PADOVA UNIVERSITY PRESS

Dolomites Research Notes on Approximation

Volume 10 · 2017 · Pages 5–14

# Caratheodory-Tchakaloff Subsampling

Federico Piazzon[a] · Alvise Sommariva[b] · Marco Vianello[c]

*Communicated by L. Bos*

**Abstract**

We present a brief survey on the compression of discrete measures by Caratheodory-Tchakaloff Subsampling, its implementation by Linear or Quadratic Programming and the application to multivariate polynomial Least Squares. We also give an algorithm that computes the corresponding Caratheodory-Tchakaloff (CATCH) points and weights for polynomial spaces on compact sets and manifolds in 2D and 3D.

**2010 AMS subject classification:** 41A10, 65D32, 93E24.

**Keywords:** multivariate discrete measures, compression, subsampling, Tchakaloff theorem, Caratheodory theorem, Linear Programming, Quadratic Programming, polynomial Least Squares, polynomial meshes.

## 1 Subsampling for discrete measures

Tchakaloff theorem, a cornerstone of quadrature theory, substantially asserts that for every compactly supported measure there exists a positive algebraic quadrature formula with cardinality not exceeding the dimension of the exactness polynomial space (restricted to the measure support). Originally proved by V. Tchakaloff in 1957 for absolutely continuous measures [31], it has then be extended to any measure with finite polynomial moments, cf. e.g. [10], and to arbitrary finite dimensional spaces of integrable functions [1].

We begin by stating a discrete version of Tchakaloff theorem, in its full generality, whose proof is based on Caratheodory theorem about finite dimensional conic combinations.

**Theorem 1.1.** *Let $\mu$ be a multivariate discrete measure supported at a finite set $X = \{x_i\} \subset \mathbb{R}^d$, with correspondent positive weights (masses) $\boldsymbol{\lambda} = \{\lambda_i\}$, $i = 1, \dots, M$, and let $S = span(\phi_1, \dots, \phi_L)$ a finite dimensional space of $d$-variate functions defined on $K \supseteq X$, with $N = dim(S|_X) \le L$.*

*Then, there exist a quadrature formula with nodes $T = \{t_j\} \subseteq X$ and positive weights $\boldsymbol{w} = \{w_j\}$, $1 \le j \le m \le N$, such that*

$$\int_X f(x)\, d\mu = \sum_{i=1}^{M} \lambda_i f(x_i) = \sum_{j=1}^{m} w_j f(t_j)\,, \quad \forall f \in S|_X\,. \tag{1}$$

*Proof.* Let $\{\psi_1, \dots, \psi_N\}$ be a basis of $S|_X$, and $V = (v_{ij}) = (\psi_j(x_i))$ the Vandermonde-like matrix of the basis computed at the support points. If $M > N$ (otherwise there is nothing to prove), existence of a positive quadrature formula for $\mu$ with cardinality not exceeding $N$ can be immediately translated into existence of a nonnegative solution with at most $N$ nonvanishing components to the underdetermined linear system

$$V^t \boldsymbol{u} = \boldsymbol{b}\,, \ \ \boldsymbol{u} \ge \boldsymbol{0}\,, \tag{2}$$

where

$$\boldsymbol{b} = V^t \boldsymbol{\lambda} = \left\{ \int_X \psi_j(x)\, d\mu \right\}\,, \ \ 1 \le j \le N\,, \tag{3}$$

is the vector of $\mu$-moments of the basis $\{\psi_j\}$.

Existence then holds by the well-known Caratheodory theorem applied to the columns of $V^t$, which asserts that a conic (i.e., with positive coefficients) combination of any numer of vectors in $\mathbb{R}^N$ can be rewritten as a conic combination of at most $N$ (linearly independent) of them; cf. [8] and, e.g., [9, §3.4.4]. □

*Remark* 1. Our main application of Theorem 1 concerns total-degree polynomial spaces, $S = \mathbb{P}_\nu^d(K)$. If $K$ is (a compact subset of) an algebraic variety of $\mathbb{R}^d$, then $dim(S|_X) \le dim(S) < L = dim(\mathbb{P}_\nu^d) = \binom{\nu+d}{d}$; if in addition $X$ is $S$-determining, i.e. polynomials vanishing on $X$ vanish everywhere on $K$, then $dim(S|_X) = dim(S)$. Indeed, a crucial step of our approach will be that of identifying, at least numerically, the right dimension and a polynomial basis of $S|_X$, starting from a standard basis of $\mathbb{P}_\nu^d$.

Since the discrete version of Tchakaloff theorem given by Theorem 1 is a direct consequence of Caratheodory theorem, we may term such an approach *Caratheodory-Tchakaloff subsampling*, and the corresponding nodes (with associated weights) a set of Caratheodory-Tchakaloff (CATCH) points.

[a]Department of Mathematics, University of Padova, Italy, e-mail: fpiazzon@math.unipd.it
[b]Department of Mathematics, University of Padova, Italy, e-mail: alvise@math.unipd.it
[c]Department of Mathematics, University of Padova, Italy, e-mail: marcov@math.unipd.it

The idea of reduction/compression of a finite measure by Tchakaloff or directly Caratheodory theorem recently arose in different contexts, for example in a probabilistic setting [17], as well as in univariate [14] and multivariate [2, 22, 27, 30] numerical quadrature, with applications to multivariate polynomial inequalities and least squares approximation [22, 30, 33]. In many situations CATCH subsampling can produce a high Compression Ratio, namely when $N \ll M$ like for example in polynomial least squares approximation [30] or in QMC (Quasi-Monte Carlo) integration [2] or in particle methods [17],

$$C_{ratio} = \frac{M}{m} \geq \frac{M}{N} \gg 1 \,, \tag{4}$$

so that the efficient computation of CATCH points and weights becomes a relevant task.

Now, while the proof of the general Tchakaloff theorem is not, that of the discrete version can be made constructive, since Caratheodory theorem itself has a constructive proof (cf., e.g., [9, §3.4.4]). On the other hand, such a proof does not give directly an efficient implementation. Nevertheless, there are at least two reasonably efficient approaches to solve the problem.

The first, adopted for example in [14] (univariate) and [30] (multivariate) in the framework of polynomial spaces, rests on *Quadratic Programming*, namely on the classical *Lawson-Hanson active set method* for *NonNegative Least Squares* (NLLS). Indeed, we may think to solve the quadratic minimum problem

$$\text{NNLS}: \quad \begin{cases} \min \|V^t \boldsymbol{u} - \boldsymbol{b}\|_2 \\ \boldsymbol{u} \geq \boldsymbol{0} \end{cases} \tag{5}$$

which exists by Theorem 1 and can be computed by standard NNLS solvers based on the Lawson-Hanson method [16], which seeks a sparse solution. Then, the nonvanishing components of such a solution give the weights $\boldsymbol{w} = \{w_j\}$ as well as the indexes of the nodes $T = \{t_j\}$ within $X$. A variant of the Lawson-Hanson method is implemented in the Matlab native function `lsqnonneg` [18], while a recent optimized Matlab implementation can be found in [28].

