Optimal Motion From Image Sequences: A Riemannian Viewpoint

Yi Ma Jana Košecká Shankar Sastry
Electronics Research Laboratory
University of California at Berkeley
Berkeley, CA 94720-1774
{mayi,janka,sastry}@robotics.eecs.berkeley.edu

Keywords: epipolar constraint, essential manifold, optical flow, Stiefel manifolds, Grassmann manifolds, Newton algorithm.

Abstract

Motion recovery from image correspondences is typically a problem of optimizing an objective function associated with the epipolar (or Longuet-Higgins) constraint. This objective function is defined on the so called essential manifold, which has a nice intrinsic Riemannian structure. Based on existing optimization techniques on Riemannian manifolds, in particular on Stiefel or Grassmann manifolds, we propose a Riemannian Newton algorithm to solve the motion recovery problem, making use of the natural geometric structure of the essential manifold. The same ideas also apply to conjugate gradient based algorithms. The proposed geometric algorithms have quadratic rates of convergence.

1 Introduction

The problem of recovering structure and motion from a sequence of images has been one of the central problems in computer vision over the past decade and has been studied extensively from various perspectives. The seminal work of Longuet-Higgins [6] on characterization of so called *epipolar constraint*, enabled decoupling of the structure and motion problems and led to the development of numerous linear and nonlinear algorithms for motion estimation (see [10, 13] for overviews). The epipolar constraint has been formulated both in discrete and differential setting and the recent work of the authors [7] demonstrated the possibility of the parallel development of linear algorithms for both cases: that of point feature measurements and optical flow.

The appeal of linear algorithms (for discrete case [10] and for differential case [7]) is the closed form solution to the problem which in the absence of noise provides true solution of the motion. Further analysis of linear techniques reveals inherent bias in the translation estimate [4]. The performance of the numerical optimization techniques which use nonlinear objective function has been shown superior to the linear ones.

Different objective function has been proposed by Horn [3], where instead of minimizing directly the essential (coplanar) constraint the objective function expresses the true errors in relative orientation (i.e. translation and rotation). Horn proposed an iterative procedure where the update of the estimate takes into account the orthonormal constraint of the unknown rotation. Horn's algorithm and the algorithm proposed by [12] are some of the few which consider explicitly the differential geometric properties of $SO(3)$. However they do not make a connection and full exploitation of the differential geometric properties of the entire space of essential matrices as characterized in our recent paper [7].

Even from the efforts which rightly concentrated on the noise related aspects of the problem (i.e. trying different nonlinear objective functions etc.) none of the existing algorithms truly took advantage of the underlying geometrical structure of the problem. The underlying search space was usually parameterized for computational convenience [3, 12, 10] instead of being consistent to its intrinsic geometric structure. Consequently, in these algorithms, solving for optimal updating direction typically involved using Lagrangian multipliers to deal with the constraints on the search space; and "walking" on such space was done approximately by the *update-then-project* procedure.

Due to recent developments in optimization techniques on Riemannian manifolds [11, 2], in this paper we give a top level mathematical view for the nonlinear optimization problem associated with the motion recovery and using Newton's method as an example, show how to apply the optimization theory on Riemannian manifolds to solve this problem by using the intrinsic Riemannian structure of the underlying search space.

This paper relies on familiarity with Edelman *et al*'s work [2] and some background of modern Riemannian geometry (a good reference for Riemannian geometry is Boothby [1]). Section 2 shows how to generalize optimization schemes on single Riemannian manifold to their product space. Section 3 then studies the intrinsic Riemannian structure of the essential manifold (the space of all essential matrices). Section 5 gives the (Riemannian) Newton algorithm for optimizing the least square objective function associated with motion recovery. More details are given in the full version of this paper Ma, Košecká and Sastry [9].

2 Optimization on Riemannian Manifold Preparation

Newton and conjugate gradient methods are classical nonlinear optimization techniques used to minimize a function $f(x)$, where x belongs to an open subset of Euclidean space $\mathbb{R}^n$. Recent developments in optimization algorithms on Riemannian manifolds have provided geometric insights for generalization of Newton and conjugate gradient methods to certain classes of Riemannian manifolds. Smith [11] gave a detailed treatment of the general theory of optimization on Riemannian manifolds; Edelman, Arias and Smith [2] then studied the case where the Riemannian manifolds are Stiefel and Grassmann manifolds, and presented a uniform geometric framework for Newton and conjugate gradient algorithms on Stiefel and Grassmann manifolds.

