AN APPROXIMATION SCHEME FOR DISTRIBUTIONALLY ROBUST NONLINEAR OPTIMIZATION

JOHANNES MILZ* AND MICHAEL ULBRICH*

Abstract. We consider distributionally robust optimization problems (DROPs) with nonlinear and nonconcave dependence on uncertain parameters. The DROP can be written as a nonsmooth, nonlinear program with a bilevel structure; the objective and each constraint function is the supremum of the expected value of a parametric function taken over an ambiguity set of probability distributions. We define ambiguity sets through moment constraints and to make the computation of first order stationary points tractable, we approximate nonlinear functions using quadratic expansions w.r.t. parameters, resulting in lower level problems defined by trust-region problems and semidefinite programs. Subsequently, we construct smoothing functions for the approximate lower level functions which are computationally tractable, employing strong duality for trust-region problems, and show that gradient consistency holds. We formulate smoothed DROPs and apply a homotopy method dynamically decreasing smoothing parameters and establish its convergence to stationary points of the approximate DROP under mild assumptions. Through our scheme, we provide a new approach to robust nonlinear optimization as well. We perform numerical simulations on a well-known test set, assuming design variables are subject to implementation errors, providing a representative set of numerical examples.

Key words. distributionally robust optimization, robust optimization, trust-region problem, semidefinite programming, smoothing functions, gradient consistency, smoothing methods

AMS subject classifications. 90C26, 90C30, 90C46, 90C59, 49M37

1. Introduction. We develop an approximation scheme for the nonlinear distributionally robust optimization problem (DROP)

$$(1.1) \qquad \min_{x \in X} \sup_{P \in \mathcal{P}} \mathbb{E}_P[f_0(x,\xi)] \quad \text{s.t.} \quad \sup_{P \in \mathcal{P}} \mathbb{E}_P[f_j(x,\xi)] \leq 0, \quad j \in J \setminus \{0\},$$

where $X \subset \mathbb{R}^n$ is the set of design variables and $f_j : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$, $j \in J \subset \mathbb{N}_0$. The ambiguity set $\mathcal{P}$ is defined through moment constraints of the random vector ξ and entropic dominance similar to [17, 20, 48]:

$$(1.2) \qquad \begin{aligned} \mathcal{P} = \{\, P \in \mathcal{M} : &\|\bar{\Sigma}^{-\frac{1}{2}}(\mathbb{E}_P[\xi] - \bar{\mu})\|_2 \leq \Delta, \quad \bar{\Sigma}_0 \preccurlyeq \operatorname{Cov}_P[\xi] \preccurlyeq \bar{\Sigma}_1, \\ &\ln \mathbb{E}_P[\exp(y^T(\xi - \mathbb{E}_P[\xi]))] \leq y^T \bar{\Sigma}_1 y \quad \text{for all} \quad y \in \mathbb{R}^p \,\}, \end{aligned}$$

where $\Delta > 0$, $\bar{\mu} \in \mathbb{R}^p$, and $\bar{\Sigma}_0$, $\bar{\Sigma}_1$, $\bar{\Sigma} \in \mathbb{R}^{p \times p}$ are symmetric, Σ_0, Σ_1 and $\Sigma_1 - \Sigma_0$ are positive semidefinite, and $\bar{\Sigma}$ is positive definite. The vector $\bar{\mu}$ and the matrices $\bar{\Sigma}_0$, $\bar{\Sigma}_1$, $\bar{\Sigma}$ are estimates for the mean and the covariance of the random vector ξ, respectively. Moreover, $\mathcal{M}$ denotes the set of probability distributions of ξ on $\mathbb{R}^p$.

In order to obtain tractable approximations of the objective and constraint functions of (1.1), we approximate $f_j(x, \cdot)$ using second order expansions $m_j(x, \cdot)$ defined by

$$(1.3) \qquad m_j(x, \xi) = a_j(x) + b_j(x)^T(\xi - \bar{\mu}) + (1/2)(\xi - \bar{\mu})^T C_j(x)(\xi - \bar{\mu}),$$

where $a_j : \mathbb{R}^n \to \mathbb{R}$, $b_j : \mathbb{R}^n \to \mathbb{R}^p$, $C_j : \mathbb{R}^n \to \mathbb{S}^p$. We formulate the approximated DROP

$$(1.4) \qquad \min_{x \in X} \sup_{P \in \mathcal{P}} \mathbb{E}_P[m_0(x, \xi)] \quad \text{s.t.} \quad \sup_{P \in \mathcal{P}} \mathbb{E}_P[m_j(x, \xi)] \leq 0, \quad j \in J \setminus \{0\}.$$

The definition of the ambiguity set $\mathcal{P}$ (see (1.2)) and

$$\mathbb{E}_P[m_j(x, \xi)] = a_j(x) + b_j(x)^T d + (1/2)d^T C_j(x)d + (1/2)C_j(x) \bullet \Sigma,$$

where $d = \mathbb{E}_P[\xi] - \bar{\mu}$ and $\Sigma = \mathrm{Cov}_P[\xi]$ implies that each lower level optimization problem of (1.4) separates into the semidefinite program (SDP)

$$(1.5) \qquad \varphi_j(x) = \max_{\Sigma \in \mathbb{S}^p} \left\{ (1/2)C_j(x) \bullet \Sigma : \quad \bar{\Sigma}_0 \preccurlyeq \Sigma \preccurlyeq \bar{\Sigma}_1 \right\},$$

and the nonconvex trust-region problem (TRP)

$$(1.6) \qquad \psi_j(x) = a_j(x) + \max_{d \in \mathbb{R}^p} \left\{ b_j(x)^T d + (1/2)d^T C_j(x)d : \quad \|\bar{\Sigma}^{-1/2}d\|_2 \leq \Delta \right\},$$

where $\psi_j : \mathbb{R}^n \to \mathbb{R}$ and $\varphi_j : \mathbb{R}^n \to \mathbb{R}$.

The optimal value functions (1.5) and (1.6) provide a tractable approximation of the lower level problems in (1.1). These functions lack higher order differentiability, motivating us to construct smoothing functions for them. We propose a homotopy method similar to smoothing methods in [15, 56] to solve a sequence of smoothed DROPs to obtain a Clarke stationary point of the approximated DROP (1.4).

The SDP in (1.5) can be solved analytically after computing the eigenvalues of a transformation of $C_j(x)$ (see [57, Thm. 2.2]). We make use of this and and apply results on spectral functions, such as statements established in [37, 53] to obtain a smoothing function of (1.5). Our approach for the value function of the TRP (1.6) utilizes strong duality for TRPs; see, e.g., [51]. We apply a reciprocal barrier function to its dual and observe that the dual is equivalent to a TRP.

Using Lagrangian duality for both (1.5) and (1.6) (see [5, Chap. 4] and [51]) we can show that (1.4) can be reformulated equivalently as a nonlinear SDP (NSDP). However, our approach allows the numerical treatment of (1.4) via a sequence of standard nonlinear programs (NLPs). Derivatives required for each NLP may be easier to obtain than those the NSDP formulation. In particular, our approach requires the derivative the function $\mathbb{R}^n \ni x \mapsto d^T C_j(x)d$, $d \in \mathbb{R}^p$, rather than of the mapping $C_j : \mathbb{R}^n \to \mathbb{S}^p$.

Distributionally robust optimization (DRO) is a popular methodology used to obtain robust solutions to optimization problems under uncertainty; cf. [20, 23, 29, 46, 55]. It "robustifies" against distributions contained in an ambiguity set. If this set is a singleton, DRO is reduced to stochastic optimization; see [47]. A very popular choice for constructing an ambiguity set is based on moment constraints of the parameters, such as the one in (1.2); cf. [20, 47, 48, 55]. Another approach is to define the set by measures close to a reference measure w.r.t. a certain distance; cf. [28, 46, 59].

Some special classes of DROPs can be transformed into one-level problems using Lagrangian duality. For example, if ambiguity sets are conic representable, maximization problems w.r.t. probability measures become conic linear programs and, therefore, can be transformed into minimization problems and concatenated with upper-level problems; cf. [20]. If suitable assumptions, such as the convexity of the objective function w.r.t. design variables, are satisfied, the resulting optimization

problem is tractable [20, 55]. The reformulation of the lower level problems of (1.4) as linear matrix inequalities has been discussed in the supplementary material of [55].

If the SDP (1.5) is removed from (1.4), we obtain the robust optimization problem (ROP)

$$(1.7) \qquad \min_{x \in X} \psi_0(x) \quad \text{s.t.} \quad \psi_j(x) \leq 0, \quad j \in J \setminus \{0\}.$$

Research on robust optimization (RO) may be divided into contributions assuming concave dependence w.r.t. parameters, see e.g., [2, 3, 5, 7], and those assuming non-concave dependence; see e.g., [21, 32, 58]. The authors of [21] and of [58] use a linearization scheme for nonlinear RO to obtain tractable approximations of lower level problems, resulting in a nonlinear second-order cone program if an ellipsoidal uncertainty set is used. Instead of linearization, second order models are applied in [34, 35]. These expansions may be more effective than linearizations and may provide a trade off between accuracy and tractability; cf. [32, 35]. This approach results in constraints such as the one in (1.7), which are reformulated using its canonical necessary and sufficient optimality conditions in [34, 35]. The resulting problem is a mathematical program with complementarity constraints (MPCCs); see e.g., [33, 50]. In addition, the constraint set contains linear matrix inequalities, requiring the Hessian matrix of the Lagrangian of each robustified constraint to be positive semidefinite. In [34, 35] the inequalities are reformulated using eigenvalue constraints, introducing nonsmooth constraint functions. Moreover, in [32] a numerical scheme for nonlinear min-max optimization problems has been developed. Nonconvex ROPs without approximation schemes have been considered in, e.g., [8, 9]. The lower level problems in (1.7) may be reformulated as SDPs; see [3, sect. 1.4 and Lem. 14.37].

Smoothing methods are popular schemes for the solution of nonconvex, nonsmooth, and Lipschitz optimization problems; see, e.g., [12, 15, 56]. Our algorithmic scheme is related to recent contributions, such as [12, 13, 15, 56], in that it provides further examples of smoothing functions and applies their concepts and methodology. We apply an NLP solver to compute stationary points of a sequence of smoothed DROPs generated by the decreasing parameters and, therefore, our algorithmic approach is similar to [15, 56].

Our scheme relies on the approximations of the lower level problems in (1.1). However, we are able to compute stationary points of the approximation (1.4) of (1.1) without the assumption that computationally available bounds on the Hessian matrix of $f_j(x, \cdot)$ as in [32] are known, and we do not require expensive numerical schemes as in [8, 9]. Our reformulation does not result in an MPCC or an NSDP, and we do not increase the dimension of the initial DROP or ROP. A further advantage is that we obtain standard NLPs with tractable objective and constraints. These conditions are all favorable from a computational point of view because, e.g., an implementation of further algorithms is not required, making our approach applicable to many problems.

Outline of the paper. In Section 2, the choice of the ambiguity set $\mathcal{P}$ (see (1.2)) is explained. Section 3 introduces the concept of smoothing function, a smoothed DROP of (1.4) and a homotopy method used for the numerical solution of (1.4). Section 4 presents our smoothing approach for the SDPs in (1.5), which utilizes theory of spectral functions. In Section 5 our smoothing scheme for the TRPs in (1.6) is presented. It is based on strong duality of TRPs. Global convergence of the homotopy method is shown in Section 6. Section 7 presents numerical examples illustrating that the approximation scheme (1.4) of (1.1) can be effective. Section 8 presents a concise summary of our contributions.

Notation. The set of symmetric $m \times m$-matrices is denoted by $\mathbb{S}^m$. We refer to $\mathbb{S}^m_{++} \subset \mathbb{S}^m$ ($\mathbb{S}^m_+ \subset \mathbb{S}^m$) as the set of positive (semi)definite matrices. The identity matrix is I. The eigenvalue mapping is $\lambda : \mathbb{S}^p \to \mathbb{R}^p$, where $\lambda(A)$ contains the eigenvalues of A in decreasing order, i.e., $\lambda_{\max}(A) = \lambda_1(A) \geq \cdots \geq \lambda_p(A) = \lambda_{\min}(A)$. Here, $A \succeq B$ ($A \succ B$) for $A, B \in \mathbb{S}^m$ means $A - B \in \mathbb{S}^m_+$ ($A - B \in \mathbb{S}^m_{++}$). We use $\bullet$ to denote the Frobenius inner-product on $\mathbb{S}^m$. The set $N(A)$ is the null space of $A \in \mathbb{S}^m$. The matrix $A^{1/2} \in \mathbb{S}^m$ is a square root of $A \in \mathbb{S}^m_+$, B^+ is the Moore-Penrose inverse of $B \in \mathbb{R}^{m \times m}$, $|J|$ is the cardinality of the set J, and $(\cdot)_+ = \max\{0, \cdot\}$. For $a \in \mathbb{R}^m$, $\mathrm{Diag}(a) \in \mathbb{S}^m$ is the diagonal matrix with $(\mathrm{Diag}(a))_{ii} = a_i$. The Euclidean norm ($\infty$-norm) on $\mathbb{R}^m$ is $\|\cdot\|_2$ ($\|\cdot\|_\infty$). The convex hull of $A \subset \mathbb{R}^{n \times m}$ is conv A. A function $h : \mathbb{R}^m \to \mathbb{R}$ is symmetric if it is invariant under coordinate permutations; see, e.g., [36]. The gradient of $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ w.r.t. x evaluated at (x, y) is denoted by $\nabla_x G(x, y)$. For $A : \mathbb{R}^n \to \mathbb{S}^p$, we denote by $DA(x)$ its derivative and with $DA(x)^*$ its adjoint operator evaluated at $x \in \mathbb{R}^n$. The set $\partial G(y)$ is the Clarke subdifferential of $G : \mathbb{R}^n \to \mathbb{R}$ at $y \in \mathbb{R}^n$ (cf. [18, p. 27]) consisting of column vectors. If $\xi \in \mathbb{R}^p$ is a random vector, we use $\mathbb{E}_P[\xi]$ and $\mathrm{Cov}_P[\xi]$ to denote its mean and covariance w.r.t $P \in \mathcal{M}$, respectively. Here, $\mathcal{M}$ is the set of probability distributions of ξ on $\mathbb{R}^p$. The normal distribution with mean $\mu \in \mathbb{R}^p$ and covariance matrix $\Sigma \in \mathbb{S}^p_+$ is $N(\mu, \Sigma)$.

2. Choice of Ambiguity Set. We comment on the choice of the ambiguity set $\mathcal{P}$ defined in (1.2), discuss conditions implying that the objective and constraint functions of the DROPs (1.1) and (1.4) are finite-valued, and suggest choices to define the quadratic model functions m_j (see (1.3)) of f_j.

We require that the mappings $\xi : \Omega \to \mathbb{R}^p$ and $f_j(x, \cdot) \circ \xi$ are random variables for all $x \in X$. The first two conditions on $\mathbb{E}_P[\xi]$ and $\mathrm{Cov}_P[\xi]$ imposed by $\mathcal{P}$ (see (1.2)), model confidence regions of the mean and the covariance of ξ under suitable assumptions, respectively; cf. [48, Thm. 9]. The condition $\ln \mathbb{E}_P[\exp(y^T(\xi - \mathbb{E}_P[\xi]))] \leq y^T \bar{\Sigma}_1 y$ for all $y \in \mathbb{R}^p$ implies that $\mathrm{Cov}_P[\xi] \preceq \bar{\Sigma}_1$; cf. [17, Thm. 2]. Furthermore, it can be shown that $\mathbb{E}_P[\|\xi\|_2^\gamma] < \infty$ for all $\gamma > 0$; cf. [11, sect. 1.1, sect. 7.1]. This implies that the objective and constraint functions of (1.1) are finite-valued for a large class of functions f_j, $j \in J$. For example, if f_j, $j \in J$, are q-times continuously differentiable, and their qth derivatives are uniformly Lipschitz continuous w.r.t. (x, ξ), we can show that the objective and constraint functions in (1.1) are finite-valued for all $x \in X$.

A worst-case distribution P_j^* of each lower level optimization problem in (1.4) exists and is contained in the ambiguity set $\mathcal{P}$. We have that $P_j^* = N(\bar{\mu} + d_j^*, \Sigma_j^*) \in \mathcal{P}$ (see [11, sect. 7.1]), where Σ_j^* is an optimal solution of (1.5) and d_j^* of (1.6).

We can choose the functions a_j, b_j and C_j as $a_j = f_j(\cdot, \bar{\mu})$, $b_j = \nabla_\xi f_j(\cdot, \bar{\mu})$ and $C_j = \nabla_{\xi\xi} f(\cdot, \bar{\mu})$, where $\nabla_{\xi\xi} f(x, \bar{\mu})$ denotes the Hessian matrix of $f(x, \cdot)$ evaluated at $(x, \bar{\mu})$. If $x \in \mathbb{R}^n$ and the second derivative of $f_j(x, \cdot)$ is Lipschitz continuous w.r.t. ξ with Lipschitz constant $L > 0$, i.e., $|f_j(x, \xi) - m_j(x, \xi)| \leq (L/6)\|\xi - \bar{\mu}\|_2^3$, for all $\xi \in \mathbb{R}^p$, it can be shown that the worst-case expected value of the truncation error

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P[|f_j(x, \xi) - m_j(x, \xi)|]$$

converges to zero as $\bar{\Sigma}_1 \to 0^+$ and $\Delta \to 0^+$. If $f_j(x, \cdot)$ are quadratic functions for each $x \in \mathbb{R}^n$ and a_j, b_j and C_j chosen as above, the functions f_j and m_j are equal and, hence, the approximation scheme is exact, i.e., (1.1) and (1.4) are equivalent.

