Trust Regions and the Quadratic Assignment Problem

by

Stefan E. Karisch

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirements for the degree of
Master of Mathematics
in
Combinatorics and Optimization

Waterloo, Ontario, Canada, 1992

Abstract

The quadratic assignment problem (QAP) belongs to the class of NP-hard combinatorial optimization problems. Although it has been studied extensively over the past 35 years, problems of size $n \geq 15$ still prove to be intractable.

Since good lower bounds are necessary to solve larger problem instances, we discuss the class of eigenvalue bounds for QAP and extend previous results of these continuous optimization approaches for QAP. Since these approaches include trust region techniques and optimization over partial orders, new results for these areas are obtained, as well.

Acknowledgements

My deepest appreciation goes to my supervisor, Henry Wolkowicz, for his guidance, patience and encouragement during my studies. No question or idea did seem to be too trivial, to not to be answered and considered by him. I was happy to be his student.

I would like to acknowledge the Austrian Bundesministerium für Wissenschaft und Forschung, the Exportakademie der Österreichischen Bundeswirtschaftskammer, and the Department of Combinatorics and Optimization for their financial support. I want to thank these institutions for making my studies at the University of Waterloo possible.

I would also like to thank Franz Rendl, Technische Universität Graz, Austria for providing parts of the test data used in this thesis.

But most of all, I wish to thank my wife Helga for her love and continuous encouragement. She gives me the strength to pursue my dreams.

To my parents.

Contents

Chapter 1

Introduction

1.1 The Quadratic Assignment Problem

The *quadratic assignment problem,* denoted *QAP*, was formulated first in 1957 by Koopmans and Beckmann [28] in the following way. Given the set $N = \{1, 2, \ldots, n\}$ and three n by n matrices $A = (a_{ij})$, $B = (b_{kl})$, and $C = (c_{ik})$, find a permutation π of the set N which minimizes

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{\pi(i)\pi(j)} + \sum_{i=1}^{n} c_{i\pi(i)}.$$

Equivalently, see [12], the *trace formulation* of QAP is to find an n by n permutation matrix X which minimizes

$$\textbf{QAP} \qquad \min_{X \in \Pi} f(X) = \operatorname{tr}(AXB^t + C)X^t,$$

where $.^t$ denotes transpose, Π denotes the set of permutation matrices, and tr is the trace.

As an illustrative example we describe an important area of application of QAPs, the backboard wiring problem. The example is due to [42] and is also described in [6] . Suppose an electric circuit should be constructed. Therefore, n modules should be placed on a board. For the n modules there are n locations $1, 2, \ldots, n$ on the board, and a_{ij} denotes the distance between locations i and j. The modules are connected by wires as given in the circuit diagram. The number of wires between modules k and l is b_{kl}. If we assign modules k and l to locations i and j, respectively, the wire length for this allocation will be $a_{ij} b_{\pi(i)\pi(j)}$, where $k = \pi(i)$ and $l = \pi(j)$. Thus, the total wire length for an allocation π is

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{\pi(i)\pi(j)} \tag{1.1.0}$$

and the problem becomes to minimize (1.1.0).

In the framework of location problems the set N describes the set of sites on which plants are to be built. The n by n matrix $A = (a_{ij})$ is the distance between the sites, whereas the connection or flow matrix $B = (b_{kl})$ describes the dependence or the interactions between plants k and l. The linear term can be interpreted as building or running costs for locating plant i at site k.

There are many applications for this allocation problem with spatial interactions. For example, [9] used QAP for hospital planning while in [5] the QAP was applied for the design of typewriter keyboards for minimizing the expected time of typing a text. For other applications, the reader is referred to a recent overview article about QAP by Burkard [6].

Many well known combinatorial optimization problems can be formulated as QAP, e.g. the linear sum assignment problem (LSAP) and also the travelling salesman problem (TSP). Since the TSP is known to be $NP-$hard, QAPs are $NP-$hard, as well. The QAP belongs to the hard core of $NP-$hard problems since even the problem of finding an $\epsilon-$approximation of the optimal solution is $NP-$hard. For the complexity proofs, see [40]. Thus from a worst case point of view, QAP is extremely difficult to solve. But also in practice QAPs of size $n \geq 15$ can prove to be intractable, see e.g. [39].

Current solution techniques employ branch and bound methods and it seems that improving the quality of the bound is an important step towards solving bigger problems.

For the QAP there are three categories of lower bounds. One includes the classical Gilmore-Lawler bound (GLB) [21, 30], and related bounds. These type of bounds seem to be the most widely used. The second category contains bounds which are based on reformulations of the QAP. These bounds generally involve solving a number of LSAPs.

The third type of bounds can be characterized as eigenvalue bounds. These bounds were first introduced for QAP in [12] and exploit the deep structure of the trace formulation of QAP. This type of bound will be discussed in this work.

The approach to obtain these bounds for QAP is to relax the constraints and solve a more tractable problem. For the quadratic part of QAP, we can relax $X \in \Pi$ to X orthogonal. This leads to a lower bound obtained from the eigenvalue decompositions of A and B, see [12], [38]. The linear part can be solved exactly as a linear assignment problem. Further improvements are obtained by 'shifts' or 'reductions' which allow changing A and B in an organized fashion and redefining C accordingly.

Another improvement is the elimination technique which was proposed in [25]. This technique involves the relaxation of $X \in \Pi$ to X orthogonal but with row and column sums 1. This approach yields a smaller dimensioned program which has only the orthogonal constraint.

The starting point of this thesis is the relaxation proposed in [25]. Here, we do not treat the quadratic and linear parts separately. It appears that a large gap between bounds and the true optimal value may occur when the quadratic and linear parts are split up [38]. We further relax this orthogonal constraint $XX^t - I = 0$ to $XX^t - I = N$, where N is negative semi-definite, equivalently

$$XX^t - I \preceq 0,$$

in the partial ordering given by the positive semi-definite matrices. This allows a characterization of a global optimum and surprisingly, under controlled assumptions, the optimum is still orthogonal.

The orthogonal constraint actually consists of $(n^2 + n)/2$ inequality constraints. So the theme of this thesis is the study of quadratic objective functions with more than one quadratic constraint. However, the orthogonal constraint $XX^t \preceq I$ resembles the simple norm constraint $x^t x \leq 1$. The latter constraint is referred to as a 'trust region' constraint, see e.g. [8]. Trust regions allow very efficient algorithms for unconstrained minimization problems. This is based on the fact that a characterization of optimality exists for the minimization of a quadratic function subject to the trust region constraint, independent of whether or not the quadratic objective function is convex or not. Note that problems with only two trust region constraints have been dealt with in [44, 45] and have proven to be very hard.

1.2 Overview

In Chapter 2 we introduce the notation which will be used in this thesis and discuss previous results concerning eigenvalue bounds for QAP. We also describe different techniques to improve these bounds which were published in the literature. The relaxation of QAP which will be studied here is given in Section 2.5.

Chapter 3 gives first and second order optimality conditions for this relaxation. The conditions are close to the optimality criteria for trust region problems and represent an extension of this theory to $(n^2 + n)/2$ inequality constraints.

In Chapter 4 we summarize the theory about the connections between trust region problems and nonsymmetric perturbation theory [43]. The results are extended for the QAP and an algorithm to solve the relaxation is shortly proposed.

Chapter 5 presents a parametric approach to solve the relaxation of QAP which is a new way to tackle a trust region problem with more than one constraint. We derive two approaches for solving this relaxation parametrically. The numerical results are given in Section 5.3. We also introduce the theory of indefinite inner products which seems to be helpful to solve the problem.

Conclusions and perspectives are given in Chapter 6. They hopefully lead to future research starting points.

Chapter 2

Preliminaries

2.1 Notations

A QAP is called *symmetric (Hermitian)*, if both matrices A and B are symmetric (Hermitian). In [12] it is shown that if one of the matrices A or B is symmetric then there exists an equivalent QAP with both matrices being symmetric.

We will consider only symmetric QAPs in this section and Section 2.2 since we will observe in Section 2.3 that it is possible to transform an arbitrary QAP into an Hermitian QAP with (complex) Hermitian A and B, see [26]. In [24] it is shown that the results in the real case also hold for the complex case. For simplicity of notation we restrict ourselves to the real symmetric case, that is

$$
\begin{array}{lll}
(\mathbf{P}) & \min & f(X) = \mathrm{tr} AXBX^t + CX^t \\
 & \text{s.t.} & X \in \Pi,
\end{array}
$$

where A, B and $X \in \Re^{n\times n}$ and $A = A^t, B = B^t$.

The following notation will be used throughout. If S is symmetric then it has an orthogonal diagonalization $S = UDU^t$, where the eigenvalues of S are ordered

$$\lambda_1(S) \geq \lambda_2(S) \geq \ldots \geq \lambda_n(S).$$

Recall that the spectrum of S is real since S is symmetric. We will use the vector $\lambda(S)$ to denote the spectrum.

The n vector of all ones is $u := (1, \ldots, 1)^t \in \Re^n$. The vector of *row sums* of the $n \times n$ matrix S is denoted by $r(S)$, while the *sum of all the entries* of S is $s(S)$, i.e.

$$r(S) := Su, \quad s(S) := u^t Su.$$

We will use the following sets of $n \times n$ matrices.

$$\mathcal{O} := \{X : X^t X = I\}$$

denotes the set of *orthogonal* matrices, while

$$\mathcal{E} := \{X : Xu = X^t u = u\}$$

contains all matrices having *row and column sums equal to one*. Finally

$$\mathcal{N} := \{X : X \geq 0\}$$

is the set of matrices with *nonnegative coefficients*. To be able to indicate the size of the matrices we sometimes write indices on the sets, for example $\mathcal{O}_n$. It is well known that the set Π of *permutation* matrices satisfies

$$\Pi = \mathcal{O} \cap \mathcal{E} \cap \mathcal{N}. \tag{2.1.1}$$

If $a, b \in \Re^n$ we denote by $< a, b >_-$ the *minimal scalar product* of a and b, which is defined by

$$< a, b >_- := \min\{\sum_i a_i b_{\pi(i)} : \pi \quad \text{permutation}\}.$$

We observe that

$$< a, b >_- \quad = \quad a^t b$$

if a is in nonincreasing and b in nondecreasing order, respectively. The *maximal scalar product* $< a, b >_+$ is defined similarly with both a and b in either nonincreasing or nondecreasing order.

The matrix P is an *orthogonal projection* if

$$P^2 = P^t = P.$$

In this case, P has eigenvalues 0 or 1. A matrix U is a *partial isometry* if

$$||Ux|| = ||x||, \ \textit{for all} \ x \in \mathcal{R}(U^t),$$

where $\mathcal{R}(.)$ denotes *range space*. In this case U has eigenvalues of modulus 1 or 0 and both UU^t and U^tU are orthogonal projections, see e.g. [2].

2.2 Eigenvalue Bounds

The orthogonal relaxation of symmetric QAPs was introduced in [12, 38] and makes use of the following result which can be traced back to [10] and was applied to the QAP in [12, 38].

Theorem 2.1 *[10, 12, 38] Let A and B be real symmetric $n \times n$ matrices. Then*

$$< \lambda(A), \lambda(B) >_- \quad \leq \quad \mathrm{tr}\ AXBX^t \quad \leq \quad < \lambda(A), \lambda(B) >_+ \qquad \forall X \in \mathcal{O}.$$

Moreover, the lower (upper) bound is attained for $X = PQ^t$, where $P, Q \in \mathcal{O}$ contain the eigenvectors of A and B, respectively, in the order prescribed by the minimal (maximal) scalar product of the eigenvalues.

Theorem 2.1 is used as a basic tool to bound the quadratic part of $f(X)$. It seems rather difficult to extend Theorem 2.1 to include also the linear term present in $f(X)$. This difficulty can be overcome by separating $f(X)$ into a quadratic term

$$q(X) := \mathrm{tr}\ AXBX^t$$

and a linear term

$$l(X) := \mathrm{tr}\ CX^t.$$

The quadratic part can be bounded using Theorem 2.1 while the linear part is solved independently as a *linear sum assignment problem* (LSAP). The linear sum assignment problem is

$$\textbf{(LSAP)} \qquad\qquad \min_{X \in \Pi}\ \mathrm{tr}\ CX^t.$$

We let QAP(A,B,C) denote the optimal objective function value of the QAP defined by matrices A, B, C while LSAP(C) denotes the optimal value of the LSAP defined by C. The following eigenvalue related bound was proposed in [12].

$$QAP(A,B,C) \quad \geq \quad < \lambda(A), \lambda(B) >_- \quad + \ LSAP(C). \tag{2.2.2}$$

This bound is based on the eigenvalues and is denoted by EVB1. Noting that all input matrices for these problems belong to $\mathcal{N}$, a trivial lower bound is of course 0. In practice (2.2.2) is often negative and so a poor bound. To improve the bound in (2.2.2), transformations are applied to A, B and C that leave $f(X)$ unchanged over Π. Two types of transformations are known to have this property.

