AN EFFICIENT, APPROXIMATE PATH-FOLLOWING ALGORITHM FOR ELASTIC NET BASED NONLINEAR SPIKE ENHANCEMENT

Max A. Little[a,b]

[b]Media Lab, Massachussetts Institute of Technology, Cambridge, MA, USA
[a]Nonlinearity and Complexity Research Group, Aston University, Birmingham, UK

ABSTRACT

Unwanted 'spike noise' in a digital signal is a common problem in digital filtering. However, sometimes the spikes are wanted and other, superimposed, signals are unwanted, and linear, time invariant (LTI) filtering is ineffective because the spikes are wideband – overlapping with independent noise in the frequency domain. So, no LTI filter can separate them, necessitating nonlinear filtering. However, there are applications in which the 'noise' includes drift or smooth signals for which LTI filters are ideal. We describe a nonlinear filter formulated as the solution to an elastic net regularization problem, which attenuates band-limited signals and independent noise, while enhancing superimposed spikes. Making use of known analytic solutions a novel, approximate path-following algorithm is given that provides a good, filtered output with reduced computational effort by comparison to standard convex optimization methods. Accurate performance is shown on real, noisy electrophysiological recordings of neural spikes.

Index Terms— Filter, regularization, nonlinear, spike, noise

1. INTRODUCTION

Consider the digital signal processing problem of removing independent and identically distributed (i.i.d., white) Gaussian noise from a known, frequency band-limited signal. Linear, time-invariant (LTI) digital filters are often used in this application [1]. An appropriate filter can pass through only the band-limited, wanted signal, removing most of the unwanted noise. However, when the signal is not band-limited and/or the noise is non-Gaussian, LTI filters are ineffective, for example, for independent (non-Gaussian) 'spike noise': that is, noise where large deviations occur more frequently than the Gaussian noise model suggests [2]. It can be shown that LTI filtering, such as the moving average filter, is highly inefficient by comparison to a simple nonlinear alternative: the moving median filter [3, 4].

In other cases, the spikes are the wanted signal and there may be i.i.d. Gaussian noise. Additionally, there may be other, band-limited components that need to be removed, for example: slow drift in equipment calibration, or power interference in electrical signals. This is an example of a *spike detection*, or *spike enhancement*, problem.

Spike enhancement is critical to electrophysiological recordings of neural cell spiking [5]. Here, LTI filters cannot simultaneously remove the i.i.d. noise and the band-limited components without corrupting the wanted spikes [2], so this problem is often addressed with specialized, heuristic algorithms. Unfortunately, many of these heuristics lack a principled mathematical framework in which to analyze their performance, and are often not robust to changing experimental conditions [7].

The purpose of this paper is to describe the use of elastic net regularization [8] in spike enhancement, and to introduce a novel algorithm for solving this convex regularization problem. The elastic net, which combines L_2-with L_1-norm regularization, is closely related to signal processing methods based on L_1-norm regularization alone, such as Lasso regression. Because of the very wide scope of such methods, it is not possible to do full justice to the concepts here, we will simply pick out some interesting examples in signal processing such as compressed sensing [9], sparse linear prediction of speech [10], and step filtering in biophysical experiments [3].

Path-following algorithms for convex regularization problems have been developed more recently in statistical machine learning [11] and nonlinear trend filtering [12] applications. These algorithms have the advantage over other numerical convex optimization methods, of allowing computation of the entire set of solutions for all values of the regularization parameters. The approximate path-following algorithm introduced in this paper exploits analytical solutions which afford considerably reduced computational complexity by comparison to numerical optimization methods such as interior-point or active set methods commonly used in sparse signal processing and machine learning [13].

2. METHODS

2.1. Spike-like signals with additive band-limited and i.i.d. noise

In this study we refer to a digital signal as the N-vector $x = (x_1, x_2 \dots x_N)^T$ with each $x_n \in \mathbb{R}$. We are considering the situation in which the measured signal x is the sum of three components: the spike-like signal u, a band-limited noise component v, and i.i.d. Gaussian noise w, so that $x = u + v + w$. The goal of the filtering operation is to recover an accurate estimate for u given x.

