Princeton University MAT 202 Spring 2008

Willie W. Wong

Apr 21 - 25

APR 21, 2008

- For right now, unless explicitly mentioned, we'll be only working with real numbers again (the complex eigenvalues we mentioned last Friday will be useful in the future, but not right now).
- Remember what we did in the last couple of weeks: we were interested in studying linear discrete dynamical systems, and we found out that given an evolution matrix A, the system can be written in closed form when A admits an eigenbasis (in other words, A is diagonalizable).
- A rather natural thing to ask (in view of geometry) is then, can the eigenbasis be also an orthonormal basis?
- Remember what it means for A to be diagonalizable: it means that there is a matrix S such that $S^{-1}AS$ is a diagonal matrix. Furthermore, remember that by the change-of-basis formulae, the column vectors of S are are the eigenvectors of A. So should the eigenbasis be an orthonormal basis, the column vectors of S are orthonormal, meaning that S is in fact an orthogonal matrix ($S^{-1} = S^T$), a fact from chapter 5.
- A definition: we say a matrix A is *orthogonally diagonalizable* if there exists an orthogonal matrix S such that $S^{-1}AS = S^TAS$ is diagonal.
- Today we study the question "which matrices are orthogonally diagonalizable?"
- Some simple examples of orthogonally diagonalizable matrices:
  - Let A be a diagonal matrix. Then for $S = I_M$ the identity matrix, $I^{-1}AI = A$ is diagonal, and I is an orthogonal matrix (corresponding to the orthonormal basis of standard vectors), so A is orthogonally diagonalizable.
  - Let A be the matrix of an orthogonal projection: let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ be an orthonormal basis of the subspace V of $\mathbb{R}^M$, then letting $U = \begin{bmatrix}\boldsymbol{v}_1 & \ldots & \boldsymbol{v}_K\end{bmatrix}$, we have that $A = UU^T$ is

the matrix of orthogonal projection onto V. Now, consider the orthogonal complement space $V^{\perp}$ and let $\boldsymbol{w}_1 \dots \boldsymbol{w}_J$ be an orthonormal basis thereof. Write the square matrix

$$S = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_K & \boldsymbol{w}_1 & \dots & \boldsymbol{w}_J \end{bmatrix}$$

then a simple calculation shows that

$$S^{-1}AS = S^{-1}UU^TS = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}$$

where there are K 1s and J 0s. It is clear that $\boldsymbol{v}_1, \dots, \boldsymbol{v}_K, \boldsymbol{w}_1, \dots, \boldsymbol{w}_J$ form an O-N basis of $\mathbb{R}^M$, so the orthogonal projection is orthogonally diagonalizable.

- Now notice one thing, in the first case, A is a diagonal matrix, and so is by definition symmetric: $A = A^T$. In the second case, $A = UU^T$, so $A^T = (UU^T)^T = (U^T)^TU^T = UU^T = A$. For both of these cases, A is symmetric. This fact is indeed more general

- Claim: if A is a matrix that is orthogonally diagonalizable then A is symmetric. To prove this statement, we note that the hypothesis implies

  $$S^{-1}AS = D \Rightarrow A = SDS^{-1} \Rightarrow A = SDS^T$$

  where D is diagonal, and the second implication follows from the fact that S is orthogonal. Now we just calculate

  $$A^T = (SDS^T)^T = (S^T)^TD^TS^T = SD^TS^T = SDS^T$$

  the last equality using the fact that D is diagonal, and hence symmetric. From this we can conclude that

  $$A = A^T$$

  and so A is symmetric.

- The rest of today will be devoted to proving the converse of the above statement, which is called the **Spectral Theorem**: *A matrix is orthogonally diagonalizable if and only if it is symmetric*.

- To prove the spectral theorem, we proceed in two steps

  - First we show that for a symmetric matrix, if $\boldsymbol{v}$ and $\boldsymbol{w}$ are two eigenvectors with different eigenvalues, then $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal.

  - Second we show that a symmetric matrix always have the full set of distinct real eigenvalues when counted with algebraic multiplicity.

