

## Chapter 4

# Nonlinear Optimization

Chapters 2 and 3 are devoted to problems with a linear structure. More specifically, the problems studied there all have an objective function and constraints that are linear in the decision variables. Although linear optimization problems are very common and can model a variety of real-world problems, we are sometimes faced with modeling a system that includes important nonlinearities. When either the constraints or objective function of an optimization problem are nonlinear in the decision variables, we say that we are faced with a nonlinear optimization problem or nonlinear programming problem (NLPP).

In this chapter, we begin by first introducing some nonlinear optimization problems, then discuss methods to solve NLPPs. The example nonlinear optimization problems that we introduce draw on a wide swath of problem domains spanning finance, design, planning, and energy systems. We then discuss an analytical approach to solving NLPPs. This method uses what are called **optimality conditions**—properties that an optimal set of decision variables has to exhibit. Optimality conditions can be likened to the technique to solve an LPP that draws on the Strong-Duality Property, which is discussed in Section 2.7.4. In Chapter 5 we introduce another approach to solving NLPPs, **iterative solution algorithms**, which are techniques used by software packages to solve NLPPs. These iterative algorithms can be likened to using the Simplex method to solve linear optimization problems or the branch-and-bound or cutting-plane methods to solve mixed-integer linear problems.

### 4.1 Motivating Examples

In this section, we present a variety of nonlinear optimization problems, which are motivated by a mixture of geometric, mechanical, and electrical systems.

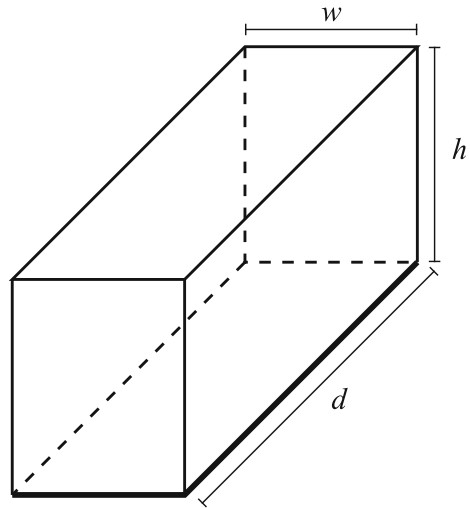
## 4.1.1 Geometric Examples

### 4.1.1.1 Packing-Box Problem

A company must determine the dimensions of a cardboard box to maximize its volume. The box can use at most  $60 \text{ cm}^2$  of cardboard. For structural reasons, the bottom and top faces of the box must be of triple weight (*i.e.*, three pieces of cardboard).

There are three decision variables in this problem,  $h$ ,  $w$ , and  $d$ , which are the height, width, and depth of the box in cm, respectively (see Figure 4.1).

**Fig. 4.1** Illustration of the Packing-Box Problem



The objective is to maximize the volume of the box:

$$\max_{h,w,d} hwd.$$

There are two types of constraints. First, we must ensure that the box uses no more than  $60 \text{ cm}^2$  of cardboard, noting that the top and bottom of the box are of triple weight:

$$2wh + 2dh + 6wd \leq 60.$$

The second type of constraint ensures that the dimensions of the box are non-negative, as negative dimensions are physically impossible:

$$w, h, d \geq 0.$$

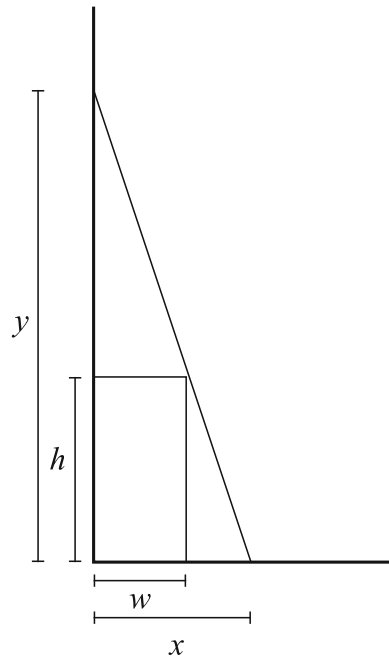
Putting all of this together, the NLPP can be written as:

$$\begin{aligned} & \max_{h,w,d} hwd \\ & \text{s.t. } 2wh + 2dh + 6wd \leq 60 \\ & \quad w, h, d \geq 0. \end{aligned}$$

### 4.1.1.2 Awning Problem

A box, which is  $h$  m high and  $w$  m wide, is placed against the side of a building (see Figure 4.2). The building owner would like to construct an awning of minimum length that completely covers the box.

**Fig. 4.2** Illustration of the Awning Problem



There are two decision variables in this problem. The first,  $x$ , measures the distance between the building wall and the point at which the awning is anchored to the ground, in meters. The second,  $y$ , measures the height of the anchor point of the awning to the building wall, also in meters.

The objective is to minimize the length of the awning. From the Pythagorean theorem this is:

$$\min_{x,y} \sqrt{x^2 + y^2}.$$

We have two types of constraints in this problem. The first ensures that the upper-right corner of the box is below the awning (which ensures that the box is wholly contained by the awning). To derive this constraint, we compute the height of the awning  $w$  m away from the wall as:

$$y - \frac{y}{x}w.$$

To ensure that the upper-right corner of the box is below the awning, this point must be at least  $h$  m high, giving our first constraint:

$$y - \frac{y}{x}w \geq h.$$

We must also ensure that the distances between the anchor points of the awning and the building and ground are non-negative:

$$x, y \geq 0.$$

Thus, our NLPP is:

$$\begin{aligned} \min_{x,y} & \sqrt{x^2 + y^2} \\ \text{s.t.} & y - \frac{y}{x}w \geq h \\ & x, y \geq 0. \end{aligned}$$

#### 4.1.1.3 Facility-Location Problem

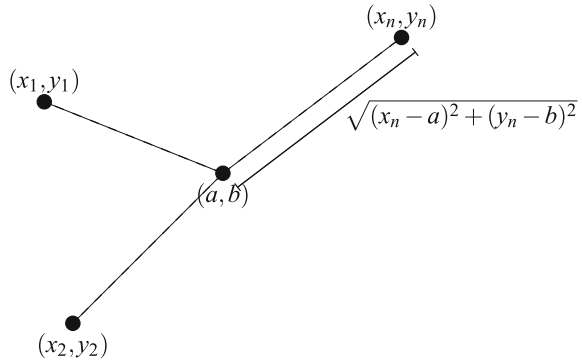
A retailer must decide where to place a single distribution center to service  $N$  retail locations in a region (see Figure 4.3). Retail location  $n$  is at coordinates  $(x_n, y_n)$ . Each week  $V_n$  trucks leave the distribution center carrying goods to retail location  $n$ , and then return to the distribution center. These trucks can all travel on straight paths from the distribution center to the retail location. The company would like to determine where to place the distribution center to minimize the total distance that all of the trucks must travel each week.

There are two variables in this problem,  $a$  and  $b$ , which denote the coordinates of the distribution center.

The objective is to minimize the total distance traveled by all of the trucks. As illustrated in Figure 4.3, each truck that serves retail location  $n$  must travel a distance of:

$$\sqrt{(x_n - a)^2 + (y_n - b)^2},$$

**Fig. 4.3** Illustration of the Facility-Location Problem



to get there from the distribution center. Because  $V_n$  trucks must travel to retail location  $n$  and they make a roundtrip, the total distance covered by all of the trucks to retail location  $n$  is:

$$2V_n\sqrt{(x_n - a)^2 + (y_n - b)^2}.$$

The objective is to minimize this total distance, or:

$$\min_{a,b} \sum_{n=1}^N 2V_n\sqrt{(x_n - a)^2 + (y_n - b)^2}.$$

Because the distribution center can be placed anywhere, this problem has no constraints.

Thus, our NLPP is simply:

$$\min_{a,b} \sum_{n=1}^N 2V_n\sqrt{(x_n - a)^2 + (y_n - b)^2}.$$

#### 4.1.1.4 Cylinder Problem

A brewery would like to design a cylindrical vat (see Figure 4.4), in which to brew beer. The material used to construct the top of the vat costs  $\$c_1$  per  $m^2$ . The material cost of the side and bottom of the vat is proportional to the volume of the vat. This is because the side and bottom must be reinforced to hold the volume of liquid that is brewed in the vat. This material cost is  $\$c_2V$  per  $m^2$ , where  $V$  is the volume of the vat. The value of the beer that is produced in the vat during its usable life is proportional to the volume of the vat, and is given by  $\$N$  per  $m^3$ . The brewery would like to design the vat to maximize the net profit earned over its usable life.

This NLPP has two decision variables,  $h$  and  $r$ , which measure the height and radius, respectively, of the vat in meters.

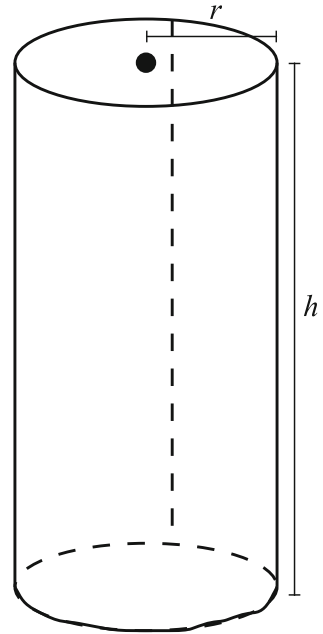
The objective is to maximize the value of the beer brewed less the cost of building the vat. The value of the beer produced by the vat is given by:

$$N\pi r^2 h.$$

The cost of top of the vat is:

$$c_1\pi r^2.$$

**Fig. 4.4** Illustration of the Cylinder Problem



The per-unit material cost (in  $\$/\text{m}^3$ ) of the side and bottom of the vat is given by:

$$c_2\pi r^2 h.$$

Thus the bottom of the vat costs:

$$c_2\pi r^2 h\pi r^2,$$

and the side costs:

$$c_2\pi r^2 h2\pi r h.$$

Thus, the objective, which is to maximize the value of the beer produced by the vat less its material costs, is:

$$\max_{h,r} N\pi r^2 h - c_1 \pi r^2 - c_2 \pi r^2 h \cdot (\pi r^2 + 2\pi r h).$$

There is one type of constraint, which ensures that the vat has non-negative dimensions:

$$r, h \geq 0.$$

Putting all of this together, the NLPP can be written as:

$$\begin{aligned} \max_{h,r} & N\pi r^2 h - c_1 \pi r^2 - c_2 \pi r^2 h \cdot (\pi r^2 + 2\pi r h) \\ \text{s.t. } & r, h \geq 0. \end{aligned}$$

## 4.1.2 Mechanical Examples

### 4.1.2.1 Machining-Speed Problem

A company produces widgets that need to be machined. Each widget needs to spend  $t_p$  minutes being prepared before being machined. The time that each widget spends being machined depends on the machine speed. This machining time is given by:

$$\frac{\lambda}{v},$$

where  $\lambda$  is a constant and  $v$  is the machine speed. The tool used to machine the widgets needs to be replaced periodically, due to wear. Each time the tool is replaced the machine is out of service (and widgets cannot be produced) for  $t_c$  minutes. The amount of time it takes for the tool to be worn down depends on the machine speed, and is given by:

$$\left(\frac{C}{v}\right)^{1/n},$$

where  $C$  and  $n$  are constants. The cost of replacement tools is negligible. Each widget sells for a price of  $\$p$  and uses  $\$m$  worth of raw materials. Moreover, every minute of production time incurs  $\$h$  of overhead costs. The company would like to determine the optimal machine speed that maximizes the average per-minute profit of the widget-machining process.

This problem has a single decision variable,  $v$ , which is the machine speed.

To derive the objective, we first note that each widget produced earns  $\$(p - m)$  of revenue less material costs. Dividing this by the number of widgets produced per minute gives an expression for net revenue per minute. Producing each widget takes

three steps. The first is preparation, which takes  $t_p$  minutes. Each widget must also be machined, which takes:

$$\frac{\lambda}{v},$$

minutes. Finally, each widget produced reduces the usable life of the machine tool by the fraction:

$$\frac{\lambda/v}{(C/v)^{1/n}}.$$

Thus, each widget contributes:

$$t_c \frac{\lambda/v}{(C/v)^{1/n}},$$

minutes toward tool-replacement time. Therefore, each widget takes a total of:

$$t_p + \frac{\lambda}{v} + t_c \frac{\lambda/v}{(C/v)^{1/n}},$$

minutes of time to produce. The objective, which is to maximize average per-minute profits of the widget-machining process, is, thus, given by:

$$\max_v \frac{p - m}{t_p + \frac{\lambda}{v} + t_c \frac{\lambda/v}{(C/v)^{1/n}}} - h,$$

where we also subtract the overhead costs.

The only constraint on this problem is that the machine speed needs to be non-negative:

$$v \geq 0.$$

Thus, our NLPP is:

$$\begin{aligned} \max_v & \frac{p - m}{t_p + \frac{\lambda}{v} + t_c \frac{\lambda/v}{(C/v)^{1/n}}} - h \\ \text{s.t. } & v \geq 0. \end{aligned}$$

#### 4.1.2.2 Hanging-Chain Problem

A chain consisting of  $N$  links, each of which is 10 cm in length and 50 g in mass, hangs between two points a distance  $L$  cm apart, as shown in Figure 4.5. The chain will naturally hang in a configuration that minimizes its potential energy. Formulate a nonlinear optimization problem to determine the configuration of the chain links when it is hanging.



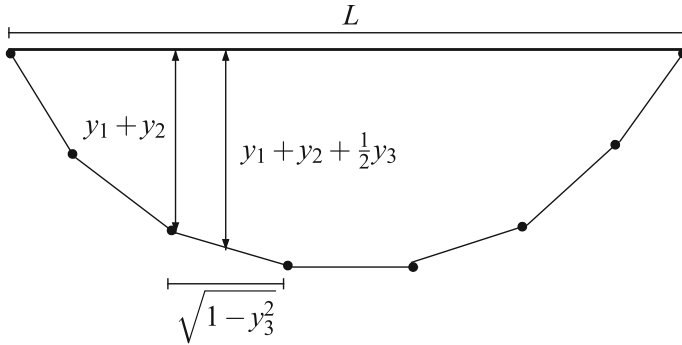


Fig. 4.5 Illustration of the Hanging-Chain Problem

To formulate this problem, we define variables,  $y_1, y_2, \dots, y_N$ . We let  $y_n$  measure the vertical displacement of the right end of the  $n$ th chain link from the right end of the  $(n - 1)$ th link in cm (see Figure 4.5).

The objective of the NLPP is to minimize the potential energy of the chain. This is, in turn, the sum of the potential energies of each chain link. The potential energy of chain link  $n$  is given by the product of its mass, the gravitational constant, and the vertical displacement of its midpoint from the ground. The displacement of the midpoint  $n$ th link is given by:

$$y_1 + y_2 + \dots + y_{n-1} + \frac{1}{2}y_n.$$

This is illustrated for the third link in Figure 4.5. Thus, the objective of the NLPP is:

$$\min_{y_1, \dots, y_N} 50g \left[ \frac{1}{2}y_1 + \left( y_1 + \frac{1}{2}y_2 \right) + \dots + \left( y_1 + \dots + y_{N-1} + \frac{1}{2}y_N \right) \right]$$

where  $g$  is the gravitational constant. This objective can be further simplified to:

$$\min_y 50g \sum_{n=1}^N \left( N - n + \frac{1}{2} \right) y_n,$$

by adding and simplifying terms.

This problem has two constraints. The first is to ensure that the vertical displacements of the chain links all sum to zero, meaning that its two anchor points are at the same height:

$$\sum_{n=1}^N y_n = 0.$$

The other constraint ensures that the horizontal displacements of the chain links sum to  $L$ , which ensures that the two anchor points of the chain are  $L$  cm apart. Because we know that chain link  $n$  has a length of 10 cm and a vertical displacement (relative to the  $(n - 1)$ th link) of  $y_n$ , we can use the Pythagorean theorem to compute its horizontal displacement (relative to the  $(n - 1)$ th link) as:

$$\sqrt{10 - y_n^2}.$$

Because this constraint requires that these horizontal displacements sum to  $L$  it can be written as:

$$\sum_{n=1}^N \sqrt{10 - y_n^2} = L.$$

Putting all of this together, the NLPP can be written as:

$$\begin{aligned} \min_y \quad & 50g \sum_{n=1}^N \left( N - n + \frac{1}{2} \right) y_n \\ \text{s.t.} \quad & \sum_{n=1}^N y_n = 0 \\ & \sum_{n=1}^N \sqrt{10 - y_n^2} = L. \end{aligned}$$

### 4.1.3 Planning Examples

#### 4.1.3.1 Return-Maximization Problem

An investor has a sum of money to invest in  $N$  different assets. Asset  $n$  generates a rate of return, which is given by  $r_n$ . Define:

$$\bar{r}_n = \mathbb{E}[r_n],$$

as the expected rate of return of asset  $n$ . The risk of the portfolio is measured by the covariance between the returns of the different assets. Let:

$$\sigma_{n,m} = \sum_{n=1}^N \sum_{m=1}^N \mathbb{E}[(r_n - \mathbb{E}[r_n])(r_m - \mathbb{E}[r_m])],$$

be the covariance between the returns of assets  $n$  and  $m$ . By convention the covariance between the return of asset  $n$  and itself,  $\sigma_{n,n}$ , is equal to  $\sigma_n^2$ , which is the variance of the return of asset  $n$ . The investor would like to determine how to allocate the sum of money to maximize the expected return on investment while limiting the variance of the portfolio to be no more than  $\bar{s}$ .

To formulate this problem, we define  $N$  decision variables,  $w_1, w_2, \dots, w_N$ , where  $w_n$  represents the fraction of the money available that is invested in asset  $n$ .

The objective function is to maximize the expected return on investment:

$$\max_{w_1, \dots, w_N} \sum_{n=1}^N \mathbb{E}[r_n w_n] = \sum_{n=1}^N \mathbb{E}[r_n] w_n = \sum_{n=1}^N \bar{r}_n w_n.$$

There are three types of constraints. First, we must ensure that the portfolio variance is no greater than  $\bar{s}$ . We can compute the variance of the portfolio as:

$$\sum_{n=1}^N \sum_{m=1}^N \mathbb{E}[(r_n w_n - \mathbb{E}[r_n w_n])(r_m w_m - \mathbb{E}[r_m w_m])].$$

Factoring the  $w$ 's out of the expectations gives:

$$\sum_{n=1}^N \sum_{m=1}^N \mathbb{E}[(r_n - \mathbb{E}[r_n])(r_m - \mathbb{E}[r_m])] w_n w_m = \sum_{n=1}^N \sum_{m=1}^N \sigma_{n,m} w_n w_m,$$

where the term on the right-hand side of the equality follows from the definition of the  $\sigma$ 's. Thus, our portfolio-variance constraint is:

$$\sum_{n=1}^N \sum_{m=1}^N \sigma_{n,m} w_n w_m \leq \bar{s}.$$

Next, we must ensure that all of the money is invested, meaning that the  $w$ 's sum to one:

$$\sum_{n=1}^N w_n = 1.$$

We, finally, must ensure that the amounts invested are non-negative:

$$w_n \geq 0, \forall n = 1, \dots, N.$$

Taking these together, our NLPP is:

$$\begin{aligned} \max_w \quad & \sum_{n=1}^N \bar{r}_n w_n \\ \text{s.t.} \quad & \sum_{n=1}^N \sum_{m=1}^N \sigma_{n,m} w_n w_m \leq \bar{s} \\ & \sum_{n=1}^N w_n = 1 \\ & w_n \geq 0, \forall n = 1, \dots, N. \end{aligned}$$

#### 4.1.3.2 Variance-Minimization Problem

Consider the basic setup in the Return-Maximization Problem, which is introduced in Section 4.1.3.1. The investor now wishes to select a portfolio that has minimal variance while achieving at least  $\bar{R}$  as the expected rate of return.

We retain the same decision variables in this problem that we define in the Return-Maximization Problem, which is introduced in Section 4.1.3.1. Specifically, define  $N$  decision variables,  $w_1, w_2, \dots, w_N$ , where  $w_n$  represents the fraction of the money available that is invested in asset  $n$ .

The objective is to minimize portfolio variance. Using the expression for portfolio variance that is derived in Section 4.1.3.1, the objective function is:

$$\min_{w_1, \dots, w_N} \sum_{n=1}^N \sum_{m=1}^N \sigma_{n,m} w_n w_m.$$

We again have three types of constraints. The first ensures that the portfolio achieves the desired minimum expected return. We can, again, use the expression in Section 4.1.3.1 for expected portfolio return to write this constraint as:

$$\sum_{n=1}^N \bar{r}_n w_n \geq \bar{R}.$$

We must also ensure that the full sum of money is invested:

$$\sum_{n=1}^N w_n = 1,$$

and that the amounts invested are non-negative:

$$w_n \geq 0, \forall n = 1, \dots, N.$$

Taking these together, our NLPP is:

$$\begin{aligned} \min_w \quad & \sum_{n=1}^N \sum_{m=1}^N \sigma_{n,m} w_n w_m \\ \text{s.t.} \quad & \sum_{n=1}^N \bar{r}_n w_n \geq \bar{R} \\ & \sum_{n=1}^N w_n = 1 \\ & w_n \geq 0, \forall n = 1, \dots, N. \end{aligned}$$

### 4.1.3.3 Inventory-Planning Problem

A store needs to plan its inventory of three different sizes of T-shirts—small, medium, and large—for the next season. Small T-shirts cost \$1 each, medium ones \$2 each, and large ones \$4 each. Small T-shirts sell for \$10 each, medium ones for \$12, and large ones for \$13. Although the store does not know the demand for the three T-shirt sizes, it forecasts that the demands during the next season for the three are independent and uniformly distributed between 0 and 3000. T-shirts that are unsold at the end of the season must be discarded, which is costless. The store wants to make its ordering decision to maximize its expected profits from T-shirt sales.

