

Hyper-renormalization: Non-minimization Approach for Geometric Estimation

KENICHI KANATANI^{1,a)} ALI AL-SHARADQAH^{2,b)} NIKOLAI CHERNOV³ YASUYUKI SUGAYA^{4,c)}

Received: December 17, 2013, Accepted: September 23, 2014, Released: December 10, 2014

Abstract: The technique of “renormalization” for geometric estimation attracted much attention when it appeared in early 1990s for having higher accuracy than any other then known methods. The key fact is that it directly specifies equations to solve, rather than minimizing some cost function. This paper expounds this “non-minimization approach” in detail and exploits this principle to modify renormalization so that it outperforms the standard reprojection error minimization. Doing a precise error analysis in the most general situation, we derive a formula that maximizes the accuracy of the solution; we call it *hyper-renormalization*. Applying it to ellipse fitting, fundamental matrix computation, and homography computation, we confirm its accuracy and efficiency for sufficiently small noise. Our emphasis is on the general principle, rather than on individual methods for particular problems.

Keywords: geometric estimation, non-minimization approach, error analysis, hyper-renormalization

1. Introduction

One of the most fundamental tasks of computer vision is to compute 2-D and 3-D shapes of objects from noisy observations based on “geometric constraints,” by which we mean properties that can be described by relatively simple equations such as the objects being lines or planes, their being parallel or orthogonal, and the camera imaging geometry being perspective projection. Many problems are formulated as follows. We observe N vector data $\mathbf{x}_1, \dots, \mathbf{x}_N$, whose true values $\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_N$ are supposed to satisfy equations in the form

$$F^{(k)}(\mathbf{x}; \boldsymbol{\theta}) = 0, \quad k = 1, \dots, L, \quad (1)$$

where $\boldsymbol{\theta}$ is an unknown parameter vector which we want to estimate. We call this type of problem simply “geometric estimation.” In the traditional domains of statistics such as agriculture, pharmacetics, and economics, observations are regarded as repeated “samples” from a parameterized probability density $p_{\boldsymbol{\theta}}(\mathbf{x})$, which explains the underlying data generating mechanism; the task is to estimate the parameter $\boldsymbol{\theta}$. We call this type of problem “statistical estimation,” for which the minimization principle has been a major tool. Namely, one chooses the value that minimizes a specified cost function. The best known approach is *maximum*

likelihood (ML), which minimizes the negative log-likelihood $l = -\sum_{\alpha=1}^N \log p_{\boldsymbol{\theta}}(\mathbf{x}_{\alpha})$. Recently, an alternative approach is more and more in use: one directly solves specified equations, called *estimating equations* [5], in the form of $\mathbf{g}(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\theta}) = \mathbf{0}$. This approach can be viewed as an extension of the minimization principle; ML corresponds to $\mathbf{g}(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} l$, known as the *score*. However, the estimating equations need not be the gradient of any function; one can modify $\mathbf{g}(\mathbf{x}_1, \dots, \mathbf{x}_N, \boldsymbol{\theta})$ as one likes so that the resulting solution $\boldsymbol{\theta}$ should have desirable properties (unbiasedness, consistency, efficiency, etc.). In this sense, the estimating equation approach is more general and flexible, having the possibility of providing a better solution than the minimization principle.

In the domain of computer vision, the minimization principle, in particular reprojection error minimization, is currently the norm and called the *Gold Standard* [7]. A notable exception is *renormalization* of Kanatani [8], [9]: instead of minimizing some cost function, it iteratively removes bias of weighted least squares. It attracted much attention because it exhibited higher accuracy than any other then known methods. However, questions were repeatedly raised as to what it minimizes, perhaps out of the deep-rooted preconception that optimal estimation should minimize something. One answer was given by Chojnacki et al. [4], who proposed an iterative scheme similar to renormalization, which they called *FNS (Fundamental Numerical Scheme)*, for minimizing what is now referred to as the Sampson error [7]. Chojnacki et al. [3] argued that renormalization can be “rationalized” if viewed as approximately minimizing the Sampson error. Leedan and Meer [19] and Matei and Meer [20] also proposed a different iterative scheme, which they called *HEIV (Heteroscedastic Errors-in-Variables)*, for minimizing the Sampson error. Kanatani and Sugaya [17] pointed out that the reprojection error can be minimized by repeated applications of Sampson error minimization: the Sampson error is iteratively modified so that it agrees with the reprojection error in the end. However, reprojection error

¹ Professor Emeritus, Okayama University, Okayama 700-8530, Japan

² Department of Mathematics, East Carolina University, Greenville, NC 27858-3453, U.S.A.

³ Formally, Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294-1170, U.S.A., deceased August 7, 2014

⁴ Department of Computer Science and Engineering, Toyohashi University of Technology, Toyohashi, Aichi 441-8580, Japan

a) kanatani2013@yahoo.co.jp

b) alsharadqaha@ecu.edu

c) sugaya@iim.cs.tut.ac.jp

minimization, which is ML in statistical terms, still has some bias. Kanatani [10], [11] and Kanatani and Sugaya [18] analytically evaluated the bias of the FNS solution and subtracted it; he called his scheme *hyperaccurate correction*. Okatani and Deguchi removed the bias of ML of particular types by analyzing the hypersurface defined by the constraint [21] and by using the method of projected scores [22].

However, bias correction is already a departure from the minimization principle. Namely, we correct the $\hat{\theta}$ that minimizes the cost function so that it approaches the true value $\bar{\theta}$ in expectation. As a result, *the value of the cost function increases (Fig. 1)*. Then, why not directly compute, by some means, that optimal value from the beginning? This is the motivation of the “non-minimization approach,” directly specifying equations to solve rather than minimizing some cost function. In this paper, we expound this principle in detail and modify renormalization so that it outperforms the standard reprojection error minimization. Doing a precise high order error analysis using the perturbation technique of Kanatani [11] and Al-Sharadqah and Chernov [1], we derive a formula that maximizes the accuracy of the solution; we call it *hyper-renormalization*. Partly, this has already been done in the single constraint case, such as ellipse fitting, by Kanatani et al. [12]. Here, we extend this principle to the most general multiple constraint case.

In Section 2, we summarize the fundamentals of geometric estimation. In Section 3, we state our noise modeling and in Section 4 introduce our non-minimization approach. In Section 5, we describe the iterative reweight, the most primitive form of the non-minimization approach. In Section 6, we reformulate Kanatani’s renormalization, which can be interpreted as an iterative improvement of the method of Taubin [26]. In Sections 7 and 8, we do a detailed error analysis of the problem. In Section 9, we derive the procedure of hyper-renormalization as an iterative improvement of what is called HyperLS [13], [14], [24]. In Section 10, we summarize the non-minimization approach. In Section 11, we apply our technique to ellipse fitting, fundamental matrix computation, and homography computation to demonstrate that it indeed outperforms reprojection error minimization. In Section 12, we conclude that hyper-renormalization is the best strategy we can take in terms of accuracy and efficiency for sufficiently small noise. Our emphasis is on the general principle, rather than on individual methods for particular problems.

2. Geometric Estimation

The functions $F^{(k)}(\mathbf{x}; \boldsymbol{\theta})$ in Eq. (1) is generally nonlinear in \mathbf{x} and $\boldsymbol{\theta}$, but in many practical problems the problem can be reparameterized so that $F^{(k)}(\mathbf{x}; \boldsymbol{\theta})$ linear in $\boldsymbol{\theta}$ (but nonlinear in \mathbf{x}), allowing us to write Eq. (1) as

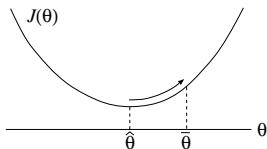


Fig. 1 If the value $\hat{\theta}$ that minimizes the cost function $J(\theta)$ is biased from the true value $\bar{\theta}$, we correct $\hat{\theta}$ so that it approaches $\bar{\theta}$ in expectation. As a result, the value of $J(\theta)$ increases.

$$\boldsymbol{\xi}^{(k)}(\mathbf{x}, \boldsymbol{\theta}) = 0, \quad k = 1, \dots, L, \quad (2)$$

where and hereafter (\mathbf{a}, \mathbf{b}) denotes the inner product of vectors \mathbf{a} and \mathbf{b} . This is made possible by regarding the “coefficients” of multiple nonlinear terms in $F^{(k)}(\mathbf{x}; \boldsymbol{\theta})$ as the unknown parameter vector $\boldsymbol{\theta}$ (see the examples below). The vector $\boldsymbol{\xi}^{(k)}(\mathbf{x})$ is some nonlinear mapping of \mathbf{x} from \mathcal{R}^m to \mathcal{R}^n , where m and n are the dimensions of the data \mathbf{x}_α and the parameter $\boldsymbol{\theta}$, respectively, and the i th component $\xi_i^{(k)}(\mathbf{x})$ of $\boldsymbol{\xi}^{(k)}(\mathbf{x})$ is the term in Eq. (1) that is multiplied by the i th component θ_i of $\boldsymbol{\theta}$. Those terms that do not involve $\boldsymbol{\theta}$ are regarded as multiplied by a constant, which we also regard as one component of $\boldsymbol{\theta}$, as shown in the examples below. We further assume that the L vectors $\boldsymbol{\xi}^{(k)}(\mathbf{x})$ need not be linearly independent. We call the number r of independent ones the *rank* of the constraint. Since the vector $\boldsymbol{\theta}$ in Eq. (2) has scale indeterminacy, we normalize it to unit norm: $\|\boldsymbol{\theta}\| = 1$.

Example 1 (Ellipse fitting). Given a point sequence (x_α, y_α) , $\alpha = 1, \dots, N$, we wish to fit an ellipse of the form (Fig. 2(a))

$$Ax^2 + 2Bxy + Cy^2 + 2f_0(Dx + Ey) + f_0^2 F = 0, \quad (3)$$

where f_0 is a fixed constant. If we let

$$\boldsymbol{\xi} = (x^2, 2xy, y^2, 2f_0x, 2f_0y, f_0^2)^\top, \quad \boldsymbol{\theta} = (A, B, C, D, E, F)^\top, \quad (4)$$

the ellipse equation has the form of Eq. (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* [7]

$$\left(\begin{array}{c} x \\ y \\ f_0 \end{array} \right), \mathbf{F} \left(\begin{array}{c} x' \\ y' \\ f_0 \end{array} \right) = 0, \quad (5)$$

where \mathbf{F} is a matrix of rank 2 called the *fundamental matrix*, from which we can compute the camera positions and the 3-D structure of the scene [9] (Fig. 2(b)), and f_0 is a fixed constant. If we let

$$\boldsymbol{\xi} = (xx', xy', f_0x, yx', yy', f_0y, f_0x', f_0y', f_0^2)^\top, \quad \boldsymbol{\theta} = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^\top, \quad (6)$$

the epipolar equation has the form of Eq. (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

$$x' = f_0 \frac{h_{11}x + h_{12}y + h_{13}f_0}{h_{31}x + h_{32}y + h_{33}f_0}, \quad y' = f_0 \frac{h_{21}x + h_{22}y + h_{23}f_0}{h_{31}x + h_{32}y + h_{33}f_0}, \quad (7)$$

where f_0 is a fixed constant. In matrix form, Eq. (7) is equivalently rewritten as

$$\left(\begin{array}{c} x' \\ y' \\ f_0 \end{array} \right) \simeq \left(\begin{array}{ccc} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{array} \right) \left(\begin{array}{c} x \\ y \\ f_0 \end{array} \right), \quad (8)$$

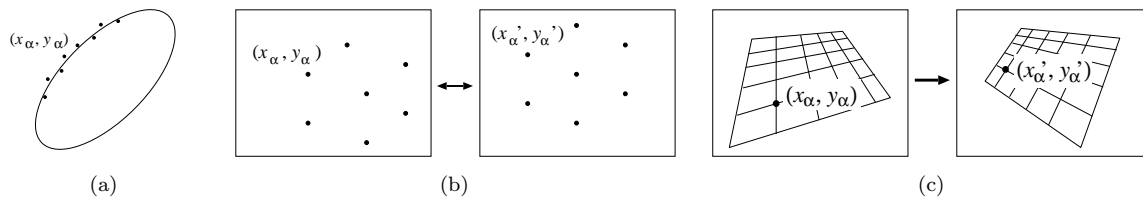


Fig. 2 (a) Fitting an ellipse to a point sequence. (b) Computing the fundamental matrix from corresponding points between two images. (c) Computing a homography between two images.

where \simeq denotes equality up to a nonzero multiplier. Noting that Eq. (8) means that the vectors on both sides are parallel to each other, we can alternatively write

$$\begin{pmatrix} x' \\ y' \\ f_0 \end{pmatrix} \times \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ f_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (9)$$

If we let

$$\begin{aligned} \boldsymbol{\xi}^{(1)} &= (0, 0, 0, -f_0x, -f_0y, -f_0^2, xy', yy', f_0y')^\top, \\ \boldsymbol{\xi}^{(2)} &= (f_0x, f_0y, f_0^2, 0, 0, 0, -xx', -yy', -f_0x')^\top, \\ \boldsymbol{\xi}^{(3)} &= (-xy', -yy', -f_0y', xx', yy', f_0x', 0, 0, 0)^\top, \end{aligned} \quad (10)$$

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

the three components of Eq. (9) have the form of Eq. (2) with $L = 3$. Note that $\boldsymbol{\xi}^{(1)}$, $\boldsymbol{\xi}^{(2)}$, and $\boldsymbol{\xi}^{(3)}$ in Eq. (10) are *linearly dependent*; only two of them are independent, so the rank is $r = 2$.

In the above examples, the constant f_0 is a reference length to stabilize numerical computation. It has been well known that care is necessary to minimize the inevitable information loss caused by computation with a limited number of significant digits. A widely known rule of thumb is to subtract a common value from the data so that each datum has as many different digits as possible and to measure all physical quantities relative to a fixed reference scale so that the values used in the computation have approximately the same order of magnitude. This was reiterated by Hartley [6] for fundamental matrix computation. Here, we choose f_0 to be of the approximate magnitude of the data x_α and y_α , for which we define an xy image coordinate system with origin at the center of the frame.

Note that our analysis is based on a particular description of the problem at hand with particular parameterization in the form of Eq. (2). For example, an ellipse can be described in many different forms other than Eq. (3), e.g., in polar coordinates, and we may be able to do better estimation with another description. In this paper, however, we start our analysis with a *given* description and try to obtain a solution as accurately as possible for *that* description. We do not go into the issue of how we can best describe the problem; we merely adopt forms commonly used in practice.

3. Noise Modeling

The goal of our geometric estimation is to infer the values that *we would obtain if observation were noiseless* based on the geometric constraints and the statistical properties of the noise. Hence, the assumption on the noise is essential. In the context of image analysis, however, “noise” means *uncertainty of image*

processing operations, rather than random fluctuations over time or space as commonly understood in physics and communications. This reflects the fact that available image processing operations are not perfect and do not necessarily output exactly what we are looking for.

In this paper, we assume that data acquisition using image processing operations such as feature extraction and edge detection is, although not perfect, fairly accurate and obtained data concentrate around their true values. In other words, we assume that the noise is fairly small. We model such situations by regarding each datum \mathbf{x}_α as deviated from its true value $\bar{\mathbf{x}}_\alpha$ by independent Gaussian noise of mean $\mathbf{0}$ and covariance matrix $V[\mathbf{x}_\alpha] = \sigma^2 V_0[\mathbf{x}_\alpha]$, where $V_0[\mathbf{x}_\alpha]$ is a known matrix that specifies the directional dependence of the noise and σ is an unknown constant that specifies the absolute magnitude; we call $V_0[\mathbf{x}_\alpha]$ the *normalized covariance matrix*, and σ the *noise level*.

