Int J Comput Vis (2010) 86: 87–110
DOI 10.1007/s11263-009-0260-y

When Discrete Meets Differential

Assessing the Stability of Structure from Small Motion

Wen-Yan Lin · Geok-Choo Tan · Loong-Fah Cheong

Received: 16 September 2008 / Accepted: 31 May 2009 / Published online: 23 June 2009

Abstract We provide a theoretical proof showing that under a proportional noise model, the discrete eight point algorithm behaves similarly to the differential eight point algorithm when the motion is small. This implies that the discrete algorithm can handle arbitrarily small motion for a general scene, as long as the noise decreases proportionally with the amount of image motion and the proportionality constant is small enough. This stability result extends to all normalized variants of the eight point algorithm. Using simulations, we show that given arbitrarily small motions and proportional noise regime, the normalized eight point algorithms outperform their differential counterparts by a large margin. Using real data, we show that in practical small motion problems involving optical flow, these discrete structure from motion (SFM) algorithms also provide better estimates than their differential counterparts, even when the motion magnitudes reach sub-pixel level. The better performance of these normalized discrete variants means that there is much to recommend them as differential SFM algorithms that are linear and normalized.

The support of the DSO grant R263-000-400-592 is gratefully acknowledged.

W.-Y. Lin (✉) · L.-F. Cheong
Electrical and Computer Engineering Department, National University of Singapore, 4 Engineering Drive 3, Singapore 117576, Singapore
e-mail: linwenyan@yahoo.com

L.-F. Cheong
e-mail: eleclf@nus.edu.sg

G.-C. Tan
Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore
e-mail: gctan@ntu.edu.sg

Keywords Structure from motion · Perturbation analysis

1 Introduction

Structure from motion (SFM) is one of the oldest problems in computer vision. Its goal is to obtain a 3-D motion and structure from multiple views of the same scene. Differential algorithms have been employed in SFM for many years. They are formulated for situations in which the motion is very small, such that the said motion can be approximated by velocity. To date, for nearly every discrete SFM algorithm, such as the seminal eight point algorithm by Longuet-Higgins (1981), there exists a differential counterpart (such as Longuet-Higgins and Prazdny 1980). Although there are recent works reporting simulation results which indicate that some discrete SFM algorithms appear capable of handling very small motions (Baumela et al. 2000; Mainberger et al. 2008; Triggs 1999), the suspicion about the stability of the discrete formulation under small motion still persists in some quarters and has not been adequately addressed theoretically. Despite the many error analyses conducted on discrete SFM (Chiuso et al. 2000; Daniilidis and Spetsakis 1997; Kanatani 2003; Luong and Faugeras 1996; Ma et al. 2001; Maybank 1992; Weng et al. 1989; Xiang and Cheong 2003), there is no work that specifically looks at the behaviour of these algorithms under increasingly smaller motions. All error analyses and comparisons are invariably taken as a snapshot at a particular instance of baseline, rather than over the entire span of baselines. What we do find are a number of anecdotal observations stemming from the following line of analysis: at a particular setting of small baseline, and given a particular set of motion/scene configuration, one obtains a certain set of empirical results and thus one accepts the hypothesis that the

discrete SFM algorithm under small baseline is stable or unstable. This lack of clear theoretical evidence continues to prompt conjectures and testing, such as: Is dense optic flow useful to compute the fundamental matrix (Mainberger et al. 2008)? Is the first order differential computation stabler when the disparity is very small (Kanatani 2003)? Would it avoid any singular behaviour potentially present in original discrete formulation? (Ma et al. 2000 showed that as far as linear formulation is concerned, the differential form is by no means simply a "first order approximation" of the discrete case in the sense that a straightforward linearization of the discrete formulation results in certain terms being not separable). Thus, the primary question that we seek to answer here is whether the discrete eight point formulation is stable even under small motion, or it faces fundamental degeneracies which the differential algorithms manage to avoid by making a first order approximation.

If the answer is the former, it calls into question the motivation (except for reason of efficiency) for a large volume of SFM literature which by and large treat the differential problem as something distinct from the discrete one. Examples of differential SFM formulations include Brooks et al. (1997), Fermüller (1995), Heeger et al. (1992), Horn and Weldon (1988), Kanatani (1993), Longuet-Higgins and Prazdny (1980), Ma et al. (2000), Nister (2007) and Viéville and Faugeras (1995). If the answer is the latter, it gives rise to the question of whether a proper understanding of the role of differential algorithms will allow us to design better discrete algorithms that can handle a larger range of baselines.

1.1 The Differential Formulation

Let us begin by considering the motivation underlying differential algorithms. As the name structure from motion suggests, one degenerate scenario common to all SFM algorithms is that of a stationary camera. This degeneracy is intrinsically insurmountable (if there is no motion, there will surely be no structure from motion). However, it brings to mind a set of very interesting questions. How large must the motion be before we can recover structure? Is it possible to recover structure from an infinitesimally small motion? If so, what are the conditions required for a reasonable structure recovery?

Differential SFM algorithms provide a very elegant answer to all of the above questions. They assert that when the motion is small, the movement of the individual feature points on the image plane can be approximated as 2D image velocity (which is in turn approximated by optical flow). After estimating the 2D optical flow, the differential algorithms seeks to compute the differential quantities defining the cameras motion (angular velocity and translation direction) and following that, the scene structure. As these algorithms are formulated in terms of the instantaneous motion, a quantity that is independent of the amount moved, it is clear that provided the image feature velocity (or optical flow) can be extracted reasonably well, the stability of the algorithm is not affected by issues of whether or not a motion is "too small".

In reality, we cannot measure the image feature velocities directly; they are actually obtained from some finite image feature displacements. This means that in order to carry out differential SFM under arbitrarily small motion, the ratio of noise to feature displacement magnitude (i.e. the percentage noise) must be sufficiently small (since if the noise is of fixed magnitude, the feature displacement measurements will eventually be overwhelmed by noise as the camera motion tends to zero). In essence, the underlying premises of the differential formulation is that one can recover structure and motion from a sufficiently small motion, provided one has a reasonable bound on the percentage noise in the feature displacement measurements (this is equivalent to saying that we need a reasonable bound on the fixed noise in the velocity measurements).

In seeking to ascertain if the differential formulation avoids an intrinsic degeneracy present in the discrete formulation we need to consider whether the associated discrete algorithm will yield a reasonable estimate for structure and motion given a sufficiently small motion and a reasonable bound on the percentage noise. Henceforth, we denote algorithms that demonstrate such behavior as being able to handle "differential conditions". We would also like to distinguish between the inherent sensitivity of the underlying problem and the error properties of a particular algorithm for solving that problem. For instance, trying to solve the SFM problem for a configuration near to the critical surface (Maybank 1992; Negahdaripour 1989) is an inherently sensitive problem. No algorithms (discrete or differential) working with finite arithmetic precision can be expected to obtain a solution that is not contaminated with large errors. In this paper, we are primarily interested in the stability of the discrete SFM algorithms under small motion, in the sense that it does not produce any more sensitivity to perturbation than is inherent in the underlying problem. Thus we would only deal with general scenes not close to an inherently ambiguous configuration.

1.2 Noise and Perturbation Analysis

We feel that a major reason for the persistent division of the two view problem into the differential and discrete domain is because it is very difficult to systematically analyze the performance of discrete algorithms when the motion is small. Some intuition into this problem can be obtained by looking at the classical discrete eight point algorithm, where the essential matrix is obtained as the solution to the least squares problem $\min \|Ax\|^2$. Since the solution is in the null space

of the symmetric matrix $A^T A$, the sensitivity of the problem can be characterized by how the eigenvalues and eigenvectors of the data matrix $A^T A$ is influenced by the amount of motion and noise. As we show later, under small motion, the data matrix can be written as:

$$A(\epsilon) \approx A_R + \epsilon A_T$$

where the data matrix $A(\epsilon)$ is now written as a function of ϵ. $A(\epsilon)$ is split into two terms: the residue term A_R when there is no motion, and the motion term ϵA_T, with $\epsilon \to 0$ as the amount of motion becomes progressively smaller. As we will show later, the rank of the matrix A_R is at most 6, and in fact, for a general scene, the rank of A_R is exactly 6. Since A_R has right nullity greater than 1, as ϵ approaches 0 and $A(\epsilon)$ approaches A_R, the problem of finding a unique solution to the right null space of $A^T(\epsilon)A(\epsilon)$ becomes increasingly ill-conditioned as the gaps between the eigenvalues become smaller. In particular, if we assume that a small fixed noise N exists in the estimation process (e.g. noise arising from finite arithmetic precision, which is 16 decimal digits for double precision):

$$\tilde{A}(\epsilon) \approx A_R + \epsilon A_T + N$$

then at a small enough motion, the noise N becomes comparable or even exceeds the motion term ϵA_T such that the legitimate solution is no longer associated with the smallest eigenvalue of $\tilde{A}^T(\epsilon)\tilde{A}(\epsilon)$. This sudden appearance of a second solution has been termed as the second eigenmotion in Ma et al. (2001). This ill-conditioning is the primary reason why vision researchers have reservations over applying the discrete formulation when faced with the problem of small motion.

However, before reaching the limit of arithmetic precision, the noise is likely to be dictated by measurement noise in the feature correspondence or the optical flow (which is actually obtained from feature displacement), and this noise is likely to obey a proportional model. In small motion, the correspondence problem is much simplified by the fact that the two views of the scene do not differ greatly from each other. There will be less hidden surfaces, smaller difference in radiometry, and less geometrical deformation. Hence, although the motion of individual feature points is small, the absolute error incurred in the matching process is also small. For really small motion, differential optical flow algorithms (Horn and Schunck 1981; Lucas and Kanade 1981) would be better placed to yield the desired measurement accuracy, especially with some of the more sophisticated recent implementations (Bruhn et al. 2005; Ho and Goecke 2008; Lempitsky et al. 2008; Nir et al. 2007; Ren 2008; Sand and Teller 2006). The error in estimating image velocity through the Brightness Constancy Equation (BCE) has been analyzed by Verri and Poggio (1989) from which it is clear that the noise in the flow measurement is also likely to be proportional to the magnitude of the motion. It was shown that error stems from various sources, such as changes in the lighting arising from non-uniform illumination or different point of view, or abrupt changes in the reflectance properties of the moving surfaces at the corresponding location in space, all of which are proportional to the magnitude of the motion. Ohta's analysis (Ohta 1996) approached from the perspective of the electronic noise in the imaging devices and also showed the same dependence of the measurement error in the optical flow on the amount of image motion. This is a consequence of the finite receptive field in real cameras, whereby the sampling function is not a Dirac's delta function but rather depends on both the image gradient and the image motion.

In this connection, it is also well to note that many algorithms for finding optical flow make errors not only due to the aforementioned sources, but also due to violation of the flow distribution model that is assumed (such as the smoothness assumption). This latter source of error might give rise to non-proportional noise and thus prevent us from obtaining structure from truly infinitesimally small motions, even if we have succeeded in proving the stability of the discrete eight point algorithm under small motion with proportional noise. However, we envisage that these algorithm-specific errors arising from flow distribution model would become smaller and smaller, especially with the recent spate of optical flow algorithms (Bruhn et al. 2005; Ho and Goecke 2008; Lempitsky et al. 2008; Nir et al. 2007; Ren 2008; Sand and Teller 2006) and together with the publication of a database for optical flow evaluation (Baker et al. 2007). Indeed, with better flows computed from these algorithms in regular usage, there is greater motivation for using flows to recover scene structure since it avoids having to solve the tricky problem of feature correspondence. It then begs the question whether we should recover structure from flow using one of the differential SFM formulations, or if inputting flows to some of the discrete normalized variants offer a better alternative (Mainberger et al. 2008).

1.3 Findings and Organization

In this paper, we carry out perturbation analysis to study the numerical stability of the discrete eight point algorithm and its variants (Chojnacki et al. 2003; Hartley, 1997; Longuet-Higgins 1981; Muhlich and Mester 1998; Torr and Murray 1997) under small motion. The noise regime that we have adopted is such that the data matrix $\tilde{A}(\epsilon)$ is given by

$$\tilde{A}(\epsilon) \approx A_R + \epsilon A_T + M(\epsilon)$$

where $M(\epsilon)$ represents the inherent measurement errors arising from various sources such as the BCE constraint and the electronic noise, both of which are proportional to the amount of motion "ϵ". We show that given a sufficiently small proportional noise $M(\epsilon)$, the discrete eight point algorithm and its variants are all capable of handling "differential conditions". For researchers who view the differential/discrete dichotomy as inviolate, this result is significant because much effort has been spent in refining the discrete eight point algorithm. It permits us to use the more intensively researched discrete algorithms without first reformulating the problem as a differential one; this can result in very large improvements over the current state-of-the-art differential algorithms. As we show later in the experimental section, the normalized discrete algorithms appear to give considerably better performance than its differential counterparts (unnormalized) even when the motion is extremely small. Thus one can indeed obtain a linear normalized differential version by using the normalized 8-point discrete algorithm. For researchers who believe that discrete algorithms can be readily applied to the small motion problem, this paper provides some theoretical explanation for their empirical results and illustrates the limits within which such an attitude may be adopted.

The theoretical portion of the paper is primarily divided into three large portions. The first third of the paper (Sects. 2 to 4) involves introducing the eight point formulation, with some minor reformulations to allow rigorous analysis of its supposed ill-conditioning in the context of small motion. The second third (Sects. 5 to 6) is primarily an adaptation of traditional perturbation theory to our problem of relating baseline to noise. Standard perturbation theory (Wilkinson 1965) usually concerns itself with a fixed data matrix and a changing amount of noise. Our scenario is somewhat different from the customary one: our data matrix $A(\epsilon)$ is also a function of ϵ. A straightforward application of standard results would not be adequate as the Gerschgorin's discs would overlap. Finally, in the last third of the paper (Sect. 7), we complete the stability analysis by tracking how the errors in the fundamental matrix estimate are propagated to the rotation and translation estimates, from which structure of the scene is finally recovered. The theoretical analysis is then followed by the experiments and the conclusion. Lastly, we also record in the appendix some theorems and results required for the perturbation analysis carried out in the paper proper.

1.4 Mathematical Notations

In this section, we explain some of the mathematical notations that a reader will frequently encounter when reading the paper.

1. A^S **symbol**
   Let
   $$A = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}.$$
   The symbol A^S denotes the vector obtained by stacking the columns of A, i.e.,
   $$A^S = \begin{bmatrix} a & b & c & d & e & f & g & h & i \end{bmatrix}^T.$$
2. $\widetilde{\mathbf{w}}$ **symbol**
   Let $\mathbf{w} = \begin{bmatrix} w_1 & w_2 & w_3 & \dots & w_9 \end{bmatrix}^T \in \mathbb{R}^9$. We denote by $\widetilde{\mathbf{w}}$ the following 3×3 matrix
   $$\widetilde{\mathbf{w}} = \begin{bmatrix} w_1 & w_4 & w_7 \\ w_2 & w_5 & w_8 \\ w_3 & w_6 & w_9 \end{bmatrix}.$$
   Clearly, we have $(\widetilde{\mathbf{w}})^S = \mathbf{w}$ and $\widetilde{A^S} = A$.
3. $\widehat{\mathbf{u}}$ **symbol**
   For each $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}^T \in \mathbb{R}^3$, we form the 3×3 skew-symmetric matrix
   $$\widehat{\mathbf{u}} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}.$$
   (a) For $\mathbf{v} \in \mathbb{R}^3$, we have
   $$\widehat{\mathbf{u}}\mathbf{v} = \mathbf{u} \times \mathbf{v}, \tag{1}$$
   where $\mathbf{u} \times \mathbf{v}$ is the vector product of $\mathbf{u}$ and $\mathbf{v}$.
   (b) For a 3×3 invertible matrix A, with $\det(A) \neq 0$, we have the following result from page 456 of Ma et al. (2003).
   $$(A^{-1})^T \widehat{\mathbf{u}} A^{-1} = \frac{1}{\det(A)} (\widehat{A\mathbf{u}}). \tag{2}$$
4. Throughout this paper, we work on the Frobenius norm of a matrix (say C) which is defined and denoted as follows:
   $$\|C\| = \sqrt{\sum_{i,j} c_{ij}^2}.$$
   It generalizes the definition of the usual norm on vectors.
5. Δx **symbol**
   Suppose the function $x(\epsilon)$ is defined for $\epsilon \geq 0$. We shall use the usual notation Δx to denote the change in x:
   $$\Delta x = x(\epsilon) - x(0).$$

Likewise, we have ΔY_i, etc. Sometimes, to avoid cumbersome notation, we denote a function $x(\epsilon)$ at $\epsilon = 0$ by just x.

6. For the ease of reading this paper, we gather in the following a table of symbols for the eigenvalues and eigenvectors of the matrices $A_R^T A_R$, $A^T(\epsilon)A(\epsilon)$ and $\tilde{A}(\epsilon)^T \tilde{A}(\epsilon)$ (to be introduced in subsequent sections).

