Processing, Modeling and Parameter Estimation of the Dynamic On-Line Handwriting Signal

Homayoon S.M. Beigi
IBM T.J. Watson Research Center
P.O. Box 218 - Yorktown Heights, New York 10598 U.S.A.
Phone: +1 (914) 945-1894 Fax: +1 (914) 945-4010
EMail: beigi@watson.ibm.com

Abstract

This paper presents the signal processing, modeling and parameter estimation of the on-line handwriting signal based on a linearization and approximation of the differential equation model of the human hand. A parameter estimation scheme has been devised for fitting a partial sinusoidal function to a set of points in the 2-dimensional plane. The estimated parameters model the output of the human motor control when constrained by moving on a plane with certain amount of variable friction. There have been attempts by other researchers to use the same model for extracting features from the cursive handwriting signal. Their estimation however assumes that the pieces of the signal are always segmented such that two consecutive zero crossings of the velocity are present in both x and y directions In the case which is solved by this paper, the pieces may be ant portion of the above segment. The method of this paper alleviates this problem by forming a penalty function which is minimized using a quadratically convergent optimization scheme. The resulting parameters have shown very good performance in modeling the data points when reconstructed and used for recognition purposes.

1 Introduction

This paper presents the signal processing, modeling and parameter estimation of the on-line handwriting signal based on a linearization and approximation of the differential equation model of the human hand. The on-line handwriting signal is generated by the movement of the tip of a stylus (pen) on a digitizer tablet. The $< x, y >$ coordinates of the movement of the stylus are sampled by the digitizer tablet at a constant frequency, typically around $100Hz$. This two dimensional signal may be thought of as the trajectory of a mechanical system constrained by a planar surface. The goal here is to model the output of this mechanical system based on observing its output in the form of a sampled on-line handwriting signal. The dynamic model of the human handwriting generation apparatus as a whole is very complicated and includes many parts. Figure 1 shows a block diagram of this control system.

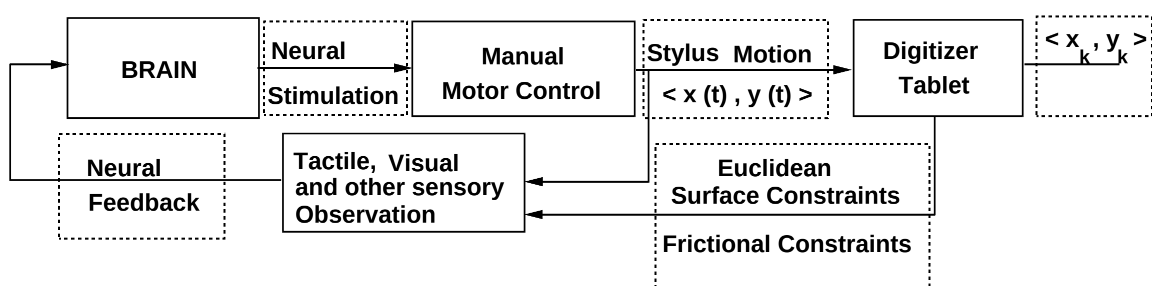

Figure 1: Process of Generation of the Handwriting Signal

The brain generates a language code based on an attempt to convey an idea. This code is perceived as an image of the written word based on experience. The hand muscles are then neurally stimulated to generate a curve on the surface based on the perceptual desired trajectory (shape of the word). In this control process, visual feedback produces corrections to the control input of the manual motor control system. In addition to visual feedback, some tactile feedback is used to remain within the constraints (planar surface of the digitizer or paper). Other parameters such as the normal force needed for inking, the tangential force needed for combating surface friction, velocity and similar measures are used as well for adapting the control input to the instantaneous situation. Please note that changing any one of these observation parameters drastically from the norm results in a very poor writing performance. This is partly due to the learning nature of handwriting. Most of the problems in using a transparent digitizer tablet (with glass as the surface), arise from the lowered friction as compared to a paper interface. Most writers experience a great difficulty of adjusting to the lower friction. This is one of the undesirable parameters haunting the performance of today's handwriting recognition systems. This shows that the handwriting generation apparatus is a finely tuned system (based on learning) and only minor deviations from the norm are tolerated. [1] presents a handwriting recognition system for on-line unconstrained handwriting. As is the case of many other systems, this recognition system generally throughs away most of the dynamic information in the handwriting signal such as the speed information. The purpose of this research is to generate features of the handwriting based on the dynamic information available in the signal. These features possess complementary information which may lead to enhanced recognition accuracies when used in conjunction with the existing recognizer. [2]

The following sections take the $< x_k, y_k >$ sequence of digitized coordinates and try to model the signal as the output of a second order dynamic system. The main effort, due to the complexity of the system and lack of enough intimate knowledge about the central controller (Brain), tries to model the generic output signal of the two-dimensional second-order dynamic system by making some assumptions in order to linearize the dynamic system within a small nominal trajectory. This may be thought of as an equivalent open-loop system whose output should be modeled.

