Adrian Townsend

MATH 582G, Application Problem

12 March 2008

My final application is a look at norm optimization problems of the form

$$minimize \left\| Ax - b \right\|, A \in \Re^{m \times n} \tag{1}$$

and comes from problems in the text [1] problems 4.11 and 6.4.

4.11: Problems of the form minimize $\|Ax - b\|$

First, consider the general case of the norm of an affine mapping as stated in the section heading. This is certainly a convex problem because norms are convex and the application of an affine mapping function to the domain does not affect convexity. So, begin with the stated problem (1) for any norm of the form $\|\cdot\|_n$ where n is a non-negative integer. Let $r = Ax - b$, where r is commonly referred to as the residual and is of interest in the case when $b \notin range(A)$. A Lagrangian may be defined as:

$$\begin{aligned} L(x, r, \lambda) &= \|r\| + \lambda^\top (r + b - Ax) \\ &= \|r\| + \lambda^\top r - \lambda^\top Ax + \lambda^\top b \\ x, r, \lambda &\in \Re^m. \end{aligned} \tag{2}$$

From here, the dual equation can be found as:

$$\begin{aligned} g(\lambda) &= \inf_{r,x \in \Re^m} L(x, r, \lambda) = \lambda^\top b + \delta_{\{0\}}(\lambda^\top A) + \inf_r (\|r\| + \lambda^\top r) \\ &= \lambda^\top b + \delta_{\{0\}}(\lambda^\top A) + \delta_{\mathbf{B}_*}(\lambda) \qquad (3) \\ &\quad then \\ \sup_\lambda g(\lambda) &= \lambda^\top b \\ \lambda^\top A = \vec{0} &\quad , \quad \|\lambda\|_* \leq 1, \end{aligned} \tag{4}$$

where δ is the indicator function and $\mathbf{B}_*$ is the unit ball on the dual (or polar) of the norm stated in (1).

At this point it is worth commenting the dual of the unit ball. By Holder's inequality, $|\langle x, y \rangle| \leq \|x\|_p \|y\|_q$, where $\frac{1}{p} + \frac{1}{q} = 1$. In these problems the focus is on the L_1 and L_∞ norms, and thus it is useful to note that $\|\cdot\|_\infty = (\|\cdot\|_1)^*$, and $\|\cdot\|_1 = (\|\cdot\|_\infty)^*$.

4.11 a

Problem:

$$minimize \; \|Ax - b\|_\infty \tag{5}$$

Let's define r to be the residual, $r = Ax - b$. This norm effectively minimizes the maximum residual value, $\max r_i$, and thus, it can be rewritten as a linear program to minimize t such that $-t\vec{1} \preceq Ax - b \preceq t\vec{1}$, where $t \in \Re_+$, and $\vec{1}$ is a vector of ones of dimension $m \times 1$. It can be seen by inspection of this problem and (4) that the dual equation for (5) is:

$$\begin{aligned} g(\lambda) &= \lambda^\top + \delta_{\{0\}}(\lambda^\top A) + \delta_{\mathbf{B}_1}(\lambda). \\ \lambda &\in \Re^m \end{aligned} \tag{6}$$

Another interesting result for the dual equation is found when one instead considers applying two dual variables to the equalities surrounding $Ax - b$.

$$\begin{aligned} L(x,t,\lambda,\nu) &= t\vec{1} + \lambda^\top(b - t\vec{1} - Ax) + \nu^\top(Ax - b - \vec{1}t) \\ &= t(\vec{1} + [\lambda + \nu]^\top \vec{1}) + (\nu - \lambda)^\top Ax + (\lambda - \nu)^\top b \\ \lambda, \nu &\in \Re_+ \end{aligned} \tag{7}$$

$$\begin{aligned} g(\lambda,\nu) &= (\lambda - \nu)^\top b + \delta_{\{0\}}[(\lambda - \nu)^\top A] + \delta_{\{1\}}([\lambda + \nu]^\top \vec{1}). \\ \lambda, \nu &\in \Re_+^m \end{aligned} \tag{8}$$

It is not surprising that the two dual equations, (6) and (8) have very similar forms. Probably the most curious and notable difference is that the indicator on the L_1 unit ball in (6) is replaced by the indicator of the L_1 norm of the sum of the dual variables in (8).

