From Banach spaces to moments to positive polynomials

Bruce Reznick
University of Illinois at Urbana-Champaign

Workshop on Inverse Moment Problems
IMS/Singapore, December 2, 2013

This talk is dedicated with great affection to

This talk is dedicated with great affection to

Moment Problems

This talk is dedicated with great affection to

Moment Problems

which have enriched my understanding of mathematics tremendously over the years.

This talk is dedicated with great affection to

Moment Problems

which have enriched my understanding of mathematics tremendously over the years.

And thanks also to the

This talk is dedicated with great affection to

Moment Problems

which have enriched my understanding of mathematics tremendously over the years.

And thanks also to the

Organizers

This talk is dedicated with great affection to

Moment Problems

which have enriched my understanding of mathematics tremendously over the years.

And thanks also to the

Organizers

and the supporting organizations for allowing me an opportunity to travel 38% of the way around the world to express this gratitude.

The story begins with Math 143, Functional Analysis, taught by Prof. W.A.J. Luxemburg at Caltech in 1972-1973, during my senior year of college. We learned about the Jordan-von Neumann Theorem in this class, and I was immediately smitten:

The story begins with Math 143, Functional Analysis, taught by Prof. W.A.J. Luxemburg at Caltech in 1972-1973, during my senior year of college. We learned about the Jordan-von Neumann Theorem in this class, and I was immediately smitten:

Theorem (Jordan-von Neumann)

Suppose X is a Banach space, and the parallelogram law holds for all $x, y \in X$:

$$||x+y||^2 + ||x-y||^2 = 2(||x||^2 + ||y||^2).$$

Then X is a Hilbert space.

Sticking to the real case, one proof is to define a prospective inner product by $(x, y) = \frac{1}{2}(||x+y||^2 - ||x||^2 - ||y||^2)$ and then establish that it has all the properties you want from an inner product. This is done by manipulating instances of the parallelogram law.

Another approach is to rewrite the parallelogram law as a second difference equation:

$$||x+y||^2 - 2||x||^2 + ||x-y||^2 = 2||y||^2$$

Another approach is to rewrite the parallelogram law as a second difference equation:

$$||x+y||^2 - 2||x||^2 + ||x-y||^2 = 2||y||^2$$
$$\implies ||x+3y||^2 - 3||x+2y||^2 + 3||x+y||^2 - ||x||^2 = 0$$

Another approach is to rewrite the parallelogram law as a second difference equation:

$$||x+y||^2 - 2||x||^2 + ||x-y||^2 = 2||y||^2$$
$$\implies ||x+3y||^2 - 3||x+2y||^2 + 3||x+y||^2 - ||x||^2 = 0$$

It is not hard to show that if this holds for all elements $x, y \in X$, then for all x, y, $||x+ty||^2 = A(x,y) + 2B(x,y)t + C(x,y)t^2$; it is immediate that $A(x,y) = ||x||^2, C(x,y) = ||y||^2$ and the task is to verify that $B(x,y)$ is an inner product. Again this can be done by formal manipulations.

Another approach is to rewrite the parallelogram law as a second difference equation:

$$||x+y||^2 - 2||x||^2 + ||x-y||^2 = 2||y||^2$$
$$\implies ||x+3y||^2 - 3||x+2y||^2 + 3||x+y||^2 - ||x||^2 = 0$$

It is not hard to show that if this holds for all elements $x, y \in X$, then for all x, y, $||x+ty||^2 = A(x,y) + 2B(x,y)t + C(x,y)t^2$; it is immediate that $A(x,y) = ||x||^2, C(x,y) = ||y||^2$ and the task is to verify that $B(x,y)$ is an inner product. Again this can be done by formal manipulations.

Attempts to generalize this result became my PhD thesis in Banach spaces. To be honest, there wasn't much functional analysis in it.

Theorem (Thesis 76; *Pac.J.Math.* 78,79)

Suppose X is a Banach space, and there exist constants $a_k, c_k(j) \in \mathbb{C}$ so that for all $x_j \in X$:

$$\sum_{k=1}^{n} a_k ||c_k(0)x_0 + \cdots + c_k(n)x_n||^p = 0.$$

Then $p = 2m$ is an even integer and for all $x, y \in X$ and real t, $||x + ty||^{2m}$ is a polynomial in t of degree p.
Conversely, if $||x + ty||^{2m}$ is a polynomial in t for all $x, y \in X$, then the identity holds iff it holds over $\mathbb{C}$; i.e, for $u_j \in \mathbb{C}$:

$$\sum_{k=1}^{n} a_k |c_k(0)u_0 + \cdots + c_k(n)u_n|^{2m} = 0.$$

The proof relies on repeatedly substituting into the identity and transforming it into an n-th difference equation (Willson, 1914).

Various other generalizations of the Jordan-von Neumann Theorem already existed, usually with $p = 2$. The next task was to come up with examples. I'll skip the identities themselves, which can be found in the indicated papers. What about spaces?

Various other generalizations of the Jordan-von Neumann Theorem already existed, usually with $p = 2$. The next task was to come up with examples. I'll skip the identities themselves, which can be found in the indicated papers. What about spaces?
Let $\mathcal{P}_{2m}$ denote the set of Banach spaces for which $||x + ty||^{2m}$ is a polynomial in t. Suppose $p = 2m$ is even; consider $L_{2m}(X, \mu)$. Then

$$||f+tg||^{2m} = \int_X |f+tg|^{2m} d\mu = \int_X (|f|^2+t(f\bar{g}+\bar{f}g)+t^2||g||^2)^m d\mu.$$

Thus, $L_p(X, \mu) \in \mathcal{P}_{2m}$. Are these the only ones (up to isometry)?

