4TE3/6TE3

Algorithms for Continuous Optimization

(Solving sparse systems with low-rank-update)

Tamás TERLAKY
Computing and Software
McMaster University

Hamilton, November 2005

`terlaky@mcmaster.ca` `Tel: 27780`

Solving Linear Systems

with Low-rank Updates

Consider the following system of linear equations with the unknown vector $p \in \mathbb{R}^n$,

$$(Q + RS^T)p = q, \tag{1}$$

where Q is an $n \times n$ matrix, R and S are $n \times k$ matrices and $q \in \mathbb{R}^n$ are given.

Assumptions:

1. The matrix Q is nonsingular.
2. The rank of both R and S are equal to k ($\leq n$).
3. The matrix $Q + RS^T$ is nonsingular.
4. The matrix Q is sparse, RS^T is dense and $k \leq n$.

Assumptions 1–4 guarantee that the system (1) is solvable efficiently.

As it is intuitively clear, solving a linear system

$$Qp = q, \tag{2}$$

where the sparse matrix Q is the coefficient matrix is computationally cheap, while directly solving equation system (1) with the dense coefficient matrix is computationally expensive.

The procedure

Goal: To utilize the sparsity of the matrix Q. First step we reformulate (1) as follows

$$Qp = q - RS^T p = q - Rz.$$

For a moment neglecting the fact that the vector $z = S^T p$ depends on the unknown p we can decompose the solution process into the following steps.

Step 1. Solve the sparse system $Qp_1 = q$.

Step 2. Solve the sparse system $QU = R$, where U is an $n \times k$ matrix of unknowns.

Step 3. Now we consider how to find an appropriate vector z. By the definition of z we have

$$z = S^T p = S^T(p_1 - Uz)$$

which is equivalent to SOLVE

$$(I + S^T U)z = S^T p_1. \tag{3}$$

Step 4. Let U be the $n \times k$ matrix with column vectors u_j. Then we have the solution

$$p = p_1 - Uz.$$

Discussion

Step 2 of the above procedure involves the solution of $k+1$ sparse linear systems, all with the same sparse coefficient matrix Q, hence these can be solved easily.

Below we will verify that the k-dimensional linear equation system (3) has a unique solution and that the above sketched way the linear equation system (1) can indeed be solved efficiently.

Before doing that, let us make a simple estimation of the computational complexity under Assumption 4. By a direct approach the linear system (1) with a dense coefficient matrix can be solved in $\mathcal{O}(n^3)$ arithmetic operations. Assume that the matrix Q is sparse and so the equation system $Qp_1 = q$ can be solved in $\rho < \mathcal{O}(n^3)$ arithmetic operations. We have to solve $k+1$ such systems and a small dense system with k unknowns. Thus the total complexity becomes $(k+1)\rho+\mathcal{O}(k^3)$. Note that, in many applications matrix Q is either (multi)diagonal, or block-diagonal with small diagonal blocks. As a consequence one has $\rho = \mathcal{O}(n)$. In many of these cases $k \leq \mathcal{O}(\sqrt{n})$ then the total computational complexity becomes $\mathcal{O}(n\sqrt{n})$, which is a factor $n\sqrt{n}$ better than the direct approach.

Verification

As it is proposed above the linear system (1) is solved by solving $k+1$ linear equation systems with the same coefficient matrix Q and a small $k \times k$ linear system as follows.

Let $p_1 \in \mathbb{R}^n$ be the solution of the linear system

$$Qp_1 = q, \tag{4}$$

$u_j \in \mathbb{R}^n$ be the solution of the linear system

$$Qu_j = r_j \tag{5}$$

for $j = 1, \cdots k$, where r_j denotes the j-th column of matrix R and finally, let $z \in \mathbb{R}^k$ be the solution of the linear system

$$(I + S^T U)z = S^T p_1 \tag{6}$$

where I denotes the k-dimensional identity matrix and $U = [u_1, \cdots, u_k]$. We prove the following theorem.

Theorem 1 *If Assumptions 1–3 hold then the unique solution of the linear system (1) is given by*

$$p = p_1 - Uz.$$

Proof

First note that by Assumption 3 the solution p exists and it is unique. By Assumption 1 the equations (4) and (5) have unique solutions. Further, by Assumption 2 the vectors $r_j, \quad j = 1, \cdots, k$ are linearly independent and then, by Assumption 1, the solution vectors $u_j, \quad j = 1, \cdots, k$ of (5) are linearly independent as well, i.e. rank$(U) = k$.

Now by proving that the coefficient matrix $I + S^T U$ of equations (6) is nonsingular we verify that equation (6) has a unique solution. Assume to the contrary that there is a nonzero vector $w \in \mathbb{R}^k$ such that

$$(I + S^T U)w = 0,$$

or equivalently $S^T U w = -w$. Then, by multiplying the nonsingular matrix $Q + RS^T$ by the nonzero vector Uw and using (5) we have

$$(Q + RS^T)Uw = (QU)w + R(S^T U w) = Rw - Rw$$

yielding a contradiction.

Finally by simple calculations the reader easily verifies that $(Q + RS^T)(p_1 - Uz) = q$ which completes the proof. □

The Sherman-Morrison formula

$$(Q+RS^T)^{-1} = Q^{-1} - Q^{-1}R(I+S^TQ^{-1}R)^{-1}S^TQ^{-1} \qquad (7)$$

- Having proved the correctness of our procedure the Sherman-Morrison formula (7) can be derived from our procedure. This can be done as follows. By definition – from (4) we have $p_1 = Q^{-1}q$, from (5) we have $U = Q^{-1}R$ and from (6) we have $z = (I+S^TU)^{-1}S^Tp_1$ – the solution p is given by

  $$\begin{aligned} p = p_1 - Uz &= Q^{-1}q - Q^{-1}R(I+S^TU)^{-1}S^Tp_1 \\ &= (Q^{-1} - Q^{-1}R(I+S^TQ^{-1}R)^{-1}S^TQ^{-1})q. \end{aligned}$$

  By Assumption 3 we can write $p = (Q+RS^T)^{-1}q$. Comparing this with the above expression and observing that Theorem 1 holds for all right-hand-side vector $q \in \mathbb{R}^n$ the Sherman-Morrison formula (7) follows.

- On the other hand, one can derive our procedure by carefully analyzing the Sherman-Morrison formula (7). Using again that the solution of (1) can be written as $p = (Q+RS^T)^{-1}q$ we have

  $$p = (Q^{-1} - Q^{-1}R(I+S^TQ^{-1}R)^{-1}S^TQ^{-1})q.$$

  Here all the "inverse matrix – vector" products have to be replaced by the solution of a linear equation system. Thus we have the expression $Q^{-1}q$ which is equivalent to the solution of (4) resulting in p_1; the expression $Q^{-1}R$ which is equivalent to the solution of (5) for each $j = 1, \cdots k$ resulting in U; and finally, having these done, the expression

  $$(I+S^TQ^{-1}R)^{-1}S^TQ^{-1}q = (I+S^TU)^{-1}S^Tp_1$$

  which is equivalent to the solution of (6) resulting in z.