Chapter 8
Convex Optimization

8.1 ▪ Definition

A *convex optimization* problem (or just a *convex problem*) is a problem consisting of minimizing a convex function over a convex set. More explicitly, a convex problem is of the form

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in C, \end{array} \tag{8.1}$$

where C is a convex set and f is a convex function over C. Problem (8.1) is in a sense implicit, and we will often consider more explicit formulations of convex problems such as convex optimization problems in *functional form*, which are convex problems of the form

$$\begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, & i = 1,2,\ldots,m, \\ & h_j(\mathbf{x}) = 0, & j = 1,2,\ldots,p, \end{array} \tag{8.2}$$

where $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are convex functions and $h_1, h_2, \ldots, h_p : \mathbb{R}^n \to \mathbb{R}$ are affine functions. Note that the above problem does fit into the general form (8.1) of convex problems. Indeed, the objective function is convex and the feasible set is a convex set since it can be written as

$$C = \left(\bigcap_{i=1}^{m} \text{Lev}(g_i, 0) \right) \bigcap \left(\bigcap_{j=1}^{p} \{\mathbf{x} : h_j(\mathbf{x}) = 0\} \right),$$

which implies that C is a convex set as an intersection of level sets of convex functions, which are necessarily convex sets, and hyperplanes, which are also convex sets.

The following result shows a very important property of convex problems: all local minimum points are also *global* minimum points.

Theorem 8.1 (local=global in convex optimization). *Let $f : C \to \mathbb{R}$ be a convex function defined on the convex set C. Let $\mathbf{x}^* \in C$ be a local minimum of f over C. Then $\mathbf{x}^*$ is a global minimum of f over C.*

Proof. Since $\mathbf{x}^*$ is a local minimum of f over C, it follows that there exists $r > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for any $\mathbf{x} \in C$ satisfying $\mathbf{x} \in B[\mathbf{x}^*, r]$. Now let $\mathbf{y} \in C$ satisfy $\mathbf{y} \neq \mathbf{x}^*$. Our objective is to show that $f(\mathbf{y}) \geq f(\mathbf{x}^*)$. Let $\lambda \in (0, 1]$ be such that $\mathbf{x}^* + \lambda(\mathbf{y} - \mathbf{x}^*) \in B[\mathbf{x}^*, r]$.

An example of such λ is $\lambda = \frac{r}{\|\mathbf{y}-\mathbf{x}^*\|}$. Since $\mathbf{x}^* + \lambda(\mathbf{y}-\mathbf{x}^*) \in B[\mathbf{x}^*, r] \cap C$, it follows that $f(\mathbf{x}^*) \le f(\mathbf{x}^* + \lambda(\mathbf{y}-\mathbf{x}^*))$, and hence by Jensen's inequality

$$f(\mathbf{x}^*) \le f(\mathbf{x}^* + \lambda(\mathbf{y}-\mathbf{x}^*)) \le (1-\lambda)f(\mathbf{x}^*) + \lambda f(\mathbf{y}).$$

Thus, $\lambda f(\mathbf{x}^*) \le \lambda f(\mathbf{y})$, and hence the desired inequality $f(\mathbf{x}^*) \le f(\mathbf{y})$ follows. ☐

A slight modification of the above result shows that any local minimum of a *strictly* convex function over a convex set is a *strict* global minimum of the function over the set.

Theorem 8.2. *Let $f : C \to \mathbb{R}$ be a strictly convex function defined on the convex set C. Let $\mathbf{x}^* \in C$ be a local minimum of f over C. Then $\mathbf{x}^*$ is a strict global minimum of f over C.*

The optimal set of the convex problem (8.1) is the set of all its minimizers, that is, $\operatorname{argmin}\{f(\mathbf{x}) : \mathbf{x} \in C\}$. This definition of an *optimal set* is also valid for general problems. An important property of convex problems is that their optimal sets are also convex.

Theorem 8.3 (convexity of the optimal set in convex optimization). *Let $f : C \to \mathbb{R}$ be a convex function defined over the convex set $C \subseteq \mathbb{R}^n$. Then the set of optimal solutions of the problem*

$$\min\{f(\mathbf{x}) : \mathbf{x} \in C\}, \tag{8.3}$$

which we denote by X^*, is convex. If, in addition, f is strictly convex over C, then there exists at most one optimal solution of the problem (8.3).

Proof. If $X^* = \emptyset$, the result follows trivially. Suppose that $X^* \neq \emptyset$ and denote the optimal value by f^*. Let $\mathbf{x}, \mathbf{y} \in X^*$ and $\lambda \in [0,1]$. Then by Jensen's inequality $f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f^* + (1-\lambda)f^* = f^*$, and hence $\lambda\mathbf{x} + (1-\lambda)\mathbf{y}$ is also optimal, i.e., belongs to X^*, establishing the convexity of X^*. Suppose now that f is strictly convex and X^* is nonempty; to show that X^* is a singleton, suppose in contradiction that there exist $\mathbf{x}, \mathbf{y} \in X^*$ such that $\mathbf{x} \neq \mathbf{y}$. Then $\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y} \in C$, and by the strict convexity of f we have

$$f\left(\frac{1}{2}\mathbf{x} + \frac{1}{2}\mathbf{y}\right) < \frac{1}{2}f(\mathbf{x}) + \frac{1}{2}f(\mathbf{y}) = \frac{1}{2}f^* + \frac{1}{2}f^* = f^*,$$

which is a contradiction to the fact that f^* is the optimal value. ☐

Convex optimization problems consist of minimizing convex functions over convex sets, but we will also refer to problems consisting of maximizing concave functions over convex sets as *convex problems*. (Indeed, they can be recast as minimization problems of convex functions by multiplying the objective function by minus one.)

Example 8.4. The problem

$$\begin{array}{ll} \min & -2x_1 + x_2 \\ \text{s.t.} & x_1^2 + x_2^2 \le 3 \end{array}$$

is convex since the objective function is linear, and thus convex, and the single inequality constraint corresponds to the convex function $f(x_1, x_2) = x_1^2 + x_2^2 - 3$, which is a convex quadratic function. On the other hand, the problem

$$\begin{array}{ll} \min & x_1^2 - x_2 \\ \text{s.t.} & x_1^2 + x_2^2 = 3 \end{array}$$

is nonconvex. The objective function is convex, but the constraint is a nonlinear equality constraint and therefore nonconvex. Note that the feasible set is the boundary of the ball with center $(0,0)$ and radius $\sqrt{3}$. ∎

8.2 ▪ Examples

8.2.1 ▪ Linear Programming

A linear programming (LP) problem is an optimization problem consisting of minimizing a linear objective function subject to linear equalities and inequalities:

$$\text{(LP)} \quad \begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \\ & \mathbf{B}\mathbf{x} = \mathbf{g}. \end{array}$$

Here $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m, \mathbf{B} \in \mathbb{R}^{p\times n}, \mathbf{g} \in \mathbb{R}^p, \mathbf{c} \in \mathbb{R}^n$. This is of course a convex optimization problem since affine functions are convex. An interesting observation concerning LP problems is based on the fact that linear functions are both convex and concave. Consider the following LP problem:

$$\begin{array}{ll} \max & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

In the literature the latter formulation is many times called the "standard formulation." The above problem is on one hand a convex optimization problem as a maximization of a concave function over a convex set, but on the other hand, it is also a problem of maximizing a convex function over a convex set. We can therefore deduce by Theorem 7.42 that if the feasible set is nonempty and compact, then there exists at least one optimal solution which is an extreme point of the feasible set. By Theorem 6.34, this means that there exists at least one optimal solution which is a basic feasible solution. A more general result dropping the compactness assumption is called the "fundamental theorem of linear programming," and it states that if the problem has an optimal solution, then it necessarily has an optimal basic feasible solution.

