Nearly optimal nested sensors location for polynomial regression on complex geometries

Alvise Sommariva[1] and Marco Vianello[1]

March 5, 2018

Abstract

We compute nearly optimal nested sensors configurations for global polynomial regression on domains with a complex shape, by resorting to the recent CATCH (Caratheodory-Tchakaloff) subsampling technique (sparse discrete moment matching via NonNegative Least Squares). For example, this allows to compress thousands of low-discrepancy sampling points on a many-sided nonconvex polygon into a small subset of weighted points, keeping the size of the uniform regression error estimates with compression ratios of 1-2 orders of magnitude. Since the ℓ^1-norm of the weights remains constant by construction, this technique differs substantially from the most popular compressed sensing methods based on ℓ^1-minimization (such as Basis Pursuit).

Keywords: sensor networks, global polynomial regression, nearly optimal sampling points, Caratheodory-Tchakaloff (CATCH) subsampling, weighted polynomial Least Squares.

1 Introduction

Polynomial regression is a commonly used technique in the analysis of sensor networks data; cf., e.g., [1, 6] and the references therein. In this paper we survey and extend a recent mathematical tool, named CATCH (Caratheodory-Tchakaloff) subsampling, for the *compression of discrete measures* [5, 8, 10, 12] and in particular of discrete polynomial Least Squares [7, 13], by sparse discrete moment matching. This method allows to select from a huge uniform discretization of a given region a much smaller number of (weighted) sampling points, even on a complex shape such as for example a nonconvex polygon with many sides, keeping practically invariant the Least Squares approximation estimates.

The potential applications concern the construction of relatively small sensor networks, which can capture the approximation power of much larger networks in recovering scalar or vector fields on complex geometries. Though the framework is that of compressed (or compressive) sensing [3], the technique differs

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

substantially from the most popular compressed sensing methods. Moreover, we shall discuss here a new feature of CATCH subsampling, the possibility of constructing a *nested family of sampling sets* corresponding to a sequence of polynomial degrees.

The theoretical base of CATCH subsampling is given by two cornerstones of cubature theory and convex analysis, that are *Tchakaloff theorem* on the existence of low cardinality positive cubature formulas exact on polynomial spaces [11], whose discrete version can be proved by *Caratheodory theorem* on finite-dimensional conic/convex combinations [2].

We begin by stating such a discrete version of Tchakaloff theorem (originally proved for integration with respect to absolutely continuous measures); cf. [8] for the literature on generalizations of Tchakaloff theorem.

In the sequel, we shall denote by $\mathbb{P}_n^d$ the space of d-variate polynomials with total degree not exceeding n, and by

$$N = N_n = \dim(\mathbb{P}_n^d) = \binom{n+d}{d} \tag{1}$$

its dimension.

Theorem 1 *Let μ be a multivariate measure whose support is a $\mathbb{P}_n^d$-determining finite set $X = \{x_i\} \subset \mathbb{R}^d$ (i.e., n-degree polynomials vanishing there vanish everywhere), with correspondent positive weights (masses) $\boldsymbol{\lambda} = \{\lambda_i\}$, $i = 1, \ldots, M$, $M = card(X) > N$.*

Then, there exist a cubature formula for the discrete measure μ, with nodes $T_n = \{t_j\} \subset X$ and positive weights $\boldsymbol{w} = \{w_j\}$, $1 \le j \le m \le N$, such that

$$\int_X p(x)\, d\mu = \sum_{i=1}^{M} \lambda_i\, p(x_i) = \sum_{j=1}^{m} w_j\, p(t_j)\ ,\quad \forall p \in \mathbb{P}_n^d\ . \tag{2}$$

Given any polynomial basis $\text{span}(p_1, \ldots, p_N) = \mathbb{P}_n^d$, and considering the *underdetermined* moment system

$$A\boldsymbol{u} = \boldsymbol{b} = A\boldsymbol{\lambda}\ ,\quad A = (V_n(X))^t = (p_k(x_i))\ ,\quad 1 \le k \le N\ ,\quad 1 \le i \le M\ , \tag{3}$$

where $V_n(X)$ is a Vandermonde-like matrix and $\boldsymbol{b}$ is the vector of moments of the polynomial basis with respect to the original discrete measure, the proof follows from the application of Caratheodory theorem to the columns of A.

