Fast Matrix Multiplication Search Using Cyclic Invariant CP-Decomposition Factor Matrices

J

1 It's Gradient Time

1.1 CP Gradient Under Normal Circumstances

For $\mathcal{X} \in \mathbb{R}^{m\times n\times p}$, $\mathbf{A} \in \mathbb{R}^{m\times r}$, $\mathbf{B} \in \mathbb{R}^{n\times r}$, and $\mathbf{C} \in \mathbb{R}^{p\times r}$, $v = \begin{bmatrix} vec(A) \\ vec(B) \\ vec(C) \end{bmatrix}$ $\in \mathbb{R}^{mnp}$;

$$f(v) = \frac{1}{2}\|\mathcal{X} - [\mathbf{A},\mathbf{B},\mathbf{C}]\|^2 \tag{1}$$

$$=$$

$$f(v) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p}\big(x_{ijk} - \sum_{l=1}^{r} a_{il}b_{jr}c_{kl})\big)^2 \tag{2}$$

The gradient of this equation is defined as follows

$$\nabla f = \begin{bmatrix} \mathsf{vec}(\frac{\partial f}{\partial \mathbf{A}}) \\ \mathsf{vec}(\frac{\partial f}{\partial \mathbf{B}}) \\ \mathsf{vec}(\frac{\partial f}{\partial \mathbf{C}}) \end{bmatrix} \in \mathbb{R}^{(m+n+p)r}$$

Where each partial derivative is defined below.

$$\frac{\partial f}{\partial \mathbf{A}} = -\mathbf{X}_{(1)}(\mathbf{C}\odot\mathbf{B}) + \mathbf{A}(\mathbf{C}^\intercal\mathbf{C} * \mathbf{B}^\intercal\mathbf{B}) \tag{3}$$

$$\frac{\partial f}{\partial \mathbf{B}} = -\mathbf{X}_{(1)}(\mathbf{C}\odot\mathbf{A}) + \mathbf{B}(\mathbf{C}^\intercal\mathbf{C} * \mathbf{A}^\intercal\mathbf{A}) \tag{4}$$

$$\frac{\partial f}{\partial \mathbf{C}} = -\mathbf{X}_{(1)}(\mathbf{B}\odot\mathbf{A}) + \mathbf{C}(\mathbf{B}^\intercal\mathbf{B} * \mathbf{A}^\intercal\mathbf{A}) \tag{5}$$

1.2 What We Are Interested In

Now define A, B, and C to be cyclic invariant.

$$\mathbf{A} = \begin{bmatrix}\mathbf{S} & \mathbf{U} & \mathbf{V} & \mathbf{W}\end{bmatrix}$$
$$\mathbf{B} = \begin{bmatrix}\mathbf{S} & \mathbf{W} & \mathbf{U} & \mathbf{V}\end{bmatrix}$$
$$\mathbf{C} = \begin{bmatrix}\mathbf{S} & \mathbf{C} & \mathbf{W} & \mathbf{U}\end{bmatrix}$$

Where $\mathbf{S} \in \mathbb{R}^{n \times R_s}$, and $\mathbf{U}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^{n \times R_c}$, where n is set by the size of the matmul tensor we are decomposing and and R_s and R_c are the sizes that we can set depending on the size of the matmul tensor. Using these matrices, our equation becomes:

$$f(v) = \frac{1}{2} \sum_{i}^{m} \sum_{j}^{n} \sum_{k}^{p} \left(x_{ijk} - \sum_{q}^{r_s} s_{iq} s_{jq} s_{kq} - \sum_{l}^{r_{uvw}} (u_{il} v_{jl} w_{kl} + w_{il} u_{jl} v_{kl} + v_{il} w_{jl} u_{kl}) \right)^2 \tag{6}$$

