Young-Household-Schoenberg Theorem

Theorem 2.2 (Classical MDS)**.** Let $D^{n\times n}$ a real symmetric matrix. $C = D - \frac{1}{2}d\cdot \mathbf{1}^T - \frac{1}{2}\mathbf{1}\cdot d^T$, $d = \mathrm{diag}(D)$. Then:

(1) $B_\alpha = -\frac{1}{2}H_\alpha D H_\alpha^T = -\frac{1}{2}H_\alpha C H_\alpha^T$ for $\forall\alpha$ signed probability measrue;
(2) $C_{i,j} = B_{i,i}(\alpha) + B_{j,j}(\alpha) - 2B_{i,j}(\alpha)$
(3) D c.n.d. $\iff$ C c.n.d.
(4) C c.n.d. $\Rightarrow$ C is a square distance matrix (i.e. $\exists Y^{n\times k}$ s.t. $C_{i,j} = \sum_{m=1}^{k}(y_{i,m} - y_{j,m})^2$)

Schoenberg Transform

Theorem 2.3 (Schoenberg Transform)**.** Given D a square distance matrix, $C_{i,j} = \Phi(D_{i,j})$. Then:

$$C \text{ is a square distance matrix} \iff \Phi \text{ is Shoenberg Transform.}$$

A *Schoenberg Transform* Φ is a transform from $\mathbb{R}^+$ to $\mathbb{R}^+$, which takes d to

$$\Phi(d) = \int_0^\infty \frac{1 - \exp(-\lambda d)}{\lambda} g(\lambda) d\lambda,$$

where $g(\lambda)$ is some nonnegative measure on $[0, \infty)$ s.t

$$\int_0^\infty \frac{g(\lambda)}{\lambda} d\lambda < \infty.$$

Examples of Schoeberg transforms include

- $\phi_0(d) = d$ with $g_0(\lambda) = \delta(\lambda)$;
- $\phi_1(d) = \dfrac{1 - \exp(-ad)}{a}$ with $g_1(\lambda) = \delta(\lambda - a)$ $(a > 0)$;
- $\phi_2(d) = \ln(1 + d/a)$ with $g_2(\lambda) = \exp(-a\lambda)$;
- $\phi_3(d) = \dfrac{d}{a(a+d)}$ with $g_3(\lambda) = \lambda \exp(-a\lambda)$;
- $\phi_4(d) = d^p$ $(p \in (0,1))$ with $g_4(\lambda) = \dfrac{p}{\Gamma(1-p)} \lambda^{-p}$ (see more in [Bav11]).

Positive Definite functions

Theorem 3.1 (Schoenberg 38)**.** A separable space M with a metric function $d(x, y)$ can be isometrically imbedded in a Hilbert space H, if and only if the family of functions $e^{-\lambda d^2}$ are positive definite for all $\lambda > 0$ (in fact we just need it for a sequence of λ_i whose accumulate point is 0).

Here a symmetric function $k(x, y) = k(y, x)$ is called *positive definite* if for all finite x_i, x_j,

$$\sum_{i,j} c_i c_j k(x_i, x_j) \geq 0, \quad \forall c_i, c_j$$

with equality $=$ holds iff $c_i = c_j = 0$. In other words the function k restricted on $\{(x_i, x_j) : i, j = 1, \ldots, n\}$ is a positive definite matrix.

Hilbertian Embedding:
Reproducing Kernel Hilbert Spaces

- k(x,y)=k(y,x) p.d.
- define functions $k_x(\cdot) = k(x, \cdot)$
- take the span
- define inner product $\langle k_x, k_y \rangle = k(x, y)$
- take the closer => RKHS

Universality of RKHS

On the other hand, every Hilbert space $\mathcal{H}$ of functions on $\mathcal{X}$ with bounded evaluation functional can be regarded as a reproducing kernel Hilbert space [**Wah90**]. By Riesz representation, for every $x \in \mathcal{X}$ there exists $E_x \in \mathcal{H}$ such that $f(x) = \langle f, E_x \rangle$. By boundedness of evaluation functional, $|f(x)| \leq \|f\|_H \|E_x\|$, one can define a reproducing kernel $k(x, y) = \langle E_x, E_y \rangle$ which is bounded, symmetric and positive definite. It is called 'reproducing' because we can reproduce the function value using $f(x) = \langle f, k_x \rangle$ where $k_x(\cdot) := k(x, \cdot)$ as a function in $\mathcal{H}$. Such an universal property makes RKHS a unified tool to study Hilbert function spaces in nonparametric statistics, including Sobolev spaces consisting of splines [**Wah90**].

