Chapter 5: Image Transforms

(from Anil. K. Jain)
2-D orthogonal and unitary transforms

- Orthogonal series expansion for an $N \times N$ image $u(m, n)$

$$v(k, l) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} u(m, n)a_{k,l}(m, n) \quad 0 \leq k, l \leq N - 1$$

$$u(m, n) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} v(k, l)a_{k,l}^*(m, n) \quad 0 \leq m, n \leq N - 1$$

- $v(k, l)$’s are the transform coefficients, $V \equiv \{v(k, l)\}$ represents the transformed image

- $\{a_{k,l}(m, n)\}$ is a set of orthonormal functions, representing the image transform
Orthonormality and completeness

\( \{a_{k,l}(m,n)\} \) must satisfy:

\[
\text{orthonormality} : \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a_{k,l}(m,n) a_{k',l'}^*(m,n) = \delta(k-k', l-l')
\]

\[
\text{completeness} : \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} a_{k,l}(m,n) a_{k,l}^*(m',n') = \delta(m-m', n-n')
\]
Matrix representation of image transform

\[ v(k,l) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} u(m,n) a_{k,l}(m,n) \]

\[ u(m,n) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} v(k,l) a_{k,l}^*(m,n) \]

\[ U = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} v(k,l) A_{k,l}^* \]

\[ v(k,l) = \langle U, A_{k,l}^* \rangle \]

- \( \langle F, G \rangle = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f(m,n) g^*(m,n) \) is the matrix inner product

- Image \( U \) can be described as a linear combination of \( N^2 \) matrix \( A_{k,l}^* \), \( k, l = 0, \ldots, N-1 \)

- \( A_{k,l}^* \) are called the basis images

- \( v(k,l) \) can be considered as the projection of \( u(m, n) \) on the \((k, l)\)-th basis image
Basis images
2-D Separable image transformation

- \( \{a_{k,l}(m,n)\} \) is separable

\[
a_{k,l}(m,n) = a_k(m)b_l(n) \equiv a(k,m)b(l,n)
\]

\[
v(k,l) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a(k,m)u(m,n)a(l,n) \rightarrow \mathbf{V} = \mathbf{AUA}^T
\]

\[
u(m,n) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} a^*(k,m)v(k,l)a^*(l,n) \rightarrow \mathbf{U} = \mathbf{A}^T\mathbf{VA}^*
\]

- \( \mathbf{A} \equiv \{a(k,m)\} \quad \mathbf{B} \equiv \{b(l,n)\} \) are unitary matrices
- Consider the above as “transforming columns of \( \mathbf{U} \) by \( \mathbf{A} \), then transforming rows of the result by \( \mathbf{A}^T \).
Example of image transform

Given \( A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \)

compute transform \( v = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ -2 \end{pmatrix} \)

Basis images \( A^*_{0,0} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \)

\( A^*_{0,1} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = A^*_{1,0} \quad A^*_{1,1} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \)

\( A^* V A^* = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 5 & -1 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \)
Properties of unitary transform

- Given \( \mathbf{v} = A \mathbf{u} \) (\( A \) is a unitary transform)
  - \( \| \mathbf{v} \|^2 = \| \mathbf{u} \|^2 \), i.e., energy-conserved
  - A unitary transformation is a rotation of the basis coordinates, or a rotation of the vector \( \mathbf{u} \) in \( N \)-Dimensional vector space
  - Unitary transform tends to concentrate most of the signal energy on a small set of transform coefficients

- \( \mathbf{\mu}_v = A \mathbf{\mu}_u \)
- \( \mathbf{R}_v = A \mathbf{R}_u A^* \)
Properties of unitary transform (cont)

- The transform coefficients tend to be decorrelated, i.e., the off-diagonal elements of $R_v$ are approaching zero.
- The entropy of a random vector is preserved under a unitary transformation (i.e., information preserved).
1-D DFT

- 1-D DFT pair (unitary transformation)
  \[ v(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} u(n)W_N^{kn} \quad k = 0, \ldots, N - 1 \]
  \[ u(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(k)W_N^{-kn} \quad n = 0, \ldots, N - 1 \]

