Spherical Averages and Applications to Spherical Splines and Interpolation

Sam Buss

Joint work with Jay Fillmore

Pixel-Cafe talk
December 18, 2002

Talk outline

I. Introduction

- Spherical averages
- Spherical splines and interpolation

II. Spherical averages

- Definition
- Uniqueness properties
- Computation algorithms
- Convex hull properties

III. Spherical Splines

- Definition
- Computation of interpolating splines
- Experimental results

Motivation: Lerping versus Slerping

Consider the problem of taking a weighted average of two points x, y on a sphere:

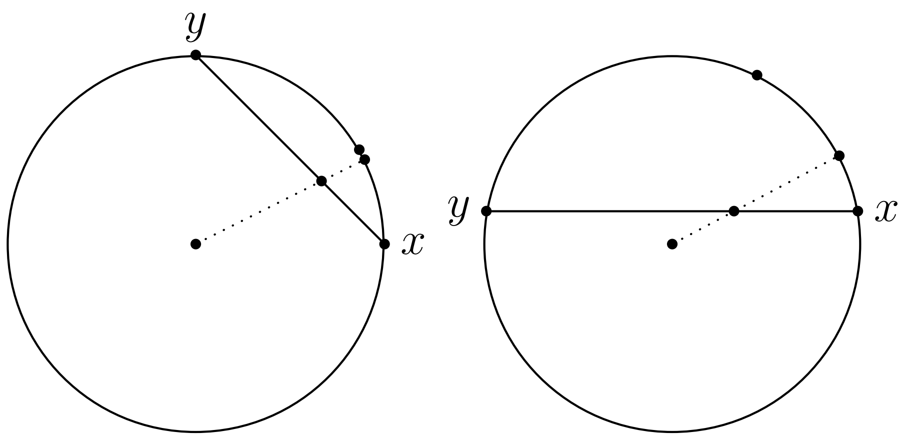
1. Form euclidean weighted average and normalize:

$$Lerp(x, y, \alpha) = \frac{(1-\alpha)x + \alpha y}{||(1-\alpha)x + \alpha y||}.$$

2. The point which is at fraction α of the way from x to y:
[Shoemake, 1985]

$$Slerp(x, y, \alpha).$$

Examples of lerp-ing versus slerping, $\alpha = 1/3$



Question: can slerping be generalized to allow forming weighted averages of more than two points at a time?

Spherical Averages

Let $p_1, \ldots, p_n$ be points on the d-sphere.
Let $w_1, \ldots, w_n$ be non-negative weights that sum to 1.

Goal: Define a weighted average

$$\sum_i w_i \cdot p_n$$

Desired properties:

- Spherical weighted averages should respect spherical distances, not Euclidean distances. They should be invariant under rotations of the sphere (no distortions near the poles or the equator, etc.)
- The averages should act qualitatively like weighted averages. In the limit (when points are close), should converge to Euclidean averages.
- If there are only two non-zero w_i weights, then it should reduce to geodesic arc interpolation ('slerp'-ing = 'spherical linear interpolation').
- The spherical weighted average should be a smooth function of w_i's and p_i's.

Application domains include: geological, astronomical, meteorological, rigid body orientation (quaternions), etc.
Prior work has mostly used Euclidean weighted averages, and occasionally stereographic projections.

Spherical splines and interpolation

Splines are a commonly used tool for generating smooth curves. A *spherical spline* is a spline curve which lies in a sphere.

Applications of spherical splines include robotics and computer graphics (using quaternions for orientation), Gondwanan apparent polar wander path, heartbeat analysis.

The most common framework for splines (Bézier curves and B-spline curves) involves choosing "control points" $p_1, \ldots, p_n$ and "blending functions" $f_1(t), \ldots, f_n(t)$ such that $f_i(t) \geq 0$ and such that $\sum_i f_i(t) = 1$ for all $t \in [0, 1]$, and letting

$$s(t) \ := \ \sum_i f_i(t) \cdot p_i$$

define the spline curve. This method is popular because the spline curve automatically inherits the smoothness of the functions f_i and since it provides a great deal of flexibility in specifying "knot positions" which allows the definition of curves with desired smoothness properties or even curves that have kinks or sharp turns.

Spherical splines have been studied extensively in the past:

Parker-Denham['79]: - normalized Euclidean curves

Thompson-Clark['82] and others: Gondwanan polar wander paths. Stereographic projections.

Fisher-Lewis['85]: blended smooth third-order segments on the sphere

Shoemake['85]: Spherical linear interpolation (slerp-ing) and Bézier-style curves.

