hexagon:2611.00918v2

When is the Gohberg–Krupnik (1992) lower bound for $aI + bH$ sharp? Exactly when $\mathrm{Re}(a\bar{b}) = 0$, for $1.2336\ldots \le p \le 5.2807\ldots$, $p \ne 2$ An expanded version, with other miscellanea

Dave Goel*

First version: November 2, 2026. This version (4.80): November 24, 2026.

Abstract

Let P_+ and P_- be the Riesz projections on the complex space $L^p(\mathbb{T})$, $1 < p < \infty$, onto the nonnegative and the negative frequencies, and let $H = -i(P_+ - P_-)$, so that $aI + bH = \alpha P_+ + \beta P_-$ with $\alpha = a - ib$ and $\beta = a + ib$. Gohberg and Krupnik gave a lower bound for $\|aI + bH\|_p$, and Hollenbeck, Kalton and Verbitsky (2003) remarked that it is not sharp in general. We characterize when it is sharp.[1] For $p \ne 2$ and $\alpha \ne 0$ the bound is not sharp whenever the ratio β/α is not real, and it is sharp whenever β/α is real and nonpositive. There is an explicitly defined $p_0 \approx 1.2336026$ such that the bound is sharp for every real ratio exactly when $p_0 \le p \le p_0/(p_0 - 1)$; outside this interval it is not sharp for any real ratio below and sufficiently close to 1. In terms of the coefficients, for $p_0 \le p \le p_0/(p_0 - 1) \approx 5.2807742$ and $p \ne 2$ the bound is sharp exactly when $\mathrm{Re}(a\bar{b}) = 0$, that is, when $\alpha = 0$ or β/α is real. For $1 < p < 2$ and each real ratio $\lambda = \beta/\alpha \in (0, 1)$ we give an exact criterion: an inequality on the positive real axis, together with the condition that λ be smaller than an explicit threshold depending on p; the symmetries $\alpha \leftrightarrow \beta$ and $p \leftrightarrow p'$, where $p' = p/(p-1)$, cover the remaining real ratios and exponents. For $1 < p < p_0$ the bound is sharp for all λ near 0, and on a band of intermediate ratios the outcome of the criterion is known only in part. The cases of H itself, of a single projection and of $p = 2$ are classical. We do not determine the norm where the bound is not sharp. Section 6 gives further results, including a wide-chord inequality that yields explicit lower bounds for the norm, strictly above the Gohberg–Krupnik bound, at two sample pairs (λ, p). The proof of the classification is computer-assisted and uses rigorous interval arithmetic on compact domains. The classification has also been proved in the Lean proof assistant, with the interval computations and two cited theorems (the Hollenbeck–Verbitsky norm of P_+ and Cole's theorem) taken as given.

2020 Mathematics Subject Classification. Primary 42A50, 47A30; Secondary 31A05, 65G30.
Key words. Riesz projection, Hilbert transform, best constants, Gohberg–Krupnik bound, subharmonic minorant, interval arithmetic.

This is the expanded version of the paper "When is the Gohberg–Krupnik (1992) lower bound for $aI + bH$ sharp? Exactly when $\mathrm{Re}(a\bar{b}) = 0$, for $1.2336\ldots \le p \le 5.2807\ldots$, $p \ne 2$" [4].

1 Introduction

Let P_+ and P_- be the Riesz projections on the complex Lebesgue space $L^p(\mathbb{T})$ of the circle with normalized angular measure, $1 < p < \infty$: the projection P_+ retains the Fourier frequencies $n \ge 0$, including the constants, and $P_- = I - P_+$ retains the frequencies $n < 0$. With $H = -i(P_+ - P_-)$, every operator $aI + bH$ with complex coefficients is a combination of the two projections,

$$aI + bH = \alpha P_+ + \beta P_-, \qquad \alpha = a - ib, \quad \beta = a + ib.$$

*The Lean 4 proof of the classification: https://github.com/david44hello/paper_gohberg-krupnik-2003-sharpness/tree/v3.90

[1] This is an expanded version of the following focused version [4]: https://hexagonmath.org/2611.00917

Our H maps the constant 1 to $-i$, whereas the usual conjugate-function operator H_0 annihilates constants; Lemma 1.2 shows that the norm of $aI + bH$ is the same under both conventions and on the real line. Gohberg and Krupnik proved the lower bound

$$\|\alpha P_+ + \beta P_-\|_p \;\geq\; \mathrm{GK}_p(\alpha,\beta) := \left\| \begin{pmatrix} \alpha & (\beta-\alpha)\cot(\pi/p) \\ 0 & \beta \end{pmatrix} \right\|_{\ell^2\to\ell^2};$$

see [6] and [7, (1.8)], where it is stated on $L^p(\mathbb{R})$ and the matrix norm is written out explicitly; Lemma 1.2 transfers it to the circle. The bound holds with equality in the classical cases: for H itself, whose norm was found by Pichorides [12] and, independently, by Cole (see [3]); for a single projection, whose norm $1/\sin(\pi/p)$ was conjectured by Gohberg and Krupnik and found by Hollenbeck and Verbitsky [8]; and for $p = 2$, where orthogonality gives $\|\alpha P_+ + \beta P_-\|_2 = \max(|\alpha|, |\beta|)$. Hollenbeck, Kalton and Verbitsky [7, p. 240] write that the bound was believed to be sharp for all complex a, b and that in general it is not; they do not determine the set of complex a, b for which it fails, and for real a, b they determine the exact norm [7, Theorem 4.1]. We call the bound *sharp* if equality holds and *not sharp* otherwise. Theorem 1.1 below characterizes when each case occurs.

For $\alpha \neq 0$, homogeneity gives $\|\alpha P_+ + \beta P_-\|_p = |\alpha| N_p(\beta/\alpha)$, where $N_p(\rho) := \|P_+ + \rho P_-\|_{L^p(\mathbb{T})\to L^p(\mathbb{T})}$, and correspondingly we write $\mathrm{GK}_p(\rho) := \mathrm{GK}_p(1,\rho)$. Throughout, $p' = p/(p-1)$, $q = p/2$ and $\theta_0 = \pi/p$. The isometry $Uf(e^{it}) = e^{-it} f(e^{-it})$ interchanges P_+ and P_-, because it sends frequency n to $-n-1$; the conjugate-linear isometry $Jf(e^{it}) = \overline{f(e^{-it})}$ preserves both frequency spaces and conjugates scalar coefficients; and both projections are self-adjoint for the sesquilinear L^p–$L^{p'}$ pairing $\langle f, g\rangle = \int_{\mathbb{T}} f\bar{g}\, dt/(2\pi)$. Consequently

$$N_p(\rho) = N_p(\bar{\rho}), \qquad N_p(\rho) = |\rho| N_p(1/\rho)\ (\rho \neq 0), \qquad N_p(\rho) = N_{p'}(\bar{\rho}).$$

Together with the corresponding symmetries of GK_p in Section 2.1, these identities reduce the question to $|\rho| \leq 1$ and $1 < p < 2$; for $\rho = 1$ both sides equal 1. If $\alpha = 0$, the norm is $|\beta|/\sin(\pi/p)$, since U conjugates P_- isometrically to P_+ and $\|P_+\|_p = 1/\sin(\pi/p)$ [8]; so the bound is sharp. When the ratio ρ is real we write it as λ. In the table below, $p_0 \approx 1.2336026$ is the exponent defined before Theorem 1.1, $p_0' = p_0/(p_0-1)$, and $\tau(p)$ is the threshold of [7, Theorem 5.5] recalled under *Relation to other work* below.

Best result before this paper	**New result**
norms of H [12, 3] and P_+ [8]; [7] (2003): not sharp in general	for $p_0 \leq p \leq p_0'$, $p \neq 2$: $N_p(\rho) = \mathrm{GK}_p(\rho) \iff \rho \in \mathbb{R}$; for $p < p_0$ or $p > p_0'$, the bound is not sharp for real ratios $\lambda < 1$ sufficiently close to 1
[7] found the exact norms for real a, b ($\lvert\rho\rvert = 1$)	$N_p(\rho) > \mathrm{GK}_p(\rho)$ for every non-real ρ, $p \neq 2$
$\lambda = -1$ [12] (1972) and $\lambda = 0$ [8] (2000) are classical endpoints	$N_p(\lambda) = \mathrm{GK}_p(\lambda)$ for every real $\lambda \leq 0$
[7, Theorem 5.5] found the threshold $\tau(p)$ of a related inequality for real-valued functions	$\lambda \in (0,1)$, $1 < p < 2$: $N_p(\lambda) = \mathrm{GK}_p(\lambda) \iff$ two explicit conditions

Results.

- **Classification.** Theorem 1.1 characterizes, for all complex α, β and all $1 < p < \infty$, when the bound is sharp, a question not settled in [7].
- **Non-real ratios.** For $p \neq 2$ the inequality $N_p(\rho) \geq \mathrm{GK}_p(\rho)$ is strict for every non-real ratio (Theorem 3.2). These include the ratios of all real pairs a, b with $ab \neq 0$ ($|\rho| = 1$, $\rho \neq \pm 1$); for these [7] found the exact norm but did not say at which of them it exceeds the bound.
- **Nonpositive ratios.** The bound is sharp for every real $\lambda \leq 0$: Theorem 5.1 treats $-1 \leq \lambda \leq 0$, recovering the classical values at $\lambda = -1$ [12] and $\lambda = 0$ [8], and $\lambda < -1$ follows by reciprocal symmetry.
- **Positive ratios.** Theorem 5.3 gives an exact criterion for each $\lambda \in (0,1)$, and Theorem 5.2 gives the exponents for which the bound is sharp for every real ratio; for real-valued functions, [7, Theorem 5.5] found a related threshold.

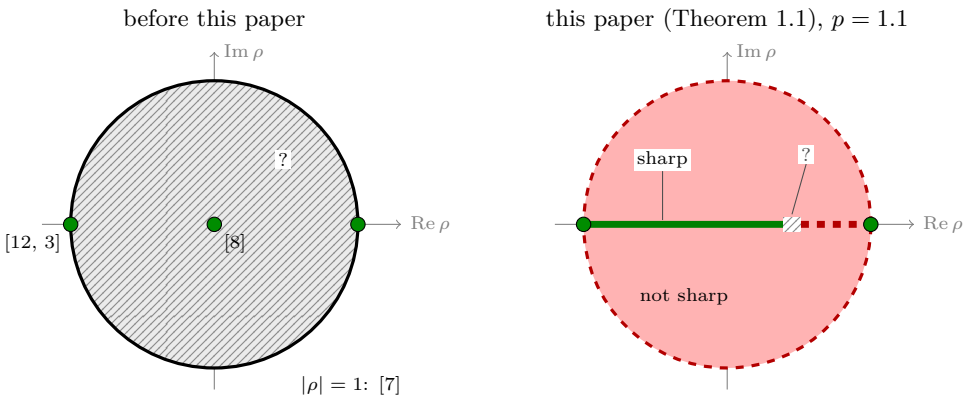


Figure 1: The disc $|\rho| \le 1$ of ratios; the other ratios reduce to it by $N_p(\rho) = |\rho| N_p(1/\rho)$. Left: what was known before this paper, for any $p \ne 2$: the ratios $\rho = -1$ [12, 3], $\rho = 0$ [8] and $\rho = 1$, where the bound is sharp, and the circle $|\rho| = 1$ of real a, b, where [7, Theorem 4.1] found the exact norm but did not say at which of these ratios it exceeds the bound. A question mark marks the ratios at which sharpness was not known before this paper (left) and the band in which it is still known only in part (right). Right: Theorem 1.1 for $p = 1.1$. For every non-real ρ, and for real λ with $\lambda_{\mathrm{loc}}(1.1) \le \lambda < 1$ ($\lambda_{\mathrm{loc}}(1.1) = 0.513\ldots$), the bound is not sharp; it is sharp for $-1 \le \lambda \le \lambda_1(1.1)$ and for $\lambda = 1$; the hatched piece of the real axis is the band of Figure 2 at $p = 1.1$. The value $\lambda_1(1.1) \approx 0.389$, defined in Section 4.3, was computed numerically. For $p_0 \le p \le p_0'$, $p \ne 2$, the bound is sharp on the whole segment $[-1, 1]$.

Figure 1 compares Theorem 1.1, for one exponent, with what was known before this paper.

The criterion for $0 < \lambda < 1$ involves a constant K and a threshold $\lambda_{\mathrm{loc}}(p)$, which come from the slope and the second variation at θ_0 of an obstacle described after Theorem 1.1. A number $\Omega(p)$ controls the first condition below as $\lambda \uparrow 1$ (Proposition 4.3). For $1 < p < 2$ put $s_0 = \sin\theta_0$, $c_0 = -\cos\theta_0 > 0$ and $A = (1 - s_0)/(2s_0)$. Let s_* be the unique root in $(0, 1)$ of $(2q-1)s^2 + (1 - q + 2qA)s - qA = 0$, and put

$$\Omega(p) = p\frac{A + s_*}{(s_*(1 - s_*))^q} - 2^q s_0^{p-2} c_0^{1-q}.$$

For $0 < \lambda < 1$ let $C = \mathrm{GK}_p(\lambda)$ and $K = 2(1-\lambda)C s_0^{p-2} c_0^{1-q}(C^2 + \lambda)^{q-1}$. Put $T(p) = s_0^2(pp'/4 - 1 - c_0^2)/c_0^4$ and

$$\lambda_{\mathrm{loc}}(p) = \begin{cases} (1 - \sqrt{T(p)})/(1 + \sqrt{T(p)}), & T(p) > 0, \\ 1, & T(p) \le 0. \end{cases}$$

The two conditions of the criterion are

$$C^p(1 + x)^p - (1 + \lambda x)^p \ge K x^q \quad \text{for every } x > 0, \qquad \lambda < \lambda_{\mathrm{loc}}(p). \tag{1.1}$$

The first condition is the restriction to $z = x > 0$ of the inequality

$$C^p|1 + z|^p - |1 + \lambda z|^p - K \operatorname{Re} z^q \ \ge \ 0 \qquad (z \in \mathbb{C}),$$

with the principal branch of z^q. Proposition 4.2, where this inequality is labeled $(*)$, shows that it is equivalent to sharpness for these λ. We call the first condition in (1.1) the *positive-axis test*; when $\Omega(p) < 0$ it fails for all λ sufficiently close to 1 (Proposition 4.3). Let $p_c \approx 1.213490304$ be the unique zero in $(1, 2)$ of $pp' - 4(1 + \cos^2(\pi/p))$ (Lemma 4.6). Then $T(p) > 0$ exactly for $1 < p < p_c$.

Propositions 4.3 and 5.14 show that Ω has exactly one zero $p_0 \approx 1.2336026$ in $(1, 1.268]$, with $\Omega < 0$ on $(1, p_0)$ and $\Omega > 0$ on $(p_0, 1.268]$. Since the bound is sharp for every real ratio when $3 - \sqrt{3} \le p < 2$ (Section 5.4) and $3 - \sqrt{3} < 1.268$, the sign of Ω beyond 1.268 is not needed.

Theorem 1.1 (Sharpness classification)**.** *Let $1 < p < \infty$ and $p_0' = p_0/(p_0 - 1)$.*

(a) If $p \ne 2$ and $\rho \notin \mathbb{R}$, then $N_p(\rho) > \mathrm{GK}_p(\rho)$.

(b) If $\lambda \le 0$ is real, then $N_p(\lambda) = \mathrm{GK}_p(\lambda)$.

(c) The equality $N_p(\lambda) = \mathrm{GK}_p(\lambda)$ holds for every real λ if and only if $p_0 \le p \le p_0'$. If $p < p_0$ or $p > p_0'$, then $N_p(\lambda) > \mathrm{GK}_p(\lambda)$ for every $\lambda < 1$ sufficiently close to 1.

(d) If $1 < p < 2$ and $0 < \lambda < 1$, then $N_p(\lambda) = \mathrm{GK}_p(\lambda)$ if and only if both conditions in (1.1) hold.

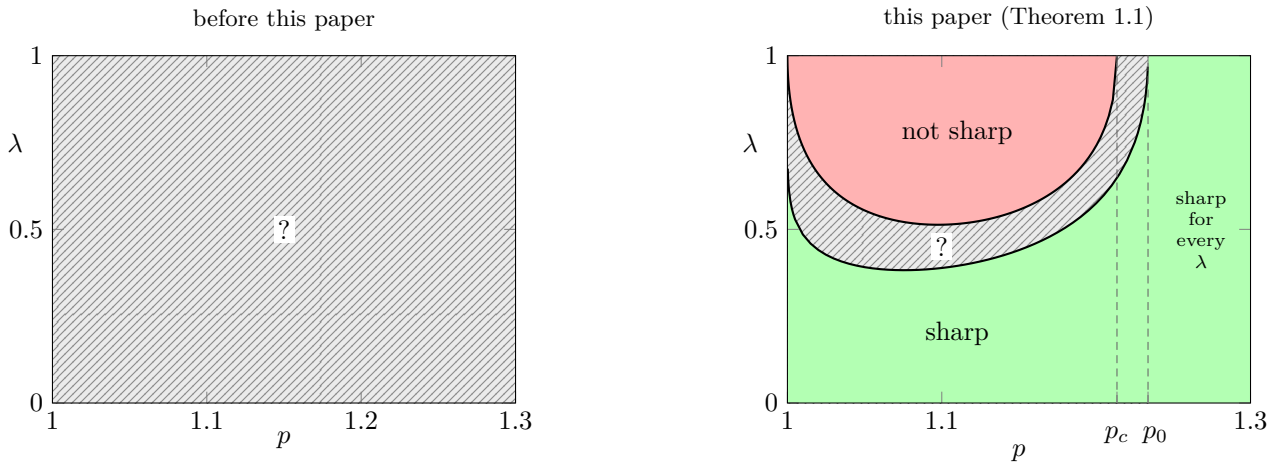


Figure 2: Sharpness of the bound for real ratios $0 < \lambda < 1$, shown for $1 < p < 1.3$. Left: before this paper, to our knowledge, no ratio $0 < \lambda < 1$ had been settled for $p \neq 2$. Right: Theorem 1.1. The bound is sharp on and below the lower curve $\lambda = \lambda_1(p)$, defined for $p < p_0$ (Section 4.3), and for every λ when $p_0 \le p < 2$. For $p < p_c$ it is not sharp on and above the upper curve $\lambda = \lambda_{\mathrm{loc}}(p)$; this curve reaches $\lambda = 1$ at $p = p_c$. In the hatched band the criterion of Theorem 1.1(d) still applies to each λ, but its outcome is known only in part; for example, when $p_c \le p < p_0$ the bound is not sharp for any $\lambda < 1$ sufficiently close to 1 (Theorem 1.1(c)). We do not know whether it is sharp anywhere in the band. The upper curve is the closed form for λ_{loc}; the lower curve was computed numerically, for illustration only, and the proofs do not use it.

Parts (a), (b) and (c) are proved as Theorem 3.2, Theorem 5.1 (with the reciprocal symmetry for $\lambda < -1$) and Theorem 5.2(a),(b), respectively, and part (d) as Theorem 5.3; for $p > 2$ part (d) applies with p' in place of p, since $N_p(\lambda) = N_{p'}(\lambda)$ and $\mathrm{GK}_p = \mathrm{GK}_{p'}$. Ratios $\lambda > 1$ reduce to $1/\lambda$ by the reciprocal symmetries above, and the bound is sharp for $\lambda = 1$, for $p = 2$ and for $\alpha = 0$. In terms of a and b: for $\alpha \neq 0$ the ratio β/α is real exactly when $\mathrm{Re}(a\bar{b}) = 0$, because $\beta\bar{\alpha} = (a + ib)(\bar{a} + i\bar{b})$ has imaginary part $2\,\mathrm{Re}(a\bar{b})$. If $\alpha = 0$, then $a = ib$ and $\mathrm{Re}(a\bar{b}) = 0$ as well. Hence, by Theorem 5.2(c), for $p_0 \le p \le p_0'$ and $p \neq 2$ the bound is sharp exactly when $\mathrm{Re}(a\bar{b}) = 0$.

When $1 < p < p_c$ we have $\lambda_{\mathrm{loc}}(p) < 1$, so the second condition of (1.1) is not automatic. Proposition 4.7 shows that the positive-axis test already fails at $\lambda = \lambda_{\mathrm{loc}}(p)$, and we do not know whether the second condition is ever needed in this range. When $p_c \le p < p_0$ the second condition is automatic, so sharpness is equivalent to the positive-axis test, and the test fails for λ close to 1. In both ranges let $(0, \lambda_1(p)]$ be the largest interval of the form $(0, \lambda]$ on which the positive-axis test holds (Section 4.3); then $\lambda_1(p) < 1$, and $\lambda_1(p) < \lambda_{\mathrm{loc}}(p)$ when $p < p_c$ (Propositions 4.3 and 4.7). So the bound is sharp for $0 < \lambda \le \lambda_1(p)$, and we do not know whether the set of ratios in $(0, 1)$ at which the bound is sharp is exactly this interval. When $p_0 \le p < 2$ the bound is sharp for every real ratio. Figure 2 shows these ranges and the two curves.

The proofs rest on a reduction to an obstacle problem. Write $f = z + \bar{w}$ with z and w analytic. Up to the positive factor $|zw|^q$, the integrand of $\int_{\mathbb{T}} \big(C^p|f|^p - |P_+ f + \rho P_- f|^p\big)\, d\theta/(2\pi)$ then depends only on $\log(|z|/|w|)$ and on the angle $\arg(zw)$, and minimizing over the first variable gives an obstacle $m_C(t)$. By Cole's theorem in the form [7, Theorem 2.1] and a plurisubharmonic argument (Theorem 2.2), $N_p(\rho) \le C$ holds if and only if some q-subtrigonometric function, that is, a 2π-periodic φ for which $|\zeta|^q \varphi(\arg \zeta)$ is subharmonic, lies below m_C. Such a function φ satisfies the arch inequality $\varphi(t) + \varphi(t + 2\pi/p) \ge 0$ [7, Proposition 3.4(1)]. At $C = \mathrm{GK}_p(\rho)$ the obstacle vanishes at two angles $2\pi/p$ apart, so a function below it satisfies the arch inequality with equality there. For a non-real ratio the two contact slopes do not balance. Moving both angles slightly then makes the sum of the obstacle values negative, and no q-subtrigonometric function can lie below m_C (Theorem 3.2).

