J Glob Optim (2008) 41:319–334
DOI 10.1007/s10898-007-9232-3

Smoothing by mollifiers. Part I: semi-infinite optimization

Hubertus Th. Jongen · Oliver Stein

Received: 7 December 2006 / Accepted: 5 September 2007 / Published online: 3 October 2007

Abstract We show that a compact feasible set of a standard semi-infinite optimization problem can be approximated arbitrarily well by a level set of a single smooth function with certain regularity properties. This function is constructed as the mollification of the lower level optimal value function. Moreover, we use correspondences between Karush–Kuhn–Tucker points of the original and the smoothed problem, and between their associated Morse indices, to prove the connectedness of the so-called min–max digraph for semi-infinite problems.

Keywords Semi-infinite optimization · Smoothing · Mollifier · Stationarity · Morse index

AMS Subject Classification 90C34 · 90C31 · 57R12

1 Introduction

We consider the semi-infinite optimization problem

$$SIP: \min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad g(x,y)\geq 0 \quad \text{for all } y\in Y$$

with objective function $f\in C^2(\mathbb{R}^n,\mathbb{R})$, constraint function $g\in C^2(\mathbb{R}^n\times\mathbb{R}^m,\mathbb{R})$, and a nonempty and compact index set $Y\subset\mathbb{R}^m$. We assume that Y is described by finitely many inequality constraints,

$$Y = \{y\in\mathbb{R}^m|\ v(y)\geq 0\}$$

with $v\in C^2(\mathbb{R}^m,\mathbb{R}^s)$ and $s\in\mathbb{N}$. Problems of this type, in which a finite-dimensional decision variable is subject to infinitely many inequality constraints, are called semi-infinite.

H. Th. Jongen
Department of Mathematics – C, RWTH Aachen University, 52056 Aachen, Germany
e-mail: jongen@rwth-aachen.de

O. Stein (✉)
School of Economics and Business Engineering, University of Karlsruhe, 76128 Karlsruhe, Germany
e-mail: stein@wior.uni-karlsruhe.de

They have been studied systematically since the 1960s. Important early contributions regarding optimality conditions and duality theory for semi-infinite problems are given in [2,10] for linear semi-infinite problems, and in [16,36] for nonlinear problems. For excellent reviews with hundreds of references on semi-infinite programming we refer to [17,30]. A standard reference for linear semi-infinite problems is [11], and [18,31,32] overview the existing numerical methods for linear and nonlinear problems. A new numerical approach was recently introduced in [9].

We denote the feasible set of *SIP* by

$$M \;=\; \{x \in \mathbb{R}^n \mid g(x,y) \geq 0 \text{ for all } y \in Y\}.$$

A basic problem in semi-infinite optimization is to check whether a point $x \in \mathbb{R}^n$ is feasible, since this involves the verification of infinitely many inequality constraints. After some preliminaries in Sect. 2, in Sect. 3 we will show that a nonempty and compact feasible set M can be approximated arbitrarily well by a level set of a single smooth function with certain regularity properties. This function will be constructed with the aid of a so-called mollifier. Moreover, we will show a correspondence between Karush–Kuhn–Tucker points of the original and the smoothed problem, along with their Morse indices. The latter result will enable us to prove the connectedness of the so-call min–max digraph for semi-infinite problems in Sect. 4. However, the result about Morse indices in Theorem 3.7(c) has a very elaborate proof on the one hand, and it is basically related to finite, and not semi-infinite programming, on the other hand. For these reasons this proof is given in [24], the second part of this article, along with more results on the finite case.

We emphasize that a smoothing procedure for *finite* optimization problems is given in [22]. There the main idea is to use the logarithmic barrier approach to approximate finitely many inequality constraints $g_i(x) \geq 0,\ i \in I,\ |I| < \infty$, by one smooth and nondegenerate constraint $\sum_{i\in I} \ln(g_i(x)) \geq \ln(\varepsilon)$ for $\varepsilon > 0$. Under mild assumptions it is shown that the approximating feasible sets M^ε converge to the original feasible set M in the Hausdorff distance, that for sufficiently small $\varepsilon > 0$ the sets M^ε are homeomorphic to M, and that there is a correspondence between the Karush–Kuhn–Tucker points of the original and the smoothed problem, along with their Morse indices. A similar approach is taken in [14] to smooth finite maximum functions. In the following we briefly explain why obvious generalizations of this approach to semi-infinite programming are not successful.

There are two standard arguments which connect semi-infinite to finite optimization problems. First, a sufficiently fine discretization of the index set leads to an arbitrarily accurate outer approximation of M by finitely many inequality constraints which could, in a next step, be smoothed by the logarithmic barrier approach. Unfortunately, the so-called second order shift-terms of semi-infinite programs (see Sect. 2.2) are ignored by the discretized problem, so that correspondences of Morse indices cannot even be established between the original and the discretized problem, let alone the smoothed discretized problem.

Second, assuming the regularity conditions of the so-called Reduction Ansatz (see Sect. 2.2) at some point $\bar{x} \in M$, the feasible set can be described by finitely many smooth inequality constraints locally around $\bar{x}$. The logarithmic barrier approach for this locally reduced semi-infinite problem is used in [20]. While Morse indices are modeled well in this approach, the assumption of the Reduction Ansatz in the whole feasible set is not generic [26] and, thus, too strong for our analysis.

Another obvious generalization of the approach from finite programming is to directly use the barrier term $\int_Y \ln(g(x,y))\,dy$ for the semi-infinite problem. For infinite quadratic programming problems a related interior point approach is presented in [28]. For nonlinear semi-infinite problems, however, this logarithmic barrier term is neither self-concordant nor does it necessarily enforce interior points, as an example from [20] shows. The main problem is that in some situations even the singularity of the logarithm is smoothed by the integral, and boundary points can become feasible for the approximation (note that $\int \ln(y) = y\ln(y) - y$ can be continuously extended to $y = 0$ with value 0). Logarithmic barrier methods for the numerical solution of semi-infinite problems, but unrelated to a smooth and nondegenerate approximation of the feasible set, are studied in [1,27].

