Journal of Mathematical Research with Applications
Jan., 2017, Vol. 37, No. 1, pp. 19–28
DOI:10.3770/j.issn:2095-2651.2017.01.002
Http://jmre.dlut.edu.cn

The Haar Wavelet Analysis of Matrices and Its Applications

Xiquan SHI
Department of Mathematical Sciences, Delaware State University, Dover 19901*, USA*

Dedicated to Professor Renhong WANG on the Occasion of His Eightieth Birthday

Abstract It is well known that Fourier analysis or wavelet analysis is a very powerful and useful tool for a function since they convert time-domain problems into frequency-domain problems. Are there similar tools for a matrix? By pairing a matrix to a piecewise function, a Haar-like wavelet is used to set up a similar tool for matrix analyzing, resulting in new methods for matrix approximation and orthogonal decomposition. By using our method, one can approximate a matrix by matrices with different orders. Our method also results in a new matrix orthogonal decomposition, reproducing Haar transformation for matrices with orders of powers of two. The computational complexity of the new orthogonal decomposition is linear. That is, for an $m \times n$ matrix, the computational complexity is $O(mn)$. In addition, when the method is applied to k-means clustering, one can obtain that k-means clustering can be equivalently converted to the problem of finding a best approximation solution of a function. In fact, the results in this paper could be applied to any matrix related problems. In addition, one can also employ other wavelet transformations and Fourier transformation to obtain similar results.

Keywords wavelet analysis; Fourier analysis; matrix decomposition; k-means clustering; linear equation

MR(2010) Subject Classification 41A30; 42C40; 62H25; 65F25; 62H30; 65T60

1. Wavelet transformation of matrix

It is well known that the Fourier series is very important and it decomposes a function into infinite sum of sine and cosine functions and the coefficients of sine and cosine functions contain the spectral information of the original function. Wavelet transformation is an update of Fourier transformation (series) and it decomposes a function into infinite sum of wavelets. Since the coefficients of the wavelets (Fourier transformation) are the spectrums of the functions, the wavelet (Fourier) transformation converts time-domain problems into frequency-domain problems. In this section a Haar-like wavelet transform is employed to set up a new matrix analyzing method. The same as the function's wavelet (Fourier) transformation, the new method can also convert time-domain problems of a matrix into its frequency-domain problems, by resulting in a new orthogonal decomposing of a matrix into a matrix composed of its spectrums.

Received August 28, 2016; Accepted September 23, 2016
E-mail address: xqshi2000@gmail.com

For an $m \times n$ matrix $A = (a_{i,j})_{0 \le i \le m-1,\ 0 \le j \le n-1}$, we define function f_A on $[0,a] \times [0,b]$ by

$$f_A(x,y) = a_{i,j},\ \frac{ia}{m} < x < \frac{(i+1)a}{m},\ \frac{jb}{n} < y < \frac{(j+1)b}{n},\ \ 0 \le i \le m-1,\ 0 \le j \le n-1. \quad (1)$$

$a = 1$, $b = \frac{n}{m}$ or $a = \frac{m}{n}$, $b = 1$ is reasonable, but we set $a = b = 1$ for convenience. That is, we assign the matrix A to a piecewise constant function with possibly break lines $x = \frac{i}{m}$, $1 \le i \le m-1$ and $y = \frac{i}{n}$, $1 \le i \le n-1$. A and f_A are called dual pair of each other.

For the value of f_A at a point on the beak lines, although definition is acceptable, we define it to be the average of the function values of its neighbors. For example, we define

$$f(\frac{i}{m}, \frac{j}{n}) = \frac{1}{4}(a_{i-1,j-1} + a_{i-1,j} + a_{i,j-1} + a_{i,j}),\ 0 < i < m,\ 0 < j < n,$$
$$f(\frac{i}{m}, y) = \frac{1}{2}(a_{i-1,j} + a_{i,j}),\ 0 < i < m,\ \frac{j}{n} < y < \frac{j+1}{n}.$$

To obtain a Haar-like wavelet, we define

$$\begin{aligned} h_0(x) =& 1,\ 0 < x < 1,\ h_0(x) = 0, x < 0 \text{ or } x > 1,\ h_0(0) = h_0(1) = 0.5, \\ h(x) =& 1,\ 0 < x < 0.5,\ h(x) = -1,\ 0.5 < x < 1,\ h_0(x) = 0, x < 0 \text{ or } x > 1, \\ & h_0(0) = 0.5,\ h(0.5) = 0,\ h_0(1) = -0.5. \end{aligned} \quad (2)$$

Then, we define

$$h_{i,j}(x) = 2^{\frac{i}{2}} h(2^i x - j) = h_{2^i + j}(x),\quad i \ge 0,\ 0 \le j \le 2^i - 1. \quad (3)$$

It is easy to see that $h_1 = h_{0,0} = h$. The following lemma is obvious.

