Supplementary Notes for Math 4/896, Section 006

Thomas Shores
Department of Mathematics
University of Nebraska
Spring 2006

Contents

1. Basics of Tikhonov Regularization

This section is in reference to Section 5.1 of the text. Here are some points of clarification. Consider the following constrained optimization problem:

(COPT) Minimize the function $f(\mathbf{x})$ over all $x \in \Omega$, the **feasible set** which is defined by $\Omega = \{\mathbf{x} \mid g(\mathbf{x}) \leq c\}$.

First, two theorems from optimization theory:

Theorem 1.1. *Let $\mathbf{x}^* \in \mathbb{R}^n$ be a local solution to (COPT), where f, g are smooth and $\nabla g(\mathbf{x}^*) \neq \mathbf{0}$. Then $(\mathbf{x}^*, \lambda^*)$ is a stationary point of the Lagrange functional*

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x})$$

with $\lambda^* \geq 0$ and $\lambda^* > 0$ exactly when $\mathbf{x}^*$ satisfies the equality constraint $g(\mathbf{x}^*) = c$.

Theorem 1.2. *If $f(\mathbf{x})$ and the the feasible set Ω are convex with f, g smooth, then every local solution to (COPT) is a global solution.*

We observe that positive semi-definite quadratic functionals like

$$f(\mathbf{m}) = \|G\mathbf{m} - \mathbf{d}\|^2 + \alpha^2 \|L(\mathbf{m} - \mathbf{m}_0)\|^2$$

are convex and the domains of positive semi-definite quadratic constraints such as $\|G\mathbf{m} - \mathbf{d}\|^2 \leq \delta$ or $\|\mathbf{m}\|^2 \leq \epsilon$ are convex.

Now here is our definition of Tikhonov regularization.

Definition 1.3. A Tikhonov regularization of the ill-posed problem

$$G(\mathbf{m}) = \mathbf{d}$$

is a problem of minimizing with respect to $\mathbf{m}$ a functional

$$\|G(\mathbf{m}) - \mathbf{d}\|^2 + \alpha^2 \|L(\mathbf{m})\|^2$$

where $\mathbf{m}, \mathbf{d}$ are vectors in a suitable space and the functional L is such that for every positive α the above problem is well posed, and for $\alpha = 0$ the resulting problem is the least squares problem associated with the equation $G(\mathbf{m}) = \mathbf{d}$.

This is the so-called *variational form* of Tikhonov regularization. What it does not attempt to account for is a strategy for choosing the regularization parameter α. This is a separate issue. We know of several methods, e.g., locating the corner of the L-curve, GCV and the discrepancy principle.

Note: strictly speaking, the TSVD and TGSVD methods that we have encountered are *not* Tikhonov regularizations, although they are regularization strategies and can be described in terms of filters, an umbrella that covers Tikhonov regularization of various orders as well. In using these methods, rather than choose a specific α one has to decide what singular values and vectors to discard in computing a solution to the regularized problem by way of SVD or GSVD.

Next consider the constrained optimization problem of minimizing $f(\mathbf{m}) = \|\mathbf{m}\|$ subject to the constraint $g(\mathbf{m}) = \|G\mathbf{m} - \mathbf{d}\| \leq \delta$. Equivalently, minimize $f(\mathbf{m}) = \|\mathbf{m}\|^2$ subject to the constraint $g(\mathbf{m}) = \|G\mathbf{m} - \mathbf{d}\|^2 \leq \delta^2$. Theorems 1.1 and 1.2 yield that a solution to the problem satisfies the constraint and equation

$$\mathbf{m} + \lambda \left(G^T G\mathbf{m} - G^T \mathbf{d}\right) = 0.$$

Note that the case $\lambda = 0$ implies that $\mathbf{m} = \mathbf{0}$, which in turn implies that $\|G\mathbf{0} - \mathbf{d}\| = \|\mathbf{d}\| \leq \delta$. Now for regularization we would normally take δ to be a measure of the error in the data. Thus this inequality is saying that the data is essentially all error, hence worthless, and there is no point in working the problem. Therefore, the only sensible possibility is that $\lambda > 0$. Set $\alpha^2 = 1/\lambda$ and obtain the Tikhonov regularization problem that is derived from minimizing the functional