We define three decision variables for this problem,  $x_s$ ,  $x_m$ , and  $x_l$ , which denote the number of small, medium, and large T-shirts ordered, respectively.

The objective is to maximize expected profits. To compute revenues from small T-shirt sales, we let  $a_s$  and  $D_s$  denote the number of small T-shirts sold and demanded, respectively. We can then express the expected revenues from small T-shirt sales, as a function of  $x_s$  as:

$$10\mathbb{E}[a_s] = 10(\mathbb{E}[a_s | D_s > x_s] \text{Prob}\{D_s > x_s\} + \mathbb{E}[a_s | D_s \leq x_s] \text{Prob}\{D_s \leq x_s\}).$$

We know that if  $D_s > x_s$ , then the number of shirts sold is  $a_s = x_s$ . Furthermore, because  $D_s$  is uniformly distributed, we know that this occurs with probability:

$$1 - \frac{x_s}{3000}.$$

Otherwise, if  $D_s \leq x_s$ , then the number of shirts sold is  $a_s = D_s$ . We compute the expected sales quantity in this case as:

$$\int_0^{x_s} \frac{1}{3000} D dD = \frac{1}{2} \cdot \frac{1}{3000} x_s^2,$$

because we know the uniformly distributed demand has a density function of  $1/3000$ . Combining these observations, we can compute the expected revenues from small T-shirt sales as:

$$\begin{aligned} 10\mathbb{E}[a_s] &= 10(\mathbb{E}[a_s | D_s > x_s] \text{Prob}\{D_s > x_s\} + \mathbb{E}[a_s | D_s \leq x_s] \text{Prob}\{D_s \leq x_s\}) \\ &= 10 \left[ x_s \cdot \left(1 - \frac{x_s}{3000}\right) + \frac{1}{6000} x_s^2 \right] \\ &= 10 \left( x_s - \frac{x_s^2}{6000} \right). \end{aligned}$$

Because the expected revenues from medium and large T-shirt sales have a similar form, the total expected revenue is given by:

$$10 \left( x_s - \frac{x_s^2}{6000} \right) + 12 \left( x_m - \frac{x_m^2}{6000} \right) + 13 \left( x_l - \frac{x_l^2}{6000} \right).$$

Thus, the objective, which is to maximize expected profits, is given by:

$$\max_{x_s, x_m, x_l} 10 \left( x_s - \frac{x_s^2}{6000} \right) + 12 \left( x_m - \frac{x_m^2}{6000} \right) + 13 \left( x_l - \frac{x_l^2}{6000} \right) - x_s - 2x_m - 4x_l.$$

The only constraints are to ensure that the number of shirts ordered are non-negative and less than 3000 (because there is a maximum possible demand of 3000 units for each type of shirt):

$$0 \leq x_s, x_l, x_m \leq 3000.$$

Thus, our NLPP is:

$$\begin{aligned} \max_x & 10 \left( x_s - \frac{x_s^2}{6000} \right) + 12 \left( x_m - \frac{x_m^2}{6000} \right) + 13 \left( x_l - \frac{x_l^2}{6000} \right) - x_s - 2x_m - 4x_l \\ \text{s.t.} & 0 \leq x_s, x_l, x_m \leq 3000. \end{aligned}$$

## 4.1.4 Energy-Related Examples

### 4.1.4.1 Economic-Dispatch Problem

An electricity transmission network consists of  $N$  nodes. There is a generator, which can produce energy, attached to each node. The cost of producing  $q_n$  MW from node  $n$  over the next hour is given by:

$$c_n(q_n) = a_{n,0} + a_{n,1}q_n + a_{n,2}q_n^2,$$

and the generator at node  $n$  must produce at least  $Q_n^-$  but no more than  $Q_n^+$  MW. Each node also has demand for energy, with  $D_n$  MW of demand over the next hour at node  $n$ . The nodes in the network are connected to each other by transmission lines and we let  $\Omega_n$  denote the set of nodes that are directly connected to node  $n$  by a transmission line. For any node  $n$  and  $m \in \Omega_n$ , the power flow over the line directly connecting nodes  $n$  and  $m$  is equal to:

$$f_{n,m} = Y_{n,m} \sin(\theta_n - \theta_m),$$

where  $Y_{n,m}$  is the (constant) electrical admittance of the line connecting nodes  $n$  and  $m$ , and  $\theta_n$  and  $\theta_m$  are the phase angles of nodes  $n$  and  $m$ , respectively. We have that  $Y_{n,m} = Y_{m,n}$  and we use the sign convention that  $f_{n,m} = -f_{m,n}$  is positive if there is a net power flow from node  $n$  to  $m$ . There is a limit on how much power can flow along each transmission line. Let  $L_{n,m} = L_{m,n} > 0$  denote the flow limit on the line directly connecting node  $n$  to node  $m$ .

The demand at node  $n$  can either be satisfied by the generator at node  $n$  or by power imported from other nodes. We assume that the network is 'lossless,' meaning that the total power produced at the  $N$  nodes must equal the total demand among the  $N$  nodes. The goal of the power system operator is to determine how much to produce from each generator and how to set the power flows and phase angles to serve customers' demands at minimum total cost.

There are three types of variables in this problem. The first are the production levels at each node, which we denote by  $q_1, \dots, q_N$ , letting  $q_n$  be the MWh of energy produced at node  $n$ . The second are the phase angles at each of the nodes, which we denote by  $\theta_1, \dots, \theta_N$ , with  $\theta_n$  representing the phase angle of node  $n$ . The third is the flow on the lines, which we denote by  $f_{n,m}$  for all  $n = 1, \dots, N$  and all  $m \in \Omega_n$ . We let  $f_{n,m}$  denote the flow, in MWh, on the line connecting nodes  $n$  and  $m$ . As noted before, we use the sign convention that if  $f_{n,m} > 0$  this means that there is a net flow from node  $n$  to node  $m$  and  $f_{n,m} < 0$  implies a net flow from node  $m$  to node  $n$ .

The objective is to minimize the total cost of producing energy to serve the system's demand:

$$\min_{q, \theta, f} \sum_{n=1}^N c_n(q) = \sum_{n=1}^N [a_{n,0} + a_{n,1}q_n + a_{n,2}q_n^2].$$

There are five set of constraints in the problem. The first set ensures that the local demand at each node is satisfied by either local supply or energy imported from other nodes:

$$D_n = q_n + \sum_{m \in \Omega_n} f_{m,n}, \forall n = 1, \dots, N.$$

Next, we need to ensure that the total amount generated equals the amount demanded. This constraint arises because the network is assumed to be lossless. Otherwise, without this constraint, energy could either be generated and not consumed anywhere in the network or it could be consumed without having been produced anywhere in the network. This constraint is written as:

$$\sum_{n=1}^N D_n = \sum_{n=1}^N q_n.$$

We also have equalities that define the flow on each line in terms of the phase angles at the end of the line:

$$f_{n,m} = Y_{n,m} \sin(\theta_n - \theta_m), \forall n = 1, \dots, N, m \in \Omega_n.$$

We must also ensure that the flows do not violate their limits:

$$f_{n,m} \leq L_{n,m}, \forall n = 1, \dots, N, m \in \Omega_n,$$

and that the power generated at each node is between its bounds:

$$Q_n^- \leq q_n \leq Q_n^+, \forall n = 1, \dots, N.$$

Thus, our NLPP is:

$$\begin{aligned} & \min_{q, \theta, f} \sum_{n=1}^N [a_{n,0} + a_{n,1}q_n + a_{n,2}q_n^2] \\ & \text{s.t. } D_n = q_n + \sum_{m \in \Omega_n} f_{m,n}, \forall n = 1, \dots, N \\ & \sum_{n=1}^N D_n = \sum_{n=1}^N q_n \\ & f_{n,m} = Y_{n,m} \sin(\theta_n - \theta_m), \forall n = 1, \dots, N, m \in \Omega_n \\ & f_{n,m} \leq L_{n,m}, \forall n = 1, \dots, N, m \in \Omega_n \\ & Q_n^- \leq q_n \leq Q_n^+, \forall n = 1, \dots, N. \end{aligned}$$



One interesting property of this optimization problem, which is worth noting, is that there are decision variables that do not appear in the objective function (specifically, the  $\theta$ 's and  $f$ 's). Despite this,  $\theta$  and  $f$  must be listed as decision variables. If they are not, that would imply that their values are fixed, meaning (in the context of this problem) that the system operator does not have the ability to decide the flows on the transmission lines or the phase angles at the nodes. This would be an overly restrictive problem formulation, given the problem description.

## 4.2 Types of Nonlinear Optimization Problems

In the following sections of this chapter, we concern ourselves with three broad classes of successively more difficult nonlinear optimization problems. Moreover, when analyzing these problems it is always helpful to put them into a standard form. By doing so, we are able to apply the same generic tools to solve these classes of NLPPs. These standard forms are akin to the standard and canonical forms of LPPs, which are introduced in Section 2.2.

We now introduce these three types of optimization problems and refer back to the motivating problems in Section 4.1 to give examples of each. We also use these examples to demonstrate how problems can be converted to these three standard forms.

### 4.2.1 Unconstrained Nonlinear Optimization Problems

An **unconstrained nonlinear optimization problem** has an objective function that is being minimized, but does not have any constraints on what values the decision variables can take. An unconstrained nonlinear optimization problem can be generically written as:

$$\min_{x \in \mathbb{R}^n} f(x),$$

where  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$  is the objective function.

Among the motivating examples given in Section 4.1, the Facility-Location Problem in Section 4.1.1.3 is an example of an unconstrained problem. The objective function of the Facility-Location Problem that is given in Section 4.1.1.3 is already a minimization. Thus, no further work is needed to convert this problem into a standard-form unconstrained problem. As discussed in Section 2.2.2.1, a maximization problem can be converted to minimization problem by multiplying the objective function through by  $-1$ .

### 4.2.2 Equality-Constrained Nonlinear Optimization Problems

An **equality-constrained nonlinear optimization problem** has an objective function that is being minimized and a set of  $m$  equality constraints that have zeros on their right-hand sides. An equality-constrained nonlinear optimization problem can be generically written as:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } h_1(x) &= 0 \\ h_2(x) &= 0 \\ &\vdots \\ h_m(x) &= 0, \end{aligned}$$

where  $f(x)$  is the objective function and  $h_1(x), h_2(x), \dots, h_m(x)$  are the  $m$  equality-constraint functions. The objective and constraint functions map the  $n$ -dimensional vector,  $x$ , to scalar values.

Among the examples that are given in Section 4.1, the Hanging-Chain Problem in Section 4.1.2.2 is an example of an equality-constrained problem. The objective function that is given in Section 4.1.2.2 is already a minimization, thus it does not have to be manipulated to put it into the standard form for an equality-constrained problem. We can convert the constraints into standard form by subtracting all of the terms from one side of the equality. This gives the following standard-form NLPP for the Hanging-Chain Problem:

$$\begin{aligned} \min_{y \in \mathbb{R}^N} f(y) &= 50g \sum_{n=1}^N \left( N - n + \frac{1}{2} \right) y_n \\ \text{s.t. } h_1(y) &= \sum_{n=1}^N y_n = 0 \\ h_2(y) &= \sum_{n=1}^N \sqrt{1 - y_n^2} - L = 0. \end{aligned}$$

### 4.2.3 Equality- and Inequality-Constrained Nonlinear Optimization Problems

An **equality- and inequality-constrained nonlinear optimization problem** has an objective function that is being minimized, a set of  $m$  equality constraints that have zeros on their right-hand sides, and a set of  $r$  less-than-or-equal-to constraints

that have zeros on their right-hand sides. An equality- and inequality-constrained problem can be generically written as:

$$\begin{aligned}
 \min_{x \in \mathbb{R}^n} & f(x) \\
 \text{s.t.} & h_1(x) = 0 \\
 & h_2(x) = 0 \\
 & \vdots \\
 & h_m(x) = 0 \\
 & g_1(x) \leq 0 \\
 & g_2(x) \leq 0 \\
 & \vdots \\
 & g_r(x) \leq 0,
 \end{aligned}$$

where  $f(x)$  is the objective function,  $h_1(x), h_2(x), \dots, h_m(x)$  are the  $m$  equality-constraint functions, and  $g_1(x), g_2(x), \dots, g_r(x)$  are the  $r$  inequality-constraint functions. The objective and constraint functions all map the  $n$ -dimensional vector,  $x$ , to scalar values. Note that one could have a problem with only inequality constraints, *i.e.*,  $m = 0$ , meaning that there are no equality constraints.

All of the other motivating examples given in Section 4.1, that are not categorized as being unconstrained or equality constrained, are examples of equality- and inequality-constrained problems. To demonstrate how a problem can be converted to the generic form, take as an example the Return-Maximization Problem, which is given in Section 4.1.3.1. We convert the maximization to a minimization by multiplying the objective function through by  $-1$ . The constraints are similarly manipulated to yield the standard form, giving the following standard-form NLPP for the Return-Maximization Problem:

$$\begin{aligned}
 \min_{w \in \mathbb{R}^N} f(w) &= - \sum_{n=1}^N \bar{r}_n w_n \\
 \text{s.t.} \quad h_1(w) &= \sum_{n=1}^N w_n - 1 = 0 \\
 g_1(w) &= \sum_{n=1}^N \sum_{m=1}^N \sigma_{n,m} w_n w_m - \bar{s} \leq 0 \\
 g_2(w) &= -w_1 \leq 0 \\
 &\vdots \\
 g_{N+1}(w) &= -w_N \leq 0.
 \end{aligned}$$

### 4.3 Global and Local Minima

When we solve an optimization problem, we want to find a feasible solution that makes the objective as small as possible among all feasible solutions. In other words, we are searching for what is known as a **global minimum**. The difficulty that we encounter is that for most problems, we can only find what are known as **local minima**. These concepts are both defined in the following subsections. As we discuss below, one can find a global minimum by exhaustively searching for all local minima and picking the one that gives the best objective-function value. This is the approach we must take to solve most nonlinear optimization problems.

#### 4.3.1 Global Minima

Given a nonlinear optimization problem, a feasible solution,  $x^*$ , is a **global minimum** if  $f(x^*) \leq f(x)$  for all other feasible values of  $x$ .

Figure 4.6 illustrates the global minimum of a parabolic objective function. Because there are no restriction on what values of  $x$  may be chosen, the example in Figure 4.6 is an unconstrained problem. Clearly,  $x^*$  gives the smallest objective value and satisfies the definition of a global minimum.

**Fig. 4.6** The global minimum of a parabolic objective function

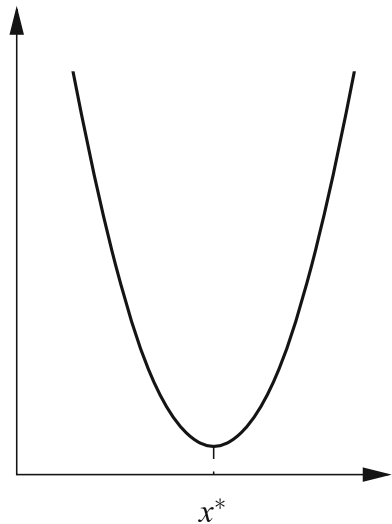
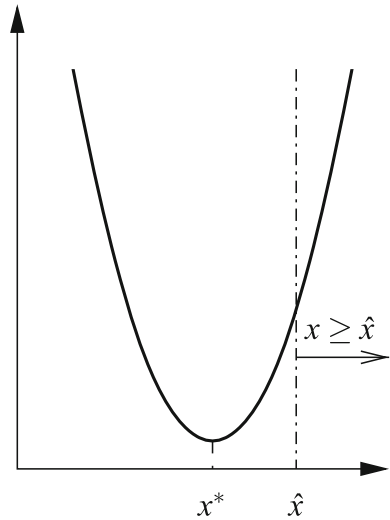


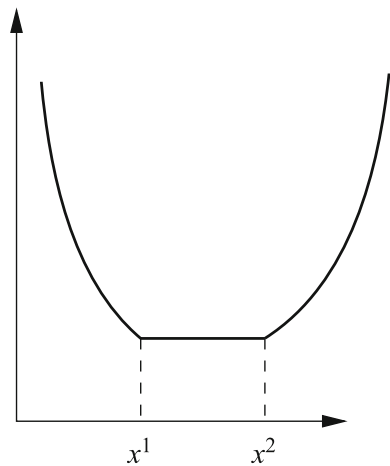
Figure 4.7 demonstrates the effect of adding the inequality constraint,  $x \geq \hat{x}$ , to the optimization problem that is illustrated in Figure 4.6. Although  $x^*$  still gives the smallest objective-function value, it is no longer feasible. Thus, it does not satisfy the definition of a global minimum. One can tell from visual inspection that  $\hat{x}$  is in fact the global minimum of the parabolic objective function when the constraint is added.

**Fig. 4.7** The global minimum of a parabolic objective function with an inequality constraint



It is also important to note that as with a linear or mixed-integer linear optimization problem, a nonlinear problem can have multiple global minima. This is illustrated in Figure 4.8, where we see that all values of  $x$  between  $x^1$  and  $x^2$  are global minima.

**Fig. 4.8** An objective function that has multiple (indeed, an infinite number of) global minima



### 4.3.2 Local Minima

Although we are typically looking for a global minimum when solving an NLPP, for most problems the best we can do is find **local minima**. This is because the methods used to find minima use local information (*i.e.*, derivatives). As discussed below, we find a global minimum of most NLPPs by exhaustively searching for all local minima and choosing the one that gives the smallest objective-function value.

We now define a local minimum and then illustrate the concept with some examples.

Given a nonlinear optimization problem, a feasible solution,  $x^*$ , is a **local minimum** if  $f(x^*) \leq f(x)$  for all other feasible values of  $x$  that are close to  $x^*$ .

Figure 4.9 shows an objective function with four local minima, labeled  $x^1$ ,  $x^2$ ,  $x^3$ , and  $x^4$ . Each of these points satisfies the definition of a local minimum, because other feasible points that are close to them give the same or higher objective-function values. Among the four local minima, one of them,  $x^4$  is also a global minimum. This illustrates the way that we normally go about finding a global minimum. We exhaustively search for all local minima and then choose the one that gives the smallest objective-function value. We must also pay attention to ensure that the problem is not unbounded—if it is then the problem does not have a global minimum (even though it may have local minima).

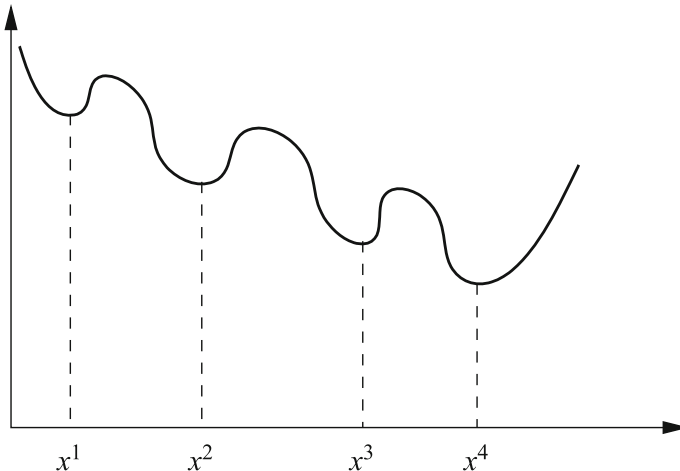


Fig. 4.9 Local and global minima of an objective function

Figure 4.10 demonstrates how adding a constraint affects the definition of a local minimum. Here we have the same objective function as in Figure 4.9, but have added the constraint  $x \geq \hat{x}$ . As in Figure 4.9,  $x^3$  and  $x^4$  are still local minima and  $x^4$  is the global minimum. However,  $x^1$  and  $x^2$  are not local minima, because they are no longer feasible. Moreover,  $\hat{x}$  is a local minimum. To see why, we first note that values of  $x$  that are close to  $\hat{x}$  but to its left give smaller objective-function values than  $\hat{x}$  does. However, these points to the left of  $\hat{x}$  are not considered in the definition of a local minimum, because we only consider *feasible* points that are close to  $\hat{x}$ . If we restrict attention to feasible points that are close to  $\hat{x}$  (*i.e.*, points to the right of  $\hat{x}$ ) then we see that  $\hat{x}$  does indeed satisfy the definition of a local minimum.

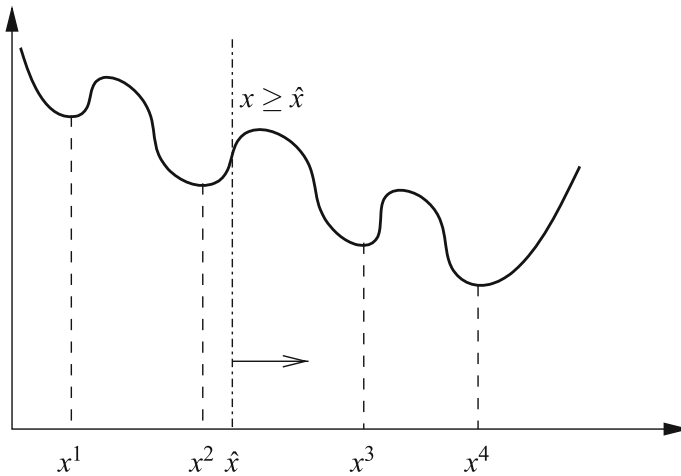


Fig. 4.10 Local and global minima of an objective function with an inequality constraint

The objective function that is shown in Figure 4.8 also demonstrates why the weak inequality in the definition of a local minimum is important. All of the points between  $x^1$  and  $x^2$  in this figure (which we argue in Section 4.3.1 are global minima) are also local minima.

