# 5. Convex Optimization and Constraints

Prof. Gordon Ritter

*October 1, 2017*

What kinds of constraints can be handled easily in portfolio optimization? Is there always a unique global optimum? In order to progress, we need a general framework to address these kinds of questions. That is the framework of convex optimization. Most of what we'll say here can be found in Boyd and Vandenberghe (2009), but that's 700 pages! We're going to try to condense it down a bit.

The line segment between $x_1$ and $x_2$ in a vector space $V$ is all points $x = \theta x_1 + (1-\theta)x_2$ with $0 \le \theta \le 1$. A set $C$ is *convex* if it contains the line segment between any two points in the set, i.e.

$$x_1, x_2 \in C \;\Rightarrow\; \theta x_1 + (1-\theta)x_2 \in C \;\forall\, \theta \in [0,1]$$

A *convex combination* of $x_1, \dots, x_k$ is any point $x$ of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \cdots + \theta_k = 1, \theta_i \ge 0$.

The *convex hull* of a set $S$, denoted $\mathrm{conv}(S)$, is the set of all convex combinations of points in $S$. The convex hull is also the smallest convex set that contains $S$, or the intersection of all convex sets containing $S$.

Figure 5.1. When $S$ is a bounded subset of the plane, the convex hull may be visualized as the shape formed by a rubber band stretched around $S$.

*Definition* 5.1. A *norm* on a vector space $V$ is a function $\|\cdot\|$ from $V \to \mathbb{R}$ that satisfies: $\|x\| \ge 0$ with equality iff $x = 0$, $\|tx\| = |t|\,\|x\|$ for $t \in \mathbb{R}$, and

$$\|x+y\| \le \|x\| + \|y\|. \tag{5.1}$$

Any norm is a convex function, as follows from the definition. Eq. (5.1) is known as the *triangle inequality.* The most familiar examples of norms are $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$ for $p > 0$. For $p = 2$ the associated distance $d(x, y) = \|x - y\|_2$ is the usual Euclidean distance, but any norm defines a distance function, or *metric*, $d(x, y) = \|x - y\|$. An *open ball* is defined to be a set of the form

$$B(x; \epsilon) := \{y : d(x, y) < \epsilon\} \quad \text{where} \quad \epsilon > 0.$$

The associated closed ball is denoted $\overline{B}(x; \epsilon)$.

To establish convexity of a set $C$, one can always try to apply the definition directly. A faster way might be to show that $C$ is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, ellipsoids, etc.) by operations that preserve convexity, such as

- intersection
- affine functions
- perspective function
- linear-fractional functions

Some more examples of convex sets are provided by familiar sets from matrix theory. $S^n$, the space of $n \times n$ symmetric, real matrices, is convex. $S^n_+$ denotes the subspace of $S^n$ with non-negative eigenvalues, or equivalently, the subspace

$$\{A \in S^n : v^T A v \geq 0 \ \forall v\}.$$

$S^n_{++}$ is defined similarly, but for *strictly positive* matrices, i.e. those satisfying $v^T A v > 0$ for any nonzero $v$. We can easily show that $S^n_{++}$ is also convex:

$$v^T[\theta A + (1 - \theta)B]v = \theta v^T A v + (1 - \theta) v^T B v \geq 0 \text{ if } A, B \in S^n_+$$

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine ($f(x) = Ax + b$ with $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$) then the image (or inverse image) of a convex set under $f$ is convex. Examples include scaling, translation, projection and the solution set of a linear matrix inequality

$$\{x \mid x_1 A_1 + \cdots + x_m A_m \preceq B\} \text{ with } A_i, B \in S^p.$$

The perspective function $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is defined by

$$P(x, t) = x/t, \quad \text{dom} P = \{(x, t) \mid t > 0\}$$

It is a fact that images and inverse images of convex sets under perspective are convex. A linear-fractional function $f : \mathbb{R}^n \to \mathbb{R}^m$ is defined as

$$f(x) = (Ax + b)/(c^T x + d),$$

with $\mathrm{dom}(f) = \{x \mid c^T x + d > 0\}$. Images and inverse images of convex sets under linear-fractional functions are convex.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if $\mathrm{dom} f$ is a convex set and

$$f\left(\theta u + (1-\theta)v\right) \le \theta f(u) + (1-\theta) f(v) \tag{5.2}$$

for all $u, v \in \mathrm{dom} f$ and $0 < \theta < 1$.

