On Commuting Elementwise Multiply with Matmul in TTIR: Necessary and Sufficient Conditions, Rewrite Patterns, and Extensions

Mihailo Đorđević

Abstract

This note studies when an elementwise product can be moved across a matrix multiplication without changing the numerical result. Two base transformations are considered: commuting the elementwise mask to the first operand of `matmul` and commuting it to the second operand. I give simple criteria that decide when each rewrite is valid. When the base rewrites fail, I discuss extensions that pre-process the left input through a fixed linear operator so that the expression can still be reordered in a useful way. Proofs and counterexamples are deferred to the appendices.

Contents

1 Introduction and scope

The TTIR graph often contains the pattern

$$(A \circ B)\, C,$$

where $A, B, C \in \mathbb{R}^{n \times n}$, $\circ$ denotes the Hadamard (elementwise) product, and juxtaposition denotes standard matrix multiplication. This work provides precise conditions for two shape preserving rewrites:

- **Transformation A** (move the mask to the left input): find B', C' depending only on B, C such that

$$(A \circ B)\, C = (A\, B') \circ C' \quad \text{for all } A.$$

- **Transformation B** (move the mask to the output of `matmul`): find B', C' depending only on B, C such that

$$(A \circ B)\, C = A \circ (B'C') \quad \text{for all } A.$$

When neither transformation is globally valid, I also describe extended rewrites that introduce a fixed linear pre-processing of A before the `matmul`.

1.1 Notation

All matrices are real $n \times n$ unless stated otherwise. For a vector $u \in \mathbb{R}^n$, $\mathrm{diag}(u)$ is the diagonal matrix with diagonal u. For a matrix X, $\mathrm{vec}(X)$ stacks its columns into a vector. The support of a vector x is $\mathrm{supp}(x) = \{i : x_i \neq 0\}$. The kth column of X is $X_{\cdot k}$. I use I for the identity, $\mathbf{1}$ for the all ones vector, and $J = \mathbf{1}\mathbf{1}^\top$ for the all ones matrix.

1.2 Elementary facts used throughout

I collect the basic identities that will be used in the analysis. Proofs are standard and omitted; they are recalled in Appendix A for completeness.

Hadamard product. For $X, Y \in \mathbb{R}^{n \times n}$, the Hadamard product $X \circ Y$ is defined entrywise, $(X \circ Y)_{ij} = X_{ij} Y_{ij}$. It is bilinear, commutative and associative. Two interactions will be repeatedly used:

$$X \circ (uv^\top) = \mathrm{diag}(u)\, X\, \mathrm{diag}(v), \tag{1}$$

$$\mathrm{vec}(X \circ Y) = \mathrm{Diag}(\mathrm{vec}(Y))\, \mathrm{vec}(X). \tag{2}$$

Identity (1) says that a rank one mask $uv^\top$ induces left and right diagonal scalings. Identity (2) converts Hadamard multiplication into a diagonal linear map in the vectorized space.

Matrix multiplication and vectorization. For conforming matrices, $\mathrm{vec}(AXB) = (B^\top \otimes A)\, \mathrm{vec}(X)$. This turns matching of matrix identities into matching of linear maps in the vectorized form. I will use this to equate coefficients with respect to the basis matrices $E^{(ik)}$ that have a single one in position (i, k).

Linearity in A. In both transformations, B and C are fixed while A is arbitrary. Each side is linear in A. Equality for all A holds if and only if the two linear maps agree on a basis, equivalently if the coefficient of each a_{ik} matches in every output position. This is the key idea behind the criteria in Section 2 and the proofs in Appendix A.

1.3 Problem statement

Given $B, C \in \mathbb{R}^{n \times n}$ and arbitrary $A \in \mathbb{R}^{n \times n}$, decide:

Q1. When do there exist B', C' depending only on B, C such that $(A \circ B)C = (AB') \circ C'$ for all A?

Q2. When do there exist B', C' depending only on B, C such that $(A \circ B)C = A \circ (B'C')$ for all A?

Q3. If neither is possible in general, what structured cases allow it?

Q4. If a base rewrite is not valid, can we insert a fixed linear pre-processing of A to obtain a useful reordered form without changing the final result?

2   Main results

This section states the conditions in plain English with minimal notation and many examples. Mathematical proofs are in Appendix A.

