Treating geospatial complex data by compression and reduced order methods[1]

Stefano De Marchi
Department of Mathematics "Tullio Levi-Civita"
University of Padova (UNIPD) – Italy
Wroclaw, September 19, 2018

DIPARTIMENTO MATEMATICA

[1] GeoEssential ERA-PLANET team at UNIPD: F. Aiolli, S. De Marchi, W. Erb, E. Perracchione, F. Piazzon, M. Polato, **M. Putti**, A. Sperduti, M. Vianello.

Complexity of data

Complexity is a tougher nut to crack!

$\hookrightarrow$ complexity of the data is as much of a difficulty in making use of the data as is their size/dimension$\hookleftarrow$

Although people understand intuitively that complexity is a real problem in data analysis
it is not always an easy notion to define.

In many cases, we recognize the complexity when we see it!

Complex and topological data analysis:
`https://www.ayasdi.com/blog/author/gunnar-carlsson/`
and plenary by Marian Mrozek on *Combinatorical topological dynamics*.

Outline

Goals of the GeoEssential project

1. Efficient method for image compression well suited for geospatial data modelling.
2. From several data (temperature, soil humidity, satellite images, ...) create a model to forecast the evolution of the dynamics in time and evaluate related uncertainties.

- For the first item, we have developed an efficient polynomial-based scheme, which enables us to compress images.
- For the second item, both Radial Basis Function (RBF)-based reduced order methods and machine learning tools are used.

Image compression

Theoretical basis [Piazzon et al. 2017]

Theorem (Discrete Caratheodory-Tchakaloff)

Let μ be a discrete multivariate measure on $\mathbb{R}^d$,

$$\mu := \sum_{i=1}^{M} \lambda_i \delta_{x_i}, \quad \lambda_i > 0, x_i \in \mathbb{R}^d,$$

supported in $X = \{x_1, \ldots, x_M\} \subset \mathbb{R}^d$ and let $S := \mathrm{span}\{\phi_1, \phi_2, \ldots, \phi_L\}$ a linear space of functions that are continuous on a compact neighborhood of X with $N = \dim(S_{|X}) \leq L$.
Then, there exists a quadrature rule for μ s.t. $\forall f \in S_{|X}$

$$\boxed{\int_X f d\mu := \sum_{i=1}^{M} f(x_i)\lambda_i = \sum_{j=1}^{m} f(t_j)\omega_j, \quad .}$$

with nodes $\{t_j\}_{j=1}^{m} \subset X$ and positive weights ω_j with $m \leq N \leq L$

Obs: this is a subsampling of disc...

Image compression

Computational aspect, I

The problem of finding the subspace is ”suggested” by the Tchakaloff's theorem (1959).

- choose any $c \in \mathbb{R}^M$ linear independent w.r.t. the columns of V^t (V=Vandermonde-like) and solve

$$\begin{cases} \min_{\tilde{\omega} \geq 0} \langle c; \tilde{\omega} \rangle, \\ V^t \tilde{\omega} = b, \end{cases}$$

  where

$$V^t \tilde{\omega} = b \; (:= V^t \lambda),$$

Obs

- The feasible region is a polytope.
- The minimum of the objective is achieved on a vertex (i.e. sparsity).

Image compression

Computational aspect, II

The minimum problem can be solved in two (alternative) ways

1. Simplex method (which is the standard solver) (or basis pursuit algorithm).

2. The Lawson-Hanson (non-negative least squares) algorithm for the relaxed problem

$$\begin{cases} \min \|V^t \tilde{\omega} - b\|_2, \\ \tilde{\omega} \geq 0, \end{cases}$$

The Lawson-Hanson algorithm finds sparse solutions ... and this is the case.

$\Longrightarrow$ Both algorithms are thinning procedures for image compression with $r = M/m \geq M/N \gg 1$.

