Math. Program., Ser. A (2017) 164:229–243
DOI 10.1007/s10107-016-1083-6

FULL LENGTH PAPER

Variational analysis of the Crouzeix ratio

Anne Greenbaum[1] · **Adrian S. Lewis**[2] ·
Michael L. Overton[3]

Received: 22 January 2016 / Accepted: 16 October 2016 / Published online: 2 November 2016

Abstract Let $W(A)$ denote the field of values (numerical range) of a matrix A. For any polynomial p and matrix A, define the Crouzeix ratio to have numerator $\max\{|p(\zeta)| : \zeta \in W(A)\}$ and denominator $\|p(A)\|_2$. Crouzeix's 2004 conjecture postulates that the globally minimal value of the Crouzeix ratio is $1/2$, over all polynomials p of any degree and matrices A of any order. We derive the subdifferential of this ratio at pairs (p, A) for which the largest singular value of $p(A)$ is simple. In particular, we show that at certain candidate minimizers (p, A), the Crouzeix ratio is (Clarke) regular and satisfies a first-order nonsmooth optimality condition, and hence that its directional derivative is nonnegative there in every direction in polynomial-matrix space. We also show that pairs (p, A) exist at which the Crouzeix ratio is not regular.

Mathematics Subject Classification 15A60 · 49J52

Dedicated to Terry Rockafellar on the Occasion of his 80th Birthday.

Anne Greenbaum: Supported in part by National Science Foundation Grant DMS-1210886. Adrian S. Lewis: Supported in part by National Science Foundation Grant DMS-1208338. Michael L. Overton: Supported in part by National Science Foundation Grant DMS-1317205.

✉ Michael L. Overton
overton@cs.nyu.edu

1 Department of Applied Mathematics, University of Washington, Seattle, WA, USA

2 Department of Operations Research and Information Engineering, Cornell University, Ithaca, NY, USA

3 Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

1 Crouzeix's conjecture

Let $\mathcal{M}^N$ denote the space of $N \times N$ complex matrices, let $\mathcal{P}_M$ denote the space of polynomials with complex coefficients and degree $\leq M$, and let $\|\cdot\|$ denote the vector or matrix 2-norm. Michel Crouzeix's 2004 conjecture [4] states that for all $A \in \mathcal{M}^N$ and all $p \in \mathcal{P}_M$, the following inequality holds regardless of the values of N and M:

$$\|p(A)\| \leq 2\|p\|_{W(A)} \tag{1}$$

where $W(A)$ is the field of values (or numerical range) of A,

$$W(A) = \{v^* A v \ : \ v \in \mathbb{C}^N, \ \|v\| = 1\},$$

and

$$\|p\|_{W(A)} = \max_{\zeta \in W(A)} |p(\zeta)| = \max_{\|v\|=1} |p(v^* A v)|.$$

Here * denotes complex conjugate transpose. The set $W(A)$ is a convex, compact subset of the complex plane [11, Ch. 1]. Clearly, the conjecture holds for $N = 1$ or if p is a constant polynomial (with the factor 2 replaced by 1) so we assume that $N \geq 2$ and p is not constant.

This conjecture, which seeks to bound the spectral norm of the polynomial of a matrix by the norm of the polynomial on the field of values of the matrix in a remarkably simple way, has been open for more than a decade. Crouzeix's 2007 theorem [5] states that the inequality (1) holds if the 2 on the right-hand side is replaced by 11.08. The conjecture postulates that the *Crouzeix ratio* $\|p\|_{W(A)}/\|p(A)\|$ is bounded below by $1/2$, while the theorem states that it is bounded below by $1/11.08$. The Crouzeix ratio is locally Lipschitz continuous on the set of all pairs (p, A) for which $p(A) \neq 0$. It is neither smooth nor convex, but it is semialgebraic.

The conjecture is known to hold for certain restricted classes of polynomials p or matrices A:

- $p(\zeta) = \zeta^M$ (from the power inequality, Berger [1] and Pearcy [17])
- $W(A)$ is a disk (Badea [4, p. 462], based on von Neumann's inequality [21] and work of Okubo and Ando [16])
- $N = 2$ (Crouzeix [4], and, more generally, if the minimum polynomial of A has degree 2 (applying results in [20])
- $N = 3$ and $A^3 = 0$ (Crouzeix [6])
- A is an upper Jordan block with a perturbation in the bottom left corner (Greenbaum and Choi [9]) or any diagonal scaling of such A (Choi [2])
- A is diagonalizable with an eigenvector matrix having condition number less than or equal to 2 (easy)
- $AA^* = A^*A$ (then the constant 2 can be improved to 1).

Extensive numerical experiments by Crouzeix [7] and Greenbaum and Overton [10] strongly support the conjecture.

