



Computing optimal discrete readout weights in reservoir computing is NP-hard

Fatemeh Hadaeghi*, Herbert Jaeger

Department of Computer Science and Electrical Engineering, Jacobs University Bremen, Bremen 28759, Germany



ARTICLE INFO

Article history:

Received 24 July 2018

Revised 6 December 2018

Accepted 6 February 2019

Available online 10 February 2019

Communicated by Dr. Chenchen Liu

Keywords:

Complexity

Linear regression

Neuromorphic hardware

Reservoir computing

Unconstrained quadratic programming

Unconventional computing

ABSTRACT

We show NP-hardness of a generalized quadratic programming problem, which we called *unconstrained n-ary quadratic programming* (UNQP). This problem has recently become practically relevant in the context of novel memristor-based neuromorphic microchip designs, where solving the UNQP is a key operation for on-chip training of the neural network implemented on the chip. UNQP is the problem of finding a vector $\mathbf{v} \in S^N$ which minimizes $\mathbf{v}^T Q \mathbf{v} + \mathbf{v}^T \mathbf{c}$, where $S = \{s_1, \dots, s_n\} \subset \mathbb{Z}$ is a given set of eligible parameters for \mathbf{v} , $Q \in \mathbb{Z}^{N \times N}$ is positive semi-definite, and $\mathbf{c} \in \mathbb{Z}^N$. In memristor-based neuromorphic hardware, S is physically given by a finite (and small) number of possible memristor states. The proof of NP-hardness is by reduction from the *unconstrained binary quadratic programming problem*, which is a special case of UNQP where $S = \{0, 1\}$ and which is known to be NP-hard.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The accustomed, apparently unbounded growth rate of digital computing technologies begins to show signs of flattening out toward a ceiling (“end of Moore’s law”). This is due to several reasons, among them the approaching of ultimate thermodynamical limits, the steeply growing costs of building foundries for the respective next-generation CMOS microchips, and technological difficulties in mastering device miniaturization. Furthermore, the energy hunger of classical digital computing technologies is increasingly becoming problematic, both for global energy management and for deploying ever more computing-intense AI algorithms on battery-powered personal devices. All of this has revived interest in “unconventional” computing research, with regards to non-digital material substrates, architectures and algorithms [1]. Among the wide diversity of approaches to unconventional computing, a leading role is emerging for non-digital implementations of artificial neural networks (ANNs). Two main trends in this arena are the exploit of low-energy, spiking neural dynamics for “deep learning” solutions [2], and “reservoir computing” (RC) methods [3]. In material RC implementations, a physical medium called the *reservoir* – it can be electronic [4], optical [5], nano-mechanical [6], macro-mechanical [7], or other [8] – is nonlinearly excited by

temporal input signals, and from the resulting, high-dimensional response signals within the medium a target output signal is delivered through a trainable readout mechanism. The work reported in this article arose within this latter line of investigation. Specifically, we are involved in a European collaborative project (NeuRAM3, www.neuram3.eu) which is concerned with the design of memristor-based, spiking neuromorphic microchips. The reservoir here is an analog VLSI recurrent neural network (RNN) – concretely, the currently available DYNAP-se [9] and its future descendants. The training of the readout mechanism amounts to solving a linear regression problem, where the target output is a trainable linear combination of the neural signals within the reservoir. Linear regression solutions find the combination weights which minimize the mean squared difference between the combined output signal and the reference target. Linear regression is easily solved by standard linear algebra algorithms when arbitrary real-valued combination weights are admitted. However, for on-chip learning, the weights will be physically realized by states of memristive “synapses”, which currently can be reliably set only to a very small number of discrete values. This situation has led us to investigate the computational nature of a “discrete” linear regression.

This article details our finding that here we are facing an NP-complete problem. In Section 2 we formalize this problem as an *unconstrained n-ary quadratic programming* (UNQP) problem and describe a known NP-complete problem, *unconstrained binary quadratic programming problem* (UBQP). In Section 3 we show

* Corresponding author.

E-mail address: f.hadaeghi@jacobs-university.de (F. Hadaeghi).

that UBQP can be reduced to UNQP in polynomial time, thereby demonstrating that UNQP is NP-hard.

2. Problem statement

In what follows, vectors are column vectors unless otherwise defined.

