

Adaptive bundle methods for nonlinear robust optimization

Martina Kuchlbauer Frauke Liers Michael Stingl

Abstract

Currently, there are few theoretical or practical approaches available for general nonlinear robust optimization. Moreover, the approaches that do exist impose restrictive assumptions on the problem structure. We present an adaptive bundle method for nonlinear and non-convex robust optimization problems with a suitable notion of inexactness in function values and subgradients. As the worst case evaluation requires a global solution to the adversarial problem, it is a main challenge in a general non-convex nonlinear setting. Moreover, computing elements of an ε -perturbation of the Clarke subdifferential in the ℓ_2 -norm sense is in general prohibitive for this class of problems. In this article, instead of developing an entirely new bundle concept, we demonstrate how existing approaches, such as Noll's bundle method for non-convex minimization with inexact information (Computational and analytical mathematics 50: 555-592, 2013) can be modified to be able to cope with this situation. Extending the non-convex bundle concept to the case of robust optimization in this way, we prove convergence under two assumptions: Firstly, that the objective function is lower C^1 and secondly, that approximately optimal solutions to the adversarial maximization problem are available. The proposed method is hence applicable to a rather general setting of nonlinear robust optimization problems. In particular, we do not rely on a specific structure of the adversary's constraints. The considered class of robust optimization problems covers the case that the worst-case adversary only needs to be evaluated up to a certain precision. One possibility to evaluate the worst case with the desired degree of precision is the use of techniques from mixed-integer linear programming (MIP).

We investigate the procedure on some analytic examples. As applications, we study the gas transport problem under uncertainties in demand and in physical parameters that affect pressure losses in the pipes. Computational results for examples in large realistic gas network instances demonstrate the applicability as well as the efficiency of the method.

1 Introduction

In this work, we consider nonlinear optimization problems that shall be protected against uncertainties. Indeed, an optimization problem may be affected by parameter uncertainties, for example due to the fact that in a real-world application not all parameters can be measured or they may fluctuate. In order to protect against such uncertainties, here we follow a robust optimization approach. A solution vector is called robust if it is feasible regardless of which parameter value is realized within an a priori defined uncertainty set. Such a solution is called robust optimum if it leads to the best guaranteed solution value among all robust solutions.

For nonlinear robust optimization, one of the main challenges consists in the evaluation of the worst case, which requires solving an adversarial problem to global optimality that includes, in general, nonlinearities and non-convexities.

Comprehensive reviews of the state-of-the-art of robust optimization including reformulation approaches for special cases such as robust linear and convex optimization are provided in [BEN09; GYH15; GS16]. In general, nonlinear robust optimization problems are intractable in theory and very difficult to solve in practice. A survey and outer approximation method for this class of problems is presented in [Ley+20]. Another outer approximation approach, using derivative-free optimization, is proposed in [MW20]. Further approximation techniques for nonlinear robust optimization include e.g. a method employing linearization [DBK06] or a convexification method [HD13].

In this paper, we consider nonlinear robust optimization problems that are covered by the following minimax formulation:

$$\min_{x \in \mathbb{R}^n} \max_{u \in \mathcal{U}} v(x, u), \quad (1)$$

with a compact uncertainty set \mathcal{U} and a jointly locally Lipschitz continuous function $v : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ that is lower C^1 [Spi81]. In the setting of (1), the inner maximization problem corresponds to an adversarial worst case evaluation on a compact uncertainty set \mathcal{U} . In Section 3, we detail the classes of nonlinear robust optimization problems that are covered by the problem statement (1). It should already be noted here that we neither impose a convexity assumption on v with respect to x nor a concavity assumption on v with respect to u .

Restrictive assumptions on robust optimization problems, such as linearity or convexity of the involved functions, or uncertainty sets of a tractable structure, allow for reformulation approaches [BHV15]. However, without such assumptions, one has to directly tackle the resulting minimax problem. The concept of bundle methods is particularly suited for the solution of the latter [Mäk02].

However, in the general non-convex situation, this is not a fully straightforward task. The main reason is that realistically, we can only assume that we have access to approximate solutions to the inner maximization problem of (1). The reason for this assumption is that we aim to find a globally optimal solution to the adversarial maximization problem. However, due to possible non-convexities, this is in general not tractable. A viable approach to compute at least an approximation of a global solution to the inner problem is to approximate nonlinearities by piecewise linear relaxation and then solve the problem up to arbitrary exactness using techniques from mixed-integer linear optimization (MIP). It is noted that relaxing the set of feasible solutions in the adversarial problem in this way, we strengthen the adversary and hence robust protection against uncertainties can be maintained. On the other hand such a relaxation however may lead to a more conservative solution. In this article, we overcome this drawback by adaptively refining the relaxations throughout the algorithm.

But this is not the only issue. Another difficulty is that having only access to approximate maximizers of the adversarial problem, we can in general neither compute an element of the exact Clarke subdifferential of the optimal value function

$$f(x) := \max_{u \in \mathcal{U}} v(x, u), \quad (2)$$

nor an element of an ε -perturbation of the latter. However, ε -perturbations of $\partial f(x)$ of the type $\partial f(x) + \varepsilon B(0, 1)$ with $B(0, 1)$ being the Euclidean norm ball, are the basis of essentially

all non-convex bundle methods able to work with inexact information. Instead, the only information which we can - in general - afford to compute is an approximate evaluation $f(x)$ with a prescribed error ε_f along with an element of the set

$$\tilde{\partial}_a f(x) = \text{conv}\{\partial_x v(x, u_x) \mid u_x \in \mathcal{U}, v(x, u_x) \geq \max_{u \in \mathcal{U}} v(x, u) - \varepsilon_f\}. \quad (3)$$

It is readily seen that the latter is only an *outer* approximation of the exact subdifferential. Moreover, given the rather weak assumptions we have imposed for the function v , nothing can be said about the Euclidean distance of an element of $\tilde{\partial}_a f(x)$ to the exact subdifferential $\partial f(x)$.

Keeping this in mind, in the following we briefly comment on the applicability of bundle methods from the literature to robust optimization problems of type (1).

A special situation worth to mention is the case in which the optimal value function f is convex. In this case, it is a straightforward exercise to show that $\tilde{\partial}_a f(x)$ is a subset of the *convex* ε -subdifferential with $\varepsilon = \varepsilon_f$. Having this said, a straightforward consequence is that in principal any convex bundle method able to work with inexact information (in both, function values and subgradients) can be applied. Such methods were first proposed by Kiwiel in [Kiw85] and were later further developed in additional studies. We refer to [OSL14] and references therein for a unified framework and overview. [MOZ17; NB10] for example analyze the case of inexact function value and subgradient evaluations with additional noise. From a theoretical point of view all these methods can be applied to minimax problems without additional convexity assumptions for the inner maximization problem in (1). However, from a practical point of view there is a big difference: while in the convex case, i. e. $v(x, \cdot)$ concave and \mathcal{U} convex, almost arbitrarily precise function values and ε -subgradients can be computed using modern convex optimization tools without much effort, in the general situation each single (approximate) evaluation of the optimal value function can be a challenge. This is why, in this case, the use of an adaptive bundle method such as in [Kiw85] should be preferred over a convex bundle method working with a fixed ε such as in [Kiw06].

Still the situation is more involved in the fully non-convex case, i. e. when we neither impose convexity of f nor convexity of the inner maximization problem. In the literature, generalizations of bundle methods to the case of non-convex functions are discussed in e.g. [HS10; LV98; SZ92]. Here, a key element to deal with non-convexities in the objective function is the so-called downshift mechanism for cutting planes which was first proposed in [SZ92]. While in [SZ92] the authors were still assuming the availability of exact information, nowadays there are also a few non-convex bundle methods available that can deal with inexact function values and subgradients [Nol13; HSS16; LPM18; HU19]. It is however important to understand that essentially all these approaches assume that – for a given $\varepsilon > 0$ – an element of the perturbed Clarke subdifferential $\partial f(x) + \varepsilon B(0, 1)$ is available. Thus, none of these algorithms can be applied for solving (1) without further modification. On the other hand, many elements of these algorithms can be reused for the efficient solution of problem (1). This is the reason why, rather than coming up with a completely new bundle concept, in this article, we demonstrate, how the non-convex inexact bundle method by Noll in [Nol13] can be turned into an adaptive bundle method for the solution of robust optimization problems of type (1). As such, our algorithm can also be seen as an extension of earlier work by Noll et al. [ANP08; ANP09; Nol10; NPR08], which constitutes the basis for the techniques described in [Nol13]. While we formally start from the concept of Noll, our ideas may – with moderate additional effort – be combined with other non-convex inexact

bundle concepts such as described in [HSS16], where a different downshift mechanism is employed or in [HU19], where a generalization of Noll’s bundle method to the case of an infinite-dimensional Hilbert space setting is proposed.

In summary, one of the major contributions of this article is the extension of non-convex bundle methods to render them applicable to nonlinear robust optimization with inexact worst case evaluations and the subdifferential approximation (3). In this, the required approximation error ε_f is dynamically decreased, depending on the current distance of trial and serious iterates. As we are able to allow for large errors in early iterations that are refined iteratively, an efficient algorithm is obtained. We also prove convergence for the resulting adaptive bundle method. The underlying formal assumption is as follows:

Assumption 1.1. *We assume that for any $x \in \mathbb{R}^n$ and any choice of $\varepsilon_f > 0$, we have access to a $u_x \in \mathcal{U}$ with $v(x, u_x) \geq \max_{u \in \mathcal{U}} v(x, u) - \varepsilon_f$. Further, we assume that we have access to $v(x, u_x)$ and to a $g_x \in \partial_x v(x, u_x)$.*

We would like to comment that this assumption essentially only says that we have to be able to solve the inner maximization problem with a given precision. Moreover it does in no way restrict the choice of the tool for the solution of the latter.

We finally note that minimax problems themselves constitute an important class of optimization problems. A review of the theory of discrete and continuous minimax problems as well as methods to solve them can be found in [RH02; DM90] and references therein. In contrast to existing approaches, our method requires neither differentiability nor low-dimensionality, for which case e.g. discretization approaches are suitable. Problem (1) can be reformulated to include uncertainties in the constraints and thus be understood as an optimization problem with infinitely many constraints. Thus, our algorithm could also be of interest for this field. We refer to [Ste12] and references therein for a general overview of semi-infinite optimization and a description of how this field relates to robust optimization. How techniques from this field can be used for minimax problems is addressed in [ŽR03]. In the context of combining these two fields, we also refer to the bundle method in [Fud+15; GGM06] which treats the convex case with inexact solutions to the inner maximization problem.

This work is structured as follows. In Section 2, we introduce the approximate convexity concepts and subdifferentials used, and in Section 3, we detail the considered setting of nonlinear robust optimization with the corresponding notion of inexactness. The adaptive bundle algorithm and convergence results are presented in Section 4. We thereby first obtain convergence of the inner and then of the outer loop. This section concludes with some notes on stopping criteria. In Section 5, we show that the algorithm is applicable to the robust protection of the gas transport problem under uncertainty. This task is in fact a two-stage robust optimization problem with a non-convex second stage. Exploiting the fact that the second-stage solution is unique, we demonstrate that the required properties for correctness of the algorithm are satisfied. We thereby approximately solve the non-convex adversarial problem using an adaptively refined piecewise linear relaxation. We also briefly outline the applicability of the method for robust protection of electricity networks in the alternating current power model. Finally, in Section 6, we demonstrate the practical applicability and efficiency of our method by presenting numerical results for an analytical example and realistic gas network instances.

2 Preliminaries

In the following, we develop the proposed methodology for nonlinear robust optimization problems. To prove convergence of the adaptive bundle method, we will assume the locally Lipschitz objective function to be lower C^1 and hence approximate convex. We clarify the term of approximate convexity by stating the definition with the help of the concept of ε' -convexity:

Definition 2.1. (*ε' -convex functions, [JLT98]*) A function $f : X \rightarrow \mathbb{R} \cup \{\pm\infty\}$ with $X \subseteq \mathbb{R}^n$ is ε' -convex if

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) + \varepsilon' \lambda(1 - \lambda) \|x - y\| \quad (4)$$

for each $x, y \in X$ and $\lambda \in [0, 1]$. This implies that for any subgradient $g \in \partial f(x)$ and any $y \in X$:

$$g^T(y - x) \leq f(y) - f(x) + \varepsilon' \|x - y\|. \quad (5)$$

Definition 2.2. (*approximate convex functions, [VLT00]*) Let $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ with $X \subseteq \mathbb{R}^n$ be lower semi-continuous and for any $\delta > 0$ let the function f_δ be defined by $f_\delta(x) = f(x)$ if $x \in B(\bar{x}, \delta)$, and $+\infty$ otherwise. f is approximate convex at $\bar{x} \in X$ if for every $\varepsilon' > 0$, there exists a $\delta > 0$ such that f_δ is ε' -convex.

The above definition and terminology suggest the assumption of approximate convexity to be relatively restrictive. However, for the case of locally Lipschitz functions, the following definition and equivalence result show that approximate convexity covers the large class of lower C^1 functions.

Definition 2.3. (*lower C^1 functions, [Spi81]*) A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is lower C^1 (LC^1) if for any \bar{x} , there are a neighborhood $U \ni \bar{x}$, a compact set Y and a function $v : U \times Y \rightarrow \mathbb{R}$ with v and $\nabla_x v$ jointly continuous in x and y , such that for every $x \in U$,

$$f(x) = \max_{y \in Y} v(x, y).$$

Proposition 2.4. (*[DG04]*) A locally Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is lower C^1 if and only if it is approximate convex.

Note that in this paper we work with locally Lipschitz functions, so that we use either the approximate convex property or the equivalent LC^1 property. In the following, when we refer to subdifferential, we mean the Clarke subdifferential:

Definition 2.5. (*[Cla90]*) The generalized directional derivative of a locally Lipschitz continuous function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined by

$$f^\circ(x; d) = \limsup_{\substack{x' \rightarrow x, \\ t \downarrow 0}} \frac{f(x' + td) - f(x')}{t}. \quad (6)$$

The generalized gradient, called Clarke subdifferential, is defined by

$$\partial f(x) = \{g \in \mathbb{R}^n \mid g^T d \leq f^\circ(x; d) \forall d \in \mathbb{R}^n\}. \quad (7)$$

In the remainder of this paper, we also make use of the following results:

Lemma 2.6. ([Cla90]) *The following properties hold for the Clarke subdifferential $\partial f(x)$ of a locally Lipschitz function $f : \mathbb{R}^n \rightarrow \mathbb{R}$:*

1. *For $x \in \mathbb{R}^n$, $\partial f(x)$ is nonempty, convex and compact.*
2. *The mapping $\partial f(\cdot)$ is locally bounded.*
3. *The mapping $\partial f(\cdot)$ is upper semi-continuous.*

Lemma 2.7. ([Cla75; Roc82]) *If a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is lower C^1 , then it is subdifferentially regular, which means that for every $x \in \mathbb{R}^n$ and $d \in \mathbb{R}^n$ the directional derivative,*

$$f'(x; d) = \limsup_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}, \quad (8)$$

exists and $f'(x; d) = f^\circ(x; d)$. Furthermore, $f'(x; d)$ is upper semi-continuous in x .

3 Nonlinear robust optimization with a notion of inexactness

We consider robust optimization problems that are formulated as a minimax problem with an objective function $v : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ that is locally Lipschitz continuous and lower C^1 , and a compact set \mathcal{U} :

$$\min_x \max_{u \in \mathcal{U}} v(x, u). \quad (9)$$

First, the problem (9) covers the case of a robust optimization problem with parameter uncertainties that affect the objective function value:

$$\begin{aligned} \min_x \quad & t \\ \text{s. t.} \quad & c(x, u) \leq t \quad \forall u \in \mathcal{U}. \end{aligned} \quad (10)$$

Choosing $v(x, u) := c(x, u)$, this problem is equivalent to (9).

Second, the following case of parameter uncertainties that affect constraints can be treated by the problem statement (9):

$$\begin{aligned} \min_x \quad & c(x) \\ \text{s. t.} \quad & g_i(x, u) \leq 0 \quad \forall u \in \mathcal{U}, i \in 1, \dots, N. \end{aligned} \quad (11)$$

Penalizing infeasibility by an l_1 penalty term with penalty parameter ψ , we obtain an unconstrained minimax problem:

$$\min_x \max_{u \in \mathcal{U}} c(x) + \psi \sum_{i=1}^N [g_i(x, u)]_+, \quad (12)$$

$[\cdot]_+$ denoting the positive part. This problem is covered by (9) by choosing $v(x, u) := c(x) + \psi \sum_{i=1}^N [g_i(x, u)]_+$.

It can be seen that e.g. for continuously differentiable functions $c(\cdot, \cdot)$ or $c(\cdot)$ and $g(\cdot, \cdot)$, the above choices of $v(\cdot, \cdot)$ meet the assumption of being locally Lipschitz continuous and lower C^1 . Furthermore, the settings in (10) and (11) can also be combined such that the objective function value as well as constraints are affected by uncertainties.

