Convex relaxation for the planted clique, biclique and cluster problems

Brendan Ames[1] Stephen Vavasis[1]

[1]Department of Combinatorics and Optimization
University of Waterloo

2010-March-2 / BIRS

Convex Relaxation

- A classic solution technique for combinatorial optimization is *convex relaxation*: enlarge the feasible region or underestimate the objective function to yield an optimization problem with convex feasible region and objective function.
- Convex optimization is usually ‘easy’ to solve.
- The solution to the relaxation yields a lower bound on the optimizer.
- More recent idea: sometimes the relaxed solution is optimal for the original problem for instances constructed in a certain manner.

Example: compressive sensing

- The *sparsest vector* problem is: find the vector $\mathbf{x}$ with the fewest number of nonzero entries satisfying underdetermined linear equations $A\mathbf{x} = \mathbf{b}$.
- This problem is NP-hard.
- [Cf. Donoho; Candès, Romberg and Tao; Zhang; others.] Suppose that A has the *spherical section* property. Suppose also that solution $\mathbf{x}^*$ is sufficiently sparse. Then the convex relaxation $\min \|\mathbf{x}\|_1$ s.t. $A\mathbf{x} = \mathbf{b}$ yields $\mathbf{x}^*$.

Maximum clique and biclique problems

- Clique: Given an undirected graph (V, E), find k vertices mutually interconnected such that k is maximized
- Biclique: Given a bipartite graph (U, V, E), find a subgraph (U^*, V^*, E^*) containing all possible $|U^*| \cdot |V^*|$ edges such that $|U^*| \cdot |V^*|$ is maximized
- Max-clique and max-biclique are both NP-hard

Biclique reformulation as rank minimization

- Existence of an *mn* biclique as rank minimization:

$$\begin{array}{lll} \min & \operatorname{rank}(X) & \\ \text{s.t.} & X(i,j) \in [0,1] & \forall (i,j) \in U \times V \\ & X(i,j) = 0 & \forall (i,j) \in (U \times V) - E \\ & \sum_{(i,j)} X(i,j) \geq mn & \end{array}$$

- Similar formulation exists for clique

Matrix rank minimization

- Matrix rank minimization is an optimization problem: $\min \operatorname{rank}(X)$ s.t. $X \in C$, where C is a convex subset of $\mathbf{R}^{m \times n}$.
- In general, the problem is NP-hard.

Matrix rank minimization and nuclear norm

- *Nuclear norm* of X, written $\|X\|_*$ is sum of X's singular values.
- Several authors, e.g., Fazel thesis (2002), suggested nuclear norm as a relaxation of rank. Nuclear norm is a convex function.
- Recht, Fazel, Parrilo (2007) showed nuclear norm relaxation is exact for an interesting class of matrix rank minimization problems

Matrix rank minimization and compressive sensing

- RFP extended compressive sensing properties to rank minimization: If $A \in \mathbf{R}^{m\times n\times p}$ satisfies a certain property, $\hat{X}$ is sufficiently low rank, and $\mathbf{b} = A\hat{X}$, then $\hat{X}$ can be recovered by minimizing $\|X\|_*$ subject to $AX = \mathbf{b}$.
- Nuclear norm minimization can be rewritten as semidefinite programming.

Nuclear norm relaxation

- Nuclear norm relaxation of biclique:

$$
(NNR)\quad \begin{array}{ll} \min & \|X\|_* \\ \text{s.t.} & X(i,j) \geq 0 \qquad \forall (i,j) \in U \times V, \\ & X(i,j) = 0 \qquad \forall (i,j) \in (U \times V) - E, \\ & \sum_{(i,j)} X(i,j) \geq mn. \end{array}
$$

- This relaxation is convex.

Our results for clique

- Consider an N-node graph G consisting of an n-node clique K_n plus diversionary edges:
  - Up to $O(n^2)$ deterministically-placed diversionary edges; at most $O(n)$ K_n-vertices adjacent to any non-K_n-vertex, or,
  - All nonclique edges inserted independently at random with probability p, and $N = O(n^2)$.
- Then the nuclear norm relaxation finds the maximum clique.
- Similar results for biclique.