Although the class of LP problems seems to be quite restrictive due to the linearity of all the involved functions, it encompasses a huge amount of applications and has a great impact on many fields in applied mathematics. Following is an example of a scheduling problem that can be recast as an LP problem.

Example 8.5. For a new position in a company, we need to schedule job interviews for n candidates numbered $1,2,\ldots,n$ in this order (candidate i is scheduled to be the ith interview). Assume that the starting time of candidate i must be in the interval $[\alpha_i, \beta_i]$, where $\alpha_i < \beta_i$. To assure that the problem is feasible we assume that $\alpha_i \leq \beta_j$ for any $j > i$. The objective is to formulate the problem of finding n starting times of interviews so that the minimal starting time difference between consecutive interviews is maximal.

Let t_i denote the starting time of interview i. The objective function is the minimal difference between consecutive starting times of interviews:

$$f(\mathbf{t}) = \min\{t_2 - t_1, t_3 - t_2, \ldots, t_n - t_{n-1}\},$$

and the corresponding optimization problem is

$$\begin{array}{ll} \max & \left[\min\{t_2 - t_1, t_3 - t_2, \ldots, t_n - t_{n-1}\}\right] \\ \text{s.t.} & \alpha_i \leq t_i \leq \beta_i, \quad i = 1,2,\ldots,n. \end{array}$$

Note that we did not incorporate the constraints that $t_i \leq t_{i+1}$ for $i = 1,2,\ldots,n-1$ since the feasibility condition will guarantee in any case that these constraints will be satisfied in an optimal solution. The problem is convex since it consists of maximizing a concave function subject to affine (and hence convex) constraints. To show that the objective function is indeed concave, note that by Theorem 7.25 the maximum of convex functions is a convex function. The corresponding result for concave functions (that can be obtained by simply looking at minus of the function) is that the minimum of concave functions is a concave function. Therefore, since the objective function is a minimum of linear (and hence concave) functions, it is a concave function. In order to formulate the problem as an LP problem, we reformulate the problem as

$$\begin{array}{ll} \max_{\mathbf{t},s} & s \\ \text{s.t.} & \min\{t_2-t_1, t_3-t_2, \ldots, t_n-t_{n-1}\} = s, \\ & \alpha_i \leq t_i \leq \beta_i, \quad i=1,2,\ldots,n. \end{array} \tag{8.4}$$

We now claim that problem (8.4) is equivalent to the corresponding problem with an inequality constraint instead of an equality:

$$\begin{array}{ll} \max_{\mathbf{t},s} & s \\ \text{s.t.} & \min\{t_2-t_1, t_3-t_2, \ldots, t_n-t_{n-1}\} \geq s, \\ & \alpha_i \leq t_i \leq \beta_i, \quad i=1,2,\ldots,n. \end{array} \tag{8.5}$$

By "equivalent" we mean that any optimal solution of (8.5) satisfies the inequality constraint as an equality constraint. Indeed, suppose in contradiction that there exists an optimal solution $(\mathbf{t}^*, s^*)$ of (8.5) that satisfies the inequality constraints strictly, meaning that $\min\{t_2^*-t_1^*, t_3^*-t_2^*, \ldots, t_n^*-t_{n-1}^*\} > s^*$. Then we can easily check that the solution $(\mathbf{t}^*, \tilde{s})$, where $\tilde{s} = \min\{t_2^*-t_1^*, t_3^*-t_2^*, \ldots, t_n^*-t_{n-1}^*\}$ is also feasible for (8.5) and has a larger objective function value, which is a contradiction to the optimality of $(\mathbf{t}^*, s^*)$. Finally, we can rewrite the inequality $\min\{t_2-t_1, t_3-t_2, \ldots, t_n-t_{n-1}\} \geq s$ as $t_{i+1}-t_i \geq s$ for any $i = 1,2,\ldots,n-1$, and we can therefore recast the problem as the following LP problem:

$$\begin{array}{ll} \max_{\mathbf{t},s} & s \\ \text{s.t.} & t_{i+1}-t_i \geq s, \quad i=1,2,\ldots,n-1, \\ & \alpha_i \leq t_i \leq \beta_i, \quad i=1,2,\ldots,n. \end{array} \qquad \blacksquare$$

8.2.2 ▪ Convex Quadratic Problems

Convex quadratic problems are problems consisting of minimizing a convex quadratic function subject to affine constraints. A general form of problems of this class can be written as

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{c}, \end{array}$$

where $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is positive semidefinite, $\mathbf{b} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}$, and $\mathbf{c} \in \mathbb{R}^m$. A well-known example of a convex quadratic problem arises in the area of linear classification and is described in detail next.

8.2.3 ▪ Classification via Linear Separators

Suppose that we are given two types of points in $\mathbb{R}^n$: type A and type B. The type A points are given by

$$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n$$

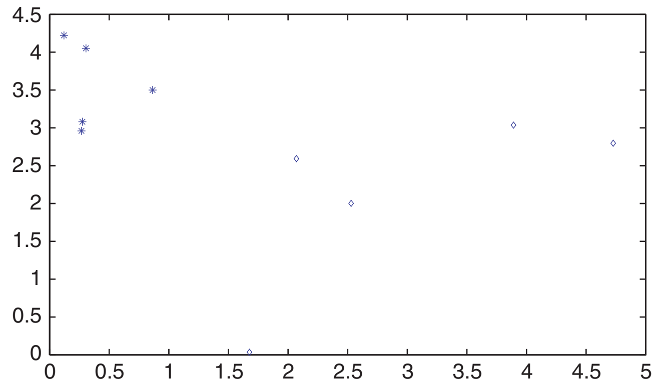

Figure 8.1. *Type A (asterisks) and type B (diamonds) points.*

and the type B points are given by

$$\mathbf{x}_{m+1}, \mathbf{x}_{m+2}, \ldots, \mathbf{x}_{m+p} \in \mathbb{R}^n.$$

For example, Figure 8.1 describes two sets of points in $\mathbb{R}^2$: the type A points are denoted by asterisks and the type B points are denoted by diamonds. The objective is to find a linear separator, which is a hyperplane of the form

$$H(\mathbf{w}, \beta) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{w}^T \mathbf{x} + \beta = 0\}$$

for which the type A and type B points are in its opposite sides:

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + \beta &< 0, \quad i = 1, 2, \ldots, m, \\ \mathbf{w}^T \mathbf{x}_i + \beta &> 0, \quad i = m+1, m+2, \ldots, m+p. \end{aligned}$$

Our underlying assumption is that the two sets of points are *linearly separable*, meaning that the above set of inequalities has a solution. The problem is not well-defined in the sense that there are many linear separators, and what we seek is in fact a separator that is in a sense farthest as possible from all the points. At this juncture we need to define the notion of the margin of the separator, which is the distance of the separator from the closest point, as illustrated in Figure 8.2. The separation problem will thus consist of finding the separator with the largest margin. To compute the margin, we need to have a formula for the distance between a point and a hyperplane. The next lemma provides such a formula, but its proof is postponed to Chapter 10 (see Lemma 10.12), where more general results will be derived.

