Mathematics of Control, Signals, and Systems
https://doi.org/10.1007/s00498-023-00370-5

ORIGINAL ARTICLE

Estimates of the size of the domain of the implicit function theorem: a mapping degree-based approach

Ashutosh Jindal[1] · Debasish Chatterjee[1] · Ravi Banavar[1]

Received: 12 September 2022 / Accepted: 6 September 2023

Abstract
In this article, we present explicit estimates of size of the domain on which the implicit function theorem and the inverse function theorem are valid. For maps that are twice continuously differentiable, these estimates depend upon the magnitude of the first-order derivatives evaluated at the point of interest, and a bound on the second-order derivatives over a region of interest. One of the key contributions of this article is that the estimates presented require minimal numerical computation. In particular, these estimates are arrived at without any intermediate optimization procedures. We then present three applications in optimization and systems and control theory where the computation of such bounds turns out to be important. First, in electrical networks, the power flow operations can be written as quadratically constrained quadratic programs, and we utilize our bounds to compute the size of permissible power variations to ensure stable operations of the power system network. Second, the robustness margin of positive-definite solutions to the algebraic Riccati equation (frequently encountered in control problems) subject to perturbations in the system matrices is computed with the aid of our bounds. Finally, we employ these bounds to provide quantitative estimates of the size of the domains for feedback linearization of discrete-time control systems.

Keywords Implicit function theorem · Numerical analysis · Degree theory · Nonlinear analysis

✉ Ashutosh Jindal
ashutosh.1@sc.iitb.ac.in

Debasish Chatterjee
dchatter@iitb.ac.in

Ravi Banavar
banavar@iitb.ac.in

[1] Systems and Control Engineering, Indian Institute of Technology, Bombay, IIT Area, Powai, Mumbai, Maharashtra 400076, India

Published online: 26 September 2023

1 Introduction

The *implicit function theorem* (ImFT) and its conjoined twin the *inverse function theorem* (IFT) constitute a cornerstone of mathematical analysis and multivariate calculus [1]. These theorems serve as the basis of several existential results in mathematics and find applications in the following areas: optimization [2–4], numerical analysis, [5, 6], control theory applications [7], ordinary differential equation [8, 9], etc. In control theory applications specifically, the ImFT allows us to show the existence of solutions of ordinary differential equations [8, 9], to assert the existence of a unique trajectory for the dynamical system, and also continuity properties of the solutions with respect to initial conditions and the control input. In particular, it appears as the driving engine behind several assertions concerning the so-called *end point map* [10, Chapter 2, 3]. ImFT also finds applications in areas such as feedback linearization of numerically discretized systems [7].

The assertions made by these theorems are of existential and local character, i.e., they hold only in a sufficiently small neighborhood around the point of interest. This feature extends to the methods developed based on these results. For instance, the well-known Frobenius's theorem[1] utilizes the IFT to compute a coordinate system that rectifies a given involutive distribution [11], and the region of such rectification is only qualitatively available in general. The ImFT does not provide us with information on the size of the domain on which these results are valid. In engineering applications, having a quantitative estimate of the size of the domain is useful and often crucial. Such a quantitative analysis of the ImFT and the IFT with explicit bounds on the sizes of the respective domains on which the theorems are applicable would be of use in many existential results in the broad area of cybernetics, especially where estimates of robustness are essential. This need serves as our chief motivation to arrive at estimates of the domain of the validity of the ImFT and the IFT.

Despite the wide applicability of the ImFT and IFT, few attempts have been made on estimating such quantitative bounds of the domains of validity of these theorems. One set of such estimates for the ImFT was provided by Holtzman [12], and the estimates there are explicit functions of the bounds on the first-order derivatives of the underlying map over a given domain of interest. Chang et al. [13] provided another set of estimates on the neighborhoods involved in ImFT based on the application of the Roche theorem [14]. These estimates were first provided for the scalar case and then applied inductively to generalize it for vector-valued maps. The bounds provided by Chang et al. [13] were based on the boundedness of the underlying complex map over a bounded domain. The accuracy of these estimates is contingent on the accuracy with which one computes the bound of the underlying map over the domain of interest and for vector-valued maps, the bounds were calculated component-wise using induction. The method needs computation of $(m-1) \times (m-1)$ sub-determinants of a certain Jacobian matrix, where m is the dimension of the underlying vector space, and since the bounds are an explicit function of these subdeterminants, they must be computed over all possible minors, due to which the method performs poorly with an increase

[1] Here we refer to the local version of the theorem; however, a global version of the Frobenius theorem also exists.

in the dimension. In the context of the ImFT for Lipschitz continuous maps, a third set of bounds based on the bounds on the generalized derivatives of the underlying map computed over a domain of interest was provided by Papi [15] for the ImFT for Lipschitz continuous functions. Although the bounds provided by Papi [15] cover a larger class of functions than those covered by Chang et al. [13], to compute these bounds one needs to ensure that the generalized derivatives of the underlying map evaluated at the point of interest are invertible, and the accuracy of the estimate is dependent on how tightly the bounds on the generalized derivatives are computed over a given domain.

For the IFT, Abraham et al. [16] provided a set of bounds based on the magnitude of the first-order derivatives evaluated at the point of interest and the boundedness of the second-order derivatives over a domain. The boundedness of the second-order derivative restricts the application of these bounds to functions that are continuously differentiable at least up to second order. These bounds require minimal numerical computations since one only needs to compute the bounds on the second-order derivatives and thus are useful in situations where limited computation capacity is available and/or coarse structural information about the maps is at hand.

In this article, we utilize the topological degree to compute estimates of the neighborhoods given by the IFT and ImFT. Degree theoretic methods are by nature flexible, depending on a few basic properties of the underlying maps, and require minimal assumptions on the premise; in particular, one only requires continuity of various maps over the underlying sets, although for computational ease one can utilize higher-order differential properties. The technique is inherently topological and permits appreciable flexibility in terms of appropriate homotopies (as we shall see in the sequel). This makes the analysis applicable to a relatively large class of functions. Although the key results presented in this article are given for $\mathcal{C}^2$ functions, these bounds are further generalized for $\mathcal{C}^1$ (see Proposition 3.7 in Sect. 3) and $\mathcal{C}^0$ functions (see Proposition A.2 in Appendix A.2).

1.1 Contributions

(i) We utilize the topological degree to compute lower bounds on the domain of validity of the ImFT. Note that our objective here is to arrive at bounds that require minimal numerical computation. For $\mathcal{C}^2$ maps, our bounds are dependent on the first-order derivatives evaluated at the point of interest and second-order derivatives on a bounded set containing the point of interest. These estimates are reported in Theorem 3.3.

(ii) Using Theorem 3.3, in Corollary 3.5, we derive estimates on the size of the domain of validity of the IFT. For finite-dimensional spaces, our estimates improve on the estimates given by Abraham et al. [16, Proposition 2.5.6].

(iii) We also demonstrate that several existing results on the estimates on the domain of applications of IFT and ImFT can also be derived by degree theoretic methods. In particular, we show that when restricted to finite-dimensional vector spaces, the bounds given by Holtzman [12] (see Proposition 3.7 in Sect. 3) and Abraham et al. [16] (see Proposition A.1 in Appendix A.1 can also be derived by elementary application of the mapping degree.

(iv) As the last theoretical contribution of this article, in Proposition A.2, we extend these bounds to the generalized ImFT for continuous maps. (However, uniqueness and regularity properties cannot be asserted in this framework.)

(v) We present two key applications of our bounds: As our first application, we investigate the robustness of the solutions of Quadratically Constrained Quadratic Problems. In this setting, we present two examples: one is the robustness margin for the stable operation of a power system network, and the other is the robustness of the solutions of the algebraic Riccati equation, a nonlinear algebraic equation, frequently encountered in optimal control theory. In the second application, we utilize our bounds to estimate the domain on which a given discrete-time control system is feedback linearizable.

Finally, through this work we also intend to draw the attention of the readers to the gamut of tools offered by the mapping degree theory; we selected the ImFT and IFT because of their central importance in much of control theory.

1.2 Notations

We use standard notations throughout the article. The set of real numbers is denoted by $\mathbb{R}$, and the set of integers is denoted by $\mathbb{Z}$. The set of positive integers is denoted by $\mathbb{N}$, and the set of all positive integers less than or equal to n is denoted by $[n]$. For $U \subset \mathbb{R}^n$ a nonempty set, $\operatorname{cl} U$ denotes the smallest closed set containing U called the closure of U; $\operatorname{int} U$ is the largest open set contained in U called the interior of U; and $\operatorname{bd} U = (\operatorname{cl} U) \setminus \operatorname{int} U$ denotes the boundary of U. For a given $r > 0$ and $x_0 \in \mathbb{R}^n$,

$$\mathsf{B}\,(x_0, r) = \{x \in \mathbb{R} \mid \|x - x_0\| < r\}$$

denotes the open ball of radius r centered at x_0, for some fixed well-defined norm $\|\cdot\|$ on $\mathbb{R}^n$. The identity matrix of order n is denoted by I_n; if the order is unambiguous from the expression, we may drop the subscript and simply write I. For given $n, m \in \mathbb{N}$, for $A \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ a linear map, $\|A\|$ is defined by

$$\|A\| = \sup\left\{\|Ax\| \mid x \in \operatorname{cl}\mathsf{B}\,(0, 1) \subset \mathbb{R}^n\right\}.$$

For a given $\nu \in \mathbb{N}$ and nonempty set U and V, $\mathcal{C}^\nu(U, V)$ denotes the class of ν-*times continuously differentiable maps* with domain and co-domain as U and V, respectively. If U and V are unambiguous from the context, we may simply write $\mathcal{C}^\nu$. The differential operator is denoted by D. For a given $f \in \mathcal{C}^\nu(U, \mathbb{R}^m)$, mapping $\mathbb{R}^n \supset U \in x \longmapsto f(x) \in \mathbb{R}^m$, $x_0 \in U$, $\mathrm{D}f(x_0)$ is the Jacobian matrix of f evaluated at x_0. The partial derivatives are denoted by adding subscript to the differential operator, i.e., $\frac{\partial}{\partial x} =: \mathrm{D}_x$.

2 Degree theory: preliminaries

In order to derive the estimates for the domains of the ImFT and IFT, we utilize several results from topological degree theory. This section serves as a rapid refresher on the topological degree. For a detailed study, one may look into [17–19].

Definition 2.1 (*Topological Degree* [18, Chapter IV, Proposition and Definition 1.1]) Let $U \subset \mathbb{R}^n$ be a nonempty and bounded open set and $f\colon \operatorname{cl} U \longrightarrow \mathbb{R}^n$ be a smooth map. Suppose $0 \in \mathbb{R}^n \backslash f(\operatorname{bd} U)$ is a regular value of f. Then, $X_f^{\mathrm{eq}} := f^{-1}(0)$ is finite (possibly empty), and we define the degree of f by

$$\operatorname{Deg}(f, U) = \sum_{x \in X_f^{\mathrm{eq}}} \operatorname{sgn}(\det(\mathrm{D} f(x))) \tag{2.1}$$

where $\mathbb{R}\backslash\{0\} \ni z \longmapsto \operatorname{sgn}(z) \in \{-1, 1\}$ is defined as:

$$\operatorname{sgn}(z) = \begin{cases} +1 & z > 0, \\ -1 & z < 0. \end{cases}$$

Definition 2.2 (*Topological Degree* (*for continuous maps*) [18, Chapter IV, Proposition and Definition 2.1]) Let $U \subset \mathbb{R}^n$ be nonempty, bounded and open. Let $f\colon \operatorname{cl} U \longrightarrow \mathbb{R}^n$ be a continuous map. Suppose $0 \in \mathbb{R}^n \backslash f(\operatorname{bd} U)$, then there exists a smooth mapping $g\colon \operatorname{cl} U \longrightarrow \mathbb{R}^n$ such that 0 is a regular value of g and for all $x \in \operatorname{bd} U$, $\|f(x) - g(x)\| < \|g(x)\|$. For all such g's, $\operatorname{Deg}(g, U)$ is the same, and we define

$$\operatorname{Deg}(f, U) := \operatorname{Deg}(g, U). \tag{2.2}$$

2.1 Properties of Deg (f, U)

Let U and f be as in Definition 2.2, then $\operatorname{Deg}(f, U)$ satisfies the following properties.

(**DEG**-a) Negative Existence Principle [18, Chapter 4 Corollary 2.5]: If f is nonvanishing on $\operatorname{cl} U$, i.e., $X_f^{\mathrm{eq}} = \emptyset$, then $\operatorname{Deg}(f, U) = 0$.

Remark 2.3 An immediate consequence of (**DEG**-a) is: $\operatorname{Deg}(f, U) \neq 0$ implies $X_f^{\mathrm{eq}} \neq \emptyset$, i.e., there exists at least one $x \in U$ such that $f(x) = 0$.

(**DEG**-b) Additivity [18, Chapter 4 Corollary 2.5]: Let $U_1, U_2 \subset U$ be nonempty and bounded open sets. Suppose $U_1 \cap U_2 = \emptyset$ and f is non vanishing on $U \backslash \operatorname{cl}(U_1 \cup U_2)$, then

$$\operatorname{Deg}(f, U) = \operatorname{Deg}(f, U_1) + \operatorname{Deg}(f, U_2).$$

Definition 2.4 (*Nonvanishing Homotopy*) Let $U \subset \mathbb{R}^n$ be bounded and open. Let $[0, 1] \times \operatorname{cl} U \ni (t, x) \longmapsto H(t, x) \in \mathbb{R}^n$ be continuous. Moreover, for each $t \in [0, 1]$,

$0 \in \mathbb{R}^n \backslash H(t, \operatorname{bd} U)$, then H defines a nonvanishing homotopy on $\operatorname{bd} U$. Two given continuous maps $f_1\colon \operatorname{cl} U \longrightarrow \mathbb{R}^n$ and $f_2\colon \operatorname{cl} U \longrightarrow \mathbb{R}^n$ are said to be nonvanishingly homotopic on $\operatorname{bd} U$ if there exists a nonvanishing homotopy H such that $H(0, \cdot) = f_1$ and $H(1, \cdot) = f_2$.

(**DEG**-c), Homotopy Invariance [18, Chapter 4 Proposition 2.4]: Let $f_1 : \operatorname{cl} U \longrightarrow \mathbb{R}^n$ and $f_2 : \operatorname{cl} U \longrightarrow \mathbb{R}^n$ be two continuous maps. Suppose f_1 and f_2 are nonvanishingly homotopic on $\operatorname{bd} U$ then

$$\operatorname{Deg}(f_1, U) = \operatorname{Deg}(f_2, U).$$

(**DEG**-c) plays an important role in computing the degree of arbitrary maps. The standard approach is as follows: for a given f, find an f_1 such that $\operatorname{Deg}(f_1, U)$ is known apriori or easily computable, and f and f_1 are nonvanishingly homotopic on $\operatorname{bd} U$. Use (**DEG**-a) along with (**DEG**-c) to comment on the existence of the equilibrium points of f on U.

We now state two key results for two functions to be nonvanishingly homotopic; these results will be useful in proving the main contributions of this article.

