1 A heuristic for the maximum-cut problem

Consider an undirected graph with \( n \) nodes and \( m \) edges. Suppose \( B \) and \( C \) partition the set of nodes: that is, \( B \cap C = \emptyset \), and \( B \cup C = \{1, \ldots, n\} \). The number of cuts associated with this partition is the number of edges with between a node in \( B \) and node in \( C \). The maximum-cut problem is the task of choosing a partition that maximizes the number of cuts for a given graph. For any partition, the number of cuts is no more than \( m \); if the number of cuts is equal to \( m \), then the graph is bipartite.

The maximum-cut problem has many applications. We describe one such application here, although you do not need to know about the application in order to solve the problem that follows. Suppose we have a communication system that operates with a two-phase clock. During the even periods \( t = 0, 2, 4, \ldots \), each node in \( B \) transmits data to its neighbors in \( C \); during the odd periods \( t = 1, 3, 5, \ldots \), each node in \( C \) transmits data to its neighbors in \( B \). The number of cuts is the number of successful transmissions that can occur in a two-period cycle. The maximum-cut problem is to assign nodes to the two groups in order to maximize the overall efficiency of communication.

The maximum-cut problem is hard to solve exactly, at least if we do not want to try all, or even most, of the \( 2^n \) possible partitions. In this problem, we explore an approximation algorithm for the maximum-cut problem (that is, a method of finding a partition that has a large number of cuts, although the corresponding partition may not have the maximum number of cuts.)

We can encode a partition using a vector \( x \in \{-1, +1\}^n \): the associated partition has \( i \in B \) if \( x_i = +1 \), and \( i \in C \) if \( x_i = -1 \). We can describe the graph using its adjacency matrix \( A \in \mathbb{R}^{n \times n} \), where

\[
A_{ij} = \begin{cases} 
1 & \text{there is an edge between nodes } i \text{ and } j, \\
0 & \text{otherwise.}
\end{cases}
\]

Note that \( A \) is symmetric, and \( A_{ii} = 0 \) for \( i = 1, \ldots, n \).

(a) Find a symmetric matrix \( P \in S^n \) such that \( P_{ii} = 0 \) for \( i = 1, \ldots, n \), and a constant \( d \in \mathbb{R} \) such that \( x^T P x + d \) is the number of cuts corresponding to the partition encoded by the vector \( x \). Then, the maximum-cut problem can be written as

\[
\begin{aligned}
\text{maximize : } & x^T P x + d \\
\text{subject to : } & x_i^2 = 1 \quad i = 1, \ldots, n.
\end{aligned}
\]

(b) A famous heuristic for the maximum-cut problem is to replace the \( n \) constraints \( x_i^2 = 1 \) with the single constraint \( \sum_{i=1}^n x_i^2 = n \). This gives us the relaxed problem

\[
\begin{aligned}
\text{maximize : } & x^T P x + d \\
\text{subject to : } & \sum_{i=1}^n x_i^2 = n.
\end{aligned}
\]
Explain how to solve the relaxed problem. Let $x^*$ be a solution of the relaxed problem. We can round $x^*$ to a vector $x \in \{-1, +1\}^n$ describing a partition by taking

$$
    x_i = \begin{cases} 
    +1 & x_i^* \geq 0, \\
    -1 & x_i^* < 0.
    \end{cases}
$$

(c) Apply your method to the data in maximum_cut_heuristic_data.m. What is the number of cuts of the partition that you find?

Another heuristic for the maximum-cut problem is to generate a large number of random partitions, and take the one that yields the largest number of cuts. In MATLAB, we can generate a random partition using the following code.

```matlab
x = sign(rand(n,1) - 0.5);
```

Generate $N = 1000$ such random partitions, and report the largest number of cuts achieved by one of the random partitions.

2 Simultaneously estimating student ability and exercise difficulty.

Each of $n$ students takes an exam that contains $m$ questions. Student $j$ receives a grade $G_{ij} \geq 0$ on question $i$. One simple model for predicting grades is to estimate $G_{ij} \approx \hat{G}_{ij} = \frac{a_j}{d_i}$, where $a_j \geq 0$ is the ability of student $j$, and $d_i > 0$ is the difficulty of question $i$. To ensure a unique model, we normalize the exam question difficulties so that the mean exam question difficulty across the $m$ questions is 1; otherwise, $a_j$ and $d_i$ would not be uniquely determined because we can scale them both by any positive constant without changing our estimates.

