

Wavelet Review

Canpei Hu

November 21, 2017

1 向量空间、张成、基底

向量空间 \mathbb{V} : ①对加法封闭: $\forall \mathbf{x}, \mathbf{y} \in \mathbb{V} \Rightarrow \mathbf{x} + \mathbf{y} \in \mathbb{V}$

②对数乘封闭: $\forall \mathbf{x} \in \mathbb{V}, \forall \lambda \in \mathbb{R} \Rightarrow \lambda \mathbf{x} \in \mathbb{V}$

线性组合: $\mathbf{z} = a\mathbf{x} + b\mathbf{y} \quad \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{V}; a, b \in \mathbb{R}$

张成: $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\} = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n | a_i \in \mathbb{R}; \mathbf{v}_i \in \mathbb{V}\}$

基底: 空间中存在的最大线性无关向量组, 即张成此空间所需的最小线性无关向量组

\mathbb{V} 的子空间: 既是 \mathbb{V} 的子集, 又是向量空间

2 正交、相关、傅里叶级数的基底

同一空间中的两个向量 \mathbf{x}, \mathbf{y} 正交, 即是 $\mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y} = 0$;

$[a, b]$ 区间上的两个函数 $f(x), g(x)$ 正交, 即是 $\int_a^b f(x)g(x)dx = 0$.

以上定义引入内积概念: 向量 \mathbf{x}, \mathbf{y} 的内积定义为 $\sum_{i=1}^n x_i y_i$; $[a, b]$ 区间上函数 $f(x), g(x)$ 的内积定义为 $\int_a^b f(x)g(x)dx$.

相关亦由内积定义, 内积的模越大, 相关性越强。正交即是相关为0的特殊情况。

三角函数系 $\{1, \cos x, \sin x, \cos 2x, \sin 2x, \dots, \cos nx, \sin nx\}$ 在基波 $\{\cos x, \sin x\}$ 的周期区间上正交, 且每一个基底与自身内积为1, 它们是傅里叶级数空间的一组标准正交基(orthonormal basis).
基底选择三原则:

①正交(orthogonal): 一组基底中的向量两两正交, 实现计算时的解耦(decouple), 并且最好是归一化(normalized)的, 称为标准正交基(orthonormal basis);

②系数紧凑(compact coefficients): 向量在少量基底上的投影系数大, 而在绝大部分基底上的投影系数小;

③特征包含: 含有一个向量想要测量的性质。

3 矩阵4个子空间、正交补

\mathbf{A} 的列空间 $\mathbb{C}(\mathbf{A})$: \mathbf{A} 的列向量张成的空间

\mathbf{A} 的零空间 $\mathbb{N}(\mathbf{A})$: 方程 $\mathbf{A}\mathbf{x} = 0$ 的解空间

\mathbf{A}^T 的列空间 $\mathbb{C}(\mathbf{A}^T)$: \mathbf{A}^T 的列向量张成的空间

\mathbf{A}^T 的零空间 $\mathbb{N}(\mathbf{A}^T)$: 方程 $\mathbf{A}^T \mathbf{x} = 0$ 的解空间

\mathbb{V}_0 在 \mathbb{V} 中的正交补: $\{\mathbf{v} \in \mathbb{V} | \forall \mathbf{v}_0 \in \mathbb{V}_0 \Rightarrow \mathbf{v} \cdot \mathbf{v}_0 = 0\}$. 子空间的正交补也是子空间。

关系:

① \mathbf{A}^T 的列空间是 \mathbf{A} 的行空间, 即 \mathbf{A} 的行向量张成的空间

② \mathbf{A}^T 的零空间是 \mathbf{A} 的左零空间, 即方程 $\mathbf{x}\mathbf{A} = 0$ 的解空间

③ $\mathbb{C}(\mathbf{A}) \perp \mathbb{N}(\mathbf{A}^T)$, 二者互为正交补, 即 $\forall \mathbf{x} \in \mathbb{C}(\mathbf{A}), \forall \mathbf{y} \in \mathbb{N}(\mathbf{A}^T) \Rightarrow \mathbf{x} \cdot \mathbf{y} = 0$

