Memoirs of the Faculty of Engineering, Okayama University, Vol. 45, pp. 15-26, January 2011

Hyper Least Squares and Its Applications

Kenichi KANATANI*, Prasanna RANGARAJAN†, Yasuyuki SUGAYA‡, and Hirotaka NIITSUMA*
∗ Department of Computer Science, Okayama University, Japan
† Department of Electrical Engineering, Southern Methodist University, U.S.A.
‡ Department of Computer Science and Engineering,
Toyohashi University of Technology, Japan

(Received November 15, 2010)

We present a new least squares (LS) estimator, called "HyperLS", specifically designed for parameter estimation in computer vision applications. It minimizes the algebraic distance under a special scale normalization, which is derived by rigorous error analysis in such a way that statistical bias is removed up to second order noise terms. Numerical experiments suggest that our HyperLS is far superior to the standard LS and comparable in accuracy to maximum likelihood (ML), which is known to produce highly accurate results in image applications but may fail to converge if poorly initialized. Our HyperLS is a perfect candidate for ML initialization. In addition, we discuss how image-based inference problems have different characteristics form conventional statistical applications, with a view to serving as a bridge between mathematicians and computer engineers.

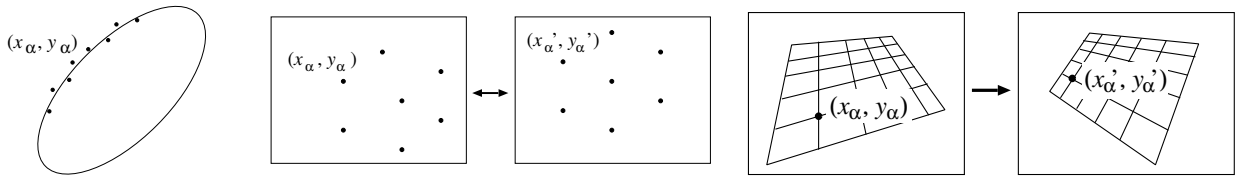
1. INTRODUCTION

We address a special class of parameter estimation problems that frequently arise in computer vision applications. The parameters are estimated from observations that should satisfy implicit polynomials in the absence of noise. The unique aspect of our approach is that we require the estimation accuracy to increase rapidly *as the noise level decreases* with a fixed number of observations. This requirement is a consequence of the fact that computer vision applications have very different characteristics from standard statistical domains. This issue is discussed in more detail later. We are hoping that this paper serves as a bridge between mathematicians and computer engineers.

An important task in computer vision is the extraction of 2-D/3-D geometric information from image data (Kanatani, 1996; Hartley and Zisserman, 2004). In this domain, maximum likelihood (ML) is known to produce highly accurate solutions, achieving the theoretical accuracy limit to a first approximation in the noise level (Kanatani, 1994; Chernov and Lesort, 2004; Kanatani, 2008). However, ML requires iterative search, which does not always converge unless started from a value sufficiently close to the solution. For this reason, various numerical schemes that produce reasonably accurate approximations have been extensively studied (Hartley and Zisserman, 2004). The simplest of such schemes is *algebraic distance minimization*, or simply *least squares* (LS), which minimizes the sum of the squares of polynomials that should be zero in the absence of noise. However, the accuracy of LS is very much limited. In this paper, we propose a new LS estimator, called *HyperLS*, by doing rigorous error analysis. The improved accuracy results from introduction of a normalization that eliminates statistical bias up to second order noise terms. Numerical experiments show that our HyperLS is far superior to the standard LS and is comparable in accuracy to ML. Thus, our HyperLS is a perfect candidate for initializing the ML iterations.

Section 2 defines the geometric fitting problem of our interest with examples that frequently appear in computer vision applications. Section 3 introduces a statistical model of observation that specifically suits image-based inference, discussing why the conventional statistical framework is not appropriate in image domains. In Sec. 4, we argue that ML is the best tool for image-based inference but that its computational cost is a major obstacle. Section 5 describes our framework of algebraic methods. In Sec. 6 and 7 we analyze errors of algebraic methods by perturbation techniques and in Sec. 8 derive expressions of covari-

*E-mail kanatani@suri.cs.okayama-u.ac.jp



(a) (b) (c)

Figure 1: (a) Fitting an ellipse to a point sequence. (b) Computing the fundamental matrix from corresponding points between two images. (c) Computing a homography between two images.

ance and bias of the solution, pointing out that we cannot control the covariance but can reduce the bias through the choice of scale normalization. In Sec. 9, we propose our HyperLS by choosing the scale normalization that eliminates the bias up to second order noise terms. In Sec. 10, we show examples of applying our HyperLS to typical computer vision problems. In Sec. 11, we conclude.

2. GEOMETRIC FITTING

The term "image data" in this paper refers to values extracted from images by image processing operations such as edge filters and interest point detectors. An example of image data includes the locations of points that have special characteristics in the images or the lines that separate image regions having different properties. We say that image data are "noisy" in the sense that image processing operations for detecting them entail uncertainty to some extent. Let $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$ be noisy image data, which we regard as perturbations in their true values $\bar{\boldsymbol{x}}_1$, ..., $\bar{\boldsymbol{x}}_N$ that satisfy implicit geometric constraints of the form

$$F^{(k)}(\boldsymbol{x};\boldsymbol{\theta}) = 0, \qquad k = 1, ..., L. \tag{1}$$

The unknown parameter $\boldsymbol{\theta}$ allows us to infer the 2-D/3-D shape and motion of the objects observed in the images (Kanatani, 1996, Hartley and Zisserman, 2004). In many important computer vision applications, we can reparameterize the problem to make the functions $F^{(k)}(\boldsymbol{x};\boldsymbol{\theta})$ linear in $\boldsymbol{\theta}$ (but nonlinear in $\boldsymbol{x}$), allowing us to write (1) as

$$(\boldsymbol{\xi}^{(k)}(\boldsymbol{x}), \boldsymbol{\theta}) = 0, \qquad k = 1, ..., L, \tag{2}$$

where and hereafter $(\boldsymbol{a}, \boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The vector $\boldsymbol{\xi}^{(k)}(\boldsymbol{x})$ represents a nonlinear mapping of $\boldsymbol{x}$.

Example 1 (Ellipse fitting). Given a point sequence (x_α, y_α), $\alpha = 1, ..., N$, we wish to fit an ellipse of the form

$$Ax^2 + 2Bxy + Cy^2 + 2(Dx + Ey) + F = 0. \tag{3}$$

(Fig. 1(a)). If we let

$$\begin{aligned} \boldsymbol{\xi} &= (x^2, 2xy, y^2, 2x, 2y, 1)^\top, \\ \boldsymbol{\theta} &= (A, B, C, D, E, F)^\top, \end{aligned} \tag{4}$$

the constraint (3) has the form of (2) with $L = 1$.

Example 2 (Fundamental matrix computation). Corresponding points (x, y) and (x', y') in two images of the same 3-D scene taken from different positions satisfy the *epipolar equation* (Hartley and Zisserman, 2004)

$$(\boldsymbol{x}, \boldsymbol{F}\boldsymbol{x}') = 0, \quad \boldsymbol{x} \equiv (x, , y, 1)^\top, \quad \boldsymbol{x}' \equiv (x', y', 1')^\top, \tag{5}$$

where $\boldsymbol{F}$ is called the *fundamental matrix*, from which we can compute the camera positions and the 3-D structure of the scene (Kanatani, 1996; Harley and Zisserman, 2004) (Fig. 1(b)). If we let

$$\begin{aligned} \boldsymbol{\xi} &= (xx', xy, x, yx', yy', y, x', y', 1)^\top, \\ \boldsymbol{\theta} &= (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^\top, \end{aligned} \tag{6}$$

the constraint (5) has the form of (2) with $L = 1$.

Example 3 (Homography computation). Two images of a planar or infinitely far away scene are related by a *homography* of the form

$$\boldsymbol{x}' \simeq \boldsymbol{H}\boldsymbol{x}, \quad \boldsymbol{x} \equiv (x, y, 1)^\top, \quad \boldsymbol{x}' \equiv (x', y', 1')^\top, \tag{7}$$

where $\boldsymbol{H}$ is a nonsingular matrix, and $\simeq$ denotes equality up to nonzero multiplier (Kanatani, 1996; Harley and Zisserman, 2004) (Fig. 1(c)). We can alternatively express (7) as the vector product equality

$$\boldsymbol{x}' \times \boldsymbol{H}\boldsymbol{x} = \boldsymbol{0}. \tag{8}$$

If we let

$$\begin{aligned} \boldsymbol{\xi}^{(1)} &= (0, 0, 0, -x, -y, -1, xy', yy', y')^\top, \\ \boldsymbol{\xi}^{(2)} &= (x, y, 1, 0, 0, 0, -xx', -yx', -x')^\top, \\ \boldsymbol{\xi}^{(3)} &= (-xy', -yy', -y', xx', yx', x', 0, 0, 0)^\top, \end{aligned} \tag{9}$$

$$\boldsymbol{\theta} = (H_{11}, H_{12}, H_{13}, H_{21}, H_{22}, H_{23}, H_{31}, H_{32}, H_{33})^\top, \tag{10}$$

the three component equations of (8) have the form of (2) with $L = 3$. Note that $\boldsymbol{\xi}^{(1)}$, $\boldsymbol{\xi}^{(2)}$, and $\boldsymbol{\xi}^{(3)}$ in (9) are linearly dependent; only two of them are independent.