Reservoir computing is an approach to train recurrent neural networks (RNNs) in temporal signal processing tasks. The objective is to find a model RNN which approximates the input-output relation presented through matched training input-output signals $\mathbf{u}_{\text{train}}(i), y_{\text{train}}(i), i = 1, \dots, L$.

The update equations of an elementary, discrete-time RC neural network are

$$\mathbf{x}(i+1) = f(\mathbf{W}\mathbf{x}(i) + \mathbf{W}^{\text{in}}\mathbf{u}(i+1)) \quad (1)$$

$$y(i) = (\mathbf{w}^{\text{out}})^T \mathbf{x}(i), \quad (2)$$

where i is discrete time, $\mathbf{u}(i) \in \mathbb{R}^K$ is the input signal, $\mathbf{x}(i) \in \mathbb{R}^N$ is the reservoir state, f is a sigmoid function applied element-wise to its argument vector, $\mathbf{W}, \mathbf{W}^{\text{in}}$ are the recurrent and input weight matrices of size $N \times N, N \times K$, $y(i) \in \mathbb{R}$ is the output signal which is obtained by linearly combining the components of the reservoir state $\mathbf{x}(i)$ with the output weights \mathbf{w}^{out} (an N -dimensional vector). Numerous variants and extensions of this basic system are being considered in RC, but for our current purpose this simple model with a scalar output signal suffices.

Training an RC network proceeds in three stages:

1. Create random recurrent and input weight matrices \mathbf{W} and \mathbf{W}^{in} .
2. Drive the network (1) with the training input $\mathbf{u}_{\text{train}}(i)$, obtaining reservoir response signals $\mathbf{x}_{\text{train}}(i)$.
3. Compute output weights \mathbf{w}^{out} which minimize a loss function

$$\mathcal{L}(\{(\mathbf{w}^{\text{out}})^T \mathbf{x}_{\text{train}}(i), y_{\text{train}}(i)\}_{i=1, \dots, L}).$$

The most popular loss function by far used in RC is the quadratic loss, which leads to solutions \mathbf{w}^{out} that solve the linear regression problem

$$\underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{w}^T \mathbf{X} - \mathbf{y}\|^2, \quad (3)$$

where $\mathbf{X} = (\mathbf{x}_{\text{train}}(1), \dots, \mathbf{x}_{\text{train}}(L))$ and $\mathbf{y} = (y_{\text{train}}(1), \dots, y_{\text{train}}(L))$.

In the RC learning paradigm, \mathbf{w}^{out} is the only item which is trained. The randomly created parameters in \mathbf{W} and \mathbf{W}^{in} are not adapted. This latter condition has rendered RC interesting for unconventional hardware realizations of RNN-like learning systems, because in principle it allows one to employ any kind of nonlinearly excitable physical medium to instantiate the “reservoir” (1).

In almost all currently realized physical reservoir computers, the output weights \mathbf{w}^{out} are represented in a classical digital computer outside the unconventional physical reservoir. This means that floating-point precision numbers can be used in \mathbf{w}^{out} , and standard linear regression algorithms can be called upon. However, a widely considered goal for further development of unconventional RC systems is to integrate the output weights into the unconventional physical substrate. Much of today’s unconventional-substrate RC research is concerned with neuromorphic hardware where the “synaptic” parameters contained in $\mathbf{W}, \mathbf{W}^{\text{in}}, \mathbf{w}^{\text{out}}$ are realized by memristors. It is currently infeasible to fabricate or tune memristive weights to anything approaching floating-point precision. In fact, one has to face extreme low-precision scenarios, where realizable values of the elements in these matrices admit only ternary settings (for instance, taking approximate values $\{-s, 0, +s\}$) or other n -ary ranges $S = \{s_1, \dots, s_n\}$ with for some quite small n .

It is easily derived that the optimization problem (3) is equivalent to

$$\underset{\mathbf{w}}{\text{minimize}} \quad \mathbf{w}^T \mathbf{Q} \mathbf{w} + \mathbf{w}^T \mathbf{c}, \quad (4)$$

where $\mathbf{Q} = \mathbf{X}\mathbf{X}^T$ and $\mathbf{c} = -2\mathbf{X}\mathbf{y}^T$. This is the format in which the linear regression objective is written in optimization theory contexts, where it is called the *quadratic programming problem*.

