This is a summary of existing manuscript, with additional goodies and directions.

1 graph matching

For each $A \in \mathbb{R}^{n\times n}$, we will view A as the adjacency matrix for a graph in the following way. The graph's vertices are the integers $\{1, 2, \ldots, n\}$ and, for all $i, j \in \{1, 2, \ldots, n\}$, there is an ij edge precisely when $A_{ij} \neq 0$; the value of nonzero A_{ij} is understood to be a multiplicity associated with the edge ij. If the graph is simple then A is $0,1$-valued, symmetric, and has zeros on the main diagonal. If we allow edges to be directed, to be loops, and/or to have real-valued multiplicity then every matrix in $\mathbb{R}^{n\times n}$ is an adjacency matrix in the above manner.

For any $A, B \in \mathbb{R}^{n\times n}$, their underlying graphs are said to be isomorphic if there is a bijection between their respective vertex sets that preserves all adjacency structure; more formally, the underlying graphs are isomorphic precisely when there exists a permutation matrix P (that is, every row and every column of P has has a single 1 and all other entries 0) such that $A = PBP^T$. The *graph matching* problem is to find the permutation matrix P that minimizes $\|A - PBP^T\|_F$, where $\|\cdot\|_F$ denotes the Froebenius norm. Some researchers describe the graph matching problem assuming that the underlying graphs are indeed isomorphic, in which case the optimal permutation matrix P provides the actual isomorphism; we (and others) also include the cases where the underlying graphs are not isomorphic since it can often be useful to match such graphs' vertices through a bijection which minimizes the least-squares discrepancy of the adjacency structure.

2 basic notation

The set of all $n \times n$ permutation matrices will be denoted $\mathcal{P}$, and the set of all $n \times n$ doubly stochastic matrices will be denoted $\mathcal{D}$. Let $\vec{1}$ denote the vector in $\mathbb{R}^n$ with 1 in each entry. Note that $\mathcal{D}$ consists of the matrices in $\mathbb{R}^{n\times n}$ which satisfy the conditions $P\vec{1} = \vec{1}$, $\vec{1}^TP = \vec{1}^T$, and $P \geq 0$ entrywise, hence $\mathcal{D}$ is defined solely by linear constraints ($\mathcal{D}$ is a polyhedral set, ie it is the intersection of a finite number of halfspaces), and $\mathcal{P}$ is precisely the binary-valued members of $\mathcal{D}$. The Birkhoff-von Neumann Theorem states that $\mathcal{D}$ is precisely the convex hull of $\mathcal{P}$.

For any matrices $A, B \in \mathbb{R}^{n\times n}$, the Euclidean inner product of A,B is $\langle A, B\rangle := \mathrm{trace} A^TB$; this is just the sum of products of corresponding entries of A and B.

3 assignment problems

We begin by defining the Linear Assignment Problem and the Quadratic Assignment Problem; the latter is precisely the graph matching problem, and we will use a subroutine solving the former problem as part of our algorithm for the latter problem.

3.1 assignment problems definitions and relaxations

Given matrices $A, B \in \mathbb{R}^{n \times n}$, the *Linear Assignment Problem* (LAP) is

$$\text{(LAP)} \quad \text{Minimize} \quad \|A - PB\|_F \ \text{ such that } \ P \in \mathcal{P}. \tag{1}$$

Expanding $\|A - PB\|_F^2$, we obtain

$$\begin{aligned}
\|A - PB\|_F^2 &= \text{trace}(A - PB)^T(A - PB) \\
&= \text{trace}A^TA - 2\text{trace}A^TPB + \text{trace}(PB)^TPB \\
&= \text{trace}A^TA - 2\text{trace}BA^TP + \text{trace}(PB)^TPB \\
&= \|A\|_F^2 - 2\langle AB^T, P\rangle + \|B\|_F^2,
\end{aligned}$$

from which we obtain an equivalent formulation of LAP as

$$\text{(LAP)} \quad \text{Maximize} \quad \langle AB^T, P\rangle \ \text{ such that } \ P \in \mathcal{P}. \tag{2}$$

("Equivalence" of problems means that a solution— ie optimal decision variables—of one problem provides a solution to the other problem.) The *Quadratic Assignment Problem* (QAP) is

$$\text{(QAP)} \quad \text{Minimize} \quad \|A - PBP^T\|_F \ \text{ such that } \ P \in \mathcal{P}, \tag{3}$$

which is precisely the graph matching problem, and it can be equivalently expressed as

$$\text{(QAP)} \quad \text{Minimize} \quad \|AP - PB\|_F \ \text{ such that } \ P \in \mathcal{P}. \tag{4}$$

