ISBN: 978-3-95450-231-8 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2023-THPL037

GRADIENT DESCENT OPTIMIZATION AND RESONANCE CONTROL OF SUPERCONDUCTING RF CAVITIES

M. Caskey, J. Diaz Cruz, S. Biedron, University of New Mexico, Albuquerque, NM, USA
A. Hagberg, Los Alamos National Lab, NM, USA

Abstract

Additive Manufacturing (AM) offers different benefits such as efficient material usage, reduced production time and design freedom. Moreover, with continuous technological developments, AM expands in versatility and different material usage capabilities. Recently new energy sources have been developed for AM – green wavelength lasers, which provide better energy absorption for pure copper. Due to high thermal and electrical conductivity of copper, this novel AM technology is highly promising for various industries, particularly, there is a huge interest to use it for accelerator applications. In particular, these AM produced accelerator components should reach the associated Ultra High Vacuum (UHV) requirements. In this study, vacuum membranes of pure copper were produced by AM using a green laser source, in different thicknesses and built angles. Furthermore, a vacuum membrane helium leak tightness test was performed at room temperature by using a high-sensitivity mass spectrometer. Comparison of these test results was performed with previously established results. Through this study, novel knowledge and initial results are provided for green laser source AM technology usage for applications for UHV accelerator components.

INTRODUCTION

When accelerating an electron beam through an SRF cavity, the cavity's high-quality factor and narrow bandwidth causes the accelerating electron beam to be susceptible to internal and external vibrations. As a result, more power is required to maintain the desired beam, thus interfering with the beam quality. Tuning the SRF cavities, using stepper motors and piezoelectric actuators, helps to keep the cavity between the power budget [1].

Microphonics are significant disturbances that have been effectively combated through active noise control methods. One such technique is the NANC algorithm, a method of gradient descent that operates the piezoelectric tuners and neutralizes narrowband microphonics. The NANC algorithm has proven useful for operating facilities such as LCLS-II, which must maintain a 10 Hz maximum cavity detuning. However, the NANC algorithm that has been used and tested appears to be open to improvement, as it requires the user to manually set parameter values [1].

The method of gradient descent is most commonly used in machine learning and data mining [2], and researchers in these fields have worked to improve the algorithm beyond the manner in which it is used in the NANC algorithm. The purpose of this paper will be to propose the application of two previously conceived methods, Adam and the Nesterov Accelerated Gradient method, so as to improve the NANC algorithm.

GRADIENT DESCENT

Gradient Descent is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function, and is often used in machine learning and data mining. As illustrated by Figure 1, the algorithm steps through an equation opposite the direction and according to the magnitude of the approximate gradient until the minimum is reached. The simple gradient descent algorithm currently utilized by the NANC algorithm takes on the form [2]:

$$a_{n+1} = a_n - \gamma \nabla F(a_n) \quad (1)$$

This simple algorithm functions to trace the graph of $\boldsymbol{F}$ to its minimum according to its derivative with respect to a and a stepsize γ. The stepsize is manually optimized, through trial and error, so as to most effectively approach the minimum. This equation can only be used for differentiable convex equations with a single minimum. If this is not the case, the algorithm may not converge to the absolute minimum [2].

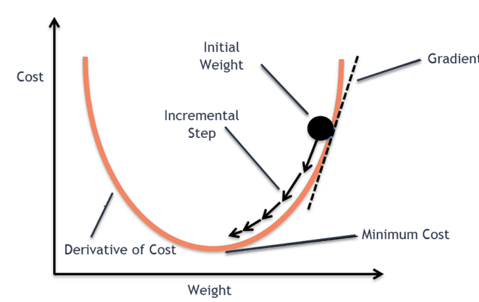

Figure 1: Illustration of gradient descent on a convex function [3]. The incremental step is representative of γ and the weight is representative of a in Equation 1.

The current NANC algorithm has worked fairly successfully for implementations like the one at Cornell [4], but it is not yet known if the algorithm works optimally as the stepsize has only been tuned manually.

