L. Vandenberghe May 16, 2018

ECE236B homework #6 solutions

1. *Exercise T5.19.*

   (a) For simplicity we assume that the elements of x are sorted in decreasing order:

   $$x_1 \geq x_2 \geq \cdots \geq x_n.$$

   It is easy to see that the optimal value is

   $$x_1 + x_2 + \cdots + x_r$$

   and is obtained by choosing $y_1 = y_2 = \cdots = y_r = 1$ and $y_{r+1} = \cdots = y_n = 0$.

   (b) We first change the objective from maximization to minimization:

   $$\begin{array}{ll} \text{minimize} & -x^T y \\ \text{subject to} & 0 \preceq y \preceq \mathbf{1} \\ & \mathbf{1}^T y = r. \end{array}$$

   We introduce a Lagrange multiplier λ for the lower bound, u for the upper bound, and t for the equality constraint. The Lagrangian is

   $$\begin{array}{rcl} L(y, \lambda, u, t) & = & -x^T y - \lambda^T y + u^T(y - \mathbf{1}) + t(\mathbf{1}^T y - r) \\ & = & -\mathbf{1}^T u - rt + (-x - \lambda + u + t\mathbf{1})^T y. \end{array}$$

   Minimizing over y yields the dual function

   $$g(\lambda, u, t) = \left\{ \begin{array}{ll} -\mathbf{1}^T u - rt & -x - \lambda + u + t\mathbf{1} = 0 \\ -\infty & \text{otherwise.} \end{array} \right.$$

   The dual problem is to maximize $g(\lambda, u, t)$ subject to $\lambda \succeq 0$ and $u \succeq 0$:

   $$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T u - rt \\ \text{subject to} & -\lambda + u + t\mathbf{1} = x \\ & \lambda \succeq 0, \quad u \succeq 0, \end{array}$$

   After changing the objective to minimization (*i.e.*, undoing the sign change we started with), we obtain

   $$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u + rt \\ \text{subject to} & u + t\mathbf{1} \succeq x \\ & u \succeq 0. \end{array}$$

   We eliminated λ by noting that it acts as a slack variable in the first constraint.

(c) The problem is equivalent to the QP

$$\begin{array}{ll}\text{minimize} & x^T\Sigma x\\ \text{subject to} & \overline{p}^T x \geq r_{\min}\\ & \mathbf{1}^T x = 1, \quad x \succeq 0\\ & \lfloor n/10 \rfloor t + \mathbf{1}^T u \leq 0.8\\ & t\mathbf{1} + u \succeq x\\ & u \succeq 0,\end{array}$$

with variables x, u, t, v.

2. *Exercise A4.14.*

(a) The KKT conditions are

$$\frac{1}{a^T x} a + \frac{1}{b^T x} b \preceq \nu \mathbf{1} \qquad x \succeq 0, \qquad \mathbf{1}^T x = 1,$$

plus the complementary slackness conditions

$$x_k\left(\nu - \frac{1}{a^T x} a_k - \frac{1}{b^T x} b_k\right) = 0, \quad k = 1, \ldots, n.$$

We show that $x = (1/2, 0, \ldots, 0, 1/2)$, $\nu = 2$ solve these equations, and hence are primal and dual optimal.
The feasibility conditions $x \succeq 0$, $\mathbf{1}^T x = 1$ obviously hold, and the complementary slackness conditions are trivially satisfied for $k = 2, \ldots, n-2$. It remains to verify the inequality

$$\frac{a_k}{a^T x} + \frac{b_k}{b^T x} \leq \nu, \quad k = 1, \ldots, n, \tag{1}$$

and the complementary slackness condition

$$x_k\left(\nu - \frac{1}{a^T x} a_k - \frac{1}{b^T x} b_k\right) = 0, \quad k = 1, n. \tag{2}$$

For $x = (1/2, 0, \ldots, 0, 1/2)$, $\nu = 2$ the inequality (1) holds with equality for $k = 1$ and $k = n$, since

$$\frac{a_1}{a^T x} + \frac{b_1}{b^T x} = \frac{2a_1}{a_1 + a_n} + \frac{2/a_1}{1/a_1 + 1/a_n} = 2,$$

and

$$\frac{a_n}{a^T x} + \frac{b_n}{b^T x} = \frac{2a_n}{a_1 + a_n} + \frac{2/a_n}{1/a_1 + 1/a_n} = 2.$$

Therefore also (2) is satisfied. The remaining inequalities in (1) reduce to

$$\frac{a_k}{a^T x} + \frac{b_k}{b^T x} = 2\frac{a_k + a_1 a_n / a_k}{a_1 + a_n} \leq 2, \quad k = 2, \ldots, n-1.$$

This is valid, since it holds with equality for $k = 1$ and $k = n$, and the function $t + a_1 a_n/t$ is convex in t, so

$$\frac{t + a_1 a_n/t}{a_1 + a_n} \leq 2$$

for all $t \in [a_n, a_1]$.