Our approach to solve the minimax problem (9) is to treat it as a minimization problem with a nonsmooth, non-convex objective function. The objective function of the minimization problem is the optimal value function of the adversarial maximization problem and is denoted by $f(x) = \max_{u \in \mathcal{U}} v(x, u)$ as defined in (2).

To evaluate a function value $f(x)$, a non-convex optimization problem hence has to be solved to global optimality. As this global optimization problem may be intractable, we make the more realistic assumption that the inner maximization problem cannot be solved exactly within a reasonable time period, but can be solved approximately in the sense that ε_f -optimal solutions can be computed for any given ε_f . This can be realized for example by piecewise linear relaxation. For a given error bound ε_f , we hence approximate the optimal value function $f(x)$ by $v(x, u_x)$ with

$$u_x \in \{u \in \mathcal{U} \mid v(x, u_x) \geq \max_{u \in \mathcal{U}} v(x, u) - \varepsilon_f\}, \quad (13)$$

so that $f(x) - \varepsilon_f \leq v(x, u_x) \leq f(x)$. For technical reasons that will be explained later, we prefer to use an overestimator of the optimal value function. In order to evaluate an approximation $f_a(x)$ of $f(x)$ for a given ε_f , we thus choose a u_x from the set in (13) and construct the following overestimator:

$$f_a(x) := v(x, u_x) + \varepsilon_f. \quad (14)$$

The exact subdifferential of f at x is calculated as (see e.g. [GD82]):

$$\partial f(x) = \text{conv}\{\partial_x v(x, u_x^*) \mid u_x^* \in \mathcal{U}, v(x, u_x^*) = \max_{u \in \mathcal{U}} v(x, u)\}. \quad (15)$$

Having at hand $u_x \in \mathcal{U}$ with $f_a(x) = v(x, u_x) + \varepsilon_f$, we choose as an approximate subgradient

$$g_x \in \partial_x v(x, u_x). \quad (16)$$

This approximate subgradient is an element of the following outer approximation of the exact subdifferential:

$$\tilde{\partial}_a f(x) := \text{conv}\{\partial_x v(x, u) \mid u \in \mathcal{U}, v(x, u) \geq v(x, u_x)\}. \quad (17)$$

We note that the approximate subdifferential $\tilde{\partial}_a f(\cdot)$ differs from an ε -subdifferential, as used e.g. in [Nol13]:

$$\partial_\varepsilon f(x) = \partial f(x) + \varepsilon B(0, 1). \quad (18)$$

By this definition, it is ensured that an element of the ε -subdifferential is in an ε -neighborhood of the exact subdifferential. Comparing $\partial_\varepsilon(\cdot)$ and $\tilde{\partial}_a f(\cdot)$, we observe that inclusion does not hold in either direction. As we consider inexact worst case evaluations of type (13), we only have the outer approximation at hand and cannot use Noll's bundle method that requires to have (18) at hand. Thereby, the exact subdifferential's property of local boundedness from Lemma 2.6 is preserved by the approximation (17):

Corollary 3.1. *The mapping $\tilde{\partial}_a f(\cdot)$ is locally bounded.*

Proof. The claim follows from Lemma 2.6, 2. applied to the function $v(x, u)$, using its local Lipschitz continuity and boundedness of \mathcal{U} . \square

In practice, to compute an approximate function value and a corresponding approximate subgradient for a given error bound ε_f , we do the following: We compute an approximate solution u_x to the adversarial maximization problem, the corresponding function value $v(x, u_x)$ and a subgradient g_x of $v(\cdot, u_x)$ at the point x . We thus require that $v(\cdot, u)$ and $\partial v(\cdot, u)$ can be evaluated exactly for each possible realization $u \in \mathcal{U}$.

An example for such a setting is a non-convex adversarial problem $\max_{u \in \mathcal{U}} v(x, u)$ which can be solved approximately using linearization techniques that provide underestimators for the optimal function value. For a given $u \in \mathcal{U}$, however, the objective value and an approximate subgradient in the sense of (17) can be evaluated exactly. In Section 5, we will see that the gas transport problem under uncertainty fits into this framework.

4 An adaptive bundle method

We now present an adaptive bundle method which is applicable to the nonlinear robust optimization problems with approximate worst case evaluations and corresponding subgradients in the sense of Section 3. Although some algorithmic concepts and proof techniques employed are inspired by [Nol13], our notion of inexactness differs, see Section 3, and requires alternative arguments at crucial steps. Therefore, to provide a complete presentation of the results, we summarize the relevant arguments from [Nol13]. Throughout this section and especially in the proofs in Appendix A in the online supplement, we will describe how each step compares to Noll’s work with a particular emphasis on where they differ. We obtain convergence results under the assumptions made in Section 3.

We start with fixing the notation. A bundle method is divided into one outer loop and a number of inner loops: The outer loop aims to find serious iterates, using inner loops. In an inner loop, new trial iterates around the current serious iterate are determined and serve as candidates for the next serious iterate.

Notation 4.1. *We denote trial iterates by $x^{k,j}$ with inner loop counter k and outer loop index j as superscripts. If the outer loop index is clear from the context, we denote trial iterates as x^k . In the remainder of this section, we define some concepts such as e.g. cutting planes. They each correspond to a certain inner loop which is indicated by an index k . We proceed with these in the same way and omit the outer loop index j from the beginning if it is clear from the context. We denote outer iterates by x_j , with outer loop counter j as subscript, or simply by x if it is clear that we mean the current serious iterate.*

We evaluate the objective function and hence solve the inner maximization problem at a trial iterate x^k with an adaptive error bound $\varepsilon_f(x^k, x)$. This error bound linearly depends on the trial iterate’s distance to the current serious iterate x with an algorithmic parameter ε'' :

$$\varepsilon_f(x^k, x) = \varepsilon'' \|x - x^k\|. \quad (19)$$

We note that the parameter ε'' can be chosen arbitrarily but also plays an important role in our convergence results. We further note that the error bound now depends not only on the point at which the function value is approximated. It also depends on the serious iterate of the outer loop within which the approximation is evaluated.

Notation 4.2. *We denote $\varepsilon_f(x^k, x)$ by $\varepsilon_f^{x^k}$ or ε_f^k if x is clear from the context. For two serious iterates $x = x_j, x_{j-1}$, we denote $\varepsilon_f(x_j, x_{j-1})$ by ε_f^x if the outer loop index is clear from the context.*

Access to an $\varepsilon_f(x^k, x)$ -exact function value approximation at any trial iterate x^k is ensured by Assumption 1.1 in Section 1. We recall that by Assumption 1.1, we have for any $x \in \mathbb{R}^n$ and any choice of $\varepsilon_f > 0$, access to a $u_x \in \mathcal{U}$ with $v(x, u_x) \geq \max_{u \in \mathcal{U}} v(x, u) - \varepsilon_f$. Following [Nol13], we now summarize some algorithmic concepts that we will require later. To find a new trial iterate, the function f is approximated around the current serious iterate x by a convex working model. Therefore, let $g_x \in \partial_x v(x, u_x)$ be the corresponding approximate subgradient of f at x . Then, the affine linear function

$$m_0(\cdot, x) := f_a(x) + g_x^T(\cdot - x) \quad (20)$$

is called the exactness plane at x . At trial iterates x^k around the serious iterate x , we define further cutting planes

$$m_k(\cdot, x) := t_k(\cdot) - s_k, \quad (21)$$

where $t_k(\cdot)$ denotes the tangent plane at x^k , and s_k denotes the downshift with respect to the serious iterate x :

$$t_k(\cdot) := f_a(x^k) + g_k^T(\cdot - x^k), \quad s_k := [t_k(x) - f_a(x)]_+ + c\|x^k - x\|^2. \quad (22)$$

Here, c denotes an algorithmic constant.

Now, as a convex working model, we choose

$$\phi_k(\cdot, x) := \max\{m_l(\cdot, x) \mid 0 \leq l \leq k-1\}. \quad (23)$$

A second-order working model is further defined by

$$\Phi_k(\cdot, x) = \phi_k(\cdot, x) + \frac{1}{2}(\cdot - x)^T Q(x)(\cdot - x), \quad (24)$$

where a symmetric and possibly non positive-definite $Q(x) \in \mathbb{R}^{n \times n}$ may be constructed using second-order information which may be available for f at the serious iterate x . The corresponding local underestimators that include second-order information are denoted by

$$M_k(\cdot, x) = m_k(\cdot, x) + \frac{1}{2}(\cdot - x)^T Q(x)(\cdot - x). \quad (25)$$

To find a new trial iterate x^k in a neighborhood of a current serious iterate x , we solve the convex program

$$\min_{z \in \mathbb{R}^n} \Phi_k(z, x) + \frac{\tau_k}{2} \|z - x\|^2, \quad (26)$$

where $\tau_k \in \mathbb{R}_+$ denotes a proximity control parameter. Throughout the algorithm, $Q(x) + \tau_k \mathcal{I} \succ 0$ is ensured by setting $\tau_k > \lambda_{\min}(Q(x))$. Since x^k solves (26), it follows that

$$g_k^* := (Q(x) + \tau_k \mathcal{I})(x - x^k) \in \partial \phi_k(x^k, x), \quad (27)$$

which is called the aggregate subgradient. In the convergence proof, we will make use of this subgradient and its properties. Furthermore, the aggregate subgradient can be used for aggregation strategies. These strategies aim to retain information but not every single cutting plane from the current and former outer loops in the convex working model and were proposed in [Kiw83]. We refer to [Nol13; NPR08; Noll10] for further explanations of aggregation strategies.

Acceptance of a trial iterate x^k as a new serious iterate x_{j+1} is decided by the following acceptance test, which depends on an algorithmic parameter $0 < \gamma < 1$:

$$\rho_k := \frac{f_a(x_j) - f_a(x^k)}{f_a(x_j) - \Phi_k(x^k, x_j)} \geq \gamma. \quad (28)$$

An x^k is thus accepted if the predicted descent (denominator) is met by the actual descent (numerator), up to a certain ratio γ . Another element of the presented bundle method is proximity control in the course of an inner loop, which we adopt from [NPR08]: Using an algorithmic parameter $\tilde{\gamma}$ with $\gamma < \tilde{\gamma} < 1$, we increase the proximity control parameter τ_k according to

$$\tilde{\rho}_k := \frac{f_a(x_j) - M_k(x^k, x_j)}{f_a(x_j) - \Phi_k(x^k, x_j)}, \quad \tau_{k+1} = \begin{cases} 2\tau_k & \tilde{\rho}_k \geq \tilde{\gamma}, \\ \tau_k & \tilde{\rho}_k < \tilde{\gamma}. \end{cases} \quad (29)$$

4.1 Algorithm

With these ingredients, we state the main algorithm in Algorithm 1. While based on the algorithm in [Nol13], it uses inexact function values and approximate subgradients in Steps 6 and 14, respectively, corresponding to our notion of inexactness as outlined in Section 3.

4.2 Convergence of the algorithm

In order to prepare a convergence proof, we now construct a global overestimator for the convex working model $\phi_k(\cdot, x)$ at a serious iterate x . In detail, we define for a serious iterate x a convex function $\phi(\cdot, x)$, a first-order model of f . It is the pointwise supremum of cutting planes that possibly arise in the current inner loop. In detail, we consider for possible trial iterates z cutting planes that are constructed with the help of an approximate worst case evaluation:

$$m_z(\cdot, x) = f_a(z) + g_z^T(\cdot - z) - s_z. \quad (30)$$

Here, $m_z(\cdot, x)$ is a cutting plane at z with an approximate function value $f_a(z)$ and an approximate subgradient $g_z \in \partial_x v(x, u_z)$. The cutting planes are downshifted by s_z with respect to x and $f_a(x)$ as defined in (21),(22). We note that for every fixed z , several cutting planes, which differ in their worst-case evaluations u_z and hence approximate function values and subgradients, may contribute to the supremum. In detail, we have

$$u_z \in \{u \in \mathcal{U} \mid v(x, u_z) \geq f(z) - \varepsilon'' \|x - z\|\} \quad (31)$$

In addition, exactness planes contribute to the supremum: By $m_0(\cdot, x)$, we denote exactness planes at x with approximate subgradients $g_x \in \tilde{\partial}_a f(x)$ without downshift. Overall, the first-order model is defined as:

$$\begin{aligned} \phi(\cdot, x) := \sup \left\{ \left\{ m_z(\cdot, x) \mid \right. \right. & u_z \in \mathcal{U}, v(x, u_z) \geq f(z) - \varepsilon'' \|x - z\|, \\ & z \in B(0, M) \setminus \{x\} \cup \\ & \left. \left. \left\{ m_0(\cdot, x) \mid u_x' \in \mathcal{U}, v(x, u_x') = v(x, u_x) \right\} \right\} \right\}. \end{aligned} \quad (32)$$

M is assumed to be chosen such that $B(0, M)$ contains x and all trial iterates which possibly occur in an inner loop corresponding to x . The existence of such an M can be justified by

Algorithm 1 Adaptive bundle method

- 1: Fix parameters: $0 < \gamma < \Gamma < 1, \gamma < \tilde{\gamma} < 1, 0 < q < T < \infty, \varepsilon'' > 0, \varepsilon_f^{x^1} > 0, \zeta > 0$.
 - 2: Choose initial values: Starting point x_1 , $Q_1 = Q_1^T$ s.t. $-q\mathcal{I} \preceq Q_1 \preceq q\mathcal{I}$, $\tau_1^\#$ s.t. $Q_1 + \tau_1^\# \mathcal{I} \succ 0$. Set $j = 1$.
 - 3: **while** $\text{dist}(0, \tilde{\partial}_a f(x_j)) \leq \tilde{\varepsilon}$ not fulfilled **do** ▷ Stopping test
 - 4: Initialize inner loop with serious iterate x_j : $\tau_1 = \tau_j^\#$. Set $k = 1$.
 - 5: Solve the program ▷ Trial step generation

$$\min_{x^k \in \mathbb{R}^n} \Phi_k(x^k, x_j) + \frac{\tau_k}{2} \|x^k - x_j\|^2.$$
 - 6: Compute $f_a(x^k)$ with $\varepsilon_f^k = \varepsilon'' \|x_j - x^k\|$. ▷ Adaptive function value approximation
 - 7: **if** $\rho_k = \frac{f_a(x_j) - f_a(x^k)}{f_a(x_j) - \Phi_k(x^k, x_j)} \geq \gamma$ **then** ▷ Acceptance test
 - 8: $x_{j+1} \leftarrow x^k$.
 - 9: $\tau_{j+1}^\# = \begin{cases} \tau_k & \rho_k < \Gamma, \\ 1/2\tau_k & \rho_k \geq \Gamma. \end{cases}$ ▷ Update $\tau^\#$
 - 10: Set Q_{j+1} s.t. $Q_{j+1} = Q_{j+1}^T, -q\mathcal{I} \preceq Q_{j+1} \preceq q\mathcal{I}$. ▷ Update Q_j
 - 11: If $Q_{j+1} + \tau_{j+1}^\# \mathcal{I} \not\succeq \zeta \mathcal{I}$: Increase $\tau_{j+1}^\#$ s.t. $Q_{j+1} + \tau_{j+1}^\# \mathcal{I} \succ \zeta \mathcal{I}$.
 - 12: $\tau_{j+1}^\# = \min\{\tau_{j+1}^\#, T\}$.
 - 13: **else**
 - 14: Generate a cutting plane $m_k(\cdot, x_j)$ with $g_k \in \tilde{\partial}_a f(x^k)$.
 - 15: **if** $\tilde{\rho}_k = \frac{f_a(x_j) - M_k(x^k, x_j)}{f_a(x_j) - \Phi_k(x^k, x_j)} \geq \tilde{\gamma}$ **then** ▷ Update τ_k
 - 16: $\tau_{k+1} \leftarrow 2\tau_k$.
 - 17: **end if**
 - 18: Build new working model ϕ_{k+1} : add $m_k(\cdot, x_j)$ to ϕ_k . ▷ Update ϕ_k
 - 19: Increase k and go to Step 5: Trial step generation.
 - 20: **end if**
 - 21: Increase j .
 - 22: **end while**
-

the proof of Lemma 4.8 in [Nol13, Lemma 4, proof part ii)] and part i) of the proof of Lemma 4.7 in Appendix A in the online supplement.

An example for the first-order model $\phi(\cdot, x)$ is depicted in Figure 1. We now transfer [Nol13, Lemma 2] to our setting. The difference is that we use (32) instead of the first-order model from [Nol13], where the exact subdifferential is approximated by (18).

Lemma 4.3. *For the first-order model $\phi(\cdot, x)$, defined in (32), the following properties hold:*

1. $\phi(x, x) = f_a(x)$,
2. $\partial\phi(x, x) \subseteq \tilde{\partial}_a f(x)$,
3. $\phi_k(\cdot, x) \leq \phi(\cdot, x)$ for the working model $\phi_k(\cdot, x)$.