Matrix	Eigenvalues	Eigenvectors
$A_R^T A_R$	λ_i	$\mathbf{r}_i$, unit vector
$A^T(\epsilon)A(\epsilon)$	$\lambda_i(\epsilon)$	$\mathbf{q}_i(\epsilon)$, unit vector
$\tilde{A}(\epsilon)^T \tilde{A}(\epsilon)$	$\tilde{\lambda}_i(\epsilon)$	$\tilde{\mathbf{q}}_i(\epsilon)$

1.5 Mathematical Expressions

The following phrases will be frequently encountered in this paper.

1. **For a sufficiently small ϵ:** If we say that a condition (or a statement) X is satisfied for a sufficiently small ϵ, it means that there exists a positive $\epsilon_0 > 0$, such that the condition (or statement) X is satisfied for all ϵ where $0 \le \epsilon < \epsilon_0$.

2. **Order ϵ^n or $O(\epsilon^n)$:** For an integer n, a function $f(\epsilon)$ is said to be of order ϵ^n if $|f(\epsilon)| \le K\epsilon^n$ for some $K > 0$ as $\epsilon \to 0$. That is, for a sufficiently small $\epsilon > 0$, $|\frac{f(\epsilon)}{\epsilon^n}|$ is uniformly bounded. In symbol, we write $f(\epsilon) = O(\epsilon^n)$. When $n = 0$, we write $O(\epsilon^0)$.

   Some special cases/notes:

   (a) For a function $f(\epsilon)$, we note that

   $$f(\epsilon) = O(\epsilon^{n+1}) \quad \Rightarrow \quad f(\epsilon) = O(\epsilon^n),$$

   but the converse is not true in general. In other words, $O(\cdot)$ is not an asymptotically tight bound.

   (b) If $F(\epsilon)$ is a matrix (or a vector), then saying it is of order ϵ^n means that each of its individual entries is of order ϵ^n. This is equivalent to saying that the norm of $F(\epsilon)$ (which is a real valued function) is of order ϵ^n.

   (c) Let k be a rational number. For a sufficiently small η, we have

   $$(1+\eta)^k = 1 + k\eta + O(\eta^2).$$

   This follows from the first order term of the Taylor expansion. In particular, for non-negative real numbers n and l, and sufficiently small ϵ and m, we have

   $$(1 + O(\epsilon^n)m^l)^k = 1 + O(\epsilon^n)m^l, \tag{3}$$

   where the constant k has been absorbed in the O-notation.

2 A Single Moving Camera Viewing a Stationary Scene

Let us assume that there is a single moving camera viewing a stationary scene consisting of N feature points $\mathbf{P}_i$, where $1 \le i \le N$.

Let $\epsilon \ge 0$ be a non-negative real number representing the elapsed time. Our goal in this section is to formulate the eight point algorithm in the form of a data matrix and a solution vector, both of which can be expressed as a series in ϵ. Subsequently, we will use matrix perturbation theory to analyze their properties when the elapsed time ϵ (and hence the motion) is small.

At time instance $\epsilon \ge 0$, a point $\mathbf{P}_i$ has its coordinates with respect to the camera reference frame given by

$$\mathbf{P}_i(\epsilon) = \begin{bmatrix} X_i(\epsilon) & Y_i(\epsilon) & Z_i(\epsilon) \end{bmatrix}^T.$$

Let us assume that the motion is smooth, with the camera positions being related to each other by the translation vector $\epsilon\mathbf{T}_c$ and a smoothly changing rotation $R(\epsilon)$. The 3×1 vector $\mathbf{T}_c$ is a constant vector representing the translational velocity, whereas the 3×3 matrix $R(\epsilon)$ is a rotation matrix which changes smoothly with ϵ and $R(0) = I$, where I is the 3×3 identity matrix.

The rotation matrix $R(\epsilon)$ can be expressed as the exponential of some skew-symmetric matrix $\widehat{\boldsymbol{\omega}}$, that is, a series of the form (Theorem 2.8, Ma et al. 2003)

$$R(\epsilon) = I + \epsilon\widehat{\boldsymbol{\omega}} + O(\epsilon^2), \tag{4}$$

where $\boldsymbol{\omega}$ is the angular velocity.

As a result of the motion, we have

$$\mathbf{P}_i(\epsilon) = R(\epsilon)(\mathbf{P}_i - \epsilon\mathbf{T}_c). \tag{5}$$

Recall from the preceding section that sometimes we shall denote $\mathbf{P}_i(0) = \mathbf{P}_i$. When projected onto the image plane of the camera, the points $\mathbf{P}_i$ and $\mathbf{P}_i(\epsilon)$ will have image coordinates $\mathbf{p}_i$ and $\mathbf{p}_i(\epsilon)$ respectively where

$$\begin{aligned} \mathbf{p}_i &= \frac{1}{Z_i}\mathbf{P}_i = \begin{bmatrix} x_i & y_i & 1 \end{bmatrix}^T, \\ \mathbf{p}_i(\epsilon) &= \frac{1}{Z_i(\epsilon)}\mathbf{P}_i(\epsilon) = \begin{bmatrix} x_i(\epsilon) & y_i(\epsilon) & 1 \end{bmatrix}^T. \end{aligned} \tag{6}$$

Using (5) and (6), we have

$$\mathbf{p}_i(\epsilon) = \mathbf{p}_i + \epsilon[\,\Delta_t x \quad \Delta_t y \quad 0\,]^T + O(\epsilon^2) \tag{7}$$

where $\Delta_t x$, $\Delta_t y$ are the x and y components of the image feature velocity respectively.

2.1 Epipolar Constraint with Normalization

Given two camera images, one at time 0 and the other at time ϵ, the epipolar constraint is

$$\mathbf{p}_i^T E(\epsilon)\mathbf{p}_i(\epsilon) = 0, \tag{8}$$

where $E(\epsilon) = \widehat{\mathbf{T}}_c R^T(\epsilon)$.

Given eight or more point matches, the above epipolar constraint is sufficient for us to determine the essential matrix $E(\epsilon)$ up to a scale factor, by solving a set of linear equations.

This is the famous eight point algorithm of Longuet-Higgins (1981). However, it is important to note that the epipolar constraint is seldom used in its raw form. Rather, for the sake of numerical stability in the presence of noise, a normalization procedure is often employed.

Let Θ be an 3×3 invertible matrix introduced for this purpose. For example, it can be of the form

$$\begin{bmatrix} a & 0 & c \\ 0 & b & d \\ 0 & 0 & 1 \end{bmatrix},$$

with $ab \neq 0$. Examples of normalization matrices taking such form are the normalization matrix in Hartley normalization (Hartley, 1997), or in the context of uncalibrated motion analysis, Θ would be the camera's intrinsic matrix.

For a sufficiently small $\epsilon \geq 0$, suppose

$$\Theta(\epsilon) = \Theta + O(\epsilon) \tag{9}$$

is invertible. Then its inverse $(\Theta(\epsilon))^{-1}$ takes the form

$$(\Theta(\epsilon))^{-1} = \Theta^{-1} + O(\epsilon). \tag{10}$$

We denote the normalized (or uncalibrated) version of the essential matrix as $F(\epsilon)$ where

$$F(\epsilon) = (\Theta^T)^{-1} E(\epsilon)(\Theta(\epsilon))^{-1}. \tag{11}$$

We will sometimes call $F(\epsilon)$ the fundamental matrix where appropriate.

Using (9) and (7), we can write

$$\Theta(\epsilon)\mathbf{p}_i(\epsilon) = [\underline{x_i} \quad \underline{y_i} \quad 1]^T + \epsilon[\underline{\Delta_t x} \quad \underline{\Delta_t y} \quad 0] + O(\epsilon)^2, \tag{12}$$

where

$$\Theta\mathbf{p}_i = [\underline{x_i} \quad \underline{y_i} \quad 1]^T,$$

$$\Theta(\epsilon)\mathbf{p}_i(\epsilon) = [\underline{x_i}(\epsilon) \quad \underline{y_i}(\epsilon) \quad 1]^T$$

and $(\underline{\Delta_t x}, \underline{\Delta_t y})$ is the image feature velocity in the normalized system. In this normalized system, the corresponding epipolar constraint (8) becomes

$$[(\Theta\mathbf{p}_i)]^T F(\epsilon)[(\Theta(\epsilon))\mathbf{p}_i(\epsilon)] = 0. \tag{13}$$

Collecting N such constraints for $i = 1, \ldots, N$, we form a system of linear equations:

$$A(\epsilon)(F(\epsilon))^S = 0, \tag{14}$$

where

$$A(\epsilon) = \begin{bmatrix} \underline{x_1(\epsilon)x_1} & \underline{x_1(\epsilon)y_1} & \underline{x_1(\epsilon)} & \underline{y_1(\epsilon)x_1 y_1(\epsilon)y_1} & \underline{y_1(\epsilon)} & \underline{x_1} & \underline{y_1} & 1 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{x_N(\epsilon)x_N} & \underline{x_N(\epsilon)y_N} & \underline{x_N(\epsilon)} & \underline{y_N(\epsilon)x_N} & \underline{y_N(\epsilon)y_N} & \underline{y_N(\epsilon)} & \underline{x_N} & \underline{y_N} & 1 \end{bmatrix} \tag{15}$$

and $(F(\epsilon))^S$ is the column vector defined in Sect. 1.4. Thus, an estimate of the matrix $F(\epsilon)$ can be obtained via the null space of $A(\epsilon)$.

Finally, we rewrite the matrix $F(\epsilon)$ in (11) into a form more amenable to analysis:

$$\begin{aligned} F(\epsilon) &= (\Theta^T)^{-1} E(\epsilon)(\Theta(\epsilon))^{-1} \\ &= (\Theta^{-1})^T \widehat{\mathbf{T}}_c \Theta^{-1} \Theta R^T(\epsilon)(\Theta(\epsilon))^{-1} \\ &= \frac{1}{\det(\Theta)}[\widehat{(\Theta\mathbf{T}_c)}][\Theta R^T(\epsilon)(\Theta(\epsilon))^{-1}] \end{aligned} \tag{16}$$

where the last step has been obtained by using (2). By (10) and (4), we have $\Theta R^T(\epsilon)(\Theta(\epsilon))^{-1} = I + O(\epsilon)$ which gives

$$F(\epsilon) = \frac{1}{\det(\Theta)}[\widehat{(\Theta\mathbf{T}_c)}][I + O(\epsilon)]. \tag{17}$$

Since $F(\epsilon)$ is defined up to a scale factor, we can write

$$F(\epsilon) = \widehat{\mathbf{T}} + O(\epsilon) \tag{18}$$

where

$$\mathbf{T} = \frac{\Theta\mathbf{T}_c}{\sqrt{2}\|\Theta\mathbf{T}_c\|}. \tag{19}$$

Here, we have set $\|\mathbf{T}\| = \frac{1}{\sqrt{2}}$ so that $(\widehat{\mathbf{T}})^S$ has unit norm. From (16) and (1), we note that $\mathbf{T}$ is in the left null space of $F(\epsilon)$.

3 The Degeneracy Affecting the Discrete Algorithm

Using (12), we rewrite the data matrix $A(\epsilon)$ as a series expansion in ϵ,

$$A(\epsilon) = A_R + \epsilon A_T + O(\epsilon^2) \tag{20}$$

where

$$A_R = \begin{bmatrix} \underline{x_1^2} & \underline{x_1 y_1} & \underline{x_1} & \underline{y_1 x_1} & \underline{y_1^2} & \underline{y_1} & \underline{x_1} & \underline{y_1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{x_N^2} & \underline{x_N y_N} & \underline{x_N} & \underline{y_N x_N} & \underline{y_N^2} & \underline{y_N} & \underline{x_N} & \underline{y_N} & 1 \end{bmatrix},$$
$$A_T = \begin{bmatrix} \underline{x_1 \Delta_t x_1} & \underline{y_1 \Delta_t x_1} & \underline{\Delta_t x_1} & \underline{x_1 \Delta_t y_1} & \underline{y_1 \Delta_t y_1} & \underline{\Delta_t y_1} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{x_N \Delta_t x_N} & \underline{y_N \Delta_t x_N} & \underline{\Delta_t x_N} & \underline{x_N \Delta_t y_N} & \underline{y_N \Delta_t y_N} & \underline{\Delta_t y_N} & 0 & 0 & 0 \end{bmatrix}. \tag{21}$$

Recall that when the motion (i.e., ϵ) is small, the discrete eight point algorithm is regarded as increasingly ill conditioned. In this section, we revisit the explanation in terms of the data matrix $A(\epsilon)$.

As ϵ tends to zero, using (20), we know that $A(\epsilon)$ tends to A_R. Let F_0 be a 3×3 matrix satisfying

$$(\Theta \mathbf{p}_i)^T F_0 (\Theta(0) \mathbf{p}_i(0)) = 0$$

i.e.,

$$(\Theta \mathbf{p}_i)^T F_0 (\Theta \mathbf{p}_i) = 0, \tag{22}$$

which is the constraint given in (13) when $\epsilon = 0$. Solving F_0 from (22) is equivalent to solving the following linear least squares system

$$A_R (F_0)^s = 0$$

whose solution space we will analyze now.

3.1 The Null Space of $A_R^T A_R$

In this subsection, we prove that for a general scene, the nullity of the 9×9 matrix $A_R^T A_R$ is 3 and we also determine the null space of $A_R^T A_R$.

Assume that the feature points on the image plane are well distributed such that we cannot fit a conic section that passes through all of them (this condition is easily satisfied, especially under small motion where the number of features which can be matched is very dense). We then have the following result.

Proposition 1 *Assume that all the feature points on the image plane do not lie on any conic section. The nullity of $A_R^T A_R$ is* 3.

Proof Since

$$A_R^T A_R \mathbf{u} = \mathbf{0} \quad \Leftrightarrow \quad A_R \mathbf{u} = \mathbf{0},$$

we shall determine the nullity of A_R.

Note that the matrix A_R in (21) contains 3 pairs of identical columns, namely columns 2 and 4, columns 3 and 7, and columns 6 and 8. Hence, the rank of A_R is at most 6.

Consider the submatrix A'_R formed from A_R by removing one copy of each repeating column pair:

$$A'_R = \begin{bmatrix} \underline{x_1^2} & \underline{x_1 y_1} & \underline{x_1} & \underline{y_1^2} & \underline{y_1} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \underline{x_N^2} & \underline{x_N y_N} & \underline{x_N} & \underline{y_N^2} & \underline{y_N} & 1 \end{bmatrix}.$$

We shall show that the 6 columns of A'_R are linearly independent so that the rank of A_R is at least 6. Suppose $A'_R \mathbf{v} = 0$, where $\mathbf{v} = [\, a \; b \; c \; d \; e \; f \,]^T \neq \mathbf{0}$. This gives,

$$a \underline{x_i^2} + b \underline{x_i y_i} + c \underline{x_i} + d \underline{y_i^2} + e \underline{y_i} + f = 0, \quad 1 \leq i \leq N,$$

which means that all the feature points lie on the conic defined by $ax^2 + bxy + cx + dy^2 + ey + f = 0$. This violates our assumption. So, we must have $\mathbf{v} = \mathbf{0}$, which implies that the rank of A'_R (and hence A_R) is at least 6. Therefore, the rank of A_R is 6. By the rank-nullity formula, the nullity of A_R is given by $(9 - (\text{rank of } A_R))$. Hence the nullity of A_R is 3, and so is that of $A_R^T A_R$. □

Proposition 2 *The null space of $A_R^T A_R$ is the following set*

$$S = \{(\widehat{\mathbf{u}})^S \mid \mathbf{u} \in \mathbb{R}^3\} = \left\{ \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix}^S \;\middle|\; u_1, u_2, u_3 \in \mathbb{R} \right\}.$$

Proof By (1), every skew-symmetric matrix $\widehat{\mathbf{u}}$ formed from $\mathbf{u} \in \mathbb{R}^3$ will satisfy (22). Thus, the set $S \subseteq$ null space of A_R.

However, the set S is a subspace of $\mathbb{R}^9$ and its dimension is 3. Since the nullity of A_R is also 3, the set S is indeed the null space of A_R. □

It is a well known fact that a real symmetric matrix of the form $Z^T Z$ has non-negative eigenvalues. Thus, we can arrange the 9 non-negative eigenvalues λ_i of the matrix $A_R^T A_R$ in a non-increasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots > \lambda_7 \geq \lambda_8 \geq \lambda_9 \geq 0.$$

Proposition 3 *Consider the real symmetric matrix $A_R^T A_R$, and let λ_i be its eigenvalue, with corresponding unit eigenvector $\mathbf{r}_i$, for $1 \leq i \leq 9$. Then we have*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_6 > \lambda_7 = \lambda_8 = \lambda_9 = 0.$$

Moreover, we may choose $\mathbf{r}_7, \mathbf{r}_8$ and $\mathbf{r}_9$ such that

$$\mathbf{r}_7 = \widehat{\mathbf{T}_7}^S, \qquad \mathbf{r}_8 = \widehat{\mathbf{T}_8}^S, \qquad \mathbf{r}_9 = \widehat{\mathbf{T}}^S$$

where $\mathbf{T}$ is defined in (19), and $\mathbf{T}_7$, $\mathbf{T}_8$ and $\mathbf{T}$ are mutually orthogonal vectors of norm $\frac{1}{\sqrt{2}}$.