Theorem 2.5 (Poincaré–Bohl) [20, Chapter 2, Theorem 2.1] *Let $U \subset \mathbb{R}^n$ be nonempty, open and bounded and f_1, f_2 be two continuous maps on $\operatorname{cl} U$ with $f_1(x) \neq 0$, and $f_2(x) \neq 0$ for all $x \in \operatorname{bd} U$. Suppose, for no $x \in \operatorname{bd} U$, $f_1(x)$ and $f_2(x)$ are* anti-parallel, *i.e.,*

$$\left\langle \frac{f_1(x)}{\|f_1(x)\|}, \frac{f_2(x)}{\|f_2(x)\|} \right\rangle \neq -1 \quad \textit{for all } x \in \operatorname{bd} U.$$

Then, f_1 and f_2 are nonvanishingly homotopic on $\operatorname{bd} U$ and the underlying homotopy is:

$$[0, 1] \times \operatorname{bd} U \ni (t, x) \longmapsto H(t, x) := t f_1(x) + (1 - t) f_2(x) \in \mathbb{R}^n.$$

Moreover, one has

$$\operatorname{Deg}(f_1, U) = \operatorname{Deg}(f_2, U).$$

Corollary 2.6 [20, Chapter 2, Theorem 2.3] *Let $U \subset \mathbb{R}^n$ be nonempty, open and bounded and $f_1 : U \longrightarrow \mathbb{R}^n$, $f_2 : U \longrightarrow \mathbb{R}^n$ be two continuous maps satisfying*

$$\|f_1(x) - f_2(x)\| < \|f_1(x)\| \quad \textit{for all } x \in \operatorname{bd} U. \tag{2.3}$$

Then, f_1 and f_2 are nonvanishingly homotopic on $\operatorname{bd} U$ with

$$\operatorname{Deg}(f_1, U) = \operatorname{Deg}(f_2, U).$$

Moreover, if $0 \notin f_1(\operatorname{bd} U)$ and $0 \notin f_2(\operatorname{bd} U)$, then one can replace (2.3) with

$$\|f_1(x) - f_2(x)\| \leq \|f_1(x)\| \quad \textit{for all } x \in \operatorname{bd} U,$$

and the assertion still holds.

3 Estimates for the IFT and ImFT

The ImFT and IFT find applications in proving several existential results in mathematical analysis and also provide the basis for several engineering algorithms. These theorems have a rich history and have been studied with various degree of generalizations by several authors in various sources [1, 16, 17, 21–24]. We provide prototypical versions of the ImFT and IFT below:

Theorem 3.1 (Implicit Function Theorem) [17, Theorem 4.B] *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open and $U \times V \ni (x, y) \longmapsto f(x, y) \in \mathbb{R}^m$ be a $\mathcal{C}^\nu$ map where $\nu \geq 1$. Suppose $(x_0, y_0) \in U \times V$ is such that $\mathrm{D}_y f(x_0, y_0) \colon \mathbb{R}^m \longrightarrow \mathbb{R}^m$ is an isomorphism. Then for any neighborhood $\mathcal{O}(y_0)$ of y_0, there is a neighborhood $\mathcal{O}(x_0)$ of x_0 and a $\mathcal{C}^\nu$ map $g \colon \mathcal{O}(x_0) \longrightarrow \mathcal{O}(y_0)$ satisfying $f(x, g(x)) = w_0 := f(x_0, y_0)$ for all $x \in \mathcal{O}(x_0)$. Furthermore, we have*

$$\mathrm{D}g(x) = -(\mathrm{D}_y f(x, g(x))^{-1}\mathrm{D}_x f(x, g(x)).$$

Theorem 3.2 (Inverse Function Theorem) [16, Theorem 2.5.2] *Let $U \subset \mathbb{R}^n$ be a nonempty open set. Let $U \ni x \longmapsto f(x) \in \mathbb{R}^n$ be a $\mathcal{C}^\nu$ map with $\nu \geq 1$. Suppose $x_0 \in U$ is such that $\mathrm{D}f(x_0) \colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is an isomorphism. Then, there exists a neighborhood $\mathcal{O}$ of x_0 such that $f|_{\mathcal{O}}$ is a $\mathcal{C}^\nu$* diffeomorphism *to its image. Moreover, for $(x, y) \in \mathcal{O} \times f(\mathcal{O})$ satisfying $y = f(x)$, one has*

$$\mathrm{D}f^{-1}(y) = (\mathrm{D}f(x))^{-1}.$$

Theorems 3.1 and 3.2 contain assertions about the existence of certain neighborhoods, but they do not give any information about the sizes of these neighborhoods. Our objective is to arrive at a lower bound on the sizes of these neighborhoods. For this, we use results from degree theory cataloged in Sect. 2. The following theorem provides one set of lower bounds on the sizes of $\mathcal{O}(x_0)$ and $\mathcal{O}(y_0)$ defined in Theorem 3.1.

Theorem 3.3 *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open and $U \times V \ni (x, y) \longmapsto f(x, y) \in \mathbb{R}^m$ be as in Theorem 3.1 and in addition let f be $\mathcal{C}^2$. Define*

$$M_y := \left\| (\mathrm{D}_y f(x_0, y_0))^{-1} \right\|, \quad \textit{and} \quad L_x := \|\mathrm{D}_x f(x_0, y_0)\|.$$

For given $R_x > 0$, and $R_y > 0$ such that $\mathsf{B}(x_0, R_x) \subset U$ and $\mathsf{B}\left(y_0, R_y\right) \subset V$ define

$$\begin{aligned} K_{xx} &:= \sup\left\{ \left\| \mathrm{D}_x^2 f(x, y) \right\| \mid (x, y) \in \mathsf{B}(x_0, R_x) \times \mathsf{B}\left(y_0, R_y\right) \right\}, \\ K_{xy} &:= \sup\left\{ \left\| \mathrm{D}_x \mathrm{D}_y f(x, y) \right\| \mid (x, y) \in \mathsf{B}(x_0, R_x) \times \mathsf{B}\left(y_0, R_y\right) \right\}, \quad \textit{and} \\ K_{yy} &:= \sup\left\{ \left\| \mathrm{D}_y^2 f(x, y) \right\| \mid (x, y) \in \mathsf{B}(x_0, R_x) \times \mathsf{B}\left(y_0, R_y\right) \right\}. \end{aligned}$$

Then, for all $0 < r_x < R_x$ *and* $0 < r_y < R_y$ *satisfying*

$$\frac{1}{2}K_{xx}r_x^2 + K_{xy}r_x r_y + \frac{1}{2}K_{yy}r_y^2 < \frac{r_y}{M_y} - r_x L_x \quad \textit{and} \tag{3.1a}$$

$$K_{xy}r_x + K_{yy}r_y < \frac{1}{M_y}, \tag{3.1b}$$

for each $x \in \mathsf{B}\,(x_0, r_x)$*, there exists a unique* $y_x \in \mathsf{B}\,(y_0, r_y)$ *satisfying* $f(x, y_x) = w_0$*. Furthermore, the map* $\mathsf{B}\,(x_0, r_x) \ni x \longmapsto y_x =: g(x) \in \mathsf{B}\,(y_0, r_y)$ *is* $\mathcal{C}^2$ *and*

$$\mathrm{D}g(x) = -(\mathrm{D}_y f(x, y_x))^{-1}\mathrm{D}_x f(x, y_x).$$

Before we prove Theorem 3.3, we state the finite-dimensional version of Lemma 2.5.4 from [16] which shall be of use in proving Theorem 3.3 and several results to follow.

Lemma 3.4 [16, Lemma 2.5.4] *Let* $\mathrm{GL}(n, \mathbb{R})$ *be the set of all linear isomorphisms on* $\mathbb{R}^n$*. Then,* $\mathrm{GL}(n, \mathbb{R})$ *is open. Moreover, for any* $M \in \mathrm{GL}(n, \mathbb{R})$ *and a linear map* $B : E \to F$ *with* $\|B\| < 1/\|M\|$, $M + B \in \mathrm{GL}(n, \mathbb{R})$.

Proof of Theorem 3.3 Assume K_{xx}, K_{xy} and L_x are not all zero. Let r_x and r_y satisfy (3.1). Fix an arbitrary $x \in \mathsf{B}\,(x_0, r_x)$. Our objective here is to show the existence of a unique $y_x \in \mathsf{B}\,(y_0, r_y)$ satisfying $f(x, y_x) = w_0$. Keeping x fixed, we define

$$V \ni y \longmapsto F(y; x) := f(x, y) - w_0 \in \mathbb{R}^m. \tag{3.2}$$

The problem of finding some y_x satisfying $f(x, y_x) = w_0$ is now equivalent to finding y_x such that $F(y_x; x) = 0$. The proof is set into two parts.
Existence: For ease of representation, we define $\tilde{x} := x - x_0$ and $\tilde{y} := y - y_0$. Define an affine approximation of F about (x_0, y_0) as

$$\begin{aligned} y \longmapsto \mathrm{Aff}(F)(y; x) := {} & f(x_0, y_0) + \mathrm{D}_x f(x_0, y_0)\tilde{x} \\ & + \mathrm{D}_y f(x_0, y_0)\tilde{y} - w_0 \in \mathbb{R}^m. \end{aligned} \tag{3.3}$$

Claim 1: $F(\cdot; x)$ *and* $\mathrm{Aff}(F)(\cdot; x)$ *are nonvanishingly homotopic on* $\mathrm{bd}\,\mathsf{B}\,(y_0, r_y)$: From the definition of $\mathrm{Aff}(F)(\cdot; x)$, we have

$$\begin{aligned} F(y; x) - \mathrm{Aff}(F)(y; x) &= f(x, y) - \big(f(x_0, y_0) \\ &\qquad + \mathrm{D}_x f(x_0, y_0)\tilde{x} + \mathrm{D}_y f(x_0, y_0)\tilde{y}\big) \\ &= (f(x, y) - f(x_0, y_0)) - \mathrm{D}f(x_0, y_0) \cdot \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} \end{aligned}$$

Using the fundamental theorem of calculus [25, Theorem 6.24], we write

$$\begin{aligned} F(y;x) - \mathrm{Aff}(F)(y;x) &= \int_0^1 \mathrm{D}f(x_0 + s\tilde{x}, y_0 + s\tilde{y}) \cdot \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} \mathrm{d}s - \mathrm{D}f(x_0, y_0) \cdot \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} \\ &= \int_0^1 \int_0^1 \mathrm{D}^2 f(x_0 + st\tilde{x}, y_0 + st\tilde{y}) \cdot ((s\tilde{x}, s\tilde{y}), (\tilde{x}, \tilde{y}))\mathrm{d}t\mathrm{d}s. \end{aligned}$$

From the bounds on $\mathrm{D}_x^2 f$, $\mathrm{D}_x \mathrm{D}_y f$, and $\mathrm{D}_y^2 f$, we have

$$\|F(y;x) - \mathrm{Aff}(F)(y;x)\| \le \frac{1}{2} K_{xx} \|\tilde{x}\|^2 + K_{xy} \|\tilde{x}\| \|\tilde{y}\| + \frac{1}{2} K_{yy} \|\tilde{y}\|^2 .$$

Since $x \in \mathsf{B}(x_0, r_x)$, we have $\|\tilde{x}\| < r_x$, and for all $y \in \mathrm{bd}\,\mathsf{B}\left(y_0, r_y\right)$, we have

$$\|F(y;x) - \mathrm{Aff}(F)(y;x)\| < \frac{1}{2} K_{xx} r_x^2 + K_{xy} r_x r_y + \frac{1}{2} K_{yy} r_y^2. \tag{3.4}$$

Moreover,

$$\begin{aligned} \|\mathrm{Aff}(F)(y;x)\| &= \left\| f(x_0, y_0) + \mathrm{D}_x f(x_0, y_0)\tilde{x} + \mathrm{D}_y f(x_0, y_0)\, \tilde{y} - w_0 \right\| \\ &= \left\| \mathrm{D}_x f(x_0, y_0)\tilde{x} + \mathrm{D}_y f(x_0, y_0)\, \tilde{y} \right\| \\ &\ge \left| \left\| \mathrm{D}_y f(x_0, y_0)\tilde{y} \right\| - \|\mathrm{D}_x f(x_0, y_0)\tilde{x}\| \right| \\ &\ge \frac{1}{M_y} \|\tilde{y}\| - L_x \|\tilde{x}\| \\ &> \frac{1}{M_y} r_y - L_x r_x \quad \text{for all} \quad y \in \mathrm{bd}\,\mathsf{B}\left(y_0, r_y\right). \end{aligned} \tag{3.5}$$

We restate (3.1a) for convenience.

$$\frac{1}{2} K_{xx} r_x^2 + K_{xy} r_x r_y + \frac{1}{2} K_{yy} r_y^2 < \frac{r_y}{M_y} - L_x r_x .$$

Thus, from (3.1a), (3.4), and (3.5), we have

$$\|F(y;x) - \mathrm{Aff}(F)(y;x)\| < \|\mathrm{Aff}(F)(y;x)\| \quad \text{for all} \quad y \in \mathrm{bd}\,\mathsf{B}\left(y_0, r_y\right), \tag{3.6}$$

and therefore from Corollary 2.6, it follows that $F(\cdot;x)$ is nonvanishingly homotopic to $\mathrm{Aff}(F)(\cdot;x)$ on $\mathrm{bd}\,\mathsf{B}\left(y_0, r_y\right)$ with

$$\mathrm{Deg}\left(F(\cdot;x), \mathsf{B}\left(y_0, r_y\right)\right) = \mathrm{Deg}\left(\mathrm{Aff}(F)(\cdot;x), \mathsf{B}\left(y_0, r_y\right)\right). \tag{3.7}$$

Claim 2: $\mathrm{Deg}\left(\mathrm{Aff}(F)(\cdot;x), \mathsf{B}\left(y_0, r_y\right)\right) \in \{-1, 1\}$: Let $y^* \in \mathbb{R}^m$ be such that

$$\mathrm{Aff}(F)(y^*;x) = 0$$

(existence and uniqueness of such a y^* is guaranteed by the fact that $\mathrm{Aff}(F)(\cdot; x)$ is an affine map in y). Then, we have

$$y^* - y_0 = -\mathrm{D}_y f(x_0, y_0)^{-1} \mathrm{D}_x f(x_0, y_0)(x - x_0).$$

Taking norms, we get

$$\begin{aligned}\left\| y^* - y_0 \right\| &= \left\| -\mathrm{D}_y f(x_0, y_0)^{-1} \mathrm{D}_x f(x_0, y_0) \tilde{x} \right\| \\ &\leq \left\| \mathrm{D}_y f(x_0, y_0)^{-1} \right\| \left\| \mathrm{D}_x f(x_0, y_0) \right\| \left\| \tilde{x} \right\| \\ &\leq M_y L_x \left\| \tilde{x} \right\| \\ &\leq M_y L_x r_x.\end{aligned}$$

Rewriting (3.1a) as

$$r_y > M_y L_x r_x + M_y \left(\frac{1}{2} K_{xx} r_x^2 + K_{xy} r_x r_y + \frac{1}{2} K_{yy} r_y^2 \right),$$

we have

$$\left\| y^* - y_0 \right\| \leq M_y L_x r_x < r_y, \tag{3.8}$$

which implies $y^* \in \mathsf{B}\left(y_0, r_y\right)$ and therefore

$$\begin{aligned}\mathrm{Deg}\left(\mathrm{Aff}(F)(\cdot; x), \mathsf{B}\left(y_0, r_y\right)\right) &= \mathrm{sgn}(\det(\mathrm{DAff}(F)(y^*; x))) \\ &= \mathrm{sgn}(\det(\mathrm{D}_y f(x_0, y_0))) \in \{-1, 1\}.\end{aligned} \tag{3.9}$$

From (3.7), we know $\mathrm{Deg}\left(F(\cdot; x), \mathsf{B}\left(y_0, r_y\right)\right) \in \{-1, 1\} \neq 0$. (**DEG**-a) now asserts the existence of a $y_x \in \mathsf{B}\left(y_0, r_y\right)$ satisfying

$$F(y_x; x) := f(x, y_x) = 0.$$

Uniqueness: To complete the proof, all that remains is to show the uniqueness of y_x. *Claim 3*: $\mathrm{D}_y f(x, y)$ *is invertible for all* $(x, y) \in \mathsf{B}(x_0, r_x) \times \mathsf{B}\left(y_0, r_y\right)$: From the fundamental theorem of calculus [25, Theorem 6.24], we have

$$\mathrm{D}_y f(x, y) = \mathrm{D}_y f(x_0, y_0)\big(\mathrm{I} + \mathrm{D}_y f(x_0, y_0)^{-1} M(x, y)\big),$$

where

$$M(x, y) := \int_0^1 \left(\mathrm{D}_x \mathrm{D}_y f(x_0 + s\tilde{x}, y_0 + s\tilde{y})\tilde{x} + \mathrm{D}_y^2 f(x_0 + s\tilde{x}, y_0 + s\tilde{y})\tilde{y} \right) \mathrm{d}s.$$

Moreover,

$$\begin{aligned}\|M(x, y)\| &= \left\| \int_0^1 \left(\mathrm{D}_{x,y} f(x_0 + s\tilde{x}, y_0 + s\tilde{y})\tilde{x} + \mathrm{D}^2_y y(x_0 + s\tilde{x}, y_0 + s\tilde{y})\tilde{y} \right) \mathrm{d}s \right\| \\ &\leq K_{xy} \|\tilde{x}\| + K_{yy} \|\tilde{y}\| \\ &< K_{xy} r_x + K_{yy} r_y \quad \text{for all} \quad (x, y) \in \mathsf{B}\,(x_0, r_x) \times \mathsf{B}\left(y_0, r_y\right).\end{aligned}$$

From (3.1b), we have

$$\|M(x, y)\| < K_{xy} r_x + K_{yy} r_y < \frac{1}{M_y} \quad \text{for all} \quad (x, y) \in \mathsf{B}\,(x_0, r_x) \times \mathsf{B}\left(y_0, r_y\right).$$

Hence, using Lemma 3.4, we see that $\mathrm{D}_y f(x, y)$ is invertible for all $x, y \in \mathsf{B}\,(x_0, r_x) \times \mathsf{B}\left(y_0, r_y\right)$.