2.1   Transformation A: when can we move the mask to the left input

Target form.   We want fixed B', C' that depend only on B, C such that

$$(A \circ B)\, C \;=\; (A\, B') \circ C' \quad \text{for every } A.$$

No shape changes and no extra operators are allowed.

Practical criterion

Think column by column through C.

1) Pick a column j of C.

2) Find the row indices where this column is nonzero:

$$\text{support}(C_{\cdot j}) = \{\, k : C_{kj} \neq 0 \,\}.$$

3) Look at those columns of B: $B_{\cdot k}$ for all k in that set.

4) **Requirement.** All these $B_{\cdot k}$ must point in the same direction, that is, they must be scalar multiples of one another.

   If the requirement holds for every column j of C, then the rewrite is valid. If it fails for any column, the rewrite is not valid.

Intuition.   Each column of C decides which columns of B are allowed to interact in the same output column. When two or more columns of B are mixed together by the same column of C, those B-columns must be proportional. If a column of C is 1-sparse, there is nothing to check for that column.

Quick checklists

- Dense C (no zeros): then every column of C mixes all columns of B, so all columns of B must be mutually proportional. Equivalently, B must be rank 1.

- Each column of C has at most one nonzero: the rewrite always passes for any B.

- Block-sparse C where columns only use indices from disjoint blocks $\mathcal{K}_1, \ldots, \mathcal{K}_r$: inside each block $\mathcal{K}_t$, all columns of B must be proportional. Different blocks may use different directions.

How to build B' and C' when it passes

For each column j of C:

- pick a unit direction $u^{(j)}$ equal to the common direction of the involved $B_{\cdot k}$,

- write $B_{\cdot k} = \alpha_k u^{(j)}$ for all k with $C_{kj} \neq 0$,

- set $C'_{\cdot j} = u^{(j)}$ and $B'_{kj} = \alpha_k C_{kj}$, with zeros elsewhere.

Then $(AB') \circ C'$ reproduces $(A \circ B)C$ for all A.

Examples that pass

Example A1. Overlapping supports, consistent proportionality.

$$B = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 3 & 6 & 3 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Column 1 of C uses $k = \{1, 2\}$, and $B_{\cdot 2} = 2 B_{\cdot 1}$. Column 2 uses $k = \{2, 3\}$, and $B_{\cdot 3} = \frac{1}{2} B_{\cdot 2}$. Column 3 uses $k = \{3\}$ only. All checks pass. The rewrite is valid.

Example A2. Dense C with rank 1 B. Let $B = uv^\top$ with $u = \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}$ and $v = \begin{bmatrix} 3 \\ 1 \\ -5 \end{bmatrix}$, and let C be any dense matrix. Every column of B is a scalar multiple of u, so the requirement holds automatically for every column of C. The rewrite is valid.

Examples that fail

Example A3. The classic counterexample.

$$B = I_2, \qquad C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Each column of C uses both $k = 1$ and $k = 2$, but $B_{\cdot 1} = [1, 0]^\top$ and $B_{\cdot 2} = [0, 1]^\top$ are not proportional. The rewrite is invalid.

Example A4. Two disjoint mixes with inconsistent directions.

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Column 1 of C uses $k = \{1, 2\}$ but $B_{\cdot 1} = [1, 0, 0]^\top$ and $B_{\cdot 2} = [1, 1, 0]^\top$ are not proportional. The rewrite is invalid, no need to check other columns.

2.2 Transformation B: when can we move the mask to the output

Target form. We want fixed B', C' that depend only on B, C such that

$$(A \circ B)\, C \;=\; A \circ (B'C') \quad \text{for every } A.$$

Practical criterion

The output entry (i, j) on the right depends only on a_{ij}. On the left, (i, j) collects contributions from all a_{ik} with weights $b_{ik} C_{kj}$. For equality for every A, all off diagonal contributions must vanish.

> For each column j of C and each row index $k \neq j$, if $C_{kj} \neq 0$ then the entire column k of B must be zero.

If the condition holds for all columns j, the rewrite is valid. Otherwise it is invalid.

Intuition. The right hand side cannot mix a_{ik} with $k \neq j$ into entry (i, j). Therefore the left hand side must not have such contributions either. The only way to kill them for every i is that the whole column k of B is zero whenever C_{kj} is nonzero with $k \neq j$.

Quick checklists

- If C is diagonal, the rewrite always passes for any B.

- If C has an off diagonal nonzero at (k, j) and the kth column of B is not identically zero, the rewrite fails immediately.

