ON THE LOCAL STABILITY OF SEMIDEFINITE RELAXATIONS

DIEGO CIFUENTES, SAMEER AGARWAL, PABLO A. PARRILO, AND REKHA R. THOMAS

Abstract. In this paper we consider a parametric family of polynomial optimization problems over algebraic sets. Although these problems are typically nonconvex, tractable convex relaxations via semidefinite programming (SDP) have been proposed. Often times in applications there is a natural value of the parameters for which the relaxation will solve the problem exactly. We study conditions (and quantitative bounds) under which the relaxation will continue to be exact as the parameter moves in a neighborhood of the original one. It suffices to restrict to quadratically constrained quadratic programs.

Our framework captures several estimation problems such as low rank approximation, camera triangulation, rotation synchronization, approximate matrix completion and approximate GCD. In these applications, a solution is easy under noiseless observations, and our results guarantee that the SDP relaxation will continue to solve the problem in the low noise regime.

1. Introduction

Polynomial optimization problems arise in many different applications across the sciences and engineering. Solving polynomial optimization problems is computationally hard, and semidefinite programming (SDP) relaxations have become a standard approach to tackle them. In particular, a hierarchy of SDP relaxations can be constructed based on the *sum-of-squares* (SOS) method [6,22]. We focus here on optimization problems over algebraic sets that involve *parameters*, and we assume that for a given value of the parameters the SDP relaxation is *tight*, i.e., it correctly solves the problem. We study the behavior of the relaxation under *small perturbations* of the special parameters, identifying sufficient conditions under which the relaxation continues to be tight.

An important class of problems we consider is that of finding the point y^* on an algebraic variety $Y \subseteq \mathbb{R}^n$ that minimizes a loss function, e.g., the Euclidean distance $\|y-\theta\|$ to some given θ. These problems arise often in statistical estimation problems, such as low rank approximation, camera triangulation, rotation synchronization, approximate matrix completion and approximate GCD. In these problems the case where $\theta \in Y$ is trivial, and SDP relaxations are tight. Our methods allow a systematic analysis of the behavior of these relaxations when θ is close to Y. In particular, we recover tightness results previously shown in the context of camera triangulation [1] and rotation synchronization [31].

Since polynomial optimization problems can always be rewritten as *quadratically constrained quadratic programs* (QCQP's) (see e.g., [26, 33]), in the rest of this paper

Date: October 11, 2017.

Key words and phrases. Parametric SDP, Stability, Sum of squares, Algebraic variety.

we will only consider QCQP's. As illustrated in the next example, there is a natural SDP relaxation of a QCQP; namely, its Lagrangian dual. Moreover, SDP relaxations coming from the SOS method correspond to the Lagrangian dual of a suitable QCQP.

Example 1.1 (Nearest point to the twisted cubic)**.** Let $Y := \{(t, t^2, t^3) : t \in \mathbb{R}\}$ be the twisted cubic curve in $\mathbb{R}^3$. Given $\theta \in \mathbb{R}^3$, the problem of finding the nearest point in Y to θ can be phrased as:

$$\min_{y\in Y} \quad \|y-\theta\|^2, \qquad \text{where} \qquad Y = \{y \in \mathbb{R}^3 : y_2 = y_1^2,\ y_3 = y_1 y_2\}. \tag{1}$$

The above is a QCQP, and its Lagrangian dual is the following SDP:

$$\max_{\gamma,\lambda_1,\lambda_2\in\mathbb{R}} \quad \gamma, \qquad \text{s.t.} \qquad \begin{pmatrix} \gamma+\|\theta\|^2 & -\theta_1 & \lambda_1-\theta_2 & \lambda_2-\theta_3 \\ -\theta_1 & 1-2\lambda_1 & -\lambda_2 & 0 \\ \lambda_1-\theta_2 & -\lambda_2 & 1 & 0 \\ \lambda_2-\theta_3 & 0 & 0 & 1 \end{pmatrix} \succeq 0. \tag{2}$$

We will show that when θ is sufficiently close to Y the above relaxation is tight. Equivalently, the *duality-gap* val (1) − val (2) is zero in a neighborhood of Y. This is illustrated in Figure 1, by showing the projection of Y onto the y_1y_3-plane, and the duality-gap for parameters θ of the form $(\theta_1, \theta_1^2, \theta_3)$. Besides the fact that there is zero-duality-gap when θ is close to Y, we will also see that we can *recover the minimizer* of (1) from the SDP.

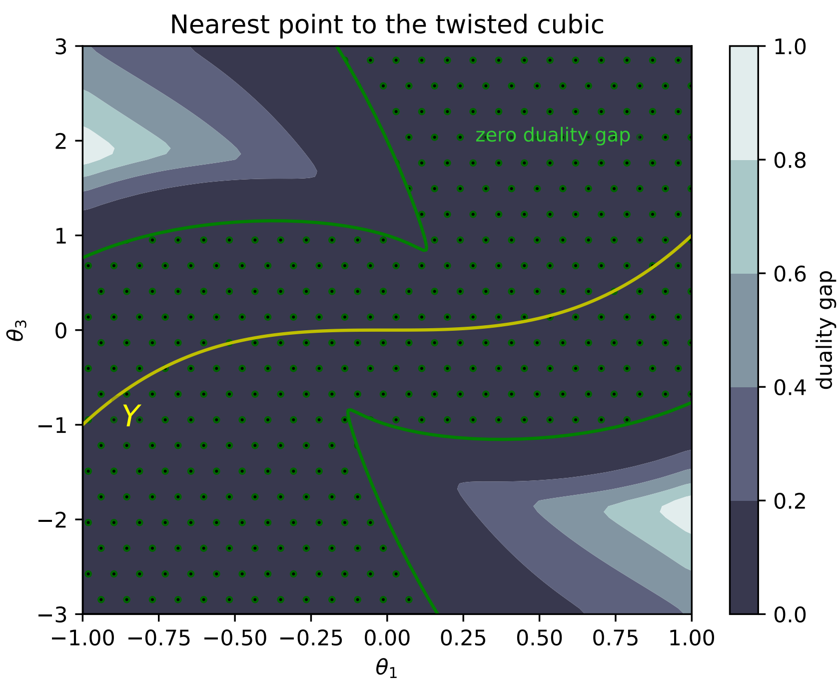

FIGURE 1. Duality gap in problem (1) for parameters θ of the form $(\theta_1, \theta_1^2, \theta_3)$. There is no duality-gap in the dotted region.

We shall say that a variety Y is *quadratic* if it is the zero set of some quadratic equations $f_i \in \mathbb{R}[y]$ (e.g., the twisted cubic (1)). As stated next, the problem of finding the nearest point from θ to a quadratic variety Y has a tight SDP relaxation when θ is close to Y. More precisely, we show that if $\bar{\theta} \in Y$ satisfies a certain regularity condition, then the relaxation is tight when θ deviates slightly from $\bar{\theta}$.

Theorem 1 (Nearest point to a quadratic variety). *Consider the problem*

$$\min_{y \in Y} \quad \|y - \theta\|^2, \qquad \textit{where} \qquad Y := \{y \in \mathbb{R}^n : f_1(y) = \cdots = f_m(y) = 0\}, \quad f_i \textit{ quadratic.}$$

Let $\bar{\theta} \in Y$ be such that $\operatorname{rank}(\nabla f(\bar{\theta})) = n - \dim_{\bar{\theta}} Y$. *Then there is zero-duality-gap for any $\theta \in \mathbb{R}^n$ that is sufficiently close to $\bar{\theta}$.*

Theorem 1 is a special case of a more general result, Theorem 8, that we will prove in Section 4. Both of these theorems rely on the objective function being strictly convex. These results can be readily applied to many estimation problems, such as rank one tensor approximation, the triangulation problem, and rotation synchronization (see Section 7.1). The proof is relatively elementary.

Unfortunately, having strictly convex objective is a strong requirement that is often not satisfied. In particular, the nearest point problem to a cubic curve cannot be written as a QCQP with a strictly convex objective.

Example 1.2. Consider the nearest point problem to the plane curve $Y := \{y \in \mathbb{R}^2 : y_2^2 = y_1^3\}$. This curve is not defined by quadratic equations in (y_1, y_2). But by introducing the auxiliary variable z_1 we can rewrite the nearest point problem as a QCQP:

$$\min_{z_1 \in \mathbb{R},\, y \in \mathbb{R}^2} \quad \|y - \theta\|^2, \qquad \text{s.t.} \qquad y_2 = y_1 z_1, \quad y_1 = z_1^2, \quad y_2 z_1 = y_1^2. \tag{3}$$

Note that the presence of the "auxiliary" variable z_1 means that the objective function is not strictly convex. Nonetheless, we will see that the resulting QCQP has zero-duality-gap when θ is close enough to the curve Y.

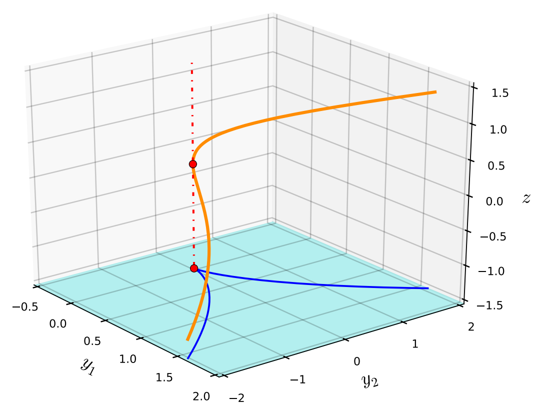

FIGURE 2. The plane curve $y_2^2 = y_1^3$ is the projection of the twisted cubic.

In this paper we consider a general family of QCQP's parametrized by θ, and we assume that there is zero-duality-gap for a fixed parameter $\bar{\theta}$. The main contribution of this paper is to identify sufficient conditions that guarantee zero-duality-gap as $\theta \to \bar{\theta}$. Our results, in particular, can be used in nearest point problems to varieties, such as the plane curve of Example 1.2. Note that the conditions we require are nontrivial, and it is easy to find problems with positive-duality-gap for θ arbitrarily close to $\bar{\theta}$ (see Example 5.3).

The organization of this paper is as follows. In Section 2 we formalize our problem of study. In Section 3 we introduce the main tools needed for our results. In Section 4 we prove Theorem 8, a generalization of Theorem 1. In Section 5 we state Theorem 14, the most general result of this paper. In Section 6 we prove Theorem 14 using the implicit function theorem. In Section 7 we show some applications of our results.

Related work. SDP relaxations of polynomial optimization problems have attracted major research in recent years. In particular, the Lagrangian dual of a QCQP has found numerous applications [6, 17, 25]. More generally, the SOS/moments method is a systematic approach to derive SDP relaxations of polynomial optimization problems; we refer to [6, 22] for an overview.

Several results concerning tightness of SDP relaxations exist. There are two main kinds of results. The first class is based on structural assumptions on the equations (or the feasible set). In particular, the Lagrangian dual of several classes of QCQP's are exact [4, 21, 37, 38], such as trust-region problems (there is a single ball constraint) [34], and S-lemma type problems [29]. Similarly, certain classes of combinatorial optimization problems are known to have tight SDP relaxations [18, 24]. We point out that there has been plenty of research in this area, and our references are far from extensive.

A second class of results, typically tailored to statistical estimation problems, shows tightness of SDP relaxations under low noise assumptions. In particular, it was shown in [1] that the first SOS relaxation of the triangulation problem is tight in the low noise regime under genericity assumptions. Similarly, it has been shown that the first SOS relaxation of the rotation synchronization problem is tight under low noise [16, 31, 36]. We generalize these types of results by systematically studying the behavior of SDP relaxations under small perturbations of the problem.

Perturbation analysis of nonlinear optimization problems is a well studied subject [7, 14, 23]. In particular, sufficient conditions for continuity and differentiability of the optimal value/solutions are known [7, 14]. Similarly, the Lipschizian stability of general optimization problems, together with concepts such as tilt/full stability, has received a lot of attention [23, 27]. The stability analysis of special classes of nonlinear programs, such as semidefinite programs [7, §5.3.6] and convex quadratic programs [19], has also been considered. In this paper we apply these techniques to the study of SDP relaxations of equality constrained QCQP's. As opposed to previous work we are mostly interested in the correctness of the SDP relaxation, i.e., whether we can recover an optimal solution of the QCQP from the relaxation.

Acknowledgements. We would like to thank Dmitriy Drusvyatskiy for many helpful conversations, and for suggesting the use of the implicit function theorem to analyze our problem.

2. Formalizing the problem

In this section we formalize our problem of study. We also begin to explain the role of Lagrange multipliers in certifying zero-duality-gap.

2.1. **Problem statement.** Consider a family of QCQP's parametrized by $\theta \in \Theta$:

$$(P_\theta) \qquad \begin{aligned} \min_{x\in\mathbb{R}^N} \quad & g_\theta(x) \\ & h_\theta^i(x) = 0, \quad \text{for } i = 1,\dots,m \end{aligned}$$

where g_θ, h_θ^i are *quadratic* in x, and the dependence on θ is *continuously differentiable*. Note that g_θ may not be convex. To simplify the notation, we will focus in this paper on the *homogeneous* case. By that we mean that g_θ, h_θ^i do not have linear terms, i.e.,

$$g_\theta(x) := x^T G_\theta x, \quad h_\theta^i(x) = x^T H_\theta^i x + b_i \quad \text{ for some } \quad G_\theta, H_\theta^i \in \mathbb{S}^N, b_i \in \mathbb{R},$$

and that at least one b_i is nonzero. In this paper $\mathbb{S}^N$ denotes the space of $N \times N$ symmetric matrices.

Remark 1 (Sign invariance)*.* Since a homogeneous problem is invariant under the involution $x \mapsto -x$, we can only recover the solution up to sign changes.

