Information and Instructions

- Open-source and citation. This is an open-source exam, on the condition that the sources be properly cited. Proper citation reflects critical selection and use of existing knowledge and eliminates intentional or unintentional plagiarism.

  You may bring in books and notes. The citation includes the page numbers as well as the book titles and authors. You may also search on the Internet, cite the exact web addresses.

- No-live communication is allowed with anyone else, in person or via any electronic media.

- Problem set and grading. Each individual selects **4** out of the 5 problems as the primary ones for the exam grading. Please mark with X the numbers for the 4 problems you choose as the primary.

  (1) ______ (2) ______ (3) ______ (4) ______ (5) ______

  The additional problem, if you wish to submit as well, will not get a score higher than the average of the primary ones.

- Concise answers and arguments are preferred; irrelevant and incorrect comments are subject to point deduction.

- The grading process is blind to individual names. Please mark every page of your answers with the ID provided to you.

- Upon completion, please submit your answers and return the problem description to the DGS office.

  If you decide not to continue and finish the exam, please inform the DGS office so, on the returned problem set.

Good luck !

Confirmation:
I have read the above instructions on Jan. ______, 2019.

The problem set in this exam is related to certain basic concepts or components broadly and frequently used for contemporary data analysis.

1. Denote by P the column-wise stochastic matrix associated with a network (or graph), $\mathbf{1}^\intercal P = \mathbf{1}^\intercal$, where $\mathbf{1}$ is the vector with all elements equal to 1.

   Assume the largest eigenvalue of P is simple. Let x_o be an arbitrary distribution, i.e., $x_0 \geq 0$ and $\mathbf{1}^\intercal x = 1$. Consider the successive random walk sequence

$$x_{k+1} = P x_k, \quad k \geq 0 \tag{1}$$

   (a) Verify that for any $k > 0$, x_k of (1) is also a distribution.

   (b) Make a statement about whether or not the walking sequence $\{x_k\}$ converges as $k \to \infty$.

   (c) Consider the m-step transition matrix P^m, $m > 1$. Make statements about the largest eigenvalue of P^m and about the successive sequence over P^m in the same fashion as with P $(m = 1)$.

   (d) (**Optional**.) Assume a special case that P is reducible, i.e., P can be partitioned into a 2×2 block matrix in certain ordering π among the network nodes (or permutation in the rows and columns of the matrix)

$$P(\pi, \pi) = \begin{pmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{pmatrix},$$

   where the diagonal blocks are square. Verify whether or not the largest eigenvalue of P_{11} is the largest eigenvalue of P.

2. Let matrix P be column-wise stochastic. Let α be a positive scalar $0 < \alpha < 1$. Consider the solution to the personalized or content-specific page-ranking equations

$$(I - \alpha P)x = (1 - \alpha)v, \tag{2}$$

   where v is a content-specific distribution, i.e., $v \neq 0$ and $\mathbf{1}^\intercal v = 1$.

   (a) Verify that under the given conditions, the solution to the ranking equations exists and is unique.

   (b) Verify that the solution x is a distribution, which is referred to as the page-rank distribution.

3. Denote by A a symmetric, positive definite (s.p.d.) matrix that is *big and sparse*. Consider the following optimization problem

$$x_{\text{opt}} = \arg\max_x f(x), \quad f(x) = \exp(-x^{\mathrm{T}} A x/2 + b^{\mathsf{T}} x). \tag{3}$$

   (a) Is the objective function $f(x)$ convex or concave or neither ?

   (b) In the case $b = 0$, find the optimum x_{opt}; in the case $b \neq 0$, verify that it is necessary to have the gradient vanished at the optimum, .i.e.,

$$\nabla_x f(x_{\text{opt}}) = 0.$$

   (c) Assume $b \neq 0$. Let x_0 be the initial point, $\nabla_x f(x_0) \neq 0$. Describe how to generate a sequence of x_{k+1} such that

$$f(x_{k+1}) > f(x_k), \tag{4}$$

   (d) Verify whether or not the sequence $\{f(x_k)\}$ converges as $k \to \infty$.

   (e) (**Optional**.) Verify whether or not the sequence $\{\nabla_x f(x_k)\}$ converges to zero as $k \to \infty$.

4. Consider the solution to the linear system of equations

$$Ax = b, \quad A = I - c\,d^{\intercal}, \quad c,\, d \neq 0$$

with a structured coefficient matrix. This is a reduced case for a system with the coefficient matrix as the sum of a non-singular sparse or structured matrix and a rank-1 matrix.

(a) Verify that if A is nonsingular, then $A^{-1} = I + \beta c\,d^{\intercal}$ for certain scalar $\beta \neq 0$. Give a specific expression of β in terms of the components of A.

(b) Describe the necessary and sufficient condition for A being nonsingular.

(c) When A is nonsingular, describe an economic solution procedure to obtain the solution to the system $Ax = b$.

(d) (**Optional**.) Extend the above results to the matrix $A = I - F_r G_r^{\intercal}$, where F_r and G_r each has r columns, with $r > 1$ but much smaller than the size of A.

5. Describe the adjacency matrix associated with each undirected, un-weighted graph G of n nodes, $n > 1$, with connections specified as follows.

(a) G is a star.

(b) G is a ring.

(c) G is a wheel, i.e., it is a star with leaf nodes connected into a ring, n is odd.

(d) G is a clique.

(e) G is a hypercube, n is a power of 2.