The separation of $V[\mathbf{x}_\alpha]$ into σ^2 and $V_0[\mathbf{x}_\alpha]$ is merely a matter of convenience; there is no fixed rule. This convention is motivated by the fact that estimation of the absolute magnitude of data uncertainty is very difficult in practice, while optimal estimation can be done, as shown shortly, only from the knowledge of $V_0[\mathbf{x}_\alpha]$. If the noise distribution is homogeneous, i.e., the same for all \mathbf{x}_α , and isotropic, i.e. the same for all directions, we can let $V_0[\mathbf{x}_\alpha] = \mathbf{I}$ (the identity).

In real applications, there are many situations to which the Gaussian noise modeling does not apply. In fact, data uncertainty may not be modelled by any probability distribution, and we need to consider erroneous measurements generally known as “outliers.” In situations where all the data are fairly accurate, on the other hand, we may safely regard the uncertainty as having Gaussian or approximately Gaussian distributions. In this paper, we consider such situations and do not go into the issue of outliers.

4. Non-minimization Approach

We reiterate our goal: appropriately assuming the statistical properties of the noise, we infer from noisy observations the values that should satisfy given constraints in the absence of noise. This problem could be reduced to minimization of some cost function $J(\boldsymbol{\theta})$, but *we need not necessarily do so*.

The minimization approach relies on the knowledge that the value $\boldsymbol{\theta}$ that minimizes the given cost function $J(\boldsymbol{\theta})$ is expected, either theoretically or experimentally, to be close to the true value $\bar{\boldsymbol{\theta}}$. However, how close it is is determined by the function $J(\boldsymbol{\theta})$ and is not of our control. In contrast, the non-minimization approach directly gives a procedure for computing $\boldsymbol{\theta}$, for which the following three stages are involved:

- (1) Devise a scheme for computing $\bar{\boldsymbol{\theta}}$ from the data \mathbf{x}_α , assuming that they have noiseless values $\bar{\mathbf{x}}_\alpha$.

- (2) Substitute $\mathbf{x}_\alpha = \bar{\mathbf{x}}_\alpha + \Delta\mathbf{x}_\alpha$ in the scheme and evaluate the error $\Delta\boldsymbol{\theta}$ of the computed value $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}} + \Delta\boldsymbol{\theta}$ in terms of the noise components $\Delta\mathbf{x}_\alpha$.
- (3) Modify the scheme so that $\Delta\boldsymbol{\theta}$ is minimized in some sense, e.g., reducing its RMS and/or bias.

The core is Stage 1, which makes the non-minimization approach possible, because *geometry is exact if the observation is exact*. Namely, we can compute the true geometry for exact data. This is a big contrast to the traditional statistical domains, such as agriculture, pharmaceuticals, and economics, in which what is going on is uncertain from the beginning, hence statistical inference is called for. As a result, such concepts as “exact observation” and “exact solution” do not make sense.

Stage 2 can be done by the well known perturbation analysis, also called “error propagation,” because *we know the underlying exact geometry*, although usually tedious and messy calculus is required. The perturbation analysis is done by expanding quantities in the noise level σ and omitting higher order terms in σ . This is justified because we are focusing on the uncertainty behavior in a small noise range.

Stage 3 is not obvious and is the most difficult part; we need to introduce ingenious and clever tricks. For example, if the scheme contains a constant that does not affect the final result if all the data are exact *yet influences the computation in the presence of noise*, its value is optimally chosen so that the highest accuracy is achieved.

One side effect is this non-minimization formalism based on error analysis is that when the final computational procedure is presented, it is often difficult to grasp its intuitive meaning, as we will see later. Perhaps, this is the main reason that this approach has been not so popular or widely accepted in the domain of computer vision.

Since we write the constraint in the form of Eq. (2), we work in the domain of $\boldsymbol{\xi}(\mathbf{x})$ rather than \mathbf{x} . Let us write $\boldsymbol{\xi}(\mathbf{x}_\alpha)$ simply as $\boldsymbol{\xi}_\alpha$. It can be expanded in the form

$$\boldsymbol{\xi}_\alpha^{(k)} = \bar{\boldsymbol{\xi}}_\alpha^{(k)} + \Delta_1\boldsymbol{\xi}_\alpha^{(k)} + \Delta_2\boldsymbol{\xi}_\alpha^{(k)} + \dots, \quad (12)$$

where and hereafter bars indicate terms without noise and the symbol Δ_m means m th order terms $O(\sigma^m)$. Using the Jacobian matrix of the mapping $\boldsymbol{\xi}^{(k)}(\mathbf{x})$, we can express the first order noise term $\Delta_1\boldsymbol{\xi}_\alpha^{(k)}$ in terms of the original noise terms $\Delta\mathbf{x}_\alpha$ as follows:

$$\Delta_1\boldsymbol{\xi}_\alpha^{(k)} = \mathbf{T}_\alpha^{(k)} \Delta\mathbf{x}_\alpha, \quad \mathbf{T}_\alpha^{(k)} \equiv \left. \frac{\partial \boldsymbol{\xi}^{(k)}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\bar{\mathbf{x}}_\alpha}. \quad (13)$$

We define the covariance matrix $V^{(kl)}[\boldsymbol{\xi}_\alpha]$ between $\boldsymbol{\xi}_\alpha^{(k)}$ and $\boldsymbol{\xi}_\alpha^{(l)}$ by

$$V^{(kl)}[\boldsymbol{\xi}_\alpha] = E[\Delta_1\boldsymbol{\xi}_\alpha^{(k)} \Delta_1\boldsymbol{\xi}_\alpha^{(l)\top}] = \sigma^2 V_0^{(kl)}[\boldsymbol{\xi}_\alpha], \quad (14)$$

where $E[\cdot]$ denotes expectation over data uncertainty. We have

$$V_0^{(kl)}[\boldsymbol{\xi}_\alpha] = \mathbf{T}_\alpha^{(k)} V_0[\mathbf{x}_\alpha] \mathbf{T}_\alpha^{(l)\top}. \quad (15)$$

Example 4 (Ellipse fitting). The first order noise term $\Delta_1\boldsymbol{\xi}_\alpha$ is

$$\Delta_1\boldsymbol{\xi}_\alpha = \mathbf{T}_\alpha \begin{pmatrix} \Delta x_\alpha \\ \Delta y_\alpha \end{pmatrix}, \quad \mathbf{T}_\alpha = 2 \begin{pmatrix} \bar{x}_\alpha & \bar{y}_\alpha & 0 & f_0 & 0 & 0 \\ 0 & \bar{x}_\alpha & \bar{y}_\alpha & 0 & f_0 & 0 \end{pmatrix}^\top. \quad (16)$$

The second order noise term $\Delta_2\boldsymbol{\xi}_\alpha$ is

$$\Delta_2\boldsymbol{\xi}_\alpha = (\Delta x_\alpha^2, 2\Delta x_\alpha \Delta y_\alpha, \Delta y_\alpha^2, 0, 0, 0)^\top. \quad (17)$$

Example 5 (Fundamental matrix computation). The first order noise term $\Delta_1\boldsymbol{\xi}_\alpha$ is

$$\Delta_1\boldsymbol{\xi}_\alpha = \mathbf{T}_\alpha (\Delta x_\alpha, \Delta y_\alpha, \Delta x'_\alpha, \Delta y'_\alpha)^\top, \quad \mathbf{T}_\alpha = \begin{pmatrix} \bar{x}'_\alpha & \bar{y}'_\alpha & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{x}'_\alpha & \bar{y}'_\alpha & f_0 & 0 & 0 \\ \bar{x}_\alpha & 0 & 0 & \bar{y}_\alpha & 0 & 0 & f_0 & 0 \\ 0 & \bar{x}_\alpha & 0 & 0 & \bar{y}_\alpha & 0 & 0 & f_0 \end{pmatrix}^\top. \quad (18)$$

The second order noise term $\Delta_2\boldsymbol{\xi}_\alpha$ is

$$\Delta_2\boldsymbol{\xi}_\alpha = (\Delta x_\alpha \Delta x'_\alpha, \Delta x_\alpha \Delta y'_\alpha, 0, \Delta y_\alpha \Delta x'_\alpha, \Delta y_\alpha \Delta y'_\alpha, 0, 0, 0)^\top. \quad (19)$$

Example 6 (Homography computation). The first order noise terms $\Delta_1\boldsymbol{\xi}_\alpha^{(k)}$ are

$$\Delta_1\boldsymbol{\xi}_\alpha^{(k)} = \mathbf{T}_\alpha^{(k)} (\Delta x_\alpha, \Delta y_\alpha, \Delta x'_\alpha, \Delta y'_\alpha)^\top, \quad \mathbf{T}_\alpha^{(1)} = \begin{pmatrix} 0 & 0 & 0 & -f_0 & 0 & 0 & \bar{y}'_\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & -f_0 & 0 & 0 & \bar{y}'_\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \bar{x}_\alpha & \bar{y}_\alpha & f_0 \end{pmatrix}^\top, \quad \mathbf{T}_\alpha^{(2)} = \begin{pmatrix} f_0 & 0 & 0 & 0 & 0 & 0 & -\bar{x}'_\alpha & 0 & 0 \\ 0 & f_0 & 0 & 0 & 0 & 0 & 0 & -\bar{x}'_\alpha & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\bar{x}_\alpha & -\bar{y}_\alpha & -f_0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^\top, \quad \mathbf{T}_\alpha^{(3)} = \begin{pmatrix} -\bar{y}'_\alpha & 0 & 0 & \bar{x}'_\alpha & 0 & 0 & 0 & 0 & 0 \\ 0 & -\bar{y}'_\alpha & 0 & 0 & \bar{x}'_\alpha & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{x}_\alpha & \bar{y}_\alpha & f_0 & 0 & 0 & 0 \\ -\bar{x}_\alpha & -\bar{y}_\alpha & -f_0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^\top. \quad (20)$$

The second order noise terms $\Delta_2\boldsymbol{\xi}_\alpha^{(k)}$ are

$$\begin{aligned} \Delta_2\boldsymbol{\xi}_\alpha^{(1)} &= (0, 0, 0, 0, 0, 0, \Delta x_\alpha \Delta y'_\alpha, \Delta y_\alpha \Delta y'_\alpha, 0)^\top, \\ \Delta_2\boldsymbol{\xi}_\alpha^{(2)} &= (0, 0, 0, 0, 0, 0, -\Delta x'_\alpha \Delta x_\alpha, -\Delta x'_\alpha \Delta y_\alpha, 0)^\top, \\ \Delta_2\boldsymbol{\xi}_\alpha^{(3)} &= (-\Delta y'_\alpha \Delta x_\alpha, -\Delta y'_\alpha \Delta y_\alpha, 0, \Delta x'_\alpha \Delta x_\alpha, \Delta x'_\alpha \Delta y_\alpha, \\ & \quad 0, 0, 0, 0)^\top. \end{aligned} \quad (21)$$

Note that the true values $\bar{\mathbf{x}}_\alpha$ are used in Eq. (15). In actual computation, they are replaced by their observations \mathbf{x}_α . It has been confirmed by many experiments that this does not affect the final result of practical problems. Also, $V_0^{(kl)}[\boldsymbol{\xi}_\alpha]$ takes only the first order noise terms into account via the Jacobian matrix, but it has been confirmed by many experiments that incorporation of higher order terms does not affect the final result.

5. Iterative Reweight

The oldest method that is not based on minimization is *iterative reweight*, which can be formulated as follows:

- (1) Let $W_\alpha^{(kl)} = \delta_{kl}$ (the Kronecker delta), $\alpha = 1, \dots, N$, and $\boldsymbol{\theta}_0 = \mathbf{0}$.
- (2) Compute the following matrix \mathbf{M} :

$$\mathbf{M} = \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} \boldsymbol{\xi}_\alpha^{(k)} \boldsymbol{\xi}_\alpha^{(l)\top}. \quad (22)$$

(3) Solve the eigenvalue problem

$$\mathbf{M}\boldsymbol{\theta} = \lambda\boldsymbol{\theta}, \quad (23)$$

and compute the unit eigenvector $\boldsymbol{\theta}$ for the smallest eigenvalue λ .

(4) If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha^{(kl)} \leftarrow \left((\boldsymbol{\theta}, V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}) \right)_r^-, \quad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \quad (24)$$

and go back to Step 2.

The expression $\left((\boldsymbol{\theta}, V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}) \right)_r^-$ in Eq. (24) denotes the (kl) element of the pseudoinverse of truncated rank r of the matrix whose (kl) element is $(\boldsymbol{\theta}, V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})$; “truncated rank” r means that all eigenvalues except the largest r are replaced by 0 in the spectral decomposition.

The motivation of this method is the weighted least squares that minimizes

$$\begin{aligned} & \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)}(\boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta})(\boldsymbol{\xi}_\alpha^{(l)}, \boldsymbol{\theta}) \\ &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} \boldsymbol{\theta}^{(k)\top} \boldsymbol{\xi}_\alpha^{(k)} \boldsymbol{\xi}_\alpha^{(l)\top} \boldsymbol{\theta} = (\boldsymbol{\theta}, \mathbf{M}\boldsymbol{\theta}). \end{aligned} \quad (25)$$

This is minimized by the unit eigenvector $\boldsymbol{\theta}$ of the matrix \mathbf{M} for the smallest eigenvalue. Note that the weight $W_\alpha^{(ki)}$ does not affect the final result if all the data are exact *yet influences the computation in the presence of noise*. According to statistics, the optimal choice of the weight $W_\alpha^{(kl)}$ in the presence of noise is the inverse of the covariance of that term up to scale. Since $(\boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta}) = (\Delta_1 \boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta}) + \dots$, the leading term of the covariance is

$$\begin{aligned} & E[(\Delta_1 \boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta})(\Delta_1 \boldsymbol{\xi}_\alpha^{(l)}, \boldsymbol{\theta})] = E[\boldsymbol{\theta}^\top \Delta_1 \boldsymbol{\xi}_\alpha^{(k)} \Delta_1 \boldsymbol{\xi}_\alpha^{(l)\top} \boldsymbol{\theta}] \\ &= (\boldsymbol{\theta}, E[\Delta_1 \boldsymbol{\xi}_\alpha^{(k)} \Delta_1 \boldsymbol{\xi}_\alpha^{(l)\top}] \boldsymbol{\theta}) = \sigma^2 (\boldsymbol{\theta}, V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}). \end{aligned} \quad (26)$$

Hence, we should choose $W_\alpha^{(kl)}$ to be the (kl) element of the inverse of the matrix whose (kl) element is the above expression. However, the inverse does not exist if the L constraint equations are not independent. In the presence of noise, the L equations are independent in appearance, so we choose r (= the number of independent equations in the absence of noise) from among the L equations. This is equivalent to computing the pseudoinverse of truncated rank r . However, $\boldsymbol{\theta}$ is not known. So, we do iterations, determining the weight $W_\alpha^{(kl)}$ from the value of $\boldsymbol{\theta}$ in the preceding step.

Let us call the first value of $\boldsymbol{\theta}$ computed with $W_\alpha^{(kl)} = \delta_{kl}$ the “initial solution.” It minimizes $\sum_{\alpha=1}^N \sum_{k=1}^L (\boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta})^2$, corresponding to what is known as *least squares (LS)*, *algebraic distance minimization*, and many other names [7]. Thus, *iterative reweight is an iterative improvement of LS*.

