

# Iterative Extensions of the Sturm/Triggs Algorithm: Convergence and Nonconvergence

John Oliensis, *Senior Member, IEEE*, and Richard Hartley, *Senior Member, IEEE*

**Abstract**—We give the first complete theoretical convergence analysis for the iterative extensions of the Sturm/Triggs algorithm. We show that the simplest extension, SIESTA, converges to nonsense results. Another proposed extension has similar problems, and experiments with “balanced” iterations show that they can fail to converge or become unstable. We present CIESTA, an algorithm that avoids these problems. It is identical to SIESTA except for one simple extra computation. Under weak assumptions, we prove that CIESTA iteratively decreases an error and approaches fixed points. With one more assumption, we prove it converges uniquely. Our results imply that CIESTA gives a reliable way of initializing other algorithms such as bundle adjustment. A descent method such as Gauss-Newton can be used to minimize the CIESTA error, combining quadratic convergence with the advantage of minimizing in the projective depths. Experiments show that CIESTA performs better than other iterations.

**Index Terms**—Structure from motion, projective geometry, factorization, projective factorization, convergence, optimization, Sturm/Triggs algorithm.

## 1 INTRODUCTION

THE Sturm/Triggs (ST) algorithm [14] is a popular example of the factorization strategy [15] for estimating 3D structure and camera matrices from a collection of matched images. The factorization part of the algorithm needs starting estimates of the *projective depths*  $\lambda_n^i$ , which [14] obtained originally from image pairs. After [14], researchers noticed that the  $\lambda_n^i$  can be taken equal or close to 1 for important classes<sup>1</sup> of camera motions [16], [1], [6], [5]. For these motions, ST computes its estimates directly from  $\lambda_n^i$  with approximately known values and comes close to being a direct method [7].

To improve ST’s results, several researchers proposed iterative extensions that initialize the  $\lambda_n^i$  (typically) at 1, estimate the structure/cameras, use these estimates to recompute the  $\lambda_n^i$ , use the new  $\lambda_n^i$  to recompute the structure/cameras, and so on [16], [1], [6], [10], [5]. One common use is for initializing bundle adjustment (BA) [5], [17]; for example, a few ST iterations can extend an affine estimate computed by Tomasi and Kanade [15] to a projective initialization. In practice, the ST iteration initially approaches the correct estimate much faster than BA does [2].

Variant iterative extensions of ST include [1], [6], [10], and [5]. Notably, [5] recommends a “balancing” step [14] after computing the  $\lambda_n^i$  to readjust their values toward 1. This keeps

1. Besides the motion classes listed in [5], this also holds for the class of small motions.

- J. Oliensis is with the Computer Science Department, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030.  
E-mail: oliensis@cs.stevens.edu.
- R. Hartley is with the Department of Information Engineering, RSISE, Australian National University, ACT 0200, Australia.  
E-mail: Richard.Hartley@anu.edu.au.

Manuscript received 1 Dec. 2005; revised 4 Nov. 2006; accepted 16 Jan. 2007; published online 16 Mar. 2007.

Recommended for acceptance by B. Triggs.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0671-1205. Digital Object Identifier no. 10.1109/TPAMI.2007.1132.

the  $\lambda_n^i$  near the correct values (for many motion classes (see footnote 1)) and may reduce the estimates’ bias [5].

Despite the widespread use of these iterations, they have not been justified theoretically. Some even lack theoretical support for the basic notion that they should improve upon the input estimate. No iteration has been shown to converge sensibly, and in this paper, we argue (and, for one variant, prove) that the current ones do not.<sup>2</sup>

This neglect of theory prevailed because: the iterations behave well in practice [13]; most researchers did not run the ST iteration to convergence, since they considered it as providing initial estimates (for example, to start BA) but not final ones, so the issue of convergence seemed irrelevant.

However, convergence is important, because it relates to stability. No one had guaranteed the stability of the ST iterations. An iteration that does not converge or that converges to nonsense results (as some ST iterations do) may be unstable, possibly producing nonsense after a single iteration. In the past, with no guarantees on its behavior, one did not know how many iterations of ST to trust or whether iterating ST some fixed number of times could serve as a reliable front end for BA. By the same token, proving that an ST iteration converges and avoids obvious nonsense confirms that it behaves reliably when applied for a small number of iterations.

In this paper, we provide the missing analysis of the ST iterations. Our initial results, based on a combination of theory and experiment, are mainly negative. We prove rigorously that the simplest iterative extension of ST converges to nonsense (generically); we argue, based on theory and experiment, that a second iteration also converges to nonsense; and we report experiments indicating that the balanced iteration [5] can fail to converge.

Given the lack of a theoretically justified iteration, we propose a simple algorithm called CIESTA, which extends ST

2. Previous convergence proofs, for example, [10], [9], [12] did not check whether the iterations converged to nonsense solutions (the trivial minima discussed later).

to a sound iteration. Instead of balancing, which can be considered a kind of regularization, CIESTA uses a more standard and better understood regularization. We prove that CIESTA descends an error function, that it iterates toward “best achievable” estimates, and that these are stationary points of the error. In effect, we show that CIESTA converges and that it avoids obvious nonsense estimates (the trivial minima). Also, since CIESTA has a known error function, one can use a standard descent method such as Gauss-Newton to minimize it instead of CIESTA. This keeps the **ST** advantage of working directly with the projective depths and gives quadratic convergence. Most of our theoretical results appeared in [11] without proof.

## 1.1 Preliminaries

### 1.1.1 Notation

Given  $N$  quantities  $\zeta_a$  indexed by  $a$ , we use  $\{\zeta\} \in \mathbb{R}^N$  to denote the column vector whose  $a$ th element is  $\zeta_a$ . If  $A \in \mathbb{R}^{M \times N}$  is a matrix, we define  $\{A\} \in \mathbb{R}^{MN}$  as the column vector obtained by concatenating the columns of  $A$ . Given two matrices  $A$  and  $B$  in  $\mathbb{R}^{M \times N}$ , we use  $\theta(A, B) \equiv \theta(\{A\}, \{B\})$  to denote the angle between the vectorized matrices.

Let  $\|\cdot\|$  denote the Frobenius norm: For an arbitrary matrix  $A$ , we have  $\|A\| \equiv (\sum_{a,b} (A_{ab}^a)^2)^{1/2} = |\{A\}|$ . We denote the  $n$ th column of  $A$  by  $A_n$ , and the range of columns from  $a$  through  $b$  by  $A_{a:b}$ . We denote the  $i$ th row of  $A$  by  $A^i$ , and the range of rows from  $a$  through  $b$  by  $A^{a:b}$ .

For multiview geometry, we use a notation like that in [5]. Let  $\mathbf{X}_n \equiv (X_n; Y_n; Z_n; 1) \in \mathbb{R}^4$  represent the homogenous coordinates<sup>3</sup> of the  $n$ th 3D point (we use “;” to indicate a column vector), with  $n = 1, 2, \dots, N_p$ . Let  $\mathbf{x}_n^i \equiv (x_n^i; y_n^i; 1) \in \mathbb{R}^3$  be the homogenous image of the  $n$ th point in the  $i$ th image, for  $i = 1, \dots, N_I$ . Let  $M^i \in \mathbb{R}^{3 \times 4}$  be the  $i$ th camera matrix. Neglecting noise, we have  $\lambda_n^i \mathbf{x}_n^i = M^i \mathbf{X}_n$ , where the constants  $\lambda_n^i$  are the *projective depths*. We use  $\lambda \in \mathbb{R}^{N_I N_p}$  to denote the vector of all the projective depths ordered in the natural way: Considering the projective depths as a matrix  $\lambda$ , with  $[\lambda]_n^i = \lambda_n^i$ , we set  $\lambda \equiv \{\lambda\}$ .

### 1.1.2 Sturm-Triggs (ST) Algorithm

Let  $\mathcal{W}(\lambda) \in \mathbb{R}^{3N_I \times N_p}$  be the scaled data matrix consisting of the  $\mathbf{x}_n^i$  multiplied by the projective depths, with  $[\mathcal{W}(\lambda)]_n^{(3i-2):3i} = \lambda_n^i \mathbf{x}_n^i$ . The  $n$ th column of  $\mathcal{W}$  corresponds to the  $n$ th 3D point, and  $i$  indexes the rows of  $\mathcal{W}$  associated with the  $i$ th image. Let  $\mathcal{M} \in \mathbb{R}^{3N_I \times 4}$  consist of the camera matrices  $M^i$  concatenated one on top of the other and define the *structure matrix*  $\mathcal{X} \in \mathbb{R}^{4 \times N_p}$  so that its  $n$ th column  $\mathcal{X}_n$  is proportional to  $\mathbf{X}_n$ . Neglecting noise, and for known  $\lambda_n^i$ , we have  $\mathcal{W}(\lambda) = \mathcal{M}\mathcal{X}$  for all points and all images. Thus, the noiseless matrix  $\mathcal{W}(\lambda)$  factors into a product of matrices with rank  $\leq 4$  and must itself have rank  $\leq 4$ . **ST** exploits this property, computing the motion and structure simultaneously by finding the best rank  $\leq 4$  approximation of the noisy  $\mathcal{W}(\lambda)$ .

## 2 SIESTA AND NONCONVERGENCE

This section introduces and analyzes theoretically the Simplest Iterative Extension of the Sturm-Triggs Algorithm

(**SIESTA**). We define an error function and show that each SIESTA iteration decreases it except at fixed points. We have two reasons for presenting this result: 1) we need it for our later proofs; 2) researchers will probably continue to use SIESTA because of its simplicity, and the result helps to explain what SIESTA does. The error function makes explicit SIESTA’s criterion for a “good” estimate; by this criterion, each SIESTA iteration “improves” upon the previous estimate. The main result of the section shows that SIESTA has an important, if not crippling, disadvantage: generically, it converges to nonsense results.

### 2.1 SIESTA: Definition

Given an arbitrary matrix  $W$  and its singular value decomposition (SVD)  $W = UDV^T$ , we define the rank  $\leq 4$  approximating matrix  $\hat{W}$  by

$$\hat{W} = U\hat{D}V^T, \quad (1)$$

where  $\hat{D}$  is obtained from the diagonal matrix  $D$  by zeroing all but the first four diagonal entries. A standard result states that  $\hat{W}$  gives the best approximation to  $W$  (under the Frobenius norm) among all matrices with rank  $\leq 4$

$$\|W - \hat{W}\| = \min_{\text{rank}(Y) \leq 4} \|W - Y\|.$$

SIESTA repeatedly adjusts the  $\lambda_n^i$  to make the scaled data matrix  $\mathcal{W}(\lambda)$  closer to rank 4. Let  $\lambda^{(k)}$  and  $\mathcal{W}^{(k)} \equiv \mathcal{W}(\lambda^{(k)})$  give the estimates of the  $\lambda_n^i$  and  $\mathcal{W}$  in the  $k$ th iteration, and let  $\hat{\mathcal{W}}^{(k)}$  denote the best rank  $\leq 4$  approximation of  $\mathcal{W}^{(k)}$  (that is,  $\hat{\mathcal{W}}^{(k)} = \hat{\mathcal{W}}^{(k)}$ ). The algorithm is

#### SIESTA

- **Initialize** the projective depths  $\lambda$ . By default, we set all the  $\lambda_n^{i(0)}$  to 1.
- **Iteration  $k$ , Stage 1.** Given the scaled data matrix  $\mathcal{W}^{(k-1)} \equiv \mathcal{W}(\lambda^{(k-1)})$ , compute its best rank  $\leq 4$  approximation  $\hat{\mathcal{W}}$ . Set  $\hat{\mathcal{W}}^{(k-1)} = \hat{\mathcal{W}}$ .
- **Iteration  $k$ , Stage 2.** Given  $\hat{\mathcal{W}}^{(k-1)}$ , choose  $\lambda^{(k)}$  to give the closest matrix of the form  $\mathcal{W}(\lambda^{(k)})$ , that is,  $\lambda^{(k)} = \arg \min_{\lambda} \|\mathcal{W}(\lambda) - \hat{\mathcal{W}}^{(k-1)}\|$ .

With  $\hat{\mathbf{w}}_n^i \equiv [\hat{\mathcal{W}}]_n^{(3i-2):3i} \in \mathbb{R}^3$ , the  $\lambda$  computed in Stage 2 is given explicitly by

$$\lambda_n^i = \arg \min_{\alpha} |\alpha \mathbf{x}_n^i - \hat{\mathbf{w}}_n^i| = \mathbf{x}_n^i \cdot \hat{\mathbf{w}}_n^i / |\mathbf{x}_n^i|^2. \quad (2)$$

The SIESTA algorithm has a simple interpretation if one thinks of the  $\mathcal{W}^{(k)}$  and  $\hat{\mathcal{W}}^{(k)}$  as points in  $\mathbb{R}^{3N_I N_p}$ . For fixed image points  $\mathbf{x}_n^i$ , the set  $\mathcal{L}^{N_I N_p}$  of all  $\{\mathcal{W}(\lambda)\}$  is a linear subspace of dimension  $N_I N_p$  since its points are indexed by the  $N_I N_p$  projective depths

$$\begin{aligned} \mathbf{V} &\in \mathcal{L}^{N_I N_p} \Leftrightarrow \mathbf{V} \\ &= \left( \lambda_1^1 \mathbf{x}_1^1; \lambda_1^2 \mathbf{x}_1^2; \dots; \lambda_1^{N_I} \mathbf{x}_1^{N_I}; \lambda_2^1 \mathbf{x}_2^1; \dots; \lambda_2^{N_I} \mathbf{x}_2^{N_I}; \dots; \lambda_{N_p}^{N_I} \mathbf{x}_{N_p}^{N_I} \right). \end{aligned}$$

Let  $\hat{\Omega}$  denote the set of all  $\{Y\}$  in  $\mathbb{R}^{3N_I N_p}$  coming from matrices  $Y \in \mathbb{R}^{3N_I \times N_p}$  of rank  $\leq 4$ . Any  $\hat{\mathcal{W}}(\lambda)$  from Stage 1 gives a point in  $\hat{\Omega}$ . We can rewrite SIESTA as (see Fig. 1a):

- **Stage 1.** Given  $\{\mathcal{W}^{(k-1)}\}$ , find the closest point  $\{\hat{\mathcal{W}}^{(k-1)}\}$  from the set  $\hat{\Omega}$ .

3. Without loss of generality, we set the focal length to 1.

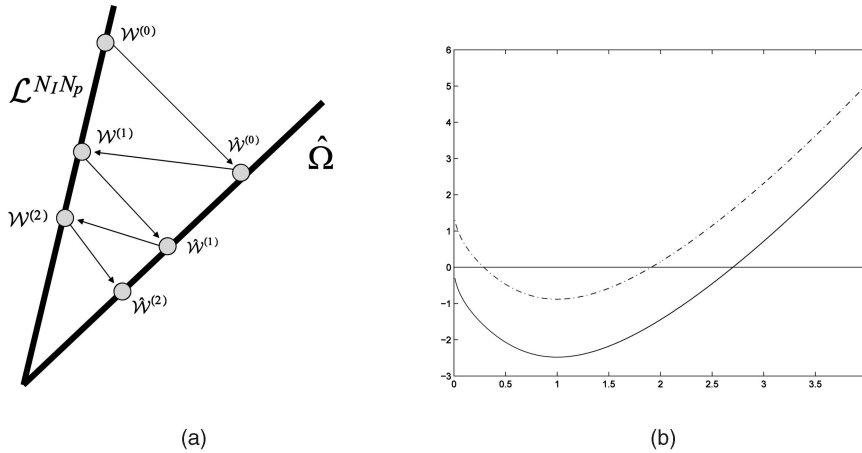


Fig. 1. (a) Illustration of SIESTA's alternating perpendicular projections into the sets  $\mathcal{L}^{N_I N_p}$  and  $\hat{\Omega}$ . The distance between successive points (from different sets) shrinks with each projection. (Our proof shows that the *angle* also shrinks.) (b) Functions  $T(\bar{a})$  and  $(1.6 + T(\bar{a}))$  (dashes) for  $f = 0.3$ .

- **Stage 2.** Given  $\{\hat{\mathcal{W}}^{(k-1)}\}$ , find the closest point  $\{\mathcal{W}^{(k)}\}$  from  $\mathcal{L}^{N_I N_p}$ .

We define the ‘‘closest point’’ using the standard Euclidean distance. Our definitions imply that the sets  $\mathcal{L}^{N_I N_p}$  and  $\hat{\Omega}$  always contain the closest point. Since  $\mathcal{L}^{N_I N_p}$  is a linear vector space, the  $\{\mathcal{W}^{(k)}\}$  computed by Stage 2 is simply the linear projection of  $\{\hat{\mathcal{W}}^{(k-1)}\}$  into  $\mathcal{L}^{N_I N_p}$ .

Next, we define an error function for SIESTA.

