Probabilities on Homographies

Evgeni Begelfor and Michael Werman *

Abstract

We present a family of "normal" distributions over a matrix group together with a simple method for estimating its parameters. In particular the mean of a set of elements can be calculated. The approach is applied to planar projective homographies, showing that using priors defined in this way improves object recognition.

1. Introduction and Outline

One reason to describe a distribution of homographies is to introduce a meaningful prior for Bayesian image recognition. Suppose we want to distinguish between N planar objects. The input we are given is a picture of an object imaged from an unknown direction. Thus the image, D should be obtained from the original by some homography ϕ. Let M be the model, in other words, one of the objects, from Bayes rule we get:

$$P(M,\phi|D) = \frac{P(D|M,\phi)P(M,\phi)}{P(D)} =$$

$$= \frac{P(D|M,\phi)P(\phi|M)P(M)}{P(D)} \qquad (1)$$

Thus to be fully Bayesian, we need to know $P(\phi|M)$, that is, a distribution on the group of the homographies of a plane.

Many papers such as [13], [5], deal with means on groups, although all of them are for subgroups of the group of euclidean motions. In [3], [4] a method for putting a distribution on a Lie group is described. The method is appropriate, though, only when the group is compact (such as $SO(3)$, the group of rotations of the 3-D space) or abelian (such as $\mathbb{R}^n$, the group of translations) or direct products of such groups. However, some models involve groups that cannot be represented in such a way, in particular the views of a planar object from different directions. These are modeled with the group of plane homographies, that is, 3-by-3 matrices where matrices differing only by scalar multiplication represent the same homography. To get rid of this ambiguity we normalize the matrices to have determinant 1. This group is usually denoted by $SL_3(\mathbb{R})$. Distributions on such groups were treated in [11], but no method for estimating expectation and the parameters of the distribution was described, and the absence of the covariance greatly diminishes the ability of the distribution to fit data. In this paper we propose a parametric distribution on such groups, together with simple methods for finding the parameters.

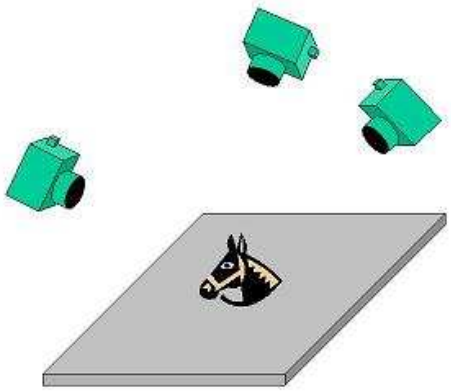

Figure 1. A planar object imaged from different directions.

The main idea needed in order to define the distribution is as in [3] for the orthogonal group, the geodesic distance on the group. The geodesic distance is used in order to define a mapping from the group G to a linear space where we estimate the parameters of the normal distribution. Another way to look at it is to say that we define an invariant distribution on the group and learn its parameters.

The next Section shows the need to define the probability using the group structure. In Section 3 the mathematical background needed is described. Section 4 describes the actual algorithms for estimating the parameters of the distribution fitted to given data. The paper finishes with a

*E. Begelfor and M. Werman are with the Computer Science department of The Hebrew University of Jerusalem.

demonstration of the methods applied to object recognition.

2. Motivation

Intuitively, a "normal" distribution on G should have a mean value μ and a covariance matrix Σ. In the case of the usual normal distributions on $\mathbb{R}^n$, we know that if a random variable X is translated by t then the probability translates

$$X \backsim N(\mu, \Sigma) \qquad \Longrightarrow \qquad (X + t) \backsim N(\mu + t, \Sigma).$$

We would like a similar property to hold for our distributions on G. For example, suppose that we have a planar object and a distribution of camera positions above it (Fig 1). If the distribution of the homographies from a set of images I to Image 1 is $N(\mu, \Sigma)$, the distributions of the homographies from the set of images I to Image 2 is $g \circ N(\mu, \Sigma)$, where g is the homography between Image 1 and Image 2. We would like the parameters to be invariant to the group action, that is

$$h \backsim N(\mu, \Sigma) \qquad \Longrightarrow \qquad (gh) \backsim N(g \circ \mu, \Sigma).$$

