Sergey Voronin
Department of Mathematics
Tufts University

December 7, 2016

HPC — Homework 2

(A) Recall as before that the *gradient* of a function $g : \mathbb{R}^n \to \mathbb{R}$ at a point $x \in \mathbb{R}^n$, if exists, is the unique vector v that satisfies

$$g(x+d) = g(x) + v^T d + o(\|d\|), \quad \forall\, d \in \mathbb{R}^n.$$

Note for example that in three dimensions with $\vec{x} = x\hat{i} + y\hat{j} + z\hat{k}$, one has that $\nabla g(\vec{x}) = \frac{\partial g}{\partial x}\hat{i} + \frac{\partial g}{\partial y}\hat{j} + \frac{\partial g}{\partial z}\hat{k}$. Notice that related to but different from the gradient is the Jacobian of $g(\vec{x})$. It is a matrix given by $J_{ij}[g(\vec{x})] = \frac{\partial g_i}{\partial x_j}$.

We like to copmpute the gradient of the functional:

$$F(x) = \|Ax - b\|^2 + \lambda\|x\|^2 + \lambda_2\|Lx\|^2 \tag{1}$$

The function is convex and has a unique global minimizer. This minimizer is obtained by setting $\nabla F(x) = 0$.

First, notice that for any vector function $h(x)$ and $f(x) = \|h(x)\|_2^2$,
we have that $\nabla f(x) = 2J[h(x)]^T h(x)$ where $J[h(x)]$ is the Jacobian of the function $h(x)$.

Recall also, that for the linear mapping $y(x) = Mx$, $y_i = \sum_{k=1}^n M_{ik}x_k$. It follows that $\frac{\partial y_i}{\partial x_j} = M_{ij}$ from which we deduce that the Jacobian of y is $J[y] = M$.

Using these results, the following computations immediately follow:

$$\begin{aligned}
&\nabla\|x\|_2^2 = 2x \\
&\nabla\|Lx\|_2^2 = 2L^T Lx \\
&\nabla\|Ax - b\|_2^2 = 2A^T(Ax - b)
\end{aligned}$$

Computing the gradient of (1),

$$\nabla F(x) = 2A^T(Ax - b) + 2\lambda x + 2\lambda_2 L^T Lx = 0 \implies (A^T A + \lambda I + \lambda_2 L^T L)x = A^T b$$

(B) The effect of Tikhonov regularization can be easily observed by setting $\lambda_2 = 0$. In that case, the minimizer satisfies the simple system $(A^T A + \lambda I)x = A^T b$. Plugging in the full SVD of

$A = USV^T$, we obtain:

$$
\begin{aligned}
x_{\text{tik}} &= \left(VS^2V^T + \lambda I\right)^{-1} VSU^Tb \\
&= \left(VS^2V^T + \lambda VV^T\right)^{-1} VSU^Tb \\
&= \left(V\left(S^2 + \lambda I\right)V^T\right)^{-1} VSU^Tb \\
&= \left(V\left(S^2 + \lambda I\right)^{-1} V^T\right) VSU^Tb \\
&= V\left(S^2 + \lambda I\right)^{-1} SU^Tb \\
&= V\text{diag}\left(\frac{\sigma_i}{\sigma_i^2 + \lambda}\right) U^Tb,
\end{aligned}
$$

where we have used properties of inverses and orthogonal matrices U and V. Notice that the generalized inverse solution $x_{\text{gen}} = A^+b$ is given by:
$x_{\text{gen}} = VS^{-1}U^Tb = V\text{diag}\left(\frac{1}{\sigma_i}\right)U^Tb$. It is thus immediately apparent that the Tikhonov regularized solution filters the effect of small singular values on the solution (and in particular, prevents the appearance of large terms in the diagonal matrix which can multiply small errors in the right hand side vector b).

Written by Sergey Voronin on December 7, 2016.