International Journal of Computational Intelligence and Applications
Vol. 4, No. 4 (2004) 401–416

Imperial College Press
www.icpress.co.uk

ROTATIONAL MATCHING PROBLEMS

GREGORY S. CHIRIKJIAN
Department of Mechanical Engineering, Johns Hopkins University
Baltimore, Maryland 21218, USA
gregc@jhu.edu.

PETER T. KIM
Department of Mathematics and Statistics, University of Guelph
Guelph, Ontario N1G 2W1, Canada
pkim@uoguelph.ca.

JA-YONG KOO
Department of Statistics, Korea University
Seoul 136-701, Korea
jykoo@korea.ac.kr

CHRISTINE H. LEE
Department of Pathology and Molecular Medicine, McMaster University
Hamilton, 50 Charlton Ave East, Ontario L8N 4A6, Canada
clee@mcmaster.ca

Received 8 May 2004
Accepted 13 August 2004

This paper addresses the issue of obtaining the optimal rotation to match two functions on the sphere by minimizing the squared error norm and the Kullback–Leibler information criteria. In addition, the accuracy in terms of the band-limited approximations in both cases are also discussed. Algorithms for fast and accurate rotational matching play a significant role in many fields ranging from computational biology to spacecraft attitude estimation. In electron microscopy, peaks in the so-called "rotation function" determine correlations in orientation between density maps of macromolecular structures when the correspondence between the coordinates of the structures is not known. In X-ray crystallography, the rotational matching of Patterson functions in Fourier space is an important step in the determination of protein structures. In spacecraft attitude estimation, a star tracker compares observed patterns of stars with rotated versions of a template that is stored in its memory. Many algorithms for computing and sampling the rotation function have been proposed over the years. These methods usually expand the rotation function in a bandlimited Fourier series on the rotation group. In some contexts the highest peak of this function is interpreted as the optimal rotation of one structure into the other, and in other contexts multiple peaks describe symmetries in the functions being compared. Prior works on rotational matching seek to maximize the correlation between two functions on the sphere. We also consider the use of the Kullback–Leibler information criteria. A gradient descent algorithm is proposed for obtaining the optimal

rotation, and a measure is defined to compare the convergence of this procedure applied to the maximal correlation and Kullback–Leibler information criteria.

Keywords: Correlations; FFT; Fourier analysis; gradient descent; Kullback–Leibler; rotations; sphere.

1. Introduction

Given two square-integrable functions $f_1(x)$ and $f_2(x)$ on $x \in \mathbb{X} \subset \mathbb{R}^3$, the general matching problem is usually defined as that of finding maxima of the function

$$\gamma(g) = \int_{x \in \mathbb{X}} f_1(x) f_2(g^{-1}x) dx. \tag{1}$$

Here g is an element of some transformation group, $\mathbb{G}$, which acts on $\mathbb{X} \subset \mathbb{R}^3$. Such problems arise in fields such as computer vision, image processing, robotics, and crystallography (among others). The particular $\mathbb{G}$ depends on the application. For example, in computer vision, the affine group plays a significant role. In the context of molecular structure determination, the group of rigid-body motions plays a significant role.[8,22] And while methods for computing matches under the full motion group are available[6,19,20] and a number of recent software algorithms perform full six-dimensional rigid-body matches,[4,30,10,31,23] it is often sufficient to perform a real-space match under rigid-body motions by first matching the centers of mass of two density functions, thereby restricting the problem to a search over rotations only.

The fast and accurate solution to the rotational matching problem is important in several areas of structural bioinformatics and biophysics.[7,21,22,3,34,2,9,39,29,36] For example, the structure of a large biological macromolecule of known sequence is often only determined experimentally as a three-dimensional density map. Then methods for altering a baseline structure[16,15] can be used to morph the structure to conform with the experimental data. Fitting an appropriate three-dimensional molecular structure model with full atomic detail to the experimentally-measured density requires fast and accurate rotational matching. Another example where rotational matching is important is in the statistical determination of binding pocket motifs and drug design.

In contrast, in X-ray crystallography, the electron density, $f(x)$, of a protein is not directly measured. Rather, if

$$F(k) = \mathcal{F}(f(x)) = \int_{\mathbb{R}^3} f(x) e^{-ik \cdot x} \, dx$$

is the Fourier transform of $f(x)$, then the experimentally measured quantity is the Patterson function $P(k) = |F(k)|^2$. Given a rigid-body motion $g = (A, a)$ where A is a rotation matrix and $a \in \mathbb{R}^3$, one observes that

$$|\mathcal{F}(f(g^{-1}x))|^2 = |F(A^t k)|^2,$$

where superscript t denotes transpose. In other words, the Patterson function does not depend on translation of the density $f(x)$. In X-ray crystallography, the matching of two Patterson functions under a pure rotation is an important tool as a step in structure determination.

While Fourier expansions on the rotation group have many important and useful properties, they do have drawbacks: bandlimited Fourier approximations of nonnegative functions need not be nonnegative everywhere; bandlimited Fourier expansions are not closed under multiplication or division. In a series of papers, Lo and Eshleman defined exponential Fourier densities on the sphere and rotation group.[24–26] These densities are defined as the exponential of bandlimited Fourier expansions on the sphere and rotation group. The motivation for studying these densities was that in the estimation of rotational processes such as spacecraft attitude estimation the use of operations such as conditioning is required. The closure of exponential Fourier densities under division makes them ideal for this use, though it is at the expense of closure under convolution. In this paper we expand on the use of exponential Fourier densities which naturally leads to using the Kullback–Leibler information divergence.

A summary of this paper goes as follows. In Sec. 2, we present the technical preliminaries followed by stating the rotational matching and informational rotational matching problems. In Sec. 3, we quantify the accuracy of the rotational matching schemes, and in Sec. 4, a discussion of the computational considerations is presented. All mathematical proofs are collected in Sec. 5.