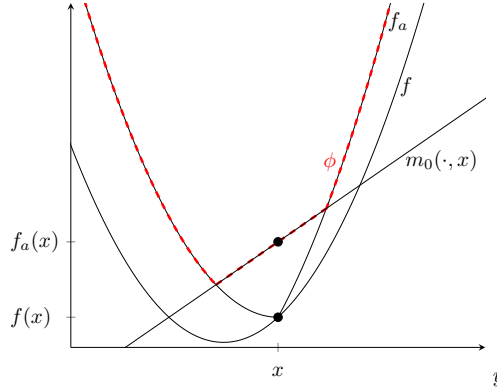


Figure 1: An example of the function $\phi(\cdot, x)$

Proof. 1.) $\phi(x, x) \geq f_a(x)$ since the exactness planes $m_{0,g}(\cdot, x)$ contribute to the supremum at x . $\phi(x, x) \leq f_a(x)$ follows from the downshift of the involved cutting planes with respect to $f_a(x)$.

2.) Let $g \in \partial\phi(x, x)$, so that for all $y \in \mathbb{R}^n$, $g^T(y - x) \leq \phi(y, x) - \phi(x, x)$, since $\phi(y, x)$ is convex in y . For an arbitrary fixed $h \in \mathbb{R}^n$ and any $t > 0$, there are $z_t \in B(0, M)$ such that

$$\phi(x + th, x) = m_{z_t, g_t}(x + th, x), \quad (33)$$

where m_{z_t, g_t} denotes the cutting plane at z_t with an approximate subgradient g_t . For a fixed t we define $y := x + th$ and $m(\cdot, x) := f_a(x) + g^T(\cdot - x)$. From convexity of ϕ and 1.), it follows that $\phi(y, x) = m_{z_t, g_t}(y, x) \geq m(y, x)$. We have from the downshift at z_t that

$$m_{z_t, g_t}(x, x) \leq f_a(x) - c\|x - z_t\|^2 = m(x, x) - c\|x - z_t\|^2. \quad (34)$$

From these inequalities for the cutting planes m and m_{z_t, g_t} at x and y , we obtain

$$g^T h \leq g_t^T h. \quad (35)$$

Moreover, let $t \rightarrow 0^+$, so that $\phi(x + th, x) \rightarrow \phi(x, x)$. Consequently, with (33) and 1.), $m_{z_t, g_t}(x + th, x) \rightarrow f_a(x)$. As $m_{z_t, g_t}(x + th, x)$ and $m_{z_t, g_t}(x, x)$ converge to the same limit, we infer from (34) that $z_t \rightarrow x$. From convergence and thus boundedness of z_t , with Corollary

3.1 we obtain boundedness of the sequence $g_t \in \tilde{\partial}_a f(z_t)$. Hence, there is a subsequence with $g_t \rightarrow \hat{g}$.

We now distinguish two cases:

i) First, we assume finite convergence of $z_t \rightarrow x$. An example of this case with $f_a(x) > f(x)$ is depicted in Figure 1. Finite convergence implies that there is a \bar{t} such that for all $t \leq \bar{t}$, $z_t = x$ and thus $g_t \in \tilde{\partial}_a f(x)$. With (35), as h was chosen arbitrarily, $g \in \tilde{\partial}_a f(x)$ and thus the claim $\partial\phi(x, x) \subseteq \tilde{\partial}_a f(x)$ follows from convexity of $\tilde{\partial}_a f(x)$.

ii) Second, we assume that there is no finite convergence of $z_t \rightarrow x$. It can be shown that this is exactly the case if $f_a(x) = f(x)$. From $f(z_t) \leq f_a(z_t) \leq f(z_t) + \varepsilon'' \|x - z_t\|$, it follows that for $t \rightarrow 0^+$, $f_a(z_t) \rightarrow f(x)$. Let u_{z_t} and u_x be such that $f_a(z_t) = v(z_t, u_{z_t}) + \varepsilon'' \|x - z_t\|$ and $f(x) = v(x, u_x)$, so that $v(z_t, u_{z_t}) \rightarrow v(x, u_x)$. As the sequence of u_{z_t} lies in \mathcal{U} and is thus bounded, we can choose a convergent subsequence with $u_{z_t} \rightarrow u^*$. Using $v(z_t, u_{z_t}) \rightarrow v(x, u^*)$, we infer $v(x, u_x) = v(x, u^*)$ and thus $\partial_x v(x, u^*) \subseteq \tilde{\partial}_a f(x)$. From $g_t \in \partial_x v(z_t, u_{z_t})$, it follows for an arbitrary $d \in \mathbb{R}^n$, denoting $v(\cdot, u)$ by $v_u(\cdot)$:

$$\begin{aligned} g_t^T d &\stackrel{\text{Def. 2.5}}{\leq} v_{u_{z_t}}^\circ(z_t; d) \stackrel{\text{Lemma 2.7}}{=} v'_{u_{z_t}}(z_t; d) = \lim_{t \downarrow 0} \frac{v_{u_{z_t}}(z_t + td) - v_{u_{z_t}}(z_t)}{t} \\ &= \lim_{t \downarrow 0} \frac{v(z_t + td, u_{z_t}) - v(z_t, u_{z_t})}{t} = v'((z_t, u_{z_t}); (d, 0)). \end{aligned}$$

Lemma 2.7 implies upper semi-continuity of $v'(\cdot, (d, 0))$ so that using convergence of $z_t \rightarrow x$ and $u_{z_t} \rightarrow u^*$, we infer

$$\limsup_{(z_t, u_{z_t}) \rightarrow (x, u^*)} v'((z_t, u_{z_t}); (d, 0)) \leq v'((x, u^*); (d, 0)). \quad (36)$$

Using $g_t^T d \leq v'((z_t, u_{z_t}); (d, 0))$, which holds for every g_t in the considered sequence and hence for the case of passing to the limit, it follows from $g_t \rightarrow \hat{g}$ that

$$\hat{g}^T d \leq v'((x, u^*); (d, 0)) = v'_{u^*}(x, d) \stackrel{\text{Lemma 2.7}}{=} v_{u^*}^\circ(x, d). \quad (37)$$

As d was chosen arbitrarily, it follows from Definition 2.5 that $\hat{g} \in \partial_x v(x, u^*) \subseteq \tilde{\partial}_a f(x)$. With (35), we have that $g^T h \leq \hat{g}^T h$. As h was chosen arbitrarily, the claim follows from convexity of $\tilde{\partial}_a f(x)$. This can be seen using the hyperplane separation theorem.

3.) Follows from the definitions of ϕ and ϕ_k . \square

In the convergence proof, we exploit perturbed subgradient inequalities that are similar to the one ensured by ε' -convexity of a function, see (5) in Definition 2.1. In detail, the assumption of $v(\cdot, \cdot)$ being lower C^1 and thus approximate convex ensures ε' -convexity for an arbitrary $\varepsilon' > 0$ on a suitable region:

Lemma 4.4. *For every $x_0 \in \mathbb{R}^n$ and $\varepsilon' > 0$, there is a $\delta > 0$ such that for every fixed $u \in \mathcal{U}$, $v(\cdot, u)$ is ε' -convex on $B(x_0, \delta)$.*

Proof. Given x_0 and $\varepsilon' > 0$, it follows from approximate convexity that for a fixed $u \in \mathcal{U}$, there is a $\delta_u > 0$ such that $v(\cdot, u)$ is ε' -convex on $B(x_0, \delta_u)$. As \mathcal{U} is compact, there exists a lower bound $\delta > 0$ with $\delta \leq \delta_u$ for all $u \in \mathcal{U}$. Hence, $v(\cdot, u)$ is ε' -convex on $B(x_0, \delta)$. \square

We will later use Lemma 4.4 to obtain ε' -convexity for arbitrary $\varepsilon' > 0$ on certain regions. If ε' -convexity is fulfilled on a suitable region, we obtain the following properties of the approximate subdifferential $\tilde{\partial}_a f(x)$ by employing the problem's minimax structure. These can be interpreted as perturbed subgradient inequalities.

Lemma 4.5. *Let x^k be a trial iterate corresponding to a serious iterate x and $g_k \in \tilde{\partial}_a f(x^k)$ be an approximate subgradient. Then, if for every fixed $u \in \mathcal{U}$, $v(\cdot, u)$ is ε' -convex on $\mathcal{X} \subseteq \mathbb{R}^n$ and $x^k \in \mathcal{X}$,*

1. $g_k^T(z - x^k) \leq f_a(z) - f_a(x^k) + \varepsilon_f^k + \varepsilon' \|x^k - z\|$ for all $z \in \mathcal{X}$,
2. $g_k^T(z - x^k) \leq f(z) - f(x^k) + \varepsilon_f^k + \varepsilon' \|x^k - z\|$ for all $z \in \mathcal{X}$.

If also $x \in \mathcal{X}$,

3. $f_a(x^k) - (f_a(x) + g_k^T(x^k - x)) \leq (\varepsilon' + \varepsilon'') \|x^k - x\|$.

Proof. 1.) Let $u_k \in \mathcal{U}$ be such that $f_a(x^k) = v(x^k, u_k) + \varepsilon_f^k$. For $g_k \in \partial_x v(x^k, u_k)$, using $v(z, u_k) \leq f(z) \leq f_a(z)$, we have

$$\begin{aligned} g_k^T(z - x^k) &\leq v(z, u_k) - v(x^k, u_k) + \varepsilon' \|x^k - z\| \\ &\leq f_a(z) - f_a(x^k) + \varepsilon_f^k + \varepsilon' \|x^k - z\|. \end{aligned} \quad (38)$$

By convexity, the desired inequality holds for all $g_k \in \tilde{\partial}_a f(x^k)$.

2.) Let $u_k \in \mathcal{U}$ be such that $f_a(x^k) = v(x^k, u_k) + \varepsilon_f^k$. For $g_k \in \partial_x v(x^k, u_k)$, using $v(z, u_k) \leq f(z)$ and $v(x^k, u_k) \geq f(x^k) - \varepsilon_f^k$, we have

$$\begin{aligned} g_k^T(z - x^k) &\leq v(z, u_k) - v(x^k, u_k) + \varepsilon' \|x^k - z\| \\ &\leq f(z) - f(x^k) + \varepsilon_f^k + \varepsilon' \|x^k - z\|. \end{aligned} \quad (39)$$

The desired inequality thus also holds for all $g_k \in \tilde{\partial}_a f(x^k)$.

3.) By substituting in a trial iterate x^k and serious iterate x , we obtain from 1.) that

$$\begin{aligned} g_k^T(x - x^k) &\leq f_a(x) - f_a(x^k) + \varepsilon_f^k + \varepsilon' \|x^k - x\| \\ &\leq f_a(x) - f_a(x^k) + \varepsilon'' \|x^k - x\| + \varepsilon' \|x^k - x\|, \end{aligned} \quad (40)$$

so the claim follows. \square

Remark 4.6. *We note that Lemma 4.5, 3., provides a bound on*

$$f_a(x^k) - (f_a(x) + g_k^T(x^k - x)) = [t_k(x) - f_a(x)]_+, \quad (41)$$

which is a part of the downshift of a cutting plane at x^k with respect to $f_a(x)$. In detail, the downshift is in (22) defined as $s_k = [t_k(x) - f_a(x)]_+ + c \|x^k - x\|^2$.

By Lemma 4.5, 3., one hence obtains a distance dependent bound on the whole downshift s_k . The term $f_a(x) - t_k(x)$ is also called linearization error in the literature (see e.g. [HS10]) and due to non-convexities it can be negative.

We provide the rather technical proofs for the following convergence results in Appendix A in the online supplement. Before stating them here, we give here a rough summary of the line of argumentation in the proofs: After showing convergence of iterates, Lemma 4.4 can be employed to ensure ε' -convexity on a region containing current and forthcoming iterates for an arbitrary $\varepsilon' > 0$. Consequently, Lemma 4.5, 3., can be used to obtain a distance dependent bound on corresponding downshifts. Together with convergence of

iterates, this implies (at least linear) convergence of the downshifts to 0. Especially, the term $[t_k(x) - f_a(x)]_+$ converges to 0. Since this term may be positive due to non-convexities, its convergence to 0 can be interpreted as follows: The impact of non-convexities vanishes in the limit.

The factor ε'' finally shows up in first convergence results (Lemma 4.7 and Theorem 4.11). Later, it vanishes by choosing ε'' to be not constant but a sequence converging to 0 and leads to sharpened convergence results in Corollaries 4.12 and 4.13.

We now start by obtaining finite convergence of the inner loop in Algorithm 1. In detail, we start with the case where $\tau_k \rightarrow \infty$ and prove the following result in Appendix A.1 in the online supplement:

Lemma 4.7. *Let x be the current serious iterate. If the inner loop does not terminate finitely and $\tau_k \rightarrow \infty$, then it holds that*

$$\text{dist}(0, \tilde{\partial}_a f(x)) \leq \eta,$$

$$\text{with } \eta = \frac{\varepsilon''}{(\bar{\gamma} - \gamma)}.$$

If the proximity parameter τ_k stays bounded, adopting an argument analogous to [Nol13, Lemma 4, Lemma 7], we obtain:

Lemma 4.8. [Nol13] *Let x be the current serious iterate. If the inner loop does not terminate finitely and τ_k is bounded, then $x^k \rightarrow x$ and*

$$0 \in \partial\phi(x, x).$$

Lemma 4.8 together with $\partial\phi(x, x) \subseteq \tilde{\partial}_a f(x)$ from Lemma 4.3, 2., leads to the following corollary.

Corollary 4.9. *Let x be the current serious iterate. If the inner loop does not terminate finitely and τ_k is bounded, then $x^k \rightarrow x$ and*

$$0 \in \tilde{\partial}_a f(x).$$

Remark 4.10. *Using Lemma 4.5, 2., we offer an interpretation of the results $\text{dist}(0, \tilde{\partial}_a f(x)) \leq \eta$ and $0 \in \tilde{\partial}_a f(x)$ from Lemma 4.7 and Corollary 4.9, respectively: Let ε_f^x be the function value approximation error at x . If $0 \in \tilde{\partial}_a f(x)$, then for any $\varepsilon' > 0$, there is a $\delta > 0$ such that for all $y \in B(x, \delta)$,*

$$\begin{aligned} 0 &\leq f(y) - f(x) + \varepsilon' \|x - y\| + \varepsilon_f^x, \\ f(x) &\leq f(y) + \varepsilon' \|x - y\| + \varepsilon_f^x. \end{aligned} \tag{42}$$

Analogously, if $\text{dist}(0, \tilde{\partial}_a f(x)) \leq \eta$, for any $\varepsilon' > 0$, there is a $\delta > 0$ such that for all $y \in B(x, \delta)$,

$$f(x) \leq f(y) + (\varepsilon' + \eta) \|x - y\| + \varepsilon_f^x. \tag{43}$$

We note that these properties also hold if ε_f^x denotes an a posteriori error, which may be available after the corresponding approximate function evaluation.

Now, we obtain convergence of the outer loop in Algorithm 1 by the following result that we prove in Appendix A.2 in the online supplement:

Theorem 4.11. *Let $\Omega := \{x \in \mathbb{R}^n : f(x) \leq f_a(x_1)\}$ be bounded and let \bar{x} be an accumulation point of the sequence x_j . Then it holds that*

$$\text{dist}(0, \partial f(\bar{x})) \leq \eta,$$

$$\text{with } \eta = \frac{\varepsilon''}{(\bar{\gamma} - \gamma)}.$$

We now sharpen the obtained results for Algorithm 1 and proceed in two steps. First, in (19), we have chosen a linear dependence of the exactness in a trial iterate and the distance to the last serious iterate: $\varepsilon_f(x^k, x) = \varepsilon'' \|x - x^k\|$. We apply instead of a fixed value ε'' a sequence $\varepsilon''_{j,k}$ which converges, passing to a subsequence, to 0 in the course of running the algorithm. We obtain convergence results that are analogous to Lemma 4.7 and Theorem 4.11. In detail, making use of $\varepsilon''_{j,k} \rightarrow 0$, we can set the constant ε'' in the optimality criterion to 0 and obtain the following results. The proofs can be found in Appendix A.3 and A.4 in the online supplement, respectively:

Corollary 4.12. *Let x be the current serious iterate and suppose the inner loop does not terminate finitely with $\tau_k \rightarrow \infty$ and $\varepsilon_f(x^k, x) = \varepsilon''_k \|x^k - x\|$ with $\varepsilon''_k \in \mathcal{O}(\tau_k^{-1}) \cup \mathcal{O}(\|x - x^k\|)$. Then it holds that*

$$0 \in \tilde{\partial}_a f(x).$$

Corollary 4.13. *Let x_1 be such that $\Omega := \{x \in \mathbb{R}^n : f(x) \leq f_a(x_1)\}$ is bounded, let \bar{x} be an accumulation point of the sequence x_j and $\varepsilon_f(x^k, x_j) = \varepsilon''_{j,k} \|x^k - x_j\|$ with $\varepsilon''_{j,k} \in \mathcal{O}(\tau_{j,k}^{-1}) \cup \mathcal{O}(\|x_j - x^k\|)$. Then it holds that*

$$0 \in \partial f(\bar{x}).$$

Second, the convergence result in Corollary 4.12 can further be sharpened by the following modification of the algorithm: When the distance between trial and serious iterate becomes small, suggesting proximity to a local optimum, re-estimate the function value at the serious iterate. In detail, during an inner loop, ε_f^x , $f_a(x)$ and $m_l(\cdot, x)$ for $l < k$ are no longer fixed, but are sequences $\varepsilon_f^{x,k}$, $f_a^k(x)$ and $m_l^k(\cdot, x)$. They are updated in the course of updating the working model in Step 18 in Algorithm 1 with

$$f_a^{k+1}(x) - \varepsilon_f^{x,k+1} \leq f(x) \leq f_a^{k+1}(x), \quad (44)$$

and $\varepsilon_f^{x,k+1} \in \mathcal{O}(\|x - x^k\|)$. In Appendix A.5 in the online supplement, we prove:

Corollary 4.14. *Let x be the current serious iterate and suppose the inner loop does not terminate finitely with $\tau_k \rightarrow \infty$ and $\varepsilon_f(x^k, x) = \varepsilon''_k \|x^k - x\|$, with $\varepsilon''_k \in \mathcal{O}(\tau_k^{-1}) \cup \mathcal{O}(\|x - x^k\|)$. Further, let $f_a^k(x)$ be a sequence that satisfies (44) with $\varepsilon_f^{x,k+1} \in \mathcal{O}(\|x - x^k\|)$. Then it holds that*

$$0 \in \partial f(x).$$

We note that making use of Corollary 4.14 would mean applying the stopping criterion $0 \in \partial f(x)$. This is contradictory to the strategy of evaluating $f_a(x)$ only with a needed, possibly increasing, exactness in the subsequent inner loop.

