Theory of Moments

Author: Han de Bruijn
Dated: 2008 June

One-dimensional Moments

Consider a collection X of arbitrary points x_k in one-dimensional space. The members of this *points cloud* can be thought as coordinate positions on a straight line:

$$X = \{x_1, x_2, x_3, ..., x_k, ..., x_{N-1}, x_N\}$$

A quantity called weight or mass m_k is associated with each of these points. The total weight or mass M of the points is given by the sum of the partial weights m_k :

$$M = \sum_{k=1}^{N} m_k = \sum_k m_k$$

It will be assumed in the sequel that the weights are always positive, meaning that they can be *normed*. Such normed weights w_k are defined by:

$$w_k = \frac{m_k}{M} \quad \Longrightarrow \quad 0 \le w_k \le 1 \quad \text{and} \quad \sum_k w_k = 1$$

It is remarked that the weights w_k can be interpreted as the components of a discrete probability distribution. Reason why we are tempted to conceive a certain spot, called center of mass, center of gravity, midpoint, middle or simply the *mean*. It is defined by:

$$\mu_x = \overline{x} = \sum_k w_k x_k$$

The midpoint takes a special position at the points cloud, since it is the weighted mean value of all positions of the points in the points cloud. It's easy to conceive a weighted mean value of other quantities, however. A most useful quantity is the so-called *second order moment*, which is also known as the *moment of inertia*, due to its applications in classical mechanics. Accordingly, the midpoint is also called a *first order moment*. The second order moment may also be called (the square of the) standard deviation or *spread*, due to the quite analogous quantity in Probability Theory:

$$\overline{x^2} = \sum_k w_k x_k^2$$

In addition to the above discrete quantities, there also exist *continuum versions* of the first and second order moments. The only difference is that the latter are defined by (definite) integrals instead of sums:

$$M = \int_a^b m(x)\,dx = \int m(x)\,dx \quad \text{and} \quad w(x) = \frac{m(x)}{M} \quad \Longrightarrow \quad \int w(x)\,dx = 1$$

$$\overline{x} = \int w(x)\, x\, dx \quad \text{and} \quad \overline{x^2} = \int w(x)\, x^2\, dx$$

It is clear from the outset, however, that such integrals are just limiting cases of discrete sums. Hence subsequent results will also be valid for the continuous version of the theory.
Second order moments may be defined with respect to a fixed, but otherwise arbitrary point p in (1-D) space:

$$\sigma_{xx}(p) = \sum_k w_k (x_k - p)^2 \quad \text{or} \quad \sigma_{xx}(p) = \int w(x)\,(x-p)^2\, dx$$

The moment of inertia is interpreted as a mean of the squared distances of the points in the cloud with respect to a fixed point p. It will be shown now that there exists a preferrable origin, which is precisely the midpoint of the points distribution.

$$\sigma_{xx}(p) = \sum_k w_k (x_k - p)^2 = \sum_k w_k x_k^2 - 2p \sum_k w_k x_k + p^2 =$$

$$\overline{x^2} - 2p\overline{x} + p^2 = \left[\overline{x^2} - \overline{x}^2\right] + \left[\overline{x}^2 - 2p\overline{x} + p^2\right]$$

The first term between square brackets [] can be worked out as follows:

$$\left[\sum_k w_k x_k^2 - \left(\sum_k w_k x_k\right)^2\right] =$$

$$\sum_k w_k x_k^2 - 2 \sum_k w_k \left(\sum_L w_L x_L\right) x_k + \sum_k w_k \left(\sum_L w_L x_L\right)^2 =$$

$$\sum_k w_k \left[x_k^2 - 2\left(\sum_L w_L x_L\right) x_k + \left(\sum_L w_L x_L\right)^2\right] =$$

$$\sum_k w_k \left[x_k - \left(\sum_L w_L x_L\right)\right]^2 = \sum_k w_k \left(x_k - \overline{x}\right)^2$$

And the second term between square brackets [] is:

$$\left[\overline{x}^2 - 2p\overline{x} + p^2\right] = \left(\overline{x} - p\right)^2$$

Conclusion:

$$\sum_k w_k (x_k - p)^2 = \sum_k w_k \left(x_k - \overline{x}\right)^2 + \left(\overline{x} - p\right)^2$$

Then we see that the first term is positive, because it is a sum of (weighted) squares. But also the second term is a square and hence positive. The latter assumes a minimum if it is exactly zero, that is if: $p = \overline{x}$. Formally:

$$\sum_k w_k (x_k - p)^2 = \text{minimum}(p) \quad \Longleftrightarrow \quad p = \overline{x} = \sum_k w_k x_k$$

