Sequential sum-of-squares programming for analysis of nonlinear systems

Torbjörn Cunis and Benoît Legat

Abstract—Numerous interesting properties in nonlinear systems analysis can be written as polynomial optimization problems with nonconvex sum-of-squares problems. To solve these problems efficiently, we propose a sequential approach of local linearizations leading to tractable, convex sum-of-squares problems. We prove local convergence under the assumption of strong regularity, a common condition in variational analysis. The new approach is applied to estimate the region of attraction of a polynomial aircraft model, where it outperforms previous methods for nonconvex sum-of-squares problems.

I. INTRODUCTION

Polynomials that can be written as a sum of squares are a strict subset of the nonnegative polynomials. While determining whether a given polynomial does not assume negative values is computationally hard, Parrillo [1] showed in his seminal paper that convex optimization over sum-of-squares polynomials can be reduced to semidefinite programming. His works, as well as the development of the dual approach via moments by Lasserre [2], and the advent of efficient algorithms for semidefinite problems laid the foundation for numerical analysis of nonlinear systems with polynomial dynamics that is today known as sum-of-squares programming.

Applications of convex sum-of-squares programming include stability verification for hybrid systems [3–7], optimization algorithms and optimization-based control [8–10], control synthesis [11–14], and many more. As these approaches often make use of Lyapunov-type functions and dissipativity inequalities, many sum-of-squares constraints for polynomial dynamics can be viewed as the natural extension of linear matrix inequalities for linear systems [15]. However, unlike in the linear case, most properties of nonlinear systems such as asymptotic stability, invariance, or controllability are often valid on a region of the system’s state-space only.

The problem of determining the region of attraction [16], for example, thus consists of finding a Lyapunov candidate \( V \) and a region \( X \) (often a sublevel set) as well as certifying that \( V \) decays strictly on \( X \). If \( V \) and the describing function of \( X \) are polynomial decision variables, estimating the region of attraction is a nonconvex, nonlinear sum-of-squares problem by the Positivstellensatz of the reals [17].

Despite nonlinear sum-of-squares problems being computationally hard, local analysis of stability and other properties of polynomial dynamics with sum-of-squares programming has been extensively studied [18–28]. Here, the (mostly bilinear) nonconvex constraints have been mitigated by bisections [29], coordinate descent [30], and combinations of both. Yet, except for quasiconvex problems, convergence is not guaranteed (see remarks in [31]). Given that the underlying semidefinite problems scale notoriously with the polynomial degree, it is desirable to limit the number of convex evaluations.

In this paper, we take inspiration from sequential convex programming [32–36] and study a sequential approach for nonlinear conic problems which we combine with a line search using a merit function from Powell [37]. The nonlinear problem is linearized around a solution candidate in order to obtain an affine conic problem. For the sum-of-squares cone, sum-of-squares toolboxes such as SPOT, SOSOPT, or SOSTOOLS are readily available to solve the local problems via reduction to a semidefinite program; and more recently, direct implementations of the sum-of-squares cone have been proposed [38, 39]. Similar to the affine case, nonlinear sum-of-squares problems could directly be reduced to a nonlinear semidefinite program; yet the authors are only aware of the toolbox SUMOFQUADS.JL [40] for that purpose, which is limited to quadratic expressions. Moreover, the semidefinite representation of a sum-of-squares polynomial is nonunique (see comments in [1, Section 3.2]); yet uniqueness of the solution is usually assumed for convergence.

We prove local convergence of the sequence of convex problems using a result from variational analysis [41] that builds upon the implicit function theorem for strongly regular generalized equations by Robinson [42]. As this result is stated for (possibly infinite dimensional) Banach spaces, our analysis works in the general setting of nonlinear conic programs with convex cones embedded in Banach spaces. Since the vector space of polynomials is not complete, we limit ourselves to optimization problems with fixed polynomial degree but our sequential algorithm can be applied to other cones as well. We further investigate the line search based on the dual theory of affine sum-of-squares optimization. Numerical results for practical engineering problems demonstrate that sequential sum-of-squares programming significantly reduces the number of convex problems to be solved and thus the computation time compared to previous, iterative approaches.

Our proof generalizes [35] in two aspects. First, we consider convex cones in arbitrary Banach spaces rather than embedded in \( \mathbb{R}^n \). Second, we show that the convergence still holds if a line search is used to improve convergence speed.
Moreover, by use of variational analysis, our paper provides a simpler proof while obtaining a tighter convergence rate.

The remainder of the paper is organized as follows: Section II introduces the tools from variational analysis and Section III motivates and states the problem of nonlinear sum-of-squares optimization. The sequential programming approach is detailed in Section IV and local convergence is proven in Section V. In Section VI, the sequential approach is applied to problems from nonlinear system analysis.