**Definition 1 (SIESTA error).** Define  $E(\mathcal{W}, Y) \equiv \|\mathcal{W} - Y\|/\|\mathcal{W}\|$  and

$$\hat{E}(\mathcal{W}) \equiv \min_{\text{rank}(Y) \leq 4} E(\mathcal{W}, Y), \quad \hat{E}(\lambda) \equiv \hat{E}(\mathcal{W}(\lambda)) = E(\mathcal{W}(\lambda), \hat{\mathcal{W}}).$$

$\hat{E}$  measures the fractional size of the nonrank 4 part of  $\mathcal{W}$ . Also, one can easily show that

$$\hat{E}(\mathcal{W}) = \sin(\theta(\mathcal{W}, \hat{\mathcal{W}})), \quad (3)$$

where  $\theta(\mathcal{W}, \hat{\mathcal{W}})$  gives the angle between  $\{\mathcal{W}\}$  and  $\{\hat{\mathcal{W}}\}$ , the vectors in  $\mathfrak{R}^{3N_I N_p}$  corresponding to the matrices  $\mathcal{W}$  and  $\hat{\mathcal{W}}$ . (Our proof of Proposition 2 in the Appendix shows (3) explicitly.)

Equation (3) shows that  $\hat{E}$  does not depend on the overall scale of  $\mathcal{W}$  or  $\lambda$ . This comes from the standard scale ambiguity of structure from motion. Similarly, one can check that SIESTA is *scale covariant*: scaling the input  $\lambda$  or  $\mathcal{W}(\lambda)$  changes SIESTA's output by the same scaling, leaving the 3D reconstruction unchanged. Instead of the space  $\mathfrak{R}^{N_I N_p}$  of all  $\lambda$ , we can consider SIESTA as operating on the sphere consisting of all unit vectors  $\lambda/|\lambda|$ .

The following show that SIESTA descends the error surface given by  $\hat{E}$ .

**Proposition 2.** The SIESTA error  $\hat{E}(\lambda^{(k)})$  is nonincreasing with  $k$ .

**Corollary 3.**  $\hat{E}(\lambda^{(k)})$  strictly decreases except at fixed points of the iteration.

**Proofs.** For easier reading, we collect all proofs into the Appendix. Proposition 2 was shown originally in [12]. Fig. 1a gives an idea of the proof.  $\square$

## 2.2 Bad Convergence of SIESTA

Generically, descent algorithms converge to local minima. We now show that there are no sensible local minima for SIESTA to converge to, which implies that SIESTA converges to a useful result only by a rare accident.

**Trivial minima.** We begin by describing trivial minima. If we choose the  $\lambda_n^i$  zero except in four columns, the matrix  $\mathcal{W}(\lambda)$  will have all but four columns composed of zeros. Then,  $\mathcal{W}(\lambda)$  will have rank  $\leq 4$  and

$$\hat{E}(\lambda) = E(\mathcal{W}(\lambda), \hat{\mathcal{W}}(\lambda)) = \frac{\|\mathcal{W}(\lambda) - \hat{\mathcal{W}}(\lambda)\|}{\|\mathcal{W}(\lambda)\|} = 0$$

because  $\mathcal{W}(\lambda)$  and its closest rank  $\leq 4$  matrix  $\hat{\mathcal{W}}(\lambda)$  are equal.

This set of  $\lambda_n^i$  gives a *trivial minimum* of the SIESTA error. Choosing all the  $\lambda_n^i$  zero except in one row also gives a trivial minimum. Trivial minima are of no interest, since they do not give reasonable interpretations of the data. Unfortunately, the proposition below shows that unless a nontrivial solution exists with exactly zero error (meaning that the data admits a noise-free solution), then the SIESTA algorithm must approach a trivial minimum, or possibly, in rare circumstances, a saddle point of the error. Experiments on small problems show that the algorithm approaches trivial minima though extremely slowly.

**Proposition 4.** Every local minimum of  $\hat{E}(\lambda)$  is a global minimum with zero error.

**Remark 5.** This result does not rule out *saddle points* with nonzero error values.

Our next lemma mitigates this negative result slightly, putting a limit on the speed with which SIESTA converges to the trivial minima. As before,  $\mathcal{W}^{(k)}$ ,  $\hat{\mathcal{W}}^{(k-1)}$ , and  $\lambda^{(k)}$  denote the output of the  $k$ th SIESTA iteration.

**Definition 6.** Let  $\theta^{(k,m)} \equiv \theta(\mathcal{W}^{(k)}, \mathcal{W}^{(m)})$  and  $\hat{\theta}^{(k,m)} \equiv \theta(\mathcal{W}^{(k)}, \hat{\mathcal{W}}^{(m)})$  (so,  $\hat{E}(\lambda^{(k)}) = \sin(\hat{\theta}^{(k,k)})$ ). Let  $\Delta_k^m \equiv \hat{\theta}^{(m-1,m-1)} - \hat{\theta}^{(k,k)}$ .

**Lemma 7.** For any  $K > 0$  and for all  $k \geq K$ ,

$$\cos(\theta^{(k,k-1)}) \geq \cos(\Delta_k^K) - \tan(\hat{\theta}^{(K-1,K-1)}) \sin(\Delta_k^K). \quad (4)$$

Also,

$$0 = \lim_{K \rightarrow \infty} \Delta_K^K = \lim_{k \rightarrow \infty} \theta^{(k,k-1)} = \lim_{k \rightarrow \infty} \theta(\lambda^{(k)}, \lambda^{(k-1)}).$$

Since  $\theta^{(k,k-1)}$  gives the angular change in  $\mathcal{W}(\lambda)$  caused by the  $k$ th SIESTA iteration, the lemma gives an upper bound on how much  $\lambda/|\lambda|$  changes in any single iteration.

**Discussion.** The fact that the error function of SIESTA has trivial minima does not by itself rule out the algorithm's usefulness. For example, the Expectation-Maximization (EM) algorithm [3] is a powerful and popular technique for image segmentation, but its error function (the inverse likelihood) also has trivial global minima at zero error. In practice, EM rarely converges to the trivial minima, and its nontrivial local minima at positive error often give good segmentations. Past experiments suggested that the situation for SIESTA was similar.

We have shown that the real situation for SIESTA is worse: Its error has *no* local minima giving good estimates; all the minima are trivial. Does it still make sense to use SIESTA? This will depend on how much reliability the user requires, and on the user's comfort level in relying on SIESTA's observed good behavior [13] as opposed to its theoretically demonstrated potential for bad behavior. Lemma 7 may help motivate the practice of applying SIESTA for a few iterations, since it gives limits on how fast SIESTA's output can depart from the "sensible" regime around the starting  $\lambda$  estimate. However, we have no proof that SIESTA converges initially toward good estimates rather than the trivial minima. Even if it does, we have no studies indicating what values of  $\mathcal{W}$  and  $\lambda$  are safe, for example, close enough to the starting estimates to give good results. These concerns also apply to the balanced iteration discussed later.

### 3 OTHER ITERATIVE EXTENSIONS OF STURM/TRIGGS

#### 3.1 The Iterations in [9], [10], [1]

Mahamud et al. [9], [10] proposed an iteration that differs from SIESTA by maintaining a normalization constraint on the columns of  $\mathcal{W}$ .<sup>4</sup> The first stage of the iteration is the same as in SIESTA, and the second stage is

- *Iteration  $k$ , Stage 2.* Given  $\hat{\mathcal{W}}^{(k-1)}$ , choose new projective depths  $\lambda^{(k)}$  so that  $\mathcal{W}(\lambda^{(k)})$  best approximates  $\hat{\mathcal{W}}^{(k-1)}$  subject to the  $N_p$  columns constraints  $|\mathcal{W}_n| = 1$ ,  $n \in \{1 \dots N_p\}$ .

The constraints fix  $\|\mathcal{W}\|^2 = N_p$ , so the SIESTA error  $\hat{E}$  reduces in effect to  $\|\mathcal{W} - \hat{\mathcal{W}}\|^2$ . It is easy to show that the iteration descends this error subject to the constraints [10]. The constrained error possibly does have nontrivial minima, but we argue below that it does not have useful minima corresponding to good structure/camera estimates.

Berthilsson et al. [1], [6] proposed a SIESTA variant roughly dual to that in [10] but did not give an error for it. A similar iteration that descends an error is SIESTA with the new Stage 2

- *Iteration  $k$ , Stage 2.* Given  $\hat{\mathcal{W}}^{(k-1)}$ , choose new projective depths  $\lambda^{(k)}$  so that  $\mathcal{W}(\lambda^{(k)})$  optimally approximates  $\hat{\mathcal{W}}^{(k-1)}$  subject to the  $N_l$  image constraints  $\|\mathcal{W}^{(3i-2):3i}\| = 1$ , where each matrix  $\mathcal{W}^{(3i-2):3i} \in \mathbb{R}^{3 \times N_p}$  gives the three rows of  $\mathcal{W}$  for image  $i$ .

It is easy to show that this iteration descends  $\|\mathcal{W} - \hat{\mathcal{W}}\|^2$ . We have not analyzed its convergence, but we expect it has problems similar to those of the Mahamud et al. iteration.

One might also consider an iteration minimizing  $\|\mathcal{W}(\lambda) - Y\|$  alternately over  $Y$  of rank  $\leq 4$  and  $\lambda$  constrained to 1 for a chosen image and for a chosen point. The constraint on  $\lambda$  eliminates its invariance under row/column rescalings, but this error still has a trivial minimum when all unconstrained  $\lambda = 0$ . Also, "fixing the gauge" based on the wrong point or image can lead to bad estimates.

**Convergence analysis for the iteration in [9] and [10]** (see footnote 4). Our results are weaker than for SIESTA, so we just summarize them; the Appendix gives more detail. For SIESTA, we ruled out nontrivial minima of the error by showing that a stationary point at an arbitrary  $\lambda$  with  $\hat{E}(\lambda) > 0$  has a direction along which one can decrease the error by infinitesimally perturbing  $\lambda$ . Such a stationary point must be a saddle, not a minimum. We could not prove this for the new algorithm, but we argue that the same conclusion holds at any "desirable" stationary point, where by this we mean one that gives a good estimate and also has  $\lambda$  values near 1 (as is appropriate for many types of motion).

The argument goes as follows: For any stationary point, we consider the error's second derivative in  $\lambda$  along a particular direction and derive two separate upper bounds on it. (We have checked these experimentally on thousands of synthetic sequences.) We then argue that the upper bounds are negative at "desirable" stationary points, so these cannot be minima.

We justify our theoretical arguments by experiments on real (see Fig. 2) and synthetic images. Table 1 shows some of the results. We obtained these by running the Mahamud et al. algorithm for 1,000 iterations; after which the error was changing so slowly that the algorithm seemed to have converged. Assuming that the algorithm had reached a stationary point, we evaluated the upper bounds. In all but one case, we found negative values for both bounds, proving the algorithm had not in fact converged to a minimum. In the exceptional case, the  $\lambda_n^i$  for a few points had shrunk toward a trivial minimum and become small; presumably, the algorithm is en route to the trivial minimum and has not reached a stationary point. Repeating the experiment without these points gave a negative bound.

This last experiment illustrates that the **ST** iterations can move toward trivial zeros by shrinking the  $\lambda$  for just one or a few points or one or a few images.

We also used standard nonlinear minimization (lsqnonlin from Matlab) to minimize the error for the Mahamud algorithm. The routine always converged to a trivial minimum. These results indicate that the convergence of the Mahamud algorithm is problematic at best.

#### 3.2 Balancing

Hartley and Zisserman [5] modifies SIESTA by adding a third "balancing" stage [16] following Stage 2 that rescales the  $\lambda_n^i$  to make them close to 1. This lessens the algorithm's bias and helps to steer it away from trivial minima. The balancing can

4. Mahamud et al. [9] also proposed a different iteration that minimizes alternately with respect to the camera and structure matrices. This loses the advantage of minimizing in the  $\lambda_n^i$ —it cannot exploit prior knowledge that the  $\lambda_n^i$  are near one.

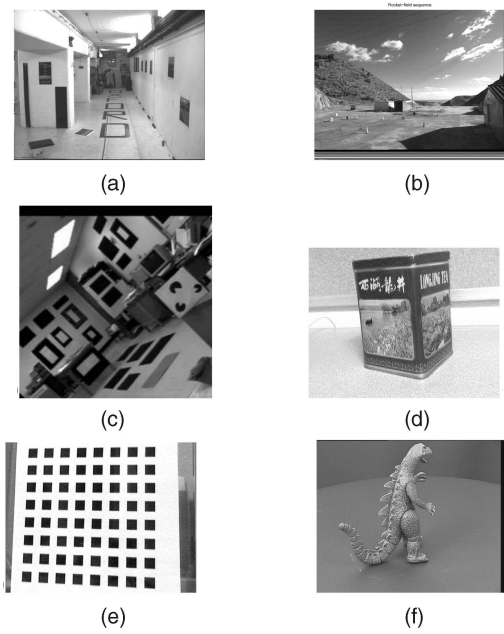


Fig. 2. Images from the six real sequences used. (a) Oxford **Corridor** sequence (11 images, 104 points). (b) **Rocket-Field** [4] (9 images, 22 points). (c) **Puma** [8] (16 images, 32 points). (d) Microsoft tea sequence [18] (2 images, 16 points). (e) Microsoft Plane calibration sequence [18] (2 images, 256 points). (f) Oxford/Hanover **Dinosaur** sequence (6 images, 31 points).

be done in two passes, by first scaling  $\lambda_n^i \rightarrow \alpha^i \lambda_n^i$  for each image  $i$  so that  $\sum_{n=1}^{N_p} |\lambda_n^i|^2 = N_p$  and then scaling  $\lambda_n^i \rightarrow \beta_n \lambda_n^i$  for each point so  $\sum_{i=1}^{N_l} |\lambda_n^i|^2 = N_l$ . Optionally, this can be repeated several times or iterated to convergence. Unfortunately, it seems likely that the rescaling conflicts with the error minimization in Stage 1 and that the balanced iteration need not converge. To exaggerate this potential conflict and make it more observable, we implemented SIESTA with a balancing stage that iterates to near convergence, with up to 10 rounds of first balancing the rows and then the columns of

$\lambda_n^i$ . In one of our small number of experiments, this algorithm apparently converged to a limit cycle that repeatedly passed through three different values for  $\lambda$  with three different values of the error  $\hat{E}(\lambda)$ . For this strong version of balancing, it seems that an iteration of SIESTA-plus-balancing does not guarantee improvement in the projective-depth estimates. There is no difference in kind between 10-fold balancing and onefold balancing, so it is possible that SIESTA plus one balancing cycle will also show nonconvergence in some situations.

## 4 CIESTA

Balancing adds into SIESTA an a priori preference for  $\lambda \approx 1$ ; it is a type of regularization. This is not always emphasized, since the regularization does not perturb the zero error estimates of SIESTA; for *zero noise*, balanced SIESTA keeps the true 3D reconstruction as a fixed point.

In the more interesting case of nonzero noise, balancing *does* perturb the estimates, dragging them toward  $\lambda \approx 1$ . It acts at cross purposes with SIESTA's Stage 2, which tries to explain the data by minimizing  $\hat{E}$ , whereas balancing imposes an a priori preference for  $\lambda \approx 1$ . Our experiments in Section 4.1 below and the limit cycle above show that balancing alters the iteration's final estimate.

One disadvantage of regularizing by balancing is that it is difficult to know how much regularization a given degree of balancing imposes. Does one cycle of balancing give a large or small regularization? Is it enough to prevent the iteration from converging to trivial minima or to prevent the  $\lambda$  for a single point becoming small? In contrast, by using a more standard regularization, we can control the amount of regularization more easily and determine exactly how much we need to rule out convergence to the trivial zeros.

