Optimization of Machine Learning Method to Improve the Management Efficiency of Heterogeneous Telecommunication Network

Viktoriia Zhebka[1], Mykola Gertsiuk[1,2], Volodymyr Sokolov[3], Vladyslav Malinov[1], and Mylana Sablina[3]

[1] *State University of Telecommunications, 7 Solomenskaya str., Kyiv, 03110, Ukraine*
[2] *The Institute of Environmental Geochemistry of National Academy of Sciences of Ukraine, 34a Academician Palladina ave., Kyiv, 03142, Ukraine*
[3] *Borys Grinchenko Kyiv University, 18/2 Bulvarno-Kudriavska str., Kyiv, 04053, Ukraine*

Abstract

This paper presents some optimization method aspects use in telecommunications networks. The use of optimization methods by machine learning means is especially important to avoid various emergencies in networks. It is advisable to use machine learning methods to obtain information about signal quality, traffic, etc. At the same time, it is possible to make various malfunctions forecasts, routing, safety control. It is determined that the Markov random field model is effective in modeling in homogeneous networks. This approach allows an exponential distribution nodes modeling in heterogeneous networks. A proximal gradient algorithm modification is presented—a method of variable metric proximal gradient. Ensuring fast convergence is achieved by diagonal step size means, which is more efficient than scalar. The article reveals an adaptive metric selection rule, i.e., a diagonal step based on the Barzilai-Borwein (BB) method. The presented algorithm combines two approaches: the standard proximal gradient method and the proximal Newton method. The establishment of clear rules for choosing the diagonal step size for convex optimization algorithms has been implemented.

Keywords

Convex optimization, optimization methods, machine learning, diagonal step size, Barzilai-Borvain method, proximal gradient.

1. Introduction

Optimization methods are traditionally used in informational technologies, more specifically in telecommunication networks, that is confirmed by positive results in a wide range of different data [1, 2]. An optimization approach use requires previous step—reality modeling and simplifying. Such approach requires a lot of work and in most cases can lead to making inefficient decisions. Thus, the use of optimization methods with machine learning is especially relevant. It will provide a possibility to predict various extraordinary cases in networks, based on analysis and learning with big array data [3].

It is relevant to use machine learning methods when it is necessary to obtain conclusions from various monitoring results, namely, traffic quality or signal, etc. Moreover, using certain However, there is a problem when heterogeneous data are used. For instance, heterogeneous network components failure is analyzed based on a set of various network parameters and factors affecting it. To model such heterogeneous networks, the Markov random field model is used [4, 5].

Pairwise exponential Markov random field belongs to multivariate exponential distributions class. The specified method based on joint distribution representation, that allows compactly introduce heterogeneous variables,

CPITS-2022: Cybersecurity Providing in Information and Telecommunication Systems, October 13, 2022, Kyiv, Ukraine
EMAIL: viktoria_zhebka@ukr.net (V. Zhebka); gertsiuk@gmail.com (M. Gertsiuk); v.sokolov@kubg.edu.ua (V. Sokolov); vladyslav1995@gmail.com (V. Malinov); m.sablina@kubg.edu.ua (M. Sablina)
ORCID: 0000-0003-4051-1190 (V. Zhebka); 0000-0003-2946-9673 (M. Gertsiuk); 0000-0002-9349-7946 (V. Sokolov); 0000-0002-0112-4975 (V. Malinov); 0000-0001-9452-1867 (M. Sablina)

which in turn will lead to a faster studying the structure method for nodes distribution with unknown parameters. In other words, this approach allows to simulate an exponential nodes distribution in heterogeneous networks. A modification of the proximal gradient algorithm – the variable metric proximal gradient method also deserves attention. Ensuring fast convergence is achieved by means of a diagonal step size, which is more efficient than a scalar one [6–8].

This article represents adaptive metric selection rule, means diagonal step, that based on the Barzilai-Borwein method (BB). Current algorithm combines two approaches: standard proximately gradient method and the proximal Newton method [9].