It appears at first sight that the above procedure minimizes

$$J = \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)}(\boldsymbol{\theta})(\boldsymbol{\xi}_\alpha^{(k)}, \boldsymbol{\theta})(\boldsymbol{\xi}_\alpha^{(l)}, \boldsymbol{\theta}), \quad (27)$$

where

$$W_\alpha^{(kl)}(\boldsymbol{\theta}) = \left((\boldsymbol{\theta}, V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}) \right)_r^-. \quad (28)$$

The function in Eq. (27) is known today as the *Sampson error* [7]. This name stems from the classical ellipse fitting scheme of P. D. Sampson [25]. However, iterative reweight does not minimize the Sampson error, because we compute at each step the value of $\boldsymbol{\theta}$ that minimizes J for a *fixed* value $W_\alpha^{(kl)}(\boldsymbol{\theta})$ with the $\boldsymbol{\theta}$ determined in the preceding step.

The perturbation analysis by Kanatani [11] shows that the covariance matrix $V[\boldsymbol{\theta}]$ of the resulting solution $\boldsymbol{\theta}$ agrees with a theoretical accuracy limit, called the *KCR (Kanatani-Cramer-Rao) lower bound* [2], [9], [11], up to $O(\sigma^4)$. Hence, further covariance reduction is practically impossible. However, it has been widely known that the iterative reweight solution has a large bias [9]. For ellipse fitting, for example, it almost always returns a smaller ellipse than the true shape. Thus, the following strategies were introduced to improve iterative reweight:

- Remove the bias of the solution.
- Exactly minimize the Sampson error in Eq. (27).

The former is Kanatani’s renormalization [8], [9] and the latter is the FNS of Chojnacki et al. [4] and the HEIV of Leedan and Meer [19] and Matei and Meer [20].

6. Renormalization

Kanatani’s renormalization [8], [9] can be described as follows:

- (1) Let $W_\alpha^{(kl)} = \delta_{kl}$, $\alpha = 1, \dots, N$, and $\boldsymbol{\theta}_0 = \mathbf{0}$.
- (2) Compute the following matrices \mathbf{M} and \mathbf{N} :

$$\mathbf{M} = \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} \boldsymbol{\xi}_\alpha^{(k)} \boldsymbol{\xi}_\alpha^{(l)\top}, \quad (29)$$

$$\mathbf{N} = \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} V_0^{(kl)}[\boldsymbol{\xi}_\alpha]. \quad (30)$$

- (3) Solve the generalized eigenvalue problem

$$\mathbf{M}\boldsymbol{\theta} = \lambda\mathbf{N}\boldsymbol{\theta}, \quad (31)$$

and compute the unit eigenvector $\boldsymbol{\theta}$ for the smallest eigenvalue λ in absolute value.

- (4) If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha^{(kl)} \leftarrow \left((\boldsymbol{\theta}, V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}) \right)_r^-, \quad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \quad (32)$$

and go back to Step 2.

This procedure has a different appearance from those in Ref. [9], in which the generalized eigenvalue problem is reduced to the standard eigenvalue problem, but the resulting solution is the same.

The motivation of renormalization is as follows. Let $\bar{\mathbf{M}}$ be the true value of the matrix \mathbf{M} in Eq. (29) defined by the true values $\bar{\boldsymbol{\xi}}_\alpha^{(k)}$. Since $(\bar{\boldsymbol{\xi}}_\alpha^{(k)}, \boldsymbol{\theta}) = 0$, we have $\bar{\mathbf{M}}\boldsymbol{\theta} = \mathbf{0}$. Hence, $\boldsymbol{\theta}$ is the eigenvector of $\bar{\mathbf{M}}$ for eigenvalue 0, but $\bar{\mathbf{M}}$ is unknown. So, we estimate it. Since $E[\Delta_1 \boldsymbol{\xi}_\alpha^{(k)}] = 0$, the expectation of \mathbf{M} is to a first approximation

$$\begin{aligned} E[\mathbf{M}] &= E\left[\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} (\bar{\boldsymbol{\xi}}_\alpha^{(k)} + \Delta_1 \boldsymbol{\xi}_\alpha^{(k)}) (\bar{\boldsymbol{\xi}}_\alpha^{(l)} + \Delta_1 \boldsymbol{\xi}_\alpha^{(l)})^\top\right] \\ &= \bar{\mathbf{M}} + \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} E[\Delta_1 \boldsymbol{\xi}_\alpha^{(k)} \Delta_1 \boldsymbol{\xi}_\alpha^{(l)\top}] \\ &= \bar{\mathbf{M}} + \frac{\sigma^2}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} V_0^{(kl)}[\boldsymbol{\xi}_\alpha] = \bar{\mathbf{M}} + \sigma^2 \mathbf{N}, \end{aligned} \quad (33)$$

where the weights $W_\alpha^{(kl)}$ are regarded as constants. Thus, $\bar{\mathbf{M}} = E[\mathbf{M}] - \sigma^2 \mathbf{N} \approx \mathbf{M} - \sigma^2 \mathbf{N}$, so instead of $\bar{\mathbf{M}}\boldsymbol{\theta} = \mathbf{0}$ we solve $(\mathbf{M} - \sigma^2 \mathbf{N})\boldsymbol{\theta} = \mathbf{0}$, or $\mathbf{M}\boldsymbol{\theta} = \sigma^2 \mathbf{N}\boldsymbol{\theta}$. Assuming that σ^2 is small, we regard it as the eigenvalue λ closest to 0. As in the case of iterative reweight, we iteratively update the weight $W_\alpha^{(kl)}$ so that it approaches $\left((\boldsymbol{\theta}, V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})\right)_r^-$.

Note that if $L = 1$, the initial solution with $W_\alpha = 1$ solves $\left(\sum_{\alpha=1}^N \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top\right)\boldsymbol{\theta} = \lambda \left(\sum_{\alpha=1}^N V_0[\boldsymbol{\xi}_\alpha]\right)\boldsymbol{\theta}$, which is nothing but the method of Taubin [26], known to be fairly accurate among algebraic methods that do not require iterations. The corresponding method for $L > 1$ has not been described, but it can be viewed as an extension of the Taubin method to multiple constraints. For simplicity, we call this extension also the *Taubin method*.

Thus, *renormalization is an iterative improvement of the Taubin solution*. According to many experiments, renormalization is shown to be more accurate than the Taubin method with nearly comparable accuracy with the FNS and the HEIV. The accuracy of renormalization is analytically evaluated by Kanatani [11], showing that the covariance matrix $V[\boldsymbol{\theta}]$ of the solution $\boldsymbol{\theta}$ agrees with the KCR lower bound up to $O(\sigma^4)$ just as iterative reweight, but the bias is much smaller.

Very small it may be, the bias is not 0. The error analysis by Kanatani [11] shows that the bias expression involves the matrix \mathbf{N} . Note that the matrix \mathbf{N} does not affect the final result if all the data are exact *yet influences the computation in the presence of noise*. This observation leads to the idea of *optimizing the matrix \mathbf{N}* in Eq. (30) to $\mathbf{N} = (1/N) \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} V_0^{(kl)}[\boldsymbol{\xi}_\alpha] + \dots$ so that *the bias is zero up to high order noise terms*. For this, we need higher order error analysis, taking the effect of the weights $W_\alpha^{(kl)}$ into consideration as well.

7. Error Analysis

Our goal is to devise an algorithm which produces a value $\boldsymbol{\theta}$ that is as close to its true value $\bar{\boldsymbol{\theta}}$ as possible. Since we are viewing the data as random variables, the solution $\boldsymbol{\theta}$ is also a random variable. Hence, we require it to be close to $\bar{\boldsymbol{\theta}}$ in the root-mean-square (RMS) sense, i.e., $\sqrt{E[\|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\|^2]}$ be small. This means that our analysis depends on the representation and parameterization of the problem, as pointed out at the end of Section 2.

Substituting Eq. (12) into the definition of the matrix \mathbf{M} in Eq. (29), we can expand it in the form

$$\mathbf{M} = \bar{\mathbf{M}} + \Delta_1 \mathbf{M} + \Delta_2 \mathbf{M} + \dots, \quad (34)$$

where $\Delta_1 \mathbf{M}$ and $\Delta_2 \mathbf{M}$ are given by

$$\begin{aligned} \Delta_1 \mathbf{M} &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(\Delta_1 \boldsymbol{\xi}_\alpha^{(k)} \bar{\boldsymbol{\xi}}_\alpha^{(l)\top} + \bar{\boldsymbol{\xi}}_\alpha^{(k)} \Delta_1 \boldsymbol{\xi}_\alpha^{(l)\top} \right) \\ &+ \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \Delta_1 \bar{W}_\alpha^{(kl)} \bar{\boldsymbol{\xi}}_\alpha^{(k)} \bar{\boldsymbol{\xi}}_\alpha^{(l)\top}, \end{aligned} \quad (35)$$

$$\begin{aligned} \Delta_2 \mathbf{M} &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(\Delta_1 \boldsymbol{\xi}_\alpha^{(k)} \Delta_1 \boldsymbol{\xi}_\alpha^{(l)\top} + \Delta_2 \boldsymbol{\xi}_\alpha^{(k)} \bar{\boldsymbol{\xi}}_\alpha^{(l)\top} \right. \\ &+ \bar{\boldsymbol{\xi}}_\alpha^{(k)} \Delta_2 \boldsymbol{\xi}_\alpha^{(l)\top} \left. \right) + \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \Delta_1 W_\alpha^{(kl)} (\Delta_1 \boldsymbol{\xi}_\alpha^{(k)} \bar{\boldsymbol{\xi}}_\alpha^{(l)\top} \\ &+ \bar{\boldsymbol{\xi}}_\alpha^{(k)} \Delta_1 \boldsymbol{\xi}_\alpha^{(l)\top}) + \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \Delta_2 W_\alpha^{(kl)} \bar{\boldsymbol{\xi}}_\alpha^{(k)} \bar{\boldsymbol{\xi}}_\alpha^{(l)\top}. \end{aligned} \quad (36)$$

Let $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}} + \Delta_1 \boldsymbol{\theta} + \Delta_2 \boldsymbol{\theta} + \dots$ be the corresponding expansion of the resulting $\boldsymbol{\theta}$. At the time of convergence, we have $W_\alpha^{(kl)} = \left((\boldsymbol{\theta}, V_0^{(kl)}[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})\right)_r^-$. Substituting the expansion of $\boldsymbol{\theta}$, we obtain the expansion $W_\alpha^{(kl)} = \bar{W}_\alpha^{(kl)} + \Delta_1 W_\alpha^{(kl)} + \Delta_2 W_\alpha^{(kl)} + \dots$, where

$$\Delta_1 W_\alpha^{(kl)} = -2 \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} (\Delta_1 \boldsymbol{\theta}, V_0^{(mn)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}), \quad (37)$$

$$\begin{aligned} \Delta_2 W_\alpha^{(kl)} &= \sum_{m,n=1}^L \Delta_1 W_\alpha^{(km)} \Delta_1 W_\alpha^{(ln)} (\bar{\boldsymbol{\theta}}, V_0^{(mn)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) \\ &- \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \left((\Delta_1 \boldsymbol{\theta}, V_0^{(mn)}[\boldsymbol{\xi}_\alpha]\Delta_1 \boldsymbol{\theta}) \right. \\ &\left. + 2(\Delta_2 \boldsymbol{\theta}, V_0^{(mn)}[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) \right). \end{aligned} \quad (38)$$

(See Appendix A.1 for the derivation.) Similarly expanding the eigenvalue λ and the matrix \mathbf{N} yet to be determined, the generalized eigenvalue problem in Eq. (31) has the form

$$\begin{aligned} &(\bar{\mathbf{M}} + \Delta_1 \mathbf{M} + \Delta_2 \mathbf{M} + \dots)(\bar{\boldsymbol{\theta}} + \Delta_1 \boldsymbol{\theta} + \Delta_2 \boldsymbol{\theta} + \dots) \\ &= (\bar{\lambda} + \Delta_1 \lambda + \Delta_2 \lambda + \dots)(\bar{\mathbf{N}} + \Delta_1 \mathbf{N} + \Delta_2 \mathbf{N} + \dots) \\ &(\bar{\boldsymbol{\theta}} + \Delta_1 \boldsymbol{\theta} + \Delta_2 \boldsymbol{\theta} + \dots). \end{aligned} \quad (39)$$

Equating the noiseless terms on both sides, we have $\bar{\mathbf{M}}\bar{\boldsymbol{\theta}} = \bar{\lambda}\bar{\mathbf{N}}\bar{\boldsymbol{\theta}}$, but since $\bar{\mathbf{M}}\bar{\boldsymbol{\theta}} = \mathbf{0}$, we have $\bar{\lambda} = 0$ (we assume that $\bar{\mathbf{N}}\bar{\boldsymbol{\theta}} \neq \mathbf{0}$ for the $\bar{\mathbf{N}}$ to be determined). Equating the first and the second order terms on both sides, we obtain the following relationships:

$$\bar{\mathbf{M}}\Delta_1 \boldsymbol{\theta} + \Delta_1 \bar{\mathbf{M}}\bar{\boldsymbol{\theta}} = \Delta_1 \lambda \bar{\mathbf{N}}\bar{\boldsymbol{\theta}}, \quad (40)$$

$$\bar{\mathbf{M}}\Delta_2 \boldsymbol{\theta} + \Delta_1 \bar{\mathbf{M}}\Delta_1 \boldsymbol{\theta} + \Delta_2 \bar{\mathbf{M}}\bar{\boldsymbol{\theta}} = \Delta_2 \lambda \bar{\mathbf{N}}\bar{\boldsymbol{\theta}}. \quad (41)$$

8. Covariance and Bias

Computing the inner product of Eq. (40) and $\bar{\boldsymbol{\theta}}$ on both sides, we have

$$(\bar{\boldsymbol{\theta}}, \bar{\mathbf{M}}\Delta_1 \boldsymbol{\theta}) + (\bar{\boldsymbol{\theta}}, \Delta_1 \bar{\mathbf{M}}\bar{\boldsymbol{\theta}}) = \Delta_1 \lambda (\bar{\boldsymbol{\theta}}, \bar{\mathbf{N}}\bar{\boldsymbol{\theta}}), \quad (42)$$

but $(\bar{\boldsymbol{\theta}}, \bar{\mathbf{M}}\Delta_1 \boldsymbol{\theta}) = (\bar{\mathbf{M}}\bar{\boldsymbol{\theta}}, \Delta_1 \boldsymbol{\theta}) = 0$ and Eq. (35) implies $(\bar{\boldsymbol{\theta}}, \Delta_1 \bar{\mathbf{M}}\bar{\boldsymbol{\theta}}) = 0$, so $\Delta_1 \lambda = 0$ (recall that we are assuming that $(\bar{\boldsymbol{\theta}}, \bar{\mathbf{N}}\bar{\boldsymbol{\theta}}) \neq 0$). The matrix $\bar{\mathbf{M}}$ has rank $n - 1$, (n is the dimension of $\boldsymbol{\theta}$), $\bar{\boldsymbol{\theta}}$ being its null vector. Hence, the product $\bar{\mathbf{M}}^- \bar{\mathbf{M}}$ equals the projection matrix $\mathbf{P}_{\bar{\boldsymbol{\theta}}}$ in the direction of $\bar{\boldsymbol{\theta}}$. It follows that by multiplying both sides of Eq. (40) by $\bar{\mathbf{M}}^-$ from left, $\Delta_1 \boldsymbol{\theta}$ is expressed as follows:

$$\Delta_1 \boldsymbol{\theta} = -\bar{\mathbf{M}}^- \Delta_1 \bar{\mathbf{M}}\bar{\boldsymbol{\theta}}. \quad (43)$$

Here, we have noted that since $\boldsymbol{\theta}$ is normalized to unit norm, $\Delta_1 \boldsymbol{\theta}$ is orthogonal to $\bar{\boldsymbol{\theta}}$ so $\mathbf{P}_{\bar{\boldsymbol{\theta}}}\Delta_1 \boldsymbol{\theta} = \Delta_1 \boldsymbol{\theta}$. From Eq. (43), we obtain the covariance matrix $V[\boldsymbol{\theta}]$ of the solution $\boldsymbol{\theta}$ is to a first approximation

$$V[\boldsymbol{\theta}] = E[\Delta_1 \boldsymbol{\theta} \Delta_1 \boldsymbol{\theta}^\top] = \frac{\sigma^2}{N} \bar{\mathbf{M}}^-. \quad (44)$$

(See Appendix A.2.) This coincides with the KCR lower bound [2], [9], [11]. Since this is a theoretical accuracy limit and since it does not include the matrix $\bar{\mathbf{N}}$, we are unable to effectively reduce the covariance by adjusting $\bar{\mathbf{N}}$. However, the total error consists

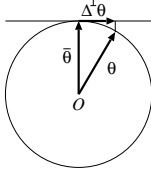


Fig. 3 The true value $\bar{\theta}$, the computed value θ , and its orthogonal component $\Delta^\perp \theta$ of θ .

of the covariance terms and the bias terms, both $O(\sigma^2)$, so we concentrate on the bias.

