A Simple First-Order Algorithm for Full-Rank Equality Constrained Optimization

S. Gratton* and Ph. L. Toint†

17 X 2025

Abstract

A very simple first-order algorithm is proposed for solving nonlinear optimization problems with nonlinear equality constraints. This algorithm adaptively selects steps in the plane tangent to the constraints or steps that reduce infeasibility, without using a merit function or filter. The tangent steps are based on the AdaGrad method for unconstrained minimization. The objective function is never evaluated by the algorithm, making it suitable for noisy problems. Its worst-case evaluation complexity is analyzed, yielding a global convergence rate in $\mathcal{O}(1/\sqrt{k})$, which matches the best known rate of first-order methods for unconstrained problems. Numerical experiments are presented suggesting that the performance of the algorithm is comparable to that of first-order methods for unconstrained problems, and that its reliability is remarkably stable in the presence of noise on the gradient.

Keywords: Equality constrained optimization, objective-function-free optimization (OFFO), first-order methods, AdaGrad, evaluation complexity.

1 Introduction

Solving nonlinear problems with nonlinear equality constraints is a central part of more general constrained optimization, and the purpose of this paper is to contribute to the subject by proposing a very simple algorithm for the case where the constraints' Jacobian is full-rank. We introduce a method drawing its inspiration both from the "trust-funnel" approaches [17, 8] and the objective-function free (OFFO) first-order algorithms that have become very popular because of their successful use in deep learning applications. Trust-funnel methods themselves exploit the older idea of decomposing the step computed within an optimization algorithm for constrained problems into a "tangential" and a "normal" components. Advocated in particular by Byrd and Omojokun [24] for sequential quadratic programming, this last idea is to consider that a useful minimization step should, on the one hand, improve the value of the objective function without deteriorating the constraints's violation too much (this is the tangential step, because it can be interpreted as a step (nearly) tangent to the constraint's surface) and, on the other hand, improve feasibility (the normal step, mostly orthogonal to the constraints' surface). This technique has been successfully used by many algorithms, and in particular by trust-funnel ones. In this case, the two types of steps are treated separately and a choice between normal and tangential steps is made at every iteration by comparing constraint violation (primal feasibility) with some dual optimality measure. Our new proposed algorithm uses a similar mechanism, although considerably simplified. Our second source of inspiration is the thriving field of first-order methods for unconstrained problems, and, more specifically, the well-known and well understood AdaGrad [10, 29, 9, 19] algorithm. Such methods are attractive because they avoid computing the value

*Université de Toulouse, INP, IRIT, Toulouse, France. Email: serge.gratton@enseeiht.fr. Work partially supported by 3IA Artificial and Natural Intelligence Toulouse Institute (ANITI), French "Investing for the Future - PIA3" program under the Grant agreement ANR-19-PI3A-0004"

†NAXYS, University of Namur, Namur, Belgium. Email: philippe.toint@unamur.be

of the objective function entirely (they only rely on gradients). This feature makes them quite robust in the presence of noise on the gradient, as is commonly the case when its evaluation involves subsampling. Again, our new method exploits this OFFO[1] strategy to achieve robustness. In effect, it consists of a simplified trust-funnel method in which the tangential step is computed using the AdaGrad algorithm.

Other OFFO methods for constrained optimization have been proposed. The stochastic algorithm proposed in [1] uses a combined tangential and normal step for a quadratic model of a merit function. A (nontrivial) stochastic analysis of its convergence (albeit not its worst-case complexity) is provided. It also allows the use of second-order information, at the cost of a somewhat complicated algorithm. [7] extends this proposal to handle inequality constraints and provides an updated convergence analysis. The algorithm of [11] is more similar to ours, but differs in significant aspects. Although it uses tangential and normal steps, it does so at every iteration. The step in the Jacobian's nullspace is computed by a standard trust-region technique and a special rule is provided to update its radius. It also relies on an explicit penalty parameter with is updated in the course of the algorithm. Convergence of the algorithm is also analyzed in a stochastic setting. The method of [4] is inspired by filter methods (another class of methods for nonlinear optimization [13, 14, 12]) and also alternates between steps which attempts to reduce constraints' infeasibilities and "optimality" steps improving the objective function value, albeit ignoring the presence of constraints in the second case. Good numerical results are reported on deep learning problems, but no convergence theory or complexity analysis is provided[2].

In view of these comments, we summarize our contributions as follows.

1. In order to solve problem (2.1), we propose an "adaptive' switching" algorithm, dubbed ADSWITCH, which can be viewed as a simplified trust-funnel method. ADSWITCH uses the purely OFFO first-order AdaGrad algorithm in the plane tangent to the constraints.

2. We provide a complete worst-case complexity analysis for the case where the objective's gradient and the constraints are deterministic, under assumptions similar to those used in [1, 7].

3. We illustrate the practical behaviour of the method on a subset of the CUTEst problems [15] as provided by the S2MPJ environment [21].

The paper is organized as follows. The ADSWITCH algorithm is introduced in Section 2 together with the necessary concepts and notation. Section 3 contains the relevant complexity analysis. Numerical illustration is finally proposed in Section 4 while some conclusions and perspectives are discussed in Section 5.

Notations: In what follows, $\|\cdot\|$ denotes the Euclidean norm on the relevant space, $\sigma_{\min}(M)$ is the smallest singular value of the matrix M.

2 An adaptative switching algorithm

We propose solving the smooth equality constrained problem

$$\min_{x\in\mathbb{R}^n} f(x) \qquad \text{such that} \qquad c(x) = 0, \tag{2.1}$$

using the purely first-order ADSWITCH method, which we now describe after establishing some notation. In (2.1), f is a smooth function from $\mathbb{R}^n$ into $\mathbb{R}$ and c is a smooth function from $\mathbb{R}^n$ into $\mathbb{R}^m$ $(m \leq n)$. For a given vector x, we assume that we can compute the gradient $g(x) = \nabla_x f(x)$, as well as the (exact) Jacobian of the constraints at x

$$J(x) = \nabla c(x) \in \mathbb{R}^{m\times n}.$$

[1] Objective-Function-Free Optimization.

[2] Since the optimality steps ignore the constraints' geometry, it seems possible that their effect might significantly decrease feasibility and that cycling could occur.

Assuming $J(x)$ has full rank, we may also compute the orthogonal projection onto its nullspace by

$$P_T(x) \stackrel{\text{def}}{=} I - J(x)^T\big(J(x)J(x)^T\big)^{-1}J(x).$$

Given this projection, the projected gradient is then

$$g_T(x) \stackrel{\text{def}}{=} P_T(x)\, g(x) \tag{2.2}$$

We may also define the (exact) least-squares Lagrange multiplier $\widehat{\lambda}(x)$ by

$$\big(J(x)J(x)^T\big)\,\widehat{\lambda}(x) \;=\; -\,J(x)\,\nabla f(x), \tag{2.3}$$

the standard Lagrangian

$$L(x,\lambda) = f(x) + \lambda^T c(x) \tag{2.4}$$

and, for ρ fixed, the associated "augmented-Lagrangian-like" Lyapunov function

$$\psi_\rho(x,\lambda) \stackrel{\text{def}}{=} L(x,\lambda) + \rho\|c(x)\|, \tag{2.5}$$

which we often abbreviate for $\lambda = \widehat{\lambda}(x)$, as

$$\psi(x) \stackrel{\text{def}}{=} \psi_\rho\big(x, \widehat{\lambda}(x)\big). \tag{2.6}$$

The definitions imply the important properties that

$$g_T(x) = \nabla f(x) + J(x)^T\widehat{\lambda}(x) \quad \text{and} \quad J(x)\,g_T(x) = 0, \tag{2.7}$$

and also that, for λ fixed to $\widehat{\lambda}(x)$,

$$\nabla_x L\big(x,\lambda\big) = g_T(x). \tag{2.8}$$

By construction, we also have that

$$c(x)^T J(x) g_T(x) = 0. \tag{2.9}$$

We are now in position to specify ADSWITCH in detail on the following page. As announced in the introduction, the algorithm uses two types of steps: "tangential steps" attempt to reduce the objective-function value (without evaluating it!) in the nullspace of the constraints, while "normal steps" aim at reducing infeasibility. Which type of step is used at a given iteration depends on the respective values of the tangential step and constraint violation.

- The tangential step is the standard AdaGrad-norm [29] step in the nullspace of the Jacobian.
- The statement of the procedure to find the normal step in Step 3 may seem abstract because κ_{nrm} may not be known, but, fortunately, there exist several standard techniques to achieve (2.15)-(2.16).

  A first technique is to use a standard Armijo backtracking linesearch along the steepest-descent direction $-J_k^T c_k$ (see Lemma 3.3). It is well-known that (2.16) holds in this case with $\kappa_{\text{nrm}} = 1/L_{JTc}$ (see [3, Lemma 2.2.1], for instance). More generally, any "gradient related" direction d_k, that is any d_k satisfying (2.15) and $d_k^T J_k^T c_k \le -\kappa_{\text{grl}}\|J_k^T c_k\|^2$ for some $\kappa_{\text{grl}} > 0$, can be used to initiate a backtracking linesearch.

  A second possibility follows along this line and is to select a positive definite matrix B_k with bounded condition number, to define $d_k = -B_k^{-1}J_k^T c_k$ and then to perform a backtracking linesearch along d_k (see [5, Section 10.3.1] for a proof that d_k is a gradient-related direction), in which case a similar result holds. For instance, using $B_k = J_k^T J_k + \delta I_n$ for some $\delta \ge 0$ yields the regularized Gauss-Newton step

$$s_{Nk} = -\gamma_k(J_k^T J_k + \delta I_n)^{-1} J_k^T c_k = -\gamma_k J_k^T(J_k J_k^T + \delta I_m)^{-1} c_k \tag{2.18}$$

Algorithm 2.1: ADSWITCH

Step 0: Initialization. A starting point x_0 is given, together with constants $\omega > 0$, $\theta > 1$, $\beta, \eta, \varsigma \in (0, 1]$. Set $\Gamma_0 = 0$ and $k = 0$.

Step 1: Evaluations. Evaluate $c_k = c(x_k)$, $J_k = J(x_k)$ and $g_k = \nabla_x f(x_k)$. Set $g_{T,k} = P_T(x_k) g_k$,

$$\Gamma_k^+ = \Gamma_k + \|g_{T,k}\|^2, \tag{2.10}$$

and

$$\alpha_{T,k} = \frac{\eta}{\sqrt{\Gamma_k^+ + \varsigma}} \tag{2.11}$$

Step 2: Tangential step: If

$$\|c_k\| \le \beta \alpha_{T,k} \|g_{T,k}\| \tag{2.12}$$

then set

$$x_{k+1} = x_k - \alpha_{T,k} g_{T,k} \tag{2.13}$$

and

$$\Gamma_{k+1} = \Gamma_k^+, \tag{2.14}$$

and go to Step 4.

Step 3: Normal step: Otherwise, find a step $s_{N,k}$ in the range space of J_k^T such that

$$\|s_{N,k}\| \le \theta \|c_k\| \tag{2.15}$$

and there exists $\kappa_{\rm nrm} > 0$ independent of k such that

$$\tfrac{1}{2}\|c(x_k + s_{N,k})\|^2 \le \tfrac{1}{2}\|c_k\|^2 - \kappa_{\rm nrm} \|J_k^T c_k\|^2, \tag{2.16}$$

and set

$$x_{k+1} = x_k + s_{N,k}. \tag{2.17}$$

Step 4: Loop. Increment k by one and go to Step 2.

where $\gamma_k > 0$ is chosen by backtracking as large as possible to ensure (2.15) and (2.16). (Note that, for $\delta > 0$, $J_k^T J_k + \delta I_n$ has bounded condition number when J_k is bounded. Moreover this is also the case for $\delta = 0$ when J_k is full-rank.)

Suitable variants of trust-region steps are also possible [5, Section 10.3.2], themselves involving several subvariants such as the truncated conjugate gradient [26, 27], CGLS [22] or LSQR [25]. Because iterates generated by these methods are linear combinations of the gradients of $\frac{1}{2}\|c_k + J_k s\|^2$, this guarantees that the final step belongs to the range of J_k^T.

