

Sparsity in images, a basis better than Fourier basis

Background knowledge: Cooley-Tukey FFT

The Fourier basis in \mathbb{R}^n is given by the DFT matrix $(F_n)_{i,j} = \omega^{(i-1)(j-1)}$, where $\omega = e^{-2\pi i/n}$ is the primitive n -th root of unity. F_n is a Vandermonde matrix:

$$F_n = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 & \dots & 1 & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{\frac{n}{2}-1} & \omega^{\frac{n}{2}} & \dots & \omega^{n-2} & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{n-2} & \omega^n & \dots & \omega^{2n-4} & \omega^{2n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \omega^{\frac{n}{2}-1} & \omega^{n-2} & \dots & \omega^{(\frac{n}{2}-1)^2} & \omega^{(\frac{n}{2}-1)\frac{n}{2}} & \dots & \omega^{(\frac{n}{2}-1)(n-2)} & \omega^{(\frac{n}{2}-1)(n-1)} \\ \hline 1 & \omega^{\frac{n}{2}} & \omega^n & \dots & \omega^{\frac{n}{2}(\frac{n}{2}-1)} & \omega^{(\frac{n}{2})^2} & \dots & \omega^{\frac{n}{2}(n-2)} & \omega^{\frac{n}{2}(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \omega^{n-1} & \omega^{2n-2} & \dots & \omega^{(n-1)(\frac{n}{2}-1)} & \omega^{(n-1)\frac{n}{2}} & \dots & \omega^{(n-1)(n-2)} & \omega^{(n-1)^2} \end{pmatrix} \cdot (1)$$

The Cooley-Tukey FFT is a divide-and-conquer algorithm for computing the DFT. For simplicity, we assume n is a power of 2, such that we can divide the matrix into 4 blocks:

- The odd columns (blue background), top half:

$$F_{\text{odd, top}} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & \omega^2 & \dots & \omega^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{n-2} & \dots & \omega^{(\frac{n}{2}-1)(n-2)} \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & (\omega^2) & \dots & (\omega^2)^{\frac{n}{2}-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (\omega^2)^{\frac{n}{2}-1} & \dots & (\omega^2)^{(\frac{n}{2}-1)(\frac{n}{2}-1)} \end{pmatrix} = F_{\frac{n}{2}} \quad (2)$$

- The even columns (white background), top half:

$$F_{\text{even, top}} = D_{\frac{n}{2}} F_{\frac{n}{2}} \quad (3)$$

where $D_n = \text{diag}(1, \omega, \omega^2, \dots, \omega^{n-1})$.

- The odd columns (blue background), bottom half:

$$F_{\text{odd, bottom}} = F_{\frac{n}{2}}. \quad (4)$$

Note $\omega^n = 1$ is ignored.

- The even columns (white background), bottom half:

$$F_{\text{even, bottom}} = -D_{\frac{n}{2}} F_{\frac{n}{2}}, \quad (5)$$

where the minus sign comes from $\omega^{\frac{n}{2}} = -1$.

Finally, we arrive at the Cooley-Tukey FFT given by:

$$F_n \mathbf{x} = \begin{pmatrix} I_{\frac{n}{2}} & D_{\frac{n}{2}} \\ I_{\frac{n}{2}} & -D_{\frac{n}{2}} \end{pmatrix} \begin{pmatrix} F_{\frac{n}{2}} & 0 \\ 0 & F_{\frac{n}{2}} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{\text{odd}} \\ \mathbf{x}_{\text{even}} \end{pmatrix} \quad (6)$$

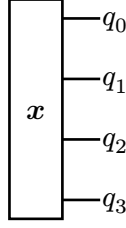
where \mathbf{x}_{odd} and \mathbf{x}_{even} contain the odd and even indexed elements of \mathbf{x} , respectively. It indicates that the discrete Fourier transformation in \mathbb{R}^n can be decomposed into two smaller discrete Fourier transformations in $\mathbb{R}^{\frac{n}{2}}$ with a diagonal matrix D_n in between. Note applying diagonal matrices can be done in $O(n)$ operations, this decomposition leads to the recurrence relation $T(n) = 2T(\frac{n}{2}) + O(n)$, which solves to $O(n \log n)$ total operations.