Principal Component Analysis (PCA)

Let $X = [X_1|X_2|\cdots|X_n] \in \mathbb{R}^{p\times n}$.

(2) $$\min_{\beta,\mu,U} I := \sum_{i=1}^{n} \|X_i - (\mu + U\beta_i)\|^2$$

where $U \in \mathbb{R}^{p\times k}$, $U^TU = I_p$, and $\sum_{i=1}^n \beta_i = 0$ (nonzero sum of β_i can be repre-

Best k-affine space approximation of data

$$\frac{\partial I}{\partial \mu} = -2\sum_{i=1}^{n}(X_i - \mu - U\beta_i) = 0 \Rightarrow \hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\frac{\partial I}{\partial \beta_i} = (x_i - \mu - U\beta_i)^T U = 0 \Rightarrow \beta_i = U^T(X_i - \mu)$$

Plug in the expression of $\hat{\mu}_n$ and β_i

$$\begin{aligned} I &= \sum_{i=1}^{n} \|X_i - \hat{\mu}_n - UU^T(X_i - \hat{\mu}_n)\|^2 \\ &= \sum_{i=1}^{n} \|X_i - \hat{\mu}_n - P_k(X_i - \hat{\mu}_n)\|^2 \\ &= \sum_{i=1}^{n} \|Y_i - P_k(y_i)\|^2, \quad Y_i := X_i - \hat{\mu}_n \end{aligned}$$

where $P_k = UU^T$ is a projection operator satisfying the idempotent property $P_k^2 = P_k$.

Denote $Y = [Y_1|Y_2|\cdots|Y_n] \in \mathbb{R}^{p\times n}$, whence the original problem turns into

$$\begin{aligned}
\min_U \sum_{i=1}^{n} \|Y_i - P_k(Y_i)\|^2 &= \min \operatorname{trace}[(Y - P_kY)^T(Y - P_kY)] \\
&= \min \operatorname{trace}[Y^T(I - P_k)(I - P_k)Y] \\
&= \min \operatorname{trace}[YY^T(I - P_k)^2] \\
&= \min \operatorname{trace}[YY^T(I - P_k)] \\
&= \min[\operatorname{trace}(YY^T) - \operatorname{trace}(YY^TUU^T)] \\
&= \min[\operatorname{trace}(YY^T) - \operatorname{trace}(U^TYY^TU)].
\end{aligned}$$

Above we use cyclic property of trace and idempotent property of projection.

Since Y does not depend on U, the problem above is equivalent to

$$(3) \qquad \max_{UU^T=I_k} Var(U^TY) = \max_{UU^T=I_k} \frac{1}{n}\operatorname{trace}(U^TYY^TU) = \max_{UU^T=I_k} \operatorname{trace}(U^T\hat{\Sigma}_nU)$$

where $\hat{\Sigma}_n = \frac{1}{n}YY^T = \frac{1}{n}(X - \hat{\mu}_n\mathbf{1}^T)(X - \hat{\mu}_n\mathbf{1}^T)^T$ is the sample variance. Assume

Principal Component Analysis (PCA)

- PCA is given by the top k eigenvector of covariance matrix

$$\widehat{\Sigma}_n = \frac{1}{n-1}\widetilde{X} \cdot \widetilde{X}^T$$

Both MDS and PCA are given by SVD of centered data matrix.

