Homework 7. Due April 10

1. **(5 pts)** Multidimensional scaling (MDS) is often used for visualization of data coming from social sciences. One such example mapping subjective similarities between a list of 12 different countries into 2D is given in here is Section 3.2, example 2, data are in Table 3.2. Read the description of the example and reproduce the plot in Figure 3.3 using the MDS.

   For your convenience, I created a file `SubjSim12Countries.mat` *with the data from Table 3.2 put into a* 12×12 *array* `S` *and symmetrized (reflected with respect to its diagonal). The names of the countries are saved in a* 12×3 *array of characters* `cname`*, in which each country is abbreviated by its first 3 letters. To read the data from it, use the following Matlab commands:*

```
data = load('SubjSim12Countries.mat');
S = data.S; % matrix of subjective similarities;
cname = data.cname; % 12-by-3 array of names of countries,
% each coutry is abbreviated with its first three letters
```

   Note that, in order to create an input for the MDS, you need to convert the matrix S into a matrix of squared distances. It won't be a matrix of squared Euclidean distances. Make up an appropriate transformation based on the description of this example. Mark the countries in your figure.

2. **(5 pts)** In this problem, we consider an example similar to the one with a noisy helix in Figure 6 in [1]. The code `MakeSpiral.m` generates a "noisy helix" dataset. Note that the data are arranged nonuniformly along the helix. To mitigate the effects of nonuniformity of data, Coifman and Lafon proposed to modify the kernel using the density of points in the dataset – see the formulas in the box in the beginning of Section 3.1 in [1]. The version of the diffusion map algorithm described in my lecture notes `data_analysis.pdf` corresponds to $\alpha = 0$ with the output most sensitive to the nonuniformity of the density of data. The version with $\alpha = 1$ reduces the effect of it. Program two versions of the diffusion map algorithm, with $\alpha = 0$ and with $\alpha = 1$, and apply them to the "noisy helix" dataset. Compare their performances. Note that `MakeSpiral.m` also estimates the density of points near each data point (the variable `q`). Experiment with the parameters on the dataset in `MakeSpiral.m`: the number of points in the set `N` (line 5) and the level of noise `sigma` (line 12). Push the diffusion map algorithm to its limits and determine for which combinations of `N` and `sigma` the algorithm no longer to produces a meaningful output.

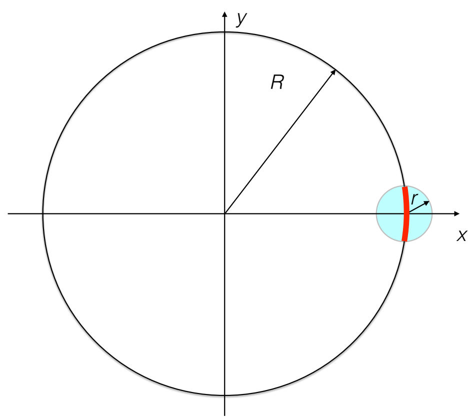
3. **(5 pts)**

(a) Suppose the dataset X, X is an $N \times 2$ matrix, consists of points sampled from the uniform distribution on a circle of radius R centered at the origin, i.e., rows of X are of the form $x_i = [R\cos(\alpha_i), R\sin(\alpha_i)]$, where $\alpha_i \sim U[0, 2\pi)$ are i.i.d. random variables, $1 \le i \le N$. Show that the eigenvalues of the covariance matrix

$$C := \frac{1}{N}(X - \bar{X})^T(X - \bar{X}), \quad \bar{X} := \frac{1}{N}\texttt{ones(N,1)}\left[\sum_{i=1}^{N} x_i(1), \sum_{i=1}^{N} x_i(2)\right]$$

can be approximated by $\lambda_1 \approx \lambda_2 \approx R^2/2$ if N is large.

(b) Suppose the dataset X, X is an $N \times 2$ matrix, consists of points sampled from the uniform distribution in the intersection of the circle of radius R centered at the origin and the ball of radius $r \ll R$ centered at $(R, 0)$, i.e., from the uniform distribution on the arc shown red in the figure below.



Show that the eigenvalues of the covariance matrix $C := \frac{1}{N}(X - \bar{X})^T(X - \bar{X})$ can be approximated by $\lambda_1 \approx \frac{r^2}{3}, \quad \lambda_2 \approx \frac{r^4}{45R^2}$ if N is large.

4. **(10 pts)** Consider the model "G12" for chaotic reversals proposed by C. Gissinger ([2], Eqs. (1)–(3)):

$$\begin{cases} \dot{Q} &= \mu Q - VD, \\ \dot{D} &= -\nu D + VQ, \\ \dot{V} &= \Gamma - V + QD. \end{cases} \tag{1}$$

At these parameter values, the chaotic dynamics of G12 is reminiscent to that of the paleomagnetic data of the Earth for the last 2 million years. This is what makes G12 interesting.

Experiment with three data assimilation algorithms applied to G12 perturbed by noise: EnKF, ETKF, and Particle Filter (Basic). Consider the following setup. The model is given by

$$v_{n+1} = \Psi(v_n) + \sigma\xi_n, \quad v_n := \begin{bmatrix} Q_n \\ D_n \\ V_n \end{bmatrix}, \quad \xi_n \sim \mathcal{N}\left(0, I_{3\times 3}\right), \tag{2}$$

where $\Psi(v_n)$ is the solution of Eq. (1) with the initial condition v_n at time $T = 3$.

```
function f = Psi(x)
f = zeros(size(x));
mu = 0.119;
nu = 0.1;
G = 0.9;
tspan = [0,3];
options = odeset('AbsTol',1e-12,'RelTol',1e-12);
G12 = @(t,x)[mu*x(1) - x(3).*x(2);-nu*x(2) + x(3)*x(1);G - x(3) + x(1)*x(2)];
for k = 1 : size(x,2)
    [~,Y] = ode45(G12,tspan,x(:,k),options);
    f(:,k) = Y(end,:)';
end
end
```

The data are given by:

$$y_{n+1} = [0, 1, 0]v_{n+1} + \gamma\eta \equiv D_{n+1} + \gamma\eta, \quad \eta \sim \mathcal{N}(0, 1), \tag{3}$$

i.e., the data are noisy measurements of the D-component of the solution of (1) done at moments of time separated by three time units.

(a) Program and run each filter for $J = 10^3$ steps for $\sigma = 0.1$ and $\gamma = 0.3$. For each filter:

- Plot each component of the actual v_n and the mean m_n, $n = 1, \ldots, J$. For the second component of v_n, also plot the observations y_n.
- Plot the trace of the covariance matrix C and the running mean of the trace of the covariance.
- Plot squared error $\|v_n - m_n\|^2$ and the running mean of $\|v_n - m_n\|^2$.

Compare the performances of the filters.

(b) Run each filter for $J = 10^3$ steps with the follwing values of σ and γ: $(\sigma, \gamma) = (0, 0.1)$ and $(0.1, 0.1)$. It is likely that you will see that some of the filters fail to work correctly at some of these settings. If the filter does not work correctly, identify the problem, i.e., find out what happens to the ensemble and causes the problem.

References

[1] R.R. Coifman, S. Lafon, Diffusion Maps, Appl. Comput. Harmon. Anal. 21 (2006) 5–30

[2] C. Gissinger, A new deterministic model for chaotic reversals Eur. Phys. J. B (2012) 85: 137. https://doi.org/10.1140/epjb/e2012-20799-5