VII.B. Bilinear forms

There are two standard "generalizations" of the dot product on $\mathbb{R}^n$ to arbitrary vector spaces. The idea is to have a "product" that takes two vectors to a scalar. Bilinear forms are the more general version, and inner products (a *kind* of bilinear form) the more specific — they are the ones that "look like" the dot product (= "Euclidean inner product") with respect to *some* basis.[2] Here we tour the zoo of various classes of bilinear forms, before narrowing our focus to inner products in the next section.

Now let $V/\mathbb{R}$ be an n-dimensional vector space:

VII.B.1. DEFINITION. A **bilinear form** is a function $B\colon V \times V \to \mathbb{R}$ such that

$$B(a\vec{u} + b\vec{v}, \vec{w}) = aB(\vec{u}, \vec{w}) + bB(\vec{v}, \vec{w})$$
$$B(\vec{u}, a\vec{v} + b\vec{w}) = aB(\vec{u}, \vec{v}) + bB(\vec{u}, \vec{w})$$

for all $\vec{u}$, $\vec{v}$, $\vec{w} \in V$. It is called **symmetric** if also $B(\vec{u}, \vec{w}) = B(\vec{w}, \vec{u})$.

Now let's lay down the following law: whenever we are considering a bilinear form B on a vector space V (even if $V = \mathbb{R}^n$!), all orthogonality (or orthonormality) is relative to B. That is, "$\vec{u} \perp \vec{w}$" by definition means $B(\vec{u}, \vec{w}) = 0$. Sounds fine, huh?

VII.B.2. EXAMPLE. Take a look at the (non-symmetric) bilinear form

$$B(\vec{x}, \vec{y}) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_2 y_1$$

on $\mathbb{R}^2$. Its simplicity belies an unpleasant nature: while allowing that $\binom{1}{0} \perp \binom{0}{1}$, it turns right around on us with $\binom{0}{1} \not\perp \binom{1}{0}$! We can avoid this sort of character if we work with symmetric[3] bilinear forms.

[2] We should warn the reader right away that the *only* bilinear form or inner product *on* $\mathbb{R}^n$ for which $B(\hat{e}_i, \hat{e}_j) = \delta_{ij}$ (i.e. under which $\hat{e}$ is an orthonormal basis) is the dot product! In order to avoid confusion coming from $\hat{e}$ we will often use the language of an abstract n-dimensional vector space V instead of $\mathbb{R}^n$.

[3] or anti-symmetric, or Hermitian symmetric — see below.

So try instead

$$\tilde{B}(\vec{x},\vec{y}) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1y_1 - x_2y_2;$$

$\tilde{B}$ is symmetric but its notion of norm is not what we are used to from the dot product: $\tilde{B}(\hat{e}_2,\hat{e}_2) = -1$, while $\tilde{B}\left(\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right),\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)\right) = 0$ — that is, under $\tilde{B}$ there exists a *nonzero* self-perpendicular vector $\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) \perp \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$. Such pathologies with general bilinear forms are routine; the point of defining inner products in §VII.C will be to get rid of them.

The Matrix of a Bilinear form. Consider a basis $\mathcal{B} = \{\vec{v}_1,\ldots,\vec{v}_n\}$ for V and write $\vec{u} = \sum u_i\vec{v}_i$, $\vec{w} = \sum w_i\vec{v}_i$. Let B be any bilinear form on V, and set

(VII.B.3) $$b_{ij} = B(\vec{v}_i,\vec{v}_j);$$

we have

$$B(\vec{u},\vec{w}) \;=\; B\left(\textstyle\sum_i u_i\vec{v}_i, \sum_j w_j\vec{v}_j\right) \;=\; \sum_{i,j} u_i\, b_{ij}\, w_j$$

$$= \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} =: {}^t[\vec{u}]_{\mathcal{B}}[B]^{\mathcal{B}}[\vec{w}]_{\mathcal{B}},$$

defining a *matrix* for B. We write the basis $\mathcal{B}$ as a superscript, because change-of-basis does not work the same way as for transformation matrices: since (for any other basis $\mathcal{C}$)

$$\begin{aligned} {}^t([\vec{u}]_{\mathcal{B}})\,[B]^{\mathcal{B}}([\vec{w}]_{\mathcal{B}}) \;&=\; {}^t(P_{\mathcal{C}\to\mathcal{B}}[\vec{u}]_{\mathcal{C}})\,[B]^{\mathcal{B}}(P_{\mathcal{C}\to\mathcal{B}}[\vec{w}]_{\mathcal{C}}) \\ &=\; {}^t[\vec{u}]_{\mathcal{C}}\left({}^tP_{\mathcal{C}\to\mathcal{B}}[B]^{\mathcal{B}}P_{\mathcal{C}\to\mathcal{B}}\right)[\vec{w}]_{\mathcal{C}} \end{aligned}$$

for all $\vec{u},\vec{w}$, we find that

(VII.B.4) $$\boxed{[B]^{\mathcal{C}} = {}^tP[B]^{\mathcal{B}}P.}$$

Matrices related in this fashion (where P is invertible) are said to be *cogredient*. Such matrices have the same rank (why?).

Symmetric Bilinear forms. Next let B be an arbitrary *symmetric* bilinear form, so that $\perp$ is a symmetric relation, and the subspace

$$V^{\perp\vec{w}} = \{\vec{v} \in V \mid B(\vec{w},\vec{v}) = 0\} = \{\vec{v} \in V \mid B(\vec{v},\vec{w}) = 0\}$$

makes sense. Now already *the matrix of such a form (with respect to any basis) is symmetric*;[4] that is,

$${}^{t}[B]^{\mathcal{B}} = [B]^{\mathcal{B}},$$

since $b_{ij} = b_{ji}$ in (VII.B.3). We now explain how to put it into an especially nice form.

VII.B.5. LEMMA. *There exists a basis $\mathcal{B}$ such that*

$$[B]^{\mathcal{B}} = \mathrm{diag}\{b_1, \ldots, b_r, 0, \ldots, 0\},$$

where the $b_i \neq 0$.