In this problem you are given the matrix $G \in \mathbb{R}^{m \times n}$ of grades. Your task is to find a set of nonnegative student abilities, and a set of positive, normalized question difficulties such that $G_{ij} \approx \hat{G}_{ij}$. In particular, choose your model in order to minimize the root mean squared error:

$$
    J = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (G_{ij} - \hat{G}_{ij})^2.
$$

(a) Explain how to solve this problem. You can ignore the constraints that the $a_j$ be nonnegative, and the $d_i$ be positive.

(b) Carry out your method on the data in ability_difficulty_data.m. Report the optimal value of $J$, the ratio of the optimal value of $J$ and the root mean squared value of $G_{ij}$; give the difficulties of the exam questions.

3 Consistent ellipsoids

Consider a measurement model

$$
    y = Ax + v,
$$

Alex Lemon
EE263: Introduction to Linear Dynamical Systems
August 6, 2015
In-class exercises 11: Page 2 of 3
where \( y \in \mathbb{R}^m \) is a measurement (which is known), \( x \in \mathbb{R}^n \) is a vector of unknown parameters (which we want to estimate), \( v \in \mathbb{R}^m \) is measurement noise (which is unknown), and \( A \in \mathbb{R}^{m \times n} \) describes the measurement system. We will assume that \( A \) is skinny and full rank, and \( v \) is norm-bounded: that is, there is a known \( \alpha > 0 \) such that \( \|v\| \leq \alpha \).

(a) First, consider the specific case when
\[
A = \begin{bmatrix} 2 & -1 \\ \end{bmatrix}, \quad y = \begin{bmatrix} 9 \\ 23 \end{bmatrix}, \quad \text{and} \quad \alpha = 25.
\]
We say that the parameter \( x \in \mathbb{R}^n \) is consistent with the model if there exists a noise vector \( v \in \mathbb{R}^m \) such that \( y = Ax + v \), and \( \|v\| \leq \alpha \). Find the set of all parameters \( x \) that are consistent with the specific model given above. You can do this by finding the set of all \( x \) such that \( \|y - Ax\| \leq \alpha \).

(b) Your friend proposes a different method for determining the set of consistent parameters. Since \( y = Ax + v \), and \( A^\dagger A = I \) for a skinny and full rank matrix, he argues that
\[
\hat{x} = A^\dagger y = A^\dagger (Ax + v) = x + A^\dagger v,
\]
and hence that
\[
x = \hat{x} - A^\dagger v,
\]
where \( \hat{x} = A^\dagger y \) is the least-squares estimate of \( x \). Therefore, the set of all parameter vectors \( x \) that are consistent with the model is
\[
\hat{x} - A^\dagger \{v \in \mathbb{R}^m \mid \|v\| \leq \alpha \}.
\]
(i) According to your friend’s analysis, what is the set of all parameter vectors \( x \) that are consistent with the specific model given in (a)? Is the consistent with the result that you derived earlier?
(ii) If your friend’s analysis does not produce the same answer you derived earlier, why is there a difference? Which do you think is correct: your friend’s analysis, or the answer you derived earlier?

(c) Now consider the general case. Let the singular-value decomposition of \( A \) be \( A = U \Sigma V^T \). By “completing the square,” show that the set of all parameters \( x \in \mathbb{R}^n \) that are consistent with the model is the empty set if \( \alpha^2 - \|\hat{\epsilon}\|^2 < 0 \), and is the set
\[
\left\{ \hat{x} + V \Sigma^{-1} z \mid \|z\| \leq \sqrt{\alpha^2 - \|\hat{\epsilon}\|^2} \right\}
\]
if \( \alpha^2 - \|\hat{\epsilon}\|^2 \geq 0 \), where \( \hat{\epsilon} = y - A\hat{x} \).

(d) The file consistent_ellipsoids_data.m contains values of \( A \), \( y \), and \( \alpha \). For this example, the vector of parameters has dimension \( n = 2 \). Plot the set of all parameter vectors \( x \) that are consistent with the model.