Least squares method, pseudoinverse. Orthogonal matrices and isometries.

Zdeněk Dvořák

March 16, 2016

Definition 1. *Let* $\mathbf{V}$ *be an inner product space and let* $\mathbf{U}$ *be its subspace of finite dimension. For* $v \in \mathbf{V}$*, the* <u>*orthogonal projection of* v *on* $\mathbf{U}$</u> *is the vector* $p \in \mathbf{U}$ *such that* $v - p \in \mathbf{U}^{\perp}$*.*

Lemma 1. *Let* $\mathbf{V}$ *be an inner product space and let* $\mathbf{U}$ *be its subspace of finite dimension. Let* $p \in \mathbf{U}$ *be the projection of* $v \in \mathbf{V}$*. Then* p *is the vector of* $\mathbf{U}$ *closest to* v*, that is,*

$$\|v - x\| > \|v - p\|$$

for every $x \in \mathbf{U} \setminus \{p\}$*.*

Lemma 2. *Let* $\mathbf{V}$ *be an inner product space and let* $\mathbf{U}$ *be its subspace of finite dimension. Let* $B = u_1, \ldots, u_k$ *be a (not necessarily othonormal) basis of* $\mathbf{U}$*. Let* $p \in \mathbf{U}$ *be the projection of* $v \in \mathbf{V}$ *on* $\mathbf{U}$*. Let*

$$G = \begin{pmatrix} \langle u_1, u_1 \rangle & \langle u_2, u_1 \rangle & \ldots & \langle u_k, u_1 \rangle \\ \langle u_1, u_2 \rangle & \langle u_2, u_2 \rangle & \ldots & \langle u_k, u_2 \rangle \\ & & \ldots & \\ \langle u_1, u_k \rangle & \langle u_2, u_k \rangle & \ldots & \langle u_k, u_k \rangle \end{pmatrix}.$$

Then G *is a regular matrix and*

$$G[p]_B^T = (\langle v, u_1 \rangle, \ldots, \langle v, u_k \rangle)^T.$$

Proof. Let $[p]_B = (\alpha_1, \ldots, \alpha_k)$. For $i = 1, \ldots, k$, we have

$$\begin{aligned} (G[p]_B^T)_i &= \langle u_1, u_i \rangle \alpha_1 + \langle u_2, u_i \rangle \alpha_2 + \ldots + \langle u_k, u_i \rangle \alpha_k \\ &= \langle \alpha_1 u_1 + \ldots + \alpha_k u_k, u_i \rangle = \langle p, u_i \rangle . \end{aligned}$$

Since $v - p \in \mathbf{U}^{\perp}$, we have $\langle v - p, u_i \rangle = 0$ for $i = 1, \ldots, k$, and thus $\langle p, u_i \rangle = \langle v, u_i \rangle$. Hence, the equality follows.

Suppose that $x = (\beta_1, \ldots, \beta_n)^T$ is a solution of the system $Gx = 0$. Let $u = \beta_1 u_1 + \ldots + \beta_k u_k$. Then, $\langle u, u_i \rangle = \beta_1 \langle u_1, u_i \rangle + \ldots + \beta_k \langle u_k, u_i \rangle = 0$ for $i = 1, \ldots, k$, and thus $u \in \{u_1, \ldots, u_k\}^\perp = \mathbf{U}^\perp$. However, $u \in \text{span}(u_1, \ldots, u_k) = \mathbf{U}$. Since $u \in \mathbf{U} \cap \mathbf{U}^\perp$, it follows that $u = o$, and thus $x = (0, \ldots, 0)$. Consequently, G is regular. □

Example 1. *Let $\mathbf{U} = span((1,1,1),(1,2,3))$ be a plane in $\mathbf{R}^3$. Determine the distance of the point $v = (3,5,1)$ from $\mathbf{U}$, without finding an orthogonal basis of $\mathbf{U}$.*

Let $u_1 = (1,1,1)$ and $u_2 = (1,2,3)$. We have

$$G = \begin{pmatrix} \langle u_1, u_1 \rangle & \langle u_2, u_1 \rangle \\ \langle u_1, u_2 \rangle & \langle u_2, u_2 \rangle \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}$$

The solution to the system $Gx = (\langle v, u_1 \rangle, \langle v, u_2 \rangle)^T = (9, 16)$ is $x = (5, -1)$. Hence, the projection of v on $\mathbf{U}$ is $5u_1 - u_2 = (4,3,2)$, and the distance from v to $\mathbf{U}$ is $\|v - (4,3,2)\| = \sqrt{6}$.