PROOF. This is by induction on $n = \dim V$ (clear for $n = 1$): we assume the result for $(n-1)$-dimensional spaces and prove it for V.

If $B(\vec{u},\vec{w}) = 0$ for all $\vec{u},\vec{w} \in V$, then $[B]^{\mathcal{B}} = 0$ in any basis and we are done. Moreover, if $B(\vec{v},\vec{v}) = 0$ for all $\vec{v} \in V$ then (using symmetry and bilinearity)

$$\begin{aligned} 2B(\vec{u},\vec{w}) &= B(\vec{u},\vec{w}) + B(\vec{w},\vec{u}) \\ &= B(\vec{u}+\vec{w},\vec{u}+\vec{w}) - B(\vec{u},\vec{u}) - B(\vec{w},\vec{w}) = 0 \end{aligned}$$

for all $\vec{u},\vec{w}$ and we're done once more. So assume otherwise: that there exists $\vec{v}_1 \in V$ such that $B(\vec{v}_1,\vec{v}_1) =: b_1 \neq 0$.

Clearly $B(\vec{v}_1,\vec{v}_1) \neq 0 \implies$ (i) $\mathit{span}\{\vec{v}_1\} \cap V^{\perp\vec{v}_1} = \{0\}$. Furthermore, (ii) $V = \mathit{span}\{\vec{v}_1\} + V^{\perp\vec{v}_1}$: any $\vec{w} \in V$ can be written

$$\vec{w} \;=\; (\vec{w} - B(\vec{w},\vec{v}_1)b_1^{-1}\vec{v}_1) \;+\; B(\vec{w},\vec{v}_1)b_1^{-1}\vec{v}_1$$

[4] The converse is also true: in fact, if $[B]^{\mathcal{B}}$ is a symmetric matrix for *any* $\mathcal{B}$, then B is symmetric. Since $B(\vec{w},\vec{u})$ is a scalar, it equals its own transpose (as a 1×1 matrix), and so $B(\vec{w},\vec{u}) = {}^{t}B(\vec{w},\vec{u}) = {}^{t}({}^{t}[\vec{w}]_{\mathcal{B}}[B]^{\mathcal{B}}[\vec{u}]_{\mathcal{B}}) = {}^{t}[\vec{u}]_{\mathcal{B}}[B]^{\mathcal{B}}[\vec{w}]_{\mathcal{B}} = B(\vec{u},\vec{w})$.

where the second term is in $span\{\vec{v}_1\}$, and the first is in $V^{\perp\vec{v}_1}$ since $B(\text{first term}, \vec{v}_1) = B(\vec{w}, \vec{v}_1) - B(\vec{w}, \vec{v}_1) b_1^{-1} B(\vec{v}_1, \vec{v}_1) = 0$. Combining (i) and (ii), $V = span\{\vec{v}_1\} \oplus V^{\perp\vec{v}_1}$.

Applying the inductive hypothesis to $V^{\perp\vec{v}_1}$ establishes the existence of a basis $\{\vec{v}_2, \ldots, \vec{v}_n\}$ with $B(\vec{v}_i, \vec{v}_j) = b_i \delta_{ij}$ $(i, j \geq 2)$. Combining this with the direct-sum decomposition of V, $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for V. By definition of $V^{\perp\vec{v}_1}$, $B(\vec{v}_1, \vec{v}_j) = B(\vec{v}_j, \vec{v}_1) = 0$ for $j \geq 2$. Therefore $B(\vec{v}_i, \vec{v}_j) = b_i \delta_{ij}$ for $i, j \geq 1$ and we're through. $\square$

Reorder the basis elements so that

$$[B]^{\mathcal{B}} = \mathrm{diag}\{ \underbrace{b_1, \ldots, b_p}_{>0}, \underbrace{b_{p+1}, \ldots, b_{p+q}}_{<0}, 0, \ldots, 0\}$$

and then normalize: take

$$\mathcal{B}' = \left\{ \frac{\vec{v}_1}{\sqrt{b_1}}, \ldots, \frac{\vec{v}_p}{\sqrt{b_p}}; \frac{\vec{v}_{p+1}}{\sqrt{-b_{p+1}}}, \ldots, \frac{\vec{v}_{p+q}}{\sqrt{-b_{p+q}}}; \vec{v}_{p+q+1}, \ldots, \vec{v}_n \right\},$$

so that

$$[B]^{\mathcal{B}'} = \mathrm{diag}\{ \underbrace{1, \ldots, 1}_{p}, \underbrace{-1, \ldots, -1}_{q}, 0, \ldots, 0\}.$$

Suppose that for some other basis $\mathcal{C} = \{\vec{w}_1, \ldots, \vec{w}_n\}$

$$[B]^{\mathcal{C}} = \mathrm{diag}\{ \underbrace{c_1, \ldots, c_{p'}}_{>0}, \underbrace{c_{p'+1}, \ldots, c_{p'+q'}}_{<0}, 0, \ldots, 0\}.$$

Are p, q and p', q' necessarily the same?

Since $[B]^{\mathcal{C}}$ and $[B]^{\mathcal{B}}$ are cogredient they must have the same rank, $p + q = p' + q'$. Now any nonzero $\vec{u} \in span\{\vec{v}_1, \ldots, \vec{v}_p\}$ has $B(\vec{u}, \vec{u}) > 0$, while a nonzero $\vec{u} \in span\{\vec{w}_{p'+1}, \ldots, \vec{w}_n\}$ has $B(\vec{u}, \vec{u}) \leq 0$. Therefore the only intersection of these two spans can be at $\{0\}$, which says that the sum of their dimensions cannot exceed $\dim V$: that is, $p + (n - p') \leq n$, which implies $p \leq p'$. An exactly symmetric argument (interchanging the $\vec{v}$'s and $\vec{w}$'s) shows that $p' \leq p$. So $p = p'$, and $q = q'$ follows immediately. We have proved:

VII.B.6. THEOREM (Sylvester). *Given a symmetric bilinear form B on $V/\mathbb{R}$, there exists a basis $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ for V such that $B(\vec{v}_i, \vec{v}_j) = 0$ for $i \neq j$, and $B(\vec{v}_i, \vec{v}_i) = \left\{ \begin{smallmatrix} 0 \\ \pm 1 \end{smallmatrix} \right.$ where the number of $+1$'s, -1's, and 0's is well-defined (another such choice of basis will not change these numbers).*

The corresponding statement for the matrix of B gives rise to the following

VII.B.7. COROLLARY. *Any symmetric real matrix is cogredient to exactly one matrix of the form* $\mathrm{diag}\{+1, \ldots, +1, -1, \ldots, -1, 0, \ldots, 0\}$.