Bézier curves and Catmull-Rum curves have been widely studied since: Shoemake['87], Duff['86], Ge-Revani['93], and many others. Generally works only for equally spaced knots.

Interpolating with (blended) spherical segments: Wang-Joe['93], Roberts-Bishop-Ganapathy['88], Kim-Nam['95], and others

Kim-Kim-Shim['95] Quaternion curves with general blending functions based on a series of slerp's. Does not respect time-symmetry, curve may exit the convex hull of the control points.

Dual quaternions - control position and orientation simultaneously. Ge-Revani['91], Juttler['94], Juttler-Wagner['96]. Not intrinsic to the sphere.

Natural splines which minimize curvature or tangential acceleration are defined by Gabriel-Kajiya['85], Jupp-Kent['87], Noakes-Heinzinger-Paden['89], Park-Ravani['97], Dam-Koch-Lillian['98], Barr-Currin-Gabriel-Hughes['92], Ramamoorthi-Barr['97], Zefran-Kumar['96]. Natural splines generally do not allow local control and tend to be fairly hard to compute.

Advantages of defining splines from spherical averages

We adapt splines based on blending function to the sphere by simply defining the spherical spline curve to be

$$s(t) \;:=\; \textstyle\bigcirc\!\!\!\!\sum f_i(t)\cdot p_i$$

This has a lot of advantages:

- Can define splines with non-interpolated control points (like B-splines). Can interpolate points by suitably choosing control points.
- Intrinsic to the sphere and invariant under rotations. Based on spherical distances, not Euclidean distances.
- Well-known strategies for choosing knot positions can be used to control the curve's characteristics.
- Knot positions do not need to be equally spaced or distinct.
- Splines can be constructed with local control.
- Computation time is fast enough for interactive applications and for many real-time applications.

Definition of Spherical Averages

In a Euclidean space, the weighted average $\sum_i w_i \cdot p_i$ can be shown to be equal to the point q where the function

$$f(q) \;=\; \sum w_i ||q - p_i||^2$$

is *minimized*. (Proof: set the first derivatives of f to zero and solve.) We define spherical weighted averages analogously:

Definition *Let* $\mathrm{dist}_S(p,q)$ *be the geodesic distance from* p *to* q *on the sphere* S*. Define*

$$f(q) \;=\; \frac{1}{2} \sum_i w_i \mathrm{dist}_S(q, p_i).$$

The spherical weighted average, denoted

$$\underset{i}{\textcircled{\scriptsize$\sum$}}\, w_i \cdot p_i$$

is equal to the point q *on* S *which minimizes* $f(q)$*.*

By compactness, f has a minimum value. In some cases the minimum is unique, and the spherical weighted average is well-defined....

Similar Reimannian barycenters have been defined for general manifolds. For these, good uniqueness and existence theorems are still lacking. [Cartan '28, Kendall '91, Corcuera-Kendall '99, ...]

Theorem *Suppose that the points $p_1, \ldots, p_n$ lie in a hemisphere H of S and at least one point in the interior of $\mathcal{H}$ has non-zero weight. Then the spherical weighted average is well-defined. Further, f has exactly one critical point in $\mathcal{H}$, and it is the spherical weighted average.*

The proof of this Theorem is surprisingly non-trivial. It uses the derivatives of f:

Let p_i be a fixed point and let T_q be the plane tangent to the sphere at a point q. Define $F_i(s)$ for $s \in T_q$ to equal $\mathrm{dist}_S(exp_q(s), p_i)$. Assume the T_q-axis x_i points from q away from p_i and that y is an orthogonal axis. Then

$$\left(\frac{\partial F_i}{\partial x}\right)_q = \rho \qquad \left(\frac{\partial F_i}{\partial y}\right)_q = 0$$

