PACIFIC JOURNAL OF MATHEMATICS
Vol. 49, No. 2, 1973

THE NUMBER OF VECTORS JOINTLY ANNIHILATED BY TWO REAL QUADRATIC FORMS DETERMINES THE INERTIA OF MATRICES IN THE ASSOCIATED PENCIL

FRANK UHLIG

Pencils of real symmetric matrices and their associated quadratic forms are interrelated. It is well known that a pencil contains a definite matrix iff the associated quadratic forms do not vanish simultaneously, provided the matrices have dimension $n \geqq 3$. This knowledge is extended here to yield the following for nonsingular pairs of real symmetric matrices of dimension $n \geqq 3$:

(I) The pencil $P(S, T)$ contains a semidefinite, but no definite matrix iff the maximal number l of lin. ind. **vectors simultaneously annihilated by the associated quadratic forms lies between 1 and $n-1$ and certain conditions on S and T hold if $l = n-1$.**

(II) The pencil $P(S, T)$ contains only indefinite matrices iff $n-1 \leqq l \leqq n$ with other (complementary to the above) conditions holding if $l = n-1$.

First we introduce the relevant notation for a pair of real symmetric (r.s.) matrices S and T of the same dimension n:

DEFINITION 1. (a) The pencil $P(S, T) = \{aS + bT \mid a, b \in \boldsymbol{R}\}$ is a *d-pencil* if $P(S, T)$ contains a definite matrix.

(b) $P(S, T)$ is a *s.d. pencil* if $P(S, T)$ contains a nonzero semidefinite, but no definite matrix.

(c) $P(S, T)$ is an *i-pencil* if $P(S, T)$ contains only indefinite matrices, except for the zero matrix.

NOTATION. We denote by Q_S the set $\{x \in \boldsymbol{R}^n \mid x'Sx = 0\}$.

DEFINITION 2. A pair of r.s. matrices S and T is called a *nonsingular pair* if S is nonsingular.

This is our main result:

MAIN THEOREM. *For a pair of* r.s. *matrices S and T of dimension $n \geqq 3$ let $l = \max\{k \mid$ there exist k* lin. ind. *vectors in $Q_S \cap Q_T\}$. Then we have:*

(a) *$P(S, T)$ is a d-pencil iff $l = 0$, and for a nonsingular pair S, T:*

(b) *$P(S, T)$ is a s.d. pencil if and only if $1 \leqq l \leqq n-1$ and* *in case of $l = n - 1$ we have that S and T are simultaneously congruent either to*

$$\text{(D):}\quad \operatorname{diag}\left(\varepsilon\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \varepsilon_3, \cdots, \varepsilon_n\right) \quad \textit{and} \quad \operatorname{diag}\left(\eta\begin{pmatrix}0 & \alpha\\ \alpha & 1\end{pmatrix}, \varepsilon_3\alpha, \cdots, \varepsilon_n\alpha\right)$$

with $\alpha \in \boldsymbol{R}$, $\varepsilon, \eta, \varepsilon_j = \pm 1$ so that $\varepsilon_m\varepsilon_k = -1$ for at least one pair of indices $3 \leqq m, k \leqq n$, or to

$$\text{(E):}\quad \operatorname{diag}(\varepsilon_1, \cdots, \varepsilon_n) \quad \textit{and} \quad \operatorname{diag}(\varepsilon_1\lambda, \cdots, \varepsilon_{n-1}\lambda, \varepsilon_n\mu)\ ,$$

where $\varepsilon_j = \pm 1$ with $\varepsilon_m\varepsilon_k = -1$ for at least one pair $1 \leqq m, k \leqq n - 1$, and $\lambda, \mu \in \boldsymbol{R}$ with $\lambda \neq \mu$.

(c) *$P(S, T)$ is an i-pencil if and only if $n - 1 \leqq l \leqq n$ and in case of $l = n - 1$ we have that S and T are simultaneously congruent to*

$$\text{(A):}\quad \pm\operatorname{diag}\left(\begin{pmatrix}0 & & 1\\ & 1 & \\ 1 & & 0\end{pmatrix}, 1, \cdots, 1\right) \quad \textit{and} \quad \pm\operatorname{diag}\left(\begin{pmatrix}0 & & \alpha\\ & \alpha & 1\\ \alpha & 1 & 0\end{pmatrix}, \alpha, \cdots, \alpha\right),$$

or

$$\text{(B):}\quad \pm\operatorname{diag}\left(\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, 1, \cdots, 1\right) \quad \textit{and} \quad \pm\operatorname{diag}\left(\begin{pmatrix}b & a\\ a & -b\end{pmatrix}, \alpha, \cdots, a\right)$$

with $a, b, \alpha \in R$, $b \neq 0$.

