ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG
INSTITUT FÜR INFORMATIK
Lehrstuhl für Mustererkennung und Bildverarbeitung
Prof. Dr.-Ing. Hans Burkhardt

Georges-Köhler-Allee
Geb. 052, Zi 01-029
D-79110 Freiburg
Tel. 0761 - 203 - 8260

Algorithmen zur digitalen Bildverarbeitung I

Pseudoinverse: A draft tutorial

Author: D. Katsoulas

1 Singular Value Decomposition

1.1 Definition

For every real matrix A of dimensions $m \times n$, there exist two *orthogonal* matrices: $U = [u_1, ..u_m] \in \mathbf{R}^{m\times m}$ and $V = [v_1, ..v_n] \in \mathbf{R}^{n\times n}$ such that

$$\Sigma = U^T AV = \text{diag}(\sigma_1, \sigma_2, .., \sigma_p) \in \mathbf{R}^{m\times n}, \tag{1}$$

where $p = \min\{m, n\}$, and $\sigma_1 \geq \sigma_2 \geq .. \geq \sigma_p$, or equivalently (due to orthogonality of U, V):

$$A = U\Sigma V^T \tag{2}$$

The $\sigma_i, i \in 1..p$ are the *singular values* of A and this method for decomposing A according to eq. (2) is widely known as the *Singular Value Decomposition* (or SVD) of the matrix A.

1.2 Properties

The SVD reveals a lot about the internals of a matrix A. If we define as σ_r the smallest non zero diagonal element of Σ, that is :

$$\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r \geq \sigma_{r+1} = .. = \sigma_p = 0, \tag{3}$$

then the following hold:

$$\text{rank}(A) = r \tag{4}$$

$$N(A) = span\{v_{r+1}, .., v_n\} \tag{5}$$

$$R(A) = span\{u_1, .., u_r\}, \tag{6}$$

where $N(A)$ and $R(A)$ the null space and the range of the matrix A respectively.

The most appealing property of SVD, is expressed by the following theorem:

Theorem: Given the linear system of equations:

$$Ax = y, \tag{7}$$

then the solution $x_{LS} \in \mathbf{R}^n$ of this system of the form:

$$x_{LS} = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i, \tag{8}$$

satisfies the following:

$$\rho_{LS} = \|Ax_{LS} - b\|_2 \quad = \quad \min_x \|Ax - b\|_2 \tag{9}$$

$$\rho_{LS}^2 = ||Ax_{LS} - b||_2^2 \quad = \quad \sum_{i=r+1}^{m} (u_i^T b)^2 \tag{10}$$

$$\forall x_1 \in \mathbf{R}^n, x_1 = \min_x \|Ax - b\|_2 \quad : \quad \|x_{LS}\|_2 \leq \|x\|_2 \tag{11}$$

Proof: Due to the orthogonality of U, V, we have that: $||U^T||_2^2 = 1$, and $VV^T = I$. If so then:

$$\|Ax - b\|_2 = \left\|U^T\right\|_2 \cdot \left\|A(VV^T)x - b\right\|_2 = \left\|(U^T AV)V^T x - U^T b\right\|_2 \tag{12}$$

Now setting $\alpha = V^T x$ and using eq. (1), eq. (12) becomes:

$$\|Ax - b\|_2^2 = \left\|\Sigma\alpha - U^T b\right\|_2^2 \tag{13}$$

but since:

$$\Sigma\alpha - U^T b = \begin{bmatrix} \sigma_1 a_1 - u_1^T b \\ \sigma_2 a_2 - u_2^T b \\ \vdots \\ \sigma_r a_r - u_r^T b \\ -u_{r+1}^T b \\ \vdots \\ -u_m^T b \end{bmatrix}_{m \times 1}, \tag{14}$$

we can write the squared 2-norm of the vector $Ax - b$ as

$$\|Ax - b\|_2^2 = \sum_{i=1}^{r} (\sigma_i \alpha_i - u_i^T b)^2 + \sum_{i=r+1}^{m} (u_i^T b)^2 \tag{15}$$

Our task is to find the vector x which minimizes $\|Ax - b\|_2^2$. Since the vector x depends on α, we can attempt to find the α minimizing $\|Ax - b\|_2^2$ instead. From eq. (15) we see that $\|Ax - b\|_2^2$ is minimized if $\alpha_i = \frac{u_i^T b}{\sigma_i}$ for $i = 1..r$. The elements α_i for $i = r + 1..m$ are not important for the minimization. However if we set them to zero, we get the vector α which minimizes $\|Ax - b\|_2^2$ on one hand, and has the smallest norm of all minimizers on the other. Since, $x = V\alpha$, and $\|x\|_2 = \|V\alpha\|_2 = \|\alpha\|_2$, the solution x_{LS} shown in eq. (8) minimizes $\rho_{LS} = ||Ax - b||_2$, so that eq. (9) is satisfied, and has the smallest 2-norm of all minimizers of ρ_{LS}, which means that eq. (11) is as well satisfied. By setting the α_i's computed above into eq. (15), we get the residual value appearing in eq. (10).