2. Formulation of the Problem

Increasing efficiency and functional system stability for heterogeneous network management provides models and machine learning methods improvement. Using machine learning methods gives a possibility to avoid destabilizing factors in networks. To simulate such networks graphical methods are used, the assessment of which is quite difficult. Thus, it is relevant to improve metric proximately gradient, scientific novelty of which is that it uses a pair-exponential Markov random field model and diagonal step selection method, which allows to provide a faster convergence of the machine learning algorithm.

3. Goal and Research Tasks

The purpose of the study is to establish clear diagonal step size rules for convex optimization algorithms and using the proposed approach to improve the machine learning method. To realize a goal, it is proposed to use an adaptive metrics selection rule, means diagonal step, that is called a Barzilai-Borwein step. Current algorithm combines two approaches: standard proximate gradient method and proximate Newton method.

4. Last Research and Publications Analysis

A design, realization, and network management are appropriate to conduct using methods of optimization and machine learning. The use of optimization methods in telecommunication network provides effective results. A relevant optimization methods support using by means of neural network, allow to analyze, and study of big data arrays and predict possible extraordinary cases in network [10, 11].

A telecommunication field is already prepared for machine learning implementation. Network operators work with big amount of data: information about clients, web performance data, Internet traffic data and social network data, etc. At the same time, operators use various applications for network planning and analysis to find patterns in the data. It causes the appearance of many machines learning programs in the telecommunications [12].

Nowadays, a machine learning can be interpreted as a withdrawal from patterns in future networks and systems design. The use of machine learning methods applies in network application for faults prediction, intrusion detection, safety control, routing, bandwidth reconfiguration considering traffic and more [13, 14].

5. Research Results

Every year, the number of devices connected to the network is constantly growing. However, this is due not only to the increase a number of smartphones, tablets, etc., but also to the emergence of new and the development of existing technologies, such as Machine-to-Machine (M2M) and Internet of Things (IoT), where all electronic devices are able to connect to a heterogeneous network. An equally important regularity is that a video traffic percentage in network will grow, which require an improvement in the quality and clarity of the image. A new technology, deployed at different levels and network parts will need interaction to meet user requirements. Moreover, user requirements will be very various: from low-latency and high data transfer rate video applications to IoT devices that has very low productivity requirements. With these conditions, networks just need to add intelligence and autonomy to adapt to this enormous heterogeneity. Networks with such characteristics are called Self-Organizing Networks –SON.

A telecommunications field is already prepared for machine learning implementation, because network operators are already having a big amount of data: the acquire and store data about clients, Internet performance data, internet traffic and socials networks data, etc. In addition, operators are already use such applications, as networks planning and root cause analysis to find patterns in the data. Therefore, it is not surprising that many machine learning programs are already starting to appear in the telecommunications sector [15].

One way to use machine learning in a network is to analyze and examine network performance data to identify idle cells and trigger automatic restarts. It can seriously affect a service quality, especially in crowded areas or at loaded times of the day. A manual resetting often results in the cellular tower being turned off for several hours. To avoid such inconvenience, it is appropriate to use machine learning methods.

Nowadays, machine learning perceived as paradigm shift to future networks and systems design. These methods must allow to make conclusions from data, acquired using different types of monitoring (e.g., signal quality, traffic examples, etc.) It is relevant to use machine learning is network applications for faults predictions, invasion detection, safety control, routing, bandwidth reconfiguration considering traffic and more.

To solve in network work issues, it is proposed a use of model in a pair-exponential Markov random field form [16].

An algorithm, based on multipliers methods of variable direction and the solution of closed form restoration for each of the sub-problems are proposed. Due to the use of these fast recoveries that the solution process is accelerated, and the method of variable direction multipliers becomes more scalable than other methods. The obtained methodology was tested both on artificially created and real data.

A convex optimization as a mathematical theory has been studied for a long time and has found its application in control synthesis, signal processing, working with big data and machine learning.

