Optimal and Suboptimal Structure From Motion

Stefano Soatto †‡ *Roger Brockett* †

† Harvard University, 29 Oxford st., Cambridge – MA 02139
‡ Università di Udine, Udine – Italy
soatto@hrl.harvard.edu

Submitted to the IEEE Int. Conf. on Computer Vision
April 15, 1997

Abstract

"Structure From Motion" (SFM) refers to the problem of estimating three-dimensional information about the environment from the motion of its two-dimensional projection onto a surface (for instance the retina). Noise plays an important role in this problem, but it has been addressed only marginally in more than twenty years of research.

We present an analysis of SFM from the point of view of noise. This analysis results in algorithms that are provably convergent and provably optimal with respect to a chosen norm. In particular, we cast SFM as a nonlinear optimization problem and define a bilinear projection iteration that converges to fixed points of a certain cost-function. We then show that such fixed points are "fundamental", i.e. are intrinsic to the problem of SFM and not an artifact introduced by our algorithms. We classify and interpret geometrically local extrema, and we argue that they correspond to phenomena observed in visual psychophysics. Finally, we show under what conditions it it is possible, given convergence to a local extremum, to "jump" to the valley containing the optimum.

1 Introduction

The problem of "Structure From Motion" (SFM) deals with extracting three-dimensional information about the environment from the motion of its projection onto a two-dimensional surface. The most outstanding example of machinery to deal with this problem is the combination of the human eye and brain: from the projection of moving objects onto the retina, we are able to gather a three-dimensional representation that is sufficient for us to reach for them, manipulate them, walk around them etc. .

At this level of generality, SFM is an extraordinarily complicated problem. However, it is possible to simplify the representation of the environment up to the point in which we can say something more precise. For instance, we can represent the environment as a set of points in 3-D space that move rigidly relative to the imaging surface (the retina or the CCD sensor of a video-camera). The goal of SFM is then to estimate the 3-D shape and motion of such feature points given either the velocity of their perspective projection onto the imaging sensor (*optical flow*), or the correspondence between projections taken from different vantage points (*feature correspondence* or *tracking*).

The restriction of SFM to feature points has been subject to the attention of the Computer Vision community for at least twenty years. To some, this is a deceptive simplification of the original problem of SFM, for the representation of the environment using rigid feature points does not account for visually complex phenomena such as the motion of the foliage of a tree or a that of a silk gown. We fully agree with this point of view. However, in partial justification of the fact that this paper *is* about feature points, we should add that there are still some important issues in SFM that have been touched upon only marginally, and we sense in the community the need for analytical results that make SFM usable beyond the few test image sequences normally employed to argue the feasibility of a particular algorithm.

A brief history of SFM

In our opinion, the first breakthrough in the recent history of SFM has been due to Longuet-Higgins [9], who proposed an algorithm to reconstruct a (point-wise) scene of 8 points or more from two projections. It triggered an entire stream of work, which we refer to collectively as "epipolar geometry", that has generated numerous contributions over the past 15 years, due to Faugeras, Maybank, Weng, Huang and many more (see Faugeras' book [4] for an overview of the results and appropriate credits). The apogee of that line of work, in our opinion, is a recent body of results due to Carlsson, Faugeras, Shashua and others, summarized in the forthcoming book of Faugeras and Luong [5].

At the same time, other significant contributions were being made in special cases of SFM. For instance, the work of Tomasi and Kanade [20] solves optimally the problem of SFM under orthographic projection. They cast SFM as an optimal fixed-rank approximation, that they solve using factorization methods. The simplicity, elegance and robustness of factorization-based methods are at the base of their widespread use under the conditions where orthographic projection is an accurate approximation of the imaging device. The results of Tomasi and Kanade were later extended to other projection models such as "para-perspective" [12].

Another line of work has also appeared recently, after the work of Heeger and Jepson [7]. It is based upon an algebraic manipulation of the optical flow equation that renders the problem of estimating SFM linear. These linear subspace methods are less widely applied owing to the fact that large optical flow is difficult to measure, while small optical flow is usually insufficient to achieve a consistent estimate. Furthermore, in the presence of noise these linear methods return a biased estimate.

Optimism and pessimism about SFM

After the work in epipolar geometry, we can safely say that the *geometry* of SFM (for feature points) is fairly well understood. In particular, under very mild general-position conditions, we know that there exists a *unique* solution for SFM. Finding it is a matter of algebra, and there are a number of ways to do that. Also, the performance of many of the algorithms has been demonstrated on (more or less controlled) sequences of real images, and this has led many in the Computer Vision community to conclude that SFM has been solved, and to declare it as no longer interesting as a scientific problem, but a mere object of "development".

On the other hand, we are witnessing the frustration of many others, especially engineers, who are implementing existing algorithms for SFM in real-world situations, and observe that they behave in a way that is much different from their declared performance. This has led some to conclude that SFM is too difficult a problem to be solved even in the case of point features [11].

How can it be possible that the same problem is trivial to some and impossible to others?

We believe there are two equally important components in SFM: geometry and noise. While the geometry has been studied extensively, the issue of noise has been touched upon only in a superficial way (there are some notable exceptions [21] upon which we will comment later). More importantly, the combined issue of noise *and* geometry has been completely ignored in 20 years of work in SFM.

Let us articulate on the frustration of the pessimists of SFM: There is some agreement that, in the presence of a *long baseline* (fast motion), SFM is quite simple to solve, and almost every algorithm gives an honest estimate [11]. On the other hand, the correspondence problem is hard to solve, if not impossible, under these circumstances (in applications where one can count on a significant assistance from the user, for instance in establishing the correspondence between few features by hand, or in presence of calibrated stereo, one needs not worry about this issue).

In order to make the correspondence problem easy in more general situations, one has to limit the baseline to be very short (image-motion in the order of one to few pixels), and then many optical-flow/feature-tracking algorithms do an decent job [1]. However, SFM becomes very hard in these situations, for the average image-motion is comparable with the localization error, and most algorithms give an estimate that bears no apparent relationship with the correct answer.

A potential solution to this situation is to increase the baseline by integrating SFM over time. This, however, cannot be achieved by mere "time-averaging", for if the solution from any two adjacent frames is wildly wrong, their average will not be much better. Numerous algorithms have been proposed to solve SFM recursively, but most of them rely on maintaining tracking of features over long periods of time (10 frames

of more). If that were possible, one could track point-features until the baseline is long enough, and then apply any SFM algorithm from 2 frames. However, it is more often the case that, due to occlusions and the intrinsically local nature of feature tracking, features are lost after a few frames. Therefore the need to integrate 3-D information over time in presence of short life-span of feature-points (or optical flow in the limit); see [17] for a discussion on the issue of time-integration. In this paper we only address the problem of SFM from 2 views (or optical flow), and seek for an optimal solution with respect to noise.

Geometry and noise

Most of the work in SFM relies on the assumption that noise is small, either to neglect it altogether, or to analyze it by means of linearization. In section 6 we will argue that, in many situations normally encountered in real life, even the most accurate optical-flow/feature-tracking algorithm leads to "signal to noise ratios" in the order of 100%. Only recently the role of noise has started to be explored, but mainly at the experimental –as opposed to analytical– level (see for instance [19]). A thorough analysis of SFM from the perspective of noise should answer at least some of the following questions:

- When can we say that the noise is "small"? (i.e. when do standard techniques break down?)
- In presence of large noise, what is the "best" one can do?
- How does one achieve the optimal solution?

In this paper we will see that solving SFM in the presence of noise entails solving a non-linear optimization problem. While it is not clear whether the optimum can be computed in closed-form, any iterative optimization algorithm is bound to have "spurious" solutions, corresponding to local minima of the cost function being optimized. Naturally, this raises some important questions:

- Are there spurious solutions (local minima) in the minimization process?
- If there are are spurious solutions, are they *fundamental*? i.e. are they intrinsic to the problem or are they an artifact of the particular optimization technique?
- Given convergence to a spurious solution, can we "jump" to the correct solution?
- If that is not possible, is there some representation that is "invariant" within different spurious solutions (so that it represents a robust quantity to estimate)?

A more speculative question one can ask is related to the human visual system:

- If there are spurious solutions that are fundamental (i.e. intrinsic to SFM), do these correspond to phenomena that can be observed in psychophysical experiments?

From the standpoint of the engineer:

- Is there a set of experiments that is "representative" to test various algorithms?

In this paper we present an analysis of SFM from the point of view of noise. This analysis results in an algorithm that provably converges to local extrema of a certain cost-function. We then prove that the extrema of this cost function are "fundamental", i.e. are intrinsic to the problem of SFM (in presence of noise), and not an artifact introduced by our algorithm. We also classify and interpret geometrically such local extrema, and we observe that they correspond to phenomena observed in visual psychophysics.

We show how it is possible, under certain conditions, to obtain the optimal solution to SFM given convergence to a local extremum.

We use a spherical projection model, because it renders the algebra sleek. However, there is nothing fundamental about that particular model, and any other representation of perspective projection would do. What is crucial, instead, is the factorization of the orthogonal projection operator (theorem 2.1), and the fact that substitution in the bilinear cost-functional does not alter the extrema (lemma 3.1).

Relation to previous work

This paper relates to many of the previous works in SFM, and some of the relationships are pointed out throughout the paper. In this section we only remark about the one that relates most closely, that is [21]. There, various general-purpose nonlinear optimization techniques are tested on various cost functions that include the epipolar constraint, as well as the average 2-norm of the image-measurements. Cramèr-Rao bounds are evaluated, and an extensive set of simulation experiments compares existing linear algorithms against the optimal.

The way local extrema are addressed in [21] is that iterative optimization schemes are initialized using a linear algorithm (such as a variation of the 8-point algorithms of Longuet-Higgins proposed by the same authors). This choice, however, defeats part of the purpose of studying SFM from the point of view of noise, for it guarantees success only where the linear algorithms work. From this standpoint, [21] can be viewed as addressing the *optimal refinement* of existing SFM algorithms. Under conditions in which the linear algorithms do not give a satisfactory answer, Weng et al. only guarantee convergence to a local minimum, with no indication as to how this is related to the optimum.

Therefore, the upshot is that, when the algorithms of Weng et al. converge to the optimal solution, their solution is identical to the one we obtain. However, we achieve useful results for noises that are one order of magnitude higher, without imposing restrictions on the initial conditions.

Organization of the paper

Section 2.3 addresses the problem of "Spherical Least-squares", that is propaedeutic to the solution of SFM proposed in section 3. That is the core section of the paper, where the "Bilinear Projection Algorithm" is proposed. After we have established that such algorithm does not introduce spurious solutions, in section 4 we address the issue of whether the original problem of SFM admits local solutions in presence of noise. We do so with the aid of random sampling techniques: the analysis of the results leads to the formalization of well-known phenomena such as the "bas-relief" ambiguity, or the "rubbery motion" perception.

In section 5 we give a recipe-like description of the algorithm to help implementation, and in section 6 we report the results of extensive simulations as well as experiments on real images.

2 Spherical Least-Squares

Suppose we are given p unit-norm vectors $\mathbf{x}_1, \ldots, \mathbf{x}_p \in \mathbf{S}^2$ and an unknown transformation $\mathbf{a} \in \mathbb{R}^3$ that acts on each $\mathbf{x}_i$, $i = 1 \ldots p$ via the cross-product $\mathbf{a} \times \mathbf{x}_i$. The vector $\mathbf{a} \times \mathbf{x}_i$ belongs to the plane orthogonal to $\mathbf{x}_i$, that is the tangent plane to the unit-sphere $\mathbf{S}^2$ at $\mathbf{x}_i$, $T_{\mathbf{x}_i}\mathbf{S}^2$ (see figure 1).

Suppose further that we can measure each transformed vector $\mathbf{a} \times \mathbf{x}_i$ up to an unknown scaling factor $\lambda_i \in (0, 1)$:

$$\mathbf{y}_i = \mathbf{a} \times \mathbf{x}_i \lambda_i + \mathbf{n}_i \quad i = 1 \ldots p \tag{1}$$

where each $\mathbf{n}_i$ represents the uncertainty (or error) in the measurement $\mathbf{y}_i$, and hence it is constrained to belong to the plane orthogonal to $\mathbf{x}_i$:

$$\mathbf{n}_i \in T_{\mathbf{x}_i}\mathbf{S}^2. \tag{2}$$

We now ask the question of – given a number p of vectors $\mathbf{x}_i$ and the corresponding measurements $\mathbf{y}_i$ – how to find the transformation $\mathbf{a} \in \mathbb{R}^3$ and the scales $\lambda = [\lambda_1, \ldots, \lambda_p]^T \in \mathbb{R}^p_+$ that minimize the norm of the uncertainty $\mathbf{n}$:

$$\min_{\mathbf{a} \in \mathbb{R}^3, \lambda \in \mathbb{R}^p_+} \sum_{i=1}^{p} ||\mathbf{n}_i||^2_{w_i} \quad \text{subject to } \mathbf{y}_i = \mathbf{a} \times \mathbf{x}_i \lambda_i + \mathbf{n}_i \quad \in T_{\mathbf{x}_i}\mathbf{S}^2 \tag{3}$$

where w_i indicate the weights chosen for the components of the cost function. We will analyze three distinct cases. In the first, w_i are chosen as $w_i(\mathbf{a}) \doteq ||\mathbf{a} \times \mathbf{x}_i|| \in [0, 1]$, which leads to a simple closed-form optimal solution. In the second case, we analyze the choice of "balanced weights" $w(\mathbf{a}) \doteq \frac{\|\mathbf{a}\times\mathbf{x}_i\|}{\|\mathbf{a}\times\bar{\mathbf{x}}\|}$, where $\bar{\mathbf{x}}$ is the centroid of the points $\mathbf{x}_i$. This case also results in a closed-form optimal solution that is only slightly more complicated than the one obtained in the first case. Finally, we analyze the case $w_i = 1 \; \forall \; i$, for which we do not have an optimal closed-form solution. However, we propose an iterative scheme that is guaranteed

to converge to a local extremum of the cost function. Obviously, in the absence of noise ($\mathbf{n} = 0$) the three problems of minimizing the cost function above under the different choices of weights have the same exact solution. In the presence of noise, typically the solution obtained with $w_i = w_i(\mathbf{a})$ approximates the solution

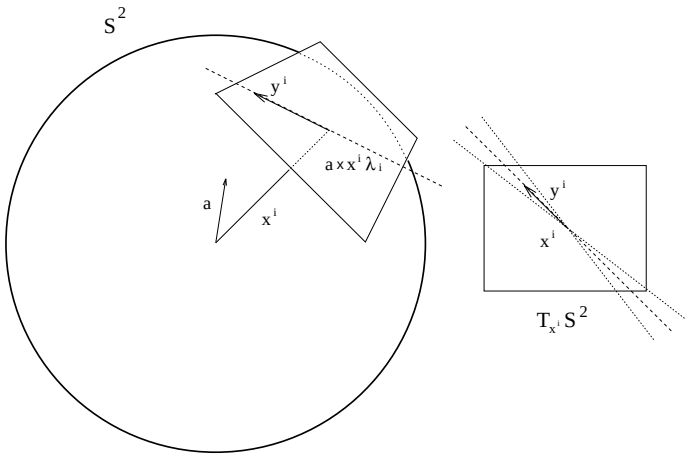

Figure 1: **Spherical projection model**

for $w_i = 1$ and can be used as a starting point for an iterative optimization.

