Proceedings of the 36th
Conference on Decision & Control
San Diego, California USA • December 1997

Singular Values and Least Squares Matching

R. W. Brockett [1]

Abstract

In this paper we investigate least squares matching problems, extending the methods of our earlier paper [1] in such a way as to make them applicable to problems involving sets of points that are so large that approximate answers are of interest. These problems are formulated in terms of continuous descent equations, and lower bounds on the quality of the best match are obtained in terms of the singular values of certain matrices determined directly by the data.

1 Introduction

Many important algorithms in computer vision involve matching points in one image, or one part of an image, with points in a second image or second part of the given image. Matching is often difficult because of the large number of possibilities. For this reason, even suboptimal solutions or solutions that are only good in some probabilistic sense are of interest. We begin with an example of the type of problem to be treated.

Example 1: Suppose we have two sets of k points in R^n and we want to act on the first set by an orthogonal transformation, so it matches the second set as closely as possible in a squared error sense. We are not told a priori which point in the first set should match which point in the second set. If the points of the first set are denoted by $\{x_i\}$ and those in the second set by $\{y_i\}$, we have the problem of finding an n by n orthogonal matrix Θ and permutation π acting on a set of k objects such that the sum

$$\eta = \sum ||\Theta x_{\pi(i)} - y_i||^2$$

is as small as possible. For a fixed choice of π the determination of Θ is not too difficult. We can rewrite the expression for η in terms of the matrices

$$X = \begin{bmatrix} | & | & .. & | \\ x_1 & x_2 & .. & x_k \\ | & | & .. & | \end{bmatrix}$$

$$Y = \begin{bmatrix} | & | & .. & | \\ y_1 & y_2 & .. & y_k \\ | & | & .. & | \end{bmatrix}$$

in the following way

$$\eta = \mathrm{tr}(\Theta X\Pi - Y)^T(\Theta X\Pi - Y)$$

with Π being a k by k permutation matrix representing the interchanges of the columns of X achieved by π. This can, in turn, be expanded to give

$$\eta = ||X||^2 + ||Y||^2 - 2\mathrm{tr}\Theta X\Pi Y^T$$

If Π were known this would be the problem of maximizing the trace of $\Theta X\Pi Y^T$ with respect to the choice of an orthogonal matrix Θ. As is well known, this is accomplished by choosing Θ so that $\Theta X\Pi Y^T$ is symmetric and positive semidefinite. Denoting the symmetric positive semidefinite square root of a positive semidefinite matrix by the usual square root symbol, we have

$$\Theta_{opt} = X\Pi Y\sqrt{(X\Pi Y)^T(X\Pi Y)}^{-1}$$

$$\eta_{opt} = \mathrm{tr}\sqrt{(X\Pi Y)^T(X\Pi Y)}$$

On the other hand, repeating this for all possible permutations is not practical for large values of k.

Example 2: Consider a collection of p line segments in R^n. The endpoints of these line segments define p pairs of points in R^n. We denote these pairs of points by $\{(a_i, b_i)\}_{i=1}^p$. We would like to find an orthogonal matrix such that the pairs $\{(\Theta a_i, \Theta b_i)\}_{i=1}^p$ match with a second "template" defined by the pairs of points $\{(s_i, t_i)\}_{i=1}^p$. If we are not given an orientation for the line segments then the pair (a_i, b_i) is equivalent to the pair (b_i, a_i). One can associate the individual pairs with the individual pairs of the template in $p!$ different ways and having done so, one can associate the points making up the pairs with the points of the templates in 2^p different ways. To express this problem in a form that can be compared with the first example we define a subgroup of the set of all permutation matrices of size $2p$ by $2p$ which consists of those matrices that can be expressed as

$$\Pi_{T^n} = \begin{bmatrix} P_{11} & 0 & .. & .. \\ 0 & P_{22} & 0 & .. \\ .. & .. & .. & .. \end{bmatrix}$$

