1 Outline

In this lecture, we study

- the power method for computing the top eigenvector,
- gradient descent for computing the top eigenvector,
- connections to the singular value decomposition (SVD).

2 Computing the Top Eigenvector

Given a positive semidefinite matrix $A \in \mathbb{R}^{d\times d}$, we want to compute the top eigenvector of A which corresponds to the largest eigenvalue of A. Let $\lambda_{\max}$ denote the largest eigenvalue of A, and let $v_{\max}$ denote the top eigenvector. The closely related notion is the **Rayleigh quotinet**, defined as

$$R(A,x) = \frac{x^\top Ax}{x^\top x} \quad \text{for any nonzero } x.$$

It is known that for any symmetric matrix A,

$$\max_{x\in\mathbb{R}^d: x\neq 0} R(A,x) = \lambda_{\max},$$
$$\operatorname*{argmax}_{x\in\mathbb{R}^d: x\neq 0} R(A,x) = v_{\max}.$$

This implies that we may compute the top eigenvector of a symmetric matrix A by solving

$$\max_{x\in\mathbb{R}^d} \quad x^\top Ax \quad \text{subject to} \quad \|x\|_2 = 1.$$

If A is positive semidefinite, then $x^\top Ax \geq 0$ for any $x \in \mathbb{R}^d$, which implies that one can replace the constraint $\|x\|_2 = 1$ by $\|x\|_2 \leq 1$ to deduce the equivalent formulation

$$\max_{x\in\mathbb{R}^d} \quad x^\top Ax \quad \text{subject to} \quad \|x\|_2 \leq 1.$$

Here, the feasible region $\{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$ is a convex set. However, the objective is to "maximize" the convex function $x^\top Ax$. Hence, the optimization problem is a nonconvex problem. Nevertheless, there exists an efficient algorithm for computing the top eigenvector.

2.1 Power Method

We present the **power method** which is known to compute the top eigenvector of a positive semidefinite matrix. The power method works with as in Algorithm 1.

To analyze the power method, we explain some necessary background. Let A be a positive semidefinite matrix with eigenvalues

$$\lambda_1 \geq \cdots \geq \lambda_d \geq 0.$$

Algorithm 1 Power Method

Initialize $\bar{x}_0 \in \mathbb{R}^d \setminus \{0\}$ and $x_0 = \bar{x}_0/\|\bar{x}_0\|_2$
for $t = 1, \ldots, T$ **do**
  Update $\bar{x}_t = A^t x_0$
  Obtain $x_t = \bar{x}_t/\|\bar{x}_t\|_2$
end for
Return x_T.

Let v_i denote the eigenvector associated with λ_i for $i \in [d]$. By the well-known **spectral theorem**, we may assume that $v_1, \ldots, v_d$ are mutually orthogonal and $\|v_i\| = 1$ for $i \in [d]$. Then $v_1, \ldots, v_d$ form an orthonormal basis of $\mathbb{R}^d$.

Theorem 10.1. *Suppose that $\lambda_1 > \lambda_2$. For any $\epsilon > 0$, if*

$$t \geq \frac{\lambda_1}{2(\lambda_1 - \lambda_2)} \log\left(\frac{1}{\epsilon(v_1^\top x_0)^2}\right),$$

then

$$(v_1^\top x_t)^2 \geq 1 - \epsilon$$

Proof. Since $v_1, \ldots, v_d$ form an orthonormal basis of $\mathbb{R}^d$, we have

$$x_0 = \sum_{i=1}^d \alpha_i v_i$$

where $\alpha_i = (v_i^\top x_0)$. Moreover, as $\bar{x}_t = A^t x_0$, it follows that

$$\bar{x}_t = \sum_{i=1}^d \alpha_i A^t v_i = \sum_{i=1}^d \alpha_i \lambda_i^t v_i.$$

Note that

$$x_t = \sum_{i=1}^d (v_i^\top x_t) v_i \quad \text{and} \quad \|x_t\|_2^2 = \sum_{i=1}^d (v_i^\top x_t)^2.$$

Since $\|x_t\|_2 = 1$, we have

$$1 - (v_1^\top x_t)^2 = \sum_{i=2}^d (v_i^\top x_t)^2 = \frac{1}{\|\bar{x}_t\|_2^2} \sum_{i=2}^d (v_i^\top \bar{x}_t)^2 = \frac{\sum_{i=2}^d \alpha_i^2 \lambda_i^{2t}}{\sum_{i=1}^d \alpha_i^2 \lambda_i^{2t}} \leq \frac{\lambda_2^{2t} \sum_{i=2}^d \alpha_i^2}{\alpha_1^2 \lambda_1^{2t}} \leq \frac{\lambda_2^{2t}}{\alpha_1^2 \lambda_1^{2t}}.$$