Proof It follows from the nullity of A_R (and hence $A_R^T A_R$) being 3 that the real symmetric matrix $A_R^T A_R$ has a zero eigenvalue, with multiplicity 3. Thus, $\lambda_7 = \lambda_8 = \lambda_9 = 0$.

The eigen-space corresponding to the zero eigenvalue is indeed the null space of $A_R^T A_R$. Since the null space of $A_R^T A_R$ is spanned by its three eigenvectors, we are free to choose $\mathbf{r}_7$, $\mathbf{r}_8$ and $\mathbf{r}_9$, as long as they belong to the subspace S in Proposition 2, and are orthonormal to each other. Therefore we choose to set

$$\mathbf{r}_9 = \widehat{\mathbf{T}}^S, \tag{23}$$

where $\mathbf{T}$ is defined in (19).

By Proposition 2, the other two eigenvectors $\mathbf{r}_7$ and $\mathbf{r}_8$ can also be written in the form $\mathbf{r}_7 = \widehat{\mathbf{T}_7}^S$, $\mathbf{r}_8 = \widehat{\mathbf{T}_8}^S$, where $\mathbf{T}_7$, $\mathbf{T}_8$ and $\mathbf{T}$ must be mutually orthogonal vectors of norm $\frac{1}{\sqrt{2}}$ to ensure the orthonormality of $\mathbf{r}_7$, $\mathbf{r}_8$ and $\mathbf{r}_9$. □

Since A_R has right nullity greater than 1, as ϵ approaches 0 and $A(\epsilon)$ approaches A_R, the problem of finding a unique solution to the right null space of $A^T(\epsilon)A(\epsilon)$ (recall that camera pose is estimated from the right null space of $A(\epsilon)$) becomes increasingly ill-conditioned. This ill-conditioning is the primary reason why vision researchers have reservations over applying the discrete formulation when faced with the problem of small motion. However, as we have argued in Sect. 1, if the noise in the flow estimation is proportional to ϵ, the question then becomes whether the noise declining proportionally to ϵ is sufficient to compensate for the increased instability due to the last three eigenvalues of $A^T(\epsilon)A(\epsilon)$ getting closer together. In the next section, we explain how this problem can be analyzed using matrix perturbation theory.

4 On the Noiseless Case $A(\epsilon)^T A(\epsilon)$

The least squares solution to (14) is given by the right null space of the 9×9 symmetric matrix $A^T(\epsilon)A(\epsilon)$. As such, the subsequent analysis is conducted on $A^T(\epsilon)A(\epsilon)$ rather than $A(\epsilon)$.

If one thinks of the eigenvectors $\mathbf{q}_i(\epsilon)$ of $A^T(\epsilon)A(\epsilon)$ as possible solutions to (14), then their corresponding eigenvalues $\lambda_i(\epsilon)$ are the residue (sum of squared error) related to these solutions. That is, we have

$$\mathbf{q}_i(\epsilon)^T A(\epsilon)^T A(\epsilon)\mathbf{q}_i(\epsilon) = \mathbf{q}_i(\epsilon)^T \lambda_i(\epsilon)\mathbf{q}_i(\epsilon) = \lambda_i(\epsilon).$$

Thus, each $\lambda_i(\epsilon)$ represents the residue of $A(\epsilon)$ associated with $\mathbf{q}_i(\epsilon)$. The larger the value of $\lambda_i(\epsilon)$, for $1 \leq i \leq 8$, the more stable is the solution as the "wrong" solution is less likely to be confused with the correct one.

In the absence of noise, using (20), the matrix $A^T(\epsilon)A(\epsilon)$ can be expressed as the following series expansion,

$$A^T(\epsilon)A(\epsilon) = A_R^T A_R + \epsilon(A_T^T A_R + A_R^T A_T) + O(\epsilon^2). \tag{24}$$

This says that the matrix $A^T(\epsilon)A(\epsilon)$ can be thought of as the "perturbation" of the matrix $A_R^T A_R$ by the matrix $\epsilon(A_T^T A_R + A_R^T A_T)$ for sufficiently small ϵ.

We shall use matrix perturbation theory to discuss the eigenvectors and eigenvalues of this "perturbed" matrix.

4.1 How the Eigenvectors of $A^T(\epsilon)A(\epsilon)$ Vary with ϵ

Let us denote the eigenvalues of the matrix $A^T(\epsilon)A(\epsilon)$ by $\lambda_i(\epsilon)$, $i = 1, 2, \ldots, 9$, where

$$\lambda_1(\epsilon) \geq \lambda_2(\epsilon) \geq \cdots \geq \lambda_9(\epsilon) \geq 0.$$

We shall now choose corresponding unit eigenvectors $\mathbf{q}_i(\epsilon)$, for $1 \leq i \leq 9$, such that $\{\mathbf{q}_1(\epsilon), \mathbf{q}_2(\epsilon), \ldots, \mathbf{q}_9(\epsilon)\}$ is an orthonormal basis of $\mathbb{R}^9$.

It is clear from (16) and the definition of $A(\epsilon)$ in (15) that the actual camera motion satisfies (14). Thus $A(\epsilon)$ and hence

$A(\epsilon)^T A(\epsilon)$ have a nullity of at least one. In other words, we have $\lambda_9(\epsilon) = 0$ and $(F(\epsilon))^S$ is the corresponding eigenvector.

It follows from both (18) and our choice of $\mathbf{r}_9$ in (23) that

$$(F(\epsilon))^S = \mathbf{r}_9 + O(\epsilon).$$

Normalizing $(F(\epsilon))^S$, we obtain the unit eigenvector $\mathbf{q}_9(\epsilon) = \frac{(F(\epsilon))^S}{\|(F(\epsilon))^S\|}$ corresponding to the eigenvalue $\lambda_9(\epsilon) = 0$. By Lemma 6 (in Appendix A.1), we have

$$\mathbf{q}_9(\epsilon) = \mathbf{r}_9 + \mathbf{z}_9(\epsilon) \text{ where } \mathbf{z}_9(\epsilon) = O(\epsilon). \tag{25}$$

Treating the matrix $A^T(\epsilon)A(\epsilon)$ as the "perturbation" of the matrix $A_R^T A_R$ by $\epsilon(A_T^T A_R + A_R^T A_T)$ for sufficiently small ϵ, we apply perturbation theory (in particular, Theorem 8 in Appendix A.1) to obtain the following result for the remaining unit eigenvectors of $A^T(\epsilon)A(\epsilon)$.

Theorem 1 *The set of unit eigenvectors of $A^T(\epsilon)A(\epsilon)$ given by*

$$\{\mathbf{q}_1(\epsilon), \mathbf{q}_2(\epsilon), \ldots, \mathbf{q}_9(\epsilon)\}$$

can be chosen such that

$$\mathbf{q}_i(\epsilon) = \mathbf{r}'_i(\epsilon) + \mathbf{z}_i(\epsilon),$$

where $\|\mathbf{r}'_i(\epsilon)\| = 1$ and $\mathbf{z}_i(\epsilon) = O(\epsilon)$. Moreover, the vectors $\mathbf{r}'_i(\epsilon)$'s have the following properties:

1. $\mathbf{r}'_9(\epsilon) = \mathbf{r}_9 = \widehat{\mathbf{T}}^S$ *(from* (25)*)*.
2. $\mathbf{r}'_i(\epsilon)$ *is a linear combination of all eigenvectors* $\mathbf{r}_j$ *of* $A_R^T A_R$*, whose associated eigenvalue* λ_j *is identical to* λ_i*, for* $i \leq 9$.
3. $\mathbf{r}'_i(\epsilon)$ *is orthogonal to* $\mathbf{r}_j$ *if* $\lambda_i \neq \lambda_j$.
4. $\mathbf{r}'_i(\epsilon)$, $1 \leq i \leq 8$ *is orthogonal to* $\mathbf{r}_9$.

Remark 1 For $7 \leq i \leq 9$, each vector $\mathbf{r}'_i(\epsilon)$ is a linear combination of $\mathbf{r}_7$, $\mathbf{r}_8$ and $\mathbf{r}_9$, and hence it is a vector in the right null space of A_R. From Proposition 2, we have

$$\widehat{\mathbf{r}'_7(\epsilon)} = \widehat{\mathbf{T}'_7(\epsilon)}, \quad \text{and} \quad \widehat{\mathbf{r}'_8(\epsilon)} = \widehat{\mathbf{T}'_8(\epsilon)}$$

for some orthogonal vectors $\mathbf{T}'_7(\epsilon)$ and $\mathbf{T}'_8(\epsilon)$ in $\mathbb{R}^3$ where

$$\|\mathbf{T}'_7(\epsilon)\| = \|\mathbf{T}'_8(\epsilon)\| = \|\mathbf{T}\| = \frac{1}{\sqrt{2}}.$$

4.2 How the Eigenvalues of $A^T(\epsilon)A(\epsilon)$ Vary with ϵ

As discussed in the preceding section, $\lambda_9(\epsilon) = 0$. For the remaining eigenvalues $\lambda_i(\epsilon)$, applying perturbation theory (Theorem 6 in Appendix A.1) to the expression for $A^T(\epsilon)A(\epsilon)$ in (24) yields the following result.

Proposition 4 *For* $1 \leq i \leq 8$,

$$\lambda_i(\epsilon) = \lambda_i + O(\epsilon).$$

From Proposition 3 we know that $\lambda_i > 0$ for $1 \leq i \leq 6$. As such, when ϵ is sufficiently small, using Proposition 4, we know that eigenvalues $\lambda_i(\epsilon)$ remains positive and hence their corresponding eigenvectors are distinct from the true solution.

However, for $7 \leq i \leq 8$, we note that Proposition 3 indicates that $\lambda_i(\epsilon)$ may be zero. From the point of view of stability, this is worrying and we must seek a more explicit expression than that offered by standard matrix perturbation theory.

Lemma 1 *For $i = 7$ or 8, if the hypothesis $\|A(\epsilon)\mathbf{q}_i(\epsilon)\| = \gamma_i\epsilon + O(\epsilon^2)$ where $\gamma_i > 0$ is true, then $\lambda_i(\epsilon) = O(\epsilon^2)$. In particular,*

$$\lambda_i(\epsilon) = \Lambda_i\epsilon^2 + O(\epsilon^3), \quad \textit{where } \Lambda_i = \gamma_i^2 > 0.$$

Proof This follows readily from the hypothesis since

$$\begin{aligned}\lambda_i(\epsilon) &= \mathbf{q}_i^T(\epsilon)A^T(\epsilon)A(\epsilon)\mathbf{q}_i(\epsilon)\\ &= \|A(\epsilon)\mathbf{q}_i(\epsilon)\|^2\\ &= \Lambda_i\epsilon^2 + O(\epsilon^3),\end{aligned}$$

where $\Lambda_i = \gamma_i^2 > 0$. □

We shall explain why the hypothesis imposed on $A(\epsilon)\mathbf{q}_i(\epsilon)$ is meaningful. Note that for $7 \leq i \leq 8$,

$$\begin{aligned}A(\epsilon)\mathbf{q}_i(\epsilon) &= (A_R + \epsilon A_T)(\mathbf{r}'_i(\epsilon) + \mathbf{z}_i(\epsilon))\\ &= A_R\mathbf{z}_i(\epsilon) + \epsilon A_T\mathbf{r}'_i(\epsilon) + O(\epsilon^2)\\ &= \epsilon\left(\frac{1}{\epsilon}A_R\mathbf{z}_i(\epsilon) + A_T\mathbf{r}'_i(\epsilon)\right) + O(\epsilon^2),\end{aligned}$$

where $A_R\mathbf{z}_i(\epsilon) + \epsilon A_T\mathbf{r}'_i(\epsilon) = O(\epsilon)$, which is the first order approximation of the residue of $A(\epsilon)$ associated with the solution $\mathbf{q}_i(\epsilon)$. The hypothesis that $\gamma_i > 0$ is intimately related to the assumption that we are dealing with a non-degenerate scene configuration in a differential setting. The reason can be seen by looking at the square of the coefficient of the first order term in the preceding equation and substituting the expressions for A_R and A_T from (21):

$$\begin{aligned}&\left\|\frac{1}{\epsilon}A_R\mathbf{z}_i(\epsilon) + A_T\mathbf{r}'_i(\epsilon)\right\|^2\\ &= \sum_{j=1}^{N}((\underline{\Delta_t\mathbf{p}_j})^T\,\underline{\widehat{\mathbf{r}'_i(\epsilon)}\,\mathbf{p}_j} + \underline{\mathbf{p}_j}^T\,\underline{\widehat{\mathbf{z}'_i(\epsilon)}\,\mathbf{p}_j})^2\end{aligned} \tag{26}$$

where

$$\mathbf{z}'_i(\epsilon) = \frac{1}{\epsilon}\mathbf{z}_i(\epsilon) = O(\epsilon^0),$$

$$\underline{\mathbf{p}_j} = [\underline{x_j} \quad \underline{y_j} \quad 1]^T, \quad \underline{\Delta_t \mathbf{p}_j} = [\underline{\Delta_t x_j} \quad \underline{\Delta_t y_j} \quad 0]^T.$$

Equation (26) should be familiar to most vision researchers: it is the sum squared error of the differential fundamental matrix (Ma et al. 2003; Viéville and Faugeras 1995) associated with the "solution" $\mathbf{r}'_i(\epsilon)$ and $\mathbf{z}'_i(\epsilon)$, with $\mathbf{r}'_i(\epsilon)$ representing the translational velocity and $\mathbf{z}'_i(\epsilon)$ related to the angular velocity.

From Theorem 1, for $i = 7, 8$, $\widehat{\mathbf{r}'_i(\epsilon)} = \widehat{\mathbf{T}'_i(\epsilon)}$ where $\mathbf{T}'_i(\epsilon)$ is orthogonal to the true translation $\mathbf{T}$ and

$$\|\mathbf{T}'_i(\epsilon)\| = \frac{1}{\sqrt{2}}.$$

Thus, the positivity hypothesis made by Lemma 1 on the first order term γ_i amounts to saying that when we substitute with a translation vector orthogonal to the true translation, the sum squared error must be greater than zero. This hypothesis must hold, otherwise the scene in view would be degenerate to the differential fundamental matrix. Therefore, using Lemma 1, we can say that

$$\lambda_i(\epsilon) = \epsilon^2 \Lambda_i + O(\epsilon^3), \quad \text{where } \Lambda_i > 0, 7 \leq i \leq 8. \tag{27}$$

Hence, in principle, under noiseless condition, there is no degeneracy in the solution to the eight point algorithm even under infinitesimal motion. The question of whether this non-degeneracy is sufficient to ensure stability under the proportional noise model is one which will investigate in the subsequent sections.

5 Eigenvalues of $A^T(\epsilon)A(\epsilon)$ under Noise

Having determined how the eigenvectors and eigenvalues of $A^T(\epsilon)A(\epsilon)$ vary with ϵ, we are now in a position to determine how they are affected by noise.

Let the corrupted data matrix $\widetilde{A}(\epsilon)$ be of the form

$$\widetilde{A}(\epsilon) = A(\epsilon) + M(\epsilon),$$

where $A(\epsilon)$ is defined in (15) and $\|M(\epsilon)\| \leq \epsilon m$ for sufficiently small ϵ. The matrix $M(\epsilon)$ represents the proportional noise model, and m is some proportionality factor which is a function of the percentage noise in the optical flow.

Now, the matrix $\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)$ is a perturbed version of $A^T(\epsilon)A(\epsilon)$ given by

$$\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon) = A^T(\epsilon)A(\epsilon) + B(\epsilon, M) \tag{28}$$

where

$$\begin{aligned} B(\epsilon, M) &= A^T(\epsilon)M(\epsilon) + M^T(\epsilon)A(\epsilon) + M^T(\epsilon)M(\epsilon) \\ &= O(\epsilon)m. \end{aligned} \tag{29}$$

The estimated solution is obtained by finding an eigenvector $\widetilde{\mathbf{q}}_9(\epsilon, M)$ of $\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)$ that corresponds to its smallest eigenvalue $\widetilde{\lambda}_9(\epsilon, M)$. Thus, we have

$$\begin{aligned} &(A^T(\epsilon)A(\epsilon) + B(\epsilon, M))\widetilde{\mathbf{q}}_9(\epsilon, M) \\ &= \widetilde{\lambda}_9(\epsilon, M)\widetilde{\mathbf{q}}_9(\epsilon, M). \end{aligned} \tag{30}$$

Note that both the eigenvalues $\widetilde{\lambda}_i(\epsilon, M)$ and eigenvectors $\widetilde{\mathbf{q}}_i(\epsilon, M)$ of $\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)$ are functions of ϵ and M. Henceforth, we rely on the $\widetilde{\ }$ notation to remind the reader of the dependence on M, suppressing M in these cases to keep our notation simple. However, in cases where the dependence on M is not clear, we will explicitly write down the dependence.

