SQ^2P, Sequential Quadratic Constrained Quadratic Programming

Serge Kruk *   Henry Wolkowicz †

August 7, 1998
University of Waterloo
Department of Combinatorics and Optimization
Waterloo, Ontario N2L 3G1, Canada
Research Report CORR 97-01

Key words: Trust-Region Subproblem, Semidefinite Programming, Sequential Quadratic Programming, Interior-point methods.

Abstract

We follow the popular approach for unconstrained minimization, i.e. we develop a local quadratic model at a current approximate minimizer in conjunction with a trust region. We then minimize this local model in order to find the next approximate minimizer. Asymptotically, finding the local minimizer of the quadratic model is equivalent to applying Newton's method to the stationarity condition.

For constrained problems, the local quadratic model corresponds to minimizing a quadratic approximation of the objective subject to quadratic approximations of the constraints (Q^2P), with an additional trust region. This quadratic model is intractable in general and is usually handled by using linear approximations of the constraints and modifying the Hessian of the objective using the Hessian of the Lagrangean, i.e. a SQP approach. Instead, we solve the Lagrangean relaxation of Q^2P using semidefinite programming. We develop this framework and present an example which illustrates the advantages over the standard SQP approach.

Contents

*Research supported by a Post-Graduate Scholarship from NSERC. E-mail sgkruk@acm.org.

†Research supported by The Natural Sciences and Engineering Research Council of Canada. E-mail henry@orion.uwaterloo.ca.

1 Introduction

A modern popular approach for unconstrained minimization of a function, $f(x), x \in \mathbb{R}^n$, is to build a quadratic model at a local estimate $x^{(k)}$. This model is usually convex; either the approximate Hessians are forced to be positive definite, as in modified-Newton and quasi-Newton methods, or a trust region is added to convexify the problem as in the Levenberg-Marquadt approach. (For unified views of unconstrained optimization, see Nazareth [27], [28], Fletcher [13], or Gill et al. [16].) In this paper we extend this modeling approach in a direct way to constrained minimization, i.e. we form a local quadratic model and then try to solve this model as best we can. We do this by solving the Lagrangean relaxation using semidefinite programming. This leads to a modification of *SQP* methods which we call SQ^2P.

In its most general formulation, constrained optimization is concerned with the nonlinear equality and inequality programs (and mixtures of both)

$$NEP \qquad \min\left\{f(x) \mid h(x) = 0, x \in \mathbb{R}^n\right\} \qquad \text{and} \quad NLP \qquad \min\left\{f(x) \mid g(x) \leq 0, x \in \mathbb{R}^n\right\},$$

where $f : \mathbb{R}^n \to \mathbb{R}$, and $h, g : \mathbb{R}^n \to \mathbb{R}^m$ and all functions are sufficiently smooth. As in the unconstrained case, a local (true) quadratic model (subproblem) is formed. It consists of a second order approximation of both objective and constraints, Q^2P. In addition, a quadratic trust region constraint can be added. This model is in general intractable due to lack of convexity. The usual approach is to approximate the quadratic constraints by linear constraints and modify the Hessian of the objective function using the Hessian of the Lagrangean, i.e. we obtain a quadratic programming *QP* subproblem and we use the well known family of solution methods, Sequential Quadratic Programming (*SQP*). (See e.g. the recent survey by Boggs and Tolle [5].)

We will argue that, even though *SQP* has been honed, over the years, into an efficient tool, the underlying model of the method can be improved. Specifically, we will show that some difficulties arising from the QP subproblem, which require special considerations in any implementation, are automatically taken care of when a different subproblem is considered.

The challenges facing an SQP implementation include the infeasibility or unboundedness of the subproblems, the accuracy of the Lagrange multiplier estimates, and the loss of superlinearity due to damped Newton steps (the so-called Maratos effect). Each of these problems has received attention since the development of SQP in the sixties and a number of solutions are known. Yet much research is still being done on the subject.

In this paper we work directly with the true quadratic subproblem Q^2P. We solve the (tractable) Lagrangean relaxation of Q^2P efficiently using semidefinite programming (SDP). This relaxation lies between the quadratic Q^2P model and the QP relaxation. In addition, it provides a surprisingly good approximation for the quadratic model. Our intention is to describe how this different subproblem can elegantly do away with some of the difficulties; we then sketch how the subproblem can be solved using interior-point methods. This is not a computational paper describing a fully defined implementation, but rather a sketch of how interior-point algorithms and semidefinite relaxations can be used in the context of constrained optimization.

Semidefinite programming has been a very successful tool for solving or approximating combinatorial optimization problems; see, for example [2, 18, 32, 20]. It also has found applications in control theory. An overview of semidefinite programming and of many of its applications is found in Vandenberghe and Boyd [40], see also [29, 31, 6]. Most of the success is related to the links between the Lagrangean and semidefinite relaxations, as discussed in [36, 32].

Moreover, efficient numerical implementations for *SDP* have appeared recently: SDPpack (Alizadeh et al.), SDPSOL (Boyd and Vandenberghe), SDPA (Fujisawa et al.), SDPT3 (Toh et al.), CSDP (Borchers) and a Matlab toolbox (Rendl et al.) URLs can be found at the *SDP* homepage of C. Helmberg http://www.zib.de/helmberg/semidef.html. Several of these packages exploit sparsity and solve exceptionally large problems. For example, Ye [4] published results on problems of the order 10000 variables.

In Section 2 we revisit a classical continuous optimization question, the Trust-Region subproblem, which we choose to view as a semidefinite program with an eye towards a generalization of the solution technique. A generalization we describe in Section 3 where the Lagrangean dual of a quadratically constrained program is solved via a semidefinite relaxation. In Section 4 we review very briefly the standard approach to sequential quadratic programming with its linearization of the constraints, which we then contrast to quadratic approximations in Section 5.

This leads, in Sections 6 and 7, to a different sequential process, based on a quadratic model of both the objective function and the constraints, and related to a higher-order Newton step.

In this paper, the results will be stated without proofs but with numerous references. More details can be found in [21]. Numerical approaches for *SDP* are discussed in [22], while more insight into the geometry of semidefinite relaxations can be found in [1].

2 The Simplest Case

Consider the unconstrained problem

$$UNC \qquad \min\left\{ f(x) \mid x \in \mathbb{R}^n \right\}.$$

When possible, the method of choice for this problem is Newton's method, which solves a quadratic model of the objective function. To ensure a solution (or convexity) of the model, Newton's method is often implemented within a Trust-Region, or Restricted-Step approach. This, very efficient, variation proceeds from an initial estimate of the solution; it develops a second-order model of the objective function, deemed valid in a region around the estimate; and solves the trust-region subproblem,

$$\textit{TRS} \qquad \min\left\{q_0(d) = d^tQd + 2b^td \mid q_1(d) = d^td \leq \delta^2, d \in \mathbb{R}^n\right\}.$$

The model is constructed from $Q = \nabla^2 f(x^{(k)})$ (or an approximation of the Hessian), $b = \nabla f(x^{(k)})$ and the parameter δ represents the radius where the model is trusted. A solution d is then used as the step to the next estimate $x^{(k+1)} = x^{(k)} + d$.

The *TRS* has been generalized [38],[25], to an arbitrary quadratic constraint and to upper and lower bounds on the trust region. Strong duality results of *TRS* are maintained and efficient implementations have been developed for these generalizations. We will review some of these results before applying them to a constrained problem.

The early results concern necessary and sufficient conditions for optimality of *TRS*. These conditions, stated in the theorem below for reference, were first established by Gay [15], and concurrently by Sorensen [37].

Theorem 2.1 *The point d is an optimal solution for TRS if and only if there is a λ such that the pair (d, λ) satisfy*

$$d^td \leq \delta^2 \qquad \textit{(primal feasibility)},$$

$$\begin{array}{ll} (Q + \lambda I) \succeq 0 & \textit{(strengthened second-order)} \\ \lambda \geq 0 & \textit{(multiplier sign)} \end{array} \qquad \textit{(dual feasibility)}$$

$$\begin{array}{ll} (Q + \lambda I)d = -b & \textit{(stationarity)} \\ \lambda(d^td - \delta^2) = 0 & \end{array} \qquad \textit{(complementarity)},$$

where $\succeq$ denotes positive semidefiniteness.