Remark 2 (Homogenization)*.* We can always get rid of linear terms by introducing a *homogenizing* variable z_0. For instance, the homogenized form of (1) is:

$$\min_{z_0\in\mathbb{R},\, y\in\mathbb{R}^3} \quad \|y-\theta z_0\|^2 \quad \text{s.t.} \quad z_0^2 = 1, \quad y_2 z_0 = y_1^2, \quad y_3 z_0 = y_1 y_2.$$

Recall that a QCQP has an associated dual pair of SDP relaxations. In the case of (P_θ) we get:

$$(P_\theta^*) \qquad \begin{aligned} \min_{S\in\mathbb{S}^N} \; & G_\theta \bullet S \\ & H_\theta^i \bullet S + b_i = 0, \quad i = 1,\dots,m \\ & S \succeq 0 \end{aligned} \qquad\qquad (D_\theta) \qquad \begin{aligned} \max_{\lambda\in\mathbb{R}^m} \; & d(\lambda) := \textstyle\sum_i \lambda_i b_i \\ & \mathcal{Q}_\theta(\lambda) \succeq 0 \end{aligned}$$

where $\bullet$ denotes the trace inner product, and $\mathcal{Q}_\theta(\lambda)$ is the Hessian of the Lagrangian function:

$$\mathcal{Q}_\theta(\lambda) := G_\theta + \sum_i \lambda_i H_\theta^i \in \mathbb{S}^N.$$

Note that (D_θ) is the Lagrangian dual of (P_θ), and that the inequalities $\operatorname{val}(P_\theta) \geq \operatorname{val}(P_\theta^*) \geq \operatorname{val}(D_\theta)$ always hold. We are concerned here with the *zero-duality-gap* condition $\operatorname{val}(P_\theta) = \operatorname{val}(D_\theta)$.

Throughout this paper we denote by $\bar\theta$ the nominal value of the parameter θ, such that $(P_{\bar\theta})$ has zero-duality-gap, and we investigate the behavior of the relaxation under small perturbations around $\bar\theta$.

2.2. **Lagrange multipliers and zero-duality-gap.** Given a feasible solution x_θ of (P_θ), recall that $\lambda \in \mathbb{R}^m$ is a *Lagrange multiplier* at x_θ if

$$\lambda^T \nabla h_\theta(x_\theta) = -\nabla g_\theta(x_\theta) \iff \sum_i \lambda_i H_\theta^i x_\theta = -G_\theta x_\theta \iff \mathcal{Q}_\theta(\lambda) x_\theta = 0.$$

We denote by $\Lambda_\theta(x_\theta)$ the affine space of *Lagrange multipliers* at x_θ.

Lemma 2. *Let $x_\theta \in \mathbb{R}^N, \lambda \in \mathbb{R}^m$. Then x_θ is optimal to (P_θ) and λ is optimal to (D_θ) with* $\operatorname{val}(P_\theta) = \operatorname{val}(D_\theta)$ *if and only if the following conditions hold:*

2(i) $h_\theta(x_\theta) = 0$ *(primal feasibility).*
2(ii) $\mathcal{Q}_\theta(\lambda) \succeq 0$ *(dual feasibility).*
2(iii) $\lambda \in \Lambda_\theta(x_\theta)$ *(complementarity).*

If furthermore $\mathcal{Q}_\theta(\lambda)$ has corank-one, then the unique optimal solution of (P^*_θ) is $x_\theta x_\theta^T$.

Proof. Assume that the conditions 2(i-iii) hold. Since $\mathcal{Q}_\theta(\lambda)x_\theta = 0$, then

$$0 = x_\theta^T \mathcal{Q}_\theta(\lambda) x_\theta = x_\theta^T G_\theta x_\theta + \sum_i \lambda_i \, x_\theta^T H_\theta^i x_\theta = g_\theta(x_\theta) - d(\lambda), \tag{4}$$

and thus (x_θ, λ) are primal/dual optimal. The converse is similar.

Assume now that $\mathcal{Q}_\theta(\lambda)$ has corank-one. Let S^* be an optimal solution of (P^*_θ). By complementary slackness we have $\mathcal{Q}_\theta(\lambda) \bullet S^* = 0$, and since both matrices lie in $\mathbb{S}^N_+$, then $\operatorname{rank}(\mathcal{Q}_\theta(\lambda)) + \operatorname{rank}(S^*) \leq N$. Thus, any optimal solution of (P^*_θ) has rank one. It follows that there is a unique optimal, namely $S^* = x_\theta x_\theta^T$. □

The above lemma gives necessary and sufficient conditions for having zero-duality-gap. Moreover, if an additional assumption holds (corank-one Hessian), then we can also recover the minimizer of the QCQP from the SDP. We point out that items 2(i-iii) correspond to the optimality conditions of a nonlinear program, and that variants of Lemma 2 have appeared before (e.g., [1, Thm 2]).

As a simple application of Lemma 2, let us see that the nearest point problem to a variety Y has zero-duality-gap for any $\bar\theta \in Y$ (in fact, the associated multiplier is $\bar\lambda = 0$).

Proposition 3 (Nearest point problem)**.** *Consider the problem:*

$$\min_{y \in Y} \|y - \theta\|^2, \quad \textit{where} \quad Y := \{y \in \mathbb{R}^n : \exists z' \in \mathbb{R}^{k-1} \textit{ s.t. } f_i(z', y) = 0, 1 \leq i \leq m\}$$

where f_i are quadratic polynomials, and z' is a set of auxiliary variables. There is zero-duality-gap for any $\bar\theta \in Y$.

Proof. In order to apply Lemma 2, we first need to consider the homogeneous version of the problem:

$$\min_{z \in \mathbb{R}^k,\, y \in \mathbb{R}^n} \quad \|y - \theta z_0\|^2 \quad \text{s.t.} \quad z_0^2 = 1, \quad h_i(z, y) = 0, \quad 1 \leq i \leq m,$$

where $z = (z_0, z')$, and h_i is the homogenization of f_i with respect to z_0:

$$h_i(z, y) = h_i(z_0, z', y) := z_0^2 \; f_i(z'/z_0, y/z_0).$$

Let $\bar x = (\bar z, \bar\theta)$ be the optimal solution, and let $g_{\bar\theta}(z, y) := \|y - \theta z_0\|^2$ be the objective function. Notice that $\bar\lambda = 0$ is a Lagrange multiplier at $\bar x$ since $\nabla g_{\bar\theta}(\bar x) = 0$. Moreover, $\bar\lambda = 0$ is dual feasible since $g_{\bar\theta}$ is convex and thus $\mathcal{Q}_\theta(\bar\lambda) = G_{\bar\theta} \succeq 0$. Then there is zero-duality-gap by Lemma 2. □

3. Continuity of Lagrange multipliers

In Lemma 2 we identified necessary and sufficient conditions for zero-duality-gap at a fixed parameter $\bar\theta$. In this section we will study conditions under which we continue to get zero-duality-gap for parameters θ close to $\bar\theta$.

Throughout this section we let $\bar\theta$ be a zero-duality gap parameter, and $\bar x$ be the minimizer of $(P_{\bar\theta})$, which we assume is *unique* (up to sign). Observe that $\bar x \neq 0$ since we assumed that the constant term b_i of some $h^i_\theta(x)$ is nonzero. Consider the *Lagrange multiplier mapping*:

$$\mathfrak{L} : \Theta \rightrightarrows \mathbb{R}^N \times \mathbb{R}^m, \qquad \begin{aligned} \theta \mapsto &\{(x_\theta, \lambda_\theta) : x_\theta \text{ primal feasible, } \lambda_\theta \in \Lambda_\theta(x_\theta)\} \\ &= \{(x_\theta, \lambda_\theta) : h_\theta(x_\theta) = 0,\ \mathcal{Q}_\theta(\lambda_\theta) x_\theta = 0\}. \end{aligned} \tag{5}$$

As we will see, *continuity* properties of $\mathfrak{L}$ play a crucial role in our analysis.

3.1. **A first stability result.** Our first stability result relies on the existence of a dual optimal solution λ satisfying the following two properties (recall that $\bar x$ is fixed).

Assumption C1H (corank-one Hessian)**.** Let λ be dual optimal at $\bar\theta$. The matrix $\mathcal{Q}_{\bar\theta}(\lambda)$ has corank-one, or equivalently, strict complementarity holds for the dual pair of SDP's.

Assumption WC (weak continuity of multipliers)**.** Let $\bar\ell = (\bar x, \lambda) \in \mathfrak{L}(\bar\theta)$ be a Lagrange multiplier pair. There exists $\ell_\theta \in \mathfrak{L}(\theta)$ such that $\ell_\theta \to \bar\ell$ as $\theta \to \bar\theta$.

Proposition 4 (C1H + WC $\implies$ stability)**.** *Let $(\bar x, \lambda)$ be primal/dual optimal at $\bar\theta$, such that Assumptions C1H and WC hold. Then for any θ sufficiently close to $\bar\theta$ there is zero-duality-gap and (P^*_θ) recovers the minimizer.*

Proof. By Assumption WC, there exist $(x_\theta, \lambda_\theta)$ with: x_θ feasible for (P_θ), $\lambda_\theta \in \Lambda_\theta(x_\theta)$, and $(x_\theta, \lambda_\theta) \to (\bar x, \lambda)$ as $\theta \to \bar\theta$. It follows that $\mathcal{Q}_\theta(\lambda_\theta) \to \mathcal{Q}_{\bar\theta}(\lambda)$, since g_θ and h^i_θ depend continuously on θ. Notice that $\mathcal{Q}_\theta(\lambda_\theta)$ has a 0-eigenvalue since $\mathcal{Q}_\theta(\lambda_\theta)x_\theta = 0$. Let us see that, as $\theta \to \bar\theta$, the remaining eigenvalues of $\mathcal{Q}_\theta(\lambda_\theta)$ are positive. This would conclude the proof because of Lemma 2. By assumption, $\mathcal{Q}_{\bar\theta}(\lambda)$ has $N-1$ positive eigenvalues. Since $\mathcal{Q}_\theta(\lambda_\theta) \to \mathcal{Q}_{\bar\theta}(\lambda)$, and by the continuity of the eigenvalues, we conclude that $\mathcal{Q}_\theta(\lambda_\theta)$ also has $N-1$ positive eigenvalues when $\theta \to \bar\theta$, as wanted. □

Proposition 4 establishes conditions to guarantee zero-duality-gap nearby $\bar\theta$. We will see a few more stability results later in this paper, but all of them rely implicitly on Proposition 4. This is illustrated in Figure 3.

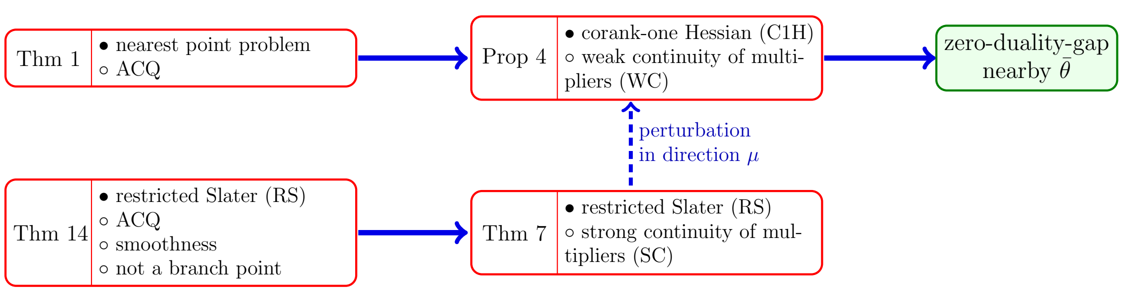

FIGURE 3. Conditions that imply zero-duality-gap nearby $\bar\theta$. The main assumption in each box is denoted with •, and regularity assumptions are marked with ○.

While it is easy to check whether a given matrix has corank-one, it is typically not easy to find a dual variable λ that satisfies Assumption C1H. This is similar to the second order optimality conditions in non-linear programming which is also stated in terms of the existence of a Lagrange multiplier that satisfies certain conditions, without any procedure to find such a multiplier

In practice there is often a natural choice $\bar{\lambda}$ for the dual variables, such as $\bar{\lambda} = 0$ in nearest point problems (Proposition 3). If Assumptions C1H and WC hold for such a nominal $\bar{\lambda}$, we will be certain that there is zero-duality-gap. Unfortunately, it is often the case that the Hessian matrix $\mathcal{Q}_{\bar{\theta}}(\bar{\lambda})$ has rank less than $N-1$, violating Assumption C1H. This situation arises, for instance, in nearest point problems to non-quadratic varieties. In the remainder of this section we will analyze how to establish stability in these cases.

3.2. **Restricted Slater.**

Assume we are given a dual variable $\bar{\lambda}$ for which Assumption C1H (corank-one Hessian) is not satisfied. The following "Slater-type" condition allows us to find a dual variable λ' for which Assumption C1H holds.

Assumption RS (restricted Slater)**.** Let $(\bar{x}, \bar{\lambda})$ be primal/dual optimal at $\bar{\theta}$, and consider the subspace

$$V := \{v \in \mathbb{R}^N : \mathcal{Q}_{\bar{\theta}}(\bar{\lambda})v = 0, \bar{x}^T v = 0\}.$$

There exists $\mu \in \mathbb{R}^m$ such that the quadratic function $\Psi_\mu(x) := \sum_i \mu_i h^i_{\bar{\theta}}(x)$ satisfies: $\nabla\Psi_\mu(\bar{x}) = 0$, and Ψ_μ is strictly convex on V.

