Rationalising the Renormalisation Method of Kanatani

Wojciech Chojnacki, Michael J. Brooks, Anton van den Hengel

Department of Computer Science, University of Adelaide
Adelaide, SA 5005, Australia

{wojtek,mjb,hengel}@cs.adelaide.edu.au

Abstract

The renormalisation technique of Kanatani is intended to iteratively minimise a cost function of a certain form while avoiding systematic bias inherent in the common method of minimisation due to Sampson. Within the computer vision community, the technique has generally proven difficult to absorb. This work presents an alternative derivation of the technique, and places it in the context of other approaches. We first show that the minimiser of the cost function must satisfy a special variational equation. A Newton-like, fundamental numerical scheme is presented with the property that its theoretical limit coincides with the minimiser. Standard statistical techniques are then employed to derive afresh several renormalisation schemes. The fundamental scheme proves pivotal in the rationalising of the renormalisation and other schemes, and enables us to show that the renormalisation schemes do not have as their theoretical limit the desired minimiser. The various minimisation schemes are finally subjected to a comparative performance analysis under controlled conditions.

Keywords: Statistical methods, surface fitting, covariance matrix, maximum likelihood, renormalisation, conic fititng, fundamental matrix estimation

1 Introduction

Many problems in computer vision are readily formulated as the need to minimise a cost function with respect to some unknown parameters. Such a cost function will often involve (known) covariance matrices characterising uncertainty of the data and will take the form of a sum of quotients of quadratic forms in the parameters. Finding the values of the parameters that minimise such a cost function is often difficult.

One approach to minimising a cost function represented as a sum of fractional expressions is attributed to Sampson. Here, an initial estimate is substituted into the denominators of the cost function, and a minimiser is sought for the now scalar-weighted numerators. This procedure is then repeated using the newly obtained estimate until convergence is obtained. It emerges that this approach is biased. Noting this, Kenichi Kanatani developed a renormalisation method whereby an attempt is made at each iteration to undo the biasing effects. Many examples may be found in the literature of problems benefiting from this approach.

In this work, we carefully analyse the renormalisation concept, and place it in the context of other approaches. We first specify the general problem form and an associated cost function to which renormalisation is applicable. We then show that the cost function minimiser must satisfy a particular variational equation. Interestingly, we observe that the renormalisation estimate is not a theoretical minimiser of the cost function, and neither are estimates obtained via some other commonly used methods. This is in contrast to a fundamental numerical scheme that we present.

New derivations are given for Kanatani's first-order and second-order renormalisation schemes, and several variations on the theme are proposed. This serves as a rationalising of renormalisation, making recourse to various statistical concepts. Experiments are carried out on the benchmark problem of estimating ellipses from synthetic data points and their covariances. The renormalisation schemes are shown to perform better than more traditional methods in the face of data exhibiting noise that is anisotropic and inhomogeneous. None of the methods outperforms the relatively straightforward fundamental numerical scheme.

2 Problem Formulation

A wide class of computer vision problems may be couched in terms of an equation of the form

$$\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}) = 0. \tag{1}$$

Here $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_l]^T$ is a vector representing unknown parameters; $\boldsymbol{x} = [x_1, \ldots, x_k]^T$ is a vector representing an element of the data (for example, the locations of a pair of corresponding points); and $\boldsymbol{u}(\boldsymbol{x}) = [u_1(\boldsymbol{x}), \ldots, u_l(\boldsymbol{x})]^T$ is a vector with the data transformed in such a way that: (i) each component is a quadratic form in the compound vector $[\boldsymbol{x}^T, 1]^T$, (ii) one component of $\boldsymbol{u}(\boldsymbol{x})$ is equal to 1. An ancillary constraint may also apply that does not involve the data, and this can be expressed as

$$\psi(\boldsymbol{\theta}) = 0 \tag{2}$$

for some scalar-valued function ψ. The estimation problem can now be stated as follows: Given a collection $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ of image data, determine $\boldsymbol{\theta} \neq \mathbf{0}$ satisfying (2) such that (1) holds with $\boldsymbol{x}$ replaced by $\boldsymbol{x}_i$ for $1 \leq i \leq n$. When $n > l$ and noise is present, the corresponding system of equations is overdetermined and as such may fail to have a non-zero solution. In this situation, we are concerned with finding $\boldsymbol{\theta}$ that best fits the data in some sense. The form of this vision problem involving (known) covariance information was first studied in detail by Kanatani [12], and later by various others (see, e.g., [4, 13, 14, 20, 21]).

Conic fitting is one problem of this kind [2, 23]. Two other conformant problems are estimating the coefficients of the *epipolar equation* [6], and estimating the coefficients of the *differential epipolar equation* [3, 22]. Each of these problems involves an ancillary *cubic constraint*. The precise way in which these example problems accord with our problem form is described in a companion work [4].

3 Cost Functions and Estimators

A vast class of techniques for solving our problem rest upon the use of cost functions measuring the extent to which the data and candidate estimates fail to satisfy (1). If—for simplicity—one sets aside the ancillary constraint, then, given a cost function $J = J(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, a corresponding estimate $\widehat{\boldsymbol{\theta}}$ is defined by

$$J(\widehat{\boldsymbol{\theta}}) = \min_{\boldsymbol{\theta} \neq 0} J(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n). \tag{3}$$

Since (1) does not change if $\boldsymbol{\theta}$ is multiplied by a non-zero scalar, it is natural to demand that $\widehat{\boldsymbol{\theta}}$ should satisfy (3) together with all the $\lambda \widehat{\boldsymbol{\theta}}$, where λ is a non-zero scalar. This is guaranteed if J is $\boldsymbol{\theta}$*-homogeneous*:

$$J(\lambda\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = J(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \quad \text{for every non-zero scalar } \lambda.$$

Henceforth only $\boldsymbol{\theta}$-homogeneous cost functions will be considered. The assignment of $\widehat{\boldsymbol{\theta}}$ (uniquely defined up to a scalar factor) to $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ will be termed the J-based *estimator* of $\boldsymbol{\theta}$.

Once an estimate has been generated by minimising a specific cost function, the ancillary constraint (if it applies) can further be accommodated via an adjustment procedure. One possibility is to use a general scheme delivering an ‘optimal correction’ described in [12, Subsec. 9.5.2]. In what follows we shall confine our attention to the estimation phase that precedes adjustment.

3.1 Algebraic Least Squares Estimator

A straightforward estimator is derived from the cost function

$$J_{\mathrm{ALS}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = ||\boldsymbol{\theta}||^{-2} \sum_{i=1}^{n} \boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta},$$

where $\boldsymbol{A}_i = \boldsymbol{u}(\boldsymbol{x}_i)\boldsymbol{u}(\boldsymbol{x}_i)^T$ and $||\boldsymbol{\theta}|| = (\theta_1^2 + \cdots + \theta_l^2)^{1/2}$. Here each summand $\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta}$ is the square of the *algebraic distance* $|\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}_i)|$. Accordingly, the J_{ALS}-based estimate of $\boldsymbol{\theta}$ is termed the *algebraic least squares* (ALS) *estimate* and is denoted $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$. It is uniquely determined, up to a scalar factor, by an eigenvector of $\sum_{i=1}^{n} \boldsymbol{A}_i$ associated with the smallest eigenvalue [4].

3.2 Approximated Maximum Likelihood Estimator

The ALS estimator treats all data as being equally valuable. When information about the measurement errors is available, it is desirable that it be incorporated into the estimation process. Here we present an estimator capable of informed weighting. It is based on the principle of maximum likelihood and draws upon Kanatani's work on geometric fitting [12, Chap. 7].

The measurement errors being generally unknowable, we regard the collective data $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ as a sample value taken on by an aggregate of vector-valued random variables $(\mathsf{x}_1, \ldots, \mathsf{x}_n)$. We assume that the distribution of $(\mathsf{x}_1, \ldots, \mathsf{x}_n)$ is not exactly specified but is an element of a collection $\{P_{\boldsymbol{\eta}} \mid \boldsymbol{\eta} \in \boldsymbol{H}\}$ of candidate distributions, with $\boldsymbol{H}$ the set of all $(n+1)$-tuples $\boldsymbol{\eta} = (\boldsymbol{\theta}; \overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_n)$ such that $\boldsymbol{\theta} \neq \mathbf{0}$ and

$$\boldsymbol{\theta}^T \boldsymbol{u}(\overline{\boldsymbol{x}}_1) = \cdots = \boldsymbol{\theta}^T \boldsymbol{u}(\overline{\boldsymbol{x}}_n) = 0. \tag{4}$$

The candidate distributions are to be such that if a distribution $P_{\boldsymbol{\eta}}$ is in effect, then each x_i $(i = 1, \ldots, n)$ is a noise-driven, fluctuating quantity around $\overline{\boldsymbol{x}}_i$.

We assume that the data come equipped with a collection $(\boldsymbol{\Lambda}_{\boldsymbol{x}_1}, \ldots, \boldsymbol{\Lambda}_{\boldsymbol{x}_n})$ of positive definite $k \times k$ *covariance matrices*. These matrices constitute repositories of prior information about the uncertainty of the data. We put the $\boldsymbol{\Lambda}_{\boldsymbol{x}_i}$ in use by assuming that, for each $\boldsymbol{\eta} \in \boldsymbol{H}$, $P_{\boldsymbol{\eta}}$ is the unique distribution satisfying the following conditions:

- for any $i, j = 1, \ldots, n$ with $i \neq j$, the random vectors x_i and x_j (or equivalently, the noises behind x_i and x_j) are stochastically independent;

- for each $i, j = 1, \ldots, n$, the random vector x_i has multivariate normal distribution with mean value vector $\overline{\boldsymbol{x}}_i$ and covariance matrix $\boldsymbol{\Lambda}_{\boldsymbol{x}_i}$, that is:

$$\mathsf{E}\left[\mathsf{x}_i\right] = \overline{\boldsymbol{x}}_i, \ \mathsf{E}\left[(\mathsf{x}_i - \overline{\boldsymbol{x}}_i)(\mathsf{x}_i - \overline{\boldsymbol{x}}_i)^T\right] = \boldsymbol{\Lambda}_{\boldsymbol{x}_i}. \tag{5}$$

Each distribution $P_{\boldsymbol{\eta}}$ will readily be described in terms of a probability density function (PDF) $(\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n) \mapsto f(\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n \mid \boldsymbol{\eta})$. Resorting to the *principle of maximum likelihood*, we give the greatest confidence to that choice of $\boldsymbol{\eta}$ for which the *likelihood function* $\boldsymbol{\eta} \mapsto f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \mid \boldsymbol{\eta})$ attains a maximum. Using the explicit form of the PDF's involved, one can show that the *maximum likelihood estimate* is the parameter $\widehat{\boldsymbol{\eta}}_{\mathrm{ML}}$ at which the cost function

$$J_{\mathrm{ML}}(\boldsymbol{\eta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)^T \boldsymbol{\Lambda}_{\boldsymbol{x}_i}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)$$

attains a minimum [4, 12]. Each term in the above summation represents the squared *Mahalanobis distance* between $\boldsymbol{x}_i$ and $\overline{\boldsymbol{x}}_i$. Note that the value of $\widehat{\boldsymbol{\eta}}_{\mathrm{ML}}$ remains unchanged if the covariance matrices are multiplied by a common scalar.