2. Expanding Functions on the Sphere and Rotation Group and the Rotational Matching Problems

In this section we will provide a brief overview of Fourier analysis on the group of 3×3 rotation matrices, $SO(3)$, and on the unit 2–sphere, S^2, see Refs. 11, 37, 14, 13, 6, 17 for details. As well, we will explicitly state the rotational matching problem.

Let

$$u(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad a(\theta) = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}.$$

The well known Euler angle decomposition says, any $g \in SO(3)$ can almost surely be uniquely represented by $g(\phi,\theta,\psi) = u(\phi)a(\theta)u(\psi)$, where the three angles (ϕ,θ,ψ) are known collectively as the Euler angles, with $\phi \in [0,2\pi)$, $\theta \in [0,\pi)$, $\psi \in [0,2\pi)$. Consider the function,

$$D^{\ell}_{q_1 q_2}(\phi,\theta,\psi) = \mathrm{e}^{-iq_1\phi} d^{\ell}_{q_1 q_2}(\cos\theta)\mathrm{e}^{-iq_2\psi},$$

where, $d^{\ell}_{q_1 q_2}$ for $-\ell \leq q_1, q_2 \leq \ell$, $\ell = 0, 1, \ldots$ are related to the Jacobi polynomials. Define the $(2\ell+1) \times (2\ell+1)$ matrix by

$$D^{\ell}(g) = [D^{\ell}_{q_1 q_2}(g)], \tag{2}$$

where $-\ell \leq q_1, q_2 \leq \ell$, $\ell \geq 0$ and $g \in SO(3)$, these constitute the collection of inequivalent irreducible representations of $SO(3)$.

Any point on S^2, can be represented by $\omega = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta)^t$, where $\theta \in [0, \pi)$, $\phi \in [0, 2\pi)$ and superscript t denotes transpose. Let

$$Y^{\ell}_{q}(\omega) = (-1)^{q} \sqrt{\frac{(2\ell+1)(\ell-q)!}{4\pi(\ell+q)!}} P^{\ell}_{q}(\cos\theta) e^{iq\phi},$$

where $\theta \in [0, \pi)$, $\phi \in [0, 2\pi)$, $-\ell \leq q \leq \ell$, $\ell = 0, 1, \ldots$ and $P^{\ell}_{q}(\cdot)$ are the Legendre functions. We can think of Y^{ℓ}_{q} as the vector entries to the $2\ell + 1$ vector $Y^{\ell}(\omega) = [Y^{\ell}_{q}(\omega)]$, $\ell \geq 0$. In this situation $\{Y^{\ell}_{q} : |q| \leq \ell, \ell = 0, 1, \ldots\}$ form a complete orthonormal basis over the space of square integrable functions on the 2-sphere, $L^2(S^2)$ and is sometimes referred to as the spherical harmonics.

In terms of the Fourier basis, the relation between $SO(3)$ and S^2 can be described in terms of the Euler angles. Furthermore, the following relation will be used

$$Y^{\ell}_{q}(g\omega) = \sum_{|j| \leq \ell} Y^{\ell}_{j}(\omega) D^{\ell}_{jq}(g^{-1}), \tag{3}$$

for $\omega \in S^2$ and $g \in SO(3)$.

Let $f \in L^2(S^2)$. We define the spherical Fourier transform on S^2 by

$$\hat{f}^{\ell}_{q} = \int_{S^2} f(\omega) \bar{Y}^{\ell}_{q}(\omega) d\omega, \tag{4}$$

where $d\omega$ is the spherical measure on S^2 and overbar denotes complex conjugation. Again we think of (4) as the vector entries of the $(2\ell+1)$ vector $\hat{f}^{\ell} = [\hat{f}^{\ell}_{q}]$, $-\ell \leq q \leq \ell$, $\ell = 0, 1, \ldots$. Now for each $\ell > 0$, define the $(2\ell+1)$-dimensional vector space $\mathcal{E}_{\ell}$ as the span of $\{Y^{\ell}_{q} : |q| \leq \ell\}$ and equip it with the usual inner product thus making $\mathcal{E}_{\ell}$ a $(2\ell+1)$-dimensional inner product space with inner product $\langle \cdot, \cdot \rangle_{\ell}$. The spherical inversion can be obtained by,

$$f(\omega) = \sum_{\ell \geq 0} \langle \hat{f}^{\ell}, \bar{Y}^{\ell}(\omega) \rangle_{\ell}, \tag{5}$$

for $\omega \in S^2$.

We would like to present our main results in terms of Sobolev spaces. Indeed, on the space $C^{\infty}(S^2)$ of infinitely continuous differentiable functions on S^2, consider the so-called Sobolev norm $\| \cdot \|_{H_s}$ of order s defined in the following way. For any function $h = \sum_{\ell} \langle \hat{h}^{\ell}, \bar{Y}^{\ell} \rangle_{\ell}$ let

$$\|h\|^2_{H_s} = \sum_{\ell > 0} (\ell(\ell+1))^{s} \|\hat{h}^{\ell}\|^2_{\ell}, \tag{6}$$

where $\|\cdot\|_\ell$ is the induced norm from the inner product. One can verify that (6) is indeed a norm. Denote by $H_s(S^2)$ the (vector-space) completion of $C^\infty(S^2)$ with respect to (6), the Sobolev norm of order $s > 1$. For some fixed constant $Q > 0$, let $H_s(S^2, Q)$ denote the smoothness class of functions $h \in H_s(S^2)$ which satisfy $\|h\|_{H_s} \leq Q$, the Sobolev ball of radius Q.

