



*Research
Report*

Symmetric Nonlinear Regression

Tamás Antal

Symmetric Nonlinear Regression

Tamás Antal
ETS, Princeton, NJ

April 2007

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

<http://www.ets.org/research/contact.html>

Copyright © 2007 by Educational Testing Service. All rights reserved.

ETS and the ETS logo are registered trademarks of
Educational Testing Service (ETS).



Abstract

An estimation tool for symmetric univariate nonlinear regression is presented. The method is based on introducing a nontrivial set of affine coordinates for diffeomorphisms of the real line. The main ingredient making the computations possible is the Connes-Moscovici Hopf algebra of these affine coordinates.

Key words: Equating, Connes-Moscovici Hopf algebra, diffeomorphisms, symmetry

Acknowledgments

I would like to extend his gratitude to Paul Holland, Alina von Davier, and Raymond Mapuranga for the lively discussions during the preparation of this manuscript. I would also like to thank Henri Moscovici for helpful comments.

1 Introduction

Usual univariate regression analysis makes a distinction between the predictor and the outcome variable. In some situations, however, a completely symmetric handling of the two variables is required. One such example, the main motivation behind this investigation, is the equating of educational tests (see von Davier, Holland, & Thayer, 2004, for an introduction to test equating). When the same sample of students are taking two different tests (Test A and Test B), there is no natural order of the two tests in the resulting data for the test scores. That is, the role of Test A and Test B are interchangeable, and this interchangeability is referred to as the *symmetry* of the data set. Consequently, any model based on this data should reflect this symmetry. Ordinary least squares linear regression will result, in general, in two different regression lines: one when Test A is fitted to Test B and the other is obtained when Test B is fitted to Test A .

For linear regression, there are known symmetric methods: one of them is obtained by measuring the distance of the points of the scatter plot and the regression line along line segments perpendicular to the regression line (Golub & Loan, 1989; Nievergelt, 1994; Sardelis & Valahas, 2004). Some statistical advantages of the symmetric view point are detailed in Sardelis and Valahas. The method was also found superior to the usual least squares approach in the field of image reconstruction see (Hamid, Bobick, & Yezzi, 2004; Kennedy, Buxton, & Gibly, 1999) and references therein.

For nonlinear regression, even the family of possible regression functions is a nontrivial question. The usual next level of generalization, polynomial regression, is not a good candidate for several reasons. First, for a higher order polynomial, the inverse does not always exist. Even when it does, it is impossible to find it, in general. Moreover, if the degree of the polynomial is larger than 1, the inverse is not going to be a polynomial, thereby prohibiting symmetric handling of the data using polynomials exclusively. Considering the larger set of functions containing invertible polynomials and their inverses would pose an unsolvable algebraic challenge, in addition to being awkward.

For these reasons, this paper introduces a solution based on *diffeomorphisms* along with their natural affine coordinatization introduced by Connes and Moscovici (1998). A real diffeomorphism is a differentiable one-to-one and onto $\mathbb{R} \rightarrow \mathbb{R}$ function with a nonzero derivative. It is easy to see that its inverse is also a diffeomorphism. With this large family of functions, symmetry is

readily taken into account. The challenge lies in finding suitable subspaces of diffeomorphisms for regression and a practical way of handling the inverse of a diffeomorphism.

2 Preliminaries

A two-dimensional *scatter plot* is a finite subset $D_{\text{obs}} = \{(x_i, y_i) \mid i = 1, \dots, N\}$ of \mathbb{R}^2 . The word *scatter plot* is used here instead of the usual terminology *data* to emphasize the geometrical nature of our problem. Often one has a *model* expressed through a family of functions $\mathcal{F} \subset \text{Function}(\mathbb{R}, \mathbb{R})$ and the problem is to find a member $f \in \mathcal{F}$ so that $D_m = \{(x_i, f(x_i)) \mid i = 1, \dots, N\}$ and D_{obs} are as closely related as possible. This regression function is denoted by $R_{\mathcal{F}}(D_{\text{obs}}) := f$. That is, $R_{\mathcal{F}}$ is defined as a map from the set of scatter plots to \mathcal{F} . For example, when the model is given by P_n (polynomials up to degree n) and closeness is defined by the distance squared,

$$d^2(D_{\text{obs}}, D_m) := \sum_{i=1}^N (y_i - p(x_i))^2, \quad p \in P_n,$$

being small, one deals with polynomial least squares regression.

