report slyvester equation

sicheng4

February 2024

Main

This section contains main theorems and proofs, as well as some errors in original paper.

Proposition 2.1

Let $\boldsymbol{S_U U}_{d+1} = \boldsymbol{Q}_{\boldsymbol{U},d+1}\boldsymbol{T}_{\boldsymbol{U},d+1}$ be a reduced QR decomposition with

$$\boldsymbol{Q}_{\boldsymbol{U},d+1} = \left[\boldsymbol{Q}_{\boldsymbol{U},d}, Q_{\boldsymbol{U},d+1}\right] \text{ and } \boldsymbol{T}_{\boldsymbol{U},d+1} = \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d} & T_{H,d+1} \\ 0^\top & \boldsymbol{\tau}_{d+1} \end{bmatrix}$$

Then, for the sketched method, the following Arnoldi-like formula holds:

$$\boldsymbol{S_U A U}_d = \boldsymbol{S_U U}_d \left(\boldsymbol{H}_d + R_H E_d^\top\right) + Q_{\boldsymbol{U},d+1}\boldsymbol{\tau}_{d+1}\boldsymbol{h}_{d+1,d}E_d^\top$$

with $R_H = \boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_H \boldsymbol{h}_{d+1,d}$ and $Q_{\boldsymbol{U},d+1} \perp \boldsymbol{S_U U}_d$. Similarly, if

$$\boldsymbol{S_V V}_{d+1} = \boldsymbol{Q}_{V,d+1}\boldsymbol{T}_{V,d+1}, \quad \boldsymbol{T}_{V,d+1} = \begin{bmatrix} \boldsymbol{T}_{V,d} & T_{G,d+1} \\ 0^\top & \boldsymbol{\theta}_{d+1} \end{bmatrix}$$

then

$$\boldsymbol{S_V B}^\top \boldsymbol{V}_d = \boldsymbol{S_V V}_d \left(\boldsymbol{G}_d + R_G E_d^\top\right) + Q_{V,d+1}\boldsymbol{\theta}_{d+1}\boldsymbol{g}_{d+1,d}E_d^\top,$$

where $R_G := \boldsymbol{T}_{V,d}^{-1} T_G \boldsymbol{g}_{d+1,d}$.

Proof

$$\begin{aligned}
\boldsymbol{S_U A U}_d &= \boldsymbol{S_U U}_{d+1}\underline{\boldsymbol{H}}_d = \boldsymbol{S_U U}_{d+1}\underline{\boldsymbol{H}}_d = \boldsymbol{S_U U}_{d+1}\left[\boldsymbol{H}_d; H^\top\right] \\
&= \boldsymbol{Q}_{\boldsymbol{U},d+1}\boldsymbol{T}_{\boldsymbol{U},d+1}\left[\boldsymbol{H}_d; H^\top\right] \\
&= \left[\boldsymbol{Q}_{\boldsymbol{U},d}, Q_{\boldsymbol{U},d+1}\right] \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d} & T_{H,d+1} \\ 0^\top & \boldsymbol{\tau}_{d+1} \end{bmatrix} \left[\boldsymbol{H}_d; H^\top\right] \\
&= \begin{bmatrix} \boldsymbol{Q}_{\boldsymbol{U},d}\boldsymbol{T}_{\boldsymbol{U},d} & \boldsymbol{Q}_{\boldsymbol{U},d}T_{H,d+1} + Q_{\boldsymbol{U},d+1}\boldsymbol{\tau}_{d+1} \end{bmatrix} \left[\boldsymbol{H}_d; H^\top\right] \\
&= \boldsymbol{Q}_{\boldsymbol{U},d}\boldsymbol{T}_{\boldsymbol{U},d}\boldsymbol{H}_d + \boldsymbol{Q}_{\boldsymbol{U},d}T_{H,d+1}H^T + Q_{\boldsymbol{U},d+1}\boldsymbol{\tau}_{d+1}H^T \\
&= \boldsymbol{Q}_{\boldsymbol{U},d}\boldsymbol{T}_{\boldsymbol{U},d}(\boldsymbol{H}_d + \boldsymbol{T}_{\boldsymbol{U},d}^{-1}T_{H,d+1}H^T) + Q_{\boldsymbol{U},d+1}\boldsymbol{\tau}_{d+1}\boldsymbol{h}_{d+1,d}E_d^\top \\
&= \boldsymbol{S_U U}_d\left(\boldsymbol{H}_d + \boldsymbol{T}_{\boldsymbol{U},d}^{-1}T_H\boldsymbol{h}_{d+1,d}E_d^\top\right) + Q_{\boldsymbol{U},d+1}\boldsymbol{\tau}_{d+1}\boldsymbol{h}_{d+1,d}E_d^\top \\
&= \boldsymbol{S_U U}_d\left(\boldsymbol{H}_d + R_H E_d^\top\right) + Q_{\boldsymbol{U},d+1}\boldsymbol{\tau}_{d+1}\boldsymbol{h}_{d+1,d}E_d^\top
\end{aligned}$$