[1] This work supported in part by Army DAAH 04-93 -G- 0330, UMD-SRC NSF EEC 94-02384, Brown Army DAAH 04-96-1-0445, Draper Learning, Airforce F33615-95-C-1759, JSEP, Navy N00014-89-J-1023 , CDCSS Army DAAG 55 97 1 0114, Draper Modeling,C.S. Draper Laboratories, DLH 496022, MIT Army DAAL03-92-G-0115

with the P_{ii} being two by two permutation matrices. When a matrix X having $2p$ columns is multiplied on the right by such a matrix, the effect is to either leave unchanged a given odd-numbered column or else to interchange it with the following even-numbered column. We also define a second subgroup of the set of all $2p$ by $2p$ permutation matrices in terms of the two by two identity matrix I_2. It consists of all permutation matrices of the form

$$\Pi_{I_2} = \begin{bmatrix} p_{11}I_2 & p_{12}I_2 & .. & .. \\ p_{21}I_2 & p_{22}I_2 & .. & .. \\ .. & .. & .. & .. \\ .. & .. & .. & .. \end{bmatrix}$$

Recast in the language of a group actions on sets, we can say that the group $(\Pi_2)^p \times \Pi_p$, where $(\Pi_2)^p$ is the product of p copies of the two element permutation group, acts on the data to preserve or reverse orientation and renumber the line segments. In terms of equations, we can express the matching problem as that of minimizing the function

$$\eta = \mathrm{tr}(\Theta X\Pi - Y)^T(\Theta X\Pi - Y)$$

with Π being expressible as the product $\Pi = \Pi_1\Pi_2$ with the first factor in Π_{T^n} and the second in Π_{I_2}. The expression for η again simplifies to

$$\eta = ||X||^2 + ||Y||^2 - 2\mathrm{tr}\Theta X\Pi Y$$

but now the group of permutations that are permitted is not the whole set of permutations that might act on the columns of X but rather only the subgroup identified above.

2 Generalities on Matching

In the next section we will establish certain lemmas that will let us link the quality of the best match available to the singular values of the rectangular matrices X and Y which appear in the statement of example one above. Recall that the singular values of a rectangular matrix, say X, are the positive square roots of the nonzero eigenvalues of either the symmetric matrix X^TX or (equivalently) the symmetric matrix XX^T. We denote these by $\mu_i(X)$. That the quality of the match and the singular values of X and Y are related is, perhaps, surprising. However, it is clear from the definition that the singular values of X and $\Theta X\Psi$ for Θ and Ψ orthogonal are the same and so the singular values have a suitable invariance.

As in reference [1], we will replace the combinatorial search involved in finding the best permutation with a calculus problem involving orthogonal matrices. However, the problems under consideration here are different and there are very significant differences between what is done in [1] and what is proposed here. The function of Θ given by $\mathrm{tr}\Theta Q\Theta^T N$, which plays a critical role in [1], is replaced here by the function of two orthogonal matrices Θ and Ψ given by $\mathrm{tr}\Theta X\Psi Y$. Only the permutation matrices corresponding to even permutations have positive determinants and belong to the set $So(n)$ of proper (determinant = +1) orthogonal matrices. The permutation matrices corresponding to odd permutations have negative determinants and so in most places we deal with the entire orthogonal group $O(n)$.

Theorem: Consider the matching problem of example one. Let $\mu_1(X) \geq \mu_2(X) \geq ...\mu_r(X)$ be the singular values of X, and let $\mu_1(Y) \geq \mu_2(Y) \geq ...\mu_r(Y)$ be the singular values of Y. The total error associated with the matching problem is lower bounded by the sum of the squares of the differences of the singular values of X and Y. That is

$$\eta \geq \sum(\mu_i(X) - \mu_i(Y))^2$$

This lower bound is achievable for particular data sets but in general it underestimates the error.

For example two, a different type of bound is available based on an averaging technique. We begin by forming an n by p matrix of centroids of the pairs

$$x_i = \frac{1}{2}(a_i + b_i) \ ; \ y_i = \frac{1}{2}(s_i + t_i)$$

Because the sum of the squares of the errors in matching the endpoints is at least twice as great as the square of the error associated with matching the centroid, we see that the above result implies the following.