- Let's see how we can prove the spectral theorem assuming those two facts:

  - Start with a symmetric $M \times M$ matrix A. By the second fact, A has at least some real eigenvalue. Let λ be a real eigenvalue, and let $\boldsymbol{v}$ be its eigenvector (remember, while the geometric multiplicity of any eigenvalue is bounded by its algebraic multiplicity, the geometric multiplicity is always at least 1). Let V be the subspace spanned by $\boldsymbol{v}$ and $V^\perp$ its orthogonal complement. We can choose $\boldsymbol{v}$ to have length 1, otherwise we'll just divide $\boldsymbol{v}$ by its length. We can construct an O-N basis of $V^\perp$, call it $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{M-1}$. Then the matrix
$$P = \begin{bmatrix} \boldsymbol{v} & \boldsymbol{u}_1 & \ldots & \boldsymbol{u}_{M-1} \end{bmatrix}$$
is an orthogonal matrix. Let's compute
$$P^{-1}AP = P^{-1}\begin{bmatrix} A\boldsymbol{v} & A\boldsymbol{u}_1 & \ldots & A\boldsymbol{u}_{M-1} \end{bmatrix} = P^T\begin{bmatrix} \lambda\boldsymbol{v} & A\boldsymbol{u}_1 & \ldots & A\boldsymbol{u}_{M-1} \end{bmatrix}$$

  - We evaluate it further and have
$$P^{-1}AP = \begin{bmatrix} \boldsymbol{v}^T \\ \boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_{M-1}^T \end{bmatrix} \begin{bmatrix} \lambda\boldsymbol{v} & A\boldsymbol{u}_1 & \ldots & A\boldsymbol{u}_{M-1} \end{bmatrix} = \begin{bmatrix} \lambda & * & * & \ldots & * \\ 0 & * & * & \ldots & * \\ 0 & * & * & \ldots & * \\ \vdots & & & & \vdots \\ 0 & * & * & \ldots & * \end{bmatrix}$$
where $*$ stands for terms we don't know exactly. However, we know that A is symmetric, so $(P^TAP)^T = P^TAP$ so that this product is also symmetric. Therefore we can conclude that
$$P^{-1}AP = \begin{bmatrix} \lambda & 0 \\ 0 & B \end{bmatrix}$$
where B is a symmetric, $M-1 \times M-1$ matrix.

  - Now, we just do the same thing to B (here we can use induction). By the above process we that there exists some orthogonal matrix Q such that $Q^{-1}BQ = \begin{bmatrix} \lambda' & 0 \\ 0 & C \end{bmatrix}$ for some $M-2 \times M-2$ matrix C. Then we have that, writing the matrix
$$R = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$$
R is orthogonal, and so is PR (remember that products of orthogonal matrices are still orthogonal), and we can write
$$(PR)^{-1}A(PR) = \begin{bmatrix} \lambda & & \\ & \lambda' & \\ & & C \end{bmatrix}$$

  - We rinse and repeat until the symmetric block matrix on the lower right hand corner is 1×1, in which case it is also diagonal, and there, we have an orthogonal diagonalization of our original matrix A.

- In the proof, we crucially used the second fact. The first fact is used more-or-less as motivation. Now we show that those two facts are true:

  - For the first fact: remember that we can write vectors as $M \times 1$ matrices and then dot products becomes

    $$\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v}^T \boldsymbol{w}$$

    Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be eigenvectors with different eigenvalues λ_v, λ_w. We notice that

    $$\boldsymbol{v} \cdot A\boldsymbol{w} = \boldsymbol{v}^T A\boldsymbol{w} = (\boldsymbol{w}^T A^T \boldsymbol{v})^T = (\boldsymbol{w} \cdot A\boldsymbol{v})^T = \boldsymbol{w} \cdot A\boldsymbol{v}$$

    where we used the fact that the transpose of a scalar is itself, and that A is symmetric. This implies that

    $$\lambda_w(\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{v} \cdot \lambda_w \boldsymbol{w} = \boldsymbol{w} \cdot \lambda_v \boldsymbol{v} = \lambda_v(\boldsymbol{w} \cdot \boldsymbol{v})$$

    If $\boldsymbol{w} \cdot \boldsymbol{v} \neq 0$, we can divide it from both sides and end up with $\lambda_v = \lambda_w$, which contradicts our assumption that $\lambda_v \neq \lambda_w$. So $\boldsymbol{w} \cdot \boldsymbol{v} = 0$ necessarily, which means that $\boldsymbol{w}$ and $\boldsymbol{v}$ are orthogonal.

