Shrinkage Formulas and Some Necessary Optimality Criteria for L1 problems

Toby Sanders

December 26, 2019

Consider the minimization of the following functional:

$$F(x; a, b, y) = a/2(x-y)^2 + bx + |x|, \tag{1}$$

for $b, y \in \mathbb{R}$ and $a > 0$. Let x^* be the minimizer. Then

$$0 \in \frac{\partial F}{\partial x}(x^*).$$

Cleanest derivation of minimizer

If $x^* \neq 0$, then

$$0 = a(x^* - y) + b + sign(x^*),$$

and so

$$x^* + sign(x^*)/a = y - b/a.$$

Observe this implies

$$sign(x^*) = sign(y - b/a)$$

and

$$|x^*| = |y - b/a| - 1/a.$$

Together these imply

$$x^* = |x^*|sign(x^*) = sign(y - b/a)(|y - b/a| - 1/a). \tag{2}$$

On the other hand, if $x^* = 0$, defining $I = [-1, 1]$, computing the subdifferential one obtains

$$0 \in -ay + b + I$$

implying

$$|y - b/a| < 1/a.$$

Combining this with (2) observe the general formula for the minimizer of F to be

$$x^* = \max(|y - b/a| - 1/a, 0) * sign(y - b/a). \tag{3}$$

More intuitive derivation

Suppose $x^* > 0$. Then one immediately arrives at

$$x^* = y - b/a - 1/a = |y - b/a| - 1/a,$$

where the second inequality follows from using $x^*, a > 0$. Suppose $x^* < 0$. Then similarly one immediately finds

$$x^* = y - b/a + 1/a = sign(y - b/a)(|y - b/a| - 1/a),$$

where the second inequality may be deduced from $a > 0 > x^*$.

Then the proof of formula (3) is complete using similar arguments to the first derivation, i.e. consider $x^* = 0$ and combining with the previous two equalities.

Complex Case

Now consider minimizing

$$f(z) = a/2|z - z_0|^2 + |z|,$$

with $a > 0$, $z, z_0 \in \mathbb{C}$. One may compute the partial derivates and set to zero to obtain

$$\begin{array}{c} f_x(z) = a(x - x_0) + x/|z| = 0 \\ f_y(z) = a(y - y_0) + y/|z| = 0 \end{array},$$

assuming $|z| \neq 0$. Rearranging these equalities yields

$$\begin{array}{c} x = \dfrac{x_0|z|}{|z| + 1/a} \\ y = \dfrac{y_0|z|}{|z| + 1/a} \end{array}. \tag{4}$$

Squaring these equations and adding together yields

$$|z|^2 = \frac{|z_0|^2|z|^2}{(|z| + 1/a)^2},$$

leading to

$$(|z| + 1/a)^2 = |z_0|^2 \ \rightarrow \ |z| = |z_0| - 1/a \tag{5}$$

Observe that this only makes sense if $|z_0| > 1/a$, and in fact the solution is $|z| = 0$ for $|z_0| \leq 1/a$.

To formally argue these cases, just take the left equation in (5) and roll with it:

$$|z_0|^2 = (|z| + 1/a)^2 \geq 1/a^2$$

Substituting (5) into (4), and combining with $|z| = 0$ for $|z_0| \leq 1/a$, we obtain the complex version of the shrinkage formula:

$$z^* = \frac{z_0}{|z_0|} \max(|z_0| - 1/a, 0) = sign(z_0) \max(|z_0| - 1/a, 0) \tag{6}$$

Complex Case with linear term

Finally, let us minimize the real part of

$$F(z) = a/2|z - z_0|^2 + \bar{b}z + |z|,$$

withe $a > 0$, $z, z_0, b \in \mathbb{C}$. The partial derivatives for $|z| \neq 0$ are given by

$$\begin{array}{c} F_x(z) = a(x - x_0) + x/|z| + b_1 \\ F_y(z) = a(y - y_0) + y/|z| + b_2 \end{array}.$$

This leads to

$$\begin{aligned} x &= \frac{(x_0 - b_1/a)|z|}{|z| + 1/a} \\ y &= \frac{(y_0 - b_2/a)|z|}{|z| + 1/a} \end{aligned} \tag{7}$$

Squaring and adding these equations leads to

$$|z|^2 = \frac{|z|^2|z_0 - b/a|^2}{(|z| + 1/a)^2},$$

hence

$$|z| = |z_0 - b/a| - 1/a.$$

Observe again, the solution is $|z| = 0$ for $|z - b/a| \leq 1/a$, and so combining the solution for $|z|$ with (7), we obtain the final compact solution

$$z^* = sign(z_0 - b/a) \max(|z_0 - b/a| - 1/a, 0).$$

Derivatives of Functions of Complex Variables

Let $z = x + iy$ and $f(z) = u(x, y) + iv(x, y)$. We say f is differentiable in the complex plane at z if

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z},$$

exists for all directions Δz. This implies that

$$\lim_{x \to 0} \frac{f(z + x) - f(z)}{x} = u_x + iv_x = \lim_{y \to 0} \frac{f(z + iy) - f(z)}{iy} = -iu_y + v_y,$$

yielding the necessary Cauchy-Riemann equations

$$u_x = v_y, \text{ and } u_y = -v_x.$$

Now suppose we want to perform something like gradient decent to find the minimizer of a function $f(z)$. Should we care about this definition of differentiability.... not really in my opinion and experience.

