Cleaning-Up Data for Sparse Model Synthesis: When Symbolic-Numeric Computation Meets Error-Correcting Codes*

Erich L. Kaltofen
Department of Mathematics,
North Carolina State University
Raleigh, North Carolina 27695-8205, USA
kaltofen@math.ncsu.edu
www4.ncsu.edu/~kaltofen

Categories and Subject Descriptors

I.1.2 [**Symbolic and Algebraic Manipulation**]: Algorithms; E.4 [**Coding and Information Theory**]: Error control codes; G.1.1 [**Numerical Analysis**]: Interpolation—*Smoothing*

General Terms

Algorithms

Keywords

Sparse Polynomial, Sparse Model, Error-Correcting Codes, Outlier Smoothing

1. EXTENDED ABSTRACT

The discipline of symbolic computation contributes to mathematical model synthesis[†] in several ways. One is the pioneering creation of interpolation algorithms that can account for sparsity in the resulting multi-dimensional models, for example, by Zippel [12], Ben-Or and Tiwari [1], and in their recent numerical counterparts by Giesbrecht-Labahn-Lee [5] and Kaltofen-Yang-Zhi [9].

The theme of our talk is the discovery of sparsity in interpolation algorithms, while at the same time allowing for erroneous input data. As shown in Figure 1, not removing erroneous input points can result in wrong, of course, but also dense outputs. Thus one may use the sparsity constraint to correct for errors, although this is easier said than done: we shall deal with the difficult case where the underlying
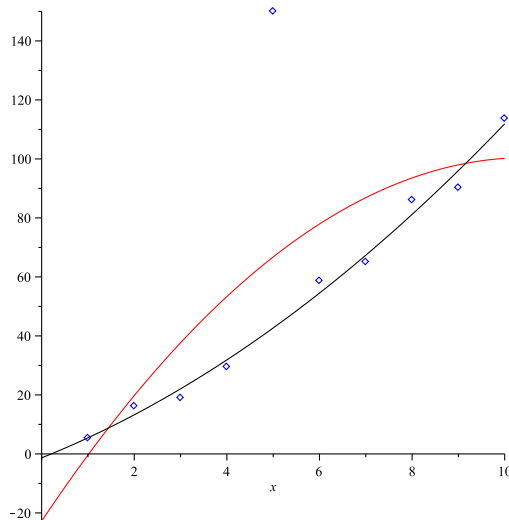

[†]Models are artifacts made by humans, not discovered. When constraining to sparse models, one has a sequence of less and less sparse models that fit the data better and better.

Figure 1: Quadratic fit without and with outlier

sparsity structure, for example the non-zero terms or non-zero entries in the model, are not known on input. Therefore we must compute two lists of discrete quantities: the supports of the output, that is, the sparsity structure, and the location of erroneous input scalars. The corresponding scalar coefficients can be considered, over the real and complex numbers, as continuous quantities. In our symbolic-numeric algorithms we allow imprecision in the input scalars, besides large outlier errors. Sparsity in many situations turns out to be a stabilizing constraint for numerical computation with floating point scalars, but outlier errors are quite destructive when not properly removed.

Sparse interpolation of polynomials or rational functions is the process of computing a sparse multivariate rational function

$$f = \sum_{j=1}^{t_f} a_j \mathbf{x}^{\vec{d}_j}, \ g = \sum_{m=1}^{t_g} b_m \mathbf{x}^{\vec{e}_m},$$

$$a_j, b_m \in \mathsf{K}, a_j \neq 0, b_m \neq 0, \quad (1)$$

from values $\gamma_\ell = (f/g)(\xi_{1,\ell}, \ldots, \xi_{n,\ell})$, where the (unknown) terms of the non-zero monomials are denoted by $\mathbf{x}^{\vec{d}_j} = x_1^{d_j,1} \cdots x_n^{d_j,n}$ and $\mathbf{x}^{\vec{e}_m} = x_1^{e_m,1} \cdots x_n^{e_m,n}$. The problem essentially constitutes sparse model recovery. We consider

both the exact and the numeric setting, the latter of which tolerates noise in the γ_ℓ's for fitting, or as we say, synthesizing a sparse model by approximation. In the numeric setting the field of coefficients is $\mathsf{K} = \mathbb{C}$, the complex numbers. Allowing for a denominator g in the model has the fortuitous side-effect of yielding an algebraic error-correcting decoder: if at $\ell = \lambda_\kappa$ one has an erroneous evaluation $\beta_{\lambda_\kappa} \neq \gamma_{\lambda_\kappa}$ for $1 \leq \kappa \leq k$, one can interpolate the unreduced $(f\Lambda(x_1))/(g\Lambda(x_1))$ where $\Lambda(x_1) = (x_1 - \xi_{1,\lambda_1}) \cdots (x_1 - \xi_{1,\lambda_k})$ is the error locator polynomial. Fitting a model (1) to a list of evaluations can be divided into several specific problems:

1. We consider 4 possible functions: univariate $(n = 1)$ and multivariate $(n \geq 2)$; polynomials $(g = 1)$ and rational functions.

2. We use dense and sparse representations, the latter in several bases: in power basis as in (1) and in Chebyshev or shifted bases with unknown shift.

3. We consider exact fitting, that is, interpolation, approximate fitting, that, is least squares solutions, exact fitting with oversampling and error removal, that is, error correcting decoding, and approximate fitting with oversampling and outlier removal.