1. Adding constants to A or B row and column-wise and appropriately modifying C.

2. Changing the main diagonal of A or B and appropriately modifying C.

To be more specific, suppose $e, f, r, s \in \Re^n$. We define

$$\begin{aligned} A(e,r) &:= A + eu^t + ue^t + \mathrm{diag}(r), \\ B(f,s) &:= B + fu^t + uf^t + \mathrm{diag}(s), \\ C(e,f,r,s) &:= C + 2Auf^t + 2eu^tB - 2nef^t - 2\sum_k e_k uf^t \\ &\quad +\mathrm{diag}(A)s^t + r\ \mathrm{diag}(B)^t - 2es^t - 2rf^t - rs^t. \end{aligned}$$

Then it can easily be verified, see [12, 16], that

$$\mathrm{tr}\ (AXB+C)X^t = \mathrm{tr}\ (A(e,r)XB(f,s) + C(e,f,r,s))X^t \quad \forall e,f,r,s \in \Re^n,\ \forall X \in \Pi. \tag{2.2.3}$$

We point out that the same constant is added to the rows and columns which preserves symmetry of A and B.

Relation (2.2.3) shows that we may choose any transformation $d := (e,f,r,s) \in \Re^{n\times 4}$ to derive bounds for QAP. We now describe two strategies proposed in the literature to make choices for d.

The first strategy is based on the observation that a small "variance" of the spectrum of A and B results in small fluctuations for the quadratic part $q(X)$ by Theorem 2.1. The *variance* of the spectrum of a symmetric matrix $A, var(A)$, can be defined as

$$\mathrm{var}(A) := \sqrt{\frac{\mathrm{tr}\ A^2}{n} - \frac{1}{n^2}(\mathrm{tr}\ A)^2}.$$

In [12] the variance is used to bound the spread of the eigenvalues from above while in [38] the variance is used as an upper bound of the minimal scalar product. In both references it is shown that

$$\min\{\mathrm{var}(A(e,r)) : e, r \in \Re^n\}$$

is attained for

$$e_k = \frac{1}{n-2}[r_k(A) - a_{kk} - \frac{1}{2n-2}(s(A) - \mathrm{tr}\ A)], \quad \forall k; \quad r_k = a_{kk} - 2e_k, \quad \forall k. \tag{2.2.4}$$

Using these transformations for A and similarly (and independently) f and s for B, one obtains the lower bound EVB2 which requires essentially the same computational effort as EVB1. The dominating part is the eigenvalue computation for A and B. Moreover, the resulting bound has considerably improved, as compared to EVB1. Since e and r is chosen independently of f and s, and independently of

LSAP(C(e,f,r,s)), it is reasonable to expect further improvements of the bound by a more careful choice of the transformation d.

As a second strategy to choose d, an iterative improvement technique is proposed in [38] that tries to choose d so that the lower bound (2.2.2), as a function of d, is as large as possible. This leads to a nonlinear, nonsmooth, nonconcave maximization problem, so that to find the best possible choice for d seems almost as difficult as solving the original QAP but provides better bounds after only a few iterations. The bound EVB3 which results from this technique constitutes a further improvement over EVB2 and often yields the best bound available for QAP. It must be pointed out however, that the computational effort to obtain EVB3 is higher than for all the previously described bounds. This makes EVB3 a computationally expensive candidate as a bounding rule in a Branch and Bound scheme to solve QAPs.

2.3 Symmetrization and the Hermitian QAP

As mentioned above it is possible to transform a QAP which is nonsymmetric. In order to apply eigenvalue bounds we need an orthogonal or unitary diagonalization of A and B. For achieving this the QAP must be symmetric or Hermitian. Restricting ourselves to real matrices implies that A or B must be symmetric. However, in the complex case we can relax the symmetric assumption to a Hermitian assumption.

In this section we show that given any QAP there exists an equivalent (complex) QAP such that the matrices A and B are Hermitian and C is real. The theory which is discussed here can be extended to the complex case when A and B are given complex. A discussion of this extension can be found in [24].

In order to describe how to transform a non-symmetric QAP to an equivalent Hermitian QAP we need to describe several classes of matrices.

For a square matrix A, let

$$A_+ = \frac{A + A^*}{2}, \quad A_- = \frac{A - A^*}{2}$$

denote the *Hermitian* and *skew-Hermitian* parts of A, respectively, and $.^*$ denotes conjugate transpose. Letting $i = \sqrt{(-1)}$, we introduce two new classes of matrices. Denote the *positive* and *negative Hermitian* parts of A by

$$\tilde{A}_+ = (A_+ + iA_-), \quad \tilde{A}_- = (A_+ - iA_-),$$

respectively. Note that $\tilde{A}_+$ and $\tilde{A}_-$ are Hermitian. Using the positive Hermitian parts of A and B, we can symmetrize the objective function

$$q(X) := \mathrm{tr} AXB^tX^t$$

of a QAP with (arbitrary) matrices A and B.

Theorem 2.2 *[26] Let A and B be two real $n \times n$ matrices. For any real $n \times n$ matrix X*

$$\mathrm{tr} AXB^*X^* = \mathrm{tr} \tilde{A}_+ X \tilde{B}_+^* X^*.$$

So $QAP(A,B,C) = QAP(\tilde{A}_+, \tilde{B}_+, C)$.

As a consequence we can bound the range of an arbitrary QAP by the minimal scalar product of $\lambda(\tilde{A}_+)$ and $\lambda(\tilde{B}_+)$.

Theorem 2.3 *[26] Let a QAP with real matrices A and B be given. Then for all permutation matrices X*

$$< \lambda(\tilde{A}_+), \lambda(\tilde{B}_+) >_- \quad \leq \ \mathrm{tr} AXB^tX^t \quad \leq \ < \lambda(\tilde{A}_+), \lambda(\tilde{B}_+) >_+ .$$

Proof. By Theorem 2.2 above we have for all permutation matrices X

$$\mathrm{tr} AXB^tX^t = \mathrm{tr} AXB^*X^* = \mathrm{tr} \tilde{A}_+ X \tilde{B}_+^* X^*,$$

because A, B and X are real. The bounds follow from Theorem 2.1 by observing that permutation matrices are unitary. □

Another approach to symmetrize the objective function $f(X)$ is by its representation as a quadratic. For this we need the notation of the *Kronecker product* of two matrices A and B which is defined by

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{pmatrix}.$$

Consider a general *real* quadratic form x^tAx in the vector variable x. It is well known that

$$x^tAx = x^tA_+x$$

for all $x \in \Re^n$. That means a quadratic form can be represented by a Hermitian matrix. The quadratic $q(X) = \mathrm{tr} AXB^tX^t$ of a QAP with arbitrary A and B can be viewed as a quadratic form in the matrix variable X. It is natural to ask for a symmetric representation of $q(X)$, just as in the vector case.

If we let $x = \mathrm{vec}(X)$ be the vector formed from unraveling X row wise, and we let $K = A \otimes B$ be the Kronecker product of A and B, then it is easily verified that

$$\mathrm{tr} AXB^tX^t = x^tKx.$$

Thus a trivial way to symmetrize $f(X)$ would be to use x^tK_+x instead. But as a consequence one would have to work with $n^2 \times n^2$ matrices which seems computationally intractable. Furthermore there is no equivalent representation for the constraints.

2.4 Projection of QAP

We now look at an arbitrary (real) $QAP(A,B,C)$, that is A and B do not have to be symmetric. This section discusses lower bounds for QAP based on a projection method. The constraints defining $\mathcal{E}$ are eliminated by providing a tractable representation of the linear manifold $\mathcal{E}$. This is a standard way of treating equality constraints in optimization, often referred to as gradient projection or reduced gradient methods. More precisely, if the linear equality constraints which define $\mathcal{E}$ are described by the linear operator L, then we need to find the linear operator T such that $X = T(Y)$, for some $Y \in \Re^k$, if and only if $L(X) = 0$. In this case

$$\mathcal{E} = \{X : X = \hat{X} + T(Y),\ Y \in \Re^k\},$$

where $\hat{X}$ is some matrix in $\mathcal{E}$. We can then substitute for X in the objective function and eliminate the constraints that define $\mathcal{E}$. However, there is no guarantee that the permutation matrices are mapped to the orthogonals by the substitution. Moreover, the operator T must have special structure to preserve the trace structure of f. This way of doing the projection preserves orthogonality of X as well as the important trace formulation of f. The technique is based on a simple characterization of permutation matrices given below.

Let the $n \times (n-1)-$matrix V be such that

$$V^tu = 0; \quad V^tV = I_{n-1}.$$

The columns of V therefore constitute an orthonormal basis of $\{u\}^\perp$. Further, let

$$v := \frac{u}{\|u\|}; \quad P = [v \,\vdots\, V] \in \mathcal{O}.$$

Thus $Q := VV^t = I - vv^t$ describes the orthogonal projection on $\{u\}^\perp$.

Lemma 2.1 *[25] Let X be $n \times n$ and Y be $(n-1) \times (n-1)$. Suppose that X and Y satisfy*

$$X = P \begin{bmatrix} 1 & 0 \\ 0 & Y \end{bmatrix} P^t. \tag{2.4.5}$$

Then

$$X \in \mathcal{E},$$

$$X \in \mathcal{N} \iff VYV^t \geq -vv^t.$$

Conversely, if $X \in \mathcal{E}$, then there is a Y such that (2.4.5) holds.

The representation in (2.4.5), which is equivalent to

$$X = vv^t + VYV^t,$$

illustrates the projection which is used. The matrix $\hat{X} = vv^t \in \mathcal{E}$, while the linear operator $T(Y) = VYV^t$ yields the null space of the linear operator L which describes the manifold $\mathcal{E}$. As a consequence we have

$$\mathcal{E} = \{vv^t + VYV^t : Y \in \Re^{(n-1)\times(n-1)}\}.$$

Substitution for X in f using this representation maintains the trace structure of f, as well, the Lemma guarantees that each permutation matrix X corresponds to an orthogonal matrix Y. We get

$$\begin{aligned} f(X) &= \operatorname{tr}[A(vv^t + VYV^t)B + C](vv^t + VY^tV^t) \\ &= \operatorname{tr}\{Avv^tBvv^t + Avv^tBVY^tV^t + AVYV^tBvv^t + AVYV^tBVY^tV^t + Cvv^t + \\ &\quad CVY^tV^t\} \\ &= \tfrac{s(A)s(B)}{n^2} + \tfrac{s(C)}{n} + \operatorname{tr}\{(V^tAV)Y(V^tB^tV) + V^tCV + \tfrac{2}{n}V^tr(A)r^t(B)V\}Y^t. \end{aligned}$$

Let $\hat{A} := V^tAV$, $\hat{B} := V^tBV, \hat{C} := V^tCV$ and $\hat{D} := \frac{2}{n}V^tr(A)r^t(B)V + \hat{C}$. We define the *projected* problem by

$$(\mathbf{PQAP}) \qquad \begin{array}{ll} \min & \operatorname{tr}[(\hat{A}Y\hat{B}^t + \hat{D})Y^t] \\ \text{s.t.} & Y \in \mathcal{O}_{n-1}, \quad VYV^t \geq -vv^t. \end{array}$$

Note that we dropped the constant $\frac{s(A)s(B)}{n^2} + \frac{s(C)}{n}$ in the objective function in (PQAP) since it does not affect the solution Y.

In view of the Lemma and the characterization of Π from (2.1.1), the following Theorem holds.

Theorem 2.4 *[25] Let X and Y be related by (2.4.5). Then X solves QAP $\iff$ Y solves PQAP.*

We note that PQAP has a structure similar to QAP, but is smaller in dimension. Moreover, there is in general a linear term in PQAP due to the projection, even if $C = 0$ in QAP.

Before closing this section we look at the rank of the linear term of (PQAP). We know that V is full rank, i.e. rank $V = n-1$, since V is constructed as an orthonormal basis of $\{u\}^\perp$. Both $r(A)$ and $r(B)$ are at most rank 1, therefore $V^t r(A) r^t(B) V$ is also at most rank 1. The rank of the second part in $\hat{D}$ depends strongly on C, since

$$\text{rank } C - 2 \leq \text{rank } (V^t C V) \leq \min(\text{rank } V, \text{rank } C).$$

We can see that if QAP is pure quadratic, that is $C = 0$, the linear part will become at most rank 1.

2.5 Further Background

As we saw above it is always possible to project a given QAP to PQAP which can then be transformed to a Hermitian QAP, if necessary. For simplicity of notation, we assume for the remaining part of this chapter that the elimination of the constraints $\mathcal{E}$, i.e. the projection, has already been done. So we use A, B, C and X for $\hat{A}, \hat{B}, \hat{C}$ and Y, respectively. We also assume A and B to be real symmetric.

Please note that we will treat from now on PQAP as an n-dimensional problem for notational simplicity, so we assume the original QAP to be $(n+1)$-dimensional.

For relaxing (PQAP) we give up the nonnegativity constraints, which are $VXV^t \geq -vv^t$. We further relax the orthogonal constraint $X \in \mathcal{O}$ which is equivalent to $XX^t - I = 0$ to $XX^t - I = N$ where N is a negative semidefinite matrix. For notional convenience we add a factor of 2 to the linear part.

We then get the following relaxation to QAP

$$\begin{array}{lll} (\mathbf{P}) & \min & f(X) = \text{tr} AXBX^t + 2CX^t \\ & \text{s.t.} & g(X) = XX^t - I \preceq 0, \end{array}$$

where $\preceq$ refers to the partial order induced by the closed convex cone of symmetric positive semi-definite matrices, denoted $\mathcal{P}$ or psd. Thus two symmetric matrices satisfying $A \preceq B$ indicates $B - A \in \mathcal{P}$. We also let nd or $\prec$ denote negative definite. Similar definitions follow for psd, $\succeq$, pd, $\succ$.

This partial order that is often called the *Loewner ordering*, see e.g. [27], makes it possible to define monotonicity of matrix-valued functions of matrices.

