1

MATHEMATICAL OPTIMIZATION FUNDAMENTALS

This chapter reviews the fundamentals of mathematical optimization and modeling. It starts with a biological network inference problem as a prototype example to highlight the basic steps of formulating an optimization problem. This is followed by a review of some basic mathematical concepts and definitions such as set and function properties and convexity analysis.

1.1 MATHEMATICAL OPTIMIZATION AND MODELING

Mathematical optimization (programming) systematically identifies the best solution out of a set of possible choices with respect to a pre-specified criterion. The general form of an optimization problem is as follows:

$$
\begin{aligned}
&\text{minimize } (\text{or maximize})\ f(\boldsymbol{x}) \\
&\text{subject to} \\
&h(\boldsymbol{x}) = 0 \\
&g(\boldsymbol{x}) \le 0 \\
&\boldsymbol{x} \in S
\end{aligned}
$$

where

- $\boldsymbol{x}$ is a N-dimensional vector referred to as, the vector of *variables*.
- S is the set from which the elements of $\boldsymbol{x}$ assume values. For example, S can be the set of real, nonnegative real or nonnegative integers. In general, variables in an optimization problem can be continuous, discrete (integer) or combinations

Optimization Methods in Metabolic Networks, First Edition. Costas D. Maranas and Ali R. Zomorrodi.

thereof. The former is used to capture the continuously varying properties of a system (e.g., concentrations), whereas the latter is used for discrete decision making (e.g., whether or not to eliminate a reaction).

- $f(\boldsymbol{x})$ is referred to as the *objective function* and serves as a mathematical description of the desired property of the system that should be optimized (i.e., maximized or minimized).
- $h(\boldsymbol{x}) = [h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \ldots, h_L(\boldsymbol{x})]^{\mathrm{T}}$ and $g(\boldsymbol{x}) = [g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_M(\boldsymbol{x})]^{\mathrm{T}}$ are constraints that must be satisfied as equalities or one-sided inequalities, respectively, and represent the feasible space of decision variables.

Any vector $\boldsymbol{x}$ that lies in S and satisfies $h(\boldsymbol{x})$ and $g(\boldsymbol{x})$ is called a feasible solution. In addition, if vector $\boldsymbol{x}$ minimizes (maximizes) the objective function, it is an optimal *solution point* to the optimization problem with an associated optimum *solution value* $f(\boldsymbol{x})$. There are different classes of optimization problems depending on the (non)linearity properties of the objective function and constraints as well as the presence or absence of discrete (i.e., binary or integer) and/or continuous variables. Standard classes of optimization problems that generally require different solution techniques are as follows:

(i) Linear programming (LP) problems involve a linear objective function $f(\boldsymbol{x})$ and constraints $h(\boldsymbol{x})$ and $g(\boldsymbol{x})$ as well as only continuous variables $\boldsymbol{x}$ (Chapter 2).

(ii) Mixed-integer LP (MILP) problems are LP problems with some of the variables assuming only discrete values (Chapter 4).

(iii) Nonlinear programming (NLP) problems involve a nonlinear objective and/or some nonlinear constraints while all variables are continuous (Chapter 9).

(iv) Mixed-integer nonlinear programming (MINLP) problems are NLPs with some variables assuming only discrete values (Chapter 11).

Mathematical optimization has been used extensively to model a wide variety of problems in science and engineering. The development of an optimization formulation modeling a real-life problem often needs to be traversed multiple times as new data, modified problem descriptions and re-interpretations due to unanticipated optimal solutions come to play. This book concentrates on mathematical optimization applications for the analysis and redesign of biological systems, with a special emphasis on metabolic networks. Example 1.1 describes the basic steps for formulating a biological network inference task as an optimization problem.

Example 1.1

Given a set of genes and time-course DNA microarray data, formulate an optimization problem to identify the regulatory interaction coefficients between genes best explaining the observed gene expression levels. The schematic representation of time-course DNA microarray data for a sample gene interaction network is given in Figure 1.1.

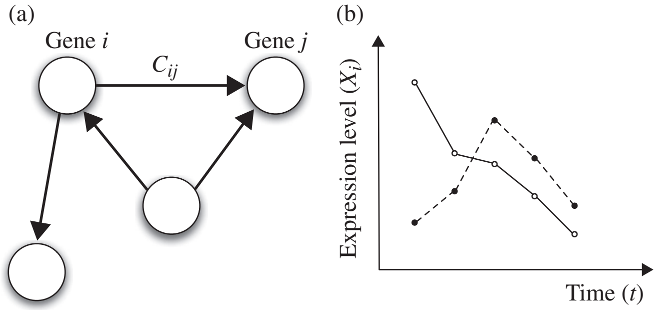

FIGURE 1.1 (a) An example of a simple gene regulatory network. Nodes and edges represent genes and interactions between genes, respectively. C_{ij} denotes the interaction coefficient of genes i and j (i.e., how gene j is affecting gene i). (b) A schematic representation of time-course DNA microarray expression data for two genes. These data are usually presented as the log ratio of the expression level of a gene at each time point with respect to a reference.

Solution: A stepwise description is provided that codifies the sequence of tasks carried out for constructing the optimization model whose solution answers the problem.

Sets:
The first task in deriving the optimization formulation of a problem is defining a number of sets indicating the essential elements of the problem over which the parameters, variables and/or constraints are defined. Two sets can be defined for this problem as follows:

$$\text{Set of genes}: \boldsymbol{I} = \{i \mid i = 1,2,\ldots,N\}$$
$$\text{Set of time points}: \boldsymbol{T} = \{t \mid t = 1,2,\ldots,T_{\mathrm{f}}\}$$

Here, N denotes the number of genes in the network.

Parameters:
Parameters (some of which are indexed over sets) encode the available data for the problem. The parameters that can be defined for this problem include the following:

X_{it}: Expression level of gene $i \in I$ at time point $t \in T$.
LB_{ij}: Lower bound on the interaction coefficient C_{ij} (see the next section for the definition of C_{ij}). Subscript j assumes values from set I.
UB_{ij}: Upper bound on the interaction coefficient C_{ij}.
Δt: Sampling interval assuming that it is constant throughout the experimental DNA microarray data. For convenience we set $\Delta t = 1$.