Example 1

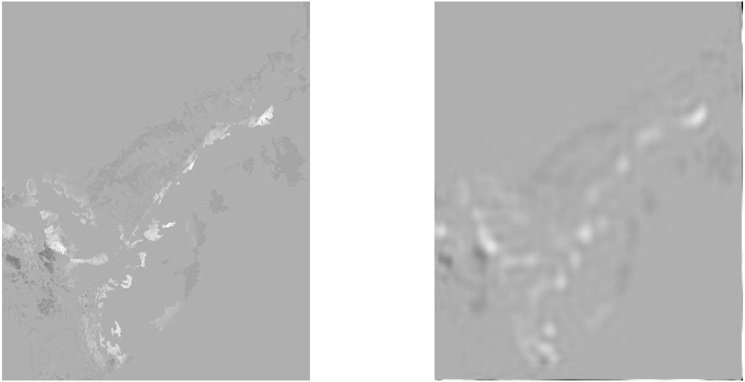


Figure: Example of image compression. The compression factor is about 70.

Example 2

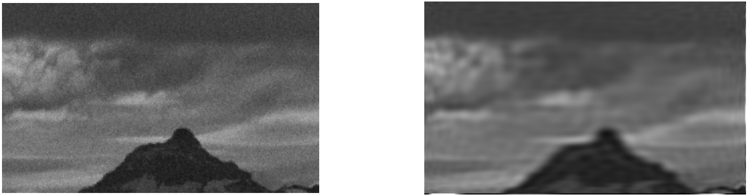


Figure: Example of image compression. The compression factor is about 100.

Example 3

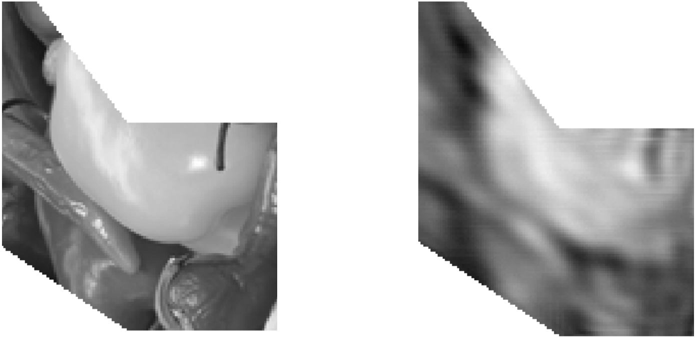


Figure: Example of image compression. The compression factor is about 100.

Time evolution, prediction by RBF-based model

Notation

- $\mathcal{X}_N = \{x_i, i = 1, \ldots, N\} \subseteq \Omega$: set of distinct, scattered data sites (nodes) of $\Omega \subseteq \mathbb{R}^M$
- $\mathcal{F}_N = \{f_i = f(x_i), i = 1, \ldots, N\}$, data values (or measurements), obtained by sampling some (unknown) function $f : \Omega \longrightarrow \mathbb{R}$ at the nodes x_i,

Scattered data interpolation problem

Find a function $R : \Omega \longrightarrow \mathbb{R}$ s.t.

$$\boxed{R_{|\mathcal{X}_N} = \mathcal{F}_N \, .}$$

RBF interpolation: consider $\phi : [0, \infty) \to \mathbb{R}$ and form

$$\boxed{R\left(x\right) = \sum_{k=1}^{N} c_k \phi\left(\|x - x_k\|_2\right), \quad x \in \Omega.}$$

Uniqueness of the solution

- The problem reduces to solving a linear system $Ac = f$, with

$$(A)_{ik} = \phi\left(\|x_i - x_k\|_2\right), \quad i,k = 1,\ldots,N.$$

- The problem is well-posed if ϕ is strictly positive definite[2]

Kernel notation

Let $\Phi : \mathbb{R}^M \times \mathbb{R}^M \longrightarrow \mathbb{R}$ be a strictly positive definite kernel. Then A becomes

$$(A)_{ik} = \Phi\left(x_i, x_k\right), \quad i,k = 1,\ldots,N.$$

[2] We remark that the uniqueness of the interpolant can be ensured also for the general case of strictly conditionally positive definite functions of order L by adding a polynomial term.