Let us look at the constraints $g(X)$ of (P) which defines an operator from $\Re^{n\times n}$ to $\Re^{n\times n}$. We define g to be $\mathcal{P}$-convex if for $X_1, X_2 \in \Re^{n\times n}$ and any $\lambda \in [0,1]$ we have

$$\lambda g(X_1) + (1-\lambda) g(X_2) \succeq g(\lambda X_1 + (1-\lambda) X_2)$$

or equivalently

$$\lambda g(X_1) + (1-\lambda)g(X_2) - g(\lambda X_1 + (1-\lambda)X_2) \in \mathcal{P}.$$

When the eigenvalues of S are ordered

$$\lambda_1(S) \geq \lambda_2(S) \geq \ldots \geq \lambda_n(S),$$

then we can see that

$$S \preceq 0 \text{ if and only if } \lambda_1(S) \leq 0$$

since the spectrum of a nsd matrix is nonpositive. Thus we can replace the cone constraint in (P) by the scalar constraint

$$\lambda_1 = \lambda_1(g(X)) \leq 0.$$

The function $\lambda_1(g(X))$ is a convex function of X. We can find the derivative of $\lambda_k(g(X))$ using the corresponding normalized eigenvector v_k of $g(X)$ and the fact that if λ_k is simple, see e.g. [29], then

$$\begin{aligned}
\frac{\partial \lambda_k(g(X))}{\partial X_{ij}} &= v_k^t \frac{\partial g(X)}{\partial X_{ij}} v_k \\
&= \frac{1}{2} v_k^t (E_{ij} X^t + X E_{ij}^t) v_k \\
&= v_k(i)(X_{.j}^t v_k),
\end{aligned} \tag{2.5.6}$$

where E_{ij} is the zero matrix with 1 in the i,j-position, $X_{.j}$ is the j-th column of X and $v_k(i)$ is the i-th component of v_k. If λ_k is not simple, then the function λ_k may not be differentiable. Since we want X orthogonal at the solution, we can expect multiple eigenvalues of 1. It is well known that the largest eigenvalue is convex and so we can obtain expressions for the subdifferentials of the largest eigenvalue if its multiplicity is > 1, see e.g. [13], [23]. These subdifferentials are just the convex hull, of the expression given above for the derivative, over all normalized eigenvectors v_k.

Note that the differentials at X in the direction h of the above functions in (P) are:

$$\begin{aligned}
df(X;h) &= \operatorname{tr} A(XBh^t + hB^tX^t) + 2Ch^t \\
dg(X;h) &= Xh^t + hX^t.
\end{aligned} \tag{2.5.7}$$

Moreover

$$df(X;h) = \operatorname{tr}\ 2AXBh^t + 2Ch^t,$$

since A and B are symmetric.

We use the inner-product

$$< S, T >= \text{ tr } ST^t$$

on the space of $n \times n$ matrices. In the space of symmetric matrices with this inner-product, $\mathcal{P}$ is a closed convex cone with nonempty interior, $int\ \mathcal{P} \neq \emptyset$. Moreover, $\mathcal{P}$ is *self polar*, i.e. the *polar cone*

$$\mathcal{P}^+ = \{T = T^t : \text{ tr } ST \geq 0, \ for\ all\ S \in \mathcal{P}\} = \mathcal{P}$$

see e.g. [31].

We will also need the *singular values* of a matrix E which we denote by

$$\sigma_1(E) \geq \ldots \geq \sigma_n(E).$$

The corresponding singular value decomposition of E is $U\Sigma V^t = E$ where U and V are orthogonal $n \times n$ matrices and Σ is a diagonal matrix containing the singular values of E.

We should mention that optimization over a partial order like the Loewner order is an important problem. There are many applications which occur for example in control theory or combinatorial problems. Some applications were presented at the Fourth SIAM Conference on Optimization in Chicago in May 1992, e.g. [1, 4]. So this work provides theory for these problems.

Chapter 3

Optimality Conditions for the Relaxed Problem

The bounds discussed in Chapter 2 dealt mainly with the quadratic and linear part of QAP separately. Only EVB3, the bound achieved by the iterative improvement technique, attempted to bring these two parts together. This outperformed, as mentioned above, the other techniques.

In the remaining part of this thesis we will treat the two parts of QAP together. To do this we start by giving first and second order optimality conditions for the relaxation of the form (P) as we defined it in the previous section.

3.1 First Order Conditions

In this section we present the first order optimality conditions for the relaxation

$$\begin{array}{lll} (\mathbf{P}) & \min & f(X) = \mathrm{tr} AXBX^t + 2CX^t \\ & \text{s.t.} & g(X) = XX^t - I \preceq 0 \end{array}$$

with $A, B, C \in \Re^{n\times n}$. For this chapter we assume A and B to be symmetric and assume that the projection was already done. The relaxation provides, under certain circumstances, conditions which guarantee that the optimal solution is orthogonal. Thus we will see that under controlled assumptions, we do not weaken our bound by relaxing the constraint $XX^t - I = 0$ to $XX^t - I \preceq 0$. First we show the following.

Lemma 3.1 *Y is an extreme point of the feasible set*

$$F = \{X : XX^t - I \preceq 0\},$$

if and only if it is an orthogonal matrix.

Proof. Suppose that the $n \times n$ matrix $Y \in F$ is not orthogonal. Let $Y = U\Sigma V^t$ be its singular value decomposition with the diagonal matrix Σ containing the singular value $0 \leq \sigma_k < 1$, i.e. we get

$$0 \preceq (Y \pm \epsilon U E_k V^t)(Y \pm \epsilon U E_k V^t) \preceq I,$$

for some ϵ, i.e. Y is not an extreme point.

Conversely, suppose Y is orthogonal, but $Y = \theta Y_1 + (1-\theta) Y_2$, for some $0 < \theta < 1$, and $Y_i \in F,\ i = 1,2$. Then the singular value decomposition of Y satisfies $\Sigma = I = \theta Z_1 + (1-\theta) Z_2$, where $Z_i = U^t Y_i V,\ i = 1,2$. Since $Z_i,\ i = 1,2$, are still feasible and so have norm ≤ 1, we conclude that $Z_i = I,\ i = 1,2$, i.e. $Y = Y_i,\ i = 1,2$. Thus Y is an extreme point. $\square$

Equivalent formulations of Lemma 3.1 can be found in [11, 37] for gaining a better insight. Although the formulations are slightly different, the resulting feasible set and its extreme points are the same.

If the objective function is concave, then the optimum of (P) is attained at an extreme point of F and is then orthogonal. It is easy to see that

$$f(X) \text{ is concave if and only if } A \otimes B \text{ is nsd.}$$

But for the spectra of $A \otimes B$ to be nonpositive either the spectrum of A must be nonnegative and the spectrum of B nonpositive or vice versa.

Note that we assumed that A and B were already projected, that is $A = V^t \tilde{A} V$ and $B = V^t \tilde{B} V$, with V defined as in Section 2.4. It is easy to see that the projection does not change the sign of the spectra of $\tilde{A}$ and $\tilde{B}$.

Suppose $\tilde{A}$ is psd, that is

$$x^t \tilde{A} x \geq 0 \qquad \forall x, \quad x^t x = 1.$$

But then

$$x^t V^t A V x = y^t A y \geq 0 \quad \forall y, \quad y^t y = x^t V^t V x = 1$$

since $V^t V = I$. So A is also psd. The argument works in the same way for negative semidefiniteness.

So one has to find appropriate r and s to shift the diagonals of $\tilde{A}$ and $\tilde{B}$ to make $\tilde{A} \otimes \tilde{B}$ nsd. This can be achieved easily by choosing $|r|$ and $|s|$ large enough. At the same time the linear term is modified as described in Section 2.2. The negative semidefiniteness of $A \otimes B$ follows.

By making $f(X)$ concave, we create many local minima. We will see in Theorem 3.2 that we can do even better and actually guarantee that X is orthogonal at the optimum without the concavity assumption.

We define the *Lagrangian* of (P)

$$\mathcal{L}(X,S) = f(X) + \mathrm{tr} S g(X)$$

and the *first order optimality conditions*

$$\begin{array}{rc} (i) & AXB + C = -SX \\ (ii) & \mathrm{tr} S(XX^t - I) = 0 \\ (iii) & XX^t \preceq I \\ (iv) & S \succeq 0. \end{array} \tag{3.1.8}$$

Note that the Lagrange multiplier is a psd symmetric matrix S, since $g(X)$ is symmetric and $\mathcal{P} = \mathcal{P}^+$.

Before we give a proof for the first order optimality conditions of (P) we look at the relation of a minimization problem like (P) to a *convex program* in which both the objective function and the constraints are convex.

Lemma 3.2 *Let $f : \Re^{n\times n} \to \Re$ be differentiable and $g : \Re^{n\times n} \to \Re^{n\times n}$ be $\mathcal{P}$-convex. Let X_0 solve*

$$\text{(P1)} \qquad \min f(X), \quad s.t \quad g(X) \preceq 0,$$

and define

$$\text{(P2)} \qquad \min \tilde{f}(X), \quad s.t \quad g(X) \preceq 0,$$

where $\tilde{f}(X) = \bigtriangledown f(X_0)(X - X_0)$. Then X_0 solves (P2), as well.

Proof. Suppose X_0 does not solve (P2). Then there exists an X_1 such that $g(X_1) \preceq 0$ and

$$\bigtriangledown f(X_0)(X_1 - X_0) < \bigtriangledown f(X_0)(X_0 - X_0) = 0.$$

But this implies that $X_1 - X_0$ is a feasible direction of decrease for (P1) which contradicts the optimality in (P1). □

Note that (P2) is a convex program since its objective function $\tilde{f}$ is linear and therefore convex. So we have a strong relationship between our problem (P) and a convex program of a form like (P2).

Recall that the *Slater condition*

$$\exists X \in \Re^{n\times n} \quad \ni \quad g(X) < 0$$

is a constraint qualification for convex programs. It guarantees the existence of Lagrange multipliers for the convex constraints. By Lemma 3.2 we know that Slater's condition is also a valid constraint qualification for (P) since each X which solves (P) also solves the convex program defined in Lemma 3.2.

Theorem 3.1 *Suppose that X solves (P). Then (3.1.8) holds for some S.*

Proof. Since $-I \in -\text{int}\ \mathcal{P}$, the zero matrix $X = 0$ satisfies the Slater constraint qualification, i.e., $g(0) \prec 0$. By Lemma 3.2 we know that the Slater condition is a valid constraint qualification for (P). Therefore, the standard Lagrange multiplier theorem, see e.g., [32], states that there exists $S \in \mathcal{P}^+$ such that complementary slackness $\text{tr}S(XX^t - I) = 0$ holds and, for all matrices h,

$$
\begin{aligned}
0 &= < \triangledown \mathcal{L}(X, S), h > \\
&= tr[A(XBh^t + hBX^t) + S(Xh^t + hX^t)] + 2Ch^t \\
&= tr2(AXB + C + SX)h^t,
\end{aligned}
$$

i..e. $AXB + C + SX = 0$. □

The first order conditions provide us with means of guaranteeing that the solution X of (P) is actually orthogonal. (Recall that we can shift A, B, and C to possibly guarantee (3.1.9) below.)

Theorem 3.2 *Suppose that A and B are nonsingular and the smallest singular value*

$$\sigma_n(A^{-1}CB^{-1}) > 1. \tag{3.1.9}$$

Then, if X solves (P), it is orthogonal and the associated Lagrange multiplier

$$S \succ 0.$$

Proof. Suppose that X solves (P). To prove that S and X are nonsingular, the first order conditions (3.1.8)(i) implies that we need only show

$$X + A^{-1}CB^{-1} \text{ is nonsingular,} \tag{3.1.10}$$

since nonsingularity of S and X implies $SX = AXB + C$ is also nonsingular. Let

$$U(A^{-1}CB^{-1})V^t = \Sigma$$

be the singular value decomposition of $A^{-1}CB^{-1}$. Then (3.1.10) holds if and only if

$$Y + \Sigma \text{ is nonsingular,}$$

where $Y = UXV^t$. Now (3.1.8)(iii) implies that

$$YY^t = UXX^tU^t \preceq UIU^t = I, \tag{3.1.11}$$

since U and V are orthogonal. Now for $||x|| = 1$ we get

$$\begin{aligned} x^t(Y+\Sigma)x &\geq x^tYx + \sigma_n \\ &> x^t\frac{(Y+Y^t)}{2}x + 1 \\ &\geq -\sigma_1(Y) + 1, \quad \text{(e.g. [34] pg. 240])} \\ &\geq 0, \end{aligned}$$

by (3.1.11). This yields (3.1.10) and so $S \succ 0$. Now 3.1.8(ii) implies that

$$\begin{aligned} 0 &= \operatorname{tr} S(X^tX - I) \\ &= \operatorname{tr} S^{1/2}(X^tX - I)S^{1/2} \end{aligned}$$

and 3.1.8(iii) that

$$S^{1/2}(X^tX - I)S^{1/2} \preceq 0.$$

Thus, $X^tX - I = 0$. □

In the pure quadratic case, i.e. $C = 0$, we do not have to apply the relaxation $XX^t \preceq I$ but can use $XX^t = I$. This yields the eigenvalue decomposition bounds of Theorem (2.1).

In the case that $A = B = 0$, the problem (P) can be solved explicitly.