We define a new iteration CIESTA that tries to minimize a regularized error  $\hat{E}_{\text{reg}}(\lambda)$ , where

TABLE 1  
Results for Five Real Sequences in Fig. 2

	$\lambda$	First	Least $\mu$	$\hat{E}^{1/2}$	Range for $\max_{r_n}$	$\max_k$	Second	
	Range	Bound	ratio	( $\times 10^{-3}$ )	$\frac{ \mathbf{w}_n ^2}{\langle \mathbf{w}  ^2} - 1$	$\frac{N_p}{D_1^2}$	$\frac{d_3^2}{\ \mathbf{w}\ ^2}$	Bound
<b>Corr 0-10</b>	0.79–1.15	−0.06	0.40	1.6	0.01–0.37	1.3	0.08	−0.09
<b>Corr 0&amp;10</b>	0.83–1.14	−0.24	0.78	1.8	0.29–0.31	1.3	0.10	0.44
<b>Corr 0&amp;8</b>	0.14–1.40	0.52	0.73	8.2	0.98–1.08	1.3	0.09	1.9
<b>Corr 0&amp;8*</b>	0.85–1.13	−0.28	0.78	1.6	0.24–0.25	1.3	0.09	0.37
<b>Corr 0&amp;1</b>	0.97–1.03	−0.46	0.75	0.5	0.04–0.04	1.3	0.12	−0.07
<b>Rock</b>	0.90–1.09	−0.07	0.46	6.0	0.03–0.20	1.4	0.08	0.01
<b>Puma</b>	0.99–1.01	−0.05	0.50	1.7	0.01–0.06	1.5	0.14	−0.21
MSTea	0.97–1.03	−0.43	0.75	0.7	0.05–0.05	1.5	0.16	−0.14
MSPlane	0.91–1.10	−0.13	0.44	0.6	0.05–0.13	1.5	0.18	−0.17

“**Corr 0-10**” is for all 11 images; other **Corr** rows are for image pairs. (“**Corr 0&8\***” is for images 0 and 8, with three problem points subtracted, see text.) “First/second bound” columns correspond to (22) and (23). The “ $\mu$  ratio” column gives the least value of  $|\hat{\mu}_3^k|^2 / \langle \hat{\mu}^k |^2 \rangle$  over images satisfying (21). Other maxima are over all points and all images (columns 5 and 7).

$$\hat{E}_{\text{reg}}(\lambda) \equiv \min_{\text{rank}(Y) \leq 4} E_{\text{reg}},$$

$$E_{\text{reg}}(\lambda, Y) \equiv \frac{\|\mathcal{W}(\lambda) - Y\|^2}{\|\mathcal{W}(\lambda)\|^2} + \mu \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} |\mathbf{x}_n^i|^2 (1 - \lambda_n^i)^2 \quad (5)$$

and  $\mu > 0$  is the regularization constant. We use this form of regularization just to simplify the calculations; others are possible. Like SIESTA, CIESTA minimizes the error alternately with respect to  $Y$  and  $\lambda$ . The algorithm is the same as SIESTA except for a new third stage.

Let  $\lambda^{(k)} \in \mathfrak{R}^{N_I N_p}$  and  $\mathcal{W}^{(k)} \equiv \mathcal{W}(\lambda^{(k)})$  now denote the output of the  $k$ th CIESTA iteration, and let  $\lambda^{(0)}$  and  $\mathcal{W}^{(0)}$  give the initialization. As before, let  $\hat{\mathcal{W}}^{(k)}$  be the best approximation of rank  $\leq 4$  to  $\mathcal{W}^{(k)}$ , and let  $\hat{w}_n^{i(k)} \equiv [\hat{\mathcal{W}}^{(k)}]_{(3i-2):3i}^n$ .

Define the constants

$$C_0 = \mu \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} |\mathbf{x}_n^i|^2, \quad C_1^{(k)} \equiv \mu \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} \mathbf{x}_n^i \cdot \hat{\mathbf{w}}_n^{i(k)},$$

$$C_2^{(k)} \equiv \mu \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} \frac{(\mathbf{x}_n^i \cdot \hat{\mathbf{w}}_n^{i(k)})^2}{|\mathbf{x}_n^i|^2}, \quad C_3^{(k)} = \mu \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} |\hat{\mathbf{w}}_n^{i(k)}|^2, \quad (6)$$

the combination  $z^{(k)}$  and the function  $b_+^{(k)}$

$$z^{(k)} \equiv C_3^{(k)} C_0 / C_2^{(k)},$$

$$b_+^{(k)}(a) \equiv a^{1/2} / \left( a^2 C_0 + 2a C_1^{(k)} + C_2^{(k)} \right)^{1/2}, \quad (7)$$

and the sixth-degree polynomial  $P^{(k)}(a)$

$$\begin{aligned} &\equiv C_0 a^6 - \left( C_0^2 - 2C_1^{(k)} \right) a^5 - \left( 2C_0 C_3^{(k)} - C_2^{(k)} \right) a^4 \\ &- \left( 4C_1^{(k)} C_3^{(k)} - 2C_2^{(k)} C_0 \right) a^3 + \left( C_0 C_3^{(k)2} - 2C_2^{(k)} C_3^{(k)} \right) a^2 \\ &+ \left( 2C_1^{(k)} C_3^{(k)2} - C_2^{(k)2} \right) a + C_2^{(k)} C_3^{(k)2}. \end{aligned} \quad (8)$$

The new third stage of CIESTA is

- **CIESTA** (iteration  $k$ , Stage 3). Find the positive roots of  $P^{(k-1)}(a)$ . When there are several: if  $z^{(k-1)} \neq 1$ , choose the one with  $a/\sqrt{C_3^{(k-1)}}$  and  $z^{(k-1)}$  on the same side of 1; if  $z^{(k-1)} = 1$ , choose  $a \neq (C_2^{(k-1)}/C_0)^{1/2}$ . Redefine  $\lambda^{(k)} \rightarrow \lambda^{(k)} = (a + \lambda^{(k)}) b_+^{(k-1)}(a)$ .

The extra cost of Stage 3 is small: it just requires finding the eigenvalues of a  $6 \times 6$  matrix.

The four propositions below address the convergence of CIESTA. They all assume

**Assumption 8 ( $\mu$  condition).** CIESTA starts with all  $\lambda_n^i = 1$ , and

$$\mu > \frac{\|\mathcal{W}^{(0)} - \hat{\mathcal{W}}^{(0)}\|^2}{\|\mathcal{W}^{(0)}\|^2 \left( \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} |\mathbf{x}_n^i|^2 \right)} = \frac{\|\mathcal{W}^{(0)} - \hat{\mathcal{W}}^{(0)}\|^2}{\|\mathcal{W}^{(0)}\|^4} \equiv \mu_{\text{bnd}}, \quad (9)$$

which is equivalent to  $C_0^2 > \left( C_0 + C_3^{(0)} - 2C_1^{(0)} \right)$ .

Starting CIESTA at  $\lambda_n^i = 1$  is not crucial; we assume it just to simplify the theorems and proofs. The main point of the Assumption is to ensure that  $|\lambda^{(k)}| > 0$  for all  $k$ , so we avoid the singularity at  $|\lambda| = 0$ . By taking  $\mu$  large enough,

we can also rule out the trivial minima.<sup>5</sup> The more  $\lambda$  with large values we want to ensure, the bigger we have to take  $\mu$ . However, the  $\mu$  bound in Assumption 8 is crude: Unless  $\lambda = 1$  makes  $\mathcal{W}$  close to rank 4, it is likely to be much bigger than necessary. The same holds for the  $\mu$  required to rule out the trivial minima, and the  $\mu$  bound needed to ensure a given number of large  $\lambda_n^i$ .

Below, we use the following definitions:  $\hat{E}_\infty$  denotes the greatest lower bound of the  $\hat{E}_{\text{reg}}(\lambda^{(k)})$ , and  $\mathcal{A}$  gives the set of accumulation points of the  $\lambda^{(k)}$ .

**Proposition 9 (Monotonic error convergence for CIESTA).**

Suppose Assumption 8 holds. The errors  $\hat{E}_{\text{reg}}(\lambda^{(k)})$  are nonincreasing with  $k$ . As  $k \rightarrow \infty$ , they converge monotonically to  $\hat{E}_\infty \geq 0$ .

**Proposition 10 (Limit behavior of the projective depths).**

Suppose Assumption 8 holds. Then, 1) Every  $\lambda_A \in \mathcal{A}$  has  $\hat{E}_{\text{reg}}(\lambda_A) = \hat{E}_\infty$ ; 2) For any  $\epsilon > 0$ , there exists a  $K$  such that  $k > K$  implies  $|\lambda^{(k)} - \lambda_A| \leq \epsilon$  for some  $\lambda_A \in \mathcal{A}$ .

**Proposition 11 (Convergence to fixed points).** Suppose

Assumption 8 holds. Let  $\lambda_A \in \mathcal{A}$ . Assume the fourth singular value of  $\mathcal{W}(\lambda_A)$  is strictly greater than the fifth, and  $z^A \neq 1$ , where  $z^A$  is the constant from (6) evaluated at  $\lambda_A$ . Then,  $\lambda_A$  is a fixed point of CIESTA and a stationary point of the error  $\hat{E}_{\text{reg}}$  (not necessarily a minimum).

**Proposition 12 (Unique convergence for CIESTA).** Suppose

the assumptions of Proposition 11 hold for some  $\lambda_A \in \mathcal{A}$ . Proposition 11 states that  $\hat{E}_{\text{reg}}$  has a stationary point at  $\lambda_A$ . If in fact  $\hat{E}_{\text{reg}}$  has a strict local minimum at  $\lambda_A$ , then CIESTA converges uniquely to  $\lambda_A$ .

**Corollary 13.** Suppose Assumption 8 holds. Then, the convergence set  $\mathcal{A}$  is connected unless every connected component of  $\mathcal{A}$  contains a  $\lambda$  violating the assumptions of Proposition 11.

**Discussion.** Propositions 9 and 10 show that each CIESTA iteration ‘‘improves’’ the estimate and that CIESTA ‘‘converges’’ (in a sense discussed below). Proposition 11 states that its end results are sensible, that is, they are stationary points of the error. Proposition 12 shows, under an additional assumption, that CIESTA converges in a strict sense to a unique result.

We have not shown that CIESTA converges to a single  $\lambda$  (except under the assumptions of Proposition 12), and it is not clear whether this always happens. However, our results have the same practical implications as a convergence proof.

A convergence proof would amount to the following guarantee: By iterating enough times, one can bring the algorithm as close as desired to a ‘‘best achievable result,’’ that is, to a  $\lambda$  with the lowest error reachable from its starting point. This does not forbid other equally good estimates with the same error as the ‘‘best result’’ though the algorithm happens not to converge to them.

Proposition 10 provides essentially the same guarantee: with enough iterations, we can bring CIESTA arbitrarily close to a ‘‘best result.’’ The difference is that the nearest ‘‘best result’’ may change from iteration to iteration. This does not matter since all are good, and we may choose any

5. The required  $\mu$  is roughly  $N_p/(N_p - 4)$  (or  $N_I/(N_I - 1)$ ) times  $\mu_{\text{bnd}}$ . This is because the trivial minima send to zero all but a fraction  $4/N_p$  (or  $1/N_I$ ) of the  $\lambda$ , so only this fraction of the terms in the regularization sum can have smaller values than for  $\lambda = 0$ . If a  $\mu$  value smaller than  $\mu_{\text{bnd}}$  suffices to rule out  $\lambda = 0$ , the  $\mu$  required to rule out trivial minima will be proportionately less.

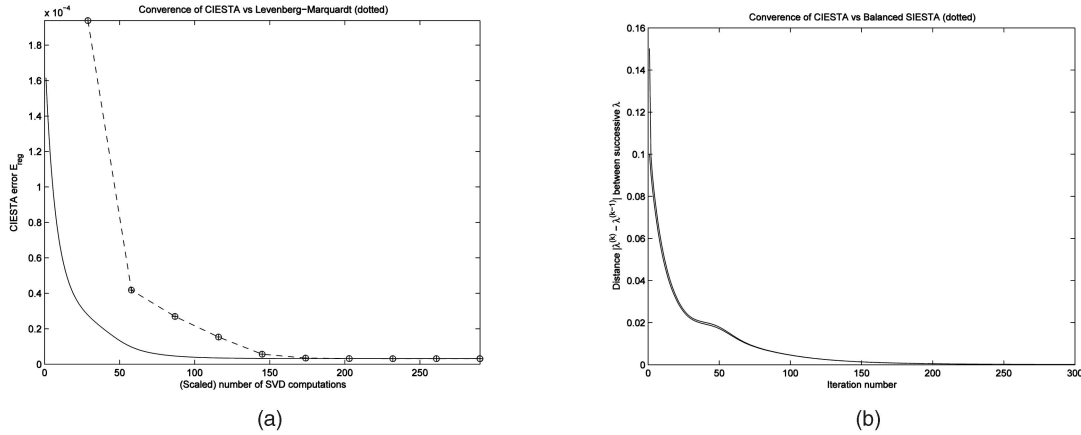


Fig. 3. (a) Convergence of CIESTA (solid line) versus LM applied to  $\hat{E}_{\text{reg}}$ . Shown is the computed  $\hat{E}_{\text{reg}}$  versus the number of SVD computations needed ( $x$ -axis). The LM curve has been scaled to fit: its  $x$  values should be *scaled up by 40* to give the true number of SVD. Circles give the LM results after each line search. (b) Similar convergence of CIESTA (solid) and SIESTA plus balancing (one round). Shown is  $|\lambda^{(k)} - \lambda^{(k-1)}|$  versus  $k$ . Both (a) and (b) were obtained on **Corr** (all images), see Table 1 and Fig. 2.

one as the final output. With Proposition 12’s assumptions, we do get strict convergence.

One can minimize  $\hat{E}_{\text{reg}}$  using a quadratically convergent technique such as Levenberg-Marquardt (LM) instead of CIESTA. This will eventually converge faster than CIESTA, but the results in [2] and our experiments (Fig. 3a) suggest that CIESTA gives much faster initial convergence. Each LM iteration has complexity  $\Omega(N_I^3 N_p^3)$ , since it minimizes in the  $(N_I N_p)\lambda$  parameters, so LM may be too slow at large  $N_I$  and/or  $N_p$ .<sup>6</sup> For comparison, each CIESTA iteration has complexity of  $O(N_p N_I^2)$  or  $O(N_I N_p^2)$ . Thus, CIESTA may be more useful for initializing BA.

In contrast with previous **ST** iterations, CIESTA can guarantee avoiding the trivial minima. We expect, and our experiments confirm, that it will often converge to a good estimate. Our reasons are as follows: For important motion classes, the true structure/motion corresponds to a rank  $\leq 4$  matrix  $\mathcal{W}_{\text{true}}$  with the exact  $\mathbf{x}_n^i$  scaled by  $\lambda_{\text{true},n}^i \approx 1$ . Since  $\|\mathcal{W}(\lambda_{\text{true}}) - \mathcal{W}_{\text{true}}\|$  is proportional to the noise  $\mathbf{x}_n^i - \mathbf{x}_{\text{exact},n}^i$  and small, it follows that  $\hat{E}_{\text{reg}}(\lambda_{\text{true}})$  is small. If there’s no ambiguity in estimating 3D from the images,  $\lambda_{\text{true}}$  causes the only deep valley in  $\hat{E}_{\text{reg}}$  near  $\lambda = 1$ . This suggests that the valley’s domain of attraction is likely to include  $\lambda = 1$  and that CIESTA started with  $\lambda = 1$  will often converge to the valley floor near the true estimate. A priori, one should have the same confidence in this that one would have in any steepest descent algorithm started from a point near the global minimum.

A potential disadvantage of CIESTA and balancing is that regularization may shift  $\lambda$  away from the best estimate. Our experiments suggest that this is a small effect. Also, CIESTA needs little regularization to stabilize it (we see no trivial minima with  $\mu = \mu_{\text{bnd}}/10$ ) and regularization effects may not matter if we do not care about convergence, for example, if we are just using a few iterations to initialize BA (see Table 5). As discussed in Section 2, the advantage of initializing with CIESTA is that we can have more confidence in its behavior even for a few iterations.

6. There is no obvious way to use sparseness as for Bundle Adjustment to reduce the cost.

If we had defined  $\hat{E}_{\text{reg}}$  without the denominator, see (5), we would have obtained a simpler iteration than CIESTA. However, this would cause serious problems. The regularization would then need to serve two purposes: besides countering the trivial minima, we would need it to prevent the  $\lambda$  and  $\hat{\mathcal{W}}$  shrinking uniformly toward 0 to reduce  $\|\mathcal{W} - \hat{\mathcal{W}}\|$ . A regularization large enough for this extra task would drag the global minimum far from the best estimate, whereas less regularization would allow small  $\lambda$  and increase the threat of trivial minima.

#### 4.1 CIESTA Experiments

Tables 2 and 3 compares the estimated  $\lambda$  and reconstruction errors<sup>7</sup> for SIESTA, CIESTA, and balancing on the real sequences in Fig. 2.<sup>8</sup> The sequence **Corr**\* consists of eight points over three images extracted from **Corr**. We computed CIESTA’s results using a standard nonlinear minimization routine (**lsqnonlin**, the Matlab implementation of LM), with  $\mu/\mu_{\text{bnd}}$  ranging from 0.1 to 10, where  $\mu$  is the regularization constant, and  $\mu_{\text{bnd}}$  is as in Assumption 8. We find

- Iterating to (apparent) convergence usually gives better estimates than a single iteration does [13]. See the odd rows (with sequence name) in Table 2.
- The different methods often give similar results with accuracies comparable to that of BA.<sup>9</sup>
- The balancing results depend little on the number of rounds of balancing.
- CIESTA is at least as accurate as the other **ST** iterations, and its accuracy does not depend strongly on  $\mu$  (odd rows in Table 2).
- CIESTA’s  $\lambda$  output depends on the  $\mu$  choice—but the dependence is weak once we exploit the invariance of the ground truth  $\lambda$  under rebalancing (even rows in Tables 2 and 3).

7. For the **Dinosaur** sequence, we computed the euclidean “ground-truth” structure  $P^{\text{GT}}$  by Bundle Adjustment.

