

Corrigendum: A proximal quasi-Newton trust-region method for nonsmooth regularized optimization

Supplementary material

A. Aravkin, R. Baraldi, G. Leconte, D. Orban

G–2021–12

March 2021

Revised: March 2024

La collection *Les Cahiers du GERAD* est constituée des travaux de recherche menés par nos membres. La plupart de ces documents de travail a été soumis à des revues avec comité de révision. Lorsqu'un document est accepté et publié, le pdf original est retiré si c'est nécessaire et un lien vers l'article publié est ajouté.

The series *Les Cahiers du GERAD* consists of working papers carried out by our members. Most of these pre-prints have been submitted to peer-reviewed journals. When accepted and published, if necessary, the original pdf is removed and a link to the published article is added.

Citation suggérée : A. Aravkin, R. Baraldi, G. Leconte, D. Orban (Mars 2021). Corrigendum: A proximal quasi-Newton trust-region method for nonsmooth regularized optimization, Supplementary material. Rapport technique, Les Cahiers du GERAD G– 2021–12, GERAD, HEC Montréal, Canada. Version révisée: Mars 2024

Suggested citation: A. Aravkin, R. Baraldi, G. Leconte, D. Orban (March 2021). Corrigendum: A proximal quasi-Newton trust-region method for nonsmooth regularized optimization, Supplementary material. Technical report, Les Cahiers du GERAD G–2021–12, GERAD, HEC Montréal, Canada. Revised version: March 2024

Avant de citer ce rapport technique, veuillez visiter notre site Web (<https://www.gerad.ca/fr/papers/G-2021-12>) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

Before citing this technical report, please visit our website (<https://www.gerad.ca/en/papers/G-2021-12>) to update your reference data, if it has been published in a scientific journal.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2021
– Bibliothèque et Archives Canada, 2021

Legal deposit – Bibliothèque et Archives nationales du Québec, 2021
– Library and Archives Canada, 2021

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Télec. : 514 340-5665
info@gerad.ca
www.gerad.ca

Corrigendum: A proximal quasi-Newton trust-region method for nonsmooth regularized optimization

Supplementary material

Aleksandr Aravkin ^a

Robert Baraldi ^b

Geoffroy Leconte ^c

Dominique Orban ^c

^a *Department of Applied Mathematics, University of Washington, Seattle WA., USA*

^b *Optimization and Uncertainty Quantification, Sandia National Laboratories, Albuquerque, NM, USA*

^c *GERAD & Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal (Québec), Canada*

saravkin@uw.edu

rjbaral@sandia.gov

geoffroy.leconte@polymtl.ca

dominique.orban@gerad.ca

March 2021

Revised: March 2024

Les Cahiers du GERAD

G–2021–12

Copyright © 2021 Aravkin, Baraldi, Leconte, Orban

Les textes publiés dans la série des rapports de recherche *Les Cahiers du GERAD* n'engagent que la responsabilité de leurs auteurs. Les auteurs conservent leur droit d'auteur et leurs droits moraux sur leurs publications et les utilisateurs s'engagent à reconnaître et respecter les exigences légales associées à ces droits. Ainsi, les utilisateurs:

- Peuvent télécharger et imprimer une copie de toute publication du portail public aux fins d'étude ou de recherche privée;
- Ne peuvent pas distribuer le matériel ou l'utiliser pour une activité à but lucratif ou pour un gain commercial;
- Peuvent distribuer gratuitement l'URL identifiant la publication.

Si vous pensez que ce document enfreint le droit d'auteur, contactez-nous en fournissant des détails. Nous supprimerons immédiatement l'accès au travail et enquêterons sur votre demande.

The authors are exclusively responsible for the content of their research papers published in the series *Les Cahiers du GERAD*. Copyright and moral rights for the publications are retained by the authors and the users must commit themselves to recognize and abide the legal requirements associated with these rights. Thus, users:

- May download and print one copy of any publication from the public portal for the purpose of private study or research;
- May not further distribute the material or use it for any profit-making activity or commercial gain;
- May freely distribute the URL identifying the publication.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Abstract : The purpose of the present note is to bring clarifications to certain concepts and surrounding notation of [1]. All results therein continue to hold. The clarifications ensure complete agreement between the paper and our implementations.

