Renormalization Returns: Hyper-renormalization and Its Applications

Kenichi Kanatani[1], Ali Al-Sharadqah[2],
Nikolai Chernov[3] and Yasuyuki Sugaya[4]

[1] Department of Computer Science, Okayama University, Okayama 700-8530 Japan
[2]Department of Mathematics, University of Mississippi, Oxford, MS 38677, U.S.A.
[3]Department of Mathematics, University of Alabama at Birmingham, AL 35294, U.S.A., [4] Department of Information and Computer Sciences,
Toyohashi University of Technology, Toyohashi, Aichi 441-8580 Japan
kanatani@suri.cs.okayama-u.ac.jp, alshalaa@gmail.com,
chernov@math.uab.edu, sugaya@iim.ics.tut.ac.jp

Abstract. The technique of "renormalization" for geometric estimation attracted much attention when it was proposed in early 1990s for having higher accuracy than any other then known methods. Later, it was replaced by minimization of the reprojection error. This paper points out that renormalization can be modified so that it outperforms reprojection error minimization. The key fact is that renormalization directly specifies equations to solve, just as the "estimation equation" approach in statistics, rather than minimizing some cost. Exploiting this fact, we do detailed error analysis of the generalized eigenvalue problem so that the renormalization solution has zero bias up to high order error terms; we call the resulting scheme *hyper-renormalization*. We apply it to ellipse fitting to demonstrate that it indeed surpasses reprojection error minimization. We conclude that it is the best method available today.

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 by using geometric constraints. Many problems are formulated as follows. We observe N vector data $\boldsymbol{x}_1$, ..., $\boldsymbol{x}_N$, whose true values $\bar{\boldsymbol{x}}_1$, ..., $\bar{\boldsymbol{x}}_N$ are supposed to satisfy a geometric constraint in the form

$$F(\boldsymbol{x};\boldsymbol{\theta}) = 0, \tag{1}$$

where $\boldsymbol{\theta}$ is an unknown parameter vector which we want to estimate. We call this type of problem simply "geometric estimation". In traditional domains of statistics such as agriculture, pharmaceutics, and economics, observations are regarded as repeated samples from a parameterized probability density model $p_{\boldsymbol{\theta}}(\boldsymbol{x})$; the task is to estimate the parameter $\boldsymbol{\theta}$. We call this type of problem simply "statistical estimation", for which the minimization principle has been a major tool: One chooses the value that minimizes a specified cost. The best known approach is *maximum likelihood* (*ML*), which minimizes the negative log likelihood $l = -\sum_{\alpha=1}^{N} \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_\alpha)$. Recently, an alternative approach is more and more in use: One directly solves specified equations, called *estimating equations* [4], in the form of $\boldsymbol{g}(\boldsymbol{x}_1, ..., \boldsymbol{x}_N, \boldsymbol{\theta}) = \mathbf{0}$. This approach can be viewed as an extension of the minimization principle; ML corresponds to $\boldsymbol{g}(\boldsymbol{x}_1, ..., \boldsymbol{x}_N, \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} l$, known as the *score*. However, the estimating equations need not be the gradient of any function, and one can modify $\boldsymbol{g}(\boldsymbol{x}_1, ..., \boldsymbol{x}_N, \boldsymbol{\theta})$ 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 the norm and is also called the *Gold Standard* [5]. A notable exception is *renormalization* of Kanatani [6, 7]: Instead of minimizing some cost, it iteratively removes bias of weighted least squares (LS). 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., who proposed in [3] an iterative scheme similar to renormalization, which they called *FNS* (*Fundamental Numerical Scheme*), for minimizing what is now referred to as the Sampson error [5]. They argued in [2] that renormalization can be "rationalized" if viewed as approximately minimizing the Sampson error. Leedan and Meer [13] and Matei and Meer [14] also proposed a different iterative scheme, which they called *HEIV* (*Heteroscedastic Errors-in-Variables*), for minimizing the Sampson error. Kanatani and Sugaya [12] pointed out that the reprojection error can be minimized by repeated applications of Sampson error minimization if the Sampson error is iteratively modified so that it agrees with the reprojection error in the end. Thus, the reprojection error minimization is accepted today as the ultimate criterion, and renormalization is regarded as a thing in the past.

In this paper, we note that renormalization is similar to the estimating equation approach for statistical estimation in the sense that it directly specifies equations to solve, which has the form of the generalized eigenvalue problem. We point out that if the form of the generalized eigenvalue problem is modified by doing high order error analysis using the perturbation technique of Kanatani [8], renormalization can achieve higher accuracy than reprojection error minimization. We call the resulting scheme *hyper-renormalization*.

Sec. 2 summarizes the fundamentals of geometric estimation. Sec. 3 describes the iterative reweight, the most primitive form of the non-minimization approach. Sec. 4 reformulates Kanatani's renormalization as an iteratively improvement of the Taubin method. In Sec. 5, we do a detailed error analysis of the generalized eigenvalue problem. In Sec. 6, the procedure of hyper-renormalization is derived as an iteratively improvement of what is called HyperLS. In In Sec. 7, we apply it to ellipse fitting to demonstrate that it indeed outperforms reprojection error minimization. In Sec. 8, we conclude that hyper-renormalization is the best method available today.

2 Geometric Estimation

Equation (1) is a general nonlinear equation in $\boldsymbol{x}$. In many practical problem, we can reparameterize the problem to make $F(\boldsymbol{x};\boldsymbol{\theta})$ linear in $\boldsymbol{\theta}$ (but nonlinear in $\boldsymbol{x}$), allowing us to write Eq. (1) as

$$(\boldsymbol{\xi}(\boldsymbol{x}),\boldsymbol{\theta}) = 0, \tag{2}$$

where and hereafter $(\boldsymbol{a},\boldsymbol{b})$ denotes the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. The vector $\boldsymbol{\xi}(\boldsymbol{x})$ is some nonlinear mapping of $\boldsymbol{x}$ from $\mathcal{R}^m$ to $\mathcal{R}^n$, where m and n are the dimensions of the data $\boldsymbol{x}_\alpha$ and the parameter $\boldsymbol{\theta}$, respectively. 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), $\alpha = 1$, ..., N, we wish to fit an ellipse of the form

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

If we let

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

Eq. (3) has the form of Eq. (2).

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* [5]

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

where $\boldsymbol{F}$ is called the *fundamental matrix*, from which we can compute the camera positions and the 3-D structure of the scene [5, 7]. If we let

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

Eq. (5) has the form of Eq. (2).

We assume that each datum $\boldsymbol{x}_\alpha$ is a deviation from its true value $\bar{\boldsymbol{x}}_\alpha$ by independent Gaussian noise of mean $\boldsymbol{0}$ and covariance matrix $\sigma^2 V_0[\boldsymbol{x}_\alpha]$, where $V_0[\boldsymbol{x}_\alpha]$ 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 $V_0[\boldsymbol{x}_\alpha]$ the *normalized covariance matrix*, and σ the *noise level*. We simply write $\boldsymbol{\xi}_\alpha$ for $\boldsymbol{\xi}(\boldsymbol{x}_\alpha)$. It can be expanded in the form

$$\boldsymbol{\xi}_\alpha = \bar{\boldsymbol{\xi}}_\alpha + \Delta_1\boldsymbol{\xi}_\alpha + \Delta_2\boldsymbol{\xi}_\alpha + \cdots, \tag{7}$$

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

$$\Delta_1\boldsymbol{\xi}_\alpha = \left.\frac{\partial\boldsymbol{\xi}(\boldsymbol{x})}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha} \Delta\boldsymbol{x}_\alpha. \tag{8}$$

We define the covariance matrix of $\boldsymbol{\xi}_\alpha$ by

$$V[\boldsymbol{\xi}_\alpha] = E[\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top] = \left.\frac{\partial\boldsymbol{\xi}(\boldsymbol{x})}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha} E[\Delta\boldsymbol{x}_\alpha\Delta\boldsymbol{x}_\alpha^\top] \left.\frac{\partial\boldsymbol{\xi}(\boldsymbol{x})}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha}^\top = \sigma^2 V_0[\boldsymbol{\xi}_\alpha], \tag{9}$$

where $E[\,\cdot\,]$ denotes expectation, and we define

$$V_0[\boldsymbol{\xi}_\alpha] \equiv \left.\frac{\partial\boldsymbol{\xi}(\boldsymbol{x})}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha} V_0[\boldsymbol{x}_\alpha] \left.\frac{\partial\boldsymbol{\xi}(\boldsymbol{x})}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha}^\top. \tag{10}$$

The true values $\bar{\boldsymbol{x}}_\alpha$ are used in this definition, but in actual computation we replace them by their observations $\boldsymbol{x}_\alpha$. It has been confirmed by many experiments that this does not affect the final result of practical problems. Also, $V_0[\boldsymbol{\xi}_\alpha]$ takes only the first order error 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. The effect of higher order error terms becomes dominant in different places, which we will discuss shortly.