8. The corresponding results reported in [11] are slightly wrong because of a bug.

9. BA’s poor result for **Puma** probably comes from inaccuracy in the original measurements rather than in BA. Its poor result on **Dino** arises from errors in the assumed calibration.

TABLE 2  
Structure Errors

	S1	S	Bal	CIESTA						GT	
				$(\mu/\mu_{\text{bnd}}) =$	0.1	0.5	1	2	5		10
<b>Corr</b>	344	53	53	54		53		54	56	55	
$E_\lambda$	2.0	.17	.17	.17		.16		.16	.17		
<b>Rock</b>	204	329 <sup>(294)</sup>	309 <sup>(293)</sup>	878	325	312		301		275	131
$E_\lambda$	0.40	.52 <sup>(.28)</sup>	.36 <sup>(.28)</sup>	.55	.47	.38		.31		.23	
<b>Puma</b>	.98	.50 <sup>(.52)</sup>	.52	.54		.48		.48		.50	166
$E_\lambda$	.05	.11	.11	.11		.10		.11		.11	
<b>Dino</b>	1.1	.22	.22	.23		.23		.23		.23	3.4
$E_\lambda$	.20	.10	.10	.09		.10		.10		.10	
<b>Corr*</b>	608	685 <sup>(703)</sup>	490 <sup>(682)</sup>	129		460	368 <sup>(1.5)</sup>			698	90
$E_\lambda$	.26	.97 <sup>(.56)</sup>	2.5 <sup>(.57)</sup>	.086		.61		.67		.54	

For each sequence (see Fig. 2), the upper row shows the fractional errors  $\sum_n |P_n^{\text{calc}} - P_n^{\text{GT}}|^2 / \sum_n |P_n^{\text{GT}}|^2$  (in units of  $10^{-4}$ ), where  $P_n^{\text{calc}} \in \mathbb{R}^3$  is computed after a projective transform to minimize the euclidean error. S1(S) are after one (10,000) SIESTA iterations, B is after 10,000 balanced iterations (but we used 2e5 for **Corr\***), and CIESTA is after LM convergence. S and Bal superscripts show results after 1,000 iterations when these differ. Balancing 1 or 10 rounds gave the same results. GT gives the error for Euclidean BA.  $E_\lambda$  (lower) rows give the error between estimated  $\lambda$  and ground truth each normalized by 10-fold balancing. **Corr\*** used  $\mu/\mu_{\text{bnd}} = 1.5$  instead of 2.

TABLE 3  
 $\lambda$  Ranges Corresponding to the Table 2 Results

Seq.	S	Bal	CIESTA						GT	
			$\frac{\mu}{\mu_{\text{bnd}}} =$	.1	.5	1	2	5		10
<b>Corr</b>	.70-1.1	.80-1.2		.42-1.2		.75-1.1		.79-1.1	.80-1.1	.81-1.2
Bal				.81-1.2		.80-1.2		.81-1.2	.81-1.2	
<b>Rock</b>	.17-1.1	.87-1.1		.05-1.6	.34-1.5	.54-1.3	.73-1.2		.90-1.1	
Bal				.77-1.2	.82-1.1	.87-1.1	.90-1.1		.90-1.1	.88-1.1
<b>Puma</b>	.72-1.0	.99-1.0		.003-1.3		.09-1.5	.18-1.3		.52-1.1	
Bal				.99-1.0		.99-1.0	.99-1.0		.99-1.0	.99-1.0
<b>Dino</b>	.80-1.0	.97-1.0		.03-1.8		.28-1.2			.83-1.0	
Bal				.97-1.0		.97-1.0			.97-1.0	.95-1.1
<b>Corr*</b>	.35-1.4	-.65-2.2		.71-1.1		.30-1.3	.37-1.3 <sup>(1.5)</sup>		.82-1.2	
Bal	.35-1.4	-.65-2.2		.87-1.1		.83-1.3	.80-1.3		.87-1.2	.83-1.1
<b>MsTEA</b>	.96-1.0	.97-1.0				.95-1.0				
<b>MsPLANE</b>	.95-1.1	.96-1.1				.96-1.0				

GT gives the ground truth  $\lambda$  range after 10 rounds of balancing. The "Bal" rows give the estimated  $\lambda$  range, also after 10 rounds. **Corr\*** used  $\mu/\mu_{\text{bnd}} = 1.5$ .

CIESTA's estimates for **Corr\*** depend more strongly on  $\mu$  than they do for the other sequences. This is as expected: a smaller amount of data constrains the reconstruction less, making it more subject to regularization. However, CIESTA does better than the other iterations over a wide  $\mu$  range.

For **Corr\***, the balancing regularization is too weak to prevent some  $\lambda$  escaping to 0 and then large negative values—even with 100 balancing rounds. This illustrates that one cannot easily adjust the amount of regularization in balancing. In our experiments, increasing the number of balancing rounds affected the iteration's progress very little—the regularization did not increase correspondingly

(for example, 1 or 10 balancing rounds gave identical  $\lambda$  after  $10^5$  iterations for **Rock**). Thus, balancing can have difficulty with sequences that require large regularization.

Our  $\mu$  values were not always big enough to prevent the converged  $\lambda$  becoming small for some points, see Table 4. We can avoid this by rebalancing the final  $\lambda$  estimate (exploiting its invariance) or by making  $\mu$  larger, as Table 3 shows. The small  $\lambda$  occurred for sequences where the camera projection is close to affine, making  $\mathcal{W}(\lambda = 1)$  nearly rank 4. From our discussion following Assumption 8, the  $\mu_{\text{bnd}}$  becomes small in this case, and fixing  $\mu \leq \mu_{\text{bnd}}$  may not be enough to prevent shrinkage in some  $\lambda_n^i$ . The  $\mu_{\text{bnd}}$  for the

TABLE 4  
 CIESTA

Thresh	Rock( $\mu/\mu_{\text{bnd}}=0.1$ )	Dino( $\mu/\mu_{\text{bnd}}=0.1$ )	Puma( $\mu/\mu_{\text{bnd}}=0.1$ )	Thresh	Puma( $\mu/\mu_{\text{bnd}}=1$ )
.4	6/22	16/31	5/32	.8	20/32
.2	9/22	23/31	5/32	.6	27/32

Number of points with median estimated  $\lambda$  above threshold shown.

near-affine **Puma** and **Dino** were  $4 \times 10^{-9}$ ,  $4 \times 10^{-8}$ , compared to  $10^{-7}$ ,  $7 \times 10^{-7}$  for **Corr** and **Rock**.

Table 5 shows that iterating LM to convergence did not have a big impact on CIESTA's results. In most cases, CIESTA's estimates after 10,000 iterations are close to those for LM. Comparing with Table 2 shows that CIESTA and balancing give similar results at 1,000 and 10,000 iterations, in accordance with Fig. 3b. The main effect of convergence is that the  $\lambda$  range is bigger than after 10,000 CIESTA iterations, especially for **Puma** (for example, for **Puma** with  $\mu/\mu_{\text{bnd}} = 1$  the range was just 0.75-1.0 at 10,000 iterations).

It is also striking in Table 5 that CIESTA's results during its first 1,000 iterations depend very weakly on the amount of regularization.

## 5 CONCLUSION

We gave the first complete theoretical convergence analysis for the iterative extensions of the ST algorithm. We showed that SIESTA, the simplest iterative extension of ST, descends an error function: Each iteration "improves" the estimates. However, we proved that SIESTA does not converge to useful results and showed that another proposed extension of ST [9] has similar problems. Hartley and Zisserman [5] advocate "balancing" to improve convergence. Our experiments suggest that the balanced iteration may not always converge and that it is sometimes unstable.

We proposed CIESTA, a new iterative extension of ST, which avoids these problems. CIESTA uses a more standard and flexible regularization than balancing. The algorithm is identical to SIESTA except for one simple extra computation. We proved that CIESTA descends an error function and, under a few weak assumptions, that it approaches nontrivial fixed points. With an additional assumption, we proved that CIESTA converges. Given a reasonable amount of regularization, one can prove that CIESTA avoids the trivial minima that plague other iterations; in practice, one can use CIESTA

with smaller regularizations. The accuracy of CIESTA's estimates does not depend strongly on the amount of regularization.

Our theoretical convergence results give CIESTA a certificate of good behavior. We argued and showed experimentally that CIESTA is likely to converge to good estimates; this implies that it heads in the right direction from its beginning iterations. Thus, CIESTA iterated a small number of times gives a *reliable* way of initializing other algorithms such as BA. There is no theoretical evidence for the reliability of any other ST iteration—even if used just for a few iterations as an initializer. Our experiments show that CIESTA converges as fast as the balanced iteration (Fig. 3b and Table 5), that its results are often more accurate (Table 2), and that it behaves stably when balancing does not (see the **Corr\*** results).

CIESTA, like other extensions of ST, has the advantage of minimizing in the  $\lambda_n^i$ , whose values are often known to be near one a priori. Thus, it often shows fast initial convergence toward estimates that are roughly correct (Fig. 3). Since it has a known error function, one can also minimize its error using a quadratically convergent method such as Gauss-Newton, which may be faster near fixed points or in narrow valleys of the error function. However, such methods are likely to be too costly when the number of points or images is large. A hybrid strategy that uses CIESTA initially and then switches to BA can combine the speed advantages of both [2].

## APPENDIX

**Proof of Proposition 2 and Corollary 3 (SIESTA decreases the error).** Each step of SIESTA selects a new point either from  $\hat{\Omega}$  or  $\mathcal{L}^{N_i N_p}$  that is closest to the point last selected. We will use the fact that these sets are both closed under scaling, for example,  $\mathbf{V} \in \hat{\Omega}$  implies that  $\sigma \mathbf{V} \in \hat{\Omega}$  for all  $\sigma \in \mathbb{R}$ .

Any set  $\mathcal{S} \in \mathbb{R}^N$  closed under scaling can be thought of as a collection of rays through the origin. Given any nonzero  $\mathbf{V} \in \mathbb{R}^N$ , consider the point in  $\mathcal{S}$  that is nearest  $\mathbf{V}$ ; we denote it  $\mathbf{V}_S = \arg \min_{\mathbf{V}' \in \mathcal{S}} |\mathbf{V}' - \mathbf{V}|$ . Since  $\mathcal{S}$  contains the whole ray through  $\mathbf{V}_S$ , clearly,  $\mathbf{V}_S$  is the nearest point on the ray to  $\mathbf{V}$ , so it is  $\mathbf{V}$ 's perpendicular projection onto the ray, with

$$\|\mathbf{V}_S - \mathbf{V}\| = |\mathbf{V}| \sin \theta_{\mathbf{V}_S}. \quad (10)$$

Here,  $\theta_{\mathbf{V}_S}$  is the angle between  $\mathbf{V}_S$  and  $\mathbf{V}$ . The angle  $\theta_{\mathbf{V}_S}$  must give the *least* angle between  $\mathbf{V}$  and any vector in  $\mathcal{S}$  (if there were an  $\mathbf{S}' \in \mathcal{S}$  with  $\theta' < \theta_{\mathbf{V}_S}$ , the least distance from  $\mathbf{V}$  to the set  $\mathcal{S}$  would be  $|\mathbf{V}| \sin \theta' < |\mathbf{V} - \mathbf{V}_S|$ , so  $\mathbf{V}_S$  would not be the closest to  $\mathbf{V}$ ).

 TABLE 5  
 CIESTA Fractional Structure Errors (Table 2) after  
 1,000 or 10,000 Iterations, or LM Convergence

$\mu/\mu_{\text{bnd}} =$	.1	1	10
<b>Corr</b>	53 53 54	53 53 53	56 56 56
<b>Rock</b>	294 327 878	292 310 312	274 275 275
<b>Dino</b>	.22 .22 .23	.22 .22 .23	.22 .23 .23
<b>Puma</b>	.52 .50 .54	.52 .50 .48	.52 .50 .50
<b>Corr*</b>	700 659 129	700 460 460	643 698 698

SIESTA chooses a point, alternately from  $\hat{\Omega}$  and  $\mathcal{L}^{N_t N_p}$ , that is the closest to the last one chosen in the other set. The most recent point chosen in  $\hat{\Omega}$  is the closest among all  $\hat{\Omega}$  choices to the most recent point in  $\mathcal{L}^{N_t N_p}$ ; in particular,  $|\{\hat{\mathcal{W}}^k\} - \{\mathcal{W}^k\}| \leq |\{\hat{\mathcal{W}}^{k-1}\} - \{\mathcal{W}^k\}|$ . The same holds for the most recent point from  $\mathcal{L}^{N_t N_p}$ . Thus, the distance between the two most recent points, one from each set, is nonincreasing; see Fig. 1a. Since the two sets are closed under scaling, it follows that the angle between the two most recent points is also nonincreasing:

$$\theta(\mathcal{W}^{(k)}, \hat{\mathcal{W}}^{(k)}) \leq \theta(\mathcal{W}^{(k)}, \hat{\mathcal{W}}^{(k-1)}) \quad (11)$$

after Stage 1, and

$$\theta(\mathcal{W}^{(k)}, \hat{\mathcal{W}}^{(k-1)}) \leq \theta(\mathcal{W}^{(k-1)}, \hat{\mathcal{W}}^{(k-1)}) \quad (12)$$

for the previous Stage 2. Together, these imply  $\theta(\mathcal{W}^{(k)}, \hat{\mathcal{W}}^{(k)})$  is nonincreasing with  $k$ .

From (10), the sine of  $\theta(\mathcal{W}^{(k)}, \hat{\mathcal{W}}^{(k)})$  relates simply to the error  $\hat{E}(\lambda^{(k)})$  in (1)

$$\begin{aligned} \sin^2\left(\theta\left(\mathcal{W}^{(k)}, \hat{\mathcal{W}}^{(k)}\right)\right) &= \frac{|\{\mathcal{W}^{(k)}\} - \{\hat{\mathcal{W}}^{(k)}\}|^2}{|\{\mathcal{W}^{(k)}\}|^2} \\ &= \min_{\text{rank}(Y) \leq 4} \frac{\|\mathcal{W}(\lambda^{(k)}) - Y\|^2}{\|\mathcal{W}(\lambda^{(k)})\|^2} = \hat{E}(\lambda^{(k)}). \end{aligned} \quad (13)$$

The quantities  $\hat{E}(\lambda^{(k)})$  are nonincreasing since the  $\theta(\mathcal{W}^{(k)}, \hat{\mathcal{W}}^{(k)})$  are. One can easily prove Corollary 3 by exploiting the linearity of Stage 2's projection onto  $\mathcal{L}^{N_t N_p}$ .  $\square$

**Proof of Proposition 4 (SIESTA nonconvergence).** We can assume without loss of generality that every 3D point has nonzero  $\lambda_n^i$  in some images, since otherwise we can eliminate these points and apply the argument below to the remaining set of points.

We suppose that the SIESTA error  $\hat{E}$  has a local minimum at  $\lambda$ . Let  $\mathcal{W} = \mathcal{W}(\lambda)$  be the corresponding scaled data matrix. By the assumption just above,  $\mathcal{W}$  has no columns consisting entirely of zeros. Under these two conditions, we will show that the error equals zero or, equivalently, that  $\mathcal{W}$  has rank  $\leq 4$ .

Consider a transformation that perturbs a matrix by multiplying its  $n$ th column by a value  $s$ . We denote this transformation by  $\tau_{n\kappa}$ , where  $\kappa = s^2 - 1$ . The reason for introducing the variable  $\kappa$  is that the subsequent computations simplify when expressed in terms of  $\kappa$ . For  $\kappa = 0$ , the transformation  $\tau_{n\kappa}$  is the identity transformation and leaves the original matrix unchanged. It is evident that applying  $\tau_{n\kappa}$  to the matrix  $\mathcal{W} = \mathcal{W}(\lambda)$  is equivalent to multiplying the  $n$ th column of the projective depths  $\lambda_n^i$  by  $s$ , so we can write  $\tau_{n\kappa}(\mathcal{W}) = \mathcal{W}(\lambda^{\tau_{n\kappa}})$ , where  $\lambda^{\tau_{n\kappa}}$  equals  $\lambda$  except for the appropriate scaling of the  $n$ th column. For small  $s$ , the transformation  $\tau_{n\kappa}$  moves  $\mathcal{W}$  and  $\lambda$  toward a trivial zero.

For the remainder of the proof, we write simply  $\mathcal{W}$ , omitting the dependence on  $\lambda$ . We denote  $\tau_{n\kappa}(\mathcal{W})$  by  $\mathcal{W}^\tau$  and the nearest<sup>10</sup> rank  $\leq 4$  matrix to  $\mathcal{W}^\tau$  by  $\hat{\mathcal{W}}^\tau$ . Recall that, similarly,  $\hat{\mathcal{W}}$  is the closest matrix to  $\mathcal{W}$  having

rank  $\leq 4$ . We may also apply the transformation  $\tau_{n\kappa}$  to  $\hat{\mathcal{W}}$ , resulting in a matrix  $(\hat{\mathcal{W}})^\tau = \tau_{n\kappa}(\hat{\mathcal{W}})$ . This matrix has the same rank as  $\hat{\mathcal{W}}$  for  $s \neq 0$  and, hence, has rank  $\leq 4$ , but, as we shall see, it is in general distinct from  $\hat{\mathcal{W}}^\tau$ . It is important to understand the difference between  $(\hat{\mathcal{W}})^\tau$  and  $\hat{\mathcal{W}}^\tau$ .

