Computers & Operations Research 37 (2010) 1027–1032

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

An efficient continuation method for quadratic assignment problems ☆

Yong Xia

LMIB of the Ministry of Education, School of Mathematics and System Sciences, Beihang University, Beijing 100191, PR China

ARTICLE INFO

Available online 15 September 2009

Keywords:
Quadratic assignment problem
Lagrangian smoothing
Frank–Wolfe algorithm
KKT
Heuristic

ABSTRACT

In this article, we propose a Lagrangian smoothing algorithm for quadratic assignment problems, where the continuation subproblems are solved by the truncated Frank–Wolfe algorithm. We establish practical stopping criteria and show the algorithm finitely terminates at a KKT point of a continuation subproblem. The quality of the returned solution is studied in detail. Finally, limited numerical results are provided.

1. Introduction

Quadratic assignment problem (QAP) is one of the great challenges in combinatorial optimization. It is known to be NP-hard. An ε-solution is NP-hard too. For comprehensive surveys of QAPs, we refer to [1,8,9,11,14]. The formulation of QAP can be written as

$$\begin{aligned} \min \quad & f(X) = \text{trace}\,(AXBX^T) \\ \text{s.t.} \quad & X \in \Pi_n, \end{aligned} \tag{1.1}$$

where A and B are $n \times n$ matrices, 'trace' denotes the sum of all diagonal elements, and Π_n is the set of $n \times n$ permutation matrices, i.e., $\Pi_n = \{X = (X_{ij}) \in \mathbb{R}^{n\times n} : Xe = X^T e = e, X_{ij} \in \{0,1\}\}$, where e is a vector with all components equal to one. Denote the continuous relaxation of Π_n by $P_n = \{X = (X_{ij}) \in \mathbb{R}^{n\times n} : Xe = X^T e = e, X_{ij} \geq 0\}$, which is also the convex hull of Π_n, i.e., $P_n = \text{conv}\{\Pi_n\}$.

In most practical applications, the QAP models are symmetric, i.e., both A and B are symmetric. Furthermore, if only one of these matrices is symmetric (say A), we can transform it to a QAP where both matrices are symmetric since

$$\text{trace}\,(AXB^TX^T) = \text{trace}\left(AX\frac{B^T+B}{2}X^T\right).$$

Therefore, we make the symmetric assumption throughout this article.

Continuation methods are new promising approaches to solving QAP. In [12] the logarithmic smoothing algorithm (LogSA) was applied to solve QAP and limited numerical results were reported. The Lagrangian smoothing algorithm for QAP was firstly devised in [16]. Numerical experiments indicated its high efficiency. In this article, we further study this algorithm. We propose a new version, establish practical stopping criteria and prove the algorithm finitely terminates at a KKT point of a continuation subproblem. We also analyze the quality of the obtained solution.

The article is organized as follows. In Section 2, a new version of Lagrangian smoothing algorithm is proposed. The convergence result is analyzed in Section 3. Implementation details and numerical results are presented in Section 4. The last section makes some concluding remarks.

Notation. For a square matrix A, vec(A) gives the column vector obtained by stacking the columns of A in increasing order of their index. The Kronecker product of matrices A and B is denoted by $A \otimes B$. Let S^n be the set of symmetric matrices of order n, i.e., $S^n = \{X \in \mathbb{R}^{n\times n} : X^T = X\}$. For $A, B \in S^n$, $A \succcurlyeq B$ denotes that $A - B$ is positive semidefinite. $\lambda_{\min}(A)$ denotes the minimal eigenvalue of A. Let C^n be the set of functions with continuous n-th order derivatives. ∇f and $\nabla^2 f$ denote the gradient and Hessian of f, respectively.

2. Lagrangian smoothing algorithm

Basically, global smoothing is to linearly combine the original objective $f(X)$ with an additional function $\Phi(X)$ such that

$$F(X;\mu) = f(X) + \mu \cdot \Phi(X) \tag{2.1}$$

is strictly convex on P_n with respect to X.