This paper is concerned with the case when the family of functions $\text{Diff}(\mathbb{R})^+$ is a subset of increasing diffeomorphisms. ¹

$$\text{Diff}(\mathbb{R})^+ = \{f : \mathbb{R} \rightarrow \mathbb{R} \mid f \text{ is bijection, } f^{(n)} \text{ exists } \forall n, f' > 0\} \subset \text{Diff}(\mathbb{R}).$$

Moreover, the goal is to find the *symmetric* regression $\phi \in \text{Diff}(\mathbb{R})^+$ for the scatter plot D_{obs} . For symmetric regression, use the following:

Definition 1 *The regression map $R_{\text{Diff}(\mathbb{R})^+}^{\text{symm}}$ from scatter plots to diffeomorphisms is a symmetric regression, if whenever $\phi = R_{\text{Diff}(\mathbb{R})^+}^{\text{symm}}(D_{\text{obs}})$ is the regression on D_{obs} , its inverse is the regression on $D_{\text{obs}}^{-1} = \{(y_i, x_i) \mid i = 1, \dots, N\}$, that is,*

$$\phi = R_{\text{Diff}(\mathbb{R})^+}^{\text{symm}}(D_{\text{obs}}) \iff \phi^{-1} = R_{\text{Diff}(\mathbb{R})^+}^{\text{symm}}(D_{\text{obs}}^{-1}).$$

Now assume that $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is an increasing diffeomorphism of the real line. First, factor the diffeomorphism as a composition of a linear function e and a diffeomorphism φ

$$\phi = e \circ \varphi \tag{1}$$

so that $\varphi(0) = 0$ and $\varphi'(0) = 1$. That is,

$$e(x) = \phi(0) + \phi'(0)x, \quad (2)$$

$$\varphi(x) = \frac{\phi(x) - \phi(0)}{\phi'(0)}. \quad (3)$$

Let $G_2 = \{\varphi \in \text{Diff}(\mathbb{R})^+ \mid \varphi(0) = 0, \varphi'(0) = 1\}$ denote the collection of all diffeomorphisms without linear part.²

The reason for this decomposition is that now it is possible to define linear and nonlinear symmetric regression. That is, the decomposition (1) can be thought of as the first step towards defining the degree of a diffeomorphism.

Following this natural decomposition, this paper first discusses the linear symmetric regression and then presents a solution for handling the nonlinear (or G_2) part of the problem.

3 Symmetric Linear Regression

First consider the case when the fitted function is a line, $e(x) = bx + a$. It is easy to see that the usual vertical least squares solution is symmetric if the correlation of the data is 1: $\rho(D) = 1$. In this case the regression is the unique line passing through all data points. If the distance between the regression line and points of the scatter plot is measured perpendicularly to the regression line, then the resulting linear regression is symmetric. More generally, if the distance between a data point and the regression line is measured along a line with slope $s(b)$ given by a *symmetric slope function* $s : \mathbb{R} \rightarrow \mathbb{R}^\times$ ($\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$ is the set of nonzero real numbers) depending on the slope of the regression line b , then to satisfy the symmetry requirement the distance for the inverse should be computed along the line $s(1/b)$. This gives the symmetry condition for s as

$$\frac{1}{s(b)} = s\left(\frac{1}{b}\right). \quad (4)$$

The above mentioned perpendicular solution is obtained by setting

$$s(b) = -\frac{1}{b}. \quad (5)$$

For the value $s(1)$ from (4), $s^2(1) = 1$ is obtained, that is, $s(1) = \pm 1$. Only $s(1) = -1$, however, is a meaningful solution. Note that (4) can be used to obtain symmetric slope functions by setting $s : (0, 1] \rightarrow \mathbb{R}^\times$ arbitrarily with the only restriction $s(1) = -1$ and defining $s(b) = \frac{1}{s(\frac{1}{b})}$ for $b > 1$.