We are then interested in getting the partial derivatives with respect to S, U, V, and W, in other words,

$$\nabla f = \begin{bmatrix} \mathsf{vec}(\frac{\partial f}{\partial \mathbf{S}}) \\ \mathsf{vec}(\frac{\partial f}{\partial \mathbf{U}}) \\ \mathsf{vec}(\frac{\partial f}{\partial \mathbf{V}}) \\ \mathsf{vec}(\frac{\partial f}{\partial \mathbf{W}}) \end{bmatrix} \in \mathbb{R}^{(m+n+p)r}$$

Then we take the derivative of f with respect to a single entry in S, namely $s_{i'q'}$.

$$\frac{\partial f}{\partial s_{i'q'}} = - \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} \big(x_{ijk} - \sum_{q=1}^{r_{\mathbf{s}}} \mathbf{s}_{iq} \mathbf{s}_{jq} \mathbf{s}_{kq} - \sum_{l=1}^{r_{\mathbf{uvw}}} (\mathbf{u}_{il} \mathbf{v}_{jl} \mathbf{w}_{kl} + \mathbf{w}_{i,l} \mathbf{u}_{j,l} \mathbf{v}_{k,l} + \mathbf{v}_{il} \mathbf{w}_{jl} \mathbf{u}_{kl}) \big) \cdot \\ (\partial_{ii'} \mathbf{s}_{jq'} \mathbf{s}_{kq'} + \mathbf{s}_{iq'} \partial_{ji'} \mathbf{s}_{kq'} + \mathbf{s}_{iq'} \mathbf{s}_{jq'} \partial_{kq'})$$

There is one tricky part that comes with the chain rule of this function. Take a look, for example, at the second term of the second line; $\mathbf{s}_{iq'} \partial_{ji'} \mathbf{s}_{kq'}$. What this means is that we took the derivative of $s_{i'q'}$ in the entirety of the $r_{\mathbf{s}}$ summation, but somewhere in it, j matches i'. Thus, we take the derivative of that specific term (thus the $\partial_{ji'}$), but what comes along are the other s factors in that term. Since this can only happen whenever q = q' we are left with the s factors of different i and k, but of equal q. This Happens often throughout this project, so it is important to understand why this happens. The steps that follow are not difficult, just tedious and lengthy, so we do it one step out of time.

$$=_1 - \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} (x_{ijk}) (\partial_{ii'} \mathbf{s}_{jq'} \mathbf{s}_{kq'} + \mathbf{s}_{iq'} \partial_{ji'} \mathbf{s}_{kq'} + \mathbf{s}_{iq'} \mathbf{s}_{jq'} \partial_{kq'})$$

$$= -\sum_{j=1}^{n}\sum_{k=1}^{p} x_{i'jk}\mathbf{s}_{jq'}\mathbf{s}_{kq'} - \sum_{i=1}^{m}\sum_{k=1}^{p} x_{ij'k}\mathbf{s}_{iq'}\mathbf{s}_{kq'} - \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ijk'}\mathbf{s}_{iq'}\mathbf{s}_{jq'}$$
$$= -\big[\mathbf{X}_{(1)}\big(\mathbf{S}\odot\mathbf{S}\big)\big]_{i'q'} - \big[\mathbf{X}_{(2)}\big(\mathbf{S}\odot\mathbf{S}\big)\big]_{j'q'} - \big[\mathbf{X}_{(3)}\big(\mathbf{S}\odot\mathbf{S}\big)\big]_{k'q'}$$
$$= -3\big[\mathbf{X}_{(1)}\big(\mathbf{S}\odot\mathbf{S}\big)\big]_{i'q'}$$