First observe that f should be real valued for it to make sense to find a minimizer. Then, one can just treat z as two variables, hence a gradient decent can take the form

$$z^{k+1} = z^k - \tau^k (f_x(z^k) + i f_y(z^k)).$$

Notice since f is real valued, so are f_x and f_y. So all we are doing is moving in the z-plane with respect to the steepest decent for the two variables, but it is simply written as a single variable z, i.e. we can just think of $f(z)$ as $u(x, y)$, and everything works the same as in vector calculus.

The same arguments easily extend to a multidimensional complex variable, $z \in \mathbb{C}^N$. Now let's take f to be

$$f(\vec{z}) = f(\vec{x} + i\vec{y}) = \|Az - c\|_2^2,$$

with $A \in \mathbb{C}^{m \times N}$ and $c \in \mathbb{C}^m$. Then as before the gradient decent should take the form

$$z^{k+1} = z^k - \tau^k (\nabla_x f(z^k) + i \nabla_y f(z^k)) \tag{8}$$

Now, it is a tedious task for me, but one can show the following

$$\begin{aligned} \nabla_x f(z) &= 2Re(A^T \overline{(Az - c)}) \\ \nabla_y f(z) &= -2Im(A^T \overline{(Az - c)}) \end{aligned}. \tag{9}$$

This leads to the gradient term in (8) taking the form

$$\nabla_x f(z) + i \nabla_y f(z) = 2A^H (Az - c)$$

Hooray!

Optimality conditions for L1

For the augmented Lagrangian optimization problem for L1, we end up with the objective function to minimize over u and w:

$$L(u, w; \sigma, \mu) = \frac{\mu}{2}\|Au - b\|_2^2 + \|w\|_1 + \frac{\beta}{2}\|Tu - w\|_2^2 - \sigma^H (Tu - w).$$

The first order optimality conditions read

$$\begin{aligned} &\mu A^H (Au - b) + \beta T^H (Tu - w) - T^H \sigma = 0 \\ &\beta(w - Tu) + \sigma \in -sign^*(w). \end{aligned}$$

Now, if we have solved the original problem we desired, the optimality condition reads

$$\mu A^H (Au - b) + T^H sign^*(Tu) \in 0. \tag{10}$$

In each of these cases, $sign^*$ is denoting the sub-differential for the absolute value. Let $v = Tu$, let $S = \{j \,|\, v_j \neq 0\}$, $R = S^C$, and T_S denote T containing only the rows from S (similarly for R). Then

$$T^H sign^*(Tu) = T_S^H sign(T_S u) + T_R^H sign^*(T_R u).$$

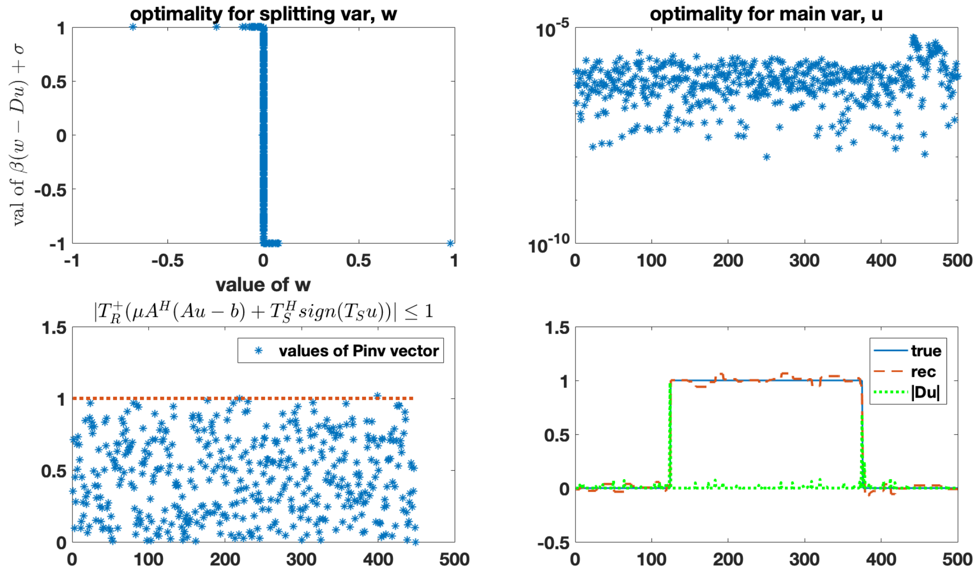

Figure 1: Optimality conditions for the augmented Lagrangian function is shown in the top. Optimality condition for original problem is shown in the bottom left.

Substituting this into (10) and rearranging (assuming $R \neq \varnothing$) yields

$$\left|(T_R^H)^+ \left(\mu A^H(Au-b) + T_S^H sign(T_S u)\right)\right| \leq 1$$

The only minor challenge in checking this condition for general problems is computing the pseudo inverse, or at least the multiplication by the pseudo inverse. To do this efficiently, first write $T_R = P_R T$, where P_R is the row selector, and recall for circulant matrices that $T = F^{-1}\Lambda_T F$. Here, F is the unitary Fourier transform and Λ_T is diagonal containing the eigenvalues of T (DFT of the first column of T). Therefore $T_R^+ = F^{-1}\Lambda_T^+ F P_R^T$. Similar arguments may be used for T_S, hence the equivalent condition (for computational purposes) may be written as

$$|F^{-1}\Lambda_T^+ F P_R^T(\mu A^H(Au-b) + T^H P_S^T sign(P_S T u))| \leq 1.$$

This is shown in the bottom left of the provided figure. The case when $R = \varnothing$ simply reduces (10) to

$$\mu A^H(Au-b) + T^H sign(Tu) = 0.$$