4.3 Stopping criteria

In general, a stopping criterion that involves exploiting the subdifferential at a given point is not applicable in practice. In this section, we list the stopping criteria that are used in our numerical tests. We motivate them by the following results.

Lemma 4.15. [Nol13] *Let x be the current serious iterate in an inner loop. If there is a trial iterate such that $x^k = x$, then $0 \in \tilde{\partial}_a f(x)$.*

Proof. Using $g_k^* = (Q(x) + \tau_k \mathcal{I})(x - x^k) \in \partial \phi_k(x^k, x)$, we infer from $x^k = x$ that $0 \in \partial \phi_k(x, x) \subseteq \partial \phi(x, x) \subseteq \tilde{\partial}_a f(x)$. \square

Lemma 4.16. *Let x be the current serious iterate in an inner loop. If $\varepsilon_f^x > 0$, then there is a $\delta > 0$ such that for any trial iterate x^k with $\varepsilon_f^k = \varepsilon'' \|x_k - x\|$:*

$$\|x^k - x\| \leq \delta \Rightarrow \tilde{\partial}_a f(x^k) \subseteq \tilde{\partial}_a f(x). \quad (45)$$

Proof. From continuity of f , we have that for any $\varepsilon_f^x > 0$ and $x \in \mathbb{R}^n$, there is a $\delta > 0$ such that for any x^k ,

$$\|x^k - x\| \leq \delta \Rightarrow \|f(x^k) - f(x)\| \leq \frac{1}{2} \varepsilon_f^x. \quad (46)$$

We choose $\delta \leq \frac{1}{2\varepsilon''} \varepsilon_f^x$, so that for any u_{x^k} with $v(x^k, u_{x^k}) \geq f(x^k) - \varepsilon'' \|x - x^k\|$, we have $v(x^k, u_{x^k}) \geq f(x) - \varepsilon_f^x$ and the claim follows from convexity of the approximate subdifferentials. \square

Both results suggest a small distance between trial and serious iterate as a motivation for stopping the algorithm. From Lemma 4.15 we have that equality of serious and trial iterate means that 0 is contained in the approximate subdifferential of the serious iterate. Moreover, from Lemma 4.16 we have that for a small enough distance between iterates, we cannot lower the subdifferential's distance to 0 by generating even closer iterates.

For practical stopping criteria, see [Nol13, Section 1.8] and [GAN13]. The proposed conditions use the distance between consecutive serious iterates or between trial and current serious iterate:

1. Let x_j be the current serious step and x_{j+1} the next accepted serious step in the inner loop of x_j . If

$$\frac{\|x_{j+1} - x_j\|}{1 + \|x_j\|} < tol, \quad (47)$$

we stop the algorithm in the outer loop and accept x_{j+1} as a solution.

2. Let x_j be the current serious step. If for three successive trial steps x^k ,

$$\frac{\|x^k - x_j\|}{1 + \|x_j\|} < tol, \quad (48)$$

then we stop the algorithm in the inner loop and accept x_j as a solution.

3. Let x_j be the current serious step and let a maximum number of inner iterations or a maximum number of outer iterations be exceeded. Then we accept x_j as a solution.

Remark 4.17. *We note that the aggregate subgradient g_k^* could also be used for a stopping criterion: In the convergence proof, we inferred the optimality criteria from convergence of the sequence of aggregated subgradients to 0, in some cases passing to a subsequence. A small $\|g_k^*\|$ thus suggests that the algorithm should be stopped.*

5 Applications

5.1 The robust gas transport problem

In this section, we concretize the adaptive bundle method for a real-world problem in order to demonstrate its applicability. The setting of nonlinear robust optimization considered so far is generic enough to cover even two-stage optimization problems with uniquely determined second-stage variables. One relevant example for this is the case of physical networks with state variables in the second stage.

Here, we consider the particular case of a robust two-stage optimization problem that arises in the context of the stationary gas transport problem. A detailed explanation of the task together with the corresponding models and solution approaches can be found in [Koc+15]. For a similar model and a decomposition method for the robust two-stage gas transport problem in the case of no active elements on cycles, see [ALS19]. We emphasize that our method is applicable to the general case of active elements on cycles and hence to a problem that has not been efficiently handled in the literature before.

Given a gas network, the task is to determine a minimum-cost control of active elements, such as compressors or control valves, so that all demands are satisfied without violating physical or technical constraints. In particular, the control of active elements is modeled as here-and-now variables at the first stage and the realization of physical states as wait-and-see variables at the second stage. Uncertainties arise in the demands and in the pressure loss coefficients. The latter is due to the fact that the pipes' frictions are uncertain and strongly influence pressure losses. The network operator therefore has to ensure feasibility of gas transport for every possible realization, meaning that a relevant robust two-stage optimization problem arises.

To keep this work self-contained, we first introduce the nominal problem, reformulate it so that it is accessible to robust optimization and derive the robust counterpart. A given gas network is modeled as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ with $|\mathcal{V}| = n$, $|\mathcal{A}| = m$ and incidence matrix $A \in \{-1, 0, 1\}^{n \times m}$. The gas flow is denoted by $q \in \mathbb{R}^m$, which is positive if flow and arc direction coincide and negative, otherwise. Pressure values at the nodes are considered by their squared values, denoted by $\pi \in \mathbb{R}^n$. The control of active elements is associated with a cost function $w(\Delta)$, which is to be minimized. Combining this objective with physical constraints and bounds leads to the following nominal optimization problem, which we will explain in more detail below:

$$\min_{\Delta, q, \pi} w(\Delta) \tag{49a}$$

$$\text{s. t. } Aq = d \tag{49b}$$

$$A^T \pi = F(q, \Delta) \tag{49c}$$

$$\pi \in [\underline{\pi}, \bar{\pi}] \tag{49d}$$

$$q \in \mathbb{R}^{|\mathcal{A}|}. \tag{49e}$$

The given demand is denoted by $d \in \mathbb{R}^n$ and is balanced: $\sum_{i=1}^n d_i = 0$. With regard to physical constraints, we first take flow conservation into account in (49b). Second, the difference between squared pressures at connected nodes depends on the type of arc: An arc $a \in \mathcal{A}$ either models a pipe if $a \in \mathcal{A}_{pi} \subseteq \mathcal{A}$, or an active element if $a \in \mathcal{A}_{ac} = \mathcal{A} \setminus \mathcal{A}_{pi}$. The pressure loss on a pipe $a = (u, v)$ depends on a pressure loss coefficient $\lambda_a > 0$. Pressure decreases in the direction of flow. Using a linear compressor model, we assign a value $\Delta_a \in \mathbb{R}$ to every active element $a \in \mathcal{A}_{ac}$. In total, denoting the pressure loss on an arc $a \in \mathcal{A}$ by $F_a(q, \Delta)$, we obtain (see e.g. [Koc+15])

$$F_a(q, \Delta) = \pi_v - \pi_u = \begin{cases} -\lambda_a q_a |q_a|, & a \in \mathcal{A}_{pi} \\ \Delta_a, & a \in \mathcal{A}_{ac}. \end{cases} \quad (50)$$

We now apply a reformulation suggested in [ALS19; Got+16]: We eliminate the constraints in (49) using the following result from [ALS19; Col+78]: Fixing the compressor variable and the pressure at an arbitrary root node, the pressure and flow values to fulfill (49b) and (49c) are unique. Further, we eliminate (49d) by penalizing violation of the pressure bounds with an exact l_1 penalty function.

We choose an arbitrary root node $r \in \mathcal{V}$ and delete the corresponding row or entry in the incidence matrix and demand vector to obtain the reduced versions \tilde{A} and \tilde{d} , respectively. Now, we divide the set of arcs \mathcal{A} into a set of basic arcs B , corresponding to a spanning tree, and a set of nonbasic arcs, $N = \mathcal{A} \setminus B$, each of them corresponding to one fundamental cycle. We can now eliminate the flow variables for basic arcs, as they are determined by the nonbasic flow variables:

$$q_B(q_N) = \tilde{A}_B^{-1}(\tilde{d} - \tilde{A}_N q_N). \quad (51)$$

For given parameters (d, λ) , the nonbasic flow $q_N(d, \lambda)$ itself can be computed by solving the cycle equation

$$\tilde{A}_N^T (\tilde{A}_B^{-1})^T F_B(q_B(q_N), \Delta) = F_N(q_N). \quad (52)$$

Here, and in the following, we assume that the root node pressure π_r is known. Given values for the active elements, we compute the corresponding state variable π as follows:

$$\pi_v(\Delta; d, \lambda) = \pi_r + \left((\tilde{A}_B^{-1})^T F_B(q_B(q_N(d, \lambda)), \Delta) \right)_v. \quad (53)$$

The pressure at a node v is thereby computed by the summed pressure loss on the path from r to v . The columns of \tilde{A}_B^{-1} describe the unique paths in the spanning tree B from the root r to the respective vertices. Finally, we can rewrite (49) with a penalty parameter $\psi \in \mathbb{R}$:

$$\min_{\Delta} w(\Delta) + \psi \sum_{v \in \mathcal{V}} \max \left\{ 0, \underline{\pi}_v - \pi_v(\Delta; d, \lambda), \pi_v(\Delta; d, \lambda) - \bar{\pi}_v \right\}. \quad (54)$$

Now we consider uncertainties in the demand d and the pressure loss coefficients λ , and protect against them in a robust sense. After a realization of the uncertain parameters, the state variables q and π are adjusted as second-stage variables so that the gas transport problem under uncertainty can be modeled as a robust two-stage optimization problem with a compact uncertainty set \mathcal{U} :

$$\min_{\Delta} \max_{(d, \lambda) \in \mathcal{U}} w(\Delta) + \psi \sum_{v \in \mathcal{V}} \max \left\{ 0, \underline{\pi}_v - \pi_v(\Delta; d, \lambda), \pi_v(\Delta; d, \lambda) - \bar{\pi}_v \right\}, \quad (55)$$

$$\mathcal{U} := \{(d, \lambda) \mid \lambda_j \in [\underline{\lambda}, \bar{\lambda}] \text{ for } j = [m], d_i \in [\underline{d}, \bar{d}] \text{ for } i \in [n], \sum_{i=1}^n d_i = 0\}. \quad (56)$$

Using the notation of Section 3, the gas transport problem is thus formulated as a minimax problem, where

$$v(\Delta, (d, \lambda)) = w(\Delta) + \psi \sum_{v \in V} \max \left\{ 0, \underline{\pi}_v - \pi_v(\Delta; d, \lambda), \pi_v(\Delta; d, \lambda) - \bar{\pi}_v \right\}, \quad (57)$$

$$f(\Delta) = \max_{(d, \lambda) \in \mathcal{U}} v(\Delta, (d, \lambda)). \quad (58)$$

Through straightforward calculation, we derive a subgradient $g_{\Delta, (d, \lambda)} \in \partial_{\Delta} v(\Delta, (d, \lambda))$. For the special case of having no compressors on cycles, the following formula thereby holds:

$$g_{\Delta, (d, \lambda)} = \nabla w(\Delta) + \psi \left(\sum_{v: \pi_v > \bar{\pi}} (\tilde{A}_{B \cap \mathcal{A}_c}^{-1})_v - \sum_{v: \pi_v < \underline{\pi}} (\tilde{A}_{B \cap \mathcal{A}_c}^{-1})_v \right). \quad (59)$$

Here we denote the v th column of (\tilde{A}_B^{-1}) , restricted to the row entries which correspond to compressor arcs, by $(\tilde{A}_{B \cap \mathcal{A}_c}^{-1})_v$.

5.1.1 Approximate solutions to the adversarial problem

The adversarial inner maximization problem in (55) is a non-convex optimization problem, which for a given Δ has to be solved with arbitrary precision, with respect to the optimal objective value $v(\Delta, (d^*, \lambda^*))$. As the non-convexity only appears in the form of one-dimensional constraints, it is efficient to apply piecewise linear relaxation to find an approximate global solution $(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)$. Piecewise linear relaxation is known to perform well in this context, see e.g. [Gei+12]. Now the question is, how to choose the error bound for the relaxation to ensure a certain precision and thus a bound on $v(\Delta, (d^*, \lambda^*)) - v(\Delta, (\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon))$. To this end, we treat the approximate solution to the adversarial problem as a black box. For a given Δ , we consider the following auxiliary optimization problem, which is a relaxation of the piecewise linear relaxation of the adversarial problem, shortening the objective to the penalty term:

$$\begin{aligned} & \max_{(d, \lambda) \in \mathcal{U}, \pi, q} \sum_{v \in V} \max \{ 0, \underline{\pi}_v - \pi_v, \pi_v - \bar{\pi}_v \} \\ & \text{s. t. } \quad Aq = d \\ & \quad (A^T \pi)_a = \Delta_a \quad \forall a \in \mathcal{A}_c \\ & \quad (A^T \pi)_a \leq -\lambda_a q_a |q_a| + \varepsilon \quad \forall a \in \mathcal{A}_p \\ & \quad (A^T \pi)_a \geq -\lambda_a q_a |q_a| - \varepsilon \quad \forall a \in \mathcal{A}_p \\ & \quad q \in \mathbb{R}^{|\mathcal{A}|} \\ & \quad \pi \in \mathbb{R}^{|V|}. \end{aligned} \quad (P^\varepsilon)$$

p^ε denotes the optimal objective value of (P^ε) and p^* denotes the optimal objective value of (P^0) , so that $v(\Delta, (d^*, \lambda^*)) = w(\Delta) + \psi p^*$. We also denote the optimal penalty term,

which is derived using piecewise linear relaxation, by \tilde{p}^ε , so that $p^* \leq \tilde{p}^\varepsilon \leq p^\varepsilon$. Now, for a given Δ we consider the following auxiliary problem, namely (P^ε) with fixed (d, λ) :

$$\begin{aligned} \max_{\pi, q} \quad & \sum_{v \in V} \max\{0, \underline{\pi}_v - \pi_v, \pi_v - \bar{\pi}_v\} \\ \text{s. t.} \quad & ((d, \lambda), \pi, q) \text{ feasible for } (P^\varepsilon). \end{aligned} \tag{P_{(d,\lambda)}^\varepsilon}$$

We denote by $p^\varepsilon(d, \lambda)$ the optimal objective value of $(P_{(d,\lambda)}^\varepsilon)$ and by $p^*(d, \lambda)$ the optimal objective value of $(P_{(d,\lambda)}^0)$, so that $v(\Delta, (d, \lambda)) = w(\Delta) + \psi p^*(d, \lambda)$. In order to bound the desired quantity $v(\Delta, (d^*, \lambda^*)) - v(\Delta, (\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon))$, we thus bound the quantity $p^* - p^*(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)$. To do this, we use the following inequalities:

$$p^*(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon) \leq p^* \leq \tilde{p}^\varepsilon \leq p^\varepsilon(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon). \tag{60}$$

We make use of the following result, omitting the proof:

Lemma 5.1. *For fixed (d, λ) and $\varepsilon > 0$, it holds that $p^\varepsilon(d, \lambda) - p^*(d, \lambda) \leq \varepsilon|V|^3$.*

With (60), we thus obtain that $p^* - p^*(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon) \leq \varepsilon|V|^3$ and it follows that

$$v(\Delta, (d^*, \lambda^*)) - v(\Delta, (\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)) \leq \psi \varepsilon |V|^3. \tag{61}$$

Using this error bound together with the subgradient formula (59), we provide an inexact function value $v(\Delta, (\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon))$ and an approximate subgradient $g_{\Delta, (\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)}$ in the sense of Section 3. However, it is not possible to derive a bound on the distance of $g_{\Delta, (\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)}$ to the exact subdifferential $\partial f(\Delta)$, as required for the ε -subdifferential (18) for the bundle method in [Nol13]. Therefore, the latter cannot be used in this context. However, for an appropriately chosen ε , we have $g_{\Delta, (\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)} \in \tilde{\partial}_a f(\Delta)$, so that the adaptive bundle method proposed here is applicable.