3 Iterative Reweight

The oldest method that is not based on minimization is the following iterative reweight:

1. Let $W_\alpha = 1$, $\alpha = 1, ..., N$, and $\boldsymbol{\theta}_0 = \boldsymbol{0}$.
2. Compute the following matrix $\boldsymbol{M}$:

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top. \tag{11}$$

3. Solve the eigenvalue problem $\boldsymbol{M}\boldsymbol{\theta} = \lambda\boldsymbol{\theta}$ and compute the unit eigenvector $\boldsymbol{\theta}$ for the smallest eigenvalue.
4. If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha \leftarrow \frac{1}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \qquad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \tag{12}$$

and go back to Step 2.

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

$$\frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha(\boldsymbol{\xi}_\alpha,\boldsymbol{\theta})^2 = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha\boldsymbol{\theta}^\top\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top\boldsymbol{\theta} = (\boldsymbol{\theta},\boldsymbol{M}\boldsymbol{\theta}). \tag{13}$$

This is minimized by the unit eigenvector $\boldsymbol{\theta}$ of the matrix $\boldsymbol{M}$ for the smallest eigenvalue. As is well known in statistics, the optimal choice of the weight W_α is the inverse of the variance of that term. Since $(\bar{\boldsymbol{\xi}}_\alpha,\boldsymbol{\theta}) = 0$, we have $(\boldsymbol{\xi}_\alpha,\boldsymbol{\theta}) = (\Delta_1\boldsymbol{\xi}_\alpha,\boldsymbol{\theta}) + \cdots$, and hence the leading term of the variance is

$$E[(\Delta_1\boldsymbol{\xi}_\alpha,\boldsymbol{\theta})^2] = E[\boldsymbol{\theta}^\top\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top\boldsymbol{\theta}] = (\boldsymbol{\theta},E[\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top]\boldsymbol{\theta}) = \sigma^2(\boldsymbol{\theta},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}). \tag{14}$$

Hence, we should choose $W_\alpha = 1/(\boldsymbol{\theta},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})$, but $\boldsymbol{\theta}$ is not known. So, we do iterations, determining the weight W_α from the value of $\boldsymbol{\theta}$ in the preceding step. Let us call the first value of $\boldsymbol{\theta}$ computed with $W_\alpha = 1$ simply the "initial solution". It minimizes $\sum_{\alpha=1}^{N}(\boldsymbol{\xi}_\alpha,\boldsymbol{\theta})^2$, corresponding to what is known as *least squares* (*LS*), *algebraic distance minimization*, and many other names [5]. This, *iterative reweight is an iterative improvement of the LS solution*.

It appears at first sight that the above procedure minimizes

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\boldsymbol{\theta})}{(\boldsymbol{\theta},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \tag{15}$$

which is known today as the *Sampson error* [5]. However, iterative reweight does not minimize it, because at each step we are computing the value of $\boldsymbol{\theta}$ that minimizes the numerator part for the fixed value of the denominator term determined in the preceding step. Hence, at the time of the convergence, the resulting solution $\boldsymbol{\theta}$ is such that

$$\frac{1}{N}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\boldsymbol{\theta})^2}{(\boldsymbol{\theta},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})} \leq \frac{1}{N}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\boldsymbol{\theta}')^2}{(\boldsymbol{\theta},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})} \tag{16}$$

for any $\boldsymbol{\theta}'$, but the following does not necessarily hold:

$$\frac{1}{N}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\boldsymbol{\theta})^2}{(\boldsymbol{\theta},V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})} \leq \frac{1}{N}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha,\boldsymbol{\theta}')^2}{(\boldsymbol{\theta}',V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}')}. \tag{17}$$

The perturbation analysis in [8] tells that the covariance matrix $V[\boldsymbol{\theta}]$ of the resulting solution $\boldsymbol{\theta}$ agrees with a theoretical accuracy limit, called *KCR lower bound* [1, 7, 8], up to $O(\sigma^4)$. Hence, it is practically impossible to reduce the variance any further. However, it has been widely known that the iterative reweight solution has a large bias [7]. For ellipse fitting, for example, it almost always fit 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. (15).

The former is Kanatani's renormalization [6, 7], and the latter is the FNS of Chojnacki et al. [3] and the HEIV of Leedan and Meer [13] and Matei and Meer [14].

4 Renormalization

Kanatani's renormalization [6, 7] can be described as follows:

1. Let $W_\alpha = 1$, $\alpha = 1, ..., N$, and $\boldsymbol{\theta}_0 = \mathbf{0}$.
2. Compute the following matrices $\boldsymbol{M}$ and $\boldsymbol{N}$:

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top, \qquad \boldsymbol{N} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha V_0[\boldsymbol{\xi}_\alpha] \tag{18}$$

3. Solve the generalized eigenvalue problem $\boldsymbol{M\theta} = \lambda \boldsymbol{N\theta}$ and compute the unit eigenvector $\boldsymbol{\theta}$ for the eigenvalue with the smallest magnitude.
4. If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha \leftarrow \frac{1}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \qquad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \tag{19}$$

   and go back to Step 2.

This has a different appearance from the procedure described in [6], in which the generalized eigenvalue problem is reduced to the standard eigenvalue problem, but the resulting solution is the same [7]. The motivation of renormalization is as follows. Let $\bar{\boldsymbol{M}}$ be the true value of the matrix $\boldsymbol{M}$ in Eq. (18) defined by the true values $\bar{\boldsymbol{\xi}}_\alpha$. Since $(\bar{\boldsymbol{\xi}}_\alpha, \boldsymbol{\theta}) = 0$, we have $\bar{\boldsymbol{M}}\boldsymbol{\theta} = \mathbf{0}$. Hence, $\boldsymbol{\theta}$ is the eigenvector of $\bar{\boldsymbol{M}}$ for eigenvalue 0. Since $\bar{\boldsymbol{M}}$ is unknown, we estimate it. Since $E[\Delta\boldsymbol{\xi}_\alpha] = 0$ to a first approximation, the expectation of $\boldsymbol{M}$ is

$$\begin{aligned} E[\boldsymbol{M}] &= E[\frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha(\bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha)(\bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha)^\top] = \bar{\boldsymbol{M}} + \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top] \\ &= \bar{\boldsymbol{M}} + \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} W_\alpha V_0[\boldsymbol{\xi}_\alpha] = \bar{\boldsymbol{M}} + \sigma^2 \boldsymbol{N}. \end{aligned} \tag{20}$$

Thus, $\bar{\boldsymbol{M}} = E[\boldsymbol{M}] - \sigma^2\boldsymbol{N} \approx \boldsymbol{M} - \sigma^2\boldsymbol{N}$, so instead of $\bar{\boldsymbol{M}}\boldsymbol{\theta} = \mathbf{0}$ we solve $(\boldsymbol{M} - \sigma^2\boldsymbol{N})\boldsymbol{\theta} = \mathbf{0}$, or $\boldsymbol{M\theta} = \sigma^2\boldsymbol{N\theta}$. Assuming that σ^2 is small, we regard it as the eigenvalue with the smallest magnitude. As in the case of iterative reweight, we iteratively update the weight W_α so that it approaches $1/(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})$.

Note that 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 [16], known to be very accurate algebraic method without requiring iterations. 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 in [8], showing that the covariance matrix $V[\boldsymbol{\theta}]$ of the renormalization solution $\boldsymbol{\theta}$ agrees with the KCR lower bound up to $O(\sigma^4)$ just as iterative reweight, but the bias is much smaller. That is the reason for the high accuracy of renormalization.

Very small it may be, the bias is not 0, however. The error analysis in [8] shows that the bias estimate expression involves the matrix $\boldsymbol{N}$. Our strategy is to *optimize* the matrix $\boldsymbol{N}$ in Eq. (18) to $\boldsymbol{N} = (1/N)\sum_{\alpha=1}^{N} W_\alpha V_0[\boldsymbol{\xi}_\alpha] + \cdots$ so that *the bias is zero up to high order error terms*. To do is the main theme of this paper.

5 Error Analysis

We now analyze the error of the generalized eigenvalue problem, using the perturbation technique of [8]. Substituting Eq. (7) into the definition of the matrix $\boldsymbol{M}$ in Eq. (18), we can expand it in the form

$$\boldsymbol{M} = \bar{\boldsymbol{M}} + \Delta_1\boldsymbol{M} + \Delta_2\boldsymbol{M} + \cdots, \tag{21}$$

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

$$\Delta_1\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N}\bar{W}_\alpha\Big(\Delta_1\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top\Big) + \frac{1}{N}\sum_{\alpha=1}^{N}\Delta_1\bar{W}_\alpha\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top, \tag{22}$$

$$\begin{aligned}\Delta_2\boldsymbol{M} = {} & \frac{1}{N}\sum_{\alpha=1}^{N}\bar{W}_\alpha\Big(\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top + \Delta_2\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha\Delta_2\boldsymbol{\xi}_\alpha^\top\Big) \\ & + \frac{1}{N}\sum_{\alpha=1}^{N}\Delta_1 W_\alpha(\Delta_1\boldsymbol{\xi}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha\Delta_1\boldsymbol{\xi}_\alpha^\top) + \frac{1}{N}\sum_{\alpha=1}^{N}\Delta_2 W_\alpha\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top.\end{aligned} \tag{23}$$

