Some Recent Results on Discrete and Nonconvex Quadratic Programming

Jiming Peng

Based on works joint with
H. Mittelmann, ASU
X. Li and R. Yang, IESE, UI

June 8, 2009

Quadratic Optimization

$$\begin{aligned} \min \quad & x^T Qx + q^T x \\ s.t. \quad & Ax = b \\ & x \geq 0, \quad x \in \{0,1\}^n or \{-1,1\}^n. \end{aligned}$$

- One of the basic optimization models widely used in many applications from experiment design and portfolio selection;

Quadratic Optimization

$$\begin{aligned} \min \quad & x^T Qx + q^T x \\ s.t. \quad & Ax = b \\ & x \geq 0, \quad x \in \{0,1\}^n or \{-1,1\}^n. \end{aligned}$$

- One of the basic optimization models widely used in many applications from experiment design and portfolio selection;
- One of the corner stones in the development of optimization theory and methodologies;

Quadratic Optimization

$$\begin{aligned} \min \quad & x^T Qx + q^T x \\ s.t. \quad & Ax = b \\ & x \geq 0, \quad x \in \{0,1\}^n or \{-1,1\}^n. \end{aligned}$$

- One of the basic optimization models widely used in many applications from experiment design and portfolio selection;
- One of the corner stones in the development of optimization theory and methodologies;
- In general, it is NP-hard. For some special cases, even getting a good approximation is hard too.

Semidefinite Programming

$$\begin{aligned} \min \quad & \mathrm{Tr}(CX) \\ s.t. \quad & \mathrm{Tr}(A_i X) = b_i, \quad i = 1, \cdots, m \\ & X \succeq 0. \end{aligned}$$

- First appeared in [Bellman & Fan, 1963];

Semidefinite Programming

$$\begin{aligned} \min \quad & \mathrm{Tr}(CX) \\ s.t. \quad & \mathrm{Tr}(A_i X) = b_i, \quad i = 1, \cdots, m \\ & X \succeq 0. \end{aligned}$$

- First appeared in [Bellman & Fan, 1963];
- Fast development in 1990s:

Semidefinite Programming

$$
\begin{aligned}
\min \quad & \mathrm{Tr}(CX) \\
s.t. \quad & \mathrm{Tr}(A_i X) = b_i, \quad i = 1, \cdots, m \\
& X \succeq 0.
\end{aligned}
$$

- First appeared in [Bellman & Fan, 1963];
- Fast development in 1990s:
  - Many applications from various disciplines [Boyd & Vandenberhe, 2004];

Semidefinite Programming

$$
\begin{aligned}
\min \quad & \mathrm{Tr}(CX) \\
s.t. \quad & \mathrm{Tr}(A_i X) = b_i, \quad i = 1, \cdots, m \\
& X \succeq 0.
\end{aligned}
$$

- First appeared in [Bellman & Fan, 1963];
- Fast development in 1990s:
  - Many applications from various disciplines [Boyd & Vandenberhe, 2004];
  - Polynomially solvable by interior-point methods [Nemirovskii & Nesterov, 1994], Handbook of SDP edited by Wolkowicz et'al, 2002;

Semidefinite Programming

$$
\begin{aligned}
\min \quad & \mathrm{Tr}(CX) \\
s.t. \quad & \mathrm{Tr}(A_i X) = b_i, \quad i = 1, \cdots, m \\
& X \succeq 0.
\end{aligned}
$$

- First appeared in [Bellman & Fan, 1963];
- Fast development in 1990s:
  - Many applications from various disciplines [Boyd & Vandenberhe, 2004];
  - Polynomially solvable by interior-point methods [Nemirovskii & Nesterov, 1994], Handbook of SDP edited by Wolkowicz et'al, 2002;
- Called new "LP" in this century.