  - For the second fact: We know from our study of complex eigenvalues that, by the fundamental theorem of algebra, any $M \times M$ matrix A has exactly M complex eigenvalues when counted with algebraic multiplicity. It we can show that all of the eigenvalues are real, then we have that the total number of real eigenvalues when counted with algebraic multiplicity must also be M. To do so, suppose $\lambda = r + si$ is a complex eigenvalue, and that $\boldsymbol{v} + i\boldsymbol{w}$ is its eigenvector. Now, remember the following fact from last Friday: that complex eigenvalues (eigenvectors) comes in pairs of complex conjugates, so we have that $\boldsymbol{v} - i\boldsymbol{w}$ is also an eigenvector, but with eigenvalues $r - si$. Now we do the same calculation as above: first we note

    $$(\boldsymbol{v} + i\boldsymbol{w}) \cdot (\boldsymbol{v} - i\boldsymbol{w}) = |\boldsymbol{v}|^2 + |\boldsymbol{w}|^2$$

    Now, from the calculation for fact one, we have

    $$(\boldsymbol{v} + i\boldsymbol{w}) \cdot A(\boldsymbol{v} - i\boldsymbol{w}) = (\boldsymbol{v} - i\boldsymbol{w}) \cdot A(\boldsymbol{v} + i\boldsymbol{w})$$

    from which we arrive at

    $$(r - si)(|\boldsymbol{v}|^2 + |\boldsymbol{w}|^2) = (r + si)(|\boldsymbol{v}|^2 + |\boldsymbol{w}|^2)$$

    which immediately implies that $s = -s$ and so $s = 0$, and the eigenvalue λ is a real number only. Thus we have proven that a symmetric matrix A cannot have complex eigenvalues, and so it has a full set of real eigenvalues.

- Now, how do we find the orthonormal eigenbasis in practice? Knowing that the spectral theorem is true, what we do is
  - First find the (real) eigenvalues of A.
  - For each eigenvalue, find its eigenspace.
  - For each eigenspace, find a basis. Now use Gram-Schmidt to find an orthonormal basis of the eigenspace.
  - Collect all the orthonormal bases together, since the eigenspaces are orthogonal to each other (by fact 1), we have that those vectors collect to form an O-N basis of $\mathbb{R}^M$.
- Example?

APR 23, 2008

- The rest of this week we'll talk about quadratic forms and a bit more about the spectral theorem.
- First, what is a quadratic form? Simply put, a quadratic form is a "homogeneous degree 2 polynomial on $\mathbb{R}^M$". Let me explain what this means
  - A polynomial on $\mathbb{R}^M$ is an expression. In particular, a polynomial is a finite sum of monomials. Letting $x_1, \ldots, x_M$ be the Cartesian coördinates on $\mathbb{R}^M$, a monomial on $\mathbb{R}^M$ is a term that looks like

    $$\text{constant} \cdot x_1^{p_1} x_2^{p_2} \cdots x_M^{p_M}$$

    i.e., a product of a constant scalar with non-negative powers of the coördinate values.
    - In $\mathbb{R}^4$,

      $$3x_1^7 - 12x_1^8 x_2^4 + x_1 x_3 x_4^2 + x_2^2 + 3$$

      is a polynomial (you can have pure constant terms: this just means that the powers of all the coördinates are 0).
    - The expression

      $$x_3^{-3}$$

      is not a polynomial nor a monomial: we cannot allow negative powers of coördinates
    - The expression

      $$\sum_{i=0}^{\infty} c_i x_2^i$$

      is not a polynomial, since the sum is infinite. However,

      $$\sum_{i=0}^{10000000000000000000} c_i x_2^i$$

      is a polynomial.

- The degree of a monomial is the sum of its powers. I.e., for a monomial

  $$\text{constant} \cdot x_1^{p_1} x_2^{p_2} \cdots x_M^{p_M}$$

  its degree is $p_1 + p_2 + \cdots p_M$. The degree of a polynomial is the largest of the degrees of its monomials.

