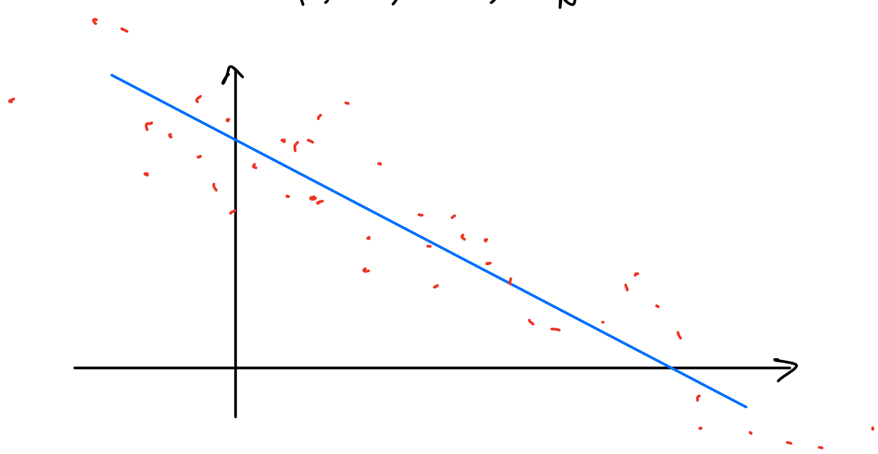
5374 Fall '22

Numerical Linear Algebra Lecture 25

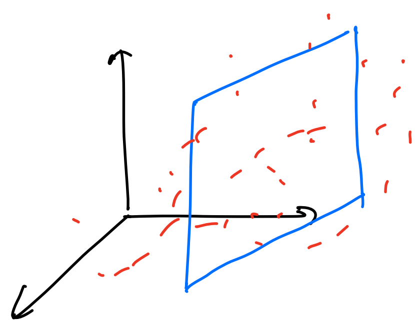
Today: A bit on Principal Components Analysis

The setup: we have a large set of points N in $\mathbb{R}^p$:

$$x_1, x_2, \ldots, x_N \in \mathbb{R}^p$$



The goal is to find a lower dimensional model that captures as much about this set as possible.

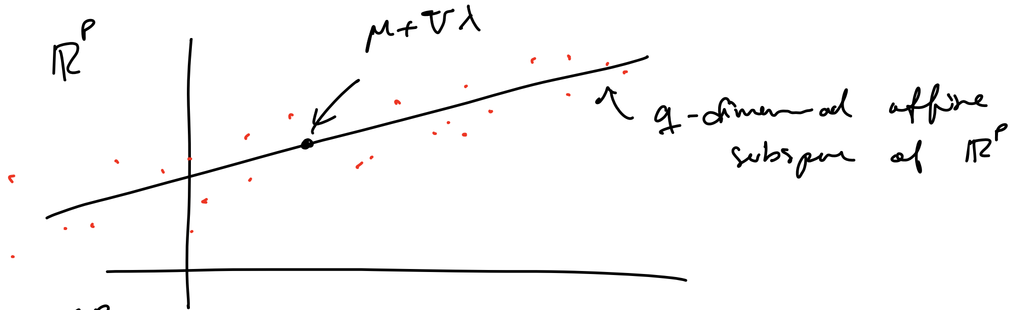


Part of the hope is that a lot of the variability in the data is due to small random errors that represent a simpler lower dimensional set.

We make the following (linear) model: to a good approximation,

$$(*) \quad x_k \approx \mu + V\lambda_k \qquad \text{for } k=1,\dots,N$$

where $\mu \in \mathbb{R}^p$, V is a $p\times q$ matrix whose columns are orthonormal (so, $q \le p$) and $\lambda_1,\dots,\lambda_N \in \mathbb{R}^q$



Fix $q \le p$

We seek out a triad $(\mu, V, \{\lambda_k\}_{k=1}^N)$ minimizing the total quadratic error:

$$\sum_{k=1}^{N} \| x_k - \mu - V\lambda_k \|_2^2$$

$$\begin{array}{c} q \\ \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{pmatrix} \end{array} \Big\} q$$

Constrained to: $\mu \in \mathbb{R}^p$

V is $p\times q$, $V^t V = I_{q\times q}$

$\lambda_k \in \mathbb{R}^q \quad (k=1,\dots,N)$

Exercise : (The case $q=1$)

$$F(\mu, v, \lambda_1, \dots, \lambda_N) = \sum_{k=1}^{N} \tfrac{1}{2} \| x_k - \mu - \lambda_k v \|_2^2$$

(when v is simply a unit vector)

Check the following:

$$\nabla_\mu F = - \sum_{k=1}^{N} x_k - \mu - \lambda_k v$$

$$\nabla_v F = - \sum_{k=1}^{N} \lambda_k (x_k - \mu - \lambda_k v)$$

$$\partial_{\lambda_j} F = -(x_j - \mu - \lambda_j v, v)$$
$$= -(x_j, v) + (\mu, v) - \lambda_j$$

Once you have checked these identities find the relation between $\mu, v, \lambda_1, \dots, \lambda_N$ when they correspond to a minimum.

The problem of finding this best q-dimensional linear fit can be dealt with using the SVD, as we know explain.

First, let's write this problem in form of matrices, where, for convenience we will an

that the best μ is $\mu = 0$
(in general, given data $x_1, \dots, x_n$, we take

$$\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x_k$$

and let $\tilde{x}_k = x_k - \bar{x}$, for this new data set, the best μ will be zero).