④ $\mathbb{C}(\mathbf{A}^T) \perp \mathbb{N}(\mathbf{A})$, 二者互为正交补, 即 $\forall \mathbf{x} \in \mathbb{C}(\mathbf{A}^T), \forall \mathbf{y} \in \mathbb{N}(\mathbf{A}) \Rightarrow \mathbf{x} \cdot \mathbf{y} = 0$

4 一个向量用另一个向量近似

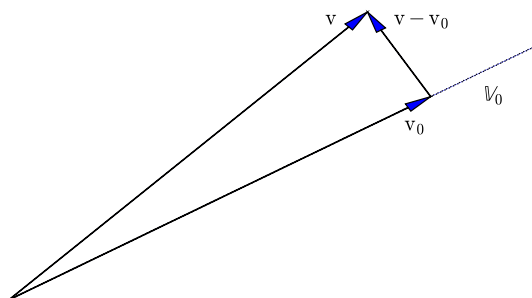


Figure 1: 一个向量用另一个向量近似

如图Fig1，用子空间 V_0 中的 αv_0 来近似表达 V 中的向量 v . 要使得误差 $v - \alpha v_0$ 最小，即是求

$$J = \|v - \alpha v_0\|^2$$

最小时的 α ，可用如下方法：

- ①J对 α 求导；
- ②用垂线最短：

$$(v - \alpha v_0) \perp v_0 \Rightarrow (v - \alpha v_0) \cdot v_0 = 0 \Rightarrow \alpha = \frac{v \cdot v_0}{v_0 \cdot v_0}$$

5 一个向量用另一组向量近似、正交投影

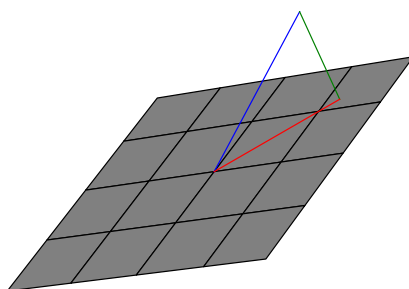


Figure 2: 一个向量用另一组向量近似

引入**正交投影**：若 \mathbb{V}_0 是 \mathbb{V} 的子空间，则 \mathbb{V} 中向量 \mathbf{v} 在 \mathbb{V}_0 中的正交投影 \mathbf{v}_0 满足：① $\mathbf{v}_0 \in \mathbb{V}_0$ ；② $\mathbf{v} - \mathbf{v}_0$ 与 \mathbb{V}_0 中所有向量都正交。

若 $\mathbf{v} \in \mathbb{V}_0$ ，则 \mathbf{v} 的正交投影就是它本身；若 $\mathbf{v} \notin \mathbb{V}_0$ ，则 $\mathbf{v} - \mathbf{v}_0$ 垂直于 \mathbb{V}_0 的所有基底。

将前文中一个向量近似的情形推广到一组向量，以两个为例。

如图，用子空间 \mathbb{V}_0 的基底 $\mathbf{v}_1, \mathbf{v}_2$ 来近似表达 \mathbb{V} 中的向量 \mathbf{v} ，使得误差 $\mathbf{v} - \alpha\mathbf{v}_1 - \beta\mathbf{v}_2$ 最小，即是求

$$J = \|\mathbf{v} - (\alpha\mathbf{v}_1 + \beta\mathbf{v}_2)\|^2$$

最小时的 α, β 。

①用高数方法求取到极值时的条件；

②用正交投影：

$$(\mathbf{v} - \alpha\mathbf{v}_1 - \beta\mathbf{v}_2) \perp \mathbb{V}_0 \Rightarrow \begin{cases} (\mathbf{v} - \alpha\mathbf{v}_1 - \beta\mathbf{v}_2) \perp \mathbf{v}_1 \\ (\mathbf{v} - \alpha\mathbf{v}_1 - \beta\mathbf{v}_2) \perp \mathbf{v}_2 \end{cases} \Rightarrow \begin{cases} \mathbf{v} \cdot \mathbf{v}_1 = \alpha\mathbf{v}_1 \cdot \mathbf{v}_1 + \beta\mathbf{v}_2 \cdot \mathbf{v}_1 \\ \mathbf{v} \cdot \mathbf{v}_2 = \alpha\mathbf{v}_1 \cdot \mathbf{v}_2 + \beta\mathbf{v}_2 \cdot \mathbf{v}_2 \end{cases}$$

