

Proof that Eigenvalues are LaGrange Multipliers

Emil Kerimov

February 7, 2017

1 Overview of matrix derivatives

$$\frac{\partial x^T A x}{\partial x} = x^T (A^T + A) \quad (1)$$

$$\frac{\partial x^T a}{\partial x} = a^T \quad (2)$$

2 Overview of LaGrange Multipliers

For our purposes we only need the simplified version of LaGrange Multipliers

If

$$\max F(\vec{x})$$

subject to

$$g(\vec{x}) = \vec{c}$$

Then

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = F(\vec{x}) - \vec{\lambda}^T (g(\vec{x}) - \vec{c})$$

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

$$\frac{\partial \mathcal{L}}{\partial \vec{x}} = 0$$

3 Proof

Consider the problem of finding a vector on the unit sphere that maximizes the quadratic form of a **symmetric** square matrix A . Written as an optimization problem we get:

$$\begin{aligned} \max_x & x^T A x \\ & x^T x = 1 \end{aligned} \quad (3)$$

Since the unit sphere is a compact and bounded space in R^n we know that a maximum (and a minimum) exist. We can use LaGrange multipliers to uniquely solve this problem. Using the method we get the function:

$$g(x, \lambda) = x^T A x - \lambda \cdot (x^T x - 1)$$

Taking derivatives:

$$\begin{aligned}\frac{\partial g(x, \lambda)}{\partial x} &= 2x^T A - 2\lambda \cdot x^T \\ \frac{\partial g(x, \lambda)}{\partial \lambda} &= -(x^T x - 1)\end{aligned}$$

Therefore the optimal values of x and λ will satisfy:

$$A^T x_{opt} = Ax_{opt} = \lambda_{opt} \cdot x_{opt} \quad (4)$$

$$x_{opt}^T x_{opt} = 1 \quad (5)$$

Denote $x_1 = x_{opt}$ and $\lambda_1 = \lambda_{opt}$. We can choose to name these as our first eigenvector and first eigenvalue respectively.

Let's attempt to find a perpendicular vector to the one we found that maximizes quadratic form of A in a different direction.

$$\begin{aligned}\max_x \quad & x^T A x \\ & x^T x = 1 \\ & x^T x_1 = 0\end{aligned} \quad (6)$$

Solving we get:

$$g(x, \lambda, \mu) = x^T A x - \lambda \cdot (x^T x - 1) - \mu \cdot x^T x_1$$

Taking derivatives:

$$\frac{\partial g(x, \lambda, \mu)}{\partial x} = 2x^T A - 2\lambda \cdot x^T - \mu \cdot x_1^T = 0$$

Taking transposes and using $A^T = A$:

$$2Ax - 2\lambda \cdot x - \mu \cdot x_1 = 0$$

Taking the dot product with respect to x_1^T we get:

$$(2Ax)x_1^T - (2\lambda x)x_1^T - (\mu x_1)x_1^T = 0$$

Using

$$\begin{aligned}(2Ax)x_1^T &= 2A^T(x x_1^T) = 0 \\ x_1 x_1^T &= 1\end{aligned}$$

We get

$$(2Ax)x_1^T - (2\lambda x)x_1^T - (\mu x_1)x_1^T = 0 + 0 - \mu = 0$$

Therefore the optimal values of x and λ will satisfy:

$$Ax_{opt} = \lambda_{opt} \cdot x_{opt} \quad (7)$$

Denote $x_2 = x_{opt}$ and $\lambda_2 = \lambda_{opt}$. We can choose to name these as our second eigenvector and second eigenvalue respectively.

The same argument extends for consecutive eigenvalues and eigenvectors. Interpreting this proof leads to the concept of Principle Component Analysis.