  - In the first example above

    $$3x_1^7 - 12x_1^8 x_2^4 + x_1 x_3 x_4^2 + x_2^2 + 3$$

    the degrees of the monomials are $7, 12, 4, 2, 0$, so the degree of the polynomial is 12.
  - In the example

    $$\sum_{i=0}^{10000000000000000000} c_i x_2^i$$

    the degree is 10000000000000000000.
- Lastly, a homogeneous polynomial is one whose monomials are all of the same degree.
  - So neither of the two examples given above are homogeneous.
  - Notice that a monomial is, by definition, homogeneous.
  - An example of a homogeneous polynomial in $\mathbb{R}^3$

    $$x_1^5 + 5x_2 x_3^4 + 3x_1^3 x_3^2 - 2x_1 x_2^2 x_3^2$$

    each of the terms has degree 5.
  - A product of homogeneous polynomials is also homogeneous: suppose $p(x)$ is a homogeneous degree m polynomial and $q(x)$ is a homogeneous degree n polynomial, by the "foil" rule for multiplying polynomials, you can check that each term in the polynomial $p \cdot q$ comes from a product of one term from p and one term from q, and therefore must have degree $m + n$.
- So in $\mathbb{R}^M$, a quadratic form, or a homogeneous degree 2 polynomial, is an expression of the form

  $$\sum_{i=1}^{M} \sum_{j=1}^{M} c_{ij} x_i x_j$$

  in other words, a sum of monomials of the form constant$\cdot x_i x_j$ where i and j are allowed to be the same. Notice that by commutativity of ordinary multiplication, $x_i x_j = x_j x_i$, so we can often simplify the quadratic form to an expression

  $$\sum_{i=1}^{M} \sum_{j=i}^{M} b_{ij} x_i x_j$$

notice now that we "order" the terms so that the index i for x_ix_j is always less than or equal to the index j. In this expression, then,

$$b_{ij} = \begin{cases} c_{ij} & i = j \\ c_{ij} + c_{ji} & i < j \\ 0 & j > i \end{cases}$$

- Notice that in the preceding definition, we can express c_{ij} and b_{ij} as square matrices over $\mathbb{R}^M$.

- So we have that: any square $M \times M$ matrix defines a quadratic form. In particular, let A be such a matrix, then its corresponding quadratic form is

$$q_A(x) = \sum_{i=1}^{M} \sum_{j=1}^{M} a_{ij} x_i x_j$$

- This association, however, is not unique! Two different matrices can define the same quadratic form as we see above, the matrices C and B are different in general: B is upper triangular, while C does not have to be.

  - A concrete example: let

$$B = \begin{bmatrix} 3 & 5 & 4 \\ 0 & -2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad , \quad C = \begin{bmatrix} 3 & 2 & 2 \\ 3 & -2 & -1 \\ 2 & 2 & 1 \end{bmatrix}$$

    then we can look at the quadratic forms defined by B and C:

$$\begin{aligned} q_B &= \sum_{i=1}^{3} \sum_{j=1}^{3} b_{ij} x_i x_j \\ &= 3x_1^2 + 5x_1x_2 + 4x_1x_3 + 0x_2x_1 - 2x_2^2 + x_2x_3 + 0x_3x_1 + 0x_3x_1 + x_3^2 \\ &= 3x_1^2 - 2x_2^2 + x_3^2 + 5x_1x_2 + 4x_1x_3 + x_2x_3 \\ q_C &= \sum_{i=1}^{3} \sum_{j=1}^{3} c_{ij} x_i x_j \\ &= 3x_1^2 + 2x_1x_2 + 2x_1x_3 + 3x_2x_1 - 2x_2^2 - x_2x_3 + 2x_3x_1 + 2x_3x_2 + x_3^2 \\ &= 3x_1^2 - 2x_2^2 + x_3^2 + 5x_1x_2 + 4x_1x_3 + x_2x_3 \end{aligned}$$

- We want to find a canonical way of associating a matrix to a quadratic form. To do so, we look at where the ambiguity lies. We notice that the terms *on the diagonal* is always the same between two matrices that define the same quadratic form, and that to define the same quadratic form, the sums $b_{ij} + b_{ij} = c_{ij} + c_{ji}$ (we are given freedom on how to distribute the constant in constant $\cdot\, x_i x_j$ to the $x_i x_j$ and $x_j x_i$ terms). This tells us that, if B and C are two matrices that define the same quadratic form,

$$b_{ij} + b_{ji} = c_{ij} + c_{ji}$$

for terms both on and off the diagonal. Written in matrix notation, this means

$$B + B^T = C + C^T$$

- Given a square matrix B, we call the matrix $(B + B^T)/2$ its *symmetrization*. Notice that
  - The symmetrization of B and the matrix B have the same diagonal terms.
  - The symmetrization of B and the matrix B define the same quadratic form.
  - The symmetrization of B is symmetric:

    $$(\frac{B + B^T}{2})^T = \frac{(B + B^T)^T}{2} = \frac{B^T + B}{2}$$

  - If B and C define the same quadratic form, their symmetrizations are equal:

    $$\frac{B + B^T}{2} = \frac{C + C^T}{2}$$

  From these we conclude that for every symmetric matrix corresponds to one and only one quadratic form; in other words, to understand quadratic forms, in is enough to understand symmetric matrices.