4.11 b

Problem:

$$minimize \; \|Ax - b\|_1 \tag{9}$$

$$\lambda \in \Re^m$$

This problem is similar to the previous, except that the L_1 norm results in the minimization of the sum of the absolute value of the residuals, $\|r\|_1$. So, a linear program may be defined to minimize $\sum_{i=1}^{m} t_i$ such that $-t \preceq Ax - b \preceq t$, where $t \in \Re_+^m$. It can be seen by inspection of this problem and (4) that the dual equation for (9) is:

$$g(\lambda) = \lambda^\top + \delta_{\{0\}}(\lambda^\top A) + \delta_{\mathbf{B}_\infty}(\lambda). \tag{10}$$

Similar to problem 4.11a, an interesting result for the dual equation is found when one assignes two dual variables to the inequalities in the linear program.

$$\begin{aligned} L(x,t,\lambda,\nu) &= t^\top \vec{1} + \lambda^\top(b - t - Ax) + \nu^\top(Ax - b - t) \\ &= (\vec{1} - \lambda - \nu)^\top t + (\lambda - \nu)^\top b + (\nu - \lambda)^\top Ax \end{aligned} \tag{11}$$

$$\begin{aligned} g(\lambda,\nu) &= (\lambda - \nu)^\top b + \delta_{\{0\}}\left[(\lambda - \nu)^\top A\right] + \delta_{\{\vec{1}\}}\left[\lambda + \nu\right] \\ \lambda, \nu &\in \Re_+^m \end{aligned} \tag{12}$$

Again, there is an apparent duality between the norm in the problem definition and the resulting constraint on the dual variables. Also, there is similarity between the constraint that the dual variable λ is in the L_∞ dual ball in (10), while in (12) the sum of the dual variables, $\lambda+\nu$, is constrained to be equal to $\vec{1}$.

In the the following three problems from 4.11, the linear program will be stated, followed by the dual equations found two ways as in 4.11 a and b.

4.11 c

Problem:

$$\begin{aligned} minimize \quad & \|Ax-b\|_1 \\ s.t. \quad & \|x\|_\infty \leq 1 \end{aligned} \tag{13}$$

A dual function may be found by a similar method as (4), where an additional dual variable, γ is needed to incorporate the L_∞ constraint on x.

$$\begin{aligned} L(x,r,\lambda,\gamma) &= \|r\|_1 + \gamma(\|x\|_\infty - 1) + \lambda^\top(b - ax + r) \quad (14) \\ \gamma \in R & \qquad r, x, \lambda \in \Re^m \\ g(\lambda,\nu) &= \lambda^\top b - \gamma + \inf_x(-\gamma\|x\|_\infty - \lambda^\top Ax) + \inf_r(\|r\|_1 + \lambda^\top r) \\ &= \lambda^\top b - \gamma - \sup_x \left[\gamma\left(\frac{1}{\gamma}\lambda^\top Ax - \|x\|_\infty\right)\right] + \sup_r(\lambda^\top r - \|r\|_1) \\ g(\lambda,\nu) &= \lambda^\top b - \gamma + \delta_{\mathbf{B}_\infty}(\lambda) - \delta_{\mathbf{B}_1}\left(\frac{1}{\gamma}\lambda^\top A\right) \quad (15) \end{aligned}$$

An equivalent linear program is to add to the linear program defined in 4.11b the constraint: $-\vec{1} \preceq x \preceq \vec{1}$. Then, an alternative Lagrangian and Dual function may be found as follows.

