Finding nice real matrix representations of real operators

Normal Operators: Let $T\colon V \to V$ be a normal operator on a real inner product space. We will pick an orthonormal basis $\mathcal{B} = \{\beta_1, \ldots, \beta_n\}$ so that $[T]_\mathcal{B}$ is as nice as possible. The general technique will be to pick any orthonormal basis $\mathcal{A}$. Then $[T]_\mathcal{A}$ is some real normal matrix A. We then think of A as being a complex normal matrix whose entries just happen to be real. Then there is an orthonormal basis in $\mathbb{C}^n$ of characteristic vectors of A, say $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. For each j we have $A\alpha_j = c_j a_j$ for some characteristic value c_j. We may write $\alpha_j = \gamma_j + i\delta_j$ where γ_j and δ_j are real. After reordering, we may suppose that c_j is real for all $j \leq k$ and c_j has positive imaginary part for all $k < j \leq \ell$ and c_j has negative imaginary part for all $\ell < j \leq n$. Now:

a) For any real matrix, normal or not, we can do the Gram-Schmidt procedure on the $2k$ real vectors $\{\gamma_1, \delta_1, \gamma_2, \delta_2, \ldots, \gamma_k, \delta_k\}$ and obtain k orthonormal characteristic vectors $\{\beta_1, \ldots, \beta_k\}$. In fact if we reorder so $c_1 \leq c_2 \leq \cdots \leq c_k$ then $T\beta_j = c_j\beta_j$.

b) Since A is normal it turns out that $\{\sqrt{2}\gamma_{k+1}, \sqrt{2}\gamma_{k+1}, \sqrt{2}\gamma_{k+2}, \ldots, \sqrt{2}\gamma_\ell, \sqrt{2}\gamma_\ell\}$ forms an orthonormal set which gives us $\{\beta_{k+1}, \ldots, \beta_{2\ell-k}\}$.

c) $n = 2\ell - k$ so we have found $\{\beta_1, \ldots, \beta_n\}$.

d) If $c_j = a_j + ib_j$ then $T\beta_{2j-k} = a_j\beta_{2j-k} - b_j\beta_{2j-k+1}$ and $T\beta_{2j-k+1} = b_j\beta_{2j-k} + a_j\beta_{2j-k+1}$ for all $k < j \leq \ell$. Thus $[T]_\mathcal{B}$ is block diagonal

$$\begin{bmatrix} D & 0 & \cdots & 0 \\ 0 & B_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & B_{\ell-k} \end{bmatrix}$$

where D is a $k \times k$ diagonal matrix with j-th diagonal entry c_j, and where B_j is the 2×2 matrix $\begin{bmatrix} a_{k+j} & b_{k+j} \\ -b_{k+j} & a_{k+j} \end{bmatrix}$.

Let us now see why all this works. First, if A is normal then characteristic vectors for different characteristic values must be orthogonal. To see this, suppose that $A\alpha = c\alpha$ and $A\beta = d\beta$ and $c \neq d$. Recall Theorem 19 on page 315 of H&K which implies $A^*\beta = \bar{d}\beta$. Then

$$c(\alpha \mid \beta) = (c\alpha \mid \beta) = (A\alpha \mid \beta) = (\alpha \mid A^*\beta) = (\alpha \mid \bar{d}\beta) = d(\alpha \mid \beta)$$

so $(\alpha \mid \beta) = 0$.

Next note that if $A\alpha = c\alpha$ then

$$A\bar{\alpha} = \bar{A}\bar{\alpha} = \overline{A\alpha} = \overline{c\alpha} = \bar{c}\bar{\alpha}$$

so $\bar{\alpha}$ is a characteristic vector with characteristic value $\bar{c}$. In particular, if c is not real then $c \neq \bar{c}$ so $(\alpha \mid \bar{\alpha}) = 0$.

Now suppose that $A\alpha = c\alpha$, and $\alpha = \gamma + i\delta$, and $c = a + bi$ where γ, δ, a, and b are real. Then $A\alpha = c\alpha$ so $A\gamma + iA\delta = (a + bi)\gamma + (-b + ai)\delta = a\gamma - b\delta + i(b\gamma + a\delta)$ so equating real and imaginary parts we have $A\gamma = a\gamma - b\delta$ and $A\delta = b\gamma + a\delta$.

