Convex Optimization

Solutions Manual

Stephen Boyd Lieven Vandenberghe

January 4, 2006

- Let c_1 be a vector in the plane defined by a_1 and a_2, and orthogonal to a_2. For example, we can take
$$c_1 = a_1 - \frac{a_1^T a_2}{\|a_2\|_2^2} a_2.$$
Then $x \in S_2$ if and only if
$$-|c_1^T a_1| \leq c_1^T x \leq |c_1^T a_1|.$$
- Similarly, let c_2 be a vector in the plane defined by a_1 and a_2, and orthogonal to a_1, *e.g.*,
$$c_2 = a_2 - \frac{a_2^T a_1}{\|a_1\|_2^2} a_1.$$
Then $x \in S_3$ if and only if
$$-|c_2^T a_2| \leq c_2^T x \leq |c_2^T a_2|.$$

Putting it all together, we can describe S as the solution set of $2n$ linear inequalities
$$\begin{array}{rcl} v_k^T x & \leq & 0, \quad k = 1, \ldots, n-2 \\ -v_k^T x & \leq & 0, \quad k = 1, \ldots, n-2 \\ c_1^T x & \leq & |c_1^T a_1| \\ -c_1^T x & \leq & |c_1^T a_1| \\ c_2^T x & \leq & |c_2^T a_2| \\ -c_2^T x & \leq & |c_2^T a_2|. \end{array}$$

(b) S is a polyhedron, defined by linear inequalities $x_k \geq 0$ and three equality constraints.

(c) S is not a polyhedron. It is the intersection of the unit ball $\{x \mid \|x\|_2 \leq 1\}$ and the nonnegative orthant $\mathbf{R}_+^n$. This follows from the following fact, which follows from the Cauchy-Schwarz inequality:
$$x^T y \leq 1 \text{ for all } y \text{ with } \|y\|_2 = 1 \iff \|x\|_2 \leq 1.$$
Although in this example we define S as an intersection of halfspaces, it is not a polyhedron, because the definition requires infinitely many halfspaces.

(d) S is a polyhedron. S is the intersection of the set $\{x \mid |x_k| \leq 1, \ \ k = 1, \ldots, n\}$ and the nonnegative orthant $\mathbf{R}_+^n$. This follows from the following fact:
$$x^T y \leq 1 \text{ for all } y \text{ with } \sum_{i=1}^n |y_i| = 1 \iff |x_i| \leq 1, \quad i = 1, \ldots, n.$$
We can prove this as follows. First suppose that $|x_i| \leq 1$ for all i. Then
$$x^T y = \sum_i x_i y_i \leq \sum_i |x_i||y_i| \leq \sum_i |y_i| = 1$$
if $\sum_i |y_i| = 1$.

Conversely, suppose that x is a nonzero vector that satisfies $x^T y \leq 1$ for all y with $\sum_i |y_i| = 1$. In particular we can make the following choice for y: let k be an index for which $|x_k| = \max_i |x_i|$, and take $y_k = 1$ if $x_k > 0$, $y_k = -1$ if $x_k < 0$, and $y_i = 0$ for $i \neq k$. With this choice of y we have
$$x^T y = \sum_i x_i y_i = y_k x_k = |x_k| = \max_i |x_i|.$$

(c) A *wedge*, *i.e.*, $\{x \in \mathbf{R}^n \mid a_1^T x \leq b_1,\ a_2^T x \leq b_2\}$.

(d) The set of points closer to a given point than a given set, *i.e.*,

$$\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\}$$

where $S \subseteq \mathbf{R}^n$.

(e) The set of points closer to one set than another, *i.e.*,

$$\{x \mid \mathbf{dist}(x, S) \leq \mathbf{dist}(x, T)\},$$

where $S, T \subseteq \mathbf{R}^n$, and

$$\mathbf{dist}(x, S) = \inf\{\|x - z\|_2 \mid z \in S\}.$$

(f) [HUL93, volume 1, page 93] The set $\{x \mid x + S_2 \subseteq S_1\}$, where $S_1, S_2 \subseteq \mathbf{R}^n$ with S_1 convex.