Indeed, Caratheodory theorem asserts that there exists a *sparse nonnegative* solution $\boldsymbol{u}$ to the system above, whose nonvanishing components (i.e., the weights $\{w_j\}$) are at most N and determine the corresponding reduced sampling points $T_n = \{t_j\}$, that we may term the *Caratheodory-Tchakaloff* (CATCH) *points* of X.

Technically, we could prove the result even on $\mathbb{P}_n^d(X)$ (the polynomial space restricted to X) whose dimension can be smaller than $\dim(\mathbb{P}_n^d)$ (for example when the points lie on an algebraic variety, such as the sphere S^2 in $\mathbb{R}^3$); for a discussion on these aspects see, e.g., [8].

2 Caratheodory-Tchakaloff subsampling

The sense of Theorem 1 from an applicative point of view, is that we can replace the support of a discrete measure by a (much) smaller one (in the applications typically $N \ll M$), if the purpose is to keep invariant the integrals of polynomials of a given degree (or equivalently to keep invariant the moments of any polynomial basis); cf. [10], and [5, 12] in a probabilistic framework. An application is the compression of high-cardinality cubature formulas, i.e. the case where the discrete measure is a cubature formula itself, exact on polynomials of a certain degree, and we replace it by a "smaller" one.

On the other hand, the point of view above gives also an immediate application to polynomial regression. Indeed, discrete polynomial Least Squares are ultimately orthogonal projections of a sampled function on polynomial spaces, with respect to a discrete measure.

We start from a basic ℓ^2-identity. Let $X = \{x_1, \dots, x_M\}$ be a discrete sampling set and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_M)^t$ a vector of corresponding positive weights. If $\mathrm{card}(X) = M > \dim(\mathbb{P}^d_{2n})$, replacing p by p^2 in (2) there are $m \le \dim(\mathbb{P}^d_{2n})$ CATCH points (and weights) such that

$$\|p\|^2_{\ell^2_{\boldsymbol{\lambda}}(X)} = \sum_{i=1}^{M} \lambda_i\, p^2(x_i) = \sum_{j=1}^{m} w_j\, p^2(t_j) = \|p\|^2_{\ell^2_{\boldsymbol{w}}(T_{2n})} \ . \tag{4}$$

Now, given a continuous function f on a compact region $\Omega \supset X$, consider the weighted Least Squares polynomial $L^{\boldsymbol{\lambda}}_X f \in \mathbb{P}^d_n$ and the standard error estimate

$$\|f - L^{\boldsymbol{\lambda}}_X f\|_{\ell^2(X)} = \min_{p \in \mathbb{P}^d_n} \|f - p\|_{\ell^2_{\boldsymbol{\lambda}}(X)} \le \sqrt{\mu(X)}\, \|f - p^*_n\|_\infty \ , \tag{5}$$

where p^*_n is the polynomial of best uniform approximation to f on Ω, and $\mu(X) = \sum_{i=1}^M \lambda_i$. Moreover, consider the weighted Least Squares polynomial $L^{\boldsymbol{w}}_{T_{2n}} f \in \mathbb{P}^d_n$

$$\|f - L^{\boldsymbol{w}}_{T_{2n}} f\|_{\ell^2_{\boldsymbol{w}}(T_{2n})} = \min_{p \in \mathbb{P}^d_n} \|f - p\|_{\ell^2_{\boldsymbol{w}}(T_{2n})} \ . \tag{6}$$

By (4) and the fact that $L^{\boldsymbol{w}}_{T_{2n}} f$ is an orthogonal projection on $\mathbb{P}^d_n$ with respect to a weighted discrete measure supported at T_{2n} (and thus by Bessel inequality $\|L^{\boldsymbol{w}}_{T_{2n}} g\|_{\ell^2_{\boldsymbol{w}}(T_{2n})} \le \|g\|_{\ell^2_{\boldsymbol{w}}(T_{2n})}$ for every g defined on T_{2n}), we easily get the following chain of inequalities

$$\|f - L^{\boldsymbol{w}}_{T_{2n}} f\|_{\ell^2_{\boldsymbol{\lambda}}(X)} \le \|f - p^*_n\|_{\ell^2_{\boldsymbol{\lambda}}(X)} + \|L^{\boldsymbol{w}}_{T_{2n}}(p^*_n - f)\|_{\ell^2_{\boldsymbol{\lambda}}(X)}$$

$$= \|f - p^*_n\|_{\ell^2_{\boldsymbol{\lambda}}(X)} + \|L^{\boldsymbol{w}}_{T_{2n}}(p^*_n - f)\|_{\ell^2_{\boldsymbol{w}}(T_{2n})} \le \|f - p^*_n\|_{\ell^2_{\boldsymbol{\lambda}}(X)} + \|p^*_n - f\|_{\ell^2_{\boldsymbol{w}}(T_{2n})}$$