The parameter $\boldsymbol{\eta} = (\boldsymbol{\theta}; \overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_n)$ naturally splits into two parts: $\pi_1(\boldsymbol{\eta}) = \boldsymbol{\theta}$ and $\pi_2(\boldsymbol{\eta}) = (\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_n)$. These parts encompass the *principal parameters* and *nuisance parameters*, respectively. We are mostly interested in the π_1-part of $\widehat{\boldsymbol{\eta}}_{\mathrm{ML}}$, which we call the maximum likelihood estimate of $\boldsymbol{\theta}$ and denote $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$. It turns out that $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ can be identified as the minimiser of a certain cost function which is directly derivable from J_{ML}. This cost function does not lend itself to explicit calculation. However, a tractable approximation [4] can be derived in the form of the function

$$J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}_i) \boldsymbol{u}(\boldsymbol{x}_i)^T \boldsymbol{\theta}}{\boldsymbol{\theta}^T \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}_i) \boldsymbol{\Lambda}_{\boldsymbol{x}_i} \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}_i)^T \boldsymbol{\theta}},$$

where $\partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{y}) = [(\partial u_i / \partial x_j)(\boldsymbol{y})]_{1 \leq i \leq l, 1 \leq j \leq k}$. If, for any k vector $\boldsymbol{y}$ and any $k \times k$ matrix $\boldsymbol{\Lambda}$, we let

$$\boldsymbol{B}(\boldsymbol{y}, \boldsymbol{\Lambda}) = \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{y}) \boldsymbol{\Lambda} \partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{y})^T, \tag{6}$$

and next, for each $i = 1, \ldots, n$, let $\boldsymbol{B}_i = \boldsymbol{B}(\boldsymbol{x}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})$, then J_{AML} can be simply written as

$$J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}}.$$

The J_{AML}-based estimate of $\boldsymbol{\theta}$ will be called the *approximated maximum likelihood* (AML) *estimate* and will be denoted $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$.

It should be observed that J_{AML} can be derived without recourse to principles of maximum likelihood by, for example, using a gradient weighted approach that also incorporates covariances. Various terms may therefore be used to describe methods that aim to minimise a cost function such as J_{AML}, although some of the terms may not be fully discriminating. Candidate labels include 'heteroscedastic regression' [13], 'weighted orthogonal regression' [1, 9], and 'gradient weighted least squares' [24].

3.3 Variational Equation

Since $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ is a minimiser of J_{AML}, we have that, when $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$, the following equation holds:

$$\partial_{\boldsymbol{\theta}} J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \mathbf{0}^T. \tag{7}$$

Here $\partial_{\boldsymbol{\theta}} J_{\mathrm{AML}}$ denotes the row vector of the partial derivatives of J_{AML} with respect to $\boldsymbol{\theta}$. We term this the *variational equation*. Direct computation shows that

$$[\partial_{\boldsymbol{\theta}} J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)]^T = 2 \boldsymbol{X}_{\boldsymbol{\theta}} \boldsymbol{\theta},$$

where $\boldsymbol{X}_{\boldsymbol{\theta}}$ is the symmetric matrix

$$\boldsymbol{X}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\boldsymbol{A}_i}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}} - \sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta}}{(\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta})^2} \boldsymbol{B}_i. \tag{8}$$

Thus (7) can be written as

$$\boldsymbol{X}_{\boldsymbol{\theta}} \boldsymbol{\theta} = \mathbf{0}. \tag{9}$$

This is a non-linear equation and is unlikely to admit solutions in closed form.

Obviously, not every solution of the variational equation is a point at which the global minimum of J_{AML} is attained. However, the solution set of the equation provides a severely restricted family of candidates for the global minimiser. Within this set, the minimiser is much easier to identify.

4 Numerical schemes

Closed-form solutions of the variational equation may be infeasible, so in practice $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ has to be found numerically. Throughout we shall assume that $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ lies close to $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$. This assumption is to increase the chances that any candidate minimiser obtained via a numerical method seeded with $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ coincides with $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$.

4.1 Fundamental Numerical Scheme

A vector $\boldsymbol{\theta}$ satisfies (9) if and only if it falls into the null space of the matrix $\boldsymbol{X}_{\boldsymbol{\theta}}$. Thus, if $\boldsymbol{\theta}_{k-1}$ is a tentative guess, then an improved guess can be obtained by picking a vector $\boldsymbol{\theta}_k$ from that eigenspace of $\boldsymbol{X}_{\boldsymbol{\theta}_{k-1}}$ which most closely approximates the null space of $\boldsymbol{X}_{\boldsymbol{\theta}}$; this eigenspace is, of course, the one corresponding to the eigenvalue closest to zero. It can be proved that as soon as the sequence of updates converges, the limit is a solution of (9) [4]. The *fundamental numerical scheme* implementing the above idea is presented in Figure (1). The algorithm can be regarded as a variant of the Newton-Raphson method.

1. Set $\boldsymbol{\theta}_0 = \widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$.
2. Assuming that $\boldsymbol{\theta}_{k-1}$ is known, compute the matrix $\boldsymbol{X}_{\boldsymbol{\theta}_{k-1}}$.
3. Compute a normalised eigenvector of $\boldsymbol{X}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the eigenvalue closest to zero and take this eigenvector for $\boldsymbol{\theta}_k$.
4. If $\boldsymbol{\theta}_k$ is sufficiently close to $\boldsymbol{\theta}_{k-1}$, then terminate the procedure; otherwise increment k and return to Step 2.

Figure 1: Fundamental numerical scheme.

4.2 Sampson's Scheme

Let

$$\boldsymbol{M}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\boldsymbol{A}_i}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}} \tag{10}$$

and let

$$J'_{\mathrm{AML}}(\boldsymbol{\theta}, \boldsymbol{\xi}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\xi}} \boldsymbol{\theta}$$

be the modification of $J_{\mathrm{AML}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ in which the variable $\boldsymbol{\theta}$ in the denominators of all the contributing fractions is "frozen" at the value $\boldsymbol{\xi}$. For simplicity, we abbreviate $J'_{\mathrm{AML}}(\boldsymbol{\theta}, \boldsymbol{\xi}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ to $J'_{\mathrm{AML}}(\boldsymbol{\theta}, \boldsymbol{\xi})$.

Sampson [18] was the first to propose a scheme aiming to minimise a function involving fractional expressions, such as J_{AML} (although his cost functions did not incorporate covariance matrices). Sampson's scheme (SMP) applied to J_{AML} takes $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ for an initial guess $\boldsymbol{\theta}_0$, and given $\boldsymbol{\theta}_{k-1}$ generates an update $\boldsymbol{\theta}_k$ by minimising the cost function $\boldsymbol{\theta} \mapsto J'_{\mathrm{AML}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{k-1})$. Assuming that the sequence $\{\boldsymbol{\theta}_k\}$ converges, the Sampson estimate is defined as $\widehat{\boldsymbol{\theta}}_{\mathrm{SMP}} = \lim_{k\to\infty} \boldsymbol{\theta}_k$. Note that each function $J'_{\mathrm{AML}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{k-1})$ is quadratic in $\boldsymbol{\theta}$. Finding a minimiser of such a function is straightforward. The minimiser $\boldsymbol{\theta}_k$ is an eigenvector of $\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the smallest eigenvalue; moreover, this eigenvalue is equal to $J'_{\mathrm{AML}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1})$, so

$$\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} \boldsymbol{\theta}_k = J'_{\mathrm{AML}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) \boldsymbol{\theta}_k. \tag{11}$$

Sampson's scheme is summarised in Figure (2).

1. Set $\boldsymbol{\theta}_0 = \widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$.
2. Assuming that $\boldsymbol{\theta}_{k-1}$ is known, compute the matrix $\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}}$.
3. Compute a normalised eigenvector of $\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the smallest (non-negative) eigenvalue and take this eigenvector for $\boldsymbol{\theta}_k$.
4. If $\boldsymbol{\theta}_k$ is sufficiently close to $\boldsymbol{\theta}_{k-1}$, then terminate the procedure; otherwise increment k and return to Step 2.

Figure 2: Sampson's scheme.

A quick glance shows that this scheme differs from the fundamental numerical scheme only in that it uses matrices of the form $\boldsymbol{M}_{\boldsymbol{\theta}}$ instead of matrices of the form $\boldsymbol{X}_{\boldsymbol{\theta}}$. These two types of matrix are related by the formula $\boldsymbol{X}_{\boldsymbol{\theta}} = \boldsymbol{M}_{\boldsymbol{\theta}} - \boldsymbol{E}_{\boldsymbol{\theta}}$, where $\boldsymbol{E}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} [(\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta})/(\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta})^2] \boldsymbol{B}_i$. Letting $k \to \infty$ in (11) and taking into account the equality $J'_{\mathrm{AML}}(\boldsymbol{\theta}, \boldsymbol{\theta}) = J_{\mathrm{AML}}(\boldsymbol{\theta})$, we see that $\widehat{\boldsymbol{\theta}}_{\mathrm{SMP}}$ satisfies

$$[\boldsymbol{M}_{\boldsymbol{\theta}} - J_{\mathrm{AML}}(\boldsymbol{\theta}) \boldsymbol{I}_l] \boldsymbol{\theta} = \mathbf{0}, \tag{12}$$

where $\boldsymbol{I}_l$ is the $l \times l$ identity matrix. We call this the *Sampson equation*. Note that it is different from the variational equation (9) and that, as a result, $\widehat{\boldsymbol{\theta}}_{\mathrm{SMP}}$ is not a genuine minimiser of J_{AML}.

5 Renormalisation

The matrices $\boldsymbol{M}_{\boldsymbol{\theta}} - \boldsymbol{E}_{\boldsymbol{\theta}}$ and $\boldsymbol{M}_{\boldsymbol{\theta}} - J_{\mathrm{AML}}(\boldsymbol{\theta}) \boldsymbol{I}_l$ underlying the variational equation (9) and the Sampson equation (12) can be viewed as modified or normalised forms of $\boldsymbol{M}_{\boldsymbol{\theta}}$. As first realised by Kanatani [12], a different type of modification can be proposed based on statistical considerations. The requirement is that the modified or *renormalised* $\boldsymbol{M}_{\boldsymbol{\theta}}$ be *unbiased* in some sense. Using the renormalised $\boldsymbol{M}_{\boldsymbol{\theta}}$, one can formulate an equation analogous to both the variational and Sampson equations. This equation can in turn be used to define an estimate of $\boldsymbol{\theta}$.

Here we rationalise the unbiasing procedure under the condition that noise in an appropriate statistical model is small. In the next section, various schemes will be presented for numerically computing the parameter estimate defined in this procedure. A later section will be devoted to the derivation of an unbiasing procedure appropriate for noise that is not necessarily small, and to the development of schemes for numerically computing the parameter estimate defined in this more general procedure.