Substituting Eq. (43) into Eq. (41), we obtain

$$\begin{aligned} \Delta_2 \lambda \bar{N} \bar{\theta} &= \bar{M} \Delta_2 \theta - \Delta_1 M \bar{M}^{-1} \Delta_1 M \bar{\theta} + \Delta_2 M \bar{\theta} \\ &= \bar{M} \Delta_2 \theta + T \bar{\theta}, \end{aligned} \quad (45)$$

where we define the matrix T to be

$$T \equiv \Delta_2 M - \Delta_1 M \bar{M}^{-1} \Delta_1 M. \quad (46)$$

Because θ is a unit vector, it has no error in the direction of itself; we are interested in the error orthogonal to it. So, we define the second order error of θ to be its orthogonal component (**Fig. 3**)

$$\Delta_2^\perp \theta \equiv P_{\bar{\theta}} \Delta_2 \theta = \bar{M}^{-1} \bar{M} \Delta_2 \theta. \quad (47)$$

Note that the first order error $\Delta_1 \theta$ in Eq. (43) is itself orthogonal to $\bar{\theta}$. Multiplying Eq. (45) by \bar{M}^{-1} on both sides from left, we obtain $\Delta_2^\perp \theta$ in the following form:

$$\Delta_2^\perp \theta = \bar{M}^{-1} (\Delta_2 \lambda \bar{N} - T) \bar{\theta}. \quad (48)$$

Computing the inner product of Eq. (45) and $\bar{\theta}$ on both sides and noting that $(\bar{\theta}, \bar{M} \Delta_2 \theta) = 0$, we obtain $\Delta_2 \lambda$ in the form

$$\Delta_2 \lambda = \frac{(\bar{\theta}, T \bar{\theta})}{(\bar{\theta}, \bar{N} \bar{\theta})}. \quad (49)$$

Hence, Eq. (48) is rewritten as follows:

$$\Delta_2^\perp \theta = \bar{M}^{-1} \left(\frac{(\bar{\theta}, T \bar{\theta})}{(\bar{\theta}, \bar{N} \bar{\theta})} \bar{N} \bar{\theta} - T \bar{\theta} \right). \quad (50)$$

From Eq. (43), we see that $E[\Delta_1 \theta] = \mathbf{0}$, i.e., the first order bias is $\mathbf{0}$, so the bias is

$$E[\Delta_2^\perp \theta] = \bar{M}^{-1} \left(\frac{(\bar{\theta}, E[T \bar{\theta}])}{(\bar{\theta}, \bar{N} \bar{\theta})} \bar{N} \bar{\theta} - E[T \bar{\theta}] \right). \quad (51)$$

9. Hyper-renormalization

From Eq. (51), we find a crucial fact: *If we can choose such an N that its noiseless value \bar{N} satisfied $E[T \bar{\theta}] = c \bar{N} \bar{\theta}$ for some constant c , we will have*

$$E[\Delta_2^\perp \theta] = \bar{M}^{-1} \left(\frac{(\bar{\theta}, c \bar{N} \bar{\theta})}{(\bar{\theta}, \bar{N} \bar{\theta})} \bar{N} \bar{\theta} - c \bar{N} \bar{\theta} \right) = \mathbf{0}. \quad (52)$$

Then, the bias will be $O(\sigma^4)$, because the expectation of odd-order noise terms is zero. In order to choose such an N , we need to evaluate the expectation $E[T \bar{\theta}]$. After a lengthy analysis (Appendix A.3), we find that $E[T \bar{\theta}] = \sigma^2 \bar{N} \bar{\theta}$ holds if we define

$$\begin{aligned} \bar{N} &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(V_0^{(kl)} [\xi_\alpha] + 2S[\bar{\xi}_\alpha^{(k)} e_\alpha^{(l)\top}] \right) \\ &\quad - \frac{1}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} \left((\bar{\xi}_\alpha^{(k)}, \bar{M}^{-1} \bar{\xi}_\alpha^{(m)}) V_0^{(ln)} [\xi_\alpha] \right. \\ &\quad \left. + 2S[V_0^{(km)} [\xi_\alpha] \bar{M}^{-1} \bar{\xi}_\alpha^{(l)} \bar{\xi}_\alpha^{(n)\top}] \right), \end{aligned} \quad (53)$$

where $S[\cdot]$ denotes symmetrization ($S[A] = (A + A^\top)/2$) and the vectors $e_\alpha^{(k)}$ are defined via

$$E[\Delta_2 \xi_\alpha^{(k)}] = \sigma^2 e_\alpha^{(k)}. \quad (54)$$

From this result, we obtain the following *hyper-renormalization*:

- (1) Let $W_\alpha^{(kl)} = \delta_{kl}$, $\alpha = 1, \dots, N$, and $\theta_0 = \mathbf{0}$.
- (2) Compute the following matrices M and N :

$$M = \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} \xi_\alpha^{(k)} \xi_\alpha^{(l)\top}, \quad (55)$$

$$\begin{aligned} N &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(kl)} \left(V_0^{(kl)} [\xi_\alpha] + 2S[\xi_\alpha^{(k)} e_\alpha^{(l)\top}] \right) \\ &\quad - \frac{1}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L W_\alpha^{(kl)} W_\alpha^{(mn)} \left((\xi_\alpha^{(k)}, M_{n-1}^{-1} \xi_\alpha^{(m)}) \right. \\ &\quad \left. V_0^{(ln)} [\xi_\alpha] + 2S[V_0^{(km)} [\xi_\alpha] M_{n-1}^{-1} \xi_\alpha^{(l)} \xi_\alpha^{(n)\top}] \right). \end{aligned} \quad (56)$$

Here, M_{n-1}^{-1} is the pseudoinverse of M of truncated rank $n-1$.

- (3) Solve the generalized eigenvalue problem

$$M \theta = \lambda N \theta, \quad (57)$$

and compute the unit eigenvector θ for the smallest eigenvalue λ in absolute value.

- (4) If $\theta \approx \theta_0$ up to sign, return θ and stop. Else, let

$$W_\alpha^{(kl)} \leftarrow \left((\theta, V_0^{(kl)} [\xi_\alpha] \theta) \right)_r^{-1}, \quad \theta_0 \leftarrow \theta, \quad (58)$$

and go back to Step 2.

In Eq. (56), the rank-truncated pseudoinverse is used, because although the true value \bar{M} should have rank $n-1$, the computed value M generally has full rank.

It turns out that the initial solution with $W_\alpha^{(kl)} = \delta_{kl}$ coincides with what is called *HyperLS* [13], [14], [24], which is derived so that the bias is removed up to second order noise terms within the framework of algebraic methods without iterations. The expression of Eq. (56) with $W_\alpha^{(kl)} = \delta_{kl}$ lacks one term as compared with the corresponding expression of HyperLS, but the same solution is produced. Thus, *hyper-renormalization is an iterative improvement of HyperLS*. We omit the details, but all the intermediate solutions θ in the hyper-renormalization are free of second order bias.

10. Summary of the Approach

We have seen that iterative reweight, renormalization, and hyper-renormalization do not minimize any cost function. In fact, *irrespective* of their original motivations, these are the methods for computing the “solution” θ of the nonlinear equation

$$M \theta = \lambda N \theta, \quad (59)$$

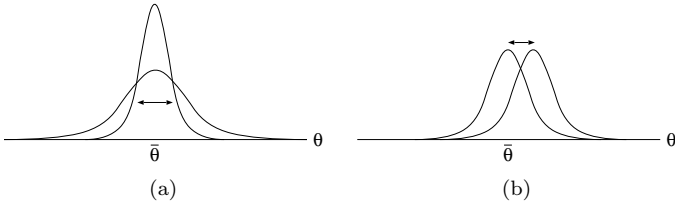


Fig. 4 (a) The matrix \mathbf{M} determines the covariance of the solution. (b) The matrix \mathbf{N} controls the bias of the solution.

Table 1 Summary of the non-minimization approach.

initial solution	final solution
least squares	iterative reweight
Taubin	renormalization
HyperLS	hyper-renormalization

where \mathbf{M} , \mathbf{N} , and λ are functions of $\boldsymbol{\theta}$. Note that Eq. (59) is a vector equation. Hence, although we can specify the symmetric matrices \mathbf{M} and \mathbf{N} as arbitrarily functions of $\boldsymbol{\theta}$, the scalar λ is indirectly determined so that Eq. (59) admits a solution; we need not specify it. Multiple values exist for such λ , and we choose the one closest to 0, because Eq. (59) is so defined that $\lambda = 0$ in the absence of noise (recall that assuming small noise, we do perturbation expansion of the function λ , see Eq. (39)).

Thus, *iterating generalized eigenvalue computation is merely a matter of convenience*; any method that solves Eq. (59) can do. In this sense, Eq. (59) corresponds to the *estimating equation* for statistical estimation, although in the traditional statistics domain it is usually a set of scalar equations, for which all terms are user specified functions. Since the data are regarded as random variables, the computed $\boldsymbol{\theta}$ is also a random variable, so it has a probability distribution $p(\boldsymbol{\theta})$. The important observation is:

- The matrix \mathbf{M} determines the covariance of the computed $\boldsymbol{\theta}$ (Fig. 4(a)).
- The matrix \mathbf{N} controls the bias of the computed $\boldsymbol{\theta}$ (Fig. 4(b)).

The analysis in Appendix A.2 applies to any method that solves Eq. (59), so the KCR lower bound is achieved up to $O(\sigma^4)$ if \mathbf{M} is chosen to be Eqs. (22), (29), and (55). The matrix \mathbf{N} is chosen to be

$$\mathbf{N} = \begin{cases} \mathbf{I}, & \text{iterative reweight} \\ \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_{\alpha}^{(kl)} V_0^{(kl)} [\boldsymbol{\xi}_{\alpha}], & \text{renormalization} \\ \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_{\alpha}^{(kl)} (V_0^{(kl)} [\boldsymbol{\xi}_{\alpha}] + \dots) & \text{hyper-renormalization} \end{cases} \quad (60)$$

We have shown that the last choice, i.e., Eq. (56), eliminates the bias up to $O(\sigma^4)$. If we iteratively solve Eq. (59), the initial solution with $W_{\alpha}^{(kl)} = \delta_{kl}$ corresponds to the least squares, the (extended) method of Taubin [26], and HyperLS [13], [14], [24], respectively. In other words, iterative reweight, renormalization, and hyper-renormalization can be viewed as an iterative improvement of least squares, the Taubin method, and HyperLS, respectively (Table 1).

Standard linear algebra routines for solving the generalized eigenvalue problem in Eq. (59) assume that \mathbf{N} is positive definite, but the matrix \mathbf{N} in Eq. (56) has both positive and negative eigenvalues. For the Taubin method and renormalization, the matrix \mathbf{N} in Eq. (30) is positive semidefinite having eigenvalue 0.

This, however, causes no difficulty, because the problem can be rewritten as

$$\mathbf{N}\boldsymbol{\theta} = \frac{1}{\lambda} \mathbf{M}\boldsymbol{\theta}. \quad (61)$$

The matrix \mathbf{M} in Eq. (55) is positive definite for noisy data, so we can use a standard routine to compute the eigenvector $\boldsymbol{\theta}$ for the largest eigenvalue in absolute value. If the matrix \mathbf{M} happens to have eigenvalue 0, it indicates that the data are all exact, so its null vector is the exact solution.

From Eq. (17), the vector $\mathbf{e}^{(k)}$ defined by Eq. (5) is $(1, 0, 1, 0, 0, 0)^{\top}$ for ellipse fitting, but we see from Eqs. (19) and (21) that it is $\mathbf{0}$ for fundamental matrix and homography computation. In general, $\mathbf{e}^{(k)} = \mathbf{0}$ for *multilinear constraints*, such as the epipolar, trifocal, and quadrifocal constraints, that are linear in each variable, because we usually assume that noise in different images are uncorrelated. Homographies are expressed as the linear fractional equations of Eq. (7), but it can be converted to the multilinear form of Eq. (9). A notable example of $\mathbf{e}^{(k)} \neq \mathbf{0}$ is ellipse fitting, but its effect is negligible. We have confirmed by experiments that the results is practically unchanged if we let $\mathbf{e}^{(k)} = \mathbf{0}$. This has also been observed in the case of ellipse fitting by HyperLS [13].

11. Experiments

11.1 Accuracy Comparison

Since the computed $\boldsymbol{\theta}$ and its true value $\bar{\boldsymbol{\theta}}$ are both unit vectors, we measure the discrepancy $\Delta\boldsymbol{\theta}$ between them by the orthogonal component to $\bar{\boldsymbol{\theta}}$ (Fig. 3),

$$\Delta^{\perp}\boldsymbol{\theta} = \mathbf{P}_{\bar{\boldsymbol{\theta}}}\boldsymbol{\theta}, \quad \mathbf{P}_{\bar{\boldsymbol{\theta}}} \equiv \mathbf{I} - \bar{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}^{\top}, \quad (62)$$

where $\mathbf{P}_{\bar{\boldsymbol{\theta}}}$ is the projection matrix along $\bar{\boldsymbol{\theta}}$. We generate M independent noise instances and evaluate the bias B and the RMS error D defined by

$$B = \left\| \frac{1}{M} \sum_{a=1}^M \Delta^{\perp}\boldsymbol{\theta}^{(a)} \right\|, \quad D = \sqrt{\frac{1}{M} \sum_{a=1}^M \|\Delta^{\perp}\boldsymbol{\theta}^{(a)}\|^2}, \quad (63)$$

where $\boldsymbol{\theta}^{(a)}$ is the solution in the a th trial. The KCR lower bound of Eq. (44) implies that the RMS error D is bounded by

$$D \geq \frac{\sigma}{\sqrt{N}} \sqrt{\text{tr}\mathbf{M}^{-}}, \quad (64)$$

where tr denotes the matrix trace. We compared the accuracy of the following eight methods:

- (1) LS,
- (2) iterative reweight,
- (3) Taubin method,
- (4) renormalization,
- (5) HyperLS,
- (6) hyper-renormalization,
- (7) ML,
- (8) ML with hyperaccurate correction.

For ML, we used the FNS of Chojnacki et al. [4]. According to our experiments, the FNS solution agrees with the ML solution up to three or four significant digits, as also observed by Kanatani and Sugaya [11], [16], so FNS can be practically identified with ML.

