INFLATION OF POORLY CONDITIONED ZEROS OF SYSTEMS OF ANALYTIC FUNCTIONS

MICHAEL BURR AND ANTON LEYKIN

ABSTRACT. Given a system of analytic functions and an approximate zero, we introduce inflation to transform this system into one with a regular quadratic zero. This leads to a method for isolating a cluster of zeros of the given system.

Let $g : \mathbb{C}^n \to \mathbb{C}^n$ be a square system of analytic functions. There exist heuristic methods to obtain approximations to some or all of the isolated zeros of g. In the context of our work, isolation of a zero is the problem of constructing a neighborhood of the given approximation that contains exactly one zero. When the approximation is close to a singular zero (with multiplicity m) of a nearby system, this problem becomes the problem of isolating a cluster of m zeros.

For regular zeros, methods derived from alpha-theory [4, 9] or interval arithmetic [11] provide effective ways to isolate zeros. These methods, however, do not apply to singular zeros. We introduce an approach, called inflation, that transforms a system with an approximate zero into a system with a regular quadratic zero — a new concept which we define. We use this new system to compute a region containing a cluster of zeros of the original system.

Most prior approaches for approximating poorly conditioned (i.e., singular or nearly singular) zeros, such as deflation [10], aim to *reduce* the multiplicity of a zero so that the resulting zeros can be found using standard methods. One new idea of our approach is, instead, to *increase* the multiplicity of the zero or cluster of zeros. By increasing the multiplicity of the zero, we regularize the zero, albeit not in a traditional sense — regular quadratic zeros are still singular, but better behaved. Our approach combines ideas from linear and nonlinear algebra, real and complex analysis, as well as convex optimization.

We refer the reader to [8] and the references therein for a discussion of other approaches to the zero clustering problem and the related problem of certifying a singular zero. To the best of our knowledge, prior to this work, the clustering problem has only been considered in the univariate case and the multivariate case where the Jacobian has corank one.

Date: October 28, 2020.
Burr was partially supported by NSF DMS #1913119.
Leykin was partially supported by NSF DMS #1719968 and #2001267.

1. Exact singular zeros

We call x a regular zero of order d of a system g if, first, the degree i part of the Taylor expansion of g at x is identically zero for $i = 0, \ldots, d-1$, and, second, the degree d part does not vanish on the unit sphere $\mathbb{S} \subset \mathbb{C}^n$. We call a regular zero of order 2 a regular quadratic zero.

Let $y \in \mathbb{C}^n$ be a zero of g and $\kappa = \dim \ker Dg(y)$. If $\kappa = 0$, then the zero is a regular zero, or, in our new terminology, regular of order one. The most common case of a multiple (nonregular and singular) zero is called a simple double zero in [7] and corresponds to a singularity of Type $\Sigma^{1,0}$ using the terminology of singularity theory [1]. We consider the more general case where y is a singularity of Type Σ^κ. We define the multiplicity of an isolated zero y to be the number of zeros near y after a generic perturbation of f [1].

We introduce inflation as a way to transform a general system with a singularity of Type Σ^κ at y into a system with a regular quadratic zero at y. Next we apply Rouché's theorem to the quadratic part of the transformed system in order to calculate the multiplicity of the zero y of the original system. Throughout this section, we assume that the singular zero y is explicitly known, and, hence, the kernel $\ker Dg(y)$ can be explicitly computed. This restriction is lifted in Section 2.

1.1. Inflation.

We first simplify the problem by applying an affine change of coordinates $A : \mathbb{C}^n \to \mathbb{C}^n$ with $A(0) = y$ so that the origin is a zero of $g \circ A$ and

$$\ker D(g \circ A)(0) = \langle e_1, \ldots, e_\kappa \rangle.$$

We then obtain the inflated system $g \circ A \circ S_\kappa$ where

$$S_\kappa(x_i) = \begin{cases} x_i, & i \in [\kappa]; \\ x_i^2, & \text{otherwise.} \end{cases}$$

Since g is generic, the quadratic part of $g \circ A \circ S_\kappa$ does not vanish on the unit sphere, so $g \circ A \circ S_\kappa$ has a regular quadratic zero at the origin. We remark that for $\kappa < n$, the resulting zero has higher multiplicity than the original zero. More precisely, from the point of view of singularity theory, we pass from Type Σ^κ to Type Σ^n. This behavior prompted us to name this method "inflation" in contrast to "deflation", see, e.g., [10].

