For all problem set throughout the semester you must do the following:

1. State your strategy for arriving at the solution.
2. Execute your strategy noting the steps.
3. ✎ Write legibly and in a logical order.

The purpose of this homework is to analyze theoretically and observe experimentally (using MATLAB) the concentration of

1. the Gaussian measure in $\mathbb{R}^n$,
2. the uniform measure on the unit sphere,
3. the uniform measure in the unit ball,
4. eigenvalues of various large matrices.

1 Facts: the volume of a ball and the surface area of a sphere

We recall that the volume of a ball of radius r for the l^2 norm in $\mathbb{R}^n$ is given by,

$$\operatorname{vol} B^n(r) = \frac{(\sqrt{\pi}\ r)^n}{\Gamma(n/2+1)} \tag{1}$$

and the surface area of the sphere is

$$\sigma_n(r) = \frac{n}{r} \operatorname{vol} B^n(r) = \frac{2(\sqrt{\pi})^n\ r^{n-1}}{\Gamma(n/2)} \tag{2}$$

2 Sampling the unit ball

We consider the following problem: how can we generate $N = 10,000$ samples of points drawn uniformly inside the unit ball in $\mathbb{R}^{400}$. This problem is equivalent to constructing 10,000 vectors with $n = 400$ co-ordinates and a norm less than 1.

Let $B^n(r)$ be the ball of radius r centered at the origin in $\mathbb{R}^n$. Because we consider the uniform measure inside the unit ball $B^n(1)$ in $\mathbb{R}^n$, each set A has a measure given by

$$\mu_n(A) = \frac{\operatorname{vol}(A)}{\operatorname{vol}(B^n(1))} \tag{3}$$

In order to generate samples according to the measure μ_n, we can use a standard rejection method. The principle of the rejection method is to sample from a uniform distribution on $[-1,1]^n$, and keep only the points that fall inside the ball. The algorithm is described in Fig. 1. The matlab function `rand` can be used to generate samples from a uniform distribution.

Assignment

1. Generate 10,000 sample uniformly in $[-1,1]^n$ with algorithm 1, for $n = 1, \cdots, 400$.
2. Plot the number of points rejected in algorithm 1 as a function of n. These points are inside the cube $[-1,1]^n$, but not inside the ball $B^n(1)$. Comment.

Algorithm 1

Input:

- N: number of samples
- n: dimension of the space

Algorithm:

for i=1 to N do

* generate n samples $u_1, \cdots, u_n$ according to the uniform distribution on $[0,1]^n$
* define $x_1 = r(1 - 2 * u_1), \cdots, x_n = r(1 - 2 * u_n)$

// $x_q, \cdots, x_n$ are uniformly distributed in $[-r, r]^n$

* if ($\|x\| \leqslant r$) then

    return x

  endif

end do

Figure 1: Generation of samples with uniform measure in the ball $B^n(r)$ with the rejection method

3 Sampling the unit sphere

The rejection method is not efficient for large n. In section 4 we will revisit the problem of generating uniform samples in the unit ball using uniform sample on the unit sphere, and drilling inside the ball. We therefore need to understand how to generate uniform samples on the unit sphere.

Before we learn how to generate samples on the unit sphere, we study the concentration of the Gaussian measure in high dimension. Our understanding of this concentration of measure will help us devise an efficient algorithm to sample the unit sphere.

3.1 Experiments

Assignment

3. Generate 10,000 samples x in $\mathbb{R}^n$ from a Gaussian distribution with mean 0 and unit variance for n being a squared integer less than 400. You can use the function `randn` in matlab.
4. Plot the histogram of the norm $\|x\|$ of the points x for $n = 4, 25, 100, 225, 400$.
5. Compute the mean and the variance of each distribution, and plot it as a function of n. What do you notice? Make a conjecture on the concentration of $\|x\|$, when x is Gaussian distributed in $\mathbb{R}^n$.