If $b = 0$ then we have $A\gamma = a\gamma$ and $A\delta = a\delta$. Thus the real and imaginary parts of α are also characteristic vectors.

Now suppose that $b \neq 0$. Then $(\alpha \mid \bar{\alpha}) = 0$ so

$$0 = (\gamma + i\delta \mid \gamma - i\delta) = (\gamma \mid \gamma) - (\delta \mid \delta) + 2i(\gamma \mid \delta)$$

which means $(\gamma \mid \delta) = 0$ and $(\gamma \mid \gamma) = (\delta \mid \delta)$. So γ and δ are orthogonal and have the same length. Note also that $(\alpha \mid \alpha) = (\gamma \mid \gamma) + (\delta \mid \delta) + i(\delta \mid \gamma) - i(\gamma \mid \delta) = 2(\gamma \mid \gamma)$ So γ and δ have length $1/sqrt2$ times the length of α. In particular if α has unit length then $\sqrt{2}\gamma, \sqrt{2}\delta$ is an orthonormal set.

These are the main ingredients which I will expand upon later but meanwhile here are some matlab calculations where we see all this in practice.

Start out with a random skew symmetric real matrix and find its eigenvectors and eigenvalues.

```
>> A = rand(7,7);
>> A = (A-A');
>> [V,D] = eig(A);
```

Let us look at the characteristic values (the .' takes the transpose to make a row vector which prints out better than a column):

```
>> diag(D).'
ans =
Columns 1 through 4
-0.0000 + 1.7173i  -0.0000 - 1.7173i        0 + 0.7516i        0 - 0.7516i
Columns 5 through 7
0.0000             0.0000 + 0.3561i   0.0000 - 0.3561i
```

We see they come in complex conjugate pairs except for the fifth one which is 0. Let's see if the fifth characteristic vector is real.

```
>> V(:,5)'
ans =
0.6173   -0.3724   -0.4482   -0.1224    0.4051   -0.0892   -0.3039
```

Okay it is real. The first, third and sixth characteristic values have positive real part. So we can take the following change of basis matrix.

```
>> P = [V(:,5) sqrt(2)*real(V(:,1)) sqrt(2)*imag(V(:,1))];
>> P= [P sqrt(2)*real(V(:,3)) sqrt(2)*imag(V(:,3)) ];
>> P= [P sqrt(2)*real(V(:,6)) sqrt(2)*imag(V(:,6))];
>> % check to see that P is orthogonal since P'*P is close to the identity.
>> norm(P'*P-eye(7))
ans =
1.6709e-15
```

Now see that $P^{-1}AP = P^{*}AP$ is in block diagonal form.

```
>> P'*A*P
ans =
0.0000    0.0000   -0.0000    0.0000   -0.0000    0.0000   -0.0000
-0.0000   -0.0000    1.7173    0.0000   -0.0000    0.0000    0.0000
0.0000   -1.7173   -0.0000    0.0000    0.0000    0.0000   -0.0000
-0.0000   -0.0000   -0.0000    0.0000    0.7516   -0.0000         0
0.0000    0.0000   -0.0000   -0.7516    0.0000   -0.0000   -0.0000
-0.0000   -0.0000   -0.0000    0.0000    0.0000   -0.0000    0.3561
0.0000   -0.0000    0.0000   -0.0000    0.0000   -0.3561   -0.0000
```

Now let us look at an orthogonal example. Start with a random 5×5 orthogonal matrix, and check to see it is in fact orthogonal.

```
>> A = orth(rand(5,5));
>> norm(A*A'-eye(5))
ans =
9.1341e-16
```

Now find its characteristic values and characteristic vectors and check that the characteristic vectors for the real characteristic values are real.

```
>> [V D] = eig(A);
>> diag(D).'
ans =
1.0000            -0.8948 + 0.4466i  -0.8948 - 0.4466i   0.1265 + 0.9920i   0.1265 - 0.9920i
>> norm(imag(V(:,1)))
ans =
0
```