As a first step, we show (under our assumptions above) that any  $\kappa \neq 0$  gives

$$\hat{E}(\mathcal{W}^\tau) \equiv E(\mathcal{W}^\tau, \hat{\mathcal{W}}^\tau) \leq E(\mathcal{W}^\tau, (\hat{\mathcal{W}})^\tau) = E(\mathcal{W}, \hat{\mathcal{W}}) \equiv \hat{E}(\mathcal{W}). \quad (14)$$

The inequality in (14) follows simply from the definition of the error  $E$ , and the fact that  $\hat{\mathcal{W}}^\tau$  is the closest matrix to  $\mathcal{W}^\tau$  having rank  $\leq 4$ . Next we show  $E(\mathcal{W}^\tau, (\hat{\mathcal{W}})^\tau) = E(\mathcal{W}, \hat{\mathcal{W}})$ . Noting that  $\mathcal{W}$  and  $\mathcal{W}^\tau$  differ only in the overall scale of their  $n$ th columns, we may compute

$$E(\mathcal{W}^\tau, (\hat{\mathcal{W}})^\tau) = \frac{\kappa |\mathcal{R}_n|^2 + \|\mathcal{R}\|^2}{\kappa |\mathcal{W}_n|^2 + \|\mathcal{W}\|^2}, \quad (15)$$

where  $\mathcal{R} = \mathcal{W} - \hat{\mathcal{W}}$ , and  $\mathcal{R}_n$  and  $\mathcal{W}_n$  are the  $n$ th columns of  $\mathcal{R}$  and  $\mathcal{W}$ . Under our assumption that  $\mathcal{W}$  gives a local minimum, the derivative of this expression with respect to  $\kappa$  must be zero. Computing the derivative at  $\kappa = 0$  and setting the numerator to zero leads to

$$|\mathcal{R}_n|^2 / |\mathcal{W}_n|^2 = \|\mathcal{R}\|^2 / \|\mathcal{W}\|^2, \quad (16)$$

that is, the left-hand ratio has the same value for any  $n$ . After substituting in (15)

$$E(\mathcal{W}^\tau, (\hat{\mathcal{W}})^\tau) = \|\mathcal{R}\|^2 / \|\mathcal{W}\|^2 = E(\mathcal{W}, \hat{\mathcal{W}}) \quad (17)$$

for all values of  $\kappa$ , as required. This proves (14).

Suppose we could make the inequality in (14) strict for arbitrarily small values of  $\kappa$ . In fact, we cannot do this, since if we could the error  $\hat{E}$  would be strictly decreasing at  $\lambda$  and  $\mathcal{W}(\lambda)$  rather than having a local minimum as assumed. Therefore, for all  $\kappa$  less than some small value, we have the equality  $E(\mathcal{W}^\tau, \hat{\mathcal{W}}^\tau) = E(\mathcal{W}^\tau, (\hat{\mathcal{W}})^\tau)$ . This means that  $(\hat{\mathcal{W}})^\tau$  is a closest rank  $\leq 4$  matrix to  $\mathcal{W}^\tau$  for all sufficiently small  $\kappa$ , regardless of which column  $n$  is scaled by the transform. We will show that this can hold only if  $\mathcal{W}$  already has rank  $\leq 4$ . First, we need

**Lemma 14.** *If a matrix  $\hat{\mathcal{W}}$  is a closest matrix having rank  $\leq r$  to a matrix  $\mathcal{W}$ , then  $\mathcal{R}^\top \hat{\mathcal{W}} = \mathcal{R} \hat{\mathcal{W}}^\top = 0$ , where  $\mathcal{R} = \mathcal{W} - \hat{\mathcal{W}}$ .*

**Proof.** To prove the lemma, we write  $\hat{\mathcal{W}} = AB^\top$ , where  $A$  and  $B$  have  $r$  columns. Since  $\hat{\mathcal{W}}$  has rank  $\leq r$ , this can always be done. We consider the derivatives of  $\|\mathcal{W} - AB^\top\|^2$  with respect to the entries of  $A$ . If any such derivative is nonzero, we can adjust  $A$  to find a closer rank  $\leq r$  matrix to  $\mathcal{W}$ . Thus, all such derivatives are zero.

It may be verified that the derivative of  $\|\mathcal{W} - AB^\top\|^2$  with respect to the entry  $A_{pq}$  is equal to the inner product  $-2\mathcal{R}^p \cdot B^q$ , where  $\mathcal{R}^p$  and  $B^q$  are rows of  $\mathcal{R} = \mathcal{W} - AB^\top$  and  $B$ , respectively. Thus, all rows of  $B$  are orthogonal to the rows of  $\mathcal{R}$ . The rows of  $\hat{\mathcal{W}} = AB^\top$  are linear combinations of the rows of  $B$  and, hence, are orthogonal to any row of  $\mathcal{R}$ . Thus,  $\mathcal{R} \hat{\mathcal{W}}^\top = 0$ .

Similarly, taking derivatives with respect to the entries of  $B$  leads to  $\mathcal{R}^\top \hat{\mathcal{W}} = 0$ .

10. The nearest matrix need not be unique.

We return to the proof of Proposition 4. Since  $\hat{\mathcal{W}}$  is a closest rank  $\leq 4$  matrix to  $\mathcal{W}$ , the lemma gives  $\mathcal{R}\hat{\mathcal{W}}^\top = 0$ . As argued above, we can choose  $\kappa \neq 0$  small enough so that  $(\hat{\mathcal{W}})^\top$  is a closest rank  $\leq 4$  matrix to  $\mathcal{W}^\top$ , regardless of what  $n$  we choose for  $\tau_{nk}$ . For such  $\kappa$ , the lemma gives  $\mathcal{R}^\top(\hat{\mathcal{W}})^\top = 0$ , where  $\mathcal{R}^\top = \mathcal{W}^\top - (\hat{\mathcal{W}})^\top$ , and it follows that  $\mathcal{R}\hat{\mathcal{W}}^\top - \mathcal{R}^\top(\hat{\mathcal{W}})^\top = 0$ . Since  $\mathcal{W}$  and  $\mathcal{W}^\top$  and similarly  $\mathcal{R}$  and  $\mathcal{R}^\top$  differ only in the scaling of their  $n$ th columns, we may easily compute the matrix  $\mathcal{R}\hat{\mathcal{W}}^\top - \mathcal{R}^\top(\hat{\mathcal{W}})^\top$ : Its  $(p, q)$ th entry equals  $\kappa \mathcal{R}_n^p \mathcal{W}_n^q$ . Since  $\kappa \neq 0$  and our arguments hold regardless of  $n$ , we have  $\mathcal{R}_n^p \mathcal{W}_n^q = 0$  for all values of  $n, p$ , and  $q$ .

We assumed that  $\mathcal{W}$  has no columns consisting entirely of zeros. Each column  $n$  of  $\mathcal{W}$  contains a nonzero entry  $\mathcal{W}_n^q$ , so for each  $n$ , we must have  $\mathcal{R}_n^p = 0$  for all  $p$ , which means that column  $n$  of  $\mathcal{R}$  is zero. Hence,  $\mathcal{R} = 0$  and  $\mathcal{W}$  gives zero error, so we have proven Proposition 4.  $\square$

**Proof of Lemma 7 (SIESTA's convergence speed).** Refer to Definition 6 and the proof of Proposition 2. From (11), we get  $\hat{\Theta}^{(k-1, k-1)} - \hat{\Theta}^{(k, k)} \geq \hat{\Theta}^{(k-1, k-1)} - \hat{\Theta}^{(k, k-1)} \equiv \bar{\Delta}_k \geq 0$ . Since  $\hat{\Theta}^{(k, k)}$  converges monotonically, we have  $\lim_{k \rightarrow \infty} (\hat{\Theta}^{(k-1, k-1)} - \hat{\Theta}^{(k, k)}) = 0$ , which with the previous equation gives

$$\lim_{k \rightarrow \infty} \bar{\Delta}_k \equiv \lim_{k \rightarrow \infty} \hat{\Theta}^{(k-1, k-1)} - \hat{\Theta}^{(k, k-1)} = 0. \quad (18)$$

As remarked in Section 2,  $\{\mathcal{W}^{(k)}\}$  equals the perpendicular projection of  $\{\hat{\mathcal{W}}^{(k-1)}\}$  onto  $\mathcal{L}^{N_I N_p}$ . That is,  $\{\mathcal{W}^{(k-1)}\} = \{\mathcal{W}^{(k)}\} + \rho$ , where  $\rho$  is orthogonal to  $\mathcal{L}^{N_I N_p}$ . The data matrix  $\{\mathcal{W}^{(k-1)}\}$  from the previous iteration lies in  $\mathcal{L}^{N_I N_p}$ , so we have equality for the dot products  $\{\mathcal{W}^{(k-1)}\} \cdot \{\mathcal{W}^{(k)}\} = \{\mathcal{W}^{(k-1)}\} \cdot \{\hat{\mathcal{W}}^{(k-1)}\}$ . Thus,

$$\begin{aligned} \cos(\theta^{(k-1, k)}) &= \cos(\hat{\Theta}^{(k-1, k-1)}) \frac{\|\{\hat{\mathcal{W}}^{(k-1)}\}\|}{\|\{\mathcal{W}^{(k)}\}\|} \\ &= \frac{\cos(\hat{\Theta}^{(k-1, k-1)})}{\cos(\hat{\Theta}^{(k, k-1)})}. \end{aligned} \quad (19)$$

Equation (18) implies that the right side of (19) converges to 1, so for the left we get  $\lim_{k \rightarrow \infty} \theta^{(k, k-1)} = 0$ . That  $\lim_{k \rightarrow \infty} \theta(\lambda^{(k-1)}, \lambda^{(k)}) = 0$  follows trivially. To derive the inequality (4), we expand  $\cos(\hat{\Theta}^{(k-1, k-1)}) = \cos(\hat{\Theta}^{(k, k-1)} + \bar{\Delta}_k)$  in (19):

$$\cos(\theta^{(k-1, k)}) = \cos(\bar{\Delta}_k) - \tan(\hat{\Theta}^{(k, k-1)}) \sin(\bar{\Delta}_k). \quad (20)$$

The inequalities from (11) and (12) that we started with and the monotone convergence of  $\hat{\Theta}^{(k, k)}$  give  $\bar{\Delta}_k \leq \hat{\Theta}^{(k-1, k-1)} - \hat{\Theta}^{(k, k)} \leq \hat{\Theta}^{(K-1, K-1)} - \hat{\Theta}^{(k, k)}$  for  $k \geq K$ , implying (4).  $\square$

**Details on the nonconvergence of the Mahamud algorithm.** As for SIESTA, we consider a transform that scales the  $\lambda_n^i$  toward a trivial minimum. We define the transform so it first scales all the projective depths for the  $k$ th image by  $s$  and then scales the column of projective depths for each 3D point to maintain the norm constraints on the columns of  $\mathcal{W}$ . As before, we apply the same transform to  $\hat{\mathcal{W}}$  as to  $\mathcal{W}$ . Assuming  $\lambda$  is a stationary point of the error, our transform also has a stationary point at  $\lambda$ , and this leads to constraints on  $\mathcal{W}$ ,  $\hat{\mathcal{W}}$  analogous to (16). Exploiting these, we try to modify the transform so it strictly decreases the error at  $\lambda$ .

This is much harder than for SIESTA. The initial transform  $\tau_{nk}$  for SIESTA gave an error that was *constant* at a stationary point, so we could make the error decrease, establishing the stationary point as a saddle, by an arbitrarily small change in  $\tau_{nk}$ . For the algorithm of [10], the error at a stationary point usually has a minimum under our initial transform. We need a *large* change in the transform to make the error decrease, so this may not always be possible. However, we argue that we can make the error decrease at “desirable” stationary points, where the estimates of the structure/cameras are roughly correct and  $\lambda_n^i \approx 1$ .

We modify the initial transform as follows. Let  $\hat{\mathcal{W}} = \hat{M} \hat{X}^T$  be the factorization that comes from the SVD of  $\mathcal{W}$ . We modify the initial transform of  $\hat{\mathcal{W}}$  by transforming  $\hat{X}$  linearly before scaling it, where we choose this linear transform to minimize the error's second derivative with respect to the transform at the stationary point. We have derived two upper bounds on the resulting second derivatives.

Define  $\mathbf{w}_n^i \equiv [\mathcal{W}_n^{(3i-2):3i}]$  and  $\hat{\mathbf{w}}_n^i \equiv [\hat{\mathcal{W}}_n^{(3i-2):3i}]$  and the residual  $\mathbf{r}_n^i \equiv \mathbf{w}_n^i - \hat{\mathbf{w}}_n^i$ ; all are vectors in  $\mathfrak{R}^3$ . Without loss of generality, take the columns of  $\hat{M}$  orthogonal and define  $\hat{\mathbf{m}}^i \equiv \hat{M}^{(3i-2):3i} \in \mathfrak{R}^{3 \times 4}$ . Let  $\hat{\mu}_a^i$  be the  $a$ th singular value of  $\hat{\mathbf{m}}^i$ , and let  $\hat{\mathbf{m}}_a^i \in \mathfrak{R}^3$  be the  $a$ th column of  $\hat{\mathbf{m}}^i$ . Choose image  $k$  so

$$\langle |\hat{\mathbf{m}}^k|^2 \rangle \geq \langle |\mathbf{w}^k|^2 \rangle, \quad (21)$$

where we use  $\langle \cdot \rangle$  to denote the average, taken over the omitted index. Such an image always exists, since our normalizations give

$$\begin{aligned} 1 &= \sum_{i=1}^{N_I} \sum_{a=1}^4 |\hat{\mathbf{m}}_a^i|^2 / 4 = \sum_{i=1}^{N_I} \langle |\hat{\mathbf{m}}^i|^2 \rangle \\ &= \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} |\mathbf{w}_n^i|^2 / N_p = \sum_{i=1}^{N_I} \langle |\mathbf{w}^i|^2 \rangle. \end{aligned}$$

Our first upper bound on the second derivative for the modified transform is

$$\begin{aligned} &2 \left( \sum_{n=1}^{N_p} |\mathbf{r}_n^k|^2 \right) \\ &\left( 2 \max_{n=1 \dots N_p} \left\| |\mathbf{w}_n^k|^2 - \langle |\mathbf{w}^k|^2 \rangle \right\| - \frac{4}{3} \frac{|\hat{\mu}_3^k|^2}{\langle |\hat{\mu}^k|^2 \rangle} \langle |\mathbf{w}^k|^2 \rangle \right) \end{aligned} \quad (22)$$

for the chosen image  $k$ . In practice, Hartley and Zisserman [5] and Sturm and Triggs [14] recommend normalizing the homogenous image points to a unit box before applying ST. Then, assuming a “desirable” stationary point with all  $\lambda_n^i$  near 1, the  $|\mathbf{w}_n^k|^2$  will be approximately constant in  $k$  and  $n$ . If the singular values  $\hat{\mu}_a^k$  all have roughly the same size, then  $|\hat{\mu}_3^k|^2 / \langle |\hat{\mu}^k|^2 \rangle \approx 1$ , and our bound is likely to be negative.

In our experiments on real sequences, the apparent convergence points of the Mahamud et al. iteration [9], [10] always have  $(\hat{\mu}_3^k)^2 / \langle |\hat{\mu}^k|^2 \rangle \approx 1$ , and they almost always make the bound in (22) negative, which rules out these “convergence points” as local minima. Note that the bound is conservative; in practice, we expect cancellations to reduce the second derivative below (22).

Why is the ratio  $\langle \hat{\mu}_3^k \rangle^2 / \langle \hat{\mu}_1^k \rangle^2$  typically near 1? One contributing factor is that, after the standard scaling to a unit box, the image submatrix  $w^i \equiv [\mathbf{w}_1^i, \mathbf{w}_2^i, \dots, \mathbf{w}_{N_p}^i] \in \mathfrak{R}^{3 \times N_p}$  typically has three singular values of the same order. Another cause is the following: Write the SVD of the scaled data matrix as  $\mathcal{W} = UDV^T$ . Writing the  $\hat{\mu}_a^k$  in terms of the image data gives

$$\langle \hat{\mu}_a^k \rangle^2 = |s_a^{kT} w^k|^2 = N_I^{-1} \frac{|s_a^{kT} w^k|^2}{\langle |wV_n|^2 \rangle} = N_I^{-1} \sum_{n=1}^{N_p} \frac{|s_a^{kT} w^k V_n|^2}{\langle |wV_n|^2 \rangle},$$

where  $s_a^k \in \mathfrak{R}^3$  represent the  $a$ th left singular vector of  $\hat{m}^k$ , and  $V_n$  denotes the  $n$ th column of  $V$ . The average in the denominator is over all images  $i$ . Thus,  $\langle \hat{\mu}_a^k \rangle^2$  is proportional to a sum of projections of the (homogeneous) image data normalized by their average values. If the camera positions are spaced roughly uniformly, as they are in many sequences, the  $k$ th image is often close to ‘‘average,’’ so the singular values  $\hat{\mu}_a^k$  all have similar sizes. One can get  $\hat{\mu}_3^k \ll \hat{\mu}_1^k$  if, for example, most of the camera positions cluster together but one is very far from the others.