One might try to define a distribution on the group G, for example $SL_3(\mathbb{R})$ by treating it as a subset of $\mathbb{R}^9$. There are a few problems with this approach. First of all, $SL_3(\mathbb{R})$ is an 8-dimensional manifold and not 9-dimensional. One might take only 8 coefficients of the 3×3 matrix, and define the distribution using those. But in this way the invariance property doesn't hold. The solution is to define the distribution using intrinsic features of the group G. We define the distribution with a given mean μ by mapping a normal distribution on the tangent space at μ to the group itself, while keeping the invariance properties.

In Figure 2 the advantage of our distribution is demonstrated. A set of homographies between a planar object and its image when the camera is randomly placed on a sphere above the plane. The dashed line is the density of one of the coefficients of the usual matrix representation of a homography (as a 3-by-3 matrix normalized to have determinant 1) and the solid line is the density of the corresponding coefficient after our transformation was applied. It can be seen that our distribution is more "normal" and informative. The distribution of the coefficients after transformation is much closer to normal, as is demonstrated by the results of Kolmogorov-Smirnov [7] tests of normality of the coefficients shown in Figure 2 (smaller numbers imply more normality, the numbers were scaled).

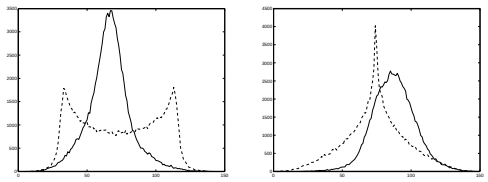

Figure 2. Distributions of two coefficients of the homography matrices in usual matrix representation (dashed line) and after transformation (solid line). Kolmogorov-Smirnov normality test of the different coefficients is shown under each graph.

3. Mathematical background

The tools used here come primely from Lie theory and differential geometry. For more information on these subjects the reader is referred to [14] or [8].

A Lie group G is a group which is also a smooth manifold, such that multiplication and inversion are smooth. For any point x on a smooth manifold one has the tangent space to the manifold at x, denoted by T_x.

Many of the examples of Lie groups are matrix groups, for example $G = SL_n(\mathbb{R})$, the set of all n-by-n matrices with real entries and determinant 1. This set has a manifold structure inherited from the natural manifold structure of $\mathbb{R}^{n^2}$, the set of all matrices. Every matrix in G has an inverse in G, and as the determinant is multiplicative, the product of two matrices in G is in G. Thus $SL_n(\mathbb{R})$ is a Lie group.

As with any smooth manifold and any point on it, if we have a Lie group G one has the tangent space to the identity element e_G which we will denote by $\mathfrak{g}$ [1] . There exists a map, called the *exponential map*, $\exp : \mathfrak{g} \to G$ such that for any $v \in \mathfrak{g}$ and any two real numbers t, s, we have $\exp(tv + sv) = \exp(tv)\exp(sv)$. Moreover, $\frac{d}{dt}\exp(tv)|_0 = v$.

Let's look again at $G = SL_n(\mathbb{R})$. The tangent space to I (the identity matrix) is a subspace of the space of all matrices. In this case, as with all matrix groups we get that the exponent map is (hence the name) $\exp(X) = e^X = \sum_{k=0}^{\infty} \frac{X^k}{k!}$

So what is $\mathfrak{g}$ for $SL_n(\mathbb{R})$? We should take all the matrices X such that $det(e^X) = 1$. As is known, $det(e^X) = e^{tr(X)}$ thus the condition is that $tr(X) = 0$. We denote the set of all n-by-n matrices with zero trace by $sl_n(\mathbb{R})$.

As the exponential map isn't onto and 1:1 in general, the inverse map, $\log$, can be defined only in a certain neighborhood of the identity. In the case of matrix group, $\log(g) = \sum_{k=1}^{\infty} \frac{-1^k}{k}(g - I)^k$.