(g) The set of points whose distance to a does not exceed a fixed fraction θ of the distance to b, *i.e.*, the set $\{x \mid \|x - a\|_2 \leq \theta \|x - b\|_2\}$. You can assume $a \neq b$ and $0 \leq \theta \leq 1$.

Solution.

(a) A slab is an intersection of two halfspaces, hence it is a convex set (and a polyhedron).

(b) As in part (a), a rectangle is a convex set and a polyhedron because it is a finite intersection of halfspaces.

(c) A wedge is an intersection of two halfspaces, so it is convex set. It is also a polyhedron. It is a cone if $b_1 = 0$ and $b_2 = 0$.

(d) This set is convex because it can be expressed as

$$\bigcap_{y \in S} \{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\},$$

i.e., an intersection of halfspaces. (For fixed y, the set

$$\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\}$$

is a halfspace; see exercise 2.9).

(e) In general this set is not convex, as the following example in $\mathbf{R}$ shows. With $S = \{-1, 1\}$ and $T = \{0\}$, we have

$$\{x \mid \mathbf{dist}(x, S) \leq \mathbf{dist}(x, T)\} = \{x \in \mathbf{R} \mid x \leq -1/2 \text{ or } x \geq 1/2\}$$

which clearly is not convex.

(f) This set is convex. $x + S_2 \subseteq S_1$ if $x + y \in S_1$ for all $y \in S_2$. Therefore

$$\{x \mid x + S_2 \subseteq S_1\} = \bigcap_{y \in S_2} \{x \mid x + y \in S_1\} = \bigcap_{y \in S_2} (S_1 - y),$$

the intersection of convex sets $S_1 - y$.

(g) The set is convex, in fact a ball.

$$\begin{aligned}
&\{x \mid \|x - a\|_2 \leq \theta \|x - b\|_2\} \\
&\quad = \{x \mid \|x - a\|_2^2 \leq \theta^2 \|x - b\|_2^2\} \\
&\quad = \{x \mid (1 - \theta^2) x^T x - 2(a - \theta^2 b)^T x + (a^T a - \theta^2 b^T b) \leq 0\}
\end{aligned}$$

Chapter 3

Convex functions

3.29 *Representation of piecewise-linear convex functions.* A function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^n$, is called *piecewise-linear* if there exists a partition of $\mathbf{R}^n$ as

$$\mathbf{R}^n = X_1 \cup X_2 \cup \cdots \cup X_L,$$

where $\mathbf{int}\, X_i \neq \emptyset$ and $\mathbf{int}\, X_i \cap \mathbf{int}\, X_j = \emptyset$ for $i \neq j$, and a family of affine functions $a_1^T x + b_1, \ldots, a_L^T x + b_L$ such that $f(x) = a_i^T x + b_i$ for $x \in X_i$.

Show that this means that $f(x) = \max\{a_1^T x + b_1, \ldots, a_L^T x + b_L\}$.

Solution. By Jensen's inequality, we have for all $x, y \in \mathbf{dom}\, f$, and $t \in [0, 1]$,

$$f(y + t(x - y)) \leq f(y) + t(f(x) - f(y)),$$

and hence

$$f(x) \geq f(y) + \frac{f(y + t(x - y)) - f(y)}{t}.$$

Now suppose $x \in X_i$. Choose any $y \in \mathbf{int}\, X_j$, for some j, and take t sufficiently small so that $y + t(x - y) \in X_j$. The above inequality reduces to

$$a_i^T x + b_i \geq a_j^T y + b_j + \frac{(a_j^T (y + t(x - y)) + b_j - a_j^T y - b_j)}{t} = a_j^T x + b_j.$$

This is true for any j, so $a_i^T x + b_i \geq \max_{j=1,\ldots,L}(a_j^T x + b_j)$. We conclude that

$$a_i^T x + b_i = \max_{j=1,\ldots,L}(a_j^T x + b_j).$$

3.30 *Convex hull or envelope of a function.* The *convex hull* or *convex envelope* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$g(x) = \inf\{t \mid (x, t) \in \mathbf{conv}\, \mathbf{epi}\, f\}.$$

Geometrically, the epigraph of g is the convex hull of the epigraph of f.