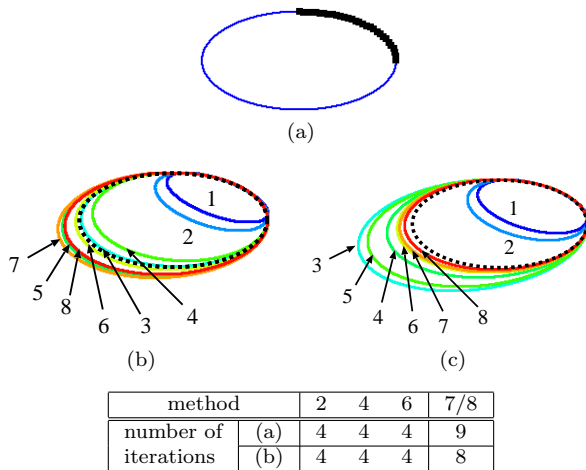


Fig. 5 (a) Thirty points on an ellipse. (b), (c) Fitted ellipses ($\sigma = 0.5$ pixels). 1) LS, 2) iterative reweight, 3) Taubin, 4) renormalization, 5) HyperLS, 6) hyper-renormalization, 7) ML, 8) ML with hyperaccurate correction. The dotted lines indicate the true shape. The table lists the number of iterations for methods 2, 4, 6, and 7/8 (methods 1, 3, and 5 are noniterative, and methods 7 and 8 have the same complexity).

For hyperaccurate correction, we used the scheme of Kanatani and Sugaya [18], which extends the scheme of Kanatani [10], [11] to multiple constraints.

The purpose of this experiment is to confirm our theoretical predictions. As shown in Eq. (44), the covariance matrix $V[\theta]$ of hyper-renormalization, as well as for iterative reweight and renormalization that use the same \mathbf{M} , coincides with the KCR lower bound up to $O(\sigma^4)$. It is known that ML also has the same order of covariance matrix [11]. However, ML has bias of $O(\sigma^2)$ [11], while hyper-renormalization has bias $O(\sigma^4)$. Hence, hyper-renormalization should have smaller RMS error than ML. We numerically confirm if this is indeed the case. On the other hand, the hyperaccurate correction of Kanatani and Sugaya [18] removes, a posteriori, the bias of $O(\sigma^2)$ from the ML solution, so it should have the same order of accuracy as hyper-renormalization, and comparing these two is also the purpose of this experiment. We also observe the convergence behavior for moderate and realistic ranges of the noise level in which our experiment is conducted.

11.2 Ellipse Fitting

We define 30 equidistant points on the ellipse shown in Fig. 5(a). The major and minor axis are set to 100 and 50 pixels, respectively. We add independent Gaussian noise of mean 0 and standard deviation σ (pixels) to the x and y coordinates of each point and fit an ellipse by different methods.

Figure 5(b), (c) show fitted ellipses for $\sigma = 0.5$ pixels; although the noise magnitude is the same, the resulting ellipses are different for different noise. The true shape is indicated by dotted lines. We can see that LS and iterative reweight have large bias, producing much smaller ellipses than the true shape. The closest ellipse is given by hyper-renormalization in Fig. 5(b) and by ML with hyperaccurate correction in Fig. 5(c).

The number of iterations for each method is also shown there. We see that ML with/without hypercorrection requires about

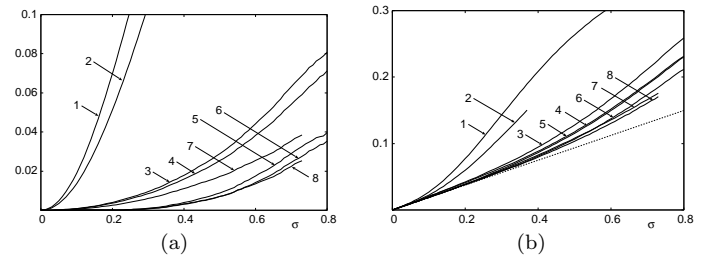


Fig. 6 The bias (a) and the RMS error (b) of the fitted ellipse for the standard deviation σ of the noise added to the data in Fig. 5(a) over 10,000 independent trials. 1) LS, 2) iterative reweight, 3) Taubin, 4) renormalization, 5) HyperLS, 6) hyper-renormalization, 7) ML, 8) ML with hyperaccurate correction. The dotted line in (b) indicates the KCR lower bound.

twice as many iterations for convergence as iterative reweight, renormalization, and hyper-renormalization; Taubin and HyperLS are noniterative algebraic methods, while hyperaccurate correction is an analytical procedure after ML has converged. The fast convergence of hyper-renormalization is a result of its initialization by HyperLS (see Table 1), while ML by FNS starts by default (i.e., by letting $\theta = \mathbf{0}$) from LS, as we did in our experiment. Evidently, FNS or other ML scheme such as the Levenberg-Marquadt method [23] should converge fast if initialized by HyperLS.

Note that all the ellipses in Fig. 5(b), (c) fit fairly well to the data points, meaning that not much difference exists among their reprojection errors, i.e., the sums of the square distances of the data points from the fitted ellipse. However, the deviation is large in the part where *no data points exist*. Since θ expresses the coefficients of the ellipse equation (see Eqs. (3) and (4)), the error $\Delta\theta$ evaluates how the “ellipse equation,” i.e., the ellipse itself, differs. This implies that the reprojection error is not a good measure of ellipse fitting; we need to evaluate the error in θ . This was also pointed out by Kanatani et al. [14] in relation to HyperLS.

Figure 6(a), (b) plot the bias B and the RMS error D , respectively, defined in Eq. (63) over 10,000 independent trials for each σ . The dotted line in Fig. 6(b) is the KCR lower bound of Eq. (64). The interrupted plots in Fig. 6 for iterative reweight, ML with/without hyperaccurate correction indicates that FNS did not converge beyond that noise level. Our convergence criterion is $\|\theta - \theta_0\| < 10^{-6}$ for the current value θ and the value θ_0 in the preceding iteration; their signs are adjusted before subtraction. If this criterion is not satisfied after 100 iterations, we stopped. For each σ , we regarded the iterations as not convergent if any among the 10,000 trials did not converge.

This disruption can be avoided by using other ML scheme such as the Levenberg-Marquadt method [23] with good initialization or even by using FNS with HyperLS initialization. If the noise is very large, however, all methods may fail or, even if they do not, convergence may be very slow, and the convergence behavior is very difficult to predict; it is beyond the scope of this paper. Here, we only point out that in our experimented noise range, which we regard as reasonable, hyper-renormalization performs very well as compared with ML with hypercorrection using FNS.

Figure 7 enlarges Fig. 6 for the small σ part. We can see from Fig. 6(a) and Fig. 7(a) that LS and iterative reweight have very large bias, in contrast to which the bias is very small for

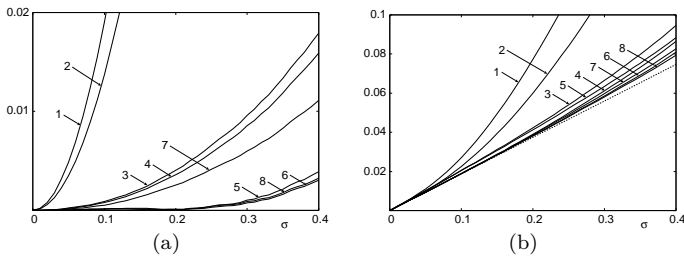
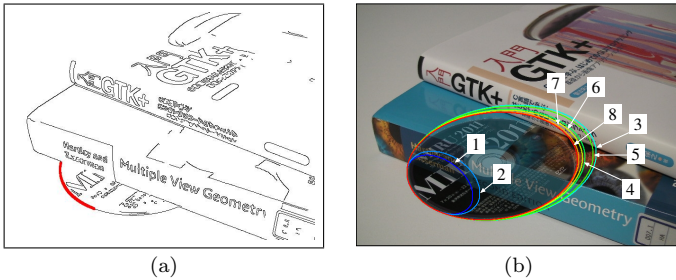


Fig. 7 (a) Enlargement of Fig. 6(a). (b) Enlargement of Fig. 6(b).



method	2	4	6	7/8
# of iter.	4	3	3	6

Fig. 8 (a) An edge image of a scene with a circular object. An ellipse is fitted to the 160 edge points indicated. (b) Fitted ellipses superimposed on the original image. The occluded part is artificially composed for visual ease. 1) LS, 2) iterative reweight, 3) Taubin, 4) renormalization, 5) HyperLS, 6) hyper-renormalization, 7) ML, 8) ML with hyperaccurate correction.

the Taubin method and renormalization. The bias of HyperLS and hyper-renormalization is still smaller and even smaller than ML. Note that the RMS error D of Eq. (63) is a measure of the distance from the true value θ , while the first order covariance matrix of Eq. (44) measure the “width” of the bell-shaped distribution from its center (see Fig. 4(a)), which is common to iterative reweight, renormalization, and hyper-renormalization. Hence, the difference in the RMS error in Fig. 6(b) and Fig. 7(b) is a direct consequence of the bias difference shown in Fig. 6(a) and Fig. 7(a).

A close examination of the small σ part (Fig. 6(b)) reveals that hyper-renormalization outperforms ML. The highest accuracy is achieved, although the difference is very small, by hyperaccurate correction of ML. However, it first requires the ML solution, and the FNS iterations for its computation may not converge above a certain noise level, as shown in Fig. 6.

We also studied the effect of $e_{\alpha}^{(k)}$ defined by Eq. (49) on hyper-renormalization and hyperaccurate correction of ML. We let $e_{\alpha}^{(k)} = \mathbf{0}$ and did the same experiments but found that the plots in Figs. 6 and 7 are unchanged.

Figure 8(a) is an edge image of a scene with a circular object. We fitted an ellipse to the 160 edge points indicated there, using various methods. Figure 8(b) shows the fitted ellipses superimposed on the original image, where the occluded part is artificially composed for visual ease. We can see that LS and iterative reweight produce much smaller ellipses than the true shape as in Fig. 5(b), (c). All other fits are very close to the true ellipse, and ML gives the best fit in this particular instance. The number of iterations before convergence for each method is also shown in Fig. 8. Again, FNS for ML with/without hypercorrection required

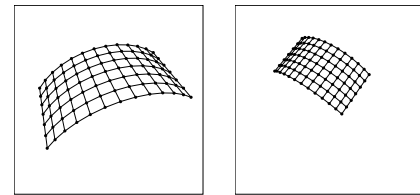


Fig. 9 Simulated images of a curved grid surface viewed from two directions.

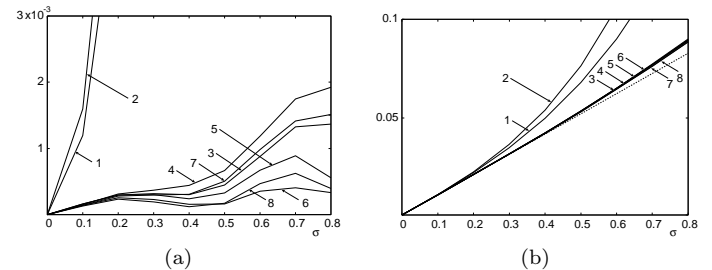


Fig. 10 The bias (a) and the RMS error (b) of the computed fundamental matrix for the standard deviation σ of the noise added to the data in Fig. 9 over 10,000 independent trials. 1) LS, 2) iterative reweight, 3) Taubin, 4) renormalization, 5) HyperLS, 6) hyper-renormalization, 7) ML, 8) ML with hyperaccurate correction. The dotted line in (b) indicates the KCR lower bound.

about twice as many iterations as other methods.

11.3 Fundamental Matrix Computation

Figure 9 shows simulated images of a curved grid surface viewed from two directions. The image size is 600×600 pixels, and the focal length is 600 pixels. We add Gaussian noise of mean 0 and standard deviation σ (pixels) to the x and y coordinates of each grid point independently and compute the fundamental matrix F . The fundamental matrix F has rank 2, so it is constrained to be $\det F = 0$ [7]. Basically, the following three approaches exist for imposing this rank constraint [16]:

- (1) A posteriori correction: The matrix F is optimally computed without considering the rank constraint and then optimally corrected so that it is satisfied.
- (2) Internal access: The matrix F is parameterized so that the rank constraint is identically satisfied and then optimized within the resulting smaller parameter space.
- (3) External access: Iterations are done in the space of unconstrained F in such a way the rank constraint is automatically satisfied at the time of convergence.

Here, we adopt the a posteriori correction approach and compare the accuracy of various methods without considering the rank constraint.

Figure 10(a), (b) plot the bias B and the RMS error D , respectively, defined in Eq. (63) over 10,000 independent trials for each σ . The dotted line in Fig. 10(b) is the KCR lower bound of Eq. (64). As we can see from Fig. 10(a), LS and iterative reweight have very large bias. As in the case of ellipse fitting, the leading covariance is common to iterative reweight, renormalization, and hyper-renormalization, and hence the RMS error reflects the influence of the bias as shown in Fig. 10(b). As seen from Fig. 10(a), ML has considerable bias, which is largely removed by the hyperaccurate correction, and hyper-renormalization directly computes

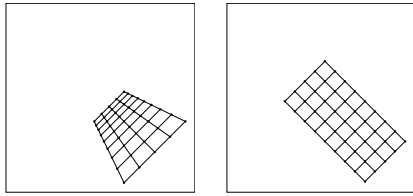


Fig. 11 Simulated images of a planar grid surface viewed from two directions.

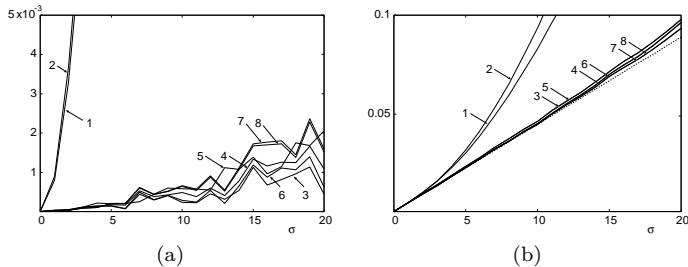


Fig. 12 The bias (a) and the RMS error (b) of the computed homography for the standard deviation σ of the noise added to the data in Fig. 11 over 10,000 independent trials. 1) LS, 2) iterative reweight, 3) Taubin, 4) renormalization, 5) HyperLS, 6) hyper-renormalization, 7) ML, 8) ML with hyperaccurate correction. The dotted line in (b) indicates the KCR lower bound.

solutions with as small bias. However, all methods except LS and iterative reweight nearly achieve the KCR lower bound, as seen from Fig. 10(b). Hence, the effect of bias reduction has little influence on the RMS error. As in the case of ellipse fitting, the best performance is obtained by hyper-renormalization and ML with hyperaccurate correction, although the difference from other methods is very small.

11.4 Homography Computation

Figure 11 shows simulated images of a planar grid surface viewed from two directions. The image size is 800×800 pixels, and the focal length is 600 pixels. We add Gaussian noise of mean 0 and standard deviation σ (pixels) to the x and y coordinates of each grid point independently and compute the homography between the two images.

Figure 12(a), (b) plot the bias B and the RMS error D , respectively, defined in Eq. (63) over 10,000 independent trials for each σ . The dotted line in Fig. 12(b) is the KCR lower bound of Eq. (64). As in the case of ellipse fitting and fundamental matrix computation, LS and iterative reweight have very large bias, resulting in large RMS errors. However, all other methods have small bias of similar magnitude, so bias reduction has little influence on their RMS errors, which nearly achieve the KCR lower bound as seen from Fig. 12(b). Yet, the best performance is obtained by hyper-renormalization and ML with hyperaccurate correction, although the difference from other methods is very small. Like ellipse fitting, however, the iterations for ML computation using FNS does not converge for large noise, although this does not occur in the noise range of Fig. 12(a), (b); the iterations did not necessarily converge beyond $\sigma = 25$ pixels. The convergence criterion is the same as in the case of ellipse fitting. In contrast, hyper-renormalization converges even in that noise range.