Notation: $\mathbb{N}$ (resp., $\mathbb{N}_0$) and $\mathbb{R}$ denote the natural numbers excluding (resp., including) zero and the reals, respectively. For some $m \in \mathbb{N}$, the set of symmetric (resp., positive semidefinite) matrices in $\mathbb{R}^{m \times m}$ is $\mathbb{S}_m$ (resp., $\mathbb{S}_m^+$).

II. Preliminaries

Let $X$, $Y$, and $P$ be Banach spaces. The dual space $X^*$ is set of linear operators $l : X \to \mathbb{R}$ is $X^*$ with evaluation $\langle \cdot, \cdot \rangle : X^* \times X \to \mathbb{R}$. Moreover, the adjoint of a linear mapping $A : X \to Y$ is the linear mapping $A^* : Y^* \to X^*$ satisfying $\langle l, A(\xi) \rangle = \langle A^*(l), \xi \rangle$ for all $\xi \in X$ and $l \in Y^*$.

A. Normals & Gradients

A convex cone is a set $C \subset X$ satisfying $r_1 \xi_1 + r_2 \xi_2 \in C$ for all $\xi_1, \xi_2 \in C$ and $r_1, r_2 \in \mathbb{R}_{\geq 0}$. The dual cone of $C$ is defined as

$$ C^* = \{ v \in X^* \mid \langle v, \xi \rangle \geq 0 \text{ for all } \xi \in X \} $$

and the dual of $C^*$ is isometric to $C$. Moreover, for a convex set $\Omega \subset X$, the normal cone mapping $N_\Omega : X \rightrightarrows X^*$ is given by

$$ N_\Omega(\xi_0) = \{ v \in X^* \mid \langle v, \xi - \xi_0 \rangle \leq 0 \text{ for all } \xi \in \Omega \} $$

if $\xi_0 \in \Omega$, and $N_\Omega(\xi_0) = \emptyset$ otherwise.

Definition 1: The (Fréchet) derivative of a nonlinear function $g : X \to Y$ at $\xi_0 \in X$ is a linear mapping $\nabla g(\xi_0) : X \to Y$ satisfying

$$ \lim_{\xi \to \xi_0} \frac{g(\xi) + \nabla g(\xi_0)(\xi - \xi_0) - g(\xi)}{\| \xi - \xi_0 \|} = 0 $$

and $g$ is (Fréchet) differentiable if and only if $\nabla g(\xi_0)$ exists for all $\xi_0 \in X$.

B. Continuity & Regularity

Let $h : X \times P \to Y$ be a function and $H : X \rightrightarrows Y$ be a set-valued mapping. $h$ is said to be Lipschitz continuous with respect to $\xi$ around $(\xi_0, \pi_0)$ in $\text{int dom } h$ if and only if

$$ \lim_{\xi, \xi' \to \xi_0, \xi \neq \xi', \pi \to \pi_0} \frac{\| h(\xi', \pi) - h(\xi, \pi) \|}{\| \xi' - \xi \|} = \kappa $$

with constant $\kappa < \infty$. Moreover, $h$ is said to have a single-valued localization $h^\prime : X' \to Y'$ around $\xi_0 \in X$ for $v_0 \in Y$ if and only if $X' \subset X$ and $X' \subset X$ are neighbourhoods of $\xi_0$ and $v_0$, respectively, $h(\xi_0) = v_0$, and $H(\xi) \cap Y' = \{ h(\xi) \}$ for all $\xi \in X'$. If the inverse $H^{-1} : v \to \{ \xi \in X \mid v \in H(\xi) \}$ has a single-valued localization around $v_0$ for $\xi_0$ that is Lipschitz continuous around $v_0$ with constant $\gamma$, then $H$ is called strongly regular at $\xi_0$ for $v_0$ with constant $\gamma$.

Theorem 1 (Theorem 8.8 of [41]): Take $\psi : X \times P \to Y$ and $N : X \rightrightarrows Y$; suppose $\psi$ is Lipschitz continuous with respect to $\pi$ around $(\xi_0, \pi_0)$ with constant $\kappa$ and $0 \in \psi(\xi_0, \pi_0) + N(\xi_0)$; if there exists $\psi_0 : X \to Y$ such that $\psi_0(\xi) = \psi(\xi, \pi)$ around $\xi_0$ if $\pi \to \pi_0$ and $\psi_0 + N$ is strongly regular at $\xi_0$ for $0$ with constant $\gamma$, then the mapping

$$ H : \pi \mapsto \{ \xi \in X \mid \psi(\xi, \pi) + N(\xi) \ni 0 \} $$

has a Lipschitz continuous, single-valued localization around $\pi_0$ for $\xi_0$ with constant $\gamma \kappa$.