For real ratios, symmetry cancels this first-order obstruction. The two contact zeros then force a q-subtrigonometric function below the obstacle to be a cosine on the arch between them. Its slope at the contact fixes the coefficient K of (4.1), and trigonometric convexity shows that such a function exists if and only if $K \cos(q|t|) \le m_C(t)$, $C = \mathrm{GK}_p(\lambda)$, on the whole circle. This reduces sharpness to the pointwise inequality $(*)$ in one complex variable (Proposition 4.2). Strict inequality for real ratios can come from the second-order term at the contact angle: along a smooth branch of radial critical points the obstacle has a local upper bound, and when $\lambda > \lambda_{\mathrm{loc}}(p)$ this bound violates the arch inequality near the contact angles (Theorem 4.4). It can also

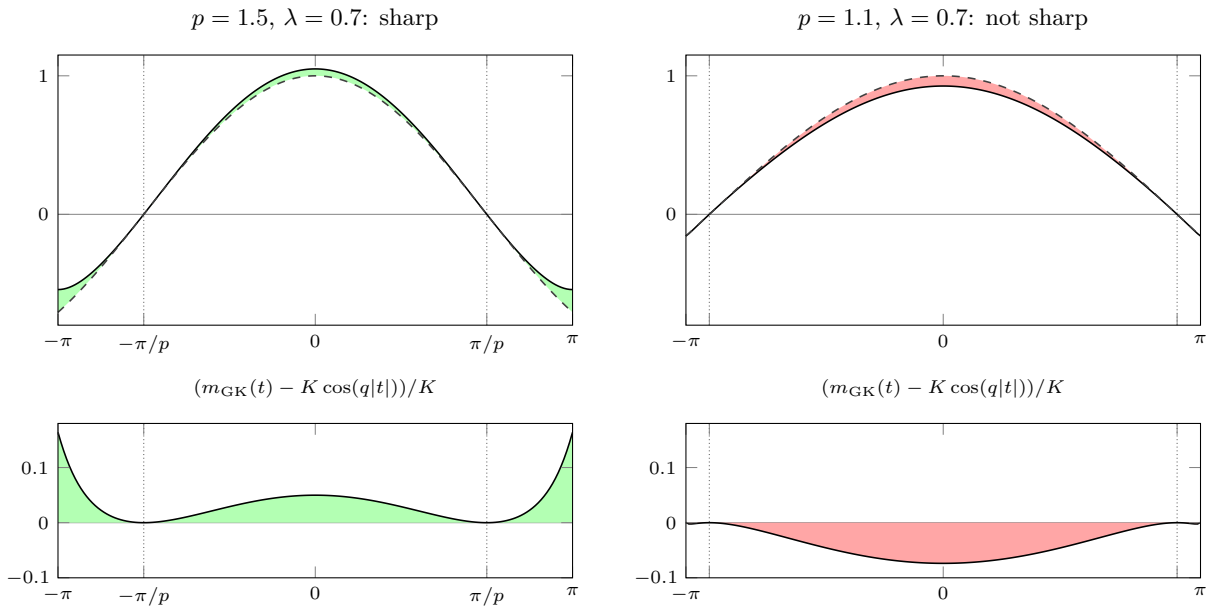


Figure 3: Proposition 4.2 at the ratio $\lambda = 0.7$, for two exponents. Green: the obstacle lies above the cosine; red: the cosine lies above the obstacle. Top row: the obstacle and the cosine, both divided by K, that is, $m_{\mathrm{GK}}(t)/K$ (solid) and $\cos(q|t|)$ (dashed), $q = p/2$, over the whole width $-\pi \le t \le \pi$. Bottom row: their difference $(m_{\mathrm{GK}}(t) - K\cos(q|t|))/K$. In the top row, both curves vanish at the contact angles $t = \pm\pi/p$ (dotted lines), and their slopes agree there (Section 4.1). By Proposition 4.2, the bound is sharp exactly when $m_{\mathrm{GK}}(t) \ge K\cos(q|t|)$ for every t. Left: $p = 1.5$, where the bound is sharp for every real ratio (Theorem 5.2(a)). Right: $p = 1.1$; since $0.7 > \lambda_{\mathrm{loc}}(1.1) = 0.513\ldots$, the bound is not sharp (Theorem 4.4). In the computed curves of this panel the cosine lies above the obstacle on almost the whole circle, including $t = 0$, where the comparison of $m_{\mathrm{GK}}(0)$ with K is the positive-axis test (Theorem 5.3). The curves were computed numerically, for illustration only, and the proofs do not use them.

come from failure of the positive-axis test. Evaluated at a suitable point as $\lambda \uparrow 1$, the test has a leading term with the sign of $\Omega(p)$, so it fails near 1 when $\Omega(p) < 0$ (Proposition 4.3); a separate estimate shows that it also fails at $\lambda = \lambda_{\mathrm{loc}}(p) < 1$ (Proposition 4.7). Figure 3 shows the obstacle and the cosine for one ratio at two exponents.

To prove sufficiency we establish $(*)$ by a critical-point analysis in the spirit of the proof of [8, Lemma 2.4]. After the Möbius change of variable $w = (1 + \lambda z)/(1 + z)$, the left side of $(*)$ equals, up to a positive factor, $(1-\lambda)^p(C^p - |w|^p) - K\operatorname{Re} Q^q$ with $Q = (w - \lambda)(1 - \bar{w})$. On a circle $|w| = \text{const}$ the first term is constant and Q runs over an ellipse, so minimizing over the circle means maximizing $\operatorname{Re} Q^q$ over that ellipse. The interior critical points form one explicit curve, together with its inversion images when $\lambda > 0$. Along the curve the derivative of the quotient of the left side of $(*)$ by $|1 + z|^p$ has the sign of $K - K_f$, where the function K_f does not involve C or K, and $K_f = K$ at the contact angle θ_0. If K_f decreases throughout, this quotient has its only minimum at θ_0, where it vanishes. If K_f first increases and then decreases, and $K_f'(\theta_0) < 0$, which is what $\lambda < \lambda_{\mathrm{loc}}(p)$ gives, then the only other possible minimum is the end of the curve on the positive real axis, which the positive-axis test controls. This gives Theorem 5.3.

For $\lambda \le 0$, and for every $\lambda \in (-1, 1)$ when $3 - \sqrt{3} \le p < 2$, the function K_f is strictly decreasing; the second case uses the product formula for the sine. For $p < 4/3$ a power series with positive coefficients shows that K_f has at most one turning point. Since $3 - \sqrt{3} < 4/3$, these two ranges of p cover $(1, 2)$ (Section 5.3). For the exponent range of Theorem 5.2 it remains to prove that the positive-axis test holds for every $\lambda \in (0, 1)$ when $p_0 \le p < 3 - \sqrt{3}$. This step uses interval arithmetic, with analytic bounds at the endpoints and in the tails (Section 5.6). It is delicate at $p = p_0$ as $\lambda \to 1$, where the leading term of the test vanishes, and there we expand in $\eta = -\log\lambda$. Interval arithmetic, again with analytic bounds at the ends of the ranges, also proves the signs of Ω, the enclosure of p_c and strict inequality at the local threshold, including the range just below p_c; Appendix B (Table 1) lists all domains.

Relation to other work. For $1 < p < 2$, [7, Theorem 5.5] determines the number $\tau(p) \in (0,1)$ such that, for real-valued u and $v = H_0 u$,

$$\|(u^2 + t^2 v^2)^{1/2}\|_p \le (1 + t^2 \tan^2(\pi/(2p)))^{1/2} \|u\|_p$$

holds for all u when $\tau(p) \le t \le 1$ and fails for some u when $0 < t < \tau(p)$. For real mean-zero u, put $\lambda = (1-t)/(1+t)$; then $u + itv = (1+t)(P_+ + \lambda P_-)u$. On this subspace the constant in the displayed inequality becomes, after division by $1+t$, $(1+t^2\tan^2(\pi/(2p)))^{1/2}/(1+t)$, whereas our criterion compares the norm on all complex inputs with $\mathrm{GK}_p(\lambda)$. For example, at $t = 1$, that is, $\lambda = 0$, the normalized constant is $1/(2\cos(\pi/(2p))) = \sin(\pi/(2p))\,\mathrm{GK}_p(0)$, strictly smaller than $\mathrm{GK}_p(0) = 1/\sin(\pi/p)$.

Ding, Grafakos and Zhu [1] give a direct proof on the real line of the norm formula for real coefficients. For real a, b, Long [11] shows that the completely bounded norm of $aI + bH$ on $L^p(\mathbb{R})$ equals the norm found in [7], and notes in [11, Remark 8.9] that this formula does not extend to arbitrary complex coefficients (see also [11, Remark 1.8]).

Section 6, which is not in the short version of this paper [4], collects further results that the classification does not use.

1.1 Circle and real-line norms

Let $T_{\mathbb{R}}$ be the Fourier multiplier on $L^p(\mathbb{R})$ with symbol $m(\xi) = \alpha$ for $\xi > 0$ and $m(\xi) = \beta$ for $\xi < 0$. The value at zero is immaterial on $\mathbb{R}$. On the circle, let T_+ and T_- be the Fourier multipliers with symbol $m(n)$ for $n \neq 0$ and with value α, respectively β, at $n = 0$, and let $T_0 = (T_+ + T_-)/2$. Thus $T_+ = \alpha P_+ + \beta P_-$, whereas $T_0 = aI + bH_0$ for the conjugate-function operator H_0 (so $H_0 1 = 0$), with $a = (\alpha+\beta)/2$ and $b = (\beta - \alpha)/(2i)$.

Lemma 1.2 (Norm correspondence)**.** *For every $1 < p < \infty$ and $\alpha, \beta \in \mathbb{C}$,*

$$\|T_+\|_{L^p(\mathbb{T})} = \|T_-\|_{L^p(\mathbb{T})} = \|T_0\|_{L^p(\mathbb{T})} = \|T_{\mathbb{R}}\|_{L^p(\mathbb{R})}.$$

Proof. Fix $0 < \nu < 1$, a trigonometric polynomial F and a nonzero Schwartz function g whose Fourier transform has compact support. For all sufficiently small $h > 0$, the Fourier supports of $g(hx)e^{i\nu x}e^{inx}$ lie entirely in the positive half-line for $n \ge 0$ and in the negative half-line for $n < 0$. Consequently the multiplier acts exactly as

$$T_{\mathbb{R}}\big(g(hx)e^{i\nu x}F(x)\big) = g(hx)e^{i\nu x}(T_+F)(x).$$

For any continuous 2π-periodic A, averaging over periods gives

$$\lim_{h\downarrow 0} h\int_{\mathbb{R}} |g(hx)|^p A(x)\,dx = \left(\int_{\mathbb{R}} |g(x)|^p\,dx\right)\left(\int_{\mathbb{T}} A(t)\,\frac{dt}{2\pi}\right).$$

Apply this identity to $A = |T_+F|^p$ and $A = |F|^p$. It follows that $\|T_+\| \le \|T_{\mathbb{R}}\|$. Using $e^{-i\nu x}$ instead proves $\|T_-\| \le \|T_{\mathbb{R}}\|$; since $T_0 = (T_+ + T_-)/2$, the triangle inequality gives the same bound for T_0.

Conversely, take a Schwartz function f with $\widehat{f} \in C_c^\infty(\mathbb{R}\setminus\{0\})$ and set

$$F_N(t) = \sum_{k\in\mathbb{Z}} f\big(N(t+2\pi k)\big).$$

Its nth Fourier coefficient is $(2\pi N)^{-1}\widehat{f}(n/N)$, where $\widehat{f}(\xi) = \int f(x)e^{-ix\xi}\,dx$. Only finitely many coefficients are nonzero, and the constant coefficient vanishes. Because $m(n/N) = m(n)$ for $n \neq 0$, applying any of T_+, T_-, T_0 to F_N gives $\sum_{k\in\mathbb{Z}}(T_{\mathbb{R}}f)\big(N(t+2\pi k)\big)$, the corresponding periodization of $T_{\mathbb{R}}f$. Both f and $T_{\mathbb{R}}f$ are Schwartz. Their rapidly decreasing tails therefore give

$$\lim_{N\to\infty} 2\pi N\|F_N\|_{L^p(\mathbb{T})}^p = \|f\|_{L^p(\mathbb{R})}^p, \qquad \lim_{N\to\infty} 2\pi N\|T_jF_N\|_{L^p(\mathbb{T})}^p = \|T_{\mathbb{R}}f\|_{L^p(\mathbb{R})}^p$$

for $j \in \{+,-,0\}$. Hence $\|T_{\mathbb{R}}f\|_p \le \|T_j\|\|f\|_p$. Such f are dense in $L^p(\mathbb{R})$ for $1 < p < \infty$ (cut off the Fourier transform smoothly, first at infinity and then near zero), which proves the reverse inequalities. □

In particular, the classical norms in [7, 8, 12] apply to the projection convention used here.

2 The Gohberg–Krupnik bound and the obstacle problem

This section records the symmetries of GK_p (Section 2.1), reduces the inequality $N_p(\rho) \le C$ to an obstacle problem (Theorem 2.2), and proves a chord inequality (Lemma 2.3); Sections 3 and 4 use the last two.

2.1 The function GK_p

The matrix that defines $\mathrm{GK}_p(\alpha,\beta)$ in the introduction gives $\mathrm{GK}_p(\alpha,\beta) = |\alpha|\mathrm{GK}_p(\beta/\alpha)$ when $\alpha \neq 0$, $\mathrm{GK}_p(0,\beta) = |\beta|/\sin(\pi/p)$, and $\mathrm{GK}_2(\alpha,\beta) = \max(|\alpha|,|\beta|)$. The squared Frobenius norm and the squared modulus of the determinant of this matrix are $|\alpha|^2 + |\beta|^2 + |\beta-\alpha|^2\cot^2(\pi/p)$ and $|\alpha\beta|^2$. They determine its singular values and are unchanged by interchanging α and β, conjugating both coefficients or replacing p by p'. Together with homogeneity, this gives

$$\mathrm{GK}_p(\bar\rho) = \mathrm{GK}_p(\rho), \qquad \mathrm{GK}_p(\rho) = |\rho|\mathrm{GK}_p(1/\rho)\ (\rho\neq 0), \qquad \mathrm{GK}_p(\rho) = \mathrm{GK}_{p'}(\rho).$$

In the normalized case, $\mathrm{GK}_p(\rho)^2$ is the larger root of

$$X^2 - \big(1 + |\rho|^2 + |1-\rho|^2\cot^2\theta_0\big)X + |\rho|^2 = 0, \tag{2.1}$$

that is, the square of the larger singular value of the matrix $\left(\begin{smallmatrix} 1 & (\rho-1)\cot\theta_0 \\ 0 & \rho \end{smallmatrix}\right)$. In particular $\mathrm{GK}_p(0) = 1/\sin(\pi/p)$ and, for $1<p<2$, $\mathrm{GK}_p(-1) = \tan(\pi/(2p))$: the norms of P_+ [8] and of H [12]. Pichorides works with real-valued functions, but for real a,b the real and complex norms of $aI+bH$ coincide [7, p. 251]. For $\rho=\lambda\in[-1,1)$ and $1<p<2$, (2.1) is equivalent to

$$(1-\lambda)\,C = s_0\,(C^2-\lambda), \qquad C = \mathrm{GK}_p(\lambda), \quad s_0 = \sin\theta_0, \tag{2.2}$$

as one checks by squaring; C is the larger root of this quadratic, namely $C = \big((1-\lambda)+\sqrt{(1-\lambda)^2+4\lambda s_0^2}\big)/(2s_0)$.

2.2 Reduction to an obstacle problem

For a trigonometric polynomial f put $z = P_+f$ and $w = \overline{P_-f}$: both are analytic polynomials, $w(0)=0$, $f = z+\bar w$ on $\mathbb{T}$, and every such pair arises. Hence, by density, $N_p(\rho)\le C$ if and only if

$$\int_{\mathbb{T}} G_C(z,w)\,\frac{d\theta}{2\pi} \ge 0, \qquad G_C(z,w) := C^p|z+\bar w|^p - |z+\rho\bar w|^p, \tag{2.3}$$

for all polynomials z,w with $w(0)=0$. Where $zw\neq 0$ put $y=\log(|z|/|w|)$ and $t=\arg(zw)$. Since $|z+\bar w|^2 = |zw|\,(2\cosh y + 2\cos t)$ and $|z+\rho\bar w|^2 = |zw|\,(e^y + |\rho|^2e^{-y} + 2\,\mathrm{Re}(\bar\rho e^{it}))$,

$$G_C(z,w) = |zw|^q\,\Gamma_C(y,t), \qquad \Gamma_C(y,t) := C^p(2\cosh y + 2\cos t)^q - \big(e^y + |\rho|^2e^{-y} + 2\,\mathrm{Re}(\bar\rho e^{it})\big)^q.$$

The *obstacle* is $m_C(t) := \inf_{y\in\mathbb{R}}\Gamma_C(y,t)$. Testing (2.3) with $z\equiv 1$, $w\equiv 0$ shows that $N_p(\rho)\ge 1$, and testing it with $z\equiv 0$, $w(\zeta)=\zeta$ shows that $N_p(\rho)\ge|\rho|$. From now on we assume that $C>\max(1,|\rho|)$. Then $\Gamma_C(y,t)\ge (C^p-1+o(1))e^{qy}$ as $y\to+\infty$ and $\Gamma_C(y,t)\ge(C^p-|\rho|^p+o(1))e^{-qy}$ as $y\to-\infty$, uniformly in t; so the infimum is attained in a fixed interval $[-Y_0,Y_0]$, and m_C is finite and continuous. If ρ is real, m_C is even. In terms of $\zeta = e^{y+it}$ and $G(\zeta) = G_C(\zeta,1) = C^p|\zeta+1|^p - |\zeta+\rho|^p$ we have $\Gamma_C(y,t) = |\zeta|^{-q}\,G(\zeta)$, so $m_C(t) = \inf_{r>0} r^{-q}G(re^{it})$.

Definition 2.1. A 2π-periodic $\varphi:\mathbb{R}\to[-\infty,\infty)$ is *q-subtrigonometric* if it is finite and $\zeta\mapsto|\zeta|^q\varphi(\arg\zeta)$ is subharmonic on $\mathbb{C}\setminus\{0\}$.

In polar coordinates $\Delta(r^q\varphi(\theta)) = r^{q-2}(\varphi''+q^2\varphi)$, so φ is q-subtrigonometric if and only if φ is continuous and $\varphi''+q^2\varphi\ge 0$ in the sense of distributions on $\mathbb{R}$ ("trigonometrically q-convex" functions, [10, Lecture 8]). Equivalently, by [7, Proposition 3.3] applied locally with q in place of p, $x\mapsto\varphi(x/q)$ is trigonometrically convex with frequency one in the sense of [7, (3.1)]; indeed $f(x)=\varphi(x/q)$ satisfies $f''+f = q^{-2}(\varphi''+q^2\varphi)(x/q)$. We apply the results of [7, Section 3] to φ through this rescaling.

Theorem 2.2 (Obstacle characterization)**.** *Let $1<p<\infty$, $\rho\in\mathbb{C}$ and $C>\max(1,|\rho|)$. Then $N_p(\rho)\le C$ if and only if there is a q-subtrigonometric φ with $\varphi\le m_C$ on $\mathbb{R}$. Hence $N_p(\rho) = \inf\{C : \text{such a } \varphi \text{ exists}\}$.*

Proof. *"If".* The function $S(\zeta) = |\zeta|^q \varphi(\arg\zeta)$ is subharmonic on $\mathbb{C}\setminus\{0\}$ and tends to 0 at 0 (φ is bounded), so it extends subharmonically to $\mathbb{C}$ with $S(0) = 0$. Then, as in [8, Remark 2.3], $\Psi(z,w) = S(zw)$ is plurisubharmonic and $\Psi \le G_C$: where $zw \neq 0$, $\Psi = |zw|^q\varphi(t) \le |zw|^q m_C(t) \le G_C$; where $zw = 0$, $\Psi = 0 \le G_C$, since $G_C(z,0) = (C^p - 1)|z|^p$ and $G_C(0,w) = (C^p - |\rho|^p)|w|^p$. For polynomials with $w(0) = 0$ the sub-mean-value property of the subharmonic function $\zeta \mapsto \Psi(z(\zeta), w(\zeta))$ on the disc gives $\int_{\mathbb{T}} G_C(z,w)\, d\theta/(2\pi) \ge \int_{\mathbb{T}} \Psi(z,w)\, d\theta/(2\pi) \ge \Psi(z(0),0) = 0$; apply (2.3).