Transformation from QP to SDP

SDP based approach for QP has been well-studied in both the continuous and discrete optimization communities since 1990s:

- Early pioneers: Lovaz (1979), Lovaz & Schrijver (1991), N. Shor (1987), Alizedah (1991);

Transformation from QP to SDP

SDP based approach for QP has been well-studied in both the continuous and discrete optimization communities since 1990s:

- Early pioneers: Lovaz (1979), Lovaz & Schrijver (1991), N. Shor (1987), Alizedah (1991);
- The well-known SDP based method for max-cut by Goemans and Williamson (1994), Nesterov (1998);

Transformation from QP to SDP

SDP based approach for QP has been well-studied in both the continuous and discrete optimization communities since 1990s:

- Early pioneers: Lovaz (1979), Lovaz & Schrijver (1991), N. Shor (1987), Alizedah (1991);
- The well-known SDP based method for max-cut by Goemans and Williamson (1994), Nesterov (1998);
- The SDP relaxation is based on the relaxation of the gram matrix $X = xx^T$ (or lifting techniques):

$$X \succeq 0, \operatorname{diag}(X) = 1 \quad \text{if } x \in \{-1, 1\}^n;$$

$$X \succeq 0 \text{ or } \begin{pmatrix} 1 & x^T \\ x & X \end{pmatrix} \succeq 0, X \geq 0 \quad \text{if } x \in \{0, 1\}^n,$$

SDP for QP

- Hundreds of papers have been published;

SDP for QP

- Hundreds of papers have been published;
  - Most of these papers follow the Goemans-Williamson approach or the Lovas-Schrijver hierarchy.

SDP for QP

- Hundreds of papers have been published;
  - Most of these papers follow the Goemans-Williamson approach or the Lovas-Schrijver hierarchy.
  - Excellent theoretical results, but scalability of the SDP approach is a concern;

SDP for QP

- Hundreds of papers have been published;
  - Most of these papers follow the Goemans-Williamson approach or the Lovas-Schrijver hierarchy.
  - Excellent theoretical results, but scalability of the SDP approach is a concern;
  - The SDP relaxation of 0-1 binary QP involves n^2 linear constraints, it is too expensive to solve such an SDP!

SDP for QP

- Hundreds of papers have been published;
  - Most of these papers follow the Goemans-Williamson approach or the Lovas-Schrijver hierarchy.
  - Excellent theoretical results, but scalability of the SDP approach is a concern;
  - The SDP relaxation of 0-1 binary QP involves n^2 linear constraints, it is too expensive to solve such an SDP!
- Some NP-hard problems can be solved relative easily if the input data has special structure:

SDP for QP

- Hundreds of papers have been published;
  - Most of these papers follow the Goemans-Williamson approach or the Lovas-Schrijver hierarchy.
  - Excellent theoretical results, but scalability of the SDP approach is a concern;
  - The SDP relaxation of 0-1 binary QP involves n^2 linear constraints, it is too expensive to solve such an SDP!
- Some NP-hard problems can be solved relative easily if the input data has special structure:
  - Many elegant results from classical graph theory and discrete optimization;

SDP for QP

- Hundreds of papers have been published;
  - Most of these papers follow the Goemans-Williamson approach or the Lovas-Schrijver hierarchy.
  - Excellent theoretical results, but scalability of the SDP approach is a concern;
  - The SDP relaxation of 0-1 binary QP involves n^2 linear constraints, it is too expensive to solve such an SDP!
- Some NP-hard problems can be solved relative easily if the input data has special structure:
  - Many elegant results from classical graph theory and discrete optimization;
  - Recent exciting developments on compressed sensing Donoho (2006), Candés and Tao (2006)!

Quadratic Assignment Problem

$$\min_{X \in \Pi} \mathrm{Tr}(AXBX^T)$$

where Π is the set of permutation matrices.

- Numerous applications from facility locations and communications (QAP-Library), TSP;

Quadratic Assignment Problem

$$\min_{X \in \Pi} \mathrm{Tr}(AXBX^T)$$

where Π is the set of permutation matrices.

- Numerous applications from facility locations and communications (QAP-Library), TSP;
- Recognized as one of the most challenging problems;

Quadratic Assignment Problem

$$\min_{X \in \Pi} \mathrm{Tr}(AXBX^T)$$

where Π is the set of permutation matrices.

- Numerous applications from facility locations and communications (QAP-Library), TSP;
- Recognized as one of the most challenging problems;
  - Optimal solution have been reported only for small-scale problems (Brixius & Anstreicher 01);

Quadratic Assignment Problem

$$\min_{X \in \Pi} \mathrm{Tr}(AXBX^T)$$

where Π is the set of permutation matrices.