Corollary: Consider the matching problem of example two. Define the centroids as above. Let $\mu_1(X) \geq \mu_2(X) \geq ...\mu_r(X)$ be the singular values of X and let $\mu_1(Y) \geq \mu_2(Y) \geq ...\mu_r(Y)$ be the singular values of Y. The total error associated with the cluster matching problem is lower bounded by twice the sum of the squares of the differences of the singular values of X and Y. That is,

$$\eta \geq 2\sum(\mu_i(X) - \mu_i(Y))^2$$

3 Some Analysis

Our approach to establishing theorem one is based on the fact that a permutation matrix is necessarily orthogonal. Instead of searching over all the permutation matrices, we search over the set of all orthogonal matrices. The advantage of enlarging the search space is that the methods of calculus are now available and the problem becomes more manageable.

Lemma 1: Given two real n by m matrices X and Y, let $\mu_1(X) \geq \mu_2(X) \geq ...\mu_r(X)$ and $\mu_1(Y) \geq \mu_2(Y) \geq ...\mu_s(Y)$ be the respective singular values. Then the value of the function

$$\phi(\Theta, \Psi) = \mathrm{tr}\Theta X \Psi Y^T$$

as Θ and Ψ range over the set of n by n and m by m orthogonal matrices, respectively, is bounded by

$$\mathrm{tr}\Theta X \Psi Y \leq \sum \mu_i \nu_i$$

The stationary values of ϕ occur when $\Theta X \Psi Y^T$ and $\Psi Y^T \Theta X$ are symmetric matrices. If the singular values of X and Y are distinct, there are $2^m m!$ distinct values of (Θ, Ψ) at which the upper bound is achieved.

Proof: We begin by replacing X and Y^T by their singular value decompositions. Our notation is

$$X = \Theta_0 D_1 \Theta_1$$

and

$$Y^T = \Psi_0 D_2 \Psi_1$$

In terms of this notation

$$\phi(\Theta, \Psi) = \mathrm{tr}\Theta\Theta_0 D_1 \Theta_1 \Psi \Psi_0 D_2 \Psi_1$$

Using the cyclic property of the trace, we rewrite this as

$$\phi(\Theta, \Psi) = \mathrm{tr}\tilde{\Theta} D_1 \tilde{\Psi} D_2$$

with $\tilde{\Theta} = \Psi_1 \Theta \Theta_0$ and $\tilde{\Psi} = \Theta_1 \Psi \Psi_0$. It is well known that for a given M, the maximum value of $\mathrm{tr}\Theta M$ with respect to a choice of orthogonal matrix Θ is obtained by choosing Θ in such a way as to make ΘM symmetric and positive semidefinite. The maximum value is, by the polar representation, simply the sum of the singular values of M. Thus we see that the stationary values of ϕ occur when both $\tilde{\Theta} D_1 \tilde{\Psi} D_2$ and $\tilde{\Psi} D_2 \tilde{\Theta} D_1$ are symmetric. This last condition also implies that $D_2 \tilde{\Theta} D_1 \tilde{\Psi}$ is symmetric. However, it is easy to see that if D_2 is diagonal with distinct eigenvalues and both $D_2 A$ and $A D_2$ are symmetric, then A must be diagonal. Thus we see that ϕ is stationary only when $\tilde{\Theta} D_1 \tilde{\Psi}$ is diagonal. By pre-multiplying $\tilde{\Theta} D_1 \tilde{\Psi}$ by its transpose we get

$$\tilde{\Psi}^T D_1 \tilde{\Theta}^T \tilde{\Theta} D_1 \tilde{\Psi} = \tilde{\Psi}^T D_1^2 \tilde{\Psi} = \Pi D_1^2 \Pi^T$$

with Π being a permutation matrix. This implies that $\tilde{\Psi}$ can be expressed as the product of a diagonal orthogonal matrix (there are 2^m such matrices) and a permutation matrix (there are $m!$ of these).

We wish to describe a descent algorithm for minimizing functions of the type appearing in the previous lemma. In order to define a gradient flow on a manifold, in this case a manifold consisting of the product of orthogonal groups $O(n)$ and $O(m)$, we need a choice of Riemannian metric. Here we use a scaled version of the standard metric on $O(n)$ defined by

$$ds^2 = \langle d\Theta\Theta^T, d\Theta\Theta^T \rangle$$

Lemma 2: Assuming the standard metric on the orthogonal group, the gradient flow for the function $\phi : O(n) \times O(n) \to R$ defined by $\phi(\Theta, \Psi) = \mathrm{tr}\Theta X \Psi Y$ is given by

