Some notes on ellipses

Ivo Ihrke

February 17, 2011

1 Ellipse decomposition

The problem of recovering midpoint, major and minor axis and radii from known parameters A, B, C, D, E, F of the implicit equation

$$f(x, y) = Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0 \tag{1}$$

of a conic in two dimensions is considered. The terms are valid for ellipses, but it will turn out, that the values are meaningful for hyperbolas as well. Formula 1 describes a general two-dimensional conic which is

- an ellipse for $B^2 - AC < 0$
- a parabola for $B^2 - AC = 0$
- a hyperbola for $B^2 - AC > 0$.

$\mathbf{p} = (x, y)$ is a point on the conic iff it satisfies equation 1. A general bi-quadratic function has an extremum or a saddle-point, the position of which can be determined by setting the gradient to zero $\nabla f = 0$.

$$\frac{\partial f}{\partial x} = 2Ax + 2By + 2D = 0 \tag{2}$$

$$\frac{\partial f}{\partial y} = 2Bx + 2Cy + 2E = 0. \tag{3}$$

Therefore the midpoint of the conic can be computed by solving the linear system

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} \begin{pmatrix} x_m \\ y_m \end{pmatrix} = -\begin{pmatrix} D \\ E \end{pmatrix}. \tag{4}$$

This point will be denoted $\mathbf{m} = (x_m, y_m)$. Using homogenous coordinates and matrix notation, Equation 1 can be written as

$$(x, y, 1)\, \mathbf{C}^{-1} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, = 0 \tag{5}$$

where matrix $\mathbf{C}^{-1}$ encodes the coefficients of the quadratic form. The entries of matrix $\mathbf{C}^{-1} \in R^{3\times 3}$ can be associated with the parameters of equation 1 as follows

$$\begin{aligned} (x,y,1)\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} &= 0 \qquad (6)\\ ax^2+dxy+gx+bxy+ey^2+hy+cx+fy+i &= \qquad (7)\\ ax^2+(d+b)xy+ey^2+(g+c)x+(h+f)y+i &= 0 \qquad (8) \end{aligned}$$

Associating the parameters A,B,C,D,E,F with equation 8 the matrix $\mathbf{C}^{-1}$ can be written as

$$\mathbf{C}^{-1} = \begin{pmatrix} A & B & D \\ B & C & E \\ D & E & F \end{pmatrix}. \qquad (9)$$

The next observation is that a general ellipse can be obtained by rotating, translating and scaling a unit circle $\mathbf{C_1}^{-1} = \mathrm{diag}(1,1,-1)$.

$$\begin{aligned} \hat{x}^T\mathbf{C_1}^{-1}\hat{x} &= \qquad (10)\\ x^T\mathbf{C}^{-1}x &= 0 \qquad (11) \end{aligned}$$

with $\hat{x} = (\mathbf{TRS})^{-1}x$, where

$$\begin{aligned} \mathbf{T} &= \begin{pmatrix} 1 & 0 & x_m \\ 0 & 1 & y_m \\ 0 & 0 & 1 \end{pmatrix} \qquad (12)\\ \mathbf{R} &= \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (13)\\ \mathbf{S} &= \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix}. \qquad (14) \end{aligned}$$

This yields the following decomposition for $\mathbf{C}^{-1}$

$$x^T\mathbf{T^{-T}R^{-T}S^{-T}C_1^{-1}S^{-1}R^{-1}T^{-1}}x = 0. \qquad (15)$$

Explicitly that is

$$\mathbf{C}^{-1} = \mathbf{T^{-T}R^{-T}S^{-T}C_1^{-1}S^{-1}R^{-1}T^{-1}} = \begin{pmatrix} \mathbf{R_2S_2^{-2}R_2^T} & \mathbf{-R_2S_2^{-2}R_2^Tm} \\ \mathbf{-m^TR_2S_2^{-2}R_2^T} & \mathbf{m^TR_2S_2^{-2}R_2^Tm} \end{pmatrix} \qquad (16)$$

where $\mathbf{R_2}$ and $\mathbf{S_2}$ on the right-hand side are the upper left 2×2 matrices of $\mathbf{R}$ and $\mathbf{S}$ respectively and $\mathbf{S^{-2}} = \mathbf{S_2}^{\mathbf{-T}}\mathbf{S_2^{-1}}$. From this decomposition it can be seen that the upper left 2×2 matrix of matrix $\mathbf{C}^{-1}$ is a product of two rotation matrices with a nonuniform scaling inbetween. This matrix is symmetric. For symmetric matrices there is a special form of the singular value decompostion (SVD) called Takagi factorization which is