## 4.4 Convex Nonlinear Optimization Problems

We note in Section 4.3 that for most optimization problems, the best we can do is find local minima. Thus, in practice, finding global minima can be tedious because it requires us to search for all local minima and pick the one that gives the smallest objective-function value. There is one special class of optimization problems, which are called **convex optimization problems**, for which it is easier to find a global minimum. A convex optimization problem has the property that any local minimum

is guaranteed to be a global minimum. Thus, finding a global minimum of a convex optimization problem is relatively easy, because we are done as soon as we find a local minimum.

In this section, we first define what a convex optimization problem is and then discuss ways to test whether a problem has the needed convexity property. We finally show the result that any local minimum of a convex optimization problem is guaranteed to be a global minimum.

#### 4.4.1 Convex Optimization Problems

To define a convex optimization problem, we consider a more generic form of an optimization problem than those given in Section 4.2. Here we write a generic optimization problem as:

$$\begin{aligned} \min_x f(x) \\ \text{s.t. } x \in X. \end{aligned}$$

As before,  $f(x)$  is the objective function that we seek to minimize. The set  $X \subseteq \mathbb{R}^n$  represents the feasible region of the problem. In the case of an unconstrained nonlinear optimization problem we would have  $X = \mathbb{R}^n$ . If we have a problem with a mixture of equality and inequality constraints, we define  $X$  as:

$$X = \{x \in \mathbb{R}^n : h_1(x) = 0, \dots, h_m(x) = 0, g_1(x) \leq 0, \dots, g_r(x) \leq 0\},$$

the set of decision-variable vectors,  $x$ , that simultaneously satisfy all of the constraints. Using this more generic form of an optimization problem, we now define a convex optimization problem.

An optimization problem of the form:

$$\begin{aligned} \min_x f(x) \\ \text{s.t. } x \in X \subseteq \mathbb{R}^n, \end{aligned}$$

is a **convex optimization problem** if the set  $X$  is convex and  $f(x)$  is a convex function on the set  $X$ .

Determining if a problem is convex boils down to determining two things: (i) is the feasible region convex and (ii) is the objective function convex on the feasible region? We discuss methods that can be used to answer these two questions in the following sections.



### 4.4.2 Determining if a Feasible Region is Convex

Recall the following definition of a convex set from Section B.1.

A set  $X \subseteq \mathbb{R}^n$  is said to be a **convex set** if for any two points  $x^1$  and  $x^2 \in X$  and for any value of  $\alpha \in [0, 1]$  we have that:

$$\alpha x^1 + (1 - \alpha)x^2 \in X.$$

One way to test whether a set is convex is to use the definition directly, as demonstrated in the following example.

*Example 4.1* Consider the feasible region of the Packing-Box Problem that is introduced in Section 4.1.1.1. The feasible region of this problem is:

$$X = \{(w \ h \ d) : 2wh + 2dh + 6wd \leq 60, w \geq 0, h \geq 0, d \geq 0\}.$$

Note that the points:

$$(w \ h \ d) = (1 \ 1 \ 29/4),$$

and:

$$(\hat{w} \ \hat{h} \ \hat{d}) = (29/4 \ 1 \ 1),$$

are both in the feasible region,  $X$ . However, if we take the midpoint of these two feasible points:

$$(\tilde{w} \ \tilde{h} \ \tilde{d}) = \frac{1}{2}(w \ h \ d) + \frac{1}{2}(\hat{w} \ \hat{h} \ \hat{d}) = (33/8 \ 1 \ 33/8),$$

we see that this point is infeasible because:

$$2\tilde{w}\tilde{h} + 2\tilde{d}\tilde{h} + 6\tilde{w}\tilde{d} \approx 118.59.$$

Thus, the set  $X$  is not convex and the Packing-Box Problem is not a convex optimization problem.  $\square$

In practice, the definition of a convex set can be cumbersome to work with. For this reason, we use the following three properties, which can make it easier to show that a feasible set is convex.

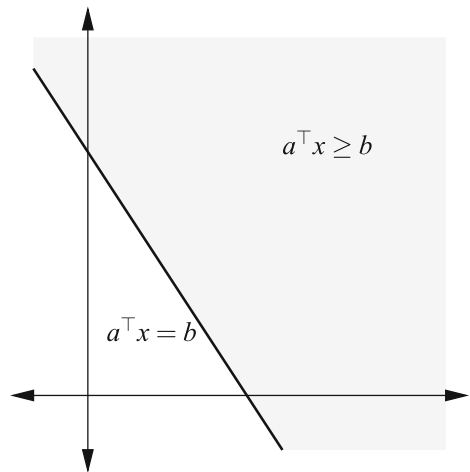
### 4.4.2.1 Linear Constraints

The first property involves equality and inequality constraints that are linear in the decision variables. We can show graphically and through a simple proof that linear constraints always define convex feasible regions.

**Linear-Constraint Property:** Any equality or inequality constraint that is linear in the decision variables defines a convex feasible set.

Before proving this result, Figure 4.11 graphically demonstrates this result. As the figure shows, a linear equality constraint defines a straight line in two dimensions, which is a convex set. We also know from the discussion in Section 2.1.1 that a linear equality constraint defines a hyperplane in three or more dimensions. All of these sets are convex. The figure also demonstrates the convexity of feasible sets defined by linear inequalities. In two dimensions, a linear inequality defines the space on one side of a line (also known as a halfspace). In higher dimensions, halfspaces generalize to the space on one side of a plane or hyperplane. These are also convex sets, as seen in Figure 4.11.

**Fig. 4.11** Illustration of the feasible region defined by linear equality and inequality constraints in two dimensions



We now prove the Linear-Constraint Property.

Consider a linear equality constraint of the form:

$$(a_1 \ a_2 \ \cdots \ a_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = a^\top x = b,$$

where  $b$  is a scalar. Suppose that there are two points,  $x^1$  and  $x^2$ , which are feasible. This means that:

$$a^\top x^1 = b, \quad (4.1)$$

and:

$$a^\top x^2 = b. \quad (4.2)$$

Take any value of  $\alpha \in [0, 1]$  and consider the convex combination of  $x^1$  and  $x^2$ :

$$\alpha x^1 + (1 - \alpha)x^2.$$

If we multiply this point by  $a$  we have:

$$a^\top [\alpha x^1 + (1 - \alpha)x^2] = a^\top \alpha x^1 + a^\top (1 - \alpha)x^2 = \alpha a^\top x^1 + (1 - \alpha)a^\top x^2. \quad (4.3)$$

Substituting (4.1) and (4.2) into (4.3) gives:

$$a^\top [\alpha x^1 + (1 - \alpha)x^2] = \alpha b + (1 - \alpha)b = b,$$

or:

$$a^\top [\alpha x^1 + (1 - \alpha)x^2] = b.$$

Thus,  $\alpha x^1 + (1 - \alpha)x^2$  is feasible and the feasible set defined by the linear equality constraint is convex.

To show the same result for the case of a linear inequality constraint, consider a linear inequality constraint of the form:

$$a^\top x \geq b.$$

Suppose that there are two points,  $x^1$  and  $x^2$ , which are feasible in this inequality constraint. This means that:

$$a^\top x^1 \geq b, \quad (4.4)$$

and:

$$a^\top x^2 \geq b. \quad (4.5)$$

Again, take any value of  $\alpha \in [0, 1]$  and consider the convex combination of  $x^1$  and  $x^2$ ,  $\alpha x^1 + (1 - \alpha)x^2$ . If we multiply this point by  $a$  we have:

$$a^\top [\alpha x^1 + (1 - \alpha)x^2] = \alpha a^\top x^1 + (1 - \alpha)a^\top x^2. \quad (4.6)$$

Substituting (4.4) and (4.5) into (4.6) and noting that  $\alpha \geq 0$  and  $1 - \alpha \geq 0$  gives:

$$a^\top [\alpha x^1 + (1 - \alpha)x^2] \geq \alpha b + (1 - \alpha)b = b,$$

or:

$$a^\top [\alpha x^1 + (1 - \alpha)x^2] \geq b.$$

Thus,  $\alpha x^1 + (1 - \alpha)x^2$  is feasible and the feasible set defined by the linear inequality constraint is convex.

The following example demonstrates how this convexity result involving linear constraints can be used.

*Example 4.2* Consider the constraints of the Variance-Minimization Problem that is introduced in Section 4.1.3.2, which are:

$$\sum_{n=1}^N \bar{r}_n w_n \geq \bar{R},$$

$$\sum_{n=1}^N w_n = 1,$$

and:

$$w_n \geq 0, \forall n = 1, \dots, N.$$

Recall, also, that the decision-variable vector in this problem is  $w$ . Each of these constraints is linear in the decision variables, meaning that each constraint individually defines a convex feasible region.

We do not, yet, know if the whole feasible region of the problem is convex. We only know that each constraint on its own gives a convex set of points that satisfies it. We discuss the Intersection-of-Convex-Sets Property in Section 4.4.2.3, which allows us to draw the stronger conclusion that this problem does indeed have a convex feasible region.  $\square$

#### 4.4.2.2 Convex-Inequality Constraints

The second property that we use to test whether an optimization problem has a convex feasible region is that less-than-or-equal-to constraints that have a convex function on the left-hand side define convex feasible regions. This property is stated as follows.

**Convex-Inequality Property:** Any inequality constraint of the form:

$$g(x) \leq 0,$$

which has a convex function,  $g(x)$ , on the left-hand side defines a convex feasible set.

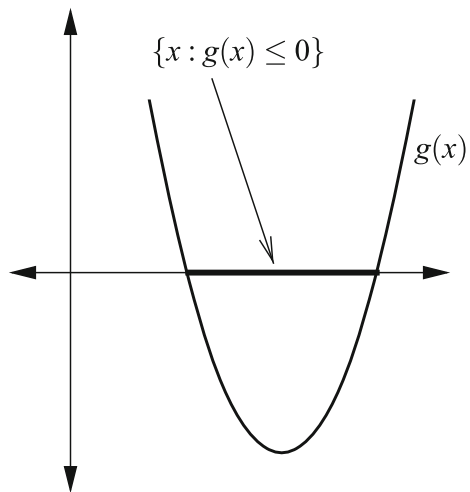
To show this property, recall the following definition of a convex function from Section B.2.

Given a convex set,  $X \subseteq \mathbb{R}^n$ , a function defined on  $X$  is said to be a **convex function** on  $X$  if for any two points,  $x^1$  and  $x^2 \in X$ , and for any value of  $\alpha \in [0, 1]$  we have that:

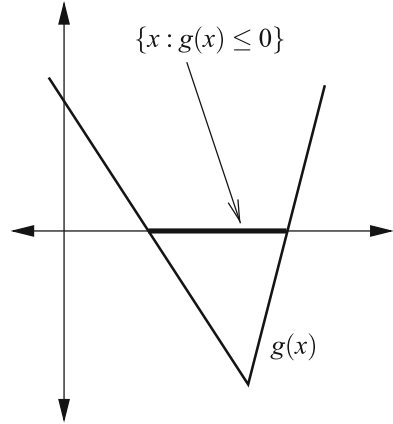
$$\alpha f(x^1) + (1 - \alpha)f(x^2) \geq f(\alpha x^1 + (1 - \alpha)x^2).$$

Figures 4.12 and 4.13 illustrate the Convex-Inequality Property graphically. Figure 4.12 shows the feasible set,  $g(x) \leq 0$ , where  $g(x)$  is a convex parabolic function while Figure 4.13 shows the case of a convex absolute value function. It is important to note that the Convex-Inequality Property only yields a one-sided implication—the feasible set defined by a constraint of the form  $g(x) \leq 0$  where  $g(x)$  is a convex function gives a convex feasible set. However, it may be the case that a constraint of the form  $g(x) \leq 0$  where  $g(x)$  is non-convex gives a convex feasible

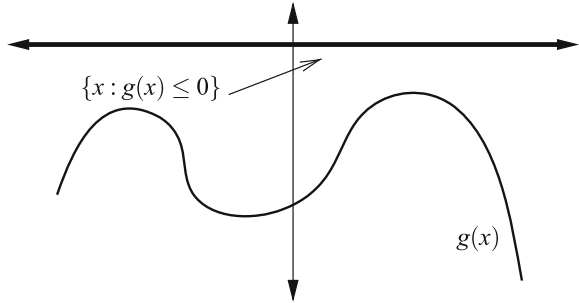
**Fig. 4.12** A convex parabola, which defines a convex feasible region when on the left-hand side of a less-than-or-equal-to constraint



**Fig. 4.13** A convex absolute value function, which defines a convex feasible region when on the left-hand side of a less-than-or-equal-to constraint



**Fig. 4.14** A non-convex function, which defines a convex feasible region when on the left-hand side of a less-than-or-equal to constraint



set. Figure 4.14 demonstrates this by showing a non-convex function,  $g(x)$ , which gives a convex feasible region when on the left-hand side of a less-than-or-equal to constraint, because the feasible region is all  $x \in \mathbb{R}$ .

We now give a proof of the Convex-Inequality Property.

Consider an inequality constraint of the form:

$$g(x) \leq 0,$$

where  $g(x)$  is a convex function. Suppose that there are two feasible points,  $x^1$  and  $x^2$ . This means that:

$$g(x^1) \leq 0, \tag{4.7}$$

and:

$$g(x^2) \leq 0. \tag{4.8}$$

Take  $\alpha \in [0, 1]$  and consider the convex combination of  $x^1$  and  $x^2$ ,  $\alpha x^1 + (1 - \alpha)x^2$ . If we plug this point into the constraint we know that:

$$g(\alpha x^1 + (1 - \alpha)x^2) \leq \alpha g(x^1) + (1 - \alpha)g(x^2), \tag{4.9}$$

because  $g(x)$  is a convex function. Substituting (4.7) and (4.8) into (4.9) and noting that  $\alpha \geq 0$  and  $1 - \alpha \geq 0$  gives:

$$g(\alpha x^1 + (1 - \alpha)x^2) \leq \alpha 0 + (1 - \alpha)0 = 0,$$

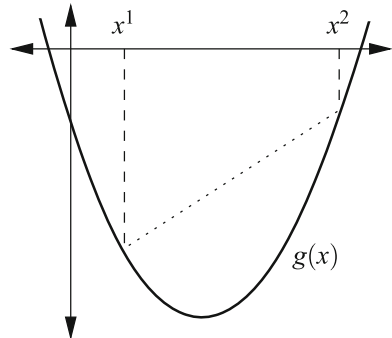
or:

$$g(\alpha x^1 + (1 - \alpha)x^2) \leq 0,$$

meaning that the point  $\alpha x^1 + (1 - \alpha)x^2$  is feasible and the constraint,  $g(x) \leq 0$ , defines a convex feasible set.

Figure 4.15 illustrates the intuition behind the Convex-Inequality Property. If two points,  $x^1$  and  $x^2$ , are feasible, this means that  $g(x^1) \leq 0$  and  $g(x^2) \leq 0$ . Because  $g(x)$  is convex, we know that at any point between  $x^1$  and  $x^2$  the function  $g(x)$  is below the secant line connecting  $g(x^1)$  and  $g(x^2)$  (cf. Section B.2 for further details). However, because both  $g(x^1)$  and  $g(x^2)$  are less than or equal to zero, the secant line is also less than or equal to zero. Hence, the function is also less than or equal to zero at any point between  $x^1$  and  $x^2$ .

**Fig. 4.15** Illustration of the Convex-Inequality Property



### 4.4.2.3 Intersection of Convex Sets

The third property that is useful to show that an optimization problem has a convex feasible region is that the intersection of any number of convex sets is convex. This is useful because the feasible region of an optimization problem is defined as

the intersection of the feasible sets defined by each constraint. If each constraint individually defines a convex set, then the feasible region of the overall problem is convex as well. To more clearly illustrate this, consider the standard-form equality- and inequality-constrained NLPP:

$$\begin{aligned}
 \min_{x \in \mathbb{R}^n} & f(x) \\
 \text{s.t.} & h_1(x) = 0 \\
 & h_2(x) = 0 \\
 & \vdots \\
 & h_m(x) = 0 \\
 & g_1(x) \leq 0 \\
 & g_2(x) \leq 0 \\
 & \vdots \\
 & g_r(x) \leq 0.
 \end{aligned}$$

The overall feasible region of this problem is:

$$X = \{x \in \mathbb{R}^n : h_1(x) = 0, \dots, h_m(x) = 0, g_1(x) \leq 0, \dots, g_r(x) \leq 0\}.$$

Another way to view this feasible region is to define the feasible region defined by each constraint individually as:

$$\begin{aligned}
 X_1 &= \{x \in \mathbb{R}^n : h_1(x) = 0\}, \\
 &\vdots \\
 X_m &= \{x \in \mathbb{R}^n : h_m(x) = 0\}, \\
 X_{m+1} &= \{x \in \mathbb{R}^n : g_1(x) \leq 0\}, \\
 &\vdots
 \end{aligned}$$

and:

$$X_{m+r} = \{x \in \mathbb{R}^n : g_r(x) \leq 0\}.$$

We can then define  $X$  as the intersection of all of these individual sets:

$$X = X_1 \cap \dots \cap X_m \cap X_{m+1} \cap \dots \cap X_{m+r}.$$



If each of the sets,  $X_1, \dots, X_{m+r}$ , is convex, then their intersection,  $X$ , is convex as well, meaning that the problem has a convex feasible region.

We now state and prove the result regarding the intersection of convex sets.

**Intersection-of-Convex-Sets Property:** Let  $X_1, \dots, X_k \subseteq \mathbb{R}^n$  be a collection of convex sets. Their intersection:

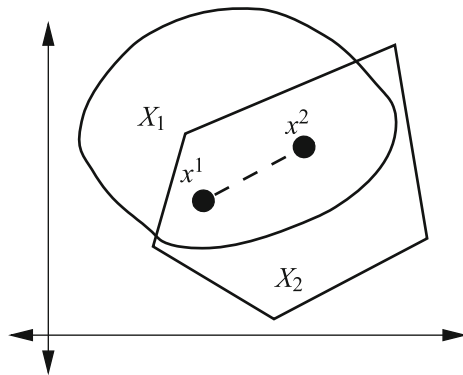
$$X = X_1 \cap \dots \cap X_k,$$

is convex.

We show this by contradiction. Suppose that the statement is not true, meaning that the set  $X$  is not convex. This means that there exist points,  $x^1, x^2 \in X$  and a value of  $\alpha \in [0, 1]$  for which  $\alpha x^1 + (1 - \alpha)x^2 \notin X$ . If  $x^1 \in X$ , then  $x^1$  must be in each of  $X_1, \dots, X_k$ , because  $X$  is defined as the intersection of these sets. Likewise, if  $x^2 \in X$  it must be in each of  $X_1, \dots, X_k$ . Because each of  $X_1, X_2, \dots, X_k$  is convex, then  $\alpha x^1 + (1 - \alpha)x^2$  must be in each of  $X_1, \dots, X_k$ . However, if  $\alpha x^1 + (1 - \alpha)x^2$  is in each of  $X_1, \dots, X_k$  then it must be in  $X$ , because  $X$  is defined as the intersection of  $X_1, \dots, X_k$ . This gives a contradiction, showing that  $X$  must be a convex set.

Figure 4.16 illustrates the idea underlying the proof of the Intersection-of-Convex-Sets Property graphically for the case of the intersection of two convex sets,  $X^1$  and  $X^2$ , in  $\mathbb{R}^2$ . The points  $x^1$  and  $x^2$  are both contained in the sets  $X_1$  and  $X_2$ , thus they are both contained in  $X$ , which is the intersection of  $X_1$  and  $X_2$ . Moreover, if we draw a line segment connecting  $x_1$  and  $x_2$  we know that this line segment must be contained in the set  $X_1$ , because  $X_1$  is convex. The line segment must also be contained in the

**Fig. 4.16** Illustration of the Intersection-of-Convex-Sets Property



set  $X_2$  for the same reason. Because  $X$  is defined as the collection of points that is common to both  $X_1$  and  $X_2$ , this line segment must also be contained in  $X$ , showing that  $X$  is a convex set.

The following example demonstrates the use of the Intersection-of-Convex-Sets Property to show that an optimization problem has a convex feasible region.

*Example 4.3* Consider the following optimization problem:

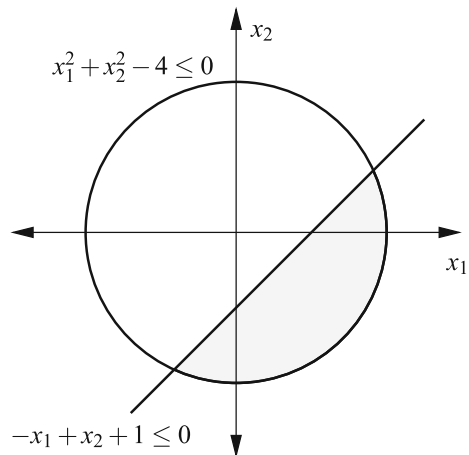
$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g_1(x) = x_1^2 + x_2^2 - 4 \leq 0 \\ & g_2(x) = -x_1 + x_2 + 1 \leq 0, \end{aligned}$$

where the objective,  $f(x)$ , is an arbitrary function. To show that the feasible region of this problem is convex, consider the first constraint. Note that we have:

$$\nabla^2 g_1(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

which is a positive-definite matrix (*cf.* Section A.2 for the definition of positive-definite matrices), meaning that  $g_1(x)$  is a convex function (*cf.* Section B.2 for the use of the Hessian matrix as a means of testing whether a function is convex). Because  $g_1(x)$  is on the left-hand side of a less-than-or-equal-to constraint, it defines a convex set. The second constraint is linear in  $x_1$  and  $x_2$ , thus we know that it defines a convex set. The overall feasible region of the problem is defined by the intersection of the feasible regions defined by each constraint, each of which is convex. Thus, the overall feasible region of the problem is convex. Figure 4.17 illustrates the feasible region defined by each constraint and the overall feasible region of the problem.  $\square$

**Fig. 4.17** Feasible region of optimization problem in Example 4.3



As another example of using the intersection property, recall that in Example 4.2 we note that each constraint of the Variance-Minimization Problem, which is introduced in Section 4.1.3.2, is linear. Thus, each constraint individually defines a convex set. Because the feasible region of the overall problem is defined as the intersection of these sets, the problem's overall feasible region is convex.