Importantly, Assumption RS can be efficiently checked. Indeed, by restating the strict convexity of Ψ_μ in terms of its Hessian, Assumption RS corresponds to the strict feasibility of an SDP (find μ s.t. $\sum_i \mu_i H^i_{\bar{\theta}}\bar{x} = 0$, $(\sum_i \mu_i H^i_{\bar{\theta}})|_V \succ 0$). We will elaborate more on Assumption RS in Section 5, illustrating concrete examples that satisfy it.

As seen next, the vector μ from Assumption RS gives us a *direction* along which we can perturb the problem to go back to the corank-one case.

Lemma 5 (RS $\implies$ "nearby" C1H)**.** *Let $(\bar{x}, \bar{\lambda})$ be primal/dual optimal at $\bar{\theta}$, and let μ be as in Assumption RS. Then there is an $\epsilon > 0$ such that $\lambda_t := \bar{\lambda} + t\mu$ is dual optimal and* corank $\mathcal{Q}_{\bar{\theta}}(\lambda_t) = 1$ *for any $0 < t < \epsilon$.*

The proof of Lemma 5 relies on the following well-known lemma.

Lemma 6 (Finsler [15])**.** *Let $A, B \in \mathbb{S}^n, A \succeq 0$ be such that $v^T Bv > 0$ for every $v \neq 0$ with $Av = 0$. Then there is some $\epsilon > 0$ such that $A + tB \succ 0$ for any $0 < t < \epsilon$.*

Proof of Lemma 5. Since $(\bar{x}, \bar{\lambda})$ is primal/dual optimal, it satisfies conditions 2(i-iii). We need to show that $(\bar{x}, \lambda_t)$ also satisfies 2(i-iii), and that corank $\mathcal{Q}_{\bar{\theta}}(\lambda_t) = 1$. It is easy to see that $\lambda_t \in \Lambda_{\bar{\theta}}(\bar{x})$, so it remains to show that $\mathcal{Q}_{\bar{\theta}}(\lambda_t) \succeq 0$ and has corank-one. Let $A := \mathcal{Q}_{\bar{\theta}}(\bar{\lambda}) \succeq 0$ and $B := \sum_i \mu_i H^i_{\bar{\theta}}$. Since $\bar{\lambda} \in \Lambda_{\bar{\theta}}(\bar{x})$ then $A\bar{x} = 0$. Similarly, since $\sum_i \mu_i H^i_{\bar{\theta}}\bar{x} = 0$ then $B\bar{x} = 0$. We may assume WLOG that $\bar{x} = (1, 0^{N-1})$. Then $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & A' \end{smallmatrix}\right)$, $B = \left(\begin{smallmatrix} 0 & 0 \\ 0 & B' \end{smallmatrix}\right)$, where $A', B' \in \mathbb{S}^{N-1}$, $A' \succeq 0$. Note that the strict convexity condition in Assumption RS means that $v^T B' v > 0$ for every nonzero $v \in \mathbb{R}^{N-1}$ with $A'v = 0$. From Lemma 6 we know that $A' + tB' \succ 0$ for all $0 < t < \epsilon$. Therefore, $A + tB \succeq 0$ and has corank-one for $0 < t < \epsilon$, as wanted. □

3.3. **Stability under Assumption RS.** Lemma 5 allows us to find some new dual variables λ' for which Assumption C1H is satisfied. In order to use Proposition 4 and conclude stability, it remains to see that λ' satisfies Assumption WC (weak continuity). This requires a *stronger* continuity condition on the original pair $(\bar{x}, \bar{\lambda})$. Before stating this assumption, we first recall a well-studied notion of continuity for set-valued-mappings. We refer to [2, 30] for a detailed introduction to set-valued-mappings.

Definition 3.1 (Painlevé-Kuratowski continuity)**.** Let $\mathfrak{F} : \Theta \rightrightarrows \mathbb{R}^n$ be a set-valued mapping, and assume that each $\mathfrak{F}(\theta) \subseteq \mathbb{R}^n$ is nonempty. A *selection* of $\mathfrak{F}$ is an assignment $y_\theta \in \mathfrak{F}(\theta)$ for each $\theta \in \Theta$. The *inner limit* of $\mathfrak{F}$ at $\bar{\theta}$ consists of all limits of selections $\{y_\theta\}_\theta$, i.e.,

$$\liminf_{\theta \to \bar{\theta}} \mathfrak{F}(\theta) := \{y \in \mathbb{R}^n : \exists y_\theta \in \mathfrak{F}(\theta) \text{ s.t. } y_\theta \xrightarrow{\theta \to \bar{\theta}} y\},$$

The *outer limit* of $\mathfrak{F}$ at $\bar{\theta}$ consists of all cluster points of selections $\{y_\theta\}_\theta$, i.e.,

$$\limsup_{\theta \to \bar{\theta}} \mathfrak{F}(\theta) := \{y \in \mathbb{R}^n : \exists \theta_i \xrightarrow{i \to \infty} \bar{\theta},\ \exists y_i \in \mathfrak{F}(\theta_i) \text{ s.t. } y_i \xrightarrow{i \to \infty} y\}.$$

The inner and outer limits are always closed sets that sandwich the closure of $\mathfrak{F}(\bar{\theta})$:

$$\liminf_{\theta \to \bar{\theta}} \mathfrak{F}(\theta) \subseteq \operatorname{cl}(\mathfrak{F}(\bar{\theta})) \subseteq \limsup_{\theta \to \bar{\theta}} \mathfrak{F}(\theta).$$

$\mathfrak{F}$ is (Painlevé-Kuratowski) *continuous*[1] at $\bar{\theta}$ if $\mathfrak{F}(\bar{\theta}) = \liminf_{\theta \to \bar{\theta}} \mathfrak{F}(\theta) = \limsup_{\theta \to \bar{\theta}} \mathfrak{F}(\theta)$.

Remark 3. When $\mathfrak{F}$ is defined by continuous equations, such as $\mathfrak{L}$, then the equation $\mathfrak{F}(\bar{\theta}) = \limsup_{\theta \to \bar{\theta}} \mathfrak{F}(\theta)$ always holds [30, Ex 5.8]. Consequently, in this paper we will focus our attention only on the inner limit.

Remark 4. Note that Assumption WC is simply that $\bar{\ell} \in \liminf_{\theta \to \bar{\theta}} \mathfrak{L}(\theta)$.

Example 3.1. Consider the mapping

$$\mathfrak{F} : \mathbb{R} \rightrightarrows \mathbb{R}, \qquad \theta \mapsto \begin{cases} \{0\}, & \text{if } \theta < 0 \\ [-1, 1], & \text{if } \theta \geq 0 \end{cases}$$

This mapping is continuous at any $\theta \neq 0$. Observe that $\liminf_{\theta \to 0} \mathfrak{F}(\theta) = \{0\}$ and $\limsup_{\theta \to 0} \mathfrak{F}(\theta) = [-1, 1]$. Thus $\mathfrak{F}$ is not continuous at 0.

Assumption SC (strong continuity of multipliers)**.** Let $\bar{\ell} \in \mathfrak{L}(\bar{\theta})$ be a Lagrange multiplier pair. There exists a closed neighborhood $U \ni \bar{\ell}$ such that $\mathfrak{L}(\bar{\theta}) \cap U \subseteq \liminf_{\theta \to \bar{\theta}} \mathfrak{L}(\theta)$, or equivalently, such that the mapping $\theta \mapsto \mathfrak{L}(\theta) \cap U$ is continuous at $\bar{\theta}$.

Theorem 7 (RS + SC $\implies$ stability)**.** *Let $(\bar{x}, \bar{\lambda})$ be primal/dual optimal at $\bar{\theta}$, such that Assumptions RS and SC hold. Then for any θ sufficiently close to $\bar{\theta}$ there is zero-duality-gap and (P^*_θ) recovers the minimizer.*

[1] Although other notions of (set-valued-mapping) continuity exist, they agree for the case of compact valued mappings [30]. Since the analysis done in this paper is local, we may always restrict the range to some closed ball. Hence, we may ignore this distinction in this paper.

Proof. Let $U \ni \bar{\ell}$ be as in Assumption SC. By Lemma 5, there is a dual optimal λ' with $\mathcal{Q}_{\bar{\theta}}(\lambda')$ of corank-one, and moreover, λ' can be arbitrarily close to $\bar{\lambda}$. Thus, we may assume that $\ell' := (\bar{x}, \lambda') \in U$. Since ℓ' also belongs to $\mathfrak{L}(\bar{\theta})$, it in fact belongs to $\mathfrak{L}(\bar{\theta}) \cap U \subseteq \liminf_{\theta \to \bar{\theta}} \mathfrak{L}(\theta)$. Then ℓ' satisfies Assumptions C1H and WC, and the theorem follows from Proposition 4. □

Although Proposition 4 and Theorem 7 are easy to interpret (particularly Proposition 4), verifying the continuity assumptions on $\mathfrak{L}$ might be complicated. In the following sections we will find simpler regularity conditions that ensure these continuity properties (see Figure 3). In particular, in Section 4 we will see that, for the restricted case of Theorem 1, a simple constraint qualification (ACQ) suffices. For the general case (Theorem 14) we will need two additional regularity assumptions.

4. An easy first case

The purpose of this section is to prove a generalized version of Theorem 1, and to obtain bounds on the magnitude of the perturbations we can tolerate. Before presenting the theorem, we recall a well-studied constraint qualification (also known as quasiregularity [5]), that guarantees the existence of Lagrange multipliers (see e.g., [3, §5.1]).

Notation. This section works with problems in homogeneous and non-homogeneous (affine) form. We will distinguish them by using different notation. In particular, y denotes variables in affine coordinates, and x in homogeneous coordinates.

Definition 4.1. Given $f : \mathbb{R}^n \to \mathbb{R}^m$, let $Y := \{y \in \mathbb{R}^n : f(y) = 0\}$. The *Abadie constraint qualification* (ACQ) holds at $\bar{y} \in Y$, denoted $\mathrm{ACQ}_Y(\bar{y})$, if Y is a smooth manifold nearby $\bar{y}$, and $\mathrm{rank}(\nabla f(\bar{y})) = n - \dim_{\bar{y}} Y$. Here $\dim_{\bar{y}} Y$ denotes the local dimension of Y at $\bar{y}$.

Remark 5. Under ACQ, the Lagrange multiplier space has dimension $m-(n-\dim_{\bar{y}} Y)$.

We will prove the following theorem.

Theorem 8. *Consider the problem*

$$(6) \qquad \min_{y \in Y} \quad q_\theta(y), \qquad \text{with} \qquad Y := \{y \in \mathbb{R}^n : f_1(y) = \cdots = f_m(y) = 0\},$$

where q_θ, f_i are quadratic, and the dependence on θ is continuous. Let $\bar{\theta}$ be such that $q_{\bar{\theta}}$ is strictly convex, and its minimizer $\bar{y}$ satisfies $\nabla q_{\bar{\theta}}(\bar{y}) = 0$, or equivalently, $\bar{y}$ is the unconstrained minimizer of $q_{\bar{\theta}}(y)$. If $\mathrm{ACQ}_Y(\bar{y})$ holds, then there is zero-duality-gap whenever θ is close enough to $\bar{\theta}$. Moreover, (P_θ^*) recovers the minimizer of (6).

Remark 6. The nearest point problem to a quadratic variety (Theorem 1) corresponds to the case $q_\theta(y) := \|y-\theta\|^2$. Indeed, this objective is strictly convex, the minimizer is $\bar{y} = \bar{\theta}$ (since $\bar{\theta} \in Y$), and thus $\nabla q_{\bar{\theta}}(\bar{y}) = 0$. Theorem 1 generalizes the main result of [1], as will be discussed in Example 7.1.

4.1. **Preparing the problem.** The proof of Theorem 8 will rely on the tools we developed in Section 3; more precisely, on Proposition 4. We will now prepare problem (6) for applying these tools.

Since the equations q_θ, f_i in (6) may involve linear terms, we need to consider its homogenized form:

$$\min_{x\in X} \quad g_\theta(x), \qquad X := \{(z_0, y) \in \mathbb{R}^{n+1} : z_0^2 = 1,\ h_1(z_0, y) = \cdots = h_m(z_0, y) = 0\} \tag{7}$$

where g_θ, h_i are the homogenizations of q_θ, f_i. The Lagrange multiplier spaces of (6) and (7) are isomorphic. Concretely, for any feasible $x_\theta = (1, y_\theta)$ it can be shown that:

$$\Lambda_\theta(y_\theta) = \{\mu \in \mathbb{R}^m : \mu^T\nabla f(y_\theta) = -\nabla q_\theta(y_\theta)\}, \tag{8}$$

$$\Lambda_\theta(x_\theta) = \{\lambda = (\lambda_0, \mu) \in \mathbb{R}^{m+1} : \lambda_0 = -g_\theta(x_\theta),\ \mu \in \Lambda_\theta(y_\theta)\}. \tag{9}$$

Also consider the Hessians of the Lagrangian functions of (6) and (7):

$$\mathcal{C}_\theta(\mu) := \nabla^2 q_\theta + \sum_i \mu_i \nabla^2 f_i, \qquad \mathcal{Q}_\theta(\lambda_0, \mu) := \nabla^2 g_\theta + \lambda_0\nabla^2(z_0^2 - 1) + \sum_i \mu_i\nabla^2 h_i. \tag{10}$$

Observe that $\mathcal{Q}_\theta(\lambda_0, \mu) \in \mathbb{S}^{n+1}$ contains $\mathcal{C}_\theta(\mu) \in \mathbb{S}^n$ as a submatrix.

We can now specialize the conditions from Proposition 4 to problem (6).