- You may keep off diagonal nonzeros in column j only in rows k whose column k of B is entirely zero.

How to build B' and C' when it passes

Let $D = \text{diag}(C_{11}, \ldots, C_{nn})$. Set $B' = B$ and $C' = D$. Then

$$A \circ (B'C') \;=\; A \circ (BD),$$

and because all off diagonal paths have been eliminated by the condition above, this equals $(A \circ B)C$ for every A.

Examples that pass

Example B1. Diagonal C, arbitrary B.

$$C = \begin{bmatrix} c_{11} & 0 & 0 \\ 0 & c_{22} & 0 \\ 0 & 0 & c_{33} \end{bmatrix}.$$

No off diagonal entries exist, so the condition holds. Set $C' = \text{diag}(c_{11}, c_{22}, c_{33})$ and $B' = B$. The rewrite is valid.

Example B2. Off diagonals target only zero columns of B.

$$B = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 5 & 0 \\ 0 & 7 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 3 \end{bmatrix}.$$

Column 1 of C has an off diagonal at $k = 3$ but column 3 of B is zero. Column 2 of C has an off diagonal at $k = 1$ but column 1 of B is zero. Column 3 of C is diagonal only. The rewrite is valid.

Examples that fail

Example B3. Single off diagonal breaks it.

$$B = I_3, \qquad C = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Column 1 of C has $C_{21} = 1$ with $k = 2 \neq 1$ and column 2 of B is not zero. The rewrite is invalid.

Example B4. Dense C with any nonzero column of B.

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Column 1 of C has $C_{21} = 1$ but column 2 of B is nonzero. The rewrite is invalid.

2.3   Summary table for quick decisions

Transformation	Pass condition
A: $(A \circ B)C = (AB') \circ C'$	For every column j in C, all B-columns indexed by the nonzeros of $C_{\cdot j}$ are proportional. Dense C forces rank 1 B. 1-sparse columns of C are always fine.
B: $(A \circ B)C = A \circ (B'C')$	For every column j in C and every off diagonal index $k \neq j$ with $C_{kj} \neq 0$, the entire column k of B must be zero. Diagonal C is always fine.

Constructing B', C'

For each column j:

1) Pick a direction $u^{(j)}$. If there exists $k \in L_j$ with $B_{\cdot k} \neq 0$, set $u^{(j)} = \frac{B_{\cdot k}}{\|B_{\cdot k}\|_2}$. If all $B_{\cdot k} = 0$, choose any unit vector $u^{(j)}$; that column of the output will be zero regardless.

2) For every $k \in L_j$, compute α_k so that $B_{\cdot k} = \alpha_k u^{(j)}$ (use $\alpha_k = u^{(j)\top} B_{\cdot k}$ if $u^{(j)}$ is unit length).

3) Set $C'_{\cdot j} = u^{(j)}$ and $B'_{kj} = \alpha_k C_{kj}$, with zeros elsewhere.

Shapes are preserved and $(AB') \circ C'$ equals $(A \circ B)C$ for all A.

3 Extensions with a fixed preprocessing of the left input

This section discusses optional rewrites that first apply a fixed linear preprocessing to A before the matrix multiplication. These rewrites always preserve the final numerical result.

3.1 Shape preserving preprocessing

Define a linear operator $\tau : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ by columns:

$$\tau(A)_{\cdot j} = \sum_k \mathrm{diag}(B_{\cdot k}) \, A_{\cdot k} \, C_{kj}.$$

Then for any A, $\tau(A)$ equals $(A \circ B)C$. Using $C' = \mathbf{1}\mathbf{1}^\top$, we can write

$$(A \circ B)C = \tau(A) \circ C'.$$

3.2 Low rank structure in the columns of B

Assume the set $\{B_{\cdot k}\}_{k=1}^n$ clusters into r proportional directions:

$$B_{\cdot k} = \alpha_k \, u^{(t)} \quad \text{for } k \in \mathcal{K}_t, \quad t = 1, \dots, r.$$

Then

$$\tau(A)_{\cdot j} = \sum_{t=1}^r \mathrm{diag}(u^{(t)}) \Big(\sum_{k \in \mathcal{K}_t} A_{\cdot k} \, \alpha_k \, C_{kj} \Big).$$

If $r = 1$ we recover a very simple schedule: one row scale by u and one multiplication with $\mathrm{diag}(v)C$ where $B = uv^\top$.