3. STATISTICAL MODEL

For statistical analysis of the above problems, we need a statistical model of observation. We regard each datum $\boldsymbol{x}_\alpha$ as perturbed from its true value $\bar{\boldsymbol{x}}_\alpha$ by $\Delta\boldsymbol{x}_\alpha$, which we assume to be independent Gaussian variables of mean 0 and covariance matrix $V[\boldsymbol{x}_\alpha]$. We do not impose any restrictions on the true values $\bar{\boldsymbol{x}}_\alpha$ except that they should satisfy (1). This is known as a *functional* model. We could alternatively introduce some statistical model according to which the true values $\bar{\boldsymbol{x}}_\alpha$ are sampled. Then, the model is called *structural*. This distinction is crucial when we consider limiting processes in the following sense (Kanatani, 2008). Conventional statistical analysis mainly focuses on the asymptotic behavior as the number of observations increases to ∞. This is based on the reasoning that the mechanism underlying noisy observations would better reveal itself as the number of observations increases (the law of large numbers) while the number of available data is limited in practice. So, the estimation accuracy vs. the number of data is a major concern. In this light, efforts have been made to obtain a *consistent* estimator for fitting an ellipse to noisy data or computing the fundamental matrix from noisy point correspondences such that the solution approaches its true value in the limit $N \to \infty$ of the number N of points (Kukush et al., 2002; 2004).

However, a peculiar characteristic of image processing applications is that one cannot "repeat" observations. One makes an inference given a single set of images, and how many times one applies image processing operations, the result is always the same, because standard image processing algorithms are deterministic; no randomness is involved. This is in a stark contrast to conventional statistical problems where we view observations as sampled from potentially infinitely many possibilities and could obtain, by repeating observations, different values originating from unknown, uncontrollable, or unmodeled causes, which we call "noise" as a whole. In image-based applications, on the other hand, the accuracy of inference deteriorates as the uncertainty of image processing operations increases. Thus, the inference accuracy vs. the uncertainty of image operations, which we call "noise" for simplicity, is a major concern. Usually, the noise is very small, often subpixel (less than one pixel) levels. In light of this observation, it has been pointed out that in image domains the "consistency" of estimators should more appropriately be defined by the behavior in the limit $\sigma \to 0$ of the noise level σ (Chernov and Lesort, 2004; Kanatani 2008).

In this paper, we are interested in image processing applications and focus on the perturbation analysis for $\sigma \approx 0$ with a fixed number N of data. Thus, the functional model suits our purpose; if we were to analyze the error behavior in the limit of $N \to \infty$, a model that specifies how the data increase would be necessary. However, we cannot imagine how they change, because the image data $\boldsymbol{x}_\alpha$, $\alpha = 1, ..., N$, including the number N, is the *property of the images under consideration*. If we observe other images with different N, the properties of the images are different: They are different scenes. Thus, the asymptotic analysis for $N \to \infty$ does not have much sense in image-based applications. This is one of the reasons why ML is regarded as the best tool for image-based estimation. Practical experience of image applications suggests that ML almost always produces a desirable solution. If one could arbitrarily increase the number N of data, as in laboratory experiments, estimation of the true values $\bar{\boldsymbol{x}}_\alpha$, called *nuisance parameters* when viewed as parameters, would not be consistent in the ML framework, as pointed out by Neyman and Scott (1948) as early as in 1948. However, the lack of consistency, which some statisticians view as a drawback of ML, has no realistic meaning in computer vision applications. On the contrary, ML has very desirable properties in the limit $\sigma \to 0$ of the noise level σ: the solution is "consistent" in the sense that it converges to the true value as $\sigma \to 0$ and "efficient" in the sense that its covariance matrix approaches a theoretical lower bound as $\sigma \to 0$ (Chernov and Lesort, 2004; Kanatani 2008).

4. MAXIMUM LIKELIHOOD

Under our Gaussian noise model, maximum likelihood (ML) is equivalent to minimizing the Mahalanobis distance

$$I = \sum_{\alpha=1}^{N} (\bar{\boldsymbol{x}}_\alpha - \boldsymbol{x}_\alpha, V[\boldsymbol{x}_\alpha]^{-1}(\bar{\boldsymbol{x}}_\alpha - \boldsymbol{x}_\alpha)), \qquad (11)$$

with respect to $\bar{\boldsymbol{x}}_\alpha$ subject to the constraint that

$$(\boldsymbol{\xi}^{(k)}(\bar{\boldsymbol{x}}_\alpha), \boldsymbol{\theta}) = 0, \qquad k = 1, ..., L, \qquad (12)$$

for some $\boldsymbol{\theta}$. If the noise is homogeneous and isotropic, (11) is the sum of the squares of the geometric distances between the observations $\boldsymbol{x}_\alpha$ and their true values $\bar{\boldsymbol{x}}_\alpha$, often referred to as the *reprojection error* in the computer vision community (Hartley and Zisserman 2004). That name originates from the following intuition: We infer the 3-D structure of the scene from its projected images, and when the inferred 3-D structure is "reprojected" onto the images, (11) measures the discrepancy between the "reprojections" of our solution and the actual observations.

In the computer vision community, ML is known to produce highly accurate solutions (Hartley and Zisserman, 1004). It can also be shown that the ML solution achieves a theoretical accuracy limit, called the *KCR lower bound*, to a first approximation in the

noise level σ (Kanatani, 1994; Chernov and Lesort, 2004; Kanatani, 2008). Hence, not much accuracy improvement can be expected any longer. As far as image processing applications are concerned, where noise is small, usually the subpixel level, ML is virtually the ultimate desirable goal, and no necessity is felt for further accuracy improvement. Rather, a major concern is its computational burden, because ML usually requires complicated nonlinear optimization. The standard approach is to search the entire parameter space of $\bar{\boldsymbol{x}}_\alpha$, $\alpha = 1, ..., N$, and $\boldsymbol{\theta}$, which usually has very high dimensions, for the minimum of (11). This strategy is called *bundle adjustment* (Triggs et al., 2000; Hartley and Zisserman 2004), a term originally used by photogrametrists. This is very time consuming, in particular if one seeks a globally optimal solution by searching the entire parameter space exhaustively (Hartley and Kahl, 2008).

A popular alternative to bundle adjustment is minimization of a function of $\boldsymbol{\theta}$ alone, called the *Sampson error* (Hartley and Zisserman, 2004), which approximates the minimum of (11) for a given $\boldsymbol{\theta}$ (the actual expression is shown later). The name "Sampson error" stems from the classical ellipse fitting scheme of Sampson (1982). Kanatani and Sugaya (2010b) showed that the exact ML solution, equivalent to bundle adjustment, can be obtained by repeating Sampson error minimization, each time modifying the Sampson error so that in the end the modified Sampson error coincides with the Mahalanobis distance. It turns out that in many practical applications the solution that minimizes the Sampson error coincides with the exact ML solution up to several significant digits; usually, two or three rounds of Sampson error modification are sufficient (Kanatani and Sugaya, 2008; 2010a; Kanatani and Niitsuma 2010).

However, minimizing the Sampson error is not straightforward. Many numerical schemes have been proposed, including the *FNS* (*Fundamental Numerical Scheme*) of Chojnacki et al. (2000), the *HEIV* (*Heteroscedastic Errors-in-Variable*) of Leedan and Meer (2000) and Matei and Meer (2006), and the *projective Gauss-Newton iterations* of Kanatani and Sugaya (2007). All these rely on local search, but the iterations do not always converge if not started from a value sufficiently close to the solution. Hence, accurate approximation schemes that do not require iterations are very much desired, even though the solution may not be optimal, and various algebraic methods have been studied in the past.