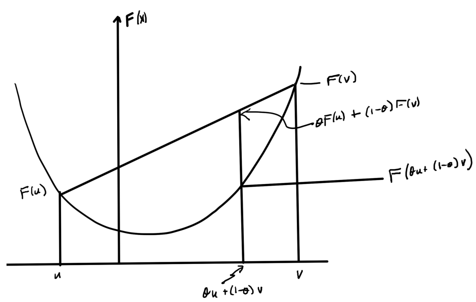


FIGURE 5.2. Illustrating the definition of a convex function.

Function $f$ is *concave* if $-f$ is convex. A function $f$ is *strictly convex* if the inequality in (5.2) is strict.

Examples of convex functions on $\mathbb{R}$ include affine, exponential $e^{ax}$ for any $a \in \mathbb{R}$, powers $x^\alpha$ on $\mathbb{R}_{++}$ for $\alpha \ge 1$ or $\alpha \le 0$, powers of absolute value $|x|^p$ on $\mathbb{R}$, for $p \ge 1$, and negative entropy $x \log x$ on $\mathbb{R}_{++}$. On $\mathbb{R}^n$ affine functions are convex and concave; all norms on any normed vector space are convex. Examples on $\mathbb{R}^{m \times n}$ include $f(X) = \mathrm{tr}(A^T X) + b$, and the spectral (maximum singular value) norm

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = \left(\lambda_{\max}(X^T X)\right)^{1/2}.$$

A key property which facilitates the proof of many theorems is the following:

*Theorem* 5.1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the function $g : \mathbb{R} \to \mathbb{R}$,

$$g(t) = f(x + tv) \quad \text{with} \quad \mathrm{dom}(g) = \{t \mid x + tv \in \mathrm{dom} f\}$$

is convex (in $t$) for any $x \in \mathrm{dom} f$ and $v \in \mathbb{R}^n$.

Hence one can check convexity of $f$ by checking convexity of functions of one variable.

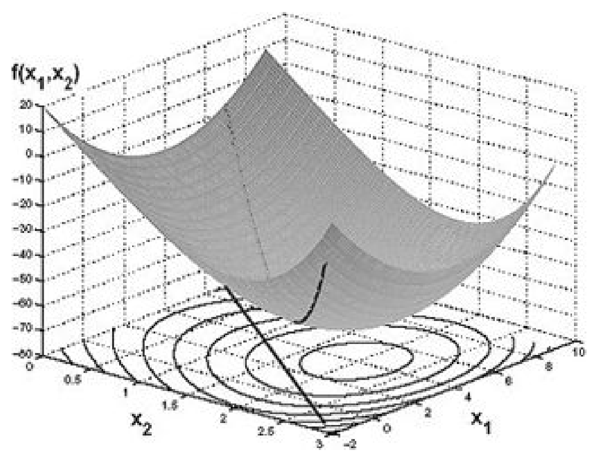


FIGURE 5.3. Two-dimensional convex functions and their 1d slices.

A function $f$ is *differentiable* if $\mathrm{dom} f$ is open and the gradient exists at each $x \in \mathrm{dom} f$. The following fact is known as the *first-order condition* for convexity.

*Theorem* 5.2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable with convex domain. Then $f$ is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \text{ for all } x, y \in \mathrm{dom} f. \tag{5.3}$$

In other words the first-order approximation of $f$ is global underestimator.

For $n = 2$, (5.3) says that the surface $z = f(x, y)$ never passes below any of its tangent planes.

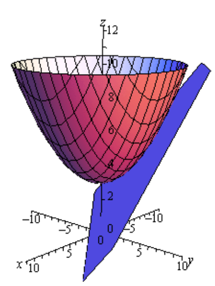

FIGURE 5.4. The *first-order condition* for convexity says the surface $z = f(x, y)$ never passes below any of its tangent planes.

The proof of Theorem 5.2 is easy given other facts presented above; for example, it suffices to check convexity on all one-dimensional subsets of the domain, and on such subsets, (5.3) simplifies.

There are also *second-order conditions* for checking convexity of differentiable functions. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *twice differentiable* if $\mathrm{dom} f$ is open and the Hessian $\nabla^2 f(x) \in S^n$ exists at each $x \in \mathrm{dom} f$.

*Theorem* 5.3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable with convex domain. Then $f$ is a convex function if and only if

$$\nabla^2 f(x) \succeq 0 \text{ for all } x \in \mathrm{dom} f.$$

If $\nabla^2 f(x) \succ 0$ for all $x \in \mathrm{dom} f$, then $f$ is strictly convex.

