ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096734

BEYOND RATE CODING: SIGNAL CODING AND RECONSTRUCTION USING LEAN SPIKE TRAINS

Anik Chattopadhyay, Arunava Banerjee

University of Florida
Computer & Information Science & Engineering
Gainesville, FL 32611

ABSTRACT

Recent years have seen a growing interest in spike based encoding of continuous time signals–a hallmark of biological computation. In this context, we present a mathematical framework for signal representation, leveraging a simple but robust mechanistic model of a biologically plausible spiking neuron. The framework considers encoding of a signal through spike trains generated by an ensemble of neurons via a standard convolve-then-threshold mechanism, albeit with a wide variety of convolution kernels. Reconstruction is posited as a convex optimization minimizing energy. Formal conditions under which perfect and approximate reconstruction of the signal from the spike trains is possible are then identified. The strength of the framework is shown in experiments on a large audio dataset, demonstrating good reconstruction at a spike rate of one fifth the Nyquist rate. Comparison against a benchmark sparse coding technique, viz convolutional orthogonal matching pursuit, shows competitive results in reconstruction with orders of magnitude improvement in runtime efficiency.

Index Terms— coding, integrate-and-fire, spike, convolution, reconstruction

1. INTRODUCTION

In most animals, sensory stimuli are communicated to the brain via ensembles of discrete spatio-temporally compact electrical events generated by neurons, known as action potentials or spikes [1]. It is widely assumed that this representation of a continuous time signal using spike trains is achieved via a rate code. In signal processing, this has been formally analyzed under the banner of pulse density coding and Δ-Σ modulators [2, 3]. There is however evidence (e.g., the H1 neuron in the fly [4]) that there are spike codes that achieve high reconstruction accuracy while being much sparser/leaner than would be warranted by a rate code. Sparse/lean spike trains are not only intrinsically energy efficient, but can also facilitate downstream computation[5, 6]. Although there has been substantial progress, an effective end to end signal processing framework that deterministically represents signals via lean spike train ensembles is yet to be laid out. Here we present a new framework for coding and reconstruction leveraging a biologically plausible coding mechanism which is a superset of the standard leaky integrate-and-fire neuron model [7].

Our proposed framework identifies reconstruction guarantees for a very general class of signals—those with *finite rate of innovation* [8]—as shown in our perfect and approximate reconstruction theorems. Most other classes, e.g. bandlimited signals, are subsets of this class. The proposed technique first formulates reconstruction as an optimization that minimizes the energy of the reconstructed signal subject to consistency with the spike train, and then solves it in closed form. We then identify a general class of signals for which reconstruction is provably perfect under certain ideal conditions. We then present a mathematical bound on the error of an approximate reconstruction when the model deviates from those ideal conditions. Finally, we present simulation experiments coding and reconstructing a large dataset of audio signals that demonstrate the efficacy of the framework.

2. CODING

For encoding, we consider the set of input signals ($\mathcal{F}$) to be the class of all finite support *bounded square integrable function* (i.e. formally $\mathcal{F} = \{X(t)|X(t) \in L^2[0,T]\}$, for some $T \in \mathbb{R}^+$, where the choice of T could be arbitrary w.l.o.g), which satisfy a finite rate of innovation bound.

We assume an ensemble of spiking neurons $K = \{K^j | j \in Z^+, j \leq n\}$, each characterized by a kernel function $K^j(t), j = 1,\ldots,n$, where $\forall j \in \{1,\ldots,n\}, K^j(t) \in C[0,T]$, $T \in \mathbb{R}^+$. Finally, we assume that K^j has a time varying threshold denoted by $T^j(t)$.