$$=_2 \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p}(\sum_{q=1}^{r_{\mathbf{s}}}\mathbf{s}_{iq}\mathbf{s}_{jq}\mathbf{s}_{kq})(\partial_{ii'}\mathbf{s}_{jq'}\mathbf{s}_{kq'} + \mathbf{s}_{iq'}\partial_{ji'}\mathbf{s}_{kq'} + \mathbf{s}_{iq'}\mathbf{s}_{jq'}\partial_{kq'})$$
$$= \sum_{j=1}^{n}\sum_{k=1}^{p}\sum_{q=1}^{r_{\mathbf{s}}}\mathbf{s}_{i'q}\mathbf{s}_{jq}\mathbf{s}_{kq}\mathbf{s}_{jq'}\mathbf{s}_{kq'} + \sum_{i=1}^{m}\sum_{k=1}^{p}\sum_{q=1}^{r_{\mathbf{s}}}\mathbf{s}_{iq}\mathbf{s}_{j'q}\mathbf{s}_{kq}\mathbf{s}_{iq'}\mathbf{s}_{kq'} + \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{q=1}^{r_{\mathbf{s}}}\mathbf{s}_{iq}\mathbf{s}_{jq}\mathbf{s}_{k'q}\mathbf{s}_{iq'}\mathbf{s}_{jq'}$$
$$= \sum_{q=1}^{r_{\mathbf{s}}}\mathbf{s}_{i'q}(\sum_{j=1}^{n}\mathbf{s}_{jq}\mathbf{s}_{jq'})(\sum_{k=1}^{p}\mathbf{s}_{kq}\mathbf{s}_{kq'}) + \sum_{q=1}^{r_{\mathbf{s}}}\mathbf{s}_{j'q}(\sum_{i=1}^{m}\mathbf{s}_{iq}\mathbf{s}_{iq'})(\sum_{k=1}^{p}\mathbf{s}_{kq}\mathbf{s}_{kq'}) + \sum_{q=1}^{r_{\mathbf{s}}}\mathbf{s}_{i'q}(\sum_{i=1}^{m}\mathbf{s}_{iq}\mathbf{s}_{iq'})(\sum_{j=1}^{n}\mathbf{s}_{jq}\mathbf{s}_{jq'})$$
$$= [S(S^{\intercal}S) * (S^{\intercal}S)]_{i'q'} + [S(S^{\intercal}S) * (S^{\intercal}S)]_{j'q'} + [S(S^{\intercal}S) * (S^{\intercal}S)]_{k'q'}$$
$$= 3[S(S^{\intercal}S) * (S^{\intercal}S]_{i'q'}$$