Lemma 1.1 *$\{h_i, i \ge 0\}$ is an orthonormal basis of $L^2[0,1]$. That is, $\{h_i, i \ge 0\}$ is a Haar wavelet-like basis of $L^2[0,1]$.*

It is obvious that $f_A \in L^2([0,1] \times [0,1])$ and thus there holds

$$f_A(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{i,j} h_i(x) h_j(y),\quad \text{in } L^2 \quad (4)$$

where

$$c_{i,j} = \int_0^1 h_j(y)\mathrm{d}y \int_0^1 f_A(x,y) h_i(x)\mathrm{d}x = \int_0^1 h_j(y)\mathrm{d}y \int_0^1 f_A(x,y) h_{s,t}(x)\mathrm{d}x.$$

It is well known that, like Fourier series, $c_{i,j}$ reflects the frequency information of f_A contained in the support region of $h_i(x)h_j(y)$, i.e., $c_{i,j}$ reflects the corresponding spectrum information of f_A. Equation (4) is called the (Harr) wavelet transformation of the matrix A. For the expansion of f_A in (4), we have the following lemma.

Lemma 1.2 *$f_A(x,y) = \sum_{i=0}^{m-1} \sum_{j=0}^{\infty} c_{i,j} h_i(x) h_j(y)$ if $m = 2^M$ for some natural number M, $f_A(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{n-1} c_{i,j} h_i(x) h_j(y)$ if $n = 2^N$ for some natural number N, and*

$$f_A(x,y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} c_{i,j} h_i(x) h_j(y) \quad (5)$$

if both $m = 2^M$ and $n = 2^N$ for some natural numbers M and N.

Proof If $i \geq m = 2^M$, then there exists $s \geq M$ and $0 \leq t \leq 2^s - 1$, such that $i = 2^s + t$. Therefore, there holds

$$c_{i,j} = \int_0^1 h_j(y)\mathrm{d}y \int_0^1 f_A(x,y)h_i(x)\mathrm{d}x = \int_0^1 h_j(y)\mathrm{d}y \int_0^1 f_A(x,y)h_{s,t}(x)\mathrm{d}x.$$

Since the support of $h_{s,t}$ is $(\frac{t}{2^s}, \frac{t+1}{2^s}) \subset (\frac{t}{2^M}, \frac{t+1}{2^M}) = (\frac{t}{m}, \frac{t+1}{m})$. According to the definition, $f_A(x,y)$ is independent of x in the region $(\frac{t}{m}, \frac{t+1}{m}) \times [0,1]$. Therefore,

$$\int_0^1 f_A(x,y)h_{s,t}(x)\mathrm{d}x = f_A(x,y)\int_0^1 h_{s,t}(x)\mathrm{d}x = 0.$$

That is $c_{i,j} = 0$. The other two conclusions can be proved similarly. □

In fact, the following result shows that (4) holds point-wisely almost everywhere.

Theorem 1.3 *Let $B_x = \{\frac{i}{m}, 0 \leq i \leq m\}$, $B_y = \{\frac{j}{n}, 0 \leq j \leq n\}$ and $D = \{\frac{t}{2^s}, s \geq 1, 0 \leq t \leq 2^s\}$. Then,*

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} c_{i,j}h_i(\bar{x})h_j(\bar{y})$$

converges if both $\bar{x} \notin B_x$ and $\bar{y} \notin B_y$. Furthermore, if both $\bar{x} \notin B_x \bigcup D$ and $\bar{y} \notin B_y \bigcup D$, then

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} c_{i,j}h_i(\bar{x})h_j(\bar{y}) = f_A(\bar{x},\bar{y}), \tag{6}$$

i.e., (4) holds point-wisely almost everywhere. If $\bar{x} \in B_x$ or $\bar{y} \in B_y$, $\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} c_{i,j}h_i(\bar{x})h_j(\bar{y})$ usually diverges, but it is a bounded sequence.

Proof Only the most important case of both $\bar{x} \notin B_x \bigcup D$ and $\bar{y} \notin B_y \bigcup D$ is taken as an example to prove Theorem 1.3. Assuming $\bar{x} = \sum_{s=1}^{\infty}\frac{\varepsilon_s}{2^s}$ and $\bar{y} = \sum_{t=1}^{\infty}\frac{\eta_t}{2^t}$ are the binary representations of $\bar{x}, \bar{y} \in (0,1)$, where both ε_s and η_t equal 0 or 1. Since $\bar{x} \notin D$ and $\bar{y} \notin D$, both ε_s and η_t contain infinite many $0's$ and $1's$. Since $\bar{x} \notin B_x$ and $\bar{y} \notin B_y$, there exist i_0 and j_0 such that

$$\frac{i_0}{m} < \bar{x} < \frac{i_0+1}{m}, \quad \frac{j_0}{n} < \bar{y} < \frac{j_0+1}{n}.$$

For $i \geq 1$, let $i = 2^a + b$ with $0 \leq b \leq 2^a - 1$. Then

$$h_i(\bar{x}) = h_{a,b}(\bar{x}) = 2^{\frac{a}{2}}h(2^a\bar{x} - b) = 2^{\frac{a}{2}}h\Big(2^a\sum_{s=1}^{a}\frac{\varepsilon_s}{2^s} - b + \sum_{s=a+1}^{\infty}\frac{\varepsilon_s}{2^{s-a}}\Big)$$

$$= \begin{cases} (-1)^{\varepsilon_{a+1}}2^{\frac{a}{2}}, & \text{if } b = 2^a\sum_{s=1}^{a}\frac{\varepsilon_s}{2^s}, \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, for $j \geq 1$, let $j = 2^s + t$ with $0 \leq t \leq 2^s - 1$. Then

$$h_j(\bar{y}) = h_{s,t}(\bar{y}) = 2^{\frac{s}{2}}h(2^s\bar{y} - t) = 2^{\frac{s}{2}}h\Big(2^s\sum_{l=1}^{s}\frac{\eta_l}{2^l} - t + \sum_{l=s+1}^{\infty}\frac{\eta_l}{2^{l-s}}\Big)$$

$$= \begin{cases} (-1)^{\eta_{s+1}}2^{\frac{s}{2}}, & \text{if } t = 2^s\sum_{l=1}^{s}\frac{\eta_l}{2^l}, \\ 0, & \text{otherwise.} \end{cases}$$