We have added the groupings of: primal feasibility, dual feasibility, complementary slackness. This corresponds to the current popular primal-dual approaches to optimization problems. We will see below that the two middle conditions and the first complementarity condition do in fact correspond to a properly chosen dual problem.

Directly from these conditions, special-purpose, very efficient algorithms have been developed for *TRS*. Some further insight into the structure of the problem is obtained by observing the equivalence of *TRS* to a convex program, i.e. to its linear semidefinite programming relaxation.

First, as was shown in [38], the Lagrangean dual of *TRS* can be written as

$$\text{NonLinDualSDP-TRS} \qquad \max\left\{-b^t(Q + \lambda I)^\dagger b - \lambda\delta^2 \mid Q + \lambda I \succeq 0, \lambda \geq 0\right\},$$

a nonlinear semidefinite program, where $(\cdot)^\dagger$ is the Moore-Penrose generalized inverse. This program illustrates the dual feasibility statements following Theorem 2.1. In addition, stationarity of the Lagrangean of the dual (NonLinDualSDP-TRS) corresponds to feasibility of the primal (TRS); while the stationarity condition in the theorem implicitly yields half of the complementary slackness, i.e. we can rewrite the stationarity as:

$$(Q+\lambda I)(d^*+z)=-b;\ d^*=-(Q+\lambda I)^{\dagger}b;\ z^t(Q+\lambda I)z=\langle Q+\lambda I, zz^t\rangle=0.$$

Note that $z\neq 0$ in the above relates to the so-called *hard case* for TRS.

In addition, the Lagrangean dual has been shown to be equivalent to the following linear semidefinite program [35],

$$\text{LinDualSDP-TRS}\qquad \min\left\{(\delta^2+1)\lambda - t \mid \begin{bmatrix} t & -b^t \\ -b & Q \end{bmatrix} \succeq \lambda I, t\in\mathbb{R}, \lambda\in\mathbb{R}_+\right\}.$$

We can take the dual of the above linear semidefinite program (LinDualSDP-TRS) and get a semidefinite program equivalent to *TRS*.

$$\text{LinPrimalSDP-TRS}\qquad \min\left\{\tilde{\mu}(Y)=\langle P_0, Y\rangle \mid \langle E_{00}, Y\rangle = 1, \langle P_I, Y\rangle \le \delta^2, Y\in\mathbb{S}_+^{n+1}\right\}.$$

The variable in this program, Y, belongs to the space of symmetric positive semidefinite matrices of dimension $n+1\times n+1$, which we denote $\mathbb{S}_+^{n+1}$. Also,

$$P_0=\begin{bmatrix} 0 & b^t \\ b & Q \end{bmatrix}, P_I=\begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}, E_{00}=\begin{bmatrix} 1 & 0 \\ 0 & \mathbf{0} \end{bmatrix},$$

and $\langle A, B\rangle$ is the usual matrix space inner product, $\operatorname{trace} BA^t$.

This pair of linear primal-dual semidefinite programs (LinDualSDP-TRS, LinPrimalSDP-TRS) are bounded and feasible. Therefore optimal solutions are attained at equal objective values. Finally, and this is crucial, part of the first column of the primal semidefinite solution, the matrix Y, is feasible for *TRS*. And, possibly with an additional displacement, chosen in the nullspace of the Lagrangean, this first column yields the same objective value for *TRS* as its dual optimal. (The next sections will provide some details in more generality.) By this procedure, usually known as *lifting*, of *TRS* to the cone of semidefinite matrices, and projecting back (by the first column), we see that there are no duality gaps for *TRS*. This was first shown by Stern and Wolkowicz [38].

Theorem 2.2 *Assuming a non-trivial trust-region, the optimal solution to TRS and to its Lagrangean dual (NonLinDualSDP-TRS) are attained and the corresponding objective values are equal.*

An interesting consequence is that polynomial-time interior-point algorithms can be used to solve *TRS* (via its semidefinite reformulation), *even if the objective function and the feasible set are non-convex.* We therefore have a tractable problem. *TRS* sits somewhere between convex and non-convex problems. What generalization of *TRS* to multiple trust-regions can we expect is now the obvious question.

3 Multiple Trust-Regions

We now move up from the one-constraint problem, since our ultimate objective is to solve *NLP* and consider a quadratic objective constrained by multiple quadratics,

$$Q^2P \qquad \min\left\{x^tQ_0x + 2b_0^tx - a_0 \mid x^tQ_kx + 2b_k^tx \le a_k, 1 \le k \le m, x \in \mathbb{R}^n\right\}.$$

In this section we present two main ideas. First we show that the feasible set of the *SDP* relaxation in matrix space actually provides a *non-convex* approximation of the feasible set of the original Q^2P. Next we exploit this geometry to obtain a good approximation of the optimum of Q^2P from the optimum of the *SDP* relaxation.

As soon as two trust-regions are considered, the standard necessary optimality conditions for Q^2P are not sufficient (as they were for *TRS*). This is reflected in the duality gap exhibited by some instances of multiple trust-region programs, an example of which follows shortly. In a certain way, the primal program satisfies the necessary conditions while the dual satisfies the sufficient.

We derive here the Lagrangean and semidefinite duals. First, introduce the vector $y = (x_0\ x)^t$. We then require $x_0^2 = 1$ or, in terms of the new variable, $y^tE_{00}y = 1$, to get an equivalent program to Q^2P with only pure quadratic forms in the objective and the constraints,

$$\text{Hom-}Q^2P \qquad \min\left\{y^tP_0y \mid y^tE_{00}y = 1, y^tP_ky \le a_k, 1 \le k \le m, y \in \mathbb{R}^{n+1}\right\},$$

where

$$E_{00} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{0} \end{bmatrix} \qquad \text{and } P_k = \begin{bmatrix} 0 & b_k^t \\ b_k & Q_k \end{bmatrix}, \qquad 0 \le k \le m.$$

The homogenization simplifies the notation and opens the way to the semidefinite relaxation since we can rewrite Hom-Q^2P using matrix variables.

$$\text{Hom-Matrix-}Q^2P \qquad \min\left\{\langle Y, P_0\rangle \mid \langle Y, E_{00}\rangle, \langle Y, P_k\rangle \le a_k, 1 \le k \le m, Y \in \mathbb{S}_+^{n+1}, Y \text{ is rank-one}\right\},$$

Dropping the rank-one condition provides a relaxation which we can justify by showing its equivalence with the Lagrangean relaxation. After some rearrangement of terms, the Lagrangean dual of Hom-Q^2P reads

$$\max\left\{\min\left\{y^t(P + \sum_{i=1}^m \lambda_iP_i + \lambda_0E_{00})y - \lambda_0 - \lambda^ta \mid y \in \mathbb{R}^{n+1}\right\} \mid \lambda_i \ge 0, \lambda_0 \in \mathbb{R}\right\}.$$

For the inner minimization to be bounded we must now have

$$P + \sum_{i=1}^m \lambda_iP_i + \lambda_0E_{00} \succeq 0.$$

Since all principal minors of a positive semidefinite matrix are positive semidefinite, this implies

$$Q + \sum_{i=1}^m \lambda_iQ_i \succeq 0.$$

This is where the duality gap arises. The standard necessary optimality conditions for Q^2P do not require the Hessian of the Lagrangean to be semidefinite. But the Lagrangean dual program we are deriving here requires the same Hessian to be semidefinite. We therefore cannot expect the primal variables corresponding to an optimal dual solution to be optimal for Q^2P. They will be optimal only in cases where the Lagrangean is convex at primal optimality.

To complete the derivation, we note that the minimum over y will be attained at $y = 0$ from which we get the dual program

$$\textit{Dual-}Q^2P \qquad \max\left\{-\lambda_0 - \lambda^t a \mid P_0 + \lambda_0 E_{00} + \sum_{i=1}^{m} \lambda_i P_i \succeq 0, \lambda \geq 0\right\}.$$

Example 3.1 Example of two trust-regions.