The physical interpretation of the above is that a moment of inertia assumes a minimal value with respect to the origin if that origin is coincident with the center of mass. A moment of inertia with respect to an origin which is different from the center of mass can be expressed as the sum of two moments: one which expresses the moment of inertia with respect to the midpoint plus one which expresses the moment of inertia of the midpoint with respect to the origin. Unless explicitly stated otherwise, it will be assumed in the sequel that all moments of inertia are defined with respect to the midpoint μ_x or all (squares of the) spreads with respect to the mean. Then we can drop the dependence on (p) in:

$$\sigma_{xx} = \sum_k w_k (x_k - \mu_x)^2$$

Two-dimensional Moments

Consider an arbitrary 2-D distribution of points (x_k, y_k) in the plane. Aagain, a quantity called weight or mass w_k is associated with each of these points. And again, we can define a spot, called the midpoint, center of mass or whatever name is to be preferred:

$$\sum_k w_k = 1$$

$$\mu_x = \overline{x} = \sum_k w_k x_k \quad \text{and} \quad \mu_y = \overline{y} = \sum_k w_k y_k$$

This is the discrete form. The continuous alternative is:

$$\iint w(x,y)\,dx\,dy = 1$$

$$\mu_x = \overline{x} = \iint w(x,y) x\,dx\,dy \quad \text{and} \quad \mu_y = \overline{y} = \iint w(x,y) y\,dx\,dy$$

Second order momenta, also called moments of inertia, are defined with respect to an origin (p, q):

$$\sigma_{xx}(p) = \sum_k w_k (x_k - p)^2 \quad \text{and} \quad \sigma_{yy}(q) = \sum_k w_k (y_k - q)^2$$

$$\sigma_{xy}(p,q) = \sum_k w_k (x_k - p)(y_k - q)$$

The continuous form is:

$$\sigma_{xx}(p) = \iint w(x,y)(x-p)^2\,dx\,dy \quad \text{and} \quad \sigma_{yy}(q) = \iint w(x,y)(y-q)^2\,dx\,dy$$

$$\sigma_{xy}(p,q) = \iint w(x,y)(x-p)(y-q)\,dx\,dy$$

It has already been shown that, at least for σ_{xx} ,there exists a preferrable origin, which is precisely the center of mass / geometric mean of the points distribution:

$$\sum_k w_k(x_k - p)^2 = \text{minimum}(p) \quad \Longleftrightarrow \quad p = \overline{x} = \sum_k w_k x_k$$

In very much the same way (method: what's in a name) we can prove for σ_{yy} :

$$\sum_k w_k(y_k - q)^2 = \text{minimum}(q) \quad \Longleftrightarrow \quad q = \overline{y} = \sum_k w_k y_k$$

How about the "mixed" second order moment σ_{xy} ?

$$\sigma_{xy}(\overline{x},\overline{y}) = \sum_k w_k(x_k - \overline{x})(y_k - \overline{y}) = \sum_k w_k x_k y_k - \sum_k w_k x_k \overline{y} - \sum_k w_k y_k \overline{x} + \overline{x}\,\overline{y} =$$

$$\sum_k x_k y_k - \overline{x}\,\overline{y} - \overline{y}\,\overline{x} + \overline{x}\,\overline{y} \quad \Longrightarrow \quad \sigma_{xy} = \overline{xy} - \overline{x}\,\overline{y}$$

Again, unless explicitly stated otherwise, it will be assumed in the sequel that all moments of inertia are with respect to the midpoint (μ_x, μ_y). Then we can drop (p, q) in:

$$\sigma_{xx} = \sum_k w_k(x_k - \mu_x)^2 \quad \text{and} \quad \sigma_{yy} = \sum_k w_k(y_k - \mu_y)^2$$

$$\sigma_{xy} = \sum_k w_k(x_k - \mu_x)(y_k - \mu_y)$$

So far, it is less clear what kind of physical meaning should be attached to the quantity σ_{xy} , which is known as a "cross correlation" in probability theory and statistics. Well, to be precise:

$$\rho = \frac{\sigma_{xy}}{\sqrt{\sigma_{xx}\sigma_{yy}}}$$

Where ρ is the so-called *cross-correlation coefficient*. Suppose however, that we don't like σ_{xy} at all and we only want to get rid of this term. How then could such a thing be accomplished ? It can certainly not be done by translation, since the origin of our coordinate system has already become fixed at the midpoint. But there is another possibility. It could be done by *rotating* the coordinate system in such a way that σ'_{xy} becomes zero in the new ('primed') system. Let's give it a try. Start with:

$$\begin{cases} x' = \cos(\theta)x + \sin(\theta)y \\ y' = -\sin(\theta)x + \cos(\theta)y \end{cases}$$