3.3 Column sparsity in C

If each column of C has at most s nonzeros, computing $\tau(A)_{\cdot j}$ touches at most s row scales. This reduces to a small gather-and-accumulate per output column.

4 Conclusion

Two base rewrites that commute elementwise multiply with matrix multiplication were studied in a shape preserving setting. For Transformation A, the necessary and sufficient pass condition is that for every column of C, all columns of B indexed by the nonzeros of that column are proportional. This includes dense C paired with rank one B and block sparse C with componentwise rank one columns in B. For Transformation B, the necessary and sufficient pass condition is that every off diagonal nonzero of C targets a zero column of B, with the diagonal case always allowed.

A   Appendix A: Mathematical analysis and proofs

This appendix presents the formal statements and proofs for the results used in Sections 2.1 and 2.2. Throughout, $A, B, C \in \mathbb{R}^{n \times n}$ are real matrices. The Hadamard product $X \circ Y$ is defined entrywise. The standard matrix product is denoted by juxtaposition.

A.1   Preliminaries

I collect standard identities and conventions.

Definition A.1 (Vectorization and Kronecker)**.** For a matrix $X \in \mathbb{R}^{n \times n}$, $\text{vec}(X) \in \mathbb{R}^{n^2}$ stacks columns of X. For conformable A, X, B,

$$\text{vec}(AXB) = (B^\top \otimes A) \; \text{vec}(X).$$

Definition A.2 (Diagonalization of elementwise multiplication)**.** For $X, Y \in \mathbb{R}^{n \times n}$,

$$\text{vec}(X \circ Y) = \text{Diag}(\text{vec}(Y)) \; \text{vec}(X).$$

Definition A.3 (Rank one mask identity)**.** For $u, v \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times n}$,

$$X \circ (uv^\top) = \text{diag}(u) \, X \, \text{diag}(v).$$

Remark A.4 (Linearity in the left argument)**.** Fixing B and C, the maps

$$\mathcal{T}_A : A \mapsto (A \circ B)C \quad \text{and} \quad \mathcal{S}_A : A \mapsto (AB') \circ C'$$

are linear in A. Equality $\mathcal{T}_A \equiv \mathcal{S}_A$ for all A is equivalent to equality on a basis of $\mathbb{R}^{n \times n}$, hence to equality of coefficients of each a_{ik} in every output position.

A.2   Transformation A

We seek conditions under which there exist B', C' depending only on B, C such that

$$(A \circ B)C = (AB') \circ C' \quad \text{for all } A \in \mathbb{R}^{n \times n}. \tag{3}$$

Lemma A.5 (Coefficient identity for A)**.** *Equality* (3) *holds for all A if and only if for all indices i, k, j,*

$$b_{ik} \, c_{kj} = b'_{kj} \, c'_{ij}. \tag{4}$$

Proof. The (i, j) entry of the left hand side is $\sum_k a_{ik} b_{ik} c_{kj}$. The (i, j) entry of the right hand side is $\left(\sum_k a_{ik} b'_{kj} \right) c'_{ij}$. Equality for all A is equivalent to equality of the coefficients of a_{ik}, hence (4). $\quad\square$

For a fixed column j, define the matrix $M^{(j)} \in \mathbb{R}^{n \times n}$ by $m_{ik}^{(j)} = b_{ik} c_{kj}$. Equation (4) is equivalent to the requirement that $M^{(j)}$ factor as an outer product of a column vector (depending on i) and a row vector (depending on k), namely $M^{(j)} = \alpha^{(j)}(\beta^{(j)})^\top$ with $\alpha_i^{(j)} = c'_{ij}$ and $\beta_k^{(j)} = b'_{kj}$.

Lemma A.6 (Rank one characterization)**.** *For fixed j, $M^{(j)} = [b_{ik}c_{kj}]_{i,k}$ has rank one if and only if all nonzero columns of $M^{(j)}$ are pairwise proportional.*

Proof. The rank of a matrix equals one if and only if all columns lie in a one dimensional subspace, equivalently all nonzero columns are proportional. $\quad\square$

Observe that the kth column of $M^{(j)}$ equals $c_{kj}B_{\cdot k}$. Therefore all nonzero columns of $M^{(j)}$ are scalar multiples of the corresponding columns of B. This yields the following necessary and sufficient condition.