III. Sum-of-Squares Optimization

Let $x = (x_1, \ldots, x_n)$ be a tuple of free variables and $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ a multi-index; a polynomial $\pi$ in $x$ up to degree $d$ is a linear combination

$$ \pi = \sum_{\| \alpha \|_1 \leq d} c_\alpha x^\alpha $$

where $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and $\| \alpha \|_1 = \sum \alpha_i$. The set $\mathbb{R}[x]$ of polynomials in $x$ with real coefficients $c_\alpha \in \mathbb{R}$ up to degree $d$ forms a vector space with norm $\| \cdot \|$. Definition 2: A polynomial $\pi \in \mathbb{R}[x]$ is a sum-of-squares polynomial $(\pi \in \Sigma_d[x])$ if and only if there exist $m \in \mathbb{N}$ and $\pi_1, \ldots, \pi_m \in \mathbb{R}[x]$ such that $\pi = \sum_{i=1}^m \pi_i^2$.

It is easy to see that $\Sigma_d[x]$ forms a convex cone in $\mathbb{R}[x]$. Moreover, its dual cone $\Sigma_d[x]^*$ is isometric to the cone of sum-of-squares polynomials $[2]$.

To avoid confusion with the Fréchet derivative $\nabla$ (with respect to the space of polynomials), we are going to use $\partial_x : \mathbb{R}[x] \to \mathbb{R}[x]^{1 \times n}$ for the differentiation operator with respect to the free variables $x$.

A convex sum-of-squares optimization problem can be reduced to a semidefinite program. This is the fundamental result of [1, Theorem 3.3], which reads as follows; denote by $s_d \in \mathbb{N}_0$ the number of monomials up to degree $d$ of a polynomial in $x$. A polynomial $\pi \in \mathbb{R}[x]$ is sum-of-squares if and only if there exists a matrix $Q \in \mathbb{S}_d^+$ satisfying $\pi = \sum_{i=1}^m \pi_i^2$, where $\pi_1 \in \mathbb{R}[x]^{n \times n}$ is the vector of monomials up to degree $d$. The solution for $Q$ is usually not unique, leading to the implicit (or kernel) and explicit (or image) relaxations of an affine sum-of-squares problem.

A. Motivation

Consider a continuous-time dynamic system defined by the differential equation

$$ \dot{x} = \phi(x) $$

(1)

where $x \in \mathbb{R}^n$ denotes the state vector and $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a polynomial function satisfying $\phi(0) = 0$. Many system-theoretic properties on a domain $D \subset \mathbb{R}^n$ can be written as polynomial dissipativity inequality of the form

$$ \forall x \in \mathbb{R}^n, x \in D \implies \partial_x V(x) \phi(x) \leq S(x) $$

(2)

where $V \in \mathbb{R}[x]$ and $S \in \mathbb{R}[x]$ are called storage function and supply rate, respectively. The set $D$ usually depends on $V$; for example, if $D$ is a sublevel set of $V$ and $V$ and $S$
are positive definite polynomials, then Eq. (2) is LaSalle’s condition for asymptotic stability [43, Theorem 2].

In the sum-of-squares literature, the dissipativity condition is rewritten using the so-called generalized S-procedure [29]. Suppose \( D = \{ x \in \mathbb{R}^n \mid \ell_V(x) \geq 0 \} \) for some \( \ell_V \in \mathbb{R}_d[x] \), then if there exists \( \zeta \in \Sigma_d[x] \) such that
\[
(S - \partial_x V \phi) - \zeta \ell_V \in \Sigma_d[x] \tag{3}
\]
then \( V \) and \( S \) satisfy (2). However, since \( \ell_V \) depends on \( V \), the sum-of-squares constraint (3) is nonlinear. Previous approaches to optimize over nonlinear sum-of-squares constraint relied on solving for one variable at a time while keeping the remaining variables fixed (see, e.g., [29–31]).

B. Nonlinear optimization problem

In general, a polynomial optimization problem with non-convex sum-of-squares constraints takes the form of a nonlinear optimization
\[
\min_{\xi \in X} \langle f, \xi \rangle \quad \text{s.t.} \quad g(\xi) \in D, \quad \xi \in C \tag{4}
\]
where \( X \) and \( Y \) are Banach spaces, \( f \in X^* \) is a linear cost, \( g : X \to Y \) is a differentiable nonlinear constraint mapping, and \( C \subseteq X \) and \( D \subseteq Y \) are convex cones. In the case of sum-of-squares, \( X \) and \( Y \) correspond to spaces of polynomials up to a finite degree with sum-of-squares cones \( C \) and \( D \) and \( g \) takes polynomial values. As nonconvex problem, (4) might have multiple optimal solutions.

Define the Lagrangian as \( L(\xi, l) = \langle f, \xi \rangle - \langle l, g(\xi) \rangle \), where \( l \in Y^* \) is a Lagrange multiplier; the Karush-Kuhn-Tucker (KKT) conditions for (4) at \( \xi_0 \in X \) are
\[
f - \nabla g(\xi_0)^* l - s = 0 \quad (s, \xi_0) = 0, \quad \langle l, g(\xi_0) \rangle = 0 \quad \zeta_0 \in C, \quad g(\zeta_0) \in D
\]
where \( s \in C^* \) and \( l \in D^* \) are dual variables associated with the cone constraints.