$$\le \left(\sqrt{\mu(X)} + \sqrt{\|\boldsymbol{w}\|_1}\right) \|f - p^*_n\|_\infty \ , \tag{7}$$

and eventually the $\ell^2_{\boldsymbol{\lambda}}(X)$ error estimate

$$\|f - L^{\boldsymbol{w}}_{T_{2n}} f\|_{\ell^2_{\boldsymbol{\lambda}}(X)} \le 2\sqrt{\mu(X)}\, \|f - p^*_n\|_\infty \ , \tag{8}$$

since $\|\boldsymbol{w}\|_1 = \sum_{j=1}^m w_j = \mu(X)$ by the exactness of (2) on the constants.

In the case where $\boldsymbol{\lambda} = (1, 1, \dots, 1)^t$, we denote by $L_X f$ the standard (unweighted) Least Squares polynomial on X, and we observe that $\mu(X) = M =$

card(X). Then, we can summarize the estimates above by considering the usual notion of *Root Mean Square Deviation* of a prediction $\hat{f}$ to f at a discrete sampling set X, namely

$$\text{RMSD}_X(\hat{f}) = \|f - \hat{f}\|_{\ell^2(X)}/\sqrt{M} \,, \tag{9}$$

writing in view of (5) and (8)

$$\text{RMSD}_X(L_X f) \leq \|f - p_n^*\|_\infty \,, \;\; \text{RMSD}_X(L_{T_{2n}}^{\boldsymbol{w}} f) \leq 2\|f - p_n^*\|_\infty \,, \tag{10}$$

which show that the natural estimates of the RMSD at X have essentially the *same size* using either the unweighted Least Squares polynomial on X, or the weighted Least Squares polynomial on the CATCH subset T_{2n}.

In other words, we expect that CATCH subsampling on $T_{2n} \subset X$ has an approximation power for polynomial regression comparable to sampling on the whole X, with a *compression ratio* $=$ card$(X)/$card(T_{2n}) that can be $\gg 1$.

2.1 Implementation

We turn now to the implementation of CATCH subsampling, that is computing a sparse solution of the moment system (3). We stress that in the application to polynomial regression we deal with $N = N_{2n} = \dim(\mathbb{P}_{2n}^d)$, $A = (V_{2n}(X))^t$ and $\boldsymbol{\lambda} = (1, 1, \ldots, 1)^t$.

Essentially two approaches have been explored: *Linear Programming* (LP) and *Quadratic Programming*. The first consists in minimizing the linear functional $\boldsymbol{c}^t\boldsymbol{u}$ for a suitable choice of the vector $\boldsymbol{c}$, subject to the constraints (3) and $\boldsymbol{u} \geq \boldsymbol{0}$. In fact, the solution is a vertex of the polytope defined by the constraints, which has (at least) $M - N$ null components, cf. e.g. [8, 9, 12].

Observe that a usual choice of compressed sensing (Basis Pursuit, cf. [3]), i.e. minimizing $\|\boldsymbol{u}\|_1$ subject to the constraints, is not feasible in the present context, since $\|\boldsymbol{u}\|_1 = \mu(X)$ for any $\boldsymbol{u}$ satisfying (3) by exactness of (2) on the constants.

Quadratic programming comes into play by solving the *NonNegative Least Squares* (NNLS) problem

$$\text{compute } \boldsymbol{u}^* \;:\; \|A\boldsymbol{u}^* - \boldsymbol{b}\|_2 = \min \|A\boldsymbol{u} - \boldsymbol{b}\|_2 \,, \;\; \boldsymbol{u} \geq \boldsymbol{0} \,, \tag{11}$$

that can be done by the well-known *Lawson-Hanson active set optimization method* [4], which automatically seeks a *sparse solution* and is implemented for example by the `lsqnonneg` native algorithm of Matlab. In practice this gives a nonzero but very small residual $\varepsilon = \|A\boldsymbol{u}^* - \boldsymbol{b}\|_2$. Indeed, it is not difficult to show that the effect of a nonzero residual translates into the substitution of the factor 2 in the second estimate of (10) with a factor $2 + \mathcal{O}\left(\varepsilon\sqrt{\mu(X)}\right)$; cf. [8] for the case $\boldsymbol{\lambda} = (1, 1, \ldots, 1)^t$.