Let $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots$ be the corresponding expansion of the resulting $\boldsymbol{\theta}$. At the time of convergence, we have $W_\alpha = 1/(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})$. Substituting the expansion of $\boldsymbol{\theta}$, we obtain the expansion $W_\alpha = \bar{W}_\alpha + \Delta_1 W_\alpha + \Delta_2 W_\alpha + \cdots$, where

$$\Delta_1 W_\alpha = -2\bar{W}_\alpha^2(\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}), \tag{24}$$

$$\Delta_2 W_\alpha = \frac{(\Delta_1 W_\alpha)^2}{\bar{W}_\alpha} - \bar{W}_\alpha^2\Big((\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\Delta_1\boldsymbol{\theta}) + 2(\Delta_2\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\Big). \tag{25}$$

(See Supplemental Material). Similarly expanding the eigenvalue λ and the matrix $\boldsymbol{N}$ yet to be determined, the generalized eigenvalue problem $\boldsymbol{M\theta} = \lambda\boldsymbol{N\theta}$ has the form

$$\begin{aligned}&(\bar{\boldsymbol{M}}+\Delta_1\boldsymbol{M}+\Delta_2\boldsymbol{M}+\cdots)(\bar{\boldsymbol{\theta}}+\Delta_1\boldsymbol{\theta}+\Delta_2\boldsymbol{\theta}+\cdots) \\ &= (\bar{\lambda}+\Delta_1\lambda+\Delta_2\lambda+\cdots)(\bar{\boldsymbol{N}}+\Delta_1\boldsymbol{N}+\Delta_2\boldsymbol{N}+\cdots)(\bar{\boldsymbol{\theta}}+\Delta_1\boldsymbol{\theta}+\Delta_2\boldsymbol{\theta}+\cdots).\end{aligned} \tag{26}$$

Equating the noiseless terms on both sides, we have $\bar{\boldsymbol{M}}\bar{\boldsymbol{\theta}} = \bar{\lambda}\boldsymbol{N}\bar{\boldsymbol{\theta}}$, but since $\bar{\boldsymbol{M}}\bar{\boldsymbol{\theta}} = \boldsymbol{0}$, we have $\bar{\lambda} = 0$. Equating the first and the second order terms on both sides, we obtain the following relationships:

$$\bar{\boldsymbol{M}}\Delta_1\boldsymbol{\theta} + \Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}} = \Delta_1\lambda\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}, \tag{27}$$

$$\bar{\boldsymbol{M}}\Delta_2\boldsymbol{\theta} + \Delta_1\boldsymbol{M}\Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{M}\bar{\boldsymbol{\theta}} = \Delta_2\lambda\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}. \tag{28}$$

Computing the inner product of Eq. (27) and $\bar{\boldsymbol{\theta}}$ on both sides, we have

$$(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{M}}\Delta_1\boldsymbol{\theta}) + (\bar{\boldsymbol{\theta}}, \Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}}) = \Delta_1\lambda(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}), \tag{29}$$

but $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{M}}\Delta_1\boldsymbol{\theta}) = (\bar{\boldsymbol{M}}\bar{\boldsymbol{\theta}}, \Delta_1\boldsymbol{\theta}) = 0$ and Eq. (22) implies $(\bar{\boldsymbol{\theta}}, \Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}}) = 0$, so $\Delta_1\lambda = 0$. The matrix $\bar{\boldsymbol{M}}$ has rank $n-1$, (n is the dimension of $\boldsymbol{\theta}$), $\bar{\boldsymbol{\theta}}$ being the null vector. Hence, if we let $\bar{\boldsymbol{M}}^-$ be the pseudoinverse of $\bar{\boldsymbol{M}}$, the product $\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}$ equals the projection matrix $\boldsymbol{P}_{\bar{\boldsymbol{\theta}}}$ in the direction of $\bar{\boldsymbol{\theta}}$. It follows that by multiplying both sides of Eq. (27) by $\bar{\boldsymbol{M}}^-$ from left, $\Delta_1\boldsymbol{\theta}$ is expressed as follows:

$$\Delta_1\boldsymbol{\theta} = -\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}}. \tag{30}$$

Here, we have noted that since $\boldsymbol{\theta}$ is normalized to unit norm, $\Delta_1\boldsymbol{\theta}$ is orthogonal to $\bar{\boldsymbol{\theta}}$ so $\boldsymbol{P}_{\bar{\boldsymbol{\theta}}}\Delta_1\boldsymbol{\theta} = \Delta_1\boldsymbol{\theta}$. Substituting Eq. (30) into Eq. (28), we obtain

$$\Delta_2\lambda\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}} = \bar{\boldsymbol{M}}\Delta_2\boldsymbol{\theta} - \Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}} + \Delta_2\boldsymbol{M}\bar{\boldsymbol{\theta}} = \bar{\boldsymbol{M}}\Delta_2\boldsymbol{\theta} + \boldsymbol{T}\bar{\boldsymbol{\theta}}, \tag{31}$$

where we define the matrix $\boldsymbol{T}$ to be

$$\boldsymbol{T} \equiv \Delta_2\boldsymbol{M} - \Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}. \tag{32}$$

Because $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$ to be the orthogonal component

$$\Delta_2^{\perp}\boldsymbol{\theta} \equiv \boldsymbol{P}_{\bar{\boldsymbol{\theta}}}\Delta_2\boldsymbol{\theta} = \bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\Delta_2\boldsymbol{\theta}. \tag{33}$$

Multiplying Eq. (31) by $\bar{\boldsymbol{M}}^-$ on both sides from left, we obtain $\Delta_2^{\perp}\boldsymbol{\theta}$ in the following form:

$$\Delta_2^{\perp}\boldsymbol{\theta} = \bar{\boldsymbol{M}}^-(\Delta_2\lambda\bar{\boldsymbol{N}} - \boldsymbol{T})\bar{\boldsymbol{\theta}}. \tag{34}$$

Computing the inner product of Eq. (31) and $\bar{\boldsymbol{\theta}}$ on both sides and noting that $(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{M}}\Delta_2\boldsymbol{\theta}) = 0$, we obtain $\Delta_2\lambda$ in the form

$$\Delta_2\lambda = \frac{(\bar{\boldsymbol{\theta}}, \boldsymbol{T}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}})}. \tag{35}$$

Hence, Eq. (34) is rewritten as follows:

$$\Delta_2^{\perp}\boldsymbol{\theta} = \bar{\boldsymbol{M}}^-\Big(\frac{(\bar{\boldsymbol{\theta}}, \boldsymbol{T}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}})}\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}} - \boldsymbol{T}\bar{\boldsymbol{\theta}}\Big). \tag{36}$$

6 Hyper-renormalization

From Eq. (30), we see that $E[\Delta_1\boldsymbol{\theta}] = \mathbf{0}$: the first order bias is $\mathbf{0}$. Thus, the bias is evaluated by the second order term $E[\Delta_2^{\perp}\boldsymbol{\theta}]$. From Eq. (36), we obtain

$$E[\Delta_2^{\perp}\boldsymbol{\theta}] = \bar{\boldsymbol{M}}^{-}\Big(\frac{(\bar{\boldsymbol{\theta}}, E[\boldsymbol{T}\bar{\boldsymbol{\theta}}])}{(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}})}\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}} - E[\boldsymbol{T}\bar{\boldsymbol{\theta}}]\Big), \tag{37}$$

which implies that *if we can choose such an* $\boldsymbol{N}$ *that* $E[\boldsymbol{T}\bar{\boldsymbol{\theta}}] = c\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}$ *for some constant* c, we will have $E[\Delta_2^{\perp}\boldsymbol{\theta}] = \mathbf{0}$. Then, the bias will be $O(\sigma^4)$, since the expectation of odd-order error terms is zero. In order to choose such an $\boldsymbol{N}$, we need to evaluate the expectation $E[\boldsymbol{T}\bar{\boldsymbol{\theta}}]$. After a lengthy analysis (see Supplemental Material), we find that $E[\boldsymbol{T}\bar{\boldsymbol{\theta}}] = \sigma^2\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}$ holds if we define

$$\begin{aligned}\bar{\boldsymbol{N}} = &\frac{1}{N}\sum_{\alpha=1}^{N}\bar{W}_\alpha\Big(V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[\bar{\boldsymbol{\xi}}_\alpha\boldsymbol{e}_\alpha^\top]\Big)\\ &-\frac{1}{N^2}\sum_{\alpha=1}^{N}\bar{W}_\alpha^2\Big((\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^{-}\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top],\Big)\end{aligned} \tag{38}$$

where $\mathcal{S}[\,\cdot\,]$ denotes symmetrization ($\mathcal{S}[\boldsymbol{A}] = (\boldsymbol{A} + \boldsymbol{A}^\top)/2$) and the vectors $\boldsymbol{e}_\alpha$ are defined via

$$E[\Delta_2\boldsymbol{\xi}_\alpha] = \sigma^2\boldsymbol{e}_\alpha. \tag{39}$$

This is the core contribution of this paper. From this result, we obtain the following *hyper-renormalization*:

1. Let $W_\alpha = 1$, $\alpha = 1$, ..., N, and $\boldsymbol{\theta}_0 = \mathbf{0}$.
2. Compute the following matrices $\boldsymbol{M}$ and $\boldsymbol{N}$:

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N}W_\alpha\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top, \tag{40}$$

$$\begin{aligned}\boldsymbol{N} = &\frac{1}{N}\sum_{\alpha=1}^{N}W_\alpha\Big(V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[\boldsymbol{\xi}_\alpha\boldsymbol{e}_\alpha^\top]\Big)\\ &-\frac{1}{N^2}\sum_{\alpha=1}^{N}W_\alpha^2\Big((\boldsymbol{\xi}_\alpha, \boldsymbol{M}_{n-1}^{-}\boldsymbol{\xi}_\alpha)V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{M}_{n-1}^{-}\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top]\Big).\end{aligned} \tag{41}$$

   Here, $\boldsymbol{M}_{n-1}^{-}$ is the pseudoinverse of $\boldsymbol{M}$ with truncated rank $n-1$, i.e., with the smallest eigenvalue replaced by 0 in the spectral decomposition.
3. Solve the generalized eigenvalue problem $\boldsymbol{M}\boldsymbol{\theta} = \lambda\boldsymbol{N}\boldsymbol{\theta}$ and compute the unit eigenvector $\boldsymbol{\theta}$ for the eigenvalue with the smallest magnitude.
4. If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha \leftarrow \frac{1}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \qquad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \tag{42}$$

   and go back to Step 2.