Consider the homogenized primal-dual pair,

$$\begin{aligned}
\mu^* &= \min\left\{y^t \begin{bmatrix} 0 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix} y \mid y_0^2 = 1, y^t \begin{bmatrix} 0 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} y \leq 1, y^t \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} y \leq 4\right\} \\
\nu^* &= \max\left\{-\lambda_0 - \lambda_1 - 4\lambda_2 \mid \begin{bmatrix} \lambda_0 & 1 - 2\lambda_1 & 0 \\ 1 - 2\lambda_1 & -2 + \lambda_1 + \lambda_2 & 0 \\ 0 & 0 & 2 + \lambda_1 + \lambda_2 \end{bmatrix} \succeq 0, \lambda \geq 0\right\}.
\end{aligned}$$

The optimal solutions are

$$y^* = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \ \textit{therefore } x^* = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \ \textit{and } \lambda^* = \begin{bmatrix} 0 \\ 1/2 \\ 3/2 \end{bmatrix},$$

with a duality gap of

$$|\mu^* - \nu^*| = \left|-4 + \frac{13}{2}\right| \neq 0.$$

The relaxation of Q^2P into a semidefinite program, done directly, by dropping the rank-one condition on the homogenized primal, or by taking the semidefinite dual of Dual-Q^2P will result in the following, which we will refer to as the relaxation of Q^2P,

$$\textit{SDP-}Q^2P \qquad \min\left\{\langle P_0, Y\rangle \mid \langle E_{00}, Y\rangle = 1, \langle P_k, Y\rangle \leq a_k, 1 \leq k \leq m, Y \in \mathbb{S}_+^{n+1}\right\}.$$

Example 3.2 Semidefinite relaxation of Example 3.1.

$$\min\left\{\langle\begin{bmatrix} 0 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, Y\rangle \mid \langle E_{00}, Y\rangle = 1, \langle\begin{bmatrix} 0 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, Y\rangle \leq 1, \langle\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, Y\rangle \leq 4\right\}$$

The optimal solution is

$$Y^* = \begin{bmatrix} 1 & 0.75 & 0 \\ 0.75 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The feasibility of the first column of the semidefinite relaxation, which is exemplified above was first shown by Fujie and Kojima [14] for an equivalent problem with linear objective function. And while a transformation of Q^2P into an equivalent program with a linear objective function is simple, it obscures the geometry of the semidefinite relaxation. In fact, this transformation is not needed.(See [1]). We define the feasible set of Q^2P,

$$\hat{F} := \{x \in \mathbb{R}^n \mid x^t Q_k x + 2b_k^t x \leq a_k, 1 \leq k \leq m\};$$

the feasible set of *SDP*-Q^2P,

$$\tilde{F} := \{Y \in \mathbb{S}_+^{n+1} \mid \langle P_k, Y\rangle \leq a_k, 1 \leq k \leq m\};$$

and the projector map,

$$P_R : \mathbb{S}_+^{n+1} \mapsto \mathbb{R}^n, P_R(\begin{bmatrix} a & x^t \\ x & X \end{bmatrix}) = x.$$

Theorem 3.3 *Suppose that Y is a feasible solution of SDP-Q^2P. The projected vector, $x = P_R(Y)$, is then feasible for all convex constraints of Q^2P.*

This is a fairly interesting result. It produces feasible points of *TRS* from feasible points of the relaxation (*SDP*-Q^2P), even when these are not rank one. Therefore, it provides a convex approximation to the set $\hat{F}$. However, *SDP* actually provides a better approximation than this would lead us to believe. And it does so using non-convex inequalities.

Let us define a valid inequality for Q^2P as

$$\sum_{i=1}^{m} \lambda_i (x^t Q_k x + 2b_k^t x - a_k) \leq 0, \quad \text{where} \quad Q + \sum_{i=1}^{m} \lambda_i Q_i \succeq 0.$$

These inequalities, an infinite number of them, are not, in general, convex. (Simply consider a *TRS* where the objective is strictly convex while the constraint is not.) However, they provide geometric insight into the *SDP* relaxation. More precisely, we have the following theorem.

Theorem 3.4 *Under Slater's constraint qualification, the closure of the set of projected first columns,*

$$\{x \mid x = P_R(Y), Y \in \tilde{F}\},$$

is equal to the set of vectors satisfying all valid inequalities,

$$\left\{x \mid \sum_{i=1}^{m} \lambda_i (x^t Q_k x + 2b_k^t x - a_k) \leq 0, \quad Q + \sum_{i=1}^{m} \lambda_i Q_i \succeq 0\right\}.$$

These valid inequalities establish the relation between the set of projected columns of *SDP* solutions and some intersection of the original constraints.

We now use the above geometric descriptions to provide an approximate solution to Q^2P from the optimum of *SDP*. We use the first column of the optimum Y but then we use the properties of the valid inequalities to improve this column, i.e. we move onto a boundary of a valid inequality, or equivalently obtain complementary slackness. A feasible pair Y, λ to the semidefinite relaxation, if Y is not rank one, will in general map to a vector x for which complementarity fails but improving the objective value is then easy. The idea is to choose a displacement along the nullspace of the Lagrangean until one or more slack constraints is satisfied with equality.

Lemma 3.5 *If the semidefinite primal optimal solution Y is not rank one, let $\tilde{x} = P_R(Y)$, (part of the first column of Y). Then there is a $\bar{x}$ chosen in $\mathcal{N}(Q_0 + \sum \lambda_i Q_i + \mu E_{00})$, the nullspace of the Lagrangean, such that $x = \tilde{x} + \bar{x}$, is feasible and will improve the primal objective value of Q^2P.*

Example 3.6 Nullspace move to optimality of Example 3.1.

$$Y^* = \begin{bmatrix} 1 & 0.75 & 0 \\ 0.75 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \quad \nabla^2 \mathcal{L}(x^*, \lambda^*) = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \quad \bar{x} \in \mathcal{N}(\nabla^2 \mathcal{L}(x^*, \lambda^*)) = \begin{bmatrix} 1.25 \\ 0 \end{bmatrix}$$

After a step in the nullspace up to a constraint, we obtain the optimal solution (compare to Example 3.1)

$$x^* = P_R(Y) + \bar{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

This additional step is a straight-forward generalization of an idea introduced by Moré and Sorensen [26] to solve *TRS* and there is an explicit expression for the step as there is for *TRS*.

4 Sequential Quadratic Programming

Sequential Quadratic Programming, denoted *SQP*, also known as Recursive Quadratic Programming, falls under the heading of Lagrange [23] or Newton-Lagrange [13] methods and is arguably the most efficient general-purpose algorithm for medium size nonlinear constrained programs [39], [5]. With solid theoretical foundations, with the appropriate quadratic subproblem, the method can be viewed as an extension of Newton or quasi-Newton algorithms to constrained optimization.

Yet the very existence of innumerable variations of the basic algorithm indicates that the last word on *SQP* has not been written. Research has produced SL_1QP [9], [13] based on a non-differentiable merit function and FSQP [30] for a method where iterates are kept within the feasible region. And much of the current research aims to apply the method to large-scale problems [17].

The original algorithm dates from Wilson's [43] dissertation in 1963 but was made better-known by Beale [3] and then Han [19] and Powell [33] a few years later. Consider again, the general, nonlinear programs with equality and inequality constraints

$$\textit{NEP} \qquad \min\left\{f(x) \mid h(x)=0, x\in\mathbb{R}^n\right\} \qquad \text{and} \quad \textit{NLP} \qquad \min\left\{f(x) \mid g(x)\le 0, x\in\mathbb{R}^n\right\},$$

where $f:\mathbb{R}^n\to\mathbb{R}$, and $h,g:\mathbb{R}^n\to\mathbb{R}^m$. We sometimes write vector-valued functions, like $h(x)$, as

$$h(x) = (h_1(x)\ h_2(x)\ \ldots\ h_m(x))^t.$$

We define the Lagrangean of *NEP* as $\mathcal{L}(x,\lambda) := f(x) + \lambda^t h(x)$. The first-order necessary conditions (under a constraint qualification) for *NEP* at an optimal point x^* guarantee the existence of a multiplier λ satisfying $\nabla_x\mathcal{L}(x^*,\lambda)=0$. Together with feasibility (equivalent to $\nabla_\lambda\mathcal{L}(x^*,\lambda)=0$), stationarity expands to

$$\begin{aligned} \nabla f(x^*) + \nabla h(x^*)\lambda &= 0,\\ h(x^*) &= 0, \end{aligned}$$

where $\lambda = (\lambda_1\ \lambda_2\ \ldots\ \lambda_m)^t$ is the vector of Lagrange multipliers. To simplify the exposition, we use $\nabla h(x)$ to denote $[\nabla h_1(x)\nabla h_2(x)\ldots\nabla h_m(x)]$, the transpose of the Jacobian of h.