5. ALGEGRAIC METHODS

For the sake of brevity, we abbreviate $\boldsymbol{\xi}^{(k)}(\boldsymbol{x}_\alpha)$ as $\boldsymbol{\xi}_\alpha^{(k)}$. *Algebraic methods* refer to those minimizing the *algebraic distance*

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}(\boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta})^2 = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}\boldsymbol{\theta}^\top\boldsymbol{\xi}_\alpha^{(k)}\boldsymbol{\xi}_\alpha^{(k)\top}\boldsymbol{\theta} = (\boldsymbol{\theta}, \boldsymbol{M}\boldsymbol{\theta}), \quad (13)$$

where we define

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}\boldsymbol{\xi}_\alpha^{(k)}\boldsymbol{\xi}_\alpha^{(k)\top}. \quad (14)$$

However, (13) is trivially minimized by $\boldsymbol{\theta} = \boldsymbol{0}$ unless some scale normalization is imposed on $\boldsymbol{\theta}$. The most common normalization is $\|\boldsymbol{\theta}\| = 1$; we call this the *standard LS*. The crucial fact is that *the solution depends on the normalization*. The aim of this paper is to find a normalization that maximizes the accuracy of the solution. This issue has been raised by Al-Sharadqah and Chernov (2009) and Rangarajan and Kanatani (2009) for circle fitting, by Kanatani and Rangarajan (2010) for ellipse fitting, and by Niitsuma et al. (2010) for homography estimation. In this paper, we generalize their results to an arbitrary number of constraints. Following Al-Sharadqah and Chernov (2009), Rangarajan and Kanatani (2009), Kanatani and Rangarajan (2010), and Niitsuma et al. (2010), we consider the class of normalizations

$$(\boldsymbol{\theta}, \boldsymbol{N}\boldsymbol{\theta}) = \text{constant}. \quad (15)$$

Traditionally, the matrix $\boldsymbol{N}$ is assumed to be positive definite, but here we allow $\boldsymbol{N}$ to be nondefinite (i.e., neither positive nor negative definite), so the constant in (15) is not necessarily positive. Given the matrix $\boldsymbol{N}$, the solution $\boldsymbol{\theta}$ that minimizes (13) subject to (15), if it exists, is obtained as the solution of the generalized eigenvalue problem

$$\boldsymbol{M}\boldsymbol{\theta} = \lambda\boldsymbol{N}\boldsymbol{\theta}. \quad (16)$$

In the absence of noise (i.e., if the image operations detected them without uncertainty), we have $(\boldsymbol{\theta}, \boldsymbol{\xi}_\alpha) = 0$ for all α and hence from (14) we have $\boldsymbol{M}\boldsymbol{\theta} = \boldsymbol{0}$, i.e., $\lambda = 0$. If $\boldsymbol{N}$ is positive definite or semidefinite, the generalized eigenvalue λ is positive in the presence of noise. The corresponding solution is obtained as the generalized eigenvector $\boldsymbol{\theta}$ for the smallest λ. Here, however, we allow $\boldsymbol{N}$ to be nondefinite, so λ may not be positive. In this paper, we do error analysis of (16) by assuming that $\lambda \approx 0$, following Kanatani (2008). So, we choose the solution to be the generalized eigenvector $\boldsymbol{\theta}$ for the λ with the smallest absolute value.

Since $\boldsymbol{M}$ is a random variable, it can become arbitrarily large with a small probability, and for a nondefinite $\boldsymbol{N}$ the solution $\boldsymbol{\theta}$ that minimize (13) subject to (16) can become arbitrarily large. As a result,

the expectation $E[\|\hat{\boldsymbol{\theta}}\|]$ of the resulting estimator $\hat{\boldsymbol{\theta}}$ could possibly diverge to ∞ for some $\boldsymbol{\theta}$ (Cheng and Kukush, 2006). In order to avoid such an anomaly, we hereafter start from (16), which we identify with the definition of our "algebraic method". Once in the form of (16), the solution $\boldsymbol{\theta}$ has scale indeterminacy, so we can adopt normalization $\|\boldsymbol{\theta}\| = 1$ rather than (16). Then, $\boldsymbol{\theta}$ is always a unit vector. The standard LS is the choice of $\boldsymbol{N} = \boldsymbol{I}$, for which (16) becomes an ordinary eigenvalue problem

$$\boldsymbol{M}\boldsymbol{\theta} = \lambda\boldsymbol{\theta}, \tag{17}$$

and the solution is the unit eigenvector $\boldsymbol{\theta}$ of $\boldsymbol{M}$ for the smallest eigenvalue λ.

6. ERROR ANALYSIS

We can expand each $\boldsymbol{\xi}_\alpha^{(k)}$ in the form

$$\boldsymbol{\xi}_\alpha^{(k)} = \bar{\boldsymbol{\xi}}_\alpha^{(k)} + \Delta_1\boldsymbol{\xi}_\alpha^{(k)} + \Delta_2\boldsymbol{\xi}_\alpha^{(k)} + \cdots, \tag{18}$$

where $\bar{\boldsymbol{\xi}}_\alpha^{(k)}$ is the noiseless value, and $\Delta_i\boldsymbol{\xi}_\alpha^{(k)}$ is the ith order term in $\Delta\boldsymbol{x}_\alpha$. The first order term is written as

$$\Delta_1\boldsymbol{\xi}_\alpha^{(k)} = \boldsymbol{T}_\alpha^{(k)}\Delta\boldsymbol{x}_\alpha, \quad \boldsymbol{T}_\alpha^{(k)} \equiv \left.\frac{\partial\boldsymbol{\xi}^{(k)}(\boldsymbol{x}_\alpha)}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha}. \tag{19}$$

We define the covariance matrices of $\boldsymbol{\xi}_\alpha^{(k)}$, $k = 1, ..., L$, by

$$\begin{aligned} V^{(kl)}[\boldsymbol{\xi}_\alpha] &\equiv E[\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\alpha^{(l)\top}] \\ &= \boldsymbol{T}_\alpha^{(k)}E[\Delta\boldsymbol{x}_\alpha\Delta\boldsymbol{x}_\alpha^\top]\boldsymbol{T}_\alpha^{(l)\top} \\ &= \boldsymbol{T}_\alpha^{(k)}V[\boldsymbol{x}_\alpha]\boldsymbol{T}_\alpha^{(l)\top}, \end{aligned} \tag{20}$$

where $E[\,\cdot\,]$ denotes expectation. The *Sampson error* that approximates the minimum of the Mahalanobis distance in (11) subject to the constraints in (12) has the following form (Kanatani, 1996; Hartley and Zisserman, 2004):

$$K = \sum_{\alpha=1}^{N}\sum_{k,l,=1}^{L} W_\alpha^{(kl)}(\boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta})(\boldsymbol{\xi}_\alpha^{(l)}, \boldsymbol{\theta}). \tag{21}$$

Here, $W_\alpha^{(kl)}$ is the (kl) element of $(\boldsymbol{V}_\alpha)_r^-$, and $\boldsymbol{V}_\alpha$ is the matrix whose (kl) element is

$$\boldsymbol{V}_\alpha = \Big((\boldsymbol{\theta}, V^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})\Big), \tag{22}$$

where the true data values $\bar{\boldsymbol{x}}_\alpha$ in the definition of $V^{(kl)}[\boldsymbol{\xi}_\alpha]$ are replaced by their observations $\boldsymbol{x}_\alpha$. The operation $(\,\cdot\,)_r^-$ denotes the pseudoinverse of truncated rank r (i.e., with all eigenvalues except the largest r replaced by 0 in the spectral decomposition), and r is the rank (the number of independent equations) of the constraint in (12).

Example 4 (Ellipse fitting). For the ellipse fitting in Example 1, the first order error $\Delta_1\boldsymbol{\xi}$ is written as

$$\Delta_1\boldsymbol{\xi}_\alpha = 2\begin{pmatrix} \bar{x}_\alpha & \bar{y}_\alpha & 0 & 1 & 0 & 0 \\ 0 & \bar{x}_\alpha & \bar{y}_\alpha & 0 & 1 & 0 \end{pmatrix}^\top \begin{pmatrix} \Delta x_\alpha \\ \Delta y_\alpha \end{pmatrix}. \tag{23}$$

The second order error $\Delta_2\boldsymbol{\xi}_\alpha$ has the following form:

$$\Delta_2\boldsymbol{\xi}_\alpha = (\Delta x_\alpha^2, 2\Delta x_\alpha\Delta y_\alpha, \Delta y_\alpha^2, 0, 0, 0)^\top. \tag{24}$$

Example 5 (Fundamental matrix computation). For the fundamental matrix computation in Example 2, the first order error $\Delta_1\boldsymbol{\xi}$ is written as

$$\Delta_1\boldsymbol{\xi}_\alpha = \begin{pmatrix} \bar{x}'_\alpha & \bar{y}'_\alpha & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{x}'_\alpha & \bar{y}'_\alpha & 1 & 0 & 0 & 0 \\ \bar{x}_\alpha & 0 & 0 & \bar{y}_\alpha & 0 & 0 & 1 & 0 & 0 \\ 0 & \bar{x}_\alpha & 0 & 0 & \bar{y}_\alpha & 0 & 0 & 1 & 0 \end{pmatrix}^\top \begin{pmatrix} \Delta x_\alpha \\ \Delta y_\alpha \\ \Delta x'_\alpha \\ \Delta y'_\alpha \end{pmatrix}. \tag{25}$$

The second order error $\Delta_2\boldsymbol{\xi}_\alpha$ has the following form:

$$\begin{aligned} \Delta_2\boldsymbol{\xi}_\alpha = (&\Delta x_\alpha\Delta x'_\alpha, \Delta x_\alpha\Delta y'_\alpha, 0, \Delta y_\alpha\Delta x'_\alpha, \Delta y_\alpha\Delta y'_\alpha, \\ &0, 0, 0)^\top. \end{aligned} \tag{26}$$

