

14.2 Congruence

The matrix of a form in Definition 14.1.2 depends on a basis, and a basis is a choice. As with linear maps in Chapter 4, the useful question is therefore not “what is the matrix?” but “which matrices arise from the same form?”. The answer here is $\mathbf{P}^\top \mathbf{A} \mathbf{P}$ rather than $\mathbf{P}^{-1} \mathbf{A} \mathbf{P}$, and one symbol’s difference changes the whole theory: the new relation keeps the rank and the symmetry type of a matrix and throws its eigenvalues away. This section proves the change-of-basis rule, names the relation, sorts the invariants from the casualties, and settles the debt Chapters 12 and 13 ran up when they kept postponing all of this.

Throughout, V is a vector space over an arbitrary field F with $\dim V = n \geq 1$, and β is a bilinear form on V . **No hypothesis on the characteristic of F is needed in this section.**

14.2.1 The change-of-basis rule for a form

Two bases $\mathcal{B}$ and $\mathcal{B}'$ of V give two matrices for one form. The link between them is the change-of-coordinates matrix of Definition 4.2.1, and the computation is three lines, because Theorem 14.1.1 (a) already expresses the form through coordinates.

Theorem 14.2.1. Change of Basis for a Form

Let β be a bilinear form on a finite-dimensional V , let $\mathcal{B}$ and $\mathcal{B}'$ be ordered bases of V , and let $\mathbf{P} = [\text{id}]_{\mathcal{B}'}^{\mathcal{B}}$ be the change-of-coordinates matrix from $\mathcal{B}'$ to $\mathcal{B}$. Then

$$[\beta]_{\mathcal{B}'}^{\mathcal{B}'} = \mathbf{P}^\top [\beta]_{\mathcal{B}}^{\mathcal{B}} \mathbf{P}.$$

Idea. There is nothing to discover: write the form in $\mathcal{B}$ -coordinates, substitute $[v]_{\mathcal{B}} = \mathbf{P}[v]_{\mathcal{B}'}$ in both slots, and read off the matrix that appears between the two coordinate vectors. The uniqueness clause of Theorem 14.1.1 then says that matrix is the one we want, with no further checking.

Proof. Write $\mathbf{A} = [\beta]_{\mathcal{B}}^{\mathcal{B}}$. By Theorem 4.2.1 (a), $[v]_{\mathcal{B}} = \mathbf{P}[v]_{\mathcal{B}'}$ for every $v \in V$. So for all $u, v \in V$, using Theorem 14.1.1 (a) and then Theorem 0.9.5 (part 4),

$$\begin{aligned} \beta(u, v) &= [u]_{\mathcal{B}}^\top \mathbf{A} [v]_{\mathcal{B}} \\ &= (\mathbf{P}[u]_{\mathcal{B}'})^\top \mathbf{A} (\mathbf{P}[v]_{\mathcal{B}'}) \\ &= [u]_{\mathcal{B}'}^\top (\mathbf{P}^\top \mathbf{A} \mathbf{P}) [v]_{\mathcal{B}'} \end{aligned}$$

Thus $\mathbf{P}^\top \mathbf{A} \mathbf{P}$ is a matrix with the property that characterizes $[\beta]_{\mathcal{B}'}^{\mathcal{B}'}$, and by the uniqueness in Theorem 14.1.1 (b) it is $[\beta]_{\mathcal{B}'}^{\mathcal{B}'}$. This proves the theorem. ■

Notice that $\mathbf{P}$ is the same matrix as in Theorem 4.2.3: the change-of-coordinates matrix from the **new** basis to the **old** one, whose columns are the new basis vectors written in old coordinates. Only the way it acts has changed, from $\mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ to $\mathbf{P}^\top \mathbf{A} \mathbf{P}$.

Example. Killing the cross term

On $\mathbb{R}^2$, let $\beta(\mathbf{x}, \mathbf{y}) = x_1y_1 + 2x_1y_2 + 2x_2y_1 + x_2y_2$, so that

$$\mathbf{A} = [\beta]_{\mathcal{E}}^{\mathcal{E}} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Compute $[\beta]_{\mathcal{B}'}^{\mathcal{B}'}$ for $\mathcal{B}' = ((1, 0), (-2, 1))$, both by [Theorem 14.2.1](#) and directly from Definition 14.1.2.

Solution. The columns of $\mathbf{P} = [\text{id}]_{\mathcal{B}'}^{\mathcal{E}}$ are the vectors of $\mathcal{B}'$ in standard coordinates, so

$$\mathbf{P} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{AP} = \begin{pmatrix} 1 & 0 \\ 2 & -3 \end{pmatrix},$$

and therefore

$$\mathbf{P}^{\top} \mathbf{AP} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -3 \end{pmatrix}.$$

Directly: with $\mathbf{v}'_1 = (1, 0)$ and $\mathbf{v}'_2 = (-2, 1)$, the four values are $\beta(\mathbf{v}'_1, \mathbf{v}'_1) = 1$, $\beta(\mathbf{v}'_1, \mathbf{v}'_2) = \beta(\mathbf{v}'_2, \mathbf{v}'_1) = 0$, and $\beta(\mathbf{v}'_2, \mathbf{v}'_2) = -3$. The two computations agree.

