Calculating Derivatives for a Least Squares Matrix Function

Yin Zhang, CAAM 454/554

Rice University, Houston, TX 77005

March 19, 2013

Given a symmetric positive semidefinite matrix $A \in \mathbb{R}^{n\times n}$, let

$$R(X) = XX^T - A, \tag{1}$$

where $X \in \mathbb{R}^{n\times k}$ for $k < n$. Consider the nonlinear least squares problem

$$\min_{X\in\mathbb{R}^{n\times k}} f(X) := \frac{1}{4}\|R(X)\|_F^2, \tag{2}$$

which should give the best rank-k approximation to A.

We know that the gradient of f has the form

$$\nabla f(X) = \frac{1}{2}\mathbf{J}(X)^T(R(X)) \ \in \ \mathbb{R}^{n\times k}, \tag{3}$$

where $\mathbf{J}(X)^T$ is the adjoint of the linear operator $\mathbf{J}(X) : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times n}$, i.e., the Jacobian of $R(X)$.

Since

$$R(X+S) = (X+S)(X+S)^T - A = R(X) + SX^T + XS^T + SS^T,$$

$\mathbf{J}(X)$ is clearly defined by

$$\mathbf{J}(X)(S) = SX^T + XS^T = SX^T + (SX^T)^T. \tag{4}$$

Let $\mathcal{P}_n$ be the permutation in $\mathbb{R}^{n^2}$ so that

$$\mathbf{vec}(M^T) = \mathcal{P}_n\mathbf{vec}(M)$$

for all $M \in \mathbb{R}^{n\times n}$. It is known that

$$\mathcal{P}_n = \mathcal{P}_n^T = \mathcal{P}_n^{-1}.$$

In view of the identity

$$\mathbf{vec}((SX^T)^T) = \mathcal{P}_n \mathbf{vec}(SX^T),$$

it is easy to derive from (4) that

$$\mathbf{vec}(\mathbf{J}(X)(S)) = \mathbf{vec}(SX^T) + \mathbf{vec}((SX^T)^T) = (I + \mathcal{P}_n)(X \otimes I)\mathbf{vec}\, S.$$

Hence the matrix representation of $\mathbf{J}(X)$ is

$$J(X) = (I + \mathcal{P}_n)(X \otimes I) \ \in \ \mathbb{R}^{n^2 \times nk} \tag{5}$$

For any symmetric matrix $R \in \mathbb{R}^{n \times n}$, we calculate

$$\begin{aligned} J(X)^T \mathbf{vec}(R) &= (X^T \otimes I)(I + \mathcal{P}_n)\mathbf{vec}(R) \\ &= (X^T \otimes I)(\mathbf{vec}(R) + \mathbf{vec}(R^T)) \\ &= 2(X^T \otimes I)\mathbf{vec}(R) \\ &= 2\mathbf{vec}(RX), \end{aligned}$$

that is, for symmetric R,

$$\mathbf{J}(X)^T(R) = 2RX. \tag{6}$$

Hence,

$$\nabla f(X) = \frac{1}{2}\mathbf{J}(X)^T(R(X)) = R(X)X = X(X^TX) - AX. \tag{7}$$

If we do Gauss-Newton method, we should examine

$$\begin{aligned} J(X)^T J(X) &= (X \otimes I)^T (I + \mathcal{P}_n)^2 (X \otimes I) \\ &= 2(X^T \otimes I)(I + \mathcal{P}_n)(X \otimes I) \\ &= 2(X^TX \otimes I) + 2(X^T \otimes I)\mathcal{P}_n(X \otimes I). \end{aligned}$$

Unfortunetely, this matrix is not easily invertible, even though the first term is,

$$(X^TX \otimes I)^{-1} = (X^TX)^{-1} \otimes I.$$

By substituting $\mathbf{J}(X)S$ in (4) into (6) for R, we obtain

$$\mathbf{J}(X)^T\mathbf{J}(X)(S) = 2(SX^TX + XS^TX). \tag{8}$$

This expression can be used to solve the normal equations by an iterative method.