Chapter 20 Problems with Equality Constraints

An Introduction to Optimization
Spring, 2014

Wei-Ta Chu

Introduction

- Solve a class of nonlinear constrained optimization problems that can be formulated as

$$\begin{aligned} &\text{minimize } f(\boldsymbol{x}) \\ &\text{subject to } h_i(\boldsymbol{x}) = 0, i = 1, \dots, m \\ &\qquad\qquad g_j(\boldsymbol{x}) \leq 0, j = 1, \dots, p \end{aligned}$$

where $\boldsymbol{x} \in R^n$, $f : R^n \rightarrow R$, $h_i : R^n \rightarrow R$, $g_j : R^n \rightarrow R$, and $m \leq n$. In vector notation, the problem above can be represented in the following *standard form*:

$$\begin{aligned} &\text{minimize } f(\boldsymbol{x}) \\ &\text{subject to } \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \\ &\qquad\qquad \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0} \end{aligned}$$

where $\boldsymbol{h} : R^n \rightarrow R^m$ and $\boldsymbol{g} : R^n \rightarrow R^p$.

Introduction

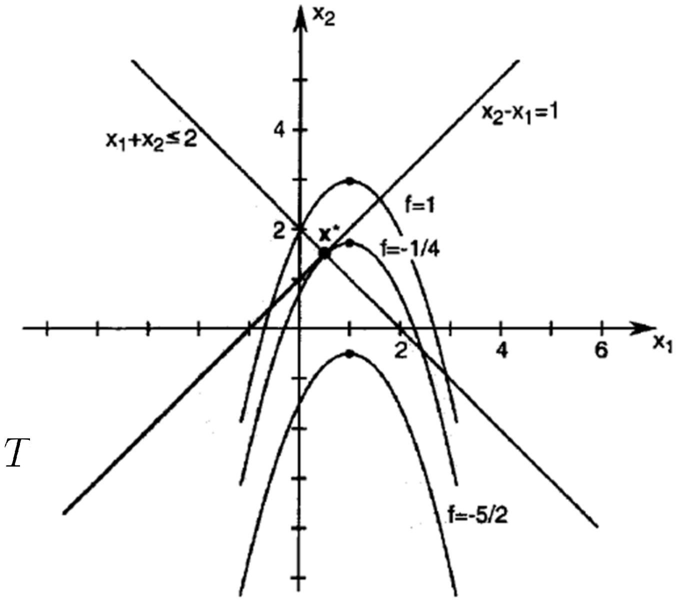
- Definition 20.1. Any point satisfying the constraints is called a *feasible point*. The set of feasible points,
$$\{\boldsymbol{x} \in R^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$$
is called a *feasible set*.
- Actually, linear programming problems have been studied.
$$\begin{aligned} \text{minimize } & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$
- For if we are confronted with a maximization problem, it can easily be transformed into the minimization problem by observing that
$$\text{maximize } f(\boldsymbol{x}) = \text{minimize } -f(\boldsymbol{x})$$

Example

- Consider the following optimization problem:

$$\begin{aligned} \text{minimize } & (x_1 - 1)^2 + x_2 + 2 \\ \text{subject to } & x_2 - x_1 = 1 \\ & x_1 + x_2 \leq 2 \end{aligned}$$

- This problem turns out to be simple enough to be solved graphically. (Figure 20.1)
- Feasible set: heavy solid line
- The inverted parabolas represent level sets of the objective function
- The minimizer lies on the level set with $f = -1/4$. The minimizer of the objective function is $\boldsymbol{x}^* = [1/2, 3/2]^T$



Problem Formulation

- The class of optimization problems we analyze in this chapter is

$$\text{minimize } f(\boldsymbol{x})$$
$$\text{subject to } \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}$$

where $\boldsymbol{x} \in R^n$, $f : R^n \to R$, $\boldsymbol{h} : R^n \to R^m$, $\boldsymbol{h} = [h_1, ..., h_m]^T$ and $m \leq n$. We assume that the function $\boldsymbol{h}$ is continuously differentiable, that is, $\boldsymbol{h} \in \mathcal{C}^1$.

- Definition 20.2. A point $\boldsymbol{x}^*$ satisfying the constraints $h_1(\boldsymbol{x}^*) = 0, ..., h_m(\boldsymbol{x}^*) = 0$ is said to be a *regular point* of the constraints if the gradient vectors $\triangledown h_1(\boldsymbol{x}^*), ..., \triangledown h_m(\boldsymbol{x}^*)$ are linearly independent.

Problem Formulation

- Let $D\boldsymbol{h}(\boldsymbol{x}^*)$ be the Jacobian matrix of $\boldsymbol{h} = [h_1, \ldots, h_m]^T$ at $\boldsymbol{x}^*$ given by

$$D\boldsymbol{h}(\boldsymbol{x}^*) = \begin{bmatrix} Dh_1(\boldsymbol{x}^*) \\ \vdots \\ Dh_m(\boldsymbol{x}^*) \end{bmatrix} = \begin{bmatrix} \nabla h_1(\boldsymbol{x}^*)^T \\ \vdots \\ \nabla h_m(\boldsymbol{x}^*)^T \end{bmatrix}$$

  Then, $\boldsymbol{x}^*$ is regular if and only if $rank(D\boldsymbol{h}(\boldsymbol{x}^*)) = m$ (i.e., the Jacobian matrix is of full rank).

- The set of equality constraint $h_1(\boldsymbol{x}) = 0, \ldots, h_m(\boldsymbol{x}) = 0$, $h_i : R^n \to R$, describes a surface

$$S = \{\boldsymbol{x} \in R^n : h_1(\boldsymbol{x}) = 0, \ldots, h_m(\boldsymbol{x}) = 0\}$$

- Assuming that the points in S are regular, the dimension of the surface S is $n - m$

Example

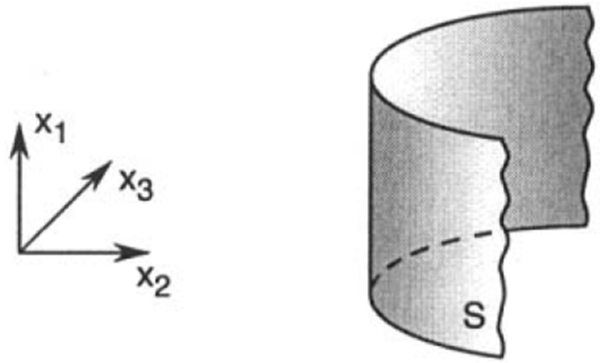
- Let *n*=3 and *m* = 1 (i.e., we are operating in R^3). Assuming that all points in *S* are regular, the set *S* is a two-dimensional surface. For example, let

$$h_1(\boldsymbol{x}) = x_2 - x_3^2 = 0$$

Note that $\nabla h_1(\boldsymbol{x}) = [0, 1, -2x_3]^T$, and hence for any $\boldsymbol{x} \in R^3$ $\nabla h_1(\boldsymbol{x}) \neq \mathbf{0}$. In this case,

$$\dim S = \dim\{\boldsymbol{x} : h_1(\boldsymbol{x}) = 0\} = n - m = 2$$

- Figure 20.2



$S=\{[x_1,x_2,x_3]^T : x_2-x_3^2=0\}$

Example

- Let n=3 and m=2. Assuming regularity, the feasible set S is a one-dimensional object (i.e., a curve in R^3). For example, let

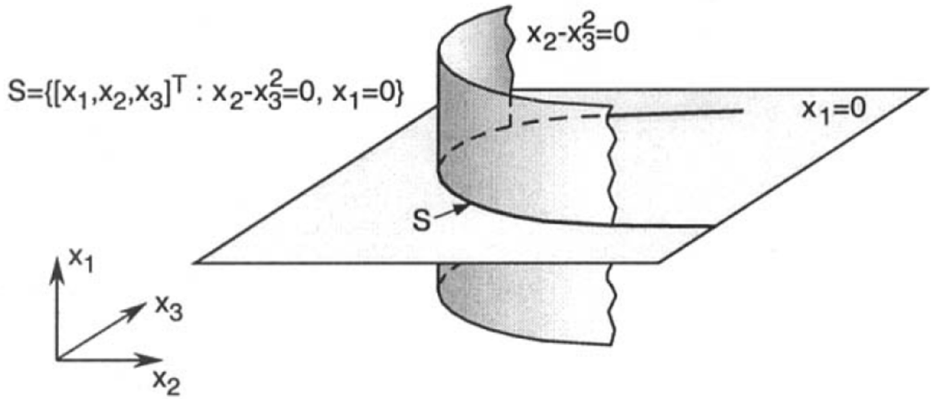
$$h_1(\boldsymbol{x}) = x_1$$

$$h_2(\boldsymbol{x}) = x_2 - x_3^2$$

In this case, $\nabla h_1(\boldsymbol{x}) = [1, 0, 0]^T$ and $\nabla h_2(\boldsymbol{x}) = [0, 1, -2x_3]^T$ Hence, the vectors $\nabla h_1(\boldsymbol{x})$ and $\nabla h_2(\boldsymbol{x})$ are linearly independent in R^3. Thus,