It turns out that the initial solution with $W_\alpha = 1$ coincides with what is called *HyperLS* [10, 11, 15], which is derived to remove the bias up to second order error terms within the framework of algebraic methods without iterations. The expression of Eq. (41) with $W_\alpha = 1$ lacks one term as compared with the corresponding expression of HyperLS, but the same solution is produced. We omit the details, but all the intermediate solutions $\boldsymbol{\theta}$ in the hyper-renormalization iterations are shown to be free of second oder bias. Thus, *hyper-renormalization is an iterative improvement of HyperLS*. As in the case of iterative reweight and renormalization, the covariance matrix $V[\boldsymbol{\theta}]$ of the hyper-renormalization solution $\boldsymbol{\theta}$ agrees with the KCR lower bound up to $O(\sigma^4)$ (see Supplemental Material).

Standard linear algebra routines for solving the generalized eigenvalue problem $\boldsymbol{M}\boldsymbol{\theta} = \lambda \boldsymbol{N}\boldsymbol{\theta}$ assume that $\boldsymbol{N}$ is positive definite, but the matrix $\boldsymbol{N}$ in Eq. (41) has both positive and negative eigenvalues. For the Taubin method and renormalization, the matrix $\boldsymbol{N}$ in Eq. (18) is positive semidefinite with a zero eigenvalue. This, however, causes no problem, because the problem can be rewritten as

$$\boldsymbol{N}\boldsymbol{\theta} = \frac{1}{\lambda}\boldsymbol{M}\boldsymbol{\theta}. \tag{43}$$

The matrix $\boldsymbol{M}$ in Eq. (40) is positive definite for noisy data, so we can use a standard routine to compute the eigenvector $\boldsymbol{\theta}$ for the eigenvalue $1/\lambda$ with the largest absolute value. If the matrix $\boldsymbol{M}$ happens to have a zero eigenvalue, it indicates that the data are all exact, so the unit eigenvector for the eigenvalue 0 is the exact solution.

7 Ellipse Fitting Experiment

We define 30 equidistant points on the ellipse shown in Fig. 1(a). The major and minor axis are set to 100 and 50 pixels, respectively. We add random Gaussian noise of mean 0 and standard deviation σ to the x and y coordinates of each point independently and fit an ellipse to the noisy point sequence using the following methods: 1. LS, 2. iterative reweight, 3. the Taubin method, 4. renormalization, 5. HyperLS, 6. hyper-renormalization, 7. ML, 8. ML with hyperaccurate correction.

For our noise, ML means reprojection error minimization, which can be computed by repeated Sampson error minimization as pointed out by Kanatani and Sugaya [12]. We used the FNS of Chojnacki et al. [3] for minimizing the Sampson error, but according to our experiments, the FNS solution agrees with the ML solution up to three or four significant digits, as also observed in [12]. So, we identified the FNS solution with the ML solution. Kanatani [9] analytically evaluated the bias of FNS to the second order terms and subtracted it from the FNS solution; he called this scheme *hyperaccurate correction*.

Figures 1(b), (c) show fitting examples for $\sigma = 0.5$; although the noise magnitude is fixed, fitted ellipses are different for different noise. The true shape is indicated by dotted lines. Iterative reweight, renormalization, and hyper-renormalization all converged after four iterations, while FNS for ML computation required nine iterations for Fig. 1(b) and eight iterations for Fig. 1(c).

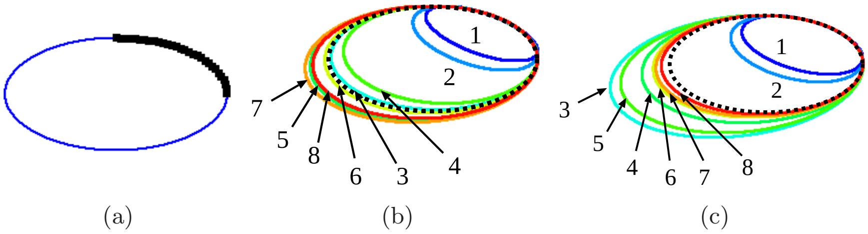


Fig. 1. (a) Thirty points on an ellipse. (b), (c) Fitted ellipses ($\sigma = 0.5$). 1. LS, 2. iterative reweight, 3. the Taubin method, 4. renormalization, 5. HyperLS, 6. hyper-renormalization, 7. ML, 8. ML with hyperaccurate correction. The dotted lines indicate the true shape.

We can see that the LS and iterative reweight solutions have large bias, producing much smaller ellipses than the true shape. The closest ellipse is given by hyper-renormalization in Fig. 1(b) and by ML with hyperaccurate correction in Fig. 1(c). Thus, the solution is different for different noise, so statistical tests are necessary for a fair comparison.

Since the computed $\boldsymbol{\theta}$ and its true value $\bar{\boldsymbol{\theta}}$ are both unit vectors, we measure their discrepancy by the orthogonal component $\Delta^{\perp}\boldsymbol{\theta} = \boldsymbol{P}_{\bar{\boldsymbol{\theta}}}\boldsymbol{\theta}$, where $\boldsymbol{P}_{\bar{\boldsymbol{\theta}}}$ ($\equiv \boldsymbol{I} - \bar{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}^{\top}$) is the orthogonal projection matrix along $\bar{\boldsymbol{\theta}}$ (Fig. 2(a)). We generated 10000 independent noise instances for each σ and evaluated the bias B (Fig. 2(b)) and the RMS (root-mean-square) error D (Fig. 2(c)) defined by

$$B = \left\| \frac{1}{10000} \sum_{a=1}^{10000} \Delta^{\perp}\boldsymbol{\theta}^{(a)} \right\|, \qquad D = \sqrt{\frac{1}{10000} \sum_{a=1}^{10000} \|\Delta^{\perp}\boldsymbol{\theta}^{(a)}\|^2}, \tag{44}$$

where $\boldsymbol{\theta}^{(a)}$ is the solution in the ath trial. The dotted line in Fig. 2(c) indicates the theoretical limit, called the *KCR lower bound* [1, 7, 8], defined by

$$D_{\mathrm{KCR}} = \frac{\sigma}{\sqrt{N}} \sqrt{\mathrm{tr}\bar{\boldsymbol{M}}^{-}}, \tag{45}$$

where $\bar{\boldsymbol{M}}^{-}$ is the pseudoinverse of the true value $\bar{\boldsymbol{M}}$ (of rank 5) of the matrix $\boldsymbol{M}$ in Eqs. (11), (18), and (40), and tr stands for the trace.

The interrupted plots in Fig. 2(b) for iterative reweight, ML, and ML with hyperaccurate correction indicate that the iterations did not converge beyond that noise level. Our convergence criterion is $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < 10^{-6}$ for the current value $\boldsymbol{\theta}$ and the value $\boldsymbol{\theta}_0$ 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 10000 trials does not converge. Figure 3 shows the enlargements of Figs. 2(b), (c) for the small σ part.

We can see from Fig. 2(b) that LS and iterative reweight have very large bias, in contrast to which the bias of the Taubin method and renormalization is very

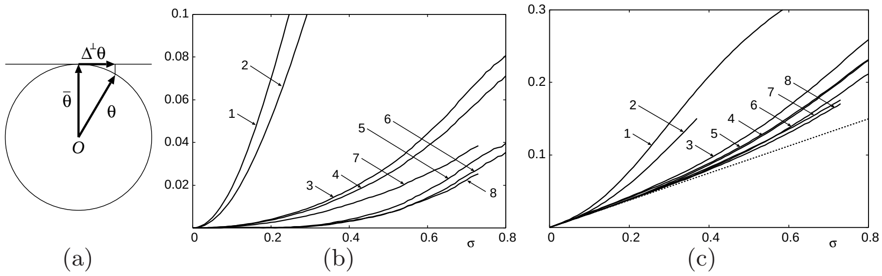


Fig. 2. (a) The true value $\bar{\boldsymbol{\theta}}$, the computed value $\boldsymbol{\theta}$, and its orthogonal component $\Delta^{\perp}\boldsymbol{\theta}$ of $\bar{\boldsymbol{\theta}}$. (b), (c) The bias (a) and the RMS error (b) of the fitted ellipse for the standard deviation σ of the added noise over 10000 independent trials. 1. LS, 2. iterative reweight, 3. the Taubin method, 4. renormalization, 5. HyperLS, 6. hyper-renormalization, 7. ML, 8. ML with hyperaccurate correction. The dotted line in (c) indicates the KCR lower bound.