Lemma 8.6. *Let $H(\mathbf{a}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} = b\}$, where $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Let $\mathbf{y} \in \mathbb{R}^n$. Then the distance between $\mathbf{y}$ and the set H is given by*

$$d(\mathbf{y}, H(\mathbf{a}, b)) = \frac{|\mathbf{a}^T \mathbf{y} - b|}{\|\mathbf{a}\|}.$$

We therefore conclude that the margin corresponding to a hyperplane $H(\mathbf{w}, -\beta)$ ($\mathbf{w} \neq \mathbf{0}$) is

$$\min_{i=1,2,\ldots,m+p} \frac{|\mathbf{w}^T \mathbf{x}_i + \beta|}{\|\mathbf{w}\|}.$$

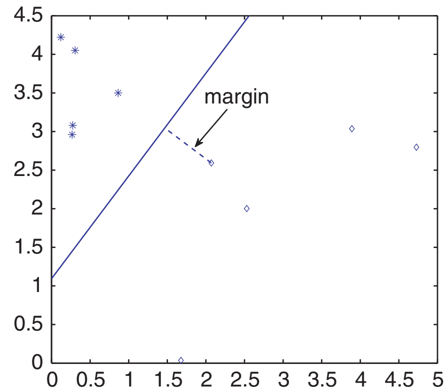

Figure 8.2. *The optimal linear seperator and its margin.*

So far, the problem that we consider is therefore

$$\begin{array}{ll} \max & \left\{\min_{i=1,2,\dots,m+p} \frac{|\mathbf{w}^T\mathbf{x}_i+\beta|}{\|\mathbf{w}\|}\right\} \\ \text{s.t.} & \mathbf{w}^T\mathbf{x}_i+\beta<0, \quad i=1,2,\dots,m, \\ & \mathbf{w}^T\mathbf{x}_i+\beta>0, \quad i=m+1,m+2,\dots,m+p. \end{array}$$

This is a rather bad formulation of the problem since it is not convex and cannot be easily handled. Our objective is to find a convex reformulation of the problem. For that, note that the problem has a degree of freedom in the sense that if $(\mathbf{w},\beta)$ is an optimal solution, then so is any nonzero multiplier of it, that is, $(\alpha\mathbf{w},\alpha\beta)$ for $\alpha\neq 0$. We can therefore decide that

$$\min_{i=1,2,\dots,m+p} |\mathbf{w}^T\mathbf{x}_i+\beta|=1,$$

and the problem can then be rewritten as

$$\begin{array}{ll} \max & \left\{\frac{1}{\|\mathbf{w}\|}\right\} \\ \text{s.t.} & \min_{i=1,2,\dots,m+p} |\mathbf{w}^T\mathbf{x}_i+\beta|=1, \\ & \mathbf{w}^T\mathbf{x}_i+\beta<0, \quad i=1,2,\dots,m, \\ & \mathbf{w}^T\mathbf{x}_i+\beta>0, \quad i=m+1,2,\dots,m+p. \end{array}$$

The combination of the first equality and the other inequality constraints implies that a valid reformulation is

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{s.t.} & \min_{i=1,2,\dots,m+p} |\mathbf{w}^T\mathbf{x}_i+\beta|=1, \\ & \mathbf{w}^T\mathbf{x}_i+\beta\leq -1, \quad i=1,2,\dots,m, \\ & \mathbf{w}^T\mathbf{x}_i+\beta\geq 1, \quad i=m+1,2,\dots,m+p, \end{array}$$

where we also used the fact that maximizing $\frac{1}{\|\mathbf{w}\|}$ is the same as minimizing $\|\mathbf{w}\|^2$ in the sense that the optimal set stays the same. Finally, we remove the problematic "min" equality constraint and obtain the following convex quadratic reformulation of the problem:

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{s.t.} & \mathbf{w}^T\mathbf{x}_i + \beta \leq -1, \quad i = 1,2,\ldots,m, \\ & \mathbf{w}^T\mathbf{x}_i + \beta \geq 1, \quad i = m+1, m+2, \ldots, m+p. \end{array} \tag{8.6}$$

The removal of the "min" constraint is valid since any feasible solution of problem (8.6) surely satisfies $\min_{i=1,2,\ldots,m+p} |\mathbf{w}^T\mathbf{x}_i + \beta| \geq 1$. If $(\mathbf{w}, \beta)$ is in addition optimal, then equality must be satisfied. Otherwise, if $\min_{i=1,2,\ldots,m+p} |\mathbf{w}^T\mathbf{x}_i + \beta| > 1$, then a better solution (i.e., with lower objective function value) will be $\frac{1}{\alpha}(\mathbf{w}, \beta)$, where $\alpha = \min_{i=1,2,\ldots,m+p} |\mathbf{w}^T\mathbf{x}_i + \beta|$.

8.2.4 ▪ Chebyshev Center of a Set of Points

Suppose that we are given m points $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ in $\mathbb{R}^n$. The objective is to find the center of the minimum radius closed ball containing all the points. This ball is called *the Chebyshev ball* and the corresponding center is *the Chebyshev center*. In mathematical terms, the problem can be written as (r denotes the radius and $\mathbf{x}$ is the center)

$$\begin{array}{ll} \min_{\mathbf{x},r} & r \\ \text{s.t.} & \mathbf{a}_i \in B[\mathbf{x}, r], \quad i = 1,2,\ldots,m. \end{array}$$

Of course, recalling that $B[\mathbf{x}, r] = \{\mathbf{y} : \|\mathbf{y} - \mathbf{x}\| \leq r\}$, it follows that the problem can be written as

$$\begin{array}{ll} \min_{\mathbf{x},r} & r \\ \text{s.t.} & \|\mathbf{x} - \mathbf{a}_i\| \leq r, \quad i = 1,2,\ldots,m. \end{array} \tag{8.7}$$

This is obviously a convex optimization problem since it consists of minimizing a linear (and hence convex) function subject to convex inequality constraints: the function $\|\mathbf{x} - \mathbf{a}_i\| - r$ is convex as a sum of a translation of the norm function and the linear function $-r$. An illustration of the Chebyshev center and ball is given in Figure 8.3.

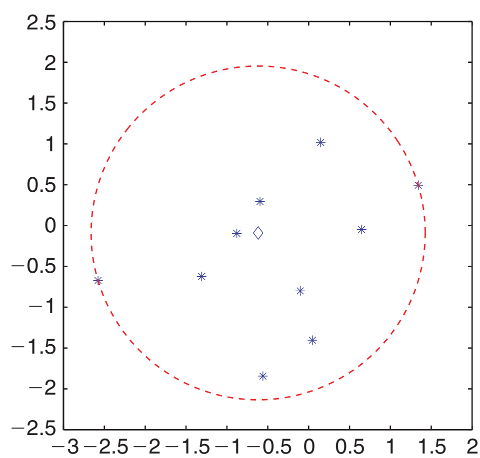

Figure 8.3. *The Chebyshev center (denoted by a diamond marker) of a set of* 10 *points (asterisks). The boundary of the Chebyshev ball is the dashed circle.*

8.2.5 ▪ Portfolio Selection

Suppose that an investor wishes to construct a portfolio out of n given assets numbered as $1,2,\dots,n$. Let $Y_j (j=1,2,\dots,n)$ be the random variable representing the return from asset j. We assume that the expected returns are known,

$$\mu_j = \mathbb{E}(Y_j), \quad j=1,2,\dots,n,$$

and that the covariances of all the pairs of variables are also known,

$$\sigma_{i,j} = COV(Y_i,Y_j), \quad i,j=1,2,\dots,n.$$

There are n decision variables $x_1,x_2,\dots,x_n$, where x_j denotes the proportion of budget invested in asset j. The decision variables are constrained to be nonnegative and sum up to 1: $\mathbf{x}\in\Delta_n$. The overall return is the random variable,

$$R=\sum_{j=1}^n x_j Y_j,$$

whose expectation and variance are given by

$$\mathbb{E}(R)=\boldsymbol{\mu}^T\mathbf{x},\ \mathbb{V}(R)=\mathbf{x}^T\mathbf{C}\mathbf{x},$$

where $\boldsymbol{\mu}=(\mu_1,\mu_2,\dots,\mu_n)^T$ and $\mathbf{C}$ is the *covariance matrix* whose elements are given by $C_{i,j}=\sigma_{i,j}$ for all $1\le i,j\le n$. It is important to note that the covariance matrix is always positive semidefinite. The variance of the portfolio, $\mathbf{x}^T\mathbf{C}\mathbf{x}$, is the *risk* of the suggested portfolio $\mathbf{x}$. There are several formulations of the portfolio optimization problem, which are all referred to as the "Markowitz model" in honor of Harry Markowitz, who first suggested this type of a model in 1952.

