problem in **NP** can be polynomially transformed to QPL. In this section we have argued that there is a polynomial-length certificate for **yes**-instances of QPL, showing that the problem is in **NP**. ∎

Notice that as a byproduct of this proof we have shown the following theorem. There seems to be no way to prove this theorem simpler than the machinery of this section.

THEOREM 4.2. *Consider the problem of minimizing a quadratic objective function* $f(\mathbf{x})$ *subject to* $A\mathbf{x} \geq \mathbf{b}$. *Then the possible outcomes are: (a) The constraints are infeasible, (b) A global minimum exists, or (c) The objective function is unbounded from below in the feasible region. In particular, the following case is not possible: The objective function is bounded below but does not achieve its minimum.*

PROOF. The preceding argument shows that for a feasible QP that does not attain its minimum, there must exist a ray such that every point on the ray is feasible, and the objective function is a decreasing quadratic function on the ray. Such a function must tend to $-\infty$. ∎

4.2. Special cases of nonconvex QP

In this section we will list some special cases of nonconvex QP that are known to be **NP**-hard. Since the general case of QP was proved to lie in **NP** when expressed as a decision problem, all of the examples of this section are also **NP**-complete when posed as decision problems.

QP with simple bounds

This is the problem of minimizing $\frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{c}^T\mathbf{x}$ subject to constraints $l_i \leq x_i \leq u_i$ for $i = 1, \ldots, n$. These constraints are sometimes called "box constraints." This problem is **NP**-hard by polynomial time transformation from SAT similar to the transformation proposed in the proof of Theorem 2.5. In that transformation we required simple bounds on the variables, namely, $0 \leq y_i \leq 1$. We also included constraints of the form $\mathbf{a}_i^T\mathbf{y} \geq b_i$ to capture the ith clause. For a setting of the variables corresponding to an assignment that satisfied the clause, we note from the reduction that $\mathbf{a}_i^T\mathbf{y} = b_i + v$ where v is a nonnegative integer between 0 and 2. We replace this constraint with an additional term in the objective function. The additional term takes the form $(\mathbf{a}_i^T\mathbf{y} - b_i - v_i)^2$ where v_i is a new variable bound to lie between 0 and 2.

This new form of QP has only simple-bound constraints on the variables. It still has the property that the objective function is nonnegative on the feasible region. One can prove that if the original boolean formula is satisfiable, then there is a setting of the variables in the quadratic program

in which the value of the objective function is zero. One can also prove conversely that setting of the variables to make the objective function zero must be an assignment to the boolean variables that satisfies all clauses.

Quadratic knapsack problem

The quadratic knapsack problem (QKP) was introduced in Section 3.1. This problem is **NP**-hard if the matrix D is not positive semidefinite, i.e., if it has negative entries on its diagonal.

We transform the subset-sum problem to QKP. The subset sum problem was described in Chapter 2 and is **NP**-complete (this was Theorem 2.4). In this decision problem, the input is a list of n nonnegative integers $a_1, \ldots, a_n$ and an integer γ. The question is whether there is a subset $J \subset \{1, \ldots, n\}$ such that

$$\sum_{j \in J} a_j = \gamma.$$

This is expressed as the following quadratic knapsack problem:

$$\begin{aligned} \text{minimize} \quad & x_1(1 - x_1) + \cdots + x_n(1 - x_n) \\ \text{subject to} \quad & a_1x_1 + \cdots + a_nx_n = \gamma, \\ & 0 \leq x_i \leq 1. \end{aligned} \tag{4.1}$$

Note that the objective function is nonnegative on the feasible region. Note also that the global minimum of the objective function is zero if and only if each x_i is set to 0 or 1, i.e., if and only if γ can be expressed as a sum of some subset of the integers $a_1, \ldots, a_n$.

