



FedUni ResearchOnline

<https://researchonline.federation.edu.au>

Copyright Notice

This is the submitted version of the following article:

Sukhorukova, N. (2019) Two curve Chebyshev approximation and its application to signal clustering. Applied Mathematics and Computation, 356, pp.42-49.

Which has been published in final form at:

<https://doi.org/10.1016/j.amc.2019.03.021>

Copyright © 2019 Elsevier Ltd. All rights reserved.

Two curve Chebyshev approximation and its application to signal clustering

Dr N. Sukhorukova, e-mail: nsukhorukova@swin.edu.au
Faculty of Science, Engineering and Technology,
Swinburne University of Technology,
PO Box 218, Hawthorn, Victoria, AUSTRALIA and
Centre for Informatics and Applied Optimization,
Federation University Australia

August 1, 2018

Abstract

In this paper we extend a number of important results of the classical Chebyshev approximation theory to the case of simultaneous approximation of two or more functions. The need for this extension is application driven, since such kind of problems appears in the area of curve (signal) clustering. In this paper we propose a new efficient algorithm for signal clustering and develop a procedure that allows one to reuse the results obtained at the previous iteration without recomputing the cluster centres from scratch. This approach is based on the extension of the classical de la Vallée-Poussin's procedure originally developed for polynomial approximation. In this paper, we also develop necessary and sufficient optimality conditions for two curve Chebyshev approximation, that is our core tool for curve clustering. These results are based on application of nonsmooth convex analysis.

Keywords: Chebyshev approximation, convex analysis, nonsmooth analysis, linear programming, signal clustering

1 Introduction

In signal processing, there is often a need for constructing cluster signal prototypes. Cluster prototypes can be interpreted as summary curves that may replace the whole group of signal segments (clusters of curves), since it is assumed that the curves from the same cluster are similar to each other

in one way or another. Signal prototypes may be used for characterising the structure of the signal segments and their common features (data analysis) and also for reducing the amount of information to be stored (data compression).

k -means is a very fast method developed for clustering points in \mathbb{R}^n . The name “ k -means” was first proposed by James MacQueen in 1967 [Mac]. This method is based on the minimisation of the total dissimilarity function: the sum of squares of the Euclidean distances between the points and the corresponding cluster centres. The theoretical and computational properties of this method as well as its applicability to data analysis, signal processing and data mining problems have been studied for decades [Sp80, BU18].

The classical k -means method contains two steps. Firstly, we assign each point to the cluster with the nearest centre. Secondly, for each cluster, we recompute the centres by minimising the total sum of dissimilarities within each cluster. Then we repeat these two steps until none of the points changes its clustering membership. In the case of the classical k -means algorithm, the second step is reduced to computing the barycentre of the cluster and therefore the algorithm is very fast.

The k -means method can be easily applied to curve clustering if the curves are discretised. This can be done, for example, by treating each time moment t_i , $i = 1, \dots, n$ as a separate coordinate in \mathbb{R}^n , where n may be large.

There have been several modifications of this method, among them the k -medoid method [KR90], where the dissimilarity function is based on other types of distances. A comprehensive review of k -medoid types of methods can be found in [HTF08]. Most k -medoid methods are slower than the classical k -means. At the same time, they may be more appropriate for some specific models and therefore there is a strong need for studying these methods as well.

It is desirable for a k -medoid algorithm, that any cluster prototype is an accurate approximation of each member of the cluster. On the top of this, it is important that the process of recomputing cluster prototypes, when groups of signals move from one cluster to another, is not computationally expensive.

In this paper we suggest a uniform (Chebyshev) approximation based model and therefore our method is from the k -medoid groups of methods. This is a convex optimisation problem. There are several advantages of the proposed implementation model. First of all, it provides an accurate approximation to the group of signals. Second, this problem can be reformulated as a linear programming problem, that can be solved efficiently. Finally,

the proposed approach allows one to compute prototype updates without recomputing from scratch.

The paper is organised as follows. In section 2, we demonstrate that this problem can be formulated as a linear programming problem and study possible ways for solving this problem efficiently. Then, in section 3, we develop the necessary and sufficient optimality conditions for curve clustering. In section 4, we extend the classical de la Vallée-Poussin's procedure to the case of two curve approximation. This procedure plays a vital role in cluster prototype computations. Finally, in section 5, we comment on the results and underline our future research directions.