The new basis was not guessed. In coordinates the form reads $q(\mathbf{x}) = x_1^2 + 4x_1x_2 + x_2^2 = (x_1 + 2x_2)^2 - 3x_2^2$, a perfect square plus a leftover, and the substitution $y_1 = x_1 + 2x_2$, $y_2 = x_2$ — that is, $\mathbf{x} = \mathbf{P}\mathbf{y}$ — is what completing the square asks for. A later section turns this into a theorem and an algorithm.

14.2.2 Congruent matrices

Every invertible $\mathbf{P}$ is a change-of-coordinates matrix between some pair of bases (Proposition 4.2.2), so [Theorem 14.2.1](#) says exactly this: the matrices of one form, over all bases, are the matrices $\mathbf{P}^{\top} \mathbf{AP}$ with $\mathbf{P}$ invertible. That set deserves a name, and the name is the organizing idea of the chapter.

Two matrices are congruent when they record the same form in two bases.

Definition 14.2.1. Congruent Matrices

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$. We say $\mathbf{A}$ is **congruent** to $\mathbf{B}$, written $\mathbf{A} \simeq \mathbf{B}$, if there **exists** an **invertible** $\mathbf{P} \in M_n(F)$ such that

$$\mathbf{B} = \mathbf{P}^{\top} \mathbf{AP}.$$

In words: congruence allows one matrix to be sandwiched by an invertible matrix and its transpose. The word **invertible** is essential; without it, $\mathbf{A} \mapsto \mathbf{P}^{\top} \mathbf{AP}$ with $\mathbf{P} = \mathbf{O}$ would relate everything to $\mathbf{O}$. The **transpose**, not the inverse, is what distinguishes the relation from similarity (Definition 4.2.2), and the symbol $\simeq$ is chosen to sit near $\sim$ without being it.

Three immediate readings. First, by [Theorem 14.2.1](#), $\mathbf{A} \simeq \mathbf{B}$ if and only if $\mathbf{A}$ and $\mathbf{B}$ are the matrices of a single bilinear form in two bases. Second, in the language of quadratic expressions on F^n , $\mathbf{A} \simeq \mathbf{B}$ says that the substitution $\mathbf{x} = \mathbf{P}\mathbf{y}$ turns $\mathbf{x}^{\top} \mathbf{Ax}$ into $\mathbf{y}^{\top} \mathbf{By}$: congruence is

invertible change of variables in a quadratic expression. Third, over $\mathbb{R}$ it is the relation $\mathbf{A} \mapsto \mathbf{S}^* \mathbf{A} \mathbf{S}$ of Chapter 13, because there the star is the transpose.

Example. Congruences and a noncongruence

- (a) $\mathbf{I}_2 \simeq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ over any field.
- (b) $\mathbf{I}_2 \simeq \text{diag}(4, 1)$ over $\mathbb{Q}$.
- (c) $\mathbf{A} \simeq \mathbf{A}$ for every $\mathbf{A}$.
- (d) $\mathbf{I}_2 \not\simeq \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ over any field.

Solution.

- (a) Take $\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, which is invertible with determinant 1. Then $\mathbf{P}^\top \mathbf{I}_2 \mathbf{P} = \mathbf{P}^\top \mathbf{P} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.
- (b) Take $\mathbf{P} = \text{diag}(2, 1)$, invertible over $\mathbb{Q}$. Then $\mathbf{P}^\top \mathbf{I}_2 \mathbf{P} = \text{diag}(4, 1)$. Scaling a basis vector by c scales the corresponding diagonal entry by c^2 , which is the single most useful congruence to have in mind.
- (c) Take $\mathbf{P} = \mathbf{I}_n$.
- (d) Suppose $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \mathbf{P}^\top \mathbf{I}_2 \mathbf{P} = \mathbf{P}^\top \mathbf{P}$ for some $\mathbf{P}$. Then, by Theorem 0.9.5, $(\mathbf{P}^\top \mathbf{P})^\top = \mathbf{P}^\top \mathbf{P}$, so the matrix would be symmetric; it is not. This is the nonexample by minimal change: one entry of $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ in (a) is altered, and congruence to $\mathbf{I}_2$ is lost — not for a subtle reason, but because the altered matrix is not the matrix of a symmetric form in any basis. The general statement is [Proposition 14.2.4](#) below.

Before using the word “congruent” in both directions, we owe the usual check.

Proposition 14.2.2. Congruence is an equivalence relation

Congruence is an equivalence relation on $M_n(F)$ (Definition 0.5.2).

Proof. Reflexive: $\mathbf{A} = \mathbf{I}_n^\top \mathbf{A} \mathbf{I}_n$.

