

A TIGHT SDP RELAXATION FOR THE CUBIC-QUARTIC REGULARIZATION PROBLEM

JINLING ZHOU, XIN LIU, JIAWANG NIE, AND XINDONG TANG

ABSTRACT. This paper studies how to compute global minimizers of the cubic-quartic regularization (CQR) problem

$$\min_{s \in \mathbb{R}^n} f_0 + g^T s + \frac{1}{2} s^T H s + \frac{\beta}{6} \|s\|^3 + \frac{\sigma}{4} \|s\|^4,$$

where f_0 is a constant, g is an n -dimensional vector, H is a n -by- n symmetric matrix, and $\|s\|$ denotes the Euclidean norm of s . The parameter $\sigma \geq 0$ while β can have any sign. The CQR problem arises as a critical subproblem for getting efficient regularization methods for solving unconstrained nonlinear optimization. Its properties are recently well studied by Cartis and Zhu [*cubic-quartic regularization models for solving polynomial subproblems in third-order tensor methods, Math. Program, 2025*]. However, a practical method for computing global minimizers of the CQR problem still remains elusive. To this end, we propose a semidefinite programming (SDP) relaxation method for solving the CQR problem globally. First, we show that our SDP relaxation is tight if and only if $\|s^*\|(\beta + 3\sigma\|s^*\|) \geq 0$ holds for a global minimizer s^* . In particular, if either $\beta \geq 0$ or H has a nonpositive eigenvalue, then the SDP relaxation is shown to be tight. Second, we show that all nonzero global minimizers have the same length for the tight case. Third, we give an algorithm to detect tightness and to obtain the set of all global minimizers. Numerical experiments demonstrate that our SDP relaxation method is both effective and computationally efficient, providing the first practical method for globally solving the CQR problem.

1. INTRODUCTION

In this paper, we focus on computing global minimizers of the following unconstrained optimization problem

$$(1.1) \quad \min_{s \in \mathbb{R}^n} M(s) := f_0 + g^T s + \frac{1}{2} s^T H s + \frac{\beta}{6} \|s\|^3 + \frac{\sigma}{4} \|s\|^4,$$

where the symmetric matrix $H \in \mathbb{R}^{n \times n}$, the vector $g \in \mathbb{R}^n$, the constant $f_0 \in \mathbb{R}$ are problem parameters, the constants $\beta \in \mathbb{R}$ and $\sigma \geq 0$ are regularization parameters, the vector $s \in \mathbb{R}^n$ stands for the decision variables and its Euclidean norm is denoted as $\|s\| = \sqrt{s^T s}$. The problem (1.1) is called the cubic-quartic regularization (CQR) problem [8]. When the regularization parameter $\sigma = 0$, the problem (1.1) is reduced to a cubic regularization problem [19]. Throughout the paper, we assume the parameters σ and β are not simultaneously zero. For a minimizer of (1.1), we mean it is a global minimizer (unless its meaning is otherwise specified).

2020 *Mathematics Subject Classification.* 90C22, 90C23, 65K05, 90C30.

Key words and phrases. regularization, polynomial, semidefinite program, tight relaxation.

The cubic-quartic regularization problem is proposed and well studied by Cartis and Zhu [8]. In fact, they consider the following more general case:

$$(1.2) \quad \min_{s \in \mathbb{R}^n} f_0 + g^T s + \frac{1}{2} s^T H s + \frac{\beta}{6} \|s\|_W^3 + \frac{\sigma}{4} \|s\|_W^4,$$

where the norm $\|s\|_W = \sqrt{s^T W s}$ is defined by a given symmetric positive-definite matrix W . When $W = I_n$ (the n -by- n identity matrix), the problem (1.2) is reduced to (1.1). If in addition $\beta = 0$, the problem (1.2) turns into a quadratic-quartic regularization (QQR) problem [9]. After imposing a linear transformation $s = W^{-\frac{1}{2}} \tilde{s}$, the problem (1.2) turns into the standard form (1.1), resulting from the fact that $\|s\|_W = \|\tilde{s}\|$. Therefore, it suffices to consider the CQR problem in the form of (1.1).

1.1. Motivations. The Newton trust-region (NTR) methods are a class of practically useful high-order approaches for computing second order stationary points of the unconstrained optimization problem

$$(1.3) \quad \min_{x \in \mathbb{R}^n} f(x),$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a nonlinear smooth function. When the Hessian of f is available, the global convergence and local quadratic convergence rate can be well established under mild conditions. However, its worst case evaluation complexity is hard to establish. To this end, the so-called adaptive regularization methods are proposed. Nesterov and Polyak [19] proposed the cubic regularization methods (AR2/ARC) and established the worst-case evaluation complexity estimates. In each iteration, it requires to solve the following cubic regularization subproblem

$$(1.4) \quad \min_{s \in \mathbb{R}^n} f_0 + g^T s + \frac{1}{2} s^T H s + \frac{\beta}{6} \|s\|^3,$$

for a parameter $\beta > 0$. Clearly, (1.4) is a special case of (1.1) with $\sigma = 0$. Later, Nesterov [17] proposed an acceleration technique to ARC for solving convex optimization problems and achieved better worst-case evaluation complexity when the subproblem (1.4) is globally solved. Moreover, Cartis, Gould and Toint [2–4] developed the adaptive cubic regularization (ARC) framework, proving global convergence and worst-case evaluation complexity results for ARC and its variants. Their analysis clarified how to adaptively select the regularization parameter β .

Cartis, Gould and Toint [6, Cor. 8.3.1] gave a characterization for the global minimizer of the ARC subproblem. When $\beta > 0$, the global minimizer s^* of (1.4) can be expressed as

$$s^* = \begin{cases} -\left(H + \frac{\beta}{2} \|s^*\| I_n\right)^{-1} g, & \text{if } \frac{\beta}{2} \|s^*\| > -\lambda_1, \\ -\left(H + \frac{\beta}{2} \|s^*\| I_n\right)^\dagger g + \alpha v_1, & \text{if } \frac{\beta}{2} \|s^*\| = -\lambda_1. \end{cases}$$

Here, λ_1 is the smallest eigenvalue of H , v_1 is any corresponding eigenvector, the superscript † denotes the Moore-Penrose Pseudoinverse, and α is one of the two roots of

$$\left\| -\left(H + \frac{\beta}{2} \|s^*\| I_n\right)^\dagger g + \alpha v_1 \right\| = \|s^*\|,$$

which results in the corresponding s^* having a lower function value. However, how to get practically efficient methods for computing the global minimizer s^* is still mostly open. ARC is now a standard cubic regularization method and has

motivated further development of general AR_p schemes. When the third-order derivatives are available, it is shown in [5, 6] that the quartic regularization model (AR3) can achieve better local convergence and worst-case evaluation complexity. However, this comes with a steep practical cost, as the corresponding subproblems are NP-hard and are generally considered computationally intractable.

Recently, Cartis and Zhu [9] generalized the cubic regularization technique to get quadratic-quartic regularization (QQR) methods. It transforms the ARC subproblem into a sequence of globally solvable quadratic-quartic approximations. By replacing the third-order term with a combination of quadratic and quartic terms, QQR generates tractable local approximations. The cubic-quartic regularization (CQR) method is a further new technique for developing better local approximations. The CQR method [8] was proposed as an efficient approach for solving unconstrained optimization problems, guaranteeing convergence to first-order stationary points. It combines a cubic component, which captures directional tensor effects, with a quartic regularizer to ensure stability, and can also be interpreted as an approximation to higher-order regularization models. In terms of computational complexity, CQR matches the worst-case iteration bound of ARC methods but does not attain the improved complexity of full quartic regularization methods. Despite its foundational importance, the lack of a practical method for computing its global minimizers still remains a key challenge. Subsequent numerical implementation and adaptive regularization techniques are provided in [7]. Global optimality conditions for general nonconvex cubic polynomials with quartic regularization terms are recently studied in [11].

In the literature, there exists much work for solving polynomial optimization problems, whose objective and constraining functions are given by polynomials. The Moment-SOS hierarchy is efficient for solving them. We refer to [12–14, 23] for introductions to this area. There exist tight relaxation methods for solving general polynomial optimization problems. For instance, tight Moment-SOS relaxations can be obtained with gradient ideals [20], optimality conditions [21], and Lagrange multipliers [22]. However, these tight relaxations may require computation with polynomials with high degrees. Their computational cost grows quickly as the degree increases.

The CQR problem is not a standard polynomial optimization problem, due to the regularization term. It can be reformulated as polynomial optimization, by introducing a new variable t and an additional constraint, say, $t^2 = s^T s$, $t \geq 0$. However, doing this is not computationally efficient, since it becomes a polynomial optimization problem of degree four. The computational complexity of Moment-SOS relaxations grows quickly as the number of variables increases, which makes it difficult to solve large-scale problems. Recently, a global convergence rate for SOS type approximations with nonconvex adaptive regularization was established by Cartis and Zhu [10]. It introduces an algorithmic framework that combines the SOS Taylor model with adaptive regularization techniques for solving nonconvex smooth optimization problems.

How to efficiently compute global optimizers of the CQR problem (1.1) is an interesting question. Cartis and Zhu [8] provided necessary and sufficient global optimality conditions. How to use these conditions to find global minimizers still

remains an open question. Most earlier methods can only guarantee to find a stationary point for (1.1). To achieve the best convergence properties of CQR methods, it is highly wanted that global minimizers of CQR problems can be computed efficiently.