1.2. Isolating zeros.

Let $\mathbb{S}_\varepsilon \subset \mathbb{C}^n$ be the sphere of radius ε and $\mathbb{B}_\varepsilon$ be the ball of radius ε centered at 0 in the Euclidean norm.

Lemma 1.1. *Let f be an arbitrary square system and Q a square system of quadratic forms. Suppose there exists $c > 0$ and $\varepsilon > 0$ such that*

(1) $\|(f - Q)(\mathbb{S}_\varepsilon)\| \leq c\varepsilon^2$, *and*
(2) $\|Q(\mathbb{S}_1)\| \geq c$,

then f has 2^n zeros in $\mathbb{B}_\varepsilon$.

Proof. Condition 2 implies that Q has no zeros on a sphere. Therefore, by homogeneity, its only zero is the origin, which has multiplicity 2^n. Since

$\|f-Q\| \leq c\varepsilon^2 \leq \|Q\|$ on $\mathbb{S}_\varepsilon$, the conclusion follows from multivariate version Rouché's theorem, see [2, Theorem 2.12]. □

Since $g \circ A \circ S_\kappa$ has a regular quadratic zero at the origin, we define $Q = Q(g)$ to be the quadratic part of $g \circ A \circ S_\kappa$. For ε a small positive number, the conditions of Lemma 1.1 hold for the pair $(g \circ A \circ S_\kappa, Q)$.

Since the inflation map is a cover with $2^{n-\kappa}$ sheets with ramification locus $\{x_1 x_2 \dots x_\kappa = 0\}$, we conclude that y is a zero of multiplicity 2^κ of g.

We illustrate our technique when the zero of g is explicitly known:

Example 1.2. *Let $\alpha \neq 0$ and g be the following system with a zero at (y_1, y_2):*

$$g(x) = [(x_1 - y_1)^2, (x_2 - y_2) - \alpha(x_1 - y_1)^2((x_1 - y_1) + 3y_1)]^T.$$

We observe that $\ker Dg(y) = \langle e_1 \rangle$ *and* $\kappa = 1$. *Applying the affine transformation $A(x) = x + y$ and inflation, we get*

$$g \circ A \circ S_1(x) = [x_1^2, x_2^2 - \alpha x_1^2(x_1 + 3y_1)]^T.$$

We observe that the quadratic part of $g \circ A \circ S_1$ is

$$Q(g \circ A \circ S_1) = [x_1^2, x_2^2 - 3\alpha y_1 x_1^2]^T,$$

which does not vanish on the unit sphere $\mathbb{S}_1$. Therefore, $g \circ A \circ S_1$ is a regular quadratic zero of multiplicity four at the origin. Near the origin, $Q(g \circ A \circ S_1)$ dominates $g \circ A \circ S_1 - Q(g \circ A \circ S_1)$.

Using Rouché's theorem [2, Theorem 2.12]*, we conclude that $g \circ A \circ S_1$ has four zeros (counted with multiplicity) in the ball $\mathbb{B}_\varepsilon$ for some $\varepsilon > 0$. Since $A \circ S_1$ is a double cover, we conclude that g has a zero of multiplicity* two *at y.*

We hint at how the computation from Example 1.2 can be used to isolate a cluster of zeros. Let $f(x) = [x_1^2 - \alpha^2, x_2 - \alpha x_1^3]^T$, which has zeros at $\pm(\alpha, \alpha^4)$. Suppose that we approximate this cluster of zeros with y. We observe that $f - f(y)$ differs from g in Example 1.2 only in the linear terms in the Taylor expansion around y. Moreover, the limit of this difference tends to 0 as $\alpha \to 0$. This behavior should be contrasted with that of the remaining terms contributing to $Q(g \circ A \circ S_1)$, which tend to a quadratic form that is nonvanishing on a sphere. Therefore, for y and α sufficiently small, $f - f(y)$ is dominated by Q on $\mathbb{S}_\varepsilon$. Thus, Rouché's theorem identifies a cluster of zeros of f near y.