Problems on a simplex

The *standard n-dimensional simplex* is defined to be the following subset of $\mathbb{R}^{n+1}$:

$$\Delta^n = \{(x_1, \ldots, x_{n+1}) : x_1 + \cdots + x_{n+1} = 1; x_i \geq 0 \text{ for } i = 1, \ldots, n+1\}.$$

Fig. 4.1 illustrates Δ^1 and Δ^2. This region is specified by one equation constraint and $n + 1$ inequalities. It has exactly $n + 1$ vertices.

The problem of minimizing $\frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{c}^T\mathbf{x}$ subject to $\mathbf{x} \in \Delta^{n-1}$ is **NP**-hard. This fact was observed by Pardalos, Ye and Han [1989], and the proof follows a theorem by Motzkin and Straus [1965]. We first recall that the following problem is **NP**-hard: Given a graph G, find the largest clique contained as a subgraph in G. A *clique* is a graph with all pairs of vertices interconnected. A clique with m vertices is denoted as K_m. The **NP**-hardness of this problem follows from Theorem 2.3.

We can transform the clique problem to QP on a simplex. The transformation is as follows. Let G be an undirected graph with n vertices numbered

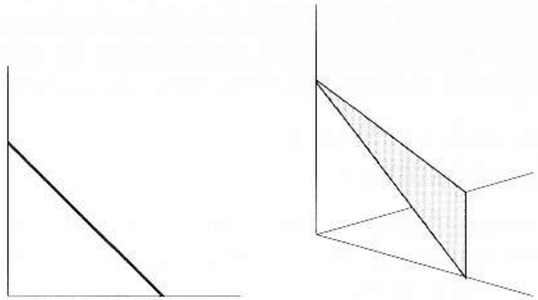

FIG. 4.1. One and two dimensional simplices.

$1, \ldots, n$. There will be a variable x_i for each vertex i. Let $E(G)$ be the edges of G, that is, a set of unordered pairs of the form (i, j) with $1 \leq i, j \leq n$. Let f be the following quadratic function from $\mathbb{R}^n$ to $\mathbb{R}$.

$$f(\mathbf{x}) = - \sum_{(i,j) \in E(G)} x_i x_j \tag{4.2}$$

(only one of the terms $x_i x_j$ and $x_j x_i$ occurs in the sum for each edge $(i, j) \in E(G)$.) We claim that the minimum value attained by $f(\mathbf{x})$ is

$$\frac{1}{2k} - \frac{1}{2}, \tag{4.3}$$

where k is the size of the maximum clique in G.

First, suppose there is a clique of size k. Then we can set the variables corresponding to clique vertices to $1/k$. The remaining variables are zero. Clearly this point is feasible. The objective function in this case will have $k(k-1)/2$ terms each of value $-1/k^2$, so the total objective function value is $-k(k-1)/(2k^2)$ which is $1/(2k) - 1/2$.

Now, we prove that the objective function cannot be lower than p where $p = 1/(2k) - 1/2$ and k is the size of the largest clique of G.

We state this as a lemma, which is due to Motzkin and Straus.

LEMMA 4.1. *The optimum value of f defined by* (4.2) *on Δ^{n-1} is given by* (4.3).

PROOF. We have already shown that the minimum is no more than the value given by (4.3).

The proof for the other direction is by induction on n, the number of vertices of G. First, if $n = 2$ and G has one edge, then the minimum objective function value is $-1/4$. If G has two vertices and no edges, then the objective function is identically zero and the maximum clique size is 1.

Assume that $n > 2$ and examine the optimum value $\mathbf{x}^*$ of the QP. Let k^* be the size of the largest clique of G. We want to prove that $f(\mathbf{x}^*) \geq 1/(2k^*) - 1/2$. There are three cases.

