Implementation of Visual Reconstruction Networks - Alternatives to Resistive Networks

D. Mansor
Telecom Research Laboratory
Clayton 3168
and
D. Suter
Dept. of Computer Science and Computer Engineering
La Trobe University
Bundoora 3083
Australia

Abstract

Visual Reconstruction networks are often based upon resistive grids. The resistive grid approach has been adopted by the Harris Coupled Depth-Slope analog network and generalised for regularization involving arbitrary degrees of smoothness. This paper considers implementations of arbitrary order regularisation networks which do not require resistive grids. The approach followed in this paper is to generalise the original formulation of Harris and then to follow alternative paths to analog circuit realisation allowed by our generalisation.

The Harris Coupled Depth-Slope model of visual reconstruction was proposed by Harris [1] as a means of independently approximating a function and various orders of derivatives of that function. The Coupled Depth-Slope model has been promoted [1] as a means for efficiently solving regularisation formulations of visual reconstruction problems [2]. This proposal was directly linked to an analog network realization that used the resistive network analogy popularised by others [3] for smoothing a derivative. Later papers, in collaboration with others, have investigated further the analog realisation of various low order smoothing networks (e.g., [4] discusses a one dimensional "thin-plate" implementation, [5] concentrates on modifying the resistive elements to allow for discontinuities, [6] presents a two dimensional first order or membrane solution, and [7] present simulations of their more general networks).

Recently [8] [9] Harris' approach has itself been generalised. The penalty based enforcement of the constraints between the derivatives is replaced with an Augmented Lagrangian formulation that can reduce to Harris' penalty based approach, or to a Lagrangian based approach, as special cases (see section 1).

More extensive details of the new setting are considered elsewhere [8] [9] [10]. The main contribution of this paper is to consider some of the actual implementation alternatives allowed by the new setting.
1 Formulation

For simplicity, the discussion in this section and in the rest of the paper is limited to one dimensional problems. Generalisations to two dimensions are generally straightforward [8] [9].

1.1 Regularization Formulations

In visual reconstruction a starting point is usually a formulation using Tikhonov regularisation [2] where one seeks a solution $\Psi$ that minimises a functional of the form:

$$
\sum_{i \in C} \frac{\beta}{2} (\Psi_i - d_i)^2 + \sum_{j \in J} \frac{1}{2} (\frac{d\Psi_j}{dx})^2 dx.
$$

In such a formulation the value of the recovered function at a point i (belonging to the set of points C in $\Omega$ at which constraints are known) is constrained to be close to the data at that point $d_i$ but be reasonably smooth. The first term in 1 encourages the solution to be close to the data, whilst the second term (penalising the roughness of $\Psi$) encourages the solution to be smooth. The degree of smoothness enforced is determined by the order of the derivatives appearing in the second term (first derivatives force the solution to be like a membrane stretched over the data points, second derivatives impose more stiffness in making the solution more like a thin-plate) and by the relative size of $\beta$ controlling the tradeoff between smoothness and fidelity to the data.

Suppose we have a formulation that involves the first and second derivatives only (even though the greater generality of the formulation 1 is elegant, for practical purposes one rarely sees a formulation involving higher than the second derivative), then Harris' penalty based formulation would replace each derivative (up to but not including the highest appearing) with an unknown and then relate these functions to the corresponding derivative with a penalty term. Such a formulation seeks the minimum $(\Psi, u)$ of the functional:

$$
\sum_{i \in C} \frac{\beta}{2} (\Psi_i - d_i)^2 + \int_{\Omega} \frac{1}{2} (\frac{d\Psi}{dx})^2 dx + \int_{\Omega} \frac{1}{2} (\frac{du}{dx})^2 dx + \int_{\Omega} \frac{1}{2} (\frac{d\Psi}{dx} - u)^2 dx
$$

The last penalty term ensures that $u$ is close to the derivative of $\Psi$ but note that exact equality between these will generally only be enforced in the limit $\rho \to \infty$. An analog network that solves such a formulation using resistive networks to smooth the derivatives, and resistors and a special tri-subtraction device [1] to enforce compatibility between the function $u$ and the derivative of $\Psi$ has been derived by Harris. Elaborations on the basic structure allow various orders of smoothing to be turned off - particularly locally to allow for isolated discontinuities in the data (e.g. [6]).

1.2 Enforcing Constraints Between Derivatives

An alternative to the penalty methods for ensuring agreement between the introduced unknowns and the derivatives is to use Lagrange multiplier techniques. The Lagrangian version of 2 is to seek an extremum (now the solution $(\Psi, u, \lambda)$ is not necessarily a minimum) of the functional:

$$
\sum_{i \in C} \frac{\beta}{2} (\Psi_i - d_i)^2 + \int_{\Omega} \frac{1}{2} (\frac{d\Psi}{dx})^2 dx + \int_{\Omega} \frac{1}{2} (\frac{du}{dx})^2 dx + \int_{\Omega} \lambda (\frac{d\Psi}{dx} - u) dx
$$