Variables:
In contrast to parameters that have known values, variables typically only have initial values and/or lower/upper bounds, and their optimal values are obtained upon solving the optimization problem. As was the case with parameters, the introduction of sets allows for the grouping of multiple unknowns under the same variable name. For the

problem in hand, we define two different categories of variables including a continuous and a discrete set of variables. The continuous variable set is defined as follows:

C_{ij}: Interaction coefficient between genes $j \in I$ and $i \in I$ (effect of gene *j* on gene *i*), where

$$\begin{cases} C_{ij} > 0 & \text{if gene } j \text{ activates gene } i \\ C_{ij} < 0 & \text{if gene } j \text{ represses gene } i \\ C_{ij} = 0 & \text{if gene } j \text{ has no effect on gene } i \end{cases}$$

The binary variables y_{ij} capture the presence or absence of an interaction between genes *i* and *j* as follows:

$$y_{ij} = \begin{cases} 1 & \text{if gene } j \text{ affects gene } i \\ 0 & \text{otherwise} \end{cases}$$

Constraints:

Constraints are defined to enforce the conditions that need to be satisfied for the problem. A constraint for this example is needed to impose the assumption that the rate of change in the expression level of each gene is a linear function of the contributions of all genes in the network (including itself):

$$\frac{dX_{it}}{dt} = \sum_{j \in I} C_{ij} X_{jt}, \quad \forall i \in I, \quad t \in T \tag{1.1}$$

Here, we approximate the derivative terms with algebraic linear constraints using a finite (forward) difference approximation:

$$\frac{dX_{it}}{dt} \approx \frac{X_{i,t+1} - X_{i,t}}{\Delta t}, \quad \forall i \in I, t \in \{1, \ldots, T_f - 1\} \tag{1.2}$$

This implies that the identification of C_{ij} requires solving the following under-determined set of linear equalities (note that the system of equations is under-determined because the number of pairwise interactions is much larger than the number of equations):

$$X_{i,t+1} - X_{i,t} = \Delta t \sum_{j \in I} C_{ij} X_{jt}, \quad \forall i \in I, t \in \{1, \ldots, T_f - 1\} \tag{1.3}$$

An additional constraint is introduced to model the presence or absence of an interaction for each pair of genes enforcing the definition of binary variables y_{ij}:

$$LB_{ij} y_{ij} \le C_{ij} \le UB_{ij} y_{ij}, \quad \forall i, j \in I \tag{1.4}$$

Observe that if y_{ij} is equal to zero, then C_{ij} is forced to assume a value of zero; whereas when y_{ij} is equal to one, then C_{ij} is free to assume any value between LB_{ij} and UB_{ij}.

Objective function:

Given that this is an under-determined system of equations, there can be infinite sets of C_{ij} all satisfying the given constraints. Optimization can be used to select one out of the many feasible values for C_{ij} that satisfies an optimality criterion. Here, we invoke the parsimony assumption whereby we accept as the most relevant solution the one that minimizes the total number of regulatory interactions. The total number of regulatory coefficients can be obtained by summation over all binary variables.

Optimization model (formulation):
By collecting all the constraints described earlier, the optimization problem is stated as follows:

$$\text{minimize} \sum_{i\in I}\sum_{j\in I} y_{ij} \qquad [P1]$$

subject to

$$X_{i,t+1} - X_{i,t} = \Delta t \sum_{j\in I} C_{ij} X_{jt}, \quad \forall i \in I, \quad t \in \{1,\ldots,T_{\mathrm{f}} - 1\} \tag{1.3}$$

$$LB_{ij}\, y_{ij} \le C_{ij} \le UB_{ij} y_{ij}, \quad \forall i, j \in I \tag{1.4}$$

$$y_{ij} \in \{0,1\}, C_{ij} \in \mathbb{R}, \qquad \forall i, j \in I$$

The solution of this problem will provide the presence or absence of a regulatory interaction between each pair of genes in the network (captured by binary variable y_{ij}) and the magnitude and sign (i.e., activation vs. inhibition) of these interactions (captured by the continuous variables C_{ij}).

Exploring trade-offs between prediction error and model complexity:
It is important to emphasize that the solution of an optimization problem always needs to be scrutinized in terms of both mathematical accuracy and the relevance to the problem. For example, a key concern for this example is whether the obtained coefficients indeed capture biologically relevant interactions or are simply artifacts of the parameter fitting process. In addition, one might be interested to know whether the identified regulatory coefficients are unique or there exists alternate optimal sets. Optimization provides ways to address these types of questions by trading-off accuracy versus model complexity (i.e., parsimony in this case). This can be accomplished for this example by exploring how the total number of non-zero C_{ij}'s decrease upon allowing for some degree of violation in the equality constraints. Introducing slack variables S_{it}^{+} and S_{it}^{-} $\left(\text{with } S_{it}^{+}, S_{it}^{-} \ge 0\right)$ for Constraint 1.3 allows for both positive and negative departures from equality:

$$X_{i,t+1} - X_{i,t} - \Delta t \sum_{j\in I} C_{ij} X_{jt} = S_{it}^{+} - S_{it}^{-}, \quad \forall i \in I, t \in \{1,\ldots,T_{\mathrm{f}} - 1\} \tag{1.5}$$

In addition, since we would like to identify a regulatory network with fewer interactions (e.g., one less than the interactions identified for the original problem represented by $y^{\max}$), we can add the following constraint:

$$\sum_{i\in I}\sum_{j\in I} y_{ij} \le y^{\max} - 1 \tag{1.6}$$

The re-formulated optimization problem aims to identify a more compact regulatory interaction network while minimizing the departure from experimental data and is described as follows:

$$\text{minimize} \sum_{i\in I} \sum_{t\in T-\{T_f\}} \left(S_{it}^{+} + S_{it}^{-}\right) \qquad [P2]$$

subject to

$$X_{i,t+1} - X_{i,t} - \Delta t \sum_{j\in I} C_{ij} X_{jt} = S_{it}^{+} - S_{it}^{-}, \quad \forall i \in I, t \in \{1,\ldots,T_f - 1\} \qquad (1.5)$$

$$\mathrm{LB}_{ij} y_{ij} \le C_{ij} \le \mathrm{UB}_{ij} y_{ij}, \quad \forall i, j \in I \qquad (1.4)$$

$$\sum_{i\in I} \sum_{j\in I} y_{ij} \le y^{\max} - 1 \qquad (1.6)$$

$$y_{ij} \in \{0,1\}, C_{ij} \in \mathbb{R}, \; S_{it}^{+}, S_{it}^{-} \ge 0, \quad \forall i, j \in I$$

Note that in contrast to [*P*1], [*P*2] minimizes the total violation of the Constraint 1.3. By solving [*P*2] for different network sizes specified by the right-hand side of Constraint 1.6 and by subsequently plotting the total error in prediction (i.e., the objective function value of [*P*2]) against the number of nonzero regulatory interactions (i.e., sum of the binary variables), a monotonically decreasing curve is obtained, as shown in Figure 1.2. The error will be quite high for very sparse models, but will approach zero as the total number of regulatory interactions approaches $y^{\max}$. In general, there tends to be a *break point* in the curve beyond which additional nonzero regulatory interactions improve the error only slightly as shown in Figure 1.2. This implies that once this point (or a desired accuracy threshold) is reached, additional parameters are likely to "overfit" rather than capture information in the data.