二元一次，解出 α, β 即可。

6 最小二乘

最小二乘的求解思路即是用一组向量近似一个向量。考虑含有两个特征的方程组：

$$\begin{cases} m_1x_1^1 + n_1x_1^2 + b_1 = y_1 \\ m_2x_2^1 + n_2x_2^2 + b_2 = y_2 \\ \dots \\ m_kx_k^1 + n_kx_k^2 + b_k = y_k \end{cases}$$

写为向量形式 $\mathbf{m} \cdot \mathbf{x}^1 + \mathbf{n} \cdot \mathbf{x}^2 + \mathbf{b} \cdot \mathbf{u} = \mathbf{y}$ 。

一般来说，这个方程组的变量维度远远高于特征数，不存在确切解。因此最小二乘的目的即是找一个方程，它能由这一组方程表示，使得表示误差最小。这等价于找一点 $\mathbf{p} = m\mathbf{x}^1 + n\mathbf{x}^2 + b\mathbf{u}$ ，使得 \mathbf{p} 与 \mathbf{y} 误差最小。

由前文向量近似表达的思想，做如下推导：

$$\begin{cases} (\mathbf{y} - \mathbf{p}) \perp \mathbf{x}^1 \\ (\mathbf{y} - \mathbf{p}) \perp \mathbf{x}^2 \\ (\mathbf{y} - \mathbf{p}) \perp \mathbf{u} \end{cases} \Rightarrow \begin{cases} (\mathbf{y} - m\mathbf{x}^1 - n\mathbf{x}^2 - b\mathbf{u}) \cdot \mathbf{x}^1 = 0 \\ (\mathbf{y} - m\mathbf{x}^1 - n\mathbf{x}^2 - b\mathbf{u}) \cdot \mathbf{x}^2 = 0 \\ (\mathbf{y} - m\mathbf{x}^1 - n\mathbf{x}^2 - b\mathbf{u}) \cdot \mathbf{u} = 0 \end{cases} \Rightarrow \begin{cases} \mathbf{x}^1 \cdot \mathbf{y} = m\mathbf{x}^1 \cdot \mathbf{x}^1 + n\mathbf{x}^2 \cdot \mathbf{x}^1 + b\mathbf{u} \cdot \mathbf{x}^1 \\ \mathbf{x}^2 \cdot \mathbf{y} = m\mathbf{x}^1 \cdot \mathbf{x}^2 + n\mathbf{x}^2 \cdot \mathbf{x}^2 + b\mathbf{u} \cdot \mathbf{x}^2 \\ \mathbf{u} \cdot \mathbf{y} = m\mathbf{x}^1 \cdot \mathbf{u} + n\mathbf{x}^2 \cdot \mathbf{u} + b\mathbf{u} \cdot \mathbf{u} \end{cases}$$

写为矩阵形式：

$$\begin{cases} \mathbf{x}^{1T}\mathbf{y} = m\mathbf{x}^{1T}\mathbf{x}^1 + n\mathbf{x}^{2T}\mathbf{x}^1 + b\mathbf{u}^T\mathbf{x}^1 \\ \mathbf{x}^{2T}\mathbf{y} = m\mathbf{x}^{1T}\mathbf{x}^2 + n\mathbf{x}^{2T}\mathbf{x}^2 + b\mathbf{u}^T\mathbf{x}^2 \\ \mathbf{u}^T\mathbf{y} = m\mathbf{x}^{1T}\mathbf{u} + n\mathbf{x}^{2T}\mathbf{u} + b\mathbf{u}^T\mathbf{u} \end{cases} \Rightarrow \begin{bmatrix} \mathbf{x}^{1T} \\ \mathbf{x}^{2T} \\ \mathbf{u}^T \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{x}^{1T} \\ \mathbf{x}^{2T} \\ \mathbf{u}^T \end{bmatrix} \begin{bmatrix} \mathbf{x}^1 & \mathbf{x}^2 & \mathbf{u} \end{bmatrix} \begin{bmatrix} m \\ n \\ b \end{bmatrix}$$

令 $\mathbf{Z} = \begin{bmatrix} \mathbf{x}^1 & \mathbf{x}^2 & \mathbf{u} \end{bmatrix}$ ，则有： $\begin{bmatrix} m \\ n \\ b \end{bmatrix} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y}$