Expanding $\|A - PBP^T\|_F^2$, we get

$$\begin{aligned}
\|A - PBP^T\|_F^2 &= \text{trace}\left[(A - PBP^T)^T(A - PBP^T)\right] \\
&= \text{trace}A^TA - 2\text{trace}A^TPBP^T + \text{trace}(PBP^T)^TPBP^T \\
&= \|A\|_F^2 - 2\langle A, PBP^T\rangle + \|B\|_F^2,
\end{aligned}$$

from which we obtain an equivalent formulation of QAP as

$$\text{(QAP)} \quad \text{Maximize} \quad \langle A, PBP^T\rangle \ \text{ such that } \ P \in \mathcal{P}. \tag{5}$$

The feasible regions of QAP and LAP are discrete, and the *relaxations* of these problems are defined by substituting nonnegativity of the decision variables in place of the binary restriction. That is, the Relaxed Linear Assignment Problem is defined (relaxing the Line 2 formulation) as

$$\text{(rLAP)} \quad \text{Maximize} \quad \langle AB^T, P \rangle \;\text{ such that }\; P \in \mathcal{D}, \tag{6}$$

and the Relaxed Quadratic Assignment Problem is defined (relaxing the Line 5 formulation) as

$$\text{(rQAP)} \quad \text{Maximize} \quad \langle A, PBP^T \rangle \;\text{ such that }\; P \in \mathcal{D}, \tag{7}$$

and a different relaxation of QAP (relaxing the Line 4 formulation) is

$$\text{(rQAP2)} \quad \text{Minimize} \quad \|AP - PB\|_F \;\text{ such that }\; P \in \mathcal{D}. \tag{8}$$

Both rQAP and rQAP2 are quadratic programming problems; ie each seeks to optimize a quadratic objective function subject to linear constraints. Yet another relaxation of QAP (relaxing Line 3) is

$$\text{(rQAP3)} \quad \text{Minimize} \quad \|A - PBP^T\|_F \;\text{ such that }\; P \in \mathcal{D}, \tag{9}$$

but rQAP3 is quartic, and not quadratic like the other two relaxations rQAP and rQAP2.

3.2 assignment problems: complexity and comparisons

Note that rLAP is a linear program (it has linear constraints and a linear objective function), and is thus efficiently solvable. LAP is an integer programming problem, and integer programming problems in general may be difficult to solve. However, interestingly, LAP is equivalent to rLAP. Indeed, the so-called Hungarian Algorithm (Kuhn 1955 and Munkres 1957, souped up by Edmonds and Karp, and independently Tomizawa to go faster) solves LAP in $O(n^3)$ operations.

Proposition 1. *The problem LAP is equivalent to rLAP, in the following way. Any solution to LAP is a solution to rLAP. Conversely, if P is any solution to rLAP, say $P = \sum_{i=1}^k \alpha_i P^{(i)}$ for positive integer k, permutation matrices $\{P^{(i)}\}_{i=1}^k$, and positive real numbers $\{\alpha_i\}_{i=1}^k$ such that $\sum_{i=1}^k \alpha_i = 1$, then it holds for all $i = 1, 2, \ldots, k$ that $P^{(i)}$ is a solution to LAP.*

Proposition 1 is usually shown by noting that the constraint matrix of LAP is totally unimodular. There is also a very straightforward argument. Suppose the permutation matrix P' is a solution to LAP, and suppose by way of contradiction that not all $P^{(i)}$ (in the notation of the proposition) are solutions to LAP. Then $\langle AB^T, P \rangle = \langle AB^T, \sum_{i=1}^k \alpha_i P^{(i)} \rangle = \sum_{i=1}^k \alpha_i \cdot \langle AB^T, P^{(i)} \rangle < \langle AB^T, P' \rangle$ contradicts the optimality of P in rLAP since P' is feasible in rLAP. The result follows. $\square$

In contrast to LAP, which is indeed efficiently solvable, QAP is NP-hard. Even the problem GRAPH ISOMORPHISM (a decision problem which just determines if a graph is isomorphic to another, but differing from Graph Matching in that it does not output any isomorphism) is not known to be in complexity class P. Interestingly, if P $\neq$ NP, then it has been shown that there are infinitely many different complexity equivalence classes between P and NP-complete, and GRAPH ISOMORPHISM COMPLETE is one of only a few suspected candidates for such intermediacy.