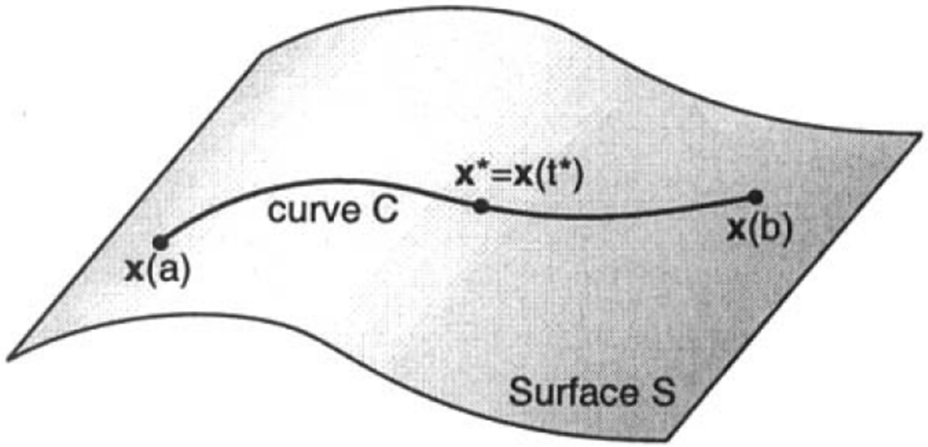
$$\dim S = \dim\{\boldsymbol{x} : h_1(\boldsymbol{x}) = 0, h_2(\boldsymbol{x}) = 0\} = n - m = 1$$

- Figure 20.3



Tangent and Normal Spaces

- Definition 20.3. A curve *C* on a surface *S* is a set of points $\{\boldsymbol{x}(t) \in S : t \in (a, b)\}$, continuously parameterized by $t \in (a, b)$ that is, $\boldsymbol{x} : (a, b) \to S$ is a continuous function.
- The definition of a curve implies that all the points on the curve satisfy the equation describing the surface. The curve *C* passes through a point $\boldsymbol{x}^*$ if there exists $t^* \in (a, b)$ such that $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$
- Figure 20. 4



Tangent and Normal Spaces

- Intuitively, we can think of a curve $C = \{\boldsymbol{x}(t) : t \in (a, b)\}$ as the path traversed by a point $\boldsymbol{x}$ traveling on the surface *S*. The position of the point as time *t* is given by $\boldsymbol{x}(t)$
- Definition 19.4. The curve $C = \{\boldsymbol{x}(t) : t \in (a, b)\}$ is differentiable if

$$\dot{\boldsymbol{x}}(t) = \frac{d\boldsymbol{x}}{dt}(t) = \begin{bmatrix} \dot{x}_1(t) \\ \vdots \\ \dot{x}_n(t) \end{bmatrix}$$

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

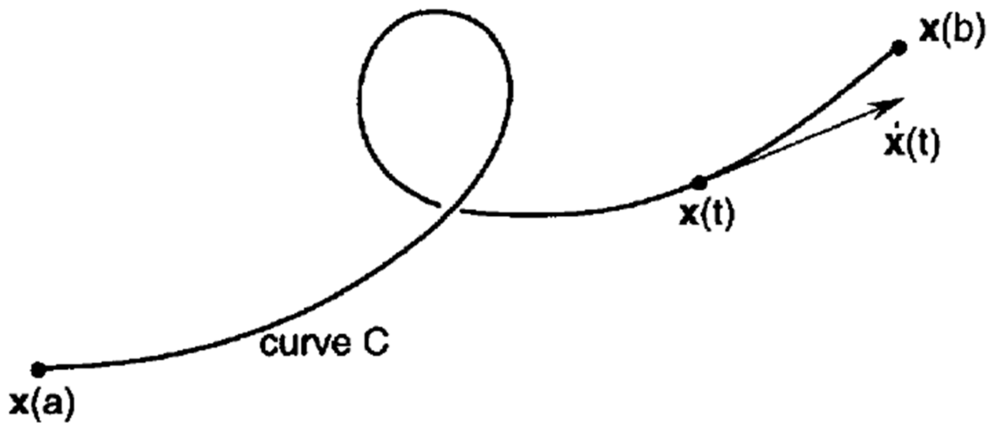
The curve $C = \{\boldsymbol{x}(t) : t \in (a, b)\}$ is twice differentiable if

$$\ddot{\boldsymbol{x}}(t) = \frac{d^2\boldsymbol{x}}{dt^2}(t) = \begin{bmatrix} \ddot{x}_1(t) \\ \vdots \\ \ddot{x}_n(t) \end{bmatrix}$$

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

Tangent and Normal Spaces

- Note that both $\dot{\boldsymbol{x}}(t)$ and $\ddot{\boldsymbol{x}}(t)$ are *n*-dimensional vectors. We can think of $\dot{\boldsymbol{x}}(t)$ and $\ddot{\boldsymbol{x}}(t)$ the velocity and acceleration, respectively, of a point traversing the curve *C* with position $\boldsymbol{x}(t)$ at time *t*. Therefore, the vector $\dot{\boldsymbol{x}}(t^*)$ is *tangent* to the curve *C* at $\boldsymbol{x}^*$
- We are now ready to introduce the notions of a tangent space. For this recall the set $S = \{\boldsymbol{x} \in R^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$, where $\boldsymbol{h} \in \mathcal{C}^1$. We think of *S* as a surface in R^n
- Figure 20.5



Tangent and Normal Spaces

- Definition 20.5. The *tangent space* at a point $\boldsymbol{x}^*$ on the surface $S = \{\boldsymbol{x} \in R^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ is the set

$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} : D\boldsymbol{h}(\boldsymbol{x}^*)\boldsymbol{y} = \boldsymbol{0}\}$$

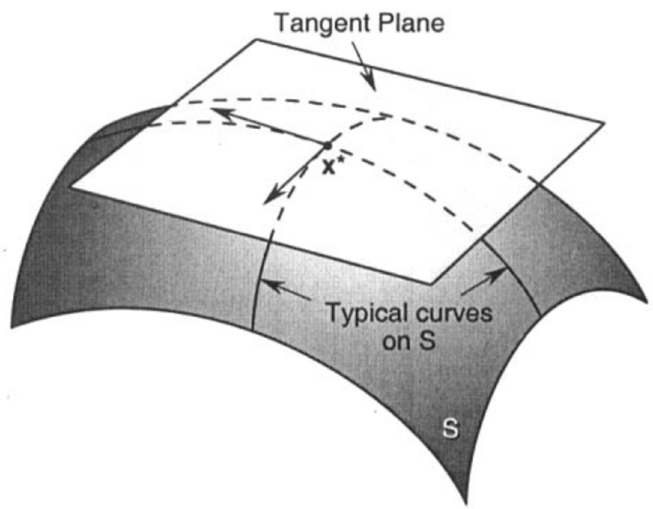
- Note that the tangent space $T(\boldsymbol{x}^*)$ is the nullspace of the matrix $D\boldsymbol{h}(\boldsymbol{x}^*)$: $T(\boldsymbol{x}^*) = \mathcal{N}(D\boldsymbol{h}(\boldsymbol{x}^*))$. The tangent space is therefore a subspace of R^n

Tangent and Normal Spaces

- Assuming that $\boldsymbol{x}^*$ is regular, the dimension of the tangent space is $n - m$, where m is the number of equality constraints $h_i(\boldsymbol{x}^*) = 0$. Note that the tangent space passes through the origin. However, it is often convenient to picture the tangent space as a plane that passes through the point $\boldsymbol{x}^*$. For this, we define the *tangent plane* at $\boldsymbol{x}^*$ to be the set