where letting $R_H = \boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_H \boldsymbol{h}_{d+1,d}$ and $Q_{\boldsymbol{U},d+1} \perp \boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{U}_d$ The same process holds for $\boldsymbol{S}_{\boldsymbol{V}}$

The second part is the verification of whitened-sketched Arnoldi relations. The bases are changed to

$$\widehat{\boldsymbol{U}}_d := \boldsymbol{U}_d \boldsymbol{T}_{\boldsymbol{U},d}^{-1}, \quad \widehat{\boldsymbol{V}}_d := \boldsymbol{V}_d \boldsymbol{T}_{V,d}^{-1}$$

. Again, we will only show the $\widehat{\boldsymbol{U}}_d$, since the other one could be derived by same process. By last proposition, we have

$$\boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{A} \boldsymbol{U}_d = \boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{U}_d \left(\boldsymbol{H}_d + \boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_H \boldsymbol{h}_{d+1,d} E_d^\top \right) + Q_{\boldsymbol{U},d+1} \boldsymbol{\tau}_{d+1} \boldsymbol{h}_{d+1,d} E_d^\top$$

Times $\boldsymbol{T}_{U,d}^{-1}$ on both side, and just do simple calculations. Note that,

$$\boldsymbol{T}_{\boldsymbol{U},d}^{-1} = \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d-1}^{-1} & -\boldsymbol{T}_{\boldsymbol{U},d-1}^{-1} T_{H,d-1} \boldsymbol{\tau}_d^{-1} \\ & -\boldsymbol{\tau}_d^{-1} \end{bmatrix}$$

This gives us:

$$\begin{aligned}
\boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{A} \boldsymbol{U}_d \boldsymbol{T}_{U,d}^{-1} &= \boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{U}_d \left(\boldsymbol{H}_d + \boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_H \boldsymbol{h}_{d+1,d} E_d^\top \right) \boldsymbol{T}_{U,d}^{-1} + Q_{\boldsymbol{U},d+1} \boldsymbol{\tau}_{d+1} \boldsymbol{h}_{d+1,d} E_d^\top \boldsymbol{T}_{U,d}^{-1} \\
&= \boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{U}_d \boldsymbol{T}_{U,d}^{-1} \boldsymbol{T}_{U,d} \left(\boldsymbol{H}_d + \boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_H \boldsymbol{h}_{d+1,d} E_d^\top \right) \boldsymbol{T}_{U,d}^{-1} + Q_{\boldsymbol{U},d+1} \boldsymbol{\tau}_{d+1} \boldsymbol{h}_{d+1,d} E_d^\top \boldsymbol{T}_{U,d}^{-1} \\
&= \boldsymbol{S}_{\boldsymbol{U}} \widehat{\boldsymbol{U}}_d \left(\widehat{\boldsymbol{H}}_d + \widehat{H} E_d^\top \right) + Q_{\boldsymbol{U},d+1} \boldsymbol{\tau}_{d+1} \boldsymbol{h}_{d+1,d} E_d^\top \boldsymbol{T}_{U,d}^{-1} \\
&= \boldsymbol{S}_{\boldsymbol{U}} \widehat{\boldsymbol{U}}_d \left(\widehat{\boldsymbol{H}}_d + \widehat{H} E_d^\top \right) + Q_{\boldsymbol{U},d+1} \boldsymbol{\tau}_{d+1} \boldsymbol{h}_{d+1,d} \boldsymbol{\tau}_d^{-1} \boldsymbol{E}_d^\top
\end{aligned}$$

where

$$\widehat{\boldsymbol{H}}_d = \boldsymbol{T}_{\boldsymbol{U},d} \boldsymbol{H}_d \boldsymbol{T}_{\boldsymbol{U},d}^{-1}, \widehat{H} = T_{H,d+1} \boldsymbol{h}_{d+1,d} \boldsymbol{\tau}_d^{-1}$$

Question1 ??? Note here, the last term is different with the item in paper. I could not figure out why the last term should be $Q_{\boldsymbol{U},d+1} \boldsymbol{h}_{d+1,d} E_d^\top$

