CS205 Review Session # 3 Notes

Linear Dependence

In order to prove that a set of nonzero vectors are linearly independent, we can assume that they are linearly *dependent* and show that this leads to a contradiction. If a set of vectors $\mathbf{V} = \{\mathbf{v}_i \neq \mathbf{0} \mid i \in \{1, \ldots, m\}\}$ is linearly dependent, the following three conditions all hold:

1. $\exists c_i \in \mathbb{R}$ (at least two of which are nonzero) such that $\sum_{i=1}^{m} c_i \mathbf{v}_i = \mathbf{0}$
2. $\exists \mathbf{v}_k \in \mathbf{V}$ such that $\mathbf{v}_k = \sum_{i \neq k} c_i \mathbf{v}_i = \mathbf{0}$ with at least one $c_i \neq 0$
3. $\exists \mathbf{v}_k \in \mathbf{V}$ that can be written as a linear combination of a *linearly independent* subset of $\mathbf{V}$, i.e. $\mathbf{v}_k = \sum_{i=1}^{\ell} c_i \mathbf{v}_i$ (with some renumbering), where the choice of $\{c_i\}$ is *unique*.

This last formulation gives, in some sense, the minimal linearly dependent subset of $\mathbf{V}$, since the removal of any vector from $\{\mathbf{v}_k, \mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ yields a linearly independent set.

Deflation Matrices

Consider a symmetric $n \times n$ matrix $\mathbf{A}$ where λ is an eigenvalue of A and $\mathbf{q}$ the corresponding (normalized) eigenvector. We define the **deflation** of A with respect to $(\lambda, \mathbf{q})$ to be:

$$\mathbf{D}_{(\lambda,\mathbf{q})} = \mathbf{A} - \lambda \mathbf{q}\mathbf{q}^T$$

We note that this matrix has the following properties:

1. $\mathbf{q} \in \text{NS}(\mathbf{A} - \lambda \mathbf{q}\mathbf{q}^T)$. Equivalently, we can say that $\mathbf{q}$ is now an eigenvector corresponding to the *zero* eigenvalue of $\mathbf{D}_{(\lambda,\mathbf{q})}$. To prove this, we have:

$$\left(\mathbf{A} - \lambda \mathbf{q}\mathbf{q}^T\right)\mathbf{q} = \mathbf{A}\mathbf{q} - \lambda \mathbf{q}\mathbf{q}^T\mathbf{q} = \lambda\mathbf{q} - \lambda\mathbf{q} = \mathbf{0}$$

2. Any other eigenvector $\mathbf{q}^*$ of $\mathbf{A}$ with associated eigenvalue λ^* (possibly, but not necessarily, distinct from λ) is now an eigenvector of $\mathbf{D}_{(\lambda,\mathbf{q})}$ with the same associated eigenvalue λ^*. To prove this, consider:

$$\left(\mathbf{A} - \lambda \mathbf{q}\mathbf{q}^T\right)\mathbf{q}^* = \mathbf{A}\mathbf{q}^* - \lambda \mathbf{q}\mathbf{q}^T\mathbf{q}^* = \lambda^*\mathbf{q}^* - \mathbf{0} = \lambda^*\mathbf{q}^*$$

   The middle step follows from the fact that the eigenvectors of a symmetric matrix are orthogonal.

3. The characteristic polynomials $P(x) = \det(\mathbf{A} - x\mathbf{I})$ and $P_D(x) = \det(\mathbf{A} - \lambda \mathbf{q}\mathbf{q}^T - x\mathbf{I})$ of $\mathbf{A}$ and $\mathbf{D}_{(\lambda,\mathbf{q})}$ respectively are identical, except that the the multiplicity of λ as a root of P has been *decreased* by one in P_D and the multiplicity of the root 0 has been *increased* by one.