Case 1. *For some i, $x_i^* = 0$.* In this case we can delete vertex i from G to yield a graph G' with $n-1$ vertices. Note that the objective function f' for this graph G' also attains value $f(\mathbf{x}^*)$ (delete the ith entry from $\mathbf{x}^*$). Let $\mathbf{x}'$ be the optimum for f'; then we have just argued that $f(\mathbf{x}^*) \geq f'(\mathbf{x}')$. By the induction hypothesis, $f'(\mathbf{x}') = 1/(2k') - 1/2$, where k' is the size of the largest clique of G'. But G also contains this same clique, so $k^* \geq k'$. Thus we get the chain of inequalities

$$f(\mathbf{x}^*) \geq f'(\mathbf{x}') = \frac{1}{2k'} - \frac{1}{2} \geq \frac{1}{2k^*} - \frac{1}{2}$$

which proves the claim.

Case 2. *For all i, $x_i^* > 0$, and $G \neq K_n$.* In this case, we look at the KKT conditions at optimum. The only active constraint is the constraint $\mathbf{e}^T\mathbf{x} = 1$ where $\mathbf{e}$ is the vector of all 1's. This means that $\nabla f(\mathbf{x}^*) = \lambda \mathbf{e}$ for some value of λ, hence the entries of $\nabla f(\mathbf{x}^*)$ are equal to one another. Since we are assuming that $G \neq K_n$, at least one possible edge is absent. Assume without loss of generality that edge $(1, 2)$ is not present in G. Then we examine the first two entries of $\nabla f(\mathbf{x}^*)$ which are,

$$\frac{\partial f}{\partial x_1}(\mathbf{x}^*) = - \sum_{(1,j) \in E(G)} x_j^*$$

and

$$\frac{\partial f}{\partial x_2}(\mathbf{x}^*) = - \sum_{(2,j) \in E(G)} x_j^*.$$

Since these are equal, we conclude that $f(x_1^* + t, x_2^* - t, x_3^*, \ldots, x_n^*)$ is equal to $f(\mathbf{x}^*)$ for all choices of t (since there is no cross-term $x_1 x_2$, the dependence on t is linear). This means that there is some value of t such that either x_1 or x_2 can be driven to zero and the value of the objective function is preserved. This reduces the problem to Case 1.

Case 3. $G = K_n$. In this case, we can compute analytically the minimum value of the objective function. Notice that the objective function in this case is equal to

$$\frac{(x_1^2 + \cdots + x_n^2) - (x_1 + \cdots + x_n)^2}{2}.$$

The second parenthesized term of the numerator is always 1 because of the constraint $\mathbf{e}^T\mathbf{x} = 1$. Therefore, the problem is to minimize $(\|\mathbf{x}\|_2^2 - 1)/2$. Using the KKT conditions, one can show that $\|\mathbf{x}\|_2$ is minimized on a simplex when each x_i is set to $1/n$. This proves the claim.

Thus, we have shown that the minimum of f is attained when each x_i for the vertices of a maximum clique of size k is set to $1/k$. ∎

4.8 Let θ be the optimal value of the objective function to (4.4), and suppose θ is a root to an integer polynomial of degree d with coefficients at most s in magnitude. Let ζ be a given rational number with numerator and denominator of magnitude at most t. Suppose an approximation θ_1 of θ is computed, such that $|\theta_1 - \theta| \leq 2^{-k}$. How large must k be in order to determine from θ_1 whether $\theta = \zeta$, $\theta < \zeta$, or $\theta > \zeta$?

5

LOCAL OPTIMIZATION

In the last chapter we saw that many versions of nonconvex quadratic programming are **NP**-hard. Indeed, practitioners have realized for years that global minimization of nonconvex problems seems computationally intractable. Accordingly, most well-known optimization packages (e.g., MINOS by Murtagh and Saunders [1987]) try to produce only local optima. In this chapter we investigate the problem of local optimization from a complexity-theoretic point of view. In the first section, we show that the general problem of local optimality for nonconvex QP is an **NP**-hard problem. This result seems surprising since it has been tacitly assumed that local optima are easy to find.

The second section characterizes local minima for QP. In the third, fourth and fifth sections we address the problem of local minima for quadratic knapsack problems. We show that this problem can be solved in polynomial time.

The practitioner of optimization ought to read this chapter with the question, "Is it reasonable to expect that the termination point of an optimization algorithm be a local minimum?" The answer to this question is not fully understood, but in this chapter we provide some positive and negative results.

