European Journal of Operational Research 268 (2018) 13–24

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Continuous Optimization

# Efficient solution of quadratically constrained quadratic subproblems within the mesh adaptive direct search algorithm☆

Nadir Amaioua[a], Charles Audet[a], Andrew R. Conn[b], Sébastien Le Digabel[a,*]

[a] GERAD and Département de mathématiques et génie industriel, École Polytechnique de Montréal, C.P. 6079, Succ. Centre-ville, Montréal, Québec H3C 3A7, Canada

[b] Mathematical Sciences, IBM T J Watson Research Center, 1101 Kitchawan Rd. Yorktown Heights NY 10598, USA

ARTICLE INFO

*Article history:*
Received 9 December 2016
Accepted 25 October 2017
Available online 20 November 2017

*Keywords:*
Nonlinear programming
Derivative-free optimization
Quadratic programming
Trust-region subproblem
Mesh adaptive direct search

ABSTRACT

The mesh adaptive direct search algorithm (MADS) is an iterative method for constrained blackbox optimization problems. One of the optional MADS features is a versatile search step in which quadratic models are built leading to a series of quadratically constrained quadratic subproblems. This work explores different algorithms that exploit the structure of the quadratic models: the first one applies an $l_1$-exact penalty function, the second uses an augmented Lagrangian and the third one combines the former two, resulting in a new algorithm. It is notable that this latter approach is uniquely suitable for quadratically constrained quadratic problems. These methods are implemented within the NOMAD software package and their impact are assessed through computational experiments on 65 analytical test problems and 4 simulation-based engineering applications.



## 1. Introduction

We consider the following constrained optimization problem:

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & f(x) \\ \text{subject to} & c_j(x) \le 0, \quad j \in [\![1,m]\!] \end{array} \tag{1}$$

where $m$ is a positive integer, $\mathcal{X}$ is a subset of $\mathbb{R}^n$, $f$ and $(c_j)_{j\in[\![1,m]\!]}$ are real-valued functions, possibly evaluated by a computer simulation, seen as a blackbox with the following characteristics: function evaluations take a long time to compute, the simulation may fail for some input values, the derivatives are not available and their approximations may be unreliable, and/or too expensive. The set $\mathcal{X}$ is often defined by bound on the variables.

Derivative-free optimization (DFO, (Conn, Scheinberg, & Vicente, 2009)) algorithms are designed to handle this type of problem. DFO methods do not use or try to approximate derivatives of the problems. Instead, they either rely on a direct search approach which uses a discretization of the solution space and generates directions to test trial points, or they use model-based methods by constructing polynomial or other approximations to mimic the functions over some specified trust-region or as a surrogate for a line-search method. Recent surveys on blackbox and derivative-free optimization appear in Audet (2014), Boukouvala, Misener, and Floudas (2016).

The present work focuses on the mesh adaptive direct search algorithm (MADS) (Audet & Dennis, Jr., 2006) with quadratic models (Conn & Le Digabel, 2013). MADS principally relies on a pair of steps, called the search and the poll, to explore the space of variables and a third step to update its parameters. Both the search and the poll are complementary: the search allows local and global exploration while the poll is local and ensures convergence. We consider a model-based approach in the search step that has no impact on the theoretical convergence analysis of MADS, but improves its practical performance. At each iteration $k$, the search builds local quadratic models near the current iterate $x^k$ for the objective function $f$ and for each of the $m$ inequality constraints. This leads to a quadratically constrained quadratic subproblem over an $l_\infty$ norm trust-region:

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & f^k(x) \\ \text{subject to} & c_j^k \le 0, \qquad\qquad j \in [\![1,m]\!] \\ & \left\|x - x^k\right\|_\infty \le \Delta^k \end{array} \tag{2}$$

where the scalar $\Delta^k \ge 0$ defines the trust-region, $f^k$ is the quadratic model of the objective function near the current iterate $x^k$ and

☆ This work is supported by the NSERC CRD RDCPJ 490744-15 grant in collaboration with Hydro-Québec and Rio Tinto.

* Corresponding author.
*E-mail addresses:* nadir.amaioua@gerad.ca (N. Amaioua), sebastien.le.digabel@gerad.ca (S. Le Digabel).
*URL:* http://www.gerad.ca/Charles.Audet (C. Audet), http://researcher.watson.ibm.com/researcher/view.php?person=us-arconn (A.R. Conn), http://www.gerad.ca/Sebastien.Le.Digabel (S. Le Digabel)

https://doi.org/10.1016/j.ejor.2017.10.058
0377-2217/

the $(c_j^k)_{j\in[\![1,m]\!]}$ are the quadratic models of the $m$ inequality constraints. We prefer an $l_\infty$ trust region because it has the same orientation as a box determined by simple bounds.

In Conn and Le Digabel (2013), to solve Problem (1), a series of subproblems of the type (2) are created and each of them is solved with a new instance of MADS. The paper concludes by stating that the quadratic structure of the subproblems is not exploited and that a blackbox optimization solver is certainly not the appropriate choice to solve a quadratically constrained quadratic problem. The main purpose of this paper is to test dedicated algorithms to solve Problem (2). We choose two widely known methods from the literature: the $l_1$-exact penalty function and the augmented Lagrangian methods. A new methodological approach combining the strengths of both methods is introduced and called the $l_1$-augmented Lagrangian: it uses an $l_1$ penalty term instead of the squared one used in the standard augmented Lagrangian.

This paper is organized as follows. Section 2 describes the MADS algorithm and in particular, its mechanisms to build and use quadratic models. The section also summarizes several methods from the literature handling quadratic subproblems. Section 3 presents the $l_1$-exact penalty function algorithm and Section 4 gives a description of the augmented Lagrangian method. Section 5 introduces the new $l_1$-augmented Lagrangian method and describes implementation choices. Section 6 compares the results of the three algorithms on a set of 61 analytical and 4 simulation-based test problems and Section 7 concludes with recommendations.

## 2. Literature review

The MADS algorithm is an iterative method that uses a discretization of the solution space, called the mesh, to select and evaluate new trial points. Each iteration of MADS consists of two main steps: the search and poll, followed by a parameter update step.

The poll and update steps are critical to the convergence proof of MADS (Audet & Dennis, Jr., 2006), but the search is optional and more flexible: it can be omitted or defined by the user. A frequently used search strategy (Conn & Le Digabel, 2013) is to automatically construct quadratic models to try and find a promising trial point. If the search succeeds, i.e. the selected trial point improves upon the current iterate, then this trial point becomes the new iterate and the poll step is skipped. Moreover, the search direction can also be exploited elsewhere, for example to prioritize poll choices (see below). On the contrary, if the search fails, the poll step becomes mandatory. Other types of search step such as VNS (Audet, Béchard, & Le Digabel, 2008) and surrogate-based (Audet, Kokkolaras, Le Digabel, & Talgorn, 2017; Booker et al., 1999) are not the topic of the present paper. In this paper, we use quadratic models in the search step since our understanding, and in fact our computational experience, suggest that it is likely to be the best, or at least one of the best, approaches.

The poll is used to choose mesh points near the current iterate and to evaluate their objective and constraint values. On the one hand, if the poll fails to find a better solution, the update step will reduce the mesh size (the parameter that scales the space discretization) and the poll size (maximum distance allowed between a trial point and the current iterate) in order to concentrate near the current iterate. On the other hand, once a better solution is found, the poll step terminates and the update step increases the mesh size. The diagram in Fig. 1 represents a description of the MADS algorithm with a search step based on quadratic models.

The quadratic models constructed in the search step are also used in the poll step: the poll produces a list of trial mesh points, and instead of sending them directly to be evaluated, quadratic models of the objective function and of the constraints are used to sort the trial points, so that the most promising ones are evaluated first. This approach is used in conjunction with the opportunistic strategy: as soon as a better solution than the current iterate is found, the poll step stops without executing the simulation at the remaining trial points. Using quadratic models in both the search and poll steps greatly improves the MADS performance, as reported in Conn and Le Digabel (2013).

MADS relies on a cache structure, which stores all the evaluated points, to select an interpolation set in order to build the models. In the search step, the interpolation set is constructed by selecting all the points from the cache that are inside the ball centered around the current iterate with a radius equal to twice the poll size parameter. In the poll step, quadratic models are used to order trial points, and interpolation points from the cache are selected within a ball centered around the current iterate with a radius $2r$, where $r$ is the smallest radius of the ball centered in the current iterate and containing all the trial points (Conn & Le Digabel, 2013).

At each iteration of MADS, a quadratic subproblem of the form (2) is constructed and solved within the search step. The resulting solution is projected on the mesh to provide a starting point that satisfies the convergence requirements of MADS. The subproblem is similar to those arising in trust-region methods called trust-region subproblems. Since the early eighties, many algorithms have been developed specifically to solve this kind of quadratic problems. The Moré and Sorenson algorithm (MS) (Moré & Sorensen, 1983) is one of the first algorithms used specifically for quadratic problems subject to an ellipsoidal constraint. It is based on solving the optimality conditions via the Newton algorithm with backtracking. The downside of this algorithm is that it uses Cholesky factorizations that become expensive for large matrices, which is not an issue in the present research since no large matrices are involved. Later, the generalized Lanczos trust-region algorithm (GLTR) (Gould, Lucidi, & Toint, 1999) was implemented by using the MS algorithm on Krylov subspaces. However, GLTR could not handle hard cases (see chapter 7 in Conn, Gould, & Toint (2000)) of the trust-region subproblems. The same principle is applied by the sequential subspace method (SSM) (Hager, 2001) that creates four dimensional subspaces instead of using the Krylov subspaces. Even if the SSM algorithm handles the hard case of the trust-region subproblem, it still solves quadratic problems over a sphere. The Gould-Robinson and Thorne algorithm (Gould, Robinson, & Thorne, 2010) improved the MS algorithm by using some high dimensional polynomial approximations that allow the Newton method to converge in fewer iterations. Another algorithm, that treats specifically quadratic problems over a convex quadratic constraint, is the Fortin-Rendl and Wolkowicz algorithm (Fortin & Wolkowicz, 2004; Rendl & Wolkowicz, 1997) which rewrites the problem into an eigenvalue subproblem that can be handled by the Newton method combined with Armijo–Goldstein conditions. This algorithm is suitable for large-scale matrix problems.