Corollary 3.1.1 *Suppose that $A = B = 0$. Then the optimum X for problem (P) is obtained from the polar decomposition of C,*

$$C = SX, \tag{3.1.12}$$

where X is orthogonal and $S \succeq 0$. Moreover, the optimal value

$$\operatorname{tr} CX^t = \sum_{i=1}^{n} \sigma_i(C).$$

Proof. Since $f(X)$ is linear, the optimum X is an extreme point of the feasible set F, i.e., it is orthogonal. Thus (3.1.12) follows from the first order optimality condition for (P). □

Note that if $f(X)$ is convex, i.e., $A \otimes B$ is psd, then Theorem (2.1) yields a first order characterization of optimality.

3.2 Second Order Conditions

We now present second order optimality conditions for our relaxation

$$
\begin{array}{llll}
(\mathbf{P}) & \min & f(X) = \operatorname{tr} AXBX^t + 2CX^t \\
 & \text{s.t.} & g(X) = XX^t - I \preceq 0
\end{array}
$$

These provide us with necessary and sufficient optimality criteria independent of whether the objective function f is convex or concave. This extends known results on trust region methods and methods for quadratic objectives with a single quadratic constraint. Note that the relaxation has $(n^2+n)/2$ constraints. Problems with only two trust region constraints have been dealt with in [44, 45] and have proven to be very hard.

We first present a test for optimality in (P) which compares different solutions of the first order optimality conditions (3.1.8). This extends the result in [18] which deals with a single real valued quadratic constraint.

Theorem 3.3 *Suppose that $X_i, S_i, i = 1, 2$, are solutions of the first order optimality conditions (3.1.8) with $S_i \succ 0, i = 1, 2$. Then*

$$4(f(X_2) - f(X_1)) = \operatorname{tr} (X_1 - X_2)(S_1 - S_2)(X_1 - X_2)^t.$$

Proof. From the first order conditions (3.1.8) we get

$$\operatorname{tr} AX_iBX_i^t + CX_i^t = -\operatorname{tr} S_iX_iX_i^t, \quad i = 1, 2,$$

and, after subtracting,

$$\operatorname{tr} AX_2BX_2^t - \operatorname{tr} AX_1BX_1^t + C(X_2 - X_1)^t = \operatorname{tr} (S_1X_1X_1^t - S_2X_2X_2^t). \qquad (3.2.13)$$

Also, (3.1.8) implies that

$$
\begin{aligned}
\operatorname{tr} AX_1BX_2^t + CX_2^t &= -\operatorname{tr} S_1X_1X_2^t \\
\operatorname{tr} AX_2BX_1^t + CX_1^t &= -\operatorname{tr} S_2X_2X_1^t = -\operatorname{tr} S_2X_1X_2^t
\end{aligned}
$$

which, after subtracting, yields

$$\operatorname{tr} C(X_2 - X_1)^t = -\operatorname{tr} S_1X_1X_2^t + \operatorname{tr} S_2X_1X_2^t. \qquad (3.2.14)$$

Subtracting (3.2.13) and (3.2.14) we get

$$
\begin{aligned}
2(f(X_2) - f(X_1)) &= \operatorname{tr} S_1X_1X_1^t - \operatorname{tr} S_2X_2X_2^t - \operatorname{tr} (S_1 - S_2)X_1X_2^t \\
&= \operatorname{tr} (S_1 - S_2)(I - X_1X_2^t) \\
&= \frac{\operatorname{tr} (S_1 - S_2)}{2}(I - X_1X_2^t + I - X_2X_1^t),
\end{aligned}
$$

since $S \succ 0$ implies $X_1X_1^t = X_2X_2^t = I$. $\square$

We now characterize the feasible directions at a feasible point X.

Lemma 3.3 *Suppose that $XX^t \preceq I$. Denote the set of feasible directions at X by*

$$\mathcal{D}_X = \{V \in \Re^{n\times n} : \exists \overline{\theta} > 0 \textit{ with } (X + \overline{\theta}V)(X + \overline{\theta}V)^t \leq I\}.$$

Then:

$$a)\ \ \emptyset \neq \textit{ int } \mathcal{D}_X = \{V : XV^t + VX^t \textit{ is nd on } \mathcal{N}(XX^t - I)\};$$
$$b)\ \ \overline{\mathcal{D}}_X = \{V : XV^t + VX^t \textit{ is nsd on } \mathcal{N}(XX^t - I)\}.$$

Proof. Note that the Hessian of tr $SXX^t = I \otimes S$ is psd and so the constraint g is $\mathcal{P}$-convex and the feasible set F is a convex set. Moreover, $g(0) \prec 0$ i.e. $X = 0$ is in the interior of F. Therefore $\mathcal{D}_0 = \Re^{n\times n}$. Thus the result holds in this trivial case. Now suppose that $0 \neq XX^t \preceq I$. Then the direction $V = -X$ points into the interior of the feasible set, i.e.

$$(X + \theta V)(X + \theta V)^t = (1-\theta)^2 XX^t \prec I,$$

for $0 < \theta \leq 1$. Thus $V = -X$ is in the interior of the convex cone $\mathcal{D}_X$.

Suppose h is a feasible direction. Then,

$$V_\lambda = \lambda h + (1-\lambda)(-X),$$

for $0 \leq \lambda < 1$, is a feasible direction pointing into the interior of the feasible set i.e. $V_\lambda \in$ int $\mathcal{D}_X$. Therefore,

$$(X + \theta V_\lambda)(X + \theta V_\lambda)^t - I = \theta(XV_\lambda^t + V_\lambda X^t) + \theta^2 V_\lambda V_\lambda^t + XX^t - I$$

is nsd for small, $\theta > 0$. Since $V_\lambda V_\lambda^t$ is psd, this implies that $XV_\lambda^t + V_\lambda X^t$ is nd on $\mathcal{N}(XX^t - I)\}$ and thus $Xh^t + hX^t$ is nsd on $\mathcal{N}(XX^t - I)\}$. Thus we have shown that int $\mathcal{D}_X \neq \phi$ and

$$\mathcal{D}_X \subset \{V : XV^t + VX^t \text{ is nd on } \mathcal{N}(XX^t - I)\}.$$

Conversely, suppose that $Xh^t + hX^t$ is nd on $\mathcal{N}(XX^t - I)\}$. Then $\theta(Xh^t + hX^t) + \theta^2 hh^t$ is nd on $\mathcal{N}(XX^t - I)\}$, for small $\theta > 0$. This implies that

$$(X + \theta h)(X + \theta h)^t = \theta(Xh^t + hX^t) + \theta^2 hh^t + XX^t - I$$

is nsd, for small $\theta > 0$, see e.g. [3] or [33], i.e. $h \in \mathcal{D}_X$. We can perturb h and still maintain that $Xh^t + hX^t$ is nd on $\mathcal{N}(XX^t - I)\}$. Therefore $h \in$ int $\mathcal{D}_X$. This proves a). Since int $\mathcal{D}_X \neq \emptyset$, b) follows from a continuity argument. $\square$

We now present second order optimality conditions for (P). Note that (3.2.15) differs from the standard conditions in the literature and allows for sufficiency to hold. In this respect, it is close to the trust region results which will be discussed in the next chapter.

Theorem 3.4 *Suppose X is feasible for (P). Define the optimality conditions*

$$\begin{array}{rc} (i) & S \succeq 0 \\ (ii) & AXB + C + SX = 0 \\ (iii) & \text{tr } S(XX^t - I) = 0 \\ (iv) & \text{tr } AhBh^t + hSh^t \geq 0, \; \textit{if } Xh^t + hX^t \textit{ is nsd on } \mathcal{N}(XX^t - I)\} \end{array} \tag{3.2.15}$$

Then:

a) X is a global minimum for (P) if and only if (3.2.15) holds for some S;

b) if (3.2.15) holds for some $S \succ 0$, then X is a global minimum of (P) and $XX^t = I$.

Proof. We first prove (a). Suppose that X solves (P) and that $XV^t + VX^t$ is negative definite on $\mathcal{N}(XX^t - I)$ with V nonsingular. Let us show that (3.2.15)(iv) holds for $h = V$. Let

$$Y_\theta = X + \theta V.$$

Note that $Y_\theta \in \text{int}F$, for small $\theta > 0$ by Lemma 3.3. Therefore,

$$Y_\theta Y_\theta^t - I = XX^t + \theta VX^t + \theta XV^t + \theta^2 VV^t - I = \theta(XV^t + VX^t) + \theta^2 VV^t + (XX^t - I)$$

is negative definite for small $\theta > 0$. If $S \neq 0$, and since V is not singular, this implies

$$\text{tr } S(Y_\theta Y_\theta^t - I) = \begin{cases} \gamma < 0, \text{for small, } \theta > 0 \\ \gamma > 0, \text{for large, } \theta > 0 \end{cases}$$

and so for any $S \succeq 0$ there exists $\overline{\theta} > 0$ such that

$$\text{tr } S(Y_{\overline{\theta}} Y_{\overline{\theta}}^t - I) = 0.$$

For simplicity of notation, let $\overline{\theta} = 1$ and

$$Y = Y_{\overline{\theta}} = X + V.$$

Then the Lagrangian

$$\begin{array}{rcl}\mathcal{L}(Y,S) & = & f(Y)+ \text{ tr } S(YY^t-I)=f(Y)\\ \mathcal{L}(X,S) & = & f(X)+ \text{ tr } S(XX^t-I)=f(X)\\ \bigtriangledown\mathcal{L}(X,S) & = & 0\end{array}$$

and, since the Lagrangian is a quadratic, we get

$$\begin{array}{rcl}f(Y)=\mathcal{L}(Y,S) & = & \mathcal{L}(X,S)+\bigtriangledown\mathcal{L}(X,S)(Y-X)+ < V, \bigtriangledown^2\mathcal{L}(X,S)V >\\ & = & f(X)+0+\text{tr } AVBV^t+\text{tr } VSV^t.\end{array}$$

Since $f(X)\leq f(Y)$ we get

$$\text{tr } AVBV^t+VSV^t\geq 0.$$

By continuity this holds for $V\in\overline{\mathcal{D}}_X$. This proves necessity. Note that (iv) implies that the standard second order optimality conditions hold.

Now suppose (3.2.15) holds. If Y is feasible for (P) and $f(Y)<f(X)$, then

$$\begin{array}{rcl}\mathcal{L}(Y,S) & = & f(Y)+ \text{ tr } S(YY^t-I)\\ & < & f(X)\\ & = & f(X)+ \text{ tr } S(XX^t-I)\\ & = & \mathcal{L}(X,S).\end{array} \tag{3.2.16}$$

We can assume that YY^t-I is negative definite. (Use a perturbation of Y and maintain $f(Y)<f(X)$, if needed.) Let $V=X-Y$ and $Y_\theta=X-\theta V$, and so Y_θ is a convex combination of X and Y for $0\leq\theta\leq 1$. Then $Y_\theta Y_\theta^t-I= -\theta(XV^t+VX^t)+\theta^2VV^t+XX^t-I$ is negative definite for small $\theta>0$. Therefore, XV^t+VX^t is positive definite on $\mathcal{N}(XX^t-I)$ which, by hypothesis, implies that

$$< V, \bigtriangledown^2\mathcal{L}(X,S)V >= \text{ tr } AVBV^t+VSV^t\geq 0.$$

But we have shown that

$$\mathcal{L}(Y,S)=\mathcal{L}(X,S)+ < V, \bigtriangledown^2\mathcal{L}(X,S)V > \;\; < \mathcal{L}(X,S),$$

by (3.2.16), a contradiction. This proves (a).

(b) follows from complementary slackness, i.e. $S\succ 0,\quad XX^t-I\preceq 0$ and $\text{tr } S(XX^t-I)=0$ imply $XX^t-I=0$. $\square$

Corollary 3.2.1 *Suppose that*

$$\sigma_n(A^{-1}CB^{-1}) > 1. \tag{3.2.17}$$

Then the second order conditions (3.2.15) characterize optimality of X for (P), with X orthogonal, i.e. X is orthogonal and solves (P) if and only if the first order conditions (3.2.15) hold and

$$\mathrm{tr} AhBh^t + hSh^t \geq 0 \ if \ Xh^t + hX^t \ is \ nsd. \tag{3.2.18}$$

Proof. Necessity follows directly from the previous theorem and holds without (3.2.17). By Theorem 3.4, the first order conditions and (3.2.18) are sufficient for X to be a global optimum for (P). But Theorem 3.2 and (3.2.17) imply that if X solves (P) then it is orthogonal and the associated Lagrange multiplier $S \succ 0$. This proves sufficiency. □

Chapter 4

Nonsymmetric Perturbations and A Trust Region Bound

4.1 Motivation

The orthogonal constraint $XX^t \preceq (=)\ I$ resembles in form the quadratic vector norm constraint, $\|x\|^2 = x^t x \leq (=)\ \delta_c, x \in \Re^n$ with $\delta_c = 1$. As we will see in this chapter, the constraint $x^t x$, when applied to the minimization of a quadratic on $\Re^n$, results in an eigenvalue problem. Therefore, the eigenvalue bounds discussed in Chapter 2 are intuitively not surprising.

When δ_c is arbitrary, the norm constraint is often called a trust region constraint. Trust region algorithms are used for unconstrained minimization. Given the real, twice continuously differentiable, function $f : \Re^n \to \Re$, a current estimate x_c, the gradient g_c at x_c, and an estimate of the Hessian H_c at x_c, the trust region approaches choose an approximate step length δ_c and find the next iterate x_+ as an approximate solution of the constrained quadratic model

$$\begin{array}{ll} \min & f(x_c) + 2g_c^t + d^t H_c d \\ s.t. & \|d\| \leq \delta_c. \end{array}$$

The radius of the trust reion δ_c is increased or decreased depending on whether we get a satisfactory decrease in the objective function.