(b) Diagonalize A using its eigenvalue decomposition $A = Q\Lambda Q^T$, and define $a_k = \lambda_k$, $b_k = 1/\lambda_k$, $x_k = (Q^T u)_k^2$. From part (a), $Q^T u = (1/\sqrt{2}, 0, \ldots, 1/\sqrt{2})$ is optimal. Therefore,

$$\begin{aligned}(u^T A u)(u^T A^{-1} u) &\leq \frac{1}{4}(\lambda_1 + \lambda_n)(\lambda_1^{-1} + \lambda_n^{-1}) \\ &= \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2.\end{aligned}$$

3. *Exercise A4.20.* The Lagrangian is

$$L(x, z) = \sum_{i=1}^{n} \left(\phi(x_i) - x_i(a_i^T z)\right) + b^T z$$

where a_i is the ith column of A. The dual function is

$$\begin{aligned}g(z) &= b^T z + \sum_{i=1}^{n} \inf_{x_i} \left(\phi(x_i) - x_i(a_i^T z)\right) \\ &= b^T z + \sum_i h(a_i^T z)\end{aligned}$$

where $h(y) = \inf_u (\phi(u) - yu)$ and the dual problem is

$$\text{maximize} \quad b^T z + \textstyle\sum_i h(a_i^T z).$$

We now work out an expression for the function h. If $|y| \leq 1/c$, the minimizer in the definition of h is $u = 0$ and $h(y) = 0$. Otherwise, we find the minimum by setting the derivative equal to zero. If $y > 1/c$, we solve

$$\phi'(u) = \frac{c}{(c-u)^2} = y.$$

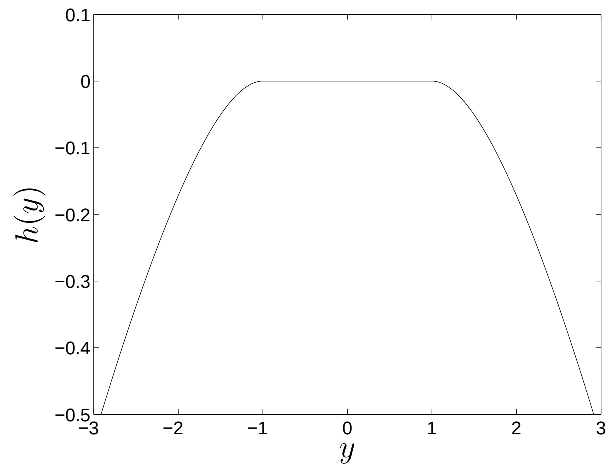
The solution is $u = c - (c/y)^{1/2}$ and $h(y) = -(1 - \sqrt{cy})^2$. If $y < -1/c$, we solve

$$\phi'(u) = -\frac{c}{(c+u)^2} = y.$$

The solution is $u = -c + (-c/y)^{1/2}$ and $h(y) = -(1 - \sqrt{-cy})^2$. Combining the different cases, we can write

$$h(u) = \begin{cases} -\left(1 - \sqrt{c|y|}\right)^2 & |y| > 1/c \\ 0 & \text{otherwise.} \end{cases}$$

The figure shows the function h for $c = 1$.