Smooth solutions should be infinitely many times differentiable at $b = 1$, resulting in a series of conditions for the derivatives $s_1^{(n)} := s^{(n)}(1)$ of s at $b = 1$:

$$s_1 = -1 \tag{6}$$

$$s_1'' = -s_1'^2 - s_1' \tag{7}$$

$$s_1^{(4)} = 3s_1'^4 + 12s_1'^3 + 15s_1'^2 - 4s_1^{(3)}s_1' + 6s_1' - 6s_1^{(3)} \tag{8}$$

\vdots

The odd degree derivatives $s_1^{(2n+1)}$ are all free. That is, there are infinitely many analytic solutions to (4) governed by the choices for the odd degree derivatives of s at $b = 1$. Analyticity further requires the convergence of the series

$$\sum_{i=0}^{\infty} \frac{s^{(i)}(1)}{i!} (b-1)^i \tag{9}$$

for any $b \in \mathbb{R}$.

The perpendicular solution (5) is obtained from the choice

$$s_1^{(2i+1)} = -(2i+1)!, \quad \text{for all } i \in \mathbb{N},$$

which makes (9) convergent only for $0 < b < 2$. Because this paper is looking for a solution in the neighborhood of the perpendicular line, this lack of analyticity should always be anticipated. To overcome this limitation in a practical setting, one could use the series expansion (9) to define a solution over $(0, 1]$ and extend it to $[1, \infty]$ by (4).

4 Nonlinear Symmetric Regression

4.1 Affine Coordinates for Diffeomorphisms

Many polynomials are diffeomorphisms, but the inverse of a polynomial is rarely a polynomial itself. That is why polynomial regression, even with diffeomorphic polynomials, is not a good candidate for symmetric regression. After the factorization $\phi = e \circ \varphi$ of a diffeomorphism ϕ , the nonlinear part φ is clearly identified. In practice, this is a two-step process. First, one fits a linear symmetric regression e to the scatter plot. Then, if the fit of e is not satisfactory, symmetric nonlinear regression is performed on the scatter plot for which the linear part e is removed. Nonlinear regression means finding the best fitting $\varphi \in G_2$ diffeomorphism.

If an arbitrary diffeomorphism $\varphi \in G_2$ were allowed during the process, the resulting regression would be a function so that $D_{\text{obs}} = D_m$. This, however, may render the regression too data driven and consequently less sample invariant. Also, there will be infinitely many solutions satisfying this equality, so the decision rule would be almost useless. Hence, the degree of the diffeomorphism for the regression should be limited, similarly to the polynomial regression case.

To overcome these problems, this paper follows Connes and Moscovici (1998) and introduces affine coordinates for the group $G_2 \subset \text{Diff}(\mathbb{R})$ by defining for a diffeomorphism $\varphi \in G_2$:

$$\delta_n(\varphi) := \log(\varphi')^{(n)}(0) \in \mathbb{R}, \quad \varphi \in G_2. \quad (10)$$

It's possible to locally reconstruct φ from these affine coordinates via

$$\varphi(x) = \int_0^x e^{\sum_n \frac{\delta_n(\varphi)}{n!} u^n} du. \quad (11)$$

General theory of affine algebraic groups (Hochschild, 1981) implies that the resulting set of affine coordinates carries the structure of a Hopf algebra.

For the coordinates of the inverse, there is

$$\tilde{\delta}_n(\varphi) := \delta_n(\varphi^{-1}) = \log((\varphi^{-1})')^{(n)}(0). \quad (12)$$

The main advantage of this Hopf algebra-based approach is that there are explicit formulae expressing the coordinates of the inverse in terms of the coordinates of the original function. The first few are listed here (see the appendix for the first 10):

$$\tilde{\delta}_1 = -\delta_1, \quad (13)$$

$$\tilde{\delta}_2 = -\delta_2 + \delta_1^2, \quad (14)$$

$$\tilde{\delta}_3 = -\delta_3 + 4\delta_1\delta_2 - 2\delta_1^3, \quad (15)$$

The interested reader should read the appendix for the details of how such expression can be derived.

4.2 Algorithm for Nonlinear Symmetric Regression

Now, a practical algorithm for computing the nonlinear symmetric regression of a scatter plot D_{obs} is provided. For a general scatter plot, the first step is to find the linear part of the

decomposition (1). That is, a function $e(x) = bx + a$ is sought, so that for the transformed scatter plot,

$$D_{\text{obs}}^{\text{non-lin}} = e^{-1}(D_{\text{obs}}) = \{(x_i, e^{-1}(y_i)) \mid i = 1, \dots, N\}$$

The symmetric linear regression would be the diagonal line

$$R_{\text{lin}}^{\text{symm}}(D_{\text{obs}}^{\text{non-lin}})(x) = x. \quad (16)$$

It is easy to see that this choice will ensure the uniqueness of the linear part, since (16) means that one cannot perform two nontrivial linear symmetric regressions one after another.