One formulation of the problem is to find a portfolio minimizing the risk under the constraint that a minimal return level is guaranteed:

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{C}\mathbf{x} \\ \text{s.t} & \boldsymbol{\mu}^T\mathbf{x}\ge\alpha, \\ & \mathbf{e}^T\mathbf{x}=1, \\ & \mathbf{x}\ge 0, \end{array} \tag{8.8}$$

where $\mathbf{e}$ is the vector of all ones and α is the minimal return value. Another option is to maximize the expected return subject to a bounded risk constraint:

$$\begin{array}{ll} \max & \boldsymbol{\mu}^T\mathbf{x} \\ \text{s.t} & \mathbf{x}^T\mathbf{C}\mathbf{x}\le\beta, \\ & \mathbf{e}^T\mathbf{x}=1, \\ & \mathbf{x}\ge 0, \end{array} \tag{8.9}$$

where β is the upper bound on the risk. Finally, a third option is to write an objective function which is a combination of the expected return and the risk:

$$\begin{array}{ll} \min & -\boldsymbol{\mu}^T\mathbf{x}+\gamma(\mathbf{x}^T\mathbf{C}\mathbf{x}) \\ \text{s.t} & \mathbf{e}^T\mathbf{x}=1, \\ & \mathbf{x}\ge 0, \end{array} \tag{8.10}$$

where $\gamma > 0$ is a penalty parameter. Each of the three models (8.8), (8.9), and (8.10) depends on a certain parameter (α, β, or γ) whose value dictates the tradeoff level between profit and risk. Determining the value of each of these parameters is not necessarily an easy task, and it also depends on the subjective preferences of the investors. The three models are all convex optimization problems since $\mathbf{x}^T\mathbf{C}\mathbf{x}$ is a convex function (its associated matrix $\mathbf{C}$ is positive semidefinite). The model (8.10) is a convex quadratic problem.

8.2.6 ▪ Convex QCQPs

A quadratically constrained quadratic problem, or QCQP for short, is a problem consisting of minimizing a quadratic function subject to quadratic inequalities and equalities:

$$\text{(QCQP)}\quad \begin{array}{lll} \min & \mathbf{x}^T\mathbf{A}_0\mathbf{x}+2\mathbf{b}_0^T\mathbf{x}+c_0 & \\ \text{s.t.} & \mathbf{x}^T\mathbf{A}_i\mathbf{x}+2\mathbf{b}_i^T\mathbf{x}+c_i \leq 0, & i=1,2,\ldots,m, \\ & \mathbf{x}^T\mathbf{A}_j\mathbf{x}+2\mathbf{b}_j^T\mathbf{x}+c_j = 0, & j=m+1,m+2,\ldots,m+p. \end{array}$$

Obviously, QCQPs are not necessarily convex problems, but when there are no equality constrainers ($p=0$) and all the matrices are positive semidefinite, $\mathbf{A}_i \succeq 0$ for $i = 0,1,\ldots,m$, the problem is convex and is therefore called a *convex QCQP*.

8.2.7 ▪ Hidden Convexity in Trust Region Subproblems

There are several situations in which a certain problem is not convex but nonetheless can be recast as a convex optimization problem. This situation is sometimes called "hidden convexity." Perhaps the most famous nonconvex problem possessing such a hidden convexity property is the *trust region subproblem*, consisting of minimizing a quadratic function (not necessarily convex) subject to an Euclidean norm constraint:

$$\text{(TRS)}\quad \min\{\mathbf{x}^T\mathbf{A}\mathbf{x}+2\mathbf{b}^T\mathbf{x}+c : \|\mathbf{x}\|^2 \leq 1\}.$$

Here $\mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$, and $\mathbf{A}$ is an $n \times n$ symmetric matrix which is not necessarily positive semidefinite. Since the objective function is (possibly) nonconvex, problem (TRS) is (possibly) nonconvex. This is an important class of problems arising, for example, as a subroutine in trust region methods, hence the name of this class of problems. We will now show how to transform (TRS) into a convex optimization problem. First, by the spectral decomposition theorem (Theorem 1.10), there exist an orthogonal matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{D} = \text{diag}(d_1, d_2, \ldots, d_n)$ such that $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, and hence (TRS) can be rewritten as

$$\min\{\mathbf{x}^T\mathbf{U}\mathbf{D}\mathbf{U}^T\mathbf{x}+2\mathbf{b}^T\mathbf{U}\mathbf{U}^T\mathbf{x}+c : \|\mathbf{U}^T\mathbf{x}\|^2 \leq 1\}, \tag{8.11}$$

where we used the fact that $\|\mathbf{U}^T\mathbf{x}\| = \|\mathbf{x}\|$. Making the linear change of variables $\mathbf{y} = \mathbf{U}^T\mathbf{x}$, it follows that (8.11) reduces to

$$\min\{\mathbf{y}^T\mathbf{D}\mathbf{y}+2\mathbf{b}^T\mathbf{U}\mathbf{y}+c : \|\mathbf{y}\|^2 \leq 1\}.$$

Denoting $\mathbf{f} = \mathbf{U}^T\mathbf{b}$, we obtain the following formulation of the problem:

$$\begin{array}{ll} \min & \sum_{i=1}^n d_i y_i^2 + 2\sum_{i=1}^n f_i y_i + c \\ \text{s.t.} & \sum_{i=1}^n y_i^2 \leq 1. \end{array} \tag{8.12}$$

The problem is still nonconvex since some of the d_is might be negative. At this point, we will use the following result stating that the signs of the optimal decision variables are actually known in advance.

Lemma 8.7. *Let $\mathbf{y}^*$ be an optimal solution of* (8.12). *Then $f_i y_i^* \leq 0$ for all $i = 1, 2, \ldots, n$.*

Proof. We will denote the objective function of problem (8.12) by $f(\mathbf{y}) \equiv \sum_{i=1}^n d_i y_i^2 + 2\sum_{i=1}^n f_i y_i + c$. Let $i \in \{1, 2, \ldots, n\}$. Define the vector $\tilde{\mathbf{y}}$ to be

$$\tilde{y}_j = \begin{cases} y_j^*, & j \neq i, \\ -y_i^*, & j = i. \end{cases}$$

Then obviously $\tilde{\mathbf{y}}$ is also a feasible solution of (8.12), and since $\mathbf{y}^*$ is an optimal solution of (8.12), it follows that

$$f(\mathbf{y}^*) \leq f(\tilde{\mathbf{y}}),$$

which is the same as

$$\sum_{i=1}^n d_i (y_i^*)^2 + 2\sum_{i=1}^n f_i y_i^* + c \leq \sum_{i=1}^n d_i (\tilde{y}_i)^2 + 2\sum_{i=1}^n f_i \tilde{y}_i + c.$$

Using the definition of $\tilde{\mathbf{y}}$, the above inequality reduces after much cancelation of terms to

$$2 f_i y_i^* \leq 2 f_i (-y_i^*),$$

which implies the desired inequality $f_i y_i^* \leq 0$. $\square$

As a direct result of Lemma 8.7 we have that for any optimal solution $\mathbf{y}^*$, the equality $\operatorname{sgn}(y_i^*) = -\operatorname{sgn}(f_i)$ holds when $f_i \neq 0$ and where the sgn function is defined to be

$$\operatorname{sgn}(x) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0. \end{cases}$$

When $f_i = 0$, we have the property that both $\mathbf{y}^*$ and $\tilde{\mathbf{y}}$ are optimal (see the proof of Lemma 8.7), and hence the sign of $\mathbf{y}^*$ can be chosen arbitrarily. As a consequence, we can make the change of variables $y_i = -\operatorname{sgn}(f_i)\sqrt{z_i} (z_i \geq 0)$, and problem (8.12) becomes

$$\begin{array}{ll} \min & \sum_{i=1}^n d_i z_i - 2\sum_{i=1}^n |f_i| \sqrt{z_i} + c \\ \text{s.t.} & \sum_{i=1}^n z_i \leq 1, \\ & z_1, z_2, \ldots, z_n \geq 0. \end{array} \tag{8.13}$$

Obviously this is a convex optimization problem since the constraints are linear and the objective function is a sum of linear terms and positive multipliers of the convex functions $-\sqrt{z_i}$. To conclude, we have shown that the nonconvex trust region subproblem (TRS) is equivalent to the convex optimization problem (8.13).