- Numerous applications from facility locations and communications (QAP-Library), TSP;
- Recognized as one of the most challenging problems;
  - Optimal solution have been reported only for small-scale problems (Brixius & Anstreicher 01);
  - Even computing a good lower bound for problems of size $n = 30$ is too expensive (the classical lifting technique leads to $\mathcal{O}(n^4)$ constraints) Hahn et'al 2007;

Existing Relaxations for QAP

- Step 1: Vectorization of X, $z = vec(X)$.

Existing Relaxations for QAP

- Step 1: Vectorization of X, $z = vec(X)$.
- Step 2: Use the Kronecker product to rewrite the problem as

$$\min_{z=vec(X)} z^T(A \otimes B)z.$$

Existing Relaxations for QAP

- Step 1: Vectorization of X, $z = vec(X)$.
- Step 2: Use the Kronecker product to rewrite the problem as

$$\min_{z=vec(X)} z^T(A \otimes B)z.$$

- Step 3: use the gram matrix $Z = zz^T$ and add constraints on Z.

Existing Relaxations for QAP

- Step 1: Vectorization of X, $z = vec(X)$.
- Step 2: Use the Kronecker product to rewrite the problem as

$$\min_{z=vec(X)} z^T (A \otimes B) z.$$

- Step 3: use the gram matrix $Z = zz^T$ and add constraints on Z.
- Many existing works on how to derive/solve these expensive relaxations, but only works for small-scale problems.

Some Observations

- Fact 1: X is a permutation matrix, not only a matrix with binary elements!

Some Observations

- Fact 1: X is a permutation matrix, not only a matrix with binary elements!
- Fact 2: The spectrum of a matrix is invariant under permutations;

Some Observations

- Fact 1: X is a permutation matrix, not only a matrix with binary elements!
- Fact 2: The spectrum of a matrix is invariant under permutations;
- Fact 3: There exist certain functions/mappings that are invariant for permutations: i.e., $f(Xv) = f(v)$ or $f(MX) = f(M), f(XM) = Xf(M)$;

Some Observations

- Fact 1: X is a permutation matrix, not only a matrix with binary elements!
- Fact 2: The spectrum of a matrix is invariant under permutations;
- Fact 3: There exist certain functions/mappings that are invariant for permutations: i.e., $f(Xv) = f(v)$ or $f(MX) = f(M), f(XM) = Xf(M)$;
- Fact 4: In many applications, the matrices A or B are associated with specific graphes, i.e., B is the Hamming distance matrix of a hypercube or the Manhattan distance matrix of a rectangular grids;

Some Observations

- Fact 1: X is a permutation matrix, not only a matrix with binary elements!
- Fact 2: The spectrum of a matrix is invariant under permutations;
- Fact 3: There exist certain functions/mappings that are invariant for permutations: i.e., $f(Xv) = f(v)$ or $f(MX) = f(M), f(XM) = Xf(M)$;
- Fact 4: In many applications, the matrices A or B are associated with specific graphes, i.e., B is the Hamming distance matrix of a hypercube or the Manhattan distance matrix of a rectangular grids;
- Fact 5: The matrices A and B have nonnegative elements, thus dominated by its first principal component.

Proposed Relaxation

- Observation: If $B \succeq 0$, then we have

$$Y = XBX^T, \quad Y \succeq 0.$$

Proposed Relaxation

- Observation: If $B \succeq 0$, then we have

$$Y = XBX^T, \quad Y \succeq 0.$$

- Question: What to do if B is not positive semidefinite?

Proposed Relaxation

- Observation: If $B \succeq 0$, then we have

$$Y = XBX^T, \quad Y \succeq 0.$$

- Question: What to do if B is not positive semidefinite?
- Answer: Positive semidefinite matrix splitting (PSD splitting) $B = B^+ - B^-$ with $B^+, B^- \succeq 0$.