Define \( \vartheta = (\xi, l) \) and \( T = X \times Y^* \). With a small abuse of notation, we identify \( g(\xi) \) with \( \ell \in T \) using the canonical isomorphism between \( D \) and \( D^* \). Then the KKT conditions are equivalent to \((-s, -g(\xi)) \) belonging to the normal cone of \( C \times D^* \) at \((\xi_0, l_0) \in T \). The KKT conditions can thus be written as generalized equation
\[
\varphi(\vartheta) + N(\vartheta) \ni 0 \tag{5}
\]
where \( \varphi : T \to T^* \) and \( N : T \to T^* \) are defined as
\[
\varphi : \vartheta \mapsto \langle f - \nabla g(\xi)^* l, g(\xi) \rangle
\]
and
\[
N : \vartheta \mapsto \{ (v, \zeta) \in T^* \mid v \in N_C(\xi), \zeta \in N_D(\ell) \}
\]
respectively. We denote the solutions to (5) by \( \Theta \subseteq T \).

Assumption 1: The set \( \Theta \) is nonempty.

Under a suitable constraint qualification, existence of a KKT point is necessary for an optimal solution of (4).

IV. SEQUENTIAL PROGRAMMING

We propose to approach the nonlinear problem (4) with a sequence of local, convex problems. To that extent, let \( \xi^k \in \mathbb{X} \) with \( k \in \mathbb{N}_0 \) be the solution of the \( k \)-th iteration and \( l^k \in Y^* \) be the associated Lagrange multiplier. Pick tolerances \( \epsilon_k, \zeta_k^* > 0 \) for the primal and dual solutions as well as a small weight \( \eta > 0 \). Our next instance \((\xi^{k+1}, l^{k+1})\) is subject to the steps:

1. Solve the convex problem at \( \xi^k \),
\[
\min_{\xi \in X, l \in Y^*} \langle f, \xi \rangle \tag{6a}
\]
\[
s.t. \quad g(\xi^k) + \nabla g(\xi^k)(\xi - \xi^k) = \zeta^k \tag{6b}
\]
and the optimal solution as \( \xi^+ \) and the associated Lagrange multiplier as \( l^+ \).

2. Solve the line search
\[
\min_{r \in \mathbb{R}} \psi(r) \tag{7}
\]
\[
s.t. \quad 0 < r \leq 1
\]
where \( \psi : r \mapsto L(r\xi^+ + (1-r)\xi^k, l^+), \) and denote the optimal solution as \( r^* \).

3. Set \( \xi^{k+1} = \tilde{r}\xi^+ + (1-\tilde{r})\xi^k \) and \( l^{k+1} = \tilde{r}l^+ + (1-\tilde{r})l^k \). We terminate the iteration if both \( \|\xi^{k+1} - \xi^k\| \leq \epsilon_k \) and \( \|l^{k+1} - l^k\| \leq \zeta_k^* \), where \( \|\ell\| \) denotes the operator norm of \( \ell : Y \to \mathbb{R} \). Otherwise, we repeat the steps for \( k + 1 \).

Remark 1: The merit function \( \psi \) in (7) is a generalization of [37, Eq. (5.13)] for convex cone constraints. Unlike other commonly used merit functions (see, e.g., [44]), \( \psi \) is differentiable and does not require to solve a sum-of-squares program.

As linear problem, (6) has a dual problem at \( \xi^k \), viz.
\[
\max_{l \in D^*, s \in C^*} \langle l, \gamma_k \rangle \tag{8a}
\]
\[
s.t. \quad f - \nabla g(\xi^k)^* l - s = 0 \tag{8b}
\]
and \( l \in D^*, \quad s \in C^* \tag{8c} \)
where \( \gamma_k \equiv \nabla g(\xi^k)^* l - g(\xi^k) \). In the following analysis, we will assume that the Lagrange multiplier \( l^+ \) is the optimal solution of (8), provided it exists. If \( (\xi_0, l_0) \in T \) satisfy the KKT conditions
\[
f - \nabla g(\xi^k)^* l_0 - s = 0 \quad (s, \xi_0) = 0, \quad \langle l_0, \nabla g(\xi^k)\xi_0 - \gamma_k \rangle = 0 \quad (\xi_0, l_0) \in C \times D^*, \quad \nabla g(\xi^k)\xi_0 - \gamma_k \in D
\]
for some \( s \in C^* \), then \( \xi_0 \) and \( l_0 \) are optimal solutions for (6) and (8), respectively, and satisfy \( \langle f, \xi_0 \rangle = \langle l_0, \gamma_k \rangle \).

