ON THE DISTANCE BETWEEN POLYTOPES*

BY

LEOPOLD B. WILLNER
Northwestern University

Abstract. This paper is written to fill an apparent gap in the literature in connection with some elementary problems of distance in finite dimensional normed linear spaces. In particular the problem of determining the distance between two polytopes in a normed n-dimensional real vector space is considered. Special consideration is given to the case in which the norm is a twice differentiable function of its arguments and for this case a convex programming algorithm is presented. In addition several other cases are considered including the well known discrete Tchebycheff approximation. The results should find application in approximation theory.

0. Notation.

R^n —The vector space of all real n-tuples $(a_1, \cdots, a_n)^T$
$||\cdot||$ —A norm on R^n
$||\cdot||_t$ —The norms defined by

$$||U||_t = \left(\sum_{i=1}^{n} |u_i|^t\right)^{1/t} \qquad \text{for} \quad 1 \leq t < \infty$$

$$||U||_\infty = \max_i |u_i| \qquad U \varepsilon R^n$$

R_t^n —The normed linear space $\{R^n, ||\cdot||_t\}$
e —The vector $(1, 1, \cdots, 1)^T \varepsilon R^n$
A^+ —The Moore–Penrose inverse of a matrix A, or generalized inverse of A, see [7].

1. Introduction. The primary purpose of this paper is to consider a feasible method of determining the distance between two polytopes in a space $\{R^n, ||\cdot||\}$, i.e.

$$\underset{X, Y \varepsilon R^n}{\text{Minimize}} \, ||X - Y|| \qquad \text{when} \quad AX \leq d, \qquad BY \leq b \tag{1}$$

where A and B are $m \times n$ and $k \times n$ real matrices respectively. In particular the objective is to determine $X^* \varepsilon C_1 \equiv \{X; AX \leq d\}$ and $Y^* \varepsilon C_2 \equiv \{Y; BY \leq b\}$ such that

$$||X^* - Y^*|| = \text{dist.}(C_1, C_2) \equiv \inf_{X \varepsilon C_1, \, Y \varepsilon C_2} ||X - Y||. \tag{2}$$

In Sec. 2 we prove that such 'optimal' vectors always exist. For a development of the concept of 'proximity maps for convex sets' in Hilbert space see [1]. In Sec. 3 problem (1) is restricted to the case in which the norm is a twice continuously differentiable function. The importance of this class of problems is mainly due to the fact that it includes the problem of distance between polytopes in the familiar spaces R_t^n, $1 < t < \infty$.[1] In this section we consider a method of solving this problem. The

*Received October 29, 1966; revised manuscript received May 23, 1967. This paper was presented in part at the American Mathematical Society Meeting in New Brunswick, New Jersey, on September 2, 1966 under the title "On the Tchebycheff Approximation."

[1] R_t^n is the space R^n with norm $||\cdot||_t$ defined by $||U||_t = (\sum_{i=1}^n |u_i|^t)^{1/t}$.

algorithm which is presented is an adaptation and extension of the method of Frank and Wolfe [2] for convex programming to this problem. For the Euclidean space R_2^n problem (1) is a quadratic programming problem and other algorithms exist for its solution e.g. [3], [4]. In Sec. 4 an approximation criterion is given in conjunction with the algorithm stated in Sec. 3, and the proof of convergence is contained in Sec. 5.

In Sec. 6 the problem of determining the distance from a point to a subspace in $\{R^n, \|\cdot\|\}$ is considered. In addition to the method of Sec. 3, linear programming formulations of problem (1) for the spaces R_1^n and R_∞^n are considered in this section. This problem can be stated in the form

$$\operatorname*{Minimize}_{X \varepsilon R^n} \|AX - d\| \tag{3}$$

where A is an $m \times n$ matrix. In R_∞^n (3) is the Tchebycheff approximation e.g. [5]. In R_1^n (3) is the L_1-norm or 'taxicab metric' approximation problem. The system (3) may in addition be subject to linear inequality constraints, say $BX \leq b$.

It is noteworthy that problem (3) in R_2^n, the least squares approximation, even when subject to linear equality constraints always possesses a closed form solution (e.g. [6]). For in this space if (3) is restricted to $\tilde{S} \equiv \{X;\ BX = b\} \neq \emptyset$ then $X^0 = B^+b + [I - B^+B](A[I - B^+B])^+(AB^+b - d)$ is a solution.[2] To see this note that $\tilde{S} = \{X;\ X = B^+b + [I - B^+B]Y,\ Y \varepsilon R^n\}$, see [7], hence $\|AX - d\|_2 = \|AB^+b + A[I - B^+B]Y - d\|_2$ and it follows that $Y^0 = (A[I - B^+B])^+(AB^+b - d)$ minimizes this expression and the fact that X^0 is a solution is established.

