

RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ
zaprasza na
OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr Anthony Nwachukwu

która odbędzie się w dniu **19.05.2026 roku**, o godzinie **13.00** w trybie zdalnym.

**Temat rozprawy: Augmented Lagrangian-Based Algorithms for Separable Non-convex Optimization
with Applications in Network Routing and Machine Learning”**

Promotor: dr hab. inż. Andrzej Karbowski – Politechnika Warszawska

Recenzenci: prof. Jacek Gondzio – Uniwersytet w Edynburgu
prof. dr hab. inż. Józef Korbicz – Uniwersytet Zielonogórski
dr hab. inż. Jerzy Respondek, prof. uczelni – Politechnika Śląska

Obrona odbędzie się w sali nr 136 w Gmachu Wydziału Elektroniki i Technik Informatycznych Politechniki Warszawskiej.

Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Warszawa, Plac Politechniki 1, oraz na stronie internetowej BIP PW: <https://www.bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego/Doktoraty/Wszczete-po-30-kwietnia-2019-r/Rada-Naukowa-Dyscypliny-Informatyka-Techniczna-i-Telekomunikacja/mgr-Anthony-Nwachukwu>

Przewodniczący Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej
Prof. dr hab. inż. Jarosław Arabas

Streszczenie

W niniejszej rozprawie doktorskiej opracowano i przeanalizowano algorytmy rozszerzonego lagranżianu (inaczej algorytmy mnożników Lagrange'a) dla dużych, separowalnych i niewypukłych zadań optymalizacji, z zastosowaniem do uogólnionego routingu sieciowego oraz uczenia maszynowego. Takie zadania obejmują globalny cel podlegający ograniczeniom wiążącym wiele podzadań (podsystemów) i pojawiają się na przykład w energooszczędnych sieciach telekomunikacyjnych, odzyskiwaniu rzadkich sygnałów i obliczeniach rozproszonych związanych z klastrami.

Najpierw przeprowadzono kompleksowe badania empiryczne istniejących metod mnożników Lagrange'a wykorzystujących dekompozycję (klasyczna dekompozycja dualna, ADMM, metody Tatjewskiego, Bertsekasa i SALA) na wybranych zadaniach niewypukłych. Przeprowadzone eksperymenty wskazują, że metoda Bertsekasa rozszerzenia lagranżianu o składnik proksymalny daje najlepszą jakość rozwiązania, skalowalność i szybkość zbieżności. Opierając się na tym odkryciu, sednem rozprawy stało się opracowanie nowego asynchronicznego wariantu algorytmu dekompozycji Bertsekasa.

Ten asynchroniczny algorytm pozwala węzłom obliczeniowym aktualizować zmienne lokalne bez globalnej synchronizacji, wykorzystując potencjalnie nieaktualne informacje. Reguły aktualizacji zostały sformalizowane w ramach modelu obliczeniowego z ograniczonym opóźnieniem, dla którego wyprowadzono rygorystyczną analizę zbieżności. Pod standardowymi warunkami regularności (gładkość, ciągłość gradientów Lipschitza i kwalifikacje ograniczeń) i przy ograniczonych opóźnieniach, iteracje są zbieżne do punktu Karusha-Kuhna-Tuckera oryginalnego zadania niewypukłego. Metoda ta została również rozszerzona w celu obsługi w ramach rozszerzonego lagranżianu ogólnych ograniczeń, w tym nierówności i mieszanych ograniczeń ze zmiennymi całkowitoliczbowymi.

Algorytmy zostały zweryfikowane poprzez obszerne eksperymenty numeryczne na syntetycznych benchmarkach i rzeczywistych zbiorach danych. Przykłady zastosowań obejmują jednoczesny routing i alokację przepustowości w sieciach energooszczędnych, rozwiązywanie dużych, rzadkich układów równań liniowych i rozproszone grupowanie metodą K-środków. W każdym przypadku asynchroniczna metoda Bertsekasa osiąga wartości docelowe porównywalne z implementacją synchroniczną, jednocześnie znacznie skracając czas obliczeń wraz ze wzrostem stopnia równoleglenia.