Various other generalizations of the Jordan-von Neumann Theorem already existed, usually with $p = 2$. The next task was to come up with examples. I'll skip the identities themselves, which can be found in the indicated papers. What about spaces?
Let $\mathcal{P}_{2m}$ denote the set of Banach spaces for which $||x + ty||^{2m}$ is a polynomial in t. Suppose $p = 2m$ is even; consider $L_{2m}(X, \mu)$. Then

$$||f+tg||^{2m} = \int_X |f+tg|^{2m} d\mu = \int_X (|f|^2 + t(f\bar{g} + \bar{f}g) + t^2||g||^2)^m d\mu.$$

Thus, $L_p(X, \mu) \in \mathcal{P}_{2m}$. Are these the only ones (up to isometry)? In order to answer this question, we need to talk about the possible polynomials $p(t) = ||f + tg||^{2m}$. Every complex $L_{2m}(X, \mu)$ can be embedded in $m + 1$ copies of a real $L_{2m}(X, \mu)$, so it will suffice to restrict our attention to the real case.

Various other generalizations of the Jordan-von Neumann Theorem already existed, usually with $p = 2$. The next task was to come up with examples. I'll skip the identities themselves, which can be found in the indicated papers. What about spaces?
Let $\mathcal{P}_{2m}$ denote the set of Banach spaces for which $||x + ty||^{2m}$ is a polynomial in t. Suppose $p = 2m$ is even; consider $L_{2m}(X, \mu)$.
Then

$$||f+tg||^{2m} = \int_X |f+tg|^{2m} d\mu = \int_X (|f|^2+t(f\bar{g}+\bar{f}g)+t^2||g||^2)^m d\mu.$$

Thus, $L_p(X, \mu) \in \mathcal{P}_{2m}$. Are these the only ones (up to isometry)? In order to answer this question, we need to talk about the possible polynomials $p(t) = ||f + tg||^{2m}$. Every complex $L_{2m}(X, \mu)$ can be embedded in $m + 1$ copies of a real $L_{2m}(X, \mu)$, so it will suffice to restrict our attention to the real case.
The following is not hard, and generalizations to more variables.

Theorem

Suppose $p \in \mathbb{R}[t]$ *is given and a proposed norm is defined on a vector space* $X = span(x, y)$ *by* $||x + ty||^{2m} = p(t)$. *Then* X *is a Banach space if and only if* $p(t) \geq 0$ *and* $p^{\frac{1}{2m}}$ *is a convex function of* t.

Theorem

Suppose $p \in \mathbb{R}[t]$ is given and a proposed norm is defined on a vector space $X = span(x, y)$ by $||x + ty||^{2m} = p(t)$. Then X is a Banach space if and only if $p(t) \geq 0$ and $p^{\frac{1}{2m}}$ is a convex function of t.

Theorem

Suppose $p \in \mathbb{R}[t]$ is given and a proposed norm is defined on a vector space $X = span(x, y)$ by $||x + ty||^{2m} = p(t)$. Then X is a Banach space if and only if $p(t) \geq 0$ and $p^{\frac{1}{2m}}$ is a convex function of t.

The first theorem is that any two dimensional space in $\mathcal{P}_4$ is an L_4 space. I will explain the attribution later.

Theorem (& Dmitriev)

If $\deg p = 4$, $p(t) \geq 0$ and $p^{1/4}$ is convex, then there exist real constants so that

$$p(t) = (b_1 + c_1 t)^4 + (b_2 + c_2 t)^4 + b_3^4.$$

If X is a space with three atoms of weight 1, and the pair (f, g) takes the values $(b_1, c_1), (b_2, c_2), (b_3, 0)$ at these atoms, then $||f + tg||^4 = p(t)$. A 2-dimensional $X \in \mathcal{P}_4$ is embeddable in L_4.

If $2m \geq 6$, then $(t^{2m} + t^2 + 1)^{1/2m}$ is convex, but

$$t^{2m} + t^2 + 1 = ||f + tg||^{2m} = \int_X (f + tg)^{2m} d\mu \implies$$

$$0 = \int_X f^{2m-4} g^4 d\mu, \qquad 1 = \binom{2m}{2} \int_X f^{2m-2} g^2 d\mu$$

Since $2m - 4 > 0$, the 1st equation implies that $fg = 0$ μ-ae, which contradicts the 2nd: the space is not embeddable in any L_{2m}.

If $2m \geq 6$, then $(t^{2m} + t^2 + 1)^{1/2m}$ is convex, but

$$t^{2m} + t^2 + 1 = ||f + tg||^{2m} = \int_X (f + tg)^{2m} d\mu \implies$$

$$0 = \int_X f^{2m-4} g^4 d\mu, \qquad 1 = \binom{2m}{2} \int_X f^{2m-2} g^2 d\mu$$

Since $2m - 4 > 0$, the 1st equation implies that $fg = 0$ μ-ae, which contradicts the 2nd: the space is not embeddable in any L_{2m}.
Also, $||x + ty + uz||^4 = 1 + 6(t^2 + u^2) + (t^2 + u^2)^2$ defines a perfectly fine Banach space X, but

$$1 + 6(t^2 + u^2) + (t^2 + u^2)^2 = \int_X (f + tg + uh)^4 d\mu \implies$$

$$\int f^4 = \int g^4 = \int h^4 = \int f^2 g^2 = \int f^2 h^2 = 1, \int g^2 h^2 = \frac{1}{3}.$$

Cauchy-Schwartz and the first five identities imply that $f^2 = g^2 = h^2$ μ-ae, which contradicts the last. Alternatively, we observe that $\int (f^2 - g^2 - h^2)^2 d\mu = 1 + 1 + 1 - 2(1) - 2(1) + 2(\frac{1}{3}) = -\frac{1}{3}$. So X is not embeddable in L_4 but every proper subspace of X is.