2 Mathematical formulation

2.1 Prototype construction

Assume that there is a group of l signals $S_1(t), \dots, S_l(t)$, whose values are measured at discrete time moments

$$t_1, \dots, t_N, t_i \in [a, b], i = 1, \dots, N.$$

We suggest to construct a group prototype in the form

$$S(\mathbf{A}, t) = \sum_{i=0}^n a_i g_i(t).$$

The functions $\{g_i(t)\}_{i=0}^n$ are called the *basis functions*. A very common choice of basis functions is the set of monomials

$$\{g_0(t) = 1, g_i(t) = t^i, i = 1, \dots, n\},$$

so that the prototype is modelled as a polynomial function. In this study we are not limiting ourselves to polynomials, but require $\{g_i(t)\}_{i=0}^n$ to form a Chebyshev system, that is

$$\det\{g_i(t_j)\}, i = 0, \dots, n, j = 1, \dots, n, a \leq t_1 < \dots < t_{n+1} \leq b \quad (1)$$

does not vanish for any choice of $t_k, k = 1, \dots, n$, such that

$$a \leq t_1 < \dots < t_{n+1} \leq b.$$

Our aim is to choose the set of parameters for the cluster prototype in such a way that the maximal deviation from each member of the group

on $[a, b]$ is minimal. That is, one has to solve the following optimisation problem:

$$\text{minimise } F(\mathbf{A}) = \sup_{t_i \in [a, b], i=1, \dots, N, j=1, \dots, l} |S_j(t_i) - S(\mathbf{A}, t_i)|, \quad (2)$$

where $\mathbf{A} = (a_0, \dots, a_n) \in \mathbb{R}^{n+1}$, a_k , $k = 0, \dots, n$ are the approximation parameters and also the decision variables.

It can be shown that $F(\mathbf{A})$ is convex, since it is a supremum of convex function. Therefore, we are working with an unconstrained convex problem with $n + 1$ variables. It is also possible to formulate this problem as a linear programming problem.

Let

$$S_{\max}(t) = \max_{j=1, \dots, l} S_j(t)$$

and

$$S_{\min}(t) = \min_{j=1, \dots, l} S_j(t).$$

Then (2) is equivalent to the following

$$\text{minimise } F(\mathbf{A}) = \sup_{t_i \in [a, b], i=1, \dots, N} \max\{S_{\max}(t_i) - S(\mathbf{A}, t_i), S(\mathbf{A}, t_i) - S_{\min}(t_i)\}$$

and therefore the problem has been reduced to a *two curve approximation problem*. Consider an additional variable

$$z = \sup_{t_i \in [a, b], i=1, \dots, N, j=1, \dots, l} |S_j(t_i) - S(\mathbf{A}, t_i)|,$$

then the following linear programming problem is equivalent to (2):

$$\min z \quad (3)$$

subject to

$$S_{\max}(t_i) - S(\mathbf{A}, t_i) \leq z, \quad j = 1, \dots, N; \quad (4)$$

$$S(\mathbf{A}, t_i) - S_{\min}(t_i) \leq z, \quad j = 1, \dots, N. \quad (5)$$

This linear programming problem has $n + 2$ variables and $2N$ constraints. Since N is the number of points where the signal segments are recorded, N may be large. However, it is still more efficient to solve (2) through its linear formulation.

There are many efficient methods for solving linear programming problems. The first efficient linear programming algorithm (simplex method)

was developed by G. Dantzig in 1947 [Dan63]. It was demonstrated in 1972 by [KM72] that the worst-case complexity of the simplex method is exponential. Despite this result, the simplex method is remarkably efficient and included in most linear programming packages. Another important group of algorithms is interior point methods developed in [Kar84], see also [NW06]. The worst-case complexity of interior point methods is polynomial.

2.2 Prototype update

Suppose now that a signal group prototype $S^*(t)$ has been constructed and

$$(\mathbf{A}^*, z^*) = (a_0^*, \dots, a_n^*, z^*)$$

is the corresponding optimal solution to (3)-(5). Let $S(t)$ be an additional signal segment that needs to be included. How can we recompute the group prototype.