We have also derived a second bound whose size is easier to estimate. Choose image  $k$  such that  $\|w^k - \hat{w}^k\|^2 \leq \hat{E} \|w^k\|^2$ , which is always possible since one can show that  $\sum_i \|w^i - \hat{w}^i\|^2 = \hat{E} \sum_i \|w^i\|^2$ . Denote the  $a$ th singular value of  $w^k$  by  $d_a^k$  and the  $a$ th singular value of  $\mathcal{W}$  by  $D_a$ . Our new bound is

$$2 \left( \sum_{n=1}^{N_p} |\mathbf{r}_n^k|^2 \right) \left( 2 \max_{n=1 \dots N_p} \left| |\mathbf{w}_n^k|^2 - \langle |\mathbf{w}^k|^2 \rangle \right| - \left( d_3^k / \|w^k\| - \hat{E}^{1/2} \right)^2 \frac{N_p}{D_1^2} \langle |\mathbf{w}^k|^2 \rangle \right). \quad (23)$$

After the standard scaling of the image data to a unit box, we expect  $d_3^k / \|w^k\| \approx 3^{-1/2} \approx 0.58$ . Even if the scene is planar, the first three singular values of  $\mathcal{W}$  are usually substantial, causing  $N_p / D_1^2 \gg 1$ . Experimentally, we find  $d_3^k / \|w^k\| \approx 0.3$  and  $N_p / D_1^2 \approx 1.3$ . Substituting the experimental values, and assuming  $\mathcal{W}$  is close to a rank  $\leq 4$  matrix so  $\hat{E} \ll 1/3$ , we can approximate the bound as

$$2 \left( \sum_{n=1}^{N_p} |\mathbf{r}_n^k|^2 \right) \left( 2 \max_{n=1 \dots N_p} \left| |\mathbf{w}_n^k|^2 - \langle |\mathbf{w}^k|^2 \rangle \right| - 0.13 \langle |\mathbf{w}^k|^2 \rangle \right). \quad (24)$$

For experiments using the two bounds above see the main text.

**Proof of Proposition 9 (CIESTA’s error convergence).** We prove the Proposition for CIESTA\*, a version of CIESTA with a slightly modified Stage 3. Define

$$E_a^{(k)}(a) \equiv 1 + C_0 + a + \frac{C_3^{(k)}}{a} - 2 \left( C_0 a + 2C_1^{(k)} + \frac{C_2^{(k)}}{a} \right)^{1/2}. \quad (25)$$

- **CIESTA\* (Stage 3).** Find the positive root of the polynomial  $P^{(k-1)}(a)$  from (8) that minimizes  $E_a^{(k-1)}(a)$ . Redefine  $\lambda^{(k)} \rightarrow \lambda^{(k)} = (a + \lambda^{(k)}) b_+^{(k-1)}(a)$ .

As we show later, in Proposition 18, CIESTA’s Stage 3 outputs the same  $a$  as this new Stage 3. Thus, CIESTA\* is equivalent to CIESTA and our proofs apply to both.

From the definitions,  $\hat{E}_{\text{reg}}(\lambda^{(k)}) = E_{\text{reg}}(\lambda^{(k)}, \mathcal{W}^{(k)})$ . We will prove the Proposition by showing that CIESTA\* minimizes  $E_{\text{reg}}(\lambda, Y)$  alternately with respect to  $Y$  of rank  $\leq 4$  and  $\lambda$ . CIESTA\*s Stage 1 carries out the first minimization as for SIESTA, and Lemma 16 below shows that Stages 2 and 3 of CIESTA\* perform the second minimization, computing  $\lambda^{(k)} = \arg \min_{\lambda} E_{\text{reg}}(\lambda, \mathcal{W}^{(k-1)})$ .

First, we need to know that we do not encounter any singularities in minimizing, especially that we never have to worry about the singularity at  $\lambda = 0$ .

**Claim 15.** Let  $\mu$  satisfy Assumption 8. Let  $\hat{\mathcal{W}}^{(j)}$  denote the approximating matrix computed in the  $(j+1)$ th CIESTA\* iteration. For all  $j \geq 0$ ,

$$\|\hat{\mathcal{W}}^{(j)}\| > 0, \quad \lambda_{\min} \equiv \arg \min_{\lambda} E_{\text{reg}}(\lambda, \hat{\mathcal{W}}^{(j)}) \neq 0$$

and  $\lambda_{\min}$  is finite and a stationary point.

**Proof (sketch).** From the Assumptions’ bound on  $\mu$ , the regularization term in  $E_{\text{reg}}(\lambda, \hat{\mathcal{W}}^{(0)})$  gets bigger as  $\lambda \rightarrow 0$  than the initial value  $E_{\text{reg}}(\lambda = 1, \hat{\mathcal{W}}^{(0)})$ . Thus, the minimum of  $E_{\text{reg}}(\lambda, \hat{\mathcal{W}}^{(0)})$  cannot occur at arbitrarily small  $\lambda$ . The same reasoning holds at large  $|\lambda|$ , so the minimizing  $\lambda$  must lie within some bounded region  $\alpha \leq |\lambda| \leq \beta$ . Since  $E_{\text{reg}}$  is smooth in this region, the minimum is a stationary point.

Consider the later iterations. We will show in Proposition 9 that no CIESTA\* iteration increases the value of  $\hat{E}_{\text{reg}}$ . Thus,  $\hat{E}_{\text{reg}}(\lambda^{(j)}) \leq \hat{E}_{\text{reg}}(\lambda^{(0)})$ , that is,  $E_{\text{reg}}(\lambda^{(j)}, \hat{\mathcal{W}}^{(j)}) \leq E_{\text{reg}}(\lambda^{(0)}, \hat{\mathcal{W}}^{(0)})$ . As before,  $E_{\text{reg}}(\lambda^{(0)} = 1, \hat{\mathcal{W}}^{(0)})$  is less than the regularization contribution<sup>11</sup> to  $E_{\text{reg}}(\lambda, \hat{\mathcal{W}}^{(j)})$  as  $\lambda$  grows big or small, so the  $\lambda$  minimizing  $\hat{E}_{\text{reg}}(\lambda, \hat{\mathcal{W}}^{(j)})$  occurs at a finite stationary point away from 0. (A full proof uses induction and proves the claim and Lemma 16 together.)  $\square$

**Lemma 16.** Suppose  $\mu$  satisfies Assumption 8. Given the output  $\hat{\mathcal{W}}^{(k-1)}$  from Stage 1 of the  $k$ th CIESTA\* iteration, Stages 2 and 3 output  $\lambda^{(k)} = \arg \min_{\lambda} E_{\text{reg}}(\lambda, \hat{\mathcal{W}}^{(k-1)})$ .

**Proof.** The proof is a straightforward calculation: we compute  $\arg \min_{\lambda} E_{\text{reg}}(\lambda, \hat{\mathcal{W}}^{(k-1)})$  and then show that CIESTA gives it as output (which will be obvious). For brevity, we omit the superscript  $(k-1)$ . From Claim 15, the minimum occurs at a stationary point of  $E_{\text{reg}}$ , where

$$0 = \frac{\partial E}{\partial \lambda_n^i} = 2 |\mathbf{x}_n^i|^2 \mu(\lambda_n^i - 1) + 2 \frac{\lambda_n^i |\mathbf{x}_n^i|^2 - \mathbf{x}_n^i \cdot \hat{\mathbf{w}}_n^i}{\|\mathcal{W}(\lambda)\|^2} - 2 \lambda_n^i |\mathbf{x}_n^i|^2 \frac{\|\mathcal{W}(\lambda) - \hat{\mathcal{W}}\|^2}{\|\mathcal{W}(\lambda)\|^4}$$

for all  $i$  and  $n$ . Solving this for the projective depths at the stationary point gives

$$\lambda_n^i = \left( a + \frac{\mathbf{x}_n^i \cdot \hat{\mathbf{w}}_n^i}{|\mathbf{x}_n^i|^2} \right) b, \quad (26)$$

11. The regularization contribution does not depend on  $\mathcal{W}^{(j)}$ .

$$a = \mu \|\mathcal{W}(\lambda)\|^2, \quad (27)$$

$$b = \frac{\|\mathcal{W}(\lambda)\|^2}{\mu \|\mathcal{W}(\lambda)\|^4 + \|\mathcal{W}(\lambda)\|^2 - \|\mathcal{W}(\lambda) - \hat{\mathcal{W}}(\lambda)\|^2}.$$

$a$  and  $b$  do not depend on the point  $n$  or image  $i$ . Since  $a$  equals  $\mu \|\mathcal{W}(\lambda)\|^2$ , consistency requires

$$a = \mu \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} |\mathbf{x}_n^i|^2 |\lambda_n^i|^2 = \mu b^2 \sum_{i=1}^{N_I} \sum_{n=1}^{N_p} |\mathbf{x}_n^i|^2 \left( a + \frac{\mathbf{x}_n^i \cdot \hat{\mathbf{w}}_n^i}{|\mathbf{x}_n^i|^2} \right)^2. \quad (28)$$

Claim 15 shows that  $0 < |\lambda| < \infty$  at the minimum, so we have  $0 < a < \infty$  and  $|b| < \infty$  there. The consistency condition simplifies to

$$b^2 = \frac{a}{a^2 C_0 + 2a C_1 + C_2}. \quad (29)$$

Since  $b^2 < \infty$  at the minimum, the denominator in (29) is nonzero. The definition of  $b_+(a)$  in (7) shows that  $b_+^2(a) = b^2$ , so  $\lambda_{\min}$ , the value of  $\lambda$  at the minimum, is given by (26) with  $b = b_+(a)s$  and  $s \in \{-1, 1\}$ . Clearly,

$$\min_{\lambda} E_{\text{reg}}(\lambda, \hat{\mathcal{W}}) \leq \min_{a>0} \min_{s=\pm 1} E_{\text{reg}} \left( \lambda = \left\{ \left( a + \frac{\mathbf{x} \cdot \hat{\mathbf{w}}}{|\mathbf{x}|^2} \right) b_+(a)s \right\}, \hat{\mathcal{W}} \right)$$

since the right side minimizes with respect to a restricted class of  $\lambda$ . The form of  $\lambda$  on the right is correct for minimizing  $E_{\text{reg}}$ , so we can replace the inequality by equality. A calculation gives  $s = 1$  at the minimum and

$$E_{\text{reg}} \left( \left\{ \left( a + \frac{\mathbf{x} \cdot \hat{\mathbf{w}}}{|\mathbf{x}|^2} \right) b_+(a) \right\}, \hat{\mathcal{W}} \right) = E_{\alpha}(a). \quad (30)$$

Summarizing,  $E_{\text{reg}}(\lambda_{\min}, \hat{\mathcal{W}}) = \min_{\lambda} E_{\text{reg}}(\lambda, \hat{\mathcal{W}}) = \min_{a>0} E_{\alpha}(a) = E_{\alpha}(a_{\min})$ , where  $\lambda_{\min}$  and  $a_{\min}$  are the minimizing values. Since  $0 < a_{\min} < \infty$  and  $b$  is finite, the form of  $E_{\alpha}$  shows that it is  $C^{\infty}$  at  $a_{\min}$ , so  $a_{\min}$  is a stationary point with

$$0 = \left. \frac{dE_{\alpha}}{da} \right|_{a_{\min}} = 1 - \frac{C_3}{a_{\min}^2} - \frac{C_0 - C_2/a_{\min}^2}{(C_0 a_{\min} + 2C_1 + C_2/a_{\min})^{1/2}}.$$

After some algebra, we find that solutions to this equation are roots of the polynomial in (8).

Looking back at CIESTA\*, we see that Stages 2 and 3 implement the minimization above, outputting  $a > 0$  and  $\lambda$  that minimize  $E_{\alpha}(a)$  and  $E_{\text{reg}}(\lambda, \hat{\mathcal{W}})$ . This gives Lemma 16.  $\square$

Both Stage 1 and the combined Stage2/Stage3 reduce  $E_{\text{reg}}$  or leaves it unchanged. Proposition 9 follows easily from this by arguments analogous to those for SIESTA.  $\square$

**Proof of Proposition 10 (Limit behavior of the projective depths under CIESTA).** We need:

**Lemma 17.** For nonzero  $\lambda \in \mathbb{R}^{N_I N_p}$ , the error  $\hat{E}_{\text{reg}}(\lambda)$  is a continuous function of  $\lambda$ .<sup>12</sup>

**Proof.**  $\hat{E}_{\text{reg}}$  in (5) is the sum of two terms. The second is continuous. Since the sum of continuous functions is

12. For a short proof, note that the error is continuous in  $\lambda$  and the singular values of  $\mathcal{W}$ , which are themselves continuous in  $\mathcal{W}$  and  $\lambda$ .

continuous, we need to only show that the first term is continuous.

We use the shorthands  $\mathcal{W} \equiv \mathcal{W}(\lambda)$  and  $\hat{\mathcal{W}} \equiv \hat{\mathcal{W}}(\lambda)$ . As discussed for SIESTA, we have  $\|\mathcal{W} - \hat{\mathcal{W}}\|^2 / \|\mathcal{W}\|^2 = \sin^2 \theta(\mathcal{W}, \hat{\mathcal{W}})$ . Let  $\lambda' \in \mathbb{R}^{N_I N_p}$  be a vector very near  $\lambda$ . We also use  $\mathcal{W}' \equiv \mathcal{W}(\lambda')$  and take  $\hat{\mathcal{W}}'$  to denote the closest rank  $\leq 4$  matrix to  $\mathcal{W}'$ . Since  $\hat{\mathcal{W}}$  minimizes the angle with  $\mathcal{W}$  among rank  $\leq 4$  matrices, we have  $\theta(\mathcal{W}, \hat{\mathcal{W}}) < 90^\circ$  and, similarly,  $\theta(\mathcal{W}', \hat{\mathcal{W}}') < 90^\circ$ . The definition of  $\hat{\mathcal{W}}$  and the triangle inequality then give  $\theta(\mathcal{W}, \hat{\mathcal{W}}) \leq \theta(\mathcal{W}, \hat{\mathcal{W}}') \leq \theta(\mathcal{W}, \mathcal{W}') + \theta(\mathcal{W}', \hat{\mathcal{W}}')$ . We can achieve  $\theta(\mathcal{W}, \hat{\mathcal{W}}) \leq \epsilon + \theta(\mathcal{W}', \hat{\mathcal{W}}')$  for any  $\epsilon$  by taking  $|\lambda' - \lambda|$  small enough to give  $\theta(\mathcal{W}, \mathcal{W}') \leq \epsilon$ . The same argument in the other direction gives  $\theta(\mathcal{W}', \hat{\mathcal{W}}') \leq \epsilon + \theta(\mathcal{W}, \hat{\mathcal{W}})$ , so we can always achieve  $|\theta(\mathcal{W}', \hat{\mathcal{W}}') - \theta(\mathcal{W}, \hat{\mathcal{W}})| \leq \epsilon$  by taking  $\lambda'$  close enough to  $\lambda$ .  $\square$

Proposition 10 follows easily:

**Result 1.** The set  $\mathcal{A}$  of accumulation points cannot be empty. From Claim 15, the possible  $\lambda$  outputted by CIESTA lie within a bounded compact set, so every sequence  $\lambda^{(k)}$  has at least one accumulation point in  $\mathcal{A}$ . Pick a  $\lambda_{\mathcal{A}} \in \mathcal{A}$  and a subsequence  $\lambda^{(k_j)}$  such that  $\lambda^{(k_j)}$  converges to  $\lambda_{\mathcal{A}}$ . Then,  $\hat{E}_{\text{reg}}(\lambda_{\mathcal{A}}) = \lim_{j \rightarrow \infty} \hat{E}_{\text{reg}}(\lambda^{(k_j)}) = \hat{E}_{\infty}$ , where the first equality comes from the lemma and the second from Proposition 9 and the definition of  $\hat{E}_{\infty}$ .