In this and the following sections we use the "hat" operator $\widehat{\mathbf{x}}$ that establishes an isomorphism between $\mathbb{R}^3$ and the Lie algebra of skew-symmetric 3×3 matrices $so(3)$ as follows:

$$\widehat{\ }: \mathbb{R}^3 \to so(3) \; ; \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \widehat{\mathbf{x}} = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}. \tag{4}$$

This operator is used to represent the cross product between two vectors $\mathbf{x}_i$ and $\mathbf{x}_j$: $\widehat{\mathbf{x}}_i \mathbf{x}_j = \mathbf{x}_i \times \mathbf{x}_j$.

Since in equation (1) the parameters $\mathbf{a}$ and λ appear as a product, it is clear that they can only be recovered up to a common scale[1]. Therefore, we choose to normalize $\mathbf{a}$, so that $\|\mathbf{a}\| = 1$, although any other choice for the normalization would do. Note, as a caveat, that we are excluding from our problem the case $\mathbf{a} = 0$, for it does not allow us to gain any knowledge about λ. Using this notation, we can formulate the "Spherical Least-Squares Problem" **(SLS)** as

$$\mathbf{a}_{opt}, \lambda_{opt} = \arg \min_{\mathbf{a} \in \mathbf{S}^2, \lambda \in \mathbb{R}^p_+} \sum_{i=1}^{p} \|\mathbf{y}_i + \widehat{\mathbf{x}}_i \mathbf{a} \lambda_i\|^2_{w_i}. \tag{5}$$

We refer to the above sum as the *cost function* of SLS.

2.1 Weighted Spherical Least Squares

In this section we consider $w_i(\mathbf{a}) \doteq \|\mathbf{a} \times \mathbf{x}_i\| \in [0, 1]$. To motivate such a choice of weights, decompose each $\mathbf{x}_i$ into a component along $\mathbf{a}$, and a component orthogonal to it: $\mathbf{x}_i = \rho \mathbf{a} \overset{\perp}{+} \epsilon \mathbf{v}_i$, with $\rho, \epsilon \in \mathbb{R}$ and $\mathbf{v}_i \in \mathbf{S}^2$. Then we have $\mathbf{y}_i = \epsilon \mathbf{a} \times \mathbf{v}_i \lambda_i + \mathbf{n}_i$. If $\mathbf{x}_i$ is nearly parallel to $\mathbf{a}$, then ϵ is small, and for each measurement $\mathbf{y}_i$ we have a noise-to-signal ratio of (assume $\epsilon \geq 0$)

$$NSR = \frac{\|\mathbf{n}_i\|}{\epsilon \|\mathbf{a} \times \mathbf{v}_i\| \lambda_i} \geq \frac{\|\mathbf{n}_i\|}{\epsilon \lambda_i} \geq \frac{\|\mathbf{n}_i\|}{\epsilon}.$$

[1] If we multiply the components of $\mathbf{a}$ by a scale $\rho \neq 0$, and the vector λ by $1/\rho$ we obtain the same measurement vector $\mathbf{y}$.

Therefore, the more $\mathbf{x}_i$ is alighed with $\mathbf{a}$, the smaller ϵ, the more severe the effects of noise. If, instead, we consider the weighted norm $\|\mathbf{n}_i\|_{w_i}$, then $NSR = \frac{\|\mathbf{n}_i\|_{w_i}}{\epsilon} = \|\mathbf{n}_i\|$.

The solution of the SLS problem using the weights w_i is easily obtained as follows

Claim 2.1 *Given $p > 3$ points $\mathbf{x}_i \in \mathbf{S}^2$, $i = 1 \ldots p$ and the corresponding measurements $\mathbf{y}_i$, under general position conditions the SLS problem (5) with $w_i = w_i(\mathbf{a})$ admits a unique solution $\mathbf{a}_{opt}$ and λ_{opt} up to a sign. If we define M to be the symmetric 3×3 matrix*

$$M = \sum_{i=1}^{p} \mathbf{y}_i \mathbf{y}_i{}^T \tag{6}$$

then the optimal unit-norm solution $\mathbf{a}_{opt}$ is the eigenvector of M corresponding to its smallest eigenvalue:

$$\mathbf{a}_{opt} = \mathbf{v}_{min}(M) \tag{7}$$

and the optimal scales λ_{opt} are obtained from $\mathbf{a}_{opt}$ via

$$\lambda_{i\,opt} = -\frac{\mathbf{y}_i{}^T \widehat{\mathbf{x}}_i \mathbf{a}_{opt}}{\mathbf{a}_{opt}^T \widehat{\mathbf{x}}_i^2 \mathbf{a}_{opt}} \qquad i = 1 \ldots p. \tag{8}$$

Before proving the claim, we recall to the attention of the reader two simple but useful facts. For the vector $\mathbf{x} \in \mathbf{S}^2$, we call the *orthogonal projector* $\mathbf{x}^\perp$ the singular 3×3 matrix that projects a vector $\mathbf{y}$ onto the plane orthogonal to $\mathbf{x}$. It is immediate to verify that this operator is given by

$$\mathbf{x}^\perp = -\widehat{\mathbf{x}}^2. \tag{9}$$

Since all points on a plane are left unchanged by a projection onto that plane, we have

$$\mathbf{x}_i^\perp \mathbf{n}_i = \mathbf{n}_i \tag{10}$$

We can now proceed with proving claim 2.1.

Proof: *Consider the cost function of SLS, defined in equation (5). For any given $\mathbf{a}$, the λ that minimizes it is given, as a function of $\mathbf{a}$, by*

$$\lambda_i(\mathbf{a}) = -(\widehat{\mathbf{x}}_i \mathbf{a})^\dagger \mathbf{y}_i = -\frac{\mathbf{a}^T \widehat{\mathbf{x}}_i \mathbf{y}_i}{\mathbf{a}^T \widehat{\mathbf{x}}_i{}^2 \mathbf{a}} \qquad i = 1 \ldots p. \tag{11}$$

Once we substitute $\lambda(\mathbf{a})$ back into (5), the components of the cost function become $\|(\widehat{\mathbf{x}}_i \mathbf{a})^\perp \mathbf{y}_i\|$. We now use (9) to replace $(\widehat{\mathbf{x}}_i \mathbf{a})^\perp \mathbf{y}_i$ with $\frac{(\widehat{\mathbf{x}}_i \mathbf{a})}{\|\widehat{\mathbf{x}}_i \mathbf{a}\|} \times \mathbf{y}_i$, ending up with minimizing for $\mathbf{a}$ the cost function $\sum_i \frac{\|(\widehat{\mathbf{x}}_i \mathbf{a}) \times \mathbf{y}_i\|^2_{w_i}}{\|\widehat{\mathbf{x}}_i \mathbf{a}\|^2}$. Since $(\widehat{\mathbf{x}}_i \mathbf{a}) \times = \mathbf{a}\mathbf{x}_i{}^T - \mathbf{x}_i \mathbf{a}^T$, and $\mathbf{y}_i$ is orthogonal to $\mathbf{x}_i$, we can further simplify the SLS problem, that becomes

$$\arg\min_{\mathbf{a} \in \mathbf{S}^2} \sum_{i=1}^{p} \|\mathbf{x}_i \mathbf{a}^T \mathbf{y}_i\|^2 = \mathbf{a}^T \left(\sum_{i=1}^{p} \mathbf{y}_i \mathbf{y}_i{}^T \right) \mathbf{a}. \tag{12}$$

It is immediate to see that the least-squares unit-norm solution for $\mathbf{a}$ is given by the eigenvector of M corresponding to the smallest eigenvalue. Note that, since M is symmetric, the eigenvalues are real and positive, and the eigenvectors are unit-norm orthogonal vectors. Once the optimal $\mathbf{a}$ has been computed, the corresponding optimal λ can be obtained as $\lambda_i(\mathbf{a})$ from equation (11). Note that there is a sign ambiguity in $\mathbf{a}$, that reflects onto the sign of the vector λ. The two remaining eigenvectors of M correspond to a maximum and a saddle of the cost function in the SLS problem (see figure 2).

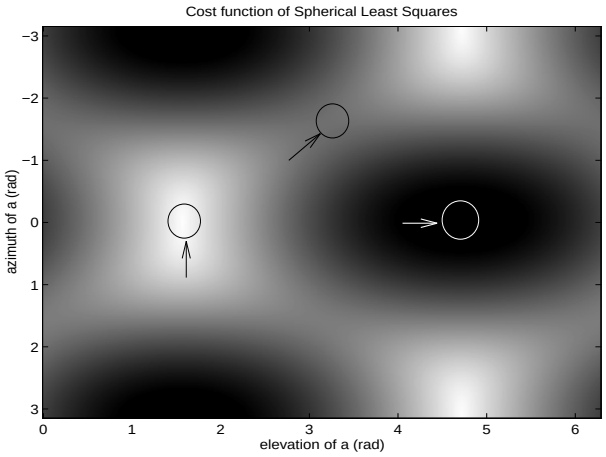

Figure 2: *Value of the cost function (5) for varying* $\mathbf{a}$*, expressed in spherical coordinates (azimuth and elevation). The plot is mirror-symmetric and therefore only one fourth of it is significant. There are three extrema indicated by arrows: a minimum (brighter region), a maximum (darker region) and a saddle (gray region connecting two bright regions). These three points identify orthogonal directions and correspond to the three eigenvectors of the matrix* M *in equation (6).*

2.2 Balanced Spherical Least Squares

Let us assume for a moment that the weights w_i are all identical to 1. We can still follow the procedure outlined in the proof of claim 2.1, but rather than ending up with minimizing the cost function $\sum_{i=1}^{p} \|\mathbf{x}_i \mathbf{a}^T \mathbf{y}_i\|^2$ in (12), we have $\sum_{i=1}^{p} \frac{\|\mathbf{x}_i \mathbf{a}^T \mathbf{y}_i\|^2}{\|\mathbf{x}_i \times \mathbf{a}\|^2}$, which can be re-written as

$$\sum_{i=1}^{p} \frac{< \mathbf{y}_i, \mathbf{a} >^2}{\|\mathbf{x}_i \times \mathbf{a}\|^2}. \tag{13}$$

If we assume that $\mathbf{x}_i$ only span a small solid angle (compared to the full sphere), then it makes sense to talk about an *average* direction $\bar{\mathbf{x}}$. If we weigh each point with $w_i = \frac{\|\mathbf{x}_i \times \mathbf{a}\|}{\|\bar{\mathbf{x}}_i \times \mathbf{a}\|}$, the cost function to be minimized becomes

$$\frac{\mathbf{a}^T M \mathbf{a}}{\mathbf{a}^T N \mathbf{a}} \tag{14}$$

where $M \doteq \sum_{i=1}^{p} \mathbf{y}_i \mathbf{y}_i^T$ and $N \doteq \hat{\bar{\mathbf{x}}}^2$. The solution for this problem has to do with Singular Rayleigh quotients and is derived in appendix A. We summarize the result in the following

Claim 2.2 *The solution* $\mathbf{a}_{opt}$ *for the problem of minimizing the cost function (13) is obtained by minimizing the corresponding Singular Rayleigh Quotient (14). The solution is given by the eigenvector of the matrix* $M_s \doteq M - \frac{M \bar{\mathbf{x}} \bar{\mathbf{x}}^T M}{\bar{\mathbf{x}}^T M \bar{\mathbf{x}}}$ *relative to the matrix* N *corresponding to the smallest non-zero eigenvalue.*

Proof: *See appendix A.*

A typical plot of the cost function (14) is shown in figure 16.

2.3 Unweighted Spherical Least Squares

When we choose all weights w_i to be identically equal to one, the problem of Spherical Least Squares can be reduced, following the proof of claim 2.1, to minimizing (13) subject to the constraints $< \mathbf{y}_i, \mathbf{x}_i >= 0$ and $\|\mathbf{x}_i\| = 1$. Naturally, in the absence of noise the solution to this problem is identical to the one of Balanced Spherical Least Squares and Weighted Spherical Least Squares seen in the previous sections. In the presence

of noise, we have not found a closed-form optimal solution for the Unweighted Spherical Least-Squares problem, nor we are aware of it having been found by others.