Show that g is the largest convex underestimator of f. In other words, show that if h is convex and satisfies $h(x) \leq f(x)$ for all x, then $h(x) \leq g(x)$ for all x.

Solution. It is clear that g is convex, since by construction its epigraph is a convex set.

Let h be a convex lower bound on f. Since h is convex, $\mathbf{epi}\, h$ is a convex set. Since h is a lower bound on f, $\mathbf{epi}\, f \subseteq \mathbf{epi}\, h$. By definition the convex hull of a set is the intersection of all the convex sets that contain the set. It follows that $\mathbf{conv}\, \mathbf{epi}\, f = \mathbf{epi}\, g \subseteq \mathbf{epi}\, h$, *i.e.*, $g(x) \geq h(x)$ for all x.

3.31 [Roc70, page 35] *Largest homogeneous underestimator.* Let f be a convex function. Define the function g as

$$g(x) = \inf_{\alpha > 0} \frac{f(\alpha x)}{\alpha}.$$

(a) Show that g is homogeneous ($g(tx) = tg(x)$ for all $t \geq 0$).

(b) Show that g is the largest homogeneous underestimator of f: If h is homogeneous and $h(x) \leq f(x)$ for all x, then we have $h(x) \leq g(x)$ for all x.

(c) Show that g is convex.

Solution.

(a) If $t > 0$,

$$g(tx) = \inf_{\alpha > 0} \frac{f(\alpha t x)}{\alpha} = t \inf_{\alpha > 0} \frac{f(\alpha t x)}{t\alpha} = tg(x).$$

For $t = 0$, we have $g(tx) = g(0) = 0$.

Exercises

Estimation

7.1 *Linear measurements with exponentially distributed noise.* Show how to solve the ML estimation problem (7.2) when the noise is exponentially distributed, with density

$$p(z) = \begin{cases} (1/a)e^{-z/a} & z \geq 0 \\ 0 & z < 0, \end{cases}$$

where $a > 0$.

Solution. Solve the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T(y - Ax) \\ \text{subject to} & Ax \preceq y. \end{array}$$

7.2 *ML estimation and ℓ_∞-norm approximation.* We consider the linear measurement model $y = Ax + v$ of page 352, with a uniform noise distribution of the form

$$p(z) = \begin{cases} 1/(2\alpha) & |z| \leq \alpha \\ 0 & |z| > \alpha. \end{cases}$$

As mentioned in example 7.1, page 352, any x that satisfies $\|Ax - y\|_\infty \leq \alpha$ is a ML estimate.

Now assume that the parameter α is not known, and we wish to estimate α, along with the parameters x. Show that the ML estimates of x and α are found by solving the ℓ_∞-norm approximation problem

$$\text{minimize} \quad \|Ax - y\|_\infty,$$

where a_i^T are the rows of A.

Solution. The log-likelihood function is

$$l(x, \alpha) = \begin{cases} m \log(1/2\alpha) & \|Ax - y\|_\infty \leq \alpha \\ -\infty & \text{otherwise.} \end{cases}$$

Maximizing over α and y is equivalent to solving the ℓ_∞-norm problem.

7.3 *Probit model.* Suppose $y \in \{0, 1\}$ is random variable given by

$$y = \begin{cases} 1 & a^T u + b + v \leq 0 \\ 0 & a^T u + b + v > 0, \end{cases}$$

where the vector $u \in \mathbf{R}^n$ is a vector of explanatory variables (as in the logistic model described on page 354), and v is a zero mean unit variance Gaussian variable.