As stated in the introduction, the rotational correlation function can be defined for two functions f, h: $S^2 \to \mathbb{R}$, on $SO(3)$ by

$$\rho(g) = \int_{S^2} f(\omega) h(g^{-1}\omega) d\omega, \tag{7}$$

for $g \in SO(3)$. Define the left regular representation Λ_g: $L^2(S^2) \to L^2(S^2)$ by

$$\Lambda_g(f)(\omega) = f(g^{-1}\omega), \tag{8}$$

for $f \in L^2(S^2)$, where the latter denotes the space of square-integrable functions on S^2.

The maximization of (7) is really equivalent to the minimization of the $L^2(S^2)$-norm and will be referred to as the rotational matching problem. In particular, our problem would be

$$\min_{g \in SO(3)} \|f - \Lambda_g h\|^2, \tag{9}$$

where $\|\cdot\|$ denotes the $L^2(S^2)$-norm.

An informational measure of deviation can also be used. Consider f_1, f_2: $S^2 \to \mathbb{R}$ and define the Kullback–Leibler information divergence by

$$\mathcal{D}(f_1 \| f_2) = \int_{S^2} f_1(\omega) \log\left(\frac{f_1(\omega)}{f_2(\omega)}\right) d\omega. \tag{10}$$

Thus in the context of (9), with respect to (10), we would have

$$\min_{g \in SO(3)} \mathcal{D}(f \| \Lambda_g h), \tag{11}$$

which we will refer to as the informational rotational matching problem.

In practice, we cannot simply minimize (11) just as we cannot minimize (9) but instead would minimize bandlimited version of (7). Therefore, some truncation has to be invoked for the minimization of (9). This will be pursued in the following two sections.

As far as minimization of the objective functions are concerned, the minimization of (9) would differ from the minimization of (10) in as much as the former would be an L^2-approximation, so that the approximating function can take negative values, while the latter would be minimization with a positivity constraint on the function in question. Some discussions related to this matter can be found in the statistical literature.[1]

2.1. *Rotational matching*

For f, h: $S^2 \to \mathbb{R}$, consider the left regular representation (8). Then using (3), we have

$$\widehat{\Lambda_g(h)}^{\ell}_{q} = \sum_{|j|\leq \ell} \hat{h}^{\ell}_{j} \overline{D^{\ell}_{jq}}(g^{-1}).$$

Making note of the fact that $\overline{D^{\ell}}(g)^t = D^{\ell}(g^{-1})$ for all $g \in SO(3)$, the matrix from of the above is

$$\widehat{\Lambda_g(h)}^{\ell} = D^{\ell}(g)\hat{h}^{\ell},$$

where $\hat{h}^{\ell}$ is the vector of Fourier transforms of $h \in L^2(S^2)$, for $\ell \geq 0$. By the Plancherel formula, (7) the rotational correlation function can be re-expressed as,

$$\rho(g) = \sum_{\ell>0} \langle \hat{f}^{\ell}, D^{\ell}(g)\hat{h}^{\ell} \rangle_{\ell}. \tag{12}$$

The rotational correlation function (12) is an infinite series and so from a practical point of view, one has to cut-off the series at some finite point. Therefore, define the truncated rotational correlation function to be

$$\rho_m(g) = \sum_{\ell=0}^{m} \langle \hat{f}^{\ell}, D^{\ell}(g)\hat{h}^{\ell} \rangle_{\ell}, \tag{13}$$

for some $m > 0$.

In terms of applications, we should mention that the above technique has relevance to the docking of atomic structures into electron-microscopy maps and the molecular-replacement problem in X-ray crystallography.[18] Further analyses related to this work will be mentioned below after we quantify the magnitude of accuracy.

2.2. *Informational rotational matching*

Let $\mathcal{S}_m$ be the collection of all functions of the following form

$$s(\omega;\beta) = \sum_{\ell=0}^{m} \langle \beta^{\ell}, \overline{Y(\omega)}^{\ell} \rangle_{\ell}, \tag{14}$$

where $\beta^{\ell} \in \mathbb{R}^{2\ell+1}$ and let $\mathcal{F}_m$ be the collection of all functions of the following form

$$f_m(\omega;\beta) = \exp\left\{ \sum_{\ell=1}^{m} \langle \beta^{\ell}, \overline{Y(\omega)}^{\ell} \rangle_{\ell} - \Psi(\beta) \right\}, \tag{15}$$

where $\beta = (\beta^{\ell})_{\ell=1}^{m}$ and $\Psi(\beta) = \log\left(\int \exp\{ \sum_{\ell=1}^{m} \langle \beta^{\ell}, \bar{Y}^{\ell}(\omega) \rangle_{\ell} \} \, d\omega \right)$. By definition,

$$\log f_m(\beta) \in \mathcal{S}_m.$$

In certain cases when the context is clear, we will write $f_m(\beta) = f_m(\omega;\beta)$.

The information projection of f onto $\mathcal{S}_m$ will be denoted by f_m^* having the property

$$\mathcal{D}(f\|f_m^*) \leq \mathcal{D}(f\|f_m(\beta)), \tag{16}$$

for all $f_m(\beta) \in \mathcal{S}_m$. This is characterized by finding the unique $f_m^* \in \mathcal{S}_m$ satisfying

$$\int_{S^2} f_m^*(\omega)\bar{Y}^\ell(\omega)d\omega = \int_{S^2} f(\omega)\bar{Y}^\ell(\omega)d\omega, \tag{17}$$

for $\ell = 1, \ldots, m$.

Thus in the context of information projection, the practical side to the minimization problem (11) would be

$$\min_{g\in SO(3)} \mathcal{D}\left(f_m^*\|\Lambda_g h_m^*\right). \tag{18}$$

3. Accuracy of Rotational Matching

Our concern then is to determine the accuracy of the bandlimited approximations to (9) and (11). Below we quantify the approximations.