The next step is then to find a symmetric nonlinear regression $\varphi \in G_2$ on $D_{\text{obs}}^{\text{non-lin}}$. A closer look at the definition of affine coordinates reveals that the affine coordinates are nothing else but the terms of the Taylor expansion of the function $\log(\varphi')$. To estimate the coordinates, one has to transform first the scatter plot $e^{-1}(D_{\text{obs}})$ to this log derivative scale. That is, observed derivatives are computed from the scatter plot using finite differences:

$$d_i := \frac{y_i - y_{i-1}}{x_i - x_{i-1}}, \quad \forall i = 2, \dots, n. \quad (17)$$

The observed log derivatives are then given by $l_i = \log(d_i) \forall i = 2, \dots, n$. For inside points, that is when $1 < i < n$; there could be another estimate d_i^g obtained via averaging the incoming and outgoing slopes. To keep the symmetry of the model, one should use the geometric mean:

$$d_i^g := \sqrt{d_i d_{i+1}}, \quad \forall i = 2, \dots, n-1. \quad (18)$$

Symmetry means that the slopes for a scatter plot D are reciprocal for the slopes computed for D^{-1} . It is an easy exercise to see that the arithmetic mean does not respect this property.

This averaging appears to be useful in resolving the anomaly introduced by the fact that the observed derivatives d_i should correspond to a mean of x_{i-1} and x_i rather than to either x_i or to x_{i-1} . This would make symmetric handling of the problem a bit awkward.

A sort of stabilization could be achieved by extending these new observations by defining

$$d_1^g := \sqrt{d_1}, \quad \text{and} \quad d_n^g := \sqrt{d_{n-1}}. \quad (19)$$

The observed log derivatives in this case are given by $l_i^g = \log(d_i^g) \forall i = 1, \dots, n$. The stabilization can be thought of as introducing two new points at the end of the scatter plot so that the resulting slopes are 1. This is not required for the method here but maybe useful in some applications. In

what follows, this paper either assumes that stabilization had been done or that the first and last points had been dropped and the data had been reindexed.

Let \tilde{l}_i^g denote the corresponding observed log derivatives that are derived from the inverse scatter plot $(D_{\text{obs}}^{\text{non-lin}})^{-1}$. By construction,

$$\tilde{l}_i^g = -l_i^g. \quad (20)$$

The problem then reduces to fitting two polynomials simultaneously with a certain maximum degree K . To this end, for a polynomial,

$$p(x) = \delta_1 x + \frac{\delta_2}{2} x^2 + \dots + \frac{\delta_K}{K!} x^K \quad (21)$$

with the k -truncated antipode by

$$\tilde{p}(x) = \tilde{\delta}_1 x + \frac{\tilde{\delta}_2}{2} x^2 + \dots + \frac{\tilde{\delta}_K}{K!} x^K, \quad (22)$$

where $\tilde{\delta}_m$ is given as in (13) to (15) and in the appendix. Truncation refers to the fact that even if $\delta_i = 0$ for $i > K$ with some K , the coordinates of the inverse $\tilde{\delta}_i$ for $i > K$ are not necessarily zero. When defining the truncated antipode, those higher order terms are omitted. Also, note that the K -truncated antipode of \tilde{p} is p itself (see the appendix for details):

$$\tilde{\tilde{p}} = p. \quad (23)$$

Then, using ordinary least squares, the polynomial fit, that is, the vector of parameter estimates $\Delta = (\delta_1, \delta_2, \dots, \delta_K)$, is found by minimizing the function

$$\begin{aligned} \ell^2(\delta_1, \delta_2, \dots, \delta_K) &= \sum_{i=1}^n (l_i^g - p(x_i))^2 + (\tilde{l}_i^g - \tilde{p}(y_i))^2 \\ &= \sum_{i=1}^n \left(l_i^g - \sum_{k=1}^K \frac{\delta_k}{k!} x_i^k \right)^2 + \left(-l_i^g - \sum_{k'=1}^K \frac{\tilde{\delta}_{k'}}{k'!} y_i^{k'} \right)^2. \end{aligned} \quad (24)$$