One way to proceed is to update $S_{\max}(t)$ and $S_{\min}(t)$ and solve another linear programming problem. There are a number of more efficient ways to approach this problem. In particular, the cluster centre does not have to be computed from scratch at each iteration.

Before moving forward, let us underline three obvious, but yet very important properties.

1. If after moving a number of signal segments in and out of a curve cluster the corresponding S_{\min} and S_{\max} remain unchanged, then the cluster centre remains the same.
2. If after moving a number of signal segments in and out of a curve cluster the corresponding S_{\min} and S_{\max} change, but the points of the maximal deviation remain the same, then the cluster centre remains the same.
3. If there exists a point $t \in [a, b]$, such that

$$S_{\max}(t) - S_{\min}(t) = 2 \sup_{t \in [a, b]} \max\{S^*(t) - S_{\min}(t), S_{\max}(t) - S^*(t)\} \quad (6)$$

then the prototype $S^*(t)$ does not require any update, since the approximation can not be improved.

Since the prototype update has to be recomputed repeatedly, one of our objectives is to demonstrate how the cluster prototype obtained at the previous iteration can be reused for the next one. This can be done, since the

updated constraint matrix contains several rows from the previous iteration. A Sherman-Morrison formula-based approach for such kind of linear programs has been proposed in [Suk15].

One possible approach is to use the solution from the previous iteration as an initial point for the next one. This approach should be exercised with care, since the final point from the previous iteration may be infeasible for the next iteration, since there are several segments moving in and out the group. In section 4 we propose a more robust approach. This approach is based on the well-known de la Vallée-Poussin's procedure [dVP11], originally developed for classical polynomial approximation.

3 Optimality conditions

In this section we develop the necessary and sufficient optimality conditions for two curve approximation, that are based on convex analysis and alternating sequence. Since the objective function is convex, we will apply convex analysis approaches from [Roc70, Zal02]. Before proceeding to two curve approximation, we provide classical results of Chebyshev approximation.

Chebyshev approximation theory is concerned with the approximation of a function f , defined on a (continuous or discrete) domain Ω , by another function s taken from a family \mathcal{F} (for example, polynomials of degree n). At any point $t \in \Omega$ the difference

$$d(t) \triangleq s(t) - f(t)$$

is called the *deviation* at t , and the *maximal absolute deviation* is defined as

$$\|s - f\| \triangleq \sup_{t \in \Omega} |s(t) - f(t)|.$$

The problem of best Chebyshev approximation is to find a function $s^* \in \mathcal{F}$ minimising the maximal absolute deviation over \mathcal{F} . Such a function s^* is called a *best approximation* of f .

The seminal result of approximation theory is Chebyshev's alternation theorem [Che54]. Let P_n be the set of polynomials of degree at most n with real coefficients.

Theorem 3.1. (*Chebyshev alternation theorem, 1854*) *A polynomial $p^* \in P_n$ is a best approximation to a continuous function f on an interval $[a, b]$ if and only if there exist $n + 2$ points $a \leq t_1 < \dots < t_{n+2} \leq b$ and a number $\sigma \in \{-1, 1\}$ such that*

$$(-1)^i \sigma (f(t_i) - p^*(t_i)) = \|f - p^*\|, \forall i = 1, \dots, n + 2.$$

The sequence of points $(t_i)_{i=1,\dots,n+2}$ is called an *alternating sequence*.

Recall that in our study we are not restricted to polynomials. We require cluster prototypes to have a form

$$S(\mathbf{A}, t) = \sum_{i=1}^n a_i g_i(t), \quad (7)$$

where vector $\mathbf{A} = (a_0, \dots, a_n)^T \in \mathbb{R}^{n+1}$ is the vector of parameters (decision variables) and functions $g_i(t)$ are the basis functions (given). The only requirement for the basis functions is to form a Chebyshev system in $[a, b]$.

Now we proceed to the Chebyshev approximation based curve clustering. Supposed that a cluster consists of m signals (S_1, \dots, S_m) , assigned with respect to the shortest distance to the cluster centres. Now we need to recompute the cluster prototype.