Niniejsza rozprawa doktorska identyfikuje najbardziej efektywny schemat dekompozycji rozszerzonego lagranżianu dla separowalnej optymalizacji niewypukłej i rozszerza go do asynchronicznej struktury równoległej z możliwymi do udowodnienia gwarancjami zbieżności. Praca ta rozwija teorię i praktykę rozproszonej optymalizacji na dużą skalę, pokazując jak asynchroniczne metody mnożników Lagrange'a mogą wykorzystać obliczenia równoległe do

bardziej efektywnego rozwiązywania złożonych zadań sieciowych i uczenia maszynowego.

Słowa kluczowe: Zbieżność asynchroniczna, metoda rozszerzonego lagranżianu Bertsekasa, optymalizacja niewypukła, obliczenia rozproszone, wydajność obliczeniowa, skalowalność, ulepszenia algorytmiczne, uczenie maszynowe, optymalizacja wielowymiarowa, rozszerzony lagranżian, optymalizacja, uczenie maszynowe, metoda mnożników o naprzemiennym kierunku, obliczenia równoległe, optymalizacja wypukła i niewypukła, ADMM, obliczenia rozproszone, klasteryzacja, maszyna wektorów nośnych, lasso, regresja, optymalizacja niewypukła, metoda mnożników, zadania separowalne, optymalizacja sieci, MINLP, zadania NP-trudne, relaksacja Lagrange'a

REVIEW OF A DOCTORAL DISSERTATION

*Augmented Lagrangian-Based Algorithms for Separable
Non-convex Optimization with Applications in Network Routing and Machine Learning*

by Anthony Chukwuemeka Nwachukwu, M.Sc.

prepared for the

Scientific Discipline Council for Technical Informatics and Telecommunications
Warsaw University of Technology

1. General overview

The dissertation addresses a research problem of high contemporary relevance, situated at the intersection of large-scale optimization, distributed computing, and applications in network routing and machine learning. The growing scale and heterogeneity of modern computational systems makes classical centralized optimization approaches increasingly impractical. In this context, the focus on distributed and, in particular, asynchronous optimization methods responds directly to the real-world challenges such as communication delays, processor heterogeneity, and scalability constraints. By targeting separable non-convex optimization problems, the dissertation tackles a class of problems that are theoretically challenging as well as practically important. The dissertation has both a scientific and applied character, proposing a solution to important and difficult optimization and IT issues.

2. The concept of the dissertation and its implementation

The dissertation, with a total of 149 pages, is written in English with a two-page summaries in Polish and English. It consists of nine chapters and a short, four-page introduction, which

initially formulates the goal of the work, describes the research problem, the key contributions and the hypothesis of the dissertation. In the last part of dissertation, the author discusses the contributions summary, the limitations of the proposed solutions as well the possible directions for further work. The attached bibliography consists of 96 items, primarily specialist journals and international conference proceedings. It is a well-curated list of relevant works related to the subject of the dissertation. The proposed bibliography is complete and confirms indirectly the fact that the doctoral student has a very good knowledge in the field of technical information technology and telecommunications.

Chapter 2 provides an introduction to the theoretical foundations covering optimization problems such as convex analysis, optimality conditions, Lagrangian duality and optimality (including the Karush-Kuhn-Tucker (KKT) condition and the augmented Lagrangian), continuity, Lipschitz conditions and stability, fixed-point iterations and convergence, as well as monotone operators and variational properties. Chapter 3 is dedicated to a review of classical decomposition methods, including the ordinary Lagrangian relaxation method, augmented Lagrangian method, Bertsekas decomposition method, Tanikawa-Mukai decomposition method, Tadjewski's decomposition method, and the SALA decomposition algorithm in the ADMM version. In Chapter 4, the convergence analysis of the asynchronous Bertsekas augmented Lagrangian method is discussed. A proof of the convergence theorem for the asynchronous algorithm is presented along with useful lemmas.

