Criteria and proof for when a new view must be appended in a Tinygrad shapeTracker not considering any masks or offsets

Peter Schillinger

September 2024

Contents

1 Definitions

Definition: A **view** is defined as a pair $(\vec{k}, \vec{s})$, where:

$$\vec{k} = (k_1, k_2, \ldots, k_n), and \quad \vec{s} = (s_1, s_2, \ldots, s_n)$$

Definition: The **view order** is an ordering of the tensor's elements associated with that view. For the first view in a shapeTracker, the view order is the order of elements stored statically in memory. For our purposes we will identify elements by their index (starting from 0) in the view order.

Definition: Given a shape $\vec{k} = (k_1, k_2, \ldots, k_n)$, define the **mixed radix conversion function** to be the function: $R_{\vec{k}} : N \to N^n$, with domain $\{0, 1, \ldots, (\prod_{l=1}^{n} k_l) - 1\}$.

$$R_{\vec{k}}(i) = \begin{pmatrix} \left\lfloor \frac{i}{\prod_{l=2}^{n} k_l} \right\rfloor \bmod k_1 \\ \vdots \\ \left\lfloor \frac{i}{k_n} \right\rfloor \bmod k_{n-1} \\ i \bmod k_n \end{pmatrix} = \left(\left\lfloor \frac{i}{\prod_{l=j+1}^{n} k_l} \right\rfloor \bmod k_j \right)_{j=1}^{n}$$

Definition: Given a view $(\vec{k}, \vec{s})$, the **reshape order** is the permutation of size n, defined as the composition of $R_{\vec{k}}$ and the matrix multiplication of $A = [s_1 \, s_2 \, \dots \, s_n]$. That is, the ith element in the reshape order is given by

$$A(R_{\vec{k}}(i))$$

for i in $\{0, 1, \dots, (\prod_{l=1}^{n} k_l) - 1\}$. If the tensor has not been permuted, the reshape order is the identity permutation.

Definition: Given a view $\vec{k} = (k_1, k_2, \dots, k_n)$, $\vec{s} = (s_1, \dots, s_n)$, define the **max merge partition** to be the partition

$$\vec{p} = (p_1 = 1, p_2, \dots, p_q)$$

to be such that for all integers $j \in \{1, \dots, n\}$ where $p_i < j < p_{i+1}$ for some $i \in \{1, \dots, q-1\}$,

$$s_j \cdot k_j = s_{j-1}$$

and for all $i \in \{1, \dots, q\}$

$$s_{p_i} \cdot k_{p_i} \neq s_{p_i - 1}$$

Definition: Given such a view and max merge partition, define the **max merge view** to have shape

$$\vec{m} = (m_1, \dots, m_q)$$

and strides

$$\vec{t} = (t_1, \dots, t_q),$$

where, for $i \in \{1, \dots, q\}$,

$$m_i = \prod_{\ell = p_i}^{p_{i+1}-1} k_\ell \quad \text{(where } p_{q+1} = n+1\text{)}$$

and

$$t_i = s_{p_{i+1}-1} \quad \text{(where again } p_{q+1} = n+1\text{).}$$

2 Proof of Mergeability Altering Within Contiguous Subblocks

2.1 Splitting Any Dimension

Given a view $\vec{k_1} = (k_1, k_2, \ldots, k_n)$ and $\vec{s_1} = (s_1, s_2, \ldots, s_n)$, we show that for any $j \in \{1, \ldots, n\}$ and any factors p_1 and p_2 of k_j, i.e., $p_1 \cdot p_2 = k_j$, reshaping to a new shape

$$\vec{k_2} = (k_1, k_2, \ldots, k_{j-1}, p_1, p_2, k_{j+1}, \ldots, k_n) := (c_1, \ldots, c_{n+1})$$

need not append a new view.