[1] The reason we pick a special name, $\mathfrak{g}$ for the tangent space which we could denote by T_e is that this space, called the Lie algebra of G, plays an important role in Lie theory, and proofs of some of following claims use Lie algebra. We chose to omit the definition of Lie algebras and their multiplication in order to keep this exposition as simple as possible.

Next, we want to put a metric structure on our group G which we will use to define the distribution. A general way to define metrics on a manifold G (not neceserily a group) is to assign to every point $g \in G$ an inner product $<,>_g$ on the tangent space T_g at g, enabling us to measure the length of a tangent vector. We denote the norm on T_g derived from $<,>_g$ by $\|\cdot\|_g$.

$$\|v\|_g = \sqrt{<v, v>_g}$$

Now we can define length of a path $\gamma : [a, b] \to G$ as follows

$$L(\gamma) = \int_a^b \|\gamma(t)'\|_{\gamma(t)} dt$$

when $\gamma(t)'$ is the derivative of $\gamma(t)$. The distance $d(g_1, g_2)$ between two points is the infimum of the lengths of paths between them, and a path where the infimum is achieved is called a geodesic. While it is not true in general, in a generic situation for every point g in G there is a unique geodesic starting from g in every direction, giving us the *exponential* map (not to be confused with the exponential map above, although sometimes they coincide) $\exp_g : T_g \to G$ such that $d(g, exp_g(v)) = \|v\|_g$ for every v in T_g. In what follows we denote the map exp_e, the Riemannian exponential map from the tangent space at the identity element of G, by $expp$.

For the family of distributions to be left-invariant, the metric has to be left-invariant. Multiplication by g defines a map $L_g : G \to G$. L_g maps e to g, and thus maps T_e to T_g. It also carries the inner product from T_e to T_g in the following way: let L_g^{-1} be the inverse map from T_g to T_e. Now the inner product defined by

$$< v, u >=< L_g^{-1}v, L_g^{-1}u >_e \quad v, u \in T_g$$

is $L_g(<,>_e)$. For the Riemannian structure to be invariant we need to have $L_g(<,>_e) =<,>_g$. We conclude that an invariant Riemannian structure on a group is determined by an inner product on T_e.

Let us demonstrate the principle with some examples. For G=$\mathbb{R}^n$, the identity element is 0. In this case we can take $<,>_0$ to be the standard scalar product. An action on the left (this group is commutative, thus it doesn't matter) by an element g of $\mathbb{R}^n$ translates the whole group, so we get that $<,>_g$= $L_g(<,>)$ is again the standard scalar product. The length of a path now is the usual length in $\mathbb{R}^n$, so geodesics are straight lines. Thus the invariant metric we get on $\mathbb{R}^n$ is the usual Euclidian metric.

The following discussion enables us to deal with the special linear group $SL_n(\mathbb{R})$ and the orthogonal group $SO_n(\mathbb{R})$. [2] We choose a certain inner product on $\mathfrak{g}$: [3]

$$< A, B >= Tr(AB^T)$$

As we'll see, the distribution we get from this definition coincides with that of [3] for the case $G = SO_n$, the group of orthogonal matrices. Additionally, we still obtain a closed form for the geodesics (the shortest paths) in this metric. It turns out that the Riemannian exponential map in this case is [15]

$$expp(X) = e^{-X^T} e^{X+X^T}$$

The map $expp$ is onto, although not 1:1 so to define the inverse map, $logg$ we need to choose X with the smallest $\|X\|$ such that $expp(X) = M$. For the case $G = SO_n(\mathbb{R})$, the group of orthogonal matrices, we get $\mathfrak{g} = so_n(\mathbb{R})$ which is the n-by-n antisymmetric matrices with zero trace. It follows that for every $X \in \mathfrak{g}$ we have $X^T = -X$, yielding

