A Fully Projective Error Model for Visual Reconstruction

Bill Triggs
LIFIA, INRIA Rhône-Alpes,
46 avenue Félix Viallet, 38031 Grenoble, France.
Bill.Triggs@imag.fr

Abstract

Measurement uncertainty is a recurrent concern in visual reconstruction. Image formation and 3D structure recovery are essentially projective processes that do not quite fit into the classical framework of affine least squares, so intrinsically projective error models must be developed. This paper describes initial theoretical work on a fully projective generalization of affine least squares. The result is simple and projectively natural and works for a wide variety of projective objects (points, lines, hyperplanes, and so on). The affine theory is contained as a special case, and there is also a canonical probabilistic interpretation along the lines of the classical least-squares/Gaussian/approximate log-likelihood connection. Standard linear algebra often suffices for practical calculations.

1 Introduction

For reliable reconstruction of 3D geometry from image measurements it is essential to take account of measurement uncertainties. Image formation and reconstruction are essentially projective processes and the errors they generate do not quite fit into the classical linear framework of error models such as affine least squares. In the absence of fully projective error models, uncertainty is currently handled on a rather *ad hoc* basis, often by simply feeding quasilinear phenomenological error estimates into a general nonlinear least squares routine. This produces numerical answers, but it obscures the underlying geometric structure of the problem and makes further theoretical (*i.e.* algebraic) development impossible.

This paper describes initial work on a fully projective generalization of affine least squares. The resulting theory is relatively simple and projectively natural, and it extends to a wide variety of projective objects: points, lines, hyperplanes and so forth. Given a choice of 'plane at infinity', the classical affine theory is contained as a special case. There is a canonical probabilistic interpretation along the lines of the potent least-squares/Gaussian/approximate log-likelihood connection, and standard linear algebra often suffices for practical calculations.

The notion that projective geometry should be 'simpler' than affine geometry is central to this work. Several aspects of projective naturality played key rôles in the development of the theory:

- It should look simple and natural in homogeneous coordinates and work equally well at all points of a projectivized space, from the origin right out to the hyperplane at infinity.
- It should generalize easily from points to hyperplanes, lines and other projective subspaces, and perhaps even to higher-degree projective varieties like quadrics and cubics.
- For projective subspaces, it should be simply expressible in terms of Grassmann coordinates (*i.e.* 'the natural parameterization').
- It should behave naturally under point/hyperplane — and hence Grassmann/dual Grassmann — duality, and also under projective transformations.

We will use tensorial notation with all indices written out explicitly, as in [7, 9]. Most of the development will apply to general projective spaces, but when we refer to the computer vision case of 2D projective images of a 3D projective world we will use indices $a = 0, \ldots, 3$ for homogeneous world vectors and $A = 0, 1, 2$ for homogeneous image vectors. The Einstein summation convention applies to corresponding covariant/contravariant index pairs, so for example $\mathbf{T}^a_b\, \mathbf{x}^b$ stands for matrix-vector multiplication $\sum_b \mathbf{T}^a_b\, \mathbf{x}^b$.

Probability densities will be denoted $\mathbf{dp}(\mathbf{x}^a|\text{Evidence})$ to emphasize that they are densities in $\mathbf{x}^a$ rather than functions. A **relative likelihood** is a function defined by dividing a probability density by a (sometimes implicit) prior $\mathbf{dp}(\mathbf{x}^a)$ or volume form ('uniform prior') $\mathbf{dV}$. **Log-unlikelihood** means -2 times the logarithm of a relative likelihood, defined up to an additive constant. χ^2 variables are log-unlikelihoods.

Although many specific error models have appeared in the literature there have been very few attempts to unify different aspects of the field. Zhang & Faugeras [10, 1] and Luong *et al* [3] respectively provide linearized least squares models for 3D point and line reconstruction and fundamental matrix estimation. Mohr *et al* [5] formulate multi-image reconstruction as a batch-mode nonlinear least squares problem, and more recently McLauchlan & Murray [4] describe a suboptimal but practically efficient linearized incremental framework for several types of reconstruction.

2 Homogenized Affine Least Squares

To motivate the projective model we will re-express classical least squares for affine points in homogeneous coordin-

[0][Draft: March 19, 1995] Submitted to *IEEE Workshop on Representations of Visual Scenes*, Cambridge, MA., June 1995.

ates. Consider a random vector $\mathbf{x} = (x^1, \ldots, x^d)^\top$ in a d-dimensional affine space, subject to some probability distribution with mean $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{X}$. We can homogenize $\mathbf{x}$ and embed it in d dimensional projective space by adding an extra component $\mathbf{x}^0 \equiv 1$ to make a $d+1$ component homogeneous vector $\mathbf{x}^a = (1, x^1, \ldots, x^d)^\top$. The mean and covariance are neatly contained in the expectation value $\langle \mathbf{x}^a \mathbf{x}^b \rangle$:

$$\left\langle \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} \begin{pmatrix} 1 & \mathbf{x}^\top \end{pmatrix} \right\rangle = \begin{pmatrix} 1 \\ \bar{\mathbf{x}} \end{pmatrix} \begin{pmatrix} 1 & \bar{\mathbf{x}}^\top \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{X} \end{pmatrix} = \begin{pmatrix} 1 & \bar{\mathbf{x}}^\top \\ \bar{\mathbf{x}} & \bar{\mathbf{x}}\bar{\mathbf{x}}^\top + \mathbf{X} \end{pmatrix}$$

Inverting this **homogenized covariance matrix** gives an equally simple **homogenized information matrix**:

$$\begin{pmatrix} 1 & \bar{\mathbf{x}}^\top \\ \bar{\mathbf{x}} & \bar{\mathbf{x}}\bar{\mathbf{x}}^\top + \mathbf{X} \end{pmatrix}^{-1} = \begin{pmatrix} 1 + \bar{\mathbf{x}}^\top \mathbf{X}^{-1} \bar{\mathbf{x}} & -\bar{\mathbf{x}}^\top \mathbf{X}^{-1} \\ \mathbf{X}^{-1} \bar{\mathbf{x}} & \mathbf{X}^{-1} \end{pmatrix}$$

Finally, contracting the information matrix with $\mathbf{x}^a$ and $\mathbf{x}^b$ gives (up to an additive constant) the chi-squared/Mahalanobis distance/Gaussian exponent/approximate log-unlikelihood of $\mathbf{x}$ with respect to $\bar{\mathbf{x}}$ and $\mathbf{X}$:

$$\begin{aligned} 1 + \chi^2(\mathbf{x}|\bar{\mathbf{x}}, \mathbf{X}) &= 1 + (\mathbf{x} - \bar{\mathbf{x}})^\top \mathbf{X}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \begin{pmatrix} 1 & \mathbf{x}^\top \end{pmatrix} \begin{pmatrix} 1 + \bar{\mathbf{x}}^\top \mathbf{X}^{-1} \bar{\mathbf{x}} & -\bar{\mathbf{x}}^\top \mathbf{X}^{-1} \\ \mathbf{X}^{-1} \bar{\mathbf{x}} & \mathbf{X}^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} \end{aligned}$$

The determinants of the homogenized covariance and information matrices are simply $\mathrm{Det}(\mathbf{X})$ and $\mathrm{Det}(\mathbf{X}^{-1})$.

The moral is that homogenization makes many Gaussian and affine least squares notions even simpler and more uniform. In fact, it is a nice way to work even when there is no question of projective space, because the parameters of the Gaussian are all kept together in one matrix. Derivations and coding become easier because equations for means fall out of those for covariances.

3 Projective Point Distributions

Now we briefly sketch the key elements of the projective least squares error model for a single projective point. For a more complete development of the theory see [8].

Consider an arbitrary probability density $\mathbf{dp}(\mathbf{x}^a)$ for an uncertain point in a d dimensional projective space $\mathcal{P}^a$. To be projectively well defined, the density must be *scale invariant*: $\mathbf{dp}(\mathbf{x}^a) = \mathbf{dp}(\lambda \mathbf{x}^a)$ for all $\mathbf{x}^a$ and all $\lambda \neq 0$. Integration against $\mathbf{dp}(\cdot)$ induces a linear expectation value operator $\langle \cdot \rangle$ on the scale-invariant functions on $\mathcal{P}^a$:

$$\langle f \rangle \equiv \int_{\mathcal{P}^a} f(\mathbf{x}^a) \, \mathbf{dp}(\mathbf{x}^a)$$

The homogenized affine analysis given above suggests that we should try to evaluate a *homogeneous covariance tensor* $\mathbf{X}^{ab} \sim \langle \mathbf{x}^a \mathbf{x}^b \rangle$, invert it to produce a *homogeneous information tensor* $\mathbf{M}_{ab} \equiv (\mathbf{X}^{-1})_{ab}$, and then take $1 + \chi^2(\mathbf{x}^a | \mathbf{X}^{ab}) \sim \mathbf{M}_{ab} \mathbf{x}^a \mathbf{x}^b$ as a measure of normalized squared error or approximate log-unlikelihood. Unfortunately, this can not quite work as it stands because $\langle \cdot \rangle$ is only defined for *scale invariant* functions of $\mathbf{x}^a$ and the moment monomials $\mathbf{x}^{a_1} \cdots \mathbf{x}^{a_k}$ all depend on the scale of $\mathbf{x}^a$. On a general projective space there is no canonical way to fix this scale, so classical means and covariances are simply not defined.

