Homework 1: Hints for solutions

1. Let $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and define $\phi\colon \mathbb{R}^n \to \mathbb{R}^m$ by $\phi(x) = Ax - b$.

   Prove that for every $i \in \{1, 2, \dots, n\}$ one has that

   $$\sum_{j=1}^{m} \frac{\partial \phi_j}{\partial x_i}(x)\phi_j(x) = a_i^T(Ax - b).$$

   Idea: Notice that this is part of the computation used to complete the proof of Lemma 1.5, which says that the orthogonal projection minimizes the distance from b to $\mathrm{Im}(A)$. By the definition of the dot product and $\phi(x) = Ax - b$ that

   $$a_i^T(Ax - b) = a_i^T(\phi(x)) = \sum_{j=1}^{m}(a_i)_j(\phi_j(x)).$$

   The proof is concluded once you write the coordinates of A and verify $\frac{\partial \phi_j}{\partial x_i} = (a_i)_j$.

2. **Hand In** Let $A \in \mathbb{R}^{m\times n}$, where $m \le n$, and assume that $r(A) = m$.

   Prove that $A^\dagger = A^T(AA^T)^{-1}$.

   Following proof of Proposition 1.8: Note since $r(A) = m$, A is surjective onto $\mathbb{R}^m$, so $\ker(A^T) = \vec{0}$, and thus $b_0 = b$ when constructing the pseudo-inverse. Thus $A^\dagger$ is the matrix so that $A^\dagger b = x$ for $x \in \mathrm{Im}(A^T)$ and so that $Ax = b$. Since $x \in \mathrm{Im}(A^T)$, there exists some $y \in \mathbb{R}^n$ so that $A^T y = x$. Applying A to each side,

   $$AA^T y = Ax = b.$$

   Since $r(A) = m$, AA^T is invertible, so

   $$y = (AA^T)^{-1}b.$$

   Thus we have

   $$A^T(AA^T)^{-1}b = A^T y = x.$$

   So we've shown that $A^T(AA^T)^{-1}$ also satisfies the properties uniquely defining $A^\dagger$, which implies $A^\dagger = A^T(AA^T)^{-1}$. □

3. Let $A \in \mathbb{R}^{m\times n}$.

   Prove that, if $A = U\Sigma V^T$ is a singular value decomposition for A, then $A^\dagger = V\Sigma^{-1}U^T$.

   Idea: Recall the columns of V and U form and orthonormal basis for $\mathrm{Im}(A^T)$ and $\mathrm{Im}(A)$, respectively. We will use the fact that in this case, UU^T is the orthogonal projection onto $\mathrm{Im}(A)$ and VV^T is the orthogonal projection onto $\mathrm{Im}(A^T)$.

   Given this information, take $b \in \mathbb{R}^m$ and $x \in \mathrm{Im}(A^T)$. Then since b_0 is the orthogonal projection of b onto $\mathrm{Im}(A)$, we can write

   $$b_0 = UU^T b.$$

   We want

   $$U\Sigma V^T x = Ax = b_0 = UU^T b.$$

   Applying U^T to each side, we have $U^T U = 1_r$ so

   $$\Sigma V^T x = U^T b.$$

   Since Σ is a diagonal matrix with positive entries, we can take Σ^{-1} to get

   $$V^T x = \Sigma^{-1}U^T b.$$

   Finally applying V to each side, since VV^T is the orthogonal projection onto $\mathrm{Im}(A^T)$, it preserves $x \in \mathrm{Im}(A^T)$. Thus

   $$x = VV^T x = V\Sigma^{-1}U^T b.$$

   This is exactly the property of $A^\dagger$, so we conclude $A^\dagger = V\Sigma^{-1}U^T$. □

4. (a) Compute by hand a singular value decomposition and the pseudoinverse of $A = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \\ -2 & 1 \end{bmatrix}$.

(b) Now try to do the same using Julia. Do you get what you expected? What happens if you compare the pseudoinverse obtained via the command `pinv` to the one obtained by taking $V\Sigma^{-1}U^T$? Produce a jupyter notebook documenting your work, including your comments on the behavior above.

See Juila Exercises– 01_Ex4b.ipynb. This is an exercise in seeing how to compute the SVD and pseudoinverse by hand which works fine, but seems to create a problem when doing this with a computer. This shows how inverting singular values which are close to zero can create numerical issues when using computer programs to solve for a pseudoinverse.