Lemma 9. *Let $\bar{x} = (1, \bar{y})$ and $\bar{\lambda} = (-g_{\bar\theta}(\bar{x}), 0)$. Then $(\bar{x}, \bar{\lambda})$ is primal/dual optimal, $\bar\lambda$ satisfies Assumption C1H, and*

$$\textit{Assumption WC holds} \iff \exists y_\theta \in Y, \mu_\theta \in \Lambda_\theta(y_\theta) \ \textit{s.t.}\ (y_\theta, \mu_\theta) \xrightarrow{\theta\to\bar\theta} (\bar{y}, 0). \tag{11}$$

Proof. The equivalence in (11) follows from the isomorphism $\Lambda_\theta(y_\theta) \cong \Lambda_\theta(x_\theta)$. Recall that $(\bar{x}, \bar\lambda)$ is optimal if and only if conditions 2(i-iii) are satisfied. Since $\nabla q_{\bar\theta}(\bar y) = 0$ then $0 \in \Lambda_{\bar\theta}(\bar y)$, and thus $\bar\lambda \in \Lambda_{\bar\theta}(\bar x)$. It remains to show that $\mathcal{Q}_{\bar\theta}(\bar\lambda) \succeq 0$ and has corank-one. Observe that $\mathcal{C}_{\bar\theta}(0) = \nabla^2 q_{\bar\theta} \succ 0$ because $q_{\bar\theta}$ is strictly convex. Note that $\mathcal{Q}_{\bar\theta}(\bar\lambda)$ contains $\mathcal{C}_{\bar\theta}(0)$ as a submatrix, and the extra row/column is such that $\mathcal{Q}_{\bar\theta}(\bar\lambda)\bar x = 0$. It follows that $\mathcal{Q}_{\bar\theta}(\bar\lambda) \succeq 0$ and has corank-one. □

4.2. **Small multipliers.** From Lemma 9, what we need now is to show the existence of some (y_θ, μ_θ) such that y_θ approaches $\bar y$ and μ_θ approaches 0.

Lemma 10. *For each θ, let y_θ be an optimal solution of (6). Then $y_\theta \to \bar y$ as $\theta \to \bar\theta$.*

Proof. See Appendix A. □

It remains to find some small multipliers μ_θ associated to y_θ. The ACQ property allows us to do so.

Lemma 11. *Let y_θ be an optimal solution of (6). Let σ_θ be the s-th largest singular value of $\nabla f(y_\theta)$, where $s := \operatorname{codim}_{y_\theta} Y$, and assume that $\sigma_\theta > 0$ (i.e., $ACQ_Y(y_\theta)$ holds). Then there exists $\mu_\theta \in \Lambda_\theta(y_\theta)$ with $\|\mu_\theta\| \leq \frac{1}{\sigma_\theta}\|\nabla q_\theta(y_\theta)\|$.*

Proof. Recall that $\Lambda_\theta(y_\theta)$ is given by the linear equation in (8). This equation has a solution, since ACQ guarantees the existence of multipliers. Then $\mu_\theta^T := -\nabla q_\theta(y_\theta)J^\dagger$ is one such solution, where $J := \nabla f(y_\theta)$ is the Jacobian and † denotes the pseudo-inverse. The lemma follows by noticing that $\|J^\dagger\| = 1/\sigma_\theta$, since σ_θ is the smallest nonzero singular value of J. □

Proof of Theorem 8. By Proposition 4 and Lemma 9, it is enough to find y_θ, μ_θ as in (11). Let y_θ be an optimal solution of (6). By Lemma 10 we know that $y_\theta \to \bar{y}$ as $\theta \to \bar{\theta}$. Since ACQ is an open condition, it holds in a neighborhood of $\bar{y}$. By Lemma 11 there exists μ_θ with $\|\mu_\theta\| \leq \frac{1}{\sigma_\theta}\|\nabla q_\theta(y_\theta)\|$. By assumption $\nabla q_{\bar{\theta}}(\bar{y}) = 0$, and by ACQ we also have $\sigma_{\bar{\theta}} > 0$. It follows that $\mu_\theta \to 0$ as $\theta \to \bar{\theta}$, as wanted. □

4.3. **Guaranteed region of zero-duality-gap.** We proceed to estimate bounds on the magnitude of the perturbations we can tolerate. Consider the set of zero-duality-gap parameters:

$$\overline{\Theta} := \{\theta \in \Theta : \mathrm{val}(P_\theta) = \mathrm{val}(D_\theta), \text{ and } (P_\theta^*) \text{ recovers the minimizer}\}.$$

Our goal is to find a neighborhood of $\bar{\theta}$ that is entirely contained in $\overline{\Theta}$.

Example 4.1. Consider one more time the twisted cubic from Example 1.1. Figure 4 illustrates the region of zero-duality-gap guaranteed by the results of this section.

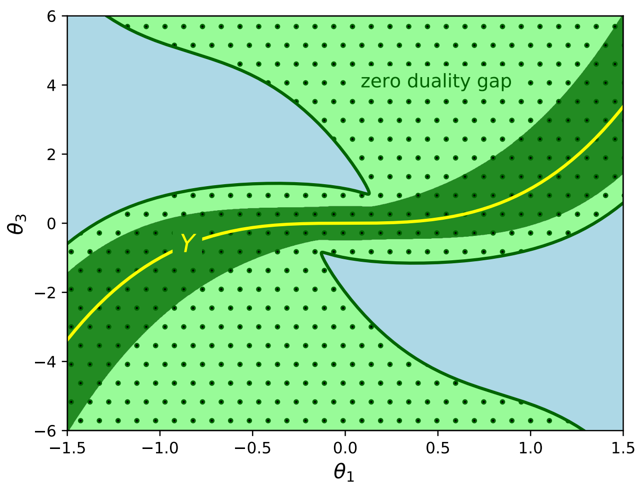

FIGURE 4. Region of zero-duality-gap from Figure 1. The darker region is the guaranteed region of zero-duality-gap (Corollary 13).

In order to find explicit bounds we need to quantify each of the assumptions we made in Proposition 4, and to impose some additional Lipschitz properties. We make the following assumptions:

12(i) (corank-one Hessian) The second smallest eigenvalue of $\mathcal{Q}_{\bar{\theta}}(\bar{\lambda})$, denoted $\nu_2(\bar{Q})$, is strictly positive.

12(ii) (weak continuity of $\mathfrak{L}$) There is a constant $K \geq 0$ such that

$$\forall \theta \in \Theta \;\exists (x_\theta, \lambda_\theta) \in \mathfrak{L}(\theta) \;\text{ s.t. }\; x_\theta \xrightarrow{\theta \to \bar{\theta}} \bar{x}, \;\; \|\lambda_\theta - \bar{\lambda}\| \leq K\|\theta - \bar{\theta}\|.$$

12(iii) (dependence on θ) There is a constant $L \geq 0$ such that

$$\|\mathcal{Q}_\theta(\bar{\lambda}) - \mathcal{Q}_{\bar{\theta}}(\bar{\lambda})\|_F \leq L\|\theta - \bar{\theta}\|$$

We will need one last assumption, which holds, in particular, when Θ is bounded.

12(iv) Consider the linear map $\mathcal{H}_\theta : \mathbb{R}^m \to \mathbb{S}^N$ that $\mu \mapsto \frac{1}{2}\sum_i \mu_i H^i_\theta$. There is some $M \geq 0$ that bounds the operator norm $\|\mathcal{H}_\theta\| \leq M$ for all $\theta \in \Theta$.

The following theorem is the quantitative version of Proposition 4.

Theorem 12. *Under Assumptions 12(i-iv),*

$$\left\{\theta \in \Theta : \|\theta - \bar\theta\| < \frac{\nu_2(\bar Q)}{KM + L}\right\} \subseteq \overline{\Theta}.$$

Proof. Let θ be such that $\|\theta - \bar\theta\| < \frac{\nu_2(\bar Q)}{KM+L}$. By Lemma 2, it is enough to show the existence of Lagrange multipliers λ_θ such that the second smallest eigenvalue of $\mathcal{Q}_\theta(\lambda_\theta)$ is positive (the first one is zero). By assumption there exists λ_θ such that $\|\lambda_\theta - \bar\lambda\| < K\|\theta - \bar\theta\|$. Note that

$$\|\mathcal{Q}_\theta(\lambda_\theta) - \mathcal{Q}_\theta(\bar\lambda)\|_F = \|\mathcal{H}_\theta(\lambda_\theta - \bar\lambda)\|_F \leq \|\mathcal{H}_\theta\|\,\|\lambda_\theta - \bar\lambda\| \leq KM\|\theta - \bar\theta\|,$$

and thus

$$\begin{aligned}\|\mathcal{Q}_\theta(\lambda_\theta) - \mathcal{Q}_{\bar\theta}(\bar\lambda)\|_F &\leq \|\mathcal{Q}_\theta(\lambda_\theta) - \mathcal{Q}_\theta(\bar\lambda)\|_F + \|\mathcal{Q}_\theta(\bar\lambda) - \mathcal{Q}_{\bar\theta}(\bar\lambda)\|_F \\ &\leq (KM + L)\|\theta - \bar\theta\| < \nu_2(\bar Q).\end{aligned}$$

By Weyl's inequality [11, Thm 5.1], we have

$$|\nu_2(\mathcal{Q}_\theta(\lambda_\theta)) - \nu_2(\mathcal{Q}_{\bar\theta}(\bar\lambda))| \leq \|\mathcal{Q}_\theta(\lambda_\theta) - \mathcal{Q}_{\bar\theta}(\bar\lambda)\| \leq \|\mathcal{Q}_\theta(\lambda_\theta) - \mathcal{Q}_{\bar\theta}(\bar\lambda)\|_F$$

and thus

$$\nu_2(\mathcal{Q}_\theta(\lambda_\theta)) \geq \nu_2(\mathcal{Q}_{\bar\theta}(\bar\lambda)) - \|\mathcal{Q}_\theta(\lambda_\theta) - \mathcal{Q}_{\bar\theta}(\bar\lambda)\|_F > \nu_2(\bar Q) - \nu_2(\bar Q) = 0,$$

as wanted. □

In the special case of the nearest point problem from Theorem 1 the above bounds can be made more explicit. The perturbation tolerance region shown in Figure 4 uses the bound in the following corollary.

Corollary 13. *Consider the (affine) setting of Theorem 1. For $\bar y \in Y$, let $\Theta(\bar y)$ consist of all θ for which $\bar y$ is the nearest point in Y, i.e., $\|\theta - \bar y\| = \operatorname{dist}(\theta, Y)$. Then*

$$\left\{\theta \in \Theta(\bar y) : \|\theta - \bar\theta\| < \frac{\sigma_s}{2M}\right\} \subseteq \overline{\Theta}$$

where:

- *σ_s is the s-th largest singular value of $\nabla f(\bar y)$, where $s := \operatorname{codim}_{\bar y} Y$.*
- *M is the operator norm of the linear map $\mathcal{H} : \mathbb{R}^m \to \mathbb{S}^n$ that $\mu \mapsto \frac{1}{2}\sum_i \mu_i \nabla^2 f_i$.*

Proof. Follows by noticing that $\nu_2(\bar Q) = 1$, $K = \frac{2}{\sigma_s}$, $L = 0$. See Appendix A. □

5. The general case

In Section 4 we observed that for nearest point problems to quadratic varieties a simple regularity condition (ACQ) guaranteed zero-duality-gap nearby $\bar\theta$. The purpose of this section is to identify regularity conditions that work for arbitrary QCQP's. Throughout this section we consider problem (P_θ) and use the following notation:

- the parameter space Θ is an open set of $\mathbb{R}^d$.

- $\bar\theta \in \Theta$ is a zero-duality-gap parameter, i.e., $\mathrm{val}(P_{\bar\theta}) = \mathrm{val}(D_{\bar\theta})$.
- $\bar x \in \mathbb{R}^N$ is optimal for $(P_{\bar\theta})$, and $\bar\lambda \in \mathbb{R}^m$ is optimal for $(D_{\bar\theta})$.
- $\bar Q := \mathcal{Q}_{\bar\theta}(\bar\lambda) \in \mathbb{S}^N$ is the Hessian of the Lagrangian at $\bar\theta$. Note that $\bar Q \succeq 0$.
- $X_\theta := \{x \in \mathbb{R}^N : h_\theta(x) = 0\}$ is the (primal) feasible set, and $\bar X := X_{\bar\theta}$.

We proceed to describe Theorem 14, our most general result. As illustrated in Figure 3, the theorem has four assumptions. The first of them is the restricted Slater assumption, which we restate for the convenience of the reader. The remaining three are regularity conditions that will be explained later in this section.

Theorem 14 (Main result)**.** *Assume that:*

RS (restricted Slater) *There exists $\mu \in \mathbb{R}^m$ such that $\mu^T \nabla h_{\bar\theta}(\bar x) = 0$ and $(\sum_i \mu_i H^i_{\bar\theta})|_V \succ 0$, where $V := \{v \in \mathbb{R}^N : \bar Q v = 0, \bar x^T v = 0\}$.*

R1 (constraint qualification) $\mathrm{ACQ}_{\bar X}(\bar x)$ *holds.*

R2 (smoothness) $\mathcal{W} := \{(\theta, x) : h_\theta(x) = 0\}$ *is a smooth manifold nearby $\bar w := (\bar\theta, \bar x)$, and $\dim_{\bar w} \mathcal{W} = \dim \Theta + \dim_{\bar x} \bar X$.*

R3 (not a branch point) *$\bar x$ is not a branch point of $\bar X$ with respect to $v \mapsto \bar Q v$.*

Then $\mathrm{val}(P_\theta) = \mathrm{val}(D_\theta)$ *whenever θ is close enough to $\bar\theta$. Moreover, (P_θ) has a unique optimal solution x_θ, and (P^*_θ) has as unique optimal solution $x_\theta x_\theta^T$.*

Remark 7 (Strictly convex Lagrangian). If $\bar Q$ has corank-one then conditions RS and R3 always hold, so it suffices to check R1 and R2.