这就是**最小二乘法**(Least Square Method, LSM)。

7 Haar小波的三种理解

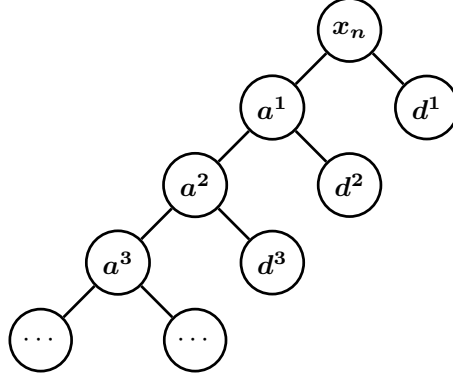


Figure 3: Haar小波分解

Fig3为小波每一层的分解图，不断分解 a 分量即可。

①: 运算观点

给定序列(向量) $\mathbf{x}_n = \{x_1, x_2, \dots, x_N\}$ ，其中 $N = 2^m$ 是2的整数幂。则 \mathbf{x}_n 的一级Haar小波分解如下：

$$\mathbf{a}^1 = \left\{ \frac{x_1 + x_2}{\sqrt{2}}, \frac{x_3 + x_4}{\sqrt{2}}, \dots, \frac{x_{N-1} + x_N}{\sqrt{2}} \right\}$$

$$\mathbf{d}^1 = \left\{ \frac{x_1 - x_2}{\sqrt{2}}, \frac{x_3 - x_4}{\sqrt{2}}, \dots, \frac{x_{N-1} - x_N}{\sqrt{2}} \right\}$$

注意其中 \mathbf{a}^1 和 \mathbf{d}^1 的长度(维度)为 $\frac{N}{2}$ 。对 \mathbf{a}^1 继续应用如上变换，得到

$$\mathbf{a}^2 = \left\{ \frac{a_1^1 + a_2^1}{\sqrt{2}}, \frac{a_3^1 + a_4^1}{\sqrt{2}}, \dots, \frac{a_{\frac{N}{2}-1}^1 + a_{\frac{N}{2}}^1}{\sqrt{2}} \right\} = \left\{ \frac{x_1 + x_2 + x_3 + x_4}{2}, \frac{x_5 + x_6 + x_7 + x_8}{2}, \dots \right\}$$

$$\mathbf{d}^2 = \left\{ \frac{a_1^1 - a_2^1}{\sqrt{2}}, \frac{a_3^1 - a_4^1}{\sqrt{2}}, \dots, \frac{a_{\frac{N}{2}-1}^1 - a_{\frac{N}{2}}^1}{\sqrt{2}} \right\} = \left\{ \frac{x_1 + x_2 - x_3 - x_4}{2}, \frac{x_5 + x_6 - x_7 - x_8}{2}, \dots \right\}$$

注意其中 \mathbf{a}^2 和 \mathbf{d}^2 的维度为 $\frac{N}{4}$ ，此即为 \mathbf{x}_n 的二级Haar小波分解

更高阶的小波分解依此类推，不断对 \mathbf{a}^i 分解，每次维度减半。

②: 基底观点

定义N维向量空间 $\mathbb{U}^1, \mathbb{V}^1, \mathbb{U}^2, \mathbb{V}^2, \dots$ ，其中 \mathbb{U}^1 的基底为：

$$\mathbf{u}_1^1 = \left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, \dots \right]^T$$

$$\mathbf{u}_2^1 = \left[0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, \dots \right]^T$$

$$\dots$$

$$\mathbf{u}_{\frac{N}{2}}^1 = \left[0, 0, \dots, 0, 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right]^T$$

\mathbb{V}^1 的基底为：

$$\mathbf{v}_1^1 = \left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0, 0, \dots \right]^T$$

$$\mathbf{v}_2^1 = \left[0, 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0, \dots \right]^T$$

$$\dots$$

$$\mathbf{v}_{\frac{N}{2}}^1 = \left[0, 0, \dots, 0, 0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right]^T$$