This implies that

$$1 - (v_1^\top x_t)^2 \leq \left(\frac{\lambda_2}{\lambda_1}\right)^{2t} \frac{1}{(v_1^\top x_0)^2}.$$

Hence, if

$$t \geq \frac{1}{2} \cdot \frac{1}{\log(\lambda_1/\lambda_2)} \log\left(\frac{1}{\epsilon(v_1^\top x_0)^2}\right),$$

then $(v_1^\top x_t)^2 \geq 1 - \epsilon$. Here,

$$\log\left(\frac{\lambda_1}{\lambda_2}\right) = \log\left(\frac{1}{1 - \frac{\lambda_1 - \lambda_2}{\lambda_1}}\right) = -\log\left(1 - \frac{\lambda_1 - \lambda_2}{\lambda_1}\right) \geq \frac{\lambda_1 - \lambda_2}{\lambda_1}.$$

Therefore, if

$$t \geq \frac{\lambda_1}{2(\lambda_1 - \lambda_2)} \log\left(\frac{1}{\epsilon(v_1^\top x_0)^2}\right),$$

then $(v_1^\top x_t)^2 \geq 1 - \epsilon$, as required. □

What does the theorem imply? Note that

$$\sum_{i=2}^{d} (v_i^\top x_t)^2 = \|x_t\|_2^2 - (v_1^\top x_t)^2 = 1 - (v_1^\top x_t)^2 \leq \epsilon.$$

Therefore, x_t is aligned with v_1, which is the top eigenvector. Moreover, the bound has the term $(v_1^\top x_0)^2$. If we sample x_0 from the multivariate normal distribution $\mathcal{N}(0, I_d)$ where I_d is the $d \times d$ identity matrix, then we can argue that $(v_1^\top x_0) = 1/\text{poly}(d)$ with high probability. In that case, the number of required iterations is

$$O\left(\frac{\lambda_1}{\lambda_1 - \lambda_2} \log\left(\frac{d}{\epsilon}\right)\right).$$

Here, the bound depends on the values of λ_1 and λ_2. The following result provides a convergence guarantee of the power method that does not depend on λ_1 and λ_2.

Theorem 10.2. *For any $\epsilon > 0$, if*

$$t \geq \frac{1}{\epsilon} \log\left(\frac{2}{\epsilon(v_1^\top x_0)^2}\right),$$

then

$$x_t^\top A x_t \geq (1 - \epsilon)\lambda_1.$$

Proof. Let k denote the largest index such that

$$\lambda_k \geq \left(1 - \frac{\epsilon}{2}\right)\lambda_1.$$

Let $\alpha_i = (v_i^\top x_0)$ for $i \in [d]$. Note that

$$1 - \sum_{i=1}^{k} (v_i^\top x_t)^2 = \sum_{i=k+1}^{d} (v_i^\top x_t)^2 = \frac{1}{\|\bar{x}_t\|_2^2} \sum_{i=k+1}^{d} (v_i^\top \bar{x}_t)^2 = \frac{\sum_{i=k+1}^{d} \alpha_i^2 \lambda_i^{2t}}{\sum_{i=1}^{d} \alpha_i^2 \lambda_i^{2t}} \leq \left(1 - \frac{\epsilon}{2}\right)^{2t} \frac{1}{(v_1^\top x_0)^2}.$$

Moreover,

$$x_t^\top A x_t = \sum_{i=1}^{d} \lambda_i (v_i^\top x_t)^2 \geq \left(1 - \frac{\epsilon}{2}\right)\lambda_1 \sum_{i=1}^{k} (v_i^\top x_t)^2.$$

Combining the two inequalities, it follows that

$$x_t^\top A x_t \geq \left(1 - \frac{\epsilon}{2}\right)\lambda_1 \left(1 - \left(1 - \frac{\epsilon}{2}\right)^{2t} \frac{1}{(v_1^\top x_0)^2}\right).$$

Here, if

$$t \geq \frac{1}{\epsilon} \log\left(\frac{2}{\epsilon(v_1^\top x_0)^2}\right),$$

then

$$x_t^\top A x_t \geq \left(1 - \frac{\epsilon}{2}\right)^2 \lambda_1 \geq (1 - \epsilon)\lambda_1,$$

as required. □

Again, If we sample x_0 from the multivariate normal distribution $\mathcal{N}(0, I_d)$, then the number of required iterations is

$$O\left(\frac{1}{\epsilon}\log\left(\frac{d}{\epsilon}\right)\right)$$

with high probability.