In the unconventional RC hardware scenarios that motivated our investigation, the parameters admissible to be used in the components of \mathbf{w} are constrained to a finite set $S = \{s_1, \dots, s_n\}$ of n real numbers. Our aim is to show that the resulting constrained quadratic programming problem is hard in the sense that it does not admit a general polynomial-time solution algorithm (provided that $P \neq NP$). In order to show this, it suffices to show hardness for parameters in $\mathbf{Q}, \mathbf{c}, S$ restricted to the integers \mathbb{Z} . That is, we consider the following optimization problem, which we will call the *unconstrained n -ary quadratic programming (UNQP) problem*

$$\underset{\mathbf{w}}{\text{minimize}} \quad \mathbf{w}^T \mathbf{Q} \mathbf{w} + \mathbf{w}^T \mathbf{c}, \quad (5)$$

subject to $\mathbf{w} \in S^N$,

where $\mathbf{Q} \in \mathbb{Z}^{N \times N}$ is positive semi-definite, $\mathbf{c} \in \mathbb{Z}^N$, $S = \{s_1, s_2, \dots, s_n\} \subset \mathbb{Z}$ and $|S| \geq 2$.

Notice that UNQP is not a single optimization problem but a family of such problems. For every choice of S , we obtain a distinct optimization problem. We will show that for every such choice, the resulting problem is NP-hard.

3. UNQP is NP-hard

We now give a proof of

Theorem 3.1. *The problem UNQP from Eq. (5) is NP-hard.*

The proof is by transformation from the *unconstrained binary quadratic programming (UBQP) problem* for vectors \mathbf{v}

$$\underset{\mathbf{v}}{\text{minimize}} \quad \mathbf{v}^T \mathbf{Q} \mathbf{v} + \mathbf{v}^T \mathbf{c}, \quad (6)$$

subject to $\mathbf{v} \in \{0, 1\}^N$,

where $\mathbf{Q} \in \mathbb{Z}^{N \times N}$ is positive semi-definite and $\mathbf{c} \in \mathbb{Z}^N$. This problem is known to be NP-hard [10].

We assume that S is ordered, that is $s_1 < \dots < s_n$. Let \mathbf{s}_j (where $j = 1, \dots, n$) denote the N -dimensional vector whose components are all equal to s_j . Consider the problem

$$\underset{\mathbf{t}}{\text{minimize}} \quad \left(\frac{\mathbf{t} - \mathbf{s}_1}{s_2 - s_1} \right)^T \mathbf{Q} \left(\frac{\mathbf{t} - \mathbf{s}_1}{s_2 - s_1} \right) + \left(\frac{\mathbf{t} - \mathbf{s}_1}{s_2 - s_1} \right)^T \mathbf{c} \quad (7)$$

subject to $\mathbf{t} \in \{s_1, s_2\}^N$.

This is equivalent to UBQP, because a solution \mathbf{t} becomes a solution \mathbf{v} of (6) by replacing components $s_1 \mapsto 0, s_2 \mapsto 1$ and vice versa. The problem (7) can be re-written as

$$\underset{\mathbf{t}}{\text{minimize}} \quad \mathbf{t}^T \tilde{\mathbf{Q}} \mathbf{t} + \mathbf{t}^T \tilde{\mathbf{c}} + D \quad (8)$$

subject to $\mathbf{t} \in \{s_1, s_2\}^N$,

where $\tilde{\mathbf{Q}} = \frac{\mathbf{Q}}{(s_2 - s_1)^2}$, $\tilde{\mathbf{c}} = \frac{\mathbf{c}}{s_2 - s_1} - \frac{2\mathbf{Q}\mathbf{s}_1}{(s_2 - s_1)^2}$, and $D = \frac{\mathbf{s}_1^T \mathbf{Q} \mathbf{s}_1}{(s_2 - s_1)^2} - \frac{\mathbf{s}_1^T \mathbf{c}}{s_2 - s_1}$. Since D is a fixed offset, this minimization problem is equivalent to

$$\underset{\mathbf{t}}{\text{minimize}} \quad \mathbf{t}^T \tilde{\mathbf{Q}} \mathbf{t} + \mathbf{t}^T \tilde{\mathbf{c}} \quad (9)$$

subject to $\mathbf{t} \in \{s_1, s_2\}^N$.