The proof, which we omit, proceeds by relating the second derivative $g''(t)$ of the 1d function $g(t) = f(x + tv)$ to $\langle v, \nabla^2 f(x) v \rangle$ and using related results from ordinary calculus.

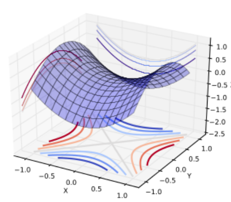

FIGURE 5.5. If $\nabla^2 f(x) \succ 0$ fails to be true for both $f$ and $-f$, then there is probably some point which has both positive and negative curvature directions. Such a point is called a *saddle point*.

Let's see some examples of multivariate functions for which we can verify convexity using the second-order condition. For our first example, consider a quadratic function:

$$f(x) = (1/2)x^T P x + q^T x + r \text{ (with } P \in S^n)$$

Then one calculates $\nabla f(x) = Px + q$, $\nabla^2 f(x) = P$. The function $f(x)$ is convex if $P \succeq 0$.

A second example is the familiar least-squares objective: $f(x) = \|Ax - b\|^2$, which has

$$\nabla f(x) = 2A^T(Ax - b), \quad \nabla^2 f(x) = 2A^T A$$

This is convex for any $A$. The quadratic-over-linear function $f(x, y) = x^2/y$ is convex for $y > 0$ (just compute the Hessian).

The log-sum-exp function $f(x) = \log \sum_k \exp x_k$ is convex; here we can compute the Hessian easily enough:

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}' z} \operatorname{diag}(z) - \frac{1}{(\mathbf{1}' z)^2} z z^T \quad \text{for } z_k = \exp x_k$$

but there are easier ways of verifying convexity in this case, such as using the composition theorem. Along similar lines, the geometric mean $f(x) = (\prod_k x_k)^{1/n}$ on $R^n_{++}$ is concave.

Define the $\alpha$-*sublevel* set of $f : \mathbb{R}^n \to \mathbb{R}$ as

$$C_\alpha = \{x \in \operatorname{dom} f \mid f(x) \leq \alpha\}$$

Fact: sublevel sets of convex functions are convex.

Define the *epigraph* of $f : \mathbb{R}^n \to \mathbb{R}$ as

$$\operatorname{epi} f = \{(x, t) \in \mathbb{R}^{n+1} \mid x \in \operatorname{dom} f, f(x) \leq t\}$$

A function $f$ is convex if and only if epi$f$ is a convex set. This is another way of checking convexity.

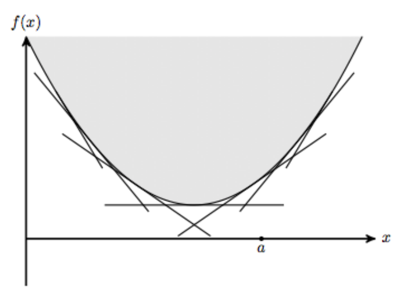


FIGURE 5.6. The shaded region is the epigraph.

If f is convex, then

$$f(\mathbb{E} z) \leq \mathbb{E} f(z)$$

for any random variable $z$. This is known as *Jensen's inequality*. The basic inequality (5.2) that defines convexity is a special case with a discrete distribution.

In practice a great way to prove convexity of $f$ is to show that $f$ is obtained from simple convex functions by operations that preserve convexity:

- nonnegative weighted sum
- composition with affine function
- pointwise maximum and supremum
- composition (sometimes; see Theorem 5.4 below)

- minimization
- perspective

Examples include the log barrier for linear inequalities

$$f(x) = -\sum \log(b_i - a_i^T x),$$

with

$$\mathrm{dom} f = \{x \mid a_i^T x < b_i, i = 1, \ldots, m\}$$

and any norm of affine function: $f(x) = \|Ax + b\|$.

If $f_1, \ldots, f_m$ are convex, then

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

is convex. For example, a piecewise-linear function

$$f(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i)$$

is convex. Similarly, if $f(x, y)$ is convex in $x$ for each $y \in A$, then

$$g(x) = \sup_{y \in A} f(x, y)$$

is convex.

Other examples include distance to farthest point in a set $C$:

$$f(x) = \sup_{y \in C} \|x - y\|$$

and the maximum eigenvalue of symmetric matrix: for $X \in S^n$,

$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^T X y.$$

and then $\lambda_{\max}(X)$ is a convex function of $X$.