Formulate the ML estimation problem of estimating a and b, given data consisting of pairs (u_i, y_i), $i = 1, \ldots, N$, as a convex optimization problem.

Solution. We have

$$\mathbf{prob}(y = 1) = Q(a^T u + b), \qquad \mathbf{prob}(y = 0) = 1 - Q(a^T u + b) = P(-a^T u - b)$$

where

$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{t^2/2}\, dt.$$

The log-likelihood function is

$$l(a, b) = \sum_{y_i = 1} \log Q(a^T u_i + b) + \sum_{y_i = 0} \log Q(-a^T u_i - b),$$

which is a concave function of a and b.

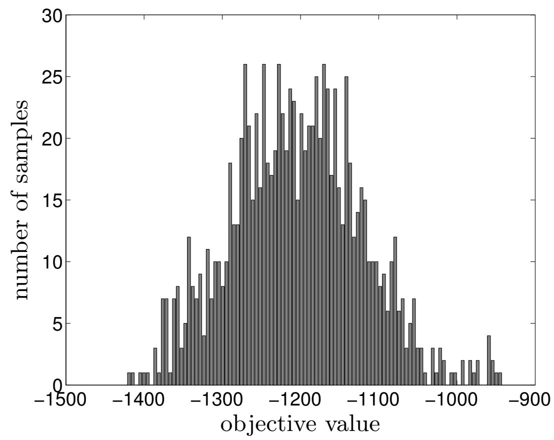
We notice that the lower bound is equal (or very close) to $p^\star$ in 10 cases, and never less than about 15% below $p^\star$.

We also generate a larger problem instance, with $n = 100$. The optimal value of the relaxation is -1687.5. The lower bound from the eigenvalue decomposition of W (see remark 5.1) is $n\lambda_{\min}(W) = -1898.4$.

(b) We first try the heuristic on the family of 100 problems with $n = 10$. The heuristic gave the correct solution in 70 instances. For the larger problem, the heuristic gives the upper bound -1336.5. At this point we can say that the optimal value of the larger problem lies between -1336.5 and -1687.5.

(c) We first try this heuristic, with $K = 10$, on the family of 100 problems with $n = 10$. The heuristic gave the correct solution in 88 instances.

We plot below a histogram of the objective obtained by the randomized heuristic, over 1000 samples.



Many of these samples have an objective value larger than the one found in part (b) above, but some have a lower cost. The minimum value is -1421.7, so $p^\star$ lies between -1421.7 and -1687.5.

(d) The contribution of x_j to the cost is $(\sum_{i=1}^n W_{ij}x_i)x_j$. If this number is positive, then switching the sign of x_j will decrease the objective by $2\sum_{i=1}^n W_{ij}x_i$.

We apply the greedy heuristic to the larger problem instance. For $x = \mathbf{1}$, the cost is reduced from 13.6 to -1344.8. For the solution from part (b), the cost is reduced from -1336.5 to -1440.6. For the solution from part (b), the cost is reduced from -1421.7 to -1440.6.

11.24 *Barrier and primal-dual interior-point methods for quadratic programming.* Implement a barrier method, and a primal-dual method, for solving the QP (without equality constraints, for simplicity)

$$\begin{array}{ll} \text{minimize} & (1/2)x^T Px + q^T x \\ \text{subject to} & Ax \preceq b, \end{array}$$

with $A \in \mathbf{R}^{m\times n}$. You can assume a strictly feasible initial point is given. Test your codes on several examples. For the barrier method, plot the duality gap versus Newton steps. For the primal-dual interior-point method, plot the surrogate duality gap and the norm of the dual residual versus iteration number.

Solution. The first figure shows the progress (duality gap) versus Newton iterations for the barrier method, applied to a randomly generated instance with $n = 100$ variables and $m = 200$ constraints. We use $\mu = 20$, $\alpha = 0.01$, $\beta = 0.5$, and $t^{(0)} = 1$. We choose $b \succ 0$, and use $x^{(0)} = 0$.