8.3 ▪ The Orthogonal Projection Operator

Given a nonempty closed convex set C, the *orthogonal projection* operator $P_C : \mathbb{R}^n \to C$ is defined by

$$P_C(\mathbf{x}) = \operatorname{argmin}\{\|\mathbf{y} - \mathbf{x}\|^2 : \mathbf{y} \in C\}. \tag{8.14}$$

The orthogonal projection operator with input $\mathbf{x}$ returns the vector in C that is closest to $\mathbf{x}$. Note that the orthogonal projection operator is defined as a solution of a convex optimization problem, specifically, a minimization of a convex quadratic function subject to a convex feasible set. The first orthogonal projection theorem states that the orthogonal projection operator is in fact well-defined, meaning that the optimization problem in (8.14) has a unique optimal solution.

Theorem 8.8 (first projection theorem). *Let C be a nonempty closed convex set. Then problem* (8.14) *has a unique optimal solution.*

Proof. Since the objective function in (8.14) is a quadratic function with a positive definite matrix, it follows by Lemma 2.42 that the objective function is coercive and hence, by Theorem 2.32, that the problem has at least one optimal solution. In addition, since the objective function is strictly convex (again, since the objective function is quadratic with positive definite matrix), it follows by Theorem 8.3 that there exists only one optimal solution. □

The distance function was already defined in Example 7.29 as

$$d(\mathbf{x},C)=\min_{\mathbf{y}\in C}\|\mathbf{x}-\mathbf{y}\|.$$

Evidently, the distance function, in the case where C is a nonempty closed and convex set, can also be written in terms of the orthogonal projection as follows:

$$d(\mathbf{x},C)=\|\mathbf{x}-P_C(\mathbf{x})\|.$$

Computing the orthogonal projection operator might be a difficult task, but there are some examples of simple sets on which the orthogonal projection can be easily computed.

Example 8.9 (projection on the nonnegative orthant). Let $C=\mathbb{R}^n_+$. To compute the orthogonal projection of $\mathbf{x}\in\mathbb{R}^n$ onto $\mathbb{R}^n_+$, we need to solve the convex optimization problem

$$\begin{array}{ll}\min & \sum_{i=1}^n (y_i-x_i)^2\\ \text{s.t.} & y_1,y_2,\ldots,y_n\geq 0.\end{array} \tag{8.15}$$

Since this problem is separable, meaning that the objective function is a sum of functions of each of the variables, and the constraints are separable in the sense that each of the variables has its own constraint, it follows that the ith component of the optimal solution $\mathbf{y}^*$ of problem (8.15) is the optimal solution of the univariate problem

$$\min\{(y_i-x_i)^2 : y_i\geq 0\},$$

which is given by $y_i^*=[x_i]_+$, where for a real number $\alpha\in\mathbb{R}$, $[\alpha]_+$ is the *nonnegative part* of α:

$$[\alpha]_+=\begin{cases}\alpha, & \alpha\geq 0,\\ 0, & \alpha<0.\end{cases}$$

We will extend the definition of the nonnegative part to vectors, and the nonnegative part of a vector $\mathbf{v}\in\mathbb{R}^n$ is defined by

$$[\mathbf{v}]_+=([v_1]_+,[v_2]_+,\ldots,[v_n]_+)^T.$$

To summarize, the orthogonal projection operator onto $\mathbb{R}^n_+$ is given by

$$P_{\mathbb{R}^n_+}(\mathbf{x})=[\mathbf{x}]_+. \quad \blacksquare$$

Example 8.10 (projection on boxes). A box is a subset of $\mathbb{R}^n$ of the form

$$B=[\ell_1,u_1]\times[\ell_2,u_2]\times\cdots\times[\ell_n,u_n]=\{\mathbf{x}\in\mathbb{R}^n:\ell_i\leq x_i\leq u_i\},$$

where $\ell_i \leq u_i$ for all $i = 1, 2, \ldots, n$. We will also allow some of the u_i's to be equal to ∞ and some of the ℓ_i's to be equal to $-\infty$; in these cases we will assume that ∞ or $-\infty$ are not actually contained in the intervals. A similar separability argument as the one used in the previous example, shows that the orthogonal projection is given by

$$\mathbf{y} = P_B(\mathbf{x}),$$

where

$$y_i = \begin{cases} u_i, & x_i \geq u_i, \\ x_i, & \ell_i < x_i < u_i, \\ \ell_i, & x_i \leq \ell_i, \end{cases}$$

for any $i = 1, 2, \ldots, n$. ∎

Example 8.11 (projection onto balls). Let $C = B[\mathbf{0}, r] = \{\mathbf{y} : \|\mathbf{y}\| \leq r\}$. The optimization problem associated with the computation of $P_C(\mathbf{x})$ is given by

$$\min_{\mathbf{y}} \{\|\mathbf{y} - \mathbf{x}\|^2 : \|\mathbf{y}\|^2 \leq r^2\}. \tag{8.16}$$

If $\|\mathbf{x}\| \leq r$, then obviously $\mathbf{y} = \mathbf{x}$ is the optimal solution of (8.16) since it corresponds to the optimal value 0. When $\|\mathbf{x}\| > r$, then the optimal solution of (8.16) must belong to the boundary of the ball since otherwise, by Theorem 2.6, it would be a stationary point of the objective function, that is, $2(\mathbf{y} - \mathbf{x}) = \mathbf{0}$, and hence $\mathbf{y} = \mathbf{x}$, which is impossible since $\mathbf{x} \notin C$. We thus conclude that the problem in this case is equivalent to

$$\min_{\mathbf{y}} \{\|\mathbf{y} - \mathbf{x}\|^2 : \|\mathbf{y}\|^2 = r^2\},$$

which can be equivalently written as

$$\min_{\mathbf{y}} \{-2\mathbf{x}^T\mathbf{y} + r^2 + \|\mathbf{x}\|^2 : \|\mathbf{y}\|^2 = r^2\},$$

The optimal solution of the above problem is the same as the optimal solution of

$$\min_{\mathbf{y}} \{-2\mathbf{x}^T\mathbf{y} : \|\mathbf{y}\|^2 = r^2\}.$$

By the Cauchy–Schwarz inequality, the objective function can be lower bounded by

$$-2\mathbf{x}^T\mathbf{y} \geq -2\|\mathbf{x}\|\|\mathbf{y}\| = -2r\|\mathbf{x}\|,$$

and on the other hand, this lower bound is attained at $\mathbf{y} = r\frac{\mathbf{x}}{\|\mathbf{x}\|}$, and hence the orthogonal projection is given by

$$P_{B[\mathbf{0},r]} = \begin{cases} \mathbf{x}, & \|\mathbf{x}\| \leq r, \\ r\frac{\mathbf{x}}{\|\mathbf{x}\|}, & \|\mathbf{x}\| > r. \end{cases} \quad ∎$$

8.4 ▪ CVX

CVX is a MATLAB-based modeling system for convex optimization. It was created by Michael Grant and Stephen Boyd [19]. This MATLAB package is in fact an interface to other convex optimization solvers such as SeDuMi and SDPT3. We will explore here some of the basic features of the software, but a more comprehensive and complete guide can be found at the CVX website (`CVXr.com`). The basic structure of a CVX program is as follows:

```
cvx_begin
{variables declaration}
minimize({objective function}) or maximize({objective function})
subject to
{constraints}
cvx_end
```

Variables Declaration

The variables are declared via the command variable or variables. Thus, for example,

```
variable x(4);
variable z;
variable Y(2,3);
```

declares three variables:

- $\mathbf{x}$, a column vector of length 4,
- z, a scalar,
- $\mathbf{Y}$, a 2×3 matrix.