\mathbb{U}^2 的基底为:

$$\begin{aligned} \mathbf{u}_1^2 &= [\frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, 0, 0, 0, 0, 0, 0, 0, \dots]^T \\ \mathbf{u}_2^2 &= [0, 0, 0, 0, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, 0, 0, 0, \dots]^T \\ &\dots \\ \mathbf{u}_{\frac{N}{4}}^2 &= [0, 0, 0, 0, \dots, 0, 0, 0, 0, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}]^T \end{aligned}$$

\mathbb{V}^2 的基底为:

$$\begin{aligned} \mathbf{v}_1^2 &= [\frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, -\frac{1}{\sqrt{4}}, -\frac{1}{\sqrt{4}}, 0, 0, 0, 0, 0, 0, 0, \dots]^T \\ \mathbf{v}_2^2 &= [0, 0, 0, 0, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, -\frac{1}{\sqrt{4}}, -\frac{1}{\sqrt{4}}, 0, 0, 0, \dots]^T \\ &\dots \\ \mathbf{v}_{\frac{N}{4}}^2 &= [0, 0, 0, 0, \dots, 0, 0, 0, 0, \frac{1}{\sqrt{4}}, \frac{1}{\sqrt{4}}, -\frac{1}{\sqrt{4}}, -\frac{1}{\sqrt{4}}]^T \end{aligned}$$

更高阶依此类推，每个基底的**支撑(support)**数量按2的指数递增，空间的基底数量按2的指数递减。注意这些向量都是N维的。

$\mathbb{U}^1, \mathbb{V}^1$ 是一对正交补，所有向量 $\{\mathbf{u}_i^1, \mathbf{v}_i^1\}$ 构成 \mathbb{R}^N 的一组单位正交基。同理， $\mathbb{U}^2, \mathbb{V}^2$ 是一对正交补，所有向量 $\{\mathbf{u}_i^2, \mathbf{v}_i^2\}$ 构成 $span\{\mathbf{u}_i^1\}$ 的一组单位正交基。在这个基础上，构建**尺度分量a**和**小波分量d**

$$\begin{aligned} \mathbf{a}^1 &= \begin{bmatrix} \mathbf{x} \cdot \mathbf{u}_1^1 & \mathbf{x} \cdot \mathbf{u}_2^1 & \dots & \mathbf{x} \cdot \mathbf{u}_{\frac{N}{2}}^1 \end{bmatrix}^T \\ \mathbf{d}^1 &= \begin{bmatrix} \mathbf{x} \cdot \mathbf{v}_1^1 & \mathbf{x} \cdot \mathbf{v}_2^1 & \dots & \mathbf{x} \cdot \mathbf{v}_{\frac{N}{2}}^1 \end{bmatrix}^T \\ \mathbf{a}^2 &= \begin{bmatrix} \mathbf{x} \cdot \mathbf{u}_1^2 & \mathbf{x} \cdot \mathbf{u}_2^2 & \dots & \mathbf{x} \cdot \mathbf{u}_{\frac{N}{4}}^2 \end{bmatrix}^T \\ \mathbf{d}^2 &= \begin{bmatrix} \mathbf{x} \cdot \mathbf{v}_1^2 & \mathbf{x} \cdot \mathbf{v}_2^2 & \dots & \mathbf{x} \cdot \mathbf{v}_{\frac{N}{4}}^2 \end{bmatrix}^T \end{aligned}$$