5.1 Eigenvalue $\widetilde{\lambda}_9(\epsilon)$

In this subsection, we shall determine the order of the eigenvalue $\widetilde{\lambda}_9(\epsilon)$ via the error $|\widetilde{\lambda}_9(\epsilon) - \lambda_9(\epsilon)|$, where $\lambda_9(\epsilon) = 0$. Specifically, we shall prove that $\widetilde{\lambda}_9(\epsilon) = O(\epsilon^2)m$ for a sufficiently small m.

Unfortunately, the standard results in perturbation theory only lead to $|\widetilde{\lambda}_i(\epsilon) - \lambda_i(\epsilon)| = O(\epsilon)m$ for each i, from which we are not able to deduce that $\widetilde{\lambda}_9(\epsilon)$ is simple since the three Gerschgorin's discs might overlap. To overcome this difficulty, we apply the techniques developed in Wilkinson (1965) and prove a modified result of Gerschgorin's Theorems, namely Proposition 8 in Appendix A.1. For readers not familiar with Gerschgorin's Theorems and the notion of Gerschgorin's discs, please refer to Theorems 5 and 6 in Appendix A.1.

Before we apply Proposition 8, let us record the following simple result which plays an important role in providing the order of eigenvalues and is also crucial for obtaining the projection $\alpha_i(\epsilon, M)$ of $\widetilde{\mathbf{q}}_9(\epsilon)$ along $\mathbf{q}_i(\epsilon)$ in Sect. 6.

Lemma 2

(a) *If either i or k is in the set $\{1, 2, 3, 4, 5, 6\}$, then*

$$\|\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)\| = O(\epsilon)m.$$

(b) *If both i and k are in the set $\{7, 8, 9\}$, then*

$$\|\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)\| = O(\epsilon^2)m.$$

Proof Part (a) is straightforward from (29). For part (b), we use $\mathbf{q}_i(\epsilon) = \mathbf{r}'_i(\epsilon) + \mathbf{z}_i(\epsilon)$ in Theorem 1, and the data matrix expression in (20) to obtain

$$\begin{aligned}
&\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)\\
&\quad = (\mathbf{r}'_k(\epsilon) + \mathbf{z}_k(\epsilon))^T(A_R^T M(\epsilon)\\
&\qquad + M^T(\epsilon)A_R)(\mathbf{r}'_i(\epsilon) + \mathbf{z}_i(\epsilon)) + O(\epsilon^2)m^2.
\end{aligned}$$

When $7 \le k \le 9$ and $7 \le i \le 9$, we have $\lambda_k = 0$ and $\lambda_i = 0$. Using Theorem 1, we have $A_R\mathbf{r}'_k(\epsilon) = \mathbf{0}$ and $A_R\mathbf{r}'_i(\epsilon) = \mathbf{0}$. Hence, we have

$$\begin{aligned}
&\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)\\
&\quad = (\mathbf{r}'_k(\epsilon) + \mathbf{z}_k(\epsilon))^T(A_R^T M(\epsilon)\\
&\qquad + M^T(\epsilon)A_R)(\mathbf{r}'_i(\epsilon) + \mathbf{z}_i(\epsilon)) + O(\epsilon^2)m^2\\
&\quad = \mathbf{z}_k^T(\epsilon)A_R^T M(\epsilon)\mathbf{r}'_i(\epsilon) + \mathbf{r}_k'^T(\epsilon)M^T(\epsilon)A_R\mathbf{z}_i(\epsilon)\\
&\qquad + \mathbf{z}_k^T(\epsilon)A_R^T M(\epsilon)\mathbf{z}_i(\epsilon) + O(\epsilon^2)m^2\\
&\quad = O(\epsilon^2)m. \qquad \square
\end{aligned}$$

If we can now specially construct an invertible matrix K such that the matrix $K^{-1}A^T(\epsilon)A(\epsilon)K$ becomes a diagonal matrix $\mathrm{Diag}(\lambda_i(\epsilon))$, then Proposition 8 provides an upper bound for $|\widetilde{\lambda}_i(\epsilon) - \lambda_i(\epsilon)|$. Our aim is to have upper bounds on $|\widetilde{\lambda}_i(\epsilon) - \lambda_i(\epsilon)|$ which enable us to isolate the 9th circular disc $\widetilde{G}_9$ from the other $\widetilde{G}_i$'s.

As mentioned above, the standard result from perturbation theory (Wilkinson 1965) is not adequate as there remains a possibility that the 9th circular disc $\widetilde{G}_9$ defined in Proposition 8 might overlap with other disc $\widetilde{G}_i$, in which case $\widetilde{\lambda}_9(\epsilon)$ would lie in the union of the discs. We need to choose K properly so that $\widetilde{G}_9$ is isolated from the other $\widetilde{G}_i$. We shall now work towards a suitable choice of an invertible matrix K.

First consider the matrix $Q^T(\epsilon)\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)Q(\epsilon)$ where $Q(\epsilon)$ is the matrix whose ith column is the unit eigenvector $\mathbf{q}_i(\epsilon)$ of the real symmetric matrix $A^T(\epsilon)A(\epsilon)$. Clearly, $Q^{-1}(\epsilon) = Q^T(\epsilon)$. From (28), we have

$$\begin{aligned}
&Q^T(\epsilon)\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)Q(\epsilon)\\
&\quad = \mathrm{Diag}(\lambda_i(\epsilon)) + Q^T(\epsilon)B(\epsilon, M)Q(\epsilon) \qquad (31)
\end{aligned}$$

where $\mathrm{Diag}(\lambda_i(\epsilon))$ is a diagonal matrix whose ith diagonal entry is $\lambda_i(\epsilon)$.

By Proposition 8, every eigenvalue $\widetilde{\lambda}_j(\epsilon)$ of $\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)$ lies in at least one of the circular discs with center $\lambda_k(\epsilon)$ and radius

$$\sum_{i=1}^{9} |\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)| = O(\epsilon)m.$$

(Note that Proposition 8 does not imply that j is necessarily equal to k.) Now, the center of the 9th circular disc is 0 while those of the 7th and 8th circular discs are $0 + O(\epsilon^2)$ (from (27)). However, all three circular discs have radii of order $O(\epsilon)$ by the preceding equation. Consequently, for a sufficiently small ϵ, these three discs may overlap with each other, and $\widetilde{\lambda}_9(\epsilon)$ lies in their union. As such, for this naive choice of $K = Q(\epsilon)$, we are not able to ascertain a good upper bound for $|\widetilde{\lambda}_9(\epsilon) - \lambda_9(\epsilon)|$.

Fortunately, by inspecting the entries in the matrix $Q(\epsilon)^T B(\epsilon, M)Q(\epsilon)$, we find that the upper bound on $|\widetilde{\lambda}_i(\epsilon) - \lambda_i(\epsilon)|$ can be improved by pre- and post-multiplying the matrix in (31) with the respective matrices $S^{-1}(\epsilon)$ and $S(\epsilon)$, where

$$S(\epsilon) = \begin{bmatrix} \epsilon I_6 & 0 \\ 0 & I_3 \end{bmatrix}.$$

Here, I_n denotes the $n \times n$ identity matrix, and 0 is a zero matrix. The effect of post-multiplying a matrix by S is the same as multiplying its first six columns by ϵ while pre-multiplying a matrix by S^{-1} is the same as multiplying its first six rows by $\frac{1}{\epsilon}$. So, we have

$$S^{-1}(\epsilon)Q^T(\epsilon)B(\epsilon, M)Q(\epsilon)S(\epsilon) = \left[\begin{array}{ccc|ccc}
\mathbf{q}_1^T(\epsilon)B\mathbf{q}_1(\epsilon) & \cdots & \mathbf{q}_1^T(\epsilon)B\mathbf{q}_6(\epsilon) & \frac{1}{\epsilon}\mathbf{q}_1^T(\epsilon)B\mathbf{q}_7(\epsilon) & \cdot & \frac{1}{\epsilon}\mathbf{q}_1^T(\epsilon)B\mathbf{q}_9(\epsilon)\\
\vdots & \vdots & \vdots & & \vdots & \\
\mathbf{q}_6^T(\epsilon)B\mathbf{q}_1(\epsilon) & \cdots & \mathbf{q}_6^T(\epsilon)B\mathbf{q}_6(\epsilon) & \frac{1}{\epsilon}\mathbf{q}_6^T(\epsilon)B\mathbf{q}_7(\epsilon) & \cdot & \frac{1}{\epsilon}\mathbf{q}_6^T(\epsilon)B\mathbf{q}_9(\epsilon)\\
\hline
\epsilon\mathbf{q}_7^T(\epsilon)B\mathbf{q}_1(\epsilon) & \cdots & \epsilon\mathbf{q}_7^T(\epsilon)B\mathbf{q}_6(\epsilon) & \mathbf{q}_7^T(\epsilon)B\mathbf{q}_7(\epsilon) & \cdot & \mathbf{q}_7^T(\epsilon)B\mathbf{q}_9(\epsilon)\\
\cdots & \cdots & \cdots & & \cdots & \\
\epsilon\mathbf{q}_9^T(\epsilon)B\mathbf{q}_1(\epsilon) & \cdots & \epsilon\mathbf{q}_9^T(\epsilon)B\mathbf{q}_6(\epsilon) & \mathbf{q}_9^T(\epsilon)B\mathbf{q}_7(\epsilon) & \cdot & \mathbf{q}_9^T(\epsilon)B\mathbf{q}_9(\epsilon)
\end{array}\right]$$

in which the diagonal sub-matrices of the above matrix $S^{-1}(\epsilon)Q^T(\epsilon)B(\epsilon, M)Q(\epsilon)S(\epsilon)$ remain the same as those of $Q^T(\epsilon)B(\epsilon, M)Q(\epsilon)$.

Note that the above transformation does not affect the eigenvalues. Thus, we have

$$S^{-1}(\epsilon)Q^T(\epsilon)\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)Q(\epsilon)S(\epsilon) = \mathrm{Diag}(\lambda_i(\epsilon)) + S^{-1}(\epsilon)Q^T(\epsilon)B(\epsilon, M)Q(\epsilon)S(\epsilon).$$

By Proposition 8, where $K = Q(\epsilon)S(\epsilon)$, every eigenvalue of $\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)$ lies in at least one of the circular discs $\widetilde{G}_k$ with center $\lambda_k(\epsilon)$ and radius

$$d_k(\epsilon, M) = \begin{cases} \sum_{i=1}^{6} |\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)| \\ \quad + \frac{1}{\epsilon}\sum_{i=7}^{9} |\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)| \\ \quad = O(\epsilon)m, \quad 1 \le k \le 6, \\ \epsilon \sum_{i=1}^{6} |\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)| \\ \quad + \sum_{i=7}^{9} |\mathbf{q}_k^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)| \\ \quad = O(\epsilon^2)m, \quad 7 \le k \le 9. \end{cases} \tag{32}$$

Using the above, we may now prove that the desired property that the 9th circular disc $\widetilde{G}_9$ is disjoint from the rest.

Proposition 5 *For a sufficiently small ϵ and m, the 9th circular disc $\widetilde{G}_9$ is disjoint from the union $\widetilde{G}_7 \cup \widetilde{G}_8$ of the 7th and 8th circular discs, which in turn is disjoint from the union $\bigcup_{i=1}^{6} \widetilde{G}_i$ of the first six circular discs.*

Proof First, the 9th circular disc $\widetilde{G}_9$ is disjoint from the union $\widetilde{G}_7 \cup \widetilde{G}_8$ of the 7th and 8th circular discs if the gap $(\Lambda_8\epsilon^2 + O(\epsilon^3))$ between the disc centers $\lambda_8(\epsilon)$ and $\lambda_9(\epsilon)$ satisfies the following:

$$\Lambda_8\epsilon^2 + O(\epsilon^3) > d_9(\epsilon, M) + \max(d_7(\epsilon, M), d_8(\epsilon, M)). \tag{33}$$

For a sufficiently small ϵ such that the left hand side is more than $\frac{1}{2}\Lambda_8\epsilon^2$, and the right hand side is $O(\epsilon^2)m$, by (32), we have a uniform bound on m (independent of ϵ) such that for a sufficiently small m, the above condition (33) is satisfied.

Next, the union $\bigcup_{i=1}^{6} \widetilde{G}_i$ of the first six circular discs is disjoint from the union $\widetilde{G}_7 \cup \widetilde{G}_8$ of the 7th and 8th circular discs, if the gap between the two nearest disc centers $\lambda_6(\epsilon)$ and $\lambda_7(\epsilon)$ satisfies the following:

$$\lambda_6(\epsilon) - \lambda_7(\epsilon) > \max_{j=7,8}(d_j(\epsilon, M)) + \max_{1\le k\le 6} d_k(\epsilon, M).$$

From Proposition 4, we note that

$$\lambda_k(\epsilon) = \lambda_k + O(\epsilon) \quad \text{where } \lambda_k > 0, \quad \text{for } 1 \le k \le 6,$$

while under a non-degenerate scene, (27) holds:

$$\lambda_j(\epsilon) = \Lambda_j\epsilon^2 + O(\epsilon^3) \quad \text{where } \Lambda_j > 0 \quad \text{for } 7 \le j \le 8.$$

Thus the above condition is satisfied if

$$\lambda_6 - \Lambda_7\epsilon^2 + O(\epsilon) > \max_{j=7,8}(d_j(\epsilon, M)) + \max_{1\le k\le 6} d_k(\epsilon, M). \tag{34}$$

However, from (32), we have

$$\max_{j=7,8}(d_j(\epsilon, M)) = O(\epsilon^2)m,$$

and

$$\max_{1\le k\le 6} d_k(\epsilon, M) = O(\epsilon)m.$$

Thus, since $\lambda_6 > 0$, the condition (34) is satisfied for a sufficiently small ϵ when m is sufficiently small (i.e., when there is a sufficiently small percentage noise).

Therefore, for a sufficiently small m, and a sufficiently small ϵ, the 9th circular disc $\widetilde{G}_9$ is disjoint from the union $\widetilde{G}_7 \cup \widetilde{G}_8$ of the 7th and 8th circular discs, which in turn is disjoint from the union $\bigcup_{i=1}^{6} \widetilde{G}_i$ of the first six circular discs. □

It follows from the second part of Proposition 8 that $\widetilde{\lambda}_9(\epsilon)$ lies in $\widetilde{G}_9$, and from (32), we record the following result:

Theorem 2 *For a sufficiently small m, and a sufficiently small ϵ, the eigenvalue $\widetilde{\lambda}_9(\epsilon)$ is simple and*

$$\widetilde{\lambda}_9(\epsilon) = O(\epsilon^2)m.$$

Moreover,

$$\begin{aligned} &\widetilde{\lambda}_i(\epsilon) = \Lambda_i\epsilon^2 + O(\epsilon^2)m, \quad i = 7, 8; \\ &\widetilde{\lambda}_i(\epsilon) = \lambda_i + O(\epsilon) + O(\epsilon)m, \quad i = 1, 2, 3, 4, 5, 6. \end{aligned} \tag{35}$$

6 Projection of $\widetilde{\mathbf{q}}_9(\epsilon)$ along $\mathbf{q}_k(\epsilon)$

From the preceding section, when m is small, $\widetilde{\lambda}_9(\epsilon)$ is simple for sufficiently small ϵ. Therefore, its corresponding eigen-space is 1-dimensional. Let $\widetilde{\mathbf{q}}_9(\epsilon)$ (which may not be a unit vector) be an eigenvector corresponding to the eigenvalue $\widetilde{\lambda}_9(\epsilon)$ and expressed in the form

$$\widetilde{\mathbf{q}}_9(\epsilon) = \sum_{i=1}^{9} \alpha_i(\epsilon, M)\mathbf{q}_i(\epsilon). \tag{36}$$

Then the perturbation introduced to $\widetilde{\mathbf{q}}_9(\epsilon)$ can be analyzed by looking at the projection coefficients $\alpha_i(\epsilon, M)$ using the same technique in Wilkinson (1965).

The following result is simple yet useful in the sequel.