2 The Pseudoinverse matrix

2.1 Definition

Given the matrix A its *pseudoinverse* A^+ is a matrix of dimensions $n \times m$, which amounts to:

$$A^+ = V\Sigma^+ U^T \tag{16}$$

In the preceding equation, the matrix Σ^+ is a diagonal matrix of the form:

$$\Sigma^+ = \text{diag}(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, .., \frac{1}{\sigma_r}, 0, .., 0) \in \mathbf{R}^{n \times m} \tag{17}$$

If so, then:

$$A^+ b = (V\Sigma^+ U^T)b = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i = x_{LS} \tag{18}$$

2.2 Insight

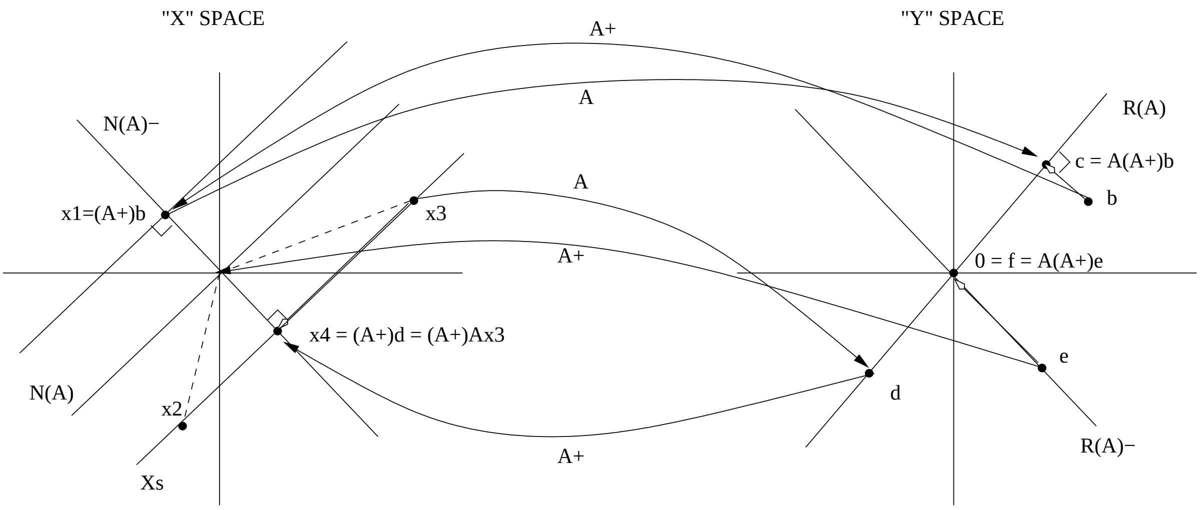

Abbildung 1: Geometric interpretation of pseudoinverse functionality.

Let's suppose we are interested in solving the linear system (7).

To ease the illustration of the problem, we suppose that A is a 2×2 matrix. We also assume that $dim(N(A)) = 1$. If X the vector space in which x belongs, and since $dim(X) = dim(R(A)) + dim(N(A))$, we deduce that $dim(R(A)) = 1$ for our example.

We suppose as well that Y is the vector space in which the vector y belongs. The two-dimensional vector spaces X and Y are depicted in Fig. 1. The one-dimensional subspaces $N(A)$ and $R(A)$ of X and Y respectively are depicted as lines in the figure. Note that in the figure, $R(A)-$ and $N(A)-$, stand for $R(A)^\perp$ and $N(A)^\perp$ respectively.