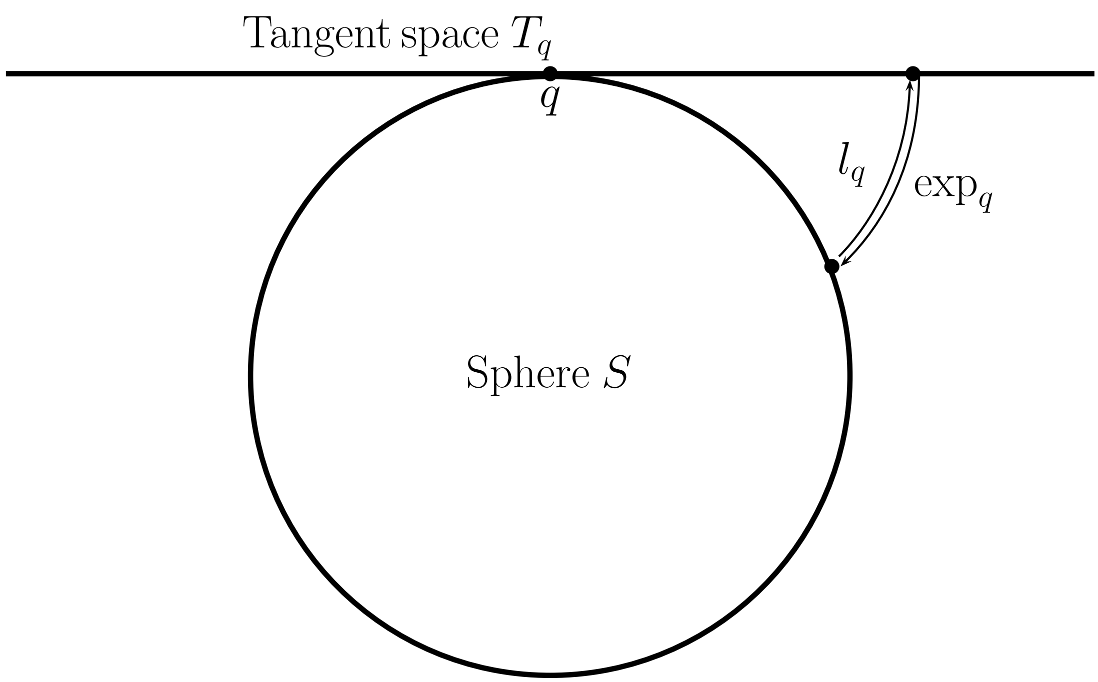
$$\left(\frac{\partial^2 F_i}{\partial x^2}\right)_q = 1 \qquad \left(\frac{\partial^2 F_i}{\partial y^2}\right)_q = \rho \cot \rho$$

$$\left(\frac{\partial^2 F_i}{\partial x \partial y}\right)_q = 0$$

where $\rho = \mathrm{dist}_S(p, q)$.

Exponential and Inverse-Exponential Maps

q - point on sphere S.
T_q - Plane tangent to S at q: the "tangent space"
Exponential map ($\exp_q$: wraps tangent plane down onto the sphere.
Inverse exponential map (l_q): the inverse mapping.



We write $\ell_q(p)$ for the inverse exponential map which maps p up to the tangent plane T_q preserving distance ($< \pi$) and angles from q.

The characterization of the first derivatives of F_i implies:

Lemma: *If q is a critical point of the function f, then*

$$\sum_i w_i \ell_q(p) \;=\; 0.$$

I.e., from q's vantage point, q appears to be the Euclidean weighted average of the p_i.

Returning the proof of the Uniqueness Theorem: The difficulty is that some of the second derivatives are negative when $\mathrm{dist}_S(p, q) > \pi/2$, so the weighted sum of distances squared function (f) is not the sum of concave-up functions. To solve this problem, we first assume we have a critical point $q \in \mathcal{H}$ where $\sum_i w_i \ell_q(p) = 0$ and then construct the line through q that goes to the sides of $\mathcal{H}$. Then points on opposite sides of the line are paired up so that their weighted perpindicular components in T_q cancel. For two such points p_1 and p_2, it can be proved that $w_1 F_1 + w_2 F_2$ is concave up at the critical point q.

Since f is concave up at all of its critical points in $\mathcal{H}$ and since f is increasing across the boundary of $\mathcal{H}$, there is a unique local (and hence global) minimum of f in $\mathcal{H}$. $\square$

