CHAPTER 4

Modeling with Nonlinear Programming

By nonlinear programming we intend the solution of the general class of problems that can be formulated as

$$\min f(x)$$

subject to the inequality constraints

$$g_i(x) \leq 0$$

for $i = 1, \ldots, p$ and the equality constraints

$$h_i(x) = 0$$

for $i = 1, \ldots, q$. We consider here methods that search for the solution using gradient information, i.e., we assume that the function f is differentiable.

EXAMPLE 4.1

Given a fixed area of cardboard A construct a box of maximum volume. The nonlinear program for this is

$$\min xyz$$

subject to

$$2xy + 2xz + 2yz = A$$

EXAMPLE 4.2

Consider the problem of determining locations for two new high schools in a set of P subdivisions N_j. Let w_{1j} be the number of students going to school A and w_{2j} be the number of students going to school B from subdivision N_j. Assume that the student capacity of school A is c_1 and the capacity of school B is c_2 and that the total number of students in each subdivision is r_j. We would like to minimize the total distance traveled by all the students given that they may attend either school A or B. It is possible to construct a nonlinear program to determine the locations (a, b) and (c, d) of high schools A and B, respectively assuming the location of each subdivision N_i is modeled as a single point denoted (x_i, y_i).

$$\min \sum_{j=1}^{P} w_{1j}\Big((a - x_j)^2 + (b - y_j)^2\Big)^{\frac{1}{2}} + w_{2j}\Big((c - x_j)^2 + (d - y_j)^2\Big)^{\frac{1}{2}}$$

subject to the constraints

$$\sum_j w_{ij} \le c_i$$

$$w_{1j} + w_{2j} = r_j$$

for $j = 1, \ldots, P$.

EXAMPLE 4.3

Neural networks have provided a new tool for approximating functions where the functional form is unknown. The approximation takes on the form

$$f(x) = \sum_j b_j \sigma(a_j x - \alpha_j) - \beta$$

and the corresponding sum of squares error term is

$$E(a_j, b_j, \alpha_j, \beta) = \sum_i \big(y_i - f(x_i)\big)^2$$

The problem of minimizing the error function is, in this instance, an unconstrained optimization problem. An efficient means for computing the gradient of E is known as the *backpropogation* algorithm.

4.1 UNCONSTRAINED OPTIMIZATION IN ONE DIMENSION

Here we begin by considering a significantly simplified (but nonetheless important) nonlinear programming problem, i.e., the domain and range of the function to be minimized are one-dimensional and there are no constraints. A necessary condition for a minimum of a function was developed in calculus and is simply

$$f'(x) = 0$$

Note that higher derivative tests can determine whether the function is a max or a min, or the value $f(x + \delta)$ may be compared to $f(x)$.

Note that if we let

$$g(x) = f'(x)$$

then we may convert the problem of finding a minimum or maximum of a function to the problem of finding a zero.

4.1.1 Bisection Algorithm

Let x^* be a root, or zero, of $g(x)$, i.e., $g(x^*) = 0$. If an initial bracket $[a, b]$ is known such that $x^* \in [a, b]$, then a simple and robust approach to determining the root is to bisect this interval into two intervals $[a, c]$ and $[c, b]$ where c is the midpoint, i.e.,

$$c = \frac{a + b}{2}$$

If

$$g(a)g(c) < 0$$

then we conclude

$$x^* \in [a, c]$$

while if

$$g(b)g(c) < 0$$

then we conclude

$$x^* \in [b, c]$$

This process may now be iterated such that the size of the bracket (as well as the actual error of the estimate) is being divided by 2 every iteration.

4.1.2 Newton's Method

Note that in the bisection method the actual value of the function $g(x)$ was only being used to determine the correct bracket for the root. Root finding is accelerated considerably by using this function information more effectively.

For example, imagine we were seeking the root of a function that was a straight line, i.e., $g(x) = ax + b$ and our initial guess for the root was x_0. If we extend this straight line from the point x_0 it is easy to determine where it crosses the axis, i.e.,

$$ax_1 + b = 0$$

so $x_1 = -b/a$. Of course, if the function were truly linear then no first guess would be required. So now consider the case that $g(x)$ is nonlinear but may be approximated locally about the point x_0 by a line. Then the point of intersection of this line with the x-axis is an estimate, or second guess, for the root x^*. The linear approximation comes from Taylor's theorem, i.e.,

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \frac{1}{2}g''(x_0)(x - x_0)^2 + \dots$$

So the linear approximation to $g(x)$ about the point x_0 can be written

$$l(x) = g(x_0) + g'(x_0)(x - x_0)$$

If we take x_1 to be the root of the linear approximation we have

$$l(x_1) = 0 = g(x_0) + g'(x_0)(x_1 - x_0)$$

Solving for x_1 gives

$$x_1 = x_0 - \frac{g(x_0)}{g'(x_0)}$$

or at the nth iteration

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}$$

The iteration above is for determining a zero of a function $g(x)$. To determine a maximum or minimum value of a function f we employ condition that $f'(x) = 0$. Now the iteration is modified as as

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$$

4.2 UNCONSTRAINED OPTIMIZATION IN HIGHER DIMENSIONS

Now we consider the problem of minimizing (or maximizing) a scalar function of many variables, i.e., defined on a vector field. We consider the extension of Newton's method presented in the previous section as well as a classical approach known as steepest descent.

4.2.1 Taylor Series in Higher Dimensions

Before we extend the search for extrema to higher dimensions we present Taylor series for functions of more than one domain variable. To begin, the Taylor series for a function of two variables is given by

$$\begin{aligned} g(x,y) =& g(x^{(0)},y^{(0)}) + \frac{\partial g}{\partial x}(x-x^{(0)}) + \frac{\partial g}{\partial y}(y-y^{(0)}) \\ &+ \frac{\partial^2 g}{\partial x^2}\frac{(x-x^{(0)})^2}{2} + \frac{\partial^2 g}{\partial y^2}\frac{(y-y^{(0)})^2}{2} + \frac{\partial^2 g}{\partial x \partial y}(x-x^{(0)})(y-y^{(0)}) \\ &+ \text{higher order terms} \end{aligned}$$

In n variables $x = (x_1, \ldots, x_n)^T$ the Taylor series expansion becomes

$$g(x) = g(x^{(0)}) + \nabla g(x^{(0)})(x - x^{(0)}) + \frac{1}{2}(x-x^{(0)})^T Hg(x^{(0)})(x-x^{(0)}) + \cdots$$

where the Hessian matrix is defined as

$$\big(Hg(x)\big)_{ij} = \frac{\partial^2 g(x)}{\partial x_i \partial x_j}$$

and the gradient is written as a row vector, i.e.,

$$\big(\nabla g(x)\big)_i = \frac{\partial g(x)}{\partial x_i}$$

4.2.2 Roots of a Nonlinear System

We saw that Newton's method could be used to develop an iteration for determining the zeros of a scalar function. We can extend those ideas for determining roots of the nonlinear system

$$\begin{aligned} g_1(x_1,\ldots,x_n) &= 0 \\ g_2(x_1,\ldots,x_n) &= 0 \\ &\vdots \\ g_n(x_1,\ldots,x_n) &= 0 \end{aligned}$$

or, more compactly,

$$g(x) = 0.$$

Now we apply Taylor's theorem to each component $g_i(x_1, \ldots, x_n)$ individually, i.e., retaining only the first order terms we have the linear approximation to g_i about the point $x^{(0)}$ as

$$l_i(x) = g_i(x^{(0)}) + \nabla g_i(x^{(0)})(x - x^{(0)})$$

for $i = 1, \ldots, n$. We can write these components together as a vector equation

$$l(x) = g(x^{(0)}) + Jg(x^{(0)})(x - x^{(0)})$$

where now

$$\big(Jg(x)\big)_{ij} = \frac{\partial g_i(x)}{\partial x_j}$$

is the $n \times n$–matrix whose rows are the gradients of the components g_i of g. This matrix is called the Jacobian of g.