An approximation of the feasible set in semi-infinite optimization by a quadratic distance function is presented in [13]. While smoothness of the approximating problem is shown, it is inherently degenerate, so that no results on Morse indices can be expected from this approach.

To our knowledge there is little work on the use of mollifiers in optimization. A basic reference for the definition of subgradients for certain discontinuous functions by mollification is [6].

2 Preliminaries

2.1 Properties of the feasible set and stationarity

Under our continuity and compactness assumptions it is easy to see that the semi-infinite constraint in *SIP* is equivalent to

$$G(x) \;:=\; \min_{y\in Y}\, g(x, y) \;\geq\; 0,$$

which means that the feasible set M is the upper level set of some auxiliary function:

$$M \;=\; \{x \in \mathbb{R}^n |\; G(x) \geq 0\}. \tag{2.1}$$

The auxiliary function G is related to the so-called *lower level problem*

$$Q(x):\; \min_{y\in\mathbb{R}^m}\, g(x, y) \quad \text{subject to} \quad v(y) \;\geq\; 0. \tag{2.2}$$

In contrast to the upper level problem which consists in minimizing f over M with the decision variable x, in the lower level problem x plays the role of an n-dimensional parameter, and y is the decision variable. In fact, the function G is the optimal value function of this parametric problem, that is, $G(x)$ is the globally minimal *value* of $Q(x)$ for $x \in \mathbb{R}^n$. We denote the globally minimal *points* of $Q(x)$ by

$$Y_\star(x) \;=\; \{y \in Y |\; g(x, y) = G(x)\}.$$

Since G is continuous [5], M is a closed set, and a feasible point $\bar{x}$ with $G(\bar{x}) > 0$ lies in the topological interior of M.

For investigations of the local structure of M or of local optimality conditions we are only interested in points from the boundary ∂M of M. In view of the continuity of G it suffices to consider the zeros of G, that is, points $x \in \mathbb{R}^n$ for which $Q(x)$ has vanishing minimal value. We denote the corresponding globally minimal points of $Q(x)$ by

$$Y_0(x) \;=\; \{y \in Y |\; g(x, y) = 0\}.$$

The set $Y_0(x)$ is also called the *active index set* of x for *SIP*. Note that each point $x \in \partial M$ satisfies $Y_0(x) \neq \emptyset$, but that the reverse is not necessarily the case.

A nice topological structure of M at its boundary points can be guaranteed under constraint qualifications. To formulate a basic constraint qualification, we need some smoothness property of G. As an optimal value function, G is not necessarily smooth, but due to a result by Danskin it is at least directionally differentiable:

Theorem 2.1 ([5]) *The optimal value function G of $Q(x)$ is directionally differentiable at each $\bar{x} \in \mathbb{R}^n$ with*

$$G'(\bar{x}, d) \;=\; \min_{y \in Y_\star(\bar{x})} D_x g(\bar{x}, y)\, d \tag{2.3}$$

for all $d \in \mathbb{R}^n$.

Here $D_x g$ denotes the row vector of partial derivatives of g with respect to x.

According to [35], for directionally differentiable problems the natural extension of the well-known and basic Mangasarian–Fromovitz constraint qualification [29] at a zero $\bar{x}$ of G is

$$\{d \in \mathbb{R}^n \,|\, G'(\bar{x}, d) > 0\} \;\neq\; \emptyset.$$

Plugging in (2.3) we obtain the following definition which is well-known for semi-infinite programs [18,26].

Definition 2.2 At $\bar{x} \in M$ the Extended Mangasarian–Fromovitz Constraint Qualification (EMFCQ) is said to hold if there exists some vector $d \in \mathbb{R}^n$ with

$$D_x g(\bar{x}, y)\, d \;>\; 0 \quad \text{for all} \;\; y \in Y_0(\bar{x}). \tag{2.4}$$

Each vector $d \in \mathbb{R}^n$ satisfying (2.4) is called EMF vector at $\bar{x}$.

Note that (2.4) is trivially satisfied for $G(\bar{x}) > 0$ since then we have $Y_0(\bar{x}) = \emptyset$. Moreover, it is not hard to see that at a zero $\bar{x}$ of G in the topological interior of M no EMF vector can exist. Hence, under EMFCQ the zero set of G and ∂M coincide or, in other words, we have $x \in \partial M$ if and only if $Y_0(x) \neq \emptyset$. In [26] it is shown that EMFCQ holds generically in semi-infinite programming, that is, for defining functions of *SIP* in general position. It is, thus, a weak assumption.

The following lemma is easy to see.

Lemma 2.3 *Let $d \in \mathbb{R}^n$ be an EMF vector at $\bar{x} \in \partial M$. Then there exists some $\bar{t} > 0$ such that for all $t \in (0, \bar{t})$ we have $G(\bar{x} + td) > 0$.*

At a local minimizer $\bar{x} \in \partial M$ of *SIP* the system

$$Df(\bar{x})d \;<\; 0, \quad G'(\bar{x}, d) \;=\; \min_{y \in Y_0(\bar{x})} D_x g(\bar{x}, y)\, d \;>\; 0 \tag{2.5}$$

can thus not be solvable, and by the Lemma of Gordan [3], the latter is equivalent to

$$0 \;\in\; \operatorname{conv}\left(-Df(\bar{x}),\; D_x g(\bar{x}, y),\; y \in Y_0(\bar{x})\right), \tag{2.6}$$

where conv stands for the convex hull. This leads to the following theorem.

Theorem 2.4 ([21]) *Let $\bar{x} \in \partial M$ be a local minimizer of SIP. Then there exist $p \in \mathbb{N}$, $y^i \in Y_0(\bar{x})$ and nontrivial multipliers $\kappa \geq 0$, $\lambda_i \geq 0$, $1 \leq i \leq p$, such that*

$$\kappa\, Df(\bar{x}) \;-\; \sum_{i=1}^{p} \lambda_i\, D_x g(\bar{x}, y^i) \;=\; 0. \tag{2.7}$$

Using the Lemma of Gordan again, it is easy to see that under EMFCQ at $\bar{x}$ one can choose $\kappa = 1$ in (2.7). Any point $\bar{x} \in M$ that satisfies (2.7) with $p \in \mathbb{N}$, $\kappa = 1$, $y^i \in Y_0(\bar{x})$, and $\lambda_i \geq 0$, $1 \leq i \leq p$, is called *Karush–Kuhn–Tucker (KKT) point* of *SIP*.