3.2 Theory: the concentration of the Gaussian measure

Let γ be the Gaussian measure. For any measurable set $A \subset \mathbb{R}^n$ γ is defined by the integral over A,

$$\gamma(A) = \frac{1}{(2\pi)^{n/2}} \int_A e^{-\|x\|^2/2} dx \tag{4}$$

Note that the integral is really computed with the density:

$$d\gamma(x) = d\gamma(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} \, dx_1 \cdots dx_n \tag{5}$$

Let $f(x)$ be a function defined on $\mathbb{R}^n$ taking values in $\mathbb{R}$,

$$f : \mathbb{R}^n \to \mathbb{R}.$$

Assignment

6. . Let $\lambda > 0$. Prove that for all $a \in \mathbb{R}$,

$$\gamma\{x \in \mathbb{R}^n : f(x) \leqslant a\} \leqslant e^{\lambda a} \int_{\mathbb{R}^n} e^{-\lambda f(x)} d\gamma(x) \tag{6}$$

   Hint: observe that

$$e^{-\lambda a}\, \gamma\{x \in \mathbb{R}^n : f(x) \leqslant a\} = \int_{\{x: f(x) \leqslant a\}} e^{-\lambda a}\, d\gamma(x). \tag{7}$$

7. Let $0 < \lambda < 1$, and $0 < \delta \leqslant n$. Apply the inequality (6) to the function $f(x) = \|x\|^2/2$, and use $a = (n-\delta)/2$, to show that

$$\gamma\left\{x \in \mathbb{R}^n : \|x\|^2 \leqslant n-\delta\right\} \leqslant e^{\lambda(n-\delta)/2} \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-(\lambda+1)\|x\|^2/2} dx \tag{8}$$

8. The change of variable $y = x\sqrt{1+\lambda}$ allows us to compute the integral above, and we have

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-(\lambda+1)\|x\|^2/2} dx = \frac{1}{(1+\lambda)^{n/2}} \tag{9}$$

   Admit Eq. (9) , and show that if we take $\lambda = \delta/(n-\delta)$ we obtain

$$\gamma\left\{x \in \mathbb{R}^n : \|x\|^2 \leqslant n-\delta\right\} \leqslant \left(\frac{n-\delta}{n}\right)^{n/2} e^{\delta/2}. \tag{10}$$

9. Let $\varepsilon = \delta/n$. Show that

$$\gamma\left\{x \in \mathbb{R}^n : \|x\|^2 \leqslant n(1-\varepsilon)\right\} \leqslant e^{-n\varepsilon^2/4} \tag{11}$$

   Hint: you can use the fact that $ln(1-x) + x \leqslant -x^2/2$.

10. Using the same approach as above, we can show that

$$\gamma\left\{x \in \mathbb{R}^n : \|x\|^2 \geqslant \frac{n}{1-\varepsilon}\right\} \leqslant e^{-n\varepsilon^2/4}.$$

   We can also show (with a slightly worse constant) that

$$\gamma\left\{x \in \mathbb{R}^n : \|x\|^2 \geqslant n(1+\varepsilon)\right\} \leqslant e^{-n\varepsilon^2/8}. \tag{12}$$

   Combine (11) and (12) to show that for every $\varepsilon > 0$, and for every $n \geqslant 0$,

$$\gamma\left\{x \in \mathbb{R}^n : \left|\frac{\|x\|^2}{n} - 1\right| \geqslant \varepsilon\right\} \leqslant 2e^{-n\varepsilon^2/8}. \tag{13}$$

11. Explain why this result shows that the Gaussian measure is concentrated on the sphere of radius $\sqrt{n}$ with a decay of $e^{-\varepsilon^2/8}$.

3.3 An efficient algorithm to generate uniform samples on the sphere $S^{n-1}(r)$

Algorithm 2

Input:

- N: number of samples
- n: dimension of the space

Algorithm:

for i=1 to N do

* generate a sample x according to the n-dimensional standard normal distribution
* $x \leftarrow \frac{x}{\|x\|}$
* return x

end do

Figure 2: Generation of N samples with uniform measure on the sphere $S^{n-1}(r)$

The algorithm in Fig. 2 generates samples that are uniformly distributed on the sphere of radius r in $\mathbb{R}^n$. We will use this algorithm to study the concentration of measure on the sphere.

Assignment

12. Generate $N = 10,000$ points on the sphere $S^{n-1}(1)$ for $n = 1, \cdots, 400$. Project the N points on the axis defined by x_1. You should get N real numbers. Plot the histogram formed by the projection, for $n = 4, 25, 100, 225, 400$.

13. Compute the mean and the variance of the distribution of the projections, and plot these values as a function of n.

14. Find $\varepsilon(n)$ such that $0.99N$ points are in the slab

$$\{x = (x_1, \cdots, x_n) \in S^{n-1}(1) : -\varepsilon(n) \leqslant x_1 \leqslant \varepsilon(n)\}.$$

Plot $\varepsilon(n)$ as function of n. What curve would you expect to find?

