Moving Forward in Structure From Motion

Andrea Vedaldi Gregorio Guidi Stefano Soatto
Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095, USA
{vedaldi,guidi,soatto}@cs.ucla.edu

Abstract

It is well-known that forward motion induces a large number of local minima in the instantaneous least-squares reprojection error. This is caused in part by singularities in the error landscape around the forward direction, and presents a challenge in using existing algorithms for structure-from-motion in autonomous navigation applications. In this paper we prove that imposing a bound on the reconstructed depth of the scene makes the least-squares reprojection error continuous. This has implications for autonomous navigation, as it suggests simple modifications for existing algorithms to minimize the effects of local minima in forward translation.

1. Introduction

Structure-from-motion (SFM), the problem of reconstructing the three-dimensional structure of a scene and the motion of a camera from a collection of images, can be counted among the success stories of Computer Vision. Over the course of the past twenty years, the geometry of the problem has been elucidated, and an extensive corpus of analysis and algorithms has been collated in several textbooks, e.g. [4]. Some have even made their way into the real-world as commercial products. However, despite all the success, *forward motion still presents a challenge to existing SFM algorithms.* This problem rarely occurs in match-moving and virtual object insertion, for which most commercial algorithms are designed, but has become painfully patent with the recent push in autonomous driving.

The difficulties with forward motion are due in part to the limited lifetime of point-feature tracks: The most informative features are in the peripheral vision and they quickly move out of the visual field. This can be addressed by enlarging the field of view, all the way to omni-directional cameras, which explains their popularity in robotic navigation. A less-easily fixed difficulty with forward motion is the presence of a *large number of local minima in the least-squares landscape of the reprojection error,* which many existing algorithms try to minimize. These local minima have been studied in some detail by Oliensis [7] and Chiuso et al. [2], using tools first introduced by Heeger and Jepson [5] and Golub and Pereyra [3]. All have shown the presence of local minima corresponding to well-known ambiguities (Necker-reversal, plane-translation and bas-relief) and, more importantly for our problem, they have shown that *the least-squares reprojection error has singularities* corresponding to translation in the direction of point features, which introduce a fine-scale topology with many local minima around the true direction of translation (see [7], figures 1 and 8, or [2] figure 7). In order to overcome this problem, semi-global approaches based on convex optimization and fractional programming have been recently proposed [1, 6], but the resulting algorithms are too slow to run in real time for navigation applications, and the actual hypotheses to guarantee global convergence are not strictly satisfied for central projection and forward motion. Thus we are left with finding simpler ways to handle the strange geometry of the reprojection error surface. *Is such a geometry a product of the mathematics we are using,* for instance the choice of L^2 to measure reprojection error, *or is it intrinsic to the problem?* Can a different choice of norm eliminate the singularities? Is there some reasonable assumption we can make on the scene or on the inference process that will make the singularities vanish?

In this paper we show that the singularities are a byproduct of the mathematics, and can be easily eliminated. Specifically, we **prove that imposing a bound on the depth of the scene makes the least-squares reprojection error continuous**. We also show how such an assumption can be easily embedded in an iterative algorithm. It does not, however, show that local minima disappear altogether. In fact, only continuity can be guaranteed analytically, not convexity, and an empirical analysis shows that some local minima still exist.

1-4244-1180-7/07/$25.00 ©2007 IEEE

2. Background

In this section we will introduce the notation and rephrase some of the results of [2] and [7] for completeness.

Let $X_1, \dots, X_N \in \mathbb{R}^3$ be points on a rigid body (the scene). Let $x_i = X_i/|X_i|$ be the projection of the i-th point taken from a given vantage point on a spherical imaging surface. Let v and ω be the instantaneous linear and angular velocities of the scene (rigid body) relative to the camera. The motion causes the feature projections x_i to move on the imaging surface with velocities z_i. Our goal is to recover both the structure and motion from the measurements $x_1, \dots, x_N$ and $z_1, \dots, z_N$. To this end we note that the i-th measured velocity is given by

$$z_i = \omega \times x_i + \lambda_i \left(v - \langle v, x_i \rangle x_i\right) + n_i \tag{1}$$

where $\lambda_i = 1/|X_i|$ is the inverse of the distance of the i-th feature from the camera center and n_i is measurement noise. As z_i, n_i and $\omega \times x_i$ belong to the tangent space $T_{x_i}(\mathbb{S}^2)$ of the spherical imaging surface, the norm of the residual does not change if we pre-multiply by $-\widehat{x}_i = -x_i\times$, obtaining

$$-\widehat{x}_i z_i = \widehat{x}_i^2 \omega - \lambda_i \widehat{x}_i v - \widehat{x}_i n_i. \tag{2}$$