Optimality of deterministic result

- If the adversary could place $\Omega(n^2)$ diversionary edges, he could create a new n-clique.
- If the adversary could insert edges to make a nonclique node adjacent to n clique nodes, he could enlarge the planted clique.

Subgradient of nuclear norm

- Proof technique: show that the maximum clique is optimal for (NNR) by showing that KKT conditions are satisfied. Furthermore show that optimal solution is unique.
- Suppose $A \in \mathbf{R}^{m \times n}$ has rank r and SVD $A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$. Then $\phi \in \partial \|A\|_*$ iff $\phi = \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \mathbf{u}_r \mathbf{v}_r^T + W$ s.t. $\|W\| \leq 1$, $\mathrm{span}(W) \perp \mathrm{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$, $\mathrm{span}(W^T) \perp \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$.

KKT conditions (biclique case)

Theorem. Suppose X is a feasible rank-one matrix $X = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$, where $\bar{\mathbf{u}}, \bar{\mathbf{v}}$ are the characteristic vectors of $U^* \subset U$, $V^* \subset V$ resp, $|U^*| = m$, $|V^*| = n$. Then X is optimal for (NNR) iff $\exists W \in \mathbf{R}^{M\times N}, \lambda \in \mathbf{R}^{M\times N}, \mu \in \mathbf{R}$ s.t.

$$\frac{\bar{\mathbf{u}}\bar{\mathbf{v}}^T}{\sqrt{mn}} + W = \mu\mathbf{e}\mathbf{e}^T + \sum_{(i,j)\in(U\times V)-E} \lambda_{ij}\mathbf{e}_i\mathbf{e}_j^T$$

with $\|W\| \leq 1$, $W^T\bar{\mathbf{u}} = \mathbf{0}$, $W\bar{\mathbf{v}} = \mathbf{0}$, $\mu \geq 0$. In this case, U^*, V^* is an optimal solution for the max biclique problem.

KKT conditions (biclique case)

Theorem. Suppose X is a feasible rank-one matrix $X = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$, where $\bar{\mathbf{u}}, \bar{\mathbf{v}}$ are the characteristic vectors of $U^* \subset U$, $V^* \subset V$ resp, $|U^*| = m$, $|V^*| = n$. Then X is optimal for (NNR) iff $\exists W \in \mathbf{R}^{M\times N}, \lambda \in \mathbf{R}^{M\times N}, \mu \in \mathbf{R}$ s.t.

$$\frac{\bar{\mathbf{u}}\bar{\mathbf{v}}^T}{\sqrt{mn}} + W = \mu\mathbf{e}\mathbf{e}^T + \sum_{(i,j)\in(U\times V)-E} \lambda_{ij}\mathbf{e}_i\mathbf{e}_j^T$$

with $\|W\| \leq 1$, $W^T\bar{\mathbf{u}} = \mathbf{0}$, $W\bar{\mathbf{v}} = \mathbf{0}$, $\mu \geq 0$. In this case, U^*, V^* is an optimal solution for the max biclique problem. If, in addition, $\mu > 0$ and $\|W\| < 1$, X is the unique optimizer.

Finding W, λ, μ

- Thus, showing that (NNR) finds the optimal biclique reduces to constructing W, λ, μ.
- Our paper gives explicit formulas for W, λ, μ.
- Proof that KKT conditions hold for W, λ, μ constructed by our formulas in the case of randomly chosen noise edges boils down to estimating the norm of a random matrix.

Norm of W: randomized case

- Theorem (Geman, 1980). Suppose $\hat{W}$ is an $M \times N$ random matrix with $M \sim N$ and with entries chosen independently from a fixed distribution whose mean is 0 (plus a few other assumptions). Then with probability exponentially close to 1, $\|\hat{W}\| \leq O(\sqrt{N})$.
- Can show $W \approx \hat{W}/\sqrt{mn}$.
- Implies that we can take M, N as large as m^2, n^2 and still obtain $\|W\| \leq 1$.