The ensemble of convolution kernels K encodes a given input signal $X(t)$ into a sequence of spikes $\{(t_i, K^{j_i})\}$, where the i^{th} spike is produced by the j_i^{th} kernel K^{j_i} at time t_i if and only if: $\int X(\tau)K^{j_i}(t_i-\tau)d\tau = T^{j_i}(t_i)$ In our experiments a simplified threshold function is assumed in which the time varying threshold $T^j(t)$ of the jth kernel remains constant at C^j until that kernel produces a spike, at which time an *after-hyperpolarization potential (ahp)* increments the threshold by a value M^j. The increment drops back to zero linearly within

Authorized licensed use limited to: IEEE Xplore. Downloaded on June 06,2023 at 14:37:40 UTC from IEEE Xplore. Restrictions apply.

a refractory period δ_j. Formally,

$$T^j(t) = C^j + \sum_{t_p^j \in [t, t-\delta_j]} M^j - \frac{M^j(t - t_p^j)}{\delta_j} \tag{1}$$

Where the sum is taken over all spike times t_p^j in the interval t to $t - \delta_j$ at which the kernel K^j generated a spike. It is worth noting that this seemingly simple threshold function carries the essence of efficient biological signal encoding as will be evident from the subsequent sections both from theory and implementation perspectives. Such a threshold allows a neuron to stay quiescent as long as the signal is uncorrelated with its kernel K^j; it starts firing when the correlation reaches a certain threshold and continues to fire at higher threshold levels communicating increasing correlation levels, only inhibited by previous spikes. This phenomena of probing signals via a lockstep series of spikes is depicted in fig1 for one kernel.

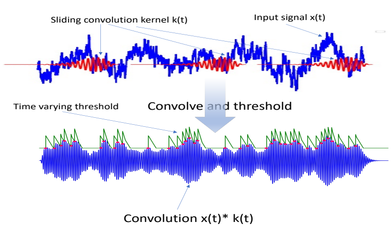

Fig. 1: The convolve and threshold mechanism described in the coding model for a single kernel. Top: a sample signal (in blue) is shown overlayed with a convolution kernel (in red). Below: the result of convolution in blue and the threshold function for the kernel in green. Spikes times are marked at the threshold crossing level with red dots.

3. DECODING

The objective of the decoding module is to reconstruct the original signal from the encoded spike trains. Considering the prospect of the invertibility of the coding scheme, we seek a signal that satisfies the same set of constraints as the original signal when generating all spikes apropos the set of kernels in ensemble K. Recognizing that such a signal might not be unique, we choose the reconstructed signal as the one with minimum $L2$-norm. Formally, the reconstruction (denoted by $X^*(t)$) of the input signal $X(t)$ is formulated to be the solution to the optimization problem:

$$\begin{aligned} X^*(t) &= \underset{\tilde{X}}{\text{argmin}} ||\tilde{X}(t)||_2^2 \\ \text{s.t.} \int \tilde{X}(\tau) K^{j_i}(t_i - \tau) d\tau &= T^{j_i}(t_i); 1 \le i \le N \end{aligned} \tag{2}$$

where $\{(t_i, K^{j_i}) | i \in \{1, ..., N\}\}$ is the set of all spikes generated by the encoder. The choice of $L2$ minimization is in congruence with the dictum of energy efficiency in biological systems. The assumption is that, of all signals, the one with the minimum energy that is consistent with the spike trains is desirable. Also, an $L2$ minimization in the objective of (2) reduces the convex optimization problem to a solvable linear system of equations as described below.

4. SIGNAL CLASS FOR PERFECT RECONSTRUCTION

We observe that in general the encoding of $L^2[0, T]$ signals into spike trains is not an injective map; the same set of spikes can be generated by different signals so as to result in the same convolved values at the spike times. Naturally, with a finite and fixed ensemble of kernels K, one cannot achieve perfect reconstruction for all $L^2[0, T]$ signals. Assuming, additionally, a finite rate of innovation, as $\mathcal{F}$ was previously defined changes the story. We now restrict ourselves to a subset $\mathcal{G}$ of $\mathcal{F}$ defined as $\mathcal{G} = \{X(t) | X(t) \in \mathcal{F}, X(t) = \sum_{p=1}^{N} \alpha_p K^{j_p}(t_p - t), j_p \in \{1, ..., n\}, \alpha_p \in \mathbb{R}, t_p \in \mathbb{R}^+, N \in Z^+\}$ and address the question of reconstruction accuracy. Essentially $\mathcal{G}$ consists of all linear combinations of arbitrarily shifted inverted kernel functions. N is bounded above by the total number of spikes that the ensemble K can generate over $[0, T]$. For the class $\mathcal{G}$ the *perfect reconstruction theorem* is presented below. The theorem is proved with the help of two lemmas.