3. Smooth DROPs, smoothing functions and a homotopy method. We outline our algorithmic scheme to compute a stationary point of (1.4). Introducing

---

Algorithm 3.1 Homotopy method

Choose parameters $t_0 \in \mathbb{R}^3_{++}$, $t_{\min} \in \mathbb{R}^3_+$, $\varepsilon_0 > 0$, $\varepsilon_{\min} \geq 0$ and $\rho \in (0,1)$.
For $k = 0, 1, \ldots$

    1. Compute an ε_k-KKT-tuple (x^k, ϑ^k) of (3.2) for $t = t^k$.
    2. If $t^k \leq t_{\min}$ and $\varepsilon_k \leq \varepsilon_{\min}$ hold, STOP and return (x^k, ϑ^k).
    3. Compute $0 < t^{k+1} \leq \rho t^k$ and $\varepsilon_{k+1} = \rho \varepsilon_k$.

---

the functions $F_j : \mathbb{R}^n \to \mathbb{R}$, $F_j(x) = \varphi_j(x) + \psi_j(x)$, $j \in J$, the DROP (1.4) becomes

$$(3.1) \qquad \min_{x \in X} F_0(x) \quad \text{s.t.} \quad F_j(x) \leq 0, \quad j \in J \setminus \{0\},$$

which is generally a nonsmooth optimization problem. In the subsequent sections, we construct smooth approximations $\widetilde{F}_j : \mathbb{R}^n \times \mathbb{R}^3_{++} \to \mathbb{R}$ of F_j parameterized by $t \in \mathbb{R}^3_{++}$. The formal definition of the functions $\widetilde{F}_j$ are given in (6.1). They are used in Algorithm 3.1 to compute a sequence of approximate KKT-points of

$$(3.2) \qquad \min_{x \in X} \widetilde{F}_0(x, t) \quad \text{s.t.} \quad \widetilde{F}_j(x; t) \leq 0, \quad j \in J \setminus \{0\},$$

as $t \to 0^+$. Since these DROPs are smooth, we can apply state of the art NLPs solvers to solve them. Throughout, let $X = \mathbb{R}^n$ hold, however, X may consist of finitely many inequality or equality constraints. Here, a point $(\bar{x}, \bar{\vartheta}) \in \mathbb{R}^n \times \mathbb{R}^{|J|-1}_+$ is referred to as KKT-tuple of (3.1) if $\bar{\vartheta}_j F_j(\bar{x}) = 0$, $F_j(\bar{x}) \leq 0$, $j \in J \setminus \{0\}$, and $0 \in \partial F_0(\bar{x}) + \sum_{j \in J \setminus \{0\}} \bar{\vartheta}_j \partial F_j(\bar{x})$. These are necessary optimality conditions for (3.1) if a constraint qualification (CQ) holds; see, e.g., [38, Cor. 5.1.8].

    We construct a smoothing function of φ_j and of ψ_j satisfying the conditions of the next definition, which is based on [15, Def. 1].

    DEFINITION 3.1. *Let* $\phi : \mathbb{R}^n \to \mathbb{R}$ *be a continuous function. The function* $\tilde{\phi} :$ $\mathbb{R}^n \times \mathbb{R}^m_{++} \to \mathbb{R}$ *is referred to as smoothing function of* ϕ *if* $\tilde{\phi}(\cdot; t)$ *is continuously differentiable for every* $t > 0$*, and for all* $x \in \mathbb{R}^n$*, it holds that*

$$\lim_{\mathbb{R}^n \ni x^k \to x, \; t^k \to 0^+} \tilde{\phi}(x^k; t^k) = \phi(x).$$

We allow for multiple smoothing parameters in Definition 3.1 as opposed to [15, Def. 1] because the smoothing function of ψ_j constructed in Subsection 5.3 depends on two. Algorithm 3.1 does not require to compute exact KKT-tuples of (3.2), which is important for an efficient numerical scheme for the DROP (3.1). Different notions of approximate KKT-points have been proposed in the literature; see, e.g., [1, 22]. We refer to (x, ϑ) as ε-KKT-tuple of (3.2) if $\chi(x, \vartheta; t) \leq \varepsilon$, where the criticality measure $\chi : \mathbb{R}^n \times \mathbb{R}^{|J|-1} \times \mathbb{R}^3_{++} \to \mathbb{R}_+$ is defined by

$$(3.3)$$
$$\chi(x, \vartheta; t) = \max_{j \in J \setminus \{0\}} \left\{ \left\| \nabla_x \widetilde{F}_0(x; t) + \sum_{j \in J \setminus \{0\}} \vartheta_j \nabla_x \widetilde{F}_j(x; t) \right\|_\infty, \; |\min\{-\widetilde{F}_j(x; t), \vartheta_j\}| \right\}.$$

    An important notion to establish convergence of Algorithm 3.1 to stationary points of (3.1) is gradient consistency. Let $\tilde{\phi} : \mathbb{R}^n \times \mathbb{R}^m_{>0} \to \mathbb{R}$ be a smoothing function of the locally Lipschitz continuous function $\phi : \mathbb{R}^n \to \mathbb{R}$. We define

$$(3.4) \quad S_{\tilde{\phi}}(x) = \text{conv}\left\{ z \in \mathbb{R}^n : \exists \mathbb{R}^n \times \mathbb{R}^m_{++} \ni (x^k, t^k) \to (x, 0), \, \nabla_x \tilde{\phi}(x^k; t_k) \to z \right\}.$$

Gradient consistency of $\tilde{\phi}$ and ϕ requires the following relation to hold; cf. [12, 13, 15]:

$$(3.5) \qquad S_{\tilde{\phi}}(x) = \partial\phi(x) \quad \text{for all} \quad x \in \mathbb{R}^n.$$

For the above setting, Clarke's subdifferential is a subset of (3.4) generalizing a remark in [15, sect. 1] to multiple smoothing parameters.

LEMMA 3.2. *Let $\tilde{\phi} : \mathbb{R}^n \times \mathbb{R}_{++}^m \to \mathbb{R}$ be a smoothing function of the locally Lipschitz continuous function $\phi : \mathbb{R}^n \to \mathbb{R}$. Then, we have that $\partial\phi(x) \subset S_{\tilde{\phi}}(x)$ for all $x \in \mathbb{R}^n$.*

Proof. Let $x \in \mathbb{R}^n$ be arbitrary and define $\tilde{\ell} : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ by $\tilde{\ell}(x;t) = \tilde{\phi}(x;te)$, which is a smoothing function of ϕ, where $e = (1,\ldots,1) \in \mathbb{R}^m$. Hence, [14, Lem. 3.1] implies $\partial\phi(x) \subset S_{\tilde{\ell}}(x)$. Using (3.4), we obtain $S_{\tilde{\ell}}(x) \subset S_{\tilde{\phi}}(x)$ concluding the proof. □

In the next two sections, we construct smoothing functions of (1.5) and (1.6) that can efficiently be evaluated as well as its gradients. Moreover, they satisfy gradient consistency.

4. Smoothing approach for the SDPs. We construct a smoothing function of φ_j (see (1.5)) satisfying the conditions stated in Section 3 for the algorithmic solution of the DROP (3.1). We use that the SDPs (1.5) can be solved analytically after computing the eigenvalues of a transformation of $C_j(x)$; cf. [57, Thm. 2.2].

PROPOSITION 4.1. *Let $C \in \mathbb{S}^p$ and $X_0, X_1 \in \mathbb{S}^p$ fulfill $X_0 \prec X_1$, and define $G = (X_1 - X_0)^{1/2} C (X_1 - X_0)^{1/2}$. Then, it holds that*

$$(4.1) \qquad C \bullet X_0 + \sum_{i=1}^{p} \min\{0, \lambda_i(G)\} = \min\left\{ C \bullet X : X_0 \preccurlyeq X \preccurlyeq X_1 \right\}.$$

Proof. The statement follows from an application of [57, Thm. 2.2]. □

Numerical simulations for dimensions $p \in \{1,\ldots,2000\}$ have indicated that this solution method is significantly faster than state of the art SDP solvers. If $\bar{\Sigma}_0 \prec \bar{\Sigma}_1$, (1.5), Proposition 4.1 and (4.1) show that

$$(4.2) \qquad \varphi_j(x) = (1/2)C_j(x) \bullet \bar{\Sigma}_0 + (1/2)\sum_{i=1}^{p}(\lambda_i(G_j(x)))_+ \quad \text{for all} \quad x \in \mathbb{R}^n,$$

where $G_j : \mathbb{R}^n \to \mathbb{S}^p$, $G_j(x) = (\bar{\Sigma}_1 - \bar{\Sigma}_0)^{1/2} C_j(x)(\bar{\Sigma}_1 - \bar{\Sigma}_0)^{1/2}$. In particular, φ_j is generally nonsmooth. Next, we show that the function $\tilde{\varphi}_j : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ defined by

$$(4.3) \qquad \tilde{\varphi}_j(x;\tau) = (1/2)C_j(x) \bullet \bar{\Sigma}_0 + (1/2)\tilde{w}(\lambda(G_j(x));\tau),$$

is a smoothing function of φ_j, where $\tilde{w} : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}$ is given by

$$(4.4) \qquad \tilde{w}(z;\tau) = \tau\sum_{i=1}^{p} \ln\left(1 + \exp\left(z_i/\tau\right)\right).$$

THEOREM 4.2. *Let $\bar{\Sigma}_0 \prec \bar{\Sigma}_1$ hold and let $C_j : \mathbb{R}^n \to \mathbb{S}^p$ be q-times continuously differentiable, where $q \geq 1$ and $j \in J$. Then, the following conditions hold true.*
1. *For all $(x,\tau) \in \mathbb{R}^n \times \mathbb{R}_{++}$, we have that*

$$(4.5) \qquad \varphi_j(x) \leq \tilde{\varphi}_j(x;\tau) \leq \varphi_j(x) + (1/2)\tau p \ln 2,$$

where φ_j and $\tilde{\varphi}_j$ is defined in (4.2) and (4.3), respectively.

2. *The function $\tilde{\varphi}_j$ is a smoothing function of φ_j, $\tilde{\varphi}_j(\cdot;\tau)$ is q-times continuously differentiable for every $\tau > 0$, and gradient consistency holds for $\tilde{\varphi}_j$ and φ_j.*

3. *If $(x^k) \subset \mathbb{R}^n$ and $(\tau_k) \subset \mathbb{R}_{++}$ are sequences such that $x^k \to x$ and $\tau_k \to 0$ as $k \to \infty$, there exists a convergent subsequence $(\nabla_x\tilde{\varphi}_j(x^k;\tau^k))_K$ of $(\nabla_x\tilde{\varphi}_j(x^k;\tau^k))$.*

Proof. 1. The estimate (4.5) follows from the inequalities (see, e.g., [44, sect. 2])

$$(z)_+ \leq \tau \ln(1 + \exp(z/\tau)) \leq (z)_+ + \tau \ln 2 \quad \text{for all} \quad z \in \mathbb{R}.$$

2. Next, we establish that $\tilde{\varphi}_j$ is a smoothing function of φ_j. Let $\tau > 0$ be arbitrary. The function φ_j is locally Lipschitz continuous as a composition of locally Lipschitz functions and $\tilde{w}(\cdot;\tau)$ is symmetric and analytic as a composition of analytic functions. Hence, [53, Thm. 2.1] implies that $\tilde{w}_\lambda(\cdot;\tau) = \tilde{w}(\cdot;\tau) \circ \lambda$ is analytic, and the classical chain rule implies that $\tilde{\varphi}_j(\cdot;\tau) = (1/2)\tilde{w}_\lambda(\cdot;\tau) \circ G_j$ is q-times continuously differentiable. Together with (4.5), we obtain that $\tilde{\varphi}_j$ is a smoothing function of φ_j.

Now, we prove that gradient consistency holds, i.e., (3.5) is fulfilled. Since $\tilde{\varphi}_j$ is locally Lipschitz continuous, it suffices to show that $S_{\tilde{\varphi}_j}(x) \subset \partial\varphi_j(x)$ for all $x \in \mathbb{R}^n$; cf. Lemma 3.2, where $S_{\tilde{\varphi}_j}(x)$ is defined in (3.4). Let $x \in \mathbb{R}^n$ be arbitrary and let $z \in \mathbb{R}^n$ be a vector such that there exists sequences $(x^k) \subset \mathbb{R}^n$ and $(\tau_k) \subset \mathbb{R}_{++}$ converging to x and 0 as $k \to \infty$, respectively, and, moreover, such that

$$\nabla_x\tilde{\varphi}_j(x^k;\tau_k) \to z \quad \text{as} \quad k \to \infty.$$

If we conclude that $z \in \partial\varphi_j(x)$, we have $S_{\tilde{\varphi}_j}(x) \subset \partial\varphi_j(x)$; see (3.4).

Now, let $k \geq 0$ be arbitrary. We compute $\nabla_x\tilde{\varphi}_j(x^k;\tau_k)$. The function $\tilde{w}(\cdot;\tau_k)$ is continuously differentiable and symmetric and, hence, the classical chain rule and [36, Thm. 1.1] imply that the directional derivative $D_x\tilde{\varphi}_j(\cdot;\tau_k)h$ of $\tilde{\varphi}_j(\cdot;\tau_k)$ w.r.t. x evaluated at x^k in direction $h \in \mathbb{R}^p$ is

$$D_x\tilde{\varphi}_j(x^k;\tau_k)h = (1/2)\bar{\Sigma}_0 \bullet DC_j(x^k)h + (1/2)(Q_{j,k}M_{j,k}Q_{j,k}^T) \bullet DG_j(x^k)h,$$

where $Q_{j,k} \in \mathbb{R}^{p \times p}$ fulfills $Q_{j,k}Q_{j,k}^T = I$ and $G_j(x^k) = Q_{j,k}\text{Diag}(\lambda(G_j(x^k)))Q_{j,k}^T$, and where $M_{j,k} = \text{Diag}(\nabla_x\tilde{w}(\lambda(G_j(x^k));\tau_k))$. Using the adjoint operators $DC_j(x^k)^*$ and $DG_j(x^k)^*$ of $DC_j(x^k)$ and $DG_j(x^k)$, we obtain that

$$(4.6) \qquad \nabla_x\tilde{\varphi}_j(x^k;\tau_k) = (1/2)DC_j(x^k)^*\bar{\Sigma}_0 + (1/2)DG_j(x^k)^*(Q_{j,k}M_{j,k}Q_{j,k}^T).$$

We have that

$$(4.7) \qquad DC_j(x)^*P = \nabla_x(C_j(x) \bullet P), \quad \text{and} \quad DG_j(x^k)^*P = \nabla_x(G_j(x^k) \bullet P)$$

for all $P \in \mathbb{S}^p$. Indeed, for any $s \in \mathbb{R}^n$ and $P \in \mathbb{S}^p$, we infer that

$$s^T DC_j(x)^*P = P \bullet DC_j(x)s = D(C_j(x) \bullet P)s = s^T\nabla_x(C_j(x) \bullet P).$$

The second equation in (4.7) can be shown similarly.