"Only if". Let g be a polynomial with $g(0) = 0$, $g \not\equiv 0$, and write $g = \zeta^k B h$ with $k \ge 1$, B a finite Blaschke product and h zero-free in the open disc. For a real trigonometric polynomial $y(\theta)$ let Y be the polynomial with $\operatorname{Re} Y = y$ on $\mathbb{T}$ and $\operatorname{Im} Y(0) = 0$, and put $z = Bh^{1/2}e^{Y/2}$, $w = \zeta^k h^{1/2} e^{-Y/2}$. Note that h is a polynomial. Write $h = h_0 \prod_j (\zeta - e^{i\gamma_j})$ with h_0 zero-free on the closed disc; then $h_0^{1/2}$ is holomorphic near the closed disc, and each $(\zeta - e^{i\gamma_j})^{1/2}$ has a branch in the disc algebra, since $\zeta - e^{i\gamma_j}$ takes values in a closed half-plane with 0 on its boundary. Thus $h^{1/2} = h_0^{1/2} \prod_j (\zeta - e^{i\gamma_j})^{1/2}$ is in the disc algebra, and hence so are z and w; moreover $w(0) = 0$, $zw = g$, and $|z|^2 = |g|e^{y}$, $|w|^2 = |g|e^{-y}$ on $\mathbb{T}$. Since $f = z + \bar{w}$ is continuous with $P_+ f = z$, $P_- f = \bar{w}$, the bound $N_p(\rho) \le C$ gives

$$\int_{\mathbb{T}} |g|^q\, \Gamma_C\big(y(\theta), \arg g(\theta)\big) \frac{d\theta}{2\pi} \ge 0 \qquad \text{for every real trigonometric polynomial } y. \tag{2.4}$$

To choose y, fix $\varepsilon > 0$. For each t choose $y_t \in [-Y_0, Y_0]$ with $\Gamma_C(y_t, t) < m_C(t) + \varepsilon$; since $\Gamma_C(y_t, \cdot)$ and m_C are continuous, this persists on an open arc around t, so there is a step function $\tilde{y}$ with values in $[-Y_0, Y_0]$ and $\Gamma_C(\tilde{y}(t), t) < m_C(t) + \varepsilon$ for all t. Let $u = \tilde{y}(\arg g)$, a bounded measurable function on $\mathbb{T}$. The Fejér means of u are real trigonometric polynomials bounded by Y_0 that converge to u almost everywhere; dominated convergence in (2.4) gives $\int_{\mathbb{T}} |g|^q (m_C(\arg g) + \varepsilon)\, d\theta/(2\pi) \ge 0$. Since $\varepsilon > 0$ is arbitrary, $\int_{\mathbb{T}} M(g)\, d\theta/(2\pi) \ge 0$ for every polynomial g with $g(0) = 0$, where $M(\zeta) = |\zeta|^q m_C(\arg\zeta)$, $M(0) = 0$, is continuous on $\mathbb{C}$.

Now apply Cole's theorem (see [3]) in the form [7, Theorem 2.1] (the case $E = \{0\}$ there): a continuous M with $\int_{\mathbb{T}} M(g)\, d\theta/(2\pi) \ge 0$ for every analytic polynomial g satisfying $g(0) = 0$ has a subharmonic minorant nonnegative at 0. Our M satisfies these hypotheses, so it has such a minorant. Take the upper-semicontinuous regularization of the supremum of all subharmonic minorants of M. The family is locally bounded above by the continuous function M, so this regularization is subharmonic and still lies below M; it is therefore the largest minorant, denoted $\underline{M}$. Then $\underline{M} \le M$ and $\underline{M}(0) \ge 0$. For $\tau > 0$, $\zeta \mapsto \tau^{-q}\underline{M}(\tau\zeta)$ is a subharmonic minorant of $\tau^{-q}M(\tau\zeta) = M(\zeta)$; so it is $\le \underline{M}$, and by symmetry $\underline{M}(\tau\zeta) = \tau^q \underline{M}(\zeta)$. Hence $\underline{M}(\zeta) = |\zeta|^q\varphi(\arg\zeta)$ with $\varphi(t) = \underline{M}(e^{it}) \le m_C(t)$. If $\varphi(t_0) = -\infty$, then $\underline{M} = -\infty$ on a ray; this is impossible, since a ray is not polar and $\underline{M} \not\equiv -\infty$. Thus φ is finite and its homogeneous extension is the subharmonic function $\underline{M}$; by Definition 2.1, φ is q-subtrigonometric. □

2.3 An angular chord inequality

Lemma 2.3 (Chord inequality)**.** *Let φ be q-subtrigonometric, $c \in \mathbb{R}$ and $0 < s \le \pi/q$. Then $\varphi(c-s)+\varphi(c+s) \ge 2\cos(qs)\,\varphi(c)$.*

Proof. Let $\nu(t) = \sin\big(q(s - |t-c|)\big)$ for $|t-c| \le s$ and $\nu = 0$ otherwise. Then $\nu \ge 0$ (as $qs \le \pi$), ν is Lipschitz with compact support, and ν' jumps by q at $c \pm s$ and by $-2q\cos(qs)$ at c, so $\nu'' + q^2\nu = q\,(\delta_{c-s} + \delta_{c+s}) - 2q\cos(qs)\,\delta_c$. Pairing the positive distribution $\varphi'' + q^2\varphi$ with ν (approximate ν by smooth functions; φ is continuous) gives $0 \le q\,[\varphi(c-s) + \varphi(c+s) - 2\cos(qs)\varphi(c)]$. □

Sections 3 and 4 use only the case $s = \pi/(2q)$, the arch inequality of [7, Proposition 3.4(1)]; for $s > \pi/(2q)$ the lemma gives, through Theorem 2.2, further necessary conditions for $N_p(\rho) \le C$.

3 Strict inequality for non-real ratios

In this section $1 < p < \infty$, $p \neq 2$ and $\rho \in \mathbb{C}$; G and m_C are as in Section 2.2, and $\mathrm{GK} = \mathrm{GK}_p(\rho)$ is given by (2.1). We show that for non-real ρ the obstacle at $C = \mathrm{GK}$ violates the arch inequality; this gives Theorem 1.1(a). If $\rho \neq 1$, then $\mathrm{GK} > \max(1, |\rho|)$: the left side of the quadratic in (2.1) equals $(X-1)(X-|\rho|^2) - |1-\rho|^2 \cot^2\theta_0\, X$, which is negative at $X = \max(1, |\rho|^2)$.

A 2π-periodic q-subtrigonometric function φ satisfies the *arch inequality* $\varphi(t) + \varphi(t + \pi/q) \ge 0$ for all t [7, Proposition 3.4(1)] (it is also Lemma 2.3 with $s = \pi/(2q)$), and $\pi/q = 2\pi/p$. With Theorem 2.2 this gives the

arch test: for every $C > \max(1, |\rho|)$,

$$m_C(t_1) + m_C(t_1 + 2\pi/p) < 0 \ \text{ for some } t_1 \quad \Longrightarrow \quad \|P_+ + \rho P_-\|_p > C. \tag{3.1}$$

3.1 The angular window

Let $C > \max(1, |\rho|)$ and $\rho \neq 1$. Expanding $|\rho + \zeta|^2 - C^2|1 + \zeta|^2 > 0$ shows that

$$U_C := \{\zeta : G(\zeta) < 0\} = \{\zeta : |\zeta + \rho| > C|\zeta + 1|\}$$

is the open disc with center $c = -(C^2 - \rho)/(C^2 - 1)$ and radius R, $R^2 = |c|^2 - (C^2 - |\rho|^2)/(C^2 - 1)$. It contains -1 and does not contain 0 (since $|\rho| < C$). Because $G(re^{it})/r^q \to +\infty$ as $r \to 0$ and as $r \to \infty$, we have $m_C(t) < 0$ exactly when the ray $\arg\zeta = t$ meets U_C, and $m_C(t) = 0$ when the ray is tangent to ∂U_C (the infimum is then attained at the tangent point). So $\{t : m_C(t) < 0\}$ is the open *window* $(\arg c - \psi, \arg c + \psi)$, where $\psi = \psi(C) \in (0, \pi/2)$ is half the angle that U_C subtends at 0: $\sin\psi = R/|c|$, that is,

$$\cos^2\psi(C) = \frac{(C^2 - |\rho|^2)(C^2 - 1)}{|C^2 - \rho|^2}. \tag{3.2}$$

Lemma 3.1 (Window width at the Gohberg–Krupnik value)**.** *Let $\rho \neq 1$. Then $C = \mathrm{GK}_p(\rho)$ is the unique $C > \max(1, |\rho|)$ at which the window has width $2\psi(C) = 2\pi/\max(p, p')$.*

Proof. Since $\psi \in (0, \pi/2)$ and $\pi/\max(p, p') \in (0, \pi/2)$, the condition is $\cos^2\psi = \cos^2\theta_0$ (note $\cos^2(\pi/p) = \cos^2(\pi/p')$). With $X = C^2$, (3.2) turns it into $(X - |\rho|^2)(X - 1) = \cos^2\theta_0\,(X^2 - 2X\,\mathrm{Re}\,\rho + |\rho|^2)$, that is, after division by $\sin^2\theta_0$,

$$X^2 - \frac{1 + |\rho|^2 - 2\cos^2\theta_0\,\mathrm{Re}\,\rho}{\sin^2\theta_0}\,X + |\rho|^2 = 0.$$

Since $(1 + |\rho|^2)/\sin^2\theta_0 = (1 + |\rho|^2)(1 + \cot^2\theta_0)$, the middle coefficient is $1 + |\rho|^2 + |1 - \rho|^2\cot^2\theta_0$: this is the quadratic of (2.1). Its roots have product $|\rho|^2$, so the smaller root is $|\rho|^2/\mathrm{GK}^2 < 1$ and is excluded by $C > 1$. □

3.2 Non-real ratios

Theorem 3.2 (Strict inequality off the real line)**.** *Let $p \neq 2$ and $\rho \in \mathbb{C} \setminus \mathbb{R}$. Then $\|P_+ + \rho P_-\|_p > \mathrm{GK}_p(\rho)$.*

Proof. Put $C = \mathrm{GK}_p(\rho)$; $C > \max(1, |\rho|)$ since $\rho \neq 1$.

The two tangent lines from 0 to ∂U_C touch it at ζ_1, ζ_2; their arguments are zeros of m_C that are $2\pi/p$ apart. Indeed, these points have the same modulus $r = (|c|^2 - R^2)^{1/2}$ and the arguments $\arg c \mp \psi$. By Lemma 3.1 the angular gap is $2\pi/\max(p, p')$. If $p > 2$ it is $2\pi/p$. If $p < 2$ it is $2\pi/p' = 2\pi - 2\pi/p$. After a relabeling, in both cases $t_2 \equiv t_1 + 2\pi/p \pmod{2\pi}$, where $t_k = \arg\zeta_k$. By the window description, $m_C(t_1) = m_C(t_2) = 0$, attained at radius r.

At the fixed radius r put $\mu(t) := G(re^{it})/r^q$. Then $m_C \le \mu$ and $\mu(t_1) = \mu(t_2) = 0$, so

$$M(t) := m_C(t) + m_C(t + 2\pi/p) \ \le \ \widetilde{M}(t) := \mu(t) + \mu(t + 2\pi/p), \qquad \widetilde{M}(t_1) = 0.$$

To compute the first-order term of $\widetilde{M}$ at t_1, write $w = (\zeta + \rho)/(\zeta + 1)$ and $L = \log C - \log|w|$. Then $G = |\rho + \zeta|^p(e^{pL} - 1)$, and $L(\zeta_k) = 0$. So at ζ_k, $(d/dt)\,G(re^{it}) = p\,|\rho + \zeta_k|^p\,dL/dt$ and $dL/dt = -\,\mathrm{Re}\,\big(i\zeta\,w'/w\big)$, with $w'/w = (1 - \rho)/\big((\zeta + \rho)(\zeta + 1)\big)$. The ray through ζ_k is tangent to the level circle $|w| = C$ at ζ_k, so $\mathrm{Re}(\zeta_k w'/w) = 0$, and therefore $|\,\mathrm{Re}(i\zeta_k w'/w)| = |\zeta_k w'/w|$. With $|\zeta_k + \rho| = C|\zeta_k + 1|$:

$$|\mu'(t_k)| = p\,|1 - \rho|\,r^{1-q}\,C^{p-1}\,|\zeta_k + 1|^{p-2} > 0.$$

The signs of $\mu'(t_1)$ and $\mu'(t_2)$ are opposite. The circle $|\zeta| = r$ meets ∂U_C orthogonally at ζ_1 and ζ_2, and its arc between them lies in U_C. So, as t increases, μ changes from positive to negative at one end of the window and from negative to positive at the other. Hence

$$\widetilde{M}'(t_1) = \pm\,p\,|1 - \rho|\,r^{1-q}\,C^{p-1}\big(|\zeta_1 + 1|^{p-2} - |\zeta_2 + 1|^{p-2}\big).$$

This first-order term vanishes only for real ρ. Since $p \neq 2$, $\widetilde{M}'(t_1) = 0$ if and only if $|\zeta_1 + 1| = |\zeta_2 + 1|$, that is, if and only if -1 lies on the perpendicular bisector of the chord $\zeta_1\zeta_2$. Both 0 and c are equidistant from ζ_1

and ζ_2, so this bisector is the line through 0 and c, and the condition becomes $c \in \mathbb{R}$, that is, $C^2 - \rho \in \mathbb{R}$, or equivalently $\rho \in \mathbb{R}$.

Hence, for $\rho \notin \mathbb{R}$, we get $\widetilde{M}'(t_1) \neq 0$ and $\widetilde{M}(t_1) = 0$. So $M(t) \leq \widetilde{M}(t) < 0$ for some t near t_1, and (3.1) gives $\|P_+ + \rho P_-\|_p > C$. □

Remark (real ratios). For real $\rho = \lambda$ the configuration is symmetric about the real axis and $\widetilde{M}'(t_1) = 0$, so the arch test must be decided at second order. Theorem 4.4 does this when $\lambda > \lambda_{\mathrm{loc}}(p)$; unlike the first-order bound above, it needs a nondegenerate minimum in the radial variable.

4 Real ratios: the criterion and strict inequality

This section proves the one-variable criterion (Proposition 4.2), on which Section 5 rests, and three strict inequalities for positive ratios: near $\lambda = 1$ when $\Omega(p) < 0$ (Proposition 4.3), for $\lambda > \lambda_{\mathrm{loc}}(p)$ (Theorem 4.4) and at $\lambda = \lambda_{\mathrm{loc}}(p)$ (Proposition 4.7). Throughout, the ratio is real and is denoted by $\lambda \in [-1, 1)$, and, unless stated otherwise, $1 < p < 2$. From now on the letter ρ denotes a modulus $|w|$ of $w = (\zeta + \lambda)/(\zeta + 1)$ (in particular a point of $(\lambda, 1)$ when $\zeta > 0$), not a ratio. Powers are principal ($-\pi < \arg z \leq \pi$), so $\operatorname{Re} z^q = |z|^q \cos(q \arg z)$ is continuous on $\mathbb{C}$, with value 0 at $z = 0$.

4.1 Contact at $\pm\pi/p$ and the constant K

For $C > 1$ the function G of Section 2.2 is negative exactly on the open disc $U_C = \{|\zeta+\lambda| > C|\zeta+1|\} = \{|\zeta+\kappa| < R\}$ with $\kappa = (C^2 - \lambda)/(C^2 - 1) > 0$ and $\kappa^2 - R^2 = (C^2 - \lambda^2)/(C^2 - 1) > 0$, so that $R = C(1 - \lambda)/(C^2 - 1)$ and $0 \notin \overline{U_C}$.

We abbreviate $\mathrm{GK} = \mathrm{GK}_p(\lambda)$ and write m_{GK} for m_C with $C = \mathrm{GK}$.

Lemma 4.1 (Contact at the Gohberg–Krupnik value)**.** *Let $\lambda \in [-1, 1)$ and $C > 1$. Then $m_C(\pm\theta_0) \geq 0$ if and only if $C \geq \mathrm{GK}_p(\lambda)$. At $C = \mathrm{GK}_p(\lambda)$ one has $m_C(\pm\theta_0) = 0$, attained only at $\zeta_* = r_* e^{\pm i\theta_0}$, $r_* = \kappa c_0$, where $|\zeta_* + \lambda| = C|\zeta_* + 1|$.*

Proof. On the ray $\arg\zeta = \theta_0$ we have $r^{-q}G \to +\infty$ as $r \to 0$ or $r \to \infty$; so $m_C(\theta_0) < 0$ if and only if the ray meets U_C, and $m_C(\theta_0) = 0$ if and only if it touches $\overline{U_C}$ without meeting U_C. Since $\cos\theta_0 < 0$, the ray meets U_C if and only if the distance $\kappa \sin\theta_0$ from the center $-\kappa$ to the line is less than R. Tangency, $\kappa^2 \sin^2\theta_0 = R^2$, reads $(C^2 - \lambda)^2 \cos^2\theta_0 = (C^2 - \lambda^2)(C^2 - 1)$, which with $X = C^2$ is (2.1); its smaller root is $\lambda^2/\mathrm{GK}^2 < 1$. So $\kappa^2 \sin^2\theta_0 - R^2 = (C^2 - \lambda^2)/(C^2 - 1) - \kappa^2 \cos^2\theta_0$, a continuous function of $C > 1$, vanishes only at $C = \mathrm{GK}$; it tends to $-\infty$ as $C \downarrow 1$ (where $\kappa \to \infty$) and to $\sin^2\theta_0 > 0$ as $C \to \infty$. Hence the ray meets U_C for $1 < C < \mathrm{GK}$ and misses $\overline{U_C}$ for $C > \mathrm{GK}$. (The foot of the perpendicular, $-\kappa\cos\theta_0\, e^{i\theta_0}$, lies on the ray, and $0 \notin U_C$.) At tangency the contact point is the foot of the perpendicular, $\operatorname{Re}\zeta_* = -r_*^2/\kappa$, so $\cos\theta_0 = -r_*/\kappa$. The case $-\theta_0$ is the mirror image. □

For every $\lambda \in [-1, 1)$ we define K by the closed form (4.1) below, with $C = \mathrm{GK}_p(\lambda)$; for $0 < \lambda < 1$ this is the constant of the introduction. It is determined by the slope of the obstacle at the contact angle: with $y_* = \log r_*$ we show that $K = -q^{-1}\partial_t\Gamma_C(y_*, \theta_0) = -m_C'(\theta_0)/q$. Uniqueness of the minimizing y_* (Lemma 4.1) and locally uniform coercivity (Section 2.2) imply that minimizers y_t of $\Gamma_C(\cdot, t)$ tend to y_* as $t \to \theta_0$. Minimality gives

$$\Gamma_C(y_t, t) - \Gamma_C(y_t, \theta_0) \leq m_C(t) - m_C(\theta_0) \leq \Gamma_C(y_*, t) - \Gamma_C(y_*, \theta_0).$$

Dividing by $t - \theta_0$ (reversing inequalities for negative increments) and using continuity of $\partial_t\Gamma_C$ gives $m_C'(\theta_0) = \partial_t\Gamma_C(y_*, \theta_0)$. It remains to show that $-q^{-1}\partial_t\Gamma_C(y_*, \theta_0)$ is given by the closed form

$$K = 2(1 - \lambda)\, C\, s_0^{p-2} c_0^{1-q} (C^2 + \lambda)^{q-1} = \frac{2(1 - \lambda)^p C^p (C^2 + \lambda)^{q-1}}{s_0\, c_0^{q-1} (C^2 - \lambda)^{p-1}}. \tag{4.1}$$

Indeed, $\partial_t\Gamma_C = -2q \sin t\, [C^p (2\cosh y + 2\cos t)^{q-1} - \lambda(e^y + \lambda^2 e^{-y} + 2\lambda\cos t)^{q-1}]$, and at ζ_* the two bases are $|\zeta_* + 1|^2/r_*$ and $|\zeta_* + \lambda|^2/r_* = C^2|\zeta_* + 1|^2/r_*$, so $-q^{-1}\partial_t\Gamma_C(y_*, \theta_0) = 2 s_0\, r_*^{1-q} C^{p-2} (C^2 - \lambda)\, |\zeta_* + 1|^{p-2}$. With $r_* = c_0(C^2 - \lambda)/(C^2 - 1)$ and

$$|\zeta_* + 1|^2 = r_*^2 - 2r_*^2/\kappa + 1 = 1 + \kappa c_0^2(\kappa - 2) = \frac{(1 - \lambda)^2 (C^2 + \lambda)}{(C^2 - 1)(C^2 - \lambda)},$$

by $\kappa^2 c_0^2 = \kappa^2 - R^2 = (C^2-\lambda^2)/(C^2-1)$ (tangency, proof of Lemma 4.1), this is

$$2 s_0 c_0^{1-q} (1-\lambda)^{p-2} C^{p-2} (C^2+\lambda)^{q-1} (C^2-\lambda)^{3-p},$$

and (2.2) turns it into both forms of (4.1). For $\lambda = 0$, $K = 2c_0^{1-q}/s_0$, the constant b_p of the Hollenbeck–Verbitsky inequality [8, (1.12)–(1.13)].