These new mathematical results solve the more general optimization problem of minimizing the function $f(x)$, where x is constrained to a Riemannian manifold (M, g) which is, in turn, usually given as a submanifold in a Euclidean Space $\mathbb{R}^n$. Previous algorithms for solving such problems were application dependent: the performance of proposed algorithms relied on particular parameterization chosen for the submanifold M and also depended on certain approximation schemes applied to update the states on the search space. However, the new results of [2] show that, on Stiefel or Grassmann manifolds, one may make use of the canonical Riemannian structure of these manifolds and systematically develop Newton and conjugate gradient methods on them. Since the parameterization and metrics are canonical and the state is updated along geodesics, the performance of so obtained algorithms is guaranteed: they typically have polynomial complexity and quadratic (super-linear) rate of convergence [11].

The purpose of this paper is to apply these new Riemannian optimization schemes to solve the computer vision problem of recovering 3D motion from image correspondences. As we will soon see the underlying Riemannian manifold for this vision problem (the so called essential manifold) is a product of Stiefel manifolds. The following theorem guarantees that Edelman et al's methods can be easily generalized to any product of Stiefel (or Grassmann) manifolds (for a detailed discussion of the Riemannian structure on a product space, one may refer to Ma, Košecká and Sastry [9]).

Theorem 1 *Consider $M = M_1 \times M_2$ the product Riemannian manifold of M_1 and M_2. Then for two vector fields $X = (X_1, X_2), Y = (Y_1, Y_2) \in \mathcal{X}(M)$, Y is parallel along X if and only if Y_1 is parallel along X_1 and Y_2 is parallel along X_2.*

As a corollary to this theorem, the geodesics in the product manifold are just the products of geodesics in the two factor manifolds. Consequently, the calculation of parallel transport and geodesics in the product space can be reduced to those for each factor manifold.

3 Riemannian Structure of the Essential Manifold

In this section we study the intrinsic Riemannian structure of the essential manifold, which plays an important role in the motion recovery from image correspondences (for details see [8]). For a vector $S = (s_1, s_2, s_3)^T \in \mathbb{R}^3$, the notation $\widehat{S}$ means the associated skew-symmetric matrix:

$$\widehat{S} = \begin{pmatrix} 0 & -s_3 & s_2 \\ s_3 & 0 & -s_1 \\ -s_2 & s_1 & 0 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

The *normalized essential manifold* is defined to be:

$$\mathcal{E}_1 = \{R\widehat{S} \mid R \in SO(3), \widehat{S} \in so(3), \frac{1}{2}tr(\widehat{S}^T\widehat{S}) = 1\}.$$

$SO(3)$ is the Lie group of 3×3 rotation matrices (special orthogonal matrices with determinant $+1$), and $so(3)$ is the Lie algebra of $SO(3)$, *i.e.* the tangent plane of $SO(3)$ at the identity. $so(3)$ then consists of all 3×3 skew-symmetric matrices. Note that $\frac{1}{2}tr(\widehat{S}^T\widehat{S}) = S^TS$. A matrix E with the form $E = R\widehat{S}, R \in SO(3), S \in S^2$ (where S^2 is the unit sphere) is then a point in this manifold. Such a matrix E is called an *normalized essential matrix.* In order to understand the Riemannian structure of the normalized essential manifold, we start with the Riemannian structure on the tangent bundle of the Lie group $SO(3)$, *i.e.* $T(SO(3))$.

