

Solving algorithm NCL's subproblems: The need for interior methods

D. Ma, D. Orban, M. A. Saunders

G-2025-33

April 2025

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.

Citation suggérée : D. Ma, D. Orban, M. A. Saunders (Avril 2025). Solving algorithm NCL's subproblems: The need for interior methods, Rapport technique, Les Cahiers du GERAD G- 2025-33, GERAD, HEC Montréal, Canada.

Avant de citer ce rapport technique, veuillez visiter notre site Web (<https://www.gerad.ca/fr/papers/G-2025-33>) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

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.

Suggested citation: D. Ma, D. Orban, M. A. Saunders (April 2025). Solving algorithm NCL's subproblems: The need for interior methods, Technical report, Les Cahiers du GERAD G-2025-33, GERAD, HEC Montréal, Canada.

Before citing this technical report, please visit our website (<https://www.gerad.ca/en/papers/G-2025-33>) 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.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2025
– Bibliothèque et Archives Canada, 2025

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.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2025
– Library and Archives Canada, 2025

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

Solving algorithm NCL's subproblems: The need for interior methods

Ding Ma ^a

Dominique Orban ^b

Michael A. Saunders ^a

^a *Systems Optimization Laboratory, Stanford University, Huang Engineering Center, Stanford, 94305–4021, CA, USA*

^b *GERAD and Dept of Mathematics and Industrial Engineering, Montréal (Qc), Canada, H3C 3A7*

dingma@alumni.stanford.edu

dominique.orban@gerad.ca

saunders@stanford.edu

April 2025
Les Cahiers du GERAD
G–2025–33

Copyright © 2025 Ma, Orban, Saunders

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 : Algorithm NCL was devised to solve a class of large nonlinearly constrained optimization problems whose constraints do not satisfy LICQ at a solution. It is mathematically equivalent to the augmented Lagrangian algorithm LANCELOT, which solves a short sequence of bound-constrained subproblems BC_k and has no LICQ difficulties. NCL's equivalent subproblems NC_k are much bigger and must be solved by a nonlinear interior method (needing first and second derivatives). We study the KKT-type systems arising within nonlinear interior methods when they are applied to the NC_k subproblems. We find that the KKT systems can sometimes be reduced to smaller SQD systems (symmetric quasi-definite) and sometimes to even smaller SPD systems (symmetric positive definite). The smaller systems have proved suitable for GPU implementation within the interior solver MadNLP when it used by MadNCL to implement Algorithm NCL.

Keywords : Constrained optimization, interior methods, sparse matrix algebra

Acknowledgements: We further acknowledge the pioneering work of Conn, Gould, and Toint [1–3, 11] in their development of the large-scale nonlinear optimizer LANCELOT. Algorithm NCL would not exist without the much earlier development of LANCELOT.

Special thanks are due to Prof Ken Judd and Dr Che-Lin Su for formulating their optimal taxation policy model [9] and finding that existing optimization solvers were not effective because of the model's failure to satisfy LICQ at a solution. Algorithm NCL is the only reliable (and efficient) solver for these optimization problems, thanks to the success of nonlinear interior methods IPOPT and KNITRO (and now MadNLP) on subproblems NC_k .

Further thanks are owed to the authors of MadNLP and MadNCL, especially Alexis Montoison, François Pacaud, and Sungho Shin, for recognizing the value of systems (K2) and (K1) for enabling GPU computation on certain large and challenging problems.

1 Introduction

For large-scale linear and nonlinear optimization, today's implementations of interior methods [7, 8, 10, 18] have arisen from the pioneering work of Fiacco and McCormick [5, 6] on barrier methods, in the sense that inequality constraints have been handled then and now by log barrier functions. Our contribution here is a tribute to the research of Tamás Terlaky and colleagues on the mathematical properties of interior methods, primarily for linear optimization but with much relevance for nonlinear extensions.

