CENG371

Scientific Computing

Fall 2022-2023

Homework 3

Due: December 30th, 2023, 23:55

Question 1 (20 points)

Denote by x the solution satisfying $Ax = b$, where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^n$ and $m > n$.

a) **(5 pts)** Let $B = \{b' \mid Ax = b + b'\}$. What is the dimensionality of B?

b) **(15 pts)** Let $A \in \mathbb{R}^{5\times 3}$ be given as $A(i,j) = \sqrt{i^2 + j^2}$, and $b \in \mathbb{R}^3$ be given as $b(i) = i$ where $1 \leq i \leq 5, 1 \leq j \leq 3$. Compute a basis of B for this problem using the algorithms you have seen in this course thus far. In your reports, refer to the name of the file where you do the computation and briefly explain your solution.

Question 2 (30 points)

Compute the singular value decomposition (you can use built-in algorithms, such as MATLAB's `svd`) of a grayscale image of your choice, $I \in \mathbb{R}^{m\times n}$ where $r = \min(m, n) > 100$.

a) **(10 pts)** Compute low-rank approximations, I_k, $k \in \{1, 2, \ldots, r\}$ of this image by discarding the smallest k singular values. In your reports, include I, $I_{r/2}$, and I_{r-10}, and make a comparison between them (2-3 sentences).

b) **(10 pts)** Denote the singular values by $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_r \mid \sigma_i < \sigma_{i+1}\}$, and let $S(k) = \sqrt{\sum_{i=1}^{k} \sigma_i^2}$. Plot $S(k)$ and the errors $\|I - I_k\|_F$ where $\|.\|_F$ is the Frobenius norm in a single log-plot (you can use `loglog` in MATLAB). Include the plots in your reports and reflect on your observations.

c) **(10 pts)** Based on your observations from part a) and part b), suggest a use case for the low-rank approximation scheme.

Question 3 (50 points)

Let $N(t) = 0.3 + 2t - 1.2t^2 + 0.5t^3 = \sum_{k=0}^{3} c_k t^k$ denote the quantity of an observable N as a function of time t. Let $N_{obs}(t)$ denote the observed **noisy** values of this system, given by

$$N_{obs}(t) = \sum_{k=0} c_k (t + \varepsilon)^k, \quad \varepsilon_k \sim \mathcal{N}(0, 0.01)$$

where $\mathcal{N}(\mu, \sigma)$ is the Gaussian distribution with mean μ and standard deviation σ, and $\varepsilon_k \sim \mathcal{P}$ means ε is sampled from the probability distribution $\mathcal{P}$. In MATLAB, you can simply use `randn*0.01` to sample the random variables. We can generate a set of observations as

$$O_n = \left\{ (t, N_{obs}(t)) \,\middle|\, t \in \left\{ \frac{1}{n+1}, \frac{2}{n+1}, \ldots, \frac{n-1}{n+1}, \frac{n}{n+1} \right\} \right\}.$$

For this problem, we want to determine the constants c_k from a set of observations.

a) **(15 pts)** Generate sets observations of this system for $n = \{5, 10, 100\}$. For each set of observations, construct the corresponding A_n and b_n.

b) **(25 pts)** Using any of the methods we have seen in this course, acquire solutions for c_k. In your reports, refer to the name of the related code file and explain your solution strategy.

c) **(10 pts)** How does the value n affect the quality of the approximations for c_k? Explain by discussing the average errors over multiple runs for each.

Regulations

1. Make sure that you reflect **your own reasoning** in a clean and concise manner.
2. Your submission should include a single PDF and your `.m` files.
3. Submission will be done via odtuclass.
4. **Late Submission:** Accepted with a penalty of $-5 \times (day)^2$.