Since $h_0(\bar{x}) = h_0(\bar{y}) = 1$, we have

$$\begin{aligned}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} c_{i,j}h_i(\bar{x})h_j(\bar{y}) =& c_{0,0} + \sum_{i=1}^{\infty} c_{i,0}h_i(\bar{x}) + \sum_{j=1}^{\infty} c_{0,j}h_j(\bar{y}) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} c_{i,j}h_i(\bar{x})h_j(\bar{y}) \\ =& c_{0,0} + \sum_{a=0}^{\infty}(-1)^{\varepsilon_{a+1}}2^{\frac{a}{2}}c_{2^a+b,0} + \sum_{j=1}^{\infty}(-1)^{\eta_{s+1}}2^{\frac{s}{2}}c_{0,2^s+t}+ \\ & \sum_{a=0}^{\infty}\sum_{s=0}^{\infty}(-1)^{\varepsilon_{a+1}+\eta_{s+1}}2^{\frac{s}{2}+\frac{s}{2}}c_{2^a+b,2^s+t}, \end{aligned} \tag{7}$$

where $b = 2^a\sum_{s=1}^{a}\frac{\varepsilon_s}{2^s}$ and $t = 2^s\sum_{l=1}^{s}\frac{\eta_l}{2^l}$. Since $\frac{i_0}{m} < \bar{x} < \frac{i_0+1}{m}$, $\frac{j_0}{n} < \bar{y} < \frac{j_0+1}{n}$, by denoting

$$x_0 = 0,\ x_a = \sum_{l=1}^{a}\frac{\varepsilon_l}{2^l},\quad y_0 = 0,\ y_s = \sum_{l=1}^{s}\frac{\eta_l}{2^l},$$

there exist $\bar{a}$ and $\bar{s}$, such that

$$\frac{i_0}{m} < x_{\bar{a}} < \bar{x} < x_{\bar{a}} + \frac{1}{2^{\bar{a}}} < \frac{i_0+1}{m},\quad \frac{j_0}{n} < y_{\bar{s}} < \bar{y} < y_{\bar{s}} + \frac{1}{2^{\bar{s}}} < \frac{j_0+1}{n}. \tag{8}$$

Thus, f_A depends only on y for $x \in [x_{\bar{a}}, x_{\bar{a}} + \frac{1}{2^{\bar{a}}}]$ and f_A depends only on x for $y \in [y_{\bar{s}}, y_{\bar{s}} + \frac{1}{2^{\bar{s}}}]$. Therefore, for $j \ge 0$ and $a \ge \bar{a}$ ($b = 2^a x_a$), it holds that

$$\begin{aligned} c_{2^a+b,j} =& \int_0^1 h_j(y)\mathrm{d}y\int_0^1 f_A(x,y)h_{a,b}(x)\mathrm{d}x = \int_0^1 h_j(y)\mathrm{d}y\int_{x_{\bar{a}}}^{x_{\bar{a}}+\frac{1}{2^{\bar{a}}}} f_A(x,y)h_{a,b}(x)\mathrm{d}x \\ =& \int_0^1 f_A(x,y)h_j(y)\mathrm{d}y\int_{x_{\bar{a}}}^{x_{\bar{a}}+\frac{1}{2^{\bar{a}}}} h_{a,b}(x)\mathrm{d}x = \int_0^1 f_A(x,y)h_j(y)\mathrm{d}y\int_0^1 h_{a,b}(x)\mathrm{d}x = 0. \end{aligned}$$

Similarly, for $i \ge 0$ and $s \ge \bar{s}$ ($t = 2^s y_s$), it holds that

$$c_{i,2^s+t} = 0.$$

Thus, (7) is reduced to

$$\begin{aligned}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} c_{i,j}h_i(\bar{x})h_j(\bar{y}) =& c_{0,0} + \sum_{a=0}^{\bar{a}-1}(-1)^{\varepsilon_{a+1}}2^{\frac{a}{2}}c_{2^a+b,0} + \sum_{s=0}^{\bar{s}-1}(-1)^{\eta_{s+1}}2^{\frac{s}{2}}c_{0,2^s+t}+ \\ & \sum_{a=0}^{\bar{a}-1}\sum_{s=0}^{\bar{s}-1}(-1)^{\varepsilon_{a+1}+\eta_{s+1}}2^{\frac{a}{2}+\frac{s}{2}}c_{2^a+b,2^s+t} \\ =& c_{0,0} + \sum_{a=0}^{\bar{a}-1}(-1)^{\varepsilon_{a+1}}2^{\frac{a}{2}}c_{2^a+b,0}+ \\ & \sum_{s=0}^{\bar{s}-1}(-1)^{\eta_{s+1}}2^{\frac{s}{2}}\Big[c_{0,2^s+t} + \sum_{a=0}^{\bar{a}-1}(-1)^{\varepsilon_{a+1}}2^{\frac{a}{2}}c_{2^a+b,2^s+t}\Big]. \end{aligned} \tag{9}$$