Acknowledgements: The research of A. Aravkin is partially supported by the Washington Research Foundation. The work of R. Baraldi was sponsored by the Department of Energy Office of Science, Office of Advanced Scientific Computing Research's John von Neumann Fellowship. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The research of D. Orban is supported by an NSERC Discovery grant.

Disclaimer: This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

In (3.2), we should have defined $\varphi(s; \mathbf{x}) := \varphi_{\text{cp}}(s; \mathbf{x}) + \frac{1}{2}\nu^{-1}\|s\|^2$ where φ_{cp} is the first order Taylor approximation

$$\varphi_{\text{cp}}(s; \mathbf{x}) = f(\mathbf{x}) + \nabla f(\mathbf{x})^\top s, \quad (1)$$

and “cp” stands for “Cauchy point”. In (3.4), $s_{k,1}$ could have been defined equivalently as the solution of

$$s_{k,1} \in \underset{s}{\operatorname{argmin}} \varphi_{\text{cp}}(s; \mathbf{x}_k) + \frac{1}{2}\nu^{-1}\|s\|^2 + \psi(s; \mathbf{x}_k) + \chi(s; \Delta_k). \quad (2)$$

The first step of the proximal-gradient algorithm applied to minimize $\varphi_{\text{cp}}(s; \mathbf{x}_k) + \psi(s; \mathbf{x}_k) + \chi(s; \Delta_k)$ with step size ν is then precisely $s_{k,1}$.

In addition to the definition of ξ in (3.6), we also define

$$\xi_{\text{cp}}(\Delta; \mathbf{x}_k, \nu) := f(\mathbf{x}_k) + h(\mathbf{x}_k) - \varphi_{\text{cp}}(s_{k,1}; \mathbf{x}_k) - \psi(s_{k,1}; \mathbf{x}_k). \quad (3)$$

Note that for $\varphi = \varphi_{\text{cp}}$,

$$\begin{aligned} \xi(\Delta_k; \mathbf{x}_k, \nu_k) &= (f + h)(\mathbf{x}_k) - (\varphi_{\text{cp}}(s_{k,1}; \mathbf{x}_k) + \frac{1}{2}\nu_k^{-1}\|s_{k,1}\|^2 + \psi(s_{k,1}; \mathbf{x}_k)) \\ &\leq (f + h)(\mathbf{x}_k) - (\varphi_{\text{cp}}(s_{k,1}; \mathbf{x}_k) + \psi(s_{k,1}; \mathbf{x}_k)) \\ &= \xi_{\text{cp}}(\Delta_k; \mathbf{x}_k, \nu_k). \end{aligned}$$

Thus $\xi_{\text{cp}}(\Delta_k; \mathbf{x}_k, \nu_k) = 0 \implies \xi(\Delta_k; \mathbf{x}_k, \nu_k) = 0$, and, by Proposition 3.3, \mathbf{x}_k is first-order stationary. Using the definition of $s_{k,1}$, we have

$$\varphi_{\text{cp}}(s_{k,1}; \mathbf{x}_k) + \frac{1}{2}\nu_k^{-1}\|s_{k,1}\|^2 + \psi(s_{k,1}; \mathbf{x}_k) \leq \varphi_{\text{cp}}(0; \mathbf{x}_k) + \psi(0; \mathbf{x}_k).$$

The above inequality, combined with the facts that $\varphi(0; \mathbf{x}_k) = f(\mathbf{x}_k)$ and $\psi(0; \mathbf{x}_k) = h(\mathbf{x}_k)$, leads to

$$\xi_{\text{cp}}(\Delta_k; \mathbf{x}_k, \nu_k) = f(\mathbf{x}_k) + h(\mathbf{x}_k) - (\varphi_{\text{cp}}(s_{k,1}; \mathbf{x}_k) + \psi(s_{k,1}; \mathbf{x}_k)) \geq \frac{1}{2}\nu_k^{-1}\|s_{k,1}\|^2. \quad (4)$$

In Step Assumption 3.1, (3.8b) should read

$$m_k(0; \mathbf{x}_k) - m_k(s_k; \mathbf{x}_k) \geq \kappa_{\text{mdc}} \xi_{\text{cp}}(\Delta_k; \mathbf{x}_k, \nu_k). \quad (5)$$