$$\begin{aligned} L(x,t,\lambda,\nu,\gamma,\theta) &= (\vec{1}-\lambda-\nu)^\top t + (\lambda-\nu)^\top b + (\nu-\lambda)^\top Ax - \gamma(x+\vec{1}) + \theta(x-\vec{1}) \quad (16) \\ &= (\vec{1}-\lambda-\nu)^\top t + (\lambda-\nu)^\top b + [(\nu-\lambda)^\top A + (\theta-\gamma)^\top]x - (\theta+\gamma)^\top\vec{1} \\ g(\lambda,\nu,\gamma,\theta) &= (\lambda-\nu)^\top b - (\gamma+\theta)^\top\vec{1} \quad (17) \\ & \quad +\delta_{\{\vec{1}\}}(\lambda+\nu) + \delta_{\{0\}}[(\gamma-\theta)^\top\vec{1} + (\lambda-\nu)^\top A] \\ \lambda,\nu,\gamma,\theta &\in \Re^m_+ \end{aligned}$$

4.11 d

Problem:

$$\begin{aligned} minimize \quad & \|x\|_1 \\ s.t. \quad & \|Ax-b\|_\infty \leq 1 \end{aligned} \tag{18}$$

A dual function may be found by a similar method as (4), which resembles (15).

$$\begin{aligned} L(x,r,\lambda,\nu) &= \|x\|_1 + \gamma(\|r\|_\infty - 1) + \lambda^\top(b - Ax + r) \qquad (19)\\ &= (\|x\|_1 - \lambda^\top Ax) + (\gamma\|r\|_\infty + \lambda^\top r) - \gamma + \lambda^\top b \\ g(\lambda,\gamma) &= \lambda^\top b + \delta_{\mathbf{B}_\infty}(-\lambda^\top A) + \delta_{\mathbf{B}_1}\left(\frac{1}{\gamma}\lambda\right) \qquad (20)\\ \lambda,\gamma &\in \Re^m_+ \end{aligned}$$

An alternate formulation of the Lagrangina and dual function may be found by assigning a dual variable to each inequality that results from expanding the norms in the objective and constraint. The problem can be stated:
minimize $t^\top\vec{1}$ such that $-t \leq x \leq t$ and $-\vec{1} \leq Ax - b \leq \vec{1}$, where $t \in \Re^m_+$.

$$\begin{aligned} L(x,t,\lambda,\nu,\gamma,\theta) &= t^\top\vec{1} - \lambda^\top(b - Ax - \vec{1}) + \nu^\top(Ax - b - \vec{1}) - \gamma^\top(t + x) + \theta^\top(x - t) \qquad (21)\\ &= [\theta^\top - \gamma^\top + (\nu-\lambda)^\top A]x + (\vec{1} - \gamma - \theta)^\top t + (\lambda-\nu)^\top b - (\lambda+\nu)^\top\vec{1} \\ g(\lambda,\nu,\gamma,\theta) &= (\lambda-\nu)^\top b - (\lambda+\nu)^\top\vec{1} \qquad (22)\\ &\quad +\delta_{\{0\}}[\gamma^\top - \theta^\top + (\lambda-\nu)^\top A] + \delta_{\{\vec{1}\}}(\gamma+\theta) \\ \lambda,\nu,\gamma,\theta &\in \Re^m_+ \end{aligned}$$

4.11 e

Problem:

$$\textit{minimize} \quad \|Ax - b\|_1 + \|x\|_\infty \qquad (23)$$

$$(24)$$

Following (4) Lagrangian and Dual equations can be found by introducing a residual variable, r.

$$\begin{aligned} L(x,r,\nu) &= \|r\|_1 + \|x\|_\infty + \lambda^\top(b - Ax + r) \qquad (25)\\ &= (\|r\|_1 + \lambda^\top r) + (\|x\|_\infty + \lambda^\top Ax) + \lambda^\top b \\ &\quad \lambda \in \Re^m \\ g(\nu) &= \lambda^\top b + \delta_{\mathbf{B}_\infty}(\lambda) + \delta_{\mathbf{B}_1}(-\lambda^\top A) \qquad (26) \end{aligned}$$

