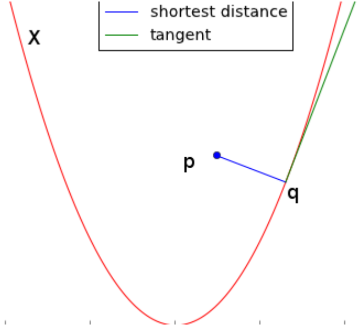
Remarks on SVD

Giorgio Ottaviani, Università di Firenze

The minimum distance problem

Let $X \subset \mathbb{A}^n_{\mathbb{R}}$ be an algebraic variety, let $p \in \mathbb{A}^n_{\mathbb{R}}$. We look for the points $q \in X$ which minimize the euclidean distance $d(p, q)$.
A necessary condition, assuming q is a smooth point of X, is that the tangent space T_qX be orthogonal to $p - q$, this is the condition to get a critical point for the distance function $d(p.-)$.



The case of matrices of bounded rank

There is one important case when this problem is solved.

Consider the affine space of $n \times m$ matrices, and let X_k be the *variety of matrices of rank* $\leq k$*.* We consider this case as a model.

X_1 is the cone over the Segre variety $\mathbb{P}^{n-1} \times \mathbb{P}^{m-1}$. X_k is the k-secant variety of X_1, denoted as $\sigma_k X_1$. The matrices in X_k which minimize the distance from A are called the *best rank k approximations of A*.

Singular Value Decomposition

If M is a (real) $m \times n$ matrix, the SVD of M is

$$M = U \Sigma V^t$$

where
U is a $m \times m$ orthogonal matrix,
V is a $n \times n$ orthogonal matrix,
$\Sigma = D(\sigma_1, \ldots, \sigma_r)$ is a $m \times n$ diagonal matrix ,with its only nonzero values appearing on the diagonal $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$, which are called the singular values of M. They are the square roots of the eigenvalues of $M^t M$. r coincides with the rank of M.

how SVD can be computed

If M is symmetric then $U = V$ and SVD reduces to the spectral theorem $M = UDU^t$.
If u_i are the columns of U and v_i are the columns of V we get $Mv_i = \sigma_i u_i$, $M^t u_i = \sigma_i v_i$. u_i are just eigenvectors of MM^t, v_i are just eigenvectors of $M^t M$.

SVD of perturbed matrix

$$\left[\begin{array}{cccc|c} .105 & .14 & .07 & .035 & .35 \\ .03 & .04 & .02 & .01 & .1 \\ .03 & .04 & .02 & .01 & .1 \\ .135 & .18 & .09 & .045 & .45 \\ \hline .3 & .4 & .2 & .1 & \end{array}\right]$$

The 4×4 block A has $\mathrm{rk} = 1$, has singular values $0.321714, 0, 0, 0$.

$$A + P_\epsilon = \begin{bmatrix} \color{red}{.106} & \color{red}{.141} & \color{red}{.068} & .035 \\ \color{red}{.035} & \color{red}{.043} & \color{red}{.012} & .01 \\ \color{red}{.024} & \color{red}{.036} & \color{red}{.03} & .01 \\ .135 & .18 & .09 & .045 \end{bmatrix}$$

$A + P_\epsilon$ has singular values $0.32196, 0.01560, 0.00034, 0$

Two singular values are small. In the Frobenius norm, the distance of A from rank 2-matrices is .00034. The distance of A from rank 1-matrices is $\sqrt{(.01560)^2 + (.00034)^2} = 0.1561$

Least square

First application of SVD is the solution of least square problem. Given M and b, let $M = U\Sigma V^t$ its SVD, then the minimum of $||Mx - b||$ is computed by $x = V\Sigma^+ U^t b$ where in Σ^+ appear $\sigma_1^{-1}, \sigma_2^{-1}, \ldots$ (indeed what we wrote is the Moore-Penrose inverse of M).
The r-th row of V is the vector which minimizes the norm $||Mx||$ under the condition $||x|| = 1$.

Best rank k approximation

Recall $X_k = \{\text{matrices of rank } \leq k\}$
so $X_1 \subset X_2 \subset X_3 \subset \ldots$
The Frobenius norm is $\|A\|_F = \sqrt{\sum_{i,j} \|a_{ij}^2\|} = \sqrt{\mathrm{tr}(A^t A)}$.

Theorem (Eckart-Young, 1936)

- *The best rank k approximation of $A = U\Sigma V^t$, where $\Sigma = D(\sigma_1, \ldots, \sigma_r)$, is $U\Sigma_k V^t$ where $\Sigma_k = D(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0)$.*
- *The distance of A from X_k in the Frobenius norm is*

$$\min_{rkB \leq k} \|A - B\|_F = \sqrt{\sum_{i \geq k+1} \sigma_i^2}$$

Examples of SVD

SVD computes the (orthogonal) decomposition of a matrix as a sum of rank 1 matrices, indeed if u_i are the columns of U, and v_i are the columns of V, we get $M = \sum_{i=1}^{r} \sigma_i (u_i v_i^t)$. To give an example, let