The nonlinear regression φ is then obtained via integration

$$\varphi(x) = \int_0^x e^{p(u)} du. \quad (25)$$

Note, that the antipode was defined so that

$$\varphi^{-1}(y) = \int_0^y e^{\tilde{p}(v)} dv, \quad (26)$$

by neglecting the error originating from the truncation. In (26), the inverse of the diffeomorphism is obtained from the antipode of the corresponding log derivative polynomial. That is, the inverse always exists, and it can be relatively easily found, unlike in the case of polynomial regression as outlined in the introduction. Moreover, by the definition of ℓ^2 and by (20) and (23),

$$\ell^2(\delta_1, \delta_2, \dots, \delta_K) = \ell^2(\tilde{\delta}_1, \tilde{\delta}_2, \dots, \tilde{\delta}_K). \quad (27)$$

Note that (27) directly implies the symmetry property of the regression φ .

For the sake of explicitness, expand (24) in the case of a degree three approximation scheme:

$$\begin{aligned} \ell^2(\delta_1, \delta_2, \delta_3) = & \sum_{i=1}^n \left(l_i^g - \delta_1 x_i - \frac{\delta_2}{2} x_i^2 - \frac{\delta_3}{6} x_i^3 \right)^2 + \\ & \left(-l_i^g + \delta_1 y_i - \frac{-\delta_2 + \delta_1^2}{2} y_i^2 - \frac{-\delta_3 + 4\delta_1\delta_2 - \delta_1^3}{6} y_i^3 \right)^2. \end{aligned} \quad (28)$$

The estimation consists of finding $(\hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3)$ in the neighborhood of $(0, 0, 0)$ so that $\ell^2(\hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3)$ of (28) is minimal.

5 Conclusion

The approach to symmetric regression based on diffeomorphisms of the real line and their Hopf algebra of affine coordinates were introduced in this paper. This method provides a practical way of handling the inverse of the regression function together with the function itself, thereby providing a tool to handle symmetric regression. While some preliminary steps towards solving the problem in general are presented, the paper should be considered as a research plan rather than a report on a finished product.

Several questions are left open. Some of them are relatively small technical matters. A notable example is the proof of the the statements about the linear symmetric slope function around (6) through (8). Some of them are potentially difficult questions. An example would be the effect of truncation of the antipode in (22).

Another large topic would be to relate the technique presented here to the usual equating methods as applied in current practice (von Davier et al., 2004).

Also, the method explicitly uses the estimates of the derivatives as derived from the data. This works as presented only when the data set is ordered in the sense that the piecewise linear

function it defines is strictly monotonic. If this is not the case, then a sort of averaging procedure should be introduced (in a symmetric fashion, of course) to estimate the slopes.

Yet another future direction could be to extend the procedure to higher dimensional diffeomorphisms via the higher order Connes-Moscovici Hopf algebras. This extension could be useful when equating tests utilizing a multidimensional item response theory model.

References

- Connes, A., & Moscovici, H. (1998). Hopf algebras, cyclic cohomology and the transverse index theorem. *Communications in Mathematical Physics*, 198(1), 199-246.
- Golub, G. H., & Loan, C. F. V. (1989). *Matrix computations* (2nd ed.). Baltimore, MD: John Hopkins University Press.
- Hamid, R., Bobick, A., & Yezzi, A. (2004). Audio-visual flow—A variational approach to multimodal flow estimation. In *Proceedings of the 12th international conference on image processing (ICIP), Singapore* (Vol. 4, pp. 2563–2566). New York: IEEE Press.
- Hochschild, G. P. (1981). *Basic theory of algebraic groups and lie algebras*. New York: Springer-Verlag.
- Kennedy, D., Buxton, B., & Gibly, J. (1999). Application of the total least squares procedure to linear view interpolation. In T. Pridmore & D. Elliman (Eds.), *Proceedings of the 10th British machine vision conference* (Vol. 1, pp. 305–314). London: BMVA.
- Menous, F. (2005). Formulas for the Connes-Moscovici Hopf algebra. *Comptes Rendus Mathématique*, 341(2), 75–78.
- Nievergelt, Y. (1994). Total least squares: State-of-the-art regression in numerical analysis. *SIAM Review*, 36(2), 258-264.
- Sardelis, D., & Valahas, T. (2004). *Least squares fitting—Perpendicular offsets*. Retrieved March 22, 2007, from *MathWorld*: <http://library.wolfram.com/infocenter/MathSource/5292/>.
- Sweedler, M. (1969). *Hopf algebras*. New York: W. A. Benjamin, Inc.
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating*. New York: Springer.