- We have mentioned an explicit formula for the projection $P_T(x)$, but it can also be computed using a rank-revealing QR factorization of J_k^T, the last columns of Q then providing a basis for the nullspace of J_k. In this case, if Q_k and R_k are the computed factors, then

$$g_{T,k} = Q_{k,[1:n,m+1:n]} Q_{k,[1:n,m+1:n]}^T g_k$$

and the Newton step (2.18) is given by

$$s_{N,k} = \begin{cases} -Q_{k,[1:n,1:r]} R_{k,[1:r,1:r]}^{-T} c_k & \text{if} \quad \delta = 0, \\ -Q_{k,[1:n,1:m]} R_{k,[1:m,1:m]} (R_{k,[1:m,1:m]}^T R_{k,[1:m,1:m]} + \delta I_m)^{-1} c_k & \text{otherwise.} \end{cases} \tag{2.19}$$

- Observe that the ADSWITCH algorithm does not involve any explicit merit function or filter to control overall progress, but rely on the simple switching condition (2.12) (hence its name). As we will see below, an implicit merit function (a modified augmented-Lagrangian with suitably chosen parameters) is however crucial for our theoretical argument.

- ADSWITCH differs in several aspects from the trust-funnel methods of [17, 8]. The first is that it avoids evaluating the objective function in order to improve its reliability in the presence of noise. The second is that it also avoids estimating Lagrange multipliers, but at the cost of computing the exact projection $P_T(\cdot)$. The third is that the trust-region approach, which is explicit in trust-funnel methods, is only implicit in ADSWITCH. Indeed it has been argued in [18] that the (unconstrained) AdaGrad algorithm can be viewed as a trust-region method with a specific radius management. Moreover, (2.15) can also be viewed as a trust-region constraint (such a condition is also imposed in [17]), but without the need to explicitly update its radius. Finally, ADSWITCH[3] is more limited than trust-funnel methods in that it does not handle approximate projections in the tangent plane.

In what follows, we denote by $\{k_\tau\} \subseteq \{k\}$ the index subsequence of "tangential iterations", that is iterations where (2.12) holds. We also denote $\{k_\nu\} = \{k\} \setminus \{k_\tau\}$ the index subsequence of "normal iterations".

3 Complexity analysis

3.1 Assumptions

Our analysis uses the following assumptions.

AS.0: f and c are continuously differentiable on $\mathbb{R}^n$.

AS.1: For all $x \in \mathbb{R}^n$, $f(x) \geq f_{\rm low}$.

AS.2: For all $x \in \mathbb{R}^n$, $\|g(x)\| \leq \kappa_g$ where $\kappa_g \geq \eta\beta$.

AS.3: For all $x \in \mathbb{R}^n$, $\|c(x)\| \leq \kappa_c$.

AS.4: For all $x \in \mathbb{R}^n$, $\|J(x)\| \leq \kappa_J$

[3] In its present incarnation.

AS.5: For all $x \in \mathbb{R}^n$, $\sigma_{\min}(J(x)) \geq \sigma_0 \in (0,1]$.

AS.6: The gradient $g(x)$ is globally Lipschitz continuous (with constant L_g).

AS.7: The Jacobian $J(x)$ is globally Lipschitz continuous (with constant L_J).

Observe that AS.1, AS.2, AS.3 and AS.4 are automatically satisfied if the iterates x_k remain in a bounded domain of $\mathbb{R}^n$. Also note that

- AS.5 ensures that $P_T(x_k)$, $g_{T,k}$ and $\widehat{\lambda}(x)$ are well-defined.
- AS.3, AS.6 and AS.7 ensure that $\widehat{\lambda}(x)$ is Lipschitz continuous (with constant L_λ).
- AS.4 ensures that c is Lipschitz continuous (with constant L_c).
- AS.4, AS.5, AS.6 and AS.7 ensure that $\nabla_x \psi_\rho(x, \widehat{\lambda})$ is Lipschitz continuous (with constant L_L).

Note that these assumptions are very similar to those of [1, 7]. In particular, AS.5 is also part of Assumption 1 in this last reference.

We finally assume, without loss of generality[4], that

$$\Gamma_{k_\tau} \geq \varsigma \quad (\tau \geq 0). \tag{3.1}$$

3.2 Tangential steps

Our analysis hinges on the fact that first-order descent can be shown on the Lyapunov $\psi(x)$, both for tangential and normal steps, despite the fact that neither $\widehat{\lambda}(x_k)$ or ρ (which we still need to define) appears in the algorithm. We start by considering tangential steps.

Lemma 3.1 Suppose that AS.4–AS.7 hold. Then

$$\psi(x_{k_\tau+1}) - \psi(x_{k_\tau}) \leq -\alpha_{T,k_\tau}\|g_{T,k_\tau}\|^2 + \eta\,\kappa_{\tan}\,\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2. \tag{3.2}$$

where

$$\kappa_{\tan} = \frac{1}{2}\Big(L_L + \rho L_c + \beta L_\lambda + L_\lambda L_c\Big).$$

Proof. We have that

$$\begin{aligned}\psi(x_{k_\tau+1}) - \psi(x_{k_\tau}) &= \underbrace{\psi_\rho(x_{k_\tau+1}, \widehat{\lambda}(x_{k_\tau})) - \psi_\rho(x_{k_\tau}, \widehat{\lambda}(x_{k_\tau}))}_{\Delta_x} \\ &\quad + \underbrace{\psi_\rho(x_{k_\tau+1}, \widehat{\lambda}(x_{k_\tau+1})) - \psi_\rho(x_{k_\tau+1}, \widehat{\lambda}(x_{k_\tau}))}_{\Delta_\lambda}.\end{aligned} \tag{3.3}$$

Now consider Δ_x and Δ_λ separately. The Lipschitz continuity of $\nabla_x \psi(x, \widehat{\lambda})$ and (2.8) give that

$$\Delta_x = -\nabla_x \psi_\rho(x_{k_\tau}, \widehat{\lambda}(x_{k_\tau}))^T(\alpha_{T,k_\tau} g_{T,k_\tau}) + r_0 + \rho(\|c(x_{k_\tau+1})\| - \|c(x_{k_\tau})\|),$$

with

$$|r_0| \leq \frac{L_L}{2}\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2.$$

Now

$$\|c(x_{k_\tau+1})\| = \|c(x_{k_\tau}) - \alpha_{T,k_\tau} J_k g_{T,k_\tau} + r_1\| \leq \|c(x_{k_\tau})\| + \|r_1\|$$

[4]The parameter ς may be re-adjusted downwards at the first tangential iteration, if necessary.

with

$$\|r_1\| \le \frac{L_c}{2}\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2,$$

giving

$$\|c(x_{k_\tau+1})\| - \|c(x_{k_\tau})\| \le \frac{L_c}{2}\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2. \tag{3.4}$$

Successively using (2.8) and (2.9), we obtain that

$$\begin{aligned}\Delta_x &= -\alpha_{T,k_\tau} g_{T,k_\tau}^T g_{T,k_\tau} + r_0 + \frac{\rho L_c}{2}\,\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2 \\ &\le -\alpha_{T,k_\tau} g_{T,k_\tau}^T g_{T,k_\tau} + \frac{1}{2}\,(L_L+\rho L_c)\,\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2 \\ &\le -\alpha_{T,k_\tau}\|g_{T,k_\tau}\|^2 + \frac{1}{2}\,(L_L+\rho L_c)\,\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2.\end{aligned} \tag{3.5}$$

Now, we may use the Lipschitz continuity of $\widehat{\lambda}$ and c to deduce that

$$\begin{aligned}\Delta_\lambda &= c_{k_\tau+1}^T\big(\widehat{\lambda}_{k_\tau+1}-\widehat{\lambda}_{k_\tau}\big) \\ &= (c_{k_\tau+1}-c_{k_\tau})^T\big(\widehat{\lambda}_{k_\tau+1}-\widehat{\lambda}_{k_\tau}\big) + c_{k_\tau}^T\big(\widehat{\lambda}_{k_\tau+1}-\widehat{\lambda}_{k_\tau}\big) \\ &\le \|c_{k_\tau}\|\,\|\widehat{\lambda}_{k_\tau+1}-\widehat{\lambda}_{k_\tau}\| + \|c_{k_\tau+1}-c_{k_\tau}\|\,\|\widehat{\lambda}_{k_\tau+1}-\widehat{\lambda}_{k_\tau}\| \\ &\le L_\lambda\|c_{k_\tau}\|\,\alpha_{T,k_\tau}\|g_{T,k_\tau}\| + L_\lambda L_c\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2,\end{aligned} \tag{3.6}$$

Taking (2.12) into account gives that $\|c_{k_\tau}\| \le \beta\alpha_{T,k_\tau}\|g_{T,k_\tau}\|$, we obtain that

$$\Delta_\lambda \le (\beta L_\lambda + L_\lambda L_c)\,\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2.$$

Thus, summing Δ_x and Δ_λ,

$$\begin{aligned}\psi(x_{k_\tau+1}) - \psi(x_{k_\tau}) \le &-\alpha_{T,k_\tau}\|g_{T,k_\tau}\|^2 + \frac{1}{2}(L_L+\rho L_c)\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2 \\ &+ (\beta L_\lambda + L_\lambda L_c)\,\alpha_{T,k_\tau}^2\|g_{T,k_\tau}\|^2\end{aligned}$$

and (3.2) follows. □

The bound (3.2) quantifies the effect of tangential steps on the Lyapunov function, and its right-hand side involves a first-order (descent) term and a second-order perturbation term. We now derive crucial bounds on these terms.

Lemma 3.2 Suppose that AS.2 and AS.5 hold. If we denote

$$\Gamma_{k_{\tau+1}} = \Gamma_{k_\tau} + \|g_{T,k_\tau}\|^2 \quad \text{and} \quad \alpha_{T,k_\tau} = \frac{\eta}{\sqrt{\Gamma_{k_{\tau+1}}+\varsigma}},$$

then, for all $0 \le \tau_0 \le \tau_1$,

$$\sum_{\tau=\tau_0}^{\tau_1} \alpha_{T,k_\tau}\,\|g_{T,k_\tau}\|^2 > \frac{\eta}{2\sqrt{2}}\,\sqrt{\Gamma_{k_{\tau_1+1}}} \tag{3.7}$$

$$\sum_{\tau=\tau_0}^{\tau_1} \alpha_{T,k_\tau}^2\,\|g_{T,k_\tau}\|^2 \le \eta^2\,\log\left(\frac{\Gamma_{k_{\tau_1+1}}+\varsigma}{\Gamma_{k_{\tau_0}}+\varsigma}\right). \tag{3.8}$$

Proof. Let $w_{k_\tau+1} = \sqrt{\Gamma_{k_{\tau+1}} + \varsigma}$. The definition (2.11) implies that

$$\begin{aligned}\sum_{\tau=\tau_0}^{\tau_1} \alpha_{T,k_\tau} \|g_{T,k_\tau}\|^2 &= \eta \sum_{\tau=\tau_0}^{\tau_1} \frac{\|g_{T,k_\tau}\|^2}{\sqrt{\Gamma_{k_{\tau+1}} + \varsigma}} \\ &> \eta \sum_{\tau=\tau_0}^{\tau_1} \frac{\|g_{T,k_\tau}\|^2}{w_{k_{\tau+1}} + w_{k_\tau}} \\ &= \eta \sum_{\tau=\tau_0}^{\tau_1} \frac{w_{k_{\tau+1}}^2 - w_{k_\tau}^2}{w_{k_{\tau+1}} + w_{k_\tau}} \\ &= \eta \sum_{\tau=\tau_0}^{\tau_1} (w_{k_{\tau+1}} - w_{k_\tau}), \\ &= \eta(w_{k_{\tau_1+1}} - w_{k_{\tau_0}}),\end{aligned}$$