The *extended-value extension* $\tilde{f}$ of $f$ is defined as:

$$\tilde{f}(x) = f(x) \text{ if } x \in \mathrm{dom} f, \text{ and } \tilde{f}(x) = \infty \text{ if } x \notin \mathrm{dom} f.$$

*Theorem* 5.4. Consider the composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$:

$$f(x) = h(g(x))$$

Then $f$ is convex if $g$ convex, $h$ convex, $\tilde{h}$ nondecreasing. Alternatively if $g$ is concave and $\tilde{h}$ nonincreasing, then again $f$ is convex.

*Proof (Sketch).* For $n = 1$ and differentiable $g, h$:

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x).$$

This needs to be $\geq 0$. So for example, if $h$ is convex then the first term is nonnegative, and if $g$ is convex and $h' \geq 0$ at all points in its domain then the second term is nonnegative too.

Applications of Theorem 5.4: $\exp g(x)$ is convex if $g$ is convex. Also $1/g(x)$ is convex if $g$ is concave and positive.

If $f(x, y)$ is convex in $(x, y)$ and $C$ is a convex set, then

$$g(x) = \inf_{y \in C} f(x, y)$$

is convex. Therefore, the distance to a set $d(x, S) = \inf_{y \in S} \|x - y\|$ is convex if $S$ is convex.

*Definition* 5.2. The *convex conjugate* of a function $f$ is

$$f^*(y) = \sup_{x \in \mathrm{dom} f} \left( y^T x - f(x) \right). \tag{5.4}$$

It is also known as the Legendre-Fenchel transformation or Fenchel transformation (after Adrien-Marie Legendre and Werner Fenchel).

This is another way of generating convex functions: $f^*$ is convex (even if $f$ is not). Also, this conjugation operation is fundamental to the construction of the *dual problem*, which we will discuss later. The dual is a very important practical tool for actually solving optimization problems numerically.

The *standard form optimization problem* is defined to be:

$$\begin{aligned} \text{minimize } & f_0(x) \\ \text{subject to } & f_i(x) \leq 0, \; i = 1, \ldots, m \\ & h_i(x) = 0, \; i = 1, \ldots, p \end{aligned}$$

The optimal value will be denoted

$$p^* = \inf \{ f_0(x) \mid f_i(x) \leq 0, \; h_j(x) = 0 \; (\forall i, j) \}$$

By convention, $p^* = \infty$ if the problem is infeasible (no x satisfies the constraints), and $p^* = -\infty$ if the problem is unbounded below. A constraint $f_i, i > 0$ is said to be *active at* $x_0$ if $f_i(x_0) = 0$.

A point $x$ is said to be *feasible* if $x \in \text{dom} f_0$ and it satisfies the constraints. The *feasible set* is the set of all feasible $x$. A feasible $x$ is said to be *optimal* if $f_0(x) = p^*$. We will use $X_{opt}$ to denote the set of optimal points. A point $z$ is *locally optimal* if it's optimal for the original problem with added constraint that $\|x - z\| \le R$ for some $R > 0$.

The standard form optimization problem has an implicit constraint

$$x \in \mathcal{D} = \bigcap_{i=0}^{m} \text{dom}(f_i) \ \cap \ \bigcap_{j=1}^{p} \text{dom}(h_j)$$

This $\mathcal{D}$ is called the domain of the problem. A problem is said to be *unconstrained* if it has no explicit constraints ($m = p = 0$). For example, minimization of

$$f_0(x) = -\sum_i \log(b_i - a_i^T x)$$

is an unconstrained problem with implicit constraints $a_i^T x < b_i$.

The standard form *convex* optimization problem is the special case of the above in which $f_0, f_1, \ldots, f_m$ are convex and equality constraints are affine. In other words, the $p$ equality constraints can be written as $Ax = b$ where $b \in \mathbb{R}^p$. We formalize these properties in the next definition.

*Definition* 5.3. The *standard form convex optimization problem* is

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

where all $f_i$ are convex.

Consider any of the functions $f_i(x)$, $i = 1, \ldots, m$ in a convex optimization problem. The domain of a convex function is convex, so $\text{dom}(f_i)$ is a convex set. Furthermore, the inverse image of a half-space under a convex function is also a convex set, so

$$\{x \in \text{dom}(f_i) : f_i(x) \le 0\}$$

is a convex set. The intersection of convex sets is convex, so the entire feasible set of a convex optimization problem is convex, as is the domain of the problem

$$\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom}(f_i).$$

*Theorem* 5.5. Any locally optimal point of a convex problem is (globally) optimal.