The theory of trust region problems was discussed in [14, 15, 17, 19, 41]. Furthermore, numerical techniques for solving trust region problems are given in [14, 35, 41].

4.2 Nonsymmetric Perturbation Theory and the Trust Region Problem

We will now look at the connection between spectral properties of a nonsymmetrically perturbed matrix to stationarity properties of a specific trust region problem. In doing this we follow [43].

Suppose that M is a real symmetric $n \times n$ matrix. Then the classical Raleigh Principle and Courant-Fischer Minimax Theorem relate the eigenvalues of M to the stationary points of the quadratic function

$$\nu(x) = x^t M x$$

with respect to the constraints

$$S_n = \{x \in \Re^n : x^t x = 1\}.$$

The *Lagrangian function* of the minimization or maximization of $\mu(x)$ with respect to S_n is

$$L(x, \lambda) = \nu(x) - \lambda x^t x$$

and so the *Lagrange equation* becomes

$$\partial_x L(x, \lambda) = 0$$

which is

$$Mx - \lambda x = 0, \tag{4.2.19}$$

and is also the eigenvector equation of M. Recall that if x and λ satisfy (4.2.19) and $x \in S_n$, then λ and x are a Lagrange multiplier and an associated stationary point of $\nu(x)$ with respect to S_n, respectively.

Thus there is a one-to-one correspondence between the eigenvalues of a symmetric matrix and the Lagrange multipliers of the corresponding quadratic form. Furthermore, the stationary points can be found by determining the unit eigenvectors of M. Note that neither convexity nor concavity is necessary in locating the constrained maximum or minimum of $\nu(\cdot)$, since they correspond to the maximal and minimal eigenvalues, respectively.

Stern and Wolkowicz examine in [43] the influence of nonsymmetric border perturbations of M on the eigenvalues of the perturbed system. Also, the connections between the spectral properties of the perturbed matrix and stationary properties of a specific trust region problem are shown.

More precisely, let K be the $(n+1)\times(n+1)$ matrix which results from a nonsymmetric border perturbation on M, that is

$$K = \begin{pmatrix} t & \eta^t \\ -\eta & M \end{pmatrix} \tag{4.2.20}$$

with $\eta \in \Re^n$ and $t \in \Re$. Now K may possess either a real or a complex spectrum depending on the choice of t. If $t = 0$ then perturbation (4.2.20) is skew-symmetric.

On the other hand consider the general quadratic function

$$\mu(x) = x^t M x - 2\eta^t x$$

on S_n. Its theory has been studied extensively in the literature. In particular, we consider a trust region problem with $\delta_c = 1$ that is

$$\textbf{(TR)} \qquad \begin{array}{ll} \min & \mu(x) = x^t M x - 2\eta^t x \\ \text{s.t.} & x^t x = 1 \end{array}$$

The Lagrange equation of (TR) is

$$(M - \lambda I)x - \eta = 0, \tag{4.2.21}$$

which is also a part of the first order necessary conditions for a solution of (TR). We denote the set of Lagrange multipliers of (TR) by Λ and assign for each $\lambda \in \Lambda$ its associated set of stationary points $S_\mu(\lambda)$.

The main results for trust region problems in the form (TR) are summarized in the following two theorems.

Theorem 4.1 *[20, 35, 41] The vector $x \in \Re^n$, with $x^t x \leq 1$, is a minimum (maximum) point of $\mu(\cdot)$ over S_n if and only if there exists a scalar λ such that x and λ together satisfy the Lagrange equation (4.2.21), $\lambda(x^t x - 1) = 0$, and the matrix $B - \lambda I$ is positive (negative) semidefinite.*

Note that Theorem 4.1 allows a characterization of optimality independent of convexity of the quadratic objective function, i.e. there is no gap between necessary and sufficient optimality conditions for a global minimum (maximum).

Theorem 4.2 *[15] The set Λ of Lagrange multipliers of $\mu(\cdot)$ with respect to S_n is finite. Let Λ be given by*

$$\lambda_1 > \lambda_2 > \ldots > \lambda_k,$$

and let $x^{\lambda_i} \in S_\mu(\lambda_i), i = 1, 2, \ldots, k$. Then

$$\mu(x^{\lambda_1}) > \mu(x^{\lambda_2}) > \ldots > \mu(x^{\lambda_k}).$$

In particular, the minimum (maximum) of $\mu(\cdot)$ over S_n is attained at a stationary point associated with λ_k (λ_1).

We denote the spectrum of M by $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_n$ and let P be an orthogonal matrix such that

$$P^t M P = D = \mathrm{diag}(\gamma_1, \gamma_2, \ldots, \gamma_n)$$

the diagonal matrix with elements $\gamma_1, \gamma_2, \ldots, \gamma_n$. Then the Lagrange equation (4.2.21) becomes

$$(D - \lambda I)\hat{x} = \hat{\eta}$$

with

$$\hat{x} = P^t x \quad \text{and} \quad \hat{\eta} = P^t \eta.$$

Note that this transformation does not change the set of Lagrange multipliers Λ and furthermore for every $x \in \Re^n$ we have

$$\hat{\mu}(\hat{x}) = \hat{x}^t D \hat{x} - \hat{\eta}^t \hat{x} = \mu(x).$$

We now assume that no eigenvector of M is orthogonal to η or equivalently

$$\hat{\eta}_i \neq 0 \qquad \forall i = 1, 2, \ldots n-1. \tag{4.2.22}$$

It can then be shown that in this case $M - \lambda I$ is invertible which allows the following results. First of all the set $S_\mu(\lambda)$ becomes the singleton

$$x^\lambda = (M - \lambda I)^{-1}\eta \tag{4.2.23}$$

and

$$\hat{x}^\lambda = (D - \lambda I)^{-1}\hat{\eta}. \tag{4.2.24}$$

Furthermore the set of Lagrange multipliers Λ is also the set of solutions of the *implicit secular equation*

$$1 - \eta^t (M - \lambda I)^{-2}\eta = 0 \tag{4.2.25}$$

which arises from the constraint $\hat{x}^{\lambda t}\hat{x}^\lambda = 1$. But (4.2.25) has the same solution set as the so called *explicit secular equation*

$$f_\mu(\lambda) = 1 - \sum_{i=1}^{n-1} \left(\frac{\hat{\eta}_i}{\gamma_i - \lambda}\right)^2 = 0. \tag{4.2.26}$$

Using the terminology of [19] we denote (4.2.26) as the *secular function* associated with $\mu(\cdot)$.

Then a unique solution of (TR) is given by x^{λ_k} from formula (4.2.23). We denote the case where condition (4.2.22) does not hold as the hard case. Note that in the hard case it may happen that the set of solutions to (TR) is not a singleton.

We now introduce the *secular antiderivative function* associated with $\mu(\cdot)$. Consider

$$g_\mu(\lambda) = \lambda - \sum_{i=1}^{n-1} (\frac{\hat{\eta}_i}{\gamma_i - \lambda})^2 = 0$$

with the convention $0/0 = 0$. Then the singularities of $g_\mu(\lambda)$ are the same as those of the secular function $f_\mu(\lambda)$, and $g'_\mu(\lambda) = f_\mu(\lambda)$ at every non-singularity λ. The following lemma establishes connections between trust region problems and perturbation theory.

Lemma 4.1 *[43] Assume that condition (4.2.22) holds and let $\lambda \in \Lambda$. Then*

$$g_\mu(\lambda) = \mu(x^\lambda)$$

where $x^\lambda = (B - \lambda I)^{-1}\eta$.

This Lemma shows that in the case in which (4.2.22) holds the secular antiderivative's values on the Lagrange multiplier set Λ are exactly the values of $\mu(\cdot)$ on the corresponding set of stationary points.

We now give sufficiency conditions for realness of the spectrum of K.

Theorem 4.3 *[43] Let M be an $n \times n$ real symmetric matrix, and consider the perturbation of K given by*

$$K = \begin{pmatrix} t & \eta^t \\ -\eta & M \end{pmatrix}$$

where $\eta \in \Re^n$. Define

$$\mu(x) = x^t M x - 2\eta^t x.$$

Let

$$\mu_1 = \max\ \{\mu(x) : x^t x = 1\}$$

and

$$\mu_k = \min\ \{\mu(x) : x^t x = 1\}.$$

Then either of the conditions

$$t \geq \mu_1$$

or

$$t \leq \mu_k$$

are sufficient for the spectrum of K to be real.

The following is another result regarding the perturbation K. This elementary result yields further connections between trust region problems and nonsymmetric perturbations. This result will be a key in the considerations of the following section.

Theorem 4.4 *[43] Let K, M and $\mu(\cdot)$ be as in Theorem 4.3. Then $\lambda \in \Lambda$, that is, λ is a Lagrange multiplier of $\mu(\cdot)$ with respect to S_n with $x \in S_\mu(\lambda)$, if and only if λ is an eigenvalue of K with associated eigenvector*

$$\begin{pmatrix} 1 \\ x \end{pmatrix}, \qquad x^t x = 1,$$

in which case $t = \mu(x)$.

Proof. After taking the inner product of the Lagrange equation

$$Mx - \eta - \lambda x = 0$$

with x and using the fact $x^t x = 1$ we obtain

$$x^t M x - \eta^t x - \lambda = 0$$

which is equivalent to

$$\mu(x) + \eta^t x - \lambda = 0.$$

The theorem follows from the eigenvector equation

$$\begin{pmatrix} t & \eta^t \\ -\eta & M \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ x \end{pmatrix}$$

which yields

$$\eta^t x + t = \lambda \Rightarrow \mu(x) = \lambda - \eta^t x \quad \text{if} \quad x^t x = 1$$

and

$$Mx - \eta = \lambda x \Rightarrow \lambda x^t x = x^t M x - \eta^t x = \mu(x) + \eta^t x.$$

□

4.3 Nonsymmetric Perturbations for the QAP

In this section we develop for the QAP a parallel approach to the nonsymmetric perturbation theory described above. For doing this we first set $D = -C$ and consider the following relaxation of QAP which is equivalent to (P).

$$\textbf{(P)} \qquad \begin{array}{ll} \min & f(X) = \mathrm{tr} AXBX^t - 2DX^t \\ \text{s.t.} & XX^t = I \end{array}$$

with the same assumptions as in the previous chapter, that is the elimination step was already done.

First let us look at the pure quadratic form of (P) which is

$$
\begin{array}{lll}
(\mathbf{R}) & \min & q(X) = \mathrm{tr} AXBX^t \\
 & \text{s.t.} & XX^t = I
\end{array}
$$

The Lagrangian function of (R) is

$$\mathcal{L}(X,S) = q(X) + \mathrm{tr} SXX^t$$

and the optimality conditions are

$$AXB = SX. \tag{4.3.27}$$

The left hand side of (4.3.27) can be interpreted as a linear operator $Q : \Re^{n\times n} \to \Re^{n\times n}$, $Q(\cdot) = A \cdot B$. We can think of $Q(\cdot)$ being associated with a matrix $\tilde{Q}$, containing the Kronecker product of A and B, that is

$$\tilde{Q} = (A \otimes B).$$

So if we consider S as an eigenvalue matrix of $Q(\cdot)$ then (4.3.27) will represent the eigenvector equation for Q, with eigenvector X.

Moreover, if we take the inner product of (4.3.27) with X and use the fact $XX^t = I$ then

$$\mathrm{tr} AXBX^t = \mathrm{tr} S.$$

That means the optimal value for (R) is equal to the trace of the Lagrange multiplier S.

This fact becomes even clearer if we look at (4.3.27) at a solution X_0 of (R). By Theorem 2.1 we know that the minimum is attained for $X_o = PQ^t$ with P and Q containing the eigenvectors of A and B, respectively, and P and Q ordering the eigenvalues in a way that we achieve a minimal scalar product. Using this fact (4.3.27) becomes

$$APQ^tB = SPQ^tI$$

or

$$PD_AP^tPQ^tQD_BQ^t = SPQ^t$$

which is

$$PD_AD_BQ^t = SPQ^t$$

where D_A and D_B denote diagonal matrices containing the eigenvalues of A and B, respectively.

Premultiplying the above by P^t and postmultiplying it with Q we get

$$D_A D_B = P^t S P.$$

This shows the connection between the optimal value of (R), which is $\mathrm{tr} D_A D_B$ and the spectrum of the Lagrange multiplier S, i.e. the eigenvalues of S are the specific eigenvalues of the Kronecker product $\tilde{Q}$ which yields the minimal scalar product.

Furthermore, we can rewrite the Lagrange function as

$$\mathcal{L}(X, S) = q(X) - \mathrm{tr} X S X^t$$

since $\mathrm{tr} S X X^t = \mathrm{tr} X S X^t$ and S is symmetric. Applying the same ideas as above to this Lagrange function yields

$$D_A D_B = Q^t S Q$$

and clearly confirms the relation between Lagrange multipliers of (R) and the spectral information of $\tilde{Q}$.