Claim 4: *$F(\cdot; x)$ has finite number of zeros in $\mathsf{B}\left(y_0, r_y\right)$*. From definition of $F(\cdot; x)$, we have

$$\mathrm{D}F(y; x) = \mathrm{D}_y f(x, y).$$

Since $\mathrm{D}_y f(x, y)$ is invertible for all $(x, y) \in \mathsf{B}\,(x_0, r_x) \times \mathsf{B}\left(y_0, r_y\right)$, 0 is a regular value of $F(\cdot; x)|_{\mathsf{B}(y_0, r_y)}$. Therefore, using Theorem 3.2 (inverse function theorem) $F^{-1}(0\,; x) := \{y \in \mathsf{B}\left(y_0, r_y\right) \mid F(y; x) = 0\}$ is discrete, i.e., around each $y_x \in \mathsf{B}\left(y_0, r_y\right)$, there exists a small enough neighborhood such that there is no other y satisfying $F(x; y) = 0$. From the continuity of $F(\cdot; x)$, $F^{-1}(0\,; x)$ is closed and is bounded since $F^{-1}(0\,; x) \subset \operatorname{cl} \mathsf{B}\left(y_0, r_y\right)$ and therefore compact and hence finite.

Let $F(\cdot; x)$ have N such zeros in $\mathsf{B}\left(y_0, r_y\right)$. Then from Definition 2.1 and the fact that $\mathrm{D}_y f(x, y)$ is invertible on $\mathsf{B}\,(x_0, r_x) \times \mathsf{B}\left(y_0, r_y\right)$, we have

$$\pm 1 = \operatorname{Deg}\left(F(\cdot; x), \mathsf{B}\left(y_0, r_y\right)\right) = \pm N$$

and thus $N = 1$. Therefore, for each $x \in \mathsf{B}\,(x_0, r_x)$ there exists a unique $y_x \in \mathsf{B}\left(y_0, r_y\right)$ such that $f(x, y_x) = 0$.

The regularity of the map $x \longmapsto g(x) := y_x$ is a consequence of Theorem 3.1 (implicit function theorem).

Suppose $K_{xx} = K_{xy} = L_x = 0$ then, for each $y \in V$, $f(\cdot, y)$ is a constant map and $f(x, y_0) = 0$ for all x. Since $f(x, \cdot) = f(x_0, \cdot)$ is now a map from $\mathbb{R}^m \longrightarrow \mathbb{R}^m$, the uniqueness of y_0 can be established using estimates on the IFT which we prove in the following corollary. □

Corollary 3.5 *Let $U \subset \mathbb{R}^n$ be nonempty and open and $U \ni x \longmapsto f(x) \in \mathbb{R}^n$ be as in Theorem 3.2 and in addition let f be $\mathcal{C}^2$. Define $L := \|\mathrm{D}f(x_0)\|$ and $M := \left\|\mathrm{D}f^{-1}(y_0)\right\|$. For a given $R > 0$ set $K := \sup\left\{\left\|\mathrm{D}^2 f(x)\right\| \mid x \in \mathsf{B}\,(x_0, R)\right\}$, and define*

$$P := \min\left\{\frac{1}{MK}, R\right\}, \quad \textit{and} \quad P' := \frac{P(2 - MKP)}{2M}.$$

Further, set $N = 8M^3K$ and define

$$Q := \min\left\{\frac{1}{NL}, \frac{P}{2M}, P\right\} \quad \textit{and} \quad Q' := \frac{Q(2-LNQ)}{L}.$$

Then, there exist

(3.5a) an open set $H \subset \mathsf{B}(x_0, P)$ such that f maps H diffeomorphically onto $\mathsf{B}\left(y_0, P'\right)$ with $y_0 := f(x_0)$, and

(3.5b) an open set $H' \subset \mathsf{B}(y_0, Q)$ such that f^{-1} maps H' diffeomorphically onto $\mathsf{B}\left(x_0, Q'\right)$.

A graphical representation of these sets is shown in Fig. 1.

Proof Define $(x, y) \longmapsto \psi(x, y) := f(x) - y$. From the definition, we have ψ a $\mathcal{C}^2$ map and $\psi(x_0, y_0) = f(x_0) - y_0 = 0$. Further, we have $\mathrm{D}_x\psi(x_0, y_0) = \mathrm{D}f(x_0)$ and is nonsingular. Therefore, ψ satisfies the assumptions of Theorems 3.1 and 3.3. Computing quantities as given in Theorem 3.3, we have

$$\begin{aligned} M := M_x &= \left\|(\mathrm{D}_x\psi(x_0, y_0)^{-1}\right\| = \left\|(\mathrm{D}f(x_0))^{-1}\right\| \\ L_y &:= \left\|\mathrm{D}_y\psi(x_0, y_0)\right\| = 1. \end{aligned}$$

and for $R := R_x > 0$ and $R_y > 0$ such that $\mathsf{B}(x_0, r_x) \subset U$ and $\mathsf{B}\left(y_0, r_y\right) \subset V$, we have

$$\begin{aligned} K_{xx} &= \sup\left\{\left\|\mathrm{D}_x^2\psi(x, y)\right\| \mid (x, y) \in \mathsf{B}(x_0, R_x) \times \mathsf{B}\left(y_0, R_y\right)\right\}, \\ &= \sup\left\{\left\|\mathrm{D}^2 f(x)\right\| \mid x \in \mathsf{B}(x_0, R_x)\right\} =: K \\ K_{xy} &:= \sup\left\{\left\|\mathrm{D}_{x,y}\psi(x, y)\right\| \mid (x, y) \in \mathsf{B}(x_0, R_x) \times \mathsf{B}\left(y_0, R_y\right)\right\} = 0, \quad \text{and} \\ K_{yy} &:= \sup\left\{\left\|\mathrm{D}_y^2\psi(x, y)\right\| \mid (x, y) \in \mathsf{B}(x_0, R_x) \times \mathsf{B}\left(y_0, R_y\right)\right\} = 0. \end{aligned}$$

Simplifying Eq. 3.1 for ψ, for each $0 < r_x < 1/MK$ and for $r_y = \frac{r_x(2-MKr_x)}{2M}$, for each $y \in \mathsf{B}\left(y_0, r_y\right)$ there exists an unique $x_y \in \mathsf{B}(x_0, r_x)$ satisfying $\psi(x_y, y) = 0 \iff f(x_y) = y$. Further, $y \longmapsto x_y =: f^{-1}(y)$ is $\mathcal{C}^2$. Defining $H := f^{-1}(\mathsf{B}\left(y_0, P'\right)) \cap \mathsf{B}(x_0, P)$ proves (3.5a).

For (3.5b), from the relation $f^{-1} \circ f =$ identity, for any $u_1, u_2 \in \mathbb{R}^n$ we have

$$\mathrm{D}^2 f^{-1}(y)(\mathrm{D}f(x)u_1, \mathrm{D}f(x)u_2) + \mathrm{D}f^{-1}(y)\mathrm{D}^2 f(x)(u_1, u_2) = 0.$$

After some rearrangement and imposing bounds on $\left\|\mathrm{D}f^{-1}(y)\right\|$, for all $x \in \mathsf{B}(x_0, P/2M)$, we arrive at

$$\left\|\mathrm{D}^2 f^{-1}(y)\right\| < 8M^3K = N \ \text{ for all } \ y \in \mathsf{B}(y_0, P/2M).$$

Applying (3.5a) on f^{-1} gives (3.5b). □

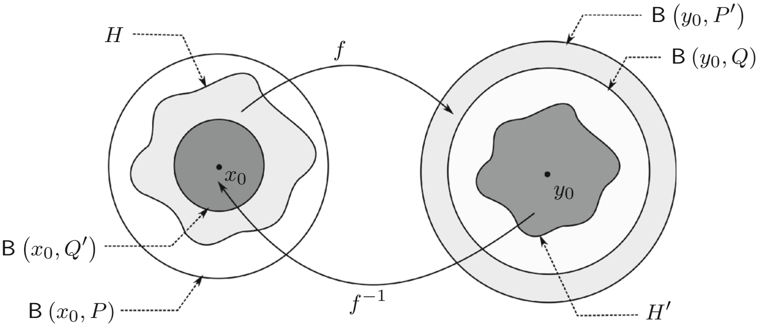

Fig. 1 Neighborhoods given in Corollary 3.5

Remark 3.6 Here are some remarks on Theorem 3.3 and Corollary 3.5.

(3.6a) Equation (3.1b) ensures uniqueness of y_x. If one is only interested in showing the existence of y_x, then one may relax (3.1) to just (3.1a).

(3.6b) For finite-dimensional spaces, the bounds given by Corollary 3.5 improve on the estimates provided by Abraham et al. [16, Proposition 2.5.6]: Proposition 2.5.6 from [16] provides the bounds for the estimates for the IFT. Bounds given by Corollary 3.5 for the IFT are applicable on a larger domain than that of [16, Proposition 2.5.6].

(3.6c) Since our objective is to come up with bounds that require minimal numerical computation, we relax several inequalities on the way and use conservative versions of them. For instance, we use Corollary 2.6, which is a coarser version of Theorem 2.5. Further, in several places, we utilize the submultiplicativity of norms to arrive at these results.

(3.6d) The proof for Theorem 3.3 relies on constructing an approximation $\mathrm{Aff}(F)(\cdot; x)$ of the map $F(\cdot; x)$ and showing both of them are nonvanishingly homotopic on bd $\mathsf{B}\left(y_0, r_y\right)$. One may improve the obtained estimates by approximating $F(\cdot; x)$ with a different function $\tilde{F}(\cdot\, ; x)$ and constructing homotopy H' such that F and $\tilde{F}$ are nonvanishingly homotopic and then use analysis on $\tilde{F}$ to show the existence of solutions of $F(y; x) = 0$. We refer the reader to [20, Chapter 1], in which Krasnosel'skij et al. have discussed several methods to come up with such homotopies.

In order to compute these estimates, one needs to compute the second-order derivatives over a bounded region. In light of this, $f \in \mathcal{C}^2$ is a necessary requirement. However, for $\mathcal{C}^1$ maps, one can alternatively use the bounds on the first-order derivatives to come up with similar estimates. Holtzman [12] provides explicit estimates for the implicit function theorem (given by the following proposition). However, for finite-dimensional spaces, one can also obtain these estimates by degree theoretic ideas. We now present these estimates and a degree theoretic proof of the same.

Proposition 3.7 [12, Theorem] *Let* $\mathbb{R}^n \times \mathbb{R}^m \supset \Omega \ni (x, y) \longmapsto P(x, y) \in \mathbb{R}^m$ *be a continuous map. Let* Ω *be open and* $(x_0, y_0) \in \Omega$. *Assume:*

(i) $P(x_0, y_0) = 0$;
(ii) $\mathrm{D}_y P$ *exists and is continuous in* Ω;
(iii) $\mathrm{D}_y P(x_0, y_0)$ *has bounded inverse* $\Gamma = (\mathrm{D}_y P(x_0, y_0))^{-1}$ *and* $\|\Gamma\| = k_1$;
(iv) $S = \{(x, y) \in \mathsf{B}(x_0, \delta) \times \mathrm{cl}\,\mathsf{B}(y_0, \epsilon)\} \subset \Omega$;
(v) *there is a real-valued function* $g_1(u, v)$ *defined for* $(u, v) \in [0, \delta] \times [0, \epsilon]$ *and nondecreasing in each argument with the other fixed such that for all* $(x, y) \in S$

$$\left\|\mathrm{D}_y P(x, y) - \mathrm{D}_y P(x_0, y_0)\right\| \leq g_1(\|x - x_0\|, \|y - y_0\|);$$

(vi) *there is a nondecreasing function* g_2 *defined on* $[0, \delta]$ *such that for all* $(x, y) \in S$

$$\|P(x, y_0)\| \leq g_2(\|x - x_0\|);$$

(vii) $k_1 g_1(\delta, \epsilon) \leq \alpha < 1$, *and* $k_1 g_2(\delta) \leq \epsilon(1 - \alpha)$.