Proposed Relaxation

- Observation: If $B \succeq 0$, then we have

$$Y = XBX^T, \quad Y \succeq 0.$$

- Question: What to do if B is not positive semidefinite?
- Answer: Positive semidefinite matrix splitting (PSD splitting) $B = B^+ - B^-$ with $B^+, B^- \succeq 0$.
- Splitting schemes:

Proposed Relaxation

- Observation: If $B \succeq 0$, then we have

$$Y = XBX^T, \quad Y \succeq 0.$$

- Question: What to do if B is not positive semidefinite?
- Answer: Positive semidefinite matrix splitting (PSD splitting) $B = B^+ - B^-$ with $B^+, B^- \succeq 0$.
- Splitting schemes:
  - Orthogonal PSD splitting: based on the SVD;

Proposed Relaxation

- Observation: If $B \succeq 0$, then we have

$$Y = XBX^T, \quad Y \succeq 0.$$

- Question: What to do if B is not positive semidefinite?
- Answer: Positive semidefinite matrix splitting (PSD splitting) $B = B^+ - B^-$ with $B^+, B^- \succeq 0$.
- Splitting schemes:
  - Orthogonal PSD splitting: based on the SVD;
  - Use the Laplacian operator: $D = \mathrm{diag}\,(\sum(B)), B = D - (D - B)$;

Proposed Relaxation

- Observation: If $B \succeq 0$, then we have

$$Y = XBX^T, \quad Y \succeq 0.$$

- Question: What to do if B is not positive semidefinite?
- Answer: Positive semidefinite matrix splitting (PSD splitting) $B = B^+ - B^-$ with $B^+, B^- \succeq 0$.
- Splitting schemes:
  - Orthogonal PSD splitting: based on the SVD;
  - Use the Laplacian operator: $D = \mathrm{diag}\left(\sum(B)\right), B = D - (D - B)$;
  - Specific splitting $B = \alpha E - B^-$, where E is the all-1 matrix.

Special Splitting Examples

- The Hamming distance matrix with a binary codebook

$$\mathcal{C} = \{c_1 = `00', c_2 = `01', c_3 = `10', c_4 = `11'\}.$$

$$B = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}, E - B = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \succeq 0.$$

Special Splitting Examples

- The Hamming distance matrix with a binary codebook

$$\mathcal{C} = \{c_1 = \text{`}00', c_2 = \text{`}01', c_3 = \text{`}10', c_4 = \text{`}11'\}.$$

$$B = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}, E - B = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \succeq 0.$$

- Manhattan distance matrix from facility location

$$B = [b_{ij}] \in \Re^{n\times n}, \quad b_{ij} = |i - j|, \quad \frac{n-1}{2} E - B \succeq 0. \qquad (1)$$

Construction of Valid Cut

Given a matrix Y and a function f, we define a mapping

$$F(Y) = [f(Y_1); f(Y_2); \cdots ; f(Y_n)].$$

Theorem: Suppose that $F(\cdot)$ is a mapping defined with a symmetric function $f(\cdot)$ and X is a permutation matrix. Then we have

$$F(XBX^T) = XF(B).$$

Constructing Cut: Choose f to be convex, and relax it to

$$F(Y) = F(XBX^T) \leq XF(B).$$

A Sample Relaxation

Let (B^+, B^-) be a PSD splitting of B and $Y^+ = XB^+X^T$, $Y^- = XB^-X^T$. Using symmetric mappings $\max, \min, \mathcal{L}_1$ and $\mathcal{L}_2$, we derive

$$
\begin{array}{ll}
\min & \mathrm{Tr}(A(Y^+ - Y^-)) \\
s.t. & Y^+ - XB^+X^T \succeq 0, \quad Y^- - XB^-X^T \succeq 0; \\
& \mathrm{diag}\,(Y^+) = X\mathrm{diag}\,(B^+), \quad Y^+e = XB^+e; \\
& \mathrm{diag}\,(Y^-) = X\mathrm{diag}\,(B^-), \quad Y^-e = XB^-e; \\
& (X\min(B^+))_i \le y^+_{i,j} \le (X\max(B^+))_i, \qquad \forall i \neq j; \\
& (X\min(B^-))_i \le y^-_{i,j} \le (X\max(B^-))_i, \qquad \forall i \neq j; \\
& \mathcal{L}_2(Y^+) \le X\mathcal{L}_2(B^+), \quad \mathcal{L}_2(Y^-) \le X\mathcal{L}_2(B^-); \\
& X \ge 0, \quad Xe = X^Te = e.
\end{array}
$$