2.2 The Reduction Ansatz

For theoretical as well as numerical purposes it is of crucial importance to keep track of the elements of the lower level solution set $Y_\star(x)$ when its argument x varies. Being solutions of $Q(x)$, under a constraint qualification they satisfy the first order necessary optimality condition of Karush–Kuhn–Tucker. Whereas the results in the previous section hold for an abstract index set Y, here we will need the description of Y by inequality constraints $v_k,\ k \in K := \{1,\ldots,s\}$. We say that the *Linear Independence Constraint Qualification (LICQ)* holds at $\bar{y}$ in Y if the gradients $Dv_k(\bar{y}),\ k \in K_0(\bar{y})$, are linearly independent. Here $K_0(\bar{y}) = \{k \in K|\ v_k(\bar{y}) = 0\}$ is the lower level active index set. Let

$$\mathcal{L}(x, y, \gamma) \ = \ g(x, y) - \gamma^\top v(y)$$

denote the Lagrangian of $Q(x)$ with multiplier vector $\gamma \in \mathbb{R}^s$. Then for $\bar{x} \in M$ and each $\bar{y} \in Y_\star(\bar{x})$ such that LICQ holds at $\bar{y}$ in $Q(\bar{x})$, there exists a unique multiplier vector $\bar{\gamma} \geq 0$ with $D_y\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma}) = 0$ and $\gamma_k \cdot v_k(\bar{y}) = 0,\ k \in K$.

Keeping track of the elements of $Y_\star(x)$ can now be achieved, for example, by means of the implicit function theorem. For $\bar{x} \in M$ a local minimizer $\bar{y}$ of $Q(\bar{x})$ is called *nondegenerate* in the sense of Jongen/Jonker/Twilt [25], if LICQ, strict complementary slackness (SCS) and the second order sufficiency condition (SOSC) $D_y^2\mathcal{L}(\bar{x}, \bar{y}, \bar{\gamma})|_{T_{\bar{y}}Y} \succ 0$ are satisfied. Here $D_y^2\mathcal{L}$ is the Hessian of $\mathcal{L}$ with respect to y, $T_{\bar{y}}Y$ is the tangent space to Y at $\bar{y}$, and $A \succ 0$ stands for the positive definiteness of a matrix A. SCS means $\bar{\gamma}_k > 0$ for all $k \in K_0(\bar{y})$.

The *Reduction Ansatz* is said to hold at $\bar{x} \in M$ if all global minimizers of $Q(\bar{x})$ are nondegenerate. Since nondegenerate minimizers are isolated, and Y is a compact set, its closed subset $Y_\star(\bar{x})$ can only contain finitely many points, say $Y_\star(\bar{x}) = \{\bar{y}^1,\ldots,\bar{y}^p\}$ with $p \in \mathbb{N}$. By a result from [8] the local variation of these points with x can be described by the implicit function theorem.

In fact, for x locally around $\bar{x}$ there exist continuously differentiable functions $y^i(x),\ 1 \leq i \leq p$, with $y^i(\bar{x}) = \bar{y}^i$ such that $y^i(x)$ is the locally unique local minimizer of $Q(x)$ around $\bar{y}^i$. Moreover, if $\bar{\gamma}^i$ is the uniquely determined multiplier vector corresponding to $\bar{y}^i$, then there exists a continuously differentiable function $\gamma^i(x)$ with $\gamma^i(\bar{x}) = \bar{\gamma}^i$ such that $\gamma^i(x)$ is the unique multiplier vector corresponding to $y^i(x),\ 1 \leq i \leq p$. It turns out that the functions $G_i(x) := g(x, y^i(x))$ are even C^2 in a neighborhood of $\bar{x}$. Their gradients are

$$DG_i(\bar{x}) \ = \ D_x g(\bar{x}, \bar{y}^i),$$

and their Hessians can be computed to be

$$D^2G_i(\bar{x}) \ = \ D_x^2 g(\bar{x}, \bar{y}^i) + S(\bar{x}, \bar{y}^i, \bar{\gamma}^i)$$

with the so-called second order shift terms

$$S(\bar{x}, \bar{y}^i, \bar{\gamma}^i) \ = \ -\begin{pmatrix} D_{yx}^2 g \\ 0 \end{pmatrix}^\top \begin{pmatrix} D_y^2\mathcal{L} & -D^\top v_{K_0^i} \\ -Dv_{K_0^i} & 0 \end{pmatrix}^{-1} \begin{pmatrix} D_{yx}^2 g \\ 0 \end{pmatrix}\Bigg|_{(\bar{x}, \bar{y}^i, \bar{\gamma}^i)},$$

where $v_{K_0^i}$ stands for the vector with entries $v_k,\ k \in K_0(\bar{y}^i)$.

The Reduction Ansatz was originally formulated in [16,36]. A major consequence of the Reduction Ansatz is the so-called Reduction Lemma: if the Reduction Ansatz holds at $\bar{x}$, then for all x from a neighborhood U of $\bar{x}$ we have

$$G(x) \ = \ \min_{1 \leq i \leq p} G_i(x).$$

In view of (2.1) this means that M can locally be described by finitely many C^2-constraints, that is, *SIP* is locally equivalent to the smooth finite optimization problem

$$SIP_{red}: \qquad \min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad G_i(x) \geq 0, \quad i = 1, \ldots, p. \tag{2.8}$$

Examples show that the Reduction Ansatz cannot be expected to hold everywhere in the feasible set of a generic semi-infinite program [26]. As nondegeneracy in the sense of Jongen/Jonker/Twilt is a local property, we can, however, define a nondegenerate KKT point of *SIP* via the locally reduced problem SIP_{red}. Let

$$L(x, \lambda) \ = \ f(x) - \sum_{i=1}^{p} \lambda_i \, G_i(x)$$

denote the Lagrangian of SIP_{red} with multiplier vector $\lambda \in \mathbb{R}^p$.