As the norm and statistics of $\pm\widehat{x}_i n_i$ and n_i are the same, we set $y_i = -\widehat{x}_i z_i$ and define the fitting cost

$$E(v, \omega, \lambda_1, ..., \lambda_N) = \sum_{i=1}^{N} \|n_i\|^2 = \\ = \sum_{i=1}^{N} \|y_i - \widehat{x}_i^2 \omega + \lambda_i \widehat{x}_i v\|^2. \tag{3}$$

The goal of SFM, for the purpose of our analysis, is to find the minimizers of the cost function E.

2.1. Reduced diagrams

Heeger and Jepson introduced subspace projection into their SFM algorithm in 1990 [5]. They showed that SFM can be reduced to a two-dimensional optimization problem. Their approach was later used by Chiuso et al. [2] to study local minima in SFM after noticing that the local extrema in the reduced optimization are related to local minima in the original one. The two-dimensional landscape can be visualized in CBS diagram. Here we rederive some of the properties of CBS diagrams since these results will be used later in proving continuity of the reduced cost function.

The squared residual $E(v, \omega, \lambda_1, \dots, \lambda_N)$ is a function of both the inverse dephts of the observed points (features) $\lambda_1, \dots, \lambda_N$ (structure) and the camera angular ω and linear v velocities (motion). The idea of [5] is to solve for all variables but linear velocity v. This gives a *reduced function* $E^*(v) = E(v, \omega^*(v), \lambda_1^*(v), \dots, \lambda_N^*(v))$ of v only. The CBS diagram is then the function

$$E^*(v) = \min_{\omega \in \mathbb{R}^3, \lambda_1, \dots, \lambda_N \in \mathbb{R}} E(v, \omega, \lambda_1, \dots, \lambda_N) \tag{4}$$

An algebraically simplified variant, called *weighted* CBS diagram, is obtained by optimizing weighted residuals $w_i \|n_i\|^2$.

2.2. Properties

The function (4) enjoys several useful properties. First we note that fixing v turns the minimization (4) into a simple least-squares problem. To make this more explicit, define $y \doteq \begin{bmatrix} y_1^T & y_2^T & \dots & y_N^T \end{bmatrix}^T$,

$$\Phi(v) \doteq \begin{bmatrix} \widehat{x}_1^2 & -\widehat{x}_1 v & 0 & \dots & 0 \\ \widehat{x}_2^2 & 0 & -\widehat{x}_2 v & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \widehat{x}_N^2 & 0 & 0 & \dots & -\widehat{x}_N v \end{bmatrix} \tag{5}$$

$a \doteq \begin{bmatrix} \omega^T & \lambda_1 & \lambda_2 & \dots & \lambda_N \end{bmatrix}^T$, so that

$$E^*(v) = \min_{a \in \mathbb{R}^{3+N}} E(v, a) = \min_{a \in \mathbb{R}^{3+N}} \|y - \Phi(v) a\|^2. \tag{6}$$

Note that the optimal value $a(v)$ of the linear variable a – angular velocity and structure – for a given linear velocity v is given by

$$a(v) = \Phi(v)^\dagger y \tag{7}$$

where $\Phi(v)^\dagger$ is the pseudo-inverse of the matrix $\Phi(v)$.

2.2.1 Singular configurations

We say that a vector u is *proportional* to a vector v, $u \propto v$ in symbols, if there exists a number $\lambda \in \mathbb{R}$ such that $u = \lambda v$. Note that $0 \propto v$ for each v.

Proposition 1. *Consider features $X_1, \dots, X_N \in \mathbb{R}^3$, $N \geq 1$ and a vector $v \in \mathbb{S}^2$, such that $X_i \not\propto v$ for $i = 1, \dots, N$. The kernel of the matrix $\Phi(v)$ is not empty if, and only if, we can find a vector $\omega \in \mathbb{S}^2$ such that*

$$X_i^\top Q(v, \omega) X_i = 0, \tag{8}$$

$$Q(v, \omega) = \frac{1}{2}(v\omega^\top + \omega v^\top - 2\langle v, \omega \rangle I). \tag{9}$$

for all $i = 1, \dots, N$.