2 Handwriting Model

The general equation of motion of most rigid-body systems, with the consideration of dynamics such as inertia, gravity, Coriolis, Centrifugal, and other forces may be written as follows:

$$T = M(\eta)\ddot{\eta} + C(\eta, \dot{\eta}) + F(\dot{\eta}) + G(\eta) + T_d \tag{1}$$

where, T is the vector of generalized forces supplied by the actuators (muscles), η is the vector of generalized coordinates, $M(\eta)$ is the equivalent mass matrix, $C(\eta, \dot{\eta})$ is the vector of generalized forces due to Coriolis and Centrifugal accelerations, $F(\dot{\eta})$ is the vector of generalized forces due to generalized viscous friction, $G(\eta)$ is the vector of generalized forces due to gravity and other potential energy such as stiffness, and T_d is the vector of disturbances, friction, and other unmodeled forces. [3] In equation 1, all vectors are of dimension $nx1$ and $M(\eta)$ has dimension nxn.

Please note that equation 1 may be linearized about a set of reference coordinates at each time instant; the non-linear equations of motion can then be approximated by a set of second order linear differential equations. If we consider the governing equations of motion for the human-hand motor control, used for handwriting generation, we may make further simplifications by ignoring all forces but the D'Alembert forces and spring stiffness. These approximations only hold true under the conditions that the speed of writing is within limits of the average writing speeds and that the hand moves in a two dimensional trajectory related to producing legible handwriting. Hollerbach, in his 1980 Doctoral Thesis [4], has made the approximation of modeling the instantaneous apparatus of the hand with two couples of springs (two along the y-direction and two along the x-direction). Figure 2 shows the approximate configuration assumed by [4] plus the extra drift parameters in the x and y directions.

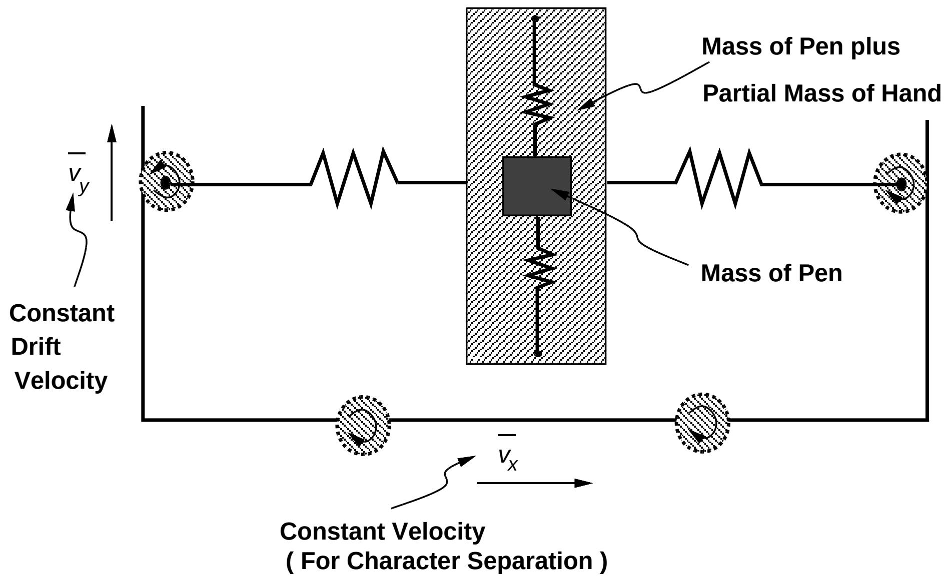

Figure 2: Approximate Model of the Hand Based on the Orthogonal Spring Model

In this model, the damping is not included. This is because of the assumption that the handwriting signal is the transient output of the hand and that there is so little damping with a time constant such that the effects of decay are not visible for at least half of period of oscillation. With this assumption, the model presumes that all the parameters are piecewise constant within the half period duration. Therefore, by modeling each half period piece of the handwriting signal separately, we are assuming that the desired trajectory changes every half period. This results in using a different approximation model for the open-loop equivalent system of each section of writing. The boundaries of these half period segments are taken to be the zero-velocity (minimum dynamics) points along the signal for the x and y directions independently (figure 3). These assumptions have been shown to hold very closely for on-line handwriting is general. [4]. Although, under certain assumptions, such as minimum jerk, these boundaries and the models may change. [4, 5]