Definition 2.5 A point $\bar{x} \in M$ is called nondegenerate Karush–Kuhn–Tucker point of *SIP* if the Reduction Ansatz holds at $\bar{x}$ and if $\bar{x}$ is a nondegenerate Karush–Kuhn–Tucker point of SIP_{red}, that is, the following three conditions hold:

(a) LICQ holds at $\bar{x}$, and there exists a (unique) multiplier vector $\bar{\lambda} \geq 0$ with $D_x L(\bar{x}, \bar{\lambda}) = 0$.
(b) The multipliers satisfy $\bar{\lambda}_i > 0, \ i = 1, \ldots, p$.
(c) The matrix $D_x^2 L(\bar{x}, \bar{\lambda})|_{T_{\bar{x}} M_{red}}$ is nonsingular.

The number of negative eigenvalues of the matrix in (c) is called the Morse index of $\bar{x}$.

For generic semi-infinite problems all Karush–Kuhn–Tucker points are nondegenerate [34,37]. In this sense, nondegeneracy of KKT points for *SIP* is a weak assumption. Clearly, a nondegenerate KKT point of *SIP* is a local minimizer if and only if its Morse index vanishes.

2.3 Mollifiers

With the Euclidean norm $||\cdot||_2$ on $\mathbb{R}^n$ the *standard mollifier* (cf., e.g., [7]) is the C^∞-function

$$\eta(x) \ = \ \begin{cases} C \exp\left(\frac{1}{||x||_2^2 - 1} \right), & ||x||_2 < 1 \\ 0, & ||x||_2 \geq 1, \end{cases}$$

where $C > 0$ is chosen such that $\int_{\mathbb{R}^n} \eta(x)\, dx = 1$. For $\varepsilon > 0$ we put

$$\eta_\varepsilon(x) \ = \ \frac{1}{\varepsilon^n} \, \eta\left(\frac{x}{\varepsilon} \right).$$

The function η_ε is also C^∞, it satisfies $\int_{\mathbb{R}^n} \eta_\varepsilon(x)\, dx = 1$, and its support $\overline{\{x \in \mathbb{R}^n \,|\, \eta_\varepsilon(x) \neq 0\}}$ is the closed ball $\overline{B(0, \varepsilon)}$ with $B(0, \varepsilon) = \{x \in \mathbb{R}^n \,|\, ||x||_2 < \varepsilon\}$, where $\bar{A}$ denotes the topological closure of a set A.

Definition 2.6 For $\varepsilon > 0$ the $\varepsilon-$mollification of a locally integrable function $F : \mathbb{R}^n \to \mathbb{R}$ is the convolution $F^\varepsilon = \eta_\varepsilon * F$ on $\mathbb{R}^n$, that is,

$$F^\varepsilon(x) \ = \ \int_{\mathbb{R}^n} \eta_\varepsilon(x - z) F(z)\, dz \ = \ \int_{B(0,\varepsilon)} \eta_\varepsilon(z) F(x - z)\, dz$$

for all $x \in \mathbb{R}^n$.

Theorem 2.7 ([7])

(a) For all $\varepsilon > 0$, the ε-mollification F^ε is in $C^\infty(\mathbb{R}^n, \mathbb{R})$.
(b) If F is continuous on $\mathbb{R}^n$, then F^ε converges to F uniformly on compact sets for $\varepsilon \to 0$.

For further details about mollifiers we refer the interested reader to [7].

3 The smoothing approach

3.1 Main results

Throughout this section we make the following three assumptions.

Assumption 3.1 The feasible set M of *SIP* is nonempty and compact.

Assumption 3.2 The EMFCQ holds everywhere in M.

Assumption 3.3 All KKT points of *SIP* are nondegenerate.

Our smoothing approach is based on the mollification of the optimal value function G:

$$G^\varepsilon = \eta_\varepsilon * G = \eta_\varepsilon * \min_{y \in Y} g(\cdot, y).$$

In view of Theorem 2.7 the function G^ε is C^∞ for each $\varepsilon > 0$, and G^ε converges to G uniformly on compact sets for $\varepsilon \to 0$.

Intuitively, for sufficiently small $\varepsilon > 0$ the set

$$M^\varepsilon = \{x \in \mathbb{R}^n | \; G^\varepsilon(x) \geq 0\},$$

and the smooth finite optimization problem

$$SIP^\varepsilon : \min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad G^\varepsilon(x) \geq 0$$

should be strongly related to M and *SIP*, respectively. We will make this statement precise in the following theorems, which hold under our general Assumptions 3.1–3.3.

Theorem 3.4 *M^ε converges to M in the Hausdorff distance for $\varepsilon \to 0$.*

Theorem 3.5 *For all sufficiently small $\varepsilon > 0$, EMFCQ holds everywhere in the set M^ε.*

Theorem 3.6 *For all sufficiently small $\varepsilon > 0$, the set M^ε is homeomorphic with M.*

Theorem 3.7

(a) The set $KKT(f, M)$ of Karush–Kuhn–Tucker points of SIP is finite.
(b) For each $\bar{x} \in KKT(f, M)$ let $U(\bar{x})$ be some neighborhood of $\bar{x}$. Then outside the sets $U(\bar{x})$, $\bar{x} \in KKT(f, M)$, the problem SIP^ε has no KKT points for sufficiently small $\varepsilon > 0$.
(c) The neighborhoods $U(\bar{x})$, $\bar{x} \in KKT(f, M)$, from part (b) can be chosen such that each $U(\bar{x})$ contains exactly one KKT point x^ε of SIP^ε for sufficiently small $\varepsilon > 0$. Moreover, x^ε is nondegenerate, and the Morse index of $\bar{x}$ in SIP and the Morse index of x^ε in SIP^ε coincide.

In the remainder of this section we will prove Theorems 3.4, 3.5, 3.6, and parts (a) and (b) of Theorem 3.7. The proof Theorem 3.7(c) is only given for a special case, since the proof for the general case would go beyond the scope of the present article. The complete proof can be found in the separate paper [24].