15. Do you think that the choice of axis (x_1 versus any other axis) is important for the results in the previous question? Justify your answer.

4 Uniform sampling in the ball

The generation of samples distributed uniformly in the ball $B^n(r)$ can be achieved by sampling uniformly a sequence of hyperspheres centered around the origin with uniformly distributed radii. This approach can be implemented with the algorithm described in Fig. 3.

Algorithm 3

Input:

- N: number of samples
- n: dimension of the space

Algorithm:

for i=1 to N do

* generate a sample x according to the n-dimensional standard normal distribution
* generate a sample u according to the uniform distribution
* $x \leftarrow r\ u^{1/n}\ \frac{x}{\|x\|}$
* return x

end do

Figure 3: Generation of N samples with uniform measure inside the sphere $B^n(r)$

Equipped with an efficient algorithm to sample uniformly inside the ball $B^n(r)$ we can study the concentration of the uniform measure in the ball.

Assignment

16. Generate $N = 10,000$ points inside the ball $B^n(\sqrt{n})$ for $n = 1, \cdots, 400$.

17. Project the points on the axis defined by the first coordinate x_1. Plot the histograms for $n = 4, 25, 100, 225, 400$.

18. Compute the mean and the variance of the distribution of the projections, and plot these values as a function of n. What do you observe?

19. Find the relative volume $w(n)$ (measured as the fraction of number of points) in the slab of thickness $1/2$ in the ball $B^n(\sqrt{n})$

$$\{x = (x_1, \cdots, x_n) \in B^n(\sqrt{n}) : -\frac{1}{2} \leqslant x_1 \leqslant \frac{1}{2}\}.$$

    Plot $w(n)$ as function of n. What curve would you expect to find? Comment.

20. Do you think that the choice of axis (x_1 versus any other axis) is important for the results in the previous question. Justify your answer.

21. Compute the histogram of the distance to the origin for all the points in $B^n(\sqrt{n})$ for $n = 1, \cdots, 400$. What do you observe? Explain the apparent paradox in terms of a concentration of measure.

5 Empirical distribution of eigenvalues and singular values

In this third section, you will experiment with another consequence of the concentration of measure: the convergence of the distribution of the eigenvalues of random matrices toward universal distributions. These distributions are independent of the specific ensembles of random matrices from which you compute the eigenvalues: they are universal. We start with a brief review of linear algebra.

5.1 Bluffer's guide to Linear Algebra

The following facts are presented here without proofs. The interested reader should consult a book on matrix analysis (e.g. [Stewart, 1998, Stewart, 2001]).

We first review the concept of eigenvalues and singular values.

Let A be an $n \times n$ matrix over the set of reals. The eigenvalues of A are of the n roots of its characteristic polynomial

$$P_A(z) = \det(A - zI). \tag{14}$$

We note that some of the roots may be in $\mathbb{C}$. We index the eigenvalues in decreasing order of magnitude,

$$\lambda_1(A), \ldots, \lambda_n(A) \qquad \text{with} \quad |\lambda_1(A)| \geqslant \ldots \geqslant |\lambda_n(A)| \tag{15}$$

The spectral radius is given by $|\lambda_1|$. If A is symmetric, then all its eigenvalues are real. If A is semi-definite positive, then all its eigenvalues are non negative.

We are also interested in a second type of spectrum formed by the "singular values". The singular values of A are defined by the singular value decomposition of A.

Theorem 1. *Let X be an $n \times p$, $n \geqslant p$ matrix, then there exists an $p \times n$ orthogonal matrix U, an $p \times p$ orthogonal matrix V, and a $p \times p$ diagonal matrix $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_p)$, such that*

$$X = U^T \Sigma V \tag{16}$$

In addition, the following properties hold

1. the ordering $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_p$ is canonical, and is expected throughout.

2. The columns of $U = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}$ and $V = \begin{bmatrix} u_1 & \cdots & u_p \end{bmatrix}$ are called the left and right singular vectors. They satisfy

$$X v_i = \sigma_i u_i, \quad X^T u_i = \sigma_i v_i, \quad i = 1, \ldots, p \tag{17}$$

The singular values are the eigenvalues of the matrix $\sqrt{X^T X}$,

$$\sigma_k(X) = \lambda_k(\sqrt{X^T X}), k = 1, \ldots, p. \tag{18}$$

3. Geometrically, the matrix $X^T X$ maps the unit sphere to an ellipsoid, the half-lengths of its principal axes being exactly the singular values of X.

