Matrix norms: connection to eigenvalues

The purpose of this document is to explain some details about the matrix 2-norm and its connections to eigenvalues. We will review some important aspects of eigenvector/eigenvalue problems.

Our starting point is the definition of the matrix 2-norm:

$$\|A\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|A\boldsymbol{x}\|_2. \tag{1}$$

We want to show that $\|A\|_2 = \sqrt{\max(\lambda_i)}$ where λ_i is an **eigenvalue** of A^tA, see Figure 1.

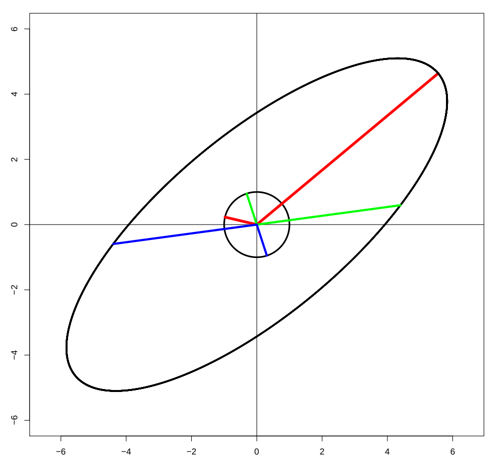

Figure 1: Geometric effect of transformation $A = \begin{pmatrix} -5 & 3 \\ -5 & -1 \end{pmatrix}$ on vectors $\boldsymbol{x}$ lying on the unit circle. The transformed vectors $A\boldsymbol{x}$ lie on an ellipse, since $A(-\boldsymbol{x}) = -A\boldsymbol{x}$, see the example vectors depicted by the smaller green and blue lines. The two red lines show the extreme case with highest geometric "stretching" effect: the pair consisting of the eigenvector $\boldsymbol{v}$ of A^tA with largest eigenvalue (the smaller red vector with length 1) and its image $A\boldsymbol{v}$, pointing into the direction of the major axis of the ellipse (as we will show below...).

In equation 1, our goal is to maximize $\sqrt{\boldsymbol{x}^t A^t A \boldsymbol{x}}$ under the constraint $\sqrt{\boldsymbol{x}^t \boldsymbol{x}} = 1$. First, lets try to simplify this problem by removing the square roots. For the constraint, this is easy: requiring that $\sqrt{\boldsymbol{x}^t \boldsymbol{x}} = 1$ is equivalent to requiring that $\boldsymbol{x}^t \boldsymbol{x} = 1$. But we can also remove the square root in $\sqrt{\boldsymbol{x}^t A^t A \boldsymbol{x}}$, since the vector $\boldsymbol{x}^*$ that maximizes $\sqrt{\boldsymbol{x}^t A^t A \boldsymbol{x}}$ is the same $\boldsymbol{x}^*$ that maximizes $\boldsymbol{x}^t A^t A \boldsymbol{x}$. To see this, note that

1. $\boldsymbol{x}^t A^t A \boldsymbol{x} \geq 0$ (because it is the length of the vector $A\boldsymbol{x}$),
2. The function $h(y) = y^2$ is **strictly monotone increasing** on $\mathbb{R}_+$.

Thus, if $\boldsymbol{x}^*$ is a maximizer of $\sqrt{\boldsymbol{x}^t A^t A \boldsymbol{x}}$, it is also a maximizer of $h(\sqrt{\boldsymbol{x}^t A^t A \boldsymbol{x}}) = \boldsymbol{x}^t A^t A \boldsymbol{x}$:

$$\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \sqrt{\boldsymbol{x}^t A^t A \boldsymbol{x}} \quad \Leftrightarrow \quad \boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \boldsymbol{x}^t A^t A \boldsymbol{x}. \tag{2}$$

Defining $M = A^t A$, the problem is then to maximize the function

$$f(\boldsymbol{x}) = \boldsymbol{x}^t M \boldsymbol{x}, \tag{3}$$

subject to the constraint

$$\boldsymbol{x}^t \boldsymbol{x} = 1. \tag{4}$$

Note that scaling the vector $\boldsymbol{x}$ by 2 (and thus using the constraint $\boldsymbol{x}^t \boldsymbol{x} = 4$) would not change anything, except that the solution would also be scaled by a factor of two: instead of observing the image $A\boldsymbol{x}$ of an input vector $\boldsymbol{x}$ with length 1, we would do the same for an input vector of length 2. The direction with largest geometric distortion, however, will not change in this case, only the length of the image $A\boldsymbol{x}$ will double. This means that the value of the constraint does not really matter, and the ratio

$$r(\boldsymbol{x}) = \frac{\boldsymbol{x}^t M \boldsymbol{x}}{\boldsymbol{x}^t \boldsymbol{x}} \tag{5}$$

remains unchanged under any such scaling operations. Therefore, we can equivalently optimize $r(\boldsymbol{x})$ in order to find the direction of largest "stretching", and then rescale the result accordingly.