We proceed to discuss each of the assumptions of Theorem 14.

RS. Restricted Slater. In order to illustrate Assumption RS in an example, we will first express it in a slightly different form. Let $n := \operatorname{rank} \bar Q$, $k := N - n$, and consider a coordinate system $x = (z, u)$, $z \in \mathbb{R}^k$, $u \in \mathbb{R}^n$ such that the Hessian has the form $\bar Q = \left(\begin{smallmatrix} 0 & 0 \\ 0 & \bar C \end{smallmatrix}\right)$ where $\bar C \in \mathbb{S}^n$ is positive definite. In these coordinates we have $V = (\bar z)^\perp \times \{0^n\}$, where $\bar x = (\bar z, \bar u)$. Then

$$(\textstyle\sum_i \mu_i H^i_{\bar\theta})|_V = \mathcal{A}(\mu)|_{(\bar z)^\perp}, \quad \text{where} \quad \mathcal{A}(\mu) := \textstyle\sum_i \mu_i \nabla^2_{zz} h^i_{\bar\theta} \in \mathbb{S}^k, \tag{12}$$

and Assumption RS becomes: $\exists \mu \in \mathbb{R}^m$ s.t. $\mu^T \nabla h_{\bar\theta}(\bar x) = 0$, $\mathcal{A}(\mu)|_{(\bar z)^\perp} \succ 0$.

Example 5.1. Consider a nearest point problem as in Proposition 3. After the change of coordinates $u := y - \theta z_0$ the problem becomes

$$\min_{z \in \mathbb{R}^k, u \in \mathbb{R}^n} \|u\|^2 \quad \text{s.t.} \quad h_i(z, u + \theta z_0) = 0$$

Since $\bar\lambda = 0$ then $\bar Q = \nabla^2(\|u\|^2) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & \mathbb{1}_n \end{smallmatrix}\right)$ has the desired form (in coordinates z, u).

Let us see that Assumption RS is satisfied in Example 1.2.

Example 5.2. Consider the nearest point problem to the curve $y_2^2 = y_1^3$. Homogenizing the equations in (3) with respect to z_0 we get

$$h_0 := z_0^2 - 1, \qquad h_1 := y_2 z_0 - y_1 z_1, \qquad h_2 := y_1 z_0 - z_1^2, \qquad h_3 := y_2 z_1 - y_1^2.$$

Let $\theta \in Y$ be a parameter on the curve, which means $\theta = (t^2, t^3)$ for some $t \in \mathbb{R}$. The minimizer of the homogenized problem is $\bar{x} = (\bar{z}, \theta)$, where $\bar{z} = (1, t)$. Let us see that for any $\theta \neq 0$ the vector $\mu := (0, t, -t^2, -1)$ satisfies Assumption RS. Observe that

$$\nabla h(\bar{x}) = \begin{pmatrix} 2 & 0 & 0 & 0 \\ \theta_2 & -\theta_1 & -z_1 & 1 \\ \theta_1 & -2t & 1 & 0 \\ 0 & \theta_2 & -2\,\theta_1 & t \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ t^3 & -t^2 & -t & 1 \\ t^2 & -2t & 1 & 0 \\ 0 & t^3 & -2\,t^2 & t \end{pmatrix},$$

and thus $\mu^T \nabla h(\bar{x}) = 0$. It remains to check the positivity condition. In order to get the matrix $\mathcal{A}(\mu)$ we consider the change of coordinates $u = y - \theta z_0$, as explained in Example 5.1. Denoting $h_i'(z, u) := h_i(z, u + \theta z_0)$, then

$$\mathcal{A}(\mu) = \sum_i \mu_i \nabla^2_{zz} h_i' = \begin{pmatrix} 2\mu_0+2\mu_1\theta_2+2\mu_2\theta_1-2\mu_3\theta_1^2 & -\mu_1\theta_1+\mu_3\theta_2 \\ -\mu_1\theta_1+\mu_3\theta_2 & -2\mu_2 \end{pmatrix} = \begin{pmatrix} t^4 & -t^3 \\ -t^3 & t^2 \end{pmatrix}.$$

Note that the orthogonal complement of $\bar{z} = (1, t)$ is spanned by $\zeta := (t, -1)$. Since $\zeta^T \mathcal{A}(\lambda) \zeta = t^2(t^2 + 1)^2$ is strictly positive, then $\mathcal{A}(\lambda)|_{(\bar{z})^\perp} \succ 0$. We conclude that Assumption RS holds for all $\theta \in Y \setminus \{0\}$.

Let us show now that Assumption RS is nontrivial. The following problem violates it, and indeed has positive-duality-gap for most values of θ.

Example 5.3 (Non-informative dual)**.** Consider the following (homogenized) formulation for the nearest point problem to the twisted cubic:

$$\min_{z\in\mathbb{R}^3, y\in\mathbb{R}^3} \quad \|y - \theta\|^2 \quad \text{s.t.} \quad z_0^2 - 1 = z_1^2 + z_2^2 - 1 = 0, \quad \begin{pmatrix} z_1 & z_2 \end{pmatrix} \begin{pmatrix} z_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \end{pmatrix} = 0$$

Let us see that $\operatorname{val}(D_\theta) = 0$ for any θ, and thus there is positive-duality-gap for most values of θ. Observe that the diagonal entries of the Lagrangian Hessian are:

$$\operatorname{diag}(\mathcal{Q}_\theta(\lambda)) = \tfrac{1}{2} \sum_i \lambda_i \operatorname{diag}(\nabla^2_{xx} h_i) = (\lambda_0, \lambda_1, \lambda_1, 0, 0, 0).$$

Then the cost function of any dual feasible λ satisfies $d(\lambda) := -\lambda_0 - \lambda_1 \leq 0$. Thus $\lambda = 0$ is dual optimal, as we claimed. Let us see now that the conditions from Theorem 14 are not met. Assuming ACQ holds, then $\dim \Lambda_{\bar{\theta}}(\bar{x}) = 5 - (6 - 1) = 0$ (see Remark 5) and thus $\Lambda_{\bar{\theta}}(\bar{x}) = \{0\}$. Since the μ from Assumption RS must belong $\Lambda_{\bar{\theta}}(\bar{x})$, then the only choice is $\mu = 0$. It follows that Assumption RS fails.

Remark 8. Examples 1.1 and 5.3 illustrate that different QCQP formulations of the same problem might have different stability properties. In general, an optimal QCQP formulation of a polynomial optimization problem can be derived by including all quadratic constraints that are valid on the variety. This is equivalent to working modulo the coordinate ring (see e.g., [10])

R1. Constraint qualification. The most basic regularity assumption we make is ACQ, i.e., that $\bar{X}$ is smooth nearby $\bar{x}$, and that the tangent space of $\bar{X}$ at $\bar{x}$ is spanned by the gradients of the constraints. Note that ACQ is needed simply to guarantee the existence of Lagrange multipliers at $\bar{x}$.

R2. Smoothness. Another natural condition to make is that the dependence of the feasible set X_θ as a function of θ is smooth (continuously differentiable). More precisely, we require that $\mathcal{W} := \{(\theta, x) : h_\theta(x) = 0\}$ is a smooth manifold nearby $\bar{w} := (\bar{\theta}, \bar{x})$, and that its local dimension is precisely $\dim \Theta + \dim_{\bar{x}} \bar{X}$.

Remark 9. For nearest point problems the feasible set X_θ is independent of θ. Therefore, $\mathcal{W} = \Theta \times \bar{X}$ and condition R2 is satisfied.

R3. Not a branch point. The ACQ property guarantees regularity of $\bar{X}$ nearby $\bar{x}$. When the Lagrangian function is not strictly convex, we also require that $\bar{x}$ remains regular after projecting onto the range of $\bar{Q}$. The following example motivates this.

Example 5.4. Consider the nearest point problem to the curve Y defined by $y_2^2 = y_1^3$. In Example 1.2 we introduced an auxiliary variable z_1 to phrase the problem as a QCQP. The lifted curve in $\mathbb{R}^3$ is the twisted cubic, which is nonsingular everywhere; see Figure 2. But curve Y has a singularity at $(0, 0)$. This singularity is problematic, since it means that the nearest point map is not uniquely defined locally.

As in the above example, when the objective function is not strictly convex (e.g., due to auxiliary variables) we need that $\bar{x}$ is regular after a suitable projection. By regularity we mean that $\bar{x}$ is not a branch point, as formalized next.

Definition 5.1 (Branch point)**.** Let $\pi : \mathbb{R}^N \to \mathbb{R}^n$ be a linear map. Let $\bar{X} \subseteq \mathbb{R}^N$ be the zero set of $\bar{h}$, and let $T_x\bar{X} := \ker \nabla \bar{h}(x)$ denote the tangent space of $\bar{X}$ at x. We say that x is a *branch point* of $\bar{X}$ with respect to π if there is a nonzero vector $v \in T_x\bar{X}$ with $\pi(v) = 0$.

Example 5.5. Let $\pi : (z, y_1, y_2) \mapsto (y_1, y_2)$, and consider the projection of the twisted cubic from Figure 2. Notice that the tangent line at the point $(0, 0, 0)$ is precisely the z-axis, and thus $(0, 0, 0)$ is a branch point.

The last regularity assumption is that $\bar{x}$ is not a branch point with respect to the map $v \mapsto \bar{Q}v$, or equivalently, with respect to the projection $\pi_{\bar{Q}}$ onto the range of $\bar{Q}$.

Example 5.6. Consider a nearest point problem as in Proposition 3. Since $\bar{\lambda} = 0$ then $\bar{Q} = \nabla^2(\|y - \theta z_0\|^2)$, and its range is $\{(z_0, z', y) \in \mathbb{R}^{n+k} : z' = 0, z_0 = -y^T\bar{\theta}\}$ of dimension n. In particular, when $\bar{\theta} = 0$ then $\pi_{\bar{Q}}$ is simply $(z, y) \mapsto y$. Moreover, Assumption R3 holds if and only if $\nabla_{z'} f(\bar{z}', \bar{y})$ is injective.

6. Proof of main theorem

Recall the Lagrange multiplier mapping $\mathfrak{L} : \Theta \rightrightarrows \mathbb{R}^N \times \mathbb{R}^m$ from (5). In Theorem 7 we showed that local continuity of this mapping (Assumption SC), together with Assumption RS, guarantee zero-duality-gap nearby $\bar{\theta}$. Thus, our goal now is to see that Assumptions (R1-R3) imply local continuity of $\mathfrak{L}$. In order to do so, we consider the following property of set-valued mappings [12, 30].

Definition 6.1 (Aubin property)**.** Let $\mathfrak{F} : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ be a set-valued mapping. $\mathfrak{F}$ has the *Aubin property* at $\bar{p} \in \mathbb{R}^d$ for $\bar{y} \in \mathbb{R}^n$ if $\bar{y} \in \mathfrak{F}(\bar{p})$ and there is a constant $\kappa \geq 0$ and neighborhoods $U \ni \bar{y}, V \ni \bar{p}$ such that

$$\mathfrak{F}(p') \cap U \subseteq \mathfrak{F}(p) + \kappa|p' - p|\mathcal{B} \quad \text{for all } p', p \in V,$$

where $\mathcal{B} \subseteq \mathbb{R}^n$ denotes the unit ball.

The Aubin property implies local continuity, as stated next.

Lemma 15. *Let $\mathfrak{F} : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ be a mapping with closed graph. Assume that $\mathfrak{F}$ has the Aubin property at $\bar{p}$ for $\bar{y}$. Then there exists a closed neighborhood $U_0 \ni \bar{y}$ such that $p \mapsto \mathfrak{F}(p) \cap U_0$ is continuous at $\bar{p}$.*

Proof. See Appendix A. □

Because of the above lemma, it is enough for us to show that the Lagrange multiplier mapping $\mathfrak{L}$ has the Aubin property. In particular, the proof Theorem 14 can be reduced to the following proposition.

Proposition 16. *Let $\bar{x}$ be optimal to ($P_{\bar{\theta}}$) and $\bar{\lambda} \in \Lambda_{\bar{\theta}}(\bar{x})$ be dual feasible. Under Assumptions (R1-R3) the mapping $\mathfrak{L}$ has the Aubin property at $\bar{\theta}$ for $(\bar{x}, \bar{\lambda})$.*

Proof of Theorem 14 (assuming Proposition 16). Since $\mathfrak{L}$ has the Aubin property at $\bar{\theta}$ for $(\bar{x}, \bar{\lambda})$, then by Lemma 15 there is a neighborhood $U \ni (\bar{x}, \bar{\lambda})$ such that the mapping $\theta \mapsto \mathfrak{L}(\theta) \cap U$ is continuous at $\bar{\theta}$. Then Assumption SC holds, and the result follows from Theorem 7. □

In the remaining of this section we will prove Proposition 16, thus concluding the proof of Theorem 14.

6.1. **The implicit function theorem.** The main technical tool we use for Proposition 16 is the implicit function theorem, which can be phrased in terms of the Aubin property (see [12, Ex. 4D.3]).

Theorem 17 (Implicit function)**.** *Given $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable, consider the solution mapping*

$$\mathfrak{F} : \mathbb{R}^d \rightrightarrows \mathbb{R}^n, \qquad p \mapsto \{y \in \mathbb{R}^n : f(p, y) = 0\}.$$

Let $\bar{p}, \bar{y}$ be such that $\bar{y} \in \mathfrak{F}(\bar{p})$. If $\nabla_y f(\bar{p}, \bar{y})$ is surjective, then $\mathfrak{F}$ satisfies the Aubin property at $\bar{p}$ for $\bar{y}$.