These examples appeared in my thesis, without reference to the moment problem, but while writing the work up for publication, my ignorance became clear. J. H. B. Kemperman was a very helpful correspondent. Later, I inhaled the relevant chapters of Akhieser and Krein and "borrowed" their ideas wherever possible.

These examples appeared in my thesis, without reference to the moment problem, but while writing the work up for publication, my ignorance became clear. J. H. B. Kemperman was a very helpful correspondent. Later, I inhaled the relevant chapters of Akhieser and Krein and "borrowed" their ideas wherever possible. We have (leaving aside the set where $f = 0$ and using Hölder to justify the convergence of the integrals),

$$p(t) = \sum_{k=0}^{2m} \binom{2m}{k} a_k t^k = \int_X (f + tg)^{2m} d\mu \iff$$

$$a_k = \int_X f^{2m-k} g^k d\mu = \int_X h^k d\nu,$$

where $h = g/f$ and $d\nu = f^{2m} d\mu$. Let $\Phi(r) = \nu(\{-\infty, r))$, then

$$a_k = \int_{-\infty}^{\infty} s^k d\Phi, \qquad a_{2m} \geq \int_{-\infty}^{\infty} s^{2m} d\Phi,$$

which is precisely the classical form of the truncated Hamburger moment problem.

We see that p comes from an L_{2m} space precisely if the Hankel matrix of its coefficients is psd. It is not hard to use Riemann approximation, Carathéodory's Theorem and Bolzano-Weierstrass to show that any solution can be realized as a finite set of point masses, so the generating functions of moment sequences are precisely the sums of powers of linear forms.

We see that p comes from an L_{2m} space precisely if the Hankel matrix of its coefficients is psd. It is not hard to use Riemann approximation, Carathéodory's Theorem and Bolzano-Weierstrass to show that any solution can be realized as a finite set of point masses, so the generating functions of moment sequences are precisely the sums of powers of linear forms.

It is also a natural question to now look at the polynomials p of degree $2m$ for which $p^{\frac{1}{2m}}$ is convex. They are in a category of cones I called "blenders" and wrote about in a collection which Mihai co-edited in memory of Julius Borcea.

We see that p comes from an L_{2m} space precisely if the Hankel matrix of its coefficients is psd. It is not hard to use Riemann approximation, Carathéodory's Theorem and Bolzano-Weierstrass to show that any solution can be realized as a finite set of point masses, so the generating functions of moment sequences are precisely the sums of powers of linear forms.
It is also a natural question to now look at the polynomials p of degree $2m$ for which $p^{\frac{1}{2m}}$ is convex. They are in a category of cones I called "blenders" and wrote about in a collection which Mihai co-edited in memory of Julius Borcea.
One of the weirdest results in my thesis was that for degree 6, the extremal sextics for which $p^{\frac{1}{6}}$ is convex are the sixth powers and polynomials derived from

$$\phi_\lambda(t) = 1 + 6\lambda t + 15\lambda^2 t^2 + 20\lambda^3 t^3 + 15\lambda^2 t^4 + 6\lambda t^5 + t^6,$$

for $0 < |\lambda| \le \frac{1}{2}$. (That is, ϕ_λ behaves like $(1+\lambda t)^6$ near 0 and $(t+\lambda)^6$ near infinity.)

In going through MathSciNet to write the blenders paper, I was astonished to find that this result (and the one on quartics) had been published a few years before my work, though fortunately for me, it wasn't reviewed in MathSciNet until after my thesis was safely deposited (and then in [52] which made it hard to find.)

In going through MathSciNet to write the blenders paper, I was astonished to find that this result (and the one on quartics) had been published a few years before my work, though fortunately for me, it wasn't reviewed in MathSciNet until after my thesis was safely deposited (and then in [52] which made it hard to find.)

MR0467523 (57 #7379) Dmitriev, V. I., The structure of a cone in a five-dimensional space. (Russian) Voronez. Gos. Univ. Trudy Naucn.-Issled. Inst. Mat. VGU Vyp. 7 (1973). (Reviewer: Ju. Saskin (RZMat 1974 #8 B618)

In going through MathSciNet to write the blenders paper, I was astonished to find that this result (and the one on quartics) had been published a few years before my work, though fortunately for me, it wasn't reviewed in MathSciNet until after my thesis was safely deposited (and then in [52] which made it hard to find.)

MR0467523 (57 #7379) Dmitriev, V. I., The structure of a cone in a five-dimensional space. (Russian) Voronez. Gos. Univ. Trudy Naucn.-Issled. Inst. Mat. VGU Vyp. 7 (1973). (Reviewer: Ju. Saskin (RZMat 1974 #8 B618)

V. I. Dmitriev, a student of Selim Krein (Mark Krein's brother) at Kharkov University, published articles on this topic in 1973 and 1991 (which our librarians found for me in Russian). There are two V. I. Dmitirievs in MathSciNet and this one seems to be at Kursk State Technical University. In the 1991 paper, he wrote "I am not aware of any other articles on this topic, except" his earlier one. I made unsuccessful efforts to contact him, both by postal mail and email, the latter via Peter Kuchment, another student of S. Krein.

Moment theory leads to some nice inequalities, especially if we think mainly of point masses. For example, the non-negativity of the determinant of the 2×2 Hankel matrix gives ...

Moment theory leads to some nice inequalities, especially if we think mainly of point masses. For example, the non-negativity of the determinant of the 2×2 Hankel matrix gives ...

Theorem (Cauchy-Schwarz)

$$0 \le \begin{vmatrix} \sum a_k^2 & \sum a_k b_k \\ \sum a_k b_k & \sum b_k^2 \end{vmatrix} = \sum_{i<j} (a_i b_j - a_j b_i)^2$$

Moment theory leads to some nice inequalities, especially if we think mainly of point masses. For example, the non-negativity of the determinant of the 2×2 Hankel matrix gives ...