Proof. In order for the kernel of $\Phi(v)$ not to be empty, the following condition must hold

$$\exists\, (\omega, \lambda_1, \dots, \lambda_n) \neq 0 : \\ \widehat{X}_i^2 \omega - \widehat{X}_i v \lambda_i = 0, \quad i = 1, \dots, N. \tag{10}$$

Since we assumed that $\widehat{X}_i v \neq 0$, if we choose $\omega = 0$ then equation (10) is satisfied if and only if all λ_i vanish, thus ω has to be different from zero. Hence, from (10) we get the following condition by taking the scalar product with v.

$$\exists \omega \in \mathbb{S}^2 : \quad \langle \widehat{X}_i^2 \omega, v \rangle = 0, \quad i = 1, \ldots, N. \tag{11}$$

The reverse implication is also true. Let us assume that (11) holds, we have $\langle \widehat{X}_i^2 \omega, v \rangle = 0$, $\langle \widehat{X}_i^2 \omega, X_i \rangle = 0$ and we know that X_i and v are linearly independent, so $\widehat{X}_i^2 \omega$ is directed along the vector $\widehat{X}_i v$, which means that there exists a λ_i (possibly null) such that (10) holds.

We rewrite (11) by using the identity $(a \times b) \times c = \langle a, c \rangle b - \langle b, c \rangle a$ to get $\widehat{X}_i^2 \omega = \langle X_i, \omega \rangle X_i - \langle X_i, X_i \rangle \omega$ and the equivalent conditions

$$\begin{aligned} v^\top (X_i X_i^\top \omega - (X_i^\top X_i)\omega) = \\ X_i^\top (v\omega^\top - v^\top \omega I) X_i = X_i^\top M X_i = 0, \end{aligned} \tag{12}$$

$$M = v\omega^\top - v^\top \omega I. \tag{13}$$

This is the equation of a quadric and is more easily studied by symmetrizing the matrix M, obtaining the equivalent constraints

$$X_i^\top Q X_i = 0, \quad i = 1, \ldots, N, \quad Q = \frac{1}{2}(M + M^\top). \tag{14}$$

□

If $v \neq w$ it is easy to verify that the eigenvectors and eigenvalues of Q are given by

$$z_1 = v + \omega, \qquad \lambda_1 = \tfrac{1}{2}(1 - \langle v, \omega \rangle), \tag{15}$$

$$z_2 = v - \omega, \qquad \lambda_2 = -\tfrac{1}{2}(1 + \langle v, \omega \rangle), \tag{16}$$

$$z_3 = v \times \omega, \qquad \lambda_3 = -\langle v, \omega \rangle. \tag{17}$$

If $v = \pm\omega$, then $Q = \pm\widehat{v}^2$ and the eignevectors-eigenvalues are given by

$$z_1 = v, \qquad \lambda_1 = 0 \tag{18}$$

$$z_{2,3} \in v^\perp, \qquad \lambda_{2,3} = \mp 1. \tag{19}$$

Equation (8) then defines a double elliptical cone which contains v and ω, directed either along $v+\omega$ with major axis along $v-\omega$ (if $\langle v, \omega \rangle > 0$), or directed along $v-\omega$ with major axis along $v + \omega$ (if $\langle v, \omega \rangle < 0$). In case $\langle v, \omega \rangle = 0$ the cone degenerates to the union of the two planes $\langle X, v \rangle = 0$ and $\langle X, \omega \rangle = 0$.

We observe that:

- $\Phi(v)$ is $3N \times 3 + N$ so for $N = 1$ the kernel is not empty. Indeed $(\omega, \lambda_1) = (X_1, 0)$ belongs to the kernel.
- For $N = 2$ the kernel is also not empty. The conditions of Proposition 1 are in fact satisfied for $\omega \propto (v \times x_1)/\langle v, x_2 \rangle - (v \times x_2)/\langle v, x_1 \rangle + x_1 \times x_2$.
- If $N \geq 3$, then $\Phi(v)$ has full rank $N + 3$ as long the quadratic equation (8) is not satisfied for some $\omega \in \mathbb{S}^2$.

The previous observations motivate the following definition:

Definition 1. The points $X_1, \ldots, X_N$ ($N \geq 3$), are said to be in *general position for the linear velocity* $v \neq 0$ if, and only if,

1. $X_i \not\propto v, \quad i = 1, \ldots, N$,
2. there exists no angular velocity ω such that $X_i^\top Q(v, \omega) X_i = 0$ for all $i = 1, \ldots, N$.