Analysis of randomization in clique case

- Follows the same lines, except in place of Geman's theorem we require Füredi and Komlós's (1981) analysis of the norm of a random symmetric matrix.
- Similar result obtained: our algorithm can find a "planted" clique of with n nodes, $n(n-1)/2$ edges, in a random graph with $O(n^2)$ vertices (and hence $O(n^4)$ edges).

The combinatorial clustering problem

- Clustering: given a sequence of data points with known pairwise distances, group them into clusters so that points in each cluster are closer to each other than to points in other clusters.
- Can be posed very generally as follows. Given a graph on n data points, where edges indicate compatibility, find a set of s disjoint cliques that cover as many nodes as possible.
- Obviously NP-hard since the $s = 1$ case is the classical max clique problem.

Our convex relaxation for combinatorial cluster problem

$$
(SDR) \quad \begin{array}{rll}
\text{maximize} & \sum \sum X_{ij} & \\
\text{s.t.} & X\mathbf{e} \leq \mathbf{e}, & \\
& \mathrm{trace}(X) = s, & \\
& X_{ij} = 0 & \forall (i,j) \notin E, \\
& X \geq 0 & \text{(semidefiniteness)}
\end{array}
$$

Solution induced by s cliques

- Suppose graph contains s disjoint cliques $C_1, \ldots, C_s$; sizes $c_1, \ldots, c_s$.
- Then

$$X = \begin{pmatrix} 1/c_1 & \cdots & 1/c_1 & & & & \\ \vdots & & \vdots & & & & \\ 1/c_1 & \cdots & 1/c_1 & & & & \\ & & & \ddots & & & \\ & & & & 1/c_s & \cdots & 1/c_s \\ & & & & \vdots & & \vdots \\ & & & & 1/c_s & \cdots & 1/c_s \end{pmatrix}$$

in which first c_1 rows/cols correspond to C_1, etc, is feasible and has objective value of $\sum_j c_j$.

Our results

- Our result (A & Vavasis, in progress) is that the relaxation described above is exact for combinatorial cluster problem contaminated by noise (extra nodes and edges).
- We assume the cliques are all within a constant factor of each other; let α denote the min clique size.
- Again, two cases: deterministic noise and random noise.

Adversary chosen (deterministic) noise

- The adversary can insert up to $O(\alpha^2)$ noise nodes (not in any clique) ...
- and $O(\alpha^2)$ noise edges, provided ...
- at most $O(\alpha)$ noise edges incident upon any node.
- These bounds are the best possible up to constants.

Random noise

- There may be as many as $O(\alpha^2)$ noise nodes.
- There may be as many as $\sqrt{\alpha}$ cliques.
- Except for clique edges, each edge inserted independently with probability p.

Proof technique

- Proof requires construction of S, the KKT multiplier of the constraint that X is positive semidefinite, satisfying linear constraints.
- Establishing that S is PSD involves norm bounds for its off-diagonal blocks.
- Thus, as in earlier theorems, proof boils down to finding an off-diagonal block (a matrix) satisfying certain linear constraints whose norm is not too large.

Our approach to finding the dual

- We parametrize the unknown matrix with a number of parameters exactly equal to the number of constraints. This yields a square system of linear equations with some noise present in the coefficients and right-hand side.
- We show that the solution to this linear system is a perturbation of an easy-to-analyze (diagonal plus rank-one) linear system.
- Finally, we use Geman to analyze the norm of the perturbation and claim that the solution to the perturbed system is similar to the solution of the easy-to-analyze system.

Conclusions and open questions

- Convex relaxation can find a clique or biclique in a graph that contains the clique and biclique plus many diversionary edges.
- If the diversionary edges are placed at random, then the algorithm can tolerate many more of them.
- Analogous result for clustering problem.
- Would be interesting to extend the technique to other information retrieval problems, e.g., nonnegative matrix factorization.
- Efficient and accurate solvers needed.