Symmetric: suppose $\mathbf{B} = \mathbf{P}^\top \mathbf{A} \mathbf{P}$ with $\mathbf{P}$ invertible, and put $\mathbf{Q} = \mathbf{P}^{-1}$, which is invertible. By Theorem 0.9.5 (part 4), $\mathbf{Q}^\top \mathbf{P}^\top = (\mathbf{P} \mathbf{Q})^\top = \mathbf{I}_n^\top = \mathbf{I}_n$, so

$$\mathbf{Q}^\top \mathbf{B} \mathbf{Q} = (\mathbf{Q}^\top \mathbf{P}^\top) \mathbf{A} (\mathbf{P} \mathbf{Q}) = \mathbf{I}_n \mathbf{A} \mathbf{I}_n = \mathbf{A}.$$

Transitive: if $\mathbf{B} = \mathbf{P}^\top \mathbf{A} \mathbf{P}$ and $\mathbf{C} = \mathbf{Q}^\top \mathbf{B} \mathbf{Q}$ with $\mathbf{P}, \mathbf{Q}$ invertible, then $\mathbf{P} \mathbf{Q}$ is invertible and $\mathbf{C} = \mathbf{Q}^\top \mathbf{P}^\top \mathbf{A} \mathbf{P} \mathbf{Q} = (\mathbf{P} \mathbf{Q})^\top \mathbf{A} (\mathbf{P} \mathbf{Q})$. This proves the proposition. ■

So the matrices in $M_n(F)$ fall into congruence classes, and the classification problem of this chapter is to describe them. The rest of the section collects the tools for telling classes apart.

14.2.3 What congruence preserves

Three invariants come out of the definition with almost no work. The first is the rank, which we already know is the co-dimension of the radical (Proposition 14.1.2).

Theorem 14.2.3. Congruence preserves rank

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ with $\mathbf{A} \simeq \mathbf{B}$. Then $\text{rank } \mathbf{A} = \text{rank } \mathbf{B}$. Consequently, whether the associated form is nondegenerate is a property of the congruence class, and it is legitimate to speak of the **rank of a bilinear form**, meaning the rank of its matrix in any basis.

Proof. Write $\mathbf{B} = \mathbf{P}^\top \mathbf{A} \mathbf{P}$ with $\mathbf{P}$ invertible. Then $\mathbf{P}^\top$ is invertible, with inverse $(\mathbf{P}^{-1})^\top$ (Theorem 0.9.5). Multiplying by invertible matrices on either side does not change the rank (Theorem 3.5.9), so $\text{rank } \mathbf{B} = \text{rank}(\mathbf{A} \mathbf{P}) = \text{rank } \mathbf{A}$. Nondegeneracy is the condition $\text{rank } \mathbf{A} = n$ by Proposition 14.1.2 (c), so it too depends only on the class. The last sentence follows because, by Theorem 14.2.1, any two matrices of one form are congruent. ■

The second invariant is the symmetry type. Here “alternating” is used in the matrix sense that Definition 7.2.2 suggests: a **zero diagonal** on top of skew-symmetry. The next section separates the three conditions carefully; for now we only need that congruence respects each of them.

Proposition 14.2.4. Congruence preserves the symmetry type

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ with $\mathbf{B} = \mathbf{P}^\top \mathbf{A} \mathbf{P}$ for an invertible $\mathbf{P}$.

- (a) If $\mathbf{A}^\top = \mathbf{A}$, then $\mathbf{B}^\top = \mathbf{B}$.
- (b) If $\mathbf{A}^\top = -\mathbf{A}$, then $\mathbf{B}^\top = -\mathbf{B}$.
- (c) If $\mathbf{A}^\top = -\mathbf{A}$ and every diagonal entry of $\mathbf{A}$ is 0, the same holds for $\mathbf{B}$.

Each implication is an equivalence, since congruence is symmetric (Proposition 14.2.2).

Proof. By Theorem 0.9.5, $\mathbf{B}^\top = (\mathbf{P}^\top \mathbf{A} \mathbf{P})^\top = \mathbf{P}^\top \mathbf{A}^\top \mathbf{P}$.

- (a) If $\mathbf{A}^\top = \mathbf{A}$, this is $\mathbf{P}^\top \mathbf{A} \mathbf{P} = \mathbf{B}$.
- (b) If $\mathbf{A}^\top = -\mathbf{A}$, it is $\mathbf{P}^\top (-\mathbf{A}) \mathbf{P} = -\mathbf{B}$.
- (c) Assume the hypothesis of (c). Then $\mathbf{B}$ is skew-symmetric by (b), so only the diagonal is left to check. For any $\mathbf{x} \in F^n$, grouping the terms of the double sum in pairs,

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = \sum_{i < j} (a_{ij} + a_{ji}) x_i x_j + \sum_i a_{ii} x_i^2 = 0,$$

since $a_{ij} + a_{ji} = 0$ by skew-symmetry and $a_{ii} = 0$ by hypothesis. Writing $\mathbf{p}_1, \dots, \mathbf{p}_n$ for the columns of $\mathbf{P}$, the i -th diagonal entry of $\mathbf{B} = \mathbf{P}^\top \mathbf{A} \mathbf{P}$ is $b_{ii} = \mathbf{e}_i^\top \mathbf{P}^\top \mathbf{A} \mathbf{P} \mathbf{e}_i = \mathbf{p}_i^\top \mathbf{A} \mathbf{p}_i$, which is 0 by the display. This proves the proposition. ■

The third invariant is the determinant, but only in a weakened sense: congruence multiplies it by a square.