An alternative approach to finding Lagrangian and Dual equations would be to expand the norms as inequalities and assign an dual variable to each. *minimize* $r^\top\vec{1} + t$, where $-r \leq Ax - b \leq r$ and $-t\vec{1} \leq x \leq t\vec{1}$ for $r \in \Re^m_+$, $t \in \Re_+$.

$$\begin{aligned} L((x,r,t,\lambda,\nu,\gamma,\theta) &= r^t\vec{1} + t + \lambda^\top(b - Ax - r) + \nu^\top(Ax - b - r) \qquad (27)\\ &\quad -\gamma^\top(x + t\vec{1}) + \theta^\top(x - t\vec{1}) \\ &= (\lambda+\nu)^\top b + (\vec{t} - \lambda - \nu)^\top r + t(1 - \gamma^\top\vec{1} - \theta^\top\vec{1}) \\ &\quad +[\theta^\top - \gamma^\top - (\lambda-\nu)^\top A]x \\ g(\lambda,\nu,\gamma,\theta) &= (\lambda-\nu)^\top b \qquad (28)\\ &\quad +\delta_{\{\vec{1}\}}(\lambda+\nu) + \delta_{\{1\}}[(\gamma+\theta)^\top\vec{1}] + \delta_{\{0\}}[A^\top(\lambda-\nu) + \gamma - \theta] \\ &\quad \lambda,\nu,\gamma,\theta \in \Re^m_+ \end{aligned}$$

6.4 Differential Approximation of $\|Ax - b\|_1$

It is natural to want to find extrema by differentiation when approaching a minimization problem. This problem considers a differentiable approximation to (9) which is often used in practice, 29.

$$|u| \approx \Phi(u) = \left(u^2 + \epsilon\right)^{\frac{1}{2}}, \tag{29}$$

where $\epsilon > 0$. Let the residual be $r = Ax - b$ where $r_i = a_i^\top - b_i$, where r_i and b_i are the i^{th} components of r and b and $a_i^\top$ is the i^{th} row of A. Then, the solution of (9) may be approximated as

$$minimize \sum_{i=1}^{m} \Phi(r_i). \tag{30}$$

It is assumed that $A \in \Re^{m \times n}$ and $rank(A) = n$. Let the optimal value of (9) be $p^\star$ and the optimal solution to (30) be $\hat{x}$, for which the residual is $\hat{r} = A\hat{x} - b$.

6.4 a

Find a lower bound for $p^\star$. Show:

$$p^\star \geq \sum_{i=1}^{m} \frac{\hat{r_i}^2}{\left(\hat{r_i}^2 + \epsilon\right)^{\frac{1}{2}}}. \tag{31}$$

Consider a lower bound or a best under-estimator of a function, f. This may be written as: $g(s) = f'(x)(s - x)$. Since L_1 is not differentiable at 0, consider using instead (29). Then, one could approximate a best under-estimator to $\|r_i\|_1$,

$$\begin{aligned} g(r_i) &= \frac{\partial\left[\Phi(\hat{r_i})\right]}{\partial \hat{r_i}}(r_i - \hat{r_i}) \\ g(r_i) &= \frac{\hat{r_i}}{\left(\hat{r_i}^2 + \epsilon\right)^{\frac{1}{2}}}(r_i - \hat{r_i}) \end{aligned} \tag{32}$$

Without loss of generality, one can let $\hat{r_i} = 0$ since (29) and $\|\cdot\|$ both equal 0 at 0. Then, the best-underestimator of (9) using our differentiable approximation to L_1 is given by

$$g(r_i) = \frac{\hat{r_i}^2}{\left(\hat{r_i}^2 + \epsilon\right)^{\frac{1}{2}}} \tag{33}$$

if one can show that the magnitude of the derivative in (32) is less than or equal to the derivative of L_1 at all points other than zero (recall, that we already showed that L_1 and its approximation are equal at zero). This is easy since $\left|\frac{\partial}{\partial r} r\right| = 1$:

$$\begin{aligned} \left|\frac{\hat{r_i}}{\left(\hat{r_i}^2 + \epsilon\right)^{\frac{1}{2}}}\right| &< 1, \\ \forall r_i &\neq 0, \\ \epsilon &> 0. \end{aligned}$$

This result may be more intuitively when inspected graphically. Figure 1 plots the L_1 norm and its approximation.