- \( N \times N \) transform matrix \( F \) for \( v = Fu \)
  \[ F = \left\{ \frac{1}{\sqrt{N}} W_N^{kn} \right\}, \quad 0 \leq k, n \leq N - 1 \]
Properties of DFT

- Symmetry $\mathbf{F} = \mathbf{F}^T \rightarrow \mathbf{F}^{-1} = \mathbf{F}^*$ (unitary : $\mathbf{F}^{-1} = \mathbf{F}^{*T}$)
- FFT needs $O(N \log_2 N)$ operations (DFT needs $O(N^2)$)
- Real DFT is conjugate symmetrical about $N/2$
  \[ \nu^* \left( \frac{N}{2} - k \right) = \nu \left( \frac{N}{2} + k \right) \]
- $x_2(n)$ is the circular convolution between $h(n)$ and $x_1(n)$
  \[ \text{DFT} \{ x_2(n) \}_N = \text{DFT} \{ h(n) \}_N \text{DFT} \{ x_1(n) \}_N \]
- Extend the length of $h(n)(N')$ and $x_1(n)(N)$ with zeros to have the same length ($M \geq N' + N - 1$), the above equation can be used to compute linear convolution
2-D DFT

- 2-D DFT is a separable transform

\[ v(k, l) = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} u(m, n) W_N^{km} W_N^{ln} \quad 0 \leq k, l \leq N - 1 \]

\[ u(m, n) = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} v(k, l) W_N^{-km} W_N^{-ln} \quad 0 \leq m, n \leq N - 1 \]

- In matrix form

\[ V = FUF^T = FUF \]
Properties of 2-D DFT

- Symmetric and unitary $\Rightarrow F = F^T$ and $F^{-1} = F^*$
- Periodic $\Rightarrow \nu(k + N, l + N) = \nu(k, l)$
- $\mathbf{V} = \mathbf{FUF}^T$ needs $2N$ times of 1-D FFT
- DFT of real images is conjugate symmetrical with respect to $\left(\frac{N}{2}, \frac{N}{2}\right)$

$$\nu(k, l) = \nu^*(N - k, N - l)$$

- $N^2$ basis images

$$\frac{1}{N} \left\{ W^{- (km + ln)}_N, 0 \leq m, n \leq N - 1 \right\} \quad 0 \leq k, l \leq N - 1$$
Plots of DFT

Figure 5.6 Two-dimensional unitary DFT of a 256 × 256 image.
Cosine transform

- **DCT** (discrete cosine transform)

\[
c(k, n) = \begin{cases} 
\frac{1}{\sqrt{N}}, & k = 0, 0 \leq n \leq N - 1 \\
\sqrt{\frac{2}{N}} \cos \frac{\pi(2n+1)k}{2N}, & 1 \leq k \leq N - 1, 0 \leq n \leq N - 1 
\end{cases}
\]

- **8x8 2-D DCT**

![Cosine transform examples of monochrome images](a.png)

![Cosine transform examples of binary images](b.png)

Figure 5.11
Properties of DCT

- DCT is real and orthogonal \( C = C^* \) and \( C^{-1} = C^T \)
- DCT \( \neq \) Real \{DFT\}
- DCT can be calculated via FFT

\[
\nu(k) = \text{Re}[\alpha(k)W_{2N}^{k/2}DFT\{\tilde{u}(n)\}_N]
\]

\[
\begin{align*}
\tilde{u}(n) &= u(2n) \quad 0 \leq n \leq (\frac{N}{2}) - 1 \\
\tilde{u}(N - n - 1) &= u(2n + 1)
\end{align*}
\]

- For highly correlated data, DCT has good energy compaction. That is, for first-order stationary Markov sequence with correlation \( \rho \cong 1 \), DCT is approximate to KL transform
Sine transform

- DST (discrete sine transform)
  \[ \psi(k,n) = \sqrt{\frac{2}{N+1}} \sin \frac{\pi(k + 1)(n + 1)}{N + 1}, \quad 0 \leq k, n \leq N - 1 \]

- Properties
  - DST is real, symmetric, and orthogonal. \(\Rightarrow\) \(\psi^* = \psi = \psi^T = \psi^{-1}\)
  - DST and DST\(^{-1}\) are the same in the form (cf. DCT)
  - DST \(\neq\) Imagery \{DFT\}
  - For first-order stationary Markov sequence with correlation \(\rho \in (-0.5, 0.5)\), the energy compaction of DST is approximate to KL transform
(Walsh-) Hadamard transform

- Hadamard matrix

\[
H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
H_n = H_{n-1} \otimes H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}
\]

- Components of HT vector contain only 1 and -1
  - The number of transitions from 1 to -1 is called sequence (like \( \omega \) in the continuous case)

\[
H_3 = \frac{1}{\sqrt{8}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}
\]
(Walsh-) Hadamard transform (cont.)