*Proof.* Suppose $x$ is locally optimal, which means $\exists\, R > 0$ such that

$$f_0(x') \geq f_0(x) \text{ for all points } x' \text{ with } x' \text{ feasible, } d(x, x') \leq R \tag{5.5}$$

Suppose also that $\exists\, y$ with $f_0(y) < f_0(x)$, hence $d(y, x) > R$ or else (5.5) would be a contradiction. Define

$$z = \theta y + (1 - \theta)x \text{ with } \theta := \frac{R}{2\|x - y\|} < \frac{1}{2}$$

Note $z$ is a convex combination of two feasible points, hence feasible itself. Note that $d(z, x) = \|z - x\| = \theta\,\|y - x\| = R/2$, so $z \in B(x; R)$. Finally,

$$f_0(z) \leq \theta f_0(y) + (1 - \theta) f_0(x) < f_0(x) \quad \text{since} \quad f_0(y) < f_0(x)$$

which contradicts our assumption that $x$ is locally optimal, so $y$ with $f_0(y) < f_0(x)$ cannot exist.

*Theorem* 5.6. If $f$ is differentiable and convex, then any point $x_0 \in \operatorname{dom} f$ satisfying $\nabla f(x_0) = 0$ is a global minimum of $f(x)$.

*Proof.* By the first-order condition for convexity,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)\ \forall\, y.$$

At $x = x_0$, $\nabla f(x_0) = 0$ and hence this reduces to $f(y) \geq f(x_0)\ \forall\, y$. $\square$

Two problems are (informally) *equivalent* if the solution of one is readily obtained from the solution of the other, and vice-versa. Convex problems can be equivalent to non-convex ones. There are some common transformations that preserve convexity, and which sometimes transform the problem into one that's easier to deal with. The standard-form convex problem in Definition 5.3 is equivalent to the following one:

$$\begin{aligned} \text{minimize (over } z) :&\ \ f_0(Fz + x_0) \\ \text{subject to } :&\ \ f_i(Fz + x_0) \leq 0,\ i = 1, ..., m \end{aligned}$$

where $F$ and $x_0$ are such that

$$Ax = b \iff x = Fz + x_0 \text{ for some } z.$$

Similarly, there is a technique known as introducing *slack variables* for linear inequalities. The problem

$$\text{minimize } f_0(x) \ \text{ subject to } \ a_i^T x \le b_i, \quad i = 1, \dots, m$$

is equivalent to

$$\begin{aligned} \text{minimize (over } x, s) : \ & f_0(x) \\ \text{subject to } : \ & a_i^T x + s_i = b_i, \quad \text{and} \quad s_i \ge 0, \\ & i = 1, \dots, m, \end{aligned}$$

For a general optimization problem in standard form (not necessarily convex), define the *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ , with $\mathrm{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$ as follows:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x).$$

The elements of $\lambda, \nu$ are called *Lagrange multipliers*. Define the *Lagrange dual* function: $g : \mathbb{R}^m \times \mathbb{R}^p \to R$, as

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$$

Note that g is concave, and can be $-\infty$ for some $\lambda, \nu$.

*Theorem* 5.7 (lower bound property). If $\lambda \succeq 0$, then $g(\lambda, \nu) \le p^*$.

*Proof.* For any feasible $\tilde{x}$, one has

$$f_0(\tilde{x}) \ge L(\tilde{x}, \lambda, \nu) \ge \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu)$$

where we used that $\lambda \succeq 0$. Minimizing over all feasible $\tilde{x}$ gives $p^* \ge g(\lambda, \nu)$. The proof is complete.

The Lagrange dual is closely related to the conjugate function which we introduced above in (5.4). Consider the problem

$$\begin{gathered} \text{minimize } f_0(x) \\ \text{subject to } Ax \preceq b, Cx = d \end{gathered}$$

The Lagrange dual function is

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \ \in \ \mathrm{dom} f_0} \big( f_0(x) + (A^T \lambda + C^T \nu)^T x - b^T \lambda - d^T \nu \big) \\ &= -f_0^*(-A^T \lambda - C^T \nu) - b^T \lambda - d^T \nu \qquad (5.6) \end{aligned}$$

Hence it greatly simplifies the computation of the dual if the conjugate of $f_0$ is known.