### 4.4.3 Determining if an Objective Function is Convex

The definition of a convex function is given in Section B.2. Although it is possible to show that an objective function is convex using this basic definition, it is typically easier to test whether a function is convex by determining whether its Hessian matrix is positive semidefinite. If it is the function is convex, otherwise the function is not (*cf.* Section B.2 for further details). We demonstrate this with the following example.

*Example 4.4* Recall the Packing-Box Problem that is introduced in Section 4.1.1.1. This problem is formulated as:

$$\begin{aligned} \max_{h,w,d} \quad & hwd \\ \text{s.t.} \quad & 2wh + 2dh + 6wd \leq 60 \\ & w \geq 0 \\ & h \geq 0 \\ & d \geq 0. \end{aligned}$$

Converting this problem to standard form it becomes:

$$\begin{aligned} \min_{h,w,d} \quad & f(h, w, d) = -hwd \\ \text{s.t.} \quad & 2wh + 2dh + 6wd - 60 \leq 0 \\ & -w \leq 0 \\ & -h \leq 0 \\ & -d \leq 0. \end{aligned}$$

The Hessian of the objective function is:

$$\nabla^2 f(h, w, d) = \begin{bmatrix} 0 & -d & -w \\ -d & 0 & -h \\ -w & -h & 0 \end{bmatrix},$$

which is not positive semidefinite. Thus, the objective function of this problem is not convex and as such this is not a convex optimization problem.  $\square$

It is important to note that the definition of a convex optimization problem only requires the objective function to be convex on the feasible region. The following example demonstrates why this is important.

*Example 4.5* Consider the unconstrained single-variable optimization problem:

$$\min_x f(x) = \sin(x).$$

We have:

$$\nabla^2 f(x) = -\sin(x),$$

which we know varies in sign for different values of  $x$ . Thus, this unconstrained optimization problem is not convex.

Suppose we add bound constraints and the problem becomes:

$$\begin{aligned} \min_x f(x) &= \sin(x) \\ \text{s.t. } \pi &\leq x \leq 2\pi. \end{aligned}$$

The Hessian of the objective function remains the same. In this case, however, we only require the Hessian to be positive semidefinite over the feasible region (*i.e.*, for values of  $x \in [\pi, 2\pi]$ ). Substituting these values of  $x$  into the Hessian gives us non-negative values. Thus, we have a convex optimization problem when the constraints are added.  $\square$

#### 4.4.4 Global Minima of Convex Optimization Problems

Having defined a convex optimization problem and discussed how to determine if an optimization problem is convex, we now turn to proving the following important Global-Minimum-of-Convex-Problem Property.

**Global-Minimum-of-Convex-Problem Property:** Consider an optimization problem of the form:

$$\begin{aligned} \min_x f(x) \\ \text{s.t. } x \in X \subseteq \mathbb{R}^n, \end{aligned}$$

where  $X$  is a convex set and  $f(x)$  is a convex function on the set  $X$ . Any local minimum of this problem is also a global minimum.

We prove this by contradiction. To do so, suppose that this is not true. That means that there is a point,  $x^*$ , which is a local minimum but not a global minimum. Thus, there is another point,  $\hat{x} \in X$ , which is a global minimum and we have that  $f(\hat{x}) < f(x^*)$ .

Now, consider convex combinations of  $x^*$  and  $\hat{x}$ ,  $\alpha x^* + (1 - \alpha)\hat{x}$ , where  $\alpha \in [0, 1]$ . Because  $X$  is convex, we know such points are feasible in the problem (as this is what it means for the set,  $X$ , to be convex). We also know, because  $f(x)$  is convex, that:

$$f(\alpha x^* + (1 - \alpha)\hat{x}) \leq \alpha f(x^*) + (1 - \alpha)f(\hat{x}).$$

Because  $\hat{x}$  is a global minimum but  $x^*$  is not, we also know that:

$$\alpha f(x^*) + (1 - \alpha)f(\hat{x}) < \alpha f(x^*) + (1 - \alpha)f(x^*) = f(x^*).$$

Combining these two inequalities gives:

$$f(\alpha x^* + (1 - \alpha)\hat{x}) < f(x^*).$$

If we let  $\alpha$  get close to 1, then  $\alpha x^* + (1 - \alpha)\hat{x}$  gets close to  $x^*$ . The last inequality says that these points (obtained for different values of  $\alpha$  close to 1), which are feasible and close to  $x^*$ , give objective-function values that are better than that of  $x^*$ . This contradicts  $x^*$  being a local minimum. Thus, it is not possible for a convex optimization problem to have a local minimum that is not also a global minimum.

## 4.5 Optimality Conditions for Nonlinear Optimization Problems

One method of finding local minima of nonlinear optimization problems is by analyzing what are known as **optimality conditions**. There are two varieties of optimality conditions that we use—necessary and sufficient conditions. **Necessary conditions** are conditions that a solution must satisfy to be a local minimum. A solution that satisfies a necessary condition could possibly be a local minimum. Conversely, a vector of decision-variable values that does not satisfy a necessary condition *cannot* be a local minimum. **Sufficient conditions** are conditions that guarantee that a solution is a local minimum. However, a solution that does not satisfy a sufficient condition *cannot* be ruled out as a possible local minimum.

With the exception of convex optimization problems, we do not typically have conditions that are *both* necessary and sufficient. This means that the best we can typically do is to use necessary conditions to find solution that *might* be local minima. Any of those solutions that also satisfy sufficient conditions are guaranteed to be local minima. However, if there are solutions that only satisfy the necessary but not the sufficient conditions, they may or may not be local minima. There is no way to definitively guarantee one way or another.

We have different sets of optimality conditions for the three types of nonlinear optimization problems that are introduced in Section 4.2—unconstrained, equality-constrained, and equality- and inequality-constrained problems. We examine each of these three problem types in turn.

### 4.5.1 Unconstrained Nonlinear Optimization Problems

An unconstrained nonlinear optimization problem has the general form:

$$\min_{x \in \mathbb{R}^n} f(x),$$

where  $f(x)$  is the objective being minimized and there are no constraints on the decision variables.

#### 4.5.1.1 First-Order Necessary Condition for Unconstrained Nonlinear Optimization Problems

We begin by stating and proving what is known as the first-order necessary condition (FONC) for a local minimum. We also demonstrate its use and limitations through some examples

**First-Order Necessary Condition for Unconstrained Nonlinear Optimization Problems:** Consider an unconstrained nonlinear optimization problem of the form:

$$\min_{x \in \mathbb{R}^n} f(x).$$

If  $x^*$  is a local minimum, then  $\nabla f(x^*) = 0$ .

Suppose  $x^*$  is a local minimum. This means that any point close to  $x^*$  must give an objective-function value that is no less than that given by  $x^*$ . Consider points,  $x^* + d$ , that are close to  $x^*$  (meaning that  $\|d\|$  is close to zero). We can use a first-order Taylor approximation (*cf.* Appendix A) to compute the objective-function value at such a point as:

$$f(x^* + d) \approx f(x^*) + d^\top \nabla f(x^*).$$

Because  $x^*$  is a local minimum, we must have  $f(x^* + d) \geq f(x^*)$ . Substituting the Taylor approximation in for  $f(x^* + d)$  this can be written as:

$$f(x^*) + d^\top \nabla f(x^*) \geq f(x^*).$$

Subtracting  $f(x^*)$  from both sides, this becomes:

$$d^\top \nabla f(x^*) \geq 0. \quad (4.10)$$

This inequality must also apply for the point  $x^* - d$ . That is,  $f(x^* - d) \geq f(x^*)$ . If we substitute the Taylor approximation for  $f(x^* - d)$  into this inequality, we have:

$$d^\top \nabla f(x^*) \leq 0. \quad (4.11)$$

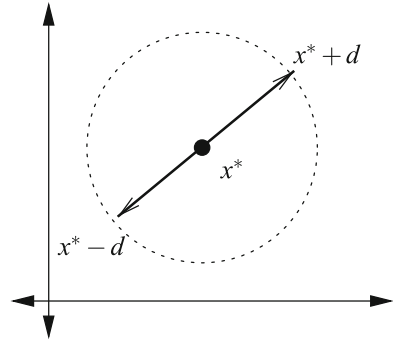
Combining (4.10) and (4.11) implies that we must have:

$$d^\top \nabla f(x^*) = 0.$$

Finally, note that we must have  $d^\top \nabla f(x^*) = 0$  for any choice of  $d$  (so long as  $\|d\|$  is close to zero). The only way that this holds is if  $\nabla f(x^*) = 0$ .

Figure 4.18 graphically illustrates the idea behind the FONC. If  $x^*$  is a local minimum, then there is a small neighborhood of points, which is represented by the area within the dotted circle, on which  $x^*$  provides the best objective-function value. If we examine the Taylor approximation of  $f(x^* + d)$  and  $f(x^* - d)$ , where  $d$  is chosen so  $x^* + d$  and  $x^* - d$  are within this dotted circle, this implies that  $d^\top \nabla f(x^*) = 0$ . Because this must hold for any choice of  $d$  within the dotted circle (*i.e.*, moving in any direction away from  $x^*$ ) this implies that we must have  $\nabla f(x^*) = 0$ .

**Fig. 4.18** Illustration of the FONC for an unconstrained nonlinear optimization problem



Points that satisfy the FONC (*i.e.*, points that have a gradient equal to zero) are also known as **stationary points**. We now demonstrate the use of the FONC with the following examples.

*Example 4.6* Consider the unconstrained problem:

$$\min_x f(x) = (x_1 - 3)^2 + (x_2 + 4)^2.$$

To find stationary points, we set the gradient of  $f(x)$  equal to zero, which gives:

$$\nabla f(x) = \begin{pmatrix} 2(x_1 - 3) \\ 2(x_2 + 4) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

or:

$$x^* = \begin{pmatrix} 3 \\ -4 \end{pmatrix}.$$

It is easy to confirm that this point is in fact a local and global minimum, because the objective function is bounded below by zero. If we plug  $x^*$  into  $f(x)$  we see that this point gives an objective-function value of zero, thus we have a local and global minimum.  $\square$

It is important to note from this example that the FONC always results in a system of  $n$  equations, where  $n$  is the number of decision variables in the optimization problem. This is because the gradient is computed with respect to the decision variables, giving one first-order partial derivative for each decision variable. The fact that the number of equations is equal to the number of variables does not necessarily mean that the FONC has a unique solution. There could be multiple solutions or no solution, as demonstrated in the following examples.

*Example 4.7* Consider the unconstrained problem:

$$\min_x f(x) = x_1 - 2x_2.$$



The gradient of this objective function is:

$$\nabla f(x) = \begin{pmatrix} 1 \\ -2 \end{pmatrix},$$

which cannot be made to equal zero. Because this problem does not have any stationary points, it cannot have any local minima. In fact, this problem is unbounded. To see this, note that:

$$\lim_{x_1 \rightarrow -\infty, x_2 \rightarrow +\infty} f(x) = -\infty. \quad \square$$

*Example 4.8* Consider the unconstrained problem:

$$\min_x f(x) = x^3 - x^2 - 4x - 6.$$

To find stationary points, we set the gradient of  $f(x)$  equal to zero, which gives:

$$\nabla f(x) = 3x^2 - 2x - 4 = 0,$$

or:

$$x^* \in \left\{ \frac{2 - \sqrt{52}}{6}, \frac{2 + \sqrt{52}}{6} \right\}.$$

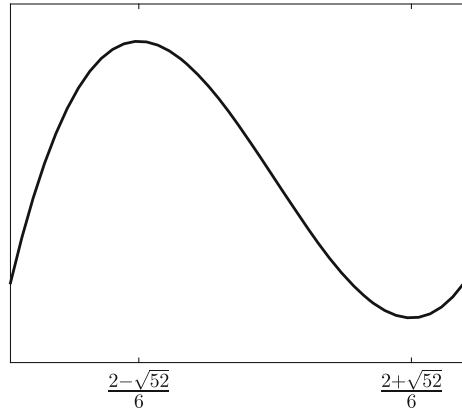
Both of these are stationary points and thus candidate local minima, based on the FONC.

Figure 4.19 shows the objective function of this problem. Based on visual inspection, it is clear that only one of these two stationary points,  $x^* = (2 + \sqrt{52})/6$ , is a local minimum, whereas  $(2 - \sqrt{52})/6$  is a local maximum. Moreover, visual inspection shows us that this objective function is also unbounded. Although  $x^* = (2 + \sqrt{52})/6$  is a local minimum, this particular problem does not have a global minimum because:

$$\lim_{x \rightarrow -\infty} f(x) = -\infty. \quad \square$$

Example 4.8 illustrates an important limitation of the FONC. Although this condition eliminates non-stationary points that cannot be local minima, it does not distinguish between local minima and local maxima. This is also apparent in Figure 4.9—the local minima that are highlighted in this figure are all stationary points where the gradient is zero. However, there are three local maxima, which are also stationary points. The following example further demonstrates this limitation of the FONC.

**Fig. 4.19** Objective function in Example 4.8



*Example 4.9* Consider the unconstrained problem:

$$\max_x f(x),$$

where the objective is an arbitrary function. To solve this problem, we convert it to a minimization of the form:

$$\min_x -f(x),$$

and search for stationary points:

$$\nabla(-f(x)) = -\nabla f(x) = 0.$$

Multiplying through by  $-1$ , the FONC can be written as:

$$\nabla f(x) = 0,$$

which also is the FONC for the following problem:

$$\min_x f(x).$$

This means that the FONCs for finding local minima and local maxima of  $f(x)$  are the same and the FONC cannot distinguish between the two.  $\square$

#### 4.5.1.2 Second-Order Necessary Condition for Unconstrained Nonlinear Optimization Problems

This limitation of the FONC—that it cannot distinguish between local minima and maxima—motivates the second-order necessary condition (SONC) for a local minimum, which we now introduce.

**Second-Order Necessary Condition for Unconstrained Nonlinear Optimization Problems:** Consider an unconstrained nonlinear optimization problem of the form:

$$\min_{x \in \mathbb{R}^n} f(x).$$

If  $x^*$  is a local minimum, then  $\nabla^2 f(x^*)$  is positive semidefinite.

Suppose  $x^*$  is a local minimum. Consider points,  $x^* + d$ , that are close to  $x^*$  (meaning that  $\|d\|$  is close to zero). We can use a second-order Taylor approximation to compute the objective-function value at such a point as:

$$f(x^* + d) \approx f(x^*) + d^\top \nabla f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d.$$

If  $x^*$  is a local minimum and  $d$  is sufficiently small in magnitude we must have:

$$f(x^* + d) \approx f(x^*) + d^\top \nabla f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d \geq f(x^*),$$

or:

$$d^\top \nabla f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d \geq 0. \quad (4.12)$$

We further know from the FONC that if  $x^*$  is a local minimum then we must have  $\nabla f(x^*) = 0$ , thus (4.12) becomes:

$$d^\top \nabla^2 f(x^*) d \geq 0,$$

when we multiply it through by 2. Because this inequality must hold for any choice of  $d$  (so long as  $\|d\|$  is close to zero), this implies that  $\nabla^2 f(x^*)$  is positive semidefinite (cf. Section A.2 for the definition of a positive semidefinite matrix).

The SONC follows very naturally from the FONC, by analyzing the second-order Taylor approximation of  $f(x)$  at points close to a local minimum. The benefit of the SONC is that it can, in many instances, differentiate between local minima and maxima, as demonstrated in the following examples.

*Example 4.10* Consider the following unconstrained optimization problem, which is introduced in Example 4.6:

$$\min_x f(x) = (x_1 - 3)^2 + (x_2 + 4)^2.$$

We know from Example 4.6 that:

$$x^* = \begin{pmatrix} 3 \\ -4 \end{pmatrix},$$

is the only stationary point. We also conclude in Example 4.6 that this point is a local and global minimum, by showing that the objective function attains its lower bound at this point. We can further compute the Hessian of the objective, which is:

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

and is positive semidefinite, confirming that the stationary point found satisfies the SONC.  $\square$

*Example 4.11* Consider the following unconstrained optimization problem, which is introduced in Example 4.8:

$$\min_x f(x) = x^3 - x^2 - 4x - 6.$$

We know from Example 4.8 that this problem has two stationary points:

$$x^* \in \left\{ \frac{2 - \sqrt{52}}{6}, \frac{2 + \sqrt{52}}{6} \right\}.$$

The Hessian of this objective function is:

$$\nabla^2 f(x) = 6x - 2.$$

If we substitute the two stationary points into this Hessian we see that:

$$\nabla^2 f((2 - \sqrt{52})/6) = -\sqrt{52} < 0,$$

and:

$$\nabla^2 f((2 + \sqrt{52})/6) = \sqrt{52} > 0.$$

The Hessian is not positive semidefinite at  $(2 - \sqrt{52})/6$ , thus this point cannot be a local minimum, which is confirmed graphically in Figure 4.19. The Hessian is positive semidefinite at  $(2 + \sqrt{52})/6$ , thus this point remains a candidate local minimum (because it satisfies both the FONC and SONC). We can graphically confirm in Figure 4.19 that this point is indeed a local minimum.  $\square$

*Example 4.12* Consider the unconstrained optimization problem:

$$\min_x f(x) = \sin(x_1 + x_2).$$

The FONC is:

$$\nabla f(x) = \begin{pmatrix} \cos(x_1 + x_2) \\ \cos(x_1 + x_2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which gives stationary points of the form:

$$x^* = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where:

$$x_1 + x_2 \in \left\{ \dots, -\frac{3\pi}{2}, -\frac{\pi}{2}, \frac{\pi}{2}, \frac{3\pi}{2}, \dots \right\}.$$

The Hessian of the objective function is:

$$\nabla^2 f(x) = \begin{bmatrix} -\sin(x_1 + x_2) & -\sin(x_1 + x_2) \\ -\sin(x_1 + x_2) & -\sin(x_1 + x_2) \end{bmatrix}.$$

The determinants of the principal minors of this matrix are  $-\sin(x_1 + x_2)$ ,  $-\sin(x_1 + x_2)$ , and 0. Thus, values of  $x^*$  for which  $-\sin(x_1 + x_2) \geq 0$  satisfy the SONC. Substituting the stationary points found above into  $-\sin(x_1 + x_2)$  gives points of the form:

$$x^* = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

where:

$$x_1 + x_2 \in \left\{ \dots, -\frac{5\pi}{2}, -\frac{\pi}{2}, \frac{3\pi}{2}, \frac{7\pi}{2}, \dots \right\},$$

that satisfy both the FONC and SONC and are candidate local minima. We further know that the sine function is bounded between  $-1$  and  $1$ . Because these values of  $x$  that satisfy both the FONC and SONC make the objective function attain its lower bound of  $-1$ , we know that these points are indeed global and local minima.  $\square$

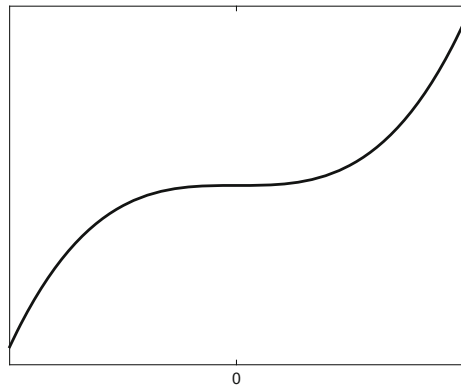
Although the SONC can typically distinguish between local minima and maxima, points that satisfy the FONC and SONC are not necessarily local minima, as demonstrated in the following example.

*Example 4.13* Consider the unconstrained optimization problem:

$$\min_x f(x) = x^3.$$

The FONC is  $\nabla f(x) = 3x^2 = 0$ , which gives  $x^* = 0$  as the unique stationary point. We further have that  $\nabla^2 f(x) = 6x$  and substituting  $x^* = 0$  into the Hessian gives a value of zero, meaning that it is positive semidefinite. Thus, this stationary point also satisfies the SONC. However, visual inspection of the objective function, which is shown in Figure 4.20, shows that this point is not a local minimum. Instead, it is a point of inflection or a saddle point.  $\square$

**Fig. 4.20** Objective function in Example 4.13



The stationary point found in Example 4.13 is an example of a **saddle point**. A saddle point is a stationary point that is neither a local minimum nor maximum. Stationary points that have an indefinite Hessian matrix (*i.e.*, the Hessian is neither positive nor negative semidefinite) are guaranteed to be saddle points. However, Example 4.13 demonstrates that stationary points where the Hessian is positive or negative semidefinite can be saddle points as well.