- Hadamard transform pair:

\[
\begin{align*}
    v(k) &= \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} u(m)(-1)^{b(k,m)}, \quad 0 \leq k \leq N - 1 \\
    u(m) &= \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(k)(-1)^{b(k,m)}, \quad 0 \leq m \leq N - 1
\end{align*}
\]

\[b(k,m) = \sum_{i=0}^{n-1} k_i m_i \quad k_i, m_i = 0,1\]

- Notice the disordering effect in the sequency in HT spectrum when filtering will be performed.
Properties of Hadamard transform

- Real, symmetric, and orthogonal $\Rightarrow \mathbf{H} = \mathbf{H}^* = \mathbf{H}^T = \mathbf{H}^{-1}$
- Fast computation (only addition is needed)
- Can be decomposed into product of sparse matrix

$$\mathbf{H} = \mathbf{H}_n = \tilde{\mathbf{H}}^n \quad n = \log_2 N$$

$$\tilde{\mathbf{H}} = \frac{1}{\sqrt{2}} \begin{bmatrix}
1 & 1 & 0 & 0 & \ldots & 0 \\
0 & 0 & 1 & 1 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & \ldots & \ldots & 1 & 1 \\
1 & -1 & 0 & 0 & \ldots & 0 \\
0 & 0 & -1 & 0 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & 0 & \ldots & \ldots & 1 & -1
\end{bmatrix}$$

$$\mathbf{v} = \mathbf{H}\mathbf{u} = \tilde{\mathbf{H}}^n\mathbf{u} = \tilde{\mathbf{H}}\tilde{\mathbf{H}}\ldots\tilde{\mathbf{H}}\mathbf{u}$$

- For highly correlated images, Hadamard transform also has good energy compaction
Plot of Hadamard transform

Figure 5.13  Examples of Hadamard transforms.
KL transform of 1-D vector

For a real $N \times 1$ random vector $u$ with autocorrelation matrix $R$, KL transform of $u$ is defined as

$$v = \Phi^* u \quad u = \Phi \cdot v$$

where $R\phi_k = \lambda_k \phi_k$, $0 \leq k \leq N-1$, i.e., $\phi_k$'s are orthonormalized eigenvectors of $R$ and $\phi_k$ is the k-th column of $\Phi$.

$\Phi$ reduces $R$ to its diagonal form, i.e.,

$$\Phi^* R \Phi = \Lambda = \text{Diag}\{\lambda_k\}$$

If we replace $R$ with $R_0 \equiv R - \mu \cdot \mu^T$, then eigenmatrix of $R_0$ is the KL transform of $u-\mu$.

KL transform is independent of the image data themselves, but related to its 2nd-order statistics.
KL transform of 2-D image

- FFT, DCT, DST all are special cases of KL transform
- For an image \( u(m, n) \) of \( N \times N \) pixels, its autocorrelation function denoted as \( r(m, n; m', n') \) (or an \( N^2 \times N^2 \mathbf{R} \) matrix), the basis images of the KL transform are \( \psi_{k,l}(m,n) : \)

\[
\sum_{m'=0}^{N-1} \sum_{n'=0}^{N-1} r(m, n; m', n') \psi_{k,l}(m', n') = \lambda_{k,l} \psi_{k,l}(m, n) \quad 0 \leq k, l \leq N - 1
\]

or

\[
\mathbf{R} \psi_i = \lambda_i \psi_i \quad i = 0, \ldots, N^2 - 1
\]

\( \psi_i \) is the \( N^2 \times 1 \) vector representation of \( \psi_{k,l}(m,n) \)
**KL transform of 2-D image (cont.)**

