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 !

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

Denote by A a symmetric, positive definite matrix that is *big and sparse.* Consider the following three computational problems and their solution methods.

- Solution to a system of linear equations with A as the coefficient matrix, b as the right hand side,

$$A\,x = b \tag{1}$$

  Optional: Describe a system of linear equations used for PageRank estimation.

- Solution to the linear Least Squares (LS) problem

$$\arg\min_x \frac{1}{2} x^{\mathrm{T}} A x - b^{\mathrm{T}} x \tag{2}$$

- Solution to the optimization problem

$$\arg\min_x \exp(-x^{\mathrm{T}} A x/2 + b^{\mathrm{T}} x) \tag{3}$$

1. (a) Recommend a ***direct*** method that is applicable, under certain conditions, to solving problem (1); specify clearly the assumed conditions; cite reference source(s).
   (b) Recommend an ***iterative*** method that is applicable, under certain conditions, to solving problem (1); specify clearly the assumed conditions; cite reference source(s).
   (c) Describe briefly advantages or limitations of each method mentioned above.
   (d) *Optional:* Describe a system of linear equations used for PageRank estimation.

2. (a) Derive the gradient of the objective function in (2).
   Describe any condition(s) for the existence of a solution. Cite reference source(s).
   (b) Derive the gradient of the objective function in (3).
   Describe any condition(s) for the existence of a solution. Cite reference source(s).
   (c) *Optional:* Describe how (3) is used in Gaussian Belief Propagation modeling and solution.

3. (a) Decide with supporting arguments whether or not problem (1) and problem (2) are equivalent in solution.
   (b) Decide with supporting arguments whether or not problem (3) and problem (2) are equivalent in solution.

4. Assuming further that $A = L - U$ is a splitting of A so that L is the lower triangular portion of the matrix and $\|L^{-1}U\| < 1$ in a well defined matrix norm.

   (a) Verify the following analytic expression of the solution,

$$x = \sum_{k=0}^{\infty} (L^{-1}U)^k (L^{-1}b) \tag{4}$$

   (b) Describe a computationally economic procedure, based on (4), for an approximate solution.

5. (a) Describe a non-linear least squares (LS) problem, and describe a solution method that is applicable; specify briefly and clearly the conditions and limitation. Cite reference source(s).
   (b) Describe any connection and fundamental difference in solution means between linear and non-linear LS problems.