5.1 Unbiasing M_θ

Regard the given data $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ as a value taken on by the random vectors $(\mathsf{x}_1, \ldots, \mathsf{x}_n)$ introduced earlier. Suppose that $(\mathsf{x}_1, \ldots, \mathsf{x}_n)$ has distribution $P_{\boldsymbol{\eta}}$ for some $\boldsymbol{\eta} = (\boldsymbol{\theta}; \overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_n)$. Form the following random version of $\boldsymbol{M}_{\boldsymbol{\theta}}$

$$\mathsf{M}_{\boldsymbol{\eta}} = \sum_{i=1}^{n} \frac{1}{\boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}} \boldsymbol{A}(\mathsf{x}_i)$$

with ‘true’ value

$$\overline{\mathsf{M}}_{\boldsymbol{\eta}} = \sum_{i=1}^{n} \frac{1}{\boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}} \boldsymbol{A}(\overline{\boldsymbol{x}}_i).$$

In view of (4), $\boldsymbol{A}(\overline{\boldsymbol{x}}_i)\boldsymbol{\theta} = 0$, for each $i = 1, \ldots, n$, so $\overline{\mathsf{M}}_{\boldsymbol{\eta}}\boldsymbol{\theta} = 0$ and further $\boldsymbol{\theta}^T \overline{\mathsf{M}}_{\boldsymbol{\eta}}\boldsymbol{\theta} = 0$. On the other hand, since each rank-one matrix $\boldsymbol{A}(\mathsf{x}_i)$ is non-negative definite, and since also each $\boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})$ is non-negative definite[1], $\mathsf{M}_{\boldsymbol{\eta}}$ is non-negative definite. As the $\boldsymbol{A}(\mathsf{x}_i)$ are independent, $\mathsf{M}_{\boldsymbol{\eta}}$ is generically positive definite, with $\mathsf{E}\left[\boldsymbol{\theta}^T \mathsf{M}_{\boldsymbol{\eta}}\boldsymbol{\theta}\right] > 0$. Thus on average $\boldsymbol{\theta}^T \mathsf{M}_{\boldsymbol{\eta}}\boldsymbol{\theta}$ does not attain its ‘true’ value of zero, and as such is biased. The bias can be removed by forming the matrix

$$\sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}(\mathsf{x}_i)\boldsymbol{\theta} - \mathsf{E}\left[\boldsymbol{\theta}^T \boldsymbol{A}(\mathsf{x}_i)\boldsymbol{\theta}\right]}{\boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}}.$$

The terms $\mathsf{E}\left[\boldsymbol{\theta}^T \boldsymbol{A}(\mathsf{x}_i)\boldsymbol{\theta}\right]$ can be calculated explicitly. There is a matrix-valued function $(\boldsymbol{x}, \boldsymbol{\Lambda}) \mapsto \boldsymbol{D}(\boldsymbol{x}, \boldsymbol{\Lambda})$, to be specified later, such that, for each $i = 1, \ldots, n$,

$$\mathsf{E}\left[\boldsymbol{\theta}^T \boldsymbol{A}(\mathsf{x}_i)\boldsymbol{\theta}\right] = \boldsymbol{\theta}^T \boldsymbol{D}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}. \tag{13}$$

[1]As $\boldsymbol{u}(\boldsymbol{x}) = [1, \boldsymbol{u}'(\boldsymbol{x})^T]^T$, we have $\partial_{\boldsymbol{x}} \boldsymbol{u}(\boldsymbol{x}) = [0, \partial_{\boldsymbol{x}} \boldsymbol{u}'(\boldsymbol{x})^T]^T$ and further

$$\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\Lambda}) = \begin{bmatrix} 0 & 0^T \\ 0 & \boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\Lambda})' \end{bmatrix}.$$

Thus $\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\Lambda})$ is not positive definite, but merely non-negative definite.

The unbiased $\mathsf{M}_{\boldsymbol{\eta}}$ can be written as

$$\mathsf{Y}_{\boldsymbol{\eta}} = \sum_{i=1}^{n} \frac{\boldsymbol{A}(\mathsf{x}_i) - \boldsymbol{D}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})}{\boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i)\boldsymbol{\theta}}. \tag{14}$$

The random matrix $\mathsf{Y}_{\boldsymbol{\eta}}$ is a raw model for obtaining a fully deterministic modification of $\boldsymbol{M}_{\boldsymbol{\theta}}$. For each $i = 1, \ldots, n$, let $\boldsymbol{D}_i = \boldsymbol{D}(\boldsymbol{x}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})$. Guided by (14), we take

$$\boldsymbol{Y}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\boldsymbol{A}_i - \boldsymbol{D}_i}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}} \tag{15}$$

for a modified $\boldsymbol{M}_{\boldsymbol{\theta}}$. Somewhat surprisingly, this choice turns out not to be satisfactory. The problem is that while the $\boldsymbol{A}_i$ do not change when the $\boldsymbol{\Lambda}_{\boldsymbol{x}_i}$ are multiplied by a common scalar, the $\boldsymbol{D}_i$ do change. A properly designed algorithm employing a modified $\boldsymbol{M}_{\boldsymbol{\theta}}$ should give as outcomes values that remain intact when all the $\boldsymbol{\Lambda}_{\boldsymbol{x}_i}$ are multiplied by a common scalar. This is especially important if we aim not only to estimate the parameter, but also to evaluate the goodness of fit. Therefore further change to the numerators of the fractions forming $\boldsymbol{Y}_{\boldsymbol{\theta}}$ is necessary.

The dependence of $\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{\Lambda})$ on $\boldsymbol{\Lambda}$ is fairly complex. To gain an idea of what needs to be changed, it is instructive to consider a simplified form of $\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{\Lambda})$. A first-order (in some sense) approximation to $\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{\Lambda})$ is, as will be shown shortly, the matrix $\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\Lambda})$ defined in (6). The dependence of $\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\Lambda})$ on $\boldsymbol{\Lambda}$ is simple: if $\boldsymbol{\Lambda}$ is multiplied by a scalar, then $\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\Lambda})$ is multiplied by the same scalar. This suggests we introduce a *compensating factor* $J_{\text{com}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, or $J_{\text{com}}(\boldsymbol{\theta})$ in short, with the property that if the $\boldsymbol{\Lambda}_{\boldsymbol{x}_i}$ are multiplied by a scalar, then $J_{\text{com}}(\boldsymbol{\theta})$ is multiplied by the inverse of this scalar. With the help of $J_{\text{com}}(\boldsymbol{\theta})$, we can form, for each $i = 1, \ldots, n$, a *renormalised* numerator $\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta} - J_{\text{com}}(\boldsymbol{\theta})\, \boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}$ and can next set

$$\boldsymbol{Y}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\boldsymbol{A}_i - J_{\text{com}}(\boldsymbol{\theta}) \boldsymbol{B}_i}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}} = \boldsymbol{M}_{\boldsymbol{\theta}} - J_{\text{com}}(\boldsymbol{\theta}) \boldsymbol{N}_{\boldsymbol{\theta}}, \tag{16}$$

where $\boldsymbol{M}_{\boldsymbol{\theta}}$ is given in (10) and $\boldsymbol{N}_{\boldsymbol{\theta}}$ is defined by

$$\boldsymbol{N}_{\boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{1}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}} \boldsymbol{B}_i. \tag{17}$$

The numerators in (16) are clearly scale invariant. Note that $J_{\text{com}}(\boldsymbol{\theta})$ plays a role similar to that played by the factors $(\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta})/(\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta})$ in the formula for $\boldsymbol{X}_{\boldsymbol{\theta}}$ given in (8). Indeed, if the $\boldsymbol{\Lambda}_{\boldsymbol{x}_i}$ are multiplied by λ, then so are the $\boldsymbol{B}_i$, and consequently the $(\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta})/(\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta})$ are multiplied by λ^{-1}. The main difference between $J_{\text{com}}(\boldsymbol{\theta})$ and the $(\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta})/(\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta})$ is that these latter fractions change with the index i, while $J_{\text{com}}(\boldsymbol{\theta})$ is common for all the numerators involved.

To find a proper expression for $J_{\text{com}}(\boldsymbol{\theta})$, we take a look at $\boldsymbol{X}_{\boldsymbol{\theta}}$ for inspiration. Note that, on account of (8), $\boldsymbol{\theta}^T \boldsymbol{X}_{\boldsymbol{\theta}} \boldsymbol{\theta} = 0$. By analogy, we demand that $\boldsymbol{\theta}^T \boldsymbol{Y}_{\boldsymbol{\theta}} \boldsymbol{\theta} = 0$. This equation together with (16) implies that

$$J_{\text{com}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{\boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{N}_{\boldsymbol{\theta}} \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}}. \tag{18}$$

It is obvious that $J_{\rm com}(\boldsymbol{\theta})$ thus defined has the property required of a compensating factor. Moreover, this form of $J_{\rm com}(\boldsymbol{\theta})$ is in accordance with the unbiasedness paradigm. Indeed, if we form the random version of $J_{\rm com}$

$$\mathsf{J}_{\rm com}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}(\mathbf{x}_i)\boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}},$$

then, insofar as $\mathsf{E}\left[\boldsymbol{\theta}^T \boldsymbol{A}(\mathbf{x}_i)\boldsymbol{\theta}\right] = \boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}$, we have, abbreviating $\mathsf{J}_{\rm com}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n)$ to $\mathsf{J}_{\rm com}(\boldsymbol{\theta})$,

$$\mathsf{E}\left[\mathsf{J}_{\rm com}(\boldsymbol{\theta})\right] = 1 \tag{19}$$

and further

$$\mathsf{E}\left[\sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}(\mathbf{x}_i)\boldsymbol{\theta} - \mathsf{J}_{\rm com}(\boldsymbol{\theta})\boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}}\right] = 0.$$

We see that $\mathbf{Y}_{\boldsymbol{\eta}}$ given by

$$\mathbf{Y}_{\boldsymbol{\eta}} = \sum_{i=1}^{n} \frac{\boldsymbol{A}(\mathbf{x}_i) - \mathsf{J}_{\rm com}(\boldsymbol{\theta})\boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})}{\boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}_i, \boldsymbol{\Lambda}_{\boldsymbol{x}_i})\boldsymbol{\theta}}$$

is unbiased, which justifies the design of $\boldsymbol{Y}_{\boldsymbol{\theta}}$.

Since, in view of (19), $J_{\rm com}(\boldsymbol{\theta})$ is equal to 1 in the mean, the difference between

$$\sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta} - J_{\rm com}(\boldsymbol{\theta})\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}}$$

and

$$\sum_{i=1}^{n} \frac{\boldsymbol{\theta}^T \boldsymbol{A}_i \boldsymbol{\theta} - \boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{B}_i \boldsymbol{\theta}}$$

is blurred on average. Thus the refined renormalisation based on (16) is close in spirit to our original normalisation based on (15).

5.2 Renormalisation Equation

The *renormalisation equation*

$$\boldsymbol{Y}_{\boldsymbol{\theta}} \boldsymbol{\theta} = \mathbf{0}. \tag{20}$$

is an analogue of the variational and Sampson equations alike. It is not naturally derived from any specific cost function, and, as a result, it is not clear whether it has any solution at all. A general belief is that in the close vicinity of $\widehat{\boldsymbol{\theta}}_{\rm ALS}$ there is a solution and only one. This solution is termed the *renormalisation estimate* and is denoted $\widehat{\boldsymbol{\theta}}_{\rm REN}$. Since the renormalisation equation is different from the variational and

Sampson equations, $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$ is distinct both from $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ and $\widehat{\boldsymbol{\theta}}_{\mathrm{SMP}}$. It should be stressed that the difference between $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$ and $\widehat{\boldsymbol{\theta}}_{\mathrm{AML}}$ may be unimportant as both these estimates can be regarded as first-order approximations to $\widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$ and hence are likely to be statistically equivalent.

In practice, $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$ is represented as the limit of a sequence of successive approximations to what $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$ should be. If, under favourable conditions, the sequence is convergent, then the limit is a genuine solution of (20). Various sequences can be taken to calculate $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$ in this way. The simplest choice results from mimicking the fundamental numerical scheme as follows. Take $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ to be an initial guess $\boldsymbol{\theta}_0$. Suppose that an update $\boldsymbol{\theta}_{k-1}$ has already been generated. Form $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}}$, compute an eigenvector of $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the eigenvalue closest to zero, and take this eigenvector for $\boldsymbol{\theta}_k$. If the sequence $\{\boldsymbol{\theta}_k\}$ converges, take the limit for $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$. As we shall see shortly, $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$ thus defined automatically satisfies (20). This recipe for calculating $\widehat{\boldsymbol{\theta}}_{\mathrm{REN}}$ constitutes what we term the *renormalisation algorithm*.