Popular radial basis functions

In Table 1, we present several RBFs. Here $r := \|\cdot\|_2$ and ε the shape parameter.

$\phi(r) = e^{-(\varepsilon r)^2}$	Gaussian C^∞	G
$\phi(r) = (1 + (\varepsilon r)^2)^{-1/2}$	Inverse MultiQuadric C^∞	IMQ
$\phi(r) = e^{-\varepsilon r}$	Matérn C^0	M0
$\phi(r) = e^{-\varepsilon r}(1 + \varepsilon r)$	Matérn C^2	M2
$\phi(r) = e^{-\varepsilon r}(3 + 3\varepsilon r + (\varepsilon r)^2)$	Matérn C^4	M4
$\phi(r) = (1 - \varepsilon r)_+^2$	Wendland C^0	W0
$\phi(r) = (1 - \varepsilon r)_+^4\,(4\varepsilon r + 1)$	Wendland C^2	W2
$\phi(r) = (1 - \varepsilon r)_+^6\left(35(\varepsilon r)^2 + 18\varepsilon r + 3\right)$	Wendland C^4	W4

Table: most popular RBFs

Model Reduction methods

Motivation

- Smaller model dimension, reduced requirements
- Similar precision, error control
- Automatic reduction, not “manual”

Applications: parametric PDEs, ODEs, adaptive grids, parallel computing and HPC....

Reference: `www.haasdonk.de/data/drwa2018` tutorial given at the*Dolomites Research Week on Approximation 2018, Canazei (I) 10-14/9/2018*.

Reduced order methods

Defintion

The problem can be visualized in a nested diagram

- Given a set of N data, the aim is finding a suitable subspace (reduced), spanned by $m \ll N$ (functions and) centers.

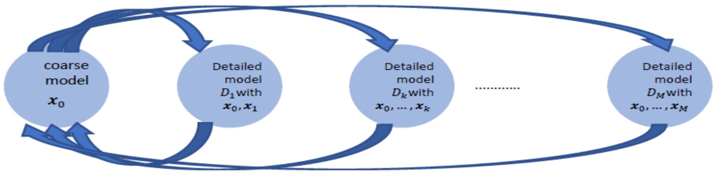


Figure: Communication diagram for macro and microscale models.

Point selection procedure

Greedy-based approach [Haasdonk, Santin 2017]

Consider a function $f : \Omega \longrightarrow \mathbb{R}$ and denote by R its RBF interpolant on X_N centers. The procedure can be summarized as follows:

starting $X_0 \neq \emptyset$

$k \geq 1$ Determine the sequence

$$x_k = \underset{x \in X_N \setminus X_{k-1}}{\arg\max} \underbrace{|f(x) - R(x)|}_{P_{X_N,\phi}(x)},$$

and form

$$X_k = X_{k-1} \cup \{x_k\}.$$

repeat Continue until a suitable maximal subspace of m terms, $m \ll N$, is found.

$P_{X_N,\phi}$: power function

Filtering by Ensemble Kalman filter (**EnKF**)

The EnKF works for non-linear models, takes into account unavoidable uncertainty (noise) in the measurements and enables us to get an estimate for the next step, say t_{k+1}.

- EnKF is a generalization of the well-known Extended KF
- When the dynamics is linear the Kalman filter provides an optimal estimate of the state, while EnKF for non-linear model is suboptimal.

More details in [GeoEssential Report 1, Sept. 2018].

VOSS

- As data set for the current study of time series, we consider the data collected in the South-Eastern part of the Veneto Region and available at

  `http://voss.dmsa.unipd.it/`.

- This data set has been created for an experimental study of the organic soil compaction and prediction of the land subsidence related to climate changes in the South-Eastern area of the Venice Lagoon catchment (VOSS - Venice Organic Soil Subsidence).
- The data were collected with the contribution of the University of Padova (UNIPD) from 2001 to 2006.