V. THEORETICAL ANALYSIS

We are going to prove local convergence of the sequential algorithm using a parametrized version of the generalized equation (5); define \( \hat{\varphi} : (\vartheta, \xi^k) \mapsto (f - \nabla g(\xi)^* l, \nabla g(\xi^k)\xi - \gamma_k) \) then \( \vartheta_0 = (\xi_0, l_0) \in \Theta \) if and only if it solves
\[
\hat{L}(\vartheta, \xi^k) = \hat{\varphi}(\vartheta, \xi^k) + N(\vartheta) \ni 0 \tag{9}
\]
at $\xi^k = \xi_0$. In other words, the generalized equation (9) can be understood as linearization of (5) around $\xi^k$. The set of KKT points of (6) at $\xi^k \in X$ is given by

$$H(\xi^k) = \{ \vartheta \in T | \dot{L}(\vartheta, \xi^k) \geq 0 \}$$

the solution map of (9).

The following result is a special case of [45, Theorem 2F.1] for solution mappings of monotone variational inequalities and proved here for completeness.

**Lemma 1:** Let $\xi^k \in X$; if $H(\xi^k)$ is nonempty, then $H(\xi^k)$ is a convex set containing $(\xi_+, l_+)$.

**Proof:** Denote the set of optimal solutions to (6) and (8) by $S \subset X \times Y^*$. Assume that $H(\xi^k)$ is nonempty, take $(\xi_0, l_0) \in H(\xi^k)$ and $(\xi_+, l_+) \in S$. By sufficiency of the KKT conditions, $(\xi_0, l_0) \in S$ and $(\tilde{f}, \tilde{\xi}_0) - (l_0, \gamma_0) = 0$. Since $(\tilde{f}, \xi_+) \leq (\tilde{f}, \xi_0)$ and $(l_+, \gamma_k) \geq (l_0, \gamma_k)$ by primal and dual optimality,

$$0 \leq (f, \xi_+) - (l_+, \gamma_k) = (\nabla g(\xi^k)^* l_+ + s, \xi_+) - (l_+, \gamma_+),$$

and a difference in the second component can be written as

$$g(\xi^k) + \nabla g(\xi^k)(\xi - \xi^k) - \nabla g(\xi^k)(\xi - \xi^k') = [\nabla g(\xi^k) - \nabla g(\xi^k)](\xi - \xi^k) - e(\xi^k, \xi^k')$$

where $e : (\xi^k, \xi^k') \rightarrow g(\xi^k') + \nabla g(\xi^k)(\xi^k - \xi^k') - g(\xi^k)$. By definition of the Fréchet derivative and Lipschitz continuity of $\nabla g(\xi)$ around $\xi_0$, we conclude that

$$\limsup_{\xi^k, \xi^k' \rightarrow \xi_0, \xi^k \neq \xi^k'} \frac{e(\xi^k, \xi^k')}{\| \xi^k - \xi^k' \|} = 0$$

and

$$\limsup_{\xi^k, \xi^k' \rightarrow \xi_0, \xi^k \neq \xi^k'} \frac{\| \nabla g(\xi^k) - \nabla g(\xi^k') \|}{\| \xi^k - \xi^k' \|} = 0.$$

Combining these results we obtain

$$\limsup_{\vartheta \rightarrow \vartheta_0} \frac{\| \varphi(\vartheta, \xi^k) - \varphi(\vartheta, \xi^k') \|}{\| \xi^k - \xi^k' \|} \leq \kappa + 0$$

which is the desired result.

We continue our theoretical analysis by proving that, by strong regularity of $(L(\cdot, \xi^k)$ and Lipschitz continuity of $\varphi_0(\cdot, \cdot)$, the KKT conditions of (6) have a locally unique solution at $\xi^k$ if $\vartheta^k = (\xi^k, l^k)$ is sufficiently close to $\Theta$.

**Proposition 1:** Let $\vartheta_0 = (\xi_0, l_0) \in \Theta$; the solution map $H$ of (9) has a single-valued localization $h : X \rightarrow T$ around $\xi_0$ for $\vartheta_0$; and $h$ is Lipschitz continuous around $\xi_0$ with constant $\gamma \kappa$.

**Proof:** Define $\varphi_0 : \vartheta \mapsto \varphi(\vartheta, \xi_0)$; then $\varphi_0(\vartheta)$ equals $\varphi(\vartheta, \xi^k)$ around $\vartheta_0$ if $\xi^k \rightarrow \xi_0$ by continuity and $\varphi_0 + N = L_0$ is strongly regular with constant $\gamma$ by Assumption 2. Moreover, $\varphi(\vartheta, \xi^k)$ is Lipschitz continuous with respect to $\xi_k$ around $(\vartheta_0, \xi_0)$ with constant $\kappa$ by Lemma 2. By virtue of Theorem 1, there exists a Lipschitz continuous, single-valued localization $h$ of $H$ around $\xi_0$ for $\vartheta_0$ with constant $\gamma \kappa$, the desired result.

In consequence, the convex problems (6) and (8) at $\xi^k$ are not only feasible but have unique solutions around $\xi_0$.