5.1.2 Algorithmic details

With the results from Section 5.1.1, we apply the adaptive bundle method, using piecewise linear relaxation by the delta method (see [MM57; Gei+12; ALS19]) to find an approximate solution $(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)$ to the adversarial problem with the error bound (61). Having at hand \tilde{p}^ε , an a posteriori error is provided by (see (60))

$$\varepsilon_p := \psi(\tilde{p}^\varepsilon - p^*(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)). \tag{62}$$

However, experimental evaluations of ε_p suggest that the bound (61) is usually not tight. As the computation of $(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)$ is also time-consuming for a small ε , we use the following strategy: We refine the linearization of pressure loss constraints only for relevant arcs and only in the case of a too large a posteriori error. When a function value approximation with an error bound ε_f is required, we compute $(\tilde{d}^\varepsilon, \tilde{\lambda}^\varepsilon)$ using the delta method parameter ε from the previous iteration. If the a posteriori error ε_p , violates the bound ε_f , we scale down ε for pressure loss constraints on arcs that were relevant for the a posteriori error in the following sense: We define

$$v_{max} := \operatorname{argmax}_{v \in \mathcal{V}} \|\operatorname{dist}([\underline{\pi}, \bar{\pi}], \tilde{\pi}_v^\varepsilon) - \operatorname{dist}([\underline{\pi}, \bar{\pi}], \pi_v(\Delta; d, \lambda))\|, \tag{63}$$

so that v_{max} is a node with the largest deviation of penalty caused by approximate worst case pressure values $\tilde{\pi}_v^\varepsilon$ to the penalty caused by exact state variables. As relevant arcs we then denote the arcs on the path $\mathcal{P}_{v_{max}}$ from the root node r to v_{max} and arcs on cycles $\mathcal{C}_i, i \in N$ that intersect with $\mathcal{P}_{v_{max}}$. The linearization is thus refined on the following set of arcs:

$$\mathcal{A}_{ref}^{v_{max}} := \{a \mid a \in (\mathcal{P}_{v_{max}} \cup \{\mathcal{C}_i \mid i \in N, (\mathcal{C}_i \cap \mathcal{P}_{v_{max}}) \neq \emptyset\})\} \quad (64)$$

In particular, we replace Step 6 in Algorithm 1 by the subroutine in Algorithm 2. At the beginning of an outer loop, ε^0 is thereby initialized, either with the value from the last inner loop or with an initial value if it is the first outer loop. In addition, if the proximity control parameter τ is decreased at Step 9 in Algorithm 1, we increase ε .

Algorithm 2 Subroutine: Adaptive function value approximation

- 1: $\varepsilon_f^k = \varepsilon'' \|x_j - x^k\|$, $\varepsilon_0 = \varepsilon^k \in \mathbb{R}^n$, $i = 0$.
 - 2: Compute $f_\varepsilon^0 = v(\Delta, (\tilde{d}^{\varepsilon_0}, \tilde{\lambda}^{\varepsilon_0}))$ and the a posteriori error ε_p .
 - 3: **while** $\varepsilon_p > \varepsilon_f^k$ **do**
 - 4: Identify v_{max} .
 - 5: $(\varepsilon_{i+1})_a = \begin{cases} 1/10(\varepsilon_i)_a & a \in \mathcal{A}_{ref}^{v_{max}}, \\ (\varepsilon_i)_a & \text{else.} \end{cases}$
 - 6: Compute $f_\varepsilon^{i+1} = v(\Delta, (\tilde{d}^{\varepsilon_{i+1}}, \tilde{\lambda}^{\varepsilon_{i+1}}))$ and update ε_p .
 - 7: Increase i .
 - 8: **end while**
 - 9: $f_a(x^k) = f_\varepsilon^i + \varepsilon_p$, $\varepsilon^{k+1} = \varepsilon_i$.
-

5.2 Application to Robust Electricity Networks in the Alternating Current Model

As argued earlier the adaptive bundle method is a general algorithm for nonlinear robust optimization problems that can also be applied for electricity networks. The latter are predominantly modeled using the alternating current optimal power flow (ACOPF) model. From the many different existing formulations [Bie+20] and applications, we focus on some references, namely [Car62; Jab08; KDS18]. There are only very few works on the robust ACOPF problem. Due to its intrinsic difficulty that mainly arises from its non-convexity, algorithms heuristically search for robust solutions [MR18], or work with convex relaxations from which it may be possible to generate robust points [Yan+21].

In its basic nominal version, the ACOPF problem is a nonlinear non-convex optimization problem. Given an electricity network modeled by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, the task is to minimize operating costs that are mainly caused by electricity generation. The cost function $\sum_{k \in \mathcal{V}} C_k(P_k)$ can be assumed to be linear or convex in the active power generation at the network nodes. It has to be ensured that given bounds on variable values and line flow limits are maintained. In addition, power flow equations that in the polar model are described by non-convex functions involving trigonometric terms necessarily need to be satisfied.

First, for each node, standard (linear) flow conservation equations model for both the active as well as the reactive demand that the net power generation at a node equals the flow that is transported along its incident lines. Non-convex power-flow equations model the active

power flow p_{kl} and the reactive power flow q_{kl} along a line $(k, l) \in \mathcal{A}$. These equations are trigonometric functions in the phase angle difference θ_{kl} and are bilinear in the node voltage magnitudes U_k and U_l . For details, see [Car62].

In electricity networks, a typical and relevant source of uncertainty consists in the fact that load profiles at a network node may be uncertain due to fluctuations or due to uncertainties in the load forecast. Typically, the corresponding uncertainty set \mathcal{U} is modeled as a box uncertainty that is composed as the cross product of an interval for each uncertain parameter. Robust protection against such uncertainties is a natural choice. In automatic generation control that is typically installed in practical network operation, the generators need to compensate for uncertainties in order to ensure that also under uncertainty the power is balanced throughout the network, see [MR18].

For the description of the robust counterpart, we focus on explaining the characteristics of the model only in the level of detail that is necessary here. Abstractly, the task consists in determining on the first stage, i.e., here-and-now, an optimal network state. This means that for each node $k \in \mathcal{V}$, the best possible active and reactive power generation P_k, Q_k and voltage magnitude U_k as well as for each line $(u, v) \in \mathcal{A}$ the phase angle difference $\theta_{u,v}$, the active and reactive power flow p_{uv}, q_{uv} need to be determined. The first-stage variables need to satisfy the corresponding non-convex power-flow equations for each line. They also need to take values between given upper and lower bounds. Let a first-stage decision be denoted by a vector $y \in \mathcal{Y}$ of appropriate dimension.

After the uncertainty has manifested itself, wait-and-see adjustable reactions that depend on $\omega \in \mathcal{U}$ need to ensure that the non-convex power flow equations and the corresponding bounds are satisfied. In more detail, for each $k \in \mathcal{V}$, adjustable active and reactive power generation P_k^ω, Q_k^ω and voltage magnitude U_k^ω as well as for each line $(u, v) \in \mathcal{A}$ the phase angle difference $\theta_{u,v}^\omega$, the active and reactive power flow $p_{uv}^\omega, q_{uv}^\omega$ need to be determined by automatic generation control. The latter fixes a decision rule in which the generators at the nodes generate the power that is necessary in order to react on fluctuations. The total power mismatch, i.e., the sum of all uncertainties in power generation is divided among the generators according to some participation factors. Again, adjustable variables also need to stay within their bounds and also need to satisfy the non-convex power-flow equations. For each $\omega \in \mathcal{U}$, let the adjustable second-stage decisions be denoted by a vector $y^\omega \in \mathcal{Y}(y, \omega)$ of appropriate dimension that also may depend on the first-stage decision $y \in \mathcal{Y}$.

As the resulting two-stage non-convex robust optimization problem is complex and in general intractable by available methods, we associate with it a robust model that is possibly somewhat more conservative however that is algorithmically tractable by the adaptive bundle method. Let us exemplify the reformulation of one of the robust constraints, for example for the voltage at a node, that reads as:

Determine feasible second-stage adjustable variables such that

$$U_k^\omega \leq \bar{U}_k \quad \forall \omega \in \mathcal{U},$$

where \bar{U}_k specifies the upper bound on the voltage at a node $k \in \mathcal{V}$. In order to model the robust counterpart, we rewrite the semi-infinite constraint as

$$\max_{\omega \in \mathcal{U}, y^\omega \in \mathcal{Y}(y, \omega)} U_k^\omega \leq \bar{U}_k,$$

where $\mathcal{Y}(y, \omega)$ denotes the set of second-stage adjustable variables. Maximization is not only performed with respect to $\omega \in \mathcal{U}$ but also with respect to $y^\omega \in \mathcal{Y}(y, \omega)$. The latter is necessary due to the fact that the ACOPF model usually does not guarantee the existence

of a unique solution on the second stage. Due to the fact that we ensure bounds for all adjustable solutions, the model outlined here may be more conservative than the original robust ACOPF model.

Writing similarly all robust constraints for ensuring power generation limits at all nodes and for transmission line limits at all line, we denote by the value $P(y^\omega) = \max\{0, U_k^\omega - \bar{U}_k, \dots\}$ the maximum bound violation in the second-stage.

The objective that is input for the adaptive bundle method then reads as

$$\min_{y \in \mathcal{Y}} \sum_{k \in \mathcal{V}} C_k(P_k) + \psi \max_{\omega \in \mathcal{U}, y^\omega \in \mathcal{Y}(y, \omega)} P(y^\omega)$$

with some penalty parameter ψ .

The model is in a form in which it can in principle be solved by the adapted bundle method outlined here. Although its structure is more complex than that of the robust gas operation problem from the last section, its close relation to the latter renders our bundle method a promising algorithm for its solution. Indeed, using MIP techniques, adaptive piecewise linear relaxations of the non-convex subproblems can be applied. Approximate subgradients in the sense of an outer approximation of the Clarke subdifferential can essentially be calculated by analytic formulas. As a result, this leads to an algorithm for a robust two-stage optimization problem with non-convex first as well as non-convex second stage where the second stage does not necessarily have a unique solution.

Similarly as for the gas application, piecewise linear relaxations strengthen the adversarial and together with the maximization over the adjustable variables, this may lead to a somewhat more conservative model than the original robust formulation. This seems acceptable as such highly complex problems are currently still beyond the state-of-the-art of available methods.

In order to be able to use our adaptive bundle method, we have to reformulate all first-stage constraints into the objective function. For the robust ACOPF problem, the first-stage decisions need to satisfy the non-convex power-flow equations in addition. Therefore, it would be beneficial to have at hand an enhancement of the method so that constrained problems can be dealt with without the need to relax the first-stage constraints. It would also be helpful for the quality of the solution to combine it with additional algorithmic enhancements such as spatial branching so that the non-convex problem on the first stage can be solved to global optimality. To conclude, we have shown in this section that the bundle method can in principle also be applied to the challenging robust ACOPF problem. The details of these algorithmic enhancements however go beyond the work performed here and are subject of future research.

6 Numerical Results

In this section, we present some numerical results, firstly for an analytical example and secondly for the robust gas transport problem from the preceding section. The adaptive bundle method was implemented in MATLAB and the experiments were conducted on a machine with an Intel Core i7-8550U (4 cores) and 16GB of RAM.

6.1 An analytical example

A two-dimensional test problem is taken from Zuhe et al. [ZNE90]. We slightly modified the example in order to obtain the non-convex problem in Example 6.1, which is depicted in Figure 2.

Example 6.1.

$$\min_{x \in [0,1]} \max_{u \in [0,1]} v(x, u) = u(1-u)(u-x)^4 - 0.2(x-0.5)^2.$$

For a fixed $x \in [0, 1]$, the local maxima of $v(x, u)$ in u are

$$u_1(x) := \frac{1}{12}(-\sqrt{4x^2 - 4x + 25} + 2x + 5), \quad u_2(x) := \frac{1}{12}(\sqrt{4x^2 - 4x + 25} + 2x + 5),$$

and the global maximum is attained at

$$u^*(x) = \begin{cases} u_2(x) & x \leq \frac{1}{2}, \\ u_1(x) & x > \frac{1}{2}. \end{cases}$$

The global solution to the robust optimization problem is attained at $x^* = 0.5$. To imitate inexactness in the function evaluations, we proceed as follows: For a given x with $\|v(x, u_1(x)) - v(x, u_2(x))\| \leq \varepsilon_f$, we choose $u_1(x)$ or $u_2(x)$ as an approximate optimal solution u_x to the inner maximization problem uniformly at random. In this way, we may approximate a negative exact subgradient by a positive subgradient, or vice versa. It should be noted that as soon as an approximation error is made at an iterate x , it holds that $0 \in \tilde{\partial}_a f(x)$, so that in the case of a serious iterate x , the theoretical stopping criterion in Step 3 of Algorithm 1 is fulfilled. In detail, when a serious iterate's function value is inexactly evaluated, the algorithm may be stopped with the following result: The aggregate subgradient is close to 0 and the serious iterate fulfills the optimality criterion $\text{dist}(0, \tilde{\partial}_a f(x)) \leq \frac{\varepsilon''}{(\bar{\gamma} - \gamma)}$ from Lemma 4.7 and Corollary 4.9.

Numerical results for different values of the parameter ε'' are summarized in Table 1, where each column with $\varepsilon'' > 0$ represents an averaged result over 50 runs. From the table, we see that the parameter value $\varepsilon'' = 1$ is not sufficiently small to yield results of the same quality as with $\varepsilon'' = 0.01$ or $\varepsilon'' = 0.1$. Figures 2 and 3 depict an example for the case $\varepsilon'' = 1$, in which the 6th serious iterate is inexactly evaluated and the optimum of $x^* = 0.5$ is not achieved by the algorithm. As at the 6th serious iterate x , a negative subgradient is approximated by a positive one, we have $0 \in \tilde{\partial}_a f(x)$. Even though the theoretical stopping criterion applies, the practical does not. Due to the positive subgradient, the algorithm however makes no progress in the sense of approaching $x = 0.5$ and after a few more inner loops, a practical stopping criterion applies.

We want to improve the results for the case of $\varepsilon'' = 1$ without globally setting ε'' to a smaller value. Therefore, we apply a strategy like that used in Corollary 4.14 and re-estimate $f_a(x)$ for a serious iterate x if $\varepsilon_f^{x,k} \geq 10^{-1} \|x - x^k\|$. An example for this case is depicted in Figure 4 and 5. Such a re-estimation also includes updating the corresponding subgradient and cutting planes. We hence do not permanently accept an inexact subgradient. This hinders the algorithm from becoming stuck at a point with an error in the subgradient approximation and thus 0 in the approximate subdifferential. The results show that this strategy indeed yields to a significant improvement in the objective and solution value, so that the optimum is reached when only inexact function values and subgradient evaluations are available.

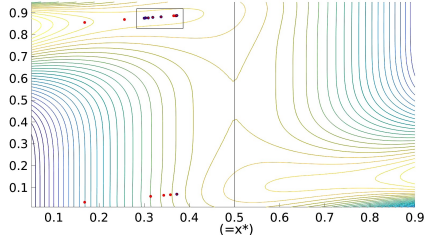


Figure 2: $v(x, u)$ from Example 6.1 with iterates from a run with $\varepsilon'' = 1$. Due to inexactness "wrong" approximate subgradients with negative slope enter the bundle for all iterates shown in the bottom left. In contrast, for exact evaluations a "correct" subgradient with positive slope is added, see iterates at top left. See Figure 3 for a zoom on the box.

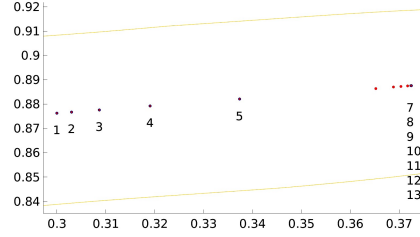


Figure 3: Zoom to the region of the subsequence of serious iterates with exact solution to the inner maximization problem. The numbers indicate the outer loop counter j .

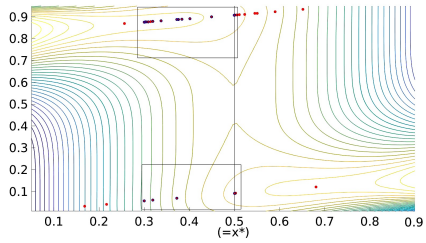


Figure 4: $v(x, u)$ from Example 6.1 with iterates from a run with re-estimation of $f_a(x)$. The optimal solution at $x^* = 0.5$ is attained. See Figure 3 for a zoom on the boxes.

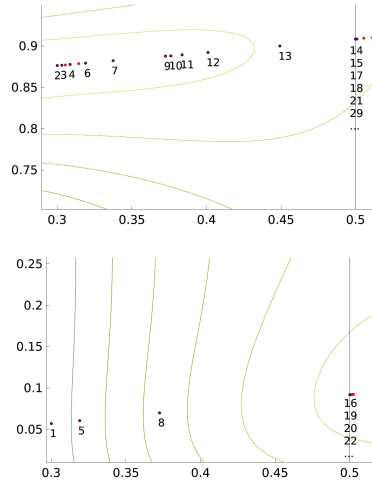


Figure 5: Zoom to the regions of the subsequences of serious iterates. The numbers indicate the outer loop counter j .

