Section 07: Solutions

1. Using the Eigenbasis

It's a very useful fact that for any symmetric $n \times n$ matrix A you can find a set of eiegenvectors $u_1, \ldots, u_n$ for A such that:

- $||u_i||_2 = 1$
- $u_i^T u_j = 0, \forall i \neq j$
- $u_1, \ldots, u_n$ form a basis of $\mathbb{R}^n$

One of the reasons this fact is useful is that facts about these matrices are easier to prove if you think about the vectors in terms of their "eigenbasis" components, instead of their components in the standard basis. As a trivial example, we'll show that you can calculate Ax for a vector x without having to do the matrix multiplication.

(a) Consider the matrix $A = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}$. Verify that $u_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $u_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ are eigenvectors and meet the definitions. Find the eigenvalues associated with u_1 and u_2

Solution:

They are eigenvectors: $Au_1 = \begin{bmatrix} 4/\sqrt{2} - 1/\sqrt{2} \\ -1/\sqrt{2} + 4/\sqrt{2} \end{bmatrix} = 3u_1$ $Au_2 = 5u_2$ by a similar calculation.

They are unit norm: $u_1^T u_1 = (1/\sqrt{2})^2 + (1/\sqrt{2})^2 = 1/2 + 1/2 = 1$. The calculation for u_2 is similar.

They are orthogonal: $u_1^T u_2 = (1/\sqrt{2})(-1/\sqrt{2}) + (1/\sqrt{2})(1/\sqrt{2}) = -1/2 + 1/2 = 0$.

They form a basis (since they're 2 linearly independent vectors in $\mathbb{R}^2$)

(b) Since $\{u_1, u_2\}$ are a basis, we can write any vector as a linear combination of them. Write $x = \begin{bmatrix} -1/\sqrt{2} \\ 3/\sqrt{2} \end{bmatrix}$ in this basis.

Solution:

$x^T u_1 = -1/2 + 3/2 = 1$. $x^T u_2 = 1/2 + 3/2 = 2$. So $x = u_1 + 2u_2$.

(c) Use the decomposition and the eigenvalues you calculated to calculate Ax wihtout doing matrix-vector multiplication.

Solution:

$$Ax = A(u_1 + 2u_2) = Au_1 + 2Au_2 = 3u_1 + 10u_2 = \begin{bmatrix} -7/\sqrt{2} \\ 13/\sqrt{2} \end{bmatrix}$$

This method of calculating a matrix vector product won't actually be more computationally efficient – but it's what's "really" happening when you do the multiplication, so this will be useful intuition under certain circumstances. Expressing vectors in an eigenbasis is also a useful proof technique, as we'll see in some later problems.

2. Sets of Eigenvectors

(a) Prove that if A is a symmetric matrix with n distinct eigenvalues, then its eigenvectors are orthogonal. *Hint: if u and v are eigenvectors, calculate $u^T Av$ two different ways.*

Solution:

Here's one possible proof. Let u, v be eigenvectors with eigenvalues λ_u and λ_v respectively (with $\lambda_u \neq \lambda_v$). Consider the quantity $u^T Av$.

On one hand,

$$u^T Av = u^T(\lambda_v v) = \lambda_v \cdot u^T v$$

On the other hand, since A is symmetric:

$$u^T Av = u^T A^T v = (Au)^T v = \lambda_u(u^T v)$$

Combinbing we have $\lambda_v u^T v = \lambda_u u^T v$, since $\lambda_u \neq \lambda_v$, we must have $u^T v = 0$ for all eigenvectors u, v.

(b) Suppose that A is a symmetric matrix. Prove, without appealing to calculus, that the solution to $\arg\max_x x^T Ax$ s.t. $||x||_2 = 1$ is the eigenvector x_1 corresponding to the largest eigenvalue λ_1 of A. (Hint: the eigenvectors of a symmetric matrix can be chosen to be an orthonormal basis, i.e. unit vectors spanning all of $\mathbb{R}^n$.)