- Notice further that, by definition, the quadratic form corresponding to A is

  $$q_A(x) = \boldsymbol{x} \cdot A\boldsymbol{x} = \boldsymbol{x}^T A\boldsymbol{x}$$

  where

  $$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix}$$

- To construct a symmetric matrix A from a quadratic form, we set a_{ii} to be the coefficient of the x_i^2 term, and $a_{ij} = a_{ji}$ for $i < j$ to be $1/2$ the coefficient of the $x_i x_j$ term.

- A quick application of the theory of symmetric matrices: on Monday we've shown that a symmetric matrix is orthogonally diagonalizable. Let $\mathcal{B}$ be the orthonormal eigenbasis for a symmetric matrix A, with eigenvalues $\lambda_1, \ldots, \lambda_M$, then

  $$A = SDS^T$$

where S is the matrix with column vectors the eigenbasis. Now we evaluate the quadratic form

$$q_A(x) = \boldsymbol{x}^T A \boldsymbol{x} = \boldsymbol{x}^T S D S^T \boldsymbol{x} = \boldsymbol{y}^T D \boldsymbol{y}$$

where

$$\boldsymbol{y} = S^T \boldsymbol{x}$$

is the expression of $\boldsymbol{x}$ in the coördinates $\mathcal{B}$. In other words, if the vectors of $\mathcal{B}$ are $\boldsymbol{u}_1, \dots, \boldsymbol{u}_M$ and

$$\boldsymbol{x} = y_1 \boldsymbol{u}_1 + \cdots + y_M \boldsymbol{u}_M$$

then

$$q_A(x) = \lambda_1 y_1^2 + \cdots + \lambda_M y_M^2$$

- This means that in a suitable basis, we can always get rid of the mixed terms $x_i x_j$ with $i \neq j$ from the quadratic form.

- This is one reason why the theorem we proved on Monday is called the *spectral* theorem. When we talk about the spectrum of light, we picture white light coming into a prism and a rainbow of colors coming out: we “unmixed” the various colors and separate them into individual components. Instead of looking at symmetric matrices, we look at quadratic forms: in the beginning it has terms mixing different directions. The spectral theorem gives us a process (like a prism) that gets rid of the mixed terms (the mixing of colors to produce white light) and leaves only pure terms (the pure colors).

APR 25, 2008

- Often, the most important thing we ask about a quadratic form is its sign and definiteness. We first recall a bit of calculus: given a quadratic form, what is its derivative?
  - We first write

    $$q(x) = \boldsymbol{x}^T A \boldsymbol{x}$$

    then the partial derivative can be evaluated using the Leibniz/product rule

    $$\begin{aligned}\frac{\partial}{\partial x_j} q &= \frac{\partial}{\partial x_j} (\boldsymbol{x}^T A \boldsymbol{x}) \\ &= \frac{\partial}{\partial x_j} \boldsymbol{x}^T A \boldsymbol{x} + \boldsymbol{x}^T \frac{\partial}{\partial x_j} A \boldsymbol{x} + \boldsymbol{x}^T A \frac{\partial}{\partial x_j} \boldsymbol{x}\end{aligned}$$

  - Noting that A is the matrix of coefficients, it does not depend on x, and so the partial derivative of A relative to x_j is always 0. On the other hand, for the vector $\boldsymbol{x}$

    $$\frac{\partial}{\partial x_j} \boldsymbol{x} = \frac{\partial}{\partial x_j} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} = \boldsymbol{e}_j$$

So the partial derivative is

$$\frac{\partial}{\partial x_j} q = \boldsymbol{e}_j^T A \boldsymbol{x} + \boldsymbol{x}^T A \boldsymbol{e}_j$$

now using the fact that A is symmetric, we have, in the end

$$\frac{\partial}{\partial x_j} q = 2(\boldsymbol{e}_j^T A)\boldsymbol{x}$$

  - Noticing that $\boldsymbol{e}_j^T A$ is an $1 \times M$ matrix, we have that the partial derivative is a linear transformation of $\mathbb{R}^M \to \mathbb{R}$, and, in particular, is 0 at the origin. So we have that
  - The origin is always a critical point for a quadratic form.