Example 6 (Homography computation). For the fundamental matrix computation in Example 2, the first order error $\Delta_1\boldsymbol{\xi}$ is written as

$$\Delta_1\boldsymbol{\xi}_\alpha^{(1)} = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 & 0 & \bar{y}'_\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & \bar{y}'_\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \bar{x}_\alpha & \bar{y}_\alpha & 1 \end{pmatrix}^\top \begin{pmatrix} \Delta x_\alpha \\ \Delta y_\alpha \\ \Delta x'_\alpha \\ \Delta y'_\alpha \end{pmatrix},$$

$$\Delta_1\boldsymbol{\xi}_\alpha^{(2)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -\bar{x}'_\alpha & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -\bar{x}'_\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\bar{x}_\alpha & -\bar{y}_\alpha & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^\top \begin{pmatrix} \Delta x_\alpha \\ \Delta y_\alpha \\ \Delta x'_\alpha \\ \Delta y'_\alpha \end{pmatrix},$$

$$\Delta_1\boldsymbol{\xi}_\alpha^{(3)} = \begin{pmatrix} -\bar{y}'_\alpha & 0 & 0 & \bar{x}'_\alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & -\bar{y}'_\alpha & 0 & 0 & \bar{x}'_\alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{x}_\alpha & \bar{y}_\alpha & 1 & 0 & 0 & 0 \\ -\bar{x}_\alpha & -\bar{y}_\alpha & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^\top \begin{pmatrix} \Delta x_\alpha \\ \Delta y_\alpha \\ \Delta x'_\alpha \\ \Delta y'_\alpha \end{pmatrix}. \tag{27}$$

The second order error $\Delta_2\boldsymbol{\xi}_\alpha^{(k)}$ has the following form:

$$\begin{aligned} \Delta_2\boldsymbol{\xi}_\alpha^{(1)} &= (0, 0, 0, 0, 0, 0, \Delta x_\alpha\Delta y'_\alpha, \Delta y_\alpha\Delta y'_\alpha, 0)^\top, \\ \Delta_2\boldsymbol{\xi}_\alpha^{(2)} &= (0, 0, 0, 0, 0, 0, -\Delta x'_\alpha\Delta x_\alpha, -\Delta x'_\alpha\Delta y_\alpha, 0)^\top, \\ \Delta\boldsymbol{\xi}_\alpha^{(3)} &= (-\Delta y'_\alpha\Delta x_\alpha, -\Delta y'_\alpha\Delta y_\alpha, 0, \Delta x'_\alpha\Delta x_\alpha, \\ &\quad \Delta x'_\alpha\Delta y_\alpha, 0, 0, 0, 0)^\top. \end{aligned} \tag{28}$$

7. PERTURBATION ANALYSIS

Substituting (18) into (14), we obtain

$$\boldsymbol{M} = \bar{\boldsymbol{M}} + \Delta_1\boldsymbol{M} + \Delta_2\boldsymbol{M} + \cdots, \tag{29}$$

where

$$\bar{\boldsymbol{M}} = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(k)\top}, \tag{30}$$

$$\Delta_1 \boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}(\bar{\boldsymbol{\xi}}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top} + \Delta_1\boldsymbol{\xi}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(k)\top}), \tag{31}$$

$$\Delta_2 \boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}(\bar{\boldsymbol{\xi}}_\alpha^{(k)}\Delta_2\boldsymbol{\xi}_\alpha^{(k)\top} + \Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top} + \Delta_2\boldsymbol{\xi}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(k)\top}). \tag{32}$$

We also expand the solution $\boldsymbol{\theta}$ and λ of (16) in the form

$$\boldsymbol{\theta} = \bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots, \quad \lambda = \bar{\lambda} + \Delta_1\lambda + \Delta_2\lambda + \cdots. \tag{33}$$

Substituting (29) and (33) into (16), we have

$$(\bar{\boldsymbol{M}} + \Delta_1\boldsymbol{M} + \Delta_2\boldsymbol{M} + \cdots)(\bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots) = (\bar{\lambda} + \Delta_1\lambda + \Delta_2\lambda + \cdots)\boldsymbol{N}(\bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots). \tag{34}$$

Equating terms of the same order, we obtain

$$\bar{\boldsymbol{M}}\bar{\boldsymbol{\theta}} = \bar{\lambda}\boldsymbol{N}\bar{\boldsymbol{\theta}}, \tag{35}$$

$$\bar{\boldsymbol{M}}\Delta_1\boldsymbol{\theta} + \Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}} = \bar{\lambda}\boldsymbol{N}\Delta_1\boldsymbol{\theta} + \Delta_1\lambda\boldsymbol{N}\bar{\boldsymbol{\theta}}, \tag{36}$$

$$\bar{\boldsymbol{M}}\Delta_2\boldsymbol{\theta} + \Delta_1\boldsymbol{M}\Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{M}\bar{\boldsymbol{\theta}} = \bar{\lambda}\boldsymbol{N}\Delta_2\boldsymbol{\theta} + \Delta_1\lambda\boldsymbol{N}\Delta_1\boldsymbol{\theta} + \Delta_2\lambda\boldsymbol{N}\bar{\boldsymbol{\theta}}. \tag{37}$$

We have $\bar{\boldsymbol{M}}\bar{\boldsymbol{\theta}} = \boldsymbol{0}$ for the true values, so $\bar{\lambda} = 0$. From (31), we have $(\bar{\boldsymbol{\theta}}, \Delta_1\bar{\boldsymbol{M}}\bar{\boldsymbol{\theta}}) = 0$. Computing the inner product of (36) and $\bar{\boldsymbol{\theta}}$ on both sides, we see that $\Delta_1\lambda = 0$. Multiplying (36) by the pseudoinverse $\bar{\boldsymbol{M}}^-$ of $\bar{\boldsymbol{M}}$ from left, we obtain

$$\Delta_1\boldsymbol{\theta} = -\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}}. \tag{38}$$

Note that since $\bar{\boldsymbol{M}}\bar{\boldsymbol{\theta}} = \boldsymbol{0}$, the matrix $\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}$ ($\equiv \boldsymbol{P}_{\bar{\boldsymbol{\theta}}}$) is the projection operator in the direction orthogonal to $\bar{\boldsymbol{\theta}}$. Also, equating the first order terms in the expansion $\|\bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots\|^2 = 1$ shows $(\bar{\boldsymbol{\theta}}, \Delta_1\boldsymbol{\theta}) = 0$ (Kanatani, 2008), hence $\boldsymbol{P}_{\bar{\boldsymbol{\theta}}}\Delta_1\boldsymbol{\theta} = \Delta_1\boldsymbol{\theta}$. Substituting (38) into (37) and computing its inner product with $\bar{\boldsymbol{\theta}}$ on both sides, we obtain

$$\Delta_2\lambda = \frac{(\bar{\boldsymbol{\theta}}, \Delta_2\boldsymbol{M}\bar{\boldsymbol{\theta}}) - (\bar{\boldsymbol{\theta}}, \Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \boldsymbol{N}\bar{\boldsymbol{\theta}})} = \frac{(\bar{\boldsymbol{\theta}}, \boldsymbol{T}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \boldsymbol{N}\bar{\boldsymbol{\theta}})}, \tag{39}$$

where we put

$$\boldsymbol{T} = \Delta_2\boldsymbol{M} - \Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}. \tag{40}$$

Next, we consider the second order error $\Delta_2\boldsymbol{\theta}$. Since $\boldsymbol{\theta}$ is normalized to have unit norm, we are interested in the error component orthogonal to $\bar{\boldsymbol{\theta}}$. So, we consider

$$\Delta_2^\perp\boldsymbol{\theta} \equiv \boldsymbol{P}_{\bar{\boldsymbol{\theta}}}\Delta_2\boldsymbol{\theta}\ (= \bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\Delta_2\boldsymbol{\theta}). \tag{41}$$

Multiplying (37) by $\bar{\boldsymbol{M}}^-$ from left and substituting (38), we obtain

$$\Delta_2^\perp\boldsymbol{\theta} = \Delta_2\lambda\bar{\boldsymbol{M}}^-\boldsymbol{N}\bar{\boldsymbol{\theta}} + \bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}} - \bar{\boldsymbol{M}}^-\Delta_2\boldsymbol{M}\bar{\boldsymbol{\theta}} = \frac{(\bar{\boldsymbol{\theta}}, \boldsymbol{T}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \boldsymbol{N}\bar{\boldsymbol{\theta}})}\bar{\boldsymbol{M}}^-\boldsymbol{N}\bar{\boldsymbol{\theta}} - \bar{\boldsymbol{M}}^-\boldsymbol{T}\bar{\boldsymbol{\theta}}. \tag{42}$$