Example I

RBF: Matern M6.

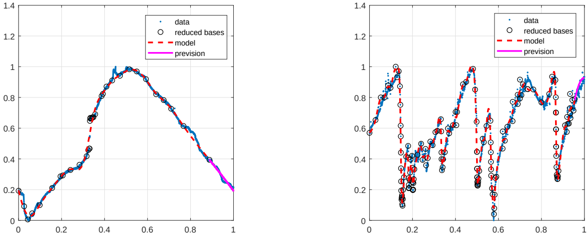


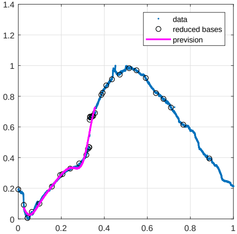
Figure: Graphical results via RBF-reduced order methods coupled with EnKF. Left: temperature data, Right: potentiometer samples.

N	M	B	$N-B$	RMSE	MAE
7819	34	6944	875	1.30E − 02	2.28E − 02
2075	158	1991	94	3.81E − 02	4.25E − 02

Figure: Accuracy for RBF-reduced order methods and Kalman filter. *B* indicates the index of the last basis extracted.

Example II

RBF: Matern M6.



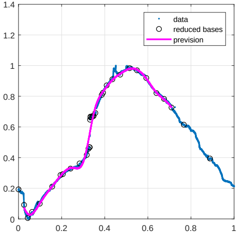


Figure: The RBF-reduced order methods and Kalman filter applied iteratively on the temperature data set. The figure shows the progresses of the algorithm for two different time steps (i.e. data assimilation).

Support Vector Machine (SVM)

Kernel-based methods are one of the most used machine learning approaches.

Support Vector Machine (SVM) is the most famous and successful kernel method!

The basic idea behind this kind of schemes is related to the so-called kernel trick which allows to implicitly compute vector similarities/classification (defined in terms of dot-product)

Time evolution-prediction by ML

Learning with kernels [Schőlkopt and Smola 2001]

- In ML kernels are defined as $\Phi(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$, where $\phi : \Omega \to H$ (future map) maps the vectors $\mathbf{x}, \mathbf{y}$ to a (higher dimensional) feature (or embedding) space H [Shawe-Taylor and Cristianini 2004].
- The main idea consists in using kernels to project data points in an higher dimensional space where the task is "easier" (for example linear): the "Kernel Trick".

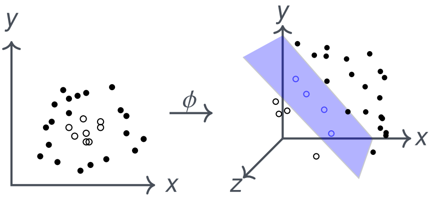


Figure: "Kernel Trick": binary classification by a future map $\phi : \mathbb{R}^2 \to \mathbb{R}^3$.

Regression with SVM: SVR machine

- When considering a regression problem, the idea is to define an ϵ-tube:
- predictions in absolute error $> \epsilon$ are linearly penalized, otherwise they are not considered as errors.

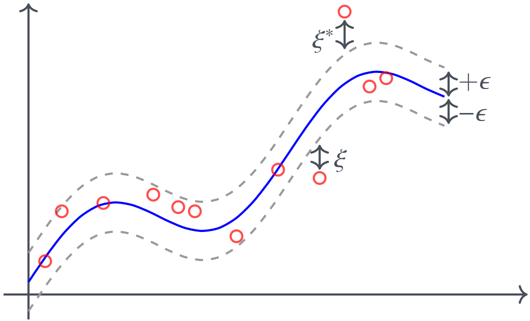


Figure: Graphical illustration of the SVR ϵ-tube. Data points inside the tube do not contribute to the loss, while points outside the tube have a cost proportional to the distance from the tube.