We also say that the points are in *general position* if they are in general position relatively to all velocitites v for which point 1 above is satisfied.

We can now rephrase Proposition 1 in the following remark.

Remark 1. If points $X_1, \ldots, X_N$ are in general position for $v \neq 0$, then the matrix $\Phi(v)$ has full rank.

2.2.2 Correspondence of minimizers

Proposition 2. *The reduced cost function $E^*(v)$ enjoys the following properties. Let $\Omega = \mathbb{S}^2 - \{x_1, \ldots, x_N\}$ the open subset of the imaging sphere that does not include the feature projections and let $\{x_1, \ldots, x_N\}$ be in general position. Then:*

- *If v is a critical point (or global minimizer) of $E^*(v)$ for $v \in \Omega$, then $(v, a(v))$ is a critical point (or global minimizer) of $E(v, a)$ for $(v, a) \in \Omega \times \mathbb{R}^{N+3}$.*
- *If (v, a) is a global minimizer of $E(v, a)$, then v is a global minimizer of $E^*(v)$.*

Proof. This kind of problems is discussed in [3]. The rank of $\Phi(v)$ is constant for all $v \in \Omega$. The result is then an immediate application of Thm. 2.1 therein. □

Prop. 2 says that all the critical points of the reduced cost function $E^*(v)$ (except at most singularities corresponding to the feature projections) correspond to critical points of the full cost function $E(v, a)$. It also says that global minimizers of the full and reduced cost functions correspond.

What the proposition does *not* say is that critical points of $E(v, a)$ are reflected by the reduced cost function $E^*(v)$. This is however easily seen. Since the points are in general position, $\Phi(v)$ never drops rank and $a(v)$ and $E^*(v)$ are differentiable. Let (v, a) be a critical point of $E(v, a)$. Then $a = a(v)$ and $\nabla E(v, a(v)) = 0$, so that

$$\frac{\partial E^*}{\partial v}(v) = \frac{\partial E}{\partial v}(v, a(v)) + \frac{\partial E}{\partial a}(v, a(v)) \frac{\partial a}{\partial v}(v) = 0. \tag{20}$$

So, except for singularities or pathological cases, the local minima of the reduced and full costs correspond.

2.3. Computation of the reduced functional

The reduced cost function (4) can be computed efficiently in two steps: Optimization of the depths given the motion, optimization of the rotation given the translation.

- **Optimizing the depths given the motion.** We fix ω and v in (3), obtaining for each λ_i its optimal value

$$\lambda_i = \frac{v^\top \widehat{x}_i^\top}{v^\top \widehat{x}_i^\top \widehat{x}_i v} \left(\widehat{x}_i^2 \omega - y_i \right), \tag{21}$$

which give the cost function

$$E(v,\omega) = \sum_{i=1}^{N} \|n_i\|^2 = \\ = \sum_{i=1}^{N} \left\| \left(I - \frac{\widehat{x}_i v v^\top \widehat{x}_i^\top}{v^\top \widehat{x}_i^\top \widehat{x}_i v} \right) \left(y_i - \widehat{x}_i^2 \omega \right) \right\|^2. \tag{22}$$

The depth λ_i of a feature x_i parallel to the motion direction v does not affect the cost function, is undetermined and the corresponding operator $\widehat{x}_i v v^\top \widehat{x}_i^\top / v^\top \widehat{x}_i^\top \widehat{x}_i v$ is null. Moreover, as v approaches x_i, the limit of such operator does not exist (the directional limit, however, is well defined). This property of the residuals n_i generates singularities in correspondence of features in the cost function.

- **Optimizing the rotation given the translation.** Assume $v \not\propto x_i$ for all features x_i. Recall that $I = uu^\top - \widehat{u}^2$. Let $u = \widehat{x}_i v / \|\widehat{x}_i v\|$, so that

$$I - \frac{\widehat{x}_i v v^\top \widehat{x}_i^\top}{v^\top \widehat{x}_i^\top \widehat{x}_i v} = -\widehat{u}^2 \tag{23}$$

$$\|n_i\|^2 = \| -\widehat{u}^2 (y_i - \widehat{x}_i^2 \omega)\|^2 = \\ = \|\widehat{u}(y_i - \widehat{x}_i^2 \omega)\|^2. \tag{24}$$