Notes

¹ The use of increasing diffeomorphisms is only to keep the connection with test equating alive. As this paper shows, the exact same procedure handles arbitrary diffeomorphisms.

² If the diffeomorphism is decreasing, then the linear part will be decreasing and the nonlinear part will be increasing; that is, it too will be in G_2 .

³ Results in this appendix are taken from Connes and Moscovici (1998).

Appendix

A.1 Connes-Moscovici Hopf Algebra

For the general theory of Hopf algebras, the reader is referred to Sweedler (1969). The one-dimensional Connes-Moscovici Hopf algebra $\mathcal{H}(1)$ was found by Connes and Moscovici (1998) while working on the transverse index theorem of foliations.³ As an algebra, it is the universal enveloping algebra of the Lie algebra generated by $X, Y, (\delta_n)_{n=1}^{\infty}$ subject to the following commutation relations:

$$[X, Y] = -X, \tag{29}$$

$$[X, \delta_n] = \delta_{n+1}, \tag{30}$$

$$[Y, \delta_n] = n\delta_n, \tag{31}$$

$$[\delta_n, \delta_m] = 0. \tag{32}$$

The coproduct, however, is not the usual one for enveloping algebras. It is defined by

$$\Delta Y = Y \otimes 1 + 1 \otimes Y, \tag{33}$$

$$\Delta X = X \otimes 1 + 1 \otimes X + \delta_1 \otimes Y, \tag{34}$$

$$\Delta \delta_1 = \delta_1 \otimes 1 + 1 \otimes \delta_1, \tag{35}$$

$$\Delta \delta_n = \Delta[X, \delta_{n-1}]. \tag{36}$$

The counit ε and the antipode S are defined on generators as follows:

$$\varepsilon(X) = \varepsilon(Y) = \varepsilon(\delta_n) = 0, \tag{37}$$

$$S(Y) = -Y, \quad S(X) = -X + \delta_1 Y, \tag{38}$$

$$S(\delta_1) = -\delta_1, \quad S(\delta_n) = S([X, \delta_{n-1}]). \tag{39}$$

The maps above extend to $\mathcal{H}(1)$ endowing it with a Hopf algebra structure.

The focus is the antipode $\tilde{\delta}_n = S(\delta_n)$ of δ_n . In particular, it is shown how these increasingly complicated expression can be derived in a systematic manner. From the definition, $S(\delta_1) = -\delta_1$.

To compute $S(\delta_2)$, proceed as follows:

$$\begin{aligned}
\tilde{\delta}_2 = S(\delta_2) &= S([X, \delta_1]) \\
&= -[S(X), S(\delta_1)] \\
&= [-X + \delta_1 Y, \delta_1] \\
&= [-X, \delta_1] + [\delta_1 Y, \delta_1] \\
&= -\delta_2 + \delta_1[Y, \delta_1] \\
&= -\delta_2 + \delta_1^2.
\end{aligned} \tag{40}$$

The interested reader may derive $\tilde{\delta}_3$ using the same line of argument. Also, it is worthwhile to note that Menous (2005) contained explicit formulae for the antipode.

Even though it is very useful to consider the whole of the Connes-Moscovici Hopf algebra, it is also worthwhile to note the commutative sub-Hopf algebra \mathcal{H}_δ generated by δ_n for $n > 0$. The commutativity of \mathcal{H}_δ implies that the antipode is involutive: $S^2(a) = S(a)$ for all $a \in \mathcal{H}_\delta$. This shows, in particular, that $S^2(\delta_{i_1} \dots \delta_{i_k}) = \delta_{i_1} \dots \delta_{i_k}$. Hence, for a polynomial p the truncated antipode of the truncated antipode is p itself: $\tilde{\tilde{p}} = p$; see (23).