Homography is a fundamental tool for panoramic image generation by image mosaicing: images of a far away scene are seamlessly

past together after warped according to the homographies between images. Usually, the images overlap very well in the part where matching points are extracted, but a large deviation may appear in the far away part with no matching points. In such applications, the reprojection error, i.e., the sum of the square distances between the points to be matched, is more or less the same among different methods as pointed out by Kanatani et al. [14]. Hence, the error in θ , which expresses the coefficients of the homography equation (see Eqs. (7) and (11)), is a better measure for evaluating the mapping deviation, although it depends on how we express the homography equation.

12. Conclusions

We have reformulated iterative reweight and renormalization of Kanatani [8], [9] as geometric estimation techniques not based on the minimization principle. We discussed our non-minimization approach in detail and did a precise error analysis in the most general situation with multiple constraints. From our analysis, we obtained a scheme which we call “hyper-renormalization” that optimizes the matrices \mathbf{M} and \mathbf{N} in the computation so that the covariance of the solution achieves the KCR lower bound up to $O(\sigma^4)$ and the bias is zero up to $O(\sigma^4)$ in the noise level σ . Doing experiments of ellipse fitting, fundamental matrix computation, and homography computation, we observed that

- (1) Iterative reweight is an iterative improvement of LS. The leading covariance of the solution agrees with the KCR lower bound [2], [9], [11], but the bias is very large, hence the accuracy is low.
- (2) Renormalization is an iterative improvement of the method of Taubin [26]. The leading covariance of the solution agrees with the KCR lower bound, and the bias is very small, hence the accuracy is high.
- (3) Hyper-renormalization is an iterative improvement of HyperLS [13], [14], [24]. The leading covariance of the solution agrees with the KCR lower bound with no bias up to $O(\sigma^4)$. It outperforms ML.
- (4) Although the difference is very small, ML with hyperaccurate correction of Kanatani [10], [11], [18] exhibits the highest accuracy. However, if FNS is used for ML computation, the iterations may not converge in the presence of large noise, while hyper-renormalization has a higher noise tolerance.

We conclude that hyper-renormalization is the best strategy for practical computations in terms of accuracy and efficiency in moderate noise ranges.

As we remarked earlier, our non-minimization approach depends on a particular representation and parameterization of the problem, so that we first need to fix the way the problem is specified. Also, our error analysis is based on Gaussian noise modeling, which only applies to fairly accurate data. Despite these limitations, our non-minimization approach is expected to play an important role in building accurate computer vision systems for various applications.

Acknowledgments The authors thank Takayuki Okatani of Tohoku University, Japan, Mike Brooks and Wojciech Chojnacki of the University of Adelaide, Australia, and Peter Meer of Rutgers University, U.S.A. This work was supported in part by JSPS Grant-in-Aid for Challenging Exploratory Research (24650086).

References

- [1] Al-Sharadqah, A. and Chernov, N.: A doubly optimal ellipse fit, *Comput. Stat. Data Anal.*, Vol.6, No.9, pp.2771–2781 (2012).
- [2] Chernov, N. and Lesort, C. : Statistical efficiency of curve fitting algorithms, *Comput. Stat. Data Anal.*, Vol.47, No.4, pp.713–728 (2004).
- [3] Chojnacki, W., Brooks, M.J. and van den Hengel, A.: Rationalising the renormalization method of Kanatani, *J. Math. Imaging Vis.*, Vol.21, No.11, pp.21–38 (2001).
- [4] Chojnacki, W., Brooks, M.J., van den Hengel, A. and Gawley, D.: On the fitting of surfaces to data with covariances, *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol.22, No.11, pp.1294–1303 (2000).
- [5] Godambe, V.P. (Ed.): *Estimating Functions*, Oxford University Press, New York, U.S.A (1991).
- [6] Hartley, R.: In defense of the eight-point algorithm, *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol.19, No.6, pp.580–593 (1997).
- [7] Hartley, R. and Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, Cambridge, U.K. (2003).
- [8] Kanatani, K.: Renormalization for unbiased estimation, *Proc. 4th Int. Conf. Computer Vision*, Berlin, Germany, pp.599–606 (1993).
- [9] Kanatani, K.: *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam, The Netherlands (1996). Reprinted by Dover, New York, U.S.A. (2005).
- [10] Kanatani, K.: Ellipse fitting with hyperaccuracy, *IEICE Trans. Inf. & Syst.*, Vol.E89-D, No.10, pp.2653–2660 (2006).
- [11] Kanatani, K.: Statistical optimization for geometric fitting: Theoretical accuracy bound and high order error analysis, *Int. J. Comput. Vis.*, Vol.80, No.2, pp.167–188 (2008).
- [12] Kanatani, K., Al-Sharadqah, A., Chernov, N. and Sugaya, Y.: Renormalization returns: Hyper-renormalization and its applications, *Proc. 12th Euro. Conf. Computer Vision*, Firenze, Italy (2012).
- [13] Kanatani, K. and Rangarajan, P.: Hyper least squares fitting of circles and ellipses. *Comput. Stat. Data Anal.*, Vol.55, No.6, pp.2197–2208 (2011).
- [14] Kanatani, K., Rangarajan, P., Sugaya, Y. and Niitsuma, H.: HyperLS and its applications, *IPSPJ Trans. Comput. Vis. Appl.*, Vol.3, pp.80–94 (2011).
- [15] Kanatani, K. and Sugaya, Y.: Compact algorithm for strictly ML ellipse fitting, *Proc. 19th Int. Conf. Pattern Recognition*, Tampa, U.S.A. (2008).
- [16] Kanatani, K. and Sugaya, Y.: Compact fundamental matrix computation, *IPSPJ Trans. Comput. Vis. Appl.*, Vol.2, pp.59–70 (2010).
- [17] Kanatani, K. and Sugaya, Y.: Unified computation of strict maximum likelihood for geometric fitting, *J. Math. Imaging Vis.*, Vol.38, No.1, pp.1–13 (2010).
- [18] Kanatani, K. and Sugaya, Y.: Hyperaccurate correction of maximum likelihood for geometric estimation, *IPSPJ Trans. Comput. Vis. Appl.*, Vol.5, pp.19–29 (2013).
- [19] Leedan, Y. and Meer, P.: Heteroscedastic regression in computer vision: Problems with bilinear constraint, *Int. J. Comput. Vis.*, Vol.37, No.2, pp.127–150 (2000).
- [20] Matei, J. and Meer, P.: Estimation of nonlinear errors-in-variables models for computer vision applications, *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol.28, No.10, pp.1537–1552 (2006).
- [21] Okatani, T. and Deguchi, K.: On bias correction for geometric parameter estimation in computer vision, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Miami Beach, FL, U.S.A., pp.959–966 (2009).
- [22] Okatani, T. and Deguchi, K.: Improving accuracy of geometric parameter estimation using projected score method, *Proc. 12th Int. Conf. Computer Vision*, Kyoto, Japan, pp.1733–1740 (2009).
- [23] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P.: *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, Cambridge, U.K. (1992).
- [24] Rangarajan, P. and Kanatani, K.: Improved algebraic methods for circle fitting, *Elec. J. Stat.*, Vol.3, pp.1075–1082 (2009).
- [25] Sampson, P. D.: Fitting conic sections to “very scattered” data: An iterative refinement of the Bookstein algorithm, *Comput. Graphics Image Process.*, Vol.18, No.1, pp.97–108 (1982).
- [26] Taubin, G.: 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.*, Vol.13, No.11, pp.1115–1138 (1991).

Appendix

A.1 Derivation of Eqs. (37) and (38)

Let \mathbf{W}_α and \mathbf{V}_α be the matrices whose (kl) element is $W_\alpha^{(kl)}$ and

$(\bar{\theta}, V_0^{(kl)}[\xi_\alpha] \bar{\theta})$, respectively. Since $\mathbf{W}_\alpha = (\mathbf{V}_\alpha)_r^-$ by definition, the identity $\mathbf{V}_\alpha \mathbf{W}_\alpha \mathbf{V}_\alpha = \mathbf{V}_\alpha$ holds. Expanding this, we obtain

$$\begin{aligned} & (\bar{\mathbf{V}}_\alpha + \Delta_1 \mathbf{V}_\alpha + \Delta_2 \mathbf{V}_\alpha + \cdots)(\bar{\mathbf{W}}_\alpha + \Delta_1 \mathbf{W}_\alpha + \Delta_2 \mathbf{W}_\alpha \\ & + \cdots)(\bar{\mathbf{V}}_\alpha + \Delta_1 \mathbf{V}_\alpha + \Delta_2 \mathbf{V}_\alpha + \cdots) \\ & = (\bar{\mathbf{V}}_\alpha + \Delta_1 \mathbf{V}_\alpha + \Delta_2 \mathbf{V}_\alpha + \cdots). \end{aligned} \quad (\text{A.1})$$

We can derive Eqs. (37) and (38) by equating the terms of the same order on both sides.

Equating the first order terms on both sides, we obtain

$$\Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha + \bar{\mathbf{V}}_\alpha \Delta_1 \mathbf{W}_\alpha \bar{\mathbf{V}}_\alpha + \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha = \Delta_1 \mathbf{V}_\alpha. \quad (\text{A.2})$$

Multiplying this by $\bar{\mathbf{W}}_\alpha$ from left and right, we have

$$\begin{aligned} & \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha + \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha \Delta_1 \mathbf{W}_\alpha \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha \\ & + \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha = \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha. \end{aligned} \quad (\text{A.3})$$

Note that the product $\bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha = \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha$ operates as the projection onto the orthogonal complement of the (common) null space of $\bar{\mathbf{V}}_\alpha$ and $\bar{\mathbf{W}}_\alpha$ and that the variations $\Delta_1 \mathbf{V}_\alpha$ and $\Delta_1 \mathbf{W}_\alpha$ take place within that domain. We also note the identities of the generalized inverse $\bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha = \bar{\mathbf{W}}_\alpha$ and $\bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha = \bar{\mathbf{V}}_\alpha$ hold. Hence, we obtain from Eq. (A.3)

$$\bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha + \Delta_1 \mathbf{W}_\alpha + \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha = \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha, \quad (\text{A.4})$$

from which we obtain $\Delta_1 \mathbf{W}_\alpha$ in the form

$$\Delta_1 \mathbf{W}_\alpha = -\bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha. \quad (\text{A.5})$$

Its (kl) element is written as

$$\begin{aligned} \Delta_1 W_\alpha^{(kl)} & = - \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{ln} \Delta_1 V_\alpha^{(mn)} \\ & = -2 \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} (\Delta_1 \theta, V_0^{(mn)}[\xi_\alpha] \bar{\theta}). \end{aligned} \quad (\text{A.6})$$

Thus, we obtain Eq. (37).

Equating the second order terms from both sides of Eq. (A.1), we have

$$\begin{aligned} & \Delta_2 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha + \bar{\mathbf{V}}_\alpha \Delta_2 \mathbf{W}_\alpha \bar{\mathbf{V}}_\alpha + \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha \Delta_2 \mathbf{V}_\alpha \\ & + \bar{\mathbf{V}}_\alpha \Delta_1 \mathbf{W}_\alpha \Delta_1 \mathbf{V}_\alpha + \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha + \Delta_1 \mathbf{V}_\alpha \Delta_1 \mathbf{W}_\alpha \bar{\mathbf{V}}_\alpha \\ & = \Delta_2 \mathbf{V}_\alpha. \end{aligned} \quad (\text{A.7})$$

Multiplying this by $\bar{\mathbf{W}}_\alpha$ from left and right, we obtain

$$\begin{aligned} & \bar{\mathbf{W}}_\alpha \Delta_2 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha + \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha \Delta_2 \mathbf{W}_\alpha \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha \\ & + \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha \Delta_2 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha + \bar{\mathbf{W}}_\alpha \bar{\mathbf{V}}_\alpha \Delta_1 \mathbf{W}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \\ & + \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha + \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \Delta_1 \mathbf{W}_\alpha \bar{\mathbf{V}}_\alpha \bar{\mathbf{W}}_\alpha \\ & = \bar{\mathbf{W}}_\alpha \Delta_2 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha, \end{aligned} \quad (\text{A.8})$$

which is rewritten as

$$\begin{aligned} & \bar{\mathbf{W}}_\alpha \Delta_2 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha + \Delta_2 \mathbf{W}_\alpha + \bar{\mathbf{W}}_\alpha \Delta_2 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \\ & + \Delta_1 \mathbf{W}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha + \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha \\ & + \bar{\mathbf{W}}_\alpha \Delta_1 \mathbf{V}_\alpha \Delta_1 \mathbf{W}_\alpha = \bar{\mathbf{W}}_\alpha \Delta_2 \mathbf{V}_\alpha \bar{\mathbf{W}}_\alpha. \end{aligned} \quad (\text{A.9})$$

This is further rewritten as follows:

$$\begin{aligned}
 & \bar{W}_\alpha \Delta_2 V_\alpha \bar{W}_\alpha + \Delta_2 W_\alpha + \bar{W}_\alpha \Delta_2 V_\alpha \bar{W}_\alpha \\
 & + \Delta_1 W_\alpha (\bar{V}_\alpha \bar{W}_\alpha) \Delta_1 V_\alpha \bar{W}_\alpha \\
 & + \bar{W}_\alpha \Delta_1 V_\alpha \bar{W}_\alpha (\bar{V}_\alpha \bar{W}_\alpha) \Delta_1 V_\alpha \bar{W}_\alpha \\
 & + \bar{W}_\alpha \Delta_1 V_\alpha (\bar{W}_\alpha \bar{V}_\alpha) \Delta_1 W_\alpha = \bar{W}_\alpha \Delta_2 V_\alpha \bar{W}_\alpha. \tag{A.10}
 \end{aligned}$$

Substituting Eq. (A.5), we obtain

$$\begin{aligned}
 & \bar{W}_\alpha \Delta_2 V_\alpha \bar{W}_\alpha + \Delta_2 W_\alpha + \bar{W}_\alpha \Delta_2 V_\alpha \bar{W}_\alpha \\
 & - \Delta_1 W_\alpha \bar{V}_\alpha \Delta_1 W_\alpha + \Delta_1 W_\alpha \bar{V}_\alpha \Delta_1 W_\alpha \\
 & - \Delta_1 W_\alpha \bar{V}_\alpha \Delta_1 W_\alpha = \bar{W}_\alpha \Delta_2 V_\alpha \bar{W}_\alpha. \tag{A.11}
 \end{aligned}$$

Hence, $\Delta_2 W_\alpha$ is written as

$$\Delta_2 W_\alpha = \Delta_1 W_\alpha \bar{V}_\alpha \Delta_1 W_\alpha - \bar{W}_\alpha \Delta_2 V_\alpha \bar{W}_\alpha, \tag{A.12}$$

whose (kl) element is

$$\begin{aligned}
 \Delta_2 W_\alpha^{(kl)} &= \sum_{m,n=1}^L \Delta_1 W_\alpha^{(km)} \bar{V}_\alpha^{(mn)} \Delta_1 W_\alpha^{(nl)} \\
 &- \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \Delta_2 V_\alpha^{(mn)} \bar{W}_\alpha^{(nl)} \\
 &= \sum_{m,n=1}^L \Delta_1 W_\alpha^{(km)} \Delta_1 W_\alpha^{(ln)} (\bar{\theta}, V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \\
 &- \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \left((\Delta_1 \theta, V_0^{(mn)} [\xi_\alpha] \Delta_1 \theta) \right. \\
 &\quad \left. + 2(\Delta_2 \theta, V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \right). \tag{A.13}
 \end{aligned}$$

Thus, we obtain Eq. (38).

