The Singular Value Decomposition

Goal: We introduce/review the singular value decompostion (SVD) of a matrix and discuss some applications relevant to vision.

Consider a matrix $M \in \mathbb{R}^{n \times k}$. For convenience we assume $n \geq k$ (otherwise consider M^T). The SVD of M is a real-valued *matrix factorization*, $M = USV^T$. The SVD can be computed using an exceptionally stable numerical algortihm.

The 'compact' SVD for tall-rectangular matrices, like M, is generated in Matlab by:

```
% When n >= k
[U, S, V] = svd(M, 0);
% Here U is n x k, S is k x k diagonal, V is k x k.
```

See also the matlab calls:

- `[U,S,V] = svd(M, 'econ');` Gives a compact form of SVD for both $n < k$ and $n \geq k$.
- `[U,S,V] = svd(M);` Gives a non-compact representation, U is $n \times n$, V is $k \times k$.

See "Singular Value Decomposition" in Wikipedia, or the classic textbook by Gilbert Strang (1993) (see Section 6.7).

Properties of the SVD

Some properties of U, S, V are:

- U, S, V provide a real-valued matrix factorization of M, i.e., $M = USV^T$.
- U is a $n \times k$ matrix with orthonormal columns, $U^TU = I_k$, where I_k is the $k \times k$ identity matrix.
- V is an orthonormal $k \times k$ matrix, $V^T = V^{-1}$.
- S is a $k \times k$ diagonal matrix, with the non-negative *singular values*, $s_1, s_2, \ldots, s_k$, on the diagonal.
- By convention the singular values are given in the sorted order $s_1 \geq s_2 \geq \ldots \geq s_k \geq 0$.

Summary: For *any square or tall-rectangular matrix* M, the SVD shows that the matrix-vector product $M\vec{x}$ can be represented as:

1. An *orthogonal change of coordinates*, $V^T\vec{x}$;
2. An *axis-aligned scaling* of the result, $S(V^T\vec{x})$; and
3. The application of the resulting *coefficients in an orthonormal basis*, $U(S(V^T\vec{x}))$.

Each of these steps is easily inverted. A similar story holds for wide-rectangular matrices, i.e., $M \in \mathbb{R}^{n \times k}$ for $n < k$.

Additional Properties of the SVD

In addition we have:

- The rank of M is given by the number of singular values s_j that are non-zero.
- If $n = k$, then U is an orthonormal matrix, $U^T = U^{-1}$, so $U^T U = UU^T = I_n$.
- The pseudo-inverse of M is defined to be $M^\dagger = VRU^T$, where R is a diagonal matrix. The the j^{th} entry on the diagonal of R is $r_j = 1/s_j$ if $s_j \neq 0$, and $r_j = 0$ if $s_j = 0$. Here R is the pseudo-inverse of the diagonal matrix S.

We consider the uniqueness of the SVD next, this can be skipped on the first reading.

Uniqueness of the SVD

Consider the SVD, $M = USV^T$, for any square or tall-rectangular matrix, i.e., $M \in \mathbb{R}^{n \times k}$ with $n \geq k$.

1. The singular values are unique and, for distinct positive singular values, $s_j > 0$, the j^{th} columns of U and V are also unique up to a sign change of both columns.

2. For any repeated and positive singular values, say $s_i = s_{i+1} = \ldots = s_j > 0$ are all the singular values equal to s_i, the corresponding columns of U and V are unique up to any rotation/reflection applied to *both* sets of columns (i.e., $U_{*,i:j} \rightarrow U_{*,i:j}W$ and $V_{*,i:j} \rightarrow V_{*,i:j}W$ for some orthogonal matrix W).

3. More care must be taken with one or more singular values at zero. Suppose $s_j > 0$ and $s_{j+1} = \ldots = s_k = 0$. Here the $(j+1)^{st}$ through the k^{th} columns of U are less constrained, and can be any set of $(k-j)$ orthonormal vectors in the $(n-j)$-dimensional left null space of M. Moreover these columns of U can be chosen independently of the last $(k-j)$ columns of V (which form a orthonormal basis for the right null space of M).