This limitation of the FONC and the SONC—that they cannot eliminate saddle points—motivates our study of second-order sufficient conditions (SOSCs). A point that satisfies the SOSC is guaranteed to be a local minimum. The limitation of the SOSC, however, is that it is not typically a necessary condition. This means that there could be local minima that do not satisfy the SOSC. Thus, points that satisfy the FONC and SONC but do not satisfy the SOSC are still candidates for being local minima. The one exception to this is a convex optimization problem, in which case we have conditions that are both necessary and sufficient. We begin with a general SOSC that can be applied to any problem. We then consider the special case of convex optimization problems.

### 4.5.1.3 General Second-Order Sufficient Condition for Unconstrained Nonlinear Optimization Problems

We now state and demonstrate the general SOSC that can be applied to any problem.

**General Second-Order Sufficient Condition for Unconstrained Nonlinear Optimization Problems:** Consider an unconstrained nonlinear optimization problem of the form:

$$\min_{x \in \mathbb{R}^n} f(x).$$

If  $x^*$  satisfies  $\nabla f(x^*) = 0$  and  $\nabla^2 f(x^*)$  is positive definite, then  $x^*$  is a local minimum.

Let  $\lambda$  be the smallest eigenvalue of  $\nabla^2 f(x^*)$ . If we consider a point  $x^* + d$  that is close to  $x^*$  (*i.e.*, by choosing a  $d$  with  $\|d\|$  that is close to zero), the objective-function value at this point is approximately:

$$f(x^* + d) \approx f(x^*) + d^\top \nabla f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d.$$

Because  $x^*$  is assumed to be stationary, we can rewrite this as:

$$f(x^* + d) - f(x^*) \approx \frac{1}{2} d^\top \nabla^2 f(x^*) d. \quad (4.13)$$

Using the Quadratic-Form Bound that is discussed in Section A.1, we further have that:

$$\frac{1}{2} d^\top \nabla^2 f(x^*) d \geq \frac{1}{2} \lambda \|d\|^2 > 0, \quad (4.14)$$

where the last inequality follows from  $\nabla^2 f(x^*)$  being positive definite, meaning that  $\lambda > 0$ . Combining (4.13) and (4.14) gives:

$$f(x^* + d) - f(x^*) > 0,$$

meaning that  $x^*$  gives an objective-function value that is strictly better than any point that is close to it. Thus,  $x^*$  is a local minimum.

The SOSC follows a similar line of reasoning to the SONC. By analyzing a second-order Taylor approximation of the objective function at points close to  $x^*$ , we can show that  $x^*$  being stationary and the Hessian being positive definite is sufficient for

$x^*$  to be a local minimum. We demonstrate the use of the SOSC with the following examples.

*Example 4.14* Consider the following unconstrained optimization problem, which is introduced in Example 4.6:

$$\min_x f(x) = (x_1 - 3)^2 + (x_2 + 4)^2.$$

This problem has a single stationary point:

$$x^* = \begin{pmatrix} 3 \\ -4 \end{pmatrix},$$

and the Hessian of the objective function is positive definite at this point. Because this point satisfies the SOSC, it is guaranteed to be a local minimum.  $\square$

*Example 4.15* Consider the following unconstrained optimization problem, which is introduced in Example 4.8:

$$\min_x f(x) = x^3 - x^2 - 4x - 6.$$

This problem has two stationary points:

$$x^* \in \left\{ \frac{2 - \sqrt{52}}{6}, \frac{2 + \sqrt{52}}{6} \right\},$$

and the Hessian of the objective is positive definite at  $(2 + \sqrt{52})/6$ . Because this point satisfies the SOSC, it is guaranteed to be a local minimum, as confirmed in Figure 4.19.  $\square$

As discussed above, a limitation of the SOSC is that it is not a necessary condition for a point to be a local minimum. This means that there can be points that are local minima, yet do not satisfy the SOSC. We demonstrate this with the following example.

*Example 4.16* Consider the following unconstrained optimization problem, which is introduced in Example 4.12:

$$\min_x f(x) = \sin(x_1 + x_2).$$

Points of the form:

$$x^* = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$



where:

$$x_1 + x_2 \in \left\{ \dots, -\frac{5\pi}{2}, -\frac{\pi}{2}, \frac{3\pi}{2}, \frac{7\pi}{2}, \dots \right\},$$

satisfy both the FONC and SONC. However, the determinant of the second leading principal minor of the Hessian of the objective function is zero at these points. Thus, the Hessian is only positive semidefinite (as opposed to positive definite) at these points. As such, these points do not satisfy the SOSC and are not guaranteed to be local minima on the basis of that optimality condition. However, we argue in Example 4.12 that because the objective function attains its lower bound of  $-1$  at these points, they must be local and global minima. Thus, this problem has local minima that do not satisfy the SOSC.  $\square$

#### 4.5.1.4 Second-Order Sufficient Condition for Convex Unconstrained Optimization Problems

In the case of a convex optimization problem, it is much easier to guarantee that a point is a global minimum. This is because the FONC alone is also sufficient for a point to be a global minimum. This means that once we find a stationary point of a convex unconstrained optimization problem, we have a global minimum.

##### **Sufficient Condition for Convex Unconstrained Optimization Problems:**

Consider an unconstrained nonlinear optimization problem of the form:

$$\min_{x \in \mathbb{R}^n} f(x).$$

If the objective function,  $f(x)$ , is convex then the FONC is sufficient for a point to be a global minimum.

A differentiable convex function has the property that the tangent line to the function at any point,  $x^*$ , lies below the function (*cf.* Section B.2 for further details). Mathematically, this means that:

$$f(x) \geq f(x^*) + \nabla f(x^*)^\top (x - x^*), \quad \forall x, x^*.$$

If we pick an  $x^*$  that is a stationary point, then this inequality becomes:

$$f(x) \geq f(x^*), \quad \forall x,$$

which is the definition of a global minimum.

This proposition follows very simply from the properties of a convex function. The basic definition of a convex function is that every secant line connecting two points on the function be above the function. Another definition for a convex function that is once continuously differentiable is that every tangent line to the function is below the function. This gives the inequality in the proof above. If we examine the tangent line at a stationary point, such as that shown in Figure 4.21, the tangent is a horizontal line. Because the function must be above this horizontal line, that means it cannot attain a value that is lower than the value given by the stationary point, which is exactly the definition of a global minimum.

We demonstrate the use of this property with the following example.

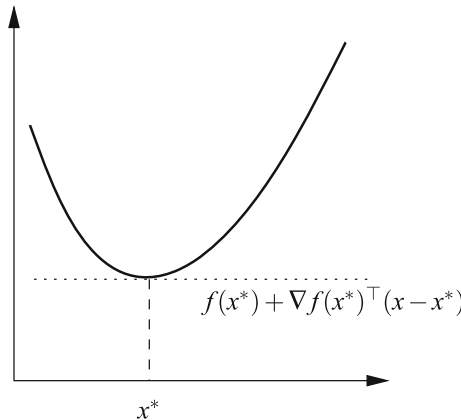
*Example 4.17* Consider the unconstrained nonlinear optimization problem:

$$\min_x f(x) = x_1^2 + x_2^2 + 2x_1x_2.$$

The FONC is:

$$\nabla f(x) = \begin{pmatrix} 2x_1 + 2x_2 \\ 2x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which gives stationary points of the form:



**Fig. 4.21** Illustration of proof of sufficiency of FONC for a convex unconstrained optimization problem

$$x^* = \begin{pmatrix} x \\ -x \end{pmatrix}.$$

We further have that:

$$\nabla^2 f(x) = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix},$$

which is positive semidefinite for all  $x$ . This means that the objective function is convex and this is a convex optimization problem. Thus, the stationary points we have found are global minima, because the FONC is sufficient for finding global minima of a convex optimization problem.  $\square$

Finally, it is worth noting that the stationary points that are found in Example 4.17 do not satisfy the general SOSC, because the Hessian of the objective is not positive definite. Nevertheless, we can conclude that the stationary points found are global minima, because the problem is convex. This, again, demonstrates the important limitation of the SOSC, which is that it is generally only a sufficient and not necessary condition for local minima.

### 4.5.2 Equality-Constrained Nonlinear Optimization Problems

We now examine optimality conditions for equality-constrained nonlinear optimization problems. As with the unconstrained case, we begin by first discussing and demonstrating the use of an FONC for a local minimum. Although the SONC and SOSC for unconstrained problems can be generalized to the constrained case, these conditions are complicated. Thus, we restrict our attention to discussing sufficient conditions for the special case of a convex equality-constrained problem. More general cases of equality-constrained problems are analyzed by Bertsekas [2] and Luenberger and Ye [7].

The FONC that we discuss also has an important technical requirement, known as **regularity**. We omit this regularity requirement from the statement of the FONC and instead defer discussing regularity to Section 4.5.2.2. This is because in many cases the regularity condition is satisfied or does not affect the optimality conditions. However, we provide an example in Section 4.5.2.2 that shows how a local minimum can fail to satisfy the FONC if the regularity condition is not satisfied.

**First-Order Necessary Condition for Equality-Constrained Nonlinear Optimization Problems:** Consider an equality-constrained nonlinear optimization problem of the form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & h_1(x) = 0 \\ & h_2(x) = 0 \\ & \vdots \\ & h_m(x) = 0. \end{aligned}$$

If  $x^*$  is a local minimum of this problem, then there exist  $m$  **Lagrange multipliers**,  $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*$ , such that:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) = 0.$$

The FONC for an equality-constrained problem require us to solve not only for values of the decisions variables in the original problem (*i.e.*, for  $x$ ) but also for an additional set of  $m$  Lagrange multipliers. Note that the number of Lagrange multipliers is always equal to the number of equality constraints in the original problem. We demonstrate the use of the FONC and Lagrange multipliers to find candidate local minima in the following example.

*Example 4.18* Consider the equality-constrained problem:

$$\begin{aligned} \min_x f(x) &= 4x_1^2 + 3x_2^2 + 2x_1x_2 + 4x_1 + 6x_2 + 3 \\ \text{s.t. } h_1(x) &= x_1 - 2x_2 - 1 = 0 \\ h_2(x) &= x_1^2 + x_2^2 - 1 = 0. \end{aligned}$$

To apply the FONC, we define two Lagrange multipliers,  $\lambda_1$  and  $\lambda_2$ , which are associated with the two constraints. The FONC is then:

$$\nabla f(x^*) + \sum_{i=1}^2 \lambda_i^* \nabla h_i(x^*) = 0,$$

or:

$$\begin{pmatrix} 8x_1^* + 2x_2^* + 4 \\ 6x_2^* + 2x_1^* + 6 \end{pmatrix} + \lambda_1^* \begin{pmatrix} 1 \\ -2 \end{pmatrix} + \lambda_2^* \begin{pmatrix} 2x_1^* \\ 2x_2^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Note that this is a system of two equations with four unknowns—the two original decision variables ( $x_1$  and  $x_2$ ) and the two Lagrange multipliers ( $\lambda_1$  and  $\lambda_2$ ). We do have two additional conditions that  $x$  must satisfy, which are the original constraints of the problem. If we add these two constraints, we now have the following system of four equations with four unknowns:

$$\begin{aligned} 8x_1^* + 2x_2^* + 4 + \lambda_1^* + 2\lambda_2^*x_1^* &= 0 \\ 6x_2^* + 2x_1^* + 6 - 2\lambda_1^* + 2\lambda_2^*x_2^* &= 0 \\ x_1^* - 2x_2^* - 1 &= 0 \end{aligned}$$

$$x_1^{*2} + x_2^{*2} - 1 = 0.$$

This system of equations has two solutions:

$$(x_1^*, x_2^*, \lambda_1^*, \lambda_2^*) = (1, 0, 4, -8),$$

and:

$$(x_1^*, x_2^*, \lambda_1^*, \lambda_2^*) = (-3/5, -4/5, 24/25, -6/5).$$

Because these are the only two values of  $x$  and  $\lambda$  that satisfy the constraints of the problem and the FONC, the candidate values of  $x$  that can be local minima are:

$$\begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

and:

$$\begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} -3/5 \\ -4/5 \end{pmatrix}.$$

We know that this problem is bounded, because the feasible region is bounded and the objective function does not asymptote. Thus, we know one of these two candidate points must be a global minimum. If we substitute these values into the objective function we have:

$$f \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 11,$$

and:

$$f \begin{pmatrix} -3/5 \\ -4/5 \end{pmatrix} = \frac{3}{25}.$$

Because it gives a smaller objective-function value, we know that:

$$\begin{pmatrix} x_1^* \\ x_2^* \end{pmatrix} = \begin{pmatrix} -3/5 \\ -4/5 \end{pmatrix},$$

is the global minimum of this problem. □

This example illustrates an important property of the FONC. When we add the original constraints of the problem, the number of equations we have is always equal to the number of unknowns that we solve for. This is because we have  $n + m$  unknowns— $n$  decision variables from the original problem and an additional  $m$  Lagrange multipliers (one for each constraint). We also have  $n + m$  equations. There are  $n$  equations that come directly from the FONC, *i.e.*, the:

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) = 0.$$

This is because the gradient vectors have one partial derivative for each of the  $n$  original problem variables. We also have an additional  $m$  equations that come from the original constraints of the problem.

Note, however, that just as in the unconstrained case, having the same number of equations as unknowns does not imply that there is necessarily a unique solution to the FONC. We could have multiple solutions, as we have in Example 4.18, no solution (which could occur if the problem is infeasible or unbounded), or a unique solution.

Just as in the unconstrained case, the FONC give us candidate solutions that could be local minima. Moreover, points that do not satisfy the FONC *cannot* be local minima. Thus, the FONC typically eliminate many possible points from further consideration. Nevertheless, the FONC cannot necessarily distinguish between local minima, local maxima, and saddle points. The SONC and SOSC for unconstrained problems can be generalized to the equality-constrained case. However the most general second-order conditions are beyond the level of this book. Interested readers are referred to more advanced texts that cover these topics [2, 7]. We, instead, focus on a sufficient condition for the special case of a convex equality-constrained problem, which we now state.

**Sufficient Condition for Convex Equality-Constrained Nonlinear Optimization Problems:** Consider an equality-constrained nonlinear optimization problem of the form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & h_1(x) = 0 \\ & h_2(x) = 0 \\ & \vdots \\ & h_m(x) = 0. \end{aligned}$$

If the constraint functions,  $h_1(x), h_2(x), \dots, h_m(x)$ , are all linear in  $x$  and the objective,  $f(x)$ , is convex on the feasible region then the FONC is sufficient for a point to be a global minimum.

This result follows because an optimization with linear equality constraints and a convex objective function is a convex optimization problem. Convex optimization problems have the property that the FONC is sufficient for a point to be a local minimum [2]. Moreover, we know that any local minimum of a convex problem is a global minimum (*cf.* the Global-Minimum-of-Convex-Problem Property that is discussed in Section 4.4.4). Taken together, these properties give the sufficiency result. We now demonstrate the use of this property with the following example.

*Example 4.19* Consider the equality-constrained problem:

$$\begin{aligned} \min_{x,y,z} f(x) &= (x - 3)^2 - 10 + (2y - 4)^2 - 14 + (z - 6)^2 - 6 \\ \text{s.t. } h_1(x) &= x + y + z - 10 = 0. \end{aligned}$$

To solve this problem we introduce one Lagrange multiplier,  $\lambda_1$ . The FONC and the original constraint of the problem give us the following system of equations:

$$2(x - 3) + \lambda_1 = 0$$

$$4(2y - 4) + \lambda_1 = 0$$

$$2(z - 6) + \lambda_1 = 0$$

$$x + y + z - 10 = 0.$$

The one solution to this system of equations is:

$$(x^* \ y^* \ z^* \ \lambda_1^*) = (23/9 \ 17/9 \ 50/9 \ 8/9).$$

The constraint of this problem is linear and the Hessian of the objective function is:

$$\nabla^2 f(x, y, z) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

which is positive definite, meaning that the objective function is convex. Thus, the solution to the FONC is guaranteed to be a global minimum.  $\square$

It is important to stress that this is only a *sufficient condition*. The problem given in Example 4.18 does not satisfy this condition, because the second constraint is not linear. Nevertheless, we are able to find a global minimum of the problem in that example by appealing to the fact that the problem is bounded and, thus, it must have a well defined global minimum which is also a local minimum. Because the FONC only gives us two candidate points that could be local minima, we know that the one that gives the smallest objective-function value is a global minimum.

#### 4.5.2.1 Geometric Interpretation of the First-Order Necessary Condition for Equality-Constrained Nonlinear Optimization Problems

A general mathematical proof of the FONC for equality-constrained problems is beyond the level of this book (interested readers are referred to more advanced texts [2] for such a proof). We can, however, provide a geometric interpretation of the

FONC for equality-constrained problems. The FONC can be rewritten as:

$$\nabla f(x^*) = - \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*),$$

which says that the gradient of the objective function at a local minimum must be a linear combination of the gradients of the constraint functions.

To understand why this is so, take a simple case of a problem with a single equality constraint and suppose that  $x^*$  is a local minimum of the problem. If so, we know that  $h_1(x^*) = 0$  (i.e.,  $x^*$  is feasible in the equality constraint). Now, consider directions,  $d$ , in which to move away from  $x^*$ . We know that a point,  $x^* + d$ , is feasible if and only if  $h_1(x^* + d) = 0$ . If we suppose that  $\|d\|$  is close to zero, we can estimate the value of the constraint function at this point using a first-order Taylor approximation as:

$$h_1(x^* + d) \approx h_1(x^*) + d^\top \nabla h_1(x^*).$$

Because  $x^*$  is feasible, we have that  $h_1(x^*) = 0$  and the Taylor approximation simplifies to:

$$h_1(x^* + d) \approx d^\top \nabla h_1(x^*).$$

Thus,  $x^* + d$  is feasible so long as  $d^\top \nabla h_1(x^*) = 0$ . Put another way, the only directions in which we can feasibly move away from  $x^*$  are directions that are perpendicular to the gradient of the constraint function.

Let us now consider what effect moving in a feasible direction,  $d$ , away from  $x^*$  would have on the objective-function value. Again, assuming that  $\|d\|$  is close to zero, we can estimate the objective-function value at this point using a first-order Taylor approximation as:

$$f(x^* + d) \approx f(x^*) + d^\top \nabla f(x^*).$$

Examining this Taylor approximation tells us that there are three possible things that can happen to the objective function if we move away from  $x^*$ . One is that  $d^\top \nabla f(x^*) < 0$ , meaning that the objective gets better. Clearly, this cannot happen if  $x^*$  is a local minimum, because that contradicts the definition of a local minimum. Along the same lines, if  $d^\top \nabla f(x^*) > 0$  then we could feasibly move in the direction  $-d$  and improve the objective function. This is because  $-d^\top \nabla h_1(x^*) = 0$ , meaning that this is a feasible direction in which to move away from  $x^*$ , and  $-d^\top \nabla f(x^*) < 0$ , meaning that the objective function improves. Clearly this cannot happen either. The third possibility is that  $d^\top \nabla f(x^*) = 0$ , meaning that the objective remains the same. This is the only possibility that satisfies the requirement of  $x^*$  being a local minimum.

In other words, if  $x^*$  is a local minimum, then we want to ensure that the only directions that we can feasibly move in are perpendicular to the gradient of the objective function. The FONC ensures that this is true, because it forces the gradient of the objective function to be a multiple of the gradient of the constraint. That way



if we have a feasible direction,  $d$ , which has the property that  $d^\top \nabla h_1(x^*) = 0$  then we also have that:

$$d^\top \nabla f(x^*) = -\lambda_1 d^\top \nabla h_1(x^*) = 0,$$

where the first equality comes from the FONC. With more than one constraint, we have to ensure that directions that are feasible to move in with respect to all of the constraints do not give an objective function improvement, which the FONC does.

Figure 4.22 graphically illustrates the FONC for a two-variable single-constraint problem. The figure shows the contour plot of the objective function and a local minimum,  $x^*$ , which gives an objective-function value of 0. The gradient of the constraint function points downward, thus the only directions that we can feasibly move away from  $x^*$  (based on the first-order Taylor approximation) is given by the dashed horizontal line. However, looking at the objective function gradient at this point, we see that these feasible directions we can move in give no change in the objective-function value.

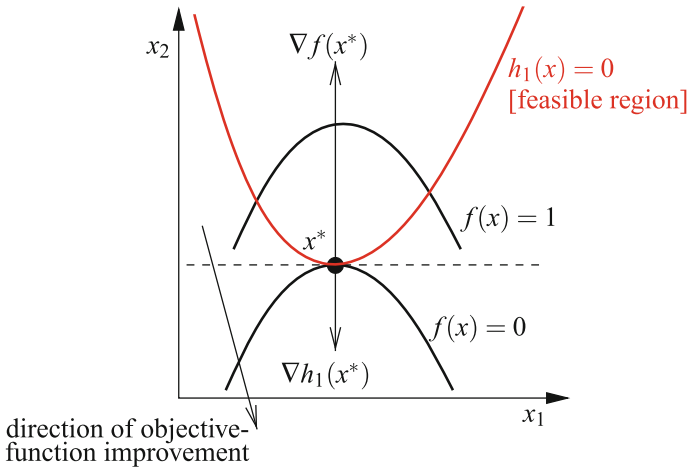
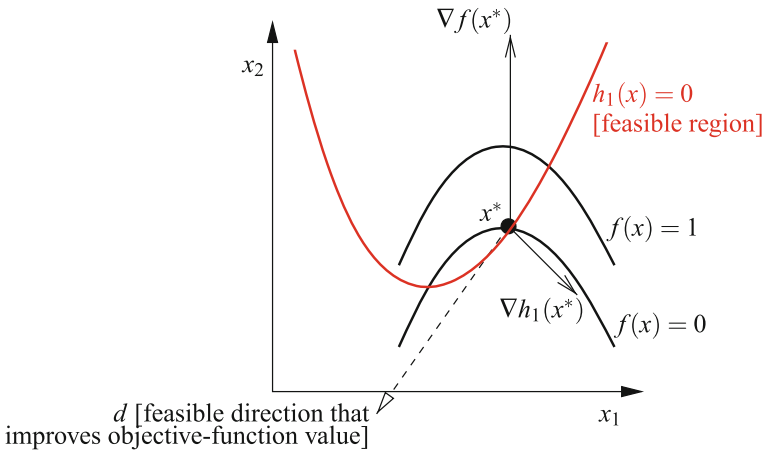


Fig. 4.22 Illustration of the FONC for an equality-constrained problem

Figure 4.23 demonstrates why a point that violates the FONC cannot be a local minimum. In this case,  $\nabla f(x^*)$  is not a multiple of  $\nabla h_1(x^*)$ . Thus, if we move away from  $x^*$  in the direction  $d$ , which is feasible based on the first-order Taylor approximation of the constraint function, the objective function decreases. This violates the definition of a local minimum.