This problem can be resolved by introducing an auxiliary normalization tensor $\mathbf{N}_{ab}$ and homogenizing with respect to it, so that quantities of the form $(\mathbf{M}_{ab} \mathbf{x}^a \mathbf{x}^b)$ are replaced by homogeneous scale invariant quantities $(\mathbf{M}_{ab} \mathbf{x}^a \mathbf{x}^b) / (\mathbf{N}_{cd} \mathbf{x}^c \mathbf{x}^d)$. We will call such functions **biquadrics** because their level surfaces are quadric: $(\mathbf{M}_{ab} \mathbf{x}^a \mathbf{x}^b) / (\mathbf{N}_{cd} \mathbf{x}^c \mathbf{x}^d) = \lambda$ implies $(\mathbf{M}_{ab} - \lambda \, \mathbf{N}_{ab}) \, \mathbf{x}^a \mathbf{x}^b = 0$. As an example, the affine normalization condition $x^0 = 1$ can be relaxed if we divide through by $\mathbf{N}^{\mathrm{aff}}_{ab} \mathbf{x}^a \mathbf{x}^b = (\mathbf{p}^\infty_a \mathbf{x}^a)^2 = (x^0)^2$, where $\mathbf{N}^{\mathrm{aff}}_{ab} \equiv \mathbf{p}^\infty_a \mathbf{p}^\infty_b = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $\mathbf{p}^\infty_a = (1 \ 0 \ \ldots \ 0)$ is the plane at infinity, At first sight the normalizer simply provides a fiducial scaling $\mathbf{N}_{ab} \mathbf{x}^a \mathbf{x}^b = 1$ with respect to which the error model can be defined, but ultimately $\mathbf{N}$ is on a par with $\mathbf{M}$ and tends to play an equally active rôle in the theory.

3.1 Basic Equations

Given a projective probability distribution $\mathbf{dp}(\mathbf{x}^a)$ and an arbitrary symmetric positive semidefinite **normalization tensor** $\mathbf{N}_{ab}$ on a projective space $\mathcal{P}^a$, we can define the **homogeneous covariance tensor**

$$\mathbf{X}^{ab} \equiv \left\langle \frac{\mathbf{x}^a \mathbf{x}^b}{\mathbf{N}_{cd} \mathbf{x}^c \mathbf{x}^d} \right\rangle$$

Note that $\mathbf{X}$ is symmetric, positive semidefinite and independent of the scale of $\mathbf{x}^a$, but it does depend on the value and scale of $\mathbf{N}$. If $\mathbf{N}$ has null directions it should be compatible with $\mathbf{dp}(\cdot)$ in the sense that the above expectation value is finite, *i.e.* the distribution should not have too much weight in the vicinity of the null space of $\mathbf{N}$. Since $\langle \cdot \rangle$ is linear, if $\mathbf{dp}(\cdot)$ is correctly normalized we have the following **covariance normalization consistency condition** on $\mathbf{X}$

$$\mathbf{N}_{ab} \mathbf{X}^{ab} = \left\langle \frac{\mathbf{N}_{ab} \mathbf{x}^a \mathbf{x}^b}{\mathbf{N}_{cd} \mathbf{x}^c \mathbf{x}^d} \right\rangle = \langle 1 \rangle = 1$$

Viewing $\mathbf{X}$ and $\mathbf{N}$ as matrices, this can be written $\mathrm{Trace}(\mathbf{N}\mathbf{X}) = 1$. If $\mathbf{dp}(\cdot)$ is not correctly normalized, we can normalize by dividing through by $\mathbf{N}_{ab} \mathbf{X}^{ab} = \langle 1 \rangle \neq 1$. The normalized covariance tensor is then

$$\left\langle \frac{\mathbf{x}^a \mathbf{x}^b}{\mathbf{N}_{cd} \mathbf{x}^c \mathbf{x}^d} \right\rangle / \langle 1 \rangle = \frac{\mathbf{X}^{ab}}{\mathbf{N}_{cd} \mathbf{X}^{cd}}$$

Usually, one can arrange to work with normalized quantities and ignore the scale factor, *i.e.* $\mathbf{N}_{ab} \mathbf{X}^{ab} = 1$.

By analogy with the homogenized affine case and assuming for the moment that $\mathbf{X}$ is nonsingular, we can invert it to produce a **homogeneous information tensor** $\mathbf{M}_{ab} \equiv (\mathbf{X}^{-1})_{ab}$

and define a corresponding **homogeneous** $1+\chi^2$ **function**

$$1+\chi^2(\mathbf{x}^a|\mathbf{X},\mathbf{N}) \equiv \frac{\mathbf{M}_{ab}\mathbf{x}^a\mathbf{x}^b}{\mathbf{N}_{cd}\mathbf{x}^c\mathbf{x}^d}$$

It is not immediately obvious that these definitions make sense, but one can argue [8] that they do in fact lead to a coherent theory of approximate least squares estimation. Two key approximations are required, both of which are *exact* in the affine case and generally accurate whenever the uncertainty is small compared to the scale of projective space. (And it is only in the limit of small uncertainty that *any* least squares technique becomes a good approximation to the more rigorous maximum relative likelihood theory).

As in the affine case, it is often useful to regard the information as the primitive quantity and derive the covariance from it. The quadratic (Gaussian) exponent $\chi^2(\mathbf{x}|\bar{\mathbf{x}},\mathbf{X}) = (\mathbf{x}-\bar{\mathbf{x}})^\top\mathbf{X}^{-1}(\mathbf{x}-\bar{\mathbf{x}})$ is the keystone of affine estimation theory because it is the leading term in the **central moment expansion** of an arbitrary distribution. The **central limit theorem** (which guarantees the asymptotic dominance of this term given 'reasonable' behaviour of the underlying distributions) is the ultimate probabilistic justification for affine least squares techniques.

Similarly, **biquadric** exponents $1+\chi^2(\mathbf{x}^a|\mathbf{X},\mathbf{N}) \equiv \left(\mathbf{M}_{ab}\mathbf{x}^a\mathbf{x}^b\right) / \left(\mathbf{N}_{cd}\mathbf{x}^c\mathbf{x}^d\right)$ lie at the heart of projective least squares. In particular, they are likely to be good asymptotic approximations to arbitrary projective log-unlikelihood functions, so that estimation theory based on them should 'work' in much the same way that conventional least squares 'works' in affine space. Given this, the uncertainties of projective points can be modelled with **biquadric probability distributions**

$$\mathbf{dp}(\mathbf{x}^a) \sim \exp\left(-\frac{1}{2}\frac{\mathbf{M}_{ab}\mathbf{x}^a\mathbf{x}^b}{\mathbf{N}_{cd}\mathbf{x}^c\mathbf{x}^d}\right)\mathbf{dV}$$

much as affine uncertainties can be modelled with Gaussians.

Several approximations are required here. Firstly, there is no canonical volume form $\mathbf{dV}$ on projective space, so it is necessary to make an 'arbitrary but reasonable' choice of this 'uniform prior'. This is annoying, but it is not specifically a problem with projective least squares: implicitly or explicitly, *every* least squares theory makes such choices. The mere existence of a uniform volume form on affine space does not make it a universally acceptable prior.

Secondly, biquadric distributions are somewhat less tractable than Gaussian ones and (except in the limit of affine normalization) there does not seem to be a closed form for their integrals. This means that we do not know the exact functional form of the relation $\mathbf{X}=\mathbf{X}(\mathbf{M},\mathbf{N})$ between the covariance and the information and normalization. However with an appropriate choice of projective basis the integral can be approximated by a Gaussian [8], with the result that *for properly normalized distributions* the 'classical' homogenized affine formula $\mathbf{X}\approx\mathbf{M}^{-1}$ is still approximately valid. Here properly normalized means that the covariance normalization condition $\mathbf{N}_{ab}\mathbf{X}^{ab}=1$ holds for $\mathbf{X}\equiv\mathbf{M}^{-1}$, so that the distribution in $\mathbf{M}$ and $\mathbf{N}$ is approximately normalized in the sense that $\langle 1\rangle \approx 1$.

It is often necessary to normalize an unnormalized biquadric distribution. Rescaling the density function amounts to **shifting** the information $\mathbf{M}$ by a multiple of $\mathbf{N}$: $\mathbf{M}\to\mathbf{M}-\lambda\mathbf{N}$. We will say that $\mathbf{M}$ is **correctly shifted** if $\mathbf{M}^{-1}$ has the correct normalization to be a covariance: $\mathbf{N}_{ab}(\mathbf{M}^{-1})^{ab}=1$. The correct shift factor can be found by solving the nonlinear **normalizing shift equation**

$$\mathbf{N}_{ab}\left((\mathbf{M}-\lambda\mathbf{N})^{-1}\right)^{ab}=1$$

This amounts to a polynomial of degree $\mathrm{Rank}(\mathbf{N})$ in λ, linear in the case of affine normalization. The desired solution is the smallest real root, which can be roughly approximated by the **approximate shift solution**

$$\lambda \approx \left(\mathbf{N}_{ab}(\mathbf{M}^{-1})^{ab}\right)^{-1}-1$$

The two main approximations required to make projective least squares 'work' are the covariance estimate $\mathbf{X}\approx\mathbf{M}^{-1}$ and the approximate shift solution. Both are *exact* for affine normalization and generally accurate for small uncertainties, but neither is very good for distributions that spread across the entire width of projective space. However, least squares is not really suitable for weak evidence (wide distributions) in any event. It makes too many assumptions about the uniformity of priors and the asymptotic shapes of distributions to be competitive with the more rigorous maximum relative likelihood theory in this case. Its main strengths are simplicity and asymptotic correctness in the limit where many moderate pieces of evidence combine to make a single strong one. And it is in exactly this limit that the additional approximations made by projective least squares become accurate.

To define a meaningful distribution, $\mathbf{M}$ and $\mathbf{N}$ need to be **non-negative**, but it is practically useful to allow them to have null directions. To guarantee the normalization condition $\mathbf{N}_{ab}(\mathbf{M}^{-1})^{ab}=1$, we will impose a **null space compatibility condition**: the null space of $\mathbf{M}$ must be contained in that of $\mathbf{N}$. This ensures that any pseudo-inverse of a singular $\mathbf{M}$ can be used to evaluate $\mathbf{N}_{ab}(\mathbf{M}^{-1})^{ab}$ (it makes no difference which). However, the covariance tensor $\mathbf{X}\approx\mathbf{M}^{-1}$ is only defined for nonsingular $\mathbf{M}$.