Note that Proposition 16 would be immediate if $\mathfrak{L}$ satisfied the hypothesis from Theorem 17. Unfortunately this is not true, since the defining equations of $\mathfrak{L}$ may have linearly dependent gradients. In order to fix this problem, we consider a maximal subset of the equations $h' \subseteq h$ such that $\{\nabla_x h^i_{\bar{\theta}}(\bar{x})\}_{h^i \in h'}$ are linearly independent. Equivalently, $h' \subseteq h$ is such that $\nabla_x h'_{\bar{\theta}}(\bar{x})$ is full rank, and has the same rank as $\nabla_x h_{\bar{\theta}}(\bar{x})$. Consider the modified solution mapping

$$\mathfrak{L}' : \theta \mapsto \{(x, \lambda) : h'_\theta(x) = 0,\ \mathcal{Q}_\theta(\lambda)x = 0\}.$$

This new mapping satisfies the assumptions of Theorem 17, as seen next.

Lemma 18. *Under Assumption R3, $\mathfrak{L}'$ has the Aubin property at $\bar{\theta}$ for $(\bar{x}, \bar{\lambda})$.*

Proof. Let us see that the surjectivity condition from Theorem 17 is satisfied. To simplify the notation we will ignore the dependence on θ, since the only parameter we consider in this proof is $\bar{\theta}$. Let $J' := \nabla_x h'(\bar{x})$, which is a submatrix of $J := \nabla_x h(\bar{x})$. Note that by construction the rows of J' are independent and $\ker J' = \ker J$. Let $f(x,\lambda) := (h'(x), \mathcal{Q}_{\bar{\theta}}(\lambda)x) \in \mathbb{R}^{m+N}$. We need to show that the rows of $\nabla f(\bar{x},\bar{\lambda})$ are independent. Denoting $\bar{Q} := \mathcal{Q}_{\bar{\theta}}(\bar{\lambda})$, then

$$\nabla_{\lambda,x} f(\bar{x},\bar{\lambda}) = \begin{pmatrix} 0 & \nabla h'(\bar{x}) \\ \nabla h(\bar{x})^T & \mathcal{Q}_{\bar{\theta}}(\bar{\lambda}) \end{pmatrix} = \begin{pmatrix} 0 & J' \\ J^T & \bar{Q} \end{pmatrix}.$$

Let (u,v) be a vector in the left kernel of $\nabla f(\bar{x},\bar{\lambda})$, i.e., $v^T J^T = 0$, $u^T J' + v^T \bar{Q} = 0$. We need to show that $(u,v) = 0$. Since $v \in \ker J = \ker J'$ then $0 = (u^T J' + v^T \bar{Q})v = v^T \bar{Q} v$, and thus $\bar{Q}v = 0$. As $v \in \ker J$ and $\bar{Q}v = 0$, then $v = 0$ by Assumption R3. Therefore $0 = u^T J' + v^T \bar{Q} = u^T J'$, and thus $u = 0$ since the rows of J' are independent. □

In order to prove Proposition 16 it remains to see that the modified mapping $\mathfrak{L}'$ agrees with $\mathfrak{L}$, at least locally. This follows from the following lemma.

Lemma 19. *Let $X_\theta \subseteq X'_\theta \subseteq \mathbb{R}^N$ be the zero sets of h_θ, h'_θ. Under Assumptions R1 and R2, there are neighborhoods $V_0 \ni \bar{\theta}$ and $U_0 \ni \bar{x}$ such that $X_\theta \cap U_0 = X'_\theta \cap U_0$ for all $\theta \in V_0$.*

The proof of Lemma 19 requires an auxiliary lemma.

Lemma 20. *Let $\mathcal{W} := \{w \in \mathbb{R}^K : h(w) = 0\}$, where $h = (h_1, \ldots, h_m)$, and assume that $\mathcal{W}$ is a smooth D-dimensional manifold nearby $\bar{w}$. Let $h' = (h_1, \ldots, h_{K-D}) \subseteq h$ be such that their gradients at $\bar{w}$ are linearly independent. Then there is a neighborhood $U \subseteq \mathbb{R}^K$ of $\bar{w}$ such that $\mathcal{W} \cap U = \mathcal{W}' \cap U$, where $\mathcal{W}' := \{w : h'(w) = 0\}$.*

Proof. By the implicit function theorem $\mathcal{W}'$ is a D-dimensional manifold nearby $\bar{w}$. Thus, there is a neighborhood $U \subseteq \mathbb{R}^K$ of $\bar{w}$ such that $\mathcal{W} \cap U$ is a submanifold of $\mathcal{W}' \cap U$. Since they have the same dimension, $\mathcal{W} \cap U$ must be an open set of $\mathcal{W}' \cap U$. □

Proof of Lemma 20. Let $\mathcal{W} := \{(\theta,x) : h_\theta(x) = 0\}$ and $\mathcal{W}' := \{(\theta,x) : h'_\theta(x) = 0\}$. We will use Lemma 20 to show the existence of a neighborhood $U \ni \bar{w}$, such that $\mathcal{W} \cap U = \mathcal{W}' \cap U$. Note that this would conclude the proof. By Assumption R2 we know that $\mathcal{W}$ is a smooth manifold nearby $\bar{w} := (\bar{x},\bar{\theta})$ of dimension $D := \dim \Theta + \dim_{\bar{x}} \bar{X}$. Recall that by construction of h' the gradients $\{\nabla h^i_{\bar{\theta}}(\bar{x})\}_{h^i \in h'}$ are linearly independent, and the number of equations is $|h'| = \operatorname{rank} \nabla h_{\bar{\theta}}(\bar{x})$. Since $\mathrm{ACQ}_{\bar{X}}(\bar{x})$ holds, then

$$|h'| = \operatorname{rank} \nabla h_{\bar{\theta}}(\bar{x}) = N - \dim_{\bar{x}} \bar{X} = (\dim \Theta + N) - D.$$

Thus the assumptions of Lemma 20 are satisfied, as wanted. □

We are finally ready to prove Proposition 16.

Proof of Proposition 16. The Aubin property is a local condition. Since $\mathfrak{L}, \mathfrak{L}'$ agree nearby $\bar{\theta}, \bar{x}$ (Lemma 19), and since $\mathfrak{L}'$ has the Aubin property (Lemma 18), then the same holds for $\mathfrak{L}$. □

7. Applications

7.1. **Estimation problems with strictly convex objective.** Let us show some immediate consequences of Theorems 1 and 8. We will use the following well-known property (see e.g., [13, §16.6]).

Lemma 21. *Let $h \subseteq \mathbb{R}[x]$ be a polynomial system and let X be its variety. If the ideal $\langle h \rangle$ is radical then ACQ holds for each smooth point of X. In particular, ACQ holds generically on X (on a dense open subset).*

We first consider two nearest point problems. We will apply Theorem 1, and thus we will need to check the ACQ property.

Example 7.1 (Triangulation)**.** Given ℓ projective cameras $P_j : \mathbb{P}^3 \to \mathbb{P}^2$ and noisy images $\hat{u}_j \in \mathbb{R}^2$ of an unknown point $z \in \mathbb{P}^3$, the triangulation problem is

$$\min_{u \in U} \quad \sum_{j=1}^{\ell} \|u_j - \hat{u}_j\|^2, \qquad U := \{u \in (\mathbb{R}^2)^\ell : \exists z \in \mathbb{P}^3 \text{ s.t. } u_j = \Pi P_j z \text{ for } 1 \leq j \leq \ell\},$$

where $\Pi : \mathbb{P}^2 \to \mathbb{R}^2$ is the dehomogenization $(y_1 : y_2 : y_3) \mapsto (y_1/y_3, y_2/y_3)$. Notice that this problem is parameterized by $\theta = (\hat{u}_1, \ldots, \hat{u}_\ell)$. U is known as the *multiview variety.* Assume that either $\ell = 2$, or $\ell \geq 4$ and the camera centers are not coplanar. Then the variety U can be described as

$$U = \{u \in (\mathbb{R}^2)^\ell : f_{ij}(u_i, u_j) = 0, \ 1 \leq i < j \leq \ell\},$$

where f_{ij} are some quadratic equations known as the epipolar constraints [20]. By using this description of U we obtain a QCQP. Moreover, these epipolar equations define a radical ideal [20], and thus ACQ holds generically (Lemma 21). It follows from Theorem 1 that the SDP relaxation of this QCQP is (generically) tight under small noise.

Remark 10. The above SDP relaxation was considered in [1], where they also showed tightness under low noise.

Example 7.2 (Rank one approximation)**.** Consider the problem of finding the nearest rank one tensor. Let $\mathbb{R}^{n_1 \times \cdots \times n_\ell}$ be the set of tensors of dimensions $(n_1, \ldots, n_\ell)$ and let $X \subseteq \mathbb{R}^{n_1 \times \cdots \times n_\ell}$ be the space of rank one tensors. X is known as the *Segre variety*, and it is also defined by quadratic equations (the 2×2 minors of the tensor flattenings). Therefore, this nearest point problem is a QCQP, and we can consider its SDP relaxation. The Segree variety is smooth at any point other than the origin. Since the ideal is radical, then ACQ holds at all these points. Thus, under low noise assumptions this problem is solved exactly by the SDP relaxation. This result also extends to the case of symmetric tensors, given that the *Veronese variety* is also defined by quadrics.

Remark 11. An essentially equivalent SDP relaxation for the nearest rank one tensor was proposed in [28]. No tightness results were known.

We will now see some applications of Theorem 8. Note that the theorem has three assumptions: the objective is strictly convex, ACQ holds, and the minimizer of the noiseless case is also the global minimum. Since in the problems below the objective function is a squared loss function, and since in the noiseless case the objective value is zero, then the last condition is always satisfied. Thus, we will only check strict convexity and ACQ.

Example 7.3 ($SO(d)$ synchronization)**.** Consider the problem of determining the absolute rotations of $n+1$ objects given (noisy) relative rotations among some pairs. Let $SO(d)$ denote the *special orthogonal group*. Given a graph $G=(V,E)$, where $V=\{0,\ldots,n\}$, and matrices $\hat{R}_{ij}\in\mathbb{R}^{d\times d}$ for each $ij\in E$, the problem is

$$\min_{R_1,\ldots,R_n\in SO(d)} \sum_{ij\in E} \|R_j-\hat{R}_{ij}R_i\|_F^2, \qquad SO(d):=\{R\in\mathbb{R}^{d\times d}: R^TR=\mathbb{1}_d,\ \det(R)=1\}, \tag{13}$$

where $R_0:=\mathbb{1}_n$ is the reference point, This problem is parametrized by $\theta=(\hat{R}_{ij})_{ij\in E}$. To obtain a QCQP, we can replace $SO(d)$ by the *orthogonal group* $O(d)$:

$$\min_{R_1,\ldots,R_n\in O(d)} \sum_{ij\in E} \|R_j-\hat{R}_{ij}R_i\|_F^2, \qquad O(d):=\{R\in\mathbb{R}^{d\times d}: R^TR=\mathbb{1}_d\}. \tag{14}$$

Observe that (13) and (14) have the same minimizer in the low noise regime, given that $SO(d)^n$ is a connected component of $O(d)^n$. Consider the SDP relaxation of the QCQP (14). Note that the objective function is strictly convex (Lemma 22), and that ACQ is satisfied everywhere since the variety is smooth and the ideal is radical (Lemma 21). Thus, under low noise, the SDP relaxation finds the true minimizer of (14), which is the same of (13).

Lemma 22. *Let $G=(V,E)$ be a connected graph, let $x_0\in\mathbb{R}^k$, and let $L_{ij}:\mathbb{R}^k\to\mathbb{R}^k$ be invertible linear maps for $ij\in E$. Then the function $f(x):=\sum_{ij\in E}\|x_j-L_{ij}x_i\|^2$, where $x=(x_1,\ldots,x_n)\in(\mathbb{R}^k)^n$, is strictly convex.*

Proof. We may assume that the reference point $x_0=0$ (otherwise, simply apply an affine transformation). Since $f(x)$ is convex and homogeneous, it suffices to see that $f(x)=0$ implies $x=0$. If $f(x)=0$ then $x_j=L_{ij}x_i$ for each $ij\in E$. Since $x_0=0$ and G is connected it is clear that each x_i must be zero. □

Remark 12. An alternative QCQP formulation for the $SO(3)$ synchronization problem can be obtained by representing rotations with quaternions [16]. The same analysis as above shows that the corresponding SDP relaxation is tight in the low noise regime, as was observed experimentally in [16].

Example 7.4 ($SE(d)$ synchronization)**.** A natural extension of the above problem is to replace rotation matrices by elements of the *special Euclidean group* $SE(d)$. Given a graph $G=(V,E)$, and $\hat{R}_{ij}\in\mathbb{R}^{d\times d}$, $\hat{u}_{ij}\in\mathbb{R}^d$ for $ij\in E$, the problem is

$$\min_{R_i\in SO(d),\,u_i\in\mathbb{R}^d} \sum_{ij\in E} \|R_j-\hat{R}_{ij}R_i\|_F^2+\|u_j-u_i-R_i\hat{u}_{ij}\|^2, \tag{15}$$

where $R_0:=\mathbb{1}_d, u_0:=0$. As before, we can replace $SO(d)$ with $O(d)$ to obtain a QCQP, and consider its SDP relaxation. An argument similar to Lemma 22 shows that the objective function is strictly convex, and thus the SDP recovers the minimizer of (15) under low noise.

Remark 13. SDP relaxations for $SE(d)$ synchronization have received considerable attention in past years and similar tightness results have been derived [31, 36].