The justification of (3.8b) (below Model Assumption 3.2) is no longer relevant with the above change. However, the following proposition shows that (5) in Step Assumption 3.1 can still be satisfied if

$$\varphi(s; \mathbf{x}_k, \mathbf{B}_k) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^\top s + \frac{1}{2}s^\top \mathbf{B}_k s = \varphi_{\text{cp}}(s; \mathbf{x}_k) + \frac{1}{2}s^\top \mathbf{B}_k s, \quad (6)$$

where $\{\mathbf{B}_k\}$ is bounded. In addition, we require the existence of a maximal trust-region radius $\Delta_{\text{max}} \geq \Delta_0$. This can be enforced in Algorithm 3.1 at the end of Step 11 with the following assignment:

$$\Delta_{k+1} \leftarrow \min(\Delta_{k+1}, \Delta_{\text{max}}). \quad (7)$$

Proposition 1. *Let Algorithm 3.1 be modified according to (7), and $\varphi(s; \mathbf{x}_k, \mathbf{B}_k)$ defined in (6) be used in Step 5. Let Model Assumption 3.2 be satisfied, and the sequence of Hessian approximations $\{\mathbf{B}_k\}$ generated by Algorithm 3.1 be bounded. Then, there exists $\kappa_{\text{mdc}} \in (0, 1)$ such that (5) holds for all k .*

Proof. By definition of s_k , computed in Step 8 of Algorithm 3.1 using (6) in Step 5,

$$\begin{aligned} \varphi(s_k; \mathbf{x}_k, \mathbf{B}_k) + \psi(s_k; \mathbf{x}_k) &= m(s_k; \mathbf{x}_k, \mathbf{B}_k) \leq m(s_{k,1}; \mathbf{x}_k, \mathbf{B}_k) = \varphi_{\text{cp}}(s_{k,1}; \mathbf{x}_k) + \frac{1}{2}s_{k,1}^\top \mathbf{B}_k s_{k,1} + \psi(s_{k,1}; \mathbf{x}_k) \\ &\leq \varphi_{\text{cp}}(s_{k,1}; \mathbf{x}_k) + \frac{1}{2}\|\mathbf{B}_k\| \|s_{k,1}\|^2 + \psi(s_{k,1}; \mathbf{x}_k). \end{aligned}$$

We multiply the above inequality by -1 and add $m(0; x_k, B_k) = m(0; x_k, \nu_k) = f(x_k) + h(x_k)$ to both sides to obtain

$$m(0; x_k, B_k) - m(s_k; x_k, B_k) \geq \xi_{\text{cp}}(\Delta_k; x_k, \nu_k) - \frac{1}{2} \|B_k\| \|s_{k,1}\|^2.$$

To satisfy (5), it is sufficient to show that there exists $\kappa_{\text{mdc}} \in (0, 1)$ such that

$$\xi_{\text{cp}}(\Delta_k; x_k, \nu_k) - \frac{1}{2} \|B_k\| \|s_{k,1}\|^2 \geq \kappa_{\text{mdc}} \xi_{\text{cp}}(\Delta_k; x_k, \nu_k),$$

i.e.,

$$(1 - \kappa_{\text{mdc}}) \xi_{\text{cp}}(\Delta_k; x_k, \nu_k) \geq \frac{1}{2} \|B_k\| \|s_{k,1}\|^2.$$

Using (4), it is also sufficient to show that there exists $\kappa_{\text{mdc}} \in (0, 1)$ such that

$$(1 - \kappa_{\text{mdc}}) \nu_k^{-1} \geq \|B_k\|. \quad (8)$$

If $\|B_k\| = 0$, (8) is satisfied for any $\kappa_{\text{mdc}} \in (0, 1)$. Otherwise, because (6) indicates that the Lipschitz constant $L(x_k)$ of $\nabla\varphi(\cdot; x_k, B_k)$ in Model Assumption 3.2 is $L(x_k) = \|B_k\|$, the definition of ν_k in Step 4 of Algorithm 3.1 yields

$$\|B_k\| \nu_k = \frac{\|B_k\|}{\|B_k\| + \alpha^{-1} \Delta_k^{-1}} \leq \frac{1}{1 + \alpha^{-1} \Delta_k^{-1} \|B_k\|^{-1}} \leq \frac{1}{\alpha^{-1} \Delta_{\max}^{-1} (\sup_{k \in \mathbb{N}} \|B_k\|)^{-1} + 1} \in (0, 1), \quad (9)$$

because $\{B_k\}$ is bounded by assumption. We deduce from (9) that (8), and therefore (5) hold. \square