	$\varepsilon'' = 0$	$\varepsilon'' = 0.01$	$\varepsilon'' = 0.1$	$\varepsilon'' = 1$	$\varepsilon'' = 1$, re-estimate $f_a(x)$
# serious iterates	37	18	29	19	48
# trial iterates	51	34	53	33	99
objective value	0.0023	0.0032	0.0030	0.0038	0.0023
solution	0.50	0.40	0.42	0.33	0.50
$\ g^*\ $	4.67e-17	1.13e-11	1.15e-12	2.53e-13	3.35e-13
# downshifts	2	8	17	6	29
running time (sec)	3.0	5.3	12.0	5.4	28.9

Table 1: Averaged results for Example 6.1 with $tol = 10^{-9}$, $\gamma = 0.1$, $\tilde{\gamma} = 0.9$. The geometric mean is used for the row of $\|g^*\|$. The last column shows results for the strategy of re-estimating the serious iterates’ function values (see Corollary 4.14).

6.2 The robust gas transport problem

For computational experiments with the robust gas transport problem, the relaxed adversarial maximization problem was solved using an implementation of the delta method in Python with Gurobi 8.1.

As test cases, we used slightly modified instances from the library of realistic gas network instances [Pfe+15], and as uncertainty sets for nominal demand and pressure loss coefficients d and λ , we used $[0.95 \cdot d, 1.05 \cdot d]$ and $[\lambda, 1.1 \cdot \lambda]$, respectively. Further, in order to fulfill bounds on Δ , we included a penalty term in the cost function $w(\Delta)$. The numerical results for the instances GasLib134 and GasLib40, modelling gas networks with 40 and 134 nodes, respectively, are summarized in Table 2 and 3, where we compare the results for different values of the algorithmic exactness parameter ε'' . When we refer to ”number of refinements”, we mean the number of re-evaluations in the sense of Algorithm 2. It should be noted from the results that the running time is determined to a large extent by the effort of approximating the adversarial problem up to the desired error: For different values of ε'' , the numbers of trial iterates and thus function evaluations are in the same order of magnitude, whereas the running time increases as the algorithm demands a smaller error in the approximation of the adversarial problem.

The results can be compared to numerical results in [ALS19]. There, the GasLib instances are also used, though with a slightly different model. We compared our results for the case of GasLib-134, which models the Greek gas network. According to a test run with the instance used in [ALS19], we attain the same solution, which makes use of one of the two active elements. The running time and iteration number of this test run were in the same order of magnitude as for our experiments that are depicted in Table 2. For our numerical experiments, we have chosen a nominal demand scenario for which both of the two active elements are used in the solution. Table 2 shows that for all tested values of ε'' , no downshifts or refinements are necessary and the algorithm requires only a few seconds. This effect can be explained by the network’s tree-structure, which makes the adversarial problem much easier to solve, leading to small a posteriori errors.

We note that the approach in [ALS19] is explicitly adapted to the structure of the robust gas transport problem and moreover cannot be applied to instances with compressors on cycles. In the instance GasLib40, there is one compressor on a cycle, so that we cannot directly compare our results. The results of our numerical experiments are depicted in Table 3. We reach for the case of $\varepsilon'' = 10$ only quite a large value for $\|g^*\|$, the aggregate

	$\varepsilon'' = 0.1$	$\varepsilon'' = 1$	$\varepsilon'' = 10$
# serious iterates	11	11	11
# trial iterates	13	13	13
objective value	343.80	343.80	343.80
$\ g^*\ $	1.30e-09	1.30e-09	1.30e-09
# refinements	0	0	0
# downshifts	0	0	0
running time (sec)	18.7	20	20.8

Table 2: Results for the instance GasLib134 (2 active elements).

	$\varepsilon'' = 0.1$	$\varepsilon'' = 1$	$\varepsilon'' = 10$	$\varepsilon'' = \min(10, 10 \cdot \ x^k - x_j\)$
# serious iterates	19	22	22	25
# trial iterates	32	37	59	50
objective value	2211.98	2211.97	2216.46	2212.14
$\ g^*\ $	1.53e-10	1.42e-09	1.00e+00	1.22e-05
# refinements	47	48	42	49
# downshifts	0	5	31	17
running time (sec)	94.1	60.4	58.5	63.0

Table 3: Results for the instance GasLib40 (6 active elements).

subgradient’s norm. To improve this result without having to solve the adversarial problem up to a lower error bound at every single function evaluation in the course of the algorithm, we apply Corollaries 4.12 and 4.13: We ensure $\varepsilon_k'' \in \mathcal{O}(\|x - x^k\|)$ by the choice of $\varepsilon'' = \min(10, 10 \cdot \|x^k - x_j\|)$, so that the enhanced convergence results of Corollaries 4.12 and 4.13 apply. Indeed, from the results in Table 3 we recognize an enhancement with respect to $\|g^*\|$ and the objective value. However, this effect is accompanied by an increase in the number of iterations and refinements, and thus function evaluations. The strategy of an adaptive ε'' thus is useful if evaluations of low precision are cheap and evaluations of high precision expensive but still tractable.

In summary, we demonstrated that our adaptive bundle method is an efficient method that is able to solve realistic instance sizes within short time, such as the robust gas transport problem for Greece. Despite the fact that this problem contains non-convexities, we demonstrated its tractability via adaptive solution of the mixed-integer linear relaxations using state-of-the-art MIP solvers. The results show that our adaptive bundle method with the appropriate notion of inexactness is able to make use of these practically efficient MIP technologies for the solution of nonlinear robust optimization problems.

7 Conclusion

We have presented an adaptive bundle method for non-convex nonlinear robust optimization problems with inexact solutions to the adversarial problem. In addition to providing

a convergence proof, we have further shown its applicability through the example of gas transport under uncertainties. In this context, a robust two-stage optimization problem with unique lower level arises from uncertainties and the need for a robust treatment of these in a real-world application.

An open question in this context is whether some of the techniques presented here could be used for the case of discrete active elements and thus global optimization. Since our framework is appropriate in the case of nonlinear robust optimization, it can also be used for further applications such as e.g. optimal water or electricity networks. Both of these applications are again network applications for which a piecewise linearization method may be the preferred choice to adaptively solve the lower level problem to global optimality. We have outlined the approach for the optimal power flow problem in the alternating current model in the application section of this article. Having this said, we would like to stress again that the bundle framework we presented, does not impose any restrictions on the choice of the solution method for the lower level problem. To give an example, the authors are currently working on a problem in the design of nanoparticles, namely robust particle synthesis. There, the lower level-problem does not exhibit a graph-based structure and thus an optimization method based on Lipschitz constants is preferred. As, however, the main purpose of this article was to introduce the bundle framework for the solution of general nonlinear robust optimization problems, we decided to report about these applications in the future rather than discussing them in full detail here.

Acknowledgments

The authors thank the DFG for their support within Project B06 in CRC TRR 154. Furthermore, we thank Denis Aßmann and Raphael Köglmaier for providing parts of the implementation.

A Proofs of convergence results from Section 4.2

In the following proofs, we employ techniques analogous to those employed by [Nol13, Lemma 3], transferred to our setting. It is noted however that first, the definitions of e.g. f_a and ϕ are not equivalent and second, we require a different argument to derive (74) due to our different notion of inexactness.

A.1 Proof of Lemma 4.7

Proof. i) We have, see (27), $g_k^* = (Q + \tau_k \mathcal{I})(x - x^k) \in \partial\phi_k(x^k, x)$, so that with $\phi_k(x, x) = f_a(x)$, with $m_0(x^k, x) \leq \phi_k(x^k, x)$ and the definition of the exactness plane $m_0(x^k, x) = f_a(x) + g_x^T(x^k - x)$ where $g_x \in \tilde{\partial}_a f(x)$,

$$g_k^{*T}(x - x^k) \leq \phi_k(x, x) - \phi_k(x^k, x) \leq f_a(x) - m_0(x^k, x) = g_x^T(x - x^k). \quad (65)$$

Using the definition of g_k^* and (65), we thus have

$$g_k^{*T}(x - x^k) = (x - x^k)^T(Q + \tau_k \mathcal{I})(x - x^k) \leq \|g_x\| \|x - x^k\|. \quad (66)$$

The matrix $(Q + \tau \mathcal{I})$ is positive definite, so that, with $\tau_k \rightarrow \infty$, we can infer $\|x - x^k\| \rightarrow 0$ and hence $x^k \rightarrow x$.

ii) We obtain $\phi_k(x^k, x) \rightarrow f_a(x)$ as follows: Using (65), there is a lower bound on $\phi_k(x, x) - \phi_k(x^k, x)$, which tends to 0, using (66) for $k \rightarrow \infty$. As $m_0(x^k, x) \rightarrow f_a(x)$ for $k \rightarrow \infty$, (65) also provides an upper bound on $\phi_k(x, x) - \phi_k(x^k, x)$, which tends to 0. With $\phi_k(x, x) = f_a(x)$, the claim follows.

iii) Let $\eta_k := \text{dist}(g_k^*, \partial\phi(x, x))$. Using $\phi(\cdot, x) \geq \phi_k(\cdot, x)$ from Lemma 4.3, 3., together with $g_k^* \in \partial\phi_k(x^k, x)$ with the corresponding subgradient inequality, it follows for all $y \in \mathbb{R}^n$ that:

$$\phi(y, x) \geq \phi_k(y, x) \geq \phi_k(x^k, x) + g_k^{*T}(y - x^k). \quad (67)$$

From part i), we have $x^k \rightarrow x$ and boundedness of g_k^* . Hence, passing to a subsequence ensures $g_k^* \rightarrow g^*$ for some g^* . Together with $\phi_k(x^k, x) \rightarrow f_a(x) = \phi(x, x)$ from ii), we obtain

$$\phi(y, x) \geq \phi(x, x) + g^{*T}(y - x),$$

so that $g^* \in \partial\phi(x, x)$. Thus, $\eta_k \leq \|g_k^* - g^*\| \rightarrow 0$.

To investigate the asymptotic behavior of ρ_k and $\tilde{\rho}_k$, we derive by using parts i)-iii) the estimates (72) in part iv) and (75) in part v). These estimates bound the terms $f_a(x) - \Phi_k(x^k, x)$ and $f_a(x^k) - M_k(x^k, x)$, respectively.

iv) Subtracting $\frac{1}{2}(x - x^k)^T Q(x - x^k)$ from the first inequality in (65), using $\phi_k(x, x) = f_a(x)$ and again the definition of g_k^* , we obtain

$$\frac{1}{2}(x - x^k)^T Q(x - x^k) + \tau_k \|x - x^k\|^2 \leq f_a(x) - \Phi_k(x^k, x). \quad (68)$$

Next, we can bound the left-hand side for a sufficiently large k and an arbitrarily fixed $0 < \zeta < 1$, using $\tau_k \rightarrow \infty$ as follows:

$$\frac{1}{2}(x - x^k)^T Q(x - x^k) + \tau_k \|x - x^k\|^2 \geq (1 - \zeta)\tau_k \|x - x^k\|^2. \quad (69)$$

Moreover, for sufficiently large k , we have from the asymptotic behavior of g_k^* , that

$$\|g_k^*\| \leq (1 + \zeta)\tau_k \|x - x^k\|. \quad (70)$$

Substituting (69) for the left-hand side of (68) and replacing τ_k using (70), for sufficiently large k we obtain

$$f_a(x) - \Phi_k(x^k, x) \geq \frac{1 - \zeta}{1 + \zeta} \|g_k^*\| \|x - x^k\|. \quad (71)$$

We define $\eta := \text{dist}(0, \partial\phi(x, x))$. From the definition of η and η_k , we obtain for $k \rightarrow \infty$ that $\|g_k^*\| \geq \eta - \eta_k > (1 - \zeta)\eta$. Putting this estimation into (71), we infer, for k large enough,

$$f_a(x) - \Phi_k(x^k, x) \geq \frac{(1 - \zeta)^2}{1 + \zeta} \eta \|x - x^k\|, \quad (72)$$

where ζ is chosen as before.

v) As defined in (21), we have $m_k(\cdot, x) = t_k(\cdot) - s_k$, with $s_k := [t_k(x) - f_a(x)]_+ + c\|x^k - x\|^2$. We first consider the case $t_k(x) > f_a(x)$. Then:

$$m_k(\cdot, x) = f_a(x) + g_k^T(\cdot - x) - c\|x - x^k\|^2. \quad (73)$$

Now, choose an arbitrary $\varepsilon' > 0$. Using Lemma 4.4, there is a $\delta > 0$ such that for every fixed $u \in \mathcal{U}$, $v(\cdot, u)$ is ε' -convex on $B(x, \delta)$ for the chosen ε' . From $x^k \rightarrow x$, we have for

sufficiently large k that $x^k \in B(x, \delta)$, so that Lemma 4.5, 3., can be applied. Using (73) we obtain:

$$\begin{aligned} f_a(x^k) - m_k(x^k, x) &= f_a(x^k) - (f_a(x) + g_k^T(x^k - x)) + c\|x^k - x\|^2 \\ &\leq (\varepsilon' + \varepsilon'')\|x^k - x\| + c\|x^k - x\|^2. \end{aligned} \quad (74)$$

It follows by subtracting $\frac{1}{2}(x - x^k)^T Q(x - x^k)$ from both sides and using $c\|x - x^k\|^2 - \frac{1}{2}(x - x^k)^T Q(x - x^k) \rightarrow 0$, that

$$f_a(x^k) - M_k(x^k, x) \leq (1 + \zeta)(\varepsilon' + \varepsilon'')\|x^k - x\| \quad (75)$$

for k large enough, again choosing ζ as before. For the case $t_k(x) \leq f_a(x)$, the estimate (75) also holds, since in this case $f_a(x^k) - m_k(x^k, x) = c\|x - x^k\|^2$. For k large enough, we then have $c\|x - x^k\|^2 \leq (\varepsilon' + \varepsilon'')\|x - x^k\|$ and $-\frac{1}{2}(x - x^k)^T Q(x - x^k) \leq \zeta\|x - x^k\|$, so that the claim follows.

vi) Now we apply the obtained results to the formulas for ρ_k and $\tilde{\rho}_k$: Rewriting $\tilde{\rho}_k$ as

$$\tilde{\rho}_k = \rho_k + \frac{f_a(x^k) - M_k(x^k, x)}{f_a(x) - \Phi_k(x^k, x)}, \quad (76)$$

and using (72) and (75), we obtain for sufficiently large k

$$\tilde{\rho}_k \leq \rho_k + \frac{(1 + \zeta)^2(\varepsilon' + \varepsilon'')\|x - x^k\|}{(1 - \zeta)^2\eta\|x - x^k\|} = \rho_k + \frac{(1 + \zeta)^2(\varepsilon' + \varepsilon'')}{(1 - \zeta)^2\eta}. \quad (77)$$

From $\rho_k < \gamma$ for all k and $\tilde{\rho}_k \geq \tilde{\gamma}$ for infinitely many k and since ζ was chosen arbitrarily, we infer $\eta < \frac{(\varepsilon' + \varepsilon'')}{(\tilde{\gamma} - \gamma)}$. As $\varepsilon' > 0$ was chosen arbitrarily, it follows that

$$\eta \leq \frac{\varepsilon''}{(\tilde{\gamma} - \gamma)}. \quad (78)$$

Finally, with $\partial\phi(x, x) \subseteq \tilde{\partial}_a f(x)$ from Lemma 4.3, 2., the claim follows. \square

A.2 Proof of Theorem 4.11

Proof. We must only consider the case where the outer loop does not terminate finitely. By k_j , we denote the inner loop counter of an outer iteration j at which a trial step is accepted as a new serious step, s.t. $x^{k_j, j} = x_{j+1}$. From the acceptance test, we thus have

$$f_a(x_j) - f_a(x_{j+1}) \geq \gamma (f_a(x_j) - \Phi_{k_j}(x_{j+1}, x_j)). \quad (79)$$

Moreover, with $g_{k_j}^* = (Q_j + \tau_{k_j}\mathcal{I})(x_j - x_{j+1}) \in \partial\phi_{k_j}(x_{j+1}, x_j)$, the following subgradient inequality, applied at x_j , holds:

$$(x_j - x_{j+1})^T (Q_j + \tau_{k_j}\mathcal{I})(x_j - x_{j+1}) \leq \phi_{k_j}(x_j, x_j) - \phi_{k_j}(x_{j+1}, x_j). \quad (80)$$

Using $\phi_{k_j}(x_j, x_j) = f_a(x_j)$ and (79), we obtain

$$\frac{1}{2}\|x_{j+1} - x_j\|_{Q_j + \tau_{k_j}}^2 \leq f_a(x_j) - \Phi_{k_j}(x_{j+1}, x_j) \leq \gamma^{-1}(f_a(x_j) - f_a(x_{j+1})). \quad (81)$$

By construction, the sequence of $f_a(x_j)$ is decreasing and due to boundedness of Ω also bounded from below, so that the right-hand side converges to 0. We thus obtain

$$\|x_{j+1} - x_j\|_{Q_j + \tau_{k_j}}^2 \xrightarrow{j \rightarrow \infty} 0, \quad (82)$$

so that there is a subsequence of serious steps x_j and a limit point \bar{x} with $x_j \rightarrow \bar{x}$ for $j \in J$. Throughout the algorithm, the accumulation point \bar{x} does not appear as a trial step, so that $f_a(\bar{x})$ is never evaluated. Therefore, we redefine the value $f_a(\bar{x}) := \lim_{j \rightarrow \infty} f_a(x_j)$. It is well-defined as the sequence is decreasing and bounded from below. From $f(x_j) \leq f_a(x_j) \leq f(x_j) + \varepsilon'' \|x_j - x_{j-1}\|$ together with (82), we have $f_a(\bar{x}) = f(\bar{x})$.