Considering the discussed approximations, let us then assume that the differential equations for the handwriting generation process may be approximated by a two dimensional second order differential equation with linear time-invariant coefficients along a piece of writing between any two consecutive extreme positions in each coordinate (x and y), given by figure 3. Under these assumptions, the solution of the approximate differential equation, in the nominal region, would be in the usual sine and cosine form. Therefore, the velocity in each coordinate will also have the same form. For the sake of modeling the handwriting it is better to consider the velocity rather than the position for the apparent reasons of robustness to noise and pre-emphasis. Thus, the velocities in x and y directions under these very crude assumptions would be of the form given in equation 2.

$$\begin{aligned} \dot{x} &= A_x sin(\omega_x(t - t_0) + \phi_x) + \bar{v}_x \\ \dot{y} &= A_y sin(\omega_y(t - t_0) + \phi_y) + \bar{v}_y \end{aligned} \quad (2)$$

where A_x, ω_x, ϕ_x and $\bar{v}_x$ are the amplitude, frequency, phase and mean velocity for the x direction. Also, A_y, ω_y, ϕ_y and $\bar{v}_y$ are the counterparts in the y direction. Figure 4 shows the plot of $\dot{y} - \bar{v}_y$ for the word *languages* written in script (see top of figure 5 for the written word). In figure 4, the segmentation boundaries have been marked. As it is apparent from the figure, the form of a cosine with constant parameters is quite an acceptable approximation for each piece in the plot. This approximation, although not as good, is also acceptable for the x direction to a certain extent. One

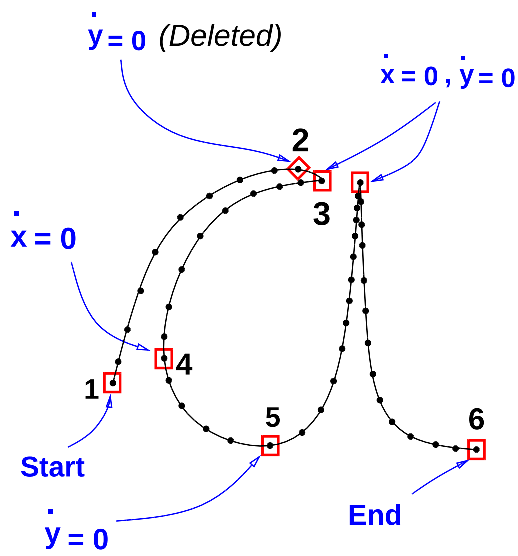

Figure 3: Segmentation of the character *a*

may at the first glance imagine that the phase does not play an important role, however, if the word is to be segmented in such a way as to use **either** $\dot{x} = 0$ **or** $\dot{y} = 0$, then the phase plays a very important role of synchronization. In fact in some cases a segment boundary is simply deleted just because it generates too small a window; in these cases, also, the phase is essential (figure 3).