- The question now is, is the critical point a minimum? a maximum? a saddle point?

- Geometrically, we can assign five different labels to a quadratic form: A quadratic form q (and its associated symmetric matrix A) is said to be
  - *positive definite* if $q(x) > 0$ whenever $\boldsymbol{x} \neq 0$
  - *positive semidefinite* if $q(x) \geq 0$ for all $\boldsymbol{x}$
  - *negative definite* if $q(x) < 0$ whenever $\boldsymbol{x} \neq 0$
  - *negative semidefinite* if $q(x) \leq 0$ for all $\boldsymbol{x}$
  - *indefinite* if $q(x)$ is positive for some x and negative for some other x.

- Notice that a positive definite matrix is also positive semidefinite. Similarly for negative definite matrices.

- Geometrically: a positive definite quadratic form means that the critical point at the origin is a unique global minimum. A negative definite quadratic form means that the critical point is a unique global maximum. A positive semidefinite quadratic form which is not positive definite means that the critical point is a degenerate global minimum: it is like a fold in a piece of paper (demonstrate). Similarly for a negative semidefinite quadratic form. An indefinite quadratic form represents a saddle point.

- Recall from Wednesday that if $\lambda_1 \ldots \lambda_M$ are the eigenvalues of A counted with multiplicity, and $\mathcal{B}$ is the orthonormal eigenbasis, with $y_1, \ldots, y_M$ the coördinates of the vector $\boldsymbol{x}$ in the $\mathcal{B}$ basis, we have that

  $$q_A(x) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_M y_M^2$$

  Now, since the squares of the coördinate values are always non-negative, the "sign" and definiteness of the matrix A is completely determined by its eigenvalues: *A symmetric matrix A is positive definite if and only if all its eigenvalues are strictly positive; it is positive semidefinite if and only if all its eigenvalues are non-negative; the negative definite and semidefinite cases are defined* mutatis mutandis*; and A is indefinite if and only if it has both positive and negative eigenvalues.*

- Unfortunately, computing exactly what are the eigenvalues can turn out to be difficult (a large degree characteristic polynomials can be really difficult to solve). Fortunately, we have the following criterion for testing positive definiteness:
  - Let A be a symmetric $M \times M$ matrix. We define the matrices $A^{(k)}$ to be the $k \times k$ matrix on the upper left corner of A. We call them the *principal submatrices* of A. Example: given
$$A = \begin{bmatrix} 1 & 5 & 4 & 2 & 6 \\ 5 & 7 & -3 & 2 & -1 \\ 4 & -3 & 3 & 5 & -2 \\ 2 & 2 & 5 & -1 & 2 \\ 6 & -1 & -2 & 2 & 1 \end{bmatrix}$$
Its principal submatrices are
$$A^{(1)} = \begin{bmatrix} 1 \end{bmatrix} , \quad A^{(2)} = \begin{bmatrix} 1 & 5 \\ 5 & 7 \end{bmatrix} , \quad A^{(3)} = \begin{bmatrix} 1 & 5 & 4 \\ 5 & 7 & -3 \\ 4 & -3 & 3 \end{bmatrix} , \quad A^{(4)} = \begin{bmatrix} 1 & 5 & 4 & 2 \\ 5 & 7 & -3 & 2 \\ 4 & -3 & 3 & 5 \\ 2 & 2 & 5 & -1 \end{bmatrix}$$
with $A^{(5)} = A$.
  - The matrix A is positive definite if and only if the determinants of all its principal submatrices are positive. Consider the matrix
$$A = \begin{bmatrix} 3 & -1 & 2 \\ -1 & 5 & 1 \\ 2 & 1 & 3 \end{bmatrix}$$
Its determinant is
$$\det(A) = 45 - 2 - 2 - 3 - 20 - 3 = 15 > 0$$
and
$$\det(A^{(1)}) = \det \begin{bmatrix} 3 \end{bmatrix} = 3$$
and
$$\det(A^{(2)}) = \det \begin{bmatrix} 3 & -1 \\ -1 & 5 \end{bmatrix} = 15 - 1 = 14$$
so A is positive definite.
  - In some cases (this one, for example) we can determine whether the eigenvalues are all positive by looking at the characteristic polynomial without solving it. Here the characteristic polynomial is
$$f_A(\lambda) = (3-\lambda)(5-\lambda)(3-\lambda)-2-2-(3-\lambda)-(3-\lambda)-4(5-\lambda) = 15-33\lambda+11\lambda^2-\lambda^3$$