The second approach is based instead on *Linear Programming* via the classical *simplex method*. Namely, we may think to solve the linear minimum problem

$$\text{LP}: \quad \begin{cases} \min \boldsymbol{c}^t \boldsymbol{u} \\ V^t \boldsymbol{u} = \boldsymbol{b} \,, \;\; \boldsymbol{u} \geq \boldsymbol{0} \end{cases} \tag{6}$$

where the constraints identify a polytope (the feasible region) in $\mathbb{R}^M$ and the vector $\boldsymbol{c}$ is chosen to be linearly independent from the rows of $V^t$ (i.e., it is not the restriction to $X$ of a function in $S$), so that the objective functional is not constant on the polytope. To this aim, if $X \subset K$ is determining on a supspace $T \supset S$ on $K$, i.e. a function in $T$ vanishing on $X$ vanishes everywhere on $K$, then it is sufficient to take $\boldsymbol{c} = \{g(x_i)\}$, $1 \leq i \leq M$, where the function $g|_K$ belongs to $T|_K \setminus S|_K$. For example, working with polynomials it is sufficient to take a polynomial of higher degree on $K$ with respect to those in $S|_K$.

Observe that in our setting the feasible region is nonempty, since $\boldsymbol{b} = V^t \boldsymbol{\lambda}$, and we are interested in any basic feasible solution, i.e., in any vertex of the polytope, that has at least $M - N$ vanishing components. As it is well-known, the solution of the Linear Programming problem is a vertex of the polytope that can be computed by the simplex method (cf., e.g., [9]). Again, the nonvanishing components of such a vertex give the weights $\boldsymbol{w} = \{w_j\}$ as well as the indexes of the nodes $T = \{t_j\}$ within $X$.

This approach was adopted for example in [27] as a basic step to compute, when it exists, a multivariate algebraic Gaussian quadrature formula (suitable choices of $\boldsymbol{c}$ are also discussed there; see Example 1 below). Linear Programming is also used in [19], to generate iteratively moment matching scenarios in view of probabilistic applications (e.g., stochastic programming).

Even though both, the active set method for (5) and the simplex method for (6), have theoretically an exponential complexity (worst case analysis), as it is well-known their practical behavior is quite satisfactory, since the average complexity turns out to be polynomial in the dimension of the problems (observe that in the present setting we deal with dense matrices); cf., e.g., [13, Ch. 9]. It is worth quoting here the extensive theoretical and computational results recently presented in the Ph.D. dissertation [32], where Caratheodory reduction of a discrete measure is implemented by Linear Programming, claiming an experimental average cost of $\mathcal{O}(N^{3.7})$.

A different combinatorial algorithm (Recursive Halving Forest), based on the SVD, is also there proposed to compute a basic feasible solution and compared with the best Linear Programming solvers, claiming an experimental average cost of $\mathcal{O}(N^{2.6})$. The methods are essentially applied to the reduction of Cartesian tensor cubature measures.

In our implementation of CATCH subsampling [23], we have chosen to work with the Octave native Linear Programming solver `glpk` [21] and the Matlab native Quadratic Programming solver `lsqnonneg` [18], that are suitable for moderate size problems, like those typically arising with total-degree polynomial spaces ($S = S_\nu = \mathbb{P}_\nu^d(K)$) in dimension $d = 2, 3$ and small/moderate degree of exactness $\nu$. On large size problems, like those typically arising in higher dimension and/or high degree of exactness, the solvers discussed in [32] could become necessary.

Now, since we may expect that the underdetermined system (2) is not satisfied exactly by the computed solution, due to finite precision arithmetic and by the effect of an error tolerance in the iterative algorithms, namely that there is a nonzero moment residual

$$\|V^t \boldsymbol{u} - \boldsymbol{b}\|_2 = \varepsilon > 0 \,, \tag{7}$$

it is then worth studying the effect of such a residual on the accuracy of the quadrature formula. We can state and prove an estimate still in the general discrete setting of Theorem 1.

**Proposition 1.2.** *Let the assumptions of Theorem 1 be satisfied, let $\boldsymbol{u}$ be a nonnegative vector such that (7) holds, where $V$ is the Vandermonde-like matrix at $X$ corresponding to a $\mu$-orthonormal basis $\{\psi_k\}$ of $S|_X$, and let $(T, \boldsymbol{w})$ be the quadrature formula corresponding to the nonvanishing components of $\boldsymbol{u}$. Moreover, let $1 \in S$ (i.e., $S$ contains the constant functions).*

*Then, for every function $f$ defined on $X$, the following error estimate holds*

$$\left| \int_X f(x)\, d\mu - \sum_{j=1}^{m} w_j f(t_j) \right| \leq C_\varepsilon E_S(f; X) + \varepsilon \|f\|_{\ell^2_\lambda(X)} \,, \tag{8}$$

*where*

$$E_S(f; X) = \min_{\phi \in S} \|f - \phi\|_{\ell^\infty(X)} \,, \;\; C_\varepsilon = 2\left(\mu(X) + \varepsilon \sqrt{\mu(X)}\right) . \tag{9}$$

*Proof.* First, observe that

$$\int_X \phi(x)\, d\mu = \langle \boldsymbol{\gamma}, \boldsymbol{b} \rangle \,, \;\; \forall \phi \in S \,, \tag{10}$$

$\boldsymbol{\gamma} = \{\gamma_k\}$, $\boldsymbol{b} = \{b_k\} = V^t\boldsymbol{\lambda}$, where $\langle\cdot,\cdot\rangle$ denotes the Euclidean scalar product in $\mathbb{R}^N$ and

$$\gamma_k = \int_X \phi(x)\psi_k(x)\,d\mu\;,\;\; b_k = \int_X \psi_k(x)\,d\mu\;,\;\; 1\le k\le N\;,$$

are the coefficients of $\phi$ in the $\mu$-orthonormal basis $\{\psi_k\}$ and the $\mu$-moments of $\{\psi_k\}$, respectively.

Take $\phi\in S$. By a classical chain of estimates in quadrature theory, we can write

$$\left|\int_X f(x)\,d\mu - \sum_{j=1}^m w_j f(t_j)\right| \le \int_X |f(x)-\phi(x)|\,d\mu$$

$$+\left|\int_X \phi(x)\,d\mu - \sum_{j=1}^m w_j\,\phi(t_j)\right| + \sum_{j=1}^m w_j\,|\phi(t_j)-f(t_j)|$$

$$\le \left(\mu(X)+\sum_{j=1}^m w_j\right)\|f-\phi\|_{\ell^\infty(X)} + \left|\int_X \phi(x)\,d\mu - \sum_{j=1}^m w_j\,\phi(t_j)\right|\,. \tag{11}$$

Now,

$$\sum_{j=1}^m w_j\,\phi(t_j) = \sum_{k=1}^N \gamma_k \sum_{j=1}^m w_j\,\psi_k(t_j) = \langle\boldsymbol{\gamma}, V^t\boldsymbol{u}\rangle\;,$$

and thus by the Cauchy-Schwarz inequality

$$\left|\int_X \phi(x)\,d\mu - \sum_{j=1}^m w_j\,\phi(t_j)\right| = |\langle\boldsymbol{\gamma}, \boldsymbol{b}-V^t\boldsymbol{u}\rangle| \le \|\boldsymbol{\gamma}\|_2\,\|\boldsymbol{b}-V^t\boldsymbol{u}\|_2 = \|\phi\|_{\ell^2_{\boldsymbol{\lambda}}(X)}\,\varepsilon\,. \tag{12}$$