Example 7.5 (Orthogonal Procrustes)**.** Given $n, k, m_1, m_2 \in \mathbb{N}$ and matrices $A \in \mathbb{R}^{m_1 \times n}$, $B \in \mathbb{R}^{m_1 \times m_2}$, $C \in \mathbb{R}^{k \times m_2}$, the weighted orthogonal Procrustes problem, also known as Penrose regression problem, is

$$\min_{X \in \mathbb{R}^{n \times k}} \quad \|AXC - B\|_F^2, \quad \text{s.t.} \quad X^T X = \mathbb{1}_k\,. \tag{16}$$

Note that the above is a QCQP parametrized by $\theta = (A, B, C)$. ACQ holds everywhere since the variety (the Stiefel manifold) is smooth and the ideal is radical. The objective function is strictly convex as long as the linear map $X \mapsto AXC$ is injective. In such cases the SDP relaxation will be tight under low noise.

Remark 14. Problem (16) may have several local optima, and thus local methods may fail [9, 35]. The above SDP relaxation was considered in [10].

7.2. **Stability of unconstrained SOS.** Let $\mathbb{R}[z]_{2d}$ be the vector space of multivariate polynomials of degree at most $2d$ in variables $z = (z_1, \ldots, z_n)$. Consider the parametric family of polynomial optimization problems:

$$(POP_\theta) \qquad \min_{z \in \mathbb{R}^n}\, p_\theta(z), \qquad \text{where } p_\theta \in \mathbb{R}[z]_{2d} \text{ depends continuously on } \theta.$$

We will analyze the stability of its *sum-of-squares* (SOS) relaxation.

We now briefly review the SOS method. A polynomial $f \in \mathbb{R}[z]_{2d}$ is SOS if it can be written in the form $f(z) = \sum_i f_i(z)^2$ for some $f_i \in \mathbb{R}[z]_d$. Consider the SOS cone

$$\Sigma_{n,2d} := \{ f \in \mathbb{R}[z]_{2d} : f(z) \text{ is SOS } \}.$$

The *SOS relaxation* of (POP_θ) is

$$(SOS_\theta) \qquad \max_{\gamma \in \mathbb{R}}\, \gamma \quad \text{s.t.} \quad p_\theta(z) - \gamma \in \Sigma_{n,2d}.$$

The above can be efficiently solved with an SDP, and it always holds that $\mathrm{val}(SOS_\theta) \leq \mathrm{val}(POP_\theta)$. The relaxation is *tight* at θ if $\mathrm{val}(SOS_\theta) = \mathrm{val}(POP_\theta)$. If the minimizer of p_θ is unique, it might also be possible to recover it from the relaxation.

Assume now that for a fixed value of $\bar\theta$ we know that the relaxation is tight. As before, we investigate the behavior of the SOS relaxation as $\theta \to \bar\theta$.

Example 7.6. For the polynomial $p_\theta(z) := z_1^4 z_2^2 + z_1^2 z_2^4 + \theta z_1^2 z_2^2 \in \mathbb{R}[z]_6$ we have:

$$\begin{aligned} \theta \geq 0 \quad &\Longrightarrow \quad \mathrm{val}(POP_\theta) = \mathrm{val}(SOS_\theta) = 0, \\ \theta < 0 \quad &\Longrightarrow \quad \mathrm{val}(POP_\theta) = \tfrac{1}{27}\theta^3 \text{ and } (SOS_\theta) \text{ is infeasible.} \end{aligned}$$

Hence the relaxation is not stable nearby $\bar\theta = 0$.

The following theorem shows stability under a certain interiority condition.

Theorem 23. *Let $\bar\theta$ be such that $\bar\gamma := \mathrm{val}(POP_{\bar\theta}) = \mathrm{val}(SOS_{\bar\theta})$ and there is a unique minimizer $\bar z$. Consider the face $K_{\bar z}$ of the cone $\Sigma_{n,2d}$ given by the vanishing at $\bar z$:*

$$K_{\bar z} := \Sigma_{n,2d} \cap L_{\bar z}, \qquad L_{\bar z} := \{ f \in \mathbb{R}[z]_{2d} : f(\bar z) = 0,\ \nabla f(\bar z) = 0 \}.$$

Note that $p_{\bar\theta} - \bar\gamma \in K_{\bar z}$. If $p_{\bar\theta} - \bar\gamma$ lies in the relative interior of $K_{\bar z}$, then the relaxation (SOS_θ) is tight and recovers the minimizer whenever θ is close enough to $\bar\theta$.

We proceed to prove Theorem 23. We may assume WLOG that $\bar\gamma = 0$, and thus $p_{\bar\theta} \in \Sigma_{n,2d}$. In order to use our methods, we need to rephrase (POP_θ) as a QCQP. Let

$$x := (z^\alpha)_{\alpha\in J} \in (\mathbb{R}[z]_d)^N, \quad \text{where} \quad J := \{\alpha \in \mathbb{N}^n : \textstyle\sum_i \alpha_i \le d\}, \quad N := \binom{n+d}{d},$$

be the vector with all N monomials of degree at most d. Notice that any $f \in \mathbb{R}[z]_{2d}$ can be written in the form $f(z) = x^TQx$ for some $Q \in \mathbb{S}^N$. We say that such a Q is a *Gram matrix* of f. Moreover, f is SOS if and only if it has a positive semidefinite Gram matrix. We need two properties of these Gram matrices.

Lemma 24. *Assume that $\bar p \in \operatorname{int} K_{\bar z}$. Then $\bar p$ has a Gram matrix $\bar Q \succeq 0$ of corank-one.*

Proof. Consider the linear map

$$\phi : \mathbb{S}^N \to \mathbb{R}[z]_{2d}, \qquad A \mapsto x^TAx. \tag{17}$$

Let $\bar x \in \mathbb{R}^N$ be given by evaluating each of the monomials in x at $\bar z$. Let

$$S := \{Q \in \mathbb{S}^N : Q \succeq 0,\ Q\bar x = 0\}$$

and observe that $K_{\bar z} = \phi(S)$. Since linear maps preserve relative interiors of convex sets, then $\operatorname{int} K_{\bar z} = \phi(\operatorname{int} S)$. It follows that $\bar p$ has a Gram matrix $\bar Q \succeq 0$ such that $\bar Q\bar x = 0$, and $\operatorname{corank} \bar Q = 1$. □

Lemma 25. *Let $\bar Q$ be a Gram matrix of $p_{\bar\theta}$. Then $Q_\theta := \phi^\dagger(p_\theta - p_{\bar\theta}) + \bar Q$ is a Gram matrix of p_θ, where $\phi^\dagger$ is the pseudo-inverse of the linear map in* (17).

Proof. Follows by noticing that Q is a Gram matrix of f if and only if $\phi(Q) = f$. □

By the above lemmas, we know that there exist Gram matrices $Q_\theta \in \mathbb{S}^N$ for each p_θ such that: $Q_{\bar\theta} \succeq 0$ and has corank-one, and the dependence on θ is continuous. Thus the parametric optimization problem (POP_θ) can be phrased as

$$\min_{x\in X} x^TQ_\theta x, \qquad \text{where} \qquad X := \{(z^\alpha)_{\alpha\in J} : z \in \mathbb{R}^n\} \subseteq \mathbb{R}^N.$$

The above is indeed a QCQP since X, the Veronese variety, is defined by quadratic equations

$$X := \{x \in \mathbb{R}^N : x_0 = 1,\ x_{\alpha_1}x_{\alpha_2} = x_{\beta_1}x_{\beta_2}\ \ \forall \alpha_1,\alpha_2,\beta_1,\beta_2 \in J \text{ s.t. } \alpha_1+\alpha_2 = \beta_1+\beta_2\}.$$

Since we have a QCQP formulation, its Lagrangian dual gives an SDP relaxation of (POP_θ). Moreover, this Lagrangian dual coincides with (SOS_θ). The proof of Theorem 23 now follows from Theorem 8.

Proof of Theorem 23. Consider the above QCQP. Since the first coordinate of x is always one, let $x = (1, y)$ with $y \in \mathbb{R}^{N-1}$. Similarly, let $\bar x = (1, \bar y)$ where $\bar y = (\bar z^\alpha)_{\alpha\in J\setminus\{0\}}$. Let $q_\theta(y) := x^TQ_\theta x$. By construction we know that $Q_{\bar\theta} \succeq 0$, $Q_{\bar\theta}\bar x = 0$, $\operatorname{corank} Q_{\bar\theta} = 1$. It follows that $\min_y q_{\bar\theta}(y) = 0$, and it is attained at $\bar y$. The corank-one assumption means that $q_{\bar\theta}$ is strictly convex. In order to apply Theorem 8 it remains to see that the ACQ assumption is satisfied. Note that the variety X is smooth since the unique singularity of the Veronese variety is the origin, but we are fixing the first coordinates to be one. Since the ideal is radical, Lemma 21 implies that ACQ holds everywhere. □

7.3. **Noisy Euclidean distance matrix completion.** We now show a simple application of Theorem 14. Consider the problem of determining the location of $n+1$ objects given the (noisy) pairwise distances among some of them. Formally, given $p \in \mathbb{N}$, a graph $G = (V, E)$, and positive numbers θ_{ij} for each $ij \in E$, the goal is to find $t_i \in \mathbb{R}^p$ such that

$$\|t_i - t_j\|^2 \approx \theta_{ij} \qquad \text{for } ij \in E.$$

The problem can be modelled as finding the nearest point to the variety of *Euclidean distance matrices*:

(18)
$$\min_{d\in D} \sum_{ij\in E} (d_{ij} - \theta_{ij})^2, \qquad D := \{d \in \mathbb{R}^{\binom{V}{2}} : \exists t_i \in \mathbb{R}^p \text{ s.t. } d_{ij} = \|t_i - t_j\|^2 \text{ for } ij \in \tbinom{V}{2}\}.$$

We point out that given a valid Euclidean distance matrix $d \in D$ recovering the locations t_i amounts to an eigenvalue decomposition.

Remark 15. It follows from [32] that the above problem is NP-hard even if $p = 1$.

Remark 16. The special case in which all pairs are observed, i.e., $E = \binom{V}{2}$, is nontrivial, and it is known as *multidimensional scaling* (see e.g., [8]).

We focus here on the one-dimensional case ($p = 1$). The variety of 1D Euclidean distance matrices is defined by some quadratics known as the *Cayler-Menger determinants*:

$$D = \{d \in \mathbb{R}^{\binom{V}{2}} : h_{ijk}(d) = 0 \text{ for } ijk \in \tbinom{V}{3}\}, \qquad h_{ijk}(d) := \det \begin{pmatrix} 0 & d_{ij} & d_{ik} & 1 \\ d_{ij} & 0 & d_{jk} & 1 \\ d_{ik} & d_{jk} & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

In particular, problem (18) is a QCQP (when $p = 1$). Notice that if all pairs are observed, i.e, $E = \binom{V}{2}$, then Theorem 1 tells us that its SDP relaxation is tight under low noise. In case that there are missing pairs this argument does not apply. The next example illustrates that Theorem 14 might be used to show tightness in such cases.

Example 7.7. Consider the graph

$$G = (V, E), \quad V = \{0, 1, 2, 3\}, \quad E = \{02, 03, 12, 13, 23\},$$

whose only missing edge is 01. Let $\bar\theta \in \mathbb{R}^E$ be a zero duality gap parameter, i.e., there is some $t \in \mathbb{R}^n$ such that $\bar\theta_{ij} = (t_i - t_j)^2$ for all $ij \in E$. We assume that $\bar\theta$ is generic (in particular, $t_i \neq t_j$). We will use Theorem 14 to show zero-duality-gap nearby $\bar\theta$. We denote $t_{ij} := t_j - t_i$ to simplify the notation. Let us split the vector $d = (z_{01}, y)$, where $z_{01} \in \mathbb{R}$ and $y \in \mathbb{R}^E$. The minimizer of (18) is $\bar d = (\bar z_{01}, \bar\theta)$, where $\bar z_{01} := t_{01}^2$, and its Jacobian is

$$\nabla h(\bar d) = \nabla_{z,y} h(\bar d) = 4 \begin{pmatrix} -t_{12}t_{02} & t_{12}t_{01} & 0 & -t_{01}t_{02} & 0 & 0 \\ -t_{13}t_{03} & 0 & t_{13}t_{01} & 0 & -t_{01}t_{03} & 0 \\ 0 & -t_{23}t_{03} & t_{23}t_{02} & 0 & 0 & -t_{02}t_{03} \\ 0 & 0 & 0 & -t_{23}t_{13} & t_{23}t_{12} & -t_{12}t_{13} \end{pmatrix}.$$

Conditions (R1-R3) from Theorem 14 are easy to check:

R1. Note that $\operatorname{rank} \nabla h(\bar d) \geq 3$ (generically), and that $\dim D = 3$. Thus, ACQ holds.
R2. The feasible set is independent of θ, so $\mathcal{W} = \Theta \times \bar X$.

R3. By Example 5.6, it is enough to check that $\nabla_{z_{01}} h(\bar{d})$ is injective. This follows by noticing that the first row of $\nabla h(\bar{d})$ is nonzero.