1.2. Contributions. To address the challenge of computing global minimizers for the CQR problem (1.1), we propose a semidefinite programming (SDP) relaxation method. We prove that this relaxation is tight under very general conditions. The tightness means that it recovers both the global minimum value and the global minimizers of (1.1). Our main contributions are fourfold:

i) We propose the following SDP relaxation for solving (1.1):

$$(1.5) \quad \begin{cases} \min_{Y, Z_1, Z_2} & \begin{bmatrix} f_0 & g^T/2 \\ g/2 & H/2 \end{bmatrix} \bullet Y + \frac{\beta}{6}(Z_2)_{22} + \frac{\sigma}{4}(Z_1)_{33} \\ \text{s.t.} & Y_{00} = 1, \quad (Z_1)_{11} = Y_{00}, \\ & (Z_1)_{12} = (Z_2)_{11}, \quad (Z_1)_{22} = (Z_2)_{12}, \\ & (Z_1)_{13} = (Z_2)_{12}, \quad (Z_1)_{23} = (Z_2)_{22}, \\ & (Z_1)_{13} = (Z_1)_{22} = Y_{11} + \cdots + Y_{nn}, \\ & Y \in \mathcal{S}_+^{n+1}, \quad Z_1 \in \mathcal{S}_+^3, \quad Z_2 \in \mathcal{S}_+^2. \end{cases}$$

In the above, the notation $\mathcal{S}_+^m$ denotes the cone of m -by- m symmetric positive semidefinite (psd) matrices and “ $\bullet$ ” denotes the Euclidean inner product for matrices. The label indices of the matrix variable Y are $0, 1, \dots, n$, and those of Z_1, Z_2 are $1, 2, 3$ and $1, 2$, respectively.

ii) We establish tightness results for the SDP relaxation (1.5). It is said to be *tight* if the optimal value ϑ^* of (1.5) equals the optimal value μ^* of (1.1). First, we prove that $\vartheta^* = \mu^*$ if and only if the regularization parameters β, σ satisfy the inequality

$$(1.6) \quad \|s^*\|(\beta + 3\sigma\|s^*\|) \geq 0,$$

where s^* is a global minimizer of (1.1). Moreover, we also show that the inequality (1.6) must hold and the SDP relaxation is tight if either $\beta \geq 0$ or H has a nonpositive eigenvalue. Additionally, we prove that all nonzero global minimizers of (1.1) must have the same length when the SDP relaxation (1.5) is tight.

iii) Our approach solves the CQR problem (1.1) via the SDP relaxation (1.5). When it is tight, we not only obtain the true global minimum value, but also obtain all the global minimizers. This is summarized in Algorithm 5.1, which is based on the dual problem of (1.5). Furthermore, the algorithm can also detect cases where the relaxation is not tight, thereby providing a built-in check for tightness.

iv) Numerical experiments demonstrate that our approach is both effective and computationally efficient, providing the first practical method for globally solving the CQR problem. Note that (1.5) is a semidefinite program. The matrix variable Y is $(n+1)$ -by- $(n+1)$, while Z_1 (resp., Z_2) is 3-by-3 (resp., 2-by-2). There are totally 8 equality constraints. Therefore, the SDP relaxation (1.5) can be solved in $O(n^{3.5} \ln(1/\epsilon))$ arithmetic operations by path-following interior-point methods [24]. In our experiments, we apply the software *Mosek* to solve (1.5). It works very efficiently. For instance, when the dimension $n = 1000$, (1.5) can be solved within around one minute.

The rest of this paper is organized as follows. In Section 2, we give a brief survey on polynomial optimization and global optimality conditions for the cubic-quartic regularization problem. In Section 3, we show how to construct the SDP relaxations and investigate their properties. In Section 4, we prove tightness of the SDP relaxations. In Section 5, we propose an algorithm for detecting tightness of the SDP relaxations and computing all global minimizers. The numerical experiments to illustrate the effectiveness and efficiency of our method for solving the CQR problem (1.1) are reported in Section 6. Finally, some conclusions are drawn in Section 7.

2. PRELIMINARIES

Notation. The symbol $\mathbb{N}$ (resp., $\mathbb{R}$) denotes the set of nonnegative integers (resp., real numbers). The superscript T denotes the transpose of a matrix or a vector. For a symmetric matrix M , $M \succeq 0$ (resp., $M \succ 0$) means that M is positive semidefinite (resp., positive definite). The cone of all N -by- N real symmetric positive semidefinite matrices is denoted as S_+^N . For $x \in \mathbb{R}^n$, its Euclidean norm is $\|x\| = \sqrt{x^T x}$. For two matrices $A, B \in \mathbb{R}^{m \times n}$, their inner product is denoted as $A \bullet B = \text{Trace}(AB^T)$.

2.1. Some basics about polynomials. For $x = (x_1, \dots, x_n)$ and $\alpha = (\alpha_1 \dots \alpha_n) \in \mathbb{N}^n$, we denote the monomial power $x^\alpha := x_1^{\alpha_1} \dots x_n^{\alpha_n}$. The degree of α is $|\alpha| := \alpha_1 + \dots + \alpha_n$. For a degree $d > 0$, we denote the power set

$$\mathbb{N}_d^n := \{\alpha \in \mathbb{N}^n : |\alpha| \leq d\}.$$

The column vector of all monomials in x and of degrees up to d is denoted as

$$[x]_d := (x^\alpha)_{\alpha \in \mathbb{N}_d^n}.$$

The length of the vector $[x]_d$ is $\binom{n+d}{d}$. The notation $\mathbb{R}[x]$ denotes the ring of polynomials in $x = (x_1, \dots, x_n)$ and with real coefficients. For a degree d , $\mathbb{R}[x]_d$ denotes the set of polynomials in $\mathbb{R}[x]$ with degrees at most d .

A polynomial $f \in \mathbb{R}[x]$ is said to be a sum of squares (SOS) if there exist polynomials $f_1, \dots, f_k \in \mathbb{R}[x]$ such that $f = f_1^2 + \dots + f_k^2$. The set of all SOS polynomial in $\mathbb{R}[x]$ is denoted as $\Sigma[x]$. Note that $\Sigma[x]$ is a convex cone in $\mathbb{R}[x]$. It is interesting to remark that each polynomial in $\Sigma[x]$ is nonnegative everywhere, while the reverse may not be true [23]. For a degree d , we denote the truncation

$$\Sigma[x]_d := \Sigma[x] \cap \mathbb{R}[x]_d.$$

In particular, we remark that $f \in \Sigma[x]_2$ if and only if $f(x) = [x]_1^T X [x]_1$ for some symmetric matrix $X \succeq 0$.

Let $\mathbb{R}[\|x\|]$ denote the set of univariate polynomials in the norm $\|x\|$, i.e.,

$$\mathbb{R}[\|x\|] = \left\{ \sum_{i=0}^d c_i \|x\|^i : c_i \in \mathbb{R}, d \in \mathbb{N} \right\}.$$

Similarly, a polynomial $p(\|x\|) \in \mathbb{R}[\|x\|]$ is said to be SOS if there exist $p_1, \dots, p_s \in \mathbb{R}[\|x\|]$ such that $p = p_1^2 + \dots + p_s^2$. The set of all SOS polynomials in $\mathbb{R}[\|x\|]$ is denoted as $\Sigma[\|x\|]$. For a degree d , we denote the truncation

$$\Sigma[\|x\|]_d := \Sigma[\|x\|] \cap \mathbb{R}[\|x\|]_d.$$

In this paper, the truncations $\Sigma[\|x\|]_2$ and $\Sigma[\|x\|]_4$ are frequently used. Note that $q(\|x\|) = q_0 + q_1\|x\| + q_2\|x\|^2 \in \Sigma[\|x\|]_2$ if and only if

$$\begin{bmatrix} q_0 & q_1/2 \\ q_1/2 & q_2 \end{bmatrix} \succeq 0.$$

Similarly, $p(\|x\|) = p_0 + p_1\|x\| + p_2\|x\|^2 + p_3\|x\|^3 + p_4\|x\|^4 \in \Sigma[\|x\|]_4$ if and only if

$$\exists t \in \mathbb{R} \quad \text{such that} \quad \begin{bmatrix} p_0 & p_1/2 & t \\ p_1/2 & p_2 - 2t & p_3/2 \\ t & p_3/2 & p_4 \end{bmatrix} \succeq 0.$$

We refer to [23, Section 2.4] for semidefinite representations for SOS polynomials.

2.2. The cubic-quartic regularization problem. We review some basic results by Cartis and Zhu [8] about the cubic-quartic regularization polynomial. Consider the general CQR polynomial

$$(2.1) \quad M_c(s) = f_i + g_i^T s + \frac{1}{2} s^T H_i s + \frac{\beta}{6} \|s\|_W^3 + \frac{\sigma_c}{4} \|s\|_W^4,$$

where $\sigma_c \geq 0$, $\beta \in \mathbb{R}$, and $\|s\|_W = \sqrt{s^T W s}$. Here W is a symmetric positive definite matrix. The gradient and Hessian of $M_c(s)$ are

$$\begin{aligned} \nabla M_c(s) &= g_i + H_i s + \frac{\beta}{2} \|s\|_W (Ws) + \sigma_c \|s\|_W^2 (Ws), \\ \nabla^2 M_c(s) &= H_i + \frac{\beta}{2} \left(W \|s\|_W + \frac{(Ws)(Ws)^T}{\|s\|_W} \right) + \sigma_c \left(\|s\|_W^2 W + 2(Ws)(Ws)^T \right). \end{aligned}$$

The necessary and sufficient conditions for global minimizers of $M_c(s)$ are given by Cartis and Zhu [8] as follows.

Theorem 2.1. [8, Theorem 2.1] *Let s_c be a global minimizer of $M_c(s)$ over $\mathbb{R}^n$ and let*

$$(2.2) \quad B(s_c) := H_i + \frac{\beta}{2} W \|s_c\|_W + \sigma_c W \|s_c\|_W^2.$$

Then, s_c satisfies

$$B(s_c)s_c = -g_i, \quad B(s_c) \succeq 0.$$

If $B(s_c)$ is positive definite, then s_c is the unique global minimizer of $M_c(s)$ in $\mathbb{R}^n$.