We now consider the sequence

$$g_{k_j}^* = (Q_j + \tau_{k_j} \mathcal{I})(x_j - x_{j+1}) \quad (83)$$

and distinguish two cases. First, we consider the case where a subsequence of $g_{k_j}^*$ converges to 0 and second, we assume that no such subsequence exists:

i) Assume first that there is an infinite subsequence J' of J such that $g_{k_j}^*$ converges to 0 for $j \in J'$. To prove $0 \in \phi(\bar{x}, \bar{x})$, we now show that the limit of $g_{k_j}^*$, which equals 0, fulfills the subgradient inequality of ϕ at \bar{x} .

First, we note that boundedness of Ω implies boundedness of the sequence of serious iterates x_j . From Corollary 3.1, it follows that the sequence of corresponding approximate subgradients $g_{x_j} \in \tilde{\partial}_a f(x_j)$ is bounded. With (66), applied at $g_{k_j}^*$, boundedness of the trial steps x^k in all the inner loops also follows. Hence, in the definition of ϕ in (32), M can be chosen such that $B(0, M)$ contains the trial iterates from all inner loops.

Now, we apply the subgradient inequality of $g_{k_j}^* \in \partial \phi_{k_j}(x_{j+1}, x_j)$ to an arbitrarily chosen vector h and use estimate (79) from the acceptance test:

$$\begin{aligned} g_{k_j}^{*T} h &\leq \phi_{k_j}(x_{j+1} + h, x_j) - \phi_{k_j}(x_{j+1}, x_j) \\ &\leq \phi(x_{j+1} + h, x_j) - f_a(x_j) + f_a(x_j) - \phi_{k_j}(x_{j+1}, x_j) \\ &= \phi(x_{j+1} + h, x_j) - f_a(x_j) + f_a(x_j) - \Phi_{k_j}(x_{j+1}, x_j) + \frac{1}{2} \|x_j - x_{j+1}\|_{Q_j}^2 \\ &\leq \phi(x_{j+1} + h, x_j) - f_a(x_j) + \gamma^{-1} (f_a(x_j) - f_a(x_{j+1})) + \frac{1}{2} \|x_j - x_{j+1}\|_{Q_j + \tau_{k_j}}^2. \end{aligned} \quad (84)$$

Choosing an arbitrary h' and using (84) with $h = x_j - x_{j+1} + h'$ and $g_{k_j}^{*T}(x_j - x_{j+1}) = \|x_j - x_{j+1}\|_{Q_j + \tau_{k_j}}^2$, we see:

$$\frac{1}{2} \|x_j - x_{j+1}\|_{Q_j + \tau_{k_j}}^2 + g_{k_j}^{*T} h' \leq \phi(x_j + h', x_j) - f_a(x_j) + \gamma^{-1} (f_a(x_j) - f_a(x_{j+1})). \quad (85)$$

With $j \rightarrow \infty$, we obtain $0 \leq \phi(\bar{x} + h', \bar{x}) - \phi(\bar{x}, \bar{x})$ and hence $0 \in \partial \phi(\bar{x}, \bar{x})$. By Lemma 4.3, 2., it follows that $0 \in \tilde{\partial}_a f(\bar{x})$ and thus, with $f_a(\bar{x}) = f(\bar{x})$, we obtain $0 \in \partial f(\bar{x})$.

ii) Second, assume that for every $j \in J$, $\|g_{k_j}^*\| = \|(Q_j + \tau_{k_j} \mathcal{I})(x_j - x_{j+1})\| \geq \mu > 0$ for a $\mu > 0$. For this case, we proceed in a manner similar to the proof of Lemma 4.7. There, we investigated a sequence of trial iterates which do not pass the acceptance test and also lead to an increase of the proximity parameter. For the limit of this sequence, we then obtained a bound on the distance of the approximate subdifferential to 0.

Now, we derive an analogous result for the accumulation point \bar{x} . To do this, we pass to a subsequence of trial iterates that do not pass the acceptance test, lead to an increase of the

proximity parameter and converge to \bar{x} . We start by identifying such a subsequence and then proceed as in the proof of Lemma 4.7: We show that an accumulation point of the sequence of corresponding aggregate subgradients lies in $\partial\phi(\bar{x}, \bar{x})$ and bound its distance to 0. We then obtain a bound on the distance of the approximate subdifferential at \bar{x} to 0.

As $\|x_j - x_{j+1}\| \rightarrow 0$, $g_{k_j}^* \not\rightarrow 0$ and the sequence of Q_j is bounded, the sequence τ_{k_j} , $j \in J$ is unbounded and $\tau_{k_j} \rightarrow \infty$. Let $J^+ := \{j \in J : \tau_{k_j} > \tau_j^\#\}$ be the set of indices of inner loops in which the proximity parameter is increased. From boundedness of the sequence $\tau_j^\#$, as ensured in Step 12 in Algorithm 1, it follows that for every $j \in J$ there is a $j' \in J^+$ with $j \leq j'$. Hence, there is an infinite subsequence $j \in J^+ \subseteq J$ with $\tau_{k_j} \rightarrow \infty$, which we now consider. Let $k_j - v_j$ be the index of the last inner step of the j th inner loop, at which the parameter τ is increased: $\tau_{k_j} = 2\tau_{k_j - v_j}$. From the acceptance test and the update rule for the proximity control parameter in the algorithm, we have $\rho_{k_j - v_j} < \gamma$ and $\tilde{\rho}_{k_j - v_j} \geq \tilde{\gamma}$, respectively.

We consider the sequence $g_{k_j - v_j}^* = (Q_j + \frac{1}{2}\tau_{k_j}\mathcal{I})(x_j - x^{k_j - v_j}) \in \partial\phi_{k_j - v_j}(x^{k_j - v_j}, x_j)$. Analogously to (65), we obtain:

$$(x_j - x^{k_j - v_j})^T (Q_j + \frac{1}{2}\tau_{k_j}\mathcal{I})(x_j - x^{k_j - v_j}) \leq g_{x_j}^T (x_j - x^{k_j - v_j}), \quad (86)$$

where $g_{x_j} \in \tilde{\partial}_a f(x_j)$. Boundedness of Ω and thus x_j together with Corollary 3.1 ensures boundedness of $\tilde{\partial}_a f(x_j)$. Hence, with (86), boundedness of $g_{k_j - v_j}^*$ follows. Using $\tau_{k_j - v_j} \rightarrow \infty$, we thus derive from the definition of $g_{k_j - v_j}^*$ that $\|x_j - x^{k_j - v_j}\| \rightarrow 0$, so that $x^{k_j - v_j} \rightarrow \bar{x}$. As the sequence $g_{k_j - v_j}^*$ is bounded, there is a subsequence J' of J such that $g_{k_j - v_j}^* \rightarrow \tilde{g}$ for some \tilde{g} . Now we apply the subgradient inequality of $g_{k_j - v_j}^* = (Q_j + \frac{1}{2}\tau_{k_j}\mathcal{I})(x_j - x^{k_j - v_j}) \in \phi_{k_j - v_j}(x^{k_j - v_j}, x_j)$ to an arbitrary h and use $\tilde{\rho}_{k_j - v_j} \geq \tilde{\gamma}$ together with $\Phi_{k_j - v_j}(x^{k_j - v_j}, x_j) = M_{k_j - v_j}(x^{k_j - v_j}, x_j)$:

$$\begin{aligned} g_{k_j - v_j}^{*T} h &\leq \phi_{k_j - v_j}(x^{k_j - v_j} + h, x_j) - \phi_{k_j - v_j}(x^{k_j - v_j}, x_j) \\ &\leq \phi(x^{k_j - v_j} + h, x_j) - f_a(x_j) + f_a(x_j) - \phi_{k_j - v_j}(x^{k_j - v_j}, x_j) \\ &= \phi(x^{k_j - v_j} + h, x_j) - f_a(x_j) + f_a(x_j) - \Phi_{k_j - v_j}(x^{k_j - v_j}, x_j) \\ &\quad + \frac{1}{2}\|x^{k_j - v_j} - x_j\|_{Q_j}^2 \\ &\leq \phi(x^{k_j - v_j} + h, x_j) - f_a(x_j) + \tilde{\gamma}^{-1}(f(x_j) - M_{k_j - v_j}(x^{k_j - v_j}, x_j)) \\ &\quad + \frac{1}{2}\|x^{k_j - v_j} - x_j\|_{Q_j}^2. \end{aligned} \quad (87)$$

To obtain a subgradient inequality implying for the limit with respect to $j \rightarrow \infty$ that $\tilde{g} \in \partial\phi(\bar{x}, \bar{x})$, we show that the last two summands tend to 0. For the last summand, this follows from $\|x_j - x^{k_j - v_j}\| \rightarrow 0$ and boundedness of Q_j .

To prove $\tilde{\gamma}^{-1}(f(x_j) - M_{k_j - v_j}(x^{k_j - v_j}, x_j)) \rightarrow 0$, we only need to show $f_a(x_j) - m_{k_j - v_j}(x^{k_j - v_j}, x_j) \rightarrow 0$ and proceed as follows: For every $j \in J'$, we infer from $f(x_j) - f_a(x_j) \leq 0$ together with $f(x^{k_j - v_j}) - f_a(x^{k_j - v_j}) \geq -\varepsilon''\|x_j - x^{k_j - v_j}\|$ that

$$f(x_j) - f_a(x_j) \leq f(x^{k_j - v_j}) - f_a(x^{k_j - v_j}) + \varepsilon''\|x_j - x^{k_j - v_j}\| \quad (88)$$

so that

$$f(x_j) - f(x^{k_j - v_j}) \leq f_a(x_j) - f_a(x^{k_j - v_j}) + \varepsilon''\|x_j - x^{k_j - v_j}\|. \quad (89)$$

Passing to the limit gives

$$\limsup_{j \in J'} f_a(x^{k_j - v_j}) \leq \lim_{j \rightarrow \infty} f_a(x_j) = f_a(\bar{x}), \quad (90)$$

and with $m_{k_j - v_j}(x^{k_j - v_j}, x_j) \leq f_a(x^{k_j - v_j})$, it follows that

$$\limsup m_{k_j - v_j}(x^{k_j - v_j}, x_j) \leq f_a(\bar{x}). \quad (91)$$

To show that $f_a(x_j) - m_{k_j - v_j}(x^{k_j - v_j}, x_j) \rightarrow 0$, it hence suffices to prove that the estimate $\liminf m_{k_j - v_j}(x^{k_j - v_j}, x_j) \geq f_a(\bar{x})$ holds. We thus need to show that the downshift $s_{k_j - v_j}$ at $x^{k_j - v_j}$ with respect to x_j and $f_a(x_j)$ tends to 0. This follows from

$$\lim_{j \rightarrow \infty} [t_{k_j - v_j}(x_j) - f_a(x_j)]_+ = \lim_{j \rightarrow \infty} [f_a(x^{k_j - v_j}) - f_a(x_j) + g_{k_j - v_j}^T(x_j - x^{k_j - v_j})]_+ = 0, \quad (92)$$

using (90), boundedness of $\tilde{\partial}_a f(x^{k_j - v_j})$ and $x^{k_j - v_j} \rightarrow x_j$. Passing in (87) to the limit, we thus infer

$$g_{k_j - v_j}^*{}^T h \leq \phi(\bar{x} + h, \bar{x}) - \phi(\bar{x}, \bar{x}). \quad (93)$$

As h was chosen arbitrarily, we obtain $\tilde{g} \in \partial\phi(\bar{x}, \bar{x})$.

In a manner similar to the proof of Lemma 4.7, parts iv) and v) in 1.), we now investigate the asymptotic behavior of $\rho_{k_j - v_j}$ and $\tilde{\rho}_{k_j - v_j}$. To this end, we derive the estimates (94) and (96) for the terms $f_a(x_j) - \Phi_{k_j - v_j}(x^{k_j - v_j}, x_j)$ and $f_a(x^{k_j - v_j}) - M_{k_j - v_j}(x^{k_j - v_j}, x_j)$, respectively.

Let $\eta := \text{dist}(0, \partial\phi(\bar{x}, \bar{x}))$ so that $\|\tilde{g}\| \geq \eta$. The following result is derived analogously to part iv) in the proof of Lemma 4.7: Using $g_{k_j - v_j}^* \in \partial\phi_{k_j - v_j}(x^{k_j - v_j}, x_j)$ and $\tau_{k_j} \rightarrow \infty$, together with $\|g_{k_j - v_j}^*\| \geq (1 - \zeta)\eta$ for j large enough, it holds that

$$f_a(x_j) - \Phi_{k_j - v_j}(x^{k_j - v_j}, x_j) \geq \frac{1 - \zeta}{1 + \zeta} \|\tilde{g}\| \|x_j - x^{k_j - v_j}\| \geq \frac{(1 - \zeta)^2}{1 + \zeta} \eta \|x_j - x^{k_j - v_j}\|. \quad (94)$$

We first assume $[t_{k_j - v_j}(x_j) - f_a(x_j)]_+ > 0$. Now, choose an arbitrary $\varepsilon' > 0$. From Lemma 4.4, we know that there is a $\delta > 0$ such that for every fixed $u \in \mathcal{U}$, $v(\cdot, u)$ is ε' -convex on $B(\bar{x}, \delta)$ for the chosen ε' . From $x^{k_j - v_j} \rightarrow \bar{x}$ and $x_j \rightarrow \bar{x}$, we obtain for sufficiently large j that $x^{k_j - v_j}, x_j \in B(\bar{x}, \delta)$. Thus, Lemma 4.5, 3., can be applied and for sufficiently large j it follows that

$$\begin{aligned} f_a(x^{k_j - v_j}) - m_{k_j - v_j}(x^{k_j - v_j}, x_j) &= f_a(x^{k_j - v_j}) - f_a(x_j) - g_{k_j - v_j}^T(x^{k_j - v_j} - x_j) \\ &\quad + c\|x^{k_j - v_j} - x_j\|^2 \\ &\leq (\varepsilon'' + \varepsilon')\|x^{k_j - v_j} - x_j\| + c\|x^{k_j - v_j} - x_j\|^2. \end{aligned} \quad (95)$$

The estimate also holds for the case $[t_{k_j - v_j}(x_j) - f_a(x_j)]_+ = 0$. Analogously to part v) in the proof of Lemma 4.7, we obtain

$$f_a(x^{k_j - v_j}) - M_{k_j - v_j}(x^{k_j - v_j}, x_j) \leq (1 + \zeta)(\varepsilon' + \varepsilon'')\|x^{k_j - v_j} - x_j\|. \quad (96)$$

Using the estimates (96) and (94), we infer

$$\begin{aligned} \tilde{\rho}_{k_j - v_j} &= \rho_{k_j - v_j} + \frac{f_a(x^{k_j - v_j}) - M_{k_j - v_j}(x^{k_j - v_j}, x)}{f_a(x) - \Phi_{k_j - v_j}(x^{k_j - v_j}, x)} \\ &\leq \rho_{k_j - v_j} + \frac{(1 + \zeta)^2(\varepsilon' + \varepsilon'')\|x^{k_j - v_j} - x_j\|}{(1 - \zeta)^2\eta\|x_j - x^{k_j - v_j}\|}, \end{aligned} \quad (97)$$

so that $\eta \leq \frac{\varepsilon' + \varepsilon''}{\tilde{\gamma} - \gamma}$, as ζ was arbitrary. Since $\varepsilon' > 0$ was chosen arbitrarily,

$$\eta \leq \frac{\varepsilon''}{\tilde{\gamma} - \gamma}. \quad (98)$$

Hence, with $\partial\phi(\bar{x}, \bar{x}) \subseteq \tilde{\partial}_a f(\bar{x})$ from Lemma 4.3, 2., we obtain $\text{dist}(0, \tilde{\partial}_a f(\bar{x})) \leq \eta$. From $f_a(\bar{x}) = f(\bar{x})$, it follows that $\text{dist}(0, \partial f(\bar{x})) \leq \eta$. \square

A.3 Proof of Corollary 4.12

Proof. Follow the proof of Lemma 4.7 in Appendix A.1 and in part v), choose not only ε' but also $\varepsilon'' > 0$ arbitrarily. As $\tau_k \rightarrow \infty$ and $x^k \rightarrow x$ has been assumed or shown, respectively, there exists a k large enough such that (74) holds. In part vi), we then replace (78) by

$$\eta \leq 0, \quad (99)$$

as $\varepsilon'' > 0$ was chosen arbitrarily, and the claims follow. \square

A.4 Proof of Corollary 4.13

Proof. Follow the proof of Theorem 4.11 in Appendix A.2 and in order to derive (95), choose not only ε' but also $\varepsilon'' > 0$ arbitrarily. As $\tau_{k_j - v_j} \rightarrow \infty$ and $\|x^{k_j - v_j} - x_j\| \rightarrow 0$ has been assumed or shown, respectively, there exists a j large enough such that (95) holds. We can then replace (98) by

$$\eta \leq 0, \quad (100)$$

as $\varepsilon'' > 0$ was chosen arbitrarily, and the claims follow. \square

A.5 Proof of Corollary 4.14

Proof. In the definition of $\phi(\cdot, x)$ in (33), for the cutting plane at x use the exact evaluation at x , such that $m_{x,g}(\cdot, x) = f(x) + g^T(\cdot - x)$ with $g \in \partial f(x)$. Then, follow the proof of Lemma 4.7 in Appendix A.1 with some minor changes:

In part 1.), i), we note that g_x is not fixed but a sequence g_x^k . Boundedness of the sequence follows from boundedness of \mathcal{U} and Lemma 2.6. We obtain $x^k \rightarrow x$ and thus $f_a^k(x) \rightarrow f(x)$.