Theorem 14.2.5. Congruence changes the determinant by a square

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ with $\mathbf{B} = \mathbf{P}^\top \mathbf{A} \mathbf{P}$ for an invertible $\mathbf{P}$. Then

$$\det \mathbf{B} = (\det \mathbf{P})^2 \det \mathbf{A},$$

with $\det \mathbf{P} \neq 0$. In particular $\det \mathbf{A} = 0$ if and only if $\det \mathbf{B} = 0$, and when they are nonzero, $\det \mathbf{B} = c^2 \det \mathbf{A}$ for some $c \in F \setminus \{0\}$.

Proof. By Theorem 7.5.4 and Theorem 7.4.5,

$$\det \mathbf{B} = \det(\mathbf{P}^\top) \det \mathbf{A} \det \mathbf{P} = (\det \mathbf{P})^2 \det \mathbf{A}.$$

Since $\mathbf{P}$ is invertible, $\det \mathbf{P} \neq 0$ (Theorem 7.5.2), so $(\det \mathbf{P})^2 \neq 0$ and the two determinants vanish together. Taking $c = \det \mathbf{P}$ gives the last claim. ■

The determinant itself is therefore **not** a congruence invariant — [Example \(Congruences and a noncongruence\)](#) (b) moved it from 1 to 4. What is invariant is the determinant read modulo squares. Write $F^\times = F \setminus \{0\}$, and for $a, b \in F^\times$ declare a and b equivalent when $a = c^2 b$ for some $c \in F^\times$. This is an equivalence relation on $F^\times$ (Definition 0.5.2): it is reflexive with $c = 1$, symmetric with c^{-1} , and transitive by multiplying the two scalars. Its classes are called **square classes**, and the set of them is commonly written $F^\times / (F^\times)^2$.

Definition 14.2.2. Discriminant

Let β be a **nondegenerate** bilinear form on an n -dimensional V . The **discriminant** of β is the square class of $\det[\beta]_{\mathcal{B}}^{\mathcal{B}}$ in $F^\times$, for any ordered basis $\mathcal{B}$ of V .

This is well defined by [Theorem 14.2.5](#) and [Theorem 14.2.1](#): another basis changes the determinant by the square $(\det \mathbf{P})^2$, hence not its class. The discriminant is coarse but genuinely useful, and how much it sees depends entirely on the field. Over $\mathbb{C}$ every nonzero number is a square, so there is only one class and the discriminant says nothing. Over $\mathbb{R}$ there are two, $\{t > 0\}$ and $\{t < 0\}$, so the discriminant records one bit: the sign of the determinant. Over $\mathbb{Q}$ there are infinitely many, and the discriminant is a strong invariant.

Quick check. Over $\mathbb{R}$, are $\text{diag}(1, 1)$ and $\text{diag}(-1, -1)$ distinguished by their discriminants?

Solution. No. Both determinants equal 1, so the discriminants agree. And yet the two matrices are not congruent over $\mathbb{R}$. Suppose $\text{diag}(-1, -1) = \mathbf{P}^\top \mathbf{I}_2 \mathbf{P} = \mathbf{P}^\top \mathbf{P}$ for an invertible $\mathbf{P}$. Evaluating both sides on $\mathbf{x} = \mathbf{e}_1$ gives

$$-1 = \mathbf{e}_1^\top \text{diag}(-1, -1) \mathbf{e}_1 = \mathbf{e}_1^\top \mathbf{P}^\top \mathbf{P} \mathbf{e}_1 = \|\mathbf{P} \mathbf{e}_1\|^2 \geq 0,$$

a contradiction. So the discriminant is not a complete invariant, even over $\mathbb{R}$ and even for 2×2 matrices. Proving that the full “sign pattern” *is* an invariant, and that it is the complete one over $\mathbb{R}$, is the law of inertia, proved later in this chapter.

14.2.4 What congruence destroys

Now the negative result of the section, and the reason the previous three chapters could not have been this one.

Example. Congruent, with no eigenvalue in common

- (a) $\mathbf{I}_2 \simeq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ by [Example \(Congruences and a noncongruence\)](#) (a). Compare their eigenvalues.
- (b) $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \simeq \text{diag}(1, -3)$ by [Example \(Killing the cross term\)](#). Compare their eigenvalues.

Solution.

- (a) $\mathbf{I}_2$ has the single eigenvalue 1, twice. The characteristic polynomial of $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ is $x^2 - 3x + 1$, whose roots are $(3 \pm \sqrt{5})/2$, that is, approximately 2.618 and 0.382. The two matrices share no eigenvalue. They do share a rank, 2; a symmetry type, symmetric; and a discriminant, since $\det = 1$ for both.
- (b) The characteristic polynomial of $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ is $x^2 - 2x - 3 = (x - 3)(x + 1)$, so its eigenvalues are 3 and -1 . Those of $\text{diag}(1, -3)$ are 1 and -3 . Again no eigenvalue in common — but this time one positive and one negative on each side, and the determinants are -3 both times. The signs survived; the numbers did not. That is the pattern the law of inertia will explain.

