Automatic Construction of a Sparse Best Basis for Potential Approximation and Inversion

Volker Michel and Doreen Fischer

Geomathematics Group, University of Siegen, Germany

Introduction

Numerous basis systems are nowadays available to represent geophysically relevant functions such as potential fields or results of data inversions. Each of the systems has its intrinsic advantages and disadvantages. Whereas classical methods such as expansions in spherical harmonics have their justifications, novel techniques such as wavelets, splines, Slepian functions, and mascons yield additional features which are tailored to improve the solution of particular problems. In practice, the choice of one of these tools is often not straightforward. For this reason, we present a novel algorithm, which is called the Regularized Functional Matching Pursuit. It allows the use of a redundant system of trial functions, where several basis systems can be combined into a so-called dictionary. This algorithm, which is an adaptation of methods developed for the Euclidean setting (see e.g. [Mallat and Zhang 1993, Vincent and Bengio 2002]), iteratively chooses a basis system out of the dictionary to build an approximation of the investigated function. For example, a combination of global basis functions and different kinds of localized basis functions is selected. As a result, large amounts of data can be processed, where a relatively low number of trial functions suffices to achieve a low approximation error. We present here the basics of the algorithm and show some numerical results, where we invert gravitational data for mass density variations. For further details of the algorithm, see [Berkel et al. 2011, Fischer 2011, Fischer and Michel 2012].

The idea

We want to solve an inverse problem $\mathcal{F}F = y$, where $y \in \mathbb{R}^l$ and $\mathcal{F} : \mathrm{L}^2(\mathcal{B}) \to \mathbb{R}^l$ (linear and continuous) are given and $F \in \mathrm{L}^2(\mathcal{B})$ is an unknown function on the Earth's interior $\mathcal{B}$. For this purpose, we intend to construct an approximate solution of the form

$$F = \sum_k \alpha_k d_k$$

in the following way:

- The expansion is developed iteratively, i.e. we move from $F_n := \sum_{k=1}^{n} \alpha_k d_k$ to $F_{n+1} := \sum_{k=1}^{n+1} \alpha_k d_k$.
- Every d_k is a function which is selected by the algorithm from a dictionary $\mathcal{D} \subset \mathrm{L}^2(\mathcal{B})$, which is a redundant system of trial functions.
- The coefficients $\alpha_k \in \mathbb{R}$ are chosen in combination with the dictionary elements d_k.
- The objective is to minimize the regularized data misfit

$$\|y - \mathcal{F}(F_n + \alpha_{n+1} d_{n+1})\|^2_{\mathbb{R}^l} + \lambda \, \|F_n + \alpha_{n+1} d_{n+1}\|^2_{\mathrm{L}^2(\mathcal{B})},$$

where λ is the regularization parameter.

The algorithm RFMP

1) Start with $F_0 := 0$, $R^0 := y$, $n := 0$.
2) Build $F_{n+1} := F_n + \alpha_{n+1} d_{n+1}$ such that

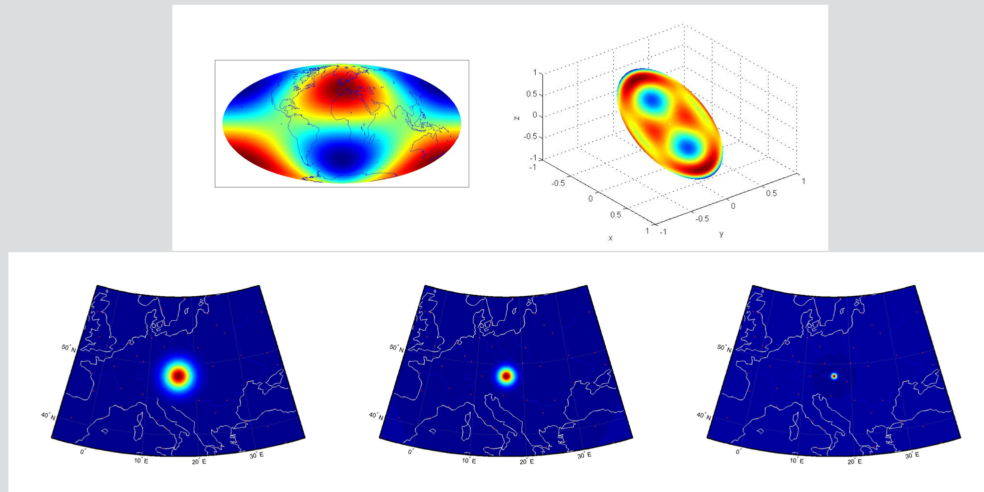
$$d_{n+1} = \operatorname*{argmax}_{d \in \mathcal{D}} \left| \frac{\langle R^n, \mathcal{F}d \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d \rangle_{\mathrm{L}^2(\mathcal{B})}}{\sqrt{\|\mathcal{F}d\|^2_{\mathbb{R}^l} + \lambda \|d\|^2_{\mathrm{L}^2(\mathcal{B})}}} \right| \quad \text{and}$$

