Mathematics of Data Science

M1 IDD

Course project - 2025/2026

- The last version of this document can be found at: **https://www.lamsade.dauphine.fr/∼croyer/ensdocs/MDS/ProjMDS.pdf**.
- Typos, questions, etc, can be sent to `clement.royer@lamsade.dauphine.fr`.
- Current version: November 14, 2025.
- **Major updates to the document**
  - 2025.11.14: First version.

Assignment

- Students should organize as groups of N people with $N \in \{1, 2, 3\}$.
- Students are expected to send their homework to `clement.royer@lamsade.dauphine.fr`. The email and homework must clearly indicate the students' first and last names (all group members should be cc'ed).
- The homework questions may suggest written or coded answers, but the final format is left open (notebook, PDF, slides,...). In particular, it is technically possible to find a solution to all optimization problems without running a solver (although using a computer may help check things faster).
- A lot of freedom is intentionally left in the formulation and solve of the optimization problems at hand, to encourage diversity of approaches among students.
- The deadline to send the sources is **January 23, 2026 AOE** (Anywhere On Earth).

Project Paris-Duchesse

Introduction

Disclaimer: This project is loosely based on the actual renovation process of Université Paris Dauphine-PSL.

Context Université Paris Duchesse is a rectangular building that will undergo renovation. Initially, the building consists of four wings A, B, C, D for a total of 10 offices. A new wing (E) will be built that will add two offices to the total.

The renovation plan will undergo the following phases in order:

1. Wing E opens and wing B is renovated.
2. Wing B re-opens and wing D is renovated;
3. Wing D re-opens and wing C is renovated;
4. Wing C re-opens and wing A is renovated;
5. Wing A re-opens (Final phase).

A key part of the renovation consists in planning the moves of personnel at every phase, so as to minimize the overall moving effort, that is the number of total moves (one person changing offices). Since all offices from wings A/B/C/D are renovated, each individual will necessary move (at least) once.

Task The intended goal of the project is to get from an initial office configuration (called "Phase 0") to a desired office configuration in the final phase ("Phase 5") in the minimum number of moves. We assume that all offices have the same size and can contain two people at most, therefore an office (and its occupants) can be moved to any other office that is not under renovation.

Problem data Figures 1 and 2 present the office allocations in the initial phase and the final phase, respectively. The people in offices belong to one of the five services: **P**residency, **S**tudents association, **O**ptimization, **T**heoretical Computer Science and **M**athematics.

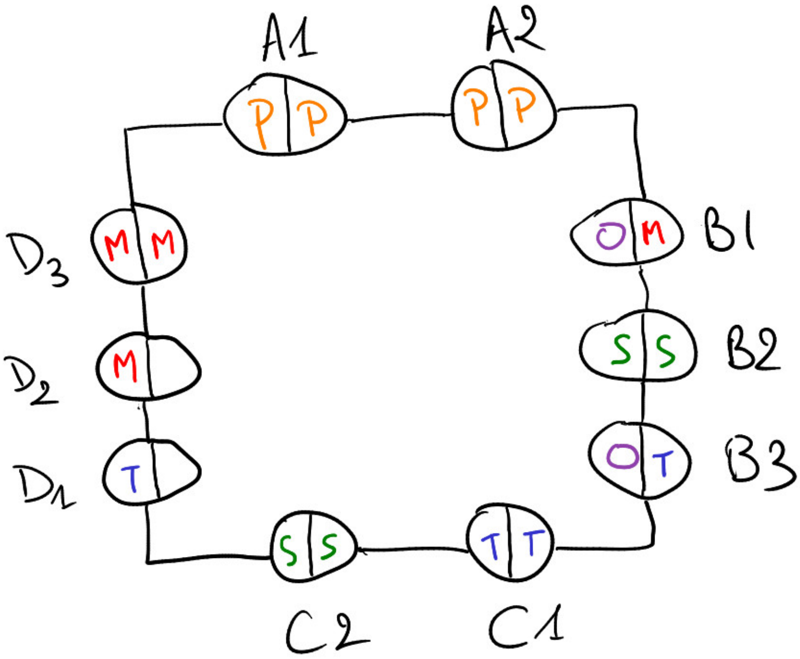

Figure 1: *Initial phase (before building wing E). Offices D1 and D2 are the only half-full offices.*