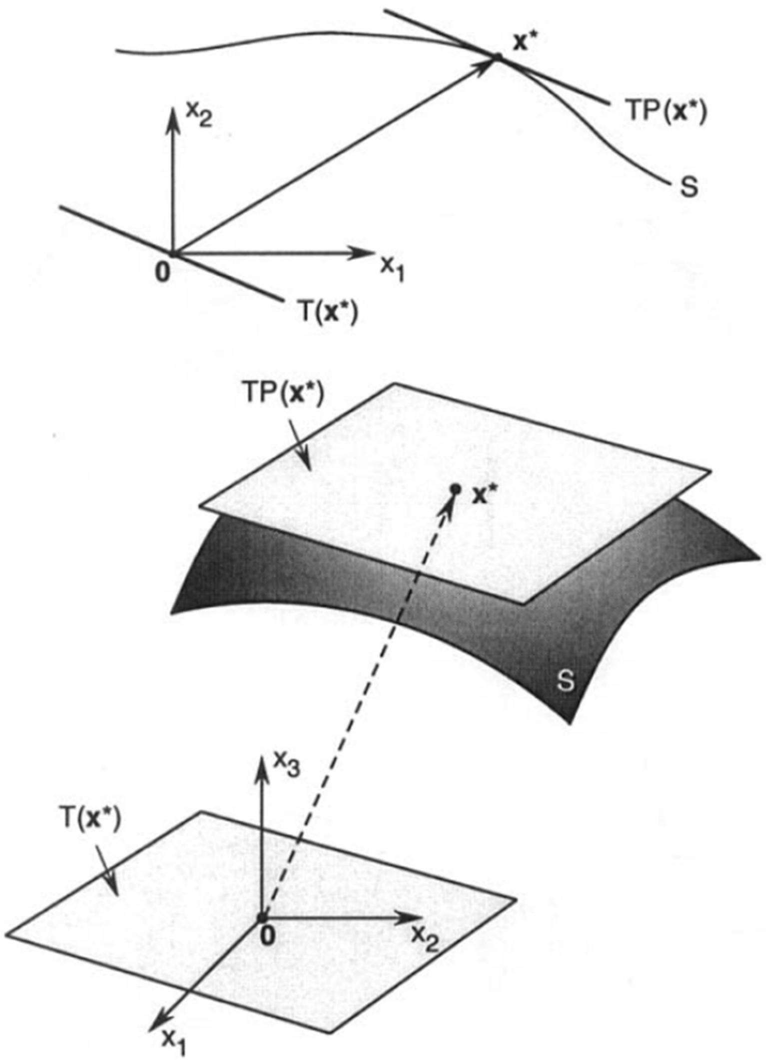
$$TP(\boldsymbol{x}^*) = T(\boldsymbol{x}^*) + \boldsymbol{x}^* = \{\boldsymbol{x} + \boldsymbol{x}^* : \boldsymbol{x} \in T(\boldsymbol{x}^*)\}$$

- Figure 20.6



Tangent and Normal Spaces

- Figure 20.7 illustrates the relationship between the tangent plane $TP(\boldsymbol{x}^*)$ and the tangent space $T(\boldsymbol{x}^*)$.



Example

- Let $S = \{\boldsymbol{x} \in R^3 : h_1(\boldsymbol{x}) = x_1 = 0, h_2(\boldsymbol{x}) = x_1 - x_2 = 0\}$
  Then, S is the x_3-axis in R^3 (Figure 20.8). We have

$$D\boldsymbol{h}(\boldsymbol{x}) = \begin{bmatrix} \nabla h_1(\boldsymbol{x}) \\ \nabla h_2(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix}$$

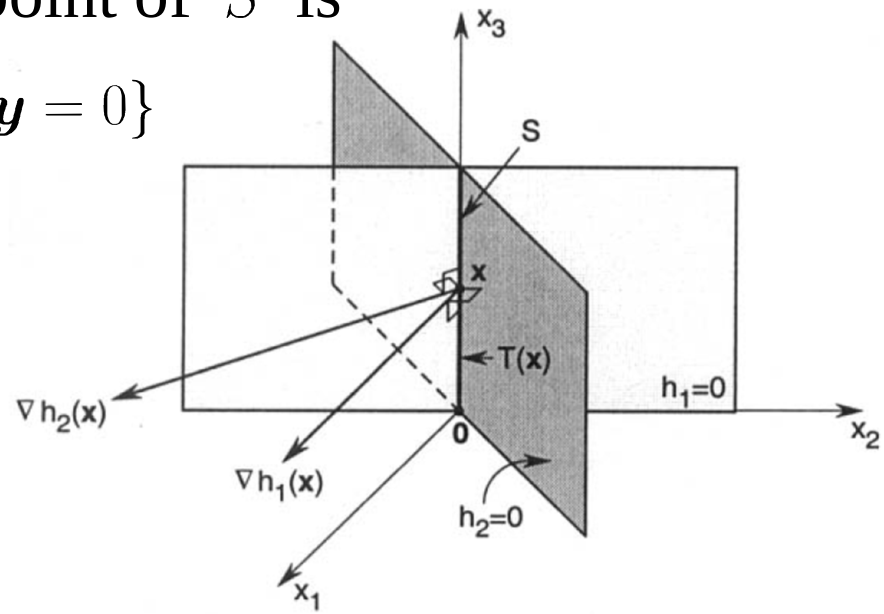
  Because ∇h_1 and ∇h_2 are linearly independent when evaluated at any $\boldsymbol{x} \in S$, all the points of S are regular. The tangent space at any arbitrary point of S is

$$T(\boldsymbol{x}) = \{\boldsymbol{y} : \nabla h_1(\boldsymbol{x})^T \boldsymbol{y} = 0, \nabla h_2(\boldsymbol{x})^T \boldsymbol{y} = 0\}$$

$$= \left\{ \boldsymbol{y} : \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \boldsymbol{0} \right\}$$

$$= \{[0, 0, \alpha]^T : \alpha \in R\}$$

$$= \text{the } x_3\text{-axis in } R^3$$



Example

- In the example, the tangent space $T(\boldsymbol{x})$ at any point $\boldsymbol{x} \in S$ is a one-dimensional subspace of R^3 .

- Intuitively, we would expect the definition of the tangent space at a point on a surface to be the collection of all "tangent vectors" to the surface at that point.
- We have seen that the derivative of a curve on a surface at a point is a tangent vector to the curve, and hence to the surface.
- The intuition above agrees with our definition whenever $\boldsymbol{x}^*$ is regular.

Tangent and Normal Spaces

- Theorem 20.1. Suppose that $\boldsymbol{x}^* \in S$ is a regular point and $T(\boldsymbol{x}^*)$ is the tangent space at $\boldsymbol{x}^*$. Then, $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ if and only if there exists a differentiable curve in S passing through $\boldsymbol{x}^*$ with derivative $\boldsymbol{y}$ at $\boldsymbol{x}^*$.
- Proof: $\Longleftarrow$: Suppose that there exists a curve $\{\boldsymbol{x}(t) : t \in (a,b)\}$ in S such that $\boldsymbol{x}(t^*) = \boldsymbol{x}^*$ and $\dot{\boldsymbol{x}}(t^*) = \boldsymbol{y}$ for some $t^* \in (a,b)$. Then, $\boldsymbol{h}(\boldsymbol{x}(t)) = \boldsymbol{0}$ for all $t \in (a,b)$. If we differentiate the function $\boldsymbol{h}(\boldsymbol{x}(t))$ with respect to t using the chain rule, we obtain
$$\frac{d}{dt}\boldsymbol{h}(\boldsymbol{x}(t)) = D(\boldsymbol{h}(\boldsymbol{x}(t)))\dot{\boldsymbol{x}}(t) = \boldsymbol{0}$$
for all $t \in (a,b)$. Therefore, at t^* we get $D(\boldsymbol{h}(\boldsymbol{x}^*))\boldsymbol{y} = \boldsymbol{0}$ and hence $\boldsymbol{y} \in T(\boldsymbol{x}^*)$

Tangent and Normal Spaces

- Definition 20.6. The normal space $N(\boldsymbol{x}^*)$ at a point $\boldsymbol{x}^*$ on the surface $S = \{\boldsymbol{x} \in R^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}\}$ is the set $N(\boldsymbol{x}^*) = \{\boldsymbol{x} \in R^n : \boldsymbol{x} = D\boldsymbol{h}(\boldsymbol{x}^*)^T \boldsymbol{z}, \boldsymbol{z} \in R^m\}$

- We can express the normal space

$$N(\boldsymbol{x}^*) = \mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*)^T)$$

that is, the range of the matrix $D\boldsymbol{h}(\boldsymbol{x}^*)^T$. Note that the normal space $N(\boldsymbol{x}^*)$ is the subspace of R^n spanned by the vectors $\triangledown h_1(\boldsymbol{x}^*), ..., \triangledown h_m(\boldsymbol{x}^*)$; that is,