8. COVARIANCE AND BIAS

8.1 Covariance Analysis

From (38), the covariance matrix $V[\boldsymbol{\theta}]$ of the solution $\boldsymbol{\theta}$ has the leading term

$$\begin{aligned}
V[\boldsymbol{\theta}] &= E[\Delta_1\boldsymbol{\theta}\Delta_1\boldsymbol{\theta}^\top] \\
&= \frac{1}{N^2}\bar{\boldsymbol{M}}^-E[(\Delta_1\boldsymbol{M}\boldsymbol{\theta})(\Delta_1\boldsymbol{M}\boldsymbol{\theta})^\top]\bar{\boldsymbol{M}}^- \\
&= \frac{1}{N^2}\bar{\boldsymbol{M}}^-E\Big[\sum_{\alpha=1}^{N}\sum_{k=1}^{L}(\Delta\boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\alpha^{(k)} \\
&\quad \sum_{\beta=1}^{N}\sum_{l=1}^{L}(\Delta\boldsymbol{\xi}_\beta^{(l)}, \boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\Big]\bar{\boldsymbol{M}}^- \\
&= \frac{1}{N^2}\bar{\boldsymbol{M}}^-\sum_{\alpha,\beta=1}^{N}\sum_{k,l=1}^{L}(\boldsymbol{\theta}, E[\Delta\boldsymbol{\xi}_\alpha^{(k)}\Delta\boldsymbol{\xi}_\beta^{(l)\top}]\boldsymbol{\theta}) \\
&\quad \bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\bar{\boldsymbol{M}}^- \\
&= \frac{1}{N^2}\bar{\boldsymbol{M}}^-\Big(\sum_{\alpha=1}^{N}\sum_{k,l=1}^{L}(\boldsymbol{\theta}, V^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}\Big)\bar{\boldsymbol{M}}^- \\
&= \frac{1}{N}\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}'\bar{\boldsymbol{M}}^-,
\end{aligned} \tag{43}$$

where we define

$$\bar{\boldsymbol{M}}' = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k,l=1}^{L}(\boldsymbol{\theta}, V^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top} \tag{44}$$

In the above derivation, we have noted that from our noise assumption we have $E[\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}] = \delta_{\alpha\beta}V^{(kl)}[\boldsymbol{\xi}_\alpha]$, where $\delta_{\alpha\beta}$ is the Kronecker delta.

8.2 Bias Analysis

The important observation is that the covariance matrix $V[\boldsymbol{\theta}]$ does not contain $\boldsymbol{N}$. Thus, *all algebraic methods have the same covariance matrix in the leading order*, as pointed out by Al-Sharadqah and Chernov (2009) for circle fitting. This observation leads us to focus on the bias. From (38), we see that the first order bias $E[\Delta_1\boldsymbol{\theta}]$ is $\boldsymbol{0}$, hence the leading bias is $E[\Delta_2^\perp\boldsymbol{\theta}]$. From (42), we have

$$E[\Delta_2^\perp\boldsymbol{\theta}] = \frac{(\bar{\boldsymbol{\theta}}, E[\boldsymbol{T}]\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \boldsymbol{N}\bar{\boldsymbol{\theta}})}\bar{\boldsymbol{M}}^-\boldsymbol{N}\bar{\boldsymbol{\theta}} - \bar{\boldsymbol{M}}^-E[\boldsymbol{T}]\bar{\boldsymbol{\theta}}. \tag{45}$$

We now evaluate the expectation $E[\boldsymbol{T}]$ of $\boldsymbol{T}$ in (40). From (32), we see that $E[\Delta_2\boldsymbol{M}]$ is given by

$$\begin{aligned} &E[\Delta_2\boldsymbol{M}] \\ &= \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}\Big(\bar{\boldsymbol{\xi}}_\alpha^{(k)}E[\Delta_2\boldsymbol{\xi}_\alpha^{(k)}]^\top \\ &\quad +E[\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top}]+E[\Delta_2\boldsymbol{\xi}_\alpha^{(k)}]\bar{\boldsymbol{\xi}}_\alpha^{(k)\top}\Big) \\ &= \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}\Big(V^{(kk)}[\boldsymbol{\xi}_\alpha]+2\mathcal{S}[\bar{\boldsymbol{\xi}}_\alpha^{(k)}\boldsymbol{e}_\alpha^{(k)\top}]\Big), \quad (46)\end{aligned}$$

where we have used (20) and defined

$$\boldsymbol{e}_\alpha^{(k)} \equiv E[\Delta_2\boldsymbol{\xi}_\alpha^{(k)}]. \quad (47)$$

The operator $\mathcal{S}[\,\cdot\,]$ denotes symmetrization ($\mathcal{S}[\boldsymbol{A}] = (\boldsymbol{A}+\boldsymbol{A}^\top)/2$). The expectation $E[\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}]$ has the following form (see Appendix):

$$\begin{aligned} &E[\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}] \\ &= \frac{1}{N^2}\sum_{\alpha=1}^{N}\sum_{k,l=1}^{L}\Big(\mathrm{tr}[\bar{\boldsymbol{M}}^-V^{(kl)}[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top} \\ &\quad +(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(l)})V^{(kl)}[\boldsymbol{\xi}_\alpha] \\ &\quad +2\mathcal{S}[V^{(kl)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}]\Big). \quad (48)\end{aligned}$$

From (46) and (48), the expectation of $\boldsymbol{T}$ is

$$\begin{aligned} E[\boldsymbol{T}] = \boldsymbol{N}_{\mathrm{T}} &- \frac{1}{N^2}\sum_{\alpha=1}^{N}\sum_{k,l=1}^{L}\Big(\mathrm{tr}[\bar{\boldsymbol{M}}^-V^{(kl)}[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top} \\ &+(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(l)})V^{(kl)}[\boldsymbol{\xi}_\alpha] \\ &+2\mathcal{S}[V^{(kl)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}]\Big), \quad (49)\end{aligned}$$

where we put

$$\boldsymbol{N}_{\mathrm{T}} = \frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{L}\Big(V^{(kk)}[\boldsymbol{\xi}_\alpha]+2\mathcal{S}[\bar{\boldsymbol{\xi}}_\alpha^{(k)}\boldsymbol{e}_\alpha^{(k)\top}]\Big). \quad (50)$$

9. HYPERLS

We propose to use as $\boldsymbol{N}$ the expression $E[\boldsymbol{T}]$ in (49), namely,

$$\begin{aligned} \boldsymbol{N} = \boldsymbol{N}_{\mathrm{T}} &- \frac{1}{N^2}\sum_{\alpha=1}^{N}\sum_{k,l=1}^{L}\Big(\mathrm{tr}[\bar{\boldsymbol{M}}^-V^{(kl)}[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top} \\ &+(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(l)})V^{(kl)}[\boldsymbol{\xi}_\alpha] \\ &+2\mathcal{S}[V^{(kl)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}]\Big), \quad (51)\end{aligned}$$

Letting $\boldsymbol{N} = E[\boldsymbol{T}]$ in (45), we see that

$$E[\Delta_2^\perp\boldsymbol{\theta}] = \bar{\boldsymbol{M}}^-\Big(\frac{(\bar{\boldsymbol{\theta}},\boldsymbol{N}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}},\boldsymbol{N}\bar{\boldsymbol{\theta}})}\boldsymbol{N}-\boldsymbol{N}\Big)\bar{\boldsymbol{\theta}} = \boldsymbol{0}. \quad (52)$$

Since the right-hand side of (49) contains the true values $\bar{\boldsymbol{\xi}}_\alpha$ and $\bar{\boldsymbol{M}}$, we replace $\bar{\boldsymbol{x}}_\alpha$ in their definition by the observation $\boldsymbol{x}_\alpha$. This does not affect the result, since the odd order noise terms have expectation 0 and hence the resulting error in $E[\Delta_2^\perp\boldsymbol{\theta}]$ is of the fourth order. Thus, the second order bias is *exactly* 0. We call this scheme *hyper least squares*, or *HyperLS* for short, after Al-Sharadqah and Chernov (2009). Note that $\boldsymbol{N}$ has scale indeterminacy: If $\boldsymbol{N}$ is multiplied by c ($\neq 0$), (16) has the same solution $\boldsymbol{\theta}$; only λ is divided by c. Thus, the noise characteristics $V^{(kl)}[\boldsymbol{\xi}_\alpha]$ in (20) and hence $V[\boldsymbol{x}_\alpha]$ need to be known only up to scale; *we need not know the absolute magnitude of the noise*.

For numerical computation, standard linear algebra routines for solving the generalized eigenvalue problem of (16) assume that $\boldsymbol{N}$ is positive definite, but here $\boldsymbol{N}$ is nondefinite. We circumvent this problem by rewriting (16) in the form

$$\boldsymbol{N}\boldsymbol{\theta} = \frac{1}{\lambda}\boldsymbol{M}\boldsymbol{\theta}. \quad (53)$$

The matrix $\boldsymbol{M}$ in (14) is positive definite except in the absence of noise, in which case the smallest eigenvalue is 0.