Our numerical experience has shown that NNLS performs better than LP in computing the CATCH weights, at least for moderate degrees n (namely, N in the order of $10^1 - 10^2$); cf. [7, 8]. In the cubature framework a third more combinatorial approach, based on a hierarchical SVD and termed "Recursive Halving Forest", has been proposed in [12]. It is there claimed that it could be more suited than Linear Programming on large scale problems (N in the order of $10^3 - 10^4$), due to an experimental average cost of $\mathcal{O}(N^{2.6})$ instead of $\mathcal{O}(N^{3.7})$, which suggests that such a method could be an alternative also to NNLS and deserve further studies.

2.2 Nested subsampling

A still unexplored feature of CATCH subsampling, is the possibility of constructing a *nested family of sampling sets*, keeping at each step the approximation power of the original discretization mesh X. We focus again on polynomial regression. Indeed, given a sequence of degrees $n_1 < n_2 < \cdots < n_k$, we can compute the nested CATCH sequence $\{T_{2n_j}\}$

$$X \supset T_{2n_k} \supset T_{2n_{k-1}} \supset \cdots \supset T_{2n_2} \supset T_{2n_1} \tag{12}$$

together with the corresponding sequence of positive weight vectors, say $\{\boldsymbol{w}_{2n_j}\}$, by solving backward the sequence of NLLS problems

$$\text{compute } \boldsymbol{u}_{j-1}^* \ : \ \|A_j \boldsymbol{u}_{j-1}^* - \boldsymbol{b}_j\|_2 = \min \|A_j \boldsymbol{u}_{j-1} - \boldsymbol{b}_j\|_2 \ , \ \ \boldsymbol{u}_{j-1} \geq \boldsymbol{0} \ , \tag{13}$$

for $j = k+1, k, \ldots, 2$, where $A_j = (V_{2n_{j-1}}(T_{2n_j}))^t$, $\boldsymbol{b}_j = A_j \boldsymbol{u}_j^*$, and we set $T_{2n_{k+1}} = X$, $\boldsymbol{u}_{k+1}^* = \boldsymbol{\lambda}$. For example, with $\boldsymbol{\lambda} = (1, 1, \ldots, 1)^t$, we have that the RMSD estimate (10) is valid with $n = n_j$ for every j. We may observe that when $\text{card}(X) \gg \dim(\mathbb{P}_{2n_k}^d)$ the computational bulk is essentially in the first step, namely the NNLS problem: $\min \|(V_{2n_k}(X))^t \boldsymbol{u}_k - (V_{2n_k}(X))^t \boldsymbol{\lambda}\|_2$, $\boldsymbol{u}_k \geq \boldsymbol{0}$, since in the other steps the extraction of CATCH points is performed on already "small" subsets.

In sensors location applications, this hierarchy of nested sampling sets could be useful for example to compare regression results obtained by different degrees, or to start from small regression degrees incrementally increasing the number of activated sensors when necessary. We stress that the whole construction of the nested sequence of CATCH sampling sets is independent of the specific function to be sampled, and can be performed once and for all on a given sampling region as a pre-processing stage.

2.2.1 A numerical example

For the purpose of illustration, in Figure 1 we plot the CATCH points extracted from a low-discrepancy set on a nonconvex polygon. Observe that the Compression Ratio (rounded to the nearest integer) is 243 for degree $n = 3$, 75 for $n = 6$ and 36 for $n = 9$. Moreover, in Table 1 we report the moment residuals

$$\varepsilon_{n_j} = \|(V_{2n_j}(T_{2n_{j+1}}))^t \boldsymbol{u}_j^* - (V_{2n_j}(X))^t \boldsymbol{\lambda}\|_2 \ , \tag{14}$$

together with the regression errors

$$E_{n_j} = \text{RMSD}_X(L_X f) \ , \ \ E_{n_j}^{catch} = \text{RMSD}_X(L_{T_{2nj}}^{\boldsymbol{w}} f) \ , \tag{15}$$

on the sequence of degrees $n_j = 3j$, $j = 1, \ldots, 5$, for the smooth test function

$$f(x) = f_1(x) = \exp(-|x - \xi|^2) \ , \tag{16}$$

and the C^2-function

$$f(x) = f_2(x) = |x - \xi|^5 \ , \tag{17}$$

with $x = (x_1, x_2)$ and $\xi = (0.6, 0.6)$.