Summary: These "symmetries" in the SVD are identical to those of the eigenvectors of a symmetric matrix, *except* for the third point above, which states there is additional freedom in the singular vectors for singular values equal to 0.

SVD, Least Squares, and Pseudo-Inverse

Applications of the SVD include solving least squares problems:

$$\vec{x} = \arg\min_{\vec{x}} ||A\vec{x} - \vec{b}||^2, \tag{1}$$

where A is $n \times k$ and $|| \cdot ||$ is the standard vector 2-norm (Euclidian length).

Let $A = USV^T$ denote the SVD of A. Then the range of A is contained in (or equal to) the subspace spanned by the orthogonal columns of U. We cannot reduce any error in $A\vec{x} - \vec{b}$ that is perpendicular to the range of A. Thus, it is equivalent to minimize

$$\begin{aligned} \vec{x} &= \arg\min_{\vec{x}} ||U^T A\vec{x} - U^T\vec{b}||^2 = \arg\min_{\vec{x}} ||U^T(USV^T)\vec{x} - U^T\vec{b}||^2 \\ &= \arg\min_{\vec{x}} ||(SV^T)\vec{x} - U^T\vec{b}||^2. \end{aligned} \tag{2}$$

From (2) it follows that an optimal solution is $\vec{x} = (VRU^T)\vec{b}$ where R is the pseudo-inverse of S (as given on p.3). Note that for large matrices, $\vec{x} = V(R(U^T\vec{b}))$ is much more efficient to compute.

Note the solution matrix used above, namely VRU^T, equals the pseudo-inverse $A^\dagger$.

SVD and the Matrix Square Root

Suppose K is a symmetric $n \times n$ matrix. Moreover, assume that K is non-negative definite, which means for every vector $\vec{x} \in \mathbb{R}^n$ we have $\vec{x}^T K \vec{x} \geq 0$.

We then compute the matrix square root of K, namely $K^{1/2}$, as follows:

1. Compute `[U, S, V] = svd(K)`.
2. Since K is symmetric and non-negative definite, it follows without loss of generality that we can set $U = V$.
3. Thus $K = VSV^T$, thus the columns of V are eigenvectors of K with the j^{th} eigenvalue being s_j.
4. Define $S^{1/2} = \text{diag}([s_1^{1/2}, s_2^{1/2}, \ldots, s_k^{1/2}])$, and note that the singular values are non-negative.
5. Therefore $J = VS^{1/2}V^T$ is a symmetric $n \times n$ matrix, such that $K = JJ$. So J is a suitable matrix square root, $K^{1/2}$.
6. Moreover, it also follows that J is non-negative definite and, as such, J is similar to the *positive square root* of a positive real number.

Covariance Matrices

Consider a multivariate normal distribution for n-dimensional vectors, $\vec{x} \sim N(\vec{m}, K)$, where:

- $N(\vec{m}, K)$ denotes the normal distribution;
- $\vec{m} \in \mathbb{R}^n$ is the mean;
- $K \in \mathbb{R}^{n \times n}$ is the covariance matrix. As such, K is symmetric (i.e., $K^T = K$) and positive definite (i.e., $\vec{u}^T K \vec{u} > 0$ for all $\vec{u} \in \mathbb{R}^n \backslash \{\vec{0}\}$).
- The probability density function for $N(\vec{m}, K)$ is,

$$\vec{x} \sim N(\vec{m}, K) \equiv \frac{1}{(2\pi|K|))^{n/2}} e^{-\frac{1}{2}(\vec{x}-\vec{m})^T K^{-1}(\vec{x}-\vec{m})}.$$

  Here $|K| = det(K)$ denotes the determinant of K.