Chapter 13 introduced this operation twice and both times deferred it to here. Section 1 of that chapter wrote that “the operation $\mathbf{A} \mapsto \mathbf{S}^* \mathbf{A} \mathbf{S}$ is called **congruence**, and Chapter 14 studies it for its own sake”, keeping only that it preserves the sign of a positive form (Proposition 13.1.2). Section 4 repeated the promise in fuller form: “Chapter 14 studies congruence in its own right, including what it does preserve; for now, note that it certainly does **not** preserve eigenvalues, since $\mathbf{S}^* \mathbf{I} \mathbf{S} = \mathbf{S}^* \mathbf{S}$ can be any positive definite matrix at all.”

Both debts are now paid, and it is worth saying the answer in one place. **What congruence preserves:** the rank ([Theorem 14.2.3](#)), and with it nondegeneracy and the dimension of the radical; the symmetry type ([Proposition 14.2.4](#)); and the square class of the determinant when that is nonzero ([Theorem 14.2.5](#)), the discriminant. Over $\mathbb{R}$ it preserves one thing more, the triple of signs, and that is the deepest of the four; it is proved in the section on inertia. **What congruence destroys:** the eigenvalues, and therefore the characteristic polynomial, the trace, and the determinant itself. Not every similarity invariant is lost — rank and invertibility, parts (a) and (b) of Proposition 4.2.6, survive both relations — but every invariant that reads the eigenvalues is. Chapter 13’s parenthetical remark is exactly right and can be made sharper: taking $\mathbf{A} = \mathbf{I}_n$ in [Definition 14.2.1](#), the congruence class of $\mathbf{I}_n$ over $\mathbb{R}$ is the set of all matrices $\mathbf{P}^T \mathbf{P}$ with $\mathbf{P}$ invertible, which by Theorem 13.1.5 (c) is the set of **all** positive definite matrices. One class, every spectrum with all eigenvalues positive.

Warning. Congruent and similar are different relations, and neither implies the other. By [Example \(Congruent, with no eigenvalue in common\)](#) (a), $\mathbf{I}_2 \simeq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$, while the only matrix similar to $\mathbf{I}_2$ is $\mathbf{I}_2$ itself, since $\mathbf{P}^{-1}\mathbf{I}_2\mathbf{P} = \mathbf{I}_2$ for every invertible $\mathbf{P}$. In the other direction, $\text{diag}(1, 2)$ is similar to $\begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix}$ — conjugate by $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ — but not congruent to it, because $\text{diag}(1, 2)$ is symmetric and the other matrix is not ([Proposition 14.2.4](#)). So do not carry any habit across. In particular, “diagonalizable” means two different things in this chapter, and a matrix that is diagonalizable by congruence need not be diagonalizable by similarity, nor conversely.

14.2.5 Where the two relations meet

There is exactly one situation in which the confusion is harmless: when the change-of-basis matrix satisfies $\mathbf{P}^\top = \mathbf{P}^{-1}$, that is, when $\mathbf{P}$ is orthogonal ([Definition 11.7.3](#)).

Proposition 14.2.6. An orthogonal change of basis is both

Let $\mathbf{A} \in M_n(\mathbb{R})$ and $\mathbf{Q} \in O(n)$. Then $\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q}$, so this matrix is at once congruent and similar to $\mathbf{A}$. Consequently, if $\mathbf{A}$ is symmetric, then $\mathbf{A}$ is congruent to the diagonal matrix $\mathbf{D}$ of its eigenvalues, listed with multiplicity in any prescribed order.

Proof. The first statement is the definition of an orthogonal matrix, $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_n$, which gives $\mathbf{Q}^\top = \mathbf{Q}^{-1}$; the matrix $\mathbf{Q}^\top \mathbf{A} \mathbf{Q}$ is then congruent to $\mathbf{A}$ by [Definition 14.2.1](#), since $\mathbf{Q}$ is invertible, and similar to $\mathbf{A}$ by [Definition 4.2.2](#).

For the second statement, let $\mathbf{A}$ be symmetric. By [Corollary 12.5.2](#) there are $\mathbf{Q} \in O(n)$ and a real diagonal $\mathbf{D}$, carrying the eigenvalues of $\mathbf{A}$ in any prescribed order, with $\mathbf{A} = \mathbf{Q} \mathbf{D} \mathbf{Q}^\top$. Multiplying on the left by $\mathbf{Q}^\top$ and on the right by $\mathbf{Q}$ gives $\mathbf{D} = \mathbf{Q}^\top \mathbf{A} \mathbf{Q}$, which is a congruence. This proves the proposition. ■

This is why the real spectral theorem is so strong, and it is what Chapter 12 meant when it said, of the principal axis theorem, that “an orthogonal change of variables is a similarity *and* a congruence, so it keeps the eigenvalues and the geometry at once”. The general theory of this chapter has no orthogonal matrices to call on — a general field has no inner product, hence no orthogonal group in that sense — so the two relations part company and only congruence remains. What [Proposition 14.2.6](#) gives over $\mathbb{R}$ is a bridge: it says that every real symmetric matrix is congruent to a diagonal matrix, and it identifies the diagonal entries as eigenvalues. The section on inertia will show that the entries are free to move but their signs are not, which is how the two pictures are finally reconciled.