3.1. *Accuracy of bandlimited rotational matching*

With respect to the rotational correlation function, one would like to estimate the degree of accuracy of the truncated rotational correlation function. In order to quantify such measures we will assume that $f, h \in H_s(S^2, Q)$. Thus we have the following which is proved in Sec. 5.

Theorem 3.1. *Suppose $f, h \in H_s(S^2, Q)$. Then*

$$|\rho(g) - \rho_m(g)| \leq Qm^{-2s}$$

where $s \geq 2$, for all $g \in SO(3)$.

Since $Q > 0$ is fixed, the Sobolev ball is fixed. We can however, vary the Q by making it go to infinity at a rate that would only marginally change the order of magnitude. Indeed, we have the following.

Corollary 3.1. *Let $0 < \epsilon < 2$ and suppose $f, h \in H_s(S^2, m^{2\epsilon})$. Then*

$$|\rho_m(g) - \rho(g)| \leq m^{-2(s-\epsilon)}$$

where $s \geq 2$, for all $g \in SO(3)$.

One can see from Corollary 3.1 that by setting $s = 2$ and $\epsilon = 1$ that the margin of error is going to be m^{-2}. Thus, for example, as relevant to the applications in Ref. 18, if we want a margin of error of 0.01, we would have to calculate to 10 eigenspaces which involves 121 Fourier coefficients. If we demand even more precision of say 0.001, we would need to calculate to 30 eigenspaces which involves 961 Fourier coefficients.

3.2. *Accuracy of informational rotational matching*

With respect to the Kullback–Leibler information criteria (10), one would like to estimate the degree of accuracy of minimizing the information projection. In order to quantify such measures we will assume that $\log f, \log h \in H_s(S^2, Q)$. Thus we have the following.

Theorem 3.2. *Suppose* $\log f, \log h \in H_s(S^2, Q)$. *Then, as* $m \to \infty$,

$$|\mathcal{D}(f\|\Lambda_g h) - \mathcal{D}(f_m^*\|\Lambda_g h_m^*)| \leq Cm^{-s}$$

where $s > 1$, *and some* $C > 0$ *for all* $g \in SO(3)$.

4. Computational and Numerical Issues in Rotational Matching

In this section we describe numerical methods for rotation matching. In Subsec. 4.1, Fourier methods based on bandlimited expansions are reviewed which would be applicable to L^2-minimizing. In Subsec. 4.2, a gradient descent algorithm is proposed which would be applicable to the minimizing of the Kullback–Leibler information criterion.

4.1. *Fast rotational matching*

If S^2 is sampled at $\mathcal{O}(B^2)$ points and $SO(3)$ is sampled at $\mathcal{O}(B^3)$ points, then the numerical evaluation of rotational correlation function in (7) would require $\mathcal{O}(B^5)$ computations when performed using brute force sampling and a numerical quadrature rule in place of the integral. If N denotes the number of points sampled in $SO(3)$, this represents an $\mathcal{O}(N^{5/3})$ computation. However, methods from computational noncommutative harmonic analysis can be used to reduce this burden.

Sampling theorems and fast Fourier (FFT) transform techniques for the rotation group have been developed.[27,28] Essentially, the double coset decomposition of $SO(3)$ corresponding to $z - x - z$ Euler angles yields matrix elements of the irreducible unitary representation matrices, (2), which in principle lend themselves to fast transforms in each coordinate (Euler angle) when exact arithmetic is used. We briefly review this and other approaches for fast computation of bandlimited Fourier series on $SO(3)$.

The whole spectrum of a bandlimited function on $SO(3)$ [such as $\rho(g)$ from (7)] can be calculated fast in principle by using the classical FFT over ϕ and ψ and fast functional transforms over θ. By using a quadrature rule, the spherical Fourier transform can be sampled in each coordinate at $\mathcal{O}(B)$ values to exactly compute the integral in (12), where B denotes the bandlimit. The whole of $SO(3)$ is then sampled at $N = \mathcal{O}(B^3)$ points. Explicitly, one first calculates

$$\tilde{\rho}_{q_1}(\theta, \psi) = \int_0^{2\pi} \rho(\phi, \theta, \psi) e^{iq_1\phi} d\phi$$

for all $q_1 \in \{-B, \ldots, B\}$ and all sample values of θ and ψ. This requires $\mathcal{O}(B^2 \cdot B \log B)$ operations. Then one calculates

$$\tilde{\tilde{\rho}}_{q_1 q_2}(\theta) = \int_0^{2\pi} \tilde{\rho}_{q_1}(\theta, \psi) e^{iq_2\psi} d\psi$$

in $\mathcal{O}(B^2 \cdot B \log B)$ operations ($\mathcal{O}(B \log B)$ for each value of n and θ). Finally, all the $SO(3)$-Fourier coefficients are calculated in $\mathcal{O}(B^2 \cdot B(\log B)^2)$ operations ($\mathcal{O}(B(\log B)^2)$ for each value of q_1 and q_2).

Since the limiting calculation is the $\mathcal{O}(B(\log B)^2)$ required for the fast functional transform in the variable θ, the whole procedure is in principle $\mathcal{O}(N(\log N)^2)$ for all values of q_1, q_2, ℓ up to the band-limit B. The reconstruction of the original function from its bandlimited spectrum,

$$\rho(g) = \sum_{\ell=0}^{B-1} (2\ell + 1) \sum_{m=-\ell}^{\ell} \sum_{n=-\ell}^{\ell} \hat{\rho}_{mn}^{\ell} D_{nm}^{\ell}(g), \tag{19}$$

requires the same $\mathcal{O}(N(\log N)^2)$ computation as was required to obtain the spectrum (since these problems are dual). But in practice, numerical instabilities associated with the computation of fast functional transforms required in the θ variable limit the applicability of this approach.