It is interesting to note that, in contrast to LAP/rLAP, the optimal objective function value of QAP in Line 3 is not the same as that of rQAP3 even for adjacency matrices associated with simple graphs; for a specific example, consider the matrices

$$A := \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad B := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

and note that for all permutation matrices P it holds that $\|A - PBP^T\|_F = \sqrt{2}$, whereas if P is the doubly stochastic 2×2 matrix all of whose entries are $\frac{1}{2}$ then here $\|A - PBP^T\|_F = 1 < \sqrt{2}$. It is also interesting to note that rQAP3 is not symmetric in its arguments even for adjacency matrices associated with simple graphs; indeed, in the above example, for any double stochastic 2×2 matrix P we have $\|B - PAP^T\|_F = \sqrt{2} > 1$.

By contrast, rLAP is symmetric in its arguments. One way to see this is through its equivalence to LAP which is symmetric in its arguments, since for any permutation matrix P it holds that $\|A - PB\|_F = \| - P^T(A - PB)\|_F = \|B - P^T A\|_F$ (and transposition is an involution on $\mathcal{P}$).

The relaxation rQAP2, unlike rQAP3, is indeed symmetric in its arguments (for adjacency matrices associated with simple graphs) since, for any doubly stochastic matrix P, $\|AP - PB\|_F = \| - (AP - PB)^T\|_F = \|BP^T - P^T A\|_F$ (and transposition is an involution on $\mathcal{D}$). Also note that the optimal objective function value of QAP in Line 4 is not the same as the optimal objective function value of rQAP2 even for adjacency matrices associated with simple graphs; Tinhoffer and Scheinerman et al (cite Ed and Dan Ullman's textbook "Fractional Graph Theory," Chapter 6, available online at Ed's webpage–see notes at end of Chapter 6) proved that for simple graphs the optimal objective function of rQAP2 is 0 if and only if the two graphs have the same iterated degree sequence and if and only if the two graphs share an equitable partition. Thus, for example, the graph consisting of two disjoint 3-cycles and the graph consisting of a 6-cycle are not isomorphic—yet the optimal objective function of rQAP2 is 0.

I am currently not sure whether rQAP is symmetric in its arguments in general, and I am not sure if the optimal objective function of rQAP can be different from the optimal objective function value of QAP in Line 5, except in the following special case.

Proposition 2. *If the matrices A, B correspond to simple graphs that are isomorphic to each other then the optimal objective function for rQAP is the same as the optimal objective function value of QAP in Line 5.*

Proof: First, note that if $P \in \mathbb{R}^{n\times n}$ is a doubly stochastic matrix and $x \in \mathbb{R}^n$ then, by submultiplicativity of the operator norm induced by the ℓ_∞ norm, $\|Px\|_\infty \leq \|P\|_{\infty,\infty}\|x\|_\infty = \|x\|_\infty$, thus if each entry of x is between 0 and 1 inclusive then each entry of Px is between 0 and 1 inclusive. Thus, since B has binary-valued entries, we get that all entries of PBP^T are between 0 and 1. This can be seen by multiplying P times B when B is broken into columns, then break $(PB)^T$ into columns and premultiply these columns by P. Finally, let P' be the permutation matrix such that $A = P'BP'^T$; we have $\langle A, P'BP'^T\rangle = \langle A, A\rangle =$ number of 1's in A, which is greater than or equal to the sum of the entries of PBP^T which correspond to the 1-entries of A (since such entries of PBP^T are between 0 and 1), which is precisely $\langle A, PBP^T\rangle$. $\square$

Of course, if the matrices A, B correspond to simple graphs that are isomorphic to each other then the optimal objective function for rQAP2 is the same as the optimal objective function value of QAP in Line 4.

4 Frank-Wolfe

In this section we describe the Frank-Wolfe Algorithm and apply it to rQAP and Graph Matching.