In the remainder of this paper, we expand upon this approach.

2. Poorly conditioned zeros

In typical applications, for system f, we do not expect that a multiple zero is given or even that f has any singular zeros. Instead, we assume that we are given an approximation y to a cluster of zeros so that f is nearly singular at y. In addition, let κ be the dimension of the numerical kernel of $Df(y)$, i.e., the number of small singular values. We construct a nearby

system g which has a zero at y and whose Jacobian has corank κ. Then, by applying the methods from Section 1, we isolate a cluster of zeros of f.

2.1. **Transformations of the system.** Suppose that $V \subset \mathbb{C}^n$ is a linear subspace with $\dim V = \kappa$. Consider

$$g(x) := f(x) - f(y) - \sum_{i=1}^{\kappa} Df(y)\pi_V(x-y) \tag{1}$$

where π_V is the orthogonal projection onto V. We observe that y is a zero of g and that $\ker Dg(y) \supset V$. When $Df(y)$ is nonsingular, this containment is an equality. In our computations, V plays the role of an approximate kernel of $Df(y)$. One practical choice for V is the span of singular vectors corresponding to the κ smallest singular values of $Df(y)$.

2.2. **Isolating a cluster of zeros.** Suppose that g is constructed from f as in Section 2.1 and that A is constructed from g as in Section 1.1. We then compare the zeros of $f \circ A \circ S_\kappa$ to those of $Q(g \circ A \circ S_1)$ using Lemma 1.1.

Define $\mathbb{B}_\varepsilon^{(\kappa)} \subset \mathbb{C}^n$ to be the ball of radius ε in the mixed norm, i.e.,

$$\|x\|_{(\kappa)} = \sqrt{\sum_{i=1}^{\kappa} |x_i|^2 + \sum_{i=\kappa+1}^{n} |x_i|}.$$

The inflation map S_κ restricted to $\mathbb{B}_\varepsilon$ is a cover of $\mathbb{B}_\varepsilon^{(\kappa)}$ with $2^{n-\kappa}$ sheets and ramification locus $\{x_1 x_2 \dots x_\kappa = 0\}$. This construction leads to the following:

Corollary 2.1. *Let f be a square system, $y \in \mathbb{C}^n$, and $V \subset \mathbb{C}^n$ be a linear subspace of dimension κ. Suppose that g is constructed as in Section 2.1. Let A be an affine map such that $g \circ A(0) = 0$ and $\ker D(g \circ A(0)) = \langle e_1, \dots, e_\kappa \rangle$. Let $Q = Q(g \circ A \circ S_\kappa)$ be the quadratic part of $g \circ A \circ S_\kappa$. If $f \circ A \circ S_\kappa$ and Q satisfy the assumptions of Lemma 1.1, then f has 2^κ zeros in $A(\mathbb{B}_\varepsilon^{(\kappa)})$.*

2.3. **Example.** Let $\alpha = 0.01$ in the system $f(x) = [x_1^2 - \alpha^2, x_1 + x_2 - \alpha x_1^3]^T$. Let $y := (0.001, -0.001)$ be a rough approximation to the center of the cluster of zeros of f. Our computation allows us to isolate the cluster of two zeros as depicted in Figure 1.

The singular values at y are 1.414 and 0.001414. Moreover, the singular vector corresponding to the smaller singular value is $v := [-0.707, 0.707]^T$. We then construct g as

$$\begin{aligned} g(x) &= f(x) - f(y) - Df(y)\pi_v(x,y) \\ &\approx [x_1^2 - 10^{-3}x_1 - 10^{-3}x_2 + 10^{-6}, -0.01x_1^3 + x_1 + x_2 - 2 \cdot 10^{-9}]^T. \end{aligned}$$

Using the affine transformation A, whose constant part is y and whose linear part is both orthogonal and sends v to e_1, we have (after suppressing terms

