimizers.

20.5 Second-Order Conditions

We assume that $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable: $f, \boldsymbol{h} \in \mathcal{C}^2$. Let

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^\top \boldsymbol{h}(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda_1 h_1(\boldsymbol{x}) + \cdots + \lambda_m h_m(\boldsymbol{x})$$

be the Lagrangian function. Let $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ be the Hessian matrix of $l(\boldsymbol{x}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{x}$:

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{F}(\boldsymbol{x}) + \lambda_1 \boldsymbol{H}_1(\boldsymbol{x}) + \cdots + \lambda_m \boldsymbol{H}_m(\boldsymbol{x}),$$

where $\boldsymbol{F}(\boldsymbol{x})$ is the Hessian matrix of f at $\boldsymbol{x}$ and $\boldsymbol{H}_k(\boldsymbol{x})$ is the Hessian matrix of h_k at $\boldsymbol{x}$, $k = 1, \ldots, m$, given by

$$\boldsymbol{H}_k(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 h_k}{\partial x_1^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 h_k}{\partial x_n \partial x_1}(\boldsymbol{x}) \\ \vdots & & \vdots \\ \frac{\partial^2 h_k}{\partial x_1 \partial x_n}(\boldsymbol{x}) & \cdots & \frac{\partial^2 h_k}{\partial^2 x_n}(\boldsymbol{x}) \end{bmatrix}.$$

We introduce the notation $[\boldsymbol{\lambda} \boldsymbol{H}(\boldsymbol{x})]$:

$$[\boldsymbol{\lambda} \boldsymbol{H}(\boldsymbol{x})] = \lambda_1 \boldsymbol{H}_1(\boldsymbol{x}) + \cdots + \lambda_m \boldsymbol{H}_m(\boldsymbol{x}).$$

Using the notation above, we can write

$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{F}(\boldsymbol{x}) + [\boldsymbol{\lambda} \boldsymbol{H}(\boldsymbol{x})].$$

Theorem 20.4 ***Second-Order Necessary Conditions.*** *Let $\boldsymbol{x}^*$ be a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$, and $f, \boldsymbol{h} \in \mathcal{C}^2$. Suppose that $\boldsymbol{x}^*$ is regular. Then, there exists $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that:*

1. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$.
2. *For all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, we have $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{y} \geq 0$.* □

Proof. The existence of $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$ follows from Lagrange's theorem. It remains to prove the second part of the result. Suppose that $\boldsymbol{y} \in T(\boldsymbol{x}^*)$; that is, $\boldsymbol{y}$ belongs to the tangent space to $S = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ at $\boldsymbol{x}^*$. Because $\boldsymbol{h} \in \mathcal{C}^2$, following the argument of Theorem 20.1, there exists a twice-differentiable curve $\{\boldsymbol{x}(t) : t \in (a, b)\}$ on S such that

$$\boldsymbol{x}(t^*) = \boldsymbol{x}^*, \quad \dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$$

for some $t^* \in (a, b)$. Observe that by assumption, t^* is a local minimizer of the function $\phi(t) = f(\boldsymbol{x}(t))$. From the second-order necessary condition for unconstrained minimization (see Theorem 6.2), we obtain

$$\frac{d^2\phi}{dt^2}(t^*) \geq 0.$$

Using the formula

$$\frac{d}{dt}(\boldsymbol{y}(t)^\top \boldsymbol{z}(t)) = \boldsymbol{z}(t)^\top \frac{d\boldsymbol{y}}{dt}(t) + \boldsymbol{y}(t)^\top \frac{d\boldsymbol{z}}{dt}(t)$$

and applying the chain rule yields

$$\begin{aligned} \frac{d^2\phi}{dt^2}(t^*) &= \frac{d}{dt}[Df(\boldsymbol{x}(t^*))\dot{\boldsymbol{x}}(t^*)] \\ &= \dot{\boldsymbol{x}}(t^*)^\top \boldsymbol{F}(\boldsymbol{x}^*)\dot{\boldsymbol{x}}(t^*) + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) \\ &= \boldsymbol{y}^\top \boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{y} + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) \geq 0. \end{aligned}$$