For the remaining condition, Assumption RS, we consider the vector $\mu \in \mathbb{R}^{\binom{V}{3}}$ with entries

$$\mu_{012} = -t_{03}t_{13}t_{23}, \quad \mu_{013} = t_{02}t_{12}t_{23}, \quad \mu_{023} = -t_{01}t_{12}t_{13}, \quad \mu_{123} = t_{01}t_{02}t_{03}. \tag{19}$$

We will see that either μ or $-\mu$ satisfy Assumption RS. It is easy to see that $\mu^T \nabla h(\bar{d}) = 0$, so it remains to verify the positivity condition. As in Example 5.1, we consider the change of coordinates $u = y - \theta z_0$, and let $h'_{ijk}(z_0, z_{01}, u) := h_{ijk}(z_{01}, u + \theta z_0)$. It can be checked that $\nabla^2_{zz} h'_{ijk} = 0$ if $ij \neq 01$, and the matrix $\mathcal{A}$ from (12) is:

$$\begin{aligned}\mathcal{A}(\mu) = \sum_{ijk} \mu_{ijk} \nabla^2_{zz} h'_{ijk} &= \mu_{012} \begin{pmatrix} t_{01}^2(t_{01}+2t_{12})^2 & -t_{02}^2-t_{12}^2 \\ -t_{02}^2-t_{12}^2 & 1 \end{pmatrix} + \mu_{013} \begin{pmatrix} t_{01}^2(t_{01}+2t_{13})^2 & -t_{03}^2-t_{13}^2 \\ -t_{03}^2-t_{13}^2 & 1 \end{pmatrix} \\ &= t_{23}^2(t_0 + t_1 - t_2 - t_3) \begin{pmatrix} t_{01}^4 & -t_{01}^2 \\ -t_{01}^2 & 1 \end{pmatrix}.\end{aligned}$$

Notice that $\mathcal{A}(\mu) \neq 0$ and that $\mathcal{A}(\mu)\bar{z} = 0$ where $\bar{z} := (1, t_{01}^2)$. It follows that either $\mathcal{A}(\mu)|_{(\bar{z})^\perp} \succ 0$ or $\mathcal{A}(-\mu)|_{(\bar{z})^\perp} \succ 0$, and thus Assumption RS holds.

Remark 17. The argument from above can be readily adapted to the case $V = \{0, \ldots, n\}$, $E = \binom{V}{2} \setminus \{01\}$. Conditions (R1-R3) are easy. Consider the vector $\mu \in \mathbb{R}^{\binom{V}{3}}$ whose only nonzero entries are the ones in (19). The matrix $\mathcal{A}(\mu)$ is then the same as before, and thus Assumption RS holds.

8. Discussion and open problems

We analyzed the stability of semidefinite relaxations of a parametric family of QCQP's. In particular, Theorems 1, 8 and 14 give sufficient conditions for stability. Theorems 1 and 8 have a single technical assumption, ACQ, which is quite easy to check. We saw several applications of both theorems in Sections 7.1 and 7.2. Theorem 14 requires Assumption RS (restricted Slater) and three additional regularity conditions. Although Assumption RS can be efficiently checked for any concrete instance (see Section 3.2), verifying it for a family of problems can sometimes be challenging. In this paper we provided an elementary illustration of the theorem (Section 7.3), but an interesting open problem is to identify larger classes of problems for which Assumption RS is satisfied.

We also established quantitative bounds on the region of stability in Theorem 12 and Corollary 13. However, as illustrated in Figure 4, the guaranteed region of stability can be considerably smaller than the actual one. This deserves further study.

Appendix A. Additional proofs

Proof of Lemma 10. Let $\gamma_\theta := q_\theta(y_\theta)$ be the optimal value. Let us first show that $\gamma_\theta \to \gamma_{\bar{\theta}}$ as $\theta \to \bar{\theta}$. Since $\gamma_\theta = q_\theta(y_\theta) \leq q_\theta(\bar{y})$ then

$$\limsup_{\theta \to \bar{\theta}} \gamma_\theta \leq \lim_{\theta \to \bar{\theta}} q_\theta(\bar{y}) = q_{\bar{\theta}}(\bar{y}) = \gamma_{\bar{\theta}}$$

Let $\rho_\theta := \min_{y\in\mathbb{R}^n} q_\theta(y)$ be the unconstrained minimum of q_θ. Clearly $\rho_\theta \leq \gamma_\theta$. Since q_θ is convex quadratic, there is an explicit formula for ρ_θ, and it can be checked that $\rho_\theta \to \rho_{\bar\theta}$. Therefore,

$$\gamma_{\bar\theta} = \rho_{\bar\theta} = \lim_{\theta\to\bar\theta} \rho_\theta \leq \liminf_{\theta\to\bar\theta} \gamma_\theta.$$

It follows that $\lim_{\theta\to\bar\theta} \gamma_\theta = \gamma_{\bar\theta}$, as we claimed.

Let us now show that $y_\theta \to \bar y$. Since $g_{\bar\theta}$ is strictly convex and $\bar y$ is the minimizer, it is sufficient to see that $g_{\bar\theta}(y_\theta) \to g_{\bar\theta}(\bar y)$. Let $t > \gamma_{\bar\theta}$ be arbitrary and let $S_\theta := \{y \in \mathbb{R}^n : g_\theta(y) \leq t\}$. Since $g_{\bar\theta}$ is strictly convex and g_θ depends continuously on θ, there is a compact set S such that $S_\theta \subseteq S$ for all θ sufficiently close to $\bar\theta$. Since $\gamma_\theta \to \gamma_{\bar\theta}$, then $g_\theta(y_\theta) < t$ when θ is close to $\bar\theta$. Therefore, we may assume that $y_\theta \in S$ for all θ. Denoting $\|\cdot\|_S$ the infinity norm on the compact set S, then

$$|g_{\bar\theta}(y_\theta) - g_{\bar\theta}(\bar y)| \leq |g_{\bar\theta}(y_\theta) - g_\theta(y_\theta)| + |g_\theta(y_\theta) - g_{\bar\theta}(\bar y)| \leq \|g_{\bar\theta} - g_\theta\|_S + |\gamma_\theta - \gamma_{\bar\theta}| \xrightarrow{\theta\to\bar\theta} 0$$

as wanted. □

Proof of Corollary 13. We will use a simple variation of Theorem 12. Recall the definitions of $\mathcal{C}_\theta(\mu)$, $\mathcal{Q}_\theta(\lambda)$ from (10). Since $\mathcal{C}_\theta(\mu)$ is a submatrix of $\mathcal{Q}_\theta(\lambda)$, where $\lambda = (\lambda_0, \mu)$, then their eigenvalues satisfy $\nu_1(\mathcal{C}_\theta(\mu)) \leq \nu_2(\mathcal{Q}_\theta(\lambda))$. The proof of Theorem 12 relied on lower bounding $\nu_2(\mathcal{Q}_\theta(\lambda))$, but we can instead bound $\nu_1(\mathcal{C}_\theta(\mu))$. Consequently, Theorem 12 can be modified by replacing $\mathcal{Q}_\theta(\lambda)$ with $\mathcal{C}_\theta(\mu)$, and $\nu_2(\bar Q)$ with $\nu_1(\mathcal{C}_{\bar\theta}(\bar\mu))$. It only remains to compute the constants from Assumptions 12(i-iii):

(i) $\nu_1(\mathcal{C}_{\bar\theta}(\bar\mu)) = 1$ since $\bar\mu = 0$ and thus $\mathcal{C}_{\bar\theta}(\bar\mu) = \nabla^2(\|y-\bar\theta\|^2) = \mathbb{1}_n$.
(ii) $K = \frac{2}{\sigma_s}$ since $\|\mu_\theta\| \leq \frac{2}{\sigma_s}\|\bar y - \theta\|$ for any $\theta \in \Theta(\bar y)$ by Lemma 11.
(iii) $L = 0$ since $\mathcal{C}_\theta(\bar\mu) = \mathbb{1}_n$ is independent of θ.

□

Proof of Lemma 15. From the definition of the Aubin property it is clear that there exists a neighborhood $U_0 \ni \bar y$ such that $\mathfrak{F}$ has the Aubin property at $\bar p$ for y, for any $y \in U_0 \cap \mathfrak{F}(\bar p)$. We may assume that U_0 is closed. Let $\mathfrak{F}_0 : p \mapsto \mathfrak{F}(p) \cap U_0$, and note that it has closed graph since $\mathfrak{F}$ does. Thus, $\mathfrak{F}_0$ is outer semicontinuous [30, Thm 5.7]). The lemma follows from [30, Thm 9.38]. □

References

[1] C. Aholt, S. Agarwal, and R. Thomas. A QCQP approach to triangulation. In *ECCV (1)*, volume 7572 of *Lecture Notes in Computer Science*, pages 654–667. Springer, 2012.
[2] J.-P. Aubin and H. Frankowska. *Set-valued analysis.* Springer Science & Business Media, 2009.
[3] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear programming: theory and algorithms.* John Wiley & Sons, 2013.
[4] A. Beck and Y. C. Eldar. Strong duality in nonconvex quadratic optimization with two quadratic constraints. *SIAM Journal on Optimization*, 17(3):844–860, 2006.
[5] D. P. Bertsekas. *Nonlinear programming.* Athena Scientific, Belmont (Mass.), 1999.
[6] G. Blekherman, P. Parrilo, and R. Thomas. *Semidefinite optimization and convex algebraic geometry*, volume 13 of *Series on Optimization.* MOS-SIAM, 2013.
[7] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems.* Springer Science & Business Media, 2013.

[8] L. Cayton and S. Dasgupta. Robust Euclidean embedding. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 169–176. ACM, 2006.
[9] M. T. Chu and N. T. Trendafilov. The orthogonally constrained regression revisited. *Journal of Computational and Graphical Statistics*, 10(4):746–771, 2001.
[10] D. Cifuentes and P. A. Parrilo. Sampling algebraic varieties for sum of squares programs. *arXiv preprint arXiv:1511.06751*, 2015.
[11] J. W. Demmel. *Applied Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997.
[12] A. L. Dontchev and R. T. Rockafellar. *Implicit functions and solution mappings: a view from variational analysis*. Springer Monographs on Mathematics. Springer, 2009.
[13] D. Eisenbud. *Commutative algebra: with a view toward algebraic geometry*, volume 150. Springer Science & Business Media, 2013.
[14] A. V. Fiacco and Y. Ishizuka. Sensitivity and stability analysis for nonlinear programming. *Annals of Operations Research*, 27(1):215–235, 1990.
[15] P. Finsler. Über das Vorkommen definiter und semidefiniter Formen in Scharen quadratischer Formen. *Commentarii Mathematici Helvetici*, 9(1):188–192, 1936.
[16] J. Fredriksson and C. Olsson. Simultaneous multiple rotation averaging using Lagrangian duality. In *Asian Conference on Computer Vision*, pages 245–258. Springer, 2012.
[17] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145, 1995.
[18] J. Gouveia, P. Parrilo, and R. Thomas. Theta bodies for polynomial ideals. *SIAM J. Optim.*, 20(4):2097–2118, 2010.
[19] A. G. Hadigheh, O. Romanko, and T. Terlaky. Sensitivity analysis in convex quadratic optimization: simultaneous perturbation of the objective and right-hand-side vectors. *Algorithmic Operations Research*, 2(2):94, 2007.
[20] A. Heyden and K. Åström. Algebraic properties of multilinear constraints. *Mathematical Methods in the Applied Sciences*, 20(13):1135–1162, 1997.
[21] S. Kim and M. Kojima. Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. *Computational Optimization and Applications*, 26(2):143–154, 2003.
[22] J. B. Lasserre. *Moments, Positive Polynomials and their Applications*, volume 1 of *Imperial College Press Optimization Series*. Imperial College Press, London, 2010.
[23] A. B. Levy, R. A. Poliquin, and R. T. Rockafellar. Stability of locally optimal solutions. *SIAM Journal on Optimization*, 10(2):580–604, 2000.
[24] L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information theory*, 25(1):1–7, 1979.
[25] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20–34, 2010.
[26] M. Mevissen and M. Kojima. SDP relaxations for quadratic optimization problems derived from polynomial optimization problems. *Asia-Pacific Journal of Operational Research*, 27(01):15–38, 2010.
[27] B. S. Mordukhovich, R. T. Rockafellar, and M. E. Sarabi. Characterizations of full stability in constrained optimization. *SIAM Journal on Optimization*, 23(3):1810–1849, 2013.
[28] J. Nie and L. Wang. Semidefinite relaxations for best rank-1 tensor approximations. *SIAM Journal on Matrix Analysis and Applications*, 35(3):1155–1179, 2014.
[29] I. Pólik and T. Terlaky. A survey of the S-lemma. *SIAM review*, 49(3):371–418, 2007.
[30] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.
[31] D. Rosen, L. Carlone, A. Bandeira, and J. Leonard. A certifiably correct algorithm for synchronization over the special Euclidean group. In *Intl. Workshop on the Algorithmic Foundations of Robotics (WAFR)*, San Francisco, CA, Dec. 2016.

[32] J. B. Saxe. Embeddability of weighted graphs in k-space is strongly NP-hard. In *Proceedings of the 17th Allerton Conference on Communications, Control, and Computing*, pages 480–489, 1979.
[33] N. Z. Shor. *Nondifferentiable optimization and polynomial problems*, volume 24. Springer Science & Business Media, 2013.
[34] R. J. Stern and H. Wolkowicz. Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations. *SIAM Journal on Optimization*, 5(2):286–313, 1995.
[35] T. Viklands. *Algorithms for the weighted orthogonal Procrustes problem and other least squares problems*. PhD thesis, Umea University, Sweden, 2006.
[36] L. Wang and A. Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference*, page iat005, 2013.
[37] Y. Ye and S. Zhang. New results on quadratic minimization. *SIAM Journal on Optimization*, 14(1):245–267, 2003.
[38] S. Zhang. Quadratic maximization and semidefinite relaxation. *Mathematical Programming*, 87(3):453–465, 2000.

Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge MA 02139, USA
E-mail address: diegcif@mit.edu

Google Inc.
E-mail address: sameeragarwal@google.com

Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge MA 02139, USA
E-mail address: parrilo@mit.edu

Deparment of Mathematics, University of Washington, Seattle WA 98195, USA
E-mail address: rrthomas@uw.edu