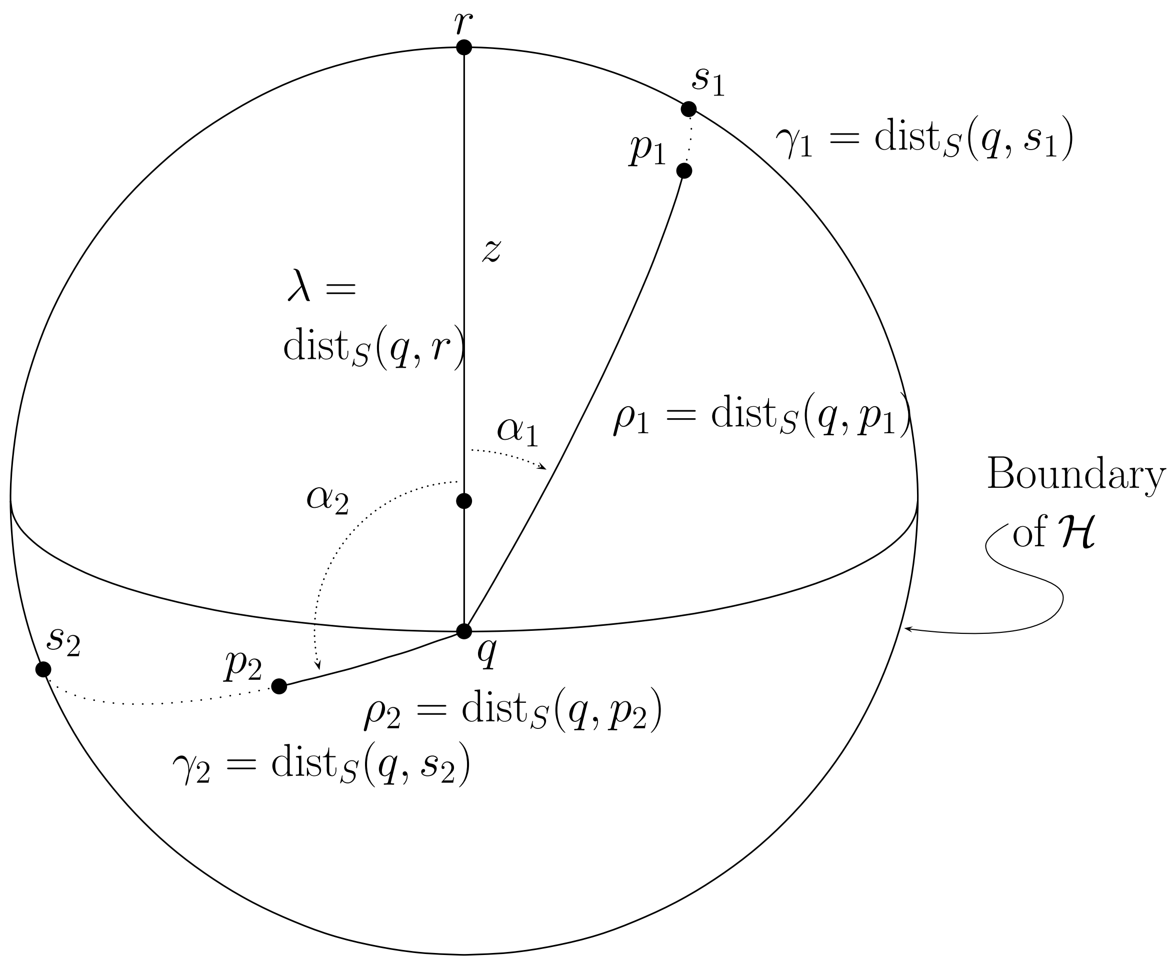


Figure 1: Top view of hemisphere $\mathcal{H}$ (orthogonal projection). All curves represent geodesics on the sphere.

We have $w_1\rho_1 \cos\alpha_1 + w_2\rho_2 \cos\alpha_2 = 0$.

Theorem *The Implicit Function Theorem can be used to prove that for points p_i in a hemisphere as above, that the spherical weighted average is a smooth (C^∞) function of the weights and the points p_i.*

Proof: The Hessian matrix is positive definite and therefore non-singular. $\square$

The partial derivatives of F_i's can be used to compute the derivatives of the spherical weighted average function.

Two Algorithms for Spherical Weighted Averages

The goal is to find a point q such that

$$\sum_i w_i \ell_q(p) \;=\; 0.$$

Algorithm A1: (Linear convergence rate)

1. Initialize: Choose initial estimate for q
2. Loop
   a. Map all points p_i to the tangent plane T_q
      (inverse exponential map)
   b. Compute Euclidean weighted average in T_q.
   c. Map down to the sphere (exponential map) as the new
      estimate for q
   d. Continue until $\sum_i w_i \ell_q(p)$ is close to zero.

Algorithm A2: (Quadratic convergence rate)
Use Newton's method to find a zero of $\sum_i w_i \ell_q(p)$.

Each iteration of Algorithm A2 requires forming the Hessian matrix by combining the second derivatives of the functions F_i and then solving the resulting matrix equation to approximate where the first derivatives of the functions F_i are all zero.