Typos on AL1

1. Updae $\widehat{\boldsymbol{H}}_{d+1}$ At d-th iteration, we could get $\widehat{\boldsymbol{H}}_{d+1}$ which would be used in next iteration as$\widehat{\boldsymbol{H}}_d$. Note, $\boldsymbol{T}_{\boldsymbol{U},d+1}^{-1} = \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d}^{-1} & -\boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_{H,d+1}\boldsymbol{\tau}_{d+1}^{-1} \\ & \boldsymbol{\tau}_{d+1}^{-1} \end{bmatrix}$

   By $\widehat{\boldsymbol{H}}_{d+1} = \boldsymbol{T}_{\boldsymbol{U},d+1}\boldsymbol{H}_{d+1}\boldsymbol{T}_{\boldsymbol{U},d+1}^{-1}$, we could imply

$$
\begin{aligned}
\widehat{\boldsymbol{H}}_{d+1} &= \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d} & T_{H,d+1} \\ & \boldsymbol{\tau}_{d+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{H}_{d+1} & H \\ 0\ldots 0, \boldsymbol{h}_{d+1,d} & \boldsymbol{h}_{d+1,d+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d}^{-1} & -\boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_{H,d+1}\boldsymbol{\tau}_{d+1}^{-1} \\ & \boldsymbol{\tau}_{d+1}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d}\boldsymbol{H}_{d+1} + T_{H,d+1}\boldsymbol{h}_{d+1,d}\boldsymbol{E}_d^\top & \boldsymbol{T}_{\boldsymbol{U},d}H + T_{H,d+1}\boldsymbol{h}_{d+1,d+1} \\ \boldsymbol{\tau}_{d+1}\boldsymbol{h}_{d+1,d}\boldsymbol{E}_d^\top & \boldsymbol{\tau}_{d+1}\boldsymbol{h}_{d+1,d+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d}^{-1} & -\boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_{H,d+1}\boldsymbol{\tau}_{d+1}^{-1} \\ & \boldsymbol{\tau}_{d+1}^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \widehat{\boldsymbol{H}}_d + T_H\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}E_d^\top & \widehat{H}_{\text{new}} \\ \boldsymbol{\tau}_{d+1}\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}E_d^\top & \boldsymbol{\tau}_{d+1}\left(-\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}E_d^\top T_{H,d+1} + \boldsymbol{h}_{d+1,d+1}\right)\boldsymbol{\tau}_{d+1}^{-1} \end{bmatrix}
\end{aligned}
$$

   where $\widehat{H}_{\text{new}} = -\left(\widehat{\boldsymbol{H}}_d + T_H\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}E_d^\top\right)T_{H,d+1}\boldsymbol{\tau}_{d+1}^{-1} + \boldsymbol{T}_{\boldsymbol{U},d}H\boldsymbol{\tau}_{d+1}^{-1} + T_{H,d+1}\boldsymbol{h}_{d+1,d+1}\boldsymbol{\tau}_{d+1}^{-1}$ and $H = \left[\boldsymbol{h}_{1,d+1}^\top, \ldots, \boldsymbol{h}_{d,d+1}^\top\right]^\top \in R^{dr\times r}$. A corresponding update for $\widehat{\boldsymbol{G}}_{d+1}$ is performed.

2. $\boldsymbol{E}_d = [zeros((d-1)*r, r); eye(r)]$ and $\boldsymbol{E}_1 = [eye(r); zeros((d-1)*r, r)]$

I could run the code, but the solution is too "bad", the residue satisfies tolerance while the solution NOT

Report on 12/2

I have tried AL1 and AL2 , as well as the direct computation to get $\boldsymbol{H}$.

Parameter

There are some parameters play important roles in AL.

1. **k** $k \neq 1$, otherwise the first for loop would be fail.
2. **s** The dimension of sketched matrix can not be too small. $dr \leq s$
3. **p** A larger p will lead to shorter running time

If we do direct computation instead of updating $\boldsymbol{H}$, and let the tol be 10^{-6}, the accuracy would be also lie in it. However, if we use a tighter tol, the accuracy tends to be worse. More specifically, it seems not converges well. The following is the my experiment results with direction computation of H and $s = 400$ and $p = 10, n1 = n2 = 1200$:

tol	residue
1e-06	3.1566e-08
1e-07	3.214e-09
1e-08	4.2937
1e-09	4.2937
1e-10	4.2937

I think the possible reason could be the limitation of s. Intutively, when you require a higher tol, the iteration should be longer. However, once s is determined, the max number of iterations could not be higher than s/r. So, I let s to be larger . I let $n1 = n2 = 120$, and tried to let $s = 100$, and 120, the accuracy is bad, while $s = 60, 80$ are good, and $s = 20, 40$ are getting worse again. This gives us hints that s should be determined very carefully
The problem may also arise because of the accuracy of lyap().

