Tchebycheff systems and extremal problems for generalized moments: a brief survey

Iosif Pinelis*

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu

Abstract: A brief presentation of basics of the theory of Tchebycheff and Markov systems of functions and its applications to extremal problems for integrals of such functions is given. The results, as well as all the necessary definitions, are stated in most common terms. This work is motivated by specific applications in probability and statistics. A few related questions are also briefly discussed, including the one on the existence of a Tchebycheff system on a given topological space.

AMS 2000 subject classifications: Primary 26D15 49K30, 49K27, 52A40; secondary 49K35, 52A20, 52A41, 60E15, 41A50, 41A52, 26A48, 26B25, 46N10, 46N30, 90C05, 90C25, 90C26, 90C47, 90C48.
Keywords and phrases: Tchebycheff systems, Markov systems, extremal problems, generalized moments, exact inequalities, generalized monotonicity, generalized convexity, stochastic orders, best approximation.

The theory of Tchebycheff systems (or, briefly, T-systems) presents powerful tools that can be used in various extremal problems in analysis, probability, and statistics; see e.g. monographs [4, 6] and the extensive bibliography there. However, whereas the Carathéodory Principle (as stated below) and the related duality principle (see e.g. [6, Theorem 4.1]) have been used extensively in probability and statistics (see e.g. [4, Chapters XII and XIII] and [12, 10, 11, 9]), there appear to have been very few (if any) such applications of the theory of T-systems, even though the latter offers significant advantages. Even in [4], I have not found applications in probability or statistics of the T-systems theory per se.

This brief survey was motivated by specific applications in [8]. Here we shall present basics of the theory of T-systems and its applications to extremal problems for the corresponding generalized moments. The results, as well as all the necessary definitions, will be stated in most common terms and thus, it is hoped, easy to use; the notions of canonical and principal representations will be avoided here. A few related questions will also be briefly discussed, including the one on the existence of a T-system on a given topological space.

***

*Supported by NSF grant DMS-0805946

For a nonnegative integer n, let $g_0,\dots,g_n$ be (real-valued) continuous functions on a compact topological space X. Let M denote the set of all (nonnegative) Borel measures on X. Take any point $\mathbf{c}=(c_0,\dots,c_n)\in\mathbb{R}^{n+1}$ such that

$$M_{\mathbf{c}}:=\Big\{\mu\in M\colon\ \int_X g_i\,\mathrm{d}\mu=c_i\quad\text{for all}\quad i\in\overline{0,n}\Big\}\neq\emptyset; \tag{1}$$

Consider also the condition that the (generalized) polynomial

$$\sum_0^n\lambda_i g_i \text{ is strictly positive on } X\text{, for some } (\lambda_0,\dots,\lambda_n)\in\mathbb{R}^{n+1}. \tag{2}$$

Carathéodory Principle. *(See e.g. [3, 5].) If the topologcal compact space X is Hausdorff and (2) holds, then the maximum of $\int_X g_{n+1}\,\mathrm{d}\mu$ over all $\mu\in M_{\mathbf{c}}$ is attained at some measure $\mu_{\max}\in M_{\mathbf{c}}$ with* $\operatorname{card}\operatorname{supp}\mu_{\max}\leqslant n+1$.

Here, as usual, card stands for the cardinality and $\operatorname{supp}\mu$ denotes the support set of the measure μ. Note that in [5] it is additionally assumed that $g_0=1$, which is used to provide for the weak compactness of $M_{\mathbf{c}}$. However, the same effect is achieved under the more general condition (2).