This shows that $\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} c_{i,j}h_i(\bar{x})h_j(\bar{y})$ converges. In addition, for any $x \notin D$, one can easily check that

$$h_0(x) + (-1)^{\varepsilon_1}h(x) = 2h_0(2x - \varepsilon_1)$$

and for any $k \geq 1$ it holds that

$$2^k h_0\Big(2^k x - \sum_{i=1}^{k} 2^{k-i}\varepsilon_i\Big) + (-1)^{\varepsilon_{k+1}} 2^k h\Big(2^k x - \sum_{i=1}^{k} 2^{k-i}\varepsilon_i\Big) = 2^{k+1} h_0\Big(2^{k+1} x - \sum_{i=1}^{k+1} 2^{k+1-i}\varepsilon_i\Big).$$

Thus, by mathematical induction, for any $k \geq 0$, it holds that

$$h_0(x) + \sum_{a=0}^{k} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} h_{a,b}(x) = 2^{k+1} h_0\Big(2^{k+1} x - \sum_{i=1}^{k+1} 2^{k+1-i}\varepsilon_i\Big). \tag{10}$$

Noting that $b = 2^a \sum_{s=1}^{a} \frac{\varepsilon_s}{2^s}$ and $t = 2^s \sum_{l=1}^{s} \frac{\eta_l}{2^l}$, one can prove that

$$\begin{aligned} &c_{0,0} + \sum_{a=0}^{\bar{a}-1} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} c_{2^a+b,0} + \sum_{s=0}^{\bar{s}-1} (-1)^{\eta_{s+1}} 2^{\frac{s}{2}} \Big[c_{0,2^s+t} + \sum_{a=0}^{\bar{a}-1} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} c_{2^a+b,2^s+t}\Big] \\ &= \int_0^1 \int_0^1 f_A(x,y) \Big[2^{\bar{s}} h_0\big(2^{\bar{s}} y - \sum_{i=1}^{\bar{s}} 2^{\bar{s}-i}\varepsilon_i\big)\Big] \Big[2^{\bar{a}} h_0\big(2^{\bar{a}} x - \sum_{i=1}^{\bar{a}} 2^{\bar{a}-i}\varepsilon_i\big)\Big] \mathrm{d}x\mathrm{d}y. \end{aligned} \tag{11}$$

In fact,

$$\begin{aligned} &c_{0,0} + \sum_{a=0}^{\bar{a}-1} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} c_{2^a+b,0} + \sum_{s=0}^{\bar{s}-1} (-1)^{\eta_{s+1}} 2^{\frac{s}{2}} [c_{0,2^s+t} + \sum_{a=0}^{\bar{a}-1} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} c_{2^a+b,2^s+t}] \\ &= \int_0^1 h_0(y)\mathrm{d}y \int_0^1 f_A(x,y) h_0(x)\mathrm{d}x + \sum_{a=0}^{\bar{a}-1} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} \int_0^1 h_0(y)\mathrm{d}y \int_0^1 f_A(x,y) h_{a,b}(x)\mathrm{d}x + \\ &\quad \sum_{s=0}^{\bar{s}-1} (-1)^{\eta_{s+1}} 2^{\frac{s}{2}} \Big[\int_0^1 h_{s,t}(y)\mathrm{d}y \int_0^1 f_A(x,y) h_0(x)\mathrm{d}x + \\ &\qquad \sum_{a=0}^{\bar{a}-1} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} \int_0^1 h_{s,t}(y)\mathrm{d}y \int_0^1 f_A(x,y) h_{a,b}(x)\mathrm{d}x\Big] \\ &= \int_0^1 h_0(y)\mathrm{d}y \int_0^1 f_A(x,y) \Big[h_0(x)\mathrm{d}x + \sum_{a=0}^{\bar{a}-1} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} h_{a,b}(x)\Big]\mathrm{d}x + \\ &\quad \sum_{s=0}^{\bar{s}-1} (-1)^{\eta_{s+1}} 2^{\frac{s}{2}} \int_0^1 h_{s,t}(y)\mathrm{d}y \int_0^1 f_A(x,y) \Big[h_0(x)\mathrm{d}x + \sum_{a=0}^{\bar{a}-1} (-1)^{\varepsilon_{a+1}} 2^{\frac{a}{2}} h_{a,b}(x)\Big]\mathrm{d}x \\ &= \int_0^1 h_0(y)\mathrm{d}y \int_0^1 f_A(x,y) \Big[2^{\bar{a}} h_0\big(2^{\bar{a}} x - \sum_{i=1}^{\bar{a}} 2^{\bar{a}-i}\varepsilon_i\big)\Big]\mathrm{d}x + \\ &\quad \sum_{s=0}^{\bar{s}-1} (-1)^{\eta_{s+1}} 2^{\frac{s}{2}} \int_0^1 h_{s,t}(y)\mathrm{d}y \int_0^1 f_A(x,y) \Big[2^{\bar{a}} h_0\big(2^{\bar{a}} x - \sum_{i=1}^{\bar{a}} 2^{\bar{a}-i}\varepsilon_i\big)\Big]\mathrm{d}x \\ &= \int_0^1 \int_0^1 f_A(x,y) \Big[h_0(y) + \sum_{s=0}^{\bar{s}-1} (-1)^{\eta_{s+1}} 2^{\frac{s}{2}}\Big] \Big[2^{\bar{a}} h_0\big(2^{\bar{a}} x - \sum_{i=1}^{\bar{a}} 2^{\bar{a}-i}\varepsilon_i\big)\Big]\mathrm{d}x\mathrm{d}y \\ &= \int_0^1 \int_0^1 f_A(x,y) \Big[2^{\bar{s}} h_0\big(2^{\bar{s}} y - \sum_{i=1}^{\bar{s}} 2^{\bar{s}-i}\varepsilon_i\big)\Big] \Big[2^{\bar{a}} h_0\big(2^{\bar{a}} x - \sum_{i=1}^{\bar{a}} 2^{\bar{a}-i}\varepsilon_i\big)\Big]\mathrm{d}x\mathrm{d}y. \end{aligned}$$