Example 7 (Ellipse fitting). If the noise in (x_α, y_α) is independent and Gaussian with mean 0 and standard deviation σ, the vector $\boldsymbol{e}_\alpha$ ($= \boldsymbol{e}^{(1)}$) in (47) is given by

$$\boldsymbol{e}_\alpha = \sigma^2(1,0,1,0,0,0)^\top. \quad (54)$$

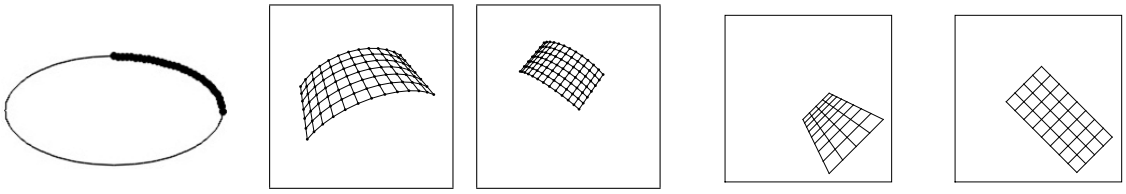
Hence, the matrix $\boldsymbol{N}_{\mathrm{T}}$ in (50) is given by

$$\boldsymbol{N}_{\mathrm{T}} = \frac{\sigma^2}{N}\sum_{\alpha=1}^{N}\begin{pmatrix} 6x_\alpha^2 & 6x_\alpha y_\alpha & x_\alpha^2+y_\alpha^2 & 6x_\alpha & 2y_\alpha & 1 \\ 6x_\alpha y_\alpha & 4(x_\alpha^2+y_\alpha^2) & 6x_\alpha y_\alpha & 4y_\alpha & 4x_\alpha & 0 \\ x_\alpha^2+y_\alpha^2 & 6x_\alpha y_\alpha & 6y_\alpha^2 & 2x_\alpha & 6y_\alpha & 1 \\ 6x_\alpha & 4y_\alpha & 2x_\alpha & 4 & 0 & 0 \\ 2y_\alpha & 4x_\alpha & 6y_\alpha & 0 & 4 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \quad (55)$$

The method due to Taubin (1991) is to use as $\boldsymbol{N}$

$$\boldsymbol{N}_{\mathrm{Taubin}} = \frac{4\sigma^2}{N}\sum_{\alpha=1}^{N}\begin{pmatrix} x_\alpha^2 & x_\alpha y_\alpha & 0 & x_\alpha & 0 & 0 \\ x_\alpha y_\alpha & x_\alpha^2+y_\alpha^2 & x_\alpha y_\alpha & y_\alpha & x_\alpha & 0 \\ 0 & x_\alpha y_\alpha & y_\alpha^2 & 0 & y_\alpha & 0 \\ x_\alpha & y_\alpha & 0 & 1 & 0 & 0 \\ 0 & x_\alpha & y_\alpha & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (56)$$

which we see is obtained by letting $\boldsymbol{e}_\alpha = \boldsymbol{0}$ in (50). As pointed out earlier, the value of σ in (55) and (56) need not be known. Hence, we can simply let $\sigma = 1$ in (16) and (53) in actual computation.



(a) (b) (c)

Figure 2: (a) 31 points on an ellipse. (b) Two views of a curved grid. (c) Two views of a planar grid.

Example 8 (Fundamental matrix computation). If the noise in (x_α, y_α) and (x'_α, y'_α) is independent and Gaussian with mean 0 and standard deviation σ, the vector $\boldsymbol{e}_\alpha$ $(= \boldsymbol{e}^{(1)})$ in (47) is $\boldsymbol{0}$, so the $\boldsymbol{N}_{\rm T}$ in (50) becomes

$$\boldsymbol{N}_{\rm T} = \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} \begin{pmatrix} x_\alpha^2 + x_\alpha'^2 & x'_\alpha y'_\alpha & x'_\alpha & x_\alpha y_\alpha & 0 \\ x'_\alpha y'_\alpha & x_\alpha^2 + y_\alpha'^2 & y'_\alpha & 0 & x_\alpha y_\alpha \\ x'_\alpha & y'_\alpha & 1 & 0 & 0 \\ x_\alpha y_\alpha & 0 & 0 & y_\alpha^2 + x_\alpha'^2 & x'_\alpha y'_\alpha \\ 0 & x_\alpha y_\alpha & 0 & x'_\alpha y'_\alpha & y_\alpha^2 + y_\alpha'^2 \\ 0 & 0 & 0 & x'_\alpha & y'_\alpha \\ x_\alpha & 0 & 0 & y_\alpha & 0 \\ 0 & x_\alpha & 0 & 0 & y_\alpha \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & x_\alpha & 0 & 0 \\ 0 & 0 & x_\alpha & 0 \\ 0 & 0 & 0 & 0 \\ x'_\alpha & y_\alpha & 0 & 0 \\ y'_\alpha & 0 & y_\alpha & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{57}$$

It turns out that the use of this matrix $\boldsymbol{N}_{\rm T}$ coincides with the well known method of Taubin (1991). As in ellipse fitting, we can let $\sigma = 1$ in (57) in actual computation.

Example 9 (Homography computation). If the noise in (x_α, y_α) and (x'_α, y'_α) is independent and Gaussian with mean 0 and standard deviation σ, the vectors $\boldsymbol{e}_\alpha^{(k)}$ in (47) are all $\boldsymbol{0}$, so the $\boldsymbol{N}_{\rm T}$ in (50) becomes

$$\boldsymbol{N}_{\rm T} = \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} \begin{pmatrix} x_\alpha^2+y_\alpha'^2+1 & x_\alpha y_\alpha & x_\alpha & -x'_\alpha y'_\alpha \\ x_\alpha y_\alpha & y_\alpha^2+y_\alpha'^2+1 & y_\alpha & 0 \\ x_\alpha & y_\alpha & 1 & 0 \\ -x'_\alpha y'_\alpha & 0 & 0 & x_\alpha^2+x_\alpha'^2+1 \\ 0 & -x'_\alpha y'_\alpha & 0 & x_\alpha y_\alpha \\ 0 & 0 & 0 & x_\alpha \\ -x'_\alpha & 0 & 0 & -y'_\alpha \\ 0 & -x'_\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & -x'_\alpha & 0 & 0 \\ -x'_\alpha y'_\alpha & 0 & 0 & -x'_\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 \\ x_\alpha y_\alpha & x_\alpha & -y'_\alpha & 0 & 0 \\ y_\alpha^2+x_\alpha'^2+1 & y_\alpha & 0 & -y'_\alpha & 0 \\ 0 & 0 & 0 & & \\ 0 & 0 & x_\alpha^2+x_\alpha'^2+y_\alpha'^2 & 2x_\alpha y_\alpha & 2x_\alpha \\ -x'_\alpha & 0 & 2x_\alpha y_\alpha & y_\alpha^2+x_\alpha'^2+y_\alpha'^2 & 2y_\alpha \\ 0 & 0 & 2x_\alpha & 2y_\alpha & 2 \end{pmatrix}, \tag{58}$$

For homography computation, the constraint is a vector equation in (8). Hence, the method of Taubin (1991) cannot be applied. However, the use of the above $\boldsymbol{N}_{\rm T}$ as $\boldsymbol{N}$ plays the same role of the method of Taubin (1991) for ellipse fitting and fundamental matrix computation, as first pointed out by Niitsuma et al. (2010). As before, we can let $\sigma = 1$ in the matrix $\boldsymbol{N}_{\rm T}$ in actual computation.

10. NUMERICAL EXPERIMENTS

We did the following three experiments:

Ellipse fitting: We fit an ellipse to the point sequence shown in Fig. 2(a). We took 31 equidistant points on the first quadrant of an ellipse with major and minor axes 100 and 50 pixels, respectively.

Fundamental matrix computation: We compute the fundamental matrix between the two images shown in Fig. 2(b), which view a cylindrical grid surface from two directions. The image size is assumed to be 600×600 (pixels) with focal lengths 600 pixels for both. The 91 grid points are used as corresponding points.

Homography computation: We compute the homography relating the two images shown in Fig. 2(c), which view a planar grid surface from two directions. The image size is assumed to be 800×800 (pixels) with focal lengths 600 pixels for both. The 45 grid points are used as corresponding points.

For each example, we compared the standard LS, our HyperLS, its Taubin approximation, and ML, for which we used the FNS of Chojnacki et al. (2000) for ellipse fitting and fundamental matrix computation and the multiconstraint FNS of Niitsuma et

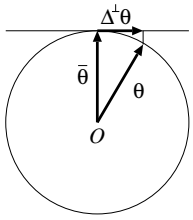


Figure 3: The true value $\bar{\boldsymbol{\theta}}$, the computed value $\boldsymbol{\theta}$, and its orthogonal component $\Delta\boldsymbol{\theta}$ to $\bar{\boldsymbol{\theta}}$.

al. (2010). As mentioned in Sec. 4, FNS and similar schemes like HEIV and projective Gauss-Newton iterations minimize not directly (11) but the Sampson error in (21), which approximates the minimum of (11), and the exact ML solution can be obtained by repeated Sampson error minimization (Kanatani and Sugaya, 2010). However, It has been observed that the solution that minimizes the Sampson error agrees with the ML solution up to several significant digits (Kanatani and Sugaya, 2008; 2010a; Kanatani and Niitsuma, 2010), and hence FNS can safely be regarded as minimizing (11).