5.1. General quadratic local minimization

We define QPLOC to be the following decision problem. Given an instance $(H, \mathbf{c}, A, \mathbf{b})$ of quadratic programming, and given a point $\mathbf{x}^* \in \mathbb{R}^n$, is $\mathbf{x}^*$ a local minimum of the problem? In other words, (1) does $\mathbf{x}^*$ satisfy $A\mathbf{x}^* \geq \mathbf{b}$, and (2) if so, does there exist an $\epsilon > 0$ such that $\frac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{c}^T\mathbf{x} \geq \frac{1}{2}\mathbf{x}^{*T} H\mathbf{x}^* + \mathbf{c}^T\mathbf{x}^*$ for all $\mathbf{x}$ satisfying $\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \epsilon$ and $A\mathbf{x} \geq \mathbf{b}$?

The main result of this section is the following theorem. This theorem is due to Murty and Kabadi [1987], although our proof is a simplification because we can take advantage of the Motzkin-Strauss results proved in the last chapter.

THEOREM 5.1. *Problem QPLOC is* **NP**-*hard.*

PROOF. Let G be an undirected graph. We show that the problem of determining whether G has a clique of size k can be reduced to an instance of QPLOC. Accordingly, assume k is given. Define $f(\mathbf{x})$ to be the quadratic objective function given by (4.2) based on graph G. Notice that this objective function may be written as $\mathbf{x}^T H_0 \mathbf{x}$ (no constant or linear term). Define $g(\mathbf{x})$ to be $f(\mathbf{x}) - c$ where

$$c = \frac{1}{2k-1} - \frac{1}{2}.$$

Notice that if the maximum clique size is k, then $g(\mathbf{x})$ will attain a negative value on Δ^{n-1}; otherwise it will always be positive on Δ^{n-1}.

Next, notice that if $\mathbf{x} \in \Delta^{n-1}$ then c is the same as $c(x_1 + \cdots + x_n)^2$. Therefore, we can introduce objective function $h(\mathbf{x})$ defined to be

$$h(\mathbf{x}) = f(\mathbf{x}) - c(x_1 + \cdots + x_n)^2;$$

again, this objective function will attain a negative value on Δ^{n-1} iff G has a clique of size at least k.

Notice that $h(\mathbf{x})$ takes the form $\mathbf{x}^T H\mathbf{x}$ (no constant or linear term). Therefore, we have produced a matrix H such that $\mathbf{x}^T H\mathbf{x}$ is negative on Δ^{n-1} if and only if G has a clique of size at least k.

Let Q be the orthant $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \geq \mathbf{0}\}$. Let a be any positive real number. Consider the problem of minimizing $\mathbf{x}^T H\mathbf{x}$ subject to the constraints $\mathbf{x} \in Q$ and $\mathbf{e}^T\mathbf{x} = a$, where $\mathbf{e}$ is the vector of all 1's. For $a = 1$ these constraints specify Δ^{n-1}. Notice, however, that objective function values scale exactly proportionally to a^2. Therefore, for any fixed choice of a, we can claim that $\mathbf{x}^T H\mathbf{x}$ is negative subject to the preceding constraints if and only if G has a clique of size at least k.

Now, finally, consider the problem of minimizing $\mathbf{x}^T H\mathbf{x}$ subject only to $\mathbf{x} \in Q$. We claim that $\mathbf{0}$ is a local (and, in fact, global) minimum of $h(\mathbf{x})$ on this region if and only if G does not have clique of size at least k. This result would prove the theorem. If G does not have a clique of size k, then the preceding argument shows that $h(\mathbf{x})$ is nonnegative at every point of Q. On the other hand, if G has a clique of size k, then there exists a point $\mathbf{x} \in Q$ such that $h(\mathbf{x}) < 0$, and $\mathbf{e}^T\mathbf{x}$ may be arbitrarily small (but positive). This means that $\mathbf{0}$ cannot be a local minimum. ∎

5.2. Characterizing local minima for quadratic programs

In this section we provide characterizations of local minima of quadratic programming problems. These characterizations are useful for the upcoming sections of this chapter.