The NANC Algorithm

$$C(t_n) = 1/N \cdot \sum_{i=n-N+1}^{n} [\delta f_{comp}(t_i)]^2 \quad (2)$$

The aim of the NANC algorithm is to minimize the above equation, the mean-square detuning of an SRF cavity. In this case, $C(t_n)$ represents the cost function at time t_n and δf_{comp} represents the effective detuning of the

ISBN: 978-3-95450-231-8 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2023-THPL037

cavity in response to the summed effects of the perturbation and the piezo-tuner. As δf is dependent upon the vibrations within the cavity, the derivative of $C(t_n)$ cannot be easily evaluated, and the equation is thus best minimized through a gradient descent algorithm of the form:

$$\tilde{A}_m(t_{n+1}) = \tilde{A}_m(t_n) - \mu_m \delta f_{comp}(t_n) e^{i(w_m t - \phi_m{}^{mod})} \quad (3)$$

Where $\tilde{A}_m$ is a complex phasor notation representative of the vibrations within the cavity, μ_m is the stepsize of the gradient descent algorithm, ${\phi_m}^{mod}$ is the phase response of the actuator, and ω is the detuning frequency of the SRF Cavity. [4]

Though it appears more complicated than the general gradient descent method, the above algorithm can also be written as:

$$\tilde{A}_m(t_{n+1}) = \tilde{A}_m(t_n) - \mu_m \boldsymbol{\nabla C(t_n)} \quad (4)$$

Where $\boldsymbol{\nabla C(t_n)}$ is the derivative of $C(t_n)$ according to $\tilde{A}_m$. This algorithm thus appears identical to the general gradient descent algorithm.

NANC OPTIMIZATION

A general gradient algorithm uses a constant stepsize which must be manually optimized through trial and error. More complicated algorithms, on the other hand, utilize variable step-sizes and stepping points for added efficiency. These algorithms tend more quickly to the minimum and also tend to be less dependent upon the initial step-size. It is the hope of this research that one or both of the algorithms explained below will improve the NANC algorithm to the point of maximum efficiency.

Nesterov Accelerated Gradient Method

Compared to the general gradient descent equation, the Nesterov accelerated gradient method alters the general method in the following way [5]:

$$\boldsymbol{a_{n+1}} = \boldsymbol{a_n} + \boldsymbol{\lambda} \cdot (\boldsymbol{a_n} - \boldsymbol{a_{n-1}}) - \gamma \boldsymbol{\nabla F(a_n)}. \quad (4)$$

This algorithm generally accelerates the process of gradient descent by qualitatively shifting the current stepping point a_n towards the minimum according to previous values [5]. In this equation, λ is a hyperparameter which is to be optimized along with γ. Even so, some have suggested that λ may be changed over time as an effect of the current n [6], such as like

$$\lambda = (n-1)/(n-2). \quad (5)$$

The Nesterov method is not as effective on equations that are not strongly convex [7]. It remains to be seen whether this is a concern with the NANC algorithm.

Adam

Different from the general gradient descent equation, the Adam method continuously modifies the stepsize in the following way:

$$m_{n+1} = (\beta_1 \cdot m_n + (1-\beta_1) \cdot \boldsymbol{\nabla F}) \,/\, (\boldsymbol{1} - \boldsymbol{\beta_1}^{\boldsymbol{n}}) \quad (6)$$

$$v_{n+1} = (\beta_2 \cdot v_n + (1-\beta_2) \cdot (\boldsymbol{\nabla F})^{\boldsymbol{2}}) \,/\, (\boldsymbol{1} - \boldsymbol{\beta_2}^{\boldsymbol{n}}) \quad (7)$$

$$\gamma_{n+1} = \alpha \cdot m_{n+1} \,/\, (\sqrt{v_{n+1}} + \epsilon). \quad (8)$$

The Adam method is useful because it more accurately follows the shape of an equation and requires less optimization than the general gradient descent algorithm. Thus, it tends to more quickly and effectively approach the minimum than general gradient descent. However, this method has additional parameters to tune: α, $\beta 1$, $\beta 2$, and ε, which would require a certain amount of trial and error to optimize [8].