Theorem A.7 (Pass condition for A). *Equality (3) holds for all A if and only if for every column j of C the family*

$$\{B_{\cdot k} : c_{kj} \neq 0\}$$

consists of proportional vectors. Equivalently, for each j with support $\mathrm{supp}(C_{\cdot j})$, the set of columns $\{B_{\cdot k} : k \in \mathrm{supp}(C_{\cdot j})\}$ spans a one dimensional subspace of $\mathbb{R}^n$.

Proof. Fix j. By Lemma A.5 and the discussion above, (3) holds if and only if $M^{(j)}$ has rank one. By Lemma A.6, this is equivalent to all nonzero columns of $M^{(j)}$ being proportional, which is equivalent to all $B_{\cdot k}$ with $c_{kj} \neq 0$ being proportional. The equivalence for each j implies the statement. □

Corollary A.8 (Dense C). *If C has no zeros, then (3) holds for all A if and only if B has rank one.*

Proof. If C has no zeros then for each j, $\mathrm{supp}(C_{\cdot j}) = \{1, \ldots, n\}$. Theorem A.7 implies that all columns of B are proportional, hence $\mathrm{rank}(B) = 1$. The converse is immediate. □

Corollary A.9 (One-sparse columns of C). *If each column of C has at most one nonzero entry, then (3) holds for all A for any B.*

Proof. If $|\mathrm{supp}(C_{\cdot j})| \leq 1$, the set in Theorem A.7 has cardinality at most one, hence is trivially proportional. □

Corollary A.10 (Block sparse C). *Suppose the indices $\{1, \ldots, n\}$ are partitioned into disjoint blocks $\mathcal{K}_1, \ldots, \mathcal{K}_r$ with the property that for every column j of C, $\mathrm{supp}(C_{\cdot j}) \subseteq \mathcal{K}_t$ for some t. Then (3) holds for all A if and only if for each block $\mathcal{K}_t$ the family $\{B_{\cdot k} : k \in \mathcal{K}_t\}$ is proportional.*

Proof. Apply Theorem A.7 within each block. □

Theorem A.11 (Construction of B' and C'). *If the pass condition in Theorem A.7 holds, one valid choice is*

$$C'_{\cdot j} = u^{(j)}, \qquad B'_{kj} = \alpha_k C_{kj},$$

where for each j we select a nonzero vector $u^{(j)}$ spanning the common direction of $\{B_{\cdot k} : c_{kj} \neq 0\}$ and scalars α_k with $B_{\cdot k} = \alpha_k u^{(j)}$ for all k in the support of $C_{\cdot j}$, with $\alpha_k = 0$ if $c_{kj} = 0$.

Proof. Fix j. For each k with $c_{kj} \neq 0$, $B_{\cdot k} = \alpha_k u^{(j)}$ by assumption. Then for all i and k,

$$b_{ik}c_{kj} = \alpha_k u_i^{(j)} c_{kj} = b'_{kj}c'_{ij}$$

with $b'_{kj} = \alpha_k c_{kj}$ and $c'_{ij} = u_i^{(j)}$. This matches (4) and concludes the proof. □

[Failure in general] Transformation A fails in general. For instance, with $n = 2$, $B = I_2$ and $C = J_2$ there are no B', C' satisfying (3) for all A.

Proof. Here each column of C has support $\{1, 2\}$, but the columns of B are e_1 and e_2, which are not proportional. By Theorem A.7 the condition fails. □

A.3 Transformation B

We seek conditions under which there exist B', C' depending only on B, C such that

$$(A \circ B)C = A \circ (B'C') \quad \text{for all } A \in \mathbb{R}^{n \times n}. \tag{5}$$

Lemma A.12 (Coefficient identity for B). *Equality* (5) *holds for all A if and only if there exists a matrix* $D = B'C'$ *such that for all* i, k, j,

$$b_{ik} c_{kj} = \delta_{kj} d_{ij}, \tag{6}$$

where δ_{kj} *is the Kronecker symbol.*

Proof. The (i, j) entry of the left hand side is $\sum_k a_{ik} b_{ik} c_{kj}$. The (i, j) entry of the right hand side is $a_{ij} d_{ij}$. Equality for all A is equivalent to matching the coefficient of each a_{ik} in entry (i, j), hence (6). $\square$

Theorem A.13 (Pass condition for B). *Equality* (5) *holds for all A if and only if for every column j of C and every* $k \neq j$ *with* $c_{kj} \neq 0$, *the kth column of B is identically zero.*

Proof. If $k \neq j$ and $c_{kj} \neq 0$, then by (6) we must have $b_{ik} = 0$ for all i, hence the entire kth column of B is zero. Conversely, if every off diagonal nonzero of c_{kj} targets a zero column of B, then (6) holds with $d_{ij} = b_{ij}c_{jj}$, which corresponds to $D = B \operatorname{diag}(C_{11}, \ldots, C_{nn})$. $\square$

Corollary A.14 (Diagonal case). *If C is diagonal then* (5) *holds for all A for any B. One may take* $B' = B$ *and* $C' = \operatorname{diag}(C_{11}, \ldots, C_{nn})$.