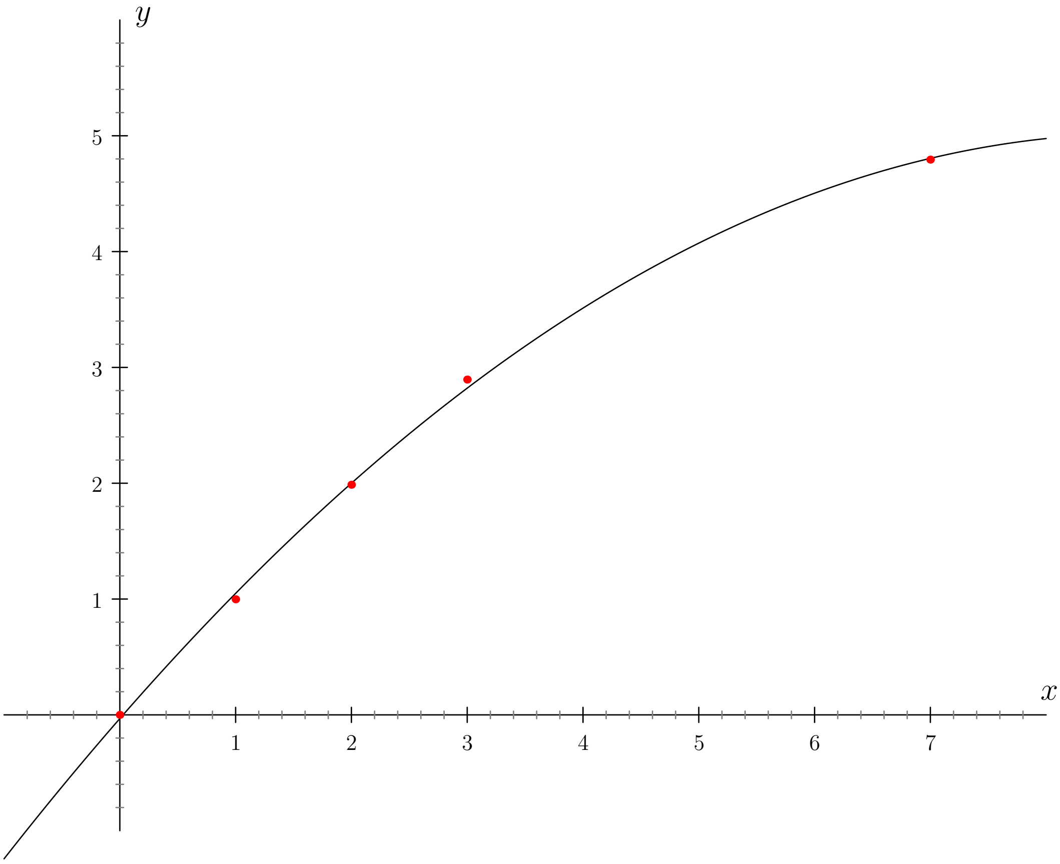
Corollary 3. *Consider the Euclidean space $\mathbf{R}^n$ with the inner product defined as the dot product, and let $\mathbf{U}$ be its subspace. Let $p : \mathbf{R}^n \to \mathbf{R}^n$ be the function that maps each vector to its projection on $\mathbf{U}$. Let $B = u_1, \ldots, u_k$ be a (not necessarily othonormal) basis of $\mathbf{U}$. Let $A = (u_1|u_2|\ldots|u_k)$. Then $p(v) = A(A^TA)^{-1}A^Tv$, and thus $A(A^TA)^{-1}A^T$ is the matrix of the function p (with respect to the canonical basis of $\mathbf{R}^n$).*

Proof. Let G be the matrix from Lemma 2. Note that $\langle u_i, u_j \rangle = u_i^T u_j$, and thus $G = A^TA$. Similarly, $b = (\langle v, u_1 \rangle, \ldots, \langle v, u_k \rangle)^T = A^Tv$. By Lemma 2, if the coordinates of $p(v)$ with respect to the basis B are $(\alpha_1, \ldots, \alpha_k)$, then $(\alpha_1, \ldots, \alpha_k)^T = G^{-1}b = (A^TA)^{-1}A^Tv$. It follows that $p(v) = \alpha_1 u_1 + \ldots + \alpha_k u_k = A(\alpha_1, \ldots, \alpha_k)^T = A(A^TA)^{-1}A^T$. □

1 Least squares method and pseudoinverse

Example 2. *Suppose we measured the following dependence of some quantity on time:*

t	*0*	*1*	*2*	*3*	*7*
$f(t)$	*0.000*	*0.998*	*1.987*	*2.855*	*4.794*



Let $S = \{0, 1, 2, 3, 7\}$. Find the approximation of f by a quadratic polynomial p such that $\sum_{t \in S}(f(t) - p(t))^2$ is minimum.