Because $\boldsymbol{h}(\boldsymbol{x}(t)) = \boldsymbol{0}$ for all $t \in (a, b)$, we have

$$\frac{d^2}{dt^2} \boldsymbol{\lambda}^{*\top} \boldsymbol{h}(\boldsymbol{x}(t)) = 0.$$

Thus, for all $t \in (a, b)$,

$$\begin{aligned}
\frac{d^2}{dt^2}\boldsymbol{\lambda}^{*\top}\boldsymbol{h}(\boldsymbol{x}(t)) &= \frac{d}{dt}\left[\boldsymbol{\lambda}^{*\top}\frac{d}{dt}\boldsymbol{h}(\boldsymbol{x}(t))\right] \\
&= \frac{d}{dt}\left[\sum_{k=1}^{m}\lambda_k^*\frac{d}{dt}h_k(\boldsymbol{x}(t))\right] \\
&= \frac{d}{dt}\left[\sum_{k=1}^{m}\lambda_k^* Dh_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t)\right] \\
&= \sum_{k=1}^{m}\lambda_k^*\frac{d}{dt}(Dh_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t)) \\
&= \sum_{k=1}^{m}\lambda_k^*\left[\dot{\boldsymbol{x}}(t)^\top \boldsymbol{H}_k(\boldsymbol{x}(t))\dot{\boldsymbol{x}}(t) + Dh_k(\boldsymbol{x}(t))\ddot{\boldsymbol{x}}(t)\right] \\
&= \dot{\boldsymbol{x}}^\top(t)[\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}(t))]\dot{\boldsymbol{x}}(t) + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}(t))\ddot{\boldsymbol{x}}(t) \\
&= 0.
\end{aligned}$$

In particular, the above is true for $t = t^*$; that is,

$$\boldsymbol{y}^\top[\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)]\boldsymbol{y} + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) = 0.$$

Adding this equation to the inequality

$$\boldsymbol{y}^\top\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{y} + Df(\boldsymbol{x}^*)\ddot{\boldsymbol{x}}(t^*) \geq 0$$

yields

$$\boldsymbol{y}^\top\left(\boldsymbol{F}(\boldsymbol{x}^*) + [\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)]\right)\boldsymbol{y} + (Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*))\ddot{\boldsymbol{x}}(t^*) \geq 0.$$

But, by Lagrange's theorem, $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$. Therefore,

$$\boldsymbol{y}^\top\left(\boldsymbol{F}(\boldsymbol{x}^*) + [\boldsymbol{\lambda}^*\boldsymbol{H}(\boldsymbol{x}^*)]\right)\boldsymbol{y} = \boldsymbol{y}^\top\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} \geq 0,$$

which proves the result. ∎

Observe that $\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ plays a similar role as the Hessian matrix $\boldsymbol{F}(\boldsymbol{x})$ of the objective function f did in the unconstrained minimization case. However, we now require that $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \geq 0$ only on $T(\boldsymbol{x}^*)$ rather than on $\mathbb{R}^n$.

The conditions above are necessary, but not sufficient, for a point to be a local minimizer. We now present, without a proof, sufficient conditions for a point to be a strict local minimizer.

Theorem 20.5 ***Second-Order Sufficient Conditions.*** *Suppose that $f, \boldsymbol{h} \in \mathcal{C}^2$ and there exists a point $\boldsymbol{x}^* \in \mathbb{R}^n$ and $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ such that:*

1. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*\top}D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^\top$.

2. *For all* $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, $\boldsymbol{y} \neq \boldsymbol{0}$, *we have* $\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} > 0$.

Then, $\boldsymbol{x}^*$ *is a strict local minimizer of* f *subject to* $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$. □

Proof. The interested reader can consult [88, p. 334] for a proof of this result. ∎

Theorem 20.5 states that if an $\boldsymbol{x}^*$ satisfies the Lagrange condition, and $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is positive definite on $T(\boldsymbol{x}^*)$, then $\boldsymbol{x}^*$ is a strict local minimizer. A similar result to Theorem 20.5 holds for a strict local maximizer, the only difference being that $\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ be negative definite on $T(\boldsymbol{x}^*)$. We illustrate this condition in the following example.