Pairs (p, A) for which the Crouzeix ratio is 0.5 are known. Given an integer n with $2 \le n \le \min(N, M+1)$, set $m = n-1$, define the polynomial $p \in \mathcal{P}_m \subset \mathcal{P}_M$ by $p(\zeta) = \zeta^m$, set the matrix $\tilde{A} \in \mathcal{M}^n$ to

$$\begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \text{ if } n = 2, \text{ or } \begin{bmatrix} 0 & \sqrt{2} & & & & \\ & \cdot & 1 & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & \cdot & 1 \\ & & & & & \cdot \ \sqrt{2} \\ & & & & & 0 \end{bmatrix} \text{ if } n > 2, \tag{2}$$

and set $A = \operatorname{diag}\big(\tilde{A}, 0\big) \in \mathcal{M}^N$. It was independently observed by Choi [2] and Crouzeix [7] that $W(A)$ is the unit disk $\mathcal{D}$, so the numerator of the Crouzeix ratio for (p, A) is one, and that $p(A) = A^{n-1}$ is a matrix with just one nonzero, namely a two in the $(1, n)$ position, so the denominator of the Crouzeix ratio is two and hence the ratio is 0.5. In fact, the experiments of Greenbaum and Overton suggest that this is essentially the *only* pair for which the Crouzeix ratio is 0.5.[1]

Crouzeix's conjecture is equivalent to saying that the pair (p, A) given above is a global minimizer of the Crouzeix ratio on $\mathcal{P}_M \times \mathcal{M}^N$. The main theorem in this paper, established in Sect. 5, is that a first-order nonsmooth necessary condition for (p, A) to be a local minimizer holds, and furthermore that the directional derivative of the Crouzeix ratio at (p, A) is nonnegative in every direction in $\mathcal{P}_M \times \mathcal{M}^N$.

2 Variational analysis

We will use the following standard notions from variational analysis. Let h map a Euclidean space $\mathbf{E}$ to $\mathbb{R}$. We say that h is *smooth* on an open set $X \subset \mathbf{E}$ if it is continuously differentiable there and that h is directionally differentiable on X if, for all $x \in X$, the directional derivative

$$h'(x; d) \equiv \lim_{t \downarrow 0} \frac{h(x + td) - h(x)}{t}$$

exists and is finite for all $d \in \mathbf{E}$. If h is locally Lipschitz and directionally differentiable on X, we say that h is (Clarke) *regular* on X when its directional derivative $x \mapsto h'(x; d)$ is upper semicontinuous (usc) on X for every fixed direction d [19, Thm. 9.16]. It is well known that for regular functions, various different notions of subgradients [19, Ch. 9] or generalized gradients [3] all coincide. We use $\partial h(x)$ to denote the set of such subgradients, or subdifferential, of h at $x \in X$. In the case we are considering (when h

[1] By this we mean, apart from making the following transformations: scaling p, scaling A, shifting the root of the monomial p and the diagonal of the matrix A by the same scalar, applying a unitary similarity transformation to A, or replacing the zero block in A by any matrix whose field of values is contained in $\mathcal{D}$. Note, however, that if the condition that p is a polynomial is relaxed to allow it to be analytic, there are many choices for (p, A) for which the ratio 0.5 is attained; for the case $N = 3$, see [8, Sec. 10].

is locally Lipschitz, directionally differentiable and regular), the subdifferential $\partial h(x)$ is a nonempty compact convex set consisting of those vectors y for which the inner product $\langle y, d\rangle$ is no greater than $h'(x; d)$ for all directions $d \in \mathbf{E}$; furthermore

$$h'(x; d) = \max_{y \in \partial h(x)} \langle y, d\rangle. \tag{3}$$

Note that the map $d \mapsto h'(x; d)$ is sublinear [19, Def. 3.18]. Hence, the nonsmooth stationarity condition $0 \in \partial h(x)$ is equivalent to the first-order optimality condition $h'(x, d) \geq 0$ for all directions $d \in \mathbf{E}$. Convex functions and smooth functions are globally regular, but nonsmooth concave functions are not.

The following nonsmooth quotient rule will be useful. It is a special case of [15, Theorem 3.45], but we include a proof for completeness.

Proposition 1 *Let $\nu : \mathbf{E} \to \mathbb{R}$ be locally Lipschitz, directionally differentiable and regular on an open set $X \in \mathbf{E}$ and let $\delta : \mathbf{E} \to \mathbb{R}$ be smooth on X with gradient $\nabla\delta$. Define the quotient h by $x \mapsto \nu(x)/\delta(x)$, assuming $\delta(x) \neq 0$ for $x \in X$. Then h is regular on X with subdifferential*

$$\partial h(x) = \frac{\delta(x)\partial\nu(x) - \nu(x)\nabla\delta(x)}{\delta(x)^2}.$$

Proof Fix $d \in \mathbf{E}$. Applying the ordinary quotient rule to the function $t \mapsto h(x + td)$, which maps $\mathbb{R}$ to $\mathbb{R}$, we find

$$\begin{aligned}
\delta(x)^2 h'(x; d) &= \delta(x)\nu'(x; d) - \nu(x)\delta'(x; d) \\
&= \delta(x) \max_{y \in \partial\nu(x)} \langle y, d\rangle - \nu(x)\big\langle\nabla\delta(x), d\big\rangle \\
&= \max_{y \in \partial\nu(x)} \big\langle\delta(x)y - \nu(x)\nabla\delta(x), d\big\rangle.
\end{aligned}$$

Since $\nu'(\cdot; d)$ is usc on X and $\nabla\delta(\cdot)$ is continuous on X, it follows that $h'(\cdot; d)$ is also usc on X and hence that h is regular there. The result now follows from (3). □

3 Parameterizing the boundary of $W(A)$

By the maximum modulus principle, $|p(\zeta)|$ must attain its maximum over $\zeta \in W(A)$ on a nonempty subset of the boundary of $W(A)$, and since p is not constant, the maximum is attained only on the boundary. The following fundamental proposition goes back to [13] and is also well known from [11,12], but the usual proof is less succinct than ours.