The tangent space of $SO(3)$ at the identity is simply its Lie algebra $so(3)$. Since $SO(3)$ is a compact Lie group, it has an intrinsic bi-invariant metric [1] (such metric is unique up to a constant scale). In the matrix form, this metric is given explicitly as:

$$g_0(\widehat{S}_1, \widehat{S}_2) = \frac{1}{2}tr(\widehat{S}_1^T\widehat{S}_2), \quad \widehat{S}_1, \widehat{S}_2 \in so(3).$$

Theorem 2 *The unit tangent bundle $T_1(SO(3))$ is equivalent to $SO(3)\times S^2$. It has a Riemannian metric g induced from the intrinsic metric g_0 on $SO(3)$ which is the same as that induced from the Euclidean metric with $T_1(SO(3))$ naturally embedded in $\mathbb{R}^{3\times 4}$ and this metric is left-invariant under the group $SO(3)$. Further, $(T_1(SO(3)), g)$ is the product Riemannian manifold of $(SO(3), g_1)$ and (S^2, g_2) with g_1 and g_2 being canonical quotient metrics for $SO(3)$ and S^2 as Stiefel manifolds.*

The proof of this theorem is referred to the full version of the paper Ma, Košcká and Sastry [9].

However, the unit tangent bundle $T_1(SO(3))$ is not exactly the normalized essential manifold $\mathcal{E}_1$. Indeed, $T_1(SO(3))$ is a double covering of the normalized essential space $\mathcal{E}_1$, *i.e.* $\mathcal{E}_1 = T_1(SO(3))/\mathbb{Z}^2$ (for details see [8]). The natural covering map from $T_1(SO(3))$ to $\mathcal{E}_1$ is:

$$\begin{aligned} h : T_1(SO(3)) &\rightarrow \mathcal{E}_1 \\ (R, R\widehat{S}) \in T_1(SO(3)) &\mapsto R\widehat{S} \in \mathcal{E}_1. \end{aligned}$$

The inverse of this map is given by:

$$h^{-1}(R\widehat{S}) = \left\{(R, R\widehat{S}),\ (R\exp(-\widehat{S}\pi), R\widehat{S})\right\}.$$

For an intrinsic geometric interpretation of this double covering, one may refer to the full paper [9].

If we take for $\mathcal{E}_1$ the Riemannian structure induced from the covering map h, the original optimization problem of optimizing a function $F(E)$ on $\mathcal{E}_1$ is equivalent to optimizing $F(R, S)$ on $T_1(SO(3))$. Due to Theorem 2, we can simply choose the induced Euclidean metric on $T_1(SO(3))$ and explicitly give out the Riemannian algorithms in terms of the matrix representation of $T_1(SO(3))$. Since this Euclidean metric is the same as the intrinsic metrics, the apparently extrinsic representation preserves all the intrinsic geometric properties of the given problems. In this sense, such algorithms will be different from other existing algorithms which make use of particular parameterizations of the underlying search space, such as quaternions [3, 12, 5, 10].

One may refer to Ma, Košcká and Sastry [9] for a detailed study of the Riemannian Newton algorithm for optimizing a general objective function on the normalized essential manifold. In the following, we will give as an example the algorithm for motion recovery by minimizing a natural objective function associated with the epipolar constraint.

4 Optimal Motion Recovery

From computer vision, we know that two corresponding image points $p, q \in \mathbb{R}^3$ satisfy the so called epipolar (or Longuet-Higgins) constraint:

$$p^T R\widehat{S}q = 0$$

where $R \in SO(3)$ and $S \in S^2$ are relative rotation and translation between the two image frames.[1] Thus to recover the motion R, S from a given set of image correspondences $p_i, q_i \in \mathbb{R}^3, i = 1, \ldots, N$, it is natural to optimize the following objective function:

$$F(R, S) = \sum_{i=1}^{N}(p_i^T R\widehat{S}q_i)^2, \quad (R, S) \in T_1(SO(3)).$$

The following theorem gives a sufficient condition such that the Hessian associated with this objective function is non-degenerate in a neighborhood of the optimal solution. Then the Riemannian Newton algorithm will have quadratic rate of convergence according to Theorem 3.4 of Smith [11].

Theorem 3 *Consider the objective function $F(R, S)$ as above. Its Hessian is non-degenerate in a neighborhood of the optimal solution if there is a unique (up to a scale) solution to the system of linear equations:*

$$p_i^T E q_i = 0, \quad E \in \mathbb{R}^{3\times 3}, \quad i = 1, \ldots, N.$$

If so, the Riemannian Newton algorithm has quadratic rate of convergence.