For a given symmetric bilinear form (or symmetric matrix), we call the # of $+1$'s and -1's in the form guaranteed by the Theorem (or Corollary) the "index of positivity"($= p$) and "index of negativity"($= q$) of B (or $[B]$). Their sum $r = p + q$ is (for obvious reasons) referred to as the rank, and the pair (p, q) (or triple $(p, q, n - r)$, or sometimes the difference $p - q$) is referred to as the **signature** of B (or $[B]$).

Anti-symmetric bilinear forms. Now if $B \colon V \times V \to \mathbb{R}$ is *anti*-symmetric (or "alternating"), i.e.

$$B(\vec{u}, \vec{v}) = -B(\vec{v}, \vec{u}),$$

for all $\vec{u}, \vec{v} \in V$, then $\perp$ is still a symmetric relation. In this case the matrix of B with respect to any basis is skew-symmetric,

$${}^t[B]^{\mathcal{B}} = -[B]^{\mathcal{B}},$$

and the analogue of Sylvester's Theorem VII.B.6 is actually simpler:

VII.B.8. PROPOSITION. *There exists a basis $\mathcal{B}$ such that (in terms of matrix blocks)*

$$[B]^{\mathcal{B}} = \begin{pmatrix} 0 & -\mathbb{I}_m & 0 \\ \mathbb{I}_m & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where $r = 2m \leq n$ is the rank of B.

VII.B.9. REMARK. This matrix can be written more succinctly as

$$[B]^{\mathcal{B}} = \begin{pmatrix} \mathbb{J}_{2m} & 0 \\ 0 & 0 \end{pmatrix},$$

where we define

$$\mathbb{J}_{2m} := \begin{pmatrix} 0 & -\mathbb{I}_m \\ \mathbb{I}_m & 0 \end{pmatrix}. \tag{VII.B.10}$$

PROOF OF PROPOSITION VII.B.8. We induce on n: clearly the result holds ($B = 0$) if $n = 1$, since $B(\vec{v}, \vec{v}) = -B(\vec{v}, \vec{v})$ must be 0.

Now assume the result is known for all dimensions less than n.

If V (of dimension n) contains any vector $\vec{v}$ with $B(\vec{v}, \vec{w}) = 0 \,\forall \vec{w} \in V$, take $\vec{v}_n := \vec{v}$. Let $U \subset V$ be any $(n-1)$-dimensional subspace with $V = \text{span}\{\vec{v}\} \oplus U$, and obtain the rest of the basis by applying induction to U.

If V does not contain such a vector, then the columns of $[B]$ (with respect to *any* basis) are independent and $\det([B]) \neq 0$, with the immediate consequence that $n = 2m$ is even (by Exercise IV.B.4). So choose $\vec{v}_{2m} \in V \backslash \{0\}$ arbitrarily, and $\vec{v}_m \in V$ so that $B(\vec{v}_{2m}, \vec{v}_m) = 1$. Writing $W := \text{span}\{\vec{v}_m, \vec{v}_{2m}\}$, and $W^\perp := \{\vec{v} \in V \mid B(\vec{v}, \vec{w}) = 0 \,\forall \vec{w} \in W\}$, we claim that $V = W \oplus W^\perp$.

Indeed, if $\vec{w} \in W \cap W^\perp$ then $\vec{w} = a\vec{v}_m + b\vec{v}_{2m}$ has $0 = B(\vec{w}, \vec{v}_m) = b$ and $0 = B(\vec{w}, \vec{v}_{2m}) = -a \implies \vec{w} = \vec{0}$. Moreover, given $\vec{v} \in V$ we may consider $\vec{w} := B(\vec{v}, \vec{v}_m)\vec{v}_{2m} - B(\vec{v}, \vec{v}_{2m})\vec{v}_m \in W$ and $\vec{w}_\perp := \vec{v} - \vec{w}$, which satisfies

$$B(\vec{w}_\perp, \vec{v}_m) = B(\vec{v}, \vec{v}_m) - B(\vec{v}, \vec{v}_m)\underbrace{B(\vec{v}_{2m}, \vec{v}_m)}_{1} + B(\vec{v}, \vec{v}_{2m})\underbrace{B(\vec{v}_m, \vec{v}_m)}_{0} = 0$$

and $B(\vec{w}_\perp, \vec{v}_{2m}) =$

$$B(\vec{v}, \vec{v}_{2m}) - B(\vec{v}, \vec{v}_m)\underbrace{B(\vec{v}_{2m}, \vec{v}_{2m})}_{0} + B(\vec{v}, \vec{v}_{2m})\underbrace{B(\vec{v}_m, \vec{v}_{2m})}_{-1} = 0.$$

Thus $\vec{w}_\perp \in W^\perp$, $\vec{v} = \vec{w} + \vec{w}_\perp$, and the claim is proved.