As in the scalar case we base our iteration on the assumption that

$$l(x^{(k+1)}) = 0$$

Hence,

$$g(x^{(k)}) + Jg(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0$$

and given $x^{(k)}$ we may determine the next iterate $x^{(k+1)}$ by solving an $n \times n$ system of equations.

4.2.3 Newton's Method

In this chapter we are interested in minimizing functions of several variables. Analogously with the scalar variable case we may modify the above root finding method to determine maxima (or minima) of a function $f(x_1, \ldots, x_n)$. To compute an extreme point we require that $\nabla f = 0$, hence we set

$$g(x) = \big(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\big)^T.$$

Substituting

$$g_i(x) = \frac{\partial f(x)}{\partial x_i}$$

into

$$g(x^{(k)}) + Jg(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0$$

produces

$$\nabla f(x^{(k)}) + Hf(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0$$

where

$$\big(Hf(x)\big)_{ij} = \big(Jg(x)\big)_{ij} = \frac{\partial g_i(x)}{\partial x_j} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Again we have a linear system for $x^{(k+1)}$.

4.2.4 Steepest Descent

Another form for Taylor's formula in n-variables is given by

$$f(x + th) = f(x) + t\nabla f(x)h \ + \ \text{higher order terms}$$

where again $(\nabla f(x))_i = \partial f(x)/\partial x_i$. Now t is a scalar and $x+th$ is a ray emanating from the point x in the direction h. We can compute the derivative of the function $f(x+th)$ w.r.t. t as

$$\frac{df}{dt}(x+th) = \nabla f(x+th)h.$$

Evaluating the derivative at the point $t=0$ gives

$$\frac{df}{dt}(x+th)|_{t=0} = \nabla f(x)h$$

This quantity, known as the directional derivative of f, indicates how the function is changing at the point x in the direction h. Recall from calculus that the direction of maximum increase (decrease) of a function is in the direction of the gradient (negative gradient). This is readily seen from the formula for the directional derivative using the identity

$$\nabla f(x)h = \|\nabla f(x)\|\|h\|\cos(\theta)$$

where θ is the angle between the vectors $\nabla f(x)$ and h. Here $\|a\|$ denotes the Euclidean norm of a vector a. We can assume without loss of generality that h is of unit length, i.e., $\|h\| = 1$. So the quantity on the right is a maximum when the vectors h and $\nabla f(x)$ point in the same direction so $\theta = 0$.

This observation may be used to develop an algorithm for unconstrained function minimization. With an appropriate choice of the scalar step-size α, the iterations

$$x^{(k+1)} = x^{(k)} - \alpha \nabla f(x^{(k)}) \tag{4.1}$$

will converge (possibly slowly) to a minimum of the function $f(x)$.

4.3 CONSTRAINED OPTIMIZATION AND LAGRANGE MULTIPLIERS

Consider the constrained minimization problem

$$\min f(x)$$

subject to

$$c_i(x) = 0$$

$i = i, \ldots, p$. It can be shown that a necessary condition for a solution to this problem is provided by solving

$$\nabla f = \lambda_1 \nabla c_1 + \cdots + \lambda_p \nabla c_p$$

where the λ_i are referred to as Lagrange multipliers. Consider the case of f, c being functions of two variables and consider their level curves. In Section 4.4 we will demonstrate that an extreme value of f on a single constraint c is given when the gradients of f and c are parallel. The equation above generalizes this to several constraints c_i: an extreme value is given if the gradient of f is a linear combination of the gradients of the c_i.

We demonstrate a solution via this procedure by recalling our earlier example.

EXAMPLE 4.4

Given a fixed area of cardboard A construct a box of maximum volume. The nonlinear program for this is

$$\min xyz$$

subject to

$$2xy + 2xz + 2yz = A$$

Now $f(x, y, z) = xyz$ and $c(x, y, z) = 2xy + 2yz + 2xz - A$. Substituting these functions into our condition gives

$$\nabla f = \lambda \nabla c$$

which produces the system of equations

$$\begin{aligned} yz - \lambda(2y + 2z) &= 0 \\ xz - \lambda(2x + 2z) &= 0 \\ xy - \lambda(2y + 2x) &= 0 \end{aligned}$$

These equations together with the constraints provide four equations for (x, y, z, λ). If we divide the first equation by the second we find $x = y$. Similarly, if the second equation is divided by the third we obtain $y = z$. From the constraint it follows then that $6x^2 = A$, hence the solution is

$$x = y = z = \sqrt{\frac{A}{6}}.$$

In this special case the nonlinear system could be solved by hand. Typically this is not the case and one must resort to numerical techniques such as Newton's method to solve the resulting $(n + m) \times (n + m)$ system

$$g(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = 0.$$

4.4 GEOMETRY OF CONSTRAINED OPTIMIZATION

4.4.1 One Equality Constraint

Consider a two variable optimization problem

$$\min f(x, y)$$

subject to

$$c(x, y) = 0.$$

Geometrically the constraint $c = 0$ defines a curve C in the (x, y)–plane, and the function $f(x, y)$ is restricted to that curve. If we could solve the constraint equation for y as $y = h(x)$, the problem would reduce to an unconstrained, single variable optimization problem

$$\min f(x, h(x)).$$

From calculus we know that a necessary condition for a minimum is

$$\frac{d}{dx}f(x,h(x)) = \frac{\partial f}{\partial x}(x,h(x)) + \frac{\partial f}{\partial y}(x,h(x))h'(x) = 0. \tag{4.2}$$

Since $c(x,h(x)) = 0$, we also have

$$\frac{d}{dx}c(x,h(x)) = \frac{\partial f}{\partial x}(x,h(x)) + \frac{\partial c}{\partial y}(x,h(x))h'(x) = 0. \tag{4.3}$$

A necessary condition for equations (4.2) and (4.3) to hold simultaneously is

$$\frac{\partial f}{\partial x}\frac{\partial c}{\partial y} - \frac{\partial f}{\partial y}\frac{\partial c}{\partial x} = 0. \tag{4.4}$$

From elementary linear algebra we know that if an equation $ad - bc = 0$ holds then the vectors (a,b) and (c,d) are linearly dependent, i.e. collinear, and so one of them is a multiple of the other. Thus there exists a constant λ such that

$$\nabla f = \lambda \nabla c. \tag{4.5}$$

Now let's look more closely at the curve C. The tangent of the curve $y = h(x)$ at a point $(x_0, y_0) = (x_0, h(x_0))$ is given by

$$y = (x - x_0)h'(x_0) + y_0.$$

We set $x - x_0 = t$ and write this equation in vector form as

$$\left[\begin{array}{c} x \\ y \end{array}\right] = \left[\begin{array}{c} x_0 \\ y_0 \end{array}\right] + t\left[\begin{array}{c} 1 \\ h'(x_0) \end{array}\right].$$

The vector $T = [1, h'(x_0)]^T$ points into the direction of the tangent line and is called a tangent vector of C at (x_0, y_0). Equation (4.3) tells that T is orthogonal to $\nabla c(x_0, y_0)$. Thus at every point on C the gradient ∇c is orthogonal to the tangent of C.