3.2 Normalizations

If we take $\mathbf{N}$ to be the **affine normalization** $\mathbf{N}^{\mathrm{aff}}_{ab}\equiv\mathbf{p}^\infty_a\mathbf{p}^\infty_b = \left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$ where $\mathbf{p}^\infty_a=(1\ 0\ \dots\ 0)$ is the hyperplane at infinity, the biquadric distribution reduces to the homogenized affine case we started from. In this case the covariance normalization condition is simply $\mathbf{X}^{00}=\mathbf{N}^{\mathrm{aff}}_{ab}\mathbf{X}^{ab}=1$ and (to the extent that the underlying distribution is well modelled by a Gaussian) the homogeneous $1+\chi^2$ function is one plus a genuine classical χ^2 variable.

On the other hand, if $\mathbf{N}$ is taken to be the identity matrix in some projective basis we have a **spherical normalization** $\mathbf{N}_{ab}\mathbf{x}^a\mathbf{x}^b=\sum_{a=0}^d(x^a)^2=1$ and the error model reduces to a spherical analogue of linear least squares, with 'distances' measured by sines of angles on the unit sphere. The two normalizations coincide for points near the origin but differ significantly near the hyperplane at infinity. The affine normalization vanishes on the plane at infinity and points there are infinitely improbable, whereas the spherical normalization is regular and well behaved for all points, including those at infinity.

These are just two of the infinitely many possible choices for $\mathbf{N}$. There is no universally 'correct' or 'canonical' normalizer. Ideally, $\mathbf{N}$ should be chosen to reflect the mechanism that generates the experimental uncertainty, although in practice numerical expediency is also a factor.

With the spherical normalization it is natural to take an eigenvalue expansion of $\mathbf{X}$ in an 'orthonormal' projective basis. The mode (maximum likelihood value) of the distribution is the maximum-eigenvalue eigenvector of $\mathbf{X}$ and the remaining eigenvectors give the principal axes of the uncertainty ellipsoids. Small eigenvalues correspond to directions with little uncertainty, while for large ones (those near the modal eigenvalue) the distribution spreads across the entire width of projective space. For the 'uniform' distribution, $\mathbf{X} = \frac{1}{d+1}\mathbf{I}$.

More generally, given any $\mathbf{M}$ and $\mathbf{N}$ there is always some projective basis in which they are in **canonical form**, *i.e.* simultaneously diagonal with $\mathbf{N}$ having entries $+1$ or 0. In this basis the global minimum of $1 + \chi^2$ is at the minimum eigenvalue eigenvector of $\mathbf{M}$ along a '1' direction of $\mathbf{N}$, and a correctly normalized distribution has $\sum 1/\lambda_i = 1$ where the sum is over the inverse eigenvalues of $\mathbf{M}$ along '1' directions of $\mathbf{N}$.

3.3 Homogeneous Chi-Squared

Except in the case of affine normalization and an underlying Gaussian distribution, the homogeneous χ^2 variable is unlikely to have a classical χ^2 distribution. However, the " χ^2 " variables used in statistics seldom *do* have exact χ^2 distributions and that does not stop them being useful error measures. Several familiar properties of the traditional χ^2 do continue to hold. Our χ^2 is nonnegative ($1+\chi^2 \geq 1$), and for nonsingular $\mathbf{M}$ its expectation value is the number of independent degrees of freedom, *i.e.* the dimension d of the projective space:

$$\left\langle 1 + \chi^2(\mathbf{x}^a) \right\rangle = \mathbf{M}_{ab} \left\langle \frac{\mathbf{x}^a \mathbf{x}^b}{\mathbf{N}_{cd}\mathbf{x}^c\mathbf{x}^d} \right\rangle = \mathbf{X}^{-1}_{ab}\mathbf{X}^{ab} = d + 1$$

Moreover, we have already seen that — in analogy to the error ellipsoids of the classical χ^2 — the level surfaces of $1 + \chi^2$ are always quadric. Near a minimum of $1 + \chi^2$ these surfaces will be ellipsoidal, but further away they may cut the plane at infinity and hence appear hyperboloidal rather than ellipsoidal.

3.4 Homogeneous Taylor Approximation

To get an idea of why biquadric functions should appear in projective least squares, consider an arbitrary smooth scale invariant function on projective space: $f(\mathbf{x}^a) = f(\lambda\mathbf{x}^a)$. $f(\cdot)$ can be approximated with a conventional Taylor series at a point, but this is not very satisfactory as the resulting truncated Taylor polynomials are not exactly scale invariant and depend on the scale of the homogeneous vector at which the derivatives are evaluated. What is needed is a *projectively invariant* analogue of the Taylor series. Once again homogenization with respect to a normalizer $\mathbf{N}_{ab}$ makes this possible.

Consider the scale-invariant function

$$f(\mathbf{x}^a) = \frac{\mathbf{M}_{a_1a_2\cdots a_{2k}}\mathbf{x}^{a_1}\mathbf{x}^{a_2}\cdots\mathbf{x}^{a_{2k}}}{(\mathbf{N}_{ab}\mathbf{x}^a\mathbf{x}^b)^k}$$

where $\mathbf{M}$ and $\mathbf{N}$ are arbitrary symmetric tensors. Multiplying out and differentiating $2k$ times using the usual iterated chain rule gives

$$\begin{aligned}
\mathbf{M}_{a_1a_2\cdots a_{2k}} &= \frac{1}{(2k)!}\frac{\partial^{2k}\left[(\mathbf{N}_{ab}\mathbf{x}^a\mathbf{x}^b)^k \cdot f(\mathbf{x}^a)\right]}{\partial\mathbf{x}^{a_1}\cdots\partial\mathbf{x}^{a_{2k}}} \\
&= \sum_{j=0}^{2k} \frac{\partial^j f}{\partial\mathbf{x}^{(a_1}\cdots\partial\mathbf{x}^{a_j}} \cdot \frac{\partial^{2k-j}(\mathbf{N}_{ab}\mathbf{x}^a\mathbf{x}^b)^k}{\partial\mathbf{x}^{a_{j+1}}\cdots\partial\mathbf{x}^{a_{2k})}}
\end{aligned}$$

Here, $(a_1 \cdots a_{2k})$ means 'take the symmetric part'. The factorial weights of the familiar iterated chain rule are subsumed by this symmetrization.

This formula gives $\mathbf{M}$ in terms of $\mathbf{N}$ and the first $2k$ derivatives of $f(\cdot)$. Now choose any $\mathbf{N}$ and let $f(\cdot)$ stand for an arbitrary scale-invariant function. The resulting $\mathbf{M}$ defines a function $(\mathbf{M}_{a_1\cdots a_{2k}}\mathbf{x}^{a_1}\cdots\mathbf{x}^{a_{2k}})/(\mathbf{N}_{ab}\mathbf{x}^a\mathbf{x}^b)^k$ that is guaranteed to agree with $f(\cdot)$ to order $2k$ at $\mathbf{x}^a$. We will say that $\mathbf{M}$ and $\mathbf{N}$ define a $(2k)^{th}$-order **homogeneous Taylor approximation** to $f(\cdot)$ at $\mathbf{x}^a$. The 'Taylor coefficients' pack neatly into the single homogeneous tensor $\mathbf{M}_{a_1\cdots a_{2k}}$. For example adding a constant to $f(\cdot)$ amounts to adding a multiple of $\mathbf{N}_{(a_1a_2}\cdots\mathbf{N}_{a_{2k-1}a_{2k})}$ to $\mathbf{M}$. With affine normalization $\mathbf{N} \equiv \mathbf{N}^{\text{aff}}$, the homogeneous Taylor series reduces to the usual inhomogeneous affine one at $\mathbf{x} = 0$.

In the present case we are mainly interested in approximating projective log-unlikelihood functions to second order near their peaks, by analogy with the Gaussian approximation to the peak of an affine distribution. The second order homogeneous Taylor approximation is a biquadric with

$$\begin{aligned}
\mathbf{M}_{ab} = \frac{1}{2} \cdot (\mathbf{N}_{cd}\mathbf{x}^c\mathbf{x}^d) \cdot \frac{\partial^2 f}{\partial\mathbf{x}^a\partial\mathbf{x}^b} \\
+ \frac{\partial f}{\partial\mathbf{x}^a} \cdot \mathbf{N}_{bc}\mathbf{x}^c + \frac{\partial f}{\partial\mathbf{x}^b} \cdot \mathbf{N}_{ac}\mathbf{x}^c + \mathbf{N}_{ab} \cdot f
\end{aligned}$$

4 Projective Least Squares for Points

We are finally ready to describe how projective least squares can be used to estimate the position of an uncertain projective point. Suppose we have collected several independent estimates of the point's position that can be summarized by a set of biquadric distributions

$$1 + \chi^2(\mathbf{x}^a|\text{Evidence}_i) = \frac{\mathbf{M}^{(i)}_{ab}\mathbf{x}^a\mathbf{x}^b}{\mathbf{N}^{(i)}_{cd}\mathbf{x}^c\mathbf{x}^d} \qquad i = 1, \ldots, k$$

Just as one might summarize the uncertainty of an experimental measurement in affine space by specifying its mean and covariance, the uncertainty of a projective measurement can be summarized by a homogeneous information tensor $\mathbf{M}$ (or alternatively by the covariance $\mathbf{X} = \mathbf{M}^{-1}$). The corresponding normalization tensor $\mathbf{N}$ should be chosen to reflect the source of the uncertainty. For example, in computer vision a spherical normalization might be appropriate for uncertainty in the 3D angular position of an incoming visual ray relative to the camera, whereas affine normalization would probably be a better model for errors due mainly to uncertainty in the measured projection of the ray on the flat image plane (*e.g.* quantization error). However when the uncertainty is small the choice of normalization is not too critical.