We consider large smooth constrained optimization problems in the form

$$\begin{array}{ll} \text{NC} & \begin{array}{l} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \phi(x) \\ \text{subject to} \quad c(x) = 0, \quad \ell \leq x \leq u, \end{array} \end{array}$$

where $\phi(x)$ is a smooth scalar function and $c(x) \in \mathbb{R}^m$ is a vector of smooth linear or nonlinear functions. We assume that first and second derivatives are available.

If the constraints represent any linear or nonlinear inequalities, we assume that slack variables have already been included as part of x , and appropriate bounds are included in ℓ and u . Problem NC is general in this sense.

1.1 Notation

Dual variables for the constraints $c(x) = 0$ are denoted by y , and duals for the simplified bounds $x \geq 0$ in section 3 are denoted by z . The Jacobian for $c(x)$ is the $m \times n$ matrix $J(x)$. If the Hessians of $\phi(x)$ and $c_i(x)$ are $H_0(x)$ and $H_i(x)$, the Hessian of the Lagrangian $L(x, y, \rho) = \phi(x) - y^T c(x)$ is $H(x, y) = H_0(x) - \sum_i y_i H_i(x)$.

2 LANCELOT and Algorithm NCL

LANCELOT [2, 3, 11] is designed to solve large, smooth constrained optimization problems. For problem NC, LANCELOT solves a sequence of about 10 BCL (Bound-Constrained augmented Lagrangian) subproblems of the form

$$\begin{array}{ll} \text{BC}_k & \begin{array}{l} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \phi(x) - y_k^T c(x) + \frac{1}{2} \rho_k c(x)^T c(x) \\ \text{subject to} \quad \ell \leq x \leq u, \end{array} \end{array}$$

where y_k is an estimate of the dual variables for the nonlinear constraints $c(x) = 0$, and $\rho_k > 0$ is a penalty parameter. After BC_k is solved (perhaps approximately) to give a subproblem solution x_k^* , the size of $\|c(x_k^*)\|$ is used to define BC_{k+1} :

- If $\|c(x_k^*)\|$ is sufficiently small, stop with “Optimal solution found”.
- If $\|c(x_k^*)\| < \|c(x_{k-1}^*)\|$ sufficiently, update $y_{k+1} = y_k - \rho_k c(x_k^*)$ and keep $\rho_{k+1} = \rho_k$.
- Otherwise, keep $y_{k+1} = y_k$ and increase the penalty (say $\rho_{k+1} = 10\rho_k$).
- If the penalty is too large (say $\rho_{k+1} > 10^{10}$), stop with “The problem is infeasible”.

2.1 Algorithm NCL

Algorithm NCL mimics LANCELOT with only one change: subproblem BC_k is replaced by the equivalent larger subproblem

$$\begin{array}{ll} \text{NC}_k & \begin{array}{l} \underset{x \in \mathbb{R}^n, r \in \mathbb{R}^m}{\text{minimize}} \quad \phi(x) + y_k^T r + \frac{1}{2} \rho_k r^T r \\ \text{subject to} \quad c(x) + r = 0, \quad \ell \leq x \leq u. \end{array} \end{array}$$

Given a subproblem solution (x_k^*, r_k^*) , the choice between updating y_k or increasing ρ_k is based on $\|r_k^*\|$. We expect $\|r_k^*\| \rightarrow 0$, so that x_k^* is increasingly close to solving NC.

The active-set solvers CONOPT [4], MINOS [14], and SNOPT [21] are nominally applicable to NC_k . Their reduced-gradient algorithms would naturally choose r as basic variables, and the x variables would be either superbasic (free to move) or nonbasic (fixed at one of the bounds). However, this is inefficient on large problems unless most variables are on a bound at the subproblem solution x_k^* .

In contrast, interior methods welcome the extra variables r in NC_k because

- subproblem NC_k is always feasible;
- the Jacobian of $c(x) + r$ always has full row rank;
- the sparse-matrix methods used for each interior iteration are affected very little by the increased subproblem size.

These properties justify part two of our title.

2.2 About 10 subproblems

A virtue of LANCELOT inherited by NCL is the limited number of subproblems NC_k , regardless of problem size. A Julia implementation of NCL [13] confirmed the value of solving subproblems NC_k with decreasing feasibility and optimality tolerances (10^{-2} to 10^{-6}). The number of subproblems remained at around 10. This contrasts with SQP methods requiring hundreds or thousands of QP subproblems on large problems.