Now apply induction to $W^\perp$ to get the remaining $2m-2$ basis elements $\vec{v}_1, \ldots, \vec{v}_{m-1}, \vec{v}_{m+1}, \ldots, \vec{v}_{2m-1}$. $\square$

Hermitian symmetric forms. Given an n-dimensional *complex* vector space $V/\mathbb{C}$, a *Hermitian-linear*[5] form $H\colon V \times V \to \mathbb{C}$ satisfies

$$H(\alpha\vec{u},\vec{w}) = \bar{\alpha}H(\vec{u},\vec{w})\,, \quad H(\vec{u},\alpha\vec{w}) = \alpha H(\vec{u}\,\vec{w})\,,$$
$$H(\vec{u}_1+\vec{u}_2,\vec{w}) = H(\vec{u}_1,\vec{w}) + H(\vec{u}_2,\vec{w})\,,$$
$$H(\vec{u},\vec{w}_1+\vec{w}_2) = H(\vec{u},\vec{w}_1) + H(\vec{u},\vec{w}_2).$$

It is *Hermitian symmetric* if it also satisfies $H(\vec{u},\vec{w}) = \overline{H(\vec{w},\vec{u})}$; we will often call such forms simply "Hermitian".

The passage to matrices (where as usual $\mathcal{B} = \{\vec{v}_1,\ldots,\vec{v}_n\}$, $\vec{u} = \sum_i u_i\vec{v}_i$, $\vec{w} = \sum_j w_j\vec{v}_j$, $h_{ij} = H(\vec{v}_i,\vec{v}_j)$) looks like

$$H(\vec{u},\vec{w}) = \textstyle\sum_{i,j} \bar{u}_i h_{ij} w_j \;=\; [\vec{u}]^*_{\mathcal{B}}[H]^{\mathcal{B}}[\vec{w}]_{\mathcal{B}};$$

the relevant version of cogredience for coordinate transformations is $S^*[H]S$. If H is Hermitian symmetric then $h_{ij} = \bar{h}_{ji}$; that is, $[H]^{\mathcal{B}} = ([H]^{\mathcal{B}})^*$ is a Hermitian symmetric *matrix*. For such forms/matrices Sylvester holds verbatim. For the purpose of studying inner product spaces in the next two sections, you can think of Hermitian symmetric matrices ($A = {}^t\bar{A}$) as being the "right" complex generalization of the real symmetric matrices (which they include), and similarly for the forms.

But there is an important caveat here: Hermitian forms aren't really bilinear over $\mathbb{C}$, due to their *conjugate*-linearity in the first argument. If we view V as a $2n$-dimensional vector space over $\mathbb{R}$, then they *are* bilinear over $\mathbb{R}$. In fact, if we go that route, then we may view

$$H = B_{\mathrm{re}} + \sqrt{-1}B_{\mathrm{im}} :\ \mathbb{R}^{2n} \times \mathbb{R}^{2n} \to \mathbb{C} = \mathbb{R} \oplus \sqrt{-1}\mathbb{R}$$

as splitting up into *two* bilinear forms over $\mathbb{R}$. Moreover, since

$$\begin{aligned} B_{\mathrm{re}}(\vec{v},\vec{u}) + \sqrt{-1}B_{\mathrm{im}}(\vec{v},\vec{u}) &= H(\vec{v},\vec{u}) = \overline{H(\vec{u},\vec{v})} \\ &= B_{\mathrm{re}}(\vec{u},\vec{v}) - \sqrt{-1}B_{\mathrm{im}}(\vec{u},\vec{v}) \end{aligned}$$

for all $\vec{u},\vec{v} \in \mathbb{R}^{2n}$, B_{re} is symmetric and B_{im} is anti-symmetric.

[5] or *sesquilinear*, from the Latin for "one and a half", since it's only conjugate-linear in the first entry.

Now given a basis $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ for V as a $\mathbb{C}$-vector space, $\tilde{\mathcal{B}} := \{\vec{v}_1, \ldots, \vec{v}_n, \sqrt{-1}\vec{v}_1, \ldots, \sqrt{-1}\vec{v}_n\} =: \{\vec{w}_1, \ldots, \vec{w}_{2n}\}$ is a basis for V as an $\mathbb{R}$-vector space. The matrix of[6]

$$\mathcal{J} := \{\text{multiplication by } \sqrt{-1}\} : \; V \to V$$

is clearly $[\mathcal{J}]_{\tilde{\mathcal{B}}} = \mathbb{J}_{2n}$ (as defined in (VII.B.10)). The complex anti-linearity of H in the first argument gives

$$\begin{aligned} B_{\mathrm{re}}(\mathcal{J}\vec{u}, \vec{v}) + \sqrt{-1}B_{\mathrm{im}}(\mathcal{J}\vec{u}, \vec{v}) &= H(\sqrt{-1}\vec{u}, \vec{v}) = -\sqrt{-1}H(\vec{u}, \vec{v}) \\ &= B_{\mathrm{im}}(\vec{u}, \vec{v}) - \sqrt{-1}B_{\mathrm{re}}(\vec{u}, \vec{v}) \end{aligned}$$

hence

$$B_{\mathrm{re}}(\mathcal{J}\vec{u}, \vec{v}) = B_{\mathrm{im}}(\vec{u}, \vec{v}) \;\; \text{and} \;\; -B_{\mathrm{im}}(\mathcal{J}\vec{u}, \vec{v}) = B_{\mathrm{re}}(\vec{u}, \vec{v}).$$

Writing $\vec{v} = \mathcal{J}\vec{w}$, this gives

$$B_{\mathrm{re}}(\mathcal{J}\vec{u}, \mathcal{J}\vec{w}) = B_{\mathrm{im}}(\vec{u}, \mathcal{J}\vec{w}) = -B_{\mathrm{im}}(\mathcal{J}\vec{w}, \vec{u}) = B_{\mathrm{re}}(\vec{w}, \vec{u}) = B_{\mathrm{re}}(\vec{u}, \vec{w})$$

and similarly $B_{\mathrm{im}}(\mathcal{J}\vec{u}, \mathcal{J}\vec{w}) = B_{\mathrm{im}}(\vec{u}, \vec{w})$. In this way one deduces that giving a Hermitian form on V is *equivalent* to giving a (real) symmetric or antisymmetric form which is "invariant" under the complex structure map J.