Moreover

$$\|\phi\|_{\ell^2_{\boldsymbol{\lambda}}(X)} \le \|\phi-f\|_{\ell^2_{\boldsymbol{\lambda}}(X)} + \|f\|_{\ell^2_{\boldsymbol{\lambda}}(X)}$$

$$\le \sqrt{\mu(X)}\,\|\phi-f\|_{\ell^\infty(X)} + \|f\|_{\ell^2_{\boldsymbol{\lambda}}(X)}\,. \tag{13}$$

On the other hand

$$\sum_{j=1}^m w_j \le \left|\sum_{j=1}^m w_j - \int_X 1\,d\mu\right| + \int_X 1\,d\mu = \left|\sum_{j=1}^m w_j - \int_X 1\,d\mu\right| + \mu(X)$$

$$\le \varepsilon\,\|1\|_{\ell^2_{\boldsymbol{\lambda}}(X)} + \mu(X) = \varepsilon\,\sqrt{\mu(X)} + \mu(X)\,, \tag{14}$$

where we have applied (12) with $\phi=1$.

Putting estimates (12)-(14) into (11 we obtain

$$\left|\int_X f(x)\,d\mu - \sum_{j=1}^m w_j f(t_j)\right| \le \left(2\mu(X)+\varepsilon\,\sqrt{\mu(X)}\right)\|f-\phi\|_{\ell^\infty(X)}$$

$$+\varepsilon\left(\sqrt{\mu(X)}\,\|\phi-f\|_{\ell^\infty(X)} + \|f\|_{\ell^2_{\boldsymbol{\lambda}}(X)}\right)\,,\;\;\forall\phi\in S\,,$$

and taking the minimum over $\phi\in S$ we finally get (8). □

It is worth observing that the assumption $1\in S$ is quite natural, being satisfied for example in the usual polynomial and trigonometric spaces. From this point of view, we can also stress that sparsity cannot be ensured by the standard Compressive Sensing approach to underdetermined systems, such as the Basis Pursuit algorithm that minimizes $\|u\|_1$ (cf., e.g., [12]), since if $1\in S$ then $\|u\|_1 = \mu(X)$ is constant.

Moreover, we notice that if $K\supset X$ is a compact set, then

$$E_S(f;X) \le E_S(f;K)\,,\;\;\forall f\in C(K)\,. \tag{15}$$

If $S$ is a polynomial space (as in the sequel) and $K$ is a "Jackson compact", $E_S(f;K)$ can be estimated by the regularity of $f$ via multivariate Jackson-like theorems; cf. [26].

To conclude this Section, we sketch the pseudo-code of an algorithm that implements CATCH subsampling, via the preliminary computation of an orthonormal basis of $S|_X$.

**Algorithm:** *(computation of CATCH points and weights):*

- *input: the discrete measure $(X,\boldsymbol{\lambda})$, the generators $(\phi_k)=(\phi_1,\dots,\phi_L)$ of $S$, possibly the dimension $N$ of $S|_X$*

1. *compute the Vandermode-like matrix $U=(u_{ik})=(\phi_k(x_i))$*
2. *if $N$ is unknown, compute $N=rank(U)$ by a rank-revealing algorithm*
3. *compute the QR factorization with column pivoting $\sqrt{\Lambda}\,U(:,\boldsymbol{\pi})=QR$, where $\Lambda=diag(\lambda_i)$ and $\boldsymbol{\pi}$ is a permutation vector (we observe that $rank(Q)=rank(\sqrt{\Lambda}\,U)=rank(U)=N$)*
4. *select the orthogonal matrix $V=Q(:,1:N)$; the first $N$ columns of $Q$ correspond to an orthonormal basis of $S|_X$ with respect to the measure $(X,\boldsymbol{\lambda})$, $(\psi_j)=(\phi_{\pi_j})R_N^{-1}$, $1\le j\le N$, where $R_N=R(1:N,1:N)$*
5. *compute a sparse solution to $V^t\boldsymbol{u}=\boldsymbol{b}=V^t\boldsymbol{\lambda}$, $\boldsymbol{u}\ge\boldsymbol{0}$, by the Lawson-Hanson method for the NNLS problem (5) or by the simplex method for the LP problem (6)*

6. *compute the residual* $\varepsilon = \|V^t \boldsymbol{u} - \boldsymbol{b}\|_2$
7. $ind = \{i : u_i \neq 0\}$, $\boldsymbol{w} = \boldsymbol{u}(ind)$, $T = X(ind)$

- *output: the CATCH compressed measure* $(T, \boldsymbol{w})$ *and the residual* $\varepsilon$ *(that appears in the relevant estimates (8)-(9))*

*Remark* 2. We observe that there are two key tools of numerical linear algebra in this algorithm, that allow to work in the right space, in view of the fact that $rank(U) = \dim(S|_X)$. The first is the computation of such a rank, that gives of course a numerical rank, due to finite precision arithmetic. Here we can resort, for example, to the SVD decomposition of $U$ in its less costly version that produces only the singular values (with a threshold on such values), which is just that used by the `rank` Matlab/Octave native function. The second is the computation of a basis of $S|_X$, namely an orthonormal basis, by the pivoting process which is aimed at selecting linearly independent generators.

An alternative approach could consist in adopting a Rank-Revealing QR factorization algorithm (RRQR), that would reduce steps 2-3 to one single matrix factorization. Such algorithms, however, are not at hand in standard Matlab and typically require the use of MEX files (cf., e.g., [11]).

## 2 Caratheodory-Tchakaloff Least Squares

The case where $(X, \boldsymbol{\lambda})$ is itself a quadrature/cubature formula for some measure on $K \supset X$, that is the compression (or reduction) of such formulas, has been till now the main application of Caratheodory-Tchakaloff subsampling, in the classical framework of algebraic formulas as well as in the probabilistic/QMC framework; cf. [14, 27, 30] and [2, 17, 32]. In this survey, we concentrate on another relevant application, that is the *compression of multivariate polynomial least squares*.

Let us consider the total-degree polynomial framework, that is

$$S = S_\nu = \mathbb{P}_\nu^d(K)\,, \tag{16}$$

the space of $d$-variate real polynomials with total-degree not exceeding $\nu$, restricted to $K \subset \mathbb{R}^d$, a compact set or a compact (subset of a) manifold. Let us define for notational convenience

$$E_n(f) = E_{\mathbb{P}_n^d(K)}(f;K) = \min_{p\in\mathbb{P}_n^d(K)} \|f - p\|_{L^\infty(K)}\,, \tag{17}$$

where $f \in C(K)$.