3.2 Convergence of M^ε to M

Lemma 3.8 *Let $\bar{x} \in \mathbb{R}^n$, $\varepsilon > 0$, $a \leq b$, and $G(x) \in [a, b]$ for all $x \in B(\bar{x}, \varepsilon)$. Then we also have $G^\varepsilon(\bar{x}) \in [a, b]$.*

Proof As the set $\bar{x} - B(0, \varepsilon)$ coincides with $B(\bar{x}, \varepsilon)$, we have $G(\bar{x} - z) \in [a, b]$ for all $z \in B(0, \varepsilon)$. The nonnegativity of η_ε implies

$$G^\varepsilon(\bar{x}) \;=\; \int_{B(0,\varepsilon)} \eta_\varepsilon(z) G(\bar{x} - z)\, dz \;\geq\; a \cdot \int_{B(0,\varepsilon)} \eta_\varepsilon(z)\, dz \;=\; a$$

and analogously $G^\varepsilon(\bar{x}) \leq b$. □

We denote the *distance* of a point $a \in \mathbb{R}^n$ from a set $B \subset \mathbb{R}^n$ by

$$\text{dist}(a, B) \;=\; \inf_{b \in B} ||a - b||_2\,,$$

the *directed distance* of a set $A \subset \mathbb{R}^n$ to a set $B \subset \mathbb{R}^n$ by

$$h(A, B) \;=\; \sup_{a \in A} \text{dist}(a, B) \;=\; \sup_{a \in A} \inf_{b \in B} ||a - b||_2\,,$$

and the *Hausdorff distance* of the sets A and B by

$$\text{dist}_H(A, B) \;=\; \max\{\, h(A, B),\; h(B, A)\,\}.$$

Proof of Theorem 3.4 We give the proof in two parts. In the first part of the proof we show that $h(M^\varepsilon, M)$ tends to zero for $\varepsilon \to 0$. Choose $\varepsilon > 0$ and some arbitrary $\bar{x} \in M^\varepsilon$, that is, $G^\varepsilon(\bar{x}) \geq 0$. Assume by contradiction that $\text{dist}(\bar{x}, M) > \varepsilon$. Then the whole ball $B(\bar{x}, \varepsilon)$ is contained in the set complement of M, that is, $G(x) < 0$ holds for all $x \in B(\bar{x}, \varepsilon)$. As an immediate consequence of Lemma 3.8 we obtain $G^\varepsilon(\bar{x}) < 0$, in contradiction to the choice of $\bar{x}$. It follows $\text{dist}(\bar{x}, M) \leq \varepsilon$ and $h(M^\varepsilon, M) = \sup_{x \in M^\varepsilon} \text{dist}(x, M) \leq \varepsilon$. This shows $\lim_{\varepsilon \to 0} h(M^\varepsilon, M) = 0$.

In the second part of the proof we show $\lim_{\varepsilon \to 0} h(M, M^\varepsilon) = 0$. Assume by contradiction that this is not the case. Then there exist some $\delta > 0$ and a sequence $\varepsilon^\nu \searrow 0$ such that for all $\nu \in \mathbb{N}$ we have

$$\delta \;<\; h(M, M^{\varepsilon^\nu}) \;=\; \sup_{x \in M} \text{dist}(x, M^{\varepsilon^\nu}).$$

Hence, for each $\nu \in \mathbb{N}$ there exists some $x^\nu \in M$ with

$$\delta \;<\; \text{dist}(x^\nu, M^{\varepsilon^\nu}). \tag{3.1}$$

By the compactness of M, the sequence $(x^\nu)_\nu$ converges to some $\bar{x} \in M$ without loss of generality (here and in the following we briefly write $(x^\nu)_\nu$ for a sequence $(x^\nu)_{\nu \in \mathbb{N}}$). The point $\bar{x}$ either satisfies $G(\bar{x}) > 0$ or $G(\bar{x}) = 0$.

Case 1: $G(\bar{x}) > 0$.
In view of Theorem 2.7(b), the sequence $(G^{\varepsilon^\nu}(\bar{x}))_\nu$ converges to $G(\bar{x}) > 0$. Therefore we have $G^{\varepsilon^\nu}(\bar{x}) > 0$ for sufficiently large $\nu \in \mathbb{N}$, that is, $\bar{x} \in M^{\varepsilon^\nu}$. It follows

$$0 \;\leq\; \text{dist}(x^\nu, M^{\varepsilon^\nu}) \;=\; \inf_{z \in M^{\varepsilon^\nu}} ||x^\nu - z||_2 \;\leq\; ||x^\nu - \bar{x}||_2 \;\to\; 0,$$

in contradiction to (3.1).

Case 2: $G(\bar{x}) = 0$.
Since EMFCQ holds at $\bar{x}$, we can choose an EMF vector $d \in \mathbb{R}^n$ and, in view of Lemma 2.3, obtain $G(\bar{x} + td) > 0$ for all sufficiently small $t > 0$. Now choose a sequence $t^k \searrow 0$. For each $k \in \mathbb{N}$ the sequence $(G^{\varepsilon^\nu}(\bar{x} + t^k d))_\nu$ converges to $G(\bar{x} + t^k d)$ which is positive for sufficiently large k. Therefore we have $G^{\varepsilon^{\nu_k}}(\bar{x} + t^k d) > 0$ for all sufficiently large $\nu_k \in \mathbb{N}$, that is, $\bar{x} + t^k d \in M^{\varepsilon^{\nu_k}}$. It follows

$$0 \le \operatorname{dist}(x^{\nu_k}, M^{\varepsilon^{\nu_k}}) \le ||x^{\nu_k} - (\bar{x} + t^k d)||_2 \le ||x^{\nu_k} - \bar{x}||_2 + t^k ||d||_2 \to 0,$$

in contradiction to (3.1).

The combination of the two parts of the proof shows that $\operatorname{dist}_H(M^\varepsilon, M)$ tends to zero for $\varepsilon \to 0$. □

For later use we remark that with the same techniques as in the first part of the above proof it is easy to show convergence of the sets $(G^\varepsilon)^{-1}(0)$ to $G^{-1}(0)$ in the directed distance, where $(G^\varepsilon)^{-1}(0)$ and $G^{-1}(0)$ denote the zeros sets of G^ε and G, respectively:

$$\lim_{\varepsilon \to 0} h\left((G^\varepsilon)^{-1}(0),\ G^{-1}(0)\right) = 0. \tag{3.2}$$

3.3 Stability of EMFCQ and homeomorphy of M^ε with M

Regarding the next lemma note that for a directionally differentiable function G the directional derivative $G'(x, d)$ is not necessarily continuous in (x, d) (take, e.g., $G(x) = -|x|$, $\bar{x} = 0$, $d_0 = 1$).