Theorem 2.2. [8, Theorem 2.2] *Let $B(s_c)$ be defined as in (2.2). Then, a point s_c is a global minimizer of $M_c(s)$ over $\mathbb{R}^n$ if the following three conditions hold:*

(1) *g is in the range of $B(s_c)$, such that*

$$B(s_c)s_c = \left(H_i + \frac{\beta}{2} W \|s_c\|_W I_n + \sigma_c W \|s_c\|_W^2 I_n \right) s_c = -g_i.$$

(2) *$B(s_c)$ is positive semidefinite:*

$$B(s_c) \succeq 0.$$

(3) *β satisfies*

$$\beta \geq -3\sigma_c \|s_c\|_W, \quad \text{or equivalently} \quad \|s_c\|_W \geq \frac{-\beta}{3\sigma_c}.$$

If, in addition to conditions (1)-(3), either $B(s_c)$ is positive definite or $\beta > -3\sigma_c \|s_c\|_W$, then s_c is the unique global minimizer of $M_c(s)$.

These conditions are very useful for studying the cubic-quartic regularization problem. We refer to [8] for more details.

3. THE SEMIDEFINITE RELAXATION

In this section, we construct an SDP relaxation for solving the cubic-quartic regularization problem (1.1).

Recall the cones $\Sigma[s]_2$, $\Sigma[\|s\|]_4$, $\Sigma[\|s\|]_2$ introduced in Subsection 2.1. Let $M(s)$ be the objective function in (1.1) and let μ^* denote the minimum value of (1.1). For a scalar $\gamma \in \mathbb{R}$, if it holds

$$M(s) - \gamma \in \Sigma[s]_2 + \Sigma[\|s\|]_4 + \|s\| \Sigma[\|s\|]_2,$$

then there exist symmetric matrices X_0, X_1, X_2 such that

$$(3.1) \quad M(s) - \gamma = \begin{bmatrix} 1 \\ s \end{bmatrix}^T X_0 \begin{bmatrix} 1 \\ s \end{bmatrix} + \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix}^T X_1 \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix} + \|s\| \begin{bmatrix} 1 \\ \|s\| \end{bmatrix}^T X_2 \begin{bmatrix} 1 \\ \|s\| \end{bmatrix},$$

$$(3.2) \quad 0 \preceq X_0 \in \mathcal{S}^{n+1}, \quad 0 \preceq X_1 \in \mathcal{S}^3, \quad 0 \preceq X_2 \in \mathcal{S}^2.$$

Note that (3.1) implies $M(s) \geq \gamma$ for all $s \in \mathbb{R}^n$, so $\gamma \leq \mu^*$.

We consider the following relaxation for solving (1.1):

$$(3.3) \quad \begin{cases} \max & \gamma \\ \text{s.t.} & \gamma \text{ satisfies (3.1) - (3.2).} \end{cases}$$

Let γ^* denote the optimal value of (3.3), then $\gamma^* \leq \mu^*$. The relaxation (3.3) is said to be *tight* if $\gamma^* = \mu^*$.

In the following, we derive the dual optimization problem of (3.3). Note $s = (s_1, \dots, s_n)$ and $M(s)$ belongs to the vector space

$$V := \text{Span} \{1, s_1, \dots, s_n, s_1^2, s_1 s_2 \dots, s_n^2, \|s\|, \|s\|^2, \|s\|^3, \|s\|^4\}.$$

Let ℓ be a linear functional on V . Let Y, Z_1, Z_2 be the matrices such that

$$(3.4) \quad Y := \begin{bmatrix} \ell(1) & \ell(s_1) & \ell(s_2) & \cdots & \ell(s_n) \\ \ell(s_1) & \ell(s_1^2) & \ell(s_1 s_2) & \cdots & \ell(s_1 s_n) \\ \ell(s_2) & \ell(s_1 s_2) & \ell(s_2^2) & \cdots & \ell(s_2 s_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell(s_n) & \ell(s_1 s_n) & \ell(s_2 s_n) & \cdots & \ell(s_n^2) \end{bmatrix}$$

$$= \begin{bmatrix} Y_{00} & Y_{01} & Y_{02} & \cdots & Y_{0n} \\ Y_{10} & Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{20} & Y_{21} & Y_{22} & \cdots & Y_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{n0} & Y_{n1} & Y_{n2} & \cdots & Y_{nn} \end{bmatrix},$$

$$(3.5) \quad Z_1 := \begin{bmatrix} \ell(1) & \ell(\|s\|) & \ell(\|s\|^2) \\ \ell(\|s\|) & \ell(\|s\|^2) & \ell(\|s\|^3) \\ \ell(\|s\|^2) & \ell(\|s\|^3) & \ell(\|s\|^4) \end{bmatrix} = \begin{bmatrix} (Z_1)_{11} & (Z_1)_{12} & (Z_1)_{13} \\ (Z_1)_{21} & (Z_1)_{22} & (Z_1)_{23} \\ (Z_1)_{31} & (Z_1)_{32} & (Z_1)_{33} \end{bmatrix},$$

$$(3.6) \quad Z_2 := \begin{bmatrix} \ell(\|s\|) & \ell(\|s\|^2) \\ \ell(\|s\|^2) & \ell(\|s\|^3) \end{bmatrix} = \begin{bmatrix} (Z_2)_{11} & (Z_2)_{12} \\ (Z_2)_{21} & (Z_2)_{22} \end{bmatrix}.$$

For ℓ to be well defined on V , the relation $\|s\|^2 = s_1^2 + \cdots + s_n^2$ poses the condition

$$\ell(\|s\|^2) = \ell(s_1^2) + \cdots + \ell(s_n^2),$$

which implies the equations

$$(3.7) \quad (Z_1)_{13} = (Z_1)_{22} = (Z_1)_{31} = (Z_2)_{12} = (Z_2)_{21} = Y_{11} + \cdots + Y_{nn}.$$

Denote the cone

$$K := \Sigma[s]_2 + \Sigma[\|s\|]_4 + \|s\|\Sigma[\|s\|]_2 \subseteq V.$$

Lemma 3.1. *Let ℓ be a linear functional on V and let Y, Z_1, Z_2 be the matrices as in (3.4), (3.5), (3.6) respectively. Then, $\ell \geq 0$ on K if and only if*

$$Y \succeq 0, \quad Z_1 \succeq 0, \quad Z_2 \succeq 0.$$

Proof. For a set $T \subseteq V$, its conic hull is denoted as

$$\text{cone}(T) := \left\{ \sum_{i=1}^N \lambda_i t_i : \lambda_i \geq 0, t_i \in T, N \in \mathbb{N} \right\}.$$

Observe that

$$\Sigma[s]_2 = \text{cone}\{p_1^2 : p_1 \in \mathbb{R}[s]_1\}, \quad \Sigma[\|s\|]_4 = \text{cone}\{p_2^2 : p_2 \in \mathbb{R}[\|s\|]_2\},$$

$$\|s\|\Sigma[\|s\|]_2 = \text{cone}\{\|s\|p_3^2 : p_3 \in \mathbb{R}[\|s\|]_1\}.$$

So $\ell \geq 0$ on K if and only if

$$\ell(p_1^2) \geq 0, \quad \ell(p_2^2) \geq 0, \quad \ell(p_3^2 \|s\|) \geq 0,$$

for all $p_1 \in \mathbb{R}[s]_1$, $p_2 \in \mathbb{R}[\|s\|]_2$ and $p_3 \in \mathbb{R}[\|s\|]_1$. One can write that

$$p_1 = v_1^T \begin{bmatrix} 1 \\ s \end{bmatrix}, v_1 \in \mathbb{R}^{n+1}, \quad p_2 = v_2^T \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix}, v_2 \in \mathbb{R}^3, \quad p_3 = v_3^T \begin{bmatrix} 1 \\ \|s\| \end{bmatrix}, v_3 \in \mathbb{R}^2.$$

Observe the relations

$$\ell(p_1^2) = v_1^T Y v_1, \quad \ell(p_2^2) = v_2^T Z_1 v_2, \quad \ell(p_3^2 \|s\|) = v_3^T Z_2 v_3.$$

Then the conclusion of the lemma follows from the above. $\square$

Define the linear function

$$\vartheta(Y, Z_1, Z_2) := \ell(M(s)) = \begin{bmatrix} f_0 & g^T/2 \\ g/2 & H/2 \end{bmatrix} \bullet Y + \frac{\beta}{6}(Z_2)_{22} + \frac{\sigma}{4}(Z_1)_{33}.$$

It is interesting to note that

$$\vartheta(Y, Z_1, Z_2) - \gamma = \ell(M(s) - \gamma) + \gamma(1 - Y_{00}).$$

Thus, if $Y_{00} = 1$, $\ell \geq 0$ on K , and γ is feasible for (3.3), then $\vartheta(Y, Z_1, Z_2) \geq \gamma$. By Lemma 3.1, the dual optimization problem of the relaxation (3.3) is

$$(3.8) \quad \begin{cases} \min_{Y, Z_1, Z_2} & \vartheta(Y, Z_1, Z_2) \\ \text{s.t.} & Y_{00} = 1, \quad (Z_1)_{11} = Y_{00}, \\ & (Z_1)_{12} = (Z_2)_{11}, \quad (Z_1)_{22} = (Z_2)_{12}, \\ & (Z_1)_{13} = (Z_2)_{12}, \quad (Z_1)_{23} = (Z_2)_{22}, \\ & (Z_1)_{13} = (Z_1)_{22} = Y_{11} + \cdots + Y_{nn}, \\ & Y \in \mathcal{S}_+^{n+1}, \quad Z_1 \in \mathcal{S}_+^3, \quad Z_2 \in \mathcal{S}_+^2. \end{cases}$$

Note that (3.8) is a semidefinite program. The matrix variable Y is $(n+1)$ -by- $(n+1)$, while Z_1 (resp., Z_2) is 3-by-3 (resp., 2-by-2). There are totally 8 equality constraints. Therefore, the relaxation (1.5) can be solved in $O(n^{3.5} \ln(1/\epsilon))$ arithmetic operations by path-following interior-point methods [24].