Now form the orthogonal coordinate change P and check that P^*AP is in block diagonal form.

```
>> P=[V(:,1) sqrt(2)*real(V(:,2)) sqrt(2)*imag(V(:,2))];
>> P=[P sqrt(2)*real(V(:,4)) sqrt(2)*imag(V(:,4))];
>> P'*A*P
ans =
1.0000   -0.0000   -0.0000    0.0000    0.0000
0.0000   -0.8948    0.4466    0.0000   -0.0000
-0.0000   -0.4466   -0.8948    0.0000    0.0000
-0.0000    0.0000   -0.0000    0.1265    0.9920
-0.0000    0.0000    0.0000   -0.9920    0.1265
```

It is interesting to note what this means in the case of orthogonal matrices. So if A is an orthogonal $n \times n$ matrix then we may write $\mathbb{R}^n = W_1 \oplus W_2 \oplus \cdots \oplus W_k$ where each W_i is invariant under A, the W_i are mutually perpendicular, A is the identity on W_1, A is minus the identity on W_2, and on W_j for $j > 2$ we have $\dim W_j = 2$ and A acts by rotating by an angle θ_j on W_j.

Real Jordan form: Now suppose we have any old operator $T\colon V \to V$ on a finite dimensional vector space V over the reals. Our goal is to find a basis $\mathcal{A}$ of V so that $[T]_{\mathcal{A}}$ has a nice form. For example, if the characteristic polynomial of T is a product of real linear factors, we can choose $\mathcal{A}$ so that $[T]_{\mathcal{A}}$ is in Jordan form. But you cannot do this if the characteristic polynomial of T is not a product of real linear factors.

By taking any basis $\mathcal{A}'$ of V and letting $A = [T]_{\mathcal{A}'}$ we may as well assume V is $\mathbb{R}^n$ and T is given by multiplying by some $n \times n$ real matrix A. We now think of A as an operator on $\mathbb{C}^n$. Then there is a basis of $\mathbb{C}^n$ which puts A in Jordan form. In particular, if the characteristic values of A are $c_1, \ldots, c_m$ and the minimal polynomial of A is $(x - c_1)^{k_1}(x - c_2)^{k_2} \cdots (x - c_m)^{k_m}$ then if $W_j = NS(A - c_jI)^{k_j}$ then we have $\mathbb{C}^n = W_1 \oplus W_2 \oplus \cdots \oplus W_m$. If we fix some j and let $c = c_j$ and $N = A - cI$, then we further decompose W_j as $W_j = Z(\alpha_1; A) \oplus Z(\alpha_2; A) \oplus \cdots \oplus Z(\alpha_k; A)$.

Recall that if c_j is real then the restriction of A to W_j can be put into Jordan form (using only real basis vectors) so we only need concern ourselves with the case where $c = c_j$ is not real.

The first thing to note is that if $\alpha \in W_j$ then $(A - \bar{c}I)^{k_j}\overline{\alpha} = \overline{(A - cI)^{k_j}\alpha} = \bar{0} = 0$ so $\bar{c}$ is also a characteristic value. If c is not real then after reordering we may as well suppose that $\bar{c} = c_{j+1}$. So we have $\overline{W_j} \subset W_{j+1}$ and likewise $\overline{W_{j+1}} \subset W_j$, so $\overline{W_j} = W_{j+1}$.

Now I claim that if c is not real and $\{\delta_1, \ldots, \delta_s\}$ is a basis of W_j and $\delta_\ell = \beta_\ell + i\gamma_\ell$ with β_ℓ and γ_ℓ real for all ℓ, then $\{\beta_1, \gamma_1, \beta_2, \gamma_2, \ldots, \beta_s, \delta_s\}$ is a basis of $W_j \oplus W_{j+1}$. To see this, take any $\alpha \in W_j \oplus W_{j+1}$. We may write $\alpha = \alpha_0 + \alpha_1$ where $\alpha_0 \in W_j$ and $\alpha_1 \in W_{j+1}$. Now $\overline{\alpha_1} \in W_j$ so we may write $\overline{\alpha_1} = d_1\delta_1 + \cdots + d_s\delta_s$ and $\alpha_0 = e_1\delta_1 + \ldots + e_s\delta_s$ so $\alpha = e_1(\beta_1 + i\gamma_1) + \overline{d_1}(\beta_1 - i\gamma_1) + \cdots + e_s(\beta_s + i\gamma_s) + \overline{d_s}(\beta_s - i\gamma_s)$ is a linear combination of the $\{\beta_1, \gamma_1, \beta_2, \gamma_2, \ldots, \beta_s, \delta_s\}$ so they span. On the other hand, if $0 = d_1\beta_1 + e_1\gamma_1 + \cdots + d_s\beta_s + e_s\gamma_s$. Let $\alpha_2 = (d_1 - ie_1)/2\delta_1 + (d_2 - ie_2)/2\delta_2 + \cdots + (d_s - ie_s)/2\delta_s \in W_j$ and $\alpha_3 = (d_1 + ie_1)/2\overline{\delta_1} + (d_2 - ie_2)/2\overline{\delta_2} + \cdots + (d_s - ie_s)/2\overline{\delta_s} \in W_{j+1}$. But