6 First-Order Renormalisation Schemes

First-order renormalisation is based on the formula

$$\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{\Lambda}) = \boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\Lambda}), \tag{21}$$

as already pointed out in Subsection 5.1. To justify this formula, we retain the sequence of independent random vectors $(\mathsf{x}_1, \ldots, \mathsf{x}_n)$ as a model for our data $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. We assume that $(\mathsf{x}_1, \ldots, \mathsf{x}_n)$ has distribution $P_{\boldsymbol{\eta}}$ for some $\boldsymbol{\eta} = (\boldsymbol{\theta}; \overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_n)$. We also make a fundamental assumption to the effect that the noise driving each x_i is small. For simplicity, denote x_i by x, and contract $\boldsymbol{\Lambda}_{\boldsymbol{x}_i}$ to $\boldsymbol{\Lambda}$. Since the noise driving x is small, we can replace $\boldsymbol{u}(\mathsf{x})$ by the first-order sum in the Taylor expansion about $\overline{\boldsymbol{x}}$, $\boldsymbol{u}(\overline{\boldsymbol{x}}) + \partial_{\boldsymbol{x}} \boldsymbol{u}(\overline{\boldsymbol{x}})(\mathsf{x} - \overline{\boldsymbol{x}})$, and next, taking into account that $\boldsymbol{\theta}^T \boldsymbol{u}(\overline{\boldsymbol{x}}) = 0$, we can write

$$\boldsymbol{\theta}^T \boldsymbol{u}(\mathsf{x}) = \boldsymbol{\theta}^T \partial_{\boldsymbol{x}} \boldsymbol{u}(\overline{\boldsymbol{x}})(\mathsf{x} - \overline{\boldsymbol{x}}).$$

Hence

$$\boldsymbol{\theta}^T \boldsymbol{A}(\mathsf{x}) \boldsymbol{\theta} = \boldsymbol{\theta}^T \partial_{\boldsymbol{x}} \boldsymbol{u}(\overline{\boldsymbol{x}})(\mathsf{x} - \overline{\boldsymbol{x}})(\mathsf{x} - \overline{\boldsymbol{x}})^T \partial_{\boldsymbol{x}} \boldsymbol{u}(\overline{\boldsymbol{x}})^T \boldsymbol{\theta}$$

and further, in view of (5) and (6),

$$\begin{aligned} \mathsf{E}\left[\boldsymbol{\theta}^T \boldsymbol{A}(\mathsf{x}) \boldsymbol{\theta}\right] &= \boldsymbol{\theta}^T \partial_{\boldsymbol{x}} \boldsymbol{u}(\overline{\boldsymbol{x}}) \boldsymbol{\Lambda} \partial_{\boldsymbol{x}} \boldsymbol{u}(\overline{\boldsymbol{x}})^T \boldsymbol{\theta} \\ &= \boldsymbol{\theta}^T \boldsymbol{B}(\overline{\boldsymbol{x}}, \boldsymbol{\Lambda}) \boldsymbol{\theta}, \end{aligned}$$

which, on account of (13), establishes (21).

With the formula (21) validated, we can safely use $\boldsymbol{Y}_{\boldsymbol{\theta}}$ in the form given in (16) (with J_{com} given in (18)). The respective renormalisation estimate will be called the *first-order renormalisation estimate* and will be denoted $\widehat{\boldsymbol{\theta}}_{\mathrm{REN1}}$.

6.1 The FORI Scheme

By introducing an appropriate stopping rule, the renormalisation algorithm can readily be adapted to suit practical calculation. In the case of first-order renormalisation, the resulting method will be termed the *first-order renormalisation scheme, Version I*, or simply the FORI scheme. It is given in Figure (3).

1. Set $\boldsymbol{\theta}_0 = \widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$.
2. Assuming that $\boldsymbol{\theta}_{k-1}$ is known, compute the matrix $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}}$ using (16).
3. Compute a normalised eigenvector of $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the eigenvalue closest to zero and take this eigenvector for $\boldsymbol{\theta}_k$.
4. If $\boldsymbol{\theta}_k$ is sufficiently close to $\boldsymbol{\theta}_{k-1}$, then terminate the procedure; otherwise increment k and return to Step 2.

Figure 3: First-order renormalisation scheme, Version I.

6.2 The FORII Scheme

The FORI scheme can be slightly modified. The resulting *first-order renormalisation scheme, Version II*, or the FORII scheme, is effectively one of two schemes proposed by Kanatani [12, Chap. 12] (the other concerns second-order renormalisation).

Introduce a function

$$J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi}; \boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \frac{\boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\xi}} \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{N}_{\boldsymbol{\xi}} \boldsymbol{\theta}}.$$

Abbreviating $J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi}; \boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ to $J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi})$, let

$$\boldsymbol{Y}_{\boldsymbol{\theta},\boldsymbol{\xi}} = \sum_{i=1}^{n} \frac{\boldsymbol{A}_i - J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi}) \boldsymbol{B}_i}{\boldsymbol{\xi}^T \boldsymbol{B}_i \boldsymbol{\xi}} = \boldsymbol{M}_{\boldsymbol{\xi}} - J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi}) \boldsymbol{N}_{\boldsymbol{\xi}}.$$

It is immediately verified that

$$\boldsymbol{\theta}^T \boldsymbol{Y}_{\boldsymbol{\theta},\boldsymbol{\xi}} \boldsymbol{\theta} = 0 \tag{22}$$

for each $\boldsymbol{\theta}$ and each $\boldsymbol{\xi} \neq \boldsymbol{0}$. We also have that

$$J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\theta}) = J_{\mathrm{com}}(\boldsymbol{\theta}) \tag{23}$$

and

$$\boldsymbol{Y}_{\boldsymbol{\theta},\boldsymbol{\theta}} = \boldsymbol{Y}_{\boldsymbol{\theta}}. \tag{24}$$

As previously, take $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ to be an initial guess $\boldsymbol{\theta}_0$. Suppose that an update $\boldsymbol{\theta}_{k-1}$ has already been generated. Then

$$\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}} = \boldsymbol{Y}_{\boldsymbol{\theta}_{k-1},\boldsymbol{\theta}_{k-1}} = \boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - c_{k-1}\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}, \tag{25}$$

where $c_{k-1} = J'_{\mathrm{com}}(\boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k-1})$. Let $\boldsymbol{\theta}_k$ be a normalised eigenvector of $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the smallest eigenvalue λ_k. In view of (25), the update $\boldsymbol{Y}_{\boldsymbol{\theta}_k}$ is straightforwardly generated from the updates $\boldsymbol{M}_{\boldsymbol{\theta}_k}$, $\boldsymbol{N}_{\boldsymbol{\theta}_k}$, and c_k. It turns out that, under a certain approximation, c_k can be updated directly from c_{k-1} rather than by appealing to the above formula.

We have

$$\boldsymbol{\theta}_k^T(\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - c_{k-1}\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}})\boldsymbol{\theta}_k = \lambda_k.$$

Substituting $\boldsymbol{\theta}_k$ for $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{k-1}$ for $\boldsymbol{\xi}$ in (22), we obtain

$$\boldsymbol{\theta}_k^T(\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1})\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}})\boldsymbol{\theta}_k = 0.$$

The last two equations imply that

$$J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) = c_{k-1} + \frac{\lambda_k}{{\boldsymbol{\theta}_k}^T\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k}. \tag{26}$$

Now, assume that $J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) = J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_k)$; since both terms are close to $J_{\mathrm{com}}(\boldsymbol{\theta}_k)$, this is a realistic assumption. Under this assumption we have $c_k = J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_k)$, and equation (26) becomes

$$c_k = c_{k-1} + \frac{\lambda_k}{{\boldsymbol{\theta}_k}^T\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k}. \tag{27}$$

This is a formula for successive updating of the c_k. Defining consecutive $\boldsymbol{Y}_{\boldsymbol{\theta}_k}$ with the help of $\boldsymbol{M}_{\boldsymbol{\theta}_k}$, $\boldsymbol{N}_{\boldsymbol{\theta}_k}$ and c_k as in (25), and proceeding as in the FORI scheme, we obtain a sequence $\{\boldsymbol{\theta}_k\}$. If it converges, we take the corresponding limit for $\widehat{\boldsymbol{\theta}}_{\mathrm{REN1}}$. It can be shown that $\widehat{\boldsymbol{\theta}}_{\mathrm{REN1}}$ thus defined satisfies the renormalisation equation. The *first-order renormalisation scheme, Version II*, or the FORII scheme, based on the above algorithm is summarised in Figure (4).

6.3 The FORIII Scheme

With the help of the function J'_{com}, yet another defining sequence can be constructed. Take $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ for an initial guess $\boldsymbol{\theta}_0$. Suppose that an update $\boldsymbol{\theta}_{k-1}$ has already been generated. Define $\boldsymbol{\theta}_k$ to be the minimiser of the function $\boldsymbol{\theta} \mapsto J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{k-1})$:

$$J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) = \min_{\boldsymbol{\theta} \neq \mathbf{0}} J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{k-1}).$$

Since

$$[\partial_{\boldsymbol{\theta}} J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi})]^T = 2\boldsymbol{Z}_{\boldsymbol{\theta},\boldsymbol{\xi}}\boldsymbol{\theta},$$

1. Set $\boldsymbol{\theta}_0 = \widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ and $c_0 = 0$.
2. Assuming that $\boldsymbol{\theta}_{k-1}$ and c_{k-1} are known, compute the matrix $\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - c_{k-1}\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}$.
3. Compute a normalised eigenvector of $\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - c_{k-1}\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the smallest eigenvalue λ_k and take this eigenvector for $\boldsymbol{\theta}_k$. Then define c_k by (27).
4. If $\boldsymbol{\theta}_k$ is sufficiently close to $\boldsymbol{\theta}_{k-1}$, then terminate the procedure; otherwise increment k and return to Step 2.

Figure 4: First-order renormalisation scheme, Version II.

where

$$\boldsymbol{Z}_{\boldsymbol{\theta},\boldsymbol{\xi}} = \frac{\boldsymbol{M}_{\boldsymbol{\xi}}}{\boldsymbol{\theta}^T\boldsymbol{N}_{\boldsymbol{\xi}}\boldsymbol{\theta}} - \frac{\boldsymbol{\theta}^T\boldsymbol{M}_{\boldsymbol{\xi}}\boldsymbol{\theta}}{(\boldsymbol{\theta}^T\boldsymbol{N}_{\boldsymbol{\xi}}\boldsymbol{\theta})^2}\boldsymbol{N}_{\boldsymbol{\xi}} = \frac{\boldsymbol{M}_{\boldsymbol{\xi}} - J'_{\mathrm{com}}(\boldsymbol{\theta},\boldsymbol{\xi})\boldsymbol{N}_{\boldsymbol{\xi}}}{\boldsymbol{\theta}^T\boldsymbol{N}_{\boldsymbol{\xi}}\boldsymbol{\theta}} = \frac{\boldsymbol{Y}_{\boldsymbol{\theta},\boldsymbol{\xi}}}{\boldsymbol{\theta}^T\boldsymbol{N}_{\boldsymbol{\xi}}\boldsymbol{\theta}},$$

it follows that $\boldsymbol{\theta}_k$ satisfies

$$\boldsymbol{Y}_{\boldsymbol{\theta},\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta} = \mathbf{0}. \tag{28}$$

Assuming that the sequence $\{\boldsymbol{\theta}_k\}$ converges, let $\widehat{\boldsymbol{\theta}}_{\mathrm{REN1}} = \lim_{k\to\infty}\boldsymbol{\theta}_k$. Then, clearly, $\widehat{\boldsymbol{\theta}}_{\mathrm{REN1}}$ satisfies

$$\boldsymbol{Y}_{\boldsymbol{\theta},\boldsymbol{\theta}}\boldsymbol{\theta} = \mathbf{0},$$

which is an equation equivalent to (20). Note that in this method $\widehat{\boldsymbol{\theta}}_{\mathrm{REN1}}$ is defined as the limit of a sequence of minimisers of cost functions. As such the algorithm is similar to Sampson's algorithm, but the latter, of course, uses different cost functions.