For a point $s = (s_1, \dots, s_n) \in \mathbb{R}^n$, let

$$Y(s) = \begin{bmatrix} 1 & s^T \\ s & ss^T \end{bmatrix} \succeq 0,$$

$$Z_1(s) = \begin{bmatrix} 1 & \|s\| & \|s\|^2 \\ \|s\| & \|s\|^2 & \|s\|^3 \\ \|s\|^2 & \|s\|^3 & \|s\|^4 \end{bmatrix} \succeq 0, \quad Z_2(s) = \begin{bmatrix} \|s\| & \|s\|^2 \\ \|s\|^2 & \|s\|^3 \end{bmatrix} \succeq 0.$$

Then, $(Y(s), Z_1(s), Z_2(s))$ is a feasible triple for (3.8) and

$$M(s) = \vartheta(Y(s), Z_1(s), Z_2(s)).$$

Let ϑ^* denote the optimal value of (3.8). The above implies that ϑ^* is less than or equal to the minimum value μ^* of (1.1), i.e., $\vartheta^* \leq \mu^*$. The relaxation (3.8) is said to be *tight* if $\vartheta^* = \mu^*$.

The following is a basic property about the relaxations (3.3) and (3.8). In particular, the relaxation (3.3) is tight if and only if (3.8) is tight.

Theorem 3.2. *Let μ^* , γ^* and v^* denote the optimal values of (1.1), (3.3) and (3.8) respectively. Then, it holds that*

$$(3.9) \quad \gamma^* = \vartheta^* \leq \mu^*.$$

Moreover, the optimal value γ^* is attainable for (3.3).

Proof. Suppose γ and (Y, Z_1, Z_2) are feasible points of (3.3) and (3.8) respectively. Then it holds

$$\vartheta(Y(s), Z_1(s), Z_2(s)) - \gamma = \text{Trace}(X_0 Y) + \text{Trace}(X_1 Z_1) + \text{Trace}(X_2 Z_2) \geq 0,$$

so $\gamma^* \leq \vartheta^*$. We show that (3.8) has a strictly feasible point. Let ν denote the normal distribution on $\mathbb{R}^n$ and

$$\hat{Y} = \int Y(s) d\nu, \quad \hat{Z}_1 = \int Z_1(s) d\nu, \quad \hat{Z}_2 = \int Z_2(s) d\nu.$$

Since $Y \in \mathcal{S}_+^{n+1}$, $Z_1 \in \mathcal{S}_+^3$, $Z_2 \in \mathcal{S}_+^2$, $\hat{Y}, \hat{Z}_1, \hat{Z}_2$ are all positive definite matrices. Then $(\hat{Y}, \hat{Z}_1, \hat{Z}_2)$ is a strictly feasible point for (3.8). By the strong duality theorem, we must have $\gamma^* = \vartheta^*$ and γ^* is attainable. The inequality $\vartheta^* \leq \mu^*$ is shown earlier. $\square$

Suppose (Y^*, Z_1^*, Z_2^*) is an optimizer of (3.8). If there exists a point $s^* = (s_1^*, \dots, s_n^*)$ such that

$$(3.10) \quad Y^* = \begin{bmatrix} 1 & s^{*T} \\ s^* & s^* s^{*T} \end{bmatrix} \succeq 0,$$

$$Z_1^* = \begin{bmatrix} 1 & \|s^*\| & \|s^*\|^2 \\ \|s^*\| & \|s^*\|^2 & \|s^*\|^3 \\ \|s^*\|^2 & \|s^*\|^3 & \|s^*\|^4 \end{bmatrix} \succeq 0, \quad Z_2^* = \begin{bmatrix} \|s^*\| & \|s^*\|^2 \\ \|s^*\|^2 & \|s^*\|^3 \end{bmatrix} \succeq 0,$$

then it holds

$$M(s^*) = \vartheta(Y^*, Z_1^*, Z_2^*) = \vartheta^*.$$

Since $\mu^* \leq M(s^*)$, the above together with (3.9) implies $\mu^* = M(s^*)$. This means that s^* is a global minimizer of $M(s)$. Indeed, if all Y^* , Z_1^* and Z_2^* are rank one, then (3.10) holds with $s^* = (Y_{1,1}^*, \dots, Y_{n,1}^*)$. Thus, we get the following conclusion.

Proposition 3.3. *Suppose (Y^*, Z_1^*, Z_2^*) is an optimizer of (3.8). If all Y^* , Z_1^* , Z_2^* are rank one, then the SDP relaxations (3.3) and (3.8) are tight (i.e., $\gamma^* = \vartheta^* = \mu^*$), and the above s^* is a global minimizer of $M(s)$.*

4. TIGHTNESS OF THE RELAXATIONS

In this section, we prove the tightness of SDP relaxations (3.3) and (3.8) under some general assumptions. Consider the CQR problem (1.1). We assume $M(s)$ has a global minimizer s^* . This can be ensured by that $\sigma > 0$, or $\sigma = 0$ but $\beta > 0$, or $\sigma = \beta = 0$ but H is positive definite.

In Subsection 2.2, when the weight matrix W is identity, the matrix function in (2.2) reduces to

$$B(s) = H + \frac{\beta}{2} \|s\| I + \sigma \|s\|^2 I.$$

By Theorem 2.1 (also see [8]), if s^* is a global minimizer of $M(s)$, then it must satisfy

$$B(s^*)s^* = -g, \quad B(s^*) \succeq 0.$$

If $B(s^*) \succ 0$, then s^* is the unique global minimizer. By Theorem 2.2 (also see [8]), the point s^* must be a global minimizer of $M(s)$ if it satisfies

$$B(s^*)s^* = -g, \quad B(s^*) \succeq 0, \quad \beta \geq -3\sigma \|s^*\|.$$

In addition to the above, if $B(s^*) \succ 0$ or $\beta > -3\sigma \|s^*\|$, then s^* is the unique global minimizer of (1.1).

First, we prove a sufficient and necessary condition for the SDP relaxations (3.3) and (3.8) to be tight. Recall that μ^* , γ^* , ϑ^* denote the optimal values of (1.1), (3.3), (3.8) respectively. It is worthy to note that (3.3) is tight if and only if (3.8) is tight. This is shown in Theorem 3.2.

Theorem 4.1. *Let s^* be a global minimizer of (1.1). Then, the SDP relaxations (3.3) and (3.8) are tight (i.e., $\gamma^* = \vartheta^* = \mu^*$) if and only if*

$$(4.1) \quad \|s^*\|(\beta + 3\sigma \|s^*\|) \geq 0.$$

Furthermore, if (4.1) holds for one minimizer s^* , then it holds for all minimizers.

Proof. “ $\Leftarrow$ ” Assume (4.1) holds, we need to show that (3.3) and (3.8) are both tight.

First, we consider the case that $\|s^*\| = 0$, i.e., $s^* = 0$. By Theorem 2.1, we have $-g = B(s^*)s^* = 0$. Since $\mu^* = M(s^*)$, we have

$$M(s) - \mu^* = \frac{1}{2} s^T H s + \frac{\beta}{6} \|s\|^3 + \frac{\sigma}{4} \|s\|^4 \geq 0 \quad \text{for all } s \in \mathbb{R}^n.$$

Let $\lambda_{\min}$ denote the smallest eigenvalue of H with the eigenvector $\hat{s}$, then

$$M(s) - \mu^* = \frac{1}{2} s^T (H - \lambda_{\min} I_n) s + \frac{1}{2} \lambda_{\min} \|s\|^2 + \frac{\beta}{6} \|s\|^3 + \frac{\sigma}{4} \|s\|^4.$$

In particular, it holds

$$M(\hat{s}) - \mu^* = \frac{1}{2} \lambda_{\min} \|\hat{s}\|^2 + \frac{\beta}{6} \|\hat{s}\|^3 + \frac{\sigma}{4} \|\hat{s}\|^4.$$

Note $M(s) - \mu^* \geq 0$ for all $s \in \mathbb{R}^n$ and the eigenvector $\hat{s}$ can be selected such that $\|\hat{s}\|$ is arbitrary positive number. So, we get

$$\sigma(z) := \frac{1}{2}\lambda_{\min}z^2 + \frac{\beta}{6}z^3 + \frac{\sigma}{4}z^4 \geq 0 \quad \text{for all } z \geq 0.$$

By Theorem 3.2.1 of [23], there exist a quadratic polynomial $p(z)$ and a linear polynomial $q(z)$ such that $\sigma(z) = p(z)^2 + zq(z)^2$. Hence,

$$M(s) - \mu^* = \frac{1}{2}s^T(H - \lambda_{\min}I_n)s + p(\|s\|)^2 + \|s\|q(\|s\|)^2.$$

Since $H - \lambda_{\min}I_n \succeq 0$, we get

$$M(s) - \mu^* \in \Sigma[s]_2 + \Sigma[\|s\|]_4 + \|s\|\Sigma[\|s\|]_2,$$

which implies $\mu^* \leq \gamma^*$.

Second, consider the case $\|s^*\| > 0$, then $\beta + 3\sigma\|s^*\| \geq 0$. Let $w = s - s^*$. As in the proof of Theorem 2.2 of [8], we have the expansion (note $B(s^*)s^* + g = 0$),

$$(4.2) \quad M(s) - \mu^* = [B(s^*)s^* + g]^T w + \frac{1}{2}w^T B(s^*)w + \mathcal{F}_2$$

$$(4.3) \quad = \frac{1}{2}w^T B(s^*)w + \mathcal{F}_2,$$

where

$$\begin{aligned} \mathcal{F}_2 &= \frac{1}{2}(\|s\| - \|s^*\|)^2 \left[\frac{\beta}{6}(\|s^*\| + 2\|s\|) + \frac{\sigma}{2}(\|s^*\| + \|s\|)^2 \right] \\ &= \frac{1}{2}(\|s\| - \|s^*\|)^2 \left[\frac{\beta + 3\sigma\|s^*\|}{6}(\|s^*\| + 2\|s\|) + \frac{\sigma}{2}\|s\|^2 \right]. \end{aligned}$$

Since $\beta + 3\sigma\|s^*\| \geq 0$, we have

$$\mathcal{F}_2 \in \Sigma[\|s\|]_4 + \|s\|\Sigma[\|s\|]_2.$$

Since s^* is a global minimizer, we have $B(s^*) \succeq 0$ by Theorem 2.1, so

$$M(s) - \mu^* \in \Sigma[s]_2 + \Sigma[\|s\|]_4 + \|s\|\Sigma[\|s\|]_2.$$

This also implies $\mu^* \leq \gamma^*$.