For minimizing the function $r(\boldsymbol{x})$, we will first find all its stationary points by computing the gradient (i.e. the vector of partial derivatives), and then set this gradient to zero.

What we now need is a formula for computing the gradient of quadratic forms $\boldsymbol{x}^t M \boldsymbol{x}$, where $M = A^t A$:

$$\nabla_{\boldsymbol{x}}\, \boldsymbol{x}^t M \boldsymbol{x} = \begin{pmatrix} \frac{\partial \boldsymbol{x}^t M \boldsymbol{x}}{\partial x_1} \\ \vdots \\ \frac{\partial \boldsymbol{x}^t M \boldsymbol{x}}{\partial x_n} \end{pmatrix} = 2M\boldsymbol{x}. \tag{6}$$

This formula can be verified by rewriting $\boldsymbol{x}^t M \boldsymbol{x}$ as $\sum_i x_i \sum_j M_{ij} x_j$, differentiating with respect to the individual components x_k, exploiting the symmetry of M, i.e. $M_{ij} = M_{ji}$, and rearranging terms.

Applying equation (6), we arrive at

$$\begin{aligned} &\nabla_{\boldsymbol{x}} r(\boldsymbol{x}) \stackrel{!}{=} 0 \\ &\Rightarrow 2\frac{M\boldsymbol{x}}{\boldsymbol{x}^t\boldsymbol{x}} - 2\boldsymbol{x}^t M \boldsymbol{x} (\boldsymbol{x}^t \boldsymbol{x})^{-2} \boldsymbol{x} = 0 \\ &\Rightarrow M\boldsymbol{x} = \lambda \boldsymbol{x}, \end{aligned} \tag{7}$$

with

$$\lambda = \frac{\boldsymbol{x}^t M \boldsymbol{x}}{\boldsymbol{x}^t \boldsymbol{x}} = r(\boldsymbol{x}). \tag{8}$$

Equation $M\boldsymbol{x} = \lambda\boldsymbol{x}$ is fulfilled for all $\boldsymbol{x}$ that are **eigenvectors** of M. Note that $M = A^t A$ is a symmetric positive (semi-)definite matrix. So M has n **orthonormal eigenvectors** $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ (i.e. the eigenvectors have length one and they are orthogonal to each other), with associated **eigenvalues** $\lambda_1, \ldots, \lambda_n \geq 0$.

Note that eigenvalue λ_j is just the value of the function $r(\boldsymbol{x})$ – and also the value of $f(\boldsymbol{x})$, evaluated at $\boldsymbol{x} = \boldsymbol{v}_j$:

$$r(\boldsymbol{v}_j) = \lambda_j = \frac{\boldsymbol{v}_j^t M \boldsymbol{v}_j}{\boldsymbol{v}_j^t \boldsymbol{v}_j} = \boldsymbol{v}_j^t M \boldsymbol{v}_j = f(\boldsymbol{v}_j), \tag{9}$$

where the last equality follows from our assumption that the eigenvectors are ortho**normal**, i.e. they have have length 1.

In summary, the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of M are the stationary points and their corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ are the associated function values $r(\boldsymbol{v}_j) = f(\boldsymbol{v}_j)$, which means that among all stationary points, the **global maximum occurs at the eigenvector $\hat{\boldsymbol{v}}$ with largest eigenvalue $\lambda_{\max}$**.

Finally, we come back to our initial problem (1), i.e.

$$\|A\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|A\boldsymbol{x}\|_2. \tag{10}$$

Since $\lambda_{\max} = \hat{\boldsymbol{v}}^t M \hat{\boldsymbol{v}} = \|A\hat{\boldsymbol{v}}\|_2^2$, the solution is

$$\|A\|_2 = \sqrt{\lambda_{\max}}. \tag{11}$$