5. Let $X \sim N(\mu, \sigma^2)$ for $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Prove that $\mathbb{E}X = \mu$ and $\mathrm{Var}(X) = \sigma^2$.

   Hint: it might be useful to recall that $\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}$.

   Idea: For the expected value, I used the substitution $u = \frac{x-\mu}{\sigma}$, so $dx = \sigma\,du$ and $x = \sigma u + \mu$ to get

$$\mathbb{E}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} x e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u e^{-\frac{u^2}{2}}\,du + \mu \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}}\,du.$$

   The left of the sum is zero, and the right of the sum uses the hint to get the expected value of μ.

   For the variance,

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}(X^2) - \mu^2.$$

   So it suffices to show $\mathbb{E}(X^2) = \sigma^2 + \mu^2$. Recall from Lemma 1.27 with $g(x) = x^2$ that we have

$$\mathbb{E}(X^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx.$$

   Using the same substitution as above with the same evaluations

$$\mathbb{E}(X^2) = \mu^2 + \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} u^2 e^{-u^2/2}\,du.$$

   Then do integration by parts for the two parts u and $ue^{-u^2/2}$ to finish the computation.

6. Let X and Y be two real random variables that are either:

   both discrete; both continuous, have respective densities f_X, f_Y and finite expected values, i.e., $\mathbb{E}(X), \mathbb{E}(Y) < \infty$.

   Prove that for all $a, b \in \mathbb{R}$ one has that $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.

   Hint: use the transformation law (Lemma 1.27 in the notes) for $g(X, Y) = X + Y$ and that for every random variable $\mathbb{E}|X| < \infty$ if and only if $\mathbb{E}X < \infty$ (see Eq. (3.1.7) in Ash's book). The concept of marginal density might also be useful.

   Idea What I want you to understand out of this is the case when $P(x, y) = f_X(x)f_Y(y)$, so the variables are independent. Then using linearity of the integral and Lemma 1.27,

$$\mathbb{E}(aX + bY) = \iint (ax + by)P(x, y)\,dx\,dy. = a \iint x f_X(x) f_Y(y)\,dx\,dy + b \iint f_X(x) y f_Y(y)\,dx\,dy = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

7. Let $\Omega := \{x_1, \dots, x_n\}$ and $p_1, \dots, p_n \geq 0$ with $p_1 + \cdots + p_n = 1$. Prove that the following algorithm generates a random variable $X \in \Omega$ with $P(X = x_i) = p_i$:

   define the numbers $w_k := \sum_{i=1}^{k} p_i$, $1 \leq k \leq n$, and $w_0 := 0$; draw $Y \sim \mathrm{Unif}([0, 1])$ (for instance, in `Julia` one can draw Y using the command `rand()`); let k be such that $w_{k-1} \leq Y < w_k$; return x_k.

   Idea The number $w_0, \dots, w_n$ form an increasing sequence of numbers between 0 and 1. Then $P(X = x_i)$ is given from the uniform distribution over the interval $[w_{i-1}, w_i]$, which is $P(X = x_i) = w_i - w_{i-1} = p_i$.

8. The element caesium-137 has a half-life of about 30,17 years. In other words, a single atom of caesium-137 has a 50 percent chance of surviving after 30,17 years, a 25 percent chance of surviving after 60,34 years, and so on.

   (a) Determine the probability that a single atom of caesium-137 decays (i.e., does not survive) after a single day. How would you model the random variable X that takes the value 1 when the atom decays and 0 otherwise?

(b) Using `Julia`, simulate 1000 times the behaviour of a collection C of 10^6 caesium-137 atoms in a single day. How would you model the random variable $Y = |\{\text{atoms in } C \text{ decaying after a single day}\}|$?

(c) The Poisson distribution with parameter λ is a discrete probability distribution that is used to "model rare events". When $Z \sim Pois(\lambda)$, one has that $P(Z = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$. Plot the Poisson distribution with $\lambda = 10^6 \cdot p$, where p is the probability computed in part (a).

(d) Compare the empirical distribution in part (b) to the theoretical distribution in part (c).

Some `Julia` packages that might be useful: `Distributions`, `StatsPlots`.

See Juila Exercises– 02_Ex4b.ipynb