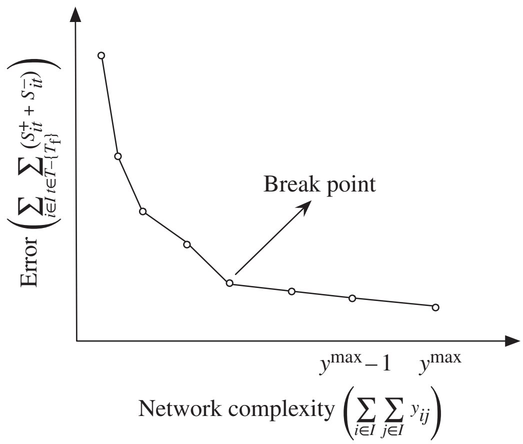

FIGURE 1.2 Schematic representation of the error as a function of the number of nonzero regulatory coefficients for the problem of Example 1.1.

In general, by adopting an optimization-based description of the problem, significant versatility is afforded in tailoring the solution to the specifics of the problem and/or exploring various trade-offs. For example, certain regulatory interactions can be excluded or pre-postulated (e.g., the interaction of known transcription factors with known genes) by setting the related binary variable y_{ij} to zero or one, respectively. Similarly, the total number of genes affecting the expression of a gene i can be restricted to a pre-specified number M using the following constraint:

$$\sum_{j \in I} y_{ij} \leq M \tag{1.7}$$

Alternate representations of this problem can be explored to account for nonlinear interactions, noise in gene expression data, time delay in regulatory interactions, and more. Interested readers are referred to the related articles [1–7] for details. □

The purpose of this illustrative example was to provide an introduction to the iterative process of formulating optimization problems, assessing their output and modifying their structure to address the follow-up questions when dealing with a real-life problem. Next, basic definitions and concepts necessary for correctly describing optimization models and assessing the existence of local and/or global optimal solution points are introduced.

1.2 BASIC CONCEPTS AND DEFINITIONS

We start by introducing basic properties of sets and functions necessary for establishing conditions for the (i) existence and (ii) uniqueness of a global optimum value. These definitions also introduce formal mathematical language and reasoning used in optimization textbooks and articles.

Let S be an arbitrary subset of $\mathbb{R}^N$. The concepts and properties for S are defined in the text.

1.2.1 Neighborhood of a Point

Given a point $\boldsymbol{x}$ in set S (i.e., $\boldsymbol{x} \in S$), an ε-*neighborhood* around x (denoted by $B(\boldsymbol{x},\varepsilon)$) is defined as follows (see Fig. 1.3):

$$B(\boldsymbol{x},\varepsilon) = \{\boldsymbol{y} \mid \boldsymbol{y} \in S \text{ and } \|\boldsymbol{x} - \boldsymbol{y}\| < \varepsilon\}$$

$B(\boldsymbol{x},\varepsilon)$ is in essence the set of points in set S within an N-dimensional sphere centered at point $\boldsymbol{x}$ with a radius of ε.

1.2.2 Interior of a Set

A point $\boldsymbol{x}$ is in the interior of a set S (denoted by int(S)) if and only if there exists an ε-neighborhood around $\boldsymbol{x}$ with $B(\boldsymbol{x},\varepsilon) \subseteq S$ for some $\varepsilon > 0$.

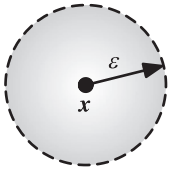

FIGURE 1.3 Schematic representation of the ε-neighborhood of a point $\boldsymbol{x} \in S$ in a two-dimensional space.

This qualitatively means that a "ball" of nonzero size can be constructed around point $\boldsymbol{x}$ so as all points within it belong to set S. This implies that interior of sets exclude "boundary" points.

1.2.3 Open Set

A set S is open if and only if $\text{int}(S) = S$.

This implies that for an open set, nonzero neighborhoods can be constructed around each point so as every point in the neighborhood belongs to the original set. Thus, open sets are identical to their interior as no boundary points are included. For example, set (0,1) is an open set as neighborhoods can be constructed for every point in it that are fully contained within the set by making ε appropriately small.

1.2.4 Closure of a Set

A point $\boldsymbol{x}$ is in the closure of set S (denoted by $cl(S)$) if and only if $\mathrm{S} \cap \mathrm{B}(\mathbf{x}, \varepsilon) \neq \varnothing$ for every $\varepsilon > 0$.

The closure of a set can be thought of as all the points in a set and all adjacent boundary points irrespective of whether they are part of the original set.

1.2.5 Closed Set

A set S is closed if and only if $cl(S) = S$.

In essence, a closed set contains all of its boundary points, and therefore its closure is identical with the original set. For example, set [1,2] is a closed set, whereas set (1,2] is neither open nor closed.

1.2.6 Bounded Set

A set S is bounded if and only if for every two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ there exists $M > 0$ such that $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| < M$.

A set is bounded if any two points within it are only a finite distance apart. This implies that the set of all real numbers $\mathbb{R}$ is unbounded.

1.2.7 Compact Set

The set S is said to be compact if and only if it is both closed and bounded.

Set compactness is important because it guarantees the existence of an optimum solution point when an objective function (that is continuous) is optimized over it. Next, we transition from set properties to function properties underlying optimality conditions for an optimization problem.

1.2.8 Continuous Functions

Function $f : S \to \mathbb{R}$ is continuous at a point $x_0 \in S$, if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$ for $\boldsymbol{x} \in S$ implies that $|f(\boldsymbol{x}) - f(x_0)| < \varepsilon$.

The definition of function continuity implies that no matter how small an arbitrary number ε is chosen, a point $\boldsymbol{x}$ close to $\boldsymbol{x}_0$ (i.e., in the δ-neighborhood of $\boldsymbol{x}_0$) can be found such that the difference between the function values at $\boldsymbol{x}_0$ and $\boldsymbol{x}$ is less than ε. Function continuity can be thought of as the absence of any "breaks" in the line that plots the function. As mentioned before, continuous functions optimized over compact, nonempty sets are guaranteed to have an optimal solution value and point.

1.2.9 Global and Local Minima

Let $f : S \to \mathbb{R}$, where S is a nonempty subset of $\mathbb{R}^N$.

- A point $\boldsymbol{x}^* \in S$ is a *global minimum point* of f if for every point $\boldsymbol{x} \in S, f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$. $f(\boldsymbol{x}^*)$ is referred to as the *global minimum value* of f over set S.
- A point $\boldsymbol{x}^* \in S$ is a *local minimum point* of f, if there exists $\varepsilon > 0$ such that for every point $\boldsymbol{x} \in N(\boldsymbol{x}^*, \varepsilon) \cap S$ we have $f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*)$.

Therefore, local minimality applies only around a neighborhood $N(\boldsymbol{x}^*, \varepsilon) \cap S$ of the minimum point, whereas global minimality applies over the entire set S. It is possible to have multiple local optimum points and values; however, there is a unique global minimum value. This value may be attainable at multiple points (alternate global minimum points). If we have strict inequalities in the earlier definitions, the point $\boldsymbol{x}^*$ is referred to as a *strict* global or local minimum, respectively. Note that *global and local maxima* are defined in a similar manner.