If $|S| = 2$, that is $S = \{s_1, s_2\}$, the conversion $s_1 \mapsto 0, s_2 \mapsto 1$ reveals that this problem is equivalent to UBQP and hence NP-hard. In the remainder we assume that $|S| \geq 3$.

In order to expand $\{s_1, s_2\}$ to S , we add to the objective function in (9) a penalty term

$$M \sum_{i=1}^N (t_i - s_1)(t_i - s_2) = M(\mathbf{t} - \mathbf{s}_1)^T(\mathbf{t} - \mathbf{s}_2) \\ = M(\mathbf{t}^T \mathbf{I} \mathbf{t} - \mathbf{t}^T(\mathbf{s}_1 + \mathbf{s}_2) + \mathbf{s}_1^T \mathbf{s}_2)$$

where $M \in \mathbb{N}$. This penalty term is zero when all components of \mathbf{t} take values in $\{s_1, s_2\}$ and positive when some components take values in $S \setminus \{s_1, s_2\}$. Adding this term to (9) and dropping the fixed offset $\mathbf{s}_1^T \mathbf{s}_2$ leads to

$$\begin{aligned} \underset{\mathbf{t}}{\text{minimize}} \quad & \mathbf{t}^T(\tilde{Q} + M\mathbf{I})\mathbf{t} + \mathbf{t}^T(\tilde{\mathbf{c}} - M(\mathbf{s}_1 + \mathbf{s}_2)) \\ \text{subject to} \quad & \mathbf{t} \in S^N. \end{aligned} \tag{10}$$

We now determine a lower bound for the penalty scaling factor M which ensures that solutions \mathbf{t} of (10) contain only components s_1 or s_2 . Note that such solutions are also solutions of (9) and hence by the replacement $s_1 \mapsto 0, s_2 \mapsto 1$ solutions of UBQP.

Introducing $H(\mathbf{t}) = \mathbf{t}^T \tilde{Q} \mathbf{t} + \mathbf{t}^T \tilde{\mathbf{c}}$ and $G(\mathbf{t}) = \sum_{i=1}^N (t_i - s_1)(t_i - s_2)$, the objective function in (10) can be written as

$$F(\mathbf{t}) = H(\mathbf{t}) + MG(\mathbf{t}) - M\mathbf{s}_1^T \mathbf{s}_2.$$

If we denote any \mathbf{t} that contains some element other than s_1 and s_2 by $\mathbf{t}^{(b+)}$ and any \mathbf{t} that contains only s_1 and s_2 by $\mathbf{t}^{(b)}$, the task is to find M such that for any $\mathbf{t}^{(b+)}$,

$$F(\mathbf{t}^{(b+)}) > \max\{F(\mathbf{t}^{(b)})\}.$$

Since $G(\mathbf{t}^{(b)}) = 0$, the task is thus find M such that for any $\mathbf{t}^{(b+)}$,

$$H(\mathbf{t}^{(b+)}) + MG(\mathbf{t}^{(b+)}) - M\mathbf{s}_1^T \mathbf{s}_2 > \max\{H(\mathbf{t}^{(b)}) - M\mathbf{s}_1^T \mathbf{s}_2\},$$

or equivalently, we have to find M such that

$$H(\mathbf{t}^{(b+)}) + MG(\mathbf{t}^{(b+)}) > \max\{H(\mathbf{t}^{(b)})\} \tag{11}$$

for any $\mathbf{t}^{(b+)}$.

An upper bound for $\max\{H(\mathbf{t}^{(b)})\}$ is obtained by noting that, since \tilde{Q} is positive semi-definite, all eigenvalues of \tilde{Q} are non-negative and $0 \leq (\mathbf{t}^{(b)})^T \tilde{Q} \mathbf{t}^{(b)} \leq \lambda_{\max}(\tilde{Q}) \|\mathbf{t}^{(b)}\|^2$ where $\lambda_{\max}(\tilde{Q})$ is the largest eigenvalue of \tilde{Q} . Let s^* be the element of $\{s_1, s_2\}$ with highest absolute value and let \mathbf{s}^* be the vector of s^* 's, then for any $\mathbf{t}^{(b)}$