Theorem (Cauchy-Schwarz)

$$0 \le \begin{vmatrix} \sum a_k^2 & \sum a_k b_k \\ \sum a_k b_k & \sum b_k^2 \end{vmatrix} = \sum_{i<j} (a_i b_j - a_j b_i)^2$$

Moment theory leads to some nice inequalities, especially if we think mainly of point masses. For example, the non-negativity of the determinant of the 2×2 Hankel matrix gives ...

Theorem (Cauchy-Schwarz)

$$0 \le \begin{vmatrix} \sum a_k^2 & \sum a_k b_k \\ \sum a_k b_k & \sum b_k^2 \end{vmatrix} = \sum_{i<j} (a_i b_j - a_j b_i)^2$$

This continues in higher degree, but the results don't seem to be as well-known.

Theorem

$$0 \le \begin{vmatrix} \sum a_k^4 & \sum a_k^3 b_k & \sum a_k^2 b_k^2 \\ \sum a_k^3 b_k & \sum a_k^2 b_k^2 & \sum a_k b_k^3 \\ \sum a_k^2 b_k^2 & \sum a_k b_k^3 & \sum b_k^4 \end{vmatrix}$$

$$= \sum_{i<j<k} (a_i b_j - a_j b_i)^2 (a_i b_k - a_k b_i)^2 (a_j b_k - a_k b_j)^2.$$

The moment characterization is also useful in proving inequalities for products of power sums. Many classical inequalities follow from convexity and don't seem to be effective in finding lower bounds of products. Power sums of real numbers represent moments with measures containing a large number of unit point masses.

Theorem (Pac.J.Math.83)

For real x_i,

$$\frac{(\sum_{j=1}^{n} x_j)(\sum_{j=1}^{n} x_j^3)}{n(\sum_{j=1}^{n} x_j^4)} > -\frac{1}{8}$$

where the constant $-\frac{1}{8}$ is best-possible and never achieved.

The moment characterization is also useful in proving inequalities for products of power sums. Many classical inequalities follow from convexity and don't seem to be effective in finding lower bounds of products. Power sums of real numbers represent moments with measures containing a large number of unit point masses.

Theorem (Pac.J.Math.83)

For real x_i,

$$\frac{(\sum_{j=1}^n x_j)(\sum_{j=1}^n x_j^3)}{n(\sum_{j=1}^n x_j^4)} > -\frac{1}{8}$$

where the constant $-\frac{1}{8}$ is best-possible and never achieved.

The moment characterization is also useful in proving inequalities for products of power sums. Many classical inequalities follow from convexity and don't seem to be effective in finding lower bounds of products. Power sums of real numbers represent moments with measures containing a large number of unit point masses.

Theorem (Pac.J.Math.83)

For real x_i,

$$\frac{(\sum_{j=1}^{n} x_j)(\sum_{j=1}^{n} x_j^3)}{n(\sum_{j=1}^{n} x_j^4)} > -\frac{1}{8}$$

where the constant $-\frac{1}{8}$ is best-possible and never achieved.

Sketch of proof: Normalize to $a_0 = a_4 = 1$. It is a calculus exercise (optimization!) to show that

$$\begin{vmatrix} 1 & a_1 & a_2 \\ a_1 & a_2 & a_3 \\ a_2 & a_3 & 1 \end{vmatrix} \ge 0 \implies a_1 a_3 \ge -\frac{1}{8}$$

Up to scaling, equality holds uniquely if $a_0 = a_4 = 1$, $a_1 = \frac{1}{\sqrt{8}}$, $a_3 = -\frac{1}{\sqrt{8}}$ and $a_2 = \frac{1}{2}$. An extremal example occurs only if μ is a measure with two atoms of measure in ratio $(2+\sqrt{3})^2 \notin \mathbb{Q}$. Thus, it can only be approximated by unit point masses but never achieved.

Up to scaling, equality holds uniquely if $a_0 = a_4 = 1$, $a_1 = \frac{1}{\sqrt{8}}$, $a_3 = -\frac{1}{\sqrt{8}}$ and $a_2 = \frac{1}{2}$. An extremal example occurs only if μ is a measure with two atoms of measure in ratio $(2+\sqrt{3})^2 \notin \mathbb{Q}$. Thus, it can only be approximated by unit point masses but never achieved.

Different methods in the same paper show that

$$\max\left(\frac{(\sum_{j=1}^n x_j)(\sum_{j=1}^n x_j^3)}{(\sum_{j=1}^n x_j^2)^2}\right) = \frac{3\sqrt{3}}{16}n^{3/2} + \frac{5}{8} + \mathcal{O}(n^{-1/2})$$

$$\min\left(\frac{(\sum_{j=1}^n x_j)(\sum_{j=1}^n x_j^3)}{(\sum_{j=1}^n x_j^2)^2}\right) = -\frac{3\sqrt{3}}{16}n^{3/2} + \frac{5}{8} + \mathcal{O}(n^{-1/2})$$

Moment methods are not useful here, because $\frac{a_1 a_3}{a_2^2}$ is unbounded.

After taking a job, and beginning to write my thesis up for publication, I read somewhere that the reason the multidimensional moment problem was hard was that there were non-negative polynomials in several variables which were not sums of squares, citing Robinson's example. *But where did I see it? Any ideas?* There wasn't much literature. Motzkin's article was available in a conference proceedings. I wrote R. M. Robinson, and he sent me a reprint of his hard-to-find article. He also told me I should write his colleague T. Y. Lam. I did, and that week changed my life in several wonderful ways.