Part (a) is well known and was first proved to this extent by Calabi [1]; for a complete bibliography regarding (a) the reader should refer to the remarks in Uhlig [2], Theorem 0.2.

Parts (b) and (c) will be proved with the help of the following theorems:

THEOREM 1. (*Uhlig* [2]).

Let S, T be a nonsingular pair of r.s. *matrices of dimension greater than* 2.

Then $P(S, T)$ is a s.d. pencil iff.

Either S and T are simultaneously congruent to $\operatorname{diag}(a_i)$ *and* $\operatorname{diag}(b_i)$ *with $b_i/a_i \neq b_j/a_j$ for at least one pair of indices (i, j) and we have*

$$\text{(i)}\quad \max_{a_i>0}\frac{b_i}{a_i} = \max_{a_i<0}\frac{b_i}{a_i} \quad \textit{or} \quad \text{(ii)}\quad \min_{a_i>0}\frac{b_i}{a_i} = \min_{a_i<0}\frac{b_i}{a_i}\ ,$$

or (iii) *S and T are simultaneously congruent to*

$$\operatorname{diag}(\varepsilon E, \cdots, \varepsilon E, \varepsilon_{k+1}, \cdots, \varepsilon_j, a_{j+1}, \cdots, a_n) \quad \textit{and}$$
$$\operatorname{diag}(\varepsilon EJ, \cdots, \varepsilon EJ, \varepsilon_{k+1}\alpha, \cdots, \varepsilon_j\alpha, b_{j+1}, \cdots, b_n)\ ,$$

where $J = \begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix}$, $E = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$; ε, $\varepsilon_i = \pm 1$, $\alpha \in \boldsymbol{R}$, $b_l \neq \alpha a_l$ *and* ε, $b_l - \alpha a_l$ *all have the same sign for* $l = j + 1, \cdots, n$.

THEOREM 2. (*Uhlig* [3]).

Let S, T *be a nonsingular pair of* r.s. *matrices of dimension greater than* 2. *Let* J *be the real Jordan normal form of* $S^{-1}T$ *and define* $l = \max\{k \mid$ *there exist* k lin. ind. *vectors in* $Q_S \cap Q_T\}$.

If (α): J *contains one* 3*-dimensional Jordan block, linear blocks else for the same eigenvalue and inertia* $S = (n-1, 1, 0)$ *or* $(1, n-1, 0)$.

Or (β): J *contains one* 2*-dimensional Jordan block corresponding to a nonreal eigenvalue, linear blocks else for identical eigenvalues and inertia* $S = (n-1, 1, 0)$ *or* $(1, n-1, 0)$, *then* $l = n - 1$.

If (γ): J *contains* $k \geqq 1$ *identical* 2*-dimensional Jordan blocks corresponding to one real eigenvalue* λ, *linear blocks else for eigenvalues* μ_i $(i = 2k+1, \cdots, n)$ *and the set*

$$\{\varepsilon_1, \cdots, \varepsilon_k, \varepsilon_i(\mu_i - \lambda) \mid i > 2k\} ,$$

where the $\varepsilon_i = \pm 1$ *are the constants in the canonical pair form of* S *and* T, *contains* $r \geqq 0$ *zeroes* $\mu_{2k+1} - \lambda = \cdots = \mu_{2k+r} - \lambda = 0$ *and only positive or only negative numbers else, and* $\varepsilon_{2k+1} = \cdots = \varepsilon_{2k+r}$, *then* $l = k$.

If (δ): *Condition* (γ) *holds except that not all* ε_i *are the same for* $2k + 1 \leqq i \leqq 2k + r$, *then* $l = k + r$.

If (η): S *and* T *can be simultaneouly diagonalised by a real congruence transformation, then* $l = 0, 2, 3, \cdots, n$ *are possible, depending on* S *and* T.

In all other cases $Q_S \cap Q_T$ *contains* $l = n$ lin. ind. *vector.*

And

THEOREM 3. (*Uhlig* [3]).

Let S, T, J, n, l, k, r, (α), (β), (γ), (δ), (η) *be as in Theorem* 2.

If $l = 0$, *then* (η) *holds for* J.

If $l = 1$, *then* (γ) *holds for* J *with* $k = 1$.

If $2 \leqq l \leqq n - 2$, *then* (γ), (δ) *or* (η) *holds for* J.

If $l = n - 1$, *then* (α), (β), (δ) (*with* $k = 1$, $r = n - 2$), *or* (η) *holds for* J.