All the algorithms above are used for a quadratic objective over a unique constraint defining the trust-region. In our case, Problem (2) is additionally constrained by $m$ quadratic constraints and, recently, one method was developed specifically for this kind of problems using an extension of the Fortin-Rendl and Wolkowicz algorithm (Pong & Wolkowicz, 2014). Solving Problem (2) can also be done by nonlinear optimization tools such as exact penalty functions and augmented Lagrangians. These latter two approaches are discussed further in the next two sections.

## 3. The $l_1$-exact penalty function ($l_1$EPF algorithm)

The $l_1$-exact penalty function starts by transforming Problem (2) into the following bound-constrained problem:

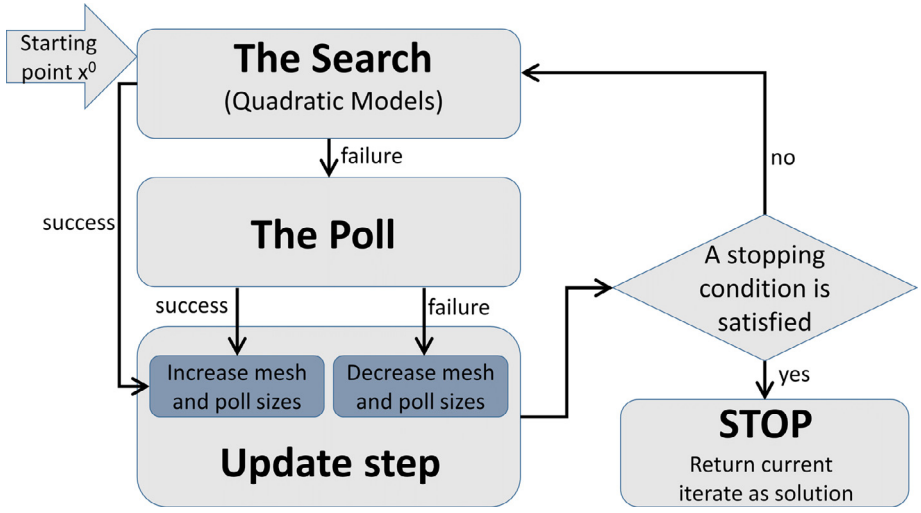
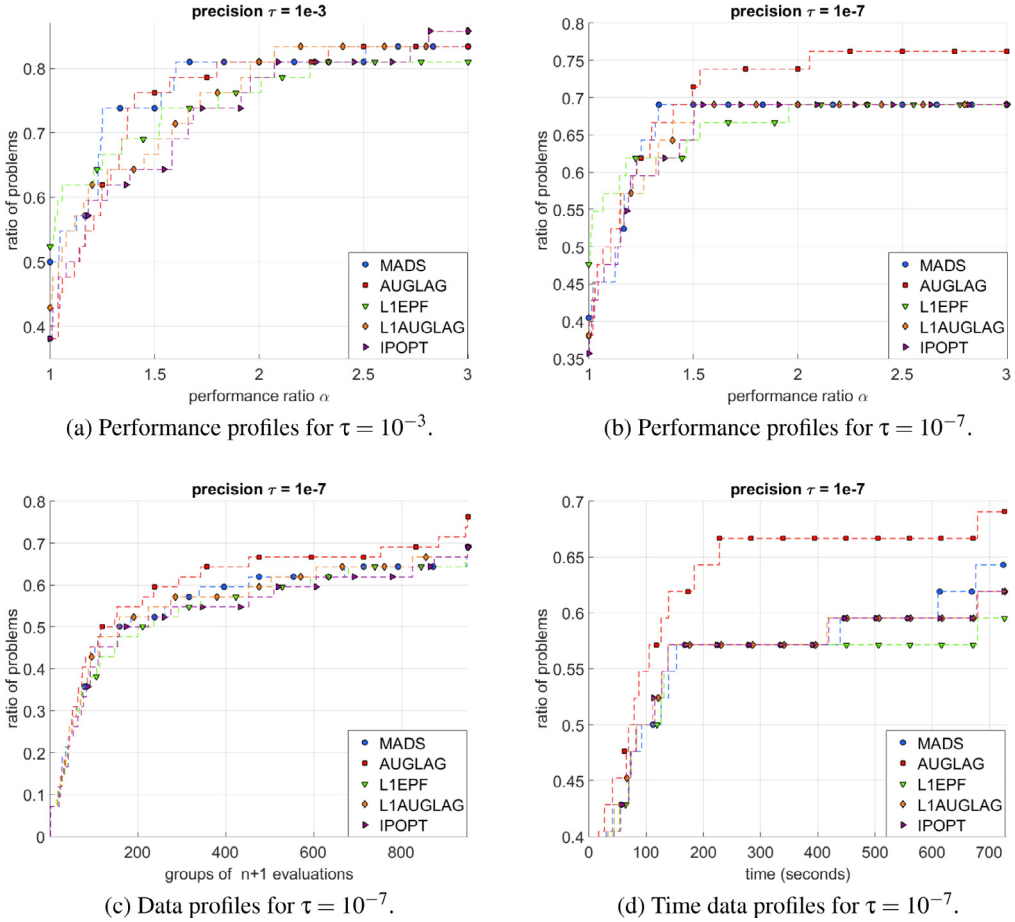



**Fig. 1.** An overview of the MADS algorithm with a search step based on quadratic models.




(a) Performance profiles for $\tau = 10^{-3}$.

(b) Performance profiles for $\tau = 10^{-7}$.

(c) Data profiles for $\tau = 10^{-7}$.

(d) Time data profiles for $\tau = 10^{-7}$.

**Fig. 2.** Performance and data profiles on the unconstrained test set.

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f^k(x) + \frac{1}{\mu} \sum_{j=1}^{m} \max\{0, c_j^k(x)\} \\ \text{subject to} \quad & \left\| x - x^k \right\|_\infty \leq \Delta^k \end{aligned} \tag{3}$$

where $\mu \in \mathbb{R}^+$ is the penalty coefficient. A very closely related method is given in Ablow and Brigham (1955). The $l_1$-exact penalty function was introduced by Zangwill (1965, 1967) and Pietrzykowski (1969) that, for $\mu$ sufficiently small, if the objective function and the constraints are continuously differentiable, any local minimum of a nonlinear problem would also be a minimizer of the $l_1$ penalty function of that problem. In Luenberger (1970) and Zangwill (1967), it is shown that, in the convex case, there exists a threshold $\mu_0$ such as when $\mu$ is smaller than $\mu_0$, a minimizer of the $l_1$ penalty function would also be a minimizer of the nonlinear problem. Charalambous (1978) generalizes this result under second order conditions. See also Conn (1973). In Coleman and Conn (1980), Coleman and Conn explore further optimality conditions of the $l_1$ penalty function. For our implementation, we concentrate on the approach provided by Coleman and Conn (1982a, 1982b).

The objective function of (3) is not differentiable on the boundary of the domain defined by $c^k(x) \leq 0$. We introduce two subsets of indices at some point $\tilde{x} \in \mathcal{X}$:

$$A_\epsilon = \left\{ j \in [\![1, m]\!] : |c_j^k(\tilde{x})| \leq \epsilon \right\} \text{ and } V_\epsilon = \left\{ j \in [\![1, m]\!] : \ c_j^k(\tilde{x}) \ > \epsilon \right\}. \tag{4}$$

The set $A_\epsilon$ contains the indices of $\epsilon$-active constraints at $\tilde{x}$, and $V_\epsilon$ groups all the indices of the $\epsilon$-violated constraints at $\tilde{x}$. The remaining constraints whose indices are in $\{j \in [\![1, m]\!] : c_j^k(\tilde{x}) \leq -\epsilon\}$ ($\epsilon$-satisfied) do not contribute locally to the objective function due to the $l_1$ term. Thus, for some neighborhood $\mathcal{N}(\tilde{x}) \subseteq \mathcal{X}$ of $\tilde{x}$, Problem (3) may be rewritten as:

$$\begin{array}{ll} \min\limits_{x \in \mathcal{N}(\tilde{x})} & p^k(x, \mu) + \frac{1}{\mu} \sum_{j \in A_\epsilon} \max\{0, c_j^k(x)\} \\ \text{subject to} & \left\| x - x^k \right\|_\infty \leq \Delta^k \end{array} \tag{5}$$

where the penalty function $p^k(x, \mu) := f^k(x) + \frac{1}{\mu} \sum_{j \in V_\epsilon} \max\{0, c_j^k(x)\}$ is differentiable at $x \in \mathcal{N}(\tilde{x})+$. The second part of the objective function of (5) is not locally differentiable. In order to overcome this difficulty, we would like to find a descent direction $h \in \mathbb{R}^n$ that would prevent any change in the $\epsilon$-active constraints up to the second order while minimizing the variation in the penalty function:

$$\begin{array}{ll} \min\limits_{h \in \mathbb{R}^n} & \nabla_x p^k(\tilde{x}, \mu)^\top h + \frac{1}{2} h^\top \nabla_{xx} p^k(\tilde{x}, \mu) h \\ \text{subject to} & \nabla c_j^k(\tilde{x})^\top h + \frac{1}{2} h^\top \nabla^2 c_j^k(\tilde{x}) h = 0, \qquad j \in A_\epsilon. \end{array} \tag{6}$$

Solving (6) approximately is detailed by Coleman and Conn (1982b) and this process circumvents the differentiability issues and explores a subspace where all the constraints are far from being satisfied. We call $h$ a range space minimization direction and it is used to reduce the penalty function while keeping the $\epsilon$-active constraints constant. That is why each time $\tilde{x} + h$ is close to a point $x^*$ that satisfies second-order sufficiency conditions, a null space minimization direction $v$ is also needed and is used to improve the action of the $\epsilon$-active constraints:

$$\phi_{A_\epsilon}^k(\tilde{x} + h + v) = 0, \tag{7}$$

with $\phi_{A_\epsilon}^k(x) = [c_j^k(x)]_{j \in A_\epsilon}$, the vector of $\epsilon$-active constraints values at $x$. Using a Taylor expansion up to the first order, the expression becomes:

$$\phi_{A_\epsilon}^k(\tilde{x} + h) + J_{A_\epsilon}^k(\tilde{x})^\top v \approx 0, \tag{8}$$

with $J_{A_\epsilon}^k(x) = [\nabla c_j^k(x)]_{j \in A_\epsilon}$, the matrix of $\epsilon$-active constraints gradients. Then $v$ can be computed by solving the linear system:

$$J_{A_\epsilon}^k(\tilde{x})^\top v = -\phi_{A_\epsilon}^k(\tilde{x} + h). \tag{9}$$

When $v$ is computed, the iterate is updated along the direction $h + v$ with a stepsize of 1 only if it presents a sufficient decrease. Algorithm 1 summarizes the inner iteration of the $l_1$EPF algorithm.