An iterative attempt at the non-linear system above by Newton's method produces

$$\begin{bmatrix} \nabla^2 f(x^{(k)}) + \sum\lambda_i^{(k)}\nabla^2 h_i(x^{(k)}) & h'(x^{(k)}) \\ h'(x^{(k)})^t & 0 \end{bmatrix} \begin{bmatrix} \delta_x \\ \delta_\lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x^{(k)}) - h'(x^{(k)})^t\lambda^{(k)} \\ -h(x^{(k)}) \end{bmatrix},$$

where $\delta_x = x^{(k+1)} - x^{(k)}$ and $\delta_\lambda = \lambda^{(k+1)} - \lambda^{(k)}$. The usual simplification, at this point, is to let $\lambda^{(k+1)} = \lambda^{(k)} + \delta_\lambda$ and $d=\delta_x$, to obtain what we will refer to as the First-Order Newton Step,

$$\textit{FONS} \quad \begin{bmatrix} \nabla^2\mathcal{L}(x^{(k)},\lambda^{(k)}) & h'(x^{(k)}) \\ h'(x^{(k)})^t & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda^{(k+1)} \end{bmatrix} = \begin{bmatrix} -\nabla f(x^{(k)}) \\ -h(x^{(k)}) \end{bmatrix}.$$

This system produces a direction d and a new vector of Lagrange multipliers estimates $\lambda^{(k+1)}$.

An important remark is that the system of equations *FONS* can also be derived as the first-order necessary conditions of the quadratic program

$$\begin{array}{llll} QP & \min & q(d) = & f(x^{(k)}) + \nabla f(x^{(k)})^t d + \frac{1}{2}d^t\nabla^2\mathcal{L}(x^{(k)},\lambda^{(k)})d \\ & \text{s.t.} & l_i(d) = & h_i(x^{(k)}) + \nabla h_i(x^{(k)})^t d = 0, \qquad 1\le i\le m, \end{array}$$

hereafter known as the *QP* subproblem. Stationarity of the Lagrangean of *QP* yields the first line of *FONS*, and feasibility yields the second line. This is why *SQP* is viewed as an extension of Newton's method to constrained optimization.

In addition, the success of the trust region strategy to unconstrained optimization has led to the addition of a trust region constraint to the *QP* subproblem, $||d||^2 \le \delta^2$. However, this can lead to infeasible subproblems. One solution is the shift proposed by Vardi [41], i.e. shifting the linearized constraint to get a relaxed problem with constraints

$$\alpha_k h_i(x^{(k)}) + \nabla h_i(x^{(k)})^t d = 0, \qquad 1 \le i \le m,$$

with relaxation parameter $0 < \alpha_k < 1$ chosen so that the feasible set is nonempty. Since the relaxation parameter has to be chosen in a heuristic fashion, another approach is to use the two trust region subproblem introduced by Celis, Dennis and Tapia [8], and used thereafter by a number of researchers (See Byrd, Schnabel and Schultz [7], Powell and Yuan [34], Yuan [44], Williamson [42], El-Alem [11], Zhang [45].) i.e. the linear constraints are replaced by the quadratic constraint

$$||h_i(x^{(k)}) + \nabla h_i(x^{(k)})^t d||^2 \le \theta_k.$$

As we will see later, in our semidefinite subproblem, the potential infeasibility is handled by relaxing the homogenization constraint $d_0^2 = 1$ to $d_0^2 \le 1$. This is related to the Vardi parameter approach to guarantee feasibility of the subproblem yet there is no required heuristic to choose the parameter. All is handled automatically, i.e. a best parameter is found when solving the *SDP* subproblem, since the program will make d_0 as close to 1 as possible.

We must recognize some characteristics of the *QP* subproblem. A Taylor first-order approximation of the constraint defines the feasible set while a second-order expansion of the objective, to which we add second-order terms of the constraints, completes the problem definition. The rational for these unexpected modifications is based on Newton's method for the optimality conditions.

We can forgo discussion of the line search either because a trust-region is used or under the assumption that a full step is taken at each iteration. This is justified only if the initial estimate x is close enough to the optimal solution x^*. In general, the *SQP* linesearch approach relies on a merit function $\varphi(x, \lambda)$, reduced at each iteration and minimized when the system of first-order conditions *FONS* is satisfied. In general, a well-behaved merit function has a local minimum where the constrained problem has a solution and it must allow the line search to accept a full step, at least asymptotically. Finding a proper merit function to achieve all the desired features of an *SQP* algorithm is still an active area of research.

This line search procedure is expressed in the following algorithms as

$$\alpha = \text{linesearch}(\varphi(x^{(k)}, \lambda^{(k)}), d).$$

This is meant to suggest that the procedure minimizes, perhaps approximately, the merit function φ, from the current iterate $(x^{(k)}, \lambda^{(k)})$, in the direction d, and returns the step length α corresponding to this one-dimensional minimization.

Alternatively, if the trust-region approach is used, then the full step returned from the trust-region subproblem is either taken or discarded with a corresponding adjustment to the trust-region radius. This modification of the trust-region is usually based on a ratio of the actual to the predicted reduction of some merit function [12].

SEQUENTIAL QUADRATIC PROGRAMMING FRAMEWORK

$\boldsymbol{SQP}(f, \nabla f, \nabla^2 f, h, \nabla h_i, \nabla^2 h_i, x^{(0)}, \lambda^{(0)})$
do
$d \in \text{argmin}\{\nabla f(x^{(k)})^t d + \frac{1}{2} d^t \nabla^2 f(x^{(k)}) d : h_i(x^{(k)}) + \nabla h_i(x^{(k)})^t d = 0, 1 \leq i \leq m, d \in \mathbb{R}^n\}$
$\alpha = \text{linesearch}(\varphi(x^{(k)}, \lambda^{(k)}), d)$
$x^{(k+1)} = x^{(k)} + \alpha d$
$k = k + 1$
Estimate new Lagrange multipliers
until convergence
return$(x^{(k)}, \lambda^{(k)})$

There are a few problems with this skeleton of an algorithm, problems that researchers have struggled with and successfully solved in a number of ways over the years. We will highlight some of these difficulties and see how a quadratically constrained subproblem handles them: First, the *QP* subproblem can be infeasible or unbounded. The infeasibility can be dealt with by taking a steepest descent step along the merit function. But the unboundedness requires some modification of the subproblem. Adding a trust region will do, moving *SQP* closer to second-order constraints and our way of thinking.

But there is more. The Lagrange multipliers are not, in the basic algorithm described above, a by-product of the subproblem. Although the multipliers resulting from *QP* can be added to the previous estimates used in the objective function, some authors suggest solving a separate least square problem [16] to get a better approximation of the "true" Lagrange multipliers. An alternative, again, is to change the subproblem.

5 Quadratic Approximations of Nonlinear Programs

Recall that the standard *SQP* subproblem approximated the objective function to second order yet approximated the constraints only to first order. Some attempt is made to include curvature information in the objective function but this is done using the Lagrange multipliers from the previous iteration.