**Fig. 4.23** Illustration of a point for which the FONC for an equality-constrained problem fails

**4.5.2.2 An Added Wrinkle—Regularity**

The FONC for equality-constrained problems has one additional technical requirement, which is known as regularity. A point is said to be **regular** if the gradients of the constraint functions at that point are all linearly independent. As the following example demonstrates, problems can have local minima that do not satisfy the regularity requirement, in which case they may not satisfy the FONC.

*Example 4.20* Consider the equality-constrained problem:

$$\begin{aligned} \min_x \quad & f(x) = 2x_1 + 2x_2 \\ \text{s.t.} \quad & h_1(x) = (x_1 - 1)^2 + x_2^2 - 1 = 0 \\ & h_2(x) = (x_1 + 1)^2 + x_2^2 - 1 = 0. \end{aligned}$$

To apply the FONC to this problem, we define two Lagrange multipliers,  $\lambda_1$  and  $\lambda_2$ , associated with the two constraints. The FONC and constraints of the problem are:

$$\begin{aligned} 2 + 2\lambda_1(x_1 - 1) + 2\lambda_2(x_1 + 1) &= 0 \\ 2 + 2\lambda_1x_2 + 2\lambda_2x_2 &= 0 \\ (x_1 - 1)^2 + x_2^2 - 1 &= 0 \\ (x_1 + 1)^2 + x_2^2 - 1 &= 0. \end{aligned}$$

Simultaneously, solving the two constraints gives  $(x_1, x_2) = (0, 0)$ . Substituting these values into the FONC gives:

$$2 - 2\lambda_1 + 2\lambda_2 = 0$$

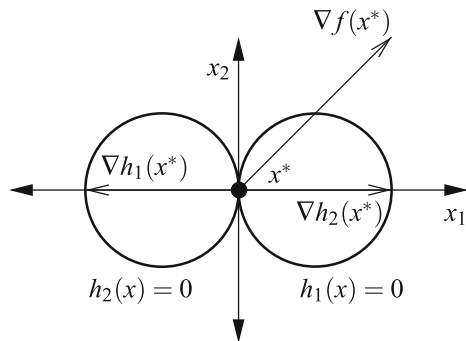
$$2 = 0,$$

which clearly has no solution.

However,  $(x_1, x_2) = (0, 0)$  is the only feasible solution in the constraints, thus it is by definition a global and local minimum. This means that this problem has a local minimum that does not satisfy the FONC. □

Figure 4.24 illustrates why the FONC fails in Example 4.20. The figure shows the feasible regions defined by each of the two constraints and  $x^*$ , which is the unique feasible solution. Because  $x^*$  is the only feasible solution, it is by definition a local and global minimum. However, at this point the gradients of the constraints are the two horizontal vectors shown in the figure, which are not linearly independent. Because the gradient of the objective function is not horizontal, it is impossible to write it as a linear combination of the constraint gradients. This is a consequence of the problem in Example 4.20 having a local minimum that violates the regularity assumption.

**Fig. 4.24** Illustration of why the FONC fails for the equality-constrained problem in Example 4.20



It is worth noting that a point that violates the regularity assumption can still satisfy the FONC. For instance, if the objective function of the problem in Example 4.20 is changed to  $f(x) = 2x_1$ , the gradient of the objective function becomes:

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

This gradient is a horizontal vector and can be written as a linear combination of the constraint function gradients. Interested readers are referred to more advanced texts [2], which discuss two important aspects of this regularity issue. One is a more general version of the FONC that does not require regularity. The other is what is known as constraint-qualification conditions. A problem that satisfies these constraint-qualification conditions are guaranteed to have local minima that satisfy the FONC.

### 4.5.3 Equality- and Inequality-Constrained Nonlinear Optimization Problems

As in the equality-constrained case, we examine equality- and inequality-constrained problems by using a FONC. We then discuss sufficient conditions for the special case of a convex equality- and inequality-constrained problem. Interested readers are referred to more advanced texts [2, 7] for the more general SONC and SOSC for equality- and inequality-constrained problems. As in the equality-constrained case, the FONC for inequality- and equality-constrained problems also have a regularity requirement. We again omit the regularity requirement from the statement of the FONC and instead defer discussion of this requirement to Section 4.5.3.2.

**First-Order Necessary Condition for Equality- and Inequality-Constrained Nonlinear Optimization Problems:** Consider an equality- and inequality-constrained nonlinear optimization problem of the form:

$$\begin{aligned}
 & \min_{x \in \mathbb{R}^n} f(x) \\
 & \text{s.t. } h_1(x) = 0 \\
 & \quad h_2(x) = 0 \\
 & \quad \vdots \\
 & \quad h_m(x) = 0 \\
 & \quad g_1(x) \leq 0 \\
 & \quad g_2(x) \leq 0 \\
 & \quad \vdots \\
 & \quad g_r(x) \leq 0.
 \end{aligned}$$

If  $x^*$  is a local minimum of this problem, then there exist  $(m + r)$  **Lagrange multipliers**,  $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*$  and  $\mu_1^*, \mu_2^*, \dots, \mu_r^*$ , such that:

$$\begin{aligned}
 \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*) &= 0 \\
 \mu_1^* &\geq 0 \\
 \mu_2^* &\geq 0 \\
 &\vdots \\
 \mu_r^* &\geq 0
 \end{aligned}$$

$$\mu_1^* g_1(x^*) = 0$$

$$\mu_2^* g_2(x^*) = 0$$

$$\vdots$$

$$\mu_r^* g_r(x^*) = 0.$$

The FONC for equality-constrained and equality- and inequality-constrained problems are very similar, in that we retain a condition involving the sum of the gradients of the objective and constraint functions and Lagrange multipliers. The FONC for equality- and inequality-constrained problems differ, however, in that the Lagrange multipliers on the *inequality* constraints must be non-negative. The third set of conditions for the equality- and inequality-constrained case is **complementary-slackness**.

To understand the complementary-slackness conditions, we first define what it means for an inequality constraint to be binding as opposed to non-binding at a solution. Note that these definitions follow immediately from analogous definitions given for linear inequality constraints in Section 2.7.6. Consider the inequality constraint:

$$g_j(x) \leq 0.$$

We say that this constraint is **non-binding** at  $x^*$  if:

$$g_j(x^*) < 0.$$

Thus, a non-binding constraint has the property that when we substitute  $x^*$  into it, there is a difference or slack between the two sides of the constraint. Conversely, this constraint is said to be **binding** at  $x^*$  if:

$$g_j(x^*) = 0.$$

Let us now examine the complementary-slackness condition, taking the case of the  $j$ th inequality constraint in the following discussion. The condition requires that:

$$\mu_j^* g_j(x^*) = 0.$$

In other words, we must have  $\mu_j^* = 0$  (*i.e.*, the Lagrange multiplier associated with the  $j$ th constraint is equal to zero),  $g_j(x^*) = 0$  (*i.e.*, the  $j$ th inequality constraint is binding), or both. This complementary-slackness condition is analogous to the complementary-slackness conditions for linear optimization problems, which are introduced in Section 2.7.6.

The complementary-slackness condition is sometimes abbreviated as:

$$\mu_j^* \geq 0 \perp g_j(x_j^*) \leq 0.$$

What this condition says is that  $\mu_j^* \geq 0$  and  $g_j(x_j^*) \leq 0$  (as before). Moreover, the  $\perp$  says that  $\mu_j^*$  must be perpendicular to  $g_j(x_j^*)$ , in the sense that their product is zero. Thus, the FONC for equality- and inequality-constrained problems are often written more compactly as:

$$\begin{aligned} \nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*) &= 0 \\ h_1(x^*) &= 0 \\ h_2(x^*) &= 0 \\ &\vdots \\ h_m(x^*) &= 0 \\ \mu_1^* \geq 0 \perp g_1(x^*) &\leq 0 \\ \mu_2^* \geq 0 \perp g_2(x^*) &\leq 0 \\ &\vdots \\ \mu_r^* \geq 0 \perp g_r(x^*) &\leq 0. \end{aligned}$$

We finally note that the FONC for equality- and inequality-constrained problems are often referred to as the **Karush-Kuhn-Tucker (KKT) conditions**. The KKT conditions are named after the three people who discovered the result. Karush first formulated the KKT conditions in his M.S. thesis. Quite a few years later Kuhn and Tucker rediscovered them independently.

*Example 4.21* Consider the equality- and inequality-constrained problem (that only has inequality constraints):

$$\begin{aligned} \min_x f(x) &= 2x_1^2 + 2x_1x_2 + 2x_2^2 + x_1 + x_2 \\ \text{s.t. } g_1(x) &= x_1^2 + x_2^2 - 9 \leq 0 \\ g_2(x) &= -x_1 + 2x_2 + 1 \leq 0. \end{aligned}$$

To write out the KKT conditions we define two Lagrange multipliers,  $\mu_1$  and  $\mu_2$ , associated with the two inequality constraints. The KKT conditions and the constraints of the original problem are then:

$$4x_1 + 2x_2 + 1 + 2\mu_1x_1 - \mu_2 = 0$$

$$2x_1 + 4x_2 + 1 + 2\mu_1x_2 + 2\mu_2 = 0$$

$$\mu_1 \geq 0 \perp x_1^2 + x_2^2 - 9 \leq 0$$

$$\mu_2 \geq 0 \perp -x_1 + 2x_2 + 1 \leq 0.$$

Note that these conditions are considerably more difficult to work with than the FONC in the unconstrained and equality-constrained cases. This is because we now have a system of equations and inequalities, the latter coming from the inequality constraints and the non-negativity restrictions on the Lagrange multipliers associated with them.

As such, we approach equality- and inequality-constrained problems by conjecturing which of the inequality constraints are binding and non-binding, and then solving the resulting the KKT conditions. We must examine all combinations of binding and non-binding constraints until we find all solutions to the KKT conditions.

With the problem at hand, let us first consider the case in which neither of the inequality constraints are binding. The complementary-slackness conditions then imply that  $\mu_1 = 0$  and  $\mu_2 = 0$ . The gradient conditions are then simplified to:

$$4x_1 + 2x_2 + 1 = 0$$

$$2x_1 + 4x_2 + 1 = 0.$$

Solving the two equations gives:

$$(x_1 \ x_2) = (-1/6 \ -1/6),$$

meaning we have found as a possible KKT point:

$$(x_1 \ x_2 \ \mu_1 \ \mu_2) = (-1/6 \ -1/6 \ 0 \ 0).$$

However, we only found these values of  $x$  and  $\mu$  by assuming which of the constraints are binding and non-binding (to determine the value of  $\mu$ ) and then solving for  $x$  in the gradient conditions. We must still check to ensure that these values satisfy all of the other conditions. If we do so, we see that the second inequality constraint is violated, meaning that this is not a solution to the KKT conditions.

We next consider the case in which the first inequality constraint is binding and the second is non-binding. The complementary-slackness conditions then imply that  $\mu_2 = 0$  whereas we cannot make any determination about  $\mu_1$ . Thus, the gradient conditions become:

$$4x_1 + 2x_2 + 1 + 2\mu_1x_1 = 0$$

$$2x_1 + 4x_2 + 1 + 2\mu_1x_2 = 0.$$

This is a system of two equations with three unknowns. We, however, have one additional equality that  $x$  must satisfy, which is the first inequality constraint. Because we are assuming in this case that this constraint is binding, we impose it as a third equation:

$$x_1^2 + x_2^2 - 9 = 0.$$

Solving this system of equations gives:

$$(x_1 \ x_2 \ \mu_1) \approx (-2.12 \ -2.12 \ -2.76),$$

$$(x_1 \ x_2 \ \mu_1) \approx (1.86 \ -2.36 \ -1.00),$$

and:

$$(x_1 \ x_2 \ \mu_1) \approx (-2.36 \ 1.86 \ -1.00),$$

meaning that:

$$(x_1 \ x_2 \ \mu_1 \ \mu_2) \approx (-2.12 \ -2.12 \ -2.76 \ 0),$$

$$(x_1 \ x_2 \ \mu_1 \ \mu_2) \approx (1.86 \ -2.36 \ -1.00 \ 0),$$

and:

$$(x_1 \ x_2 \ \mu_1 \ \mu_2) \approx (-2.36 \ 1.86 \ -1.00 \ 0),$$

are candidate KKT points. However, because  $\mu_1$  is negative in all three of these vectors, these are not KKT points.

The third case that we examine is the one in which the first inequality constraint is non-binding and the second inequality is binding. The complementary-slackness conditions imply that  $\mu_1 = 0$  whereas we cannot make any determination regarding the value of  $\mu_2$ . Thus, the simplified gradient conditions and the second inequality constraint (which we impose as an equality) are:

$$4x_1 + 2x_2 + 1 - \mu_2 = 0$$

$$2x_1 + 4x_2 + 1 + 2\mu_2 = 0$$

$$-x_1 + 2x_2 + 1 = 0.$$



Solving these three equations gives:

$$(x_1 \ x_2 \ \mu_2) = (1/14 \ -13/28 \ 5/14),$$

meaning that we have:

$$(x_1 \ x_2 \ \mu_1 \ \mu_2) = (1/14 \ -13/28 \ 0 \ 5/14),$$

as a possible solution KKT point. Moreover, when we check the remaining conditions, we find that they are all satisfied, meaning that this is indeed a KKT point.

The last possible case that we examine is the one in which both of the inequality constraints are binding. In this case, complementary slackness does not allow us to fix any of the  $\mu$ 's equal to zero. Thus, we solve the following system of equations:

$$4x_1 + 2x_2 + 1 + 2\mu_1x_1 - \mu_2 = 0$$

$$2x_1 + 4x_2 + 1 + 2\mu_1x_2 + 2\mu_2 = 0$$

$$x_1^2 + x_2^2 - 9 = 0$$

$$-x_1 + 2x_2 + 1 = 0,$$

which has the solutions:

$$(x_1 \ x_2 \ \mu_1 \ \mu_2) \approx (-2.45 \ -1.73 \ -2.66 \ 0.81),$$

and:

$$(x_1 \ x_2 \ \mu_1 \ \mu_2) \approx (2.85 \ 0.93 \ -2.94 \ -2.49).$$

Clearly neither of these are KKT points, because both of them have negative values for  $\mu_1$ .

Thus, the only solution to the KKT conditions and the only candidate point that could be a local minimum is  $(x_1^*, x_2^*) = (1/14, -13/28)$ .  $\square$

Below we give an algorithm for finding KKT points. We first, in Step 2, conjecture which inequalities are binding and non-binding. We next, in Step 3, fix the  $\mu$ 's associated with non-binding constraints equal to zero (due to the complementary-slackness requirement). Next, in Step 4, we solve the system of equations given by the gradient conditions, all of the equality constraints, and any inequality constraints that are assumed to be binding. Inequalities that are assumed to be binding are written as equal-to-zero constraints. We finally check in Step 5 that  $x$  satisfies the inequality constraints that we assume to be non-binding in Step 2 (and which are, thus, ignored when solving for  $x$  and  $\mu$ ). We also check that  $\mu$  is non-negative. If both of these conditions are true, then the values found for  $x$  and  $\mu$  in Step 4 of the algorithm give a KKT point. Otherwise, they do not and the point is discarded from further consideration.

---

**Algorithm for Finding KKT Points**


---

- 1: **procedure** KKT FIND
  - 2:   Fix which inequalities are binding and non-binding
  - 3:   Fix  $\mu$ 's associated with inequalities assumed to be non-binding to zero
  - 4:   Solve system of equations given by gradient conditions, equality constraints, and binding inequalities (written as equalities)
  - 5:   Check that  $x$  satisfies constraints assumed to be non-binding and  $\mu$  is non-negative
  - 6: **end procedure**
- 

Although the Algorithm for Finding KKT Points provides an efficient way to handle the inequalities in the KKT conditions, it is still quite cumbersome. This is because finding all KKT points typically requires the process be repeated for *every* possible combination of binding and non-binding inequality constraints. The problem in Example 4.21 has two inequality constraints, which gives us four cases to check. A problem with  $r$  inequality constraints would require checking  $2^r$  cases. Clearly, even a small problem can require many cases to be tested to find all KKT points. For instance, with  $r = 10$  we must check 1024 cases whereas a 100-inequality problem would require about  $1.27 \times 10^{30}$  cases to be examined.

There are two ways that we can reduce the search process. First, we use the fact that for a convex equality- and inequality-constrained problem, the KKT conditions are sufficient for a global minimum. This implies that as soon as a KKT point is found, we need not search any further. This is because we know the point that we have is a global minimum. The second is that we can use knowledge of a problem's structure to determine if a set of constraints would be binding or not in an optimum. We begin by first discussing the convex case.

**Sufficient Condition for Convex Equality- and Inequality-Constrained Nonlinear Optimization Problems:** Consider an equality- and inequality-constrained nonlinear optimization problem of the form:

$$\begin{aligned}
 \min_{x \in \mathbb{R}^n} \quad & f(x) \\
 \text{s.t.} \quad & h_1(x) = 0 \\
 & h_2(x) = 0 \\
 & \vdots \\
 & h_m(x) = 0 \\
 & g_1(x) \leq 0 \\
 & g_2(x) \leq 0 \\
 & \vdots \\
 & g_r(x) \leq 0.
 \end{aligned}$$

If the equality constraint functions,  $h_1(x), h_2(x), \dots, h_m(x)$ , are all linear in  $x$  and the inequality-constraint functions,  $g_1(x), g_2(x), \dots, g_r(x)$ , and the objective function,  $f(x)$ , are all convex on the feasible region then the KKT condition is sufficient for a point to be a global minimum.

*Example 4.22* Consider the following equality- and inequality-constrained problem, which is given in Example 4.21:

$$\begin{aligned} \min_x \quad & f(x) = 2x_1^2 + 2x_1x_2 + 2x_2^2 + x_1 + x_2 \\ \text{s.t.} \quad & g_1(x) = x_1^2 + x_2^2 - 9 \leq 0 \\ & g_2(x) = -x_1 + 2x_2 + 1 \leq 0. \end{aligned}$$

Note that the Hessian of the objective function is:

$$\nabla^2 f(x) = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix},$$

which is positive definite, meaning that the objective function is a convex function. Moreover, the second inequality constraint is linear, which we know defines a convex feasible region. The Hessian of the first inequality-constraint function is:

$$\nabla^2 g_1(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

which is also positive definite, meaning that this constraint function is also convex. Thus, this problem is convex and any KKT point that we find is guaranteed to be a global minimum. This means that once we find the KKT point  $(x_1, x_2, \mu_1, \mu_2) = (1/14, -13/28, 0, 5/14)$ , we can stop and ignore the fourth case, because we have a global minimum.

It is also worth noting that when we have a convex equality- and inequality-constrained problem, our goal is to find a KKT point as quickly as possible. Because we first try the case in which neither constraint is binding and find that the second constraint is violated, it could make sense to next examine the case in which the second constraint is binding and the first constraint is non-binding (the third case that we examine in Example 4.21). This shortcut—assuming that violated constraints are binding in an optimal solution—will often yield a KKT point more quickly than randomly examining different combinations of binding and non-binding inequality constraints.  $\square$

Another approach to reducing the number of cases to examine in finding KKT points is to use knowledge of a problem's structure to determine if some constraints are binding or non-binding in an optimum. We demonstrate this approach with the following example.

*Example 4.23* Consider the Packing-Box Problem, which is formulated as:

$$\begin{aligned} \max_{h,w,d} \quad & hwd \\ \text{s.t.} \quad & 2wh + 2dh + 6wd \leq 60 \\ & w, h, d \geq 0, \end{aligned}$$

in Section 4.1.1.1.

This problem has four inequality constraints, thus exhaustively checking all combinations of binding and non-binding inequalities would result in examining 16 cases. As opposed to doing this, let us argue that some of the constraints must be binding or non-binding in an optimal solution. We begin by arguing that each of the non-negativity constraints must be non-binding. To see this, note that if any of  $h$ ,  $w$ , or  $d$  equals zero, then we have a box with a volume of  $0 \text{ cm}^3$ . Setting each of  $h$ ,  $w$ , and  $d$  equal to one gives a box with a larger volume and does not violate the restriction on the amount of cardboard that can be used. Thus, a box with a volume of zero cannot be optimal. Knowing that these three constraints must be non-binding in an optimum has reduced the number of cases that we must examine from 16 to two.