Recall that by Theorem 5.7, if $\lambda \succeq 0$, then $g(\lambda, \nu) \leq p^*$. This provides all sorts of lower bounds for $p^*$, but some are better than others. What if we maximized $g(\lambda, \nu)$ over the space of all $\lambda \succeq 0$? Indeed, the *Lagrange dual problem* is the following:

$$\text{maximize } g(\lambda, \nu) \text{ subject to } \lambda \succeq 0. \tag{5.7}$$

If we could solve this problem, we would know the best lower bound on $p^*$ that could be obtained by applying the lower bound property of the dual function. Eq. (5.7) is generally a convex optimization problem, even if the original one wasn't! The optimal value of the dual problem is typically denoted $d^*$. The pair $\lambda, \nu$ are said to be *dual feasible* if $\lambda \succeq 0$, and $(\lambda, \nu) \in \operatorname{dom}(g)$.

As an example, consider the standard form LP and its dual:

$$\text{minimize } c^T x \quad \text{subject to} \quad Ax = b,\ x \succeq 0$$

The Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x \end{aligned}$$

$L$ is affine in $x$, hence

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & c + A^T \nu - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Hence the dual problem is

$$\text{maximize} \quad -b^T \nu \quad \text{subject to} \quad A^T \nu + c \succeq 0$$

The statement $d^* \leq p^*$ always holds, and is known as *weak duality.* The weak duality bound is sometimes useful for finding nontrivial lower bounds for difficult problems. The equality $d^* = p^*$ does not usually hold, but when it holds, it is known as *strong duality.* Strong duality usually does hold for convex problems, under conditions that we'll make precise very soon. The difference $p^* - d^*$ is known as the *duality gap*, and is always positive.

More generally, dual feasible points allow us to bound how suboptimal a given feasible point is, without knowing the exact value of $p^*$. Indeed, if $x$

is primal feasible and $(\lambda, \nu)$ is dual feasible, then

$$f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu).$$

We refer to the gap between primal and dual objectives, $f_0(x) - g(\lambda, \nu)$, as the duality gap associated with the primal feasible point $x, (\lambda, \nu)$. A primal-dual feasible pair $x, (\lambda, \nu)$ localizes the optimal value of the primal (and dual) problems to an interval:

$$p^*, d^* \in [g(\lambda, \nu), f_0(x)].$$

Any condition that guarantees strong duality in convex problems is called a *constraint qualification*. The most well-known one is *Slater's condition*, for which we need a few definitions. The *affine hull* $\mathrm{aff}(S)$ of a set $S$ in a vector space is the smallest affine set containing $S$, or equivalently, the intersection of all affine sets containing $S$. The *relative interior* of a set $S$, defined as:

$$\mathrm{relint}(S) := \{x \in S : \exists \epsilon > 0, B(x, \epsilon) \cap \mathrm{aff}(S) \subseteq S\},$$

If this definition seems abstract, just think of the set

$$\{(x, y, 0) \in \mathbb{R}^3 \mid x^2 + y^2 \leq 1\}.$$

Its interior (with respect to $\mathbb{R}^3$) is empty, but its relative interior is an open unit disk in the $xy$-plane.

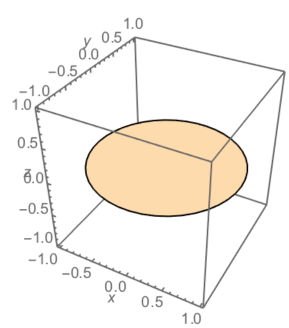


FIGURE 5.7. The shaded region is the relative interior of the set $\{(x, y, 0) \in \mathbb{R}^3 \mid x^2 + y^2 \leq 1\}$.

*Theorem* 5.8 (Slater's condition)*.* Strong duality holds for a convex problem if

(5.8) $\exists\, x \in \operatorname{relint}(\mathcal{D})$ such that: $f_i(x) < 0$ for $i = 1, \ldots, m$, and $Ax = b$.

Under these conditions, the dual optimum is attained if $p^* > -\infty$.

When (5.8) holds, the problem is said to be *strictly feasible*, so a simple mnemonic for Slater's condition is: "strictly feasible implies strongly dual." The theorem can be sharpened in various ways. For example, affine inequalities do not need to hold with strict inequality. There exist non-convex problems with strong duality, it just isn't guaranteed in general.

We'll now begin some considerations which will lead us up to the important *Karush-Kuhn-Tucker (KKT) conditions*, which are practically what one uses to actually solve many optimization problems.