In Chapter 5, the author considers an especially important issue, as the network optimization problem of simultaneous routing and bandwidth allocation in energy-aware networks is presented. Here, selected implemented methods are tested on the network problems of different sizes. The tested networks consist of loosely connected node clusters with strongly connected nodes. The experimental results presented in summary tables and figures illustrate well the studied optimization algorithms. In Chapter 6, the implemented two variants of the Bertsekas methods (the synchronous and the asynchronous) are investigated. Using the same dataset as in the Chapter 5, numerical experiments were conducted with the use of Python. The experimental results are given for small, large and extra-large problem objective values (routing). The short Chapter 7 presents the solution of regularized linear systems with augmented Lagrangian algorithms. In numerical experiments, three linear systems, designed to reflect varying data structures and matrix conditions, are implemented. Among them are: an image inpainting task, using randomized diagonal measurement

operators, and a biomedical dataset. Chapter 8 explores both synchronous and asynchronous update schemes within the Bertsekas augmented Lagrangian framework, highlighting their algorithmic structures and practical performance across diverse datasets. In the two last chapters (9 and 10), the author presents the distributed formulation of the constrained K-Means problem and investigates the impact of asynchrony on K-Means clustering through empirical comparisons on large-scale datasets. A summary of contributions and suggestions for future research are presented in the Chapter 11.

On the basis of a brief overview of the content of the dissertation, I conclude that Anthony Chukwuemeka Nwachukwu, M.Sc., has demonstrated a high degree of ability to formulate scientific problems, depending on the analysis of the current state of knowledge available in the contemporary international literature. The research question: „Which Lagrangian-based optimization algorithms are most effective for solving large-scale, distributed non-convex problems, and how can we design and rigorously analyze asynchronous versions of these methods? has been solved using appropriate methods and algorithms of the non-convex optimization theory. Extensive experimental studies were carried out using both available synthetic benchmarks and real-world datasets. The thesis is of a theoretical nature and the description of the problem and its solution uses the mathematical formalism appropriate for the discipline of technical informatics and telecommunications.

3. Original achievements and the author's contribution

By answering the research question and achieving the formulated goals, the author obtained several interesting and original scientific results confirmed both by formal proof and by experimental computer research, using available synthetic benchmarks and real-world datasets. The important scientific achievements of Anthony Chukwuemeka Nwachukwu, M.Sc., and his main contribution to the discipline of *Technical informatics and telecommunications* include, among others, the following:

a) The major achievement of the dissertation is the development of an original asynchronous variant of the Bertsekas augmented Lagrangian decomposition algorithm. While the Bertsekas method is a well-established and powerful approach in convex and certain non-convex settings, its extension to an asynchronous computational model is

nontrivial. The author successfully reformulates the algorithmic framework in order to allow local computational nodes to update their variables without resorting to global synchronization and with no risk of using potentially outdated information. This constitutes a genuine methodological contribution that goes beyond straightforward adaptations of existing synchronous algorithms.

b) A special significance of the work lies in the rigorous theoretical analysis accompanying the proposed algorithm. The dissertation provides a detailed convergence analysis under a bounded-delay asynchronous model, which is well aligned with standard assumptions used in the analysis of asynchronous iterative methods. Under appropriate regularity conditions, including smoothness, Lipschitz continuity of gradients, and suitable constraint qualifications, the author proves convergence of the proposed method to Karush–Kuhn–Tucker points of the original non-convex problem. Given the inherent difficulty of establishing convergence guarantees for asynchronous algorithms in non-convex settings, this analysis represents a substantial theoretical contribution.

c) The dissertation is distinguished by its comprehensive and well-organized review of augmented Lagrangian–based decomposition methods. Classical dual decomposition, the method of multipliers, ADMM, the Bertsekas method, Tatjewski’s approach, and the SALA algorithm are presented in a systematic manner. Importantly, this review is not merely descriptive; it serves as a foundation for a comparative empirical study that validates the choice of the Bertsekas method as the algorithmic core of the dissertation. This critical and selective use of literature demonstrates the author’s strong understanding of the field and the ability to position their work within existing research.

d) The breadth and depth of the numerical experiments constitutes another achievement. The proposed algorithms are validated on a wide range of problems, including simultaneous routing and bandwidth allocation in energy-aware networks, the solution of large regularized systems of linear equations, and distributed K-means clustering. These applications cover both network optimization and machine learning, thereby illustrating the universality of the proposed framework. The experiments are performed on both synthetic benchmarks and real-world datasets, enhancing the credibility and practical relevance of the results.

e) The dissertation demonstrates that the asynchronous Bertsekas algorithm achieves objective values comparable to those obtained by synchronous implementations, while significantly reducing wall-clock computation time as the degree of parallelism increases. This

empirical evidence strongly supports the main claim of the dissertation, namely that asynchronous augmented Lagrangian methods can exploit modern parallel computing environments more efficiently than their synchronous counterparts.

f) The work also succeeds in bridging the gap between theory and practice. The mathematical analysis is closely connected to the algorithmic design, and the assumptions made in the theoretical part are reflected in the implementation and experiments. This coherence between theoretical development and practical validation is a notable strength of the dissertation.