Consider the space $\mathbf{V}$ of functions $S \to \mathbf{R}$, with inner product $\langle g_1, g_2 \rangle = \sum_{t \in S} g_1(t)g_2(t)$. Let $p_1(t) = 1$, $p_2(t) = t$ and $p_3(t) = t^2$ be elements of $\mathbf{V}$. Any quadratic polynomial is a linear combination of p_1, p_2 and p_3. Hence, p is the projection of f on span(p_1, p_2, p_3). Let

$$G = \begin{pmatrix} \langle p_1, p_1 \rangle & \langle p_2, p_1 \rangle & \langle p_3, p_1 \rangle \\ \langle p_1, p_2 \rangle & \langle p_2, p_2 \rangle & \langle p_3, p_2 \rangle \\ \langle p_1, p_3 \rangle & \langle p_2, p_3 \rangle & \langle p_3, p_3 \rangle \end{pmatrix} = \begin{pmatrix} 5 & 11 & 63 \\ 11 & 63 & 379 \\ 63 & 379 & 2499 \end{pmatrix}$$

and

$$b = (\langle f, p_1 \rangle, \langle f, p_2 \rangle, \langle f, p_3 \rangle)^T = (10.634, 47.095, 269.547)^T.$$

By Lemma 2, the coordinates of p with respect to the basis p_1, p_2, p_3 are the solution to the system $Gx = b$, which is

$$x \approx (-0.032, 1.146, -0.065)^T.$$

Hence, $p \approx -0.032 + 1.146t - 0.065t^2$.

Example 3. *Another way of viewing Example 2: Suppose that* $p = \alpha_0 + \alpha_1 t + \alpha_2 t^2$ *for some (unknown) coefficients* α_0, α_1 *and* α_2. *Let*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 7 & 49 \end{pmatrix}.$$

Then

$$(p(0), p(1), p(2), p(3), p(7))^T = A(\alpha_0, \alpha_1, \alpha_2)^T.$$

Ideally, we would like to have $p = f$, *and thus* $(\alpha_0, \alpha_1, \alpha_2)$ *would be a solution to the system*

$$A(\alpha_0, \alpha_1, \alpha_2)^T = (0.000, 0.998, 1.987, 2.855, 4.794)^T.$$

However, this system has no solution, and thus we want to find $(\alpha_0, \alpha_1, \alpha_2)$ *so that* $A(\alpha_0, \alpha_1, \alpha_2)^T$ *differs from* $(0.000, 0.998, 1.987, 2.855, 4.794)^T$ *(in the Euclidean norm) as little as possible.*

Lemma 4. *Let* A *be an* $m \times n$ *real matrix of rank* n, *let* b *be a column vector of* m *real numbers, and let* x *be such* $\|Ax - b\|$ *is minimum. Then* $x = (A^T A)^{-1} A^T b$.

Proof. Observe that Ax is the projection of b on the column space of A, and thus by Corollary 3,

$$Ax = A(A^T A)^{-1} A^T b.$$

By comparing the sides, we see that we can choose $x = (A^T A)^{-1} A^T b$. Note that x is unique, by the uniqueness property from Lemma 1 and the assumption that A has full column rank. □

Let us remark that the previous lemma can be modified to handle the case when A does not have full column rank: then, x can be chosen as any of the (infinitely many) solutions to the system $A^T Ax = A^T b$.

The matrix $(A^T A)^{-1} A^T$ is the <u>pseudoinverse</u> to A (and if A is regular, its pseudoinverse is equal to A^{-1}). The pseudoinverse can be defined (in a somewhat more complicated way) even if A does not have full column rank.

2 Orthogonal matrices and isometries

Definition 2. *Let* $\mathbf{V}$ *be an inner product space over* $\mathbf{R}$. *A function* $f : \mathbf{V} \to \mathbf{V}$ *is an* *isometry* *if* $\|f(x) - f(y)\| = \|x - y\|$ *for every* $x, y \in \mathbf{V}$.

Examples: rotations, reflections, translations, ...

Proposition 5. *Any isometry of an inner product space is an affine function; and thus, if* $f : \mathbf{V} \to \mathbf{V}$ *is an isometry and* $f(o) = o$, *then* f *is a linear function.*

We skip the proof of this proposition, which requires a bit of math analysis.

Lemma 6. *Let* $\mathbf{V}$ *be an inner product space over* $\mathbf{R}$. *Let* $f : \mathbf{V} \to \mathbf{V}$ *be a linear function. The following claims are equivalent:*

1. f *is an isometry*
2. f *preserves the norm, that is,* $\|f(x)\| = \|x\|$ *for every* $x \in \mathbf{V}$.
3. f *preserves the inner product, that is,* $\langle f(x), f(y)\rangle = \langle x, y\rangle$ *for every* $x, y \in \mathbf{V}$.