Now observe that, using (3.1),

$$\begin{aligned} w_{k_{\tau_1+1}} - w_{k_{\tau_0}} &= \sqrt{\Gamma_{k_{\tau_1+1}} + \varsigma} - \sqrt{\Gamma_{k_{\tau_0}} + \varsigma} \\ &= \frac{\Gamma_{k_{\tau_1+1}} - \Gamma_{k_{\tau_0}}}{\sqrt{\Gamma_{k_{\tau_1+1}} + \varsigma} + \sqrt{\Gamma_{k_{\tau_0}} + \varsigma}} \\ &\geq \frac{\Gamma_{k_{\tau_1+1}} - \Gamma_{k_{\tau_0}}}{\sqrt{2\Gamma_{k_{\tau_1+1}}} + \sqrt{2\Gamma_{k_{\tau_0}}}} \\ &> \frac{\Gamma_{k_{\tau_1+1}}}{2\sqrt{2\Gamma_{k_{\tau_1+1}}}} = \frac{\sqrt{\Gamma_{k_{\tau_1+1}}}}{2\sqrt{2}}\end{aligned}$$

which gives (3.7). Using the concavity and the increasing nature of the logarithm, we also have from (2.11) that

$$\sum_{\tau=\tau_0}^{\tau_1} \alpha_{T,k_\tau}^2 \|g_{T,k_\tau}\|^2 = \eta^2 \sum_{\tau=\tau_0}^{\tau_1} \frac{\|g_{T,k_\tau}\|^2}{\Gamma_{k_{\tau+1}} + \varsigma} = \eta^2 \sum_{\tau=\tau_0}^{\tau_1} \frac{\Gamma_{k_{\tau+1}} - \Gamma_{k_\tau}}{\Gamma_{k_{\tau+1}} + \varsigma} \leq \eta^2 \sum_{\tau=\tau_0}^{\tau_1} \log(\Gamma_{k_{\tau+1}} + \varsigma) - \log(\Gamma_{k_\tau} + \varsigma),$$

giving (3.8). □

3.3 Normal steps

Having considered tangential steps, we now show that a decrease in the Lyapunov function ψ may also be proved if a normal step is taken, provided the penalty parameter ρ is chosen large enough. We start with a fairly simple observation.

Lemma 3.3 Suppose that AS.5 holds and that a normal step is used at iteration k_ν. Then

$$\|c_{k_\nu+1}\| \leq \left(1 - \kappa_{\rm nrm}\sigma_0^2\right) \|c_{k_\nu}\|. \tag{3.9}$$

Proof. We have from (2.16) that $\|c_{k_\nu+1}\| < \|c_{k_\nu}\|$ Then

$$\|c_{k_\nu}\|(\|c_{k_\nu+1}\| - \|c_{k_\nu}\|) \leq (\|c_{k_\nu+1}\| + \|c_{k_\nu}\|)(\|c_{k_\nu+1}\| - \|c_{k_\nu}\|) = \|c_{k_\nu+1}\|^2 - \|c_{k_\nu}\|^2$$

and therefore, using (2.16) again, that

$$\|c_{k_\nu+1}\| - \|c_{k_\nu}\| \leq -\frac{\kappa_{\rm nrm}\|J_{k_\nu}^T c_{k_\nu}\|^2}{\|c_{k_\nu}\|} \leq -\frac{\kappa_{\rm nrm}\sigma_0^2\|c_{k_\nu}\|^2}{\|c_{k_\nu}\|} = -\kappa_{\rm nrm}\sigma_0^2\|c_{k_\nu}\|.$$

□

Lemma 3.4 Suppose that AS.3–AS.7 hold and that a normal step is used at iteration k_ν. Define

$$\rho = \frac{1}{\kappa_{\rm nrm}\,\sigma_0^2}\Big[\,\kappa_c\,\theta\,L_\lambda \;+\; \frac{1}{2}\,\kappa_c\,\theta^2\,(L_L + L_\lambda L_c) \;+\; \eta\,\Big] \tag{3.10}$$

Then we have that

$$\psi(x_{k_\nu+1}) - \psi(x_{k_\nu}) \le -\eta\|c_{k_\nu}\|. \tag{3.11}$$

Proof. The first part of (2.7) and the fact that $s_{N,k}$ belongs to the orthogonal of the Jacobian's nullspace, $\nabla_x L(x_{k_\nu},\widehat{\lambda}_{k_\nu}) = g_{T,k_\nu}$ and $g_{T,k_\nu}^T s_{N,k_\nu} = 0$. Thus, using (2.6), (3.9), (2.15), the Lipschitz continuity of $\nabla_x\psi(x,\widehat{\lambda})$ (ρ is fixed in (3.10)) and the definition of Δ_x and Δ_λ in (3.3), we obtain that

$$\begin{aligned}
\Delta_x &= \psi_\rho(x_{k_\nu+1},\widehat{\lambda}_{k_\nu}) - \psi_\rho(x_{k_\nu},\widehat{\lambda}_{k_\nu})\\
&= L(x_{k_\nu+1},\widehat{\lambda}_{k_\nu}) - L(x_{k_\nu},\widehat{\lambda}_{k_\nu}) + \rho\big(\|c_{k_\nu+1}\| - \|c_{k_\nu}\|\big)\\
&\le (\nabla_x L(x_{k_\nu},\widehat{\lambda}_{k_\nu})^T s_{N,k_\nu} + r_3 - \rho\kappa_{\rm nrm}\sigma_0^2\|c_{k_\nu}\|\\
&\le -\rho\kappa_{\rm nrm}\sigma_0^2\|c_{k_\nu}\| + \frac{\theta^2 L_L}{2}\|c_{k_\nu}\|^2.
\end{aligned} \tag{3.12}$$

Using AS.6, the Lipschitz continuity of $\widehat{\lambda}$ and c and AS.5, we then deduce that

$$\begin{aligned}
\Delta_\lambda &= \psi_\rho(x_{k_\nu+1},\widehat{\lambda}_{k_\nu+1}) - \psi_\rho(x_{k_\nu+1},\widehat{\lambda}_{k_\nu})\\
&\le (\|c_{k_\nu}\| + \|c_{k_\nu+1} - c_{k_\nu}\|)\,\|\widehat{\lambda}_{k_\nu+1} - \widehat{\lambda}_{k_\nu}\|\\
&\le L_\lambda\,\|s_{N,k_\nu}\|\,\|c_{k_\nu}\| + \frac{L_\lambda L_c}{2}\|s_{N,k_\nu}\|^2\\
&\le L_\lambda\,\theta\|c_{k_\nu}\|^2 + \frac{L_\lambda L_c\theta^2}{2}\|c_{k_\nu}\|^2
\end{aligned} \tag{3.13}$$

and thus, summing (3.12) and (3.13), that

$$\psi(x_{k_\nu+1}) - \psi(x_{k_\nu}) \le -\rho\kappa_{\rm nrm}\sigma_0^2\|c_{k_\nu}\| + L_\lambda\,\theta\kappa_c\,\|c_{k_\nu}\| + \frac{1}{2}\,(L_L + L_\lambda L_c)\,\theta^2\kappa_c\|c_{k_\nu}\|.$$

The bound (3.11) then follows from (3.10).
□

3.4 Telescoping sum

We now combine the effects of tangential and normal steps to accumulate in a telescoping sum the relevant decreases on $\psi(x)$ across all iterations.

Lemma 3.5 Suppose that AS.0–AS.7 hold. Then

$$\sqrt{\Gamma_{k_{\tau_1+1}}} + \sum_{\nu=\nu_0}^{\nu_1} \|c_{k_\nu}\| \leq \kappa_{\rm gap} + \kappa_{\rm tan} \log\left(1 + \frac{\Gamma_{k_{\tau_1+1}}}{\varsigma}\right), \tag{3.14}$$

where

$$\kappa_{\rm gap} = \frac{2\sqrt{2}}{\eta}\left(\psi(x_0) + \frac{\kappa_J\kappa_g\kappa_c}{\sigma_0^2} + \rho\kappa_c - f_{\rm low}\right).$$

Proof. Combining (3.2), (3.7) and (3.8), we obtain that

$$\sum_{\tau=\tau_0}^{\tau_1} \left(\psi(x_{k_\tau+1}) - \psi(x_{k_\tau})\right) \leq -\frac{\eta}{2\sqrt{2}}\sqrt{\Gamma_{k_{\tau_1+1}}} + \eta^2\kappa_{\rm tan}\log\left(\frac{\Gamma_{k_{\tau_1+1}}+\varsigma}{\Gamma_{k_{\tau_0}}+\varsigma}\right) \tag{3.15}$$

Then, if $\min[k_{\nu_0}, k_{\tau_0}] = 0$ and $\max[k_{\nu_1}, k_{\tau_1}] = k$, we have that $w_{k_{\tau_0}} = \varsigma$ and $\Gamma_{k_{\tau_0}} = 0$. The bounds (3.11) and (3.15) therefore give that

$$\begin{aligned}
&\psi(x_{k+1}) - \psi(x_0)\\
&\quad= \sum_{\tau=\tau_0}^{\tau_1}\big(\psi(x_{k_\tau+1}) - \psi(x_{k_\tau})\big) + \sum_{\nu=\nu_0}^{\nu_1}\big(\psi(x_{k_\nu+1}) - \psi(x_{k_\nu})\big)\\
&\quad\leq -\frac{\eta}{2\sqrt{2}}\sqrt{\Gamma_{k_{\tau_1+1}}} - \eta\sum_{\nu=\nu_0}^{\nu_1}\|c_{k_\nu}\| + \eta^2\kappa_{\rm tan}\log\left(\frac{\Gamma_{k_{\tau_1+1}}+\varsigma}{\Gamma_{k_{\tau_0}}+\varsigma}\right)\\
&\quad\leq -\frac{\eta}{2\sqrt{2}}\sqrt{\Gamma_{k_{\tau_1+1}}} - \frac{\eta}{2\sqrt{2}}\sum_{\nu=\nu_0}^{\nu_1}\|c_{k_\nu}\| + \eta\kappa_{\rm tan}\log\left(1+\frac{\Gamma_{k_{\tau_1+1}}}{\varsigma}\right)
\end{aligned} \tag{3.16}$$

where we used the fact that $\eta \leq 1$. But (2.3), AS.2, AS.4, and AS.5 ensure that

$$\|\widehat{\lambda}(x)\| \leq \frac{\kappa_J}{\sigma_0^2}\|g(x)\| \leq \frac{\kappa_J\kappa_g}{\sigma_0^2}.$$

Hence, using (2.5), AS.1, AS.2, AS.3, we have that

$$\psi(x_{k+1}) - \psi(x_0) \geq f_{\rm low} - \frac{\kappa_J\kappa_g\kappa_c}{\sigma_0^2} - \rho\kappa_c - \psi(x_0) \stackrel{\rm def}{=} -\eta\kappa_{\rm gap},$$

so that (3.16) implies (3.14). □

This result is central in our analysis as it provides consistent upper bounds on a quantity related to the optimality mesure $\|g_{T,k}\| + \|c_k\|$, which we now exploit separately.

3.5 Tangential complexity

Our next step is to use (3.14) to derive a global rate of convergence over tangential steps. We start by proving a useful technical result.

Lemma 3.6 Suppose that $at \leq b + c\,\log(t)$ for $t \geq 1$. Then

$$t \leq \max\left[e^{b/c}, \frac{4c^2}{a^2}\right].$$

Proof. Suppose that $t > e^{b/c}$. Then $b \le c\,\log(t)$ and thus (see [28, eq. (14)])

$$at \le 2c\,\log(t) \le 2c\,\log(1+t) \le \frac{2ct}{\sqrt{1+t}}.$$

Hence $a\sqrt{1+t} \le 2c$, which is to say that $t \le (2c/a)^2 - 1$, yielding the desired bound. □

We now consider the rate of convergence for tangential steps proper.