Theorem 2.1 (*Bertsekas [3] and Ng [12]*). *Suppose $\Phi : P_n \to \mathbb{R}$ is a C^2 function such that the minimum eigenvalue of $\nabla^2\Phi$ is greater than a positive number ε for all $X \in P_n$. Then there exists a real $M > 0$ such that if $\mu > M$, then $f + \mu \cdot \Phi$ is a strictly convex function on P_n.*

☆ Supported by Chinese NSF Grant 10671199.
E-mail address: dearyxia@gmail.com

0305-0548/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2009.09.002

The logarithmic barrier function

$$\Phi^{Log}(X) = -\sum_{i=1}^{n}\sum_{j=1}^{n}\ln(X_{ij}) - \sum_{i=1}^{n}\sum_{j=1}^{n}\ln(1-X_{ij}) \tag{2.2}$$

was introduced as a smoothing function in [12]. It is well-defined when $X \in (0,1)^{n\times n}$. If any value of X_{ij} is 0 or 1, then $\Phi^{Log}(X) = +\infty$.

Theorem 2.2 (*Ng [12]*). *There exists a real $M > 0$ such that if $\mu \geq M$, then $f + \mu \cdot \Phi^{Log}$ is a strictly convex function on $(0,1)^{n\times n}$.*

Notice that QAP (1.1) is equivalent to

$$\min \quad \text{trace}\,(AXBX^T) \tag{2.3}$$
$$\text{s.t.} \quad \text{trace}\,(XX^T) = n, \tag{2.4}$$
$$X \in P_n. \tag{2.5}$$

Introducing a Lagrangian multiplier for the constraint (2.4), we obtain the corresponding Lagrangian function

$$L(X;\mu_0) = \text{trace}\,(AXBX^T) + \mu_0 \cdot \text{trace}\,(XX^T) - n\mu_0, \tag{2.6}$$

which motivates the following global smoothing function [16]:

$$\Phi^{Lag}(X) = \text{trace}\,(XX^T). \tag{2.7}$$

Theorem 2.3. *Let the $n \times (n-1)$ matrix V be such that*

$$V^Te = 0, \quad V^TV = I_{n-1}.$$

Define $\hat{A} = V^TAV$, $\hat{B} = V^TBV$ and

$$\hat{\lambda}_{\min} = \min\{\lambda_{\min}(\hat{B})\lambda_{\max}(\hat{A}), \lambda_{\max}(\hat{B})\lambda_{\min}(\hat{A}), \lambda_{\min}(\hat{B})\lambda_{\min}(\hat{A}), \lambda_{\max}(\hat{B})\lambda_{\max}(\hat{A})\}. \tag{2.8}$$

Then $f + \mu_0 \cdot \Phi^{Lag}$ is a strictly convex function on P_n for any $\mu_0 > -\hat{\lambda}_{\min}$.

Proof. Let

$$Q = \left[\frac{e}{\sqrt{n}} : V\right] \in O_n := \{X \in \mathbb{R}^{n\times n} : XX^T = I_n\}.$$

We introduce a well-known result due to Hadley et al. [10].

Lemma 2.1 (*Hadley et al. [10]*). *Let X be $n \times n$ and Y be $(n-1)\times(n-1)$. Suppose that X and Y satisfy*

$$X = Q\begin{bmatrix} 1 & 0 \\ 0 & Y \end{bmatrix}Q^T = \frac{1}{n}ee^T + VYV^T. \tag{2.9}$$

Then

$$Xe = X^Te = e,$$
$$X \in O_n \Longleftrightarrow Y \in O_{n-1}.$$

Taking Eq. (2.9) into $f(X)$ and $\Phi^{Lag}(X)$, respectively, we have

$$\text{trace}\,(AXBX^T) = \text{trace}\,(\hat{A}Y\hat{B}Y^T) + \frac{2}{n}\text{trace}\,(V^TAee^TBVY^T) + \frac{1}{n^2}(e^TAe)(e^TBe),$$
$$\text{trace}\,(XX^T) = \text{trace}\,(YY^T) + 1.$$

Consequently, the reduced Hessian of $f + \mu_0 \cdot \Phi^{Lag}$ is $2\hat{B}\otimes\hat{A} + 2\mu_0 I$. It is positive definite if and only if $\mu_0 > -\lambda_{\min}(\hat{B}\otimes\hat{A}) = -\hat{\lambda}_{\min}$, where $\hat{\lambda}_{\min}$ is defined in (2.8). □

Linearly combining $f + \mu_0 \cdot \Phi^{Lag}$ with the exact penalty function (see Theorem 2.4):

$$P(X;\mu_\infty) = \text{trace}\,(AXBX^T) + \mu_\infty \cdot \text{trace}\,(XX^T), \tag{2.10}$$

we obtain a sequence of functions

$$PL(X;\mu) = \text{trace}\,(AXBX^T + \mu XX^T), \tag{2.11}$$

and define

$$H(X;\mu) = AXBX^T + \mu XX^T. \tag{2.12}$$

Then we solve the following parametric optimization problem, denoted by $QAP(\mu)$:

$$\min \quad PL(X;\mu)$$
$$\text{s.t.} \quad X \in P_n, \tag{2.13}$$

for a decreased sequence $\{\mu\} \subseteq [\mu_\infty, \mu_0]$.