A.2 Covariance Matrix of the Solution

Substituting Eq. (35) into Eq. (43) and noting that $\xi_\alpha^{(k)\top} \theta = 0$, we can write $\Delta_1 \theta$ as follows:

$$\begin{aligned}
 \Delta_1 \theta &= -\bar{M}^{-1} \Delta_1 M \bar{\theta} \\
 &= -\bar{M}^{-1} \left(\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right). \tag{A.14}
 \end{aligned}$$

We can evaluate $V[\theta] = E[\Delta_1 \theta \Delta_1 \theta^\top]$, the covariance matrix of θ to a first approximation, by eliminating $\Delta_1 \xi_\alpha^{(k)}$ in the expectation, using our assumption that the noise in $\xi_\alpha^{(k)}$ is independent for each α and hence

$$E[\Delta_1 \xi_\alpha^{(k)} \Delta_1 \xi_\beta^{(l)\top}] = \delta_{\alpha\beta} V_0^{(kl)} [\xi_\alpha], \tag{A.15}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta. We also use the identity

$$\sum_{l,n=1}^L \bar{W}_\alpha^{(kl)} (\bar{\theta}, V_0^{(ln)} [\xi_\alpha] \bar{\theta}) \bar{W}_\alpha^{(mn)} = \bar{W}_\alpha^{(km)}, \tag{A.16}$$

which means $\bar{W}_\alpha \bar{W}_\alpha^\top \bar{W}_\alpha = \bar{W}_\alpha$ in the matrix form, to simplify the expression.

Using Eq. (A.15), we see that

$$\begin{aligned}
 V[\theta] &= E[\Delta_1 \theta \Delta_1 \theta^\top] \\
 &= E[\bar{M}^{-1} \left(\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \bar{M}^{-1}]
 \end{aligned}$$

$$\begin{aligned}
 & \frac{1}{N} \sum_{\beta=1}^N \sum_{m,n=1}^L \bar{W}_\beta^{(mn)} (\Delta_1 \xi_\beta^{(n)}, \bar{\theta}) \bar{\xi}_\beta^{(m)\top} \bar{M}^{-1} \\
 &= E[\bar{M}^{-1} \left(\frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(mn)} \right. \\
 &\quad \left. (\bar{\theta}, \Delta_1 \xi_\alpha^{(l)}) (\Delta_1 \xi_\beta^{(n)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \bar{\xi}_\beta^{(m)\top} \right) \bar{M}^{-1}] \\
 &= \bar{M}^{-1} \left(\frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(mn)} \right. \\
 &\quad \left. (\bar{\theta}, E[\Delta_1 \xi_\alpha^{(l)} \Delta_1 \xi_\beta^{(n)\top}] \bar{\theta}) \bar{\xi}_\alpha^{(k)} \bar{\xi}_\beta^{(m)\top} \right) \bar{M}^{-1} \\
 &= \bar{M}^{-1} \left(\frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(mn)} \right. \\
 &\quad \left. (\bar{\theta}, \sigma^2 \delta_{\alpha\beta} V_0^{(ln)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(k)} \bar{\xi}_\beta^{(m)\top} \right) \bar{M}^{-1} \\
 &= \bar{M}^{-1} \left(\frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,m=1}^L \left(\sum_{l,n=1}^L \bar{W}_\alpha^{(kl)} (\bar{\theta}, V_0^{(ln)} [\xi_\alpha] \bar{\theta}) \right. \right. \\
 &\quad \left. \left. \bar{W}_\alpha^{(mn)} \right) \bar{\xi}_\alpha^{(k)} \bar{\xi}_\alpha^{(m)\top} \right) \bar{M}^{-1} \\
 &= \frac{\sigma^2}{N} \bar{M}^{-1} \left(\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,m=1}^L \bar{W}_\alpha^{(km)} \bar{\xi}_\alpha^{(k)} \bar{\xi}_\alpha^{(m)\top} \right) \bar{M}^{-1} \\
 &= \frac{\sigma^2}{N} \bar{M}^{-1} \bar{M} \bar{M}^{-1} = \frac{\sigma^2}{N} \bar{M}^{-1}. \tag{A.17}
 \end{aligned}$$

A.3 Evaluation of $E[T\bar{\theta}]$

We evaluate the expectation of $T\theta = \Delta_2 M \theta - \Delta_1 M \bar{M}^{-1} \Delta_1 M \bar{\theta}$ term by term. The basic strategy is elimination of $\Delta_1 \xi_\alpha^{(k)}$ and $\Delta_2 \xi_\alpha^{(k)}$ in the expectation by using Eqs. (A.15) and (54), respectively. We also use Eq. (A.16) for simplifying the expression.

A.3.1 Evaluation of $E[\Delta_2 M \bar{\theta}]$

Consider the expectation of $\Delta_2 M \bar{\theta}$. From $(\bar{\xi}_\alpha, \bar{\theta}) = 0$ and Eq. (36), we obtain

$$\begin{aligned}
 \Delta_2 M \bar{\theta} &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(\Delta_1 \xi_\alpha^{(k)} \Delta_1 \xi_\alpha^{(l)\top} + \Delta_2 \xi_\alpha^{(k)} \bar{\xi}_\alpha^{(l)\top} \right. \\
 &\quad \left. + \bar{\xi}_\alpha^{(k)} \Delta_2 \xi_\alpha^{(l)\top} \right) \bar{\theta} + \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \Delta_1 W_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(k)} \bar{\xi}_\alpha^{(l)\top} \\
 &\quad + \bar{\xi}_\alpha^{(k)} \Delta_1 \xi_\alpha^{(l)\top}) \bar{\theta} + \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \Delta_2 W_\alpha^{(kl)} \bar{\xi}_\alpha^{(k)} \bar{\xi}_\alpha^{(l)\top} \bar{\theta} \\
 &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left((\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \Delta_1 \xi_\alpha^{(k)} + (\Delta_2 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &\quad + \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \Delta_1 W_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)}. \tag{A.18}
 \end{aligned}$$

Hence,

$$\begin{aligned}
 E[\Delta_2 M \bar{\theta}] &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(E[\Delta_1 \xi_\alpha^{(k)} \Delta_1 \xi_\alpha^{(l)\top}] \bar{\theta} \right. \\
 &\quad \left. + (E[\Delta_2 \xi_\alpha^{(l)}], \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &\quad + \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L (E[\Delta_1 W_\alpha^{(kl)} \Delta_1 \xi_\alpha^{(l)}], \bar{\theta}) \bar{\xi}_\alpha^{(k)}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{\sigma^2}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(V_0^{(kl)} [\xi_\alpha] \bar{\theta} + (e_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &+ \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L (E[\Delta_1 W_\alpha^{(kl)} \Delta_1 \xi_\alpha^{(l)}], \bar{\theta}) \bar{\xi}_\alpha^{(k)}. \quad (\text{A.19})
 \end{aligned}$$

Consider the expectation of $\Delta_1 W_\alpha \Delta_1 \xi_\alpha$. From Eqs. (35), (37), and (43), we see that

$$\begin{aligned}
 \Delta_1 W_\alpha^{(kl)} &= -2 \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} (\bar{M}^- \Delta_1 \bar{M} \bar{\theta}, V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \\
 &= 2 \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} (\bar{M}^- \Delta_1 \bar{M} \bar{\theta}, V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \\
 &= 2 \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} (\bar{M}^- \left(\frac{1}{N} \sum_{\beta=1}^N \sum_{p,q=1}^L \bar{W}_\beta^{(pq)} \right. \\
 &\quad \left. (\Delta_1 \xi_\beta^{(p)} \bar{\xi}_\beta^{(q)\top} + \bar{\xi}_\beta^{(p)} \Delta_1 \xi_\beta^{(q)\top}) \right. \\
 &\quad \left. + \frac{1}{N} \sum_{\beta=1}^N \sum_{p,q=1}^L \Delta_1 \bar{W}_\beta^{(pq)} \bar{\xi}_\beta^{(p)} \bar{\xi}_\beta^{(q)\top} \right) \bar{\theta}, V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \\
 &= 2 \sum_{m,n=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} (\bar{M}^- \left(\frac{1}{N} \sum_{\beta=1}^N \sum_{p,q=1}^L \bar{W}_\beta^{(pq)} \bar{\xi}_\beta^{(p)} \right. \\
 &\quad \left. (\Delta_1 \xi_\beta^{(q)}, \bar{\theta}) \right), V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \\
 &= \frac{2}{N} \sum_{\beta=1}^N \sum_{m,n,p,q=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \bar{W}_\beta^{(pq)} (\Delta_1 \xi_\beta^{(q)}, \bar{\theta}) \\
 &\quad (\bar{M}^- \bar{\xi}_\beta^{(p)}, V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \\
 &= \frac{2}{N} \sum_{\beta=1}^N \sum_{m,n,p,q=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \bar{W}_\beta^{(pq)} (\bar{\xi}_\beta^{(p)}, \\
 &\quad \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) (\Delta_1 \xi_\beta^{(q)}, \bar{\theta}). \quad (\text{A.20})
 \end{aligned}$$

Hence,

$$\begin{aligned}
 &E[\Delta_1 W_\alpha^{(kl)} \Delta_1 \xi_\alpha^{(l)}] \\
 &= E[\frac{2}{N} \sum_{\beta=1}^N \sum_{m,n,p,q=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \bar{W}_\beta^{(pq)} (\bar{\xi}_\beta^{(p)}, \\
 &\quad \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) (\Delta_1 \xi_\beta^{(q)}, \bar{\theta}) \Delta_1 \xi_\alpha^{(l)}] \\
 &= \frac{2}{N} \sum_{\beta=1}^N \sum_{m,n,p,q=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \bar{W}_\beta^{(pq)} (\bar{\xi}_\beta^{(p)}, \\
 &\quad \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) E[\Delta_1 \xi_\alpha^{(l)} \Delta_1 \xi_\beta^{(q)\top}] \bar{\theta} \\
 &= \frac{2}{N} \sum_{\beta=1}^N \sum_{m,n,p,q=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \bar{W}_\beta^{(pq)} (\bar{\xi}_\beta^{(p)}, \\
 &\quad \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \sigma^2 \delta_{\alpha\beta} V_0^{(lq)} [\xi_\alpha] \bar{\theta} \\
 &= \frac{2\sigma^2}{N} \sum_{m,n,p,q=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \bar{W}_\alpha^{(pq)} (\bar{\xi}_\alpha^{(p)}, \\
 &\quad \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) V_0^{(lq)} [\xi_\alpha] \bar{\theta}. \quad (\text{A.21})
 \end{aligned}$$

It follows that

$$\begin{aligned}
 &\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L (E[\Delta_1 W_\alpha^{(kl)} \Delta_1 \xi_\alpha^{(l)}], \bar{\theta}) \bar{\xi}_\alpha^{(k)} \\
 &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \left(\left(\frac{2\sigma^2}{N} \sum_{m,n,p,q=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \bar{W}_\alpha^{(pq)} (\bar{\xi}_\alpha^{(p)}, \right. \right. \\
 &\quad \left. \left. \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) V_0^{(lq)} [\xi_\alpha] \bar{\theta} \right), \bar{\theta} \right) \bar{\xi}_\alpha^{(k)} \\
 &= \frac{2\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n,p,q=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(ln)} \bar{W}_\alpha^{(pq)} (\bar{\xi}_\alpha^{(p)}, \\
 &\quad \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) (\bar{\theta}, V_0^{(lq)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(k)} \\
 &= \frac{2\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,m,n,p=1}^L \bar{W}_\alpha^{(km)} (\bar{\xi}_\alpha^{(p)}, \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \\
 &\quad \left(\sum_{l,q=1}^L \bar{W}_\alpha^{(nl)} (\bar{\theta}, V_0^{(lq)} [\xi_\alpha] \bar{\theta}) \bar{W}_\alpha^{(qp)} \right) \bar{\xi}_\alpha^{(k)} \\
 &= \frac{2\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,m,n,p=1}^L \bar{W}_\alpha^{(km)} (\bar{\xi}_\alpha^{(p)}, \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \bar{W}_\alpha^{(np)} \bar{\xi}_\alpha^{(k)} \\
 &= \frac{2\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,m,n,p=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(np)} (\bar{\xi}_\alpha^{(p)}, \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(k)}. \quad (\text{A.22})
 \end{aligned}$$

Substituting Eq. (A.22) into Eq. (A.19), we obtain the following expression for $E[\Delta_2 \bar{M} \bar{\theta}]$:

$$\begin{aligned}
 E[\Delta_2 \bar{M} \bar{\theta}] &= \frac{\sigma^2}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(V_0^{(kl)} [\xi_\alpha] \bar{\theta} + (e_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &+ \frac{2\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,m,n,p=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(np)} (\bar{\xi}_\alpha^{(p)}, \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(k)}. \quad (\text{A.23})
 \end{aligned}$$

A.3.2 Evaluation of $E[\Delta_1 \bar{M} \bar{M}^- \Delta_1 \bar{M} \bar{\theta}]$

We evaluate the expectation of $\Delta_1 \bar{M} \bar{M}^- \Delta_1 \bar{M} \bar{\theta}$. From Eq. (35), we can write

$$\begin{aligned}
 \Delta_1 \bar{M} \bar{\theta} &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(\Delta_1 \xi_\alpha^{(k)} \bar{\xi}_\alpha^{(l)\top} + \bar{\xi}_\alpha^{(k)} \Delta_1 \xi_\alpha^{(l)\top} \right) \bar{\theta} \\
 &+ \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \Delta_1 \bar{W}_\alpha^{(kl)} \bar{\xi}_\alpha^{(k)} \bar{\xi}_\alpha^{(l)\top} \bar{\theta} \\
 &= \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)}. \quad (\text{A.24})
 \end{aligned}$$

