Journal of MATHEMATICAL
www.elsevier.com/locate/jmaa

On the spans of polynomials and their derivatives

Rajesh Pereira

University of Saskatchewan, Saskatoon SK, Canada S7N 5E6

Received 14 June 2004

Available online 23 February 2005

Submitted by H.M. Srivastava

Abstract

We prove that a majorization-type relation among the root sets of three polynomials implies that the same relation holds for the root sets of their derivatives. We then use this result to give a unified derivation of the classical results due to Sz.-Nagy, Robinson, Meir and Sharma which relate the span of a polynomial to the spans of its first or higher derivatives. We also show how this relation can be generated by interlacing polynomials.

Keywords: Spans of polynomials; Differentiators; Geometry of polynomials; Majorization order; Interlacing

1. An ordering

Whenever roots or eigenvalues are listed in this paper, any root or eigenvalue of multiplicity $m > 1$ will always be listed m times.

We begin with a brief review of majorization, a key topic in the theory of inequalities and matrix analysis which has recently been playing an increasing role in the geometry of polynomials ([1] and [7] are good examples of this). [4] is the standard reference for majorization and contains a wealth of information about this subject.

E-mail address: rjxpereira@yahoo.com.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2005.01.010

Definition 1.1. Let $(a_1, a_2, \dots, a_n)$ and $(b_1, b_2, \dots, b_n)$ be two n-tuples of real numbers arranged in descending order. Then we say $(a_1, a_2, \dots, a_n)$ is majorized by $(b_1, b_2, \dots, b_n)$ (and we write $(a_1, a_2, \dots, a_n) \prec (b_1, b_2, \dots, b_n)$) if

(1) $\sum_{i=1}^{k} a_i \leqslant \sum_{i=1}^{k} b_i$ for $\forall k$, $1 \leqslant k \leqslant n-1$, and
(2) $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i$.

We note that we could replace condition (1) with the equivalent condition

(1′) $\sum_{i=k}^{n} a_i \geqslant \sum_{i=k}^{n} b_i$ for $\forall k$, $2 \leqslant k \leqslant n$.

Roughly speaking, $(a_1, a_2, \dots, a_n)$ is majorized by $(b_1, b_2, \dots, b_n)$ means that the n-tuple $(a_1, a_2, \dots, a_n)$ is less spread out than $(b_1, b_2, \dots, b_n)$.

In what follows, S_n is the group of permutations on n elements.

Proposition 1.2 [4]. *Let I be any interval in $\mathbb{R}$ and let $(a_1, a_2, \dots, a_n)$ and $(b_1, b_2, \dots, b_n)$ be two n-tuples of real numbers in I. Then the following are equivalent*:

(1) $(a_1, a_2, \dots, a_n) \prec (b_1, b_2, \dots, b_n)$.
(2) $(a_1, a_2, \dots, a_n)$ *is in the convex hull of* $\{(b_{\sigma(1)}, b_{\sigma(2)}, \dots, b_{\sigma(n)})\}_{\sigma \in S_n}$.
(3) $\sum_{i=1}^{n} \phi(a_i) \leqslant \sum_{i=1}^{n} \phi(b_i)$ *for all convex functions* $\phi : I \to \mathbb{R}$.

Definition 1.3. Let $x = (x_1, x_2, \dots, x_n)$ be an n-tuple of real numbers, and let $k \leqslant n-2$. Then $x^{(k)} = (y_1, y_2, \dots, y_{n-k})$, where $\{y_i\}_{i=1}^{n-k}$ are the roots of the kth derivative of the polynomial $p(x) = \prod_{i=1}^{n}(x - x_i)$ listed in descending order.

We are now ready to state the main theorem of our paper.

Theorem 1.4. *Let x, y and z be three n-tuples of real numbers listed in descending order, and let $k \leqslant n-2$. If $x \prec y + z$, then $x^{(k)} \prec y^{(k)} + z^{(k)}$.*

We defer proof of this result to section three of this paper. We note that if $z = (0, 0, \dots, 0)$, this theorem reduces to a result of Borcea [1, Corollary 1.3]. We will also need the following result which shows that the ordering in Theorem 1.4 can be used to prove some seminorm inequalities. We recall that a seminorm $\|\cdot\|$ on $\mathbb{R}^n$ is permutation-invariant if $\|Px\| = \|x\|$ for all n by n permutation matrices P and all $x \in \mathbb{R}^n$.