We wish a better balanced, yet tractable, subproblem where the feasible region is also a second-order approximation. As the original subproblem considered was called the *QP* subproblem, we will call this program the Q^2P subproblem. Consider a vector $x^{(k)} \in \mathbb{R}^n$, an estimate of the primal solution. Expand the functions of *NLP* by second-order Taylor polynomials and express

$$\begin{array}{llll} \textit{NLP-}Q^2P & \min & q_0(d) & = \nabla f(x^{(k)})^t d + \frac{1}{2} d^t \nabla^2 f(x^{(k)}) d \\ & \text{s.t.} & q_j(d) & = g_j(x^{(k)}) + \nabla g_j(x^{(k)})^t d + \frac{1}{2} d^t \nabla^2 g_j(x^{(k)}) d \leq 0, \qquad 1 \leq j \leq m. \end{array}$$

Such a straightforward subproblem has often been considered, but has, just as often, been discarded as unsolvable. One notable exception is an algorithm by Maany [24] developed, interestingly enough, because the standard *SQP* approach failed on the highly nonlinear orbital trajectory problems they were studying. (See Dixon, Hersom and Maany [10].)

Before we attempt to solve the Q^2P subproblem, we will precisely construct it and analyze the properties it possesses that make it an attractive approximation to a nonlinear program.

5.1 Feasible region

The subproblem above differs from the traditional *QP* subproblem mostly in the feasible region it describes. The objective function is correspondingly simplified not to include what would be redundant constraint information.

Even if the feasible region of the quadratic approximation does not always include the original feasible region, it is closer in some sense to that region, *for the Taylor residual is smaller.* Also, since the second-order feasible region is within the linearly enclosed region, a bounded *QP* subproblem implies a bounded second-order subproblem. But note that the reverse is false so that Q^2P may be bounded while *QP* is not.

As as aside, there is a sense in which the semidefinite relaxation yields a feasible set somewhere in between the linear and the quadratic approximations, since it is true that the projected first column of the semidefinite feasible solutions is isomorphic to a relaxation of the feasible set of the quadratically constrained program, as shown above in Section 3.

5.2 Second-order Lagrange multiplier estimates

In traditional *SQP*, the multipliers are essential in the formulation of the objective function, they must therefore be reasonably accurate. Yet they are based on the previous iteration, unless they have been updated after the linesearch.

We recall that a pair of vectors x^* and λ^*, optimal for *NLP*, are related by the stationarity equation,

$$\nabla f(x^*) + \sum \lambda_i \nabla g_i(x^*) = 0.$$

This condition suggests that the optimal solution λ of the least-square problem,

$$\min \left\{ \|\nabla f(x^{(k)}) + \sum \lambda_i \nabla g_i(x^{(k)})\|_2^2 \mid \lambda \in \mathbb{R}^m \right\},$$

might provide an appropriate estimate of the true multipliers. An estimate which improves as $x^{(k)}$ approaches feasibility and the right active set is identified.

In the section of their book devoted to the identification of accurate multipliers, Gill, Murray and Wright [16] pursue this further and suggest aiming for second-order multiplier estimates: The approach is to let $d = x^* - x^{(k)}$ and expand the stationarity condition of *NLP*, around $x^{(k)}$, by a Taylor polynomial of first order to get

$$\nabla f(x^{(k)}) + \nabla^2 f(x^{(k)})d + \sum \lambda_i(\nabla g_i(x^{(k)}) + \nabla^2 g_i(x^{(k)})d) + o(\|d\|^2) = 0,$$

or, using the Lagrangean,

$$\nabla f(x^{(k)}) + \nabla^2 \mathcal{L}(x^{(k)}, \lambda^*)d + \sum \lambda_i^* \nabla g_i(x^{(k)}) + o(\|d\|^2) = 0.$$

Gill, Murray and Wright note at this point that it is impossible to estimate λ^* directly from the above equation for two reasons: First, d is unknown; second, components of λ^* are buried inside the Hessian of the Lagrangean. They reason that the best available multipliers λ and an approximating step d used in a least-square problem such as

$$\min \left\{ \|\nabla f(x^{(k)}) + \nabla^2 \mathcal{L}(x^{(k)}, \lambda) d + \sum \eta_i \nabla g_i(x^{(k)})\|_2^2 \mid \eta \in \mathbb{R}^m \right\},$$

would provide a vector η, deemed a second-order estimate of λ^* if d is sufficiently small and λ is, at least, a first-order estimate of λ^*.

This is where the Q^2P subproblem yields interesting information. From stationarity of *NLP-*Q^2P, at optimal vectors d and λ, we obtain

$$\nabla f(x^{(k)}) + \nabla^2 \mathcal{L}(x^{(k)}, \lambda) d + \sum \lambda_i \nabla g_i(x^{(k)}) = 0.$$

These optimal multipliers λ therefore *solve the second-order least-square problem* for the given d. One of the two concerns of Gill, Murray and Wright, namely that the correct multipliers are buried in the Hessian of the Lagrangean is implicitly taken care of. We need only to assume that $x^{(k)}$ is close to x^* to conclude that the multipliers obtained from the Q^2P subproblem are second-order estimates of the true optimal multipliers. Without solving an additional least-square problem, Q^2P yields valuable dual variables in tandem with primal updates.

5.3 Optimality conditions

We now turn our attention to the vector d obtained from Q^2P, to qualify its value as a primal update. In traditional *SQP*, the usual guarantee is convergence to a point satisfying first-order conditions for *NLP*. In SQ^2P, we can claim a somewhat stronger result involving second-order conditions. In this section we assume that the strong constraint qualification holds for *NLP*: at optimality, the gradients of the active constraints are linearly independent.

Lemma 5.1 *Assume that $x^{(k)}$ is feasible for NLP. If the NLP-Q^2P subproblem is solved by $d = 0$ with multipliers λ, then the pair of vectors $x^{(k)}$ and λ satisfies the first-order conditions and second-order conditions of NLP. Conversely, if $x^{(k)}$ and λ satisfy the first and second-order necessary conditions of NLP. Then the pair of vectors $d = 0$, λ satisfy the first and second-order conditions of NLP-Q^2P.*

This implies that the Q^2P subproblem does better than the *QP* subproblem since they both solve the first-order conditions but only the former guarantees second-order optimality conditions. This is expected of a trust-region approach.

At this point we have some of the characteristics of the *NLP-*Q^2P subproblem. It may be worthwhile to repeat that an algorithm iterating exclusively on feasible points is possible. But one strength of *SQP*, in most of its variations, is not to require feasibility until convergence.

5.4 Additional constraint

There is an interesting avenue to explore, the addition of another constraint, a trust-region, around our best current solution, excluding the previous stationary point to which the algorithm had converged.

The additional constraint, a trust-region constructed to exclude the spurious stationary point, convexifies the Lagrangean and reduces the gap between our convex primal-dual approach and the original non-convex problem.

Lemma 5.2 *Suppose that x is a primal optimal solution to NLP-Q^2P with associated Lagrange multipliers $\lambda_1, \ldots, \lambda_m$. Then there exists a quadratic constraint that, added to the problem, will yield a convex Hessian of the Lagrangean while retaining x as an optimal solution.*

This is barely scratching the surface of what can be done with this approach and is not meant as a proof that the additional constraints guarantees convergence, especially since the right choice of radius for the additional trust-region has not been found. But the fact that is may be possible to eliminate the duality gap, and therefore solve the original problem by solving the relaxation, is appealing.

6 Quadratically Constrained Quadratic Programming

Now that a reasonable subproblem is defined and its solution is known to be useful, we combine it to our previous work on semidefinite relaxations to fully describe the SQ^2P approach.

The original problem under study is

$$\textit{NLP} \qquad \min\left\{ f(x) \mid g(x) \le 0, x \in \mathbb{R}^n \right\}.$$

At some point $x^{(k)}$, possibly infeasible, we expand every function by second-order Taylor polynomials and construct the subproblem

$$\begin{array}{llrcl} \textit{NLP-}Q^2P & \min & q_0(d) & = & \nabla f(x^{(k)})^t d + \frac{1}{2} d^t \nabla^2 f(x^{(k)}) d \\ & \text{s.t.} & q_i(d) & = & g_i(x^{(k)}) + \nabla g_i(x^{(k)})^t d + \frac{1}{2} d^t \nabla^2 g_i(x^{(k)}) d \le 0, \qquad 1 \le i \le m \\ & & & & d^t d \le \delta^2. \end{array}$$

We added a trust-region to guarantee a bounded subproblem, in cases of non-convex objective functions.

Note that, for simplicity, we assume that our constraints are nonlinear. Linear constraints have to be treated differently, essentially squared, see [32]. Equivalently, linear constraints can be eliminated or mapped to a linear constraint in matrix space.