However, it is very simple to write a Newton-type iteration for the cost function (13), which is guaranteed to converge to a local extremum. For instance the following simple iteration can be easily proven to be contractive and therefore to converge to a fixed point[2]:

$$\mathbf{a}_{k+1} = \mathrm{pr}_{\mathbf{S}^2}\left(I - H^{-1}D^T\right)\mathbf{a}_k \tag{15}$$

where $H = \sum \frac{\mathbf{y}\mathbf{y}^T - 2\mathbf{a}\mathbf{a}^T\widehat{\mathbf{x}}^2 + \mathbf{a}^T\mathbf{y}\mathbf{y}^T\mathbf{a}\widehat{\mathbf{x}}^2 - 2\mathbf{a}\mathbf{a}^T\mathbf{y}\mathbf{y}^T}{(\mathbf{a}^T\widehat{\mathbf{x}}^2\mathbf{a})^2} - 4\frac{-\mathbf{y}\mathbf{y}^T\mathbf{a}\mathbf{a}^T\widehat{\mathbf{x}}^2 - \widehat{\mathbf{x}}^2\mathbf{a}\mathbf{a}^T\mathbf{y}\mathbf{y}^T\mathbf{a}\mathbf{a}^T\widehat{\mathbf{x}}^2}{(\mathbf{a}^T\widehat{\mathbf{x}}^2\mathbf{a})^3}$, $D = \sum \frac{\mathbf{y}\mathbf{y}^T(\mathbf{a}^T\widehat{\mathbf{x}}^2\mathbf{a}) - \widehat{\mathbf{x}}^2(\mathbf{a}^T\mathbf{y}\mathbf{y}^T\mathbf{a})}{(\mathbf{a}^T\widehat{\mathbf{x}}^2\mathbf{a})^2}$ and $\mathrm{pr}_{\mathbf{S}^2}$ denotes normalization (projection onto the sphere). Of course this iteration is only guaranteed to converge to a local solution of the Unweighted SLS. However, we have noticed that using a Weighted SLS or a Balanced SLS as an initialization step usually puts the iteration in the basin of attraction of the global minimum. In particular, for small levels of noise (up to 50% of the average signal), we have observed that both the Weighted SLS and the Balanced SLS approximate well the solution of the unweighted SLS, and therefore running an iteration of the type above improves only marginally the solution. For higher noise levels we have observed that when the deviation of the $\mathbf{x}_i$ is large, the solution to the Weighted SLS provides an accurate initialization, while when deviation of the $\mathbf{x}_i$ is small (on the order of 40^o or less), the solution of a Balanced SLS is more accurate. In both cases, however, the iteration above converges in few steps (less than 10).

3 Structure From Motion

Consider a variation of the Spherical Least-Squares problem of equation (5), where we add to the cost function the term $\widehat{\mathbf{x}}_i^2\mathbf{b}$, with $\mathbf{b} \in \mathbb{R}^3$ unknown. We call the resulting optimization problem *"Structure From Motion"* **(SFM)**:

$$\min_{\mathbf{a}\in\mathbf{S}^2, \mathbf{b}\in\mathbb{R}^3 \lambda\in\mathbb{R}^p_+} \sum_{i=1}^{p} ||\mathbf{n}_i||^2_{w_i} \quad \text{subject to} \quad \mathbf{y}_i = -\widehat{\mathbf{x}}_i\mathbf{a}\lambda_i + \widehat{\mathbf{x}}_i^2\mathbf{b} + \mathbf{n}_i \in T_{\mathbf{x}_i}\mathbf{S}^2, \quad i = 1\ldots p. \tag{16}$$

We refer to the function

$$r_0(\mathbf{a}, \mathbf{b}, \lambda) = \sum_{i=1}^{p} ||\mathbf{y}_i + \widehat{\mathbf{x}}_i\mathbf{a}\lambda_i - \widehat{\mathbf{x}}_i^2\mathbf{b}||^2 \tag{17}$$

as the *cost function* of SFM, and we neglect the subscripts w_i that indicate the choice of weights. This problem can be interpreted as that of estimating the direction of translation $\mathbf{a} \in \mathbf{S}^2$, the rotational velocity $\mathbf{b} \in \mathbb{R}^3$ and the depth $1/\lambda_i$ $i = 1\ldots p$ of a number p of moving points in 3-D, from the noisy projection of their velocity onto the retina, modeled as a unit-sphere. The estimation criterion is to minimize the "reprojection error", i.e. the difference between the measured optical flow and the one predicted by the choice of parameters, in a phylosophy similar to Prediction Error Methods [16].

In fact, the 3-D velocity of a point of coordinates $\mathbf{X}_i \in \mathbb{R}^3$ rotating about an axis parallel to $\mathbf{b}$ with angular velocity $||\mathbf{b}||$, and translating in the direction of $\mathbf{a}$ is given by $\mathbf{V}_i = \mathbf{b} \times \mathbf{X}_i + \mathbf{a}$. The projection of the point $\mathbf{X}_i$ onto the retina is $\mathbf{x}_i \doteq \frac{\mathbf{X}_i}{\|\mathbf{X}_i\|}$, and the velocity of such projection is given by

$$\mathbf{v}_i = \mathbf{x}_i^\perp\left(\mathbf{b} \times \mathbf{x}_i + \frac{\mathbf{a}}{||\mathbf{X}_i||}\right). \tag{18}$$

In order to obtain the equation in (16) it is sufficient to multiply both sides of (18) on the left by $-\widehat{\mathbf{x}}_i$, and define $\mathbf{y}_i \doteq -\widehat{\mathbf{x}}_i\mathbf{v}_i$ and $\lambda_i \doteq ||\mathbf{X}_i||^{-1}$ [3].

[2] A more sound way to proceed is to specify the iteration directly on the Sphere, by defining gradient flows on the Orthogonal group. This can be done (see for instance [3]), but is beyond the scope of this paper.

[3] We have adopted a spherical projection model for simplicity of notation. If a traditional pin-hole model is adopted instead, then $\mathbf{x}_\pi \doteq \frac{\mathbf{X}}{\mathbf{X}_3}$, and $\mathbf{y}_\pi = \mathcal{A}(\mathbf{x}_\pi)\frac{\mathbf{a}}{\mathbf{X}_3} + \mathcal{B}(\mathbf{x}_\pi)\mathbf{b}$, where $\mathcal{A}(\mathbf{x}_\pi) \doteq \begin{bmatrix} 1 & 0 & -x_{\pi 1} \\ 0 & 1 & -x_{\pi 2} \end{bmatrix}$ and $\mathcal{B}(\mathbf{x}_\pi) \doteq$

3.1 Pure translation

In the special case $\mathbf{b} = 0$ (no rotation), (16) reduces to a SLS problem. Depending upon the weights chosen, we have shown in section 2.3 how to obtain a solution (up to a sign) for the direction of translation $\mathbf{a}$ and the scaling parameters (inverse depths) λ_i.

3.2 General motion

In the presence of rotation $\mathbf{b} \neq 0$, the problem defined by equation (16) is no longer a standard SLS problem. Many in the Computer Vision literature have proposed methods to "undo" the rotation, by either warping the image (see for instance [15, 13, 2]), by applying a linear transformation to the measured data (for instance [18, 7]), or by estimating a rotational velocity assuming small translation (for instance [11]). In all cases, however, the transformation acts on the noise as well as on the data, therefore spoiling the goal of achieving an optimal estimate. The main reason behind these methods is that they lead to *linear* algorithms that work fairly well in the presence of small noise (for the case of [11] the assumption is used to initialize a recursive algorithm for multi-frame SFM). Here we take a different approach, that is somewhat more aware of noise and results in an optimal (although non-linear) estimator.

Following the derivation of the solution of SLS in claim 2.1, we can transform the problem of SFM into

$$\arg \min_{\mathbf{a}\in\mathbf{S}^2, \mathbf{b}\in\mathbb{R}^3} \sum_{i=1}^{p} \|\mathbf{x}_i \mathbf{a}^T (\mathbf{y}_i - \widehat{\mathbf{x}}_i^2 \mathbf{b})\|^2 \tag{19}$$

Now, for any given $\mathbf{b}$, the (unique) $\mathbf{a}$ that minimizes the norm of the cost function of SFM in (16) can be obtained by solving the SLS problem (5) with $\mathbf{y}_i$ being substituted by

$$\tilde{\mathbf{y}}_i = \mathbf{y}_i - \widehat{\mathbf{x}}_i^2 \mathbf{b}. \tag{20}$$

In the same fashion, given $\mathbf{a}$, the (unique) $\mathbf{b}$ that minimizes the norm of the $\mathbf{n}$ is obtained immediately from (19) as

$$\mathbf{b}(\mathbf{a}) = \left(\sum_{i=1}^{p} \widehat{\mathbf{x}}_i^2 \mathbf{a}\mathbf{a}^T \widehat{\mathbf{x}}_i^2 \right)^{-1} \sum_{i=1}^{p} \widehat{\mathbf{x}}_i^2 \mathbf{a}\mathbf{a}^T \mathbf{y}_i. \tag{21}$$

Therefore, the "conditional" problems of estimating $\mathbf{a}$ given $\mathbf{b}$ – or $\mathbf{b}$ given $\mathbf{a}$ – are particularly simple in that they admit a unique solution that can be computed in closed-form using linear algebra. Based on the simplicity of the conditional problems, one may be tempted to try the following

Bilinear Projection Algorithm (BPA)

- let $k = 0$ and choose any initial value for $\mathbf{b} \in \mathbb{R}^3$
- iterate the following (linear) computations:
  - $\mathbf{a}_k = \arg\min_{\mathbf{a}\in\mathbf{S}^2} \sum_{i=1}^{p} \|\mathbf{x}_i \mathbf{a}^T (\mathbf{y}_i - \widehat{\mathbf{x}}_i^2 \mathbf{b}_k)\|^2 \qquad \text{(SLS)}$
  - $\mathbf{b}_{k+1} = \left(\sum_{i=1}^{p} \widehat{\mathbf{x}}_i^2 \mathbf{a}_k \mathbf{a}_k^T \widehat{\mathbf{x}}_i^2\right)^{-1} \sum_{i=1}^{p} \widehat{\mathbf{x}}_i^2 \mathbf{a}_k \mathbf{a}_k^T \mathbf{y}_i.$
  - $k = k + 1$

We are implicitly excluding the case $\mathbf{a} = 0$, for that case can be detected and treated separately, as we show in section 5. It is straightforward to prove that this iteration converges "somewhere". It is less obvious to make sure that the iteration converges to some meaningful local extrema. Luckily we have the following

$\begin{bmatrix} -x_{\pi 1}x_{\pi 2} & 1+x_{\pi 1}^2 & -x_{\pi 2} \\ -1-x_{\pi 2}^2 & x_{\pi 1}x_{\pi 2} & x_{\pi 1} \end{bmatrix}$. Given a number of points and the corresponding vectors in the pin-hole coordinates, $(\mathbf{x}_\pi, \mathbf{y}_\pi)$, their equivalent in the spherical model can be easily obtained as $\mathbf{x} = \frac{\mathbf{x}_\pi}{\|\mathbf{x}_\pi\|}$, and $\mathbf{y} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \\ x_{\pi 2} & -x_{\pi 1} \end{bmatrix} \frac{\mathbf{y}_\pi}{\|\mathbf{x}_\pi\|^2}$.

Claim 3.1 *Given $p > 5$ points in general position, the Bilinear Projection Algorithm converges to a local extremum of the bilinear cost function* **(BCF)**

$$r(\mathbf{a}, \mathbf{b}) \doteq \sum_{i=1}^{p} \|\mathbf{x}_i \mathbf{a}^T (\mathbf{y}_i - \widehat{\mathbf{x}}_i^2 \mathbf{b})\|^2.$$

Extrema of the bilinear cost function r correspond to those of the original cost function of SFM in (16).

In order to prove the claim we need to establish that the bilinear iteration does not introduce "phantom" stationary points to the cost function. This is guaranteed by the following lemma:

Lemma 3.1 *Let $\psi(\mathbf{a})$ be the p-dimensional vector with i-th component $\mathbf{x}_i \mathbf{a}^T \mathbf{y}_i$, and ϕ the $p \times 3$ matrix with i-th row equal to $\mathbf{x}_i \mathbf{a}^T \widehat{\mathbf{x}}_i^2$. Then $r(\mathbf{a}, \mathbf{b}) = \|\psi(\mathbf{a}) - \phi(\mathbf{a})\mathbf{b}\|$ and define $r_2(\mathbf{a}) \doteq \|\phi(\mathbf{a})^{\perp} \psi(\mathbf{a})\|$ [4]. Furthermore, assume that ϕ has constant rank $\rho = 3$ in some open subset Ω of $\mathbb{R}^p$.*

- *If $\mathbf{a}^*$ is a critical point (or a global minimizer) of r_2, and $\mathbf{b}^* \doteq \phi^{\dagger}(\mathbf{a}^*)\psi(\mathbf{a}^*)$, then $(\mathbf{a}^*, \mathbf{b}^*)$ is a critical point (or a global minimizer) of r and $r_2(\mathbf{a}^*) = r(\mathbf{a}^*, \mathbf{b}^*)$.*
- *If $(\mathbf{a}^*, \mathbf{b}^*)$ is a global minimizer of r, then $\mathbf{a}^*$ is a global minimizer of r_2 and $r_2(\mathbf{a}^*) = r(\mathbf{a}^*, \mathbf{b}^*)$.*

Here $\dagger$ denotes the (least-squares) pseudo-inverse.

Proof: *This is a a subcase of theorem 2.1 on page 416 of [6], with the simple extension of allowing the affine term ψ to depend on $\mathbf{a}$.*
Proof of claim 3.1: *The Bilinear Projection Algorithm is a Gauss-Newton iteration for the cost function $r(\mathbf{a}, \mathbf{b})$. In fact, the lemma guarantees that the iteration performed by alternating the variables $\mathbf{a}$ and $\mathbf{b}$ has the same fixed points of the iteration performed simultaneously on $\mathbf{a}$ and $\mathbf{b}$. The first part of the claim follows from standard properties of Gauss-Newton iterations. That such extrema correspond to those of the original cost function in (16) follows by applying the lemma again to the SLS problem.*

Remark 3.1 *Owing to the bilinear nature of the cost function r, each iteration of the Bilinear Projection Algorithm is not just a* step *along the gradient of the cost r but, rather, the minimization restricted to a slice of the parameter set. Normally the convergence rate of Gauss-Newton iterations is quadratic. However, due to the bilinear nature of the cost function r, the BPA may exhibit higher convergence rates.*

Some comments are now in order.