For the convex optimization study, the following model was chosen as a basis:

$$\min_{x \in R^n} \quad F(x) = f(x) + q(x) \qquad (1)$$

where $x \in R^n$ is solution variable, function $f : R^n \to R$ convexed and differential, function $q : R^n \to R \cup \{\infty\}$ is convexed and can be not differentiated. Function q may be used can be used to encode constraints on a variable x.

A proposed structured model of (1) convexed optimization using for a lot of machine learning tasks: regression, classification, matrices build, etc. To solve such optimization tasks, presented in the form, the proximal gradient algorithms are very often used, which allows to improve measurement process, provide practical rules for step selection, theoretical guarantees under moderate conditions, increasing the accuracy of the result, etc.

Most proximal gradient algorithm modifications subject to the same form, that is called a metric proximate gradient method.

When selection an algorithm step, that provides a fast convergence, it is established that the diagonal step size is more effective than the scalar one. Clear rules for diagonal step size for convex optimization algorithms selection should be established.

This article proposes to use an adaptive metrics selection rule, means diagonal step, that is called a Barzilai-Borwein step. Proposed algorithm combines two approaches: a standard proximate gradient method, and a proximal Newton method. Metrical proximal gradient with diagonal step BB provides low measurement steps loses, a much better Hessian approximation at each iteration, and consequently—a fast convergence of the algorithm (in comparing with proximate gradient method with scalar step). Conducted empirical research determined, that proposed method with diagonal metrics provides improved convergence in comparing with proximal algorithm method with scalar step.

The most popular diagonal step determining should be attributed the following: spectral scalar step size, variable non-scalar metric, diagonal metric.

Usually, a spectral step method is used for gradient BB type methods. The additive rule for spectral metric selection uses a spectral step method. To reduce the specified limitations, a new adaptive diagonal metric selection strategy with convergence guarantees using string search is proposed [17, 18].

A proximate gradient step can be considered as the minimization of the function *F*, where differentiated part *f* is approximated to its second-order form for x^n, relatively $M^n \in C^h_{++}$ [19].

$$prox_{q,M^n}\left(x^n - \left(M^n\right)^{-1}\nabla f\left(x^n\right)\right) =$$

$$\arg\min_x q(x) + f\left(x^n\right) + \nabla f\left(x^n\right)^T\left(x - x^n\right) + \frac{1}{2}\left\|x - x^n\right\|^2_{M^n}$$

This approximation shows that $M^n = \nabla^2 f\left(x^n\right)$ is optimal choice after proximal Newton method. However, the Hessian use usually leads to a high cost of iterations. An alternative option and Hessian approximation with use of secant state:

$$M^n c^n \approx y^n, \tag{2}$$

for step $c^n = x^n - x^{n-1}$ and gradient change $y^n = \nabla f\left(x^n\right) - \nabla f\left(x^{n-1}\right)$.

A Barzilai-Borwein method is an approach, that estimates a scalar Hessian approximation, by setting $M^n = \left(\alpha^n\right)^{-1}$, that satisfies formula (2). The method is quite popular.

The most common steps in the BB method are the following:

$$\beta^n_{BB1} = c^{n2}_{\ 2} / \left(c^n y^n\right); \quad \beta^n_{BB2} = \left(c^n, y^n\right) / y^{n2}_{\ 2} \tag{3}$$

where $\beta^n_{ББ2} \le \beta^n_{ББ1}$ is always performed.

For this purpose, several modifications and protective measures have been adopted to the initial stage of BB. One of such numerical measures for c^n and y^n a hybrid choice between these two steps is suggested:

$$\beta^n_{BB} = \beta_{BB}\left(c^n, y^n\right) = \begin{cases} \beta^n_{BB2}, if\, \beta^n_{BB1} < \partial\beta^n_{BB2} \\ \beta^n_{BB1} - \frac{1}{\delta}\beta^n_{BB2}, in\ other\ cases \end{cases} \tag{4}$$

where hyperparameter $\partial \in R$ usually equals to 2. If β^n_{BB} in equation (4) is negative, then previous step is chosen $\beta^n_{BB} = \beta^{n-1}_{BB}$.