Homogenization, obtained by adding a component d_0 to the vector d, together with the constraint $d_0^2 = 1$, allow the semidefinite relaxation,

$$\textit{PSDP} \qquad \min\left\{ \langle P_0, Y \rangle \mid \langle E_{00}, Y \rangle = 1, \langle P_i, Y \rangle \le a_i, 1 \le i \le m, \langle P_I, Y \rangle \le \delta^2, Y \in \mathbb{S}_+^{n+1} \right\},$$

where

$$P_0 = \begin{bmatrix} 0 & \nabla f(x^{(k)})^t & 0 \\ \nabla f(x^{(k)}) & \nabla^2 f(x^{(k)}) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; P_i = \begin{bmatrix} 0 & \nabla g_i(x^{(k)})^t & 0 \\ \nabla g_i(x^{(k)}) & \nabla^2 g_i(x^{(k)}) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \; a_i = -2h_i(x^{(k)}),$$

and where E_{00} and P_I have their usual definitions,

$$E_{00} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, P_I = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}.$$

But this relaxation, can possibly be infeasible if the current estimate is too far from the feasible region. To overcome this difficulty in *SQP*, Vardi suggested a heuristic shift of the linear constraints. We can do a related shift of our second-order constraints by allowing the additional component d_0 to take values between zero and one. That is, we change $d_0^2 = 1$ to $d_0^2 \leq 1$.

This additional relaxation allows for a feasible subproblem. Of course we would want d_0 to be as close to 1 as possible and examination of the subproblem shows that it automatically tries to make d_0 'large'. We need no heuristic to choose a Vardi-type parameter.

Once into the semidefinite cone, this additional relaxation corresponds to a change from $\langle E_{00}, Y\rangle = 1$ to $\langle E_{00}, Y\rangle \leq 1$.

The dual program is then either

$$DSDP \qquad \max\Big\{-\mu - \lambda^t a \mid P_0 + \mu E_{00} + \sum_{i=1}^{m} \lambda_i P_i + \lambda_I P_I \succeq 0, \mu \in \mathbb{R}, \lambda \geq 0\Big\},$$

or

$$DSDP \qquad \max\Big\{-\lambda_0 - \lambda^t a \mid P_0 + \lambda_0 E_{00} + \sum_{i=1}^{m} \lambda_i P_i + \lambda_I P_I \succeq 0, \lambda \geq 0\Big\},$$

whether the Vardi shift is included in the primal.

Solving one of the above primal-dual pair *PSDP*,*DSDP*, in the case of gap-free *NLP*, is enough since, as we have seen, the first column is optimal for the quadratic approximation. But, in general, we need an appropriate merit function to ensure sufficient decrease at each step and guarantee global convergence of the algorithm, whether we use a line search or a trust-region strategy.

The choice of merit function for *SQP* algorithms varies considerably. For infeasible iterates there is a need to balance improvement in the objective function and movement towards feasibility.

We will come back, briefly, to the merit function when we investigate convergence of the algorithm but we first complete its description. After solving the Q^2P subproblem for a direction $d \neq 0$, the next iterate is obtained by $x^{(k+1)} = x^{(k)} + d$. This new point serves for the expansion of a new problem by second-order polynomials and we iterate until the subproblem yields $d = 0$. As with any trust-region based algorithm, we adjust the trust-region radius according to the ratio of predicted improvement to actual improvement. At the end, we have a solution satisfying both first and second-order conditions of *NLP*. Somewhat more formally, here is the SQ^2P algorithm.

Sequential Quadratically Constrained Programming Framework

$\boldsymbol{SQ^2P}(f, \nabla f, \nabla^2 f, g_i, \nabla g_i, \nabla^2 g_i, x^{(0)})$
  do
    $Y \in \text{argmin}\{\langle P_0, Y\rangle : \langle P_i, Y\rangle \leq a_i, \langle E_{00}, Y\rangle = 1, Y \in \mathbb{S}_+^{n+1}\}$
    $(\mu, \lambda^{(k+1)}) \in \text{argmax}\{-\mu - \sum \lambda_i a_i : P + \sum \lambda P_i + \mu E_{00} \succeq 0, \lambda \geq 0\}$
    $d = P_R(Y)$
    $x^{(k+1)} = x^{(k)} + d$
    $r^k = \frac{\varphi(x^{(k)}) - \varphi(x^{(k+1)})}{q_0(x^{(k)}) - q_0(x^{(k+1)})}$
    if $(r^k < \frac{1}{4})$
      $\delta = \delta/4$
    elseif $(r^k > \frac{3}{4})$ and $\|x^{(k+1)} - x^{(k)}\| = \delta$
      $\delta = 2\delta$
    fi
    $k = k + 1$
  while $(\|d\| > \varepsilon)$
  Find maximal $d \in \mathcal{N}(\nabla^2 \mathcal{L})$ such that $g(x^{(k)} + d) \leq 0$
  $x^{(k)} = x^{(k)} + d$
return$(x^{(k)}, \lambda^{(k)})$

6.1 Semidefinite relaxation solution

If the *NLP-Q^2P* subproblem is convex, or more generally, if it is an instance without duality gaps, then solving the semidefinite relaxation, which can be done in polynomial time, will be enough since the primal semidefinite solution will be rank one. We will have a pair of primal-dual vectors satisfying the sufficient conditions for optimality of Q^2P.

This takes care of the convex case and of many non-convex cases. In other cases, we can move along the nullspace of the Lagrangean until we hit one of the constraints. This is possible since the the first column of the semidefinite relaxation is feasible for *NLP-Q^2P*. This nullspace-restricted step improves the objective value even if it does not lead to an optimal solution.

7 SQ^2P Viewed as a Higher-Order Newton Method

We now investigate the convergence of an iterative algorithm developed within the SQ^2P framework. The interesting point is that, as *SQP* could be viewed as a Newton method applied to the optimality conditions of *NLP*, so can SQ^2P; except that the Q^2Psubproblem corresponds to a higher-order Newton step.

To sketch the asymptotic convergence rate, we will make the standard simplifying assumption: When $x^{(k)}$ is close to x^*, the active constraints of the Q^2P subproblem are the same as the active constraints of *NLP*. That we have identified the active, and therefore the inactive constraints for *NLP* at x^*, allows us to ignore inactive constraints and change the active constraints to equalities. Therefore, under this assumption, we need consider only the equality-constrained program

$$NEP \qquad \min\Big\{ f(x) \mid h(x) = 0, x \in \mathbb{R}^n \Big\}.$$

We rewrite the stationarity condition of the Lagrangean as

$$\nabla\mathcal{L}(x^{(k)}, \lambda^{(k)}) = \begin{bmatrix} \nabla_x\mathcal{L}(x^{(k)}, \lambda^{(k)}) \\ \nabla_\lambda\mathcal{L}(x^{(k)}, \lambda^{(k)}) \end{bmatrix} = \begin{bmatrix} \nabla f(x^{(k)}) + h'(x^{(k)})^t\lambda \\ h(x^{(k)}) \end{bmatrix} = 0.$$

Newton's method can be used to solve a system of nonlinear equations as the one above. From the definition of $\mathcal{L}$, a second-order Newton step can be written as a nonlinear system of equation in δ_x and δ_λ. The manner in which we write it has little to do with a method of solution. It has everything to do with the comparison we wish to make between three systems of equations: from a second-order Newton method, from the SQ^2P step, and from the standard *SQP* step.