We now use the fact that $\widehat{u} = (v x_i^\top - x_i v^\top)/\|\widehat{x}_i v\|$, and that the residual $y_i - \widehat{x}_i^2 \omega$ is orthogonal to x_i to write

$$E(v,\omega) = \sum_{i=1}^{N} \|n_i\|^2 = \\ = \sum_{i=1}^{N} \frac{\|x_i v^\top (y_i - \widehat{x}_i^2 \omega)\|^2}{\|\widehat{x}_i v\|^2} = \\ = \sum_{i=1}^{N} \frac{(v^\top (y_i - \widehat{x}_i^2 \omega))^2}{\|\widehat{x}_i v\|^2}. \tag{25}$$

The latter expression is a standard least-squares estimation problem for ω. Note that estimating the velocity v is much harder, as the weights $\|\widehat{x}_i v\|$ depend on it. Note also that the residual $y_i - \widehat{x}_i^2 \omega$ is orthogonal to x_i, so that when the denominator $\|\widehat{x}_i v\|$ is small, because v approaches x_i, the numerator is small too. Solving for ω yields

$$\omega = \left(\sum_{i=1}^{N} \frac{\widehat{x}_i^2 v v^\top \widehat{x}_i^2}{\|\widehat{x}_i v\|^2} \right)^{-1} \sum_{i=1}^{N} \frac{\widehat{x}_i^2 v v^\top y_i}{\|\widehat{x}_i v\|^2}. \tag{26}$$

3. Bounded depth

In this section we extend the previous results and present the main contribution of this paper, which is a characterization of the reduced cost function when the reconstructed depths are bounded.

Singularities (discontinuities) at feature locations $v = x_1, \ldots, x_N$ of the reduced cost function $E^*(v)$ are possible because the feature dephts can be made arbitrarily small. This is reflected in equation (3): when a coefficient $\widehat{x}_i v \to 0$ as $v \to x_i$, the corresponding variable (inverse depth) λ_i approaches infinity to counter-balance. Thus we propose to work with a "regularized" reduced cost function $E^*_\alpha(v)$ defined as

$$E^*_\alpha(v) = \min_{\omega \in \mathbb{R}^3,\ |\lambda_i| \leq \alpha} E(v, \omega, \lambda_1, \ldots, \lambda_N). \tag{27}$$

The following lemmas are meant to establish the continuity of $E^*_\alpha(v)$ and the correspondence of global and local minimizers, by showing results in a more general setting (extending Proposion 2 with respect to correspondence of minimizers).

3.1. Continuity

Consider a function $E : V \times A \to \mathbb{R}$, where V is a topological space and A is a subset of $\mathbb{R}^k$, and let $E^*(v) = \min_{a \in A} E(v,a)$. In general, the continuity of E is not sufficient for E^* to be continuous, as it is only sufficient for the upper semicontinuity of E^*. However, if A is compact then E^* is indeed continuous, as shown in the following lemma.

Lemma 1. *Let E and E^* be defined as above. If E is continuous, then E^* is upper semicontinuous; moreover, if A is compact then E^* is also continuous.*

Proof. Consider the level sets

$$L^s_{E^*} = \{ v \in V : E^*(v) < s \}, \tag{28}$$

$$L^s_E = \{ (v,a) \in V \times A : E(v,a) < s \}. \tag{29}$$

Then

$$L^s_{E^*} = \pi_V \left(L^s_E \right), \tag{30}$$

where $\pi_V : V \times A \to V$ is the canonical projection. Since the level set L^s_E is open (E is continuous), its projection is open, too. The openess of $L^s_{E^*}$ for every s proves the upper semicontinuity of E^*.

Now if A is compact, we consider the level sets

$$W^s_{E^*} = \{v \in V : E^*(v) \leq s\}, \quad (31)$$
$$W^s_E = \{(v,a) \in V \times A : E(v,a) \leq s\}, \quad (32)$$

we also have

$$W^s_{E^*} = \pi_V (W^s_E), \quad (33)$$

since $E(v,\cdot)$ attains its minimum value in A. Since the level set W^s_E is closed (E is continuous) and A is compact, the projection is closed, too. This can be easily seen by taking a point $v \notin W^s_{E^*}$ and a covering of $\{v\} \times A$ with open subsets belonging to ${W^s_E}^c$, then extracting a finite covering and considering the intersection of their projections on V to obtain a neighborhood of v in ${W^s_{E^*}}^c$.

Thus, if A is compact E^* is both upper and lower semi-continuous, completing the proof. □

The result stated in the last lemma applies to the regularized cost function E^*_α defined in (27), since the bound on λ_i restricts the space of parameters to a compact space (there is a natural bound on the parameter ω, given the bound α on the paramters λ_i — see Lemma 5) . The use of the cost function (27) is in this sense justified.