Algorithm 1 provides an overview of the methodology used to select a descent direction. The complete method for selecting descent direction relies on estimating the Lagrange multipliers to decide whether to take a vertical step or not. These estimates help also to include the curvature information of the constraints to approximate the Hessian matrix of the penalty function. See Coleman and Conn (1982b) for the complete description. A line-search algorithm tailored for piecewise differentiable functions is described in Coleman and Conn (1982b). The overall algorithm tries simply, at each iteration, to find a minimizer of the penalty function for a fixed value of $\mu$ whilst attempting to maintain the near active constraints near active, followed by a step that makes near active constraints more active (Algorithm 1). Then the algorithm proceeds to reduce $\mu$ if the selected minimizer is infeasible for (2) in order to put more weight on the constraints, if necessary. Algorithm 2 describes the outer iteration of the $l_1$EPF method.

**Algorithm 1:** $l_1$EPF algorithm (inner iteration) [17].

Initialization: $x$, $\epsilon \leftarrow 1$, $A_\epsilon$, $V_\epsilon$, $\tau \leftarrow 10^{-5}$
**if** $x$ is close to a point satisfying 1$st$ order conditions and $\left\| \phi_{A_\epsilon}^k(x) \right\| \leq \tau$ **then**
  Stop
**end if**
Compute $h$ (by approximately solving (6))
**if** $x + h$ is close to a point satisfying 1$st$ order conditions **then**
  Compute $v$ (by solving (9))
  **if** $x + h + v$ decreases $p$ sufficiently **then**
    $x \leftarrow x + h + v$
  **else**
    $\epsilon \leftarrow \frac{\epsilon}{2}$
    Update $A_\epsilon$ and $V_\epsilon$
    Compute $h$ (by approximately solving (6))
    Compute the step size $\beta$ (Line-search algorithm)
    $x \leftarrow x + \beta h$
  **end if**
**else**
  Compute the step size $\beta$ (Line-search algorithm)
  $x \leftarrow x + \beta h$
**end if**
Go to 2

**Algorithm 2:** $l_1$EPF algorithm (outer iteration) [17].

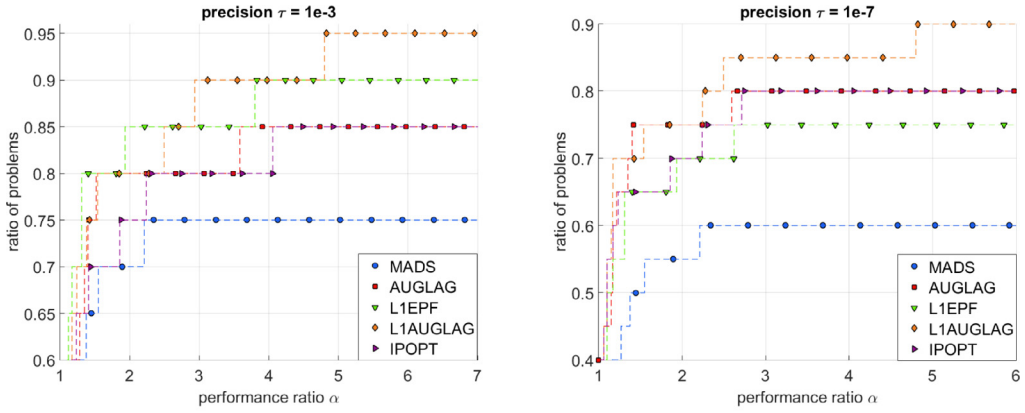
Initialization: $x$, $\mu \leftarrow 1$
**while** $x$ is not feasible **do**
  Minimize $p$ to find new iterate $x$ (Algorithm 1)
  Update: $\mu = \mu/10$
**end while**

## 4. Augmented Lagrangian (AugLag)

The augmented Lagrangian method was introduced by Powell (1969) and Hestenes (1969) in the late sixties. Since that time, the method was widely studied, improved and supported by convergence analyses (Griva, Nash, & Sofer, 2009; Nocedal & Wright, 1999) and different versions of the algorithm were implemented. A well-known implementation is the LANCELOT package (Conn, Gould, & Toint, 1992, 1996). When using slack variables in the implementation, it is possible to exploit the structure of these variables to increase the efficiency of the algorithm (Conn, Gould, & Toint, 1994; Conn, Vicente, & Visweswariah, 1999). An augmented Lagrangian for constrained blackbox problems is designed in Gramacy et al. (2016). The paper Audet, Le Digabel, and Peyrega (2016) introduces a derivative-free trust-region algorithm (DFTR) that uses the augmented Lagrangian approach to solve quadratically constrained quadratic subproblems and provides a method to update the trust-region radius. In the present work, we follow the approach described in Arreckx, Lambe, Martins, and Orban (2016) which is similar to the LANCELOT implementation.

First, slack variables $s \in \mathbb{R}^m$ are added to Problem (2):

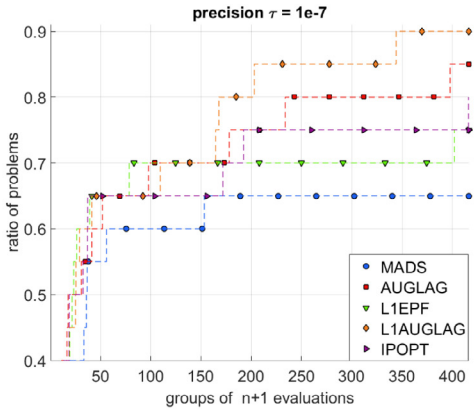
$$\begin{array}{lll} \min\limits_{x \in \mathcal{X},\ s \in \mathbb{R}^m} & f^k(x) & \\ \text{subject to} & c_j^k(x) + s_j = 0, & j \in [\![1, m]\!] \\ & \left\| x - x^k \right\|_\infty \leq \Delta^k, & \\ & s_j \geq 0, & j \in [\![1, m]\!] \end{array} \tag{10}$$



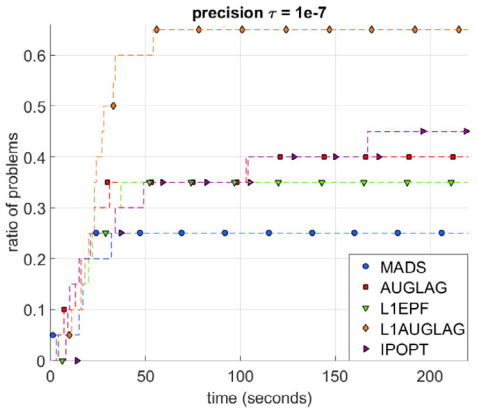

(a) Performance profiles for $\tau = 10^{-3}$.




(b) Performance profiles for $\tau = 10^{-7}$.




(c) Data profiles for $\tau = 10^{-7}$.




(d) Time data profiles for $\tau = 10^{-7}$.

**Fig. 3.** Performance and data profiles on the constrained test set.

and the bound constrained augmented Lagrangian is defined as follows:

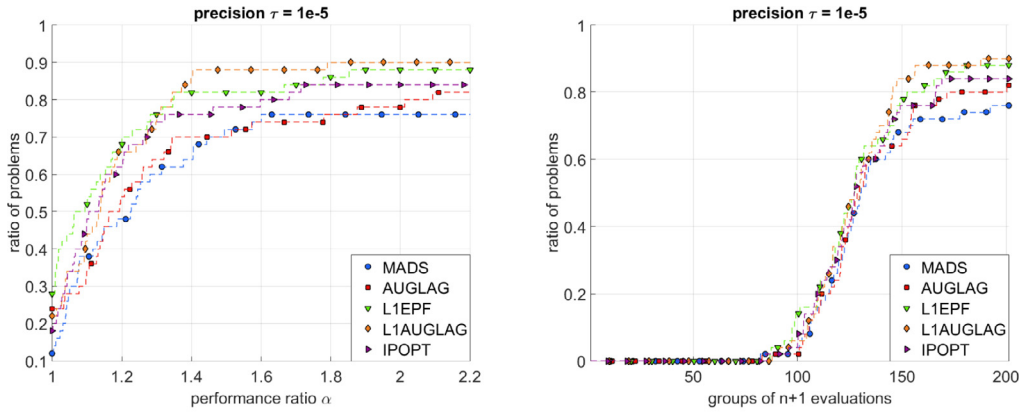
$$\begin{aligned} \min_{x\in\mathcal{X},\ s\in\mathbb{R}^m} \quad & \mathcal{L}_a^k(x,s,\lambda,\mu) \quad = \quad f^k(x) - \sum_{j=1}^{m}\lambda_j(c_j^k(x)+s_j) \\ & \qquad\qquad + \frac{1}{2\mu}\sum_{j=1}^{m}(c_j^k(x)+s_j)^2 \\ \text{subject to} \quad & \left\|x-x^k\right\|_\infty \quad \leq \quad \Delta^k, \\ & s_j \geq 0, \qquad\qquad\quad j \in [\![1,m]\!] \end{aligned} \tag{11}$$

where $\lambda \in \mathbb{R}^m$ and $\mu \in \mathbb{R}^+$. The augmented Lagrangian algorithm solves (11) at each iteration $l$ of the outer algorithm for fixed values $\lambda^l$ and $\mu^l$ to compute the next iterate $(x^{l+1}, s^{l+1})$. The index $l$ counts the outer iterations of the augmented Lagrangian method while the index $k$ characterizes the subproblem (2). The outer iteration proceeds to update one of the parameters $\lambda^l$ or $\mu^l$: when all the constraints values are below a given tolerance value, $\eta^l$, the iterate is judged locally feasible and the algorithm proceeds to update the multipliers $\lambda^l$. In the other case, the penalty parameter $\mu$ is reduced in order to give more importance to the constraints and try to satisfy them on the next iteration. Algorithm 3 gives an overview of the outer iteration of the augmented Lagrangian method where $\mathcal{L}^k(x,s,\lambda) = f^k(x) - \sum_{j=1}^m \lambda_j(c_j^k(x)+s_j)$ and $\mathcal{P}_{\mathcal{X}\times\mathbb{R}_+^m}(y)$ the projection of $y$ into the set $\mathcal{X}\times\mathbb{R}_+^m = \{(x,s) : x\in\mathcal{X}, s\in\mathbb{R}_+^m\}$.