Using (4.4), we obtain

$$(4.8) \qquad (\nabla_x\tilde{w}(z;\tau))_i = \frac{1}{1 + \exp(-z_i/\tau)}$$

for all $(z,\tau) \in \mathbb{R} \times \mathbb{R}_{++}$ and $i = 1,\ldots,p$. We deduce that $(\nabla_x\tilde{w}(\lambda(G_j(x^k));\tau_k))$ is bounded. Moreover, $(Q_{j,k})$ is bounded. Hence, we can assume w.l.o.g. that there exist $\bar{u}^j \in \mathbb{R}^p$ and $\bar{Q}_j \in \mathbb{R}^{p \times p}$ such that

$$\nabla_x\tilde{w}(\lambda(G_j(x^k));\tau_k) \to \bar{u}^j, \quad \text{and} \quad Q_{j,k} \to \bar{Q}_j \quad \text{as} \quad k \to \infty,$$

with $\bar{Q}_j\bar{Q}_j^T = I$ and $G_j(x) = \bar{Q}_j\mathrm{Diag}(\lambda(G_j(x)))\bar{Q}_j^T$, where we have used that λ is continuous; cf. [31, Cor. 6.3.8]. In addition, (4.8) implies for $i = 1, \ldots, p$, that

$$(\nabla_x\tilde{w}(\lambda(G_j(x^k)); \tau_k))_i \to (\bar{u}^j)_i \in \begin{cases} \{0\} & \text{if} \quad \lambda_i(G_j(x)) < 0, \\ [0,1] & \text{if} \quad \lambda_i(G_j(x)) = 0, \quad \text{as} \quad k \to \infty. \\ \{1\} & \text{if} \quad \lambda_i(G_j(x)) > 0, \end{cases}$$

Hence, (4.6) and the continuity of both DC_j and DG_j show that

$$\nabla_x\tilde{\varphi}_j(x^k; \tau_k) \to (1/2)DC_j(x)^*\bar{\Sigma}_0 + (1/2)DG_j(x)^*\bar{Q}_j\mathrm{Diag}(\bar{u}^j)\bar{Q}_j^T = z \quad \text{as} \quad k \to \infty.$$

To verify that $z \in \partial\varphi_j(x)$, we compute $\partial\varphi_j(x)$ using (4.2). The function $\mathbb{S}^p \ni G \mapsto \sum_{i=1}^p (\lambda_i(G))_+$ is regular (cf. [37, Cor. 4]), sums of regular functions are regular, and continuously differentiable functions are regular; cf. [19, Prop. 2.3.6]. Hence, through applications of the chain rule [18, Thm. 2.3.10], and [37, Thm. 8], we obtain that

$$\partial\varphi_j(x) = \left\{ \tfrac{1}{2}DC_j(x)^*\bar{\Sigma}_0 + \tfrac{1}{2}DG_j(x)^*Q\mathrm{Diag}(u)Q^T : Q \in O_j(x), u \in \partial w(\lambda(G_j(x))) \right\},$$

where $O_j(x) = \{Q \in \mathbb{R}^{p \times p} : QQ^T = I, G_j(x) = Q\mathrm{Diag}(\lambda(G_j(x)))Q^T\}$ and $w : \mathbb{R}^p \to \mathbb{R}$ is defined by $w(z) = \sum_{i=1}^p (z)_+$. For each $z \in \mathbb{R}^p$, and for all $i \in \{1, \ldots, p\}$ and $g \in \partial w(z)$, it holds that $g_i = 0$ if $z_i < 0$, $g_i \in [0,1]$ if $z_i = 0$, and $g_i = 1$ if $z_i > 0$. Hence, we infer $\bar{u}^j \in \partial w(\lambda(G_j(x)))$ and, finally, that $z \in \partial\varphi_j(x)$.

3. We can adapt the above reasoning to deduce that $(\nabla_x\tilde{\varphi}_j(x^k; \tau^k))$ has a convergent subsequence if $(x^k) \subset \mathbb{R}^n$ and $(\tau_k) \subset \mathbb{R}_{++}$ fulfill $x^k \to x$, $\tau_k \to 0$ as $k \to \infty$.□

Based on an eigendecomposition of $G_j(x)$, the computation of $\nabla_x\tilde{\varphi}_j(x; \tau)$ is cheap; cf. (4.6). The next step in order to solve the DROP (3.1) efficiently is to construct a computationally tractable smoothing function of (1.6).

5. Smoothing approach for the TRPs. We derive a smoothing function of the optimal value function defined in (1.6) based on constructing one of the function $v : \mathbb{R}^n \to \mathbb{R}$ defined by

$$(5.1) \qquad v(x) = \min_{s \in \mathbb{R}^p} \left\{ (1/2)s^T H(x)s + g(x)^T s : \quad (1/2)\|s\|_2^2 \le (1/2)\Delta^2 \right\},$$

where $g : \mathbb{R}^n \to \mathbb{R}^p$ and $H : \mathbb{R}^n \to \mathbb{S}^p$. Throughout, let $\Delta > 0$ be satisfied. We obtain a smoothing function of (5.1) as a value function of a "lifted" TRP. The lifted TRP results from a barrier formulation of a Lagrangian dual of (5.1). Since TRPs are theoretically and practically tractable (see [6, sect. 2] and [40, sect. 5]), our construction implies that the smoothing function of v can be evaluated efficiently. Moreover, based on Danskin's theorem, we can deduce that the evaluations of derivatives of the smoothing function are computationally tractable as well. In addition, we establish gradient consistency and, thus, the smoothing function meets the conditions stated in Section 3. In particular, we infer that the DROP (3.1) can be solved by Algorithm 3.1. Our approximation and smoothing scheme can be applied to nonlinear ROPs as an alternative to methods used in, e.g., [21, 35].

5.1. Lagrangian dual of TRPs. Before we review properties of the Lagrangian dual of the nominal TRP

$$(5.2) \qquad \min_{s \in \mathbb{R}^p} (1/2)s^T Hs + g^T s \quad \text{s.t.} \quad (1/2)\|s\|_2^2 \le (1/2)\Delta^2,$$

where $g = g(x_0) \in \mathbb{R}^p$, $H = H(x_0) \in \mathbb{S}^p$, and $x_0 \in \mathbb{R}^n$, we state necessary and sufficient optimality conditions of (5.2); see, e.g., [49, Lem. 2.4, Lem. 2.8].

THEOREM 5.1. *The TRP* (5.2) *has an optimal solution* $s^* \in \mathbb{R}^p$. *Moreover, the vector* $s^* \in \mathbb{R}^p$ *is an optimal solution of* (5.2) *iff there exists* $\lambda^* \in \mathbb{R}$ *such that*

$$(5.3) \quad (H + \lambda^* I)s^* = -g, \quad \|s^*\|_2 \leq \Delta, \quad \lambda^*(\|s^*\|_2 - \Delta) = 0, \quad \lambda^* \geq 0, \quad H + \lambda^* I \succcurlyeq 0.$$

In addition, if (s^*, λ^*) *fulfills* (5.3) *and* $\lambda^* > -\lambda_{\min}(H)$, *then* s^* *is the unique optimal solution of* (5.2). *Moreover, if* (s_1^*, λ_1^*) *and* (s_2^*, λ_2^*) *fulfill* (5.3), *it holds that* $\lambda_1^* = \lambda_2^*$.

If (s^*, λ^*) satisfies (5.3), we refer to it as optimal primal-dual solution of (5.2). Next, we provide a definition of the hard case of the TRP (5.2).

DEFINITION 5.2. *Let* (s^*, λ^*) *be an optimal primal-dual solution of* (5.2). *If* $\lambda^* = -\lambda_{\min}(H)$ *holds, the* hard case *occurs for* (5.2), *and otherwise the* easy case.

The term "hard case" is due to [40] and the terminology of the "easy case" has been used in, e.g., [51]. If the hard case occurs for (5.2), i.e., if $(s^*, -\lambda_{\min}(H))$ is an optimal primal-dual solution of (5.2), we have that $g \perp N(H - \lambda_{\min}(H)I)$. Indeed, (5.3) implies for all $v \in N(H - \lambda_{\min}(H)I)$ that $v^T g = -v^T(H - \lambda_{\min}(H)I)s^* = 0$. Now, we state a result on Lagrangian duality of (5.2); cf. [6, 25, 27, 51, 52].

THEOREM 5.3 ([52, Prop. 3.1, Thm. 3.3, Cor. 3.4]). *A Lagrangian dual problem of* (5.2)*—phrased as a minimization problem—is given by*

$$(5.4) \qquad \min_{\lambda \in \mathbb{R}} d(\lambda) \quad \text{s.t.} \quad H + \lambda I \succcurlyeq 0, \quad \lambda \geq 0,$$

where $d : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ *is defined by*

$$(5.5) \quad d(\lambda) = \begin{cases} \frac{1}{2}g^T(H + \lambda I)^+ g + \frac{1}{2}\Delta^2 \lambda & \text{if} \quad \lambda \geq (-\lambda_{\min}(H))_+, \ g \perp N(H + \lambda I), \\ \infty & \text{else.} \end{cases}$$

Moreover, (5.4) *has a unique optimal solution* λ^*, *which is the unique Lagrange multiplier associated to* (5.2). *In addition, strong duality holds, i.e., the optimal value of* (5.2) *equals* $-d^*$, *where* d^* *denotes the optimal value of* (5.4).

We define the solution mapping $s : \mathbb{R} \to \mathbb{R}^p$ by

$$(5.6) \qquad s(\lambda) = -(H + \lambda I)^+ g$$

and summarize properties of the dual function d.

LEMMA 5.4. *The following conditions hold true.*
1. *The function* d *defined in* (5.5) *is convex and* $d(\lambda) \to \infty$ *as* $\lambda \to \infty$.
2. *If* $\lambda > (-\lambda_{\min}(H))_+$, *then* d *is twice continuously differentiable at* λ, *and*

$$(5.7) \qquad d'(\lambda) = -(1/2)\|s(\lambda)\|_2^2 + (1/2)\Delta^2.$$

3. *If* $g \neq 0$, *then* $d''(\lambda) > 0$ *for all* $\lambda > (-\lambda_{\min}(H))_+$.

Proof. The statements follow from [52, Prop. 3.2] and the proof of [52, Thm. 3.3]. □

5.2. Barrier formulation for the dual of TRPs. We state a barrier problem of (5.4) using a reciprocal barrier and show that an optimal solution of it is an approximate solution to (5.4). In Subsection 5.3, it is shown that the barrier problem corresponds to a "lifted" TRP justifying the use of a reciprocal barrier instead of a self-concordant one. Hence, it can be solved with any TRP solver enabling us to define

and evaluate a smoothing function of ψ_j (see (1.6)) and its derivatives efficiently and, subsequently, to solve the DROP (3.1). The barrier problem associated to (5.4) is

$$(5.8) \qquad \min_{\lambda \in \mathbb{R}} d(\lambda) + \nu B_\eta(\lambda) \quad \text{s.t.} \quad \lambda > E(-H; \eta), \quad \lambda > 0,$$

where $\nu, \eta > 0$ and the reciprocal barrier $B_\eta : ((E(-H; \eta))_+, \infty) \to \mathbb{R}$ is defined by

$$(5.9) \qquad B_\eta(\lambda) = \frac{1}{\lambda} + \frac{1}{\lambda - E(-H; \eta)},$$

see, e.g., [26, sect. 3.1]. Here, $E : \mathbb{S}^p \times \mathbb{R}_{++} \to \mathbb{R}$ is an entropy function defined by

$$(5.10) \qquad E(A; \eta) = \eta \ln \sum_{i=1}^{p} \exp(\lambda_i(A)/\eta).$$

It has successfully been used in the context of nonsmooth optimization, see, e.g., [16, 41], E is a smoothing function of $\lambda_{\max}$ and fulfills

$$(5.11) \qquad \lambda_{\max}(A) \leq E(A; \eta) \leq \lambda_{\max}(A) + \eta \ln p,$$

for all $A \in \mathbb{S}^p$ and every $\eta > 0$; cf. [41, eq. (17) and eq. (18)], and [31, Cor. 6.3.8]. In particular, for all $A \in \mathbb{S}^p$ and any $\eta > 0$, we have that

$$(5.12) \qquad \lambda_{\min}(A) = -\lambda_{\max}(-A) \geq -E(-A; \eta).$$

We could use the barrier function $((-\lambda_{\min}(H))_+, \infty) \ni \lambda \mapsto -\ln \lambda - \ln \det(H + \lambda I)$ in (5.8), which does not require to compute $\lambda_{\min}(H)$ and to smooth $\lambda_{\min}$. However, the resulting primal problem would not be a TRP and requires, e.g., an adapted version of [40, Alg. 3.2] for its numerical solution. Next, we show that (5.8) has a unique optimal solution for any $\nu, \eta > 0$.

LEMMA 5.5. *For every $\nu, \eta > 0$, the barrier problem (5.8) has a unique optimal solution $\lambda^*(\nu, \eta)$ and it holds $\lambda^*(\nu, \eta) > (E(-H; \eta))_+$, where E is defined in (5.10).*

Proof. Let $\nu, \eta > 0$ be arbitrary. Define the objective function of (5.8) by

$$(5.13) \qquad B_{\nu,\eta} : ((E(-H; \eta))_+, \infty) \to \mathbb{R}, \quad B_{\nu,\eta} = d + \nu B_\eta,$$

where d and B_η is defined in (5.5) and (5.9), respectively. Let $\lambda > (E(-H; \eta))_+$ be arbitrary. Since $(E(-H; \eta))_+ \geq (-\lambda_{\min}(H))_+$ holds (cf. (5.12)), we have that

$$B_{\nu,\eta}(\lambda) = \frac{1}{2} g^T (H + \lambda I)^{-1} g + \frac{1}{2} \Delta^2 \lambda + \frac{\nu}{\lambda} + \frac{\nu}{\lambda - E(-H; \eta)} \geq \frac{1}{2} \Delta^2 \lambda$$

showing that $B_{\nu,\eta}(\lambda) \to \infty$ as $\lambda \to \infty$. From (5.5), (5.9), and (5.13), we infer that

$$B_{\nu,\eta}(\lambda) \geq \frac{\nu}{\lambda} + \frac{\nu}{\lambda - E(-H; \eta)} \to \infty \quad \text{as} \quad \lambda \to (E(-H; \eta))_+.$$

Thus, (5.8) has an optimal solution $\lambda^*(\nu, \eta)$ and there holds $\lambda^*(\nu, \eta) > (E(-H; \eta))_+$.

Now, we show that $B_{\nu,\eta}$ is strictly convex. Lemma 5.4 implies that $B_{\nu,\eta}$ (cf. (5.13)) is twice continuously differentiable at λ with

$$(5.14) \qquad B'_{\nu,\eta}(\lambda) = -\frac{1}{2} g^T (H + \lambda I)^{-2} g - \frac{\nu}{\lambda^2} - \frac{\nu}{(\lambda - E(-H; \eta))^2} + \frac{1}{2} \Delta^2,$$

and

$$B''_{\nu,\eta}(\lambda) = g^T (H + \lambda I)^{-3} g + \frac{2\nu}{\lambda^3} + \frac{2\nu}{(\lambda - E(-H; \eta))^3} > 0,$$

implying that $B_{\nu,\eta}$ is strictly convex. Hence, $\lambda^*(\nu, \eta)$ is the unique solution of (5.8). □

For ν, $\eta > 0$, we denote by $\lambda^*(\nu, \eta)$ the optimal solution of (5.8); cf. Lemma 5.5.

THEOREM 5.6. *Let ν, $\eta > 0$ be arbitrary. Then, the following conditions hold.*
1. *We have that*

$$(5.15) \qquad \lambda^*(\nu, \eta) \geq \sqrt{2\nu}/\Delta, \quad and \quad \lambda^*(\nu, \eta) - E(-H; \eta) \geq \sqrt{2\nu}/\Delta,$$

where $\lambda^*(\nu, \eta)$ is the optimal solution of (5.8) and E is defined in (5.10).
2. *The point $\lambda^*(\nu, \eta)$ is an $(\sqrt{2\nu}\Delta + (1/2)\Delta^2\eta \ln p)$-optimal solution of (5.4), i.e.,*

$$(5.16) \qquad d^* \leq d(\lambda^*(\nu, \eta)) \leq d^* + \sqrt{2\nu}\Delta + (1/2)\Delta^2\eta \ln p,$$

where d^* denotes the optimal value of (5.4) and d is defined in (5.5).
3. *It holds that*

$$(5.17) \qquad d^* \leq d(\lambda^*(\nu, \eta)) + \nu B_\eta(\lambda^*(\nu, \eta)) \leq d^* + 2\sqrt{2\nu}\Delta + (1/2)\Delta^2\eta \ln p,$$

where the barrier function B_η is defined in (5.9).

We apply the following result to prove Theorem 5.6.

LEMMA 5.7. *Let η, $\epsilon > 0$ be arbitrary, and consider*

$$(5.18) \qquad \min_{\lambda \in \mathbb{R}} d(\lambda) \quad s.t. \quad \lambda \geq \epsilon, \quad \lambda \geq E(-H; \eta) + \epsilon.$$

Then, problem (5.18) has a unique optimal solution $\bar{\lambda}_{\eta,\epsilon}$. Moreover, it holds that

$$(5.19) \qquad d^* \leq d(\bar{\lambda}_{\eta,\epsilon}) = d^*_{\eta,\epsilon} \leq d^* + (1/2)\Delta^2(\eta \ln p + \epsilon),$$

where d^* denotes the optimal value of (5.4) and $d^*_{\eta,\epsilon}$ the one of (5.18).

Proof. We establish existence and uniqueness of solutions of (5.18). If $g = 0$, we obtain $d(\lambda) = (1/2)\Delta^2\lambda$. Hence, the optimal solution $\bar{\lambda}_{\eta,\epsilon}$ of (5.18) is given by $\bar{\lambda}_{\eta,\epsilon} = (E(-H; \eta))_+ + \epsilon$. If $g \neq 0$, Lemma 5.4 and (5.12) imply that the objective of (5.18) is coercive, twice continuously differentiable in an open neighborhood of the feasible set of (5.18), and $d''(\lambda) > 0$ for all $\lambda > (E(-H; \eta))_+$. Hence, there exists a unique optimal solution $\bar{\lambda}_{\eta,\epsilon}$ of (5.18).