For level contours $f(x,y) = f_0$ at level f_0 (an arbitrary constant) the situation is analogous, i.e., at each point on the contour the gradient ∇f is orthogonal to the tangent. Moreover, it is shown in multivariable calculus that ∇f points into the region in which f is increasing as illustrated in Figure 4.1. Note that the vector $(\partial f/\partial y, -\partial f/\partial x)$ is orthogonal to ∇f and so is a tangent vector.

At a point (x_0, y_0) on C for which (4.5) holds, the level contour of $f_0 = f(x_0, y_0)$ intersects the curve C. Since the gradients of f and c are collinear at this point, the tangents of the contour $f = f_0$ and the curve $c = 0$ coincide, hence the two curves meet tangentially. Thus the condition (4.5) means geometrically that we search for points at which a level contour and the constraint curve C have a tangential contact.

EXAMPLE 4.5

Consider the problem of finding all maxima and minima of

$$f(x,y) = x^2 - y^2$$

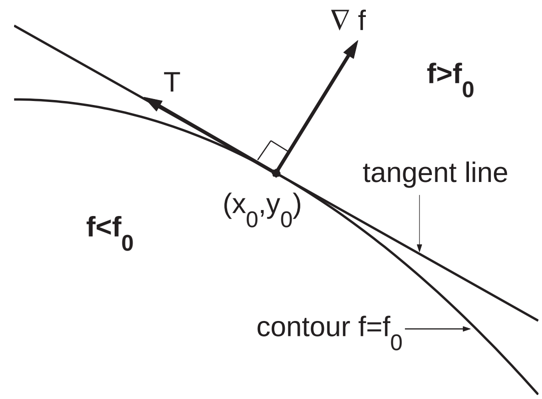

FIGURE 4.1: The gradient of f is orthogonal to the tangent of a level contour and points into the region of increasing f.

subject to

$$x^2 + y^2 = 1. \tag{4.6}$$

The equation (4.5) becomes

$$\begin{aligned} 2x &= 2\lambda x & (4.7) \\ 2y &= -2\lambda y, & (4.8) \end{aligned}$$

and (4.6)–(4.8) are three equations for (x, y, λ). Equation (4.7) has the solution $x = 0$ and the solution $\lambda = 1$ if $x \neq 0$. If $x = 0$, (4.6) leads to $y = \pm 1$ giving the solution points $(0, \pm 1)$ with values $f(0, \pm 1) = -1$. If $x \neq 0$ and $\lambda = 1$, (4.8) implies $y = 0$ and so $x = \pm 1$ from (4.6). This leads to the solution points $(\pm 1, 0)$ with values $f(\pm 1, 0) = 1$. Hence the points $(0, \pm 1)$ yield minima and $(\pm 1, 0)$ yield maxima.

In Figure 4.2 (a) some level contours of f and the constraint circle (4.6) are shown. The contours $f = 1$ and $f = -1$ are the only contours that meet this circle tangentially. The points of tangency are the maximum and minimum points of f restricted to the unit circle.

A slightly more complicated objective function is

$$f(x, y) = x^3 + y^2.$$

Again we seek all maxima and minima of f subject to the constraint (4.6). The equation (4.5) now results in

$$\begin{aligned} 3x^2 &= 2\lambda x & (4.9) \\ 2y &= 2\lambda y. & (4.10) \end{aligned}$$

Equation (4.9) has the solution $x = 0$ and $\lambda = 3x/2$ if $x \neq 0$. If $x = 0$ we find $y = \pm 1$ from (4.6) giving the solutions $(0, \pm 1)$ with values $f(0, \pm 1) = 1$. If $\lambda = 3x/2 \neq 0$, equation (4.10) has the solutions $y = 0$ and $\lambda = 1$ if $y \neq 0$. Now if $y = 0$ we find

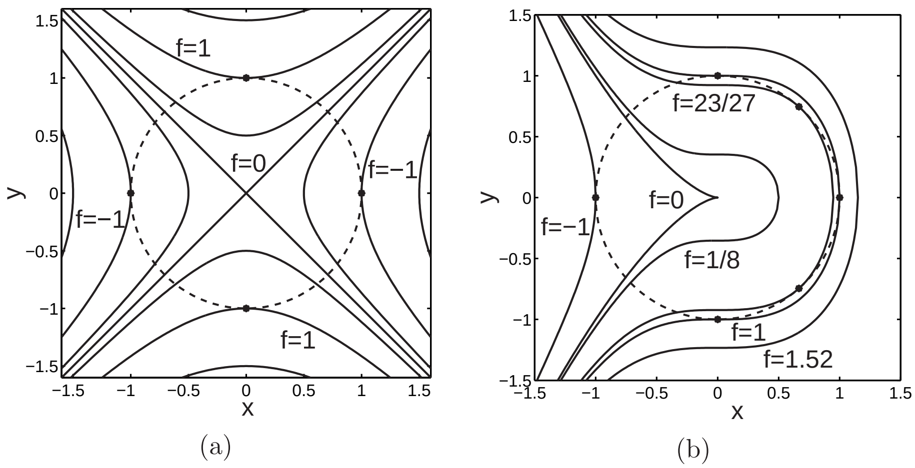

FIGURE 4.2: Unit circle $x^2 + y^2 = 1$ (dashed) and level contours of (a): $f(x, y) = x^2 - y^2$, (b): $f(x, y) = x^3 + y^2$. The points of tangency are the extreme points of $f(x, y)$ restricted to the unit circle.

$x = \pm 1$ from (4.6) giving the solutions $(\pm 1, 0)$ with values $f(\pm 1, 0) = \pm 1$. If $y \neq 0$ it follows that $\lambda = 1$, hence $x = 2/3$, and so $y = \pm\sqrt{5}/3$ from (4.6). The f-values of the solution points $(2/3, \pm\sqrt{5}/3)$ are both $23/27 < 1$. Thus there is a single global minimum $f = -1$ at $(-1, 0)$, and three global maxima $f = 1$ at $(0, \pm 1)$ and $(1, 0)$.

Some level contours of f and the constraint curve (4.6) are shown in Figure 4.2 (b). Note that the zero contour forms a cusp, $y = \pm(-x)^{3/2}$, $x \leq 0$. The points of tangency of a level contour and the constraint curve are again identified with extreme points. Since the points $(2/3, \pm\sqrt{5}/3)$ are located between the global maximum points they must correspond to local minima.

In three dimensions the equation $\nabla f = \lambda \nabla c$, resulting from an optimization problem with a single constraint, implies that at a solution point a level surface $f(x, y, z) = f_0$ is tangent to the constraint surface $c(x, y, z) = 0$.