Such modification and activities designed to eliminate instability in the primary BB with step (4) for the poorly conditioned *f*. However, even with such modification, a scalar BB still addicted to inconsistences. It should be notes, that $(\beta^n_{BB1})^{-1}J$ and $(\beta^n_{BB2})^{-1}J$ can be considered, and Hessian approximation in Euclidean space. However, in poorly conditioned conditions these scalar approximations may be far from the true non-Euclidean geometry of Hesse. In other case, e.g., after such approximations, as step direction projection c^n and gradient change y^n they may be close to orthogonality. This leads to degenerate scenarios with $\beta_{BB1} \to \infty$ or $\beta_{BB2} \to 0$. For such cases, a scalar estimation may be significantly different from the secant condition (2) and Hesse's geometry.

The proposed diagonal step of the BB is considered in the paper.

To display Hessian geometry *f* we enter the diagonal metric M^n, that at each iteration *n* is calculated as follows:

$$\min_{m\in R^n} \left\|Mc^n - y^n\right\|^2_2 + \eta\left\|M - M^{n-1}\right\|^2_F \tag{5}$$

$$\left(\beta^n_{BB1}\right)^{-1} J \le M \le \left(\beta^n_{BB2}\right)^{-1} J$$

$$M = diag\left(m\right)$$

where hyperparameter $\eta > 0$ manages the compromise between the satisfaction of the state (2) and consistency with previous metrics M^{n-1}. If Hessian changes so fast, it is necessary to select a low enough value μ. In this case this parameter numerical protection will be used. If Hessian won't change much during iterations, there is a necessity to choose big value μ. Eventually, diagonal elements are restricted, means guaranteed with BB step (3).

One of proposed form (5) peculiarity is, that it has a simple closed solution form. For $M_n = diag\left(m_n\right)$, where $m^n = \left[m^n_1, ..., m^n_h\right] \in R^n$ task (5) solving is given as follows:

$$m^n_i = \begin{cases} \frac{1}{\beta^n_{ББ1}}, & \frac{c^n_i y^n_i + \eta m^{n-1}_i}{\left(c^n_i\right)^2 + \eta} < \frac{1}{\beta^n_{ББ1}} \\ \frac{1}{\beta^n_{ББ2}}, & \frac{c^n_i y^n_i + \eta m^{n-1}_i}{\left(c^k_i\right)^2 + \eta} > \frac{1}{\beta^n_{ББ2}} \\ \frac{c^n_i y^n_i + \eta m^{n-1}_i}{\left(c^n_i\right)^2 + \eta}, & in\ other\ cases. \end{cases} \tag{6}$$

where c^n_i and y^n_i are *i* element c^n and y^n.

A metric proximation gradient algorithm with diagonal metrics has next steps (Fig. 1).