Assume strong duality holds, $x^*$ is primal optimal, $(\lambda^*, \nu^*)$ is dual optimal. Then

$$
\begin{aligned}
f_0(x^*) = g(\lambda^*, \nu^*) &= \inf_x \left( f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_j \nu_j^* \, h_j(x) \right) \\
&\le f_0(x^*) + \sum_i \lambda_i^* \, f_i(x^*) + \sum_j \nu_j^* \, h_j(x^*) \\
&= f_0(x^*)
\end{aligned}
$$

To get from the second line to the third, note that all $h_j(x^*) = 0$, and $\lambda_i^* \, f_i(x^*) \le 0$ by our conventions and definitions, but if it were strictly negative, then we'd have $f_0(x^*) = f_0(x^*) + \text{negative}$, which is a contradiction. Hence the inequality is actually an equality.

To summarize: if strong duality holds, $x^*$ is primal optimal, $(\lambda^*, \nu^*)$ is dual optimal, then

$$\lambda_i^* \, f_i(x^*) = 0 \quad \text{for all } i = 1, \ldots, m.$$

This gives us a way of seeing which constraints are active at the solution. If some $\lambda_i^* > 0$ then $f_i(x^*)$ must be zero, meaning the constraint is *active*, or alternatively if $f_i(x^*) < 0$ (meaning the constraint isn't active) then $\lambda_i^* = 0$. So any active constraints will have nonzero Lagrange multipliers in the solution to the dual problem. Any non-active constraints will have their corresponding Lagrange multipliers in the dual-optimal solution equal to zero. The above chain of inequalities also directly imply that $x^*$ minimizes $L(x, \lambda^*, \nu^*)$.

*Definition* 5.4. A *primal-dual pair* is a pair

$$\zeta = \{x, (\lambda, \nu)\} \in \mathcal{D} \times \mathrm{dom}(g).$$

In a problem with differentiable $f_i, h_i$, the primal-dual pair $\zeta$ is said to satisfy the *KKT conditions* if the following four things hold:

(1) primal constraints: $f_i(x) \leq 0$, $i = 1, \ldots, m$, $h_i(x) = 0$, $i = 1, ..., p$
(2) dual constraints: $\lambda \succeq 0$
(3) complementary slackness: $\lambda_i f_i(x) = 0$, $i = 1, \ldots, m$
(4) first-order condition:

$$\nabla f_0(x) + \sum_i \lambda_i \nabla f_i(x) + \sum_j \nu_j \nabla h_j(x) = 0\,.$$

From our discussion above, if strong duality holds and $x, \lambda, \nu$ are optimal, then they must satisfy the KKT conditions. This is the case for convex and nonconvex problems alike.

Furthermore, if $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ satisfy KKT for a convex problem, then they are optimal. To see this, note that the first two KKT conditions state that $\tilde{x}$ is primal feasible. Since $\tilde{\lambda} \succeq 0$, it follows that $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in $x$; the last KKT condition states that its gradient with respect to $x$ vanishes at $x = \tilde{x}$, so it follows that $\tilde{x}$, minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ by the first-order condition. Furthermore,

$$f_0(\tilde{x}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = g(\tilde{\lambda}, \tilde{\nu}).$$

In summary, for any convex optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

If Slater's condition is satisfied in a convex problem, then $x$ is optimal if and only if there exist $\lambda, \nu$ that satisfy the KKT conditions with the given $x$. This can be viewed as a generalization of the optimality condition $\nabla f_0(x) = 0$ for unconstrained problems.

A point $x_0$ is said to be *regular* if the gradients of the active constraints at $x_0$ are linearly independent. For problems without strong duality, the KKT conditions are necessary for *local* optimality under a mild assumption: If $x^*$ is locally optimal and $x^*$ is a regular point then there exists $\lambda^*, \nu^*$ such that the KKT conditions hold.

An important example in which the KKT conditions can be written down, and solved, explicitly is *equality constrained convex quadratic minimization*:

$$\text{minimize } (1/2)x^T P x + q^T x + r \qquad \text{subject to } Ax = b,$$

where $P \in S_+^n$. The KKT conditions for this problem are

$$Ax^* = b, \quad Px^* + q + A^T\nu^* = 0$$

which we can write equivalently as a single matrix equation,

$$\begin{pmatrix} P & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \nu^* \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix}$$

Solving this set of $m+n$ equations in the $m+n$ variables $(x^*, \nu^*)$ gives the optimal primal and dual variables. Assuming the relevant inverses exist, one could apply the block inversion formula to get an explicit formula for $x^*$ that doesn't depend on $\nu^*$ (exercise).