Proof. If C is diagonal then for $k \neq j$ we have $c_{kj} = 0$, so (6) imposes no constraints beyond $d_{ij} = b_{ij}c_{jj}$. $\square$

[Failure with dense C] If C has an off diagonal nonzero at (k, j) and the kth column of B is nonzero, then Transformation B fails.

Proof. Immediate from Theorem A.13. $\square$

A.4 Vectorized proofs

For completeness, I give vectorized derivations that encode the proofs above as equalities of linear maps.

Transformation A

We require

$$\operatorname{vec}((A \circ B)C) = \operatorname{vec}((AB') \circ C')$$

for all A. Using identities,

$$\operatorname{vec}((A \circ B)C) = (C^\top \otimes I) \operatorname{Diag}(\operatorname{vec}(B)) \operatorname{vec}(A),$$

and

$$\operatorname{vec}((AB') \circ C') = \operatorname{Diag}(\operatorname{vec}(C')) (B'^\top \otimes I) \operatorname{vec}(A).$$

Therefore equality for all A is equivalent to

$$(C^\top \otimes I) \, \text{Diag}(\text{vec}(B)) = \text{Diag}(\text{vec}(C')) \, (B'^\top \otimes I). \tag{7}$$

Inspecting the $(i, k; j)$ block of (7) yields $b_{ik}c_{kj} = b'_{kj}c'_{ij}$, which is (4). The rank one characterization follows from the observation that for fixed j the block column equals $[c_{1j}B._1, \ldots, c_{nj}B._n]$, which has rank one if and only if all its nonzero columns are proportional.

Transformation B

We require

$$\text{vec}((A \circ B)C) = \text{vec}(A \circ (B'C')) = \text{Diag}(\text{vec}(B'C')) \, \text{vec}(A)$$

for all A, which is equivalent to

$$(C^\top \otimes I) \, \text{Diag}(\text{vec}(B)) = \text{Diag}(\text{vec}(B'C')).$$

For a fixed j, the jth block column of the left-hand side consists of blocks

$$\left[c_{1j} \, \text{Diag}(B._1), \ \ldots, \ c_{nj} \, \text{Diag}(B._n) \right].$$

For this operator to be diagonal, every off-diagonal block $c_{kj} \, \text{Diag}(B._k)$ with $k \neq j$ must be the zero matrix, hence if $c_{kj} \neq 0$ then $B._k = 0$. The diagonal block equals $c_{jj} \, \text{Diag}(B._j)$, which identifies

$$\text{Diag}(\text{vec}(B'C')) = \text{Diag}\big(\text{vec}(B \, \text{diag}(C_{11}, \ldots, C_{nn})) \big),$$

so one may take $D = B'C' = B \, \text{diag}(C_{11}, \ldots, C_{nn})$.

A.5   Constructive counterexamples

Example A.15 (A fails with $B = I_2$, $C = J_2$)**.** Assume B', C' satisfy (3) for all A. Comparing coefficients in entry $(1, 1)$ yields $b'_{11}c'_{11} = 1$ and $b'_{21}c'_{11} = 0$. In entry $(2, 2)$ we obtain $b'_{22}c'_{22} = 1$ and $b'_{12}c'_{22} = 0$. This forces both $b'_{21} = 0$ and $b'_{22} \neq 0$, and simultaneously $c'_{11} \neq 0$ and $c'_{22} \neq 0$, which conflicts when matching the mixed entries $(1, 2)$ and $(2, 1)$. No solution exists.

Example A.16 (B fails with $B = I_2$, $C = J_2$)**.** Equation (6) requires $b_{i2}c_{21} = \delta_{21}d_{i1} = 0$ for all i. Since $c_{21} = 1$ and $b_{i2} = [0, 1]^\top$ is not zero, this is impossible.