Let $\bar{\boldsymbol{\theta}}$ be the true value of the parameter $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}$ its computed value. We consider the following error:

$$\Delta^{\perp}\boldsymbol{\theta} = \boldsymbol{P}_{\bar{\boldsymbol{\theta}}}\hat{\boldsymbol{\theta}}, \qquad \boldsymbol{P}_{\bar{\boldsymbol{\theta}}} \equiv \boldsymbol{I} - \bar{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}^{\top}. \tag{59}$$

Recall that matrix $\boldsymbol{P}_{\bar{\boldsymbol{\theta}}}$ represents the orthogonal projection onto the space orthogonal to $\bar{\boldsymbol{\theta}}$. Since the computed value $\hat{\boldsymbol{\theta}}$ is normalized to a unit vector, it distributes around $\bar{\boldsymbol{\theta}}$ on the unit sphere. Hence, the meaningful deviation is its component orthogonal to $\bar{\boldsymbol{\theta}}$, and we measure the error component in the tangent space to the unit sphere at $\bar{\boldsymbol{\theta}}$ (Fig. 3). The theoretical accuracy limit, called the *KCR lower bound* (Kanatani, 1996; Chernov and Lesort, 2004; Kanatani, 2008), is given by

$$E[\Delta^{\perp}\boldsymbol{\theta}\Delta^{\perp}\boldsymbol{\theta}^{\top}] \succ \frac{1}{N}\Big(\frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k,l}\bar{W}_{\alpha}^{(kl)}\bar{\boldsymbol{\xi}}_{\alpha}^{(k)}\bar{\boldsymbol{\xi}}_{\alpha}^{(l)\top}\Big)^{-} \equiv V_{\rm KCR}[\boldsymbol{\theta}], \tag{60}$$

where $\bar{W}_{\alpha}^{(kl)}$ is the value of $W_{\alpha}^{(kl)}$ in (21) evaluated by using the true values $\bar{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\xi}}_{\alpha}^{(kl)}$. The relation $\succ$ means that the left-hand side minus the right-hand side is a positive semidefinite symmetric matrix, and the operation $(\,\cdot\,)^{-}$ denotes pseudoinverse.

We added independent Gaussian noise of mean 0 and standard deviation σ to the x and y coordinates of data each point and repeated the fitting M times for each σ, using different noise. We let $M = 10000$ for ellipse fitting and fundamental matrix computation and $M = 1000$ for homography computation. Then, we evaluated the root-mean-square (RMS) error

$$E = \sqrt{\frac{1}{M}\sum_{a=1}^{M}\|\Delta^{\perp}\boldsymbol{\theta}^{(a)}\|^2}, \tag{61}$$

where $\Delta\boldsymbol{\theta}^{(a)}$ is the value of $\Delta\boldsymbol{\theta}$ in the ath trial. We compared this to the following bound obtained by computing the trace (60):

$$\sqrt{E[\|\Delta^{\perp}\boldsymbol{\theta}\|^2]} \geq \sqrt{{\rm tr}V_{\rm KCR}[\boldsymbol{\theta}]}. \tag{62}$$

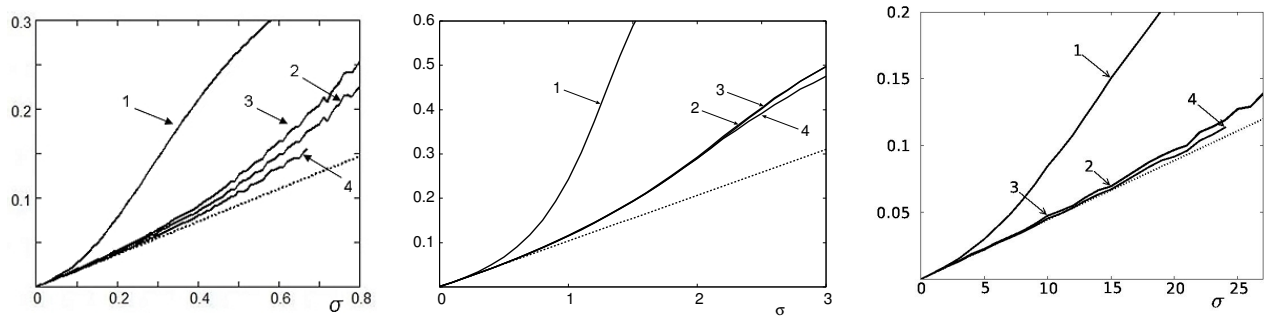
Figure 4 plots for σ the RMS error of (61) for each method and the KCR lower bound of (62). We observe the following:

Ellipse fitting: The standard LS performs poorly, while ML exhibits the highest accuracy, almost reaching the KCR lower bound. However, ML computation fails to converge above a certain noise level. In contrast, HyperLS produces, without iterations, an accurate solution close to ML. The accuracy of its Taubin approximation is practically the same as the traditional Taubin method and is slightly lower than HyperLS.

Fundamental matrix computation: Again, the standard LS is poor, while ML has the highest accuracy, almost reaching the KCR lower bound. The accuracy of HyperLS is very close to ML. Its Taubin approximation (= the traditional Taubin method) has practically the same accuracy as HyperLS. The fundamental matrix has the constraint that its rank be 2. The comparison here is done before the rank constraint is imposed.

Homography computation: In this case, too, the standard LS is poor, while ML has the highest accuracy, almost reaching the KCR lower bound. However, ML computation fails to converge above a certain noise level. The accuracy of HyperLS is very close to ML. Its Taubin approximation has practically the same accuracy as HyperLS.

In all examples, the standard LS performs poorly, while ML provides the highest accuracy. We also see that ML computation may fail in the presence of large noise. The convergence of ML critically depends on the accuracy of the initialization. In the above experiments, we used the standard LS to start the FNS iterations. We confirmed that the use of our HyperLS to start the iterations significantly extends the noise range of convergence, though the computation fails sooner or later. On the other hand, HyperLS is algebraic and hence immune to the convergence problem, producing a solution close in accuracy to ML in any noise level. The Taubin approximation is clearly



(a) (b) (c)

Figure 4: RMS error vs. the standard deviation σ of noise added to each point. 1. standard LS, 2. HyperLS, 3. Taubin approximation, 4. ML. The dotted lines indicate the KCR lower bound. (a) Ellipse fitting. (b) Fundamental matrix computation. (c) Homography computation.

inferior to HyperLS for ellipse fitting but is almost equivalent to HyperLS for fundamental matrices and homographies. This reflects that fact that while $\boldsymbol{\xi}$ is *quadratic* in x and y for ellipses (see (4)), the corresponding $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^{(k)}$ are *bilinear* in x, y, x', and y' for fundamental matrices (see (6)) and homographies (see (9)), so $\boldsymbol{e}_\alpha^{(k)}$ in (47) is $\mathbf{0}$.

In computer vision applications, we frequently do inference from multiple images based on "multilinear" constraints involving homographies, fundamental matrices, trifocal tensors, and other geometric quantities (Hartley and Zisserman, 2004). In such a situation, although the constraint itself is nonlinear, it is linear in observations of each image, so we have $\boldsymbol{e}_\alpha^{(k)} = \mathbf{0}$, because noise in different images are assumed to be independent. In such a problem, the accuracy of HyperLS is nearly the same as its Taubin approximation. However, HyperLS is expected to be clearly superior if the constraint involves nonlinear terms in observations of the same image.

11. CONCLUSING REMARKS

We presented a new form of least squares (LS), which we call "HyperLS", for geometric problems that appear in computer vision applications. Doing rigorous error analysis, we maximized the accuracy by introducing a normalization that eliminates statistical bias up to second order noise terms. Numerical experiments for computing ellipses, fundamental matrices, and homographies show that our method yields a solution far superior to the standard LS and comparable in accuracy to ML, which is known to produce highly accurate solutions but may fail to converge if poorly initialized. Thus, our HyperLS is a perfect candidate for ML initialization. We compared the performance of HyperLS and its Taubin approximation and attributed the performance differences to the structure of the problem. We have also discussed in detail how image-based inference problems have characteristics very different form conventional statistical applications, with a view to serving as a bridge between mathematicians and computer engineers.

In this paper, we analyzed only the leading covariance term, which is $O(\sigma^2)$, and the leading bias term, which is also $O(\sigma^2)$. Due to technical difficulties, we are unable to evaluate at this stage how higher order terms of $O(\sigma^4)$ affect the solution. This is a remaining issue to be studied in the future.

Acknowledgments. The authors thank Ali Al-Sharadqah and Nikolai Chernov of the University of Alabama at Birmingham, U.S.A, Wolfgang Förstner of the University of Bonn, Germany, and Alexander Kukush of National Taras Shevchenko University of Kyiv, Ukraine, for helpful discussions. This work was supported in part by the Ministry of Education, Culture, Sports, Science, and Technology, Japan, under a Grant in Aid for Scientific Research (C 21500172).

REFERENCES

Al-Sharadqah, A. Chernov, N. (2009). Error analysis for circle fitting algorithms. Elec. J. Stat. 3:886–911.

Cheng, C.-L., A. Kukush, A. (2006). Non-existence of the first moment of the adjusted least squares estimator in multivariate errors-in-variables model. Metrika 64:41–46.