For instance, the univariate dense exact polynomial interpolation problem with error removal constitutes a Reed-Solomon decoder [11, 2]. Our SPINO (sparse polynomial interpolation with noise and outliers) algorithm [4] solves the univariate sparse polynomial approximate fitting problem with outlier removal. Our sparse multivariate rational function interpolation algorithms [7, 8] are based on dense univariate algorithms and can also approximate noisy points and remove outliers. From a dense univariate algorithm we can compute a scalar shift σ for the variable $x = y + \sigma$ to obtain a sparse model in y [3].

Our talk addresses the number of evaluations by which our algorithms oversample. For reducing the number of evaluations, in Kaltofen and Pernet [6] we propose to use list-decoding, and in Kaltofen and Yang [8] row subset selection. List-decoding allows for more errors and outliers while computing efficiently a list of interpolants that contains the original sparse function which was evaluated. A further problem is the recovery of univariate sparse rational functions, $(f\Lambda)/(g\Lambda)$ where f and $g \neq 1$ are sparse, which can be possibly unreduced, as in $(x^{20} - 1)/(x - 1)$. We have deployed, so far unsuccessfully, sparse signal recovery via Candès's and Tao's ℓ^1-norm optimization to the problem. The above list allows for more combinations: for example, dense multivariate polynomial interpolation with error removal can be solved by a multivariate sparse interpolation algorithm in the manner of Blahut's [2] Reed-Solomon decoder.

Our algorithms are doubly hybrid: they combine exact with numerical methods, in fact, constitute a numerical version of the algebraic error-correcting decoders, and recover both discrete outputs, the term degrees in the sparse support and the outlier locations, and continuous data, the complex coefficients that fit the data.

This is joint work with Matthew T. Comer (North Carolina State University, now Wofford University, USA), Clément Pernet (Univ. Grenoble, France) and Zhengfeng Yang (East China Normal University, Shanghai, China).

2. REFERENCES

[1] Ben-Or, M., and Tiwari, P. A deterministic algorithm for sparse multivariate polynomial interpolation. In *Proc. Twentieth Annual ACM Symp. Theory Comput.* (New York, N.Y., 1988), ACM Press, pp. 301–309.

[2] Blahut, R. E. A universal Reed-Solomon decoder. *IBM J. Res. Develop. 18*, 2 (Mar. 1984), 943–959.

[3] Boyer, B., Comer, M. T., and Kaltofen, E. L. Sparse polynomial interpolation by variable shift in the presence of noise and outliers in the evaluations. In *Proc. Tenth Asian Symposium on Computer Mathematics (ASCM 2012)* (Oct. 2013). Submitted to SLNCS; URL: http://www.math.ncsu.edu/~kaltofen/bibliography/13/BCK13.pdf.

[4] Comer, M. T., Kaltofen, E. L., and Pernet, C. Sparse polynomial interpolation and Berlekamp/Massey algorithms that correct outlier errors in input values. In *ISSAC 2012 Proc. 37th Internat. Symp. Symbolic Algebraic Comput.* (New York, N. Y., July 2012), J. van der Hoeven and M. van Hoeij, Eds., Association for Computing Machinery, pp. 138–145. URL: http://www.math.ncsu.edu/~kaltofen/bibliography/12/CKP12.pdf.

[5] Giesbrecht, M., Labahn, G., and Lee, W. Symbolic-numeric sparse interpolation of multivariate polynomials. *J. Symbolic Comput. 44* (2009), 943–959.

[6] Kaltofen, E., and Pernet, C. Sparse polynomial interpolation codes and their decoding beyond half the minimal distance. In Nabeshima [10]. URL: http://www.math.ncsu.edu/~kaltofen/bibliography/14/KaPe14.pdf.

[7] Kaltofen, E., and Yang, Z. Sparse multivariate function recovery from values with noise and outlier errors. In *ISSAC 2013 Proc. 38th Internat. Symp. Symbolic Algebraic Comput.* (New York, N. Y., 2013), M. Kauers, Ed., Association for Computing Machinery, pp. 219–226. URL: http://www.math.ncsu.edu/~kaltofen/bibliography/13/KaYa13.pdf.

[8] Kaltofen, E., and Yang, Z. Sparse multivariate function recovery with a high error rate in evaluations. In Nabeshima [10]. URL: http://www.math.ncsu.edu/~kaltofen/bibliography/14/KaYa14.pdf.

[9] Kaltofen, E., Yang, Z., and Zhi, L. On probabilistic analysis of randomization in hybrid symbolic-numeric algorithms. In *SNC'07 Proc. 2007 Internat. Workshop on Symbolic-Numeric Comput.* (New York, N. Y., 2007), J. Verschelde and S. M. Watt, Eds., ACM Press, pp. 11–17. URL: http://www.math.ncsu.edu/~kaltofen/bibliography/07/KYZ07.pdf.

[10] Nabeshima, K., Ed. *ISSAC 2014 Proc. 39th Internat. Symp. Symbolic Algebraic Comput.* (New York, N. Y., 2014), Association for Computing Machinery.

[11] Welch, L. R., and Berlekamp, E. R. Error correction of algebraic block codes. US Patent 4,633,470, 1986. Filed 1983; see http://patft.uspto.gov/.

[12] Zippel, R. E. *Probabilistic algorithms for sparse polynomials.* PhD thesis, Massachusetts Inst. of Technology, Cambridge, USA, Sept. 1979.