4.2 The one-variable criterion and strict inequality for ratios near 1

Proposition 4.2 (One-variable criterion)**.** *Let $1 < p < 2$ and $\lambda \in [-1,1)$, and put $C = \mathrm{GK}_p(\lambda)$. Then $\|P_+ + \lambda P_-\|_p = C$ if and only if $m_{\mathrm{GK}}(t) \ge K\cos(q|t|)$ for all $|t| \le \pi$, and if and only if*

$$\mathcal{D}_C(z) := C^p|1+z|^p - |1+\lambda z|^p - K \operatorname{Re} z^q \;\ge\; 0 \qquad (z \in \mathbb{C}). \tag{$*$}$$

Proof. The second equivalence is the definition of m_C after $z \mapsto 1/z$: for $\zeta = 1/z$ one has $|\zeta|^p \mathcal{D}_C(1/\zeta) = C^p|\zeta+1|^p - |\zeta+\lambda|^p - K|\zeta|^q\cos(q\arg\zeta)$, so $(*)$ holds if and only if $r^{-q}\big(C^p|1+re^{it}|^p - |\lambda+re^{it}|^p\big) \ge K\cos(qt)$ for all $r > 0$ and $|t| \le \pi$. *"If".* The 2π-periodic extension of $t \mapsto K\cos(q|t|)$, $|t| \le \pi$, is q-subtrigonometric: it solves $\varphi'' + q^2\varphi = 0$ except at the odd multiples of π, and there its derivative jumps from $-Kq\sin(q\pi)$ to $+Kq\sin(q\pi)$, an upward jump. It lies below m_{GK}, so Theorem 2.2 gives $\|P_+ + \lambda P_-\|_p \le \mathrm{GK}$; the reverse inequality is the Gohberg–Krupnik bound. *"Only if".* By Theorem 2.2 there is a 2π-periodic q-subtrigonometric $\varphi \le m_{\mathrm{GK}}$. Since $m_{\mathrm{GK}}(\pm\theta_0) = 0$ (Lemma 4.1) and $\varphi(-\theta_0) + \varphi(\theta_0) \ge 0$ by Lemma 2.3 with $s = \pi/(2q)$, we have $\varphi(\pm\theta_0) = 0$. Hence, on the arch $[-\theta_0, \theta_0]$ of length π/q, $\varphi(t) = A_\varphi\cos(qt) + B_\varphi\sin(qt)$ for some constants A_φ, B_φ [7, Proposition 3.4(2)]; since $\varphi(\theta_0) = 0$ and $q\theta_0 = \pi/2$, we get $B_\varphi = 0$, so $\varphi = A_\varphi\cos(qt)$. By Section 4.1, m_{GK} is differentiable at θ_0 with $m'_{\mathrm{GK}}(\theta_0) = -Kq$. Since $m_{\mathrm{GK}} - \varphi \ge 0$ with equality at θ_0, comparing left derivatives gives $m'_{\mathrm{GK}}(\theta_0) + A_\varphi q \le 0$, so $A_\varphi \le K$. On the right, [7, Lemma 3.1] gives $\varphi'(\theta_0^+) \ge \varphi'(\theta_0^-)$, and $\varphi'(\theta_0^-) = -A_\varphi q$, so $\varphi'(\theta_0^+) \ge -A_\varphi q$; and tangency gives $\varphi'(\theta_0^+) \le m'_{\mathrm{GK}}(\theta_0) = -Kq$; hence $A_\varphi = K$ and $\varphi'(\theta_0^+) = -Kq$. A q-subtrigonometric function lies, on any interval of length at most π/q, above the sinusoid with its value and right derivative at the left end [7, Lemma 3.2]; at θ_0 this sinusoid is $K\cos(qt)$, since $\varphi(\theta_0) = 0$ and $\varphi'(\theta_0^+) = -Kq$, and $[\theta_0, \pi]$ has length $\pi - \pi/p < \pi/q$. So $m_{\mathrm{GK}} \ge \varphi \ge K\cos(qt)$ on $[\theta_0, \pi]$, on the arch $m_{\mathrm{GK}} \ge \varphi = K\cos(qt)$, and the rest follows by evenness. □

Proposition 4.3 (Strict inequality near $\lambda = 1$)**.** *Let $1 < p < 2$, and let A, s_* and Ω be as defined before Theorem 1.1. If $\Omega(p) < 0$, then $\|P_+ + \lambda P_-\|_p > \mathrm{GK}_p(\lambda)$ for all $\lambda < 1$ sufficiently close to 1. Moreover $\Omega < 0$ on $(1, 1.2336025]$. With Proposition 5.14 it follows that the zero p_0 of Ω in $(1, 1.268]$ (Section 1) satisfies $1.2336025 < p_0 < 1.2336027$, and that $\Omega < 0$ on $(1, p_0)$.*

Proof. Put $\delta = 1 - \lambda$. On the positive axis, the map $r \mapsto \rho = (r+\lambda)/(r+1)$ sends $(0,\infty)$ onto $(\lambda, 1)$ and $m_C(0) = \inf_{\lambda<\rho<1}(1-\lambda)^p(C^p - \rho^p)/Q(\rho)^q$, $Q(\rho) = (1-\rho)(\rho-\lambda)$. Equation (2.2) gives the exact relations $s_0(C^2-1) = \delta(C - s_0)$ and $C^2 - \lambda = \delta C/s_0$. Hence, as $\delta \to 0$, $C \to 1$, $(C^p-1)/\delta \to pA$, and the closed form of K becomes $K/\delta = 2C s_0^{p-2} c_0^{1-q}(C^2+\lambda)^{q-1} \to 2^q s_0^{p-2} c_0^{1-q}$. At the point $\rho_* = 1 - s_*\delta$ we have $Q(\rho_*) = s_*(1-s_*)\delta^2$ exactly and $(C^p - \rho_*^p)/\delta \to p(A + s_*)$, so

$$\limsup_{\delta\to 0} \frac{m_{\mathrm{GK}}(0)}{\delta} \le p\,\frac{A + s_*}{(s_*(1-s_*))^q}.$$

If $\Omega(p) < 0$ then $m_{\mathrm{GK}}(0) < K$ for small δ, so $(*)$ fails on the positive real axis and Proposition 4.2 gives $\|P_+ + \lambda P_-\|_p > \mathrm{GK}$. (The value s_* minimizes $s \mapsto (A+s)/(s(1-s))^q$, which tends to $+\infty$ at both ends and has one critical point; any s would do for the upper bound.) It remains to determine the sign of Ω. On $[1.001, 1.2336025]$ interval arithmetic shows that $\Omega < 0$ (Appendix B). On $(1, 1.001]$ we have $p\,(A+s_*)/(s_*(1-s_*))^q \le p\,2^{p-1}/s_0$ (take $s = 1/2$), so $\Omega < 0$ if $p\,2^{q-1} < s_0^{p-1} c_0^{1-q}$. The left side is at most 0.7081. By Jordan's inequality $s_0 \ge 2(p-1)/p$, so $s_0^{p-1} \ge (p-1)^{p-1} \ge 0.001^{0.001} > 0.9931$, since x^x decreases on $(0, 1/e)$. Together with $c_0 \ge \cos(0.001\pi)$, this makes the right side larger than 0.993, which exceeds 0.7081. Hence $\Omega < 0$ on $(1, 1.001]$. □

4.3 Strict inequality at second order and the local threshold

In this subsection $\lambda \in (-1, 1)$; for $p > 2$ the statements below hold with p' in place of p, by duality (Section 1). We write $\zeta = e^{y+it}$ and $w = (\zeta+\lambda)/(\zeta+1)$, and use G, Γ_C and m_C as in Section 2.2, with $C = \mathrm{GK}_p(\lambda) > 1$.

Theorem 4.4 (Strict inequality at second order)**.** *Let $1 < p < 2$ and $0 < \lambda < 1$. If*

$$\frac{pp'}{4}\sin^2(\pi/p) + \frac{4\lambda\cos^4(\pi/p)}{(1+\lambda)^2} > 1, \tag{4.2}$$

then $\|P_+ + \lambda P_-\|_p > \mathrm{GK}_p(\lambda)$. When $T(p) > 0$, condition (4.2) is equivalent to $\lambda > \lambda_{\mathrm{loc}}(p)$, with T and λ_{loc} as in Section 1.

Proof. Since λ is real, $\Gamma_C(y,-t) = \Gamma_C(y,t)$, so m_C is even and $m_C(-\theta_0+s) + m_C(\theta_0+s) = m_C(\theta_0-s) + m_C(\theta_0+s)$; the two angles are $2\pi/p$ apart. By the arch test (3.1) it suffices to find a small $s \neq 0$ with $m_C(\theta_0-s) + m_C(\theta_0+s) < 0$.

On the ray $t = \theta_0$, $G \geq 0$ if and only if $|w| \leq C$, and $C = \mathrm{GK}$ is the maximum of $|w|$ on this ray (Lemma 4.1). With $\ell = \log w$ as a holomorphic function of $u = y + it$, $\ell' = (1-\lambda)\zeta/((\zeta+\lambda)(\zeta+1))$, and $\partial_y \log|w| = \mathrm{Re}\,\ell' = 0$ if and only if $\mathrm{Re}[(\zeta+\lambda)(\zeta+1)/\zeta] = 1 + \lambda - c_0(r + \lambda/r) = 0$, where $\zeta = re^{i\theta_0}$. So the contact point is $\zeta_* = re^{i\theta_0}$ with $r + \lambda/r = \sigma := (1+\lambda)/c_0$, $r > 1$ the larger root (by the uniqueness in Lemma 4.1, $r = r_*$). The sign of $\partial_y \log|w|$ is the sign of $1 + \lambda - c_0(r + \lambda/r)$: it is negative, positive, negative across the two positive roots. Thus the smaller root is a minimum and the larger a maximum. Since $1 + \lambda - \sigma < 0$, the quadratic $r^2 - \sigma r + \lambda$ is negative at $r = 1$, placing 1 between the roots. At the contact point, $\Gamma_C = \partial_y \Gamma_C = 0$ at (y_*, θ_0), $y_* = \log r$.

We show below that $\partial_y^2 \Gamma_C(y_*, \theta_0) > 0$. By the implicit function theorem there is a smooth branch $y_*(t)$ of $\partial_y \Gamma_C = 0$ near θ_0, and $m_C(t) \leq \mu_*(t) := \Gamma_C(y_*(t), t)$ with $\mu_*(\theta_0) = 0$ and $\mu_*'' = \partial_t^2 \Gamma_C - (\partial_t \partial_y \Gamma_C)^2/\partial_y^2 \Gamma_C$. The odd terms cancel, so $m_C(\theta_0 - s) + m_C(\theta_0 + s) \leq \mu_*''(\theta_0) s^2 + O(s^4)$. It therefore suffices to prove that (4.2) implies $\mu_*''(\theta_0) < 0$.

To compute $\mu_*''(\theta_0)$, write $G = \mathcal{P}(e^{pL} - 1)$ with $\mathcal{P} = e^{p\chi}$, $\chi = \log|\zeta + \lambda|$, $L = \log C - \log|w|$. Both χ and L are harmonic in (y,t), so $L_{tt} = -L_{yy}$. At the contact, $L = L_y = 0$, hence $G_t = p\mathcal{P}L_t$, $G_{yy} = p\mathcal{P}L_{yy}$, $G_{ty} = p\mathcal{P}(L_{ty} + p\chi_y L_t)$, $G_{tt} = p\mathcal{P}(L_{tt} + pL_t^2 + 2p\chi_t L_t)$. Since also $G = G_y = 0$ there, $e^{qy}\partial_t^2 \Gamma_C = G_{tt}$, $e^{qy}\partial_y^2 \Gamma_C = G_{yy}$ and $e^{qy}\partial_t \partial_y \Gamma_C = G_{ty} - qG_t$. Therefore

$$\frac{e^{qy_*}\,\mu_*''(\theta_0)}{p\mathcal{P}} = \frac{L_{yy}\,(2p\chi_t L_t + pL_t^2) - L_{yy}^2 - \big(L_{ty} + L_t(p\chi_y - q)\big)^2}{L_{yy}}. \tag{4.3}$$

At ζ_* put $v = r - \lambda/r > 0$. There $(\zeta+\lambda)(\zeta+1)/\zeta = is_0 v$, so $\ell' = -i(1-\lambda)/(s_0 v)$ and $\ell'' = (1-\lambda)(\zeta - \lambda/\zeta)/(s_0 v)^2$ with $\zeta_* - \lambda/\zeta_* = -c_0 v + is_0\sigma$. Since $L = \log C - \mathrm{Re}\,\ell$,

$$L_t = -\frac{1-\lambda}{s_0 v}, \quad L_{yy} = \frac{(1-\lambda)c_0}{s_0^2 v} > 0, \quad L_{ty} = \frac{(1-\lambda)\sigma}{s_0 v^2}, \quad \chi_y = \frac{1 - (\lambda/r)c_0}{D_1}, \quad \chi_t = -\frac{(\lambda/r)s_0}{D_1}, \tag{4.4}$$

with $D_1 = 1 - 2(\lambda/r)c_0 + (\lambda/r)^2$. In particular $\partial_y^2 \Gamma_C = e^{-qy} p\mathcal{P} L_{yy} > 0$, as used above.

With $v = r - \lambda/r$ as above, the contact identities give

$$c_0 = \frac{(1+\lambda)r}{r^2+\lambda}, \qquad s_0^2 = \frac{(r^2-1)(r^2-\lambda^2)}{(r^2+\lambda)^2}, \qquad D_1 = \frac{(r^2-\lambda)(r^2-\lambda^2)}{r^2(r^2+\lambda)}.$$

Substituting (4.4) into (4.3) first yields

$$\begin{aligned}\frac{e^{qy_*}\mu_*''(\theta_0)}{p\mathcal{P}} = {} & \frac{2p\lambda(1-\lambda)}{rvD_1} + \frac{p(1-\lambda)^2}{s_0^2 v^2} - \frac{(1-\lambda)c_0}{s_0^2 v} \\ & - \frac{1-\lambda}{c_0 v}\left(\frac{\sigma}{v} - p\chi_y + \frac{p}{2}\right)^2.\end{aligned}$$

Write $X = r^2$ and $\delta = 1 - \lambda$. The contact identities give

$$\chi_y = \frac{X}{X-\lambda}, \qquad \frac{\sigma}{v} - p\chi_y + \frac{p}{2} = \left(1 - \frac{p}{2}\right)\frac{\sigma}{v}, \qquad \frac{s_0^2 v}{rD_1} = \frac{X-1}{X+\lambda}.$$

Hence

$$2\lambda\frac{s_0^2 v}{rD_1} + \delta = \frac{(1+\lambda)(X-\lambda)}{X+\lambda} = c_0 v, \qquad c_0^2 v^2 = (1+\lambda)^2 - 4\lambda c_0^2.$$

The first identity combines the two positive terms above into $p\delta c_0/(s_0^2 v)$. Put $B_0 = \delta(1+\lambda)^2/(c_0^3 s_0^2 v^3) > 0$. Dividing by B_0 then gives

$$\frac{e^{qy_*}\mu_*''(\theta_0)}{p\mathcal{P}B_0} = (p-1)\frac{c_0^4 v^2}{(1+\lambda)^2} - s_0^2\left(1-\frac{p}{2}\right)^2.$$

The second identity and $c_0^2 + s_0^2 = 1$ now yield

$$\frac{e^{qy_*}\mu_*''(\theta_0)}{p\mathcal{P}} = B_0\left[(p-1)\left(1-\frac{4\lambda c_0^4}{(1+\lambda)^2}\right) - \frac{p^2 s_0^2}{4}\right]. \tag{4.5}$$

Thus $\mu_*''(\theta_0) < 0$ if and only if $(pp'/4)s_0^2 + 4\lambda c_0^4/(1+\lambda)^2 > 1$, which is (4.2). By Lemma 4.6(b), this inequality holds exactly when $\lambda > \lambda_{\mathrm{loc}}(p)$. □

Corollary 4.5 (Nondegenerate contact for $\lambda \le 0$)**.** *Let $1 < p < 2$ and $-1 < \lambda \le 0$, and let μ_* be defined as in the proof of Theorem 4.4. Then $\mu_*''(\theta_0) > 0$.*

Proof. For $\lambda \le 0$ the map $r \mapsto r + \lambda/r$ increases on $(0,\infty)$, so the contact is unique and $r > 1 > |\lambda|$. All factors in (4.5) are nonzero, and (4.5) holds for every $\lambda \in (-1,1)$ (the same computation). Now $4\lambda c_0^4/(1+\lambda)^2 \le 0$, and $(pp'/4)\sin^2(\pi/p) < 1$: with $u = 1/2 - 1/p \in (-1/2, 1/2)\setminus\{0\}$ this reads $\sin^2(\pi u) > 4u^2$, which is Jordan's inequality. So the left side of (4.2) is < 1, and the bracket in (4.5) equals $(p-1)[1-(pp'/4)s_0^2 - 4\lambda c_0^4/(1+\lambda)^2] > 0$. Since $B_0 > 0$, this gives $\mu_*''(\theta_0) > 0$. □

So for $\lambda \le 0$ the second-order mechanism of Theorem 4.4 gives no obstruction, in agreement with Theorem 5.1. The inequality $(pp'/4)\sin^2(\pi/p) < 1$ proved above is used again in the proof of Lemma 4.6(b).

Lemma 4.6 (The local-threshold exponent)**.** *(a) Some $\lambda \in (0,1)$ satisfies (4.2) if and only if $\mathcal{F}(p) := pp' - 4(1+\cos^2(\pi/p)) > 0$, that is, if and only if $p < p_c$ or $p > p_c'$, where $p_c \in [1.213490304, 1.213490305]$ is the only zero of $\mathcal{F}$ in $(1,2)$.*

(b) For $0 < \lambda < 1$ the left side of (4.2) is less than 1, equal to 1 or greater than 1 according as $\lambda < \lambda_{\mathrm{loc}}(p)$, $\lambda = \lambda_{\mathrm{loc}}(p)$ or $\lambda > \lambda_{\mathrm{loc}}(p)$. Moreover $0 < \lambda_{\mathrm{loc}}(p) \le 1$.

Proof. (a) The left side of (4.2) increases in λ, with supremum $(pp'/4)s_0^2 + c_0^4$ at $\lambda \to 1$; this exceeds 1 if and only if $pp's_0^2 > 4(1-c_0^4) = 4s_0^2(1+c_0^2)$, i.e. $\mathcal{F} > 0$. Here $T(p) = s_0^2\mathcal{F}(p)/(4c_0^4)$, so the two functions have the same sign. For the zero set, write $\mathcal{F} = (p-2)^2/(p-1) - 4\sin^2(\pi(p-2)/(2p))$. On $(1, 1+10^{-6}]$, $\mathcal{F} \ge 1/(p-1) - 8 > 0$, since $(p-2)^2/(p-1) = 1/(p-1) + p - 3$. On $[3/2, 2)$, with $\eta = 2-p$ and $x = \pi\eta/(2p) \le \pi/6$, $\mathcal{F}/\eta^2 \le 1/(p-1) - (\pi/p)^2(1-x^2/6)^2 \le 2 - (\pi^2/4)(1-\pi^2/216)^2 < -0.24$. On $[1+10^{-6}, 3/2]$ we use interval arithmetic: $\mathcal{F} > 0$ on $[1+10^{-6}, 1.2134]$, $\mathcal{F} < 0$ on $[1.2136, 3/2]$, and on $[1.2134, 1.2136]$, $\mathcal{F}' < 0$ and $\mathcal{F}$ changes sign within the stated bracket.

(b) Put $x = (1-\lambda)/(1+\lambda)$, so that $4\lambda/(1+\lambda)^2 = 1 - x^2$, and x decreases from 1 to 0 as λ increases from 0 to 1. Since $c_0^4 + s_0^2(1+c_0^2) = 1$, the left side of (4.2) minus 1 equals $c_0^4\,(T(p) - x^2)$. If $T(p) \le 0$, this is negative for every $\lambda \in (0,1)$, and $\lambda_{\mathrm{loc}}(p) = 1$. If $T(p) > 0$, its sign is that of $\sqrt{T(p)} - x$, which is negative, zero or positive according as $\lambda < \lambda_{\mathrm{loc}}(p)$, $\lambda = \lambda_{\mathrm{loc}}(p)$ or $\lambda > \lambda_{\mathrm{loc}}(p)$. Finally, at $x = 1$ the identity gives $c_0^4(T(p) - 1) = (pp'/4)s_0^2 - 1 < 0$, by the proof of Corollary 4.5; so $T(p) < 1$ and $\lambda_{\mathrm{loc}}(p) > 0$. □

Let $h(\lambda, p) := m_{\mathrm{GK}}(0) - K$. The positive-axis test is the condition $h(\lambda,p) \ge 0$, and the proof of Proposition 4.3 estimates h. We have $h(0,p) > 0$: by Theorem 5.1, equality in $(*)$ at $\lambda = 0$ holds only at $z = e^{\pm i\theta_0}/r_*$, and $r^{-q}G(r) \to +\infty$ as $r \to 0$ and as $r \to \infty$ (Section 2.2). By Proposition 4.2, $h(\lambda,p) < 0$ implies $\|P_+ + \lambda P_-\|_p > \mathrm{GK}_p(\lambda)$.

For fixed $1 < p < 2$, the functions $\lambda \mapsto \mathrm{GK}_p(\lambda)$ and $\lambda \mapsto K$ are continuous on $(-1,1)$. The coercivity in Section 2.2 is locally uniform in λ with $C = \mathrm{GK}_p(\lambda) > \max(1, |\lambda|)$, so the infimum defining $m_{\mathrm{GK}}(0)$ can locally be restricted to a common compact interval in y. Joint continuity of $\Gamma_C(y,0)$ then implies that $\lambda \mapsto m_{\mathrm{GK}}(0)$, and hence $\lambda \mapsto h(\lambda,p)$, is continuous. For $1 < p < p_0$ let $(0, \lambda_1(p)]$ be the largest interval of the form $(0,\lambda]$ on which $h(\cdot,p) \ge 0$. It exists, and $\lambda_1(p) < 1$, because $h(\cdot,p)$ is continuous, $h(0,p) > 0$, and $h(\lambda,p) < 0$ for λ close to 1 (proof of Proposition 4.3, since $\Omega(p) < 0$).

Proposition 4.7 (Strict inequality at the local threshold)**.** *For every $1 < p < p_c$, $h(\lambda_{\mathrm{loc}}(p), p) < 0$. Consequently $h(\cdot,p)$ has a smallest zero in $(0, \lambda_{\mathrm{loc}}(p))$ (by continuity, $h(0,p) > 0$ and Lemma 4.6(b)), and $\|P_+ + \lambda P_-\|_p > \mathrm{GK}_p(\lambda)$ also at $\lambda = \lambda_{\mathrm{loc}}(p)$.*

Proof. Put $\delta = 1-\lambda$. On the positive axis, $\rho = \lambda + \vartheta(1-\lambda)$ with $\vartheta \in (0,1)$ gives

$$m_{\mathrm{GK}}(0) \;\le\; \frac{C^p - \rho^p}{(\vartheta(1-\vartheta))^q} \qquad \text{for every } \vartheta \in (0,1), \tag{4.6}$$

since the factor $(1-\lambda)^p$ cancels. We bound the right side of (4.6) minus K, with K in the closed form (4.1), from above at $\lambda = \lambda_{\mathrm{loc}}(p)$, in three ranges of p.

(a) For $p \in [1+10^{-6}, 1.2134]$ we use interval arithmetic on cells in p, with a value of ϑ chosen in each cell. The quantities are written in closed forms in $a = \pi/p'$ that avoid cancellation, for example $d = \sqrt{T(p)}$, written as $d^2 = \sin^2 a\,(pp'/4 - 1 - \cos^2 a)/\cos^4 a$, and $\lambda_{\mathrm{loc}} = (1-d)/(1+d)$. Every cell gives an upper bound < 0; the domains and precision are in Appendix B, and the programs in [5].

(b) For $p \in (1, 1+10^{-6}]$, Lemma A.1 gives $h(\lambda_{\mathrm{loc}}(p), p) < -1.88 < 0$.