$$=_3 \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p}(\sum_{l=1}^{r_{\mathbf{uvw}}}(\mathbf{u}_{il}\mathbf{v}_{jl}\mathbf{w}_{kl} + \mathbf{w}_{i,l}\mathbf{u}_{j,l}\mathbf{v}_{k,l} + \mathbf{v}_{il}\mathbf{w}_{jl}\mathbf{u}_{kl}))(\partial_{ii'}\mathbf{s}_{jq'}\mathbf{s}_{kq'} + \mathbf{s}_{iq'}\partial_{ji'}\mathbf{s}_{kq'} + \mathbf{s}_{iq'}\mathbf{s}_{jq'}\partial_{kq'})$$
$$= \sum_{j=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{r_{uvw}}\mathbf{u}_{i'l}\mathbf{v}_{jl}\mathbf{w}_{kl}\mathbf{s}_{jq'}\mathbf{s}_{kq'} + \sum_{i=1}^{m}\sum_{k=1}^{p}\sum_{l=1}^{r_{uvw}}\mathbf{u}_{il}\mathbf{v}_{j'l}\mathbf{w}_{kl}\mathbf{s}_{iq'}\mathbf{s}_{kq'} + \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{r_{uvw}}\mathbf{u}_{il}\mathbf{v}_{jl}\mathbf{w}_{k'l}\mathbf{s}_{iq'}\mathbf{s}_{jq'}$$
$$+ \sum_{j=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{r_{uvw}}\mathbf{w}_{i'l}\mathbf{u}_{jl}\mathbf{v}_{kl}\mathbf{s}_{jq'}\mathbf{s}_{kq'} + \sum_{i=1}^{m}\sum_{k=1}^{p}\sum_{l=1}^{r_{uvw}}\mathbf{w}_{il}\mathbf{u}_{j'l}\mathbf{v}_{kl}\mathbf{s}_{iq'}\mathbf{s}_{kq'} + \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{r_{uvw}}\mathbf{w}_{il}\mathbf{u}_{jl}\mathbf{v}_{k'l}\mathbf{s}_{iq'}\mathbf{s}_{jq'}$$
$$+ \sum_{j=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{r_{uvw}}\mathbf{v}_{i'l}\mathbf{w}_{jl}\mathbf{u}_{kl}\mathbf{s}_{jq'}\mathbf{s}_{kq'} + \sum_{i=1}^{m}\sum_{k=1}^{p}\sum_{l=1}^{r_{uvw}}\mathbf{v}_{il}\mathbf{w}_{j'l}\mathbf{u}_{kl}\mathbf{s}_{iq'}\mathbf{s}_{kq'} + \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l=1}^{r_{uvw}}\mathbf{v}_{il}\mathbf{w}_{jl}\mathbf{u}_{k'l}\mathbf{s}_{iq'}\mathbf{s}_{jq'}$$
$$= \sum_{l=1}^{r_{uvw}}\mathbf{u}_{i'l}(\sum_{j=1}^{n}\mathbf{v}_{jl}\mathbf{s}_{jq'})(\sum_{k=1}^{p}\mathbf{w}_{kl}\mathbf{s}_{kq'}) + \sum_{l=1}^{r_{uvw}}\mathbf{v}_{j'l}(\sum_{i=1}^{m}\mathbf{u}_{il}\mathbf{s}_{iq'})(\sum_{k=1}^{p}\mathbf{w}_{kl}\mathbf{s}_{kq'}) + \sum_{l=1}^{r_{uvw}}\mathbf{w}_{k'l}(\sum_{i=1}^{m}\mathbf{u}_{il}\mathbf{s}_{iq'})(\sum_{j=1}^{n}\mathbf{v}_{jl}\mathbf{s}_{jq'})$$
$$+ \sum_{l=1}^{r_{uvw}}\mathbf{w}_{i'l}(\sum_{j=1}^{n}\mathbf{u}_{jl}\mathbf{s}_{jq'})(\sum_{k=1}^{p}\mathbf{v}_{kl}\mathbf{s}_{kq'}) + \sum_{l=1}^{r_{uvw}}\mathbf{u}_{j'l}(\sum_{i=1}^{m}\mathbf{w}_{il}\mathbf{s}_{iq'})(\sum_{k=1}^{p}\mathbf{v}_{kl}\mathbf{s}_{kq'}) + \sum_{l=1}^{r_{uvw}}\mathbf{v}_{k'l}(\sum_{i=1}^{m}\mathbf{w}_{il}\mathbf{s}_{iq'})(\sum_{j=1}^{n}\mathbf{u}_{jl}\mathbf{s}_{jq'})$$

$$\frac{\partial f}{\partial \mathbf{S}} = \quad 3\cdot\Big(S\big((S^{\intercal}S) * (S^{\intercal}S)\big) + U\big((V^{\intercal}S) * (W^{\intercal}S)\big) + V\big((U^{\intercal}S) * (W^{\intercal}S)\big) + W\big((U^{\intercal}S) * (V^{\intercal}S)\big)$$
$$-(\mathbf{X}_{(1)} + \mathbf{X}_{(2)} + \mathbf{X}_{(3)})(S\odot S)\Big)$$
$$\frac{\partial f}{\partial \mathbf{U}} = \quad 3\cdot\Big(S\big((S^{\intercal}V) * (S^{\intercal}W)\big) + U\big((V^{\intercal}V) * (W^{\intercal}W)\big) + V\big((W^{\intercal}V) * (U^{\intercal}W)\big) + W\big((U^{\intercal}V) * (V^{\intercal}W)\big)\Big)$$

$$
-\mathbf{X}_{(1)}(V \odot W) - \mathbf{X}_{(2)}(W \odot V) - \mathbf{X}_{(3)}(V \odot W)
$$

$$
\frac{\partial f}{\partial \mathbf{V}} = 3 \cdot \Big(S\big((S^\intercal U) * (S^\intercal W)\big) + U\big((W^\intercal U) * (V^\intercal W)\big) + V\big((U^\intercal U) * (W^\intercal W)\big) + W\big((V^\intercal U) * (U^\intercal W)\big) \Big)
$$

