Algebraic Solution for Quadratic and Tikhonov-Regularized Forms

Sean Borman

July 18 2004

1 Preliminaries

1.1 Differentiable Scalar Functions

Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a scalar function of the vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^\top$. The gradient vector is the vector of partial derivatives,

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}. \tag{1}$$

1.2 The Gradient of a Scalar Product

Let $f = \boldsymbol{a}^\top \boldsymbol{x}$ where $\boldsymbol{a}$ and $\boldsymbol{x}$ are n-vectors, then $\nabla f(\boldsymbol{x}) = \boldsymbol{a}$. This is seen by expanding the inner product $f(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x}$ as $f(\boldsymbol{x}) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$, and then differentiating to find $\nabla f(\boldsymbol{x}) = [a_1, a_2, \ldots, a_n]^\top = \boldsymbol{a}$.

Let $f = \boldsymbol{y}^\top \boldsymbol{A} \boldsymbol{x}$ with m-vector $\boldsymbol{y}$ and $m \times n$ matrix $\boldsymbol{A}$, then $\nabla f(\boldsymbol{x}) = \boldsymbol{A}^\top \boldsymbol{y}$. This may be seen by noting that $\boldsymbol{y}^\top \boldsymbol{A}$ is an n-vector and applying the result for the scalar product above.

1.3 The Gradient of a Quadratic Form

If $f = \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}$ with n-vector $\boldsymbol{x}$ and real-symmetric $n \times n$ matrix $\boldsymbol{A}$, then $\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}$. We can show this by writing f in terms of the on-diagonal and off-diagonal products,

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} x_j \\ &= \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{i \neq j}^{n} \sum^{n} a_{ij} x_i x_j, \end{aligned}$$

then differentiating f with respect to x_k for $k \in \{1, 2, \ldots, n\}$ yields,

$$\begin{aligned} \frac{\partial f(\boldsymbol{x})}{x_k} &= 2a_{kk} x_k + 2 \sum_{i \neq k} a_{ki} x_i \\ &= 2 \sum_{i=1}^{n} a_{ki} x_i, \end{aligned}$$

so that

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}.$$

2 Minimization of Quadratic Forms

Here we consider the problem of finding $\hat{\boldsymbol{x}}$ which minimizes the objective function with quadratic form,

$$f(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2,$$

with $m \times n$ real matrix $\boldsymbol{A}$. This may be rewritten as,

$$\begin{aligned} f(\boldsymbol{x}) &= (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) \\ &= (\boldsymbol{y}^\top - \boldsymbol{x}^\top \boldsymbol{A}^\top) \cdot (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) \\ &= \boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} \\ &= \boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}. \end{aligned}$$

Noting that $\boldsymbol{A}^\top A$ is real-symmetric and differentiating the objective function with respect to $\boldsymbol{x}$ yields the gradient vector,

$$\nabla f(\boldsymbol{x}) = -2\boldsymbol{A}^\top \boldsymbol{y} + 2\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}$$

A necessary requirement for $\hat{\boldsymbol{x}}$ to be a minimum of $f(\boldsymbol{x})$ is that $\nabla f(\hat{\boldsymbol{x}}) = 0$. In this case we have that,

$$\boldsymbol{A}^\top \boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^\top \boldsymbol{y}$$

and assuming that $\boldsymbol{A}^\top \boldsymbol{A}$ is invertible,

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{y}.$$

The expression $(\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top$ is known variously as the generalized inverse, pseudoinverse, or Moore-Penrose inverse of $\boldsymbol{A}$.

3 Optimization for Tikhonov Regularized Problems

In Tikhonov regularized solutions to inverse problems we are faced with the problem of finding $\hat{\boldsymbol{x}}$ which minimizes the objective function containing a quadratic form and regularizing functional,

$$f(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|^2 + \lambda \|\boldsymbol{x}\|^2,$$

where $\lambda > 0$ is the regularization parameter. This may be rewritten as,

$$\begin{aligned} f(\boldsymbol{x}) &= (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \lambda \boldsymbol{x}^\top \boldsymbol{x} \\ &= (\boldsymbol{y}^\top - \boldsymbol{x}^\top \boldsymbol{A}^\top)(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}) + \lambda \boldsymbol{x}^\top \boldsymbol{x} \\ &= \boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{x} \\ &= \boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \lambda \boldsymbol{x}^\top \boldsymbol{x}. \end{aligned}$$

Noting that $\boldsymbol{A}^\top A$ is real-symmetric and differentiating the objective function with respect to $\boldsymbol{x}$ yields the gradient vector,

$$\begin{aligned} \nabla f(\boldsymbol{x}) &= -2\boldsymbol{A}^\top \boldsymbol{y} + 2\boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + 2\lambda \boldsymbol{x} \\ &= 2\boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) + 2\lambda \boldsymbol{x} \end{aligned}$$

A necessary requirement for $\hat{\boldsymbol{x}}$ to be a minimum of $f(\boldsymbol{x})$ is that $\nabla f(\hat{\boldsymbol{x}}) = 0$. In this case we have that,

$$(\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I})\, \hat{\boldsymbol{x}} = \boldsymbol{A}^\top \boldsymbol{y}$$

and assuming that $\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I}$ is invertible,

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}^\top \boldsymbol{A} + \lambda \boldsymbol{I})^{-1} \boldsymbol{A}^\top \boldsymbol{y}.$$