3 The linear system

For simplicity, we assume in this section that the bounds $\ell \leq x \leq u$ are simply $x \geq 0$. When a nonlinear primal-dual interior method such as IPOPT or KNITRO is applied to NC_k , each search direction is obtained from a linear system of the form

$$\begin{pmatrix} -(H + X^{-1}Z) & J^T \\ J & -\rho_k I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta r \\ \Delta y \end{pmatrix} = \begin{pmatrix} r_2 \\ r_3 \\ r_1 \end{pmatrix}, \quad (\text{K3})$$

where $X = \text{diag}(x)$, $Z = \text{diag}(z)$, and r_1, r_2, r_3 are residuals for the system of nonlinear equations being solved by the interior method. Although (K3) is a large system, Δr does not damage its sparsity, and both IPOPT and KNITRO have been efficient as solvers for NC_k on some large models of taxation policy [12].

For all NC_k , $\rho_k \geq 1$ (and ultimately $\rho_k \gg 1$), it is stable to eliminate Δr from (K3) to obtain

$$\begin{pmatrix} -(H + X^{-1}Z) & J^T \\ J & \frac{1}{\rho_k} I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \begin{pmatrix} r_2 \\ r_1 + \frac{r_3}{\rho_k} \end{pmatrix}, \quad \Delta r = \frac{1}{\rho_k}(\Delta y - r_3). \quad (\text{K2})$$

If H is positive semidefinite or definite, (K2) is SQD (symmetric quasi-definite) [22] and can be solved efficiently using sparse indefinite Cholesky-type LDL^T factorization, where D is diagonal but indefinite.

We can eliminate Δy from (K2) to give the smaller system

$$(H + X^{-1}Z + \rho_k J^T J) \Delta x = -r_2 + J^T(r_3 + \rho_k r_1), \quad \Delta y = r_3 + \rho_k(r_1 - J \Delta x), \quad (\text{K1})$$

as noted by Montoison et al. in connection with the solver MadNLP [17]. If H is SPD, system (K1) can often be solved by sparse LDL^T factorization, ideally using GPUs.

If J contains one or more rather dense rows, the system defining Δx in (K1) would also be rather dense. As noted by a referee, it may help to use the Sherman-Morrison-Woodbury formula to solve for Δx . This would require $H + X^{-1}Z$ to be sparse.

Alternatively, and again if $H + X^{-1}Z$ is sparse, it may be more efficient to eliminate Δx from (K2) to give

$$(I + \rho_k J(H + X^{-1}Z)^{-1} J^T) \Delta y = \rho_k r_1 + r_3, \quad (H + X^{-1}Z) \Delta x = J^T \Delta y - r_2. \quad (\text{K1}')$$

The system defining Δy in (K1') is smaller than the one defining Δx in (K1), but will be dense if J contains any dense columns.

3.1 MadNLP

MadNLP [17] is a Julia-based optimizer that implements a primal-dual interior method for large linear and nonlinear optimization problems. With system (K3) in mind, MadNLP has options for solving the reduced systems (K2) or (K1). Such reductions would require recoding of IPOPT and KNITRO (which is not likely to happen).

3.2 MadNCL

Algorithm NCL has been implemented as MadNCL [16] with MadNLP as the solver for subproblems NC_k . For convex problems, systems (K2) and (K1) can use a GPU implementation of sparse LDL^T factorization (indefinite and definite respectively). This allows MadNCL to solve subproblems NC_k many times faster than IPOPT or KNITRO, and thus provides a GPU solver for large examples of problem NC. MadNCL has already proved capable of solving very large SCOPF problems (Security Constrained Optimal Power Flow), showing that Algorithm NCL is able to solve problems involving MPEC constraints—another challenging class of problems that fail LICQ.

4 Tribute

Tamás Terlaky, with colleagues Cornelis Roos and Jean-Philippe Vial, pioneered the analysis of interior methods for linear optimization [19, 20]. We pay tribute to this innovative work here by studying the linear algebra of Algorithm NCL, an approach to nonlinear optimization that would not exist if linear and then nonlinear interior methods had not been invented.