$$(d_\ell - ie_\ell)/2\delta_\ell + (d_\ell + ie_\ell)/2\overline{\delta_\ell} = (d_\ell - ie_\ell)/2(\beta_\ell + i\gamma_\ell) + (d_\ell + ie_\ell)/2(\beta_\ell - i\gamma_\ell)$$

$$= (1/2)(d_\ell\beta_\ell + id_\ell\gamma_\ell - ie_\ell\beta_\ell + e_\ell\gamma_\ell + d_\ell\beta_\ell - id_\ell\gamma_\ell + ie_\ell\beta_\ell + e_\ell\gamma_\ell) = d_\ell\beta_\ell + e_\ell\gamma_\ell$$

So we have $0 = \alpha_2 + \alpha_3$ which means $\alpha_2 = \alpha_3 = 0$ since W_j and W_{j+1} are independent. But then linear independence of $\delta_1, \ldots, \delta_s$ implies $d_\ell - ie_\ell = 0$ for all ℓ. Also $0 = \overline{2\alpha_3} = \sum \overline{d_\ell + ie_\ell}\delta_\ell$ so $d_\ell + ie_\ell = 0$ for all ℓ. Thus $d_\ell = e_\ell = 0$ for all ℓ and thus $\{\beta_1, \gamma_1, \beta_2, \gamma_2, \ldots, \beta_s, \delta_s\}$ is linearly independent.

Suppose the Jordan form of A has a $k \times k$ Jordan block with a nonreal diagonal entry $c = a + bi$. This Jordan block corresponds to a cyclic subspace $Z(\alpha; A)$ with basis $\{\alpha, N\alpha, \ldots, N^{k-1}\alpha\}$. Write $N^j\alpha = \beta_j + i\gamma_j$ where β_j and γ_j are real. Then

$$A\beta_j + iA\gamma_j = A(\beta_j + i\gamma_j) = (N + aI + biI)(\beta_j + i\gamma_j) = N(\beta_j + i\gamma_j) + (a + bi)(\beta_j + i\gamma_j)$$

$$= \beta_{j+1} + i\gamma_{j+1} + a\beta_j - b\gamma_j + i(a\gamma_j + b\beta_j)$$

So we get $A\beta_j = \beta_{j+1} + a\beta_j - b\gamma_j$ and $A\gamma_j = \gamma_{j+1} + a\gamma_j + b\beta_j$ So the matrix of the restriction of A to the subspace spanned by $\beta_0, \gamma_0, \beta_1, \gamma_1, \ldots, \beta_{k-1}, \gamma_{k-1}$ using the basis $\{\beta_0, \gamma_0, \beta_1, \gamma_1, \ldots, \beta_{k-1}, \gamma_{k-1}\}$ is the $2k \times 2k$ matrix $RJ_k(c) = \begin{bmatrix} B & 0 & \cdots & 0 & 0 \\ I & B & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & B & 0 \\ 0 & 0 & \cdots & I & B \end{bmatrix}$ where I is the 2×2 identity and $B = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$.

Thus in the end, for any linear operator $T\colon V \to V$ there is a basis $\mathcal{A}$ of V so that $[T]_{\mathcal{A}}$ is block diagonal with each diagonal block either $J_k(c)$ for some real characteristic value c, or $RJ_k(c)$ for some non-real c.