The same declaration can be written as

```
variables x(4) z Y(2,3);
```

Atoms

CVX accepts only convex functions as objective and constraint functions. There are several basic convex functions, called "atoms," which are embedded in CVX. Some of these atoms are given in the following table.

function	meaning	attributes
norm(x,p)	$\sqrt[p]{\sum_{i=1}^n \lvert x_i\rvert^p} (p \geq 1)$	convex
square(x)	x^2	convex
sum_square(x)	$\sum_{i=1}^n x_i^2$	convex
square_pos(x)	$[x]_+^2$	convex, nondecreasing
sqrt(x)	$\sqrt{x}$	concave, nondecreasing
inv_pos(x)	$\frac{1}{x}(x > 0)$	convex, nonincreasing
max(x)	$\max\{x_1, x_2, \ldots, x_n\}$	convex, nondecreasing
quad_over_lin(x,y)	$\frac{\|\mathbf{x}\|^2}{y}\ (y > 0)$	convex
quad_form(x,P)	$\mathbf{x}^T\mathbf{P}\mathbf{x}\ (\mathbf{P} \succeq 0)$	convex

In addition, CVX is aware that the function x^p for an even integer p is a convex function and that affine functions are both convex and concave.

Operations Preserving Convexity

Atoms can be incorporated by several operations which preserve convexity:

- addition,
- multiplication by a nonnegative scalar,

- composition of a nondecreasing convex function with a convex function,
- composition of a convex function with an affine transformation.

CVX is also aware that minus a convex function is a concave function. The constraints that CVX is willing to accept are inequalities of the forms

```
f(x)<=g(x)
g(x)>=f(x)
```

where f is convex and g is concave. Equality constraints must be affine, and the syntax is (h and s are affine functions)

```
h(x)==s(x)
```

Note that the equality must be written in the format ==. Otherwise, it will be interpreted as a substitution operation.

Example 8.12. Suppose that we wish to solve the least squares problem

$$\min ||\mathbf{A}\mathbf{x}-\mathbf{b}||^2,$$

where

$$\mathbf{A}=\begin{pmatrix}1 & 2\\ 3 & 4\\ 5 & 6\end{pmatrix}, \qquad \mathbf{b}=\begin{pmatrix}7\\ 8\\ 9\end{pmatrix}.$$

We can find the solution of this least squares problem by the MATLAB commands

```
>> A=[1,2;3,4;5,6];b=[7;8;9];
>> x=(A'*A)\(A'*b)
x =
   -6.0000
    6.5000
```

To solve this problem via CVX, we can use the function `sum_square`:

```
cvx_begin
variable x(2)
minimize(sum_square(A*x-b))
cvx_end
```

The obtained solution is as expected:

```
>> x
x =
   -6.0000
    6.5000
```

We can also solve the problem by noting that

$$||\mathbf{A}\mathbf{x}-\mathbf{b}||^2=\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}-2\mathbf{b}^T\mathbf{A}\mathbf{x}+||\mathbf{b}||^2$$

and writing the following commands:

```
cvx_begin
variable x(2)
minimize(quad_form(x,A'*A)-2*b'*A*x)
cvx_end
```

However, the following program is *wrong* and CVX will not accept it:

```
cvx_begin
variable x(2)
minimize(norm(A*x-b)^2)
cvx_end
```

The reason is that the objective function is written as a composition of the square function which is *not* nondecreasing with the function `norm(Ax-b)`. Of course, we know that the image of $||\mathbf{A}\mathbf{x}-\mathbf{b}||$ consists only of nonnegative values and that the square function is nondecreasing over that domain. However, CVX is not aware of that. If we insist on making such a decomposition we can use the function `square_pos` – the scalar function $\varphi(x)=\max\{x,0\}^2$, which is convex and nondecreasing, and write the legitimate CVX program:

```
cvx_begin
variable x(2)
minimize(square_pos(norm(A*x-b)))
cvx_end
```

It is also worth mentioning that since the problem of minimizing the norm is equivalent to the problem of minimizing the squared norm in the sense that both problems have the same optimal solution, the following CVX program will also find the optimal solution, but the optimal value will be the square root of the optimal value of the original problem:

```
cvx_begin
variable x(2)
minimize(norm(A*x-b))
cvx_end
```

∎

Example 8.13. Suppose that we wish to write a CVX code that solves the convex optimization problem

$$\begin{array}{ll} \min & \sqrt{x_1^2+x_2^2+1}+2\max\{x_1,x_2,0\} \\ \text{s.t.} & |x_1|+|x_2|+\frac{x_1^2}{x_2}\leq 5 \\ & \frac{1}{x_2}+x_1^4\leq 10 \\ & x_2\geq 1 \\ & x_1\geq 0. \end{array} \tag{8.17}$$

In order to write the above problem in CVX, it is important to understand the reason why $\sqrt{x_1^2+x_2^2+1}$ is convex since writing `sqrt(x(1)^2+x(2)^2+1)` in CVX will result in an error message. Since the expression is written as a composition of an increasing concave function with a convex function, in general it does not result in a convex function. A valid reason why $\sqrt{x_1^2+x_2^2+1}$ is convex is that it can be rewritten as $||(x_1,x_2,1)^T||$. That is, it is a composition of the norm function with an affine transformation. Correspondingly, the correct syntax in CVX will be `norm([x;1])`. Overall, a CVX program that solves (8.17) is

```
cvx_begin
variable x(2)
minimize(norm([x;1])+2*max(max(x(1),x(2)),0))
subject to
norm(x,1)+quad_over_lin(x(1),x(2))<=5
inv_pos(x(2))+x(1)^4<=10
x(2)>=1
x(1)>=0
cvx_end
```

∎

Example 8.14. Suppose that we wish to find the Chebyshev center of the 5 points

$$(-1,3), \quad (-3,10), \quad (-1,0), \quad (5,0), \quad (-1,-5).$$

Recall that the problem of finding the Chebyshev center of a set of points $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ is given by (see Section 8.2.4)

$$\begin{array}{ll} \min_{\mathbf{x},r} & r \\ \text{s.t.} & \|\mathbf{x}-\mathbf{a}_i\| \le r, \quad i = 1,2,\ldots,m, \end{array}$$

and thus the following code will solve the problem:

```
A=[-1,-3,-1,5,-1;3,10,0,0,-5];
cvx_begin
variables x(2) r
minimize(r)
subject to
for i=1:5
   norm(x-A(:,i))<=r
end
cvx_end
```

This results in the optimal solution

```
>> x
x =
   -2.0002
    2.5000
>> r
r =
    7.5664
```

To plot the 5 points along with the Chebyshev circle and center we can write

```
plot(A(1,:),A(2,:),'*')
hold on
plot(x(1),x(2),'d')
t=0:0.001:2*pi;
xx=x(1)+r*cos(t);
```

```
yy=x(2)+r*sin(t);
plot(xx,yy)
axis equal
axis([-11,7,-6,11])
hold off
```