Discrete LS approximation by total-degree polynomials of degree at most $n$ on $X \subset K$ is ultimately an orthogonal projection of a function $f$ on $\mathbb{P}_n^d(X)$, with respect to the scalar product of $\ell^2(X)$, namely

$$\|f - \mathcal{L}_n f\|_{\ell^2(X)} = \min_{p\in\mathbb{P}_n^d(K)} \|f - p\|_{\ell^2(X)} = \min_{p\in\mathbb{P}_n^d(X)} \|f - p\|_{\ell^2(X)}\,. \tag{18}$$

Recall that for every function $g$ defined on $X$

$$\|g\|_{\ell^2(X)}^2 = \sum_{i=1}^{M} g^2(x_i) = \int_X g^2(x)\,d\mu\,, \tag{19}$$

where $\mu$ is the discrete measure supported at $X$ with unit masses $\boldsymbol{\lambda} = (1,\ldots,1)$.

Taking $p^* \in \mathbb{P}_n^d(X)$ such that $\|f - p^*\|_{\ell^\infty(X)}$ is minimum (the polynomial of best uniform approximation of $f$ in $\mathbb{P}_n^d(X)$), we get immediately the classical LS error estimate

$$\|f - \mathcal{L}_n f\|_{\ell^2(X)} \le \|f - p^*\|_{\ell^2(X)} \le \sqrt{M}\,\|f - p^*\|_{\ell^\infty(X)} \le \sqrt{M}\,E_n(f)\,, \tag{20}$$

where $M = \mu(X) = \mathrm{card}(X)$. In terms of the Root Mean Square Error (RMSE), an indicator widely used in the applications, we have

$$\mathrm{RMSE}_X(\mathcal{L}_n f) = \frac{1}{\sqrt{M}}\,\|f - \mathcal{L}_n f\|_{\ell^2(X)} \le E_n(f)\,. \tag{21}$$

Now, if $M > N_{2n} = \dim(\mathbb{P}_{2n}^d(X))$ (we stress that here polynomials of degree $2n$ are involved), by Theorem 1 there exist $m \le N_{2n}$ Caratheodory-Tchakaloff (CATCH) points $T_{2n} = \{t_j\}$ and weights $\boldsymbol{w} = \{w_j\}$, $1 \le j \le m$, such that the following basic $\ell^2$ identity holds

$$\|p\|_{\ell^2(X)}^2 = \sum_{i=1}^{M} p^2(x_i) = \sum_{j=1}^{m} w_j\,p^2(t_j) = \|p\|_{\ell^2_w(T_{2n})}^2\,,\quad \forall p \in \mathbb{P}_n^d(X)\,. \tag{22}$$

Notice that the CATCH points $T_{2n} \subset X$ are $\mathbb{P}_n^d(X)$-determining, i.e., a polynomial of degree at most $n$ vanishing there vanishes everywhere on $X$, or in other terms $\dim(\mathbb{P}_n^d(T_{2n})) = \dim(\mathbb{P}_n^d(X))$, or equivalently any Vandermonde-like matrix with a basis of $\mathbb{P}_n^d(X)$ at $T_{2n}$ has full rank. This also entails that, if $X$ is $\mathbb{P}_n^d(K)$-determining, then such is $T_{2n}$.

Consider the $\ell_w^2(T_{2n})$ LS polynomial $\mathcal{L}_n^c f$, namely

$$\|f - \mathcal{L}_n^c f\|_{\ell^2_w(T_{2n})} = \min_{p\in\mathbb{P}_n^d(K)} \|f - p\|_{\ell^2_w(T_{2n})} = \min_{p\in\mathbb{P}_n^d(X)} \|f - p\|_{\ell^2_w(T_{2n})}\,. \tag{23}$$

Notice that $\mathcal{L}_n^c$ is a *weighted* least squares operator; reasoning as in (21) and observing that $\sum_{j=1}^m w_j = M$ since $1 \in \mathbb{P}_n^d$, we get immediately

$$\|f - \mathcal{L}_n^c f\|_{\ell^2_w(T_{2n})} \le \sqrt{M}\,E_n(f)\,. \tag{24}$$

On the other hand, we can also write the following estimates

$$\|f - \mathcal{L}_n^c f\|_{\ell^2(X)} \le \|f - p^*\|_{\ell^2(X)} + \|\mathcal{L}_n^c(p^* - f)\|_{\ell^2(X)}$$

and

$$\|\mathcal{L}_n^c(p^* - f)\|_{\ell^2(X)} = \|\mathcal{L}_n^c(p^* - f)\|_{\ell^2_w(T_{2n})} \le \|p^* - f\|_{\ell^2_w(T_{2n})}\,,$$

where we have used the basic $\ell^2$ identity (22), the fact that $\mathcal{L}_n^c p^* = p^*$ and that $\mathcal{L}_n^c f$ is an orthogonal projection. By the two estimates above we get eventually

$$\|f - \mathcal{L}_n^c f\|_{\ell^2(X)} \leq \sqrt{M}\left(\|f-p^*\|_{\ell^\infty(X)} + \|f-p^*\|_{\ell^\infty(T_{2n})}\right) \leq 2\sqrt{M}\,E_n(f)\,, \tag{25}$$

or, in RMSE terms,

$$\mathrm{RMSE}_X(\mathcal{L}_n^c f) \leq 2E_n(f)\,, \tag{26}$$

which shows the most relevant feature of the "compressed" least squares operator $\mathcal{L}_n^c$ at the CATCH points (CATCHLS), namely that

- *the LS and compressed CATCHLS RMSE estimates (21) and (26) have substantially the same size.*

This fact, in particular the appearance of the factor 2 in the estimate for the compressed operator, is reminiscent of *hyperinterpolation theory* [29]. Indeed, what we are constructing here is a sort of hyperinterpolation in a fully discrete setting. Roughly summarizing, hyperinterpolation ultimately approximates a (weighted) $L^2$ projection on $\mathbb{P}_n^d$ by a discrete weighted $\ell^2$ projection, via a quadrature formula of exactness degree $2n$. Similarly, here we are approximating a $\ell^2$ projection on $\mathbb{P}_n^d$ by a weighted $\ell^2$ projection with a smaller support, again via a quadrature formula of exactness degree $2n$.

The estimates above are valid by the theoretical exactness of the quadrature formula. In order to take into account a nonzero moment residual as in (7), we state and prove the following

**Proposition 2.1.** *Let $\mu$ be the discrete measure supported at $X$ with unit masses $\boldsymbol{\lambda} = (1,\ldots,1)$, let $\boldsymbol{u}$ be a nonnegative vector such that (7) holds, where $V$ is the orthogonal Vandermonde-like matrix at $X$ corresponding to a $\mu$-orthonormal basis $\{\psi_k\}$ of $\mathbb{P}_{2n}^d(X)$, and let $(T_{2n},\boldsymbol{w})$ be the quadrature formula corresponding to the nonvanishing components of $\boldsymbol{u}$. Then the following polynomial inequalities hold for every $p \in \mathbb{P}_n^d(X)$*

$$\|p\|_{\ell^2(X)} \leq \alpha_M(\varepsilon)\,\|p\|_{\ell^2_w(T_{2n})} \leq \sqrt{M}\,\beta_M(\varepsilon)\,\|p\|_{\ell^\infty(T_{2n})}\,, \tag{27}$$

*where*

$$\alpha_M(\varepsilon) = \left(1-\varepsilon\sqrt{M}\right)^{-1/2}\,,\quad \beta_M(\varepsilon) = \alpha_M(\varepsilon)\left(1+\varepsilon/\sqrt{M}\right)^{1/2}\,, \tag{28}$$