4. *Exercise A4.26.*

   (a) The Lagrangian is $L(x, y, z) = \|y\|_2 + \gamma\|x\|_1 + z^T(Ax - b - y)$. The infimum over x and y is

   $$\inf_{x,y} L(x, y, z) = \begin{cases} -b^T z & \|z\|_2 \leq 1, \ \|A^T z\|_\infty \leq \gamma \\ -\infty & \text{otherwise.} \end{cases}$$

   The dual problem is

   $$\begin{array}{ll} \text{minimize} & -b^T z \\ \text{subject to} & \|A^T z\|_\infty \leq \gamma \\ & \|z\|_2 \leq 1. \end{array}$$

   (b) The KKT conditions for optimality of x, y, z are:

      i. *Primal feasibility:* $y = Ax - b$.

      ii. *Dual feasibility:* $\|z\|_2 \leq 1$ and $\|A^T z\|_\infty \leq \gamma$.

      iii. *Minimum of Lagrangian:* x, y minimize $L(\tilde{x}, \tilde{y}, z)$. Therefore

      $$\|y\|_2 - z^T y = \inf_{\tilde{y}} \left(\|\tilde{y}\|_2 - z^T \tilde{y} \right) = 0,$$

      and

      $$\gamma\|x\|_1 + z^T Ax = \inf_{\tilde{x}} \left(\gamma\|\tilde{x}\|_1 + z^T A\tilde{x} \right) = 0.$$

   We apply these conditions to $x = x^\star$ and $y = Ax^\star - b$. The first part of condition (iii), combined with $\|z\|_2 \leq 1$ and $y \neq 0$, implies that $z = y/\|y\|_2 = r$. From condition (ii), we have $\|A^T r\|_\infty \leq \gamma$ and from the second part of condition (iii), $\gamma\|x\|_1 + r^T Ax = 0$.

   (c) The condition $-(A^T r)^T x^\star = \gamma\|x^\star\|_1$ with $\|A^T r\|_\infty \leq \gamma$, holds only if, for each column a_k,

   $$a_k^T r = \gamma \text{ if } x_k^\star < 0, \qquad a_k^T r = -\gamma \text{ if } x_k^\star > 0, \qquad |a_k^T r| \leq \gamma \text{ if } x_k^\star = 0.$$

Now, from the Cauchy-Schwarz inequality, since $\|r\|_2 = 1$, we have

$$|a_i^T r| \le \|a_i\|_2 \|r\|_2 = \|a_i\|_2.$$

Therefore if $\|a_i\|_2 < \gamma$, we must have $x_i^\star = 0$.

5. *Exercise T5.29.* The Lagrangian is

$$L(x,\nu) = (-3-\nu)x_1^2 + (1-\nu)x_2^2 + (2+\nu)x_3^2 + 2(x_1+x_2+x_3) = \nu.$$

The KKT conditions are

- Primal feasibility: $x_1^2 + x_2^2 + x_3^2 = 1$.
- The gradient of the Lagrangian with respect to x is zero:

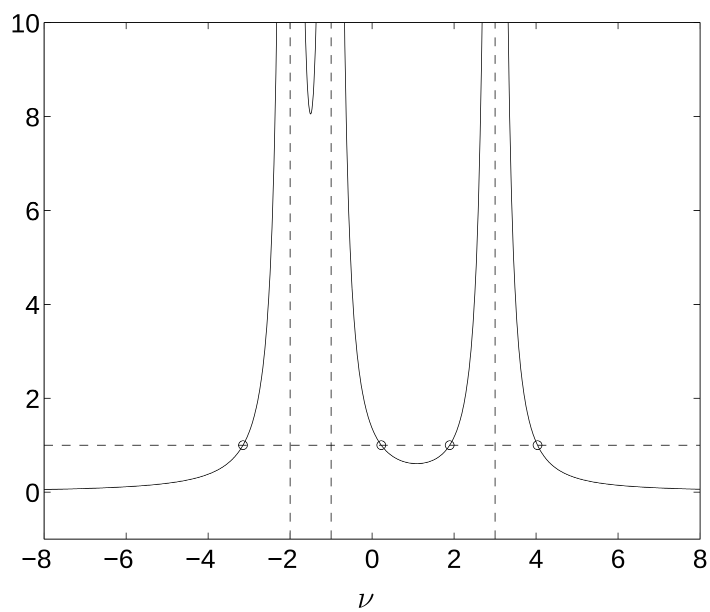
$$(-3+\nu)x_1 + 1 = 0, \qquad (1+\nu)x_2 + 1 = 0, \qquad (2+\nu)x_3 + 1 = 0.$$

Since strong duality holds, these conditions are necessary conditions for optimality of x, ν (see page 243). They are not sufficient because the problem is not convex. Therefore not all solutions of the KKT conditions are necessarily optimal.