(c) Let $p \in [1.2134, p_c)$. As $p \uparrow p_c$, we have $\lambda_{\mathrm{loc}}(p) \to 1$ and $h \to 0$, so we rescale. With $\rho = 1 - s\delta$ ($s \in (0,1)$ fixed), (4.6) gives $h \le \delta\, g(\delta, p, s)$, where

$$g = \frac{Z_1 + s\,\psi_p(s\delta)}{(s(1-s))^q} - \frac{K}{\delta}, \qquad Z_1 = \frac{C^p - 1}{C-1}\cdot\frac{C-1}{\delta}, \qquad \psi_p(x) = \frac{1-(1-x)^p}{x}.$$

Here $(C-1)/\delta = (C-s_0)/(s_0(C+1))$, by $s_0(C^2-1) = \delta(C-s_0)$; the right side is regular at $\delta = 0$. Both removable quotients, $(C^p-1)/(C-1)$ and $\psi_p(s\delta)$, equal $p\,\xi^{p-1}$ for some ξ between their endpoints (mean value theorem), so $\delta = 0$ may lie inside a cell. The quotient K/δ is regular at $\delta = 0$. Interval arithmetic gives $g < 0$ for $(p,\delta) \in [1.2134, 1.2134904] \times [0, 0.03]$ with $s = 1371/5000$ (at $\delta = 0$, $g = p(A+s)/(s(1-s))^q - 2^q s_0^{p-2} c_0^{1-q}$, whose minimum over s is $\Omega(p)$). Interval arithmetic also gives $T(p) \le 0.015^2$ on $[1.2134, 1.2134904]$ (Appendix B); so for $1.2134 \le p < p_c$ we have $0 < T(p)$, $d = \sqrt{T(p)} \le 0.015$ and $1 - \lambda_{\mathrm{loc}}(p) = 2d/(1+d) \le 2d \le 0.03$. Since $p_c < 1.2134904$, this covers $[1.2134, p_c)$. □

5 Sharpness for real ratios

By Proposition 4.2, sharpness for a ratio $\lambda \in [-1,1)$ is equivalent to the inequality $(*)$, and in this section we prove $(*)$ in the cases, with $1 < p < 2$, where Theorems 5.1–5.3 assert sharpness. Sections 5.1–5.4 prove Theorem 5.1 and the case $3 - \sqrt{3} \le p < 2$ of Theorem 5.2(a), Section 5.5 proves Theorem 5.3, and Section 5.6 treats the remaining exponents $p_0 \le p < 3 - \sqrt{3}$; part (b) of Theorem 5.2 is Proposition 4.3.

Theorem 5.1 (Sharpness for nonpositive ratios)**.** *For every $1 < p < \infty$ and every $\lambda \in [-1, 0]$, $\|P_+ + \lambda P_-\|_p = \mathrm{GK}_p(\lambda)$. For $1 < p < 2$ equality in $(*)$ holds only at $z = e^{\pm i\pi/p}/r_*$, with r_* the contact modulus of Lemma 4.1.*

Theorem 5.2 (The exponent range for all real ratios)**.** *Let p_0 and p_0' be as in Theorem 1.1 ($p_0 \approx 1.2336026$, $p_0' \approx 5.2807742$).*

(a) If $p_0 \le p \le p_0'$, then $\|P_+ + \lambda P_-\|_p = \mathrm{GK}_p(\lambda)$ for every $\lambda \in \mathbb{R}$.

(b) If $p < p_0$ or $p > p_0'$, then $\|P_+ + \lambda P_-\|_p > \mathrm{GK}_p(\lambda)$ for every $\lambda < 1$ sufficiently close to 1.

(c) If $p_0 \le p \le p_0'$, $p \ne 2$ and $\alpha \ne 0$, then $\|\alpha P_+ + \beta P_-\|_p = \mathrm{GK}_p(\alpha, \beta)$ if and only if $\beta/\alpha \in \mathbb{R}$. If $\alpha = 0$, the bound is always sharp.

Theorem 5.3 (An exact positive-ratio criterion)**.** *Let $1 < p < 2$ and $0 < \lambda < 1$. Then $\|P_+ + \lambda P_-\|_p = \mathrm{GK}_p(\lambda)$ if and only if $m_{\mathrm{GK}}(0) \ge K$ (the first condition in (1.1)) and $\lambda < \lambda_{\mathrm{loc}}(p)$, where λ_{loc} is the threshold defined in Section 1 (see Theorem 4.4 and Lemma 4.6; so $\lambda_{\mathrm{loc}} = 1$ for $p \ge p_c$).*

5.1 The Möbius transformation and ellipses

Throughout Sections 5.1–5.4 we assume $1 < p < 2$ and $\lambda \in [-1, 1)$, and we put $\mu = -\lambda$ and $C = \mathrm{GK}_p(\lambda)$; K is given by (4.1). Put $\varepsilon = 2q - 1 = p - 1 \in (0,1)$, and, for $\theta \in (0, \pi)$, write $s_1 = \sin((1-q)\theta)$ and $s_2 = \sin(q\theta)$; below, θ will be the argument of Q. As in Section 4, powers are principal; note that $x \mapsto \operatorname{Re} x^q$ is nondecreasing on $\mathbb{R}$, since $\cos(q\pi) < 0$.

For $z \ne -1$ let $w = (1+\lambda z)/(1+z)$, a bijection of $\mathbb{C} \setminus \{-1\}$ onto $\mathbb{C} \setminus \{\lambda\}$ with $\operatorname{Im} w = -(1-\lambda)\operatorname{Im} z/|1+z|^2$. From $w - \lambda = (1-\lambda)/(1+z)$ and $1 - w = (1-\lambda)z/(1+z)$,

$$Q := (w-\lambda)(1-\bar{w}) = \frac{(1-\lambda)^2\,\bar{z}}{|1+z|^2}, \qquad \operatorname{Re} Q^q = \frac{(1-\lambda)^p \operatorname{Re} z^q}{|1+z|^p}, \qquad \mathcal{D}_C(z) = \frac{|1+z|^p}{(1-\lambda)^p}\,\Phi(w), \tag{5.1}$$

with $\Phi(w) := (1-\lambda)^p(C^p - |w|^p) - K\,\mathrm{Re}\,Q(w)^q$, as $|w| = |1+\lambda z|/|1+z|$. For $w = \rho e^{i\phi}$ (the angle ϕ is unrelated to the function φ of Definition 2.1), $Q = -(\rho^2+\lambda) + (1+\lambda)\rho\cos\phi + i(1-\lambda)\rho\sin\phi$ runs over an ellipse E_ρ (center $-(\rho^2+\lambda)$, semi-axes $(1\pm\lambda)\rho$). With $f_\rho(\phi) := \mathrm{Re}\,Q(\rho e^{i\phi})^q$ and $Q(\bar w) = \overline{Q(w)}$,

$$\min_{|w|=\rho}\Phi = (1-\lambda)^p(C^p-\rho^p) - K\max_{0\le\phi\le\pi} f_\rho(\phi). \tag{5.2}$$

For $\phi\in(0,\pi)$, $\mathrm{Im}\,Q > 0$ and f_ρ is smooth; we locate its critical points. Let $\theta := \arg Q \in (0,\pi)$, $\omega := (1-q)\theta$ and $v := e^{i\omega} + \lambda e^{-i\omega} = (1+\lambda)\cos\omega + i(1-\lambda)\sin\omega$. Since $dQ/d\phi = i(w-\lambda\bar w)$, $Q^{q-1} = |Q|^{q-1}e^{-i\omega}$, $z = re^{-i\theta}$ by (5.1), and $w - \lambda\bar w = (1-\lambda)[1+\lambda r^2 + (1+\lambda)\bar z]/|1+z|^2$,

$$f_\rho'(\phi) = -q|Q|^{q-1}\,\mathrm{Im}(\bar v w) = -\frac{q(1-\lambda)|Q|^{q-1}}{|1+z|^2}\,D(\theta,r), \qquad D(\theta,r) := (1+\lambda)\,r\,s_2 - (1+\lambda r^2)\,s_1. \tag{5.3}$$

As v and w lie in the upper half-plane, $f_\rho'(\phi) = 0$ if and only if $w = \tau v$ with $\tau = \rho/|v| > 0$, if and only if $D(\theta,r) = 0$. At such a point, with $|v|^2 = 1+\lambda^2+2\lambda\cos 2\omega$,

$$|Q|\sin\theta = (1-\lambda)\,\mathrm{Im}\,w = (1-\lambda)^2\tau\sin\omega, \quad \rho = \tau|v|, \quad w-\lambda\bar w = (1-\lambda^2)\tau e^{i\omega}, \quad \mathrm{Re}\left[e^{-i\omega}(w+\lambda\bar w)\right] = \tau|v|^2. \tag{5.4}$$

The critical points lie on one curve. Fix $\theta\in(0,\pi)$; then $0 < s_1 < s_2$, as $s_2 - s_1 = 2\cos(\theta/2)\sin(\varepsilon\theta/2)$. For $\lambda > -1$, $D(\theta,0) = -s_1 < 0 < (1+\lambda)(s_2-s_1) = D(\theta,1)$ and $\partial_r D = (1+\lambda)s_2 - 2\lambda r s_1 > 0$ on $[0,1]$ (if $\lambda > 0$ use $s_2 > s_1$), so $D(\theta,\cdot)$ has exactly one zero $R(\theta)\in(0,1)$; for $\lambda = -1$, $D = -(1-r^2)s_1$. In both cases

$$R = \frac{2s_1}{(1+\lambda)s_2 + \sqrt{(1+\lambda)^2 s_2^2 - 4\lambda s_1^2}},$$

with discriminant $\ge (1-\lambda)^2 s_1^2 > 0$; so R is real-analytic, $R(\pi^-) = 1$, and $r_0 := R(0^+)\in(0,1]$. For $\lambda = 0$, $D(\theta,\cdot)$ is linear. For $\lambda\ne 0$ its zeros have product $1/\lambda$: for $\lambda < 0$ the other zero is negative; for $\lambda > 0$ the other one, $1/(\lambda R)$, gives the image of $Re^{-i\theta}$ under the inversion $z\mapsto 1/(\lambda\bar z)$ of Lemma 5.6. The *critical curve* is $z_\theta := R(\theta)e^{-i\theta}$, $0<\theta<\pi$; put $w(\theta) := w(z_\theta)$ and $\rho(\theta) := |w(\theta)|$. So the critical points of f_ρ in $(0,\pi)$ are the points $w(\theta)$ with $\rho(\theta) = \rho$ and, for $\lambda > 0$, their inversion images $\lambda/\overline{w(\theta)}$ with $\lambda/\rho(\theta) = \rho$. On the critical curve, with $r = R(\theta)$, $A_\theta := |1+z_\theta|^2 = 1 + 2r\cos\theta + r^2$ and $B := |1+\lambda z_\theta|^2$,

$$|Q| = \frac{(1-\lambda)^2 r}{A_\theta}, \quad \rho^2 = \frac{B}{A_\theta}, \quad \tau = \frac{r\sin\theta}{A_\theta s_1}, \quad 2\tau\cos\omega - 1 = \frac{Y}{A_\theta s_1}, \qquad Y := 2rs_2 - (1+r^2)s_1 = (1-\lambda)\,r\,(s_2 - rs_1) > 0. \tag{5.5}$$

The third identity follows from (5.4); the fourth from $2\sin\theta\cos((1-q)\theta) = \sin((2-q)\theta) + s_2$ and $2\cos\theta\, s_1 = \sin((2-q)\theta) - s_2$; the last from $D = (rs_2 - s_1) + \lambda r(s_2 - rs_1) = 0$ and $r\le 1$.

Lemma 5.4 (Strict maxima on the ellipses)**.** *At every critical point of f_ρ in $(0,\pi)$,*

$$f_\rho'' = q|Q|^{q-2}\tau^2(1-\lambda)^2\Big[(1-q)(1+\lambda)^2\cos(q\theta) - |v|^2\,\frac{\sin\omega}{\sin\theta}\Big] < 0. \tag{5.6}$$

Proof. By (5.4), $Q^{q-1}Q' = i(1-\lambda^2)\tau|Q|^{q-1}$ is purely imaginary, and $Q'' = -(w+\lambda\bar w)$. So $f_\rho'' = \mathrm{Re}[q(q-1)(Q^{q-1}Q')^2/Q^q + qQ^{q-1}Q''] = q(1-q)|Q|^{q-2}\tau^2(1-\lambda^2)^2\cos(q\theta) - q|Q|^{q-1}\tau|v|^2$, and $|Q| = (1-\lambda)^2\tau\sin\omega/\sin\theta$ gives the formula. If $\theta\ge\theta_0$, then $q\theta\in[\pi/2,\pi)$ and the bracket is negative. Let $\theta<\theta_0$. *Case* $\lambda\le 0$. Here $|v|^2 \ge (1+\lambda)^2$, and $\sin\omega - (1-q)\cos(q\theta)\sin\theta = q\sin\theta\cos(q\theta) - \cos\theta\sin(q\theta) > 0$: for $\theta\ge\pi/2$ both terms are ≥ 0, the first > 0; for $\theta<\pi/2$ it says $\tan(q\theta)/(q\theta) < \tan\theta/\theta$, which holds because $y\mapsto\tan y/y$ increases on $(0,\pi/2)$. So $(1-q)(1+\lambda)^2\cos(q\theta)\sin\theta \le (1+\lambda)^2\sin\omega \le |v|^2\sin\omega$, the first inequality strict if $\lambda > -1$, the second if $\lambda = -1$. *Case* $\lambda > 0$. As $|v|^2 > (1+\lambda)^2\cos^2\omega$, it suffices that $g := \cos^2\omega\sin\omega - (1-q)\cos(q\theta)\sin\theta \ge 0$. By $\cos^2\omega\sin\omega = (\sin\omega + \sin 3\omega)/4$, $2\sin\theta\cos(q\theta) = \sin((1+q)\theta) + \sin\omega$ and $\sin 3\omega - \sin((1+q)\theta) = -2\cos((2-q)\theta)\sin((2q-1)\theta)$, we get $4g = J(\theta) := 2(2q-1)\sin\theta\cos(q\theta) - 2\cos((2-q)\theta)\sin((2q-1)\theta)$. On $(0,\theta_0)$, $\sin\theta$, $\cos(q\theta)$, $\sin((2q-1)\theta) > 0$. If $\theta\ge\pi/(2(2-q))$, then $(2-q)\theta\in[\pi/2, 3\pi/2)$ (as $(2-q)\theta_0 = \pi/q - \pi/2$): both terms of J are ≥ 0, the first > 0. If $\theta < \pi/(2(2-q))$, then $J(0) = 0$ and $J' = (1-q)[2(2q-1)\sin\theta\sin(q\theta) + 6\sin((2-q)\theta)\sin((2q-1)\theta)] > 0$, all angles lying in $(0,\pi)$. So $J > 0$. □

We record three consequences of Lemma 5.4. (E1) For each $\rho > 0$, f_ρ has at most one critical point in $(0,\pi)$: all are strict local maxima, and between two of them there would be an interior minimum. (E2) $\rho(\theta)$ is continuous, injective by (E1), and $\rho \to \infty$ as $\theta \to \pi$ ($z_\theta \to -1$); so ρ increases strictly from $\rho_0 := \rho(0^+) = (1+\lambda r_0)/(1+r_0) \in [0,1)$ to ∞, and $\rho' > 0$ since θ is a smooth function of ρ (implicit function theorem, $f_\rho'' \neq 0$) and $\rho'(\theta)\,\theta'(\rho) = 1$. If $\lambda > 0$, $r_0 \le 1$ gives $\rho_0 \ge (1+\lambda)/2 > \sqrt{\lambda}$, so the inversion images lie on circles $\rho = \lambda/\rho(\theta) < \sqrt{\lambda} < \rho_0$: the critical curve and its inversion images never share a circle. (E3) If $\rho > \rho_0$, f_ρ has exactly one critical point $\phi_c \in (0,\pi)$, at $w(\theta)$ with $\rho(\theta) = \rho$; $f_\rho' > 0$ on $(0,\phi_c)$, $f_\rho' < 0$ on (ϕ_c,π), and $\max_{[0,\pi]} f_\rho$ is attained only at ϕ_c. If $\rho \le \rho_0$ (and $\rho \ge \sqrt{\lambda}$ when $\lambda > 0$), f_ρ is monotone on $[0,\pi]$ (by (E2) no critical point of f_ρ lies on such a circle: the critical curve has $\rho(\theta) > \rho_0$ and its inversion images have $\rho < \sqrt{\lambda}$), and $Q(\rho) - Q(-\rho) = 2(1+\lambda)\rho \ge 0$ (both real) gives $\max f_\rho = f_\rho(0)$, at $w = \rho$.

5.2 The function K_f

On the critical curve put $\Lambda(\theta) := \Phi(w(\theta))/(1-\lambda)^p = C^p - \rho(\theta)^p - K\,\mathrm{Re}\, z_\theta^q/|1+z_\theta|^p = \mathcal{D}_C(z_\theta)/|1+z_\theta|^p$. In the upper half-plane, $\partial_{\bar w} Q = \lambda - w$ and $\partial_w Q = 1 - \bar w$ give $(\partial_x + i\partial_y)\,\mathrm{Re}\, Q^q = q[Q^{q-1}(\lambda - w) + \overline{Q^{q-1}}(1-w)] = q|Q|^{q-1}(v - 2w\cos\omega)$, and the gradient of $|w|^p$ is $p|w|^{p-2}w$. On the critical curve $w = \tau v$, so $\nabla\Phi(w(\theta)) = q|Q|^{q-1}(2\tau\cos\omega - 1)\,(K - K_f(\theta))\,v$ with

$$K_f(\theta) := \frac{2(1-\lambda)^p \rho^{p-2}\,\tau\,|Q|^{1-q}}{2\tau\cos\omega - 1} = \frac{2(1-\lambda)^2 \sin\theta\; r^{2-q} B^{q-1}}{Y} \qquad (r = R(\theta)), \tag{5.7}$$

the second form by (5.5); K_f contains neither C nor K. By (5.8) below, $K_f(\theta)$ is the value of K at which Λ is stationary at θ. Since $\mathrm{Re}(\bar v w') = (\rho^2)'/(2\tau)$,

$$\frac{d}{d\theta}\Phi(w(\theta)) = \mathrm{Re}\,\big(\overline{\nabla\Phi}\,w'\big) = q|Q|^{q-1}(2\tau\cos\omega - 1)\,\frac{\rho\rho'}{\tau}\,\big(K - K_f(\theta)\big), \tag{5.8}$$

and by (5.5) and (E2), $\Lambda'(\theta) = c(\theta)\,(K - K_f(\theta))$ with $c(\theta) > 0$.

Lemma 5.5 (The contact point)**.** *We have $\rho(\theta_0) = C$, $\Lambda(\theta_0) = 0$ and $K_f(\theta_0) = K$. Moreover $z_* := z_{\theta_0} = 1/\zeta_*$ with $\zeta_* = r_* e^{i\theta_0}$ (Lemma 4.1), so $|z_*| = 1/r_*$.*

Proof. At $\theta = \theta_0$ we have $\sin\omega = c_0$ and $\cos\omega = s_0$. As $\theta_0 = \arg Q$, $s_0\,\mathrm{Re}\, Q = -c_0\,\mathrm{Im}\, Q$; with $Q = -(\rho^2+\lambda) + (1+\lambda)\,\mathrm{Re}\, w + i(1-\lambda)\,\mathrm{Im}\, w$ and $w = \tau v$ this is $(\rho^2+\lambda)s_0 = \tau|v|^2 = \rho|v|$, where $|v|^2 = (1+\lambda)^2 s_0^2 + (1-\lambda)^2 c_0^2$. Squaring, $s_0^2(\rho^2-1)(\rho^2-\lambda^2) = (1-\lambda)^2 c_0^2 \rho^2$: this is (2.1) for the ratio λ, with $X = \rho^2$; its roots are C^2 and λ^2/C^2. The second is excluded: for $\lambda \le 0$ it gives $\rho^2 + \lambda = \mu(\mu/C^2 - 1) \le 0 < \rho|v|/s_0$, with equality on the left only for $\lambda = 0$, where the second root $\rho = 0$ is excluded since $\rho(\theta_0) > \rho_0 > 0$ by (E2); for $\lambda > 0$, $\rho(\theta_0) > \rho_0 > \sqrt{\lambda} > \lambda/C$ by (E2). So $\rho(\theta_0) = C$, and $\Lambda(\theta_0) = 0$ as $\mathrm{Re}\, z_*^q = |z_*|^q \cos(q\theta_0) = 0$. Next, $\tau = \rho^2/((\rho^2+\lambda)s_0) = C^2/((C^2+\lambda)s_0)$, $|Q| = (1-\lambda)^2\tau c_0/s_0$ by (5.4), and $2\tau s_0 - 1 = (C^2-\lambda)/(C^2+\lambda)$; in (5.7), with the first form of (4.1), $K_f(\theta_0)/K = (1-\lambda)C/(s_0(C^2-\lambda)) = 1$ by (2.2). Finally, $|\zeta|^p \mathcal{D}_C(1/\zeta) = C^p|\zeta+1|^p - |\zeta+\lambda|^p$ on $\arg\zeta = \theta_0$; it vanishes at $1/z_*$, so Lemma 4.1 gives $1/z_* = \zeta_*$. □

Lemma 5.6 (Inversion)**.** *Let $0 < \lambda < 1$, $z \neq 0$ and $z' := 1/(\lambda\bar z)$. Then*

$$\mathcal{D}_C(z) - \lambda^q |z|^p \mathcal{D}_C(z') = (C^p + \lambda^q)\big(|1+z|^p - \lambda^{-q}|1+\lambda z|^p\big),$$

which is positive exactly when $|z| > 1/\sqrt{\lambda}$, i.e. when $|w| < \sqrt{\lambda}$. So $\mathcal{D}_C \ge 0$ on $\{|z| \le 1/\sqrt{\lambda}\}$ implies $\mathcal{D}_C > 0$ on $\{|z| > 1/\sqrt{\lambda}\}$. In the w-plane, $z \mapsto z'$ is $w \mapsto \lambda/\bar w$.