*provided that $\varepsilon\sqrt{M} < 1$.*

**Corollary 2.2.** *Let the assumptions of Proposition 2 be satisfied. Then the following error estimate holds for every $f \in C(K)$*

$$\|f - \mathcal{L}_n^c f\|_{\ell^2(X)} \leq (1+\beta_M(\varepsilon))\,\sqrt{M}\,E_n(f)\,. \tag{29}$$

*Proof of Proposition 2 and Corollary 1.* First, observe that

$$\|p\|^2_{\ell^2(X)} = \int_X p^2(x)\,d\mu = \sum_{j=1}^m w_j\,p^2(t_j) + \varepsilon_{2n}$$

$$\leq \sum_{j=1}^m w_j\,p^2(t_j) + |\varepsilon_{2n}| = \|p\|^2_{\ell^2_w(T_{2n})} + |\varepsilon_{2n}|\,,$$

where by Proposition 1

$$|\varepsilon_{2n}| \leq \varepsilon\|p^2\|_{\ell^2(X)}\,.$$

Now, using the fact that we are in a fully discrete setting, we get

$$\|p^2\|_{\ell^2(X)} \leq \sqrt{M}\,\|p^2\|_{\ell^\infty(X)} = \sqrt{M}\,\|p\|^2_{\ell^\infty(X)} \leq \sqrt{M}\,\|p\|^2_{\ell^2(X)}\,,$$

and finally putting together the three estimates above

$$\|p\|^2_{\ell^2(X)} \leq \|p\|^2_{\ell^2_w(T_{2n})} + \varepsilon\sqrt{M}\,\|p\|^2_{\ell^2(X)}\,,$$

that is the first inequality in (27), provided that $\varepsilon\sqrt{M} < 1$. To get the second inequality in (27), we simply observe that for every function $g$ defined on $X$

$$\|g\|^2_{\ell^2_w(T_{2n})} \leq \left(\sum_{j=1}^m w_j\right)\|g\|^2_{\ell^\infty(T_{2n})} \leq M\left(1+\varepsilon/\sqrt{M}\right)\|g\|^2_{\ell^\infty(T_{2n})} \tag{30}$$

in view of (14) (here $\mu(X) = M$). We notice incidentally that the estimates in [30, §4] must be corrected, since the factor $(1+\varepsilon/\sqrt{M})^{1/2}$ is missing there.

Concerning Corollary 1, take $p^* \in \mathbb{P}_n^d(X)$ such that $\|f-p^*\|_{\ell^\infty(X)}$ is minimum (the polynomial of best uniform approximation of $f$ in $\mathbb{P}_n^d(X)$). Then we can write, in view of Proposition 1 and the fact that $\mathcal{L}_n^c$ is an orthogonal projection operator in $\ell^2_{\boldsymbol{w}}(T_{2n})$,

$$\|f-\mathcal{L}_n^c f\|_{\ell^2(X)} \leq \|f-p^*\|_{\ell^2(X)} + \|\mathcal{L}_n^c(p^*-f)\|_{\ell^2(X)}$$

$$\leq \sqrt{M}\,\|f-p^*\|_{\ell^\infty(X)} + \alpha_M(\varepsilon)\,\|\mathcal{L}_n^c(p^*-f)\|_{\ell^2_w(T_{2n})}$$

$$\leq \sqrt{M}\,\|f-p^*\|_{\ell^\infty(X)} + \alpha_M(\varepsilon)\,\|p^*-f\|_{\ell^2_w(T_{2n})}$$

$$\leq \sqrt{M}\,\|f-p^*\|_{\ell^\infty(X)} + \sqrt{M}\,\beta_M(\varepsilon)\,\|p^*-f\|_{\ell^\infty(T_{2n})}$$

$$\leq \sqrt{M}\,(1+\beta_M(\varepsilon))\,E_n(f)\,, \tag{31}$$

that is (29). □

*Remark* 3. Observe that $\beta_M(\varepsilon) \to 1$ as $\varepsilon \to 0$, and quantitatively, $\beta_M(\varepsilon) \approx 1$ for $\varepsilon\sqrt{M} \ll 1$. Then we can write the approximate estimate

$$\mathrm{RMSE}_X(\mathcal{L}_n^c f) \lesssim (2+\varepsilon\sqrt{M}/2)\,E_n(f)\,,\quad \varepsilon\sqrt{M} \ll 1\,, \tag{32}$$

i.e., we substantially recover (26), as well as the size of (21), with a mild requirement on the moment residual error (7).

**Table 1:** Cardinality $m$, Compression Ratio, moment residual and RMSE$_X$ by LS and CATCHLS for the Gaussian $f_1(\rho) = \exp(-\rho^2)$ and the power function $f_2(\rho) = (\rho/2)^5$, $\rho = \sqrt{x^2+y^2}$, where $X$ is the Halton point set of Fig. 1.

| deg $n$ | 3 | 6 | 9 | 12 | 15 | 18 |
|---|---|---|---|---|---|---|
| $N_{2n}$ | 28 | 91 | 190 | 325 | 496 | 703 |
| NNLS: $m$ | 28 | 91 | 190 | 325 | 493 | 693 |
| LP: $m$ | 28 | 91 | 190 | 325 | 493 | 691 |
| $C_{ratio}$ | 200 | 62 | 29 | 17 | 11 | 8 |
| NNLS: residual $\varepsilon$ | 4.9e-14 | 1.2e-13 | 3.4e-13 | 4.3e-13 | 8.8e-13 | 2.5e-12 |
| LP: residual $\varepsilon$ | 2.0e-14 | 3.0e-14 | 9.1e-14 | 9.8e-14 | 7.7e-14 | 7.6e-14 |
| NNLS/LP (cputime ratio) | 0.38 | 0.23 | 0.19 | 0.27 | 0.74 | 0.70 |
| $f_1$: LS | 3.6e-02 | 4.8e-03 | 2.3e-04 | 3.1e-06 | 2.0e-07 | 2.2e-09 |
| NNLS-CATCHLS | 4.1e-02 | 4.9e-03 | 2.3e-04 | 3.2e-06 | 2.0e-07 | 2.2e-09 |
| LP-CATCHLS | 5.0e-02 | 6.1e-03 | 2.7e-04 | 3.5e-06 | 2.0e-07 | 2.3e-09 |
| $f_2$: LS | 2.8e-01 | 2.4e-03 | 1.5e-04 | 2.6e-05 | 6.7e-06 | 2.2e-06 |
| NNLS-CATCHLS | 3.1e-01 | 2.4e-03 | 1.6e-04 | 2.7e-05 | 6.8e-06 | 2.2e-06 |
| LP-CATCHLS | 3.9e-01 | 3.0e-03 | 1.8e-04 | 3.0e-05 | 6.7e-06 | 2.2e-06 |

**Example 2.1.** An example of reconstruction of two bivariate functions with different regularity by LS and CATCHLS on a nonstandard domain (union of four disks) is displayed in Table 1 and Figure 1, where $X$ is a low-discrepancy point set, namely the about 5600 Halton points of the domain taken from 10000 Halton points of the minimal surrounding rectangle. Polynomial least squares on low-discrepancy point sets have been recently studied for example in [20], in the more general framework of Uncertainty Quantification.