First of all, we need to construct two curves:

- $S_{\max}(t) = \max_{i=1,\dots,m} S_i(t)$;
- $S_{\min}(t) = \min_{i=1,\dots,m} S_i(t)$.

Then the parameters of the cluster prototype are the solution of the following optimisation problem:

$$\text{minimise } F(A) = \max_{t \in [a,b]} \{S_{\max}(t) - S(\mathbf{A}, t), S(\mathbf{A}, t) - S_{\min}(t)\}.$$

If the interval $[a, b]$ is discretised ($t_i \in [a, b]$, $i = 1, \dots, N$), a solution can be obtain by solving a linear programming problem (see section 2).

Let

$$\Delta = \max_{t \in [a,b]} \{S_{\max}(t) - S(\mathbf{A}, t), S(\mathbf{A}, t) - S_{\min}(t)\}.$$

Definition 3.1. A point t_k where

$$S_{\max}(t_k) - S(\mathbf{A}, t_k) = \Delta \text{ or } S(\mathbf{A}, t_k) - S_{\min}(t_k) = \Delta$$

is called a *maximal deviation point*.

Definition 3.2. A *maximal deviation point* t_k , such that

$$S_{\max}(t_k) - S(\mathbf{A}, t_k) = \Delta$$

is called a *positive deviation alternating point*, while a point t_k , such that

$$S(\mathbf{A}, t_k) - S_{\min}(t_k) = \Delta$$

is called a *negative deviation alternating point*.

Theorem 3.2. *An approximation $S(\mathbf{A}^*, t)$ is a best approximation to a pair of curves S_{\max} and S_{\min} on an interval $[a, b]$ if and only if at least one of the following conditions holds:*

1. *there exists a time moment $t_k \in [a, b]$, such that*

$$\Delta = S_{\max}(t_k) - S(\mathbf{A}^*, t_k) = S(\mathbf{A}^*, t_k) - S_{\min}(t_k);$$

2. *there exist $n + 2$ points $a \leq t_1 < \dots < t_{n+2} \leq b$ and*

- $\Delta = S_{\max}(t_k) - S(\mathbf{A}^*, t_k) = S(\mathbf{A}^*, t_{k+1}) - S_{\min}(t_{k+1})$, $k = 1, \dots, n$
or
- $\Delta = S(\mathbf{A}^*, t_k) - S_{\min}(t_k) = S_{\max}(t_{k+1}) - S(\mathbf{A}^*, t_{k+1})$, $k = 1, \dots, n$.

Proof:

Since the objective function is convex, its necessary and sufficient optimality condition is as follows:

$$0_{n+1} \in \partial F(\mathbf{A}^*), \quad (8)$$

where \mathbf{A}^* is an optimal set approximation parameters. The subdifferential

$$\partial F(\mathbf{A}^*) = \Delta \text{co} \left\{ \begin{pmatrix} g_0(t_k^+) \\ g_1(t_k^+) \\ \vdots \\ g_n(t_k^+) \end{pmatrix}, - \begin{pmatrix} g_0(t_k^-) \\ g_1(t_k^-) \\ \vdots \\ g_n(t_k^-) \end{pmatrix} \right\}, \quad t_k^+ \in T^+, \quad t_k^- \in T^-, \quad (9)$$

where T^+ is the set of positive deviation alternating points and T^- is the set of negative deviation alternating points. This condition is equivalent to the existence of a positive solution of the following linear system:

$$\mathbf{M}\boldsymbol{\Lambda} = 0_{n+1}, \quad (10)$$

where \mathbf{M} is a matrix whose columns are the gradients at the maximal deviation points (extreme points of the subdifferential ∂F) and $\boldsymbol{\Lambda}$ is a vector whose components are non-negative and the sum of all the components is 1 (that is, there is at least one strictly positive component).

Due to Caratheodory's theorem, there exists a system of at most $n + 2$ points from the subdifferential whose convex combination gives 0_{n+1} . Therefore, the number of columns in matrix \mathbf{M} is at most $n + 2$. Since the basis functions $g_i(t)$, $i = 1, \dots, n$ form a Chebyshev system, the number of

columns in \mathbf{M} can not be less than $n + 2$, otherwise the only solution to (10) is the trivial solution (all components are zeros).