Lemma 3.9 *For $\bar{x}, d_0 \in \mathbb{R}^n$ let $G'(\bar{x}, d_0) > 0$. Then there exist neighborhoods U and D of $\bar{x}$ and d_0, respectively, such that for all $x \in U$ and all $d \in D$ we have $G'(x, d) > 0$.*

Proof Assume the contrary. Then there exist sequences $x^\nu \to \bar{x}$ and $d^\nu \to d_0$ with

$$0 \ge G'(x^\nu, d^\nu) = \min_{y \in Y_\star(x^\nu)} D_x g(x^\nu, y)\, d^\nu.$$

Hence for each $\nu \in \mathbb{N}$ there exists some $y^\nu \in Y_\star(x^\nu)$ with

$$0 \ge D_x g(x^\nu, y^\nu)\, d^\nu. \tag{3.3}$$

As the sequence $(y^\nu)_\nu$ is contained in the compact set Y it converges to $\bar{y} \in Y$ without loss of generality. The closedness of the solution point mapping $Y_\star(\cdot)$ [19] implies even $\bar{y} \in Y_\star(\bar{x})$. Taking the limit in (3.3) thus leads to the contradiction

$$0 \ge D_x g(\bar{x}, \bar{y})\, d_0 \ge \min_{y \in Y(\bar{x})} D_x g(\bar{x}, y)\, d_0 = G'(\bar{x}, d_0) > 0.$$

□

The following lemma shows that the directional derivative of the smooth function G^ε can be written as the mollification of the directional derivative of G.

Lemma 3.10 *For all $x, d \in \mathbb{R}^n$ and $\varepsilon > 0$ we have*

$$\begin{aligned} DG^\varepsilon(x)\, d &= \int_{B(0,\varepsilon)} \eta_\varepsilon(z)\, G'(x - z, d)\, dz \\ &= \int_{B(0,\varepsilon)} \eta_\varepsilon(z) \min_{y \in Y_\star(x-z)} (D_x g(x - z, y)\, d)\, dz. \end{aligned}$$

Proof Consider the functions

$$F^k_{x,d}(z) \;=\; \eta_\varepsilon(z)\,\frac{G(x+(1/k)d-z)-G(x-z)}{1/k}$$

on $\mathbb{R}^n$ with $k \in \mathbb{N}$. It is not hard to see that

$$DG^\varepsilon(x)\,d \;=\; \lim_{k\to\infty}\int_{B(0,\varepsilon)} F^k_{x,d}(z)\,dz$$

holds. Moreover, due to the directional differentiability of G the pointwise limit of $F^k_{x,d}$ exists and is given by $F^\star_{x,d}(z) = \eta_\varepsilon(z)\,G'(x-z,d)$, so that we may write

$$\int_{B(0,\varepsilon)} \eta_\varepsilon(z)\,G'(x-z,d)\,dz \;=\; \int_{B(0,\varepsilon)} \lim_{k\to\infty} F^k_{x,d}(z)\,dz\,.$$

Hence, the first equation of the assertion is shown if in the above formulas taking limits and integrals may be interchanged. We will prove the latter using Lebesgue's dominated convergence theorem.

In fact, since G is Lipschitz continuous on compact sets (cf., e.g., [4]), we may choose a Lipschitz constant L for G on $B(x,2\varepsilon)$. Then for all sufficiently large k and all $z \in B(0,\varepsilon)$ we obtain

$$\left|\frac{G(x+(1/k)d-z)-G(x-z)}{1/k}\right| \;\le\; L||d||$$

and, thus, for all $z \in \mathbb{R}^n$

$$\left|F^k_{x,d}(z)\right| \;\le\; L||d||\,\eta_\varepsilon(z)$$

with $\int_{\mathbb{R}^n} L||d||\,\eta_\varepsilon(z)\,dz \;=\; L||d|| \;<\; \infty$. Due to the dominated convergence theorem this shows the first asserted equation. The second equation immediately follows by plugging in (2.3). □

Lemma 3.11 *For $\bar{x}, d_0 \in \mathbb{R}^n$ let $G'(\bar{x},d_0) > 0$. Then there exist neighborhoods U and D of $\bar{x}$ and d_0, respectively, and some $\bar{\varepsilon} > 0$ such that for all $x \in U$, $d \in D$ and $\varepsilon \in (0,\bar{\varepsilon})$ we have $DG^\varepsilon(x)\,d > 0$.*

Proof From Lemma 3.9 let $\widetilde{U}$ and D be the neighborhoods of $\bar{x}$ and d_0, respectively, such that $G'(x,d) > 0$ holds for all $x \in \widetilde{U}$ and $d \in D$. Choose $\delta > 0$ with $B(\bar{x},\delta) \subset \widetilde{U}$, and put $\bar{\varepsilon} = \delta/2$ as well as $U = B(\bar{x},\bar{\varepsilon})$. Then for all $x \in U$, $d \in D$, $\varepsilon \in (0,\bar{\varepsilon})$, and $z \in B(0,\varepsilon)$ we have $x - z \in B(\bar{x},\delta) \subset \widetilde{U}$ and, thus, $G'(x-z,d) > 0$. Lemma 3.10 and the positivity of η_ε on $B(0,\varepsilon)$ now imply

$$DG^\varepsilon(x)\,d \;=\; \int_{B(0,\varepsilon)} \eta_\varepsilon(z)\,G'(x-z,d)\,dz \;>\; 0.$$

□

Proof of Theorem 3.5 For later use in the proof of Theorem 3.6 we prove a slightly stronger result than required in Theorem 3.5. In fact, let $d \in C^1(\mathbb{R}^n,\mathbb{R}^n)$ be a bounded vector field such that for all $x \in \partial M$, $d(x)$ is an EMF vector at x (cf. [26,33] for the construction of such a field). We will actually show that for all sufficiently small $\varepsilon > 0$ and all $x \in \partial M^\varepsilon$, the (unperturbed) vector $d(x)$ can be used as an EMF vector at x.