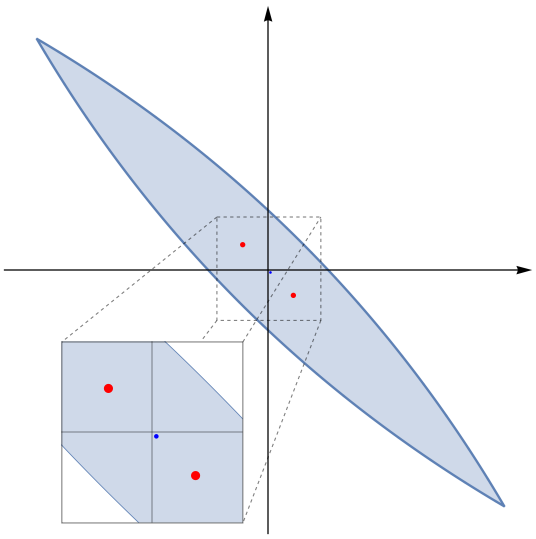

FIGURE 1.
The zeros of $f_{0.01} = 0$ are the red points. An approximation $y = (0.001, -0.001)$ to the center of the cluster is the blue point. The shaded region is $A(\mathbb{B}_{0.17}^{(\kappa)})$, which isolates the zeros of $f_{0.01}$.

which are nonzero only due to round-off error),

$$g \circ A \approx \begin{bmatrix} 0.5x_1^2 + x_1x_2 + 0.5x_2^2 - 0.001414x_2 \\ 0.003536x_2^3 + 0.01061x_1^2x_2 + 0.01061x_1x_2^2 + 0.003536x_1^3 \\ -1.5 \cdot 10^{-5}x_1^2 - 3 \cdot 10^{-5}x_1x_2 - 1.5 \cdot 10^{-5}x_2^2 - 1.414x_2 \end{bmatrix}.$$

We observe that in this system, there are no constant terms and only x_2 appears as a linear term. We apply the inflation map S_1, which replaces x_2 by its square, and extract the quadratic part

$$Q = Q(g \circ A \circ S_1) \approx [0.5x_1^2 - 0.001414x_2^2, -1.5 \cdot 10^{-5}x_1^2 - 1.414x_2^2]^T$$

Using a sum of squares computation, see Section 3.1 for details, we find that a lower bound on $q = \|Q\|^2$ on the unit sphere is 0.2221. Hence, in Lemma 1.1, $c \approx 0.4713$.

We now compute ε for Condition 1 of Lemma 1.1 as follows (omitting terms with coefficients less than 10^{-16}):

$$f \circ A \circ S_1 - Q \approx \begin{bmatrix} 0.5x_2^4 + x_1x_2^2 - 0.001414x_1 - 10^{-4} \\ 0.003536x_2^6 + 0.01061x_1x_2^4 + 0.01061x_1^2x_2^2 - 1.5 \cdot 10^{-5}x_2^4 \\ +0.003536x_1^3 - 3 \cdot 10^{-5}x_1x_2^2 + 1.414 \cdot 10^{-6}x_1 - 10^{-11} \end{bmatrix}.$$

Via crude upper bounds, namely, by using the triangle inequality and bounding the magnitude of all variables from above by ε, we find that Condition 1 of Lemma 1.1 holds for $\varepsilon \in [0.017, 0.39]$. Therefore, by Corollary 2.1, f has 2 zeros in $A(\mathbb{B}_{\varepsilon}^{(1)})$.

3. Lower bounds on quadratic forms

A key step in the application of our method is finding a $c > 0$ so that Condition 2 of Lemma 1.1 holds. Computing c is a polynomial optimization problem.

Let $Q = (Q_1, \ldots, Q_n)$ be a vector of quadratic forms. Condition 2 of Lemma 1.1 requires a positive lower bound on $\|Q\|$ on the unit sphere. We use a sum-of-squares certificate of positivity approach to compute this bound, but also discuss alternatives. A full study of theoretical efficiency as well as practicality of these approaches is left to future work.