Finally, if $[H]^{\mathcal{B}} = \mathrm{diag}\{+1, \ldots, +1, -1, \ldots, -1, 0 \ldots, 0\} =: D$ then one finds that

$$[B_{\mathrm{re}}]^{\tilde{\mathcal{B}}} = \begin{pmatrix} D & 0 \\ 0 & D \end{pmatrix} \;\; \text{and} \;\; [B_{\mathrm{im}}]^{\tilde{\mathcal{B}}} = \begin{pmatrix} 0 & D \\ -D & 0 \end{pmatrix};$$

that is, the signature of B_{re} is exactly "double" that of H.

Quadratic Forms. Now let's return to unambiguously real vector spaces.

VII.B.11. DEFINITION. A *quadratic form* is a function $Q : \; V \to \mathbb{R}$ of the form

$$Q(\vec{u}) = B(\vec{u}, \vec{u}),$$

for some *symmetric* bilinear form B.

[6]Conversely, given a $2n$-dimensional real vector space V with a linear transformation J with $J^2 = -\mathrm{Id}$ (or equivalently, matrix $\mathbb{J}_{2n}$ in some basis), we call J a *complex structure on* V.

Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$, use this basis to expand a given vector $\vec{u} = \sum_i x_i \vec{v}_i$, and set $b_{ij} = B(\vec{v}_i, \vec{v}_j) =$ entries of $[B]^{\mathcal{B}}$; then

$$Q(\vec{u}) = {}^t[\vec{u}]_{\mathcal{B}} [B]^{\mathcal{B}} [\vec{u}]_{\mathcal{B}} = \textstyle\sum_{i,j} x_i \, b_{ij} \, x_j \, .$$

Basically this is a homogeneous quadratic function of several variables, with cross-terms (like $x_2 x_3$). Sylvester effectively says "so long" to the cross-terms. It gives us a new basis $\mathcal{C} = \{\vec{w}_1, \ldots, \vec{w}_n\}$ such that

$$[B]^{\mathcal{C}} = \mathrm{diag}\{ \underbrace{+1, \ldots, +1}_{p} , \underbrace{-1, \ldots, -1}_{q} , 0, \ldots, 0\};$$

expanding $\vec{u} = \sum y_i \vec{w}_i$ in this new basis, we have
(VII.B.12)
$$Q(\vec{u}) = {}^t[\vec{u}]_{\mathcal{C}} [B]^{\mathcal{C}} [\vec{u}]_{\mathcal{C}} = y_1^2 + \ldots + y_p^2 - (y_{p+1}^2 + \ldots + y_{p+q}^2).$$

Application to Calculus. Say $f \colon \mathbb{R}^n \to \mathbb{R}$ (nonlinear) has a stationary (or "critical") point at $0 = (0, \ldots, 0)$,

$$\left.\frac{\partial f}{\partial x_1}\right|_0 = \ldots = \left.\frac{\partial f}{\partial x_n}\right|_0 = 0.$$

Consider its Taylor expansion about 0:

$$f(x_1, \ldots, x_n) = \text{constant} + \sum_{i,j=1}^{n} x_i x_j b_{ij} + \text{higher-order terms},$$

where

$$b_{ij} = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_0 .$$

These are the entries of the so-called "Hessian" matrix of f (evaluated at 0); sometimes $Hess(f)$ is just called the "matrix of second partials" in calculus texts.

Now Sylvester's theorem, applied exactly as above, gives us a linear change of coordinates $(y_1, \ldots, y_n)$ so that the second term in the Taylor series takes exactly the form (VII.B.12). This is otherwise known as completing squares! If $p = n$ then we have $y_1^2 + \cdots + y_n^2$ and a local minimum; if $q = n$ then $-y_1^2 - \cdots - y_n^2$ and a local maximum. The other cases where $p + q = n$ are saddle points, and if

$p+q<n$ (i.e., $Hess(f)|_0$ is not of maximal rank) then one must look to the higher-order terms. Incidentally in this latter case we say 0 is a *degenerate* critical point.

Is there any way you can tell what situation you're in without finding the basis guaranteed by Sylvester's theorem? Certainly if $\det(Hess(f)|_0) \neq 0$ then $p+q=n$ — we have a *nondegenerate* critical point. Where do we go from there? If $n=2$ (notice that this is the situation in your calculus text), then we can use the fact that determinants of cogredient matrices have the same sign: ${}^tSAS = B \implies \det A \cdot (\det S)^2 = \det B$. If $\det(Hess(f)|_0) > 0$ then either $(p,q) = (2,0)$ or $(0,2)$ (a "$++$" or "$--$" situation), which means a local extremum; if $\det(Hess(f)|_0) < 0$ then we have $p=q=1$ ("$+-$") and a saddle point.

More generally ($n \geq 2$) if $Hess(f)|_0$ is diagonalizable via an orthogonal change of basis, that could substitute for Sylvester (how?) *and* tell us the signs. In fact, in the next section we will see that real symmetric matrices can *always* be "orthogonally diagonalized"!

Nondegenerate bilinear forms. A bilinear (or Hermitian) form B on a vector space V is said to be *nondegenerate* if and only if, for every vector $\vec{v} \in V$, there exists $\vec{w} \in V$ such that $B(\vec{v},\vec{w}) \neq 0$. (Equivalently, the matrix of the form has nonzero determinant.)

- A nondegenerate symmetric form B is called **orthogonal**.[7] If the signs in Sylvester's theorem are all positive (resp. negative), B is *positive* (resp. *negative*) *definite*; otherwise, B is *indefinite* of signature (p,q), where $p+q=n$.
- A nondegenerate anti-symmetric form B is termed **symplectic**. In this case $\dim(V)$ is necessarily even: $n=2m$, and (in some basis $\mathcal{B}$) $[B]^{\mathcal{B}} = \mathbb{J}_{2m}$.
- A nondegenerate Hermitian form H is called **unitary**. The same terminology applies as in the orthogonal case.

[7]Calling a *form* "orthogonal" or "unitary" is a bit nonstandard, but is more consistent with the standard use of "symplectic".