We have implemented CATCH subsampling by NonNegative Least Squares (via the `lsqnonneg` Matlab native function) and by Linear Programming (via the `glpk` Octave native function). In the Linear Programming approach, one has to choose a vector $\boldsymbol{c}$ in the target functional. Following [27], we have taken $\boldsymbol{c} = \{x_i^{2n+1} + y_i^{2n+1}\}$, where $X = \{(x_i, y_i)\}$, $1 \le i \le M$, i.e., the vector $\boldsymbol{c}$ corresponds to the polynomial $x^{2n+1} + y^{2n+1}$ evaluated at $X$. There are two reasons for this choice. The first is that (only) in the univariate case, as proved in [27], it leads to $2n+1$ Gaussian quadrature nodes. The second is that the polynomial $x^{2n+1} + y^{2n+1} \notin \mathbb{P}^2_{2n}$, and thus we avoid that $\boldsymbol{c}^t\boldsymbol{u}$ be constant on the polytope defined by the constraints (recall, for example, that for $\boldsymbol{c}^t = (1, \dots, 1)$ we have $\boldsymbol{c}^t\boldsymbol{u} = \sum u_i = M$).

Observe that the CATCH points computed by NNLS and LP show quite different patterns, as we can see in Figure 1. On the other hand they both give a compressed LS operator with practically the same RMSEs as we had sampled at the original points, with remarkable Compression Ratios. The moment residuals appear more stable with LP, but are in any case extremely small with both solvers. On the other hand, at least with the present degree range and implementation (Matlab 7.7.0 (2008) and Octave 3.0.5 (2008) with an Athlon 64 X2 Dual Core 4400+ 2.40GHz processor), NNLS turn out to be more efficient than LP (the cputime varies from the order of $10^{-1}$ sec. at degree $n = 3$ to the order of $10^2$ sec. at degree $n = 18$).

We stress that the compression procedure is function independent, thus we can preselect the re-weighted CATCH sampling sites on a given region, and then apply the compressed CATCHLS formula to different functions. This approach to polynomial least squares could be very useful in applications where the sampling process is difficult or costly, for example to place a small/moderate number of accurate sensors on some region of the earth surface, for the measurement and reconstruction of a scalar or vector field.

## 2.1 From the discrete to the continuum

In what follows we study situations where the sampling sets are discrete models of "continuous" compact sets, in the framework of polynomial approximation. In particular, we have in mind the case where $K$ is the *closure of a bounded open subset* of $\mathbb{R}^d$ (or of a bounded open subset of a lower-dimensional manifold in the induced topology, such as a subarc of the circle in $\mathbb{R}^2$ or a subregion of the sphere in $\mathbb{R}^3$). The so-called "Jackson compacts", that are compact sets where a Jackson-like inequality holds, are of special interest, since there the best uniform approximation error $E_n(f)$ can be estimated by the regularity of $f$; cf. [26].

Such a connection with the continuum has already been exploited in the previous sections, namely on the right-hand side of the LS error estimates, e.g. in (21) and (29). Now, to get a connection also on the left-hand side, we should give some structure to the discrete sampling set $X$. We shall work within the theory of *polynomial meshes*, introduced in [7] and later developed by various authors; cf., e.g., [3, 4, 6, 15, 24] and the references therein.

We recall that a *weakly admissible polynomial mesh* of a compact set $K$ (or of a compact subset of a manifold) in $\mathbb{R}^d$ (or $\mathbb{C}^d$, we restrict here to the real case), is a sequence of finite subsets $X_n \subset K$ such that

$$\|p\|_{L^\infty(K)} \le C_n \|p\|_{\ell^\infty(X_n)}, \ \forall p \in \mathbb{P}^d_n(K), \tag{33}$$

where $C_n = \mathcal{O}(n^\alpha)$, $M_n = \operatorname{card}(X_n) = \mathcal{O}(N^\beta)$, with $\alpha \ge 0$, and $\beta \ge 1$. Indeed, since $X_n$ is automatically $\mathbb{P}^d_n(K)$-determining, then $M_n \ge N = \dim(\mathbb{P}^d_n(K)) = \dim(\mathbb{P}^d_n(X_n))$. In the case where $\alpha = 0$ (i.e., $C_n \le C$) we speak of an *admissible polynomial mesh*, and such a mesh is termed *optimal* when $\operatorname{card}(X_n) = \mathcal{O}(N)$.

Polynomial meshes have interesting computational features (cf. [6]), e.g.

- extension by algebraic transforms, finite union and product
- contain computable near optimal interpolation sets [4, 5]
- are near optimal for uniform LS approximation, namely [7, Thm. 1]

$$\|\mathcal{L}_n\| = \sup_{f \in C(K), f \ne 0} \frac{\|\mathcal{L}_n f\|_{L^\infty(K)}}{\|f\|_{L^\infty(K)}} \le C_n \sqrt{M_n}, \tag{34}$$

where $\mathcal{L}_n$ is the $\ell^2(X_n)$-orthogonal projection operator $C(K) \to \mathbb{P}^d_n(K)$.

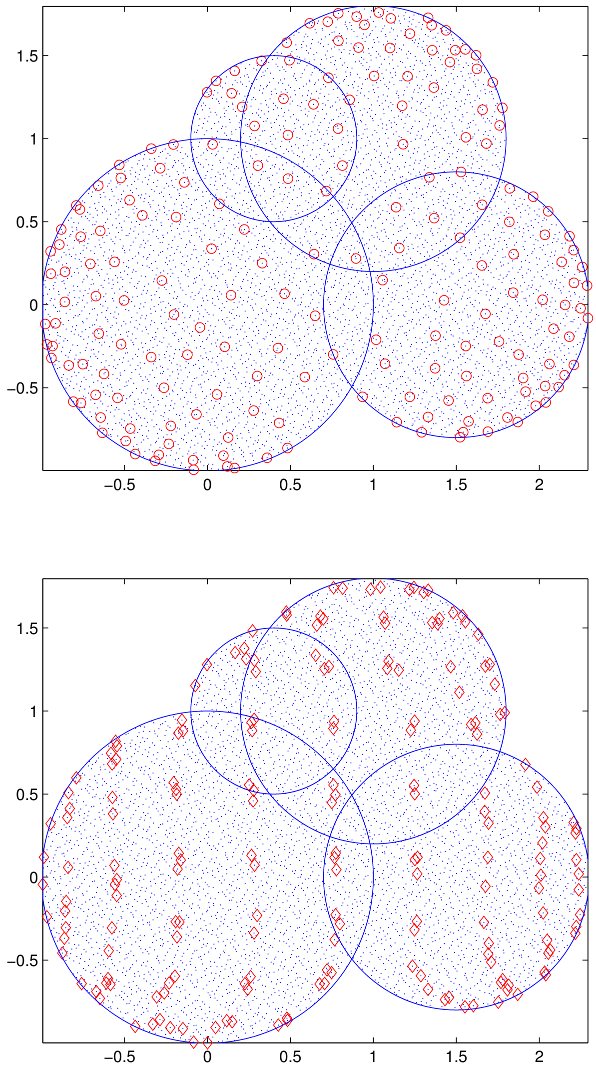



**Figure 1:** Extraction of 190 points for CATCHLS ($n = 9$) from $M \approx 5600$ Halton points on the union of 4 disks: $C_{ratio} = M/m \approx 29$; top: by NonNegative Least Squares as in (5); bottom: by Linear Programming as in (6).