Quick check. Over $\mathbb{R}$, which matrices are congruent to $\mathbf{I}_n$?

Solution. Exactly the positive definite symmetric matrices. If $\mathbf{B} = \mathbf{P}^\top \mathbf{I}_n \mathbf{P} = \mathbf{P}^\top \mathbf{P}$ with $\mathbf{P}$ invertible, then $\mathbf{B}$ is symmetric and $\mathbf{x}^\top \mathbf{B} \mathbf{x} = \|\mathbf{P} \mathbf{x}\|^2 > 0$ for $\mathbf{x} \neq \mathbf{0}$, since $\mathbf{P} \mathbf{x} \neq \mathbf{0}$. Conversely every positive definite $\mathbf{B}$ is $\mathbf{P}^\top \mathbf{P}$ for an invertible $\mathbf{P}$ by [Theorem 13.1.5](#). Compare

the similarity class of $\mathbf{I}_n$, which contains $\mathbf{I}_n$ alone: one relation collapses this class to a point, the other spreads it over every positive definite matrix there is.

14.2.6 Exercises

14.2.6.1 A. Check your understanding

Exercise. A1

- State the change-of-basis rule for the matrix of a bilinear form, saying which matrix $\mathbf{P}$ appears.
- Define what it means for $\mathbf{A}, \mathbf{B} \in M_n(F)$ to be congruent.
- For each of rank, symmetry, determinant, eigenvalues, and the square class of a nonzero determinant, say whether it is a congruence invariant.
- Determine whether the following is correct, with a reason: congruent matrices are similar.
- State the one circumstance in which a change of basis is simultaneously a similarity and a congruence.

Solution.

- [Theorem 14.2.1](#): $[\beta]_{\mathcal{B}'}^{\mathcal{B}'} = \mathbf{P}^\top [\beta]_{\mathcal{B}}^{\mathcal{B}} \mathbf{P}$, where $\mathbf{P} = [\text{id}]_{\mathcal{B}'}^{\mathcal{B}}$ is the change-of-coordinates matrix from the new basis to the old one.
- [Definition 14.2.1](#): $\mathbf{B} = \mathbf{P}^\top \mathbf{A} \mathbf{P}$ for some invertible $\mathbf{P} \in M_n(F)$.
- Invariant: rank ([Theorem 14.2.3](#)); symmetry ([Proposition 14.2.4](#)); the square class of a nonzero determinant ([Theorem 14.2.5](#)). Not invariant: the determinant itself, and the eigenvalues ([Example \(Congruent, with no eigenvalue in common\)](#)).
- Incorrect. $\mathbf{I}_2 \simeq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$, but the only matrix similar to $\mathbf{I}_2$ is $\mathbf{I}_2$.
- When the change-of-coordinates matrix $\mathbf{P}$ satisfies $\mathbf{P}^\top = \mathbf{P}^{-1}$, that is, when $\mathbf{P}$ is orthogonal ([Proposition 14.2.6](#)).

14.2.6.2 B. Practice

Exercise. B1: A congruence in $\mathbb{Q}$

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, both over $\mathbb{Q}$.

- Compute $\mathbf{P}^\top \mathbf{A} \mathbf{P}$, and hence write down a matrix congruent to $\mathbf{A}$ that is diagonal.
- Find the eigenvalues of both matrices and compare them.
- Compute the discriminant of the form $\mathbf{x}^\top \mathbf{A} \mathbf{y}$ and check that it agrees for the two matrices.

Solution.

- (a) $\mathbf{P}$ is invertible, with $\det \mathbf{P} = -2 \neq 0$. Then

$$\mathbf{A}\mathbf{P} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{P}^\top \mathbf{A}\mathbf{P} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

So $\mathbf{A} \simeq \text{diag}(2, -2)$ over $\mathbb{Q}$.

- (b) The characteristic polynomial of $\mathbf{A}$ is $x^2 - 1$, so its eigenvalues are 1 and -1 ; those of $\text{diag}(2, -2)$ are 2 and -2 . No eigenvalue is shared, and each matrix has one positive and one negative eigenvalue.
- (c) $\det \mathbf{A} = -1$ and $\det \text{diag}(2, -2) = -4 = (-2)^2 \cdot (-1)$, which is the relation $\det \mathbf{B} = (\det \mathbf{P})^2 \det \mathbf{A}$ of [Theorem 14.2.5](#). The two determinants differ by the square 4, so they lie in the same square class of $\mathbb{Q}^\times$ and the discriminants agree, as [Definition 14.2.2](#) requires.