For each $x \in \partial M$ we choose the neighborhoods $U(x)$ and $D(x)$ as well as $\bar{\varepsilon}(x)$ from Lemma 3.11. Since ∂M is compact, we can choose finitely many points $x^i \in \partial M$, $1 \leq i \leq m \in \mathbb{N}$ with $\partial M \subset \bigcup_{i=1}^m U(x^i)$. With the abbreviations $U^i = U(x^i)$, $D^i = D(x^i)$, and $\bar{\varepsilon}^i = \bar{\varepsilon}(x^i)$, Lemma 3.11 yields $DG^\varepsilon(x)d > 0$ for all $x \in U^i$, $d \in D^i$, $\varepsilon \in (0, \bar{\varepsilon}^i)$, and $1 \leq i \leq m$.

As the compact set ∂M has a positive distance from the open set $\bigcup_{i=1}^m U^i$, we can find some sufficiently small $\alpha > 0$ so that $\{x \in \mathbb{R}^n \mid \mathrm{dist}(x, \partial M) < \alpha\}$, the "$\alpha$-tube" around ∂M, is contained in $\bigcup_{i=1}^m U^i$. Due to the continuity of the vector field d, for sufficiently small $\alpha > 0$ and after possibly increasing m, we have $DG^\varepsilon(x)d(x) > 0$ for all $x \in U^i$, $\varepsilon \in (0, \bar{\varepsilon}^i)$, and $1 \leq i \leq m$.

Recall that under EMFCQ the sets ∂M and $G^{-1}(0)$ coincide. In view of (3.2) we can choose $\varepsilon > 0$ so small that $(G^\varepsilon)^{-1}(0)$, the zero set of G^ε, is contained in the latter α-tube and, hence, also in $\bigcup_{i=1}^m U^i$. This means that for each $x \in (G^\varepsilon)^{-1}(0)$ we can find some $i \in \{1, \ldots, m\}$ with $x \in U^i$ and, thus, $DG^\varepsilon(x)d(x) > 0$ for all $\varepsilon \in (0, \bar{\varepsilon}^i)$. In particular, x cannot be an interior point of M^ε. Now the assertion follows for all $\varepsilon \in (0, \min_{1 \leq i \leq m} \bar{\varepsilon}^i)$. □

We emphasize that for M^ε the notions of EMFCQ and LICQ coincide.

Proof of Theorem 3.6 The bounded vector field d from the proof of Theorem 3.5 is completely integrable and defines a flow on $\mathbb{R}^n$. As we have seen in the above proof, for sufficiently small $\varepsilon > 0$ and $x \in \partial M^\varepsilon$ we have $DG^\varepsilon(x)d(x) \neq 0$, so that the trajectories of this flow intersect ∂M^ε transversally, with a change of sign in G^ε while passing ∂M^ε. The proof can now be completed along the lines of the corresponding proofs in [12,26]. □

3.4 Correspondences between KKT points

It remains to show Theorem 3.7.

Proof of Theorem 3.7(a) Since all elements of $KKT(f, M)$ are nondegenerate, they are isolated elements of the compact set M. As the set $KKT(f, M)$ is closed, it can thus contain only finitely many elements. □

Proof of Theorem 3.7(b) For each $\bar{x} \in KKT(f, M)$ let $U(\bar{x})$ be some neighborhood of $\bar{x}$. By part a) the set $KKT(f, M)$ is finite, so that

$$V = \bigcup_{\bar{x} \in KKT(f,M)} U(\bar{x})$$

is an open set. It follows that $M \setminus V$ is compact and contains no KKT points of *SIP*.

Now consider any point $\bar{x} \in \partial M$ which is not a KKT point of *SIP*. Since EMFCQ holds at $\bar{x}$, (2.7) does not have a nontrivial solution $\kappa \geq 0$, $\lambda_i \geq 0$, $1 \leq i \leq p$ with $p \in \mathbb{N}$, neither. This means that (2.6) is violated and, by the Lemma of Gordan, that (2.5) has a solution $d_0 \in \mathbb{R}^n$. In particular, we have $Df(\bar{x})d_0 < 0$, and d_0 is an EMF vector at $\bar{x}$.

According to Lemma 3.11 we can choose a neighborhood U of $\bar{x}$ and some $\bar{\varepsilon}$ such that $DG^\varepsilon(x)d_0 > 0$ holds for all $x \in U$ and $\varepsilon \in (0, \bar{\varepsilon})$. After possibly shrinking U we also have $Df(x)d_0 < 0$ for all $x \in U$. By the Lemma of Gordan, for all $x \in U$ and $\varepsilon \in (0, \bar{\varepsilon})$ we find $0 \notin \mathrm{conv}(-Df(x), DG^\varepsilon(x))$. The latter in particular holds for all $x \in U \cap M^\varepsilon$. For $x \in U \cap \partial M^\varepsilon$ it means that x does not even satisfy the Fritz John condition, let alone the KKT condition, and no $x \in U$ from the interior of M^ε can be stationary since $Df(x)$ cannot vanish. We have thus shown that for each $\bar{x} \in \partial M$ we can construct a neighborhood $U(\bar{x})$

and a scalar $\bar{\varepsilon}(\bar{x}) > 0$ such that for all $\varepsilon \in (0, \bar{\varepsilon}(\bar{x}))$ no element of $U(\bar{x}) \cap M^\varepsilon$ is a KKT point for SIP^ε.

An analogous result can be shown for any point $\bar{x}$ from the topological interior of M which is not a KKT point of *SIP*. In fact, such a point is simply characterized by the condition $Df(\bar{x}) \neq 0$, and by continuity there exists some neighborhood $U(\bar{x})$ lying in the interior of M with $Df(x) \neq 0$ for all $x \in U(\bar{x})$. In view of Theorem 3.4 there exists some $\bar{\varepsilon}(\bar{x})$ such that for all $\varepsilon \in (0, \bar{\varepsilon}(\bar{x}))$ the neighborhood $U(\bar{x})$ is also contained in the interior of M^ε, that is, the KKT condition of SIP^ε is $Df(x) = 0$ for any $x \in U(\bar{x})$. We have thus shown that for all $\varepsilon \in (0, \bar{\varepsilon}(\bar{x}))$ no element of $U(\bar{x}) \cap M^\varepsilon$ is a KKT point for SIP^ε.