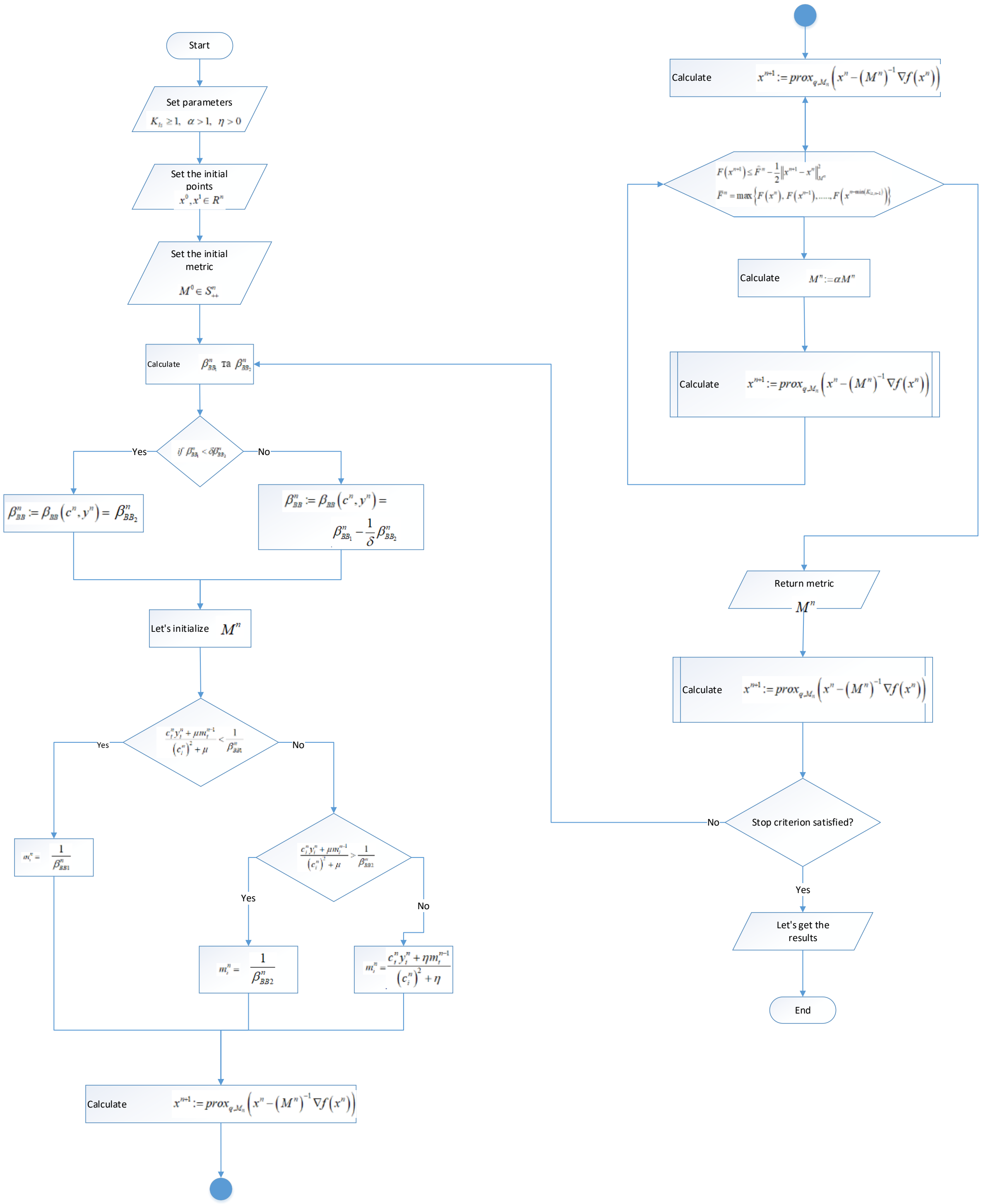

Figure 1: A metric proximation gradient algorithm

A metrics proximation gradient with diagonal step BB significantly goes beyond standard proximal gradient ($\sim 20\%$ of measurements) as poorly conditioned.

6. Conclusions

Proposed diagonal metric provides better estimate for poorly conditioned local Hessian in comparing with standard scalar Barzilai-Borwein step BB, which leads to faster algorithm convergence. A metric proximal gradient method is developed, scientific novelty of which is in, that it use a pair-exponential Markov random field model and selection diagonal step method, that allows to provide faster machine learning algorithm convergency. A proposed block implementation in heterogeneous network management system allows early react to overloaded network with help of short- and medium-term forecasts build and strengthen an intelligent network management block.

A proposed metric proximate gradient method implementation in heterogenous telecommunication network will provide efficient decentralized management of heterogenous network management and reduce amount of service information in the network. It will allow to avoid network overload when extraordinary cases are appeared. However, the question of network overload on the equipment during its management, when many users present, remains open.

7. References

[1] V. Astapenya, et al., Last Mile Technique for Wireless Delivery System using an Accelerating Lens, in IEEE International Conference on Problems of Infocommunications. Science and Technology, 2020. doi: 10.1109/ picst51311.2020.9467886.