$$(\mathbf{t}^{(b)})^T \tilde{Q} \mathbf{t}^{(b)} \leq \lambda_{\max}(\tilde{Q}) \|\mathbf{t}^{(b)}\|^2 \leq \lambda_{\max}(\tilde{Q}) \|\mathbf{s}^*\|^2$$

and

$$(\mathbf{t}^{(b)})^T \tilde{\mathbf{c}} \leq |(\mathbf{t}^{(b)})^T| |\tilde{\mathbf{c}}| \leq |(\mathbf{s}^*)^T| |\tilde{\mathbf{c}}|,$$

where the operation $|\mathbf{z}|$ sets each element in a vector \mathbf{z} to its absolute value. Thus,

$$K := \lambda_{\max}(\tilde{Q}) \|\mathbf{s}^*\|^2 + |\tilde{\mathbf{c}}^T| \|\mathbf{s}^*\| \geq H(\mathbf{t}^{(b)}) \tag{12}$$

is an upper bound of $\max\{H(\mathbf{t}^{(b)})\}$. We have to find M such that, for any $\mathbf{t}^{(b+)}$,

$$H(\mathbf{t}^{(b+)}) + MG(\mathbf{t}^{(b+)}) > K,$$

or equivalently, such that for any $\mathbf{t}^{(b+)}$,

$$M > \frac{K - H(\mathbf{t}^{(b+)})}{G(\mathbf{t}^{(b+)})}. \tag{13}$$

(13) is well-defined since $G(\mathbf{t}^{(b+)}) > 0$. We proceed to find a lower positive bound L_G for $G(\mathbf{t}^{(b+)})$ and a lower bound L_H for $H(\mathbf{t}^{(b+)})$.

Observing that we assumed S to be ordered, an obvious positive lower bound L_G is given by taking for $\mathbf{t}^{(b+)}$ a vector made only of s_1 's and s_2 's except one element set to s_3 , resulting in $L_G = (s_3 - s_1)(s_3 - s_2)$.

For L_H , let s^{**} be the member of $S = \{s_1, \dots, s_n\}$ with highest absolute value and \mathbf{s}^{**} the vector of all s^{**} 's. A positive upper bound K' for $|H(\mathbf{t}^{(b+)})|$ is given (re-using (12)) as $K' = \lambda_{\max}(\tilde{Q}) \|\mathbf{s}^{**}\|^2 + |\tilde{\mathbf{c}}^T| \|\mathbf{s}^{**}\|$. Consequently, $L_H = -K'$ gives a lower bound of $H(\mathbf{t}^{(b+)})$.

In summary,

$$M = \frac{K - L_H}{L_G} \tag{14}$$

ensures that solutions \mathbf{t} of the optimization problem (10) contain only components from $\{s_1, s_2\}$.

It is easy to see that the transformation from (6) to (10) with M determined by (14) can be accomplished in polynomial time.

Some core ideas used in this proof were motivated by [11] and [12].

4. Discussion

We only showed that UNQP is NP-hard, but did not show that it is in NP. If true, this may turn out to be surprisingly difficult to demonstrate. Related quadratic programming problems have been proven to be in NP only long after NP-hardness was established, and those proofs are rather involved [13,14].

The binary problem UBQP has numerous applications. There exists an extensive literature on exact and approximate methods capable of producing practical solutions for this problem in a large variety of circumstances (survey in [15]). Some of the existing methods for coming to practical terms with UBQP may turn out to be adaptable to our more general problem UNQP, but this remains to be worked out when the occasion arises. With regards to reservoir computing, initial investigations indicate that very simple approximations already work quite satisfactorily.

Acknowledgments

This work was supported by European H2020 collaborative project NeuRAM3 [grant number 687299]. Fatemeh Hadaeghi would also like to thank Adam N. Letchford, who provided insight and expertise that greatly assisted this research.