The inverse transformation is given by $F_n^\dagger \mathbf{x}/n$. The DFT matrix is unitary up to a scale factor:

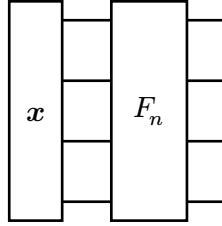
$$F_n F_n^\dagger = nI.$$

Tensor network representation of the Cooley-Tukey FFT

This section requires preliminary knowledge of tensor networks (TODO: add reference). In tensor network diagram, a vector of size $n = 2^k$ can be represented as a tensor with k indices, denoting the basis index $i = 2^0 q_0 + 2^1 q_1 + \dots + 2^{k-1} q_{k-1}$.

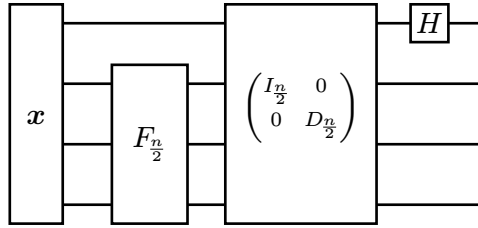


In the following, we aim to find a tensor network decomposition for the linear map F_n :



Step 1: To start, the equation Equation 6 can be represented as the following tensor network:

$$F_n x = \left(\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes I_{\frac{n}{2}} \right) \begin{pmatrix} I_{\frac{n}{2}} & 0 \\ 0 & D_{\frac{n}{2}} \end{pmatrix} \begin{pmatrix} F_{\frac{n}{2}} x_{\text{odd}} \\ F_{\frac{n}{2}} x_{\text{even}} \end{pmatrix}, \quad (7)$$

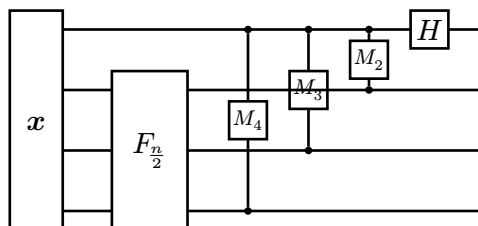


where $H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is a Hadamard matrix (upto a constant factor).

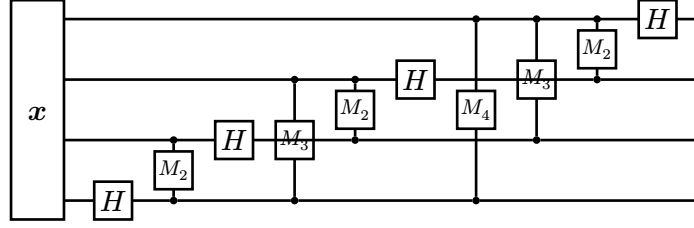
Step 2: Then, we will decompose the diagonal matrix $\begin{pmatrix} I_{\frac{n}{2}} & 0 \\ 0 & D_{\frac{n}{2}} \end{pmatrix}$ into a tensor network. This diagonal matrix corresponds to operation: if q_0 is 0 (odd index), the no operation is applied, otherwise (even index) the operation $D_{\frac{n}{2}}$ is applied. Observe that $D_n = \text{diag}(1, \omega^{\frac{n}{2}}) \otimes \text{diag}(1, \omega, \omega^2, \dots, \omega^{\frac{n}{2}-1}) = \text{diag}(1, \omega^{\frac{n}{2}}) \otimes \text{diag}(1, \omega^{\frac{n}{4}}) \otimes \dots \otimes \text{diag}(1, \omega)$. We have

$$\begin{pmatrix} I_{\frac{n}{2}} & 0 \\ 0 & D_{\frac{n}{2}} \end{pmatrix} = \text{ctrl}_0(\text{diag}(1, \omega^{\frac{n}{4}})_1) \text{ctrl}_0(\text{diag}(1, \omega^{\frac{n}{8}})_2) \dots \text{ctrl}_0(\text{diag}(1, \omega)_{\log_2 n}), \quad (8)$$

where $\text{ctrl}_i(A_j)$ means the target operation applied on A_j is applied only if bit q_i is 1. Here, since the controlled gate is diagonal, it can be represented as a matrix connecting two variables:



In this diagram, $M_k = \begin{pmatrix} 1 & 1 \\ 1 & e^{i\pi/2^{k-1}} \end{pmatrix}$ connects the two qubits involved in the controlled operation, which effectively multiplies a phase factor $e^{i\pi/2^{k-1}}$ if two bits are both in state 1. By recursively decomposing the $F_{\frac{n}{2}}$ tensor, we can obtain the following tensor network.



Direct evaluation of this tensor network takes $O(n \log^2 n)$ operations. By respecting the fact that the *controlled phase* operation is a diagonal matrix, we can merge these operations and further reduce the complexity to $O(n \log(n))$.

Entangled Fourier Basis: XY Correlation

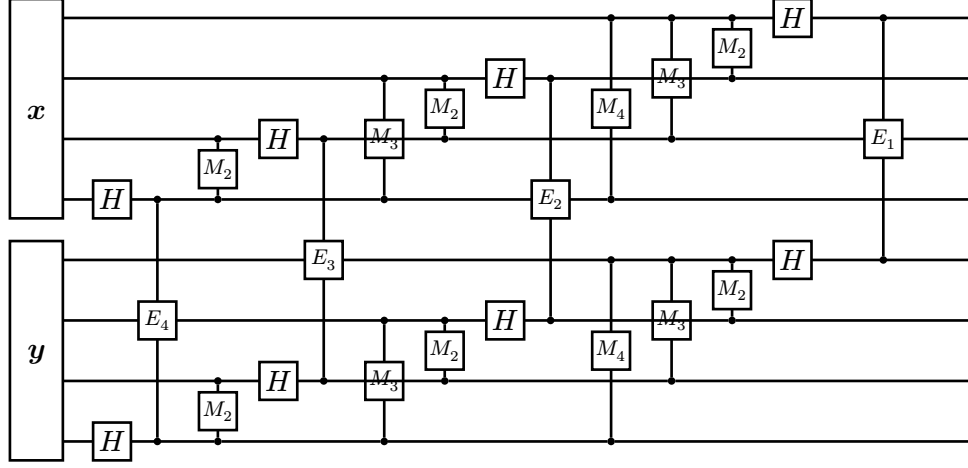
In the standard 2D Fourier transform, the x and y coordinates are processed independently. For an image of size $2^n \times 2^n$ (i.e., square images with $m = n$), we apply QFT on the n row qubits and separately on the n column qubits. This independence assumption is often suboptimal for natural images where spatial correlations exist between rows and columns.

We propose an *entangled QFT basis* that introduces controlled-phase gates between x and y qubits after each layer of the QFT circuit. For the square case $m = n$, we use a *one-to-one* entanglement structure where each x qubit x_k is coupled with the corresponding y qubit y_k . The entanglement gate E_k has exactly the same form as the M gate defined in the previous section:

$$E_k = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\varphi_k} \end{pmatrix} \quad (9)$$

acting on qubits (x_{n-k}, y_{n-k}) , where φ_k is a learnable phase parameter. Similar to the M_k gate which multiplies a phase factor when both control and target qubits are in state $|1\rangle$, the entanglement gate E_k multiplies a phase $e^{i\varphi_k}$ when both $x_{n-k} = 1$ and $y_{n-k} = 1$. The key difference is that while M_k uses the fixed phase $\pi/2^{k-1}$ determined by the QFT structure, E_k uses a learnable phase φ_k that can be optimized to capture image-specific correlations.