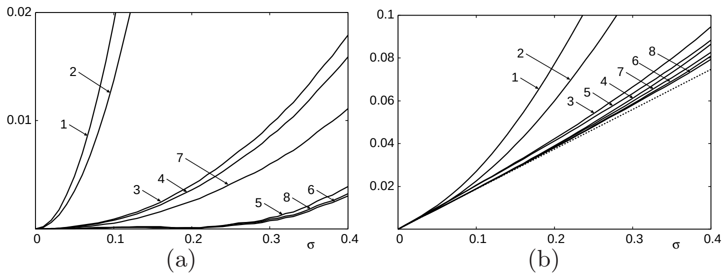
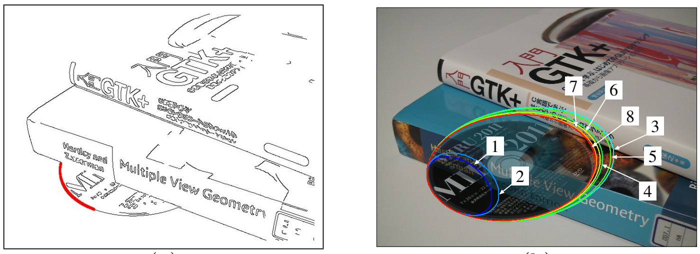


Fig. 3. (a) Enlargement of Fig. 2(b). (b) Enlargement of Fig. 2(c).

small. The bias of HyperLS and hyper-renormalization is still smaller and even smaller than ML. Since the leading covariance is common to iterative reweight, renormalization, and hyper-renormalization, the RMS reflects the magnitude of the bias as shown in Fig. 2(c). Because the hyper-renormalization solution does not have bias up to high order error terms, it has nearly the same accuracy as ML, or reprojection error minimization. A close examination of the small σ part (Fig. 3(b)) reveals that hyper-renormalization outperforms ML. The highest accuracy is achieved, although the difference is very small, by Kanatani's hyperaccurate correction of ML [9]. 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 Figs. 2(b), (c). On the other hand, hyper-renormalization is very robust to noise. This is because the initial solution is HyperLS, which is itself highly accurate already as shown in Figs. 2 and 3. For this reason, we conclude that it is the best method for practical computations.

Figure 4(a) is an edge image of a scene with a circular object. We fitted an ellipse to the 160 edge points indicated in red, using various methods. Figure 4(b)



(a) (b)

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

shows the fitted ellipses superimposed on the original image, where the occluded part is also artificially composed for visual ease. In this case, iterative reweight converged after four iterations, and renormalization and hyper-renormalization converged after three iterations, while FNS for ML computation required six iterations. We can see that LS and iterative reweight produce much smaller ellipses than the true shape as in Fig. 1(b), (c). All other fits are very close to the true ellipse, and ML gives the best fit in this particular instance.

8 Conclusions

We have reformulated iterative reweight and renormalization as geometric estimation techniques not based on the minimization principle and optimized the matrix $\boldsymbol{N}$ that appears in the renormalization computation so that the resulting solution has no bias up to the second order noise terms. We called the resulting scheme "hyper-renormalization" and applied it to ellipse fitting (see Supplemental Material for fundamental matrix computation). We observed:

1. Iterative reweight is an iterative improvement of LS. The leading covariance of the solution agrees with the KCR lower bound, but the bias is very large, so the accuracy is low.
2. Renormalization is an iterative improvement of the Taubin method. The leading covariance of the solution agrees with the KCR lower bound, and the bias is very small, so the accuracy is high.
3. Hyper-renormalization is an iterative improvement of HyperLS. The leading covariance of the solution agrees with the KCR lower bound with no bias up to high order error terms. Its accuracy outperforms ML (reprojection error minimization).
4. Although the difference is very small, ML with hyperaccurate correction exhibits the highest accuracy, but the iterations for its computation may

not converge in the presence of large noise, while hyper-renormalization is robust to noise.

We conclude that hyper-renormalization is the best method for practical computations.

References

1. N. Chernov and C. Lesort, Statistical efficiency of curve fitting algorithms, *Comp. Stat. Data Anal.*, **47**-4 (2004-11), 713–728.
2. W. Chojnacki, M. J. Brooks and A. van den Hengel, Rationalising the renormalization method of Kanatani, *J. Math. Imaging Vis.*, **21**-11 (2001-2), 21–38.
3. W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley, On the fitting of surfaces to data with covariances, *IEEE Trans. Patt. Anal. Mach. Intell.*, **22**-11 (2000-11), 1294–1303.
4. V. P. Godambe (Ed.), *Estimating Functions*, Oxford University Press, New York, U.S.A., 1991.
5. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, U.K., 2000.
6. K. Kanatani, Renormalization for unbiased estimation, *Pro. 4th Int. Conf. Comput. Vis. (ICCV'93)*, May 1993, Berlin, Germany, pp. 599–606.
7. K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice* Elsevier, Amsterdam, the Netherlands, 1996; reprinted, Dover, York, NY, U.S.A., 2005.
8. K. Kanatani, Statistical optimization for geometric fitting: Theoretical accuracy analysis and high order error analysis, *Int. J. Comput. Vis.*, **80**-2 (2008-11), 167–188.
9. K. Kanatani, Ellipse fitting with hyperaccuracy, *IEICE Trans. Inf. & Syst.*, **E89-D**-10 (2006-10), 2653–2660.
10. K. Kanatani and P. Rangarajan, Hyper least squares fitting of circles and ellipses, *Comput. Stat. Data Anal.*, **55**-6 (2011-6), 2197–2208.
11. K. Kanatani, P. Rangarajan, Y. Sugaya and H. Niitsuma, HyperLS and its applications, *IPSJ Trans. Comput. Vis. Appl.*, **3** (2011-10), 80–94.
12. K. Kanatani and Y. Sugaya, Unified computation of strict maximum likelihood for geometric fitting, *J. Math. Imaging Vis.*, **38**-1 (2010-9), 1–13.
13. Y. Leedan and P. Meer, Heteroscedastic regression in computer vision: Problems with bilinear constraint, *Int. J. Comput. Vis.*. **37**-2 (2000-6), 127-150.
14. J. Matei and P. Meer, Estimation of nonlinear errors-in-variables models for computer vision applications, *IEEE Trans. Patt. Anal. Mach. Intell.*, **28**-10 (2006-10), 1537–1552.
15. P. Rangarajan and K. Kanatani, Improved algebraic methods for circle fitting, *Electronic J. Stat.*, **3** (2009-10), 1075–1082.
16. G. Taubin, Estimation of planar curves, surfaces, and non-planar space curves defined by implicit equations with applications to edge and range image segmentation, *IEEE Trans. Patt. Anal. Mach. Intell.*, **13**-11 (1991-11), 1115–1138.

Supplemental Material

Contents

A Expansion of the weights W_α

In order to evaluate $\Delta_1 W_\alpha$ and $\Delta_2 W_\alpha$ in Eqs. (22) and (23), we write $W_\alpha = 1/(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})$ as $W_\alpha(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}) = 1$ and substitute the expansions of W_α and $\boldsymbol{\theta}$:

$$(\bar{W}_\alpha + \Delta_1 W_\alpha + \Delta_2 W_\alpha + \cdots)\Big((\bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots, V_0[\boldsymbol{\xi}_\alpha](\bar{\boldsymbol{\theta}} + \Delta_1\boldsymbol{\theta} + \Delta_2\boldsymbol{\theta} + \cdots)\Big)$$
$$= 1. \tag{46}$$

Equating the noiseless terms on both sides, we obtain $\bar{W}_\alpha(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) = 1$. Equating the first order terms on both sides, we obtain

$$\Delta_1 W_\alpha(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) + \bar{W}_\alpha(\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) + \bar{W}_\alpha(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\Delta_1\bar{\boldsymbol{\theta}}) = 0, \tag{47}$$

from which we obtain

$$\Delta_1 W_\alpha = -\frac{2\bar{W}_\alpha(\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})} = -2\bar{W}_\alpha^2(\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}). \tag{48}$$

Equating the second order terms on both sides of Eq. (46), we obtain

$$\Delta_2 W_\alpha(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) + \Delta_1 W_\alpha\Big((\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) + (\bar{\boldsymbol{\theta}}V_0[\boldsymbol{\xi}_\alpha], \Delta_1\boldsymbol{\theta})\Big)$$
$$+\bar{W}_\alpha\Big((\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\Delta_1\boldsymbol{\theta}) + (\Delta_2\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) + (\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\Delta_2\boldsymbol{\theta})\Big) = 0, \tag{49}$$

from which we obtain

$$\begin{aligned}
\Delta_2 W_\alpha &= -\frac{1}{(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})}\Big(2\Delta_1 W_\alpha(\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) + \bar{W}_\alpha\Big((\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\Delta_1\boldsymbol{\theta}) \\
&\quad +2(\Delta_2\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\Big)\Big) \\
&= -\bar{W}_\alpha\Big(2\Delta_1 W_\alpha\Big(-\frac{\Delta_1 W_\alpha}{2\bar{W}_\alpha^2}\Big) + \bar{W}_\alpha\Big((\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\Delta_1\boldsymbol{\theta}) + 2(\Delta_2\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\Big)\Big) \\
&= \frac{(\Delta_1 W_\alpha)^2}{\bar{W}_\alpha} - \bar{W}_\alpha^2\Big((\Delta_1\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\Delta_1\boldsymbol{\theta}) + 2(\Delta_2\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\Big).
\end{aligned} \tag{50}$$

Thus, we obtain Eqs. (24) and (25).