The original scientific results presented in the thesis and the contribution of Anthony Chukwuemeka Nwachukwu, M.Sc., to the discipline of *Technical informatics and telecommunications*, confirmed to a large extent by theoretical and experimental research, in some part have already been published in several collaborative scientific works, namely in the following journals: *International Journal of Electronics and Telecommunications* (2025, 70 points), *Energies* (2024, 140 points) and *Task Quarterly* (2020, 20 points).

Finally, the dissertation is well structured and clearly written. The progression from theoretical foundations, through a review of existing methods, to the development of a new algorithm and its applications is logical and easy to follow. The figures, tables, and numerical results are generally well presented, and the conclusions are supported by the given evidence. Overall, the dissertation demonstrates the author's maturity as a researcher and their ability to conduct independent, high-quality scientific work.

4. Remarks and comments

Overall, I rate the entire dissertation positively in terms of content and editing. Some of the comments below are of substantive and some of editorial nature:

a) One of the main issues regards the assumptions underlying the theoretical convergence analysis:

- While the assumptions of smoothness, Lipschitz continuity of gradients, and appropriate constraint qualifications are a standard in the literature, they are relatively strong and may not always be satisfied in practical large-scale applications. The dissertation could have benefited from a more extensive discussion of how restrictive

these assumptions are in practice and to what extent is the proposed method inclined to their violation.

- The convergence results establish convergence to KKT points, which is the strongest result one can typically expect in a non-convex optimization. However, the practical implications of convergence to stationary points, as opposed to global optima, are discussed nonexhaustively. A more thorough discussion of the quality of the obtained solutions in non-convex settings, supported, preferably, by additional empirical observations, would strengthen the practical interpretation of the theoretical results.

b) Although the dissertation includes comparisons with several classical and well-established methods, such as ADMM and other augmented Lagrangian-based approaches, the set of benchmark algorithms could be expanded. In particular, recent developments in asynchronous non-convex optimization, including novel variants of asynchronous ADMM or stochastic methods, are considered not extensively. Inclusion of such methods could provide a broader perspective on the relative performance of the proposed algorithm.

c) Editorial remarks:

- The theoretical background chapter, while thorough and mathematically sound, is relatively extensive. Some of the material presented is well known to experts in optimization and could have been condensed. A more concise presentation of the standard results might have allowed for greater emphasis to be put on the novel theoretical contributions, specific to the asynchronous algorithm developed in the dissertation.

- From a visual perspective, the numerical results section includes a very large number of figures and tables. While this reflects the breadth of the experimental study, it can make it difficult at times for the reader to identify quickly the most important trends and conclusions. The readability of those sections could be improved by more frequent use of summary tables or explicit statements highlighting the key empirical findings.

- Although the dissertation includes a section on limitations and future work, the limitations of the proposed method could have been more explicitly emphasized throughout the text. Issues such as sensitivity to parameter selection, the impact of

very large communication delays or scalability limits in extremely large or highly heterogeneous systems are mentioned only briefly. A more critical discussion of these aspects would provide a more balanced assessment of the applicability of the method.

- In doctoral dissertations, it is recommended to include summaries of individual chapters, in which the author introduces new content and results. The lack of such summaries somewhat hinders the substantive and logical conclusions that could be drawn between chapters.

4. Summary

In conclusion, the identified weaknesses do not undermine the overall quality or significance of the dissertation. Rather, they point to natural directions for further research and refinement of the theses. The dissertation remains a strong and valuable contribution to the field of distributed optimization.

Considering all aspects of the doctoral dissertation by Anthony Chukwuemeka Nwachukwu, M.Sc., including the noted scientific and applied achievements in particular, as well as the list of co-authored publications, and taking into account the substantive and formal discussion comments, that do not have a significant impact on the final assessment of the dissertation:

- a) I declare that the doctoral dissertation meets all the requirements of the Law on Higher Education and Science of 20 July 2018 (Article 168) in the field of engineering and technology in the discipline of *Technical informatics and telecommunications*;
- b) I request that the dissertation be accepted by the *Discipline Council for Technical Informatics and Telecommunications* of Warsaw University of Technology.