**Result 2.** Suppose Result 2 is false. Then, there exists an infinite subsequence  $\lambda^{(k_j)}$  such that every member differs from every  $\lambda_{\mathcal{A}} \in \mathcal{A}$  by at least  $\epsilon$ . The sequence  $\lambda^{(k_j)}$  must have at least one accumulation point  $\lambda_{\infty}$ , which means that we can choose  $\lambda^{(k_j)}$  arbitrarily close to  $\lambda_{\infty}$ . However, the definition of  $\mathcal{A}$  implies  $\lambda_{\infty} \in \mathcal{A}$ , contradicting our assumption that the  $\lambda^{(k_j)}$  differ from all  $\lambda_{\mathcal{A}} \in \mathcal{A}$  by at least  $\epsilon$ .  $\square$

**Proposition 18.** CIESTA and CIESTA\* (defined in the proof of Proposition 9) are equivalent.

**Proof.** We present this proof partly because our proof of Proposition 11 depends on it. First, we define

$$\chi^{(k)} \equiv 4C_3^{(k)} C_0^{1/4} / (C_2^{(k)})^{3/4}, \quad f^{(k)} \equiv C_1^{(k)} / \sqrt{C_0 C_2^{(k)}}. \quad (31)$$

**Remark 19.** One can easily show<sup>13</sup> that CIESTA gives the following constraints on these constants and the ones defined earlier in (6) and (7): for some  $\beta > 0$ , we have  $C_0 > \beta$ ,  $C_3^{(k)} \geq C_2^{(k)}$ ,  $C_2^{(k)} > \beta$ ,  $z^{(k)} > \beta$ ,  $\chi^{(k)} > \beta$ ,  $|f^{(k)}| \leq 1$ .

For convenience, we omit the iteration superscript  $(k)$  from now on. We also define

**Definition 20** ( $\bar{a}$ ,  $E_{\bar{a}}$ , and  $\bar{b}_+$ ). Let  $\bar{a} \equiv a\sqrt{C_0/C_2}$ . We rewrite  $E_{\alpha}$  and  $b_+$  (see (25) and (7)) in terms of the rescaled  $\bar{a}$

13. For  $C_2^{(k)}$ : Since  $|\mathcal{W}^{(k)}| > 0$  and  $\hat{\mathcal{W}}^{(k)}$  gives its best rank  $\leq 4$  approximation,  $|\mathcal{W}^{(k)} - \hat{\mathcal{W}}^{(k)}|^2 \leq (1 - 4/N)|\mathcal{W}^{(k)}|^2$ , where  $N = \min(3N_I, N_p)$ . Then,  $(4/N)|\mathcal{W}^{(k)}|^2 \leq \text{trace}(\mathcal{W}^{(k)} \hat{\mathcal{W}}^{(k)T}) = \bar{\lambda} \cdot \bar{W} \leq |\bar{\lambda}| |\bar{W}|$ , with  $\bar{W} \equiv \{(\mathbf{x}/|\mathbf{x}|) \cdot \hat{\mathbf{w}}_n^{(k)}\} \in \mathbb{R}^{N_p N_I}$  and  $\bar{\lambda} \equiv \{\lambda^{(k)} |\mathbf{x}|\} \in \mathbb{R}^{N_I N_p}$  (for example,  $\bar{\lambda}$  has entries  $\lambda_n^{(k)} |\mathbf{x}_n^i|$ ). Also,  $C_2^{(k)} = \mu |\bar{W}|^2$  and  $|\mathcal{W}^{(k)}| = |\bar{\lambda}|$ , so  $C_2^{(k)} \geq \mu |\mathcal{W}^{(k)}|^2 16/N^2$ , where Claim 15's proof shows  $|\mathcal{W}^{(k)}|$  is bounded. The  $f^{(k)}$  constraint comes from the Schwartz inequality.

$$E_{\bar{\alpha}}(\bar{a}) \equiv 1 + C_0 + \sqrt{\frac{C_2}{C_0}} \left( \bar{a} + \frac{z}{\bar{a}} - 2 \frac{C_0^{3/4}}{C_2^{1/4}} \left( \bar{a} + \frac{1}{\bar{a}} + 2f \right)^{1/2} \right) = E_{\alpha}(a)$$

$$\bar{b}_+(\bar{a}) \equiv C_0^{-1/4} C_2^{-1/4} \left( \bar{a} + \frac{1}{\bar{a}} + 2f \right)^{1/2} = b_+(a). \quad (32)$$

To connect CIESTA\* that minimizes  $E_{\bar{\alpha}}$  and CIESTA, we must explore the properties of the global minima of  $E_{\bar{\alpha}}$ .

**Proposition 21.** *Assume  $\bar{a} > 0$  and the constant constraints from Remark 19 (valid for CIESTA/CIESTA\*). For  $z \neq 1$ , the global minimum of  $E_{\bar{\alpha}}(\bar{a})$  occurs at a stationary point at a unique  $\bar{a}$ . This  $\bar{a}$  is the only stationary point on the same side of 1 as  $z$ . When  $z = 1$ ,  $E_{\bar{\alpha}}$  has either a single global minimum at  $\bar{a} = 1$ , with no other stationary points, or two global minima at  $\bar{a}$  and  $1/\bar{a}$ , with one additional stationary point (a maximum) at  $\bar{a} = 1$ .*

**Proof.** We will assume  $f > -1$  for this proof. The theorems do not require this, but we assume it for simplicity since the case  $f = -1$  needs to be handled separately.

The proof has two aspects: 1) We must show that the minima occur at interior stationary points and not at the boundaries  $\bar{a} = 0$  or  $\infty$ . 2) We must determine the location of the global minimum and show that it is unique (unless  $z = 1$ ).

For the first, easier part, consider the derivative of  $dE_{\bar{\alpha}}/d\bar{a}$  near the boundaries. As  $\bar{a} \rightarrow 0$ , clearly,  $dE_{\bar{\alpha}}/d\bar{a} \rightarrow -(C_2/C_0)^{1/2} z \bar{a}^{-2} < 0$ . Since  $E_{\bar{\alpha}}$  is decreasing near  $\bar{a} = 0$ , no minimum or stationary point occurs at  $\bar{a} = 0$ . Similarly, it is easy to see that  $dE_{\bar{\alpha}}/d\bar{a} > 0$  as long as  $\bar{a}$  is sufficiently large. Thus, the minimum of  $E_{\bar{\alpha}}$  must occur at an internal stationary point, and there are no stationary points at the boundaries.

It will turn out that  $\bar{a} = 1$  also acts as a kind of boundary. Since our assumption  $f > -1$  implies that  $\bar{a} + 1/\bar{a} + 2f > 0$  at  $\bar{a} = 1$ , it is simple to compute

$$\left. \frac{dE_{\bar{\alpha}}}{d\bar{a}} \right|_{\bar{a}=1} = (C_2/C_0)^{1/2} (1 - z). \quad (33)$$

For  $z < 1$ ,  $E_{\bar{\alpha}}$  is increasing at  $\bar{a} = 1$ . We already know that it is decreasing near  $\bar{a} = 0$ , so it must have a stationary point local minimum at some  $\bar{a} \in (0, 1)$ . Similarly,  $z > 1$  implies that  $E_{\bar{\alpha}}$  has a local minimum at an  $\bar{a} \in (1, \infty)$ . Next, we locate the global minimum.

**The case  $z = 1$ .** For this case,  $E_{\bar{\alpha}}$  becomes a function of  $A \equiv \bar{a} + \bar{a}^{-1} \geq 2$ , where  $A = 2$  corresponds to  $\bar{a} = 1$ . The first derivative is  $dE_{\bar{\alpha}}/dA = \sqrt{C_2/C_0} (1 - C_0^{3/4} C_2^{-1/4} (A + 2f)^{-1/2})$ , and setting this to 0 gives a single stationary point at  $A^s \equiv -2f + C_0^{3/2} C_2^{-1/2}$ . One can verify that  $\text{sign}(dE_{\bar{\alpha}}/dA) = \text{sign}(A - A^s)$ . Then,

- $A^s \leq 2$  implies  $dE_{\bar{\alpha}}/dA > 0$  for all  $A \geq 2$ , so the global minimum of  $E_{\bar{\alpha}}$  occurs at the boundary  $A = 2$ , translating to a single minimum at  $\bar{a} = 1$ .
- $A^s > 2$  implies that the global minimum occurs at  $A = A^s$ . This corresponds to *two* global minima at  $\bar{a}^s$  and  $1/\bar{a}^s$ . The only other stationary point

comes from  $dA/d\bar{a}|_{\bar{a}=1} = 0$  and occurs at  $A = 2$ . This stationary point is a maximum since  $dE_{\bar{\alpha}}/dA|_{A=2} < 0$ .

The above proves Proposition 21 for  $z = 1$ .

**The case  $z > 1$ .** We treat  $E_{\bar{\alpha}}$  as a function of  $z$  and  $\bar{a}$  and write

$$E_{\bar{\alpha}}(\bar{a}, z) - E_{\bar{\alpha}}(\bar{a}, z = 1) = \sqrt{\frac{C_2}{C_0}} \frac{z - 1}{\bar{a}}. \quad (34)$$

For  $z > 1$ , the correction  $(z - 1)/\bar{a}$  in (34) decreases with  $\bar{a}$ , so  $E_{\bar{\alpha}}$  becomes relatively smaller at  $\bar{a} \geq 1$  than at  $\bar{a} \leq 1$ . We will show that this produces a global minimum with  $\bar{a} > 1$ .

Let  $\bar{a}_z^{(-)}$  denote the minimum of  $E_{\bar{\alpha}}(\bar{a}, z)$  over all  $\bar{a} \in (0, 1]$ . For  $z = 1$ , let  $a_1^{(+)} \equiv 1/a_1^{(-)}$  denote the second global minimum, and let  $\epsilon_1 \equiv \min_{\hat{a} > 0} E_{\bar{\alpha}}(\hat{a}, z = 1)$  denote the minimum value of  $E_{\bar{\alpha}}$  for  $z = 1$ . Equation (34) implies

$$E_{\bar{\alpha}}(\bar{a}_z^{(-)}, z) - E_{\bar{\alpha}}(\bar{a}_z^{(-)}, z = 1) \geq E_{\bar{\alpha}}(\bar{a}_1^{(+)}, z) - E_{\bar{\alpha}}(\bar{a}_1^{(+)}, 1) = E_{\bar{\alpha}}(\bar{a}_1^{(+)}, z) - \epsilon_1, \quad (35)$$

which gives

$$E_{\bar{\alpha}}(\bar{a}_z^{(-)}, z) \geq E_{\bar{\alpha}}(\bar{a}_z^{(-)}, 1) + E_{\bar{\alpha}}(\bar{a}_1^{(+)}, z) - \epsilon_1 \geq \epsilon_1 + E_{\bar{\alpha}}(\bar{a}_1^{(+)}, z) - \epsilon_1 = E_{\bar{\alpha}}(\bar{a}_1^{(+)}, z) \geq \min_{\hat{a} \geq 1} E_{\bar{\alpha}}(\hat{a}, z). \quad (36)$$

Suppose that the inequality in (35) is strict. Then, (36) gives

$$\min_{\bar{a} \leq 1} E_{\bar{\alpha}}(\bar{a}, z) > \min_{\bar{a} \geq 1} E_{\bar{\alpha}}(\bar{a}, z). \quad (37)$$

We know from (33) and the discussion following that  $E_{\bar{\alpha}}(\bar{a}, z > 1)$  achieves its least value over  $\bar{a} \geq 1$  at a local minimum. Equation (37) implies that this is the global minimum over all  $\bar{a} > 0$ .

Now, suppose that the inequality in (35) is actually an equality. Since  $z > 1$  implies that  $(z - 1)/\bar{a}$  in (34) decreases strictly with  $\bar{a}$ , this can happen only if  $\bar{a}_z^{(-)} = \bar{a}_1^{(+)} = 1$ . Equation (33) shows that  $E_{\bar{\alpha}}(\bar{a}, z)$  descends strictly at  $\bar{a} = 1$ , so

$$\min_{\bar{a} \leq 1} E_{\bar{\alpha}}(\bar{a}, z) \equiv E_{\bar{\alpha}}(\bar{a}_z^{(-)}, z) = E_{\bar{\alpha}}(1, z) > \min_{\bar{a} \geq 1} E_{\bar{\alpha}}(\bar{a}, z).$$

We conclude that  $E_{\bar{\alpha}}$  has its global minimum at  $\bar{a} > 1$  whether (35) is an equality or not.

**The case  $z < 1$ .** An argument similar to that above shows that the global minimum is at  $\bar{a} < 1$ .

Last, we must prove that the global minimum is unique for  $z \neq 1$ . We do this by showing that there is just a single stationary point on the side of 1 where the global minimum occurs. Our previous arguments show that this is the unique global minimum.

We consider the second derivative of  $E_{\bar{\alpha}}$ , which turns out to behave very simply. We have

$$\begin{aligned} \frac{d^2 E_{\bar{\alpha}}}{d\bar{\alpha}^2} &= \frac{(C_0 C_2)^{1/4}}{2\bar{\alpha}^3} (\chi + T(\bar{\alpha})), \\ T(\bar{\alpha}) &\equiv \bar{\alpha}^{1/2} \frac{(\bar{\alpha}^4 - 6\bar{\alpha}^2 - 8f\bar{\alpha} - 3)}{(\bar{\alpha}^2 + 1 + 2f\bar{\alpha})^{3/2}}. \end{aligned} \tag{38}$$

$d^2 E_{\bar{\alpha}}/d\bar{\alpha}^2$  has the same zeros as  $\chi + T(\bar{\alpha})$ . We study  $T$  by analyzing its stationary points at

$$0 = \frac{dT(\bar{\alpha})}{d\bar{\alpha}} = \frac{3}{2}(\bar{\alpha} - 1)(\bar{\alpha} + 1) \frac{\bar{\alpha}^4 + 4\bar{\alpha}^3 f + 6\bar{\alpha}^2 + 4f\bar{\alpha} + 1}{\sqrt{\bar{\alpha}}(\bar{\alpha}^2 + 1 + 2f\bar{\alpha})^{5/2}}. \tag{39}$$

The factor  $\bar{\alpha}^4 + 4\bar{\alpha}^3 f + 6\bar{\alpha}^2 + 4f\bar{\alpha} + 1$  has zeros at  $f = -1 - (\bar{\alpha} - 1)^4/(4\bar{\alpha}(\bar{\alpha}^2 + 1))$ , outside the allowed range for  $f$ , so this factor is positive. Similarly, the factor  $\bar{\alpha}^2 + 1 + 2f\bar{\alpha}$  is positive. Then,  $T(\bar{\alpha} > 0)$  has a unique stationary point at  $\bar{\alpha} = 1$ . Also, (39) implies that  $\text{sign}(dT/d\bar{\alpha}) = \text{sign}(\bar{\alpha} - 1) < 0$ , so  $T(\bar{\alpha})$  achieves its global minimum at  $\bar{\alpha} = 1$ . A calculation gives  $T(1) = \min_{\bar{\alpha}} T(\bar{\alpha}) = -2^{3/2}/(1 + f)^{1/2} < 0$ . Fig. 1b plots  $T$  for  $f = 0.3$ .

Suppose  $z > 1$ . Then, we know that the global minimum occurs at  $\bar{\alpha} > 1$ , so there is at least one stationary point in this region. From (33), we know that  $dE_{\bar{\alpha}}/d\bar{\alpha} < 0$  at  $\bar{\alpha} = 1$ . At the stationary point with smallest  $\bar{\alpha} > 1$ , the derivative has increased to 0, which means that the second derivative must have become positive somewhere in the interval  $(1, \bar{\alpha})$ . Our results above on  $T$  show that the second derivative remains positive at all  $\bar{\alpha}$  larger than the one where it crosses 0. Thus,  $dE_{\bar{\alpha}}/d\bar{\alpha} > 0$  at all  $\bar{\alpha}' > \bar{\alpha}$ , and  $\bar{\alpha}$  gives the unique stationary point with  $\bar{\alpha} > 1$ .

