Computational methods-Lecture 11

Eigenvalues; singular value decomposition; power method

1 Eigenvalues and eigenvectors

Recall:
For a square matrix A, if there exists λ and $x \neq 0$ such that

$$Ax = \lambda x,$$

then λ is called an eigenvalue and x is called an eigenvector. The pairm (λ, x) is called an eigen-pair.

The eigenvalues are roots of the characteristic polynomial

$$\varphi(\lambda) = \det(\lambda I - A).$$

Theorem 1 (Schur theorem)**.** *For a real square matrix A, there exists an orthogonal matrix Q such that*

$$Q^T A Q = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \cdots & \vdots \\ & & & R_{mm} \end{pmatrix}$$

where R_{ii} is a number or a 2×2 matrix with two complex eigenvalues, conjugate with each other.

If A is real symmetric, all R_{ii} are numbers and the matrix on the right hand side is in fact diagonal.

The largest eigenvalues or smallest eigenvalues can often be reduced to the max or min of the **Rayleigh quotient**:

$$R(x) := \frac{\langle Ax, x\rangle}{|x|^2}.$$

2 Singular value decomposition (SVD)

For non-square matrix, there is no eigenvalues. However, we can generalize the idea to the so-called **singular values**. In particular, consider A of size $m \times n$ and $A^T A$.

Since A^TA is positive semi-definite, we can find orthogonal matrix V of size $n \times n$ such that

$$A^TA = VDV^T.$$

Here, the columns of V are eigenvectors of A^TA and

$$V^T(A^TA)V = D = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \cdots & \\ & & & \sigma_n^2 \end{pmatrix}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

Define the matrix Σ of size $m \times n$. Let $\Sigma_{ii} = \sigma_i$ for $i \leq \min(m, n)$, and other entries are zero.

Suppose $\sigma_k > 0$ and $\sigma_{k+1} = 0$ (if all nonzero, then $k = n$; if all zero, then $k = 0$). Consider

$$u_i = \frac{1}{\sigma_i} Av_i, \quad i = 1, \cdots, k$$

where v_i's are the columns of V.

We have the following observation:

Lemma 1. *$k \leq \min(m, n)$ and u_i's are ortho-normal.*

The first is clear. The reason is that k equals the rank of A^TA and thus the rank of A. To show that they are orthonormal, we just note

$$u_i^T u_j = \frac{1}{\sigma_i^2} v_i^T A^T A v_j,$$

and use the fact that v_i's are eigenvectors.

Then, we can add $m - k$ orthonormal vectors into the list $\{u_1, \cdots, u_k\}$ to form an orthonormal basis of $\mathbb{R}^m$. Let U be the matrix whose columns are these vectors. Then, one has:

Theorem 2 (The singular value decomposition)**.** *$A = U\Sigma V^T$.*

Proof. One only has to verify

$$AV = U\Sigma.$$

For $i \leq k$, $Av_i = \sigma_i u_i$ holds by definition. For $i > k$, $A^TAv_i = 0$, so $v_i^T A^T A v_i = 0$. Hence, $Av_i = 0$. Clearly, $\sigma_i u_i = 0$ since $\sigma_i = 0$. □

The SVD tells us that

$$A = \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

Exercise: Show that u_i's are eigenvectors of AA^T.

Exercise: Show that the 2-norm of A is the largest singular value and the Frobenius norm is the square root of the sum of squares of all singular values.

Example find the singular value decomposition for

$$A = \begin{pmatrix} 1 & 0 \\ 2 & 0 \\ 0 & -1 \\ 0 & 2 \end{pmatrix}$$

3 Power method

The idea of power method is as follows: suppose λ_1 is the eigenvalue with largest magnitude and the multiplicity is 1:

$$|\lambda_1| > |\lambda_2| \geq \cdots |\lambda_n| > 0.$$

Suppose first that A has n independent eigenvectors. Then, for any x_0

$$x_0 = \sum_{j=1}^{n} \beta_j v_j.$$

If we multiply A, we have

$$Ax_0 = \sum_{j=1}^{n} \beta_j \lambda_j v_j.$$

If we do this k times

$$A^k x_0 = \sum_{j=1}^{n} \beta_j \lambda_j^k v_j.$$

The observation is that $|\lambda_1^k| \gg |\lambda_j|^k$ for $j \geq 2$. Hence, the vector

$$A^k x_0$$

will be close to the leading eigenvector. This then motivates the power method for finding the leading eigen-pair.

If $|\lambda_1| < 1$, the iteration $x^{(k)} = A^k x_0 \to 0$. If $|\lambda_1| > 1$, it diverges. Hence, to get a regular result, we need to normalize the vectors.

1. Choose $x_0 \neq 0$; $x \leftarrow x_0$
2. For $k = 1, \cdots$

   $u \leftarrow Ax$

   $x = u/u_{i_0}$, where i_0 is chosen such that $|u_{i_0}| = \|u\|_\infty$.
3. Lastly, compute the eigenvalue

$$\lambda \leftarrow \frac{x^T Ax}{x^T x}$$

There are some observations:

- If the matrix A is sparse, then the matrix-vector multiplication can be fast.
- If there no enough eigenvectors, the method still works. (Try to show this using Jordan block).
- If λ_1 has multiplicity larger than 1, the method still works.

Clearly, if we set $\mu_k = u_{i_0}$ in the kth iteration, then

$$x^{(m)} = \frac{A^m x_0}{\prod_{k\le m} \mu_k}$$

Using this, it is easy to see

Theorem 3. *If in the expansion of x_0 $\beta_1 \neq 0$, then $x^{(m)}$ converges to the leading eigenvector and $\lambda^{(k)}$ converges to the eigenvalue. Moreover, the error estimate holds*

$$|\lambda^{(k)} - \lambda_1| \le C|\frac{\lambda_2}{\lambda_1}|^k.$$

If the matrix A is real symmetric, the error can be improved to $|\lambda_2/\lambda_1|^{2k}$

If one desires other eigenvalues, one can do shifting or do deflation (like the classical Wielandt deflation). We omit these here.

4 The Householder's method

Let $w \in \mathbb{R}^n$ with $|w| = 1$. We call the following matrix

$$P = I - 2ww^T$$

the Householder transform.

The Householder matrix is orthogonal. Its eigenvalues are 1 and -1.

$$P^{-1} = P^T = P.$$

The Householder transform is in fact the reflection about the hyperplane perpendicular to w. To see this, we decompose v into $v = x + y$ where $x \parallel w$ while $y \perp w$. Then,

$$Pv = P(x + y) = x - y.$$

Due to this geometric meaning, for any x, y with $|x| = |y|$, we can find such a reflection operator that maps x into y. In fact, the difference between x and y is perpendicular to the symmetric plane, so parallel to $x - y$. Hence, we can choose

$$w = \frac{x - y}{\|x - y\|_2}.$$

In other words,

$$P_{x \to y} = I - 2\frac{(x - y)(x - y)^T}{\|x - y\|_2^2} \left(= I - 2\frac{(x - y) \otimes (x - y)}{|x - y|^2} \right).$$

With this observation, we can get a useful result:

Proposition 1. *For any nonzero x, there exists a Householder transform H such that*

$$Hx = -\sigma e_1,$$

where

$$\sigma = \operatorname{sgn}(x_1)\|x\|_2.$$

According to the construction above,

$$H = I - \beta^{-1}uu^T, \quad u = x + \sigma e_1, \quad \beta = \frac{1}{2}\|u\|^2 = \sigma(\sigma + x_1).$$