2.2. Filtering by convex optimization

The approach taken in this study is to pose the digital filtering operation as the solution to a convex optimization problem. The value of this approach is that we can exploit convex regularization as a tool to develop an efficient filtering method. The following objective functional is introduced:

$$H[x, y] = E[x, y] + \lambda J_L[y] + \gamma J_N[y] \qquad (1)$$

where y is the filter output signal, $E[x, y] = \|x - y\|_2^2$ is the *error term*, $J_L[y] = \|By\|_2^2$ is the *linear penalty*, and $J_N[y] = \|y\|_1$ is the *nonlinear penalty*, with corresponding scalar regularization parameters $\lambda, \gamma \geq 0$. Here, $\|\cdot\|_p$ refers to the L_p-norm. The $N \times N$ matrix B determines a linear filtering operation, as described below. The functional H is the sum of three terms: an error term E that is minimized (exactly zero) when $x = y$, and two penalty terms (J_L, J_N) that are minimized (and both zero) when $y = 0$. The goal of the filter is to take as input, the noisy signal x and return $u' = y$ as the de-noised signal.

Minimization of so-called elastic net functionals [8] such as (1), which involve a sum of L_2- and L_1-norm terms, are known as quadratic programs; state-of-the-art optimization methods include primal-dual interior-point and active sets [13]. In our experiments, we used the MATLAB tool CVX which automatically selects an appropriate minimization algorithm [14].

2.3. Linear filtering term in isolation

To analyze the behaviour of the filtering operation, it is instructive to consider the effect of the linear and nonlinear penalty terms of (1) in isolation. By setting $\gamma = 0$, the choice of the matrix B is directly related to the LTI filtering operation applied to the signal x. Then (1) reduces to a sum of two L_2-norm terms, and the minimizer can be computed analytically:

$$u' = (I + \lambda B^T B)^{-1} x = B' x \qquad (2)$$

where I is the $N \times N$ identity matrix. Because we are only considering the time-invariant case, the rows of B will consist of copies of the same linear filtering row vector, time shifted by one place. As a result of this shift-invariant structure, the equivalent filtering operation B' is also shift-invariant. In fact, the rows of B' can be understood as the coefficient sequence of a single FIR filter, and B' as a convolution operator based upon this FIR filter [1].

In practice, we often want to design a specific set of FIR filter coefficients [1]. For example, we may know which frequencies in the input signal we wish to attenuate. We can then insert these FIR coefficients into the rows of B', and from this solve for B:

$$B = \lambda^{-\frac{1}{2}} C \qquad (3a)$$

$$C = \left[B'^{-1} - I\right]^{\frac{1}{2}} \qquad (3b)$$

where $[\cdot]^{\frac{1}{2}}$ is any matrix square-root operation. Using (3a,b), we can utilize any given FIR filter design to calculate C. Note that B is simply a rescaling of C, so changing λ just involves rescaling the elements of C. A very useful set of FIR coefficients are those that define a whitening filter that attempts to make the power spectrum of the signal constant [1]. These coefficients can be estimated from x using a variety of techniques including linear prediction analysis [1].

It is important to note the analysis of this section only holds if the nonlinear term $\gamma = 0$: if not, the relationship between the minimizer for (1) and the matrix B' cannot be calculated analytically.

2.4. Nonlinear filtering term in isolation

We next consider the case when $\lambda = 0$, so that (1) reduces to a sum of L_2-norm error and L_1-norm penalty terms. As above, there is a simple analytical solution [11]:

$$u_n' = \left(|x_n| - \tfrac{1}{2}\gamma\right)^+ \operatorname{sgn}(x_n) \qquad (4)$$

for $n = 1,2 \ldots N$, where $(\cdot)^+$ indicates the positive part. This *shrinkage* operation independently reduces the absolute value of each element in the signal by $\gamma/2$. Because the assumption is that each element of the signal is the sum of a spike and an independent noise term with small spread relative to the spike magnitude, this operation typically attenuates the noise leaving the spike (although the magnitude of the spike is also reduced).