Proposition 2 *For $\theta \in [0, 2\pi)$, define the Hermitian matrix*

$$H_\theta = \frac{1}{2}\left(e^{i\theta}A + e^{-i\theta}A^*\right). \tag{4}$$

A point z is a boundary point of $W(A)$ if and only if $z = v^*Av$ where v is a unit eigenvector of H_θ corresponding to $\lambda_{\max}(H_\theta)$, the largest eigenvalue of H_θ, for some $\theta \in [0, 2\pi)$.

Proof We use the real inner product on $\mathbb{C}$ defined by $\langle \xi, \eta\rangle = \mathrm{Re}(\xi^*\eta)$. Since $W(A)$ is closed and convex, z is a boundary point of $W(A)$ if and only if it lies on a supporting hyperplane, namely, a line L_θ described by the conditions $\langle e^{-i\theta}, y - z\rangle = 0$ for $y \in L_\theta$ and $\langle e^{-i\theta}, y - z\rangle \leq 0$ for $y \in W(A)$, for some $\theta \in [0, 2\pi)$. Such a boundary point satisfies $z = v_\theta^* A v_\theta$ where v_θ maximizes, over all unit vectors $v \in \mathbb{C}^N$,

$$\left\langle e^{-i\theta}, v^*Av\right\rangle = \mathrm{Re}\left(v^*(e^{i\theta}A)v\right) = v^*H_\theta v.$$

Hence, v_θ is a unit eigenvector corresponding to $\lambda_{\max}(H_\theta)$. □

Note that if $\lambda_{\max}(H_\theta)$ is simple, then v_θ is uniquely defined up to a unimodular scalar.

4 The subdifferential of the Crouzeix ratio

Let us identify $p \in \mathcal{P}_M$ with its coefficient vector $c = [c_0, c_1, \ldots, c_M]^T \in \mathbb{C}^{M+1}$, with $c_j \neq 0$ for at least one $j \in \{1, \ldots, M\}$, and define the function $q : \mathbb{C}^{M+1} \times \mathbb{C} \to \mathbb{C}$ by

$$q(c, \zeta) = \sum_{j=0}^{M} c_j \zeta^j.$$

Depending on the context, we will also interpret q as a function mapping $C^{M+1} \times \mathcal{M}^N$ to $\mathcal{M}^N$, defined by substituting $A \in \mathcal{M}^N$ for $\zeta \in \mathbb{C}$ above. We write the Crouzeix ratio as

$$f(c, A) = \frac{\tau(c, A)}{\beta(c, A)}$$

where

$$\tau(c, A) = \max\left\{|q(c, z)| : z \in W(A)\right\}, \tag{5}$$

and

$$\beta(c, A) = \|p(A)\| = \sigma_{\max}\big(q(c, A)\big), \tag{6}$$

the largest singular value of $\sum_{j=0}^{M} c_j A^j$. Thus f maps the Euclidean space $\mathbb{C}^{M+1} \times \mathcal{M}^N$, with real inner product

$$\langle (c, A), (d, B)\rangle = \mathrm{Re}\big(c^*d + \mathrm{tr}(A^*B)\big),$$

to $\mathbb{R}$. We address the case where the denominator is zero below. The notations τ and β were chosen to indicate the "top" and "bottom" components of the ratio.

We begin our analysis with the numerator. We can rewrite τ as

$$\tau(c, A) \;=\; \max\bigl\{\phi(c, A, \omega, v) : |\omega| = 1,\; \|v\| = 1\bigr\}, \tag{7}$$

where the function $\phi\colon \mathbb{C}^{M+1} \times \mathcal{M}^N \times \mathbb{C} \times \mathbb{C}^N \to \mathbb{R}$ is defined by

$$\phi(c, A, \omega, v) \;=\; \mathrm{Re}\bigl(\omega^* q(c, v^* A v)\bigr).$$

Let $Z(c, A)$ denote the set of points $z \in W(A)$ attaining the maximum in (5) and let $\Omega(c, A)$ denote the set of pairs (ω, v) attaining the maximum in (7). Clearly

$$\Omega(c, A) \;=\; \bigl\{(\omega, v) : |\omega| = 1, \|v\| = 1, z \in Z(c, A), v^* A v = z, \omega^* z = |q(c, z)|\bigr\}.$$

By [19, Thm. 10.31], τ is everywhere locally Lipschitz, directionally differentiable and regular, with subdifferential

$$\partial\tau(c, A) \;=\; \mathrm{conv}\bigl\{\nabla_{(c,A)}\phi(c, A, \omega, v) : (\omega, v) \in \Omega(c, A)\bigr\}. \tag{8}$$

By definition, the gradient vector satisfies

$$\begin{aligned}&\phi(c + \delta c, A + \delta A, \omega, v) - \phi(c, A, \omega, v)\\ &\quad = \Bigl\langle \nabla_{(c,A)}\phi(c, A, \omega, v), (\delta c, \delta A)\Bigr\rangle \;+\; o(\delta c, \delta A).\end{aligned}$$

The left-hand side is

$$\begin{aligned}&\mathrm{Re}\Bigl(\omega^*\bigl(q(c + \delta c, v^*(A + \delta A)v) - q(c, v^* A v)\bigr)\Bigr)\\ &\qquad = \;\mathrm{Re}\Bigl(\omega^*\bigl(\langle \nabla q(c, v^* A v), \bigl(\delta c, v^*(\delta A)v\bigr)\rangle\bigr)\Bigr) \;+\; o(\delta c, \delta A).\end{aligned}$$