- If $\mathbf{R}$ matrix is separable, then $\Psi = \{\psi_i\}$ of $N^2 \times N^2$ is also separable

  $$\mathbf{R} = \mathbf{R}_1 \otimes \mathbf{R}_2 \quad \Psi = \Phi_1 \otimes \Phi_2$$

  where

  $$\Phi_j \mathbf{R}_j \Phi_j^T = \Lambda_j, \quad j = 1, 2$$

  $\Rightarrow$ KL transform : $\mathbf{v} = \Psi^T \mathbf{u}$ or $\mathbf{V} = \Phi_1^T \mathbf{U} \Phi_2^T$

- Advantages of this separability : reduce the eigenvalue problem of $\mathbf{R}$ matrix ($N^2 \times N^2$) into two eigenvalue problems of $\mathbf{R}_1, \mathbf{R}_2$ ($N \times N$)
Properties of the KL transform

- **Decorrelation**
  - \( \{v(k), k = 0, ..., N - 1\} \) are uncorrelated and have zero mean
  - \( \Phi \) is not unique, not necessarily be the eigenmatrix of \( R \)

- **Minimum basis restriction error**

\[
\text{Signal } u \text{ is transformed via } A \text{ to be } v, \text{ clip the first } m \text{ values by } I_m \\
\text{to form } w, \text{ and transformed back to } z \text{ via } B. \text{ The basis restriction error is defined as}
\]

\[
J_m \equiv \frac{1}{N} E \left( \sum_{n=0}^{N-1} |u(n) - z(n)|^2 \right) = \frac{1}{N} Tr[E\{(u - z)(u - z)^T\}]
\]
Properties of the KL transform (cont.)

- Find \( A_{KL} \) and \( B_{KL} \) (KL transform) such that \( J_m \) is minimized for each value of \( m \in [1, N] \). (KL is effective in compacting the energy)

- Select \( A = \Phi^* \) and \( B = \Phi \), \( AB = I \), and columns of \( \Phi \) are sorted according to their corresponding eigenvalues \( \lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{N-1} \)

- Minimum transmission rate at a given distortion

  - In a noisy channel, given a distortion specification,
    \[
    D = \frac{1}{N} E[(u - \hat{u})^T(u - \hat{u})]
    \]

  KL transform can, among all unitary transformation \( A \) (include \( A = I \)), achieve the minimum transmission rate 
  \[
  R(\Phi^*) \leq R(A)
  \]
An example of KL-transform

- A 2×1 vector $\mathbf{u}$, whose covariance matrix is
  \[
  \mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad |\rho| < 1
  \]

  The KL transform is
  \[
  \mathbf{\Phi}^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}
  \]

  \[
  \mathbf{v} = \mathbf{\Phi} \cdot \mathbf{u} \implies E\{[v(0)]^2\} = \lambda_0 = 1 + \rho, \quad E\{[v(1)]^2\} = 1 - \rho
  \]

  \[
  R(\mathbf{\Phi}) = \frac{1}{2} \left[ \max(0, \frac{1}{2} \log \frac{1+\rho}{\theta}) + \max(0, \frac{1}{2} \log \frac{1-\rho}{\theta}) \right]
  \]

  comparing
  \[
  R(\mathbf{I}) = \frac{1}{4} [-2 \log \theta], \quad 0 < \theta < 1
  \]

  \[
  \implies R(\mathbf{\Phi}) < R(\mathbf{I}) \quad \text{Let } \theta \text{ be small, say } \theta < 1 - |\rho|
  \]
Comparison of energy distribution between transforms