After taking a job, and beginning to write my thesis up for publication, I read somewhere that the reason the multidimensional moment problem was hard was that there were non-negative polynomials in several variables which were not sums of squares, citing Robinson's example. *But where did I see it? Any ideas?* There wasn't much literature. Motzkin's article was available in a conference proceedings. I wrote R. M. Robinson, and he sent me a reprint of his hard-to-find article. He also told me I should write his colleague T. Y. Lam. I did, and that week changed my life in several wonderful ways.

M. D. Choi and T. Y. Lam defined the cones $P_{n,2m}$ and $\Sigma_{n,2m}$ of psd forms of degree $2m$ in n variables and sums of squares. They proved that these were closed convex cones. Hilbert's 1888 proof that there exist forms in $P_{3,6} \setminus \Sigma_{3,6}$ needs that $\Sigma_{3,6}$ is closed. The proof specifically used the argument that any finite sum of squares of ternary cubics is a sum of $28 = \binom{6+3-1}{6}$ squares of ternary cubics. This is "Carathéodory's Theorem". Carathéodory, a future PhD student at Göttingen, was 15 in 1888.

It is natural when looking at cones to consider their duals, but one needs an inner product. The moment problem provided an obvious candidate. I defined the cone $Q_{n,2m}$ and "proved" that $P_{n,2m}$ and $Q_{n,2m}$ are dual under the Fischer inner product, but this fact had been staring me in the face from my reading in Akhieser-Krein and elsewhere, which needed only trivial modifications: replace the given moments with their generating polynomial.

It is natural when looking at cones to consider their duals, but one needs an inner product. The moment problem provided an obvious candidate. I defined the cone $Q_{n,2m}$ and "proved" that $P_{n,2m}$ and $Q_{n,2m}$ are dual under the Fischer inner product, but this fact had been staring me in the face from my reading in Akhieser-Krein and elsewhere, which needed only trivial modifications: replace the given moments with their generating polynomial.

Let $\mathcal{I}(n,d) = \{(i_1,\dots,i_n) : 0 \le i_k \in \mathbb{Z}, \sum i_k = d\}$ and for $i \in \mathcal{I}(n,d)$ write the multinomial coefficient $c(i) = \frac{d!}{i_1!\dots i_n!}$. Write $x^i = x_1^{i_1}\dots x_n^{i_n}$ and for a real form p of degree d in n variables, scale the coefficients by:

$$p = \sum_{i\in\mathcal{I}(n,d)} c(i)a(p;i)x^i$$

Then the *Fischer inner product* is defined by

$$[p,q] = \sum_{i\in\mathcal{I}(n,d)} c(i)a(p;i)a(q;i).$$

If q is the generating function of (X, μ), then $[p, q] = \int p d\mu$. This inner product has a huge number of algebraic properties as well. If you think about point masses and for $\alpha \in \mathbb{R}^n$ define $(\alpha \cdot)^d = (\alpha \cdot x)^d$, then it is easy to see that

$$[p, (\alpha \cdot)^d] = p(\alpha).$$

Define the d-th order differential operator $q(D)$ in the obvious multinomial way:

$$q(D) = \sum_{i \in \mathcal{I}(n,e)} c(i) a(q; i) \left(\frac{\partial}{\partial x_1} \right)^{i_1} \cdots \left(\frac{\partial}{\partial x_n} \right)^{i_n}.$$

For $i, j \in \mathcal{I}(n, d)$, it is easy to see that $D^i(x^j) = 0$ if $i \neq j$ and

$$D^i(x^i) = \prod_{i=1}^{n} i_k!.$$

It follows that $q(D)p = d![p, q]$. The advantage of this definition is that it is meaningful when p and q have different degrees.

Classically, p is apolar to q if $q(D)p = 0$. The inner product satisfies these properties:

- If $f \in H_e(\mathbb{C}^n)$ and $g \in H_{d-e}(\mathbb{C}^n)$, then

$$d![fg, p] = (fg)(D)p = f(D)g(D)p = e![f, g(D)p]$$

Classically, p is apolar to q if $q(D)p = 0$. The inner product satisfies these properties:

- If $f \in H_e(\mathbb{C}^n)$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$d![fg, p] = (fg)(D)p = f(D)g(D)p = e![f, g(D)p]$$
- If $e \leq d$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$g(D)(\alpha\cdot)^d = \frac{d!}{e!} g(\alpha)(\alpha\cdot)^e.$$

Classically, p is apolar to q if $q(D)p = 0$. The inner product satisfies these properties:

- If $f \in H_e(\mathbb{C}^n)$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$d![fg, p] = (fg)(D)p = f(D)g(D)p = e![f, g(D)p]$$
- If $e \leq d$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$g(D)(\alpha\cdot)^d = \frac{d!}{e!} g(\alpha)(\alpha\cdot)^e.$$
- p is apolar to $(\alpha\cdot)^d$ iff $p(\alpha) = 0$.

Classically, p is apolar to q if $q(D)p = 0$. The inner product satisfies these properties:

- If $f \in H_e(\mathbb{C}^n)$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$d![fg, p] = (fg)(D)p = f(D)g(D)p = e![f, g(D)p]$$
- If $e \le d$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$g(D)(\alpha\cdot)^d = \frac{d!}{e!}g(\alpha)(\alpha\cdot)^e.$$
- p is apolar to $(\alpha\cdot)^d$ iff $p(\alpha) = 0$.
- p is apolar to $g(x)(\alpha\cdot)^{d-e}$ for every $g \in H_e(\mathbb{C}^n)$ iff all e-th order derivatives of p vanish at α iff p vanishes to e-th order at α.