Proof. Since f is linear, $\|f(x) - f(y)\| = \|f(x-y)\|$. If f preserves the norm, then $\|f(x - y)\| = \|x - y\|$ as required. Conversely, if f is an isometry, then $\|f(x)\| = \|f(x) - f(o)\| = \|x - o\| = \|x\|$.

If f preserves the inner product, then $\|f(x)\| = \sqrt{\langle f(x), f(x)\rangle} = \sqrt{\langle x, x\rangle} = \|x\|$, and thus it preserves the norm. Conversely, if f preserves the norm, then $\langle f(x), f(y)\rangle = \frac{1}{2}\left(\|f(x + y)\|^2 - \|f(x)\|^2 - \|f(y)\|^2\right) = \frac{1}{2}\left(\|x + y\|^2 - \|x\|^2 - \|y\|^2\right) = \langle x, y\rangle$. □

Lemma 6 shows that isometries also preserve angles.

Definition 3. *A square matrix* Q *is* *orthogonal* *if* $Q^T Q = I$.

Lemma 7. *Let* $\mathbf{V}$ *be an inner product space over* $\mathbf{R}$. *Let* $f : \mathbf{V} \to \mathbf{V}$ *be a linear function. Let* $B = v_1, \dots, v_n$ *and* C *be orthonormal bases of* $\mathbf{V}$. *Then* f *is an isometry if and only if* $[f]_{B,C}$ *is an orthogonal matrix.*

Proof. Recall that if $(\alpha_1, \dots, \alpha_n)$ and $(\beta_1, \dots, \beta_n)$ are the coordinates of vectors x and y with respect to an orthonormal basis, then $\langle x, y\rangle = \alpha_1\beta_1 + \ldots + \alpha_n\beta_n$. Hence, $\langle x, y\rangle = [x]_B[y]_B^T = [x]_C[y]_C^T$. Let $Q = [f]_{B,C}$.

Suppose that Q is orthogonal. Then $\langle f(x), f(y)\rangle = [f(x)]_C[f(y)]_C^T = (Q[x]_B^T)^T(Q[y]_B^T) = [x]_B(Q^TQ)[y]_B^T = [x]_B[y]_B^T = \langle x, y\rangle$, hence f preserves the inner product, and thus f is an isometry.

Conversely, suppose that f is an isometry, and thus f preserves the inner product. Hence, $\langle x, y\rangle = \langle f(x), f(y)\rangle = [x]_B(Q^TQ)[y]_B^T$ for every $x, y \in \mathbf{V}$, and in particular, $\langle v_i, v_j\rangle = [v_i]_B(Q^TQ)[v_j]_B^T = e_i(Q^TQ)e_j^T = (Q^TQ)_{ij}$. Since $\langle v_i, v_j\rangle = 0$ if $i \neq j$ and $\langle v_i, v_i\rangle = 1$ for all i, it follows that Q^TQ is the identity matrix, and thus Q is orthogonal. $\square$

Since id is an isometry, this implies that the transition matrix $[\mathrm{id}]_{B,C}$ between orthonormal bases is orthogonal.

Lemma 8. *For any $n \times n$ matrix Q, the following claims are equivalent:*

1. *Q is orthogonal.*
2. *Q is regular and $Q^{-1} = Q^T$.*
3. *Q^T is orthogonal.*
4. *$QQ^T = I$.*
5. *Q is regular and Q^{-1} is orthogonal.*
6. *The rows of Q form an orthonormal basis of $\mathbf{R}^n$.*
7. *The columns of Q form an orthonormal basis of $\mathbf{R}^n$.*

Proof. From the definition of the orthogonal matrix, $Q^TQ = I$, and thus $Q^{-1} = Q^T$. Hence, $I = QQ^{-1} = QQ^T = (Q^T)^TQ^T$, and thus Q^T is orthogonal. Also, $(Q^{-1})^T = (Q^T)^T = Q$, and thus $(Q^{-1})^TQ^{-1} = QQ^{-1} = I$ and Q^{-1} is orthogonal. The reverse implications follow by symmetry.

Also, note that $(Q^TQ)_{ij}$ is equal to the dot product of i-th and the j-th column of Q. Hence, $Q^TQ = I$ if and only if the set of columns of Q is orthonormal, and similarly $QQ^T = I$ if and only if the set of rows of Q is orthonormal. $\square$

Lemma 9. *The product of two orthogonal matrices is orthogonal.*

Proof. If $Q_1^TQ_1 = I$ and $Q_2^TQ_2 = I$, then $(Q_1Q_2)^T(Q_1Q_2) = Q_2^TQ_1^TQ_1Q_2 = Q_2^TQ_2 = I$. $\square$