The circuit structure for $n = 4$ qubits per dimension is shown below. Each row qubit x_k first passes through its QFT circuit (Hadamard gate H followed by controlled-phase gates M_j with other row qubits), and similarly for column qubits y_k . The entanglement gates E_k (shown as boxes connecting x and y qubit lines) are applied after the Hadamard layers, coupling the corresponding row and column qubits:



Summarizing the gate applications for each qubit:

- **Row qubit x_k** : Hadamard gate H , followed by controlled-phase gates M_j with qubits x_0, \dots, x_{k-1} (standard QFT structure), then entanglement gate E_{n-k} with column qubit y_k
- **Column qubit y_k** : Hadamard gate H , followed by controlled-phase gates M_j with qubits y_0, \dots, y_{k-1} (standard QFT structure), then entanglement gate E_{n-k} with row qubit x_k

For a square $2^n \times 2^n$ image encoded with n row qubits and n column qubits (total $2n$ qubits), we add exactly n entanglement gates $\{E_1, E_2, \dots, E_n\}$, one for each pair of corresponding row/column qubits.

The total transformation becomes:

$$\mathcal{T}_{\text{entangled}} = U_{\text{entangle}} \cdot (F_n \otimes F_n) \quad (10)$$

where $U_{\text{entangle}} = \prod_{k=1}^n E_k$ is the product of all entanglement gates acting on qubit pairs (x_{n-k}, y_{n-k}) , and F_n is the n -qubit QFT applied along each spatial dimension.

Key advantages of this approach:

- Captures diagonal features and cross-dimensional patterns common in natural images
- Maintains $O(n \log n)$ computational complexity in the number n of qubits per spatial dimension (equivalently $O(N \log N)$ for linear image size $N = 2^n$), matching the standard QFT
- Adds exactly n additional real-valued learnable parameters φ_k (one phase per qubit pair), i.e., $O(n)$ in n
- Reduces to standard 2D QFT when all entanglement phases $\varphi_k = 0$

Alternative Basis: Time Evolving Block Decimation (TEBD)

Time Evolving Block Decimation (TEBD) is a tensor network ansatz originally developed for simulating 1D quantum many-body systems. It employs a *brickwork* pattern of nearest-neighbor two-qubit gates applied in alternating layers.

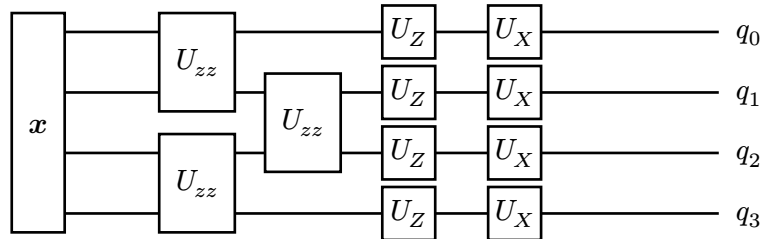


Figure 8: TEBD matrix block for Trotter approximation.

The Trotter-Suzuki decomposition approximates time evolution as:

$$e^{-i\hat{H}\eta} \approx \prod_{i=1}^N e^{-ih_X X_i \eta} \prod_{i=1}^{N-1} e^{-iJ Z_i Z_{i+1} \eta} \prod_{i=1}^N e^{-ih_Z Z_i \eta} \quad (11)$$

Notation: \otimes = tensor product (independent subsystems); \prod = matrix multiplication (composition).

Exact U_{ZZ} Decomposition of CP Gates

The QFT controlled-phase gate $CP_k = \text{diag}(1, 1, 1, e^{i\varphi_k})$ with $\varphi_k = \pi/2^{k-1}$ decomposes exactly as:

$$CP_k = e^{i\varphi_k/4} \left(R_Z^{(i)}(\varphi_k/2) \otimes R_Z^{(j)}(\varphi_k/2) \right) U_{ZZ}(-\varphi_k/4) \quad (12)$$

where $R_Z(\theta) = e^{-i\theta Z/2}$ and $U_{ZZ}(\theta) = e^{-i\theta Z_i \otimes Z_j}$. The $R_Z \otimes R_Z$ is a tensor product; the full expression is their composition with U_{ZZ} .

