DSC 140B - Homework 02

Due: Wednesday, January 21

Instructions:

- Write your solutions to the following problems **by hand**, either on on another piece of paper that you scan or using a tablet. Typed solutions will not be accepted for credit!
- Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer.
- Homework problems are graded pass/fail on completeness and effort, not correctness.
- Homeworks are due via Gradescope at 11:59 PM.

Problem 1. (1 credit)

Suppose (just like in the last homework) that in a group of 1000 people, 600 currently live in California and 400 currently live in Texas. In any given year, 5% of the people living in California move to Texas, and 3% of the people living in Texas move to California. You may assume that the people do not move to any other states.

We can represent the current number of people living in California and Texas with a *population vector*:

$$\vec{p} = (\#\text{ in California}, \#\text{ in Texas})^T.$$

The initial situation described above is represented by the population vector $(600, 400)^T$.

a) Let $\vec{f}(\vec{p})$ be the linear transformation which takes in a current population vector, $\vec{p} = (c, t)^T$, and returns the population vector after one year has passed. In part (b) of the corresponding problem on the last homework, you should have found the following formula for $\vec{f}$ with respect to the standard basis:

$$\vec{f}(\vec{p}) = (.95c + .03t, \quad .05c + .97t)^T$$

Write the *matrix* A representing $\vec{f}$ with respect to the standard basis.

b) Using a matrix multiplication, find the population vector after one year has passed, given that the initial population vector is $(600, 400)^T$. Your result should not contain decimals.

c) In part (f) of the last homework, we saw that two eigenvectors of A are

$$\vec{u}^{(1)} = \begin{pmatrix} 375 \\ 625 \end{pmatrix} \qquad \vec{u}^{(2)} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Verify that these are eigenvectors of the matrix A by performing the matrix multiplication.

d) Write the matrix $A_{\mathcal{U}}$ of the linear transformation $\vec{f}$ with respect to the basis $\mathcal{U} = \{\vec{u}^{(1)}, \vec{u}^{(2)}\}$.

e) In part (f) of the last homework, you found that the initial population vector $\vec{x} = (600, 400)^T$ expressed in the new basis has coordinates $[\vec{x}]_{\mathcal{U}} = (1, 225)^T$.

Compute $A_{\mathcal{U}}[\vec{x}]_{\mathcal{U}}$ and then convert the resulting to a coordinate vector in the standard basis.

Hint: your result should be familiar.

Problem 2. (1.5 credits)

In this problem, we will prove that for a symmetric matrix A, the unit vector $\vec{x}$ that maximizes $\|A\vec{x}\|^2$ is the eigenvector corresponding to the largest eigenvalue (in absolute value).[1]

Let A be a $d \times d$ symmetric matrix and let $\hat{u}^{(1)}, \hat{u}^{(2)}, \ldots, \hat{u}^{(d)}$ be d of its eigenvectors. Assume that they are all mutually orthogonal and have unit norm (the spectral theorem guarantees this is possible to assume). Let $\lambda_1, \lambda_2, \ldots, \lambda_d$ be the corresponding eigenvalues, and assume that they are in decreasing order by magnitude. That is: $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_d|$.

a) The eigenvectors $\hat{u}^{(1)}, \ldots, \hat{u}^{(d)}$ form an orthonormal basis for $\mathbb{R}^d$. This means that any vector $\vec{x} \in \mathbb{R}^d$ can be written as a linear combination:

$$\vec{x} = a_1 \hat{u}^{(1)} + a_2 \hat{u}^{(2)} + \cdots + a_d \hat{u}^{(d)}$$

for some coefficients $a_1, \ldots, a_d \in \mathbb{R}$. This is called the *eigendecomposition* of $\vec{x}$ with respect to the eigenvectors of A.

If $\vec{x}$ is a unit vector (i.e., $\|\vec{x}\| = 1$), show that the coefficients must satisfy $a_1^2 + a_2^2 + \cdots + a_d^2 = 1$.

Hint: Use the fact that the eigenvectors are orthonormal, meaning $\hat{u}^{(i)} \cdot \hat{u}^{(j)} = 0$ if $i \neq j$ and $\hat{u}^{(i)} \cdot \hat{u}^{(i)} = 1$ for all i.

b) Again let $\vec{x} = a_1 \hat{u}^{(1)} + a_2 \hat{u}^{(2)} + \cdots + a_d \hat{u}^{(d)}$ be the eigendecomposition of $\vec{x}$. Show that:

$$A\vec{x} = a_1 \lambda_1 \hat{u}^{(1)} + a_2 \lambda_2 \hat{u}^{(2)} + \cdots + a_d \lambda_d \hat{u}^{(d)}$$

c) Again let $\vec{x} = a_1 \hat{u}^{(1)} + a_2 \hat{u}^{(2)} + \cdots + a_d \hat{u}^{(d)}$ be the eigendecomposition of $\vec{x}$. Show that:

$$\|A\vec{x}\|^2 = \lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 + \cdots + \lambda_d^2 a_d^2$$

d) Remember our original goal: we want to find a unit vector $\vec{x}$ that maximizes $\|A\vec{x}\|^2$.