这样，就在向量空间的基础上理解了小波分解。

③：图像观点

将序列下标改为连续变量，作图如下：

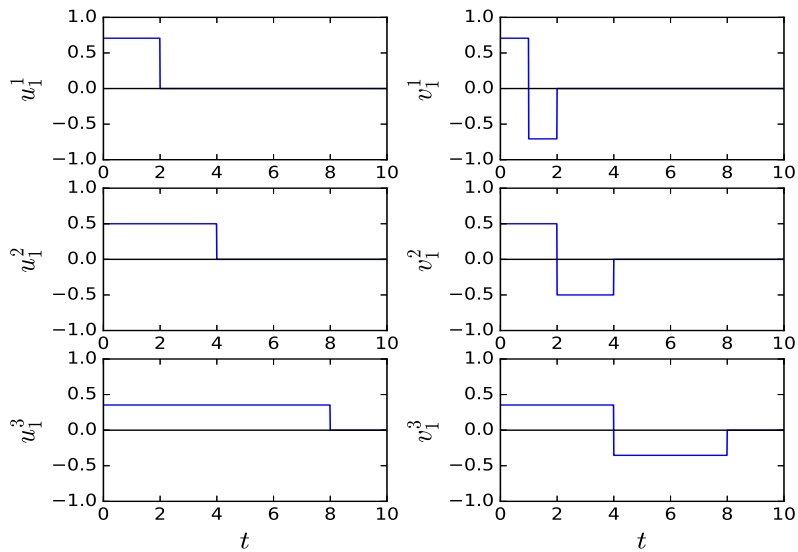


Figure 4: Haar小波基

从前文可以看出， u_1^1 向右平移2得到 u_2^1 ， u_1^1 向右平移4得到 u_3^1 ； u_1^2 向右平移4得到 u_2^2 ， u_1^2 向右平移8得到 u_3^2 ，其余规律以此类推。

上面连续形式的 u_j^i, v_j^i 构成一组小波基，对于任何均方收敛的函数，都可以表示为若干小波基的线性组合，即若干小波基波形的叠加。

8 二维Haar小波基

前面已经引入连续形式的一级小波基，这里介绍尺度函数和小波函数，仅讨论一级分解(实际上二级分解即是取一级分解得到的尺度分量继续分解)。

尺度函数(父小波)即是前文中的 u_j^i ，用来描述小波变换的尺度，连续情形下记作函数 $\phi(x)$

小波函数(母小波)即是前文中的 v_j^i ，用来描述该尺度下的小波变换，连续情形下记作函数 $\psi(x)$

二维情形下，只有两个一维尺度函数相乘才能得到二维尺度函数，即

$$\phi(x, y) = \phi(x)\phi(y)$$

而其他三种乘积均得到不同方向上的二维小波函数

$$\psi^1(x, y) = \phi(x)\psi(y), \quad \psi^2(x, y) = \psi(x)\phi(y), \quad \psi^3(x, y) = \psi(x)\psi(y)$$

将其写为离散形式并忽略所有零元素(实际上得到的矩阵是稀疏的，这里我只取非零元)：

尺度函数(水平父、垂直父)(LL)：

$$\phi(x, y) = \phi(x)\phi(y) = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

水平小波函数(水平母、垂直父)(HL)：

$$\psi^1(x, y) = \phi(x)\psi(y) = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

垂直小波函数(水平父、垂直母)(LH)：

$$\psi^2(x, y) = \psi(x)\phi(y) = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

对角小波函数(水平母、垂直母)(HH)：

$$\psi^3(x, y) = \psi(x)\psi(y) = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

一般来说，对图像(矩阵)做Haar分解：LL保留原图像的主要内容，HL保留水平方向的高频信息，LH保留垂直方向的高频信息，HH保持对角线方向的高频信息。

9 FT、STFT、WT的关系

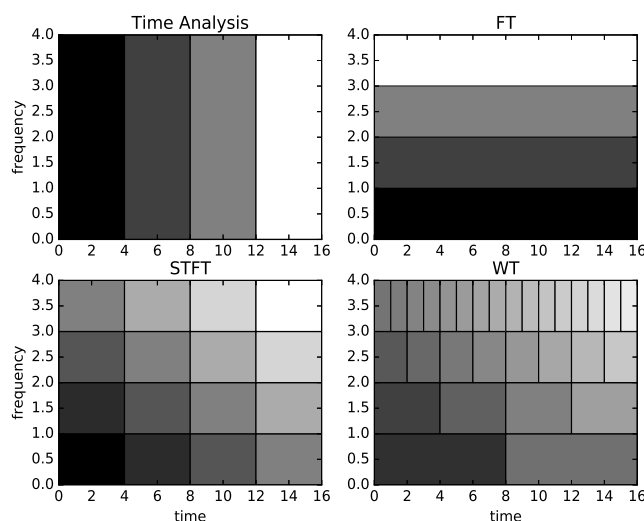


Figure 5: FT,STFT,WT的关系

在时域分析(Time Analysis)中无法得到频域信息,而傅里叶变换(Fourier Transform)无法得到时域信息(如图中上面两幅图,左图只具备时间分辨率,右图只具备频率分辨率)。基于这两种方法的局限性,后来提出了短时傅里叶变换(Short-Time Fourier Transform)和小波变换(Wavelet Transform)。