Proposition 1.5. *Let x, y and z be three n-tuples of real numbers listed in descending order and let $\|\cdot\|$ be a permutation-invariant seminorm on $\mathbb{R}^n$. If $x \prec y + z$, then $\|x\| \leqslant \|y\| + \|z\|$.*

Proof. Let $\{P_\sigma\}_{\sigma \in S_n}$ be the set of n by n permutation matrices. By Proposition 1.2, $x \prec y + z$ if and only if there exists non-negative real numbers $\{\lambda_\sigma\}_{\sigma \in S_n}$ with $\sum_{\sigma \in S_n} \lambda_\sigma = 1$ such that $x = \sum_{\sigma \in S_n} \lambda_\sigma P_\sigma (y + z)$. Hence $\|x\| \leqslant \sum_{\sigma \in S_n} \lambda_\sigma \|P_\sigma (y + z)\| = \|y + z\| \leqslant \|y\| + \|z\|$. $\square$

2. Spans of hyperbolic polynomials

We say that a polynomial p is hyperbolic if all of its roots are real. For any nth degree hyperbolic polynomial p, $Z(p)$ denotes the n-tuple consisting of the roots of p listed in descending order.

The span of a hyperbolic polynomial p is the difference between the largest root of p and the smallest root of p. Equivalently, the span of a hyperbolic polynomial p is the length of the smallest interval which contains all the roots of p which immediately gives us $\operatorname{span}(p') \leqslant \operatorname{span}(p)$. Some (less trivial) classical results on the spans of hyperbolic polynomials can be easily derived from Theorem 1.4. (See [8, Chapter 6] for a compilation of these and many other results on hyperbolic polynomials.)

We can use our machinery to prove the following result of Sz.-Nagy [10] (independently rediscovered by Meir and Sharma [6]) which relates the span of a hyperbolic polynomial to that of its first derivative.

Proposition 2.1. *Let p be an nth degree hyperbolic polynomial where $n \geqslant 2$, then $\sqrt{\frac{n-2}{n}}\operatorname{span}(p) \leqslant \operatorname{span}(p')$.*

Proof. We first note that $\operatorname{span}(p(x+c)) = \operatorname{span}(p(x))$. So without loss of generality, we may assume that $k = x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n = -k$ are the roots of p where $k = \frac{1}{2}\operatorname{span}(p)$. Let $w_1 \geqslant w_2 \geqslant \cdots \geqslant w_{n-1}$ be the roots of the derivative of p. We note that

$$(2k, 0, 0, \ldots, 0, 0, -2k) \prec (k, x_2, x_3, \ldots, x_{n-1}, -k) + (k, -x_{n-1}, -x_{n-2}, \ldots, -x_2, -k).$$

We now can use Theorem 1.4 and the fact that $(2k, 0, 0, \ldots, 0, 0, -2k)$ are the roots of the polynomial $(x^2 - 4k^2)x^{n-2}$ to obtain

$$\left(2\sqrt{\frac{n-2}{n}}k, 0, 0, \ldots, 0, 0, -2\sqrt{\frac{n-2}{n}}k\right) \prec (w_1, w_2, \ldots, w_{n-1}) + (-w_{n-1}, -w_{n-2}, \ldots, -w_1).$$

Hence $\operatorname{span}(p') = w_1 - w_{n-1} \geqslant \sqrt{\frac{n-2}{n}}\operatorname{span}(p)$. □

Definition 2.2. Let $x = (x_1, x_2, \ldots, x_n)$ be an n-tuple of real numbers listed in descending order, and let $k \leqslant n-2$. Then $\|x\|_{(k)} = \operatorname{span}(p^{(k)})$ where $p(x) = \prod_{i=1}^{n}(x - x_i)$.

We now justify our use of the norm symbol.