Since the biquadric $1 + \chi^2$ functions represent log-unlikelihoods, the proper way to combine them into a single estimate of the position of $\mathbf{x}^a$ is to add them and then correct the constant offset term to normalize the combined distribution. First consider the **commensurable** case in which all of the normalizations $\mathbf{N}^{(i)} = \mathbf{N}$ are identical. The sum of log-unlikelihoods reduces to a sum of information tensors, as in the affine theory:

$$1 + \sum_{i=1}^{k} \chi^2(\mathbf{x}^a | \text{Evidence}_i) = \frac{\mathbf{M}_{ab}\mathbf{x}^a\mathbf{x}^b}{\mathbf{N}_{cd}\mathbf{x}^c\mathbf{x}^d}$$

where

$$\mathbf{M}_{ab} \equiv \sum_{i=1}^{k} \mathbf{M}_{ab}^{(i)} \; - \; (k-1)\,\mathbf{N}_{ab}$$

The term $(k-1)\,\mathbf{N}$ prevents the '1's of the $1 + \chi^2$ terms from accumulating, but a further correction to the shift of $\mathbf{M}$ is still needed. This can be found by solving the normalizing shift equation either exactly or approximately. The shifted $\mathbf{M}$ then defines a correctly normalized posterior distribution for $\mathbf{x}^a$ given all the evidence, and its inverse $\mathbf{X} = \mathbf{M}^{-1}$ gives the covariance in the usual way. The mode (maximum likelihood estimate) for $\mathbf{x}^a$ is the global minimum of the biquadric, *i.e.* the minimum eigenvalue eigenvector of $\mathbf{M}$ along a non-null direction of $\mathbf{N}$. The shift correction step is dispensable if only the mode is required.

Now consider the **incommensurable** case in which all of the normalizers $\mathbf{N}^{(i)}$ are different. This case is much less tractable. In general the combined log-unlikelihood is a complicated rational function and analytical or numerical approximations are required.

Many nonlinear optimization techniques can be used to find the mode. One possible way to proceed is to make a commensurable reapproximation of the combined distribution by choosing some suitable common normalization $\mathbf{N}$ and approximating each log-unlikelihood to second order with a biquadric in $\mathbf{N}$. This is straightforward except for the choice of the point(s) at which the approximations are to be based. To ensure self-consistency, the log-unlikelihoods should ideally be expanded about the true mode of the combined distribution. Since this is not known until the end of the calculation, it is necessary to start with approximations based at some sensible estimate of the mode (or perhaps at the mode of each distribution), find the resulting approximate combined mode, and if necessary iterate, at each step basing a new approximation at the latest mode estimate. Each iteration must accumulate a new approximate unshifted information tensor $\mathbf{M}$ from the component distributions and find its minimum eigenvalue eigenvector (the updated estimate of the combined mode). There is no need adjust the shift of $\mathbf{M}$ until the end of the calculation. Once the mode has been found, a second order biquadric reapproximation gives an estimate of the combined information and covariance.

There is no guarantee that this nonlinear procedure will converge correctly. Indeed, combinations of incommensurable distributions are often multi-modal, although the secondary peaks are usually negligible unless there is strongly conflicting evidence. However preliminary experiments suggest that convergence is reasonable in some realistic cases. A possible explanation for this is the fact that biquadrics are typically convex within quite a wide radius of their global minimum. They become non-convex near their non-minimal eigenvectors, but these critical points are usually far from the minimum in the standard projective basis unless $\mathbf{N}$ is particularly 'squashed'.

It might be suggested that the need to resort to approximations in the incommensurable case is a flaw of the projective least squares method, but that is not quite fair. It arises because the biquadric form is significantly richer than the Gaussian one, and even 'linear' least squares produces nonlinear equations in all but the simplest situations (*e.g.* orthogonal regression, *c.f.* section 6). In fact, except for problems with the nonlinear normalizing shift equation, the projective model is not significantly less tractable than the affine one. And even for incommensurable distributions, projective least squares provides an attractive intermediate analytical form for problems that might otherwise have produced completely 'opaque' end-to-end numerical formulations.

5 Behaviour under Projections

Now we discuss the behaviour of projective least squares under projective mappings. First consider a general situation in which some event $\mathbf{x}$ 'causes' an event $\mathbf{y}$ in the sense that $\mathbf{y} = \mathbf{f}(\mathbf{x})$ for some function $\mathbf{f}(\cdot)$, and $\mathbf{y}$ in turn gives rise to some measured evidence E. The conditional independence of E on $\mathbf{x}$ given $\mathbf{y}$ results in the classical Bayesian formula

$$\frac{\mathbf{dp}(\mathbf{x}|E)}{\mathbf{dp}(\mathbf{x})} = \left.\frac{\mathbf{dp}(\mathbf{y}|E)}{\mathbf{dp}(\mathbf{y})}\right|_{\mathbf{y}=\mathbf{f}(\mathbf{x})}$$

which says that E augments the prior likelihood $\mathbf{dp}(\mathbf{x})$ of $\mathbf{x}$ to the same degree that it enhances that of $\mathbf{y} = \mathbf{f}(\mathbf{x})$. In other words, the relative likelihood function on $\mathbf{y}$-space simply pulls back to the correct relative likelihood function on $\mathbf{x}$-space under $\mathbf{f}(\cdot)$. If several $\mathbf{x}$ are mapped to the same $\mathbf{y}$, their relative weightings are determined by the prior $\mathbf{dp}(\mathbf{x})$.

If $\mathbf{f}(\cdot)$ has unknown internal parameters μ, *i.e.* $\mathbf{y} = \mathbf{f}(\mathbf{x}, \mu)$, the data space $\mathbf{x}$ can be extended to include these and the above $\mathbf{dp}(\mathbf{x}|\cdot)$ factors become $\mathbf{dp}(\mathbf{x},\mu|\cdot)$. Integrating over all possible values of $\mathbf{x}$ and applying the conditional probability definition $\mathbf{dp}(\mathbf{x},\mu) = \mathbf{dp}(\mathbf{x}|\mu)\cdot\mathbf{dp}(\mu)$ gives

$$\frac{\mathbf{dp}(\mu|E)}{\mathbf{dp}(\mu)} = \int_{\mathbf{x}} \frac{\mathbf{dp}(\mathbf{x},\mu|E)}{\mathbf{dp}(\mu)} = \int_{\mathbf{x}} \mathbf{dp}(\mathbf{x}|\mu)\cdot\frac{\mathbf{dp}(\mathbf{x},\mu|E)}{\mathbf{dp}(\mathbf{x},\mu)}$$
$$= \int_{\mathbf{x}} \mathbf{dp}(\mathbf{x}|\mu)\cdot\left.\frac{\mathbf{dp}(\mathbf{y}|E)}{\mathbf{dp}(\mathbf{y})}\right|_{\mathbf{y}\,=\,\mathbf{f}(\mathbf{x},\mu)}$$

This says that the posterior likelihood for μ is proportional to the total probability for *any* corresponding $\mathbf{x}$ to give the observation via $\mathbf{y} = \mathbf{f}(\mathbf{x},\mu)$. In other words the log-unlikelihood of μ given E is proportional to the logarithmic 'shift factor' required to normalize the distribution of $\mathbf{x}$ given μ and E.

The above analysis applies directly to a projective mapping $\mathbf{x}^a \to \mathbf{y}^A = \mathbf{P}^A_a\mathbf{x}^a$ between projective spaces $\mathcal{P}^a$ and $\mathcal{P}^A$. If we assume that the relative likelihood on $\mathcal{P}^A$ can be approximated by a biquadric $\{\mathbf{M}_{AB},\mathbf{N}_{AB}\}$ and that the prior on $\mathcal{P}^a$ is sufficiently 'uniform', the pulled back density on $\mathcal{P}^a$ is the biquadric

$$\mathbf{dp}(\mathbf{x}|\text{Evidence}) \;\approx\; \exp\left(-\frac{1}{2}\frac{(\mathbf{M}_{AB}\mathbf{P}^A_a\mathbf{P}^B_b)\mathbf{x}^a\mathbf{x}^b}{(\mathbf{N}_{CD}\mathbf{P}^C_c\mathbf{P}^D_d)\mathbf{x}^c\mathbf{x}^d}\right)\mathbf{dV}$$

In matrix notation, the information $\mathbf{M}$ and normalization $\mathbf{N}$ are pulled back respectively to $\mathbf{P}^\top\mathbf{M}\mathbf{P}$ and $\mathbf{P}^\top\mathbf{N}\mathbf{P}$. The preservation of the biquadric functional form under projective transformations implies that image space error models are directly pulled back to source space ones. However it should also be clear that there is little hope of obtaining commensurable distributions when combining observations pulled back from distinct image spaces $\mathcal{P}^{A_1},\ldots,\mathcal{P}^{A_k}$: the pulled-back normalizations $\mathbf{N}_{A_iB_i}\mathbf{P}^{A_i}_a\mathbf{P}^{B_i}_b$ will usually all be different.

In general the pulled-back $\mathbf{M}$ needs to be shifted by a multiple of the pulled-back $\mathbf{N}$ to produce a correctly normalized probability density on $\mathcal{P}^a$. The shift required is proportional to the logarithm of the total probability for *any* point in $\mathcal{P}^a$ to project to the observation, and hence depends on $\mathbf{P}^A_a$. As mentioned above, if the transformation is uncertain the posterior log-unlikelihood for a particular value $\mathbf{P}^A_a$ given the observation $\{\mathbf{M}_{AB},\mathbf{N}_{AB}\}$ is proportional to the shift $\lambda(\mathbf{P}^A_a)$ required to normalize the pulled-back distribution. In the next section we will use this to derive estimation techniques for uncertain projective subspaces, but for the remainder of this section we assume that $\mathbf{P}^A_a$ is a fixed known transformation.