3.1. **Sum-of-squares relaxation.** We identify $\mathbb{C}^n$ with the real space $\mathbb{R}^{2n}$ of twice as many dimensions. For each variable x_i, we let $a_i = \mathrm{Re}(x_i)$ and $b_i = \mathrm{Im}(x_i)$. We are interested in a positive lower bound on the following quartic over the real unit sphere $\mathbb{S}_1 \subset \mathbb{R}^{2n}$:

$$q(a,b) = \|Q(a+\mathbf{i}b)\|^2 = \sum_{j=1}^{n} \mathrm{Re}(Q_j(a+\mathbf{i}b))^2 + \mathrm{Im}(Q_j(a+\mathbf{i}b))^2.$$

In addition, we define

$$s(a,b) = \|a+\mathbf{i}b\|^2 = \sum a_i^2 + \sum b_i^2,$$

where the real zero set of $s(a,b)-1$ is the unit sphere. Our goal is to compute

$$q^* = \min_{\mathbb{S}_1} q = \min\{q(a,b) \mid s(a,b) = 1\}.$$

Let Σ_{SOS} be the convex cone of polynomials that are sums of squares. The sums-of-squares moment relaxation of level r computes a lower bound on q^*, see, e.g., [3, §3.2.4] and [13]. In our case, this bound is

$$c_r = \max\{c \mid q - c = \sigma + h,\ \sigma \in \Sigma_{SOS},\ h \in (s-1),\ \max(\deg\sigma, \deg h) \le 2r\}.$$

Since $\lim_{r\to\infty} c_r = q^*$, for r large enough, $c_r > 0$ when $q^* > 0$. We remark that, should a relaxation of level r succeed with $c_r > 0$, a rational certificate of positivity for q can be produced as in [12].

3.2. **Alternative approaches.** Isolating all critical points of q on the sphere $\mathbb{S}_1$ is one alternate approach. This can be accomplished provided the zeros of a Lagrange multiplier system are regular and their number matches the generic number of zeros of this system. In full generality, the zeros may be singular, "at infinity", or poorly conditioned, which would hamper the practicality of this approach.

Interval arithmetic, see [11], may be used to estimate a lower bound on q over the sphere $\mathbb{S}_1$. This method estimates q on each region of a partition of the sphere. As the partition becomes finer, this estimate converges to q^*. Subdivision-based schemes involving interval arithmetic sometimes have exponential complexities, see [5], but may be efficient in practice [6].

4. Summary and remarks

Given a heuristic to determine an approximate rank and an approximate kernel of linear map, we summarize our approach below.

Input: A polynomial system $f : \mathbb{C}^n \to \mathbb{C}^n$ and a point $y \in \mathbb{C}^n$ that approximates a cluster of zeros.
Output: An integer $0 \leq \kappa \leq n$, a real number $\varepsilon > 0$, and the image of a mixed-norm ball $\mathbb{B}_\varepsilon^{(\kappa)}$ that contains 2^κ zeros of f.

(1) Compute κ, the dimension of an approximate kernel of $Df(y)$.
(2) Use Equation (1) to construct the system g which has a zero at y of Type Σ^κ.
(3) Construct an affine transformation A so that $g \circ A$ has a zero at 0 and the kernel of $g \circ A$ at 0 is $\langle e_1, \ldots, e_\kappa \rangle$.
(4) Use the inflation operator S_κ so that $g \circ A \circ S_\kappa$ has no constant or linear terms.
(5) Find a lower bound $c > 0$ for the norm of the quadratic part Q of the system $g \circ A \circ S_1$ on the unit sphere using an approach from Section 3.
(6) Find a small $\varepsilon > 0$ so that $\|f - Q\| < c\varepsilon^2$ on the sphere $\mathbb{S}_\varepsilon$.
(7) Return κ, A, and ε.

4.1. **Application to systems with exact singular zeros.** Suppose f has an exact zero x of singularity Type Σ^κ inflatable to a regular quadratic zero. As long as there is an efficient way to refine the approximation y of x, our approach is guaranteed to succeed. Indeed, one can use SVD to recover κ and and an approximation of $\ker Df(x)$ robustly, since exactly κ singular values approach 0 as $y \to x$.

This procedure does not certify the fact that there is a singular zero in the constructed region, but it isolates and certifies the existence of a cluster of multiplicity equal to that of x.

4.2. **Approximating the cluster.** While we consider computing an approximation y to a cluster as a separate problem, one algorithm we have in mind is that of deflation, as described in [10]. This approach provides a way to refine an approximation to an exact singular zero as required in Section 4.1. In the case of a cluster of several distinct zeros, deflation still delivers an augmented (often overdetermined) system of equations that can be used for our purpose. The fixed point of the operator in Newton's method is a good heuristic approximation of the cluster even if this fixed point is not an actual zero of the augmented system.