Then, a unique map F *exists defined on* $\mathsf{B}(x_0, \delta)$ *mapping into* $\mathrm{cl}\,\mathsf{B}(y_0, \epsilon)$ *and possessing the following properties:*

(a) $P(x, F(x)) = 0$ *for all* $x \in \mathsf{B}(x_0, \delta)$.
(b) $F(x_0) = y_0$.
(c) $x \longmapsto F(x)$ *is continuous.*

Proof We first prove the strict version of (vii) and then using this we achieve the nonstrict version (vii). To this end, let $\epsilon > 0$ and $\delta > 0$ be such that $k_1 g_1(\delta, \epsilon) < \alpha < 1$ and $k_1 g_2 < \epsilon(1 - \alpha)$ and $\mathsf{B}(x_0, \delta) \times \mathrm{cl}\,\mathsf{B}(y_0, \epsilon) \subset \Omega$. Fix an $x \in \mathsf{B}(x_0, \delta)$ and define $\mathrm{cl}\,\mathsf{B}(y_0, \epsilon) \ni y \longmapsto P(y; x) := P(x, y) - P(x_0, y_0)$. Further, define an approximation of $P(\cdot; x)$ as

$$y \longmapsto \mathrm{Aff}(P)(y; x) := P(x, y_0) + \mathrm{D}_y P(x_0, y_0)\tilde{y} \tag{3.10}$$

where $\tilde{y} := y - y_0$ and $\tilde{x} := x - x_0$. From the definition of $P(\cdot; x)$, we have

$$\begin{aligned} P(y; x) - \mathrm{Aff}(P)(y; x) &= (P(x, y) - P(x, y_0)) - \mathrm{D}_y P(x_0, y_0)\tilde{y}. \\ &= \int_0^1 \Big(\mathrm{D}_y P(x, y_0 + t\tilde{y}) - \mathrm{D}_y P(x, y_0)\Big)\tilde{y}\mathrm{d}t, \end{aligned}$$

and

$$\|P(y; x) - \mathrm{Aff}(P)(y; x)\| \leq \left(\int_0^1 \left\|\mathrm{D}_y P(x, y_0 + t\tilde{y}) - \mathrm{D}_y P(x, y_0)\right\| \mathrm{d}t\right) \|\tilde{y}\|.$$

Therefore, from (v), for all $y \in \mathrm{bd}\,\mathsf{B}(y_0, \epsilon)$, we have

$$\|P(y; x) - \mathrm{Aff}(P)(y; x)\| \leq g_1(\delta, \epsilon)\epsilon \leq \frac{\alpha\epsilon}{k_1}. \tag{3.11}$$

Similarly, using (iii), (vi) and (3.10), for all $y \in \operatorname{bd} \mathsf{B}(y_0, \epsilon)$, we have

$$\|\operatorname{Aff}(P)(y;x)\| \geq \frac{\epsilon}{k_1} - g_2(\delta). \tag{3.12}$$

Since by assumption $k_1 g_2(\delta) < \epsilon(1-\alpha)$, from (3.11) and (3.12), we have

$$\|P(y;x) - \operatorname{Aff}(P)(y;x)\| < \|\operatorname{Aff}(P)(y;x)\|.$$

Thus, using Corollary 2.6, we have

$$\operatorname{Deg}(P(\cdot;x), \mathsf{B}(y_0,\epsilon)) = \operatorname{Deg}(\operatorname{Aff}(P)(\cdot;x), \mathsf{B}(y_0,\epsilon)) \in \{-1, 1\}.$$

Therefore, there exists a $y_x \in \mathsf{B}(y_0, \epsilon)$ such that $P(x, y_x) = 0$.
Following arguments similar to those of Theorem 3.3, for uniqueness, it suffices to show that $\mathrm{D}_y P(x,y)$ is nonsingular for all $(x,y) \in S$. Rewriting $\mathrm{D}_y P(x,y)$ as

$$\mathrm{D}_y P(x,y) = \mathrm{D}_y P(x_0,y_0)\Big(\mathrm{I}_n - \Gamma(\mathrm{D}_y P(x_0,y_0) - \mathrm{D}_y P(x,y))\Big)$$

Using (vii), for all $(x,y) \in S$,

$$\begin{aligned}\Big\|\Gamma\Big(\mathrm{D}_y P(x_0,y_0) - \mathrm{D}_y P(x,y)\Big)\Big\| &\leq \|\Gamma\| \, \Big\|\mathrm{D}_y P(x_0,y_0) - \mathrm{D}_y P(x,y)\Big\| \\ &\leq k_1\alpha < 1.\end{aligned}$$

Hence using Lemma 3.4, $\mathrm{D}_y P(x,y)$ is nonsingular for all $(x,y) \in S$. Therefore for all $\epsilon > 0$, $\delta > 0$ such that $k_1 g_1(\delta,\epsilon) < \epsilon(1-\alpha)$, for all $x \in \mathsf{B}(x_0,\delta)$ there exists a unique $y_x \in \mathsf{B}(y_0,\epsilon)$ such that $P(x,y_x) = 0$.
For given $\delta > 0$, define

$$\epsilon^* := \sup\{\epsilon \mid k_1 g_1(\delta,\epsilon) \leq \alpha, \ k_1 g_2(\delta) \leq \epsilon(1-\alpha)\} > 0.$$

Choose an $\alpha < \alpha' < 1$ and extend g_1, g_2 if necessary, then there exists an ϵ' such that $k_1 g_1(\delta,\epsilon') < \alpha'$ and $k_1 g_2(\delta) < \epsilon'(1-\alpha')$ and $\mathsf{B}(x_0,\delta) \times \mathsf{B}\left(y_0,\epsilon'\right) \in \Omega$. One can see that $\epsilon' > \epsilon^*$. Set

$$\epsilon''(\alpha') = \inf\left\{\epsilon' \mid k_1 g_1(\delta,\epsilon') < \alpha', \ k_1 g_2(\delta) < \epsilon'(1-\alpha')\right\}.$$

Then for each $x \in \mathsf{B}(x_0,\delta)$, there exist a unique $y_x \in \mathsf{B}\left(y_0, \epsilon''(\alpha')\right)$ such that $P(x,y_x) = 0$. Let $\{\alpha_n \mid n \in \mathbb{N}\}$ be a monotone decreasing sequence with $\alpha_n \leq \alpha'$ and $\alpha_n \longrightarrow \alpha$. Suppose $g_1(\delta,\cdot)$ is continuous on $[\epsilon^*, \epsilon']$[2] then we have a sequence

[2] Such an extension of $g_1(\delta,\cdot)$ exists and one possible choice is

$$g_1(\delta,r) = \begin{cases} g_1(\delta,r) & r \leq \epsilon^*, \\ g_1(\delta,\epsilon^*) + \sup\left\{\left\|\mathrm{D}_y P(x,y) - \mathrm{D}_y P(x_0,y_0)\right\| \mid S_r\right\} & \\ -\sup\left\{\left\|\mathrm{D}_y P(x,y) - \mathrm{D}_y P(x_0,y_0)\right\| \mid S\right\} & r > \epsilon^*, \end{cases}$$

where $S_r = \mathsf{B}(x_0,\delta) \times \mathsf{B}(y_0,r)$.

$\epsilon_n := \epsilon''(\alpha_n)$, such that $\epsilon_n \longrightarrow \epsilon^*$. Furthermore, for each $x \in \mathsf{B}(x_0, \delta)$, there exists a unique y_x such that $P(x, y_x) = 0$ and for each n, $y_x \in \mathsf{B}(y_0, \epsilon_n)$. Therefore,

$$y_x \in \bigcap_{n \in \mathbb{N}} \mathsf{B}(y_0, \epsilon_n) = \operatorname{cl} \mathsf{B}\left(y_0, \epsilon^*\right).$$

Thus, for each δ, ϵ satisfying (vii), for each $x \in \mathsf{B}(x_0, \delta)$ there exists a unique $y_x \in \operatorname{cl} \mathsf{B}(y_0, \epsilon)$ satisfying $P(x, y_x) = 0$. Defining $x \longmapsto F(x) = y_x$ proves (a) and (b), while (c) is a consequence of Theorem 3.2 (ImFT). □

In Proposition 3.7, in addition to the above assumptions, if $\mathrm{D}_x P(x, y)$ exists and is continuous then one has the following corollary.

Corollary 3.8 *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open sets and $U \times V \ni (x, y) \longmapsto f(x, y) \in \mathbb{R}^n$ be $\mathcal{C}^1$. Suppose for some $(x_0, y_0) \in U \times V$, $\mathrm{D}_y f(x_0, y_0)$ is an isomorphism. Define $M_y := \left\|(\mathrm{D}_y f(x_0, y_0))^{-1}\right\|$ and $L_x := \|\mathrm{D}_x f(x_0, y_0)\|$. For any given $r_x > 0$ and $r_y > 0$, define*

$$\mathbf{L}_x(r_x) := \sup\left\{\|\mathrm{D}_x f(x, y_0) - \mathrm{D}_x f(x_0, y_0)\| \mid (x, y) \in \mathsf{B}(x_0, r_x)\right\}, \quad \textit{and}$$
$$\mathbf{L}_y(r_x, r_y) := \sup\left\{\left\|\mathrm{D}_y f(x, y) - \mathrm{D}_y f(x_0, y_0)\right\| \mid (x, y) \in \mathsf{B}(x_0, r_x) \times \mathsf{B}\left(y_0, r_y\right)\right\}.$$

Then for all $r_x > 0$ and $r_y > 0$ satisfying

$$\mathbf{L}_x(r_x) r_x + \mathbf{L}_y(r_x, r_y) r_y < \frac{r_y}{M_y} - L_x r_x, \quad \textit{and} \tag{3.13a}$$
$$M_y \mathbf{L}_y(r_x, r_y) \leq \alpha < 1, \tag{3.13b}$$

for each $x \in \mathsf{B}(x_0, r_x)$ there exist a unique $y_x \in \mathsf{B}\left(y_0, r_y\right)$ satisfying $f(x, y_x) = w_0 =: f(x_0, y_0)$. Furthermore, $\mathsf{B}(x_0, r_x) \ni x \longmapsto g(x) := y_x \in \mathsf{B}\left(y_0, r_y\right)$ is continuously differentiable.

Proof Define $g_1(r_x, r_y) := \mathbf{L}_y(r_x) r_y$ and $g_2(r_x) := \mathbf{L}_x(r_x)$. Applying Proposition 3.7 with inequality (vii) replaced with its strict version and rearranging terms provide (3.13a) and (3.13b). The differentiability of g is a consequence of Theorem 3.1 (ImFT). □

Corollary 3.9 *Let $U \subset \mathbb{R}^n$ be open and $U \ni x \longmapsto f(x) \in \mathbb{R}^n$ be a $\mathcal{C}^1$ map. Let $x_0 \in U$ be such that $\mathrm{D} f(x_0)$ is nonsingular. Define $L := \|\mathrm{D} f(x_0)\|$ and $M := \left\|(\mathrm{D} f(x_0))^{-1}\right\|$. For each $r > 0$, define*

$$\mathbf{L}(r) := \sup\left\{\|\mathrm{D} f(x) - \mathrm{D} f(x_0)\| \mid x \in \mathsf{B}(x_0, r)\right\}.$$

Then for all $r_x > 0$ and $r_y > 0$ satisfying

$$\mathbf{L}(r_x) < \frac{1}{M} \quad \textit{and} \quad r_y = \frac{r_x(1 - M\mathbf{L}(r_x))}{M},$$

there exist an open set $H_{r_x} \subset \mathsf{B}\,(x_0, r_x)$ *such that* f *maps* H_{r_x} *onto* $\mathsf{B}\left(y_0, r_y\right)$ $\mathcal{C}^1$-*diffeomorphically. Further, for each* $0 < r < r_y$, *define*

$$\mathbf{M}(r) = \sup\left\{\left\|\mathrm{D}f^{-1}(y) - \mathrm{D}f^{-1}(y_0)\right\| \mid y \in \mathsf{B}\,(y_0, r)\right\}.$$

Then, for each $r'_y > 0$ *and* $r'_x > 0$ *satisfying*

$$\mathbf{M}(r'_y) < \frac{1}{L} \quad \textit{and} \quad r'_x = \frac{r_y{}'(1 - L\mathbf{M}(r'_y)}{L},$$

there exists an open set $H'_{r'_y} \subset \mathsf{B}\left(x_0, r'_y\right)$ *such that* f^{-1} *maps* $H'_{r'_y}$, $\mathcal{C}^1-$ *diffeomorphically onto* $\mathsf{B}\left(x_0, r'_x\right) \subset H_{r_x}$.

Remark 3.10 Estimating $\mathbf{M}$ requires explicit expression of $\mathrm{D}f^{-1}(y_0)$. This can be overcome by using the relation $\mathrm{D}f^{-1}(y) = (\mathrm{D}f(x))^{-1}$, where $x = f^{-1}(y)$. Let $r' < r_x$, be such that $r = r'(1 - M\mathbf{L}(r'))/M$, then we have

$$\mathbf{M}(r) \leq \sup\left\{\left\|(\mathrm{D}f(x))^{-1} - (\mathrm{D}f(x_0))^{-1}\right\| \mid x \in \mathsf{B}\left(x_0, r'\right)\right\}.$$

Corollary 3.5 and Corollary 3.9 consider an unconstrained variation around y_0. However, often we are interested to know how the preimage of y under f varies when y is varied along a particular direction. A slightly improved version of Corollary 3.5 is presented in the following proposition.

Proposition 3.11 *Let* $U \subset \mathbb{R}^n$ *be a nonempty open set and* $f : U \longrightarrow \mathbb{R}^n$ *satisfy the assumptions of Theorem* 3.2 *and* L, M, K, P *and* R *be as defined in Corollary* 3.5. *Let* $\mathsf{W} \subset \mathbb{R}^n$ *be a subspace, define*

$$\mathsf{B}_{\mathsf{W}}\,(y_0, r_{\mathsf{W}}) := \{y \subset \mathbb{R}^n \mid y - y_0 \in \mathsf{W} \textit{ and } \|y - y_0\| < r_{\mathsf{W}}\}$$

and

$$M_{\mathsf{W}} := \sup\left\{\left\|\mathrm{D}f^{-1}(y_0)\tilde{y}\right\| \mid \tilde{y} \in \mathsf{B}_{\mathsf{W}}\,(0, 1)\right\}.$$

Then, for any

$$0 < r_x < \min\left\{\frac{1}{MK}, R\right\} \quad \textit{and} \quad r_{\mathsf{W}} = \frac{r_x(2 - r_x MK)}{2M_{\mathsf{W}}},$$

for each $y \in \mathsf{B}_{\mathsf{W}}\,(y_0, r_{\mathsf{W}})$, *there exist a unique* $x_y \in \mathsf{B}\,(x_0, r_x) \subset \mathbb{R}^n$ *such that* $f(x_y) = y$.

Proof First fix a $y \in \mathsf{B}_{\mathsf{W}}\,(y_0, r_{\mathsf{W}})$ and define $F(x; y) = (\mathrm{D}f x_0)^{-1}(f(x) - y)$. We have $F(x; y) = 0 \iff f(x) = y$. Define an approximation of $F(\cdot; y)$ as

$$\mathrm{Aff}(F)(x; y) = (x - x_0) - \mathrm{D}f(x_0)^{-1}(y - y_0).$$

Setting $\tilde{x} := x - x_0$ and $\tilde{y} := y - y_0$, we have

$$\begin{aligned} F(x; y) - \mathrm{Aff}(F)(x; y) &= \mathrm{D}f(x_0)^{-1}(f(x-y)) - (\tilde{x} - \mathrm{D}f(x_0)^{-1}\tilde{y}) \\ &= \mathrm{D}f(x_0)^{-1}\Big(f(x) - \big(f(x_0) + \mathrm{D}f(x_0)\tilde{x}\big)\Big). \end{aligned}$$

Therefore,

$$\|F(x; y) - \mathrm{Aff}(F)(x; y)\| < \frac{1}{2}MKr_x^2 \ \text{ for all } \ x \in \mathsf{B}\,(x_0, r_x)\,.$$

Similarly,

$$\begin{aligned} \|\mathrm{Aff}(F)(x; y)\| &= \left\|\tilde{x} - \mathrm{D}f(x_0)^{-1}\tilde{y}\right\| \\ &\geq \left|\|\tilde{x}\| - \left\|\mathrm{D}f(x_0)^{-1}\tilde{y}\right\|\right| \\ &> r_x - M_{\mathsf{W}}r_{\mathsf{W}} \ \text{ for all } \ x \in \mathrm{bd}\,\mathsf{B}\,(r_x, .) \end{aligned}$$

Then for all r_x and r_{W} satisfying

$$\frac{1}{2}MKr_x^2 \leq r_x - M_{\mathsf{W}}r_{\mathsf{W}}, \tag{3.14}$$

for all $y \in \mathsf{B}_{\mathsf{W}}\,(y_0, r_{\mathsf{W}})$, we have

$$\|F(x; y) - \mathrm{Aff}(F)(x; y)\| < \|\mathrm{Aff}(F)(x; y)\| \ \text{ for all } \ x \in \mathrm{bd}\,\mathsf{B}\,(x_0, r_x)\,,$$

and therefore are nonvanishingly homotopic, and

$$\mathrm{Deg}\,(F(\cdot; y), \mathsf{B}\,(x_0, r_x)) = \mathrm{Deg}\,(\mathrm{Aff}(F)(\cdot; y), \mathsf{B}\,(x_0, r_x)) \in \{-1, 1\}.$$

Thus, there exists an $x_y \in \mathsf{B}\,(x_0, r_x)$ such that $F(x_y; y) = 0 \iff f(x_y) = y$. This proves existence of such an $x_y \in \mathsf{B}\,(x_0, r_x)$ for each $y \in \mathsf{B}_{\mathsf{W}}\,(y_0, \mathsf{r}_{\mathsf{W}})$. Since for all $r_x < \min\{\frac{1}{MK}, R\}$, $\mathrm{D}F(x; y) = \mathrm{D}f(x_0)^{-1}\mathrm{D}f(x)$ is invertible for all $x \in \mathsf{B}\,(x_0, r_x)$, the uniqueness of x_y is asserted following arguments similar to Proposition 3.5. For a given r_x, maximizing r_{W} while satisfying (3.14) results in the equality

$$r_{\mathsf{W}} = \frac{r_x(2 - r_xMK)}{2M_{\mathsf{W}}}, \tag{3.15}$$

and thereby completing the proof. □

4 Applications

ImFT and IFT find their applications in proving several results in nonlinear analysis and form the basis of several results in system theory and control, such as the robustness of solutions, and the existence, and uniqueness of solutions of ordinary differential equations. Several existential results on robustness are a consequence of the ImFT. One can apply the bounds derived in this article on the ImFT and IFT to obtain quantitative variants of the aforementioned methods. In this section, we include two such applications. First, we look into the robustness of the solutions of the nonlinear equations with respect to parametric variations and particularly the quadratically constrained quadratic programs (QCQP). A second application of these bounds is in geometric control methods, where we utilize these bounds to give explicit estimates on the domain of the feedback linearization of discrete-time dynamical systems.

