Updating Matrix Inverse

Arian Maleki

August 15, 2018

1 Block Matrix Inversion

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

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

$$\boldsymbol{F} = \boldsymbol{X}_{S\cup T}^\top \boldsymbol{X}_{S\cup T} = \begin{pmatrix} \boldsymbol{X}_S^\top \\ \boldsymbol{X}_T^\top \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_S & \boldsymbol{X}_T \end{pmatrix} = \begin{pmatrix} \overbrace{\underbrace{\begin{matrix}\boldsymbol{X}_S^\top \boldsymbol{X}_S \\ \boldsymbol{X}_T^\top \boldsymbol{X}_S\end{matrix}}_{\boldsymbol{C}}}^{\boldsymbol{A}} & \overbrace{\underbrace{\begin{matrix}\boldsymbol{X}_S^\top \boldsymbol{X}_T \\ \boldsymbol{X}_T^\top \boldsymbol{X}_T\end{matrix}}_{\boldsymbol{D}}}^{\boldsymbol{B}} \end{pmatrix}.$$

Hence

$$\boldsymbol{F}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{E}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{E} \\ -\boldsymbol{E}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{E} \end{pmatrix},$$

where $\boldsymbol{E} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$. Note that $\boldsymbol{E}$ 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 $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}$ is given, how should we calculate the inverse of $\boldsymbol{A}$? First note that if

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

then $\boldsymbol{T}_{22} = (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}$. Also, $\boldsymbol{T}_{12} = -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{T}_{22}$ so $-\boldsymbol{A}^{-1}\boldsymbol{B} = \boldsymbol{T}_{12}\boldsymbol{T}_{22}^{-1}$. Then,

$$\boldsymbol{T}_{11} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{T}_{22}\boldsymbol{C}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1} + (-\boldsymbol{T}_{12})(-\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{21}).$$

Thus,

$$\boldsymbol{A}^{-1} = \boldsymbol{T}_{11} - \boldsymbol{T}_{12}\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{21}.$$

2 Approximation

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

Now

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

Hence,

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

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 want to find the inverse of $\boldsymbol{A}$ I can run the folling iteration:

$$
\boldsymbol{V}_{k+1} = \boldsymbol{V}_k + \underbrace{\alpha}_{\text{step size}}(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{V}_k).
$$

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

$$
\boldsymbol{V}_0 = \hat{\boldsymbol{F}}^{-1} \to \boldsymbol{V}_{k+1} = \boldsymbol{V}_k + \alpha(\boldsymbol{I} - \boldsymbol{F}\boldsymbol{V}_k).
$$

The only challenge is how we should pick α? In general, all the eigenvalues of $\boldsymbol{I} - \alpha\boldsymbol{F}$ 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

$$
\boldsymbol{V}_{k+1} = \boldsymbol{A}^{-1} + \overbrace{\boldsymbol{E}_{k+1}}^{\text{error}},
$$

then $\boldsymbol{A}\boldsymbol{V}_{k+1} = \boldsymbol{I} + \boldsymbol{A}\boldsymbol{E}_{k+1}$. Also

$$\begin{aligned}\boldsymbol{A}\boldsymbol{V}_{k+1} &= \boldsymbol{A}\boldsymbol{V}_k + \alpha(\boldsymbol{A} - \boldsymbol{A}^2\boldsymbol{V}_k) \\ &= \boldsymbol{A}\boldsymbol{V}_k + \alpha[\boldsymbol{A} - \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{E}_k)] \\ &= \boldsymbol{I} + \boldsymbol{A}\boldsymbol{E}_k - \alpha\boldsymbol{A}^2\boldsymbol{E}_k.\end{aligned}$$

Hence,

$$\boldsymbol{A}\boldsymbol{E}_{k+1} = \boldsymbol{A}(\boldsymbol{I} - \alpha\boldsymbol{A})\boldsymbol{E}_k \implies \boldsymbol{E}_{k+1} = (\boldsymbol{I} - \alpha\boldsymbol{A})\boldsymbol{E}_k.$$

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

$$\boldsymbol{Q}^\top\boldsymbol{E}_{k+1} = (\boldsymbol{I} - \alpha\Delta)\boldsymbol{Q}^\top\boldsymbol{E}_k.$$

Note that $\|\boldsymbol{Q}^\top\boldsymbol{E}_{k+1}\|_F = \|\boldsymbol{E}_{k+1}\|_F$. 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 $\boldsymbol{V}_0 = \hat{\boldsymbol{A}}^{-1}$, we iterates through

$$\boldsymbol{V}_{k+1} = \boldsymbol{V}_k\left(2\boldsymbol{I} - \boldsymbol{A}\boldsymbol{V}_k\right).$$

This method is quadratically convergent if $\|\boldsymbol{I} - \boldsymbol{V}_0\|_2 < 1$. For Hermitian matrix this means the spectral radius $\rho(\boldsymbol{V}_0) < \sqrt{2}$.

Chebyshev-Sen-Prabhu Iteration Starts with $\boldsymbol{V}_0 = \boldsymbol{A}/\|\boldsymbol{A}\|_F^2$, we iterates through

$$\boldsymbol{V}_{k+1} = \boldsymbol{V}_k\left[3\boldsymbol{I} - \boldsymbol{F}\boldsymbol{V}_k(3\boldsymbol{I} - \boldsymbol{A}\boldsymbol{V}_k)\right].$$

This method is cubically convergent.

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

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

can we easily calculate the inverse of

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

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

$$\left(\boldsymbol{X}_{\tilde{S}}^\top \operatorname{diag}[\ddot{\ell}(\hat{\boldsymbol{\beta}}_\lambda)]\boldsymbol{X}_{\tilde{S}} + (1-\alpha)\tilde{\lambda}\boldsymbol{I}_{\tilde{S}}\right) = \begin{pmatrix}\boldsymbol{X}_S^\top \operatorname{diag}[\ddot{\ell}_S]\boldsymbol{X}_S + (1-\alpha)\lambda\boldsymbol{I}_S & \boldsymbol{X}_T^\top \operatorname{diag}[\ddot{\ell}_S]\boldsymbol{X}_S \\ \boldsymbol{X}_S^\top \operatorname{diag}[\ddot{\ell}_T]\boldsymbol{X}_T & \boldsymbol{X}_T^\top \operatorname{diag}[\ddot{\ell}_T]\boldsymbol{X}_T + (1-\alpha)\lambda\boldsymbol{I}_T\end{pmatrix}.$$

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