$M = \begin{pmatrix} 2 & 10 & 2 \\ 2 & -1 & -3 \\ 1 & 7 & 0 \end{pmatrix}$. Then the SVD decomposition of M is

$$M = \begin{bmatrix} -.82631 & .0446582 & -.561442 \\ .0883261 & -.974242 & -.207488 \\ -.556246 & -.22104 & .801082 \end{bmatrix} \cdot \begin{bmatrix} 12.5599 & 0 & 0 \\ 0 & 3.66274 & 0 \\ 0 & 0 & .912972 \end{bmatrix} \cdot \begin{bmatrix} -.161802 & -.974941 & -.152676 \\ -.567938 & -.0345247 & .822347 \\ -.807011 & .219768 & -.54812 \end{bmatrix}$$

$$= 12.5599 \begin{bmatrix} -.82631 \\ .0883261 \\ -.556246 \end{bmatrix} \cdot \begin{bmatrix} -.161802 & -.974941 & -.152676 \end{bmatrix} +$$

$$3.66274 \begin{bmatrix} .0446582 \\ -.974242 \\ -.22104 \end{bmatrix} \cdot \begin{bmatrix} -.567938 & -.0345247 & .822347 \end{bmatrix} +$$

$$.912972 \begin{bmatrix} -.561442 \\ -.207488 \\ .801082 \end{bmatrix} \cdot \begin{bmatrix} -.807011 & .219768 & -.54812 \end{bmatrix}$$

Best rank 2 approximation, according to Eckart-Young Theorem.

The best rank 2 approximation forgets the third singular value (and its successors), so forget the blue.
It is

$$M' = \begin{bmatrix} -.82631 & .0446582 \\ .0883261 & -.974242 \\ -.556246 & -.22104 \end{bmatrix} \cdot \begin{bmatrix} 12.5599 & 0 \\ 0 & 3.66274 \end{bmatrix} \cdot \begin{bmatrix} -.161802 & -.974941 & -.152676 \\ -.567938 & -.0345247 & .822347 \end{bmatrix}$$

$$= 12.5599 \begin{bmatrix} -.82631 \\ .0883261 \\ -.556246 \end{bmatrix} \cdot \begin{bmatrix} -.161802 & -.974941 & -.152676 \end{bmatrix} +$$

$$3.66274 \begin{bmatrix} .0446582 \\ -.974242 \\ -.22104 \end{bmatrix} \cdot \begin{bmatrix} -.567938 & -.0345247 & .822347 \end{bmatrix}$$

M' has rank two and $||M - M'|| = .912972$

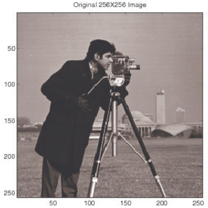
Knowledge of rank saves informations collecting data

A random matrix of format $r \times c$ needs rc entries to be known, and you cannot reduce these informations.
For a matrix of format $r \times c$ of rank one, the expression $a_{ij} = x_i y_j$ saves memory and requires just $r + c$ informations.
Let's visualize it

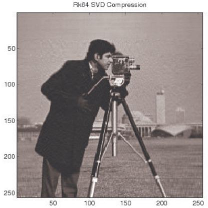
$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} y_0 & y_1 & y_2 & y_3 & y_4 \end{pmatrix} = \begin{pmatrix} x_0 y_0 & x_0 y_1 & x_0 y_2 & x_0 y_3 & x_0 y_4 \\ x_1 y_0 & x_1 y_1 & x_1 y_2 & x_1 y_3 & x_1 y_4 \\ x_2 y_0 & x_2 y_1 & x_2 y_2 & x_2 y_3 & x_2 y_4 \\ x_3 y_0 & x_3 y_1 & x_3 y_2 & x_3 y_3 & x_3 y_4 \end{pmatrix}$$

Images created by Emanuele Frandi, Alessandra Papini (Universita' di Firenze).

SVD can be used to compress images.
Original image is 256×256

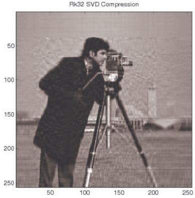


rank 256

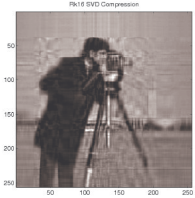


rank 64

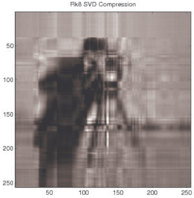
Images of small rank



rank 32



rank 16



rank 8

SVD and critical points, the geometric point of view.

Lemma (SVD revisited)

Let $A = XDY$ be the SVD decomposition of a matrix A. The critical points of the distance function $d_A = d(A, -)$ from A to the variety of rank one matrices are given by $\sigma_i x_i \otimes y_i$, where x_i are the columns of X, y_j are the rows of Y and D has σ_i on the diagonal.

(x_i, y_i) are called singular pairs of A.

Critical points in any rank.

Lemma (SVD revisited, again.)

Let $A = XDY$ be the SVD decomposition of A. The critical points of the distance function $d_A = d(A, -)$ from A to the variety X_k of rank k matrices are given by

$$\Sigma_{j\in\{i_1,\ldots,i_k\}}\sigma_j x_j \otimes y_j$$

for any subset of indices $\{i_1, \ldots, i_k\}$, where x_i are the columns of X, y_j are the rows of Y and D has σ_i on the diagonal.

Tangent space at $\Sigma_{j\in\{i_1,\ldots,i_k\}}\sigma_j x_j \otimes y_j$ is the sum of individual tangent spaces, according to Terracini Lemma.

The number of critical points for A of rank $r \geq k$ is $\binom{r}{k}$. For a general $m \times n$ matrix, assuming $m \leq n$, it is $\binom{m}{k}$.

General conclusions.

- The sums of the critical points on X_1 give all the critical points on the k-secant variety $\sigma_k X_1$. This fact does NOT generalize to other varieties X.

- If X is a critical point of d_A on $\sigma_k X_1$, then $A - X$ is a critical point of d_A on $\sigma_{r-k} X_1$. This fact GENERALIZES to other varieties X, in a proper way that we will see.