4.1 Robustness of nonlinear equations

Solving nonlinear equations is challenging, and one often needs to employ numerical methods to compute a solution for a system of nonlinear equations. Estimates of the bounds in the ImFT and IFT help in proving the existence of solutions for a system of nonlinear equations on a given set, as we now demonstrate.

Let $f\colon \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^n$ be a $\mathcal{C}^2$ map. Let $X \subset \mathbb{R}^n$, $U \subset \mathbb{R}^m$ be nonempty open sets. A system of nonlinear equations in (x, u) on $X \times U$ is given by

$$f(x,u) = 0 \tag{4.1}$$

is *solvable* if there exists an $(x_0, u_0) \in X \times U$ such that $f(x_0, u_0) = 0$.

Definition 4.1 Let $\Omega \subset X$ have a nonempty interior. (4.1) is said to be *robustly solvable* on Ω, if there exists an $r_u > 0$, such that for all $u \in \mathsf{B}(u_0, r_u) \subset U$, there exists a (unique) $x_u \in \Omega$ satisfying $f(x_u, u) = 0$. Moreover, for a given r_x, the supremum over all such r_us is called the *robustness margin* for (4.1) over Ω.

Remark 4.2 Definition 4.1 is similar to [26, Definition 2.1]. However, unlike [26, Definition 2.1], Ω can be arbitrarily chosen and need not arise from linear constraints given by Equation (2) in [27].

Theorem 4.3 *Let $X \subset \mathbb{R}^n$, $U \subset \mathbb{R}^m$ be nonempty set and $f : X \times U \longrightarrow \mathbb{R}^n$ be a $\mathcal{C}^2$ map. Let $(x_0, u_0) \in X \times U$ be such that $f(x_0, u_0) = 0$ and $\mathrm{D}_x f(x_0, u_0)$ is nonsingular. Define*

$$L_u := \|\mathrm{D}_u f(x_0, u_0)\| \quad \textit{and} \quad M_x := \left\| \left(\mathrm{D}_x f(x_0, u_0)^{-1} \right) \right\|,$$

and for a given $R_x > 0$, $R_u > 0$ such that $\mathsf{B}(x_0, R_x) \subset X$ and $\mathsf{B}(u_0, R_u) \subset U$ set

$$\begin{aligned}
K_{xx} &:= \sup \left\{ \left\| \mathrm{D}_x^2 f(x,u) \right\| \mid (x,u) \in \mathsf{B}(x_0, r_x) \times \mathsf{B}(u_0, r_u) \right\}, \\
K_{uu} &:= \sup \left\{ \left\| \mathrm{D}_u^2 f(x,u) \right\| \mid (x,u) \in \mathsf{B}(x_0, R_x) \times \mathsf{B}(u_0, R_u) \right\}, \quad \textit{and} \\
K_{xu} &:= \sup \left\{ \left\| \mathrm{D}_{x,u} f(x,u) \right\| \mid (x,u) \in \mathsf{B}(x_0, R_x) \times \mathsf{B}(u_0, R_u) \right\}.
\end{aligned}$$

Then for all r_x and r_u satisfying

$$\frac{1}{2}K_{xx}r_x^2 + K_{xu}r_x r_u + \frac{1}{2}K_{uu}r_u^2 < \frac{r_x}{M_x} - L_u r_u, \tag{4.2a}$$

$$K_{xu}r_u + K_{xx}r_x < \frac{1}{M_x}, \tag{4.2b}$$

$$0 < r_u \leq R_x, \text{ and} \tag{4.2c}$$

$$0 < r_x \leq R_y, \tag{4.2d}$$

(4.1) is robustly solvable on $\mathsf{B}\,(x_0, r_x)$ *and robustness margin is bounded below by* r_u.

Remark 4.4 Similar to Theorem 3.3, (4.2b) is necessary for the uniqueness of the solution. If the uniqueness is not required, then one can relax (4.2) to (4.2a), (4.2c) and (4.2d).

The proof follows directly from applying Theorem 3.3 on f and is therefore omitted. With the general nonlinear case described, we now consider the QCQPs.

4.1.1 Robustness of solutions of QCQPs

Let $u \in \mathbb{R}^n$ and define

$$\mathbb{R}^n \ni x \longmapsto F(x) := Q(x) + Lx \in \mathbb{R}^n,$$

where $L \in \mathbb{R}^{n\times n}$ and $Q(x)$ is a quadratic function with its i^{th} component defined as

$$[Q(x)]_i := x^\top Q_i x \text{ for all } i \in [n]. \tag{4.3}$$

where for all $i \in [n]$, $Q_i \in \mathbb{R}^{n\times n}$ is a symmetric matrix. For a given matrix $A \in \mathbb{R}^{m\times n}$ and a vector $b \in \mathbb{R}^m$, a QCQP is defined by a system of quadratic equations

$$F(x) = Q(x) + Lx = u, \tag{4.4}$$

subjected to m constraints

$$x \in \Omega := \{x \in \mathbb{R}^n \ \mid \ [Ax]_i \leq b_i \ i \in [m]\}, \tag{4.5}$$

where $[Ax]_i$ and b_i denote the i^{th} components of the vectors Ax and b. The constraint set Ω defines a polyhedron as shown in Fig. 2. The dashed lines represent the equalities $[Ax]_i = b_i$, and the shaded interior along with the boundary defines the constraint set.

It is assumed that the constraints are not redundant, and x and u have the same dimension. We are interested in the following problem.

Problem 4.5 For a given u_0, let there be an $x_0 \in \Omega$ such that (x_0, u_0) solves (4.4). Find the *robustness margin*, i.e., the largest $r_u>0$ such that for all u satisfying $\Big|[u_0]_i - [u]_i\Big| < r_u$ for all $i \in [n]$, there exists an $x \in \Omega$ (not necessarily unique) such that (x, u) solves (4.4).

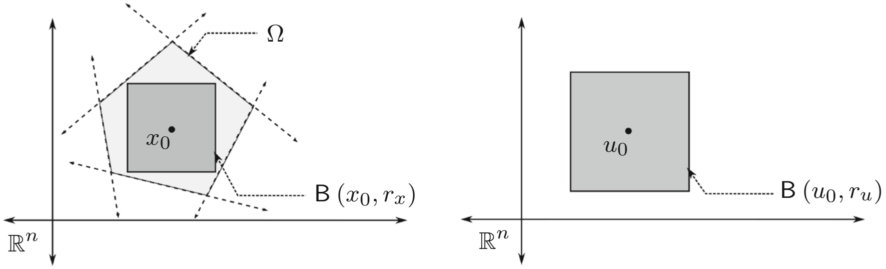

Fig. 2 Constraint set Ω and the *n-cube* around x_0 and u_0

First we utilize the ImFT to show that a nonzero robustness margin exists and then use the bounds derived for ImFT to come up with a lower bound on r_u. Casting (4.4) as (4.1), define

$$\mathbb{R}^n \times \mathbb{R}^n \ni (x, u) \longmapsto f(x, u) := Q(x) + Lx - u \in \mathbb{R}^n. \tag{4.6}$$

The Jacobian for f at (x_0, u_0) is given by

$$\mathrm{D}f(x_0, u_0) = \big(\mathrm{D}_x f(x_0, u_0),\, \mathrm{D}_u f(x_0, u_0)\big) = \big(2Q'(x_0) + L,\, \mathrm{I}_n\big) \tag{4.7}$$

where $Q'(x_0)$ denotes an $n \times n$ matrix with its i^{th} row as $x_0^\top Q_i$ and I_n denotes an identity matrix of order n. Since the inequalities are placed component-wise, i.e., we are looking for an *n-cube* around u_0, and Ω is defined by (4.5). We shall choose the infinity norm to work with, for a vector $v \in \mathbb{R}^n$ the infinity norm is defined as

$$\|v\|_\infty := \max\{|v_i| \mid i \in [n]\}.$$

The corresponding induced norm is defined as follows: let $A\colon \mathbb{R}^n \times \mathbb{R}^m$ be a linear operator, then $\|A\|_\infty$ is defined as

$$\|A\|_\infty := \max\{\|Ax\|_\infty \mid \|x\|_\infty \leq 1\}.$$

Let $\mathbb{R}^n \times \mathbb{R}^n \ni (u, v) \longmapsto B(u, v) \in \mathbb{R}^m$ be a bilinear map. The induced norm on B is defined as:

$$\|B\|_\infty := \max\{\|B(u, v)\|_\infty \mid \|u\|_\infty \leq 1,\ \|v\|_\infty \leq 1\}.$$

With this premise set, we have the following result.

Theorem 4.6 *Suppose* $Q \in \mathbb{R}^{n\times n}$ *and* $L \in \mathbb{R}^{n\times n}$ *is such that* $2Q'(x_0) + L$ *is non-singular, and* Ω *is nondegenerate, i.e., has a nonempty interior. Then* (4.4) *is robustly solvable on* Ω*. Moreover, define* $M_x := \left\|(2Q'(x_0) + L)^{-1}\right\|_\infty$ *and* $K_{xx} := \max\left\{|x^\top Q_i x| \mid \|x\|_\infty = 1, i \in [n]\right\}$*. Then, for a given* $0 < r_x < \frac{1}{M_x K_{xx}}$ *such that* $\mathsf{B}(x_0, r_x) \subset \Omega$*, the robustness margin* r_u *is lower bounded by* $\frac{r_x(2 - M_x K_{xx} r_x)}{2M_x}$.

The proof is a straightforward application of Theorem 3.3 on f as defined in (4.6) and hence omitted.

Remark 4.7 The condition $0 < r_x < 1/M_x K_{xx}$ ensures uniqueness of $x \in \mathsf{B}\,(x_0, 1/M_x K_{xx})$. However, one cannot assert the uniqueness of x on Ω, using Theorem 4.6.

Example 4.8 In order to illustrate the above-calculated bounds, we present a simple example. The example is taken from [26] so as to establish comparisons. The data are as follows:

$$A = \begin{bmatrix} -1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix},\ B = \begin{bmatrix} -0.5 \\ 3 \\ -0.5 \\ 3 \end{bmatrix},\ Q_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ Q_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},\ L = \begin{bmatrix} 1 & -3 \\ 2 & -1 \end{bmatrix},\quad \text{and}$$

$$u_0 = \begin{bmatrix} -2 \\ 4 \end{bmatrix}.$$

The underlying expression is then given by:

$$\begin{aligned} F(x) &= \begin{bmatrix} x_1^2 + x_1 - 3x_2 \\ x_2^2 + 2x_1 - x_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \end{bmatrix} \quad \text{with} \\ \Omega &= \left\{ x \colon \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \leq \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leq \begin{bmatrix} 3 \\ 3 \end{bmatrix} \right\}. \end{aligned} \tag{4.8}$$

The unique solution is given by $x_0 \approxeq \begin{bmatrix} 1.36 & 1.74 \end{bmatrix}^\top$. One can check that $\mathrm{D}F(x_0)$ is nonsingular. The corresponding quantities are $L_x = 6.7204$, $M_x = 0.3763$, $L_u = 1$, and $K_{xx} = 2$. For $r_x = 0.86$, $\mathsf{B}\,(x_0, 0.86) \subset \Omega$. The robustness margin for (4.8) comes out to be, $r_u \geq 1.546$. This is greater than the estimates given by LP-feasibility routine ($r_u \geq 1.2054$) by Dvijotham et al. [26]. The bounds can be further improved by changing the map $(x, u) \longmapsto (F(x) - u)$ to $\bar{F}(x, u) := (\mathrm{D}F(x_0)^{-1}) \cdot (F(x) - u)$. This gives us an updated bound of $r_u \geq 1.5781$; however, this is smaller than the estimate given by the bound tightening procedure ($r_u \geq 1.7069$) in [26]. Figure 3 shows simulation results for the improved bounds. The blue scatter plot shows $F(\Omega)$. The yellow rectangle shows the bd $\mathsf{B}\,(u_0, r_u)$. One can see that $\mathsf{B}\,(u_0, r_u)$ is contained in $F(\Omega)$ and thus for each $u \in \mathsf{B}\,(u_0, r_u)$ there will be an $x \in \Omega$ such that $F(x) - u = 0$. The bounds can be improved if one restricts the variation along a particular subspace. This is shown by the purple-colored parallelogram. The u was successively allowed to vary along an arbitrary unit vector (w_1, w_2), and the maximum perturbation was computed. This is denoted by $\mathrm{bd}(B_W)$ in the figure.