One approach which is numerically stable is to expand the Wigner D-functions, (2), in bandlimited Fourier series and use the FFT not only on ϕ and ψ, but in all three rotational dimensions after a change of coordinates. Such an approach was originally introduced in Ref. 33 and adopted by Ref. 18. While this approach does make use of a three-dimensional FFT, and appears to be of computational value, it still scales as $\mathcal{O}(B^4) = \mathcal{O}(N^{4/3})$ because of the cost of transforming to and from the new coordinates in which the FFT is computed.

In contrast to both of the above recent approaches, if the rotation function is computed directly in the θ dimension as proposed by Ref. 7 (i.e. by storing the $d_{q_1 q_2}^{\ell}(\cos\theta)$ values in advance and using $\mathcal{O}(B^2)$ operations instead of using an $\mathcal{O}(B(\log B)^2)$ fast functional transform), then the $SO(3)$ Fourier transforms for all q_1, q_2, ℓ up to the band-limit can still be performed in $\mathcal{O}(B^4) = \mathcal{O}(N^{4/3})$ arithmetic operations. Hence, the original method for fast rotation function evaluation scales as well as the more recent approach. And while it does not scale as well as approaches based on fast functional transforms, it does not suffer from numerical instabilities either.

4.2. *Gradient descent procedures*

If two functions on S^2 are evaluated at $\mathcal{O}(B^2)$ points and stored, then the computation of the integral for either the rotational correlation function (7) or the informational measure (10) requires $\mathcal{O}(B^2)$ evaluations/interpolations for each orientation $g \in SO(3)$. Therefore, instead of computing the value of the functions in (7) or (10) at each point in a fine grid in $SO(3)$, we consider sampling at a very

sparse grid, and driving the value of these integrals to their minimal values using a gradient descent algorithm. Similar methods have been used elsewhere in the context of satellite attitude estimation[32] and problems in robotics.[5,12,35]

Given a function $f(g)$ which we seek to minimize, the gradient descent procedure is to seek a direction which reduces the value. Infinitesimal rotational motions are captured by the basis elements of the Lie algebra of $SO(3)$.

Let $\{X_i\}$ denote the set of basis elements of the Lie algebra $\mathfrak{so}(3)$:

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}; \quad X_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}; \quad X_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

One can define derivatives of the form

$$X_i^R f(g) = \frac{d}{dt}(f(g \circ \exp(tX_i))|_{t=0}.$$

The collection of these directional derivatives is in effect a gradient vector pointing in the direction of steepest ascent. This may be used to update the current group element as

$$g \rightarrow g \circ \exp\left(-\epsilon \sum_{i=1}^{3} X_i [X_i^R f(g)]\right),$$

where ϵ is a small step size which is chosen. Iterating this process will lead to the convergence to the nearest local minimum of the function.

A potential benefit of this approach is that it does not require the evaluation of the integral or function at all values of g in a fine grid on $SO(3)$. Even if the path meanders, or if the gradient descent method must be restarted from several different initial values of g, it is very plausible that it would converge in fewer than the $\mathcal{O}(B^2)$ steps required to make it as expensive as most of the methods described in the previous section.

However, having said this, the quality of the function on $SO(3)$ being minimized has a large influence over this behavior. A function with many deep local minima which divert the gradient descent path away from the global minimum is not as desirable as one with few minima. This inspires us to define a "goodness criterion" for cost functions. Let g_0 denote the global minimum of $f(g)$. If all of the gradient vectors point directly toward this global minimum, the gradient descent procedure will converge without being diverted. The geodesic in $SO(3)$ connecting g and g_0 is defined by the path

$$g_p(t) = g \exp(t \log(g^{-1} g_0)).$$

The tangent vector to this path at $t = 0$ can be compared with the direction $\sum_{i=1}^{3} X_i [X_i^R f(g)]$ used in the gradient descent procedure. The inner product of the normalized versions of these two vectors in the tangent to $SO(3)$ at g can be computed for each value of $g \in SO(3)$. Integrating this inner product over all of $SO(3)$ gives a "badness measure" which is zero for a function with a single minimum.

In future work we will perform numerical tests to compare the relative goodness/badness of the rotational correlation function and Kullback–Leibler information criteria for $SO(3)$.

5. Proofs

This section provides proofs of our early mathematical statements.

5.1. *Proof of Theorem 3.1*

The proof of Theorem 3.1 goes as follows. By (12) and (13), we have

$$\begin{aligned}
|\rho(g) - \rho_m(g)| &\le \sum_{\ell>m} |\langle \hat{f}^\ell, D^\ell(g)\hat{h}^\ell\rangle_\ell| \\
&\le \sum_{\ell>m} \|\hat{f}^\ell\|_\ell \|\hat{h}^\ell\|_\ell \\
&= \sum_{\ell>m} (\ell(\ell+1))^{-s} (\ell(\ell+1))^{s/2} \|\hat{f}^\ell\|_\ell (\ell(\ell+1))^{s/2} \|\hat{h}^\ell\|_\ell \\
&\le m^{-2s} \left\{ \sum_{\ell>m} (\ell(\ell+1))^s \|\hat{f}^\ell\|_\ell^2 \right\}^{1/2} \left\{ \sum_{\ell>m} (\ell(\ell+1))^s \|\hat{h}^\ell\|_\ell^2 \right\}^{1/2} \\
&\le Q m^{-2s},
\end{aligned}$$

where the first line is by the triangle inequality, the second is by the Cauchy–Schwarz inequality and the fact that D^ℓ is a unitary operator, the third is definition, the fourth is another application of the Cauchy–Schwarz inequality and the last line is due to the fact that $f, h \in H_s(S^2, Q)$. □

5.2. *Proof of Theorem 3.2*

Let

$$\gamma_2(f) = \inf_{s\in\mathcal{S}_m} \|\log f - s\|_2 \quad \text{and} \quad \gamma_\infty(f) = \inf_{s\in\mathcal{S}_m} \|\log f - s\|_\infty$$

be the L^2 and L^∞ error approximations for a density function f by some $s \in \mathcal{S}_m$. Define A by

$$\|s\|_\infty \le A\|s\|_2 \quad \text{for all } s \in \mathcal{S}_m.$$

Let $\|\beta\|$ denote the Euclidean norm of a vector β. Set $\mathcal{I} = \{(\ell, k)\colon k = -\ell, \ldots, \ell, \ell \ge 0\}$, $\mathcal{I}_m = \{(\ell, k)\colon k = -\ell, \ldots, \ell, \ell = 0, \ldots, m\}$ and $\mathcal{I}_m^c = \{(\ell, k)\colon k = -\ell, \ldots, \ell, \ell > m\}$.