- The above claim guarantees that, following the BPA outlined in this paragraph, we do not introduce spurious solutions to the problem of SFM.
- It has been known for a long time in the Computer Vision community (see for instance [4]) that the problem of SFM *in the absence of noise* has a *unique solution* up to a sign and a scale ambiguity for the parameters $\mathbf{a}$ and λ (under the usual general position conditions).
- However, it is still possible that the *original problem* of SFM defined in (16) admits local minima. In which case the Bilinear Projection Algorithm could converge to such minima. Again, we emphasize that these minima are *intrinsic* to the problem, and not artifacts introduced by the BPA.
- The existence of unexpected local minima of the BPA would be bad news from the point of view of Computational Vision, because it would mean that SFM is intrinsically hard. However, they would be good news from the perspective of Visual Psychophysics, for local interpretations provided by the Bilinear Projection Algorithm could be tested on human subjects.

Therefore, it remains to be established whether the original problem of SFM, as formulated in (16), admits spurious local solutions *in the presence of noise* $\mathbf{n}$. This is the subject of the next section.

[4] Here the notation $\phi^{\perp}$ stands for the projector operator onto the orthogonal complement to the range space of ϕ, defined as $\phi^{\perp} = I - \phi\phi^{\dagger}$, where $\dagger$ denotes the pseudo-inverse.

4 "Bas-relief ambiguity", "rubbery motion percept" and a robust representation of shape

Let us summarize our current understanding in the problem of optimal SFM (16). In the absence of noise, we know it has a unique solution (up to the usual sign and scale ambiguity) [4]. In the presence of noise, we have a technique (the Bilinear Projection Algorithm) that is guaranteed to converge to a local extremum of the cost function of SFM (16). We also know that such local extrema, if they exist, are "fundamental", and not an artifact of the particular algorithm we have proposed in section 3.

The most natural questions that are left to answer are

1. does the *problem of SFM* admit spurious solutions in the presence of noise?
2. if so, is it possible to *classify* them?
3. is it possible to *detect* the fact that a local extremum is not the correct solution to SFM?
4. are there particular representations of the unknown parameters that are *invariant*, i.e. common to the spurious extrema as well as the correct solution?
5. given a spurious solution, is it possible to recover the correct solution?

In order to answer some of these questions, one can set up a random sampling simulation and check the cost function for local minima in the presence of noise levels up to 100%. Note that, by virtue of lemma 3.1, it is equivalent to check any of the cost functions $r_0(\mathbf{a}, \mathbf{b}, \lambda)$, $r(\mathbf{a}, \mathbf{b})$ or $r_2(\mathbf{a})$, defined in (17) and in lemma 3.1 respectively. This makes the random search particularly favorable for $r_2(\mathbf{a})$, since it only depends upon 2 parameters, and can therefore be visualized, as we will see in section 6.

In an extensive session of Montecarlo simulations, we have tested the convergence of the bilinear iteration for random parameters $\mathbf{a}, \mathbf{b}$, starting from initial conditions distributed uniformly around the true parameters in increasingly large intervals. Noise up to 100% was added to the measurements.

We have indeed experienced convergence to local extrema of the BPA, and we have verified that these local extrema are common to the original cost function r_0 of (16), as expected from claim 3.1. The answer to question 1 is therefore positive: there do exist local extrema.

Furthermore, we have observed that local extrema tend to aggregate into 8 groups, 7 of which occur more frequently. One of the extrema obviously corresponds to the true parameters, 4 are local minima, and the remaining 2 divide into a maximum and a saddle.

In addition to visualizing the residual cost functions r_2 (or two-dimensional slides of the 5-dimensional r, or $(5+p)$-dimensional r_0), which we do in the experimental section, we are going to interpret analytically the experimental results, by analyzing the the local extrema.

Let us consider the bilinear cost function described in claim 3.1. Since for $\mathbf{b} = 0$ the weighted SLS problem has 3 extrema (section 2.3 and figure 2), it is tempting to conjecture that there may be two extrema for $\mathbf{b}$ (one corresponding to the true value, and one to $\mathbf{b} = 0$), and corresponding to each extremum of $\mathbf{b}$ there could be three extrema for $\mathbf{a}$, associated with the solution of the corresponding weighted SLS. We will see that this is indeed the case.

Remark 4.1 *The claims contained in this section are "weak", in the sense that they rely on our observation that, for noise levels up to* 100%*, local extrema tend to cluster into 8 groups. In the next few sections we are going to give an analytical interpretation of these extrema. This does not prove that, for higher noise levels or for particular configurations of the points in space, these will be the only local extrema. Indeed, in all what follows we assume that points and measurements are generic, and they do not fall into singular configurations.*

Remark 4.2 *The presence of local minima to the problem of SFM has been long known in the Computer Vision community (see for instance [21]). The existence of few local minima has been reported by few other researchers. For instance, [19] observe convergence to* "typically fewer than 5"*. They also report convergence of their algorithms to a translational velocity orthogonal to the true one, and give a qualitative explanation of this phenomenon. Furthermore, visual illusions experimented in psychophysical experiments provide a basis to study the local minima of SFM.*

4.1 The "bas-relief" ambiguity

The first situation we consider describes a phenomenon well-known as the "bas-relief ambiguity". Consider looking at an object that occupies a small portion of the visual field, and suppose that it rotates about an axis orthogonal to the line of sight passing through the centroid of the object (see figure 3). This

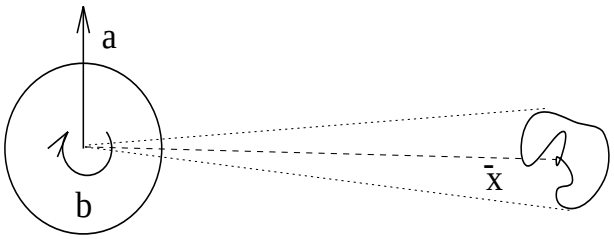

Figure 3: **Conditions for the bas-relief ambiguity**

situation gives raise to a faulty percept, in that "the rotational component of motion is confused with its translational component". As a particular case, when a "thin" object rotates with constant velocity under these conditions, it is perceived as moving with non-constant velocity: slower when the facing the viewer, faster when orthogonal to it (the so-called "rotating billboard effect"). Even more strikingly, when two such objects are connected rigidly and oriented perpendicularly to each other, they are perceived as disconnected and moving somewhat independently (see figure 4). This effect, commonly observed in psychophysical

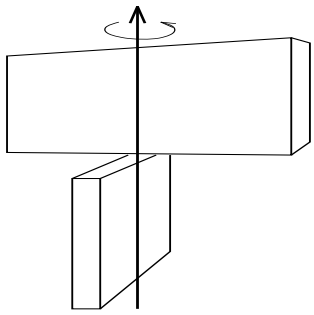

Figure 4: **Rotating billboards**

experiments, has been analyzed previously by assuming an affine approximation of the perspective projection model [8]. The fact that the solution of SFM in the case of (noiseless) perspective does not account for this effect (for the solution is unique) has led some to speculate that the human visual system somehow encodes an affine projection model. In this section we show that this needs not be the case, for the bas-relief ambiguity can be explained as a side-effect of noise, without need to resort to approximations of the projection model.

Remark 4.3 *One of the common complaints to SFM algorithms from perspective is that they become unreliable in the presence of small fields of view, and when the rotational component of motion is "confused" with the translational component (see for instance [21]). It is our goal in this section to formalize this concept and establish that this effect is "intrinsic", and not an artifact of the algorithm. We will also discuss what information can be robustly recovered under these conditions.*

Remark 4.4 *The statements in [21] regarding the intrinsic difficulty in solving SFM under the conditions just described seem to be in contrast with the observation made in [19], that rotation is irrelevant as far as the estimation of SFM is concerned. We will see that the two statements are not contradictory, and can be reconciled by analyzing the problem from the perspective of the measurement noise.*

Noise and the bas-relief ambiguity

When an object occupies a small portion of the visual field, $\mathbf{x}_i$ tend to be similar. We call their average direction $\bar{\mathbf{x}}$, which corresponds to the line of sight. Rotation about an axis orthogonal to the line of sight passing through the object corresponds to $\mathbf{a}$ and $\mathbf{b}$ being orthogonal to each other and both orthogonal to $\bar{\mathbf{x}}$ (see figure 3). We will now show that these conditions, in presence of noise, give rise to a minimum of the cost function (16).

Claim 4.1 *Let* $\mathbf{y}_i = -\widehat{\mathbf{x}}_i\mathbf{a}\lambda_i + \widehat{\mathbf{x}}_i^2\mathbf{b} + \mathbf{n}_i$, *and* $\bar{\mathbf{x}} \perp \mathbf{a} \perp \mathbf{b}$ *as described above. Furthermore, let* $\mathbf{b} = \mathbf{b}_0 + \delta\mathbf{b}$, *with* $\|\delta\mathbf{b}\| \cong \|\mathbf{n}\|$, *the average norm of the measurement error. Then the cost function* $r_0(\mathbf{a}, \mathbf{b}, \lambda) \doteq \sum_{i=1}^{p} \|\mathbf{y}_i + \widehat{\mathbf{x}}_i\mathbf{a}\lambda_i - \widehat{\mathbf{x}}_i^2\mathbf{b}\|^2$ *has a local extremum at* $\tilde{\mathbf{b}} = \mathbf{b}_0$, $\tilde{\lambda} = \lambda - \bar{\lambda}$, *and* $\tilde{\mathbf{a}} = \mathbf{v}_{min}(\sum_{i=1}^{p} \mathbf{y}_i\mathbf{y}_i{}^T)$, *where the bar denotes the average, and* $\mathbf{v}_{min}$ *denotes the eigenvector corresponding to the smallest eigenvalue.*

Proof *Without loss of generality, assume* $\mathbf{b}_0 = 0$, *so that* $\mathbf{b} = \delta\mathbf{b}$ *is of size comparable to the noise:* $\|\mathbf{b}\| \cong \|\mathbf{n}\|$. *Since* $\bar{\mathbf{x}} \perp \mathbf{b}$ *and* $\mathbf{x}_i \in \mathbf{S}^2$, *we have*

$$\|\widehat{\mathbf{x}}_i^2\mathbf{b}\| \cong \|\mathbf{b}\| \quad \forall i = 1 \dots p \tag{22}$$

and therefore the terms $\widehat{\mathbf{x}}_i^2\mathbf{b}$ *are comparable with the noise* $\mathbf{n}_i$, *and they can be lumped as a bias into*

$$\tilde{\mathbf{n}}_i = \widehat{\mathbf{x}}_i^2\mathbf{b} + \mathbf{n}_i. \tag{23}$$

The value of $\mathbf{b}$ *that minimizes the norm of* $\tilde{\mathbf{n}}$ *is* $\tilde{\mathbf{b}} = 0$, *and the corresponding* $\tilde{\mathbf{a}}$ *is obtained as the solution to the SLS problem (5), e.g.* $\tilde{\mathbf{a}} = \mathbf{v}_{min}(\sum_{i=1}^{p} \mathbf{y}_i\mathbf{y}_i{}^T)$. *In order to evaluate the corresponding* $\tilde{\lambda}$, *we observe that*

$$\begin{aligned} \lambda_i &= -(\widehat{\mathbf{x}}_i\mathbf{a})^\dagger\mathbf{y}_i + (\widehat{\mathbf{x}}_i\mathbf{a})^\dagger\widehat{\bar{\mathbf{x}}}^2\mathbf{b} \\ &\cong \lambda_i^0 + (\widehat{\bar{\mathbf{x}}}\mathbf{a})^\dagger\widehat{\bar{\mathbf{x}}}^2\mathbf{b} \\ &= \lambda_i^0 + \bar{\lambda} \end{aligned}$$

where λ_i^0 *are obtained assuming* $\mathbf{b} = 0$ *and* $\bar{\lambda}$ *is the average of the scales* λ_i. *Therefore, the scales corresponding to the local solution* $\mathbf{b} = 0$ *are the zero-mean version of the original ones*

$$\tilde{\lambda}_i \cong \lambda_i - \bar{\lambda}. \tag{24}$$

Remark 4.5 *As a consequence of the above claim, we can conclude that the representation of the "shape" of the points using the centered inverse depths* $\tilde{\lambda}_i$, $i = 1 \dots p$ *is robust to the bas-relief ambiguity, since it is invariant with respect to* $\delta\mathbf{b}$.