Prof. PŚ dr hab. inż. Jerzy S. Respondek
Politechnika Śląska
Wydział Automatyki, Elektroniki i Informatyki
Katedra Informatyki Stosowanej
ul. Akademicka 16
44-100 Gliwice

Erlangen, Germany, January 27, 2026

Review of doctoral dissertation

Anthony Chukwuemeka Nwachukwu, M.Sc.

Entitled

Augmented Lagrangian-Based Algorithms for Separable Non-convex Optimization with Applications in Network Routing and Machine Learning

Implemented in the unit

*Warsaw University of Technology
Faculty of Electronics and Information Technology*

1. The research problem, its solution, and significance

In mathematical terms, the dissertation focuses on non-convex static optimization problems solved using Lagrange multipliers. There are constraints in the form of inequality systems. Importantly, they are separated to a certain extent, the significance of which will be discussed later.

In terms of the IT tools used, numerical methods are applied, importantly, in a parallel and distributed version.

In this context, the development of a new Bertsekas decomposition method is a particularly important added value provided by the doctoral student. From a purely mathematical point of view, the significance of this innovation may be underestimated. However, when combined with a distributed implementation, it fully proves its innovativeness and importance. Namely, it allows for a particularly effective implementation of the constructed optimization method in a parallel version, as it obviously significantly

reduces the need for frequent synchronization of individual computing nodes, where each can, in a sense, autonomously deal with its part of a larger but decomposable optimization problem. Thus, we have the benefits of less frequent data updates and less time spent transferring network data for synchronization in individual nodes.

Chapter 1 precisely formulates the research problem.

Chapter 2 presents the mathematical foundations of calculations in the field of convexity, optimal solution conditions, Lagrange's method, and duality of the problem. It also discusses the convergence conditions of the iterative methods used in the dissertation.

Chapter 3 discusses Lagrange methods in more detail, and in Chapter 4, the doctoral student moves on to separable problems.

In Chapter 4, he independently formulates and develops Bertsekas' asynchronous method.

Chapter 5 applies the developed methods of computational mathematics to the simultaneous routing and bandwidth allocation problem.

Chapter 6 addresses the same technical problem in synchronous and asynchronous versions of the algorithms.

Chapters 7 and 8 apply the developed Lagrange methods to regularized linear systems.

Another technical application is discussed in chapters 9 and 10, namely K-means clustering.

In conclusion, the doctoral student accurately defines the scope of possible applications of his methods and is aware of their limitations. In the summary, he accurately diagnoses open problems. What is more, one of the applications, in network throughput and routability research, has potential for research on the efficiency of parallel and distributed algorithms, in particular those presented in earlier chapters of the thesis, but not only those.

Therefore, the work should be considered thematically coherent, which is not easy to achieve due to its significant interdisciplinary nature.

The above tools and methods are interdisciplinary in nature and cover mathematics, computational informatics (in parallel version), and applications related to technical informatics.

The dissertation is carefully written not only in terms of content, but also in terms of editing. The results obtained are illustrated with numerous one- and two-dimensional graphs and diagrams.

Importantly, the results have been published in two different reputable journals. The choice of journals is relevant to the subject matter of the dissertation, thus providing an additional guarantee that the published results have been competently verified through peer review. The work is interdisciplinary, covering mathematics and computational informatics, and its contribution to the discipline is significant.

In summary, I highly appreciate the innovative level of the work, in particular the development and adaptation of classical mathematical methods in such a way that they are effective in parallel algorithmic implementations without losing or even compromising their accuracy.

Therefore, the work is an example of a successful combination of the results of several disciplines and, thus, interdisciplinary research. Interestingly, the relevance of the selected applications is evidenced by the fact that they have the potential to increase the effectiveness of the theoretical subject matter of the work itself; increasing the efficiency of the network will also increase the efficiency of the implementation itself, which in turn can also be applied to research on further increasing the efficiency of the network. This is a kind of positive feedback in the development of the subject, in the positive sense of the word.