3.2. Correspondence of minimizers

Now we see which results on correspondence of local and global minima we can obtain in this general setting.

Let a^* be a function defined from V to A such that

$$E^*(v) = \min_{a \in A} E(v,a) = E(v, a^*(v)). \quad (34)$$

Lemma 2. *$\bar{v}$ is a global minimizer for E^* if, and only if $(\bar{v}, a^*(\bar{v}))$ is a global minimizer for E.*

Proof. It follows trivially from the equation $\min_{v \in V} E^*(v) = \min_{v \in V, a \in A} E(v,a)$. □

Lemma 3. *If $\bar{v}$ is a local minimizer for E^*, then $(\bar{v}, a^*(\bar{v}))$ is a local minimizer for E.*

Proof. Let $U \ni \bar{v}$ be a neighborhood such that $\bar{v}$ is a minimizer for E^* in U. Then for $v \in U$ and $a \in A$ we have

$$E(v,a) \geq E(v, a^*(v)) = E^*(v) \geq \\ \geq E^*(\bar{v}) = E^*(\bar{v}, a^*(\bar{v})), \quad (35)$$

so $(\bar{v}, a^*(\bar{v}))$ is a minimizer for E in $U \times A$. □

The converse is not true in general, so we consider an additional constraint on the regularity of a^*.

Lemma 4. *If $(\bar{v}, a^*(\bar{v}))$ is a local minimizer for E and a^* is continuous in $\bar{v}$, then $\bar{v}$ is a local minimizer for E^*.*

Proof. Let $W \ni (\bar{v}, a^*(\bar{v}))$ be a neighborhood such that $(\bar{v}, a^*(\bar{v}))$ is a minimizer for E in W. Since the mapping $\phi : v \mapsto (v, a^*(v))$ is continuous in $\bar{v}$, then $\phi^{-1}(W)$ contains a neighborhood U of $\bar{v}$. For $v \in U$, we have

$$E^*(v) = E(v, a^*(v)) \geq E(\bar{v}, a^*(\bar{v})) = E^*(\bar{v}), \quad (36)$$

so $\bar{v}$ is a minimizer for E^* in U. □

Under the assumptions of Proposition 2, with $a(v) = \Phi(v)^\dagger y$, continuity of $a(v)$ holds, so all lemmas 2, 3, 4 apply and we obtain correspondence of minimizers for the regularized reduced cost function (27).

This sequence of lemmas constitutes the contribution of this paper. In the next section we will illustrate empirically the effect of bounded depth on CBS diagrams.

4. Experiments

In this section we illustrate the behavior of the CBS diagrams when bounding the underlying depths. Notice that this is not straightforward, because the CBS diagrams *do not depend on depth*, at least not directly, because it has been eliminated by subspace projection [5].

In figure 1 we show the CBS diagrams computed using various bounds on the scene depth. Each plot is divided into two parts, for each one the cost function is represented through a mapping of the domain $\mathbb{S}^2$ onto a square (by projecting onto a octahedron) and with color encoding. The left part shows just the function values and the feature points in green, while the right part shows the function with simulated 3D appearance. The four plots start from no boundary (top left) and go to an extremely high boundary value (bottom right), which actually shows that imposing a bound on depth that is too strict eventually produces deep changes in the structure of the cost function, as expected.

The structure of the cost function near singularities, and its modifications with bounded depths, are better appreciated in the closeups in figure 2, where the opposite of the function is shown (so that minima are represented as peaks). It is clear that the bound on depths has the effect of removing the singularities (as proved) and also that many of the local minima disappear, even though local peaks are still present.

We evaluated the improvement obtained by minimizing a regularized cost function when the algorithm used for minimization is a simple gradient descent, in order to provide a quantitative appraisal of the change in structure that the cost function undergoes when considering bounded depths. For these experiments, we generated 10000 scenes of 25 random points each, selected according to a Gaussian distribution with standard deviation 0.5, at a distance of 2.5 from the camera. The measurements were generated by an instantaneus motion of the camera along the forward direction