Remark 4.6 *The Bilinear Projection Algorithm does not enforce the fact that* $\lambda_i > 0$. *As a result, the algorithm can converge to a local interpretation of the parameters where some of the* λ_i *are negative. This is indeed the case in the bas-relief ambiguity, as we have seen in the previous claim. We can use this as a test to detect whether a local extremum corresponds to the bas-relief ambiguity, as we will see later.*

Remark 4.7 *From the claim we can conclude that, in presence of high noise levels, a portion of the rotational velocity* $\mathbf{b}$ *can be confused with noise and compensated by a "bias" in the translational velocity* $\mathbf{a}$, *and the corresponding inverse depths* λ_i *are scaled towards the origin. This statement confirms the observations of Weng et al. [21] – that there exist conditions for which small rotations are confused with small translations – although they attribute the effect to the geometry of the epipolar constraint. Similar observations were also made by Oliensis [11].*

Remark 4.8 *The consequence of claim 4.1 seem in contrast with the observations of [19], that the axis of rotation has no impact on the bias of the estimate of translation. However, in order for the bas-relief effect to show, the conditions of claim 4.1 must be met, in particular the aperture angle must be small, the noise level must be high and the algorithm must be initialized far away from the true solution. Some of these conditions are not explored in the experimental setup of [19] (the aperture angle was of* 90^o *throughout the*

experiments, and the noise level of 0.5 *pixels), and therefore the effect is not observed. In section 6, we have tested the residual of the cost function that is common to some of the algorithms considered in [19] under the conditions of claim 4.1 (aperture angle of* 20^o*, 5 pixel noise), and it did show a bias corresponding to the bas-relief ambiguity (see section 6). In so-called "linear methods" that eliminate rotation (as in the schemes described in [19]), the manifestation of the bas-relief ambiguity is a "bias" of the direction of translation. In [18], it is observed that their (linear) algorithm is unbiased in presence of fields of view of* 180^o*, although the statement is not motivated analytically. The explanation lies in the analysis of the bas-relief ambiguity, that does not manifest itself with large fields of view.*

4.2 Other extrema

In the proof of claim 4.1 we have computed the local minimum $\tilde{\mathbf{a}}$ associated with $\tilde{\mathbf{b}} = 0$ as the solution of the corresponding weighted SLS problem. Such a solution is the eigenvector of the matrix M (defined in equation (6)) corresponding to its smallest eigenvalue. In the absence of noise, such eigenvalue is 0. For small noise levels, it still is distinguishably smaller than the remaining two. However, in the presence of large noises, the eigenvalues become comparable and therefore the actual solution of the SLS problem can be any of the three eigenvectors of M. This indeed happens – for large noise levels – and it accounts for the three extrema of the cost function found experimentally.

4.3 Detecting local minima and switching between extrema

We have established that there are (at least) two extrema of the bilinear cost function (BCF) for $\mathbf{b}$: one corresponding to the true solution, and one corresponding to the bas-relief ambiguity $\tilde{\mathbf{b}} = 0$. Correspondingly, there are three extrema for $\mathbf{a}$, the eigenvectors of the matrix M. Out of these three extrema, there is a minimum, a saddle, and a maximum, as we have seen in section 2.3.

It is in fact possible to detect whether a stationary point $\tilde{\mathbf{a}}$ is a local minimum or not. In fact, given $\tilde{\mathbf{a}}$, we can compute its orthogonal complement (the remaining two eigenvectors of M), and compute the corresponding residual. We then just choose the eigenvector that carries the smallest residual. By doing so, we rule out 4 local extrema, and we are left with 2 possibilities: $\tilde{\mathbf{b}} = \mathbf{b}$, or $\tilde{\mathbf{b}} \cong 0$, the bas-relief ambiguity.

As it turns out, this situation can also be detected easily, even without knowing the noise level $\|\mathbf{n}\|$ (which gives a lower bound on the residual). In fact, as we have observed in the previous section, the correct $\mathbf{b}$ leads to reconstructed scales λ that are positive, while in the bas-relief ambiguity they are biased towards the origin and, as long as not all λ_i are equal (i.e. when the structure is a perfect fronto-parallel plane), some $\tilde{\lambda}_i$ will be negative, as a consequence of claim 4.1. Note that even in the latter case, which corresponds to the bas-relief ambiguity, it is still possible to retrieve a useful representation of shape, for $\tilde{\lambda}$ can be re-scaled by choosing ρ so that $\tilde{\lambda} + \rho > 0$, which leads to the *bas-relief*. These scaled parameters can be chosen to effectively re-initialize the algorithm, as we will see in the experimental section.

4.4 "Rubbery motion" percept

An interesting phenomenon that has been observed in psychophysical experiments occurs under the same conditions of the bas-relief ambiguity. However, instead of rotational velocity being underestimated, it is perceived as being the opposite of the true one. For instance, a convex object rotating clockwise is perceived as being a concave object rotating counter-clockwise (this phenomenon is heavily exploited in visual illusions), or a flat object (such as a plane or a Necker cube) can be perceived as "flipped" (the farther face is seen as being upfront and viceversa).

In order to analyze this phenomenon from the point of view of noise, let us consider the same conditions of the bas-relief ambiguity as expressed in claim 4.1, and assume that $\tilde{\mathbf{b}} = -\mathbf{b}$, and $\tilde{\lambda}_i = -\lambda_i^0 - \bar{\lambda}$. In order for this to be a legitimate local extremum, as a consequence of lemma 3.1, there must exist some $\tilde{\mathbf{a}}$ that makes the noise $\tilde{\mathbf{n}}_i$ small, where

$$\mathbf{y}_i = -\widehat{\mathbf{x}}_i \tilde{\mathbf{a}} \tilde{\lambda}_i + \widehat{\mathbf{x}}_i^2 \tilde{\mathbf{b}} + \tilde{\mathbf{n}}_i. \tag{25}$$

if we substitute the expressions for $\tilde{\lambda}_i$ and $\tilde{\mathbf{b}}$ we get

$$\mathbf{y}_i \cong \widehat{\mathbf{x}}_i \tilde{\mathbf{a}} \lambda_i^0 - (\widehat{\tilde{\mathbf{a}}}\bar{\mathbf{x}})(\widehat{\mathbf{a}}\bar{\mathbf{x}})^\dagger \widehat{\tilde{\mathbf{x}}}^2 \mathbf{b} - \widehat{\tilde{\mathbf{x}}}^2 \mathbf{b} + \tilde{\mathbf{n}}_i. \tag{26}$$

From that expression is it possible to see that, under the assumptions of the bas-relief ambiguity, $\tilde{\mathbf{a}} = -\mathbf{a}_0$, where $\mathbf{a}_0$ is the solution to the SLS problem obtained by assuming $\mathbf{b} = 0$. In fact, in that case we have

$$\mathbf{y}_i \cong \widehat{\mathbf{x}}_i \tilde{\mathbf{a}} \lambda_i^0 - (\widehat{\mathbf{a}}\bar{\mathbf{x}})^\perp \widehat{\bar{\mathbf{x}}}^2 \mathbf{b} + \tilde{\mathbf{n}}_i \tag{27}$$

but $\widehat{\bar{\mathbf{x}}}^2\mathbf{b} \cong \mathbf{b}$ and $\widehat{\mathbf{a}}\bar{\mathbf{x}} \cong \frac{\mathbf{b}}{\|\mathbf{b}\|}$, so that $(\widehat{\mathbf{a}}\bar{\mathbf{x}}) \times (\widehat{\bar{\mathbf{x}}}^2\mathbf{b}) = 0$ and the second term in the above expression is negligible:

$$\mathbf{y}_i \cong \widehat{\mathbf{x}}_i \tilde{\mathbf{a}} \lambda_i^0 + \tilde{\mathbf{n}}_i. \tag{28}$$

The expression for $\tilde{\mathbf{a}}$ that minimizes the sum of the norms of $\tilde{\mathbf{n}}_i$ is obtained as the solution of the SLS problem associated to (28).

Remark 4.9 *Note that this solution, which we call the "rubbery motion" effect, is* not *just the correct solution up to a sign, for that would correspond to $\tilde{\lambda}_i = -\lambda_i$, while here we have $\tilde{\lambda}_i = \lambda_i - 2\lambda_i^0$. The reason for the name "rubbery" comes from the fact that this solution is not consistent with the hypothesis of rigid motion, and therefore the noise $\tilde{\mathbf{n}}_i$ must be large enough to accomodate deviations from this hypothesis.*

4.5 Enforcing positive depth

An immediate consequence of the above considerations is that, if we can devise an algorithm that enforces the fact that the scale parameters are positive, then it would eliminate the "rubbery" percept and the "bas-relief" ambiguity since they both produce (some or all, respectively) negative scales.

However, enforcing $\lambda_i > 0$ requires solving an optimization problem with inequality constraints, which cannot be implemented in the simple and fast way of the Bilinear Projection Algorithm.

4.6 A robust representation of shape

The averaged inverse depths is invariant under the bas-relief ambiguity. The rubbery motion percept consists in a sign change of the averaged inverse depths. Therefore, the averaged inverse depths represent shape up to a global scaling factor and a sign, and it is invariant under the bas-relief ambiguity and the rubbery motion percept.

5 Algorithm and Implementation

5.1 Recipe Algorithm

For the purpose of reference, we report here recipe for the implementation of the algorithm. We are given p unit-norm vectors $\mathbf{x}_1, \ldots, \mathbf{x}_p$, and the corresponding measurements $\mathbf{y}_1, \ldots, \mathbf{y}_p$.

Check for data consistency: Choose a threshold ϵ close to the machine precision and check that $\mathbf{x}_i^T \mathbf{y}^i < \epsilon$. If not, the measurement are given incorrectly or have been generated by a different projection model.

Check for pure rotation: Let ϕ be the $3 \times (3p)$ matrix with blocks $\widehat{\mathbf{x}}_i^2$, and $\mathbf{Y}$ the vector obtained by stacking $\mathbf{y}_i$ on top of each other. Choose a threshold γ and Compute $\mathbf{b} = \phi^\dagger \mathbf{Y}$. If the residual $\phi^\perp \mathbf{Y}$ is less than γ stop, for pure rotation is a good fit. Otherwise proceed with the initialization.

Initialization: let $k = 0$ and initialize the algorithm with $\mathbf{b}_0$. Among the possible choices are

- $\mathbf{b}_0 = 0$ (this puts the algorithm close to the solution corresponding to the bas-relief ambiguity)
- $\mathbf{b}_0$ as computed by assuming pure rotation (see "Check for pure rotation" above)
- Compute $\mathbf{b}_0$ and $\mathbf{a}_0$ assuming $\lambda_i = 1 \ \forall i$ (i.e. all points are on the sensor's surface)
- use the output of any "linear algorithm"
- choose $\mathbf{b}_0$ at random.

We have noticed no significant difference between these choices, for all of them are equally bad in the case of large noise. In the case of small noise, using the output of a linear algorithm helps because it places the starting point close to the true solution. But the important point is that the rest of the algorithm is fairly insensitive to initialization.

Bilinear Iteration: Choose a threshold δ for convergence, and iterate until the difference of the residuals at successive iterations is less than δ or in any case for no more than 100 iterations

- $\mathbf{a}_k = \arg\min_{\mathbf{a}\in\mathbf{S}^2} \sum_{i=1}^p \|\mathbf{x}_i \mathbf{a}^T (\mathbf{y}_i - \widehat{\mathbf{x}}_i^2 \mathbf{b}_k)\|_{w_i}^2$ (SLS)
- $\mathbf{b}_{k+1} = \left(\sum_{i=1}^p \widehat{\mathbf{x}}_i^2 \mathbf{a}_k \mathbf{a}_k^T \widehat{\mathbf{x}}_i^2\right)^{-1} \sum_{i=1}^p \widehat{\mathbf{x}}_i^2 \mathbf{a}_k \mathbf{a}_k^T \mathbf{y}_i$.
- $k = k + 1$

Check for local extrema: Once the algorithm has converged after, say, $k = conv$ steps, the algorithm returns the matrix $M = \sum_{i=1}^p (\mathbf{y}_i - \widehat{\mathbf{x}}_i^2 \mathbf{b}_k)(\mathbf{y}_i - \widehat{\mathbf{x}}_i^2 \mathbf{b}_k)^T$ used to compute $\mathbf{a}_k$. Compute the residual corresponding to all 3 eigenvectors of M and choose $\mathbf{a}$ as the one that generates the smallest residual.

Check for bas-relief ambiguity: check $\lambda_{conv} > 0$. If so, stop. Otherwise, λ_k corresponds to the bas-relief estimate of depth and/or to the rubbery motion interpretation.

- choose ρ such that $\lambda_\rho = \rho + \lambda_{conv} > 0$. This is a normalized estimate of shape.
- use the normalized λ_ρ to estimate a normalized translation $\mathbf{a}_\rho$ and a rotation $\mathbf{b}_\rho$ by solving the linear least-squares problem defined by the cost function r_0 in (17).
- use $\mathbf{b}_\rho$ to initialize the algorithm again (go back to "Initialization").

Verdict: If the algorithm has converged to a solution that admits positive depth, then stop. If it converges to the same estimate as before re-initialization, it means that the local extremum corresponding to the bas-relief ambiguity is dominant (observability of $\|\mathbf{b}\|$ is lost), and the normalized shape estimates corresponding to λ_ρ are the only useful outcome of the algorithm.

6 Performance Assessment

How much noise is too much noise?

Typical feature-tracking/optical-flow algorithms declare accuracy in locating corresponding feature-points in the order of 0.1 pixels [1]. It is our experience that this is indeed the case for about 30% of the feature-points extracted automatically according to a SSD criterion [10]. However, when considering all features, a more realistic figure for the localization error is 1 pixel.

Now consider a camera with a 30^o field of view and an imaging sensor of 512×512 pixels, translating forward at $0.5m/s$. Depending upon the scene being viewed, the average norm of the flow vectors on the imaging sensor is in the order of 1-2 pixels (for a 15 frames/second capture rate). Therefore, an error in the order of 1 pixel corresponds to 50%-100% of the measurements.

Consider again the camera just described, but now looking at an object that is $2m$ ahead of the camera and rotating about an axis passing through its centroid at $1^o/s$. In this case the average norm of flow vectors is 0.1 pixels/frame, and 1 pixel error corresponds to an untolerable 1000%.

The point of these simple calculations is to motivate the study of SFM from the point of view of noise: even the use of the most accurate feature-tracker does not dispense us from dealing with noise in scenarios that are very often encountered in real-world situations.

6.1 Sample tests on real images

Most image sequences publically available on the Web have been used to demonstrate the performance of existing algorithms. Since the algorithm described in section 3 is optimal by construction, we should expect it to work at least as well as any other algorithm. Most real image sequences available do not have a reliable ground-truth, and therefore a fair comparison is impossible. In section 6.5, however, we report the results

of a simulation to compare the optimal algorithm versus schemes based upon epipolar geometry and linear subspace constraints.

Here we just report the use of the algorithm on a real image sequence for the sake of example. We have chosen as a sample experiment a "box-sequence" that was available in Matlab format with calibration data (figure 5), since the motion pattern is the one leading to the bas-relief ambiguity. The box rotates about a vertical axis at a rate of $3^o/frame$, which is a fairly large motion. Under these conditions even the 8-point algorithm of Longuet-Higgins works.

