On solving Khatri-Rao systems of equations

David Eriksson

Center for Applied Mathematics
Cornell University

dme65@cornell.edu

November 20, 2017

Joint work with Charles Van Loan and Alex Townsend

The Kronecker product

- $A \otimes B$ is a block matrix where the ijth block is $a_{ij}B$:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{32} & a_{32} & a_{33} \end{bmatrix} \otimes B = \left[\begin{array}{c|c|c} a_{11}B & a_{12}B & a_{13}B \\ \hline a_{21}B & a_{22}B & a_{23}B \\ \hline a_{32}B & a_{32}B & a_{33}B \end{array}\right]$$

- $A \otimes B$ is data-sparse
- If A is m-by-n, B is p-by-q then:
  - $A \otimes B$ has $mpnq$ entries
  - ...but can be represented by $mn + pq$ entries

Basic algebraic properties

For $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times n}$:

$$
\begin{aligned}
(A \otimes B)^T &= A^T \otimes B^T \\
(A \otimes B)^{-1} &= A^{-1} \otimes B^{-1} \\
(A \otimes B)^{\dagger} &= A^{\dagger} \otimes B^{\dagger} \\
(A \otimes B)(C \otimes D) &= (AC) \otimes (BD) \\
A \otimes (B \otimes C) &= (A \otimes B) \otimes C \\
A \otimes B &= (\text{Perfect Shuffle})^T (B \otimes A)\, (\text{Perfect Shuffle}) \\
\det(A \otimes B) &= \det(A)^n \det(B)^m \\
\operatorname{tr}(A \otimes B) &= \operatorname{tr}(A)\, \operatorname{tr}(B) \\
\operatorname{rank}(A \otimes B) &= \operatorname{rank}(A) \operatorname{rank}(B)
\end{aligned}
$$

Basic properties

- Computing dense factorizations is cheap!
- Only need to compute factorizations of A and B separately

$$\begin{aligned}(A \otimes B) &= (L_A L_A^T) \otimes (L_B L_B^T) \\ &= (L_A \otimes L_B)(L_A \otimes L_B)^T\end{aligned}$$

$$\begin{aligned}(A \otimes B) &= (P_A L_A U_A) \otimes (P_B L_B U_B) \\ &= (P_A \otimes P_B)(L_A \otimes L_B)(U_A \otimes U_B)\end{aligned}$$

Reshaping KP computations

- Assume $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{n\times n}$
- Computing $y = (A \otimes B)x$ is $\mathcal{O}(m^2n^2)$ flops:

```
y = kron(A, B)*x
```

- The equivalent operation $y = \mathrm{vec}(BXA^T)$ is $\mathcal{O}(mn(m+n))$ flops:

```
y = reshape(B*reshape(x, n, m)*A', m*n, 1)
```

- For A, B triangular, solving $(A \otimes B)x = y$ is $\mathcal{O}(mn(m+n))$ flops:

$$\left(\begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \otimes B\right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}B & 0 & 0 \\ a_{21}B & a_{22}B & 0 \\ a_{31}B & a_{32}B & a_{33}B \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

  - For $k = 1 : m$ do $z_k = \frac{y_k - \sum_{i=1}^{k-1} a_{ki} z_i}{a_{kk}}$, $x_k = B^{-1} z_k$

The Khatri-Rao product

- Definition:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} \qquad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{bmatrix}$$

$$C = A \circledast B = \left[\begin{array}{c|c} A_{11} \otimes B_{11} & A_{12} \otimes B_{12} \\ \hline A_{21} \otimes B_{21} & A_{22} \otimes B_{22} \\ \hline A_{31} \otimes B_{31} & A_{32} \otimes B_{32} \end{array}\right]$$

- Assume $A \in \mathbb{R}^{(mp)\times(mp)}$, $B \in \mathbb{R}^{(np)\times(np)}$
- The resulting matrix C is of size $(mnp) \times (mnp)$
- MVMs and triangular solves are $\mathcal{O}(p^2 mn(m+n))$ flops

Matrix equations

- Solve KR system ⇔ Solve system of matrix equations:

$$B_{i1}X_1A_{i1}^T + \cdots + B_{ip}X_pA_{ip}^T = D_i, \qquad i = 1, \ldots, p$$

- Follow from $\sum_{j=1}^{p}(A_{ij} \otimes B_{ij})x_j = \text{vec}\left(\sum_{i=1}^{p} B_{ij}X_jA_{ij}^T\right)$
- Can embed generalized Sylvester equations $\sum_{i=1}^{p}(A_i \otimes B_i)x = f$:

$$\begin{bmatrix} A_1 \otimes B_1 & A_2 \otimes B_2 & A_3 \otimes B_3 \\ I \otimes I & -I \otimes I & 0 \\ 0 & I \otimes I & -I \otimes I \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} f \\ 0 \\ 0 \end{bmatrix} \quad \text{(for } p = 3\text{)}$$