4.1 general Frank-Wolfe

We now state the Frank-Wolfe Algorithm in general (before we apply it to our case of interest). The general optimization problem is

$$\text{(FWP)} \quad \text{Minimize} \quad f(x) \ \text{ such that } \ x \in S, \tag{10}$$

where $S \subset \mathbb{R}^m$ is a polyhedral set (ie is described by linear constraints) and the function $f : S \to \mathbb{R}$ is continuously differentiable. (When the Franke-Wolfe Algorithm was introduced, f was assumed to be quadratic, but nonquadratic functions are also currently treated with the algorithm. In our application here f will be quadratic.) A starting point $x^{(1)} \in S$ is chosen in some fashion, perhaps arbitrarily. For $i = 1, 2, 3, \ldots$, the following is done. The function $\tilde{f}^{(i)} : S \to \mathbb{R}$ is defined to be the first order (ie linear) approximation to f at $x^{(i)}$—that is, $\tilde{f}^{(i)}(x) := f(x^{(i)}) + \nabla f(x^{(i)})^T(x - x^{(i)})$; then solve the linear program: minimize $\tilde{f}^{(i)}(x)$ such that $x \in S$ (this can be done efficiently since it is a linear objective function with linear constraints, and note that, by ignoring additive constants, the objective function of this subproblem can be abbreviated: minimize $\nabla f(x^{(i)})^T x$ such that $x \in S$), say the solution is $\tilde{x}^{(i)} \in S$. Now, the point $x^{(i+1)} \in S$ is defined as the solution to: minimize $f(x)$ such that x is on the line segment from $x^{(i)}$ to $\tilde{x}^{(i)}$ in S. (This is a just a one dimensional optimization problem; in the case where f is quadratic this can be analytically solved exactly.) Go to the next i, and terminated this iterative procedure when the sequence of iterates $x^{(1)}, x^{(2)}, x^{(3)}, \ldots$ stops changing much or develops a gradient close enough to zero.

4.2 applying Frank-Wolfe to graph matching

We are interested in solving Graph Matching for given adjacency matrices A and B; that is, we want to solve QAP. We will attempt this by using Frank-Wolfe to quickly get an inexact solution to rQAP or rQAP2, and then output the "closest" permutation matrix to our inexact solution of rQAP or rQAP2, considering it to be an inexact solution to QAP, ie Graph Matching.

Which shall we use, rQAP or rQAP2? We will look at advantages of each. A significant advantage of rQAP2 is that the objective function to be minimized is "more convex" than the analogous objective function of rQAP: Indeed, the gradient of $-\langle A, XBX^T\rangle$ (we henceforth consider rQAP as a minimization problem by negating the objective function) is $-AXB^T - A^TXB$ which has Hessian $-B \otimes A \;-\; B^T \otimes A^T$ which has no indication at all of necessarily being positive semidefinite. In contrast, the gradient of $\|AX - XB\|_F^2$ is $2\left(A^TAX + XBB^T - A^TXB - AXB^T\right)$, which has Hessian $2\left(I \otimes A^TA \;+\; BB^T \otimes I \;-\; B \otimes A \;-\; B^T \otimes A^T\right)$ which (ignoring the "2") is the same as the Hessian of rQAP with the addition of two positive semidefinite matrices! We are ignoring rQAP3 for the reasons mentioned previously.

Now, to run Frank-Wolfe to try to solve rQAP or rQAP2: Start by arbitrarily selecting a doubly stochastic matrix $P^{(1)} \in \mathcal{D}$. For $i = 1, 2, 3, \ldots$ successively, let $G^{(i)}$ be the gradient of the objective function at $P^{(i)}$ (as mentioned, the gradient of $-\langle A, XBX^T\rangle$ is $-AXB^T - A^TXB$ and the gradient of $\|AX - XB\|_F^2$ is $2\left(A^TAX + XBB^T - A^TXB - AXB^T\right)$, which are used according as you are working with rQAP or rQAP2) and then solve the problem: minimize $\langle G^{(i)}, P\rangle$ such that $P \in \mathcal{D}$. Of course, this is exactly an rLAP problem (negating the objective to change rLAP from maximization to minimization), which is equivalent to LAP, and a solution is obtained by the Hungarian Algorithm. Denote this solution $\tilde{P}^{(i)}$. Then minimize the objective function on the line segment from $P^{(i)}$ to $\tilde{P}^{(i)}$, and denote the solution (to this one dimensional problem) as $P^{(i+1)}$.