In either case $\|s^*\| = 0$ or $\|s^*\| > 0$, we have shown $\mu^* \leq \gamma^*$. On the other hand, it holds $\gamma^* = \vartheta^* \leq \mu^*$ by Theorem 3.2. Therefore, we can further get $\gamma^* = \vartheta^* = \mu^*$, i.e., the SDP relaxations (3.3) and (3.8) are tight.

“ $\Rightarrow$ ” Assume the SDP relaxations (3.3) and (3.8) are tight. We need to show that (4.1) holds. Note that if $s^* = 0$, then (4.1) automatically holds. So we consider the general case that $s^* \neq 0$. Since the relaxation (3.3) is tight, its optimal value $\gamma^* = \mu^* = M(s^*)$, which is achievable for (3.3) by Theorem 3.2. So,

$$M(s) - M(s^*) = [s]_1^T Q[s]_1 + \sigma(\|s\|),$$

where Q is a symmetric psd matrix and $\sigma(\|s\|) \in \Sigma[\|s\|]_4 + \|s\|\Sigma[\|s\|]_2$. Note that σ is a univariate polynomial in $\|s\|$ and $\sigma(z) \geq 0$ for all $z \geq 0$. By Theorem 3.2.1 of [23], we get

$$\sigma(\|s\|) = p(\|s\|)^2 + \|s\|q(\|s\|)^2$$

for a quadratic polynomial $p(z)$ and a linear polynomial $q(z)$, hence

$$M(s) - M(s^*) = [s]_1^T Q[s]_1 + p(\|s\|)^2 + \|s\|q(\|s\|)^2.$$

Note that

$$0 = [s^*]_1^T Q[s^*]_1 + p(\|s^*\|)^2 + \|s^*\|q(\|s^*\|)^2.$$

Because all the terms in the right hand side above are nonnegative,

$$[s^*]_1^T Q[s^*]_1 = p(\|s^*\|)^2 = \|s^*\|q(\|s^*\|)^2 = 0.$$

Since $s^* \neq 0$, we have $\|s^*\| > 0$ and hence

$$p(\|s^*\|) = q(\|s^*\|) = 0.$$

Without loss of generality, assume the leading coefficients of p and q are nonnegative. The coefficient of $\|s\|^4$ of $p(\|s\|)^2$ is $\frac{\sigma}{4}$, so we can factorize p, q as

$$\begin{aligned} p(\|s\|) &= \frac{\sqrt{\sigma}}{2}(\|s\| - \|s^*\|)(\|s\| + \tau), \\ q(\|s\|) &= \sqrt{\eta}(\|s\| - \|s^*\|), \end{aligned}$$

for some real scalars $\tau \in \mathbb{R}$ and $\eta \geq 0$. Hence, it holds

$$\begin{aligned} p(\|s\|)^2 + \|s\|q(\|s\|)^2 &= (\|s\| - \|s^*\|)^2 \left[\frac{\sigma}{4}(\|s\| + \tau)^2 + \eta\|s\| \right] \\ &= \left(\|s\|^2 - 2\|s\|\|s^*\| + \|s^*\|^2 \right) \left[\frac{\sigma}{4}\|s\|^2 + \left(\frac{\sigma\tau}{2} + \eta \right)\|s\| + \frac{\sigma\tau^2}{4} \right]. \end{aligned}$$

To match the representation of $M(s) - M(s^*)$, the coefficient of $\|s\|$ in $p(\|s\|)^2 + \|s\|q(\|s\|)^2$ must be zero, so

$$\|s^*\|^2 \left(\frac{\sigma\tau}{2} + \eta \right) - \|s^*\| \frac{\sigma\tau^2}{2} = 0.$$

The coefficient of $\|s\|^3$ in $p(\|s\|)^2 + \|s\|q(\|s\|)^2$ must be $\frac{\beta}{6}$, so

$$\left(\frac{\sigma\tau}{2} + \eta \right) - \frac{\|s^*\|\sigma}{2} = \frac{\beta}{6}.$$

The above equations imply that

$$\|s^*\|^2 \left(\frac{\sigma\tau}{2} + \eta \right) = \|s^*\| \frac{\sigma\tau^2}{2}, \quad \left(\frac{\sigma\tau}{2} + \eta \right) = \frac{1}{6}(\beta + 3\sigma\|s^*\|).$$

Therefore, it holds

$$\frac{\|s^*\|^2}{6}(\beta + 3\sigma\|s^*\|) = \|s^*\| \frac{\sigma\tau^2}{2} \geq 0.$$

Since $\|s^*\| \geq 0$ and $\sigma \geq 0$, the condition (4.1) must hold.

If (4.1) holds for one minimizer s^* , then the relaxations (3.3) and (3.8) are tight. This further implies that (4.1) must also hold for all other minimizers, if they exist. Therefore, if (4.1) holds for one minimizer, then it holds for all minimizers. $\square$

When does the condition (4.1) hold? An interesting case is that H has a non-positive eigenvalue. Since s^* is a global minimizer, the expansion formula (4.2) shows

$$(4.4) \quad \frac{1}{2}w^T B(s^*)w + \mathcal{F}_2 \geq 0 \quad \text{for all } w \in \mathbb{R}^n.$$

By Theorem 2.1, the global optimality condition implies

$$(4.5) \quad B(s^*) = H + \frac{\beta}{2}\|s^*\|I_n + \sigma\|s^*\|^2 I_n \succeq 0.$$

Suppose H has an eigenvalue $\lambda \leq 0$. Then, the above implies

$$\frac{\beta}{2}\|s^*\| + \sigma\|s^*\|^2 \geq -\lambda \geq 0.$$

Since $\sigma \geq 0$, one can see that

$$\|s^*\|(\beta + 3\sigma\|s^*\|) \geq 2\left(\frac{\beta}{2}\|s^*\| + \sigma\|s^*\|^2\right) \geq 0.$$

So, (4.1) holds, and we get the following corollary.

Corollary 4.2. *If either $\beta \geq 0$ or H has a nonpositive eigenvalue, then the SDP relaxations (3.3) and (3.8) are tight, i.e., $\gamma^* = \vartheta^* = \mu^*$.*

Theorem 4.1 implies that the SDP relaxations (3.3) and (3.8) are tight if and only if $\|s^*\|(\beta + 3\sigma\|s^*\|) \geq 0$ for every minimizer s^* of (1.1). Indeed, if this inequality holds for one minimizer, then it holds for all minimizers. Moreover, when the relaxations are tight, we can show that all nonzero global minimizers have the same length.

Theorem 4.3. *Assume at least one of σ and β is nonzero and the SDP relaxations (3.3) and (3.8) are tight. If $s^*, \hat{s}$ are two minimizers of (1.1), then*

$$(4.6) \quad \|\hat{s}\| \cdot \|s^*\| \cdot (\|\hat{s}\| - \|s^*\|) = 0.$$

Proof. If one of $\|\hat{s}\|$ and $\|s^*\|$ is zero, then (4.6) automatically holds. Now suppose they are both nonzero. By Theorem 3.2, the optimal value of (3.3) is achievable. Since (3.3) and (3.8) are tight relaxations, as in the proof of Theorem 4.1, we can show that there exist a psd matrix Q , a quadratic polynomial $p(z)$ and a linear polynomial $q(z)$ such that

$$M(s) - \mu^* = [s]_1^T Q [s]_1 + p(\|s\|)^2 + \|s\|q(\|s\|)^2.$$

Since $\hat{s}, s^*$ are minimizers of (1.1), one can see

$$M(s^*) - \mu^* = M(\hat{s}) - \mu^* = 0.$$

Both $\|\hat{s}\|$ and $\|s^*\|$ are positive, so

$$p(\|\hat{s}\|) = p(\|s^*\|) = q(\|\hat{s}\|) = q(\|s^*\|) = 0.$$

We discuss in two cases.

- Suppose q is not identically zero. Then, both $\|\hat{s}\|$ and $\|s^*\|$ are roots of $q(z) = 0$. Since the degree of q is one, we can get $\|\hat{s}\| = \|s^*\|$.
- Suppose q is identically zero. As in the proof of Theorem 4.1, we can factorize that

$$p(\|s\|) = \frac{\sqrt{\sigma}}{2}(\|s\| - \|s^*\|)(\|s\| + \tau)$$

for some real scalar τ . Since $q = 0$ and the coefficient of $\|s\|^3$ (resp., $\|s\|$) in $p(\|s\|)^2$ is $\frac{\beta}{6}$ (resp., 0), we have

$$\frac{\sigma}{2}(\tau - \|s^*\|) = \frac{\beta}{6}, \quad \frac{\sigma}{2}\tau\|s^*\|(\|s^*\| - \tau) = 0.$$

Note $\|s^*\| > 0$ and the above implies $\beta = 0$ if $\sigma = 0$. Since σ and β are not both zero, we must have $\sigma > 0$. If $\tau \neq 0$, then $\tau = \|s^*\|$ and hence

$$p(\|s\|) = \frac{\sqrt{\sigma}}{2}(\|s\|^2 - \|s^*\|^2).$$

Since $p(\|\hat{s}\|) = 0$, we get $\|\hat{s}\| = \|s^*\|$. If $\tau = 0$, then

$$p(\|s\|) = \frac{\sqrt{\sigma}}{2}\|s\|(\|s\| - \|s^*\|).$$

Since $p(\|\hat{s}\|) = 0$ and $\|\hat{s}\| > 0$, we also get $\|\hat{s}\| = \|s^*\|$.