2.5. An approximate path-following minimization algorithm

As explained above, the functional (1) can be minimized using numerical convex optimization methods. Such methods have good computational properties but they are complex. Here we propose a simple approach that makes use of the isolated linear and nonlinear analytical solutions discussed in the previous two sections. The basic idea is to take small 'steps' towards the optimal filtering solution to (1) by alternately applying equation (2) followed by (4). This is related to so-called path-following (or homotopy) algorithms [11], which minimize the functional (1) by incrementally calculating the solution for increasing (or decreasing) values of the regularization parameters.

To motivate this numerical method, we first denote the operation of equation (2) on x by $F_{\lambda}[x]$, and of equation (4) by $G_{\gamma}[x]$. It is straightforward to show that $F_{\lambda}^{k}[x] = F_{k\lambda}[x]$ and $G_{\gamma}^{k}[x] = G_{k\gamma}[x]$, i.e. if we apply (2) or (4) repeatedly, this will be the same as applying each equation once but with λ replaced by $k\lambda$ (similarly for γ). It is also true that if λ is small, then the minimizer for (1) will be very close to that with $\lambda = 0$, and similarly for γ small. Putting these facts together, if, for example, we set $\lambda_1 = \lambda/R, \gamma_1 = \gamma/R$ and choose R sufficiently large, we can see that it is possible to come close to the minimizer for (1) by alternately assuming that $\gamma_1 = 0$ and $\lambda_1 = 0$ and iterating approximately R times.

Summarizing, the algorithm is as follows:

1. Initialization: set $\mu^0 = x$, compute $E[x, \mu^0]$ and choose $R \gg 1$. Set $\lambda_1 = \lambda/R, \gamma_1 = \gamma/R$ and $k = 1$. Compute $B' = (I + \lambda_1 B^T B)^{-1}$.
2. Set $\mu^k = G_{\gamma_1}\left[F_{\lambda_1}[\mu^{k-1}]\right]$.
3. Set $k = k + 1$.
4. Compute $E[x, \mu^k]$. If $|E[x, \mu^k] - E[x, \mu^{k-1}]| > \epsilon$, go to step 2.
5. Otherwise, set $u' = \mu^k$ and finish.

In step 4, we monitor the interim solution μ^k for convergence: the tolerance ϵ should be chosen to detect when the iteration has achieved sufficient accuracy.

2.5. A simple peak-picking heuristic

As a benchmark in testing the performance of the novel filter, we describe a peak-picking heuristic often used as a basis for more sophisticated spike detection algorithms: simply retain a fraction p of the largest magnitude elements of x, and set the rest to zero. This is related to simple thresholding methods which are ubiquitous in electrophysiological recordings, for example [7].

2.6. Measuring filtering accuracy

In order to quantify the accuracy of these filtering algorithms, we compute the *mean absolute error* (MAE) at recovering the spike signal u given the filtered output u':

$$MAE[u', u] = \frac{1}{N} \sum_{n=1}^{N} |u'_n - u_n|$$

This quantity cannot be negative and is zero only if the filter is perfect (i.e. $u' = u$), therefore, the smaller the MAE, the better the accuracy.