Comments on New SDP Relaxation of QAPs

- A substantial reduction in model complexity;

Comments on New SDP Relaxation of QAPs

- A substantial reduction in model complexity;
- Not dominated by existing expensive relaxations;

Comments on New SDP Relaxation of QAPs

- A substantial reduction in model complexity;
- Not dominated by existing expensive relaxations;
- Easy integration of other cheap relaxations;

Comments on New SDP Relaxation of QAPs

- A substantial reduction in model complexity;
- Not dominated by existing expensive relaxations;
- Easy integration of other cheap relaxations;
- Strong bounds for large scale instances have been obtained;

Comments on New SDP Relaxation of QAPs

- A substantial reduction in model complexity;
- Not dominated by existing expensive relaxations;
- Easy integration of other cheap relaxations;
- Strong bounds for large scale instances have been obtained;
- Future directions:

Comments on New SDP Relaxation of QAPs

- A substantial reduction in model complexity;
- Not dominated by existing expensive relaxations;
- Easy integration of other cheap relaxations;
- Strong bounds for large scale instances have been obtained;
- Future directions:
  - Using dimension reduction techniques to get a good approximation;

Comments on New SDP Relaxation of QAPs

- A substantial reduction in model complexity;
- Not dominated by existing expensive relaxations;
- Easy integration of other cheap relaxations;
- Strong bounds for large scale instances have been obtained;
- Future directions:
  - Using dimension reduction techniques to get a good approximation;
  - Develop new solving techniques for these new relaxation models.

0-1 Binary QP

We consider the following binary QP

$$(\mathrm{StQP}) \quad \max \quad x^T Qx$$
$$s.t. \quad \sum_{i=1}^{n} x_i = k, \quad x \in \{0,1\}^n.$$

- Applications: the densest k-subgraph, feature selection in learning;

0-1 Binary QP

We consider the following binary QP

$$(\mathrm{StQP}) \quad \max \quad x^T Qx$$
$$s.t. \quad \sum_{i=1}^{n} x_i = k, \quad x \in \{0,1\}^n.$$

- Applications: the densest k-subgraph, feature selection in learning;
- Extra constraints can be added;

0-1 Binary QP

We consider the following binary QP

$$(\mathrm{StQP}) \quad \max \quad x^T Qx$$
$$s.t. \quad \sum_{i=1}^{n} x_i = k, \quad x \in \{0,1\}^n.$$

- Applications: the densest k-subgraph, feature selection in learning;
- Extra constraints can be added;
- NP-hard, even a good approximation is hard unless P=NP. PTAS have been ruled out recently(S. Khot, SIAM J. Computing, 2006, Best paper award in SIAM).

Existing Approaches

- Classical approach based on graph theory and discrete optimization;

Existing Approaches

- Classical approach based on graph theory and discrete optimization;
- SDP relaxation by Lovaz and Schrijver (2001):

$$
\begin{aligned}
\max \quad & \mathrm{Tr}(QX) \\
s.t. \quad & Xe = k * \mathrm{diag}(X) \\
& \sum_{i=1}^{n} x_{ii} = k \\
& X \succeq 0, \quad X \geq 0.
\end{aligned}
$$

Issues in Existing Approaches

- The conventional SDP relaxation uses only the binary structure of the variables;
  The structure of the graph has been ignored!

Issues in Existing Approaches

- The conventional SDP relaxation uses only the binary structure of the variables;
  The structure of the graph has been ignored!
- Only applicable to medium size problem (Helmberg & Rendl);

Issues in Existing Approaches

- The conventional SDP relaxation uses only the binary structure of the variables;
  The structure of the graph has been ignored!
- Only applicable to medium size problem (Helmberg & Rendl);
- Constant approximations have been reported only for special classes of instances. Even most heuristics have a hard time in local search!

Issues in Existing Approaches

- The conventional SDP relaxation uses only the binary structure of the variables;
  The structure of the graph has been ignored!
- Only applicable to medium size problem (Helmberg & Rendl);
- Constant approximations have been reported only for special classes of instances. Even most heuristics have a hard time in local search!
- Question: What's wrong?