To compute the next iterate in step 3 of Algorithm 3, a quadratic model is used to approximate (11) at each iteration $l$ of

**Algorithm 3:** Augmented Lagrangian to solve Problem (10): outer iteration [3].

1: Initializations: $l \leftarrow 0, \lambda^l \leftarrow 0, \mu^l \leftarrow 1.2, \eta^l \leftarrow 1$, $y^{l^\top} = [x^{l^\top} s^{l^\top}], \tau \leftarrow 10^{-5}, \omega^l \leftarrow 1$
2: **if** $\left\|y^l - \mathcal{P}_{\mathcal{X}\times\mathbb{R}_+^m}(y^l - \nabla_y\mathcal{L}(x^l,s^l,\lambda^l))\right\| \leq \tau$ and $\left\|c(x^l)+s^l\right\| \leq \tau$ **then**
3: Stop
4: **else**
5: Go to 7
6: **end if**
7: Compute the next iterate $(x^{l+1}, s^{l+1})$ by solving (11) for fixed values of $\lambda^l, \mu^l$ and $\omega^l$ (Algorithm 5)
8: **if** $\left\|c(x^{l+1})+s^{l+1}\right\|_\infty \leq \eta^l$ **then**
9: $\lambda^{l+1} \leftarrow \lambda^l - \frac{1}{\mu^l}(c^k(x^l)+s^l)$
10: $\mu^{l+1} \leftarrow \mu^l$
11: $\eta^{l+1} \leftarrow \eta^l(\mu^l)^{0.9}$
12: $\omega^{l+1} \leftarrow \omega^l\mu^{l+1}$
13: **else**
14: $\lambda^{l+1} \leftarrow \lambda^l$
15: $\mu_{l+1} \leftarrow \frac{\mu_l}{10}$
16: $\eta^{l+1} \leftarrow \frac{(\mu^{l+1})^{0.1}}{10}$
17: $\omega^{l+1} \leftarrow \mu^{l+1}$
18: **end if**
19: $l \leftarrow l+1$
20: Go to 2




(a) Performance profiles. (b) Data profiles.

**Fig. 4.** Performance and data profiles of the AIRCRAFT_RANGE problem.

the outer loop:

$$\begin{aligned} \min_{d^x\in\mathbb{R}^n,\ d^s\in\mathbb{R}^m} \quad & \nabla_x\mathcal{L}_a^k(x^l,s^l,\lambda^l,\mu^l)^\top d^x + \tfrac{1}{2}d^{x\top}\nabla_{xx}\mathcal{L}_a^k(x^l,s^l,\lambda^l,\mu^l)d^x \\ & +\nabla_s\mathcal{L}_a^k(x^l,s^l,\lambda^l,\mu^l)^\top d^s+\tfrac{1}{2}d^{s\top}\nabla_{ss}\mathcal{L}_a^k(x^l,s^l,\lambda^l,\mu^l)d^s \\ \text{subject to} \quad & \left\|x^l+d^x-x^k\right\|_\infty \le \Delta^k, \\ & s^l+d^s\ge 0, \\ & \|d^x\|_\infty \le \delta^l, \\ & \|d^s\|_\infty \le \delta^l \end{aligned} \tag{12}$$

with $\delta^l$ the trust-region of the model. To simplify, let $d\in\mathbb{R}^{n+m}$ be the concatenation of $d^x$ and $d^s$ ($d^\top=[d^{x\top}d^{s\top}]$) and $y\in\mathbb{R}^{n+m}$ the concatenation of $x$ and $s$ ($y^\top=[x^\top s^\top]$). Since the constraints of (12) represent the intersection of the box defined by the simple bounds with the box defined by the $l_\infty$ trust region, it is possible to redefine the feasible set as a box $\mathcal{B}$ with lower and upper bounds $l\in\mathbb{R}^{n+m}$ and $u\in\mathbb{R}^{n+m}$:

$$\mathcal{B} = \quad \{d\in\mathbb{R}^{n+m} : l\le d\le u\}. \tag{13}$$

Problem (12) is rewritten as:

$$\min_{d\in\mathcal{B}} \quad \nabla_y\mathcal{L}_a^k(y^l,\lambda^l,\mu^l)^\top d + \tfrac{1}{2}d^\top\nabla_{yy}\mathcal{L}_a^k(y^l,\lambda^l,\mu^l)d. \tag{14}$$

Solving (14) relies on an iterative method developed by Moré and Toraldo to treat box constrained quadratic problems (see Arreckx et al. (2016), Moré and Toraldo (1991)). This method was an option for LANCELOT B in the GALAHAD package (Gould, Orban, & Toint, 2003b). We chose this approach due to some preliminary results on unconstrained problems that favored this trust-region method over the line-search method used in the previous section. The Moré and Toraldo algorithm manipulates the active-set $\mathcal{A}(\tilde{d})=\{i\in[\![1,n+m]\!] : \tilde{d}_i=\ell_i \text{ or } \tilde{d}_i=u_i\}$ to look for a descent direction $p$ that would be used to update its current iterate $\tilde{d}$ via a line-search. In the first stage of the algorithm, a promising face containing $\tilde{d}$ is selected via a projected gradient method. A face of $\mathcal{B}$ is defined as a set of vectors for which active-set $\mathcal{A}(\tilde{d})$ coordinates coincide with $\tilde{d}$:

$$\mathcal{F} = \quad \{w\in\mathcal{B} : w_i=\tilde{d}_i \text{ for } i\in\mathcal{A}(\tilde{d})\}. \tag{15}$$

The second part of the Moré and Toraldo method concentrates on finding a descent direction $p$ on the selected face with a conjugate gradient algorithm and update $\tilde{d}$. Algorithm 4 summarizes the Moré and Toraldo method to compute the solution $d$ to (14):

Once a direction $d$ is computed, the inner iteration of the augmented Lagrangian method uses a typical trust-region approach: the ratio $r$ (defined in Step 4 of Algorithm 5 ) determines whether the model fits the augmented Lagrangian function well. If the ratio is close to 1, the model fits the function well and, in that case, the trust-region radius must increase. On the contrary, if the ratio is negative or close to zero, the trust-region must be reduced. The ratio $r$ also allows one to decide if the iterate $x$ gets updated or not. Algorithm 5 describes the inner iteration of the augmented Lagrangian method (where $\mathcal{L}^k(x,s,\lambda)=f^k(x)-\sum_{j=1}^m\lambda_j(c_j^k(x)+s_j)$ and $\mathcal{P}_{\mathcal{X}\times\mathbb{R}_+^m}(y)$ the projection of $y$ into the set $\mathcal{X}\times\mathbb{R}_+^m=\{(x,s) : x\in\mathcal{X}, s\in\mathbb{R}_+^m\}$).

**Algorithm 4:** algorithm PG/CG [47].

1: **while** no solution $d$ is found **do**
2: Use the projected gradient (PG) method to select a promising face of the feasible set $\mathcal{B}$
3: Use the conjugate gradient (CG) method to explore the selected face and find a descent direction $p$
4: Use a projected line-search alongside $p$ to choose the next iterate $d$
5: **end while**

**Algorithm 5:** Augmented Lagrangian: inner iteration [3].

1: Initializations: d, $\epsilon_1\leftarrow 0.1, \epsilon_2\leftarrow 0.9, \gamma_1\leftarrow 0.01, \gamma_2\leftarrow 100$
2: **if** $\left\|y^l-\mathcal{P}_{\mathcal{X}\times\mathbb{R}_+^m}(y^l-\nabla_y\mathcal{L}(x^l,s^l,\lambda^l))\right\|\le\omega^l$ **then**
3: Stop
4: **else**
5: Go to 7
6: **end if**
7: Solve (14) to find $d$
8: Compute the ratio $r=\dfrac{\mathcal{L}_a^k(y^l+d,\lambda^l,\mu^l)-\mathcal{L}_a^k(y^l,\lambda^l,\mu^l)}{\nabla\mathcal{L}_a^k(y^l,\lambda^l,\mu^l)^\top d+\frac{1}{2}d^\top\nabla^2\mathcal{L}_a^k(y^l,\lambda^l,\mu^l)d}$
9: **if** $r\ge\epsilon_1$ **then**
10: $x^l\leftarrow x^l+d$
11: **if** $r\ge\epsilon_2$ **then**
12: $\delta\leftarrow\gamma_2\max\{\|d\|_\infty,\delta\}$
13: **end if**
14: **else**
15: $\delta\leftarrow\gamma_1\|d\|_\infty$
16: **end if**
17: Go to 2

## 5. Augmented Lagrangian with $l_1$ penalty term ($l_1$AugLag)

The present section proposes a new method that combines the algorithms from the previous two sections. Since the constraints in Problem (2) are quadratic rather than general nonlinear, the augmented Lagrangian loses the quadratic structure of the subproblem due to the squared penalty term. The idea is to replace this part with a $l_1$ penalty term in order to get a piecewise quadratic function. We define the $l_1$-augmented Lagrangian problem in the following manner:

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & f^k(x) - \sum\limits_{j=1}^{m}\lambda_j c_j^k(x) + \dfrac{1}{\mu}\sum\limits_{j=1}^{m}\max\{0, c_j^k(x)\} \\ \text{subject to} & \left\|x - x^k\right\|_\infty \leq \Delta^k. \end{array} \tag{16}$$

By using the same definition as in Eq. (4) of $\epsilon$-active and $\epsilon$-violated sets for a point $\tilde{x} = x^l$ (at iteration $l$ of the outer approximation), it is possible to restrict the analysis of Problem (16) in a neighborhood $\mathcal{N}(x^l) \subseteq \mathcal{X}$ of $x^l$ (same approach as in (5)):

$$\begin{array}{ll} \min\limits_{x\in\mathcal{N}(x^l)} & D^k(x,\lambda,\mu) + \frac{1}{\mu}\sum_{j\in A_\epsilon}\max\{0, c_j^k(x)\} \\ \text{subject to} & \left\|x - x^k\right\|_\infty \leq \Delta^k \end{array} \tag{17}$$

where $D^k(x,\lambda,\mu) = f^k(x) - \sum_{j=1}^m \lambda_j c_j^k(x) + \frac{1}{\mu}\sum_{j\in V_\epsilon} c_j^k(x)$ is differentiable, but the second term involving $A_\epsilon$ is not. This is similar to the $l_1$-exact penalty approach: At each iteration $l$, we need to handle the differentiability issues by concentrating on computing a direction that minimizes the variation in the $l_1$-augmented Lagrangian expression and keeps the $\epsilon$-active constraints from varying for fixed values of $\lambda$ and $\mu$:

$$\begin{array}{ll} \min\limits_{h\in\mathbb{R}^n} & \nabla_x D^k(x^l,\lambda^l,\mu^l)^\top h + \frac{1}{2}h^\top \nabla_{xx} D^k(x^l,\lambda^l,\mu^l)h \\ \text{subject to} & \nabla c_j^k(x^l)^\top h + \frac{1}{2}h^\top \nabla^2 c_j^k(x^l)h = 0, \qquad j \in A_\epsilon. \end{array} \tag{18}$$