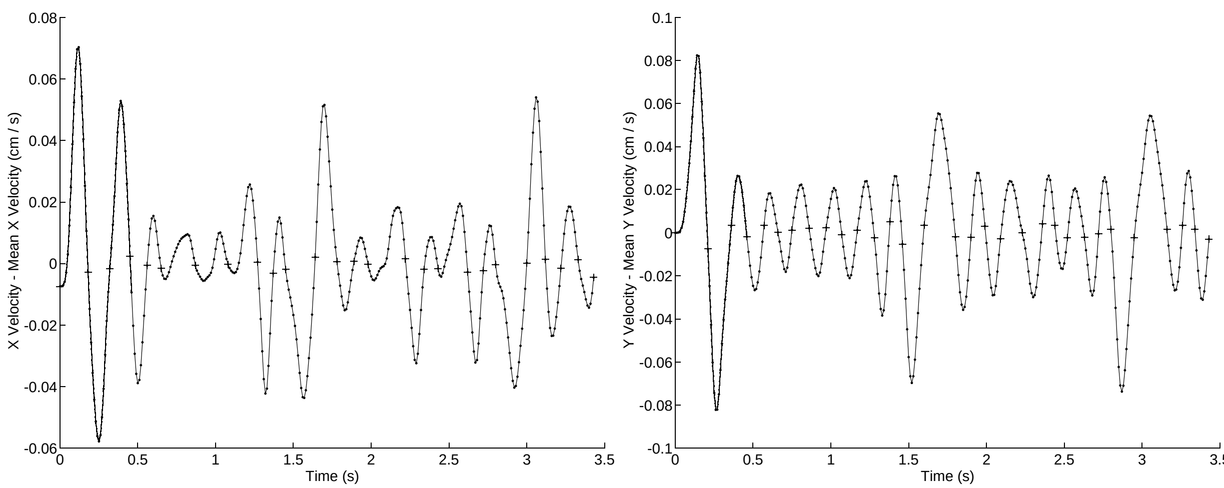

Figure 4: x (left) and y (right) Velocity Plots for the Word *languages*

3 Pre-Processing

In a practical sense, however, if segments of the writing are written very slowly, a couple of problems may arise. The first problem comes up in the computation of the segmentation points where $\dot{x}$ and $\dot{y}$ are nearly zero. The finite difference approximation of the velocities in x and y assumes a Δt which is small in comparison to the nominal Δx or Δy. However, in slow speeds, since the Δt is a fixed number equal to the inverse of the sampling frequency, the velocity is not

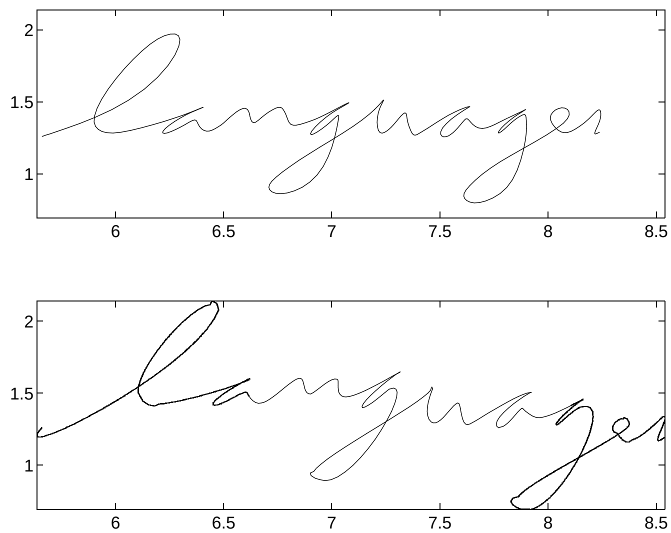

Figure 5: The Original (top) and Reconstruction (bottom) of the word *languages* using the Dynamic Parameters

correctly estimated. To alleviate this problem, an approximate time normalization is done to generate writing with nearly similar nominal speeds.

Figures 6 through 9 show the plots of x and y versus time for four different samples of the character a, written by the same writer. Shape of the character is the same for all the four samples. In these figures, the darker lines show the original points and the lighter lines show plots of the points after time normalization. The a in figure 8 is an example written in a nominal speed (empirically determined). The sample of figure 6 was written in a close to nominal speed with certain variations and as it may be seen from the graph, after normalization, the plot looks a lot more like those in figure 8. The sample of figure 7 was written with a uniform speed, but in general much slower than the nominal speed of writing. After time normalization, again, the plots are very close to those shown in figure 8. The character of figure 7 is shown in figure 10. It may be noted that in the addition of the problem in velocity computation, slowly written characters also contain a fair amount of noise. The noise is reduced by using a 10th order low-pass Yulewalk filter, with zero phase shift, on the signal and the results are shown in figure 10 in a lighter line. This filtering is done before proceeding with the time normalization. Finally, figure 9 shows the most complicated case which is the case when the velocity of writing changes within a single stroke. This often happens when person is preoccupied by a thought while writing. An example is when the person puts the pen on the pad and meanwhile thinks of the statement to write. The application of this time normalization is not only important in regards to segmentation, but it also allows the approximation of all the mean velocities within the segments by the same mean velocity over the whole stroke. This is particularly useful for reducing the number of parameters to be saved for a follow up reconstruction using these parameters.