Now let us examine the characteristics of the pulled-back distributions a little more closely. If $\mathbf{P}^A_a$ is a projective isomorphism — a nonsingular mapping between spaces of the same dimension, possibly from $\mathcal{P}^a$ to itself — its effect is analogous to that of a projective change of basis and there are no essentially new features.

If $\mathbf{P}^A_a$ is a nonsingular *injection* — *i.e.* a one-to-one mapping of $\mathcal{P}^a$ onto a projective subspace of $\mathcal{P}^A$ — the pulled-back likelihood is isomorphic to the restriction of the parent likelihood to the range subspace in $\mathcal{P}^A$. The only new feature is that the injected subspace may happen to 'miss' the mode of the parent distribution by a substantial margin, so that the pulled back likelihood has a shape and range of values much attenuated compared to those of the parent function on $\mathcal{P}^A$.

Finally, consider the case where $\mathbf{P}^A_a$ is a singular surjection onto a projective space of lower dimension. In this case each point of $\mathcal{P}^A$ has a nontrivial 'preimage' in $\mathcal{P}^a$ (*i.e.* the projective subspace of $\mathcal{P}^a$ that projects onto it), and $\mathcal{P}^a$ also necessarily contains a null subspace of points that project to nothing at all: $\mathbf{P}^A_a\mathbf{x}^a = 0$. The pulled-back likelihood is constant on each preimage space but is undefined on the null space as the pulled back $\mathbf{M}$ and $\mathbf{N}$ both vanish there. The pulled back equi-probability surfaces are degenerate quadrics with singularities on the null space, and generally look more like elliptical cones than ellipsoids.

The singular surjective situation occurs for the usual 3D$\to$2D perspective projection in computer vision. In that case the null space is the centre of projection, the preimage spaces are the optical rays, and the equi-probability surfaces — the sets of world points that are equally likely to have produced the given image measurement — are elliptical cones centred on the centre of projection and generated by the optical rays, that project to the experimental error ellipses in the image plane. The considerable representative power of the projective least squares framework is illustrated by its ability to deal with error models for perspective projection out-of-hand.

It was to accommodate surjective projections that we insisted on allowing $\mathbf{M}$ to be *semi*-definite. Note that the null space compatibility condition is maintained: if the null space of $\mathbf{M}_{AB}$ is a subset of that of $\mathbf{N}_{AB}$, the same is true of the pulled-back tensors $\mathbf{M}_{AB}\mathbf{P}^A_a\mathbf{P}^B_b$ and $\mathbf{N}_{AB}\mathbf{P}^A_a\mathbf{P}^B_b$. The normalization condition $\mathbf{N}_{AB}(\mathbf{M}^{-1})^{AB} = 1$ (with $\mathbf{M}^{-1}$ interpreted as a pseudo-inverse) is also preserved under surjective pull-backs, so the shift factor of $\mathbf{M}$ does not usually need to be corrected in this case.

6 Subspace Estimation

The results of the previous section can be used to develop projective least squares error models for projective subspaces. Given a number of uncertain points, we are interested in 'fitting' a projective subspace to them and estimating its uncertainty.

Suppose we have measured a single point $\mathbf{x}^a$, whose uncertainty is characterized by a biquadric distribution in $\mathbf{M}_{ab}$ and $\mathbf{N}_{ab}$. A k dimensional projective subspace in d dimensions can be specified by choosing a set of $k+1$ independent points that span it, *i.e.* by giving a $(d+1)\times(k+1)$ rank $k+1$ matrix $\mathbf{U}^a_A$ whose columns span the subspace ($A = 0,\ldots,k$). $\mathbf{U}^a_A$ can be thought of as a nonsingular projective injection from an abstract k dimensional projective space $\mathcal{P}^A$ to $\mathcal{P}^a$. As discussed in the previous section, if $\mathbf{U}$ is uncertain its relative likelihood given the observation $\{\mathbf{M},\mathbf{N}\}$ is proportional to the total probability in the subspace it generates, and hence to the total probability in the pulled back distribution on $\mathcal{P}^A$. In fact, up to an additive constant the log-unlikelihood of $\mathbf{U}$ given $\{\mathbf{M},\mathbf{N}\}$

is precisely the shift factor $\lambda(\mathbf{U}^\top\mathbf{M}\mathbf{U}, \mathbf{U}^\top\mathbf{N}\mathbf{U})$ required to normalize the pulled back distribution $\{\mathbf{U}^\top\mathbf{M}\mathbf{U}, \mathbf{U}^\top\mathbf{N}\mathbf{U}\}$:

$$1 + \chi^2(\mathbf{U}|\mathbf{M},\mathbf{N}) \overset{+\text{const}}{\approx} 1 + \lambda(\mathbf{U}^\top\mathbf{M}\mathbf{U}, \mathbf{U}^\top\mathbf{N}\mathbf{U})$$

At this point our approximate shift solution $1 + \lambda(\mathbf{M},\mathbf{N}) \approx \text{Trace}^{-1}(\mathbf{N}\mathbf{M}^{-1})$ comes into its own. Without a tractable analytic approximation to $\lambda(\mathbf{U}^\top\mathbf{M}\mathbf{U}, \mathbf{U}^\top\mathbf{N}\mathbf{U})$ it would be impossible to develop explicit methods for the least squares fitting of subspaces. The abstract theory would still exist, but there would be no closed-form formulae. Adopting this approximation we have the remarkably simple estimate

$$\begin{aligned} 1 + \chi^2(\mathbf{U}|\mathbf{M},\mathbf{N}) &\overset{+\text{const}}{\approx} \text{Trace}^{-1}\left(\mathbf{U}^\top\mathbf{N}\mathbf{U}\cdot(\mathbf{U}^\top\mathbf{M}\mathbf{U})^{-1}\right) \\ &= \text{Trace}^{-1}\left(\mathbf{N}\cdot\mathbf{U}(\mathbf{U}^\top\mathbf{M}\mathbf{U})^{-1}\mathbf{U}^\top\right) \end{aligned}$$

Note the invariance of this formula under redefinitions $\mathbf{U} \to \mathbf{U}\mathbf{A}$ of the spanning basis of the subspace, where $\mathbf{A}$ is any nonsingular $(k+1)\times(k+1)$ matrix.

Dually, a subspace can be specified as the intersection of $d-k$ hyperplanes, *i.e.* by a $(d-k)\times(d+1)$ rank $d-k$ matrix $\mathbf{W}_a^C$ that determines a set of $d-k$ independent homogeneous linear equations $\mathbf{W}_a^C\mathbf{x}^a = 0$ ($C = k+1,\ldots,d+1$). $\mathbf{W}$ and $\mathbf{U}$ specify the same subspace if and only if $\mathbf{W}\mathbf{U} = 0$ and the $(d+1)\times(d+1)$ matrix $\binom{\mathbf{U}^\top}{\mathbf{W}}$ is nonsingular. For any such pair $\{\mathbf{U},\mathbf{W}\}$ and any nonsingular symmetric $(d+1)\times(d+1)$ matrix $\mathbf{X}$ we have the standard decomposition

$$\mathbf{X} = \mathbf{U}\left(\mathbf{U}^\top\mathbf{X}^{-1}\mathbf{U}\right)^{-1}\mathbf{U}^\top + \mathbf{X}\mathbf{W}^\top\left(\mathbf{W}\mathbf{X}\mathbf{W}^\top\right)^{-1}\mathbf{W}\mathbf{X}$$

Applying this at the covariance $\mathbf{X} \equiv \mathbf{M}^{-1}$ gives the approximate log-unlikelihood of the subspace in terms of dual coordinates

$$\begin{aligned} &1 + \chi^2(\mathbf{W}|\mathbf{X},\mathbf{N}) \\ &\qquad \overset{+\text{const}}{\approx} \text{Trace}^{-1}\left[\mathbf{N}\cdot\left(\mathbf{X} - \mathbf{X}\mathbf{W}^\top(\mathbf{W}\mathbf{X}\mathbf{W}^\top)^{-1}\mathbf{W}\mathbf{X}\right)\right] \end{aligned}$$

Since $\mathbf{X} = \mathbf{M}^{-1}$ is normalized, the leading term is just $\text{Trace}(\mathbf{N}\cdot\mathbf{X}) = 1$.