Perfect Reconstruction Theorem: Let $X(t) \in \mathcal{G}$ be an input signal. Then for appropriately chosen time-varying thresholds of the kernels, the reconstruction, $X^*(t)$, resulting from the proposed coding-decoding framework is accurate with respect to the $L2$ metric, i.e., $||X^*(t) - X(t)||_2 = 0$.

Lemma1: The solution $X^*(t)$ to the reconstruction problem given by (2) can be written as: $X^*(t) = \sum_{i=1}^{N} \alpha_i K^{j_i}(t_i - t)$ where the coefficients $\alpha_i \in \mathbb{R}$ can be uniquely solved from a system of linear equations if the shifted kernel functions $K^{j_i}(t_i - t)$ are *linearly independent*.

Proof: An argument similar to that of the Representer Theorem [9] on (2) directly results in: $X^*(t) = \sum_{i=1}^{N} \alpha_i K^{j_i}(t_i - t)$ where the α_i's are real valued coefficients. This holds true because any component of $X^*(t)$ orthogonal to the span of the $K^{j_i}(t_i - t)$'s does not contribute to the convolution (inner product) constraints. In essence, $X^*(t)$ is an orthogonal projection of $X(t)$ on the span of shifted kernels $\{K^{j_i}(t_i - t) | i \in 1, 2, ..., N\}$. Therefore, the coefficients can be derived by solving the linear system: $P\alpha = T$ where P is the $N \times N$ Gram matrix of the shifted kernels $\{K^{j_i}(t_i - t) | i \in 1, 2, ..., N\}$ in the Hilbert space with standard inner product, $[P]_{ik} = \langle K^{j_i}(t_i - t), K^{j_k}(t_k - t) \rangle$, and $T = \langle T^{j_1}(t_1), ..., T^{j_N}(t_N) \rangle^T$. Furthermore, the system has a unique solution since the Gram Matrix P is invertible. This happens because the set of constituent spiking kernels $K^{j_i}(t_i - t)$ grows as a linearly independent set because of the ahp effect which ensures separation in time and therefore a new kernel cannot be fully represented in the span of previous kernels.

Authorized licensed use limited to: IEEE Xplore. Downloaded on June 06,2023 at 14:37:40 UTC from IEEE Xplore. Restrictions apply.

Lemma2: Let $X^*(t)$ be the reconstruction of an input signal $X(t)$ and $\{(t_i, K^{j_i})\}_{i=1}^N$ be the set of spikes generated. Then, for any arbitrary signal $\tilde{X}(t)$ within the span of $\{K^{j_i}(t_i - t)|i \in \{1, 2, ..., N\}\}$, i.e., the set of shifted inverted kernels at respective spike times, given by $\tilde{X}(t) = \sum_{i=1}^N a_i K^{j_i}(t_i - t)$ the following holds: $||X(t) - X^*(t)|| \leq ||X(t) - \tilde{X}(t)||$
Proof: This follows from the fact that $X^*(t)$ is an orthogonal projection on the span.

Exploring further, for a given input signal $X(t)$ if S_1 and S_2 are two sets of spike trains where $S_1 \subset S_2$ produced by two different kernel ensembles, the second a superset of the first, then Lemma 2 further implies that the reconstruction due to S_2 is at least as good as the reconstruction due to S_1 because the reconstruction due to S_1 is in the span of the shifted kernel functions of S_2 as $S_1 \subset S_2$. This immediately leads to the conclusion that for a given input signal the more kernels we add to the ensemble the better the reconstruction, provided the kernels maintain linear independence.
Proof of the Theorem: The proof of the theorem follows directly from Lemma 2. Since the input signal $X(t) \in \mathcal{G}$, let $X(t)$ be given by: $X(t) = \sum_{p=1}^N \alpha_p K^{j_p}(t_p - t) \quad (\alpha_p \in \mathbb{R}, t_p \in \mathbb{R}^+, N \in Z^+)$ Assume that the time varying thresholds of the kernels in our kernel ensemble K are set in such a manner that the following conditions are satisfied: $\langle X(t), K^{j_p}(t_p - t)\rangle = T^{j_p}(t_p) \quad \forall p \in \{1, ..., N\}$ i.e., each of the kernels K^{j_p} at the very least produces a spike at time t_p against $X(t)$ (regardless of other spikes at other times). Clearly then $X(t)$ lies in the span of the appropriately shifted and inverted response functions of the spike generating kernels. Applying Lemma 2 it follows that: $||X(t) - X^*(t)||_2 \leq ||X(t) - X(t)||_2 = 0$