Then:

$$\sigma'_{xy} = 0 \quad \Longleftrightarrow \quad \sum_k w_k x'_k y'_k =$$

$$\sum_k w_k \left[\cos(\theta)x_k + \sin(\theta)y_k\right]\left[-\sin(\theta)x_k + \cos(\theta)y_k\right] =$$

$$-\cos(\theta)\sin(\theta)\sum_k w_k x_k^2 + \sin(\theta)\cos(\theta)\sum_k w_k y_k^2$$

$$+\left[\cos^2(\theta) - \sin^2(\theta)\right]\sum_k w_k x_k y_k$$

Two goniometric formulas should me memorized here:

$$\sin(2\theta) = 2\sin(\theta)\cos(\theta) \quad \text{and} \quad \cos(2\theta) = \cos^2(\theta) - \sin^2(\theta)$$

Herewith:

$$\sum_k w_k x'_k y'_k = -\frac{1}{2}\sin(2\theta)(\sigma_{xx} - \sigma_{yy}) + \cos(2\theta)\sigma_{xy} = 0$$

Resulting in:

$$\tan(2\theta) = \frac{2\sigma_{xy}}{\sigma_{xx} - \sigma_{yy}} \qquad \text{for} \quad \sigma_{xx} \neq \sigma_{yy}$$

And as special cases:

$$\sin(2\theta) = 0 \quad \Longrightarrow \quad \theta = k.\frac{\pi}{2} \quad k = 0, 1, 2, 3\ldots$$

$$\text{for} \quad \sigma_{xx} \neq \sigma_{yy} \quad \text{and} \quad \sigma_{xy} = 0$$

Meaning that the situation where $\sigma_{xy} = 0$ is found back with every rotation of the coordinate system over 90 degrees.

$$\cos(2\theta) = 0 \quad \Longrightarrow \quad \theta = \frac{\pi}{4} + k.\frac{\pi}{2} \quad k = 0, 1, 2, 3\ldots$$

$$\text{for} \quad \sigma_{xx} = \sigma_{yy} \quad \text{and} \quad \sigma_{xy} \neq 0$$

Meaning that the situation where $\sigma_{xx} = \sigma_{yy}$ occurs every time θ is precisely in the middle between two angles where $\sigma_{xy} = 0$.
If both $\sigma_{xx} = \sigma_{yy}$ and $\sigma_{xy} = 0$ then the choice of the angle θ is arbitrary.
The other two moments of inertia, σ'_{xx} and σ'_{yy}, are expressed into the angle θ as follows:

$$\sigma'_{xx} = \sum_k w_k (x'_k)^2 = \sum_k w_k \left[\cos(\theta)x_k + \sin(\theta)y_k\right]^2$$

$$= \cos^2(\theta)\sum_k w_k x_k^2 + \sin^2(\theta)\sum_k w_k y_k^2 + 2\sin(\theta)\cos(\theta)\sum_k w_k x_k y_k$$

$$\Longrightarrow \quad \sigma'_{xx} = \cos^2(\theta)\sigma_{xx} + \sin^2(\theta)\sigma_{yy} + 2\sin(\theta)\cos(\theta)\sigma_{xy}$$

And:

$$\sigma'_{yy} = \sum_k w_k (y'_k)^2 = \sum_k w_k \left[-\sin(\theta)x_k + \cos(\theta)y_k\right]^2$$

$$= \sin^2(\theta)\sum_k w_k x_k^2 + \cos^2(\theta)\sum_k w_k y_k^2 - 2\sin(\theta)\cos(\theta)\sum_k w_k x_k y_k$$

$$\Longrightarrow \quad \sigma'_{yy} = \sin^2(\theta)\sigma_{xx} + \cos^2(\theta)\sigma_{yy} - 2\sin(\theta)\cos(\theta)\sigma_{xy}$$

Working out the latter formula somewhat further:

$$\sigma'_{yy} = \left[1 - \cos^2(\theta)\right]\sigma_{xx} + \left[1 - \sin^2(\theta)\right]\sigma_{yy} - 2\sin(\theta)\cos(\theta)\sigma_{xy}$$

$$= \sigma_{xx} + \sigma_{yy} - \sigma'_{xx}$$

It is thus seen that the sum of the two "main" moments of inertia is entirely *invariant* for an orthogonal coordinate transformation:

$$\sigma'_{xx} + \sigma'_{yy} = \sigma_{xx} + \sigma_{yy}$$

We conclude that, indeed, the "unwanted" σ_{xy} can be eliminated by a suitable rotation of the coordinate system, while the sum of the other "main" moments of inertia $(\sigma_{xx} + \sigma_{yy})$ remains independent of any such transformation. The question may be raised for what values of θ the transformed moments of inertia σ'_{xx} and/or σ'_{yy} attain an extreme value, a maximum or a minimum. In order to find out, derivatives to θ will be calculated. First we repeat:

$$\sigma'_{xx} = \cos^2(\theta)\sigma_{xx} + \sin^2(\theta)\sigma_{yy} + 2\sin(\theta)\cos(\theta)\sigma_{xy}$$

$$\sigma'_{yy} = \sin^2(\theta)\sigma_{xx} + \cos^2(\theta)\sigma_{yy} - 2\sin(\theta)\cos(\theta)\sigma_{xy}$$

Giving:

$$\frac{d}{d\theta}\sigma'_{xx} = -2\sin(\theta)\cos(\theta)\sigma_{xx} + 2\cos(\theta)\sin(\theta)\sigma_{yy} + 2\cos^2(\theta)\sigma_{xy} - 2\sin^2(\theta)\sigma_{xy}$$

$$= -\sin(2\theta)(\sigma_{xx} - \sigma_{yy}) + \cos(2\theta)2\sigma_{xy} = 0$$

$$\frac{d}{d\theta}\sigma'_{yy} = +2\cos(\theta)\sin(\theta)\sigma_{xx} - 2\sin(\theta)\cos(\theta)\sigma_{yy} - 2\cos^2(\theta)\sigma_{xy} + 2\sin^2(\theta)\sigma_{xy}$$

$$= +\sin(2\theta)(\sigma_{xx} - \sigma_{yy}) - \cos(2\theta)2\sigma_{xy} = 0$$

It is seen that exactly the same equations are obtained as with $\sigma'_{xy} = 0$. Meaning that σ'_{xx} and σ'_{yy} attain their extreme values, both at the same time, when and only when $\sigma'_{xy} = 0$. Opposite signs indicate that one of the two extremes, σ'_{xx} or σ'_{yy}, must be a minimum while the other must be a maximum.
Alternative viewpoints may be obtained by just reversing the whole story. Take the 'primed' coordinate system for granted. And transform back to 'unprimed' coordinates. In order to accomplish this, it is not necessary to solve the above equations for the unprimed moments of inertia. Instead, simply reverse the angle of rotation and you're done:

$$\sigma_{xx} = \cos^2(-\theta)\sigma'_{xx} + \sin^2(-\theta)\sigma'_{yy} + 2\sin(-\theta)\cos(-\theta)\sigma'_{xy}$$

$$\sigma_{yy} = \sin^2(-\theta)\sigma'_{xx} + \cos^2(-\theta)\sigma'_{yy} - 2\sin(-\theta)\cos(-\theta)\sigma'_{xy}$$

$$\sigma_{xy} = -\frac{1}{2}\sin(-2\theta)(\sigma'_{xx} - \sigma'_{yy}) + \cos(-2\theta)\sigma'_{xy}$$

Remember, however, that the cross correllation moment is zero by definition in the primed system. At last, consider the primed system as the standard one. Herewith it is expressed that the coordinate system where the cross correllation moment is zero is to be considered in the sequel as the *preferred* system of coordinates. The x- and y-axis, associated with the preferred system, are known in Physics as the *main axes of inertia*. They are attached to the points cloud, as body fitted coordinates so to speak. Any other system is now the result of a rotation of the main axes of inertia, over a certain angle θ . And the transformed moments of inertia can always be derived from the main moments of inertia:

$$\sigma_{xx} = \cos^2(\theta)\sigma'_{xx} + \sin^2(\theta)\sigma'_{yy} \quad \text{and} \quad \sigma_{yy} = \sin^2(\theta)\sigma'_{xx} + \cos^2(\theta)\sigma'_{yy}$$