1.2.10 Existence of an Optimal Solution

After introducing concepts related to set compactness, function continuity and definitions of optimality, the following optimum solution existence criterion can be formally stated. Consider the following general unconstrained optimization problem:

$$\underset{x \in S}{\min \text{ or } \max} f(x)$$

where $S \subseteq \mathbb{R}^N$ and $f : S \to \mathbb{R}$ is continuous on S. An optimal solution x^* exists if S is a nonempty and compact set.

This implies that unconstrained optimization problems over compact sets are guaranteed to have an optimal solution. In practice, many (constrained or unconstrained) optimization problems are originally described over unbounded or open sets with sometimes discontinuities in the objective function. It is a good practice to set finite lower and upper bounds on all variables (i.e., set compactness) and eliminate discontinuities in the objective function to ensure the existence of an optimal solution.

The derivation of uniqueness criteria for an optimum solution value (i.e., a single local optimum that is also a global optimum) hinges upon the concept of convexity. Testing for convexity is facilitated by the establishment of differentiability properties for the objective function (and constraints).

1.3 CONVEX ANALYSIS

The concept of convexity is central in optimization because it provides the means for proving the existence of a global optimal solution value (or point). Here, we provide a brief description of convexity of (i) a set, (ii) a function at a point and (iii) a function over an entire set. Interested readers are encouraged to refer to optimization textbooks such as Refs. [8.14] for more details.

1.3.1 Convex Sets and Their Properties

Convex Combination of Two Points Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$. Any point $(1-\lambda)\boldsymbol{x}_1 + \lambda \boldsymbol{x}_2$ with $\lambda \in [0,1]$ is referred to as a *convex combination* of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. If $\lambda \in (0,1)$, then it is a *strict* convex combination.

Convex Set Set S is convex if the line segment connecting any two points in the set also lies completely within the set. In mathematical language, S is a convex set if and only if for every two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ their convex combinations $(1-\lambda)\boldsymbol{x}_1 + \lambda \boldsymbol{x}_2$ for every $\lambda \in [0,1]$ is also within S. If this condition holds for every strict convex combination of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, then the set S is a *strict* convex set. Any set not satisfying these requirements is a nonconvex set. Examples of convex and nonconvex sets are shown in Figure 1.4.

Special cases of convex sets frequently arise in the treatment of LP problems (i.e., extreme points, hyperplanes, half-spaces, rays, extreme directions and convex cones). The following provides their definitions.

Extreme Points A point $\boldsymbol{x} \in S$ where S is convex, is an *extreme* point of S if it cannot be represented as the strict convex combination of two distinct points in S. Therefore, if $\boldsymbol{x} = (1-\lambda)\boldsymbol{x}_1 + \lambda \boldsymbol{x}_2$ for $\lambda \in (0,1)$ and $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$, then $\boldsymbol{x} = \boldsymbol{x}_1 = \boldsymbol{x}_2$ [8].

Hyperplanes Hyperplanes are an extension of straight lines in $\mathbb{R}^2$. A hyperplane H in $\mathbb{R}^N$ is defined as $H = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = k\}$, where $\boldsymbol{a} \neq 0$ and $k \in \mathbb{R}$. The vector $\boldsymbol{a}$ is

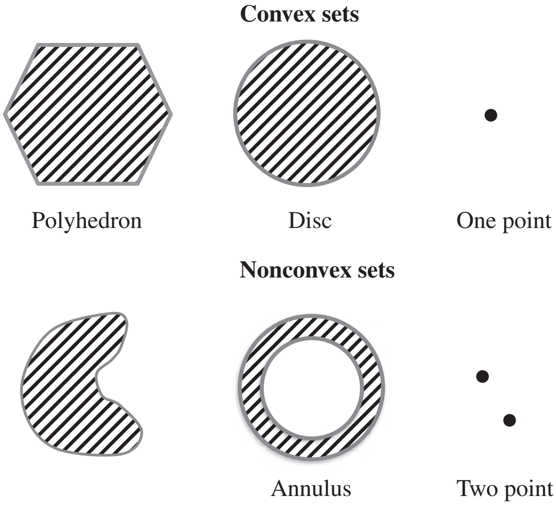

FIGURE 1.4 Examples of convex and nonconvex sets.

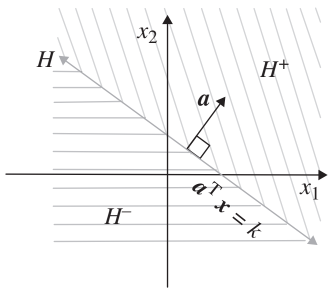

FIGURE 1.5 A schematic representation of a hyperplane and its corresponding half-spaces in a two-dimensional space.

called the *normal* of H as it is the gradient of the linear function $f(\boldsymbol{x}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$ and is thus normal to the hyperplane (see Fig. 1.5). Hyperplanes are a central concept in LP (Chapter 2) and in the analysis of metabolic networks arising in metabolite balances under steady state (Chapter 6).

Half-Spaces Each hyperplane divides $\mathbb{R}^N$ into two half-spaces. If H is a hyperplane, as defined earlier, then sets $H^- = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^N,\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \leq k\}$ and $H^+ = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^N,\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \geq k\}$ are the corresponding half-spaces and are convex (see Fig. 1.5). Sets $H^- = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^N,\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} < k\}$ and $H^+ = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^N,\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} > k\}$ are *open* half-spaces. The imposition of any bounds (lower or upper) on the total metabolic flow through a metabolite (see Chapter 6) gives rise to a half-space constraint.

Example 1.2

H as defined in the following is an example of a hyperplane:

$$H = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \middle| \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3, x_1 + x_2 + x_3 = 10 \equiv \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 10 \right\}$$

The normal for H is as follows:

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

The half-spaces defined by H are as follows:

$$H^+ = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \middle| \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3, x_1 + x_2 + x_3 \geq 10 \equiv \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \geq 10 \right\}$$

$$H^- = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \middle| \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3, x_1 + x_2 + x_3 \leq 10 \equiv \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \leq 10 \right\}$$

□

Rays A ray is a set of points on a line defined as $\{\boldsymbol{y} \mid \boldsymbol{y} = \boldsymbol{x}_0 + \lambda \boldsymbol{d}, \lambda \geq 0, \boldsymbol{d} \neq 0\}$, where $\boldsymbol{x}_0$ and $\boldsymbol{d}$ are the *vertex* and *direction* of the ray, respectively. It is easy to verify that the set describing a ray is convex.

Direction of a Convex Set A nonzero vector $\boldsymbol{d}$ is called a direction of the convex set S, if a ray with vertex $\boldsymbol{x}_0$ and direction $\boldsymbol{d}$ is contained in S for every $x_0 \in S$. Obviously, if the set S is bounded, it has no directions.