$$\mathbf{A} = \mathbf{U^T D U} \tag{17}$$

where $\mathbf{U}$ is an orthogonal[1] and $\mathbf{D}$ a diagonal matrix. This is exactly what is needed to recover $\mathbf{R}$, therefore $\mathbf{R} = \mathbf{U}^T$ when applying the SVD to $\mathbf{R_2 S_2^{-2} R_2{}^T}$. It should be pointed out, that $\mathbf{S}$ is not recovered yet although it may seem so. The problem is, that equation 1 can be multiplied by an arbitrary factor without changing the equation. Therefore the assumption in equation 16, that the homogenous part of the coordinates is equal to 1 does not hold. This is not a real problem since

$$\mathbf{C}^{-1} = \mathbf{T^{-T} R^{-T} S^{-T} C_1^{-1} S^{-1} R^{-1} T^{-1}} \tag{18}$$

with known $\mathbf{T}$ and $\mathbf{R}$. Therefore

$$\mathbf{R^T T^T C^{-1} T R} = \mathbf{S^{-T} C_1^{-1} S^{-1}} = \mathbf{S_3^{-2}} \tag{19}$$

which is a diagonal matrix that can be divided by its last component $\mathbf{S_{3,3}^{-2}}$ to enforce the assumption made above. After this normalization step s_x and s_y can be extracted from $\mathbf{S_3^{-2}}$:

$$s_x = \sqrt{\left|\frac{1}{\mathbf{S_{1,1}^{-2}}}\right|} \tag{20}$$

$$s_y = \sqrt{\left|\frac{1}{\mathbf{S_{2,2}^{-2}}}\right|} \tag{21}$$

The major and the minor axis are now the two columns of matrix $\mathbf{R}$ scaled by s_x and s_y respectively and the midpoint is $\mathbf{m}$.

It should be noted that $\mathbf{S_{1,1}^{-2}}$ and $\mathbf{S_{2,2}^{-2}}$ are negative due to division by $\mathbf{S_{3,3}^{-2}}$ which again is negative because of the last component of $\mathbf{C_1^{-1}} = \mathrm{diag}(1, 1, -1)$. This implies that the absolute value in equation 21 could be replaced by a minus sign. However, if matrix $\mathbf{C}^{-1}$ describes a hyperboloid either of $\mathbf{S_{1,1}^{-2}}$ and $\mathbf{S_{2,2}^{-2}}$ can be positive or negative. Using the absolute value results in a stable computation. The rotation, translation and scaling components are still valid.

Generalization to Higher Dimensions

Even though the derivation has been carried out for ellipses in two dimensions, the resulting decomposition is valid in higher dimensions as well. The only change occurs in equation 21 where there are now N components for N dimensional conics that have to be divided by the $(N+1)$th diagonal entry of matrix $\mathbf{S_{(N+1)}^{-2}}$.

[1]Orthogonal transformations preserve a symmetric inner product. Thus an orthogonal matrix has an orthonormal basis.

2 Application to covariance matrix and error ellipsoid

Suppose the covariance matrix $\mathbf{\Sigma}$ of a multivariate gaussian distribution is given. It can be visualized by an error ellipsoid that takes the form

$$(\theta - \overline{\theta})^T \mathbf{\Sigma}^{-1}(\theta - \overline{\theta}) = k \tag{22}$$

where k is a confidence factor determined by the χ^2 distribution and $\overline{\theta}$ the mean value of the gaussian distribution mentioned above. The probability that θ appears in the interior of the ellipsoid defined by equation 22 is equal to $P_{\chi^2}(k,2)$ [2]. Some values for k are $k = 2.41$ (70%) and $k = 5.99$ (95%). Writing equation 22 in the form

$$(\theta^T \mathbf{\Sigma}^{-1} - \overline{\theta}^T \mathbf{\Sigma}^{-1})(\theta - \overline{\theta}) - k = 0 \tag{23}$$

$$\theta^T \mathbf{\Sigma}^{-1}\theta - \overline{\theta}^T \mathbf{\Sigma}^{-1}\theta - \theta^T \mathbf{\Sigma}^{-1}\overline{\theta} + \overline{\theta}^T \mathbf{\Sigma}^{-1}\overline{\theta} - k = 0 \tag{24}$$

shows, that it can be written in homogenous coordinates as

$$\tilde{\theta}^T \begin{pmatrix} \mathbf{\Sigma}^{-1} & -\mathbf{\Sigma}^{-1}\overline{\theta} \\ -\overline{\theta}^T \mathbf{\Sigma}^{-1} & \overline{\theta}^T \mathbf{\Sigma}^{-1}\overline{\theta} - k \end{pmatrix} \tilde{\theta} = 0 \tag{25}$$

where $\tilde{\theta} = \begin{pmatrix} \theta \\ 1 \end{pmatrix}$. This is exactly the same form as in equation 16 which makes it feasible to compute the main axes and scalings (matrices $\mathbf{R}$ and $\mathbf{S}$) to extract the main directions of error. $\mathbf{T}$ is already known. It is derived from the mean value of the distribution. In two or three dimensions it also allows the drawing of the ellipsoid. Generate points $\hat{x}_i$ on the unit circle/sphere and transform them using $x_i = \mathbf{TRS}\hat{x}_i$.