Remark 1. The condition that X be compact can oftentimes be circumvented by using, for instance, an appropriate compactification, say $\overline{X}$, of X if X is (say) only locally compact (as, for instance, $\mathbb{R}^k$ is). At that, one may be able to find a function h, which is positive and continuous on X and such that the functions $\frac{g_0}{h},\dots,\frac{g_n}{h}$ can be continuously extended from X to $\overline{X}$; sometimes one of the g_i's can play the role of h; or, more generally, h can be constructed as (or based on) a polynomial $\sum_0^n\lambda_i g_i$. Replacing then $g_0,\dots,g_n$ by the continuous extensions of the functions $\frac{g_0}{h},\dots,\frac{g_n}{h}$ to $\overline{X}$ and, correspondingly, replacing the measure μ by the measure ν (on X and hence on $\overline{X}$) defined by the formula $\mathrm{d}\nu=h\,\mathrm{d}\mu$, one will largely reduce the original optimization problem on X to one on the compact space $\overline{X}$.

The essential fact is that the upper bound $n+1$ on the cardinality of the support of an extremal measure μ given in the Carathéodory Principle can be approximately halved in the presence of the Tchebycheff or, especially, Markov property.

Definition 1. The sequence $(g_0,\dots,g_n)$ of functions is a *T-system* if the restrictions of these $n+1$ functions to any subset of X of cardinality $n+1$ are linearly independent. If, for each $k\in\overline{0,n}$, the initial subsequence $(g_0,\dots,g_k)$ of the sequence $(g_0,\dots,g_n)$ is a T-system, then $(g_0,\dots,g_n)$ is said to be an *M-system* (where M refers to Markov).

By Haar's theorem, linearly independent functions $g_0,\dots,g_n$ on X form a T-system on X if and only the problem of best uniform approximation of any given continuous function on X by a polynomial $\sum_0^n\lambda_i g_i$ has a unique solution; see e.g. [13].

Any T-system satisfies the condition (2); see e.g. [6, Theorem II.1.4].

For any $n \geqslant 1$ and any topological space X of cardinality $\geqslant n+1$, if there exists a T-system $(g_0,\dots,g_n)$ of continuous functions on X, then X is necessarily Hausdorff. Indeed, take any distinct x_0 and x_1 in X. Let $x_2,\dots,x_n$ be any points in X such that $x_0,\dots,x_n$ are distinct. The restrictions of the functions $g_0,\dots,g_n$ to the set $\{x_0,\dots,x_n\}$ are linearly independent and hence $g_i(x_0) \neq g_i(x_1)$ for some $i \in \overline{0,n}$. Take now any disjoint open sets O_0 and O_1 in $\mathbb{R}$ containing $g_i(x_0)$ and $g_i(x_1)$, respectively. Then the pre-images $g_i^{-1}(O_0)$ and $g_i^{-1}(O_1)$ of O_0 and O_1 under the mapping g_i are disjoint open sets in X containing x_0 and x_1, respectively. Thus, X is a Hausdorff topological space.

In connection with Remark 1, it should be noted that, clearly, if $(g_0,\dots,g_k)$ is a T-system or an M-system, then the same is true of the sequence $(\frac{g_0}{h},\dots,\frac{g_n}{h})$, for any positive continuous function h. Note also that, if $(g_0,\dots,g_n)$ is a T-system on a set X' containing X and such that $\operatorname{card}(X' \setminus X) \geqslant n$, then $(h_0,\dots,h_n) := A(g_0,\dots,g_n)$ is an M-system on X for some linear (necessarily nonsingular) transformation A of $\mathbb{R}^{n+1}$; cf. [6, Theorem II.4.1].

A T-system $(g_0,\dots,g_n)$ with $n \geqslant 1$ on (the compact topological space) X exists only if X is one-dimensional (which will be the case in many applications). More precisely, if for some $n \geqslant 1$ there exists a T-system of $n+1$ functions on X, then X must be homeomorphic to a subset of a circle; for $X \subseteq \mathbb{R}^k$ this was proved in [7], and for general X in [2] (with an additional restriction) and in [16]; a further extension of this result to complex T-systems was given in [14], where one can also find yet another proof of the real-valued version.