The minimisers $\boldsymbol{\theta}_k$ can be directly calculated. To see this, rewrite (28) as

$$\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta} = J'_{\mathrm{com}}(\boldsymbol{\theta},\boldsymbol{\theta}_{k-1})\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}.$$

We see that $\boldsymbol{\theta}_k$ is an eigenvector and $J'_{\mathrm{com}}(\boldsymbol{\theta},\boldsymbol{\theta}_{k-1})$ is a corresponding eigenvalue of the linear pencil $\mathcal{P}_{k-1}\colon \mu \mapsto \mathcal{P}_{k-1}(\mu)$ defined by

$$\mathcal{P}_{k-1}(\mu) = \boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - \mu\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}} \qquad (\mu \text{ a real number}).$$

If $\boldsymbol{\xi}$ is any eigenvector of $\mathcal{P}_{k-1}$ with eigenvector λ

$$\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}}\boldsymbol{\xi} = \lambda\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}\boldsymbol{\xi},$$

then, necessarily, $J'_{\mathrm{com}}(\boldsymbol{\xi},\boldsymbol{\theta}_{k-1}) = \lambda$. Since $J'_{\mathrm{com}}(\boldsymbol{\theta}_k,\boldsymbol{\theta}_{k-1}) \leq J'_{\mathrm{com}}(\boldsymbol{\xi},\boldsymbol{\theta}_{k-1})$ we conclude that $J'_{\mathrm{com}}(\boldsymbol{\theta}_k,\boldsymbol{\theta}_{k-1}) \leq \lambda$. Thus $\boldsymbol{\theta}_k$ is an eigenvector of $\mathcal{P}_{k-1}$ corresponding

1. Set $\boldsymbol{\theta}_0 = \widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$.
2. Assuming that $\boldsymbol{\theta}_{k-1}$ is known, compute the matrices $\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}}$ and $\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}$.
3. Compute a normalised eigenvector of the eigenvalue problem

$$\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}}\boldsymbol{\xi} = \mu \boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}\boldsymbol{\xi}$$

corresponding to the smallest eigenvalue and take this eigenvector for $\boldsymbol{\theta}_k$.
4. If $\boldsymbol{\theta}_k$ is sufficiently close to $\boldsymbol{\theta}_{k-1}$, then terminate the procedure; otherwise increment k and return to Step 2.

Figure 5: First-order renormalisation scheme, Version III.

to the smallest eigenvalue. This observation leads to the *first-order renormalisation, Version III,* or the FORIII scheme, given in Figure 5. The matrices $\boldsymbol{N}_{\boldsymbol{\theta}_{k-1}}$ are singular, so the eigenvalue problem for $\mathcal{P}_{k-1}$ is degenerate. A way of reducing this problem to a non-degenerate one, based on the special form of the matrices $\boldsymbol{M}_{\boldsymbol{\theta}}$ and $\boldsymbol{N}_{\boldsymbol{\theta}}$, is presented in [4].

7 Second-Order Renormalisation

Second-order renormalisation rests on knowledge of the exact form of $\boldsymbol{D}(\boldsymbol{x})$. Here we first determine this form and next use it to evolve a second-order renormalisation estimate and various schemes for calculating it.

7.1 Calculating $\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{\Lambda})$

Determining the form of $\boldsymbol{D}(\boldsymbol{x}, \boldsymbol{\Lambda})$ is tedious but straightforward. We commence by introducing some notation.

Let $\boldsymbol{x} = [x^1, \ldots, x^k]$ be the vector of variables. Append to this vector a unital component, yielding

$$\boldsymbol{y} = [\boldsymbol{x}^T, 1]^T = [y_1, \ldots, y_{k+1}]^T.$$

Let $\boldsymbol{u}(\boldsymbol{x}) = [u_1(\boldsymbol{x}), \ldots, u_l(\boldsymbol{x})]^T$ be the vector of carriers. Given the special form of $\boldsymbol{u}(\boldsymbol{x})$ as described in Section 2, each $u_\gamma(\boldsymbol{x})$ $(\gamma = 1, \ldots, l)$ can be expressed as

$$u_\gamma(\boldsymbol{x}) = \boldsymbol{y}^T \boldsymbol{\mathcal{K}}_\gamma \boldsymbol{y}, \tag{29}$$

where $\boldsymbol{\mathcal{K}}_\gamma = [k_{\gamma,\alpha\beta}]$ is a symmetric $(k+1) \times (k+1)$ matrix. In what follows we adopt Einstein's convention according to which the summation sign for repeated indices is omitted. With this convention, equation (29) becomes

$$u_\gamma(\boldsymbol{x}) = k_{\gamma,\alpha\beta}\, y_\alpha y_\beta. \tag{30}$$

Let $\mathbf{x}$ be a random Gaussian k vector with mean $\overline{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{\Lambda} = [\sigma_{\alpha\beta}]$. Define a random vector $\mathbf{y} = [\mathbf{x}^T, 1]^T$. Clearly, $\mathbf{y}$ has

$$\overline{\boldsymbol{y}} = [\overline{\boldsymbol{x}}^T, 1]^T = [\overline{y}_1, \ldots, \overline{y}_{k+1}]^T$$

for the mean, and the $(k+1)\times(k+1)$ matrix $[s_{\alpha\beta}]$ defined by

$$s_{\alpha\beta} = \begin{cases} \sigma_{\alpha\beta} & \text{if } 1 \le \alpha, \beta \le k, \\ 0 & \text{otherwise} \end{cases}$$

for the covariance matrix. Let $\Delta\mathbf{y} = \mathbf{y} - \overline{\boldsymbol{y}}$. Suppose that $\boldsymbol{\theta}^T = [\theta_1, \ldots, \theta_l]$ satisfies

$$\boldsymbol{\theta}^T \boldsymbol{u}(\overline{\boldsymbol{x}}) = \theta_\gamma k_{\gamma,\alpha\beta}\, \overline{y}_\alpha \overline{y}_\beta = 0. \tag{31}$$

Since

$$\boldsymbol{\theta}^T \boldsymbol{A}(\mathbf{x})\boldsymbol{\theta} = \boldsymbol{\theta}^T \boldsymbol{u}(\mathbf{x})\boldsymbol{u}(\mathbf{x})^T \boldsymbol{\theta} = \theta_\gamma \theta_{\gamma'} k_{\gamma,\alpha\beta} k_{\gamma',\alpha'\beta'}\, \mathsf{y}_\alpha \mathsf{y}_\beta \mathsf{y}_{\alpha'} \mathsf{y}_{\beta'},$$

we have

$$\mathsf{E}\left[\boldsymbol{\theta}^T \boldsymbol{A}(\mathbf{x})\boldsymbol{\theta}\right] = \theta_\gamma \theta_{\gamma'} k_{\gamma,\alpha\beta} k_{\gamma',\alpha'\beta'}\, \mathsf{E}\left[\mathsf{y}_\alpha \mathsf{y}_\beta \mathsf{y}_{\alpha'} \mathsf{y}_{\beta'}\right].$$

Now

$$\begin{aligned} \mathsf{E}\left[\mathsf{y}_\alpha \mathsf{y}_\beta \mathsf{y}_{\alpha'} \mathsf{y}_{\beta'}\right] = {} & \overline{y}_\alpha \overline{y}_\beta \overline{y}_{\alpha'} \overline{y}_{\beta'} + \overline{y}_\alpha \overline{y}_\beta s_{\alpha'\beta'} + \overline{y}_\alpha \overline{y}_{\alpha'} s_{\beta'\beta'} + \overline{y}_\alpha \overline{y}_{\beta'} s_{\alpha'\beta} \\ & + \overline{y}_{\alpha'} \overline{y}_\beta s_{\alpha\beta'} + \overline{y}_\beta \overline{y}_{\beta'} s_{\alpha\alpha'} + \overline{y}_{\alpha'} \overline{y}_{\beta'} s_{\alpha\beta} \\ & + \mathsf{E}\left[\Delta\mathsf{y}_\alpha \Delta\mathsf{y}_{\alpha'} \Delta\mathsf{y}_\beta \Delta\mathsf{y}_{\beta'}\right]. \end{aligned}$$

By a standard result about moments of the multivariate normal distribution,

$$\mathsf{E}\left[\Delta\mathsf{y}_\alpha \Delta\mathsf{y}_\beta \Delta\mathsf{y}_{\alpha'} \Delta\mathsf{y}_{\beta'}\right] = s_{\alpha\beta} s_{\alpha'\beta'} + s_{\alpha\alpha'} s_{\beta\beta'} + s_{\alpha\beta'} s_{\alpha'\beta}.$$

In view of (31),

$$\theta_\gamma k_{\gamma,\alpha\beta}\, \overline{y}_\alpha \overline{y}_\beta = 0 \qquad \text{and} \qquad \theta_{\gamma'} k_{\gamma',\alpha'\beta'}\, \overline{y}_{\alpha'} \overline{y}_{\beta'} = 0,$$

and so

$$\begin{aligned} \theta_\gamma \theta_{\gamma'} k_{\gamma,\alpha\beta} k_{\gamma',\alpha'\beta'}\, \mathsf{E}\left[\mathsf{y}_\alpha \mathsf{y}_\beta \mathsf{y}_{\alpha'} \mathsf{y}_{\beta'}\right] = \theta_\gamma \theta_{\gamma'} k_{\gamma,\alpha\beta} k_{\gamma',\alpha'\beta'} \big(& \overline{y}_\alpha \overline{y}_{\alpha'} s_{\beta'\beta'} + \overline{y}_\alpha \overline{y}_{\beta'} s_{\alpha'\beta} \\ & + \overline{y}_{\alpha'} \overline{y}_\beta s_{\alpha\beta'} + \overline{y}_\beta \overline{y}_{\beta'} s_{\alpha\alpha'} \\ & + s_{\alpha\beta} s_{\alpha'\beta'} + s_{\alpha\alpha'} s_{\beta\beta'} + s_{\alpha\beta'} s_{\alpha'\beta} \big). \end{aligned}$$

Hence

$$\mathsf{E}\left[\boldsymbol{\theta}^T\boldsymbol{A}(\mathbf{x})\boldsymbol{\theta}\right] = \theta_\gamma\theta_{\gamma'}k_{\gamma,\alpha\beta}k_{\gamma',\alpha'\beta'}\left(\overline{y}_\alpha\overline{y}_{\alpha'}s_{\beta'\beta'} + \overline{y}_\alpha\overline{y}_{\beta'}s_{\alpha'\beta} + \overline{y}_{\alpha'}\overline{y}_\beta s_{\alpha\beta'} + \overline{y}_\beta\overline{y}_{\beta'}s_{\alpha\alpha'} + s_{\alpha\beta}s_{\alpha'\beta'} + s_{\alpha\alpha'}s_{\beta\beta'} + s_{\alpha\beta'}s_{\alpha'\beta}\right). \tag{32}$$

Let $\boldsymbol{D}_1(\overline{\boldsymbol{x}},\boldsymbol{\varLambda}) = [d_{1,\gamma\gamma'}]$ and $\boldsymbol{D}_2(\overline{\boldsymbol{x}},\boldsymbol{\varLambda}) = [d_{2,\gamma\gamma'}]$ be the $l\times l$ matrices defined by

$$\begin{aligned} d_{1,\gamma\gamma'} &= k_{\gamma,\alpha\beta}k_{\gamma',\alpha'\beta'}\left(\overline{y}_\alpha\overline{y}_{\alpha'}s_{\beta\beta'} + \overline{y}_\alpha\overline{y}_{\beta'}s_{\alpha'\beta} + \overline{y}_{\alpha'}\overline{y}_\beta s_{\alpha\beta'} + \overline{y}_\beta\overline{y}_{\beta'}s_{\alpha\alpha'}\right), \\ d_{2,\gamma\gamma'} &= k_{\gamma,\alpha\beta}k_{\gamma',\alpha'\beta'}\left(s_{\alpha\beta}s_{\alpha'\beta'} + s_{\alpha\alpha'}s_{\beta\beta'} + s_{\alpha\beta'}s_{\alpha'\beta}\right). \end{aligned} \tag{33}$$

With these matrices, (32) can be written as

$$\mathsf{E}\left[\boldsymbol{\theta}^T\boldsymbol{A}(\boldsymbol{x})\boldsymbol{\theta}\right] = \boldsymbol{\theta}^T(\boldsymbol{D}_1(\overline{\boldsymbol{x}},\boldsymbol{\varLambda}) + \boldsymbol{D}_2(\overline{\boldsymbol{x}},\boldsymbol{\varLambda}))\boldsymbol{\theta}.$$

Hence

$$\boldsymbol{D}(\boldsymbol{x},\boldsymbol{\varLambda}) = \boldsymbol{D}_1(\boldsymbol{x},\boldsymbol{\varLambda}) + \boldsymbol{D}_2(\boldsymbol{x},\boldsymbol{\varLambda}), \tag{34}$$

which is the desired formula.