2. Author profile

The doctoral candidate graduated in mathematics from the Federal University of Technology Owerri in Nigeria. He obtained a Master of Science in Nanotechnology from Gdańsk University of Technology. In addition, he obtained a second Master of Science degree in mathematics from the Italian University of L'Aquila. Therefore, the candidate is very well prepared theoretically, and his two master's degrees—one in mathematics and the other in the more physical field of nanotechnology—should be appreciated and complement each other.

He has participated in numerous projects, including fast matrix multiplication, which was achieved using parallel algorithms in Fortran.

The doctoral student also participated in several summer schools in Berlin, Rennes, and Oxford. As can be seen, these include prestigious institutions.

There is no information about the scientific conferences he attended.

His teaching activity is at a good level, which is important as it is consistent with the subject of his work. He taught classes in Numerical Methods and Optimization Techniques.

The author's contribution is confirmed by two publications in reputable journals. Below is an analysis of the candidate's bibliometric data:

According to the Scopus database, the candidate is the co-author of two publications. So far, they have not been independently cited, but they were published very recently, i.e., in 2024 and 2025, respectively. These journals are:

- "*International Journal of Electronics and Telecommunications*" – a journal included in the so-called Philadelphia Journal List. Although it is a national journal published by the Polish Academy of Sciences Committee of Electronics and Telecommunications, with a ministerial score of 70 points, it is indexed by the prestigious Web of Science database in the "Core" collection. In addition, another publication in this journal has been accepted for printing.

- "*Energies*" which is published by MDPI, but has a ministerial score of 140 points and is also indexed by both Scopus and the Core collection of the Web of Science database.

In addition, he is the author of one national publication in the quarterly journal of the TriCity Academic Computer Network Information Center.

The above bibliometric indicators are auxiliary in assessing the Author's achievements. The number of significant publications should be assessed as good at the stage of the scientific career at which the Candidate currently finds himself.

3. Summary

Pursuant to the Act of July 20, 2018, Law on Higher Education and Science (Polish Journal of Laws 2023, item 742), **I hereby declare that the doctoral dissertation submitted by Mr. Anthony Chukwuemeka Nwachukwu meets the requirements for doctoral dissertations in the**

Discipline of science: Information and Communications Technology,

Field of Science: Engineering and Technology,

and I request that it be admitted for public defense.



Edinburgh, 21 March 2026.

Professor Jacek Gondzio
School of Mathematics
University of Edinburgh
James Clerk Maxwell Building
King's Buildings
Edinburgh, EH9 3JZ, UK

Reviewer's Report on the PhD thesis:
**Augmented Lagrangian-Based Algorithms for Separable Non-convex Optimization
with Applications in Network Routing and Machine Learning**
by Mr Anthony Chukwuemeka Nwachukwu, submitted for a degree of PhD
at the Faculty of Electronics and Information Technology, Warsaw University of Technology

The PhD dissertation of Mr Anthony Nwachukwu is concerned with the asynchronous version of a proximal Augmented Lagrangian method. Several variants of the method are analysed and implemented and their practical behaviour is compared.

Augmented Lagrangian methods are well suited to solving very large loosely coupled optimization problems. The key difficulty when dealing with such models is the ability to exploit the near-separability between almost independent blocks. This usually requires ignoring some of the connections (couplings) between the otherwise independent models and using a sophisticated coordination scheme which maintains a 'dialogue' between the small models. Decomposition of a large model into blocks offers an advantage that (possibly many) small subproblems will be dealt with independently, and this clearly opens many opportunities to exploit parallelism. However, it comes with a challenge of how information from such distributed optimization models is coordinated to guarantee a convergence to a global optimal solution. In particular, an important decision has to be made regarding how information arriving from subproblems is processed and two options are possible:

- (a) a *synchronous* processing which requires gathering information from all subproblems at the same time,
- (b) an *asynchronous* processing in which the coordinator updates its decision as soon as it receives fresh information from any single subproblem.

The theoretical and practical convergence of the latter is often problematic. Its analysis is significantly more complicated than that of the former. The PhD dissertation of Mr Anthony Nwachukwu addresses this challenging problem and delivers an original solution to it.

General Comments and Observations:

The dissertation promises to address nonconvex optimization problems and to develop a convergent asynchronous algorithm to solve such problems.