The gradient of q at the pair (c, ζ) is defined by

$$\langle \nabla q(c, \zeta), (\delta c, \delta\zeta)\rangle \;=\; (\delta c)_0 + \sum_{j=1}^{M}\Bigl((\delta c)_j \zeta^j + j c_j \zeta^{j-1}(\delta\zeta)\Bigr).$$

Setting $z = v^* A v$, we deduce

$$\begin{aligned}
&\Big\langle \nabla_{(c,A)}\phi(c, A, \omega, v), (\delta c, \delta A)\Big\rangle \\
&\quad = \mathrm{Re}\Big(\omega^*(\delta c)_0 + \omega^* \sum_{j=1}^{M} \big((\delta c)_j z^j + j c_j z^{j-1} v^*(\delta A) v\big)\Big) \\
&\quad = \Big\langle \omega\big((z^*)^j\big)_{j=0}^{M}, \delta c\Big\rangle + \Big\langle \omega \sum_{j=1}^{M} j c_j^* (z^*)^{j-1} v v^*, \delta A\Big\rangle,
\end{aligned}$$

so

$$\nabla_{(c,A)}\phi(c, A, \omega, v) \;=\; \Big(\omega\big((z^*)^j\big)_{j=0}^{M}, \; \omega \sum_{j=1}^{M} j c_j^* (z^*)^{j-1} v v^*\Big).$$

Assuming $\tau(c, A) \neq 0$ and applying (8), we find that

$$\begin{aligned}
\partial\tau(c, A) \;=\; \mathrm{conv}\Big\{ & \frac{q(c, z)}{|q(c, z)|}\Big(\big((z^*)^j\big)_{j=0}^{M}, \; \sum_{j=1}^{M} j c_j^* (z^*)^{j-1} v v^*\Big) : \\
& z = v^* A v \in Z(c, A), \;\; \|v\| = 1\Big\}.
\end{aligned} \tag{9}$$

Recall from Sect. 3 that, exploiting the maximum modulus principle together with Proposition 2, we know an explicit formula for the unit vectors v satisfying $v^* A v \in Z(c, A)$: they are eigenvectors corresponding to the maximum eigenvalue of H_θ for some $\theta \in [0, 2\pi)$.

Now we turn to the denominator $\beta(c, A) = \sigma_{\max}\big(q(c, A)\big)$. The largest singular value of a matrix X is characterized by

$$\begin{aligned}
\sigma_{\max}(X) &= \max\Big\{\mathrm{Re}\big(u^* X w\big) : \|u\| = \|w\| = 1\Big\} \\
&= \max\Big\{\big\langle X, u w^*\big\rangle : \|u\| = \|w\| = 1\Big\}.
\end{aligned}$$

Assume that $\sigma_{\max}\big(q(c, A)\big)$ is simple, with corresponding left and right unit singular vectors $u, w \in \mathbb{C}^N$, so that the denominator is smooth with gradient

$$\nabla\beta(c, A) = u w^*.$$

It follows that

$$\begin{aligned}
&\beta(c + \delta c, A + \delta A) - \beta(c, A) \\
&\quad = \sigma_{\max}\big(q(c + \delta c, A + \delta A)\big) - \sigma_{\max}\big(q(c, A)\big) \\
&\quad = \mathrm{Re}\Big(u^*\big(q(c + \delta c, A + \delta A) - q(c, A)\big) w\Big) \; + \; o(\delta c, \delta A)
\end{aligned}$$

$$= \mathrm{Re}\Big(u^* \sum_{j=0}^{M} \big((c_j + \delta c_j)(A + \delta A)^j - c_j A^j\big) w\Big) \; + \; o(\delta c, \delta A)$$

$$= \mathrm{Re} \sum_{j=0}^{M} (\delta c_j)(u^* A^j w) + \mathrm{Re}\ \mathrm{tr}\Big(\sum_{j=1}^{M} c_j \sum_{l=0}^{j-1} A^l (\delta A) A^{j-l-1}\Big) w u^* \; + \; o(\delta c, \delta A)$$

$$= \mathrm{Re} \sum_{j=0}^{M} (\delta c_j)(u^* A^j w) + \mathrm{Re}\ \mathrm{tr}\Big(\sum_{j=1}^{M} c_j \sum_{l=0}^{j-1} A^{j-l-1} w u^* A^l\Big) \delta A \; + \; o(\delta c, \delta A),$$

and hence

$$\nabla \beta(c, A) \; = \; \Big(\big(w^* A^{*j} u\big)_{j=0}^{M}, \sum_{j=1}^{M} c_j^* \sum_{l=0}^{j-1} A^{*l} u w^* A^{*(j-l-1)}\Big). \tag{10}$$

Since $N \geq 2$, it follows from the assumption on the simplicity of the maximum singular value of $q(c, A)$ that $\beta(c, A)$ is nonzero, and therefore that $\tau(c, A)$ is nonzero (because if it were zero, $W(A)$ would consist of a single point λ with $\sum c_j \lambda^j = 0$, and this would imply that $A = \lambda I$ and hence $\beta(c, A) = 0$).

This discussion leads to the following result.