Please refer to the full version of the paper [9] for the proof. It is not just a coincidence that the conditions for the Hessian to be non-degenerate are exactly the same as that for the eight-point linear algorithm (see [10, 8]) to have a unique solution. A heuristic explanation is that the objective function here is a quadratic form of the epipolar constraint which the linear algorithm is directly based on.

Now assume that the Hessian is non-degenerate. We need to solve for the optimal updating vector $\Delta \in \mathbb{R}^{3\times 4}$. Pick five linearly independent tangent vectors of $T_1(SO(3))$, *i.e.* a basis of $T_R(SO(3)) \times T_S(S^2)$: $E^j, j = 1, \ldots, 5$. One obtains five linear equations:

$$\text{Hess } F(E^j, \Delta) = -dF(E^j), \quad j = 1, \ldots, 5.$$

In general, these five linear equations uniquely determine Δ. In particular, one can choose the simplest basis such that for $j = 1, 2, 3$: $E^j = (R\widehat{e}_j, 0)$ with $e_j, j = 1, 2, 3$ the standard basis for $\mathbb{R}^3$, and for $j = 4, 5$: $E^j = (0, e_j)$ such that $\{S, e_4, e_5\}$ form an orthonormal basis for $\mathbb{R}^3$. The vectors e_4, e_5 can be obtained using Gram-Schmidt process.

Define the 5×5 matrix $A \in \mathbb{R}^{5\times 5}$ and the 5 dimensional vector $\mathbf{b} \in \mathbb{R}^5$ to be:

$$(A)_{jk} = \text{Hess } F(E^j, E^k), \quad (\mathbf{b})_j = -dF(E^j),$$

for $1 \le j, k \le 5$. Then solve for the vector $\mathbf{a} \in \mathbb{R}^5$ to be $\mathbf{a} = A^{-1}\mathbf{b}$. Let $\omega = (a_1, a_2, a_3)^T \in \mathbb{R}^3$ and $v = a_4 e_4 + a_5 e_5 \in \mathbb{R}^3$. Then for the optimal updating vector $\Delta = (\Delta_1, \Delta_2) \in \mathbb{R}^{3\times 4}$, we have $\Delta_1 = R\widehat{\omega}$ and $\Delta_2 = v$. We now summarize the Riemannian Newton algorithm for the optimal motion recovery, which can be directly implemented.

The Newton Algorithm for Optimal Motion Recovery from the Objective Function $F(R, S)$

1. At the point $(R, S) \in SO(3) \times S^2$,

a) compute the gradient $G \in \mathbb{R}^{3\times 4}$ equal to:

$$\sum_{i=1}^{N} p_i^T R\widehat{S}q_i \left(p_i q_i^T \widehat{S}^T - R\widehat{S}q_i p_i^T R,\ \widehat{q}_i R^T p_i - S p_i^T R\widehat{q}_i^T S\right);$$

[1] In the literature, for different definitions of the rotation R, the matrix R in the above expression might differ by a transpose.

b) compute the optimal updating vector $\Delta = -\text{Hess}^{-1}G$:

(i) compute the vectors e_4, e_5 from S using Gram-Schmidt process and obtain the five basis tangent vectors $E^j \in T_R(SO(3)) \times T_S(S^2), 1 \leq j \leq 5$ as defined in the above,

(ii) compute the 5×5 matrix $(A)_{jk} = \text{Hess } F(E^j, E^k), 1 \leq j, k \leq 5$ using

$$
\begin{aligned}
& \text{Hess } F(X, Y) \\
= & \sum_{i=1}^{N} p_i^T (X_1 \widehat{S} + R\widehat{X}_2) q_i p_i^T (Y_1 \widehat{S} + R\widehat{Y}_2) q_i \\
+ & \; p_i^T R\widehat{S} q_i p_i^T \left(-\frac{1}{2}(X_1 Y_1^T + Y_1 X_1^T) R\widehat{S} \right. \\
- & \left. X_2^T Y_2 R\widehat{S} + (X_1 \widehat{Y}_2 + Y_1 \widehat{X}_2) \right) q_i,
\end{aligned}
$$