Proof: To show that a new view need not be appended, we must show that there exists a linear function from the indices of the new shape to the reshape order. Let $A_1 : N^n \to N$ be the linear function mapping the old indices to the reshape order, i.e., $A_1 = [s_1 \ \ldots \ s_n]$. Then we are to show that there exists some A_1 such that, for all i in $\{0, 1, \ldots, (\prod_{l=1}^{n} k_l) - 1\}$,

$$A_1(R_{\vec{k_1}}(i)) = A_2(R_{\vec{k_2}}(i))$$

. Computing the left hand side,

$$A_1(R_{\vec{k_1}}(i)) = A_1(\begin{pmatrix} \left\lfloor \frac{i}{\prod_{l=2}^{n} k_l} \right\rfloor \bmod k_1 \\ \vdots \\ \left\lfloor \frac{i}{k_n} \right\rfloor \bmod k_{n-1} \\ i \bmod k_n \end{pmatrix}) = \sum_{r=1}^{n} s_r \left(\left\lfloor \frac{i}{\prod_{l=r+1}^{n} k_l} \right\rfloor \bmod k_r \right)$$

We claim that $A_2 = \vec{s_2}^\top = [s_1, s_2, \ldots, s_{j-1}, p_2 \cdot s_j, s_j, s_{j+1}, \ldots, s_n] := [t_1, \ldots, t_{n+1}]$ works.

Computing the right hand side gives

$$A_2(R_{\vec{k_2}}(i)) = A_2(\begin{pmatrix} \left\lfloor \frac{i}{\prod_{l=2}^{n+1} c_l} \right\rfloor \bmod c_1 \\ \vdots \\ \left\lfloor \frac{i}{c_{n+1}} \right\rfloor \bmod c_n \\ i \bmod c_{n+1} \end{pmatrix}) = \sum_{r=1}^{n+1} t_r \left(\left\lfloor \frac{i}{\prod_{l=r+1}^{n+1} k_l} \right\rfloor \bmod c_r \right)$$

Cancelling like terms on each side, we find we are only left with

$$s_j \left(\left\lfloor \frac{i}{\prod_{l=j+1}^{n} k_l} \right\rfloor \bmod k_j \right) = t_j \left(\left\lfloor \frac{i}{\prod_{l=j+1}^{n+1} c_l} \right\rfloor \bmod p_1 \right) + t_{j+1} \left(\left\lfloor \frac{i}{\prod_{l=j+2}^{n+1} c_l} \right\rfloor \bmod p_2 \right)$$

$$= p_2 \cdot s_j \left(\left\lfloor \frac{i}{\prod_{l=j+1}^{n+1} c_l} \right\rfloor \bmod p_1 \right) + s_j \left(\left\lfloor \frac{i}{\prod_{l=j+1}^{n} k_l} \right\rfloor \bmod p_2 \right)$$

Dividing out by s_j and setting $w = \frac{i}{\prod_{l=j+1}^{n} k_l}$ we get

$$\lfloor w \rfloor \bmod k_j = p_2 \left(\left\lfloor \frac{w}{p_2} \right\rfloor \bmod p_1 \right) + (\lfloor w \rfloor \bmod p_2) \tag{1}$$

Let $r_2 = \lfloor w \rfloor \bmod p_2$. Then

$$\lfloor w \rfloor = p_2 \left\lfloor \frac{w}{p_2} \right\rfloor + r_2$$

Let $\left\lfloor \frac{w}{p_2} \right\rfloor = r_1 \bmod p_1$. Then

$$\left\lfloor \frac{w}{p_2} \right\rfloor = p_1 m + r_1 \quad \text{for some integer } m.$$

Therefore,

$$\begin{aligned}
\lfloor w \rfloor &= p_2 \lfloor \frac{w}{p_2} \rfloor + r_2 \\
&= p_2 (p_1 m + r_1) + r_2 \\
&= p_1 p_2 m + p_2 r_1 + r_2 \\
&= k_j m + p_2 r_1 + r_2 \\
&= p_2 r_1 + r_2 \bmod k_j
\end{aligned}$$

But

$$p_2 \left(\left\lfloor \frac{w}{p_2} \right\rfloor \bmod p_1 \right) + (\lfloor w \rfloor \bmod p_2) = p_2 r_1 + r_2 \quad \text{by definition.}$$

$$\Rightarrow \quad \text{LHS} = \text{RHS}$$

Done #

2.2 Combining Dimension Within Contiguous Subblock

Given view

$$\vec{k}_1 = (k_1, k_2, \ldots, k_n), \qquad \vec{s}_1 = (s_1, s_2, \ldots, s_n),$$

we show that for any $j \in \{2, \ldots, n\}$, if $s_j k_j = s_{j-1}$, then reshaping to

$$\vec{k}_2 = (k_1, \ldots, k_{j-2}, k_{j-1} \cdot k_j, k_{j+1}, \ldots, k_n) := (c_1, \ldots, c_{n-1}),$$

need not append a new view, i.e., there exists a linear function from indices in $\vec{k}_2$ to reshape order.