Lemma 5.1. *Suppose that $A\gamma_2 = o(1)$ and $\gamma_\infty < \infty$. If $C_1^{-1} \le f \le C_1$ then for n sufficiently large*

$$D(f\|f_m^*) \le \frac{C_1}{2} e^{\gamma_\infty} \gamma_2^2.$$

Proof. For $g = \log f$, let

$$s_m(g) = \sum_{\mathcal{I}_m} \hat{g}_k^\ell Y_k^\ell.$$

Set $\theta^* = \left(\int f \bar{Y}^\ell f\right)$ and $\theta = \left(\int Y_k^\ell f(\delta)\right)$, where $\delta = \left(\hat{g}_k^\ell\right)$. The entries in the vector $\theta^* - \theta$ are seen to be coefficients in the $L^2(S^2)$ orthonormal projection of $f - f(\delta)$ onto $\mathcal{S}_m$. By Bessel's inequality, the boundedness of f, we have

$$\begin{aligned}\|\theta^* - \theta\|^2 &\leq \|f - f(\delta)\|_2^2 \\ &\leq C_1^2 \int \frac{(f - f(\delta))^2}{f} \\ &\leq C_1^2 \|g - s(g)\|_2^2 \exp\left\{2\|g - s_m(g)\|_\infty - 2\{\hat{g}_0^0 + \Psi(\delta)\}\right\} \\ &\leq C_1^2 \mathrm{e}^{4\gamma_\infty} \gamma_2^2.\end{aligned}$$

For the last inequality we have used the fact that $|\Psi(\delta) + \hat{g}_0^0| \leq \|g - s_m(g)\|_\infty$, since $\Psi(\delta) + \hat{g}_0^0 = \log\left\{\int \exp(s_m(g) - g) f\right\}$. From this same fact we have

$$\|\log f / f(\delta)\|_\infty \leq 2\|g - s_m(g)\|_\infty = 2\gamma_\infty$$

and together with $C_1^{-1} \leq f \leq C_1$, we obtain

$$\|\log f(\delta)\|_\infty \leq \log C_1 + 2\gamma_\infty.$$

Now, if $C_1 \mathrm{e}^{2\gamma_\infty} \gamma_2 \leq 1/(4ebA)$, that is, if $A\gamma_2 = o(1)$, then it follows from Lemma 5[1] that the solution β^* to the equation $\left(\int Y_k^\ell f(\beta)\right) = \theta^*$ exists and that

$$\|\log f_m^* / f(\delta)\|_\infty \leq \epsilon,$$

where $\epsilon = 4C_1^2 \exp(4\gamma_\infty + 1) A\gamma_2$. So by the triangle inequality, we obtain

$$\|\log f / f_m^*\|_\infty \leq 2\gamma_\infty + \epsilon,$$

and

$$\|\log f_m^*\|_\infty \leq \log C_1 + 2\gamma_\infty + \epsilon. \tag{20}$$

Therefore, it follows from the boundedness of f and Lemma 1,[1] we have

$$D(f\|f_m^*) \leq D(f\|f(\delta)) \leq \frac{1}{2}\mathrm{e}^{\|g - s(g)\|_\infty} C_1 \|g - s(g)\|_2^2 \leq \frac{1}{2} C_1 \mathrm{e}^{\gamma_\infty} \gamma_2^2. \qquad \square$$

Lemma 5.2. *If $\log f \in H_s(S^2, Q)$, then there exists a constant C_1 such that*

$$C_1^{-1} \leq f \leq C_1.$$

Proof. Write $g = \log f = \sum_{\mathcal{I}} \hat{g}_k^\ell Y_k^\ell$. Observe that

$$\begin{aligned}|g(\omega)|^2 &\leq \left(\sum_{\mathcal{I}} (\ell(\ell+1))^s |\hat{g}_k^\ell|^2\right)\left(\sum_{\mathcal{I}} (\ell(\ell+1))^{-s} |Y_k^\ell|^2\right) \\ &\leq Q \sum_{\ell=0}^{\infty} (\ell(\ell+1))^{-s} (2\ell+1).\end{aligned}$$

Since $s > 1$, the series $\sum_{\ell=0}^{\infty} (\ell(\ell+1))^{-s}(2\ell+1)$ converges thus giving us a bound. $\square$

Lemma 5.3.

$$A \le Cm \tag{21}$$

$$\gamma_2(f) \le Cm^{-s} \tag{22}$$

$$\gamma_\infty(f) \le Cm^{-s+1}, \tag{23}$$

where C is a generic constant independent of m.

Proof. To determine A, choose any element $s = \sum_{\mathcal{I}_m} \beta_k^\ell Y_k^\ell$ in $\mathcal{S}_m$. By the Cauchy–Schwarz and Parseval's inequality, we have that, uniformly in $\omega \in S^2$,

$$\begin{aligned}
|s(\omega)| &\le \left(\sum_{\mathcal{I}_m} |Y_k^\ell(\omega)|^2 \right)^{1/2} \left(\sum_{\mathcal{I}_m} |\beta_k^\ell|^2 \right)^{1/2} \\
&\le \left(\sum_{\mathcal{I}_m} |Y_k^\ell(\omega)|^2 \right)^{1/2} \|s\|_2 \\
&\le C \left(\sum_{l=0}^{m} (2\ell+1) \right)^{1/2} \|s\|_2 \\
&\le Cm\|s\|_2,
\end{aligned}$$

which proves (21).