- For first-order stationary Markov process with large \( \rho \), DCT is comparable to KL
- With small \( \rho \), DST is comparable to KL
- For any \( \rho \), we can find a faster sinusoidal transform (DCT or DST) to substitute for the optimal KL transform which needs covariance matrix to compute transform basis vectors
Figure 5.18 Distribution of variances of the transform coefficients (in decreasing order) of a stationary Markov sequence with $N = 16, \rho = 0.95$ (see Example 5.9).
<table>
<thead>
<tr>
<th>$k$</th>
<th>KL</th>
<th>Cosine</th>
<th>Sine</th>
<th>Unitary DFT</th>
<th>Hadamard</th>
<th>Haar</th>
<th>Slant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.946</td>
<td>1.943</td>
<td>1.688</td>
<td>1.100</td>
<td>1.644</td>
<td>1.644</td>
<td>1.904</td>
</tr>
<tr>
<td>2</td>
<td>0.615</td>
<td>0.648</td>
<td>1.352</td>
<td>0.292</td>
<td>0.544</td>
<td>0.487</td>
<td>0.641</td>
</tr>
<tr>
<td>3</td>
<td>0.292</td>
<td>0.295</td>
<td>0.421</td>
<td>0.139</td>
<td>0.431</td>
<td>0.487</td>
<td>0.233</td>
</tr>
<tr>
<td>4</td>
<td>0.171</td>
<td>0.174</td>
<td>0.463</td>
<td>0.086</td>
<td>0.153</td>
<td>0.144</td>
<td>0.173</td>
</tr>
<tr>
<td>5</td>
<td>0.114</td>
<td>0.114</td>
<td>0.181</td>
<td>0.062</td>
<td>0.152</td>
<td>0.144</td>
<td>0.172</td>
</tr>
<tr>
<td>6</td>
<td>0.082</td>
<td>0.083</td>
<td>0.216</td>
<td>0.051</td>
<td>0.149</td>
<td>0.144</td>
<td>0.072</td>
</tr>
<tr>
<td>7</td>
<td>0.063</td>
<td>0.063</td>
<td>0.098</td>
<td>0.045</td>
<td>0.121</td>
<td>0.144</td>
<td>0.072</td>
</tr>
<tr>
<td>8</td>
<td>0.051</td>
<td>0.051</td>
<td>0.116</td>
<td>0.043</td>
<td>0.051</td>
<td>0.050</td>
<td>0.051</td>
</tr>
<tr>
<td>9</td>
<td>0.043</td>
<td>0.043</td>
<td>0.060</td>
<td>0.045</td>
<td>0.051</td>
<td>0.050</td>
<td>0.051</td>
</tr>
<tr>
<td>10</td>
<td>0.037</td>
<td>0.037</td>
<td>0.067</td>
<td>0.051</td>
<td>0.051</td>
<td>0.050</td>
<td>0.051</td>
</tr>
<tr>
<td>11</td>
<td>0.033</td>
<td>0.033</td>
<td>0.040</td>
<td>0.062</td>
<td>0.051</td>
<td>0.050</td>
<td>0.051</td>
</tr>
<tr>
<td>12</td>
<td>0.030</td>
<td>0.030</td>
<td>0.042</td>
<td>0.086</td>
<td>0.051</td>
<td>0.050</td>
<td>0.031</td>
</tr>
<tr>
<td>13</td>
<td>0.028</td>
<td>0.028</td>
<td>0.031</td>
<td>0.139</td>
<td>0.051</td>
<td>0.050</td>
<td>0.031</td>
</tr>
<tr>
<td>14</td>
<td>0.027</td>
<td>0.027</td>
<td>0.029</td>
<td>0.292</td>
<td>0.050</td>
<td>0.050</td>
<td>0.031</td>
</tr>
<tr>
<td>15</td>
<td>0.026</td>
<td>0.026</td>
<td>0.026</td>
<td>1.100</td>
<td>0.043</td>
<td>0.050</td>
<td>0.031</td>
</tr>
</tbody>
</table>

**TABLE 5.2** Variances $\sigma_k^2$ of Transform Coefficients of a Stationary Markov Sequence with $\rho = 0.95$ and $N = 16$. See Example 5.9.
Singular Value Decomposition (SVD)

- For an \( N \times M \) matrix \( U \) (\( M \leq N \)), we can find \( r \) orthogonal \( M \times 1 \) eigenvector \( \{ \phi_m \} \) by using \( U^T U \), or find \( r \) orthogonal \( N \times 1 \) eigenvector \( \{ \psi_m \} \) by using \( UU^T \) (\( r \) is the rank of \( U \))

\[
U^T U \phi_m = \lambda_m \phi_m, \quad m = 1, \ldots, r
\]
\[
UU^T \psi_m = \lambda_m \psi_m, \quad m = 1, \ldots, r
\]

Matrix \( U \) can be described as

\[
U = \Psi \cdot \Lambda^{1/2} \cdot \Phi^T = \sum_{m=1}^{r} \sqrt{\lambda_m} \psi_m \phi_m^T
\]

\[
\Lambda^{1/2} = \begin{bmatrix}
\sqrt{\lambda_1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & \sqrt{\lambda_r}
\end{bmatrix}
\]