SVR model problem

Given the training set $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$, ($\mathbf{x}_i$, support vectors) the ϵ-tube idea leads to the following primal linear problem:

$$R(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{w} + b.$$

To determine $\boldsymbol{w}$ and b, we solve the constrained optimization problem with regularization parameter C and slack variables ξ_i, $i = 1, \dots, N$,

$$\min_{\boldsymbol{w}, b, \xi, \xi^*} \left[\frac{1}{2} \boldsymbol{w}^\top \boldsymbol{w} + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \right],$$

subject to

$$\boxed{\begin{array}{ll} R(\boldsymbol{x}_i) - f_i \leq \epsilon + \xi_i, & \\ f_i - R(\boldsymbol{x}_i) \leq \epsilon + \xi_i^* & i = 1, \dots, N, \\ \xi_i \, \xi_i^* \geq 0, & \end{array}}$$

Remarks

- $C \geq 0$ represents the so-called *trade-off parameter* and it is indeed a smoothing parameter.
- The *parameter* $\epsilon \geq 0$ indicates the width of the *tube* in which the samples fall into without being counted as errors.

Data pre-processing: Multi SVR, sliding window

- Take $t_1, \ldots, t_n$ and a window size $k > 0 \in \mathbb{N}$,
- Determine $n - k + 1$ training vectors s.t. $\boldsymbol{t}_i = (t_i, \ldots, t_{j+k-1})^\intercal$ for $j = 1, \ldots, n - k + 1$.
- The predicted value, say $\boldsymbol{t}_i$ with associated target value f_i, is t_{j+k}.

In general, if we want to build a model for predicting the data point at Δt steps in the future, the associated target value to $\boldsymbol{t}_i$ will be $f_i = t_{j+k+\Delta t-1}$.

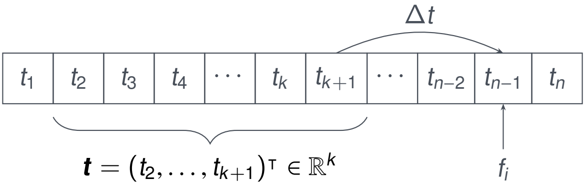


Figure: Illustration of the sliding window mechanism.

An experiment

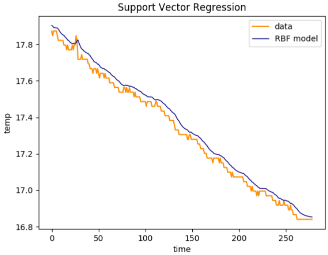


Figure: Prediction using single SVR model, with RMSE=0.04.

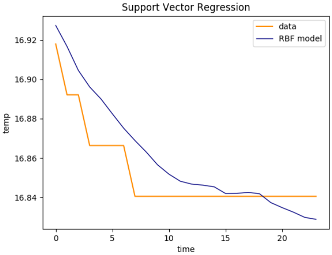
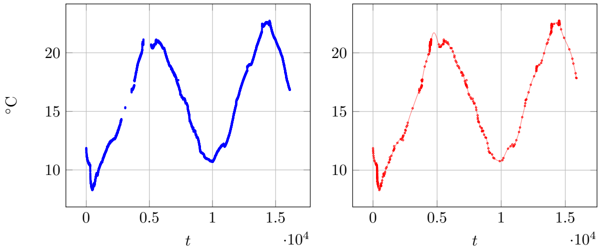


Figure: Prediction using 24 SVR models, with RMSE=0.14.

Note: the RBF used is the Gaussian.

Another experiment

- The total number data points is 16114 in which 12 are not considered in the training phase since are the points we want to forecast.
- From the total of 16102 training points only 14637 are valued, the remaining ones are missing.
- Using RBF we extract 393 basis.