A first observation is that the KKT conditions imply $\nu \neq 2$, $\nu \neq -1$, $\nu \neq 3$. We can therefore eliminate x and reduce the KKT conditions to a nonlinear equation in ν:

$$\frac{1}{(-3+\nu)^2} + \frac{1}{(1+\nu)^2} + \frac{1}{(2+\nu)^2} = 1$$

The left-hand side is plotted in the figure.



There are four solutions:

$$\nu = -3.15, \qquad \nu = 0.22, \qquad \nu = 1.89, \qquad \nu = 4.04,$$

corresponding to

$$x = (0.16, 0.47, 0.87), \qquad x = (0.36, -0.82, -0.45),$$

$$x = (0.90, -0.35, -0.26), \qquad x = (-0.97, -0.20, -0.17).$$

$\nu^\star$ is the largest of the four values: $\nu^\star = 4.0352$. This can be seen several ways. The simplest way is to compare the objective values of the four solutions x, which are

$$f_0(x) = 4.64, \qquad f_0(x) = -1.13, \qquad f_0(x) = -1.59, \qquad f_0(x) = -5.33.$$

We can also evaluate the dual objective at the four candidate values for ν.

We can also explain from the theory why the largest ν is the correct one. On page 244 we have seen that for a *convex* problem, the KKT conditions are sufficient for optimality. For a nonconvex problem the argument on page 244 fails because the gradient condition,

$$\nabla_x L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = \nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0,$$

does not necessarily mean that $\tilde{x}$ is a minimizer of $L(x, \tilde{\lambda}, \tilde{\nu})$. Therefore we cannot conclude that $g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$. However, if we replace the last (gradient) condition by

$$\tilde{x} = \operatorname*{argmin}_x L(x, \tilde{\lambda}, \tilde{\nu}),$$

then we can say that for a nonconvex problem these modified KKT conditions are sufficient for optimality. In this exercise L is a quadratic function of x, so $\tilde{x}$ is a minimizer of $L(x, \tilde{\nu})$ if the gradient of L is zero and

$$\nabla^2 L(\tilde{x}, \tilde{\nu}) = \nabla^2 f_0(\tilde{x}) + \tilde{\nu} \nabla^2 f_1(\tilde{x}) \succeq 0.$$

In other words

$$\begin{bmatrix} -3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} + \tilde{\nu} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \succeq 0,$$

and therefore the optimal $\nu^\star$ must be greater than 3.

6. *Exercise T5.17.* The problem can be expressed as

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & f_i(x) \le b_i, \quad i = 1, \ldots, m \end{array}$$

if we define $f_i(x)$ as the optimal value of the LP

$$\begin{array}{ll} \text{maximize} & x^T a \\ \text{subject to} & C_i a \preceq d, \end{array}$$

where a is the variable, and x is treated as a problem parameter. It is readily shown that the Lagrange dual of this LP is given by

$$\begin{array}{ll} \text{minimize} & d_i^T z \\ \text{subject to} & C_i^T z = x \\ & z \succeq 0. \end{array}$$

The optimal value of this LP is also equal to $f_i(x)$, so we have $f_i(x) \leq b_i$ if and only if there exists a z_i with

$$d_i^T z \leq b_i, \qquad C_i^T z_i = x, \qquad z_i \succeq 0.$$

7. *Exercise T5.21 (a)–(c).*

   (a) $p^\star = 1$.

   (b) The Lagrangian is $L(x, y, \lambda) = e^{-x} + \lambda x^2/y$. The dual function is

   $$\begin{aligned} g(\lambda) &= \inf_{x,y>0} (e^{-x} + \lambda x^2/y) \\ &= \begin{cases} 0 & \lambda \geq 0 \\ -\infty & \lambda < 0. \end{cases} \end{aligned}$$

   Therefore the dual problem is

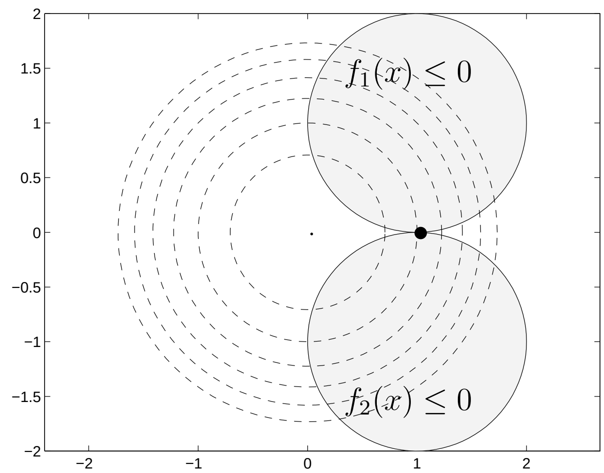
   $$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & \lambda \geq 0, \end{array}$$

   and the optimal value of the dual is $d^\star = 0$.