First assume that there is no point where both positive and negative maximal deviation is reached. Arrange maximal deviation points in ascending order:

$$a \leq t_1 < t_2 < \dots < t_{n+2} \leq b.$$

Hence, the system is as follows:

$$\left(\sigma_0 \begin{pmatrix} g_0(t_1) \\ g_1(t_1) \\ \vdots \\ g_n(t_1) \end{pmatrix} \dots \sigma_n \begin{pmatrix} g_0(t_{n+1}) \\ g_1(t_{n+1}) \\ \vdots \\ g_n(t_{n+1}) \end{pmatrix} \right) \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = -\sigma_{n+1} \alpha_{n+1} \begin{pmatrix} g_0(t_{n+2}) \\ g_1(t_{n+2}) \\ \vdots \\ g_n(t_{n+2}) \end{pmatrix}, \quad (11)$$

where $\sigma_k = 1$ for positive alternating points and $\sigma_k = -1$ for negative alternating points ($k = 0, \dots, n + 1$). Since the functions $g_i(t)$, $i = 0, \dots, n$ form a Chebyshev system (the corresponding determinants do not vanish and therefore the sign remains unchanged), applying Cramer's rules, obtain that

$$\sigma_i = -\sigma_{i+1}, \quad i = 0, \dots, n - 1.$$

Now assume that there are points where both positive and negative maximal deviation are reached, that is the first condition holds, the approximation can not be improved, since the maximal deviation can not be made any smaller than

$$\Delta = \frac{1}{2}(S_{\max}(t_k) - S_{\min}(t_k)).$$

This proves the theorem. □

Therefore, the results for classical Chebyshev approximation and uniform approximation based clustering are very similar to each other. In the next section we demonstrate that, despite all these similarities, there are several fundamental differences. Therefore, a careful analysis is required for the extension of the classical results to the case of uniform approximation based clustering.

4 Modelling

In the case of classical Chebyshev approximation there are two important properties.

1. If the basis functions form a Chebyshev system, then the optimal solution is unique.
2. By increasing the degree of the polynomial the maximal error can be made arbitrary small.

The following simple example demonstrate that these two properties are not true for the case of uniform approximation based clustering (two simultaneous curves approximation).

Example 4.1. Let $[a, b] = [0, 1]$, $S_{\max}(t) = 1 - 0.5t$ and $S_{\min}(t) = 0.5t$. Find a best linear approximation for these two curves.

$S(A, t) = 0.5$ is optimal. Moreover, any line

$$0.5 + kt, \quad k \in [-0.25, 0.25]$$

is optimal.

Regardless of the degree of the polynomial, the maximal deviation can not be made below 0.5.

The highlighted fundamental differences between the classical uniform approximation and two curves simultaneous approximation demonstrate that not all the classical results can be generalised. In the rest of this section we show that one of the fundamental results of the classical Chebyshev approximation, namely, de la Vallée-Poussin's procedure can be extended.

The classical de la Vallée-Poussin's procedure for polynomial approximation [dIVP11] has been extended to any basis functions, providing that they form a Chebyshev system (see [KS66] for details). One starts with an initial basis (a system of $n + 2$ points from $[a, b]$) then updates this basis by replacing one or more basis points by some other points from $[a, b]$ and eventually constructs an approximation that satisfies the necessary and sufficient optimality conditions. This procedure consists of two main steps.

1. Construct an approximation that deviates at the basis points from the original function by the same absolute deviation, the signs of the deviations are alternating. This approximation is also called the *Chebyshev interpolation approximation* (also Chebyshev interpolation polynomial). In the case of Chebyshev systems, such an approximation is unique [KS66, Rem57].
2. If there exists a non-basis point whose maximal absolute deviation is greater than it is at the basis points, then this point should be included into the basis, while one of the basis points should be removed.

Namely, the removal has to be done in such a way that the deviation signs at the new basis are alternating. It is enough to remove a neighbouring basis point with the same deviation sign or, if there is only one neighbouring point and the deviation sign at this point is opposite, remove the most extreme basis point from the opposite side of the interval.