A Convexity Property

Definition *A subset C of S^d is* convex *iff for any two points $x, y \in C$, there is a shortest geodesic from x to y, which lies entirely in C.*

C is the convex hull *of $D \subset S^d$ iff C is the unique smallest convex set containing D.*

Definition A spherical weighted average is *proper* if there is a hemisphere $\mathcal{H}$ containing all the non-zero weighted points in its interior.

Theorem *Let $x_1, \ldots, x_n \in S^d$ and suppose that either $n \neq 2$ or x_1 and x_2 are not antipodal. Then $\{x_1, \ldots, x_n\}$ has a convex hull and it is equal to the set of points which can be written as proper weighted averages of $x_1, \ldots, x_n$.*

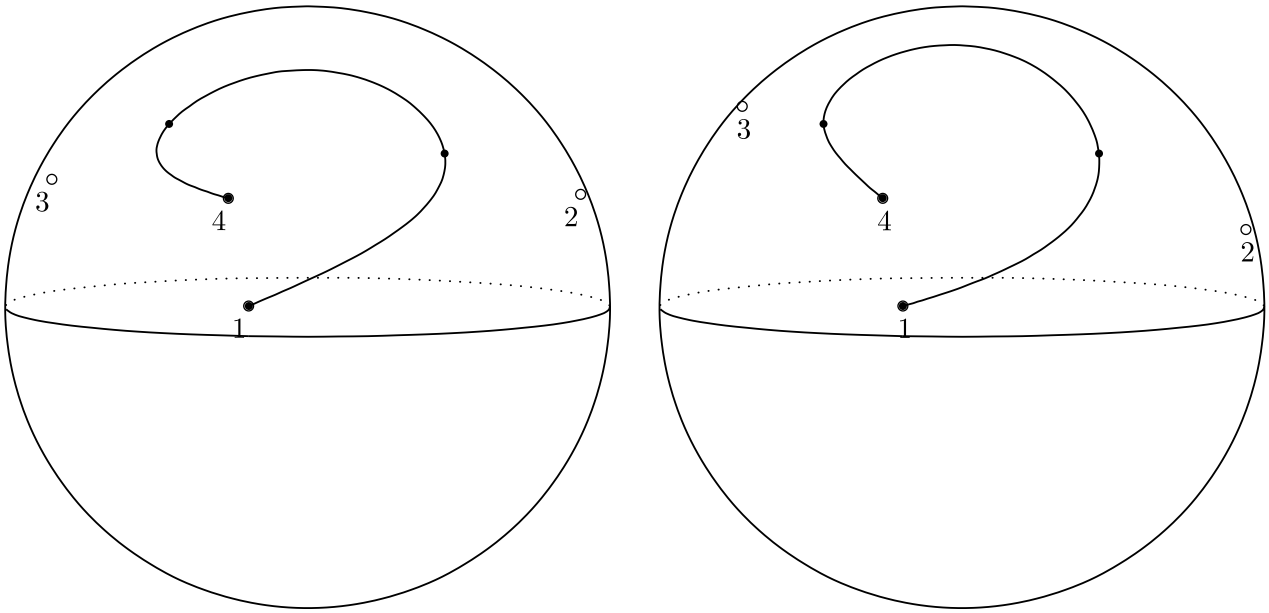
Splines based on Spherical Averages

Recall that we use blending functions $f_i(t)$ which are non-negative and sum to 1. Let $p_1, \ldots, p_n$ be control points. The spline curve is defined by

$$s(t) \; = \; \sum_i f_i(t) \cdot p_i.$$

For our experiments we used standard B-spline blending functions with doubled knots at the beginning and ending knot position so that the first and last control points are interpolated. The functions we used are piecewise degree 3 polynomials with continuous second derivatives everywhere.

Normally B-splines curves do not interpolate the control points. However, one can form a spline curve which interpolates given points $c_1, \ldots, c_n$ by suitably choosing control points $p_1, \ldots, p_n$. We used a linear convergence rate iterative algorithm to determine control points for interpolation. (This was faster than a quadratic convergence rate algorithm for 17 digits of accuracy!)



(a) (b)

Figure 2: Interpolating four points on S^2. (a) shows the curve generated with equally spaced knots. (b) is the curve generated with knots spaced proportionally to the spherical distance between interpolation points. The interpolated points are drawn as small disks; the control points are drawn as small circles.