In fact, the general case of (real-valued) T-systems can be easily reduced to the special case with $X \subseteq \mathbb{R}^k$. Indeed, for any natural n consider the mapping $x \mapsto r(x) := g(x)/\|g(x)\|$ of X into the unit sphere S^n in $\mathbb{R}^{n+1}$, where $g(x) := (g_0(x),\dots,g_n(x))$ and $\|\cdot\|$ is the Euclidean norm. In view of the T-property of $(g_0,\dots,g_n)$ and the compactness of X, this mapping is correctly defined $\big($since $g(x)$ is nonzero for any $x \in X\big)$, one-to-one, and continuous, and hence a homeomorphism of the compact Hausdorff set X onto the image in S^n of X under the mapping r. In the case $n=1$, this also proves the mentioned result of [7, 2, 16]. Another elementary observation in this regard, presented in [1], is that a T-system $(g_0,\dots,g_n)$ with $n \geqslant 1$ on X may exist only if X does not contain a "tripod", that is a set homeomorphic to the set $\{(s,0) \in \mathbb{R}^2\colon |s| < 1\} \cup \{(0,t) \in \mathbb{R}^2\colon 0 < t < 1\}$.

We shall henceforth consider the case when $X = [a,b]$ for some a and b such that $-\infty < a < b < \infty$. Let $(g_0,\dots,g_n)$ be a T-system on $[a,b]$. Let $\det\big(g_i(x_j)\big)_0^n$ denote the determinant of the matrix $\big(g_i(x_j)\colon i \in \overline{0,n},\ j \in \overline{0,n}\big)$. This determinant is continuous in $(x_0,\dots,x_n)$ in the (convex) simplex (say Σ) defined by the inequalities $a \leqslant x_0 < \dots < x_n \leqslant b$ and does not vanish anywhere on Σ. So, $\det\big(g_i(x_j)\big)_0^n$ is constant in sign on Σ.

Definition 2. The sequence $(g_0,\dots,g_n)$ is said to be a T_+-*system* on $[a,b]$ if $\det\big(g_i(x_j)\big)_0^n > 0$ for all $(x_0,\dots,x_n) \in \Sigma$. If $(g_0,\dots,g_k)$ is a T_+-system on $[a,b]$ for each $k \in \overline{0,n}$, then the sequence $(g_0,\dots,g_n)$ is said to be an M_+-*system* on $[a,b]$.

Clearly, if $(g_0,\dots,g_n)$ is a T-system on $[a,b]$, then either $(g_0,\dots,g_n)$ or

$(g_0,\dots,g_{n-1},-g_n)$ is a T_+-system on $[a,b]$. Similarly, if $(g_0,\dots,g_n)$ is an M-system on $[a,b]$ then, for some sequence $(s_0,\dots,s_n)\in\{-1,1\}^{n+1}$ of signs, $(s_0g_0\dots,s_ng_n)$ is an M_+-system on $[a,b]$.

In the case when the functions $g_0,\dots,g_n$ are n times differentiable at a point $x\in(a,b)$, consider also the *Wronskians*

$$W_0^k(x) := \det\big(g_i^{(j)}(x)\big)_0^k,$$

where $k\in\overline{0,n}$ and $g_i^{(j)}$ is the jth derivative of g_i, with $g_i^{(0)}:=g_i$; in particular, $W_0^0(x)=g_0(x)$.

Proposition 1. *Suppose that the functions $g_0,\dots,g_n$ are (still continuous on $[a,b]$ and) n times differentiable on (a,b). Then, for the sequence $(g_0,\dots,g_n)$ to be an M_+-system on $[a,b]$, it is necessary that $W_0^k\geqslant 0$ on (a,b) for all $k\in\overline{0,n}$, and it is sufficient that $u_0>0$ on $[a,b]$ and $W_0^k>0$ on (a,b) for all $k\in\overline{1,n}$.*

Thus, verifying the M_+-property largely reduces to checking the positivity of several functions of only one variable.