Solution:

Let $u_1, \dots, u_n$ be an orthonormal set of unit vectors (which are guaranteed to exist by symmetry of A. Let x be a unit vector. We can write x as $\sum_{i=1}^n \alpha_i u_i$. We claim that $\sum \alpha_i^2 = 1$. Indeed:

$$||x||^2 = ||\sum \alpha_i u_i||^2 \stackrel{*}{=} \sum ||\alpha_i u_i||^2 = \sum \alpha_i^2 ||u_i||^2 = \sum \alpha_i^2$$

Where the starred equality is a result of observing that any cross-terms are 0 by orthogonality of u_i (see a more detailed explanation in Section 4 of the solution).

Now let's examine $x^T Ax$.

$$\begin{aligned} x^T Ax &= x^T A \left(\sum \alpha_i u_i\right) \\ &= x^T \left(\sum \alpha_i \lambda_i u_i\right) \\ &= \left(\sum \alpha_i u_i^T\right)\left(\sum \alpha_i \lambda_i u_i\right) \\ &\stackrel{*}{=} \sum \alpha_i^2 \lambda_i ||u_i||^2 \\ &= \sum \alpha_i^2 \lambda_i \end{aligned}$$

Where again the starred equality uses that cross terms are 0 by orthoganality. Since $\sum \alpha_i^2 = 1$, we are just taking a convex combination of the λ_i. This is clearly maximized by making $\alpha_1 = 1$ where λ_1 is the maximum eigenvalue. Observe that this is indeed possible by setting $x = u_1$ as claimed.

(c) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $\mathbb{R}^{n\times n}$ symmetric matrices. Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ have the exact same set of eigenvectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ with the corresponding eigenvalues $\alpha_1, \alpha_2, \cdots, \alpha_n$ for $\boldsymbol{A}$, and $\beta_1, \beta_2, \cdots, \beta_n$ for $\boldsymbol{B}$. Please write down the eigenvectors and their corresponding eigenvalues for the following matrices:

(i) $\boldsymbol{D} = \boldsymbol{A} - \boldsymbol{B}$

Solution:

Eigenvectors u_i with eigenvalues $\alpha_i - \beta_i$ Since $(A-B)x = Ax - Bx = (\alpha_i - \beta_i)x$.

(ii) $\boldsymbol{E} = \boldsymbol{AB}$ **Solution:**

Eigenvecutors u_i with eigenvalues $\alpha_i\beta_i$ Since $ABu_i = A\beta_i u_i = \beta_i A u_i = \beta_i \alpha_i u_i$.

(iii) $\boldsymbol{F} = \boldsymbol{A}^{-1}\boldsymbol{B}$ (assume $\boldsymbol{A}$ is invertible) **Solution:**

Observe that $A^{-1}u_i = \frac{1}{\alpha_i}u_i$.
To show this we examine $A^{-1}Au_i$.
On the one hand: $A^{-1}Au_i = A^{-1}\alpha_i u_i = \alpha_i A^{-1}u_i$.
On the other hand: $A^{-1}Au_i = Iu_i = u_i$.
Setting both equal to each other, we have $\alpha_i A^{-1}u_i = u_i$, so $A^{-1}u_i = \frac{1}{\alpha_i}u_i$.

Then we have $A^{-1}Bu_i = A^{-1}\beta_i u_i = \beta_i A^{-1}u_i = \frac{\beta_i}{\alpha_i}u_i$
Thus $A^{-1}B$ has eigenvectors u_i with eigenvalues $\frac{\beta_i}{\alpha_i}$.

3. Positive Semi-Definite Matrices

A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is *positive-semidefinite (PSD)* if $x^T\boldsymbol{A}x \geq 0$ for all $x \in \mathbb{R}^n$.

(a) For any $y \in \mathbb{R}^n$, show that yy^T is PSD.

Solution:

Let x be any vector, then the quadratic form we need to examine is $x^T(yy^T)x$. Regrouping, we care about $(x^Ty)(y^Tx)$. Since x^Ty is just a number, and dot products are symmetric, we can rewrite this number as $(x^Ty)^2$, and it is clearly non-negative as required.

(b) Let X be a random vector in $\mathbb{R}^n$ with covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\mathrm{T}]$. Show that $\boldsymbol{\Sigma}$ is PSD.