EXAMPLE 4.6

Find the maxima and minima of

$$f(x, y, z) = 5x + y^2 + z$$

subject to

$$x^2 + y^2 + z^2 = 1. \tag{4.11}$$

The equation $\nabla f = \lambda \nabla c$ now leads to

$$5 = 2\lambda x \tag{4.12}$$
$$2y = 2\lambda y \tag{4.13}$$
$$1 = 2\lambda z. \tag{4.14}$$

From (4.12) and (4.14) we infer that $x = 5z$, and (4.13) has the solutions $y = 0$ and $\lambda = 1$ if $y \neq 0$. Assume first $y = 0$. The constraint (4.11) implies $x^2+z^2 = 26z^2 = 1$, hence $z = \pm 1/\sqrt{26}$, $x = \pm 5/\sqrt{26}$, and $f(\pm 5/\sqrt{26}, 0, \pm 1/\sqrt{26}) = \pm\sqrt{26}$.

Now assume $y \neq 0$, hence $\lambda = 1$, and so $x = 5/2$, $z = 1/2$. The constraint (4.11) then yields $26/4 + y^2 = 1$ which has no solution. Thus there is a unique maximum at $(5/\sqrt{26}, 0, 1/\sqrt{26})$ and a unique minimum at $(-5/\sqrt{26}, 0, -1/\sqrt{26})$.

EXAMPLE 4.7

Find the maxima and minima of

$$f(x, y, z) = 8x^2 + 4yz - 16z \tag{4.15}$$

subject to the constraint

$$4x^2 + y^2 + 4z^2 = 16. \tag{4.16}$$

Note that (4.16) defines an ellipsoid of revolution. The equation $\nabla f = \lambda \nabla c$ yields

$$\begin{aligned} 16x &= 8\lambda x && (4.17)\\ 4z &= 2\lambda y && (4.18)\\ 4y - 16 &= 8\lambda z. && (4.19) \end{aligned}$$

From (4.18) we find $z = \lambda y/2$ and then from (4.19) $4y - 16 = 4\lambda^2 y$, i.e.

$$y = \frac{4}{1-\lambda^2}, \quad z = \frac{2\lambda}{1-\lambda^2}.$$

Equation (4.17) has the solutions $x = 0$ and $\lambda = 2$ if $x \neq 0$. Assume first $x = 0$. Substituting y, z and $x = 0$ into (4.16) yields a single equation for λ which can be manipulated to $\lambda^2(3 - \lambda^2) = 0$, i.e. $\lambda = 0$ or $\lambda^2 = 3$. Setting $\lambda = 0$ leads to $y = 4$, $z = 0$, and $f(0, 4, 0) = 0$. For $\lambda = \mp\sqrt{3}$ we find $y = 2$ and $z = \pm\sqrt{3}$, with values $f(0, -2, \pm\sqrt{3}) = \mp 24\sqrt{3}$.

If $x \neq 0$ we have $\lambda = 2$ and so $y = z = -4/3$. The missing value of x is again found from (4.16) as $x = \pm 4/3$. The values of f at these points are both $128/3$. Thus the maxima and minima of f are

$$f_{max} = f(\pm 4/3, -4/3, -4/3) = 128/3, \quad f_{min} = f(0, -2, \sqrt{3}) = -24\sqrt{3}.$$

The level surfaces for the minimum and maximum values of f and the constraint ellipsoid are shown in Figure 4.3. We see in this figure that the solution points are points of tangency of a level surface and the constraint surface.

4.4.2 Several Equality Constraints

If several constraints are present, the situation is trivial when the number of (independent) constraints equals the number of variables. In this case all constraints typically are satisfied only by a finite number of points, if any, and one merely

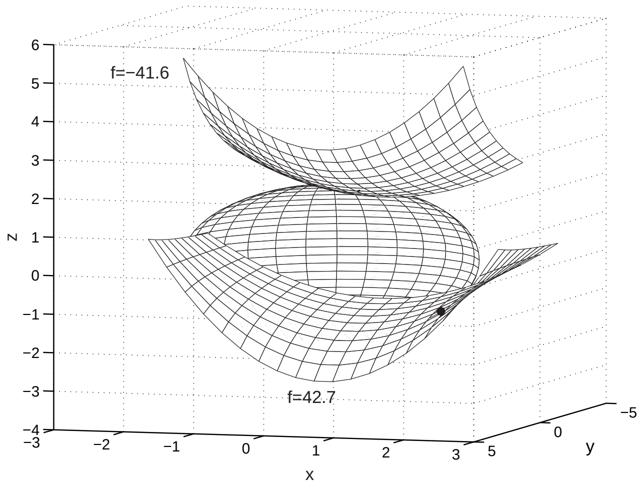

FIGURE 4.3: Level surfaces $f = f_{min} \approx -41.6$ and $f = f_{max} \approx 42.7$ for $f(x, y, z)$ defined by equation (4.15). Both level surfaces have a tangential contact with the constraint ellipsoid (4.16).

has to evaluate the objective function at these points to find the global maxima or minima. Lagrange multiplies are needed if the number of constraints is smaller than the number of variables.

Consider for simplicity the case of three variables (x, y, z) and two constraints $c_1(x, y, z) = 0$, $c_2(x, y, z) = 0$. Each of the two constraints defines a surface in three dimensional (x, y, z)–space, and both constraints together define a curve C, the intersection of the two constraint surfaces. (Two non–parallel planes in three dimensional space intersect in a straight line. Likewise, two curved surfaces typically intersect in a curve.) Now a level set $f(x, y, z) = f_0$ also defines a surface, and the condition for f to have an extreme point when restricted to C is again that a level surface and C meet tangentially at some point (x_0, y_0, z_0). This condition means that the tangent line of C at the point of contact is entirely in the tangent plane of the level surface. Since the tangent line of C is the intersection of the tangent planes of the two constraint surfaces, the tangency condition means that all three tangent planes intersect in a line. This is a special condition because in general three planes in three dimensional space have only a single point in common.

As in two dimensions, the gradient $\nabla f(x_0, y_0, z_0)$ is orthogonal to the tangent plane of the level surface $f(x, y, z) = f(x_0, y_0, z_0)$ at (x_0, y_0, z_0). The same holds for the tangent planes of the constraint surfaces $c_1 = 0$ and $c_2 = 0$. The condition that these planes intersect in a line implies that the three gradient vectors to which they are orthogonal are all located in the normal plane of that line and hence are linearly dependent as illustrated in Figure 4.4. Thus one of these gradient vectors is a linear combination of the other two, which we write as $\nabla f = \lambda_1 \nabla c_1 + \lambda_2 \nabla c_2$. For more variables and constraints the situation is similar.