We start by giving *second order* conditions for quadratic programming problem:

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T H\mathbf{x} + \mathbf{c}^T\mathbf{x} \\ &\text{subject to} \quad A\mathbf{x} \geq \mathbf{b}. \end{aligned} \tag{5.1}$$

Let $\mathbf{x}^*$ be a feasible point in $\mathbb{R}^n$. Let $\mathbf{g}$ denote $H\mathbf{x}^* + \mathbf{c}$, the gradient of f at $\mathbf{x}^*$. Let $(A_U, \mathbf{b}_U)$ denote the subset of constraints active at $\mathbf{x}^*$, so that $A_U\mathbf{x}^* = \mathbf{b}_U$. We claim that the following two conditions are necessary and sufficient for local minimality of (5.1):

1. For all $\mathbf{d}$ such that $A_U\mathbf{d} \geq \mathbf{0}$, $\mathbf{g}^T\mathbf{d} \geq 0$.
2. For all $\mathbf{d}$ such that $A_U\mathbf{d} \geq \mathbf{0}$ and $\mathbf{g}^T\mathbf{d} = 0$, $\mathbf{d}^T H\mathbf{d} \geq 0$.

Observe that $\mathbf{d}$ is a feasible direction at $\mathbf{x}^*$ if and only if $A_U\mathbf{d} \geq \mathbf{0}$. Note that Condition 1 combined with feasibility is equivalent to the KKT conditions. Condition 2 is the additional *second order condition*.

Recall that a *feasible direction* for an optimization problem at a feasible point $\mathbf{x}^*$ is a vector $\mathbf{d}$ such that there is an $\epsilon > 0$ such that $\mathbf{x}^* + t\mathbf{d}$ is feasible for all $t \in [0, \epsilon]$. A *descent direction* is a vector $\mathbf{d}$ such that there is an $\epsilon > 0$ such that $f(\mathbf{x}^* + t\mathbf{d}) < f(\mathbf{x}^*)$ for all $t \in (0, \epsilon)$. Finally, a vector $\mathbf{d}$ is called a *feasible descent direction* if it is both a feasible direction and a descent direction. Clearly, the presence of a feasible descent direction means that $\mathbf{x}^*$ is not a local minimum.

Before proving that Conditions 1 and 2 are equivalent to local minimality, we first prove the following simpler auxiliary theorem that gives an alternative characterization of the conditions.

THEOREM 5.2. *Conditions* 1 *and* 2 *are equivalent to the condition that there are no feasible descent directions at* $\mathbf{x}^*$.

PROOF. First, suppose Conditions 1 and 2 hold. Then we claim that there are no feasible descent directions. Let $\mathbf{d}$ be a feasible direction at $\mathbf{x}^*$, so that $A_U\mathbf{d} \geq \mathbf{0}$. Then

$$f(\mathbf{x}^* + t\mathbf{d}) - f(\mathbf{x}^*) = t\mathbf{g}^T\mathbf{d} + t^2\mathbf{d}^T H\mathbf{d}/2. \tag{5.2}$$

Condition 1 implies that $\mathbf{g}^T\mathbf{d} \geq 0$. If $\mathbf{g}^T\mathbf{d} > 0$, then the first term of the right-hand side of (5.2) is positive and dominates the second term for t small enough, so $\mathbf{d}$ cannot be a descent direction. If $\mathbf{g}^T\mathbf{d} = 0$ then the second term is nonnegative by Condition 2, hence $\mathbf{d}$ once again is not a descent direction.

Conversely, suppose Condition 1 or Condition 2 fail. Then it is easily seen that the $\mathbf{d}$ that violates either condition is a feasible descent direction. ∎

We now prove the main theorem for this section. This theorem is due to Contesse [1980], but our proof is different.