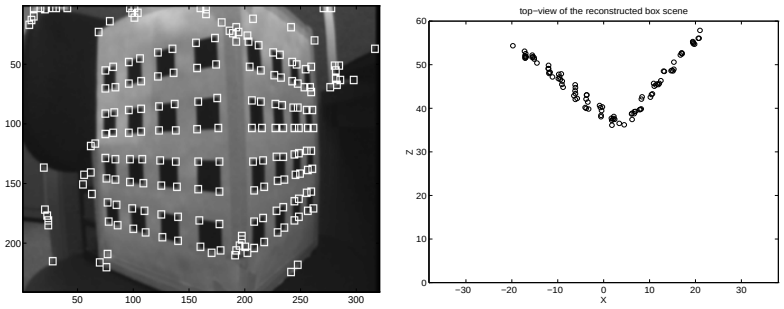

Figure 5: **Box scene** *A box rotates about a vertical axis at $3^o/frame$. The top-view of the reconstructed scene is shown on the right in normalized units (units of translational velocity). No reliable ground-truth is available.*

6.2 Sample of convergence behavior during simulations

In figure 6 we show a typical case where the iteration converges to the global minimum. The residual decreases up to the level of the noise (left), and both the parameters $\mathbf{a}, \mathbf{b}$ (center) and the scales λ (right) are within the noise level from the true values. However, sometimes the residual stabilizes at a level different

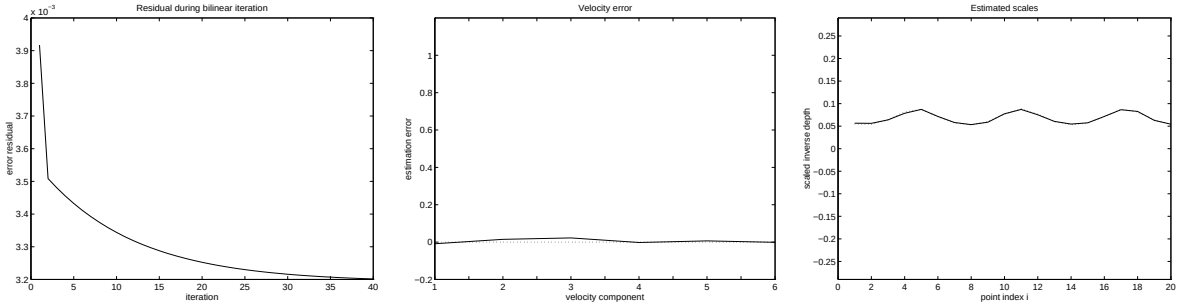

Figure 6: **Convergence to the global minimum** *(left) residual cost function, (center) parameter estimation error, (right) estimated scales. Ground truth is in dotted lines.*

from the noise level, as in figure 7 (top-row left), but both the parameters (center) and the scales (right) are quite far from their true values. This is a clear sign that the algorithm has converged to a local extremum. However, if we check all three eigenvectors of the matrix M in (6) and the scales λ that they generate (7 middle-row), we see that one of them (center) produces an estimate that corresponds to the averaged version of the correct scales. Therefore, there has been a switch of the eigenvalues of M. We can now switch to the solution for $\mathbf{a}$ and λ corresponding to the eigenvector that generates the smallest residual, and use that as an initial condition for a second run of the algorithm, that converges in 5 iterations to the correct estimate (7 bottom-row). In figure 8 we show convergence to the bas-relief ambiguity.

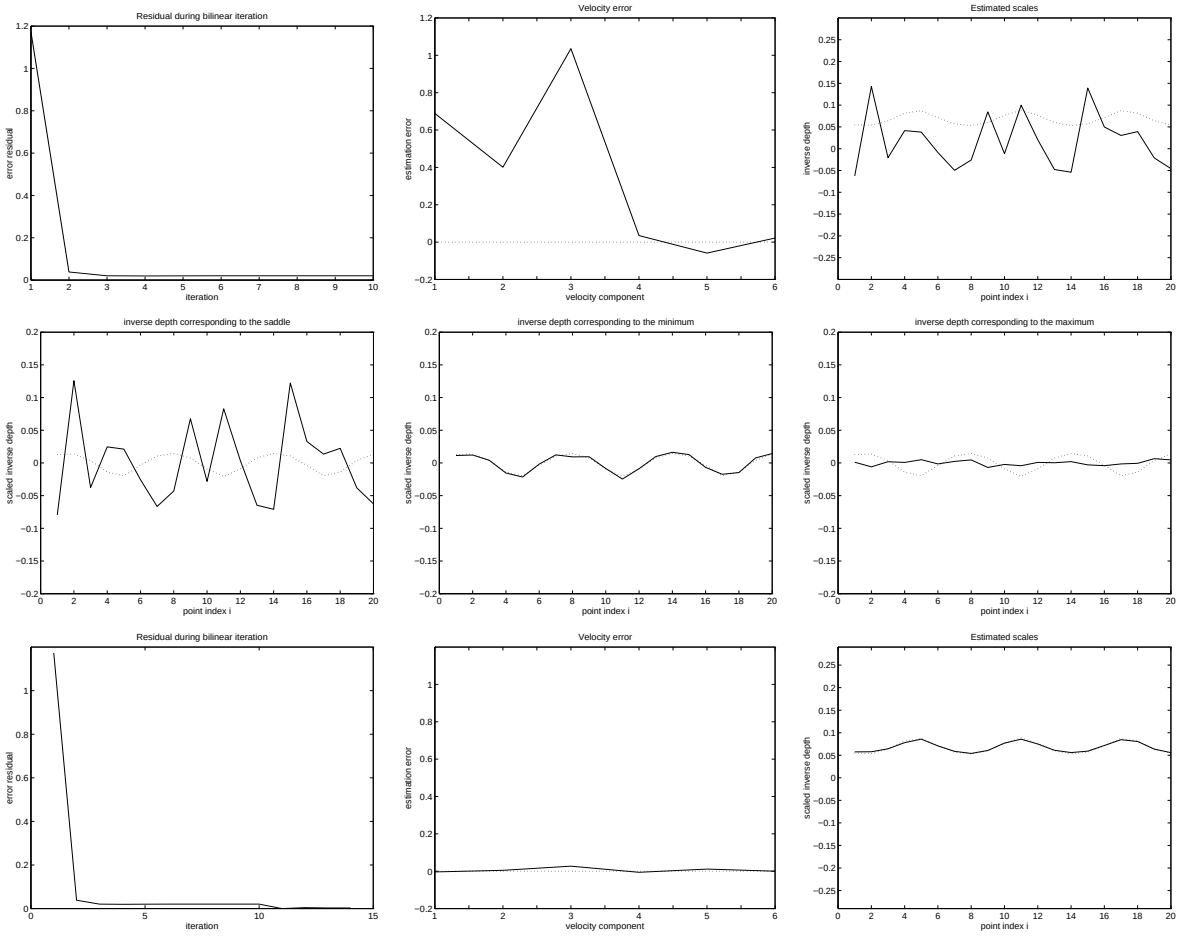

Figure 7: **Convergence to a saddle** *(top) the residual stabilizes (left), but the parameter error (center) and scales (right) are far away from their true values (in dotted lines). The normalized scales corresponding to the 3 eigenvectors of M, plotted in the middle row, show that one of them corresponds to the correct estimate. We may then switch to the correct solution and re-initialize the algorithm, that converges to the correct solution within 5 steps (bottom row).*

6.3 Predicted behavior based on the analysis

Based upon the analysis carried out in section 3, we know that local extrema of the original function of SFM r in (16) are in correspondence with the local extrema of the reduced function $r_2(\mathbf{a})$. Since $\|\mathbf{a}\| = 1$, we can represent $\mathbf{a}$ in spherical coordinates and plot the cost function r_2, as we do in figure 9. The motion that generated this plot was a fixating motion similar to the one of the box experiment (figure 5). In addition to the global minimum, corresponding to the coordinates $(0, \pi/2)$ (azimuth, elevation), we expect a maximum and a saddle in the orthogonal direction. These are showed in figure 9 (top-left). The saddle coincides with the singularity of the spherical coordinates: in fact, the two lines $(-\pi/2, \alpha)$, and $(\pi/2, \alpha)$ correspond to a point on the sphere. The rubbery interpretation corresponds to the local minimum diametrically opposed to the true motion (figure 9 top-right). the local extrema corresponding to the bas-relief ambiguity are reported in figure 9 (bottom-left), while in (bottom-right) we show the location of the extrema corresponding to the "rubbery" bas-relief ambiguity. We expect that our simulations will show convergence to some or all of these local extrema. In figure 10 we show the mesh-plot of the residual cost function in absence of noise (left) and with 100% noise (right). Notice that noise shifts the global minimum and makes the true solution a local

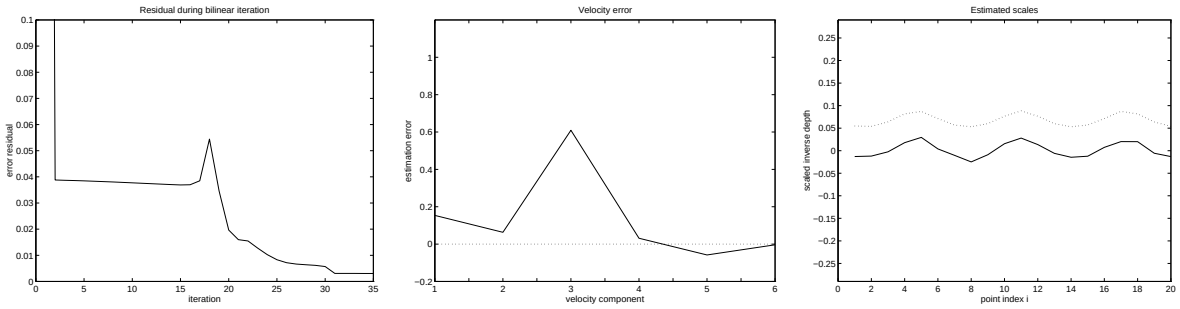

Figure 8: **Convergence to the bas-relief ambiguity** *The residual stabilizes (left), the parameters (center) are far away from the true values, but the estimated scales (right) are an averaged version of the true ones.*

minimum. This is a manifestation of the bas-relief ambiguity.

6.4 Experimental trials

We have considered $p = 20$ points on a volume of side $1m$ centered $2m$ from the center of projection. The points have images on the unit sphere. First we have considered forward translation at $0.2m/frame$, and noise levels of 0.1, 1 and 10 pixels, corresponding to 4%, 40% and 400% of the measurements respectively. For each noise level we have performed 200 trials. In figure 11 we show the plot of the residual of the cost function superimposed to the point where the Bilinear Projection iteration converged (a black asterisk). On the right, the same points have been checked against local minima, and the global minimum has been chosen correctly in all cases. Note that the cost function is periodic, so that only half of the figure is relevant. It can be seen that this type of motion is really simple and even large noise levels are easily tolerated by the algorithm. The situation is very different for a fixating motion. We have considered the same situation just described, but where the cloud of dots rotates of $1^o/frame$ about an axis passing through the centroid. In this case we have considered noise of 0.05, 0.5 and 5 pixels, corresponding to 2%, 20% and 200% of the measurements. Already at 2% noise we notice that the algorithm converges to local minima corresponding to both the saddle, the bas-relief ambiguity, and the rubbery motion perception. If we check for local minima, however, we can get to the correct estimate in all 200 trials (right). The same holds for 20% noise. For 200% noise, however, the rubbery motion perception becomes stable, so that about 40% of the trials return the rubbery motion solution even after correction.

6.5 Comparison with other algorithms

In this section we compare the optimal algorithm proposed in section 3 with other approaches based upon epipolar geometry [4], and upon linear subspace methods [7]. We use as a representative of the first class the algorithm described in [22], and as a representative of linear subspace methods the one in [18]. We consider $p = 20$ points distributed uniformly in a cube of side $1m$ centered at $2m$, rotating at $1^o/frame$, with measurements on the unit sphere corrupted by 50% noise. In figure 13 we show the result of 100 trials. The upshot is, not surprisingly, that the optimal algorithm works better. The interesting effect displayed by the algorithms based upon the epipolar constraint is that, despite the estimation error being large, the residual of the optimization is small. In fact, it is smaller than the noise level, which is a clear symptom that the epipolar constraint does not minimize the reprojection error. The outcome of the experiments for linear subspace methods show that the estimates are biased, as reported in [18].

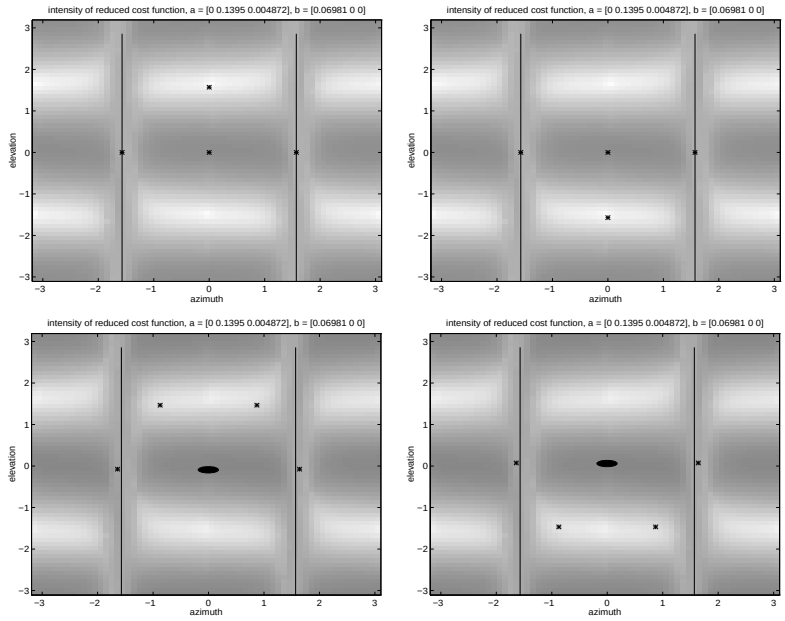

Figure 9: **Predicted extrema for fixating motion** *Local extrema are plotted as black asterisks superimposed to the residual of the cost function r_2 (left). We also show the local extrema corresponding to the rubbery interpretation (top-right), the bas-relief ambiguity (bottom-left), and the rubbery bas-relief ambiguity (bottom-right). Since there is a π-periodicity, only half of the plots is significant.*

7 Conclusions

The assumption of "small noise" is often illegitimate in conditions normally encountered in real-world experiments with SFM. Therefore, SFM needs to be addressed from the perspective of noise. We have proposed a bilinear projection iteration that provably converges to the optimal solution of SFM, and given analytical conditions for when this does not happen.