According to the definition of h_0,

$$\begin{aligned}
&\int_0^1\int_0^1 f_A(x,y)\Big[2^{\bar{s}}h_0\big(2^{\bar{s}}y-\sum_{i=1}^{\bar{s}}2^{\bar{s}-i}\varepsilon_i\big)\Big]\Big[2^{\bar{a}}h_0\big(2^{\bar{a}}x-\sum_{i=1}^{\bar{a}}2^{\bar{a}-i}\varepsilon_i\big)\Big]\mathrm{d}x\mathrm{d}y\\
&=\int_{x_{\bar{a}}}^{x_{\bar{a}}+\frac{1}{2^{\bar{a}}}}\mathrm{d}x\int_{y_{\bar{s}}}^{y_{\bar{s}}+\frac{1}{2^{\bar{s}}}} f_A(x,y)\Big[2^{\bar{s}}h_0\big(2^{\bar{s}}y-\sum_{i=1}^{\bar{s}}2^{\bar{s}-i}\varepsilon_i\big)\Big]\Big[2^{\bar{a}}h_0\big(2^{\bar{a}}x-\sum_{i=1}^{\bar{a}}2^{\bar{a}-i}\varepsilon_i\big)\Big]\mathrm{d}y\\
&=\int_{x_{\bar{a}}}^{x_{\bar{a}}+\frac{1}{2^{\bar{a}}}}\mathrm{d}x\int_{y_{\bar{s}}}^{y_{\bar{s}}+\frac{1}{2^{\bar{s}}}} f_A(x,y)2^{\bar{s}}2^{\bar{a}}\mathrm{d}y=f_A(\bar{x},\bar{y}). \qquad (12)
\end{aligned}$$

In the last step, the fact that, according to (8), restricted to the rectangle region $(x,y)\in[x_{\bar{a}},\ x_{\bar{a}}+\frac{1}{2^{\bar{a}}}]\times[y_{\bar{s}},\ y_{\bar{s}}+\frac{1}{2^{\bar{s}}}]$, $f_A(x,y)=a(i_0,j_0)=f_A(\bar{x},\bar{y})$ is used. (6) is thus obtained by (9), (11) and (12). Theorem 1.3 is proved. □

The following result is a direct conclusion of Theorem 1.3.

Corollary 1.4 *For any* p, $1\le p<\infty$, *it holds that*

$$f_A(x,y)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}c_{i,j}h_i(x)h_j(y),\quad \text{in } L^p. \qquad (13)$$

Since

$$f_{M,N}(x,y)=\sum_{i=0}^{2^M-1}\sum_{j=0}^{2^N-1}c_{i,j}h_i(x)h_j(y) \qquad (14)$$

is piecewise constant with possible break lines $x=\frac{i}{2^M}$, $1\le i\le 2^M-1$ and $y=\frac{i}{2^N}$, $1\le i\le 2^N-1$, its dual pair matrix $A_{M,N}$ can be obtained by

$$A_{M,N}=\big(f_{M,N}(\frac{2i+1}{2^{M+1}},\frac{2j+1}{2^{N+1}})\big)_{0\le i\le 2^M-1,\ 0\le j\le 2^N-1}. \qquad (15)$$

Since $f_{M,N}$ is an approximation of f_A, $A_{M,N}$ is thus an approximation of the matrix A, with the property that

Lemma 1.5 $A=A_{M,N}$ *if* $m=2^M$ *and* $n=2^N$.

Lemma 1.5 is a direct conclusion of Lemma 1.2.