Now we shall perturb the above argument. We consider the linear operator

$$K(X, Y) = \begin{pmatrix} AXB - IYD \\ IXD^t + IYT \end{pmatrix} \tag{4.3.28}$$

which is a nonsymmetric perturbation of $Q(\cdot)$. So $K(\cdot, \cdot)$ can be associated with a 2×2 block matrix made up of the Kronecker product blocks in the following way

$$\tilde{K} = \begin{pmatrix} A \otimes B & -I \otimes D \\ I \otimes D^t & I \otimes T \end{pmatrix}. \tag{4.3.29}$$

Let $\mathcal{S}$ denote the set of Lagrange multiplier matrices with associated set of stationary points $\mathcal{S}_\mu(S)$, i.e. $X \in \mathcal{S}_\mu(S)$ means $XX^t = I$, $AXB - D = SX$ which are the optimality conditions for (P). This leads to the following result which is an analogue to Theorem 4.4.

Theorem 4.5 *Let Q, K and $f(\cdot)$ be defined as above. Then $S \in \mathcal{S}$ with $X \in \mathcal{S}_\mu(S)$, if and only if S is an eigenmatrix of K with associated eigenvector*

$$\begin{pmatrix} X \\ I \end{pmatrix}, \quad XX^t = I,$$

in which case $f(X) = \mathrm{tr}\ T$.

Proof. It is easy to see that

$$\tilde{K} \cdot \begin{pmatrix} X \\ I \end{pmatrix} = S \cdot \begin{pmatrix} X \\ I \end{pmatrix}$$

is equivalent to

$$\begin{aligned} AXB - D &= SX \\ XD^T + T &= S. \end{aligned}$$

This and $XX^t = I$ in turn mean $X \in \mathcal{S}_\mu(S)$. Furthermore,

$$T = S - XD^t = AXBX^t - DX^t - XD^t,$$

since $XX^t = I$, so that tr $T = f(X)$. □

Since we have a characterization of optimality for (P) it is possible to develop a More-Sorensen type algorithm [35]. It is well known that this algorithm solves the trust region problem very efficiently. To generate an algorithm like that for solving (P) we have to look at tr T and have to maintain at the same time $XX^t = I$, i.e. we have to minimize tr T while solving the eigenvector equation.

We should also mention that we can easily relax (P) to a problem with only one constraint and then apply a More-Sorensen type algorithm directly. Note that

$$XX^t = I \Rightarrow \text{ tr } XX^t = n. \tag{4.3.30}$$

Therefore, after the elimination, one could consider a relaxed problem for (P) of the form

$$\begin{array}{ll} \min & f(X) = \text{ tr } AXBX^t + 2CX^t \\ s.t. & g_t(X) = \ tr\ XX^t = n \end{array}$$

This is now a trust region problem with a quadratic objective and a single quadratic constraint.

Chapter 5

Parametric Trust Region Bounds

This chapter proposes a parametric approach to solve (P). With border perturbations on A and B we will include the linear part of the relaxation in a quadratic problem. The numerical results are given in Section 5.3 and are quite promising. We conclude this chapter with the theory of indefinite inner products that might be helpful in getting a closer approximation to (P).

5.1 Symmetric Border Perturbations

We still consider the following relaxation of QAP throughout this chapter.

$$\begin{array}{lll} (\mathbf{P}) & \min & f(X) = \mathrm{tr} AXBX^t + 2CX^t \\ & \text{s.t.} & XX^t = I \end{array}$$

where A, B are (real) symmetric $n \times n$ matrices and $C \in \Re^{n \times n}$ with $r := \text{rank } C$. We assume again that the elimination of the constraints was already done.

Let

$$C = ab^t$$

be a full rank decomposition of C with $a, b \in \Re^{n \times r}$. With this factorization of C we define the following symmetric border perturbations

$$\hat{A}_S = \begin{pmatrix} \alpha & a^t \\ a & A \end{pmatrix} \text{ and } \hat{B}_S = \begin{pmatrix} \beta & b^t \\ b & B \end{pmatrix}$$

on A and B, respectively. $\alpha, \beta \in \Re^{r \times r}$ and a, b come from the full rank decomposition described above. Furthermore, denote Z as

$$Z = \begin{pmatrix} w & u^t \\ v & Y \end{pmatrix} \tag{5.1.31}$$

with $u, v \in \Re^{n\times r}$ and $w \in \Re^{r\times r}$. Then we can introduce the following $(n+r)$ dimensional eigenvalue problem

$$(\mathbf{Q_S}) \qquad \begin{array}{ll} \min & q_S(X) = \mathrm{tr}\hat{A}_S Z \hat{B}_S Z^t \\ \text{s.t.} & ZZ^t = I_{n+r}. \end{array}$$

Since Z depends on the choice of α and β we will also denote it by $Z(\alpha, \beta)$. The following will show the relation between (P) and (Q_S).

But first we state a lemma regarding the limiting behavior of $Z(\alpha, \beta)$. Therefore, we look at α and β in the following form. Let again $r = \mathrm{rank}\ C$ and define

$$\alpha = \begin{pmatrix} rt_1 & & & & \\ & (r-1)t_1 & & 0 & \\ & & \ddots & & \\ & 0 & & 2t_1 & \\ & & & & t_1 \end{pmatrix} \tag{5.1.32}$$

and

$$\beta = \begin{pmatrix} rt_2 & & & & \\ & (r-1)t_2 & & 0 & \\ & & \ddots & & \\ & 0 & & 2t_2 & \\ & & & & t_2 \end{pmatrix}, \tag{5.1.33}$$

with $t_1, t_2 \in \Re$.

Lemma 5.1 *Suppose that $Z(\alpha, \beta)$ solves (Q_S). Then for either*

(i) $t_1 \to -\infty$ and $t_2 \to +\infty$ or

(ii) $t_1 \to +\infty$ and $t_2 \to -\infty$

we have

$$Z(\alpha, \beta) \to \begin{pmatrix} I_r & 0 \\ 0 & Y \end{pmatrix}.$$

Proof. We consider case (i), that is $t_1 \to -\infty$ and $t_2 \to +\infty$. Since (Q_S) is a symmetric eigenvalue problem we know by Theorem 2.1 that the optimal objective function value becomes

$$< \lambda(\hat{A}_S), \lambda(\hat{B}_S) >_-$$

where $\lambda(\hat{A}_S)$ and $\mu(\hat{B}_S)$ are the spectrums of $\hat{A}_S$ and $\hat{B}_S$ in nondecreasing and nonincreasing order, respectively. Furthermore the optimum of (Q_S) is attained at

$$Z = PQ^t$$

where P and Q contain the eigenvectors corresponding to $\lambda(\hat{A}_S)$ and $\lambda(\hat{B}_S)$, respectively.

If we choose t_1 sufficiently small and t_2 sufficiently large then we have by Gersgorin's Theorem, see e.g. [27], that the eigenvalues corresponding to the entries of α and β lie in disjoint discs. So

$$\lambda(\hat{A}_S) = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_r, \lambda_{r+1}(\hat{A}_S), \ldots, \lambda_{n+r}(\hat{A}_S))^t$$

is ordered nondecreasingly and

$$\lambda(\hat{B}_S) = (\tilde{\beta}_1, \ldots, \tilde{\beta}_r, \lambda_{r+1}(\hat{B}_S), \ldots, \lambda_{n+r}(\hat{B}_S))^t$$

is ordered nonincreasingly with

$$\alpha_{ii} - r(|a|_i) \leq \tilde{\alpha}_i \leq \alpha_{ii} + r(|a|_i)$$

and

$$\beta_{ii} - r(|b|_i) \leq \tilde{\beta}_i \leq \beta_{ii} + r(|b|_i),$$

where $r(|a|_i)$ and $r(|b|_i)$ denote the sums of the absolute values of row (or column) i, $i = 1, \ldots, r$, of a and b, respectively. Then the eigenvalue decompositions of $\hat{A}_S$ and $\hat{B}_S$ are

$$P \begin{pmatrix} \tilde{\alpha}_1 & & & & & \\ & \ddots & & & 0 & \\ & & \tilde{\alpha}_r & & & \\ & & & \lambda_{r+1}(\hat{A}_S) & & \\ & 0 & & & \ddots & \\ & & & & & \lambda_{n+r}(\hat{A}_S) \end{pmatrix} P^t$$

and

$$Q \begin{pmatrix} \tilde{\beta}_1 & & & & & \\ & \ddots & & & 0 & \\ & & \tilde{\beta}_r & & & \\ & & & \lambda_{r+1}(\hat{B}_S) & & \\ & 0 & & & \ddots & \\ & & & & & \lambda_{n+r}(\hat{B}_S) \end{pmatrix} Q^t,$$

and for $t_1 \to -\infty$

$$P \to \begin{pmatrix} I_r & 0 \\ 0 & P' \end{pmatrix}$$

and for $t_2 \to +\infty$

$$Q \to \begin{pmatrix} I_r & 0 \\ 0 & Q' \end{pmatrix}$$

Therefore,

$$\lim_{\substack{t_1 \to -\infty \\ t_2 \to +\infty}} Z(\alpha, \beta) = PQ^t = \begin{pmatrix} I_r & 0 \\ 0 & Y \end{pmatrix}$$

where $Y = P'Q'^t$.

The same argument works for Case (ii), as well. $\square$

Since we want to make use of this result we will only look at α and β in the form given above. We also assume Case (i), that is $t_1 \ll 0$ and $t_2 \gg 0$.

Let us consider the objective function of (Q_S). By the definition of (Q_S) we have

$$\begin{aligned}
q_S(Z) &= \operatorname{tr} \ \hat{A}_S Z \hat{B}_S Z^t \\
&= \operatorname{tr} \ \begin{pmatrix} \alpha & a^t \\ a & A \end{pmatrix} \begin{pmatrix} w & u^t \\ v & Y \end{pmatrix} \begin{pmatrix} \beta & b^t \\ b & B \end{pmatrix} \begin{pmatrix} w^t & v^t \\ u & Y^t \end{pmatrix} \\
&= \operatorname{tr} \ \alpha w \beta w^t + \alpha w b^t u + a^t v \beta w^t + a^t v b^t u + \\
&\quad\ \alpha u^t b w^t + \alpha u^t B u + a^t Y b w^t + a^t Y B u + \\
&\quad\ a w \beta v^t + a w b^t Y^t + A v \beta v^t + A v b^t Y^t + \\
&\quad\ a u^t b v^t + a u^t B Y^t + A Y b v^t + A Y B Y^t \\
&= \operatorname{tr} \ AYBY^t + 2awb^tY^t + \alpha w \beta w^t + \\
&\quad\ 2\alpha w b^t u + 2a^t v \beta w + 2a^t v b^t u + \alpha u^t B u + \\
&\quad\ A v \beta v^t + 2a^t Y B u + 2AYbv^t \\
&= \operatorname{tr} \ AYBY^t + 2awb^tY^t + \alpha w \beta w^t + g_S(w, u, v, Y)
\end{aligned}$$

with $g_S(w, u, v, Y) = \operatorname{tr} \ 2\alpha w b^t u + 2a^t v \beta w + 2a^t v b^t u + \alpha u^t B u + A v \beta v^t + 2a^t Y B u + 2AYbv^t$, which we call the error term of (Q_S).

We can see that the first two terms in q_S are almost in the form of the objective function of (P) which is

$$f(X) = \operatorname{tr} \ AXBX^t + 2CX^t$$

since as shown before $w \to I_r$ for $t_1 \ll 0$ and $t_2 \gg 0$ and $C = ab^t$.

As solving (Q_S) for fixed t_1 and t_2 is straight forward, we try to use a solution Z of (Q_S) for getting an approximate solution for (P).

Theorem 5.1 *Let (P) and (Q_S) be given as above. Suppose Z solves (Q_S) and $\bar{X}$ solves (P). Then*

$$f(\bar{X}) \geq \mathrm{tr} AYBY^t + 2awb^tY^t + g_S(w,u,v,Y) + \alpha(w - I_r)\beta(w + I_r)^t. \qquad (5.1.34)$$

Proof. Let $\bar{X}$ be the solution of (P) with $\bar{X}\bar{X}^t = I$. Then

$$\bar{Z} = \begin{pmatrix} I_r & 0 \\ 0 & \bar{X} \end{pmatrix}$$

is feasible but not necessarily optimal for (Q_S) and so

$$q(\bar{Z}) = f(\bar{X}) + \alpha\beta \geq q(Z)$$

by optimality of Z .

Thus

$$f(\bar{X}) + \alpha\beta \geq \mathrm{tr} AYBY^t + 2awb^tY^t + g_S(w,u,v,Y) + \alpha w \beta w^t$$

and hence

$$f(\bar{X}) \geq \mathrm{tr} AYBY^t + 2awb^tY^t + g_S(w,u,v,Y) + \alpha(w - I_r)\beta(w + I_r)^t.$$

□

In order to get a good approximation for (P) and so a good bound for the QAP, one has to maximize the right hand side of (5.2.37), which we denote by $h(\alpha, \beta)$, with respect to α and β. So

$$h(\alpha,\beta) = \mathrm{tr} AYBY^t + 2awb^tY^t + g_S(w,u,v,Y) + \alpha(w - I_r)\beta(w + I_r)^t. \qquad (5.1.35)$$

But since α and β are variables in t_1 and t_2, respectively, we can maximize $h(\alpha, \beta)$ with respect to t_1 and t_2. So for getting a good bound for (P) we have to

$$\max_{t_1,t_2} h(\alpha,\beta)$$

which we can do by Newton's method.