$$expp(X) = e^{-X^T} e^{X+X^T} = e^X$$

and respectivly, $logg(g) = \log(g)$, giving the same as in [3]. Now we make the step from metric to distributions. Recall that the usual normal distribution in $\mathbb{R}^n$ with mean μ and covariance Σ has density

$$\phi(x) \propto e^{-(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

When trying to mimic that distribution in the case of Lie groups, we have a small technical complication, namely, to define the covariance we need to look at points in $\mathfrak{g}$ as vectors, not matrices, by picking some basis $v_1, \ldots, v_m$ for $\mathfrak{g}$ and taking f to be the map from $\mathfrak{g}$ to $\mathbb{R}^m$ giving the coefficients according to that basis. This done, we define "Lie-normal" [4] distribution on G with mean $\mu \in G$ and covariance Σ (a m-by-m matrix) as having the density

$$\phi(g) \propto e^{-f(logg(\mu^{-1}g))^T \Sigma^{-1} f(logg(\mu^{-1}g))}$$

and this family of distributions is left-invariant by construction.

4. Algorithms

The algorithms described in this Section follow the algorithms of [1] differing in the substitution of $expp$ and $logg$ in place of $\exp$ and $\log$.

Our goal is to find a Lie-normal distribution on G that fits the data. First we should estimate the expectation of

[2] This discussion is true in general for semi-simple Lie groups of which these two are examples

[3] This inner product is invariant under orthogonal transformations: $< OA, OB >= tr(OAB^T O^T) = tr(O^T OAB^T) = tr(AB^T) = =< A, B >$. This definition is also very natural in the setting of semi-simple groups, see [15].

[4] Strictly speaking, we have no right to call these distributions *normal* as a notion of truly, that is, fitting the central limit theorem normal distributions on groups exists [6], although, these distributions aren't feasible for computational purposes.

the distribution, by finding the "mean" of the data. In this Section we define the mean on a Lie group. The notion of a center of mass on Riemannian manifold was intoduced by Cartan, see [9], ch. 8 and [12] for further information. Remember from the previous Section that the distance in G is given by

$$d(g,h) = \|logg(g^{-1}h)\|$$

Thus the mean of $g_1, ...g_n \in G$ is [9]

$$\mu = \arg\min_{h\in G} \sum_1^n d(g_i,h)^2 =$$

$$= \arg\min_{h\in G} \sum_1^n \|logg(g_i^{-1}h)\|^2$$

The first order approximation to the mean is given by

$$\hat{\mu} = expp(\frac{1}{n}\sum_{i=1}^n logg(g_i))$$

The error in this approximation is larger when points are far from the identity. Thus we left-multiply all points by $\hat{\mu}^{-1}$ so that $\hat{\mu}$ is moved to identity. Now we compute the mean of these residual points and combine this with $\hat{\mu}$ to get a new approximation of the mean. This process is repeated until the mean of the residuals is sufficiently near the identity.

Algorithm 1:
Input: $g_1, \ldots, g_n$
Output: $\mu \in G$, the mean
$\mu = g_1$
Repeat
$\Delta g_i = \mu^{-1} g_i$
$\Delta\mu = expp(\frac{1}{n}\sum_{i=1}^n logg(\Delta g_i))$
$\mu = \mu\Delta\mu$
Until $\|logg(\Delta\mu)\| < \epsilon$

As this is a gradient decent method, it will converge to a local minima. The uniqueness cannot be guaranteed in general, for example there is no unique mean between the north and the south pole of a sphere. In the Appendix we show that for close enough homographies the computation converges to the global minima.