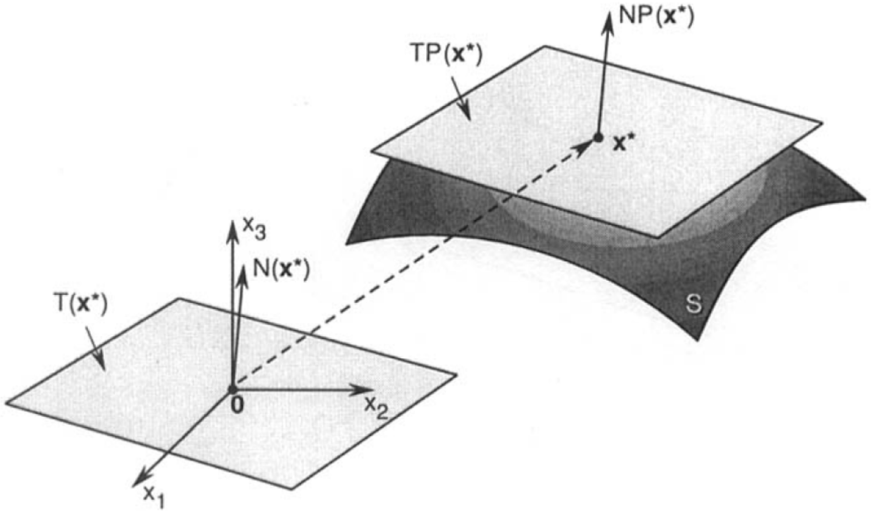
$$N(\boldsymbol{x}^*) = \mathrm{span}[\triangledown h_1(\boldsymbol{x}^*), ..., \triangledown h_m(\boldsymbol{x}^*)]$$
$$= \{\boldsymbol{x} \in R^n : \boldsymbol{x} = z_1 \triangledown h_1(\boldsymbol{x}^*) + \cdots + z_m \triangledown h_m(\boldsymbol{x}^*), z_1, ..., z_m \in R\}$$

Tangent and Normal Spaces

- Note that the normal space contains the zero vector. Assuming that $\boldsymbol{x}^*$ is regular, the dimension of the normal space $N(\boldsymbol{x}^*)$ is m . As in the case of the tangent space, it is often convenient to picture the normal space $N(\boldsymbol{x}^*)$ as passing through the point $\boldsymbol{x}^*$ (rather than through the origin of R^n). For this, we define the normal plane at $\boldsymbol{x}^*$ as the set

$$NP(\boldsymbol{x}^*) = N(\boldsymbol{x}^*) + \boldsymbol{x}^* = \{\boldsymbol{x} + \boldsymbol{x}^* \in R^n : \boldsymbol{x} \in N(\boldsymbol{x}^*)\}$$

- Figure 20.9



Tangent and Normal Spaces

- Lemma 20.1. We have $T(\boldsymbol{x}^*) = N(\boldsymbol{x}^*)^\perp$ and $T(\boldsymbol{x}^*)^\perp = N(\boldsymbol{x}^*)$
- Proof: By definition of $T(\boldsymbol{x}^*)$, we may write

$$T(\boldsymbol{x}^*) = \{\boldsymbol{y} \in R^n : \boldsymbol{x}^T\boldsymbol{y} = 0 \text{ for all } \boldsymbol{x} \in N(\boldsymbol{x}^*)\}$$

Hence, by definition of $N(\boldsymbol{x}^*)$, we have $T(\boldsymbol{x}^*) = N(\boldsymbol{x}^*)^\perp$
By Exercise 3.11 we also have $T(\boldsymbol{x}^*)^\perp = N(\boldsymbol{x}^*)$

Tangent and Normal Spaces

- By Lemma 20.1, we can write R^n as the direct sum decomposition (see Section 3.3):

$$R^n = N(\boldsymbol{x}^*) \oplus T(\boldsymbol{x}^*)$$

that is, given any vector $\boldsymbol{v} \in R^n$, there are unique vectors $\boldsymbol{w} \in N(\boldsymbol{x}^*)$ and $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ such that

$$\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{y}$$

Lagrange Condition

- Consider functions of two variables and only one equality constraint. Let $h : R^2 \to R$ be the constraint function. Recall that at each point $\boldsymbol{x}$ of the domain, the gradient vector $\bigtriangledown h(\boldsymbol{x})$ is orthogonal to the level set that passes through that point. Indeed, let us choose a point $\boldsymbol{x}^* = [x_1^*, x_2^*]^T$ such that $h(\boldsymbol{x}^*) = 0$ and assume that $\bigtriangledown h(\boldsymbol{x}) \neq \mathbf{0}$. The level set through the point $\boldsymbol{x}^*$ is the set $\{\boldsymbol{x} : h(\boldsymbol{x}) = 0\}$. We then parameterize this level set in a neighborhood of $\boldsymbol{x}^*$ by a curve $\{\boldsymbol{x}(t)\}$, that is, a continuously differentiable vector function $\boldsymbol{x} : R \to R^2$ such that

$$\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \quad t \in (a,b) \quad \boldsymbol{x}^* = \boldsymbol{x}(t^*) \quad \boldsymbol{x}(t^*) \neq \mathbf{0} \quad t^* \in (a,b)$$

Lagrange Condition

- We can now show that $\triangledown h(\boldsymbol{x}^*)$ is orthogonal to $\dot{\boldsymbol{x}}(t^*)$. Indeed, because h is constant on the curve $\{\boldsymbol{x}(t) : t \in (a,b)\}$ we have that for all $t \in (a,b)$, $h(\boldsymbol{x}(t)) = 0$
- Hence, for all $t \in (a,b)$, $\frac{d}{dt}h(\boldsymbol{x}(t)) = 0$
- Applying the chain rule, we get

$$\frac{d}{dt}h(\boldsymbol{x}(t)) = \triangledown h(\boldsymbol{x}(t))^T \dot{\boldsymbol{x}}(t) = 0$$

Therefore, $\triangledown h(\boldsymbol{x}^*)$ is orthogonal to $\dot{\boldsymbol{x}}(t^*)$

Lagrange Condition

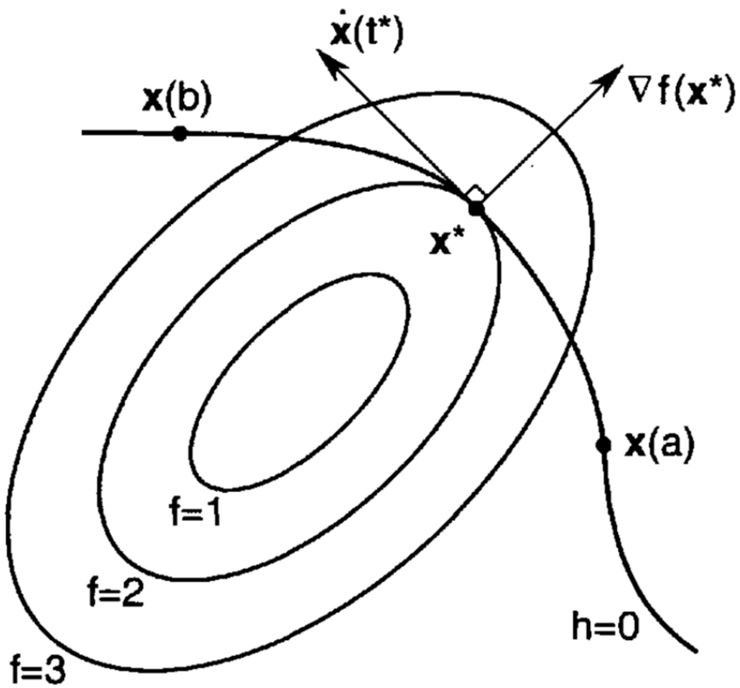
- Now suppose that $\boldsymbol{x}^*$ is a minimizer of $f : R^2 \to R$ on the set $\{\boldsymbol{x} : h(\boldsymbol{x}) = 0\}$. We claim that $\nabla f(\boldsymbol{x}^*)$ is orthogonal to $\dot{\boldsymbol{x}}(t^*)$ To see this, it is enough to observe that the composite function of t given by $\phi(t) = f(\boldsymbol{x}(t))$ achieves a minimum at t^* Consequently, the first-order necessary condition for the unconstrained extremum problem implies that $\frac{d\phi}{dt}(t^*) = 0$ Applying the chain rule yields

$$0 = \frac{d\phi}{dt}(t^*) = \nabla f(\boldsymbol{x}(t^*))^T \dot{\boldsymbol{x}}(t^*) = \nabla f(\boldsymbol{x}^*)^T \dot{\boldsymbol{x}}(t^*)$$

Thus, $\nabla f(\boldsymbol{x}^*)$ is orthogonal to $\dot{\boldsymbol{x}}(t^*)$. The fact that $\dot{\boldsymbol{x}}(t^*)$ is tangent to the curve $\{\boldsymbol{x}(t)\}$ at $\boldsymbol{x}^*$ means that $\nabla f(\boldsymbol{x}^*)$ is orthogonal to the curve at $\boldsymbol{x}^*$

Lagrange Condition

- Recall that $\triangledown h(\boldsymbol{x}^*)$ is also orthogonal to $\dot{\boldsymbol{x}}(t^*)$. Therefore, the vectors $\triangledown h(\boldsymbol{x}^*)$ and $\triangledown f(\boldsymbol{x}^*)$ are parallel; that is, $\triangledown f(\boldsymbol{x}^*)$ is a scalar multiple of $\triangledown h(\boldsymbol{x}^*)$. The observations allow us now to formulate ***Lagrange's theorem*** for functions of two variables with one constraint.