$$\|G\mathbf{m} - \mathbf{d}\|^2 + \alpha^2 \|\mathbf{m}\|^2 .$$

Of course, as we have noted above, a solution $\mathbf{m}_\alpha$ to this problem assumes a definite value of α rather than determining it. The suitable value of α is determined by noting that $\lambda > 0$ means that the equality constraint is satisfied, that is,

$$\|G\mathbf{m}_\alpha - \mathbf{d}\| = \delta.$$

It is this extra condition that is used to determine α. What we have just described is none other than the discrepancy principle!

Finally, let's consider the constrained optimization problem of minimizing $f(\mathbf{m}) = \|G\mathbf{m} - \mathbf{d}\|$ subject to the constraint $g(\mathbf{m}) = \|\mathbf{m}\| \leq \epsilon$. Equivalently, minimize $f(\mathbf{m}) = \|G\mathbf{m} - \mathbf{d}\|^2$ subject to the constraint $g(\mathbf{m}) = \|\mathbf{m}\| \leq \epsilon$. Observe that if $G\mathbf{m} = \mathbf{d}$ actually has a solution which satisfies the constraint, then there is no regularization involved here at all. So the only situations of interest are those in which no exact solution satisfies the constraint. Therefore, we have to form a Lagrange functional, calculate its gradient and set it equal to zero. This yields

$$\|G\mathbf{m} - \mathbf{d}\|^2 + \alpha^2 \|\mathbf{m}\|^2$$

where we took the Lagrange multiplier to be $\lambda = \alpha^2$, since we know it will be positive. The condition $\|\mathbf{m}\| = \epsilon$ is used to determine α in this case. By itself, this does not provide a particularly useful criterion for choosing α. The point here is that the constrained optimization problem leads to a Tikhonov regularization formula.

2. Significant Digits

Here is the discussion that appeared in the midterm:

Suppose that you are given a nonzero number in scientific notation, say

$$x_A = d_1.d_2d_3d_4d_5\ldots \times 10^m$$

and you are told that this number is only accurate to about 3 digits. This means roughly that

$$|x_A - x_T| \approx 0.00e_4e_5\ldots \times 10^m = e_4.e_5\ldots \times 10^{m-3}.$$

Thus, we can say that the relative error is

$$\frac{|x_A - x_T|}{|x_T|} = \frac{e_4.e_5\ldots \times 10^{m-3}}{d_1.d_2d_3d_4d_5\ldots \times 10^m} \approx 10^{-3},$$

which gives us a nice rule of thumb: if a datum has only n significant digits, then the difference between this number and the true value it is supposed to represent is a relative error of about 10^{-n} and, conversely, if the relative error in an approximation is about 10^{-n} then the approximating number has about n significant digits. Now look again at Equation (4.91) on page 66 of the text and you can see the connection between singular values and significant digits.

This line of thought gives a crude estimate. The fact is that there are more precise definitions that are commonly used. One definition goes like this: x_A approximates x_T to m significant digits if the the absolute error $|x_A - x_T|$ at most 5 in the $(m+1)$th digit, counting to the right from the first nonzero digit of x_T. Thus, 4995 approximates 5000 to 3 significant digits, but 4994 ony approximates 5000 to 2 significant digits.

Numerical analysts prefer a continuous version of significant digits for purposes of error analysis, so the customary definition is given in terms of relative error: x_A approximates x_T to m significant digits if

$$\frac{|x_A - x_T|}{|x_T|} \leq 5 \times 10^{-(m+1)}.$$

By this definition, 4996 would approximate 5000 to 2 significant digits since $4/5000 = 0.0008$. (BTW, some authors use m in place of $m+1$.)