VII.B.13. EXAMPLE. On $\mathbb{R}^4$ with "spacetime" coordinates (x, y, z, t), we define an indefinite orthogonal form of signature $(3, 1)$ via its quadratic form $x^2 + y^2 + z^2 - t^2$. The resulting "Minkowski space" is closely associated with the theory of special relativity.

Nondegenerate symplectic and indefinite-orthogonal forms also play a central role in the topology of projective manifolds.

*   *   *

We conclude this section with a brief (and somewhat sketchy) account of the simplest of the so-called "classical groups". This material will not be used in the remainder of the text.

Linear algebraic groups. These are those subgroups of $\mathrm{GL}_n(\mathbb{C})$ or $\mathrm{GL}_n(\mathbb{R})$, the **general linear groups** of invertible $n \times n$ matrices with entries in $\mathbb{C}$ or $\mathbb{R}$, that are defined by "linear algebraic conditions". We'll only be interested in examples of *reductive* such groups, which are the ones with the property that whenever a vector subspace W is closed under their action on a vector space V, there is another subspace W' also closed under this action, such that $V = W \oplus W'$. In general, one can show (Chevalley's Theorem) that the reductive linear algebraic groups are the subgroups of GL_n defined by the property of fixing a subalgebra of the tensor algebra[8]

$$\oplus_{a,b} V^{\otimes a} \otimes (V^{\vee})^{\otimes b},$$

where $V = \mathbb{C}^n$ resp. $\mathbb{R}^n$. But that is a subject for a different course; here we'll just define a few of these groups, using only determinants and nondegenerate bilinear forms.

For B orthogonal resp. symplectic, the corresponding **orthogonal** resp. **symplectic group** consists of all $g \in \mathrm{GL}_n(\mathbb{F})$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$)

[8]We have not discussed tensor products of vector spaces, but they're easy to define: given vector spaces V and W with bases $\{\vec{v}_i\}_{i=1}^n$ and $\{\vec{w}_j\}_{j=1}^m$, you simply take $V \otimes W$ to be the nm-dimensional vector space with basis $\{\vec{v}_i \otimes \vec{w}_j\}_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$. (These symbols obey the distributive property.) These aren't unfamiliar objects either. You could try to prove, for instance, that $V^{\vee} \otimes W$ is the vector space of linear transformations from V to W, or that a bilinear form is an element of $(V^{\vee})^{\otimes 2}$.

satisfying

(VII.B.14) $$B(g\vec{v}, g\vec{w}) = B(\vec{v}, \vec{w}) \;\; \forall \vec{v}, \vec{w} \in \mathbb{F}^n.$$

(We write $\mathrm{Sp}_n(\mathbb{F}, B)$ resp. $\mathrm{O}_n(\mathbb{F}, B)$ for these groups.) By Proposition VII.B.8, all the symplectic groups are isomorphic (via conjugation by a change-of-basis matrix, see Exercise 8) to

$$\mathrm{Sp}_{2m}(\mathbb{F}) := \{g \in \mathrm{GL}_{2m}(\mathbb{F}) \mid {}^t g \mathbb{J}_{2m} g = \mathbb{J}_{2m}\}$$

for $\mathbb{F} = \mathbb{C}$ resp. $\mathbb{R}$. Similarly, by Sylvester's Theorem, the real orthogonal groups are isomorphic to one of the

$$\mathrm{O}(p, q) := \{g \in \mathrm{GL}_{p+q}(\mathbb{R}) \mid {}^t g \mathbb{I}_{p,q} g = \mathbb{I}_{p,q}\},$$

where $\mathbb{I}_{p,q} := \mathrm{diag}\{\mathbb{I}_p, -\mathbb{I}_q\}$. Note that $\mathrm{O}(p, q) \cong \mathrm{O}(q, p)$, and $\mathrm{O}(n, 0)$ is written $\mathrm{O}_n(\mathbb{R})$; these are called *indefinite* resp. *definite* orthogonal groups. All the complex orthogonal groups are isomorphic to

$$\mathrm{O}_n(\mathbb{C}) := \{g \in \mathrm{GL}_n(\mathbb{C}) \mid {}^t g g = \mathbb{I}_n\}.$$

The **unitary groups** $\mathrm{U}_n(H)$ are defined by the property of preserving a Hermitian form: they consist of those $g \in \mathrm{GL}_n(\mathbb{C})$ with

(VII.B.15) $$H(g\vec{u}, g\vec{v}) = H(\vec{u}, \vec{v}) \quad \forall \vec{u}, \vec{v} \in \mathbb{C}^n,$$

and are each conjugate to one of the

$$\mathrm{U}(p, q) := \{g \in \mathrm{GL}_{p+q}(\mathbb{C}) \mid g^* \mathbb{I}_{p,q} g = \mathbb{I}_{p,q}\}.$$

(Write $\mathrm{U}(n, 0) =: U(n)$ for the *definite* unitary group.) But this is a little deceptive, as (VII.B.15) is not a $\mathbb{C}$-linear condition: the $\mathrm{U}(p, q)$ are actually *real* linear algebraic groups. If (as above) we think of $\mathbb{C}^n$ as $\mathbb{R}^{2n}$, and write $H = B_{\mathrm{re}} + \sqrt{-1} B_{\mathrm{im}}$, then I claim that

$$\mathrm{U}_n(H) = \mathrm{O}_{2n}(\mathbb{R}, B_{\mathrm{re}}) \cap \mathrm{Sp}_{2n}(\mathbb{R}, B_{\mathrm{im}})$$

as a subgroup of $\mathrm{GL}_{2n}(\mathbb{R})$. The "$\subseteq$" inclusion is clear, since to preserve H, $\gamma \in \mathrm{GL}_{2n}(\mathbb{R})$ must preserve its real and imaginary parts. Conversely, we just need to show that preserving B_{re} and B_{im} also implies that γ "comes from" an element of $\mathrm{GL}_n(\mathbb{C})$, or equivalently commutes with $\mathcal{J}$ (cf. Exercise 6 below). This follows from

$$B_{\text{re}}(g\mathcal{J}\vec{u},g\vec{v}) = B_{\text{re}}(\mathcal{J}\vec{u},\vec{v}) = B_{\text{im}}(\vec{u},\vec{v}) = B_{\text{im}}(g\vec{u},g\vec{v}) = B_{\text{re}}(\mathcal{J}g\vec{u},g\vec{v})$$

and nondegeneracy of B_{re}.