(iii) compute the 5 dimensional vector $(\mathbf{b})_j = -dF(E^j), 1 \leq j \leq 5$ using:

$$
dF(X) = \sum_{i=1}^{N} p_i^T R\widehat{S} q_i (p_i^T X_1 \widehat{S} q_i + p_i^T R\widehat{X}_2 q_i),
$$

(iv) compute the vector $\mathbf{a} \in \mathbb{R}^5$ such that $\mathbf{a} = A^{-1}\mathbf{b}$; define $\omega = (a_1, a_2, a_3)^T \in \mathbb{R}^3$ and $v = a_4 e_4 + a_5 e_5 \in \mathbb{R}^3$, then the optimal updating vector $\Delta = (R\widehat{\omega}, v)$;

2. move (R, S) in the direction Δ along the geodesic to $(\exp(R, \Delta_1), \exp(S, \Delta_2))$, using the formula for geodesics on $SO(3)$ and $\mathbb{P}^2$ respectively:

$$
\begin{aligned}
\exp(R, \Delta_1) &= R(I + \widehat{\omega} \sin t + \widehat{\omega}^2 (1 - \cos t)), \\
\exp(S, \Delta_2) &= S \cos \sigma + U \sin \sigma,
\end{aligned}
$$

where $t = \frac{1}{2}\text{tr}(\Delta_1^T \Delta_1), \omega = R^T \Delta_1 / t$ and $\sigma = ||\Delta_2||$, $U = \Delta_2 / \sigma$;

3. repeat if $||G|| \geq \epsilon$ for some pre-determined $\epsilon > 0$.

One may refer to [9] for the derivation of this algorithm as well as a discussion of applying the Riemannian Newton algorithm to the differential case.

5 Discussions and Future Work

We only applied Newton's method which to the motion recovery problem has the fastest convergence rate (among algorithms using second order information, see [2] for a comparison). In fact, the application of other conjugate gradient algorithms would be easier since they usually only involve the calculation of the first order information (the gradient, not Hessian), at the cost of slower convergence rate.

Like most iterative search algorithm, Newton and conjugate gradient algorithms are local methods, *i.e.* they do not guarantee convergence to the global minimum. For the motion recovery problem, one can use linear algorithms to get a close guess for the optimal motion and use it for initializing the proposed nonlinear algorithm. Simulation results of the proposed algorithms can be found in the full report [9]. Detailed study of the performance of the proposed algorithm and its sensitivity to noise on synthetic and real images is currently in progress.

References

[1] William M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry.* Academic Press, second edition, 1986.

[2] Alan Edelman, Tomas Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Analysis Applications*, to appear.

[3] B. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.

[4] A. D. Jepson and D. J. Heeger. Linear subspace methods for recovering translation direction. *Spatial Vision in Humans and Robots, Cambridge Univ. Press*, pages 39–62, 1993.

[5] Kenichi Kanatani. *Geometric Computation for Machine Vision.* Oxford Science Publications, 1993.

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

[7] Yi Ma, Jana Košecká, and Shankar Sastry. Motion recovery from image sequences: Discrete viewpoint vs. differential viewpoint. In *Proceeding of European Conference on Computer Vision, Volume II*, pages 337–53, 1998.

[8] Yi Ma, Jana Košecká, and Shankar Sastry. Motion recovery from image sequences: Discrete viewpoint vs. differential viewpoint. *Electronic Research Laboratory Memorandum, UC Berkeley*, UCB/ERL M98/11, February 1998.

[9] Yi Ma, Jana Košecká, and Shankar Sastry. Optimal motion from image sequences: A riemannian viewpoint. *Electronic Research Laboratory Memorandum, UC Berkeley*, UCB/ERL M98/37, June 1998.

[10] Stephen Maybank. *Theory of Reconstruction from Image Motion.* Springer-Verlag, 1993.

[11] Steven T. Smith. *Geometric Optimization Methods for Adaptive Filtering.* PhD thesis. Harvard University, Cambridge, Massachusetts, 1993.

[12] C. J. Taylor and D. J. Kriegman. Structure and motion from line segments in multiple images. *IEEE Transactions on PAMI*, 17(11):1021–32, 1995.

[13] J. Weng, T.S. Huang, and N. Ahuja. *Motion and Structure from Image Sequences.* Springer Verlag, 1993.