Journal of Computational Physics 352 (2018) 213–245

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Scientific data interpolation with low dimensional manifold model ☆

Wei Zhu [a,*,1], Bao Wang [a], Richard Barnard [b], Cory D. Hauck [b,c,2], Frank Jenko [d], Stanley Osher [a,3]

[a] Department of Mathematics, University of California Los Angeles, Los Angeles, CA 90095, USA
[b] Computational and Applied Mathematics Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[c] Department of Mathematics, University of Tennessee, Knoxville, TN 37996-1320, USA
[d] Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received 21 June 2017
Received in revised form 18 September 2017
Accepted 25 September 2017
Available online 28 September 2017

Keywords:
Low dimensional manifold model (LDMM)
Scientific data interpolation
Data compression
Regular and irregular sampling
Weighted graph Laplacian

ABSTRACT

We propose to apply a low dimensional manifold model to scientific data interpolation from regular and irregular samplings with a significant amount of missing information. The low dimensionality of the patch manifold for general scientific data sets has been used as a regularizer in a variational formulation. The problem is solved via alternating minimization with respect to the manifold and the data set, and the Laplace–Beltrami operator in the Euler–Lagrange equation is discretized using the weighted graph Laplacian. Various scientific data sets from different fields of study are used to illustrate the performance of the proposed algorithm on data compression and interpolation from both regular and irregular samplings.

1. Introduction

Interpolation and reconstruction of scientific data sets from sparse sampling is of great interest to many researchers from various communities. In many situations, data are only partially sampled due to logistic, economic, or computational constraints: limited number of sensors in seismic data or hyperspectral data acquisition, low-dose radiographs in medical imaging, coarse-grid solutions of partial differential equations due to computational complexity, etc. Moreover, sometimes

☆ The submitted manuscript has been authored by a contractor of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

* Corresponding author.
E-mail addresses: mrzw731@gmail.com (W. Zhu), wangbaonj@gmail.com (B. Wang), barnardrc@ornl.gov (R. Barnard), hauckc@ornl.gov (C.D. Hauck), jenko@physics.ucla.edu (F. Jenko), sjo@math.ucla.edu (S. Osher).

[1] This work is supported by ONR Grant N00014-14-1-0444.
[2] This material is based, in part, upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing and by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. De-AC05-00OR22725.
[3] This material is based, in part, upon work supported by the U.S. Department of Energy, Office of Science and by National Science Foundation, under Grant Numbers DOE-SC0013838 and DMS-1554564.

https://doi.org/10.1016/j.jcp.2017.09.048

one may also intentionally sample partial information of the scientific data set as a straightforward data compression technique. As a result, it has become an important topic to reconstruct the original data set from regular or irregular samplings.

There are typically two ways to approach this problem. The first one is to use the underlying physics to infer the missing data [1–5]. The drawback is that such techniques are usually problem-specific and not generally applicable to similar problems in other fields of study. Signal and data processing techniques, on the other hand, usually do not require too much prior information of the governing physics. These models intend to fill in the missing information by the properties manifested by the sampled data themselves, while implicitly enforcing common structures from physical intuition in the regularization.

Many signal processing approaches to data interpolation have been studied in the context of image inpainting and seismic data interpolation. Popular interpolation models have been proposed through total variation [6,7], wavelets [8,9], and curvelets [10–12,1]. After the introduction of the nonlocal mean by Buades et al. in [13], patch-based techniques exploiting similarity and redundancy of local patches have been extensively studied for inpainting and reconstruction [14–16]. This also leads to a wide variety of sparse-signal models which assume that patches can be sparsely represented by atoms in a prefixed or learned dictionary [8,17]. Patch-based Bayesian models have also been proposed in image and data interpolation [18,19]. However, as reported in [18], some of the algorithms can only be applied to the interpolation of randomly selected samples, and fail to achieve satisfactory results for uniform grid interpolation. Moreover, most of the methods perform poorly when a significant amount of information ($\geq 95\%$) is missing.

Recently, a low dimensional manifold model (LDMM) has been proposed for general image processing problems [20]. In particular, it achieved state-of-the-art results for image interpolation problems with a significant number of missing pixels. The main idea behind LDMM is that the patch manifold (to be explained in Section 2) of a real-world 2D image has a much lower intrinsic dimension than that of the ambient space. Based on this observation, the authors used the dimension of the patch manifold as a regularizer in the variational formulation, and the optimization problem is solved using alternating minimization with respect to the image and the manifold. The key step in the algorithm, which involves solving a Laplace–Beltrami equation over an unstructured point cloud sampling the patch manifold, is solved via either the point integral method [21] or the weighted graph Laplacian [22].

In this work, we apply LDMM to the interpolation of 2D and 3D scientific data sets from either regular or irregular samplings, and demonstrate its superiority when compared to other methods. Moreover, we also compare the performance of LDMM as a sampling-based data compression technique to other standard compression methods. Unlike the other compression methods, sampling-based methods do not require access to the full data set. Although the results of sampling-based algorithms are generally inferior to standard compression methods, they have the advantage of easy implementation in the compression step, and they are also faster in the reconstruction step if only the reconstruction of a small portion of the data set is required. A useful by-product of this comparison is that the standard compression methods are implicitly compared against one another on a set of physically meaningful test cases that can be used for future benchmarks.

The rest of the paper is organized as follows. Section 2 reviews the low dimensional manifold model and justifies its application to scientific data interpolation through a dimension analysis. Section 3 outlines the detailed numerical implementation of LDMM with weighted graph Laplacian which was missing in [22]. A comparison of the numerical results on various scientific data interpolation and compression is reported in Section 4. Finally, we draw our conclusion in Section 5.

2. Low dimensional manifold model

Low dimensional manifold model (LDMM) is a recently proposed mathematical image processing technique which performs particularly well on natural image inpainting [20,23]. The main observation is that the intrinsic dimension of the patch manifold of a natural image is much smaller than that of the ambient Euclidean space. Therefore it is intuitive to use the dimension of the patch manifold as a regularizer to recover the degraded image. We argue that the same property holds true for scientific data sets. Throughout the entire paper, we present our analysis and algorithm for 3D scientific data sets. The formulation for 2D and higher dimensional data sets follows in a natural way.

2.1. Patch manifold and dimension analysis

Consider a 3D datacube $f \in \mathbb{R}^{m\times n\times r}$. For any voxel $\boldsymbol{x} \in \bar{\Omega} = \{1,2,\ldots,m\} \times \{1,2,\ldots,n\} \times \{1,2,\ldots,r\}$,[4] the patch $\mathcal{P}f(\boldsymbol{x})$ is defined as a vector storing the data values in a 3D cube of size $s_1 \times s_2 \times s_3$, with $\boldsymbol{x}$ being the first voxel of the 3D cube in the lexicographic order, i.e. $\boldsymbol{x}$ is in one particular corner of the cube.[5] The patch set $\mathcal{P}(f)$ of f is the collection of all patches:

$$\mathcal{P}(f) = \left\{\mathcal{P}f(\boldsymbol{x}) : \boldsymbol{x} \in \bar{\Omega}\right\} \subset \mathbb{R}^d, \quad d = s_1 \times s_2 \times s_3.$$

[4] The notation Ω is reserved for the sampled subset of $\bar{\Omega}$.

[5] One can also choose $\boldsymbol{x}$ to be the center of the cube, and the result will be similar. The reason is that the reconstruction is performed on patches instead of on voxels. This will be clear in Section 3.