Classically, p is apolar to q if $q(D)p = 0$. The inner product satisfies these properties:

- If $f \in H_e(\mathbb{C}^n)$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$d![fg, p] = (fg)(D)p = f(D)g(D)p = e![f, g(D)p]$$
- If $e \leq d$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$g(D)(\alpha\cdot)^d = \frac{d!}{e!} g(\alpha)(\alpha\cdot)^e.$$
- p is apolar to $(\alpha\cdot)^d$ iff $p(\alpha) = 0$.
- p is apolar to $g(x)(\alpha\cdot)^{d-e}$ for every $g \in H_e(\mathbb{C}^n)$ iff all e-th order derivatives of p vanish at α iff p vanishes to e-th order at α.
- If $\deg q \leq \deg p$ and $q(D)p = 0$, then all multiples of q in $H_d(\mathbb{C}^n)$ are apolar to p.

Classically, p is apolar to q if $q(D)p = 0$. The inner product satisfies these properties:

- If $f \in H_e(\mathbb{C}^n)$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$d![fg, p] = (fg)(D)p = f(D)g(D)p = e![f, g(D)p]$$
- If $e \le d$ and $g \in H_{d-e}(\mathbb{C}^n)$, then
$$g(D)(\alpha\cdot)^d = \frac{d!}{e!} g(\alpha)(\alpha\cdot)^e.$$
- p is apolar to $(\alpha\cdot)^d$ iff $p(\alpha) = 0$.
- p is apolar to $g(x)(\alpha\cdot)^{d-e}$ for every $g \in H_e(\mathbb{C}^n)$ iff all e-th order derivatives of p vanish at α iff p vanishes to e-th order at α.
- If $\deg q \le \deg p$ and $q(D)p = 0$, then all multiples of q in $H_d(\mathbb{C}^n)$ are apolar to p.
- If q is irreducible and $\deg p = d$, then $q(D)p = 0$ iff $p = \sum(\alpha_k\cdot)^d$, where $\alpha_k \in \{q = 0\}$. (The "Fundamental Theorem of Apolarity"; known pre-Hilbert, but needs Nullstellensatz for a rigorous proof.)

We have that $\Sigma_{n,m} \subsetneq P_{n,m}$ unless $n = 2$, $m = 2$ or $(n, m) = (3, 4)$. It follows that $Q_{n,2m} = P^*_{n,m} \subsetneq \Sigma^*_{n,2m}$ in those cases. But what is $\Sigma^*_{n,2m}$. Again, this is almost trivial: $p \in \Sigma^*_{n,2m}$ iff $[p, h^2] \geq 0$ for all squares of forms h of degree m.

$$h(x) = \sum_{\ell \in \mathcal{I}(n,m)} t(\ell) x^\ell \implies [p, h^2] = \sum_{\ell, \ell'} a(p; \ell + \ell') t(\ell) t(\ell'),$$

Holy Hankel! Sylvester had already defined this quadratic form for ternary quartics and called it the catalecticant.
An even symmetric sextic can be written as

$$p(x_1, \dots, x_n) = a \left(\sum x_i^2 \right)^3 + b \left(\sum x_i^2 \right) \left(\sum x_i^4 \right) + c \left(\sum x_i^6 \right)$$

Let $\hat{p}(t) = at^3 + bt^2 + ct$; if x_k has k 1's and $n - k$ 0's, then $p(x_k) = \hat{p}(k)$.

Theorem (CLR,87)

- $p \in P_{n,6} \iff \hat{p}(t) \geq 0$ *for* $t \in \{1, 2, \ldots, n\}$
- $p \in \Sigma_{n,6} \iff \hat{p}(t) \geq 0$ *for* $t \in \{1\} \cup [2, n]$

For example, the Robinson form $R \in P_{3,6} \setminus \Sigma_{3,6}$ has $\hat{R}(t) = \frac{1}{2}(t-2)(t-3)$. With a bit more work, we can use this result to calculate the even symmetric sextics in the dual cone.

Theorem (Memoir,92)

- $(x^2 + y^2 + z^2) - \lambda(x^6 + y^6 + z^6) \in P^*_{3,6} = Q_{3,6} \iff \lambda \leq \frac{2}{3}$.
- $(x^2 + y^2 + z^2) - \lambda(x^6 + y^6 + z^6) \in \Sigma^*_{3,6} \iff \lambda \leq \frac{7}{10}$.

Here are a few auto-plagiarized slides on Sylvester's theorem about the expression of binary forms as sums of powers of linear forms. This is the one dimensional moment problem with complex weights and measures were allowed. Nevertheless, if the Sylvester method is used carefully, one can solve moment problems in one variable on the real line, but with negative weights not ruled out, *a priori*.

In one very special family of case, we are able to get quite detailed information about the solutions of moment problems with discrete weights over $\mathbb{C}$. This ties in with a beautiful subject in combinatorics called "spherical designs".

Theorem (Sylvester, 1851)

Suppose $p(x,y) = \sum_{j=0}^{d} \binom{d}{j} a_j x^{d-j} y^j \in F[x,y] \subset \mathbb{C}[x,y]$ *and* $h(x,y) = \sum_{t=0}^{r} c_t x^{r-t} y^t = \prod_{j=1}^{r} (\beta_j x - \alpha_j y)$ *is a product of pairwise distinct linear factors,* $\alpha_j, \beta_j \in F$. *Then there exist* $\lambda_k \in F$ *so that*

$$p(x,y) = \sum_{k=1}^{r} \lambda_k (\alpha_k x + \beta_k y)^d$$

if and only if

$$\begin{pmatrix} a_0 & a_1 & \cdots & a_r \\ a_1 & a_2 & \cdots & a_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d-r} & a_{d-r+1} & \cdots & a_d \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Some notes on the proof:

Some notes on the proof:

- This is an algorithm! Given p, for increasing r, write the coefficients of p in the Hankel matrix, and look for null vectors c corresponding to polynomials with distinct roots in F. (Hankel was 12 years old in 1851.)