6.1 Affine Limit

In the affine case the approximate shift formula is exact and the biquadric distributions become Gaussians, so the projective error model reduces to the standard affine one. Making the standard decompositions

$$\mathbf{M} \equiv \begin{pmatrix} 1 + \bar{\mathbf{x}}^\top\mathbf{X}^{-1}\bar{\mathbf{x}} & -\bar{\mathbf{x}}^\top\mathbf{X}^{-1} \\ -\mathbf{X}^{-1}\bar{\mathbf{x}} & \mathbf{X}^{-1} \end{pmatrix}, \qquad \mathbf{N} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \mathbf{U}\begin{pmatrix} 1 \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{A}\mathbf{y} + \mathbf{b} \end{pmatrix}, \quad \mathbf{U} \equiv \begin{pmatrix} 1 & 0 \\ \mathbf{b}^\top & \mathbf{A} \end{pmatrix}$$

we have

$$\mathbf{U}^\top\mathbf{M}\mathbf{U} = \begin{pmatrix} 1 + (\bar{\mathbf{x}}-\mathbf{b})^\top\mathbf{X}^{-1}(\bar{\mathbf{x}}-\mathbf{b}) & -(\bar{\mathbf{x}}-\mathbf{b})^\top\mathbf{X}^{-1}\mathbf{A} \\ \mathbf{A}^\top\mathbf{X}^{-1}(\bar{\mathbf{x}}-\mathbf{b}) & \mathbf{A}^\top\mathbf{X}^{-1}\mathbf{A} \end{pmatrix}$$

$$\mathbf{U}^\top\mathbf{N}\mathbf{U} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Using the fact that an incorrectly shifted affine information tensor has an inverse with 00 coefficient $1/(1+\lambda)$

$$\begin{aligned} &\begin{pmatrix} 1 + \lambda + \bar{\mathbf{x}}^\top\mathbf{X}^{-1}\bar{\mathbf{x}} & -\bar{\mathbf{x}}^\top\mathbf{X}^{-1} \\ -\mathbf{X}^{-1}\bar{\mathbf{x}} & \mathbf{X}^{-1} \end{pmatrix}^{-1} \\ &\qquad = \frac{1}{1+\lambda}\begin{pmatrix} 1 \\ \bar{\mathbf{x}} \end{pmatrix}\begin{pmatrix} 1 & \bar{\mathbf{x}}^\top \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{X} \end{pmatrix} \end{aligned}$$

a short calculation gives

$$\begin{aligned} &\chi^2(\mathbf{U}|\mathbf{X},\bar{\mathbf{x}}) \overset{+\text{const}}{=} \\ &\quad (\bar{\mathbf{x}}-\mathbf{b})^\top\left(\mathbf{X}^{-1} - \mathbf{X}^{-1}\mathbf{A}(\mathbf{A}^\top\mathbf{X}^{-1}\mathbf{A})^{-1}\mathbf{A}^\top\mathbf{X}^{-1}\right)(\bar{\mathbf{x}}-\mathbf{b}) \end{aligned}$$

This is the standard affine formula for the log-unlikelihood of an affine subspace $\mathbf{A}\mathbf{y} + \mathbf{b}$ given an uncertain observation of a point on it. The matrix vanishes on vectors $\mathbf{A}\mathbf{y}$ in the subspace and hence measures the 'orthogonal Mahalanobis distance' of the mean $\bar{\mathbf{x}}$ from the subspace.

In terms of dual coordinates the affine subspace is

$$\mathbf{W}\begin{pmatrix} 1 \\ \mathbf{x} \end{pmatrix} = \mathbf{D}\mathbf{x} + \mathbf{c} = 0 \qquad\qquad \mathbf{W} \equiv \begin{pmatrix} \mathbf{c} & \mathbf{D} \end{pmatrix}$$

where $\mathbf{D}\mathbf{A} = 0$ and $\mathbf{c} = -\mathbf{D}\mathbf{b}$. In this case the affine log-unlikelihood is simply

$$\chi^2(\mathbf{W}|\mathbf{X},\bar{\mathbf{x}}) \overset{+\text{const}}{=} (\mathbf{D}\bar{\mathbf{x}} + \mathbf{c})^\top\left(\mathbf{D}\mathbf{X}\mathbf{D}^\top\right)^{-1}(\mathbf{D}\bar{\mathbf{x}} + \mathbf{c})$$

This is easily verified from the non-dual-form affine log-unlikelihood given above, or with a little more effort from the projective dual-form log-unlikelihood. Basically, it says that the information in constraint violation space is measured by the inverse of the classical constraint covariance matrix.

6.2 Grassmann Coordinates

We promised that projective least squares would look natural in Grassmann coordinates, and now we verify this. The k dimensional projective subspace spanned by the column vectors of $\mathbf{U}_A^a$ has Grassmann coordinates [2, 7]

$$\mathbf{u}^{[a_0\cdots a_k]} \;\sim\; \mathbf{U}_{A_0}^{a_0}\cdots\mathbf{U}_{A_k}^{a_k}\,\varepsilon^{A_0\cdots A_k}$$

Alternatively, a k dimensional subspace can be specified by $d-k$ linear constraints $\mathbf{W}_a^C\mathbf{x}^a = 0$ (the rows of the matrix $\mathbf{W}$, labelled by $C = k+1,\ldots,d+1$) to give dual Grassmann coordinates

$$\mathbf{w}_{[a_{k+1}\cdots a_d]} \;\sim\; \varepsilon_{C_{k+1}\cdots C_d}\,\mathbf{W}_{a_{k+1}}^{C_{k+1}}\cdots\mathbf{W}_{a_d}^{C_d}$$

Here, $\mathbf{u}^{a_0\cdots a_k}$ and $\mathbf{w}_{a_{k+1}\cdots a_d}$ are respectively the $(k+1)\times(k+1)$ minors of $\mathbf{U}$ and the $(d-k)\times(d-k)$ minors of $\mathbf{W}$. They are only defined up to scale, and if $\mathbf{U}$ and $\mathbf{W}$ specify the same subspace they are tensor duals of one another.

The subspace log-unlikelihood $1 + \chi^2(\mathbf{U}|\mathbf{M},\mathbf{N}) \approx \text{Trace}^{-1}\left(\mathbf{N}\cdot\mathbf{U}(\mathbf{U}^\top\mathbf{M}\mathbf{U})^{-1}\mathbf{U}^\top\right)$ can be rewritten in terms of the Grassmann coordinates $\mathbf{u}^{a_0\cdots a_k}$ by expanding the inverse $(\mathbf{U}^\top\mathbf{M}\mathbf{U})^{-1}$ by cofactors and rearranging. The result is

$$1+\chi^2(\mathbf{u}|\mathbf{M},\mathbf{N}) \overset{+\text{const}}{\approx} \frac{\mathbf{M}_{a_0b_0}\cdots\mathbf{M}_{a_kb_k}\cdot\mathbf{u}^{a_0\cdots a_k}\;\mathbf{u}^{b_0\cdots b_k}}{(k+1)\,\mathbf{N}_{c_0d_0}\mathbf{M}_{c_1d_1}\cdots\mathbf{M}_{c_kd_k}\cdot\mathbf{u}^{c_0c_1\cdots c_k}\;\mathbf{u}^{d_0d_1\cdots d_k}}$$

Once again we recognize the familiar form of the biquadric, this time in the Grassmann coordinates $\mathbf{u}^{a_0\cdots a_k}$ rather than the point coordinates $\mathbf{x}^a$, with information[1] $\mathbf{M}_{[\![a_0b_0}\cdots\mathbf{M}_{a_kb_k]\!]}$ and normalization $(k+1)\cdot\mathbf{N}_{[\![a_0b_0}\mathbf{M}_{a_1b_1}\cdots\mathbf{M}_{a_kb_k]\!]}$. If $k=0$ we get back the original point distribution, as would be expected.

The space of k dimensional projective subspaces in d dimensions is locally parameterized by $(d-k)\times(k+1)$ matrices and therefore has dimension $(d-k)(k+1)$. The Grassmann coordinatization embeds it as a projective subvariety of the $\binom{d+1}{k+1}$ dimensional homogeneous space $\mathcal{P}^{[a_0\cdots a_k]}$ of $k+1$ index skew tensors. The constraint equations that determine this subvariety are the quadratic **Grassmann simplicity constraints**

$$\mathbf{u}^{a_0\cdots a_{k-1}[a_k}\mathbf{u}^{b_0\cdots b_k]}=0$$

Hence, although the Grassmann coordinates $\mathbf{u}^{a_0\cdots a_k}$ are linearly independent, they are quadratically highly redundant.

The subspace information and normalization tensors can be viewed as symmetric matrices on the large $\binom{d+1}{k+1}$ dimensional space $\mathcal{P}^{[a_0\cdots a_k]}$. They are nonsingular whenever the underlying $\mathbf{M}_{ab}$ and $\mathbf{N}_{ab}$ are, however there are linear (non-matricial) relations among their components that enforce the Grassmann simplicity constraints. Any product of symmetric tensors of the form $\mathbf{T}^{[a_0a_1\cdots a_k]\cdot[b_0b_1\cdots b_k]}\equiv\mathbf{M}_0^{[\![a_0b_0}\mathbf{M}_1^{a_1b_1}\cdots\mathbf{M}_k^{a_kb_k]\!]}$ is 'simple' in the sense that $\mathbf{T}^{a_0\cdots a_{k-1}[a_k\,\cdot\,b_0\cdots b_k]}=0$ because the antisymmetrization always includes a pair of symmetric indices. A biquadric built with such 'simple' Grassmann tensors "projects on to the simple part of $\mathbf{u}^{a_0\cdots a_k}$" in the sense that it is insensitive to the 'non-simple part' $\mathbf{u}^{a_0\cdots a_{k-1}[a_k}\mathbf{u}^{b_0\cdots b_k]}\neq 0$.

A similar process can be applied to the dual-form matricial log-unlikelihood $1+\chi^2(\mathbf{W}|\mathbf{X},\mathbf{N})$ given above, to derive the dual Grassmann log-unlikelihood

$$1+\chi^2(\mathbf{w}|\mathbf{X},\mathbf{N}) \overset{+\text{const}}{\approx} \frac{\mathbf{X}^{a_{k+1}b_{k+1}}\cdots\mathbf{X}^{a_db_d}\cdot\mathbf{w}_{a_{k+1}\cdots a_d}\;\mathbf{w}_{b_{k+1}\cdots b_d}}{\big(\mathbf{X}-(d-k)\mathbf{XNX}\big)^{c_{k+1}d_{k+1}}\mathbf{X}^{c_{k+2}d_{k+2}}\cdots\mathbf{X}^{c_dd_d}\cdot\mathbf{w}_{c_{k+1}\cdots c_d}\;\mathbf{w}_{d_{k+1}\cdots d_d}}$$

where $\mathbf{X}\equiv\mathbf{M}^{-1}$ and $\mathbf{N}_{ab}\mathbf{X}^{ab}=1$. Once again the log-unlikelihood has the biquadric form, this time in the dual coordinates $\mathbf{w}_{a_{k+1}\cdots a_d}$. The information and normalization tensors are again 'simple' in the Grassmann sense. This can also be derived by tensor dualization of the contravariant Grassmann formula. Note that in the affine case $\mathbf{XNX}=\binom{1}{\bar{\mathbf{x}}}(1\;\;\bar{\mathbf{x}}^\top)$.