4. the MATLAB command `[U,S,V] = svd(X)` compute the singular value decomposition (SVD) of a matrix `X`.

5.2 Some classical random matrix ensembles

We list here some simple models of matrices with random entries. Such models are encountered frequently in physics, graph theory, probability, etc. We restrict our description to square matrices.

5.3 IID matrix ensembles

Such matrices are composed of random entries drawn independently from the same probability distribution. The distribution is often normalized to have zero mean and variance one. Examples include:

- random sign matrices (Bernoulli ensemble), where a matrix M has independent entries $m_{i,j} \in \{-1, 1\}$ with probability p and $1 - p$ respectively.
- Gaussian matrix ensemble, where a matrix M has independent entries $m_{i,j}$ drawn from a centered unit variance Gaussian distribution, $m_{i,j} \sim N(0, 1)$.

5.4 Wigner matrix ensembles

Such matrices are always symmetric (and so the lower triangular coefficients are not independent of the upper triangular ones). The upper triangular coefficients and the diagonal coefficients are all independent, but may have different distributions. Examples include:

- symmetric Bernoulli ensemble. The upper triangular and the diagonal entries are all signed Bernoulli variables in $\{-1, 1\}$.
- Gaussian Orthogonal Ensemble (GOE). The upper triangular entries have distribution $N(0, 1)$ and the diagonal entries have distribution $N(0, 2)$.

Assignment

22. Write MATLAB functions to create the following random Wigner matrices:
    - symmetric Bernoulli ensemble with $p = \frac{1}{2}$.
    - Gaussian Orthogonal Ensemble (GOE).
23. For each value of $n = 10, 50, 100, 500, 1000$ generate 100 random realization of the Wigner Gaussian and Bernoulli random matrices, and compute the smallest $\lambda_{\min}$ and the largest $\lambda_{\max}$ eigenvalues. You will build a scatter plot: the different realizations of $\lambda_{\min}$ or $\lambda_{\max}$ on the y-axis are plotted as a function of n on the x-axis.
24. Find the line of best fit to guess the dependency of $\lambda_{\min}$ and $\lambda_{\max}$ on the dimension n.

5.5 Wigner semi-circle law

We now consider the normalized empirical spectral distribution (ESD for short) defined by

$$\mu([a, b]) = \frac{1}{n} \# \left\{ i : \frac{\lambda_i}{\sqrt{n}} \in [a, b] \right\} \tag{19}$$

The ESD can be written in terms of a sum of point masses at the (normalized) eigenvalues,

$$\mu = \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k/\sqrt{n}}. \tag{20}$$

In 1955, Wigner proved that the ESD of any Wigner matrix converges almost surely (and hence also in probability and in expectation) to the semicircular distribution

$$\begin{cases} \frac{1}{2\pi}\sqrt{4-|x|^2} & \text{if } |x| \leqslant 2 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

Assignment

25. For each value of $n = 10, 50, 100, 500, 1000$ generate 100 random realization of the Wigner Gaussian and Bernoulli random matrices, and compute the ESD. Plot the ESD and the semi-circle distribution.

5.6 Circular law

We now consider more general IID matrices, and compute the eigenvalues of such matrices. Since the matrices are no longer symmetric, some (half) of the eigenvalues are complex. The ESD is now defined in the complex plane as

$$\mu(z) = \frac{1}{n} \sum_{k=1}^{n} \delta(z)_{\lambda_k/\sqrt{n}}. \tag{22}$$

Assignment

26. For each value of $n = 10, 50, 100, 500, 1000$ generate 100 random realization of the Gaussian and Bernoulli random matrices, and compute the eigenvalues of the corresponding matrices. Display the normalized eigenvalues $\lambda_k/\sqrt{n}$ in the complex plane using a scatter plot. distribution.

27. Formulate a conjecture about the limit distribution of the spectrum of random square matrices with independent entries, which are non symmetric.

References

[Stewart, 1998] Stewart, G. (1998). *Matrix Algorithms: Volume 1, Basic Decompositions*, volume 1. Cambridge University Press.

[Stewart, 2001] Stewart, G. W. (2001). *Matrix Algorithms: Volume 2, Eigensystems*, volume 2. Society for Industrial and Applied Mathematics.