First, here is the second-order Newton step,

$$\begin{bmatrix} \sum \lambda_i \nabla h_i(x^{(k)}) + \left(\nabla^2 f(x^{(k)}) + \sum \lambda_i \nabla^2 h_i(x^{(k)})\right)\delta_x + H_3(\delta_x, \delta_\lambda) \\ h'(x^{(k)})\delta_x + \frac{1}{2}\delta_x^t h''(x^{(k)})\delta_x \end{bmatrix} = \begin{bmatrix} -\nabla f(x^{(k)}) \\ -h(x^{(k)}) \end{bmatrix},$$

where we have grouped the third-order derivatives of f and h under the name H_3. We can contrast this step to stationarity of the Lagrangean of *NLP*-Q^2P,

$$\begin{bmatrix} \sum \lambda_i \nabla h_i(x^{(k)}) + \left(\nabla^2 f(x^{(k)}) + \sum \lambda_i \nabla^2 h_i(x^{(k)})\right)\delta_x \\ h'(x^{(k)})\delta_x + \frac{1}{2}\delta_x^t h''(x^{(k)})\delta_x \end{bmatrix} = \begin{bmatrix} -\nabla f(x^{(k)}) \\ -h(x^{(k)}) \end{bmatrix},$$

and to stationarity of the Lagrangean of the *QP* subproblem or, equivalently, of a first-order Newton step,

$$\begin{bmatrix} \sum \lambda_i \nabla h_i(x^{(k)}) + \left(\nabla^2 f(x^{(k)}) + \sum \lambda_i \nabla^2 h_i(x^{(k)})\right)\delta_x \\ h'(x^{(k)})\delta_x \end{bmatrix} = \begin{bmatrix} -\nabla f(x^{(k)}) \\ -h(x^{(k)}) \end{bmatrix}.$$

The main difference between the stationarity of *NLP*-Q^2P and a second-order Newton's method lies in the third derivative terms missing in the former. But the second-order terms related to the curvature of the constraints are present and this is where we expect SQ^2P to overtake *SQP*, namely when the original problem has highly curved constraints. Viewed differently, the *NLP*-Q^2P subproblem *produces a first-order step towards stationarity and a second-order step towards feasibility.*

Under suitable conditions, the asymptotic q-quadratic convergence rate follows from stationarity of *NLP*-Q^2P, as expressed above, and from the convergence of a second-order Newton method.

For global convergence of the algorithm from an arbitrary starting point we use the following simple merit function

$$\varphi(x^{(k)}, \lambda^{(k)}) = \frac{1}{2}\left\| \begin{matrix} \nabla\mathcal{L}(x^{(k)}, \lambda^{(k)}) \\ g^+(x^{(k)}) \end{matrix} \right\|_2^2,$$

where

$$g_i^+(x^{(k)}) = \begin{cases} g_i(x^{(k)}), & \text{if } g_i(x^{(k)}) > 0; \\ 0, & \text{if } g_i(x^{(k)}) \le 0, \end{cases}$$

i.e. we strive for stationarity and feasibility.

The derivative, in the direction d, satisfies

$$d^t \nabla \varphi(x^{(k)}, \lambda^{(k)}) = d^t \nabla^2 \mathcal{L}(x^{(k)}, \lambda^{(k)}) + \sum d^t g_i^+(x^{(k)}) \nabla g_i(x^{(k)}),$$

which, since the solution of *NLP-Q*^{2}*P*, d satisfies the system

$$\begin{aligned} \nabla^2 \mathcal{L}(x^{(k)}, \lambda^{(k)}) d + \nabla \mathcal{L}(x^{(k)}, \lambda^{(k)}) &= 0 \\ \frac{1}{2} d^t \nabla^2 g_i(x^{(k)}) d + \nabla g_i(x^{(k)}) d + g_i(x^{(k)}) &\le 0, \end{aligned}$$

implies $d^t \nabla \varphi(x^{(k)}, \lambda^{(k)}) \le 0$, i.e., d is a descent direction for the merit function. Notice that this descent property does not rely on convexity. It therefore applies to general nonlinear programs.

8 Illustration

Example 8.1 contrasts the iterations of *SQP* and *SQ*^{2}*P* on a very small problem. The full step was taken at each iteration for the traditional *SQP*, and the trust-region used in *SQ*^{2}*P* was always large enough not to be binding, to better illustrate the directions.

Example 8.1 *Illustrative comparison of SQP and SQ*^{2}*P.*

$$\min \left\{ -x_1 - x_2 \mid x_1^3 - x_2 \le 0,\ x_1^3 + x_2^2 - 1 \le 0 \right\}$$

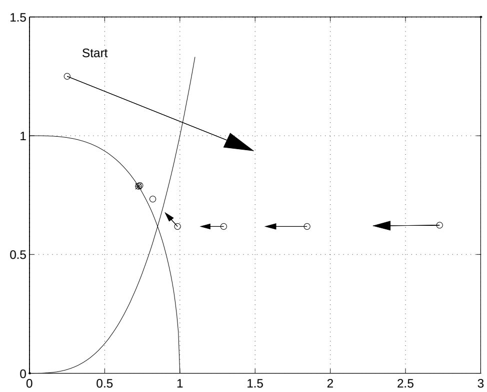

Figure 1: Iterations of SQP on Example 8.1, from initial point $(\frac{1}{4}\ \frac{5}{4})^t$. *As the first iteration demonstrates, the direction given by the QP subproblem can be poor.*

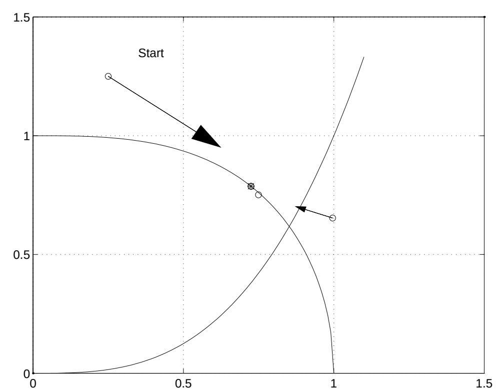

Figure 2: Iterations of SQ^2*P on the same example. The horizontal scale is changed to highlight the value of the direction provided by the Q*^{2}*P subproblem.*

9 Conclusion

The modern popular approach to optimization is via a local quadratic model. However, this (true) quadratic model is a quadratic constrained quadratic program and is generally intractable. Current approaches make two modifications to obtain a tractable subproblem: they use linear approximations to the constraints rather than quadratic; and they use an approximation of the Hessian of the objective function in order to maintain the information from the current Lagrange multiplier estimates and also obtain a true convex subproblem. Thus they solve a convex *QP* subproblem. This well known approach is called *SQP*. This approach has the added advantage that it can be considered to be Newton's method acting on the optimality conditions. However, there are several drawbacks and difficulties that researchers have had to deal with.

In this paper we have developed an approach which deals directly and transparently with the true quadratic model. We solve a relaxation of the true quadratic model using the best tractable approximation that we know of, i.e. we solve the Lagrangean relaxation of the quadratic model. We do this using *SDP*. We call this the SQ^2P approach.

This approach can be considered to be a higher order Newton method on the optimality conditions. More importantly, this approach avoids many of the difficulties of *SQP*, e.g. the Maratos effect, unboundedness and infeasibility of the subproblem, sufficient accuracy of the Lagrange multipliers. We have presented a simple example that illustrates how SQ^2P outperforms *SQP*.

There still remains many questions before this becomes a viable alternative to *SQP*. Semidefinite programming is a relatively new area and only recently have large sparse problems been solved, e.g. [46] and [4] where problems with 10000 variables have been solved.

Acknowledgement

The authors are indebted to an anonymous referee for many helpful comments that helped clarify the exposition of this paper.

References

[1] A. ALFAKIH, S. KRUK, and H. WOLKOWICZ. A note on geometry of semidefinite relaxations. Technical Report in progress, University of Waterloo, Waterloo, Canada, 1996.

[2] F. ALIZADEH. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5:13–51, 1995.

[3] E.M.L. BEALE. Numerical methods. In J. Abadie, editor, *Nonlinear Programming*. North-Holland, Amsterdam, 1967.

[4] S. J. BENSON, Y. YE, and X. ZHANG. Solving large-scale sparse semidefinite programs for combinatorial optimization. Technical report, The University of Iowa, 1997.

[5] P.T. BOGGS and J.W. TOLLE. Sequential quadratic programming. *Acta Numerica*, 4:1–50, 1995.

[6] S. BOYD, L. El GHAOUI, E. FERON, and V. BALAKRISHNAN. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, June 1994.

[7] R.H. BYRD, R.B. SCHNABEL, and G.A. SCHULTZ. A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties. *SIAM J. Numer. Anal.*, 22:47–67, 1985. Also available as Technical Report CU-CS-216-82, Department of Computer Science, University of Colorado, Boulder, CO.