B Derivation of Hyper-renormalization

We evaluate the expectation of $\boldsymbol{T}\bar{\boldsymbol{\theta}} = \Delta_2 \boldsymbol{M}\bar{\boldsymbol{\theta}} - \Delta_1 \boldsymbol{M}\bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M}\bar{\boldsymbol{\theta}}$ and derive the hyper-renormalization procedure in several steps.

B.1 Evaluation of $E[\boldsymbol{\Delta_2 M}\bar{\boldsymbol{\theta}}]$

From Eq. (22) and $(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{\theta}}) = 0$, we can express $\Delta_2 \boldsymbol{M}\bar{\boldsymbol{\theta}}$ in the following form:

$$\begin{aligned}
\Delta_2 \boldsymbol{M}\bar{\boldsymbol{\theta}} &= \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(\Delta_1 \boldsymbol{\xi}_\alpha \Delta_1 \boldsymbol{\xi}_\alpha^\top + \Delta_2 \boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta_2 \boldsymbol{\xi}_\alpha^\top \Big) \bar{\boldsymbol{\theta}} \\
&\quad + \frac{1}{N}\sum_{\alpha=1}^{N} \Delta_1 W_\alpha (\Delta_1 \boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta_1 \boldsymbol{\xi}_\alpha^\top) \bar{\boldsymbol{\theta}} + \frac{1}{N}\sum_{\alpha=1}^{N} \Delta_2 W_\alpha \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top \bar{\boldsymbol{\theta}} \\
&= \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha ((\Delta_1 \boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}}) \Delta_1 \boldsymbol{\xi}_\alpha + (\Delta_2 \boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}}) \bar{\boldsymbol{\xi}}_\alpha) \\
&\quad + \frac{1}{N}\sum_{\alpha=1}^{N} \Delta_1 W_\alpha (\Delta_1 \boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}}) \bar{\boldsymbol{\xi}}_\alpha. \qquad (51)
\end{aligned}$$

Hence, $E[\Delta_2 \boldsymbol{M}\bar{\boldsymbol{\theta}}]$ is

$$\begin{aligned}
E[\Delta_2 \boldsymbol{M}\bar{\boldsymbol{\theta}}] &= \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(E[\Delta_1 \boldsymbol{\xi}_\alpha \Delta_1 \boldsymbol{\xi}_\alpha^\top] \bar{\boldsymbol{\theta}} + (E[\Delta_2 \boldsymbol{\xi}_\alpha], \bar{\boldsymbol{\theta}}) \bar{\boldsymbol{\xi}}_\alpha \Big) \\
&\quad + \frac{1}{N}\sum_{\alpha=1}^{N} (E[\Delta_1 W_\alpha \Delta_1 \boldsymbol{\xi}_\alpha], \bar{\boldsymbol{\theta}}) \bar{\boldsymbol{\xi}}_\alpha \\
&= \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(V_0[\boldsymbol{\xi}_\alpha] \bar{\boldsymbol{\theta}} + (\boldsymbol{e}_\alpha, \bar{\boldsymbol{\theta}}) \bar{\boldsymbol{\xi}}_\alpha \Big) + \frac{1}{N}\sum_{\alpha=1}^{N} (E[\Delta_1 W_\alpha \Delta_1 \boldsymbol{\xi}_\alpha], \bar{\boldsymbol{\theta}}) \bar{\boldsymbol{\xi}}_\alpha. \qquad (52)
\end{aligned}$$

We now consider the expectation of $\Delta_1 W_\alpha \Delta_1 \boldsymbol{\xi}_\alpha$. From Eqs. (22), (24), (30), we obtain the following expressions:

$$\begin{aligned}
\Delta_1 W_\alpha &= -2\bar{W}_\alpha^2 (\Delta_1 \boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha] \bar{\boldsymbol{\theta}}) = 2\bar{W}_\alpha^2 (\bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M}\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha] \bar{\boldsymbol{\theta}}) \\
&= 2\bar{W}_\alpha^2 (\bar{\boldsymbol{M}}^- \Big(\frac{1}{N}\sum_{\beta=1}^{N} \bar{W}_\beta \Big(\Delta_1 \boldsymbol{\xi}_\beta \bar{\boldsymbol{\xi}}_\beta^\top + \bar{\boldsymbol{\xi}}_\beta \Delta_1 \boldsymbol{\xi}_\beta^\top \Big) + \frac{1}{N}\sum_{\beta=1}^{N} \Delta_1 \bar{W}_\beta \bar{\boldsymbol{\xi}}_\beta \bar{\boldsymbol{\xi}}_\beta^\top \Big) \bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha] \bar{\boldsymbol{\theta}}) \\
&= \frac{2}{N}\sum_{\beta=1}^{N} \bar{W}_\alpha^2 \bar{W}_\beta (\Delta_1 \boldsymbol{\xi}_\beta, \bar{\boldsymbol{\theta}}) (\bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\beta, V_0[\boldsymbol{\xi}_\alpha] \bar{\boldsymbol{\theta}}) \\
&= \frac{2}{N}\sum_{\beta=1}^{N} \bar{W}_\alpha^2 \bar{W}_\beta (\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha] \bar{\boldsymbol{\theta}}) (\Delta_1 \boldsymbol{\xi}_\beta, \bar{\boldsymbol{\theta}}), \qquad (53)
\end{aligned}$$

$$E[\Delta_1 W_\alpha \Delta_1 \boldsymbol{\xi}_\alpha] = E[\frac{2}{N}\sum_{\beta=1}^{N} \bar{W}_\alpha^2 \bar{W}_\beta (\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})(\Delta_1 \boldsymbol{\xi}_\beta, \bar{\boldsymbol{\theta}})\Delta_1 \boldsymbol{\xi}_\alpha]$$

$$= \frac{2}{N}\sum_{\beta=1}^{N} \bar{W}_\alpha^2 \bar{W}_\beta (\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) E[\Delta_1 \boldsymbol{\xi}_\alpha \Delta_1 \boldsymbol{\xi}_\beta^\top]\bar{\boldsymbol{\theta}}$$

$$= \frac{2}{N}\sum_{\beta=1}^{N} \bar{W}_\alpha^2 \bar{W}_\beta (\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) \sigma^2 \delta_{\alpha\beta} V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}$$

$$= \frac{2\sigma^2}{N} \bar{W}_\alpha^3 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}, \tag{54}$$

$$\frac{1}{N}\sum_{\alpha=1}^{N} (E[\Delta_1 W_\alpha \Delta_1 \boldsymbol{\xi}_\alpha], \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha = \frac{1}{N}\sum_{\alpha=1}^{N} (\frac{2\sigma^2}{N} \bar{W}_\alpha^3 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}) V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha$$

$$= \frac{2\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^3 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha$$

$$= \frac{2\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha. \tag{55}$$

Here, we have noted that error in $\boldsymbol{\xi}_\alpha$ is independent for different α and hence $E[\Delta_1 \boldsymbol{\xi}_\alpha \Delta_1 \boldsymbol{\xi}_\beta^\top] = \delta_{\alpha\beta} V_0[\boldsymbol{\xi}_\alpha]$, where $\delta_{\alpha\beta}$ is the Kronecker delta. From the above expressions, we can write $E[\Delta_2 \boldsymbol{M}\bar{\boldsymbol{\theta}}]$ as follows:

$$E[\Delta_2 \boldsymbol{M}\bar{\boldsymbol{\theta}}] = \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} + (\boldsymbol{e}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha\Big) + \frac{2\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha. \tag{56}$$

B.2 Evaluation of $E[\Delta_1 \boldsymbol{M}\bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M}\bar{\boldsymbol{\theta}}]$