Now we want to find the covariance matrix. First we find the mean μ of the data using the algorithm above, then map the data to the tangent space of μ and finally compute the covariance matrix. Pick any basis $\{v_1, v_2, \ldots, v_m\}$ for $\mathfrak{g}$ and let $f : \mathfrak{g} \to \mathbb{R}^m$ be a map that takes $X \in \mathfrak{g}$ to its representation according to the basis as in the previous Section. Now we can present the algorithm:

Algorithm 2:
Input: $g_1, \ldots, g_n \in G$
Output: mean μ and covariance matrix Σ
$\mu =$ mean of $\{g_i\}$
$x_i = logg(\mu^{-1} g_i)$
$\Sigma_{kl} = \frac{1}{n}\sum_{i=1}^n f_k(x_i) f_l(x_i)$

Any basis for $\mathfrak{g}$ will do. For $G = SL_n(\mathbb{R})$ we can take, for example, the following basis:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

leading to the following f:

$$f(\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}) =$$

$$= (x_{11} - x_{22}, x_{22} - x_{33}, x_{12}, x_{13}, x_{21}, x_{23}, x_{31}, x_{32})$$

5. Results

In this Section the results of applying our methods are demonstrated. The computations of $logg$ and $expp$ were done with Matlab.

In Figure 4 we see a set of images of a wallet with the average view chosen by the algorithm highlighted.

We chose a contrived, but nontrivial example to test our new distribution. For the object recognition task we took as the three models the same cutter but opened to three different angles. Each one of the models was imaged from different directions. In Figure 5 sample pictures of each model is shown in a different row. The goal was to recognize the model, opening angle, from a single image. A training set of 100 images for each angle was given and Algorithm 1 was used to find the "average" view of each class. The distribution parameters for both the usual normal distribution (see Section 1) and Lie-normal (Section 3)distribution were

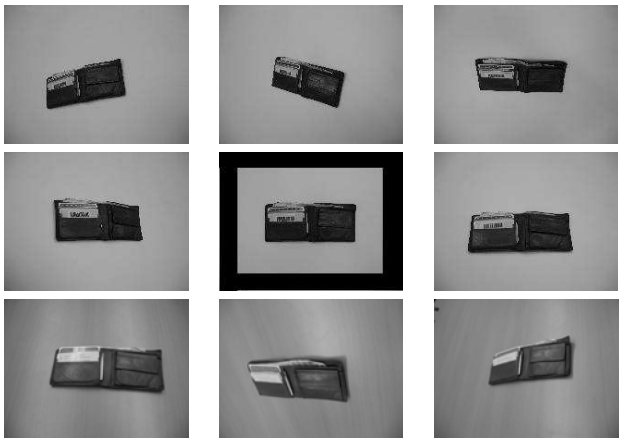

Figure 3. Different views of a wallet with the mean view in the middle.

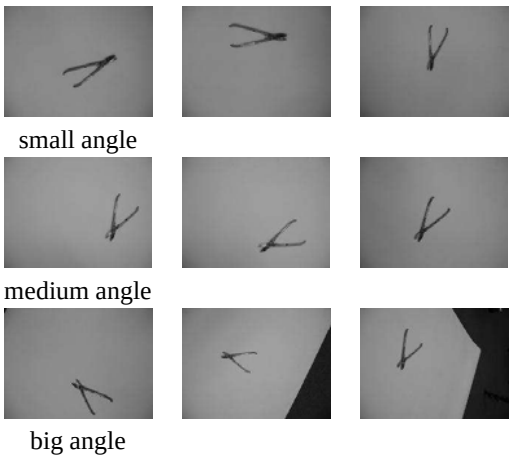

Figure 4. Each row shows a few images with the same opening angle.

computed. The test set contained 60 images, different from the training set (20 for each opening angle). The recognition stage was implemented in the following way: for a given image, the best homographies to the three average views were computed, using Lowe's [10] feature detector, with the homography between points of interest computed by the RANSAC algorithm [2].

The different scoring methods were: choosing the model with the smallest fitting error obtained by a homography, (thus using only the $P(D|M,\phi)$ term in Eq.[1]), choosing the object with the homography having the highest probability (using the $P(\phi|M)$ term in Eq.[1]) and combining the two, thus using the full Bayes formula. The results in the table below show that the fitting error performs poorly (in fact, it is not better than random) which is due to the fact that homography has many degrees of freedom and can align images well even if they are not images of the same object. Combining fitting error with the usual distribution is better, but as the results show, the Lie-normal distribution combined with fitting error outperforms the other methods.