$$\sigma_{xy} = \frac{1}{2}\sin(2\theta)(\sigma'_{xx} - \sigma'_{yy})$$

It is seen that $\sigma_{xy} = 0$, for $\theta = k.\pi/2 \quad k = 1, 2, ...$. The same angle values cause σ_{xx} to become equal to σ'_{xx}, for $\theta = k.\pi$, or equal to σ'_{yy}, for $\theta = \pi/2 + k.\pi$. And the same angle values cause σ_{yy} to become equal to σ'_{yy}, for $\theta = k.\pi$, or equal to σ'_{xx}, for $\theta = \pi/2 + k.\pi$. It all means that σ'_{xx} and σ'_{yy} exchange roles with every increase of the angle θ with 90^o , while $\sigma_{xy} = 0$ at the same time.
On the other hand, σ_{xx} and σ_{yy} become equal for $\theta = \pi/4 + k.\pi/2$. Then the cross correllation moment σ_{xy} attains a minimum (negative) or maximum (positive) value of $\pm(\sigma'_{xx} - \sigma'_{yy})/4$.
If $\sigma'_{xx} = \sigma'_{yy}$, that is when the main moments of inertia are equal to each other, then also the transformed main moments of inertia always will be equal to each other and the cross correllation moment σ_{xy} will always be zero. This special case is often induced by symmetry.
Last but not least. The fact that the above formulas are entirely insensitive to an increase of the angle θ with 180^o also means that it is impossible to detect the *orientation* of the main axes coordinate system, with help of moments up to the second order alone. However, provided that one has to deal with the common non-symmetric case, only *two possibilities* will remain, and they can differ only by a rotation over 180 degrees.

Schwarz inequality

Inclining towards Linear Algebra, but with the theory of point clouds in mind, we will redefine the inner product of two vectors as:

$$(\vec{a} \cdot \vec{b}) := \sum_k w_k a_k b_k$$

It can be shown easily that such an inner product obeys all of the usual rules:

$$(\vec{a} \cdot \vec{b}) = (\vec{b} \cdot \vec{a})$$
$$(\vec{a} \cdot (\vec{b} + \vec{c})) = (\vec{a} \cdot \vec{b}) + (\vec{a} \cdot \vec{c})$$

$$(\vec{a} \cdot \lambda\vec{b}) = \lambda(\vec{a} \cdot \vec{b})$$
$$(\vec{a} \cdot \vec{a}) \geq 0$$

For the last rule to be obeyed, it is necessary that masses w_k be positive (or zero). Schwartz inequality can now be conjectured for this inner product:

$$(\vec{a} \cdot \vec{b})^2 \leq (\vec{a} \cdot \vec{a})(\vec{b} \cdot \vec{b})$$

Proof:

$$(\lambda\vec{a} - \vec{b} \cdot \lambda\vec{a} - \vec{b})^2 \geq 0 \quad \Longrightarrow$$
$$\lambda^2(\vec{a} \cdot \vec{a}) - 2\lambda(\vec{a} \cdot \vec{b}) + (\vec{b} \cdot \vec{b}) \geq 0$$

This is a quadratic inequality in λ. In order for this inequality to hold, its discriminant must be negative or zero:

$$4(\vec{a} \cdot \vec{b})^2 - 4(\vec{a} \cdot \vec{a})(\vec{b} \cdot \vec{b}) \leq 0 \quad \Longrightarrow$$
$$(\vec{a} \cdot \vec{b})^2 \leq (\vec{a} \cdot \vec{a})(\vec{b} \cdot \vec{b})$$

Which completes the proof. Returning now to our original problem, define for example:

$$\vec{a} = (x_1, x_2, ..., x_k, ..., x_N) \quad \text{and} \quad \vec{b} = (1, 1, ..., 1, ..., 1)$$

Then:

$$(\vec{a} \cdot \vec{b}) = \sum_k w_k x_k \quad \text{and} \quad (\vec{a} \cdot \vec{a}) = \sum_k w_k x_k^2 \quad \text{and} \quad (\vec{b} \cdot \vec{b}) = \sum_k w_k = 1$$

Therefore:

$$\left(\sum_k w_k x_k\right)^2 \leq \left(\sum_k w_k x_k^2\right) \quad \Longrightarrow \quad \sum_k w_k x_k^2 - \left(\sum_k w_k x_k\right)^2 \geq 0$$

This result is equivalent with our previous finding that:

$$\sum_k w_k (x_k - \overline{x})^2 \geq 0$$

New results are obtained when applying Schwarz inequality to the quantity known formerly as cross correllation moment. For the sake of simplicity, the midpoint is set as the origin $(0, 0)$ of the coordinate system. Now define:

$$\vec{a} = (x_1, x_2, ..., x_k, ..., x_N) \quad \text{and} \quad \vec{b} = (y_1, y_2, ..., y_k, ..., y_N)$$

Then:

$$(\vec{a} \cdot \vec{b}) = \sum_k w_k x_k y_k = \sigma_{xy}$$

$$(\vec{a} \cdot \vec{a}) = \sum_k w_k x_k^2 = \sigma_{xx} \quad \text{and} \quad (\vec{b} \cdot \vec{b}) = \sum_k w_k y_k^2 = \sigma_{yy}$$

Therefore, according to Schwarz inequality, the following relationship must hold:

$$\sigma_{xy}^2 \le \sigma_{xx}\sigma_{yy} \quad \text{or} \quad \sigma_{xx}\sigma_{yy} - \sigma_{xy}^2 \ge 0$$

As a consequence, the following quantity is bound like a (co)sinus-function:

$$\rho := \frac{\sigma_{xy}}{\sqrt{\sigma_{xx}\sigma_{yy}}} \qquad \text{where} \qquad 0 \le |\rho| \le 1$$

The thus defined quantity ρ has already been mentioned as the cross-correlation, in a narrower sense.