In fact the developments are somewhat restricted.

- The dissertation claims to provide the analysis of the method which allows to deal with nonconvex objective. The analysis relies on the use of strong regularization which convexifies the objective and therefore allows to generalize standard techniques developed for proximal algorithms applied in convex case to the nonconvex one.
- The nonconvexity of the objective function is remedied by the use of proximal term. For seriously nonconvex objectives these proximal terms may have to be very significant. To satisfy Assumption 1 and to make the cost function $Q_\rho(x, s)$ (equation (57)) strictly convex one would need ρ in the proximal terms $\rho\|x - s\|^2$ to be (very) large. Of course this is achievable for sufficiently large ρ , but it would alter the problem very significantly.
- This work aimed at delivering a complete analysis and evaluating a practical performance of an asynchronous proximal Augmented Lagrangian method. As it has been stated in Section 6.3.1, the asynchronous variants of the algorithms have been emulated by introducing artificial delays. It would be interesting to know how this simulated asynchronous environment differs from the real one.
- Regardless my mild criticism in the earlier item, this dissertation presents strong and exhaustive computational evidence. Several variants of algorithms have been implemented and tested when applied to various relevant data science applications, such as simultaneous routing and bandwidth allocation problem, solving large systems of linear equations, and k-means clustering. The Author has demonstrated his ability to model the problem and to apply a sophisticated optimization technique to its solution.

Detailed Comments and Small Corrections:

It is not possible to avoid misprints and minor errors when writing a book of more than a hundred pages, especially when this is done in a foreign language. The list of a few deficiencies noticed while reading the dissertation is given below. They do not affect my overall positive evaluation of this dissertation.

The Author uses shortcuts to mention various methods/algorithms, for example, calling them "Bertsekas", "Tatjewski", etc. I would suggest replacing them with more complete names such as "Algorithm of Bertsekas", "Algorithm of Tatjewski", respectively.

Page 19, line -3: The word 'asynchrony' has been used here (and then twice on page 21). I'm not a native English speaker hence my views of the use of words may not always be correct. However, I think that the word 'asynchronicity' might fit better in this context.

Page 20, equation (3): There is an inconsistency of notation: the image of the function g is \mathcal{R}^q while the indices run through $j = 1, 2, \dots, m$.

Page 24: A subdifferential is mentioned here. It would be good to define the notions of subgradient and subdifferential before they are used.

Page 24: There is a clash of notation. Capital C is used to denote a convex set and then to denote a matrix in the linear constraints $Cx = d$.

Page 24: A sentence "These facts ensure ... in decomposition methods" is repeated twice on this page.

Page 25: The KKT conditions are introduced without prior mentioning of differentiability of f, g and h .

Page 26: You may add \hyphenation{Lipschitz-continuous} to improve the printout.

Page 35 (in the middle): In a good style of writing it is better to avoid statements like "it can be easily proved that ..."

Here, in particular, it does not seem obvious why all optimal λ_j^* should be the same. Please would the Author elaborate on that?

Page 37, the last line before Section 4.1: Replace: "and will prove its convergence." with: "its convergence will be proved."

Page 38: Assumption 1 is a bit unclear. Do you mean that the cost function $Q_\rho(x, s)$ (equation (57)) is strictly convex? This would be a rather strong assumption that a possibly nonconvex function f is very strongly regularized by the proximal terms of form $\rho\|x - s\|^2$. Of course this is achievable for sufficiently large ρ , but it would alter the problem very significantly.

Page 41: It would be helpful to add a comment/explanation of the role played by the coefficients $a_i(k', k)$ and $b_j(k', k)$ (equations (81) and (87), respectively).

Page 42, line 2 of "i's Algorithm": Replace: "and compute $\hat{\mu}^k$." with: "and computes $\hat{\mu}^k$."

Page 42, line 2 of "j's Algorithm": Replace: "and compute \hat{x}_i^{k+1} ." with: "and computes \hat{x}_i^{k+1} ."

Page 42, line between equations (88) and (89): What do you mean by "the corresponding exact is"? Is it "the corresponding exact variable is"?