Proof. Insert $|1+z'| = |1+\lambda z|/(\lambda|z|)$, $|1+\lambda z'| = |1+z|/|z|$ and $\mathrm{Re}\, z'^q = \mathrm{Re}\, z^q/(\lambda^q|z|^p)$ ($\arg z' = \arg z$, $|z'| = 1/(\lambda|z|)$) in $\mathcal{D}_C(z')$. The sign follows from $|1+\lambda z|^2 - \lambda|1+z|^2 = (1-\lambda)(1-\lambda|z|^2)$; if $|z| > 1/\sqrt{\lambda}$, then $|z'| < 1/\sqrt{\lambda}$ and $\mathcal{D}_C(z) > \lambda^q|z|^p\mathcal{D}_C(z') \ge 0$. Finally $w(z') = \lambda(1+\bar z)/(1+\lambda\bar z)$. □

5.3 Monotonicity of K_f

In this subsection we assume $-1 < \lambda < 1$; the endpoint $\lambda = -1$ is treated directly in the proof of Theorem 5.1. Fix $\theta \in (0,\pi)$ and put the angles $a = \theta/2$, $b = \varepsilon a$ (so $a + b = q\theta$, $a - b = (1-q)\theta$). On the critical curve the radius $r = R(\theta) \in (0,1)$ is written as $r = \sin(\gamma - b)/\sin(\gamma + b)$ with $\gamma \in (b, \pi/2)$, and $X := \cos 2\gamma$. Write $D =$

$(rs_2-s_1)+\lambda r(s_2-rs_1)$. With $r=\sin(\gamma-b)/\sin(\gamma+b)$, product-to-sum gives $s_1-rs_2=\sin 2b\,\sin(a-\gamma)/\sin(\gamma+b)$ and $s_2-rs_1=\sin 2b\,\sin(\gamma+a)/\sin(\gamma+b)$, so $D=0$ becomes

$$\mu=-\lambda=\frac{\sin(\gamma-a)\sin(\gamma+b)}{\sin(\gamma+a)\sin(\gamma-b)}. \tag{5.9}$$

Hence $\lambda<0$ for $\gamma>a$, $\lambda=0$ at $\gamma=a$, and $\lambda>0$ for $\gamma<a$. The identities

$$\lambda r=-\frac{\sin(\gamma-a)}{\sin(\gamma+a)},\qquad s_2-rs_1=\frac{\sin(\gamma+a)\sin 2b}{\sin(\gamma+b)},$$
$$|1+\lambda z_\theta|=\frac{\sin 2a}{\sin(\gamma+a)}$$

follow from (5.9), the formula for r and the sine addition formulas; the last identity follows by squaring, since both sides are positive. Inserting them into the second form of (5.7) for K_f gives $K_f=2(1+\mu)\exp L(a,\gamma)$ with $L=(1-q)\log\sin(\gamma-b)+q\log\sin(\gamma+b)+\varepsilon\log\sin 2a-\varepsilon\log\sin(\gamma+a)-\log\sin 2b$. With $u_1=\theta$, $u_2=\varepsilon\theta$, $u_3=(1-\varepsilon)\theta$ let

$$\begin{aligned}\Pi(X)=\ &-(1-\varepsilon)^2\sin u_1\sin u_2\,X^2+2\varepsilon\sin u_3(\sin u_1-\sin u_2)\,X\\ &+(1+\varepsilon^2)\sin u_1\sin u_2-2\varepsilon\sin u_1\sin u_2\cos u_3-2\varepsilon\sin^2 u_3.\end{aligned}$$

Lemma 5.7 (Sign formula)**.** *Along the level set $\lambda=\text{const}$ of (5.9), for every $\lambda\in(-1,1)$,*

$$\frac{d\log K_f}{da}=\frac{\Pi(X)}{2\sin\theta\,\sin(\varepsilon\theta)\,\sin(a-b)\,[\cos(a-b)-X\cos(a+b)]},$$

and the denominator is positive. Hence K_f increases exactly where $\Pi(X)>0$.

Proof. First suppose $\lambda\neq 0$. Let $M=\log|\sin(\gamma-a)|+\log\sin(\gamma+b)-\log\sin(\gamma+a)-\log\sin(\gamma-b)$, so the level set is $M=\log|\mu|$ and, by implicit differentiation along it, $d\log K_f/da=\mathcal{E}/M_\gamma$ with $\mathcal{E}=L_aM_\gamma-L_\gamma M_a$. With p fixed and $b=\varepsilon a$, differentiation gives

$$\begin{aligned}L_a&=\varepsilon\big[q\cot(\gamma+b)-(1-q)\cot(\gamma-b)-\cot(\gamma+a)+2\cot 2a-2\cot 2b\big],\\ L_\gamma&=(1-q)\cot(\gamma-b)+q\cot(\gamma+b)-\varepsilon\cot(\gamma+a),\\ M_a&=-\cot(\gamma-a)-\cot(\gamma+a)+\varepsilon\big[\cot(\gamma+b)+\cot(\gamma-b)\big],\\ M_\gamma&=\cot(\gamma-a)+\cot(\gamma+b)-\cot(\gamma+a)-\cot(\gamma-b).\end{aligned}$$

To clear the cotangents, put

$$\begin{aligned}U&=\sin(\gamma-a)\sin(\gamma+a)=\frac{\cos 2a-X}{2},\\ V&=\sin(\gamma-b)\sin(\gamma+b)=\frac{\cos 2b-X}{2},\\ N&=\sin 2a\,V-\sin 2b\,U,\qquad S=UV.\end{aligned}$$

Cotangent sums and differences give

$$\begin{aligned}SM_\gamma&=N,&&&SM_a&=(\varepsilon U-V)\sin 2\gamma,\\ SL_\gamma&=\tfrac12[(U-\varepsilon V)\sin 2\gamma+\varepsilon N],\\ SL_a&=\tfrac{\varepsilon}{2}[(\varepsilon U-V)\sin 2\gamma+N]+2\varepsilon UV(\cot 2a-\cot 2b).\end{aligned}$$

In $S^2\mathcal{E}=(SL_a)(SM_\gamma)-(SL_\gamma)(SM_a)$ the terms $\pm\frac{\varepsilon}{2}(\varepsilon U-V)N\sin 2\gamma$ cancel. Since $\sin^2 2\gamma=1-X^2$,

$$\mathcal{E}S=\frac{\varepsilon N^2-(U-\varepsilon V)(\varepsilon U-V)(1-X^2)}{2UV}+2\varepsilon(\cot 2a-\cot 2b)N.$$

Write the first term as $\Pi_1(X)/2$, where $\Pi_1:=[\varepsilon N^2-(U-\varepsilon V)(\varepsilon U-V)(1-X^2)]/(UV)$. Its numerator vanishes where $U=0$ and where $V=0$: there $1-X^2$ equals $\sin^2 2a$, respectively $\sin^2 2b$, and N equals $V\sin 2a$, respectively $-U\sin 2b$. So Π_1 is a quadratic polynomial in X. Its leading coefficient is $-(1-\varepsilon)^2$, and at $X=\pm1$ one finds $\Pi_1(\pm1)=4\varepsilon\sin^2(a-b)=2\varepsilon(1-\cos u_3)$. Hence $\Pi_1(X)=-(1-\varepsilon)^2X^2+1+\varepsilon^2-2\varepsilon\cos u_3$. Since

$\sin 2a \sin 2b(\cot 2a - \cot 2b) = -\sin u_3$ and $2N = \sin u_3 - (\sin u_1 - \sin u_2)X$, this gives $2\sin u_1 \sin u_2\, \mathcal{E}S = \Pi(X)$. From $\cot x - \cot y = \sin(y-x)/(\sin x \sin y)$, $M_\gamma = \sin(a-b)\,[\sin(\gamma+a)\sin(\gamma+b) + \sin(\gamma-a)\sin(\gamma-b)]/S = \sin(a-b)[\cos(a-b) - X\cos(a+b)]/S$. The factor S cancels. Finally, $\cos(a-b) + \cos(a+b) = 2\cos a \cos b$ and $\cos(a-b) - \cos(a+b) = 2\sin a \sin b$ are positive, so $\cos(a-b) - X\cos(a+b) > 0$ for $|X| \le 1$. For $\lambda = 0$ we pass to the limit in the final identity. The explicit root $R(\theta)$ in Section 5.1 is real-analytic in λ near zero, and K_f is positive and real-analytic there for each $0 < \theta < \pi$. The coordinate γ also varies smoothly: the derivative of $\sin(\gamma-b)/\sin(\gamma+b)$ with respect to γ is $\sin(2b)/\sin^2(\gamma+b) > 0$. Thus both sides of the formula in Lemma 5.7 extend continuously to $\lambda = 0$, where its denominator remains strictly positive. □

As in Section 5.1, let $\omega = (1-q)\theta$. Then $\omega \in (0, \pi/(m+1))$, where $m := q/(1-q) = p/(2-p) > 1$, so that $q\theta = m\omega$. Put

$$\Psi_m(\omega) := \frac{m^2 \sin^2\omega - \sin^2(m\omega)}{\sin^4\omega}, \qquad W(\theta) := (1-\varepsilon)^2 \sin\theta \sin(\varepsilon\theta) - \varepsilon \sin^2((1-\varepsilon)\theta).$$

Lemma 5.8 (One-variable form)**.** *On the critical curve at parameter $\lambda \in (-1,1)$, $K_f'(\theta)$ has the sign of $4(m^2-1)\lambda/(1+\lambda)^2 - \Psi_m(\omega)$. In particular, K_f increases at θ if and only if*

$$\Psi_m(\omega) \;<\; \frac{4(m^2-1)\,\lambda}{(1+\lambda)^2}.$$

Proof. Put $c_q = \cos(q\theta)$, $c_\omega = \cos\omega > 0$, $W_0 = (1-\varepsilon)^2 \sin\theta \sin(\varepsilon\theta)$. Using $c_\omega^2 - c_q^2 = \sin\theta\sin(\varepsilon\theta)$, $\sin u_3\, c_q/c_\omega = \sin u_1 - \sin u_2$ and, for the constant term, $\sin^2 u_3 = \sin^2 u_1 + \sin^2 u_2 - 2\sin u_1 \sin u_2 \cos u_3$, one checks, by comparing the coefficients of X^2, X and 1, the identity

$$\Pi(X) = \frac{W}{c_\omega^2}\,(c_q^2 + c_\omega^2 - 2c_q c_\omega X) - W_0\Big(X - \frac{c_q}{c_\omega}\Big)^2. \tag{5.10}$$

Let $\sigma = \sqrt{1-X^2}\,\sin\omega$ and $\nu = c_q - Xc_\omega$. Since $\cos(2\gamma \mp \omega) = Xc_\omega \pm \sigma$, (5.9) reads $\mu = (\nu-\sigma)/(\nu+\sigma)$, i.e. $\nu = ((1-\lambda)/(1+\lambda))\,\sigma$; here $\nu + \sigma = 2\sin(\gamma+a)\sin(\gamma-b) > 0$. Then $c_q^2 + c_\omega^2 - 2c_q c_\omega X = \nu^2 + \sigma^2\cot^2\omega$, $c_q/c_\omega - X = \nu/c_\omega$ and $W_0 - W = \varepsilon\sin^2 2\omega$, so (5.10) becomes

$$\Pi(X) = \frac{\sigma^2\cot^2\omega}{c_\omega^2}\,\Big[\,W - 4\varepsilon\Big(\frac{1-\lambda}{1+\lambda}\Big)^2 \sin^4\omega\,\Big].$$

With $\theta = (m+1)\omega$, $\varepsilon\theta = (m-1)\omega$, $(1-\varepsilon)\theta = 2\omega$ and $\sin((m+1)\omega)\sin((m-1)\omega) = \sin^2(m\omega) - \sin^2\omega$ one finds $W/(4\varepsilon\sin^4\omega) = 1 - \Psi_m(\omega)/(m^2-1)$. Since $1 - (1-\lambda)^2/(1+\lambda)^2 = 4\lambda/(1+\lambda)^2$, the sign of Π is the sign of $4(m^2-1)\lambda/(1+\lambda)^2 - \Psi_m(\omega)$, and the lemma follows from Lemma 5.7. □

Lemma 5.9 (Monotonicity for $\lambda \le 0$)**.** *For every $1 < p < 2$ and $\lambda \in (-1, 0]$, K_f is strictly decreasing on $(0,\pi)$.*

Proof. The right side in Lemma 5.8 is ≤ 0, while $\Psi_m(\omega) > 0$: since $\sin x/x$ decreases on $(0,\pi)$ and $\omega < m\omega < \pi - \omega$, we have $0 < \sin(m\omega) < m\sin\omega$ (the first inequality as $m\omega < \pi$). □

Lemma 5.10 (Monotonicity for $p \ge 3 - \sqrt{3}$)**.** *If $3 - \sqrt{3} \le p < 2$, then K_f is strictly decreasing on $(0,\pi)$ for every $\lambda \in (-1,1)$.*

Proof. By Lemma 5.8 and $4\lambda/(1+\lambda)^2 < 1$ it suffices to show that $\Psi_m \ge m^2 - 1$, that is, $W \le 0$ on $(0,\pi)$. All of θ, $\varepsilon\theta$, $(1-\varepsilon)\theta$ lie in $(0,\pi)$; by $\sin y = y\prod_n (1 - y^2/(n\pi)^2)$ and $\log(1-u) = -\sum_k u^k/k$, with $\xi_n = \theta/(n\pi) \in (0,1)$,

$$\log\frac{\varepsilon\sin^2((1-\varepsilon)\theta)}{(1-\varepsilon)^2\sin\theta\sin(\varepsilon\theta)} = \sum_{n\ge1}\sum_{k\ge1}\frac{\xi_n^{2k}}{k}\,\big[1 + \varepsilon^{2k} - 2(1-\varepsilon)^{2k}\big]$$

(the prefactors $\varepsilon(1-\varepsilon)^2\theta^2$ cancel). The $k=1$ bracket is $-(\varepsilon^2 - 4\varepsilon + 1) \ge 0$ for $\varepsilon \ge 2 - \sqrt{3}$; for $k \ge 2$, $2(1-\varepsilon)^{2k} \le 2(4 - 2\sqrt{3})^2 < 0.58$. So the logarithm is positive and $W < 0$. □

Lemma 5.11 (A single turning point for $p < 4/3$)**.** *For $1 < m < 2$ (i.e. $1 < p < 4/3$), Ψ_m is strictly increasing on $(0, \pi/2)$. Consequently, for every $\lambda \in (0,1)$, K_f is either strictly decreasing on $(0,\pi)$, or, for some $\theta_\lambda \in (0,\pi)$, strictly increasing on $(0,\theta_\lambda]$ and strictly decreasing on $[\theta_\lambda, \pi)$.*

Proof. Let $t = \sin^2\omega$ and $Y_m(t) = \cos(2m\omega)$. Since $(dt/d\omega)^2 = 4t(1-t)$ and $d^2t/d\omega^2 = 2-4t$, the equation $d^2Y_m/d\omega^2 + 4m^2Y_m = 0$ becomes

$$t(1-t)Y_m'' + (\tfrac{1}{2} - t)Y_m' + m^2Y_m = 0.$$

Since $\cos(2m\omega)$ is even in ω, it is an analytic function of ω^2, and so of $t = \sin^2\omega$, near 0. Writing $Y_m = \sum_{n\geq 0} y_n t^n$ gives $y_0 = 1$ and

$$y_{n+1} = \frac{n^2 - m^2}{(n+1)(n+\frac{1}{2})} y_n.$$

The resulting series has radius of convergence at least 1; both $\sum_n y_n \sin^{2n}\omega$ and $\cos(2m\omega)$ are analytic for $|\omega| < \pi/2$ and agree near 0, so they agree on $(-\pi/2, \pi/2)$. This gives

$$\sin^2(m\omega) = \tfrac{1}{2}(1 - Y_m(t)) = \sum_{n\geq 1} \frac{2^{2n-1} m^2 \prod_{j=1}^{n-1}(j^2 - m^2)}{(2n)!} t^n.$$

Consequently

$$\Psi_m(\omega) = \frac{m^2 t - \sin^2(m\omega)}{t^2} = \sum_{n\geq 0} \beta_n t^n, \qquad \beta_n = \frac{2^{2n+3}\, m^2(m^2-1)\prod_{j=2}^{n+1}(j^2 - m^2)}{(2n+4)!}.$$

For $1 < m < 2$ every factor is positive, so Ψ_m is a power series in t with positive coefficients, and t increases on $(0, \pi/2) \supset (0, \pi/(m+1))$. By Lemma 5.8, the set on which K_f increases is an initial interval. At $\omega_e = \pi/(m+1)$, $\sin(m\omega_e) = \sin\omega_e$, so

$$\Psi_m(\omega_e) = \frac{m^2 - 1}{\sin^2\omega_e} > m^2 - 1 > \frac{4(m^2-1)\lambda}{(1+\lambda)^2}.$$

Thus K_f decreases near $\theta = \pi$, which gives the dichotomy. □

5.4 Proofs of Theorems 5.1 and 5.2

Proposition 5.12 (Reduction)**.** *Let $1 < p < 2$ and $\lambda \in [-1, 1)$. If K_f is strictly decreasing on $(0, \pi)$, then $\mathcal{D}_C \geq 0$ on $\mathbb{C}$, with equality only at z_* and $\bar{z}_*$.*

Proof. As $K_f(\theta_0) = K$, by (5.8) the function Λ decreases strictly on $(0, \theta_0]$ and increases strictly on $[\theta_0, \pi)$. So $\Lambda(\theta) > \Lambda(\theta_0) = 0$ for $\theta \neq \theta_0$, the limit $\Lambda(0^+)$ exists and is positive, and $\Phi(\rho_0) = (1-\lambda)^p\Lambda(0^+) > 0$, since $w(\theta) \to \rho_0$ and Φ is continuous.

Let $I := [0, \rho_0]$ if $\lambda \leq 0$ and $I := [\sqrt{\lambda}, \rho_0]$ if $\lambda > 0$. In the interior of I, $Q_{\mathbb{R}} := Q(\rho) = (\rho - \lambda)(1-\rho) > 0$ and

$$\Phi'(\rho) = -(1-\lambda)^p p\rho^{p-1} - Kq Q_{\mathbb{R}}^{q-1}(1 + \lambda - 2\rho) = qQ_{\mathbb{R}}^{q-1}(2\rho - 1 - \lambda)\big[K - K_{\mathbb{R}}(\rho)\big], \quad K_{\mathbb{R}} := \frac{2(1-\lambda)^p \rho^{p-1} Q_{\mathbb{R}}^{1-q}}{2\rho - 1 - \lambda}.$$

For $\rho \leq \rho_1 := (1+\lambda)/2$ both terms of the middle expression are ≤ 0, the first < 0. On $(\rho_1, 1)$, $(\log K_{\mathbb{R}})' = [\varepsilon/\rho - 2/(2\rho - 1 - \lambda)] - (1-q)(2\rho - 1 - \lambda)/Q_{\mathbb{R}} < 0$, since $\varepsilon(2\rho - 1 - \lambda) < 2\rho$. Suppose $\rho_0 > \rho_1$. As $\theta \to 0^+$, we have $w(\theta) \to \rho_0$, $\operatorname{Re} v > 0$ and $K - K_f(\theta) < 0$, so $\operatorname{Re}\nabla\Phi(w(\theta)) < 0$. Since Φ is C^1 near ρ_0 (because $Q_{\mathbb{R}}(\rho_0) > 0$), it follows that $\Phi'(\rho_0) \leq 0$, i.e. $K_{\mathbb{R}}(\rho_0) \geq K$; hence $K_{\mathbb{R}} > K$ and $\Phi' < 0$ on (ρ_1, ρ_0). Thus $\Phi' < 0$ in the interior of I (for $\lambda = -1$, $I = \{0\}$) and Φ is continuous on I, so $\Phi \geq \Phi(\rho_0) > 0$ on I.

Now let $\rho \geq 0$, and $\rho \geq \sqrt{\lambda}$ if $\lambda > 0$. If $\rho > \rho_0$, (5.2) and (E3) give $\min_{|w|=\rho} \Phi = (1-\lambda)^p\Lambda(\theta)$ with $\rho(\theta) = \rho$: this is nonnegative, and it vanishes only for $\rho = C$, at $w(\theta_0)$ and $\overline{w(\theta_0)}$. If $\rho \in I$, the minimum is $\Phi(\rho) > 0$ by (E3) and the bound on I just proved. By (5.1), $\mathcal{D}_C \geq 0$ on $\mathbb{C} \setminus \{-1\}$ (on $\{|z| \leq 1/\sqrt{\lambda}\} \setminus \{-1\}$ if $\lambda > 0$), with zeros only at z_* and $\bar{z}_*$.

At the point $z = -1$, continuity gives $\mathcal{D}_C(-1) \geq 0$. As $C^p|1+z|^p - |1+\lambda z|^p$ is C^1 ($p > 1$) and $\operatorname{Re}(-1+iy)^q = \cos(q\pi) + q\sin(q\pi)|y| + O(y^2)$, $\mathcal{D}_C(-1+iy) = \mathcal{D}_C(-1) + cy - Kq\sin(q\pi)|y| + o(y)$ for some real c; so -1 is not a local minimum. If $\mathcal{D}_C(-1)$ were 0, then -1 would be a minimum, since $\mathcal{D}_C \geq 0$; hence $\mathcal{D}_C(-1) > 0$.