Now we use Descartes' rule of signs: the number of sign changes of consecutive coefficients when a polynomial of one variable is ordered in descending degree is equal to 2 times a non-negative number plus the number of positive roots.

$$f_A(\lambda) = -\lambda^3 + 11\lambda^2 - 33\lambda + 15$$

has 3 sign changes, which means that it has either 3 or 1 positive root.

$$f_A(-\lambda) = \lambda^3 + 11\lambda^2 + 33\lambda + 15$$

means that it has 0 negative roots. Since we also know that $f_A(0) = 15 \neq 0$, we have that all of the roots of f_A, which we know by the spectral theorem must be real, are positive, and so we confirm that A must be positive definite.

- Lastly, we give an application of quadratic forms and symmetric matrices relating to graphing conics. Suppose we are asked to describe the graph of

  $$5x^2 + 4xy + 2y^2 = 2$$

  The left hand side is a quadratic form in $\mathbb{R}^2$. So we write it as

  $$\boldsymbol{x} \cdot A\boldsymbol{x} = 1$$

  where

  $$A = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

  what we can do now is to orthogonally diagonalize A: first we solve for A's eigenvalues:

  $$f_A(\lambda) = \lambda^2 - 7\lambda + 6 = (\lambda - 6)(\lambda - 1)$$

  so A has two distinct eigenvalues $6, 1$. Its eigenvectors can be found by considering

  $$E_6 = \ker \begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \quad , \quad E_1 = \ker \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

  We find that E_6 is spanned by $(2, 1)$, and E_1 spanned by $(1, -2)$. We normalize the two vectors and take

  $$\boldsymbol{u} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix} \quad , \quad \boldsymbol{v} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$$

  then $\boldsymbol{u}, \boldsymbol{v}$ forms our orthonormal eigenbasis. In this coördinates, our equation reads

  $$6c_u^2 + c_v^2 = 1$$

  which is an ellipse with major axis in the $\boldsymbol{v}$ direction and semi-major axis length 1, and minor axis in the $\boldsymbol{u}$ direction and semi-minor axis length $1/\sqrt{6}$.

- For a quadratic form q in $\mathbb{R}^2$, the graph of $q = k$ where k is a positive number can be classified by the eigenvalues of q's corresponding matrix:
  - If both eigenvalues are positive and distinct, the graph is an ellipse with the principal axes given by directions of the eigenbasis. The semi-major and semi-minor axis are given by $\sqrt{k}/\sqrt{\lambda_1}$ and $\sqrt{k}/\sqrt{\lambda_2}$.
  - If we have a positive repeated eigenvalue, the graph is a circle.
  - If one eigenvalue is positive and the other negative, the graph is an hyperbola with the principal axes given by directions of the eigenbasis. The semi-major axis is the inverse of the square root of the positive eigenvalue $\sqrt{k}/\sqrt{\lambda_+}$, while the semi-minor axis is the inverse of the square root of the absolute value of the negative eigenvalue $\sqrt{k}/\sqrt{|\lambda_-|}$.
  - Exercise! What happens when both eigenvalues are negative, and what happens when one of the eigenvalue is positive and the other 0?
- For higher dimensions: let $q(x) = \boldsymbol{x} \cdot A\boldsymbol{x}$ be a quadratic form for A a symmetric $M \times M$ matrix. If A has M distinct eigenvalues, we call the eigenspaces of A the *principal axes* of q. They are the axes along which the graph $q(x) = k$ where k is a positive number is aligned. (In particular, the direction with the largest positive eigenvalue will be the direction of the closest approach of the graph to the origin, and the distance to the origin will be $1/\sqrt{\lambda}$ where λ is the largest positive eigenvalue.)

HOMEWORK FOR THIS WEEK

8.1: 10, 12, 14, 22, 24

8.2: 6, 14, 18, 22, 28, 29, 40