Lemma 3.7 Suppose that AS.0–AS.7 hold. Then

$$\sqrt{\Gamma_{k_{\tau_1+1}}} \le \kappa_T \stackrel{\rm def}{=} \sqrt{\frac{\varsigma}{2}}\max\left[e^{\frac{\kappa_{\rm gap}}{2\kappa_{\rm tan}}}, \frac{32\kappa_{\rm tan}^2}{\varsigma}\right] \tag{3.17}$$

and

$$\sum_{\tau=\tau_0}^{\tau_1} \|g_{T,k_\tau}\| + \|c_{k_\tau}\| \le \kappa_T\sqrt{\tau_1+1}\left(1+\frac{\beta\eta}{\sqrt{\varsigma}}\right). \tag{3.18}$$

Proof. The bound (3.14) implies that

$$\sqrt{\Gamma_{k_{\tau_1+1}}} \le \kappa_{\rm gap} + \kappa_{\rm tan}\log\left(1+\frac{\Gamma_{k_{\tau_1+1}}}{\varsigma}\right).$$

Now (3.1) implies that

$$1+\frac{\Gamma_{k_{\tau_1+1}}}{\varsigma} \le \frac{2}{\varsigma}\Gamma_{k_{\tau_1+1}}$$

and thus that

$$\sqrt{\frac{\varsigma}{2}}\sqrt{\frac{2\Gamma_{k_{\tau_1+1}}}{\varsigma}} \le \kappa_{\rm gap} + \kappa_{\rm tan}\log\left(\frac{2\Gamma_{k_{\tau_1+1}}}{\varsigma}\right) = \kappa_{\rm gap} + 2\kappa_{\rm tan}\log\left(\sqrt{\frac{2\Gamma_{k_{\tau_1+1}}}{\varsigma}}\right).$$

Using Lemma 3.6, we then obtain that

$$\sqrt{\Gamma_{k_{\tau_1+1}}} = \sqrt{\frac{\varsigma}{2}}\sqrt{\frac{2\Gamma_{k_{\tau_1+1}}}{\varsigma}} \le \sqrt{\frac{\varsigma}{2}}\max\left[e^{\frac{\kappa_{\rm gap}}{2\kappa_{\rm tan}}}, \frac{32\kappa_{\rm tan}^2}{\varsigma}\right]$$

This is (3.17). We may now invoke the inequality

$$\sum_{j=0}^k a_j \le \sqrt{k+1}\sqrt{\sum_{j=0}^k a_j^2}$$

for nonnegative $\{a_j\}_{j=0}^k$ to deduce from (2.10) and (3.17) that

$$\sum_{\tau=\tau_0}^{\tau_1} \|g_{T,k_\tau}\| \le \sqrt{\tau_1+1}\sqrt{\sum_{\tau=\tau_0}^{\tau_1}\|g_{T,k_\tau}\|^2} = \sqrt{\tau_1+1}\,\sqrt{\Gamma_{k_{\tau_1+1}}} \le \sqrt{\tau_1+1}\,\kappa_T. \tag{3.19}$$

Using the switching condition (2.12) and the fact that $\alpha_{T,k_\tau} \le \eta/\sqrt{\varsigma}$, we also deduce that

$$\sum_{\tau=\tau_0}^{\tau_1}\|c_{k_\tau}\| \le \sum_{\tau=\tau_0}^{\tau_1}\beta\alpha_{T,k_\tau}\|g_{T,k_\tau}\| \le \frac{\beta\eta}{\sqrt{\varsigma}}\sum_{\tau=\tau_0}^{\tau_1}\|g_{T,k_\tau}\| \le \frac{\beta\eta\kappa_T\sqrt{\tau_1+1}}{\sqrt{\varsigma}}.$$

Summing this bound with (3.19) then gives (3.18). □

Note that a marginally tighter bound on $\Gamma_{k_{\tau_1}}$ can obtained, as in [2], by using the Lambert W_{-1} function (see [6]) instead of applying Lemma 3.6, but we have chosen the latter for simplicity of exposition.

3.6 Normal complexity

The analysis of the complexity of normal step also uses the switching condition, but in the other direction.

Lemma 3.8 Suppose that AS.0–AS.7 hold. Then

$$\sum_{\nu=\nu_0}^{\nu_1} \|g_{T,k_\nu}\| + \|c_{k_\nu}\| < \kappa_N \left(1 + \frac{2\sqrt{2}}{\beta\eta}\right), \tag{3.20}$$

where

$$\kappa_N = \kappa_{\rm gap} + \kappa_{\rm tan} \log\left(\frac{2\kappa_T}{\varsigma}\right). \tag{3.21}$$

Proof. The bound (3.14) ensures that

$$\sum_{\nu=\nu_0}^{\nu_1} \|c_{k_\nu}\| \le \kappa_{\rm gap} + \kappa_{\rm tan} \log\left(1 + \frac{\Gamma_{k_{\tau_1+1}}}{\varsigma}\right) \tag{3.22}$$

where k_{τ_1} is the index of the last tangential iteration before k_{ν_1}. As in Lemma 3.7, we may now use the assumption that $\Gamma_{k_{\tau_1}} \ge \varsigma$ and the monotonicity of the logarithm to obtain that

$$\sum_{\nu=\nu_0}^{\nu_1} \|c_{k_\nu}\| \le \kappa_{\rm gap} + 2\kappa_{\rm tan} \log\left(\sqrt{\frac{2\Gamma_{k_{\tau_1}}}{\varsigma}}\right) \le \kappa_{\rm gap} + 2\kappa_{\rm tan} \log\left(\sqrt{\frac{2}{\varsigma}\kappa_T}\right) = \kappa_N. \tag{3.23}$$

Using the switching condition (2.12), we obtain that, for $\nu \in \{\nu_0, \dots, \nu_1\}$,

$$\|c_{k_\nu}\| \ge \beta\, \alpha_{T,k_\nu} \|g_{T,k_\nu}\|. \tag{3.24}$$

Observe now that, using the assumption that $\Gamma_{k_\nu} \ge \varsigma$, and (3.17),

$$\alpha_{T,k_\nu} = \frac{\eta}{\sqrt{\Gamma_{k_\nu} + \varsigma}} \ge \frac{\eta}{\sqrt{2\Gamma_{k_\nu}}} \ge \frac{\eta}{\sqrt{2\kappa_T}}.$$

Thus, (3.24) yields that, for $\nu \in \{\nu_0, \dots, \nu_1\}$

$$\|c_{k_\nu}\| \;\ge \frac{\beta\eta}{\sqrt{2\kappa_T}} \|g_{T,k_\nu}\|. \tag{3.25}$$

With (3.23), this implies that

$$\sum_{\nu=\nu_0}^{\nu_1} \|g_{T,k_\nu}\| \le \frac{\sqrt{2\kappa_T}}{\beta\eta} \sum_{\nu=\nu_0}^{\nu_1} \|c_{k_\nu}\| \le \frac{\kappa_N \sqrt{2\kappa_T}}{\beta\eta}.$$

Summing this bound with (3.23) gives (3.20). □

3.7 Combined complexity

The combined complexity may now be derived by assembling the above results.

Theorem 3.9 Suppose that AS.0-AS.7 hold. Then

$$\frac{1}{k+1}\sum_{j=0}^{k}\Big(\|g_{T,j}\|+\|c_j\|\Big) \le \frac{\kappa_{\rm ADSW,1}}{\sqrt{k+1}}+\frac{\kappa_{\rm ADSW,2}}{k+1} = \mathcal{O}\left(\frac{1}{\sqrt{k+1}}\right), \tag{3.26}$$

where

$$\kappa_{\rm ADSW,1} = \kappa_T\left(1+\frac{\beta\eta}{\sqrt{\varsigma}}\right)+\frac{4\kappa_N\,\kappa_g}{\beta\eta} \quad\text{and}\quad \kappa_{\rm ADSW,2} = \kappa_N\left(1+\frac{2\sqrt{2}}{\beta\eta}\right).$$

Proof. Now consider iterations from 0 to k of both types (tangential and normal) by setting $\min[k_{\nu_0},k_{\tau_0}]=0$ and $\max[k_{\nu_1},k_{\tau_1}]=k$ (as in Lemma 3.5). We then obtain, by combining (3.18) and (3.20), that

$$\begin{aligned}\sum_{j=0}^{k}\Big(\|g_{T,j}\|+\|c_j\|\Big) &= \sum_{\tau=\tau_0}^{\tau_1}\Big(\|g_{T,k_\tau}\|+\|c_{k_\tau}\|\Big)+\sum_{\nu=\nu_0}^{\nu_1}\Big(\|g_{T,k_\nu}\|+\|c_{k_\nu}\|\Big)\\ &\le \kappa_T\sqrt{k+1}\left(1+\frac{\beta\eta}{\sqrt{\varsigma}}\right)+\kappa_N\left(1+\frac{2\sqrt{2}}{\beta\eta}\right),\end{aligned}$$

where we used the inequalities $\tau_1 \le k_{\tau_1}\le k$ and $k_{\nu_1}\le k$. The bound (3.26) is finally obtained by dividing both sides by $k+1$. □

4 Numerical illustration

To illustrate the behaviour of ADSWITCH, we coded the algorithm in Matlab and applied it on a set of small dimensional problems from the CUTEst collection [15] as supplied in Matlab by S2MPJ [21]. Our implementation uses the Newton normal step (2.19) and the constants

$$\beta = 0.01,\quad \eta=1,\quad \theta=1000,\quad \delta=\varsigma=10^{-5}\ \text{and}\ \omega=1.$$

For a given $\epsilon\in(0,1)$ a run on a given test problem is deemed successful at iteration k if

$$\max[\|g_{T,k}\|,\|c_k\|]\le\epsilon\ \ \text{(convergence)} \tag{4.1}$$

or

$$\|J_k^Tc_k\|\le\epsilon\ \text{ but }\ \|c_k\|>\epsilon\ \ \text{(infeasible critical point)} \tag{4.2}$$

or if a sufficiently optimal function value was found in the sense that

$$\|c_k\|\le\epsilon\ \text{ and }\ \left\{\begin{array}{ll}|f(x_k)|\le|f_*|+10^{-7} & \text{if }\ |f_*|<10^{-7}\\ |f(x_k)-f_*|\le 10^{-7}|f_*| & \text{if }\ |f_*|\ge 10^{-7}\end{array}\right. \tag{4.3}$$

where f_* is the best known feasible function value for the problem. Optimization was also stopped after a maximum of 100000 iterations.

We first ran all test problems with accurate gradient values and $\epsilon=10^{-5}$. The complete results are reported in appendix, and may be broadly summarized as follows.

1. The algorithm appears to converge as predicted. Despite the fact that our current analysis does not cover the case where the Jacobian may lose rank, the algorithm did solve a number of instances where this occurs, sometimes finding an infeasible critical point of the constraints' violation[5].

2. Its performance and reliability is (as could be anticipated) dominated by that of AdaGrad for the tangential step. Since this method is purely first-order, it often performs well, but may fail to solve ill-conditioned problems in a reasonable number of iterations. This is also the case for ADSWITCH. Over the 71 problems in our test set, it solves 44 (62%) of them within 750 iterations, 58 (81%) of them within 100000 iterations and fails on 13. While reliability is not a strong point of AdaGrad and ADSWITCH in the deterministic case, the situation is very different when noise is present, as we show below.

We illustrate a few case of satisfactory convergence in Figures 1 and 2. In these figures, one can clearly see the difference in speed of convergence between tangential steps (AdaGrad-like) and normal steps (Newton), and thus that the overall performance is dominated by that of the first-order method defining the tangential step. This is even clearer when convergence is too slow, as shown in Figure 3 where constraint violation remains very small[6], while very slow convergence of the projected gradient results in a large number of iterations. The left panel of Figure 1 (ORTHREGA with $n = 133$ and $m = 64$) is also interesting because it shows the switching condition (2.12) in action: the algorithm first approaches an infeasible critical point of the constraints' violation until the switching rule allows the tangential step to take over, causing the iterates to escape.