Some notes on the proof:

- This is an algorithm! Given p, for increasing r, write the coefficients of p in the Hankel matrix, and look for null vectors c corresponding to polynomials with distinct roots in F. (Hankel was 12 years old in 1851.)
- Since $(\beta \frac{\partial}{\partial x} - \alpha_j \frac{\partial}{\partial y})$ kills $(\alpha x + \beta y)^d$, if $h(D)$ is defined to be $\prod_{j=1}^r (\beta_j \frac{\partial}{\partial x} - \alpha_j \frac{\partial}{\partial y})$, then

$$h(D)p = \sum_{m=0}^{d-r} \frac{d!}{(d-r-m)!m!} \left(\sum_{i=0}^{d-r} a_{i+m} c_i \right) x^{d-r-m} y^m$$

The coefficients of $h(D)p$ are, up to multiple, the rows in the matrix product, so the matrix condition is $h(D)p = 0$. Each linear factor in $h(D)$ kills a different summand, and dimension counting takes care of the rest.

Some notes on the proof:

- This is an algorithm! Given p, for increasing r, write the coefficients of p in the Hankel matrix, and look for null vectors c corresponding to polynomials with distinct roots in F. (Hankel was 12 years old in 1851.)
- Since $(\beta\frac{\partial}{\partial x} - \alpha_j\frac{\partial}{\partial y})$ kills $(\alpha x + \beta y)^d$, if $h(D)$ is defined to be $\prod_{j=1}^r(\beta_j\frac{\partial}{\partial x} - \alpha_j\frac{\partial}{\partial y})$, then

$$h(D)p = \sum_{m=0}^{d-r}\frac{d!}{(d-r-m)!m!}\left(\sum_{i=0}^{d-r}a_{i+m}c_i\right)x^{d-r-m}y^m$$

  The coefficients of $h(D)p$ are, up to multiple, the rows in the matrix product, so the matrix condition is $h(D)p = 0$. Each linear factor in $h(D)$ kills a different summand, and dimension counting takes care of the rest.
- If h has repeated factors, see Gundelfinger's Theorem (1886). A factor $(\beta x - \alpha y)^\ell$ gives a summand $(\alpha x + \beta y)^{d-(\ell-1)}q$, where q is an arbitrary form of degree $\ell - 1$.

Here is an example of Sylvester's Theorem in action. Let

$$p(x,y) = x^3 + 12x^2y - 6xy^2 + 10y^3 =$$

$$\binom{3}{0}\cdot 1\ x^3 + \binom{3}{1}\cdot 4\ x^2y + \binom{3}{2}\cdot(-2)xy^2 + \binom{3}{3}\cdot 10\ y^3$$

We have

$$\begin{pmatrix} 1 & 4 & -2 \\ 4 & -2 & 10 \end{pmatrix}$$

Here is an example of Sylvester's Theorem in action. Let

$$p(x,y) = x^3 + 12x^2y - 6xy^2 + 10y^3 =$$

$$\binom{3}{0} \cdot 1\ x^3 + \binom{3}{1} \cdot 4\ x^2y + \binom{3}{2} \cdot (-2)xy^2 + \binom{3}{3} \cdot 10\ y^3$$

We have

$$\begin{pmatrix} 1 & 4 & -2 \\ 4 & -2 & 10 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Here is an example of Sylvester's Theorem in action. Let

$$p(x,y) = x^3 + 12x^2y - 6xy^2 + 10y^3 =$$

$$\binom{3}{0} \cdot 1\ x^3 + \binom{3}{1} \cdot 4\ x^2y + \binom{3}{2} \cdot (-2)xy^2 + \binom{3}{3} \cdot 10\ y^3$$

We have

$$\begin{pmatrix} 1 & 4 & -2 \\ 4 & -2 & 10 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and $2x^2 - xy - y^2 = (2x+y)(x-y)$, so that

$$p(x,y) = \lambda_1(x-2y)^3 + \lambda_2(x+y)^3.$$

In fact, $p(x,y) = -(x-2y)^3 + 2(x+y)^3$.

The next simple example is $p(x,y) = 3x^2y$. Note that

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies c_0 = c_1 = 0$$

so that h would have to have repeated factors, and p is not a sum of two cubes. Similarly, $x^{d-1}y$ requires d d-th powers.

The next simple example is $p(x,y) = 3x^2y$. Note that

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies c_0 = c_1 = 0$$

so that h would have to have repeated factors, and p is not a sum of two cubes. Similarly, $x^{d-1}y$ requires d d-th powers.

As a side-note, Sylvester's Theorem allows one to compute the rank of a form over different fields: for example, the quintic $3x^5 - 20x^3y^2 + 10xy^4 = x^5 + (x+iy)^5 + (x-iy)^5$ is a sum of three 5-th powers over $\mathbb{Q}[i]$, four 5-th powers over $\mathbb{Q}[\sqrt{-2}]$ and five 5-th powers over any real field. This last fact follows from a different theorem of Sylvester related to Descartes' Rule of Signs, which can be found in Pólya-Szegö.

There is a complete parameterization of $(x^2 + y^2)^m$ as a sum of $m + 1$ $2m$-th powers; $m + 1$ is the minimal number.

There is a complete parameterization of $(x^2+y^2)^m$ as a sum of $m+1$ $2m$-th powers; $m+1$ is the minimal number.