$$\alpha_{n+1} := \frac{\langle R^n, \mathcal{F}d_{n+1} \rangle_{\mathbb{R}^l} - \lambda \langle F_n, d_{n+1} \rangle_{\mathrm{L}^2(\mathcal{B})}}{\|\mathcal{F}d_{n+1}\|^2_{\mathbb{R}^l} + \lambda \|d_{n+1}\|^2_{\mathrm{L}^2(\mathcal{B})}}$$

3) Update the residual $R^{n+1} := R^n - \mathcal{F}(\alpha_{n+1} d_{n+1})$.
4) Stop or increase n by 1 and go to Step 2.

The dictionary

Examples of basis functions for the 3d-ball, plotted on the surface or a planar cut

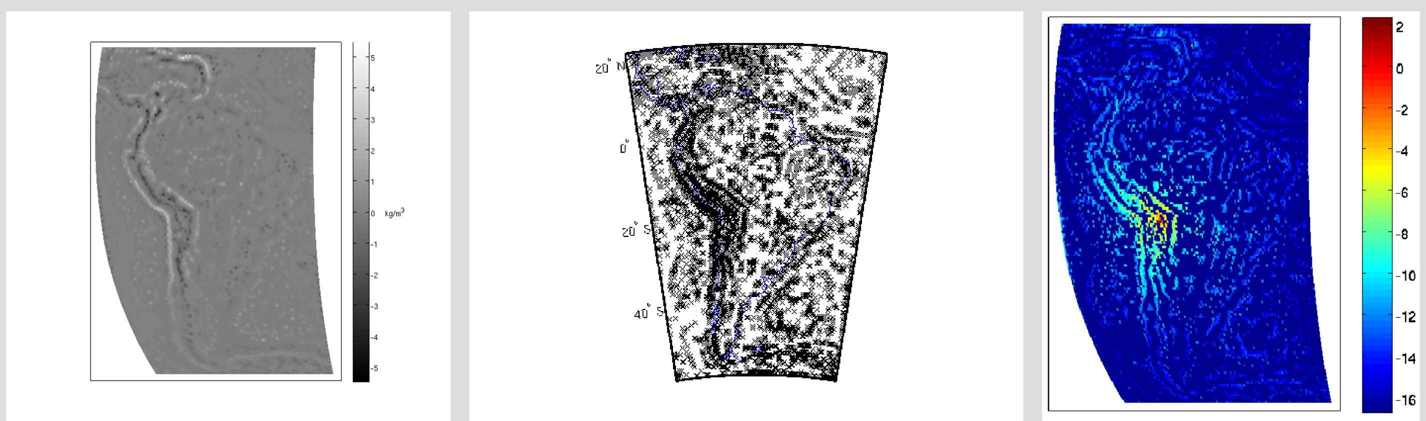


Application: Inversion of gravitational data

- 25,440 (simulated) data of the EGM2008 potential on a point grid 7 km above the Earth
- Dictionary of the form described above
- The regularization parameter $\lambda = 4.6416$ is chosen via the L-curve method.
- The algorithm is truncated after 20,000 summands were chosen for the expansion.