- The last $p-1$ rows make sure $x_1 = x_2 = x_3$

The failed factorization attempt

- **Compute a factorization with KR structure**

$$\begin{bmatrix} A_{11} \otimes B_{11} & (A_{21} \otimes B_{21})^T \\ A_{21} \otimes B_{21} & A_{22} \otimes B_{22} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{bmatrix}.$$

- Equating blocks:

$$\begin{aligned} A_{11} \otimes B_{11} &= L_{11} L_{11}^T \\ A_{21} \otimes B_{21} &= L_{21} L_{11}^T \\ A_{22} \otimes B_{22} &= L_{21} L_{21}^T + L_{22} L_{22}^T \end{aligned}$$

- L_{11} and L_{21} clearly have KP structure
- ...but $L_{22} L_{22}^T = A_{22} \otimes B_{22} - L_{21} L_{21}^T$ does not!
- Dense matrix factorizations do not preserve KR structure

The approximate KR idea

- We can get an *approximate* KR factorization if we solve

$$\min_{\substack{\tilde{A}_{22}\in\mathbb{R}^{m\times m}\\ \tilde{B}_{22}\in\mathbb{R}^{n\times n}}} \left\| (A_{22}\otimes B_{22} - L_{21}L_{21}^T) - \tilde{A}_{22}\otimes\tilde{B}_{22} \right\|_F$$

- Compute the Cholesky factorizations:

$$\tilde{A}_{22} = \tilde{L}_{22}^{(A)}\left[\tilde{L}_{22}^{(A)}\right]^T, \qquad \tilde{B}_{22} = \tilde{L}_{22}^{(B)}\left[\tilde{L}_{22}^{(B)}\right]^T$$

- Gives the approximate factorization:

$$\begin{bmatrix} A_{11}\otimes B_{11} & (A_{21}\otimes B_{21})^T \\ A_{21}\otimes B_{21} & A_{22}\otimes B_{22} \end{bmatrix} \approx \left(\tilde{L}^{(A)} \circledast \tilde{L}^{(B)}\right)\left(\tilde{L}^{(A)} \circledast \tilde{L}^{(B)}\right)^T$$

where

$$\tilde{L}^{(A)} = \begin{bmatrix} L_{11}^{(A)} & 0 \\ L_{21}^{(A)} & \tilde{L}_{22}^{(A)} \end{bmatrix}, \qquad \tilde{L}^{(B)} = \begin{bmatrix} L_{11}^{(B)} & 0 \\ L_{21}^{(B)} & \tilde{L}_{22}^{(B)} \end{bmatrix}$$

Approximate KR-Cholesky factorization

Algorithm 1 Approximate KR-Cholesky: Finds lower triangular matrices $L^{(A)}$ and $L^{(B)}$ such that $C = A \circledast B \approx (L^{(A)} \circledast L^{(B)})(L^{(A)} \circledast L^{(B)})^T$

for i=1:p **do**
  for j=i:p **do**
    $[\tilde{A}_{ij}, \tilde{B}_{ij}] \leftarrow$
      $\textsc{nearest\_kp}\left(A_{ij} \otimes B_{ij} - \sum_{\ell=1}^{i-1} (L_{i\ell}^{(A)} [L_{j\ell}^{(A)}]^T) \otimes (L_{i\ell}^{(B)} [L_{j\ell}^{(B)}]^T)\right)$
    if $i == j$ **then**
      $\tilde{A}_{ii} \leftarrow \textsc{closest\_spd}(\tilde{A}_{ii})$
      $\tilde{B}_{ii} \leftarrow \textsc{closest\_spd}(\tilde{B}_{ii})$
      $L_{ii}^{(A)} \leftarrow \textsc{chol}(\tilde{A}_{ii})$
      $L_{ii}^{(B)} \leftarrow \textsc{chol}(\tilde{B}_{ii})$
    else
      $L_{ji}^{(A)} \leftarrow \left(L_{ii}^{(A)} \setminus \tilde{A}_{ij}\right)^T$
      $L_{ji}^{(B)} \leftarrow \left(L_{ii}^{(B)} \setminus \tilde{B}_{ij}\right)^T$
    end if
  end for
end for