4.3. **Geometry of a cluster.** Inflation changes the geometry around a cluster. As a result, the mixed-norm ball of Corollary 2.1 approximates the geometry of a cluster better than a Euclidean ball, as in the cluster isolation statement in [7, Theorem 4]. It may be possible to pursue a strategy similar to that of [7] in the case of a regular quadratic zero. We remark that even

if this strategy succeeds, approaches inspired by Smale's α-theory tend to provide very conservative bounds.

4.4. **Reduction to regular zeros of higher order.** Our isolation technique can be readily adopted to the case of a regular zero of arbitrary order. In fact, our method has a straightforward application to simple multiple zeroes of arbitrary multiplicity as studied, for instance, in [8]. In this case, $\kappa = 1$ and one may generalize inflation to produce a system with a regular zero of order μ when the multiplicity μ is known. This construction would also result in a mixed norm ball that may give a tighter isolation region than the Euclidean ball produced by the method of [8].

A system with a singular zero with arbitrary multiplicity structure could be more complicated, however. We suspect that a more intricate inflation procedure may be needed in the general case.

4.5. **Applications to square polynomial systems.** Given a system of n polynomial equations in n unknowns, it is common to have both a heuristic method to approximate all zeros and a way to count them with multiplicity. When, in addition, all zeros are either regular or inflatable to regular quadratic zeros, our approach certifies the completeness of the approximate computation by producing a computer derived, but explicitly checkable proof.

References

[1] Valadimir I. Arnold, Sabir M. Gusein-Zade, and Aleksadr N. Varchenko. *Singularities of differentiable maps. Volume 1.* Modern Birkhäuser Classics. Birkhäuser/Springer, New York, 2012.

[2] Carlos A. Berenstein, Alekos Vidras, Roger Gay, and Alain Yger. *Residue Currents and Bezout Identities.* Birkhäuser Basel, Basel, 1993.

[3] Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite optimization and convex algebraic geometry.* SIAM, Philadelphia, PA, 2012.

[4] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale. *Complexity and Real Computation.* Springer-Verlag New York, 1998.

[5] Michael Burr, Shuhong Gao, and Elias Tsigaridas. The complexity of subdivision for diameter-distance tests. *Journal of Symbolic Computation*, 101:1–27, 2020.

[6] Felipe Cucker, Alperen A. Ergür, and Josue Tonelli-Cueto. Plantinga-Vegter algorithm takes average polynomial time. In *ISSAC'19—Proceedings of the 2019 ACM International Symposium on Symbolic and Algebraic Computation*, pages 114–121. ACM, New York, 2019.

[7] Jean-Pierre Dedieu and Mike Shub. On simple double zeros and badly conditioned zeros of analytic functions of n variables. *Mathematics of computation*, pages 319–327, 2001.

[8] Zhiwei Hao, Wenrong Jiang, Nan Li, and Lihong Zhi. On isolation of simple multiple zeros and clusters of zeros of polynomial systems. *Mathematics of Computation*, 89(322):879–909, 2020.

[9] Jonathan D. Hauenstein and Frank Sottile. Algorithm 921: Alphacertified: Certifying solutions to polynomial systems. *ACM Transactions on Mathematical Software*, 38(4):1–20, 2012.

[10] Anton Leykin, Jan Verschelde, and Ailing Zhao. Newton's method with deflation for isolated singularities of polynomial systems. *Theoretical Computer Science*, 359(1-3):111–122, 2006.

[11] Ramon E. Moore, R. Baker Kearfott, and Michael J. Cloud. *Introduction to interval analysis*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.
[12] Helfried Peyrl and Pablo A Parrilo. Computing sum of squares decompositions with rational coefficients. *Theoretical Computer Science*, 409(2):269–281, 2008.
[13] Konrad Schmüdgen. The K-moment problem for compact semi-algebraic sets. *Mathematische Annalen*, 289(2):203–206, 1991.

Clemson University, 220 Parkway Drive, Clemson, SC 29634
Email address: burr2@clemson.edu

School of Mathematics, Georgia Tech, Atlanta, Georgia
Email address: leykin@math.gatech.edu