Now, we establish (5.19). Since $\bar{\lambda}_{\eta,\epsilon} \geq (E(-H; \eta))_+ + \epsilon$, we have that $d^* \leq d(\bar{\lambda}_{\eta,\epsilon})$. Moreover, if $\lambda^* > (E(-H; \eta))_+ + \epsilon$ holds, we infer $d^* = d^*_{\eta,\epsilon}$, where λ^* denotes the optimal solution of (5.4). Hence, it remains to consider the case where

$$(-\lambda_{\min}(H))_+ \leq \lambda^* \leq (E(-H; \eta))_+ + \epsilon.$$

We define $\bar{\lambda} = \lambda^* + \eta \ln p + \epsilon$, and observe that $\bar{\lambda} \geq \epsilon$. From (5.11), we infer that

$$E(-H; \eta) \leq -\lambda_{\min}(H) + \eta \ln p \leq \lambda^* + \eta \ln p$$

showing that $\bar{\lambda} \geq E(-H; \eta) + \epsilon$. Hence, $\bar{\lambda}$ is feasible for (5.18). Lemma 5.4 implies that d is convex and differentiable at $\bar{\lambda}$. Therefore, we have that

$$d(\lambda^*) - d(\bar{\lambda}) \geq d'(\bar{\lambda})(\lambda^* - \bar{\lambda}) = -d'(\bar{\lambda})(\eta \ln p + \epsilon)$$

resulting in

$$d(\lambda^*) + d'(\bar{\lambda})(\eta \ln p + \epsilon) \geq d(\bar{\lambda}) \geq d(\bar{\lambda}_{\eta,\epsilon}).$$

Now, (5.6), Lemma 5.4 and (5.7) imply $d'(\bar{\lambda}) \leq (1/2)\Delta^2$ and, hence, (5.19) holds. $\square$

To prove the estimates in (5.16), we use that the functions $G_1 : (0, \infty) \to \mathbb{R}$ and $G_2 : (E(-H; \eta), \infty) \to \mathbb{R}$ defined by

$$G_1(\lambda) = -\ln \lambda, \quad \text{and} \quad G_2(\lambda) = -\ln(\lambda - E(-H; \eta))$$

are 1-self-concordant barrier functions of their domains; cf. [42, sect. 2.3.1, Ex. 2].

Proof of Theorem 5.6. 1. We establish (5.15). Recall that the objective of (5.8) is $B_{\nu,\eta}$; cf. (5.13). Lemma 5.5 implies that $B'_{\nu,\eta}(\lambda^*(\nu, \eta)) = 0$ and (5.14) results in

$$g^T(H + \lambda^*(\nu, \eta)I)^{-2}g + \frac{2\nu}{\lambda^*(\nu, \eta)^2} + \frac{2\nu}{(\lambda^*(\nu, \eta) - E(-H; \eta))^2} = \Delta^2.$$

Lemma 5.5 and (5.12) further yield $H + \lambda^*(\nu, \eta)I \in \mathbb{S}^p_{++}$ and, hence, we infer that

$$\frac{2\nu}{\lambda^*(\nu, \eta)^2} \leq \Delta^2, \quad \text{and} \quad \frac{2\nu}{(\lambda^*(\nu, \eta) - E(-H; \eta))^2} \leq \Delta^2$$

showing the estimates in (5.15).

2. Next, we verify (5.16). The point $\lambda^*(\nu, \eta)$ is feasible for (5.4) by (5.15) and, therefore, we have $d^* \leq d(\lambda^*(\nu, \eta))$. Now, let $\lambda > (E(-H; \eta))_+$ be arbitrary. Both functions G_1 and G_2 defined prior the proof are 1-self-concordant for their domains. Hence, we obtain from [42, Prop. 2.3.2] that

(5.20)
$$-\frac{1}{\lambda^*(\nu, \eta)}(\lambda - \lambda^*(\nu, \eta)) = G'_1(\lambda^*(\nu, \eta))(\lambda - \lambda^*(\nu, \eta)) \leq 1,$$

$$-\frac{1}{\lambda^*(\nu, \eta) - E(-H; \eta)}(\lambda - \lambda^*(\nu, \eta)) = G'_2(\lambda^*(\nu, \eta))(\lambda - \lambda^*(\nu, \eta)) \leq 1.$$

Further, $B'_{\nu,\eta}(\lambda^*(\nu, \eta)) = 0$ results in

$$d'(\lambda^*(\nu, \eta)) = -\nu B'_\eta(\lambda^*(\nu, \eta))$$

showing with (5.15), (5.20), and $\lambda^*(\nu, \eta) > (E(-H; \eta))_+$ that

$$d'(\lambda^*(\nu, \eta))(\lambda - \lambda^*(\nu, \eta)) = -\nu B'_\eta(\lambda^*(\nu, \eta))(\lambda - \lambda^*(\nu, \eta))$$

$$= \frac{\nu}{\lambda^*(\nu, \eta)^2}(\lambda - \lambda^*(\nu, \eta)) + \frac{\nu}{(\lambda^*(\nu, \eta) - E(-H; \eta))^2}(\lambda - \lambda^*(\nu, \eta))$$

$$\geq -\frac{\nu}{\lambda^*(\nu, \eta)} - \frac{\nu}{\lambda^*(\nu, \eta) - E(-H; \eta)}.$$

Next, the convexity of d (cf. Lemma 5.4), the above formula, and (5.15) yield that

(5.21)
$$d(\lambda^*(\nu, \eta)) - d(\lambda) \leq d'(\lambda^*(\nu, \eta))(\lambda^*(\nu, \eta) - \lambda)$$

$$\leq \frac{\nu}{\lambda^*(\nu, \eta)} + \frac{\nu}{\lambda^*(\nu, \eta) - E(-H; \eta)} \leq \frac{2\nu}{\sqrt{2\nu}}\Delta = \sqrt{2\nu}\Delta.$$

Now, we denote by $\bar{\lambda}_{\eta,\epsilon}$ the optimal solution of (5.18) for an arbitrary $\epsilon > 0$, which fulfills $\bar{\lambda}_{\eta,\epsilon} \geq (E(-H; \eta))_+ + \epsilon$; cf. Lemma 5.7. Furthermore, Lemma 5.7, (5.19) and (5.21) with $\lambda = \bar{\lambda}_{\eta,\epsilon}$ show that

$$d(\lambda^*(\nu, \eta)) \leq d(\bar{\lambda}_{\eta,\epsilon}) + \sqrt{2\nu}\Delta \leq d^* + \sqrt{2\nu}\Delta + (1/2)\Delta^2(\eta \ln p + \epsilon).$$

The latter inequalities hold for all $\epsilon > 0$ and, hence, we obtain (5.16).

3. We show (5.17). Using (5.9) and (5.15), we infer that $\nu B_\eta(\lambda^*(\nu, \eta)) > 0$ $\nu B_\eta(\lambda^*(\nu, \eta)) \leq \sqrt{2\nu}\Delta$, and $\lambda^*(\nu, \eta)$ is feasible for (5.4). Hence, (5.16) implies (5.17).□

The error estimates presented in Theorem 5.6 depend on $\ln p$ and on the prescribed trust-region radius Δ. Therefore, the data dependence is weak.

5.3. Smoothing function for TRPs. We show that the function $\tilde{v} : \mathbb{R}^n \times \mathbb{R}^2_{++} \to \mathbb{R}$ defined by

$$(5.22) \quad \tilde{v}(x; \nu, \eta) = \min_{\tilde{s} \in \mathbb{R}^{p+2}} \left\{ (1/2)\tilde{s}^T \widetilde{H}_\eta(x) \tilde{s} + \tilde{g}_\nu(x)^T \tilde{s} : \ (1/2)\|\tilde{s}\|_2^2 \le (1/2)\Delta^2 \right\}.$$

is a smoothing function of v (see (5.1)) and establish gradient consistency, where

$$(5.23) \quad \widetilde{H}_\eta(x) = \begin{bmatrix} H(x) & & \\ & 0 & \\ & & -E(-H(x); \eta) \end{bmatrix} \in \mathbb{S}^{p+2}, \quad \text{and} \quad \tilde{g}_\nu(x) = \begin{bmatrix} g(x) \\ \sqrt{2\nu} \\ \sqrt{2\nu} \end{bmatrix} \in \mathbb{R}^{p+2},$$

and $E(\cdot; \eta)$ is defined in (5.10). Subsequently, we apply these results to define a smoothing function of ψ_j (see (1.6)), to infer its gradient consistency, and to deduce computationally tractability—crucial properties for an efficient solution of approximated DROPs using Algorithm 3.1. To prove these properties, we use that a Lagrangian dual of (5.22) is

$$(5.24) \quad \min_{\lambda \in \mathbb{R}} d(\lambda; x) + \frac{\nu}{\lambda} + \frac{\nu}{\lambda - E(-H(x); \eta)} \quad \text{s.t.} \quad \lambda > E(-H(x); \eta), \quad \lambda > 0,$$

where $x \in \mathbb{R}^n$ and $d : ((-E(H(x); \eta))_+, \infty) \times \mathbb{R}^n \to \mathbb{R}$ is defined by

$$(5.25) \quad d(\lambda; x) = (1/2)g(x)^T (H(x) + \lambda I)^{-1} g(x) + (1/2)\Delta^2 \lambda.$$

LEMMA 5.8. *Let $x \in \mathbb{R}^n$ and $\nu, \eta > 0$ be arbitrary. Then, the problem (5.24) has a unique optimal solution $\tilde{\lambda}(x; \nu, \eta)$ and it holds that $\tilde{\lambda}(x; \nu, \eta) > (E(-H(x); \eta))_+$. Moreover, the optimal value of (5.22) equals the negative of the one of (5.24), the hard case does not occur for (5.22), and*

$$(5.26) \quad \tilde{v}(x; \nu, \eta) = -(1/2)\tilde{g}_\nu(x)^T (\widetilde{H}_\eta(x) + \tilde{\lambda}(x; \nu, \eta)I)^{-1} \tilde{g}_\nu(x) - (1/2)\Delta^2 \tilde{\lambda}(x; \nu, \eta).$$

Proof. Lemma 5.5 implies that (5.24) has a unique optimal solution $\tilde{\lambda}(x; \nu, \eta)$ and it holds that $\tilde{\lambda}(x; \nu, \eta) > (E(-H(x); \eta))_+$. Using (5.12), we infer that $\lambda_{\min}(H(x)) \ge -E(-H(x); \eta)$ and (5.23) shows $\lambda_{\min}(\widetilde{H}_\eta(x)) = -(E(-H(x); \eta))_+$.

If $E(-H(x); \eta) > 0$, we have that $y = (0, \ldots, 0, 1) \in N(\widetilde{H}_\eta(x) - \lambda_{\min}(\widetilde{H}_\eta(x))I)$ and $y^T \tilde{g}_\nu(x) \ne 0$. If $E(-H(x); \eta) \le 0$, we get that $w = (0, \ldots, 0, 1, 0) \in N(\widetilde{H}_\eta(x) - \lambda_{\min}(\widetilde{H}_\eta(x))I)$ and $w^T \tilde{g}_\nu(x) \ne 0$. Hence, we obtain $\tilde{g}_\nu(x) \not\perp N(\widetilde{H}_\eta(x) - \lambda_{\min}(\widetilde{H}_\eta(x))I)$.

Next, for all $\lambda > (E(-H(x); \eta))_+$, we infer from (5.23) and (5.25) that

$$d(\lambda; x) + \frac{\nu}{\tilde{\lambda}(x; \nu, \eta)} + \frac{\nu}{\tilde{\lambda}(x; \nu, \eta) - E(-H(x); \eta)} = \frac{1}{2}\tilde{g}_\nu(x)^T (\widetilde{H}_\eta(x) + \lambda I)^{-1} \tilde{g}_\nu(x) + \frac{1}{2}\Delta^2 \lambda.$$

Hence, Theorem 5.3 shows that strong duality holds and (5.26) is satisfied. The hard case does not occur for (5.22) since $\tilde{\lambda}(x; \nu, \eta) > (E(-H(x); \eta))_+ = -\lambda_{\min}(\widetilde{H}_\eta(x))$. $\square$

We establish an error estimate on $\tilde{v}$ (see (5.22)) and show that it is a smoothing function of v (see (5.1)). We define, similar to (5.6), the mapping $s : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ by

$$(5.27) \quad s(\lambda; x) = -(H(x) + \lambda I)^+ g(x).$$

For $\nu, \eta > 0$, we denote by $(\tilde{s}(x; \nu, \eta), \tilde{\lambda}(x; \nu, \eta))$ an optimal primal-dual solution of (5.22), where

$$(5.28) \quad \tilde{\lambda}(\cdot; \nu, \eta) : \mathbb{R}^n \to \mathbb{R} \quad \text{and} \quad \tilde{s}(\cdot; \nu, \eta) : \mathbb{R}^n \to \mathbb{R}^p.$$

From (5.3), Lemma 5.8, the block structure of $\widetilde{H}_\eta(x)$ (see (5.23)) and (5.27), we infer that for all $x \in \mathbb{R}^n$ it holds that

$$(5.29) \qquad \widetilde{s}(x; \nu, \eta) = (s(\tilde{\lambda}(x; \nu, \eta); x), \widetilde{s}_{p+1}(x; \nu, \eta), \widetilde{s}_{p+2}(x; \nu, \eta)).$$

In particular, the first p components of $\widetilde{s}(x; \nu, \eta)$ are given by $s(\tilde{\lambda}(x; \nu, \eta); x)$. By applying (5.3) and (5.23), we obtain that

$$(5.30) \quad \widetilde{s}_{p+1}(x; \nu, \eta) = \frac{\sqrt{2\nu}}{\tilde{\lambda}(x; \nu, \eta)}, \quad \text{and} \quad \widetilde{s}_{p+2}(x; \nu, \eta) = \frac{\sqrt{2\nu}}{\tilde{\lambda}(x; \nu, \eta) - E(-H(x); \eta)}.$$

THEOREM 5.9. *Let ν, $\eta > 0$ be arbitrary, and let the mappings $g : \mathbb{R}^n \to \mathbb{R}^p$ and $H : \mathbb{R}^n \to \mathbb{S}^p$ be q-times continuously differentiable, where $q \geq 1$. Then, the following conditions hold true.*
1. *For every $x \in \mathbb{R}^n$, we have that*

$$(5.31) \qquad v(x) \geq \tilde{v}(x; \nu, \eta) \geq v(x) - 2\sqrt{2\nu}\Delta - (1/2)\Delta^2 \eta \ln p,$$

  where v is defined in (5.1) and $\tilde{v}$ in (5.22).
2. *The mappings $\widetilde{s}(\cdot; \nu, \eta)$ and $\tilde{\lambda}(\cdot; \nu, \eta)$ defined in (5.28) are $q-1$-times continuously differentiable, and $\tilde{v}(\cdot; \nu, \eta)$ is q-times continuously differentiable. We have that*

$$(5.32) \quad \nabla_x \tilde{v}(x; \nu, \eta) = \nabla_x \wp(x, s)|_{s=s(\tilde{\lambda}; x)} + (1/2)(\widetilde{s}_{p+2})^2 \nabla_x(-E(-H(x); \eta)),$$

  where $\wp : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$ is defined by

$$(5.33) \qquad\qquad \wp(x, s) = g(x)^T s + (1/2)s^T H(x)s$$

  and $(\widetilde{s}, \tilde{\lambda}) = (\widetilde{s}(x; \nu, \eta), \tilde{\lambda}(x; \nu, \eta))$ is the optimal primal-dual solution of (5.22).
3. *The function $\tilde{v}$ is a smoothing function of v.*

Proof. 1. Let $x \in \mathbb{R}^n$ be arbitrary. Theorem 5.6 and Lemma 5.8 yield with (5.17) and (5.26) that (5.31) holds.

2. Lemma 5.8 further shows that $\tilde{\lambda}(x; \nu, \eta) > (E(-H(x); \eta))_+$ implying that strict complementarity slackness holds for (5.22). Moreover, the function $E(\cdot; \eta)$ (see (5.10)) is analytic as $z \mapsto \eta \ln \sum_{i=1}^p \exp(z_i/\eta)$ is analytic (see [53, Thm. 3.1]) and, therefore, the mapping $\widetilde{H}_\eta$ (see (5.23)) is q-times continuously differentiable. Hence, the implicit function theorem applies to the first-order optimality conditions (5.3) of (5.22) and implies that $\tilde{\lambda}(\cdot; \nu, \eta)$ and $\widetilde{s}(\cdot; \nu, \eta)$ are $q-1$-times continuously differentiable.

Now, the equations (5.22), (5.23), (5.29), (5.33) together with Danskin's theorem [10, Thm. 4.13, Rem. 4.14] yield that $\tilde{v}(\cdot; \nu, \eta)$ is differentiable and show that its gradient is given by (5.32). Next, [30, Cor. 8.2] implies that $\widetilde{s}(\cdot; \nu, \eta)$ is continuous showing that $\nabla_x \tilde{v}(\cdot; \nu, \eta)$ is continuous. Moreover, the chain rule and (5.22) imply that $\tilde{v}(\cdot; \nu, \eta)$ is q-times continuously differentiable.

3. The function v is continuous by [30, Thm. 7], $\tilde{v}(\cdot; \nu, \eta)$ is continuously differentiable and, hence, (5.31) shows that $\tilde{v}$ is a smoothing function of v. □

The next result asserts gradient consistency of the function $\tilde{v}$ defined in (5.22).