Algorithm

Algorithm 1 Sketched-and-truncated Arnoldi method for Sylvester equations

Require: $\boldsymbol{A} \in R^{n_1\times n_1}, \boldsymbol{B} \in R^{n_2\times n_2}, C_1 \in R^{n_1\times r}, C_2 \in R^{n_2\times r}, \boldsymbol{S_U} \in R^{s\times n_1}, \boldsymbol{S_V} \in R^{s\times n_2}$, integers $0 < k \leq \text{maxit} \ll \min\{n_1, n_2\}$, $\text{tol} > 0, p \geq 1$

Ensure: $\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}$ such that $\boldsymbol{X}^{(1)}(\boldsymbol{X}^{(2)})^\top = \boldsymbol{X}_d$ approximately solves $\boldsymbol{AX} + \boldsymbol{XB} = C_1 C_2^\top$

Compute skinny QRs: $U_1\ell = C_1, V_1 s = C_2, \boldsymbol{Q}_{\boldsymbol{U},1}\beta_1 = \boldsymbol{S_U}C_1, \boldsymbol{Q}_{V,1}\beta_2 = \boldsymbol{S_U}C_2$, and set $\boldsymbol{T}_{\boldsymbol{U},1} = \beta_1, \boldsymbol{T}_{V,1} = \beta_2$
for $d = 1, \ldots, \text{maxit}$ **do**
    Compute $\widetilde{U} = \boldsymbol{A}U_d, \widetilde{V} = \boldsymbol{B}^\top V_d$
    for $i = \max\{1, d-k+1\}, \ldots, d$ **do**
        Set $\widetilde{U} = \widetilde{U} - U_i\boldsymbol{h}_{i,d}$ with $\boldsymbol{h}_{i,d} = U_i^\top \widetilde{U}$
        Set $\widetilde{V} = \widetilde{V} - V_i\boldsymbol{g}_{i,d}$ with $\boldsymbol{g}_{i,d} = V_i^\top \widetilde{V}$
    end for
    Compute skinny QRs: $U_{d+1}\boldsymbol{h}_{d+1,d} = \widetilde{U}$ and $V_{d+1}\boldsymbol{g}_{d+1,d} = \widetilde{V}$
    Update QRs: $\boldsymbol{Q}_{\boldsymbol{U},d+1}\boldsymbol{T}_{\boldsymbol{U},d+1} = \boldsymbol{S_U}[\boldsymbol{U}_d, U_{d+1}]$, $\boldsymbol{Q}_{V,d+1}\boldsymbol{T}_{V,d+1} = \boldsymbol{S_V}[\boldsymbol{V}_d, V_{d+1}]$
    Update $\widehat{\boldsymbol{H}}_d = \boldsymbol{T}_{\boldsymbol{U},d}\boldsymbol{H}_d\boldsymbol{T}_{\boldsymbol{U},d}^{-1}, \widehat{H} = T_{H,d+1}\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}$,
$\widehat{\boldsymbol{G}}_d = \boldsymbol{T}_{V,d}\boldsymbol{G}_d\boldsymbol{T}_{V,d}^{-1}, \widehat{G} = T_{G,d+1}\boldsymbol{g}_{d+1,d}\boldsymbol{\theta}_d^{-1}$
    if $\text{mod}(d,p) = 0$ **then**
        Solve $(\widehat{\boldsymbol{H}}_d + \widehat{H}E_d^\top)\boldsymbol{Y} + \boldsymbol{Y}(\widehat{\boldsymbol{G}}_d + \widehat{G}E_d^\top)^\top = E_1\boldsymbol{\beta}_1\boldsymbol{\beta}_2^\top E_1^\top$ for $\boldsymbol{Y}$
        Compute $\rho = \sqrt{\|\boldsymbol{h}_{d+1,d}E_d^\top \boldsymbol{Y}\|_F^2 + \|\boldsymbol{Y}E_d\boldsymbol{g}_{d+1,d}^\top\|_F^2}$
        if $\rho < \text{tol}$ **then**
            break
        end if
    end if
end for
Compute (possibly low-rank) factors $\boldsymbol{Y}_1, \boldsymbol{Y}_2$ such that $\boldsymbol{Y} \approx \boldsymbol{Y}_1\boldsymbol{Y}_2^\top$
Retrieve $\boldsymbol{X}^{(1)} = \boldsymbol{U}_d\boldsymbol{T}_{\boldsymbol{U},d}^{-1}\boldsymbol{Y}_1, \boldsymbol{X}^{(2)} = \boldsymbol{V}_d\boldsymbol{T}_{V,d}^{-1}\boldsymbol{Y}_2$ by the two-pass step