A.6   Stability under perturbations

The pass conditions are structurally stable. For Transformation A, small perturbations of B and C preserve proportionality when the proportional groups in B are separated by a uniform angle margin and when zero entries of C remain below the verifier threshold. For Transformation B, small off diagonal entries of C must remain below the threshold or the corresponding columns of B must be sufficiently close to zero.

A.7   Uniqueness and degrees of freedom

In Transformation A, when the pass condition holds, B' and C' are not unique. For each column j, any scaling $C'._j \leftarrow \gamma_j C'._j$ and $B'._j \leftarrow \gamma_j^{-1} B'._j$ preserves (4). In Transformation B, once the matrix $\tilde{D} := B \, \text{diag}(c_{11}, \ldots, c_{nn})$ is fixed, any factorization $\tilde{D} = B'C'$ works; a canonical choice is $B' = B$, $C' = \text{diag}(c_{11}, \ldots, c_{nn})$.

A.8 Extension note

The extended constructions with preprocessing τ in Section 3 can also be derived in vectorized form by factoring $(C^\top \otimes I)\,\mathrm{Diag}(\mathrm{vec}(B))$ into a product of block diagonal and skinny mixing operators. The detailed algebra is given in Appendix B.

Proof. By definition of τ the equality is entrywise true for every (i, j):

$$[\tau(A) \circ C']_{ij} = \tau(A)_{ij} = \sum_k a_{ik}\, b_{ik}\, c_{kj} = \big[(A \circ B)C\big]_{ij}.$$

$\square$

Theorem A.17 (Correctness of grouped-direction lift). *Assume $B_{\cdot k} = \alpha_k u^{(t)}$ for $k \in \mathcal{K}_t$, $t = 1, \ldots, r$. Let τ be as in (??), $B^\diamond$ as described above and $C^\diamond = \mathbf{1}\mathbf{1}^\top$. Then (??) holds for all A.*

Proof. The proof is already given after (??). It is a regrouping of terms indexed by the partition of $\{1, \ldots, n\}$ into $\{\mathcal{K}_t\}$. $\square$

Theorem A.18 (Correctness of column-sparsity lift). *Assume every column of C has at most s nonzeros. Let τ be as in (??), with $B^\diamond$ routing nonzero row indices of C to distinct blocks and pre-multiplying by $\mathrm{diag}(B_{\cdot k})$. With $C^\diamond = \mathbf{1}\mathbf{1}^\top$, (??) holds for all A.*

Proof. Immediate from the construction in the proof sketch for F3. $\square$

A.9 Limitations and non-goals

These extensions are not substitutes for the base shape-preserving rewrites in Sections 2.1–2.2. They are useful only when the verifier detects structural reuse.

No creation of new commutativity. If one insists on a final form where raw A appears unchanged on the right hand side, i.e.

$$(A \circ B)C \overset{?}{=} A \circ (B'C'),$$

no preprocessing τ can help unless Theorem A.13 holds, because the right hand side depends only on a_{ij} while the left depends on all a_{ik} in the same row.

Avoid wide lifts without reuse. The n^2-width lift that splits by every column and row is always possible algebraically but is not a practical rewrite. The grouped-direction and column-sparsity lifts are bounded by small r or s and are the only viable variants.

A.10 Stability and tolerances

In practice one works with floating point arithmetic and near-proportional directions. The grouped-direction lift is stable if the angle between directions is bounded away from zero and the assignment of columns to directions respects a margin. The column-sparsity lift is stable if the small entries of C can be safely treated as zeros by a numerics policy.

A.11 Worked examples

I give two concrete instances with explicit matrices to illustrate the constructions.

Example B.1. Rank one B. Let $B = uv^\top$ with $u = [1, -2, 4]^\top$, $v = [3, 1, -5]^\top$ and let C be arbitrary. Then $r = 1$ with $u^{(1)} = u$ and $\alpha_k = v_k$. The lift uses

$$\tau(A) = \begin{bmatrix} \operatorname{diag}(u)A \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad B^\diamond = \operatorname{diag}(v)\,C \in \mathbb{R}^{n \times n}, \quad C^\diamond = \mathbf{1}\mathbf{1}^\top.$$

Hence

$$(A \circ B)C = \left(\operatorname{diag}(u)A \right)\left(\operatorname{diag}(v)C \right),$$

followed by a Hadamard with all ones, which is a no-op. This is a very cheap schedule: one row-scale and one matmul.