We can, further, argue that the first constraint must be binding, which gives us only a single case to examine. To see this, note that if the constraint is non-binding, this means that there is unused cardboard. In such a case, we can increase the value of any one of  $h$ ,  $w$ , or  $d$  by a small amount so as not to violate the  $60 \text{ cm}^2$  restriction, and at the same time increase the volume of the box. Thus, a box that does not use the full  $60 \text{ cm}^2$  of cardboard cannot be optimal. Knowing this, the number of cases that we must examine is reduced to one.

To solve for an optimal solution, we convert the problem to standard form, which is:

$$\begin{aligned} \min_{h,w,d} \quad & f(h, w, d) = -hwd \\ \text{s.t.} \quad & g_1(h, w, d) = 2wh + 2dh + 6wd - 60 \leq 0 \\ & g_2(h, w, d) = -h \leq 0 \\ & g_3(h, w, d) = -w \leq 0 \\ & g_3(h, w, d) = -d \leq 0. \end{aligned}$$

If we assign four Lagrange multipliers,  $\mu_1$ ,  $\mu_2$ ,  $\mu_3$ , and  $\mu_4$ , to the inequality constraints, the KKT conditions are:

$$\begin{aligned} -wd + \mu_1 \cdot (2w + 2d) - \mu_2 &= 0 \\ -hd + \mu_1 \cdot (2h + 6d) - \mu_3 &= 0 \\ -hw + \mu_1 \cdot (2h + 6w) - \mu_4 &= 0 \end{aligned}$$

$$\mu_1 \geq 0$$

$$\mu_2 \geq 0$$

$$\mu_3 \geq 0$$

$$\mu_4 \geq 0$$

$$\mu_1 \cdot (2wh + 2dh + 6wd - 60) = 0$$

$$-\mu_2 h = 0$$

$$-\mu_3 w = 0$$

$$-\mu_4 d = 0$$

$$2wh + 2dh + 6wd - 60 \leq 0$$

$$-h \leq 0$$

$$-w \leq 0$$

$$-d \leq 0.$$

Based on the argument just presented, we must only consider one case in which the first constraint is binding and the others non-binding. The complementary-slackness and gradient conditions and binding constraint give us the following system of equations:

$$-wd + \mu_1 \cdot (2w + 2d) = 0$$

$$-hd + \mu_1 \cdot (2h + 6d) = 0$$

$$-hw + \mu_1 \cdot (2h + 6w) = 0$$

$$2wh + 2dh + 6wd - 60 = 0,$$

which has the solution:

$$(h \ w \ d \ \mu_1) \approx (5.48 \ 1.83 \ 1.83 \ 0.46).$$

Because this problem has a bounded feasible region and the objective does not asymptote, the point:

$$(h \ w \ d) \approx (5.48 \ 1.83 \ 1.83),$$

which is the only candidate local minimum, must be a local and global minimum of this problem.  $\square$

#### 4.5.3.1 Geometric Interpretation of the Karush-Kuhn-Tucker Condition for Equality- and Inequality-Constrained Problems

It is helpful to provide some intuition behind the KKT condition for equality- and inequality-constrained problems. We specifically examine the gradient and complementary-slackness conditions and the sign restriction on Lagrange multipliers for inequality constraints. Thus, we examine the case of an equality- and inequality-constrained problem that only has inequality constraints.

To understand how the KKT condition is derived, consider a two-variable problem with two inequality constraints:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g_1(x) \leq 0 \\ & g_2(x) \leq 0, \end{aligned}$$

and suppose that  $x^*$  is a local minimum of this problem. Figure 4.25 shows the constraints and feasible region of this problem and where  $x^*$  lies in relation to them. As shown in the figure, the first constraint is binding at  $x^*$ . This is because  $x^*$  is on the boundary defined by the first constraint, meaning that there is no slack in the two sides of the constraints. Conversely, the second constraint is non-binding at  $x^*$ . This is because  $x^*$  is not on the boundary of the constraint. Thus, there is slack between the two sides of the constraint. Because  $x^*$  is a local minimum, we know that there is a neighborhood of feasible points around  $x^*$  with the property that  $x^*$  has the smallest objective-function value on this neighborhood. This neighborhood is denoted by the dotted circle centered around  $x^*$  in Figure 4.25. All of the points in the shaded region that are within the dotted circle give objective-function values that are greater than or equal to  $f(x^*)$ .

Now consider the problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g_1(x) = 0. \end{aligned}$$

Figure 4.26 shows the feasible region of this problem and  $x^*$ . This problem has the same objective function as the problem shown in Figure 4.25, but the feasible region differs. Specifically, the binding constraint from the original problem is now an equality constraint and the non-binding constraint is removed. Figure 4.26 shows the same neighborhood of points around  $x^*$ , denoted by the dotted circle. Note that if  $x^*$  gives the best objective-function value in the neighborhood shown in Figure 4.25 then it also gives the best objective-function value in the neighborhood shown in

**Fig. 4.25** A local minimum of a two-variable problem with two inequality constraints

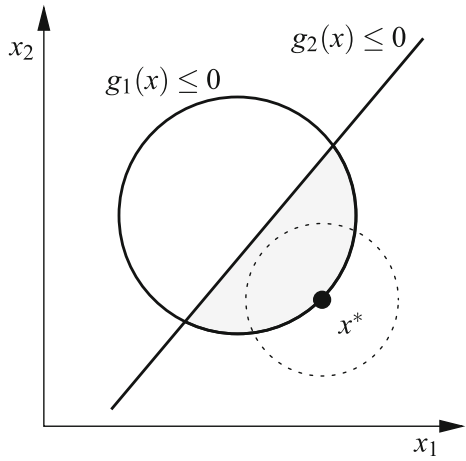
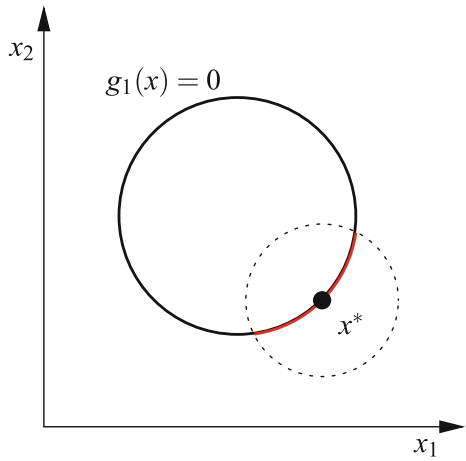


Figure 4.26. This is because the neighborhood in Figure 4.26 has fewer points (only those on the boundary where  $g_1(x) = 0$ , which is highlighted in red in Figure 4.26) and the objective function of the two problems are identical.

**Fig. 4.26** A local minimum of a two-variable problem with one equality constraint that is equivalent to the problem illustrated in Figure 4.25



We, thus, conclude that if  $x^*$  is a local minimum of the problem:

$$\begin{aligned} \min_x & f(x) \\ \text{s.t. } & g_1(x) \leq 0 \\ & g_2(x) \leq 0, \end{aligned}$$

and that only the first constraint is binding at  $x^*$ , then it must also be a local minimum of the problem:

$$\begin{aligned} \min_x & f(x) \\ \text{s.t.} & g_1(x) = 0. \end{aligned}$$

This second problem is an equality-constrained problem. Thus, we can apply the FONC for equality-constrained problems, which is discussed in Section 4.5.2, to it. Doing so gives:

$$\nabla f(x^*) + \mu_1^* g_1(x^*) = 0,$$

where we are letting  $\mu_1^*$  denote the Lagrange multiplier on the equality constraint. If we define  $\mu_2^* = 0$ , we can write this gradient condition as:

$$\nabla f(x^*) + \mu_1^* g_1(x^*) + \mu_2^* g_2(x^*) = \nabla f(x^*) + \sum_{j=1}^r \mu_j^* g_j(x^*) = 0, \quad (4.15)$$

which is the gradient condition we have if we apply the KKT condition to the original equality- and inequality-constrained problem. We further have the complementary-slackness condition that the KKT condition requires. This is because we fix  $\mu_2^* = 0$  when deriving equation (4.15). Note, however, that  $\mu_2^*$  is the Lagrange multiplier on the second inequality constraint. Moreover, the second inequality constraint is the one that is non-binding at the point  $x^*$ , as shown in Figure 4.25.

Thus, the gradient and complementary-slackness requirements of the KKT condition can be derived by applying FONC to the equivalent equality-constrained problem.

We can also provide some intuition around the sign restriction on Lagrange multipliers by conducting this type of analysis. Again, if we take the two-constraint problem:

$$\begin{aligned} \min_x & f(x) \\ \text{s.t.} & g_1(x) \leq 0 \\ & g_2(x) \leq 0, \end{aligned}$$

then the gradient condition is:

$$\nabla f(x^*) + \mu_1^* g_1(x^*) + \mu_2^* g_2(x^*) = \nabla f(x^*) + \mu_1^* g_1(x^*) = 0,$$

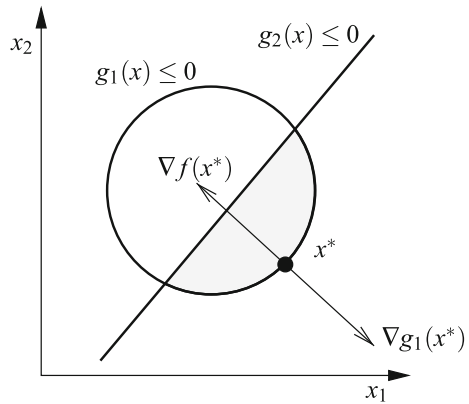
because we are assuming that the second inequality constraint is non-binding and by complementary slackness we have that  $\mu_2^* = 0$ . This condition can be rewritten as:

$$\nabla f(x^*) = -\mu_1^* g_1(x^*).$$



This condition has the same interpretation that is discussed for equality-constrained problems in Section 4.5.2.1. Namely, it says that the gradient of the objective function must be linearly dependent with the gradient of the binding inequality constraint at a local minimum. However, if we further restrict  $\mu_1^* \geq 0$ , then the gradient condition further says that the gradient of the objective function must be a *non-positive* multiple of the gradient of the binding inequality constraint at a local minimum. Figures 4.27 and 4.28 show why this sign restriction is important.

**Fig. 4.27**  $\nabla f(x^*)$  and  $\nabla g_1(x^*)$  of a two-variable problem with two inequality constraints if  $\mu_1^* \geq 0$



**Fig. 4.28**  $\nabla f(x^*)$  and  $\nabla g_1(x^*)$  of a two-variable problem with two inequality constraints if  $\mu_1^* \leq 0$

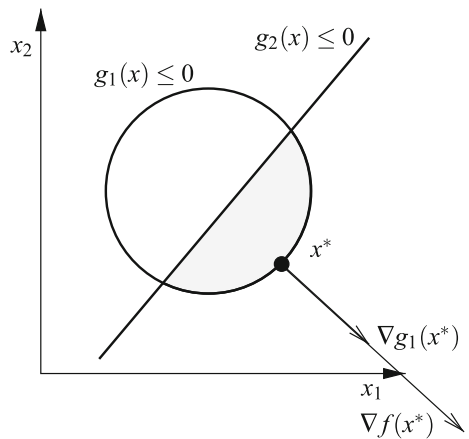


Figure 4.27 shows the gradient of the binding inequality constraint at the local minimum. We know that this gradient points outward from the feasible region, because that is the direction in which  $g_1(x)$  increases. Recall that when we provide a geometric interpretation of the Lagrange multipliers for equality-constrained problems in Section 4.5.2.1, we find that the only directions in which we can move

away from a local minimum are perpendicular to the gradient of the constraint function. This is no longer true when we have inequality constraints. Indeed, we can move in any direction away from  $x^*$  into the shaded region that is shown in Figure 4.27. We know, however, that we *cannot* move away from  $x^*$  in the direction of  $\nabla g_1(x^*)$ . This is because  $g_1(x^*) = 0$  (because the first constraint is binding at  $x^*$ ) and because  $\nabla g_1(x^*)$  is a direction in which  $g_1(x)$  increases. Thus, moving in the direction of  $\nabla g_1(x^*)$  would violate the constraint.

Figure 4.27 also shows that if  $\mu_1^* \geq 0$ , then the gradient of  $f(x^*)$  is pointing inward to the feasible region. This is desirable because we know that  $\nabla f(x^*)$  is a direction in which the objective function increases, meaning that  $-\nabla f(x^*)$  is a direction in which the objective function decreases. To see this, note that if we move in a direction,  $d = -\nabla f(x^*)$ , away from  $x^*$ , the first-order Taylor approximation of the objective function at the new point is:

$$f(x^* + d) = f(x^* - \nabla f(x^*)) \approx f(x^*) - \nabla f(x^*)^\top \nabla f(x^*) < f(x^*).$$

However, because  $-\nabla f(x^*)$  points in the same direction as  $\nabla g_1(x^*)$ , this direction that decreases the objective function is an infeasible direction to move in.

Figure 4.28 also shows the gradient of the binding inequality constraint at the local minimum. It further shows that if  $\mu_1^* \leq 0$ , then  $\nabla f(x^*)$  and  $\nabla g_1(x^*)$  point in the same direction. However,  $x^*$  cannot be a local minimum in this case. The reason is that if we move in the direction of  $-\nabla f(x^*)$  away from  $x^*$  (which is a feasible direction to move in, because  $\nabla f(x^*)$  and  $\nabla g_1(x^*)$  now point in the same direction), the objective function decreases.

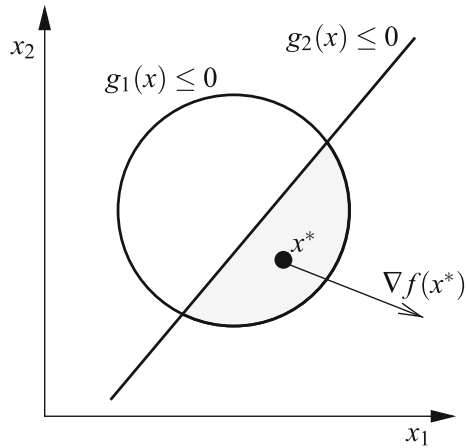
Finally, we can use this same kind of analysis to show the complementary-slackness condition in another way. Figure 4.29 shows a problem with the same feasible region as that shown in Figures 4.25–4.28, however the objective function is now different and the local minimum,  $x^*$ , is interior to both inequality constraints. The gradient condition for this problem would be:

$$\nabla f(x^*) + \mu_1^* g_1(x^*) + \mu_2^* g_2(x^*) = \nabla f(x^*) = 0,$$

because we are assuming that both inequality constraints are non-binding and by complementary slackness we have that  $\mu_1^* = 0$  and  $\mu_2^* = 0$ . In some sense, the complementary-slackness condition says that the gradient condition should ignore the two non-binding inequality constraints and find a point at which the gradient of the objective function is equal to zero. Figure 4.29 shows the logic of this condition by supposing that  $\nabla f(x^*) \neq 0$ . If the gradient of the objective function is as shown in the figure, then  $x^*$  cannot be a local minimum, because moving a small distance in the direction  $d = -\nabla f(x^*)$  away from  $x^*$  (which is a feasible direction to move in) reduces the objective function compared to  $x^*$ . This, however, contradicts the definition of a local minimum.

All of these derivations can be generalized to problems with any number of variables and equality and inequality constraints. However, we focus on problems with two variables and only two inequality constraints to simplify this discussion.

**Fig. 4.29**  $\nabla f(x^*)$  when inequality constraints are non-binding



### 4.5.3.2 Regularity and the Karush-Kuhn-Tucker Condition

When applied to equality- and inequality-constrained problems, the KKT condition has the same regularity requirement that we have with equality-constrained problems. However, the definition of a regular point is slightly different when we have inequality constraints. We say that a point,  $x^*$ , is **regular** in an equality- and inequality-constrained problem if the gradients of the equality constraints and all of the inequality constraints that are *binding* at  $x^*$  are linearly independent.

As with equality-constrained problems, we can have equality- and inequality-constrained problems, such as in Example 4.20, that have local minima that do not satisfy this regularity condition. In such a case, the local minimum may not satisfy the KKT condition. Interested readers are referred to more advanced texts [2] that further discuss this regularity issue.

## 4.6 Sensitivity Analysis

The subject of **sensitivity analysis** is concerned with estimating how changes to a nonlinear optimization problem affect the optimal objective-function value. Thus, this analysis is akin to that carried out in Section 2.6 for linear optimization problems. The following Sensitivity Property explains how this sensitivity analysis is conducted with nonlinear problems.

**Sensitivity Property:** Consider an equality- and inequality-constrained nonlinear optimization problem of the form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & h_1(x) = 0 \\ & h_2(x) = 0 \\ & \vdots \\ & h_m(x) = 0 \\ & g_1(x) \leq 0 \\ & g_2(x) \leq 0 \\ & \vdots \\ & g_r(x) \leq 0. \end{aligned}$$

Suppose  $x^*$  is a local minimum and  $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*, \mu_1^*, \mu_2^*, \dots, \mu_r^*$  are Lagrange multipliers associated with the equality and inequality constraints.

Consider the alternate equality- and inequality-constrained nonlinear optimization problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & h_1(x) = u_1 \\ & h_2(x) = u_2 \\ & \vdots \\ & h_m(x) = u_m \\ & g_1(x) \leq v_1 \\ & g_2(x) \leq v_2 \\ & \vdots \\ & g_r(x) \leq v_r, \end{aligned}$$

and let  $\hat{x}$  be a local minimum of this problem. So long as  $|u_1|, |u_2|, \dots, |u_m|, |v_1|, |v_2|, \dots, |v_r|$  are sufficiently small, we can estimate the objective-function value of the new problem as:

$$f(\hat{x}) \approx f(x^*) - \sum_{i=1}^m \lambda_i^* u_i - \sum_{j=1}^r \mu_j^* v_j.$$

The Sensitivity Property says that the Lagrange multipliers found in the FONC of constrained nonlinear optimization problems provide the same sensitivity information that the sensitivity vector (which are equal to dual variables) give for linear optimization problems. It is also important to stress that although the Sensitivity Property is stated for problems with both equality and inequality constraints, it can clearly be applied to problems with only one type of constraint. A problem with only equality constraints would not have any  $\mu$ 's, because those Lagrange multipliers are associated with inequality constraints. It can similarly be applied to problems with only inequality constraints.

The Sensitivity Property does require the changes to the right-hand sides of the constraints to be small in magnitude, however it does not specify how large a value of  $u$  and  $v$  can be used. This is an unfortunate limitation of the theorem and is a difference compared to sensitivity analysis for linear optimization problems. For linear optimization problems, we can explicitly determine how much the right-hand side of constraints can change before the optimal basis changes using condition (2.49). We have no such result for nonlinear problems.

We now demonstrate the use of the Sensitivity Property with an example.

*Example 4.24* Consider the Packing-Box Problem, which is examined in Example 4.23:

$$\begin{aligned} \max_{h,w,d} \quad & hwd \\ \text{s.t.} \quad & 2wh + 2dh + 6wd \leq 60 \\ & h \geq 0 \\ & w \geq 0 \\ & d \geq 0. \end{aligned}$$

In standard form this problem is:

$$\begin{aligned} \min_{h,w,d} \quad & f(h, w, d) = -hwd \\ \text{s.t.} \quad & g_1(h, w, d) = 2wh + 2dh + 6wd - 60 \leq 0 && (\mu_1) \\ & g_2(h, w, d) = -h \leq 0 && (\mu_2) \\ & g_3(h, w, d) = -w \leq 0 && (\mu_3) \\ & g_4(h, w, d) = -d \leq 0, && (\mu_4) \end{aligned}$$

where the Lagrange multiplier associated with each constraint is indicated in the parentheses to right of it. We know from the analysis in Example 4.23 that:

$$(h \ w \ d \ \mu_1 \ \mu_2 \ \mu_3 \ \mu_4) \approx (5.48 \ 1.83 \ 1.83 \ 0.46 \ 0 \ 0 \ 0),$$

is the unique solution to the KKT condition and is a local and global minimum of the problem.

We wish to know the effect of increasing the amount of cardboard available to  $62\text{ cm}^2$  and requiring the box to be at least  $0.4\text{ cm}$  wide. In other words, we want to estimate the optimal objective-function value of the following problem:

$$\begin{aligned} \max_{h,w,d} \quad & hwd \\ \text{s.t.} \quad & 2wh + 2dh + 6wd \leq 62 \\ & h \geq 0 \\ & w \geq 0.4 \\ & d \geq 0. \end{aligned}$$

To apply the Sensitivity Property to answer this question, we must convert the constraints of this problem to have the same left-hand sides as the standard-form problem that is solved in Example 4.23. This is because the Sensitivity Property only tells us how to estimate the effect of changes to the right-hand side of constraints. The objective function must also be changed to a minimization. We can write the problem with additional cardboard and the minimum-width requirement as:

$$\begin{aligned} \min_{h,w,d} \quad & f(h, w, d) = -hwd \\ \text{s.t.} \quad & g_1(h, w, d) = 2wh + 2dh + 6wd - 60 \leq 2 \\ & g_2(h, w, d) = -h \leq 0 \\ & g_3(h, w, d) = -w \leq -0.4 \\ & g_4(h, w, d) = -d \leq 0. \end{aligned}$$

Applying the Sensitivity Property, we can estimate the new optimal objective-function value as:

$$f(\hat{x}) \approx f(x^*) - 2\mu_1^* - 0\mu_2^* + 0.4\mu_3^* - 0\mu_4^* = -18.26 - 0.92 = -19.18.$$

The Sensitivity Property shows the objective function decreasing when we increase the amount of available cardboard. Recall, however, that the original problem is a maximization. We change the objective function to a minimization by multiplying the objective through by  $-1$  to apply the KKT condition. Thus, when we take this into account, we conclude that the volume of the box *increases* by approximately  $0.92\text{ cm}^3$  when we add the cardboard and impose the minimum-width requirement.  $\square$

### 4.6.1 Further Interpretation of the Karush-Kuhn-Tucker Condition

As a final note, we can use the Sensitivity Property to gain some more insights into the complementary-slackness and sign restrictions in the KKT condition for equality-

and inequality-constrained problems. For this discussion, consider a simple problem with one constraint:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } g_1(x) \leq 0. \end{aligned}$$

Suppose that we have a local minimum,  $x^*$ , and a Lagrange multiplier,  $\mu_1^*$ , that satisfy the KKT condition. Suppose that we change the right-hand side of the constraint so the problem becomes:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) \\ \text{s.t. } g_1(x) \leq v_1, \end{aligned}$$

where  $v_1 < 0$  but  $|v_1|$  is small (*i.e.*, we change the right-hand side to a negative number that is small in magnitude). We can intuitively determine what happens to the objective-function value when we make this change.