From part (c), we know that we can write $\|A\vec{x}\|^2$ as

$$\lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 + \cdots + \lambda_d^2 a_d^2,$$

and from part (a), we know that if $\vec{x}$ is a unit vector, then

$$a_1^2 + a_2^2 + \cdots + a_d^2 = 1.$$

So, maximizing $\|A\vec{x}\|^2$ over unit vectors $\vec{x}$ is equivalent to maximizing $\lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 + \cdots + \lambda_d^2 a_d^2$. subject to the constraint $a_1^2 + a_2^2 + \cdots + a_d^2 = 1$.

What choice of $a_1, a_2, \ldots, a_d$ maximizes this quantity? You don't need to *rigorously* prove that your choice is the best one; just explain your reasoning.

Hint: think of the constraint as a "budget". That is, you have a total of 1 unit to distribute among $a_1^2, a_2^2, \ldots, a_d^2$. You'll want to use the fact that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_d|$ to figure out the best allocation.

e) The last part effectively proved that the unit vector maximizing $\|A\vec{x}\|^2$ is the eigenvector corresponding to the largest eigenvalue (in absolute value). Explain (in just a sentence or two) why this follows.

f) Now consider a related but different problem: maximizing $vecx^T A\vec{x}$ subject to $\|\vec{x}\| = 1$. Using a similar approach as above, show that this is maximized by taking $\vec{x}$ to be the eigenvector corresponding to the largest eigenvalue (*not* in absolute value). What is the maximum value of $\vec{x}^T A\vec{x}$ in this case?

Note: Here we care about the largest eigenvalue itself, not the largest in absolute value. For instance, if $\lambda_1 = 5$ and $\lambda_2 = -10$, the maximum of $\vec{x}^T A\vec{x}$ is 5, not 10.

[1]This implies, by the way, that the unit vector that maximizes $\|A\vec{x}\|$ is also the eigenvector corresponding to the largest eigenvalue (in absolute value), since the square root function is monotonically increasing.

Problem 3. (1.5 credits)

Let $g(\vec{x}) = g(x_1, x_2) = 4x_1^2 + 3x_2^2 + 10x_1x_2$, where we've defined $\vec{x} = (x_1, x_2)^T$. In this problem, we will consider maximizing g subject to the constraint $x_1^2 + x_2^2 = 1$.

You saw how to solve optimization problems like this in your multivariate calculus class using the method of *Lagrange multipliers*. Informally-speaking, the idea behind Lagrange multipliers is that the gradient vector of g and the gradient of the constraint $x_1^2 + x_2^2 - 1$ should be parallel at a constrained optimum. Since two vectors $\vec{a}$ and $\vec{b}$ are parallel if and only if $\vec{a} = \lambda\vec{b}$ for some λ, and since the gradient of the constraint is simply $(2x_1, 2x_2)^T = 2\vec{x}$, this means that a local optimum should satisfy $\nabla g(\vec{x}) = 2\lambda\vec{x}$. This looks similar to the eigenvector equation $A\vec{x} = \lambda\vec{x}$; in this problem we'll make the connection clearer.

a) The Lagrange multiplier approach says that we should define the *Lagrangian*:

$$\mathcal{L}(x_1, x_2, \lambda) = g(\vec{x}) - \lambda(x_1^2 + x_2^2 - 1)$$

We then solve the system of three equations in three unknowns:

$$\frac{\partial \mathcal{L}}{\partial x_1} = 0$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

Write out and solve this system for x_1, x_2, and λ.

Hint: Try to get a formula for x_1^2 in terms of λ only, and same for x_2^2. When you get to this point, you will be able to substitute your formulas for x_1^2 and x_2^2 into $\partial\mathcal{L}/\partial\lambda = 0$ to get a function of the form

$$\frac{a_1}{(b_1\lambda + c_1)^2 + d_1} + \frac{a_2}{(b_2\lambda + c_2)^2 + d_2} - 1 = 0,$$

where the a, b, c, d's are all constants. We want to solve this for λ, which is not easy to do analytically. Instead, solve it numerically using `scipy.optimize.fsolve`, or similar. Once you've solved for λ, you can plug it back in to your equations for x_1^2 and x_2^2. Some of the possible combinations of x_1 and x_2 you get may not actually solve the original system of equations; be sure to check which ones do by plugging them back into the original equations and making sure they equal zero.

Note: when you use code (like `fsolve`) to solve part of the problem, you do *not* need to include your code in your writeup. Just describe what you did.

b) The equation $g(\vec{x}) = 4x_1^2 + 3x_2^2 + 10x_1x_2$ can be written in matrix-vector form as $g(\vec{x}) = \vec{x}^T A\vec{x}$ for an appropriately-defined matrix, A. Find this matrix A, and show that the matrix form is equivalent to the original form.

You can assume that A is symmetric.

c) Using whatever method you choose (e.g., numpy), compute the eigenvectors and eigenvalues of A. Show that the eigenvectors are the same as your solution to part (a).

d) We saw in lecture that a matrix can be interpreted as the representation of a linear transformation $\vec{f}(\vec{x})$. It turns out that A represents the *gradient* of $\vec{g}$.

Show that A represents the linear transformation $\vec{f}(\vec{x}) = \frac{1}{2}\nabla g(\vec{x})$, $\nabla g(\vec{x}) = (\partial g/\partial x_1,$ where $\partial g/\partial x_2)^T$ is the *gradient* of g.

Therefore, for a function g of the form $ax_1^2 + bx_2^2 + cx_1x_2$, the gradient is a linear transformation that can be computed by a matrix multiplication, and the method of Lagrange multipliers is equivalent to finding an eigenvector of the matrix representing the gradient.