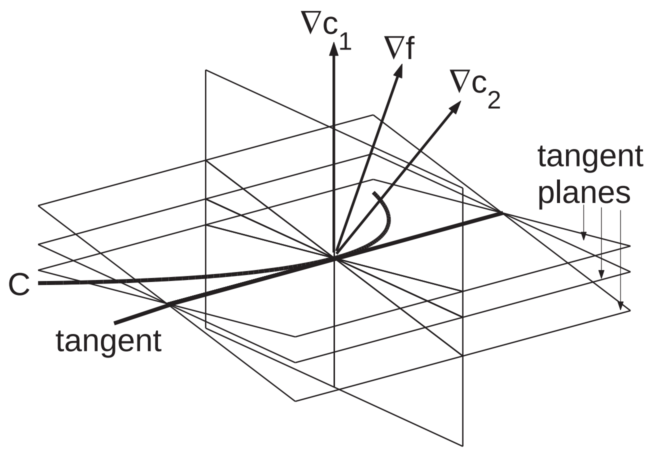

FIGURE 4.4: At a solution point of a three–variable optimization problem with two constraints the tangent plane of the level surface of f and the tangent planes of the two intersecting constraint surfaces $c_1 = 0$ and $c_2 = 0$ intersect in the tangent of the constraint curve C. As a consequence all three gradients are in the normal plane of C and so are linearly dependent.

EXAMPLE 4.8

Find the maxima and minima of

$$f(x, y, z) = x^2 + y^2 - z$$

subject to

$$\begin{aligned} x^2 + y^2 &= 1 \\ x^2 + z^2 &= 1. \end{aligned}$$

Here we can find a parametric representation of the constraint curve C. Substituting $x^2 = 1 - z^2$ from the second constraint equation into the first constraint equation yields $y^2 = z^2$, i.e. $z = \pm y$. The first constraint defines a circle which we parametrize as $x = \cos\varphi$, $y = \sin\varphi$, where $-\pi \le \varphi \le \pi$. Thus the constraints define two curves

$$C_\pm : \quad (x, y, z) = (\cos\varphi, \sin\varphi, \pm\sin\varphi).$$

Note that the two curves intersect if $z = 0$, i.e., at $\varphi = 0$ and $\varphi = \pi$.

To solve the constrained optimization problem we substitute the parametric representation of $C_\pm$ into f and set

$$f_\pm(\varphi) = 1 \mp \sin\varphi.$$

The extreme points are determined by $df_\pm/d\varphi = \mp\cos\varphi = 0$, hence $\varphi = \pm\pi/2$, with values $f_\pm(\mp\pi/2) = 2$ and $f_\pm(\pm\pi/2) = 0$. Thus there are two maxima at $(0, \pm1, -1)$ and two minima at $(0, \pm1, 1)$ with values 2 and 0, respectively. The intersecting

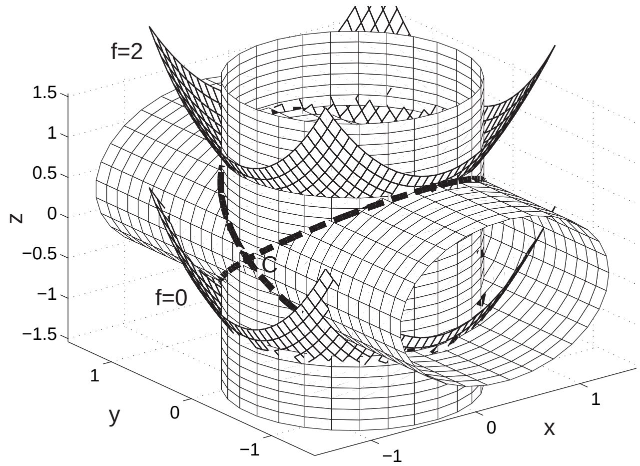

FIGURE 4.5: Intersecting constraint cylinders and level surfaces for Example 4.8.

constraint cylinders and the level surfaces for the maximum and minimum values are shown in Figure 4.5. It can be easily verified a posteriori that the gradient of f and the gradients of the two constraint functions are linearly dependent at the four extreme points.

4.4.3 Inequality Constraints

Finally consider the case of inequality constraints for a problem with n variables. Inequality constraints define a feasible region S in n–dimensional space, and the objective function is restricted to S. Extreme points can be located in the interior of S as well as on the boundary. If there are no solutions to $\nabla f = 0$ in the interior, all extreme points are on the boundary. Assume that $c(x) \geq 0$ is one of the inequality constraints. The boundary of this constraint is the hypersurface defined by $c(x) = 0$. Finding an extreme point on this boundary amounts to solving an optimization problem with a single equality constraint (and possibly an additional set of inequality constraints). If two inequality constraints $c_1 \geq 0$, $c_2 \geq 0$ are present, the optimal solution may also be located on the intersection of the two boundary hypersurfaces $c_1 = c_2 = 0$ which leads to a problem with two equality constraints etc. The situation is naturally much more complicated than in linear programming. Linear programming problems do not have solutions in the interior of the feasible region.

EXAMPLE 4.9

Consider the problem of minimizing the objective function

$$f(x, y) = \frac{x^4}{4} - \frac{x^2}{2} + \frac{y^2}{2}.$$

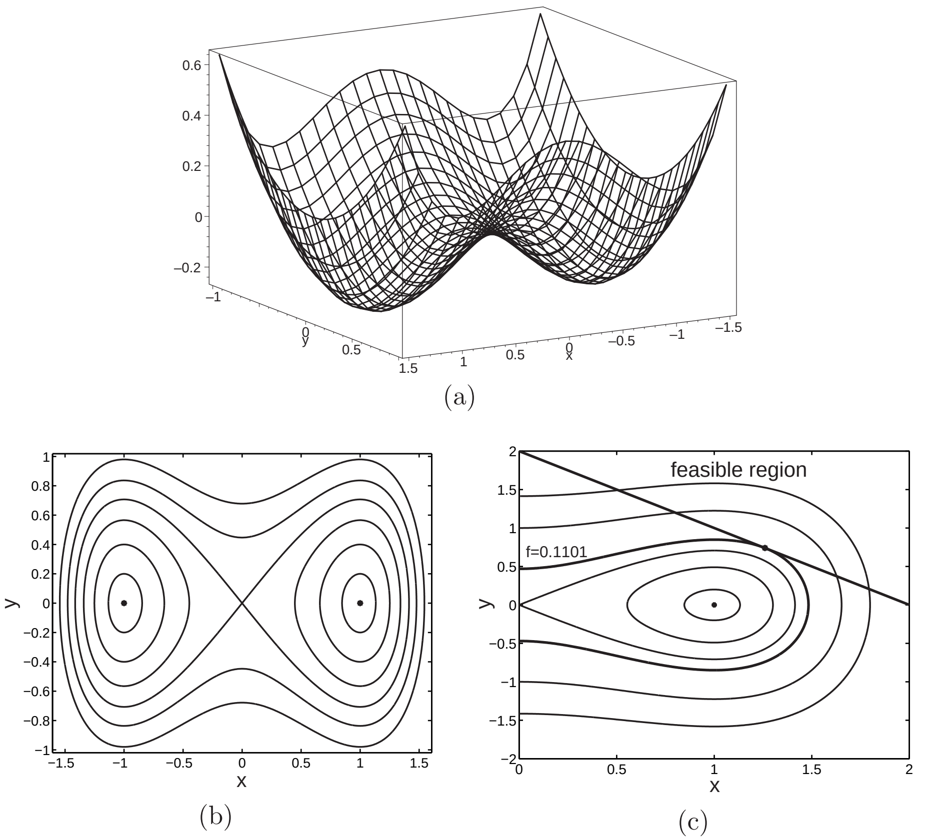

FIGURE 4.6: (a): Three dimensional plot of $f(x,y) = x^4/4 - x^2/2 + y^2/2$. (b): Level contours of f. (c): Contours of f in the right half plane and the constraint boundary $x + y = 2$.