4.1.2 Application to power systems

Power Flow studies and Optimal Power Flow studies require us to calculate operating voltages and currents by balancing the load demand and the power generated. Integration of renewable sources like solar and wind energy in the power system network

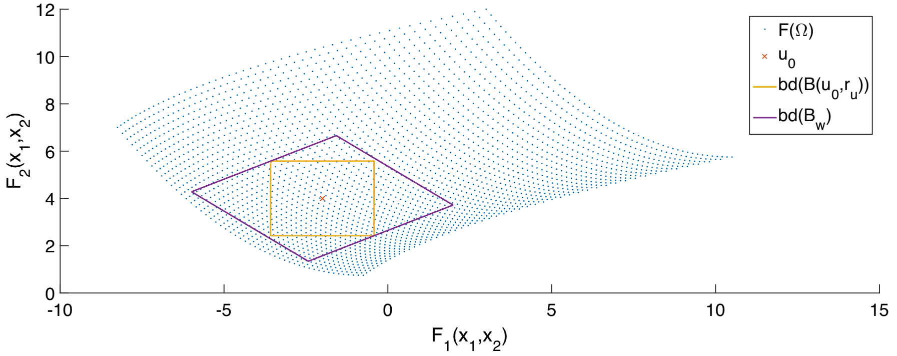

Fig. 3 Robustness Margin for Example 4.8

leads to uncertainty in the generated supply. Due to this uncertain supply of renewable sources, it is difficult to ensure that there will be sufficient power generation to meet the power demand while adhering to the stable operation limits of the power system network. The power flow equations involve balancing the generated power supply and the required demand. As described by Dvijotham et al. [27], the AC power-flow equations are written as follows:

$$\begin{aligned}
\mathrm{Re}\left(V_i\overline{Y_{i0}V_0}+\sum_{k=1}^{n}V_i\overline{Y_{ik}V_k}\right)&=p_i \quad \text{for all } i\in \mathrm{PQ},\\
\mathrm{Im}\left(V_i\overline{Y_{i0}V_0}+\sum_{k=1}^{n}V_i\overline{Y_{ik}V_k}\right)&=q_i \quad \text{for all } i\in \mathrm{PQ},\\
\mathrm{Re}\left(V_i\overline{Y_{i0}V_0}+\sum_{k=1}^{n}V_i\overline{Y_{ik}V_k}\right)&=p_i \quad \text{for all } i\in \mathrm{PV},\\
|V_i|^2&=v_i^2 \quad \text{for all } i\in \mathrm{PV},
\end{aligned} \tag{4.9}$$

where V_i denotes the complex voltage phasor, $p_i,\ q_i$ denotes the active and reactive power injection at the node i, and Y denotes the admittance matrix. PV denotes the set of generator buses, and PQ denotes the load bus. v_i denotes the root-mean-square (rms) voltage at the i^{th} bus. V_0 denotes the reference slack bus and is kept fixed at $1+j0$ per unit magnitude, where $j=\sqrt{-1}$. Power flow as given by 4.9 is quadratic. Further, for stable operation of the power system network, one needs to maintain the bus voltage magnitude within prescribed limits. This imposes a quadratic constraint on (4.9). Setting

$$\begin{aligned}
x&=\left[\mathrm{Re}(V_1)\ \dots\ \mathrm{Re}(V_n)\ \mathrm{Im}(V_1)\ \dots\ \mathrm{Im}(V_n)\right] \text{ and}\\
u&=\left[p_1\ \dots\ p_n\ q_1\ \dots\ q_n\ v_1^2\ \dots\ v_n^2\right]
\end{aligned}$$

the powerflow equations given by (4.9) can be written as QCQP of type (4.5) with $\mathbb{R}^{2n} \ni x \longmapsto F(x) \in \mathbb{R}^{2n}$ with its i^{th} component given by

$$\begin{aligned} [F(x)]_i &:= \operatorname{Re}\left(\sum_{k=1}^{n} V_i \overline{Y_{i0} V_0} + \overline{V_i Y_{ik} V_k}\right) \quad \text{for all } \ i \in [n], \\ [F(x)]_{n+i} &:= \operatorname{Im}\left(V_i \overline{Y_{i0} V_0} + \sum_{k=1}^{n} V_i \overline{Y_{ik} V_k}\right) \quad \text{for all } \ i \in \text{PQ}, \\ [F(x)]_{n+i} &:= \operatorname{Re}(V_i)^2 + \operatorname{Im}(V_i)^2 \quad \text{for all } \ i \in \text{PV}. \end{aligned} \tag{4.10}$$

For stable operation, the voltage magnitudes are to be kept within tolerance limits. For a fixed $r_x \in]0, 1[$, the constraint set is

$$\Omega = \{x \in \mathbb{R}^{2n} \mid 1 - r_x \le v_i^2 \le 1 + r_x \text{ for all } i \in [n]\}.$$

The constraint set Ω can be relaxed to the following

$$\Omega' = \{x \in \mathbb{R}^{2n} \mid -r_x \le \left\|x - x^*\right\|_\infty \le r_x \text{ for all } i \in [n]\}$$

where $x^* \in \mathbb{R}^{2n}$ be the nominal operating point satisfying $(\operatorname{Re} V_i^*)^2 + \operatorname{Im}(V_i^*)^2 = 1$ for all $i \in [n]$.

The test cases are obtained from the dataset given in the MatPower package found in MATLAB software [28]. The package contains well-defined libraries and a set of routines for solving problems like Power Flow analysis, Optimal Load Flow, and DC Power Flow analysis. The package is open source and is available online. The problem is simulated for several test cases from MatPower. The maximum allowable r_x and r_u are recorded in P.U. magnitude. In order to simulate real-life scenarios, we only consider the variation in the first five dimensions of u. The results are tabulated in Table 1, where $M_f := \left\|\mathrm{D}F^{-1}(u_0)\right\|$, $M_F' := \sup\left\{\mathrm{D}F^{-1}(u_0) \cdot \tilde{u} \mid u \in \mathbb{R}^{2n},\ \tilde{u}_i = 0 \text{ for all } i > 5\right\}$, and $K_{\bar{F}} = \sup\left\{\left\|\mathrm{D}^2 \bar{F}(x)\right\| \mid x \in \mathbb{R}^{2n}\right\}$ where $\bar{F} = (\mathrm{D}F(x_0))^{-1} \cdot F$. For Case 5, 9, 14, and 30, we also plot r_x with respect to r_u and compare it with the bounds given by the bound tightening method given by Dvijotham et al. [26]. These are shown in Fig. 4.

Remark 4.9 From Table 1 and Fig. 4, one can see, for lower values of r_x, our bounds are comparable to those of [26]. However, with an increase in the dimension, our estimates start to perform poorly as compared to those given in [26]. The one key reason for this is the way we bound K_{xx}. In order to keep computation minimum, we bound $\max\left\{|x^\top Q x| \mid \|x\|_\infty \le 1\right\}$ by $\sum_{i,j\in[n]}|Q_{ij}|$. For a general matrix Q, this grows quadratically with respect to n. However, if Q is a sparse matrix, then one can obtain tighter estimates. One can alternatively utilize the inequlaity

$$\max\left\{|x^\top Q x| \mid \|x\|_\infty \le 1\right\} \le n \left\|Q\right\|_2,$$

Table 1 Lower bounds on the robustness margin r_u for various cases from MatPower

Case	M_F	M'_F	$K_{\bar{F}}$	Max r_x	Max r_u
5	0.5154	0.0512	12.971	0.0771	0.7529
9	1.3802	0.7968	39.065	0.0256	0.0161
14	2.4795	0.5291	91.066	0.0110	0.0104
30	6.2576	0.3225	0.330×10^3	3.303×10^{-3}	5.120×10^{-3}
57	12.153	0.2657	1.032×10^3	0.969×10^{-3}	1.823×10^{-3}
85	5.1119	0.0140	13.48×10^3	0.074×10^{-3}	2.643×10^{-3}
141	22.049	0.2371	1.958×10^3	0.512×10^{-3}	1.078×10^{-3}

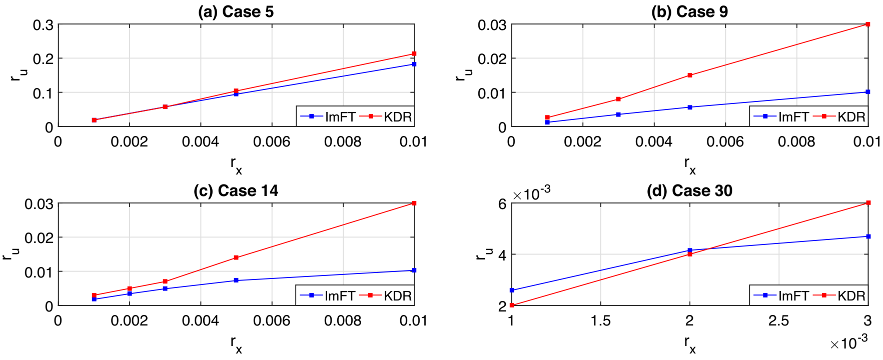

Fig. 4 Robustness margin estimates obtained using ImFT (Blue) for **a** Case 5, **b** Case 9, **c** Case 14 and **d** Case (30) compared with estimates obtained using bound tightening (Red) from [26]

to estimate which grows linearly with n.

Further, with an increase in n, the problem dimension increases and norm-based inequalities such as triangular inequality and sub-multiplicative inequality become more conservative. Our estimates are conservative when compared to those provided in [26]. However, the numerical computation required to compute these bounds is much lower than that required by [26]. In particular, no optimization routine is required to compute these bounds. The bounds given by Dvijotham et al. [26] require us to perform nontrivial optimizations on matrices and solve LP feasibility programs. In comparison with that, we only require to compute for the inverse of an $n \times n$ matrix, i.e., $\mathrm{D}F(x_0)$, for which there exist efficient algorithms and packages in the literature. Since our objective is not to tackle matrix inversion methods, we do not investigate methods for computing matrix inversion. Some literature on matrix inversion can be found in [29–32]. One key limitation of this method is that it can only provide us with a lower bound on the robustness margin.

4.1.3 Robustness of the algebraic Riccati equation

Consider a linear time-invariant control system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Ax(t) + Bu(t) \tag{4.11}$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$ for all $t > 0$. For any given $n \in \mathbb{N}$, $\mathsf{Sym}(n)$, $\mathsf{PD}(n)$, and $\mathsf{PSD}(n)$ denote the set of symmetric matrices, positive-definite matrices, and positive-semidefinite matrices of order $n \times n$, respectively. Let

$$\mathcal{U} = \left\{ [0, \infty[\ni t \longmapsto u(t) \in \mathbb{R}^m \mid \int_0^\infty \|u(t)\|^2 \, \mathrm{d}t < \infty \right\},$$

be the set of square-integrable controls over infinite horizon. Define the infinite horizon linear quadratic regulator (LQR) as the following optimal control problem.

Problem 4.10 For given $Q \in \mathsf{PSD}(n)$, and $R \in \mathsf{PD}(m)$ compute a u^*, if it exists, such that it solves

$$\begin{cases} \underset{u \in \mathcal{U}}{\text{minimize}} & J(u) = \frac{1}{2} \int_0^\infty x(t)^\top Q x(t) + u(t)^\top R u(t), \\ \text{subject to} & x(0) = x_0, \text{ and (4.11) for almost all } t \in [0, \infty[\,. \end{cases}$$

From results on LQR [33, Chapter 6], the optimal control is of the type

$$u^*(t) = -R^{-1} B^\top P x^*(t) \quad \text{for almost all } t \in [0, \infty[,$$

where $P \in \mathsf{PD}(n)$ solves the algebraic Riccati equation (ARE)

$$A^\top P + PA + Q - PBR^{-1}B^\top P = 0. \tag{ARE}$$

For given $A_0 \in \mathbb{R}^{n\times n}$, and $B_0 \in \mathbb{R}^{n\times m}$ such that (A_0, B_0) forms a stabilizable pair, consider a linear control system given by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = (A_0 + \Lambda)x + B_0 u \tag{4.12}$$

where parameter $\Lambda \in \mathbb{R}^{n\times n}$ denotes the unmodeled part of system dynamics.

Problem 4.11 Find a bound on $\|\Lambda\|$ such that (4.12) is stabilizable, i.e., find an r_Λ such that $(A_0 + \Lambda, B_0)$ is stabilizable for all $\Lambda \in \mathsf{B}\,(0, r_\Lambda)$.

Remark 4.12 Note that we may have very-well-considered variations in B_0 as well, however for illustration purposes, we restrict to variations in A_0.

Lemma 4.13 [33, Chapter 6] *For each $Q \in \mathsf{PD}(n)$ and $R \in \mathsf{PD}(m)$, a unique positive-definite solution for* (ARE) *exists if and only if (A, B) forms a stabilizable pair.*[3]

[3] Two given matrices $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times n}$ are said to form a stabilizable pair if there exists a $K \in \mathbb{R}^{n\times m}$ such that $A + BK$ is Hurwitz, i.e., have eigenvalues with strictly negative real parts.

Using the above lemma, for any given Λ, $(A_0+\Lambda, B_0)$ being stabilizable is equivalent to showing that (ARE) has a positive-definite solution with $A = A_0+\Lambda$. Further, since $\mathsf{PD}(n)$ is open in $\mathsf{Sym}(n)$, there exists an r_P such that for all $P \in \mathsf{B}(P_0, r_P) \subset \mathsf{Sym}(n)$, P is positive-definite. Problem 4.11 can now be formulated as the following robustness problem.

Problem 4.14 For given A_0, B_0, Q, R, let there be a $P_0 \in \mathsf{PD}(n)$ such that it solves (ARE). For a given $r_P > 0$ such that $\mathsf{B}(P_0, r_P) \subset \mathsf{PD}(n)$, find an $r_\Lambda > 0$ such that for each $A \in \mathsf{B}(A_0, r_\Lambda) \subset \mathbb{R}^{n\times n}$ there exists a $P \in \mathsf{B}(P_0, r_P)$ solving

$$f(A, P) := A^\top P + PA + Q - PB_0R^{-1}B_0^\top P = 0.$$

Note that Problem 4.14 is just investigating the robustness of solutions of nonlinear equations. Therefore, we may use the bounds derived for ImFT to compute r_Λ. To this note, we have $\mathrm{D}_P f(A, P))$ defined as

$$\begin{aligned}\mathsf{Sym}(n) \ni \mu \longmapsto \mathrm{D}_P f(A, P)\cdot\mu &= A^\top\mu + \mu A \\ &\quad -(\mu B_0R^{-1}B_0^\top P + PB_0R^{-1}B_0^\top\mu)\end{aligned} \tag{4.13}$$

and $\mathrm{D}_A f(A, P)$ is defined as

$$\mathbb{R}^{n\times n} \ni \nu \longmapsto \mathrm{D}_A f(A, P)\cdot\nu = \nu^\top P + P\nu. \tag{4.14}$$

Evaluating $\mathrm{D}_P f$ at (A_0, P_0), we have the following linear map:

$$\begin{aligned}\mu \longmapsto \mathrm{D}_P f(A_0, P_0)\cdot\mu &= A_0^\top\mu + \mu A_0 - (\mu B_0R^{-1}B_0^\top P_0 + P_0B_0R^{-1}B_0^\top\mu) \\ &= (A_0 - B_0R^{-1}B_0^\top P_0)^\top\mu + \mu(A_0 - B_0R^{-1}B_0^\top P_0) \\ &= A_c^\top\mu + \mu A_c\end{aligned} \tag{4.15}$$

with $A_c = A_0 - B_0R^{-1}B_0^\top P_0$ Since (A_0, B_0) is a stabilizable pair and $u^* = -R^{-1}B_0^\top P_0x^*$ is the stabilizing control, A_c is a stable matrix with its eigenvalues having strictly negative real parts. For any $v \in \mathsf{Sym}(n)$, define

$$v \longmapsto \mu(v) := -\int_0^\infty e^{tA_c^\top} v e^{tA_c}\mathrm{d}t,$$

then we have

$$A_c^\top\mu(v) + \mu(v)A_c = v.$$

Therefore, $\mathrm{D}_P f(A_0, P_0) : \mathsf{Sym}(n) \longrightarrow \mathsf{Sym}(n)$ is a surjective map. Furthermore, using rank-nullity theorem we can conclude that it is an isomorphism with inverse

$$v \longmapsto (\mathrm{D}_P f(A_0, P_0)^{-1})\cdot v = -\int_0^\infty e^{tA_c^\top} v e^{tA_c}\mathrm{d}t. \tag{4.16}$$

Since $\mathrm{D}_P f(A_0, P_0)$ is an isomorphism, we can now apply Lemma 3.8. Define

$$M_P := \|\mathrm{D}_P f(A_0, P_0)\| \quad \text{and} \quad L_A := \|\mathrm{D}_A f(A_0, P_0)\| \,.$$

Further for any given $r > 0$ and $r^* > 0$, define

$$\mathbf{L}_P(r, r^*) := \sup\left\{\mathrm{D}_P f(A, P) - \mathrm{D}_P f(A_0, P_0) \mid (A, P) \in \mathsf{B}\,(A_0, r) \times \mathsf{B}\left(P_0, r^*\right)\right\}$$

and

$$\mathbf{L}_A(r) := \sup\left\{\mathrm{D}_A f(A, P_0) - \mathrm{D}_A f(A_0, P_0) \mid A \in \mathsf{B}\,(A_0, r)\right\}.$$

Then for all $r_P > 0, r_\Lambda > 0$ with $\mathsf{B}\,(P_0, r_P) \subset \mathsf{PD}(n)$ satisfying

$$\mathbf{L}_A(r_\Lambda) r_\Lambda + \mathbf{L}_P(r_\Lambda, r_P) r_P < \frac{r_P}{M_P} - L_A r_\Lambda, \text{ and} \tag{4.17a}$$

$$M_P \mathbf{L}_P(r_\Lambda, r_P) < 1, \tag{4.17b}$$

for each $A \in \mathsf{B}\,(A_0, r_\Lambda)$, there exists a $P \in \mathsf{B}\,(P_0, r_P)$ such that it solves ARE and consequently (A, B) is stablizable for all $A \in \mathsf{B}\,(A_0, r_\Lambda)$.