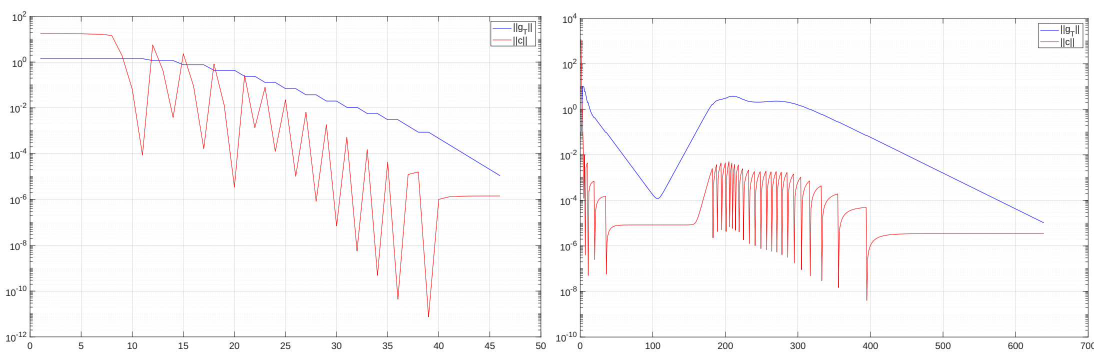

Figure 1: Evolution of the optimality measure (left: BYRDSPHR, right: ORTHREGA)

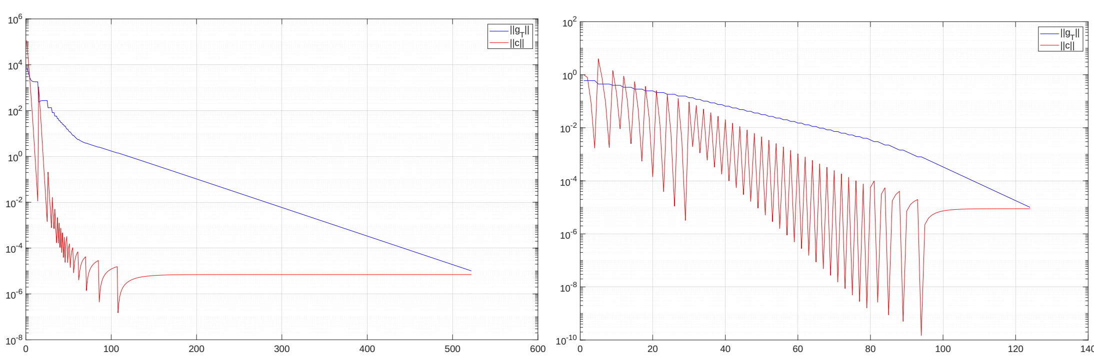

Figure 2: Evolution of the optimality measure (left: LUKVLE2, right: BT1)

[5] This occurs, for instance, at the starting point of HS61.
[6] HS50 has linear equality constraints only.

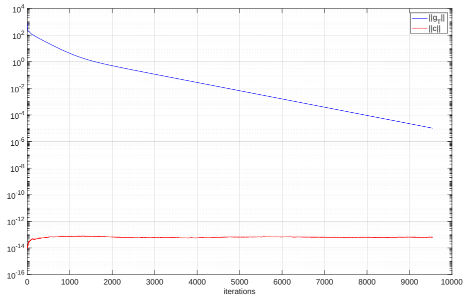

Figure 3: Evolution of the optimality measure (HS50)

relative noise level	number of total failures	number of total successes
5%	6	61
15%	7	53
25%	7	53
50%	12	51

Figure 4: Reliability of the ADSWITCH algorithm in the presence of relative Gaussian noise on the gradient.

In order to illustrate the resilience of ADSWITCH when noise is present, we also ran the algorithm on all instances of our test problems while adding relative Gaussian noise (of zero mean and unit variance) to the gradient. Each test problem was run ten times independently, and for four levels (5%, 15%, 25% and 50%) of noise. The termination ϵ was set to 10^{-3} in (4.1)–(4.3). The table in Figure 4 reports, for each of these levels, the number of problems for which all ten runs failed (middle column) and the number of problems for which all ten runs converged (rightmost column). This table shows that the reliability of the algorithm is remarkably insensitive to the noise level. It is indeed noteworthy that around two thirds of the considered test problems could be solved with top reliability even if their gradients are perturbed by 50% relative noise, which means that barely one significant digit is correct. This stable behaviour is also visible in Figure 5 showing the performance of the algorithm on the BT1 problem for increasing noise levels.

5 Conclusions and perspectives

false We have proposed a very simple first-order algorithm for solving nonlinear optimization problems with nonlinear equality constraints. This algorithm adaptively selects steps in the plane tangent to the constraints or steps that reduce infeasibility. It does so without using a merit function or a filter to enforce convergence, but instead relies on the simple switching condition (2.12). The tangent steps are based on the AdaGrad method for unconstrained minimization. As is the case for AdaGrad, the objective function is never evaluated. We have analyzed its worst-case evaluation complexity and obtain global convergence rates which match the best known rates for unconstrained problems. Numerical experiments have been presented indicating that the performance of the algorithm is comparable to that of first-order methods for unconstrained problems.

At this stage, several theoretical questions remain open for further research. These include a stochastic convergence analysis, covering the cases of rank-deficient Jacobians and/or unbounded gradients, the use of alternative minimization methods in the tangential step (such as ADAM [23], ASTR1 [19], OFFAR [20] or more standard unconstrained minimizing techniques using objective function's values). On the practical side, further numerical experiments are clearly necessary for assessing the true potential of the new method. The handling of inequality constraints is also of obvious interest.

Acknowledgement

Philippe Toint is grateful for the continued and friendly support of the Algo team at Toulouse IRIT. The authors also thank Benedetta Morini and Stefania Bellavia for pointing out a number of typos.

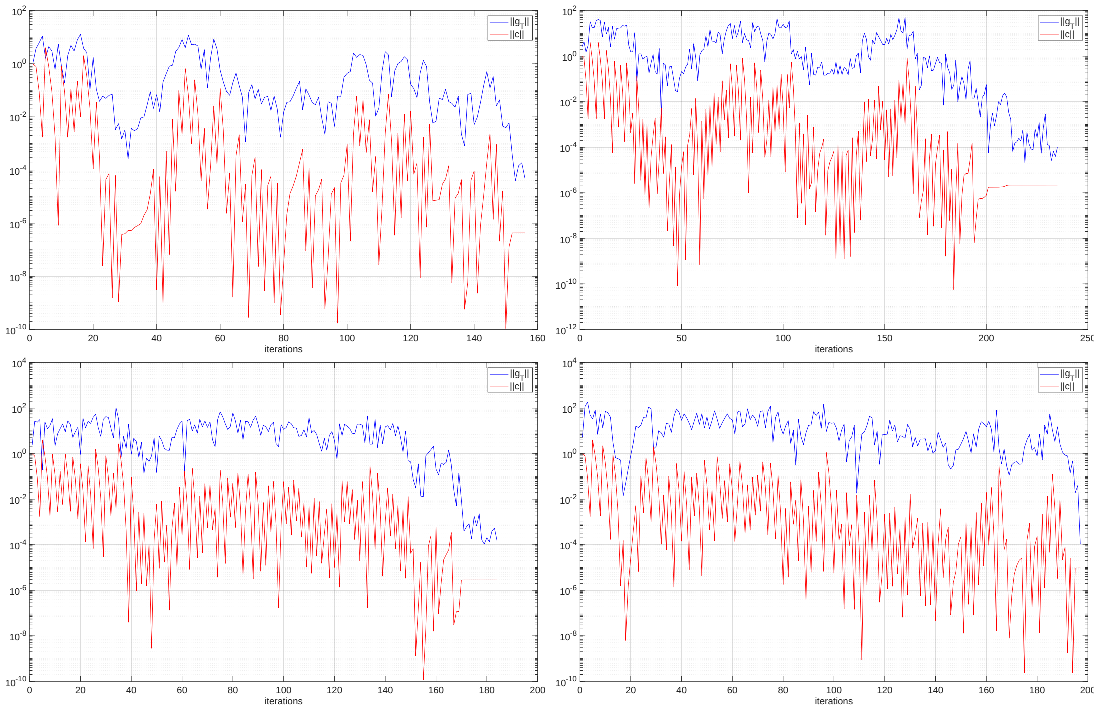

Figure 5: Evolution of the two components of the optimality measure for a run of the noisy BT1 problem, with 5% (top left), 15% (top right), 25% (bottom left) and 50% (bottom right) relative Gaussian noise on the gradient

References

[1] F. Curtis A. Berahas, D. Robinson, and B. Zhou. Sequential quadratic optimization for nonlinear equality constrained stochastic optimization. SIAM Journal on Optimization, 31(2):1352–1379, 2021.

[2] S. Bellavia, G. Gratton, B. Morini, and Ph. L. Toint. Fast stochastic Adagrad for nonconvex bound-constrained optimization. arXiv:2505:06374, 2025.

[3] C. Cartis, N. I. M. Gould, and Ph. L. Toint. Evaluation complexity of algorithms for nonconvex optimization. Number 30 in MOS-SIAM Series on Optimization. SIAM, Philadelphia, USA, June 2022.

[4] C. Coelho, M. Fernanda, P. Costa, and L. L. Ferras. A two-stage training method for modeling constrained systems with neural networks. arXiv:2403.02730, 2024.

[5] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. Trust-Region Methods. Number 1 in MOS-SIAM Optimization Series. SIAM, Philadelphia, USA, 2000.

[6] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. Advances in Computational Mathematics, 5:329—359, 1996.

[7] F. Curtis, D. Robinson, and B. Zhou. Sequential quadratic optimization for stochastic optimization with deterministic nonlinear inequality and equality constraints. SIAM Journal on Optimization, 34(4):3592–3622, 2024.

[8] F. E. Curtis, D. P. Robinson, and M. Samadi. Complexity analysis of a trust funnel algorithm for equality constrained optimization. SIAM Journal on Optimization, 28(2):1533–1563, 2018.

[9] A. Défossez, L. Bottou, F. Bach, and N. Usunier. A simple convergence proof for Adam and Adagrad. Transactions on Machine Learning Research, October 2022.

[10] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12, July 2011.

[11] Y. Fang, S. Na, M. Mahoney, and M. Kolar. Fully stochastic trust-region sequential quadratic programming for equality constrained optimization problems. SIAM Journal on Optimization, 34(2):2007–2037, 2024.

[12] R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint, and A. Wächter. Global convergence of trust-region SQP-filter algorithms for nonlinear programming. SIAM Journal on Optimization, 13(3):635–659, 2002.

[13] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. Mathematical Programming, 91(2):239–269, 2002.

[14] R. Fletcher, S. Leyffer, and Ph. L. Toint. On the global convergence of a filter-SQP algorithm. SIAM Journal on Optimization, 13(1):44–59, 2002.

[15] N. I. M. Gould, D. Orban, and Ph. L. Toint. CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. Computational Optimization and Applications, 60(3):545–557, 2015.

[16] N. I. M. Gould, D. P. Robinson, and Ph. L. Toint. Corrigendum: Nonlinear programming without a penalty function or a filter. Mathematical Programming, Series A, 131(1):403–404, 2012.

[17] N. I. M. Gould and Ph. L. Toint. Nonlinear programming without a penalty function or a filter. Mathematical Programming, Series A, 122(1):155–196, 2010. See also [16].

[18] S. Gratton, S. Jerad, and Ph. L. Toint. Complexity of a class of first-order objective-function-free optimization algorithms. Optimization Methods and Software, 1:1–31, 2024.

[19] S. Gratton, S. Jerad, and Ph. L. Toint. Complexity and performance for two classes of noise-tolerant first-order algorithms. Optimization Methods and Software, (to appear), 2025.

[20] S. Gratton, S. Jerad, and Ph. L. Toint. A stochastic objective-function-free adaptive regularization method with optimal complexity. Open Journal of Mathematical Optimization, 6(5), 2025.

[21] S. Gratton and Ph. L. Toint. S2MPJ and CUTEst optimization problems for Matlab, Python and Julia. Optimization Methods and Software, (to appear), 2025.

[22] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. Journal of the National Bureau of Standards, 49:409–436, 1952.

[23] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In Proceedings in the International Conference on Learning Representations (ICLR), 2015.

[24] E. O. Omojokun. Trust region algorithms for optimization with nonlinear equality and inequality constraints. PhD thesis, University of Colorado, Boulder, Colorado, USA, 1989.

[25] C. C. Paige and M. A. Saunders. LSQR: an algorithm for sparse linear equations and sparse least squares. ACM Transactions on Mathematical Software, 8:43–71, 1982.