In part ii), we can show $\phi_k(x^k, x) \rightarrow f(x)$.

In part iii), we use that for every $y \in \mathbb{R}^n$, there is a \bar{k} such that for every $k > \bar{k}$, $\phi(y, x) \geq \phi_k(y, x)$. It follows that $g^* \in \partial\phi(x, x) \subseteq \partial f(x)$.

The rest of the proof can be carried out analogously to the proofs of Lemma 4.7 and Corollary 4.12 in Appendix A.1 and A.3, respectively, yielding $0 \in \partial f(x)$. \square

B Parameters for numerical results

Table 4: Parameters for the tests with GasLib40 and GasLib134, leading to the results in Tables 3 and 2.

	parameter	value
0	Γ	0.9
1	T	1e+100
2	c	1e-05
3	γ	0.1
4	$\tilde{\gamma}$	0.9
5	$\tau_1^\#$	1
6	tol	1e-07
7	ψ	10
8	ψ_Δ	1e+04

References

- [ALS19] Aßmann, D., Liers, F., and Stingl, M. “Decomposable robust two-stage optimization: an application to gas network operations under uncertainty”. In: *Networks* 74.1 (2019), pp. 40–61. ISSN: 0028-3045. DOI: 10.1002/net.21871.
- [ANP08] Apkarian, P., Noll, D., and Prot, O. “A trust region spectral bundle method for nonconvex eigenvalue optimization”. In: *SIAM J. Optim.* 19.1 (2008), pp. 281–306. ISSN: 1052-6234. DOI: 10.1137/060665191.
- [ANP09] Apkarian, P., Noll, D., and Prot, O. “A proximity control algorithm to minimize nonsmooth and nonconvex semi-infinite maximum eigenvalue functions”. In: *J. Convex Anal.* 16.3-4 (2009), pp. 641–666. ISSN: 0944-6532.
- [BEN09] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. *Robust optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ, 2009, pp. xxii+542. ISBN: 978-0-691-14368-2. DOI: 10.1515/9781400831050.
- [BHV15] Ben-Tal, A., Hertog, D. den, and Vial, J.-P. “Deriving robust counterparts of nonlinear uncertain inequalities”. In: *Math. Program.* 149.1-2, Ser. A (2015), pp. 265–299. ISSN: 0025-5610. DOI: 10.1007/s10107-014-0750-8.
- [Bie+20] Bienstock, D., Escobar, M., Gentile, C., and Liberti, L. *Mathematical Programming formulations for the Alternating Current Optimal Power Flow problem*. 2020. arXiv: 2007.05334 [math.OC].
- [Car62] Carpentier, J. “Contribution a l’etude du dispatching economique”. In: *Bulletin de la Societe Francaise des Electriciens* 3.1 (1962), pp. 431–447.
- [Cla75] Clarke, F. H. “Generalized gradients and applications”. In: *Trans. Amer. Math. Soc.* 205 (1975), pp. 247–262. ISSN: 0002-9947. DOI: 10.2307/1997202.

- [Cla90] Clarke, F. H. *Optimization and nonsmooth analysis*. Second. Vol. 5. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990, pp. xii+308. ISBN: 0-89871-256-4. DOI: 10.1137/1.9781611971309.
- [Col+78] Collins, M., Cooper, L., Helgason, R., Kennington, J., and LeBlanc, L. “Solving the pipe network analysis problem using optimization techniques”. In: *Management Sci.* 24.7 (1977/78), pp. 747–760. ISSN: 0025-1909. DOI: 10.1287/mnsc.24.7.747.
- [DBK06] Diehl, M., Bock, H. G., and Kostina, E. “An approximation technique for robust nonlinear optimization”. In: *Math. Program.* 107.1-2, Ser. B (2006), pp. 213–230. ISSN: 0025-5610. DOI: 10.1007/s10107-005-0685-1.
- [DG04] Daniilidis, A. and Georgiev, P. “Approximate convexity and submonotonicity”. In: *J. Math. Anal. Appl.* 291.1 (2004), pp. 292–301. ISSN: 0022-247X. DOI: 10.1016/j.jmaa.2003.11.004.
- [DM90] Demyanov, V. F. and Malozëmov, V. N. *Introduction to minimax*. Translated from the Russian by D. Louvish, Reprint of the 1974 edition. Dover Publications, Inc., New York, 1990, pp. viii+307. ISBN: 0-486-66423-6.
- [Fud+15] Fuduli, A., Gaudioso, M., Giallombardo, G., and Miglionico, G. “A partially inexact bundle method for convex semi-infinite minmax problems”. In: *Commun. Nonlinear Sci. Numer. Simul.* 21.1-3 (2015), pp. 172–180. ISSN: 1007-5704. DOI: 10.1016/j.cnsns.2014.07.033.
- [GAN13] Gabarrou, M., Alazard, D., and Noll, D. “Design of a flight control architecture using a non-convex bundle method”. In: *Math. Control Signals Systems* 25.2 (2013), pp. 257–290. ISSN: 0932-4194. DOI: 10.1007/s00498-012-0093-z.
- [GD82] Gauvin, J. and Dubeau, F. “Differential properties of the marginal function in mathematical programming”. In: *Math. Programming Stud.* 19 (1982). Optimality and stability in mathematical programming, pp. 101–119. ISSN: 0303-3929. DOI: 10.1007/bfb0120984.
- [Gei+12] Geißler, B., Martin, A., Morsi, A., and Schewe, L. “Using piecewise linear functions for solving MINLPs”. In: *Mixed integer nonlinear programming*. Vol. 154. IMA Vol. Math. Appl. Springer, New York, 2012, pp. 287–314.
- [GGM06] Gaudioso, M., Giallombardo, G., and Miglionico, G. “An incremental method for solving convex finite min-max problems”. In: *Math. Oper. Res.* 31.1 (2006), pp. 173–187. ISSN: 0364-765X. DOI: 10.1287/moor.1050.0175.
- [Got+16] Gotzes, C., Heitsch, H., Henrion, R., and Schultz, R. “On the quantification of nomination feasibility in stationary gas networks with random load”. In: *Math. Methods Oper. Res.* 84.2 (2016), pp. 427–457. ISSN: 1432-2994. DOI: 10.1007/s00186-016-0564-y.
- [GS16] Goerigk, M. and Schöbel, A. “Algorithm engineering in robust optimization”. In: *Algorithm engineering*. Vol. 9220. Lecture Notes in Comput. Sci. Springer, Cham, 2016, pp. 245–279.
- [GYH15] Gorissen, B. L., Yanıkoğlu, İ., and Hertog, D. den. “A practical guide to robust optimization”. In: *Omega* 53 (2015), pp. 124–137. ISSN: 0305-0483. DOI: <https://doi.org/10.1016/j.omega.2014.12.006>.

- [HD13] Houska, B. and Diehl, M. “Nonlinear robust optimization via sequential convex bilevel programming”. In: *Math. Program.* 142.1-2, Ser. A (2013), pp. 539–577. ISSN: 0025-5610. DOI: 10.1007/s10107-012-0591-2.
- [HS10] Hare, W. and Sagastizábal, C. “A redistributed proximal bundle method for nonconvex optimization”. In: *SIAM J. Optim.* 20.5 (2010), pp. 2442–2473. ISSN: 1052-6234. DOI: 10.1137/090754595.
- [HSS16] Hare, W., Sagastizábal, C., and Solodov, M. “A proximal bundle method for nonsmooth nonconvex functions with inexact information”. In: *Comput. Optim. Appl.* 63.1 (2016), pp. 1–28. ISSN: 0926-6003. DOI: 10.1007/s10589-015-9762-4.
- [HU19] Hertlein, L. and Ulbrich, M. “An inexact bundle algorithm for nonconvex nonsmooth minimization in Hilbert space”. In: *SIAM J. Control Optim.* 57.5 (2019), pp. 3137–3165. ISSN: 0363-0129. DOI: 10.1137/18M1221849.
- [Jab08] Jabr, R. A. “Optimal Power Flow Using an Extended Conic Quadratic Formulation”. In: *IEEE Transactions on Power Systems* 23.3 (Aug. 2008), pp. 1000–1008. DOI: 10.1109/TPWRS.2008.926439.
- [JLT98] Jofré, A., Luc, D. T., and Théra, M. “ ε -subdifferential and ε -monotonicity”. In: *Nonlinear Anal.* 33.1 (1998), pp. 71–90. ISSN: 0362-546X. DOI: 10.1016/S0362-546X(97)00511-7.
- [KDS18] Kocuk, B., Dey, S. S., and Sun, X. A. “Matrix minor reformulation and SOCP-based spatial branch-and-cut method for the AC optimal power flow problem”. In: *Mathematical Programming Computation* 10.4 (Dec. 2018), pp. 557–596. DOI: 10.1007/s12532-018-0150-9.
- [Kiw06] Kiwiel, K. C. “A proximal bundle method with approximate subgradient linearizations”. In: *SIAM J. Optim.* 16.4 (2006), pp. 1007–1023. ISSN: 1052-6234. DOI: 10.1137/040603929.
- [Kiw83] Kiwiel, K. C. “An aggregate subgradient method for nonsmooth convex minimization”. In: *Math. Programming* 27.3 (1983), pp. 320–341. ISSN: 0025-5610. DOI: 10.1007/BF02591907.
- [Kiw85] Kiwiel, K. C. “An algorithm for nonsmooth convex minimization with errors”. In: *Math. Comp.* 45.171 (1985), pp. 173–180. ISSN: 0025-5718. DOI: 10.2307/2008055.
- [Koc+15] Koch, T., Hiller, B., Pfetsch, M. E., and Schewe, L. *Evaluating Gas Network Capacities*. Ed. by M. E. Pfetsch, T. Koch, L. Schewe, and B. Hiller. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015. DOI: 10.1137/1.9781611973693. eprint: <https://epubs.siam.org/doi/pdf/10.1137/1.9781611973693>.
- [Ley+20] Leyffer, S., Menickelly, M., Munson, T., Vanaret, C., and Wild, S. M. “A survey of nonlinear robust optimization”. In: *INFOR Inf. Syst. Oper. Res.* 58.2 (2020), pp. 342–373. ISSN: 0315-5986. DOI: 10.1080/03155986.2020.1730676.
- [LPM18] Lv, J., Pang, L.-P., and Meng, F.-Y. “A proximal bundle method for constrained nonsmooth nonconvex optimization with inexact information”. In: *J. Global Optim.* 70.3 (2018), pp. 517–549. ISSN: 0925-5001. DOI: 10.1007/s10898-017-0565-2.

- [LV98] Lukšan, L. and Vlček, J. “A bundle-Newton method for nonsmooth unconstrained minimization”. In: *Math. Programming* 83.3, Ser. A (1998), pp. 373–391. ISSN: 0025-5610. DOI: 10.1016/S0025-5610(97)00108-1.
- [Mäk02] Mäkelä, M. M. “Survey of bundle methods for nonsmooth optimization”. In: *Optim. Methods Softw.* 17.1 (2002), pp. 1–29. ISSN: 1055-6788. DOI: 10.1080/10556780290027828.
- [MM57] Markowitz, H. M. and Manne, A. S. “On the solution of discrete programming problems”. In: *Econometrica* 25 (1957), pp. 84–110. ISSN: 0012-9682. DOI: 10.2307/1907744.
- [MOZ17] Malick, J., Oliveira, W. de, and Zaourar, S. “Uncontrolled inexact information within bundle methods”. In: *EURO J. Comput. Optim.* 5.1-2 (2017), pp. 5–29. ISSN: 2192-4406. DOI: 10.1007/s13675-015-0060-9.
- [MR18] Molzahn, D. K. and Roald, L. A. “Towards an AC optimal power flow algorithm with robust feasibility guarantees”. In: *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018, pp. 1–7.
- [MW20] Menickelly, M. and Wild, S. M. “Derivative-free robust optimization by outer approximations”. In: *Math. Program.* 179.1-2, Ser. A (2020), pp. 157–193. ISSN: 0025-5610. DOI: 10.1007/s10107-018-1326-9.
- [NB10] Nedić, A. and Bertsekas, D. P. “The effect of deterministic noise in subgradient methods”. In: *Math. Program.* 125.1, Ser. A (2010), pp. 75–99. ISSN: 0025-5610. DOI: 10.1007/s10107-008-0262-5.
- [Nol10] Noll, D. “Cutting plane oracles to minimize non-smooth non-convex functions”. In: *Set-Valued Var. Anal.* 18.3-4 (2010), pp. 531–568. ISSN: 1877-0533. DOI: 10.1007/s11228-010-0159-3.
- [Nol13] Noll, D. “Bundle method for non-convex minimization with inexact subgradients and function values”. In: *Computational and analytical mathematics*. Vol. 50. Springer Proc. Math. Stat. Springer, New York, 2013, pp. 555–592. DOI: 10.1007/978-1-4614-7621-4_26.
- [NPR08] Noll, D., Prot, O., and Rondepierre, A. “A proximity control algorithm to minimize nonsmooth and nonconvex functions”. In: *Pac. J. Optim.* 4.3 (2008), pp. 571–604. ISSN: 1348-9151.
- [OSL14] Oliveira, W. de, Sagastizábal, C., and Lemaréchal, C. “Convex proximal bundle methods in depth: a unified analysis for inexact oracles”. In: *Math. Program.* 148.1-2, Ser. B (2014), pp. 241–277. ISSN: 0025-5610. DOI: 10.1007/s10107-014-0809-6.
- [Pfe+15] Pfetsch, M. E. et al. “Validation of nominations in gas network optimization: models, methods, and solutions”. In: *Optim. Methods Softw.* 30.1 (2015), pp. 15–53. ISSN: 1055-6788. DOI: 10.1080/10556788.2014.888426.
- [RH02] Rustem, B. and Howe, M. *Algorithms for worst-case design and applications to risk management*. Princeton University Press, Princeton, NJ, 2002, pp. xviii+389. ISBN: 0-691-09154-4.
- [Roc82] Rockafellar, R. T. “Favorable classes of Lipschitz-continuous functions in subgradient optimization”. In: *Progress in nondifferentiable optimization*. Vol. 8. IIASA Collaborative Proc. Ser. CP-82. Internat. Inst. Appl. Systems Anal., Laxenburg, 1982, pp. 125–143.

- [Spi81] Spingarn, J. E. “Submonotone subdifferentials of Lipschitz functions”. In: *Trans. Amer. Math. Soc.* 264.1 (1981), pp. 77–89. ISSN: 0002-9947. DOI: 10.2307/1998411.
- [Ste12] Stein, O. “How to solve a semi-infinite optimization problem”. In: *European J. Oper. Res.* 223.2 (2012), pp. 312–320. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2012.06.009.
- [SZ92] Schramm, H. and Zowe, J. “A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results”. In: *SIAM J. Optim.* 2.1 (1992), pp. 121–152. ISSN: 1052-6234. DOI: 10.1137/0802008.
- [VLT00] Van Ngai, H., Luc, D. T., and Théra, M. “Approximate convex functions”. In: *J. Nonlinear Convex Anal.* 1.2 (2000), pp. 155–176. ISSN: 1345-4773.
- [Yan+21] Yang, H., Morton, D. P., Bandi, C., and Dvijotham, K. “Robust Optimization for Electricity Generation”. In: *INFORMS Journal on Computing* 33.1 (2021), pp. 336–351. DOI: 10.1287/ijoc.2020.0956. eprint: <https://doi.org/10.1287/ijoc.2020.0956>.
- [ZNE90] Zuhe, S., Neumaier, A., and Eiermann, M. C. “Solving minimax problems by interval methods”. In: *BIT* 30.4 (1990), pp. 742–751. ISSN: 0006-3835. DOI: 10.1007/BF01933221.
- [ŽR03] Žaković, S. and Rustem, B. “Semi-infinite programming and applications to minimax problems”. In: vol. 124. Applied mathematical programming and modelling. 2003, pp. 81–110. DOI: 10.1023/B:ANOR.0000004764.76984.30.