The result can be seen in Figure 8.4. ∎

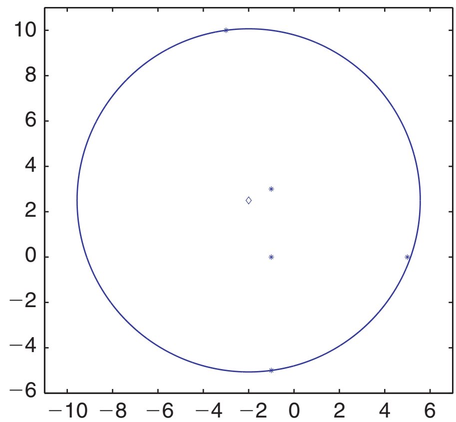

Figure 8.4. *The Chebyshev center (diamond marker) of* 5 *points (in asterisks).*

Example 8.15 (robust regression). Suppose that we are given 21 points in $\mathbb{R}^2$ generated by the MATLAB commands

```
randn('seed',314);
x=linspace(-5,5,20)';
y=2*x+1+randn(20,1);
x=[x;5];
y=[y;-20];
plot(x,y,'*')
hold on
```

The resulting plot can be seen in Figure 8.5. Note that the point $(5,-20)$ is an outlier; it is far away from all the other points and does not seem to fit into the almost-line structure of the other points. The least squares line, also called *the regression line*, can be found by the commands (see also Chapter 3)

```
A=[x,ones(21,1)];
b=y;
u=A\b;
alpha=u(1);beta=u(2);
plot([-6,6],alpha*[-6,6]+beta);
hold off
```

resulting in the line plotted in Figure 8.6. As can be clearly seen in Figure 8.6, the least squares line is very much affected by the single outlier point, which is a known drawback of the least squares approach. Another option is to replace the l_2-based objective function $\|\mathbf{Ax}-\mathbf{b}\|_2^2$ with an l_1-based objective function; that is, we can consider the optimization problem

$$\min \|\mathbf{Ax}-\mathbf{b}\|_1.$$

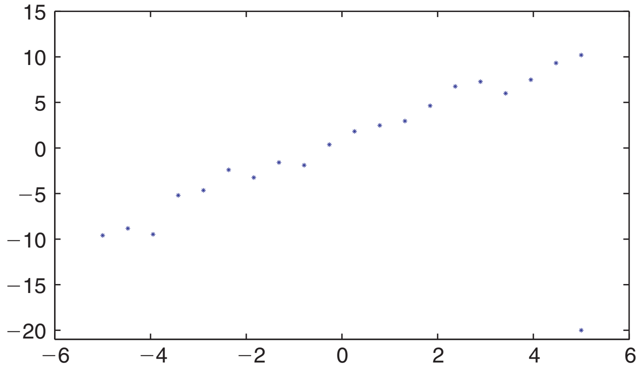

Figure 8.5. 21 *points in the plane. The point* $(5,-20)$ *is an outlier.*

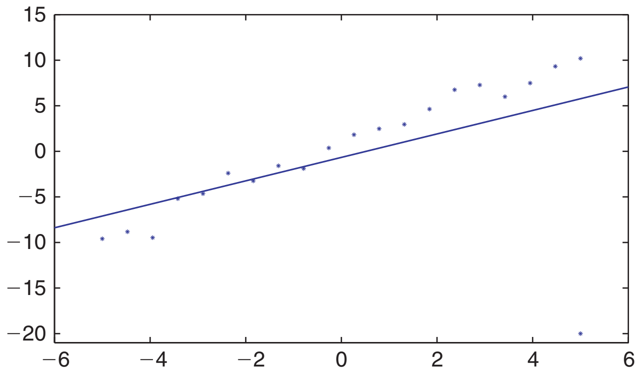

Figure 8.6. 21 *points in the plane along with their least squares (regression) line.*

This approach has the advantage that it is less sensitive to outliers since outliers are not as severely penalized as they are penalized in the least squares objective function. More specifically, in the least squares objective function, the distances to the line are squared, while in the l_1-based function they are not. To find the line using CVX, we can run the commands

```
plot(x,y,'*')
hold on
cvx_begin
variable u_l1(2)
minimize(norm(A*u_l1-b,1))
cvx_end
alpha_l1=u_l1(1);
beta_l1=u_l1(2);
plot([-6,6],alpha_l1*[-6,6]+beta_l1);
axis([-6,6,-21,15])
hold off
```

and the corresponding plot is given in Figure 8.7. Note that the resulting line is insensitive to the outlier. This is why this line is also called *the robust regression line*. ∎

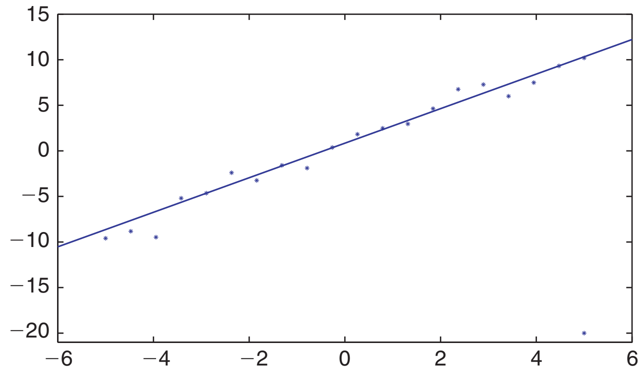

Figure 8.7. 21 *points in the plane along with their robust regression line.*

Example 8.16 (solution of a trust region subproblem). Consider the trust region subproblem (see Section 8.2.7)

$$\begin{array}{ll} \min & x_1^2 + x_2^2 + 3x_3^2 + 4x_1x_2 + 6x_1x_3 + 8x_2x_3 + x_1 + 2x_2 - x_3 \\ \text{s.t.} & x_1^2 + x_2^2 + x_3^2 \le 1, \end{array}$$

which is the same as

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} \\ \text{s.t.} & \|\mathbf{x}\|^2 \le 1, \end{array}$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 3 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} \frac{1}{2} \\ 1 \\ -\frac{1}{2} \end{pmatrix}.$$

The problem is nonconvex since the matrix $\mathbf{A}$ is not positive definite:

```
>> A=[1,2,3;2,1,4;3,4,3];
>> b=[0.5;1;-0.5];
>> eig(A)

ans =

   -2.1683
   -0.8093
    7.9777
```

It is therefore not possible to solve the problem directly using CVX. Instead, we will use the technique described in Section 8.2.7 to convert the problem into a convex problem, and then we will be able to solve the transformed problem via CVX. We begin by computing the spectral decomposition of $\mathbf{A}$,

```
[U,D]=eig(A);
```

and then compute the vectors $\mathbf{d}$ and $\mathbf{f}$ in the convex reformulation of the problem:

```
f=U'*b;
d=diag(D);
```

We can now use CVX to solve the equivalent problem (8.13):

```
cvx_begin
variable z(3)
minimize(d'*z-2*abs(f)'*sqrt(z))
subject to
sum(z)<=1
z>=0
cvx_end
```

The optimal solution is then computed by $y_i = -\text{sgn}(f_i)\sqrt{z_i}$ and then $\mathbf{x} = \mathbf{U}\mathbf{y}$:

```
>> y=-sign(f).*sqrt(z);
>> x=U*y
x =

   -0.2300
   -0.7259
    0.6482
```

$\blacksquare$

Exercises

8.1. Consider the problem

$$(\mathrm{P}) \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & g(\mathbf{x}) \leq 0 \\ & \mathbf{x} \in X, \end{array}$$

where f and g are convex functions over $\mathbb{R}^n$ and $X \subseteq \mathbb{R}^n$ is a convex set. Suppose that $\mathbf{x}^*$ is an optimal solution of (P) that satisfies $g(\mathbf{x}^*) < 0$. Show that $\mathbf{x}^*$ is also an optimal solution of the problem

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in X. \end{array}$$

8.2. Let $C = B[\mathbf{x}_0, r]$, where $\mathbf{x}_0 \in \mathbb{R}^n$ and $r > 0$ are given. Find a formula for the orthogonal projection operator P_C.