[26] T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. SIAM Journal on Numerical Analysis, 20(3):626–637, 1983.

[27] Ph. L. Toint. Towards an efficient sparsity exploiting Newton method for minimization. In I. S. Duff, editor, Sparse Matrices and Their Uses, pages 57–88, London, 1981. Academic Press.

[28] F. Topsoe. Some bounds for the logarithmic function. Inequality Theory and Applications, 4(01), 2007.

[29] R. Ward, X. Wu, and L. Bottou. Adagrad stepsizes: sharp convergence over nonconvex landscapes. In Proceedings in the International Conference on Machine Learning (ICML2019), 2019.

Appendix; Details of the Numerical Results

The following tables give the details of the numerical results discussed in Section 4.

- Table A.1 reports the results of the noiseless runs. Columns 2 and 3 indicate the number of variables and the number of constraints. The column "#its" gives the number of iterations and the column "exitc" the termination condition ("convg" = convergence, "infeas" = infeasible critical point, "maxit" = maximum number of iterations reached).

- Tables A.2, A.3, A.4 and A.5 report the average statistics of 10 independent runs for each problem, for increasing relative Gaussian noise levels. The last column of the table gives the number of successful runs (among 10).

Problem	n	m	$f(x)$	$\|g_T(x)\|$	$\|c(x)\|$	#its	exitc
BT1	2	1	-9.999918e-01	9.64e-07	8.25e-08	141	convg
BT2	3	1	+3.256821e-02	9.34e-07	6.29e-07	186	convg
BT3	5	3	+4.093023e+00	8.58e-07	1.76e-08	70	convg
BT4	3	2	-4.551055e+01	4.49e-07	1.16e-07	16	convg
BT5	3	2	+9.617152e+02	7.00e-07	1.86e-07	46	convg
BT6	5	2	+2.770448e-01	8.75e-07	1.55e-08	107	convg
BT7	5	3	—	—	—	10572	infeas
BT8	5	2	—	—	—	21	infeas
BT9	4	2	-1.000000e+00	9.95e-07	4.92e-07	2481	convg
BT10	2	2	-1.000000e+00	0.00e+00	5.85e-09	6	convg
BT11	5	3	+8.248909e-01	9.55e-07	7.76e-07	351	convg
BT12	5	3	+6.188119e+00	9.46e-07	4.53e-07	240	convg
BYRDSPHR	3	2	-4.683300e+00	8.94e-07	3.33e-08	50	convg
DIXCHLNG	10	5	+2.471898e+03	9.98e-07	2.67e-07	2722	convg
EIGENA2	110	55	+2.448710e-13	9.90e-07	2.06e-09	696	convg
EIGENACO	110	55	+2.436158e-13	9.87e-07	9.41e-16	389	convg
EIGENB2	110	55	+1.800000e+01	6.82e-07	1.78e-15	18	convg
EIGENBCO	110	55	+9.000000e+00	6.82e-07	1.66e-15	18	convg
ELEC	75	25	+2.438128e+02	1.00e-06	4.79e-07	65380	convg
GENHS28	10	8	+9.271737e-01	9.27e-07	1.36e-11	85	convg
HS100LNP	7	2	+6.806301e+02	8.57e-07	2.47e-07	113	convg
HS6	2	1	+9.238528e-13	8.60e-07	8.53e-07	55	convg
HS7	2	1	-1.732051e+00	9.26e-07	1.96e-07	284	convg
HS8	2	2	-1.000000e+00	0.00e+00	1.32e-08	4	convg
HS9	2	1	-5.000000e-01	8.87e-07	7.11e-15	22	convg
HS26	3	1	+1.303916e-09	1.00e-06	5.10e-07	18098	convg
HS27	3	1	+3.999998e-02	9.79e-07	5.23e-07	271	convg
HS28	3	1	+9.687529e-13	9.02e-07	1.33e-15	137	convg
HS39	4	2	-1.000000e+00	9.95e-07	4.92e-07	2481	convg
HS40	4	3	-2.500002e-01	9.74e-07	3.24e-07	600	convg
HS42	4	2	+1.385786e+01	9.85e-07	8.75e-07	92	convg
HS46	5	2	+8.808661e-09	1.00e-06	2.35e-07	13907	convg
HS47	5	3	+4.650485e-10	1.00e-06	6.88e-07	26784	convg
HS48	5	2	+2.912966e-13	9.31e-07	2.81e-15	177	convg
HS50	5	3	+2.888537e-13	9.99e-07	6.63e-14	11166	convg
HS51	5	3	+8.485715e-14	5.67e-07	1.46e-15	19	convg
HS52	5	3	+5.326648e+00	9.87e-07	1.76e-09	150	convg
HS61	3	2	—	—	—	2	infeas
HS77	5	2	+2.415051e-01	9.27e-07	1.18e-07	103	convg
HS78	5	3	-2.919700e+00	9.31e-07	1.97e-08	95	convg
HS79	5	3	+7.877683e-02	9.71e-07	2.29e-07	357	convg
LUKVLE1	20	18	+6.232459e+00	9.11e-07	1.20e-08	98	convg
LUKVLE2	20	13	+4.372199e+02	9.93e-07	3.19e-08	604	convg
LUKVLE3	20	2	+2.758657e+01	9.94e-07	8.67e-07	2319	convg
LUKVLE4	20	9	+1.060660e+02	1.36e+02	3.15e-07	100000	maxit
LUKVLE6	21	10	+1.120010e+03	2.10e-01	7.64e-07	100000	maxit
LUKVLE7	20	4	-5.352208e+00	8.90e-07	3.91e-08	111	convg
LUKVLE8	20	18	+2.099621e+03	8.65e-07	3.18e-09	55	convg
LUKVLE9	20	6	+6.560168e+00	1.51e+01	8.57e-07	100000	maxit
LUKVLE10	20	18	+6.639332e+00	7.23e-07	3.27e-07	40	convg
LUKVLE11	18	10	+5.359296e-07	3.53e-05	1.69e-09	100000	maxit
LUKVLE12	17	12	+2.286872e+02	4.44e+01	2.23e+00	100000	maxit
LUKVLE13	18	10	+5.483076e+01	4.24e-06	1.20e-08	100000	maxit
LUKVLE14	18	10	+4.180657e+04	7.86e-01	7.02e-07	100000	maxit
LUKVLE15	17	12	+7.527144e+01	1.14e+01	9.10e-07	100000	maxit
LUKVLE16	17	12	+8.040898e+01	2.71e+01	3.79e+00	100000	maxit
LUKVLE17	17	12	+1.190180e+02	5.48e+01	8.37e+00	100000	maxit
LUKVLE18	17	12	+2.627594e+01	1.71e+01	8.37e+00	100000	maxit
LUKVLI4	20	9	+1.060660e+02	1.36e+02	3.15e-07	100000	maxit
MARATOS	2	1	-1.000000e+00	9.27e-07	2.26e-07	168	convg
MWRIGHT	5	3	+2.497881e+01	9.23e-07	1.04e-08	114	convg
ORTHRDM2	9	3	+1.087113e-13	6.51e-07	7.14e-07	31	convg
ORTHRDS2	9	3	+5.921433e-14	4.81e-07	2.63e-08	33	convg
ORTHREGA	133	64	+3.503002e+02	9.82e-07	1.66e-08	705	convg
ORTHREGB	27	6	+3.057742e-14	3.50e-07	1.02e-08	15	convg
ORTHREGC	15	5	+1.125393e-12	9.24e-07	7.22e-08	50	convg
ORTHREGD	43	20	+2.179046e+02	1.74e+01	1.09e-03	100000	maxit
ORTHRGDM	43	20	+5.067350e+01	8.20e+00	1.97e-03	100000	maxit
ORTHRGDS	43	20	+6.210193e+00	9.97e-07	8.69e-07	474	convg
S316m322	2	1	—	—	—	0	infeas
SPINOP	11	9	+8.435060e-02	7.91e-04	2.20e-06	100000	maxit

Table A.1: Results of running ADSWITCH in the abscence of noise

Problem	n	m	avr $f(x)$	avr $\|g_T(x)\|$	avr $\|c(x)\|$	avr #its	#success
BT1	3	1	-5.681296e-01	7.73e-04	3.20e-04	8.92e+01	10
BT2	5	3	+3.257271e-02	1.51e-03	3.49e-04	1.18e+02	10
BT3	3	2	+4.089376e+00	5.05e-02	6.34e-04	2.50e+04	10
BT4	3	2	-4.551215e+01	1.87e-02	5.03e-04	8.88e+01	10
BT5	5	2	+9.617275e+02	1.60e-01	9.36e-05	1.10e+03	10
BT6	5	3	+2.770621e-01	9.19e-03	2.69e-04	3.50e+03	10
BT7	5	2	—	—	—	2.51e+03	10
BT8	4	2	+1.000238e+00	6.34e-04	3.41e-04	1.47e+01	10
BT9	2	2	-1.000535e+00	9.87e-04	5.37e-04	1.15e+03	10
BT10	5	3	-1.000038e+00	0.00e+00	5.91e-05	5.00e+00	10
BT11	5	3	+8.246795e-01	4.64e-02	4.07e-04	1.29e+04	10
BT12	3	2	+6.188179e+00	6.06e-03	6.49e-04	1.29e+02	10
BYRDSPHR	10	5	-4.683084e+00	1.21e-02	1.63e-05	1.85e+02	10
DIXCHLNG	110	55	+2.471678e+03	1.25e+01	3.47e-04	—	9
EIGENA2	110	55	+2.507664e-07	1.00e-03	4.03e-04	3.98e+02	10
EIGENACO	110	55	+2.410396e-07	9.81e-04	9.35e-16	2.31e+02	10
EIGENB2	110	55	+3.530739e-06	1.00e-03	6.16e-04	1.10e+03	10
EIGENBCO	75	25	+5.356264e-06	9.58e-04	5.29e-04	1.07e+03	10
ELEC	10	8	+2.438048e+02	1.86e-01	4.33e-04	1.00e+05	10
GENHS28	7	2	+9.274072e-01	2.82e-02	6.75e-06	3.01e+03	10
HS100LNP	2	1	+6.806369e+02	5.73e-01	1.91e-04	—	1
HS6	2	1	+9.637182e-07	8.72e-04	2.76e-04	2.96e+01	10
HS7	2	2	-1.732173e+00	9.72e-04	4.26e-04	1.65e+02	10
HS8	2	1	-1.000000e+00	0.00e+00	1.32e-08	4.00e+00	10
HS9	3	1	-4.998315e-01	3.11e-03	5.86e-15	1.41e+01	10
HS26	3	1	+1.374024e-05	1.01e-03	7.19e-04	2.19e+02	10
HS27	3	1	+3.999576e-02	9.84e-04	2.58e-04	6.39e+01	10
HS28	4	2	+1.022606e-06	9.25e-04	6.55e-16	7.09e+01	10
HS39	4	3	-1.000759e+00	9.86e-04	7.61e-04	1.08e+03	10
HS40	4	2	-2.501555e-01	1.42e-02	4.17e-04	2.71e+02	10
HS42	5	2	+1.386050e+01	1.79e-01	2.38e-04	9.08e+02	10
HS46	5	3	+4.853021e-06	9.96e-04	5.22e-04	1.31e+02	10
HS47	5	2	+1.474970e-05	1.03e-03	4.84e-04	7.36e+02	10
HS48	5	3	+3.056976e-07	9.53e-04	1.88e-15	9.29e+01	10
HS50	5	3	+3.332153e-07	1.07e-03	1.07e-13	6.29e+03	10
HS51	5	3	+1.176429e-07	6.55e-04	1.69e-15	1.03e+01	10
HS52	3	2	+5.326494e+00	1.14e-01	6.34e-05	3.00e+04	10
HS61	5	2	—	—	—	1.00e+00	10
HS77	5	3	+2.415176e-01	1.14e-02	3.55e-04	2.04e+03	10
HS78	5	3	-2.918759e+00	6.58e-02	3.58e-04	—	6
HS79	20	18	+7.879160e-02	2.26e-03	3.19e-04	1.24e+02	10
LUKVLE1	20	13	+6.234044e+00	2.71e-03	3.84e-04	1.13e+02	10
LUKVLE2	20	2	+4.371168e+02	3.46e+00	5.67e-04	1.00e+05	10
LUKVLE3	20	9	+2.757471e+01	1.64e-01	6.39e-04	1.00e+05	10
LUKVLE4	21	10	+1.058513e+02	1.35e+02	5.50e-06	1.00e+05	10
LUKVLE6	20	4	+1.119942e+03	4.69e-01	5.08e-04	1.00e+05	10
LUKVLE7	20	18	-5.353248e+00	1.18e-01	3.97e-04	—	0
LUKVLE8	20	6	—	—	—	1.13e+04	10
LUKVLE9	20	18	+6.555645e+00	1.51e+01	1.66e-04	1.00e+05	10
LUKVLE10	18	10	+6.639743e+00	1.44e-02	2.76e-04	4.84e+03	10
LUKVLE11	17	12	+3.438763e-05	1.09e-03	4.80e-04	6.96e+03	10
LUKVLE12	18	10	+2.286872e+02	4.44e+01	2.23e+00	1.00e+05	10
LUKVLE13	18	10	+5.483466e+01	8.14e-02	4.72e-04	1.00e+05	10
LUKVLE14	17	12	+4.180658e+04	8.46e-01	7.29e-04	1.00e+05	10
LUKVLE15	17	12	+7.511104e+01	1.14e+01	1.81e-04	—	0
LUKVLE16	17	12	+8.040898e+01	2.71e+01	3.79e+00	—	0
LUKVLE17	17	12	+1.190180e+02	5.48e+01	8.37e+00	—	0
LUKVLE18	20	9	+2.627594e+01	1.71e+01	8.37e+00	—	0
LUKVLI4	2	1	+1.063358e+02	1.38e+02	5.60e-06	1.00e+05	10
MARATOS	5	3	-1.000295e+00	9.53e-04	5.91e-04	9.98e+01	10
MWRIGHT	9	3	+2.497543e+01	1.45e-01	4.88e-04	—	9
ORTHRDM2	9	3	+1.034714e-07	6.19e-04	3.51e-04	2.16e+01	10
ORTHRDS2	133	64	+8.522014e-08	5.48e-04	3.05e-04	2.14e+01	10
ORTHREGA	27	6	+3.502947e+02	6.49e-02	7.05e-04	1.00e+05	10
ORTHREGB	15	5	+5.311601e-08	3.99e-04	3.80e-05	1.06e+01	10
ORTHREGC	43	20	+6.775424e-07	8.16e-04	5.67e-04	2.69e+01	10
ORTHREGD	43	20	—	—	—	2.13e+02	10
ORTHRGDM	43	20	—	—	—	1.86e+03	10
ORTHRGDS	2	1	+6.210324e+00	2.21e-02	5.60e-04	—	0
S316m322	11	9	—	—	—	0.00e+00	10
SPINOP	11	9	+9.941025e-02	1.10e-03	5.74e-04	6.14e+04	10