Page 42, lines -3 and -1: Replace: "converges to zero" with: "converge to zero"

Page 43, proof of Lemma 1: There is a slight inaccuracy here. Theoretically, we could have the situation that (92) is symmetric positive definite, but there is a sequence of μ and s (for a given ρ) such that the Hessian (92) would converge to a singular matrix meaning that its inverse would not be bounded (actually, it would not exist). The Author may fix this easily by choosing large enough ρ .

Page 46 (equation (115)): You should add that you use $\|\cdot\|_1$ here.

In general, please would you indicate clearly what norms are used in various equations (unless it is a default Euclidean norm).

Page 47, proof of Lemma 4: The Assumptions 5 and 6 involve Euclidean norms, hence the constants Φ_L and Φ_g in equation (116) may need to be scaled to reflect norm equivalence.

Page 47, Lemma 5: A new variable π^k has appeared here without any explanation what it means. Generally, Lemma 5 is stated without prior definition of the notation used. This makes the reading and checking of this part of the dissertation particularly challenging, practically impossible.

Page 50, line below equation (134): The Author says: "... and the definition of π we have:" Please would you be more precise: where was π defined?

Page 52, line below equation (145): Replace: "Hence, " with: "we obtain:"

Page 59, line 6 in the third paragraph: Replace: "for bigger networks" with: "for larger networks"

Page 61: When notation is introduced, please would you also explain the meaning of $y_{w,e(i,j)}$ used in equation (176).

Page 62, line -3: Should "with respect to flows $w \in W$ " be replaced with "with respect to demands $w \in W$ "?

Page 78, line -2: Replace: "and calculate $\hat{\mu}_i^k$ " with: "and calculates $\hat{\mu}_i^k$ "

Page 86, the last paragraph before Section 7.1.1: The sentence: "These methods enable ... and machine learning." suggests that ℓ_2 -regularization is commonly used in compressed sensing and machine learning. I would not agree with that. The applications arising in compressed sensing and machine learning usually look for sparse solutions, which involves ℓ_1 -regularization.

Page 86, the line above equation (243): Replace: "the dual variables are updated ..." with: "the variables are updated ..."

Page 86, the line before equation (246): Replace: "the dual variables are updated ..." with: "the variables are updated ..."

Page 87, the line before equation (248): Replace: "the dual variables are updated ..." with: "the variables are updated ..."

Page 87, the line before equation (252): Replace: "the dual variables are updated ..." with: "the variables are updated ..."

Page 89, line 7: The Author says: "The corresponding observation vector ... and positive semi-definite system." Why should this linear system involve a positive semi-definite matrix?

Page 90, line -7: The Author says: "... while ADMM and Tadjewski exhibit modestly higher computational cost." Actually, reading Table 7.1, ADMM has the same performance measures as the method of Bertsekas (when 12 partitions are used).

Page 95, the second paragraph: The Author says: "... the equality constraints are equivalently rewritten as a pair of inequalities ..." Such a transformation is extremely dangerous. It doubles the number of constraints and introduces (near) linear dependence of them. This might lead to serious rank deficiency of matrices involved in the linear algebra operations.

Page 99, line 4 above Table 8.1: The Author says: "These results suggest that asynchronous distributed optimization is especially effective in scenarios involving low-dimensional or poorly conditioned systems..." This observation is based on the performance on just one randomly generated test example. Such a statement is extremely risky and should be eliminated unless substantial computational evidence is provided to support it.

Page 106, equation (291): Using a quadratic penalty for any excess in the inequality (282) does not seem correct.

Page 108, line 1 in Section 9.2.1: Replace: "generalizability" with: "generality"

References [25], [34], [35], [49], [58]: The names or geographical names should be written using capital letters: [25] - Rockafellar, [34] - Lagrangian, [35] - Lagrange, [49] - ADMM, [58] - Wisconsin.

Conclusion

This dissertation documents a considerable effort made by the Author to develop a new asynchronous variant of the proximal Augmented Lagrangian algorithm. The method has been thoroughly analysed, its convergence has been established, its implementation has been developed and computational evidence has been provided to demonstrate its practical behaviour. The dissertation is mostly mathematically sound and generally well written. It contains original contributions to the optimization theory and practice.

This dissertation meets all requirements for Mr Anthony Chukwuemeka Nwachukwu to be awarded the PhD Degree.

It is a quality work which deserves being awarded a distinction.



Professor Jacek Gondzio