$$\dot{\Theta} = (\Theta X \Psi Y^T - Y \psi^T X^T \Theta^T)\Theta$$

$$\dot{\Psi} = (\Psi Y^T \Theta X - X^T \Theta^T Y \Psi^T)\Psi$$

Expressed in terms of the variables $T = \Theta X$ and $V = \Psi Y^T$, these equations take the form

$$\dot{T} = (TV - V^T T^T)T$$

$$\dot{V} = (VT - T^T V^T)V$$

Proof: The expression for the gradient flow follows from the Taylor series expansion of

$$f = \mathrm{tr}(I + \Omega\Theta X(I + S)\Psi Y^T$$

The first terms are

$$f(\Omega, S) \approx \mathrm{tr}\Theta X \Psi Y^T + \mathrm{tr}(\Omega\Theta X \Psi Y^T + S\Psi Y^T \Theta X$$

from which we get the equations for $\dot{\Theta}$ and $\dot{\Psi}$. The expression for the derivatives of ΘX and ΨY^T follow from these by substitution.

Matching problems involving more structure can lead to situations in which one does not want to investigate the minimum over all orthogonal matrices but rather only the minimum over some subgroup. This is illustrated by example two, in which case it is natural to consider the subgroup consisting of matrices of the following form.

Example: let $\mathcal{G}_1$ be the full group of orthogonal matrices and let $\mathcal{G}_a$ be the "maximal torus" consisting of orthogonal matrices of the form

$$G_a = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 & 0 & .. & .. \\ -\sin\theta_1 & \cos\theta_1 & 0 & 0 & .. & .. \\ 0 & 0 & \cos\theta_1 & \sin\theta_2 & .. & .. \\ 0 & 0 & -\sin\theta_1 & \cos\theta_2 & .. & .. \\ .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & .. & .. \end{bmatrix}$$

Let $\mathcal{G}_b$ consist of orthogonal matrices of the form

$$G_b = \begin{bmatrix} 0 & \omega_{12} I_2 & \omega_{13} I_2 & .. \\ -\omega_{12} I_2 & 0 & \omega_{23} & .. \\ -\omega_{13} & -\omega_{23} & 0 & .. \\ .. & .. & .. & .. \end{bmatrix}$$

The subalgebra consisting of those matrices that can be expressed as sums of matrices of these two types is of relevance in treating example two.

Such cases require a modification of the results of lemma 2.

Lemma 3: Let $\mathcal{L}_1$ be a Lie subalgebra contained in the set of n by n skew-symmetric matrices, and let $\mathcal{L}_2$ be a Lie subalgebra contained in the set of m by m skew-symmetric matrices. Suppose that their corresponding groups, denoted here as $\mathcal{G}_1$ and $\mathcal{G}_2$, which are closed subgroups of $O(n)$ and $O(m)$, respectively. Let π_i denote the orthogonal projection of the set of real skew-symmetric matrices onto $\mathcal{L}_i$. Then the gradient descent equation for

$$\phi(\Theta, \Psi) = \mathrm{tr}\Theta X \Psi Y^T$$

with Θ in $\mathcal{G}_1$ and Ψ in $\mathcal{G}_2$ is

$$\dot{\Theta} = \frac{1}{2}\pi_1(\Theta X \Psi Y^T - \Psi^T Y \Theta^T X^T)\Theta$$

$$\dot{\Psi} = \frac{1}{2}\pi_2(\Psi Y^T \Theta X - \Psi^T X^T Y \Theta^T)\Psi$$

Expressed in terms of the variables $T = \Theta X$ and $V = \Psi Y$ these equations take the form

$$\dot{T} = \frac{1}{2}\pi_1(TV - V^T T^T)T$$

$$\dot{V} = \frac{1}{2}\pi_2(VT - T^T V^T)V$$

Proof: We have an inner product defined on the space of skew-symmetric matrices by $\langle \Omega_1, \Omega_2 \rangle = \mathrm{tr}\Omega_1^T \Omega_2$. Relative to this inner product, there is an orthogonal projection onto any subspace. The given equations are simply the gradient flow on the given manifold.

4 References

1. R. W. Brockett, "Least Squares Matching Problems," *Linear Algebra and Its Applications*, Vols. 122/123/124 (1989) pp. 761-777.