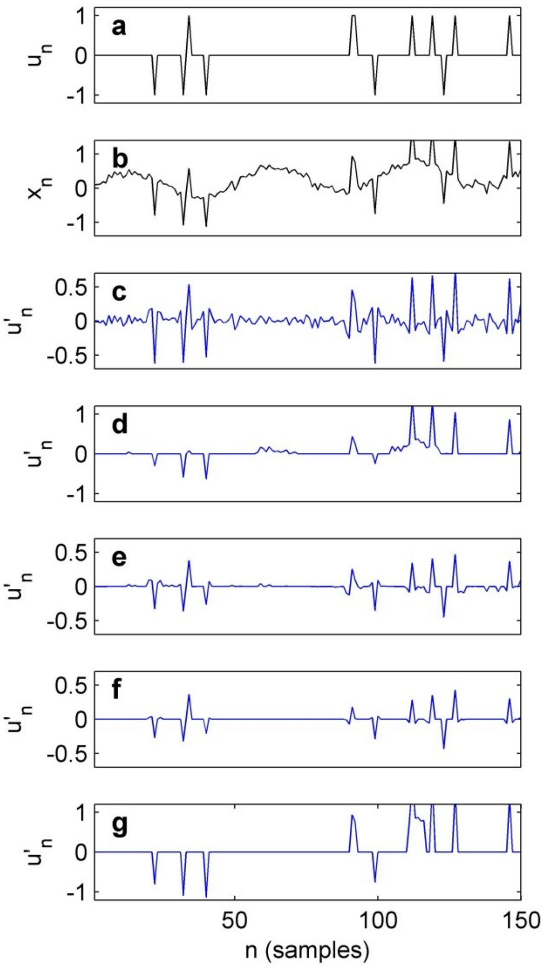

Fig. 1. (a) A synthetic spike-like signal ($N = 150$), showing equal-magnitude positive and negative spikes. (b) The same signal with additive Gaussian i.i.d. noise, linear drift (from 0 at $n = 0$ to 0.5 at $n = N$), and sinusoidal noise ($0.4\sin[6\pi/N]$). (c) The signal in (b) after linear filtering with equation (2), B is the first-order discrete integrator matrix, $\lambda = 0.8$, MAE = 0.095. (d) After nonlinear filtering with equation (4), $\gamma = 0.5$, MAE = 0.070. (e) After solving for the optimal solution to the convex problem posed in (1), regularization parameters as above, MAE = 0.066. (f) After approximate path-following, MAE = 0.061. (g) After peak-picking heuristic algorithm, $p = 0.1$, MAE = 0.064.

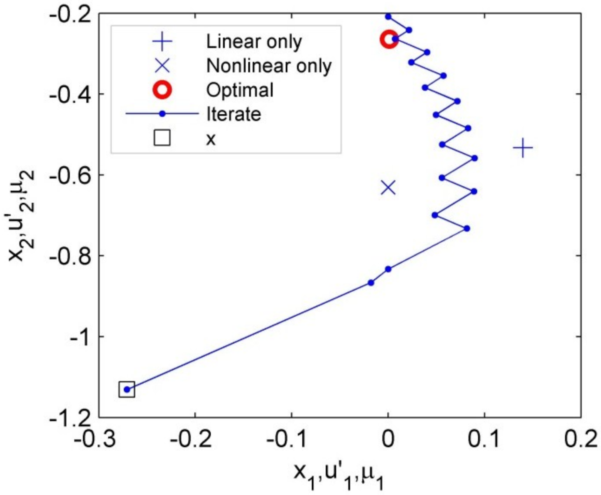

Fig. 2. The signal space projected onto two dimensions, before and after filtering. Input signal x (black square) is the starting point for the approximate path-following algorithm, which comes close to the minimizer for (1) (red circle) obtained by solving the quadratic programming problem. The linear only (+) solution is obtained using (2), and nonlinear only (×) obtained using (4). Synthetic signal and algorithm parameters as in Fig. 1.

3. EXPERIMENTAL DATA

In order to test the filter on real-world spike-like data, neural signals were obtained from tetrodes (a cluster of four 12.5 mm Ni-Cr wires twisted together) implanted and advanced towards the striatum of a freely-moving, awake rat. Signals were band-pass filtered in hardware (1Hz–9kHz) and sampled at a rate of 31.25kHz. Signals were downsampled offline using polyphase anti-alias filtering to 3kHz. One signal from one wire of a tetrode near a single neuron (a presumed Fast Spiking Interneuron [5]) was used for analysis.

Table 1. Mean absolute error (MAE) of the peak-picking benchmark, minimizer for the functional (1) obtained by convex optimization, and approximate minimizer for (1). Synthetic signal and algorithm parameters as in Fig. 1.