MDS and PCA=SVD

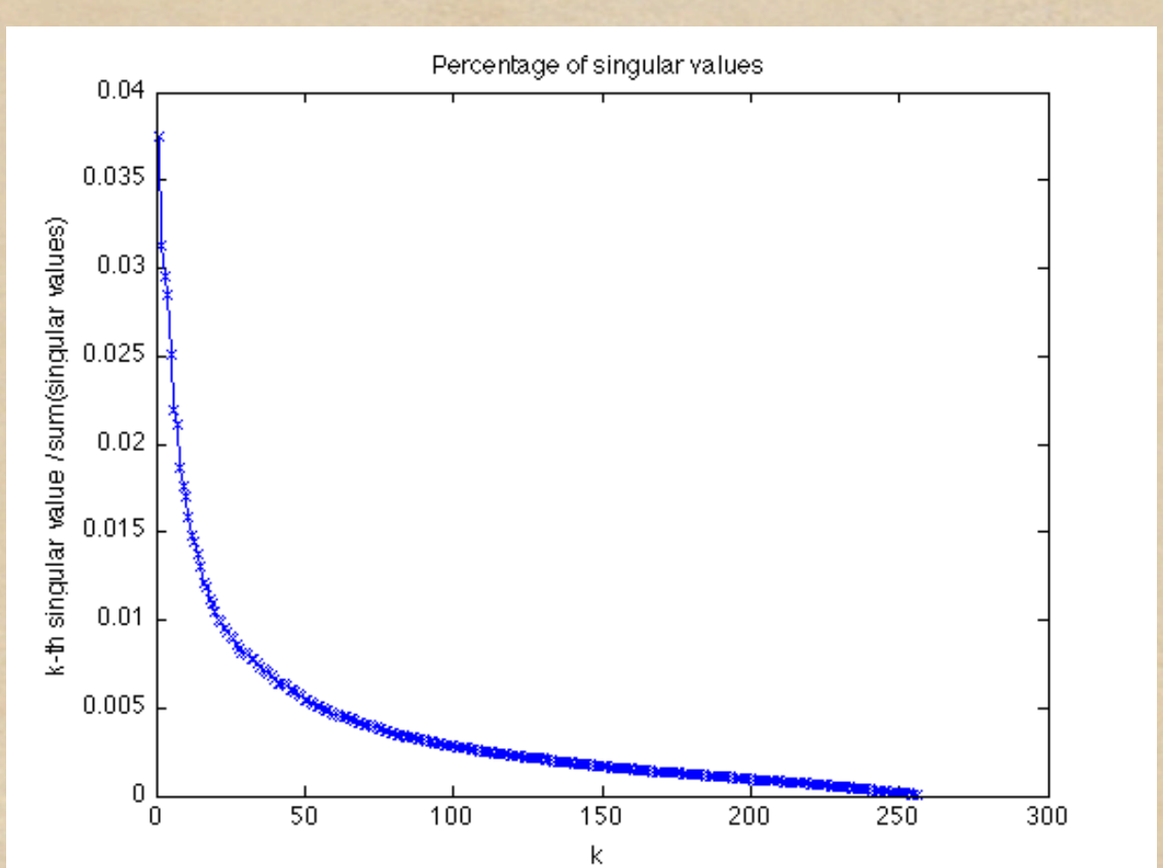
(SVD) of $X = [x_1, \ldots, x_n]^T \in \mathbb{R}^{n \times p}$ in the following sense,

$$Y = X - \frac{1}{n}\mathbf{1}\mathbf{1}^T X = \tilde{U}\tilde{S}\tilde{V}^T, \quad \mathbf{1} = (1, \ldots, 1)^T \in \mathbb{R}^n$$

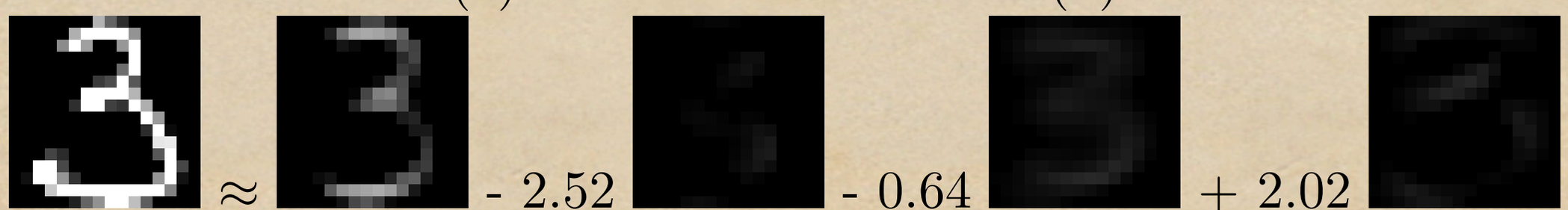
- top k left singular vectors give MDS (Kernel spectrum)
- top k right singular vectors give PCA (Covariance spectrum)

Example of PCA

(a)



(b)



(c)