Extreme Direction and Extreme Ray of a Convex Set The concept of an extreme direction is similar to that of an extreme point. A direction of a convex set S is called an *extreme* direction if it cannot be represented as a positive combination of any two distinct directions of S, that is, if $\boldsymbol{d} = \lambda_1 \boldsymbol{d}_1 + \lambda_2 \boldsymbol{d}_2,\ \lambda_1, \lambda_2 \geq 0$ then $\boldsymbol{d} = \boldsymbol{d}_1 = \boldsymbol{d}_2$ (see Fig. 1.6). Any ray whose vertex is an extreme point and direction is an extreme direction is called an extreme ray.

Convex Cone A convex set C is called a convex cone if $\lambda \boldsymbol{x} \in C$ for each $\boldsymbol{x} \in C$ for all $\lambda \geq 0$ (see Fig. 1.7). Each cone contains the origin (for $\lambda = 0$) and at least one ray with vertex at the origin. A convex cone can be viewed as a convex set where all points on a line linking the origin and any point in the set also belongs to the set.

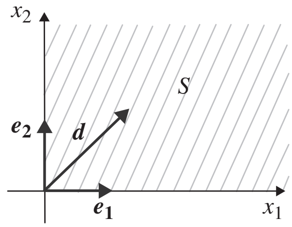

FIGURE 1.6 Examples of a direction (i.e., $\boldsymbol{d}$) and extreme directions (i.e., $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$) of a convex set S.

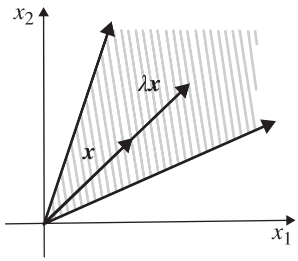

FIGURE 1.7 An example of a convex cone. For any $\boldsymbol{x} \in \boldsymbol{C}$ and $\lambda \geq 0$, we have $\lambda \boldsymbol{x} \in \boldsymbol{C}$.

The concepts of the extreme direction/ray of a convex cone are cornerstones in the extreme pathway analysis of metabolic networks [9–13].

Polyhedral Set and Polyhedral Cone A polyhedral set is the intersection of a finite number of half-spaces. A polyhedral cone is a polyhedral set, whose half-spaces pass through the origin.

For example, set $S = \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{Ax} \leq \boldsymbol{b}, \boldsymbol{x} \geq 0\}$ is a polyhedral set as it is the intersection of N half-spaces defined by $\boldsymbol{Ax} \leq \boldsymbol{b}$ and the half-space defined by $\boldsymbol{x} \geq 0$. Extreme points of this polyhedral set are the intersections of the half-spaces [8]. The feasible regions of LP problems correspond to polyhedral sets. As we will see in Chapter 2, the solution of LP problems always lies on an extreme point of this polyhedral set.

1.3.2 Convex Functions and Their Properties

Set convexity is important as it ensures that every point within the set (except extreme points) is reachable as a linear combination of others. This has implications for the design of algorithms that search for the optimum within convex sets. Proving set convexity is cumbersome as every two point combination must be tested. Functions provide a way of circumventing this challenge by testing for set convexity through an equivalent function convexity criterion.

Convex and Concave Function Definitions A function defined over a convex set S is *convex over set S* if a line connecting any two points on the function lies above the function. Stated formally, function $f : S \to \mathbb{R}$ is convex in S if and only if for every two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ and every $\lambda \in [0,1]$ we have:

$$f\left(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2\right) \le \lambda f\left(\boldsymbol{x}_1\right) + (1-\lambda) f\left(\boldsymbol{x}_2\right) \tag{1.8}$$

- Function f is *strictly* convex if we have a strict inequality in Equation 1.8.
- A function $f : S \to \mathbb{R}$ is concave if and only if $-f$ is convex.
- A function f is *convex at the point* $\bar{\boldsymbol{x}} \in S$, if $f(\lambda \bar{\boldsymbol{x}} + (1-\lambda)\boldsymbol{x}) \le \lambda f(\bar{\boldsymbol{x}}) + (1-\lambda) f(\boldsymbol{x})$ for each $\lambda \in (0,1)$ and each $\boldsymbol{x} \in S$.

Figure 1.8 illustrates some examples of convex and nonconvex functions. It is possible for a function to be nonconvex over a set but convex within a defined subset. For example, the function shown in Figure 1.9 is nonconvex in set S_2, but is convex within set S_1. Another example is $f(x) = x^3$, which is nonconvex in $\mathbb{R}$, but is convex for $x \ge 0$.

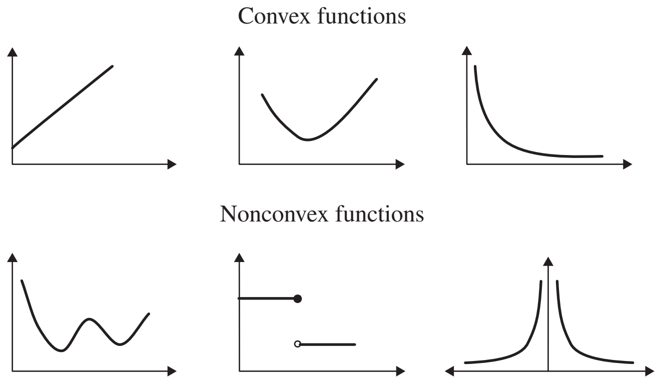

FIGURE 1.8 Examples of convex and nonconvex functions.

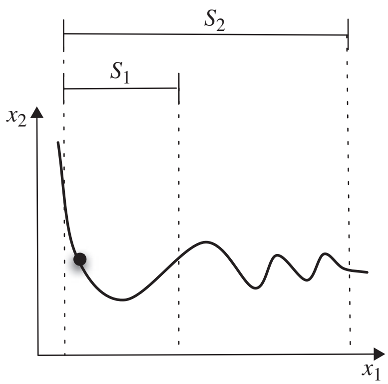

FIGURE 1.9 Convexity of a function within different sets. This function is nonconvex in S_2 but is convex within S_1 and at the point indicated on the graph.

Properties of Convex and Concave Functions If $f, g: S \to \mathbb{R}$ are convex functions in S:

- $f + g$ is convex in S.
- λf is convex in S if $\lambda > 0$.
- max $(f(\boldsymbol{x}), g(\boldsymbol{x}))$ is convex.
- min $(f(\boldsymbol{x}), g(\boldsymbol{x}))$ is generally nonconvex.

Connection of Set Convexity with Function Convexity Let $f(\boldsymbol{x})$ be a convex function in set S, then set $S_c = \{\boldsymbol{x} \in S \mid f(\boldsymbol{x}) \le c\}$, where c is an arbitrary scalar, is a convex set. The statement is true in the reverse direction as well. Set S_c is also referred to as the *level set* of function f.