Solving the nearest Kronecker product (NKP) problem

- Consider $V \in \mathbb{R}^{m_1 \times n_1}$, $W \in \mathbb{R}^{m_2 \times n_2}$, $U \in \mathbb{R}^{(m_1 m_2) \times (n_1 n_2)}$
- How do we minimize $\varphi(V, W) = \|U - V \otimes W\|_F$?
- Can be reshaped into $\varphi(V, W) = \|\mathcal{R}(U) - \operatorname{vec}(V) \operatorname{vec}(W)^T\|_F$
- If u, v are the singular vectors corresponding to $\sigma_1(\mathcal{R}(U))$:

$$\operatorname{vec}(V_{opt}) = \sqrt{\sigma_1}\, u, \qquad \operatorname{vec}(W_{opt}) = \sqrt{\sigma_1}\, v.$$

- In the special case $U = \sum_{i=1}^{k} \tilde{U}_i \otimes \hat{U}_i$ we have

$$\mathcal{R}(U) = \sum_{i=1}^{k} \operatorname{vec}(\tilde{U}_i) \operatorname{vec}(\hat{U}_i)^T$$

- Can solve the NKP problem in $\mathcal{O}(k(m_1 n_1 + m_2 n_2))$ in this case

Putting it all together

- Can compute an approximate KR-Cholesky factorization efficiently
- Can also compute an approximate KR-LU factorization:

$$C \approx (P^{(A)} \otimes P^{(B)})(L^{(A)} \otimes L^{(B)})(U^{(A)} \otimes U^{(B)})$$

- Both factorizations cost a total of $\mathcal{O}(p^3(m^3 + n^3))$ flops to compute
- Compare to $\mathcal{O}((mnp)^3)$ flops for a dense factorization
- Fast MVMs and approximate factorization $\implies$ try Krylov method

Preconditioner impact on spectrum

Random SPD matrices A and B with $p = 5$ and $m = n = 20$.

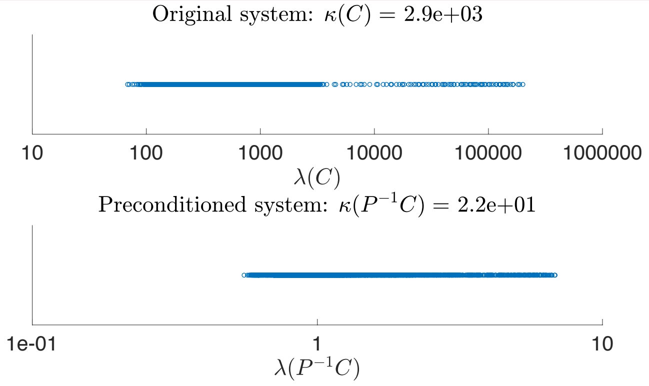

Figure: *Eigenvalue spectrums of C and $P^{-1}C$.*

Numerical experiment

- Random SPD matrices A and B with $p = 10$ and $m = n = 100$.
- Use KR-Cholesky as a preconditioner for CG

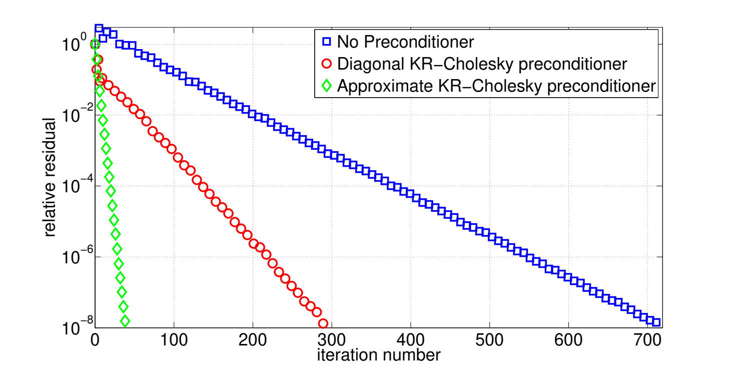

Figure: *Convergence plot for CG.*

Numerical experiment

Preconditioner	Construction time	Iteration time	Iterations
No preconditioner	0 s	14.86 s	724
Block-diagonal KR-Cholesky	0.011 s	16.17 s	298
KR-Cholesky	1.15 s	3.15 s	40

Table: Number of seconds necessary to compute each preconditioner and number of seconds spent on CG iterations until the specified tolerance was achieved.

Conclusions

- Dense factorizations do not preserve KR structure
- Can modify the block-Cholesky/LU algorithms to compute approximate factorizations
- Involves solving an NKP problem in each block position
- Works well as a preconditioner with CG/GMRES on toy problems
- I am still looking for interesting applications