2.2 Gradient Descent

Recall that for a positive semidefinite matrix A, the top eigenvector of A can be computed by solving

$$\max_{x\in\mathbb{R}^d} \quad x^\top Ax \quad \text{subject to} \quad \|x\|_2 \leq 1.$$

Note that this problem is equivalent to

$$\min_{x\in\mathbb{R}^d} \quad -\frac{1}{2}x^\top Ax \quad \text{subject to} \quad \|x\|_2 \leq 1.$$

We may apply projected gradient descent to solve this minimization problem. Note that the projection of any vector u to the ball $\{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$ is given by $u/\|u\|_2$. Then projected gradient descent proceeds with

$$\begin{aligned} \tilde{x}_{t+1} &= x_t + \eta Ax_t = (I + \eta A)x_t, \\ x_{t+1} &= \frac{\tilde{x}_{t+1}}{\|\tilde{x}_{t+1}\|_2}. \end{aligned}$$

Note that this update rule is equivalent to

$$\begin{aligned} \bar{x}_t &= (I + \eta A)^t x_0, \\ x_t &= \frac{\tilde{x}_t}{\|\tilde{x}_t\|_2}. \end{aligned}$$

This corresponds to the power method applied to compute the top eigenvector of matrix $I + \eta A$. In fact, we have

$$\underbrace{\lambda_i(I + \eta A)}_{\text{The } i\text{th largest eigenvalue of } I+\eta A} = (1 + \eta) \underbrace{\lambda_i}_{\text{The } i\text{th largest eigenvalue of } A}.$$

Moreover, when v_i is the eigenvector that corresponds to λ_i, we have that v_i is the eigenvector that corresponds to the ith largest eigenvalue of $I + \eta A$. This implies that the gradient descent method indeed computes the top eigenvector of A, and the covergence rate is

$$O\left(\frac{(1+\eta)\lambda_1}{(1+\eta)\lambda_1 - (1+\eta)\lambda_2}\log\left(\frac{d}{\epsilon}\right)\right) = O\left(\frac{\lambda_1}{\lambda_1 - \lambda_2}\log\left(\frac{d}{\epsilon}\right)\right).$$

2.3 Computing the Largest Singular Value

Let $A \in \mathbb{R}^{n\times p}$ be an $n \times p$ matrix. The famous **singular value decomposition (SVD)** theorem states that A can written as

$$A = U\Sigma V^\top$$

where

- $r = \min\{n, p\}$,
- $U \in \mathbb{R}^{n\times r}$ is an $n \times r$ matrix with orthonormal columns,
- $\Sigma \in \mathbb{R}^{r\times r}$ is a $r \times r$ diagonal matrix with entries $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$,
- $V \in \mathbb{R}^{p\times r}$ is a $p \times r$ matrix with orthonormal columns.

Here, $\sigma_1, \ldots, \sigma_r$ are the **singular values** of A. Let u_1 denote the **top left singular vector** that corresponds to σ_1, and let v_1 denote the **top right singular vector** corresponding to σ_1. Note that

$$A^\top A = V\Sigma^2 V^\top \quad \text{and} \quad AA^\top = U\Sigma^2 U^\top$$

which implies that

$$\sigma_1 = \sqrt{\lambda_{\max}(A^\top A)} = \sqrt{\lambda_{\max}(AA^\top)}.$$

Therefore, by applying the power method to $A^\top A$, one can obtain the top right singular vector v_1. Similarly, by applying the power method to $AA^\top$, we can deduce the top left singular vector u_1.

Algorithm 2 Power Method for the Top Left Singular Vector

Initialize $\bar{u}_0 \in \mathbb{R}^n \setminus \{0\}$ and $u_0 = \bar{u}_0/\|\bar{u}_0\|_2$
for $t = 1, \ldots, T$ **do**
  Update $\bar{u}_t = (AA^\top)^t x_0$
  Obtain $u_t = \bar{u}_t/\|\bar{u}_t\|_2$
end for
Return u_T.

Algorithm 3 Power Method for the Top Right Singular Vector

Initialize $\bar{v}_0 \in \mathbb{R}^p \setminus \{0\}$ and $v_0 = \bar{v}_0/\|\bar{v}_0\|_2$
for $t = 1, \ldots, T$ **do**
  Update $\bar{v}_t = (A^\top A)^t v_0$
  Obtain $v_t = \bar{v}_t/\|\bar{v}_t\|_2$
end for
Return v_T.