Theorem 3 *Let $c = [c_0, c_1, \ldots, c_M]^T$, with c_j nonzero for at least one $j > 0$, and $A \in \mathcal{M}^N$ be given, with $N \geq 2$. Assume that the largest singular value of $\sum_j c_j A^j$ is simple. Then the Crouzeix ratio f is regular on a neighborhood of (c, A) with subdifferential*

$$\partial f(c, A) = \frac{\beta(c, A) \partial \tau(c, A) - \tau(c, A) \nabla \beta(c, A)}{\beta(c, A)^2}, \tag{11}$$

where $\partial \tau(c, A)$ and $\nabla \beta(c, A)$ are given by (9) and (10) respectively.

Proof The proof follows from the analysis above, using the nonsmooth quotient rule in Proposition 1. □

5 Local optimality conditions at candidate minimizers

We are now in a position to study nonsmooth stationarity of our candidate minimizers. As at the end of Sect. 1, given an integer n with $2 \leq n \leq \min(N, M+1)$, set $m = n-1$ and define the polynomial p as the monomial $\zeta \mapsto \zeta^m$ with coefficients

$$c = [0, \ldots, 0, 1, 0, \ldots 0]^T \in \mathbb{C}^{M+1}. \tag{12}$$

Let

$$A = \mathrm{diag}\big(\tilde{A}, 0\big) \in \mathcal{M}^N \tag{13}$$

where $\tilde{A} \in \mathcal{M}^n$ is given in (2). Then, as observed following (2), $W(A)$ is the unit disk $\mathcal{D}$, and $p(A) = A^{n-1}$, which has norm two, so $\tau(c, A) = 1$, $\beta(c, A) = 2$ and the Crouzeix ratio $f(c, A) = 0.5$. Hence, the pair (c, A) is a candidate minimizer of f, and is a global minimizer if Crouzeix's conjecture is true.

Theorem 4 *Let c, A be given by* (12), (13). *The subdifferential of the Crouzeix ratio at (c, A) is*

$$\partial f(c, A) = \mathrm{conv}_{\theta \in [0,2\pi)} \left\{ \left(y_\theta, Y_\theta \right) \right\}$$

where

$$y_\theta = \frac{1}{2} \left[z^m, z^{m-1}, \ldots, z, 0, z^{-1}, z^{-2}, \ldots, z^{m-M} \right]^T$$

and Y_θ is the block diagonal matrix $\mathrm{diag}\left(\tilde{Y}_\theta, 0\right)$, *where $\tilde{Y}_\theta$ is the $n \times n$ matrix*

$$\tilde{Y}_\theta = \frac{1}{4} \begin{bmatrix} z & 0 & \sqrt{2}z^{-1} & \sqrt{2}z^{-2} & \cdots & \sqrt{2}z^{3-n} & z^{2-n} \\ \sqrt{2}z^2 & 2z & 0 & 2z^{-1} & \cdots & 2z^{4-n} & \sqrt{2}z^{3-n} \\ \vdots & & & & & & \vdots \\ \sqrt{2}z^{n-2} & 2z^{n-3} & 2z^{n-4} & 2z^{n-5} & \cdots & 0 & \sqrt{2}z \\ \sqrt{2}z^{n-1} & 2z^{n-2} & 2z^{n-3} & 2z^{n-4} & \cdots & 2z & 0 \\ z^n & \sqrt{2}z^{n-1} & \sqrt{2}z^{n-2} & \sqrt{2}z^{n-3} & \cdots & \sqrt{2}z^2 & z \end{bmatrix}$$

with $z = e^{-i\theta}$. When $n = 2$, these should be interpreted as

$$y_\theta = \left[z, 0, z^{-1}, \ldots, z^{1-M} \right]^T \quad \text{and} \quad \tilde{Y}_\theta = \frac{1}{4} \begin{bmatrix} z & 0 \\ z^2 & z \end{bmatrix}.$$

Corollary 5 *Let c, A be given by* (12), (13). *Then*

$$0 \in \partial f(c, A).$$

This says that for any n and m satisfying $2 \leq n \leq N$ and $m = n - 1 \leq M$, the pair (c, A) is a nonsmooth stationary point of f. As explained in Sect. 2, together with regularity this implies that the directional derivative of the Crouzeix ratio is nonnegative in every direction—a new result for $N > 2$. It was implicitly already known for $N = n = 2$, because Crouzeix's conjecture is known to hold for 2×2 matrices. It was also implicitly known previously that, for fixed c given in (12), $0 \in \partial f(c, \cdot)(A)$, since Crouzeix's conjecture is known to hold when p is a fixed monomial.

The proof of Corollary 5 is immediate, as the convex combination

$$\frac{1}{N+1} \sum_{k=0}^{N} \left(y_{2k\pi/(N+1)}, Y_{2k\pi/(N+1)} \right)$$

is zero. Alternatively, note that the integral

$$\frac{1}{2\pi}\int_0^{2\pi}\left(y_\theta, Y_\theta\right)d\theta$$

is zero.