**Lemma 3:** Let $\vartheta_0 = (\xi_0, l_0) \in \Theta$; then $h(\xi^k) = (\xi_+, l_+)$ around $\xi_0$.

**Proof:** By Proposition 1, there exists a single-valued localization $h(\xi^k)$ of $H$ around $\xi_0$: that is, $H(\xi^k)$ is nonempty and, by Lemma 1, a convex set that contains $(\xi_+, l_+)$ around $\xi_0$. On the other hand, $H(\xi^k)$ contains the isolated point $h(\xi^k)$ and thus is a singleton. Hence, $h(\xi^k) = (\xi_+, l_+)$ around $\xi_0$.

It rests to prove that the next iterate, subject to the line search, converges towards a KKT point as well.

**Proposition 2:** Let $(\xi_0, l_0) \in \Theta$; if $\xi^k, \xi_+$, and $l_+$ are sufficiently close to $(\xi_0, l_0)$, then the solution of (7) satisfies $\dot{r} \geq \eta / \kappa$.

**Proof:** If $\dot{r} < 1$, then $\psi'(\dot{r}) - \eta = 0$. The derivative is

$$\psi'(r) = (f - \nabla g(\xi(r))^* l_+ , \xi_+ - \xi^k)$$
where \( \xi(r) \triangleq r\xi^* + (1 - r)\xi_k^* \). Since \((\xi^*, l^*) \in H(\xi_k^*)\),
\[
\psi'(0) = \langle f, \xi^* \rangle - \langle f - \nabla g(\xi_k^*)^* l^*, \xi^* \rangle - \langle \nabla g(\xi_k^*)^* l^*, \xi^* \rangle - \langle s, \xi^* \rangle
\]
\[
= -\langle \nabla g(\xi_k^*)^* l^*, \xi^* - \xi_k^* \rangle - \langle l^*, g(\xi_k^*) \rangle - \langle s, \xi^* \rangle
\]
\[
= -\langle s, \xi^* \rangle - \langle l^*, \xi^* \rangle \leq 0
\]
where the equalities follow from (6) and (8) as well as \( \langle f, \xi^* \rangle = \langle l^*, \gamma_k \rangle \), and the inequality follows from the definition of the dual cone. Since \( \nabla g(\xi^*)^* l^* \) is Lipschitz continuous around \((\xi_0, l_0)\) by Assumption 2, we have that
\[
|\psi'(r) - \psi'(0)| \leq \kappa \| \xi^* - \xi_k^* \|^2 r.
\]
With \( \psi'(0) \leq 0 \) and \( \psi'(\hat{r}) = \eta > 0 \) as well as \( \| \xi^* - \xi_k^* \| \leq \| \xi^* - \xi_0 \| + \| \xi_0 - \xi_k^* \| < 1 \), if \( \hat{r} < 1 \) and \( \xi^* \) and \( \xi_k^* \) are sufficiently close to \( \xi_0 \), we conclude that \( \kappa \hat{r} \geq \eta \).

We combine our results into a local convergence property of the sequential approach.

**Theorem 2:** Let \( \vartheta_0 = (\xi_0, l_0) \in \Theta \) and suppose that \( \gamma \kappa < 1 \); there exists a constant \( \alpha \in (0, 1) \) such that
\[
\| \vartheta^{k+1} - \vartheta_0 \| \leq \alpha \| \vartheta^k - \vartheta_0 \| \tag{10}
\]
if \( \vartheta^k \in \mathcal{T} \) is sufficiently close to \( \Theta \).

**Proof:** Let \( \vartheta^k = (\xi_k, l_k) \); by Proposition 1, there exists a single-valued localization \( h \) of the solution mapping \( H \) around \( \xi_0 \) for \( \vartheta_0 \) satisfying
\[
\| h(\xi^k) - h(\xi^k') \| \leq \gamma \kappa \| \xi^k - \xi^k' \|
\]
for any \( \xi^k, \xi^k' \in X \) in a neighbourhood of \( \xi_0 \). Then \( \vartheta_0 = h(\xi_0) \) as well as \( \vartheta_+ = (\xi_+, l_+) = h(\xi^* \) by Lemma 3 and
\[
\| \vartheta^{k+1} - \vartheta_0 \| = \| \hat{r}(\vartheta_+ - \vartheta_0) + (1 - \hat{r})(\vartheta^k - \vartheta_0) \|
\]
\[
\leq \hat{r} \gamma \kappa \| \xi^* - \xi_0 \| + (1 - \hat{r}) \| \vartheta^k - \vartheta_0 \|
\]
\[
\leq (1 - \hat{r} + \hat{r} \gamma \kappa) \| \vartheta^k - \vartheta_0 \|
\]
as \( \| \xi^k - \xi_0 \| \leq \| \vartheta^k - \vartheta_0 \| \), since \( \hat{r} \geq \omega > 0 \) by Proposition 2, setting \( \alpha = 1 - \omega(1 - \gamma \kappa) \in (\gamma \kappa, 1) \) is the desired result.