Note that

$$f_{M,N}(x,y)=\sum_{i=0}^{2^M-1}\sum_{j=0}^{2^N-1}c_{i,j}h_i(x)h_j(y)=\tilde{h}_M^T(x)C_{M,N}\tilde{h}_N(y),$$

where

$$\tilde{h}_M^T(x)=(h_0(x),h_1(x),h_2(x),\dots,h_{2^M-1}(x)),\ \tilde{h}_N^T(y)=(h_0(y),h_1(y),h_2(y),\dots,h_{2^N-1}(y))$$

and $C_{M,N}=(c_{i,j})_{0\le i\le 2^M-1,\ 0\le j\le 2^N-1}$, it holds that

$$A_{M,N}=\big(f_{M,N}(\frac{2i+1}{2^{M+1}},\frac{2j+1}{2^{N+1}})\big)_{0\le i\le 2^M-1,\ 0\le j\le 2^N-1}=H_M^TC_{M,N}H_N, \qquad (16)$$

where $H_t=(\tilde{h}_t(\frac{1}{2^{t+1}})\ \tilde{h}_t(\frac{3}{2^{t+1}})\cdots\tilde{h}_t(\frac{2^{t+1}-1}{2^{t+1}}))$ is a $2^t\times 2^t$ matrix. The following lemma is well known.

Lemma 1.6 *$Q_t = 2^{-\frac{t}{2}} H_t$ is a Haar orthogonal matrix for any natural number $t \geq 1$.*

According to the above results, we obtain the following wavelet decomposition of matrix.

Theorem 1.7 *Let A be an $m \times n$ matrix and $f_A(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_{i,j} h_i(x) h_j(y)$ be wavelet expansion of its corresponding dual pair function. Then $A_{M,N}$, an approximation matrix of the matrix A defined in (16), has the following orthogonal decomposition.*

$$A_{M,N} = H_M^T C_{M,N} H_N,$$

where $Q_t = 2^{-\frac{t}{2}} H_t$ is an orthogonal matrix. In addition,

$$A = A_{M,N} = H_M^T C_{M,N} H_N,$$

if $m = 2^M$ and $n = 2^N$.

We should note that H_t is also a sparse matrix with each column having equal $t+1$ non-zero entries. H_1, H_2, H_3 are given as follows.

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad H_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{pmatrix},$$

$$H_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{pmatrix}.$$

Since H_t is known, all the computational complexity of the wavelet decomposition comes from $c_{i,j}$.

Theorem 1.8 *The calculation of $c_{i,j}$ involves at most 69 multiplications. Therefore, the wavelet decomposition (16) has linearly computational complexity $O(2^M 2^N)$.*

Proof To calculate $c_{i,j}$, for $0 \leq x_1 < x_2 \leq 1$ and $0 \leq y_1 < y_2 \leq 1$, we first calculate

$$I(x_1, x_2; y_1, y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_A(x,y) h_i(x) h_j(y) \mathrm{d}x \mathrm{d}y.$$

Let

$$m_1 = [mx_1], \ m_2 = [mx_2], \ n_1 = [ny_1], \ n_2 = [ny_2],$$

where $[x]$ is the greatest integer of less then or equal to x. Then,

$$I(x_1,x_2;y_1,y_2)=\begin{cases}(x_2-x_1)(y_2-y_1)a_{m_1,n_1}, \text{if } x_2\le \frac{m_1+1}{m} \text{ and } y_2\le\frac{n_1+1}{n},\\ (x_2-x_1)((\frac{n_1+1}{n}-y_1)a_{m_1,n_1}+\frac{1}{n}\sum_{j=n_1+1}^{n_2-1}a_{m_1,j}+(y_2-\frac{n_2}{n})a_{m_1,n_2}),\\ \text{if } x_2\le\frac{m_1+1}{m} \text{ and } y_2>\frac{n_1+1}{n},\\ (y_2-y_1)((\frac{m_1+1}{m}-x_1)a_{m_1,n_1}+\frac{1}{m}\sum_{i=m_1+1}^{m_2-1}a_{i,n_1}+(x_2-\frac{m_2}{m})a_{m_2,n_1}),\\ \text{if } \ x_2>\frac{m_1+1}{m} \text{ and } y_2\le\frac{n_1+1}{n},\\ (y_2-y_1)((\frac{m_1+1}{m}-x_1)a_{m_1,n_1}+\frac{1}{m}\sum_{i=m_1+1}^{m_2-1}a_{i,n_1}+(x_2-\frac{m_2}{m})a_{m_2,n_1}),\\ \text{if } \ x_2>\frac{m_1+1}{m} \text{ and } y_2\le\frac{n_1+1}{n},\\ I(x_1,\frac{m_1+1}{m};y_1,y_2)+I(\frac{m_2}{m},x_2;y_1,y_2)+I(\frac{m_1+1}{m},\frac{m_2}{m};y_1,\frac{n_1+1}{n})+\\ I(\frac{m_1+1}{m},\frac{m_2}{m};\frac{n_2}{n},y_2)+\frac{1}{mn}\sum_{i=m_1+1}^{m_2-1}\sum_{j=n_1+1}^{n_2-1}a_{i,j},\\ \text{if } \ x_2>\frac{m_1+1}{m} \text{ and } y_2>\frac{n_1+1}{n}.\end{cases} \tag{17}$$