Solution:

Let y be an arbitrary vector in $\mathbb{R}^n$. We show $y^T\Sigma y \geq 0$.
Expanding the definition of Σ we have

$$y^T\mathbb{E}\left[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T\right]y = \mathbb{E}\left[y^T(X-\mathbb{E}[X])(X-\mathbb{E}[X])^Ty\right]$$

where the equality follows from y being non-random. Observe that for any realization of X this becomes y^Tww^Ty for some vector w, but this is the exact form we looked at in part a, so we know that for any w this is non-negative. The expectation of a random variable must be non-negative if all of its realizations are non-negative, thus Σ is PSD.

(c) Assume $\boldsymbol{A}$ is a symmetric matrix so that $\boldsymbol{A} = \boldsymbol{U}\text{diag}(\alpha)\boldsymbol{U}^T$ where $\text{diag}(\alpha)$ is an all zeros matrix with the entries of α on the diagonal and $\boldsymbol{U}^T\boldsymbol{U} = I$. Show that $\boldsymbol{A}$ is PSD if and only if $\min_i \alpha_i \geq 0$. (Hint: compute $x^T\boldsymbol{A}x$ and consider values of x proportional to the columns of $\boldsymbol{U}$, i.e., the orthonormal eigenvectors).

Solution:

We show first that if all α_i are positive that the resulting matrix is PSD:
Let x be an arbitrary vector in $\mathbb{R}^n$, and consider $x^TU\text{diag}(\alpha)U^Tx$. Let $y = U^Tx$ Then we can rewrite our previous expression as $y^T\text{diag}(\alpha)y = \sum \alpha_i y_i^2 \geq 0$ by the assumption that all α_i are non-negative.

We now show if some α_i is negative, that the matrix is not PSD by finding a vector y such that $y^TAy < 0$.

Let i be an index such that α_i is negative. Let x be the vector which is 1 in entry i and 0 everywhere else, and let $y = Ux$. Then $y^T A y = x^T U^T U \text{diag}(\alpha) U^T U x = x^T \text{diag}(\alpha) x = \alpha_i < 0$. Since we have exhibited a vector y, such that $y^T A y < 0$, A is not PSD.

(d) Show that a real symmetrix matrix is PSD if and only if all of its eigenvalues are non-negative. **Solution:**

Let A be a PSD matrix, and x be an eigenvector, then $x^T A x = x^T \lambda x = \lambda ||x||_2^2$, where λ is the associated eigenvalue, since the quadratic form must be non-negative, and norms are non-negative, λ cannot be negative.

Conversely let $u_1, \ldots, u_n$ be some orthonormal set of eigenvectors, with non-negative eigenvalues $\lambda_1, \ldots, \lambda_n$. By rewriting any vectors in the orthonormal basis, we can easily see what happens:

$$\begin{aligned} x^T A x &= \left(\sum_{i=1}^n \alpha_i u_i\right)^T A \left(\sum_{i=1}^n \alpha_i u_i\right) \\ &= \sum_{i=1}^n (\alpha_i u_i)^T A (\alpha_i u_i) \\ &= \sum_{i=1}^n (\alpha_i u_i)^T (\lambda_i \alpha_i u_i) \\ &= \sum_{i=1}^n \lambda_i \alpha_i^2 \end{aligned}$$

Where again we use the fact that since u_i and u_j are orthogonal means there won't be cross terms. Since $\lambda_i \geq 0$, the quantity is indeed non-negative so the matrix is PSD.

4. Cross term proof

Solution:

$$\begin{aligned} ||\sum_i \alpha_i u_i||^2 &= (\sum_i \alpha_i u_i)^T (\sum_i \alpha_i u_i) \\ &= \sum_i \alpha_i^2 u_i^T u_i + \sum_{i \neq j} \alpha_i \alpha_j u_i^T u_j \\ &= \sum_i \alpha_i^2 u_i^T u_i && (u_i \text{ and } u_j \text{ are perpendicular}) \\ &= \sum_i ||\alpha_i u_i||^2 \end{aligned}$$