To prove (34), we can write the chain of inequalities

$$\|\mathcal{L}_n f\|_{L^\infty(K)} \leq C_n \|\mathcal{L}_n f\|_{\ell^\infty(X_n)} \leq C_n \|\mathcal{L}_n f\|_{\ell^2(X_n)}$$

$$\leq C_n \|f\|_{\ell^2(X_n)} \leq C_n \sqrt{M_n} \|f\|_{\ell^\infty(X_n)} \leq C_n \sqrt{M_n} \|f\|_{L^\infty(K)} , \tag{35}$$

where we have used the polynomial inequality (33) and the fact that $\mathcal{L}_n f$ is a discrete orthogonal projection. From (34) we get in a standard way the uniform error estimate

$$\|f - \mathcal{L}_n f\|_{L^\infty(K)} \leq (1 + \|\mathcal{L}_n\|)\, E_n(f) \leq \left(1 + C_n \sqrt{M_n}\right) E_n(f) , \tag{36}$$

valid for every $f \in C(K)$.

These properties show that polynomial meshes are good models of multivariate compact sets, in the context of polynomial approximation. Unfortunately, several computable meshes have high cardinality.

In [7, Thm. 5] it has been proved that admissible polynomial meshes can be constructed in any compact set satysfying a Markov polynomial inequality with exponent $r$, but these have cardinality $\mathcal{O}(n^{rd})$. For example, $r = 2$ on convex compact sets with nonempty interior. Construction of optimal admissible meshes has been carried out for compact sets with various geometric structures, but still the cardinality can be very large already for $d = 2$ or $d = 3$, for example on polygons/polyhedra with many vertices, or on star-shaped domains with smooth boundary; cf., e.g., [15, 25].

As already observed, in the applications of LS approximation it is very important to reduce the sampling cardinality, especially when the sampling process is difficult or costly. Thus we may think to apply CATCH subsampling to polynomial meshes, in view of CATCHLS approximation, as in the previous section. In particular, it results that we can substantially keep the uniform approximation features of the polynomial mesh. We give the main result in the following

**Proposition 2.3.** *Let $X_n$ be a polynomial mesh (cf. (33)) and let the assumptions of Proposition 2 be satisfied with $X = X_n$. Then, the following estimate hold*

$$\|\mathcal{L}_n^c\| = \sup_{f\in C(K), f\neq 0} \frac{\|\mathcal{L}_n^c f\|_{L^\infty(K)}}{\|f\|_{L^\infty(K)}} \le C_n \sqrt{M_n}\, \beta_{M_n}(\varepsilon)\,, \tag{37}$$

*provided that $\varepsilon\sqrt{M_n} < 1$, where $\mathcal{L}_n^c f$ is the least squares polynomial at the Caratheodory-Tchakaloff points $T_{2n} \subseteq X_n$. Moreover,*

$$\|p\|_{L^\infty(K)} \le C_n \sqrt{M_n}\, \beta_{M_n}(\varepsilon)\, \|p\|_{\ell^\infty(T_{2n})}\,, \quad \forall p \in \mathbb{P}_n^d(K)\,. \tag{38}$$

*Proof.* To prove (37), we can write the estimates

$$\begin{aligned} \|\mathcal{L}_n^c f\|_{L^\infty(K)} &\le C_n \|\mathcal{L}_n^c f\|_{\ell^\infty(X_n)} \le C_n \|\mathcal{L}_n^c f\|_{\ell^2(X_n)} \\ &\le C_n\, \alpha_{M_n}(\varepsilon)\, \|\mathcal{L}_n^c f\|_{\ell^2_w(T_{2n})} \le C_n\, \alpha_{M_n}(\varepsilon)\, \|f\|_{\ell^2_w(T_{2n})}\,, \end{aligned}$$

using the first estimate in (27) for $p = \mathcal{L}_n^c f$ and the fact that $\mathcal{L}_n^c f$ is a discrete orthogonal projection, and then conclude by (30) applied to $f$.

Concerning (38), we can write

$$\|p\|_{L^\infty(K)} \le C_n \|p\|_{\ell^\infty(X_n)} \le C_n \|p\|_{\ell^2(X_n)}\,,$$

and then apply (27) to $p$. □

By Proposition 3 and (28), we have that the (estimate of) the uniform norm of the CATCHLS operator has substantially the same size of (34), as long as $\varepsilon\sqrt{M_n} \ll 1$. On the other hand, inequality (38) with $\varepsilon = 0$ says that

- *the $2n$-deg CATCH points of a polynomial mesh are a polynomial mesh*

$$\|p\|_{L^\infty(K)} \le C_n \sqrt{M_n}\, \|p\|_{\ell^\infty(T_{2n})}\,, \quad \forall p \in \mathbb{P}_n^d(K)\,. \tag{39}$$

Moreover, (38) shows that such CATCH points, computed in finite-precision arithmetic, are still a polynomial mesh in the degree range where $\varepsilon\sqrt{M_n} \ll 1$. For a discussion of the consequences of (39) in the theory of polynomial meshes see [33].

In order to make an example, in Figure 2 we consider the (high cardinality) optimal polynomial mesh constructed on a smooth convex set ($C^2$ boundary), by the rolling ball theorem as described in [25] (the set boundary corresponds to a level curve of the quartic $x^4 + 4y^4$). The CATCH points have been computed by NNLS as in (5), and the LS and CATCHLS uniform operator norms have been numerically estimated on a fine control mesh via the corresponding discrete reproducing kernels, as discussed in [6, §2.1]. In Figure 2-bottom, we see that the CATCHLS operator norm is close to the LS operator norm, as we could expect from (34) and (37), which however turn out to be large overestimates of the actual norms.

**Acknowledgments.** Work partially supported by the DOR funds and the biennial projects CPDA143275 and BIRD163015 of the University of Padova, and by the GNCS-INdAM. This research has been accomplished within the RITA (Research ITalian network on Approximation).

## References


[1] G. Berschneider and Z. Sasvári. On a theorem of Karhunen and related moment problems and quadrature formulae, Spectral theory, mathematical system theory, evolution equations, differential and difference equations. *Oper. Theory Adv. Appl.*, 221:173–187, 2012.

[2] L. Bittante, S. De Marchi and G. Elefante. A new quasi-Monte Carlo technique based on nonnegative least-squares and approximate Fekete points. *Numer. Math. Theory Methods Appl.*, to appear.

[3] T. Bloom, L. Bos, J.-P. Calvi and N. Levenberg. Polynomial interpolation and approximation in $\mathbb{C}^d$. *Ann. Polon. Math.*, 106:53–81, 2012.

[4] L. Bos, J.-P. Calvi, N. Levenberg, A. Sommariva and M. Vianello. Geometric Weakly Admissible Meshes, Discrete Least Squares Approximation and Approximate Fekete Points. *Math. Comp.*, 80:1601–1621, 2011.

