EIGENVALUES AND THE POWER METHOD

David F. Gleich
August 21, 2023

Learning objectives
1. See a variety of ways to think about eigenvalues
2. Look at the power method

Note, these notes are still being edited. There are a huge diversity of perspectives and geometric interpretations of eigenvalues and eigenvectors, so it's challenging to know how to show them. I'm working on some pictures to help.

There are a variety of ways to derive and define the eigenvalues of a matrix $\boldsymbol{A}$. The most general definition of an eigenvalue of a matrix is a value λ such that $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$. This definition, however, obscures much of the utility of eigenvalues of symmetric matrices (which are extremely common).

1 CRITICAL DIRECTIONS

** Still working on this section. Skip it now! **

The eigenvalues and eigenvectors of a symmetric, positive definite matrix $\boldsymbol{A}$ are critical directions in the quadratic function

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \boldsymbol{A}\mathbf{x}$$

that are invariant to transformations.

For a symmetric matrix $\boldsymbol{A}$, then the eigenvalues of $\boldsymbol{A}$ are the *stationary points* of the following optimization problem:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \mathbf{x}^T \boldsymbol{A}\mathbf{x} \\ \text{subject to} & \|\mathbf{x}\|^2 = 1 \end{array} \tag{1}$$

2 STATIONARY POINTS

To go ahead and define something in terms of another definition: stationary points are those points where the Lagrangian of the problem has zero derivative. And what is the Lagrangian? It's a function that balances tradeoffs between the objective function $\mathbf{x}^T \boldsymbol{A}\mathbf{x}$ and the constraint $\|\mathbf{x}\|^2 = 1$

$$\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{x}^T \boldsymbol{A}\mathbf{x} - \lambda \cdot (\mathbf{x}^T \mathbf{x} - 1).$$

The gradient of this function is just

$$\partial \mathcal{L} / \partial \mathbf{x} = 2\boldsymbol{A}\mathbf{x} - 2\lambda \mathbf{x}$$

$$\partial \mathcal{L} / \partial \lambda = \mathbf{x}^T \mathbf{x} - 1.$$

So at a stationary point, by definition, we have

$$\boldsymbol{A}\mathbf{x} = \lambda \mathbf{x} \qquad \mathbf{x}^T \mathbf{x} = 1.$$

Conclusion: any stationary point of (1) is a pair:

$$(\mathbf{x}, \lambda) \text{ where } \boldsymbol{A}\mathbf{x} = \lambda \mathbf{x}$$

which implies that $(\boldsymbol{A} - \lambda \boldsymbol{I})\mathbf{x} = 0$ and also that $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$.

Note that this analysis gives the same result for *minimizing* the problem instead of maxing

3 THE POWER METHOD TO FIND EIGENVALUES

Given that we have an optimization problem, one strategy to produce an algorithm is to seek a maximizer of $\mathbf{x}^T \boldsymbol{A}\mathbf{x}$ where $\|\mathbf{x}\| = 1$. Because the goal is a maximizer, we would do gradient *ascent* instead of gradient *descent*. However, this time we have a constraint that makes the problem more complicated. A simplistic strategy to handle this constraint is just to take a gradient step:

$$\mathbf{y} = \mathbf{x}^{(k)} + 2\gamma_k \boldsymbol{A}\mathbf{x}^{(k)}$$

and to project it back onto the feasible set:

$$\mathbf{x}^{(k+1)} = \text{argmin}_{\mathbf{z}} \|\mathbf{z} - \mathbf{y}\| \text{ where } \|\mathbf{z}\| = 1.$$

A quick analysis similar to~(1) shows that $\mathbf{z} = \gamma\mathbf{y}$ for some γ such that $\|\mathbf{z}\| = 1$. That is to say, we just take $\mathbf{y}$ and normalize it.

This gives us the iteration:

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{x}^{(k)} + 2\gamma_k \boldsymbol{A}\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)} + 2\gamma_k \boldsymbol{A}\mathbf{x}^{(k)}\|}.$$

Again, we are interested in maximizing $\mathbf{x}^{(k+1)^T} \boldsymbol{A}\mathbf{x}^{(k+1)}$. This suggests taking γ_k large. In the limit as $\gamma_k \to \infty$ [1] then we find that

[1] — TODO – work out this derivation more. Can we show that $\gamma_k \to \infty$ is a natural step?

$$\mathbf{x}^{(k+1)} = (\boldsymbol{A}\mathbf{x})/\|\mathbf{x}\|.$$

This is the power method!

DEFINITION 1 (the power method) *Let $\mathbf{x}^{(0)}$ be any vector. Then the power method is the iteration*

$$\mathbf{x}^{(k+1)} = (\boldsymbol{A}\mathbf{x}^{(k)})/\|\mathbf{x}^{(k)}\|.$$

There are no *eigenvalues* in the power method. Instead, there are only eigenvectors. To get the eigenvalue, we need to look at the Rayleigh quotient

$$\lambda^{(k)} = \mathbf{x}^{(k)^T} \boldsymbol{A}\mathbf{x}^{(k)}.$$

This quantity can often be computed with minimal overhead because we need to compute the vector $\boldsymbol{A}\mathbf{x}^{(k)}$ to get the next iterate of the power method.

4 CONVERGENCE OF THE POWER METHOD

First, we need to show that the power method is really a simple algorithm. That is, we need to show that $\mathbf{x}^{(k)} = \boldsymbol{M}^k\mathbf{x}^{(0)}$ for some matrix. This type of simple statement will not quite be possible, we just need one slight correction to handle a sticky situation with the norm.

THEOREM 2 *Let $\mathbf{x}^{(k)}$ be the kth iterate of the power method starting from $\mathbf{x}^{(0)}$. Then $\mathbf{x}^{(k)} = \boldsymbol{A}^k\mathbf{x}^{(0)}/\|\boldsymbol{A}^k\mathbf{x}^{(0)}\|$.*

Proof This holds for $\mathbf{x}^{(1)}$ given that this is the explicit iteration. To show that it holds for all future iterations, we proceed inductively. Assume that it is true for the kth iteration: $\mathbf{x}^{(k)} = \boldsymbol{A}^k\mathbf{x}^{(0)}/\|\boldsymbol{A}^k\mathbf{x}^{(0)}\|$. This also means that $\mathbf{x}^{(k)} = \rho_k\boldsymbol{A}^k\mathbf{x}^{(0)}$ for some scalar ρ_k. Thus,

$$\mathbf{x}^{(k+1)} = \rho_k\boldsymbol{A}^{k+1}\mathbf{x}^{(k)}/\|\rho_k\boldsymbol{A}^{k+1}\mathbf{x}^{(k)}\|.$$

And now ρ_k cancels out of the equation and we are done. ■

This shows that we do have a simple algorithm and also that the long term behavior will be governed by matrix powers.

5 TERMINATION

A good way to terminate the iteration is to check if we satisfy the eigenvalue residual

$$\boldsymbol{A}\mathbf{x}^{(k)} - \lambda^{(k)}\mathbf{x}^{(k)} \approx 0$$