5. APPROXIMATE RECONSTRUCTION:

A practical challenge on which the reconstruction accuracy depends is whether one can generate spikes at the correct temporal locations. The lockstep time-varying threshold (1) alludes to the fact that spikes could be produced arbitrarily close to the desired time points, as adopted in our experiments in Section 6, by setting the C^j's, M^j's and the δ_js at a reasonably low value. At this point we need to evaluate the deviation of the spike times from certain points of interest in the setting of the proposed thresholding scheme, as addressed next.
Lemma 4: Let $X(t)$ be an input signal. Let K^p be a kernel for which we want to generate a spike at time t_p. Let the inner product $\langle X(t), K^p(t_p - t)\rangle = I^p$. Then, if the baseline threshold of the kernel K^p is $C^p \leq I^p$ and the absolute refractory period is δ as modeled in Equation 1, the kernel K^p must produce a spike in the interval $[t_p - \delta, t_p]$ according to the threshold model defined in Equation 1.
Proof: The proof of the lemma appeals to the nature of the lockstep threshold function defined in 1 and follows from the intermediate value theorem.

The next obvious question that arises is: how much error does one perceive in reconstruction as the spikes deviate from the location as desired in the Perfect Reconstruction Theorem for factors such as noise, the signal not being perfectly represented in the span of kernel functions, etc. We try to identify those conditions in the following lemma:
Lemma 5: Let $X(t)$ be represented as $X(t) = \sum_{i=1}^N \alpha_i f_{p_i}(t_i - t), \alpha_i \in \mathcal{R}^+$, where $f_{p_i}(t)$ are bounded functions on finite support that constitute the input signal. Assume that there is at least one kernel function in the ensemble for which $||f_{p_i}(t) - K^{j_i}(t)||_2 < \delta \ \forall i \in \{1, ..., N\}$. Also assume that the framework is able to produce a spike within γ (for some δ and $\gamma \in \mathcal{R}^+$) interval of $t_i, \forall i$. Also assume that the constituents f_{p_i} satisfy a frame bound type of condition: $\sum_{k \neq i} \langle f_{p_i}(t - t_i), f_{p_k}(t - t_k)\rangle \ \leq \eta \ \forall i \in \{1, ..., N\}$ and the kernel functions are Lipschitz continuous. Under such conditions, the $L2$ error in reconstruction of $X^*(t)$ is bounded.
Proof: The proof of this theorem follows from continuity arguments and the use of bounds on the eigen values of the Gram matrix P.

Effect of Ahp, Stability of the solution and Windowing: The combination of Lemmas 4 & 5 shows that even under non ideal conditions, the reconstruction of our technique only suffers from bounded error, although this comes at the cost of increasing spike rates, as dictated by Lemma 4. High spike rates has the adverse effect of worsening the condition number of P. An instability in the P matrix practically renders the solution useless from an application standpoint where spike times are represented in finite precision floating point numbers, leading to quantization error. Indeed one can show that for a general positive definite matrix P the condition number can grow exponentially large as the size of the matrix grows. This problem is partially mitigated by the effect of the ahp for a finite size of P by ensuring linear independence among the spikes. But it can still get worse as we process longer signals and the size of P grows arbitrarily large. This is where the combined effect of causality of the kernels and the ahp comes to our defense. We observe that the addition of the $n + 1^{th}$ spike on an existing set of n spikes can only affect the solution substantially within a finite time window which in turn is facilitated by the effect of the ahp, which ensures that new spikes maintain reasonable separation with previous spikes and therefore have a fading affect on the reconstruction back into the past. This simple observation enables us to encode and reconstruct signals in an online mode within a finite window of spikes, leading to remarkable efficiency when tested on real datasets as shown in Section 6.