A.2 Antipode of δ_n up to $n = 10$

$$\tilde{\delta}_1 = -\delta_1,$$

$$\tilde{\delta}_2 = -\delta_2 + \delta_1^2,$$

$$\tilde{\delta}_3 = -\delta_3 + 4\delta_1\delta_2 - 2\delta_1^3,$$

$$\tilde{\delta}_4 = 6\delta_1^4 - 18\delta_2\delta_1^2 + 7\delta_3\delta_1 + 4\delta_2^2 - \delta_4,$$

$$\tilde{\delta}_5 = -24\delta_1^5 + 96\delta_2\delta_1^3 - 46\delta_3\delta_1^2 - 52\delta_2^2\delta_1 + 11\delta_4\delta_1 + 15\delta_2\delta_3 - \delta_5,$$

$$\begin{aligned}
\tilde{\delta}_6 &= 120\delta_1^6 - 600\delta_2\delta_1^4 + 326\delta_3\delta_1^3 + 548\delta_2^2\delta_1^2 - 101\delta_4\delta_1^2 - 271\delta_2\delta_3\delta_1 + 16\delta_5\delta_1 \\
&\quad - 52\delta_2^3 + 15\delta_3^2 + 26\delta_2\delta_4 - \delta_6, \\
\tilde{\delta}_7 &= -720\delta_1^7 + 4320\delta_2\delta_1^5 - 2556\delta_3\delta_1^4 - 5688\delta_2^2\delta_1^3 + 932\delta_4\delta_1^3 + 3700\delta_2\delta_3\delta_1^2 - \\
&\quad 197\delta_5\delta_1^2 + 1408\delta_2^3\delta_1 - 361\delta_3^2\delta_1 - 629\delta_2\delta_4\delta_1 + 22\delta_6\delta_1 - 427\delta_2^2\delta_3 + \\
&\quad 56\delta_3\delta_4 + 42\delta_2\delta_5 - \delta_7, \\
\tilde{\delta}_8 &= 5040\delta_1^8 - 35280\delta_2\delta_1^6 + 22212\delta_3\delta_1^5 + 61416\delta_2^2\delta_1^4 - 9080\delta_4\delta_1^4 - \\
&\quad 47500\delta_2\delta_3\delta_1^3 + 2311\delta_5\delta_1^3 - 26920\delta_2^3\delta_1^2 + 6227\delta_3^2\delta_1^2 + 10899\delta_2\delta_4\delta_1^2 - \\
&\quad 351\delta_6\delta_1^2 + 14613\delta_2^2\delta_3\delta_1 - 1743\delta_3\delta_4\delta_1 - 1317\delta_2\delta_5\delta_1 + 29\delta_7\delta_1 + 1408\delta_2^4 - \\
&\quad 1215\delta_2\delta_3^2 + 56\delta_4^2 - 1056\delta_2^2\delta_4 + 98\delta_3\delta_5 + 64\delta_2\delta_6 - \delta_8, \\
\tilde{\delta}_9 &= -40320\delta_1^9 + 322560\delta_2\delta_1^7 - 212976\delta_3\delta_1^6 - 703008\delta_2^2\delta_1^5 + 94852\delta_4\delta_1^5 + \\
&\quad 613892\delta_2\delta_3\delta_1^4 - 27568\delta_5\delta_1^4 + 461024\delta_2^3\delta_1^3 - 97316\delta_3^2\delta_1^3 - 171012\delta_2\delta_4\delta_1^3 + \\
&\quad 5119\delta_6\delta_1^3 - 340164\delta_2^2\delta_3\delta_1^2 + 37297\delta_3\delta_4\delta_1^2 + 28368\delta_2\delta_5\delta_1^2 - 583\delta_7\delta_1^2 - \\
&\quad 65104\delta_2^4\delta_1 + 51400\delta_2\delta_3^2\delta_1 - 2191\delta_4^2\delta_1 + 44859\delta_2^2\delta_4\delta_1 - 3844\delta_3\delta_5\delta_1 - \\
&\quad 2531\delta_2\delta_6\delta_1 + 37\delta_8\delta_1 - 1215\delta_3^3 + 20245\delta_2^3\delta_3 - 6285\delta_2\delta_3\delta_4 - \\
&\quad 2373\delta_2^2\delta_5 + 210\delta_4\delta_5 + 162\delta_3\delta_6 + 93\delta_2\delta_7 - \delta_9, \\
\tilde{\delta}_{10} &= 362880\delta_1^{10} - 3265920\delta_2\delta_1^8 + 2239344\delta_3\delta_1^7 + 8584992\delta_2^2\delta_1^6 - 1066644\delta_4\delta_1^6 - \\
&\quad 8208900\delta_2\delta_3\delta_1^5 + 342964\delta_5\delta_1^5 - 7664256\delta_2^3\delta_1^4 + 1489736\delta_3^2\delta_1^4 + \\
&\quad 2627260\delta_2\delta_4\delta_1^4 - 73639\delta_6\delta_1^4 + 6900116\delta_2^2\delta_3\delta_1^3 - 701317\delta_3\delta_4\delta_1^3 - \\
&\quad 536596\delta_2\delta_5\delta_1^3 + 10366\delta_7\delta_1^3 + 1969008\delta_2^4\delta_1^2 - 1434876\delta_2\delta_3^2\delta_1^2 + \\
&\quad 57016\delta_4^2\delta_1^2 - 1256931\delta_2^2\delta_4\delta_1^2 + 100261\delta_3\delta_5\delta_1^2 + 66504\delta_2\delta_6\delta_1^2 - 916\delta_8\delta_1^2 + \\
&\quad 62335\delta_3^3\delta_1 - 1122949\delta_2^3\delta_3\delta_1 + 323677\delta_2\delta_3\delta_4\delta_1 + 122952\delta_2^2\delta_5\delta_1 - \\
&\quad 10116\delta_4\delta_5\delta_1 - 7833\delta_3\delta_6\delta_1 - 4534\delta_2\delta_7\delta_1 + 46\delta_9\delta_1 - 65104\delta_2^5 + \\
&\quad 112135\delta_2^2\delta_3^2 - 8476\delta_2\delta_4^2 + 210\delta_5^2 + 65104\delta_2^3\delta_4 - 9930\delta_3^2\delta_4 - 14875\delta_2\delta_3\delta_5 - \\
&\quad 4904\delta_2^2\delta_6 + 372\delta_4\delta_6 + 255\delta_3\delta_7 + 130\delta_2\delta_8 - \delta_{10}.
\end{aligned} \tag{41}$$

