where $a = (a_1, \ldots, a_d)$.

Thus, we are looking for a unit vector a solving $(X - \mu)a = 0$ in the least squares sense, that is, some a such that $a^\top a = 1$ minimizing

$$a^\top (X - \mu)^\top (X - \mu)a.$$

Compute some SVD $VDU^\top$ of $X - \mu$, where the main diagonal of D consists of the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$ of $X - \mu$ arranged in descending order. Then

$$a^\top (X - \mu)^\top (X - \mu)a = a^\top U D^2 U^\top a,$$

where $D^2 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2)$ is a diagonal matrix, so pick a to be *the last column in* U (corresponding to the smallest eigenvalue σ_d^2 of $(X - \mu)^\top (X - \mu)$). This is a solution to our best fit problem.

Therefore, if U_{d-1} is the linear hyperplane defined by a, that is,

$$U_{d-1} = \{u \in \mathbb{R}^d \mid \langle u, a \rangle = 0\},$$

where a is the last column in U for some SVD $VDU^\top$ of $X - \mu$, we have shown that the affine hyperplane $A_1 = \mu + U_{d-1}$ is a best approximation of the data set $X_1, \ldots, X_n$ in the least squares sense.

It is easy to show that this hyperplane $A_1 = \mu + U_{d-1}$ minimizes the sum of the square distances of each X_i to its orthogonal projection onto A_1. Also, since U_{d-1} is the orthogonal complement of a, the last column of U, we see that U_{d-1} is spanned by the first $d-1$ columns of U, that is, the first $d - 1$ principal directions of $X - \mu$.

All this can be generalized to a *best* $(d-k)$*-dimensional affine subspace* A_k *approximating* $X_1, \ldots, X_n$ *in the least squares sense* $(1 \leq k \leq d-1)$. Such an affine subspace A_k is cut out by k independent hyperplanes H_i (with $1 \leq i \leq k$), each given by some equation

$$a_{i\,1}x_1 + \cdots + a_{i\,d}x_d + c_i = 0.$$

If we write $a_i = (a_{i\,1}, \cdots, a_{i\,d})$, to say that the H_i are independent means that $a_1, \ldots, a_k$ are linearly independent. In fact, we may assume that $a_1, \ldots, a_k$ form an *orthonormal system.*

Then finding a best $(d - k)$-dimensional affine subspace A_k amounts to solving the homogeneous linear system

$$\begin{pmatrix} X & \mathbf{1} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & X & \mathbf{1} \end{pmatrix} \begin{pmatrix} a_1 \\ c_1 \\ \vdots \\ a_k \\ c_k \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$