Table A.2: Results of running ADSWITCH with 5% relative Gaussian noise

Problem	n	m	avr $f(x)$	avr $\|g_T(x)\|$	avr $\|c(x)\|$	avr #its	#success
BT1	3	1	-1.720654e-01	2.87e-04	2.80e-04	2.19e+02	10
BT2	5	3	+3.257574e-02	2.86e-03	5.14e-04	1.45e+02	10
BT3	3	2	+4.089932e+00	6.74e-02	6.34e-04	7.06e+04	10
BT4	3	2	-4.551083e+01	5.03e-02	1.67e-04	3.36e+02	10
BT5	5	2	+9.617453e+02	2.74e-01	4.39e-05	4.57e+03	10
BT6	5	3	+2.771418e-01	1.71e-02	3.62e-04	—	7
BT7	5	2	—	—	—	1.97e+03	10
BT8	4	2	+1.000238e+00	3.77e-04	3.41e-04	1.38e+01	10
BT9	2	2	-1.000468e+00	7.93e-04	4.70e-04	6.71e+02	10
BT10	5	3	-1.000038e+00	0.00e+00	5.91e-05	5.00e+00	10
BT11	5	3	+8.248886e-01	6.47e-02	4.69e-04	—	7
BT12	3	2	+6.188227e+00	1.06e-02	4.39e-04	1.90e+02	10
BYRDSPHR	10	5	-4.680516e+00	3.82e-02	1.76e-04	4.58e+02	10
DIXCHLNG	110	55	+2.471747e+03	3.29e+01	5.13e-04	—	8
EIGENA2	110	55	+2.864156e-07	1.07e-03	5.67e-04	3.97e+02	10
EIGENACO	110	55	+2.500993e-07	9.99e-04	9.37e-16	2.35e+02	10
EIGENB2	110	55	+2.989956e-06	1.02e-03	2.20e-04	1.05e+03	10
EIGENBCO	75	25	+1.674676e-06	8.43e-04	4.82e-04	1.35e+03	10
ELEC	10	8	+2.438072e+02	4.00e-01	5.36e-04	—	6
GENHS28	7	2	+9.275303e-01	3.38e-02	6.75e-06	3.59e+04	10
HS100LNP	2	1	+6.806511e+02	1.04e+00	3.81e-04	—	0
HS6	2	1	+8.822774e-07	8.34e-04	5.77e-04	2.94e+01	10
HS7	2	2	-1.732173e+00	9.28e-04	4.25e-04	1.45e+02	10
HS8	2	1	-1.000000e+00	0.00e+00	1.32e-08	4.00e+00	10
HS9	3	1	-4.993365e-01	7.17e-03	6.76e-15	2.20e+01	10
HS26	3	1	+1.632328e-05	1.15e-03	4.97e-04	2.15e+02	10
HS27	3	1	+3.999706e-02	9.14e-04	2.16e-04	4.66e+01	10
HS28	4	2	+1.738691e-06	1.20e-03	7.54e-16	6.87e+01	10
HS39	4	3	-1.000411e+00	1.07e-03	4.13e-04	5.98e+02	10
HS40	4	2	-2.493358e-01	3.90e-02	4.84e-06	4.18e+02	10
HS42	5	2	+1.386567e+01	2.88e-01	1.42e-04	1.38e+03	10
HS46	5	3	+7.154893e-06	1.06e-03	3.65e-04	1.24e+02	10
HS47	5	2	+1.763663e-05	1.17e-03	5.01e-04	6.97e+02	10
HS48	5	3	+3.810065e-07	1.05e-03	1.86e-15	9.47e+01	10
HS50	5	3	+4.796120e-07	1.29e-03	9.03e-14	6.20e+03	10
HS51	5	3	+1.153904e-07	6.35e-04	1.57e-15	9.90e+00	10
HS52	3	2	+5.326745e+00	1.30e-01	6.34e-05	—	5
HS61	5	2	—	—	—	1.00e+00	10
HS77	5	3	+2.415599e-01	1.79e-02	2.39e-04	5.11e+04	10
HS78	5	3	-2.918552e+00	6.92e-02	4.09e-04	—	4
HS79	20	18	+7.879634e-02	2.28e-03	5.15e-04	1.54e+02	10
LUKVLE1	20	13	+6.233718e+00	7.71e-03	3.12e-04	1.20e+02	10
LUKVLE2	20	2	+4.371783e+02	7.13e+00	4.04e-04	—	8
LUKVLE3	20	9	+2.757958e+01	5.72e-01	4.40e-04	—	9
LUKVLE4	21	10	+1.063795e+02	1.41e+02	8.44e-06	1.00e+05	10
LUKVLE6	20	4	+1.119853e+03	1.16e+00	5.46e-04	1.00e+05	10
LUKVLE7	20	18	-5.350163e+00	4.22e-01	2.45e-04	—	0
LUKVLE8	20	6	—	—	—	1.12e+03	10
LUKVLE9	20	18	+6.707674e+00	1.60e+01	2.84e-04	1.00e+05	10
LUKVLE10	18	10	+6.473732e+00	2.91e-02	3.32e-04	1.19e+04	10
LUKVLE11	17	12	+2.849669e-01	5.36e-03	5.57e-04	—	9
LUKVLE12	18	10	+2.286872e+02	4.44e+01	2.23e+00	1.00e+05	10
LUKVLE13	18	10	+5.483592e+01	1.08e-01	5.89e-04	1.00e+05	10
LUKVLE14	17	12	+4.180658e+04	1.02e+00	3.89e-04	1.00e+05	10
LUKVLE15	17	12	+7.514427e+01	1.15e+01	1.81e-04	—	0
LUKVLE16	17	12	+8.040898e+01	2.71e+01	3.79e+00	—	0
LUKVLE17	17	12	+1.190180e+02	5.48e+01	8.37e+00	—	0
LUKVLE18	20	9	+2.627594e+01	1.71e+01	8.37e+00	—	0
LUKVLI4	2	1	+1.094837e+02	1.56e+02	6.13e-06	1.00e+05	10
MARATOS	5	3	-1.000232e+00	7.86e-04	4.65e-04	9.35e+01	10
MWRIGHT	9	3	+2.497649e+01	2.93e-01	5.48e-04	—	7
ORTHRDM2	9	3	+9.737407e-08	5.56e-04	4.05e-04	2.27e+01	10
ORTHRDS2	133	64	+1.610065e-07	7.37e-04	4.02e-04	2.24e+01	10
ORTHREGA	27	6	+3.502994e+02	1.19e-01	4.88e-04	—	5
ORTHREGB	15	5	+7.691602e-08	4.91e-04	8.40e-05	1.09e+01	10
ORTHREGC	43	20	+9.146250e-07	8.78e-04	3.23e-04	2.67e+01	10
ORTHREGD	43	20	—	—	—	1.41e+02	10
ORTHRGDM	43	20	—	—	—	1.66e+03	10
ORTHRGDS	2	1	+6.210852e+00	5.11e-02	5.33e-04	—	0
S316m322	11	9	—	—	—	0.00e+00	10
SPINOP	11	9	+1.134528e-01	1.43e-03	6.21e-04	4.12e+04	10

