1. Consider the **collaborative filtering** problem: given a set of $m$ items and $n$ users, and observations of a subset of the scores $\{a_{ij}\}$ that the $j$th user assigns to the $i$th item, predict the remaining unknown scores.

A popular method for solving this problem is to assume that the $m \times n$ score matrix $A$ is of low-rank. Recall that we say $A$ has rank $r$ if it can be written in the form $A = XY$ for some $X \in \mathbb{R}^{m \times r}$ and $Y \in \mathbb{R}^{r \times n}$, and such a factorization doesn’t exist for any number smaller than $r$. If $A$ has rank $r$, then it only has $O(r(n + m))$ degrees of freedom — the number of entries it takes to specify $X$ and $Y$ — rather than the potentially much larger $mn$ degrees of freedom it takes to specify all the entries of $A$. Thus, if we assume that $A$ has a rank $r$ that is much smaller than $\min\{m, n\}$, then we can expect that we can completely determine $A$ once we have seen a small subset of its entries.

One can show that, if $A$ indeed has rank $r$, and it is not too “spiky”, then with high probability, once we have seen $O(rn \log(n))$ observations selected i.i.d. uniformly at random from $A$, we can recover all of $A$ exactly by solving a (suprisingly tractable!) optimization problem $\mathbf{1}$.

In practical applications (say to predicting item ratings on Amazon), $A$ is not exactly low-rank, but this method still works very well empirically; intuitively, this is because $A$ is well-approximated by a low-rank matrix.

Verify the intuition that underlies the good performance of these algorithms: show that, given any matrix $A \in \mathbb{R}^{m \times n}$, there is a matrix $B \in \mathbb{R}^{m \times n}$ that has rank $O(\varepsilon^{-2} \log(m + n))$ and is entrywise close to $A$. Specifically, show that $B$ satisfies

$$\|A - B\|_\infty \leq \varepsilon \max_{i=1, \ldots, n} \|a_i\|_2.$$  

Here $a_i$ is the $i$th column of $A$ and $\|A\|_\infty = \max_{i,j} |a_{ij}|$ is the largest absolute entry.

Hint: use the fact that $A = I_m \times A$.

2. Gas molecules move about randomly in a box that is divided into two halves symmetrically by a partition; there is a hole in the partition. Suppose there are $n$ molecules in the box. Molecular motion can be modeled by keeping the molecules in their positions with probability $1/2$, and with probability $1/2$ choosing a number between $1$ and $n$ at random and moving the corresponding molecule to the other side of the partition.

(a) Argue that the number of molecules on one side of the partition evolves as a Markov chain. What are the states, and the transition probabilities?

(b) Show that the chain is ergodic, and find its stationary distribution. Hint: you can assume that detailed balance is satisfied. If you use this hint, the identities

$$\prod_{i=0}^{k-1} \binom{n-i}{i+1} = \binom{n}{k} \quad \text{and} \quad \sum_{i=0}^{n} \binom{n}{i} = 2^n$$

may be helpful.

---

1This observation and the randomized algorithms that arise from it have driven a lot of applied and theoretical work in applied mathematics, statistics, and machine learning over the past decade.
3. Consider a random vector $\mathbf{V} = (x, y)$ that is distributed according to the pdf

$$p(x, y) = c \exp(-(x^2 y^2 + x^2 + y^2 - 8x - 8y)/2),$$

where $c \approx 1/20216.335877$ is the constant required to make this a pdf.

(a) Show that $x|y$ and $y|x$ are normal random variables, and determine their mean and variances.

(b) Implement a Gibbs sampler to draw samples from this distribution. Use this sampler to draw $10^4$ samples, and plot the last 1000 along with several contour lines of the density $p$.

(c) Attach your code\(^2\) for both of the MCMC samplers above to your homework submission.

\(^2\)Make your code coherent and easy to follow. I don’t care what language you use