2-D Moment Invariants

We start with repeating some results from a preceding section:

$$\sigma'_{xx} = \cos^2(\theta)\sigma_{xx} + \sin^2(\theta)\sigma_{yy} + 2\sin(\theta)\cos(\theta)\sigma_{xy}$$

$$\sigma'_{yy} = \sin^2(\theta)\sigma_{xx} + \cos^2(\theta)\sigma_{yy} - 2\sin(\theta)\cos(\theta)\sigma_{xy}$$

Combining these formulas by addition results in:

$$\sigma'_{xx} + \sigma'_{yy} = \sigma_{xx} + \sigma_{yy}$$

It is thus seen that the sum of the two main moments of inertia is *invariant* for an orthogonal coordinate transformation. Substracting the formulas for σ'_{xx} and σ'_{yy} results in:

$$\sigma'_{xx} - \sigma'_{yy} = \left[\cos^2(\theta) - \sin^2(\theta)\right](\sigma_{xx} - \sigma_{yy}) + 2\sin(\theta)\cos(\theta)2\sigma_{xy}$$

$$\sigma'_{xx} - \sigma'_{yy} = \cos(2\theta)(\sigma_{xx} - \sigma_{yy}) + \sin(2\theta)2\sigma_{xy}$$

On the other hand we found:

$$2\sigma'_{xy} = -\sin(2\theta)(\sigma_{xx} - \sigma_{yy}) + \cos(2\theta)2\sigma_{xy} = 0$$

In matrix form:

$$\begin{bmatrix} \sigma'_{xx} - \sigma'_{yy} \\ 2\sigma'_{xy} \end{bmatrix} = \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) \end{bmatrix} \begin{bmatrix} \sigma_{xx} - \sigma_{yy} \\ 2\sigma_{xy} \end{bmatrix}$$

Thus a *rotation* of the vector $(\sigma_{xx} - \sigma_{yy}, 2\sigma_{xy})$ over an angle 2θ results in another vector of the same kind. It also means that the vector and its transform have the same length:

$$(\sigma'_{xx} - \sigma'_{yy})^2 + (2\sigma'_{xy})^2 = (\sigma_{xx} - \sigma_{yy})^2 + (2\sigma_{xy})^2$$

$$\left[\sigma'_{xx} + \sigma'_{yy}\right]^2 - 4\left[\sigma'_{xx}\sigma'_{xx} - (\sigma'_{xy})^2\right] = \left[\sigma_{xx} + \sigma_{yy}\right]^2 - 4\left[\sigma_{xx}\sigma_{xx} - (\sigma_{xy})^2\right]$$

But we know that $\sigma'_{xx} + \sigma'_{yy} = \sigma_{xx} + \sigma_{yy}$ is an invariant. Therefore the following quantity must be an invariant as well:

$$\sigma'_{xx}\sigma'_{xx} - (\sigma'_{xy})^2 = \sigma_{xx}\sigma_{xx} - \sigma_{xy}^2 \ge 0$$

Idea and Material

Suppose we have a (handwritten) cipher "3" at our disposal, which is the kind of rough *material* we have to live with, most of the time. Our purpose is to recognize the *material* neatly as a "3", by a digital computer.

But in order to establish what a neat "3" is, we also must have an *idea* what a cipher "3" should look like. It will be assumed that such an *idea* is always much simpler than the material. The idea of a cipher "3" could, for example, be defined like this:

```
 XXX
X   X
    X
  XX
    X
    X
X   X
 XXX
```

Question: how can we "see" that material of a "3" is the same as the idea of a "3" ? Answer: by matching somehow the idea with the material. It is clear that the material cannot be changed; it must be taken as it is. Ideas are far more flexible. We can decorate our idea "3" with a couple of parameters, such that it can be skewed, scaled, rotated and translated in (2-D) space. Let (x_I, y_I) be the coordinates of our idea, then the transformed idea may read as follows:

$$
\begin{aligned}
x'_I &= a.x_I + b.y_I + g \\
y'_I &= p.x_I + q.y_I + h
\end{aligned}
$$

Wee seek a correspondence with the theory of moments. As a good initialization, we can define the local origins of both the idea and the material in such a way that they coincide with the respective midpoints:

$$(\sum_I x_I, \sum_I y_I) = (0,0)_I \quad \text{and} \quad (\sum_M x_M, \sum_M y_M) = (0,0)_M$$