When satisfied that the iterates $P^{(1)}, P^{(2)}, P^{(3)}, \ldots$ have gotten close enough to a local optima of rQAP, say the current iterate is P^*, then project P^* to the class of permutation matrices by maximizing $\langle P^*, P\rangle$ such that $P \in \mathcal{P}$; this problem is LAP, and the optimal P is the output of our procedure, and provides the vertex bijection that is our inexact graph matching.

5 a linear programming approach: alternative to quadratic approach

Suppose A and B are $n \times n$ adjacency matrices corresponding to graphs that happen to be isomorphic. Then the rQAP2 problem is to find the $n \times n$ doubly stochastic P such that $AP = PB$ or, in other words, to solve $AP = PB$, $P\vec{1}_n = \vec{1}_n$, $\vec{1}_n^T P = \vec{1}_n^T$, $P \geq 0$. Let $x \in \mathbb{R}^{n^2}$ denote the vector consisting of the columns of P concatenated into a long vector. Then we are equivalently trying to solve $Mx = r$, $x \geq 0$ where

$$M = \begin{bmatrix} I_n \otimes A - B^T \otimes I_n \\ I_n \otimes \vec{1}_n^T \\ \vec{1}_n^T \otimes I_n \end{bmatrix} \quad \text{and} \quad r = \begin{bmatrix} \vec{0}_{n^2} \\ \vec{1}_n \\ \vec{1}_n \end{bmatrix}$$

Now, this can be solved by solving the linear program:
minimize $\left[\begin{smallmatrix} \vec{0}_{n^2} \\ \vec{1}_{n^2+2n} \end{smallmatrix}\right]^T \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ such that $[M | I_{n^2+2n}] \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = r$ and $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] \geq 0_{2n^2+2n}$ for variables $x \in \mathbb{R}^{n^2}$ and the artificial variables $y \in \mathbb{R}^{n^2+2n}$; running the simplex method from the basis corresponding to the identity matrix I_{n^2+2n}, the artificial variables y will all eventually become 0 and the resulting x corresponds to the desired doubly stochastic P above. Then "project" P to the class of permutation matrices by the LAP problem $\max_{P' \in \mathcal{P}} \langle P, P' \rangle$ and adopt the bijection underlying the optimal P' as the bijection which is the approximate isomorphism.

6 stray notes to self

(Check these out for non 2×2!) has been (mostly) checked for 2×2
If $X, A, B \in \mathbb{R}^{n \times n}$ then
$\nabla \text{trace} X = I$
$\nabla \text{trace} A^T X = A$
$\nabla \text{trace} X^T X = 2X$
$\nabla \text{trace} X^T A X = AX + A^T X$
$\nabla \text{trace} A X^T X = XA + XA^T$
$\nabla \text{trace} B X^T A^T X = AXB^T + A^T XB$
$\nabla \text{trace} A^T X B X^T = AXB^T + A^T XB$
In general, if $g(X) := \nabla f(X)$ and $Y = X^T$ then $\nabla f(Y) = \left(g(X^T)\right)^T$

$\nabla\|AX - XB\|_F^2 = \nabla\text{trace}(AX - XB)^T(AX - XB)$
$= \nabla\text{trace}X^TA^TAX - 2\nabla\text{trace}B^TX^TAX + \nabla\text{trace}B^TX^TXB$
$= \nabla\text{trace}A^TAXX^T - 2\nabla\text{trace}B^TX^TAX + \nabla\text{trace}BB^TX^TX$
$= 2A^TAX - 2(A^TXB + AXB^T) + 2XBB^T = 2[A^TAX + XBB^T - A^TXB - AXB^T]$

If $\nabla = AXB$ then, after concatenating columns of X into a long column, $\nabla^2 = B^T \otimes A$ where

$$B^T \otimes A := \begin{bmatrix} b_{11}A & b_{21}A & \dots \\ b_{12}A & b_{22}A & \dots \\ \vdots & \vdots & \end{bmatrix}$$

Gradient of $-\langle A, XBX^T\rangle$ is $-AXB^T - A^TXB$ with Hessian $-B \otimes A \quad - \quad B^T \otimes A^T$.

Gradient of $\|AX - XB\|_F^2$ is $2\left(A^TAX + XBB^T - A^TXB - AXB^T\right)$,
with Hessian $2\left(I \otimes A^TA \quad + \quad BB^T \otimes I \quad - \quad B \otimes A \quad - \quad B^T \otimes A^T\right)$