We next consider the expectation of $\Delta_1 \boldsymbol{M}\bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M}\bar{\boldsymbol{\theta}}$. From Eq. (22), we can write

$$\Delta_1 \boldsymbol{M}\bar{\boldsymbol{\theta}} = \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(\Delta_1 \boldsymbol{\xi}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top + \bar{\boldsymbol{\xi}}_\alpha \Delta_1 \boldsymbol{\xi}_\alpha^\top\Big)\bar{\boldsymbol{\theta}} + \frac{1}{N}\sum_{\alpha=1}^{N} \Delta_1 \bar{W}_\alpha \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top \bar{\boldsymbol{\theta}}$$

$$= \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha (\Delta_1 \boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha, \tag{57}$$

$$\Delta_1 \boldsymbol{M}\bar{\boldsymbol{M}}^- \Delta_1 \boldsymbol{M}\bar{\boldsymbol{\theta}} = \Delta_1 \boldsymbol{M}\bar{\boldsymbol{M}}^- \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha (\Delta_1 \boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha$$

$$= \Big(\frac{1}{N}\sum_{\beta=1}^{N} \bar{W}_\beta \Big(\Delta_1 \boldsymbol{\xi}_\beta \bar{\boldsymbol{\xi}}_\beta^\top + \bar{\boldsymbol{\xi}}_\beta \Delta_1 \boldsymbol{\xi}_\beta^\top\Big) + \frac{1}{N}\sum_{\beta=1}^{N} \Delta_1 \bar{W}_\beta \bar{\boldsymbol{\xi}}_\beta \bar{\boldsymbol{\xi}}_\beta^\top\Big)\bar{\boldsymbol{M}}^- \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha (\Delta_1 \boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta \Big(\mathit{\Delta}_1\boldsymbol{\xi}_\beta \bar{\boldsymbol{\xi}}_\beta^\top + \bar{\boldsymbol{\xi}}_\beta \mathit{\Delta}_1\boldsymbol{\xi}_\beta^\top\Big)\bar{\boldsymbol{M}}^-(\mathit{\Delta}_1\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha \\
&\quad + \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \mathit{\Delta}_1 \bar{W}_\beta \bar{\boldsymbol{\xi}}_\beta \bar{\boldsymbol{\xi}}_\beta^\top \bar{\boldsymbol{M}}^-(\mathit{\Delta}_1\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta (\mathit{\Delta}_1\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})\Big(\mathit{\Delta}_1\boldsymbol{\xi}_\beta \bar{\boldsymbol{\xi}}_\beta^\top + \bar{\boldsymbol{\xi}}_\beta \mathit{\Delta}_1\boldsymbol{\xi}_\beta^\top\Big)\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha \\
&\quad + \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \mathit{\Delta}_1 \bar{W}_\beta (\mathit{\Delta}_1\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\beta \bar{\boldsymbol{\xi}}_\beta^\top \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta (\mathit{\Delta}_1\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})(\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\mathit{\Delta}_1\boldsymbol{\xi}_\beta \ \ \text{(Let this term be } \boldsymbol{t}_1\text{)} \\
&\quad + \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta (\mathit{\Delta}_1\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})(\mathit{\Delta}_1\boldsymbol{\xi}_\beta, \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\beta \ \ \text{(Let this term be } \boldsymbol{t}_2\text{)} \\
&\quad + \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \mathit{\Delta}_1 \bar{W}_\beta (\mathit{\Delta}_1\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})(\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\beta. \ \ \text{(Let this term be } \boldsymbol{t}_3\text{)}
\end{aligned}
\tag{58}
$$

Consider the expectation of each term. The expectation of the first term $\boldsymbol{t}_1$ is

$$
\begin{aligned}
E[\boldsymbol{t}_1] &= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta (\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha) E[\mathit{\Delta}_1\boldsymbol{\xi}_\beta \mathit{\Delta}_1\boldsymbol{\xi}_\alpha^\top]\bar{\boldsymbol{\theta}} \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta (\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\sigma^2\delta_{\alpha\beta}V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} \\
&= \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}.
\end{aligned}
\tag{59}
$$

The expectation of the second term $\boldsymbol{t}_2$ is

$$
\begin{aligned}
E[\boldsymbol{t}_2] &= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta (\bar{\boldsymbol{\theta}}, E[\mathit{\Delta}_1\boldsymbol{\xi}_\alpha \mathit{\Delta}_1\boldsymbol{\xi}_\beta^\top]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\beta \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta (\bar{\boldsymbol{\theta}}, \sigma^2\delta_{\alpha\beta}V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\beta \\
&= \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\alpha = \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^-V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha.
\end{aligned}
\tag{60}
$$

In order to evaluate the expectation of the last term $\boldsymbol{t}_3$, we note that we obtain from Eqs. (53) and (54) the following expressions:

$$\varDelta_1 W_\beta = \frac{2}{N}\sum_{\gamma=1}^{N} \bar{W}_\beta^2 \bar{W}_\gamma (\varDelta_1 \boldsymbol{\xi}_\gamma, \bar{\boldsymbol{\theta}})(\bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\gamma, V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}})$$
$$= \frac{2}{N}\sum_{\gamma=1}^{N} \bar{W}_\beta^2 \bar{W}_\gamma (\bar{\boldsymbol{\xi}}_\gamma, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}})(\varDelta_1 \boldsymbol{\xi}_\gamma, \bar{\boldsymbol{\theta}}), \tag{61}$$

$$E[\varDelta_1 W_\beta \varDelta_1 \boldsymbol{\xi}_\alpha] = E[\frac{2}{N}\sum_{\gamma=1}^{N} \bar{W}_\beta^2 \bar{W}_\gamma (\bar{\boldsymbol{\xi}}_\gamma, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}})(\varDelta_1 \boldsymbol{\xi}_\gamma, \bar{\boldsymbol{\theta}})\varDelta_1 \boldsymbol{\xi}_\alpha]$$
$$= \frac{2}{N}\sum_{\gamma=1}^{N} \bar{W}_\beta^2 \bar{W}_\gamma (\bar{\boldsymbol{\xi}}_\gamma, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}) E[\varDelta_1 \boldsymbol{\xi}_\alpha \varDelta_1 \boldsymbol{\xi}_\gamma^\top]\bar{\boldsymbol{\theta}}$$
$$= \frac{2}{N}\sum_{\gamma=1}^{N} \bar{W}_\beta^2 \bar{W}_\gamma (\bar{\boldsymbol{\xi}}_\gamma, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}) \sigma^2 \delta_{\alpha\gamma} V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}$$
$$= \frac{2\sigma^2}{N} \bar{W}_\beta^2 \bar{W}_\alpha (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}) V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}, \tag{62}$$

$$(E[\varDelta_1 \bar{W}_\beta \varDelta_1 \boldsymbol{\xi}_\alpha], \bar{\boldsymbol{\theta}}) = (\frac{2\sigma^2}{N} \bar{W}_\beta^2 \bar{W}_\alpha (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}) V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})$$
$$= \frac{2\sigma^2}{N} \bar{W}_\beta^2 \bar{W}_\alpha (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}})(\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})$$
$$= \frac{2\sigma^2}{N} \bar{W}_\beta^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}). \tag{63}$$

Hence, the expectation of $\boldsymbol{t}_3$ is

$$E[\boldsymbol{t}_3] = \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \Big(\frac{2\sigma^2}{N} \bar{W}_\beta^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}})\Big)(\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha)\bar{\boldsymbol{\xi}}_\beta$$
$$= \frac{2\sigma^2}{N^3}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta^2 (\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha)(\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\beta$$
$$= \frac{2\sigma^2}{N^3}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha \bar{W}_\beta^2 \bar{\boldsymbol{\xi}}_\beta^\top \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}\bar{\boldsymbol{\xi}}_\beta$$
$$= \frac{2\sigma^2}{N^2}\sum_{\beta=1}^{N} \bar{W}_\beta^2 \bar{\boldsymbol{\xi}}_\beta^\top \bar{\boldsymbol{M}}^- \Big(\frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top\Big) \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}\bar{\boldsymbol{\xi}}_\beta$$
$$= \frac{2\sigma^2}{N^2}\sum_{\beta=1}^{N} \bar{W}_\beta^2 \bar{\boldsymbol{\xi}}_\beta^\top \bar{\boldsymbol{M}}^- \bar{\boldsymbol{M}} \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}\bar{\boldsymbol{\xi}}_\beta = \frac{2\sigma^2}{N^2}\sum_{\beta=1}^{N} \bar{W}_\beta^2 \bar{\boldsymbol{\xi}}_\beta^\top \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}}\bar{\boldsymbol{\xi}}_\beta$$
$$= \frac{2\sigma^2}{N^2}\sum_{\beta=1}^{N} \bar{W}_\beta^2 (\bar{\boldsymbol{\xi}}_\beta, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\beta]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\beta. \tag{64}$$

From the above results, we can write the expectation of $\varDelta_1 \boldsymbol{M}\bar{\boldsymbol{M}}^- \varDelta_1 \boldsymbol{M}\bar{\boldsymbol{\theta}}$ as follows:

$$
E[\varDelta_1 \boldsymbol{M}\bar{\boldsymbol{M}}^- \varDelta_1 \boldsymbol{M}\bar{\boldsymbol{\theta}}] = \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha) V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}}
$$

$$
+\frac{3\sigma^2}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha. \tag{65}
$$

B.3 Hyper-renormalization

From the above results, the expectation of $\boldsymbol{T}\bar{\boldsymbol{\theta}}$ is

$$
\begin{aligned}
&E[\boldsymbol{T}\bar{\boldsymbol{\theta}}] \\
&= \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} + (\boldsymbol{e}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha \Big) - \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha) V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} \\
&\quad -\frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{\xi}}_\alpha \\
&= \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} + \bar{\boldsymbol{\xi}}_\alpha \boldsymbol{e}_\alpha^\top \bar{\boldsymbol{\theta}} \Big) - \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha) V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} \\
&\quad -\frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} \\
&= \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} + (\bar{\boldsymbol{\xi}}_\alpha \boldsymbol{e}_\alpha^\top + \boldsymbol{e}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top)\bar{\boldsymbol{\theta}} \Big) - \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha) V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}} \\
&\quad -\frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 \Big(\bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top \bar{\boldsymbol{M}}^- V_0[\boldsymbol{\xi}_\alpha] + V_0[\boldsymbol{\xi}_\alpha] \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top \Big)\bar{\boldsymbol{\theta}},
\end{aligned} \tag{66}
$$

where $\boldsymbol{e}_\alpha$ is defined via Eq. (39). Note that since the matrix $\bar{\boldsymbol{N}}$ in the generalized eigenvalue problem should be symmetric, we made use of the identity $(\bar{\boldsymbol{\xi}}, \boldsymbol{\theta}) = 0$ and added extra terms that are zero so that a symmetric matrix results. Thus, with the matrix $\bar{\boldsymbol{N}}$ defined by Eq. (38), the above expression is written as $E[\boldsymbol{T}\bar{\boldsymbol{\theta}}] = \sigma^2 \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}$. It follows that if $\boldsymbol{N}$ is defined by Eq. (41), its noiseless value $\bar{\boldsymbol{N}}$ satisfies $E[\boldsymbol{T}\bar{\boldsymbol{\theta}}] = \sigma^2 \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}$.