STFT是将时间轴分段,每一段上分别进行FT,这样做得到的时频域分辨率如左下图。然而由时间和频率的不确定性原理,更低的频率需要更长的时间块来描述,而一旦时间块变长,对于较高频率而言STFT的意义也就不存在。这就产生了一对矛盾。

WT也是将时间轴分段,但与STFT不同的是,WT是非均匀分段,在较高频率处使用小的时间块刻画,在较低频率处使用大的时间块刻画,进行了一定程度上的折中。这一良好性质的理论基础是各级小波分解中的尺度函数。

不过需注意小波仍然没有克服时间和频率的不确定性,克服这一不确定性的变换有希尔伯特-黄变换(HHT)等。

10 WT的应用

①图像去噪、压缩(指纹图片压缩、JPEG2000编码): 舍弃高频成分,保留尺度分量。压缩比高,速度快,信息丢失少,传输过程抗干扰;

②数值计算: 存在快速变换方法(滤波器组实现),对N点序列的时间复杂度为 $O(N)$,可用于计算傅里叶变换等;

③多分辨率处理、压缩感知;

④信号的时频分析;

⑤……

11 SVD分解、PCA

回顾特征值分解(Eigen Value Decomposition,EVD): 若对于矩阵 A 和向量 x 、标量 λ ,有 $Ax = \lambda x$,则称 x 为 A 的特征向量, λ 为 A 的特征值(Eigen Value)。特征向量的含义是,该向量 x 在该变换 A 下的结果方向不变,尺度变为原来的 λ 倍。

将特征向量排成一个矩阵的列,记作 P ;将相应的特征值写为对角阵,记作 Λ ,则有 $A = P\Lambda P^{-1}$ 。对于实对称阵 A ,所有特征向量组成一组正交基,即排成正交阵 P ,使得 $A = P\Lambda P^{-1} = P\Lambda P^T$ 。

但是EVD的局限性在于，只有方阵才能做EVD，而实际使用的矩阵(如最小二乘的系数矩阵)往往不是方阵。因此提出问题：对任意的 $m \times n$ 矩阵 $\mathbf{A}_{m \times n}$ ，能否找到 n 维空间中的一组正交基，使其经过 \mathbf{A} 变换后，成为 m 维空间中的一组正交基？

对任意 $m \times n$ 矩阵 $\mathbf{A}_{m \times n}$ ， $\mathbf{A}^T \mathbf{A}$ 是 $n \times n$ 实对称阵，因而其特征向量是 n 维空间的一组正交基，记为 $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ ，则 \mathbf{A} 将其映射为 $\{\mathbf{A}\mathbf{v}_1, \mathbf{A}\mathbf{v}_2, \dots, \mathbf{A}\mathbf{v}_n\}$ 。由于 $\{\mathbf{v}\}$ 本身正交，有 $\mathbf{v}_i^T \mathbf{v}_j = 0$ ，且

$$(\mathbf{A}\mathbf{v}_i)^T \mathbf{A}\mathbf{v}_j = \mathbf{v}_i^T \mathbf{A}^T \mathbf{A}\mathbf{v}_j = \mathbf{v}_i^T \lambda_j \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j = 0$$

这样，就找到了经过 \mathbf{A} 映射后仍然正交的一组向量，现在将映射后的正交向量单位化。由于

$$\|\mathbf{A}\mathbf{v}_j\|^2 = (\mathbf{A}\mathbf{v}_j)^T \mathbf{A}\mathbf{v}_j = \mathbf{v}_j^T \mathbf{A}^T \mathbf{A}\mathbf{v}_j = \lambda_j$$

故对于 $\lambda_j \neq 0$ 取

$$\mathbf{u}_j = \frac{\mathbf{A}\mathbf{v}_j}{\|\mathbf{A}\mathbf{v}_j\|} = \frac{1}{\sqrt{\lambda_j}} \mathbf{A}\mathbf{v}_j$$

