JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 105, No. 1, pp. 249–252, APRIL 2000

TECHNICAL NOTE

Nearest q-Flat to m Points[1,2]

P. TSENG[3]

Communicated by O. L. Mangasarian

Abstract. Recently, Bradley and Mangasarian studied the problem of finding the nearest plane to m given points in $\Re^n$ in the least square sense. They showed that the problem reduces to finding the least eigenvalue and associated eigenvector of a certain $n \times n$ symmetric positive-semidefinite matrix. We extend this result to the general problem of finding the nearest q-flat to m points, with $0 \leq q \leq n-1$.

Key Words. Nearest point, clustering, eigenvalues, eigenvectors.

1. Problem Description

The problem of finding the nearest point to m given points in $\Re^n$ in the least square sense (i.e., minimizing the sum of the squares of the Euclidean distances to m points) has a well-known solution, namely, the mean of the m points. Recently, Bradley and Mangasarian (Ref. 1) studied a variant of this problem involvng finding the nearest plane to m given points. They showed that the problem reduces to finding the least eigenvalue and associated eigenvector of a certain $n \times n$ symmetric positive-semidefinite matrix. The above two problems are of interest, since they arise as subproblems in algorithms for data clustering (Refs. 1–3). In this note, we consider the general problem of finding the nearest q-flat (i.e., q-dimensional affine set; see Ref. 4, p. 3) to m points in $\Re^n$, with $0 \leq q \leq n-1$ integer. We show that this problem reduces to finding the $p = n - q$ least eigenvalues and associated eigenvectors of the same $n \times n$ matrix considered by Bradley and Mangasarian (see Theorem 2.1). Accordingly, the k-mean algorithm (Ref. 2, Section

[1]This work was partially supported by the National Science Foundation, Grant CCR-9731273.
[2]The author thanks Professor Olvi Mangasarian for encouragement and suggestions in this work.
[3]Professor, Department of Mathematics, University of Washington, Seattle, Washington.

3.3.2 and Ref. 3) and the k-plane algorithm of Ref. 1 can be extended by replacing the mean ($q=0$) and the plane ($q=n-1$) with a q-flat.

We first formulate the problem. Recall that the square of the distance from any $x\in\Re^n$ to the q-flat, $0\le q\le n-1$ integer,

$$\{y\in\Re^n: W'y=\gamma'\},\tag{1}$$

where $p=n-q$, $W\in\Re^{n\times p}$ has rank p, $\gamma\in\Re^{1\times p}$, is given by

$$\min_{y:W'y=\gamma'}\|y-x\|^2=\|W(W'W)^{-1}(W'x-\gamma')\|^2.$$

Throughout, a prime denotes transpose and $\|\cdot\|$ denotes the 2-norm. Then, given m points $x_1,\ldots,x_m$ in $\Re^n$, the square of the distance from each x_i to the set (1) is

$$\sum_{i=1}^{m}\|W(W'W)^{-1}(W'x_i-\gamma')\|^2.\tag{2}$$

The problem is to minimize this over all W of rank p and γ.

Notice that, without loss of generality, we can assume that the columns of W in (1) form an orthonormal set, i.e.,

$$W'W=I.$$

Then, the expression (2) simplifies to

$$\begin{aligned}\sum_{i=1}^{m}\|(W'x_i-\gamma')\|^2&=\sum_{i=1}^{m}\sum_{j=1}^{p}(w_j'x_i-\gamma_j)^2\\&=\sum_{j=1}^{p}\sum_{i=1}^{m}(x_i'w_j-\gamma_j)^2\\&=\sum_{j=1}^{p}\|Aw_j-e\gamma_j\|^2\\&=\|AW-e\gamma\|_F^2,\end{aligned}$$

where w_j denotes the jth column of W, γ_j denotes the jth component of γ, A is the $m\times n$ matrix whose ith row is x_i', e is the m-dimensional vector of 1's, and $\|\ \|_F$ denotes the Frobenius norm, i.e.,

$$\|B\|_F=\sqrt{\langle B,B\rangle_F}\quad\text{and}\quad\langle B,C\rangle_F=\mathrm{tr}[B'C].$$

Thus, the problem of minimizing (2) over all W of rank p, $1\le p\le n$ integer, and γ may be reformulated as

$$\min_{\substack{W\in\Re^{n\times p}:W'W=I\\ \gamma\in\Re^{1\times p}}}\|AW-e\gamma\|_F^2.\tag{3}$$

Problem (3) has a global optimal solution, since its feasible set is nonempty and compact in W, while its objective function is continuous in (W, γ) and coercive in γ. Moreover, the set of global optimal solutions of (3) is invariant under right multiplication by orthogonal matrices. Specifically, if $W \in \Re^{n \times p}$ and $\gamma \in \Re^{1 \times p}$ form a global optimal solution of (3), then for any orthogonal $O \in \Re^{p \times p}$, we have

$$(WO)'(WO) = I \quad \text{and} \quad \|A(WO) - e(\gamma O)\|_F^2 = \|AW - e\gamma\|_F^2,$$

so WO and γO form a global optimal solution of (3). This is not too surprising, since right-multiplying W and γ by a $p \times p$ orthogonal matrix does not change the set (1).