To further satisfy the $\epsilon$-active constraints, we can use (9) to compute a range space minimization direction $v$. Algorithm 1 still applies: we just have to compute the null space minimization direction $h$ in Steps 3 and 10 by solving (18) instead. To update $\lambda$ or $\mu$, we use a method similar to the outer iteration algorithm of the augmented Lagrangian, however we can keep updating the penalty parameter $\mu$ in exactly the same way, but the $\lambda$ update needs some adjustment. We can determine how to update $\lambda$ by using the optimality test:

$$\nabla_x \mathcal{L}^k(x^l,\lambda^l) = 0 \tag{19}$$

with $\mathcal{L}^k(x,\lambda) = f^k(x) - \sum_{j=1}^m \lambda_j c_j^k(x)$, the Lagrangian of Problem (2). So, at iteration $l$, if $x^{l+1}$ is a minimizer of (17), then there exists a vector $\bar{\lambda}$ for which:

$$\nabla f^k(x^{l+1}) - \sum_{j=1}^{m}\lambda_j \nabla c_j^k(x^{l+1}) + \frac{1}{\mu}\sum_{j\in V_\epsilon}\nabla c_j^k(x^{l+1}) - \sum_{j\in A_\epsilon}\bar{\lambda}_j \nabla c_j^k(x^{l+1}) = 0. \tag{20}$$

If we define $\lambda^{l+1}$ to be:

$$\lambda^{l+1} = \begin{cases} \lambda_j^l + \bar{\lambda}_j, & \text{if } j \in A_\epsilon \\ \lambda_j^l - \frac{1}{\mu}, & \text{if } j \in V_\epsilon \\ \lambda_j^l, & \text{if } j \notin A_\epsilon \cup V_\epsilon \end{cases} \tag{21}$$

then the optimality test (19) becomes:

$$\begin{aligned} \nabla_x \mathcal{L}^k(x^{l+1},\lambda^{l+1}) &= \nabla f^k(x^{l+1}) - \sum_{j=1}^{m}\lambda_j^{l+1}\nabla c_j^k(x^{l+1}) \\ &= \nabla f^k(x^{l+1}) - \sum_{j\notin A_\epsilon\cup V_\epsilon}\lambda_j^l \nabla c_j^k(x^{l+1}) \\ &\quad - \sum_{j\in V_\epsilon}(\lambda_j^l - \frac{1}{\mu})\nabla c_j^k(x^{l+1}) - \sum_{j\in A_\epsilon}(\lambda_j^l + \bar{\lambda}_j) \\ &\quad \nabla c_j^k(x^{l+1}) \\ &= \nabla f^k(x^{l+1}) - \sum_{j=1}^{m}\lambda_j \nabla c_j^k(x^{l+1}) + \frac{1}{\mu}\sum_{j\in V_\epsilon} \\ &\quad \nabla c_j^k(x^{l+1}) \\ &\quad - \sum_{j\in A_\epsilon}\bar{\lambda}_j \nabla c_j^k(x^{l+1}) \\ &= 0 \end{aligned} \tag{22}$$

which confirms our update method of $\lambda$. The only remaining issue is how to compute $\bar{\lambda}$. Eq. (20) is rewritten as:

$$\sum_{j\in A_\epsilon}\bar{\lambda}_j \nabla c_j^k(x^{l+1}) = \nabla f^k(x^{l+1}) - \sum_{j=1}^{m}\lambda_j \nabla c_j^k(x^{l+1}) + \frac{1}{\mu}\sum_{j\in V_\epsilon}\nabla c_j^k(x^{l+1}). \tag{23}$$

By using matrix notations, Eq. (23) becomes:

$$J_{A_\epsilon}^k \bar{\lambda} = \nabla D^k(x^{l+1},\lambda^{l+1},\mu^{l+1}) \tag{24}$$

where $J_{A_\epsilon}^k$ is the matrix of $\epsilon$-active constraint gradients defined in Section 3. By using a QR decomposition, there exists an orthogonal matrix $Q$ and an upper triangular matrix $U$ such that:

$$J_{A_\epsilon}^k = QR = [Q_1\ Q_2]\begin{bmatrix} U \\ 0 \end{bmatrix} = Q_1 U \tag{25}$$

where the number of columns of $Q_1$ is the same as the number of rows of $U$ and corresponds to the number of $\epsilon$-active constraints near $\tilde{x}$. We can use this to transform (24) into:

$$U\ \bar{\lambda} = Q_1^\top \nabla D^k(x^{l+1},\lambda^{l+1},\mu^{l+1}). \tag{26}$$

It is now easy to obtain $\bar{\lambda}$ by using a backward substitution algorithm. Algorithm 6 displays the outer iteration of the $l_1$AUGLAG method (where $\mathcal{L}^k(x,\lambda) = f^k(x) - \sum_{j=1}^m \lambda_j c_j^k(x)$ and $\mathcal{P}_\mathcal{X}(y)$ the projection of $y$ into the set $\mathcal{X}$).

## 6. Computational results

In this section, we introduce three sets of test problems for the comparison between five variants of NOMAD: each version uses a different approach to solve Subproblem (2). The results of this comparison are detailed in Sections 6.2, 6.3 and 6.4.

### 6.1. Test problems

We use three sets of optimization problems: the first set is composed of 42 unconstrained analytic problems, the second consists of 19 constrained analytic problems and the third contains 4 simulation-based constrained applications (constraints are also simulation-based): two MDO problems called AIRCRAFT_RANGE and SIMPLIFIED_WING, a structural engineering design problem of a welded beam called WELDED and a blackbox optimization problem (LOCKWOOD) which minimizes the cost of managing wells in the Lockwood Solvent Groundwater Plume Site (Montana) that prevents the contamination of the Yellowstone River. Table 1 groups the name, dimension and number of constraints of the 65 test problems.

We implemented the three methods of Sections 3, 4 and 5 into the NOMAD package (Le Digabel, 2011) (www.gerad.ca/nomad)

and compare five algorithmic variants. The first variant is denoted MADS and uses the current version of NOMAD (version 3.7) to solve Subproblem (2). The three variants $l_1$EPF, AUGLAG and $l_1$AUGLAG respectively use the $l_1$-exact penalty function, the augmented Lagrangian and the $l_1$-augmented Lagrangian to treat (2). Finally, the last variant uses the IPOPT package (Wächter & Biegler, 2006) which is software for nonlinear optimization. The tests were conducted on a Linux machine with an Intel(R) Core(TM) i7-2600 (3.40 GigaHertz) processor.

To compare these five methods, we rely on the Moré and Wild performance and data profiles (Moré & Wild, 2009) by considering the following formula as the main convergence test:

$$f(x_f) - f(x) \geq (1-\tau)(f(x_f) - f^*), \tag{27}$$

where $\tau$ is the convergence precision, $f^*$ is the best solution found by all solvers for a specific problem and a given budget (of either evaluations or time) and $x_f$ is the first feasible point of the problem found by any solver. If a solver cannot find a feasible solution, the convergence test will always fail which will favor all the solvers that find at least one feasible point. Performance profiles show the ratio of solved problems according to a performance ratio $\alpha$ defined as the upper bound of the number of evaluations needed to satisfy the convergence test. These profiles allow one to compare the relative performances of the algorithms. Data profiles, on the other hand, indicate which algorithms is best for a fixed number of function evaluations: they show the ratio of solved problems to groups of $n+1$ function evaluations.

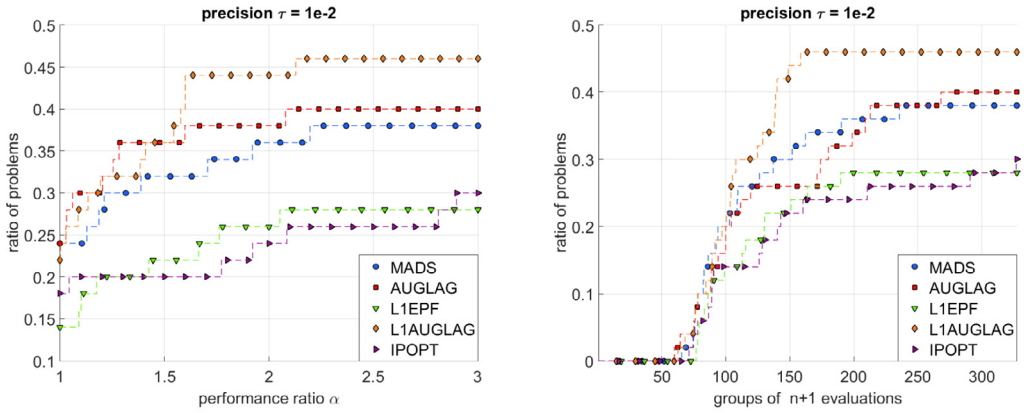
### 6.2. Unconstrained test set

It may seem strange that we are testing unconstrained problems when we compare methods relying on multipliers and penalty terms, nevertheless, in the absence of constraints, our methods reduce to their inner iterations: the $l_1$EPF and $l_1$AUGLAG become an iterative line-search method that uses matrix decompositions to select the descent direction, while the AUGLAG method solves a trust-region subproblem with a combination of projected and conjugate gradient methods. We thought it was important to keep the unconstrained test set since the results shown in this section lead us to recommend different approaches for the constrained and unconstrained cases.

Fig. 2(a) groups performance profiles for the unconstrained set with $\tau = 10^{-3}$. It is difficult to set a winner apart from this graph: for example, IPOPT starts slowly, but emerges on top for a performance ratio $\alpha = 3$. Fig. 2(b) and (c) represent performance and data profiles for the first set of problems with a precision $\tau = 10^{-7}$. AUGLAG clearly outperforms the other methods in this case.