Equation (17) shows that the calculation of $I(x_1,x_2;y_1,y_2)$ needs at most seventeen multiplications. Let $i=2^s+t, 0\le t<2^s$ and $j=2^u+v, 0\le v<2^u$. Then

$$\begin{aligned}c_{i,j}=&\int_0^1\int_0^1 f_A(x,y)h_i(x)h_j(y)\mathrm{d}x\mathrm{d}y=\int_0^1\int_0^1 f_A(x,y)h_{s,t}(x)h_{u,v}(y)\mathrm{d}x\mathrm{d}y\\ =&2^{\frac{s+u}{2}}\big(I(\frac{t}{2^s},\frac{2t+1}{2^{s+1}};\frac{v}{2^u},\frac{2v+1}{2^{u+1}})-I(\frac{t}{2^s},\frac{2t+1}{2^{s+1}};\frac{2v+1}{2^{u+1}},\frac{v+1}{2^u})-\\ &I(\frac{2t+1}{2^{s+1}},\frac{t+1}{2^s};\frac{v}{2^u},\frac{2v+1}{2^{u+1}})+I(\frac{2t+1}{2^{s+1}},\frac{t+1}{2^s};\frac{2v+1}{2^{u+1}},\frac{v+1}{2^u})\big).\end{aligned}$$

Above equation shows that the calculation of $c_{i,j}$ involves at most 69 multiplications. □

2. Some applications of wavelet decomposition of matrix

A k-partition of $\{0,1,\dots,m-1\}$ is to decompose $\{0,1,\dots,m-1\}$ into k non-empty sets $S=\{S_1,S_2,\dots,S_k\}$ such that $\bigcup_{i=1}^k S_i=\{0,\ 1,\dots,m-1\}$ and $S_i\bigcap S_j$ is empty if $i\ne j$. The k-means clustering of a given set of d-dimensional observation data $x_j=(a_{0,j},a_{1,j},\dots,a_{d-1,j})^T\in\mathbf{R}^d$, $0\le j\le m-1$, is to find a k-partition $\tilde S=\{\tilde S_1,\tilde S_2,\dots,\tilde S_k\}$ such that for any other k-partition $S=\{S_1,S_2,\dots,S_k\}$, it holds that

$$\sum_{j=1}^k\sum_{i\in\tilde S_j}\|x_i-\tilde m_j\|^2\le\sum_{j=1}^k\sum_{i\in S_j}\|x_i-m_j\|^2,$$

where

$$\tilde m_j=\frac{1}{|\tilde S_j|}\sum_{i\in\tilde S_j}x_i:=(\tilde m_{0,j},\tilde m_{1,j},\dots,\tilde m_{d-1,j})^T,$$

$$m_j=\frac{1}{|S_j|}\sum_{i\in S_j}x_i:=(m_{0,j},m_{1,j},\dots,m_{d-1,j})^T$$

and $|A|$ is the cardinality of the set A. k-means clustering is an NP-hard and very active problem [1,2]. Different from all the current methods, in this section a new method is presented by converting the k-means clustering to an equivalent new form.

For any k-partition $S = \{S_1, S_2, \ldots, S_k\}$ of $\{0, 1, \ldots, m-1\}$, denote $D_j = \bigcup_{i \in S_j} (\frac{i}{m}, \frac{i+1}{m})$. Then $D = \{D_1, D_2, \ldots, D_k\}$ is called a k-partition of interval $(0,1)$ based on intervals $\{(\frac{i}{m}, \frac{i+1}{m}), 0 \le i \le m-1\}$. It is easy to see that any k-partition of $\{0, 1, \ldots, m-1\}$ is pairing to a k-partition of interval $(0,1)$ based on intervals $\{(\frac{i}{m}, \frac{i+1}{m}), 0 \le i \le m-1\}$, and vice versa.

Assume that $A = (a_{i,j})_{0 \le i \le d-1, 0 \le j \le m-1}$ and its dual pair function is $f_A(x,y)$. For a k-partition $S = \{S_1, S_2, \ldots, S_k\}$ of $\{0, 1, \ldots, m-1\}$, denote $E = (b_{i,j})_{0 \le i \le d-1, 0 \le j \le m-1}$ where $b_t = (b_{0,t}, b_{1,t}, \ldots, b_{d-1,t})^T = m_j$ if $t \in S_j$. Then it holds that

$$\sum_{j=1}^{k} \sum_{i \in S_j} \|x_i - m_j\|^2 = \sum_{j=0}^{m-1} \|x_j - b_j\|^2 = \|A - E\|_F^2, \tag{18}$$

where $\|A\|_F$ is the Frobenius norm of the matrix A. Denote also that $D_j = \bigcup_{i \in S_j} (\frac{i}{m}, \frac{i+1}{m})$ and let $f_E(x,y)$ be the dual pair function of E. Then,

$$f_E(x,y) = m_{i,j}, \text{ for } \frac{i}{d} < x < \frac{i+1}{d} \text{ and } y \in D_j.$$

In fact, it holds that

$$m_{i,j} = \frac{1}{|S_j|} \sum_{l \in S_j} a_{i,l} = \frac{1}{m|D_j|_{\text{meas}}} \sum_{l \in S_j} a_{i,l} = \frac{1}{|D_j|_{\text{meas}}} \int_{D_j} f_A(x,y) \mathrm{d}y, \quad \frac{i}{d} < x < \frac{i+1}{d},$$

where $|D_j|_{\text{meas}}$ is the Lebesgue measure of D_j. That is,

$$f_E(x,y) = \frac{1}{|D_j|_{\text{meas}}} \int_{D_j} f_A(x,y) \mathrm{d}y, \text{ for } y \in D_j \tag{19}$$

is the mean of of f_A on the set D_j. This shows that the k-means clustering of vectors is corresponding to the following k-means clustering of a piecewise constant function.