2. Existence. We are given two polytopes in R^n: $C_1 = \{X;\ AX \leq d\}$ and $C_2 = \{Y;\ BY \leq b\}$; and the distance between any two corresponding points $X \varepsilon C_1$, $Y \varepsilon C_2$ is defined *vis à vis* a specific norm $\|\cdot\|$ to be $\|X - Y\|$. We seek two corresponding points which are closest among all such pairs of points. The fact that C_1 and C_2 are closed convex bodies is not enough to guarantee the existence of such points; however, for polytopes it follows nonetheless. The following will suffice to establish this fact.

LEMMA 1. *If S_1 and S_2 are nonempty subsets of R^n such that S, the algebraic difference of these sets, is closed then $\exists X^0 \varepsilon S_1$ and $Y^0 \varepsilon S_2 \ni \|X^0 - Y^0\| = \text{dist}(S_1, S_2)$.*

LEMMA 2. *The algebraic difference of two polytopes in R^n is a polytope in R^n.*

Proof. We wish to show that $C = \{U;\ U = X - Y,\ X \varepsilon C_1,\ Y \varepsilon C_2\}$ is a polytope in R^n whenever C_1 and C_2 are polytopes in R^n. We may write $C_1 = \{X;\ AX \leq a\}$, $C_2 = \{Y;\ BY \leq b\}$ for some matrices A and B, and vectors a and b. $\tilde{C} \equiv \{(X, Y);\ AX \leq a,\ BY \leq b\}$ is a polytope in R^{2n} and C is the image of $\tilde{C}$ under the linear transformation whose matrix in the standard ordered basis is $[I, -I]$. Since the image under a linear transformation of a polytope is also a polytope (to see this let P be a polytope then

$$P = \left\{U;\ U = \sum_{i=1}^{r} \lambda_i U^i + \sum_{j=1}^{s} \mu_j V^j,\ \lambda_i,\ \mu_i \geq 0,\ \sum_{i=1}^{r} \lambda_i = 1\right\}$$

where U^i are the extreme points of P and V^j are the extreme rays; the image $T(P)$ of P under a linear transformation T can be expressed as

$$T(P) = \left\{\bar{U};\ \bar{U} = \sum_{i=1}^{r} \lambda_i \bar{U}^i + \sum_{j=1}^{s} \mu_j \bar{V}^j,\ \lambda_i,\ \mu_j \geq 0\ \sum_{i=1}^{r} \lambda_i = 1\right\}$$

where $\bar{U}^i = TU^i$ and $\bar{V}^j = TV^j$, hence $T(P)$ is a polytope), C is a polytope.

[2] B^+ is the generalized inverse of B, see [7].

THEOREM. *The distance between any two polytopes in* $\{R^n, ||\cdot||\}$ *is attained.*

3. A method of feasible directions. In this section we consider an algorithm which is suitable for determining the solution to problem (1) when the norm possesses continuous second derivatives. This is a method of 'feasible directions' in the sense of Zoutendijk [8] and also very much in the spirit of 'generalized programming' *à la* Dantzig and Wolfe [4], [9], [10]. The proof of convergence follows readily from and is an adaptation of a proof in Frank and Wolfe [2]. The class of norms considered here includes the norms $||\cdot||_t$, $1 < t < \infty$.

For simplicity we can assume without loss of generality that $X^1 \varepsilon C_1$ and $Y^1 \varepsilon C_2$ are on hand to start the process. Solutions to systems of linear inequalities are readily available by various methods including: Dantzig, Orden, Wolfe [11], Goldstein and Cheney [5], Agmon [12], and Motzkin and Schoenberg [13].

In general at the kth iteration:

Phase 1.

(a) Given $X^k \varepsilon C_1$, $Y^k \varepsilon C_2$, let $U^k = X^k - Y^k$ and determine $\pi^k = \nabla\ ||U||\ |_{U=U^k}$

(b) Solve the two linear programs

$$\text{(I) Minimize}_{V \varepsilon C_1}\ \pi^k V \qquad \text{(II) Maximize}_{W \varepsilon C_2}\ \pi^k W$$

to obtain V^k and W^k extreme points of C_1 and C_2 respectively.

(c) Let $P^k = V^k - X^k$, $Q^k = W^k - Y^k$.

Phase 2.