Sampling from a Multivariate Normal Distribution

To sample from the Normal distribution $N(\vec{m}, K)$ we do the following:

1. Generate a $n \times 1$ vector $\vec{u}$ where each element u_j is independently sampled from $N(0, 1)$ (i.e., the 1D Normal distribution with mean 0 and covariance 1).
2. Compute the matrix square root of K, namely $K^{1/2}$, as defined on p.6.
3. Then $\vec{d} = K^{1/2}\vec{u}$ generates a fair sample from $N(0, K)$.
4. Set $\vec{x} = \vec{m} + \vec{d}$, we claim this is a fair sample from $N(\vec{m}, K)$.

To check that the covariance of $\vec{d}$ is actually K, first note that the mean of $\vec{d} = \vec{0}$. Then, by definition, the covariance of $\vec{d}$ is:

$$C \equiv E(\vec{d}\vec{d}^T) = E(K^{1/2}\vec{u}\vec{u}^T K^{T/2}) = K^{1/2}E(\vec{u}\vec{u}^T)K^{1/2} = K^{1/2}I_n K^{1/2} = K,$$

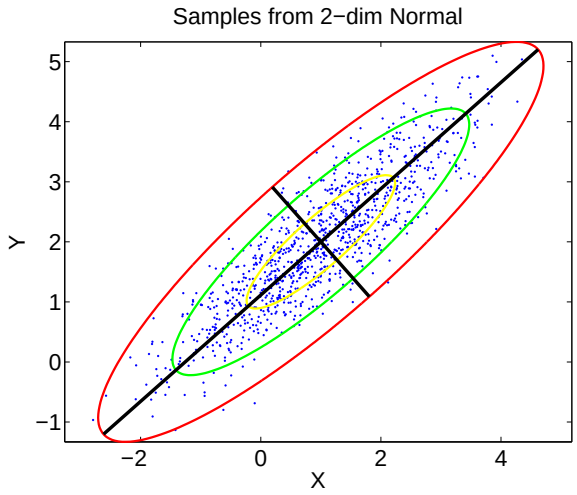
confirming that $\vec{d}$ has the covariance K, as desired.

Sample Covariance and Principal Directions

Given a set of sample vectors $\{\vec{x}_j\}_{j=1}^k$, with each $\vec{x}_j \in \mathbb{R}^n$, the sample mean and covariance are defined to be:

$$\vec{m}_s = \frac{1}{k}\sum_{j=1}^{k} \vec{x}_j, \quad \text{and } C_s = \frac{1}{(k-1)}\sum_{j=1}^{k}(\vec{x}_j - \vec{m}_s)(\vec{x}_j - \vec{m}_s)^T, \tag{3}$$

A 2D example is shown to the right. The blue points denote the samples $\vec{x}_k$. The ellipses denote curves of constant standard deviation, when measured in terms of the sample covariance C_s. That is, the curve $\vec{m} + d(\vec{\theta})$ satisfies $\vec{d}(\theta)^T C_s^{-1} \vec{d}(\theta) = \rho$, for $\rho = 1, 2,$ or 3 (corresponding to the yellow, green and red ellipses, respectively).



The black lines above indicate the `principal directions` from the sample mean, i.e., the major and minor axes of the ellipses. These are the directions of eigenvectors of C_s. The length of the j^{th} line segment, from the sample mean to the red ellipse, is equal to $\rho s_j^{1/2}$, where $\rho = 3$ and s_j is the j^{th} singular value of C_s.

Minimum Residual Variance Bases

Given a set of sample vectors $\{\vec{x}_j\}_{j=1}^k$, with each $\vec{x}_j \in \mathbb{R}^n$, form the matrix $X \in \mathbb{R}^{n\times k}$. As before, the sample mean is $\frac{1}{k}\sum_{j=1}^k \vec{x}_k = \vec{m}_s$.