Upon concluding, we would like to anticipate some of the criticisms that a quick reading of the paper may have triggered. We invite the reader to seek confirmation of these statements in the paper.

- The constraint we use is **not** the equivalent of the epipolar constraint masked under spherical coordinates.
- The fact that we use a spherical projection model does **not** imply that we need a wide field of view.
- We do **not** assume small (or large) motions. We assume that velocity is measured, but nowhere we make the assumption that such a velocity needs to be small. Naturally, if discrete displacements are used to compute velocities using first-order differences, such an approximation is valid only for small *displacements*.
- The fact that in the spherical projection model the measurements live in tangent planes is associated with the geometry of the imaging model, and is **not** an approximation.

While it is not our interest to test the algorithm on the few and simple sequences publically available, we are very interested in an experimental platform that allows serious and systematic testing of different algorithms under repeatable conditions. We are working towards this end.

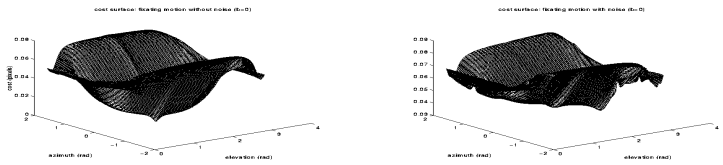

Figure 10: **Residual error corresponding to bas-relief ambiguity** *In the absence of noise (left), there is a global minimum corresponding to the true solution. In presence of* 100% *noise (right), the true solution becomes a local minimum, and the global minimum corresponds to the bas-relief ambiguity.*

Acknowledgements

We wish to thank John Oliensis for his suggestions on a first version of the manuscript, and Carlo Tomasi for his comments on the role of noise.

References

[1] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *Int. J. of Computer Vision*, 12(1):43–78, 1994.

[2] J. Bergen, R. Kumar, P. Anandan, and M. Irani. Representation of scenes from collections of images. In *Proc. of the IEEE Workshop on Visual Scene Representation*, Boston, June 1995.

[3] R. W. Brockett. Least Squares Matching Problems. *Linear Algebra and Its Applications*, 122-124:761-777, 1989.

[4] O. D. Faugeras. *Three Dimensional Vision, a geometric viewpoint.* MIT Press, 1993.

[5] O. D. Faugeras and Q. T. Luong. in preparation, 1997.

[6] G. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least-squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10 (2):413–432, 1973.

[7] A. Jepson and D. Heeger. Linear subspace methods for recovering rigid motion. *Spatial Vision in Humans and Robots, Cambridge University Press*, 1992.

[8] J. J. Koenderink and A. J. Van Doorn. Affine structure from motion. *J. Optic. Soc. Am.*, 8(2):377–385, 1991.

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

[10] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. 7th Int. Joint Conf. on Art. Intell.*, 1981.

[11] J. Oliensis. Provably correct algorithms for multi-frame structure from motion. *Proc. of the IEEE Conf. on Compupter Vision and Pattern Recognition*, 1996.

[12] C. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *Proc. of the 3 ECCV, LNCS Vol 810, Springer Verlag*, 1994.

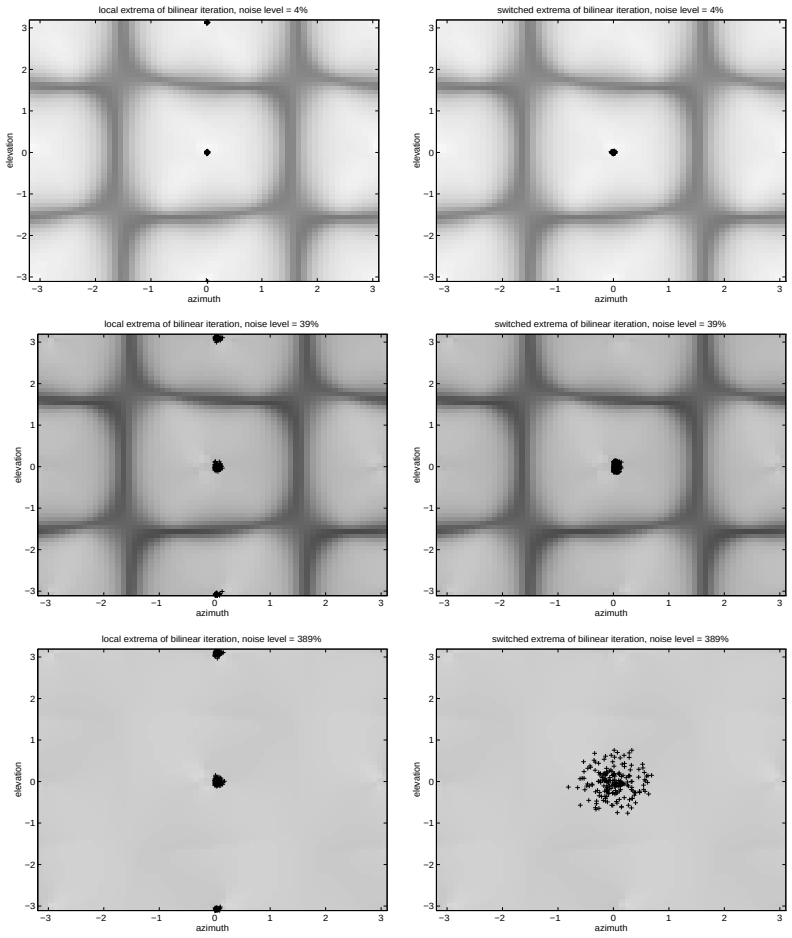

Figure 11: **Forward translation** *The residual of the cost function r_2 is superimposed to the fixed points of the Bilinear Projection iteration before (left) and after (right) correction for local extrema. Noise is* 4% *(top),* 40% *(center) and* 400% *(bottom) of the measurements. Convergence to the valley of the global optimum is achieved in all 200 trials.*

[13] P. Anandan R. Kumar and K. Hanna. Shape recovery from multiple views: a parallax based approach. *Proc. of the Image Understanding Workshop*, 1994.

[14] H. S. Sawhney. Simplifying motion and structure analysis using planar parallax and image warping. *Proc. of the Int. Conf. on Pattern Recognition*, Seattle, June 1994.

[15] H. S. Sawhney. Simplifying motion and structure analysis using planar parallax and image warping. *Proc. of the Int. Conf. on Pattern Recognition*, Seattle, June 1994.

[16] T. Soderstrom and P. Stoica *System Identification*. Prentice Hall, 1989.

[17] S. Soatto, R. Frezza, P. Perona. Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control*, 41(3):393-413, 1996.

[18] I. Thomas and E. Simoncelli. Linear Structure from Motion. *Technical Report IRCS 94-26*, University of Pennsylvania, 1994.

[19] T. Tian, C. Tomasi, and D. Heeger. Comparison of approaches to egomotion computation. In *proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1996.

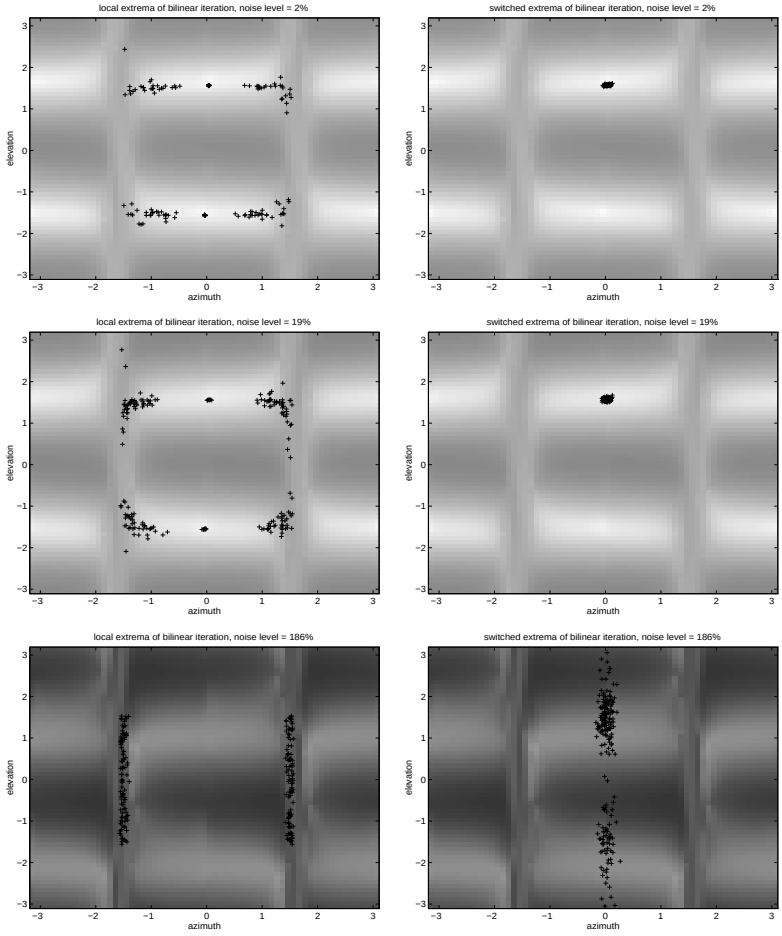

Figure 12: **Bas-relief ambiguity** *The Bilinear Projection algorithm converges to local minima corresponding to both the bas-relief ambiguity and the rubbery interpretation (left). For noises of* 2% *(top row) and* 20% *(center row), checking for local extrema is sufficient to achieve the correct solution in all 200 trials. For* 200% *noise (bottom), the rubbery interpretation is stable in* 30% *of the trials.*

[20] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. of Computer Vision*, 9(2):137–154, 1992.

[21] J. Weng, N. Ahuja, and T. Huang. Optimal motion and structure estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15:864–884, 1993.

[22] J. Weng, T. S. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(5):451–476, 1989.

A Rayleigh quotients

Let $\mathbf{v} \in \mathbb{R}^n$ be a non-zero vector and M, N two $n \times n$ symmetric matrices, with N positive-definite. The function

$$R(\mathbf{v}) \doteq \frac{\mathbf{v}^T M \mathbf{v}}{\mathbf{v}^T N \mathbf{v}} \qquad M = M^T; \quad N = N^T > 0; \quad \mathbf{v} \neq 0 \tag{29}$$

is called a *Rayleigh quotient*. $R(\mathbf{v})$ is homogeneous, $R(\alpha\mathbf{v}) = R(\mathbf{v})\ \forall \alpha \neq 0$, and therefore all the values taken by R are represented by the values on a compact (normalized) domain, for instance the sphere $\{\|\mathbf{v}\| = 1\}$, or the ellipsoid $\{\mathbf{v}^T N\mathbf{v} = 1\}$. We recall that the eigenvector of a (symmetric) matrix M *relative to* a (symmetric) matrix N is defined as the vector $\mathbf{v}$ for which there exists a (real) scalar λ, called the relative eigenvalue such that $M\mathbf{v} = \lambda N\mathbf{v}$.

We are now interested in finding critical points of Rayleigh quotients. First observe that $R(\mathbf{v}) \geq 0$, and that there exist maximum and minimum of R on its (compact) support. If M has a non-trivial null-space, then $\mathbf{v} \in \text{Null}(M)$ clearly minimizes $R(\mathbf{v})$. Therefore we will assume $M > 0$.

Claim A.1 *If $\mathbf{v}$ is an eigenvector of M relative to N with corresponding eigenvalue λ, then $\mathbf{v}$ is a critical point of the Rayleigh quotient $R(\mathbf{v})$, and λ is the corresponding critical value.*

Proof: $dR = 2\frac{M\mathbf{v} - R(\mathbf{v})N\mathbf{v}}{\mathbf{v}^T N\mathbf{v}} \iff M\mathbf{v} = R(\mathbf{v})N\mathbf{v}$. *The equality is satisfied by the relative eigenvalue/eigenvector pair $(R(\mathbf{v}), \mathbf{v})$.*

A more interesting situation arises when the matrix N is singular, in which case R may or may not be bounded. We study these two cases in the next two paragraphs. Before proceeding, however, note that R is not defined for $\mathbf{v} = 0$, and the limit $R(\mathbf{v}); \mathbf{v} \to 0$ is also not defined. However, for any *fixed* direction $\mathbf{v}$, the limit $R(\alpha\mathbf{v}); \alpha \to 0$ is defined. If σ_m (σ_M) are the smallest (largest) eigenvalue of M, and ρ_m (ρ_M) the smallest (largest) eigenvalue of N, then

$$\frac{\sigma_m}{\rho_M} \leq R(\mathbf{v}) \leq \frac{\sigma_M}{\rho_m}. \tag{30}$$

A typical plot of the value of $R(\mathbf{v})$ for $\mathbf{v} = \mathbf{v}(\theta, \phi)$ on a sphere is shown in figure 14; the asterisks indicate the critical points.