3 Mahalanobis distance

Let us denote the matrix of equation 25 as $\mathbf{C}^{-1}$ and call $\mathbf{C}$ *homogenous covariance matrix*, because it contains not only information about the covariances, but also about the mean value and the scale. The Mahalanobis norm is defined as:

$$|\mathbf{x}|_M^2 = \mathbf{x}^T \mathbf{\Sigma}^{-1}\mathbf{x} \tag{26}$$

where $\mathbf{\Sigma}$ is the covariance matrix as in the previous section. $\mathbf{x}$ is a non-homogenous vector and the distance measure should be taken with respect to the mean value of the distribution, i.e. the vector $\mathbf{x}$ should have been moved by $-\overline{\theta}$ before taking the distance measure. That is because the covariance matrix $\mathbf{\Sigma}$ is defined around the mean value of the distribution. The *homogenous covariance matrix* takes this already into acount but $\mathbf{x}$ has to be extended to

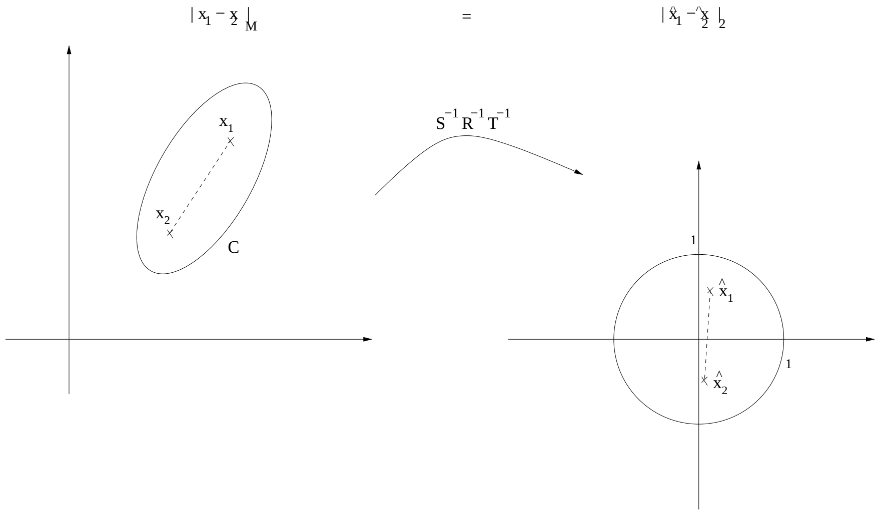

Figure 1: Relation between Mahalanobis distance and Euclidean distance

a homogenous vector with its homogenous part set to 1 ($x = \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix}$). The Mahalanobis distance reads then:

$$|x|_M^2 = x^T \mathbf{C}^{-1} x \tag{27}$$

According to equation 16, the matrix $\mathbf{C}^{-1}$ can be decomposed into translation, rotation and non-uniform scaling parts. Therefore equ. 27 reads:

$$|x|_M^2 = x^T \mathbf{T}^{-\mathbf{T}} \mathbf{R}^{-\mathbf{T}} \mathbf{S}^{-\mathbf{T}} \mathbf{S}^{-\mathbf{1}} \mathbf{R}^{-\mathbf{1}} \mathbf{T}^{-\mathbf{1}} x \tag{28}$$

which is the same as

$$|x|_M^2 = (\mathbf{S}^{-\mathbf{1}} \mathbf{R}^{-\mathbf{1}} \mathbf{T}^{-\mathbf{1}} x)^T (\mathbf{S}^{-\mathbf{1}} \mathbf{R}^{-\mathbf{1}} \mathbf{T}^{-\mathbf{1}} x) \tag{29}$$

This is the square of the Euclidean distance in a transformed space. The situation is depicted in figure 1. Therefore if the notation of equation **??** is adopted,

$$\hat{x} = \mathbf{S}^{-\mathbf{1}} \mathbf{R}^{-\mathbf{1}} \mathbf{T}^{-\mathbf{1}} x \tag{30}$$

the following equality can be stated:

$$|x|_M = |\hat{x}|_2 = |\mathbf{S}^{-\mathbf{1}} \mathbf{R}^{-\mathbf{1}} \mathbf{T}^{-\mathbf{1}} x|_2 \tag{31}$$