Optimal Basis Selection Problem: Select a p-dimensional basis $\{\vec{b}_j\}_{j=1}^p$ that minimizes the following:

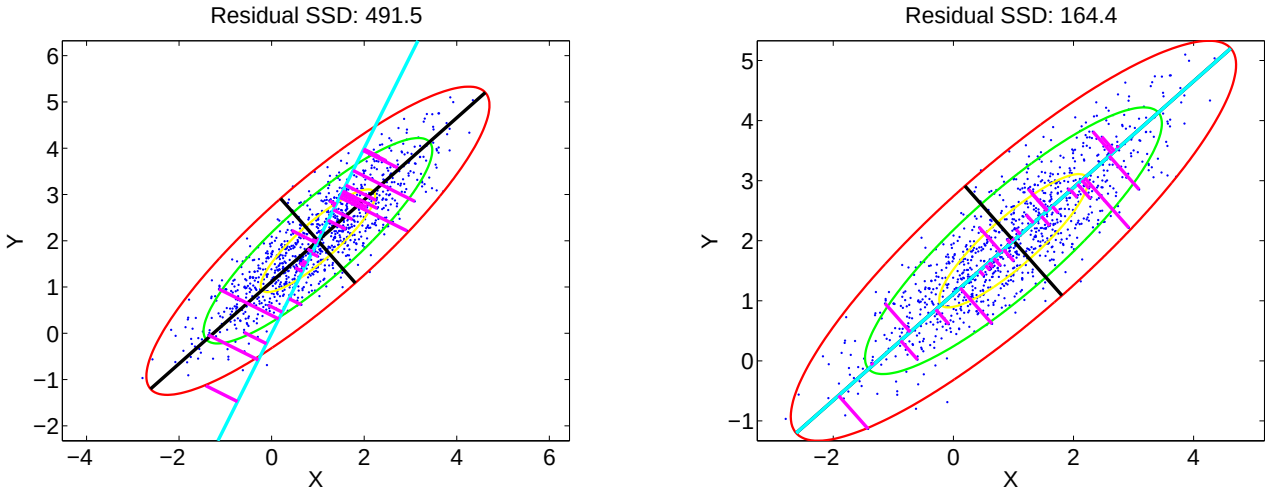
$$SSD_p = \min_{B\in\mathbb{R}^{n\times p}} \sum_{j=1}^{k} \min_{\vec{a}_j} ||\vec{x}_j - (\vec{m}_s + B\vec{a}_j)||^2. \tag{4}$$

Here $B = (\vec{b}_1, \ldots, \vec{b}_p)$ is the $n \times p$ matrix formed from the selected basis. The right-most minima above indicates that (for a given basis B), we choose the coefficients $\vec{a}_j$ which minimize the least squares error, $E_j^2 = ||\vec{x}_j - (\vec{m}_s + B\vec{a}_j)||^2$. The basis selection problem is then to choose B to minimize the sum of these least-squares errors $\sum_{j=1}^k E_j^2$ (aka, the sum of squared differences (SSD)).

An optimal choice of the p-dimensional basis, B, makes $SSD_p = \sum_{j=1}^k E_j^2$ as small as possible, and SSD_p is called the *minimum residual variance* for any basis of dimension p.

Example: Minimum Residual Variance Basis

Consider choosing an optimal 1D basis for the previous 2D example:



The cyan lines above indicate two choices for the basis direction $\vec{b}$. The mauve lines connect selected samples $\vec{x}_j$ with their best approximation $\vec{m}_s + \vec{b}a_j$. The squared length of these mauve lines are the least squares errors $E_j^2 = \min_{a_j} ||\vec{x}_j - (\vec{m}_s + \vec{b}a_j)||^2$. The residual SSD equals $\sum_{j=1}^{k} E_j^2$, and is given in the title of each plot.

In the right plot above we have set $\vec{b}$ to be the first principal direction. That is, $\vec{b}$ is the first column of U where $C_s = USV^T$. The figure illustrates that this choice minimizes the residual variance.

Principal Component Analysis: PCA

The following Theorem provides the general result.