Method	success rate
Fitting error	0.34
Normal distribution only	0.3
Lie-normal distribution only	0.47
Normal dis. and fitting error	0.44
Lie-normal dis. and fitting error	0.54

6. Summary

We proposed a new family of probability distributions on the group of homographies. The advantage of this new approach is the invariance of the family, thus these distributions are more suitable to describe group-invariant distributions arising in computer vision. The parameters of these distributions are easily estimated and the density simply computed. The advantage of the method as opposed to ad hoc approach is demonstrated in a toy example of object recognition.

For a set contained in a convex ball (every two points in the ball have a unique geodesic between them that is contained in the ball) the uniqueness of the mean and the convergence of Algorithm 1 follows from [1], p.160. We show that a ball with radius $\frac{1}{5}$ in the Frobenius norm ($\|X\| = \sqrt{tr(XX^T)}$, which is the euclidian norm in $\mathbb{R}^{n^2}$) around I in $G = SL_n(\mathbb{R})$ sits in a convex ball in the sense of Section 4, thus the mean of every set contained in it is uniquely defined and reached by the algorithm. In what follows d will be the invariant metric on the group as defined above.

Lemma 1 *The map $expp$ is one-to-one on $\{X \in sl_n(\mathbb{R}), \|X\| < \frac{1}{2}\}$*

Suppose $expp(X)$=$expp(Y)$, then

$$e^{-X^T}e^{X+X^T} = e^{-Y^T}e^{Y+Y^T}$$

$$e^{-X^T}e^{X+X^T}e^{-Y^T-Y}e^{Y^T} = I$$

$$\log(e^{-X^T}e^{X+X^T}e^{-Y^T-Y}e^{Y^T}) = 2\pi ikI \quad k \in \mathbb{Z}$$

By CBH formula [14]

$$X + X^T - X^T - Y^T - Y + Y^T + O(\|X\|, \|Y\|)(X - Y)$$

$$= 2\pi ikI \quad k \in \mathbb{Z}$$

with the constant in $O()$ term smaller than 1. As $\|X\|, \|Y\| < \frac{1}{2}$ we see that k must be zero and we have

$$\|X - Y\| < \|X - Y\| \max(\|X\|, \|Y\|)$$

but then $X = Y$ and the injectivity is proved.

Lemma 2 *The Sectional curvature of $SL_n(\mathbb{R})$ with the Riemannian structure as above is bounded by 24.*

This bound is extremely loose, but it's not the bottleneck of our calculation, so it suffices. The norm is submultiplicative, thus for any $X, Y \in sl_n(\mathbb{R})$ we have

$$\|[X, Y]\| \leq 2\|X\|\|Y\|$$

By [8],p. 277 the formula for the affine connection on our manifold is

$$\nabla_X Y = \frac{1}{2}([X, Y] + [X^T, Y] + [Y^T, X])$$

and by the previous equation

$$\|\nabla_X Y\| \leq 3\|X\|\|Y\|$$

Now the Sectional curvature in the plane defined by X, Y is

$$\sigma(X,Y) = \frac{< \nabla_Y \nabla_X X - \nabla_X \nabla_Y X + \nabla_{[X,Y]} X, Y >}{\|X\|^2 \|Y\|^2 - < X, Y >^2}$$

Without lost of generality assume $\|X\| = \|Y\| = 1$ and $< X, Y >= 0$. We obtain the claimed bound on the curvature:

$$|\sigma(X,Y)| \leq 3 \cdot 3 + 3 \cdot 3 + 3 \cdot 2 = 24$$

By [1], if the Sectional curvature of a Riemannian manifold is bounded from above by K and the injectivity radius is bigger than R then

$$ConRad \geq \min\{\frac{1}{2}R, \frac{1}{2}\frac{\pi}{\sqrt{K}}\}$$

Thus we have $ConRad > \frac{1}{4}$, that is, the convergence of the algorithm is assured on the ball of radius $\frac{1}{4}$ in the invariant metric. To see what it means in the usual euclidian metric in $SL_n(\mathbb{R})$