   (c) Slater's condition is not satisfied.

8. *Exercise T5.26.*

   (a) The figure shows the feasible set (the intersection of the two shaded disks) and some contour lines of the objective function. There is only one feasible point, $(1, 0)$, so it is optimal for the primal problem and we have $p^\star = 1$.



(b) The Lagrangian is

$$
\begin{aligned}
&L(x_1, x_2, \lambda_1, \lambda_2) \\
&\quad = \; x_1^2 + x_2^2 + \lambda_1((x_1-1)^2 + (x_2-1)^2 - 1) + \lambda_2((x_1-1)^2 + (x_2+1)^2 - 1) \\
&\quad = \; (1+\lambda_1+\lambda_2)x_1^2 + (1+\lambda_1+\lambda_2)x_2^2 - 2(\lambda_1+\lambda_2)x_1 - 2(\lambda_1-\lambda_2)x_2 + \lambda_1 + \lambda_2.
\end{aligned}
$$

The KKT conditions are the following.

- x is primal feasible:

$$(x_1-1)^2 + (x_2-1)^2 \le 1, \qquad (x_1-1)^2 + (x_2+1)^2 \le 1.$$

- The multipliers for the inequality constraints are nonnegative: $\lambda_1 \ge 0$, $\lambda_2 \ge 0$.
- Complementary slackness:

$$\lambda_1((x_1-1)^2 + (x_2-1)^2 - 1) = \lambda_2((x_1-1)^2 + (x_2+1)^2 - 1) = 0.$$

- The gradient of the Lagrangian at x is zero:

$$
\begin{aligned}
2x_1 + 2\lambda_1(x_1-1) + 2\lambda_2(x_1-1) &= 0 \qquad (3)\\
2x_2 + 2\lambda_1(x_2-1) + 2\lambda_2(x_2+1) &= 0.
\end{aligned}
$$

At $x = (1, 0)$, there exist no λ_1, λ_2 that satisfy these equations, because (3) requires $2 = 0$ for all λ_1 and λ_2.

(c) The Lagrange dual function is given by

$$g(\lambda_1, \lambda_2) = \inf_{x_1, x_2} L(x_1, x_2, \lambda_1, \lambda_2).$$

L has a minimum at

$$x_1 = \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 + \lambda_2}, \qquad x_2 = \frac{\lambda_1 - \lambda_2}{1 + \lambda_1 + \lambda_2}$$

if $1+\lambda_1+\lambda_2 \geq 0$, and is unbounded below otherwise. Therefore

$$g(\lambda_1,\lambda_2) \;=\; \left\{ \begin{array}{ll} -\frac{(\lambda_1+\lambda_2)^2+(\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2} + \lambda_1+\lambda_2 & 1+\lambda_1+\lambda_2 \geq 0 \\ -\infty & \text{otherwise,} \end{array} \right.$$

where we interpret $a/0 = 0$ if $a=0$ and as $-\infty$ if $a<0$. The dual problem is

$$\begin{array}{ll} \text{maximize} & \dfrac{\lambda_1+\lambda_2-(\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2} \\ \text{subject to} & \lambda_1 \geq 0, \quad \lambda_2 \geq 0. \end{array}$$

Since g is symmetric ($g(\lambda_1,\lambda_2) = g(\lambda_2,\lambda_1)$) and concave, we have

$$g(\lambda_1,\lambda_2) = \frac{1}{2}(g(\lambda_1,\lambda_2)+g(\lambda_2,\lambda_1)) \leq g(\frac{\lambda_1+\lambda_2}{2}, \frac{\lambda_1+\lambda_2}{2})$$

for all λ_1 and λ_2. We can therefore take $\lambda_1=\lambda_2$ in the dual. The dual function

$$g(\lambda_1,\lambda_1) = \frac{2\lambda_1}{1+2\lambda_1}$$

tends to the maximum value of 1 as $\lambda_1=\lambda_2 \to \infty$.