THEOREM 5.10. *Let the conditions of Theorem 5.9 be fulfilled. Then, the following conditions are satisfied.*
1. *Gradient consistency holds for $\tilde{v}$ and v, where v is defined in (5.1) and $\tilde{v}$ in (5.22).*
2. *Let $x \in \mathbb{R}^n$ be given, $(x^k) \subset \mathbb{R}^n$ and (ν_k), $(\eta_k) \subset \mathbb{R}_{++}$ be sequences converging to x and 0 as $k \to \infty$, respectively. Then, there exists a convergent subsequence $(\nabla_x \tilde{v}(x^k; \nu_k, \eta_k))_K$ of $(\nabla_x \tilde{v}(x^k; \nu_k, \eta_k))$.*

We need the following result to prove Theorem 5.10.

LEMMA 5.11. *Let $(\eta_k) \subset \mathbb{R}_{++}$ be a sequence such that $\eta_k \to 0$ as $k \to \infty$. Furthermore, let $A : \mathbb{R}^n \to \mathbb{S}^p$ be continuously differentiable and let $(x^k) \subset \mathbb{R}^n$ be a sequence such that $x^k \to x \in \mathbb{R}^n$ as $k \to \infty$. Then, there exists a subsequence $(\nabla_x(E(\cdot; \eta_k) \circ A)(x^k))_K$ of $(\nabla_x(E(\cdot; \eta_k) \circ A)(x^k))$ such that*

$$\nabla_x(E(\cdot; \eta_k) \circ A)(x^k) \to \sum_{i=1}^{r} \theta_i DA(x)^*[u_i u_i^T] \in DA(x)^* \partial \lambda_{\max}(A(x)) \text{ as } K \ni k \to \infty,$$

where E is defined in (5.10), $1 \leq r \leq r(A(x))$, $r(A(x))$ denotes the multiplicity of $\lambda_{\max}(A(x))$, $\theta_i \in [0,1]$, $\sum_{i=1}^{r} \theta_i = 1$, and $u_i \in \mathbb{R}^p$, $\|u_i\|_2 = 1$, are pairwise orthogonal eigenvectors of $A(x)$ corresponding to $\lambda_{\max}(A(x))$.

Proof. The mapping A is continuously differentiable and $\lambda_{\max}$ is convex and, hence, regular in the sense of [19, Def. 2.3.4]; see [19, Prop. 2.3.6]. Moreover, A and $\lambda_{\max}$ are locally Lipschitz continuous. The chain rule [19, Thm. 2.3.9] implies that

$$(5.34) \qquad \partial(\lambda_{\max} \circ A)(x) = DA(x)^* \partial \lambda_{\max}(A(x)).$$

The function $E(\cdot; \eta_k)$ is analytic (see [53, Thm. 3.1]) and, hence, the chain rule implies

$$(5.35) \qquad \nabla_x(E(\cdot; \eta_k) \circ A)(x^k) = DA(x^k)^* \nabla_A E(A(x^k); \eta_k).$$

We define $A_k = A(x^k)$ and $A = A(x)$. Next, we show that there exists a subsequence $(\nabla_A E(A_k; \eta_k))_K$ of $(\nabla_A E(A_k; \eta_k))$ such that

$$(5.36) \qquad \nabla_A E(A_k; \eta_k) \to \sum_{i=1}^{r} \theta_i u_i u_i^T \in \partial \lambda_{\max}(A) \quad \text{as} \quad K \ni k \to \infty.$$

For all $k \geq 0$, we have that

$$\nabla_A E(A_k; \eta_k) = \sum_{i=1}^{p} \theta_{i,k} u_i(A_k) u_i(A_k)^T, \quad \text{and} \quad \theta_{i,k} = \frac{\exp \frac{\lambda_i(A_k) - \lambda_{\max}(A_k)}{\eta_k}}{\sum_{i=1}^{p} \exp \frac{\lambda_i(A_k) - \lambda_{\max}(A_k)}{\eta_k}},$$

where $A_k u_i(A_k) = \lambda_{\max}(A_k) u_i(A_k)$, $\|u_i(A_k)\|_2 = 1$, and the vectors $u_i(A_k)$ are pairwise orthogonal for $i = 1, \ldots p$; cf. [41, sect. 4]. We have that $\sum_{i=1}^{p} \theta_{i,k} = 1$ and $\theta_{i,k} \in [0,1]$. Hence, we can assume w.l.o.g. that for all $i \in \{1, \ldots, p\}$, it holds that $u_i(A_k) \to u_i \in \mathbb{R}^p$, $\theta_{i,k} \to \theta_i \in [0,1]$ as $k \to \infty$, $\|u_i\|_2 = 1$, and $\sum_{i=1}^{p} \theta_i = 1$. We have $A_k u_i(A_k) = \lambda_i(A_k) u_i(A_k)$ for all $k \geq 0$, $A_k \to A$ as $k \to \infty$ and λ is continuous (cf. [31, Cor. 6.3.8]) showing that u_i is an eigenvector of A corresponding to $\lambda_i(A)$. Moreover, $0 = u_i(A_k)^T u_j(A_k) \to u_i^T u_j$ as $k \to \infty$ for all $i \neq j$ implies that u_i are pairwise orthogonal.

Now, let $i \in \{1, \ldots, p\}$ be an index such that $\lambda_i(A) < \lambda_{\max}(A)$, i.e., $i > r(A)$. We obtain that $\lambda_i(A_k) - \lambda_{\max}(A_k) \leq (\lambda_i(A) - \lambda_{\max}(A))/2 < 0$ for all $k \geq 0$ sufficiently large. Hence, we infer $\theta_{i,k} \to 0$ as $k \to \infty$ resulting in $\theta_i = 0$. Moreover, it holds that

$$\text{conv} \left\{ uu^T : Au = \lambda_{\max}(A)u, \ \|u\|_2 = 1, \ u \in \mathbb{R}^p \right\} = \partial \lambda_{\max}(A)$$

(cf. [41, sect. 4]) and, hence, we conclude that (5.36) holds. We have that $DA(x^k) \to DA(x)$ as $k \to \infty$ and, therefore, (5.34) and (5.35) imply the assertion. $\square$

We use the notation $(\nu_k, \eta_k)_{\mathbb{N}_0}$ to indicate a sequence distinguishing it from its elements (ν_k, η_k) and to avoid using $((\nu_k, \eta_k))$, and $(\nu_k, \eta_k)_K$ to denote a subsequence of $(\nu_k, \eta_k)_{\mathbb{N}_0}$. In addition to Lemma 5.11, we apply the next result to prove Theorem 5.9.

LEMMA 5.12. *Let the conditions of Theorem 5.9 be fulfilled. Moreover, let $\bar{x} \in \mathbb{R}^n$ be given, $(x^k) \subset \mathbb{R}^n$ and (ν_k), $(\eta_k) \subset \mathbb{R}_{++}$ be sequences converging to $\bar{x}$ and 0 as $k \to \infty$, respectively. We denote $(\tilde{s}^k, \tilde{\lambda}_k) = (\tilde{s}(x^k; \nu_k, \eta_k), \tilde{\lambda}(x^k; \nu_k, \eta_k))$, where $(\tilde{s}(x; \nu, \eta), \tilde{\lambda}(x; \nu, \eta))$ is defined in (5.28). Then, the following conditions hold true.*
1. *The sequence $(\tilde{s}^k, \tilde{\lambda}_k)_{\mathbb{N}_0}$ has a convergent subsequence $(\tilde{s}^k, \tilde{\lambda}_k)_K$. In particular, there exist $(\bar{s}, \bar{\lambda}) \in \mathbb{R}^p \times \mathbb{R}_+$ and $\bar{\alpha}, \bar{\beta} \in \mathbb{R}$ such that*

$$(5.37) \quad \tilde{s}^k = (s(\tilde{\lambda}_k; x^k), \tilde{s}^k_{p+1}, \tilde{s}^k_{p+2}) \to (\bar{s}, \bar{\beta}, \bar{\alpha}) \quad and \quad \tilde{\lambda}_k \to \bar{\lambda} \quad as \quad K \ni k \to \infty.$$

2. *If $\bar{\lambda} > -\lambda_{\min}(H(\bar{x}))$ holds, the easy case occurs for (5.1) with $x = \bar{x}$, $(\bar{s}, \bar{\lambda})$ is an optimal primal-dual solution of (5.1) for $x = \bar{x}$, and $\bar{\alpha} = 0$.*
3. *If $\bar{\lambda} = -\lambda_{\min}(H(\bar{x}))$ holds, the hard case occurs for (5.1) with $x = \bar{x}$. Moreover, let $w_i \in \mathbb{R}^p$, $\|w_i\|_2 = 1$, $i = 1, \ldots, r$, be pairwise orthogonal eigenvectors of $H(\bar{x})$ corresponding to $\lambda_{\min}(H(\bar{x}))$, where $r \in \mathbb{N}$. Then, the vectors $(\bar{s} + \gamma_i^+ w_i, \bar{\lambda})$ and $(\bar{s} + \gamma_i^- w_i, \bar{\lambda})$ are optimal primal-dual solutions of (5.1) for $x = \bar{x}$, where*

$$(5.38) \quad \gamma_i^+ = -w_i^T \bar{s} + \sqrt{(w_i^T \bar{s})^2 + \bar{\alpha}^2}, \quad and \quad \gamma_i^- = -w_i^T \bar{s} - \sqrt{(w_i^T \bar{s})^2 + \bar{\alpha}^2}.$$

Proof. 1. Let $k \geq 0$ be arbitrary. We show that $(\tilde{s}^k, \tilde{\lambda}_k)_{\mathbb{N}_0}$ is bounded. Since $\|\tilde{s}^k\|_2 \leq \Delta$ holds, $(\tilde{s}^k)$ is bounded. Lemma 5.8 shows that $\tilde{\lambda}_k = \tilde{\lambda}(x^k; \nu_k, \eta_k) > (E(-H(x^k); \eta))_+$ and, hence, (5.12) implies

$$(5.39) \qquad\qquad\qquad \tilde{\lambda}_k > -(\lambda_{\min}(H(x^k)))_+.$$

Now, (5.26), Lemma 5.8 and (5.39) yield that

$$\tilde{v}(x^k; \nu_k, \eta_k) = -\frac{1}{2}\tilde{g}_{\nu_k}^T(x^k)(\widetilde{H}_{\eta_k}(x^k) + \tilde{\lambda}_k I)^{-1}\tilde{g}_{\nu_k}(x^k) - \frac{1}{2}\Delta^2 \tilde{\lambda}_k \leq -\frac{1}{2}\Delta^2 \tilde{\lambda}_k \leq 0,$$

The left-hand side of the above inequality converges to $v(\bar{x})$ as $k \to \infty$ by Theorem 5.9 and $\Delta > 0$ holds implying that $(\tilde{\lambda}_k)$ is bounded. In particular, $(\tilde{s}^k, \tilde{\lambda}_k)_{\mathbb{N}_0}$ is bounded and it has a convergent subsequence $(\tilde{s}^k, \tilde{\lambda}_k)_K$. Hence, (5.29) implies that (5.37) holds for some $(\bar{s}, \bar{\lambda}) \in \mathbb{R}^p \times \mathbb{R}_+$ and $\bar{\alpha}, \bar{\beta} \in \mathbb{R}$.

Next, (5.14) shows that a necessary optimality condition of (5.22) is

$$\Delta^2 = \|s(\tilde{\lambda}_k; x^k)\|_2^2 + \frac{2\nu_k}{\tilde{\lambda}_k^2} + \frac{2\nu_k}{(\tilde{\lambda}_k - E(-H(x^k); \eta_k))^2} = \|\tilde{s}^k\|_2^2,$$

where we have used (5.30) and (5.30) to establish the second equality. Hence, by applying (5.37) we obtain that

$$(5.40) \qquad\qquad \Delta^2 = \|\tilde{s}^k\|_2^2 \to \|\bar{s}\|_2^2 + \bar{\beta}^2 + \bar{\alpha}^2 \quad as \quad K \ni k \to \infty.$$

Moreover, from (5.39), we infer that

$$(5.41) \qquad\qquad\qquad H(\bar{x}) + \bar{\lambda} I \succcurlyeq 0, \quad and \quad \bar{\lambda} \geq 0.$$

Using (5.27) and (5.39), we have that

$$(5.42) \quad 0 = (H(x^k) + \tilde{\lambda}_k I)s(\tilde{\lambda}_k; x^k) + g(x^k) \to (H(\bar{x}) + \bar{\lambda} I)\bar{s} + g(\bar{x}) \text{ as } K \ni k \to \infty.$$

2. Now, we verify that $(\bar{s}, \bar{\lambda})$ is an optimal primal-dual solution of (5.1) for $x = \bar{x}$ and $\bar{\alpha} = 0$ if $\bar{\lambda} > -\lambda_{\min}(H(\bar{x}))$. By assumption $H(\bar{x}) + \bar{\lambda}I$ is invertible and, hence, (5.42) implies that $\bar{s}$ is the unique solution to $(H(\bar{x}) + \bar{\lambda}I)\bar{s} = -g(\bar{x})$. Therefore, (5.27) and (5.42) result in $s(\bar{\lambda}; \bar{x}) = \bar{s}$. Moreover, (5.40) implies that $\|\bar{s}\|_2 \leq \Delta$.

If $\bar{\lambda} > 0$, then continuity of $\lambda_{\min}$, $\bar{\lambda} > -\lambda_{\min}(H(\bar{x}))$, $\tilde{\lambda}_k \to \bar{\lambda}$ as $K \ni k \to \infty$ and (5.11) imply that $\tilde{\lambda}_k \geq \bar{\lambda}/2 > 0$ and $\tilde{\lambda}_k - E(-H(x^k); \eta_k) \geq (\bar{\lambda} + \lambda_{\min}(H(\bar{x})))/2 > 0$ for all $k \in K$ sufficiently large. Therefore, we obtain from (5.30) that

$$(5.43) \quad \tilde{s}_{p+1}^k = \frac{\sqrt{2\nu_k}}{\tilde{\lambda}_k} \to 0, \quad \text{and} \quad \tilde{s}_{p+2}^k = \frac{\sqrt{2\nu_k}}{\tilde{\lambda}_k - E(-H(x^k); \eta_k)} \to 0 \quad \text{as} \quad K \ni k \to \infty,$$

and, therefore, $\bar{\alpha}, \bar{\beta} = 0$. Now, (5.40) implies that $\Delta^2 = \|\bar{s}\|_2^2$.

Hence, $(s(\bar{\lambda}; \bar{x}), \bar{\lambda})$ satisfies $\bar{\lambda}(\|\bar{s}\|_2^2 - \Delta^2) = 0$ and, therefore, it fulfills (5.3) implying that it is an optimal primal-dual solution of (5.1) for $x = \bar{x}$ by Theorem 5.1. Theorem 5.1 further implies that the easy case occurs.

3. Next, we establish that the vectors $(\bar{s} + \gamma_i^+ w_i, \bar{\lambda})$ and $(\bar{s} + \gamma_i^- w_i, \bar{\lambda})$ are optimal primal-dual solutions of (5.1) for $x = \bar{x}$ if $\bar{\lambda} = -\lambda_{\min}(H(\bar{x}))$. Let $i \in \{1, \dots, r\}$ be arbitrary. The numbers γ_i^+ and γ_i^- solve

$$\gamma_i^2 + 2\gamma_i w_i^T \bar{s} - \bar{\alpha}^2 = 0.$$

Using $\|w_i\|_2 = 1$ and (5.40), we obtain for $\gamma_i \in \{\gamma_i^-, \gamma_i^+\}$ that

$$(5.44) \quad \|\bar{s} + \gamma_i w_i\|_2^2 = \|\bar{s}\|_2^2 + 2\gamma_i w_i^T \bar{s} + \gamma_i^2 = \Delta^2 - \bar{\alpha}^2 - \bar{\beta}^2 + 2\gamma_i w_i^T \bar{s} + \gamma_i^2 \leq \Delta^2$$

with equality if $\bar{\beta} = 0$ and, moreover, (5.42) and $(H(\bar{x}) + \bar{\lambda}I)w_i = 0$ results in

$$(5.45) \quad (H(\bar{x}) + \bar{\lambda}I)(\bar{s} + \gamma_i w_i) = (H(\bar{x}) + \bar{\lambda}I)\bar{s} = -g(\bar{x}).$$

If $\bar{\lambda} > 0$, (5.43) shows that $\bar{\beta} = 0$. Hence, (5.44) implies that $\bar{\lambda}(\|\bar{s} + \gamma_i w_i\|_2^2 - \Delta^2) = 0$.

Moreover, (5.41), (5.42), (5.44) and (5.45), and the above complementarity condition yield that $(\bar{s} + \gamma_i w_i, \bar{\lambda})$, $\gamma_i \in \{\gamma_i^-, \gamma_i^+\}$, fulfill (5.3) and, hence, are optimal primal-dual solutions of (5.1) for $x = \bar{x}$ by Theorem 5.1. Theorem 5.1 further implies that the hard case occurs. □

The proof of Theorem 5.9 requires the gradient of $\wp$ (see (5.33)), which is given by

$$(5.46) \quad \nabla_x \wp(x, s) = \nabla_x g(x)^T s + (1/2)\nabla_x s^T H(x)s = \nabla_x g(x)^T s + (1/2)DH(x)^*[ss^T]$$

Indeed, the first equality in (5.46) follows from the chain rule and the second using a similar derivation as in (4.7).