**Table 1**
Description of the test problems.

| Name | $n$ | $m$ | Source | Name | $n$ | $m$ | Source | Name | $n$ | $m$ | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Unconstrained | | | | | | | | Constrained | | | |
| ACKLEY | 10 | 0 | Hedar (0000) | POWELLSG | 12 | 0 | Gould, Orban, and Toint (2003a) | CRESCENT10 | 10 | 2 | Audet and Dennis (2009) |
| ARWHEAD | 10 | 0 | Gould et al. (2003a) | POWELLSG | 20 | 0 | Gould et al. (2003a) | DISK10 | 10 | 1 | Audet and Dennis (2009) |
| ARWHEAD | 20 | 0 | Gould et al. (2003a) | RADAR | 7 | 0 | Mladenović, Petrović, Kovačević-Vujčić, and Čangalović (2003) | G1 | 13 | 9 | Michalewicz and Schoenauer (1996) |
| BDQRTIC | 10 | 0 | Gould et al. (2003a) | RANA | 2 | 0 | Jamil and Yang (2013) | G2 | 10 | 2 | Michalewicz and Schoenauer (1996) |
| BDQRTIC | 20 | 0 | Gould et al. (2003a) | RASTRIGIN | 2 | 0 | Hedar (0000) | G2 | 20 | 2 | Michalewicz and Schoenauer (1996) |
| BIGGS6 | 6 | 0 | Gould et al. (2003a) | RHEOLOGY | 3 | 0 | Audet and Hare (2017) | G4 | 5 | 6 | Michalewicz and Schoenauer (1996) |
| BRANIN | 2 | 0 | Hedar (0000) | ROSENBROCK | 2 | 0 | Jamil and Yang (2013) | G6 | 2 | 2 | Michalewicz and Schoenauer (1996) |
| BROWNAL | 10 | 0 | Gould et al. (2003a) | SCHWEFEL | 2 | 0 | Schwefel (1981) | G7 | 10 | 8 | Michalewicz and Schoenauer (1996) |
| BROWNAL | 20 | 0 | Gould et al. (2003a) | SHOR | 5 | 0 | Lukšan and Vlček (2000) | G8 | 2 | 2 | Michalewicz and Schoenauer (1996) |
| DIFF2 | 2 | 0 | Conn and Le Digabel (2013) | SROSENBR | 10 | 0 | Gould et al. (2003a) | G9 | 7 | 4 | Michalewicz and Schoenauer (1996) |
| ELATTAR | 6 | 0 | Lukšan and Vlček (2000) | SROSENBR | 20 | 0 | Gould et al. (2003a) | G10 | 8 | 6 | Michalewicz and Schoenauer (1996) |
| EVD61 | 6 | 0 | Lukšan and Vlček (2000) | TREFETHEN | 2 | 0 | Jamil and Yang (2013) | HS44 | 4 | 6 | Hock and Schittkowski (1981) |
| FILTER | 9 | 0 | Lukšan and Vlček (2000) | TRIDIA | 10 | 0 | Gould et al. (2003a) | HS64 | 3 | 1 | Hock and Schittkowski (1981) |
| GRIEWANK | 2 | 0 | Hedar (0000) | TRIDIA | 20 | 0 | Gould et al. (2003a) | HS93 | 6 | 2 | Hock and Schittkowski (1981) |
| GRIEWANK | 10 | 0 | Hedar (0000) | VARDIM | 10 | 0 | Gould et al. (2003a) | HS108 | 9 | 13 | Hock and Schittkowski (1981) |
| HS78 | 5 | 0 | Lukšan and Vlček (2000) | VARDIM | 20 | 0 | Gould et al. (2003a) | HS114 | 9 | 6 | Lukšan and Vlček (2000) |
| OSBORNE2 | 11 | 0 | Lukšan and Vlček (2000) | WATSON | 12 | 0 | Jamil and Yang (2013) | MAD6 | 5 | 7 | Lukšan and Vlček (2000) |
| PBC1 | 5 | 0 | Lukšan and Vlček (2000) | WONG1 | 7 | 0 | Lukšan and Vlček (2000) | PENTAGON | 6 | 15 | Lukšan and Vlček (2000) |
| PENALTY1 | 10 | 0 | Gould et al. (2003a) | WONG2 | 10 | 0 | Lukšan and Vlček (2000) | SNAKE | 2 | 2 | Audet and Dennis (2009) |
| PENALTY1 | 20 | 0 | Gould et al. (2003a) | WOODS | 12 | 0 | Gould et al. (2003a) | | | | |
| POLAK2 | 10 | 0 | Lukšan and Vlček (2000) | WOODS | 20 | 0 | Gould et al. (2003a) | | | | |
| Simulation-based MDO applications | | | | | | | | | | | |
| AIRCRAFT_RANGE | 10 | 10 | Agte, Sobieszczanski-Sobieski, and Sandusky (1999); Sobieszczanski-Sobieski, Agte, and Sandusky, Jr. (1998) | SIMPLIFIED_WING | 7 | 3 | Tribes, Dubé, and Trépanier (2005) | WELDED | 4 | 6 | Garg (2014) |
| LOCKWOOD | 6 | 4 | Kannan and Wild (2012); Matott, Rabideau, and Craig (2006) | | | | | | | | |




(a) Performance profiles. (b) Data profiles.

**Fig. 5.** Performance and data profiles of the LOCKWOOD problem.

**Algorithm 6:** $l_1$-augmented Lagrangian: outer iteration.

1: Initializations: $l \leftarrow 0, \lambda^l \leftarrow 0, \mu^l \leftarrow 1, \eta^l \leftarrow 1,$
$x^l, \tau \leftarrow 10^{-5}, \omega^l \leftarrow 1$
2: **if** $\left\| x^l - \mathcal{P}_{\mathcal{X}}(x^l - \nabla_x \mathcal{L}(x^l, \lambda^l)) \right\| \leq \tau$ **then**
3: Stop
4: **else**
5: Go to 7
6: **end if**
7: Compute the next iterate $x^{l+1}$ by minimizing (17) for fixed values
of $\lambda^l$, $\mu^l$ and $\omega^l$
8: **if** $c_j^k(x^{l+1}) \leq \eta^l$ for all $j \in V_\epsilon$ **then**
9: $\lambda^{l+1} \leftarrow \begin{cases} \lambda_j^l + \bar{\lambda_j} & j \in A_\epsilon \\ \lambda_j^l - \frac{1}{\mu} & j \in V_\epsilon \\ \lambda_j^l & j \notin A_\epsilon \cup V_\epsilon \end{cases}$
10: $\mu^{l+1} \leftarrow \mu^l$
11: $\eta^{l+1} \leftarrow \eta^l (\mu^l)^{0.9}$
12: $\omega^{l+1} \leftarrow \omega^l \mu^{l+1}$
13: **else**
14: $\lambda^{l+1} \leftarrow \lambda^l$
15: $\mu_{l+1} \leftarrow \frac{\mu_l}{10}$
16: $\eta^{l+1} \leftarrow \frac{(\mu^{l+1})^{0.1}}{10}$
17: $\omega^{l+1} \leftarrow \mu^{l+1}$
18: **end if**
19: $l \leftarrow l + 1$
20: Go to 2

Fig. 2(d) groups time data profiles: they show the ratio of solved problems with precision $\tau = 10^{-7}$ for a given time budget. This figure takes into account the effort deployed to solve the quadratic subproblems and confirms the superiority of the Moré and Toraldo method for our unconstrained test set. We conclude that a trust-region is preferable to a line-search method to solve the unconstrained subproblems that arises in MADS when using quadratic models.

Even if the $l_1$EPF and $l_1$AUGLAG reduce to the same inner iteration, we can see that there are differences in the profiles because these algorithms choose different starting points for the subproblem. This implementation choice was done after several preliminary tests on constrained problems to select the best starting points of the subproblem for our three methods. This choice was kept for the unconstrained case. Still, the variations between $l_1$EPF and $l_1$AUGLAG are small and the two methods seem equivalent.

### 6.3. Constrained test set

The problems of the second set are constrained. Fig. 3 shows performance and data profiles for precisions $\tau = 10^{-3}$ and $\tau = 10^{-7}$. For this set of problems, all the new methods outperform MADS, especially $l_1$AUGLAG which gives the best results and solves almost 95% of the problems while the MADS-based NOMAD solves only 75% when $\tau = 10^{-7}$ as illustrated in the performance profiles of Fig. 3(a).

Fig. 3(d) represents time data profiles for the second set of test problems: using IPOPT to solve subproblems seems to be efficient timewise, but still, the figure confirms that $l_1$AUGLAG is the best approach for the constrained case.

### 6.4. Simulation-based applications

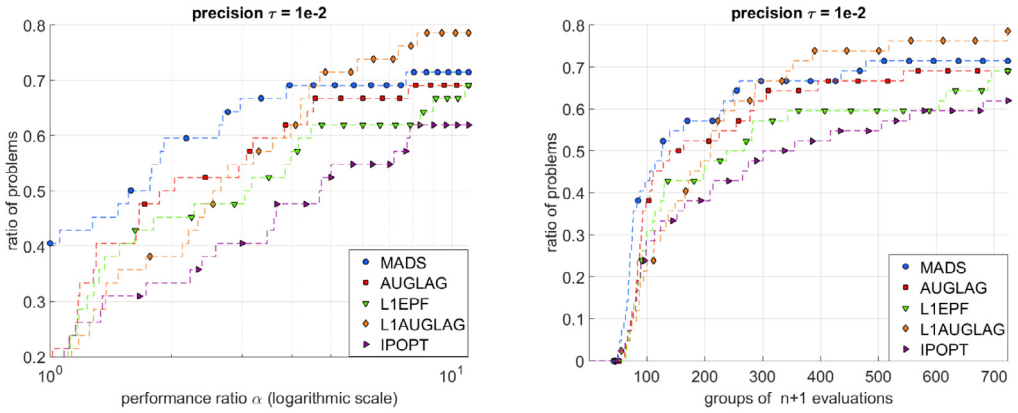
For the last set, we create 50 different instances for each of the four applications: for a given problem, a Latin hypercube sampling generates 50 starting points and each of these instances is solved by the five variants of NOMAD. Some of these starting points are feasible, others are not. These simulated-based problems tend to be timewise computationally expensive: the 750 tests of AIRCRAFT_RANGE, SIMPLIFIED_WING and WELDED ran for over five days and the 250 runs of LOCKWOOD, took almost two weeks of execution time.

For each application, we draw data and performance profiles to compare the different methods (we consider that all the instances are different problems). It is important to carefully select an appropriate value of the precision parameter $\tau$ for each problem since, on one hand, if $\tau$ is too small, the profiles tends to pile up with a small proportion of problems solved. On the other hand, if $\tau$ is too large, the curves converge to 100% problem solved quickly. For clarity, we chose values of $\tau$ as the power of 10 that spreads out as much as possible the profiles.