Gaussian i.i.d. noise standard deviation σ	Peak-picking MAE	Optimal solution to (1), convex optimization MAE	Path-following algorithm MAE
0.08	0.060	0.063	0.046
0.20	0.062	0.064	0.049
0.50	0.061	0.064	0.071
1.00	0.073	0.064	0.048
1.50	0.080	0.067	0.054

4. RESULTS AND DISCUSSION

Synthetic spike-like signals (Fig. 1a), with additive noise (Fig. 1b) are, as expected, clearly very difficult to reconstruct using purely linear (Fig. 1c) or purely nonlinear (Fig. 1d) filtering. Neither of these two filters alone can obtain good estimates of the real spike-like signal (Fig. 1a). The approximate path-following solution (Fig. 1e) comes very close to the optimal solution (Fig. 1e), although it overshoots somewhat (Fig. 2). This close approximation of the approximate algorithm to the optimal solution can also be seen in the signal space (Fig. 2). While the peak-picking estimate (Fig. 1g) removes most of the i.i.d. noise, it is unable to effectively handle the drift and sinusoidal noise.

As the spread of Gaussian i.i.d. noise increases, the approximate solver achieves the smallest MAE in reconstructing the spike signal u (Table 1). Although the optimal solution to (1) has higher MAE than the iterative solver, among the three filtering algorithms, it is the most resistant to increasing i.i.d. noise spread. The peak-picking benchmark achieves good performance at low noise spread, but deteriorates rapidly as this spread increases.

Here, the algorithm parameters $(\lambda, \gamma, B, p, R, \epsilon)$ have been chosen to achieve good performance at low i.i.d. noise spread, and kept constant as the noise spread increases; in practice these parameters might be adapted to the increasing noise. This could be achieved with appropriate cross-validation methods, for example. Nonetheless, this investigation shows that the minimizer for (1) is quite robust to changing i.i.d. noise spread, whereas, the naïve peak-picking benchmark can only function effectively if the parameters are chosen carefully. This occurs because the peak-picking algorithm cannot effectively distinguish between false spikes caused by high noise spread and genuine spikes. Similarly, although the approximate algorithm can achieve good performance at low noise spread, because it only approximately minimizes (1), it has variable performance that might require controlled adaptation.

The novel filter in (1) is very effective at removing band-limited noise from the experimental neural spike data (Fig. 3a,c), using linear prediction analysis to estimate a whitening filter in B'. Most of the i.i.d. noise is also removed while retaining all of the largest spikes (together with some smaller fluctuations that are probably spurious). The iterative solver is also similarly effective (Fig. 3d), converging in 19 iterations. The simple peak-picking heuristic, however, fails to identify most of the major spikes and also introduces a large number of spurious spikes (Fig. 3b).

Experimentation shows that by comparison to the 'gold standard' primal-dual interior point method, the approximate algorithm presented here achieves useful convergence with a similar number of iterations, but each iteration is extremely simple.

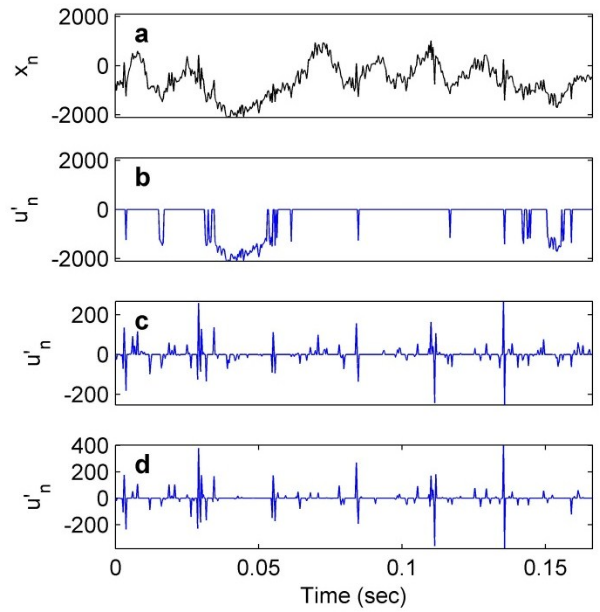

Fig. 3. Nonlinear filtering of experimental tetrode recordings of neural spikes. (a) Recorded signal: spikes visibly obscured by low-frequency drift and i.i.d. noise. (b) Recorded signal after applying peak-picking heuristic algorithm, $p = 0.2$. (c) Signal after solving for optimal solution to the convex problem (1), regularization parameters $\lambda = 100, \gamma = 500$. Here B' is a whitening filter of length 256 estimated using linear prediction analysis from the recorded signal, B derived from B' using (3). (d) After applying the path-following algorithm, parameters as above, with $R = 50$ and $\epsilon = 1 \times 10^5$.