The weighted Least Squares polynomial regression corresponding to CATCH subsampling has been implemented in Matlab, solving each step of (13) by the

Table 1: Compression ratio, moment residual and RMSD at X on two test functions with different regularity, where X is the Halton point set of Fig. 1.

deg	$n_1 = 3$	$n_2 = 6$	$n_3 = 9$	$n_4 = 12$	$n_5 = 15$
card(T_{2n_j})	28	91	190	325	496
compr ratio	243	75	36	21	14
res ε_{n_j}	5.9e-11	3.3e-11	1.7e-11	8.8e-12	2.8e-12
$E_{n_j}\ f_1$	1.2e-03	8.1e-07	3.1e-09	1.0e-12	1.6e-15
$E^{catch}_{n_j}\ f_1$	1.5e-03	8.8e-07	3.4e-09	1.1e-12	2.1e-15
$E_{n_j}\ f_2$	1.8e-03	2.1e-05	1.8e-06	3.2e-07	9.6e-08
$E^{catch}_{n_j}\ f_2$	2.2e-03	2.3e-05	1.9e-06	3.3e-07	1.0e-07

`lsqnonneg` native algorithm, and then computing (as a solution of an overdetermined system) the coefficient vectors $\boldsymbol{c}_j^*$ in the weighted Least Squares problem

$$\text{compute } \boldsymbol{c}_j^* \ : \ \ \|\sqrt{W_j}\,(B_j\boldsymbol{c}_j^* - \boldsymbol{f}_j)\|_2 = \min \|\sqrt{W_j}\,(B_j\boldsymbol{c}_j - \boldsymbol{f}_j)\|_2 \ , \tag{18}$$

where $\boldsymbol{c}_j \in \mathbb{R}^{N_j}$ with $N_j = \dim(\mathbb{P}^2_{n_j})$, $B_j = V_{n_j}(T_{2n_j})$, $W_j = diag(\boldsymbol{w}_{2n_j})$ is the diagonal matrix of the CATCH weights and $\boldsymbol{f}_j = f(T_{2n_j})$.

We stress that *the RMSD at X remains practically invariant by CATCH subsampling*, which means that we are substantially keeping the approximation power of polynomial regression on a fine discretization of the domain, *using a much lower number of sampling locations*. This feature, together with the nested structure (12)-(13), makes us confident in the usefulness of CATCH subsampling to the construction of nearly optimized sensor networks for global polynomial regression, e.g. in the framework of geospatial analysis.

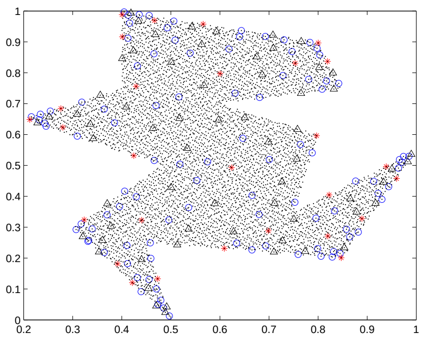

Figure 1: Nested CATCH points for polynomial regression on a 14-side nonconvex polygon, from 6800 Halton points (dots): degree $n_1 = 3$ (28 stars), $n_2 = 6$ (91 = 28 stars plus 63 triangles), $n_3 = 9$ (190 = 28 stars plus 63 triangles plus 99 circles).

References

[1] T. Banerjee, K.R. Chowdhury, D.P. Agrawal, Using polynomial regression for data representation in wireless sensor networks, Int. J. Commun. Syst. 20 (2007), 829–856.

[2] C. Caratheodory, Über den Variabilittsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen", Mathematische Annalen 64 (1907), 95–115.

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

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

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

[6] K. Ohba et al., Environmental Data Recovery using Polynomial Regression for Large-scale Wireless Sensor Networks, in: Proc. of the 5th Intern. Conf. on Sensor Networks - Vol. 1: SENSORNETS, 161-168, 2016, Rome, Italy.

[7] F. Piazzon, A. Sommariva and M. Vianello, Caratheodory-Tchakaloff Least Squares, 2017 International Conference on Sampling Theory and Applications (SampTA 2017), IEEE Xplore Digital Library, DOI: 10.1109/SAMPTA.2017.8024337.

[8] F. Piazzon, A. Sommariva and M. Vianello, Caratheodory-Tchakaloff Subsampling, Dolomites Res. Notes Approx. DRNA 10 (2017), 5–14.

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

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

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

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

[13] M. Vianello, Compressed sampling inequalities by Tchakaloff's theorem, Math. Inequal. Appl. 19 (2016), 395–400.