Proof of Theorem 4 Since $W(A)=\mathcal{D}$ and $\sum_j c_j\zeta^j=\zeta^m$, we have that $Z(c,A)$ is the unit circle $\left\{e^{i\theta} : \theta\in[0,2\pi)\right\}$. The Hermitian matrix defined in (4) is $H_\theta=\operatorname{diag}(\tilde{H}_\theta,0)$, where

$$\tilde{H}_\theta=\frac{1}{2}\left(e^{i\theta}A+e^{-i\theta}A^*\right)=\frac{1}{2}\begin{bmatrix} 0 & \sqrt{2}e^{i\theta} & & & & \\ \sqrt{2}e^{-i\theta} & 0 & e^{i\theta} & & & \\ & e^{-i\theta} & 0 & e^{i\theta} & & \\ & & \ddots & \ddots & \ddots & \\ & & & e^{-i\theta} & 0 & \sqrt{2}e^{i\theta} \\ & & & & \sqrt{2}e^{-i\theta} & 0 \end{bmatrix}.$$

Some calculations show that, for all θ, its largest eigenvalue is simple with unit eigenvector

$$v_\theta=\frac{1}{\sqrt{n-1}}\begin{bmatrix}\tilde{x}_\theta\\0\end{bmatrix}\quad\text{where}\quad\tilde{x}_\theta=\left[\frac{e^{(n-1)i\theta}}{\sqrt{2}},e^{(n-2)i\theta},\ldots,e^{i\theta},\frac{1}{\sqrt{2}}\right]^T\tag{14}$$

and with $v_\theta^*Av_\theta=e^{-i\theta}$. In what follows we write z as an abbreviation for $e^{-i\theta}$.

Let us consider the numerator τ. Equation (9) gives

$$\partial\tau(c,A)=\operatorname{conv}_{\theta\in[0,2\pi)}\left\{\left(s_\theta,S_\theta\right)\right\}$$

where

$$s_\theta=z^m\left[1,z^{-1},\ldots,z^{-M}\right]^T=\left[z^m,\ldots,1,\ldots,z^{m-M}\right]^T$$

and, using (12) and noting that $m=n-1$,

$$S_\theta=z^{n-1}(n-1)z^{2-n}v_\theta v_\theta^*=(n-1)zv_\theta v_\theta^*.$$

Using (14) we find that $S_\theta=\operatorname{diag}\left(\tilde{S}_\theta,0\right)$ where, if $n=2$,

$$\tilde{S}_\theta=\frac{1}{2}\begin{bmatrix} z & 1\\ z^2 & z\end{bmatrix}$$

and otherwise

$$\tilde{S}_\theta = \frac{1}{2}\begin{bmatrix} z & \sqrt{2} & \sqrt{2}z^{-1} & \sqrt{2}z^{-2} & \cdots & \sqrt{2}z^{3-n} & z^{2-n} \\ \sqrt{2}\,z^2 & 2z & 2 & 2z^{-1} & \cdots & 2z^{2-n} & \sqrt{2}z^{3-n} \\ \vdots & & & & & & \vdots \\ \sqrt{2}z^{n-2} & 2z^{n-3} & 2z^{n-4} & 2z^{n-5} & \cdots & 2 & \sqrt{2}z \\ \sqrt{2}z^{n-1} & 2z^{n-2} & 2z^{n-3} & 2z^{n-4} & \cdots & 2z & \sqrt{2} \\ z^n & \sqrt{2}z^{n-1} & \sqrt{2}z^{n-2} & \sqrt{2}z^{n-3} & \cdots & \sqrt{2}z^2 & z \end{bmatrix}.$$

Now we turn to the denominator. Let e_j denote the jth coordinate vector. Since $p(A) = \mathrm{diag}(2e_1e_n^*, 0)$, its maximum singular value is simple, with corresponding left and right singular vectors $u = [e_1; 0]$ and $v = [e_n; 0]$. Hence, using (10), we have

$$\nabla\beta(c, A) = (r_\theta, R_\theta)$$

where

$$r_\theta = [0, \ldots, 0, 2, 0, \ldots, 0]^T$$

since $u^*A^kw = 0$ for $k = 0, \ldots, m-1$, $u^*A^mw = 2$, and $A^k = 0$ for $k > m$, and, using (12) and (13), $R_\theta = \mathrm{diag}\big(\tilde{R}_\theta, 0\big)$, where $\tilde{R}_\theta = e_1e_2^*$ if $n = 2$, and otherwise

$$\tilde{R}_\theta = \sum_{\ell=0}^{n-2} \tilde{A}^{*\ell}e_1e_n^*\tilde{A}^{*(n-2-\ell)} = \sqrt{2}e_1e_2^* + 2\sum_{l=2}^{n-2} e_\ell e_{\ell+1}^* + \sqrt{2}e_{n-1}e_n^*$$
$$= \begin{bmatrix} 0 & \sqrt{2} & & & & \\ & \cdot & 2 & & & \\ & & \cdot & \cdot & & \\ & & & \cdot & \cdot & \\ & & & & \cdot & 2 \\ & & & & & \cdot & \sqrt{2} \\ & & & & & & 0 \end{bmatrix}.$$

Finally, since the assumptions of Theorem 3 hold, we obtain from (11) that

$$\partial f(c, A) = \mathrm{conv}_{\theta\in[0,2\pi)}\Big\{\big(y_\theta, Y_\theta\big)\Big\}$$

where, since $\tau(c, A) = 1$ and $\beta(c, A) = 2$,

$$y_\theta = \frac{1}{4}\,(2s_\theta - r_\theta) \quad \text{and} \quad Y_\theta = \frac{1}{4}\,(2S_\theta - R_\theta)\,.$$

The proof is completed by combining the equations given above. □

A crucial point in the proof is that the twos in $2s_\theta$ and r_θ cancel *and* the first superdiagonals in $2S_\theta$ and in R_θ cancel. Since these quantities are independent of θ, Corollary 5 could not hold without their cancellation.

6 Breakdown of regularity

In this section we show that pairs (c, A) exist at which the Crouzeix ratio f is not regular. The numerator τ is regular everywhere, even without the assumptions in Theorem 3. The same is true of the denominator β, as it is the composition of a convex function (the maximum singular value) with a polynomial. However, Proposition 1 does not apply when the denominator is not smooth. So, we focus on the directional derivative instead.