Proof:

Let A_1 be a linear function for shape $\vec{k}_1$. We show there exists a linear function A_2 such that

$$A_1\left(R_{\vec{k}_1}(i)\right) = A_2\left(R_{\vec{k}_2}(i)\right) \quad \text{for all } i \in \left\{0, \ldots, \prod_{i=1}^{n} k_i - 1\right\}.$$

Let $A_2 = \vec{s}_2^{\,T} = \left[s_1, \ldots, s_{j-2}, s_j, s_{j+1}, \ldots, s_n\right] := \left[t_1, \ldots, t_{n-1}\right]$.
We want to show

$$A_1\left(R_{\vec{k}_1}(i)\right) = A_2\left(R_{\vec{k}_2}(i)\right) \quad \text{for all } i \in \left\{0, \ldots, \prod_{i=1}^{n} k_i - 1\right\}$$

Simplifying both sides gives

$$\sum_{r=1}^{n} s_r\left(\left\lfloor \frac{i}{\prod_{l=r+1}^{n} k_l}\right\rfloor \mod k_r\right) = \sum_{r=1}^{n-1} t_r\left(\left\lfloor \frac{i}{\prod_{l=r+1}^{n-1} c_l}\right\rfloor \mod c_r\right)$$

And cancelling like terms gives

$$s_{j-1}\left(\left\lfloor \frac{i}{\prod_{l=j}^{n} k_l}\right\rfloor \mod k_{j-1}\right) + s_j\left(\left\lfloor \frac{i}{\prod_{l=j+1}^{n} k_l}\right\rfloor \mod k_j\right) = t_{j-1}\left(\left\lfloor \frac{i}{\prod_{l=j}^{n-1} c_l}\right\rfloor \mod k_{j-1}k_j\right)$$

But because $t_{j-1} = s_j$, $s_{j-1} = s_j k_j$, and letting $w = \frac{i}{\prod_{l=j+1}^{n} k_l}$, we recover an equivalent equation to (1):

$$k_j\left(\left\lfloor \frac{w}{k_j}\right\rfloor \mod k_{j-1}\right) + \left(\lfloor w \rfloor \mod k_j\right) = \lfloor w \rfloor \mod k_{j-1}k_j$$

which was shown to hold.

Done #

2.3 Proof of Mergeability of Arbitrary Altering Inside of Contiguous Subblocks

Given a view $\vec{k} = (k_1, \ldots, k_n)$, $\vec{s} = (s_1, \ldots, s_n)$ and a new shape $\vec{u} = (u_1, \ldots, u_n)$ for which there exists a partition $\vec{p} = (p_1 = 1, p_2, \ldots, p_q)$ such that for all $i \in \{1, \ldots, q\}$,

$$\prod_{l=p_i}^{p_{i+1}-1} \mu_l = m_i,$$

where m_i is the ith size in the max merge view's shape $\vec{m} = (m_1, \ldots, m_q)$ associated with $(\vec{k}, \vec{s})$.
Proof:
By iteratively combining all dimensions to the view's max merge shape (shown in 2.2), and then splitting m_i into arbitrary factors $\mu_{p_i}, \mu_{p_i+1}, \ldots, \mu_{p_{i+1}-1}$ (shown in 2.1), we can reshape to $\vec{u}$ without ever appending a view.

Done #

3 Disproof of Mergeability Altering Outside of Contiguous Subblocks

Given view $(\vec{u}, \vec{v})$ and associated max merge view $(\vec{m}, \vec{s})$, let $\vec{k} = (k_1, \ldots, k_n)$ be such that $\prod k_\ell = \prod u_\ell$, but there exists no partition $\vec{p} = (p_1, \ldots, p_\ell)$ such that

$$m_i = \prod_{l=p_i}^{p_{i+1}-1} k_l \quad \forall i \in \{1, \ldots, q\}.$$

We show that reshaping to $\vec{k}$ necessarily appends view, that is, that the function from indices in $\vec{k}$ to reshape order is nonlinear.
Proof:
Imagine accumulating the k's in $\vec{k}$, starting at k_n and going backwards and keeping track of the cumulative product.
If there is no partition, then there must exist some first index j such that

$$\prod_{\ell=j+1}^{n} k_\ell < \prod_{\ell=i}^{q} m_\ell < \prod_{\ell=j}^{n} k_\ell \quad \text{where } i \text{ is chosen to be maximal.}$$