Convex QP: Relaxation or Geometric Embedding: I

We rewrite the problem as QAP:

$$\begin{aligned} \max \quad & x^T(Q - \lambda_{\min}(Q)I)x \\ s.t \quad & \sum_{i=1}^{n} x_i = k, x_i \in \{0,1\}. \end{aligned} \tag{2}$$

$\lambda_{\min}(Q)$ denotes the minimal eigenvalue of Q.

- Relaxation for a cheap bound?

Convex QP: Relaxation or Geometric Embedding:II

- Let $\bar{Q} = (Q - \lambda_{\min}(Q)I) \succeq 0$. We can interpret each element of $\bar{Q}$ as the inner product of two data points in a data set on the surface of a unit sphere in a certain dimensional space;

$$\mathcal{V} = \{v_i : \|v_i\|^2 = -\lambda_{\min}(Q),\ i = 1, \cdots, v_n\}, \bar{Q}_{ij} = v_i^T v_j.$$

Convex QP: Relaxation or Geometric Embedding:II

- Let $\bar{Q} = (Q - \lambda_{\min}(Q)I) \succeq 0$. We can interpret each element of $\bar{Q}$ as the inner product of two data points in a data set on the surface of a unit sphere in a certain dimensional space;

$$\mathcal{V} = \{v_i : \|v_i\|^2 = -\lambda_{\min}(Q),\ i = 1, \cdots, v_n\}, \bar{Q}_{ij} = v_i^T v_j.$$

- Geometric Embedding: Consider a specific clustering problem of finding a single cluster of fixed size whose within cluster sum of squared distances is minimal:

$$\min_{|\mathcal{V}_1|=k} \sum_{v \in \mathcal{V}_1} \|v - \frac{\sum_{v \in \mathcal{V}_1} v}{k}\|^2. \tag{3}$$

Here $|\mathcal{V}_1|$ denotes the cardinality of the subset $\mathcal{V}_1$.

Approximation to A Simple Clustering Problem

Theorem: Problem (StQP) and the clustering problem (3) share the same optimal solution set.
Problem (3) is equivalent to

$$\min_{c} \min_{|\mathcal{V}_1|=k} \sum_{v\in\mathcal{V}_1} \|v-c\|^2. \tag{4}$$

Some simple and effective approximation algorithms/heuristics:

- Use an iterative scheme as in the classical K-means clustering that subsequently update c and cluster;

Approximation to A Simple Clustering Problem

Theorem: Problem (StQP) and the clustering problem (3) share the same optimal solution set.
Problem (3) is equivalent to

$$\min_{c} \min_{|\mathcal{V}_1|=k} \sum_{v \in \mathcal{V}_1} \|v - c\|^2. \tag{4}$$

Some simple and effective approximation algorithms/heuristics:

- Use an iterative scheme as in the classical K-means clustering that subsequently update c and cluster;
- Try different starting points and select the best one as final output (provable 2-approximation);

Approximation to A Simple Clustering Problem

Theorem: Problem (StQP) and the clustering problem (3) share the same optimal solution set.
Problem (3) is equivalent to

$$\min_{c} \min_{|\mathcal{V}_1|=k} \sum_{v \in \mathcal{V}_1} \|v - c\|^2. \tag{4}$$

Some simple and effective approximation algorithms/heuristics:

- Use an iterative scheme as in the classical K-means clustering that subsequently update c and cluster;
- Try different starting points and select the best one as final output (provable 2-approximation);
- Use the first eigenvector of Q to find the cluster...