Called the spectral representation, outer product expansion, or singular value decomposition (SVD) of \( U \)

Note. In KL, \( R \) is symmetrical, but \( U \) is not
Singular Value Decomposition (cont)

- SVD of $\mathbf{U}$ (diagonalization): $\Lambda^{1/2} = \Psi^T \mathbf{U} \Phi$
  - $\mathbf{U}$ can be described as the composition of $r$ matrix (corresponding to different $\lambda_m$, called $\mathbf{U}$’s singular values)

- If $\mathbf{U}$ is an $N \times M$ image, $\mathbf{U}$ can be described as an $NM \times 1$ vector, also as $\{\lambda_m^{1/4} \psi_m, \lambda_m^{1/4} \phi_m; m = 1, ..., r\}$

 needling totally $(M+N)r$ data and be more efficient than $N \times M$ or $NM \times 1$ if $r << M$

Note: $\mathbf{U}$ is not symmetrical
Properties of SVD

- We can compute \( \psi_m \equiv \frac{1}{\sqrt{\lambda_m}} u \phi_m \), \( m = 1, \ldots, r \), when knowing \( \phi_m \) due to the same eigenvalue.

- Energy conservation

\[
E = \sum_m \sum_n u^2(m, n) = \sum_m \lambda_m
\]

- Denote \( U_k \equiv \sum_{m=1}^{k} \sqrt{\lambda_m} \psi_m \phi_{m}^T \), \( k \leq r \), as \( U \)'s \( k \) partial-sum (\( \lambda_m \) is ordered)

\( U_k \) is the optimal rank-\( k \) description, with approximation error

\[
\varepsilon_k^2 = \sum_{m=1}^{M} \sum_{n=1}^{N} |u(m, n) - u_k(m, n)|^2 = \sum_{m=k+1}^{r} \lambda_m, \quad k = 1, \ldots, r
\]

- The energy of each SVD spectral band \( \sqrt{\lambda_m} \psi_m \phi_{m}^T \) is \( \lambda_m \)
Properties of SVD (cont)

- For a given image, SVD can compact the largest energy for a specified number of transform coefficients (optimized independently).
  - The maximal number of basis images for SVD is $N$, while for other transforms, it is $N^2$.
- For KL transform, it optimizes one set of ensemble with the same covariance matrix, resulting in the largest “average energy” in probabilistic sense.
Properties of SVD (cont)

- Example:

\[
\begin{bmatrix}
1 & 2 \\
2 & 1 \\
1 & 3 \\
\end{bmatrix}
\rightarrow
\begin{cases}
\lambda_1 = 18.06 \\
\lambda_2 = 1.94
\end{cases}
\]

\[
\begin{bmatrix}
\lambda_1 & \phi_1 \\
\lambda_2 & \phi_2
\end{bmatrix} =
\begin{bmatrix}
0.5019 \\
0.8649
\end{bmatrix},
\begin{bmatrix}
0.8649 \\
-0.5019
\end{bmatrix}
\rightarrow
\begin{bmatrix}
4.25 & 0 \\
0 & 1.39
\end{bmatrix}
\]

\[
U_1 = \sqrt{\lambda_1} \psi_1 \phi_1^T =
\begin{bmatrix}
1.120 & 1.94 \\
0.935 & 1.62 \\
1.549 & 2.70
\end{bmatrix}
\]

Using DCT: \( V = C_3 U C_2^T \) =
\[
\frac{1}{\sqrt{12}}
\begin{bmatrix}
\sqrt{2} & \sqrt{2} & \sqrt{2} \\
\sqrt{3} & 0 & -\sqrt{3} \\
1 & -2 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 2 \\
1 & 3 \\
1 & -1
\end{bmatrix}
= \frac{1}{\sqrt{12}}
\begin{bmatrix}
10\sqrt{2} & -2\sqrt{2} \\
-\sqrt{3} & \sqrt{3} \\
-1 & -5
\end{bmatrix}
\]

\[
\sum_k \sum_l \psi^2 (k, l) = \lambda_1 + \lambda_2
\]