Unconstrained optimization leads to the equations

$$\begin{aligned} \frac{\partial f}{\partial x} &= x^3 - x = 0 \quad \Rightarrow \quad x = 0 \text{ or } x = \pm -1 \\ \frac{\partial f}{\partial y} &= y = 0. \end{aligned}$$

To check the types of the extreme points $(0,0)$ and $(\pm 1, 0)$ we compute the Hessean matrices,

$$Hf(0,0) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Hf(\pm 1, 0) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

From the form of these matrices it follows that $(\pm 1, 0)$ are minimum points ($f = -1/4$), and $(0,0)$ is a saddle point ($f = 0$). A three–dimensional surface plot of f is shown in Figure 4.6 (a), and some level contours are displayed in Figure 4.6 (b).

Now consider the problem of minimizing $f(x,y)$ subject to the inequality constraint

$$c(x,y) = x + y \geq 2.$$

Since $c(\pm 1, 0) < 2$, the global minima of f are not in the feasible region, hence the optimal solution must be on the boundary. We are then led to the problem of minimizing f subject to the equality constraint

$$x + y = 2.$$

The equation (4.5) leads to

$$x^3 - x = \lambda, \quad y = \lambda \quad \Rightarrow \quad x^3 - x - y = 0.$$

Substituting $y = 2 - x$ from the constraint equation into this equation gives $x^3 - 2 = 0$, with the solution $x = 2^{1/3} = 1.2600$, and hence $y = 2 - 2^{1/3} = 0.7401$. The numerical value of f at this point is 0.11012. Note that the equation for x also follows directly from the unconstrained, single variable optimization problem associated with $f(x, 2 - x)$.

In Figure 4.6 (c) the constraint line and some level contours are shown. The solution point is again revealed as point of tangency.

4.5 MODELING EXAMPLES

EXAMPLE 4.10

A manufacturerer of colored TV's is planning the introduction of two new products: a 19–inch stereo color set with a manufacturerer's suggested retail price of \$339 per year, and a 21–inch stereo color set with a suggested retail price of \$339 per year. The cost of the company is \$195 per 19–inch set and \$225 per 21–inch set, plus additional fixed costs of \$400, 000 per year. In the competitive market the number of sales will affect the sales price. It is estimated that for each type of set, the sales price drops by one cent for each additional unit sold. Furthermore, sales of the 19–set will affect sales of the 21–inch set and vice versa. It is estimated that the price for the 19–inch set will be reduced by an additional 0.3 cents for each 21–inch sold, and the price for 21–inch sets will decrease for by 0.4 cents for each 19–inch set sold. The company believes that when the number of units of each type produced is consistent with these assumptions all units will be sold. How many units of each type of set should be manufactured such the profit of the company is maximized?

The relevant variables of this problem are:

s_1: number of units of the 19–inch set produced per year,
s_2: number of units of the 21–inch set produced per year,
p_1: sales price per unit of the 19–inch set (\$),
p_2: sales price per unit of the 21–inch set (\$),
C: manufacturing costs (\$ per year),
R: revenue from sales (\$ per year),
P: profit from sales (\$ per year).

The market estimates result in the following model equations,

$$
\begin{aligned}
p_1 &= 339 - 0.01s_1 - 0.003s_2 \\
p_2 &= 399 - 0.04s_1 - 0.01s_2 \\
R &= s_1p_1 + s_2p_2 \\
C &= 400,000 + 195s_1 + 225s_2 \\
P &= R - C.
\end{aligned}
$$

The profit then becomes a nonlinear function of (s_1, s_2),

$$P(s_1, s_2) = -400,000 + 144s_1 + 174s_2 - 0.01s_1^2 - 0.007s_1s_2 - 0.01s_2^2. \qquad (4.20)$$

If the company has unlimited resources, the only constraints are $s_1, s_2 \geq 0$.

Unconstrained Optimization. We first solve the unconstrained optimization problem. If P has a maximum in the first quadrant this yields the optimal solution. The condition for an extreme point of P leads to a linear system of equations for (s_1, s_2),

$$
\begin{aligned}
\frac{\partial P}{\partial s_1} &= 144 - 0.02s_1 - 0.007s_2 = 0 \\
\frac{\partial P}{\partial s_2} &= 174 - 0.007s_1 - 0.02s_2 = 0.
\end{aligned}
$$

The solution of these equations is $s_1^* = 4735$, $s_2^* = 7043$ with profit value $P^* = P(s_1^*, s_2^*) = 553,641$. Since s_1^*, s_2^* are positive, the inequality constraints are satisfied. To determine the type of the extreme point we inspect the Hessean matrix,

$$HP(s_1^*, s_2^*) = \begin{bmatrix} -0.02 & -0.007 \\ -0.007 & -0.02 \end{bmatrix}.$$

A sufficient condition for a maximum is that $(HP)_{11} < 0$ and $\det(HP) > 0$. Both of these conditions are satisfied and so our solution point is indeed a maximum, in fact a global maximum. In Figure 4.7 (a) a three–dimensional plot of $P(s_1, s_2)$ is shown. Some level contours are displayed in Figure 4.7 (b). The level contours play here the role of isoprofit lines. Because P is a nonlinear function, the isoprofit lines form closed curves that surround the maximum at (s_1^*, s_2^*).

Constrained Optimization. Now assume the company has limited resources which restrict the number of units of each type produced per year to

$$s_1 \leq 5,000, \quad s_2 \leq 8,000, \quad s_1 + s_1 \leq 10,000.$$

The first two constraints are satisfied by (s_1^*, s_2^*), however $s_1^* + s_2^* = 11,778$. The global maximum point of P is now no longer in the feasible region, thus the optimal solution must be on the boundary. We therefore solve the constrained optimization problem