For solving (7), we consider the following two cases:

$\mathbf{y = d}$ and $\mathbf{d \in R(A)}$: The system has an infinite number of solutions. If x_3 a solution to the system, then $x_s = x_3 + x_N$, where $x_N \in N(A)$, is as well a solution. All those solutions belong to a one-dimensional translate of $N(A)$, illustrated by the line X_s in the figure. This line passes through x_3, and is parallel to the line defined by $N(A)$. As implied by equations (18), (11), from all the solutions of the system, the solution delivered by the pseudoinverse $x_4 = A^+d$ will be the one with the minimum norm, that is the minimum distance from the origin of the X space. The distance of x_4 to the origin is minimized only if x_4 belongs to the orthogonal space of $N(A)$, that is, $x_4 \in N(A)^\perp$.

Note that since x_3 a solution of $Ax = d$, we have that $Ax_3 = d$. In addition, $x_4 = A^+d$. Thereby, $x_4 = A^+Ax_3$. This means that the matrix A^+A, projects vectors of X-space to $N(A)^\perp$. This projection is orthogonal, since $N(A)^\perp$ is orthogonal to the X_s line.

$\mathbf{y = b}$ **and** $\mathbf{b \notin R(A)}$. In this case the system has no solution. However, as implied by equations (18), (9), the "solution" delivered by the pseudoinverse: $x_1 = A^+b$, when mapped via A to the Y-space, gives a vector $c = Ax_1$, on the R(A), so that the distance of c to b is minimized.

Since $c = Ax_1$ and $x_1 = A^+b$, we get $c = AA^+b$. Since the vector c is an element of $R(A)$ with the minimum distance to b, it is the orthogonal projection of b to $R(A)$. The matrix AA^+, defines therefore an orthogonal projection of a vector of the Y space to the R(A).

Note that in every case, the matrix A^+, takes vectors of the Y- space and brings them to vectors of $N(A)^\perp$. This implies that: $R(A^+) = N(A)^\perp$. Those vectors of $N(A)^\perp$, solve the system $Ax = y_R$, where y_R the orthogonal projection of y to R(A) on one hand, and have minimum distance from the origin of the X- space on the other.

Lets consider now the case, where $Ax = e$, $e \in R(A)^\perp$. The orthogonal projection of e to $R(A)$ is the null vector in this case. The vector which solves the system $Ax = 0$ with minimum norm is the null vector. In other words the pseudoinverse maps vectors of $R(A)^\perp$ to the null vector. That is : $N(A^+) = R(A)^\perp$.

Based on these observations, we could conclude the following regarding the properties of the pseudoinverse matrix:

2.3 Properties

1. $R(A^+) = N(A)^\perp$. Since $N(A)^\perp = R(A^T)$ (why?) we finally get: $\mathbf{R(A^+) = N(A)^\perp = R(A^T)}$.

2. $N(A^+) = R(A)^\perp$. Since $R(A)^\perp = N(A^T)$ (why?) we finally get: $\mathbf{N(A^+) = R(A)^\perp = N(A^T)}$.

3. A^+A, defines the orthogonal projection of an arbitrary vector of the X space to $N(A)^\perp$. This has the following consequences:
   (a) $\mathbf{(A^+A)^T = A^+A}$ (why?)
   (b) The matrix $I - A^+A$ defines the orthogonal projection of the vector to $N(A)$.
   (c) Since $I - A^+A \in N(A)$, we have: $A(I - A^+A) = 0$, or $\mathbf{AA^+A = A}$.

4. AA^+, defines the orthogonal projection of an arbitrary vector of the Y-space to R(A). This has the following consequences:
   (a) $\mathbf{(AA^+)^T = AA^+}$ (why?)
   (b) The matrix $I - AA^+$ is the orthogonal projection of the vector to $R(A)^\perp$.
   (c) We have already shown that: $R(A)^\perp = N(A^+)$. Thereby, $A^+(I - AA^+) = 0$, or $\mathbf{A^+AA^+ = A^+}$.

3 Exercises

Based on what discussed in the previous sections show that:

1. $(A^T)^+ = (A^+)^T$
2. $A^{++} = A$
3. $A^T A A^+ = A^T$
4. $A^+ A A^T = A^+$
5. If $r = n$ then $A^+ = (A^T A)^{-1} A^T$
6. If $r = m$ then $A^+ = A^T (A A^T)^{-1}$

References:

[1] G. H. Golub and C. F. Van Loan *Matrix Computations.* John Hopkins University Press, 1989.

[2] W. Press, S. Teukolsky, W. Vetterling, B. Flannery. *Numerical Recipes in C++ (The Art of Scientific Computing), Second Edition.* Cambridge University Press, 2001.