1899
An even more useful (numerically) approach that also enforces the constraints exactly is to combine the Lagrangian and penalty based methods (but now $\rho$ no longer needs to be large):

$$
\sum_{i \in G} \beta (\Psi_i - d_i)^2 + \int \left( \frac{1}{2} \frac{d\Psi}{dz} \right)^2 \, dx + \int \left( \frac{1}{2} \frac{du}{dz} \right)^2 \, dx + \int \lambda \left( \frac{d\Psi}{dz} - u \right) \, dx + \frac{\rho}{2} \int \left( \frac{d\Psi}{dz} - u \right)^2 \, dx
$$

From what is known of Augmented Lagrangian methods in general [11] it is to be expected that this type of formulation will offer the stability of penalty based methods with the accuracy of Lagrangian methods for enforcing constraints. Indeed it is well known that penalty methods only accurately enforce the constraints in the limit where the penalty parameters become infinitely large. Large penalty parameters can be a problem for digital implementations as well as for analog implementations (in the Harris analog implementation approach this translates to the need for relatively large resistances coupling each "derivative layer" compared with the size of the resistances coupling between layers [7]).

It seems that we pay a price of using Lagrangian based formulations in that we must now find the Lagrange multipliers as well. However, it is often the case [8][9] that the Lagrange multipliers are useful quantities themselves or can be eliminated from the expression.

In all of our simulations to date, the Lagrangian methods have proved stable and thus, in view of their simplicity relative to the Augmented Lagrangian methods, and also to contrast with the penalty based approach of Harris, we investigate only these formulations here.

2 Implementation

In this section of the paper we introduce the basic method of deriving analog implementations that do not rely on the resistive grid approach.

2.1 The Analog Principle

Our starting point is to use a gradient type approach. For the penalty based formulations, we can use simple gradient descent upon the variables. For the Lagrangian or Augmented Lagrangian approach the solution is usually a saddle point rather than a minimum of the functional. In numerical optimization it is well known that one can often perform gradient descent on the primary variables (generally our derivative approximations) since the functional is minimised with respect to these variables, and gradient ascent on the secondary variables (generally the Lagrange multipliers) [12]. This approach as a basis for analog neural networks was proposed by Platt [13].

To illustrate the approach we will consider a simple example. Consider the Lagrangian formulation for regularisation using first order smoothness only. We seek the extremum of the functional:

$$
L = \sum_{i \in G} \beta (\Psi_i - d_i)^2 + \int \frac{1}{2} u^2 \, dx + \int \lambda \left( \frac{d\Psi}{dz} - u \right) \, dx
$$

It is easy to show that at the extremum $\lambda = u$, and that, if we substitute for $\lambda$ in our formulation, we obtain a new functional that has the same extremum.
but where \( u \) is now a secondary variable \( [8] \):

\[
L = \sum_{i=0}^{n-1} \frac{\beta}{2} (\Psi_i - d_i)^2 - \int_0^1 \frac{1}{2} u^2 \, dx + \int_0^1 u \frac{d\Psi}{dx} \, dx
\]  

(6)

This formulation leads to much more simple discrete form (arbitrarily setting the nodal spacing to one unit):

\[
L = \sum_{i=0}^{n-1} \frac{\beta}{2} (\Psi_i - d_i)^2 + \sum_{i=0}^{n-2} (-u_i^2 + u_i(\Psi_{i+1} - \Psi_i))
\]  

(7)

and much more simple analog description:

\[
\frac{d\Psi_i}{dt} = \frac{\partial L}{\partial \Psi_i} = \left\{ \begin{array}{ll}
-\beta(\Psi_i - d_i) + u_i - u_{i-1} & i = 1 \ldots n - 2 \\
-\beta(\Psi_0 - d_0) + u_0 & i = 0 \\
-\beta(\Psi_{n-1} - d_{n-1}) - u_{n-2} & i = n - 1
\end{array} \right.
\]

(8)

\[
\frac{du_i}{dt} = \frac{\partial L}{\partial u_i} = -u_i + \Psi_{i+1} - \Psi_i \quad i = 0 \ldots n - 2
\]

(9)

2.2 The Analog Circuit

Equations 8 and 9 describe the time evolution of functions \( \Psi_i \) and \( u_i \) from some arbitrary starting point to the final solution. This requires analog subtraction, addition, multiplication and integration functions. These are implemented through the use of simple transconductance amplifiers based on MOS transistors \([14]\) and capacitors.

For small signal operation, the transfer function of a transconductance amplifier, may be approximated by the following linear equation:

\[
I_{out} = G_m(V_1 - V_2)
\]

(10)

Here, \( G_m \) is the transconductance of the amplifier and \( I_b \) is a bias current that can be controlled by a bias voltage \( V_b \).

Equation 10 shows how the transconductance amplifier may be used to perform subtraction and addition of analog signals. The addition of analog current signals may be achieved through connecting the outputs of more than one transconductance amplifiers together. In addition, the transconductance may be controlled by varying the bias voltage \( V_b \). This effectively provides a means of adjusting the constant that multiplies the differential voltage input. The integration function is achieved through current integration using capacitors.