7.2 Redefining $J_{\rm com}(\boldsymbol{\theta})$ and $\boldsymbol{Y}_{\boldsymbol{\theta}}$

We retain the framework of Subsection 5.1, but use the full expression for $\boldsymbol{D}(\boldsymbol{x},\boldsymbol{\varLambda})$ instead of the first-order approximation. We aim to modify, for each $i = 1,\ldots,n$, the numerator $\boldsymbol{\theta}^T\boldsymbol{A}_i\boldsymbol{\theta}$ into a term similar to $\boldsymbol{\theta}^T\boldsymbol{A}_i\boldsymbol{\theta} - \boldsymbol{\theta}^T\boldsymbol{D}_i\boldsymbol{\theta}$ (recall that $\boldsymbol{D}_i = \boldsymbol{D}(\boldsymbol{x}_i,\boldsymbol{\varLambda}_{\boldsymbol{x}_i})$), remembering the need for suitable compensation for scale change. The main problem now is that $\boldsymbol{D}(\boldsymbol{x},\boldsymbol{\varLambda})$ does not change equivariantly with $\boldsymbol{\varLambda}$. Under the scale change $\boldsymbol{\varLambda}\mapsto\lambda\boldsymbol{\varLambda}$, the two components of $\boldsymbol{D}(\boldsymbol{x},\boldsymbol{\varLambda})$ defined in (34) undergo two different transformations: $\boldsymbol{D}_1(\boldsymbol{x},\boldsymbol{\varLambda})\mapsto\lambda\boldsymbol{D}_1(\boldsymbol{x},\boldsymbol{\varLambda})$ and $\boldsymbol{D}_2(\boldsymbol{x},\boldsymbol{\varLambda})\mapsto\lambda^2\boldsymbol{D}_2(\boldsymbol{x},\boldsymbol{\varLambda})$. We adopt a solution as follows. We introduce a compensating factor $J_{\rm com}(\boldsymbol{\theta};\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$, or $J_{\rm com}(\boldsymbol{\theta})$ in short, with the property that if the $\boldsymbol{\varLambda}_{\boldsymbol{x}_i}$ are multiplied by λ, then $J_{\rm com}(\boldsymbol{\theta})$ is multiplied by λ^{-1}. We place this factor in front of $\boldsymbol{D}_{1,i} = \boldsymbol{D}_1(\boldsymbol{x}_i,\boldsymbol{\varLambda}_{\boldsymbol{x}_i})$, its square in front of $\boldsymbol{D}_{2,i} = \boldsymbol{D}_2(\boldsymbol{x}_i,\boldsymbol{\varLambda}_{\boldsymbol{x}_i})$, and form a modified numerator as follows:

$$\boldsymbol{\theta}^T\boldsymbol{A}_i\boldsymbol{\theta} - J_{\rm com}(\boldsymbol{\theta})\,\boldsymbol{\theta}^T\boldsymbol{D}_{1,i}\boldsymbol{\theta} - J_{\rm com}(\boldsymbol{\theta})^2\,\boldsymbol{\theta}^T\boldsymbol{D}_{2,i}\boldsymbol{\theta}. \tag{35}$$

This numerator is obviously invariant with respect to scale change. In analogy to (17), we introduce

$$\boldsymbol{N}_{1,\boldsymbol{\theta}} = \sum_{i=1}^n \frac{1}{\boldsymbol{\theta}^T\boldsymbol{B}_i\boldsymbol{\theta}}\boldsymbol{D}_{1,i}, \qquad \boldsymbol{N}_{2,\boldsymbol{\theta}} = \sum_{i=1}^n \frac{1}{\boldsymbol{\theta}^T\boldsymbol{B}_i\boldsymbol{\theta}}\boldsymbol{D}_{2,i} \tag{36}$$

and, in analogy to (16), let

$$\boldsymbol{Y}_{\boldsymbol{\theta}} = \boldsymbol{M}_{\boldsymbol{\theta}} - J_{\rm com}(\boldsymbol{\theta})\,\boldsymbol{N}_{1,\boldsymbol{\theta}} - J_{\rm com}(\boldsymbol{\theta})^2\,\boldsymbol{N}_{2,\boldsymbol{\theta}}. \tag{37}$$

Demanding again that $\boldsymbol{\theta}^T \boldsymbol{Y}_{\boldsymbol{\theta}} \boldsymbol{\theta} = 0$, we obtain the following quadratic equation for $J_{\text{com}}(\boldsymbol{\theta})$:

$$\boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta} - J_{\text{com}}(\boldsymbol{\theta})\, \boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta} - J_{\text{com}}(\boldsymbol{\theta})^2\, \boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\theta}} \boldsymbol{\theta} = 0.$$

This equation has two solutions

$$J^{\pm}_{\text{com}}(\boldsymbol{\theta}) = \frac{\pm\sqrt{(\boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta})^2 + 4\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\theta}} \boldsymbol{\theta} \cdot \boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta}} - \boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta}}{2\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\theta}} \boldsymbol{\theta}}.$$

As $\boldsymbol{M}_{\boldsymbol{\theta}}$ is positive definite, and $\boldsymbol{N}_{1,\boldsymbol{\theta}}$ and $\boldsymbol{N}_{2,\boldsymbol{\theta}}$ are non-negative definite, we have $J^{-}_{\text{com}}(\boldsymbol{\theta}) \leq 0$ and $J^{+}_{\text{com}}(\boldsymbol{\theta}) \geq 0$. Since the compensating factor used in the first-order renormalisation is non-negative, we take, by analogy, $J^{+}_{\text{com}}(\boldsymbol{\theta})$ to be a compensating factor and denote it by $J_{\text{com}}(\boldsymbol{\theta})$; thus

$$J_{\text{com}}(\boldsymbol{\theta}) = \frac{\sqrt{(\boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta})^2 + 4\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\theta}} \boldsymbol{\theta} \cdot \boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta}} - \boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta}}{2\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\theta}} \boldsymbol{\theta}}. \tag{38}$$

Multiplying both numerator and denominator of J_{com} by $[(\boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta})^2 + 4\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\theta}} \boldsymbol{\theta} \cdot \boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta}]^{1/2} + \boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta}$, we see that

$$J_{\text{com}}(\boldsymbol{\theta}) = \frac{2\boldsymbol{\theta}^T \boldsymbol{M} \boldsymbol{\theta}}{\sqrt{(\boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta})^2 + 4\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\theta}} \boldsymbol{\theta} \cdot \boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta}} + \boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta}}. \tag{39}$$

If $\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\theta}} \boldsymbol{\theta} \cdot \boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta}$ is small compared to $(\boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta})^2$, then we may readily infer that

$$J_{\text{com}}(\boldsymbol{\theta}) \approx \frac{\boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\theta}} \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\theta}} \boldsymbol{\theta}}.$$

This expression is very similar to formula (18) for $J_{\text{com}}(\boldsymbol{\theta})$, which indicates that the solution adopted is consistent with the first-order renormalisation.

Inserting $J_{\text{com}}(\boldsymbol{\theta})$ given in (38) into (37), we obtain a well-defined expression for $\boldsymbol{Y}_{\boldsymbol{\theta}}$. We can now use it to define a renormalisation estimate using the renormalisation equation (20). We call this estimate the *second-order renormalisation estimate* and denote it $\widehat{\boldsymbol{\theta}}_{\text{REN2}}$.

7.3 The SORI Scheme

Mimicking the FORI scheme, we can readily advance a scheme for numerically finding $\widehat{\boldsymbol{\theta}}_{\text{REN2}}$. We call this the *second-order renormalisation scheme, Version I*, or the SORI scheme. Its steps are given in Figure (6).

1. Set $\boldsymbol{\theta}_0 = \widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$.
2. Assuming that $\boldsymbol{\theta}_{k-1}$ is known, compute the matrix $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}}$ using (37) and (38).
3. Compute a normalised eigenvector of $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the eigenvalue closest to zero and take this eigenvector for $\boldsymbol{\theta}_k$.
4. If $\boldsymbol{\theta}_k$ is sufficiently close to $\boldsymbol{\theta}_{k-1}$, then terminate the procedure; otherwise increment k and return to Step 2.

Figure 6: Second-order renormalisation scheme, Version I.

7.4 The SORII Scheme

The SORI scheme can be modified in a similar way to that employed with the FORI scheme. The resulting *second-order renormalisation scheme, Version II,* or the SORII scheme, is effectively the second of the two schemes originally proposed by Kanatani.

Introduce

$$J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi}; \boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \frac{\sqrt{(\boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\xi}} \boldsymbol{\theta})^2 + 4\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\xi}} \boldsymbol{\theta} \cdot \boldsymbol{\theta}^T \boldsymbol{M}_{\boldsymbol{\xi}} \boldsymbol{\theta}} - \boldsymbol{\theta}^T \boldsymbol{N}_{1,\boldsymbol{\xi}} \boldsymbol{\theta}}{2\boldsymbol{\theta}^T \boldsymbol{N}_{2,\boldsymbol{\xi}} \boldsymbol{\theta}}. \tag{40}$$

Abbreviating $J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi}; \boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ to $J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi})$, let

$$\boldsymbol{Y}_{\boldsymbol{\theta},\boldsymbol{\xi}} = \boldsymbol{M}_{\boldsymbol{\xi}} - J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi})\, \boldsymbol{N}_{1,\boldsymbol{\xi}} - [J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\xi})]^2\, \boldsymbol{N}_{2,\boldsymbol{\xi}}. \tag{41}$$

It is immediately verified that equations (22) and (24) hold, as does $J_{\mathrm{com}}(\boldsymbol{\theta}) = J'_{\mathrm{com}}(\boldsymbol{\theta}, \boldsymbol{\theta})$, the counterpart of (23).