8.3. Let f be a strictly convex function over $\mathbb{R}^m$ and let g be a convex function over $\mathbb{R}^n$. Define the function

$$h(\mathbf{x}) = f(\mathbf{A}\mathbf{x}) + g(\mathbf{x}),$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$. Assume that $\mathbf{x}^*$ and $\mathbf{y}^*$ are optimal solutions of the unconstrained problem of minimizing h. Show that $\mathbf{A}\mathbf{x}^* = \mathbf{A}\mathbf{y}^*$.

8.4. For each of the following optimization problems (a) show that it is convex, (b) write a CVX code that solves it, and (c) write down the optimal solution (by running CVX).

(i)

$$\begin{array}{ll} \min & x_1^2 + 2x_1x_2 + 2x_2^2 + x_3^2 + 3x_1 - 4x_2 \\ \text{s.t.} & \sqrt{2x_1^2 + x_1x_2 + 4x_2^2 + 4} + \frac{(x_1 - x_2 + x_3 + 1)^2}{x_1 + x_2} \leq 6 \\ & x_1, x_2, x_3 \geq 1. \end{array}$$

(ii)

$$\begin{array}{ll} \max & x_1+x_2+x_3+x_4 \\ \text{s.t.} & (x_1-x_2)^2+(x_3+2x_4)^4 \leq 5 \\ & x_1+2x_2+3x_3+4x_4 \leq 6 \\ & x_1,x_2,x_3,x_4 \geq 0. \end{array}$$

(iii)

$$\begin{array}{ll} \min & 5x_1^2+4x_2^2+7x_3^2+4x_1x_2+2x_2x_3+|x_1-x_2| \\ \text{s.t.} & \frac{x_1^2+x_2^2}{x_3}+(x_1^2+x_2^2+1)^4 \leq 10 \\ & x_3 \geq 10. \end{array}$$

(iv)

$$\begin{array}{ll} \min & \sqrt{x_1^2+x_2^2+2x_1+5}+x_1^2+2x_1x_2+x_2^2+2x_1+3x_2 \\ \text{s.t.} & \frac{x_1^2}{x_1+x_2}+\left(\frac{x_1^2}{x_2}+1\right)^8 \leq 100 \\ & x_1+x_2 \geq 4 \\ & x_2 \geq 1. \end{array}$$

(v)

$$\begin{array}{ll} \min & |2x_1+3x_2+x_3|+x_1^2+x_2^2+x_3^2+\sqrt{2x_1^2+4x_1x_2+7x_2^2+10x_2+6} \\ \text{s.t.} & \frac{x_1^2+1}{x_2}+2x_1^2+5x_2^2+10x_3^2+4x_1x_2+2x_1x_3+2x_2x_3 \leq 7 \\ & x_1 \geq 0 \\ & x_2 \geq 1. \end{array}$$

For this problem also show that the expression inside the square root is always nonnegative, i.e., $2x_1^2+4x_1x_2+7x_2^2+10x_2+6 \geq 0$ for all x_1,x_2.

(vi)

$$\begin{array}{ll} \min & \frac{1}{2x_2+3x_3}+5x_1^2+4x_2^2+7x_3^2+\frac{x_1^2+x_1+1}{x_2+x_3} \\ \text{s.t.} & \max\left\{x_1+x_2,x_3^2\right\}+(x_1^2+4x_1x_2+5x_2^2+1)^2 \leq 10 \\ & x_1,x_2,x_3 \geq 0.1. \end{array}$$

(vii)

$$\begin{array}{ll} \min & \sqrt{2x_1^2+3x_2^2+x_3^2+4x_1x_2+7}+(x_1^2+x_2^2+x_3^2+1)^2 \\ \text{s.t.} & \frac{(x_1+x_2)^2}{x_3+1}+x_1^8 \leq 7 \\ & x_1^2+x_2^2+4x_3^2+2x_1x_2+2x_1x_3+2x_2x_3 \leq 10 \\ & x_1,x_2,x_3 \geq 0. \end{array}$$

(viii)

$$\begin{array}{ll} \min & \frac{x_1^4+2x_1^2x_2^2+x_2^4}{x_1^2+2x_1x_2+x_2^2}+\sqrt{x_3^2+1} \\ \text{s.t.} & x_1^2+x_2^2+2x_3^2+2x_1x_2+2x_1x_3+2x_2x_3 \leq 100 \\ & x_1+x_2+x_3=2 \\ & x_1+x_2 \geq 1. \end{array}$$

(ix)

$$\begin{array}{ll} \min & \frac{x_1^4}{x_2^2}+\frac{x_2^4}{x_1^2}+2x_1x_2+|x_1+5|+|x_2+5|+|x_3+5| \\ \text{s.t.} & \left(\left(x_1^2+x_2^2+x_3^2+1\right)^2+1\right)^2+x_1^4+x_2^4+x_3^4 \leq 200 \\ & \max\left\{x_1^2+4x_1x_2+9x_2^2,x_1,x_2\right\} \leq 40 \\ & x_1 \geq 1 \\ & x_2 \geq 1. \end{array}$$

(x)

$$\begin{array}{ll} \min & (x_1+x_2+x_3)^8+x_1^2+x_2^2+3x_3^2+2x_1x_2+2x_2x_3+2x_1x_3 \\ \text{s.t.} & (|x_1-2x_2|+1)^4+\frac{1}{x_3}\leq 10, \\ & 2x_1+2x_2+x_3\leq 1, \\ & 0\leq x_3\leq 1. \end{array}$$

8.5. Suppose that we are given 40 points in the plane. Each of these points belongs to one of two classes. Specifically, there are 19 points of class 1 and 21 points of class 2. The points are generated and plotted by the MATLAB commands

```
rand('seed',314);
x=rand(40,1);
y=rand(40,1);
class=[2*x<y+0.5]+1;
A1=[x(find(class==1)),y(find(class==1))];
A2=[x(find(class==2)),y(find(class==2))];
plot(A1(:,1),A1(:,2),'*','MarkerSize',6)
hold on
plot(A2(:,1),A2(:,2),'d','MarkerSize',6)
hold off
```

The plot of the points is given in Figure 8.8. Note that the rows of $\mathbf{A}_1\in\mathbb{R}^{19\times 2}$ are the 19 points of class 1 and the rows of $\mathbf{A}_2\in\mathbb{R}^{21\times 2}$ are the 21 points of class 2. Write a CVX-based code for finding the maximum-margin line separating the two classes of points.

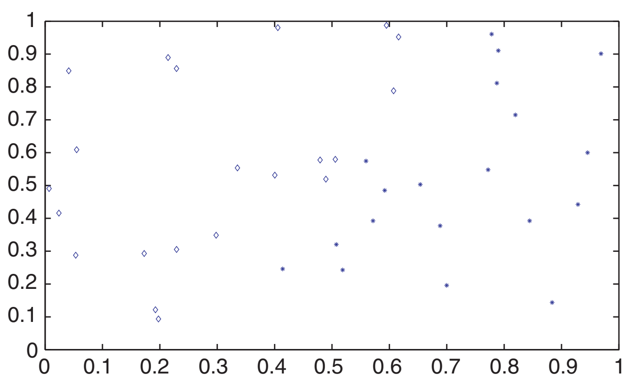

Figure 8.8. 40 *points of two classes: class* 1 *points are denoted by asterisks, and class* 2 *points are denoted by diamonds.*