By repeating these steps, one eventually arrives to the situation where there is no point that should be moved into the basis and therefore the current approximation is optimal, since the necessary and sufficient optimality conditions are satisfied. The basis exchange rule implies that each updated basis leads to the Chebyshev interpolation approximation, whose absolute deviation at the basis points is at least as large as the absolute deviation at the previous basis.

In the case of two curve approximation, the definition of basis remains the same: any set of distinct $n + 2$ points from $[a, b]$ forms a basis. Basis points are also called *nodes*.

In our study we are working with two curves and therefore there may be points whose absolute deviation is maximal and both deviation signs are active. To be able to work with this kind of points, we introduce the following definitions.

Definition 4.1. *A node where the absolute deviation is maximal and both positive and negative deviation is reached is called a double node.*

Lemma 4.1. *Assume that the following two properties hold.*

1. *An original basis is chosen in such a way that there is no double node.*
2. *The replacement of one of the basis point by a maximal absolute deviation point leads to a basis without double nodes.*

Then the classical de la Vallée-Poussin's procedure can be extended to the case of two curve approximation for Chebyshev systems and terminate at a point where the condition 2 of Theorem 3.2 is satisfied.

Proof:

We start with an arbitrary collection of $n + 2$ points from $[a, b]$ as the initial basis $T^0 = \{t_0^0, \dots, t_{n+1}^0\}$. For each point $t \in T^0$ from this basis assign one of the values $S_{\max}(t)$ or $S_{\min}(t)$ in such a way that there is no pair of neighbouring points assigned to the same curve (that is, construct $F(t)$). In this case there exists a unique Chebyshev interpolation approximation that

deviates at the basis points from the assigned values by the same absolute value and the signs of the deviations are alternating (similar to one curve approximation).

The generalisation of the basis update step requires to consider two different possibilities.

1. The absolute maximal deviation outside of the basis does not exceed the absolute deviation at the basis points. In this case the condition 2 of Theorem 3.2 is satisfied and therefore the obtained approximation is optimal.
2. There exists a point outside of the basis where the absolute deviation is higher than it is at the basis points. The basis update rule is the same as it is for one curve approximation: the maximal absolute deviation point replaces the adjacent basis point with the same deviation sign or, if there is no same sign adjacent basis point, this point replaces the furthest basis point regardless of its deviation sign. Similar to one curve approximation [KS66], this basis update leads to a Chebyshev interpolation approximation with a higher absolute deviation. This can be demonstrated by using $F(t)$.

Since we only consider the situation where there is no point which is both positive and negative alternation point, the procedure is fully extended.

□

Our next step is to demonstrate how the assumptions from Lemma 4.1 can be removed. In Theorem 4.1 we demonstrate how to construct an optimal approximation where the condition 2 of Theorem 3.2 is satisfied or, if this is not possible, how to find a basis, such that the assumptions of Lemma 4.1 are satisfied.

Theorem 4.1. *The classical de la Vallée-Poussin's procedure can be extended to the case of two curve approximation for Chebyshev systems.*

Proof: We start by identifying points t , where the difference

$$S_{\max}(t) - S_{\min}(t)$$

reaches its maximal value in $[a, b]$. We will call these points maximal difference points. Note that there exists at least one maximal difference point. Let

$$\Delta^* = 0.5 \max_{t \in [a, b]} (S_{\max}(t) - S_{\min}(t)),$$

where Δ^* is a lower bound for the optimal maximal deviation.

Consider two possibilities.

1. Assume that the number of maximal difference points is $l \geq n + 1$. Choose any $n + 1$ maximal distance points. Since the basis functions form a Chebyshev system, there exists a unique approximation that passes through

$$0.5(S_{\max}(t) - S_{\min}(t))$$

at the chosen $n + 1$ maximal distance points. If there is no point whose maximal absolute deviation is strictly higher than Δ^* , then the obtained approximation is optimal.

Now assume that there exists a point t^* whose absolute maximal deviation is greater than Δ^* . This point, together with the chosen $n + 1$ points, form a basis and the alternation order between $S_{\max}(t)$ and $S_{\min}(t)$ is determined by the maximal absolute deviation at t^* . By construction, this point can not have both positive and negative maximal deviation and therefore the alternation order is determined uniquely. The absolute maximal deviation is higher than Δ^* , therefore, there is no double node and any possible basis update can not lead to the presence of double nodes. By Lemma 4.1, the procedure is extended.