First consider the case in which the constraint is non-binding in the original problem. If we change the constraint to  $g_1(x) \leq v_1$  where  $v_1$  is sufficiently small in magnitude, then the same  $x^*$  is still feasible and optimal in the new problem. Thus, the objective-function value does not change at all. The Sensitivity Property tells us that we can estimate the change in the objective-function value as:

$$f(\hat{x}) \approx f(x^*) - \mu_1^* v_1.$$

Because we reasoned that the objective-function value is the same, we must have  $\mu_1^* v_1 = 0$  or  $\mu_1^* = 0$ . This, however, is precisely what complementary slackness requires. Because the constraint is non-binding, the Lagrange multiplier associated with it must be zero. The Sensitivity Property further tells us that changing the right-hand side of a constraint that is not binding by a small amount will not change the optimal objective-function value.

Now, consider the case in which the constraint is binding in the original problem. If we change the constraint to  $g_1(x) \leq v_1$  the objective-function value must get worse (*i.e.*, larger). This is because the feasible region is reduced in size when the right-hand side of the constraint is changed. Before the constraint is changed, solutions for which  $g_1(x) = 0$  are feasible. These solutions are no longer feasible when the right-hand side of the constraint is changed. Thus, the objective function cannot be better when we make this change. Again, the Sensitivity Property tells us that we can estimate the change in the objective-function value as:

$$f(\hat{x}) \approx f(x^*) - \mu_1^* v_1.$$

Because we reasoned that the objective-function value gets worse when we make the change, this means  $-\mu_1^* v_1 \geq 0$  or  $\mu_1^* \geq 0$ , because we have that  $v_1 < 0$ . This, however, is the sign restriction that the KKT condition places on Lagrange

multipliers associated with inequality constraints. Thus, the Sensitivity Property tells us that when we change the right-hand sides of inequality constraints that are binding, the objective function changes in a specific direction.

This interpretation of the complementary-slackness property required by the KKT condition is analogous to the derivation of complementary slackness between a primal linear optimization problem and its dual in Section 2.7.6.

## 4.7 Final Remarks

This chapter introduces analytic methods of solving nonlinear optimization problems. These rely on analyzing optimality conditions, which are a powerful tool for certain types of problems. However, in some cases optimality conditions may yield systems of equations or inequalities that are too difficult to solve. For this reason, iterative solution algorithms, which is the topic of Chapter 5, are often used. These algorithms are implemented in software packages and can be likened to using the Simplex method to solve linear optimization problems.

Our discussion of optimality conditions does not include the more general second-order conditions for constrained problems. Such conditions are beyond the level of this book. Interested readers are referred to more advanced texts for a treatment of such conditions [2, 7]. More advanced texts [1] also provide alternate optimality conditions to the KKT condition that can handle problems that do not satisfy the regularity requirement and specialized treatment of optimality conditions for convex optimization problems [3].

## 4.8 GAMS Codes

This section provides GAMS [4] codes for the main problems considered in this chapter. GAMS can use a variety of different software packages, among them MINOS [8], CONOPT [6], and KNITRO [5], to actually solve an NLPP.

### 4.8.1 Packing-Box Problem

The Packing-Box Problem, which is introduced in Section 4.1.1.1, has the following GAMS formulation:

---

```

1 variable z ;
2 positive variables h, w, d ;
3 equations of, cardBoard ;
4 of .. z =e= h*w*d ;

```



```

5  cardBoard .. 2*w*h+2*h*d+6*w*d =l= 60;
6  model box /all/;
7  solve box using nlp maximizing z;

```

Lines 1 and 2 declare variables, Line 3 gives names to the model equations, Line 4 defines the objective function, Line 5 specifies the constraint, Line 6 defines the model, and Line 7 directs GAMS to solve it.

The GAMS output that provides information about the optimal solution is:

---

		LOWER	LEVEL	UPPER	MARGINAL
3	---- VAR z	-INF	18.257	+INF	.
4	---- VAR h	.	5.477	+INF	.
5	---- VAR w	.	1.826	+INF	.
6	---- VAR d	.	1.826	+INF	.

---

### 4.8.2 Awning Problem

An instance of the Awning Problem, which is introduced in Section 4.1.1.2, which has  $h = 2$  and  $w = 3$  has the following GAMS formulation:

---

```

1  scalars h /2/, w /3/;
2  variable z;
3  positive variables x, y;
4  equations of, box;
5  of .. z =e= sqrt(x**2+y**2);
6  box .. y-w*y/x =g= h;
7  model awning /all/;
8  solve awning using nlp minimizing z;

```

---

Line 1 declares and sets the values of the scalar parameters, Lines 2 and 3 declare variables, Line 4 gives names to the model equations, Line 5 defines the objective function, Line 6 specifies the constraint, Line 7 defines the model, and Line 8 directs GAMS to solve it.

The GAMS output that provides information about the optimal solution is:

---

		LOWER	LEVEL	UPPER	MARGINAL
3	---- VAR z	-INF	7.023	+INF	.
4	---- VAR x	.	5.289	+INF	.
5	---- VAR y	.	4.621	+INF	.

---

### 4.8.3 Facility-Location Problem

An instance of the Facility-Location Problem, which is introduced in Section 4.1.1.3, that has three retail locations at coordinates (1, 1), (-1, 2), and (3, 0) that each receive 10 trucks, has the following GAMS formulation:

---

```

1  set ret /1*3/;
2  parameters
3      V(ret)
4          /1 10
5            2 10
6            3 10/
7      x(ret)
8          /1 1
9            2 -1
10           3 3/
11     y(ret)
12         /1 1
13           2 2
14           3 0/;
15  variables z, a, b;
16  equations of;
17  of .. z =e= sum(ret, 2*V(ret)*sqrt(power((x(ret)-a),2) +
18           power((y(ret)-b),2)));
19  model facLoc /all/;
20  solve facLoc using nlp minimizing z;
```

---

Line 1 declares and defines the set of retail locations. Sets are a construct in GAMS that allow us to create data, variables, or constraints that are assigned to different entities being modeled. For instance, in the Facility-Location Problem each retail location has a pair of coordinates and a fixed number of trucks that must make deliveries to it as model data. The set allows these to be modeled without having to individually write out each piece of data individually in the model. Lines 2–14 declare and set the values of the problem parameters, Line 15 declares variables, Line 16 gives a name to the model equation, Line 17 defines the objective function, Line 18 defines the model, and Line 19 directs GAMS to solve it.

The GAMS output that provides information about the optimal solution is:

---

		LOWER	LEVEL	UPPER	MARGINAL
3	---- VAR z	-INF	89.443	+INF	.
4	---- VAR a	-INF	1.000	+INF	.
5	---- VAR b	-INF	1.000	+INF	.

---

### 4.8.4 Cylinder Problem

An instance of the Cylinder Problem, which is introduced in Section 4.1.1.4, in which  $N = 10$ ,  $c_1 = 2$  and  $c_2 = 0.5$ , has the following GAMS formulation:

---

```

1 scalars N /10/, c1 /2/, c2 /0.5/;
2 variable z;
3 positive variables r, h;
4 equations of;
5 of .. z =e= N*pi*h*r**2-c1*pi*r**2-c2*pi*h*(pi*r**2+2*
      pi*r*h)*r**2;
6 model cyl /all/;
7 solve cyl using nlp maximizing z;

```

---

Line 1 declares and sets the values of scalar parameters, Lines 2 and 3 declare variables, Line 4 gives a name to the model equation, Line 5 defines the objective function, Line 6 defines the model, and Line 7 directs GAMS to solve it.

The GAMS output that provides information about the optimal solution is:

---

		LOWER	LEVEL	UPPER	MARGINAL
3	---- VAR z	- <b>INF</b>	12.174	+ <b>INF</b>	.
4	---- VAR r	.	0.826	+ <b>INF</b>	.
5	---- VAR h	.	1.721	+ <b>INF</b>	.

---

### 4.8.5 Machining-Speed Problem

An instance of the Machining-Speed Problem, that is introduced in Section 4.1.2.1, in which  $p = 10$ ,  $m = 1$ ,  $t_p = 1$ ,  $\lambda = 0.1$ ,  $t_c = 1.1$ ,  $C = 1$ ,  $n = 2$ , and  $h = 0.4$ , has the following GAMS formulation:

---

```

1 scalars p /10/, m /1/, tp /1/, lambda /0.1/, tc /1.1/, C /1/, n
      /2/, h /0.4/;
2 variable z;
3 positive variable v;
4 equations of;
5 of .. z =e= (p-m)/(tp+lambda/v+tc*((C/v)**(1/n))/(lambda/v))-h;
6 model machine /all/;
7 solve machine using nlp maximizing z;

```

---

Line 1 declares and sets the values of scalar parameters, Lines 2 and 3 declare variables, Line 4 gives a name to the model equation, Line 5 defines the objective function, Line 6 defines the model, and Line 7 directs GAMS to solve it.

The GAMS output that provides information about the optimal solution is:

---

		LOWER	LEVEL	UPPER	MARGINAL
1					
3	---- VAR z	- <b>INF</b>	1.286	+ <b>INF</b>	.
4	---- VAR v	.	0.069	+ <b>INF</b>	.

---

### 4.8.6 Hanging-Chain Problem

An instance of the Hanging-Chain Problem, which is introduced in Section 4.1.2.2, in which the chain has 10 links and  $L = 4$ , has the following GAMS formulation:

---

```

1 set links /1*10/;
2 scalars g /9.80665/, L /4/;
3 variables z, y(links);
4 equations of, height, width;
5 of .. z =e= 50*g*sum(links, (card(links)-ord(links)
   +0.5)*y(links));
6 height .. sum(links, y(links)) =e= 0;
7 width .. sum(links, sqrt(10-power(y(links),2))) =e= L;
8 model chain /all/;
9 solve chain using nlp minimizing z;

```

---

Line 1 declares and defines the set of chain links, Line 2 declares and sets the values of scalar parameters, Line 3 declares variables, Line 4 gives names to the model equations, Line 5 defines the objective function, Lines 6 and 7 declare the constraints, Line 8 defines the model, and Line 9 directs GAMS to solve it.

The GAMS output that provides information about the optimal solution is:

---

		LOWER	LEVEL	UPPER	MARGINAL
1					
3	---- VAR z	- <b>INF</b>	-3.859E+4	+ <b>INF</b>	.
5	---- VAR y				
7		LOWER	LEVEL	UPPER	MARGINAL
9	1	- <b>INF</b>	-3.160	+ <b>INF</b>	.
10	2	- <b>INF</b>	-3.158	+ <b>INF</b>	.
11	3	- <b>INF</b>	-3.154	+ <b>INF</b>	.
12	4	- <b>INF</b>	-3.139	+ <b>INF</b>	.
13	5	- <b>INF</b>	-2.969	+ <b>INF</b>	.
14	6	- <b>INF</b>	2.969	+ <b>INF</b>	.
15	7	- <b>INF</b>	3.139	+ <b>INF</b>	.
16	8	- <b>INF</b>	3.154	+ <b>INF</b>	.
17	9	- <b>INF</b>	3.158	+ <b>INF</b>	.
18	10	- <b>INF</b>	3.160	+ <b>INF</b>	.

---

### 4.8.7 Return-Maximization Problem

The Return-Maximization Problem, which is introduced in Section 4.1.3.1, has the following GAMS formulation:

---

```

1  set asset;
2  alias (asset, assetA);
3  parameters ret(asset), cov(asset, asset);
4  scalar s;
5  variable z;
6  positive variable w(asset);
7  equations of, risk, alloc;
8  of .. z =e= sum(asset, ret(asset)*w(asset));
9  risk .. sum((asset, assetA), cov(asset, assetA)*w(asset)*
      w(assetA)) =l= s;
10 alloc .. sum(asset, w(asset)) =e= 1;
11 model invest /all/;
12 solve invest using nlp maximizing z;

```

---

Line 1 declares the set of assets and Line 2 declares an alias of this set. Lines 3 and 4 declare the problem parameters, Lines 5 and 6 declare variables, Line 7 gives names to the model equations, Line 8 defines the objective function, Lines 9 and 10 declare the constraints, Line 11 defines the model, and Line 12 directs GAMS to solve it.

Note that this GAMS code will not compile without values being assigned to the set `asset` and to the parameters `ret(asset)`, `cov(asset, asset)`, and `s`.

### 4.8.8 Variance-Minimization Problem

The Variance-Minimization Problem, which is introduced in Section 4.1.3.2, has the following GAMS formulation:

---

```

1  set asset;
2  alias (asset, assetA);
3  parameters ret(asset), cov(asset, asset);
4  scalar R;
5  variable z;
6  positive variable w(asset);
7  equations of, earn, alloc;
8  of .. z =e= sum((asset, assetA), cov(asset, assetA)*w(asset)*
      w(assetA));
9  earn .. sum(asset, ret(asset)*w(asset)) =g= R;
10 alloc .. sum(asset, w(asset)) =e= 1;
11 model invest /all/;
12 solve invest using nlp minimizing z;

```

---

Line 1 declares the set of assets and Line 2 declares an alias of this set. Lines 3 and 4 declare the problem parameters, Lines 5 and 6 declare variables, Line 7 gives names to the model equations, Line 8 defines the objective function, Lines 9 and 10 declare the constraints, Line 11 defines the model, and Line 12 directs GAMS to solve it.

Note that this GAMS code will not compile without values being assigned to the set `asset` and to the parameters `ret(asset)`, `cov(asset, asset)`, and `R`.

### 4.8.9 Inventory-Planning Problem

The Inventory-Planning Problem, which is introduced in Section 4.1.3.3, has the following GAMS formulation:

---

```

1 variable z ;
2 positive variables xs, xm, xl ;
3 equations of ;
4 of .. z =e= 10*(xs-(xs**2)/6000)+12*(xm-(xm**2)/6000)
      +13*(xl-(xl**2)/6000)-xs-2*xm-4*xl ;
5 model inventory /all/ ;
6 solve inventory using nlp maximizing z ;

```

---

Lines 1 and 2 declare variables, Line 3 gives a name to the model equation, Line 4 defines the objective function, Line 5 defines the model, and Line 6 directs GAMS to solve it.

The GAMS output that provides information about the optimal solution is:

---

		LOWER	LEVEL	UPPER	MARGINAL
3	---- VAR z	- <b>INF</b>	33996.154	+ <b>INF</b>	.
4	---- VAR xs	.	2700.000	+ <b>INF</b>	.
5	---- VAR xm	.	2500.000	+ <b>INF</b>	.
6	---- VAR xl	.	2076.923	+ <b>INF</b>	.

---

### 4.8.10 Economic-Dispatch Problem

The Economic-Dispatch Problem, which is introduced in Section 4.1.4.1, has the following GAMS formulation:

---

```

1 set node ;
2 alias (node, nodeA) ;
3 parameters a0(node), a1(node), a2(node), D(node), Y(node,node),
      link(node,node), L(node,node), minQ(node), maxQ(node) ;
4 variables z, theta(node), f(node,node), q(node) ;
5 positive variable q(node) ;
6 equations of, demand, balance, flow ;

```

---

```

7 of .. z =e= sum(node, a0(node)+a1(node)*q(node)+a2(node)*q(node)
  **2);
8 demand(node) .. D(node) =e= q(node) + sum(nodeA$(link(node,
  nodeA)), f(nodeA, node));
9 balance .. sum(node, D(node)) =e= sum(node, q(node));
10 flow(node, nodeA) .. f(node, nodeA)$(link(node, nodeA)) =e= Y(node
  , nodeA)*sin(theta(node)-theta(nodeA));
11 f.up(node, nodeA) = L(node, nodeA);
12 q.lo(node) = minQ(node);
13 q.up(node) = maxQ(node);
14 model dispatch /all/;
15 solve dispatch using nlp minimizing z;

```

---

Line 1 declares the set of nodes and Line 2 declares an alias of this set. Line 3 declares the problem parameters, Lines 4 and 5 declare variables, Line 6 gives names to the model equations, Line 7 defines the objective function, and Lines 8–10 declare the constraints. Note that the constraint in Line 8 only adds flow from `nodeA` to `node` in determining the supply/demand balance constraint for `node` if the two nodes are directly linked by a transmission line (which is what the parameter `link(node, nodeA)` indicates). Similarly, the constraint in Line 10 only computes the flow on lines that are directly linked. Line 11 imposes the upper bounds on the flow variables and Lines 12 and 13 impose the lower and upper bounds on production at each node. Line 14 defines the model and Line 15 directs GAMS to solve it.

Note that this GAMS code will not compile without values being assigned to the set `node` and to the parameters `a0(node)`, `a1(node)`, `a2(node)`, `D(node)`, `Y(node, node)`, `link(node, node)`, `L(node, node)`, `minQ(node)`, and `maxQ(node)`.

## 4.9 Exercises

**4.1** Jose builds electrical cable using two types of metallic alloys. Alloy 1 is 55% aluminum and 45% copper, while alloy 2 is 75% aluminum and 25% copper. The prices at which Jose can buy the two alloys depends on the amount he purchases. The total cost of buying  $x_1$  tons of alloy 1 is given by:

$$5x_1 + 0.01x_1^2,$$

and the total cost of buying  $x_2$  tons of alloy 2 is given by:

$$4x_2 + 0.02x_2^2.$$

Formulate a nonlinear optimization problem to determine the cost-minimizing quantities of the two alloys that Jose should use to produce 10 tons of cable that is at least 30% copper.

**4.2** Emma is participating in an  $L$  km bicycle race. She is planning on carrying a hydration bladder on her back to keep herself hydrated during the race. If we let  $v$  denote her average speed during the race in km/h and  $w$  the volume of the hydration bladder in liters, then she consumes water at an average rate of  $cv^3 \cdot (w + 1)^2$  liters per hour. Formulate a nonlinear optimization problem to determine how much water Emma should carry and the average speed at which she should bike to minimize her race time.

**4.3** Vishnu has \$35 to spend on any combination of three different goods—apples, oranges, and bananas. Apples cost \$2 each, oranges \$1.50 each, and bananas \$5 each. Vishnu measures his happiness from consuming apples, oranges, and bananas using a utility function. If Vishnu consumes  $x_a$  apples,  $x_o$  oranges, and  $x_b$  bananas, then his utility is given by:

$$3 \log(x_a) + 0.4 \log(x_o + 2) + 2 \log(x_b + 3).$$

Formulate a nonlinear optimization problem to determine how Vishnu should spend his \$35.

**4.4** Convert the models formulated for Exercises 4.1–4.3 into standard form for the type of nonlinear optimization problems that they are.

**4.5** Is the model formulated for the Facility-Location Problem that is introduced in Section 4.1.1.3 a convex optimization problem?

**4.6** Is a local minimum of Exercise 4.1 guaranteed to be a global minimum? Explicitly explain why or why not.

**4.7** What difficulties could arise in applying FONC, SONC, and SOSC to the model formulated for the Facility-Location Problem that is in Section 4.1.1.3?

**4.8** Find all of the KKT points for the model formulated in Exercise 4.3. Are any of these KKT points guaranteed to be global optima? Explicitly explain why or why not.

**4.9** Using the solution to Exercise 4.8, approximate how much Vishnu’s utility increases if he has an additional \$1.25 to spend and must purchase at least one orange. Compare your approximation to the actual change in Vishnu’s utility.

**4.10** Write a GAMS code for the model formulated in Exercise 4.1.

## References

1. Bazaraa MS, Sherali HD, Shetty CM (2006) Nonlinear programming: theory and algorithms, 3rd edn. Wiley-Interscience, Hoboken
2. Bertsekas D (2016) Nonlinear programming, 3rd edn. Athena Scientific, Belmont



3. Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge University Press, Cambridge
4. Brook A, Kendrick D, Meeraus A (1988) GAMS – a user’s guide. ACM, New York. [www.gams.com](http://www.gams.com)
5. Byrd RH, Nocedal J, Waltz RA (2006) KNITRO: an integrated package for nonlinear optimization. In: di Pillo G, Roma M (eds) Large-scale nonlinear optimization, pp 35–59. [www.artelys.com/en/optimization-tools/knitro](http://www.artelys.com/en/optimization-tools/knitro)
6. Drud AS (1994) CONOPT – a large-scale GRG code. ORSA J Comput 6(2):207–216. [www.conopt.com](http://www.conopt.com)
7. Luenberger DG, Ye Y (2016) Linear and nonlinear programming, 4th edn. Springer, New York
8. Murtagh BA, Saunders MA (1995) MINOS 5.4 user’s guide. Technical report SOL 83-20R, Systems Optimization Laboratory, Department of Operations Research, Stanford University, Stanford, California. [www.sbsi-sol-optimize.com/asp/sol\\_product\\_minos.htm](http://www.sbsi-sol-optimize.com/asp/sol_product_minos.htm)