4 Parameter Estimation

Consider the parameter estimation problem when the spring model of figure 2 is used. Equations 2 give the solution to the x and y velocities of the signal. The $\bar{v}_x$ term of these equations is the mean x velocity which results in the separation of the characters in the writing. If this velocity were equal to zero, the hand would stag stagnant and all the characters would be formed overlapping

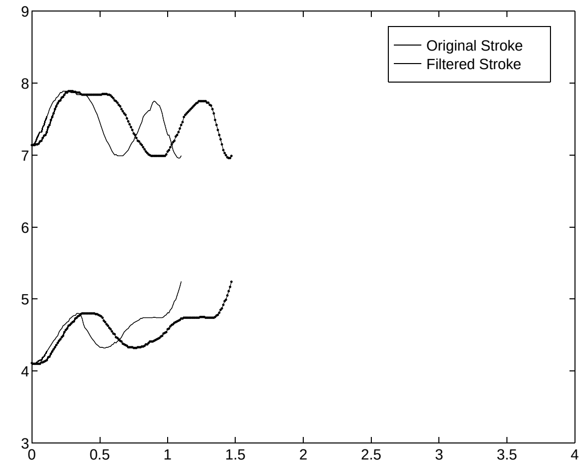

Figure 6: *a*: normal speed with small variations

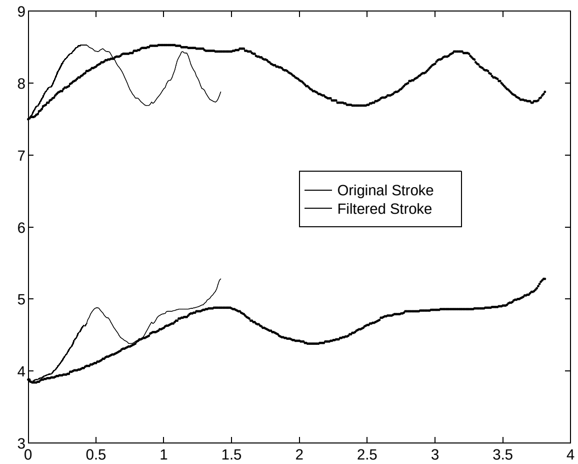

Figure 7: *a*: written very slowly

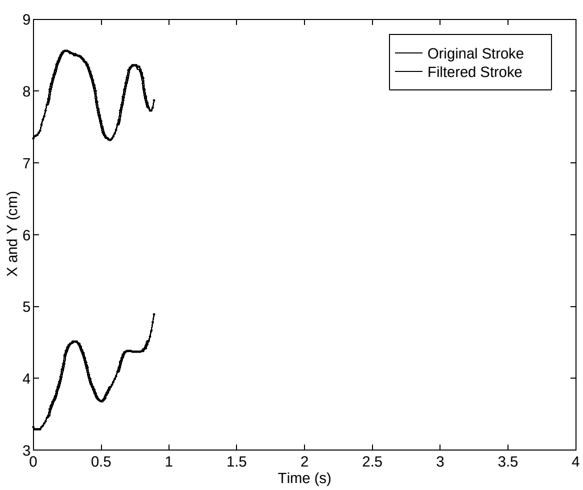

Figure 8: *a*: nominal speed

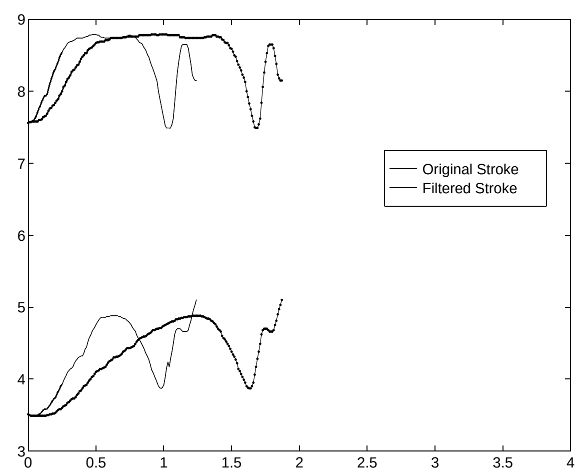

Figure 9: *a*: very slowly (first ligature); nominal speed (other parts)