Our observations in the proof of Theorem 2 are similar to the assumptions in [35], which proved local convergence of sequential convex programming in the Euclidean space and without a line search. However, if we omit the line search (\( \hat{r} \equiv 1 \)) the rate of convergence we obtain in the proof, namely \( \alpha = \gamma \kappa \), is better than this previous result.

**VI. NUMERICAL EXAMPLES**

The region of attraction of a nonlinear dynamic system \( \dot{x} = \phi(x) \) is defined as the set of initial conditions for which the system trajectories converge to an equilibrium point, here the origin. Estimating the region of attraction is a classical problem in nonlinear systems analysis and a recurrent application of sum-of-squares methods, provided that the dynamics are represented by polynomial equations of motion. One aims to find a polynomial Lyapunov candidate function \( v \in \mathbb{R}^{d_1}[x] \) that decays along trajectories starting in a sublevel set of \( v \), that is, there exists \( \iota > 0 \) such that
\[
\dot{v}(x) = \partial_x v(x) \phi(x) < 0 \tag{11}
\]
for all \( x \neq 0 \) satisfying \( v(x) \leq \iota \). If (11) is satisfied, then \( \{ x \in \mathbb{R}^n \mid v(x) \leq \iota \} \) is an invariant subset of the region of attraction [43, Theorem 2].

In order to lower bound the volume of the region of attraction estimate, a polynomial shape \( p \in \mathbb{R}[x] \) is introduced. For any \( \iota > 0 \), the nonlinear sum-of-squares problem is given as
\[
\min_{v \in \mathbb{R}^{d_1}[x], c \in \mathbb{R}^{d_1}[x], v \in \mathbb{R}^{d_1}[x]} \{ -b \}
\]
\[
s_{2}(v - \iota) - \partial_x v \phi - \varrho \in \mathbb{S}_{d_1}[x] \]
\[
s_{1}(p - b) - v + \iota \in \mathbb{S}_{d_1}[x] \]
\[
v - \varrho \in \mathbb{S}_{d_1}[x] \tag{12}
\]
where \( \varrho \) is small, positive-definite polynomial; we choose \( \varrho = 10^{-6} \sum_{i=1}^{n} x_i^2 \). Eq. (12) has been considered for region of attraction estimation [20, 31, 34, 46, 47] as well as, with minor modifications, for reachability [11], control synthesis [11, 20], or robust stability [22, 26]. In these works, the nonlinearities are split into an iteration of convex and quasiconvex problems via coordinate descent.

We have applied sequential sum-of-squares programming to estimate the region of attraction of the short-period (two states) and longitudinal (four states) dynamics of an airplane. The equations of motion \( \phi \) are cubic polynomials in the two-state case and quintic polynomials in the four-state case; we fix \( \iota \equiv 1 \); and solve for quadratic and quartic Lyapunov functions in both cases. The degrees of the decision variables \( s_{1} \) and \( s_{2} \) as well as number of iterations, computation time, and final value of the objective are detailed in Tab. I. The initial guess for \( v \) has been the quadratic Lyapunov function of the linearized dynamics; \( b = 1 \); and the variables \( s_{1} \) and \( s_{2} \) have been initialized to homogeneous polynomials given in the appendix. The computations have been terminated once the change in the primal variables was below a absolute tolerance \( \epsilon_k \equiv 10^{-6} \) and the change of the dual variables was below a relative tolerance \( \epsilon_k \equiv 10^{-6} \| b \|_\ast \).

For comparison, running the iterative approach of [31] took considerably longer without reaching the same final values (Tab. II); the differences are particularly noticeable for quartic Lyapunov functions. The computationally most expensive part of both algorithms are the semidefinite relaxations of the convex sum-of-squares subproblems. As shown

| Dynamics | \( d_0 \) | \( d_1 \) | \( d_2 \) | Final value | Iterations | Time |
|----------|--------|--------|--------|------------|-----------|-----|
| Short-period | 2 | 0 | 2 | -1.515 | 8 | 2.77 s |
| 4 | 2 | 4 | -1.772 | |
| Longitudinal | 2 | 0 | 4 | -0.354 | 20 | 9.28 s |
| 4 | 2 | 4 | -2.788 | 16 | 25.24 s |

1Terminated early due to numerical issues of Mosek; the nonlinear sum-of-squares constraint was satisfied with a tolerance of \( 1.52 \times 10^{-5} \).