Lemma 3

$$(\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon))\alpha_j(\epsilon, M)$$
$$= \sum_{i=1}^{9} \alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon). \quad (37)$$

Proof Substituting (36) into (30), and using

$$A^T(\epsilon)A(\epsilon)\mathbf{q}_i(\epsilon) = \lambda_i(\epsilon)\mathbf{q}_i(\epsilon),$$

we have

$$\sum_{i=1}^{9} \alpha_i(\epsilon, M)\lambda_i(\epsilon)\mathbf{q}_i(\epsilon) + \sum_{i=1}^{9} \alpha_i(\epsilon, M)B(\epsilon, M)\mathbf{q}_i(\epsilon)$$
$$= \widetilde{\lambda}_9(\epsilon)\left(\sum_{i=1}^{9} \alpha_i(\epsilon, M)\mathbf{q}_i(\epsilon)\right).$$

Pre-multiplying the above equation with $\mathbf{q}_j^T(\epsilon)$, we obtain the required relation (37) by noting that

$$\mathbf{q}_j^T(\epsilon)\mathbf{q}_i(\epsilon) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$ □

Lemma 4 *Suppose the maximum projection is given by*

$$\max\{|\alpha_i(\epsilon, M)| \mid 1 \leq i \leq 9\} = |\alpha_{i_*}(\epsilon, M)|$$

for some i_* *in* $1 \leq i \leq 9$. *Then*

$$|\widetilde{\lambda}_9(\epsilon) - \lambda_{i_*}(\epsilon)| \leq \sum_{i=1}^{9} |\mathbf{q}_j^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)| = O(\epsilon)m.$$

Proof Dividing (37) by $\alpha_{i_*}(\epsilon, M)$ yields

$$(\widetilde{\lambda}_9(\epsilon) - \lambda_{i_*}(\epsilon)) = \sum_{i=1}^{9} \frac{\alpha_i(\epsilon, M)}{\alpha_{i_*}(\epsilon, M)}\mathbf{q}_j^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)$$

with $|\frac{\alpha_i(\epsilon,M)}{\alpha_{i_*}(\epsilon,M)}| \leq 1$ for $1 \leq i \leq 9$. Thus we have

$$|\widetilde{\lambda}_9(\epsilon) - \lambda_{i_*}(\epsilon)| \leq \sum_{i=1}^{9} \left|\frac{\alpha_i(\epsilon, M)}{\alpha_{i_*}(\epsilon, M)}\mathbf{q}_j^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)\right|$$
$$\leq \sum_{i=1}^{9} |\mathbf{q}_j^T(\epsilon)B(\epsilon, M)\mathbf{q}_i(\epsilon)| = O(\epsilon)m.$$

The order follows from Lemma 2. □

Theorem 3 *For a sufficiently small noise* m *(and hence* M*) and a sufficiently small* ϵ*, the maximum projection is given by*

$$\max\{|\alpha_i(\epsilon, M)| \mid 1 \leq i \leq 9\} = |\alpha_9(\epsilon, M)|.$$

Proof We first prove that for every j in $1 \leq j \leq 6$,

$$|\alpha_j(\epsilon, M)| \neq \max\{|\alpha_i(\epsilon, M)| \mid 1 \leq i \leq 9\}$$

for a sufficiently small ϵ (for a given M).

Suppose on the contrary that, for some j_* in $1 \leq j \leq 6$, there is a sequence $\{\epsilon_s\}$ with $\lim_{s\to\infty} \epsilon_s = 0$ and $|\alpha_{j_*}(\epsilon_s, M)| = \max\{|\alpha_i(\epsilon_s, M)| \mid 1 \leq i \leq 9\}$.

By Lemma 4, we have

$$|\widetilde{\lambda}_9(\epsilon_s) - \lambda_{j_*}(\epsilon_s)| \leq \sum_{i=1}^{9} |\mathbf{q}_j^T(\epsilon_s)B(\epsilon_s, M)\mathbf{q}_i(\epsilon_s)|$$

As $s \to \infty$, we note that the right hand side of the above inequality approaches 0, by Lemma 4. Using Theorem 2 and Proposition 4, the left hand side approaches λ_{j_*}, which is positive. This yields a contradiction.

Therefore, $|\alpha_j(\epsilon, M)|$ is non-maximal for $1 \leq j \leq 6$, when ϵ is sufficiently small.

Now, we shall prove, again by contradiction, that for every j where $7 \leq j \leq 8$,

$$|\alpha_j(\epsilon, M)| \neq \max\{|\alpha_i(\epsilon, M_s)| \mid 1 \leq i \leq 9\}$$

for a sufficiently small M and a sufficiently small ϵ.

Suppose for some $j_* \in \{7, 8\}$, there is a sequence $\{M_s\}$ with $\lim_{s\to\infty} \|M_s\| = 0$ and $|\alpha_{j_*}(\epsilon, M_s)| = \max\{|\alpha_i(\epsilon, M_s)| \mid 1 \leq i \leq 9\}$. By Lemma 4, we have

$$|\widetilde{\lambda}_9(\epsilon) - \lambda_{j_*}(\epsilon)| \leq \sum_{i=1}^{9} |\mathbf{q}_j^T(\epsilon)B(\epsilon, M_s)\mathbf{q}_i(\epsilon)|.$$

As $s \to \infty$, we note that the right hand side approaches 0, since $\lim_{s\to\infty} \|M_s\| = 0$. The left hand side approaches $\Lambda_{j_*}\epsilon^2$, by Theorem 2 and Proposition 1. However, $\Lambda_{j_*}\epsilon^2$ is positive for a sufficiently small $\epsilon > 0$. This yields a contradiction.

Therefore, for a sufficiently small m (and hence M) and a sufficiently small ϵ, the projection $|\alpha_j(\epsilon, M)|$, for $7 \leq j \leq 8$, is non-maximal.

We conclude that for a sufficiently small m and a sufficiently small ϵ, the projection $|\alpha_9(\epsilon, M)|$ is maximal. □

Note that the following vector

$$\frac{1}{\alpha_9(\epsilon, M)}\widetilde{\mathbf{q}}_9(\epsilon) = \sum_{i=1}^{9} \frac{\alpha_i(\epsilon, M)}{\alpha_9(\epsilon, M)}\mathbf{q}_i(\epsilon)$$

is an eigenvector of $\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)$ corresponding to the eigenvalue $\widetilde{\lambda}_9(\epsilon)$ with $|\frac{\alpha_i(\epsilon,M)}{\alpha_9(\epsilon,M)}| \leq 1$ for $1 \leq i \leq 8$.

Thus, from now on, for a sufficiently small ϵ and m, we may assume that

$$\widetilde{\mathbf{q}}_9(\epsilon) = \sum_{i=1}^{9} \alpha_i(\epsilon, M)\mathbf{q}_i(\epsilon), \quad (38)$$

where $\alpha_9(\epsilon, M) = 1$ and $|\alpha_i(\epsilon, M)| \le 1$ for $1 \le i \le 8$. Note that with $\alpha_9(\epsilon, M) = \alpha_{i*}(\epsilon, M) = 1$, both Lemmas 3 and 4 still apply. We shall now proceed to determine the upper bounds on $\alpha_i(\epsilon, M)$ for $1 \le i \le 8$.

Proposition 6 *For $1 \le j \le 6$, for sufficiently small m and ϵ, we have*

$$\alpha_j(\epsilon, M) = O(\epsilon)m.$$

Proof From Lemma 3, we have

$$|\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon)|\,|\alpha_j(\epsilon, M)| \le \sum_{i=1}^{9} |\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)|,$$

as each $|\alpha_i(\epsilon, M)| \le 1$. By Theorem 2 and Proposition 1, we have $|\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon)| = \lambda_j + O(\epsilon)$ while

$$\sum_{i=1}^{9} |\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)| = O(\epsilon)m,$$

by Lemma 2. Therefore, we have

$$\alpha_j(\epsilon, M) = O(\epsilon)m$$

for a sufficiently small ϵ and m. □

Lemma 5 *For $j = 7$ or 8, when m is sufficiently small, we have*

$$\left| \sum_{i=1}^{9} \alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon) \right| = O(\epsilon^2)m.$$

Proof Note that

$$\begin{aligned} &\sum_{i=1}^{9} |\alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)| \\ &\quad \le \sum_{i=1}^{6} |\alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)| \\ &\qquad + \sum_{i=7}^{9} |\alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)|. \end{aligned}$$

For $7 \le i \le 9$, we use $|\alpha_i(\epsilon, M)| \le 1$ and Lemma 2 to obtain

$$\sum_{i=7}^{9} |\alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)| = O(\epsilon^2)m.$$

For $1 \le i \le 6$, by Proposition 6 and Lemma 2, we have

$$|\alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)| = O(\epsilon^2)m^2.$$

Therefore,

$$\sum_{i=1}^{9} |\alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)| = O(\epsilon^2)m.$$

□

We have the following bound on $\alpha_j(\epsilon, M)$ for $j = 7$ or 8.

Proposition 7 *For $7 \le j \le 8$, for sufficiently small m and ϵ, we have*

$$|\alpha_j(\epsilon, M)| = O(\epsilon^0)m.$$

Proof Using Lemma 3, we obtain

$$|\alpha_j(\epsilon, M)| \le \frac{\sum_{i=1}^{9} |\alpha_i(\epsilon, M)\mathbf{q}_j^T(\epsilon) B(\epsilon, M)\mathbf{q}_i(\epsilon)|}{|\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon)|}.$$

By Lemma 1 and Theorem 2, we have, for sufficiently small ϵ and m,

$$\begin{aligned} |\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon)| &= \Lambda_j\epsilon^2 + O(\epsilon^3)m \\ &= \Lambda_j\epsilon^2(1 + O(\epsilon)m) > 0, \end{aligned}$$

where $\Lambda_j > 0$. Therefore, together with Lemma 5, we conclude that $|\alpha_j(\epsilon, M)| = O(\epsilon^0)m$. □

As a consequence of the preceding proposition, we are able to set a uniform bound on m, i.e., independent of ϵ, such that $|\alpha_j(\epsilon, M)| < 1$ (and are as small as we like) for $7 \le j \le 8$. With this m, for a sufficiently small $\epsilon > 0$, we will also have $|\alpha_j(\epsilon, M)| < 1$ (and are as small as we like) for $1 \le j \le 6$. The above result proves that stability of the estimated solution $\widetilde{\mathbf{q}}_9(\epsilon)$ under the differential condition of small motion and a bounded percentage noise.

For comparison with other statistical analysis, we would like to find an explicit expression for the lowest order noise terms (i.e., the m terms) of $\alpha_j(\epsilon, M)$, via (37) in conjunction with the results obtained in Propositions 6 and 7 and Lemma 2. We shall state and prove the result in our last theorem:

Theorem 4 *Given a sufficiently small m, and a sufficiently small ϵ,*

$$\widetilde{\mathbf{q}}_9(\epsilon) = \sum_{i=1}^{9} \alpha_i(\epsilon, M)\mathbf{q}_i(\epsilon)$$

where $\alpha_9(\epsilon, M) = 1$,

$$\begin{aligned} \alpha_i(\epsilon, M) &= -\frac{\mathbf{q}_i^T(\epsilon) B(\epsilon, M)\mathbf{q}_9(\epsilon)}{\lambda_i(\epsilon)} + O(\epsilon)m^2 \\ &= O(\epsilon)m, \quad \textit{for } 1 \le i \le 6, \\ \alpha_i(\epsilon, M) &= -\frac{\mathbf{q}_i^T(\epsilon) B(\epsilon, M)\mathbf{q}_9(\epsilon)}{\lambda_i(\epsilon)} + O(\epsilon^0)m^2 \\ &= O(\epsilon^0)m, \quad \textit{for } 7 \le i \le 8. \end{aligned}$$

Proof From (37) and (38) and Propositions 6 and 7, we have

$$(\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon))\alpha_j(\epsilon, M)$$

$$
\begin{aligned}
&= \mathbf{q}_j^T(\epsilon) B(\epsilon, M) \mathbf{q}_9(\epsilon) \\
&\quad + \sum_{i=1}^{8} \alpha_i(\epsilon, M) \mathbf{q}_j^T(\epsilon) B(\epsilon, M) \mathbf{q}_i(\epsilon) \\
&= \begin{cases} \mathbf{q}_j^T(\epsilon) B(\epsilon, M) \mathbf{q}_9(\epsilon) + O(\epsilon) m^2, & 1 \le j \le 6, \\ \mathbf{q}_j^T(\epsilon) B(\epsilon, M) \mathbf{q}_9(\epsilon) + O(\epsilon^2) m^2, & 7 \le j \le 8. \end{cases}
\end{aligned} \tag{39}
$$

Applying Proposition 4 and Lemma 1, we have

$$
\begin{aligned}
\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon) &= -\lambda_j(\epsilon)\left(1 - \frac{\widetilde{\lambda}_9(\epsilon)}{\lambda_j(\epsilon)}\right) \\
&= \begin{cases} -\lambda_j(\epsilon)(1 + O(\epsilon^2) m), & \text{if } 1 \le j \le 6, \\ -\lambda_j(\epsilon)(1 + O(\epsilon^0) m), & \text{if } 7 \le j \le 8. \end{cases}
\end{aligned}
$$

Hence, by (3), we have

$$
\frac{1}{\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon)} = \begin{cases} \frac{1}{-\lambda_j(\epsilon)}(1 + O(\epsilon^2) m), & \text{if } 1 \le j \le 6, \\ \frac{1}{-\lambda_j(\epsilon)}(1 + O(\epsilon^0) m), & \text{if } 7 \le j \le 8. \end{cases}
$$

Dividing (39) throughout by $\widetilde{\lambda}_9(\epsilon) - \lambda_j(\epsilon)$ and using Lemma 2, we have the desired expression stated in the theorem. □

The above result shows that the lowest order terms in noise are the same as those derived in Wilkinson (1965), pages 70, 71, 83, where the noise is denoted by ϵ. It allows us to extend much of the unbiasedness/noise whitening analysis carried out on the discrete eight point algorithm to the differential case, because the foundation of such analysis is the lowest order noise terms in the perturbation analysis. One example is the work of Muhlich and Mester (1998), which showed that for the so-called TLS-FC normalized variant of the discrete eight point algorithm,[1] the expected value of the $\alpha_i(\epsilon, M)$ is zero for $i \neq 9$, that is, the estimated solution is unbiased.

7 Obtaining the Rotation and Translation Parameters

To complete our investigation, we need to ensure that the subsequent decomposition of the fundamental matrix $\widetilde{F}(\epsilon)$ into the translation and the rotation estimates are stable. Several intermediate steps are also involved, including the correcting of $\widetilde{F}(\epsilon)$ to the nearest rank-two matrix, and the correcting of the recovered essential matrix $\widetilde{E}(\epsilon)$ to the nearest matrix with the desired property of having the first two singular values being equal.

In addition, to establish a proper comparison between the discrete and the differential formulation, we need to convert the rotational and translational displacements recovered from the discrete algorithm into the corresponding velocity formulations. As mentioned previously, the differential two view formulation converts the SFM problem into one independent of ϵ but involving differential entities like velocity. Accordingly, the required orders in the errors of the discrete estimates so that the corresponding velocity estimates have errors independent of ϵ are given by:

$$
\begin{aligned}
&\text{Error in the translation direction} = O(\epsilon^0) m, \\
&\text{Error in the rotation estimate} = O(\epsilon) m.
\end{aligned} \tag{40}
$$

The above means that when the discrete rotation estimate is divided by the time ϵ to get the rotational velocity, the latter's error would be independent of ϵ. The error in the translation estimate only needs $O(\epsilon^0) m$ instead of $O(\epsilon) m$ because the terms $\mathbf{T}_c$ and $\mathbf{T}$ in our formulation in fact represent velocities already (see (5) and (19)). This is related to the fact that we can only recover the translation direction anyway.

The overall order of the error provided in Theorem 4 is $O(\epsilon^0) m$ and superficially, does not give us much hope that the discrete eight point algorithm can meet the condition set out in (40) for rotation, which requires error of the order $O(\epsilon) m$. Fortunately, Theorem 4 also shows that the orders of the perturbation coefficients $\alpha_i(\epsilon, M)'s$ are not all equal. In fact, only two coefficients, $\alpha_7(\epsilon, M)$ and $\alpha_8(\epsilon, M)$, are of order $O(\epsilon^0) m$, whereas the rest are of order $O(\epsilon) m$.

We can obtain a better bound if we split the recovered fundamental matrix $\widetilde{F}(\epsilon) = \widetilde{\mathbf{q}_9(\epsilon)}$ into a sum of two terms, such that the large $O(\epsilon^0) m$ noise only perturbs the translation vector. The rest of this section and Appendix A.2 are devoted to doing just such a split and keeping track of how the errors are propagated and apportioned in the subsequent decomposition into the translation and rotation estimates.

7.1 Some Preliminaries

Before proceeding further, the following short note on 'nearest matrix' will be used extensively in the discussion of the various errors throughout this section.