Let $\gamma(t)$ be a path in R^m. If $\|\gamma''(t)\| < C$ for $0 \leq t \leq 1$ then $\|\gamma(1) - \gamma(0)\| \geq \|\gamma'(0)\| - \frac{1}{2}C$

$$\|\gamma(1) - \gamma(0)\| \geq \frac{< \gamma(1) - \gamma(0), \gamma'(0) >}{\|\gamma'(0)\|}$$

$$< \gamma(1) - \gamma(0), \gamma'(0) >= \int_0^1 < \gamma'(t), \gamma'(0) > dt =$$

$$= \int_0^1 < \int_0^t \gamma''(s) ds + \gamma'(0), \gamma'(0) > dt =$$

$$=< \gamma'(0), \gamma'(0) > + \int_0^1 \int_0^s < \gamma''(s), \gamma'(0) >\geq$$

$$\geq < \gamma'(0), \gamma'(0) > -\frac{1}{2}C\|\gamma'(0)\|$$

Collecting the pieces:

$$\|\gamma(1) - \gamma(0)\| \geq \|\gamma'(0)\| - \frac{1}{2}C$$

If we take $\gamma(t) = expp(tX)$ for $\|X\| = \frac{1}{4}$ we can easily bound the second derivative from above by $\frac{1}{10}$, thus using the last claim we obtain: if $\|G - I\| < \frac{1}{5}$ then $d(G, I) < \frac{1}{4}$. In other words, the euclidian ball of radius $\frac{1}{5}$ sits in a convex ball in the Riemannian metric.

References

[1] M.Berger, *A Panoramic View of Riemannian Geometry*, Berlin : Springer, 2003

[2] M.A.Fischler and R.C.Bolles. "Random sampling consensus: A paradigm for model fitting with applications to image analysis and automated cartography." *Comm. Assoc. Comp. Mach.*, 24(6) pp. 381:395,1981

[3] P. Fletcher, C. Lu and S. Joshi, "Statistics of Shape via Principal Geodesic Analysis on Lie Groups" *Proc. IEEE CVPR 2003* vol. 1, pp. 95-101

[4] P. Fletcher, C. Lu, S. Joshi and S.Pizer, "Gaussian Distributions on Lie Groups and Their Application to Statistical Shape Analysis" *Proc. IPMI 2003* LNCS 2732 pp. 450-462

[5] V. M. Govindu, "Lie-Algebraic Averaging for Globally Consistent Motion Estimation" *Proc. IEEE CVPR 2004,* vol. 1, pp. 684-691

[6] U.Grenander,*Probabilities on Algebraic Structures*, New York : Wiley, 1963

[7] W. Hays and R. Winkler, *Statistics: probability, inference, and decision*, New York : Holt, Rinehart and Winston, 1970

[8] S. Helgason *Differential geometry, Lie groups, and Symmetric spaces*, Academic Press, 1978.

[9] P. Buser and H. Karcher, *Gromov's almost flat manifolds*, Asterisque, 1981

[10] David G. Lowe, "Distinctive image features from scale - invariant keypoints," *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110

[11] E.Miller and C. Chef'dhotel, "Practical Non-Parametric Density Estimation on a Transformation Group for Vision." *Proc. IEEE CVPR 2003,* vol. 2, pp. 114-121

[12] X. Pennec. "Probabilities and Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements." *Proc. of Nonlinear Signal and Image Processing 99*, pp. 194-198

[13] R.Rao, D. Ruderman, "Learning Lie Groups for Invariant Visual Perception" *(Advances in NIPS 11,* pp. 810-816, 1999

[14] V.S. Varadarajan, *Lie groups, Lie algebras, and their representations*, Englewood Cliffs, N.J. : Prentice-Hall, 1974

[15] H.C. Wang, "Discrete Nilpotent Subgroups of Lie Groups", *J. Differential Geometry*, vol. 3, pp. 481-492, 1969