The partial derivatives of h are

$$\frac{\partial h}{\partial t_1} = 2wb^tu + u^tBu + (w - I_r)(w + I_r)^t\beta$$

and

$$\frac{\partial h}{\partial t_2} = 2wa^t v + v^t A v + \alpha(w - I_r)(w + I_r)^t.$$

Then the Hessian of h becomes

$$H = \begin{pmatrix} 0 & (w - I_r)(w + I_r)^t \\ (w - I_r)(w + I_r)^t & 0 \end{pmatrix}$$

and since $(w - I_r)(w + I_r)^t \to 0$ for $t_1 \ll 0$ and $t_2 \gg 0$ we see that h is almost linear.

Newton's direction of ascent for maximizing the bound is $H^{-1} \bigtriangledown h$, with

$$H^{-1} = \frac{1}{(w - I_r)(w + I_r)^t} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Thus a Newton step for $(\alpha, \beta)^t$ is

$$\begin{aligned} \begin{pmatrix} \alpha_{k+1} \\ \beta_{k+1} \end{pmatrix} &= \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix} + \begin{pmatrix} \frac{2wa^t v + v^t A v}{(w-I_r)(w+I_r)^t} - \alpha_k \\ \frac{2wb^t u + u^t B u}{(w-I_r)(w+I_r)^t} - \beta_k \end{pmatrix} \\ &= \frac{1}{(w - I_r)(w + I_r)^t} \begin{pmatrix} 2wa^t v + v^t A v \\ 2wb^t u + u^t B u \end{pmatrix}. \end{aligned}$$

We use this method in our numerical experiments which are described in Section 5.3. Note that $h(\alpha, \beta)$ seems to have many local solutions what makes this approach not very satisfying. It mainly depends on the initial guess whether a good local optimum is found. One way to achieve this is shortly discussed in Section 5.3.

5.2 Nonsymmetric Border Perturbations

As we saw in the previous section the main problem we have in using symmetric border perturbations is the large error term which is mainly caused by the terms

$$\operatorname{tr} 2\alpha w b^t u + 2a^t v \beta w \tag{5.2.36}$$

since these do not tend to zero. We can observe that these terms are products of values that go linearly to infinity and linearly to zero. The remaining parts in the error term have a quadratic part that tends to zero what makes them less dominating.

This raises the question whether one can construct the border perturbations in a way to get rid of the terms in (5.2.36). We shall show that this can be done by nonsymmetric perturbations.

Let the full rank decomposition of C be now given by

$$C = -ab^t$$

with $a, b \in \Re^n$. We then perturb A and B nonsymmetrically that is

$$\hat{A}_N = \begin{pmatrix} \alpha & a^t \\ -a & A \end{pmatrix} \quad \text{and} \quad \hat{B}_N = \begin{pmatrix} \beta & b^t \\ -b & B \end{pmatrix}$$

and define the $(n+r)$ dimensional minimization problem

$$\begin{array}{lll} (\mathbf{Q_N}) \qquad & \min & q_N(Z) = \mathrm{tr} \hat{A}_N Z \hat{B}_N^t Z^t \\ & \text{s.t.} & ZZ^t = I_{n+r} \end{array}$$

with Z defined as in (5.1.31).

Since $\hat{A}_N$ and $\hat{B}_N$ are not symmetric we can not state a limiting result for $Z(\alpha, \beta)$ in the same way as we did in Lemma 5.1 and furthermore (Q_N) is not directly solvable like (Q_S).

But the advantage of this approach is that the objective function of (Q_N) becomes

$$\begin{aligned} q_N(Z) &= \mathrm{tr} \quad \hat{A}_N Z \hat{B}_N^t Z^t \\ &= \mathrm{tr} \quad \begin{pmatrix} \alpha & a^t \\ -a & A \end{pmatrix} \begin{pmatrix} w & u^t \\ v & Y \end{pmatrix} \begin{pmatrix} \beta & b^t \\ -b & B \end{pmatrix} \begin{pmatrix} w^t & v^t \\ u & Y^t \end{pmatrix} \\ &= \mathrm{tr} \quad \alpha w \beta w^t + \alpha w b^t u + a^t v \beta w + a^t v b^t u - \\ &\qquad \alpha u^t b w + \alpha u^t B u - a^t Y b w + a^t Y B u - \\ &\qquad a w \beta v^t - a w b^t Y^t + A v \beta v^t + A v b^t Y^t + \\ &\qquad a u^t b v^t - a u^t B Y^t - A Y b v^t + A Y B Y^t \\ &= \mathrm{tr} \quad AYBY^t - 2awb^t Y^t + \alpha w \beta w^t + \\ &\qquad 2a^t v b^t u + \alpha u^t B u + A v \beta v^t \\ &= \mathrm{tr} \quad AYBY^t + 2awb^t Y^t + \alpha w \beta w^t + g_N(w, u, v, Y) \end{aligned}$$

with $g_N(w, u, v, Y) = \mathrm{tr}\ 2a^t v b^t u + \alpha u^t B u + \beta v^t A v$.

We see that some terms cancelled and so the error term g_N becomes at least smaller than g_S in number of terms. It also seems that we got rid of the dominating terms of g_S. This fact was shown by extensive numerical experiments where the errors g_S and g_N were compared.

We can restate Theorem 5.1 for the nonsymmetric case in the following form.

Theorem 5.2 *Let (P) and (Q_N) be given as above. Suppose Z solves (Q_N) and $\bar{X}$ solves (P). Then*

$$f(\bar{X}) \geq \mathrm{tr} AYBY^t + 2awb^t Y^t + g_N(w, u, v, Y) + \alpha(w - I_r)\beta(w + I_r)^t. \tag{5.2.37}$$

Proof. The proof of Theorem 5.1 applies. $\square$

We now have to find a way to solve (Q_N) since it is not an eigenvalue problem as it is not symmetric. One way would be to apply the symmetrization techniques described in Section 2.3. For $\hat{A}$ and $\hat{B}$ we compute

$$\hat{A}_+ = \begin{pmatrix} \alpha & 0 \\ 0 & A \end{pmatrix}, \quad \hat{A}_- = \begin{pmatrix} 0 & a^t \\ -a & 0 \end{pmatrix}$$

and

$$\hat{B}_+ = \begin{pmatrix} \beta & 0 \\ 0 & B \end{pmatrix}, \quad \hat{B}_- = \begin{pmatrix} 0 & b^t \\ -b & 0 \end{pmatrix}.$$

Then the positive Hermitian parts of A and B are

$$\tilde{A}_+ = \begin{pmatrix} \alpha & ia^t \\ -ia & A \end{pmatrix}$$

and

$$\tilde{B}_+ = \begin{pmatrix} \beta & ib^t \\ -ib & B \end{pmatrix},$$

respectively. Following Theorem 2.3 we can bound (Q_N) by

$$(\tilde{\mathbf{Q}}_+) \qquad \begin{array}{ll} \min & q_N(Z) = \mathrm{tr}\tilde{A}_+ Z_c \tilde{B}_+ Z_c^* \\ \text{s.t.} & Z_c Z_c^* = I_{n+r} \end{array}$$

since we minimize over unitary matrices now. The subscript c on Z_c denotes that Z_c is complex.

By solving $(\tilde{Q}_+)$ as Hermitian $(n+r)$ dimensional eigenvalue problem we get Z_c for which

$$q_+(Z_c) \leq q_N(Z)$$

where Z solves (Q_N). Since Z_c is complex the error term which results from using Z_c for (Q_N) is different to $g_N(w,u,v,X)$. By denoting $\tilde{a} = ia$ and $\tilde{b} = ib$ we get for $q_N(Z_c)$

$$\begin{aligned} q_N(Z_c) &= \mathrm{tr} \quad \hat{A}_+ Z_c \hat{B}_+ Z_c^* \\ &= \mathrm{tr} \quad \begin{pmatrix} \alpha & \tilde{a}^t \\ \tilde{a} & A \end{pmatrix} \begin{pmatrix} w_c & u_c^* \\ v_c & Y_c \end{pmatrix} \begin{pmatrix} \beta & \tilde{b}^t \\ \tilde{b} & B \end{pmatrix} \begin{pmatrix} w_c^* & v_c^* \\ u_c & Y_c^* \end{pmatrix} \end{aligned}$$

and we are exactly where we have already been with the symmetric border perturbations. In order to minimize the amount of the error we can apply the same method as we did for the symmetric border perturbations in the previous section.

Size	Opt.	GLB	EVB1	EVB2	EVB3	EVB4	EVB5	
12	578	439	-909	446	498	472	488	(525)
15	1150	963	-1745	927	1002	973	1122	(1024)
20	2570	2057	-3198	2075	2286	2196	2185	(2436)
30	6124	4539	-7836	4982	5443	5266	5229	(5834)
10	4954	3586	-12791	2774	4541	4079	**4960**	(6106)
10	8082	6139	-13750	6365	7617	7211	7113	(9749)
10	8649	7030	-14647	6869	8233	7837	7763	(8685)
10	8843	6840	-13546	7314	8364	8006	7670	(9171)
10	9571	7627	-15748	8095	8987	8672	8479	(9709)
10	936	878	-777	885	895	887	887	(996)
12	1652	1536	-1407	1562	1589	1573	1578	(1633)
14	2724	2492	-2488	2574	2630	2609	2600	(2678)
16	3720	3358	-3271	3518	3594	3560	3556	(3732)
18	5358	4776	-4422	5035	5150	5104	5094	(5357)
20	6922	6166	-5785	6533	6678	6625	6607	(6954)

Table 5.3.0: Lower Bounds for QAP

5.3 Numerical Results

We now discuss the numerical experiments we did to compare the new parametric bound with several existing bounding strategies. Therefore, we denote the new bound which results from the symmetric border perturbations by EVB5. The numerical results are summarized in Table 5.3. The first group of QAP instances is taken from [36] and represents the most used data for numerical experiments for QAP. The second group of instances, all of size $n = 10$, is taken from [7]. The problems were randomly generated with the following characteristics: $a_{ij} = a_{ji}$ are uniformly generated in $[0, 10]$; b_{kl} represents the squared euclidean distance of x_k, x_l where x_i, $i = 1, ..., n$ are uniformly distributed points in $[0, 10] \times [0, 10]$; the c_{ik} are again uniformly distributed in $[0, 10]$. The last group of instances is taken from [25] and is randomly generated in the following way. The instances are pure quadratic, that is $C = 0$. B is a symmetric matrix with entries drawn uniformly from $1, ..., 10$. Finally, A represents the l_1-distance of a connected cellular complex of n cells in the plane. These data therefore imitate the QAP as a floor planning problem.

The meaning of the columns in Table 5.3 is as follows. The first column gives the size of the problem instances while the second column provides the best known solution for the problem. For problems with $n \leq 15$ this solution is provably optimal. Then we include the classical bound of Gilmore and Lawler (GLB). The following columns show the eigenvalue bounds discussed in this work. EVB1, EVB2, EVB3 represent the bounds introduced in Section 2.2. EVB4 is the bound which bases on the projection to PQAP. The bound derived in Section 5.1. and given by (5.1.34) is denoted by EVB5.

We should give some technical notes about the computation of EVB5. We calculated the lower bound given by (5.1.34) and then maximized $h(\alpha, \beta)$ in (5.1.35) with respect to α and β. As initial guess we chose an arbitrary t_1 and then took $t_2 = t_1 \frac{||a||}{||b||}$. This rather heuristic initialization for α and β proved to be very good. This relies on the fact that we know a and b of the dominating terms (5.2.36) in advance. By this choice of α and β the two terms become of the same size which makes the error of smaller size. We did this procedure for five different initializations for t_1. In each maximization we did five iterations and chose the best value for the lower bound (5.1.34). So EVB5 represents the best lower bound we found. The value in parentheses describes the bound without correction by $g_S(w, u, v, Y)$ and tr $\alpha(w - I_r)\beta(w - I_r)^t$ as given in Theorem 5.1. The value is

$$\mathrm{tr} AYBY^t + 2awb^tY^t$$

and therefore an upper bound of (P).

Note that the first value of EVB5 in the second group of problem instances is larger than the optimum. This seems to come from round off errors which normally only occur for large α and β.

The comparisons of the different bounds show that EVB5 is a competitive bound compared to EVB1, EVB2 and EVB3. Unfortunately, it is not often an improvement to EVB4. But the values in parentheses let hope that a smaller correction could yield bounds which are very close to the optimal values. Moreover, we have not used any kind of shifts or reductions yet to improve EVB5. So these first numerical results are still promising and encouraging for future work into this direction.

We will introduce some theory which might be helpful for this in the next section.

5.4 Indefinite Inner Products and (Q_N)

We tried several things to get an at least approximate solution for (Q_N). The key observation for this was that for

$$K = \begin{pmatrix} -I_r & 0 \\ 0 & I_n \end{pmatrix} \tag{5.4.38}$$

$\hat{A}_N$ and $\hat{B}_N$ are $K-$*symmetric*, that is

$$\tilde{A} = \hat{A}K = \begin{pmatrix} -\alpha & a^t \\ a & A \end{pmatrix}$$

and

$$\tilde{B} = \hat{B}K = \begin{pmatrix} -\beta & b^t \\ b & B \end{pmatrix}.$$

So multiplying $\hat{A}_N$ and $\hat{B}_N$ with a matrix which is very similar to I, $|K| = I$, leads to symmetry.

One can reformulate (Q_N) as

$$(\tilde{\mathbf{Q}}_{\mathbf{N}}) \qquad \begin{array}{ll} \min & \mathrm{tr}\tilde{A}KZ\tilde{B}KZ^t \\ \text{s.t.} & ZZ^t = I_{n+r} \end{array}$$

since $KK = I$.