Let $\widetilde{C}(\epsilon)$ be the noise corrupted version of a matrix $C(\epsilon)$. Due to the noise, $\widetilde{C}(\epsilon)$ may lack some desired properties which are present in $C(\epsilon)$ (an example of such property is

[1] In the TLS-FC variant (Muhlich and Mester 1998), matrix perturbation analysis was used to formulate a new data matrix $\widetilde{A}'(\epsilon)$ given by $\widetilde{A}'(\epsilon) = \widetilde{A}(\epsilon) + M'(\epsilon) = A(\epsilon) + \epsilon M + M'(\epsilon)$, where $M'(\epsilon)$ is chosen such that $\widetilde{A}'(\epsilon)$ satisfies the rank 8 constraint without making any changes to the columns of $\widetilde{A}(\epsilon)$ in which noise is not present, and that $\|M'(\epsilon)\|$ is minimized. A straightforward application of the result in (41) in Sect. 7 shows that this new $\widetilde{A}'(\epsilon)$ is bounded by the same proportional noise regime, and thus the results from our paper are applicable to this TLS-FC variant.

that the first two singular values are identical or the rank is 2).

As is often the case, we use the 'nearest' $\widetilde{C}'(\epsilon)$ to $\widetilde{C}(\epsilon)$ (if it exists) instead of $\widetilde{C}(\epsilon)$ in the following sense:

(a) $\widetilde{C}'(\epsilon)$ possesses the desired properties, and
(b) $\|\widetilde{C}'(\epsilon) - \widetilde{C}(\epsilon)\| = \min\{\|K - \widetilde{C}(\epsilon)\| \mid K$ possesses the desired properties$\}$.

Thus, we have,

$$\|\widetilde{C}'(\epsilon) - \widetilde{C}(\epsilon)\| \leq \|C(\epsilon) - \widetilde{C}(\epsilon)\|. \tag{41}$$

This ensures that in using the nearest matrix, the 'correction' introduced has the same order of error.

Note that the essential/fundamental matrix is only defined up to a scale factor. We regard the 'true' fundamental matrix $F_t(\epsilon)$ as one having unit Frobenius norm and given by

$$F_t(\epsilon) = \widehat{\mathbf{q}_9(\epsilon)} = \widehat{\mathbf{T}}_t \Theta R^T(\epsilon)(\Theta(\epsilon))^{-1} \tag{42}$$

where $\mathbf{T}_t$ is parallel to $\mathbf{T}$ defined in (19) but scaled such that $\|F_t(\epsilon)\| = 1$. The true essential matrix $E_t(\epsilon)$ is defined as the de-normalized version of $F_t(\epsilon)$,

$$E_t(\epsilon) = \Theta^T F_t(\epsilon)\Theta(\epsilon) = \Theta^T \widehat{\mathbf{T}}_t \Theta R^T(\epsilon). \tag{43}$$

The estimated fundamental matrix is given by

$$\widetilde{F}(\epsilon) = \widehat{\widetilde{\mathbf{q}}_9(\epsilon)}.$$

With noise, $\widehat{\widetilde{\mathbf{q}}_9(\epsilon)}$ may not be a unit vector and thus may not have unit norm. Letting $\widetilde{F}(\epsilon)$ stay un-normalized has the virtue of keeping the following proof simple while still obtaining error expressions that suffice for our purpose.

7.2 Splitting the Fundamental Matrix

We know from Theorems 1 and 4 that our estimated solution vector $\widetilde{\mathbf{q}}_9(\epsilon)$ can be expressed as

$$\begin{aligned}\widetilde{\mathbf{q}}_9(\epsilon) &= \mathbf{q}_9(\epsilon) + \sum_{i=7}^{8} \alpha_i(\epsilon, M)\mathbf{r}'_i(\epsilon) + \sum_{i=7}^{8} \alpha_i(\epsilon, M)\mathbf{z}_i(\epsilon) \\ &\quad + \sum_{i=1}^{6} \alpha_i(\epsilon, M)\mathbf{q}_i(\epsilon) \\ &= \mathbf{q}_9(\epsilon) + \sum_{i=7}^{8} \alpha_i(\epsilon, M)\mathbf{r}'_i(\epsilon) + O(\epsilon)m,\end{aligned}$$

where

$$\widehat{\mathbf{r}'_i(\epsilon)} = \widehat{\mathbf{T}'_i(\epsilon)}, \quad 7 \leq i \leq 8.$$

Therefore, using the definition of $F_t(\epsilon)$ in (42), we have

$$\begin{aligned}\widetilde{F}(\epsilon) &= \widehat{\widetilde{\mathbf{q}}_9(\epsilon)} \\ &= F_t(\epsilon) + \sum_{i=7}^{8} \alpha_i(\epsilon, M)\widehat{\mathbf{T}'_i(\epsilon)} + O(\epsilon)m.\end{aligned} \tag{44}$$

Utilizing the relation $\Theta R^T(\epsilon)(\Theta(\epsilon))^{-1} - I = O(\epsilon)$, we can modify (44) such that

$$\begin{aligned}\widetilde{F}(\epsilon) &= F_t(\epsilon) + \sum_{i=7}^{8} \alpha_i(\epsilon, M)\widehat{\mathbf{T}'_i(\epsilon)}\{\Theta R^T(\epsilon)(\Theta(\epsilon))^{-1} \\ &\quad + O(\epsilon)\} + O(\epsilon)m \\ &= F_a(\epsilon, M) + O(\epsilon)m\end{aligned}$$

where

$$\begin{aligned}F_a(\epsilon, M) &= F_t(\epsilon) + \left(\sum_{i=7}^{8} \alpha_i(\epsilon, M)\widehat{\mathbf{T}'_i(\epsilon)}\right) \\ &\quad \times \Theta R^T(\epsilon)(\Theta(\epsilon))^{-1} \\ &= F_t(\epsilon) + O(\epsilon^0)m\end{aligned} \tag{45}$$

is a part of $\widetilde{F}(\epsilon)$ that contains the true rotation but an incorrect translation. As $\widetilde{F}(\epsilon)$ may lack the rank two property associated with a fundamental matrix, we apply the algorithm described in Ma et al. (2003) that chooses a rank 2 matrix $\widetilde{F}'(\epsilon)$ with the minimum $\|\widetilde{F}(\epsilon) - \widetilde{F}'(\epsilon)\|$. If we consider $\widetilde{F}(\epsilon)$ to be a perturbed version of the valid fundamental matrix $F_a(\epsilon, M)$ (i.e., having rank 2), then from (41),

$$\|\widetilde{F}(\epsilon) - \widetilde{F}'(\epsilon)\| \leq \|\widetilde{F}(\epsilon) - F_a(\epsilon, M)\| = O(\epsilon)m.$$

Hence, the error in $\widetilde{F}'(\epsilon)$ takes the form

$$\widetilde{F}'(\epsilon) = F_a(\epsilon, M) + O(\epsilon)m.$$

7.3 Errors in the Motion Estimates

The essential matrix $\widetilde{E}(\epsilon)$ is obtained by de-normalizing $\widetilde{F}'(\epsilon)$:

$$\widetilde{E}(\epsilon) = \Theta^T \widetilde{F}'(\epsilon)\Theta(\epsilon) = E_a(\epsilon, M) + O(\epsilon)m \tag{46}$$

where using (2), (43) and (45), we have

$$\begin{aligned}&E_a(\epsilon, M) \\ &= \Theta^T F_a(\epsilon, M)\Theta(\epsilon) \\ &= \left(\Theta^T \widehat{\mathbf{T}}_t \Theta + \det(\Theta)\sum_{i=7}^{8} \alpha_i(\epsilon, M)\widehat{\Theta^{-1}\mathbf{T}'_i(\epsilon)}\right) R^T(\epsilon) \\ &= E_t(\epsilon) + O(\epsilon^0)m.\end{aligned} \tag{47}$$

Observe that from (47), $E_a(\epsilon)$ is a valid essential matrix (in the sense that it has rank 2 and two identical non-zero singular values), since it is a product of a skew symmetric matrix and a rotation matrix $R(\epsilon)$ in $SO(3)$.

We can treat $\widetilde{E}(\epsilon)$ as a perturbed version of $E_a(\epsilon)$. Therefore, using the algorithm in Ma et al. (2003) to enforce on $\widetilde{E}(\epsilon)$ the condition of having the first two singular values being equal, we can obtain a valid essential matrix $\widetilde{E}'(\epsilon)$, where from (41) and (47), we have

$$\|\widetilde{E}'(\epsilon) - \widetilde{E}(\epsilon)\| \leq \|E_a(\epsilon, M) - \widetilde{E}(\epsilon)\| = O(\epsilon)m. \tag{48}$$

Using triangle inequality and the orders from (46) and (48), we have

$$\begin{aligned}&\|\widetilde{E}'(\epsilon) - E_a(\epsilon, M)\| \\ &\quad \leq \|\widetilde{E}'(\epsilon) - \widetilde{E}(\epsilon)\| + \|E_a(\epsilon, M) - \widetilde{E}(\epsilon)\| \\ &\quad = O(\epsilon)m.\end{aligned} \tag{49}$$

Similarly, using the orders from (46), (47) and (48), we have

$$\begin{aligned}&\|\widetilde{E}'(\epsilon) - E_t(\epsilon, M)\| \\ &\quad \leq \|\widetilde{E}'(\epsilon) - \widetilde{E}(\epsilon)\| + \|E_t(\epsilon, M) - \widetilde{E}(\epsilon)\| \\ &\quad = O(\epsilon^0)m.\end{aligned} \tag{50}$$

The rest of the proof basically keeps track of how the errors are propagated when one uses singular value decomposition on $\widetilde{E}'(\epsilon)$ to obtain the rotation and translation estimates. The steps are nontrivial but the arguments are straightforward. Interested readers can refer to Appendix A.2 for the details. In particular, using the order in (50) and Proposition 9 in Appendix A.2, we obtain:

$$\text{Error in the unit translational vector} = O(\epsilon^0)m.$$

With regards to rotation, if one considers $\widetilde{E}'(\epsilon)$ to be a perturbed version of $E_a(\epsilon)$ which contains the true rotation, then using the order in (49) and Proposition 10 in Appendix A.2, we obtain:

$$\text{Error in the rotational matrix} = O(\epsilon)m,$$

completing the requirements set out in (40).

8 Simulation Results

We present simulation results for the following linear algorithms:

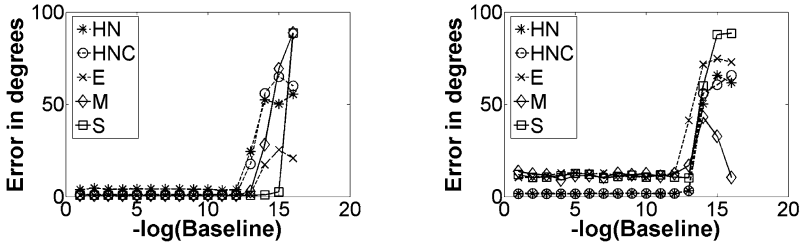
HN denotes the eight point algorithm using Hartley normalization (Hartley, 1997),
HNC denotes the eight point algorithm with Hartley normalization and estimated by Total Least Squares—Fixed Column (TLS-FC) (Muhlich and Mester 1998),
E denotes the un-normalized eight point algorithm (Longuet-Higgins 1981),
M denotes the differential essential matrix (Ma et al. 2000), and
S denotes the linear subspace differential algorithm (Heeger et al. 1992).

8.1 Decreasing Baseline

Simulation results for decreasing baseline are given in Figs. 1 and 2, with those of the discrete algorithms represented by dotted lines and those of the differential algorithms by solid lines. The simulation conditions are as follows. The "scene" consisted of a point cloud containing 1000 points with an average depth of 10 units. The points were uniformly distributed between depths of 7 and 13 units. The simulated camera had a 45° field of view (FoV) with a focal length of 1 unit. The initial translation was set at $(0, 0.1, 0)$ unit, and the initial rotation at $(0.01, 0, 0.01)$ radians. Both the baseline and rotation were steadily decreased by factors of 10 to simulate increasingly small motion. This baseline was decreased until 10^{-16}, the limit of arithmetic precision, in order to verify our theoretical prediction. The optical flow noise was 3.5% of the average magnitude of the optical flow. The rotational errors presented in Fig. 2 have been normalized such that

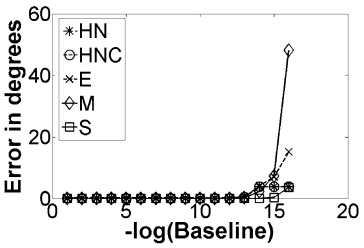
$$\text{normalized rotational error} = \frac{\text{rotational error in degrees}}{\text{baseline}}.$$

As such, a constant normalized rotation error in the graphs indicates that the actual error is decreasing proportionally to the amount moved by the camera.

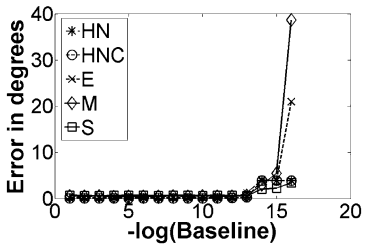


(a) Translation Direction=(0, 0, 1) (b) Translation Direction=(1, 0, 0)

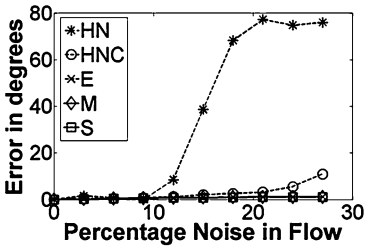
Fig. 1 Error in estimating the translation direction with decreasing baseline. For lateral translation (**b**), the errors in E, M, and S are large



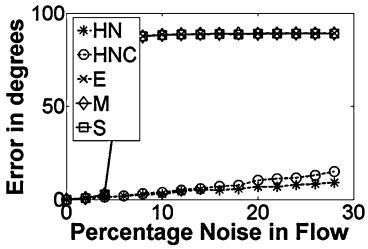
(a) Translation Direction=(0, 0, 1)



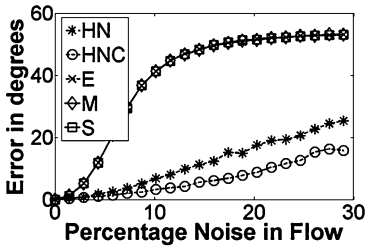
(b) Translation Direction=(1, 0, 0)



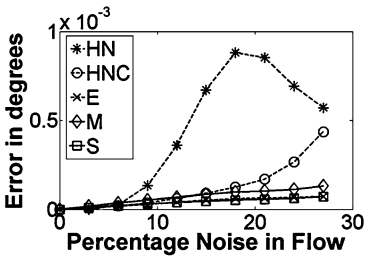
(a) Translation=(0, 0, 0.001)



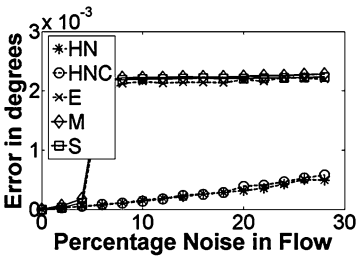
(b) Translation=(0.001, 0, 0)



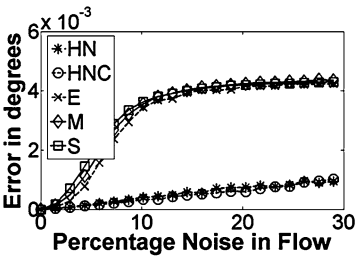
(c) Translation=(0.001, 0.001, 0.001)



(d) Translation=(0, 0, 0.001)



(e) Translation=(0.001, 0, 0)



(f) Translation=(0.001, 0.001, 0.001)

Fig. 2 Error in estimating the rotation with decreasing baseline

Fig. 3 This figure illustrates the performances of various linear algorithms as the noise increases. The rotation parameters for all simulations in this figure is given by Rotation = (0 0 0). Figures (**a**) to (**c**) show the error in estimating the translation direction. Figures (**d**) to (**f**) show the error in estimating the rotation

8.2 Increasing Noise

The scene is similar to that in Sect. 8.1. However, in this scenario, we fix the translation (see Fig. 3 for the translation) and rotation while increasing the amount of noise. The results are presented in Fig. 3, with each column representing different types of translational motions.

8.3 Observations

1. From Figs. 1 and 2, one can see that there was no deterioration in the computation of the motion parameters using the discrete eight point algorithms (E, HN and HNC) despite reductions in the baseline to the limit of arithmetic precision. Note that the errors for the discrete algorithms shot up at about 10^{-12} or 10^{-13}: At this small baseline, the magnitude of the optical flow, being two to three orders of magnitude smaller than the baseline, reached the limit of arithmetic precision, rendering the proportional noise model invalid and thus resulting in the breakdown of the discrete algorithms. This simulation clearly verifies that the theoretical predictions made in the previous section are correct.