We can also write

$$\begin{aligned}
 &\Delta_1 \bar{M} \bar{M}^- \Delta_1 \bar{M} \bar{\theta} \\
 &= \Delta_1 \bar{M} \bar{M}^- \left(\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &= \left(\frac{1}{N} \sum_{\beta=1}^N \sum_{p,q=1}^L \bar{W}_\beta^{(pq)} \left(\Delta_1 \xi_\beta^{(p)} \bar{\xi}_\beta^{(q)\top} + \bar{\xi}_\beta^{(p)} \Delta_1 \xi_\beta^{(q)\top} \right) \right. \\
 &\quad \left. + \frac{1}{N} \sum_{\beta=1}^N \sum_{p,q=1}^L \Delta_1 \bar{W}_\beta^{(pq)} \bar{\xi}_\beta^{(p)} \bar{\xi}_\beta^{(q)\top} \right) \bar{M}^-
 \end{aligned}$$

$$\begin{aligned}
 & \left(\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &= \frac{1}{N} \sum_{\beta=1}^N \sum_{p,q=1}^L \bar{W}_\beta^{(pq)} \left(\Delta_1 \xi_\beta^{(p)} \bar{\xi}_\beta^{(q)\top} + \bar{\xi}_\beta^{(p)} \Delta_1 \xi_\beta^{(q)\top} \right) \bar{M}^- \\
 & \left(\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &+ \frac{1}{N} \sum_{\beta=1}^N \sum_{p,q=1}^L \Delta_1 \bar{W}_\beta^{(pq)} \bar{\xi}_\beta^{(p)} \bar{\xi}_\beta^{(q)\top} \bar{M}^- \\
 & \left(\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &= \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(pq)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \left(\Delta_1 \xi_\beta^{(p)} \bar{\xi}_\beta^{(q)\top} \right. \\
 & \left. + \bar{\xi}_\beta^{(p)} \Delta_1 \xi_\beta^{(q)\top} \right) \bar{M}^- \bar{\xi}_\alpha^{(k)} + \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \\
 & \Delta_1 \bar{W}_\beta^{(pq)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\beta^{(p)} \bar{\xi}_\beta^{(q)\top} \bar{M}^- \bar{\xi}_\alpha^{(k)} \\
 &= \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(pq)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) (\bar{\xi}_\beta^{(q)}, \\
 & \bar{M}^- \bar{\xi}_\alpha^{(k)}) \Delta_1 \xi_\beta^{(p)} \quad (\equiv \mathbf{t}_1) \\
 &+ \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(pq)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) (\Delta_1 \xi_\beta^{(q)}, \\
 & \bar{M}^- \bar{\xi}_\alpha^{(k)}) \bar{\xi}_\beta^{(p)} \quad (\equiv \mathbf{t}_2) \\
 &+ \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \Delta_1 \bar{W}_\beta^{(pq)} (\Delta_1 \xi_\alpha^{(l)}, \bar{\theta}) (\bar{\xi}_\beta^{(q)}, \\
 & \bar{M}^- \bar{\xi}_\alpha^{(k)}) \bar{\xi}_\beta^{(p)} \quad (\equiv \mathbf{t}_3). \tag{A.25}
 \end{aligned}$$

Consider the three terms \mathbf{t}_1 , \mathbf{t}_2 , and \mathbf{t}_3 separately. The expectation of \mathbf{t}_1 is

$$\begin{aligned}
 E[\mathbf{t}_1] &= \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(pq)} (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \\
 & E[\Delta_1 \xi_\beta^{(p)} \Delta_1 \xi_\alpha^{(l)\top}] \bar{\theta} \\
 &= \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(pq)} (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \sigma^2 \delta_{\alpha\beta} V_0^{(pl)} [\xi_\alpha] \bar{\theta} \\
 &= \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(pq)} (\bar{\xi}_\alpha^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) V_0^{(pl)} [\xi_\alpha] \bar{\theta}. \tag{A.26}
 \end{aligned}$$

The expectation of \mathbf{t}_2 is

$$\begin{aligned}
 E[\mathbf{t}_2] &= \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(pq)} (\bar{\theta}, E[\Delta_1 \xi_\alpha^{(l)} \Delta_1 \xi_\beta^{(q)\top}]) \\
 & \bar{M}^- \bar{\xi}_\alpha^{(k)} \bar{\xi}_\beta^{(p)} \\
 &= \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(pq)} (\bar{\theta}, \sigma^2 \delta_{\alpha\beta} V_0^{(lq)} [\xi_\alpha]) \\
 & \bar{M}^- \bar{\xi}_\alpha^{(k)} \bar{\xi}_\beta^{(p)} \\
 &= \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(pq)} (\bar{\xi}_\alpha^{(k)}, \bar{M}^- V_0^{(lq)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(p)}. \tag{A.27}
 \end{aligned}$$

Finally, we consider the expectation of \mathbf{t}_3 . From Eq. (A.20), we can write

$$\begin{aligned}
 \Delta_1 W_\beta^{(pq)} &= \frac{2}{N} \sum_{\gamma=1}^N \sum_{m,n,r,s=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\gamma^{(rs)} (\bar{\xi}_\gamma^{(r)}, \\
 \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) (\Delta_1 \xi_\gamma^{(s)}, \bar{\theta}). \tag{A.28}
 \end{aligned}$$

Hence,

$$\begin{aligned}
 & E[\Delta_1 W_\beta^{(pq)} \Delta_1 \xi_\alpha^{(l)}] \\
 &= E\left[\frac{2}{N} \sum_{\gamma=1}^N \sum_{m,n,r,s=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\gamma^{(rs)} (\bar{\xi}_\gamma^{(r)}, \right. \\
 & \left. \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) (\Delta_1 \xi_\gamma^{(s)}, \bar{\theta}) \Delta_1 \xi_\alpha^{(l)} \right] \\
 &= \frac{2}{N} \sum_{\gamma=1}^N \sum_{m,n,r,s=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\gamma^{(rs)} (\bar{\xi}_\gamma^{(r)}, \\
 & \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) E[\Delta_1 \xi_\alpha^{(l)} \Delta_1 \xi_\gamma^{(s)\top}] \bar{\theta} \\
 &= \frac{2}{N} \sum_{\gamma=1}^N \sum_{m,n,r,s=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\gamma^{(rs)} (\bar{\xi}_\gamma^{(r)}, \\
 & \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) \sigma^2 \delta_{\alpha\gamma} V_0^{(ls)} [\xi_\alpha] \bar{\theta} \\
 &= \frac{2\sigma^2}{N} \sum_{m,n,r,s=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\alpha^{(rs)} (\bar{\xi}_\alpha^{(r)}, \\
 & \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) V_0^{(ls)} [\xi_\alpha] \bar{\theta}. \tag{A.29}
 \end{aligned}$$

It follows that

$$\begin{aligned}
 & (E[\Delta_1 \bar{W}_\beta^{(pq)} \Delta_1 \xi_\alpha^{(l)}], \bar{\theta}) (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \\
 &= \left(\frac{2\sigma^2}{N} \sum_{m,n,r,s=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\alpha^{(rs)} (\bar{\xi}_\alpha^{(r)}, \right. \\
 & \left. \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) V_0^{(ls)} [\xi_\alpha] \bar{\theta}, \bar{\theta} \right) (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \\
 &= \frac{2\sigma^2}{N} \sum_{m,n,r,s=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\alpha^{(rs)} (\bar{\xi}_\alpha^{(r)}, \\
 & \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) (\bar{\theta}, V_0^{(ls)} [\xi_\alpha] \bar{\theta}) (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}). \tag{A.30}
 \end{aligned}$$

Thus, the expectation of \mathbf{t}_3 is

$$\begin{aligned}
 E[\mathbf{t}_3] &= \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} (E[\Delta_1 \bar{W}_\beta^{(pq)} \Delta_1 \xi_\alpha^{(l)}], \bar{\theta}) \\
 & (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \bar{\xi}_\beta^{(p)}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{N^2} \sum_{\alpha,\beta=1}^N \sum_{k,l,p,q=1}^L \bar{W}_\alpha^{(kl)} \left(\frac{2\sigma^2}{N} \sum_{m,n,r,s=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\alpha^{(rs)} \right. \\
 &\quad \left. (\bar{\xi}_\alpha^{(r)}, \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) (\bar{\theta}, V_0^{(ls)} [\xi_\alpha] \bar{\theta}) (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \right) \bar{\xi}_\beta^{(p)} \\
 &= \frac{2\sigma^2}{N^3} \sum_{\alpha,\beta=1}^N \sum_{k,l,m,n,p,q,r,s=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} \bar{W}_\alpha^{(rs)} \\
 &\quad (\bar{\xi}_\alpha^{(r)}, \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) (\bar{\theta}, V_0^{(ls)} [\xi_\alpha] \bar{\theta}) (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \bar{\xi}_\beta^{(p)} \\
 &= \frac{2\sigma^2}{N^3} \sum_{\alpha,\beta=1}^N \sum_{k,m,n,p,q,r=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} (\bar{\xi}_\alpha^{(r)}, \\
 &\quad \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) \left(\sum_{l,s=1}^L \bar{W}_\alpha^{(kl)} (\bar{\theta}, V_0^{(ls)} [\xi_\alpha] \bar{\theta}) \bar{W}_\alpha^{(sr)} \right) \\
 &\quad (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \bar{\xi}_\beta^{(p)} \\
 &= \frac{2\sigma^2}{N^3} \sum_{\alpha,\beta=1}^N \sum_{k,m,n,p,q,r=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} (\bar{\xi}_\alpha^{(r)}, \\
 &\quad \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) \bar{W}_\alpha^{(kr)} (\bar{\xi}_\beta^{(q)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) \bar{\xi}_\beta^{(p)} \\
 &= \frac{2\sigma^2}{N^2} \sum_{\beta=1}^N \sum_{m,n,p,q=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} (\bar{\xi}_\beta^{(q)}, \bar{M}^- \\
 &\quad \left(\frac{1}{N} \sum_{\alpha=1}^N \sum_{k,r=1}^L \bar{W}_\alpha^{(kr)} \bar{\xi}_\alpha^{(k)} \bar{\xi}_\alpha^{(r)\top} \right) \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) \bar{\xi}_\beta^{(p)} \\
 &= \frac{2\sigma^2}{N^2} \sum_{\beta=1}^N \sum_{m,n,p,q=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} (\bar{\xi}_\beta^{(q)}, \\
 &\quad \bar{M}^- \bar{M} \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) \bar{\xi}_\beta^{(p)} \\
 &= \frac{2\sigma^2}{N^2} \sum_{\beta=1}^N \sum_{m,n,p,q=1}^L \bar{W}_\beta^{(pm)} \bar{W}_\beta^{(qn)} (\bar{\xi}_\beta^{(q)}, \bar{M}^- V_0^{(mn)} [\xi_\beta] \bar{\theta}) \bar{\xi}_\beta^{(p)}.
 \end{aligned} \tag{A.31}$$

Adding the expectations of \mathbf{t}_1 , \mathbf{t}_2 , and \mathbf{t}_3 together, we obtain the expectation of $\Delta_1 \mathbf{M} \bar{\mathbf{M}}^- \Delta_1 \mathbf{M} \bar{\theta}$ in the following form:

$$\begin{aligned}
 E[\Delta_1 \mathbf{M} \bar{\mathbf{M}}^- \Delta_1 \mathbf{M} \bar{\theta}] &= \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(m)}, \\
 &\quad \bar{M}^- \bar{\xi}_\alpha^{(k)}) V_0^{(ln)} [\xi_\alpha] \bar{\theta} + \frac{3\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(k)}, \\
 &\quad \bar{M}^- V_0^{(lm)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(n)}.
 \end{aligned} \tag{A.32}$$

A.3.3 Evaluation of $E[\mathbf{T} \bar{\theta}]$

Combining the above expectations $E[\Delta_2 \mathbf{M} \bar{\theta}]$ and $E[\Delta_1 \mathbf{M} \bar{\mathbf{M}}^- \Delta_1 \mathbf{M} \bar{\theta}]$, we obtain the expectation of $\mathbf{T} \bar{\theta}$ in the form

$$\begin{aligned}
 E[\mathbf{T} \bar{\theta}] &= E[\Delta_2 \mathbf{M} \bar{\theta}] - E[\Delta_1 \mathbf{M} \bar{\mathbf{M}}^- \Delta_1 \mathbf{M} \bar{\theta}] \\
 &= \frac{\sigma^2}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(V_0^{(kl)} [\xi_\alpha] \bar{\theta} + (e_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &\quad + \frac{2\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,m,n,p=1}^L \bar{W}_\alpha^{(km)} \bar{W}_\alpha^{(np)} (\bar{\xi}_\alpha^{(p)}, \bar{M}^- V_0^{(mn)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(k)}
 \end{aligned}$$

$$\begin{aligned}
 &- \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(m)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) V_0^{(ln)} [\xi_\alpha] \bar{\theta} \\
 &- \frac{3\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(k)}, \bar{M}^- V_0^{(lm)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(n)} \\
 &= \frac{\sigma^2}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(V_0^{(kl)} [\xi_\alpha] \bar{\theta} + (e_\alpha^{(l)}, \bar{\theta}) \bar{\xi}_\alpha^{(k)} \right) \\
 &- \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(m)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) V_0^{(ln)} [\xi_\alpha] \bar{\theta} \\
 &- \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(k)}, \bar{M}^- V_0^{(lm)} [\xi_\alpha] \bar{\theta}) \bar{\xi}_\alpha^{(n)} \\
 &= \frac{\sigma^2}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(V_0^{(kl)} [\xi_\alpha] \bar{\theta} + \bar{\xi}_\alpha^{(k)} e_\alpha^{(l)\top} \bar{\theta} \right) \\
 &- \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(m)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) V_0^{(ln)} [\xi_\alpha] \bar{\theta} \\
 &- \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} \bar{\xi}_\alpha^{(n)} \bar{\xi}_\alpha^{(k)\top} \bar{M}^- V_0^{(lm)} [\xi_\alpha] \bar{\theta} \\
 &= \frac{\sigma^2}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L \bar{W}_\alpha^{(kl)} \left(V_0^{(kl)} [\xi_\alpha] \bar{\theta} + (\bar{\xi}_\alpha^{(k)} e_\alpha^{(l)\top} + e_\alpha^{(l)} \bar{\xi}_\alpha^{(k)\top}) \bar{\theta} \right) \\
 &- \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(m)}, \bar{M}^- \bar{\xi}_\alpha^{(k)}) V_0^{(ln)} [\xi_\alpha] \bar{\theta} \\
 &- \frac{\sigma^2}{N^2} \sum_{\alpha=1}^N \sum_{k,l,m,n=1}^L \bar{W}_\alpha^{(kl)} \bar{W}_\alpha^{(mn)} (\bar{\xi}_\alpha^{(n)} \bar{\xi}_\alpha^{(k)\top} \\
 &\quad \bar{M}^- V_0^{(lm)} [\xi_\alpha] + V_0^{(lm)} [\xi_\alpha] \bar{M}^- \bar{\xi}_\alpha^{(k)} \bar{\xi}_\alpha^{(n)\top}) \bar{\theta} = \sigma^2 \bar{N} \bar{\theta},
 \end{aligned} \tag{A.33}$$

where we define the matrix \bar{N} by Eq. (53).



Kenichi Kanatani received his B.E., M.S., and Ph.D. in applied mathematics from the University of Tokyo, Japan, in 1972, 1974 and 1979, respectively. After serving as Professor of computer science at Gunma University, Japan, and Okayama University, Japan, he retired in 2013 and is now Professor Emeritus of Okayama University. He is the author of many books on computer vision and received many awards including the Best Paper Awards from IPSJ (1987) and IEICE (2005). He is a Fellow of IEICE and IEEE.



Ali Al-Sharadqah received his B.Sc. and M.Sc. in from Jordan University of Science and Technology, Jordan, in 2001 and 2004, respectively, and his Ph.D. in applied mathematics from the University of Alabama at Birmingham, U.S.A., in 2011. He was Visiting Assistant Professor at the University of Mississippi, U.S.A. and currently is Assistant Professor of mathematics and statistics at East Carolina University, U.S.A. His research interests extend from statistical analysis of geometric problems to signal processing, computer vision, and pattern recognition.



Nikolai Chernov received his M.S. and Ph.D. in mathematics from Moscow University in 1979 and 1984, respectively. He worked at Moscow University, JINR, and MIREA in USSR, and UCLA, Georgia Institute of Technology, and Princeton University in U.S.A. He was Professor of mathematics at the University of Alabama at Birmingham, U.S.A. until he died on August 7, 2014. His research interests included ergodic theory, dynamical systems, chaos, statistical mechanics, probability theory, statistics, data processing, and computer programming.



Yasuyuki Sugaya received his B.E., M.S., and Ph.D. in computer science from University of Tsukuba, Japan, 1996, 1998, and 2001, respectively. After serving as Assistant Professor of computer science at Okayama University, Japan, he is currently Associate Professor of computer science and engineering at Toyohashi University of Technology, Japan. His research interests include image processing and computer vision. He received the IEICE Best Paper Award in 2005.

(Communicated by *Akihiro Sugimoto*)