Consider two covariance matrices $\mathbf{A}_{n\times n}$ and $\mathbf{B}_{n\times n}$. We say that $\mathbf{A}$ is *bigger* than $\mathbf{B}$, often denoted by $\mathbf{A} \geq \mathbf{B}$ or $\mathbf{A} \succsim \mathbf{B}$, if $\mathbf{A} - \mathbf{B}$ is semi-positive definite. Why do we use the "definiteness" of a matrix to compare the size of two covariance matrices?

First, notice that a covariance matrix is not only symmetrical, but also semi-positive definite. Consider a random vector $\mathbf{x} = (x_1, ..., x_n)^\top$. The covariance matrix is defined by

$$\mathbf{K} := \mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^\top].$$

Given any constant vector $\mathbf{v}$ of length n, we have

$$\mathbf{v}^\top\mathbf{K}\mathbf{v} = \mathbb{E}[\mathbf{v}^\top(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{v}^\top(\mathbf{x} - \mathbb{E}[\mathbf{x}]))^\top] \geq 0$$

by the definition of $\mathbf{K}$. Therefore, the covariance matrix $\mathbf{K}$ is semi-positive definite. In fact, $\mathbf{v}^\top\mathbf{K}\mathbf{v}$ is zero iff $\mathbf{x}$ has no variance at all.

There is another intuitive way of interpreting the definiteness described above. Consider the same vector $\mathbf{v}$ and the random vector $\mathbf{x}$. The dot product $\mathbf{v}^\top\mathbf{x}$ is the projection of the random vector from n-dimensional space on a one-dimensional space along the direction of $\mathbf{v}$, i.e., this collapse the n-dimensional random variable to a one-dimensional random variable through some linear combination. If we calculate the variance of the one-dimensional random variable $\mathbf{v}^\top\mathbf{x}$, we obtain

$$\begin{aligned}\mathrm{Var}[\mathbf{v}^\top\mathbf{x}] &= \mathbb{E}[\mathbf{v}^\top\mathbf{x}(\mathbf{v}^\top\mathbf{x})^\top] - \mathbb{E}[\mathbf{v}^\top\mathbf{x}]\,\mathbb{E}[\mathbf{v}^\top\mathbf{x}]^\top \\ &= \mathbf{v}^\top\left(\mathbb{E}[\mathbf{x}\mathbf{x}^\top] - \mathbb{E}[\mathbf{x}]\,\mathbb{E}[\mathbf{x}]^\top\right)\mathbf{v} \\ &= \mathbf{v}^\top\mathbf{K}\mathbf{v}.\end{aligned}$$

Notice that the variance assumes the exact form as before. And since variance is non-negative, it is clear that the covariance matrix must be semi-positive definite. That is, for any direction $\mathbf{v}$, the variance of "$\mathbf{x}$ projected on that direction" is (clearly) non-negative.

Motivated by the intuitive interpretation, lets now compare two covariance matrices. Let $\mathbf{x} = (x_1, ..., x_n)^\top$ and $\mathbf{y} = (y_1, ..., y_n)^\top$ be random vectors with mean $(0, ..., 0)^\top$ for simplicity. Let $\mathbf{A} = \mathbb{E}[\mathbf{x}\mathbf{x}^\top]$ and $\mathbf{B} = \mathbb{E}[\mathbf{y}\mathbf{y}^\top]$ be the covariance matrices. Our goal is to compare $\mathbf{A}$ and $\mathbf{B}$ in some meaningful way. We can project $\mathbf{x}$ and $\mathbf{y}$ on a vector $\mathbf{v}$, and then compare the variance (non-negative real number) of the two projections. To make the comparison meaningful, it is reasonable to compare *all* possible projections, i.e., consider all possible choices of $\mathbf{v}$.

Formally, consider any vector $\mathbf{v}$. The projection of $\mathbf{x}$ on $\mathbf{v}$ is $\mathbf{v}^\top\mathbf{x}$. The variance of $\mathbf{v}^\top\mathbf{x}$ is

$$\begin{aligned}\mathbb{E}[(\mathbf{v}^\top\mathbf{x})^2] &= \mathbb{E}[\mathbf{v}^\top\mathbf{x}\mathbf{x}^\top\mathbf{v}] \\ &= \mathbf{v}^\top\,\mathbb{E}[\mathbf{x}\mathbf{x}^\top]\mathbf{v} = \mathbf{v}^\top\mathbf{A}\mathbf{v}\end{aligned}$$

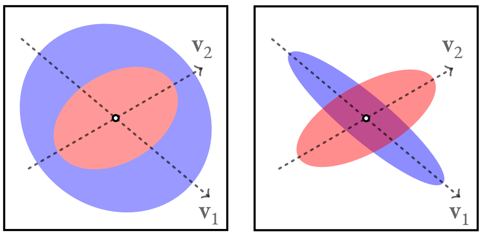
where $\mathbf{A}$ is the covariance matrix. Similarly, consider the same for $\mathbf{y}$. If we find that $\forall\mathbf{v}$,

$$\mathbf{v}^\top\mathbf{A}\mathbf{v} - \mathbf{v}^\top\mathbf{B}\mathbf{v} = \mathbf{v}^\top(\mathbf{A} - \mathbf{B})\mathbf{v} \geq 0,$$

then, by definition, $\mathbf{A} - \mathbf{B}$ is semi-positive definite. Now we know why we say $\mathbf{A}$ is *larger* than $\mathbf{B}$ when $\mathbf{A} - \mathbf{B}$ is positive definite:

> If $\mathbf{A} - \mathbf{B}$ is positive definite, then *for all possible directions* $\mathbf{v}$, the variance of $\mathbf{x}$ is larger than $\mathbf{y}$'s. [a]
>
> [a] This order of semi-positive definite matrices is called the Löwner ordering.

This interpretation of the partial ordering can be understood easily through visualisation. The following are representations of the distributions $\mathbf{x}$ and $\mathbf{y}$ where the two random vectors are two-dimensional:



case 1 case 2

Let $\mathbf{x}$ with covariance matrix $\mathbf{A}$ be the blue distribution and $\mathbf{y}$ with covariance matrix $\mathbf{B}$ be the red distribution. It is clear that in case 1, $\mathbf{A}$ is *bigger* than $\mathbf{B}$ since the variance of $\mathbf{x}$ is bigger that $\mathbf{y}$'s in *every* direction. (every possible direction of projection) However, the same statement is not true in case 2. In some directions (e.g. $\mathbf{v}_1$), the variance of $\mathbf{x}$ is larger; in other directions (e.g. $\mathbf{v}_2$), the variance of $\mathbf{y}$ is larger. Thus, $\mathbf{A}$ and $\mathbf{B}$ are not comparable by the partial order in case 2. #