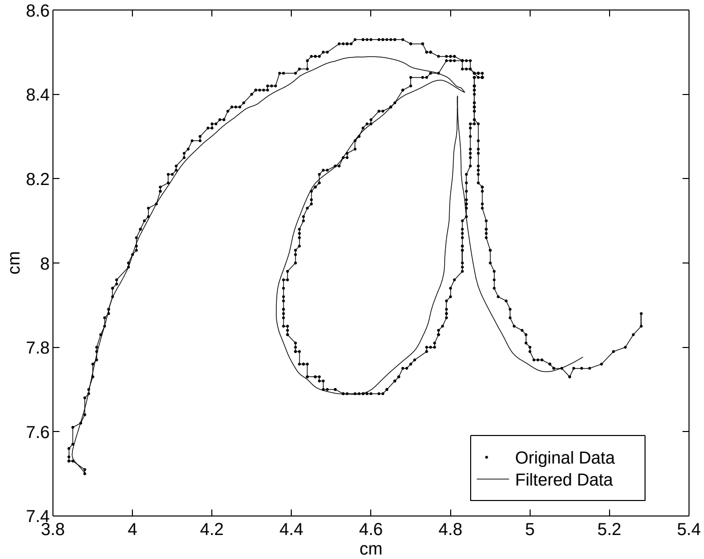

Figure 10: character a written with a very slow speed

one another. For estimating $\bar{v}_x$, we assume that the mean velocity is constant within each stroke. Therefore, the value of $\bar{v}_x$ is estimated to be the mean value of v_x computed within a stroke. Similarly, we notice that if the writer is to write on a horizontal line (rule), the mean y velocity should be zero. In fact, if this value is positive, then the writing would start drifting upward and it would drift downward in case this velocity is negative. $\bar{v}_y$ may also be estimated to be the mean value of the y velocity. Using these estimates of $\bar{v}_x$ and $\bar{v}_y$ and subtracting these values from $\dot{x}$ and $\dot{y}$ in equations 2, the new x and y velocities will be given by:

$$\dot{\hat{x}} = A_x sin(\omega_x(t - t_0) + \phi_x)$$
$$\dot{\hat{y}} = A_y sin(\omega_y(t - t_0) + \phi_y) \quad (3)$$

Now the estimation problem reduces to the estimation of the parameters of a sine curves fit given a few data points. The following parameter estimation scheme may be used for fitting a sinusoidal curve to any set of points. An optimization problem is formulated for estimating these parameters. A very good set of initial conditions are picked for the amplitude, the frequency and the phase. Due to the reliability of these initial conditions, a penalty is imposed on deviating from the initial amplitude and frequency values. From this point on, due to the similarity of the equations for $\dot{\hat{x}}$ and $\dot{\hat{y}}$, the procedure for estimating the generic amplitude A_η, the generic frequency ω_η and the generic phase ϕ_η are presented. Here, η is considered to be a generalized coordinate which may be replaced by x or y. Consider an n point segment of a sampled stroke using the segmentation scheme of last section. If the sample interval is denoted by Δt, the following n-dimensional vectors may be defined:

$$\vec{t} : t_k = (k-1)\Delta t \qquad 1 \le k \le n \quad (4)$$

$$\vec{s} : s_k = sin(\omega_\eta t_k + \phi_\eta) \qquad 1 \le k \le n \quad (5)$$

$$\vec{\xi} : \xi_k = t_k sin(\omega_\eta t_k + \phi_\eta) \qquad 1 \le k \le n \quad (6)$$

$$\vec{\Xi} : \Xi_k = {t_k}^2 sin(\omega_\eta t_k + \phi_\eta) \qquad 1 \le k \le n \quad (7)$$

$$\vec{c} : c_k = cos(\omega_\eta t_k + \phi_\eta) \qquad 1 \leq k \leq n \tag{8}$$

$$\vec{e} : e_k = \dot{\vec{\eta}} - A_\eta \vec{s} \tag{9}$$

The objective is to find a segment of a sine wave which would model the data points in this segment with minimal sum of squares of errors. The parameters to be estimated for each coordinate are the amplitude, A_η, the frequency, ω_η, and the phase ,ϕ_η. Let us denote the local velocity vector (with the mean velocity subtracted) of the points in the segment as $\dot{\vec{\eta}}_d$. Then, the simplest objective function for the optimization problem of estimating the system parameters is the sum of squares of errors along the segment, given by,

$$E_\eta = (\dot{\vec{\eta}}_d - A_\eta \vec{s})^T (\dot{\vec{\eta}}_d - A_\eta \vec{s}) \tag{10}$$

The problem with minimizing equation 10 is that by increasing the frequency, it is possible to reduce the value of E_η, however, a high frequency sinusoidal curve will not fit the general continuous signal well. It is imperative to introduce a penalty on the size of the frequency. This is especially true since a very good initial value for the frequency may be obtained by estimating the zero crossings of the data. A good estimate of the amplitude, A_η may also be obtained by using the value of the velocity point with the maximum absolute value. Since the data is segmented at the zero velocity points, an initial value of 0 is very appropriate for ϕ_η. One may be satisfied with these initial values and not want to compute the actual values. However, here is where the major contribution of this paper lies. In picking the segments, any combination of $\dot{x} = 0$ and $\dot{y} = 0$ may be used. Assume that a segment is between the points $\dot{x} = 0$ and $\dot{y} = 0$ and we are estimating the parameters for the y velocity. In this case, the y velocity does not start from 0. Therefore, the above initial estimates may be way off.