Chernov, N., Lesort, C. (2004). Statistical efficiency of curve fitting algorithms. Comput. Stat. Data Anal. 47:713–728.

Chojnacki, W., Brooks, M.J., van den Hengel, A., Gawley, D. (2000). On the fitting of surfaces to data with covariances. IEEE Trans. Patt. Anal. Mach. Intell. 22:1294–1303.

Hartley, R., Kahl, F. (2007). Optimal algorithms in multiview geometry. In: Proc. 8th Asian Conf. Computer Vision, Tokyo, Japan 1:13–34.

Hartley, R., Zisserman, A. (2004). Multiple View Geometry in Computer Vision. 2nd ed., Cambridge: Cambridge University Press.

Kanatani, K. (1996). Statistical Optimization for Geometric Computation: Theory and Practice. Amsterdam: Elsevier Science; reprinted, New York: Dover (2005).

Kanatani, K. (2006). Ellipse fitting with hyperaccuracy. IEICE Trans. Inf. & Syst. E89-D:2653–2660.

Kanatani, K. (2006). Statistical optimization for geometric fitting: Theoretical accuracy analysis and high order error analysis. Int. J. Comput. Vision. 80:167–188.

Kanatani, K., Niitsuma, H. (2010). Optimal two-view planar scene triangulation. In: Proc. 10th Asian Conf. Comput. Vision. Queenstown, New Zealand.

Kanatani, K., Rangarajan, P. (2010). Hyperaccurate algebraic ellipse fitting without iterations. In: Proc. 5th Int. Conf. Comput. Vision Theory Appl. Angers, France.

Kanatani, K., Sugaya, Y. (2007). Performance evaluation of iterative geometric fitting algorithms. Comp. Stat. Data Anal. 52:1208–1222.

Kanatani, K., Sugaya, Y. (2008). Compact algorithm for strictly ML ellipse fitting. In: Proc. 19th Int. Conf. Pattern Recog. Tampa, FL, U.S.A.

Kanatani, K., Sugaya, Y. (2010a). Compact fundamental matrix computation. IPSJ Trans. Comput. Vision Appl. 2:59–70.

Kanatani, K., Sugaya, Y. (2010b). Unified computation of strict maximum likelihood for geometric fitting. J. Math. Imaging Vision. 38:1–13.

Kukush, K., Markovski, I., Van Huffel, S. (2002). Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. Comp. Stat. Data Anal. 41:3–18.

Kukush, K., Markovski, I., Van Huffel, S. (2004). Consistent estimation in an implicit quadratic measurement error model. Comp. Stat. Data Anal. 47:123–147.

Leedan, Y., and Meer, P. (2000). Heteroscedastic regression in computer vision: Problems with bilinear constraint. Int. J. Comput. Vision. 37:127–150.

Matei, B.C., and Meer, P. (2006). Estimation of nonlinear errors-in-variables models for computer vision applications. IEEE Trans. Patt. Anal. Mach. Intell. 28:1537–1552.

Neyman, J., Scott, E.L. (1948). Consistent estimates based on partially consistent observations. Econometrica. 16:1–32.

Niitsuma, H., Rangarajan, P., Kanatani, K. (2010). High accuracy homography computation without iterations. In: Proc. 16th Symp. Sensing Imaging Inf. Yokohama, Japan.

Rangarajan, P., Kanatani, K. (2009). Improved algebraic methods for circle fitting. Elec. J. Stat. 3:1075–1082.

Sampson, P.D. (1982). Fitting conic sections to "very scattered" data: An iterative refinement of the Bookstein algorithm. Comput. Graphics Image Process. 18:97–108.

Taubin, G. (1991). Estimation of planar curves, surfaces, and non-planar space curves defined by implicit equations with applications to edge and range image segmentation. IEEE Trans. Patt. Anal. Mach. Intell. 13:1115–1138.

Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A. (2000). Bundle adjustment—A modern synthesis. In: in Triggs, B, Zisserman, A., Szeliski, R. eds., Vision Algorithms: Theory and Practice. Springer, 298–375.

Appendix

The term $E[\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}]$ is computed as follows:

$$
\begin{aligned}
&E[\Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}]\\
&= E[\frac{1}{N}\sum_{\alpha=1}^{N}\sum_{k=1}^{3}\Big(\bar{\boldsymbol{\xi}}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top}+\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(k)\top}\Big)\bar{\boldsymbol{M}}^-\frac{1}{N}\sum_{\beta=1}^{N}\sum_{l=1}^{3}\Big(\bar{\boldsymbol{\xi}}_\beta^{(l)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\Delta_1\boldsymbol{\xi}_\beta^{(l)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\Big)]\\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N}\sum_{k,l=1}^{3}E[(\bar{\boldsymbol{\xi}}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top}+\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(k)\top})\bar{\boldsymbol{M}}^-(\bar{\boldsymbol{\xi}}_\beta^{(l)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\Delta_1\boldsymbol{\xi}_\beta^{(l)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top})]\\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N}\sum_{k,l=1}^{3}E[\bar{\boldsymbol{\xi}}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top}\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\bar{\boldsymbol{\xi}}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{\xi}_\beta^{(l)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\\
&\quad+\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(k)\top}\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(k)\top}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{\xi}_\beta^{(l)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}]\\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N}\sum_{k,l=1}^{3}E[\bar{\boldsymbol{\xi}}_\alpha^{(k)}(\Delta_1\boldsymbol{\xi}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)})\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\bar{\boldsymbol{\xi}}_\alpha^{(k)}(\Delta_1\boldsymbol{\xi}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{\xi}_\beta^{(l)})\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\\
&\quad+\Delta_1\boldsymbol{\xi}_\alpha^{(k)}(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)})\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\Delta_1\boldsymbol{\xi}_\alpha^{(k)}(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{\xi}_\beta^{(l)})\bar{\boldsymbol{\xi}}_\beta^{(l)\top}]\\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N}\sum_{k,l=1}^{3}E[(\Delta_1\boldsymbol{\xi}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)})\bar{\boldsymbol{\xi}}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+(\Delta_1\boldsymbol{\xi}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{\xi}_\beta^{(l)})\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\\
&\quad+(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)})\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\Delta_1\boldsymbol{\xi}_\alpha^{(k)}(\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{\xi}_\beta^{(l)},\bar{\boldsymbol{\xi}}_\alpha^{(k)})\bar{\boldsymbol{\xi}}_\beta^{(l)\top}]\\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N}\sum_{k,l=1}^{3}E[\bar{\boldsymbol{\xi}}_\alpha^{(k)}((\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)})^\top\Delta_1\boldsymbol{\xi}_\alpha^{(k)})\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\mathrm{tr}[\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{\xi}_\beta^{(l)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top}]\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\\
&\quad+(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)})\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}+\Delta_1\boldsymbol{\xi}_\alpha^{(k)}(\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(k)})\bar{\boldsymbol{\xi}}_\beta^{(l)\top}]\\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N}\sum_{k,l=1}^{3}\Big(\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\bar{\boldsymbol{M}}^-E[\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}]+\mathrm{tr}[\bar{\boldsymbol{M}}^-E[\Delta_1\boldsymbol{\xi}_\beta^{(l)}\Delta_1\boldsymbol{\xi}_\alpha^{(k)\top}]]\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\\
&\quad+(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)})E[\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}]+E[\Delta_1\boldsymbol{\xi}_\alpha^{(k)}\Delta_1\boldsymbol{\xi}_\beta^{(l)\top}]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\Big)\\
&= \frac{\sigma^2}{N^2}\sum_{\alpha,\beta=1}^{N}\sum_{k,l=1}^{3}\Big(\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\bar{\boldsymbol{M}}^-\delta_{\alpha\beta}V_0^{(kl)}[\boldsymbol{\xi}_\alpha]+\mathrm{tr}[\bar{\boldsymbol{M}}^-\delta_{\alpha\beta}V_0^{(kl)}[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\\
&\quad+(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta^{(l)})\delta_{\alpha\beta}V_0^{(kl)}[\boldsymbol{\xi}_\alpha]+\delta_{\alpha\beta}V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\beta^{(l)\top}\Big)\\
&= \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N}\sum_{k,l=1}^{3}\Big(\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}\bar{\boldsymbol{M}}^-V_0^{(kl)}[\boldsymbol{\xi}_\alpha]+\mathrm{tr}[\bar{\boldsymbol{M}}^-V_0^{(kl)}[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}+(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(l)})V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\\
&\quad+V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}\Big)\\
&= \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N}\sum_{k,l=1}^{3}\Big(\mathrm{tr}[\bar{\boldsymbol{M}}^-V_0^{(kl)}[\boldsymbol{\xi}_\alpha]]\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}+(\bar{\boldsymbol{\xi}}_\alpha^{(k)},\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(l)})V_0^{(kl)}[\boldsymbol{\xi}_\alpha]+2\mathcal{S}[V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha^{(k)}\bar{\boldsymbol{\xi}}_\alpha^{(l)\top}]\Big).
\end{aligned}
\tag{63}
$$

Thus, (48) is obtained.