2. The performances of the differential essential matrix algorithm (M) and the linear subspace algorithm (S) were extremely poor (Figs. 1(b) and 3), especially in the lateral motion configuration which is susceptible to the bas-relief ambiguity (Chiuso et al. 2000; Daniilidis and Spetsakis 1997; Ma et al. 2001; Xiang and Cheong 2003). Their performances were more or less comparable to that of the un-normalized discrete approach (E). In contrast, the normalized discrete eight point algorithm with Total Least Squares—Fixed Column estimation (*HNC*) appeared to give very much superior results even when the motion was small, with HN's not far behind.
3. In Fig. 2, the absolute rotational error declined proportionally with the baseline as predicted (i.e. the rotational error was of the order $O(\epsilon)m$).
4. Referring to Fig. 3, the impact of noise was keenly felt for the un-normalized discrete approach (E) and the differential algorithms (M and S). Under all motion types tested, the well-known forward bias (Chiuso et al. 2000; Daniilidis and Spetsakis 1997; Ma et al. 2001) reared its ugly head at even a low level of noise. For instance, in Fig. 3(c), when the noise was about 10%, the forward-biased solution of 0° for the translation resulted in an error of about 45° (the true translation vector lies in the 45° direction). In the same token, for the forward translation case (Fig. 3(a)), the excellent results of E, M, and S should be treated with caution. These algorithms had a strong forward bias and irrespective of the true motion, tended to give a forward translation estimate whenever the noise was moderately large. On the other hand, the normalized discrete algorithms (*HN* and *HNC*) exhibited much less sensitivity to noise under all conditions tested, except in the translational estimate of *HN* under forward translation (Fig. 3(a)) with the noise level greater than 10%. The results of this simulation imply that we could expect a stable performance from the discrete *HNC* algorithm when dealing with small motions, provided that the proportional noise in the optical flow computation is small enough.

9 Results on Real Image Sequences

With conventional CCD imaging technology and the mechanical stability of the measurement apparatus, it is clearly impossible to replicate with real image sequences the extremely small baseline scenario in the preceding section. Our goal in this section is to show that over a practical range of decreasing baselines, the normalized discrete algorithms can perform as well, if not better than the differential counterparts. The range of flow magnitude simulated is indicated in the first row of Table 1; our smallest baseline corresponds to the case where the average flow magnitude is of the order 10^{-1} pixel. This limit is reasonable as at the current technology level, the imaging noise expected for a high-quality, 12-bit, scientific imaging system may cause flow variation on the order of 0.01 pixels to 0.001 pixels, depending on the image content (Timoner and Freeman 2001). Such noise level would already constitute a 10% noise for a subpixel image motion of the order 10^{-1} pixel, which would be a problem for both the discrete and the differential algorithms.

Three sequences were taken by moving a camera along a linear rail using two different consumer-grade cameras. For sequence A in Fig. 4(a), the FoV was 31°. For sequences B and C in Figs. 4(b) and (c), the FoV was 53°. Optical flow was estimated using the state-of-the-art algorithm provided by Sand and Teller (2006). 4000 flows were obtained from sub-sampling the available flows and filtered using RANSAC to remove obvious outliers. In most scenes, 99% of the tested flows were considered inliers and different RANSAC trials gave little variation in the results. There was no scene-specific tuning of either the RANSAC thresh-

(a) Sequence A (b) Sequence B (c) Sequence C

Fig. 4 Scenes that have been tested. Field of view ranges from 31 to 53°. Sequences A and B involve a pure lateral translation, while Sequence C involves a pure forward translation

Table 1 Translation errors for sequences in Fig. 4. (*NL*) in the bottom half of the table indicates that a nonlinear bundle adjustment step was used to refine the results obtained by the corresponding linear algorithm in the top half. The first row indicates the average magnitude of the optical flow for the sequence in that particular column

Error (°)	A1	A2	A3	B1	B2	B3	C1	C2	C3
Flow Mag.	0.53	1.0	4.1	0.73	1.17	1.35	0.8	1.1	2.8
HNC	16.5	22.3	5.2	6.5	24.2	14.3	4.4	4.1	3.8
M	89.1	48.5	7.0	87.1	80.6	88.6	2.6	2.5	2.8
S	89.1	48.5	7.0	87.1	80.6	88.6	2.5	2.5	2.8
HNC (NL)	1.8	6.0	1.1	6.7	2.2	7.3	3.3	4.3	2.8
M (NL)	83.0	4.8	1.0	87.3	7.2	84.5	3.2	2.8	2.2
S (NL)	87.2	38.0	5.6	82.9	7.4	87.8	2.3	2.4	2.9

olds or the parameters in the optical flow estimation algorithm. Here, we gave the average error over three trials. For computational efficiency, the number of flows were further reduced to 2000 by sub-sampling before being used for camera pose recovery. For comparison purpose, the camera pose estimated from all the linear algorithms was also refined using the bundle adjustment algorithm (Hartley and Zisserman 2000). The results of the linear algorithms are tabulated in the top half of Table 1, with the corresponding results refined by bundle adjustment reported in the bottom half of Table 1. As we have no means of accurately measuring the ground truth for small rotations, only the translational error is reported.

Sequences in Figs. 4(a) and (b) involve a pure lateral translation, while that in Fig. 4(c) involves a pure forward translation. Observe that the discrete linear estimator *HNC* of Muhlich and Mester (1998) performed much better than the differential estimator *M* from Ma et al. (2000). For example in the lateral motion sequences (Figs. 4(a) and (b)), the discrete algorithm was able to give a good estimate even under circumstances in which its differential counterpart failed completely. These experimental results show clearly that for SFM problems involving a practical range of small motions, the normalized discrete linear algorithms out-performed their differential counterparts by a large margin, especially in lateral motion configuration which are liable to the bas-relief ambiguity. For forward translation (Fig. 4(c)), the performance of the normalized discrete algorithms remained on par with the differential ones and was stable over decreasing baseline. We also note that random noises have apparently substantial effects on the performances of all algorithms, as can be seen from the non-smooth error figures over changing baseline in Table 1. This means that the subsequent step of bundle adjustment to refine the pose estimate is especially important. Given a normalized discrete algorithm that can provide an initial estimate stably over a large range of motion and over different motion configurations, the non-linear bundle adjustment step would have a higher chance of finding the global minimum and will do so more quickly (see bottom half of Table 1).

10 Conclusion

We have proven that the eight point algorithm and its variants are "differential algorithms" in the sense that they can handle arbitrarily small motions given a sufficiently tight bound on the percentage noise. This proof was done using tools from matrix perturbation analysis. It shows that for a sufficiently small percentage noise proportional to the feature displacement magnitude, the eigenvalues of the data matrix remain separate and the solution vector can be recovered well even under very small motion. That is, there is no degeneracy inherent in the discrete linear formulation as the baseline approaches zero. Using both real and simulation results, we have validated the theoretical analysis and shown that even under small motion, the normalized discrete eight point algorithms can perform well and indeed significantly outperform their linear unnormalized differential counterparts. Given that much efforts have been spent in improving the discrete algorithms, and in view of our theoretical and experimental results, it seems that for linear algorithms at least, a properly normalized eight point algorithm should be used for SFM even in small motion.

Having obtained the theoretical result that there is no degeneracy for the two-view SFM case, it would also be interesting to investigate whether the many so-called instabilities associated with small motion in various other problems are due to the instability of the specific discrete algorithms rather than the inherent sensitivity of small motion. For instance, Triggs (1999) considered the case where the third view of a trifocal tensor is obtained by an infinitesimal change of a discrete two-view system. The additional constraint was obtained by differentiating the discrete epipolar constraint $\mathbf{p}^T E\mathbf{p}' = 0$ with both E and $\mathbf{p}'$ changing, which yields $\mathbf{p}^T E\dot{\mathbf{p}}' + \mathbf{p}^T \dot{E}\mathbf{p}' = 0$. While such formulation has the virtue of simplicity, the additional differential information $\mathbf{p}^T E\dot{\mathbf{p}}' + \mathbf{p}^T \dot{E}\mathbf{p}'$ can be drowned out when combined with the existing epipolar constraint $\mathbf{p}^T E\mathbf{p}'$, leading to apparent degeneracy under small changes in E and $\mathbf{p}'$. The problem is not inherently sensitive however; rather, a proper weighing and normalization scheme can do much to enhance the usefulness of the differential information and generally improve the stability of the algorithm. A full treatment of this question is beyond the scope of this paper and presents a very interesting subject for future research.

Appendix A

A.1 Perturbation of Eigenvalues and Eigenvectors

We record some results on perturbation theory from Wilkinson (1965). The first two results are due to Gerschgorin. These Gerschgorin Disc Theorems give us a method of estimating the eigenvalues of a matrix based solely on the entries of the matrix.

Theorem 5 (Wilkinson 1965, Theorem 3, page 71) *Every eigenvalue λ of an $n \times n$ matrix C lies in at least one of the circular discs with centers c_{ii} and radii $\sum_{j \neq i} |c_{ij}|$, where c_{ij} is the entry of the matrix C on its ith row and jth column.*

The above circular disc is called a Gerschgorin disc.

Theorem 6 (Wilkinson 1965, Theorem 4, page 71) *If k of the Gerschgorin disc form a connected domain which is isolated from the other discs, then there are precisely k eigenvalues of C within this connected domain.*

The next result is a slight modification of the above Gerschgorin's Theorems. It is applied to the matrix $\widetilde{A}^T(\epsilon)\widetilde{A}(\epsilon)$ in (28) in Sect. 5.1.

Proposition 8 *Let $\widetilde{C} = C + H$, where $\widetilde{C}$, C and H are $n \times n$ matrices. Suppose there is an invertible matrix K such that $K^{-1}CK = D$, where $D = \mathrm{Diag}(d_i)$ is a diagonal matrix with diagonal entries d_i. Then every eigenvalue $\widetilde{\lambda}$ of $\widetilde{C}$ lies in at least one of the circular discs $\widetilde{G}_i$ with center d_i and radius $\sum_{j=1}^{n} |(K^{-1}HK)_{ij}|$, where $(K^{-1}HK)_{ij}$ is the (ij)-entry of the matrix $K^{-1}HK$.*

Moreover, if k of the above circular discs form a connected domain which is isolated from the other discs, then there are precisely k eigenvalues of $\widetilde{C}$ within this connected domain.

Proof Firstly, we note that the matrices $K^{-1}\widetilde{C}K$ and $\widetilde{C}$ have the same set of eigenvalues. Now, by the first Gerschgorin's result, namely Theorem 5, every eigenvalue of $K^{-1}\widetilde{C}K$, and hence of $\widetilde{C}$, lies in one of its Gerschgorin discs. The ith Gerschgorin disk G_i of the matrix $K^{-1}\widetilde{C}K$ is given by

$$G_i = \left\{ \lambda \,|\, |\lambda - (d_i + (K^{-1}HK)_{ii})| \leq \sum_{j \neq i} |(K^{-1}HK)_{ij}| \right\}. \tag{51}$$

Applying the triangle inequality to the inequality in (51) gives

$$|\lambda - d_i| \leq \sum_{j} |(K^{-1}HK)_{ij}|,$$

which defines a circular disc $\widetilde{G}_i$ centered at d_i and with radius $\sum_j |(K^{-1}HK)_{ij}|$. This circular disc contains the ith Gerschgorin disk G_i. Consequently, every eigenvalue of $\widetilde{C}$ lies in one of such circular discs. The second part of the proposition now follows readily from Theorem 6. □

Before we proceed to obtain the perturbation of eigenvectors, we first include a simple proof of the next lemma which is used to provide us with a unit vector.

Lemma 6 *Suppose $\mathbf{q}(\epsilon) = \mathbf{r}(\epsilon) + \mathbf{z}(\epsilon)$ where $\mathbf{r} = O(\epsilon^0)$ and $\mathbf{z}(\epsilon) = O(\epsilon)$. Then the unit vector*

$$\breve{\mathbf{q}}(\epsilon) = \frac{\mathbf{q}(\epsilon)}{\|\mathbf{q}(\epsilon)\|}$$

can be expressed as

$$\breve{\mathbf{q}}(\epsilon) = \breve{\mathbf{r}}(\epsilon) + \mathbf{w}(\epsilon)$$

where $\breve{\mathbf{r}}(\epsilon) = \frac{\mathbf{r}(\epsilon)}{\|\mathbf{r}(\epsilon)\|}$ and $\mathbf{w}(\epsilon) = O(\epsilon)$.

Proof Note that $\|\mathbf{q}(\epsilon)\| = \|\mathbf{r}(\epsilon)\| + O(\epsilon)$. Thus,

$$\begin{aligned}
\breve{\mathbf{q}}(\epsilon) &= \frac{\mathbf{q}(\epsilon)}{\|\mathbf{q}(\epsilon)\|} \\
&= \frac{1}{\|\mathbf{r}(\epsilon)\| + O(\epsilon)}(\mathbf{r}(\epsilon) + \mathbf{z}(\epsilon)) \\
&= \frac{1}{\|\mathbf{r}(\epsilon)\|(1 + O(\epsilon))}(\mathbf{r}(\epsilon) + \mathbf{z}(\epsilon)) \\
&= \breve{\mathbf{r}}(\epsilon) + \mathbf{w}(\epsilon)
\end{aligned}$$

where $\mathbf{w}(\epsilon) = O(\epsilon)$. We have made use of $\frac{1}{1+O(\epsilon)} = 1 + O(\epsilon)$ from (3). □

For a perturbed symmetric matrix, we first have the following result on its perturbed eigenvalues from Wilkinson (1965).

Theorem 7 (Wilkinson 1965, Wielandt-Hoffman Theorem, page 104) *Suppose $\widetilde{C}(H) = C + H$, where $\widetilde{C}(H)$, C and H are $n \times n$ real symmetric matrices. If $\widetilde{C}(H)$ and C have eigenvalues $\widetilde{\lambda}_i(H)$ and λ_i respectively and they are arranged in non-increasing order, then*

$$\sum_{i=1}^{n} (\widetilde{\lambda}_i(H) - \lambda_i)^2 \leq \|H\|^2.$$

It follows that for each i,

$$|\widetilde{\lambda}_i(H) - \lambda_i| \leq \|H\|. \tag{52}$$

In the above statement, we have used the symbol $\widetilde{C}(H)$ instead of $\widetilde{C}$ for $C + H$ to stress the dependence of its eigenvalue $\widetilde{\lambda}_i(H)$ on H.

To obtain the perturbed eigenvectors, we may apply a technique in Wilkinson (1965) which we have also used in Sect. 6. The idea is quite simple and we thus state the result without proof.

Lemma 7 *Let $\widetilde{C}(H) = C + H$, where $\widetilde{C}(H)$, C and H are $n \times n$ real symmetric matrices. Suppose $\{\mathbf{r}_i | 1 \leq i \leq n\}$ is a basis of eigenvectors of C, where each $\mathbf{r}_i$ is an eigenvector that corresponds to eigenvalue λ_i. For a fixed k, let*

$$\mathbf{q}_k(H) = \sum_{i=1}^{n} \alpha_i(H)\mathbf{r}_i$$

be an eigenvector of $\widetilde{C}(H)$ *corresponding to eigenvalue* $\widetilde{\lambda}_k(H)$. *Suppose* j *is such that* $\lambda_j \neq \lambda_k$. *Then for a sufficiently small* $\|H\|$, *the projection* $\alpha_j(H)$ *of* $\mathbf{q}_k(H)$ *on* $\mathbf{r}_j$ *is non-maximal, i.e.,*

$$|\alpha_j(H)| \neq \max\{|\alpha_i(H)| \mid 1 \leq i \leq n\}.$$

Theorem 8 *Suppose* $\widetilde{C}(H) = C + H$, *where* $\widetilde{C}(H), C$ *and* H *are* $n \times n$ *real symmetric matrices. The unit eigenvectors* $\widetilde{\mathbf{q}}_k(H)$ *and* $\mathbf{r}_k$ *of* $\widetilde{C}(H)$ *and* C *corresponding to* $\widetilde{\lambda}_k(H)$ *and* λ_k *respectively are related by*

$$\widetilde{\mathbf{q}}_k(H) = \mathbf{q}'_k(H) + O(\|H\|),$$

where $\mathbf{q}'_k(H)$ *is a unit vector and is a linear combination of all those eigenvectors* $\mathbf{r}_j$ *of* C, *whose associated eigenvalue* λ_j *is identical to* λ_k.

A.2 Errors in the Translation Vector and Rotation Matrix

In this subsection, we show that the decomposition of an essential matrix into its rotational and translational terms is stable. This means that in general, lowering the amount of noise will improve the rotational as well as the translational estimate, rather than only one or the other.