Again, take $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ to be an initial guess $\boldsymbol{\theta}_0$. Suppose that an update $\boldsymbol{\theta}_{k-1}$ has already been generated. Note that

$$\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1},\boldsymbol{\theta}_{k-1}} = \boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - c_{k-1} \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}} - c_{k-1}^2 \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}, \tag{42}$$

where

$$c_{k-1} = J'_{\mathrm{com}}(\boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k-1}). \tag{43}$$

Let $\boldsymbol{\theta}_k$ be a normalised eigenvector of $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1}}$ corresponding to the smallest eigenvalue λ_k. We intend to find an update c_k appealing directly to c_{k-1}. To this end, observe that

$$\boldsymbol{\theta}_k{}^T (\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - c_{k-1} \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}} - c_{k-1}^2 \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}) \boldsymbol{\theta}_k = \lambda_k. \tag{44}$$

Substituting $\boldsymbol{\theta}_k$ for $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{k-1}$ for $\boldsymbol{\xi}$ in (22), taking into account (41), assuming $J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) = J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_k)$ (as explained analogously when deriving the FORII scheme), and taking into account that, by (43), $J'_{\mathrm{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_k) = c_k$, we obtain

$$\boldsymbol{\theta}_k{}^T(\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - c_k \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}} - c_k^2 \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}})\boldsymbol{\theta}_k = 0.$$

Combining this equation with (44) yields

$$\lambda_k - (c_k - c_{k-1})\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k - (c_k^2 - c_{k-1}^2)\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k = 0. \qquad (45)$$

Let $\Delta c_{k-1} = c_k - c_{k-1}$. Taking into account that $c_k^2 - c_{k-1}^2 = \Delta c_{k-1}(\Delta c_{k-1} + 2c_{k-1})$, we can rewrite (45) as the quadratic constraint on Δc_{k-1} given by

$$\lambda_k - \Delta c_{k-1}\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k - \Delta c_{k-1}(\Delta c_{k-1} + 2c_{k-1})\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k = 0. \qquad (46)$$

Let

$$D_{k-1} = \left[\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k + 2c_{k-1}\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k\right]^2 + 4\lambda_k \boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k.$$

Equation (46) has two solutions

$$\Delta^{\pm} c_{k-1} = \frac{\pm\sqrt{D_{k-1}} - \boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k - 2c_{k-1}\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k}{2\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k}, \qquad (47)$$

which are real when $D_{k-1} \geq 0$.

Suppose that $D_{k-1} \geq 0$. If c_k were directly defined by (43), it would be non-negative. It is therefore reasonable to insist that c_k obtained by updating c_{k-1} also be non-negative. This requirement can be met by setting $c_0 = 0$ and next by ensuring that $\Delta c_{k-1} \geq 0$ for all k. For this reason, we select $\Delta^+ c_{k-1}$ to be Δc_{k-1}, obtaining

$$c_k = c_{k-1} + \frac{\sqrt{D_{k-1}} - \boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k - 2c_{k-1}\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k}{2\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k}. \qquad (48)$$

To treat the case $D_{k-1} < 0$, we first multiply the numerator and denominator of the fractional expression in (48) by $\sqrt{D_{k-1}} + \boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k + 2c_{k-1}\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k$, obtaining

$$c_k = c_{k-1} + \frac{2\lambda_k}{\sqrt{D_{k-1}} + \boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k + 2c_{k-1}\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k}.$$

Next we note that if $\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k$ is small compared to $\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k$, then $D_{k-1} \approx (\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k)^2$, whence

$$\sqrt{D_{k-1}} + \boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k + 2c_{k-1}\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k \approx 2\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k$$

and further

$$c_k \approx c_{k-1} + \frac{\lambda_k}{\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}}\boldsymbol{\theta}_k}.$$

This formula is very similar to (27). We use it with the equality sign instead of the approximation sign to generate c_k in the case $D_{k-1} < 0$.

In this way, we arrive at the following update formula:

$$c_k = \begin{cases} c_{k-1} + \frac{\sqrt{D_{k-1}} - \boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}} \boldsymbol{\theta}_k - 2c_{k-1} \boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}} \boldsymbol{\theta}_k}{2\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}} \boldsymbol{\theta}_k} & \text{if } D_{k-1} \geq 0, \\ c_{k-1} + \frac{\lambda_k}{\boldsymbol{\theta}_k{}^T \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}} \boldsymbol{\theta}_k} & \text{if } D_{k-1} < 0. \end{cases} \tag{49}$$

The SORII scheme can now be formulated as in Figure (7).

1. Set $\boldsymbol{\theta}_0 = \widehat{\boldsymbol{\theta}}_{\text{ALS}}$ and $c_0 = 0$.
2. Assuming that $\boldsymbol{\theta}_{k-1}$ and c_{k-1} are known, compute the matrix $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1},\boldsymbol{\theta}_{k-1}}$ by using (42).
3. Compute a normalised eigenvector of $\boldsymbol{Y}_{\boldsymbol{\theta}_{k-1},\boldsymbol{\theta}_{k-1}}$ corresponding to the smallest eigenvalue λ_k and take this eigenvector for $\boldsymbol{\theta}_k$. Then define c_k by using (49).
4. If $\boldsymbol{\theta}_k$ is sufficiently close to $\boldsymbol{\theta}_{k-1}$, then terminate the procedure; otherwise increment k and return to Step 2.

Figure 7: Second-order renormalisation scheme, Version II.

7.5 The SORIII Scheme

The estimate $\widehat{\boldsymbol{\theta}}_{\text{REN2}}$ can be represented as a limit of a sequence of minimisers of cost functions as follows. Take $\widehat{\boldsymbol{\theta}}_{\text{ALS}}$ for an initial guess $\boldsymbol{\theta}_0$. Suppose that an update $\boldsymbol{\theta}_{k-1}$ has already been generated. Define $\boldsymbol{\theta}_k$ to be the minimiser of the function $\boldsymbol{\theta} \mapsto J'_{\text{com}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{k-1})$:

$$J'_{\text{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) = \min_{\boldsymbol{\theta} \neq \boldsymbol{0}} J'_{\text{com}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{k-1}).$$

Assuming that the sequence $\{\boldsymbol{\theta}_k\}$ converges, take $\lim_{k\to\infty} \boldsymbol{\theta}_k$ for $\widehat{\boldsymbol{\theta}}_{\text{REN2}}$. It can readily be shown that $\widehat{\boldsymbol{\theta}}_{\text{REN2}}$ thus defined satisfies $\boldsymbol{Y}_{\boldsymbol{\theta}} \boldsymbol{\theta} = 0$. It also can be shown that each $\boldsymbol{\theta}_k$ satisfies

$$\boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} \boldsymbol{\theta} = J'_{\text{com}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{k-1}) \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}} \boldsymbol{\theta} + [J'_{\text{com}}(\boldsymbol{\theta}, \boldsymbol{\theta}_{k-1})]^2 \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}} \boldsymbol{\theta}. \tag{50}$$

Here $\boldsymbol{\theta}_k$ is an eigenvector and $J'_{\text{com}}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1})$ is a corresponding eigenvalue of the *quadratic pencil* $\mathcal{P}_{k-1} : \mu \mapsto \mathcal{P}_{k-1}(\mu)$ defined by

$$\mathcal{P}_{k-1}(\mu) = \boldsymbol{M}_{\boldsymbol{\theta}_{k-1}} - \mu \boldsymbol{N}_{1,\boldsymbol{\theta}_{k-1}} - \mu^2 \boldsymbol{N}_{2,\boldsymbol{\theta}_{k-1}} \qquad (\mu \text{ a real number}).$$

In fact, $\boldsymbol{\theta}_k$ is an eigenvector of $\mathcal{P}_{k-1}$ corresponding to the smallest eigenvalue. This observation leads to the *second-order renormalisation scheme, Version III*, or the SORIII scheme, given in Figure (8).

1. Set $\theta_0 = \widehat{\theta}_{\mathrm{ALS}}$.
2. Assuming that θ_{k-1} is known, compute the matrices $M_{\theta_{k-1}}$, $N_{1,\theta_{k-1}}$ and $N_{2,\theta_{k-1}}$.
3. Compute a normalised eigenvector of the eigenvalue problem

$$M_{\theta_{k-1}}\xi = \lambda N_{1,\theta_{k-1}}\xi + \lambda^2 N_{2,\theta_{k-1}}\xi$$

corresponding to the smallest eigenvalue and take this eigenvector for θ_k.
4. If θ_k is sufficiently close to θ_{k-1}, then terminate the procedure; otherwise increment k and return to Step 2.

Figure 8: Second-order renormalisation scheme, Version III.

The eigenvalue problem for a quadratic pencil can readily be reduced to the eigenvalue problem for a linear pencil. Indeed, ξ and λ satisfy

$$M\xi = \lambda N_1\xi + \lambda^2 N_2\xi \tag{51}$$

if and only if there exists ξ' such that

$$\begin{bmatrix} M & -N_1 \\ 0 & I_l \end{bmatrix} \begin{bmatrix} \xi \\ \xi' \end{bmatrix} = \lambda \begin{bmatrix} 0 & N_2 \\ I_l & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \xi' \end{bmatrix},$$

in which case, necessarily, $\xi' = \lambda\xi$. The matrices N_1 and N_2 appropriate for the SORIII scheme are non-negative definite, but not necessarily positive definite. A way of reducing the problem (51) to a similar problem involving positive definite matrices M', N'_1 N'_2 is given in [4] This method takes advantage of the special form of the matrices M_θ, $N_{1,\theta}$, and $N_{2,\theta}$.

8 Experimental Results

The previously derived algorithms were tested on the problem of conic fitting, which constitutes a classical benchmark problem in the literature [2, 5, 7, 8, 10, 11, 15–19, 23]. Specifically, the fitting algorithms were applied to contaminated data arising from a portion of an ellipse.

Synthetic testing is employed here as this enables precise control of the nature of the data and their associated uncertainties. Results obtained in real world testing, in applications domains described earlier, are presented in subsequent work.

Our tests proceeded as follows. A randomly oriented ellipse was generated such that the ratio of its major to minor axes was in the range $[2, 3]$, and its major axis was approximately 200 pixels in length. One third of the ellipse's boundary was chosen as the base curve, and this included the point of maximum curvature of the ellipse. A set of *true points* was then randomly selected from a distribution uniform along the length of the base curve.

For each of the true points, a covariance matrix was randomly generated (using a method described below) in accordance with some chosen *average level of noise*, σ. The true points were then perturbed randomly in accordance with their associated covariance matrices, yielding the *data points*. In general, the noise conformed to an inhomogeneous and anisotropic distribution. Figure 9 shows a large ellipse, some selected true points, a small ellipse for each of these points and the data points. Each of the smaller ellipses represents a level set of the probability density function used to generate the datum, and as such captures graphically the nature of the uncertainty described by its covariance matrix.

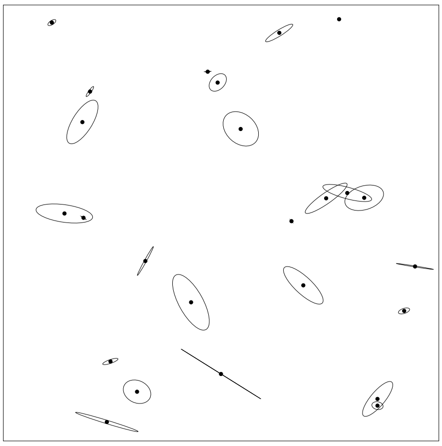

Figure 9: True ellipse, data, and associated covariance ellipses

The following procedure was adopted for generating covariance matrices associated with image points, prescribing (anisotropic and inhomogeneous) noise at a given average level σ. The scale α of a particular covariance matrix was first selected from a uniform distribution in the range $[0, 2\sigma]$. (Similar results were obtained using other distributions about σ.) Next, a skew parameter β was generated from a uniform distribution between 0 and 0.5. An intermediate covariance matrix was then formed by setting

$$\boldsymbol{\Lambda}' = \alpha \begin{bmatrix} \beta & 0 \\ 0 & 1 - \beta \end{bmatrix}.$$

This matrix was then 'rotated' by an angle γ selected from a uniform distribution between 0 and 2π to generate the final covariance

$$\boldsymbol{\Lambda} = \boldsymbol{O}_\gamma \boldsymbol{\Lambda}' \boldsymbol{O}_\gamma{}^T$$

with

$$\boldsymbol{O}_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix}.$$

Let $\mathrm{Tr}\,\boldsymbol{A}$ denote the trace of the matrix $\boldsymbol{A}$. Since $\mathrm{Tr}\,\boldsymbol{\Lambda} = \mathrm{Tr}\,\boldsymbol{\Lambda}' = \alpha$ and $\mathsf{E}\,[\alpha] = \sigma$, it is clear that the above procedure ensures that $\mathsf{E}\,[\mathrm{Tr}\,\boldsymbol{\Lambda}] = \sigma$.