Proof of Theorem 5.10. 1. Let $\bar{x} \in \mathbb{R}^n$ be arbitrary. The function v is locally Lipschitz continuous (cf. [24, Thm. 4.1]), and, hence, $\partial v(\bar{x})$ is well-defined. From (5.1), (5.33) and [18, Thm. 2.1], we have that

$$(5.47) \quad \partial v(\bar{x}) = \text{conv} \left\{ \nabla_x \wp(\bar{x}, s^*) : s^* \in \mathcal{S}_{\text{TR}}^*(\bar{x}) \right\},$$

where $\mathcal{S}_{\text{TR}}^*(\bar{x})$ denotes the set of optimal solutions of (5.1) for $x = \bar{x}$.

Next, we establish that gradient consistency holds, i.e., that (3.5) holds distinguishing if the easy or the hard case occurs for (5.1) with $x = \bar{x}$. The inclusion $\partial v(\bar{x}) \subset S_{\tilde{v}}(\bar{x})$ follows from v being locally Lipschitz continuous, where $S_{\tilde{v}}(x)$ is defined in (3.4); cf. Lemma 3.2. Let $z \in \mathbb{R}^p$ be such that there exist sequences $(x^k) \subset \mathbb{R}^n$ and $(\nu_k), (\eta_k) \subset \mathbb{R}_{++}$ fulfilling $x^k \to \bar{x}$ and $\nu_k, \eta_k \to 0$, and

$$(5.48) \quad \nabla_x \tilde{v}(x^k; \nu_k, \eta_k) \to z \quad \text{as} \quad k \to \infty.$$

Lemma 5.12 implies that the sequence $(\tilde{s}^k, \tilde{\lambda}_k)_{\mathbb{N}_0}$ of optimal primal-dual solutions $(\tilde{s}^k, \tilde{\lambda}_k)$ of (5.22) for $(x, \nu, \eta) = (x^k, \nu_k, \eta_k)$ has a convergent subsequence $(\tilde{s}^k, \tilde{\lambda}_k)_K$. Moreover, the sequence $(s(\tilde{\lambda}_k; x^k), \tilde{\lambda}_k)_K$ converges to $(\bar{s}, \bar{\lambda})$ and $\tilde{s}_{p+2} \to \bar{\alpha}$ as $K \ni k \to \infty$, where $\bar{s} \in \mathbb{R}^p$, $\bar{\lambda} \geq 0$ and $\bar{\alpha} \in \mathbb{R}$, and $s(\lambda; x)$ is defined in (5.27).

In addition, Lemma 5.11 applies with $A = -H$ and shows that there exists a subsequence $(\nabla_x(E(-H(x^k); \eta_k)))_{K'}$ of $(\nabla_x(E(-H(x^k); \eta_k)))_K$ such that

$$(5.49) \qquad \nabla_x(E(-H(x^k); \eta_k)) \to -\sum_{i=1}^{r} \theta_i DH(\bar{x})^*[w_i w_i^T] \quad \text{as} \quad K' \ni k \to \infty,$$

where $1 \leq r \leq r(A(\bar{x}))$, $r(A(\bar{x}))$ denotes the multiplicity of $\lambda_{\max}(A(\bar{x}))$, $\theta_i \in [0, 1]$, and $\sum_{i=1}^{r} \theta_i = 1$. Moreover, $\|w_i\|_2 = 1$ and w_i are pairwise orthogonal eigenvectors of $A(\bar{x}) = -H(\bar{x})$ corresponding to $\lambda_{\max}(A(\bar{x})) = -\lambda_{\min}(H(\bar{x}))$. We define $r = r(A(\bar{x}))$.

Hence, (5.32), (5.49), and g and H being continuously differentiable show that

$$(5.50) \quad \nabla_x \tilde{v}(x^k; \nu_k, \eta_k) \to \nabla_x \wp(\bar{x}, \bar{s}) + (\bar{\alpha}^2/2) \sum_{i=1}^{r} \theta_i DH(\bar{x})^*[w_i w_i^T] \text{ as } K' \ni k \to \infty.$$

If the easy case occurs for (5.1) with $x = \bar{x}$, Lemma 5.12 further implies that $\bar{s} \in \mathcal{S}_{\mathrm{TR}}^*(\bar{x})$ and $\bar{\alpha} = 0$. By applying (5.47), (5.48) and (5.50), we infer that $z \in \partial v(\bar{x})$.

If the hard case occurs for (5.1), Lemma 5.12 further implies that $\bar{s} + \gamma_i^+ w_i$ and $\bar{s} + \gamma_i^- w_i$ are optimal solutions of (5.1) for $x = \bar{x}$, where γ_i^+ and γ_i^- are defined in (5.38). If $\bar{\alpha} = 0$, (5.38) implies that either γ_i^+ or γ_i^- is zero and, hence, $\bar{s}$ is an optimal solution of (5.1) for $x = \bar{x}$, and, hence, (5.47), (5.48) and (5.50) imply that $z \in \partial v(\bar{x})$. If $\bar{\alpha} > 0$, (5.38) results in $\gamma_i^+ - \gamma_i^- = 2\sqrt{(w_i^T \bar{s})^2 + \bar{\alpha}^2} > 0$. We define

$$(5.51) \qquad\qquad \tau_i^+ = \frac{-\gamma_i^-}{\gamma_i^+ - \gamma_i^-}, \quad \text{and} \quad \tau_i^- = \frac{\gamma_i^+}{\gamma_i^+ - \gamma_i^-}.$$

Furthermore, (5.38) implies that $\gamma_i^+ > 0$ and $\gamma_i^- < 0$ and, hence, (5.51) shows that

$$(5.52) \quad \begin{aligned} &\tau_i^+ > 0, \quad \tau_i^- > 0, \quad \tau_i^+ + \tau_i^- = 1, \\ &\tau_i^+ \gamma_i^+ + \tau_i^- \gamma_i^- = \frac{-\gamma_i^- \gamma_i^+ + \gamma_i^+ \gamma_i^-}{\gamma_i^+ - \gamma_i^-} = 0, \quad \text{and} \quad \tau_i^+ (\gamma_i^+)^2 + \tau_i^- (\gamma_i^-)^2 = \bar{\alpha}^2. \end{aligned}$$

Using (5.33) and (5.46), we obtain for $\gamma_i \in \{\gamma_i^-, \gamma_i^+\}$ that

$$\begin{aligned} \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i w_i) &= \nabla_x g(\bar{x})^T \bar{s} + (1/2)DH(\bar{x})^*[\bar{s}\bar{s}^T] + \gamma_i \nabla_x g(\bar{x})^T w_i \\ &\quad + (1/2)\gamma_i DH(\bar{x})^*[w_i \bar{s}^T + \bar{s} w_i^T] + (1/2)(\gamma_i)^2 DH(\bar{x})^*[w_i w_i^T] \end{aligned}$$

resulting in

$$\begin{aligned} &\tau_i^+ \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^+ w_i) + \tau_i^- \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^- w_i) \\ &\quad = (\tau_i^- + \tau_i^+) \nabla_x g(\bar{x})^T \bar{s} + (1/2)(\tau_i^- + \tau_i^+)DH(\bar{x})^*[\bar{s}\bar{s}^T] \\ &\qquad + (\tau_i^+ \gamma_i^+ + \tau_i^- \gamma_i^-) \nabla g(\bar{x})^T w_i + (1/2)(\tau_i^+ \gamma_i^+ + \tau_i^- \gamma_i^-)DH(\bar{x})^*[w_i \bar{s}^T + \bar{s} w_i^T] \\ &\qquad + (1/2)(\tau_i^+ (\gamma_i^+)^2 + \tau_i^- (\gamma_i^-)^2)DH(\bar{x})^*[w_i w_i^T]. \end{aligned}$$

Hence, (5.52) implies that

$$\begin{aligned} &\tau_i^+ \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^+ w_i) + \tau_i^- \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^- w_i) \\ &\quad = \nabla_x g(\bar{x})^T \bar{s} + (1/2)DH(\bar{x})^*[\bar{s}\bar{s}^T] + (\bar{\alpha}^2/2)DH(\bar{x})^*[w_i w_i^T], \end{aligned}$$

implying with $\sum_{i=1}^{r} \theta_i = 1$ and (5.46) that

$$
\begin{aligned}
\sum_{i=1}^{r} \theta_i \tau_i^+ \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^+ w_i) &+ \sum_{i=1}^{r} \theta_i \tau_i^- \nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^- w_i) \\
&= \nabla_x \wp(\bar{x}, \bar{s}) + (\bar{\alpha}^2/2) \sum_{i=1}^{r} \theta_i DH(\bar{x})^*[w_i w_i^T].
\end{aligned}
$$
(5.53)

Moreover, using (5.52), we have that $\sum_{i=1}^{r} \theta_i \tau_i^+ + \sum_{i=1}^{r} \theta_i \tau_i^- = \sum_{i=1}^{r} \theta_i(\tau_i^+ + \tau_i^-) = 1$ the limit in (5.50) equals (5.53). Now, we use that $\nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^+ w_i)$ and $\nabla_x \wp(\bar{x}, \bar{s} + \gamma_i^- w_i)$ are contained in $\partial v(\bar{x})$ (cf. Lemma 5.12) implying that (5.53) is a convex combination of elements of $\partial v(\bar{x})$. Hence, (5.47), (5.48) and (5.50) yield $z \in \partial v(\bar{x})$.

2. Adapting the above reasoning and using (5.32) we obtain that $(\nabla_x \tilde{v}(x^k; \nu_k, \eta_k))$ has a converging subsequence if $x^k \to x$ and $\nu_k, \eta_k \to 0^+$ as $k \to \infty$. □

Theorem 5.9 and Theorem 5.10 imply that the function $\tilde{\psi}_j : \mathbb{R}^n \times \mathbb{R}_{>0}^2 \to \mathbb{R}$ given by

$$
(5.54) \quad \tilde{\psi}_j(x; \nu, \eta) = h_j(x) - \min_{\widetilde{s} \in \mathbb{R}^{p+2}} \left\{ (1/2)\widetilde{s}^T \widetilde{H}_{\eta,j}(x)\widetilde{s} + \widetilde{g}_{\nu,j}(x)^T \widetilde{s} : \ \|\widetilde{s}\|_2 \le \Delta \right\},
$$

is a smoothing function of ψ_j (see (1.6)), where $h_j(x) = a_j(x)$, $g_j(x) = -\bar{\Sigma}^{1/2} b_j(x)$, and $H_j(x) = -\bar{\Sigma}^{1/2} C_j(x) \bar{\Sigma}^{1/2}$. Moreover, $\widetilde{H}_{\eta,j}$ and $\widetilde{g}_{\nu,j}$ are defined as in (5.23) with H and g replaced by H_j and g_j, respectively. The representation of $\tilde{\psi}_j$ results from (1.6) being transformed to the TRP (5.1) using $d \mapsto s = \bar{\Sigma}^{-1/2} d$.

THEOREM 5.13. *Let $\bar{\Sigma} \in \mathbb{S}_{++}^p$, and $a_j : \mathbb{R}^n \to \mathbb{R}$, $b_j : \mathbb{R}^n \to \mathbb{R}^p$ and $C_j : \mathbb{R}^n \to \mathbb{R}^p$ be q-times continuously differentiable, where $q \ge 1$ and $j \in J$. Then, the following conditions hold true.*
1. *The function $\tilde{\psi}_j$ defined in (5.54) is a smoothing function of ψ_j, $\tilde{\psi}_j(\cdot; \nu, \eta)$ is q-times continuously differentiable for every ν, $\eta > 0$, and gradient consistency holds.*
2. *Let $x \in \mathbb{R}^n$ be given and $(x^k) \subset \mathbb{R}^n$ and (ν_k), $(\eta_k) \subset \mathbb{R}_{++}$ be sequences converging to x and 0 as $k \to \infty$, respectively. Then, there exists a convergent subsequence $(\nabla_x \tilde{\psi}_j(x^k; \nu_k, \eta_k))_K$ of $(\nabla_x \tilde{\psi}(x^k; \nu_k, \eta_k))$.*

The computational cost of evaluating (5.54) are essentially the same as the evaluation of (1.6) since $\widetilde{H}_{\eta,j}(x)$ (see (5.23)) is a block-diagonal matrix for $x \in \mathbb{R}^n$ implying that our smoothing approach is tractable both theoretically and practically.

6. Convergence of the homotopy method. We show that a sequence of KKT-tuples of (3.2) generated by Algorithm 3.1 converges to a stationary point of the DROP (3.1) under mild assumptions. We define $\widetilde{F}_j : \mathbb{R}^n \times \mathbb{R}_{++}^3 \to \mathbb{R}$ by

$$
(6.1) \qquad \widetilde{F}_j(x; t) = \tilde{\varphi}_j(x; \tau) + \tilde{\psi}_j(x; \nu, \eta)
$$

and recall that $F_j : \mathbb{R}^n \to \mathbb{R}$, $F_j(x) = \varphi_j(x) + \psi_j(x)$ for all $j \in J$, where we set $t = (\tau, \nu, \eta)$, and $\tilde{\varphi}_j$ and $\tilde{\psi}_j$ is defined in (4.3) and (5.54), respectively. Suitable assumptions on (1.4) imply that the DROP (3.2) has feasible points.

PROPOSITION 6.1. *Let $z \in \mathbb{R}^n$ be a strictly feasible point for (3.1) and let the conditions of Theorem 4.2 and Theorem 5.13 be fulfilled for any $j \in J \setminus \{0\}$. Then, z is a strictly feasible point to (3.2) for all sufficiently small $t > 0$.*

Proof. Theorem 4.2, Theorem 5.13, and (6.1) imply that

$$
\widetilde{F}_j(z; t) = \tilde{\varphi}_j(z; \tau) + \tilde{\psi}_j(z; \nu, \eta) \to F_j(z) \quad \text{as} \quad t = (\tau, \nu, \eta) \to 0^+
$$

for all $j \in J \setminus \{0\}$ establishing the assertion. □

Next, we provide a global convergence result of Algorithm 3.1.

THEOREM 6.2. *Let the conditions of Theorem 4.2 and Theorem 5.13 hold for all $j \in J$. Choose $\varepsilon_{\min}$, $t_{\min} = 0$ and let the sequence $(x^k, \vartheta^k)_{\mathbb{N}_0}$ be generated by Algorithm 3.1. Then, every accumulation point of $(x^k, \vartheta^k)_{\mathbb{N}_0}$ is a KKT-point of (3.1).*

Proof. Let $(\bar{x}, \bar{\vartheta})$ be an accumulation point of $(x^k, \vartheta^k)_{\mathbb{N}_0}$. Then, there exists a subsequence $(x^k, \vartheta^k)_K$ of $(x^k, \vartheta^k)_{\mathbb{N}_0}$ converging to $(\bar{x}, \bar{\vartheta})$ as $K \ni k \to \infty$. Further, it holds that $0 \le \chi(x^k, \vartheta^k; t^k) \le \varepsilon_k$ for all $k \ge 0$, where χ is defined in (3.3). Since $\varepsilon_k \to 0$ as $k \to \infty$, we obtain from (6.1), Theorem 4.2 and Theorem 5.13 that

$$\varepsilon_k \ge |\min\{-\widetilde{F}_j(x^k; t^k), \vartheta_j^k\}| \to |\min\{-F_j(\bar{x}), \bar{\vartheta}_j\}| = 0 \ \text{ as } \ K \ni k \to \infty, \ \ \forall j \in J \backslash \{0\}.$$

Because $(a, b) \mapsto \min\{a, b\}$ is a complementarity function, we have that $\bar{\vartheta}_j F_j(\bar{x}) = 0$, $F_j(\bar{x}) \le 0$ and $\bar{\vartheta}_j \ge 0$ for all $j \in J \setminus \{0\}$. We can assume w.l.o.g. that the sequences $(\nabla_x \tilde{\varphi}_j(x^k; \tau_k))_K$, $j \in J$, and $(\nabla_x \tilde{\psi}_j(x^k; \nu_k, \eta_k))_K$, $j \in J$, are convergent; cf. Theorem 4.2 and Theorem 5.13. Hence, there exist $v_j, w_j \in \mathbb{R}^n$ such that

$$\nabla_x \tilde{\varphi}_j(x^k; \tau_k) \to v_j, \quad \nabla_x \tilde{\psi}_j(x^k; \nu_k, \eta_k) \to w_j \ \text{ as } \ K \ni k \to \infty, \quad \text{for all} \quad j \in J.$$

Now, let $j \in J$ be arbitrary. We verify that $v_j + w_j \in \partial F_j(\bar{x})$. Theorem 4.2 and Theorem 5.13 apply and yield that $v_j \in \partial \varphi_j(\bar{x})$ and $w_j \in \partial \psi_j(\bar{x})$ due to gradient consistency. Next, [18, Thm. 2.1] and [19, Prop. 2.3.6] show that φ_j and ψ_j are regular according to [19, Def. 2.3.4] and, therefore, [19, Cor. 3 on p. 40] results in $\partial F_j(\bar{x}) = \partial \varphi_j(\bar{x}) + \partial \psi_j(\bar{x})$ showing $v_j + w_j \in \partial F_j(\bar{x})$. Hence, we have that

$$v_0 + w_0 + \sum_{j \in J \backslash \{0\}} \bar{\vartheta}_j (v_j + w_j) \in \partial F_0(\bar{x}) + \sum_{j \in J \backslash \{0\}} \bar{\vartheta}_j \partial F_j(\bar{x}).$$

Moreover, $\chi(x^k, \vartheta^k; t^k) \to 0$ as $k \to \infty$, where χ is defined in (3.3), implies that

$$\nabla_x \widetilde{F}_0(x^k; t^k) + \sum_{j \in J \backslash \{0\}} (\vartheta^k)_j \nabla_x \widetilde{F}_j(x^k; t^k) \to 0 \quad \text{as} \quad K \ni k \to \infty,$$

and, therefore, we infer that $0 \in \partial F_0(\bar{x}) + \sum_{j \in J \backslash \{0\}} \bar{\vartheta}_j \partial F_j(\bar{x})$.   □

If we only assume (x^k) to have a convergent subsequence, we need to impose a suitable CQ for (3.1) to infer convergence of a subsequence of (ϑ^k); cf. [56, Thm. 3.2]. Moreover, the existence of KKT-tuples of the DROP (3.2) may be verified under suitable CQs for (3.1); cf. [56].