In implementing the analog circuit for the 1st order regularisation network, the basic building blocks are the \( \Psi_i \) and \( u_i \) nodes in figure 1. Equations 8 and 9 govern the operation of these neurons respectively. The special expressions for cases \( i=0 \) and \( i=n-1 \) in equation 8 may lead to structurally different \( \Psi_i \) nodes at either end of the network (i.e. \( \Psi_0 \) and \( \Psi_{n-1} \)). This irregularity may be overcome we define \( u_{-1} = 0 \) and \( u_{n+1} = 0 \).

The circuit described in section 2.2 was simulated using Spice3b1. The transient response of the circuit with a step input was analysed. Step signals were applied to every \( d_i \) input simultaneously, and the magnitude of each step signal corresponds to magnitude of the input signal for that node.

The circuit has then tested with several inputs including a ten point noisy step input function. The results are shown in figure 2. To a certain extent, the circuit has succeeded in smoothing out the noise. As expected, a negative peak appears in the first derivative the the transition point of the step function.
Figure 1: The analog circuit

Figure 2: Reconstructed noisy step function
2.2.1 Effect of varying $\beta$

The bias voltage ($V_p$) of transconductance amplifier $X_0$ in Fig. 1 corresponds to the constant $\beta$ in equation 8. In theory, $\beta$ determines how close the outputs $\Psi_i$ are to their corresponding inputs $d_i$. The smaller $\beta$ is, the more the output may deviate from the original input, hence better smoothing for noisy data. Larger $\beta$ tends to have the opposite effect.

The bias voltage of the transconductance amplifiers used here controls the bias current $I_b$ according to equation 10. In general, increasing $V_p$ has the effect of increasing $\beta$ of equation 8. The effects of varying $\beta$ on the output of the noisy step input function is shown in Fig. 3. Several simulations with different $V_p$ values were carried out. As predicted, the output function $\Psi$ becomes smoother as $V_p$ decreases but with possible large deviations from the input function $d$.

2.2.2 Performance of the circuit

The performance of the circuit here is defined in terms of the time taken for the step response to reach an acceptable final value (less than 1% change in the outputs). Using the noisy step input for the test, the times for the output to settle down to an acceptable level for varying $V_p$ values are summarised in the following table:

<table>
<thead>
<tr>
<th>$V_p$</th>
<th>Time to reach final value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15V</td>
<td>2.4 $\mu$sec</td>
</tr>
<tr>
<td>0.25V</td>
<td>1.7 $\mu$sec</td>
</tr>
<tr>
<td>0.50V</td>
<td>1.1 $\mu$sec</td>
</tr>
<tr>
<td>0.75V</td>
<td>0.6 $\mu$sec</td>
</tr>
<tr>
<td>1.00V</td>
<td>0.3 $\mu$sec</td>
</tr>
</tbody>
</table>

The results clearly indicate another penalty that is incurred when more smoothing is required; the time taken for the circuit to arrive to a final result increases. Even so, convergence still occurs within a few microseconds. Thus this circuit far exceeds the capability of a digital computer when applied to this problem, and a viable solution for real time applications.
2.2.5 Extension to allow discontinuities

The first order regularisation circuit described previously may be extended to allow discontinuities. Here, a simple extension is described to show that the concept is realisable. This circuit will allow discontinuity to occur between any two data points.

To allow for discontinuities, only the \( u_i \) neural circuit requires modification. This is shown in figure 4. The basic extension of the circuit (without the digital interface) only requires an additional two transistors for every \( u_i \) node, one NMOS and the other PMOS. Together they provide a switch which selects either the \( u_i \) circuit output or signal zero \( (V_{zero}) \). The selection is done digitally whereby a digital "1" (5 volts) at the input \( select \) selects \( V_{zero} \) as the output of this modified circuit, therefore allowing discontinuity to occur between data points \( i \) and \( i+1 \). A signal "0" (0 volts) at that input turns off the facility for discontinuity at that point by selecting the original \( u_i \) output.

To test the circuit, the noisy step input used previously is applied to this circuit. Here, the bias voltage \( V_{b} \) was set to 0.25 volts and a discontinuity is allowed between data points \( d_4 \) and \( d_5 \) by applying 5 volts to input \( select \). The results are shown in figure 3. Comparing this to figure 2, it is obvious that here, a sudden transition exists between outputs \( \Psi_4 \) and \( \Psi_5 \) and that the discontinuity has no effect on the neighbouring points.

3 Conclusion

A novel analog technique that is an alternative to resistive grid approaches has been applied to surface reconstruction using a first order regularisation formulation. In addition, a simple extension to permit discontinuity to occur in the output resulting function has also been demonstrated. Through circuit simulation using Spice3, correct operation of the proposed circuit has been demonstrated.

The approach followed here can be generalised to higher order regularisation formulations and there are a number of advantages of the circuit described here compared with other circuits to achieve similar aims [6]. Harris's circuit required complex tri-directional subtracting devices. In addition, this circuit does not require resistors which are expensive on silicon. Our more general starting point allows the derivation many other alternative networks - the evaluation of the
implementation cost and performance of all of many of these alternatives is the subject of current work.

References