Fix $M = m = 1$ and p by $p(\zeta) = \zeta$, equivalently $c = [0, 1]^T$, and write

$$\check{f}(A) = f(c, A) = \frac{\check{\tau}(A)}{\check{\beta}(A)} = \frac{\tau(c, A)}{\beta(c, A)}.$$

Then immediately from the definition,

$$\check{f}(A) = \frac{\max_{\|v\|=1} |v^* A v|}{\max_{\|u\|=\|w\|=1} |u^* A w|}.$$

If, for some A, $\sigma_{\max}(A)$ has multiplicity greater than one, $\check{\beta}$ is nonsmooth at A, and hence (11) does not apply. However, by the ordinary quotient rule, the directional derivative of $\check{f}$ at A in a direction $D \in \mathcal{M}^N$ is

$$\check{f}'(A; D) = \frac{\check{\beta}(A)\check{\tau}'(A; D) - \check{\tau}(A)\check{\beta}'(A; D)}{\check{\beta}(A)^2}. \tag{15}$$

Since the numerator and denominator are both regular, we have from (3) that

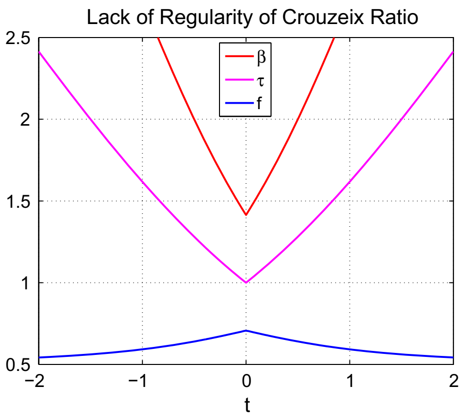
$$\check{\tau}'(A; D) = \max_{G \in \partial\check{\tau}(A)} \langle G, D\rangle \quad \text{and} \quad \check{\beta}'(A; D) = \max_{G \in \partial\check{\beta}(A)} \langle G, D\rangle. \tag{16}$$

Let $N = n = 3$ and fix A to be given by $\tilde{A}$ in (2), that is, a 3×3 Jordan block with zero on the diagonal, scaled by $\sqrt{2}$. Note that $W(A)$ is the unit disk $\mathcal{D}$, so the numerator $\check{\tau}(A) = 1$, but the denominator $\check{\beta}(A) = \sqrt{2}$, not 2 as in Theorem 4, because now $p(A) = A$, not A^2. So, $\check{f}(A) = 1/\sqrt{2}$.

We can derive $\partial\check{\tau}(A)$ using (9). We find

$$\partial\check{\tau}(A) = \mathrm{conv}_{\theta\in[0,2\pi)}\{T_\theta\} \tag{17}$$

Fig. 1 Plot of the denominator β, the numerator τ and the Crouzeix ratio f evaluated at $(c, A + tA^2)$, where $c = [0, 1]^T$ (so $p(\zeta) = \zeta$)) and A is the 3×3 Jordan block scaled by $\sqrt{2}$, for $t \in [-2, 2]$. This example shows that f is not regular at (c, A)



where, noting that $q(z) = z$,

$$T_\theta = z v_\theta v_\theta^* = \frac{1}{4} \begin{bmatrix} z & \sqrt{2} & z^{-1} \\ \sqrt{2}z^2 & 2z & \sqrt{2} \\ z^3 & \sqrt{2}z^2 & z \end{bmatrix}$$

with $z = e^{-i\theta}$.

Since $\sigma_{\max}(A)$ has multiplicity two, the denominator is not smooth at A, but it is convex and hence regular and its subdifferential is [22]

$$\begin{aligned} \partial \check{\beta}(A) &= \operatorname{conv} \left\{ uw^* : u^* A w = \sigma_{\max}(A) = \sqrt{2}, \|u\| = \|w\| = 1 \right\} \\ &= \operatorname{conv} \left\{ \begin{bmatrix} 0 & |\mu|^2 & \mu\overline{\nu} \\ 0 & \overline{\mu}\nu & |\nu|^2 \\ 0 & 0 & 0 \end{bmatrix} : \ |\mu|^2 + |\nu|^2 = 1 \right\}. \end{aligned} \tag{18}$$

Now, let $D = A^2 = 2e_1 e_3^*$. Then, it follows from (16), (17) and (18) that

$$\check{\tau}'(A; D) = \frac{2}{4} \max_{\theta \in [0, 2\pi)} \cos(\theta) = \frac{1}{2} \quad \text{and} \quad \check{\beta}'(A; D) = 2 \max_{|\mu|^2 + |\nu|^2 = 1} \operatorname{Re}(\mu\overline{\nu}) = 1.$$

So, using (15), we find

$$\check{f}'(A; D) = \frac{\frac{\sqrt{2}}{2} - 1}{2} < 0.$$

A similar argument shows that $\check{f}'(A; -D) = \check{f}'(A; D) < 0$, so the directional derivative $\check{f}'(A; \cdot)$ is not sublinear: if it were, we would arrive at the contradiction

$$0 > \check{f}'(A; D) + \check{f}'(A; -D) \geq \check{f}'(A; 0) = 0.$$

Hence, it follows from the discussion in Sect. 2 that $\check{f}$ is not regular at A, and so f is not regular at (c, A). Figure 1 shows plots of $\check{\beta}$, $\check{\tau}$ and $\check{f}$ evaluated at $A + tD$ for $t \in [-2, 2]$.