A similar argument works for  $z < 1$ . This concludes the proof of Proposition 21. As noted earlier, the proof extends to the special case  $f = -1$ .  $\square$

Finally, we use Proposition 21 to show the equivalence of CIESTA and CIESTA\*. A calculation of the derivative  $dE_{\bar{\alpha}}/d\bar{\alpha}$  shows that  $\text{sign}(1 - \bar{\alpha}^{-2}) = \text{sign}(1 - z\bar{\alpha}^{-2})$  at a stationary point. For  $z > 1$ , the global minimum stationary point at  $\bar{\alpha} > 1$  then satisfies  $\bar{\alpha} > z^{1/2} > 1$ . For  $z < 1$ , the global minimum satisfies  $\bar{\alpha} < z^{1/2} < 1$ . These inequalities translate into the conditions given in CIESTA. We derived the polynomial  $P(a)$  in (8) by squaring the equation for the stationary point, and one can verify that any zero of  $P(a)$  not corresponding to a stationary point fails the sign check above. Thus, the procedure given in CIESTA does select the root of  $P(a)$  giving the global minimum of  $E_{\bar{\alpha}}$  or  $E_{\alpha}$ , so CIESTA is equivalent to CIESTA\*. This proves Proposition 18.  $\square$

**Proof of Proposition 11 (CIESTA converges to fixed points).** We need:

**Lemma 22.** *Under the assumptions of Proposition 11, the  $\lambda$  output by Stage 3 of CIESTA depends continuously on  $\hat{W}(\lambda)$  and  $\lambda$  over a neighborhood of  $\lambda_A$ .*

**Proof.** Since the fourth singular value is strictly greater than the fifth, the properties of the SVD imply that  $\hat{W}(\lambda)$  is continuous in  $\lambda$  at  $\lambda_A$ . The constants  $C_m$  from (6) are continuous in  $\hat{W}(\lambda)$  and, thus, in  $\lambda$ . Remark 19 states that  $C_0, C_2$ , and  $C_3$  have  $\beta > 0$  as a lower bound, so continuity

implies that they are strictly positive at  $\lambda_A$ . Then,  $z$  is finite and positive and continuous in  $\lambda$  at  $\lambda_A$ . Since  $z \neq 1$  at  $\lambda_A$  by assumption, we have  $z \neq 1$  over a neighborhood of  $\lambda_A$ , and Proposition 21 implies that  $E_{\bar{\alpha}}$  has a unique global minimum at  $\bar{\alpha} \neq 1$  everywhere in this neighborhood. Let  $\bar{\alpha}_{GM}$  denote the  $\bar{\alpha}$  giving the global minimum. Consider the expression (32) for  $E_{\bar{\alpha}}$ . Since  $\bar{\alpha}_{GM} \neq 1$  and  $|f| \leq 1$ , see Remark 19, the square root in (32) has  $\bar{\alpha}_{GM} + \bar{\alpha}_{GM}^{-1} + 2f > 0$ . Then, the form of  $E_{\bar{\alpha}}$  shows that it is  $C^\infty$  at  $\lambda_A$  with respect to  $\bar{\alpha}$ ,  $f$ , and  $C_2$ .

Proposition 21 and its proof show that the global minimum of  $E_{\bar{\alpha}}$  occurs at a local minimum stationary point with  $d^2 E_{\bar{\alpha}}/d\bar{\alpha}^2 > 0$ . Let  $\bar{\alpha}_{LM}$  give the local minimum. Then,  $\bar{\alpha}_{LM}$  is a differentiable (thus, continuous) function of  $f$  and  $C_2$  over a neighborhood of their values at  $\lambda_A$ . In fact,

$$\frac{\partial \bar{\alpha}_{LM}}{\partial C_2} = - \left( \frac{\partial^2 E_{\bar{\alpha}}}{\partial \bar{\alpha}^2} \right)^{-1} \frac{\partial^2 E_{\bar{\alpha}}}{\partial \bar{\alpha} \partial C_2}$$

with a similar equation for  $f$ . The derivatives exist over a neighborhood of  $\lambda_A$  since  $d^2 E_{\bar{\alpha}}/d\bar{\alpha}^2 > 0$  and  $E_{\bar{\alpha}}$  is  $C^\infty$  in  $\bar{\alpha}$ ,  $f$ , and  $C_2$ . Since the global minimum of  $E_{\bar{\alpha}}$  occurs at a unique  $\bar{\alpha}$ , it is easy to show that it coincides with  $\bar{\alpha}_{LM}$  over a neighborhood of the values of  $f$  and  $C_2$  at  $\lambda_A$  (the continuity of  $E_{\bar{\alpha}}$  implies that competing local minima continue to have larger errors than  $E_{\bar{\alpha}}(\bar{\alpha}_{LM})$  over some neighborhood). Thus,  $\bar{\alpha}_{GM} = \bar{\alpha}_{LM}$  over this neighborhood, and  $\bar{\alpha}_{GM}$  depends continuously on  $f$ ,  $C_2$  and, therefore, on  $\lambda$  over a neighborhood of  $\lambda_A$ .

Since  $\bar{\alpha}_{GM}$  and  $\hat{W}(\lambda)$  are continuous in  $\lambda$ , and the Stage-3 output of CIESTA is clearly continuous in  $\bar{\alpha}_{GM}$  and  $\hat{W}(\lambda)$ , Lemma 22 follows.  $\square$

We return to the proof of Proposition 11. Let  $\text{CIESTA}(\lambda) \in \mathfrak{R}^{N_t N_p}$  denote the projective-depth vector obtained by applying one CIESTA iteration to  $\lambda$ . Choose a subsequence  $\lambda^{(k_j)}$  that converges to  $\lambda_A$ . Since CIESTA is continuous at  $\lambda_A$ , the sequence  $\text{CIESTA}(\lambda^{(k_j)}) \equiv \lambda^{(k_j+1)}$  converges to  $\text{CIESTA}(\lambda_A)$ . Hence, both  $\lambda_A$  and  $\text{CIESTA}(\lambda_A)$  are in  $\mathcal{A}$ , so Proposition 10 implies that they have the same value of the error:  $\hat{E}_{\text{reg}}(\lambda_A) = \hat{E}_{\text{reg}}(\text{CIESTA}(\lambda_A)) = \hat{E}_\infty$ . Let Stage (2/3) denote the combination of Stages 2 and 3. As usual, Stage 1 of CIESTA minimizes the error  $E_{\text{reg}}(\lambda, Y)$  with respect to  $Y$  of rank  $\leq 4$ . Lemma 16 shows that Stage (2/3) minimizes it with respect to  $\lambda$  and outputs the minimizing  $\lambda$ . Since CIESTA leaves the error unchanged at  $\lambda_A$ , both Stage 1 and Stage (2/3) leave it unchanged at  $\lambda_A$ . For Stage (2/3), this gives  $E_{\text{reg}}(\lambda_A, \hat{W}(\lambda_A)) = \min_{\lambda} E_{\text{reg}}(\lambda, \hat{W}(\lambda_A))$ . Proposition 21 shows that  $\arg \min_{\lambda} E_{\text{reg}}(\lambda, \hat{W}(\lambda_A))$  is unique, so  $\lambda_A$  gives the minimum and is the output of CIESTA—it is a fixed point as claimed.

Last, we show that  $\hat{E}_{\text{reg}}$  has a stationary point at  $\lambda_A$ . Our discussion above shows that  $E_{\text{reg}}(\lambda = \lambda_A, Y = \hat{W}(\lambda_A))$  has minima separately with respect to  $\lambda$  and  $Y$ , where  $Y$  is a matrix of rank  $\leq 4$ . Thus,  $E_{\text{reg}}$  has a stationary point at  $(\lambda_A, \hat{W}(\lambda_A))$ , with  $\partial E_{\text{reg}}/\partial \lambda|_{\lambda_A} = 0$  and  $\partial E_{\text{reg}}/\partial Y_M|_{\hat{W}(\lambda_A)} = 0$ , where  $Y_M$  represents the coordinates on the manifold of matrices with rank  $\leq 4$ . Then,  $\partial \hat{E}_{\text{reg}}/\partial \lambda|_{\lambda_A}$

$$= \frac{\partial E_{\text{reg}}(\lambda, \hat{W}(\lambda))}{\partial \lambda} \Big|_{\lambda_A} = \frac{\partial E_{\text{reg}}(\lambda, \hat{W}(\lambda_A))}{\partial \lambda} \Big|_{\lambda_A} + \lim_{\varepsilon \rightarrow 0} \frac{E_{\text{reg}}(\lambda_A, \hat{W}(\lambda_A + \varepsilon)) - E_{\text{reg}}(\lambda_A, \hat{W}(\lambda_A))}{\varepsilon}.$$

The first term on the right vanishes since  $E_{\text{reg}}$  is stationary in  $\lambda$ , and the second vanishes since it is stationary in  $Y$  and since  $\hat{W}$  is continuous in  $\lambda$  at  $\lambda_A$ . This proves Proposition 11.  $\square$

#### Proof of Proposition 12 (CIESTA's unique convergence).

Recall that  $\lambda^{(k)} \in \mathfrak{R}^{N_i N_p}$  gives the output of the  $k$ th CIESTA iteration,  $\mathcal{A}$  is the set of accumulation points for the sequence  $\lambda^{(k)}$ , and  $\hat{E}_\infty$  gives the greatest lower bound of the errors  $\hat{E}(\lambda^{(k)})$ . Also, recall that  $\hat{E}_\infty = \hat{E}_{\text{reg}}(\lambda_A)$  for any accumulation point  $\lambda_A \in \mathcal{A}$ .

Seeking a contradiction, we assume that CIESTA has at least two accumulation points  $\lambda_A$  and  $\lambda_B$ . By assumption, the error  $\hat{E}_{\text{reg}}$  has a *strict* local minimum at  $\lambda_A$ , so there is a neighborhood around  $\lambda_A$  for which  $\hat{E}_{\text{reg}}(\lambda \neq \lambda_A) > \hat{E}_\infty$ , and it follows from result 1 of Proposition 10 that this neighborhood contains no accumulation points except for  $\lambda_A$ . This means that the distance between  $\lambda_A$  and any other accumulation point has the greatest lower bound  $\delta > 0$ . Since the limit of accumulation points is also an accumulation point, there exists an accumulation point that is closest to  $\lambda_A$  and whose distance to  $\lambda_A$  equals  $\delta$  exactly (there may be more than one). Without loss of generality, we choose  $\lambda_B$  to be a closest accumulation point to  $\lambda_A$ , with  $|\lambda_A - \lambda_B| = \delta$ .

We choose a decreasing sequence of numbers  $\epsilon_j$  with  $0 < \epsilon_j < \delta/2$  and a subsequence  $\lambda^{(k_j)}$  of the  $\lambda^{(k)}$  with the following properties:

1. The  $\epsilon_j$  converge monotonically to 0.
2. The subsequence converges to  $\lambda_A$ , with  $|\lambda^{(k_j)} - \lambda_A| < \epsilon_j$ .
3. For all iterations  $\mu$  with  $\mu \geq k_j$ , CIESTA's output  $\lambda^{(\mu)}$  falls within a distance  $\epsilon_j$  of some accumulation point, that is,  $\exists \lambda_\alpha \in \mathcal{A} : |\lambda^{(\mu)} - \lambda_\alpha| < \epsilon_j$ .
4. CIESTA applied to  $\lambda^{(k_j)}$  gives  $\lambda^{(k_{j+1})}$  with  $|\lambda_A - \lambda^{(k_{j+1})}| > \delta/2$ .

Since  $\lambda_A$  is an accumulation point of the  $\lambda^{(k)}$ , clearly, we can enforce Property 2. We can ensure Property 3 because of Proposition 10, which states that, for any  $\epsilon$ , we can make  $\lambda^{(k)}$  fall within  $\epsilon$  of  $\mathcal{A}$  by taking  $k$  larger than some value. To achieve Property 4, we choose each  $\lambda^{(k_j)}$  so that applying CIESTA gives a  $\lambda^{(k_{j+1})}$  with a closest accumulation point  $\lambda_\alpha^{(j)} \neq \lambda_A$ . (If we were unable to do this, the  $\lambda^{(k)}$  would not accumulate at any point besides  $\lambda_A$ , contrary to our assumption that  $\mathcal{A}$  contains at least two points.) For any  $\lambda$ , the triangle inequality gives  $|\lambda_A - \lambda^{(k_{j+1})}| \geq |\lambda_A - \lambda| - |\lambda^{(k_{j+1})} - \lambda|$ , so we have

$$|\lambda_A - \lambda^{(k_{j+1})}| \geq |\lambda_A - \lambda_\alpha^{(j)}| - |\lambda^{(k_{j+1})} - \lambda_\alpha^{(j)}| > \delta - \epsilon_j > \delta/2,$$

where in the last line we used Property 3 and the assumed upper bound on  $\epsilon_j$ . This shows that we can achieve Property 4.

Because our assumptions are the same as for Proposition 11, Lemma 22 in the proof of Proposition 11 dictates that CIESTA's output  $\lambda$  depends continuously on the input  $\lambda$  over a neighborhood of  $\lambda_A$ . Then,

$$\lim_{j \rightarrow \infty} \lambda^{(k_{j+1})} \equiv \lim_{j \rightarrow \infty} \text{CIESTA}(\lambda^{(k_j)}) = \text{CIESTA}(\lambda_A) = \lambda_A$$

by Property 2, the continuity of CIESTA, and the result of Proposition 11 that  $\lambda_A$  is a fixed point. However, this implies  $\lim_{j \rightarrow \infty} |\lambda_A - \lambda^{(k_{j+1})}| = 0$ , which contradicts Property 4 as desired.

A similar argument establishes the Corollary.  $\square$

## ACKNOWLEDGMENTS

NICTA is a research centre funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council, through Backing Australia's Ability and the ICT Research Centre of Excellence programs.

## REFERENCES

- [1] R. Berthilsson, A. Heyden, and G. Sparr, "Recursive Structure and Motion from Image Sequences Using Shape and Depth Spaces," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 444-449, 1997.
- [2] A. Buchanan and A. Fitzgibbon, "Damped Newton Algorithms for Matrix Factorization with Missing Data," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. II, pp. 316-322, 2005.
- [3] A.P. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc.*, vol. 39, no. B, pp. 1-38, 1977.
- [4] R. Dutta, R. Manmatha, L.R. Williams, and E.M. Riseman, "A Data Set for Quantitative Motion Analysis," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 159-164, 1989.
- [5] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2000.
- [6] A. Heyden, R. Berthilsson, and G. Sparr, "An Iterative Factorization Method for Projective Structure and Motion from Image Sequences," *Image and Vision Computing*, vol. 17, pp. 981-991, 1999.
- [7] B.K.P. Horn and E.J. Weldon, Jr., "Direct Methods for Recovering Motion," *Int'l J. Computer Vision*, vol. 2, no. 1, pp. 51-76, 1988.
- [8] R. Kumar and A.R. Hanson, "Sensitivity of the Pose Refinement Problem to Accurate Estimation of Camera Parameters," *Proc. Int'l Conf. Computer Vision*, pp. 365-369, 1990.
- [9] S. Mahamud, M. Hebert, Y. Omori, and J. Ponce, "Provably-Convergent Iterative Methods for Projective Structure from Motion," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. I, pp. 1018-1025, 2001.
- [10] S. Mahamud and M. Hebert, "Iterative Projective Reconstruction from Multiple Views," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. II, pp. 430-437, 2000.
- [11] J. Oliensis and R. Hartley, "Iterative Extensions of the Sturm/Triggs Algorithm: Convergence and Nonconvergence," *Proc. European Conf. Computer Vision*, vol. IV, pp. 214-227, 2006.
- [12] J. Oliensis, "Fast and Accurate Self-Calibration," *Proc. Int'l Conf. Computer Vision*, pp. 745-752, 1999.
- [13] J. Oliensis and Y. Genc, "Fast and Accurate Algorithms for Projective Multi-Image Structure from Motion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 546-559, June 2001.
- [14] P. Sturm and B. Triggs, "A Factorization Based Algorithm for Multi-Image Projective Structure and Motion," *Proc. European Conf. Computer Vision*, vol. II, pp. 709-720, 1996.
- [15] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method," *Int'l J. Computer Vision*, vol. 9, pp. 137-154, 1992.
- [16] B. Triggs, "Factorization Methods for Projective Structure and Motion," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 845-851, 1996.

- [17] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle Adjustment—A Modern Synthesis," *Proc. Workshop Vision Algorithms: Theory and Practice*, pp. 298-372, 1999.
- [18] Z. Zhang, "A Flexible New Technique for Camera Calibration," Microsoft Technical Report MSR-TR-98-71, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, Nov. 2000.



**John Oliensis** received the PhD degree in theoretical particles physics from the University of Chicago and carried out research in physics at Princeton University, the Fermi National Accelerator Laboratory, and the Argonne National Laboratory. He began research in computer vision in 1988, joining the University of Massachusetts at Amherst as a member of the research faculty. From 1994 to 2003, he was a research scientist at the NEC Research Insti-

tute, where he organized three workshops bringing together researchers in computer vision, human vision, neuroscience, and learning. Since 2003, he has been an associate professor in the computer science department at Stevens Institute of Technology. His research interests include the estimation of object shape from images, perceptual organization, the recognition of objects, and human vision. He is a senior member of the IEEE and IEEE Computer Society and an associate editor of *IEEE Transactions on Pattern Analysis and Machine Intelligence*.



**Richard Hartley** received the degree from the University of Toronto in 1976 with a thesis in knot theory and worked in this area for several years before joining the General Electric (GE) Research and Development Center, where he developed Computer-Aided Electronic Design tools and created a very successful design system called the Parsifal Silicon Compiler, described in his book *Digit Serial Computation*.

He works in Computer Vision at the Australian National University and in the National ICT Australia, a government-funded research institute. In 1991, he was awarded GE's Dushman Award for this work. Around 1990, he developed an interest in Computer Vision, and in 2000, he coauthored (with Andrew Zisserman) a book on multiple view geometry. He has written papers in knot theory, geometric voting theory, computational geometry, computer-aided design, and computer vision and holds 32 US patents. He is a senior member of the IEEE and IEEE Computer Society.

▷ **For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).**