It will be decided that the first order moments of both the transformed idea and the material should be the same. For the idea, this means the following:

$$\sum_I x'_I \quad = \quad a.\sum_I x_I + b.\sum_I y_I + g\sum_I 1 = g\sum_I 1$$

$$\sum_I y'_I \quad = \quad p.\sum_I x_I + q.\sum_I y_I + h\sum_I 1 = h\sum_I 1$$

It follows that the translation vector (g, h) is equal to the midpoint of the transformed idea:

$$g = \sum_I x'_I / \sum_I 1 \quad \text{and} \quad h = \sum_I y'_I / \sum_I 1$$

Which in turn is equal to the midpoint of the material:

$$g = \sum_M x_M / \sum_M 1 \quad \text{and} \quad h = \sum_M y_M / \sum_M 1$$

Use upper indices instead of lower where appropriate and calculate second order moments with respect to the midpoint.

$$(\sigma'_{xx})_I = \sum_I (x'_I - g)^2 = \sum_I (a.x_I + b.y_I)^2 = a^2\sigma^I_{xx} + 2ab\ \sigma^I_{xy} + b^2\sigma^I_{yy}$$

$$(\sigma'_{yy})_I = \sum_I (y'_I - h)^2 = \sum_I (p.x_I + q.y_I)^2 = p^2\sigma^I_{xx} + 2pq\ \sigma^I_{xy} + q^2\sigma^I_{yy}$$

$$(\sigma'_{xy})_I = \sum_I (x'_I - g)(y'_I - h) = \sum_I (a.x_I + b.y_I)(p.x_I + q.y_I) =$$

$$ap\ \sigma^I_{xx} + (aq + bp)\sigma^I_{xy} + bq\ \sigma^I_{yy}$$

In matrix form:

$$\begin{bmatrix} (\sigma'_{xx})_I & (\sigma'_{xy})_I \\ (\sigma'_{xy})_I & (\sigma'_{yy})_I \end{bmatrix} = \begin{bmatrix} a & b \\ p & q \end{bmatrix} \begin{bmatrix} \sigma^I_{xx} & \sigma^I_{xy} \\ \sigma^I_{xy} & \sigma^I_{yy} \end{bmatrix} = \begin{bmatrix} a & p \\ b & q \end{bmatrix}$$

Indeed:

$$= \begin{bmatrix} a\sigma^I_{xx} + b\sigma^I_{xy} & a\sigma^I_{xy} + b\sigma^I_{yy} \\ p\sigma^I_{xx} + q\sigma^I_{xy} & p\sigma^I_{xy} + q\sigma^I_{yy} \end{bmatrix} \begin{bmatrix} a & p \\ b & q \end{bmatrix} =$$

$$\begin{bmatrix} a^2\sigma^I_{xx} + ba\sigma^I_{xy} + ab\sigma^I_{xy} + b^2\sigma^I_{yy} & ap\sigma^I_{xx} + bp\sigma^I_{xy} + aq\sigma^I_{xy} + bq\sigma^I_{yy} \\ pa\sigma^I_{xx} + qa\sigma^I_{xy} + pb\sigma^I_{xy} + qb\sigma^I_{yy} & p^2\sigma^I_{xx} + qp\sigma^I_{xy} + pq\sigma^I_{xy} + q^2\sigma^I_{yy} \end{bmatrix}$$

The determinant of a product of matrices is the product of the determinants:

$$\begin{vmatrix} (\sigma'_{xx})_I & (\sigma'_{xy})_I \\ (\sigma'_{xy})_I & (\sigma'_{yy})_I \end{vmatrix} = \begin{vmatrix} a & b \\ p & q \end{vmatrix} \begin{vmatrix} \sigma^I_{xx} & \sigma^I_{xy} \\ \sigma^I_{xy} & \sigma^I_{yy} \end{vmatrix} = \begin{vmatrix} a & b \\ p & q \end{vmatrix}^T$$

Hence:

$$\left[\sigma^I_{xx}\sigma^I_{yy} - (\sigma^I_{xy})^2\right]' = \left[\sigma^I_{xx}\sigma^I_{yy} - (\sigma^I_{xy})^2\right](aq - bp)^2$$

But suppose that the second order moments of the transformed idea are made equal to the second order moments of the material, then:

$$(aq - bp)^2 = \frac{\left[\sigma^M_{xx}\sigma^M_{yy} - (\sigma^M_{xy})^2\right]}{\left[\sigma^I_{xx}\sigma^I_{yy} - (\sigma^I_{xy})^2\right]}$$