3.1 Mahalanobis distance between point and line

In homogenous coordinates (2D) a line l is given by the implicit equation $l^T x = 0$. A line l connecting two points x_1 and x_2 is given by:

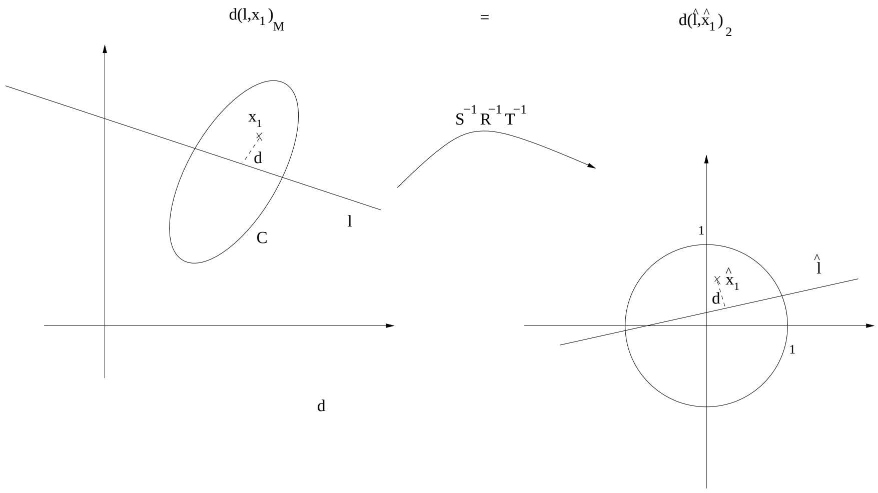

Figure 2: Relation between Mahalanobis distance between point and line and Euclidean distance

$$l = x_1 \times x_2$$

and the point x at the intersection of two lines l_1 and l_2 is given by:

$$x = l_1 \times l_2$$

if a Euclidean distance measure $d(\hat{x}, \hat{l})_2$ between a point and a line is available, this can be used to define a Mahalanobis distance measure for this point-line pair. The result in equ. 31 is used for this purpose. Point x and line l have to be transformed to normalized space and the Euclidean distance measure has to be taken there. For the 2D-homogenous case this is especially simple since:

$$\mathbf{A}^{-1} = \mathbf{S^{-1}R^{-1}T^{-1}} \tag{32}$$

$$(\mathbf{A}^{-1}x_1 \times \mathbf{A}^{-1}x_2) = (\mathbf{A}^{-1})^*(x_1 \times x_2) \tag{33}$$

where $\mathbf{M}^*$ is the matrix of cofactors of $\mathbf{M}$. If $\mathbf{M}$ is invertible then

$$\mathbf{M}^* = det(\mathbf{M})\mathbf{M}^{-T} \tag{34}$$

Equation (33) and (34) are given in [1], p. 555 without proof. Using equations (33,34) the transformation for a homogenous line in 2D can be written as:

$$\hat{l} = (\mathbf{A}^{-1}x_1 \times \mathbf{A}^{-1}x_2) = det(\mathbf{A}^{-1})\mathbf{A}^T(x_1 \times x_2) \hat{=} \mathbf{A}^T l \tag{35}$$

with $\hat{=}$ being equality up to scale. Since a scale factor doesn't change the implicit equation for line l, the transformation $\mathbf{A}^T$ can be applied to the homogenous coordinates of l directly. Thus similarly to equ. 31:

$$d(x,l)_M = d(\mathbf{S^{-1}R^{-1}T^{-1}}x, \mathbf{S}^T\mathbf{R}^T\mathbf{T}^T l)_2 = d(\hat{x},\hat{l})_2 \tag{36}$$

This situation is shown in figure 2.

4 TODO

- explain relation to general conic
- check what happens for degenerate conics (e.g. parabola - equ. (4) will be singular)
- distance point - conic (http://www-sop.inria.fr/robotvis/personnel/zzhang/Publis/Tutorial-Estim/node11.html)

5 Thanks

Thanks to Andreas Kapp, Ady Ecker, Christof Kraus, and Jason Filos who pointed out mistakes in earlier versions of this text. I have to apologize for my slow response in removing the erroneous section on the distance between a conic and a point and on the negative last entry in the unit circle matrix.

References

[1] R. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 2000.

[2] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. Technical Report 2927, INRIA, Sophia-Antipolis Cedex, France, 1996.