Therefore, we get $\|\hat{s}\| = \|s^*\|$ for both cases. $\square$

The following are some exposition examples for the above conclusions.

Example 4.4. (i) Consider the CQR problem

$$\min_{s \in \mathbb{R}^1} \|s\|^4 - 4\|s\|^3 + 6s^2 - 4s + 1.$$

In terms of the formulation of (1.1), this corresponds to $\sigma = 4$, $\beta = -24$, $H = 12$, $g = -4$ and $f_0 = 1$. The unique global minimizer is $s^* = 1$. By solving (3.8), we can get the optimizer

$$Y^* = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad Z_1^* = \begin{bmatrix} 1 & \frac{1}{2} & 1 \\ \frac{1}{2} & 1 & 2 \\ 1 & 2 & 4 \end{bmatrix}, \quad Z_2^* = \begin{bmatrix} \frac{1}{2} & 1 \\ 1 & 2 \end{bmatrix}.$$

The optimal values of (3.3) and (3.8) are $\gamma^* = \vartheta^* = -1$. However, the minimum value $\mu^* = 0$. The SDP relaxations (3.3) and (3.8) are not tight. The condition (4.1) fails to hold.

(ii) Consider the CQR problem

$$\min_{s \in \mathbb{R}^1} \|s\|^4 - 6\|s\|^3 + 13s^2 - 12s + 4.$$

This corresponds to parameters $\sigma = 4$, $\beta = -36$, $H = 26$, $g = -12$ and $f_0 = 4$. There are two global optimizers: $s^* = 1$ or 2 . By solving (3.8), we get the optimizer

$$Y^* = \begin{bmatrix} 1 & 1.5 \\ 1.5 & 2.25 \end{bmatrix}, \quad Z_1^* = \begin{bmatrix} 1 & 0.75 & 2.25 \\ 0.75 & 2.25 & 6.75 \\ 2.25 & 6.75 & 20.25 \end{bmatrix}, \quad Z_2^* = \begin{bmatrix} 0.75 & 2.25 \\ 2.25 & 6.75 \end{bmatrix}.$$

The optimal values of (3.3) and (3.8) are $\gamma^* = \vartheta^* = -5$, while the minimum value $\mu^* = 0$. The relaxations (3.3) and (3.8) are not tight. The condition (4.1) fails to hold. This can also be implied by Theorem 4.3.

(iii) When $s^* = 0$ is a global minimizer of CQR problem, the condition $\beta \geq -3\sigma\|s^*\|$ may not be necessary for (3.3) and (3.8) to be tight. Consider the CQR problem

$$\min_{s \in \mathbb{R}^n} \|s\|^2(\|s\| - 1)^2 + \epsilon\|s\|^3,$$

for a scalar $0 < \epsilon < 12$. The corresponding parameters are: $\sigma = 4$, $\beta = -12 + \epsilon$, and $H = 2I_n$. The minimum value $\mu^* = 0$ and the unique minimizer is $s^* = 0$. Clearly, the relaxations (3.3) and (3.8) are tight, i.e., $\gamma^* = \vartheta^* = 0$. The condition (4.1) holds.

(iv) Consider the CQR problem

$$\min_{s \in \mathbb{R}^n} \|s\|^4 - 4\|s\|^3 + 4\|s\|^2 = \|s\|^2(\|s\| - 2)^2.$$

The corresponding parameters are: $\sigma = 4$, $\beta = -24$, and $H = 8I_n$. The minimum value $\mu^* = 0$, and the global minimizers are 0 and all points s^* such that $\|s^*\| = 2$. Clearly, the relaxations (3.3) and (3.8) are tight. The condition (4.1) holds.

5. CERTIFYING TIGHTNESS AND EXTRACT MINIMIZERS

This section discusses how to detect tightness of the relaxations (3.3) and (3.8), and how to extract global minimizers of (1.1).

As shown in Theorem 3.2, the optimal value γ^* of (3.3) is attainable. There exist symmetric matrices X_0^*, X_1^*, X_2^* such that

$$M(s) - \gamma^* = \begin{bmatrix} 1 \\ s \end{bmatrix}^T X_0^* \begin{bmatrix} 1 \\ s \end{bmatrix} + \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix}^T X_1^* \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix} + \|s\| \begin{bmatrix} 1 \\ \|s\| \end{bmatrix}^T X_2^* \begin{bmatrix} 1 \\ \|s\| \end{bmatrix},$$

$$X_0^* \in \mathcal{S}_+^{n+1}, \quad X_1^* \in \mathcal{S}_+^3, \quad X_2^* \in \mathcal{S}_+^2.$$

Suppose (Y^*, Z_1^*, Z_2^*) is an optimizer of (3.8). Let ℓ^* be the linear functional determined by (Y^*, Z_1^*, Z_2^*) as in (3.4)-(3.6). Note that

$$0 = \vartheta^* - \gamma^* = \vartheta(Y^*, Z_1^*, Z_2^*) - \gamma^* = \ell^*(M(s) - \gamma^*) = Y^* \bullet X_0^* + Z_1^* \bullet X_1^* + Z_2^* \bullet X_2^*.$$

Since all the matrices are psd, we can further get

$$Y^* \bullet X_0^* = Z_1^* \bullet X_1^* = Z_2^* \bullet X_2^* = 0,$$

$$\text{rank } Y^* + \text{rank } X_0^* \leq n + 1,$$

$$\text{rank } Z_1^* + \text{rank } X_1^* \leq 3, \quad \text{rank } Z_2^* + \text{rank } X_2^* \leq 2.$$

We can write $X_0^* = R^T R$ for some $R \in \mathbb{R}^{r \times (n+1)}$ with $r \leq n+1$. Note $\text{rank } Z_1^* \geq 1$, so $r_1 := \text{rank } X_1^* \leq 2$. Also note $r_2 := \text{rank } X_2^* \leq 2$. Then, there exist $a_1, a_2 \in \mathbb{R}^3$ and $b_1, b_2 \in \mathbb{R}^2$ such that

$$\begin{aligned} X_1^* &= a_1 a_1^T + a_2 a_2^T, \\ X_2^* &= b_1 b_1^T + b_2 b_2^T. \end{aligned}$$

In the above, we can let $a_2 = 0$ if $r_1 = 1$, $b_2 = 0$ if $r_2 = 1$, and $b_1 = b_2 = 0$ if $r_2 = 0$. Then

$$\begin{aligned} \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix}^T X_1^* \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix} &= \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix}^T (a_1 a_1^T + a_2 a_2^T) \begin{bmatrix} 1 \\ \|s\| \\ \|s\|^2 \end{bmatrix} \\ &= (a_{10} + a_{11}\|s\| + a_{12}\|s\|^2)^2 \\ &\quad + (a_{20} + a_{21}\|s\| + a_{22}\|s\|^2)^2. \end{aligned}$$

Similarly,

$$\begin{aligned} \begin{bmatrix} 1 \\ \|s\| \end{bmatrix}^T X_2^* \begin{bmatrix} 1 \\ \|s\| \end{bmatrix} &= \begin{bmatrix} 1 \\ \|s\| \end{bmatrix}^T (b_1 b_1^T + b_2 b_2^T) \begin{bmatrix} 1 \\ \|s\| \end{bmatrix} \\ &= (b_{10} + b_{11}\|s\|)^2 + (b_{20} + b_{21}\|s\|)^2. \end{aligned}$$

Denote the univariate polynomials

$$(5.1) \quad p_i(z) = a_{i0} + a_{i1}z + a_{i2}z^2, \quad i = 1, 2,$$

$$(5.2) \quad q_j(z) = b_{j0} + b_{j1}z, \quad j = 1, 2.$$

Then, it holds

$$(5.3) \quad M(s) - \gamma^* = [s]_1^T R^T R [s]_1 + \sum_{i=1}^2 p_i(\|s\|)^2 + \|s\| \sum_{j=1}^2 q_j(\|s\|)^2.$$

Suppose s^* is a global minimizer of (1.1). Note that

$$M(s^*) - \gamma^* = [s^*]_1^T R^T R[s^*]_1 + \sum_{i=1}^2 p_i(\|s^*\|)^2 + \|s^*\| \sum_{j=1}^2 q_j(\|s^*\|)^2.$$

If the relaxation (3.3) is tight, i.e., $\gamma^* = M(s^*)$, then s^* is a solution to

$$(5.4) \quad \begin{cases} R[s]_1 = 0, \\ p_1(\|s\|) = p_2(\|s\|) = 0, \\ \|s\|q_1(\|s\|) = \|s\|q_2(\|s\|) = 0. \end{cases}$$

Conversely, if $s^* \in \mathbb{R}^n$ satisfies (5.4), then

$$M(s^*) - \gamma^* = 0.$$

Since $\gamma^* = \mu^* \leq M(s^*)$, the above implies s^* is a global minimizer of (1.1) and the relaxation (3.3) is tight. The latter further implies (3.8) is also tight. Moreover, by Theorems 4.1 and 4.3, there exists $z^* > 0$ with $\beta \geq -3\sigma z^*$ such that for all nonzero solutions s^* to (5.4), if there are any, it holds

$$\|s^*\| = z^*.$$

Summarizing the above, we get the following algorithm for checking tightness of (3.3) and extracting global minimizers for (1.1).

Algorithm 5.1. *Let $\mathcal{S} := \emptyset$ and do the following:*

Step 1. Solve the SDP relaxation (3.3) and (3.8), get the optimal values γ^ , ϑ^* , and the representation (5.3).*

Step 2. If the zero vector 0 satisfies (5.4), let $\mathcal{S} := \mathcal{S} \cup \{0\}$.