7. Numerical examples. We construct DROPs from the Moré-Garbow-Hillstrom test set [39] consisting of standard NLPs modeling design variables as uncertain, which has been considered in, e.g., [8, 35], for RO:

$$(7.1) \qquad\qquad \min_{x \in \mathbb{R}^n} \ \sup_{P \in \mathcal{P}_\epsilon} \ \mathbb{E}_P[f_0(x + \xi)],$$

where $\mathcal{P}_\epsilon$ is defined by

$$(7.2) \qquad\qquad \mathcal{P}_\epsilon = \{ P \in \mathcal{M} : \ \|\mathbb{E}_P[\xi]\|_2 \le \epsilon, \ 0 \preccurlyeq \mathrm{Cov}_P[\xi] \preccurlyeq \epsilon I \}.$$

We choose $\epsilon = \{10^{-3}, 10^{-2}\}$ and refer to Appendix A for a description of how we selected test problems. The goals of our numerical results are to show that our algorithmic scheme is effective in that Algorithm 3.1 is an efficient method to solve

$$(7.3) \qquad\qquad \min_{x \in \mathbb{R}^n} \ F_0(x),$$

where F_0 is defined in (6.1) and $a_0(x) = f_0(x)$, $b_0(x) = \nabla f_0(x)$, $C_0(x) = \nabla^2 f_0(x)$ is chosen in (1.5) and (1.6), and it allows to compute stationary points of (7.3), which are more robust than stationary points of the deterministic

$$(7.4) \qquad \min_{x \in \mathbb{R}^n} \ f_0(x),$$

and the stochastic program

$$(7.5) \qquad \min_{x \in \mathbb{R}^n} \ \mathbb{E}_{\bar{P}_\epsilon}[f_0(x + \xi)],$$

for $\bar{P}_\epsilon = N(0, (\epsilon/10)I)$. We choose $\bar{P}_\epsilon = N(0, (\epsilon/10)I)$ to mimic the set up of the application considered in [20, sect. 4.3].

All problems are solved using `IPOPT` [54] and its `Julia` interface `Ipopt.jl` without modifying options except of the overall termination, which was set to 10^{-4} for all iterations of Algorithm 3.1, and it was chosen as 10^{-5} for the solution of the nominal problems (7.4) and the sample average approximation of (7.5). We use exact Hessian information for nominal and stochastic programs and L-BFGS in Algorithm 3.1. Derivatives are computed with automatic differentiation using the `Julia` package `ForwardDiff` [45] including the gradients of the smoothing functions $\tilde{\varphi}_0$ (see (4.3)) and $\tilde{\psi}_0$ (see (5.54)). Theorem 4.2, Theorem 5.6 and the termination tolerance used in Algorithm 3.1 motivate the choices $\nu_{\min} = 10^{-8}$ and $\eta_{\min}, \tau_{\min} = \sqrt{\nu_{\min}}$, and, moreover, $\nu_0 = 10^{-2}$, and $\eta_0, \tau_0 = \sqrt{\nu_0}$. We define $\nu_{k+1} = \rho^2 \nu_k$, and $\eta_{k+1}, \tau_{k+1} = \rho \eta_k$, where $\rho = 0.1$.

To solve (7.3), we choose x_N^* as initial value in Algorithm 3.1, where x_N^* is the stationary point computed by `IPOPT` for the nominal problem (7.4), which results in less iterations than using the initial values of the test problems. To initialize the solution of (7.3) in the $(k+1)$st iteration of Algorithm 3.1, we use the one obtained in the kst iteration. Numerical values are displayed with four significant figures and we approximated expected values using Monte Carlo with 1000 independent samples.

Our choices for the smoothing parameters imply that Algorithm 3.1 performs five outer iterations. Table 1 lists the median number of corresponding objective function, gradient and Hessian evaluations used by `IPOPT` to compute a stationary point of (7.3) using Algorithm 3.1, of (7.4) and of the sample average approximation of (7.5). Evaluating the smoothing function $\widetilde{F}_0$ (see (6.1)) of the cost function F_0 of (7.3) at (x, t) requires $f_0(x)$ (see (7.4)), $\nabla f_0(x)$ and $\nabla^2 f_0(x)$. To obtain $\nabla_x \widetilde{F}_0(x; t)$, we reuse the gradient $\nabla f_0(x)$, the Hessian $\nabla^2 f_0(x)$, and compute the gradients of $x \mapsto \widetilde{s}^T \nabla^2 f_0(x) \widetilde{s}$, $x \mapsto \nabla f_0(x)^T \widetilde{s}$, where $\widetilde{s}$ is optimal solution of the TRP (5.54), and of two mapping of the form $x \mapsto \nabla^2 f_0(x) \bullet R$, where $R \in \mathbb{S}^p$; cf. (4.6) and (5.32). For efficiency, we exploit that the gradient of $x \mapsto \nabla f_0(x)^T \widetilde{s}$ equals $\nabla^2 f_0(x) \widetilde{s}$. We believe that Table 1 indicates that Algorithm 3.1 is an efficient method for (7.3).

The solution of the TRPs (5.54) using [40, Alg. 3.14] for all iterations of Algorithm 3.1 required less than six iterations making it an effective method. In particular, they are insensitive w.r.t. the choice of smoothing parameters η_k, ν_k. The evaluation of (1.5) using (4.2) instead of applying SDP solvers is about three orders of magnitudes faster, e.g., for $p = 20$, the quotient of the median run time over 100 randomly generated SDPs of the form (1.5) using (4.2) and `SCS` [43] is $4.340 \cdot 10^{-4}$.

Table 2 lists for mgh01 and mgh03 the number of iterations. Moreover, it displays the KKT-error of `IPOPT`, the distance of the stationary point of the current iteration to the one of the previous iteration and the smoothing parameter ν_k for each outer iteration k of Algorithm 3.1. We deduce that empirically the distance

Median number of objective function, gradient, and Hessian evaluations required by IPOPT *for nominal (N), distributionally robust (DR) and stochastic optimization problem (S) of all selected test problems. The number of evaluations for the approximate DROPs* (7.3) *are the sum of all evaluations used within* Algorithm 3.1.

	N	DR	S	N	DR	S	N	S
ϵ	#-f_0	#-$\widetilde{F}_0$	#-f_0	#-∇f_0	#-$\nabla_x \widetilde{F}_0$	#-∇f_0	#-$\nabla^2 f_0$	#-$\nabla^2 f_0$
10^{-3}	14	120	$1.25 \cdot 10^4$	14	37.5	$1.2 \cdot 10^4$	13	$1.1 \cdot 10^4$
10^{-2}	14	190.5	$1.0 \cdot 10^4$	14	56	$1.0 \cdot 10^4$	13	$0.9 \cdot 10^3$

For each outer iteration of Algorithm 3.1 *applied to* (7.1) *and* $\epsilon = 10^{-3}$, *the number of iterations to a compute stationary point of* (3.2), *the final KKT-error, relative distance of the initial point and the stationary point, and value of the smoothing parameter* ν_k.

Problem	k	#-iter	KKT-error	$\frac{\|x^k - x^{k-1}\|_2}{\max\{1, \|x^{k-1}\|_2\}}$	ν_k
mgh01	1	17	$2.272 \cdot 10^{-7}$	0.3329	0.1
	2	10	$2.756 \cdot 10^{-6}$	$9.633 \cdot 10^{-2}$	$1.0 \cdot 10^{-3}$
	3	2	$4.34 \cdot 10^{-5}$	$5.013 \cdot 10^{-5}$	$1.0 \cdot 10^{-5}$
	4	2	$6.355 \cdot 10^{-5}$	$3.975 \cdot 10^{-5}$	$1.0 \cdot 10^{-7}$
	5	2	$9.689 \cdot 10^{-5}$	$3.345 \cdot 10^{-5}$	$1.0 \cdot 10^{-8}$
mgh03	1	25	$7.654 \cdot 10^{-5}$	0.9994	0.1
	2	7	$6.938 \cdot 10^{-7}$	$5.542 \cdot 10^{-5}$	$1.0 \cdot 10^{-3}$
	3	7	$2.303 \cdot 10^{-7}$	$5.495 \cdot 10^{-6}$	$1.0 \cdot 10^{-5}$
	4	5	$4.997 \cdot 10^{-7}$	$5.494 \cdot 10^{-7}$	$1.0 \cdot 10^{-7}$
	5	5	$1.015 \cdot 10^{-6}$	$4.07 \cdot 10^{-8}$	$1.0 \cdot 10^{-8}$

of subsequent stationary points (3.2) computed by Algorithm 3.1 converges to zero and that the number of inner iterations decreases monotonically indicating that the homotopy method is computationally efficient.

For each selected problem, we compare the stationary points x^*_{DR} of (7.3), x^*_N of (7.4) and x^*_S of (7.5) using the following two quantities:

$$(7.6) \quad V_{\mathbb{E}}(x) = \max_{1 \leq i \leq 10} \mathbb{E}_{P_i}[f_0(x + \xi_i)], \quad \text{and} \quad V_{\text{StD}}(x) = \max_{1 \leq i \leq 10} \text{StD}_{P_i}[f_0(x + \xi_i)],$$

where $P_i = N(\mu_i, \sigma_i^2 I) \in \mathcal{P}_\epsilon$, and μ_i and σ_i are independent and uniformly distributed on $\{\mu \in \mathbb{R}^p : \|\mu\|_2 \leq \Delta\}$ and $\{\sigma \in \mathbb{R} : 0 \leq \sigma^2 \leq \epsilon\}$, respectively. Here, StD denotes the standard deviation. The quantities in (7.6) mimic maximum mean and standard deviations of repeated implementations of x and $V_{\mathbb{E}}(x)$ is a lower bound on the objective function value of (7.1) evaluated at x. Table 3 and Table 4 display $V_{\mathbb{E}}(x)$ and $V_{\text{StD}}(x)$, for $x \in \{x^*_{DR}, x^*_N, x^*_S\}$, and $\epsilon \in \{10^{-3}, 10^{-2}\}$. We conclude that in most cases the distributionally robust stationary point has lower mean and standard deviation than nominal and stochastic stationary points.

The problems mgh33 and mgh34 are quadratic w.r.t. ξ (cf. [39, sect. 3]) and, hence, the approximation scheme is exact, i.e., (7.1) is equivalent to (7.3). For the problems mgh10, mgh11 and mgh17, we obtain very different orders of magnitude of $V_{\mathbb{E}}(x)$ and of $V_{\text{StD}}(x)$ for $x \in \{x^*_N, x^*_{DR}, x^*_S\}$, resulting from exponential terms in the corresponding objective functions; cf. [39, sect. 3].

8. Conclusion and outlook. We have provided a new algorithmic scheme for both DRO and RO. The main advantages of our approach are, that the number of constraints of the DROP is the same as for the nominal problem, MPCCs and NSDPs are avoided, and any NLP solver can be used to compute stationary points of the DROPs in Algorithm 3.1. Moreover, it is applicable to a large class of problems

TABLE 3

Quantities $V_{\mathbb{E}}$ and V_{StD} (see (7.6)) evaluated at x_N^*, x_{DR}^*, x_S^* for $\epsilon = 10^{-3}$.

Problem	$V_{\mathbb{E}}(x_N^*)$	$V_{\mathbb{E}}(x_{DR}^*)$	$V_{\mathbb{E}}(x_S^*)$	$V_{\mathrm{StD}}(x_N^*)$	$V_{\mathrm{StD}}(x_{DR}^*)$	$V_{\mathrm{StD}}(x_S^*)$
mgh01	0.1867	0.1536	0.1761	0.2559	0.1528	0.2399
mgh03	$3.175 \cdot 10^6$	$3.135 \cdot 10^1$	$2.899 \cdot 10^1$	$4.507 \cdot 10^6$	$8.083 \cdot 10^1$	$7.328 \cdot 10^1$
mgh04	$3.756 \cdot 10^8$	$3.754 \cdot 10^8$	$3.754 \cdot 10^8$	$5.256 \cdot 10^8$	$5.246 \cdot 10^8$	$5.252 \cdot 10^8$
mgh06	$1.884 \cdot 10^2$	$1.798 \cdot 10^2$	$1.863 \cdot 10^2$	$1.076 \cdot 10^2$	$8.388 \cdot 10^1$	$1.028 \cdot 10^2$
mgh07	0.1778	0.1778	0.1779	0.2089	0.2086	0.209
mgh10	$9.626 \cdot 10^{10}$	$1.356 \cdot 10^6$	$2.134 \cdot 10^6$	$1.303 \cdot 10^{11}$	$3.482 \cdot 10^5$	$1.993 \cdot 10^6$
mgh11	$6.237 \cdot 10^{278}$	$3.283 \cdot 10^1$	$2.258 \cdot 10^{133}$	∞	0.8311	$7.134 \cdot 10^{134}$
mgh13	$4.387 \cdot 10^{-2}$	$4.387 \cdot 10^{-2}$	$4.385 \cdot 10^{-2}$	$5.662 \cdot 10^{-2}$	$5.662 \cdot 10^{-2}$	$5.66 \cdot 10^{-2}$
mgh14	0.7525	0.7492	0.752	0.7223	0.7144	0.7229
mgh17	$7.9421 \cdot 10^{17}$	1.133	$1.735 \cdot 10^{11}$	$2.19 \cdot 10^{19}$	$3.551 \cdot 10^{-2}$	$4.959 \cdot 10^{12}$
mgh20	0.1318	0.1291	0.1309	0.1461	0.1425	0.1453
mgh21	3.92	3.19	3.702	1.723	1.045	1.621
mgh22	0.2164	0.2163	0.2163	0.1219	0.1219	0.1219
mgh25	0.3078	0.3078	0.3073	0.6784	0.6784	0.6768
mgh27	$4.855 \cdot 10^{-2}$	$4.853 \cdot 10^{-2}$	$4.85 \cdot 10^{-2}$	$6.864 \cdot 10^{-2}$	$6.854 \cdot 10^{-2}$	$6.859 \cdot 10^{-2}$
mgh30	0.1408	0.1406	0.1408	$7.52 \cdot 10^{-2}$	$7.485 \cdot 10^{-2}$	$7.519 \cdot 10^{-2}$
mgh31	0.194	0.1924	0.1936	$9.437 \cdot 10^{-2}$	$8.959 \cdot 10^{-2}$	$9.399 \cdot 10^{-2}$
mgh33	$4.514 \cdot 10^2$	$4.514 \cdot 10^2$	$4.508 \cdot 10^2$	$6.369 \cdot 10^2$	$6.369 \cdot 10^2$	$6.365 \cdot 10^2$
mgh34	$2.394 \cdot 10^2$	$2.394 \cdot 10^2$	$2.391 \cdot 10^2$	$3.203 \cdot 10^2$	$3.203 \cdot 10^2$	$3.204 \cdot 10^2$
mgh35	$6.772 \cdot 10^{-2}$	$5.266 \cdot 10^{-2}$	0.1244	0.3531	$2.726 \cdot 10^{-2}$	1.383

TABLE 4

Quantities $V_{\mathbb{E}}$ and V_{StD} (see (7.6)) evaluated at x_N^*, x_{DR}^*, x_S^* for $\epsilon = 10^{-2}$.