Complexity Analysis

We beak down the whole process into three steps:

- Use the eigenvalue decomposition of the input matrix Q to construct the data set $\mathcal{V}$ ($O(n^3)$);

Complexity Analysis

We beak down the whole process into three steps:

- Use the eigenvalue decomposition of the input matrix Q to construct the data set $\mathcal{V}$ $(O(n^3))$;
- Find the $k-1$ closest data points for any given data point in $\mathcal{V}$ $(O(n^2 k \log n))$;

Complexity Analysis

We beak down the whole process into three steps:

- Use the eigenvalue decomposition of the input matrix Q to construct the data set $\mathcal{V}$ ($O(n^3)$);
- Find the $k-1$ closest data points for any given data point in $\mathcal{V}$ ($O(n^2 k \log n)$);
- Perform a local search ($O(n^2)$);

Complexity Analysis

We beak down the whole process into three steps:

- Use the eigenvalue decomposition of the input matrix Q to construct the data set $\mathcal{V}$ ($O(n^3)$);
- Find the $k-1$ closest data points for any given data point in $\mathcal{V}$ ($O(n^2 k \log n)$);
- Perform a local search ($O(n^2)$);
- In total $O(n^3 + n^2 k \log n + n^2)$.

Comments

- A new role of convex QP relaxation;

Comments

- A new role of convex QP relaxation;
- A bridge to simple and effective clustering model;

Comments

- A new role of convex QP relaxation;
- A bridge to simple and effective clustering model;
- Future direction: Extensions to other binary QPs such as $x^T Qx + q^T x$; and faster approximation algorithms based on the spectrum of Q.

Sparse Solutions to Linear Equation System

$$\begin{array}{ll} \min & \|x\|_0 \\ s.t. & Ax = b, x \geq 0. \end{array}$$

Replacing the objective in the above model by $\|x\|_1$, we end up with an LP problem. As proved by Candés and Tao (2006), Donoho (2006):

Theorem: If the input data matrix A follows certain distribution and there exists a sparse solution, then the solution from the LP problem is also optimal for the original L_0 optimization problem with a high probability.

From Linear Equation to QP

The LP problem can be equivalently stated as

$$\begin{array}{ll} \min & \|Ax-b\|^2 \\ s.t. & \sum_{i=1}^{n} x_i = 1, x \geq 0. \end{array}$$

Let us consider a generalized case:

$$\begin{array}{lll} \min & x^T Q x + q^T x & (5) \\ s.t. & \sum_{i=1}^{n} x_i = 1, x \geq 0. & (6) \end{array}$$

Question: Under what conditions, the above problem admits sparse solutions?

Checking the Co-positivity of Matrices

Question: Given a matrix Q, is there a nonnegative vector x such that $x^T Qx < 0$?
Mathematically, we can address the above problem by solving the following problem:

$$\begin{array}{ll} \min & x^T Qx \\ s.t. & \sum_{i=1}^{n} x_i = 1, x \geq 0. \end{array} \quad (7)$$

Such a model arise also from learning and feature selection. The problem has been proved to be NP-hard (Murty and Kabadi, 1987). It is also called standard quadratic programming problem in the literature.

A Simple SDP Relaxation

$$\begin{aligned} \min \quad & \mathrm{Tr}(QX) \\ s.t. \quad & \sum_{i,j=1}^{n} x_{ij} = 1, X \succeq 0, X \geq 0. \end{aligned} \tag{8}$$

Observation: My simple matlab code always gives me rank-one solution, which implies the SDP relaxation solved the original problem precisely!

Checking the Co-positivity of Random Matrices

We have proved the following result.

Theorem: If the matrix Q is random following certain distributions, then with a high probability that the optimal solution of problem (7) is sparse and it can be found in polynomial time.

Comments on Probabilistic Analysis of QP

- A new analysis for a special class of QP;

Comments on Probabilistic Analysis of QP

- A new analysis for a special class of QP;
- Verified numerically by using SDP relaxation;

Comments on Probabilistic Analysis of QP

- A new analysis for a special class of QP;
- Verified numerically by using SDP relaxation;
- Future directions:

Comments on Probabilistic Analysis of QP

- A new analysis for a special class of QP;
- Verified numerically by using SDP relaxation;
- Future directions:
  - Possible extension to other NP hard problems;

Comments on Probabilistic Analysis of QP

- A new analysis for a special class of QP;
- Verified numerically by using SDP relaxation;
- Future directions:
  - Possible extension to other NP hard problems;
  - For SDP, under what conditions, the SDP problem has rank 1 solution? How can we use this information to develop more effective resolution techniques?

Questions

For reference, please refer to my personal web site
https://netfiles.uiuc.edu/pengj/www/
or contact pengj@illinois.edu