k-means clustering of a given piecewise function f defined on the unit square with possible break lines $x = \frac{i}{d}$, $1 \le i \le d-1$ and $y = \frac{j}{m}$, $1 \le j \le m-1$ is to find a k-partition $D = \{D_1, D_2, \ldots, D_k\}$ of interval $(0,1)$ based on intervals $\{(\frac{i}{m}, \frac{i+1}{m}), 0 \le i \le m-1\}$, such that

$$\|f_A - f_E\|^2 = \int_0^1 \int_0^1 |f_A(x,y) - f_E(x,y)|^2 \mathrm{d}x \mathrm{d}y$$

is minimum among all possible such kind k-partitions, where

$$f_E(x,y) = \frac{1}{|D_j|_{\text{meas}}} \int_{D_j} f_A(x,y) \mathrm{d}y, \text{ for } y \in D_j, \ 1 \le j \le k.$$

Let $E = (b_{i,j})_{0 \le i \le d-1, 0 \le j \le m-1}$ denote the pair matrix of f_E. Then, for $t \in S_j$, it holds that

$$b_{i,t} = \frac{1}{|S_j|} \sum_{l \in S_j} a_{i,l}, \text{ for } t \in S_j$$

where $S = \{S_1, S_2, \ldots, S_k\}$ is the pair k-partition of $\{0, 1, \ldots, m-1\}$ corresponding to $D =$

$\{D_1, D_2, \ldots, D_k\}$. According to the definitions of f_A and f_E, it holds that

$$\|f_A - f_E\|^2 = \int_0^1 \int_0^1 |f_A(x,y) - f_E(x,y)|^2 \mathrm{d}x\mathrm{d}y = \frac{1}{\mathrm{d}m} \sum_{i=0}^{d-1} \sum_{j=0}^{m-1} |a_{i,j} - b_{i,j}|^2 = \frac{1}{\mathrm{d}m} \|A - E\|_F^2. \quad (20)$$

According to (18) and (20), the following equivalent theorem is obtained.

Theorem 2.1 *The k-means clustering of vectors $x_j = (a_{0,j}, a_{1,j}, \ldots, a_{d-1,j})^T \in \mathbf{R}^d, 0 \le j \le m-1$, is equivalent to the k-means clustering of the piecewise function f_A, where $A = (a_{i,j})_{0 \le i \le d-1, 0 \le j \le m-1}$.*

The above equivalent theorem provides a new method to study k-means clustering. The following lemma is obvious.

Lemma 2.2 *If Q is an orthogonal matrix of order d and c is a constant, then the k-means clustering of vectors $x_j = (a_{0,j}, a_{1,j}, \ldots, a_{d-1,j})^T \in \mathbf{R}^d$, $0 \le j \le m-1$ is equivalent to the k-means clustering of vectors $cQx_j \in \mathbf{R}^d$, $0 \le j \le m-1$.*

Assuming that $A_{M,N}$ is a suitable approximation of A, according to (16), it holds that

$$H_M A_{M,N} = 2^M C_{M,N} H_N.$$

Since usually $C_{M,N}$ is an approximate sparse matrix and H_N is a sparse matrix, above formula produces a new method for dimensional reduction other than principal component analysis. In addition, this skill can be repeatedly used to achieve better results. Considering dimension reduction is a key step for k-means clustering, a new method is presented for k-means clustering.

Theorem 2.3 *Assume that the k-partition $\tilde{S} = \{\tilde{S}_1, \tilde{S}_2, \ldots, \tilde{S}_k\}$ of $\{0, 1, \ldots, 2^N - 1\}$ is a k-means solution of the data $A_{M,N}$ and denote that $\tilde{D}_j = \bigcup_{i \in \tilde{S}_j} (\frac{i}{2^N}, \frac{i+1}{2^N})$. Then $S = \{S_1, S_2, \ldots, S_k\}$ is an approximate solution of the k-means clustering of the original data $x_j = (a_{0,j}, a_{1,j}, \ldots, a_{d-1,j})^T \in \mathbf{R}^d$, $0 \le j \le m-1$, where S_j is the collection of all the index i such that $(\frac{i}{m}, \frac{i+1}{m}) \bigcap \tilde{D}_j$ has the maximum Lebesgue measure, i.e.,*

$$S_j = \big\{i;\ |(\frac{i}{m}, \frac{i+1}{m}) \bigcap \tilde{D}_j|_{\mathrm{meas}} = \max\{|(\frac{i}{m}, \frac{i+1}{m}) \bigcap \tilde{D}_l|_{\mathrm{meas}};\ 1 \le l \le k\}\big\}.$$

In addition, this method is also useful in feature extraction, such as image processing or signal processing. The study of this aspect is still investigating.

References

[1] J. A. HARTIGAN, M. A. WONG. *Algorithm AS 136: A K-Means Clustering Algorithm*. Journal of the Royal Statistical Society, Series C, 1979, 28(1): 100108.

[2] V. ALON, S. S. SHAI. *K-means Recovers ICA Filters when Independent Components are Sparse*. Proc. of Int'l Conf. Machine Learning (ICML 2014).