In order to be able to proceed, we choose a special form of the generic affine transformation, namely a scaling in the x-direction and in the y-direction (with $S_x > 0$ and $S_y > 0$) followed by a rotation over an angle θ :

$$\begin{bmatrix} a & b \\ p & q \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix}$$

Herewith we find:

$$\left[\sigma^I_{xx} + \sigma^I_{yy}\right]' = S_x^2 \sigma^I_{xx} + S_y^2 \sigma^I_{yy} \quad \Longrightarrow \quad S_x^2 \sigma^I_{xx} + S_y^2 \sigma^I_{yy} = \sigma^M_{xx} + \sigma^M_{yy}$$

Because the trace is an invariant for rotations. On the other hand, the above-mentioned expression for the determinant is further simplified to:

$$(S_x S_y)^2 = \frac{\left[\sigma^M_{xx}\sigma^M_{yy} - (\sigma^M_{xy})^2\right]}{\left[\sigma^I_{xx}\sigma^I_{yy} - (\sigma^I_{xy})^2\right]}$$

Resulting in two equations with two unknowns:

$$\lambda_x = S_x^2 \sigma^I_{xx} \quad \text{and} \quad \lambda_y = S_y^2 \sigma^I_{yy}$$

Namely:

$$\lambda_x \lambda_y = \text{Rdet} \quad \text{and} \quad \lambda_x + \lambda_y = \text{Sp}$$

Where:

$$\text{Rdet} = \sigma^I_{xx}\sigma^I_{yy} \frac{\left[\sigma^M_{xx}\sigma^M_{yy} - (\sigma^M_{xy})^2\right]}{\left[\sigma^I_{xx}\sigma^I_{yy} - (\sigma^I_{xy})^2\right]} \approx \sigma^M_{xx}\sigma^M_{yy} - (\sigma^M_{xy})^2 \quad \text{and} \quad \text{Sp} = \sigma^M_{xx} + \sigma^M_{yy}$$

Where it is tacitly assumed that the ideas are already at their main axes. The result is a quadratic equation:

$$\lambda_x \lambda_y = \lambda_x(\text{Sp} - \lambda_x) = \text{Rdet} \quad \Longrightarrow \quad \lambda_x^2 - \text{Sp}\,\lambda_x + \text{Rdet} = 0$$

Where the roots must be positive. So we can solve for $\lambda_{x,y}$ and calculate $S_{x,y}$ herefrom. It is noted, however, that the solution method is insensitive for the magnitudes of S_x and S_y. This means that they can, and will, be the other way around. So we must try the ordered pair (S_x, S_y) first. Then switch them into the ordered pair (S_y, S_x). And find out which of the matchings is the best.
At last, we should devise a means to calculate the angle of rotation θ. The following result is quite appropriate for our purpose:

$$\begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) \end{bmatrix} \begin{bmatrix} \sigma^I_{xx} - \sigma^I_{yy} \\ 2\sigma^I_{xy} \end{bmatrix} = \begin{bmatrix} (\sigma^I_{xx})' - (\sigma^I_{yy})' \\ 2(\sigma^I_{xy})' \end{bmatrix} = \begin{bmatrix} \sigma^M_{xx} - \sigma^M_{yy} \\ 2\sigma^M_{xy} \end{bmatrix}$$

It says that the angle between $(\sigma_{xx} - \sigma_{yy}, 2\sigma_{xy})$ of the idea, and the corresponding vector of the material, is twice the angle of rotation we are looking for. But mind the ambiguity, again. The angle between two vectors can only be determined apart from a rotation over 180^o. Therefore the angle θ can only be determined apart from a rotation over 90^o. So we must try four angles $\theta + k.\pi/2$, $k = 0, 1, 2, 3$ and find out which of the matchings is the best. Well, not entirely. We cannot even distinguish between an angle θ and its reverse $-\theta$, because the roots can be written as follows:

$$\lambda_{x,y} = \frac{\sigma_{xx}^M + \sigma_{yy}^M}{2} \pm \frac{\sigma_{xx}^M - \sigma_{yy}^M}{2} \sqrt{1 + \tan^2(2\theta)}$$

Where:

$$\tan(2\theta) = \frac{2\sigma_{xy}^M}{\sigma_{xx}^M - \sigma_{yy}^M}$$

So the roots are insensitive to the sign of $\tan(2\theta)$ and thus insensitive to the sign of θ itself. Consequently, we must investigate eight cases for the angles and two for the scalings, sixteen cases in total, when we're stretching and matching with moments.

Disclaimers

Anything free comes without referee :-(
My English may be better than your Dutch :-)