Note that as stated earlier, the frequency tends to converge to large numbers to reduce the error E_η in equation 10, and that we have a good initial estimate of ω_η based on zero-crossings. We also have a good initial estimate of the amplitude A_η. Based on these arguments, we should not allow the frequency and amplitude to deviate from their initial conditions by a lot. This may be done by modifying equation 10 as follows,

$$E_\eta = (\dot{\vec{\eta}}_d - A_\eta \vec{s})^T (\dot{\vec{\eta}}_d - A_\eta \vec{s}) + \alpha(\omega_\eta - \tilde{\omega}_\eta)^2 + \beta(A_\eta - \tilde{A}_\eta)^2 \tag{11}$$

where $\tilde{\omega}_\eta$ and $\tilde{A}_\eta$ are the initial estimates of ω_η and A_η respectively and α and β are weighting factors.

Considering the objective function of equation 11, a Newton's method may be used to iteratively solve for the set of parameters which minimize E_η. To apply Newton's method to solving this optimization problem, let us define the state vector for the parameters to be estimated, as,

$$\zeta_\eta = \begin{bmatrix} A_\eta \\ \phi_\eta \\ \omega_\eta \end{bmatrix} \tag{12}$$

Then the gradient of E_η, g, is written as,

$$g = \nabla_\zeta E_\eta = \begin{bmatrix} -2\vec{s}^T(\dot{\vec{\eta}} - A_\eta \vec{s}) + 2\beta(A_\eta - \tilde{A}_\eta) \\ -2A_\eta \vec{c}^T(\dot{\vec{\eta}} - A_\eta \vec{s}) \\ -2A_\eta \vec{\xi}^T E_\eta + 2\alpha(\omega_\eta - \tilde{\omega}) \end{bmatrix} \tag{13}$$

Similarly, the symmetric $3x3$ Hessian matrix, G, may be written as follows,

$$G = \nabla_{\zeta}{}^2 E_\eta = \begin{bmatrix} 2(\vec{s}^T\vec{s} + 2n\beta & 2\vec{c}^T(A_\eta\vec{s} - \vec{e}_\eta) & 2\vec{c}^T(A_\eta\vec{\xi} - \vec{e}_\eta) \\ G_{12} & 2(A_\eta\vec{c}^T(A_\eta\vec{c} + \vec{e}_\eta) & A_\eta\left[A_\eta n(n+1)\Delta t + 2\vec{\xi}^T(\dot{\vec{\eta}} - 2A_\eta\vec{s})\right] \\ G_{13} & G_{23} & 2A_\eta\left[A_\eta\vec{t}^T\vec{t} - 2A_\eta\vec{\xi}^T\vec{\xi} + \vec{\Xi}^T\dot{\vec{\eta}}\right] \end{bmatrix} \tag{14}$$

Newton's method assumes that the surface of the objective function is quadratic and provides and iterative direction toward the minimum value of this function. [6] However, the direction that is provided by Newton's method, requires the inverse of the Hessian matrix. In this problem, due to the small size of the Hessian matrix, the inverse may be solved for analytically in terms of elements of the Hessian matrix, G. Equation 15 gives the inverse Hessian matrix based on the elements of the Hessian given in equation 14.

$$H = \nabla_{\zeta}{}^2 E_\eta{}^{-1} = \frac{1}{|G|}\begin{bmatrix} G_{22}G_{33} - G_{23}{}^2 & G_{23}G_{13} - G_{12}G_{33} & G_{12}G_{23} - G_{22}G_{13} \\ H_{12} & G_{11}G_{33} - G_{13}{}^2 & G_{12}G_{13} - G_{11}G_{23} \\ H_{13} & H_{23} & G_{11}G_{22} - G_{12}{}^2 \end{bmatrix} \tag{15}$$

where, $|G|$ is the determinant of the Hessian matrix given by,

$$|G| = G_{11}G_{12}G_{13} + 2G_{12}G_{23}G_{13} - (G_{13}{}^2G_{22} + G_{12}{}^2G_{33} + G_{23}{}^2G_{11}) \tag{16}$$

Given the above equations, it is simple to solve for the parameter estimate vector, $\vec{\zeta_\eta}$ using the Newton step,

$$\vec{\zeta}_\eta^{(i+1)} = \vec{\zeta}_\eta^{(i)} - \gamma^{(i)}H^{(i)}g^{(i)} \qquad i = 0, 1, 2, \cdots \tag{17}$$