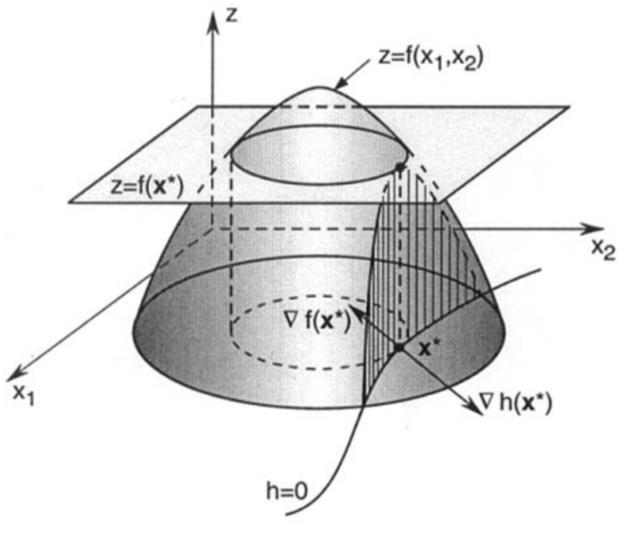


Lagrange Theorem

- Theorem 20.2 **Lagrange's Theorem** for $n = 2, m = 1$. Let the point $\boldsymbol{x}^*$ be a minimizer of $f : R^2 \to R$ subject to the constraint $h(\boldsymbol{x}) = 0, h : R^2 \to R$. Then, $\triangledown f(\boldsymbol{x}^*)$ and $\triangledown h(\boldsymbol{x}^*)$ are parallel. That is, if $\triangledown h(\boldsymbol{x}^*) \neq \mathbf{0}$, then there exists a scalar λ^* such that

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

- The scalar λ^* is called the ***Lagrange multiplier***. Note that the theorem also holds for maximizers.



Lagrange Theorem

- Lagrange's theorem provides a first-order necessary condition for a point to be a local minimizer. This condition, which we call the ***Lagrange condition***, consists of two equations:

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

- Note that the Lagrange condition is necessary but not sufficient. Figure 20.12 illustrates a variety of points where the Lagrange condition is satisfied, including a case where the point is not an extremizer.

Lagrange Theorem

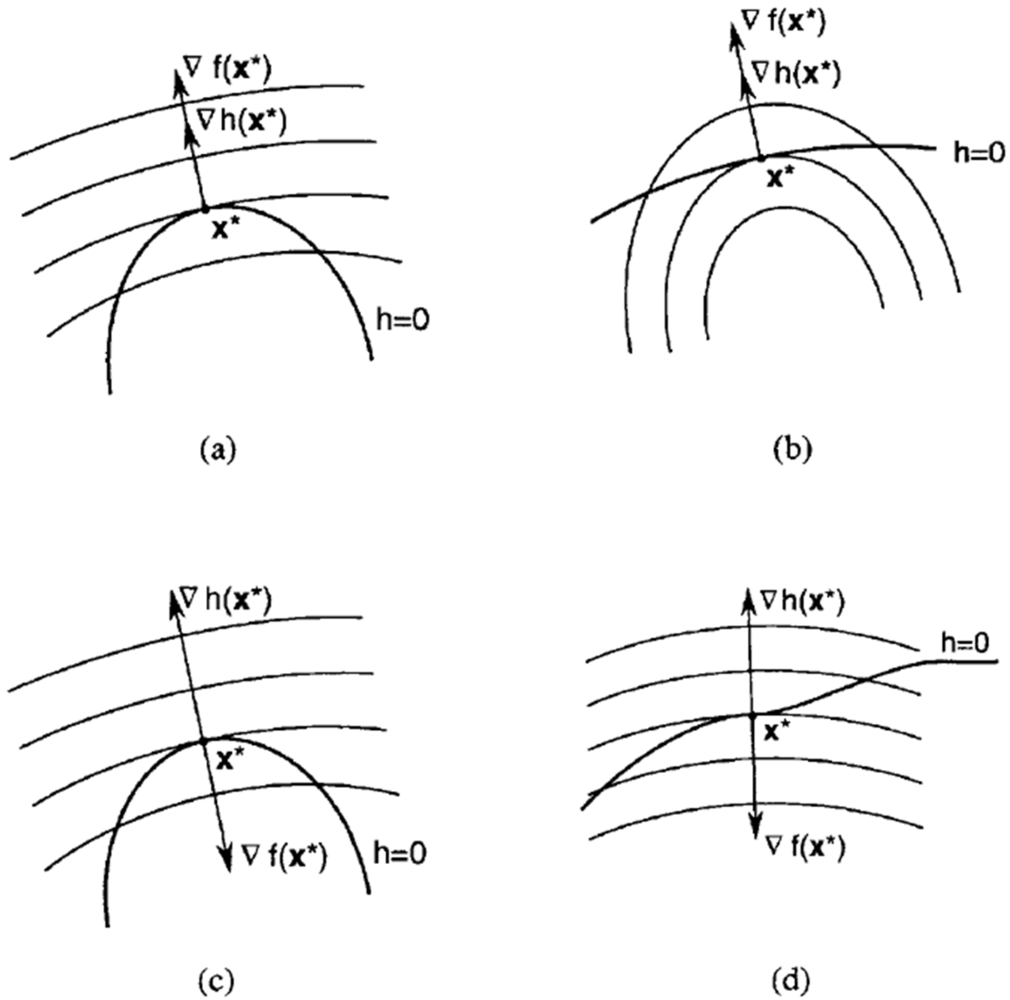


Figure 19.12 Four examples where the Lagrange condition is satisfied: (a) maximizer, (b) minimizer, (c) minimizer, (d) not an extremizer (adapted from [87])

Lagrange Theorem

- Theorem 20.3 Lagrange's Theorem. Let $\boldsymbol{x}^*$ be a local minimizer (or maximizer) of $f: R^n \to R$, subject to $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$ $\boldsymbol{h}: R^n \to R^m, m \leq n$. Assume that $\boldsymbol{x}^*$ is a regular point. Then, there exists $\boldsymbol{\lambda}^* \in R^m$ such that

$$Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}^T$$

- Proof. We need to prove that

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

for some $\boldsymbol{\lambda}^* \in R^m$; that is, $\nabla f(\boldsymbol{x}^*) \in \mathcal{R}(D\boldsymbol{h}(\boldsymbol{x}^*)^T) = N(\boldsymbol{x}^*)$. But by Lemma 20.1, $N(\boldsymbol{x}^*) = T(\boldsymbol{x}^*)^\perp$. Therefore, it remains to show that $\nabla f(\boldsymbol{x}^*) \in T(\boldsymbol{x}^*)^\perp$

Lagrange Theorem

- Proof. Suppose that $\boldsymbol{y} \in T(\boldsymbol{x}^*)$. Then, by Theorem 20.1, there exists a differentiable curve $\{\boldsymbol{x}(t) : t \in (a,b)\}$ such that for all $t \in (a,b)$, $\boldsymbol{h}(\boldsymbol{x}(t)) = \boldsymbol{0}$, and there exists $t^* \in (a,b)$ satisfying