Now consider a version of the above in which we have linear inequality constraints. For simplicity, we'll drop the linear term in $f_0$. So the "primal" or original problem is, for $P \in S_{++}^n$,

$$\text{minimize } x^T P x \ \text{ subject to } Ax \preceq b.$$

The dual function is

$$g(\lambda) = \inf_x \left(x^T P x + \lambda^T (Ax - b)\right) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda\,.$$

The dual problem is then maximizing a quadratic-plus-linear function $g(\lambda)$ over the space $\lambda \succeq 0$, which seems simpler. Slater's condition specifies that $p^* = d^*$ if $A\tilde{x} \prec b$ for some $\tilde{x}$, but in fact more is true: for this class of problems, $p^* = d^*$ in any case.

We want to be able to analyze how sensitive the optimal value is to changes in the constraints. To this end, define the *perturbed problem* as

$$\text{minimize } f_0(x) \text{ subject to } \ f_i(x) \le u_i, h_j(x) = v_j$$

for some vectors $u \in \mathbb{R}^m, v \in \mathbb{R}^p$. In other words, the perturbed problem is just like the original, but with potentially non-zero parameters $u_i$ and $v_j$ on the right-hand sides, where formerly one had zeros. When $u_i$ is positive it means that we have relaxed the $i$-th inequality constraint; when $u_i$ is negative, it means that we have tightened the constraint. Define $p^*(u, v)$

as the optimal value of the perturbed problem, noting that $p^*(0,0)$ equals what we called $p^*$ before.

Suppose now that $p^*(u,v)$ is differentiable at $u=0, v=0$. Then, provided strong duality holds, the optimal dual variables $\lambda^*, \nu^*$ are related to the gradient of $p^*(\,\cdot\,,\,\cdot\,)$ at $u=0, v=0$:

$$\lambda_i^* = -\frac{\partial p^*}{\partial u_i}(0,0), \quad \nu_j^* = -\frac{\partial p^*}{\partial v_i}(0,0)\,. \tag{5.9}$$

Thus, assuming the conditions above (differentiability and strong duality), the optimal Lagrange multipliers are exactly the local sensitivities of the optimal value with respect to constraint perturbations.Tightening the $i$-th inequality constraint a small amount (i.e., taking $u_i$ small and negative) yields an increase in $p^*$ of approximately $-\lambda_i^* u_i$. In the same way, loosening the $i$-th constraint a small amount (i.e., taking $u_i$ small and positive) yields a decrease in $p^*$ of approximately $-\lambda_i^* u_i$.

There is an economic interpretation of (5.9). We consider (for simplicity) a convex problem with no equality constraints, which satisfies Slaters condition. The variable $x$ determines how a firm operates, and the objective $f_0$ is $-1$ times the profit. Each constraint $f_i(x) \le 0$ represents a limit on some resource such as labor, steel, or warehouse space. Then $-p^*(u)$ tells us how much more or less profit could be made if more, or less, of each resource were made available to the firm.

According to (5.9), $\lambda_i^*$ tells us approximately how much more profit the firm could make, for a small increase in availability of resource $i$. It follows that $\lambda_i^*$ would be the natural or equilibrium price for resource $i$, if it were possible for the firm to buy or sell it. If the firm could buy resource $i$, at a price that is less than $\lambda_i^*$, then it should do so because it could then operate in a way that increases its profit more than the cost of buying the resource. For this reason $\lambda_i^*$ is called the *shadow price*.

5.1. **Computational Frameworks.** Nowadays there is a nice software package for solving convex optimization problems developed by Diamond and Boyd (2016), called CVXPY. It automatically transforms the problem into standard form, calls a solver, and unpacks the results. We show a basic example involving constrained regression.

```python
# Solves a bounded least-squares problem.

from cvxpy import *
import numpy

# Problem data.
m = 10
n = 5
numpy.random.seed(1)
A = numpy.random.randn(m, n)
b = numpy.random.randn(m, 1)

# Construct the problem.
x = Variable(n)
objective = Minimize(sum_entries(square(A*x - b)))
constraints = [0 <= x, x <= 1]
prob = Problem(objective, constraints)

print "Optimal value", prob.solve()
print "Optimal var"
print x.value # A numpy matrix.
```

## References


Boyd, Stephen and Lieven Vandenberghe (2009). *Convex optimization.* Cambridge university press.

Diamond, Steven and Stephen Boyd (2016). "CVXPY: A Python-Embedded Modeling Language for Convex Optimization". In: *Journal of Machine Learning Research* 17.83, pp. 1–5.