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 papge 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 $j = 1, 2, \ldots, m$.

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 $\|.\|_1$ 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 $\hat{\mu}_l^k$" with: "and calculates $\hat{\mu}_l^k$"

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 Tatjewski 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.

J Gondzio

Professor Jacek Gondzio