Example 20.9 Consider the following problem:

$$\text{maximize} \quad \frac{\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}},$$

where

$$\boldsymbol{Q} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{P} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

As pointed out earlier, we can represent this problem in the equivalent form

$$\begin{aligned} &\text{maximize} && \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ &\text{subject to} && \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 1. \end{aligned}$$

The Lagrangian function for the transformed problem is given by

$$l(\boldsymbol{x}, \lambda) = \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \lambda(1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x}).$$

The Lagrange condition yields

$$(\lambda \boldsymbol{I} - \boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0},$$

where

$$\boldsymbol{P}^{-1}\boldsymbol{Q} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

There are only two values of λ that satisfy $(\lambda \boldsymbol{I} - \boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}$, namely, the eigenvalues of $\boldsymbol{P}^{-1}\boldsymbol{Q}$: $\lambda_1 = 2$, $\lambda_2 = 1$. We recall from our previous discussion of this problem that the Lagrange multiplier corresponding to the solution is the maximum eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$, namely, $\lambda^* = \lambda_1 = 2$. The corresponding eigenvector is the maximizer—the solution to the problem. The eigenvector corresponding to the eigenvalue $\lambda^* = 2$ satisfying the constraint $\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 1$ is $\pm \boldsymbol{x}^*$, where

$$\boldsymbol{x}^* = \left[\frac{1}{\sqrt{2}}, 0\right]^\top.$$

At this point, all we have established is that the pairs $(\pm\boldsymbol{x}^*, \lambda^*)$ satisfy the Lagrange condition. We now show that the points $\pm\boldsymbol{x}^*$ are, in fact, strict local maximizers. We do this for the point $\boldsymbol{x}^*$. A similar procedure applies to $-\boldsymbol{x}^*$. We first compute the Hessian matrix of the Lagrangian function. We have

$$\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) = 2\boldsymbol{Q} - 2\lambda\boldsymbol{P} = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}.$$

The tangent space $T(\boldsymbol{x}^*)$ to $\{\boldsymbol{x} : 1 - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} = 0\}$ is

$$\begin{aligned} T(\boldsymbol{x}^*) &= \{\boldsymbol{y} \in \mathbb{R}^2 : \boldsymbol{x}^{*\top}\boldsymbol{P}\boldsymbol{y} = 0\} \\ &= \{\boldsymbol{y} : [\sqrt{2}, 0]\boldsymbol{y} = 0\} \\ &= \{\boldsymbol{y} : \boldsymbol{y} = [0, a]^\top,\ a \in \mathbb{R}\}. \end{aligned}$$

Note that for each $\boldsymbol{y} \in T(\boldsymbol{x}^*)$, $\boldsymbol{y} \neq \boldsymbol{0}$,

$$\boldsymbol{y}^\top \boldsymbol{L}(\boldsymbol{x}^*, \lambda^*)\boldsymbol{y} = [0, a] \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 0 \\ a \end{bmatrix} = -2a^2 < 0.$$

Hence, $\boldsymbol{L}(\boldsymbol{x}^*, \lambda^*) < 0$ on $T(\boldsymbol{x}^*)$, and thus $\boldsymbol{x}^* = [1/\sqrt{2}, 0]^\top$ is a strict local maximizer. The same is true for the point $-\boldsymbol{x}^*$. Note that

$$\frac{\boldsymbol{x}^{*\top}\boldsymbol{Q}\boldsymbol{x}^*}{\boldsymbol{x}^{*\top}\boldsymbol{P}\boldsymbol{x}^*} = 2,$$

which, as expected, is the value of the maximal eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$. Finally, we point out that any scalar multiple $t\boldsymbol{x}^*$ of $\boldsymbol{x}^*$, $t \neq 0$, is a solution to the original problem of maximizing $\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{x}/\boldsymbol{x}^\top\boldsymbol{P}\boldsymbol{x}$. ∎

20.6 Minimizing Quadratics Subject to Linear Constraints

Consider the problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\boldsymbol{x}^\top\boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \end{aligned}$$

where $\boldsymbol{Q} > 0$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m < n$, $\operatorname{rank}\boldsymbol{A} = m$. This problem is a special case of what is called a *quadratic programming problem* (the general form of a quadratic programming problem includes the constraint $\boldsymbol{x} \geq \boldsymbol{0}$). Note that the constraint set contains an infinite number of points (see Section 2.3). We now show, using Lagrange's theorem, that there is a unique solution to the optimization problem above. Following that, we provide an example illustrating the application of this solution to an optimal control problem.