[8] M.R. CELIS, J.E. DENNIS Jr., and R.A. TAPIA. A trust region strategy for nonlinear equality constrained optimization. In *Proceedings of the SIAM Conference on Numerical Optimization, Boulder, CO*, 1984. Also available as Technical Report TR84-1, Department of Mathematical Sciences, Rice University, Houston, TX.

[9] T. F. COLEMAN and A. R. CONN. Nonlinear programming via an exact penalty function: Asymptotic analysis. *Math. Prog.*, 24:123–136, 1982.

[10] L.C.W. DIXON, S.E. HERSOM, and Z.A. MAANY. Initial experience obtained solving the low thrust satellite trajectory optimisation problem. Technical Report T.R. 152, The Hatfield Polytechnic Numerical Optimization Centre, 1984.

[11] M. EL-ALEM. A global convergence theory for the CDT trust-region algorithm for constrained optimization. *SIAM Journal of Numerical Analysis*, 28:266–290, 1991.

[12] M. ELHALLABI. A global convergence theory for SLP and SQP trust-region algorithms. Technical Report CRPC-TR94371, Department of Mathematical Sciences, Rice University, Houston, TX, 1994.

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

[14] T. FUJIE and M. KOJIMA. Semidefinite programming relaxation for nonconvex quadratic programs. *J. Global Optim.*, 10(4):367–380, 1997.

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

[16] P.E. GILL, W. MURRAY, and M.H. WRIGHT. *Practical Optimization*. Academic Press, New York, London, Toronto, Sydney and San Francisco, 1981.

[17] P.E. GILL, M.A. SAUNDERS, and W. MURRAY. SNOPT: An SQP algorithm for large-scale constrained optimization. Technical Report NA 96-2, University of California, San Diego, 1996.

[18] M.X. GOEMANS and D.P. WILLIAMSON. .878-approximation algorithms for MAX CUT and MAX 2SAT. In *ACM Symposium on Theory of Computing (STOC)*, 1994.

[19] S.P. HAN. A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Applications.*, 22:297–309, 1977.

[20] C. HELMBERG, S. POLJAK, F. RENDL, and H. WOLKOWICZ. Combining semidefinite and polyhedral relaxations to integer programs. In *Proceedings of the 4th International IPCO Conference*, volume 920 of *Lecture Notes in Computer Science*, pages 124–134. Springer, 1995.

[21] S. KRUK. Semidefinite programming applied to nonlinear programming. Master's thesis, University of Waterloo, 1996.
URL: http://orion.uwaterloo.ca/~hwolkowi/henry/software/sqqp.d/krukmsthesis.ps.gz.

[22] S. KRUK, M. MURAMATSU, F. RENDL, R.J. VANDERBEI, and H. WOLKOWICZ. The Gauss-Newton direction in linear and semidefinite programming. Technical Report CORR 98-16, University of Waterloo, Waterloo, Canada, 1998. Detailed Web Version at
URL: ftp://orion.uwaterloo.ca/pub/henry/reports/gnsdplong.ps.gz.

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

[24] Z.A. MAANY. A new algorithm for highly curved constrained optimization. Technical Report T.R. 161, The Hatfield Polytechnic Numerical Optimization Centre, 1985.

[25] J.J. MORÉ. Generalizations of the trust region problem. Technical Report MCS-P349-0193, Argonne National Labs, Argonne, IL, 1993.

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

[27] J. L. Nazareth. The Newton and Cauchy perspectives on computational nonlinear optimization. *SIAM Rev.*, 36(2):215–225, 1994.

[28] J. L. Nazareth. *The Newton-Cauchy framework*, volume 769 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1994. A unified approach to unconstrained nonlinear minimization.

[29] M. OVERTON and H. WOLKOWICZ, editors. *Semidefinite programming.* North-Holland Publishing Co., Amsterdam, 1997. Dedicated to the memory of Svatopluk Poljak, Math. Programming **77** (1997), no. 2, Ser. B.

[30] E.R. PANIER and A.L. TITS. On combining feasibility, descent and superlinear convergence in inequality constrained optimization. *Mathematical Programming*, 59:261–276, 1993.

[31] P. PARDALOS and H. WOLKOWICZ (Editors). *Special Issue on Semidefinite Programming*, volume 2.1 of *J. Combinatorial Optimization.* Kluwer Academic Publishers, 1998.

[32] S. POLJAK, F. RENDL, and H. WOLKOWICZ. A recipe for semidefinite relaxation for (0,1)-quadratic programming. *Journal of Global Optimization*, 7:51–73, 1995.

[33] M. J. D. Powell. Algorithms for nonlinear constraints that use Lagrangian functions. *Math. Prog.*, 14, 1978.

[34] M.J.D. POWELL and Y. YUAN. A trust region algorithm for equality constrained optimization. Technical Report DAMTP 1986/NA2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England, 1986.

[35] F. RENDL and H. WOLKOWICZ. A semidefinite framework for trust region subproblems with applications to large scale minimization. *Mathematical Programming*, 77(2):273–299, 1997.

[36] N.Z. SHOR. Quadratic optimization problems. *Soviet J. of Computer and Systems Science*, 25(6):1–11, 1987.

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

[38] R. STERN and H. WOLKOWICZ. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM J. Optimization*, 5(2):286–313, 1995.

[39] J. STOER. Principles of sequential quadratic programming methods for solving nonlinear programs. In K. Schittkowski, editor, *Computational Mathematical Programming*, pages 165–207. Springer-Verlag, Berlin, 1985.

[40] L. VANDENBERGHE and S. BOYD. Positive definite programming. In *Mathematical Programming: State of the Art, 1994*, pages 276–308. The University of Michigan, 1994.

[41] A. VARDI. A trust region algorithm for equality constrained minimization: convergence properties and implementation. *SIAM J. Numer. Anal.*, 22:575–591, 1985.

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

[43] R.B. WILSON. *A simplicial algorithm for concave programming.* PhD thesis, Harvard University, 1963.

[44] Y. YUAN. On a subproblem of trust region algorithms for constrained optimization. *Mathematical Programming*, 47:53–63, 1990.

[45] Y. ZHANG. Computing a Celis-Dennis-Tapia trust-region step for equality constrained optimization. *Mathematical Programming*, 55:109–124, 1992.

[46] Q. ZHAO, S.E. KARISCH, F. RENDL, and H. WOLKOWICZ. Semidefinite programming relaxations for the quadratic assignment problem. *J. Combinatorial Optimization*, 2.1, 1998. CORR 95-27,
URL: ftp://orion.uwaterloo.ca/pub/henry/reports/qapsdp.ps.gz.

10 Notation

Key	Description	Section
NEP	General nonlinear equality program.	1
$\succeq$	Partial order of the semidefinite cone.	1
NLP	General nonlinear inequality program.	1
UNC	General unconstrained program.	2
TRS	Trust-Region Subproblem.	2
NonLinDualSDP-TRS	Nonlinear dual semidefinite formulation of *TRS*.	2
LinDualSDP-TRS	Linear dual semidefinite formulation of *TRS*.	2
$(\cdot)^\dagger$	Generalized Moore-Penrose inverse	2
$\mathbb{S}_+^{n+1}$	Space of positive semidefinite matrices	2
LinPrimalSDP-TRS	Linear primal semidefinite formulation of *TRS*.	2
Q^2P	Quadratic objective, quadratic constraint program in $\mathbb{R}^{n+1}$.	3
Hom-Q^2P	Q^2P homogenized, in $\mathbb{R}^{n+1}$.	3
Hom-Matrix-Q^2P	Q^2P homogenized, in $\mathbb{S}_+^{n+1}$.	3
Dual-Q^2P	Dual of Q^2P.	3
SDP-Q^2P	Semidefinite relaxation of Q^2P.	3
P_R	Projection of semidefinite relaxation solution to original space.	3
FONS	First-Order Newton Step.	4
QP	Traditional subproblem.	4
NLP-Q^2P	Quadratic subproblem for *NLP*.	5
PSDP	Semidefinite relaxation of *NLP*-Q^2P.	6
DSDP	Semidefinite dual to the relaxation of *NLP*-Q^2P.	6