But because j is the first index where this happens and i is maximal,

$$\prod_{\ell=i+1}^{q} m_\ell = \prod_{\ell=p}^{n} k_\ell \quad \text{for some } p \in \{j+1, \ldots, n\}.$$

And of course, because i is maximal

$$\prod_{\ell=i+1}^{q} m_\ell \leq \prod_{\ell=j+1}^{n} k_\ell.$$

Let $A_0 = \vec{v}^\top$ be the given linear function for the shape $\vec{u}$.
Towards a contradiction, assume there exists some linear function A_2 such that

$$A_0\left(R_{\vec{u}}(i)\right) = A_2\left(R_{\vec{k}}(i)\right) \quad \forall i \in \left\{0, \ldots, \prod_{\ell=1}^{n} k_\ell - 1\right\}.$$

Because we have shown mergeability to max merged shape, we know, for $A_1 = \vec{s}^T = [s_1, \ldots, s_\ell]$,

$$A_0\left(R_{\vec{u}}(i)\right) = A_1\left(R_{\vec{m}}(i)\right) \quad \text{where } \vec{u}$$

$$\Rightarrow A_1\left(R_{\vec{m}}(i)\right) = A_2\left(R_{\vec{k}}(i)\right)$$

Let $A_2 = (t_1, \ldots, t_n)$.
For $a \in \{j+1, j+2, \ldots, p\}$, let the number

$$n_a = \prod_{l=a}^{n} k_l = (\prod_{l=i+1}^{q} m_l)(\prod_{l=a}^{p-1} k_l) < \prod_{l=i}^{q} m_l \quad (\text{where } \prod_{l=p}^{p-1} k_l = 1 \text{ of course}).$$

Then

$$R_{\vec{m}}(n_a) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \left(\left\lfloor \frac{(\prod_{l=i+1}^{q} m_l)(\prod_{l=a}^{p-1} k_l)}{\prod_{l=i+1}^{q} m_l} \right\rfloor \bmod m_i \right) \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \prod_{l=a}^{p-1} k_l \\ \vdots \\ 0 \end{pmatrix}$$

and

$$R_{\vec{k}}(n_a) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \text{where the 1 appears in the (a-1)th entry}$$

Of course

$$A_1 \left(R_{\vec{m}}(n_a) \right) = A_2 \left(R_{\vec{k}}(n_a) \right)$$

still holds.
But

$$A_1 \left(R_{\vec{m}}(n_a) \right) = s_i \prod_{l=a}^{p-1} k_l \quad \text{and} \quad A_2 \left(R_{\vec{k}}(n_a) \right) = t_{a-1}.$$

$$\Rightarrow \quad t_{a-1} = s_i \prod_{l=a}^{p-1} k_l.$$

Now consider the number

$$n_0 = \prod_{l=i}^{q} m_l = m_i \prod_{l=p}^{n} k_l.$$

$$R_{\vec{m}}(n_0) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \quad (i\text{-th spot}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and

$$R_{\vec{k}}(n_0) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \left\lfloor \frac{m_i}{\prod_{l=j+1}^{p-1} k_l} \right\rfloor \bmod k_j \\ \vdots \\ \left\lfloor \frac{m_i}{k_{p-1}} \right\rfloor \bmod k_{p-2} \\ m_i \bmod k_{p-1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

We again have $A_1\left(R_{\vec{m}}(n_0)\right) = A_2\left(R_{\vec{k}}(n_0)\right)$.
But

$$A_1\left(R_{\vec{m}}(n_0)\right) = s_{i-1}$$

and

$$\begin{aligned} A_2\left(R_{\vec{k}}(n_0)\right) &= \sum_{r=p-1}^{j} t_r \left(\left\lfloor \frac{m_i}{\prod_{l=r+1}^{p-1} k_l} \right\rfloor \bmod k_r \right) \\ &= s_i \sum_{r=p-1}^{j} \prod_{l=r+1}^{p-1} k_l \left(\left\lfloor \frac{m_i}{\prod_{l=r+1}^{p-1} k_l} \right\rfloor \quad \bmod k_r \right) \\ &= s_i \cdot m_i. \end{aligned}$$

$\Rightarrow \quad s_{i-1} = s_i m_i$, which violates the definition of a max merged shape

$\Rightarrow \quad A_2$ is not linear, and therefore reshaping to $\vec{k}$ without appending a view is impossible.

Done #