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

$$R(X) = XX^T - A,$$

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

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

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

We know that the gradient of $f$ has the form

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

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

Since

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

$J(X)$ is clearly defined by

$$J(X)(S) = SX^T + XS^T = SX^T + (SX^T)^T.$$

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

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

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
\[ \text{vec}((SX^T)^T) = \mathcal{P}_n \text{vec}(SX^T), \]
it is easy to derive from (4) that
\[ \text{vec}(J(X)(S)) = \text{vec}(SX^T) + \text{vec}((SX^T)^T) = (I + \mathcal{P}_n)(X \otimes I)\text{vec} S. \]

Hence the matrix representation of \( J(X) \) 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
\[ J(X)^T \text{vec}(R) = (X^T \otimes I)(I + \mathcal{P}_n)\text{vec}(R) = (X^T \otimes I)(\text{vec}(R) + \text{vec}(R^T)) = 2(X^T \otimes I)\text{vec}(R) = 2\text{vec}(RX), \]
that is, for symmetric \( R \),
\[ J(X)^T(R) = 2RX. \tag{6} \]
Hence,
\[ \nabla f(X) = \frac{1}{2} J(X)^T(R(X)) = R(X)X = X(X^TX) - AX. \tag{7} \]

If we do Gauss-Newton method, we should examine
\[ J(X)^T J(X) = (X \otimes I)^T (I + \mathcal{P}_n)^2 (X \otimes I) = 2(X^T \otimes I)(I + \mathcal{P}_n)(X \otimes I) = 2(X^TX \otimes I) + 2(X^T \otimes I)\mathcal{P}_n(X \otimes I). \]

Unfortunately, this matrix is not easily invertible, even though the first term is,
\[ (X^TX \otimes I)^{-1} = (X^TX)^{-1} \otimes I. \]

By substituting \( J(X)S \) in (4) into (6) for \( R \), we obtain
\[ J(X)^T J(X)(S) = 2(SX^TX + XS^TX). \tag{8} \]

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