Remark 4.15 The bounds r_P and r_Λ depend upon the choice of the nominal point (A_0, P_0), which itself depends upon the choice of Q and R. One can then optimize these bounds over all possible choices of $Q \in \mathsf{PD}(n)$ and $R \in \mathsf{PD}(m)$. However, for each Q and R, to compute the corresponding P, we need to solve (ARE) which is a nonlinear equation making the optimization of r_P, r_Λ with respect to Q and R at least difficult, if not intractable.

We now present a simple illustrative example on the double integrator.

Example 4.16 Consider the (perturbed) double integrator system defined by

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \Lambda\right)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u$$

with $\Lambda \in \mathbb{R}^{2\times 2}$. Set $Q = \mathrm{I}_2$ and $R = 1$, then the corresponding solution of (ARE) is

$$P_0 = \begin{pmatrix} \sqrt{3} & 1 \\ 1 & \sqrt{3} \end{pmatrix}.$$

For any $\begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} =: v \in \mathbb{R}^{2\times 2}$ and $\begin{pmatrix} \mu_1 & \mu_2 \\ \mu_2 & \mu_3 \end{pmatrix} =: \mu \in \mathsf{Sym}(2)$, we have

$$\mathrm{D}_A f(A_0, P_0) \cdot v = \begin{pmatrix} 2(\sqrt{3}v_{11} + v_{21}) & v_{11} + v_{22} + \sqrt{3}(v_{21} + v_{12}) \\ v_{11} + v_{22} + \sqrt{3}(v_{21} + v_{12}) & 2(\sqrt{3}v_{22} + v_{12}) \end{pmatrix}$$

and

$$\mathrm{D}_P f(A_0, P_0) \cdot \mu = \begin{pmatrix} -2\mu_2 & \mu_1 - \sqrt{3}\mu_2 - \mu_3 \\ \mu_1 - \sqrt{3}\mu_2 - \mu_3 & 2(\mu_2 - \sqrt{3}\mu_3) \end{pmatrix}.$$

Since $A_c = (A_0 - B_0 B_0^\top P_0)$ is Hurwitz, $\mathrm{D}_P f(0, P_0)$ is invertible and for any $\begin{pmatrix} \nu_1 & \nu_2 \\ \nu_2 & \nu_3 \end{pmatrix} =: \nu \in \mathsf{Sym}(2)$, we have

$$\mathrm{D}_P f(A_0, P_0)^{-1} \cdot \nu = \begin{pmatrix} -1.1574\nu_1 + \nu_2 - 0.2887\nu_3 & -0.5\nu_1 \\ -0.5\nu_1 & -0.2887(\nu_1 + \nu_3) \end{pmatrix}. \tag{4.18}$$

Using the infinity norm, we have

$$L_A = \|\mathrm{D}_A f(A_0, P_0)\| = 6.928 \quad \text{and} \quad M_P = \left\| \mathrm{D}_P f(A_0, P_0)^{-1} \right\| = 3.0207$$

Further, we have

$$\mathbf{L}_A(r_\Lambda) = 0 \quad \text{and} \quad \mathbf{L}_P(r_\Lambda, r_P) \leq 2(r_P + r_\Lambda).$$

Substituting in Eq. (4.17), we get that for all $r_P > 0$, and $r_\Lambda > 0$ satisfying

$$\begin{aligned} 2r_P(r_\Lambda + r_P) &< 0.3310 r_P - 6.928 r_\Lambda \quad \text{and} \\ r_P + r_\Lambda &< 0.1655 \end{aligned}$$

for each $A \in \mathsf{B}(A_0, r_\Lambda)$, there exists a $P \in \mathsf{B}(P_0, r_P)$ such that it solves the (ARE).

Remark 4.17 The purpose of these examples is not to compute the optimum bounds for which we require nonlinear optimization (often global), but to rather show that several fundamental problems in systems and control applications can be recast into the ImFT and IFT framework.

4.2 Estimating the domain of feedback linearizability

Let $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{U} \subset \mathbb{R}^m$ be nonempty open sets and $\mathcal{X} \times \mathcal{U} \ni (x, u) \longmapsto f(x, u) \in \mathcal{X}$ be an analytic map. Define a discrete-time control system as

$$x(k+1) = f(x(k), u(k)) \tag{4.19}$$

where for each $k \in \mathbb{N}$, $x(k) \in \mathcal{X}$ denotes the system state and $u(k) \in \mathcal{U}$ is the control input. A point $(x_0, u_0) \in \mathcal{X} \times \mathcal{U}$ is called the *equilibrium point* of (4.19) if it satisfies $f(x_0, u_0) = x_0$.

Definition 4.18 Let (x_0, u_0) be an equilibrium point of (4.19) and $\mathcal{O}(x_0) \ni x_0$, $\mathcal{O}(u_0) \ni u_0$ be nonempty and open. Suppose

$$\mathcal{O}(x_0) \ni x \longmapsto \phi(x) \in \mathbb{R}^n$$

is a diffeomorphism onto its image and

$$\mathcal{O}(x_0) \times \mathcal{O}(u_0) \ni (x, u) \longmapsto \psi(x, u) =: v \in \mathbb{R}^m$$

is such that for each $x \in \mathcal{O}(x_0)$, $\psi(x, \cdot) : \mathcal{O}(u_0) \longrightarrow \mathbb{R}^m$ is an injective map. Then, (4.19) is said to be *locally feedback linearizable* on $\mathcal{O}(x_0) \times \mathcal{O}(u_0)$ if ϕ and ψ transforms (4.19) to an equivalent controllable linear system of type

$$z(k+1) := Az(k) + Bv(k) \tag{4.20}$$

with $z(k) = \phi(x(k))$ and $v(k) = \psi(x(k), u(k))$ for all $k \in \mathbb{N}$.

Since feedback-linearization-based methods are local, i.e., the results hold only in a sufficiently small neighborhood of the operating point, for practical implementation of such methods one needs to know an apriori estimate on the domain on which (4.19) is feedback linearizable. Using the necessary and sufficient conditions for feedback linearizability, one can ensure the existence of such open sets and maps. However, these methods do not provide estimates about how large these sets are. For systems with f being analytic, this is equivalent to finding the domains on which ϕ is a diffeomorphism and $\psi(x, \cdot)$ is injective. One can therefore utilize the bounds on the ImFT and IFT to come up with these estimates. The following propositions provide first estimates on the domain on which ϕ is a diffeomorphism and $\psi(x, \cdot)$ is injective.

Proposition 4.19 *Let ϕ be the linearizing coordinate change for (4.19). Define*

$$L_\phi := \|\mathrm{D}\phi(x_0)\| \; M_\phi := \left\|\mathrm{D}\phi^{-1}(z_0)\right\|.$$

For any given $r > 0$ define

$$\mathbf{L}_\phi(r) := \sup\{\mathrm{D}\phi(x) - \mathrm{D}\phi(x_0) \mid x \in \mathsf{B}(x_0, r)\}$$

and define

$$P_\phi := \arg\max\left\{\epsilon(r) \mid r > 0, \mathbf{L}_\phi(r) \leq \frac{1}{M}\right\} \; \textit{and} \; P'_\phi = \epsilon(P_\phi),$$

where for all $r > 0$, $\epsilon(r) := r(1 - M_\phi \mathbf{L}_\phi(r))/M_\phi$. Then there exists an open set $H_\phi \subset \mathsf{B}\left(x_0, P_\phi\right)$ such that ϕ maps H_ϕ onto $\mathsf{B}\left(x_0, P'_\phi\right)$ diffeomorphically.

Proposition 4.20 *Let $u \longmapsto \psi(x,u) := v$ be the new linearized control. Suppose $\mathrm{D}_u\psi(x_0,u_0)$ is nonsingular, then define*

$$M_\psi^u := \left\|\mathrm{D}_u\psi(x_0,u_0)^{-1}\right\| \; L_\psi^x := \|\mathrm{D}_x\psi(x_0,u_0)\| .$$

For a given $r_x > 0$, $r_u > 0$ and $r_v > 0$ let $\mathcal{B}_1 := \mathsf{B}(x_0,r_x) \times \mathsf{B}(u_0,r_u)$ and define

$$\mathbf{L}_\psi^x(r_x) := \sup\left\{\|\mathrm{D}_x\psi(x_0,u_0) - \mathrm{D}_x\psi(x,u_0)\| \mid (x,u_0) \in \mathcal{B}_1\right\}$$

and

$$\mathbf{L}_\psi^u(r_x,r_u) := \sup\left\{\|\mathrm{D}_u\psi(x_0,u_0) - \mathrm{D}_u\psi(x,u)\| \mid (x,u) \in \mathcal{B}_1\right\}.$$

Then for any given $r_x > 0$, $r_u > 0$ and $r_v > 0$ satisfying

$$\mathbf{L}_\psi^x(r_x)r_x + \mathbf{L}_\psi^u(r_x,r_u)r_u < \frac{r_u}{M_\psi^u} - r_x L_\psi^x - r_v, \text{ and} \tag{4.21a}$$

$$M_\psi^u \mathbf{L}_\psi^u(r_x,r_u) < 1 \tag{4.21b}$$

for each $(x,v) \in \mathsf{B}(x_0,r_x) \times \mathsf{B}(v_0,r_v)$ there exists a unique $u \in \mathsf{B}(u_0,r_u)$ satisfying

$$\psi(x,u) = v.$$

Further, define $\mathsf{B}(x_0,r_x) \times \mathsf{B}(v_0,r_v) \ni (x,v) \longmapsto \gamma(x,v){:=}u \in \mathsf{B}(u_0,r_u)$, then γ is an analytic map.

Propositions 4.19 and 4.20 are direct applications of bounds computed on the domain of IFT and ImFT. Note that, Propositions 4.19 and 4.20 provide the domain on which ϕ and ψ are well defined. However, for the control system (4.19) to be feedback linearizable one also needs to ensure that the trajectory belongs to $\mathcal{O}(x_0)$ (or equivalently in $\mathcal{O}(z_0)$) for all $k \in \mathbb{N}$. This limits the choice of the control input $v(k)$. From the linearized dynamics, we have

$$z(k+1) - z_0 = A(z(k) - z_0) + B(v(k) - v_0).$$

Taking norms, we have

$$\begin{aligned}\|z(k+1) - z_0\| &= \|A(z(k) - z_0) + B(v(k) - v_0)\| \\ &\le \|A\|\,\|z(k) - z_0\| + \|B\|\,\|v(k) - v_0\|.\end{aligned}$$

For a given $0 < r_z < P'_\phi$, for all $0 < \epsilon_v < \frac{1}{\|B\|}\left(P'_\phi - \|A\|\,r_z\right)$ we have $\|z(k+1) - z_0\| < r_z$. Thus, for all $z(k) \in \mathsf{B}(z_0,r_z)$ and $v(k) \in \mathsf{B}(v_0,\epsilon_v)$, $z(k+1) \in \mathsf{B}(z_0,r_z)$. Combining bounds from Proposition (4.19) and (4.20), along with ensuring that trajectories stay in $\mathcal{O}(z_0)$ we can find an $\mathcal{O}(x_0)$ and $\mathcal{O}(u_0)$ such that System (4.19) is feedback linearizable on $\mathcal{O}(x_0) \times \mathcal{O}(u_0)$.

Remark 4.21 The bounds are given by Propositions 4.19 and 4.20 utilize minimum information of ϕ and ψ. In particular, they only rely on the bounds on the first-order derivatives. This leads to often conservative estimates. However, the bounds can be significantly improved if we utilize the structure of ϕ and ψ as we demonstrate in the next example.

Example 4.22 Consider the following discrete system

$$\begin{pmatrix} x_1(k+1) \\ x_2(k+1) \end{pmatrix} = \begin{pmatrix} x_2(k) \\ (1+x_1(k))^2 u(k) \end{pmatrix} \tag{4.22}$$

with an equilibrium point at (0, 0, 0). Setting $\phi(x_1, x_2) = (x_1, x_2)$ and $\psi(x_1, x_2, u) = (1+x_1^2)u = v$, one can linearize (4.22) about (0, 0, 0). In order to compute the bounds on the domain of feedback linearizability of (4.22) we employ Propositions 4.19 and 4.20. Let us first compute bounds for ϕ. Since ϕ is the identity map, it is a global diffeomorphism on $\mathbb{R}^2$. A similar assertion is obtained from Proposition 4.19. From definition of ψ, we have

$$\mathrm{D}_u \psi(x_1, x_2, u) = (1 + x_1)^2 \tag{4.23}$$

and

$$\mathrm{D}_x \psi(x_1, x_2, u) = \begin{pmatrix} 2(1+x_1)u & 0 \end{pmatrix} \tag{4.24}$$

Computing quantities as defined in Proposition 4.20, we have $M_\phi^u := \|\mathrm{D}_u \psi(0, 0, 0)\| = 1$, $L_\phi^x := \|\mathrm{D}_x \psi(0, 0, 0)\| = 0$. For any given $r_x > 0, r_u > 0$, $\mathcal{B}_1 =: \mathsf{B}(x_0, r_x) \times \mathsf{B}(x_0, r_u)$, we have

$$\mathbf{L}_\psi^x(r_x) := \sup\{\|\mathrm{D}_x \psi(x_0, u_0) - \mathrm{D}_x \psi(x, u_0)\| \mid (x, u_0) \in \mathcal{B}_1\} = 0$$

and

$$\mathbf{L}_\psi^u(r_x, r_u) := \sup\{\|\mathrm{D}_x \psi(x_0, u_0) - \mathrm{D}_x \psi(x, u)\| \mid (x, u) \in \mathcal{B}_1\} = r_x(r_x + 2)$$

Substituting these quantities in to Eq. (3.1), we get that, for $r_x > 0, r_u > 0, r_v > 0$ satisfying

$$r_v < r_u(1 - r_x(r_x + 2)), \tag{4.25a}$$

$$(2 + r_x)r_x < 1, \tag{4.25b}$$

for each $(x, v) \in \mathsf{B}(x_0, r_x) \times \mathsf{B}(u_0, r_v)$, there exists a unique $u \in \mathsf{B}(u_0, r_u)$ such that $\psi(x, u) = v$. Therefore, (4.22) is feedback linearizable on $\mathsf{B}(x_0, r_x) \times \mathsf{B}(u_0, r_u)$. For $r_x = 0.2$, we have $r_v \leq 0.56 r_u$.

However, the bounds given by Eq. (4.25) are conservative in nature. From direct observation, for $x_1 \neq -1$, one has $u = v/(1 + x_1)^2$ and hence the map $\psi(x, \cdot)$ is globally invertible, further $\mathbb{R}\backslash\{-1\}\times\mathbb{R}\times\mathbb{R} \ni (x_1, x_2, v) \longmapsto \gamma(x, v) = v/(1+x_1)^2$ is a smooth map and thus (4.22) is feedback linearizable for all $(x_1, x_2, u) \in \mathbb{R}\backslash\{-1\}\times\mathbb{R}\times\mathbb{R}$.

