Updating Matrix Inverse

Arian Maleki

August 9, 2018

1 Block Matrix Inversion

For the moment I assume that we are working with LASSO.

Adding Columns Suppose that $A = X_S^\top X_S$ and we have access to A^{-1}. Our goal is to calculate F^{-1}, where

$$F = X_{S\cup T}^\top X_{S\cup T} = \begin{pmatrix} X_S^\top \\ X_T^\top \end{pmatrix} \begin{pmatrix} X_S & X_T \end{pmatrix} = \begin{pmatrix} \overbrace{X_S^\top X_S}^{A} & \overbrace{X_S^\top X_T}^{B} \\ \underbrace{X_T^\top X_S}_{C} & \underbrace{X_T^\top X_T}_{D} \end{pmatrix}.$$

Hence

$$F^{-1} = \begin{pmatrix} A^{-1} + A^{-1}BECA^{-1} & -A^{-1}BE \\ -ECA^{-1} & E \end{pmatrix},$$

where $E = (D - CA^{-1}B)^{-1}$. Note that E is hopefully a small matrix. Hence, this inverse is easy. The rest of the calculations are straightforward.

Removing Columns Now suppose that the inverse of $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is given, how should we calculate the inverse of A? First note that if

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix},$$

then $T_{22} = (D - CA^{-1}B)^{-1}$. Also, $T_{12} = -A^{-1}BT_{22}$ so $-A^{-1}B = T_{12}T_{22}^{-1}$. Then,

$$T_{11} = A^{-1} + A^{-1}BT_{22}CA^{-1} = A^{-1} + (-T_{12})(-T_{22}^{-1}T_{21}).$$

Thus,

$$A^{-1} = T_{11} - T_{12}T_{22}^{-1}T_{21}.$$

2 Approximation

Now let's discuss the following: $\left(X_S^\top \operatorname{diag}[\ddot{\ell}(\hat{\beta}_\lambda)] X_S + (1-\alpha)\lambda I_S \right)^{-1}$ is given, where S is the active set for the pair (α, λ). We are interested in $\left(X_{\tilde{S}}^\top \operatorname{diag}[\ddot{\ell}(\hat{\beta}_{\tilde{\lambda}})] X_{\tilde{S}} + (1-\alpha)\tilde{\lambda} I_{\tilde{S}} \right)^{-1}$, where $\tilde{\lambda}$ is very close to λ. Hence S and $\tilde{S}$ are close.

Now

$$
\begin{aligned}
F &= \left(X_{\tilde{S}}^\top \operatorname{diag}[\ddot{\ell}(\hat{\beta}_{\tilde{\lambda}})] X_{\tilde{S}} + (1-\alpha)\tilde{\lambda} I_{\tilde{S}} \right) \\
&= \left(\underbrace{X_{\tilde{S}}^\top \operatorname{diag}[\ddot{\ell}(\hat{\beta}_{\tilde{\lambda}})] X_{\tilde{S}} + (1-\alpha)\tilde{\lambda} I_{\tilde{S}}}_{A} + X_{\tilde{S}}^\top \overbrace{\Delta}^{\operatorname{diag}[\ddot{\ell}(\hat{\beta}_{\tilde{\lambda}})] - \operatorname{diag}[\ddot{\ell}(\hat{\beta}_\lambda)]} X_{\tilde{S}} + (1-\alpha)(\tilde{\lambda} - \lambda) I_{\tilde{S}} \right) \\
&= A \left(I + A^{-1}\left[X_{\tilde{S}}^\top \Delta X_{\tilde{S}} + (1-\alpha)(\tilde{\lambda} - \lambda) I_{\tilde{S}} \right] \right).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
F^{-1} &= \left(I + A^{-1}\left[X_{\tilde{S}}^\top \Delta X_{\tilde{S}} + (1-\alpha)(\tilde{\lambda} - \lambda) I_{\tilde{S}} \right] \right)^{-1} A^{-1} \\
&\approx A^{-1} + A^{-1}\left[X_{\tilde{S}}^\top \Delta X_{\tilde{S}} + (1-\alpha)(\tilde{\lambda} - \lambda) I_{\tilde{S}} \right] A^{-1}
\end{aligned}
$$

by Taylor expansion.

