

Caratheodory-Tchakaloff Least Squares

Federico Piazzon, Alvise Sommariva and Marco Vianello
Department of Mathematics, University of Padova (Italy)
Email: fpiazzon, alvise, marcov@math.unipd.it

Abstract—We discuss the Caratheodory-Tchakaloff (CATCH) subsampling method, implemented by Linear or Quadratic Programming, for the compression of multivariate discrete measures and polynomial Least Squares.

I. 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 [21], it has then been extended to any measure with finite polynomial moments, cf. e.g. [6].

In this paper we focus on polynomial Least Squares (LS), that is ultimately orthogonal projections with respect to a discrete measure. In Section II and III we show that the original sampling set can be replaced by a smaller one, keeping practically invariant the LS approximation estimates.

We begin by stating a discrete version of Tchakaloff theorem, in its full generality, whose proof is based on Caratheodory theorem itself has a constructive proof (cf., e.g., [5, §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 [9] (univariate) and [20] (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

\[
\text{NNLS : } \begin{cases} 
\min \|V^t u - b\|_2 \\
u \geq 0
\end{cases}
\]

which exists by Theorem 1 and can be computed by standard NNLS solvers based on the Lawson-Hanson method [11], which seeks a sparse solution. Then, the nonvanishing components of such a solution give the weights \(w = \{w_j\}\) 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 lsqlin, with a recent optimization in [18].
The second approach is based instead on Linear Programming via the classical simplex method. Namely, we may think to solve the linear minimum problem

$$\text{LP}: \begin{cases} \min c^T u \\ V^T u = b, \ u \geq 0 \end{cases}$$

(5)

where the constraints identify a polytope (the feasible region) in $\mathbb{R}^M$ and the vector $c$ 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 subspace $T \supset S$ on $K$, i.e. a function in $T$ vanishing on $X$ vanishes everywhere on $K$, then it is sufficient to take $c = \{g(x_i)\}, 1 \leq i \leq M$, 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 $b = V^t\lambda$, 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., [5]). Again, the nonvanishing components of such a vertex give the weights $w = \{w_j\}$ as well as the indexes of the nodes $T = \{t_j\}$ within $X$. This approach was adopted for example in [17] as a basic step to compute, when it exists, a multivariate algebraic Gaussian quadrature formula.

Even though both, the active set method for (4) and the simplex method for (5), 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., [8, Ch. 9]. It is worth quoting here the extensive theoretical and computational results recently presented in the Ph.D. dissertation [22], where Caratheodory reduction of a discrete measure is implemented by Linear Programming, claiming an experimental average cost of $O(N^{3.7})$.

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 $O(N^{2.9})$. The methods are essentially applied to the reduction of Cartesian tensor cubature measures.

In our implementation of CATCH subsampling [15], we have chosen to work with the Octave native Linear Programming solver glpk and the Matlab native Quadratic Programming solver lsqnonneg, that are suitable for moderate size problems, like those typically arising with polynomial spaces ($S = S_\nu = \mathbb{P}_d^\nu$) 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 [22] 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

$$\|V^t u - b\|_2 = \varepsilon > 0,$$

(6)

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:** [14] *Let the assumptions of Theorem 1 be satisfied, let $u$ be a nonnegative vector such that (6) holds, where $V$ is the Vandermonde-like matrix at $X$ corresponding to a $\mu$-orthonormal basis $\{\psi_k\}$ of $S|_X$, and let $(T, w)$ be the quadrature formula corresponding to the nonvanishing components of $u$. 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

$$\left| I_\mu(f) - \sum_{j=1}^m w_j f(t_j) \right| \leq C_\varepsilon E_S(f; X) + \varepsilon \|f\|_2^2(X),$$

(7)

where $C_\varepsilon = 2 (\mu(X) + \varepsilon \sqrt{\mu(X)})$ and

$$E_S(f; X) = \min_{\phi \in S} \|f - \phi\|_{\ell^\infty(X)}.$$  

(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., [7]), 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

$$E_S(f; X) \leq E_S(f; K), \ \forall f \in C(K).$$

(9)

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. [16].

II. CARATHEREODORY-TCHAKALOFF LEAST SQUARES

The case where $(X, \lambda)$ 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. [9], [17], [20] and [1], [12], [22]. 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 $S = S_\nu = \mathbb{P}_d^\nu(K)$, the space of $d$-variante 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 $E_n(f) = E_{p_n^d}(f; K) = \min_{p \in p_n^d(K)} \|f - p\|_{L^\infty(K)}$, for $f \in C(K)$.

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 $p_n^d(X)$, with respect to the scalar product of $\ell^2(X)$, namely
\[
\|f - \mathcal{L}_n f\|_{\ell_2(X)} = \min_{p \in p_n^d(X)} \|f - p\|_{\ell_2(X)}.
\]

Recall that $\|g\|_{\ell_2(X)}^2 = \sum_{i=1}^{M} g^2(x_i) = I_\mu(g^2)$ for every function $g$ defined on $X$, where $\mu$ is the discrete measure supported at $X$ with unit masses $\lambda = (1, \ldots, 1)$.