To prove this, we first recall that similar matrices have the same characteristic polynomial. Next, we consider the Householder transformation $\mathbf{H}$ that reflects $\mathbf{q}$ into $\mathbf{e}_1$, i.e. $\mathbf{Hq} = \mathbf{e}_1$. Since $\mathbf{H}$ is an orthogonal matrix by construction, it follows that $\mathbf{HAH}^T = \mathbf{HAH}$ is similar to $\mathbf{A}$. Additionally, $\mathbf{HAH}$ has a very specific structure. In particular, its first column will have the following form:

$$\mathbf{HAHe}_1 = \mathbf{HAHHq} = \mathbf{HAq} = \lambda\mathbf{Hq} = \lambda\mathbf{e}_1$$

By symmetry, the first row of $\mathbf{HAH}$ will be $\lambda\mathbf{e}_1^T$. More generally, we have that:

$$\mathbf{HAH} = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{M} & \\ 0 & & & \end{pmatrix}$$

where $\mathbf{M}$ is a symmetric $(n-1) \times (n-1)$ matrix. It is easy to see that $\mathbf{H}(\lambda\mathbf{qq}^T)\mathbf{H} = \lambda\mathbf{e}_1\mathbf{e}_1^T$, which implies that:

$$\mathbf{H}(\mathbf{A} - \lambda\mathbf{qq}^T)\mathbf{H} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \mathbf{M} & \\ 0 & & & \end{pmatrix}$$

Since similarity transforms do not affect the characteristic polynomial of a matrix, we observe that the deflation changed exactly one root of P from λ to 0. All other roots are unaffected and given by the characteristic polynomial of $\mathbf{M}$.

Traces and Determinants

The characteristic polynomial of an $n \times n$ matrix $\mathbf{A}$ is defined as $P(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$, and the roots of this polynomial are the eigenvalues of $\mathbf{A}$. To prove this, note that if $P(\lambda_i) = \det(\mathbf{A} - \lambda_i\mathbf{I}) = 0$, then the matrix $\mathbf{A} - \lambda_i\mathbf{I}$ is singular, and thus the equation $(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{x} = \mathbf{0}$ has a nontrivial solution. Rearranging terms, we see that $\mathbf{Ax} = \lambda_i\mathbf{x}$, and thus λ_i is an eigenvalue of $\mathbf{A}$. Repeated application of the **Fundamental Theorem of Algebra** allows us to rewrite $P(\lambda)$ as follows:

$$\begin{aligned} P(\lambda) &= (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_1)\cdots(\lambda - \lambda_n) \\ &= (-1)^n\left(\lambda^n - \lambda^{n-1}\sum_i \lambda_i + \lambda^{n-2}\sum_{i \neq j}\lambda_i\lambda_j - \cdots + (-1)^n\lambda_1\lambda_2\cdots\lambda_n\right) \\ &= (-1)^n\lambda^n + (-1)^{n-1}\lambda^{n-1}\sum_i \lambda_i + (-1)^{n-2}\lambda^{n-2}\sum_{i \neq j}\lambda_i\lambda_j + \cdots + \lambda_1\lambda_2\cdots\lambda_n \end{aligned}$$

where each λ_i is an eigenvalue of $\mathbf{A}$. This formulation tells us that we can derive some useful information directly from the characteristic polynomial without needing to solve for its roots. Namely, we see that the sum of all eigenvalues is equal to the coefficient of λ^{n-1} in the characteristic polynomial modulo the sign term $(-1)^{n-1}$. Similarly, the product of the eigenvalues is equal to the polynomial's constant term. These values have specific meanings for symmetric $n \times n$ matrices: they are the **trace** and the **determinant** of the matrix, respectively.

Frobenius Norm

An alternative, but equivalent, definition for the Frobenius norm is:

$$\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}(\mathbf{A}^T\mathbf{A})} = \sqrt{\mathrm{tr}(\mathbf{A}\mathbf{A}^T)}$$

To prove this equality, note that:

$$(\mathbf{A}^T\mathbf{A})_{ij} = \sum_{k=1}^{m} (\mathbf{A}^T)_{ik}\mathbf{A}_{kj} = \sum_{k=1}^{m} \mathbf{A}_{ki}\mathbf{A}_{kj}$$

Therefore:

$$\mathrm{tr}(\mathbf{A}^T\mathbf{A}) = \sum_{i=1}^{n} (\mathbf{A}^T\mathbf{A})_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{m} \mathbf{A}_{ki}\mathbf{A}_{ki} = \sum_{i=1}^{n}\sum_{k=1}^{m} \mathbf{A}_{ki}^2 = \|\mathbf{A}\|_F^2$$

The proof for $\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}(\mathbf{A}\mathbf{A}^T)}$ follows similarly.