# Feature Extraction using Legendre-Sobolev Representation for Handwritten Mathematical Characters

Parisa Alvandi and Stephen M. Watt

**Abstract.** Earlier work has shown how handwritten characters may be represented as plane curves with $x(\lambda)$ and $y(\lambda)$ polynomial functions, and how efficient recognition can be achieved when the polynomials are written in a Legendre-Sobolev basis. In mathematical handwriting recognition there is no fixed baseline and it is necessary to identify certain character features in order to determine spatial relationships. We show how this may be done by operating on the polynomial curves in the Legendre-Sobolev representation.

Modern pen-based devices usually capture digital pen movement a sequence of $(x, y)$ points, sampled at some fixed time frequency. These sample points may be used to provide a continuous functional approximations to the coordinates of the pen movement.

It is useful and convenient to use approximating polynomial functions with time or arc length as the parameter [3]. Most of the work to find these polynomials may be performed in real time, as the pen is moving, by numerically integrating moment functions [4]. Instead of representing the polynomials in a monomial basis, we may choose as basis functions the degree ordered Legendre-Sobolev polynomials that are orthogonal with respect to the inner product

$$\langle f, g\rangle_{LS} = \int_{-1}^{1} f(\lambda)g(\lambda)\mathrm{d}\lambda + \mu \int_{-1}^{1} f'(\lambda)g'(\lambda)\mathrm{d}\lambda, \qquad \lambda \geq 0. \tag{1}$$

This special case of Legendre-Sobolev polynomnials has been studied by Althammer [1]. The basis polynomials $S_i^\mu(\lambda)$ may be obtained by orthogonalization of the monomials $\lambda^i$ with respect to the inner product (1). In what follows, we will omit the explicit $\mu$ when writing $S_i(\lambda)$.

For mathematical handwriting recognition, it has been shown [5, 6] that polynomial approximants of degree $d = 10$ to 12 are sufficient and we may write $x(\lambda) \approx \sum_{i=0}^{d} x_i S_i(\lambda), y(\lambda) \approx \sum_{i=0}^{d} y_i S_i(\lambda)$. In this representation, after shifting the origin and scaling, the curves are points in a $(2d-1)$-dimensional vector space. Character classes are almost completely linearly separable and may be classified using support vector machines [10, 9]. The Euclidean distance in this space corresponds to a variational integral between curves. This allows very fast comparison of characters to known models. As the character classes are convex, linear homotopies between two points in a class will also lie in the class. This may be used to map feature-defining points from known models to new points, e.g. to determine the handwriting baseline [7, 8].

In determining certain features for handwriting, we must find locations where the coordinate curves achieve locally extreme values, e.g. relative maxima or mina, or critical points where the derivatives of the coordinate curves are simultaneously zero.

This motivates us to the problem of finding extrema and approximate polynomial GCD for polynnomials in Legendre-Sobolev representation. In this talk we present mathematical tools for these problems. We demonstrate matrix methods to:

- Compute the coefficients of the Legendre-Sobolev approximation from integrated moments,
- Convert between Legendre-Sobolev and Legendre bases,
- Compute derivatives in Legenre-Sobolev representation,
- Compute the GCD of polynomials in Legendre-Sobolev basis.

## References

[1] P. Althammer. Eine Erweiterung des Orthogonalitätsbegriffes bei Polynomen und deren Anwendung auf die beste approximation. 211:192–204, 1962.

[2] P. Alvandi and S. M. Watt. Real-time computation of Legendre-Sobolev approximations. In *Manuscript*, page 8pp, 2018.

[3] B. W. Char and S. M. Watt. Representing and Characterizing Handwritten Mathematical Symbols through Succinct Functional Approximation. In *9th International Conference on Document Analysis and Recognition (ICDAR 2007), 23-26 September, Curitiba, Paraná, Brazil*, pages 1198–1202, 2007.

[4] O. Golubitsky and S. M. Watt. Online stroke modeling for handwriting recognition. In *Proceedings of the 2008 Conference of the Centre for Advanced Studies on Collaborative Research (CASCON)*, pages 72–80, 2008.

[5] O. Golubitsky and S. M. Watt. Online computation of similarity between handwritten characters. In *Document Recognition and Retrieval XVI, part of the IS&T-SPIE Electronic Imaging Symposium, San Jose, CA, USA, January 20-22, 2009. Proceedings*, volume 7247 of *SPIE Proceedings*, pages C1–C10. SPIE, 2009.

[6] O. Golubitsky and S. M. Watt. Distance-based classification of handwritten mathematical characters. *International Journal on Document Analysis and Recognition*, 13(2):133–146, 2010.

[7] R. Hu and S. M. Watt. Optimization of Point Selection on Digital Ink Curves. In *2012 International Conference on Frontiers in Handwriting Recognition, ICFHR 2012, Bari, Italy, September 18-20, 2012*, pages 527–532, 2012.

[8] R. Hu and S. M. Watt. Identifying Features via Homotopy on Handwritten Mathematical Symbols. In *15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2013, Timisoara, Romania, September 23-26, 2013*, pages 61–67. IEEE Computer Society, 2013.

[9] B. Keshari and S. M. Watt. Hybrid mathematical symbol recognition using support vector machines. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR 2007)*, pages 859–863. IEEE Press, 2007.

[10] B. Keshari and S. M. Watt. Online mathematical symbol recognition using SVMs with features from functional approximation. In *Proceedings of the Mathematical User Interface Workshop (MathUI 2008)*, 2008.

Parisa Alvandi
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo ON, Canada N2L 3G1
e-mail: `palvandi@uwaterloo.ca`

Stephen M. Watt
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo ON, Canada N2L 3G1
e-mail: `smwatt@uwaterloo.ca`