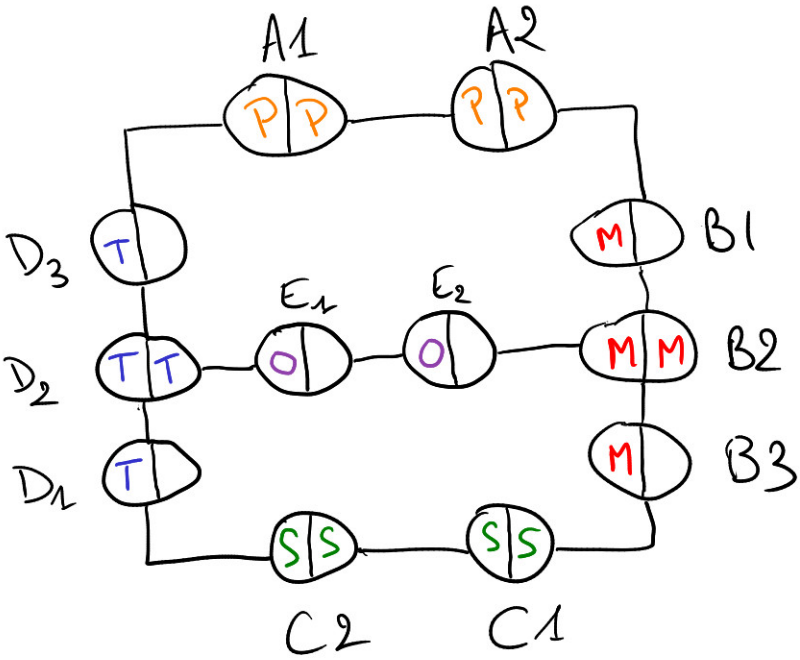

Figure 2: *Final allocation. Offices B1, B3, D1, D3, E1 and E2 are half full.*

1 Linear programming relaxation

A linear relaxation of the proposed approach consists in replacing binary decision variables by continuous ones. As an example[1], given a set of offices $\{1, 2, \ldots, n\}$ and a set of phases $\{0, \ldots, 5\}$, consider x_{ijp} that indicates whether the current occupants of office i at phase p should move to office j during phase p. In real life, this is a binary variable, and thus its value should be in $\{0, 1\}$. To make use of the course's material, we will however consider a relaxation of the *integer* constraint $x_{ijp} \in \{0, 1\}$ into the continuous constraint $0 \leq x_{ijp} \leq 1$.

Question 1 *Write a linear program (LP) corresponding to this problem. Indicate the number of variables and constraints, and write down the associated KKT conditions.*

Question 2 *Compute a solution for your problem and its dual.*

Question 3 *An ideal allocation would avoid putting the students' association offices next to that of the presidency throughout the renovation. Modify your problem accordingly, and compute (if possible) a solution. Compare the solution with that of Question 2.*

2 Penalized problem

Since we know the final allocation of the offices, we may want to try to match the final allocation as early as possible in the renovation process. For this purpose, we add a term to the objective that quantifies how the allocation differs from the final one, of the form

$$\lambda \sum_{p=0}^{5} \|\boldsymbol{z}^p - \boldsymbol{z}^F\|_1 = \lambda \sum_i |[\boldsymbol{z}^p]_i - [\boldsymbol{z}^F]_i|, \tag{1}$$

where $\boldsymbol{z}^p$ represents the vector of office allocations[2] at phase p, $\boldsymbol{z}^F$ is the corresponding vector of final affectations, and $\lambda > 0$ is a weight put on this penalty.

Question 4 *Adapt your linear program of Part 1 to encode the penalty (1). Give the dual of the resulting problem.*

Question 5 *Compute a solution of the resulting problem and its dual for $\lambda = 100$. Do you observe changes compared to the solution obtained in Section 1?*

[1] Other variables can -and likely should- be used in the model. In addition, students are free not to use the variables x_{ijp}.

[2] This notation assumes that this information is stored in a vector. The formula should be adapted to fit the variable representation chosen in Part 1.

3 Semidefinite programming relaxation

We now come back to the problem studied in Section 1. Recall that we provided a linear programming relaxation of the binary constraints that applied to every variable. The goal of this section is to explore another relaxation based on semidefinite programming. As seen in class, optimizing over a vector of binary variables [3] $\boldsymbol{u} \in \{-1,1\}^N$ amounts to considering a matrix $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ with constraints

$$\boldsymbol{U} = \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}, \quad \boldsymbol{U}_{ii} = 1 \ \forall i = 1, \dots, N.$$

Indeed, the constraints on the diagonal elements of $\boldsymbol{U}$ will imply that the elements of $\boldsymbol{u}$ have values in $\{-1,1\}$.

The problem is then relaxed into

$$\boldsymbol{U} \succeq \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}, \quad \boldsymbol{U}_{ii} = 1 \ \forall i = 1, \dots, N. \tag{2}$$

The system (2) can be reformulated in standard SDP form.

Question 6 *Form a semidefinite programming model of the problem based on the relaxation idea* (2)*. Write its dual and KKT conditions.*

Question 7 *Compute the solution to your semidefinite program.*

Question 8 *Apply the randomized approach seen in class to obtain an integer solution for the allocation problem. Is the resulting allocation feasible? Does it seem optimal?*

[3] Note that any vector $\boldsymbol{v} \in \{0,1\}^N$ can be converted into $\boldsymbol{u} \in \{-1,1\}^N$ through the linear transformation $\boldsymbol{u} = 2\boldsymbol{v} - 1$.