A special case of Proposition 1 (with $n=1$ and $g_0=1$) is the following well-known fact: if a function g_1 is continuous on $[a,b]$ and has a positive derivative on (a,b), then g_1 is (strictly) increasing on $[a,b]$; vice versa, if g_1 is increasing on $[a,b]$, then the derivative of g_1 (if exists) must be nonnegative on (a,b).

As in this special case, the proof of Proposition 1 in general can be based on the mean-value theorem; cf. e.g. [4, Theorem 1.1 of Chapter XI], which states that the requirement for W_0^k to be strictly positive on the *closed* interval $[a,b]$ for all $k\in\overline{0,n}$ is equivalent to a condition somewhat stronger than being an M_+-system on $[a,b]$; in connection with this, one may also want to look at [6, Theorem IV.5.2]. Note that, in the applications to the proofs of [8, Lemmas 2.2 and 2.3], the relevant Wronskians vanish at the left endpoint of the interval.

The proof of Proposition 1 can be obtained by induction on n using the recursive formulas for the determinants $\det\big(g_i(x_j)\big)_0^n$ and W_0^n as displayed right above [4, (5.5) in Chapter VIII] and in [4, (5.6) in Chapter VIII], where we use g_i in place of ψ_i.

Proposition 2. *Suppose that $(g_0,\dots,g_{n+1})$ is an M_+-system on $[a,b]$ or, more generally, each of the sequences $(g_0,\dots,g_n)$ and $(g_0,\dots,g_{n+1})$ is a T_+-system on $[a,b]$. Suppose also that condition (1) holds. Let $m:=\lfloor\frac{n+1}{2}\rfloor$. Then one has the following.*

(I) *The maximum (respectively, the minimum) of $\int_a^b g_{n+1}\,\mathrm{d}\mu$ over all $\mu\in M_{\mathbf{c}}$ is attained at a unique measure $\mu_{\max}$ (respectively, $\mu_{\min}$) in $M_{\mathbf{c}}$. Moreover, the measures $\mu_{\max}$ and $\mu_{\min}$ do not depend on the choice of g_{n+1}, as long as g_{n+1} is such that $(g_0,\dots,g_{n+1})$ is a T_+-system on $[a,b]$.*

(II) *There exist subsets $X_{\max}$ and $X_{\min}$ of $[a,b]$ such that $X_{\max}\supseteq\operatorname{supp}\mu_{\max}$, $X_{\min}\supseteq\operatorname{supp}\mu_{\min}$, and*

   (a) *if $n=2m$ then $\operatorname{card}X_{\max}=\operatorname{card}X_{\min}=m+1$, $X_{\max}\ni b$, and $X_{\min}\ni a$;*

(b) *if* $n = 2m-1$ *then* $\operatorname{card} X_{\max} = m+1$*,* $\operatorname{card} X_{\min} = m$*, and* $X_{\max} \supseteq \{a, b\}$.

Whereas the maximizer $\mu_{\max}$ and the minimizer $\mu_{\min}$ are each unique in $M_{\mathbf{c}}$ for each given $\mathbf{c}$ with $M_{\mathbf{c}} \neq \emptyset$, in particular applications such as the ones in [8], one may want to allow the vector $\mathbf{c}$ to vary, and then the measures $\mu_{\max}$ and $\mu_{\min}$ will vary with $\mathbf{c}$, and thus the corresponding subsets $X_{\max}$ and $X_{\min}$ of $[a, b]$ may vary. Then the number of real variables needed to describe each of the sets $X_{\max}$ and $X_{\min}$ will be about $\frac{n+1}{2}$, that is, half the number of restrictions on the measure μ and also half the upper bound on $\operatorname{card} \operatorname{supp} \mu_{\max}$ in the Carathéodory Principle; here one should also take into account that, as described in part (II) of Proposition 2, the sets $X_{\max}$ and $X_{\min}$ may have to contain at least one of the endpoints a and b of the interval $[a, b]$, with the corresponding reduction in the required number of variables. On the other hand, the Carathéodory Principle holds for more general systems of functions, defined on a set X of a much more general class.