Semi-singular Rayleigh quotients

Now let $\text{Null}(M) \neq \emptyset$, but assume that all vectors that anihilate M, also anihilate N. If there are vectors in $\text{Null}(M)$ that do not belong to $\text{Null}(N)$, those trivially minimize the Rayleigh quotient. Therefore we will consider the case $\text{Null}(M) = \text{Null}(N)$. We define the *semi-singular Rayleigh quotient* as follows

$$R(\mathbf{v}) \doteq \frac{\mathbf{v}^T M\mathbf{v}}{\mathbf{v}^T N\mathbf{v}} \quad M = M^T; \quad N = N^T; \quad \mathbf{v} \notin \text{Null}(M) = \text{Null}(N). \tag{31}$$

Note that, in addition to being homogeneous, the semi-singular Rayleigh quotient is invariant along the null-space of N: $R(\mathbf{v} + \alpha\mathbf{n}) = R(\mathbf{v})\ \forall\ \mathbf{n} \in \text{Null}(N), \alpha \neq 0$. Therefore, we can further restrict the domain to be the orthogonal complement of the null-space of N:

$$\min_{\mathbf{v} \in \text{Null}(N)^\perp} R(\mathbf{v}). \tag{32}$$

All vectors in $\text{Null}(N)$ represent critical points, and their corresponding critical value is undefined, since the limit of $R(\mathbf{v})$ for $\mathbf{v} \to \text{Null}(N)$ is undefined. However, R is bounded in the whole domain. If we call σ_m the smallest *non-zero* eigenvalue of M, and ρ_m the smallest *non-zero* eigenvalue of N, then relation (30) holds unchanged.

Claim A.2 *Let $N = U\Sigma U^T$, where $U \in \mathcal{O}(n)$ is an $n \times n$ orthonormal matrix. Let $U_r = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r \end{bmatrix}$ be obtained from U by deleting the $n - r$ columns corresponding to the zero eigenvalues of N, and similarly with Σ_r. Then the critical points of the semi-singular Rayleigh quotient $R(\mathbf{v})$ for $\mathbf{v} \in \text{Null}(N)^\perp$ are obtained as the critical points of the Rayleigh quotient*

$$\frac{\mathbf{w}^T U_r^T M U_r \mathbf{w}}{\mathbf{w}^T \Sigma_r \mathbf{w}}: \quad \mathbf{w} \in \mathbb{R}^r; \quad \mathbf{w} \neq 0. \tag{33}$$

Proof: *Follows immediately from the fact that the columns of U_r span the subspace* $\text{Null}(N)^\perp$.

Therefore, semi-singular Rayleigh quotients can be treated as regular (non-singular) Rayleigh quotients on a subspace. A typical plot of the value of a semi-singular Rayleigh quotient for $n = 3$ and $r = 2$ is plotted in figure 15. The asterisks are the extrema of R, the 'o' represents the null-space of N, and the curve represents its orthogonal complement. Note that two of the extrema (a maximum and a minimum) are on the admissible subspace, while the third critical point is aligned with the null-space of N, and its value is undefined. We are now left with considering the case when $\text{Null}(M) \supset \text{Null}(N)$.

Singular Rayleigh quotients

In this section we restrict ourselves to consider N-matrices of co-rank 1 (whose null-space have dimension 1), and we require that $M > 0$. We define the *singular Rayleigh quotient* (of degree 1) as

$$R(\mathbf{v}) \doteq \frac{\mathbf{v}^T M \mathbf{v}}{\mathbf{v}^T N \mathbf{v}} \qquad M = M^T > 0; \quad N = N^T; \quad (\text{Null}(N) =< \mathbf{n} >) \tag{34}$$

Naturally, R is no longer bounded for $\mathbf{v} \to \text{Null}(N)$. Furthermore, R is no longer invariant along the null-space of N. In fact, if we decompose $\mathbf{v}$ into a component along $\text{Null}(N)^\perp$ and a component along $\text{Null}(N)$, $\alpha\mathbf{w} \overset{\perp}{+} \beta\mathbf{n}$, then we have

$$R(\alpha\mathbf{w} \overset{\perp}{+} \beta\mathbf{n}) = R(\mathbf{w}) + \frac{2\alpha\beta\mathbf{w}^T M\mathbf{n} + \beta^2\mathbf{n}^T M\mathbf{n}}{\alpha^2\mathbf{w}^T N\mathbf{w}}. \tag{35}$$

As a consequence, we can no longer restrict our attention to $\text{Null}(N)^\perp$, and we have to consider the full-fledged problem

$$\min_{\mathbf{v} \notin \text{Null}(N)} R(\mathbf{v}). \tag{36}$$

However, it is easy to see that one can still reduce the problem of finding extrema of R to analyzing a semi-singular Rayleigh quotient. To this end, define the matrix M_s as

$$M_s \doteq M - \frac{M\mathbf{n}\mathbf{n}^T M}{\mathbf{n}^T M\mathbf{n}}. \tag{37}$$

We have then the following

Claim A.3 *The critical points of R correspond to the critical points of the semi-singular Rayleigh quotient $\frac{\mathbf{v}^T M_s \mathbf{v}}{\mathbf{v}^T N\mathbf{v}}$ for $\mathbf{v} \notin \text{Null}(N)$.*

Proof A.1 *First note that $\text{Null}(M_s) = \text{Null}(N) =< \mathbf{n} >$. Let $\mathbf{v} = \mathbf{w} \overset{\perp}{+} \rho\mathbf{n}$, and $N = U\Sigma U^T$ as in the previous paragraph. $dR = \left[2\frac{M\mathbf{w}+\rho M\mathbf{n}-RN\mathbf{w}}{\mathbf{w}^T N\mathbf{w}}, 2\frac{\mathbf{w}^T M\mathbf{n}+\rho\mathbf{n}^T M\mathbf{n}}{\mathbf{w}^T N\mathbf{w}}\right]$, so at a critical point we must have*

$$\rho = -\frac{\mathbf{w}^T M\mathbf{n}}{\mathbf{n}^T M\mathbf{n}} \tag{38}$$

and $\mathbf{w}$ must satisfy

$$M\mathbf{w} - \frac{M\mathbf{n}\mathbf{n}^T M}{\mathbf{n}^T M\mathbf{n}}\mathbf{w} = \lambda N\mathbf{w} \tag{39}$$

which corresponds to $\mathbf{w}$ being an eigenvector of M_s relative to N with corresponding eigenvalue R.

A typical plot of the value of R on a sphere is reported in figure 16. Notice that the extrema, represented by asterisks, do not lie on the orthogonal complement to the null-space of N, represented by the solid curve.

Remark A.1 *Note that the extrema to singular Rayleigh quotients cannot be found simply by multiplying M by the pseudo-inverse of N, for that would give extrema on the orthogonal complement of the range of N.*

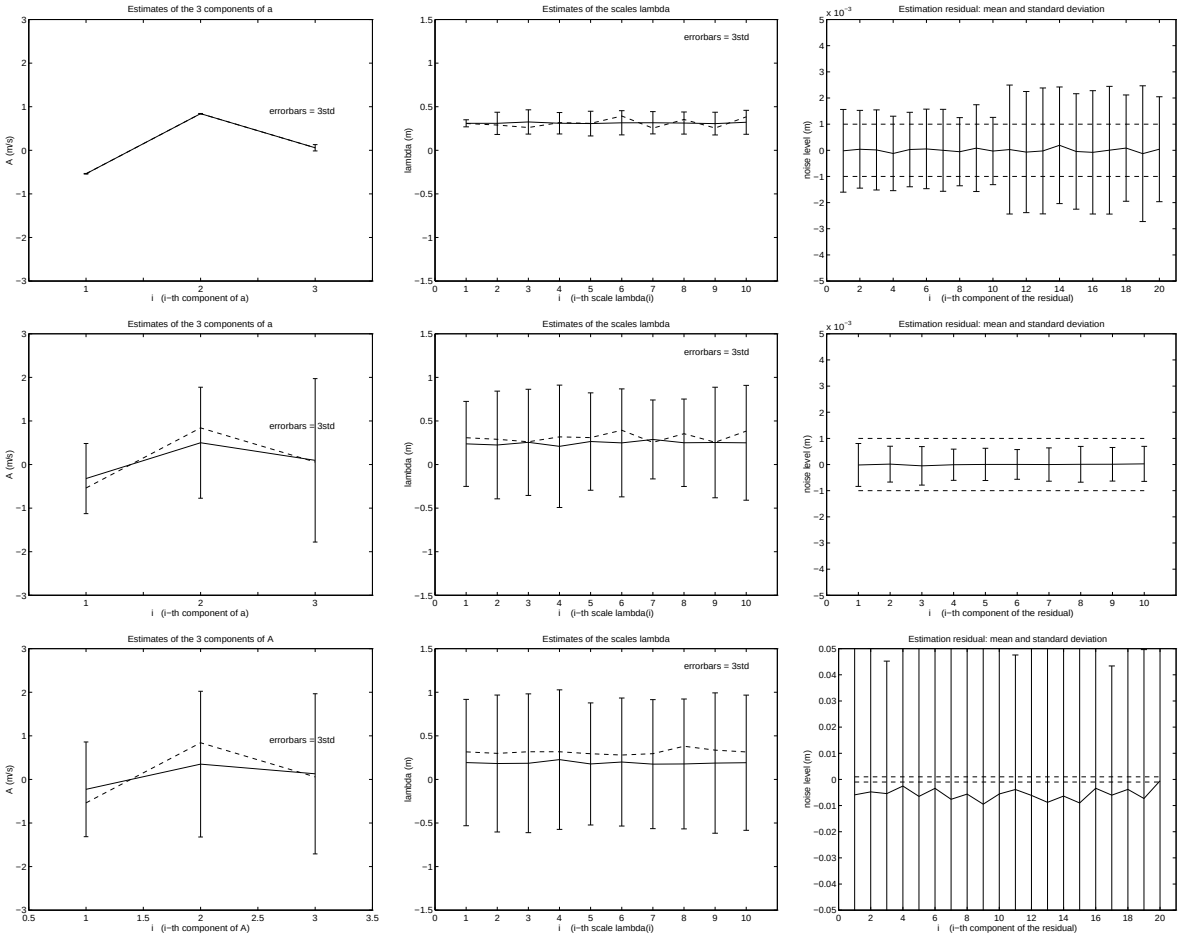

Figure 13: **Comparison with epipolar geometry and linear subspace methods** *(Top row) the estimates of translation (left) and the scales (center) are plotted with errorbars for 100 trials of the experiment. Noise is 50% of the measurements. The residual (right) is small, but it can be no smaller than the noise level, plotted in dashed lines. (Center row) algorithms based on the epipolar constraint perform considerably worse (left and center), despite the fact that the residual is smaller (right). The fact that the residual is smaller than the noise level is indeed a consequence of the fact that the epipolar constraint does not minimize the reprojection error. Linear subspace methods (bottom row) exhibit a bias both in the estimates (left and center) and in the residual (right).*

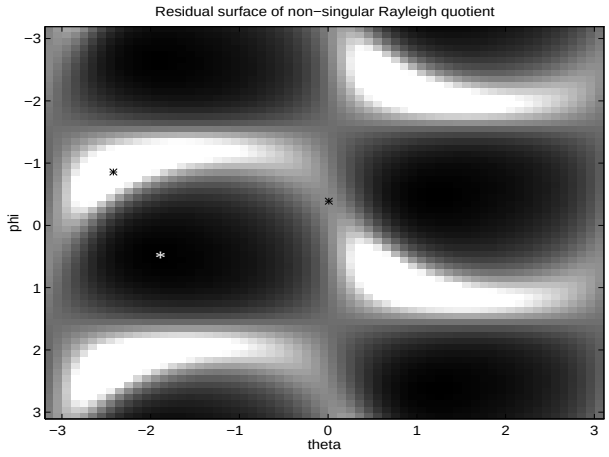

Figure 14: *Typical plot of the value of a Rayleigh quotient on a sphere, represented in local coordinates by an azimuth (theta) and an elevation angle (phi). Extrema are indicated by asterisks. The plot is periodic, so only half of it is relevant.*

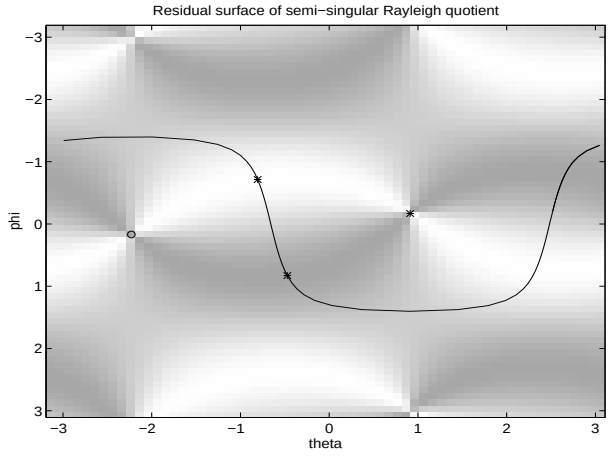

Figure 15: *Typical plot of a semi-singular Rayleigh quotient. The solid curve is the locus of points on the space orthogonal to* Null(N). *Extrema are indicated by asterisks.*

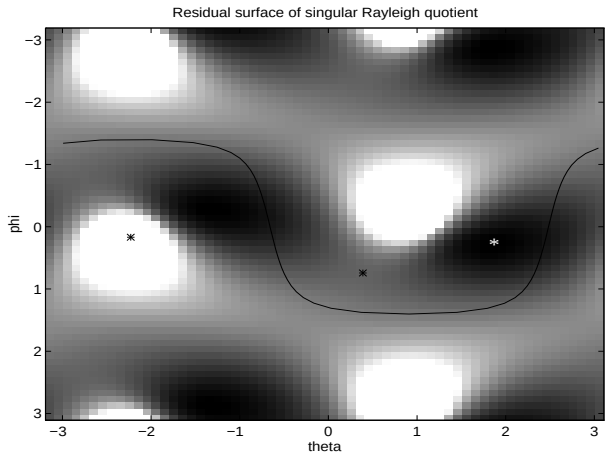

Figure 16: *Typical plot of a singular Rayleigh quotient on a phere. The solid curve is the locus of point on the orthogonal complement of* $\mathrm{Null}(N)$. *Extrema are indicated by asterisks.*