C Evaluation of the Covariance Matrix of the Solution

Iterative reweight, renormalization, and hyper-renormalization all solve the generalized eigenvalue problem $\boldsymbol{M}\boldsymbol{\theta} = \lambda\boldsymbol{N}\boldsymbol{\theta}$, where $\boldsymbol{N}$ is the unit matrix $\boldsymbol{I}$ for iterative reweight, the matrix in Eq. (18) for renormalization, and the matrix in Eq. (41) for hyper-renormalization, while $\boldsymbol{M}$ is common (Eqs. (11), (18), and (40)). If we closely examine the analysis in Sec. 5, we obtain Eq. (30) irrespective of the definition of $\boldsymbol{N}$. Hence, the covariance of the solution $\boldsymbol{\theta}$ is written up to $O(\sigma^4)$ as follows:

$$
\begin{aligned}
V[\boldsymbol{\theta}] &= E[\Delta\boldsymbol{\theta}\Delta\boldsymbol{\theta}^\top] \\
&= E[\Big(\frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha(\Delta\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha\Big)\Big(\frac{1}{N}\sum_{\beta=1}^{N} \bar{W}_\beta(\Delta\boldsymbol{\xi}_\beta, \bar{\boldsymbol{\theta}})\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\beta\Big)^\top] \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha\bar{W}_\beta E[(\Delta\boldsymbol{\xi}_\alpha, \bar{\boldsymbol{\theta}})(\Delta\boldsymbol{\xi}_\beta, \bar{\boldsymbol{\theta}})]\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top\bar{\boldsymbol{M}}^- \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha\bar{W}_\beta (\bar{\boldsymbol{\theta}}, E[\Delta\boldsymbol{\xi}_\alpha\Delta\boldsymbol{\xi}_\beta^\top]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top\bar{\boldsymbol{M}}^- \\
&= \frac{1}{N^2}\sum_{\alpha,\beta=1}^{N} \bar{W}_\alpha\bar{W}_\beta (\bar{\boldsymbol{\theta}}, \sigma^2\delta_{\alpha\beta}V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\beta^\top\bar{\boldsymbol{M}}^- \\
&= \frac{\sigma^2}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 (\bar{\boldsymbol{\theta}}, V_0[\boldsymbol{\xi}_\alpha]\bar{\boldsymbol{\theta}})\bar{\boldsymbol{M}}^-\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\bar{\boldsymbol{M}}^- = \frac{\sigma^2}{N}\bar{\boldsymbol{M}}^-\Big(\frac{1}{N}\sum_{\alpha=1}^{N}\bar{W}_\alpha\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top\Big)\bar{\boldsymbol{M}}^- \\
&= \frac{\sigma^2}{N}\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\bar{\boldsymbol{M}}^- = \frac{\sigma^2}{N}\bar{\boldsymbol{M}}^-,
\end{aligned}
\tag{67}
$$

where we have noted that error in $\boldsymbol{\xi}_\alpha$ is independent for different α and hence $E[\Delta_1\boldsymbol{\xi}_\alpha\Delta_1\boldsymbol{\xi}_\beta^\top] = \delta_{\alpha\beta}V_0[\boldsymbol{\xi}_\alpha]$, $\delta_{\alpha\beta}$ being the Kronecker delta. We have also used the identity for pseudoinverse: $\bar{\boldsymbol{M}}^-\bar{\boldsymbol{M}}\bar{\boldsymbol{M}}^- = \bar{\boldsymbol{M}}^-$.

from the above results, we see that the leading term of the covariance matrix of the solution is the same for iterative reweight, renormalization, and hyper-renormalization, independent of $\boldsymbol{N}$. The last term of Eq. (67) coincides with the accuracy limit called the *KCR lower bound* [1, 7, 8], meaning that the solution of iterative reweight, renormalization, and hyper-renormalization all achieve the KCR lower bound in the leading order. The trace of Eq. (67) is $\mathrm{tr}V[\boldsymbol{\theta}] = \mathrm{tr}E[\Delta\boldsymbol{\theta}\Delta\boldsymbol{\theta}^\top] = E[\|\Delta\boldsymbol{\theta}\|^2]$, whose square root gives the bound ton the RMS error in Eq. (45).

D Fundamental Matrix Computation Experiment

Figure 5 shows two synthetic images of a cylindrical grid placed in the scene. The image size is assumed to be 600×600 pixels with focal length 600 pixels for both. We added Gaussian noise of mean 0 and standard deviation σ pixels the x and y coordinates of each grid point independently and computed the fundamental matrix. The fundamental matrix $\boldsymbol{F}$ is constrained to $\det \boldsymbol{F} = 0$ [5], and there exist three approaches to enforce it: 1) A posteriori correction. The fundamental matrix is first computed without considering the constraint and is modified a posteriori so as to satisfy it in an optimal manner. 2) Internal access. The fundamental matrix is parameterized so that the constraint is identically satisfied and is optimized in the ("internal") parameter space. 3) External access. We do iterations in the ("external") 9-D space of the fundamental matrix in such a way that an optimal solution that satisfies the constraint automatically results. Here, we adopt the a posteriori correction approach and compare the accuracy of computation before that correction, using the following methods: 1. LS, 2. iterative reweight, 3. the Taubin method, 4. renormalization, 5. HyperLS, 6. hyper-renormalization, 7. ML, 8. ML with hyperaccurate correction.

Then, we computed the bias B and the RMS error D defined by Eq. (44), where $\boldsymbol{\theta}$ is the 9-D vector as defined in Eq. (6). The KCR lower bound is given by Eq. (45), where the matrix $\bar{\boldsymbol{M}}$ has rank 8 for the fundamental matrix computation. The computed bias and the RMS error are plotted in Fig. 2 together with the KCR lower bound. As in the case of ellipse fitting, LS and iterative reweight have large bias, resulting in large RMS error. As we see in Fig. 6(a), ML has considerable bias, but it is mostly removed by hyperaccurate correction, resulting in nearly the same bias as hyper-renormalization. As we see in Fig. 2(b), the RMS error of all methods except LS and iterative reweight is already very close to the KCR lower bound, so the effect of bias reduction is very small. A close examination reveals, however, that the highest accuracy is exhibited by hyper-renormalization and ML with hyperaccurate correction, just as in the case of ellipse fitting.

References

[17] K. Kanatani and Y. Sugaya, Compact fundamental matrix computation, *IPSJ Tran. Comput. Vis. Appl.*, **2** (2010-3), 59–70.

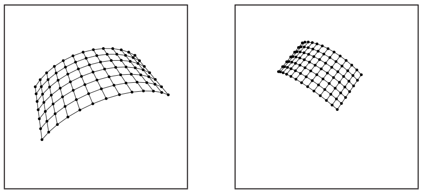


Fig. 5. Two synthetic images of a cylindrical grid.

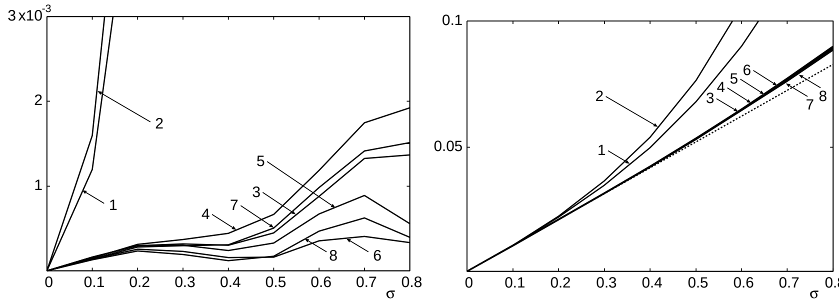


Fig. 6. The bias (a) and the RMS error (b) of the computed fundamental matrix for the standard deviation σ of the added noise for 10000 independent trials. 1. LS, 2. iterative reweight, 3. the Taubin method, 4. renormalization, 5. HyperLS, 6. hyper-renormalization, 7. ML, 8. ML with hyperaccurate correction. The dotted line in (c) indicates the KCR lower bound.