Unfortunately, we could not find a way to bound (Q_N) or $(\tilde{Q}_N)$ from below in a way, that the resulting bound is close to the optimal solution. It is simple to bound (Q_N) and $(\tilde{Q}_N)$ from above by just using $Z = PQ^t$ where P and Q come from the orthogonal eigenvalue decompositions of $\tilde{A}$ and $\tilde{B}$, respectively. But an upper bound - especially if we do not know whether it is sharp - does not help to bound QAP. We shall mention here, that extensive numerical results showed that this upper bound could be sharp.

The lower bounds we achieved came from relaxing the minimization of (Q_N) or $(\tilde{Q}_N)$ from one variable Z to two variables Z_1 and Z_2, e.g. minimizing

$$\mathrm{tr}\hat{A}Z_1\hat{B}Z_2^t$$

such that $Z_iZ_i^t = I, \;\; i = 1, 2$. But as mentioned above, these bounds are very poor.

An other approach for tackling (Q_N), which uses the fact that $\hat{A}$ and $\hat{B}$ are almost symmetric, is the theory of indefinite inner products. In the remainder of this chapter we want to introduce the concept of indefinite inner products and thereby follow Gohberg et. al. [22].

In traditional linear algebra the concept of orthogonality, length and angles is defined by a definite inner product. We replace this definite scalar product by an indefinite one, which clearly changes the geometry substantially. Note that we restrict ourselves to the real case.

The standard scalar product in $\Re^n$ is defined by $(\cdot, \cdot)$. Thus for $x, y \in \Re^n$

$$(x, y) \equiv y^t x.$$

We call a function $[\cdot, \cdot]$ from $\Re^n \times \Re^n$ to $\Re^n$ an *indefinite inner product* if the following axioms are satisfied:

(i) $[\alpha x_1 + \beta x_2, y] = \alpha[x_1, y] + \beta[x_2, y]$ for all $\alpha, \beta \in \Re$ and $x_1, x_2, y \in \Re^n$;

(ii) $[x, y] = [y, x]$ for all $x, y \in \Re^n$;

(iii) if $[x, y] = 0$ for all $y \in \Re^n$, then $x = 0$.

The axioms guarantee linearity in the first argument (i), antisymmetry (ii) and non-degeneracy (iii). Thus the indefinite scalar products satisfies all the properties of the standard scalar product. One possible exception is that $[x, x]$ may be non-positive for $x \neq 0$.

An indefinite scalar product can be characterized by

$$[x, y] = (Hx, y), \qquad x, y, \in \Re^n$$

where $H \in \Re^n \times \Re^n$ is an invertible symmetric matrix. Based on this characterization we develop an analogous theory as for the standard scalar products and show their application for (Q_N).

When a scalar product is introduced on $\Re^n$ then certain matrices have symmetries defined by the inner product. A definite inner product leads to the usual classes of symmetric, orthogonal and normal matrices. We now define corresponding classes for the indefinite scalar product. Be $[\cdot, \cdot]$ again the indefinite inner product associated with H and be A an n by n matrix considered as linear transformation on $\Re^n$. The *H-transpose* of A is the unique $n \times n$ matrix, denoted by $A^{[t]}$, which satisfies

$$[Ax, y] = [x, A^{[t]}y] \tag{5.4.39}$$

for all $x, y \in \Re^n$. We can also describe the H-transpose of A as the transpose with respect to $[\cdot, \cdot]$. Expressing (5.4.39) in terms of H we can derive that

$$A^{[t]} = H^{-1}A^tH,$$

where A^t is the usual transpose of A.

It is natural to define a matrix A as *H-selftranspose* if

$$A = A^{[t]}.$$

In a similar way, a matrix U is said to be *H-orthogonal* if it is invertible and

$$U^{-1} = U^{[t]},$$

and a matrix N is *H-normal* if

$$NN^{[t]} = N^{[t]}N.$$

Analogously to results for definite inner products, H-selftranspose and H-orthogonal matrices are also H-normal.

As mentioned in the beginning of this section, the theory of indefinite inner products seems to be helpful for solving (Q_N). So it is easy to see that K as defined in (5.4.38) is symmetric and invertible and can thus be used for an indefinite inner product. For this scalar product we can make the following simple observation.

Lemma 5.2 *$\hat{A}$ and $\hat{B}$ are K-selftranspose.*

Proof. Since $\hat{A}$ and $\hat{B}$ have the same structure, we prove the lemma for $\hat{A}$. First observe that $K^{-1} = K$ and so

$$\begin{aligned} K\hat{A}K &= \begin{pmatrix} I_r & 0 \\ 0 & I_n \end{pmatrix} \begin{pmatrix} \alpha & a^t \\ -a & A \end{pmatrix} \begin{pmatrix} I_r & 0 \\ 0 & I_n \end{pmatrix} \\ &= \begin{pmatrix} \alpha & a^t \\ -a & A \end{pmatrix} \\ &= A^{[t]}. \end{aligned}$$

□

So in some sense (Q_N) is a symmetric problem with respect to an indefinite inner product. This gives hope that a closer study of indefinite scalar products in connection with nonsymmetrically perturbed eigenvalue problems like (Q_N) yields interesting theory and better results.

Chapter 6

Conclusions and Perspectives

We studied the application of trust region ideas to the quadratic assignment problem. It seems that although the quadratic constraints of the relaxation (P) resemble in form the single quadratic constraint of a trust region problem, that trust region methods cannot be applied directly for solving (P). But we were able to extend existing results for one quadratic constraint to $(n^2 + n)/2$ quadratic constraints.

One way to solve (P) is via nonsymmetric border perturbations as proposed in Chapter 4. Since we could characterize optimality for (P), similar algorithms for solving (P) as for solving trust region problems can be derived. This was shortly discussed at the end of Chapter 4 and leads to starting points for future research.

A promising approach is the parametric one introduced in Chapter 5. Despite a relatively rough approximation of (P), the numerical results were surprisingly good. It seems that future work in this direction, especially in connection with indefinite inner products, can yield closer approximations of (P). This would be an important step towards solving quadratic objective functions with quadratic constraints even though the constraints here have a very special structure. We did not only obtain new theoretical results for the trust region problem but also for the theory of optimization over partial orders.

This thesis also showed that spectral information is a key in optimization. Since this fact is very often unconsidered we think that it should be mentioned here.

Bibliography

[1] F. ALIZADEH. Semidefinite programming: duality theory, eigenvalue optimization and comniatorial applications. Presented at the Fourth SIAM Conference on Optimization, 1992.

[2] A. BEN-ISRAEL and T.N.E. GREVILLE. *Generalized Inverses: Theory and Applications.* Wiley-Interscience, 1974.

[3] D.P. BERTSEKAS. *Constrained optimization and Lagrange multipliers.* Academic Press, New York, NY, 1982.

[4] S. BOYD. Method of centers of minimizing generalized eigenvalues. Presented at the Fourth SIAM Conference on Optimization, 1992.

[5] R. E. BURKARD and J. OFFERMANN. Entwurf von Schreibmaschinentastaturen mittels quadratischer Zuordnungsprobleme. *Z. Operations Res.*, 21:B121–B132, 1977.

[6] R.E. BURKARD. Locations with spatial interactions: the quadratic assignment problem. In P.B. Mirchandani and R.L. Francis, editors, *Discrete Location Theory.* John Wiley, Berlin, 1991.

[7] P. CARRARESI and F. MALUCELLI. A new lower bound for the quadratic assignment problem. Technical Report TR-7/88, Universita di Pisa, 1988.

[8] J.E. DENNIS Jr. and R.B. SCHNABEL. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, NJ, 1983. Russian edition, Mir Publishing Office, Moscow, 1988, O. Burdakov, translator.

[9] A. N. ELSHAFEI. Hospital lay-out as a quadratic assignment problem. *Operational Research Quarterly*, 28:167–179, 1977.

[10] K. FAN. Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci. U.S.A.*, 35:652–655, 1951.

[11] P.A. FILLMORE and J.P. WILLIAMS. Some convexity theorems for matrices. *Glasgow Mathematical Journal*, pages 110–117, 1971.

[12] G. FINKE, R.E. BURKARD, and F. RENDL. Quadratic assignment problems. *Annals of Discrete Mathematics*, 31:61–82, 1987.

[13] R. FLETCHER. Semi-definite matrix constraints in optimization. *SIAM J. Control and Optimization*, 23:493–513, 1985.

[14] R. FLETCHER. *Practical Methods of Optimization.* John Wiley & Sons, New York, NY, 1987.

[15] G.E. FORSYTHE and G.H. GOLUB. On the stationary values of a second-degree polynomial on the unit sphere. *SIAM J. Applied Mathematics*, 13:1050–1068, 1965.

[16] A. M. FRIEZE and J. YADEGAR. On the quadratic assignment problem. *Discrete Applied Mathematics*, 5:89–98, 1983.

[17] W. GANDER. On the linear least squares problem with a quadratic constraint. Technical Report STAN-CS-78-697, Department of Computer Science, Stanford University, Stanford, CA, 1978.

[18] W. GANDER. Least squares with a quadratic constraint. *Numer. Math.*, 36:291–307, 1981. This is the abbreviated version of the technical report published in 1978.

[19] W. GANDER, G.H. GOLUB, and U. von MATT. A constrained eigenvalue problem. *Linear Algebra and its Applications*, 114/115:815–839, 1989.

[20] D.M. GAY. Computing optimal locally constrained steps. *SIAM J. Sci. Statist. Comput.*, 2:186–197, 1981.

[21] P.C. GILMORE. Optimal and suboptimal algorithms for the quadratic assignment problem. *SIAM Journal on Applied Mathematics*, 10:305–313, 1962.

[22] I. GOHBERG, P. LANCASTER, and L. RODMAN. *Matrices and indefinite scalar products.* Birkhaeuser, Basel, 1983.

[23] B. GOLLAN. Eigenvalue perturbations and nonlinear parametric optimization. *Mathematical Programming Studies*, 30:67–81, 1987.

[24] S.W. HADLEY. *Continuous optimization approaches for the quadratic assignment problem.* PhD thesis, University of Waterloo, Waterloo, Canada, 1989.

[25] S.W. HADLEY, F. RENDL, and H. WOLKOWICZ. A new lower bound via projection for the quadratic assignment problem. *Mathematics of Operations Research*, 1989. to appear.

[26] S.W. HADLEY, F. RENDL, and H. WOLKOWICZ. Symmetrization of nonsymmetric quadratic assignment problems and the Hoffman-Wielandt inequality. *Linear Algebra and its Applications*, 58:109–124, 1992.

[27] R. HORN and C. JOHNSON. *Matrix Analysis.* Cambridge University Press, New York, 1985.

[28] T. C. KOOPMANS and M. J. BECKMANN. Assignment problems and the location of economic activities. *Econometrica*, 25:53–76, 1957.

[29] P. LANCASTER and M. TISMENTESKY. *The Theory of Matrices: with applications.* Academic Press, New York, second edition, 1985.

[30] E. LAWLER. The quadratic assignment problem. *Management Science*, 9:586–599, 1963.

[31] R. LOEWY and H. SCHNEIDER. Positive operators on the n-dimensional ice cream cone. *J. Math. Anal. Appl.*, 49:375–392, 1975.

[32] D.G. LUENBERGER. *Optimization by Vector Space Methods.* John Wiley, 1969.

[33] D.G. LUENBERGER. *Linear and Nonlinear Programming.* Reading. Addison-Wesley, Massachusetts, second edition, 1984.

[34] A.W. MARSHALL and I. OLKIN. *Inequalities: Theory of Majorization and its Applications.* Academic Press, New York, NY, 1979.

[35] J.J. MORÉ and D.C. SORENSEN. Computing a trust region step. *SIAM J. Sci. Statist. Comput.*, 4:553–572, 1983.

[36] C.E. NUGENT, T.E. VOLLMAN, and J. RUML. An experimental comparison of techniques for the assignment of facilities to locations. *Operations Research*, 16:150–173, 1968.

[37] M.L. OVERTON and R.S. WOMERSLEY. On the sum of the largest eigenvalues of a symmetric matrix. *SIAM J. Matrix Analysis and Applications*, 13:41–45, 1992.

[38] F. RENDL and H. WOLKOWICZ. Applications of parametric programming and eigenvalue maximization to the quadratic assignment problem. *Mathematical Programming*, 1989. to appear.

[39] C. ROUCAIROL. A parallel branch and bound algorithm for the quadratic assignm ent problem. *Discrete Applied Mathematics*, 15:211–225, 1987.

[40] S. SAHNI and T. GONZALES. P-complete approximation problems. *Journal of ACM*, 23:555–565, 1976.

[41] D.C. SORENSEN. Newton's method with a model trust region modification. *SIAM J. Numer. Anal.*, 19:409–426, 1982.

[42] L. STEINBERG. The backboard wiring problem: a placement algorithm. *SIAM Review.*, 3:37–50, 1961.

[43] R.J. STERN and H. WOLKOWICZ. Trust region problems and nonsymmetric perturbation theory. Technical Report CORR 91-10, University of Waterloo, Waterloo, Canada, 1991.

[44] K.A. WILLIAMSON. *A robust trust region algorithm for nonlinear programming.* PhD thesis, Department of Mathematical Sciences, Rice University, Houston, TX, 1990.

[45] Y. YUAN. A dual algorithm for minimizing a quadratic function with two quadratic constraints. *Mathematical of Computational Mathematics*, 9:348–359, 1991.