The **special linear groups** $\mathrm{SL}_n(\mathbb{F})$ simply consist of the elements of $\mathrm{GL}_n(\mathbb{F})$ with determinant 1. They contain the symplectic groups (see Exercise 5), from which it also follows that $\mathrm{SL}_{2n}(\mathbb{R})$ contains $\mathrm{U}_n(H)$. On the other hand, $\mathrm{SL}_n(\mathbb{C})$ does *not* contain $\mathrm{U}_n(H)$; *intersecting* them gives the **special unitary groups** $\mathrm{SU}_n(H)$ (and $\mathrm{SU}(p,q)$). The **special orthogonal groups** (denoted $\mathrm{SO}_n(\mathbb{F},B)$, $\mathrm{SO}(p,q)$, etc.) are given by intersecting the orthogonal and special linear groups.

Jordan decomposition. Let G be one of the above linear algebraic groups (over $\mathbb{R}$ or $\mathbb{C}$).[9] Since G is a subgroup of $\mathrm{GL}_n(\mathbb{F})$ for some n, we may think of elements of G as invertible matrices (with additional constraints) acting on $\mathbb{R}^n$ or $\mathbb{C}^n$. We close this section with a fundamental result which is closely related to Jordan normal form.

VII.B.16. DEFINITION. An element $g \in G$ is
- **semisimple** if g is diagonalizable over $\mathbb{C}$
- **unipotent** if a power of $(g - \mathbb{I}_n)$ is zero.

VII.B.17. THEOREM. *Every $g \in G$ may be written* UNIQUELY *as a product*

$$g = g_{\text{ss}} g_{\text{un}} \tag{VII.B.18}$$

of COMMUTING *semisimple and unipotent elements* OF G.

PROOF. Begin with the case of $G = \mathrm{GL}_n(\mathbb{C})$. A Jordan form matrix J admits such a decomposition, since each block decomposes:

$$\begin{pmatrix} \sigma & & & \\ 1 & \sigma & & \\ & \ddots & \ddots & \\ & & 1 & \sigma \end{pmatrix} = \begin{pmatrix} \sigma & & & \\ & \sigma & & \\ & & \ddots & \\ & & & \sigma \end{pmatrix} \begin{pmatrix} 1 & & & \\ \sigma^{-1} & 1 & & \\ & \ddots & \ddots & \\ & & \sigma^{-1} & 1 \end{pmatrix}.$$

[9]Even more generally, Theorem VII.B.17 will hold for all reductive linear algebraic groups over a perfect field, though that is obviously beyond our scope here.

By Theorem VI.E.11 (the existence part), we have

$$g = \gamma J(g)\gamma^{-1} = \gamma J_{\text{ss}}J_{\text{un}}\gamma^{-1} = \underbrace{\gamma J_{\text{ss}}\gamma^{-1}}_{=:g_{\text{ss}}}\underbrace{\gamma J_{\text{un}}\gamma^{-1}}_{=:g_{\text{un}}},$$

and $g_{\text{ss}}, g_{\text{un}}$ commute because $J_{\text{ss}}, J_{\text{un}}$ do. Conversely, *given* a decomposition (VII.B.18) and $\vec{v} \in E_\sigma(g_{\text{ss}})$, we have

$$(g - \sigma\mathbb{I}_n)^n\vec{v} = (g_{\text{un}}g_{\text{ss}} - \sigma\mathbb{I}_n)^n\vec{v} = \sigma^n(g_{\text{un}} - \mathbb{I}_n)^n\vec{v} = \vec{0}$$

since g_{ss} and g_{un} commute, and so $\vec{v} \in \widetilde{E}_\sigma(g)$. Writing

$$\mathbb{C}^n = \oplus_{j=1}^s E_{\sigma_j}(g_{\text{ss}}),$$

this shows that $E_{\sigma_j}(g_{\text{ss}}) \subseteq \widetilde{E}_{\sigma_j}(g)$. Since the dimensions of $\widetilde{E}_{\sigma_j}(g)$ cannot sum to *more* than n, these inclusions are equalities. Therefore, the eigenspaces of g_{ss}, and so g_{ss} itself (and thus $g_{\text{un}} = gg_{\text{ss}}^{-1}$), are determined uniquely by g.

It remains to show that if g belongs to one of the classical groups $G \leq \text{GL}_n(\mathbb{C})$ above, then g_{ss} and g_{un} belong as well. First, if g has real entries, then $\widetilde{\ker}(\sigma\mathbb{I} - g)\,(= E_\sigma(g_{\text{ss}}))$ and $\widetilde{\ker}(\bar{\sigma}\mathbb{I} - g)\,(= E_{\bar{\sigma}}(g_{\text{ss}}))$ are complex-conjugate, from which one deduces that g_{ss} (hence g_{un}) is real. Next, if $\det(g) = 1$, then since the determinant of a unipotent matrix is *always* 1, $\det(g_{\text{ss}}) = \det(g_{\text{un}}^{-1}) = 1$. Finally, if g preserves a nondegenerate symmetric or alternating bilinear form B, we claim that $g_{\text{ss}}, g_{\text{un}}$ do too. (This will finish the proof, as all the above groups are "cut out" of $\text{GL}_n(\mathbb{C})$ by some combination of these constraints.)