5 Conclusion

In this article, we provide a lower bound on the domain of the applicability of the ImFT. We utilize degree theoretic ideas to come up with these estimates. One key advantage that is gained by utilizing the degree theoretic ideas is the applicability of these results. For $\mathcal{C}^2$ functions, the bounds are given as functions of the magnitude of the first-order derivatives evaluated over a point and the bounds on the second-order derivatives calculated over a region of interest. These ideas can be suitably extended to $\mathcal{C}^1$ and $\mathcal{C}^0$ maps. These bounds can be suitably extended to IFT and the bounds derived by Corollary 3.5 surpass those given by Abraham et al. [16].

The ImFT and IFT have several applications in system theory and control. In this article, we have addressed a few of them. We utilize these bounds to investigate the robustness of the solutions of the nonlinear equations with parametric variations. The method is adapted to address quadratically constrained quadratic programs (QCQPs). In control theory, the algebraic Riccati equation (ARE) can be formulated as a QCQP. We utilize the bounds on ImFT to compute the robustness of the solutions of the ARE under variations in the system matrix. This helps us compute bounds on system matrices so that the system remains stable. We also apply these bounds on the power flow equations to compute margins on the allowable power variations that ensure the stable operation of the power system network. We validate our results on the benchmark systems provided in the MatPower package of MATLAB software.

Another important application of these bounds is in the feedback linearization methods, where we use these bounds to come up with an estimate on the domains of the feedback linearizability of discrete-time systems.

Note that we do not claim presented bounds to be optimal, rather we tend to seek estimates that require minimal numerical computation and can be used in scenarios where limited computation capacity is available on board. As a future extension, one may look at improving these estimates by exploiting the structures of the underlying mapping f and finding different applications of these bounds in engineering problems.

Acknowledgements We would like to thank Professor Mayank Baranwal for his input and suggestions.

Author Contributions **Ashutosh Jindal** performed formal analysis, investigation, and writing—original draft; **Debasish Chatterjee** did conceptualization, methodology, and supervision; **Ravi Banavar** done supervision and critical analysis.

Funding The authors have no funding information to declare.

Availability of data and materials Not applicable.

Declarations

Ethical approval Not applicable.

Conflict of interest The authors declare that they have no competing interests whatsoever.

Appendix A: Some more results on the ImFT using degree theory

A.1 Showing [16, Proposition 2.5.4] as a corollary of Proposition 3.7

For $\mathcal{C}^2$ maps, Abraham et al. [16] provide explicit estimates on the size of the neighborhoods involved in the IFT. These bounds are given in the following proposition.

Proposition A.1 [16, Proposition 2.5.6] *Suppose* $f : U \subset E \longrightarrow F$ *is* $\mathcal{C}^r$, $r \geq 2$, $x_0 \in U$, *and* $\mathrm{D} f(x_0)$ *is an isomorphism. Let* $L = \|\mathrm{D} f(x_0)\|$ *and* $M = \left\|(\mathrm{D} f(x_0))^{-1}\right\|$. *Assume* $\left\|\mathrm{D}^2 f(x)\right\| \leq K$ *for all* $x \in \mathsf{B}(x_0, R) \subset U$, *for some* $R > 0$. *Define* $P := \min\{\frac{1}{2KM}, R\}$ *and* $P' = \frac{P}{2M}$. *Further, let* $N = 8M^3K$ *and* $Q := \min\{\frac{1}{2NL}, \frac{P}{2M}, P\}$ *and* $Q' := \frac{Q}{2L}$. *Then there exist*

(A.1a) *an open set* $H \subset \mathsf{B}(x_0, P)$ *such that* f *maps* H *diffeomorphically onto* $\mathsf{B}\left(x_0, P'\right)$ *and*

(A.1b) *an open set* $H' \subset \mathsf{B}(f(x_0), Q)$ *such that* f^{-1} *maps* H' *diffeomorphically onto* $\mathsf{B}\left(x_0, Q'\right)$.

Proposition A.1 can be shown as a corollary of Proposition 3.7 as we now demonstrate.

Proof of Proposition A.1 Define $(x, y) \longmapsto F(x, y) = f(x) - y$. Then we have $F(x, y) = 0 \iff f(x) = y$. Furthermore, $\mathrm{D}_x F(x_0, f(x_0)) = \mathrm{D} f(x_0)$ is nonsingular. Therefore F satisfies assumptions of Proposition 3.7. Define $M = \left\|\mathrm{D} f(x_0)^{-1}\right\|$ and let $R > 0$ be such that $\mathsf{B}(x_0, R) \subset U$ and define

$$K = \sup\left\{\left\|\mathrm{D}^2 f(x)\right\| \mid x \in \mathsf{B}(x_0, R)\right\}.$$

For a given $0 < r_x < R$ and r_y, define

$$g_1(r_x) = K r_x \text{ and } g_2(r_y) = r_y.$$

Then for all $(x, y) \in \operatorname{cl} \mathsf{B}(x_0, r_x) \times \mathsf{B}\left(y_0, r_y\right)$, we have

$$\|\mathrm{D}_x F(x, y) - \mathrm{D}_x F(x_0, y_0)\| = \|\mathrm{D} f(x) - \mathrm{D} f(x_0)\| \leq g_1(r_x), \text{ and}$$
$$\|F(x_0, y)\| = \|y_0 - y\| \leq r_y.$$

Setting $\alpha = 0.5$, for all r_x, r_y satisfying

$$MKr_x \leq 0.5 \text{ and } Mr_y \leq 0.5 r_x$$

For each $y \in \mathsf{B}\left(y_0, r_y\right)$, there exists a unique x_y in $\operatorname{cl} \mathsf{B}(x_0, r_x)$ satisfying $f(x_y) = y$. Defining $y \longmapsto g(y) := x_y$, provides the local inverse, further g is $\mathcal{C}^2$ in view of

Theorem 3.2. Setting $r_x = P$, $r_y = P'$, and $H = g(\mathsf{B}\left(y_0, P'\right))$ proves (A.1a). From the expression of $\mathrm{D}f^{-1}(y)$, we have

$$\sup\left\{\mathrm{D}^2 f^{-1}(y) \mid y \in \mathsf{B}\left(y_0, P'\right)\right\} \le 8M^3K = N.$$

Therefore, replacing f with g, M with L and K with N we get $Q = \min\{\frac{1}{2NL}, \frac{P}{2M}, P\}$ and $Q' = \frac{Q}{2L}$, such that for each $x \in \mathsf{B}\left(x_0, Q'\right)$ there exists a $y_x \in \mathrm{cl}\,\mathsf{B}\,(y_0, Q)$ such that $f(x) = y_x$. Defining $H' = f(\mathsf{B}\left(x_0, Q'\right))$ proves (A.1b). □

A.2 Generalized ImFT for $\mathcal{C}^0$ maps

A generalized version of ImFT for continuous maps is presented by Halkin [34]. We can extend Proposition 3.7 to obtain estimates for such maps as follows.

Proposition A.2 *Let $\Omega \subset \mathbb{R}^n \times \mathbb{R}^m$ be open and $(x_0, y_0) \in \Omega$. Let $\Omega \ni (x, y) \longmapsto P(x, y)$ be a continuous map. Assume:*

(i) $P(x_0, y_0) = 0$*;*

(ii) *P is differentiable with respect to y at (x_0, y_0) and $\mathrm{D}_y P(x_0, y_0)$ has a bounded inverse Γ and $\|\Gamma\| = k_1$.*

(iii) $S = \{(x, y) \in \mathsf{B}\,(x_0, \delta) \times \mathrm{cl}\,\mathsf{B}\,(y_0, \epsilon)\} \subset \Omega.$

(iv) *there exists a real-valued function $[0, \delta] \times [0, \epsilon] \ni (u, v) \longmapsto g_3(u, v)$ such that g_3 is nondecreasing in each argument with the other fixed and for all $(x, y) \in S$*

$$\left\| P(x, y) - (P(x, y_0) + \mathrm{D}_y P(x_0, y_0) \cdot (y - y_0))\right\| \le g_3(\|x - x_0\|, \|y - y_0\|);$$

(v) *there is a nondecreasing function $[0, \delta] \ni x \longmapsto g_2(x)$ such that for all $(x, y_0) \in S$*

$$\|P(x, y_0)\| \le g_2(\|x - x_0\|);$$

(vi) $k_1(g_2(\delta) + g_3(\delta, \epsilon)) \le \epsilon.$

Then for all $x \in \mathsf{B}\,(x_0, \delta)$ there exists a y_x (not necessarily unique) in $\mathrm{cl}\,\mathsf{B}\,(y_0, \epsilon)$ such that $P(x, y_x) = 0$.

Proof Existence of such a g_1, g_2, $\epsilon > 0$ and $\delta > 0$ satisfying (i)–(vi) is ensured from the continuity of P and properties of the derivative $\mathrm{D}_y P(x_0, y_0)$.

Fix an $x \in \mathsf{B}\,(x_0, \delta)$ and define

$$\mathrm{cl}\,\mathsf{B}\,(y_0, \epsilon) \ni y \longmapsto P(y; x) := P(x, y) - P(x_0, y_0). \tag{A1}$$

then $P(x, y) = 0 \iff P(y; x) = P(x_0, y_0) = 0$. If $P(\cdot; x)$ vanishes on $\mathrm{bd}\,\mathsf{B}\,(y_0, \epsilon)$ then there is nothing to prove. Otherwise, assume $P(\cdot; x)$ is nonvanishing for all $y \in \mathrm{bd}\,\mathsf{B}\,(y_0, \epsilon)$. Define an approximation of $P(\cdot; x)$ as

$$\mathrm{Aff}(P)(y; x) = P(x, y_0) + \mathrm{D}_y P(x_0, y_0) \cdot (y - y_0). \tag{A2}$$

Using (ii), (v) and A2, for all $y \in \operatorname{bd} \mathsf{B}(y_0, \epsilon)$, we have

$$\|\mathrm{Aff}(P)(y; x)\| \leq \frac{\epsilon}{k_1} - g_2(\delta).$$

Furthermore, from (iv), for all $y \in \operatorname{bd} \mathsf{B}(y_0, \epsilon)$, we have

$$\|P(y; x) - \mathrm{Aff}(P)(y; x)\| \leq g_3(\epsilon, \delta).$$

From (vi), for all $y \in \operatorname{bd} \mathsf{B}(y_0, \epsilon)$ we have

$$\|P(y; x) - \mathrm{Aff}(P)(y; x)\| \leq \|\mathrm{Aff}(P)(y; x)\|.$$

Furthermore, $\mathrm{Aff}(P)(\cdot; x)$ and $P(\cdot; x)$ are nonvanishing on $\operatorname{bd} \mathsf{B}(y_0, \epsilon)$. Therefore, from Corollary 2.6 we have

$$\mathrm{Deg}\,(P(\cdot; x), \mathsf{B}(y_0, \epsilon)) = \mathrm{Deg}\,(\mathrm{Aff}(P)(\cdot; x), \mathsf{B}(y_0, \epsilon)) \in \{-1, 1\},$$

and thus there exists a (not necessarily unique) $y_x \in \operatorname{cl} \mathsf{B}(y_0, \epsilon)$ such that $P(y_x; x) = P(x, y_x) = 0$. □

References

1. Krantz SG, Parks HR (2002) The implicit function theorem: history, theory, and applications. In: Modern Birkhäuser classics. Springer
2. Bertsekas DP, Hager W, Mangasarian O (1999) Nonlinear programming. Athena Scientific, Belmont
3. Dontchev AL (1999) Lipschitzian stability of Newton's method for variational inclusions. In: IFIP conference on system modeling and optimization. Springer, pp 119–147
4. Fletcher R (2000) Practical methods of optimization, 2nd edn. Wiley, Chichester
5. Mordukhovich B (1994) Lipschitzian stability of constraint systems and generalized equations. Nonlinear Anal Theory Methods Appl 22(2):173–206
6. Nash SG, Sofer A (1996) Linear and nonlinear programming. McGraw-Hill Science, Engineering and Mathematics, New York
7. Jindal A, Banavar R, Chatterjee D (2021) Feedback linearization of implicit discrete systems. IFAC-PapersOnLine 54(19):340–345. https://doi.org/10.1016/j.ifacol.2021.11.100. 7th IFAC Workshop on Lagrangian and Hamiltonian Methods for nonlinear control LHMNC 2021
8. Eldering J (2013) Normally hyperbolic invariant manifolds: the noncompact case. In: Atlantis series in dynamical systems, vol 2. Atlantis Press, Paris
9. Duistermaat JJ, Kolk JA (2000) Lie groups. Universitext, Springer
10. Bressan A, Piccoli B (2007) Introduction to the mathematical theory of control. American Institute of Mathematical Sciences
11. Isidori A (1985) Nonlinear control systems. Springer
12. Holtzman JM (1970) Explicit ε and δ for the implicit function theorem. SIAM Rev 12(2):284–286
13. Chang H, He W, Prabhu N (2003) The analytic domain in the implicit function theorem. J Inequal Pure Appl Math 4(1)
14. Ash RB (2014) Complex variables. Academic Press
15. Papi M (2005) On the domain of the implicit function and applications. J Inequal Appl 2005(3):1–14
16. Abraham R, Marsden JE, Ratiu T (2007) Manifolds, tensor analysis, and applications. In: Applied mathematical sciences. Springer
17. Zeidler E (1993) Nonlinear functional analysis and its applications, vol I: Fixed—point theorems/Transl. by Peter R. Wadsack. Springer

18. Outerelo E, Ruiz JM (2009) Mapping degree theory, vol 108. American Mathematical Society
19. Dinca G, Mawhin J (2021) Brouwer degree : the core of nonlinear analysis. Springer
20. Krasnosel'skij MA, Zabrejko PP (1984) Geometrical methods of nonlinear analysis, vol 263. Springer
21. Zeidler E (1995) Applied functional analysis: applications to mathematical physics. In: Applied mathematical sciences, vol 108. Springer
22. Spivak M (1995) Calculus on manifolds: a modern approach to classical theorems of advanced calculus. CRC Press
23. Clarke F (1976) On the inverse function theorem. Pac J Math 64(1):97–102
24. Lang S (1997) Undergraduatae analysis. In: Undergraduate texts in mathematics. Springer
25. Rudin W (1976) Principles of mathematical analysis, 3rd edn. McGraw-Hill, New York
26. Dvijotham K, Krishnamoorthy B, Luo Y, Rapone B (2023) Robust feasibility of systems of quadratic equations using topological degree theory. Optim Lett. https://doi.org/10.1007/s11590-023-02015-7
27. Dvijotham K, Nguyen H, Turitsyn K (2017) Solvability regions of affinely parameterized quadratic equations. IEEE Control Syst Lett 2(1):25–30
28. Zimmerman RD, Murillo-Sánchez CE, Thomas RJ (2010) Matpower: steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans Power Syst 26(1):12–19
29. Chern MY, Murata T (1983) Fast algorithm for concurrent LU decomposition and matrix inversion
30. Martinsson PG, Rokhlin V, Tygert M (2005) A fast algorithm for the inversion of general Toeplitz matrices. Comput Math Appl 50(5):741–752. https://doi.org/10.1016/j.camwa.2005.03.011
31. Li S (2009) Fast algorithms for sparse matrix inverse computations. Ph.D. Thesis. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated-2023-03-02. https://www.proquest.com/dissertations-theses/fast-algorithms-sparse-matrix-inverse/docview/305001365/se-2
32. Liu J, Liang Y, Ansari N (2016) Spark-based large-scale matrix inversion for big data processing. IEEE Access 4:2166–2176. https://doi.org/10.1109/ACCESS.2016.2546544
33. Liberzon D (2011) Calculus of variations and optimal control theory: a concise introduction. Princeton University Press
34. Halkin H (1974) Implicit functions and optimization problems without continuous differentiability of the data. SIAM J Control 12(2):229–236. https://doi.org/10.1137/0312017