If $l = n$, *then none of* (α), (β), (γ) *or* (δ) *may hold for* J.

Proof of the Main Theorem.

Part (b): All *s.d.* pencils have been characterized in Theorem 1. So if S and T are simultaneously congruent to $D_1 = \operatorname{diag}(\varepsilon_i)$ and $D_2 = \operatorname{diag}(\varepsilon_i \lambda_i)$ then (i) or (ii) of Theorem 1 must hold. Assume (i) holds and let

$$(1) \qquad \lambda = \max_{\varepsilon_i > 0} \lambda_i = \max_{\varepsilon_i < 0} (-\lambda_i)\ .$$

Then $x'D_1x = x_m^2 - x_k^2 + \sum \varepsilon_i x_i^2 = 0$ and $x'D_2x = \lambda x_m^2 - \lambda x_k^2 + \sum \varepsilon_i \lambda_i x_i^2 = 0$ implies

$$(2) \qquad \sum_{i \neq m,k} \varepsilon_i(\lambda_i - \lambda)x_i^2 = 0\ .$$

Now for $\varepsilon_i > 0$ we have $\lambda_i \leqq \lambda$ while $\varepsilon_i < 0$ implies $\lambda_i \geqq \lambda$. Hence (2) is a semidefinite quadratic form in $x_1, \cdots, x_n$ with at least one nonzero coefficient $\varepsilon_i(\lambda_i - \lambda)$ by Theorem 1. Thus all vectors $x \in Q_{D_1} \cap Q_{D_2}$ must have their ith component equal to zero for all i with $\lambda_i \neq \lambda$ and hence $2 \leqq l \leqq n - 1$ and $l = n - 1$ exactly in case of (η). The proof in case of (ii) is similar.

If S and T are not simultaneously diagonalizable, then by Theorem 1 we must have case (γ) or (δ) of Theorem 2. We conclude that $1 \leqq l \leqq n - 1$ and $l = n - 1$ only if (δ) holds with $k = 1$, $r = n - 2$, hence if (D) holds. Now for the converse: If $l = n - 1$ and (D) holds, then $P(S, T)$ is a *s.d.* pencil by Theorem 1. If $1 \leqq l \leqq n - 2$ then by Theorem 3 only (γ), (δ) or (η) may hold for S and T. We are done in case of (γ) or (δ) due to Theorem 1. For arbitrary l the case of (η) will be settled by the following Lemma:

LEMMA 1. *For two real diagonal matrices* $D_1 = \operatorname{diag}(\varepsilon_i)$ *and* $D_2 = \operatorname{diag}(\varepsilon_i\lambda_i)$, *where* $\varepsilon_i = \pm 1$, $i = 1, \cdots, n$ *let* $l = \max\{k \mid$ *there exist* k lin. ind. *vectors in* $Q_{D_1} \cap Q_{D_2}\}$.

Then $P(D_1, D_2)$ is an i-pencil iff $l = n$.

Proof. Assume $P(D_1, D_2)$ is not an i-pencil. Then it must be a d-pencil or a *s.d.* pencil, in which cases $l = 0$ or $1 \leqq l \leqq n - 1$ as shown above.

Conversely if $P(D_1, D_2)$ is an i-pencil, then $L = \{(\varepsilon_i, \varepsilon_i\lambda_i)\}$ is contained in no open or closed halfspace of $\boldsymbol{R}^2$ and hence we may assume without loss of generality that $\lambda_3 < \lambda_1 < \lambda_2$ and $\varepsilon_1 = -1$, $\varepsilon_2, \varepsilon_3 = 1$.

Then the quadratic forms read like:

$$x'D_1x = -x_1^2 + x_2^2 + x_3^2 + \sum_{i=3}^{n} \varepsilon_i x_i^2$$

and

$$x'D_2x = -\lambda_1x_1^2 + \lambda_2x_2^2 + \lambda_3x_3^2 + \sum_{i=3}^{n} \varepsilon_i\lambda_i x_i^2\ .$$

Let $f\colon \boldsymbol{R}^3 \to \boldsymbol{R}^2$; $f(x_1, x_2, x_3) = (-x_1^2 + x_2^2 + x_3^2, -\lambda_1x_1^2 + \lambda_2x_2^2 + \lambda_3x_3^2)$. We will show that f is surjective:

First if $a, b \in \operatorname{Im} f$, then $\alpha a + \beta b \in \operatorname{Im} f$ for all $\alpha, \beta \geqq 0$. For if $f(x_1, x_2, x_3) = a$, $f(y_1, y_2, y_3) = b$, then

$$f(\sqrt{\alpha x_1^2 + \beta y_1^2}, \sqrt{\alpha x_2^2 + \beta y_2^2}, \sqrt{\alpha x_3^2 + \beta y_3^2}) = \alpha a + \beta b .$$

