DS-GA.1013 Mathematical Tools for Data Science
Homework 1
Yves Greatti - yg390

1. (Rotation) For a symmetric matrix A, can there be a nonzero vector x such that Ax is nonzero and orthogonal to x? Either prove that this is impossible, or explain under what condition on the eigenvalues of A such a vector exists. Let $x \in V$, an inner product space , by the spectral theorem there exists an orthonormal basis of V, consisting of eigenvectors of A, let $u_1, \ldots, u_n$ be the eigenvectors of A, and $\lambda_i, \ldots, \lambda_n$ the eigenvalues for each of these eigenvectors. $x \in \text{span}\{u_1, \ldots, u_n\} \Rightarrow x = \sum_{i=1,n} \alpha_i u_i, \alpha_i \neq 0$. $x^T(Ax) = (\sum_{i=1,n} \alpha_i u_i)(\sum_{j=1,n} \alpha_j A u_j) = (\sum_{i=1,n} \alpha_i u_i)(\sum_{j=1,n} \alpha_j \lambda_j u_j) = \sum_{i=1,n} \alpha_i^2 \lambda_i$ since $u_i^T u_j = 0$ for $i \neq j$ and $u_i^T u_i = 1$. x and Ax are nonzero and Ax is orthogonal to x: $x^T(Ax) = 0 \Rightarrow \sum_{i=1,n} \alpha_i^2 \lambda_i = 0$.

2. (Matrix decomposition) The trace can be used to define an inner product between matrices:

$$\langle A, B \rangle := \text{tr}\left(A^T B\right), \quad A, B \in \mathbb{R}^{m \times n}, \tag{1}$$

where the corresponding norm is the Frobenius norm $||A||_{\text{F}} := \langle A, A \rangle$.

   (a) Express the inner product in terms of vectorized matrices and use the result to prove that this is a valid inner product. $(AB)_{ij} = (\sum_k A_{ik} B_{kj})_{ij}$, and $(A^T B)_{ij} = (\sum_k A_{ki} B_{kj})_{ij}$. $\text{tr}(A) = \sum_i A_{ii} \Rightarrow \text{tr}\left(A^T B\right) = \sum_i \sum_k A_{ki} B_{ki} = \sum_i \sum_j A_{ij} B_{ij} = \text{vec}(A)^T \text{vect}(B) = \langle \text{vec}(A), \text{vec}(B) \rangle$. The trace is then the inner product between vectors in $\mathbb{R}^{m \times n}$ thus is a valid inner product.

   (b) Prove that for any $A, B \in \mathbb{R}^{m \times n}$, $\text{tr}\left(A^T B\right) = \text{tr}\left(B A^T\right)$. $\text{tr}\left(B A^T\right) = \sum_i \sum_k B_{ik} A_{ik} = \sum_i \sum_j A_{ij} B_{ij} = \text{tr}\left(A^T B\right)$.

   (c) Let $u_1, \ldots, u_n$ be the eigenvectors of a symmetric matrix A. Compute the inner product between the rank-1 matrices $u_i u_i^T$ and $u_j u_j^T$ for $i \neq j$, and also the norm of $u_i u_i^T$ for $i = 1, \ldots, n$. For $i \neq j$, $\langle u_i u_i^T, u_j u_j^T \rangle = \text{tr}\left(u_i u_i^T u_j u_j^T\right) = 0$ since $u_i \perp u_j$, u_i, u_j being two eigenvectors for different eigenvalues of the symmetric matrix A. if $i = j$ then $\langle u_i u_i^T, u_i u_i^T \rangle = \text{tr}\left(u_i u_i^T u_i u_i^T\right) = \text{tr}\left((u_i^T u_i)^2\right) = (u_i^T u_i)^2 \Rightarrow |\left|\left|u_i u_i^T\right|\right|_{\text{F}} = u_i^T u_i$.

   (d) What is the projection of A onto $u_i u_i^T$? The projection of A onto $u_i u_i^T$ is $\langle A, u_i u_i^T \rangle$, A is a symmetric matrix, by the spectral theorem, $A = UDU^T$ where $D = \textbf{diag}(\lambda)$. $\langle A, UU^T \rangle = \text{tr}\left(UDU^T UU^T\right) = \text{tr}\left(UU^T(UDU^T)\right) = \text{tr}\left(UU^T UU^T D\right) = \text{tr}\left((U^T U)^2 D\right)$ thus $\langle A, u_i u_i^T \rangle = \lambda_i (u_i^T u_i)$.

   (e) Provide a geometric interpretation of the matrix $A' := A - \lambda_1 u_1 u_1^T$, which we defined in the proof of the spectral theorem, based on your previous answers. From the previous question the orthogonal projection of A in $u_i u_i^T$ is $\lambda_i u_i u_i^T$ so A' has row or column subspaces contained in $(u_1)^\perp$.

3. (Quadratic forms) Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix, and let $f(x) := x^T A x$ be the corresponding quadratic form. We consider the 1D function $g_v(t) = f(tv)$ obtained by restricting the quadratic form to lie in the direction of a vector v with unit ℓ_2 norm.

(a) Is $g_v(t)$ a polynomial? If so, what kind?

(b) What is the curvature (i.e. the second derivative) of $g_v(t) = f(tv)$ at an arbitrary point t?

(c) What are the directions of maximum and minimum curvature of the quadratic form? What are the corresponding curvatures equal to?

4. (Projected gradient ascent) Projected gradient descent is a method designed to find the maximum of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ in a constraint set $\mathcal{S}$. Let $\mathcal{P}_\mathcal{S}$ denote the projection onto $\mathcal{S}$, i.e.

$$\mathcal{P}_\mathcal{S}(x) := \arg\min_{y \in \mathcal{S}} ||x - y||_2^2 . \tag{2}$$

The kth update of projected gradient ascent equals

$$x^{[k]} := \mathcal{P}_\mathcal{S}(x^{[k-1]} + \alpha \nabla f(x^{[k-1]})), \qquad k = 1, 2, \ldots, \tag{3}$$

where α is a positive constant and $x^{[0]}$ is an arbitrary initial point.

(a) Use the same arguments we used to prove Lemmas 5.1 and 5.2 in the notes on PCA to derive the projection of a vector x onto the unit sphere in n dimensions.

(b) Derive an algorithm based on projected gradient ascent to find the maximum eigenvalue of a symmetric matrix $A \in \mathbb{R}^{n \times n}$.

(c) Let us express the iterations in the basis of eigenvectors of A: $x^{[k]} := \sum_{i=1}^n \beta_i^{[k]} u_i$. Compute the ratio between the coefficient corresponding to the largest eigenvalue and the rest $\frac{\beta_1^{[k]}}{\beta_i^{[k]}}$ as a function of k, α, and $\beta_1^{[0]}$, …, $\beta_n^{[0]}$. Under what conditions on α and the initial point does the algorithm converge to the eigenvector u_1 corresponding to the largest eigenvalue? What happens if α is extremely large (i.e. when $\alpha \to \infty$)?

(d) Implement the algorithm derived in part (b). Support code is provided in `main.py` within `Q4.zip`. Observe what happens for different sizes of α. Report the plots generated by the script.