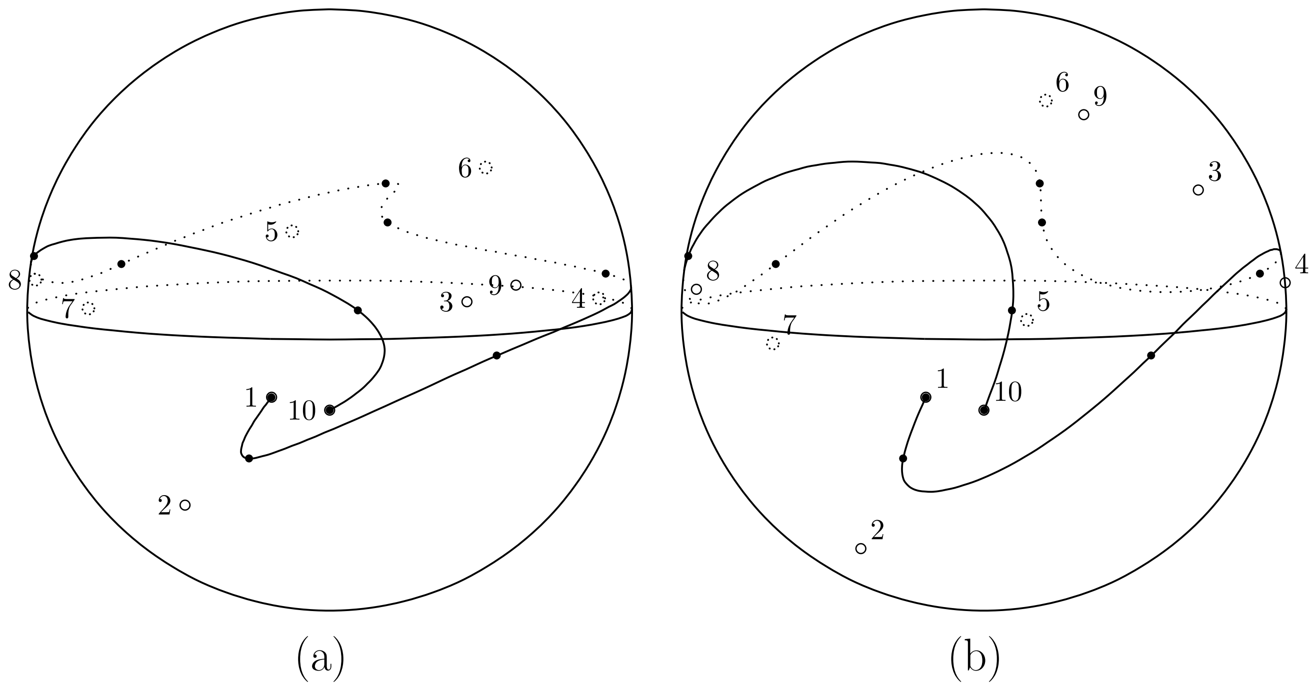


Figure 3: Interpolating ten points on S^2. (a) shows the curve generated with equally spaced knots. (b) is the curve generated with knots spaced proportionally to the spherical distance between interpolation points. Lines and control points that are on the back side of the sphere are drawn dotted.

Run times

Algorithm	Averaging 4 points on S^2	Averaging 4 points on S^3	Averaging 12 points on S^2	Averaging 12 points on S^3
A1	0.1538	0.196	0.3626	0.423
A2	0.0396	0.103	0.0846	0.265

Table 1: Run times in milliseconds of the computation of spherical weighted averages to approximately 16 digits of accuracy. Algorithm A1 is the linear convergence rate algorithm. Algorithm A2 is the quadratic convergence rate algorithm. The time reported is the average elapsed real time achieved over 5,000 computations.

A2's speed advantage erodes in higher dimensions since it has to solve a $d \times d$ matrix equation in each iteration. A1 merely adds up $(d+1)$-vectors.

Number of interpolated points	Computation of control points	Computation also of 64 curve points	Computation also of 256 curve points
4 points on S^2	0.276	2.716	7.801
12 points on S^2	1.768	4.213	10.943
4 points on S^3	0.318	6.388	20.133
12 points on S^3	2.582	9.037	27.107

Table 2: Run times in milliseconds of the computation of interpolating spherical spline curves to approximately 16 digits of accuracy. The first column reports the time needed to compute the control points. The second and third columns report this time plus the time needed to compute 64 and 256 points on the interpolating curve for equally spaced values of t. Algorithm S2 was used to compute the control points, and Algorithm A2 used to compute the points equally spaced on the curve. Algorithm A2 was seeded with good estimates for the successive points.