[5] L. Bos, S. De Marchi, A. Sommariva and M. Vianello. Computing multivariate Fekete and Leja points by numerical linear algebra. *SIAM J. Numer. Anal.*, 48:1984–1999, 2010.

[6] L. Bos, S. De Marchi, A. Sommariva and M.Vianello. Weakly Admissible Meshes and Discrete Extremal Sets. *Numer. Math. Theory Methods Appl.*, 4:1–12, 2011.

[7] J.P. Calvi and N. Levenberg. Uniform approximation by discrete least squares polynomials. *J. Approx. Theory*, 152:82–100, 2008.

[8] C. Caratheodory. Uber den Variabilittsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen, *Rend. Circ. Mat. Palermo*, 32:193–217, 1911.

[9] M. Conforti, G. Cornuéjols and G. Zambelli,. Integer programming. Graduate Texts in Mathematics, 271, Springer, Cham, 2014.

[10] R.E. Curto and L.A. Fialkow. A duality proof of Tchakaloff's theorem. *J. Math. Anal. Appl.*, 269:519–532, 2002.

[11] L. Foster. Rank Revealing Code. *available online at*: http://www.math.sjsu.edu/~foster/rankrevealingcode.html.

[12] S. Foucart and H. Rahut. A Mathematical Introduction to Compressive Sensing. Birkhäuser, 2013.

[13] I. Griva, S.G. Nash and A. Sofer. Linear and Nonlinear Optimization, 2nd Edition. SIAM, 2009.

[14] D. Huybrechs. Stable high-order quadrature rules with equidistant points. *J. Comput. Appl. Math.*, 231:933–947, 2009.

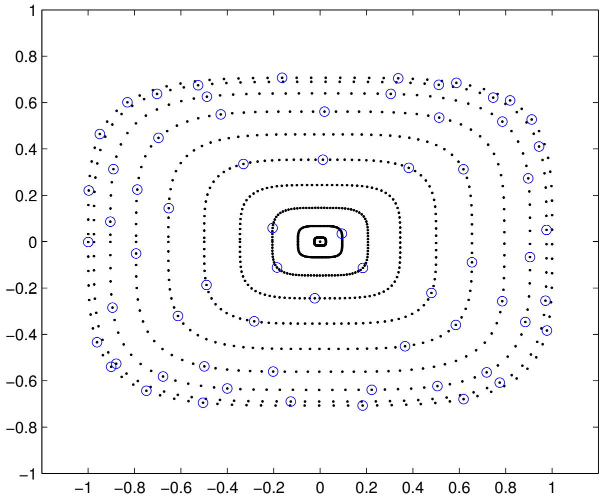
[15] A. Kroó. On optimal polynomial meshes. *J. Approx. Theory* 163:1107–1124, 2011.

[16] C.L. Lawson and R.J. Hanson. Solving leastsquares problems. Revised reprint of the 1974 original. Classics in Applied Mathematics 15, SIAM, Philadelphia, 1995.

[17] C. Litterer and T. Lyons. High order recombination and an application to cubature on Wiener space. *Ann. Appl. Probab.*, 22:1301–1327, 2012.

[18] MATLAB, online documentation and manual (2016), available at: http://www.mathworks.com.

[19] S. Mehrotra and D. Papp. Generating moment matching scenarios using optimization techniques. *SIAM J. Optim.* 23: 963–999, 2013.



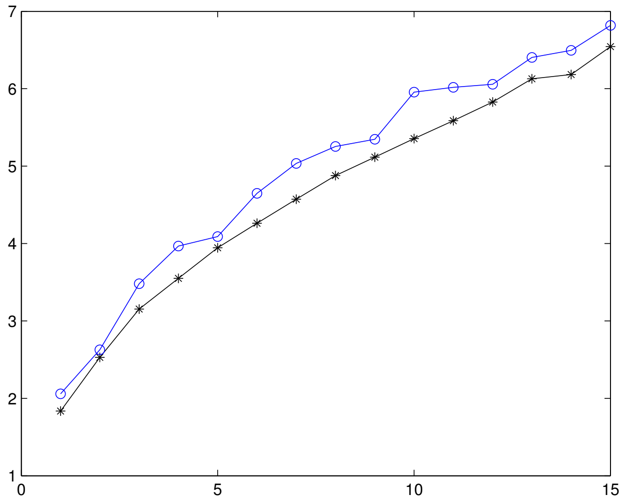




**Figure 2:** Top: polynomial mesh and extracted CATCH points on a smooth convex set ($n = 5$, $C_{ratio} = 971/66 \approx 15$); bottom: numerically evaluated LS ($*$) and CATCHLS ($\circ$) uniform operator norms, for degree $n = 1, \ldots, 15$.

[20] G. Migliorati and F. Nobile. Analysis of discrete least squares on multivariate polynomial spaces with evaluations at low-discrepancy point sets. *J. Complexity*, 31: 517–542, 2015.

[21] GNU Octave. online documentation and manual (2016), avalaible at: https://www.gnu.org/software/octave.

[22] F. Piazzon, A. Sommariva and M. Vianello. Quadrature of Quadratures: Compressed Sampling by Tchakaloff Points. poster presented at the 4th Dolomites Workshop on Constructive Approximation and Applications (DWCAA16), Alba di Canazei (Trento, Italy), Sept. 2016; available online at: http://events.math.unipd.it/dwcaa16/?q=node/6.

[23] F. Piazzon, A. Sommariva and M. Vianello. CATCH: a Matlab/Octave code for Caratheodory-Tchakaloff compression of multivariate discrete measures via Linear or Quadratic Programming, Nov. 2016, available online at: www.math.unipd.it/~marcov/CAAsoft.html.

[24] F. Piazzon and M. Vianello. Small perturbations of polynomial meshes. *Appl. Anal.*, 92:1063–1073, 2013.

[25] F. Piazzon and M. Vianello. Constructing optimal polynomial meshes on planar starlike domains. *Dolomites Res. Notes Approx. DRNA*, 7:22–25, 2014.

[26] W. Pleśniak. Multivariate Jackson Inequality, *J. Comput. Appl. Math.*, 233:815–820, 2009.

[27] E.K. Ryu and S.P. Boyd. Extensions of Gauss quadrature via linear programming. *Found. Comput. Math.*, 15:953–971, 2015.

[28] M. Slawski. Nonnegative least squares: comparison of algorithms (paper and code), available online at: https://sites.google.com/site/slawskimartin/code.

[29] I.H. Sloan. Interpolation and Hyperinterpolation over General Regions. *J. Approx. Theory*, 83:238–254, 1995.

[30] A. Sommariva and M. Vianello. Compression of multivariate discrete measures and applications. *Numer. Funct. Anal. Optim.*, 36:1198–1223, 2015.

[31] V. Tchakaloff. Formules de cubatures mécaniques à coefficients non négatifs. (French) *Bull. Sci. Math.* 81:123–134, 1957.

[32] M. Tchernychova. “Caratheodory” cubature measures. Ph.D. dissertation in Mathematics (supervisor: T. Lyons), University of Oxford, 2015.

[33] M. Vianello. Compressed sampling inequalities by Tchakaloff’s theorem. *Math. Inequal. Appl.*, 19: 395–400, 2016.