In Figs. 4 and 5, we can immediately see that $l_1$AUGLAG is the best approach for the AIRCRAFT_RANGE and LOCKWOOD problems. But, while $l_1$EPF and IPOPT seem to give good results on the AIRCRAFT_RANGE application (they are even overlapping $l_1$AUGLAG for $\alpha \leq 1.3$), AUGLAG and MADS are more efficient than $l_1$EPF and IPOPT on the LOCKWOOD problem.

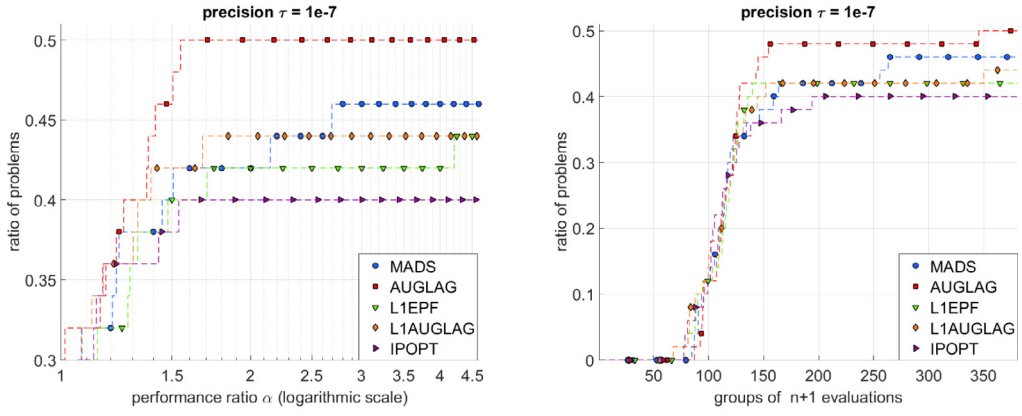
In Fig. 6, MADS shows some good results when $\alpha \leq 4.5$ and the budget is limited to 300 $(n+1)$ evaluations. When we allow more evaluation budget, $l_1$AUGLAG can almost reach 80% of problems solved.

In Fig. 7, the AUGLAG approach outperforms the rest of the algorithms with IPOPT coming last again. we are not sure why




(a) Performance profiles. (b) Data profiles.

**Fig. 6.** Performance and data profiles of the SIMPLIFIED_WING problem.




(a) Performance profiles. (b) Data profiles.

**Fig. 7.** Performance and data profiles of the WELDED problem.

AUGLAG wins here but we thought it would be useful to show that $l_1$AUGLAG does not always win.

In all the previous figures of the simulation-based problems, IPOPT performs poorly compared to the other methods, except for the instances of AIRCRAFT_RANGE. To recommend an alternative for MADS among the three proposed methods, we build performance and data profiles with $\tau = 10^{-2}$ and $\tau = 10^{-5}$ for the combined 200 instances of the four simulation-based applications.

In Fig. 8, $l_1$AUGLAG seems to be the clear winner. As suspected, IPOPT is not a viable option to solve the subproblem, while the three remaining approaches, MADS, $l_1$EPF and AUGLAG, are competitive with each other on the simulation-based problems. This was predictable since IPOPT is used for general nonlinear optimization while methods such as the $l_1$AUGLAG have an advantage when solving quadratically constrained quadratic problems.

## 7. Discussion

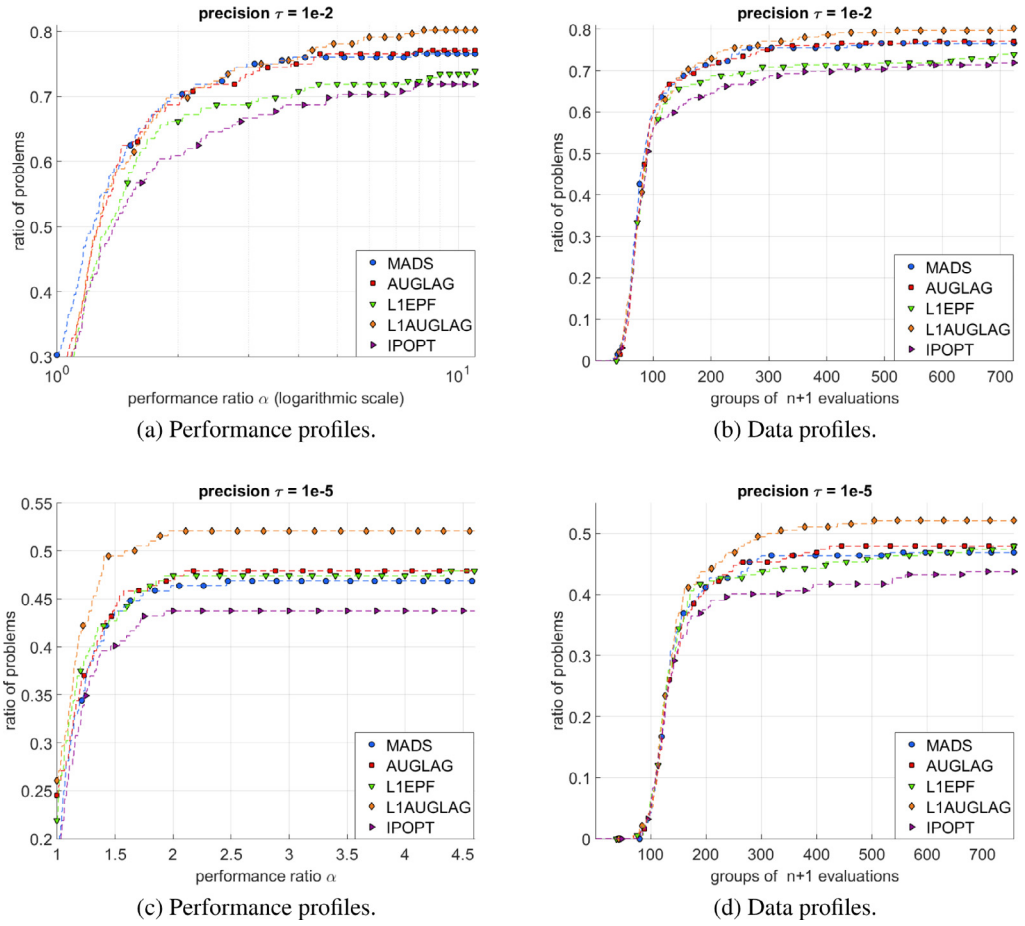
This work details the implementation of two existing nonlinear optimization methods ($l_1$EPF and AUGLAG) and the development of a new algorithm called the $l_1$-augmented Lagrangian method ($l_1$AUGLAG). These methods are used within a derivative-free optimization framework to solve quadratically constrained quadratic subproblems that arise when using quadratic models of the objective function and of the constraints. It would be possible to use the implemented methods within algorithms that rely on treating subproblems to solve general nonlinear problems with derivatives. Another avenue would be to use the augmented Lagrangian method to transform a DFO problem with general equality constraints into an unconstrained problem that MADS can deal with. And this approach could be incorporated into the broader multiobjective framework (Audet, Savard, & Zghal, 2010).

This paper shows that using a dedicated algorithm to solve quadratic subproblems improves the overall performance of MADS with quadratic models. The results show that, in the unconstrained case, using a trust-region approach (inner iteration of the augmented Lagrangian based on the Moré and Toraldo method) yields better result than a line-search method (which is the case of the inner iteration for both the $l_1$-exact penalty function and the $l_1$-augmented Lagrangian).

The $l_1$AUGLAG approach gives the best results in the constrained case as, we believe, it combines the strengths of both AUGLAG and $l_1$EPF methods: the Lagrange multipliers terms improve the computation performances and the $l_1$ penalty term allows Problem (16) to have a piecewise quadratic structure.

In all cases, there is at least one method that improves over MADS in regards of both results and time and this is the main goal of the paper. The computational results show also that at least one of the three suggested methods is performing better than IPOPT which validates the choice of implementing a dedicated method for NOMAD.

We conclude with recommendations for the NOMAD software package. For solving quadratic subproblems in the search step, NOMAD should use the Moré and Toraldo augmented Lagrangian algorithm in the unconstrained case and the $l_1$-augmented Lagrangian method for constrained problems. Finally, our implementation of the $l_1$-augmented Lagrangian handles inequalities directly.




(a) Performance profiles. (b) Data profiles.

(c) Performance profiles. (d) Data profiles.

**Fig. 8.** Performance and data profiles for all 200 instances of the simulation-based applications.

It would be interesting to see how adding slacks as in the $l_2$-augmented Lagrangian and treating all constraints as equalities would compare with our current implementation.

## Acknowledgments

We would like to thank the editor and two anonymous referees for their constructive remarks and comments.

## References

Ablow, C., & Brigham, G. (1955). An analog solution of programming problems. *Journal of the Operations Research Society of America, 3*(4), 388–394. doi:10.1287/opre.3.4.388.

Agte, J., Sobieszczanski-Sobieski, J., & Sandusky, R. (1999). Supersonic business jet design through bilevel integrated system synthesis. In *Proceedings of the world aviation conference*. San Francisco, CA: MCB University Press, Bradford, UK. vol. SAE Paper No. 1999-01-5622.

Arreckx, S., Lambe, A., Martins, J., & Orban, D. (2016). A matrix-free augmented lagrangian algorithm with application to large-scale structural design optimization. *Optimization and Engineering, 17*(2), 359–384. doi:10.1007/s11081-015-9287-9.

Audet, C. (2014). A survey on direct search methods for blackbox optimization and their applications. In P. Pardalos, & T. Rassias (Eds.), *Mathematics without boundaries: Surveys in interdisciplinary research. chapter 2* (pp. 31–56). Springer. http://www.springer.com/mathematics/analysis/book/978-1-4939-1123-3.

Audet, C., Béchard, V., & Le Digabel, S. (2008). Nonsmooth optimization through mesh adaptive direct search and variable neighborhood search. *Journal of Global Optimization, 41*(2), 299–318. doi:10.1007/s10898-007-9234-1.

Audet, C., & Dennis, J., Jr. (2006). Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization, 17*(1), 188–217. doi:10.1137/040603371.

Audet, C., & Dennis, J., Jr. (2009). A progressive barrier for derivative-free nonlinear programming. *SIAM Journal on Optimization, 20*(1), 445–472. doi:10.1137/070692662.