Theorem: *(Minimum residual variance.)* For $0 \leq p \leq n$, the basis B formed from the first p principal components of the sample covariance matrix C_s (i.e., the first p columns of the matrix U of an SVD of $C_s = USV^T$) minimizes the residual variance

$$SSD(B) = \sum_{j=1}^{k} \min_{\vec{a}_j} ||\vec{x}_j - (\vec{m}_s + B\vec{a}_j)||^2, \tag{5}$$

over all possible choices of (p dimensional) bases B. Moreover, the optimal value SSD_p is given by

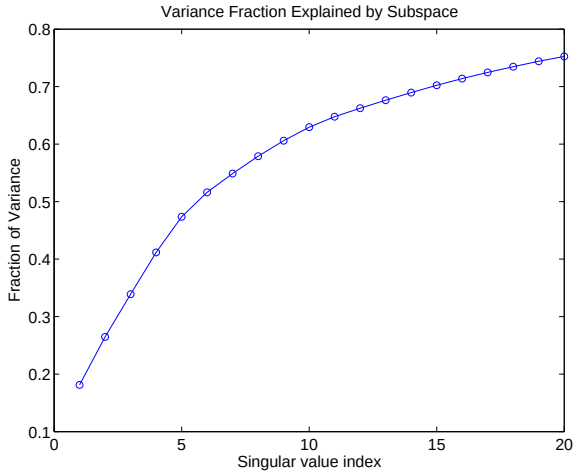
$$SSD_p = \sum_{j=p+1}^{n} s_j, \tag{6}$$

where s_j is the j^{th} singular value of the sample covariance C_s.

Note $SSD_0 = \sum_{j=1}^{n} s_j$ is the total variance in the original data set. And SSD_p monotonically decreases to 0 as p increases to n. A useful statistic is the fraction of the total variance that can be explained by a p-dimensional basis, $Q_p = (SSD_0 - SSD_p)/SSD_0$.

PCA Applied to Eyes

Subset of 1196 eye images (25×20 pixels, rewritten as 500-dimensional sample vectors $\vec{x}_j$):

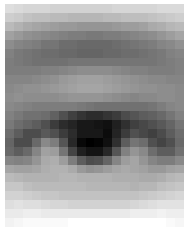


The fraction of the total variance, Q_p, captured by an optimal p dimensional subspace is plotted above. The basis is formed from the first p principal components of the 500-dimensional data set. Note Q_p approaches 75% for a 20-dimensional basis.

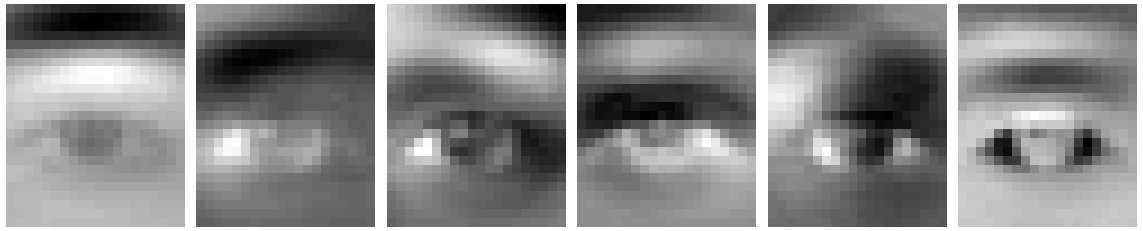
Eigen-Eye Subspace Model

The mean iamge and some principal components for the eye dataset are shown below:

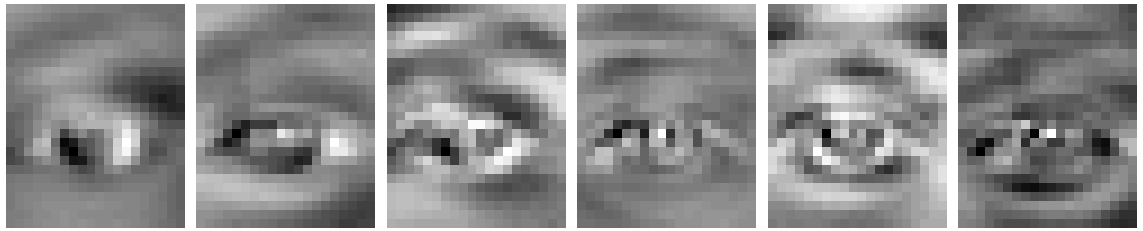
Mean Eye



Basis Images 1 through 6



Basis Images 10, 15, 20, 25, 30, 35



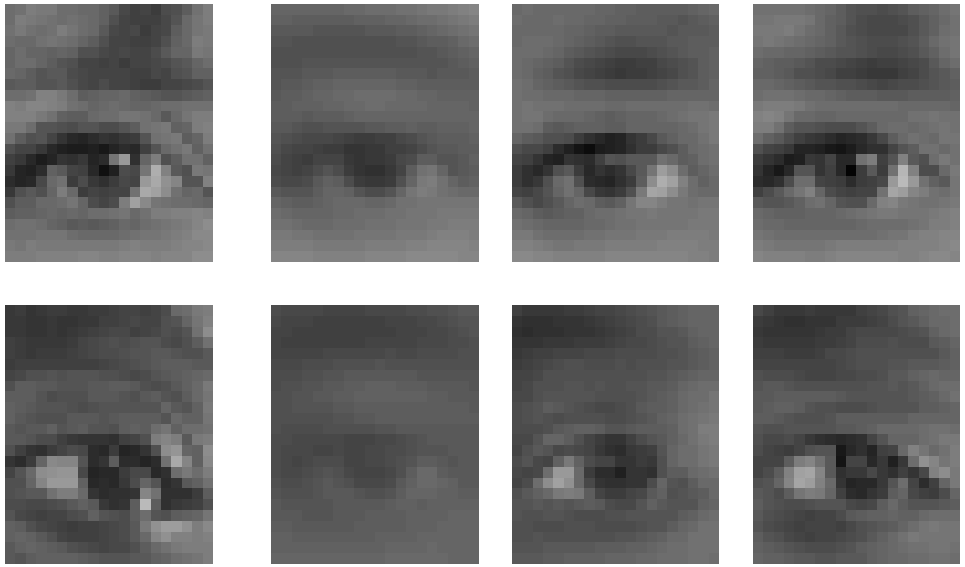
The first few principal directions (representing the dominant directions of variation in the data set) are shown on the top row. These appear to correspond to large scale shading effects and the variation around the eyebrow.

The higher order principal directions appear to capture variations at a smaller spatial scale.

Eye Reconstructions

Given the sample covariance C_s of the data, and the SVD $C_s = USV^T$, let U_p denote the first p columns of U. Then, according to the theorem on p.12, this choice of basis minimizes the residual variance.

Original Image Reconstructions for p = 5, 20, 50:



Given a new eye image $\vec{x}$ (centered and scaled), we can represent this image in the basis U_p, by solving the least squares problem $\vec{a}_0 = \arg\min_{\vec{a}} ||\vec{x} - (\vec{m}_s + U_p\,\vec{a})||^2$.

The reconstructed image $\vec{r}(\vec{a}_0) = \vec{m}_s + U_p\,\vec{a}_0$ is shown for two cases above. Note that the reconstruction is reasonable for a 20-dimensional basis, and improves as the dimension (i.e., p) increases.

References

Gilbert Strant, Introduction to Linear Algebra, 2^{nd} Edition, Wellesley-Cambridge Press, August 1993.

J. Cohen, Dependency of the spectral reflectance curves of the Munsell color chips, Psychonomic Sci 1964, 1, pp.369-370.

A. Kimball Romney, and Tarow Indow, Munsell Reflectance Spectra Represented in Three-Dimensional Euclidean Space, Color Research and Application, 28(3), 2003, pp.182-196.

Matlab tutorial: `utvisToolbox/tutorials/colourTutorial.m`

Matthew Turk and Alex Pentland, Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, 3(1), 1991, pp.71-86.

Brett Allen, Brian Curless, and Zoran Popovic, The space of human body shapes: reconstruction and parameterization from range scans, in ACM SIGGRAPH 2003, pp.27-31.