[1] For convenience we introduce the notation $[\![a_0b_0a_1b_1\cdots a_kb_k]\!]$ to denote $[a_0a_1\cdots a_k][b_0b_1\cdots b_k]$ on the index pairs a_ib_i of a set of 2 index tensors, *i.e.* antisymmetrize separately over the first indices and the second indices of the pairs.

6.3 Hyperplanes

Hyperplanes (codimension one subspaces) are a particularly important special case of the above. The log-unlikelihood for the location of a hyperplane $\mathbf{w}_a\mathbf{x}^a=0$ given an uncertain point on it follows immediately from the above dual-form matrix or Grassmann formulae:

$$1+\chi^2(\mathbf{w}_a|\mathbf{X},\mathbf{N}) \overset{+\text{const}}{\approx} \frac{\mathbf{X}^{ab}\,\mathbf{w}_a\mathbf{w}_b}{(\mathbf{X}-\mathbf{XNX})^{cd}\,\mathbf{w}_c\mathbf{w}_d}$$

Dually to the point case, the log-unlikelihood is a biquadric in the hyperplane coordinates. For an affine distribution this becomes

$$\chi^2(\mathbf{w}_a|\mathbf{X},\mathbf{N}) \overset{+\text{const}}{=} \frac{\left(\mathbf{d}^\top\bar{\mathbf{x}}+c\right)^2}{\mathbf{d}^\top\bar{\mathbf{X}}\mathbf{d}}$$

where $\mathbf{w}_a$ is $(c\;\;\mathbf{d}^\top)$, and $\bar{\mathbf{x}}$ and $\bar{\mathbf{X}}$ are the classical mean and covariance.

The denominator plays a much more active rôle in hyperplane and k-subspace estimation than it did in the point fitting problem. Let us examine the hyperplane case a little more closely to find out why.

First of all, there is nothing intrinsically wrong with hyperplane distributions with 'simple' normalizers $\overline{\mathbf{N}}^{ab}$. It is just that in the case of point-plane fitting the correct answer can not be quite so simple. Consider a hyperplane distribution with a 'slowly varying' denominator $\overline{\mathbf{N}}^{ab}\mathbf{w}_a\mathbf{w}_b$. For example, $\overline{\mathbf{N}}^{ab}$ could be the $(d+1)\times(d+1)$ unit matrix in some basis, or the affine hyperplane normalizer $\begin{pmatrix}0&0\\0&\mathbf{I}\end{pmatrix}$, where $\mathbf{I}$ is the $d\times d$ unit matrix[2]. If the plane passes exactly through the mode of the point distribution, we would expect its likelihood to depend only weakly on its orientation: any plane passing right through the observation should be about equally good as far as the least squares error is concerned. Since the denominator was chosen to be almost independent of orientation, the numerator must also depend only weakly on orientation. But this implies that the rate of decay of the likelihood as the plane moves away from the point is *also* independent of orientation: the only remaining parameter is a direction-independent scalar peak width. However in general the point distribution is not spherically symmetric and the rate of decay of the plane distribution ought to be different in different directions. In summary, it is not possible to have all three of: *(i)* an isotropic likelihood *at* the observation; *(ii)* an anisotropic decay *away from* the observation; *(iii)* an isotropic normalizer $\overline{\mathbf{N}}^{ab}$. The first two are essential to represent the data correctly, so we are forced to deal with non-isotropic normalizers $\mathbf{N}^{ab}$ and hence (if the plane is being fitted to several points) incommensurable distributions. This is not simply a problem with the projective

[2] This gives the conventional normalization for Euclidean hyperplanes, with a constant offset and a unit direction vector.

theory: classical affine least squares also gives incommensurable distributions for subspace fitting (*e.g.* orthogonal regression). In fact, the projective point of view makes the situation clearer by unifying the classically separate theories of point and plane fitting.

6.4 Normalization & Covariance

The above formulae for subspace log-unlikelihoods are only correct up to an additive constant. The modes (maximum likelihood values) of the subspace distributions can be found directly from the unshifted information tensors, but if subspace covariances are required the correct shift factors must be estimated.

In fact, it is straightforward to show that the normalization sum $\mathrm{Trace}(\mathbf{N}\cdot\mathbf{M}^{-1})$ for the subspace-fitted-to-point distributions is always $\binom{d}{k}$ instead of 1. The reason is simply that even when the point distribution is narrow the resulting subspace distribution always has a $\binom{d}{k}$ dimensional modal (*i.e.* maximum likelihood) subspace in $\mathcal{P}^{[a_0\cdots a_k]}$, corresponding to the $\binom{d}{k}$ different 'directions' in which the k-subspace can pass right through the centre of the point distribution[3].

Since $\mathrm{Trace}(\mathbf{N}\cdot\mathbf{M}^{-1}) = \binom{d}{k}$ for the $\binom{d+1}{k+1}$ dimensional subspace information tensor, the approximate shift equation predicts a normalizing shift of $\mathbf{M} \to \mathbf{M} + (\binom{d}{k} - 1)\cdot\mathbf{N}$. However this approximation is not recommended as it is likely to be quite inaccurate for such large shift factors. On the other hand, whenever subspace-through-point likelihoods from several points are combined, the resulting distribution tends to be much better localized because the null directions from different points tend to cancel each other out, leaving a single reasonably well defined mode. In this case (and modulo the usual correction for the accumulation of the '1's in the $(1+\chi^2)$'s) the shift factor required to normalize the combined subspace distribution tends to be much smaller.

The $\binom{d+1}{k+1}$ dimensional Grassmann parameterization is global but redundant and it is often convenient to re-express the mode and covariance in terms of some minimal local parameterization, say $\mathbf{z}^\alpha$ where $\alpha = 1,\ldots,(d-k)(k+1)$. Given Grassmann information and normalization matrices $\mathbf{M}$ and $\mathbf{N}$, the Grassmann mode can be found by the usual minimum eigenvector procedure. The expression $\mathbf{u}^{a_0\cdots a_k}(\mathbf{z}^\alpha)$ for the Grassmann parameterization in terms of $\mathbf{z}^\alpha$ must then be inverted at the Grassmann mode to find the $\mathbf{z}^\alpha$-space mode. (This may require the solution of nonlinear equations). Finally, the $\mathbf{z}^\alpha$-space information matrix can be found by evaluating the second derivatives of $1+\chi^2\left(\mathbf{u}^{a_0\cdots a_k}(\mathbf{z}^\alpha)|\mathbf{M},\mathbf{N}\right)$ at the $\mathbf{z}^\alpha$-space mode.

[3] The 'directions' of the modal subspace are generated by the $\binom{d}{k}$ choices of k directions $\mathbf{u}_1^a,\ldots,\mathbf{u}_k^a$ among the d in any hyperplane not passing through the point mode $\hat{\mathbf{x}}^a$. The corresponding k-subspace is the span $\hat{\mathbf{x}}^{[a_0}\mathbf{u}_1^{a_1}\cdots\mathbf{u}_k^{a_k]}$. The $\binom{d}{k}$ dimensional modal subspace intersects the $(d-k)(k-1)$ dimensional Grassmann variety of k-subspaces (*i.e.* simple tensors) in the $(d-k)(k-2)$ dimensional variety of k-subspaces through $\hat{\mathbf{x}}^a$.

7 Fundamental Matrix Estimation

As another example of the use of projective least squares, consider the problem [3, 1] of estimating the fundamental matrix between two images from a set of corresponding point pairs. Given any two 2D projective images $\mathcal{P}^A$ and $\mathcal{P}^{A'}$ of a 3D scene taken from different positions, a pair $\{\mathbf{x}^A,\mathbf{x}^{A'}\}$ of image points corresponds to some 3D point if and only if the **epipolar constraint** $\mathbf{F}_{AA'}\,\mathbf{x}^A\mathbf{x}^{A'} = 0$ is satisfied, where $\mathbf{F}_{AA'}$ is the 3×3 rank 2 **fundamental matrix**. The point $\mathbf{x}^A$ gives rise to a corresponding **epipolar line** $\mathbf{F}_{AA'}\mathbf{x}^A$ in the opposite image $\mathcal{P}^{A'}$ and all potentially matching $\mathbf{x}^{A'}$ lie on this line. The epipolar lines all pass through a point called the **epipole** $\mathbf{e}^{A'}$. This is the image of the projection centre of the opposite camera and satisfies $\mathbf{F}_{AA'}\mathbf{e}^{A'} = 0$. Similarly for $\mathbf{x}^{A'}$, $\mathbf{F}_{AA'}\mathbf{x}^{A'}$ and $\mathbf{e}^A$.

We can estimate $\mathbf{F}$ from a set of corresponding uncertain point pairs by viewing the epipolar constraint from each pair as a single linear constraint on $\mathbf{F}$. Intuitively, the smaller the deviations $|\mathbf{F}_{AA'}\mathbf{x}^A\mathbf{x}^{A'}|$ are, the better the fit will be, but we want to make this into a more rigorous approximate maximum likelihood estimate. The situation is analogous to that of hyperplane estimation: $\mathbf{F}_{AA'}$ can be viewed as defining a projective hyperplane in the $3\times 3 - 1 = 8$ dimensional projective space of tensors $\mathcal{P}^{AA'}$, and the data can be mapped bilinearly into this space via $\{\mathbf{x}^A,\mathbf{x}^{A'}\}\to\mathbf{x}^A\mathbf{x}^{A'}$. In fact, it turns out that we can re-use our projective least squares equations for hyperplanes.