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

- Now consider the composite function $\phi(t) = f(\boldsymbol{x}(t))$. Note that t^* is a local minimizer of this function. By the first-order necessary condition for unconstrained local minimizers (see Theorem 6.1)

$$\frac{d\phi}{dt}(t^*) = 0$$

Lagrange Theorem

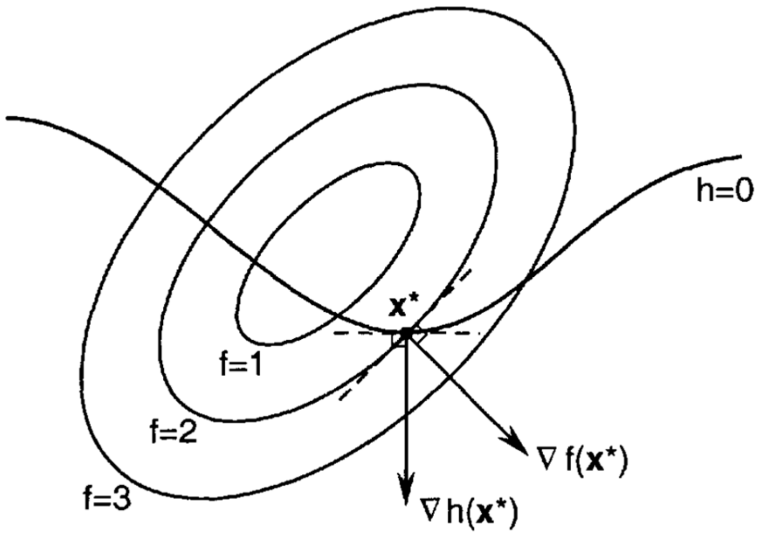
- Proof. Applying the chain rule yields

$$\frac{d\phi}{dt}(t^*) = Df(\boldsymbol{x}^*)\dot{\boldsymbol{x}}(t^*) = Df(\boldsymbol{x}^*)\boldsymbol{y} = \bigtriangledown f(\boldsymbol{x}^*)^T\boldsymbol{y} = 0$$

  So all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$ satisfy $\bigtriangledown f(\boldsymbol{x}^*)^T\boldsymbol{y} = 0$
  that is, $\bigtriangledown f(\boldsymbol{x}^*) \in T(\boldsymbol{x}^*)^{\perp}$
  This completes the proof.

Lagrange's Theorem

- Lagrange's theorem states that if $\boldsymbol{x}^*$ is an extremizer, then the gradient of the objective function f can be expressed as a linear combination of the gradients of the constraints. We refer to the vector $\boldsymbol{\lambda}^*$ as the ***Lagrange multiplier vector***, and its component as ***Lagrange multipliers***.
- A compact way to write the necessary condition is $\nabla f(\boldsymbol{x}^*) \in N(\boldsymbol{x}^*)$. If this condition fails, then $\boldsymbol{x}^*$ cannot be an extremizer.
- Figure 20.13



Lagrange's Theorem

- Consider the following problem:

$$\text{minimize } f(x)$$
$$\text{subject to } h(x) = 0$$

where $f(x) = x$ and

$$h(x) = \begin{cases} x^2 & \text{if } x < 0 \\ 0 & \text{if } 0 \leq x \leq 1 \\ (x-1)^2 & \text{if } x > 1 \end{cases}$$

The feasible set is evidently [0, 1]. Clearly, $x^* = 0$ is a local minimizer. However, $f'(x^*) = 1$ and $h'(x^*) = 0$. Therefore, x^* does not satisfy the necessary condition in Lagrange's theorem. Note, however, that x^* is not a regular point, which is why Lagrange's theorem does not apply here.

Lagrange's Theorem

- It is convenient to introduce the ***Lagrangian function*** $l : R^n \times R^m \to R$ given by

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) \triangleq f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x})$$

  The Lagrange condition for a local minimizer $\boldsymbol{x}^*$ can be represented using the Lagrangian function as

$$Dl(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}^T$$

  for some $\boldsymbol{\lambda}^*$, where the derivative operation *D* is with respect to the entire argument $[\boldsymbol{x}^T, \boldsymbol{\lambda}^T]^T$. In other words, the necessary condition in Lagrange's theorem is equivalent to the first-order necessary condition for unconstrained optimization applied to the Lagrangian function.

Lagrange's Theorem

- Denote the derivative of l with respect to $\boldsymbol{x}$ as $D_x l$ and the derivative of l with respect to $\boldsymbol{\lambda}$ as $D_\lambda l$. Then,
$$Dl(\boldsymbol{x}, \boldsymbol{\lambda}) = [D_x l(\boldsymbol{x}, \boldsymbol{\lambda}), D_\lambda l(\boldsymbol{x}, \boldsymbol{\lambda})]$$
Note that $D_x l(\boldsymbol{x}, \boldsymbol{\lambda}) = Df(\boldsymbol{x}) + \boldsymbol{\lambda}^T D\boldsymbol{h}(\boldsymbol{x})$ and $D_\lambda l(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{h}(\boldsymbol{x})^T$
Therefore, Lagrange's theorem for a local minimizer $\boldsymbol{x}^*$ can be stated as
$$D_x l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}^T$$
$$D_\lambda l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}^T$$
for some $\boldsymbol{\lambda}^*$, which is equivalent to $Dl(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}^T$
In other words, the Lagrange condition can be expressed as $Dl(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}^T$

Lagrange's Theorem

- The Lagrange condition is used to find possible extremizers. This entails solving the equations

$$D_x l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}^T$$

$$D_\lambda l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{0}^T$$

  The above represents $n+m$ equations in $n+m$ unknowns. Keep in mind that the Lagrange condition is necessary but not sufficient; that is, a point $\boldsymbol{x}^*$ satisfying the equations above need not be an extremizer.

Example

- Given a fixed area of cardboard, we wish to construct a closed cardboard box with maximum volume. We can formulate and solve this problem using the Lagrange condition.
- Denote the dimension of the box with maximum volume by x_1, x_2, x_3 and let the given fixed area of cardboard be A. The problem can then be formulated as

$$\begin{aligned} &\text{maximize } x_1x_2x_3 \\ &\text{subject to } x_1x_2 + x_2x_3 + x_3x_1 = \frac{A}{2} \end{aligned}$$

Example

- 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 $\bigtriangledown f(\boldsymbol{x}) = -[x_2x_3, x_1x_3, x_1x_2]^T$ and $\bigtriangledown h(\boldsymbol{x}) = [x_2 + x_3, x_1 + x_3, x_1 + x_2]^T$. Note that all feasible points are regular in this case. By the Lagrange condition, the dimension of the box with maximum volume satisfies

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

where $\lambda \in R$

Example

- We now solve these equations. First, we show that x_1, x_2, x_3 and λ are all nonzero. Suppose that $x_1 = 0$. By the constraints, we have $x_2 x_3 = A/2$. 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 $x_2 x_3 = 0$ This contradicts the constraints. A similar argument applies to x_2, x_3
- Next, suppose that $\lambda = 0$. Then, the sum of the three Lagrange equations gives $x_2 x_3 + x_1 x_3 + x_1 x_2 = 0$, which contradicts the constraints.

Example

- We now solve for x_1, x_2, 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 $x_1 = x_2 = x_3 = \sqrt{A/6}$

Example

- Notice that we have ignored the constraints that x_1, x_2, 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, x_3 , then this solution must necessarily be equal to the solution above obtained by ignoring the positivity constraints.

Example

- Consider the problem of extremizing the objective function $f(\boldsymbol{x}) = x_1^2 + x_2^2$ on the ellipse

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

We have

$$\bigtriangledown f(\boldsymbol{x}) = [2x_1, 2x_2]^T$$

$$\bigtriangledown h(\boldsymbol{x}) = [2x_1, 4x_2]^T$$

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

$$2x_1 + 2\lambda x_1 = 0$$

$$2x_2 + 4\lambda x_2 = 0$$

$$x_1^2 + 2x_2^2 = 1$$

$$2x_1 + 2\lambda x_1 = 0$$
$$2x_2 + 4\lambda x_2 = 0$$
$$x_1^2 + 2x_2^2 = 1$$

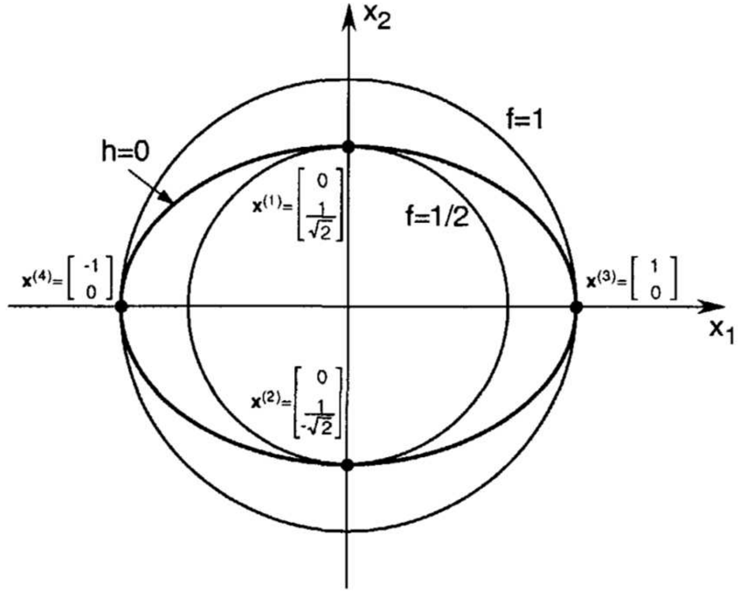
Example

- 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

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

Example

- Because
$$f(\boldsymbol{x}^{(1)}) = f(\boldsymbol{x}^{(2)}) = \frac{1}{2}$$
$$f(\boldsymbol{x}^{(3)}) = f(\boldsymbol{x}^{(4)}) = 1$$
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 (Figure 20.14)