Taking $p^* \in p_n^d(X)$ such that $\|f - p^*\|_{L^\infty(X)}$ is minimum (the polynomial of best uniform approximation of $f$ in $p_n^d(X)$), we get immediately the classical LS error estimate
\[
\|f - \mathcal{L}_n f\|_{\ell_2(X)} \leq \|f - p^*\|_{\ell_2(X)} \leq \sqrt{M} E_n(f),
\]
where $M = \mu(X) = \text{card}(X)$. In terms of the Root Mean Square Error (RMSE), an indicator widely used in the applications, we have
\[
\text{RMSE}_X(\mathcal{L}_n f) = \frac{1}{\sqrt{M}} \|f - \mathcal{L}_n f\|_{\ell_2(X)} \leq E_n(f).
\]

Now, if $M > N_{2n}$, we stress that here polynomials of degree $2n$ are involved, by Theorem 1 there exist $m \leq N_{2n}$ Caratheodory-Tchakaloff (CATCH) points $T_{2n} = \{t_j\}$ and weights $w = \{w_j\}$, $1 \leq j \leq m$, such that the following basic $\ell^2$ identity holds for every $p \in p_n^d(X)$
\[
\|p\|_{\ell_2(X)}^2 = \sum_{i=1}^{M} p^2(x_i) = \sum_{j=1}^{m} w_j p^2(t_j) = \|p\|_{\ell_2(T_{2n})}^2.
\]

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

Consider the $\ell_n^d(T_{2n})$ LS polynomial $\mathcal{L}_n f$, namely
\[
\|f - \mathcal{L}_n f\|_{\ell_2(T_{2n})} = \min_{p \in p_n^d(X)} \|f - p\|_{\ell_2(T_{2n})}.
\]

Notice that $\mathcal{L}_n$ is a weighted least squares operator; reasoning as in (12) and observing that $\sum_{j=1}^{m} w_j = M$ since $1 \in p_n^d$, we get immediately
\[
\|f - \mathcal{L}_n f\|_{\ell_2(T_{2n})} \leq \sqrt{M} E_n(f).
\]

On the other hand, we can also write the following estimates
\[
\|f - \mathcal{L}_n f\|_{\ell_2(X)} \leq \|f - p^*\|_{\ell_2(X)} + \|\mathcal{L}_n(p^* - f)\|_{\ell_2(X)}
\]
\[
\|\mathcal{L}_n(p^* - f)\|_{\ell_2(X)} = \|\mathcal{L}_n(p^* - f)\|_{\ell_2(T_{2n})} \leq \|f - p^*\|_{\ell_2(T_{2n})},
\]

which shows the most relevant feature of the “compressed” least squares operator $\mathcal{L}_n$ at the CATCH points (CATCHLS), namely that the LS and compressed CATCHLS RMSE estimates (12) and (16) 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 [19]. 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 $p_n^d$ by a discrete weighted $\ell^2$ projection, via a quadrature formula of exactness degree $2n$.

Similarly, here we are approximating an $\ell^2$ projection on $p_n^d$ 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 (6), we state and prove the following

Proposition 2: [14] Let $\mu$ be the discrete measure supported at $X$ with unit masses $\lambda = (1, \ldots, 1)$, let $u$ be a nonnegative vector such that (6) holds, where $V$ is the orthogonal Vandermonde-like matrix at $X$ corresponding to a $\mu$-orthonormal basis $\{\psi_k\}$ of $p_n^d(X)$, and let $(T_{2n}, u)$ be the quadrature formula corresponding to the nonvanishing components of $u$. Then the following polynomial inequality hold for every $p \in p_n^d(X)$
\[
\|p\|_{\ell_2(X)} \leq \sqrt{M} A_M(\varepsilon) \|p\|_{\ell_2(T_{2n})},
\]
\[
A_M(\varepsilon) = \left( 1 - \varepsilon \sqrt{M} \right)^{1/2} \left( 1 + \varepsilon / \sqrt{M} \right)^{1/2},
\]
provided that $\varepsilon \sqrt{M} < 1$.

Corollary 1: [14] Let the assumptions of Proposition 2 be satisfied. Then the following error estimate holds for every $f \in C(K)$
\[
\|f - \mathcal{L}_n f\|_{\ell_2(X)} \leq (1 + A_M(\varepsilon)) \sqrt{M} E_n(f).
\]

Remark 1: Observe that $A_M(\varepsilon) \to 1$ as $\varepsilon \to 0$, and quantitatively, $A_M(\varepsilon) \approx 1$ for $\varepsilon \sqrt{M} \ll 1$. Then we can write the approximate estimate
\[
\text{RMSE}_X(\mathcal{L}_n f) \lesssim (2 + \varepsilon \sqrt{M} / 2) E_n(f), \quad \varepsilon \sqrt{M} \ll 1,
\]
i.e., we substantially recover (16), as well as the size of (12), with a mild requirement on the moment residual error (6).