Write $V_i := \widetilde{E}_{\sigma_i}(g) = E_{\sigma_i}(g_{\text{ss}})$. Since any vector decomposes into a sum of vectors in these V_i, it will suffice to show that

(VII.B.19) $$B(g\vec{v}, g\vec{w}) = B(\vec{v}, \vec{w}) \quad \forall i,\, j,\, \vec{v} \in V_i,\, \vec{w} \in V_j$$

implies

(VII.B.20) $$B(g_{\text{ss}}\vec{v}, g_{\text{ss}}\vec{w}) = B(\vec{v}, \vec{w}) \quad \forall i,\, j,\, \vec{v} \in V_i,\, \vec{w} \in V_j.$$

The latter is equivalent to

(VII.B.21) $$B(V_i, V_j) = 0 \text{ if } \sigma_i\sigma_j \neq 1,$$

since LHS(VII.B.20) is just $\sigma_i\sigma_j B(\vec{v},\vec{w})$. Suppose $\sigma_i\sigma_j \neq 1$; then $(\sigma_i^{-1}\mathbb{I} - g)$ is invertible on V_j, and g invertible on V_i. Given $\vec{v} \in V_i$, $\vec{w} \in V_j$, write $\vec{w}_\ell := (\sigma_i^{-1}\mathbb{I} - g)^{-\ell}\vec{w} \in V_j$ and $\vec{v}_\ell := g^{-\ell}\vec{v} \in V_i$; and note that $(g - \sigma_i\mathbb{I})^k V_i = \{\vec{0}\}$ for some k. Applying (VII.B.19),

$$\begin{aligned} B(\vec{v},\vec{w}) &= B\left(\vec{v}, (\sigma_i^{-1}\mathbb{I} - g)\vec{w}_1\right) = \sigma_i^{-1}B(\vec{v},\vec{w}_1) - B(\vec{v}, g\vec{w}_1) \\ &= \sigma_i^{-1}B(\vec{v},\vec{w}_1) - B(g^{-1}\vec{v},\vec{w}_1) = \sigma_i^{-1}B\left((g-\sigma_i\mathbb{I})\vec{v}_1, \vec{w}_1\right) \\ &= \cdots = \sigma_i^{-k}B\left((g-\sigma_i\mathbb{I})^k\vec{v}_k, \vec{w}_k\right) = 0 \end{aligned}$$

establishes (VII.B.21) and we are done. $\square$

Exercises

(1) A quadratic form on $\mathbb{R}^3$ is given by $Q(x,y,z) = x^2 + 3y^2 + z^2 - 4xy + 2xz - 2yz$.
(a) Write the matrix of the corresponding symmetric bilinear form B with respect to $\hat{e}$. (Be careful: if -4 is an entry in your matrix, it isn't quite correct.)
(b) Find S such that ${}^tS[B]^{\hat{e}}S$ is diagonal by following the steps in the proof of Sylvester's theorem.
(c) What is the signature of B?

(2) Consider the vector space $P_2(\mathbb{R})$ of polynomials of degree ≤ 2 with (symmetric) bilinear form

$$B(f,g) := \int_0^1 f(t)g(t)dt.$$

(a) Find a basis $\mathcal{B}$ of $P_2(\mathbb{R})$ in which $[B]^{\mathcal{B}} = \mathbb{I}_3$. [Hint: work as if B were the dot product and use "Gram-Schmidt", starting from the basis $\{1, t, t^2\}$.]
(b) Let $T : P_2(\mathbb{R}) \to P_2(\mathbb{R})$ be the "shift" operator $Tf(t) = f(t-1)$. Compute $[T]_{\mathcal{B}}$ (where $\mathcal{B}$ is the basis found in part (a)).

(3) Find a basis of the vector space(!) of all alternating bilinear forms on $\mathbb{R}^n$. What's the dimension? [Hint: you could use matrices.]

(4) Write out the proof of "Sylvester's theorem" for Hermitian forms.

(5) Show that $\mathrm{Sp}_{2m}(\mathbb{F})$ is in fact a subgroup of $\mathrm{SL}_{2m}(\mathbb{F})$. [Hint: use Jordan to reduce to g_{ss}.]

(6) Deduce that the image of $\mathrm{GL}_n(\mathbb{C}) \to \mathrm{GL}_{2n}(\mathbb{R})$, given by regarding $\mathbb{C}^n$ as $\mathbb{R}^{2n}$ via

$$(z_1, \ldots, z_n) \mapsto (\mathrm{Re}(z_1), \ldots, \mathrm{Re}(z_n), \mathrm{Im}(z_1), \ldots, \mathrm{Im}(z_n)),$$

consists of those elements commuting with $\mathbb{J}_{2n}$.

(7) (i) Let B be an orthogonal form on $\mathbb{R}^n$, and $\vec{w} \in \mathbb{R}^n$ be such that $B(\vec{w}, \vec{w}) = 2$. Show that the (matrix of the) **reflection** $\mu(\vec{v}) := \vec{v} - B(\vec{w}, \vec{v})\vec{w}$ belongs to $\mathrm{O}_n(\mathbb{R}, B)$.
(ii) Let B be a symplectic form on $\mathbb{R}^{2m}$, $\vec{w} \in \mathbb{R}^{2m}$, and $c \in \mathbb{R}$. Show that the (matrix of the) **transvection** $\tau(\vec{v}) := \vec{v} - cB(\vec{w}, \vec{v})\vec{w}$ belongs to $\mathrm{Sp}_{2m}(\mathbb{R}, B)$.
[Note: standard results in abstract algebra (beyond our scope here) state that every element of an orthogonal resp. symplectic group is a product of reflections resp. transvections.]

(8) (a) Let B be a symplectic form on $\mathbb{R}^{2m}$. Use Prop. VII.B.8 to show that there exists an invertible matrix $S \in M_{2m}(\mathbb{R})$ such that sending $g \mapsto SgS^{-1}$ produces a bijection[10] between $\mathrm{Sp}_{2m}(\mathbb{R}, B)$ and $\mathrm{Sp}_{2m}(\mathbb{R})$ (as defined above).
(b) Let B be an orthogonal form on $\mathbb{R}^n$. Use Sylvester's Theorem VII.B.6 to produce an isomorphism (as in (a)) between $\mathrm{O}_n(\mathbb{R}, B)$ and some $\mathrm{O}(p, q)$ (with $p + q = n$).

[10] In fact, a group isomorphism, meaning that it respects products and inverses.