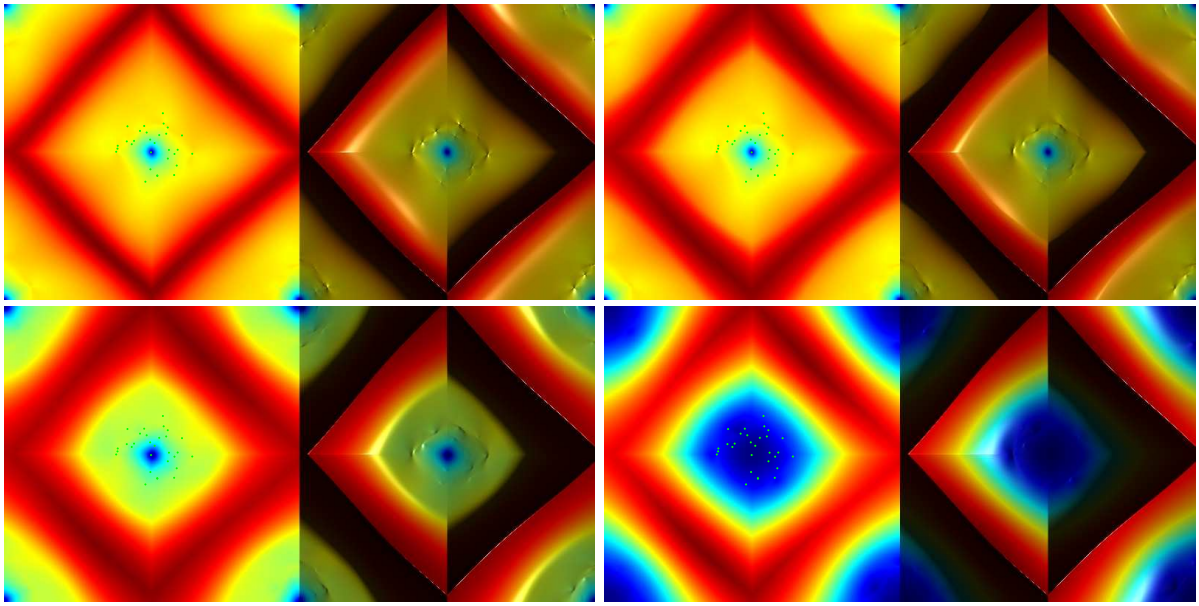

Figure 1. CBS diagrams with various depth constraints. Top-left: no constraint; top-right: $|X_i| > 0.5L$ (L is the minimum distance of features from camera); bottom-left: $|X_i| > 0.8L$; bottom-right: $|X_i| > L$.

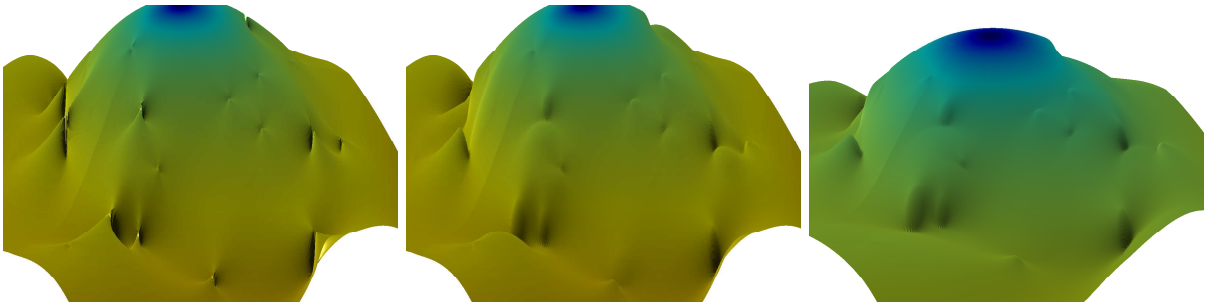

Figure 2. 3D representation of CBS diagrams with various depth constraints. Top-left: no constraint; top-right: $|X_i| > 0.5L$; bottom: $|X_i| > 0.8L$.

no bound	86.2%
$\|X_i\| > 0.01L$	87.1%
$\|X_i\| > 0.05L$	87.1%
$\|X_i\| > 0.1L$	87.2%
$\|X_i\| > 0.5L$	87.1%
$\|X_i\| > 0.9L$	88.6%
$\|X_i\| > 1.5L$	69.2%

Table 1. Percentage of successes for gradient descent algorithm with bounded depths. L is the minimum distance of the scene points from the camera (note that the unit of measure of distances is chosen so that the instantaneous velocity has unit norm).

with random rotational component, and the descent algorithm was initialized randomly. The step on the direction of greatest descent was selected using a backtracking strategy, choosing the first point where the cost function decreases.