Issue 1 When we change λ from a large value to a small value in multiple steps. At every step our approximation introduces extra new error. This means that the error may accumulate and our apporximation may become worse as we proceed.

Idea 1 There are iterative algorithms for finding inverse for instance, if I wnat to find the inverse of A I can run the folling iteration:

$$
V_{k+1} = V_k + \underbrace{\alpha}_{\text{step size}} (I - AV_k).
$$

So, ideally for $\tilde{\lambda}$, after obtaining a reasonably good estimation of F^{-1}, we can run one or two iterations of

$$
V_0 = \hat{F}^{-1} \to V_{k+1} = V_k + \alpha(I - FV_k).
$$

The only challenge is how we should pick α? In general, all the eigenvalues of $I - \alpha F$ should be between 0 and 1. However, I am not sure how people set this step size in practice. We should do some research.

Some thoughts on this: let

$$
V_{k+1} = A^{-1} + \overbrace{E_{k+1}}^{\text{error}},
$$

then $AV_{k+1} = I + AE_{k+1}$. Also

$$AV_{k+1} = AV_k + \alpha(A - A^2 V_k)$$
$$= AV_k + \alpha[A - A(I + AE_k)]$$
$$= I + AE_k - \alpha A^2 E_k.$$

Hence,

$$AE_{k+1} = A(I - \alpha A)E_k \implies E_{k+1} = (I - \alpha A)E_k.$$

Take the spectral decomposition $A = Q\Delta Q^\top$:

$$Q^\top E_{k+1} = (I - \alpha\Delta)Q^\top E_k.$$

Note that $\|Q^\top E_{k+1}\|_F = \|E_{k+1}\|_F$. So I think a good value of α should decrease the Frobenius norm. Can we choose α that minimizes the Frobenius norm, and is it a good idea?

Newton-Schulz Iteration   Starts with $V_0 = \hat{F}^{-1}$, we iterates through

$$V_{k+1} = \frac{1}{2}V_k \left(3I - V_k^2\right).$$

This method is quadratically convergent if $\|I - V_0\|_2 < 1$. For Hermitian matrix this means the spectral radius $\rho(V_0) < \sqrt{2}$.

Idea 2   If the above idea does not work, then maybe we should do full inversion every now and then to avoid the error accumulation. For instance we should use approximation technique for 10 different values of λ and then do full inversion for the 11-th λ so on and so forth.

Issue 2   If we have the inverse of

$$A = \left(X_S^\top \operatorname{diag}[\ddot{\ell}(\hat{\beta}_\lambda)]X_S + (1 - \alpha)\lambda I_S\right),$$

can we easily calculate the inverse of

$$\left(X_{\tilde{S}}^\top \operatorname{diag}[\ddot{\ell}(\hat{\beta}_{\tilde{\lambda}})]X_{\tilde{S}} + (1 - \alpha)\tilde{\lambda} I_{\tilde{S}}\right)?$$

The answer is affirmative. Suppose $\tilde{S} \supseteq S$, i.e. $\tilde{S} = S \cup T$. Then

$$\left(X_{\tilde{S}}^\top \operatorname{diag}[\ddot{\ell}(\hat{\beta}_{\tilde{\lambda}})]X_{\tilde{S}} + (1 - \alpha)\tilde{\lambda} I_{\tilde{S}}\right) = \begin{pmatrix} X_S^\top \operatorname{diag}[\ddot{\ell}_S]X_S + (1 - \alpha)\lambda I_S & X_T^\top \operatorname{diag}[\ddot{\ell}_S]X_S \\ X_S^\top \operatorname{diag}[\ddot{\ell}_T]X_T & X_T^\top \operatorname{diag}[\ddot{\ell}_T]X_T + (1 - \alpha)\lambda I_T \end{pmatrix}.$$

Then, we can use the tricks of block matrices to do the inversion.