Since

$$(m(m+1))^s \sum_{\mathcal{I}_m^c} |\hat{g}_k^\ell|^2 \le \sum_{\mathcal{I}} (\ell(\ell+1))^s |\hat{g}_k^\ell|^2 < Q,$$

we have the bound on γ_2 as follows:

$$\gamma_2^2 \le \sum_{\mathcal{I}_m^c} |\hat{g}_k^\ell|^2 < Q(m(m+1))^{-s} \le Cm^{-2s}.$$

It follows from the Cauchy–Schwarz inequality and the addition formula for spherical harmonics that the error γ_∞^2 is bounded by

$$\begin{aligned}
\left| \sum_{\mathcal{I}_m^c} \hat{g}_k^\ell Y_k^\ell(\omega) \right|^2 &\le \left(\sum_{\mathcal{I}_m^c} |Y_k^\ell(\omega)|^2 (\ell(\ell+1))^{-s} \right) \left(\sum_{\mathcal{I}_m^c} (\ell(\ell+1))^s |\hat{g}_k^\ell|^2 \right) \\
&\le Q \sum_{l=m+1}^{\infty} (\ell(\ell+1))^{-s} (2\ell+1) \\
&\le Cm^{-2s+2}.
\end{aligned} \tag{24}$$

□

Lemma 5.4.

$$\mathcal{D}(f \| f_m^*) \le Cm^{-2s}. \tag{25}$$

Proof. As $m \to \infty$, we have that $\gamma_\infty \le Cm^{-s+1} = o(1)$, $\gamma_2 \le Cm^{-s}$ and $A\gamma_2 \le Cm^{-s+1} = o(1)$. Consequently, from Lemma 5.1, we obtain the bound (25). □

Lemma 5.5. *For any $f_m \in \mathcal{F}_m$, we have a Pythagorean-like identity*

$$\mathcal{D}(f\|f_m) = \mathcal{D}(f|f_m^*) + \mathcal{D}(f_m^*\|f_m).$$

Proof. Observe that

$$\begin{aligned}
\mathcal{D}(f\|f_m) &= \int f \log \frac{f}{f_m} \\
&= \int f \left(\log \frac{f}{f_m^*} + \log \frac{f_m^*}{f_m} \right) \\
&= \mathcal{D}(f|f_m^*) + \int f \log \frac{f_m^*}{f_m} \\
&= \mathcal{D}(f|f_m^*) + \int f_m^* \log \frac{f_m^*}{f_m} \\
&= \mathcal{D}(f|f_m^*) + \mathcal{D}(f_m^*\|f_m).
\end{aligned}$$

The third equality follows from the relation (17). □

Now we prove Theorem 3.2. Since the proof is for some fixed $g \in SO(3)$, without loss of generality, we will assume $g = 1$, the unit element so that $\Lambda_1 h = h$. Let h_m^* be the information projection of h onto $\mathcal{F}_m$. From Lemma 5.5, we have

$$\mathcal{D}(f\|h_m^*) = \mathcal{D}(f|f_m^*) + \mathcal{D}(f_m^*\|h_m^*),$$

which implies that

$$\mathcal{D}(f\|h) - \mathcal{D}(f_m^*\|h_m^*) = \mathcal{D}(f\|h) - \mathcal{D}(f\|h_m^*) + \mathcal{D}(f\|f_m^*). \tag{26}$$

Since $\log h \in H_s(S^2, Q)$ with $s > 1$, we have that

$$\|\log h\|_\infty \le C \quad \text{and} \quad \|\log h_m^*\|_\infty \le C. \tag{27}$$

Observe that

$$\begin{aligned}
\left|\mathcal{D}(f\|h) - \mathcal{D}(f\|h_m^*)\right|^2 &= \left| \int_{S^2} f(\log h_m^* - \log h) \right|^2 \\
&\le \int_{S^2} f^2 \int_{S^2} \left(\log \frac{h}{h_m^*} \right)^2 \\
&\le C \int_{S^2} h \left(\log \frac{h}{h_m^*} \right)^2 \\
&\le C \exp(\|\log h - \log h_m^*\|_\infty) \mathcal{D}(h\|h_m^2) \\
&\le C\mathcal{D}(h\|h_m^2).
\end{aligned} \tag{28}$$

The second inequality comes from the Cauchy–Schwarz inequality, and the third from Lemma 1.[1] It follows from (26) and (28) that

$$\begin{aligned}
\left|\mathcal{D}(f\|h) - \mathcal{D}(f_m^*\|h_m^*)\right| &\le \left|\mathcal{D}(f\|h) - \mathcal{D}(f\|h_m^*)\right| + \mathcal{D}(f\|f_m^*) \\
&\le C_1 m^{-s} + C_2 m^{-2s}.
\end{aligned}$$

□

Acknowledgements

Research supported in part by NSERC (Canada) and by KOSEF (Korea) through The Statistical Research Center for Complex Systems at Seoul National University.