$$\max P$$

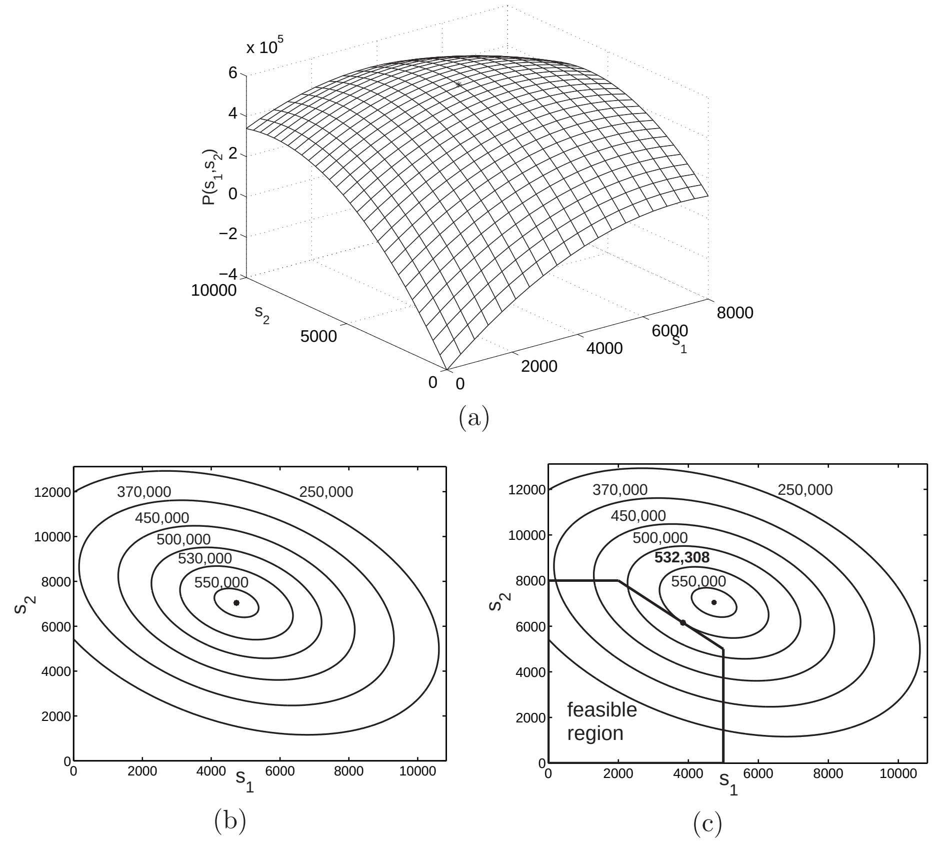

FIGURE 4.7: (a): Three dimensional plot of $P(s_1, s_2)$, equation (4.20). (b): Level contours of P. (c): Level contours of P and feasible region for the constrained optimization problem.

subject to

$$c(s_1, s_2) = s_1 + s_2 - 10,000 = 0.$$

We can either substitute s_2 or s_1 from the constraint equation into P and solve an unconstrained one–variable optimization problem, or use Lagrange multipliers. Choosing the second approach, the equation $\nabla P = \lambda \nabla c$ becomes

$$\begin{aligned} 144 - 0.02s_1 - 0.007s_2 &= \lambda \\ 174 - 0.007s_1 - 0.02s_2 &= \lambda, \end{aligned}$$

which reduces to a single equation for s_1, s_2. Together with the constraint equation we then have again a system of two linear equations,

$$\begin{aligned} -0.013s_1 + 0.013s_2 &= 30 \\ s_1 + s_2 &= 10,000. \end{aligned}$$

The solution is $s_1^* = 3846$, $s_2^* = 6154$, with profit value $P^* = 532,308$. In Figure 4.7 (c) the feasible region and some contour levels are shown. The optimal solution

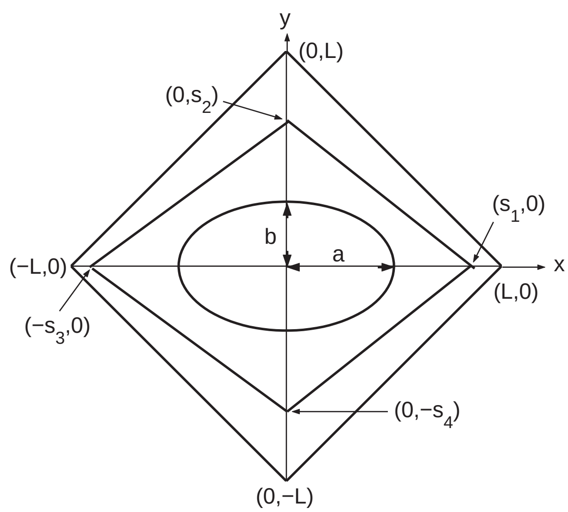

FIGURE 4.8: Geometry of the problem of Example 4.11.

is revealed as point of tangency of the isoprofit line $P = P^*$ and the constraint line. It is also clear from the figure that the solution point cannot be located on one of the two other boundary lines $s_1 = 5,000$ or $s_2 = 8,000$.

EXAMPLE 4.11

A fish farm has a fish lake on a square area. The length of the diagonal of the square is $2L$. The fish lake has the shape of an ellipse with semi–axes a and b. The center of the lake is at the center of the square and the semi–axes are on the diagonals. The owner of the fish farm has fencing material of length l where $l < 4\sqrt{2}L$. She wants to surround the lake by a fence in the form of a quadrilateral whose corner points are on the diagonals of the square. In order that the owner has enough space to work at the lake, the distance between fence and lake must not be smaller than a given distance d_m. What is the position of the corner points of the fence such that the enclosed area is maximal?

To formulate this problem as a nonlinear program, we introduce a (x, y)–coordinate whose origin is at the center of the square. The corner points of the square are $(\pm L, 0)$ and $(0, \pm L)$. The equation of the fish lake's boundary is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

The corner points of the fence's quadrilateral have coordinates $(s_1, 0)$, $(0, s_2)$, $(-s_3, 0)$, and $(0, -s_4)$ $(0 \le s_j \le L)$ with (s_1, s_2, s_3, s_4) to be determined, see Figure 4.8.

To invoke the distance restriction, we have to compute the minimal distance between the ellipse and the four edges of the quadrilateral. Consider the edge in the first quadrant. The equation of this edge is $y = (s_2/s_1)(s_1 - x)$. Some thought reveals that the minimal distance between this straight line and the ellipse is given by

$$(s_1 s_2 - d(s_1, s_2))/\sqrt{s_1^2 + s_2^2}, \tag{4.21}$$

where

$$d(s_1, s_2) = a^2 s_2^2 + b^2 s_1^2,$$

provided $s_1 s_2 \geq d(s_1, s_2)$. Thus the minimum distance condition for this edge can be formulated as

$$s_1 s_2 - d(s_1, s_2) \geq d_m \sqrt{s_1^2 + s_2^2}.$$

The minimum distance conditions for the other three edges are obtained by replacing (s_1, s_2) in this inequality by (s_3, s_2), (s_3, s_4), and (s_1, s_4), respectively.

The area enclosed by the fence is

$$A(s_1, s_2, s_3, s_4) = \frac{1}{2}(s_1 s_2 + s_2 s_3 + s_3 s_4 + s_4 s_1).$$

Now the optimization problem can be formulated as

$$\max A(s_1, s_2, s_3, s_4)$$

subject to the inequality constraints

$$\begin{aligned}
s_1 s_2 - d(s_1, s_2) &\geq d_m \sqrt{s_1^2 + s_2^2} \\
s_3 s_2 - d(s_3, s_2) &\geq d_m \sqrt{s_3^2 + s_2^2} \\
s_3 s_4 - d(s_3, s_4) &\geq d_m \sqrt{s_3^2 + s_4^2} \\
s_1 s_4 - d(s_1, s_4) &\geq d_m \sqrt{s_1^2 + s_4^2} \\
s_j &\leq L \ \ (1 \leq j \leq 4),
\end{aligned}$$

and the equality constraint

$$\sqrt{s_1^2 + s_2^2} + \sqrt{s_2^2 + s_3^2} + \sqrt{s_3^2 + s_4^2} + \sqrt{s_4^2 + s_1^2} = l.$$

Note that we don't need to impose the constraints $s_1 \geq a$ and $s_2 \geq b$. The minimum distance requirement for (s_1, s_2) implies $s_1 s_s \geq d(s_1, s_2)$ and this can be only satisfied if $s_1 \geq a$ and $s_2 \geq b$.