Proof: To prove that S_c is convex, we need to show that for any two arbitrary points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in S_c we have $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in S$ for $\lambda \in [0,1]$. In other words, we need to show that $f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \le c$. Since $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S_c$, we have $f(\boldsymbol{x}_1) \le c$ and $f(\boldsymbol{x}_2) \le c$. Therefore, $\lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2) \le \lambda c + (1-\lambda)c = c$. Also, since $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ and f is convex, $f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \le \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2)$. The proof then follows directly from the last two inequalities. The proof for the reverse direction is derived in a similar fashion. □

This is a very important result because it allows testing for set convexity by testing the convexity properties of the functions defining the set as their level set. Note that since the intersection of convex sets is a convex set, the feasible region described by a number of constraints corresponding to convex level sets is also a convex set. This implies that the convexity of the region defined by a set of inequality constraints can be inferred by testing the convexity of each individual function. Figure 1.10 illustrates convex sets arising from level sets associated with convex functions.

Testing for the convexity of a function based on the definitions already provided is often cumbersome. Much more tractable representations of convexity can be drawn by using the partial (first and second-order) derivatives of the function.

Differentiable Functions Let S be a nonempty subset of $\mathbb{R}^N$. A function $f: S \to \mathbb{R}$ is *differentiable* at $\bar{\boldsymbol{x}} \in \text{int}(S)$ if and only if f is continuous at $\bar{\boldsymbol{x}}$ and for each $\Delta \boldsymbol{x}$,

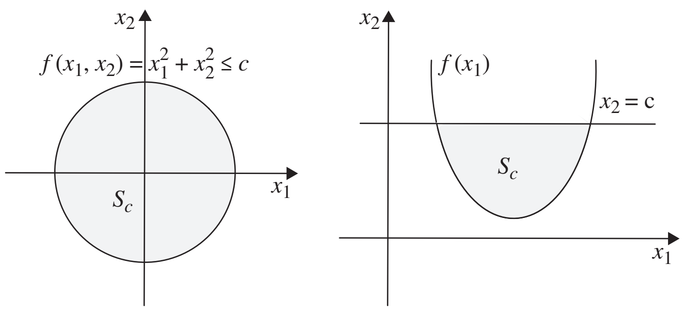

FIGURE 1.10 Examples of convex level sets associated with convex functions.

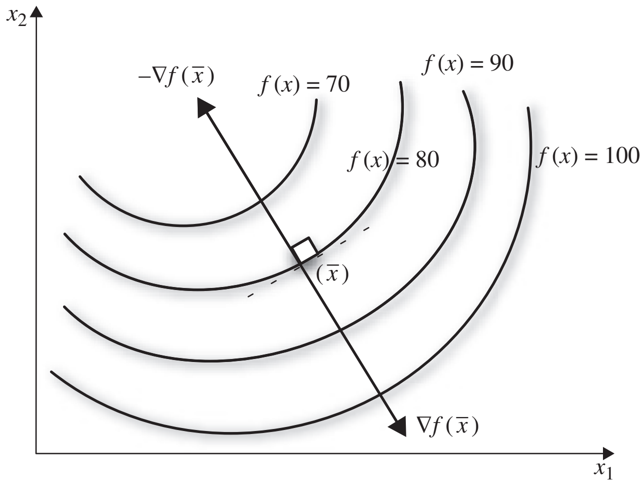

FIGURE 1.11 The gradient vector at any given point is normal to the level set of the function at that point and represents the direction of the steepest ascent.

where $\bar{\boldsymbol{x}} + \Delta \boldsymbol{x} \in S$, there exists a vector $\nabla f(\bar{\boldsymbol{x}})$ (called the *gradient vector*) and a function $\alpha : \mathbb{R}^{\mathrm{N}} \to \mathbb{R}$ such that

$$f(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla^{\mathrm{T}} f(\bar{\boldsymbol{x}}) \Delta \boldsymbol{x} + \alpha(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}) \|\Delta x\| \tag{1.9}$$

where

$$\nabla f(\bar{\boldsymbol{x}}) = \begin{pmatrix} \dfrac{\partial f(\bar{\boldsymbol{x}})}{\partial x_1} & \dfrac{\partial f(\bar{\boldsymbol{x}})}{\partial x_2} & \cdots & \dfrac{\partial f(\bar{\boldsymbol{x}})}{\partial x_N} \end{pmatrix}^{\mathrm{T}} \tag{1.10}$$

and $\lim_{\Delta x \to 0} \alpha(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}) = 0$. This definition implies that the linear approximation (or first-order Taylor expansion) of the function f at any point $\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}$ becomes equal to $f(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x})$ as $\Delta \boldsymbol{x}$ approaches zero from any direction. Function f is differentiable on an open set $S_{\mathrm{o}} \subset S$ if it is differentiable for every point in S_{o}. The gradient of a function points to the direction of greatest increase (*steepest ascent*). Similarly, the negative of the gradient vector represents the direction of the *steepest descent* (see Fig. 1.11). In addition, the gradient of a function at any given point is normal to the level sets of the function at that point.

Twice Differentiable Functions Let $f : S \to \mathbb{R}$, where S is a nonempty set in $\mathbb{R}^N$. f is *twice differentiable* at $\bar{\boldsymbol{x}} \in int(S)$ if and only if f is continuous at $\bar{\boldsymbol{x}}$ and for each $\Delta \boldsymbol{x}$, where $\bar{\boldsymbol{x}} + \Delta \boldsymbol{x} \in S$, there exists a gradient vector $\nabla^{\mathrm{T}} f(\bar{\boldsymbol{x}})$, a $N \times N$ (symmetric) matrix $\boldsymbol{H}(\bar{\boldsymbol{x}})$ (*Hessian* matrix) and a function $a : \mathbb{R}^N \to \mathbb{R}$ such that

$$f(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla^{\mathrm{T}} f(\bar{\boldsymbol{x}}) \Delta \boldsymbol{x} + \frac{1}{2} (\Delta \boldsymbol{x})^{\mathrm{T}} \boldsymbol{H}(\bar{\boldsymbol{x}}) \Delta \boldsymbol{x} + \alpha(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}) \|\Delta \boldsymbol{x}\|^2 \tag{1.11}$$

where

$$\boldsymbol{H}(\bar{\boldsymbol{x}}) = \begin{pmatrix} \frac{\partial}{\partial x_1}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_1}\right) & \frac{\partial}{\partial x_1}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_2}\right) & \cdots & \frac{\partial}{\partial x_1}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_N}\right) \\ \frac{\partial}{\partial x_2}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_1}\right) & \frac{\partial}{\partial x_2}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_2}\right) & \cdots & \frac{\partial}{\partial x_2}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_N}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_N}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_1}\right) & \frac{\partial}{\partial x_N}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_2}\right) & \cdots & \frac{\partial}{\partial x_N}\left(\frac{\partial f(\bar{\boldsymbol{x}})}{\partial x_N}\right) \end{pmatrix} \tag{1.12}$$

and $\lim_{\Delta x \to 0} \alpha(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}) = 0$. As was the case for first-order differentiability, second-order differentiability implies that the quadratic approximation (or second-order Taylor expansion) of the function *f* at any point $\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}$ becomes equal to $\boldsymbol{f}(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x})$ as $\Delta \boldsymbol{x}$ approaches zero from any direction.