Connection to the Diagonal Matrix D_n

The twiddle factor matrix has **tensor product** form:

$$D_n = \text{diag}(1, \omega^{n/2}) \otimes \text{diag}(1, \omega^{n/4}) \otimes \dots \otimes \text{diag}(1, \omega) \quad (13)$$

In the FFT, D_n appears controlled:

$$\begin{pmatrix} I_{n/2} & 0 \\ 0 & D_{n/2} \end{pmatrix} = \prod_k \text{ctrl}_0 \left(\text{diag}(1, \omega^{n/2^{k+1}})_k \right) \quad (14)$$

Each ctrl_0 operation is a CP gate, decomposable via Equation 12:

$$\text{ctrl}_0 \left(\text{diag}(1, e^{i\varphi})_j \right) = e^{i\varphi/4} \left(R_Z^{(0)}(\varphi/2) \otimes R_Z^{(j)}(\varphi/2) \right) U_{ZZ}^{(0,j)}(-\varphi/4) \quad (15)$$

This gives an **exact** TEBD-style implementation of the QFT using U_{ZZ} gates and R_Z rotations:

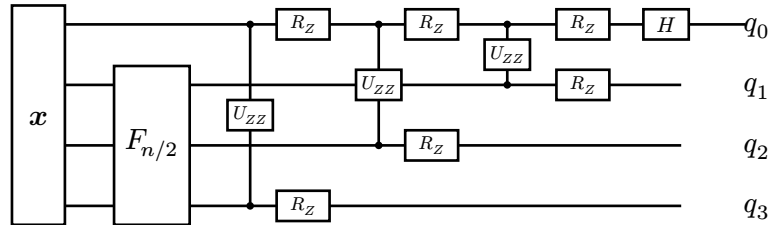


Figure 9: TEBD-style decomposition of one QFT recursion level. Input \mathbf{x} feeds into $F_{n/2}$ for qubits q_1, q_2, q_3 . Qubit q_0 passes through U_{ZZ} gates (with paired R_Z rotations) connecting to each lower qubit, followed by a final Hadamard gate H .

The remained $F_{n/2}$ could also be decomposed recursively, which is not drawn here explicitly.

Learning a better Fourier basis

Observing that in this representation, tensor parameters can be tuned without affecting the computational complexity, e.g. the parameters in M_k and H . Can we find a transformation better than the Fourier basis? Or is Fourier basis already optimal for image processing?

Intuitively, the fourier basis is not optimal for image processing, because:

- the fourier basis assumes periodic boundary condition, which is not suitable for image processing.
- the 2d fourier basis assumes the X and Y coordinates are independent, which is not suitable for image processing.

Tasks

- Create an image dataset $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$.

- Create a tensor network transformation based on the above QFT circuit, denoted as $\mathcal{T}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the parameters of the tensor network.
- Variationally optimize the circuit parameters to capture the sparsity of the image. The cost function is

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^N \|\mathbf{x}_i - \mathcal{T}(\boldsymbol{\theta})^{-1}(\text{truncate}(\mathcal{T}(\boldsymbol{\theta})(\mathbf{x}_i), k))\|_2^2 \quad (16)$$

Here, we can choose a different loss function to capture details in the image, e.g. the edges. For simplicity, we use the l_1 -norm instead:

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^N \|\mathcal{T}(\boldsymbol{\theta})(\mathbf{x}_i)\|_1 \quad (17)$$

This loss will encourage the tensor network to output a sparse pattern in the “moment space”. It is a standard trick that widely used in *compressed sensing*.

- In the 2D Fourier transformation, the X and Y coordinates are independent. Here we allow X and Y coordinates to correlate with each other in the tensor network basis.
- Add edge detection features.
- Compare the performance with the Fourier basis.