To illustrate Proposition 2, one may consider the simplest two special cases when the conditions of the proposition hold and its conclusion is obvious:

(i) $n = 0$, $g_0(x) \equiv 1$, g_1 is increasing on $[a, b]$, and $c_0 \geqslant 0$; then $\operatorname{supp} \mu_{\max} \subseteq \{b\}$ and $\operatorname{supp} \mu_{\min} \subseteq \{a\}$; in fact, $\mu_{\max} = c_0 \delta_b$ and $\mu_{\min} = c_0 \delta_a$; here and in what follows, δ_x denotes the Dirac probability measure at point x.

(ii) $n = 1$, $g_0(x) \equiv 1$, $g_1(x) \equiv x$, g_2 is strictly convex on $[a, b]$, $c_0 \geqslant 0$, and $c_1 \in [c_0 a, c_0 b]$; then $\operatorname{supp} \mu_{\max} \subseteq \{a, b\}$ and $\operatorname{card} \operatorname{supp} \mu_{\min} \leqslant 1$; in fact, $\mu_{\max} = \frac{c_0 b - c_1}{b-a} \delta_a + \frac{c_1 - c_0 a}{b-a} \delta_b$, and $\mu_{\min} = c_0 \delta_{c_1/c_0}$ if $c_0 > 0$ and $\mu_{\min} = 0$ if $c_0 = 0$.

These examples also show that the T-property of systems of functions can be considered as generalized monotonicity/convexity; see e.g. [15] and bibliography there.

Proof of Proposition 2. Consider two cases, depending on whether $\mathbf{c}$ is strictly or singularly positive; in equivalent geometric terms, this means, respectively, that $\mathbf{c}$ belongs to the interior or the boundary of the smallest closed convex cone containing the subset $\{(g_0(x), \dots, g_n(x)) \colon x \in [a, b]\}$ of $\mathbb{R}^{n+1}$ [6, Theorem IV.6.1].

In the first case, when $\mathbf{c}$ is strictly positive, both statements of Proposition 2 follow by [6, Theorem IV.1.1]; at that, one should let $X_{\max} = \operatorname{supp} \mu_{\max}$ and $X_{\min} = \operatorname{supp} \mu_{\min}$. (The condition that $\mathbf{c}$ be strictly positive appears to be missing in the statement of the latter theorem; cf. [4, Theorem 1.1 of Chapter 1.1].)

In the other case, when $\mathbf{c}$ is singularly positive, use [6, Theorem III.4.1], which states that in this case the set $M_{\mathbf{c}}$ consists of a single measure (say μ_*), and its support set $X_* := \operatorname{supp} \mu_*$ is of an index $\leqslant n$; that is, $\ell_- + 2\ell + \ell_+ \leqslant n$, where ℓ_-, ℓ, and ℓ_+ stand for the cardinalities of the intersections of X_* with the sets $\{a\}$, (a, b), and $\{b\}$. It remains to show that this condition on the index of X_* implies that there exist subsets $X_{\max}$ and $X_{\min}$ of $[a, b]$ satisfying the conditions (IIa) and (IIb) of Proposition 2 and such that $X_{\max} \cap X_{\min} \supseteq X_*$.

If $n = 2m$ then $\operatorname{card}(X_* \cap (a,b)) = \ell \leqslant \lfloor \frac{2m-\ell_- -\ell_+}{2} \rfloor \leqslant \lfloor \frac{2m-\ell_-}{2} \rfloor = m - \ell_-$; so, $\operatorname{card}(X_* \cup \{b\}) \leqslant \ell_- + (m - \ell_-) + 1 = m+1$. Adding now to the set $X_* \cup \{b\}$ any $m+1-\operatorname{card}(X_* \cup \{b\})$ points of the complement of $X_* \cup \{b\}$ to $[a,b]$, one obtains a subset $X_{\max}$ of $[a,b]$ such that $X_{\max} \supseteq X_*$, $X_{\max} \ni b$, and $\operatorname{card} X_{\max} = m+1$. Similarly, there exists a subset $X_{\min}$ of $[a,b]$ such that $X_{\min} \supseteq X_*$, $X_{\min} \ni a$, and $\operatorname{card} X_{\min} = m+1$.