For $\lambda > 0$, Lemma 5.6 extends $\mathcal{D}_C \geq 0$ from $\{|z| \leq 1/\sqrt{\lambda}\}$ to $\mathbb{C}$, without new zeros. □

Proof of Theorem 5.1. Let $1 < p < 2$. For $\lambda \in (-1, 0]$, K_f is strictly decreasing by Lemma 5.9. For $\lambda = -1$, $R \equiv 1$, $B = 4\sin^2(\theta/2)$, $Y = 4\cos(\theta/2)\sin(\varepsilon\theta/2)$, and (5.7) gives $K_f = 2^p \sin^\varepsilon(\theta/2)/\sin(\varepsilon\theta/2)$, with $(\log K_f)' = (\varepsilon/2)[\cot(\theta/2) - \cot(\varepsilon\theta/2)] < 0$. Propositions 5.12 and 4.2 give $\|P_+ + \lambda P_-\|_p = \mathrm{GK}_p(\lambda)$, with equality in $(*)$ only at z_* and $\bar{z}_*$. For $p > 2$ use $N_p(\lambda) = N_{p'}(\lambda)$ and $\mathrm{GK}_p = \mathrm{GK}_{p'}$ (Section 2.1); for $p = 2$ both sides are $\max(1, |\lambda|)$. □

Proof of Theorem 5.2. We first prove (a) for $3 - \sqrt{3} \le p \le 3 + \sqrt{3}$. Let $3 - \sqrt{3} \le p < 2$. For $\lambda \in (-1, 1)$, K_f is strictly decreasing by Lemma 5.10, and Propositions 5.12 and 4.2 give $N_p(\lambda) = \mathrm{GK}_p(\lambda)$; the case $\lambda = -1$ is covered by Theorem 5.1, and $\lambda = 1$ is trivial ($P_+ + P_- = I$, $\mathrm{GK}_p(1) = 1$). For $|\lambda| > 1$, $N_p(\lambda) = |\lambda| N_p(1/\lambda)$ (Section 1) and $\mathrm{GK}_p(\lambda) = |\lambda|\mathrm{GK}_p(1/\lambda)$ (in (2.1), $\lambda \mapsto 1/\lambda$ amounts to $X \mapsto X/\lambda^2$). Duality covers $2 < p \le 3 + \sqrt{3} = (3 - \sqrt{3})'$, and $p = 2$ is trivial. Part (a) for $p_0 \le p < 3 - \sqrt{3}$ (and, by duality, $3 + \sqrt{3} < p \le p_0'$) is Proposition 5.14 in Section 5.6; part (b) is Proposition 4.3 and duality. Part (c) follows from (a) and Theorem 3.2. □

5.5 An exact criterion for positive ratios

Proof of Theorem 5.3. Let $1 < p < 2$ and $0 < \lambda < 1$. Recall from Section 5.2 that the function $\Lambda(\theta) = \mathcal{D}_C(z_\theta)/|1 + z_\theta|^p$ satisfies $\Lambda'(\theta) = c(\theta)\,(K - K_f(\theta))$ with $c(\theta) > 0$, $\Lambda(\theta_0) = 0$ and $K_f(\theta_0) = K$.

"If". Assume $h(\lambda, p) \ge 0$ and $\lambda < \lambda_{\mathrm{loc}}(p)$. By Lemma 4.6(b), the left side of (4.2) is less than 1. We first show that $K_f'(\theta_0) < 0$. At $\theta = \theta_0$ we have $m\omega = q\theta_0 = \pi/2$ and $\sin\omega = c_0$, so $\Psi_m(\omega) = (m^2 c_0^2 - 1)/c_0^4$. Since $m^2/(m^2 - 1) = pp'/4$ and $1/(m^2 - 1) = pp'/4 - 1$,

$$\frac{\Psi_m(\omega)}{m^2 - 1} = \frac{1 - (pp'/4)s_0^2}{c_0^4},$$

so $\Psi_m(\omega) > 4(m^2 - 1)\lambda/(1 + \lambda)^2$ is equivalent to $(pp'/4)s_0^2 + 4\lambda c_0^4/(1 + \lambda)^2 < 1$. Lemma 5.8 therefore gives $K_f'(\theta_0) < 0$.

Since $3 - \sqrt{3} < 4/3$, Lemma 5.11 (for $p < 4/3$) and Lemma 5.10 (for $p \ge 3 - \sqrt{3}$) together cover $1 < p < 2$. They show that K_f is either decreasing, or increasing on $(0, \theta_\lambda]$ and decreasing afterwards; in the second case $\theta_\lambda < \theta_0$, because $K_f'(\theta_0) < 0$ whereas $K_f' \ge 0$ on $(0, \theta_\lambda]$. Hence $\{K_f < K\} = (0, \theta_a) \cup (\theta_0, \pi)$ for some $\theta_a \in [0, \theta_0)$, so Λ increases on $(0, \theta_a)$, decreases to $\Lambda(\theta_0) = 0$ on (θ_a, θ_0), and increases on (θ_0, π). Thus $\Lambda \ge \min(\Lambda(0^+), 0)$ and $\Lambda > 0$ away from θ_0 unless $\Lambda(0^+) < 0$.

But $\Lambda(0^+) = \Phi(\rho_0)/(1 - \lambda)^p$ with ρ_0 real, i.e. a value of $\mathcal{D}_C/|1 + z|^p$ on the positive axis, which is ≥ 0 because $h \ge 0$ means $\mathcal{D}_C(x) \ge 0$ for every $x > 0$. On a circle $|w| = \rho > \rho_0$ the minimum of Φ is attained at the critical point (Lemma 5.4, (E3)) and equals $(1 - \lambda)^p \Lambda(\theta(\rho)) \ge 0$. For $\sqrt{\lambda} \le \rho \le \rho_0$, by (E3) the minimum of Φ on $|w| = \rho$ is attained at $w = \rho$. These points form the real segment $[\sqrt{\lambda}, \rho_0]$ and correspond to positive $z = (1 - w)/(w - \lambda)$, so $h \ge 0$ gives $\Phi \ge 0$ on these circles; Lemma 5.6 treats $|w| < \sqrt{\lambda}$. So $(*)$ holds (at $z = -1$ by continuity), and Proposition 4.2 gives $\|P_+ + \lambda P_-\|_p = \mathrm{GK}_p(\lambda)$.

"Only if". If $h(\lambda, p) < 0$, then $(*)$ fails on the positive axis and the norm exceeds GK (Proposition 4.2). If $\lambda > \lambda_{\mathrm{loc}}(p)$, Theorem 4.4 gives the same. At $\lambda = \lambda_{\mathrm{loc}}(p)$ $(1 < p < p_c)$, $h < 0$ by Proposition 4.7, so again the norm exceeds GK. □

5.6 The exponents $p_0 \le p < 3 - \sqrt{3}$

By Theorem 5.3, for $p_c \le p < 2$ (where $\lambda_{\mathrm{loc}} = 1$) the bound is sharp at $\lambda \in (0, 1)$ exactly when $h(\lambda, p) \ge 0$. We prove that $h(\lambda, p) > 0$ for every $\lambda \in (0, 1)$ when $p_0 \le p < 3 - \sqrt{3}$.

Put $\delta = 1 - \lambda$ and $\rho = 1 - s\delta$ with $s \in (0, 1)$. The exact relations in the proof of Proposition 4.3 give $Q(\rho) = s(1 - s)\delta^2$ and $h = \delta \min_{0<s<1} g(\delta, s)$, where

$$g(\delta, s) := \frac{Z_1(\delta) + s\,\psi_p(s\delta)}{(s(1 - s))^q} - k(\delta), \qquad Z_1 := \frac{C^p - 1}{\delta},$$

$$\psi_p(x) := \frac{1 - (1 - x)^p}{x}, \qquad k := \frac{K}{\delta} = 2C s_0^{p-2} c_0^{1-q} (C^2 + \lambda)^{q-1},$$

and $C = C(\delta)$ is the root near 1 of $s_0 C^2 - \delta C - s_0(1 - \delta) = 0$; the identity

$$Z_1 = \frac{C^p - 1}{C - 1} \cdot \frac{C - s_0}{s_0(C + 1)}$$

involves no division by δ. All of these are real-analytic near $\delta = 0$. Lemma 5.13 also uses these functions for $\delta < 0$, through their analytic continuations; there $C(\delta)$ is still the root near 1, not $\mathrm{GK}_p(1-\delta)$. At $\delta = 0$, $g(0,s) = p(A+s)/(s(1-s))^q - 2^q s_0^{p-2} c_0^{1-q}$, whose minimum over s is $\Omega(p)$, attained only at s_*.

Lemma 5.13 (The involution $\lambda \mapsto 1/\lambda$)**.** *Let $\delta^\sharp = -\delta/(1-\delta)$ (so $1-\delta^\sharp = 1/\lambda$), $s^\sharp = s(1-\delta)/(1-s\delta)$ and $\Theta(\delta) = (1-\delta)^{q-1} C(\delta)^p$. Then $C(\delta)C(\delta^\sharp) = 1$ and $g(\delta,s) = \Theta(\delta)\, g(\delta^\sharp, s^\sharp)$. Consequently, in the coordinates $\lambda = e^{-\eta}$, $\rho = e^{-\sigma\eta}$, the function $\Xi(\eta,\sigma) := g/V$, $V := e^{(1-q)\eta/2} C^{p/2}$, is even in η for each $\sigma \in (0,1)$.*

Proof. The number $D := 1/C(\delta^\sharp)$ solves $s_0 D^2 - \delta D - s_0(1-\delta) = 0$ (divide the equation of $C(\delta^\sharp)$ by $C(\delta^\sharp)^2$ and multiply by $1-\delta$, using $\delta^\sharp(1-\delta) = -\delta$); the other root is negative, so $D = C(\delta)$. Next, $1 - s^\sharp\delta^\sharp = 1/\rho$, $s^\sharp(1-s^\sharp) = \lambda s(1-s)/\rho^2$ and $\delta = -\lambda\delta^\sharp$, so $C^p - \rho^p = -C^p\rho^p\big(C(\delta^\sharp)^p - (1/\rho)^p\big)$ and the powers of ρ cancel ($p = 2q$): the first term of g is $\Theta(\delta)$ times its value at $(\delta^\sharp, s^\sharp)$. Also $C(\delta^\sharp)^2 + 1/\lambda = (C^2+\lambda)/(C^2\lambda)$ gives $k(\delta) = C^{2q}\lambda^{q-1} k(\delta^\sharp) = \Theta(\delta)\, k(\delta^\sharp)$. In the coordinates (η,σ) the involution is $(\eta,\sigma) \mapsto (-\eta,\sigma)$ and $\Theta(\delta) = V(\eta)/V(-\eta)$, so Ξ is even. □

Proposition 5.14 (The remaining exponent strip)**.** *Let $p_0 \le p \le 1.268$ and $0 < \lambda < 1$. Then $h(\lambda,p) > 0$. Hence $\|P_+ + \lambda P_-\|_p = \mathrm{GK}_p(\lambda)$ for every real λ when $p_0 \le p < 3 - \sqrt{3}$.*

Proof. By Proposition 4.3, $\Omega < 0$ on $(1, 1.2336025]$; interval arithmetic gives $\Omega' > 0$ on $[1.2336025, 1.2336027]$ and $\Omega > 0$ on $[1.2336027, 1.268]$, with the computational record in Appendix B. Hence p_0 is the only zero of Ω in $(1, 1.268]$, and $\Omega \ge 0$ on $[p_0, 1.268]$.

Case 1: $p \in [p_0, 1.2341027]$, $0 < \delta \le 0.05$. At $p = p_0$, $\delta = 0$ and $s = s_*$, the limiting minimum is $\Omega(p_0) = 0$, so g has no positive lower bound near this point, and we expand in η instead. By Lemma 5.13, $\partial_\eta \Xi(0,\sigma) = \partial_\eta^3 \Xi(0,\sigma) = 0$, so for $0 < \eta \le \eta_1 := -\log 0.95$,

$$\Xi(\eta,\sigma) = \Xi(0,\sigma) + e_2(\sigma)\eta^2 + c_4(\xi,\sigma)\eta^4, \qquad \xi \in (0,\eta).$$

Here $e_2(\sigma) = \frac{1}{2}\partial_\eta^2 \Xi(0,\sigma)$ and $c_4 = \frac{1}{24}\partial_\eta^4 \Xi$. Near (p_0, s_*), where $\Xi(0,\sigma)$ is close to 0, positivity must come from the term $e_2(\sigma)\eta^2$.

Also $\Xi(0,\sigma) = g(0,\sigma) \ge \Omega(p) \ge 0$. With $\tau_4 := \eta_1^2 \sup_{[0,\eta_1]} |c_4|$, every (σ,p) cell satisfies either $e_2 > \tau_4$ (then $\Xi > \Xi(0,\sigma) \ge 0$) or $\Xi(0,\sigma) > \eta_1^2(\max(0,-e_2) + \tau_4)$ (then $\Xi > 0$). The quotients are evaluated without division by η through the entire function $E(z) = (1-e^{-z})/z$: $\delta = \eta E(\eta)$, $s = \sigma E(\sigma\eta)/E(\eta)$, $1 - s = e^{-\sigma\eta}(1-\sigma)E((1-\sigma)\eta)/E(\eta)$, $s\psi_p(s\delta) = p\sigma E(p\sigma\eta)/E(\eta)$, $(C^p - 1)/(C-1) = p\,E(-pL)/E(-L)$ with $L = \log C$. The bound τ_4 comes from power series centered on subintervals of $[0,\eta_1]$, and e_2, $\Xi(0,\sigma)$ from the series at $\eta = 0$ ($\sigma \in [0.045, 0.95]$, which covers $s \in [0.05, 0.95]$ since $sE(\eta) \le \sigma \le s$). Indeed, $\sigma = -\log(1-s\delta)/\eta \ge s\delta/\eta = sE(\eta)$ by $-\log(1-x) \ge x$, whereas concavity gives $\log(1-s\delta) \ge s\log(1-\delta) = -s\eta$, hence $\sigma \le s$. After bisection one of the two alternatives holds on every final cell, with $\tau_4 \le 0.0191$; Appendix B gives the numbers of cells. Near the ends $s \le 0.05$ and $s \ge 0.95$ we have $g \ge Z_1/0.05^q - k$ and $g \ge (Z_1 + 0.95)/0.05^q - k$, respectively (as $\psi_p \ge 1$), and interval arithmetic shows that, for (p,δ) in the range of Case 1, these lower bounds exceed 0.2920 and 6.320.

Case 2: $\delta \ge 0.05$ or $p \ge 1.2341027$. Here we use interval arithmetic on boxes in (p,δ,s), with the same treatment of the ends in s and of $\delta = 0$ (the binomial series of $((1+v)^p - 1)/v$ with a rigorous tail); the value $\delta = 1$ ($\lambda = 0$) is included in the boxes. A mean-value form proves $g > 0$ on boxes with $0.05 \le s \le 0.95$, and the bounds at the ends in s prove it on rectangles in (p,δ) for $s \le 0.05$ and $s \ge 0.95$. The repository [5] also contains independent programs that check these inequalities.

For $p_0 \le p < 3 - \sqrt{3}$ we have $p > p_c$, so $\lambda_{\mathrm{loc}} = 1$ and Theorem 5.3 turns $h > 0$ into equality for $0 < \lambda < 1$. For $\lambda \in [-1, 0]$ use Theorem 5.1; $\lambda = 1$ is trivial, and $|\lambda| > 1$ follows from reciprocal symmetry. So the bound is sharp for every real ratio when $p_0 \le p < 3 - \sqrt{3}$. □

6 Further results

This section collects further results; none of them is used in Sections 1–5. The chord inequality of Lemma 2.3 holds for half-widths up to π/q, not only for the arch; applied to the obstacle, wider chords give lower bounds for the norm, strictly above the Gohberg–Krupnik bound, each proved at one explicit choice of C, c and s (Proposition 6.3). We also show that the window of Section 3 already yields the Gohberg–Krupnik bound, compare the two ways in which sharpness fails for positive ratios, show that the exponent $3 - \sqrt{3}$ in Lemma 5.10 is optimal, give a second proof of Proposition 5.14, and report runs of the programs on false claims, which they did not certify. In this section ρ again denotes a complex ratio, as in Sections 1–3.

6.1 Wide chords and explicit lower bounds

The arch test (3.1) uses the chord inequality of Lemma 2.3 with half-width $s = \pi/(2q) = \pi/p$. Wider chords give further necessary conditions.

Lemma 6.1 (Wide chords)**.** *Let $1 < p < 2$, $\rho \in \mathbb{C}$ and $C > \max(1, |\rho|)$, and suppose that $N_p(\rho) \le C$. Then for every $c \in \mathbb{R}$ and every $s \in [\pi/p, \pi]$,*

$$m_C(c-s) + m_C(c+s) \;\ge\; 2\cos(qs)\, m_C(c). \tag{6.1}$$

Proof. By Theorem 2.2 there is a q-subtrigonometric $\varphi \le m_C$. Since $q < 1$, we have $s \le \pi < \pi/q$, so Lemma 2.3 applies; and $\cos(qs) \le 0$ because $qs \in [\pi/2, q\pi]$. Hence

$$m_C(c-s) + m_C(c+s) \;\ge\; \varphi(c-s) + \varphi(c+s) \;\ge\; 2\cos(qs)\,\varphi(c) \;\ge\; 2\cos(qs)\, m_C(c). \qquad \square$$

For $s = \pi/p$ the right side of (6.1) vanishes and the lemma is the arch test. If (6.1) fails for some choice of c and s, then $N_p(\rho) > C$. Because m_C enters the left side minus the right side of (6.1) with the nonnegative coefficients 1, 1 and $-2\cos(qs)$, a failure can be certified by *upper* bounds for m_C: by definition $m_C(t) \le \Gamma_C(y, t)$ for every real y. Such upper bounds are explicit values of an elementary function and can be enclosed by interval arithmetic. For $p > 2$ one applies the lemma at p', using $N_p(\rho) = N_{p'}(\bar\rho)$.

Taking $c = 0$ and s slightly larger than θ_0 gives a second proof that the positive-axis test is necessary (the implication $h(\lambda, p) < 0 \Rightarrow N_p(\lambda) > \mathrm{GK}_p(\lambda)$ of Section 4.3).

Corollary 6.2 (First-order wide chord)**.** *Let $1 < p < 2$ and $\lambda \in (-1, 1)$. If $h(\lambda, p) < 0$, then $N_p(\lambda) > \mathrm{GK}_p(\lambda)$.*

Proof. Put $C = \mathrm{GK}_p(\lambda)$, so that $h(\lambda, p) = m_C(0) - K$; since $\lambda \ne 1$, $C > \max(1, |\lambda|)$ (Section 3), so Lemma 6.1 applies. The obstacle is even, since λ is real. Take $c = 0$ and $s = \theta_0 + d$ with small $d > 0$. By Lemma 4.1 and Section 4.1, $m_C(\theta_0) = 0$ and m_C is differentiable at θ_0 with $m_C'(\theta_0) = -qK$, so $m_C(s) = -qKd + o(d)$. Also $\cos(q\theta_0) = 0$ and the derivative of $\cos(qs)$ at $s = \theta_0$ is $-q$, so $\cos(qs) = -qd + o(d)$. Therefore

$$m_C(-s) + m_C(s) - 2\cos(qs)\, m_C(0) \;=\; 2qd\,\big(m_C(0) - K\big) + o(d),$$

which is negative for small $d > 0$. This violates (6.1), so $N_p(\lambda) > C$ by Lemma 6.1. $\square$

This proof does not use the results of [7, Section 3], which the proof of Proposition 4.2 needs.

Proposition 6.3 below treats two sample pairs. For $\lambda = \frac12$ and $p = 21/20$ the ratio lies below $\lambda_{\mathrm{loc}}(21/20) = 0.5545\ldots$, so Theorem 4.4 does not apply; strict inequality follows from Proposition 4.2, since (4.6) with $\vartheta = \frac12$ gives $h(\frac12, \frac{21}{20}) < -0.15$. For $\lambda = \frac35$ and $p = 8$, strict inequality follows from Theorem 4.4 at $p' = 8/7$, since $\frac35 > \lambda_{\mathrm{loc}}(8/7) = 0.5447\ldots$. The two bounds of Proposition 6.3 also bound the gap $N_p(\lambda) - \mathrm{GK}_p(\lambda)$ from below. Each is proved by showing that (6.1) fails at one explicit choice of C, c and s, using upper bounds for m_C.

Proposition 6.3 (Two explicit lower bounds)**.**

$$\|P_+ + \tfrac12 P_-\|_{21/20} \;>\; 3.53 \;>\; \mathrm{GK}_{21/20}(\tfrac12) = 3.4977039\ldots, \qquad \|P_+ + \tfrac35 P_-\|_8 \;>\; 1.468 \;>\; \mathrm{GK}_8(\tfrac35) = 1.4570\ldots$$

Proof. For the first bound take $p = 21/20$, $C = 353/100$, $c = 0$, $s = \pi - 51741/10^6$, $y_0 = 41/250$ and $y_1 = 3/250$. Since m_C is even, $m_C(\pm s) \le \Gamma_C(y_1, s)$; also $m_C(0) \le \Gamma_C(y_0, 0)$ and $\cos(qs) < 0$. Hence the quantity

$$S := 2\,\Gamma_C(y_1, s) - 2\cos(qs)\,\Gamma_C(y_0, 0)$$

is an upper bound for the left side of (6.1) minus its right side. Interval arithmetic at 160 bits gives $S < -0.00196$, so (6.1) fails and $\|P_+ + \frac12 P_-\|_{21/20} > 353/100 = 3.53$ by Lemma 6.1.

For the second bound, $N_8(\frac35) = N_{8/7}(\frac35)$ by duality. Take $p = 8/7$, $C = 367/250$, $c = 0$, $s = \pi - 27361/200000$, $y_0 = 13/25$ and $y_1 = 101/1000$; the same evaluation gives $S < -0.00074$, so $\|P_+ + \frac35 P_-\|_{8/7} > 367/250 = 1.468$. The repository [5] also contains two further programs, not part of the recorded runs, that prove both bounds again. $\square$

6.2 The Gohberg–Krupnik bound from the window

The window of Section 3.1 also gives the lower bound $N_p(\rho) \geq \mathrm{GK}_p(\rho)$ itself. Let $p \neq 2$, $\rho \neq 1$ and $\max(1, |\rho|) < C < \mathrm{GK}_p(\rho)$ (for $p = 2$ or $\rho = 1$ we have $\mathrm{GK}_p(\rho) = \max(1, |\rho|)$, and the bound is the one in Section 2.2). As C decreases to $\max(1, |\rho|)$, the right side of (3.2) tends to 0 (its denominator stays positive, since $\rho \neq 1$), so $\psi(C) \to \pi/2$. Since ψ is continuous and, by Lemma 3.1, $\psi(C) = \pi/\max(p, p')$ only at $C = \mathrm{GK}_p(\rho)$, we have $\psi(C) > \pi/\max(p, p')$ for every such C: the open window $\{t : m_C(t) < 0\}$ is longer than $2\pi/\max(p, p')$.