2See [31] and its appendix for details.
TABLE II
COMPUTATION DETAILS OF THE ITERATIVE APPROACH OF [31] FOR REGION-OF-ATTraction ESTIMATION.

| Dynamics  | \(d_0\) | \(d_1\) | \(d_2\) | Final value | Iterations | Time  |
|-----------|---------|---------|---------|-------------|------------|-------|
| Short-period | 2       | 0       | 2       | -1.514      | 20         | 10.94 s |
|           | 4       | 2       | 4       | -1.761      | 22         | 26.25 s |
| Longitudinal | 2      | 0       | 4       | -0.333      | 22         | 113.21 s |
|           | 4       | 2       | 4       | -2.749      | 66         | 465.50 s |

in [48, Section 6.6.3], in order to solve an semidefinite program of \(M\) equality constraints and a positive semidefinite matrix with a block diagonal structure with blocks of size \(N_i\), a second-order interior point solver will have a number of iterations scaling with \(\sqrt{\sum N_i}\) and the number of arithmetic operations per iteration scaling with \(M(M^2 + M \sum N_i^2 + \sum N_i^3)\). As the polynomial in (3) will have the largest degree over all Sum-of-Squares constraints, it will correspond to the larger block in the semidefinite program. The size of the block will be

\[
\left(\frac{n + d/2}{d/2}\right)
\]

where \(N\) is the number of variables and \(d\) is the degree of the left-hand side of (3) which is at least the degree of the storage function added with the degree of \(\phi\).

Fig. 1 details how the total computation time to solve the semidefinite programs in each iteration grows with the number of states and degree of polynomials for the region of attraction estimation. In the iterative approach, in each convex program, either the Lyapunov function or the level sets are fixed which slightly decreases \(M\) or removes small blocks in the semidefinite program. However, since it does not affect the size of the block corresponding to (3), the semidefinite programs are not considerably easier to solve than the semidefinite programs solved by the sequential approach. Therefore, the total time depends mostly on the number of semidefinite programs solved in each approach. Besides, the iterative approach of [31] includes quasiconvex programs, which are solved through bisection, at each iteration. In addition to increasing the number of semidefinite programs to solve, this bisection requires solving semidefinite programs that are close to infeasible which might induces numerical issues. Overall, the computation time for the sequential approach in each iteration is about ten times lower than the computation time for the iterative approach.

VII. CONCLUSIONS

This paper studies the solution of nonlinear convex programs by a sequential algorithm with the addition of a line search. Theorem 2 shows that, under standard regularity assumptions, the algorithm converges linearly with and without the line search. Further investigation of the line search is left for future work.

While the paper analyses the sequential algorithms on arbitrary (possibly infinite dimensional) Banach spaces, we only apply it to the finite-degree sum-of-squares cone of polynomials with finite degree. In [39], the authors shows how to leverage special properties of the sum-of-squares cone for solving the nonconvex Burer-Monteiro formulation.

We are currently investigating whether such refined analysis could allow the sequential algorithm to exploit the structure of the sum-of-squares cone as well.

APPENDIX

For the computations detailed in Tab. I, the following initializations of the multipliers \(s_1\) and \(s_2\) were used:

\[
\begin{align*}
\text{n} & = 2, \quad d_0 = 2, \quad s_1 = 1, \quad s_2 = x_1^2 + x_2^2 \\
\text{n} & = 2, \quad d_0 = 4, \quad s_1 = x_1^2 + x_2^2, \quad s_2 = x_1^2 + x_2^2 \\
\text{n} & = 4, \quad d_0 = 2, \quad s_1 = 1, \quad s_2 = \sum_{i=1}^{n} x_i^2 \\
\text{n} & = 4, \quad d_0 = 4, \quad s_1 = \sum_{i=1}^{n} x_i^2, \quad s_2 = (\sum_{i=1}^{n} x_i^2)^2
\end{align*}
\]

where \(n\) is the number of states and \(d_0\) the degree of the Lyapunov function.

All numerical examples were performed on a 2.8 GHz quad-core Intel Core i7 processor with 16 GB of memory. We use SOSOPT for the convex sum-of-squares problem and MOSEK for semidefinite programming. Source code is available at https://github.com/tcunis/bisosprob.

ACKNOWLEDGMENT

The authors wish to thank Frank Permenter for initiating their collaboration.

REFERENCES

[1] P. A. Parrilo, “Semidefinite programming relaxations for semialgebraic problems,” Mathematical Programming, Series B, vol. 96, no. 2, pp. 293–320, 2003.
[2] J. B. Lasserre, “Global Optimization with Polynomials and the Problem of Moments,” SIAM Journal on Optimization, vol. 11, no. 3, pp. 796–817, 2001.
[3] S. Prajna and a. Papachristodoulou, “Analysis of switched and hybrid systems - beyond piecewise quadratic methods,” Proceedings of the 2003 American Control Conference, 2003., vol. 4, no. Xi, pp. 2779–2784, 2003.
[4] A. A. Ahmadi and P. A. Parrilo, “Non-monotonic Lyapunov functions for stability of discrete time nonlinear and switched systems,” in 47th IEEE Conference on Decision and Control, 2008, pp. 614–621.
[5] A. Papachristodoulou and S. Prajna, “Robust Stability Analysis of Nonlinear Hybrid Systems,” IEEE Transactions on Automatic Control, vol. 54, no. 5, pp. 1035–1041, 2009.
[6] H. Ichihara, “Sum of Squares Based Input-to-State Stability Analysis of Polynomial Nonlinear Systems,” SICE Journal of Control, Measurement, and System Integration, vol. 5, no. 4, pp. 218–225, 2012.
Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.