6. EXPERIMENTS ON REAL SIGNALS

The proposed framework was tested on audio signals. **Dataset:** We chose the Freesound Dataset Kaggle 2018, an audio dataset of natural sounds referred in [10], containing 18,873 audio files.

Authorized licensed use limited to: IEEE Xplore. Downloaded on June 06,2023 at 14:37:40 UTC from IEEE Xplore. Restrictions apply.

All audio samples in this dataset are provided as uncompressed PCM 16bit, 44.1kHz, mono audio files. **Set of Kernels:** We chose gammatone filters ($at^{n-1}e^{-2\pi bt}\cos(2\pi ft + \phi)$) in our experiments since they are widely used as a reasonable model of cochlear filters in auditory systems [11].

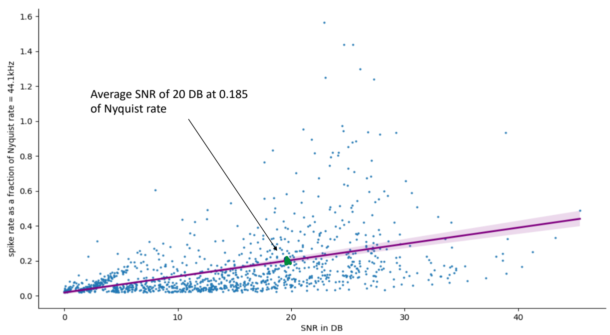

Fig. 2: Comprehensive result of an experiment with 50 kernels reconstructing 200 sound snippets. The graph shows a scatter plot of reconstructions (each dot represents a reconstruction) with spike-rate of the ensemble (x-axis) against corresponding SNR value of the reconstructions (y-axis). The trend line is generated using seaborn regression fit.

Results: In all experiments, kernels were normalized, and the parameters of the time-varying threshold function (1) were initialized through a systematic grid search. We processed each audio snippet of length $\approx 1 - 2s$ with a fixed set of parameter values, except for the refractory period which was decreased gradually leading to improvement in reconstructions at increasingly higher spike rates.

The results of our experiment, as shown in figure 2, demonstrate that by increasing the spike rate sufficiently by tuning the refractory period, near perfect reconstruction is feasible which is in agreement with our theoretical analysis. We see some variability in reconstruction accuracy across signals, which could be attributed to the idiosyncrasies of the dataset–certain audio samples could be noisy or ill-represented in the kernels. But the overall trend on reconstruction shows great promise, on average showing $\approx 20dB$ at $1/5$th Nyquist Rate. This in conjunction with the fact that signals are represented in this scheme only via set of spike times and kernel indexes (thresholds can be inferred) holds potential for an extremely efficient coding mechanism. Since generation of spikes requires scanning through convolutions in one pass, encoding is very efficient. Decoding is slightly time consuming because we solve a linear equation $P\alpha = T$ to derive the coefficients. But then reconstruction is done in an online manner on a finite window as described in Section 5. So the overall process still remains linear making it a suitable choice for lengthy continuous time signals.

Comparison With Convolutional Orthogonal Matching Pursuit:

Our framework is capable of discovering sparse representations of signals as demonstrated by the reconstruction theorems and thus becomes a competitor of existing sparse coding methods.

In another set of experiments we compared our technique against convolutional orthogonal matching pursuit. The results of the comparison are promising and are furnished in figure 3. Since OMP is inherently computationally inefficient, for these set of experiments we kept the length of signals relatively small ($\approx 100 - 200ms$) and used a reduced set of kernels (≈ 10 kernels). Figure 3 reports the results of ≈ 20 randomly chosen audio snippets from the dataset and our framework showed better reconstruction accuracy for majority of the snippets. This is likely attributable to the greedy nature of COMP which may not always capture the locally important features, unlike our technique. In terms of runtime complexity, our technique far outshines COMP, as in COMP one needs to constantly recalculate the convolutions and perform Gram-Schmidt. This effect is more pronounced as the signal grows longer. Admittedly, there are many state-of-the art sparse convolutional techniques based on $L1$ optimization, as shown in [12], [13], [14], [15], that perform faster than COMP, but our technique is supposed to outperform those asymptotically as it remains linear in the length of the signal due to the windowing mechanism described in section 5.