Example B.2. Two-column sparsity per output column. Let $n = 4$. Suppose each column of C has exactly two nonzeros. Then $s = 2$ and

$$\tau(A) = [\, A \mid A \,] \in \mathbb{R}^{4 \times 8}.$$

For the jth output column, let the nonzeros of $C_{.j}$ occur at rows k_1, k_2 with values γ_1, γ_2. The corresponding column of $B^\diamond$ places $\operatorname{diag}(B_{.k_1})$ in the first block and $\operatorname{diag}(B_{.k_2})$ in the second block, with weights γ_1, γ_2. Then the jth column of $(\tau(A)B^\diamond)$ equals

$$\operatorname{diag}(B_{.k_1})A_{.k_1}\,\gamma_1 + \operatorname{diag}(B_{.k_2})A_{.k_2}\,\gamma_2,$$

which is exactly the jth column of $(A \circ B)C$.

A.12   Uniqueness and degrees of freedom

The factorizations are not unique. In F2, any rescaling $u^{(t)} \leftarrow \gamma_t u^{(t)}$ paired with $\alpha_k \leftarrow \alpha_k/\gamma_t$ for $k \in \mathcal{K}_t$ leaves the construction invariant. In F3, the assignment of nonzero rows to blocks is arbitrary up to permutations and zero padding. One may choose assignments that align with memory layout or kernel preferences.

A.13   Relation to the base criteria

When the pass condition for Transformation A holds, the grouped-direction lift with r equal to the number of connected components in the co-support graph of C degenerates to the exact B', C' construction in Theorem A.11. When Transformation B holds with diagonal C, the shape-preserving τ reduces to a single diagonal post-scaling and becomes redundant.

A.14   Proof of vec-factorization claims

For completeness I record a compact derivation of (**??**) and the factorizations.

Lemma A.19. *For all $A, B, C \in \mathbb{R}^{n \times n}$,*

$$\operatorname{vec}\left((A \circ B)C\right) = (C^\top \otimes I)\, \operatorname{Diag}(\operatorname{vec}(B))\, \operatorname{vec}(A).$$

Proof. We have $\operatorname{vec}((A \circ B)C) = (C^\top \otimes I)\,\operatorname{vec}(A \circ B) = (C^\top \otimes I)\,\operatorname{Diag}(\operatorname{vec}(B))\,\operatorname{vec}(A)$ by the standard properties of vectorization and the diagonalization of the Hadamard product. $\quad\square$

Lemma A.20 (Grouped-direction factorization)**.** *Under the assumptions of F2,*

$$\mathcal{M}(B, C) = (B^{\diamond\top} \otimes I)\left(I \otimes \mathcal{U}\right)\Pi,$$

where $\mathcal{U}$ *is the block diagonal operator with blocks* $\mathrm{diag}(u^{(t)})$, Π *is a permutation that replicates columns according to the partition* $\{\mathcal{K}_t\}$, *and* $B^\diamond$ *is as defined earlier.*

Proof. The kth column of $\mathrm{Diag}(\mathrm{vec}(B))$ applies the row scale $\mathrm{diag}(B_{\cdot k})$. For $k \in \mathcal{K}_t$ this equals $\alpha_k \, \mathrm{diag}(u^{(t)})$. Grouping by t and pulling out the common operators yields the claimed factorization, with the mixing weights collected in $B^\diamond$. $\qquad\square$

Lemma A.21 (Column-sparsity factorization)**.** *Under the assumptions of F3,*

$$\mathcal{M}(B, C) = (B^{\diamond\top} \otimes I) \, \Pi_s \, \mathrm{Diag}(\mathrm{vec}(B)),$$

where Π_s *replicates* $\mathrm{vec}(A)$ *into* s *blocks and* $B^\diamond$ *routes the at most s nonzeros per column of C into distinct blocks.*

Proof. Replication by Π_s builds s copies of columns so that at most s terms are needed per output column. The routing encoded by $B^\diamond$ collects the correct pairs of $(B_{\cdot k}, C_{kj})$ into the jth output column. $\qquad\square$

A.15 Concluding remark for Appendix B

The extended rewrites follow from linear-algebraic factorizations of the operator $(C^\top \otimes I) \, \mathrm{Diag}(\mathrm{vec}(B))$. Their practical value depends on the number of reusable directions among $\{B_{\cdot k}\}$ and on the column sparsity of C. These are precisely the two structural features that the verifiers and cost models should detect before emitting an extended rewrite