As can be seen from Table 1, the fraction of gradient descent runs that correctly find the global minimum is consistently greater when the bound on the depths is enforced, even for loose bounds. Eventually, imposing a very strict boundary disrupts entirely the structure of the cost function, as shown by the last row of the table.

5. Discussion

We have proven that imposing a bound on the depth of the reconstructed scene makes the least-squares reprojection error continuous. This result shows that many of the local minima induced by singularities in the least-squares cost function that plague existing SFM algorithms when applied to autonomous navigation, where most of the motion is forward, do not really exist and can be simply avoided

by altering the main iteration in the algorithms to enforce bounded depth. [1]

Our theoretical results can be visualized on the two-dimensional CBS diagrams. Note that we are imposing bounds on depth, and the CBS diagram does not depend on depth, so the reader will have to actually read the proofs to appreciate how this comes about.

Because the local structure is affected by the location of feature points used in the computation of the reduced cost, one could conceive sampling strategies where only subsets of the available measures are chosen. This way spurious minima due to the configuration of points change, whereas the only stable minimum under the sampling procedure should be the one due to the actual motion. This, however, is beyond the scope of this paper and is the subject of ongoing work.

Acknowledgments

This research was supported by AFOSR FA9550-06-1-0138 and ONR N00014-03-1-0850:P0001.

6. Appendix

Lemma 5. *Let E be as defined in* (3) *and fix a bound $\alpha > 0$ on the inverse depths, i.e. restric the domain of E to the set $\mathbb{S}^2 \times A$, where $A = \mathbb{R}^3 \times [-\alpha, \alpha]^N$. Assume that there exists two features X_i and X_j which are linearly independent. Then there exists a constant B such that, for any given unit velocity v, any minimizer $(\omega, \lambda_1, \ldots, \lambda_N) \in A$ of $E(v, \cdot) : A \to \mathbb{R}$ must satisfy $\|\omega\| \leq B$.*

Proof. Each summand in (3) satisfies $\|y_i - \widehat{x}_i^2\omega + \lambda_i \widehat{x}_i v\| \geq \|\widehat{x}_i^2\omega\| - \|y_i\| - \alpha\|\widehat{x}_i v\|$. Moreover, since there are at least two linearly independent features X_i and X_j, we can find $\gamma > 0$ such that for all $\omega \in \mathbb{R}^3$ we have $\max\{\|\widehat{x}_i^2\omega\|, \|\widehat{x}_j^2\omega\|\} \geq \gamma\|\omega\|$. Hence we get the bound

$$E(v, \omega, \lambda_1, \ldots, \lambda_N) \geq \\ \geq \sum_i \|\widehat{x}_i^2\omega\| - C_1 \geq \gamma\|w\| - C_1 \quad (37)$$

for some constant C_1. Therefore $\lim_{\|\omega\|\to\infty} E = \infty$ uniformly in v and the minimization problem can be restricted to a limited subset of A. □

References

[1] S. Agarwal, M. Chandraker, F. Kahl, D. Kriegman, and S. Belongie. Practical global optimization for multiview geometry. In *Proc. of the Europ. Conf. on Comp. Vision*, 2006.

[2] A. Chiuso, R. Brockett, and S. Soatto. Optimal structure from motion: local ambiguities and global estimates. *Intl. J. of Computer Vision*, 39(3):195–228, September 2000.

[3] G. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least-squares problems whose variables separate. *SIAM J. Numer. Anal.*, 10(2):413–532, 1973.

[4] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2000.

[5] D.J. Heeger and A.D. Jepson. Subspace methods for recovering rigid motion, part ii: Theory. In *RBCV-TR*, 1990.

[6] F. Kahl and D. Henrion. Globally optimal estimates for geometric reconstruction. In *Proc. of the Intl. Conf. on Comp. Vision*, 2005.

[7] J. Oliensis. The least-squares error for structure from infinitesimal motion. *Int. J. of Computer Vision*, 61(3):259–299, 2005.

[8] P. D. Sampson. Fitting conic section to "very scattered" data: An iterative refinement of the bookstein algorithm. *Computer Vision, Graphics and Image Processing*, 18:97–108, 1982.

[1] An anonymous reviewer suggested that our analysis should be extended to other cost functionals, such as Sampson's distance [8], that might be smooth without the need for boundedness or positivity constraints. However, we have focused on the least-squares reprojection error since this is the cost functional that was used to analyze the structure of local minima.