In the proof of Theorem 3.4, (4) guarantees that the inequality

$$\xi_{\text{cp}}(\Delta_k; x_k, \nu_k) \geq \frac{1}{2} \nu_k^{-1} \|s_{k,1}\|^2 \geq \frac{1}{2} (L(x_k) + \alpha^{-1} \Delta_k^{-1}) \|s_{k,1}\|^2 \geq \frac{1}{2} \alpha^{-1} \Delta_k^{-1} \|s_{k,1}\|^2, \quad (10)$$

continues to hold, where $L(x_k)$ is the Lipschitz constant of the gradient of $\varphi(\cdot; x_k)$. Thus, the theorem remains valid.

In the simple case where $h = 0$ and $s_{k,1} = -\nu_k \nabla f(x_k)$ is inside the trust-region,

$$\xi_{\text{cp}}(\Delta_k; x_k, \nu_k) = -\nabla f(x_k)^T s_{k,1} = \nu_k \|\nabla f(x_k)\|^2, \quad (11)$$

which suggests that the appropriate criticality measure (analogous to the gradient norm in the smooth case) should be $\nu_k^{-1/2} \xi_{\text{cp}}(\Delta_k; x_k, \nu_k)^{1/2}$ instead of $\nu_k^{-1} \xi_{\text{cp}}(\Delta_k; x_k, \nu_k)^{1/2}$. The first-order optimality condition in (3.13) then becomes

$$\nu_k^{-1/2} \xi_{\text{cp}}(\Delta_{\min}; x_k, \nu_k)^{1/2} \leq \epsilon \quad (0 < \epsilon < 1). \quad (12)$$

This changes the bound in (3.15) to

$$|S(\epsilon)| \leq \frac{(f+h)(x_0) - (f+h)_{\text{low}}}{\eta_1 \kappa_{\text{mdc}} \nu_{\min} \epsilon^2} = O(\epsilon^{-2}), \quad (13)$$

i.e., ν_{\min}^2 becomes ν_{\min} , but does not alter the order of complexity.

Similarly, the limits considered in Corollary 3.9 and Theorem 3.11 should be $\liminf_{k \rightarrow \infty} \nu_k^{-1/2} \xi_{\text{cp}}(\Delta_k; x_k, \nu_k)^{1/2}$ and $\lim_{k \rightarrow \infty} \nu_k^{-1/2} \xi_{\text{cp}}(\Delta_k; x_k, \nu_k)^{1/2}$, respectively.

In addition to the definition of $\xi(\sigma; x)$ in (6.4), we should have also defined

$$\xi_{\text{cp}}(\sigma; x) := f(x) + h(x) - (\varphi(\sigma; x) + \psi(\sigma; x)), \quad (14)$$

where

$$s \in \arg \min_s m(s; x, \sigma). \quad (15)$$

However, (6.6) is still valid.

After (6.10), by analogy with Algorithm 3.1, the stationarity measure should be $\sigma_k^{1/2} \xi_{\text{cp}}(\sigma_{\max}; x_k)^{1/2}$.

In Section 7, all our numerical experiments use $\xi_{\text{cp}}(\Delta_k; x_k, \nu_k)^{1/2}$ and $\xi(\sigma_k; x_k)^{1/2}$ as proxies for the stationarity measures. Even though the factors $\nu_k^{-1/2}$ and $\sigma_k^{1/2}$, respectively, are not included in the proxies, the measures remain valid as long as $\nu_{\min} > 0$ and $\sigma_{\max} < \infty$, which is the case under our assumptions.

References

- [1] A. Y. Aravkin, R. Baraldi, and D. Orban. [A proximal quasi-Newton trust-region method for nonsmooth regularized optimization](#). *SIAM J. Optim.*, 32(2):900–929, 2022.