PROBLEMS

4.1. Extend Example 4.2 for a collection of S schools.

4.2. Show how Newton's method for root finding can be used to calculate $\sqrt{3}$. Compute numerically an iterated sequence that converges to this value. Stop the iteration if $|x_{n+1} - x_n| \leq 10^{-5}$. What is the effect of changing the initial condition?

4.3. Use Newton's method to find the positive root of

$$g(x) = x - \tanh(2x) = 0$$

up to five decimal places.

4.4. Plot $f(x) = x\sin(x)$ in $0 \leq x \leq 15$ and convince yourself that $f(x)$ has three local maxima in that range. Compute these maxima up to five decimal places using Newton's method.

4.5. Let

$$f(x, y) = x^4 + y^3 + xy^2 + x^2 - y + 1.$$

Find the quadratic approximation of $f(x, y)$ at the points
(a) $x_0 = y_0 = 0$,
(b) $x_0 = 1,\ y_0 = 0$,
(c) $x_0 = y_0 = 2$.

4.6. Compute the Jacobian of

$$g(x) = \begin{bmatrix} x_1x_2 - x_1 - 1 \\ x_1x_2x_3 - 2x_2 \\ e^{-x_1^2} - 3x_3 - 1 \end{bmatrix}$$

at $x_0 = [0, 0, 0,]^T$ and $x_0 = [1, 1, 1]^T$.

4.7. Minimize the objective function

$$f(x_1, x_2) = 7x_1^2 + 2x_1x_2 + x_2^2 + x_1^4 + x_2^4$$

using 50 iterations of
(a) Newton's method
(b) Steepest Descent
with starting value $x_0 = (3, 3)^T$. Plot the values of the iterates for each method on the same graph. You may experiment with the value of α in Equation (4.1). *Hint:* start small.

4.8. Assume a farmer has L feet of fencing for a rectangular area with lengths x and y. Determine these lengths such that the enclosed area is a maximum.

4.9. Consider an ellipse with semi-axes $a \geq b$. The area enclosed by the ellipse is $A = \pi ab$ and the circumference is $L = 4aE(a)$, where $e = \sqrt{1 - b^2/a^2}$ is the eccentricity and $E(e)$ is the complete elliptic integral of the second kind – a given function of e. Show that the constrained optimization problem

$$\max(\pi ab)$$

subject to

$$4aE(e) = L$$

leads to the following equation for e,

$$\frac{e}{1 - e^2} = -\frac{2E'(e)}{E(e)},$$

where $E'(e) = dE(e)/de$. *Note:* It turns out that the only solution of this equation is $e = 0$, i.e. $a = b$. Thus the area of an ellipse with prescribed circumference is a maximum if the ellipse degenerates to a circle.

4.10. Find all extreme points (local maxima and minima) of

$$f(x, y) = x^3 + y^2$$

subject to

$$y^2 - x^2 = 1.$$

Make a sketch showing the constraint curve, some level curves of f, and the extreme points as points of tangencies.

4.11. Find the minimum distance of the surface

$$2x^2 + y^2 - z^2 = 1$$

to the origin.

4.12. Find the points on the unit sphere

$$x^2 + y^2 + z^2 = 1,$$

for which the function

$$f(x, y, z) = 2x^2 + y^2 - z^2 - x$$

has a global maximum and a global minimum, respectively.

4.13. A manufacturer of personal computers currently sells $10,000$ units per month of a basic model. The manufacture cost per unit is \$700 and the current sales price is \$950. During the last quarter the manufacturer lowered the price by \$100 in a few test markets, and the result was a 50% increase in orders. The company has been advertising its product nationwide at a cost of \$$50,0000$ per month. The advertising agency claims that increasing the advertising budget by \$$10,000$ per month would result in a sales increase of 200 units per month. Management has agreed to consider an increase in the advertising budget to no more than \$$100,000$ per month.

Determine the price and the advertising budget that will maximize the profit. Make a table comparing the maximal profit and the corresponding values of the price, the advertising budget, and the number of sales to their current values, and to the optimal values that would result without advertisement.

Hint: Let N be the number of sales per month. Write $N = N_0 + \Delta N_p + \Delta N_a$, where N_0 is the current value of N, ΔN_p is the increase of N due to price reduction, and ΔN_a is the increase of N due to increasing the advertising budget. Note: If you don't find a solution in the interior of the feasible region, the optimal solution is on a boundary.

4.14. A local newspaper currently sells for \$1.50 per week and has a circulation of $80,000$ subscribers. Advertising sells for 250/page, and the paper currently sells 350 pages per week (50 pages/day). The management is looking for ways to increase profit. It is estimated that an increase of 10 cents/week in the subscription price will cause a drop of $5,000$ subscribers. Increasing the price of advertising by \$100/page will cause the paper to lose approximately 50 pages of advertising in a week. The loss of advertising will also affect circulations, since one of the reasons people by the newspaper is the advertisement. It is estimated that a loss of 50 pages of advertisement per week will reduce circulation by $1,000$ subscribers.

(a) Find the weekly subscription price and advertisement price that will maximize the profit.

(b) Same as (a), but now with the constraint that the advertising price cannot be increased beyond \$400.

Hint: Let M be the number of advertising pages per week. Write $M = M_0 + \Delta M_a$, where M_0 is the current value of M, and ΔM_a is the change caused by increasing the advertising price. Proceed similarly for N, the number of subscribers. Here you have to consider two causes of change.

4.15. Verify the expression (4.21) in Example 4.11.

In Exercises 4.16–4.19 use an optimization software such as the *fmincon* function of Matlab to find the optimal solution.

4.16. Redo the problem of Example 4.10, but now choose as objective function the marginal profit, i.e., the ratio $(R-C)/C$ of the profit and the total manufacturing costs.

4.17. Maximize the volume xyz of a cardboard subject to the equality constraint $xy + xz + yz = 4$ and the inequality constraints

$$\begin{aligned} 0 \le x &\le 0.5 \\ 2 \le y &\le 3 \\ z &\ge 1. \end{aligned}$$

4.18. Solve the fencing problem of Example 4.11 for $L = 4$, $a = 1.5$, $b = 2.5$, and

(a) $l = 20$, $d_m = 0.3$,

(b) $l = 20$, $d_m = 0.4$,

(c) $l = 17$, $d_m = 0.1$.

Hint: A good starting value for s_1 is $(a+L)/2$.

4.19. Solve the school problem of Example 4.2 for five districts with coordinates

x_j	0	0	0	-100	100
y_j	0	100	-100	0	0

and

(a) $r_1 = 200$, $r_2 = 300$, $r_3 = 200$, $r_4 = 500$, $r_5 = 300$, $c_1 = 1500$, $c_2 = 1500$,

(b) $r_1 = 200$, $r_2 = 400$, $r_3 = 200$, $r_4 = 500$, $r_5 = 300$, $c_1 = 700$, $c_2 = 2000$.

Hint: A reasonable starting value for w_{ij} is $r_j/2$. For the coordinates (a, b, c, d) you may try $(0, 0, 0, 0)$, $(100, 0, -100, 0)$, or $(50, 50, -50, -50)$.