Problem	$V_{\mathbb{E}}(x_N^*)$	$V_{\mathbb{E}}(x_{DR}^*)$	$V_{\mathbb{E}}(x_S^*)$	$V_{\mathrm{StD}}(x_N^*)$	$V_{\mathrm{StD}}(x_{DR}^*)$	$V_{\mathrm{StD}}(x_S^*)$
mgh01	1.866	0.7566	1.174	2.581	0.5774	1.548
mgh03	$3.178 \cdot 10^7$	$2.829 \cdot 10^3$	$2.81 \cdot 10^3$	$4.511 \cdot 10^7$	$7.443 \cdot 10^3$	$7.414 \cdot 10^3$
mgh04	$3.752 \cdot 10^9$	$3.728 \cdot 10^9$	$3.744 \cdot 10^9$	$5.246 \cdot 10^9$	$5.142 \cdot 10^9$	$5.233 \cdot 10^9$
mgh06	$1.903 \cdot 10^3$	$8.186 \cdot 10^2$	$1.528 \cdot 10^3$	$5.762 \cdot 10^3$	$2.055 \cdot 10^3$	$4.626 \cdot 10^3$
mgh07	1.793	1.781	1.792	2.129	2.09	2.125
mgh10	$9.616 \cdot 10^{11}$	$9.616 \cdot 10^{11}$	$3.502 \cdot 10^6$	$1.294 \cdot 10^{12}$	$1.294 \cdot 10^{12}$	$3.261 \cdot 10^6$
mgh11	$8.08 \cdot 10^{256}$	$3.283 \cdot 10^1$	$7.044 \cdot 10^{129}$	∞	0.7971	$2.228 \cdot 10^{131}$
mgh13	0.4413	0.4413	0.4412	0.5675	0.5675	0.5674
mgh14	7.537	7.262	7.466	7.318	6.657	7.245
mgh17	$3.857 \cdot 10^{68}$	1.374	$8.32 \cdot 10^{46}$	$1.135 \cdot 10^{70}$	0.3561	$2.527 \cdot 10^{48}$
mgh20	1.321	1.262	1.279	1.487	1.427	1.445
mgh21	$3.941 \cdot 10^1$	$1.556 \cdot 10^1$	$2.496 \cdot 10^1$	$1.758 \cdot 10^1$	3.972	$1.083 \cdot 10^1$
mgh22	2.184	2.183	2.183	1.219	1.219	1.219
mgh25	$1.647 \cdot 10^1$	$1.647 \cdot 10^1$	$1.643 \cdot 10^1$	$5.02 \cdot 10^1$	$5.021 \cdot 10^1$	$5.0 \cdot 10^1$
mgh27	0.4891	0.4876	0.488	0.7019	0.6896	0.6995
mgh30	1.408	1.382	1.402	0.7588	0.7236	0.755
mgh31	2.063	1.846	2.01	1.217	0.8228	1.166
mgh33	$4.516 \cdot 10^3$	$4.516 \cdot 10^3$	$4.506 \cdot 10^3$	$6.392 \cdot 10^3$	$6.392 \cdot 10^3$	$6.381 \cdot 10^3$
mgh34	$2.378 \cdot 10^3$	$2.378 \cdot 10^3$	$2.372 \cdot 10^3$	$3.207 \cdot 10^3$	$3.207 \cdot 10^3$	$3.204 \cdot 10^3$
mgh35	$1.221 \cdot 10^3$	$3.846 \cdot 10^3$	$3.717 \cdot 10^2$	$2.563 \cdot 10^4$	$4.247 \cdot 10^4$	$7.365 \cdot 10^3$

without the need of implementing further algorithms.

Through second order expansions m_j (see (1.3)) of f_j, we obtain tractable approximations to the worst-case functions of (1.1), as TRPs and SDPs are computationally tractable; cf. [5, 6]. The use of second order expansions may bee viewed as a trade off between accuracy and tractability, and may provide more accurate approximations than linearizations.

Appendix A. We selected problems from the Moré-Garbow-Hillstrom test set [39], which is available in `Julia` through the package `NLSProblems.jl` (version as of

TABLE 5

$Z_\epsilon(x_N^*)$ and number of parameters p of problems from the Moré-Garbow-Hillstrom test set with $Z_\epsilon(x_N^*)$ exceeding 10^{-1}, where $\epsilon = 10^{-3}$.

Problem	p	$Z_\epsilon(x_N^*)$	Problem	p	$Z_\epsilon(x_N^*)$	Problem	p	$Z_\epsilon(x_N^*)$
mgh01	2	0.5778	mgh13	4	0.1262	mgh27	10	0.1403
mgh03	2	$1.006 \cdot 10^7$	mgh14	4	1.911	mgh30	10	0.2737
mgh04	2	$1.234 \cdot 10^9$	mgh17	5	$8.83 \cdot 10^{24}$	mgh31	10	0.3551
mgh06	2	1.903	mgh20	6	0.3353	mgh33	10	$5.505 \cdot 10^2$
mgh07	3	0.4727	mgh21	20	7.056	mgh34	10	$2.243 \cdot 10^2$
mgh10	3	$3.524 \cdot 10^9$	mgh22	20	0.4509	mgh35	10	0.8223
mgh11	3	$2.567 \cdot 10^{127}$	mgh25	10	1.282			

Nov 16, 2018) using its default set up as follows: We compute for each test problem a stationary point x_N^* of the nominal problem (7.4) and $Z_\epsilon(x_N^*)$ defined by

$$Z_\epsilon(x_N^*) = \mathbb{E}_{N(0,\epsilon I)}[X(x_N^*)] + \mathrm{StD}_{N(0,\epsilon I)}[X(x_N^*)], \quad X(x_N^*)(\xi) = \frac{f_0(x_N^* + \xi) - f_0(x_N^*)}{\max\{1, |f_0(x_N^*)|\}},$$

and select problems fulfilling $Z_\epsilon(x_N^*) \geq 10^{-1}$ for $\epsilon = 10^{-3}$; cf. Table 5.

A related approach has been used in [4] to investigate uncertain linear programs.

REFERENCES

[1] R. ANDREANI, G. HAESER, AND J. M. MARTÍNEZ, *On sequential optimality conditions for smooth constrained optimization*, Optimization, 60 (2011), pp. 627–641, https://doi.org/10.1080/02331930903578700.

[2] A. BEN-TAL, D. DEN HERTOG, AND J.-P. VIAL, *Deriving robust counterparts of nonlinear uncertain inequalities*, Math. Program., 149 (2015), pp. 265–299, https://doi.org/10.1007/s10107-014-0750-8.

[3] A. BEN-TAL, L. EL GHAOUI, AND A. NEMIROVSKI, *Robust Optimization*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2009.

[4] A. BEN-TAL AND A. NEMIROVSKI, *Robust solutions of linear programming problems contaminated with uncertain data*, Math. Program., 88 (2000), pp. 411–424, https://doi.org/10.1007/PL00011380.

[5] A. BEN-TAL AND A. NEMIROVSKI, *Lectures on Modern Convex Optimization*, MOS-SIAM Series on Optimization, SIAM, Philadelphia, 2001, https://doi.org/10.1137/1.9780898718829.

[6] A. BEN-TAL AND M. TEBOULLE, *Hidden convexity in some nonconvex quadratically constrained quadratic programming*, Math. Program., 72 (1996), pp. 51–63, https://doi.org/10.1007/BF02592331.

[7] D. BERTSIMAS, D. B. BROWN, AND C. CARAMANIS, *Theory and Applications of Robust Optimization*, SIAM Rev., 53 (2011), pp. 464–501, https://doi.org/10.1137/080734510.

[8] D. BERTSIMAS, O. NOHADANI, AND T. K., *Nonconvex Robust Optimization for Problems with Constraints*, INFORMS J. Comput., 22 (2010), pp. 44–58, https://doi.org/10.1287/ijoc.1090.0319.

[9] D. BERTSIMAS, O. NOHADANI, AND K. M. TEO, *Robust Optimization for Unconstrained Simulation-Based Problems*, Oper. Res., 58 (2010), pp. 161–178, https://doi.org/10.1287/opre.1090.0715.

[10] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer Science & Business Media New York, Berlin, Heidelberg, New York, 2013, https://doi.org/10.1007/978-1-4612-1394-9.

[11] V. V. BULDYGIN AND Y. V. KOZACHENKO, *Metric Characterization of Random Variables and Random Processes*, vol. 188, American Mathematical Society, Providence, Rhode Island, 2000.

[12] J. V. BURKE AND T. HOHEISEL, *Epi-convergent Smoothing with Applications to Convex Composite Functions*, SIAM J. Optim., 23 (2013), pp. 1457–1479, https://doi.org/10.1137/120889812.

[13] J. V. BURKE AND T. HOHEISEL, *Epi-convergence Properties of Smoothing by Infimal Convolution*, Set-Valued Var. Anal., 25 (2017), pp. 1–23, https://doi.org/10.1007/s11228-016-0362-y.

[14] J. V. Burke, T. Hoheisel, and C. Kanzow, *Gradient Consistency for Integral-convolution Smoothing Functions*, Set-Valued Var. Anal., 21 (2013), pp. 359–376, https://doi.org/10.1007/s11228-013-0235-6.

[15] X. Chen, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program., 134 (2012), pp. 71–99, https://doi.org/10.1007/s10107-012-0569-0.

[16] X. Chen, H. Qi, L. Qi, and K.-L. Teo, *Smooth Convex Approximation to the Maximum Eigenvalue Function*, J. Global Optim., 30 (2004), pp. 253–270, https://doi.org/10.1007/s10898-004-8271-2.

[17] Z. Chen, M. Sim, and H. Xu, *Distributionally Robust Optimization with Infinitely Constrained Ambiguity Sets*, History, (2017).

[18] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., 205 (1975), pp. 247–262, https://doi.org/10.1090/S0002-9947-1975-0367131-6.

[19] F. H. Clarke, *Optimization and Nonsmooth Analysis*, vol. 5 of Classics in Applied Mathematics, SIAM, Philadelphia, 1990, https://doi.org/10.1137/1.9781611971309.

[20] E. Delage and Y. Ye, *Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems*, Oper. Res., 58 (2010), pp. 595–612, https://doi.org/10.1287/opre.1090.0741.

[21] M. Diehl, H. G. Bock, and E. Kostina, *An approximation technique for robust nonlinear optimization*, Math. Program., 107 (2006), pp. 213–230, https://doi.org/10.1007/s10107-005-0685-1.

[22] J. Dutta, K. Deb, R. Tulshyan, and R. Arora, *Approximate KKT points and a proximity measure for termination*, J. Global Optim., 56 (2013), pp. 1463–1499, https://doi.org/10.1007/s10898-012-9920-5.

[23] P. M. Esfahani and D. Kuhn, *Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations*, Math. Program., 171 (2018), pp. 115–166, https://doi.org/10.1007/s10107-017-1172-1, https://doi.org/10.1007/s10107-017-1172-1.

[24] A. V. Fiacco and Y. Ishizuka, *Sensitivity and stability analysis for nonlinear programming*, Ann. Oper. Res., 27 (1990), pp. 215–235, https://doi.org/10.1007/BF02055196.

[25] O. E. Flippo and B. Jansen, *Duality and sensitivity in nonconvex quadratic optimization over an ellipsoid*, Eur. J. Oper. Res., 94 (1996), pp. 167–178, https://doi.org/10.1016/0377-2217(95)00199-9.

[26] A. Forsgren, P. E. Gill, and M. H. Wright, *Interior methods for nonlinear optimization*, SIAM Rev., 44 (2002), pp. 525–597, https://doi.org/10.1137/S0036144502414942.

[27] A. L. Fradkov and V. A. Yakubovich, *The S-procedure and duality relations in nonconvex problems of quadratic programming*, Vestnik Leningrad. Univ. Math., 6 (1979), pp. 101–109. In Russian, 1973.

[28] R. Gao and A. J. Kleywegt, *Distributionally robust stochastic optimization with Wasserstein distance*, arXiv preprint arXiv:1604.02199, (2016).

[29] J. Goh and M. Sim, *Distributionally robust optimization and its tractable approximations*, Oper. Res., 58 (2010), pp. 902–917, https://doi.org/10.1287/opre.1090.0795.

[30] W. W. Hogan, *Point-to-set maps in mathematical programming*, SIAM Rev., 15 (1973), pp. 591–603, https://doi.org/10.1137/1015073.

[31] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2 ed., 2013, https://doi.org/10.1017/9781139020411.

[32] B. Houska and M. Diehl, *Nonlinear Robust Optimization via Sequential Convex Bilevel Programming*, Math. Program., 142 (2013), pp. 539–577, https://doi.org/10.1007/s10107-012-0591-2.

[33] C. Kanzow and A. Schwartz, *The price of inexactness: convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited*, Math. Oper. Res., 40 (2014), pp. 253–275, https://doi.org/10.1287/moor.2014.0667.

[34] P. Kolvenbach, O. Lass, and S. Ulbrich, *An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty*, Optim. Eng., 19 (2018), pp. 697–731, https://doi.org/10.1007/s11081-018-9388-3.

[35] O. Lass and S. Ulbrich, *Model order reduction techniques with a posteriori error control for nonlinear robust optimization governed by partial differential equations*, SIAM J. Sci. Comput., 39 (2017), pp. S112–S139, https://doi.org/10.1137/16M108269X.

[36] A. S. Lewis, *Derivatives of Spectral Functions*, Math. Oper. Res., 21 (1996), pp. 576–588, https://doi.org/10.1287/moor.21.3.576.

[37] A. S. Lewis, *Nonsmooth analysis of eigenvalues*, Math. Program., 84 (1999), pp. 1–24, https://doi.org/10.1007/s10107980004a.

[38] M. M. Mäkelä and P. Neittaanmäki, *Nonsmooth Optimization: Analysis and Algorithms with Applications to Optimal Control*, World Scientific Singapore, Singapore, 1992, https://doi.org/10.1142/1493.

[39] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, *Testing unconstrained optimization software*, ACM Trans. Math. Softw., 7 (1981), pp. 17–41.

[40] J. J. Moré and D. C. Sorensen, *Computing a Trust Region Step*, SIAM J. Sci. and Stat. Comput., 4 (1983), pp. 553–572, https://doi.org/10.1137/0904038.

[41] Y. Nesterov, *Smoothing technique and its applications in semidefinite optimization*, Math. Program., 110 (2007), pp. 245–259, https://doi.org/10.1007/s10107-006-0001-8.

[42] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994, https://doi.org/10.1137/1.9781611970791.

[43] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, *Conic Optimization via Operator Splitting and Homogeneous Self-Dual Embedding*, J. Optim. Theory Appl., 169 (2016), pp. 1042–1068, https://doi.org/10.1007/s10957-016-0892-3.

[44] E. Polak, J. O. Royset, and R. Womersley, *Algorithms with Adaptive Smoothing for Finite Minimax Problems*, J. Optim. Theory Appl., 119 (2003), pp. 459–484, https://doi.org/10.1023/B:JOTA.0000006685.60019.3e.

[45] J. Revels, M. Lubin, and T. Papamarkou, *Forward-Mode Automatic Differentiation in Julia*, arXiv:1607.07892, (2016), https://arxiv.org/abs/1607.07892.

[46] A. Shapiro, *Distributionally robust stochastic programming*, SIAM J. Optim., 27 (2017), pp. 2258–2275, https://doi.org/10.1137/16M1058297.

[47] A. Shapiro and A. Kleywegt, *Minimax analysis of stochastic problems*, Optim. Methods Softw., 17 (2002), pp. 523–542, https://doi.org/10.1080/1055678021000034008.

[48] A. M.-C. So, *Moment inequalities for sums of random matrices and their applications in optimization*, Math. Program., 130 (2011), pp. 125–151, https://doi.org/10.1007/s10107-009-0330-5.

[49] D. Sorensen, *Newton's Method with a Model Trust Region Modification*, SIAM J. Numer. Anal., 19 (1982), pp. 409–426, https://doi.org/10.1137/0719026.

[50] S. Steffensen and M. Ulbrich, *A new relaxation scheme for mathematical programs with equilibrium constraints*, SIAM J. Optim., 20 (2010), pp. 2504–2539, https://doi.org/10.1137/090748883.

[51] R. J. Stern and H. Wolkowicz, *Indefinite Trust Region Subproblems and Nonsymmetric Eigenvalue Perturbations*, SIAM J. Optim., 5 (1995), pp. 286–313, https://doi.org/10.1137/0805016.

[52] P. D. Tao and L. T. H. An, *Lagrangian stability and global optimality in nonconvex quadratic minimization over Euclidean balls and spheres*, J. Convex Anal., 2 (1995), pp. 263–276.

[53] N.-K. Tsing, M. K. Fan, and E. I. Verriest, *On analyticity of functions involving eigenvalues*, Linear Algebra Appl., 207 (1994), pp. 159–180, https://doi.org/10.1016/0024-3795(94)90009-4.

[54] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Math. Program., 106 (2006), pp. 25–57, https://doi.org/10.1007/s10107-004-0559-y.

[55] W. Wiesemann, D. Kuhn, and M. Sim, *Distributionally Robust Convex Optimization*, Oper. Res., 62 (2014), pp. 1358–1376, https://doi.org/10.1287/opre.2014.1314.

[56] M. Xu, J. J. Ye, and L. Zhang, *Smoothing SQP methods for Solving Degenerate Nonsmooth Constrained Optimization Problems with Applications to Bilevel Programs*, SIAM J. Optim., 25 (2015), pp. 1388–1410, https://doi.org/10.1137/140971580.

[57] Y. Xu, W. Sun, and L. Qi, *A feasible direction method for the semidefinite program with box constraints*, Appl. Math. Lett., 24 (2011), pp. 1874–1881, https://doi.org/10.1016/j.aml.2011.05.010.

[58] Y. Zhang, *General Robust-Optimization Formulation for Nonlinear Programming*, J. Optim. Theory Appl., 132 (2007), pp. 111–124, https://doi.org/10.1007/s10957-006-9082-z.

[59] C. Zhao and Y. Guan, *Data-driven risk-averse stochastic optimization with Wasserstein metric*, Oper. Res. Lett., 46 (2018), pp. 262–267, https://doi.org/10.1016/j.orl.2018.01.011.