Let $E(\epsilon)$ be an essential matrix with a finite Frobenius norm. Recall $\widetilde{E}'(\epsilon)$ is the corrupted version of $E(\epsilon)$ but it has been corrected to possess the desired properties of an essential matrix. The rotation and translation estimates are obtainable from the SVD of $\widetilde{E}'(\epsilon)$ using the algorithm in Hartley and Zisserman (2000). The SVD process involves the eigenvalues and eigenvectors of the real symmetric matrices $\widetilde{E}'^T(\epsilon)\widetilde{E}'(\epsilon)$ and $\widetilde{E}'(\epsilon)\widetilde{E}'^T(\epsilon)$, of whom we note the following (ϵ is here suppressed temporarily):

$$\begin{aligned}\|\widetilde{E}'^T\widetilde{E}' - E^TE\| &= \|\widetilde{E}'^T(\widetilde{E}' - E) + (\widetilde{E}'^T - E^T)E\| \\ &\leq \|\widetilde{E}'^T\|\,\|\widetilde{E}' - E\| + \|(\widetilde{E}'^T - E^T)\|\,\|E\| \\ &\leq \|\widetilde{E}' - E\|(2\|E\| + \|\widetilde{E}' - E\|), \qquad (53)\end{aligned}$$

which has the same order as $\|\widetilde{E}'(\epsilon) - E(\epsilon)\|$, since $\|E(\epsilon)\|$ is finite. The same result can be obtained for $\|\widetilde{E}'(\epsilon)\widetilde{E}'^T(\epsilon) - E(\epsilon)E^T(\epsilon)\|$. Thus, both errors have the same order as $\|\widetilde{E}'(\epsilon) - E(\epsilon)\|$.

Consider the SVD of the matrix $E(\epsilon)$

$$E(\epsilon) = U(\epsilon)\begin{bmatrix}\sqrt{\lambda(\epsilon)} & 0 & 0\\ 0 & \sqrt{\lambda(\epsilon)} & 0\\ 0 & 0 & 0\end{bmatrix}V^T(\epsilon) \qquad (54)$$

where $\lambda(\epsilon) = \lambda + O(\epsilon)$ is a positive real number, and $U(\epsilon)$ and $V(\epsilon)$ are orthogonal matrices. Each ith column $\mathbf{v}_i(\epsilon)$ of $V(\epsilon)$ is a unit eigenvector of $E^T(\epsilon)E(\epsilon)$ that corresponds to the eigenvalue $\lambda(\epsilon)$ for $i = 1, 2$ and 0 for $i = 3$.

Likewise, we have the corresponding SVD of $\widetilde{E}'(\epsilon)$:

$$\widetilde{E}'(\epsilon) = U'(\epsilon)\begin{bmatrix}\sqrt{\lambda'(\epsilon)} & 0 & 0\\ 0 & \sqrt{\lambda'(\epsilon)} & 0\\ 0 & 0 & 0\end{bmatrix}(V'(\epsilon))^T, \qquad (55)$$

where the ith column of $V'(\epsilon)$ is the unit eigenvector of $\widetilde{E}'^T(\epsilon)\widetilde{E}'(\epsilon)$.

Using (54) and (55), for $i = 1, 2$, the ith columns $\mathbf{u}_i(\epsilon)$, $\mathbf{v}_i(\epsilon)$, $\mathbf{u}'_i(\epsilon)$ and $\mathbf{v}'_i(\epsilon)$ of the respective matrices $U(\epsilon)$, $V(\epsilon)$, $U'(\epsilon)$ and $V'(\epsilon)$ are related as follows,

$$\begin{aligned}\sqrt{\lambda(\epsilon)}\mathbf{v}_i(\epsilon) &= E^T(\epsilon)\mathbf{u}_i(\epsilon),\\ \sqrt{\lambda'(\epsilon)}\mathbf{v}'_i(\epsilon) &= \widetilde{E}'^T(\epsilon)\mathbf{u}'_i(\epsilon).\end{aligned} \qquad (56)$$

From Hartley and Zisserman (2000), the translation directions associated with $\widetilde{E}'(\epsilon)$ and $E(\epsilon)$ are given by the third columns $\mathbf{u}'_3(\epsilon)$ and $\mathbf{u}_3(\epsilon)$ respectively. The next result gives the error involved in these translation vector estimates.

Proposition 9 *For the unit translational vectors* $\mathbf{u}_3(\epsilon)$ *and* $\mathbf{v}_3(\epsilon)$, *the errors* $\|\mathbf{u}'_3(\epsilon) - \mathbf{u}_3(\epsilon)\|$ *and* $\|\mathbf{v}'_3(\epsilon) - \mathbf{v}_3(\epsilon)\|$ *have the same order as* $\|\widetilde{E}'(\epsilon) - E(\epsilon)\|$.

Proof The vectors $\mathbf{u}_3(\epsilon)$, $\mathbf{v}_3(\epsilon)$ $\mathbf{u}'_3(\epsilon)$ and $\mathbf{v}'_3(\epsilon)$ are unit eigenvectors corresponding to the simple eigenvalue 0 of the real symmetric matrices $E^T(\epsilon)E(\epsilon)$, $E(\epsilon)E^T(\epsilon)$, $\widetilde{E}'^T(\epsilon)\widetilde{E}'(\epsilon)$ and $\widetilde{E}'(\epsilon)\widetilde{E}'^T(\epsilon)$ respectively.

The result now follows readily from Theorem 8 in Appendix A.1 and the error obtained in (53). □

Next, we relate $U(\epsilon)$ to $U'(\epsilon)$ and $V(\epsilon)$ to $V'(\epsilon)$ when ϵ is sufficiently small. This relationship is then used to determine the error in the estimate of the rotation matrix.

Lemma 8 *Both* $\|U'(\epsilon) - U(\epsilon)\|$ *and* $\|V'(\epsilon) - V(\epsilon)\|$ *have the same order as* $\|\widetilde{E}'(\epsilon) - E(\epsilon)\|$.

Proof The non-zero eigenvalue of the real symmetric matrix $E^T(\epsilon)E(\epsilon)$ (and hence also $E(\epsilon)E^T(\epsilon)$) is repeated twice. Hence, the corresponding eigen space has dimension 2. Therefore, we choose $\mathbf{u}_2(\epsilon)$ and $\mathbf{u}'_2(\epsilon)$ such that

$$\mathbf{u}'_2(\epsilon) = \mathbf{u}_2(\epsilon) + O(\|\widetilde{E}'(\epsilon) - E(\epsilon)\|). \qquad (57)$$

(Here we have used (53).)

Now, we view $\widetilde{E}'^T(\epsilon)\widetilde{E}'(\epsilon)$ as a perturbation of $E^T(\epsilon)E(\epsilon)$ with $\|\widetilde{E}'^T(\epsilon)\widetilde{E}'(\epsilon) - E^T(\epsilon)E(\epsilon)\| = O(\|\widetilde{E}'(\epsilon) - E(\epsilon)\|)$. By the Wielandt-Hoffman Theorem (recorded as Theorem 7 in Appendix A.1), the perturbed eigenvalue $\lambda'_i(\epsilon)$ is

$$\lambda'_i(\epsilon) = \lambda_i(\epsilon) + O(\|\widetilde{E}'(\epsilon) - E(\epsilon)\|), \quad i = 1, 2,$$

so that, using (56), (57) and (3), we obtain

$$\mathbf{v}_2'(\epsilon) = \frac{(\widetilde{E}'(\epsilon))^T \mathbf{u}_2'(\epsilon)}{\sqrt{\lambda(\epsilon) + O(\|\widetilde{E}'(\epsilon) - E(\epsilon)\|)}} = \frac{1}{\sqrt{\lambda(\epsilon)}}(\widetilde{E}'(\epsilon))^T \mathbf{u}_2(\epsilon) + O(\|\widetilde{E}'(\epsilon) - E(\epsilon)\|).$$

Therefore, we have

$$\|\mathbf{v}_2'(\epsilon) - \mathbf{v}_2(\epsilon)\| \leq \|\widetilde{E}'(\epsilon) - E(\epsilon)\| \frac{\|\mathbf{u}_2(\epsilon)\|}{\sqrt{\lambda(\epsilon)}} + \|O(\|\widetilde{E}'(\epsilon) - E(\epsilon)\|)$$

which is of the same order as $\|\widetilde{E}'(\epsilon) - E(\epsilon)\|$.

Now, the vector $\mathbf{u}_1(\epsilon)$ (respectively $\mathbf{v}_1(\epsilon)$), being orthogonal to both $\mathbf{u}_2(\epsilon)$ and $\mathbf{u}_3(\epsilon)$ (respectively $\mathbf{v}_2(\epsilon)$ and $\mathbf{v}_3(\epsilon)$), can be obtained by taking the unit vector along $\widehat{\mathbf{u}_2(\epsilon)\mathbf{u}_3(\epsilon)}$ (respectively $\widehat{\mathbf{v}_2(\epsilon)\mathbf{v}_3(\epsilon)}$).

Similarly, we have $\mathbf{u}_1'(\epsilon)$ as the unit vector along $\widehat{\mathbf{u}_2'(\epsilon)\mathbf{u}_3'(\epsilon)}$. The error

$$\|\mathbf{u}_1'(\epsilon) - \mathbf{u}_1(\epsilon)\| = \|\widehat{\mathbf{u}_2'(\epsilon)\mathbf{u}_3'(\epsilon)} - \widehat{\mathbf{u}_2(\epsilon)\mathbf{u}_3(\epsilon)}\|$$

can be shown to have the same order as $\|\widetilde{E}'(\epsilon) - E(\epsilon)\|$. The same argument applies for the error

$$\|\mathbf{v}_1'(\epsilon) - \mathbf{v}_1(\epsilon)\|.$$

Hence, we have proven the result. □

Finally, we discuss the error in the rotation matrix.

Proposition 10 *Let the rotation matrices associated with $E(\epsilon)$ and $\widetilde{E}'(\epsilon)$ be denoted as $R(\epsilon)$ and $R'(\epsilon)$ respectively. Then the Frobenius norm of the error in estimating the rotation matrix is of the same order as $\|\widetilde{E}'(\epsilon) - E(\epsilon)\|$.*

Proof The rotation matrices associated with $E(\epsilon)$ and $\widetilde{E}'(\epsilon)$ are $R(\epsilon)$ and $R'(\epsilon)$ respectively. Using the algorithm in Hartley and Zisserman (2000), $R(\epsilon)$ and $R'(\epsilon)$ are given by

$$\begin{aligned} R(\epsilon) &= U(\epsilon)W(V(\epsilon))^T, \\ R'(\epsilon) &= U'(\epsilon)W(V'(\epsilon))^T, \end{aligned} \tag{58}$$

where W may take the form

$$\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The correct form of W can be identified by enforcing the positive depth constraint.

Assume that we have identified the true W and we call it W_0. Using (58), the difference between $R(\epsilon)$ and $R'(\epsilon)$ is given by (ϵ is again suppressed temporarily)

$$\begin{aligned} \|R - R'\| &= \|U'W_0V'^T - UW_0V^T\| \\ &\leq \|U'W_0V'^T - U'W_0V^T\| \\ &\quad + \|U'W_0V^T - UW_0V^T\| \\ &\leq \|U'\|\,\|W_0\|\,\|V'^T - V^T\| \\ &\quad + \|U' - U\|\,\|W_0\|\,\|V^T\| \end{aligned}$$

which has the same order as $\|\widetilde{E}'(\epsilon) - E(\epsilon)\|$. □

References

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M. J., & Szeliski, R. (2007). Database and evaluation methodology for optical flow. In *IEEE 11th int. conf. on computer vision*.

Baumela, L., Agapito, L., Reid, I., & Bustos, P. (2000). Motion estimation using the differential epipolar equation. In *International conf. on pattern recognition* (Vol. 3, pp. 840–843).

Brooks, M. J., Chojnacki, W., & Baumela, L. (1997). Determining the egomotion of an uncalibrated camera from instantaneous optical flow. *Journal of Optical Society of America A*, *14*(10), 2670–2677.

Bruhn, A., Weickert, J., & Schnörr, C. (2005). Lucas/Kanade meets Horn/Schunck: combining local and global optic flow methods. *International Journal of Computer Vision*, *61/3*, 211–231.

Chiuso, A., Brockett, R., & Soatto, S. (2000). Optimal structure from motion: local ambiguities and global estimates. *International Journal of Computer Vision*, *39*(3), 195–228.

Chojnacki, W., Brooks, M. J., van den Hengel, A., & Gawley, D. (2003). Revisiting Hartley's normalised eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(9), 1172–1177.

Daniilidis, K., & Spetsakis, M. E. (1997). Understanding noise sensitivity in structure from motion. In Y. Aloimonos (Ed.), *Visual navigation: from biological systems to unmanned ground vehicles* (pp. 61–88). Hillsdale: Lawrence Erlbaum.

Fermüller, C. (1995). Passive navigation as a pattern recognition problem. *International Journal of Computer Vision*, *14*, 147–158.

Hartley, R. (1997). In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(6), 580–593.

Hartley, R., & Zisserman, A. (2000). *Multiple view geometry in computer vision*. Cambridge: Cambridge University Press.

Heeger, D. J., Kosecka, J., & Sastry, S. (1992). Subspace methods for recovering rigid motion I: algorithm and implementations. *International Journal of Computer Vision*, *7*(2), 95–117.

Ho, H. T., & Goecke, R. (2008). Optical flow estimation using Fourier Mellin transform. In *IEEE conf. computer vision and pattern recognition*.

Horn, B. K. P., & Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, *17*(1–3), 185–203.

Horn, B. K. P., & Weldon, E. J. Jr. (1988). Direct method for recovering motion. *International Journal of Computer Vision*, *2*, 51–76.

Kanatani, K. (1993). 3d interpretation of optical flow by renormalization. *International Journal of Computer Vision*, *11*(3), 267–282.

Kanatani, K. (2003). Comparing optimal 3-D reconstruction for finite motion and optical flow. *Journal of Electronic Imaging*, *12*(3), 478–488.

Lempitsky, V., Roth, S., & Rother, C. (2008). Fusionflow: discrete-continuous optimization for optical flow estimation. In *IEEE conf. computer vision and pattern recognition*.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, *293*, 133–135.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London, Series B*, *208*, 385–397.

Lucas, B., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of DARPA image understanding workshop* (pp. 121–130).

Luong, Q. T., & Faugeras, O. (1996). The fundamental matrix: theory, algorithms and stability analysis. *International Journal of Computer Vision*, *17*(1), 43–76.

Ma, Y., Kosecka, J., & Sastry, S. (2000). Linear differential algorithm for motion recovery: a geometric approach. *International Journal of Computer Vision*, *36*(1), 71–89.

Ma, Y., Kosecka, J., & Sastry, S. (2001). Optimization criteria, sensitivity and robustness of motion and structure estimation. *International Journal of Computer Vision*, *44*(3), 219–249.

Ma, Y., Soatto, S., Kosecka, J., & Sastry, S. S. (2003). *An invitation to 3-D vision*. New York: Springer.

Mainberger, M., Bruhn, A., & Weickert, J. (2008). Is dense optic flow useful to compute the fundamental matrix? In *Lecture notes in computer science: Vol. 5112. Proceedings of the 5th international conference on image analysis and recognition* (pp. 630–639). Berlin: Springer.

Maybank, S. (1992). *Theory of reconstruction from image motion*. Berlin: Springer.

Muhlich, M., & Mester, R. (1998). The role of total least squares in motion analysis. In *European conf. on computer vision*.

Negahdaripour, S. (1989). Critical surface pairs and triplets. *International Journal of Computer Vision*, *3*, 293–312.

Nir, T., Bruckstein, A. M., & Kimmel, R. (2007). Over-parameterized variational optical flow. *International Journal of Computer Vision*, *76*(2), 205–216.

Nister, D. (2007). An efficient solution for infinitesimal camera motion. In *IEEE conf. computer vision and pattern recognition*.

Ohta, A. (1996). Uncertainty models of the gradient constraint for optical flow computation. *IEICE Transactions on Information and Systems*, *7*, 958–964.

Ren, X. (2008). Local grouping for optical flow. In *IEEE conf. computer vision and pattern recognition*.

Sand, P., & Teller, S. J. (2006). Particle video: long-range motion estimation using point trajectories. In *IEEE conf. computer vision and pattern recognition*.

Timoner, S. J., & Freeman, D. M. (2001). Multi-image gradient-based algorithms for motion estimation. *Optical Engineering*, *40*(9), 2003–2016.

Torr, P., & Murray, D. (1997). The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, *24*(3), 271–300.

Triggs, B. (1999). Differential matching constraints. In *IEEE international conf. computer vision* (pp. 370–376).

Verri, A., & Poggio, T. (1989). Motion field and optical flow: qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 490–498.

Viéville, T., & Faugeras, O. (1995). Motion analysis with a camera with unknown and possibly varying intrinsic parameters. In *IEEE international conf. computer vision* (pp. 750–756).

Weng, J., Huang, T., & Ahuja, N. (1989). Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 451–476.

Wilkinson, J. H. (1965). *The algebraic eigenvalue problem*. Oxford: Clarendon Press.

Xiang, T., & Cheong, L. F. (2003). Understanding the behavior of SFM algorithms: a geometric approach. *International Journal of Computer Vision*, *51*(2), 111–137.