2. Solution Description

By using a generalized Rayleigh–Ritz theorem [Eq. (4.3.19) in Ref. 5], we have the following main result which shows that a global optimal solution of (3) occurs at an orthonormal set of eigenvectors corresponding to the p least eigenvalues of $A'(I - ee'/m)A$. This result may be viewed as an extension of Theorem 5 in Ref. 1 for the case $p = 1$.

Theorem 2.1. A global optimal solution of (3) occurs at a $W \in \Re^{n \times p}$ whose columns are the orthonormal eigenvectors corresponding to the p least eigenvalues of $B = A'(I - ee'/m)A$ and $\gamma = e'AW/m$. The optimal objective value equals the sum of the p least eigenvalues of B.

Proof. The proof is patterned partly after that of Theorem 5 in Ref. 1. The global optimal solution of (3) occurs at where the partial derivatives of the Lagrangian of (3) vanish. More precisely, for

$$L(W, \gamma, \Lambda) = \|AW - e\gamma\|_F^2 - \langle \Lambda, W'W - I \rangle_F,$$

where $\Lambda \in \Re^{p \times p}$ is symmetric, we obtain after some calculation that

$$(1/2)\nabla_W L(W, \gamma, \Lambda) = A'(AW - e\gamma) - W\Lambda = 0, \tag{4}$$

$$-(1/2)\nabla_\gamma L(W, \gamma, \Lambda) = e'(AW - e\gamma) = 0, \tag{5}$$

$$-\nabla_\Lambda L(W, \gamma, \Lambda) = W'W - I = 0. \tag{6}$$

Equation (5) yields

$$\gamma = e'AW/m,$$

so that

$$AW - e\gamma = AW - ee'AW/m = (I - ee'/m)AW.$$

Then, the idempotent property of $I-ee'/m$ (as in Lemma 2 of Ref. 1) implies

$$\|AW-e\gamma\|_F^2=\mathrm{tr}[W'A'(I-ee'/m)^2AW]=\mathrm{tr}[W'BW].$$

Since $W'W=I$, we also have from Eq. (4.3.19) in Ref. 5 that

$$\mathrm{tr}[W'BW]\geq\lambda_1+\cdots+\lambda_p, \tag{7}$$

where $\lambda_1\leq\lambda_2\leq\cdots\leq\lambda_n$ are the eigenvalues of B in ascending order. Thus, a lower bound on the optimal objective value of (3) is $\lambda_1+\cdots+\lambda_p$. This lower bound is attained by the W whose columns are the orthonormal eigenvectors of B corresponding to $\lambda_1,\ldots,\lambda_p$, with $\gamma=e'AW/m$. □

As is noted in Ref. 1, since $B=A'(I-ee'/m)^2A$, the eigenvalues and eigenvectors of B may be obtained from the singular value decomposition of $(I-ee'/m)A$. Also, instead of using Eq. (4.3.19) in Ref. 5, which is proven in Ref. 5 via the Courant–Fischer theorem, we can prove (7) directly using the observation that (4)–(6) imply

$$BW-W(W'BW)=0,$$

and then showing that

$$W=VO,$$

where the columns of $V\in\mathfrak{R}^{n\times p}$ are the orthonormal eigenvectors of B and $O\in\mathfrak{R}^{p\times p}$ is orthogonal.

References

1. Bradley, P. S., and Mangasarian, O. L., *k-Plane Clustering*, Computer Sciences Report 9808, University of Wisconsin, Madison, Wisconsin, 1998; Journal of Global Optimization (to appear).
2. Jain, A. K., and Dubes, R. C., *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
3. Selim, S. Z., and Ismail, M. A., *K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality*, IEEE Transactions on Pattern Analysis and Machine Intellience, Vol. 6, pp. 81–87, 1984.
4. Rockafellar, R. T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
5. Horn, R. A., and Johnson, C. R., *Matrix Analysis*, Cambridge University Press, Cambridge, United Kingdom, 1985.