Example

- Consider the following problem:

$$\text{maximize } \frac{\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x}}$$

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

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

Indeed,

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

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

$$\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} = 1$$

Example

- The optimization problem becomes

$$\text{maximize } \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$$
$$\text{subject to } \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} = 1$$

- Let us write $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}$, $h(\boldsymbol{x}) = 1 - \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x}$
- Any feasible point for this problem is regular. We now apply Lagrange's method. We first form the Lagrangian function

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

Applying the Lagrange condition yields

$$D_x l(\boldsymbol{x}, \lambda) = 2\boldsymbol{x}^T \boldsymbol{Q} - 2\lambda \boldsymbol{x}^T \boldsymbol{P} = \boldsymbol{0}^T$$
$$D_\lambda l(\boldsymbol{x}, \lambda) = 1 - \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} = 0$$

$$D_x l(\boldsymbol{x}, \lambda) = 2\boldsymbol{x}^T \boldsymbol{Q} - 2\lambda \boldsymbol{x}^T \boldsymbol{P} = \boldsymbol{0}^T$$

Example

$$D_\lambda l(\boldsymbol{x}, \lambda) = 1 - \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} = 0$$

- The first of the equations above can be represented as

$$\boldsymbol{Q}\boldsymbol{x} - \lambda \boldsymbol{P}\boldsymbol{x} = \boldsymbol{0} \quad \text{or} \quad (\lambda \boldsymbol{P} - \boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0}$$

This representation is possible because $\boldsymbol{P} = \boldsymbol{P}^T$ and $\boldsymbol{Q} = \boldsymbol{Q}^T$
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 exists, is an eigenvector of $\boldsymbol{P}^{-1}\boldsymbol{Q}$ and the Lagrange multiplier is the corresponding eigenvalue.

Example

- As usual, let $\boldsymbol{x}^*$ and λ^* be the optimal solution. Because $\boldsymbol{x}^{*T}\boldsymbol{P}\boldsymbol{x}^* = 1$ and $\boldsymbol{P}^{-1}\boldsymbol{Q}\boldsymbol{x}^* = \lambda^*\boldsymbol{x}^*$, we have
$$\lambda^* = \boldsymbol{x}^{*T}\boldsymbol{Q}\boldsymbol{x}^*$$
Hence, λ^* is the maximum of the objective function, and therefore is, in fact, the maximal eigenvalue of $\boldsymbol{P}^{-1}\boldsymbol{Q}$

Second-Order Conditions

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

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

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 $\boldsymbol{x}$, $k = 1, ..., m$, given by

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

Second-Order Conditions

- We introduce the notation $[\boldsymbol{\lambda H}(\boldsymbol{x})]$:
$$[\boldsymbol{\lambda H}(\boldsymbol{x})] = \lambda_1 \boldsymbol{H}_1(\boldsymbol{x}) + \cdots + \lambda_m \boldsymbol{H}_m(\boldsymbol{x})$$
- Using the notation above, we can write
$$\boldsymbol{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{F}(\boldsymbol{x}) + [\boldsymbol{\lambda H}(\boldsymbol{x})]$$
- Theorem 20.4. **Second-Order Necessary Conditions**. Let $\boldsymbol{x}^*$ be a local minimizer of $f : R^n \to R$ subject to $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{h} : R^n \to R^m, m \leq n$, and $f, \boldsymbol{h} \in \mathcal{C}^2$. Suppose that $\boldsymbol{x}^*$ is regular. Then, there exists $\boldsymbol{\lambda}^* \in R^m$ such that:
  - 1. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T$
  - 2. For all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$. We have $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \boldsymbol{y} \geq 0$

Second-Order Conditions

- 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 $\boldsymbol{T}(\boldsymbol{x}^*)$ rather than on R^n
- These 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.

Second-Order Conditions

- Theorem 20.5. **Second-Order Sufficient Conditions**. Suppose that $f, \boldsymbol{h} \in \mathcal{C}^2$ and there exists a point $\boldsymbol{x}^* \in R^n$ and $\boldsymbol{\lambda}^* \in R^m$ such that:
  - 1. $Df(\boldsymbol{x}^*) + \boldsymbol{\lambda}^{*T} D\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}^T$
  - 2. For all $\boldsymbol{y} \in T(\boldsymbol{x}^*)$. We have $\boldsymbol{y}^T \boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{y} > 0$

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

Second-Order Conditions

- Consider the following problem:

$$\text{maximize} \ \frac{\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}} \qquad \boldsymbol{Q} = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{P} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

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

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

- The Lagrangian function for the transformed problem is given by $l(\boldsymbol{x},\lambda) = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \lambda(1 - \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x})$
  The Lagrange condition yields

$$(\lambda\boldsymbol{I} - \boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{x} = \boldsymbol{0} \qquad \boldsymbol{P}^{-1}\boldsymbol{Q} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Second-Order Conditions

- There are only two values of λ that satisfy $(\lambda \boldsymbol{I} - \boldsymbol{P}^{-1}\boldsymbol{Q})\boldsymbol{x} = 0$ 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 $\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} = 1$ is $\pm \boldsymbol{x}^*$, where

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

Second-Order Conditions

- 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 $\{\boldsymbol{x} : 1 - \boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x} = 0\}$ is

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

Second-Order Conditions

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

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

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

- Note that

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

  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 $\dfrac{\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{P}\boldsymbol{x}}$

Minimizing Quadratics Subject to Linear Constraints

- Consider the problem

$$\begin{aligned} &\text{minimize } \tfrac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

  where $\boldsymbol{Q} > 0, \boldsymbol{A} \in R^{m\times n}, m < n, rank(\boldsymbol{A}) = m$. 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 $\boldsymbol{x} \geq \mathbf{0}$).
- Note that the constraint set contains an infinite number of points.

Minimizing Quadratics Subject to Linear Constraints

- To solve the problem, we first form the Lagrangian function

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{\lambda}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$$

The Lagrange condition yields

$$D_x l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{x}^{*T}\boldsymbol{Q} - \boldsymbol{\lambda}^{*T}\boldsymbol{A} = \boldsymbol{0}^T$$

Rewriting, we get

$$\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^*$$

Premultiplying both sides of the above by $\boldsymbol{A}$ gives

$$\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^*$$

Using the fact that $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$, and noting that $\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T$ is invertible because $\boldsymbol{Q} > 0$ and $rank(\boldsymbol{A}) = m$, we can solve for $\boldsymbol{\lambda}^*$ to obtain $\boldsymbol{\lambda}^* = (\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}$. Therefore, we obtain

$$\boldsymbol{x}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}$$

Minimizing Quadratics Subject to Linear Constraints

- The point $\boldsymbol{x}^*$ is the only candidate for a minimizer. To establish that $\boldsymbol{x}^*$ is indeed a minimizer, we verify that it satisfies the second-order sufficient conditions.
- For this, we first find the Hessian matrix of the Lagrangian function at $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$. We have

$$\boldsymbol{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{Q}$$

  which is positive definite. Thus, the point $\boldsymbol{x}^*$ is a strict local minimizer.

Minimizing Quadratics Subject to Linear Constraints

- The special case where $\boldsymbol{Q} = \boldsymbol{I}_n$ reduces to the problem considered in Section 12.3. Specifically, the problem in Section 12.3 is to minimize the norm $\|\boldsymbol{x}\|$ subject to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ The objective function here is $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$, which is not differentiable at $\boldsymbol{x} = \boldsymbol{0}$. This precludes the use of Lagrange's theorem because the theorem requires differentiability of the objective function.
- We can overcome this difficulty by considering an equivalent optimization problem

$$\begin{aligned} &\text{minimize } \tfrac{1}{2}\|\boldsymbol{x}\|^2 \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Minimizing Quadratics Subject to Linear Constraints

- The objective function $\|\boldsymbol{x}\|^2/2$ has the same minimizer as the previous objective function $\|\boldsymbol{x}\|$. Indeed, if $\boldsymbol{x}^*$ is such that for all $\boldsymbol{x} \in R^n$ satisfying $\boldsymbol{Ax} = \boldsymbol{b}$, $\|\boldsymbol{x}^*\| \leq \|\boldsymbol{x}\|$, then $\|\boldsymbol{x}^*\|^2/2 \leq \|\boldsymbol{x}\|^2/2$ subject to $\boldsymbol{Ax} = \boldsymbol{b}$ is simply the problem considered above with $\boldsymbol{Q} = \boldsymbol{I}_n$, we easily deduce the solution to be $\boldsymbol{x}^* = \boldsymbol{A}^T(\boldsymbol{AA}^T)^{-1}\boldsymbol{b}$, which agrees with the solution in Section 12.3.