Theorem 2.3. *Let W be the one-dimensional vector space generated by $e = (1, 1, \ldots, 1)$ and $k \leqslant n-2$. Then $\|x\|_{(k)}$ is a norm on the quotient space $\mathbb{R}^n/W$.*

Proof. It is clear that $\|\cdot\|_{(k)}$ is a well-defined function on $\mathbb{R}^n/W$, since $\|x + \alpha e\|_{(k)} = \|x\|_{(k)}$ for all $\alpha \in \mathbb{R}$. It is also clear that $\|x\|_{(k)} \geqslant 0$, $\|0\|_{(k)} = 0$ and $\|cx\|_{(k)} = |c|\|x\|_{(k)}$. Let x, y be n-tuples. Then since $(x+y)_\downarrow \prec x_\downarrow + y_\downarrow$ ($x_\downarrow$ denotes the n-tuple formed by rearranging the entries of x in descending order.), Theorem 1.4 and Proposition 1.5 together imply the triangle inequality $\|x+y\|_{(k)} \leqslant \|x\|_{(k)} + \|y\|_{(k)}$. Finally, repeated use of Proposition 2.1 shows us that $\|x\|_{(k)} = 0$ implies that $x = 0$. $\square$

Corollary 2.4. *Let K be any compact convex subset of $\mathbb{R}^n/W$, then $\|x\|_{(k)}$ achieves its maximum on K at an extreme point of K.*

This corollary includes some known results about the spans of derivatives as special cases. If we take $K = [-1, 1]^n$ in the above corollary, we get the following result of Robinson [9].

Corollary 2.5. *Let $\mathcal{P}_n$ be the set of all nth degree polynomials having all of their roots in the interval $[-1, 1]$. Then any polynomial p which maximizes $\operatorname{span}(p^{(k)})$ over $\mathcal{P}_n$ for some $k \leqslant n-2$ must be of the form $(x-1)^a(x+1)^b$ where $a+b=n$.*

If we let K instead be the unit ball of l^1 norm, we get the following strengthening of a result of Meir and Sharma [5].

Corollary 2.6. *Let $\mathcal{P}_n^1$ be the set of all nth degree polynomials whose roots $x_1, x_2, \ldots, x_n$ are all real and satisfy the inequality $\sum_{i=1}^n |x_i| \leqslant 1$ and let $q(x) = x^{n-1}(x-1)$. Then $\operatorname{span}(p^{(k)}) \leqslant \operatorname{span}(q^{(k)})$ for all $p \in \mathcal{P}_n^1$.*

We note that we can use the method of proof of Theorem 2.3 to generate an entire class of permutation-invariant norms on $\mathbb{R}^n$.

Proposition 2.7. *Let $k \leqslant n-2$ and $\|\cdot\|$ be a permutation-invariant norm on $\mathbb{R}^{n-k}$. Let $\|\cdot\|'$ be defined as follows: $\|x\|' = \|x^{(k)}\|$ for all $x \in \mathbb{R}^n$. Then $\|\cdot\|'$ is a permutation-invariant norm on $\mathbb{R}^n$.*

3. Proof of the main theorem

We begin this paper by reviewing a key concept discussed in [2,7]. (Readers unfamiliar with some of the terminology may wish to refer to [7, Section 2].)

Definition 3.1. Let $\mathcal{H}$ be an n-dimensional Hilbert space, $A \in L(\mathcal{H})$ and P a projection from $\mathcal{H}$ onto a subspace of $\mathcal{H}$ having co-dimension one, set $B = PAP|_{P\mathcal{H}}$. Let $p_A(x) = \det(xI - A)$ and $p_B(x) = \det(xI - B)$. Then we shall say that P is a differentiator of A if $p_B(x) = \frac{1}{n}\frac{d}{dx}p_A(x)$.

In particular, the eigenvalues of B are the roots of the derivative of the characteristic polynomial of A. We can use [7, Theorem 2.5] to construct a differentiator.