Secondly $0 \neq x \in \boldsymbol{R}^3$ can always be found s.t. $f(x) = (\varepsilon, \delta)$, whenever $(\varepsilon, \delta) \in \{(0, 0), (1, 0), (-1, 0), (0, 1), (0, -1)\}$, which will complete the proof that $\operatorname{Im} f = \boldsymbol{R}^2$:

To solve $f(x) = (\varepsilon, \delta)$ means to solve the following linear system of equations in x_1^2, x_2^2, x_3^2:

$$\left(\begin{array}{ccc|c} -1 & 1 & 1 & \varepsilon \\ -\lambda_1 & \lambda_2 & \lambda_3 & \delta \end{array}\right) .$$

This is equivalent to

$$\left(\begin{array}{ccc|c} -1 & 1 & 1 & \varepsilon \\ 0 & \lambda_2 - \lambda_1 & \lambda_3 - \lambda_1 & \delta - \varepsilon\lambda_1 \end{array}\right) . \tag{3}$$

Since $\lambda_3 < \lambda_1 < \lambda_2$, the second equation in (3) represents a line L with positive slope in the quarter plane $\{(x_2^2, x_3^2)\}$. Hence this line will have points in common with the set $A = \{(x_2^2, x_3^2) \mid x_2^2 + x_3^2 > 1\}$. But $\varepsilon \leqq 1$, hence taking $(x_2^2, x_3^2) \in L \cap A \neq \emptyset$, then the system (3) can be solved with $x_1^2 > 0$.

The lemma is proved if we exhibit n lin. ind. vectors in $Q_{D_1} \cap Q_{D_2}$.

Let $y_1 = (x_1, x_2, x_3, 0, \cdots, 0)$ with $x_1, x_2, x_3 \neq 0$ s.t. $f(x_1, x_2, x_3) = 0$,

$$\begin{aligned} y_2 &= (-x_1, x_2, x_3, 0, \cdots, 0) \\ y_3 &= (-x_1 - x_2, x_3, 0, \cdots, 0) \quad \text{and} \\ y_i &= (\alpha_i, \beta_i, \gamma_i, 0, \cdots, 0) + e_i \quad \text{for} \quad i \geqq 4 , \end{aligned}$$

where e_i is the ith unit vector and $\alpha_i, \beta_i, \gamma_i$ are such that $f(\alpha_i, \beta_i, \gamma_i) = (-\varepsilon_i, -\varepsilon_i\lambda_i)$.

Part (c): This is obvious once parts (a) and (b) have been proved. Thus the Main Theorem is proved.

Outlook. The Main Theorem contains three "if and only if" statements relating the type of pencil a nonsingular pair of r.s. matrices S and T generates to the maximal number l of lin. ind. vectors simultaneously annihilated by the two associated real quadratic forms. On the one hand the distinction of *d*-, *s.d.* and *i*-pencils is easily made, while on the other it is also enough to know l for a specific pair of matrices except for the case that $l = n - 1$. Then the further conditions on the finest simultaneous block diagonal structure of S and T are somewhat hard to verify; one needs to know the real Jordan normal form of $S^{-1}T$ for example.

It is an open question whether this difficulty could be avoided either by changing the concepts of d-, $s.d.$ or i-pencil slightly or by using a characterization of pairs of quadratic forms different from our number l.

ACKNOWLEDGMENT. Dr. Olga Taussky-Todd, my thesis advisor at the California Institute of Technology, had suggested to try to generalize the theorems about simultaneous diagonalization and about d-pencils by studying $Q_S \cap Q_T$. Moreover, Dr. H. F. Bohnenblust suggested to find properties of $P(S, T)$ that relate to l, the maximal number of lin. ind. vectors in $Q_S \cap Q_T$, rather than relating l to the real Jordan normal form of $S^{-1}T$ as done in [3]. I am grateful to both, for both suggestions have been intertwined here.

REFERENCES

1. E. Calabi, *Linear systems of real quadratic forms*, Proc. Amer. Math. Soc., **15** (1964), 844-846.

2. F. Uhlig, *Definite and semidefinite matrices in a real symmetric matrix pencil*, Pacific J. Math., **49** (1973), 561-568.

3. ———, *On the maximal number of linearly independent real vectors simultaneously annihilated by two real quadratic forms*, Pacific J. Math., **49** (1973), 543-560.

Received July 12, 1972.

UNIVERSITÄT WÜRZBURG