(a) $\text{Minimize}_{0 \le \lambda_1, \lambda_2 \le 1} ||U^k + \lambda_1 P^k - \lambda_2 Q^k||$ to obtain λ_1^k and λ_2^k .

(b) Let $X^{k+1} = X^k + \lambda_1^k P^k$, $Y^{k+1} = Y^k + \lambda_2^k Q^k$ return to Phase 1 with k replaced by $k + 1$ unless $\lambda_1^k = \lambda_2^k = 0$ in which case terminate the process.

In case both C_1 and C_2 are convex polyhedra the above algorithm will suffice. If this is not the case then at some iteration it may happen that (I) or (II) yields an edge instead of an extreme point, say

$$V^k(t) = V_1^k + tV_2^k$$

in the polytope C_1 along which $\pi^k V \to -\infty$ as $t \to \infty$. In this case (2a) should be modified to

$$\underset{0 \le \lambda_1, \lambda_2 \le 1}{\text{Minimize}}\ ||U^k + \lambda_1 P^k(t) - \lambda_2 Q^k|| \tag{4}$$

where $P^k(t) = V^k(t) - X^k$ to obtain $\lambda_1^k(t)$ and λ_2^k . Determine max $t \ni \lambda_1^k(t) = 1$, say t^k, and let $\lambda_1^k = \lambda_1^k(t^k)$. The fact that t^k is finite is a direct consequence of the existence theorem. An unbounded edge of (II) is treated similarly.

4. An approximation criterion. For any feasible $U = X - Y$, $U^0 = X^0 - Y^0$, $X, X^0 \varepsilon C_1$, $Y, Y \varepsilon C_2$

$$\begin{aligned} ||U|| &\ge ||U^0|| + \nabla\ ||U^0||\ (U - U^0) \quad \text{by convexity} \\ &= ||U^0|| + \nabla\ ||U^0||\ ([X - Y] - [X^0 - Y^0]) \\ &= ||U^0|| + \nabla\ ||U^0||\ (X - X^0) - \nabla\ ||U^0||\ (Y - Y^0) \\ &\ge ||U^0|| + \nabla\ ||U^0||\ (V - X^0) - \nabla\ ||U^0||\ (W - Y^0) \\ &= ||U^0|| + \nabla\ ||U^0||\ (P - Q) \end{aligned}$$

thus for optimal U^* we have the bounds

$$||U^0|| + \nabla ||U^0|| (P - Q) \leq ||U^*|| \leq ||U^0||$$

where V, W are the solutions of

$$\begin{array}{llll} \text{Min} & \pi V & \text{Max} & \pi W \\ \text{s.t.} & AV \leq d & \text{s.t.} & BW \leq b \end{array}$$

here $\pi = \nabla ||U^0||$, $P = V - X^0$, $Q = W - Y^0$. This is precisely the approximation criterion of [2].

5. Convergence. At any iteration we suppose that U^k, V^{k+1}, W^{k+1} hence P^{k+1} and Q^{k+1} are known. Using the convex programming method of [2] we would obtain $\bar{U}^{k+1}$ which satisfies

$$\frac{||\bar{U}^{k+1}|| - ||U^*||}{L} \leq \frac{||U^k|| - ||U^*||}{L} \operatorname{Max} \left\{1 - \frac{||U^k|| - ||U^*||}{L}, \frac{1}{2}\right\} \tag{5}$$

where U^* is the optimal solution and $L > 0$ is a constant which depends on the constraint set (since the solution of (1) exists it follows that we can restrict the search to $C_1 \cap S$ and $C_2 \cap S$ for some sphere S, thus L exists). Now

$$\begin{aligned} ||\bar{U}^{k+1}|| \geq ||\hat{U}^{k+1}|| &\equiv \underset{1 \leq \lambda \leq 1}{\text{Minimum}} ||U^k + \lambda(P^{k+1} - Q^{k+1})|| \\ &\geq \underset{0 \leq \lambda_1, \lambda_2 \leq 1}{\text{Minimum}} ||U^k + \lambda_1 P^{k+1} - \lambda_2 Q^{k+1}|| \\ &= ||U^{k+1}|| \end{aligned}$$

hence

$$\frac{||U^{k+1}|| - ||U^*||}{L} \leq \frac{||\bar{U}^{k+1}|| - ||U^*||}{L} \tag{6}$$

and since (5) is sufficient for convergence (see [2]) it follows from (6) that the algorithm of Sec. 3 converges to an optimal solution.