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 [13], 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
TABLE I: Cardinality $m$, Compression Ratio, moment residual and RMSE of $X$ by LS and CATCHLS for the Gaussian $f_1(\rho) = \exp(-\rho^2)$ and the power function $f_2(\rho) = (\rho/2)^3$, $\rho = \sqrt{x^2 + y^2}$, where $X$ is the Halton point set of Figure 1.

<table>
<thead>
<tr>
<th>Degree</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{2n}$</td>
<td>66</td>
<td>231</td>
<td>496</td>
<td>861</td>
</tr>
<tr>
<td>NNLS: $m$</td>
<td>66</td>
<td>231</td>
<td>493</td>
<td>838</td>
</tr>
<tr>
<td>LP: $m$</td>
<td>66</td>
<td>231</td>
<td>493</td>
<td>837</td>
</tr>
<tr>
<td>$C_{ratio} = M/m$</td>
<td>85</td>
<td>24</td>
<td>11</td>
<td>6.7</td>
</tr>
<tr>
<td>NNLS: residual $\epsilon$</td>
<td>8.3e-14</td>
<td>3.4e-13</td>
<td>1.1e-12</td>
<td>2.7e-12</td>
</tr>
<tr>
<td>LP: residual $\epsilon$</td>
<td>3.6e-14</td>
<td>3.9e-14</td>
<td>7.7e-14</td>
<td>7.3e-14</td>
</tr>
<tr>
<td>CPU ratio: NNLS/LP</td>
<td>0.34</td>
<td>0.23</td>
<td>0.64</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Observe in Table 1 that the CATCH points, computed by NNLS and LP, 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. In this example, NNLS turn out to be more efficient than LP.

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.

### III. 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. [16].

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 (12) and (18). 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 [3] and later developed by various authors; cf., e.g., [2], [3], [10] 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

$$
\|p\|_{L^\infty(K)} \leq C_n \|p\|_{L^\infty(X_n)} , \forall p \in P_n^d(K),
$$

(20)

where $C_n = O(n^\alpha)$, $M_n = \text{card}(X_n) = O(N^\beta)$, with $\alpha \geq 0$, and $\beta \geq 1$. Indeed, since $X_n$ is automatically $P_n^d(K)$-determining, then $M_n \geq N = \text{dim}(P_n^d(K)) = \text{dim}(P_n^d(X_n))$.

In the case where $\alpha = 0$ (i.e., $C_n \leq C$) we speak of an admissible polynomial mesh, and such a mesh is termed optimal when $\text{card}(X_n) = O(N)$.

Polynomial meshes have interesting computational features (cf. [2]), e.g. can be extended by algebraic transforms, finite union and product, contain computable near optimal interpolation sets, and are near optimal for uniform LS approximation, namely [3, Thm. 1]

$$
\|L_n\| = \sup_{f \in C(K), \|f\|_{L^\infty(K)} = 1} \|L_n f\|_{L^\infty(K)} \leq C_n \sqrt{M_n},
$$

(21)

where $L_n$ is the $L^2(X_n)$-orthogonal projection operator $C(K) \to P_n^d(K)$. From (21) we get in a standard way the uniform error estimate for $f \in C(K)$

$$
\|f - L_n f\|_{L^\infty(K)} \leq (1 + C_n \sqrt{M_n}) E_n(f).
$$

(22)

These properties show that polynomial meshes are good models of multivariate compact sets, in the context of polynomial approximation. Unfortunately, several computable
meshes, even optimal meshes, have high cardinality already for \( d = 2 \) or \( d = 3 \), e.g. on Markov compact sets [3, Thm. 5], on polygons/polyhedra with many vertices, or star-shaped domains with smooth boundary [10]. 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 3:** [14] Let \( X_n \) be a polynomial mesh (cf. (20)) and let the assumptions of Proposition 2 be satisfied with \( X = X_n \).

Then, the following estimate hold for the uniform norm of the CATCHLS operator

\[
\| \mathcal{L}_n \| \leq C_n(\varepsilon) \sqrt{M_n}, \quad C_n(\varepsilon) = C_n A M_n(\varepsilon), \quad (23)
\]

provided that \( \varepsilon \sqrt{M_n} < 1 \), where \( \mathcal{L}_n f \) is the Least Squares polynomial at the Caratheodory-Tchakaloff points \( T_{2n} \subseteq X_n \). Moreover, for every \( p \in E_n(K) \)

\[
\| p \|_{L^\infty(K)} \leq C_n(\varepsilon) \sqrt{M_n} \| p \|_{L^\infty(T_{2n})}. \quad (24)
\]

By Proposition 3, we have that the (estimate of) the uniform norm of the CATCHLS operator has substantially the same size of (21), and the the \( 2n \)-degree CATCH points of a polynomial mesh are a polynomial mesh, as long as \( \varepsilon \sqrt{M_n} \ll 1 \).

In the example of Figure 2 we see that the CATCHLS operator norm is close to the LS operator norm, as we could expect from (21) and (23), 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**