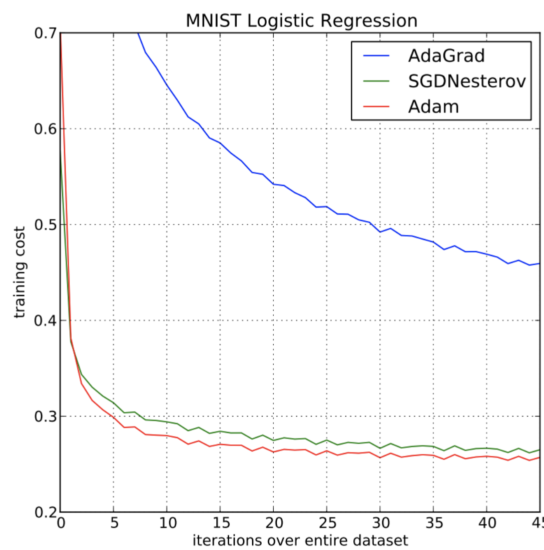

Figure 2: Logistic regression training negative log likelihood on MNIST images and IMDB movie reviews with 10,000 bag-of-words (BoW) feature vectors, using three separate gradient descent techniques: AdaGrad, Nesterov Accelerated Gradient Descent, and Adam [8].

As shown in Figure 2, the Nesterov and Adam methods may perform similarly well depending on the given situation, and generally outperform other algorithms. They may even be made to perform better when combined together. It is the purpose of future work to test this for the case of the NANC algorithm.

SUMMARY AND FUTURE WORK

Currently, the general gradient descent algorithm as used in the NANC algorithm works well as demonstrated in [4], but the algorithm could be improved. Experimentation must be done in order to determine whether the proposed

ISBN: 978-3-95450-231-8 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2023-THPL037

optimization algorithms, the Nesterov accelerated gradient descent algorithm and Adam, will further enhance the NANC implementation. It is the ultimate goal that these experiments will answer: whether the current stepsize truly requires further optimization, whether the cost equation encounters too much noise for the Nesterov gradient descent algorithm to work most efficiently, or whether it is possible for the NANC algorithm to more quickly and efficiently approach the minimum of the cost function.

ACKNOWLEDGEMENTS

This work was supported by the U.S. National Science Foundation under Award PHY-1549132, the Center for Bright Beams

REFERENCES

[1] J. Diaz Cruz *et al*, "Machine Learning Based LLRF and Resonance Control of Superconducting Cavities," in *Proc. IPAC'21,* Campinas, Brazil, May 2021, pp. 920-923: `doi: 10.18429/JACoW-IPAC2021-MOPAB290`

[2] J. Lu, "Gradient Descent, Stochastic Optimization and Other Tales," *arXiv,* 2022. `doi: 10.48550/ARXIV.2205.00832`

[3] R. Menon, "The Ascent of Gradient Descent," *Clairvoyant Blog*, 2019, `https://blog.clairvoyantsoft.com/the-ascent-of-gradient-descent-23356390836f`

[4] N. Banerjee *et al*., "Active suppression of microphonics detuning in high QL cavities", *Phys. Rev. Accel. Beams*, vol. 22, no. 5, p. 052002, May, 2019. `doi:10.1103/PhysRevAccelBeams.22.052002`

[5] J. Melville, "Nesterov Accelerated Gradient and Momentum," Dec. 2016. `https://jlmelville.github.io/mize/nesterov.html#NAG`

[6] G. Gordon and R. Tibshirani, "Accelerated First-Order Methods," presented at Carnegie Mellon University.

[7] H. Li, C. Fang, and Z. Lin, "Accelerated First-Order Optimization Algorithms for Machine Learning", *Proceedings of the IEEE,* vol. 108, no. 11, pp. 2067-2082, Nov. 2020. `doi: 10.1109/jproc.2020.3007634`

[8] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv,* 2014. `doi: 10.48550/ARXIV.1412.6980`