Table A.3: Results of running ADSWITCH with 15% relative Gaussian noise

Problem	n	m	avr $f(x)$	avr $\|g_T(x)\|$	avr $\|c(x)\|$	avr #its	#success
BT1	3	1	-1.628888e-01	1.65e-05	3.72e-04	1.86e+02	10
BT2	5	3	+3.257851e-02	5.28e-03	3.90e-04	2.48e+02	10
BT3	3	2	+4.089948e+00	6.60e-02	6.34e-04	9.29e+04	10
BT4	3	2	-4.551030e+01	1.04e-01	2.53e-04	5.51e+02	10
BT5	5	2	+9.617367e+02	2.09e-01	1.93e-04	6.10e+03	10
BT6	5	3	+2.771881e-01	2.53e-02	6.25e-04	—	1
BT7	5	2	—	—	—	1.51e+03	10
BT8	4	2	+1.000238e+00	2.43e-04	3.41e-04	1.42e+01	10
BT9	2	2	-1.000286e+00	6.60e-04	2.91e-04	4.16e+02	10
BT10	5	3	-1.000038e+00	0.00e+00	5.91e-05	5.00e+00	10
BT11	5	3	+8.259372e-01	1.09e-01	4.43e-04	—	6
BT12	3	2	+6.188639e+00	2.01e-02	3.28e-04	2.05e+02	10
BYRDSPHR	10	5	-4.672953e+00	8.00e-02	9.61e-05	5.88e+02	10
DIXCHLNG	110	55	+2.471810e+03	3.74e+01	4.88e-04	—	6
EIGENA2	110	55	+2.814117e-07	1.06e-03	3.87e-04	4.03e+02	10
EIGENACO	110	55	+2.651002e-07	1.03e-03	9.35e-16	2.43e+02	10
EIGENB2	110	55	+2.653621e-06	1.02e-03	5.90e-04	1.27e+03	10
EIGENBCO	75	25	+6.332410e-07	8.51e-04	4.14e-04	1.56e+03	10
ELEC	10	8	+2.438104e+02	5.26e-01	5.69e-04	—	6
GENHS28	7	2	+9.281926e-01	5.42e-02	6.75e-06	2.64e+04	10
HS100LNP	2	1	+6.806694e+02	1.50e+00	6.92e-04	—	0
HS6	2	1	+9.704023e-07	8.32e-04	4.74e-04	3.12e+01	10
HS7	2	2	-1.732196e+00	7.94e-04	5.04e-04	1.51e+02	10
HS8	2	1	-1.000000e+00	0.00e+00	1.32e-08	4.00e+00	10
HS9	3	1	-4.981924e-01	1.09e-02	2.06e-14	7.99e+01	10
HS26	3	1	+1.680639e-05	1.18e-03	5.93e-04	2.00e+02	10
HS27	3	1	+4.000247e-02	9.90e-04	1.26e-04	4.02e+01	10
HS28	4	2	+2.764455e-06	1.42e-03	8.77e-16	7.06e+01	10
HS39	4	3	-1.000434e+00	7.07e-04	4.45e-04	3.96e+02	10
HS40	4	2	-2.232808e-01	6.51e-02	1.23e-04	5.08e+02	10
HS42	5	2	+1.387811e+01	4.89e-01	3.55e-04	3.79e+03	10
HS46	5	3	+7.617645e-06	1.14e-03	6.62e-04	1.16e+02	10
HS47	5	2	+2.249967e-05	1.37e-03	3.85e-04	6.16e+02	10
HS48	5	3	+5.445390e-07	1.25e-03	1.39e-15	8.88e+01	10
HS50	5	3	+7.538284e-07	1.61e-03	9.79e-14	6.08e+03	10
HS51	5	3	+1.182718e-07	6.25e-04	1.67e-15	1.05e+01	10
HS52	3	2	+5.327150e+00	1.37e-01	6.34e-05	—	1
HS61	5	2	—	—	—	1.00e+00	10
HS77	5	3	+2.416327e-01	2.47e-02	2.62e-04	—	8
HS78	5	3	-2.918216e+00	8.49e-02	3.36e-04	—	1
HS79	20	18	+7.879582e-02	3.00e-03	6.26e-04	2.09e+02	10
LUKVLE1	20	13	+4.363520e+00	6.83e-03	3.41e-04	1.25e+02	10
LUKVLE2	20	2	+4.371531e+02	5.88e+00	6.31e-04	—	9
LUKVLE3	20	9	+2.758052e+01	1.14e+00	6.23e-04	—	7
LUKVLE4	21	10	+1.146093e+02	1.73e+02	6.14e-06	1.00e+05	10
LUKVLE6	20	4	+1.119798e+03	2.06e+00	4.77e-04	1.00e+05	10
LUKVLE7	20	18	-5.349869e+00	4.53e-01	3.32e-04	—	0
LUKVLE8	20	6	—	—	—	3.95e+02	10
LUKVLE9	20	18	+6.467339e+00	1.45e+01	1.53e-04	1.00e+05	10
LUKVLE10	18	10	+6.464724e+00	2.88e-02	5.32e-04	4.79e+04	10
LUKVLE11	17	12	+6.422930e-05	1.84e-03	3.02e-04	5.17e+03	10
LUKVLE12	18	10	+2.286872e+02	4.44e+01	2.23e+00	1.00e+05	10
LUKVLE13	18	10	+5.501071e+01	1.89e-01	3.69e-04	1.00e+05	10
LUKVLE14	17	12	+4.180664e+04	1.77e+00	4.53e-04	1.00e+05	10
LUKVLE15	17	12	+7.384004e+01	1.14e+01	1.82e-04	—	0
LUKVLE16	17	12	+8.040898e+01	2.71e+01	3.79e+00	—	0
LUKVLE17	17	12	+1.190180e+02	5.48e+01	8.37e+00	—	0
LUKVLE18	20	9	+2.627594e+01	1.71e+01	8.37e+00	—	0
LUKVLI4	2	1	+1.094630e+02	1.55e+02	9.07e-06	1.00e+05	10
MARATOS	5	3	-9.998270e-01	8.83e-03	2.92e-04	3.76e+01	10
MWRIGHT	9	3	+2.497751e+01	2.90e-01	5.26e-04	—	7
ORTHRDM2	9	3	+1.408678e-07	6.56e-04	4.15e-04	2.25e+01	10
ORTHRDS2	133	64	+1.867325e-07	7.67e-04	4.32e-04	2.25e+01	10
ORTHREGA	27	6	+3.503036e+02	1.71e-01	4.19e-04	—	3
ORTHREGB	15	5	+6.680941e-08	4.66e-04	2.55e-04	1.21e+01	10
ORTHREGC	43	20	+5.552793e-07	7.69e-04	3.77e-04	2.81e+01	10
ORTHREGD	43	20	—	—	—	1.44e+02	10
ORTHRGDM	43	20	—	—	—	1.60e+03	10
ORTHRGDS	2	1	+6.211177e+00	6.17e-02	6.38e-04	—	0
S316m322	11	9	—	—	—	0.00e+00	10
SPINOP	11	9	+1.331340e-01	1.97e-03	5.60e-04	2.56e+04	10

Table A.4: Results of running ADSWITCH with 25% relative Gaussian noise

Problem	n	m	avr $f(x)$	avr $\|g_T(x)\|$	avr $\|c(x)\|$	avr #its	#success
BT1	3	1	-8.595453e-01	9.99e-02	4.59e-04	2.67e+02	10
BT2	5	3	+3.258385e-02	7.70e-03	4.76e-04	7.86e+02	10
BT3	3	2	+4.090844e+00	1.06e-01	6.34e-04	—	9
BT4	3	2	-4.550471e+01	2.77e-01	2.28e-04	1.19e+03	10
BT5	5	2	+9.618963e+02	6.57e-01	2.61e-04	5.57e+03	10
BT6	5	3	+2.773363e-01	2.99e-02	5.98e-04	—	0
BT7	5	2	—	—	—	1.14e+03	10
BT8	4	2	+1.000243e+00	1.98e-03	3.41e-04	1.46e+01	10
BT9	2	2	-1.000035e+00	1.17e-02	2.21e-04	1.98e+02	10
BT10	5	3	-1.000038e+00	0.00e+00	5.91e-05	5.00e+00	10
BT11	5	3	+8.258014e-01	1.25e-01	5.58e-04	—	5
BT12	3	2	+6.189020e+00	2.82e-02	3.48e-04	5.23e+02	10
BYRDSPHR	10	5	-4.664093e+00	1.11e-01	1.19e-04	8.67e+02	10
DIXCHLNG	110	55	+2.472434e+03	6.46e+01	5.25e-04	—	5
EIGENA2	110	55	+9.973427e-07	1.86e-03	3.57e-04	4.15e+02	10
EIGENACO	110	55	+2.948147e-07	1.08e-03	9.41e-16	2.81e+02	10
EIGENB2	110	55	+3.675137e-06	1.07e-03	5.37e-04	8.56e+02	10
EIGENBCO	75	25	+2.125972e-06	7.45e-04	5.15e-04	1.94e+03	10
ELEC	10	8	+2.438199e+02	6.84e-01	5.62e-04	—	2
GENHS28	7	2	+9.280133e-01	5.07e-02	6.75e-06	—	1
HS100LNP	2	1	+6.806891e+02	1.84e+00	8.94e-04	—	0
HS6	2	1	+1.921030e-03	1.60e-02	3.73e-04	2.96e+01	10
HS7	2	2	-1.732161e+00	8.73e-04	3.80e-04	9.28e+01	10
HS8	2	1	-1.000000e+00	0.00e+00	1.32e-08	4.00e+00	10
HS9	3	1	-4.968118e-01	1.36e-02	1.10e-14	7.22e+01	10
HS26	3	1	+3.827215e-05	2.15e-03	4.57e-04	1.67e+02	10
HS27	3	1	+4.007659e-02	2.90e-03	2.31e-04	9.16e+01	10
HS28	4	2	+2.028082e-05	2.71e-03	8.44e-16	7.52e+01	10
HS39	4	3	-1.000374e+00	5.82e-03	4.58e-04	1.94e+02	10
HS40	4	2	-1.981579e-01	4.81e-02	1.35e-04	3.81e+02	10
HS42	5	2	+1.391016e+01	5.52e-01	2.13e-04	4.87e+03	10
HS46	5	3	+4.013015e-05	1.67e-03	5.27e-04	1.56e+02	10
HS47	5	2	+5.386687e-05	2.46e-03	5.13e-04	4.49e+02	10
HS48	5	3	+8.988756e-06	4.31e-03	1.39e-15	8.36e+01	10
HS50	5	3	+8.049833e-05	1.25e-02	1.08e-13	5.38e+03	10
HS51	5	3	+2.340881e-07	9.39e-04	1.98e-15	1.68e+01	10
HS52	3	2	+5.328581e+00	1.57e-01	6.34e-05	—	0
HS61	5	2	—	—	—	1.00e+00	10
HS77	5	3	+2.416904e-01	3.22e-02	4.66e-04	—	0
HS78	5	3	-2.914874e+00	1.41e-01	2.74e-04	—	0
HS79	20	18	+7.880086e-02	4.41e-03	5.19e-04	9.58e+02	10
LUKVLE1	20	13	+3.740180e+00	1.06e-02	2.27e-04	1.11e+02	10
LUKVLE2	20	2	+4.371965e+02	8.04e+00	5.47e-04	—	6
LUKVLE3	20	9	+2.758515e+01	1.13e+00	4.96e-04	—	5
LUKVLE4	21	10	+1.859096e+02	2.56e+02	9.65e-05	1.00e+05	10
LUKVLE6	20	4	+1.119752e+03	4.22e+00	5.48e-04	1.00e+05	10
LUKVLE7	20	18	-5.347145e+00	5.86e-01	4.82e-04	—	0
LUKVLE8	20	6	—	—	—	1.90e+02	10
LUKVLE9	20	18	+7.100923e+00	1.90e+01	3.53e-04	1.00e+05	10
LUKVLE10	18	10	+6.459921e+00	2.75e-02	4.97e-04	9.69e+04	10
LUKVLE11	17	12	+1.000256e-04	2.75e-03	5.29e-04	3.72e+03	10
LUKVLE12	18	10	+2.286872e+02	4.44e+01	2.23e+00	1.00e+05	10
LUKVLE13	18	10	+5.501757e+01	3.78e-01	3.38e-04	1.00e+05	10
LUKVLE14	17	12	+4.180692e+04	4.10e+00	3.72e-04	1.00e+05	10
LUKVLE15	17	12	+7.963970e+01	1.27e+01	1.81e-04	—	0
LUKVLE16	17	12	+8.040898e+01	2.71e+01	3.79e+00	—	0
LUKVLE17	17	12	+1.190180e+02	5.48e+01	8.37e+00	—	0
LUKVLE18	20	9	+2.627594e+01	1.71e+01	8.37e+00	—	0
LUKVLI4	2	1	+1.641033e+02	2.16e+02	2.26e-04	1.00e+05	10
MARATOS	5	3	-1.000135e+00	4.39e-03	3.68e-04	3.51e+01	10
MWRIGHT	9	3	+2.498269e+01	4.85e-01	4.89e-04	—	4
ORTHRDM2	9	3	+1.938152e-07	7.12e-04	2.85e-04	2.68e+01	10
ORTHRDS2	133	64	+2.075502e-07	7.50e-04	2.86e-04	2.67e+01	10
ORTHREGA	27	6	+3.503053e+02	2.04e-01	6.55e-04	—	0
ORTHREGB	15	5	+1.903767e-07	7.21e-04	2.66e-04	1.51e+01	10
ORTHREGC	43	20	+7.509051e-07	8.31e-04	5.59e-04	3.67e+01	10
ORTHREGD	43	20	—	—	—	1.65e+02	10
ORTHRGDM	43	20	—	—	—	1.84e+03	10
ORTHRGDS	2	1	+6.212538e+00	9.40e-02	7.00e-04	—	0
S316m322	11	9	—	—	—	0.00e+00	10
SPINOP	11	9	+3.658931e-01	1.59e-02	4.10e-04	1.53e+03	10

Table A.5: Results of running ADSWITCH with 50% relative Gaussian noise