Step 3. Solve the following system of univariate equations:

$$p_1(z) = p_2(z) = q_1(z) = q_2(z) = 0.$$

If they have a common real zero z^ , then go to the next step.*

Step 4. Output the following set and stop:

$$\mathcal{S} := \mathcal{S} \cup \{s \in \mathbb{R}^n : \|s\| = z^*, R[s]_1 = 0\}.$$

In the above, we have seen that the system (5.4) has a solution if and only if the relaxations (3.3) and (3.8) are tight, for which case we can get all minimizers of (1.1) by Algorithm 5.1. Thus, we get the following theorem.

Theorem 5.2. *The SDP relaxations (3.3) and (3.8) are tight if and only if the system (5.4) has a solution. If they are tight, the set $\mathcal{S}$ output by Algorithm 5.1 consists of all global minimizers of (1.1); if otherwise they are not tight, the output set $\mathcal{S}$ is empty.*

The following is an illustrative example for Algorithm 5.1.

Example 5.3. Consider the CQR problem

$$\min_{s \in \mathbb{R}^{10}} 7s_1^2 - \sum_{i=2}^{10} s_i^2 + \sum_{i=1}^5 s_1 s_{2i} - \sum_{i=1}^4 s_1 s_{2i+1} + \|s\|^4.$$

The corresponding parameters are $\sigma = 4$, $\beta = 0$. By solving (3.8), we get the optimal triple (Y^*, Z_1^*, Z_2^*) with $\text{rank } Y^* = 2$, $\text{rank } Z_1^* = \text{rank } Z_2^* = 1$. It took

around 0.03 second. Algorithm 5.1 produces

$$\begin{aligned} p_1(z) &\approx -0.7071 + 0.7071z^2, \\ p_2(z) &\approx 0.4082 - 0.8165z + 0.4083z^2, \\ q_1(z) &\approx -0.7071 + 0.7071z, \end{aligned}$$

and $R \in \mathbb{R}^{9 \times 11}$, which has rank 9. For neatness, we do not display R here. Solving $p_1(z) = p_2(z) = q_1(z) = 0$, we can get $z^* = 1$. In Step 4, the minimizers are given as

$$\begin{aligned} s^* &= t(-1, 1, -1, 1, -1, 1, -1, 1, -1, 1), \\ \|s^*\| &= 1, \quad t \in \mathbb{R}. \end{aligned}$$

Therefore, we get two global minimizers

$$s^* = \frac{\pm 1}{\sqrt{10}}(-1, 1, -1, 1, -1, 1, -1, 1, -1, 1).$$

The optimal values are $\gamma^* = \vartheta^* = \mu^* = -1$. The relaxations (3.3) and (3.8) are tight.

6. NUMERICAL EXPERIMENTS

We present some numerical experiments for solving the cubic-quartic regularization problem (1.1). The SDP relaxations (3.3) and (3.8) are solved by the software **Mosek** [1]. The computations are implemented in MATLAB R2022b on a Lenovo Laptop with CPU@2.10GHz and RAM 16.0G. For neatness of presentation, all computational results are displayed in four decimal digits.

Recall that μ^* denotes the minimum value of (1.1), γ^* denotes the maximum value of (3.3), ϑ^* denotes the minimum value of (3.8), and s^* denotes the computed global minimizer of (1.1). We remark that by Theorem 3.2, γ^* gives a lower bound for μ^* , and $\mu^* \leq M(s)$ for all $s \in \mathbb{R}^n$. The SDP relaxations (3.3) and (3.8) are tight if $M(s^*) = \gamma^*$. In computational practice, one typically cannot have $M(s^*)$ equal to γ^* exactly, due to numerical errors. To measure the numerical accuracy of s^* , we use the absolute and relative errors

$$\mathbf{err-abs} = |M(s^*) - \gamma^*|, \quad \mathbf{err-rel} = \left| \frac{M(s^*) - \gamma^*}{M(s^*)} \right|.$$

Example 6.1. Consider the CQR problem

$$\min_{s \in \mathbb{R}^5} \sum_{1 \leq i < j \leq 5} s_i s_j - \frac{5}{2} \sum_{i=1}^5 s_i^2 - \|s\|^3 + \|s\|^4.$$

Solving (3.3) and (3.8), we get $\text{rank } Y^* = 5$, $\text{rank } Z_1^* = \text{rank } Z_2^* = 1$. It took around 0.04 second. By Algorithm 5.1, we get

$$\begin{aligned} p_1(z) &\approx -0.9501 + 0.0703z + 0.3040z^2, \\ p_2(z) &\approx 0.0926 - 0.8669z + 0.4898z^2, \\ q_1(z) &\approx -0.8560 + 0.5170z, \\ R &\approx [0, -0.4472, -0.4472, -0.4472, -0.4472]. \end{aligned}$$

In Step 3, we get $z^* \approx 1.6559$. The set of global minimizers can be parameterized as $s^* = t_1 \xi_1 + t_2 \xi_2 + t_3 \xi_3 + t_4 \xi_4$ with $\|s^*\| = 1.6559$, where

$$\begin{aligned} \xi_1 &= (1, -1, 0, 0, 0), & \xi_2 &= (1, 0, -1, 0, 0), \\ \xi_3 &= (1, 0, 0, -1, 0), & \xi_4 &= (1, 0, 0, 0, -1). \end{aligned}$$

The optimal values are $\gamma^* = \vartheta^* = \mu^* \approx -5.2479$. The relaxations (3.3) and (3.8) are tight. Particularly, $\hat{s}^* = (1.1709, -1.1709, 0, 0, 0)$ is a global minimizer, at which the errors are

$$\mathbf{err-abs} = 1.08 \cdot 10^{-8}, \quad \mathbf{err-rel} = 2.06 \cdot 10^{-9}.$$

We compare our method with some classical nonlinear optimization methods, e.g., the function `fminunc` in MATLAB with default setting of parameters. The convergence of `fminunc` highly depends on initial point. In our experiments, we used several initial points for `fminunc`. Let s_0 denote the initial point and f^* denote the optimal value by `fminunc`. For each s_0 , we report the final iterate $\hat{s}$, and the corresponding value. On the other hand, at the minimizer $s^* = (1.1709, -1.1709, 0, 0, 0)$, the

TABLE 1. Convergence of `fminunc` for different initial points in Example 6.1.

s_0	$\hat{s}$	f^*
(0, 0, 0, 0, 0)	(0, 0, 0, 0, 0)	0.0000
(1, 1, 1, 1, 1)	(0, 0, 0, 0, 0)	0.0000
(10, 10, 10, 10, 10)	(0.4501, 0.4501, 0.4501, 0.4501, 0.4501)	-0.4999
(-10, -10, -10, -10, -10)	(-0.4501, -0.4501, -0.4501, -0.4501, -0.4501)	-0.4999
(100, 100, 100, 100, 100)	(1.0850, 1.0850, 1.0850, 1.0850, 1.0850)	17.4262

objective value $M(s^*) \approx -5.2479$. As shown in Table 1, `fminunc` can only get a critical point for some initial points.

Example 6.2. Consider the CQR problem

$$\min_{s \in \mathbb{R}^3} \|s\|^4 - 10\|s\|^3 - \left(\frac{5}{2}s_1^2 + 2s_2^2 + 3s_3^2\right) + s_1s_2 + 2s_2s_3 + \sum_{i=1}^3 is_i.$$

By solving (3.3) and (3.8), we get $\text{rank } Y^* = \text{rank } Z_1^* = \text{rank } Z_2^* = 1$. It took around 0.01 second. By Proposition 3.3, we get the minimizer

$$s^* = ((Y^*)_{10}, (Y^*)_{20}, (Y^*)_{30}) \approx (-1.8131, 3.6458, -6.5873).$$

By Algorithm 5.1, we get $z^* \approx 7.7441$ and the matrix $R \in \mathbb{R}^{3 \times 4}$ has rank 3. It confirms s^* is the unique minimizer. The optimal values are

$$\gamma^* = \vartheta^* = \mu^* \approx -1281.5926.$$

The relaxations (3.3) and (3.8) are tight. The absolute and relative errors are

$$\mathbf{err-abs} = 2.12 \cdot 10^{-6}, \quad \mathbf{err-rel} = 1.65 \cdot 10^{-9}.$$

We compare our method with some classical nonlinear optimization methods, e.g., the function `fminunc` in MATLAB with the default setting of parameters. The convergence of `fminunc` highly depends on initial point. In our experiments, we used several initial points for `fminunc`. Let s_0 denote the initial point and f^* denote the optimal value by `fminunc`. For each s_0 , we report the final iterate $\hat{s}$, and the corresponding value. On the other hand, at the minimizer $s^* = (-1.8131, 3.6458, -6.5873)$, the objective value $M(s^*) \approx -1281.5926$. As shown in Table 2, `fminunc` can only get a critical point for some initial points.