[2] V. Astapenya, et al., Analysis of Ways and Methods of Increasing the Availability of Information in Distributed Information Systems, in IEEE 8th International Conference on Problems of Infocommunications, Science and Technology, 2021. doi: 10.1109/ picst54195.2021.9772161.

[3] F. Kipchuk, et al. Investigation of Availability of Wireless Access Points based on Embedded Systems, in IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), 2019. doi: 10.1109/picst47496.2019.9061551.

[4] W. Zucchini, I. L. MacDonald, R. Langrock, Hidden Markov Models for Time Series: An Introduction Using R. Chapman and Hall, 2016.

[5] P. Anakhov, et al., Systematization of Measures on Lightning Protection of the Objects of Telecommunications Network, International Journal of Avanced Trends in Computer Science and Engineering, vol. 9, no. 5, 2020, pp. 7870–7877.

[6] A. Canale, N. Lunardon, Churn Prediction in Telecommunications Industry. A Study based on Bagging Classifiers Telecom, Carlo Alberto Notebooks, vol. 350, 2014, pp. 1–11.

[7] DSTU (State standard of Ukraine) 3899:2013. Dyzain i erhonomika. Terminy ta vyznachennia poniat [Design and ergonomics. Terms and definitions].

[8] J. Hughes, M. Haran, P. C. Caragea, Autologistic Models for Binary Data on a Lattice, Environmetrics, vol. 22, no. 7, 2011, pp. 857–871.

[9] ITU-T Recommendation G.602. Transmission Media Characteristics. Reliability and Availability of Analogue Cable Transmission Systems and Associated Equipments.

[10] A. A. Khan, J. Sanjay, M. M. Sepehri, Applying Data Mining to Customer Churn Prediction in an Internet Service Provider, Int. J. Comput. Appl., vol. 9, no. 7, 2010, pp. 8–14.

[11] O. Klymovych, et al., The Diagnostics Methods for Modern Communication Tools in the Armed Forces of Ukraine Based on Neural Network Approach, in Modern Machine Learning Technologies Workshop, pp. 198–208.

[12] M. J. Wainwright, M. I. Jordan, Graphical Models, Exponential Families, and Variational Inference, Found. and Tr. in Mach. Learn., vol. 1, no. 1–2, 2008, pp. 1–305.

[13] P. Babarczi, et al., A Mathematical Framework for Measuring Network Flexibility, Computer Communications, vol. 164, 2020, pp. 13–24.

[14] V. Mukhin, et al., Models for Analysis and Prognostication of the Indicators of the Distributed Computer Systems' Characteristics, International Review on Computers and Software (IRECOS), vol. 10, no. 12, 2015, pp. 1216–1224.

[15] W. Kellerer, et al., How to Measure Network Flexibility? A Proposal for Evaluating Softwarized Networks, IEEE Communications Magazine, vol. 56, no. 10, 2018, pp. 186–192.

[16] Z. Hu, et al., Distributed Computer System Resources Control Mechanism based on Network-Centric Approach, International Journal of Intelligent Systems and Applications, vol. 9, no. 7, 2017, pp. 41–51.

[17] V. Zhebka, E. Negodenko, A. Aronov, Algoritm Maksimalno Effektivnogo Ispolzovaniya Pamyati dlya Poparnogo

Markovskogo Sluchaynogo Polya, Actual Problems of Economic, vol. 1, no. 223, 2020, pp. 180–191.
[18] V. Zhebka, et al., Optimizatsiya Rabotyi Algoritma Gradientnogo Bustinga s Pomoschyu Perekrestnoy Proverki, Actual Problems of Economic, vol. 12, no. 222, 2019, pp. 189–197.
[19] V. Zhebka, Modeliuvannia Markivskoho Vypadkovoho Polia z Metoiu Yoho Podalshoi Optymizatsii ta Zastosuvannia, Zviazok, vol. 5, 2020, pp. 35–40.