$$
-\mathbf{X}_{(1)}(W \odot U) - \mathbf{X}_{(2)}(U \odot W) - \mathbf{X}_{(3)}(W \odot U)
$$

$$
\frac{\partial f}{\partial \mathbf{W}} = 3 \cdot \Big(S\big((S^\intercal U) * (S^\intercal V)\big) + U\big((V^\intercal U) * (W^\intercal V)\big) + V\big((W^\intercal U) * (U^\intercal V)\big) + W\big((U^\intercal U) * (V^\intercal V)\big) \Big)
$$

$$
-\mathbf{X}_{(1)}(U \odot V) - \mathbf{X}_{(2)}(V \odot U) - \mathbf{X}_{(3)}(U \odot V)
$$

2 It's Jacobian Time

$$J = \begin{bmatrix} J_s & J_u & J_v & J_w \end{bmatrix} \in \mathbb{R}^{n^3 \times n(r_s + 3r_{uvw})} \tag{7}$$

$$J_s = (S \odot S) \otimes I_n + \Pi_2^\intercal \cdot (S \odot S) \otimes I_n + \Pi_3^\intercal \cdot (S \odot S) \otimes I_n \tag{8a}$$
$$J_u = (V \odot W) \otimes I_n + \Pi_2^\intercal \cdot (W \odot V) \otimes I_n + \Pi_3^\intercal \cdot (V \odot W) \otimes I_n \tag{8b}$$
$$J_v = (W \odot U) \otimes I_n + \Pi_2^\intercal \cdot (U \odot W) \otimes I_n + \Pi_3^\intercal \cdot (W \odot U) \otimes I_n \tag{8c}$$
$$J_w = (U \odot V) \otimes I_n + \Pi_2^\intercal \cdot (V \odot U) \otimes I_n + \Pi_3^\intercal \cdot (U \odot V) \otimes I_n \tag{8d}$$

$$J^\intercal J \cdot vec(M) = \begin{bmatrix} J_s^\intercal J_s & J_s^\intercal J_u & J_s^\intercal J_v & J_s^\intercal J_w \\ J_u^\intercal J_s & J_u^\intercal J_u & J_u^\intercal J_v & J_u^\intercal J_w \\ J_v^\intercal J_s & J_v^\intercal J_u & J_v^\intercal J_v & J_v^\intercal J_w \\ J_w^\intercal J_s & J_w^\intercal J_u & J_w^\intercal J_v & J_w^\intercal J_w \end{bmatrix} \begin{bmatrix} vec(M_s) \\ vec(M_u) \\ vec(M_v) \\ vec(M_w) \end{bmatrix} \tag{9}$$

$$J_s vec(M_s) = vec\Big(M_s (S \odot S)^\intercal + S(S \odot M_s)^\intercal + S(M_s \odot S)^\intercal \Big) \tag{10a}$$
$$J_u vec(M_u) = vec\Big(M_u (V \odot W)^\intercal + V(W \odot M_u)^\intercal + W(M_u \odot V)^\intercal \Big) \tag{10b}$$
$$J_v vec(M_v) = vec\Big(M_v (W \odot U)^\intercal + W(U \odot M_v)^\intercal + U(M_v \odot W)^\intercal \Big) \tag{10c}$$
$$J_w vec(M_w) = vec\Big(M_w (U \odot V)^\intercal + U(V \odot M_w)^\intercal + V(M_w \odot U)^\intercal \Big) \tag{10d}$$