Proposition 3.2. *Let $v = (1, 1, \ldots, 1)$ be the all ones vector in $\mathbb{R}^n$ and P be the orthogonal projection onto the orthogonal complement of the span of v. Then P is a differentiator for any n by n diagonal matrix.*

We also need the following result of Ky Fan [3]:

Proposition 3.3. *Let $\{A_i\}_{i=1}^k$ be a set of n by n Hermitian matrices and let $A = \sum_{i=1}^k A_i$. Let $\lambda(A)$ and $\lambda(A_i)$ be n-tuples whose roots are the eigenvalues of A and A_i listed in descending order. Then $\lambda(A) \prec \sum_{i=1}^k \lambda(A_i)$.*

Now we can prove Theorem 1.4. We only need to prove the $k = 1$ case. Let $x = (x_1, x_2, \dots, x_n)$, $y = (y_1, y_2, \dots, y_n)$, and $z = (z_1, z_2, \dots, z_n)$ be three n-tuples who entries are listed in descending order.

Let $X = \mathrm{diag}(x_1, x_2, \dots, x_n)$ and for each $\sigma \in S_n$, let $Y_\sigma = \mathrm{diag}(y_{\sigma(1)}, y_{\sigma(2)}, \dots, y_{\sigma(n)})$ and $Z_\sigma = \mathrm{diag}(z_{\sigma(1)}, z_{\sigma(2)}, \dots, z_{\sigma(n)})$.

Suppose $x \prec y + z$, then x is in the convex hull of $\{(y_{\sigma(1)} + z_{\sigma(1)}, y_{\sigma(2)} + z_{\sigma(2)}, \dots, y_{\sigma(n)} + z_{\sigma(n)})\}_{\sigma \in S_n}$ by Proposition 1.2. Which means there exists $\lambda_\sigma \geqslant 0$ with $\sum_{\sigma \in S_n} \lambda_\sigma = 1$ such that $X = \sum_{\sigma \in S_n} \lambda_\sigma (Y_\sigma + Z_\sigma)$. Now let P be as in Proposition 3.2. Then $PXP|_{P\mathcal{H}} = \sum_{\sigma \in S_n} \lambda_\sigma (PY_\sigma P|_{P\mathcal{H}} + PZ_\sigma P|_{P\mathcal{H}})$. Ky Fan's result now gives us $x^{(1)} \prec y^{(1)} + z^{(1)}$.

4. Linear combinations of interlacing polynomials

We begin this section by introducing the well-known concept of interlacing polynomials.

Definition 4.1. Let p, q be two hyperbolic polynomials and let $m = \deg(p) + \deg(q)$. Recall that $Z(pq) = (x_1, x_2, \dots, x_m)$ are the roots of pq listed in descending order. Then p and q are said to be interlacing if p is some non-zero multiple of one of $\prod_{i=1}^{\lfloor \frac{m}{2} \rfloor} (x - x_{2i})$ and $\prod_{i=1}^{\lfloor \frac{m}{2} \rfloor} (x - x_{2i+1})$.

In other words, two hyperbolic polynomials are interlacing iff their roots alternate. If p and q are interlacing and have no zeros in common, then p and q are said to be strictly interlacing. It is obvious that any two interlacing polynomials either have the same degree or have degrees which differ by one.

Given any two hyperbolic polynomials p and q, we would like to investigate any possible majorization inequalities between p, q and $p + q$. One immediate problem is that $p + q$ may have non-real roots. (Consider, for instance, $p(x) = x^3 - x$ and $q(x) = 2x$.) However, this never happens when p and q are interlacing as the following result shows.

Proposition 4.2. *Let p and q be two interlacing hyperbolic polynomials. Then $\alpha p + \beta q$ is a hyperbolic polynomial for all real α and β such that $\alpha^2 + \beta^2 \neq 0$.*

Proof. If p and q are strictly interlacing, this result is simply one direction of the Hermite–Kakeya theorem [8, Theorem 6.3.8]. By dividing out any common factors of p and q, we can always reduce the general case to the strictly interlacing case. $\square$

We will be working with companion matrices. The companion matrix of a polynomial $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ is the following n by n matrix:

$$C_p = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}.$$

The eigenvalues of a companion matrix are the roots of its associated polynomial. Companion matrices have an affine structure in the sense that if $\alpha + \beta = 1$ and $r(x) = \alpha p(x) + \beta q(x)$, then $C_r = \alpha C_p + \beta C_q$. Unfortunately, companion matrices are not Hermitian. We can get around this using the following generalization of a result of Lax.