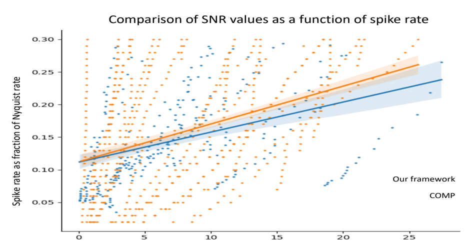

Fig. 3: Comparison between our framework vs COMP in terms of overall reconstruction accuracy and spike rate.

7. CONCLUSION

We have formulated a framework that identifies the precise conditions under which continuous time signals can be represented using an ensemble of spike trains, from which the signal can be recovered perfectly. Although aligned in their goals, this framework is very different from that investigated in Niquist-Shanon theory. The primary difference between the two lies in their respective modes of representation/coding. Instead of sampling the value of a function at uniform or non-uniform prespecified sample points the new coding scheme reports the (non-uniform) sample points where the function takes specific convolved values. Coding is intimately related to compression and our experimental results indicate great potential in this regard.

Authorized licensed use limited to: IEEE Xplore. Downloaded on June 06,2023 at 14:37:40 UTC from IEEE Xplore. Restrictions apply.

References

[1] Fred Rieke, Davd Warland, Rob de Ruyter van Steveninck, and William Bialek, *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, MA, USA, 1999.

[2] H Inose, Y Yasuda, and J Murakami, “A telemetering system by code modulation: δ-σ modulation,,” *IRE Transactions on Space Electronics Telemetry*, vol. 8, pp. 204–209, 1962.

[3] Ingrid Daubechies and Ron DeVore, “Approximating a bandlimited function using very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order,” *Annals of Mathematics*, vol. 158, no. 2, pp. 679–710, 2003.

[4] Ilya Nemenman, Geoffrey D Lewen, William Bialek, and Rob R de Ruyter van Steveninck, “Neural coding of natural stimuli: information at sub-millisecond resolution,” *PLoS computational biology*, vol. 4, no. 3, pp. e1000025, 2008.

[5] P. Földiák, “Forming sparse representations by local anti-hebbian learning,” *Biological Cybernetics*, vol. 64, no. 2, pp. 165–170, Dec 1990.

[6] Daniel Graham and David Field, “Sparse coding in the neocortex,” *Evolution of Nervous Systems*, vol. 3, pp. 181–187, 2007.

[7] Peter Dayan and L. F. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, The MIT Press, 2005.

[8] M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[9] Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola, “A generalized representer theorem,” in *International Conference on Computational Learning Theory*. Springer, 2001, pp. 416–426.

[10] Eduardo Fonseca, Manoj Plakal, Frederic Font, Daniel P. W. Ellis, Xavier Favory, Jordi Pons, and Xavier Serra, “General-purpose tagging of freesound audio with audioset labels: Task description, dataset, and baseline,” 2018.

[11] R. Patterson, Ian Nimmo-Smith, J. Holdsworth, and P. Rice, “An efficient auditory filterbank based on the gammatone function,” 01 1988.

[12] Garcia-Cardona et. al., “Convolutional dictionary learning: A comparative review and new algorithms,” *IEEE Transactions on Comp. Imaging*, vol. PP, pp. 1–1, 05 2018.

[13] Brendt Wohlberg, “Efficient convolutional sparse coding,” in *2014 IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2014, pp. 7173–7177.

[14] Heide et. al., “Fast and flexible convolutional sparse coding,” in *2015 IEEE Conf. on CVPR*, 2015, pp. 5135–5143.

[15] Brendt Wohlberg, “Boundary handling for convolutional sparse representations,” in *2016 IEEE ICIP*, 2016, pp. 1833–1837.

Authorized licensed use limited to: IEEE Xplore. Downloaded on June 06,2023 at 14:37:40 UTC from IEEE Xplore. Restrictions apply.