2. Assume that the number of maximal difference points $l \leq n$. Add any arbitrary chosen $n + 1 - l$ points. Assign all these $n + 1$ points to $S_{\max}(t)$ or $S_{\min}(t)$ in an alternating way. Since the basis functions form a Chebyshev system, there exists a unique approximation that passes through the specified $(n + 1)$ points, the absolute deviation is Δ^* . If there is no points where the maximal absolute deviation exceeds Δ^* , the current approximation is optimal. Otherwise, the maximal deviation point together with the current $(n + 1)$ points form a basis ($n + 2$ points in total, since none of the points can coincide). Therefore, there exists a unique Chebyshev interpolation approximation, whose absolute deviation exceeds Δ^* at the basis points. Therefore, there is no double node and any possible basis update can not lead to the presence of double nodes. By Lemma 4.1, the procedure is extended.

Therefore, the procedure is fully extended.

□

Coming back to the cluster prototype update, the basis point, obtained at the previous iteration can be used as an initial basis for the following one. This approach is similar to the usage of the solution obtain at a previous iteration as an initial point for the next one, but there is no risk of getting an

infeasible point. The proposed approach is very efficient when the number of curves moving in or out the curve cluster is not very large.

5 Conclusions and further research directions

This paper extends the classical Chebyshev approximation results to the case of curve clustering. The idea to extend these results comes from the application of optimisation and approximation to signal clustering and can be viewed as an extension of the well-known k -means method to curve clustering in uniform metric (k -medoid). Therefore, apart from the theoretical significance for optimisation and approximation, the results have potential applications in the area of signal processing and other areas of engineering and science.

There are still a number of open problems. The following conjecture is one of our future research directions.

Conjecture 5.1. *The two curve approximation de la Vallée-Poussin's procedure is equivalent to the dual simplex method applied to the corresponding linear programming problem.*

This conjecture is of both theoretical and applied significance, since it underlines an efficient approach for curve clustering.

We also intend to study the theoretical properties of the original objective function for k -medoids, similar the study proposed in [BU18] for the Euclidean norm-based distances.

References

- [BU18] Adil Bagirov and Julien Ugon. Nonsmooth dc programming approach to clusterwise linear regression: optimality conditions and algorithms. *Optimization methods and software*, 33(1):194–219, 2018.
- [Che54] PA Chebyshev. Théorie des mécanismes connus sous le nom de parallélogrammes, 1ère partie. *Mémoires présentés à l'Académie impériale des sciences de Saint-Pétersbourg par divers savants*, 1854.
- [Dan63] George B. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, New Jersey, 1963.
- [dlVP11] de la Valle Poussin. C. sur la methode de l'approximation minimum. *Ann. Soc. Sci. Bruxelles*, 35:1–16, 1911.

- [HTF08] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Springer Series in Statistics, 2008.
- [Kar84] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, December 1984.
- [KM72] V. Klee and G. J. Minty. How good is the simplex algorithm? In O. Shisha, editor, *Inequalities III*, pages 159–175. Academic Press Inc., New York, 1972.
- [KR90] L. Kaufman and P.J. Rousseeuw. *Finding Groups in Data: an introduction to cluster analysis*. Wiley, 1990.
- [KS66] Samuel Karlin and William Studden. *Tchebycheff systems, with applications in analysis and statistics*. Interscience Publishers New York, 1966.
- [Mac] James MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*.
- [NW06] J. Nocedal and S. Wright. *Numerical Optimization*. Springer-Verlag, Inc., New York, 2006.
- [Rem57] E.Ya Remez. General computational methods of Chebyshev approximation. *Atomic Energy Translation*, 4491, 1957.
- [Roc70] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.
- [Sp80] H. Späth. *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*. Ellis Horwood Limited, Chichester, 1980.
- [Suk15] Nadezda Sukhorukova. An interior point method and sherman–morrison formula for solving large scale convex quadratic problems with diagonal hessians. *ANZIAM J.*, 56(E):E1–E21, 2015.
- [Zal02] C. Zalinescu. *Convex Analysis in General Vector Spaces*. World Scientific, 2002.