Proposition 4.3 [4, Theorem 9.G.3]. *Let A and B be two matrices such that $\alpha A + \beta B$ has no non-real eigenvalues for all $\alpha, \beta \in \mathbb{R}$. Let $\lambda(A)$, $\lambda(B)$ and $\lambda(A+B)$ be the eigenvalues of A, B and $A+B$ respectively listed in descending order. Then $\lambda(A+B) \prec \lambda(A) + \lambda(B)$.*

Now let p and q be two interlacing nth degree monic hyperbolic polynomials. Let α and β be two real constants which sum to one. By setting $A = \alpha C_p$ and $B = \beta C_q$ in the previous proposition, we obtain the following result:

Theorem 4.4. *Let p and q be two interlacing nth degree monic hyperbolic polynomials. Let $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1$ and let $r(x) = \alpha p(x) + \beta q(x)$, then $Z(r) \prec (\alpha Z(p))_\downarrow + (\beta Z(q))_\downarrow$.*

The $\downarrow$s signify that the n-tuples are written in descending order even after multiplication by a possibly negative constant.

So, for instance, with α, β, p, q, r as in previous theorem, we have $\operatorname{span}(r) \leqslant |\alpha| \operatorname{span}(p) + |\beta| \operatorname{span}(q)$.

We may also consider the case of two interlacing polynomials whose degrees differ by one. In what follows, let $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ and $q(x) = x^{n-1} + b_{n-1}x^{n-1} + \cdots + b_1x + b_0$ be two interlacing hyperbolic polynomials. The following matrix will also be useful:

$$M = \begin{pmatrix} 0 & 0 & \cdots & 0 & -b_0 \\ 0 & 0 & \cdots & 0 & -b_1 \\ 0 & 0 & \cdots & 0 & -b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -b_{n-1} \\ 0 & 0 & \cdots & 0 & -1 \end{pmatrix}.$$

Let α be an arbitrary real number and $s(x) = p(x) - \alpha q(x)$, then $C_s = C_p - \alpha M$. Using the same reasoning as before, we get the following result.

Theorem 4.5. *Let p and q be two interlacing monic hyperbolic polynomials with* $\deg(p) = \deg(q) + 1$. *Let α be an arbitrary real number and $s(x) = p(x) - \alpha q(x)$, then $Z(s) \prec Z(p) + (\max(\alpha, 0), 0, 0, \ldots, 0, 0, \min(\alpha, 0))$.*

References

[1] J. Borcea, B. Shapiro, Hyperbolic polynomials and spectral order, C. R. Acad. Sci. Paris Ser. I 337 (2003) 693–698.

[2] C. Davis, Eigenvalues of compressions, Bull. Math. de la Soc. Sci. Math. Phys. de la R.P.R. 51 (1959) 3–5.

[3] K. Fan, On a theorem of Weyl concerning eigenvalues of linear transformations I, Proc. Nat. Acad. Sci. USA 35 (1949) 652–655.

[4] A.W. Marshall, I. Olkin, Inequalities: Theory of Majorization and its Applications, Academic Press, New York, 1979.

[5] A. Meir, A. Sharma, Span of derivatives of polynomials, Amer. Math. Monthly 74 (1967) 527–531.

[6] A. Meir, A. Sharma, On the zeros of derivatives of polynomials, Canad. Math. Bull. 11 (1968) 443–445.

[7] R. Pereira, Differentiators and the geometry of polynomials, J. Math. Anal. Appl. 285 (2003) 336–348.

[8] Q.I. Rahman, G. Schmeisser, Analytic Theory of Polynomials, Oxford Univ. Press, Oxford, 2002.

[9] R.M. Robinson, On the spans of derivatives of polynomials, Amer. Math. Monthly 71 (1964) 504–508.

[10] J. Sz.-Nagy, Uber algebraische Gleichungen mit lauter reellen Wurzeln, Jahresber. Deutsch. Math.-Verein. 27 (1918) 37–43.