Convexity Check for Differentiable Functions Let $f: S \to \mathbb{R}$, where *S* is a nonempty open convex set in $\mathbb{R}^N$. If *f* is differentiable for every point $\boldsymbol{x}$ in *S*, then it is *convex at a point* $\bar{\boldsymbol{x}} \in S$ if and only if:

$$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \nabla^{\mathrm{T}} f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}), \quad \forall \boldsymbol{x} \in S \tag{1.13}$$

Similarly, *f* is *concave at a point* $\bar{\boldsymbol{x}} \in S$ if and only if

$$f(\boldsymbol{x}) \leq f(\bar{\boldsymbol{x}}) + \nabla^{\mathrm{T}} f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}), \quad \forall \boldsymbol{x} \in S \tag{1.14}$$

In other words, a convex (concave) function always lies above (below) its first-order (linear) approximation at any point $\bar{\boldsymbol{x}} \in S$, respectively. Strict convexity or concavity can be established in a similar manner by converting the inequality signs in Constraints 1.13 and 1.14 to strict inequalities. Convexity or concavity of a function *f* at a given point $\bar{\boldsymbol{x}} \in S$ can be extended for set S, if Constraints 1.13 or 1.14, respectively, apply for every point $\bar{\boldsymbol{x}} \in S$.

Convexity Check for Twice Differentiable Functions Based on Hessian Matrix If a function *f* is twice differentiable within a set *S*, then the information contained within the second-order partial derivatives can be used to test/prove the convexity of function *f* over the set. Let $f: S \to \mathbb{R}$, where *S* is a nonempty open convex set in $\mathbb{R}^N$ and *f* is twice differentiable in *S*. Function *f* is convex in *S* if and only if its Hessian matrix $\boldsymbol{H}(\boldsymbol{x})$ is positive semidefinite (psd) for every point in *S*. (Note that a matrix $\boldsymbol{M}$ is psd if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in S \subseteq \mathbb{R}^N, \boldsymbol{x} \neq \boldsymbol{0}$.)

Proof: (a) We first provide the proof in the forward direction, that is, we show that if *f* is convex, then $\boldsymbol{H}(\boldsymbol{x})$ is psd for every point in *S*. Let $\bar{\boldsymbol{x}}$ be an arbitrary point in *S*.

It follows from the convexity of f that $f(\bar{\boldsymbol{x}}+\lambda\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}})+\nabla^{\mathrm{T}} f(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}}),\ \forall \boldsymbol{x} \in S$ (see Equation 1.13). Also, since f if twice differentiable, we have (see Equation 1.11):

$$f(\boldsymbol{x})=f(\bar{\boldsymbol{x}})+\nabla^{\mathrm{T}} f(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})+\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}} \boldsymbol{H}(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}})+\alpha(\boldsymbol{x}-\bar{\boldsymbol{x}})\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^{2}$$

Therefore, by combining the last two expressions, we have $\frac{1}{2}(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}} \boldsymbol{H}(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}}) +\alpha(\boldsymbol{x})\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^{2} \geq 0$. As $\boldsymbol{x}$ approaches $\bar{\boldsymbol{x}}$, function $\alpha(\boldsymbol{x}-\bar{\boldsymbol{x}})$ approaches zero implying that $(\boldsymbol{x}-\bar{\boldsymbol{x}})^{\mathrm{T}} \boldsymbol{H}(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}}) \geq 0$ which completes the proof. (b) The proof in the reverse direction proceeds in a similar fashion. □

A concave, strictly convex or strictly concave function is associated with a *negative semidefinite (nsd)*, *positive definite (pd)* or *negative definite (nd)* Hessian matrix, respectively. A matrix is psd, nsd, pd or nd if all of its eigenvalues are nonnegative, nonpositive, positive or negative, respectively. A matrix that is neither psd nor nsd is called *indefinite*. This result enables checking the convexity properties of a function (and consequently of a set) by inspecting the eigenvalues of the corresponding Hessian matrix.

Example 1.3

Check whether the following function is concave or convex:

$$f(x_1, x_2, x_3)=x_1^2+x_2^2+x_3^2+2x_1x_2$$

Solution: This Hessian matrix for this quadratic function is as follows:

$$\boldsymbol{H}=\begin{bmatrix} 2 & 2 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

The eigenvalues of the Hessian matrix can be obtained by solving the characteristic equation $\det(\boldsymbol{H}-\lambda\boldsymbol{I})=0$, where det represents the determinant of a matrix. The eigenvalues of the Hessian matrix are $\lambda=0,2,4 \geq 0$ implying that it is psd and the function is convex. □

The convexity properties of the following frequently encountered functions can be established by checking the eigenvalues of their Hessian matrix:

- $f(x)=\log(x)$ is concave.
- $f(x,y)=xy$ is neither convex nor concave.
- $f(x,y,z)=xyz$ is neither convex nor concave.
- $f(x,y,z)=\dfrac{x}{y}$ is neither convex nor concave.
- $f(x,y)=\dfrac{x^2}{y}$ is convex.
- $f(x,y)=\sqrt{xy}$ is concave.
- $f(x,y)=\dfrac{1}{xy}$ is convex.

As described earlier, the multiplication of a convex function with a positive scalar and the sum of convex functions yield convex functions. Therefore, complex expressions can be analyzed for their convexity by disassembling them into smaller terms and analyzing each one separately. For example, function $f(x,y)=x^2/y+x\log(x)$ is convex as the sum of two convex functions. In some cases, establishing (or refuting) convexity may require the recombination of various terms. For example, the convexity of function $g(x,y)=(x+y)^2-2xy$ as the sum of a convex $(x+y)^2$ and a nonconvex $(-2xy)$ function cannot be initially determined. However, upon combining the two functions, a single convex function $(x-y)^2$ emerges.

1.3.3 Convex Optimization Problems

Upon establishing a connection between set convexity and function convexity and an eigenvalue-based test, the following optimum solution value uniqueness test can be applied. An optimization problem is convex if the objective function and all inequality constraints are convex and all equality constraints are linear. It can be schematically represented as follows:

$$\begin{aligned}
&\text{minimize}\ (\text{convex function})\\
&\text{subject to}\\
&(\text{convex function}) \le 0\\
&(\text{linear function}) = 0\\
&x \in \text{compact set}
\end{aligned}$$

A key attribute of convex optimization problems is that there exists a unique local (and thus global) *minimum value* (see Chapter 9). In addition, if the objective function is strictly convex, then the point associated with the global minimum value is also unique (*global minimum point*). Consequently, LP problems described by linear objective functions and constraints are convex optimization problems. This implies that the optimal solution value of an LP problem (if one exists) is unique (see also Chapter 2).