令 $r = \text{rank}(\mathbf{A})$, $\sigma_i = \sqrt{\lambda_i}$, $0 \leq i \leq r$ ，上式变为

$$\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{A}\mathbf{v}_i$$

由于矩阵的rank只有 r ，故对应 $\lambda \neq 0$ 的特征向量 \mathbf{v} 只有 r 个(对应 $\lambda = 0$ 的 $n-r$ 个是零空间 $\mathcal{N}(\mathbf{A}^T \mathbf{A})$ 的正交基，即 $\mathbf{A}^T \mathbf{A}\mathbf{v}_j = 0$)，因此对应的 \mathbf{u} 只有 r 个，即 $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$ 。因此需要补充向量 $\{\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \dots, \mathbf{u}_m\}$ 使 $\{\mathbf{u}_i\}$ 成为 \mathbb{R}^m 空间的正交基。

在 $\mathcal{N}(\mathbf{A}^T \mathbf{A})$ 中选取 $\{\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_n\}$ 使其同时是 $\mathcal{N}(\mathbf{A})$ 的正交基，并令 $\sigma_i = 0, r < i \leq m$ 。这样做的可行性是：若 \mathbf{v} 满足 $\mathbf{A}\mathbf{v} = 0$ ，则必满足 $\mathbf{A}^T \mathbf{A}\mathbf{v} = 0$ 。

这样，就得到了 $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}, \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ ，使得

$$\mathbf{A} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r & \mathbf{v}_{r+1} & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & & 0 \\ & \ddots & & & & & \\ & & \sigma_r & & & & \\ \hline & & & 0 & & & 0 \end{bmatrix}$$

上式记作 $\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma}$ ，即

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{-1} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ \hline & & & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} \\ &= \sum_{i=1}^r \delta_i^2 \mathbf{u}_i \mathbf{v}_i^T \end{aligned}$$

这就是 \mathbf{A} 的奇异值分解(Singular Value Decomposition, SVD)。其中， \mathbf{U} 是 $m \times m$ 正交阵， \mathbf{V} 是 $n \times n$ 正交阵， $\mathbf{\Sigma}$ 是 $m \times n$ 对角阵。 \mathbf{u}_i 是 $\mathbf{A}\mathbf{A}^T$ 的特征向量，称为 \mathbf{A} 的左奇异向量； \mathbf{v}_i 是 $\mathbf{A}^T \mathbf{A}$ 的特征向量，称为 \mathbf{A} 的右奇异向量； σ_i 是 $\mathbf{A}^T \mathbf{A}$ 的特征值开方，称为 \mathbf{A} 的奇异值。奇异值为0表明对应的维度缺少信息，奇异值越大表明对应维度容纳的信息方差越大，即信息量越大。

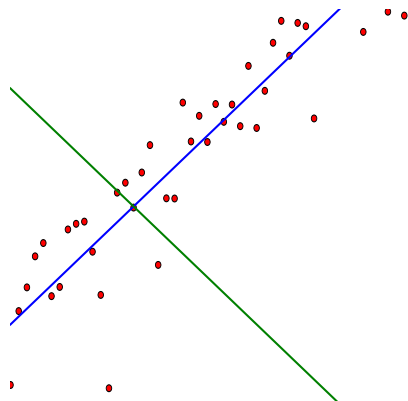


Figure 6: PCA:最大重构与最大可分

主成分分析(Principal Component Analysis,PCA): 寻找一个超平面, 使其具有两个很好的性质:

①**最大重构性:** 样本点到这个超平面的距离足够近(近似误差小);

②**最大可分性:** 样本点在超平面的投影尽可能分散(投影方差大)。

对Fig5中的点, 蓝色直线即是PCA所求的超平面, 绿色直线是该超平面的正交。最大重构性表明点到蓝色直线的距离之和最近, 即到绿色直线的正交投影之和最小; 最大可分性表明点到蓝色直线的正交投影之和最大, 即到绿色直线的距离之和最大。

PCA的一种实现手段即是SVD, 取出最大奇异值, 它对应的维度容纳信息量最大, 即是主成分。

SVD作为线性代数理论的集大成者, 有诸多应用:

①求矩阵的四个子空间;

②数值计算: 判断方程是否有解, 并求近似解(回归);

③压缩去噪: 取少量较大奇异值, 便可得到大部分信息;

④向量组的模式识别(Pattern Recognition): 取最大奇异值;

⑤潜在语义索引(Latent Semantic Indexing,LSI): 取较大奇异值, 每个对应一种语义

⑥.....