Exercise. B2: A noncongruence over $\mathbb{Q}$

- (a) Prove that $\text{diag}(1, 1)$ and $\text{diag}(1, 2)$ are **not** congruent over $\mathbb{Q}$.
- (b) Prove that they **are** congruent over $\mathbb{R}$, by exhibiting $\mathbf{P}$.
- (c) Explain in one sentence why (a) and (b) are consistent.

Solution.

- (a) Suppose $\text{diag}(1, 2) = \mathbf{P}^\top \text{diag}(1, 1)\mathbf{P}$ with $\mathbf{P} \in M_2(\mathbb{Q})$ invertible. By [Theorem 14.2.5](#), $2 = (\det \mathbf{P})^2 \cdot 1$, so $\det \mathbf{P} \in \mathbb{Q}$ would satisfy $(\det \mathbf{P})^2 = 2$. No rational number squares to 2 (Theorem 0.2.3), a contradiction. Hence the two are not congruent over $\mathbb{Q}$.
- (b) Take $\mathbf{P} = \text{diag}(1, 1/\sqrt{2})$, which is invertible over $\mathbb{R}$. Then $\mathbf{P}^\top \text{diag}(1, 2)\mathbf{P} = \text{diag}(1, 1)$, and congruence is symmetric ([Proposition 14.2.2](#)).
- (c) Congruence is a relation **over a specified field**: the matrix $\mathbf{P}$ is required to have entries in that field, and $1/\sqrt{2}$ is available in $\mathbb{R}$ but not in $\mathbb{Q}$. Equivalently, 2 is a square in $\mathbb{R}^\times$ and not in $\mathbb{Q}^\times$, so the discriminant separates the two matrices over $\mathbb{Q}$ and not over $\mathbb{R}$.

Exercise. B3: Completing the square, again

On $\mathbb{R}^2$, let $\beta(\mathbf{x}, \mathbf{y}) = x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$.

- (a) Write down $\mathbf{A} = [\beta]_{\mathcal{E}}^{\mathcal{E}}$ and expand $q(\mathbf{x}) = \beta(\mathbf{x}, \mathbf{x})$ as a polynomial in x_1, x_2 .
- (b) Complete the square to find a basis $\mathcal{B}'$ with $[\beta]_{\mathcal{B}'}^{\mathcal{B}'} = \mathbf{I}_2$. Verify by computing $\mathbf{P}^\top \mathbf{A}\mathbf{P}$.
- (c) Hence find the eigenvalues of $\mathbf{A}$ and compare them with those of $\mathbf{I}_2$.

Solution.

- (a) $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$, and $q(\mathbf{x}) = x_1^2 - 2x_1x_2 + 2x_2^2$.
- (b) $q(\mathbf{x}) = (x_1 - x_2)^2 + x_2^2$. So the substitution $y_1 = x_1 - x_2$, $y_2 = x_2$, that is $x_1 = y_1 + y_2$

and $x_2 = y_2$, makes $q = y_1^2 + y_2^2$. This is $\mathbf{x} = \mathbf{P}\mathbf{y}$ with

$$\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{A}\mathbf{P} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad \mathbf{P}^\top \mathbf{A}\mathbf{P} = \mathbf{I}_2.$$

By [Theorem 14.2.1](#) the basis is $\mathcal{B}' = ((1, 0), (1, 1))$, the columns of $\mathbf{P}$.

- (c) The characteristic polynomial of $\mathbf{A}$ is $x^2 - 3x + 1$, with roots $(3 \pm \sqrt{5})/2$; $\mathbf{I}_2$ has the eigenvalue 1 twice. The two matrices are congruent with no eigenvalue in common. What they share is that all eigenvalues are positive on both sides, which by the check above is no accident: the congruence class of $\mathbf{I}_2$ over $\mathbb{R}$ consists of exactly the positive definite matrices.

14.2.6.3 C. Going deeper

Exercise. C1: Congruence versus equivalence

- (a) Prove that if $\mathbf{A} \simeq \mathbf{B}$ then $\mathbf{A}$ and $\mathbf{B}$ are equivalent in the sense of Definition 4.3.1.
- (b) Give two matrices in $M_2(\mathbb{Q})$ that are equivalent but not congruent, and justify your example.
- (c) Hence explain what “the congruence classes refine the equivalence classes” means, and say what extra information a congruence class carries that a rank does not.

Solution.

- (a) If $\mathbf{B} = \mathbf{P}^\top \mathbf{A}\mathbf{P}$ with $\mathbf{P}$ invertible, then $\mathbf{P}^\top$ is invertible (Theorem 0.9.5), so $\mathbf{B} = \mathbf{Q}\mathbf{A}\mathbf{P}$ with $\mathbf{Q} = \mathbf{P}^\top$ and $\mathbf{P}$ both invertible, which is Definition 4.3.1.
- (b) Take $\text{diag}(1, 1)$ and $\text{diag}(1, 2)$. Both have rank 2, so they are equivalent by Theorem 4.3.3 (c). They are not congruent over $\mathbb{Q}$ by [Exercise \(B2: A noncongruence over \$\mathbb{Q}\$ \)](#) (a).
- (c) By (a), each congruence class lies inside a single equivalence class, and by (b) an equivalence class may split into more than one congruence class; that is what “refine” means. An equivalence class is named by the rank alone (Theorem 4.3.3 (c)). A congruence class remembers the rank, and in addition the symmetry type and the discriminant — and over $\mathbb{R}$, the signs. Equivalence allows two independent changes of basis, one in each slot; congruence allows only one, used twice, and the extra invariants are the price of that constraint.