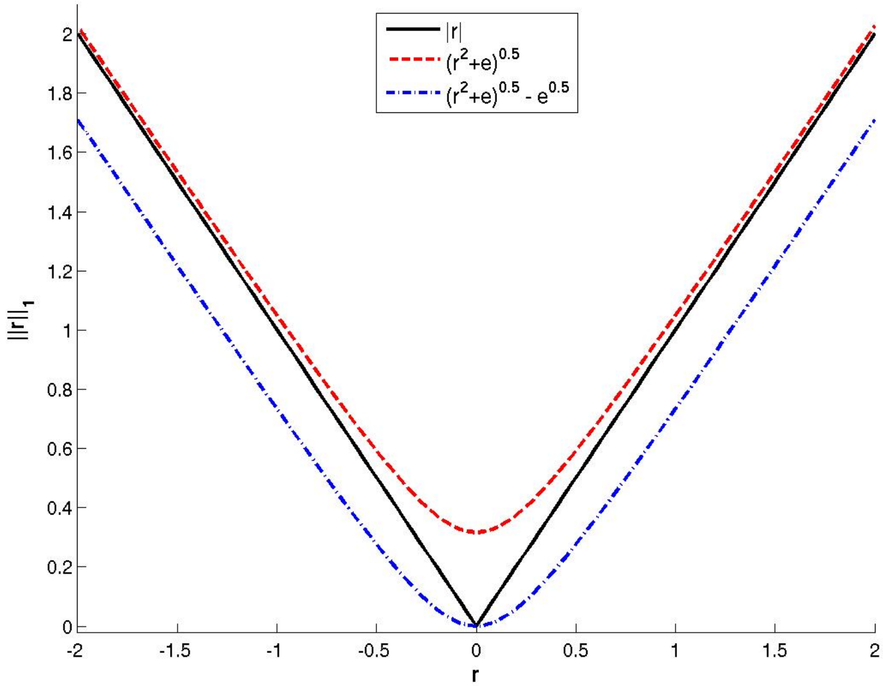

Figure 1: Plots of the L_1 norm, the approximation (29).

6.4 b

The next interesting result may be show by rearranging (31) and adding $|r_i|$'s to each side and noting that: $\|A\hat{x} - b\|_1 = \sum_{i=1}^{m} |\hat{r_i}|$.

$$
\begin{aligned}
\sum_{i=1}^{m} |\hat{r_i}| &\leq p^\star + \sum_{i=1}^{m} |\hat{r_i}| - \sum_{i=1}^{m} \frac{\hat{r_i}^2}{\left(\hat{r_i}^2 + \epsilon\right)^{\frac{1}{2}}} \\
\|A\hat{x} - b\|_1 &\leq p^\star + \sum_{i=1}^{m} \left(|\hat{r_i}| - \frac{|\hat{r_i}||\hat{r_i}|}{\left(\hat{r_i}^2 + \epsilon\right)^{\frac{1}{2}}} \right) \\
\|A\hat{x} - b\|_1 &\leq p^\star + \sum_{i=1}^{m} |\hat{r_i}| \left(1 - \frac{|\hat{r_i}|}{\left(\hat{r_i}^2 + \epsilon\right)^{\frac{1}{2}}} \right) \qquad (34)
\end{aligned}
$$

Once $\hat{x}$ and $\hat{r}$ are ocmputed, then on can use (34) to examine a bound on how sub-optimal $\hat{x}$ is for the L_1 norm approximation problem. The form of (34) is interesting, because the result in the parenthesis on the right hand side may be thought of as the error of the approximation (29).

References

[1] Boyd, S., Vandenberghe, L. "Convex Optimization. "Cambridge University Press, New York, 2004.

[2] Hiriart-Urruty, J., Lemaréchal, C., "Fundamentals of Convex Analysis. "Springer-Verlag Berlin Heidelberg, Germany, 2001