Suppose that the uncertainties in the positions of $\mathbf{x}^A$ and $\mathbf{x}^{A'}$ can be modelled by independent normalized biquadric distributions $\{\mathbf{M}_{AB},\mathbf{N}_{AB}\}$ and $\{\mathbf{M}_{A'B'},\mathbf{N}_{A'B'}\}$ with covariances $\mathbf{X}^{AB} = (\mathbf{M}^{-1})^{AB}$ and $\mathbf{X}^{A'B'} = (\mathbf{M}^{-1})^{A'B'}$. Since the distributions are independent their moments can be factorized. In particular

$$\begin{aligned}
&\left\langle \frac{(\mathbf{x}^A\mathbf{x}^{A'})(\mathbf{x}^B\mathbf{x}^{B'})}{(\mathbf{N}_{CD}\mathbf{N}_{C'D'})(\mathbf{x}^C\mathbf{x}^{C'})(\mathbf{x}^D\mathbf{x}^{D'})}\right\rangle \\
&\qquad = \left\langle \frac{\mathbf{x}^A\mathbf{x}^B}{\mathbf{N}_{CD}\mathbf{x}^C\mathbf{x}^D}\cdot\frac{\mathbf{x}^{A'}\mathbf{x}^{B'}}{\mathbf{N}_{C'D'}\mathbf{x}^{C'}\mathbf{x}^{D'}}\right\rangle \\
&\qquad = \left\langle \frac{\mathbf{x}^A\mathbf{x}^B}{\mathbf{N}_{CD}\mathbf{x}^C\mathbf{x}^D}\right\rangle\cdot\left\langle\frac{\mathbf{x}^{A'}\mathbf{x}^{B'}}{\mathbf{N}_{C'D'}\mathbf{x}^{C'}\mathbf{x}^{D'}}\right\rangle \\
&\qquad = \mathbf{X}^{AB}\,\mathbf{X}^{A'B'}
\end{aligned}$$

Viewing $\mathbf{M}_{AB}\mathbf{M}_{A'B'}$, $\mathbf{N}_{AB}\mathbf{N}_{A'B'}$ and $\mathbf{X}^{AB}\mathbf{X}^{A'B'}$ as 9×9 homogeneous symmetric matrices on the 8 dimensional projective space $\mathcal{P}^{AA'}$, we have $\mathbf{N}_{AB}\mathbf{N}_{A'B'}\cdot\mathbf{X}^{AB}\mathbf{X}^{A'B'} = 1$ and (since $\mathbf{X}^{AB}\mathbf{X}^{A'B'}\cdot\mathbf{M}_{BC}\mathbf{M}_{B'C'} = \delta_C^A\delta_{C'}^{A'}$ is the identity operator on $\mathcal{P}^{AA'}$) $\mathbf{X}^{AB}\mathbf{X}^{A'B'} = (\mathbf{M}_{AB}\mathbf{M}_{A'B'})^{-1}$. So rather remarkably, $\mathbf{M}_{AB}\mathbf{M}_{A'B'}$ and $\mathbf{N}_{AB}\mathbf{N}_{A'B'}$ define a correctly shifted biquadric distribution with covariance $\mathbf{X}^{AB}\mathbf{X}^{A'B'}$ on $\mathcal{P}^{AA'}$, that correctly models the uncertainty of the tensor-product image point $\mathbf{x}^A\mathbf{x}^{A'}$ to second order. This is notwithstanding the fact that the space of all possible $\mathbf{x}^A\mathbf{x}^{A'}$ is only a 4 dimensional quadratic subvariety of the 8 dimensional projective tensor space $\mathcal{P}^{AA'}$. Since the epipolar constraint $\mathbf{F}_{AA'}\mathbf{x}^A\mathbf{x}^{A'} = 0$ defines a projective hyperplane in $\mathcal{P}^{AA'}$ and we know how to fit projective hyperplanes to points, we can immediately write down the log-unlikelihood of $\mathbf{F}$ given $\mathbf{x}^A$ and $\mathbf{x}^{A'}$:

$$1 + \chi^2(\mathbf{F}_{AA'}|\mathbf{x}^A, \mathbf{x}^{A'}) \overset{+\text{const}}{\approx} \frac{\mathbf{X}^{AB}\mathbf{X}^{A'B'}\,\mathbf{F}_{AA'}\mathbf{F}_{BB'}}{\left(\mathbf{X}^{CD}\mathbf{X}^{C'D'} - (\mathbf{XNX})^{CD}(\mathbf{X'N'X'})^{C'D'}\right)\mathbf{F}_{CC'}\mathbf{F}_{DD'}}$$

Writing the information and normalization tensors as 9×9 symmetric matrices on the 9 dimensional space of components of $\mathbf{F}$, the biquadric log-unlikelihoods for different point pairs can be combined in the usual way. As in the hyperplane case, they are always incommensurable so nonlinear techniques are required.

If both of the points have affine distributions, converting to 3×3 matrix notation and denoting the homogeneous mean $\binom{1}{\bar{\mathbf{x}}}$ by $\hat{\mathbf{x}}$ and the homogeneous affine covariance $\begin{pmatrix} 0 & 0 \\ 0 & \bar{\mathbf{X}} \end{pmatrix}$ by $\hat{\mathbf{X}}$, we can re-express this as follows:

$$\chi^2(\mathbf{F}|\hat{\mathbf{x}}, \hat{\mathbf{X}}, \hat{\mathbf{x}}', \hat{\mathbf{X}}') \overset{+\text{const}}{=} \frac{(\hat{\mathbf{x}}^\top\mathbf{F}\hat{\mathbf{x}}')^2}{\hat{\mathbf{x}}^\top\mathbf{F}\hat{\mathbf{X}}'\mathbf{F}^\top\hat{\mathbf{x}} + \hat{\mathbf{x}}'^\top\mathbf{F}^\top\hat{\mathbf{X}}\mathbf{F}\mathbf{x}' + \mathrm{Trace}(\mathbf{F}\hat{\mathbf{X}}'\mathbf{F}^\top\hat{\mathbf{X}})}$$

This formula can also be derived by classical maximum likelihood calculations. The term $\mathrm{Trace}\big(\mathbf{F}\hat{\mathbf{X}}'\mathbf{F}^\top\hat{\mathbf{X}}\big)$ is second order in the uncertainty and is often ignored relative to the first order terms: with this approximation the formula has been used for nonlinear estimation of the fundamental matrix with good results [3]. Roughly, it says that the 'primitive' error measure $(\hat{\mathbf{x}}'^\top\mathbf{F}\hat{\mathbf{x}})^2$ needs to be normalized by dividing by the sum of the variance of each measured point orthogonal to the opposite epipolar line. When one or both of the measured points lie near an epipole, the second order trace term is sometimes significant relative to the other terms and tends to have a stabilizing effect on the fit, so it should probably not be omitted if the epipoles lie within the images (*e.g.* frontal motion).

8 Discussion & Future Work

The results we have presented are obviously still at the theoretical level and it remains to be seen how useful projective least squares will turn out to be in practice. However, it is becoming clear that error modelling will become a central issue in visual reconstruction, not only to ensure the accuracy of the final results, but also because the efficiency of intermediate stages such as correspondence and database indexing depends critically on the uncertainties involved. Given that projective least squares is both 'projectively correct' and relatively tractable (notwithstanding the length of some of the equations we have written), it seems likely that it will have a part to play in all this.

On the technical level there are still many loose ends. Analytical work is needed to clarify the status of the two approximations made in deriving the basic error model, and the development of a 'central moment expansion' based on the homogeneous Taylor series could be mathematically fruitful. More practically it would be useful to have projective least squares methods for quadrics and higher order projective varieties, and for further types of subspace-subspace intersection and union (*e.g.* intersection of subspaces at a point). It is also unclear how to extend the fundamental matrix estimation model to the trilinear and quadrilinear constraints that exist when there are additional images [6, 7, 9]. Although the relation between the multilinear data tensors $\mathbf{x}^{A_i}\mathbf{x}^{A_j}\cdots\mathbf{x}^{A_k}$ and the corresponding constraint tensor is still linear, it is no longer a simple scalar and it is not yet clear how to capture it correctly in a projective least squares error model.

Acknowledgments

This work was supported by the European Community through Esprit programs HCM and SECOND.

References

[1] O. Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.

[2] W. V. D. Hodge and D. Pedoe. *Methods of Algebraic Geometry*, volume 1. Cambridge University Press, 1947.

[3] Q.-T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo. On determining the fundamental matrix: Analysis of different methods and experimental results. Technical Report RR-1894, INRIA, Sophia Antipolis, France, 1993.

[4] P. F. McLauchlan and D. W. Murray. A unifying framework for structure and motion recovery from image sequences. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, Cambridge, MA, June 1995.

[5] R. Mohr, B. Boufama, and P. Brand. Accurate relative positioning from multiple images. Technical Report RT-102, LIFIA, INRIA Rhône-Alpes, Grenoble, France, 1993. Submitted to *Journal of Artificial Intelligence*.

[6] A. Shashua. Algebraic functions for recognition. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 1994.

[7] B. Triggs. The geometry of projective reconstruction I: Matching constraints and the joint image. Submitted to *Int. J. Computer Vision*, 1995.

[8] B. Triggs. Least squares estimation in projective spaces. To appear, 1995.

[9] B. Triggs. Matching constraints and the joint image. In E. Grimson, editor, *IEEE Int. Conf. Computer Vision*, Cambridge, MA, June 1995.

[10] Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysis*. Springer-Verlag, 1992.