If $n = 2m-1$ then $\operatorname{card}(X_* \cap (a,b)) = \ell \leqslant \lfloor \frac{2m-1-\ell_- -\ell_+}{2} \rfloor \leqslant m-1$ and hence $\operatorname{card}(X_* \cup \{a,b\}) \leqslant 1 + (m-1) + 1 = m+1$. So, there exists a subset $X_{\max}$ of $[a,b]$ such that $X_{\max} \supseteq X_*$, $X_{\max} \supseteq \{a,b\}$, and $\operatorname{card} X_{\max} = m+1$. One also has $\operatorname{card} X_* = \ell_- + \ell + \ell_+ \leqslant \lfloor \frac{2m-1+\ell_- +\ell_+}{2} \rfloor \leqslant \lfloor \frac{2m+1}{2} \rfloor = m$. So, there exists a subset $X_{\min}$ of $[a,b]$ such that $X_{\min} \supseteq X_*$ and $\operatorname{card} X_{\min} = m$. □

References

[1] R. C. Buck. Linear spaces and approximation theory. In *On numerical approximation. Proceedings of a Symposium, Madison, April 21-23, 1958*, Edited by R. E. Langer. Publication no. 1 of the Mathematics Research Center, U.S. Army, the University of Wisconsin, pages 11–23. The University of Wisconsin Press, Madison, 1959.

[2] P. C. Curtis, Jr. n-parameter families and best approximation. *Pacific J. Math.*, 9:1013–1027, 1959.

[3] W. Hoeffding. The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist.*, 26:268–275, 1955.

[4] S. Karlin and W. J. Studden. *Tchebycheff systems: With applications in analysis and statistics*. Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

[5] A. F. Karr. Extreme points of certain sets of probability measures, with applications. *Math. Oper. Res.*, 8(1):74–85, 1983.

[6] M. G. Kreĭn and A. A. Nudel′man. *The Markov moment problem and extremal problems*. American Mathematical Society, Providence, R.I., 1977. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50.

[7] J. C. Mairhuber. On Haar's theorem concerning Chebychev approximation problems having unique solutions. *Proc. Amer. Math. Soc.*, 7:609–615, 1956.

[8] I. Pinelis. An asymptotically Gaussian bound on the Rademacher tails (preprint), arXiv:1007.2137 [math.PR].

[9] I. Pinelis. Exact lower bounds on the exponential moments of Winsorized and truncated random variables (preprint), arXiv:1001.2901 [math.PR].

[10] I. Pinelis. Optimal tail comparison based on comparison of moments. In *High dimensional probability (Oberwolfach, 1996)*, volume 43 of *Progr. Probab.*, pages 297–314. Birkhäuser, Basel, 1998.

[11] I. Pinelis. On normal domination of (super)martingales. *Electron. J. Probab.*, 11:no. 39, 1049–1070, 2006.

[12] I. S. Pinelis and S. A. Utev. Sharp exponential estimates for sums of independent random variables. *Theory Probab. Appl.*, 34(2):340–346, 1989.
[13] I. J. Schoenberg. On the question of unicity in the theory of best approximations. *Ann. New York Acad. Sci.*, 86:682–692 (1960), 1960.
[14] I. J. Schoenberg and C. T. Yang. On the unicity of solutions of problems of best approximation. *Ann. Mat. Pura Appl. (4)*, 54:1–12, 1961.
[15] M. Shaked and J. G. Shanthikumar. *Stochastic orders*. Springer Series in Statistics. Springer, New York, 2007.
[16] K. Sieklucki. Topological properties of sets admitting the Tschebycheff systems. *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astr. Phys.*, 6:603–606, 1958.