Exercise. C2: Orthogonal congruence

Let $\mathbf{A} \in M_n(\mathbb{R})$ be symmetric with eigenvalues $\lambda_1, \dots, \lambda_n$, listed with multiplicity.

- (a) Prove that $\mathbf{A} \simeq \text{diag}(\lambda_1, \dots, \lambda_n)$.
- (b) Deduce that if every λ_i is nonzero, then $\mathbf{A}$ is congruent to a diagonal matrix all of whose entries are ± 1 .
- (c) Give two real symmetric 2×2 matrices that are congruent and have no eigenvalue in common.

Hint for (b): recall from [Example \(Congruences and a noncongruence\)](#) (b) what scaling a

basis vector does to a diagonal entry.

Solution.

- (a) This is [Proposition 14.2.6](#): by Corollary 12.5.2 there is $\mathbf{Q} \in O(n)$ with $\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = \mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$, and $\mathbf{Q}$ is invertible, so $\mathbf{A} \simeq \mathbf{D}$.
- (b) Let $\mathbf{S} = \text{diag}(c_1, \dots, c_n)$ with $c_i = 1/\sqrt{|\lambda_i|}$, which is defined and nonzero because $\lambda_i \neq 0$. Then $\mathbf{S}$ is invertible and $\mathbf{S}^\top \mathbf{D} \mathbf{S} = \text{diag}(c_1^2 \lambda_1, \dots, c_n^2 \lambda_n)$, whose i -th entry is $\lambda_i/|\lambda_i| = \pm 1$ according to the sign of λ_i . By (a) and transitivity ([Proposition 14.2.2](#)), $\mathbf{A}$ is congruent to that matrix.
- (c) By (b), $\text{diag}(1, -3)$ and $\text{diag}(2, -1)$ are both congruent to $\text{diag}(1, -1)$, hence to each other by [Proposition 14.2.2](#); explicitly, $\mathbf{P} = \text{diag}(\sqrt{2}, 1/\sqrt{3})$ gives $\mathbf{P}^\top \text{diag}(1, -3) \mathbf{P} = \text{diag}(2, -1)$. Their eigenvalues are $\{1, -3\}$ and $\{2, -1\}$, which are disjoint. Note that the **signs** match, one positive and one negative on each side; part (b) is the reason, and the law of inertia will say it is forced.

Exercise. C3: How much the discriminant sees

- (a) Show that $\mathbb{C}^\times$ has exactly one square class, and that $\mathbb{R}^\times$ has exactly two.
- (b) List the squares in $\mathbb{F}_5^\times$ and show that $\mathbb{F}_5^\times$ has exactly two square classes.
- (c) Hence determine, for each of $\mathbb{C}$, $\mathbb{R}$, and $\mathbb{F}_5$, the largest number of pairwise noncongruent invertible matrices in M_n that the discriminant alone can distinguish.

Solution.

- (a) Over $\mathbb{C}$, every $a \neq 0$ has a square root (Example (Square Roots of $5 - 12i$)), so $a = c^2 \cdot 1$ and every element is equivalent to 1: one class. Over $\mathbb{R}$, squares of nonzero reals are exactly the positive reals, so a and b are equivalent precisely when $a/b > 0$, which happens exactly when they have the same sign. The classes are the positive reals and the negative reals: two classes, and they are distinct since 1 and -1 are not related.
- (b) The squares are $1^2 = 1$, $2^2 = 4$, $3^2 = 4$, $4^2 = 1$, so the set of nonzero squares is $\{1, 4\}$. Multiplying $\{1, 4\}$ by 2 gives $\{2, 3\}$, so the classes are $\{1, 4\}$ and $\{2, 3\}$: two classes, each of size 2, covering all four elements of $\mathbb{F}_5^\times$.
- (c) The discriminant takes one value per square class, so it can separate at most as many classes as there are square classes: one over $\mathbb{C}$, two over $\mathbb{R}$, and two over $\mathbb{F}_5$. Over $\mathbb{C}$ it is therefore useless, and over $\mathbb{R}$ and $\mathbb{F}_5$ it sorts invertible matrices into at most two piles. By [Exercise \(C2: Orthogonal congruence\)](#) (b), over $\mathbb{R}$ every invertible symmetric matrix is congruent to one of the $n + 1$ matrices $\mathbf{I}_p \oplus -\mathbf{I}_{n-p}$, whose determinant is $(-1)^{n-p}$; so the discriminant sees only the parity of $n - p$, and for $n \geq 2$ it cannot separate, say, $\mathbf{I}_2$ from $-\mathbf{I}_2$. That the $n + 1$ matrices really are pairwise noncongruent is the law of inertia, not anything the discriminant can deliver.