Audet, C., & Hare, W. (2017). Derivative-free and blackbox optimization. *Springer series in operations research and financial engineering*. Springer International Publishing. doi:10.1007/978-3-319-68913-5. http://www.springer.com/us/book/9783319689128; in press

Audet, C., Kokkolaras, M., Le Digabel, S., & Talgorn, B. (2017). Order-based error for managing ensembles of surrogates in derivative-free optimization. *Technical Report G-2016-36*. Les cahiers du GERAD. To appear in *Journal of Global Optimization*. doi:10.1007/s10898-017-0574-1.

Audet, C., Le Digabel, S., & Peyrega, M. (2016). A derivative-free trust-region augmented Lagrangian algorithm. *Technical Report G-2016-53*. Les cahiers du GERAD.

Audet, C., Savard, G., & Zghal, W. (2010). A mesh adaptive direct search algorithm for multiobjective optimization. *European Journal of Operational Research, 204*(3), 545–556. doi:10.1016/j.ejor.2009.11.010.

Booker, A., Dennis, J., Jr., Frank, P., Serafini, D., Torczon, V., & Trosset, M. (1999). A rigorous framework for optimization of expensive functions by surrogates. *Structural and Multidisciplinary Optimization, 17*(1), 1–13. doi:10.1007/BF01197708.

Boukouvala, F., Misener, R., & Floudas, C. (2016). Global optimization advances in mixed-integer nonlinear programming, MINLP, and constrained derivative-free optimization, CDFO. *European Journal of Operational Research, 252*(3), 701–727. doi:10.1016/j.ejor.2015.12.018.

Charalambous, C. (1978). A lower bound for the controlling parameters of the exact penalty functions. *Mathematical programming, 15*(1), 278–290. doi:10.1007/BF01609033.

Coleman, T., & Conn, A. (1980). Second-order conditions for an exact penalty function. *Mathematical Programming, 19*(1), 178–185. doi:10.1007/BF01581639.

Coleman, T., & Conn, A. (1982a). Nonlinear programming via an exact penalty function: Asymptotic analysis. *Mathematical programming, 24*(1), 123–136. doi:10.1007/BF01585100.

Coleman, T., & Conn, A. (1982b). Nonlinear programming via an exact penalty function: Global analysis. *Mathematical Programming, 24*(1), 137–161. doi:10.1007/BF01585101.

Conn, A. (1973). Constrained optimization using a nondifferentiable penalty function. *SIAM Journal on Numerical Analysis, 10*(4), 760–784. doi:10.1137/0710063.

Conn, A., Gould, N., & Toint, P. (1992). *LANCELOT: A Fortran package for large-scale nonlinear optimization (release A)*. New-York: Springer.

Conn, A., Gould, N., & Toint, P. (1994). A note on exploiting structure when using slack variables. *Mathematical Programming, 67*(1), 89–97. doi:10.1007/BF01582214.

Conn, A., Gould, N., & Toint, P. (1996). Numerical experiments with the LANCELOT package (release A) for large-scale nonlinear optimization. *Mathematical Programming, 73*(1), 73–110. doi:10.1007/BF02592099.

Conn, A., Gould, N., & Toint, P. (2000). *Trust region methods*. SIAM. doi:10.1137/1.9780898719857.

Conn, A., & Le Digabel, S. (2013). Use of quadratic models with mesh-adaptive direct search for constrained black box optimization. *Optimization Methods and Software, 28*(1), 139–158. doi:10.1080/10556788.2011.623162.

Conn, A., Scheinberg, K., & Vicente, L. (2009). Introduction to derivative-free optimization. *MOS-SIAM series on optimization*. Philadelphia: SIAM. doi:10.1137/1.9780898718768.

Conn, A., Vicente, L., & Visweswariah, C. (1999). Two-step algorithms for nonlinear optimization with structured applications. *SIAM Journal on Optimization, 9*(4), 924–947. doi:10.1137/S1052623498334396.

Fortin, C., & Wolkowicz, H. (2004). The trust region subproblem and semidefinite programming. *Optimization Methods and Software, 19*(1), 41–67. doi:10.1080/10556780410001647186.

Garg, H. (2014). Solving structural engineering design optimization problems using an artificial bee colony algorithm. *Journal of Industrial and Management Optimization, 10*(3), 777–794. doi:10.3934/jimo.2014.10.777.

Gould, N., Lucidi, S., & Toint, P. (1999). Solving the trust-region subproblem using the Lanczos method. *SIAM Journal on Optimization, 9*(2), 504–525. doi:10.1137/S1052623497322735.

Gould, N., Orban, D., & Toint, P. (2003a). CUTEr (and SifDec): A constrained and unconstrained testing environment, revisited. *ACM Transactions on Mathematical Software, 29*(4), 373–394. doi:10.1145/962437.962439.

Gould, N., Orban, D., & Toint, P. (2003b). Galahad, a library of thread-safe fortran 90 packages for large-scale nonlinear optimization. *ACM Transactions on Mathematical Software, 29*(4), 353–372. doi:10.1145/962437.962438.

Gould, N., Robinson, D., & Thorne, H. (2010). On solving trust-region and other regularised subproblems in optimization. *Mathematical Programming Computation, 2*(1), 21–57. doi:10.1007/s12532-010-0011-7.

Gramacy, R., Gray, G., Le Digabel, S., Lee, H., Ranjan, P., Wells, G., & Wild, S. (2016). Modeling an augmented lagrangian for blackbox constrained optimization. *Technometrics, 58*(1), 1–11. doi:10.1080/00401706.2015.1014065.

Griva, I., Nash, S., & Sofer, A. (2009). *Linear and nonlinear optimization*. Society for Industrial and Applied Mathematics.

Hager, W. (2001). Minimizing a quadratic over a sphere. *SIAM Journal on Optimization, 12*(1), 188–208. doi:10.1137/S1052623499356071.

Hedar, A.-R. Global optimization test problems. http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar_files/TestGO.htm. [last accessed on 20 October 2017] http://goo.gl/0vxil.

Hestenes, M. (1969). Multiplier and gradient methods. *Journal of Optimization Theory and Applications, 4*(5), 303–320. doi:10.1007/BF00927673.

Hock, W., & Schittkowski, K. (1981). Test examples for nonlinear programming codes. *Lecture notes in economics and mathematical systems*: vol. 187. Berlin, Germany: Springer.

Jamil, M., & Yang, X.-S. (2013). A literature survey of benchmark functions for global optimisation problems. *International Journal of Mathematical Modelling and Numerical Optimisation, 4*(2), 150–194. doi:10.1504/IJMMNO.2013.055204.

Kannan, A., & Wild, S. (2012). Benefits of deeper analysis in simulation-based groundwater optimization problems. In *Proceedings of the XIX international conference on computational methods in water resources (CMWR 2012)*.

Le Digabel, S. (2011). Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm. *ACM Transactions on Mathematical Software, 37*(4), 44:1–44:15. doi:10.1145/1916461.1916468.

Luenberger, D. (1970). Control problems with kinks. *IEEE Transactions on Automatic Control, 15*(5), 570–575.

Lukšan, L., & Vlček, J. (2000). Test problems for nonsmooth unconstrained and linearly constrained optimization. *Technical Report V-798*. ICS AS CR.

Matott, L., Rabideau, A., & Craig, J. (2006). Pump-and-treat optimization using analytic element method flow models. *Advances in Water Resources, 29*(5), 760–775. doi:10.1016/j.advwatres.2005.07.009.

Michalewicz, Z., & Schoenauer, M. (1996). Evolutionary algorithms for constrained parameter optimization problems. *Evolutionary computation, 4*(1), 1–32. doi:10.1162/evco.1996.4.1.1.

Mladenović, N., Petrović, J., Kovačević-Vujčić, V., & Čangalović, M. (2003). Solving spread spectrum radar polyphase code design problem by tabu search and variable neighbourhood search. *European Journal of Operational Research, 151*(2), 389–399. doi:10.1016/S0377-2217(02)00833-0.

Moré, J., & Sorensen, D. (1983). Computing a trust region step. *SIAM Journal on Scientific Computing, 4*(3), 553–572. doi:10.1137/0904038.

Moré, J., & Toraldo, G. (1991). On the solution of large quadratic programming problems with bound constraints. *SIAM Journal on Optimization, 1*(1), 93–113. doi:10.1137/0801008.

Moré, J., & Wild, S. (2009). Benchmarking derivative-free optimization algorithms. *SIAM Journal on Optimization, 20*(1), 172–191. doi:10.1137/080724083.

Nocedal, J., & Wright, S. (1999). Numerical optimization. *Springer series in operations research*. New York: Springer.

Pong, T., & Wolkowicz, H. (2014). The generalized trust region subproblem. *Computational Optimization and Applications, 58*(2), 273–322. doi:10.1007/s10589-013-9635-7.

Powell, M. (1969). A method for nonlinear constraints in minimization problems. In R. Fletcher (Ed.), *Optimization* (pp. 283–298). New York: Academic Press. http://www.ams.org/mathscinet-getitem?mr=42:7284

Rendl, F., & Wolkowicz, H. (1997). A semidefinite framework for trust region subproblems with applications to large scale minimization. *Mathematical Programming, 77*(1), 273–299. doi:10.1007/BF02614438.

Schwefel, H.-P. (1981). *Numerical optimization of computer models*. New York, NY, USA: John Wiley & Sons, Inc.

Sobieszczanski-Sobieski, J., Agte, J., & Sandusky, R., Jr. (1998). Bi-level integrated system synthesis (BLISS). *Technical Report NASA/TM-1998-208715*. NASA, Langley Research Center.

Pietrzykowski, T. (1969). An exact potential method for constrained maxima. *SIAM Journal on numerical analysis, 6*(2), 299–304.

Tribes, C., Dubé, J.-F., & Trépanier, J.-Y. (2005). Decomposition of multidisciplinary optimization problems: formulations and application to a simplified wing design. *Engineering Optimization, 37*(8), 775–796. doi:10.1080/03052150500289305.

Wächter, A., & Biegler, L. T. (2006). On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming, 106*(1), 25–57. doi:10.1007/s10107-004-0559-y.

Zangwill, W. (1965). Nonlinear programming by sequential unconstrained maximization. *Technical Report Working Paper 131*. University of California, Berkeley.

Zangwill, W. (1967). Non-linear programming via penalty functions. *Management science, 13*(5), 344–358. doi:10.1287/mnsc.13.5.344.