With the data points and their associated covariances prepared, each method under test was then challenged to determine the coefficients of the best fitting conic. (Note, therefore, that it was not assumed that the conic was an ellipse.) The methods were supplied with the data points, and if a specific method was able to utilise uncertainty information, it was also supplied with the data points' covariance matrices. Then estimates were generated, and for each of these a measure of the error was computed using a recipe given below. Testing was repeated many times using newly generated data points (with the covariance matrices and true data points remaining intact). The average errors were then displayed for each method.

The error measure employed was as follows. Assume that a particular method has estimated an ellipse. The error in this estimate was declared to be the sum of the shortest (Euclidean) distances of each *true point* from the estimated ellipse. Note that this measure takes advantage of the fact that the underlying true points are known. Were these unknown, an alternative measure might be the sum of the Mahalanobis distances from the data points to the estimated ellipses.

The methods tested were as follows:

- **ALS** = Algebraic least squares scheme,
- **SMP** = Sampson's scheme,
- **FORI** = first-order renormalisation scheme 1,
- **FORII** = first-order renormalisation scheme 2,
- **FORIII** = first-order renormalisation scheme 3,
- **SORI** = second-order renormalisation scheme 1,
- **SORII** = second-order renormalisation scheme 2,
- **SORIII** = second-order renormalisation scheme 3,
- **FNS** = Fundamental numerical scheme.

Table 1 shows the average error obtained when each method was applied to 500 sets of data points, with σ varying from 1 to 10 pixels in steps of 1. Each set of data points was obtained by perturbing 60 given (true) points. Figure 10 shows the tabular data in graphical form. The algebraic least squares method performs worst while some of the renormalisation schemes and the fundamental numerical scheme perform best. Sampson's method is systematically deficient, generating average errors up to 22% greater than the best methods. The SORI and SORII schemes are similarly deficient; however, they are best seen as incremental developments leading to SORIII. Finally, the FORI, FORII, FORIII and SORIII schemes are seen to trail FNS only very slightly.

AVERAGE	SCHEME								
NOISE LEVEL	ALS	SMP	FOR I	FOR II	FOR III	SOR I	SOR II	SOR III	FNS
1.0	2.710	1.093	1.072	1.072	1.071	1.093	1.093	1.071	1.075
2.0	5.579	2.078	1.990	1.987	1.987	2.078	2.078	1.987	1.976
3.0	8.340	3.169	3.077	3.067	3.067	3.169	3.169	3.067	3.049
4.0	11.889	4.515	4.153	4.147	4.147	4.513	4.514	4.147	4.136
5.0	15.091	5.662	5.129	5.092	5.092	5.655	5.661	5.092	5.054
6.0	19.135	7.099	6.128	6.136	6.136	7.091	7.098	6.137	6.102
7.0	22.919	8.103	6.970	6.898	6.898	8.089	8.101	6.898	6.893
8.0	26.036	9.294	8.115	8.037	8.037	9.254	9.288	8.037	7.966
9.0	31.906	10.791	9.036	8.948	8.948	10.748	10.776	8.950	8.827
10.0	34.118	12.403	10.658	10.571	10.571	13.242	12.387	10.573	10.485

Table 1: Error results obtained for all methods

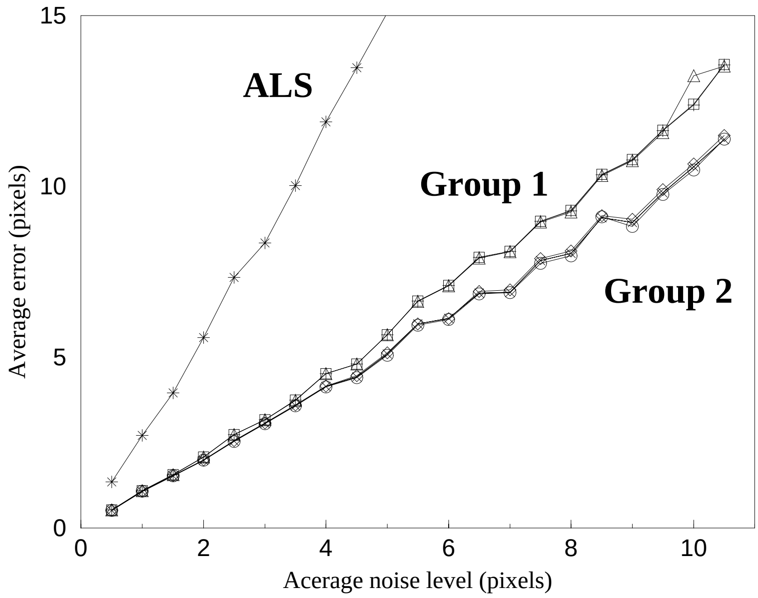

Figure 10: Error results against average noise level depicted graphically. ALS refers to the algebraic least squares method (with some errors out of range). Group 1 comprises SMP, SORI, SORII. Group 2 comprises FORI, FORII, FORIII, SORIII, FNS. See tabulated results for details.

9 Conclusion

The statistical approach to parameter estimation problems of Kenichi Kanatani occupies an important place within the computer vision literature. However, a critical component of this work, the so-called renormalisation method, concerned with minimising particular cost functions, has proven difficult for the vision community to absorb. Our major aim in this paper has been to clarify a number of issues relating to this renormalisation method.

For a relatively general problem form, encompassing many vision problems, we first derived a practical cost function for which claims of optimality may be advanced. We then showed that a Sampson-like method of minimisation generates estimates which are statistically biased. Renormalisation was rationalised as an approach to undoing this bias, and we generated several novel variations on the theme.

Pivotal in the establishing of a framework for comparing selected iterative minimisation schemes was the devising of what we called the fundamental numerical scheme. It emerges that this scheme is not only considerably simpler to derive and implement than its renormalisation-based counterparts, but it also exhibits marginally superior performance.

Acknowledgements

The authors are grateful for the insightful comments of Marino Ivancic, Kenichi Kanatani, Garry Newsam, Naoya Ohta and Robyn Owens. In addition, the authors would like to thank two anonymous referees for providing suggestions that led to improvements in the presentation of the paper. This work was in part funded by the Australian Research Council and the Cooperative Research Centre for Sensor Signal and Information Processing.

References

[1] P. T. Boggs and J. E. Rodgers, *Orthogonal distance regression*, Statistical Analysis of Measurement Error Models and Applications (Arcata, California, June 10–16, 1989), Contemporary Mathematics, vol. 112, American Mathematical Society, Providence, RI, 1989, pp. 183–194.

[2] F. Bookstein, *Fitting conic sections to scattered data*, Computer Graphics and Image Processing **9** (1979), no. 1, 56–71.

[3] M. J. Brooks, W. Chojnacki, and L. Baumela, *Determining the egomotion of an uncalibrated camera from instantaneous optical flow*, Journal of the Optical Society of America A **14** (1997), no. 10, 2670–2677.

[4] W. Chojnacki, M. J. Brooks, and A. van den Hengel, *Fitting surfaces to data with covariance information: Fundamental methods applicable to computer vision*, Technical Report TR 99-03, Department of Computer Sci-

ence, University of Adelaide, South Australia, August 1999, Available from http://www.cs.adelaide.edu.au/~mjb/mjb_abstracts.html.

[5] T. Ellis, *Ellipse detection and matching with uncertainty*, Image and Vision Computing **10** (1992), no. 2, 271–276.

[6] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, The MIT Press, Cambridge, Mass., 1993.

[7] A. Fitzgibbon, M. Pilu, and R. B. Fisher, *Direct least square fitting of ellipses*, IEEE Transactions on Pattern Analysis and Machine Intelligence **21** (1999), no. 5, 476–480.

[8] A. W. Fitzgibbon and R. B. Fisher, *A buyer's guide to conic fitting*, Proceedings of the Sixth British Machine Vision Conference (Birmingham, England, 11-14 September, 1995) (D. Pycock, ed.), BMVA, 1995.

[9] W. A. Fuller, *Measurement Error Models*, John Wiley & Sons, New York, 1987.

[10] W. Gander, G. H. Golub, and R. Strebel, *Least-square fitting of circles and ellipses*, BIT **43** (1994), 558–578.

[11] K. Kanatani, *Statistical bias of conic fitting and renormalisation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **16** (1994), no. 3, 320–326.

[12] ———, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam, 1996.

[13] Y. Leedan and P. Meer, *Heteroscedastic regression in computer vision: problems with bilinear constraint*, International Journal of Computer Vision **37** (2000), no. 2, 127–150.

[14] M. Mühlich and R. Mester, *The role of total least squares in motion analysis*, Computer Vision—ECCV'98, Fifth European Conference on Computer Vision (Freiburg, Germany, June 2–6, 1998) (H. Burkhardt and B. Neumann, eds.), Lecture Notes in Computer Science, vol. 1407, Springer, Berlin, 1998, pp. 305–321.

[15] J. Porill, *Fitting ellipses and predicting confidence envelopes using a bias corrected kalman filter*, Image and Vision Computing **8** (1990), no. 1, 37–41.

[16] P. L. Rosin, *A note on the least square fitting of ellipses*, Pattern Recognition Letters **14** (1993), no. 10, 799–808.

[17] P. L. Rosin and G. A. West, *Nonparametric segmentation of curves into various representations*, IEEE Transactions on Pattern Analysis and Machine Intelligence **17** (1995), no. 12, 1140–1153.

[18] P. D. Sampson, *Fitting conics sections to 'very scattered' data: An iterative refinement of the Bookstein algorithm*, Computer Graphics and Image Processing **18** (1982), no. 1, 97–108.

[19] G. Taubin, *Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations, with applications to edge and range image segmentation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991), no. 11, 1115–1138.

[20] B. Triggs, *A new approach to geometric fitting*, Available from `http://www.inrialpes.fr/movi/people/Triggs`.

[21] ______, *Optimal estimation of matching constraints*, 3D Structure from Multiple Images of Large-Scale Environments (SMILE), workshop organised by the CUMULI, PANORAMA and VANGUARD projects in conjunction with ECCV'98 (Freiburg, Germany, June 6–7, 1998) (R. Koch and L. Van Gool, eds.), Freiburg University, 1998, Full version available from `http://www.inrialpes.fr/movi/people/Triggs`, pp. 30–32.

[22] T. Viéville and O. D. Faugeras, *Motion analysis with a camera with unknown, and possibly varying intrinsic parameters*, Proceedings of the Fifth International Conference on Computer Vision (Cambridge, MA, June 20–23, 1995), IEEE Computer Society Press, Los Alamitos, CA, 1995, pp. 750–756.

[23] Z. Zhang, *Parameter estimation techniques: a tutorial with application to conic fitting*, Image and Vision Computing **15** (1997), no. 1, 57–76.

[24] ______, *On the optimization criteria used in two-view motion analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (1998), no. 7, 717–729.