(a) Temperature raw data points

(b) Extracted basis

SVR

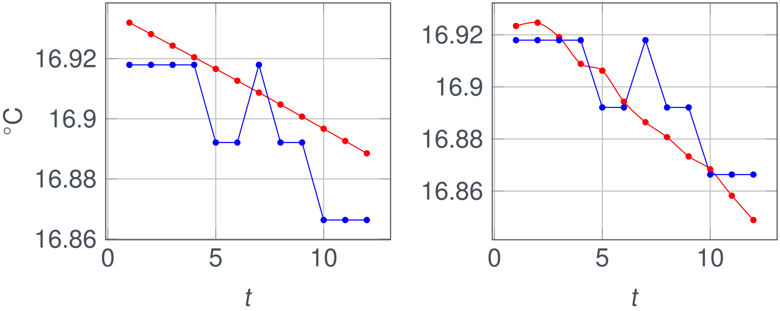


Figure: SVR with reduced (left) and classical (right) training for environmental data.

MSVR

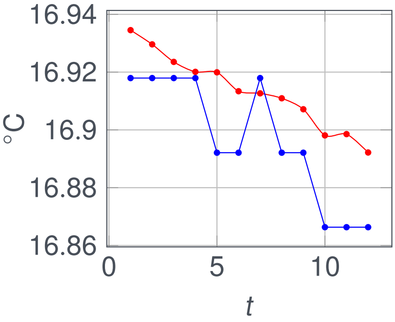


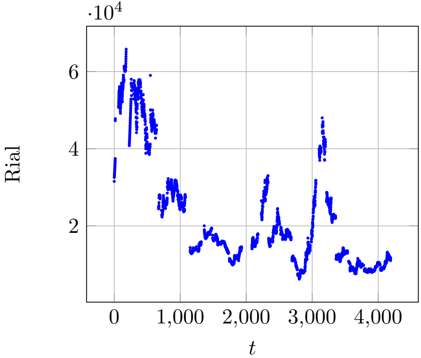
Figure: MSVR, 48h as timing window, with reduced (left) and classical (right) training for environmental data.

Obs: due "moderate" smoothness of data, SVR performs slightly better than MSVR.

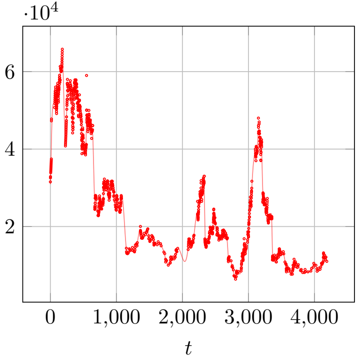
Financial example

- Data comes from http://tsetmc.ir, Tehran Securities Exchange Technology Management Co., daily closing price of a stock named Behran Oil
- Period: 16/04/2001 to 01/04/2018, for a total of 4172 data points, of which only 3369 are valued.
- Prediction time window the last 10 data points.
- For the training data, we extract 1471 basis using the RBF algorithm.
- For MSVR we employ a 30 weekdays window. (In this case, since financial data vary quickly and without any real trend, a larger sliding window would not be useful.)

Graphs



(a) Financial market raw data points



(b) Extracted basis

Figure: Left: financial data. Right: the extracted basis and the RB model.

To do

- Apply the compression algorithm on satellite images
- Use less kernel predictors and integrate with interpolation.
- Use data generation approach to get smoother function and apply SVR on top of that.
- Use this approach to mix data from observation with data from physically-based model simulations.

Main references

1. Image compression [Piazzon et al. 2017].
2. RBF-based reduced order methods (Greedy approach) [Wirtz et al. 2015].
3. Learning with kernels [Fasshauer & McCourt 2015].

F. Piazzon, A. Sommariva, M. Vianello, *Caratheodory-Tchakaloff Subsampling*, Dolom. Res. Notes Approx. **10** (2017), pp. 5–14.

D. Wirtz, N. Karajan, B. Haasdonk, *Surrogate Modelling of multiscale models using kernel methods*, Int. J. Numer. Met. Eng. **101** (2015), pp. 1–28.

G.E. Fasshauer, M.J. McCourt, *Kernel-based Approximation Methods Using Matlab*, World Scientific, Singapore, 2015.

Thank you, Grazie, Dziękuję

Rete ITaliana di Approssimazione

Italian Network on Approximation

https://sites.google.com/site/italianapproximationnetwork/