7 Concluding remarks

If the polynomial-matrix pair (c, A) described by Eqs. (12) and (13) is indeed a global minimizer of the Crouzeix ratio, as numerical evidence strongly suggests, then Crouzeix's conjecture is true. In this work we have shown, in contrast, just a local stationarity property of (c, A): the ratio has nonnegative directional derivative in every direction. Even in classical smooth optimization, this property does not certify a local minimizer, let alone a global one.

However, perhaps we have somewhat understated our progress towards proving that the pair(c, A) is at least a local minimizer. One variational analytic approach to establishing local optimality [14, Cor. 4.13] would need three properties of the Crouzeix ratio f at (c, A):

- prox-regularity of f [19, Def. 13.27]
- zero lying in the relative interior [19, Sec. 2.H] of the subdifferential $\partial f(c, A)$
- when f is restricted to a certain "active" manifold, on which it is smooth, (c, A) is a local minimizer.

The first two properties follow from the results established above, as we now explain.

Theorem 3 gives conditions under which the ratio f is (Clarke) regular, and Theorem 4 confirms that f is regular at (c, A) given by (12) and (13). However, under the same conditions, it follows from the representation (7) of the numerator and the smoothness of the denominator that the ratio has the stronger property of prox-regularity at (c, A). Indeed, it can be written locally as the sum of a continuous convex function and a C^2 smooth function [19, Thm. 10.33].

The second property follows from our second proof of Corollary 5 and the following observation: for any continuous map F from the unit interval into a Euclidean space, the integral of F lies in the relative interior of the convex hull of its range. To see this, denote the integral by x, and consider any normal vector y to the convex hull at x. By definition, the inner product of y with $F(\cdot) - x$ is everywhere nonnegative, but its integral is zero, so it must be identically zero. Hence $-y$ is also a normal vector. The result then follows from the definition of relative interior [18, Cor. 11.6.2].

The third property mentioned above would need a second-order analysis beyond our current scope. Nonetheless, the progress we have presented is a striking showcase of the variational-analytic toolkit for investigating local optimality conditions for nonsmooth functions, as well as a reassuring test of Crouzeix's conjecture.

References

1. Berger, C.: A strange dilation theorem. Not. Am. Math. Soc. 12, 590 (1965). Abstract 625–152
2. Choi, D.: A proof of Crouzeix's conjecture for a class of matrices. Linear Algebra Appl. **438**(8), 3247–3257 (2013)

3. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983). Reprinted by SIAM, Philadelphia, 1990
4. Crouzeix, M.: Bounds for analytical functions of matrices. Integral Equ. Oper. Theory **48**, 461–477 (2004)
5. Crouzeix, M.: Numerical range and functional calculus in Hilbert space. J. Funct. Anal. **244**(2), 668–690 (2007)
6. Crouzeix, M.: Spectral sets and 3 × 3 nilpotent matrices. In: Topics in Functional and Harmonic Analysis, vol. 14 of Theta Ser. Adv. Math., pp. 27–42. Theta, Bucharest (2013)
7. Crouzeix, M.: Private communication (2015)
8. Crouzeix, M.: Some constants related to numerical ranges. SIAM J. Matrix Anal. Appl. **37**, 420–442 (2016)
9. Greenbaum, A., Choi, D.: Crouzeix's conjecture and perturbed Jordan blocks. Linear Algebra Appl. **436**(7), 2342–2352 (2012)
10. Greenbaum, A., Overton, M.L.: Numerical investigation of Crouzeix's conjecture. (2016). http://cs.nyu.edu/overton/papers/pdffiles/NumerInvestCrouzeixConj.pdf. Accessed 1 Nov 2016
11. Horn, R.A., Johnson, C.R.: Topics in Matrix Analysis. Cambridge University Press, Cambridge (1991)
12. Johnson, C.R.: Numerical determination of the field of values of a general complex matrix. SIAM J. Numer. Anal. **15**(3), 595–602 (1978)
13. Kippenhahn, R.: Über den Wertevorrat einer Matrix. Math. Nachr., 6, 193–228, 1951. English translation by P.F. Zachlin and M.E. Hochstenbach, Linear and Multilinear Algebra 56, pp. 185–225 (2008)
14. Lewis, A.S., Zhang, S.: Partial smoothness, tilt stability, and generalized Hessians. SIAM J. Optim. **23**, 74–94 (2013)
15. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation. I, Basic Theory. Springer, Berlin (2006)
16. Okubo, K., Ando, T.: Constants related to operators of class C_ρ. Manuscr. Math. **16**(4), 385–394 (1975)
17. Pearcy, C.: An elementary proof of the power inequality for the numerical radius. Mich. Math. J. **13**, 289–291 (1966)
18. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
19. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, New York (1998)
20. Tso, S.-H., Wu, P.-Y.: Matricial ranges of quadratic operators. Rocky Mt. J. Math. **29**(3), 1139–1152 (1999)
21. von Neumann, J.: Eine Spektraltheorie fuer allgemeine Operatoren eines unitaeren Raumes. Math. Nachr., 4, 258–281, Collected Works, vol. IV, Pergamon, Oxford 1962, pp. 341–364 (1951)
22. Watson, G.A.: Characterization of the subdifferential of some matrix norms. Linear Algebra Appl. **170**, 33–45 (1992)