Example 1.4

The following optimization problem is an example of a convex problem where the objective function is convex and all constraints are linear and thus are convex.

$$\begin{aligned}
&\text{minimize} \quad x^2+y^2\\
&\text{subject to}\\
&x+y \le 3\\
&y \ge \frac{1}{x}\\
&x \ge 0
\end{aligned}$$

A graphical presentation of the feasible space and the objective function is given in Figure 1.12. □

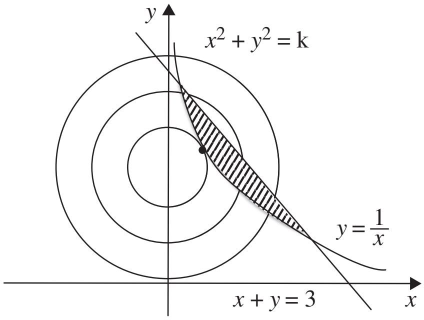

FIGURE 1.12 Graphical representation of the convex optimization problem in Example 1.4. The convexity of the problem implies that we have a global minimum.

1.3.4 Generalization of Convex Functions

Generalized forms of convexity such as quasi- and pseudoconvexity suffice to guarantee global minimality even if the function is nonconvex under certain conditions (see Chapter 9). In the following, we provide only the definitions of these generalized forms. Interested readers are encouraged to refer to standard nonlinear optimization textbooks (e.g., Ref. [14]) for more details.

Quasiconvex Functions f is quasiconvex in S if for every two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ and each $\lambda \in (0,1)$, we have

$$f\left(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2\right) \leq \max\left\{f\left(\boldsymbol{x}_1\right), f\left(\boldsymbol{x}_2\right)\right\} \tag{1.15}$$

This definition implies that a function is quasiconvex within a set, if for any two points within the set the value of the function at any of the points in the line segment connecting the two points, is less than or equal to the larger of the values that the function attains at the two examined points.

Pseudoconvex Functions A function f (differentiable within S) is pseudoconvex if for every two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ with $\nabla^{\mathrm{T}} f(\boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2) > 0$, we have $f(\boldsymbol{x}_1) > f(\boldsymbol{x}_2)$, or equivalently for every two points $\boldsymbol{x}_1, \boldsymbol{x}_2 \in S$ with $f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2)$, we have $\nabla^{\mathrm{T}} f(\boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2) \leq 0$.

EXERCISES

1.1 Show whether the following functions are convex, concave or neither:

(a) $f(x) = \sin(x), x \in [0, \pi]$.

(b) $f(x) = x^k, x \geq 0, k \geq 0$.

(c) $f(x) = x \ln(x), x > 0$.

1.2 Consider a single-period inventory model where demand is a random variable with density f; that is, $P(\text{demand} \le x) = \int_0^x f(x')dx'$. Also, all stock-outs are lost sales, C is the unit cost of each item, p is the loss due to inability to fulfill orders (includes loss of revenue and customer goodwill), r is the selling price per unit and l is the salvage value of each unsold item at the end of the period. The problem is to determine the optimal order quantity Q that will maximize the expected net revenue for the season. The expected net revenue, denoted by $\Pi(Q)$, is given by the following equation:

$$\prod(Q) = r\mu + l\int_0^Q (Q-x)f(x)dx - (p+r)\int_Q^x (x-Q)f(x)dx - CQ$$

where μ is the expected demand as follows:

$$\mu = \int_0^x xf(x)dx$$

(a) Show that $\Pi(Q)$ is a concave function in $Q(\ge 0)$.

(b) Based on the result of (a) explain how you will find the optimal ordering policy.

(c) Compute the optimal policy for the following data:

$$C = \$2.50 \quad r = \$5.00 \quad l = 0 \quad p = \$2.50$$

$$f(x) = \begin{cases} \dfrac{1}{400} & 100 \le x \le 500 \\ 0 & \text{otherwise} \end{cases}$$

Hint: Use Leibniz rule for differentiation under the integral sign.

1.3 Identify the relations that constants a and b must satisfy so that the function

$$f(x,y) = x^a y^b \quad x, y \ge 0$$

is (a) convex and (b) concave for every $x, y \ge 0$. Generalize the result for the n-dimensional case:

$$f(x) = \prod_{i=1}^N x_i^{a_i} \quad x_i \ge 0$$

1.4 Let $g: \mathbb{R}^N \to \mathbb{R}$ be a concave function, and let function f be defined by $f(x) = \dfrac{1}{g(x)}$. Show that f is convex over $S = [x \mid g(x) > 0]$.

REFERENCES

1. Chemmangattuvalappil N, Task K, Banerjee I: An integer optimization algorithm for robust identification of non-linear gene regulatory networks. *BMC Syst Biol* 2012, **6**:119.
2. Dasika M, Gupta A, Maranas C, Altman R, Dunker A, Hunter L, Jung T, Klein T: A mixed integer linear programming (MILP) framework for inferring time delay in gene regulatory networks. Pacific Symposium on Biocomputing 2004, 2003: pp. 474–485.
3. Foteinou P, Yang E, Saharidis G, Ierapetritou M, Androulakis I: A mixed-integer optimization framework for the synthesis and analysis of regulatory networks. *J Global Optim* 2009, **43**(2–3):263–276.
4. Gupta A, Varner J, Maranas C: Large-scale inference of the transcriptional regulation of Bacillus subtilis. *Comput Chem Eng* 2005, **29**(3):565–576.
5. Lee W, Tzou W: Computational methods for discovering gene networks from expression data. *Briefings Bioinformat* 2009, **10**(4):408–423.
6. Thomas R, Paredes CJ, Mehrotra S, Hatzimanikatis V, Papoutsakis ET: A model-based optimization framework for the inference of regulatory interactions using time-course DNA microarray expression data. *BMC Bioinformat* 2007, **8**:228.
7. Xuan NV, Chetty M, Coppel R, Wangikar PP: Gene regulatory network modeling via global optimization of high-order dynamic Bayesian network. *BMC Bioinformat* 2012, **13**:131.
8. Bazaraa MS, Jarvis JJ, Sherali HD: *Linear programming and network flows*, 4th edn. Hoboken, N.J.: John Wiley & Sons; 2010.
9. Papin JA, Price ND, Edwards JS, Palsson BB: The genome-scale metabolic extreme pathway structure in Haemophilus influenzae shows significant network redundancy. *J Theor Biol* 2002, **215**(1):67–82.
10. Wiback SJ, Palsson BO: Extreme pathway analysis of human red blood cell metabolism. *Biophys J* 2002, **83**(2):808–818.
11. Schilling CH, Edwards JS, Letscher D, Palsson B: Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol Bioeng* 2000, **71**(4):286–306.
12. Schilling CH, Letscher D, Palsson BO: Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* 2000, **203**(3):229–248.
13. Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO: Metabolic pathways in the post-genome era. *Trends Biochem Sci* 2003, **28**(5):250–258.
14. Bazaraa MS, Sherali HD, Shetty CM: *Nonlinear programming: theory and algorithms*, 3rd edn. Hoboken, N.J.: Wiley-Interscience; 2006.