Numerical result

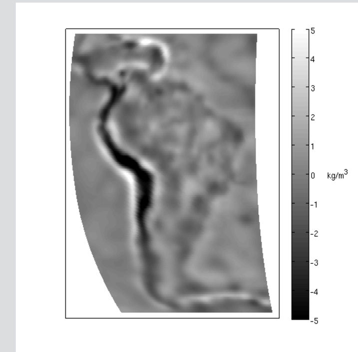
Harmonic density variations (left-hand), chosen centres of the localized basis functions (middle), and local influence of the basis functions (right-hand)



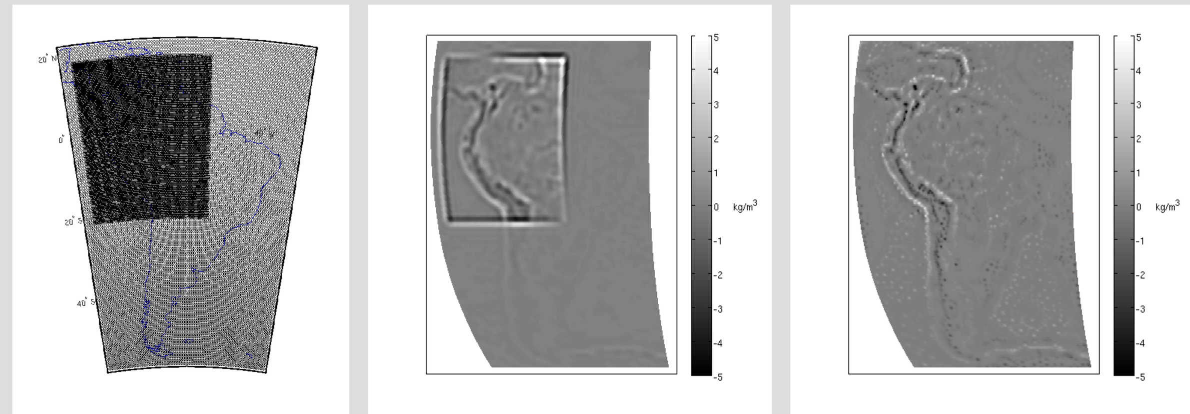
result | centres | local influence

Comparison to other methods

We compare the RFMP to a spline and a wavelet method (see [Michel 2010]). In the case of splines, the numerical limit for the size of the basis is approximately 10^4 which yields a lower resolution (right-hand).



In the case of wavelets, irregular data grids cause problems (left-hand: data grid, middle: wavelet-based approximation, right-hand: approximation by the RFMP)



Conclusions

The novel algorithm has several new features in comparison to previous approaches:

- The algorithm locally adapts the resolution of the result based on the local detail structure of the solution.
- Different kinds of basis functions can be combined, in particular, global functions and wavelet/spline bases.
- There is no numerical limit any more for the size of the basis.
- Heterogeneous data can be inverted.

Acknowledgements

We are grateful for valuable discussions with Frederik J. Simons (Princeton). The project is supported by the German Research Foundation (DFG, projects MI 655/2-2 and MI 655/7-1).

References

- P. Berkel, D. Fischer, V. Michel (2011): Spline multiresolution and numerical results for joint gravitation and normal-mode inversion with an outlook on sparse regularization, *Int. J. Geomath.*, 1:167-204.
- D. Fischer (2011): *Sparse Regularization of a Joint Inversion of Gravitational Data and Normal Mode Anomalies*, PhD thesis, Geomathematics Group, University of Siegen.
- D. Fischer, V. Michel (2012): Sparse regularization of inverse gravimetry — case study: spatial and temporal mass variations in South America, submitted to *Inverse Probl.*
- S.G. Mallat, Z. Zhang (1993): Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.*, 41:3397-3415.
- V. Michel (2010): Tomography — problems and multiscale solutions, in: *Handbook of Geomathematics* (W. Freeden, M.Z. Nashed, T. Sonar, eds.), pp. 949-972.
- P. Vincent, Y. Bengio (2002): Kernel matching pursuit, *Machine Learning*, 48:169-191.