Example

- Consider the discrete-time linear system model

$$x_k = ax_{k-1} + bu_k, \quad k \geq 1$$

with initial condition x_0 given. We can think of $\{x_k\}$ as a discrete-time signal that is controlled by an external input signal $\{u_k\}$. In the control literature, x_k is called the *state* at time k . For a given x_0 , our goal is to choose the control signal $\{u_k\}$ so that the state remains "small" over a time interval $[1, N]$, but at the same time the control signal is "not too large."

Example

- To express the desire to keep the state $\{x_k\}$ small, we choose the control sequence to minimize

$$\frac{1}{2}\sum_{i=1}^{N} x_i^2$$

  On the other hand, maintaining a control signal that is not too large, we minimize

$$\frac{1}{2}\sum_{i=1}^{N} u_i^2$$

- The two objectives above are conflicting in the sense that they cannot, in general, be achieved simultaneously – minimizing the first may result in a large control effort, while minimizing the second may result in large states.

Example

- One way to approach the problem is to minimize a weighted sum of the two functions above. Specifically, we can formulate the problem as

$$\text{minimize } \frac{1}{2}\sum_{i=1}^{N}(qx_i^2 + ru_i^2)$$

$$\text{subject to } x_k = ax_{k-1} + bu_k, \quad k = 1, \ldots, N, x_0 \text{ given}$$

  where the parameters q and r reflect the relative importance of keeping the state small versus keeping the control effort not too large. This problem is an instance of the *linear quadratic regulator* (LQR) problem. Combining the two conflicting objectives of keeping the state small while keeping the control effort small is an instance of the weighted *sum* approach.

Example

- To solve the problem, we can rewrite it as a quadratic programming problem. Define

$$\boldsymbol{Q} = \begin{bmatrix} q\boldsymbol{I}_n & \boldsymbol{O} \\ \boldsymbol{o} & r\boldsymbol{I}_n \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & & \cdots & 0 & -b & & \cdots & 0 \\ -a & 1 & & \vdots & & -b & & \vdots \\ & \ddots & \ddots & & \vdots & & \ddots & \\ 0 & & -a & 1 & 0 & \cdots & & -b \end{bmatrix}$$

$$\boldsymbol{b} = \begin{bmatrix} ax_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \boldsymbol{z} = [x_1, \ldots, x_N, u_1, \ldots, u_N]^T$$

Example

- With these definitions, the problem reduces to the previously considered quadratic programming problem

$$\text{minimize } \frac{1}{2}\boldsymbol{z}^T\boldsymbol{Q}\boldsymbol{z}$$
$$\text{subject to } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}$$

  where $\boldsymbol{Q}$ is $2N \times 2N$, $\boldsymbol{A}$ is $N \times 2N$, and $\boldsymbol{b} \in R^N$. The solution is $\boldsymbol{z}^* = \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}\boldsymbol{b}$
- The first N components of $\boldsymbol{z}^*$ represent the optimal state signal in the interval $[1, N]$, whereas the second N components represent the optimal control signal.

Example

- In practice, computation of the matrix inverses in the formula for z^* above may be too costly. There are other ways to tackle the problem by exploiting its special structure. This is the study of *optimal control*.

Example

- Credit-Card Holder Dilemma. Suppose that we currently have a credit-card debt of $10,000. Credit-card debts are subject to a monthly interest rate of 2%, and the account balance is increased by the interest amount every month. Each month we have the option of reducing the account balance by contributing a payment to the account. Over the next 10 months, we plan to contribute a payment every month in such a way as to minimize the overall debt level while minimizing the hardship of making monthly payments.

Example

- We solve our problem using the LQR framework. Let the current time be 0, x_k the account balance at the end of month k, and u_k our payment in month k. We have

  $$x_k = 1.02x_{k-1} - u_k, \quad k = 1, ..., 10$$

  that is, the account balance in a given month is equal to the account balance in the previous month plus the monthly interest on that balance minus our payment that month. Our optimization problem is then

  $$\begin{aligned} &\text{minimize } \tfrac{1}{2}\textstyle\sum_{i=1}^{10}(qx_i^2 + ru_i^2) \\ &\text{subject to } x_k = 1.02x_{k-1} - u_k, \quad k = 1, ..., 10, x_0 = 10,000 \end{aligned}$$

  which is an instance of the LQR problem. The parameters q and r reflect our priority in trading off between debt reduction and hardship in making payments.

Example

- The more anxious we are to reduce our debt, the larger the value of q relative to r. On the other hand, the more reluctant we are to make payments, the larger the value of r relative to q.
- The solution to the problem is given by the formula derived in previous example. This figure plots the monthly account balances and payments over the next 10 months using $q = 1$ and $r = 10$. We can see here that our debt has been reduced to less than $1000 after 10 months, but with a first payment close to $3000.

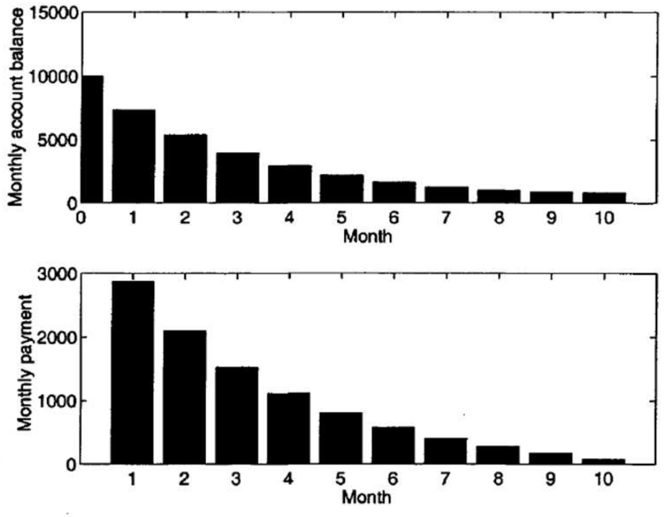


Figure 19.15 Plots for Example 19.10 with $q = 1$ and $r = 10$

Example

- If we feel that a payment of \$3000 is too high, then we can try to reduce this amount by increasing the value of r relative to q. However, going too far along these lines can lead to trouble. Indeed, if we use $q = 1, r = 300$, although the monthly payments do not exceed \$400, the account balance is never reduced by much below \$10,000. In this case, the interest on the account balance eats up a significant portion of our monthly payments. In fact, our debt after 10 months will be higher than \$10,000.

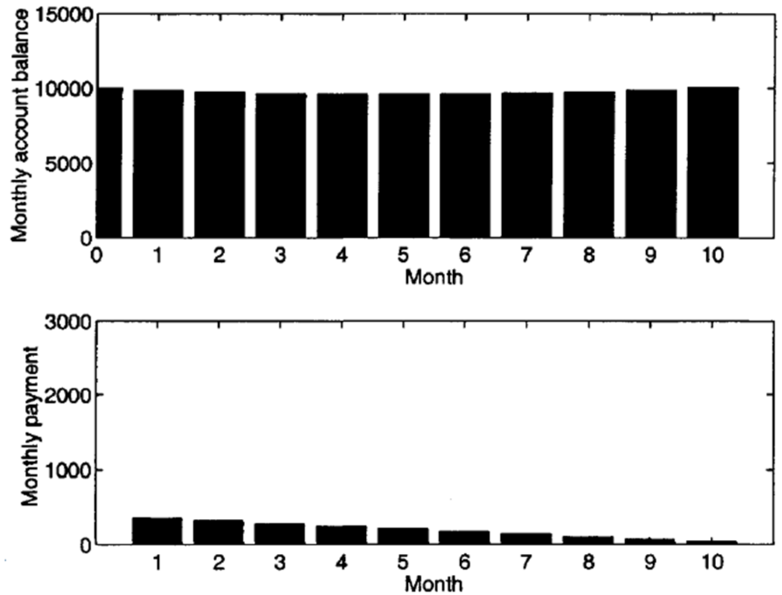


Figure 19.16 Plots for Example 19.10 with $q = 1$ and $r = 300$