In recent years as director of the Quantum Computing and Optimization Laboratory (QCOL) at Lehigh University, Tamás has turned his attention to a radically new and challenging subject: Quantum Computing Optimization (e.g., [15]). We can be sure that this research field is in the very best hands.

References

- [1] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Numer. Anal.*, 28:545–572, 1991.
- [2] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *LANCELOT: A Fortran Package for Large-scale Nonlinear Optimization (Release A)*. Lecture Notes in Computational Mathematics 17. Springer Verlag, Berlin, Heidelberg, New York, London, Paris and Tokyo, 1992.
- [3] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. MOS-SIAM Ser. Optim. 1. SIAM, Philadelphia, 2000.
- [4] CONOPT home page. <https://www.gams.com/products/conopt/>, accessed July 13, 2024.
- [5] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1968. Republished as [6].

- [6] A. V. Fiacco and G. P. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Classics in Applied Mathematics, Vol. 4. SIAM, Philadelphia, 1990. Republication of [5].
- [7] Gurobi optimization software. <http://www.gurobi.com>, accessed July 12, 2024.
- [8] IPOPT open source optimization software. <https://projects.coin-or.org/Ipopt>, accessed July 12, 2024.
- [9] K. L. Judd, D. Ma, M. A. Saunders, and C.-L. Su. Optimal income taxation with multidimensional taxpayer types. Working paper, Hoover Institution, Stanford University, 2017.
- [10] KNITRO optimization software. <https://www.artelys.com/solvers/knitro/>, accessed July 12, 2024.
- [11] LANCELOT optimization software. <https://github.com/ralna/LANCELOT>, accessed July 12, 2024.
- [12] D. Ma, K. L. Judd, D. Orban, and M. A. Saunders. Stabilized optimization via an NCL algorithm. In M. Al-Baali, Lucio Grandinetti, and Anton Purnama, editors, *Numerical Analysis and Optimization, NAO-IV, Muscat, Oman, January 2017*, volume 235 of Springer Proceedings in Mathematics and Statistics, pages 173–191. Springer International Publishing Switzerland, 2018.
- [13] D. Ma, D. Orban, and M. A. Saunders. Algorithm NCL for constrained optimization and its use of SQD systems. Presented at SIAM Conference on Linear Algebra, Paris, France, May 13–17, 2024, May 13–17, 2024. <https://www.siam.org/conferences/cm/conference/la24>.
- [14] MINOS sparse nonlinear optimization solver. https://www.gams.com/latest/docs/S_MINOS.html, accessed July 12, 2024.
- [15] Mohammadhossein Mohammadisiahroudi, Ramin Fakhimi, and Tamás Terlaky. Efficient use of quantum linear system algorithms in interior point methods for linear optimization. <https://arxiv.org/html/2205.01220>, 2024.
- [16] A. Montoison, F. Pacaud, M. A. Saunders, S. Shin, and D. Orban. MadNCL: GPU implementation of algorithm NCL. Working paper on Overleaf, 2024.
- [17] A. Montoison, F. Pacaud, S. Shin, et al. MadNLP: a solver for nonlinear programming. <https://github.com/MadNLP/MadNLP.jl>, 2024.
- [18] MOSEK optimization software. <http://mosek.com/>, accessed July 12, 2024.
- [19] C. Roos, T. Terlaky, and J.-Ph. Vial. Theory and Algorithms for Linear Optimization: An Interior Point Approach. Wiley, 1997.
- [20] C. Roos, T. Terlaky, and J.-Ph. Vial. Interior Point Methods for Linear Optimization. Springer, revised edition of Wiley 1997 edition, 2005.
- [21] SNOPT sparse nonlinear optimization solver. <http://ccom.ucsd.edu/~optimizers/solvers/snopt/>, accessed July 12, 2024.
- [22] R. J. Vanderbei. Symmetric quasi-definite matrices. SIAM J. Optim., 5:100–113, 1995.