Then, we want to minimize

$$\sum_{k=1}^{N} \| x_k - V \lambda_k \|_2^2$$

Let $X = (x_1 \ x_2 \ \dots \ x_N)$ be a $p \times N$ matrix, and let $\Lambda = (\lambda_1 \ \dots \ \lambda_N)$ be a $q \times N$ matrix. Then observe

$$x_k = X e_k$$

$$V \lambda_k = (V \Lambda) e_k$$

$$e_k = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \end{pmatrix} \leftarrow k\text{-th pos}$$

$(e_k \in \mathbb{R}^N)$

Then the error can be written as

$$= \sum_{k=1}^{N} \| (X - V\Lambda) e_k \|_2^2$$

$$= \| X - V \Lambda \|_{Fr}^2 \quad \left(=: \operatorname{tr}\left((X - V\Lambda)^{T} (X - V\Lambda) \right) \right)$$

Moreover, since by constraint, $V^t V = I_{q\times q}$, the matrix V has rank q, and thus $V\Lambda$ has rank $\leq q$.

Last time we learned that the SVD provides for every $q \leq \operatorname{rank}(A)$ the best approximat of rank q:

$$\text{If} \quad A \overset{(**)}{=} \sigma_1\, u_1 \otimes v_1 + \cdots + \sigma_r\, u_r \otimes v_r \qquad (r = \operatorname{rank}(A))$$

$$A_q := \sigma_1\, u_1 \otimes v_1 + \cdots + \sigma_q\, u_q \otimes v_q \qquad (\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r)$$

Then $\|A - A_q\|_{Fr} \leq \|A - B\|_{Fr}$ for every matrix of rank q.

A bit of (non-elementary) computations can show that the best q-linear fit is given by the first q column of the right singular vector in the SVD, i.e. $v_1, v_2, \ldots, v_q$ in the SVD $(**)$

Iterative Methods

Gradient Descent

This is a method apt for solving equations of the form $Ax = b$ when A is symmetric or self-adjoint ($A^t = A$ or $A^* = A$).

Earlier this semester we looked at things like

$$f(x) = \frac{1}{2}(Ax-b, Ax-b) = \frac{1}{2}\|Ax-b\|_2^2$$

Now we shall look at something very similar

$$f(x) = \frac{1}{2}(Ax, x) - (b, x) \quad , \quad x \in \mathbb{R}^n$$

Unlike the function above which is interesting for any matrix A, this last function is of interest mostly when A is symmetric.

Exercise: Show that if $A^t = A$, then

$$\nabla f(x) = Ax - b \quad , \quad D^2 f(x) = A$$

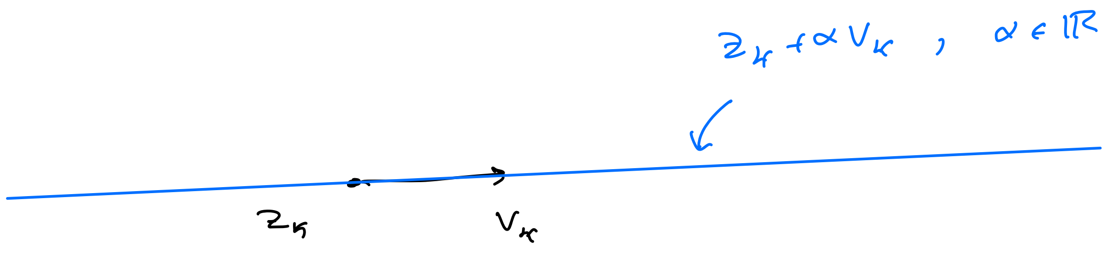
From the exercise we see that if A is symmetric and positive definite then the unique solution x_* to $Ax=b$, is the unique minimizer of $f(x)$. This observation is the starting point for all descent/optimization methods for solving linear equations.

Line searches

Fix A, b as before and let us write

$$f(z) = \frac{1}{2}(Az, z) - (z, b) \quad , \quad z \in \mathbb{R}^n$$

I want to generate a sequence $z_1, z_2, \ldots$ of points where $f(z)$ is becoming smaller and smaller, suppose that I have a way of choosing directions $v_1, v_2, \ldots, v_k, \ldots$. At stage k, I start at z_k and move in the direction v_k to find a smaller value of f.



Let me consider the following function of the single variable α:

$$f_{z_k, v_k}(\alpha) = f(z_k + \alpha v_k)$$

$$= \frac{1}{2}\left(A(z_k + \alpha v_k), z_k + \alpha v_k\right) - (z_k + \alpha v_k, b)$$

$$= \frac{1}{2}(A z_k, z_k) + (A \alpha v_k, z_k) + \frac{1}{2}(A \alpha v_k, \alpha v_k) - (z_k, b) - \alpha (v_k, b)$$

$$= \underbrace{\frac{1}{2}(A z_k, z_k) - (z_k, b)}_{f(z_k)} + \alpha\left((A v_k, z_k) - (v_k, b)\right) + \frac{1}{2}(A v_k, v_k)\alpha^2$$

Note: $(A v_k, z_k) - (v_k, b) = (v_k, A z_k - b)$
(since $A = A^+$)

Then, $f_{z_k, v_k}(\alpha) = f(z_k) + (A z_k - b, v_k)\alpha + \frac{1}{2}(A v_k, v_k)\alpha^2$

This is a convex parabola in α, and the min is obtained at α_k s.t.

$$\left(\frac{d}{d\alpha} f_{z_k, v_k}\right)(\alpha_k) = 0$$

$$(A z_k - b, v_k) + (A v_k, v_k)\alpha_k = 0$$

$\Longrightarrow$

$$a_k = - \frac{(Az_k - b, v_k)}{(Av_k, v_k)}$$