$$
\begin{aligned}
J_s^\intercal J_s vec(M_s) &= 3 \cdot vec\Big(M_s(S^\intercal S) * (S^\intercal S) + 2 \cdot S(M_s^\intercal S) * (S^\intercal S)\Big) \\
J_s^\intercal J_u vec(M_u) &= 3 \cdot vec\Big(M_u(V^\intercal S) * (W^\intercal S) + V(M_u^\intercal S) * (W^\intercal S) + W(M_u^\intercal S) * (V^\intercal S)\Big) \\
J_s^\intercal J_v vec(M_v) &= 3 \cdot vec\Big(M_v(U^\intercal S) * (W^\intercal S) + U(M_v^\intercal S) * (W^\intercal S) + W(M_v^\intercal S) * (U^\intercal S)\Big) \\
J_s^\intercal J_w vec(M_w) &= 3 \cdot vec\Big(M_w(U^\intercal S) * (V^\intercal S) + U(M_w^\intercal S) * (V^\intercal S) + V(M_w^\intercal S) * (U^\intercal S)\Big)
\end{aligned}
$$

$$
\begin{aligned}
J_u^\intercal J_s vec(M_s) &= 3 \cdot vec\big(M_s(S^\intercal V) * (S^\intercal W) + S\big((M_s^\intercal W) * (S^\intercal V) + (M_s^\intercal V) * (S^\intercal W)\big)\big) \\
J_u^\intercal J_u vec(M_u) &= 3 \cdot vec\Big(M_u(V^\intercal V) * (W^\intercal W) + V(M_u^\intercal W) * (W^\intercal V) + W(M_u^\intercal V) * (V^\intercal W)\Big) \\
J_u^\intercal J_v vec(M_v) &= 3 \cdot vec\Big(M_v(W^\intercal V) * (U^\intercal W) + W(M_v^\intercal W) * (U^\intercal V) + U(M_v^\intercal V) * (W^\intercal W)\Big) \\
J_u^\intercal J_w vec(M_w) &= 3 \cdot vec\Big(M_w(U^\intercal V) * (V^\intercal W) + U(M_w^\intercal W) * (V^\intercal V) + V(M_w^\intercal V) * (U^\intercal W)\Big)
\end{aligned}
$$

$$
\begin{aligned}
J_v^\intercal J_s vec(M_s) &= 3 \cdot vec\Big(M_s(S^\intercal U) * (S^\intercal W) + S\big((M_s^\intercal U) * (S^\intercal W) + (M_s^\intercal W) * (S^\intercal U)\big)\Big) \\
J_v^\intercal J_u vec(M_u) &= 3 \cdot vec\Big(M_u(V^\intercal W) * (W^\intercal U) + V(M_u^\intercal U) * (W^\intercal W) + W(M_u^\intercal W) * (V^\intercal U)\Big) \\
J_v^\intercal J_v vec(M_v) &= 3 \cdot vec\Big(M_v(W^\intercal W) * (U^\intercal U) + W(M_v^\intercal U) * (U^\intercal W) + U(M_v^\intercal W) * (W^\intercal U)\Big) \\
J_v^\intercal J_w vec(M_w) &= 3 \cdot vec\Big(M_w(U^\intercal W) * (V^\intercal U) + U(M_w^\intercal U) * (V^\intercal W) + V(M_w^\intercal W) * (U^\intercal U)\Big)
\end{aligned}
$$

$$
\begin{aligned}
J_w^\intercal J_s vec(M_s) &= 3 \cdot vec\Big(M_s(S^\intercal U) * (S^\intercal V) + S\big((M_s^\intercal U) * (S^\intercal V) + (M_s^\intercal V) * (S^\intercal U)\big)\Big) \\
J_w^\intercal J_u vec(M_u) &= 3 \cdot vec\Big(M_u(V^\intercal U) * (W^\intercal V)) + V(M_u^\intercal V) * (W^\intercal U)) + W(M_u^\intercal U) * (V^\intercal V)\Big) \\
J_w^\intercal J_v vec(M_v) &= 3 \cdot vec\Big(M_v(W^\intercal W) * (U^\intercal V) + W\big(2 \cdot (M_v^\intercal V) * (U^\intercal U) + U(M_v^\intercal U) * (W^\intercal V)\Big) \\
J_w^\intercal J_w vec(M_w) &= 3 \cdot vec\Big(M_w(U^\intercal U) * (V^\intercal V) + U(M_w^\intercal V) * (V^\intercal U) + V(M_w^\intercal U) * (U^\intercal V)\Big)
\end{aligned}
$$