If $p > 2$, the window then contains two angles t and $t + 2\pi/p$. If $p < 2$, the two angles t and $t + 2\pi/p$ split the circle into arcs of lengths $2\pi/p$ and $2\pi/p'$, and t can be chosen so that the window contains the closed arc of length $2\pi/p'$, hence both of its endpoints. In both cases $m_C(t) + m_C(t + 2\pi/p) < 0$, so the arch test (3.1) gives $\|P_+ + \rho P_-\|_p > C$. Letting $C \uparrow \mathrm{GK}_p(\rho)$ gives $\|P_+ + \rho P_-\|_p \geq \mathrm{GK}_p(\rho)$. Thus the circle form of the Gohberg–Krupnik bound follows from Theorem 2.2 and Lemma 3.1, without [6] or Lemma 1.2.

6.3 The two failure mechanisms for positive ratios

Both ways in which sharpness fails for $0 < \lambda < 1$ are chord conditions at θ_0: Theorem 4.4 is the arch test at second order, and Corollary 6.2 is the wide-chord inequality (6.1) at first order. Let $1 < p < p_c$, and let $\lambda_1(p)$ be as in Section 4.3. The first-order condition starts to fail at smaller ratios: $\lambda_1(p) < \lambda_{\mathrm{loc}}(p)$ by Proposition 4.7, so the positive-axis test already fails at ratios where the second-order condition still holds. For example, at $p = 1.05$, numerically $\lambda_1 \approx 0.391$ while $\lambda_{\mathrm{loc}} = 0.5545\ldots$, and the ratio $\lambda = \frac{1}{2}$ of the first bound in Proposition 6.3 lies between the two.

We evaluated $h(\cdot, p)$ in double precision at 200 equally spaced ratios in $[0.01, 0.999]$, for eleven values of p between 1.0005 and 1.2134 (this scan is not part of the supporting calculations). In each case $h(\cdot, p)$ changed sign exactly once on this grid, so the set $\{\lambda \in (0, 1) : h(\lambda, p) \geq 0\}$ appeared to be an interval $(0, \lambda_1(p)]$. If this holds for every $1 < p < p_c$, then $h \geq 0$ forces $\lambda \leq \lambda_1(p) < \lambda_{\mathrm{loc}}(p)$, so the second condition of Theorem 5.3 is redundant there (the question left open after Theorem 1.1), and the sharp positive ratios are exactly $(0, \lambda_1(p)]$. We do not prove this interval property.

6.4 The exponent $3 - \sqrt{3}$ is optimal in Lemma 5.10

Proposition 6.4 (Optimality of $3 - \sqrt{3}$)**.** *If $1 < p < 3 - \sqrt{3}$, then for every $\lambda < 1$ sufficiently close to 1 the function K_f is not monotone on $(0, \pi)$.*

Proof. Write $\varepsilon = p - 1$, so $p < 3 - \sqrt{3}$ means $\varepsilon < 2 - \sqrt{3}$. In the proof of Lemma 5.10, the logarithm of $\varepsilon \sin^2((1-\varepsilon)\theta)/\big((1-\varepsilon)^2 \sin\theta \sin(\varepsilon\theta)\big)$ is the double series with brackets $1 + \varepsilon^{2k} - 2(1-\varepsilon)^{2k}$. For $\varepsilon < 2 - \sqrt{3}$ the bracket with $k = 1$, namely $-(\varepsilon^2 - 4\varepsilon + 1)$, is negative. The terms with $k = 1$ sum to $-(\varepsilon^2 - 4\varepsilon + 1)\theta^2/6$, since $\sum_n (n\pi)^{-2} = 1/6$. Each bracket has absolute value at most 2 and $\sum_{k \geq 2} \xi^{2k}/k \leq \xi^4/(2(1-\xi^2))$, so the remaining terms are bounded in absolute value by $(\pi^4/90)(\theta/\pi)^4/(1 - (\theta/\pi)^2) = O(\theta^4)$. Hence the logarithm is negative for small $\theta > 0$, that is, $W(\theta) > 0$. By the identity $W/(4\varepsilon \sin^4\omega) = 1 - \Psi_m(\omega)/(m^2 - 1)$ from the proof of Lemma 5.8, this means $\Psi_m(\omega) < m^2 - 1$ at the corresponding $\omega = (1-q)\theta$. Since $4(m^2-1)\lambda/(1+\lambda)^2 \to m^2 - 1$ as $\lambda \to 1$, Lemma 5.8 shows that K_f increases at this θ for every λ sufficiently close to 1. On the other hand, at $\omega_e = \pi/(m+1)$ we have $\sin(m\omega_e) = \sin\omega_e$, so $\Psi_m(\omega_e) = (m^2 - 1)/\sin^2\omega_e > m^2 - 1$, and K_f decreases near $\theta = \pi$ for every $\lambda \in (0, 1)$, as in the proof of Lemma 5.11. So K_f is not monotone. □

Since $p_0 < 3 - \sqrt{3}$, Proposition 5.12 does not apply on $[p_0, 3 - \sqrt{3})$ for λ near 1; this is why Section 5.6 proves the positive-axis test directly.

6.5 A second proof of Proposition 5.14

In the proof of Proposition 5.14, the Case 1 corner $p \in [p_0, 1.2341027]$, $0 < \delta \leq 0.05$ is handled by the expansion in η that Lemma 5.13 provides. An independent program proves $g > 0$ on the smaller corner $[p_0, 1.2337027] \times (0, 0.01]$ in (p, δ) by Taylor's formula in the original coordinates. Fix p, put $s_* = s_*(p)$ and $\mathbf{v} = (s - s_*, \delta)^{\mathsf{T}}$, and let $\nabla^2 g(\delta, s)$ be the Hessian of g with respect to the ordered pair (s, δ). At the base point $s = s_*$, $\delta = 0$ the derivative of g in s vanishes, since s_* minimizes $g(0, \cdot)$. The derivative in δ is $\frac{1}{2}(pA + 1 - q)\,\Omega(p)$, with $pA + 1 - q > 0$. Indeed, differentiating $g(\delta, s_*) = \Theta(\delta) g(\delta^\sharp, s^\sharp)$ (Lemma 5.13) at $\delta = 0$, with $\Theta(0) = 1$,

$\Theta'(0) = pA + 1 - q$ (because $C'(0) = A$), $d\delta^\sharp/d\delta = -1$ and $\partial_s g(0, s_*) = 0$, gives $2\,\partial_\delta g(0, s_*) = (pA + 1 - q)\,\Omega(p)$. Thus

$$g(\delta, s) = \Omega(p)\,\big(1 + \tfrac{1}{2}(pA + 1 - q)\,\delta\big) + \int_0^1 (1 - t)\,\mathbf{v}^{\mathsf{T}} \nabla^2 g\big(t\delta, s_* + t(s - s_*)\big)\mathbf{v}\,dt.$$

Since $\Omega(p) \geq 0$ for $p \geq p_0$, it suffices that the Hessian is positive definite. The program proves this at 120 bits on the rectangle $1/4 \leq s \leq 27/100$, $0 \leq \delta \leq 0.01$, for every $p \in [1.2336025, 1.2337027]$, using 1385 boxes. Since $[1/4, 27/100]$ contains $s_*(p)$, the rectangle contains every segment from $(s_*, 0)$ to one of its points. Values of s outside $[1/4, 27/100]$ are covered by mean-value boxes and the tail bounds in s, and a second, independent program covers the rest of the strip, $[1.2336025, 1.268] \times [0.01, 1]$ and $[1.2337027, 1.268] \times [0, 0.01]$ (7938 boxes and 920 tail rectangles). These two programs are in the repository [5] but are not part of the recorded runs of Appendix B. At $p = p_0$, $s = s_*$, $\delta = 0$ the Schur complement $(g_{\delta\delta} - g_{s\delta}^2/g_{ss})/2$ equals $e_2(\sigma_*) \approx 0.1092$, where $\sigma_* = s_*(p_0)$: at $p = p_0$ both are the coefficient of δ^2 in $\min_s g(\delta, s)$, since $\delta = \eta + O(\eta^2)$ and $V = 1 + O(\eta)$. So the two methods agree there.

6.6 Checks of the programs

The programs were also run on false claims, and they did not certify them. For instance, the program for case (c) of the proof of Proposition 4.7 did not certify $g < 0$ when $s = 1/50$ replaces $s = 1371/5000$; the program for the sign of Ω did not certify $\Omega < 0$ on $[1.2336027 - 10^{-9}, 1.2336027]$, where $\Omega > 0$; and the program for Case 2 of Proposition 5.14 did not certify $g > 3 \times 10^{-4}$ on a box where the numerical minimum of g is 2.93×10^{-4}.

A An estimate near the endpoint $p = 1$

This appendix proves the estimate used in case (b) of the proof of Proposition 4.7.

Lemma A.1 (Endpoint estimate)**.** *For $1 < p \leq 1 + 10^{-6}$, $h(\lambda_{\mathrm{loc}}(p), p) < -1.88$.*

Proof. Let $1 < p \leq 1 + 10^{-6}$, $a = \pi/p'$, and $\bar{a} = 3.2 \times 10^{-6}$; then $0 < a < \bar{a}$, $s_0 = \sin a = \sin(\pi/p)$ and $c_0 = \cos a = -\cos(\pi/p)$. Put

$$d^2 = \frac{s_0^2}{c_0^4}\left(\frac{p\pi}{4a} - 1 - c_0^2\right), \qquad \lambda = \frac{1 - d}{1 + d}, \qquad \delta = 1 - \lambda = \frac{2d}{1 + d}, \qquad Q_A = \frac{\delta}{s_0}.$$

Here $\lambda = \lambda_{\mathrm{loc}}(p)$. Since $p = \pi/(\pi - a)$,

$$p - 1 \leq a/3, \quad a - a^3/6 \leq s_0 \leq a, \quad 1 - a^2/2 \leq c_0 \leq 1, \quad \frac{\pi a}{4}(1 - 3a) \leq d^2 \leq \frac{\pi a}{4}(1 + a).$$

For the last enclosure, its lower bound follows by using $s_0^2 \geq a^2(1 - a^2/3)$ and $p\pi/(4a) - 1 - c_0^2 \geq \pi/(4a) - 2$; the upper bound follows from $c_0^{-4} \leq 1 + 3a^2$ and $(1 + a/3)(1 + 3a^2) \leq 1 + a$ on this interval. These bounds imply

$$\sqrt{a} \leq \delta \leq 1.8\sqrt{a}, \qquad a^{-1/2} \leq Q_A \leq 2a^{-1/2}, \qquad 2d \leq 1.8\sqrt{a}.$$

Write $C = \mathrm{GK}_p(\lambda)$. Equation (2.2), a quadratic in C, gives $C = (Q_A + \sqrt{Q_A^2 + 4\lambda})/2$, hence $C^2 \leq Q_A^2 + 2$ and $C \leq Q_A + Q_A^{-1} \leq 2Q_A$. Inequality (4.6) with $\vartheta = 1/2$, that is, $\rho = (1 + \lambda)/2$, gives

$$h(\lambda, p) \leq U_1 - U_2 - K, \qquad U_1 = (2C)^p, \qquad U_2 = (2\rho)^p.$$

With $x = (p - 1)\log(2C)$, we have $0 \leq x \leq (a/3)\log(8/\sqrt{a}) < 1$. The estimate $e^x \leq 1 + 2x$ therefore yields

$$U_1 \leq 2Q_A + \frac{2}{Q_A} + \frac{8}{3}Q_A a \log\frac{8}{\sqrt{a}}, \qquad U_2 \geq 2\rho = \frac{2}{1 + d} \geq 2 - 2d.$$

For the lower bound on K, let r be the contact modulus and put $v = r - \lambda/r = \sqrt{\delta^2 + 4\lambda s_0^2}/c_0$. The contact equation $r + \lambda/r = (1 + \lambda)/c_0$ implies

$$\frac{1}{c_0} \leq r \leq \frac{1 + s_0^2/\delta}{c_0}, \qquad v \leq \frac{\delta(1 + 2s_0^2/\delta^2)}{c_0}, \qquad |re^{i\pi/p} + \lambda| \geq r - \lambda \geq \delta.$$

Write $G = \mathcal{P}(e^{pL} - 1)$ with $\mathcal{P} = |\zeta + \lambda|^p$ and $L = \log C - \log|w|$, as in the proof of Theorem 4.4. At the contact $L = 0$, so $e^{qy_*}\partial_t\Gamma_C = G_t = p\mathcal{P}L_t$, and $L_t = -\delta/(s_0 v)$ by (4.4); so $\partial_t\Gamma_C = -pr^{-q}|\zeta_* + \lambda|^p\delta/(s_0 v)$. With $K = -\partial_t\Gamma_C/q$ from Section 4.1, $q = p/2 < 1$ and $Q_A = \delta/s_0$, this yields

$$K = 2Q_A\, r^{-q}\,\frac{|re^{i\pi/p} + \lambda|^p}{v} \geq 2Q_A \frac{c_0^2\delta^{p-1}}{(1 + s_0^2/\delta)(1 + 2s_0^2/\delta^2)} \geq 2Q_A(1 - E_K),$$

where

$$E_K = (p-1)\log(1/\delta) + a^2 + \frac{s_0^2}{\delta} + \frac{2s_0^2}{\delta^2} \leq \frac{a}{6}\log\frac{1}{a} + a^2 + a^{3/2} + 2a.$$

In the second lower bound for K we used $\delta^{p-1} \geq 1 - (p-1)\log(1/\delta)$, $c_0^2 \geq 1 - a^2$, and $(1+u)^{-1} \geq 1 - u$ for $u \geq 0$; all the subtracted terms are less than one here. Combining the bounds gives

$$h(\lambda_{\mathrm{loc}}(p), p) \leq -2 + R(a), \qquad R(a) = \sqrt{a}\left[2 + \frac{16}{3}\log\frac{8}{\sqrt{a}} + 1.8 + \frac{2}{3}\log\frac{1}{a} + 8 + 4\sqrt{a} + 4a\right].$$

Every term in R is increasing on $(0, \bar{a}]$: in particular $\sqrt{a}\log(1/a)$ increases for $a < e^{-2}$. Finally $\sqrt{\bar{a}} < 0.001789$, $\log(8/\sqrt{\bar{a}}) < 8.5$, and $\log(1/\bar{a}) < 12.66$ give $R(\bar{a}) < 0.12$. Thus $h(\lambda_{\mathrm{loc}}(p), p) < -1.88$ throughout the endpoint interval. □

B Interval-arithmetic verifications

This appendix records the interval calculations used in the proof of Theorem 1.1 and lists their domains in Table 1. The interval calculations use outward-rounded real balls. Let f be an explicitly defined real function. If the enclosure of f on a box B lies in $(0, \infty)$, then $f > 0$ on B. Boxes on which the enclosure does not give the required sign are bisected, and a calculation succeeds only when every box is certified. Removable quotients are evaluated by the power series and mean-value formulas described in the corresponding proofs, with rigorous remainder bounds. In the calculations of Table 1, floating-point optimization is used only to propose rational test points for the positive-axis test; those points are then checked by interval arithmetic.

The symbols h, g, δ, s refer to the rescaled positive-axis formulas in Proposition 4.7 and Section 5.6; the two formulas for g coincide (in Section 5.6 the argument p is suppressed). For case (c) of the proof of Proposition 4.7 put $s_f = 1371/5000$ and $g_f(p, \delta) = g(\delta, p, s_f)$. Write $p_0^- = 1.2336025$, $p_0^+ = 1.2336027$ and $p_1 = 1.2341027$.

For $1.2134 \leq p < p_c$ the rows for g_f and T combine as in case (c) of the proof of Proposition 4.7: $0 < \delta_{\mathrm{loc}}(p) = 1 - \lambda_{\mathrm{loc}}(p) \leq 0.03$, so $h(\lambda_{\mathrm{loc}}(p), p) \leq \delta_{\mathrm{loc}}(p) g_f(p, \delta_{\mathrm{loc}}(p)) < 0$.

The ends of the ranges are treated analytically: Proposition 4.3 covers $1 < p \leq 1.001$ for $\Omega < 0$, Lemma 4.6 covers the ends for $\mathcal{F}$, and Appendix A covers $1 < p \leq 1 + 10^{-6}$ for case (b) of the proof of Proposition 4.7. The Case 1 cells begin at $p_0^- < p_0$, but their conclusion is used only for $p \geq p_0$, where $\Omega(p) \geq 0$ by Proposition 5.14. The tails $s \to 0, 1$ are covered by the analytic bounds in Section 5.6.

The interval proofs use Arb ball arithmetic [9] through python-flint 0.9.0 (FLINT 3.6.0) [2]. The programs, with the exact commands, inputs and outputs, are in the repository [5]; it also contains independent programs for the inequalities $\Omega' > 0$, $\Omega > 0$ and $g > 0$, which are not part of the recorded runs.

Quantity	Domain	Verified conclusion	Precision; size
Ω (Prop. 4.3)	$p \in [1.001, p_0^-]$	$\Omega < 0$	120 bits; 141 cells
Ω', Ω (Prop. 5.14)	$p \in [p_0^-, p_0^+]$, respectively $[p_0^+, 1.268]$	$\Omega' > 0$, respectively $\Omega > 0$	120 bits; 1 cell and 35 cells
$\mathcal{F}, \mathcal{F}'$ (Lemma 4.6)	$p \in [1.000001, 1.2134]$, $[1.2136, 1.5]$ and $[1.2134, 1.2136]$	$\mathcal{F} > 0$, $\mathcal{F} < 0$ and $\mathcal{F}' < 0$, respectively	192 bits; 12 cells, 13 cells and 1 cell
Root of $\mathcal{F}$ (Lemma 4.6)	the two points 1.213490304 and 1.213490305	$\mathcal{F}(1.213490304) > 0 > \mathcal{F}(1.213490305)$	192 bits; 2 points
$h(\lambda_{\mathrm{loc}}(p), p)$ (Prop. 4.7)	$p \in [1.000001, 1.2134]$	$(C^p - \rho^p)/(\vartheta(1-\vartheta))^q - K < 0$ at $\lambda = \lambda_{\mathrm{loc}}(p)$, $\rho = \lambda + \vartheta(1-\lambda)$, one rational ϑ per cell; hence $h(\lambda_{\mathrm{loc}}(p), p) < 0$ by (4.6)	192 bits; 17785 cells
g_f (Prop. 4.7)	$(p, \delta) \in [1.2134, 1.2134904] \times [0, 0.03]$	the removable extension of g_f is negative	160 bits; 1023 cells
T (Prop. 4.7)	$p \in [1.2134, 1.2134904]$	$T(p) \le 0.015^2$	160 bits; 200 cells
g, Case 1 (Prop. 5.14)	$p \in [p_0^-, p_1]$, $0 < \delta \le 0.05$, $0 < s < 1$	$e_2 > \tau_4$ or $\Xi(0, \sigma) > \eta_1^2(\max(0, -e_2) + \tau_4)$ on every cell; tail bounds > 0.2920 and > 6.320	128 bits; 2729 final cells (1929 by the first alternative, 800 by the second), 64 coefficient subintervals
g, Case 2 (Prop. 5.14)	$[p_0^-, 1.268] \times [0.05, 1] \cup [p_1, 1.268] \times [0, 0.05]$ in (p, δ)	$g > 0$ for every $0 < s < 1$	120 bits; 13168 mean-value boxes and 714 tail rectangles

Table 1: Interval domains used in the proof. All decimal endpoints are exact rational numbers, and $1.000001 = 1 + 10^{-6}$. The domains of Cases 1 and 2 of Proposition 5.14 share boundary segments and together cover the required region. Here $p_0^- = 1.2336025$, $p_0^+ = 1.2336027$ and $p_1 = 1.2341027$.

Acknowledgments

This work was done overwhelmingly by AI models, steered by the author in a continuous critical dialogue, variously with Google's Gemini, Anthropic's Claude and OpenAI's ChatGPT. The author takes full responsibility for the paper.

References

[1] Y. Ding, L. Grafakos, K. Zhu, On the norm of the operator $aI + bH$ on $L^p(\mathbb{R})$, Bull. Korean Math. Soc. 55 (2018), 1209–1219.

[2] The FLINT team, FLINT: Fast Library for Number Theory, version 3.6.0 (2026), https://flintlib.org; python-flint, version 0.9.0 (2026), https://github.com/flintlib/python-flint

[3] T. W. Gamelin, *Uniform algebras and Jensen measures*, London Math. Soc. Lecture Note Ser. 32, Cambridge Univ. Press, Cambridge, 1978.

[4] D. Goel, When is the Gohberg–Krupnik (1992) lower bound for $aI + bH$ sharp? Exactly when $\mathrm{Re}(a\bar{b}) = 0$, for $1.2336\ldots \le p \le 5.2807\ldots$, $p \ne 2$, Zenodo (2026), https://doi.org/10.5281/zenodo.23732817

[5] D. Goel, Supporting calculations for the Gohberg–Krupnik sharpness classification (2026), software repository, https://github.com/david44hello/paper_gohberg-krupnik-2003-sharpness/tree/v3.90

[6] I. Gohberg, N. Krupnik, *One-dimensional linear singular integral equations*, Vols. I, II, Oper. Theory Adv. Appl. 53, 54, Birkhäuser, Basel, 1992 (Russian original 1973).

[7] B. Hollenbeck, N. J. Kalton, I. E. Verbitsky, Best constants for some operators associated with the Fourier and Hilbert transforms, Studia Math. 157 (2003), 237–278.

[8] B. Hollenbeck, I. E. Verbitsky, Best constants for the Riesz projection, J. Funct. Anal. 175 (2000), 370–392.

[9] F. Johansson, Arb: efficient arbitrary-precision midpoint-radius interval arithmetic, IEEE Trans. Comput. 66 (2017), 1281–1292.

[10] B. Ya. Levin, *Lectures on entire functions*, Transl. Math. Monogr. 150, Amer. Math. Soc., Providence, RI, 1996.

[11] Z. Long, Sharp noncommutative Hilbert and Riesz inequalities: classification and spectral corrections, arXiv:2610.10158v2 (2026).

[12] S. K. Pichorides, On the best values of the constants in the theorems of M. Riesz, Zygmund and Kolmogorov, Studia Math. 44 (1972), 165–179.