6. The distance from a point to a subspace. In this case the problem has the form

$$\begin{aligned} \text{Minimize } & ||AX - d|| \\ \text{s.t. } & BX \geq b \end{aligned} \tag{7}$$

and may also be represented by

$$\begin{aligned} \text{Minimize } & ||Y - d|| \\ \text{s.t. } & DY \geq f \end{aligned} \tag{8}$$

for some D and f, however (7) is more common. The method of Sec. 3 applies to (8) if $||\cdot||$ has continuous second derivatives.

(A) *A Discrete Tchebycheff Approximation.* When the norm is $||\cdot||_\infty$, (7) is equivalent to

$$\begin{aligned} \text{Minimize } & t \\ \text{s.t. } & AX + te \geq d \\ & -AX + te \geq -d \\ & BX \qquad \geq b \end{aligned} \tag{9}$$

see [5], [14], [15], [16]. Similarly (1) is equivalent to

$$\begin{aligned} &\text{Minimize} \quad t \\ &\text{s.t.} \quad X - Y + te \geq 0 \\ &\quad -X + Y + te \geq 0 \\ &\quad -AX \geq -d \\ &\quad -BY \geq -b. \end{aligned} \tag{10}$$

Both (9) and (10) are linear programs and the method of [11] is highly efficient for their solution or for the solution of their duals.

(B) *A Taxicab Metric Approximation.* When the norm is $\|\cdot\|_1$, (7) is equivalent to

$$\begin{aligned} &\text{Minimize} \quad e^T Y^+ + e^T Y^- \\ &\text{s.t.} \quad AX + Y^+ - Y^- = d, \\ &\quad Y^+, Y^- \geq 0 \end{aligned} \tag{11}$$

see [16] and [17]. Similarly (1) is equivalent to

$$\begin{aligned} &\text{Minimize} \quad e^T Y^+ + e^T Y^- \\ &\quad X - Z + Y^+ - Y^- = 0, \\ &\quad -AX \geq -d, \\ &\quad -BZ \geq -b. \end{aligned} \tag{12}$$

(C) *A Least Squares Approximation.* When the norm is $\|\cdot\|_2$, (7) is the well known least squares approximation, a quadratic programming problem, and may be solved by the methods of [3] or [4].

References

[1] W. Cheney and A. A. Goldstein, *Proximity maps for convex sets*, Proc. Amer. Math. Soc. **10**, 448–450 (1959)

[2] M. Frank and P. Wolfe, *An algorithm for quadratic programming*, NRLQ **3**, 95–110 (1956)

[3] P. Wolfe, *The simplex method for quadratic programming*, Econometrica **27**, 382–398 (1959)

[4] G. Dantzig, *Linear programming and extensions*, Princeton University Press, Princeton, New Jersey, 1963.

[5] A. A. Goldstein and W. Cheney, *A finite algorithm for the solution of consistent linear equations and inequalities and for the Tchebycheff approximation of consistent linear equations*, Pacific J. Math. **3**, 415–427 (1958)

[6] A. Ben-Israel, *Generalized inverses and the Bott-Duffin network analysis*, J. Math Anal. Appl. **7** (1963)

[7] R. Penrose, *A generalized inverse for matrices*, Proc. Cambridge Philos. Soc. **51**, 406–413 (1955)

[8] G. Zoutendijk, *Methods of feasible directions*, Elsevier Publishing Co., New York, 1960.

[9] P. Wolfe, *Computational techniques for non linear programs*, Princeton University Conference, 1957

[10] G. Dantzig and P. Wolfe, *Decomposition principle for linear programs*, Operations Research **8**, 101–111 (1960)

[11] G. Dantzig, A. Orden and P. Wolfe, *Notes on linear programming: Part I. The generalized simplex method for minimizing a linear form under linear inequality restraints*, Pacific J. Math. **5**, 183–195 (1955)

[12] S. Agmon, *The relaxation method for linear inequalities*, Canada J. Math. **6**, 382–392 (1954)

[13] T. S. Motzkin and I. J. Shoenberg, *The relaxation method for linear inequalities*, Canada J. Math. **6**, 393–404 (1954)

[14] J. E. Kelley, Jr., *An application of linear programming to curve fitting*, J. SIAM 15–22 (1958)

[15] E. Stiefel, *Note on Jordan elimination, linear programming and Tchebycheff approximation*, Numerische Mathematik **2**, 1–17 (1960)

[16] I. Barrodale and A. Young, *Algorithms for best L_1 and L_∞ linear approximations on a discrete set*, Numerische Mathematik **8**, 295–306 (1966)

[17] A. Charnes, W. W. Cooper and R. O. Ferguson, *Optimal estimation of executive compensation by linear programming*, Management Science **1**, 138–151 (1955)