References

- [1] H. Broersma, S. Stepney, G. Wendin, *Computability and complexity of unconventional computing devices*, in: *Computational Matter*, Springer, 2018, pp. 185–229.
- [2] W.-Y. Tsai, D.R. Barch, A.S. Cassidy, M.V. DeBole, A. Andreopoulos, B.L. Jackson, M.D. Flickner, J.V. Arthur, D.S. Modha, J. Sampson, et al., Always-on speech recognition using truorth, a reconfigurable, neurosynaptic processor, *IEEE Trans. Comput.* 66 (6) (2017) 996–1007.
- [3] H. Jaeger, H. Haas, *Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication*, *Science* 304 (5667) (2004) 78–80.
- [4] B. Schrauwen, M. D'Haene, D. Verstraeten, J. Van Campenhout, *Compact hardware liquid state machines on fpga for real-time speech recognition*, *Neural networks* 21 (2) (2008) 511–523.
- [5] M. Freiburger, A. Katumba, P. Bienstman, J. Dambre, *On-chip passive photonic reservoir computing with integrated optical readout*, in: *Proceedings of the 2017 IEEE International Conference on Rebooting Computing (ICRC)*, IEEE, 2017, pp. 1–4.
- [6] J.C. Coulombe, M.C. York, J. Sylvestre, *Computing with networks of nonlinear mechanical oscillators*, *PLoS One* 12 (6) (2017) e0178663.
- [7] K. Nakajima, H. Hauser, T. Li, R. Pfeifer, *Information processing via physical soft body*, *Sci. Rep.* 5 (2015) 10487.
- [8] A. Goudarzi, M.R. Lakin, D. Stefanovic, *DNA reservoir computing: a novel molecular computing approach*, in: *International Workshop on DNA-Based Computers*, Springer, 2013, pp. 76–89.
- [9] S. Moradi, N. Qiao, F. Stefanini, G. Indiveri, *A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs)*, *IEEE Trans. Biomed. Circ. Syst.* 12 (1) (2018) 106–122.
- [10] P.M. Pardalos, S. Jha, *Complexity of uniqueness and local search in quadratic 0-1 programming*, *Oper. Res. Lett.* 11 (2) (1992) 119–123.
- [11] L. Galli, A.N. Letchford, *A compact variant of the QCR method for quadratically constrained quadratic 0-1 programs*, *Optim. Lett.* 8 (4) (2014) 1213–1224.
- [12] A. Billionnet, S. Elloumi, *Using a mixed integer quadratic programming solver for the unconstrained quadratic 0-1 problem*, *Math. Program.* 109 (1) (2007) 55–68.

- [13] S.A. Vavasis, Quadratic Programming is in NP, Technical Report, Cornell University, 1990.
- [14] A. Del Pia, S.S. Dey, M. Molinaro, Mixed-integer quadratic programming is in np, *Math. Program.* 162 (1-2) (2017) 225–240.
- [15] G. Kochenberger, J.-K. Hao, F. Glover, M. Lewis, Z. Lü, H. Wang, Y. Wang, The unconstrained binary quadratic programming problem: a survey, *J. Comb. Optim.* 28 (1) (2014) 58–81.



Fatemeh Hadaeghi studied Electrical Engineering and Biomedical Engineering and obtained her Ph.D. degree in Biomedical Engineering (Computational Neuroscience) at Amirkabir University of Technology (AUT), Tehran, Iran. Currently, she is a postdoctoral fellow for computational science in MINDS (Modeling Intelligent Dynamical Systems) research group at Jacobs University (Bremen, Germany). Her research focus is on studying the neuronal mechanisms through which the brain achieves the cognitive and affective functions such as learning, decision-making, and emotional and cognitive information processing. She also works at the interface of neuroscience and technology where computational models are developed to investigate the dynamical aspects of information processing in the brain and to assist engineers to design brain-inspired computing architectures.



Herbert Jaeger studied mathematics and psychology at the University of Freiburg and obtained his Ph.D. in Computer Science (Artificial Intelligence) at the University of Bielefeld. After a 5-year postdoctoral fellowship at the German National Research Center for Computer Science (Sankt Augustin, Germany) he headed the “Intelligent Dynamical Systems” group at the Fraunhofer Institute for Autonomous Intelligent Systems AIS (Sankt Augustin, Germany). Since 2003 he is Associate Professor for Computational Science at Jacobs University Bremen. His research targets the computational modelling of hierarchical, non-linear, neural (or just plain computational) learning systems with applications in pattern recognition, engineering and cognitive/robot modelling, and non-digital devices.