Theorem 2.4. *The optimal function values of QAP (1.1) and Problem (2.13) are equal for all $\mu = \mu_\infty \leq -\hat{\lambda}_{\max}$, where*

$$\hat{\lambda}_{\max} = \max\{\lambda_{\min}(\hat{B})\lambda_{\max}(\hat{A}), \lambda_{\max}(\hat{B})\lambda_{\min}(\hat{A}), \lambda_{\min}(\hat{B})\lambda_{\min}(\hat{A}), \lambda_{\max}(\hat{B})\lambda_{\max}(\hat{A})\}. \tag{2.14}$$

Proof. Since $P_n = \text{conv}\{\Pi_n\}$ and Π_n is the extreme point set of P_n, it is sufficient to show that $PL(X;\mu)$ is concave on P_n for all $\mu \leq -\hat{\lambda}_{\max}$. Similarly to the proof of Theorem 2.3, we conclude that the reduced Hessian of $PL(X;\mu)$ is $2\hat{B}\otimes\hat{A} + 2\mu I$ and it is negative semidefinite if and only if $\mu \leq -\lambda_{\max}(\hat{B}\otimes\hat{A}) = -\hat{\lambda}_{\max}$, where $\hat{\lambda}_{\max}$ is defined in (2.14). □

Given any parameter μ, we use the canonical Frank–Wolfe algorithm [5] to solve the subproblem (2.13). It approximates the objective function with its first order Taylor expansion at any given iteration point X_k, resulting in the linear programming subproblem (omitting the constant terms)

$$\min \quad \text{trace}\,(\nabla_x H(X_k;\mu)\cdot X^T)$$
$$\text{s.t.} \quad X \in P_n, \tag{2.15}$$

where $\nabla_x H(X_k;\mu) \in \mathbb{R}^{n\times n}$ denotes the gradient of $H(X;\mu)$ with respect to X at X_k, i.e.,

$$\nabla_x H(X_k;\mu) = 2AX_kB + 2\mu X_k.$$

Furthermore, Problem (2.15) is a typical linear assignment problem (LAP) and can be solved in $O(n^3)$ time, for example, using Hungarian method.

The optimal solution of the LAP (2.15), X_k^*, is used to construct the descent search direction $D_k = X_k^* - X_k$. A line search

$$\alpha^* = \arg\min_{\alpha\in[0,1]} PL(X_k + \alpha D_k;\mu) \tag{2.16}$$

furnishes the next iteration

$$X_{k+1} = X_k + (1-\beta+\beta\alpha^*)\alpha^* D_k, \tag{2.17}$$

where $\beta \in [0,1]$ is a fixed parameter and the process is repeated. Generally, the new step length satisfies $(1-\beta+\beta\alpha^*)\alpha^* \leq \alpha^*$, where the equality holds when $\beta = 0$ or $\alpha^* = 1$.

It is easy to verify that the point sequence $\{X_k\}$ generated by the above Frank–Wolfe algorithm converges to X^*, a KKT point of (2.13). But the convergence is slow and hence it is quite time-consuming to obtain X^*. This phenomenon motivates us to approximate X^* using truncated Frank–Wolfe algorithm, which only generates the first several iterative points.

The above process is repeated for a decreased sequence $\{\mu_i\}$, the so-called outer iteration.

To introduce practical stopping criteria, we need some lemmas.

Lemma 2.2. *Suppose $X_k \in \Pi_n$ is an optimal solution of the LAP (2.15), i.e., $\text{trace}\,(\nabla_x H(X_k;\mu)(X_k - X_k^*)^T) = 0$. Then it is also a KKT point of QAP(μ) (2.13).*

Proof. The KKT system of $QAP(\mu)$ (2.13) is

$$(\nabla_x H(X;\mu))_{ij} - \lambda_i - \sigma_j \geq 0, \quad i,j = 1,\ldots,n, \tag{2.18}$$

$$X_{ij}((\nabla_x H(X;\mu))_{ij} - \lambda_i - \sigma_j) = 0, \quad i,j = 1,\ldots,n, \tag{2.19}$$

$$X \in P_n. \tag{2.20}$$