Although we have strong duality ($d^\star = p^\star = 1$), the dual optimum is not attained and therefore the KKT conditions are not solvable.

9. *Exercise A4.3.* The Lagrangian is

$$L(x,z,\mu) = \sum_k x_k \log(x_k/y_k) + b^T z - z^T A x + \mu - \mu \mathbf{1}^T x.$$

Minimizing over x_k gives the conditions

$$1 + \log(x_k/y_k) - a_k^T z - \mu = 0, \quad k=1,\ldots,n,$$

with solution

$$x_k = y_k e^{a_k^T z + \mu - 1}.$$

Plugging this in in L gives the Lagrange dual function

$$g(z,\mu) = b^T z + \mu - \sum_{k=1}^n y_k e^{a_k^T z+\mu-1}$$

and the dual problem

$$\text{maximize} \quad b^T z + \mu - \sum_{k=1}^n y_k e^{a_k^T z+\mu-1}.$$

This can be simplified a bit if we optimize over μ by setting the derivative equal to zero:

$$\mu = 1 - \log \sum_{k=1}^n y_k e^{a_k^T z}.$$

After this simplification the dual problem reduces to the problem in the assignment.

10. *Exercise A4.5.* The Lagrangian is

$$\begin{aligned} L(x,\lambda) &= \frac{1}{2}\|x-a\|_2^2 + \lambda\|x\|_1 - \lambda \\ &= \sum_{k=1}^{n}\left(\frac{(x_k-a_k)^2}{2} + \lambda|x_k|\right) - \lambda. \end{aligned}$$

The Lagrangian is easy to minimize over x because it is separable:

$$\begin{aligned} g_k(\lambda) &= \inf_{x_k}\left(\frac{(x_k-a_k)^2}{2} + \lambda|x_k|\right) \\ &= \begin{cases} -\lambda^2/2 + \lambda|a_k| & \lambda \leq |a_k| \\ a_k^2/2 & \lambda > |a_k|. \end{cases} \end{aligned}$$

This is a differentiable concave function with derivative

$$g_k'(\lambda) = \max\{|a_k| - \lambda, 0\}.$$

The dual problem is

$$\begin{array}{ll} \text{maximize} & g(\lambda) = \sum_k g_k(\lambda) - \lambda \\ \text{subject to} & \lambda \geq 0. \end{array}$$

g is differentiable and concave, with derivative

$$g'(\lambda) = \sum_{k=1}^{n}\max\{|a_k| - \lambda, 0\} - 1.$$

The derivative varies from $\|a\|_1 - 1$ for $\lambda = 0$, to -1 if $\lambda \geq \max|a_k|$. If $\|a\|_1 \leq 1$, then g is decreasing on $\mathbf{R}_+$, the optimal λ is zero, and the optimal x is $x = a$. If $\|a\|_1 > 1$, we can find the optimal λ by solving the piecewise-linear equation

$$\sum_{k=1}^{n}\max\{|a_k| - \lambda, 0\} = 1. \tag{4}$$

From the optimal λ, we obtain the optimal x as follows:

$$x_k = \begin{cases} 0 & \lambda \geq |a_k| \\ a_k - \lambda & \lambda < |a_k|,\ a_k > 0 \\ a_k + \lambda & \lambda < |a_k|,\ a_k < 0. \end{cases} \tag{5}$$

This method can also be derived from the KKT conditions. The conditions are

(a) *Primal feasibility.* $\|x\|_1 \leq 1$.

(b) *Dual feasibility.* $\lambda \geq 0$.

(c) *Complementary slackness.* $\lambda(1 - \|x\|_1) = 0$.

(d) *Minimality of Lagrangian.* The optimal x minimizes

$$L(x, \lambda) = \sum_{k=1}^{n} \left(\frac{(x_k - a_k)^2}{2} + \lambda |x_k| \right) - \lambda.$$

This is equivalent to (5).

To solve these conditions, we first note that (d) implies $|x_k| = \max\{|a_k| - \lambda, 0\}$, so

$$\|x\|_1 = \sum_{k=1}^{n} \max\{|a_k| - \lambda, 0\}.$$

The other three conditions imply that $\lambda = 0$ only if $\|a\|_1 \leq 1$ and $\lambda > 0$ only if it satisfies the equation (4).