In equation 17, γ is a weight which may be computed using any line search technique such as Golden Section, Fletcher, etc. [6]

5 Reconstruction and Recognition

The parameters of equation 2 after estimation may be used for several purposes including reconstruction (hence compression) and recognition of the writing. Figures 5 and 11 show sample reconstructions using the estimates of parameters for equation 2. Figure 11 also shows the location of segmentation points. These segmentation points may be further reduced by using the cusp point for the two segmentation points in the vicinity of a cusp and removing the extraneous point. This is also important in removing noisy parameters from those used for the purpose of recognition. [2] presents some results of recognition based on the parameters extracted by the techniques of this paper. The recognition results show that these parameters indeed capture some extra information compared to those obtained by considering only the shape of the written word. [1]

6 Conclusion

In conclusion, a parameter estimation scheme has been devised for fitting a partial sinusoidal function to a set of points in the 2-dimensional plane. The estimated parameters model the output of the human motor control when constrained by moving on a plane with certain amount of variable friction. Fitting a sinusoidal function to a set of points comes up very often considering the nature of the mechanical world. These parameters may be used to compress the data for the set of points they model. This compression is highly desirable given today's high rate of data to be transmitted. There have been attempts by other researchers to use the same model for extracting features from the cursive handwriting signal. Their estimation however assumes that the pieces of the signal are

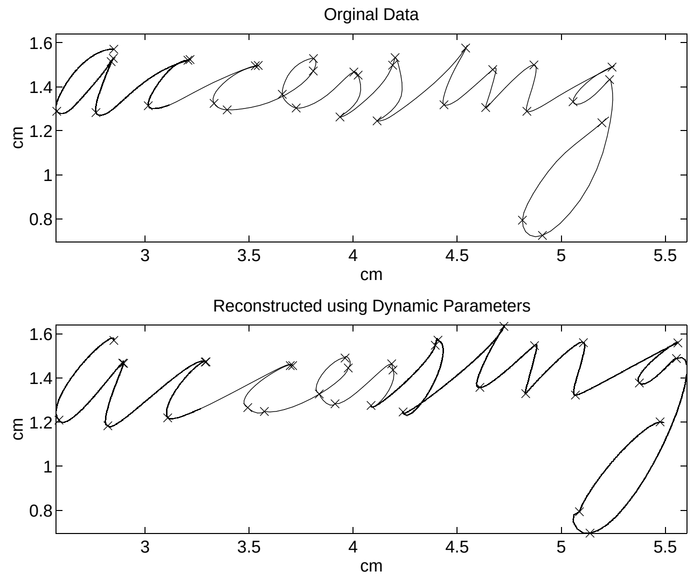

Figure 11: The Reconstruction of the word *accessing* using the Dynamic Parameters

always segmented such that two consecutive zero crossings of the velocity are present in both x and y directions In the case which is solved by this paper, the pieces may be ant portion of the above segment. This presents hurdles in estimating the amplitude, frequency and phase of the approximating sinusoid. The method of this paper alleviates this problem by forming a penalty function which is minimized using a quadratically convergent optimization scheme. The resulting parameters have shown very good performance in modeling the data points when reconstructed and used for recognition purposes. An active effort is on for enhancing the information extraction from these parameters to aid the handwriting recognition process. [2]

References

[1] Krishna S. Nathan, Homayoon S.M. Beigi, Gregory J. Clary, Jayashree Subrahmonia and Hiroshi Maruyama, "Real Time On-Line Unconstrained Handwriting Recognition using Statistical Methods," ICASSP-95, Detroit, Michigan, May 8-12, 1995.

[2] Homayoon S.M. Beigi, "Pre-Processing the Dynamics of On-Line Handwriting, Feature Extraction and Recognition," Submitted to the International Workshop on Frontiers of Handwriting Recognition, Colchester, England, Sep. 2-5, 1996.

[3] J. J. Craig, Adaptive Control of Mechanical Manipulators, Addison-Wesley Publishing Company New York, 1987.

[4] John M. Hollerbach, "An Oscillation Theory of Handwriting," Doctoral Dissertation, Massachusetts Institute of Technology, March 1980.

[5] Adel M. Alimi and Réjean Plamondon, "Performance Analysis of Handwritten Strokes Generation Models," Pre-Proc. of IWFHRIII, Buffalo, New York, May 25-27, 1993, pp.272-283.

[6] Homayoon S.M. Beigi, "Neural Network Learning and Learning Control Through Optimization Techniques," Doctoral Thesis, School of Engineering and Applied Science, Columbia University, New York City, New York, 1991.