A.3 Affine Coordinates and the Antipode

There is, of course, a deeper reason for the introduction of the affine coordinates $\delta_n(\varphi)$. To keep the paper more accessible, however, only a small justification is provided for them by relating the antipode formulae (39) to the affine coordinates of the inverse of the diffeomorphism. First write the first few coordinates explicitly:

$$\delta_1(\varphi) = \varphi''(0), \quad (42)$$

$$\delta_2(\varphi) = -\varphi''(0)^2 + \varphi^{(3)}(0), \quad (43)$$

$$\delta_3(\varphi) = 2\varphi''(0)^3 - 3\varphi^{(3)}(0)\varphi''(0) + \varphi^{(4)}(0). \quad (44)$$

Moreover, by definition, $\delta_n(\varphi^{-1}) = S(\delta_n)(\varphi)$. To see how this compares to $S(\delta_n)$ introduced before, first observe that

$$\begin{aligned} \delta_n(\varphi^{-1}) &= (\log(\varphi^{-1}))^{(n)}(0) \\ &= \left(\log \frac{1}{\varphi' \circ \varphi^{-1}} \right)^{(n)}(0). \end{aligned} \quad (45)$$

Using this, one obtains

$$\delta_1(\varphi^{-1}) = -\varphi''(0), \quad (46)$$

$$\delta_2(\varphi^{-1}) = 2\varphi''(0)^2 - \varphi^{(3)}(0), \quad (47)$$

$$\delta_3(\varphi^{-1}) = -8\varphi''(0)^3 + 7\varphi^{(3)}(0)\varphi''(0) - \varphi^{(4)}(0). \quad (48)$$

It is an easy exercise to see that

$$\delta_1(\varphi^{-1}) = -\delta_1(\varphi), \quad (49)$$

$$\delta_2(\varphi^{-1}) = -\delta_2(\varphi) + \delta_1(\varphi)^2, \quad (50)$$

$$\delta_3(\varphi^{-1}) = -\delta_3(\varphi) + 4\delta_1(\varphi)\delta_2(\varphi) - 2\delta_1(\varphi)^3, \quad (51)$$

which is in accordance with (13) to (15).