The updating of $\widehat{\boldsymbol{H}}_{d+1}$ and $\widehat{\boldsymbol{G}}_{d+1}$ should be as followings: $\boldsymbol{T}_{\boldsymbol{U},d+1}^{-1} = \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{U},d}^{-1} & -\boldsymbol{T}_{\boldsymbol{U},d}^{-1}T_H\boldsymbol{\tau}_{d+1}^{-1} \\ & \boldsymbol{\tau}_{d+1}^{-1} \end{bmatrix}$ so that

$$\widehat{\boldsymbol{H}}_{d+1} = \begin{bmatrix} \widehat{\boldsymbol{H}}_d + T_H\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}E_d^\top & \widehat{H}_{\text{new}} \\ \boldsymbol{\tau}_{d+1}\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}E_d^\top & \boldsymbol{\tau}_{d+1}\left(-\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}E_d^\top T_H + \boldsymbol{h}_{d+1,d+1}\right)\boldsymbol{\tau}_{d+1}^{-1} \end{bmatrix},$$

where $\widehat{H}_{\text{new}} = \left(-\widehat{\boldsymbol{H}}_d + T_H\boldsymbol{h}_{d+1,d}\boldsymbol{\tau}_d^{-1}E_d^\top\right)T_H\boldsymbol{\tau}_{d+1}^{-1} + \boldsymbol{T}_dH\boldsymbol{\tau}_{d+1}^{-1} + T_H\boldsymbol{h}_{d+1,d+1}\boldsymbol{\tau}_d^{-1}$ and $H = \left[\boldsymbol{h}_{1,d+1}^\top, \ldots, \boldsymbol{h}_{d,d+1}^\top\right]^\top \in R^{dr\times r}$. A corresponding update for $\widehat{\boldsymbol{G}}_{d+1}$ is performed.

Report on 26/2

By the equation, times $\widehat{\boldsymbol{U}_d}' \boldsymbol{S}_{\boldsymbol{U}}'$ on both side

$$\boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{A} \boldsymbol{U}_d = \boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{U}_d \left(\boldsymbol{H}_d + \boldsymbol{T}_{\boldsymbol{U},d}^{-1} T_H \boldsymbol{h}_{d+1,d} E_d^\top \right) + Q_{\boldsymbol{U},d+1} \boldsymbol{\tau}_{d+1} \boldsymbol{h}_{d+1,d} E_d^\top$$

, we get

$$\widehat{\boldsymbol{U}}_d^\top \boldsymbol{S}_{\boldsymbol{U}}^\top \boldsymbol{S}_{\boldsymbol{U}} \boldsymbol{A} \widehat{\boldsymbol{U}}_d = \widehat{\boldsymbol{H}}_d + \widehat{H} E_d^\top,$$

, and we use this formulation to get $\widehat{\boldsymbol{H}}_d$

Tolerance (tol)	Approximation Error $(X_d - X)$	Residue
1×10^{-6}	1.1784×10^{-5}	1.2272×10^{-7}
1×10^{-7}	1.2591×10^{-6}	3.1516×10^{-8}
1×10^{-8}	1.2591×10^{-6}	3.1516×10^{-8}
1×10^{-9}	1.2591×10^{-6}	3.1516×10^{-8}
1×10^{-10}	1.2591×10^{-6}	3.1516×10^{-8}

Table 1: Approximation error and residue for different tolerances. n1=n2=5000, s=600, maxit=300

Tolerance (tol)	Approximation Error $(X_d - X)$	Residue
1×10^{-6}	2.7757×10^{-8}	7.2089×10^{-10}
1×10^{-7}	1.591×10^{-9}	3.827×10^{-11}
1×10^{-8}	1.4575×10^{-10}	5.1717×10^{-12}
1×10^{-9}	1.7549×10^{-11}	8.3189×10^{-13}
1×10^{-10}	7.0832×10^{-12}	1.48×10^{-13}

Table 2: Approximation error and residue for different tolerances . n1=n2=5000, s=3000, maxit = 1500