5. SUMMARY AND CONCLUSIONS

In this study we introduced a novel, linear-nonlinear digital filter for recovering spike-like signals corrupted by band-limited and i.i.d. Gaussian noise. The filter is posed as a convex optimization problem, minimized by an approximate path-following algorithm. By comparison to a widely-used benchmark peak-picking heuristic, this new filter is robust to a wide range of i.i.d. noise spread. The filter also demonstrates good performance at recovering neural firing spikes from noisy electrophysiological recordings obtained from tetrodes implanted in a rat brain.

One of the advantages of this elastic net filter formulation, is the availability of a computationally simple, approximate path-following approach based upon known analytical solutions. This allows us to efficiently process very long signals, whereas, more general techniques from convex optimization such as primal-dual interior point (applied to, for example, dictionary-based sparse recovery which might be attempted in this application [9]) would typically require far heavier computational resources. Therefore, we anticipate this technique being useful in the common experimental circumstances where getting a quick, approximate result for large databases of very long signals is required.

Further comparisons against the performance of other spike-enhancement techniques [7,15] applied to databases of very long signals would be valuable to explore the practical trade-offs between accuracy and computational effort for this problem of spike enhancement.

6. ACKNOWLEDGEMENTS

The author would like to thank Greg Gage for providing the electrophysiology data used in this paper. Max Little is funded by the Wellcome Trust, grant number WT090651MF.

7. REFERENCES

[1] Proakis, J.G. and D.K. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*, Prentice Hall, Upper Saddle River, NJ, 2006.
[2] M.A. Little and N.S. Jones, "Generalized methods and solvers for noise removal from piecewise constant signals: Part I – Background theory", *Proceedings of the Royal Society A*, 467(2135) pp. 3088-3114, 2011.
[3] M.A. Little and N.S. Jones, "Sparse Bayesian step-filtering for high-throughput analysis of molecular machine dynamics", *2010 IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings* pp. 4162-4165, 2010.
[4] Arce, G.R. *Nonlinear Signal Processing: A Statistical Approach*, Wiley-Blackwell, Hoboken, NJ, 2004.
[5] G.J. Gage, C.R. Stoetzner, A.B. Wiltschko and J.D. Berke, "Selective Activation of Striatal Fast-Spiking Interneurons during Choice Execution", *Neuron*, 67 pp. 466–479, 2010.
[6] J. Berke, M. Okatan, J. Skurski, and H. Eichenbaum, "Oscillatory entrainment of striatal neurons in freely moving rats", *Neuron*, 43(6) pp. 883–96, 2004.
[7] S. B. Wilson and R. Emerson, "Spike detection: a review and comparison of algorithms", *Clinical Neurophysiology*, 113, pp. 1873–1881, 2002.
[8] H. Zou and T. Hastie, "Regularization and Variable Selection via the Elastic Net". *Journal of the Royal Statistical Society*, *Series B*, 67(2) pp. 301–320, 2005.
[9] D.L. Donoho, "Compressed sensing", *IEEE Transactions on Information Theory*, 52(4) pp. 1289–1306.
[10] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen and M. Moonen, "Sparse Linear Prediction and Its Applications to Speech Processing", *IEEE Transactions in Audio, Speech and Language Processing*, 20(5) pp. 1644–1657, 2012.
[11] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2nd ed., Springer, 2008.
[12] R.J. Tibshirani and J. Taylor, "The solution path of the generalized lasso", *Annals of Statistics*, 39(3) pp. 1335–1371.
[13] Boyd, S. and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
[14] Grant, M. and S. Boyd. *CVX: Matlab software for disciplined convex programming*, Version 2.0 Beta. http://cvxr.com/cvx, 2012.
[15] M. Mboup, "Neural spike detection and localisation via Volterra filtering", *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*. IEEE, 2012.