TABLE 2. Convergence of `fminunc` for different initial points in Example 6.2.

s_0	$\hat{s}$	f^*
(0, 0, 0)	(-1.8131, 3.6458, -6.5873)	-1281.5926
(1, 1, 1)	(1.7782, -4.8411, 5.7600)	-1257.7205
(-0.5, 0, 0.5)	(1.7782, -4.8411, 5.7601)	-1257.7205
(10, 10, 10)	(1.7780, -4.8412, 5.7599)	-1257.7205

Example 6.3. Consider the CQR problem

$$\min_{s \in \mathbb{R}^6} \|s\|^4 - 5\|s\|^3 - 2s_1^2 - s_3^2 - \frac{3}{2}s_4^2 - \frac{1}{2}s_5^2 - 2s_1s_2 - s_1s_3 - 3s_1s_4 + s_2s_3 - 2s_2s_4 - s_2s_5 - 2s_3s_5 - s_4s_5 + 2\sum_{i=1}^5 s_i.$$

By solving (3.3) and (3.8), we get $\text{rank } Y^* = \text{rank } Z_1^* = \text{rank } Z_2^* = 1$. It took around 0.03 second. By Proposition 3.3, we get the minimizer

$$s^* = ((Y^*)_{10}, \dots, (Y^*)_{50}) \approx (-2.8277, -1.4802, -0.7917, -2.5252, -0.9839).$$

By Algorithm 5.1, we get $z^* \approx 4.2612$ and the matrix $R \in \mathbb{R}^{5 \times 6}$ has rank 5. This confirms s^* is the unique minimizer. The optimal values are

$$\gamma^* = \vartheta^* = \mu^* \approx -144.8805.$$

The relaxations (3.3) and (3.8) are tight. The absolute and relative errors are

$$\text{err-abs} = 3.25 \cdot 10^{-7}, \quad \text{err-rel} = 2.24 \cdot 10^{-9}.$$

We compare our method with some classical nonlinear optimization methods, e.g., the function `fminunc` in `MATLAB` with the default setting of parameters. The convergence of `fminunc` highly depends on initial point. In our experiments, we choose different initial points for `fminunc`. Let s_0 denote the initial point and f^* denote the optimal value by `fminunc`. For each s_0 , we report the final iterate $\hat{s}$, and the corresponding value. On the other hand, at the minimizer s^* , the objective value

TABLE 3. Convergence of `fminunc` for different initial points in Example 6.3.

s_0	$\hat{s}$	f^*
(0, 0, 0, 0, 0)	(-2.8277, -1.4802, -0.7917, -2.5252, -0.9839)	-144.8805
(1, 1, 1, 1, 1)	(2.9580, 1.3600, 0.0610, 2.5694, 0.3495)	-112.6193
(1, 0, 0, 0, 0)	(2.9580, 1.3600, 0.0610, 2.5694, 0.3495)	-112.6193
(0.3, 0.1, -0.7, 0.8, 0.6)	(2.9580, 1.3600, 0.0610, 2.5694, 0.3495)	-112.6193
(10, 10, 10, 10, 10)	(2.9573, 1.3592, 0.0621, 2.5709, 0.3478)	-112.6193

$M(s^*) \approx -144.8805$. As shown in Table 3, `fminunc` can only get a critical point for some initial points.

In the following, we explore the computational performance of the SDP relaxations (3.3)-(3.8). Up to the linear transformation $s = (\frac{4}{\sigma})^{1/4} \tilde{s}$, the CQR problem can be transformed to a new one with $\sigma = 4$, i.e., the coefficient of $\|s\|^4$ is one.

Example 6.4. Consider some randomly generated CQR problem in the form

$$(6.1) \quad \min_{s \in \mathbb{R}^n} \|s\|^4 + \frac{\beta}{6} \|s\|^3 + \frac{1}{2} s^T H s + g^T s,$$

where $\sigma = 4$. The vector g and matrix H are generated as in **MATLAB** as in [8]:

$$g = \text{randn}(n, 1), \quad H_1 = \text{randn}(n, n), \quad H = (H_1 + H_1^T)/2.$$

For each n and β , we generate 20 instances. For each instance, we solve the CQR by SDP relaxations (3.3)-(3.8). For all instances, the CQR problems are solved successfully by (3.3)-(3.8). The accuracy error **err-abs** is around 10^{-9} for all instances. We report the average computational time (in seconds) for each case of (n, β) . The numerical performance is reported in Table 4. As we can see, these CQR problems can be solved efficiently by the relaxations. For instance, when the dimension $n = 1000$, the relaxations can be solved within around one minute.

TABLE 4. Computational time for solving (6.1) by (3.3) and (3.8).

$\beta \backslash n$	10	1	0	-1	-10	-100
100	0.11	0.12	0.11	0.12	0.11	0.24
200	0.34	0.30	0.30	0.30	0.45	0.97
300	1.72	1.76	1.74	1.77	1.63	3.49
400	5.01	5.84	5.79	5.96	5.20	11.30
500	6.15	6.19	6.27	6.21	6.45	21.82
600	11.60	11.97	11.54	11.30	11.30	35.78
700	18.72	18.43	18.07	18.82	18.84	43.14
800	23.64	23.98	22.91	23.85	23.36	79.03
900	37.09	35.87	36.54	34.78	35.54	67.53
1000	52.92	52.74	53.46	54.15	54.33	60.93

7. CONCLUSIONS

In this paper, we give an SDP relaxation method for finding global minimizers of the cubic-quartic regularization problem. This method transforms the original non-convex optimization problem into a tractable semidefinite program. We prove that the relaxations are tight when the regularization parameters satisfy

$$\|s^*\|(\beta + 3\sigma\|s^*\|) \geq 0,$$

where s^* is a global minimizer of the CQR problem. This condition may be used as guidelines for parameter selections in numerical optimization methods. Moreover, we show that if the relaxations are tight, then all nonzero minimizers of the CQR problem have the same length. Algorithm 5.1 is given to obtain the set of all global minimizers, when the SDP relaxations are tight. Numerical experiments demonstrate that our SDP relaxation method is efficient for solving CQR problems.

Acknowledgement Jiawang Nie is partially supported by the NSF grant DMS-2513254.

REFERENCES

- [1] M. ApS.: Mosek optimization toolbox for matlab: User’s Guide and Reference Manual. Version 4.1, (2019)
- [2] Cartis, C., Gould, N.I., Toint, P.L.: Adaptive cubic regularisation methods for unconstrained optimization. part I: motivation, convergence and numerical results. *Math. Program.* 127(2), 245–295, (2011)
- [3] Cartis, C., Gould, N.I., Toint, P.L.: Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function and derivative-evaluation complexity. *Math. Program.* 130(2), 295–319, (2011)
- [4] Cartis, C., Gould, N.I., Toint, P.L.: On the complexity of steepest descent Newton’s and regularized Newton’s methods for nonconvex unconstrained optimization problems. *SIAM J. Optim.* 20(6), 2833–2852, (2010)
- [5] Cartis, C., Gould, N.I., Toint, P.L.: A concise second-order complexity analysis for unconstrained optimization using high-order regularized models. *Optim. Methods and Soft.* 35(2), 243–256, (2020)
- [6] Cartis, C., Gould, N.I., Toint, P.L.: Evaluation complexity of algorithms for nonconvex optimization. *MOS-SIAM Series on Optimization.* (2022)
- [7] Cartis, C., Hauser, R., Liu, Y., Welzel, K., Zhu, W.: Efficient implementation of third-order tensor methods with adaptive regularization for unconstrained optimization. *arXiv:2501.00404.* (2024)
- [8] Cartis, C., Zhu, W.: Cubic-quartic regularization models for solving polynomial subproblems in third-order tensor methods. *Math. Program.* 1–53, (2025)
- [9] Cartis, C., Zhu, W.: Second-order methods for quartically-regularised cubic polynomials, with applications to high-order tensor methods. *Math. Program.* 1–47, (2025)
- [10] Cartis, C., Zhu, W.: Global Convergence of High-Order Regularization Methods with Sums-of-Squares Taylor Models. *arXiv:2404.03035.* (2024)
- [11] Cartis, C., Zhu, W.: Global Optimality Characterizations and Algorithms for Minimizing Quartically-Regularized Third-Order Taylor Polynomials. *arXiv:2504.20259.* (2025)
- [12] Henrion, D., Korda, M., Lasserre, J.B.: *The Moment-sos Hierarchy.* World Scientific, Singapore, (2020)
- [13] Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. *Emerging applications of algebraic geometry*, pp. 157–270, (2009)
- [14] Lasserre, J.B.: *An introduction to polynomial and semi-algebraic optimization*, vol. 52. Cambridge University Press, Cambridge, (2015)
- [15] Levenberg, K.: A method for the solution of certain non-linear problems in least squares. *Appl. Math.* 2(2), 164–168, (1944)
- [16] Marquardt, D. W.: An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441, (1963)
- [17] Nesterov, Y.: Accelerating the cubic regularization of Newton’s method on convex problems. *Math. Program.* 112(1), 159–181, (2008)
- [18] Nesterov, Y.: Quartic regularity. *Vietnam J. Math.* 1–23, 2025.
- [19] Nesterov, Y., Polyak, B.T.: Cubic regularization of newton method and its global performance. *Math. Program.* 108(1), 177–205, (2006)

- [20] Nie, J., Demmel, J., Sturmfels, B.: Minimizing polynomials via sum of squares over the gradient ideal. *Math. Program.* 106(3), 587–606 (2006)
- [21] Nie, J.: Optimality conditions and finite convergence of Lasserre’s hierarchy. *Math. Program.* 146(1-2), 97–121 (2014)
- [22] Nie, J.: Tight relaxations for polynomial optimization and Lagrange multiplier expressions. *Math. Program.* 178 (1-2), 1–37 (2019)
- [23] Nie, J.: *Moment and Polynomial Optimization*, SIAM, 2023.
- [24] Todd, M.: *Semidefinite Optimization*. *Acta Numerica.* 10, 515–560, (2001)

JINLING ZHOU, SCHOOL OF MATHEMATICS AND COMPUTATIONAL SCIENCE, XIANGTAN UNIVERSITY, XIANGTAN, HUNAN, 411105, CHINA.

Email address: jinlingzhou@smail.xtu.edu.cn

XIN LIU, ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCE, CHINESE ACADEMY OF SCIENCES, 55 ZHONG GUAN CUN EAST ROAD, BEIJING 100190, CHINA

Email address: liuxin@lsec.cc.ac.cn

JIAWANG NIE, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA SAN DIEGO, 9500 GILMAN DRIVE, LA JOLLA, CA, USA, 92093.

Email address: njw@math.ucsd.edu

XINDONG TANG, DEPARTMENT OF MATHEMATICS, HONG KONG BAPTIST UNIVERSITY, KOWLOON TONG, KOWLOON, HONG KONG.

Email address: xdtang@hkbu.edu.hk