Theorem ("Length",13)

The representations of $(x^2+y^2)^m$ as a sum of $m+1$ $2m$-th powers over $\mathbb{C}$ are given by

$$\binom{2m}{m}(x^2+y^2)^m$$

$$= \frac{1}{m+1}\sum_{j=0}^{m}\left(\cos\left(\tfrac{j\pi}{m+1}+\theta\right)x + \sin\left(\tfrac{j\pi}{m+1}+\theta\right)y\right)^{2m},$$

$$\theta \in \mathbb{C}.$$

There is a complete parameterization of $(x^2+y^2)^m$ as a sum of $m+1$ $2m$-th powers; $m+1$ is the minimal number.

Theorem ("Length",13)

The representations of $(x^2+y^2)^m$ as a sum of $m+1$ $2m$-th powers over $\mathbb{C}$ are given by

$$\binom{2m}{m}(x^2+y^2)^m$$

$$= \frac{1}{m+1}\sum_{j=0}^{m}\left(\cos\left(\tfrac{j\pi}{m+1}+\theta\right)x+\sin\left(\tfrac{j\pi}{m+1}+\theta\right)y\right)^{2m},$$

$$\theta \in \mathbb{C}.$$

There is a complete parameterization of $(x^2+y^2)^m$ as a sum of $m+1$ $2m$-th powers; $m+1$ is the minimal number.

Theorem ("Length",13)

The representations of $(x^2+y^2)^m$ as a sum of $m+1$ $2m$-th powers over $\mathbb{C}$ are given by

$$\binom{2m}{m}(x^2+y^2)^m$$

$$= \frac{1}{m+1}\sum_{j=0}^{m}\left(\cos\left(\tfrac{j\pi}{m+1}+\theta\right)x+\sin\left(\tfrac{j\pi}{m+1}+\theta\right)y\right)^{2m},$$

$$\theta \in \mathbb{C}.$$

The earliest version I have found of this identity is for real θ only, by Avner Friedman, from the 1950s.

This allows for the complete solution of certain moment problems over $\mathbb{R}$. For example, if $m = 2$ and

$$a_0 = 1, \quad a_1 = 0, \quad a_2 = \frac{1}{3}, \quad a_3 = 0, \quad a_4 = 1,$$

then $a_k = c_1 r_1^k + c_2 r_2^k + c_3 r_3^k$ for $0 \le k \le 4$ precisely when (up to permutation):

$$c_1 = \frac{16}{18(1+T^2)^2}, \quad c_2 = \frac{(1-\sqrt{3}T)^4}{18(1+T^2)^2}, \quad c_3 = \frac{(1+\sqrt{3}T)^4}{18(1+T^2)^2}$$

and

$$r_1 = T, \quad r_2 = \frac{\sqrt{3}+T}{1-\sqrt{3}T}, \quad r_3 = \frac{-\sqrt{3}+T}{1+\sqrt{3}T}$$

The identities are formally true whether T is real or complex.

In higher variables, there is a generally non-constructive theorem of Hilbert, which is essential to his analysis of Waring's Problem:

Theorem (Hilbert,1909)

For all n, r, let $N = \binom{n+2r-1}{n-1}$. Then there exist $0 < \lambda_k \in \mathbb{Q}$ and $\alpha_{kj} \in \mathbb{Z}, 1 \le k \le N, 1 \le j \le n$, such that

$$\sum_{k=1}^{N} \lambda_k(\alpha_{k1}x_1 + \cdots + \alpha_{kn}x_n)^{2r} = (x_1^2 + \cdots + x_n^2)^r$$

There are "constructive" versions, by Hausdorff and Stridsberg, which require knowing the roots of Hermite polynomials. The fundamental reason this is true is that the "average" value of $(\alpha \cdot x)^{2r}$ (as α is integrated over the unit sphere in the natural measure) turns out to be a multiple of $(x_1^2 + \cdots + x_n^2)^r$.

It can be shown that the smallest possible N for which there exists a (real) identity:

$$\sum_{k=1}^{N}(\alpha_{k1}x_1+\cdots+\alpha_{kn}x_n)^{2r}=(x_1^2+\cdots+x_n^2)^r$$

is $\binom{n+r-1}{n-1}$ and that, in the case this bound is met, a rescaled copy of $\{\alpha_k\}$ lives on S^{n-1}. In the theory of spherical designs, as developed by Delsarte, Goethals and Seidel in the 1970s, these are called "tight spherical $(n, 2r+1)$-designs". Tight spherical designs are known to exist when $n=2$, $2r+1\le 3$ and $(n, 2r+1) = (3,5), (7,5), (23,5), (8,7), (23,7), (24,11)$, and probably, but not yet provably, no others. When they exist, they are unique up to rotation. These are always deeply interesting combinatorial sets; the $(11, 24)$ tight spherical design consists of the 196,560 minimal vectors of the Leech Lattice in $\mathbb{R}^{24}$.

The unique tight spherical $(3,5)$-design is the icosahedron.

The unique tight spherical $(3,5)$-design is the icosahedron.

Theorem (Memoir,92)

The equation

$$(x^2+y^2+z^2)^2 = \sum_{k=1}^{6} (a_k x + b_k y + c_k z)^4 \tag{1}$$

holds if and only if the 12 points $\pm(a_k, b_k, c_k)$ *are the vertices of a regular icosahedron inscribed in a sphere with center 0 and radius* $(5/6)^{1/4}$.

As an explicit implementation of this (up to scaling) using the Schönemann coordinates for the icosahedron, we write $\Phi = \frac{1+\sqrt{5}}{2}$, and note that $\Phi^4 + 1 = 3\Phi^2$. Then

$$(x+\Phi y)^4 + (x-\Phi y)^4 + (y+\Phi z)^4 + (y-\Phi z)^4$$
$$+(z+\Phi x)^4 + (z-\Phi x)^4 = 6\Phi^2(x^2+y^2+z^2)^2$$

Thank you for your attention and your patience.