References

1. A. R. Barron and C. H. Sheu, Approximation of density functions by sequences of exponential families, *Ann. Statist.* **19** (1991) 1347–1369.
2. A. T. Brunger, Patterson correlation searches and refinement, *Macromolecular Crystallography, Part A, Methods in Enzymology* **276** (1997) 558–580.
3. V. I. Burdina, Symmetry of rotation function, *Soviet Phys. Crystallogr., USSR* **15** (1971) 545.
4. E. E. Castellano, G. Oliva and J. Navaza, Fast rigid-body refinement for molecular-replacement techniques, *J. Appl. Crystallogr.* **25** (1992) 281–284.
5. G. S. Chirikjian and D. Stein, Kinematic design and commutation of a spherical stepper motor, *IEEE/ASME Trans. Mechatronics* **4** (1999) 342–353.
6. G. S. Chirikjian and A. B. Kyatkin, *Engineering Applications of Noncommutative Harmonic Analysis* (CRC Press, Boca Raton, 2001).
7. R. A. Crowther, The fast rotation function, in *The Molecular Replacement Method*, ed. M. G. Rossmann (Gordon and Breach, New York, 1972), pp. 173–178.
8. R. A. Crowther and D. M. Blow, A method of positioning a known molecule in an unknown crystal structure, *Acta Crystallographica* **23** (1967) 544.
9. W. L. Delano and A. T. Brunger, The direct rotation function — rotational Patterson correlation search applied to molecular replacement, *Acta Crystallographica Section D-Biological Crystallography* **51** (1995) 740–748.
10. K. Diederichs, Structural superposition of proteins with unknown alignment and detection of topological similarity using a 6-dimensional search algorithm, *Proteins-Structure Funct. Genet.* **23** (1995) 187–195.
11. I. M. Gel'fand, R. A. Minlos and Z. Y. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Macmillan, New York, 1963).
12. S. Gwak, J. Kim and F. C. Park, Numerical optimization on the Euclidean group with applications to camera calibration, *IEEE Trans. Robotics Automation* **19** (2003) 65–74.
13. D. M. Healy, H. Hendriks and P. T. Kim, Spherical deconvolution, *J. Multivariate Anal.* **67** (1998) 1–22.
14. D. M. Healy and P. T. Kim, An empirical Bayes approach to directional data and efficient computation on the sphere, *Ann. Statist.* **24** (1996) 232–254.
15. M. Kim, G. S. Chirikjian and R. L. Jernigan, Elastic models of conformational transitions in macromolecules, *J. Mol. Graphics Modelling* **21** (2002) 151–160.
16. M. Kim, R. L. Jernigan and G. S. Chirikjian, Efficient generation of feasible pathways for protein conformational transitions, *Biophys. J.* **83** (2002) 1620–1630.
17. P. T. Kim and J. Y. Koo, Optimal spherical deconvolution, *J. Multivariate Anal.* **80** (2002) 21–42.
18. J. A. Kovacs and W. Wriggers, Fast rotational matching, *Acta Crystallographica* **D58** (2002) 1282–1286.
19. A. B. Kyatkin and G. S. Chirikjian, Algorithms for fast convolutions on motion groups, *Appl. Comput. Harmonic Anal.* **9** (2000) 220–241.
20. A. B. Kyatkin and G. S. Chirikjian, Pattern matching as a correlation on the discrete motion group, *Comput. Vision Image Understanding* **74** (1999) 22–35.

21. E. A. Lattman, Optimal sampling of rotation function, *Acta Crystallographica Section B-Structural Crystallography and Crystal Chemistry* **B28** (1972) 1065.
22. E. Lattman, Use of the rotation and translation functions, *Meth. Enzymol.* **115** (1985) 55–77.
23. C. Lemmen, C. Hiller and T. Lengauer, RigFit: A new approach to superimposing ligand molecules, *J. Computer-Aided Mol. Design* **12** (1998) 491–502.
24. J. T. Lo and L. R. Eshleman, Exponential Fourier densities and optimal estimation for axial processes, *IEEE Trans. Inf. Theory* **25** (1979) 463–470.
25. J. T. Lo and L. R. Eshleman, S^2 and optimal estimation and detection for directional processes, *IEEE Trans. Inf. Theory* **23** (1977) 321–336.
26. J. T. Lo and L. R. Eshleman, Exponential Fourier densities on $SO(3)$ and optimal estimation and detection for rotational processes, *SIAM J. Appl. Math.* **36** (1979) 73–82.
27. D. K. Maslen, Efficient computation of Fourier transforms on compact groups, *J. Fourier Anal. Appl.* **4** (1998) 19–52.
28. D. K. Maslen and D. N. Rockmore, Generalized FFTs — A survey of some recent results, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **28** (1997) 183–237.
29. J. Navaza, On the fast rotation function, *Acta Crystallographica Sec.* **A43** (1987) 645–653.
30. J. Navaza, AMoRe — An automated package for molecular replacement, *Acta Crystallographica Sec.* **A50** (1994) 157–163.
31. J. Navaza and P. Saludjian, AMoRe: An automated molecular replacement program package, *Macromolecular Crystallography, Part A, Methods in Enzymology* **276** (1997) 581–594.
32. F. C. Park, J. Kim and C. Kee, Geometric descent algorithms for attitude determination using the global positioning system, *J. Guidance Contr. Dynam.* **23** (2000) 26–33.
33. T. Risbo, Fourier transform summation of Legendre series and D-functions, *J. Geodesy* **70** (1996) 383–396.
34. M. G. Rossmann and D. M. Blow, Detection of sub-units within crystallographic asymmetric unit, *Acta Crystallographica* **15** (1962) 24.
35. D. Stein, E. R. Scheinerman and G. S. Chirikjian, Mathematical models of binary spherical-motion encoders, *IEEE/ASME Trans. Mechatronics* **8** (2003) 234–244.
36. L. Tong and M. G. Rossmann, Rotation function calculations with GLRF program, *Macromolecular Crystallography, Part A, Methods in Enzymology* **276** (1997) 594–611.
37. D. A. Varshalovich, A. N. Moskalev and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1998).
38. N. J. Vilenkin and A. U. Klimyk, *Representation of Lie Groups and Special Functions* Vols. 1–3 (Kluwer, Dordrecht, 1991).
39. T. O. Yeates, Statistics for rotation functions, *J. Appl. Crystallogr.* **26** (1993) 448–449.