Since the compact set $M \setminus V$ can be covered by finitely many of such neighborhoods $U(x^i)$, $1 \leq i \leq m$, with $m \in \mathbb{N}$ and $x^i \in M \setminus V$, the assertion follows for all $\varepsilon \in (0, \min_{1 \leq i \leq m} \bar{\varepsilon}(x^i))$. □

We prove Theorem 3.7(c) only for a special case in the present article. Let $\bar{x}$ be one of the finitely many (nondegenerate) KKT points of *SIP*. We have to show that there exists a neighborhood U of $\bar{x}$ such that for sufficiently small $\varepsilon > 0$ the set U contains exactly one KKT point x^ε of SIP^ε. Moreover, x^ε has to be nondegenerate and to possess the same Morse index as $\bar{x}$.

As the Reduction Ansatz holds at $\bar{x}$, locally around $\bar{x}$ the problem *SIP* is equivalent to the locally reduced problem (2.8). It is thus sufficient to show the assertion for the *finite* optimization problem

$$SIP_{red}: \min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad G_i(x) \geq 0, \; i = 1, \dots, p$$

with C^2-constraints G_i, $i = 1, \dots, p$, all of which are active at $\bar{x}$. With $I = \{1, \dots, p\}$, $G(x) = \min_{i \in I} G_i(x)$, and

$$G^\varepsilon(x) = \int_{B(0,\varepsilon)} \eta_\varepsilon(z) G(x - z)\, dz \tag{3.4}$$

its smoothing is

$$SIP^\varepsilon: \min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad G^\varepsilon(x) \geq 0.$$

In order to facilitate the exploration of relations between SIP_{red} and SIP^ε at $\varepsilon = 0$, we transform the integral in (3.4) as follows:

$$\begin{aligned} G^\varepsilon(x) &= \int_{B(0,\varepsilon)} \frac{1}{\varepsilon^n} \eta\left(\frac{z}{\varepsilon}\right) G(x - z)\, dz = \frac{1}{\varepsilon^n} \int_{B(0,1)} \eta\left(\frac{\varepsilon\zeta}{\varepsilon}\right) G(x - \varepsilon\zeta)\, \varepsilon^n\, d\zeta \\ &= \int_{B(0,1)} \eta(z) G(x - \varepsilon z)\, dz. \end{aligned}$$

From this reformulation it is clear that $G^\varepsilon(x)$ depends smoothly on ε, and that $G^0(x) = G(x)$ holds. We emphasize that $G^\varepsilon(x)$ is now even well-defined for *negative* values of ε.

Proof of Theorem 3.7(c) for $\boldsymbol{p = 1}$ In the case $p = 1$ only one smooth constraint appears in SIP_{red}. Then for $k \in \{1, 2\}$ we can write

$$D^k G^\varepsilon(x) = \int_{B(0,1)} \eta(z) D^k G(x - \varepsilon z)\, dz$$

and consider the function

$$H(\varepsilon, x, \lambda) = \begin{pmatrix} D^\top f(x) - \lambda\, D^\top G^\varepsilon(x) \\ G^\varepsilon(x) \end{pmatrix}$$
$$= \begin{pmatrix} D^\top f(x) - \lambda \int_{B(0,1)} \eta(z) D^\top G(x - \varepsilon z)\, dz \\ \int_{B(0,1)} \eta(z) G(x - \varepsilon z)\, dz \end{pmatrix}.$$

For the nondegenerate KKT point $\bar{x}$ of SIP_{red} with multiplier $\bar{\lambda} > 0$ we find $H(0, \bar{x}, \bar{\lambda}) = 0$, and the Jacobian

$$D_{(x,\lambda)} H(0, \bar{x}, \bar{\lambda}) \;=\; \begin{pmatrix} D^2 L(\bar{x}, \bar{\lambda}) & -D^\top G(\bar{x}) \\ DG(\bar{x}) & 0 \end{pmatrix}$$

is nonsingular. Hence, by the implicit function theorem there exist some $\bar{\varepsilon} > 0$ and functions $x \in C^1((-\bar{\varepsilon}, \bar{\varepsilon}), \mathbb{R}^n)$, $\lambda \in C^1((-\bar{\varepsilon}, \bar{\varepsilon}), \mathbb{R})$ with $x(0) = \bar{x}$ and $\lambda(0) = \bar{\lambda}$ such that for all $\varepsilon \in (-\bar{\varepsilon}, \bar{\varepsilon})$, $(x(\varepsilon), \lambda(\varepsilon))$ is the locally unique solution of $H(x, \lambda, \varepsilon) = 0$. The latter means that G^ε is active at $x(\varepsilon)$ and that $x(\varepsilon)$ is critical for SIP^ε. By continuity arguments, for sufficiently small $\bar{\varepsilon} > 0$, $\lambda(\varepsilon)$ is positive, and the Morse indices of $\bar{x}$ and $x(\varepsilon)$ coincide. Altogether, there exists a neighborhood U of $\bar{x}$ such that for all $\varepsilon \in (0, \bar{\varepsilon})$ the point $x^\varepsilon := x(\varepsilon)$ is the only KKT point of SIP^ε. Moreover, x^ε is nondegenerate and possesses the same Morse index as $\bar{x}$. □

Due to the nonsmoothness of G, in the case $p > 1$ the proof of Theorem 3.7(c) is significantly more elaborate. Its details are presented in [24].

4 Connectedness of the min–max digraph

Assume that at any $x \in M$ we can define ascent and descent directions for f. Then these define ascent and descent flows for f, respectively. For compact M suppose that all local minima and maxima of f on M are isolated critical points. Starting in a neighborhood of a local minimum we follow the ascent flow and might reach a local maximum. From there we step downwards via the descent flow and might reach a local minimum. Perhaps the latter minimum is different from the former one, and we repeat the afore-standing procedure. In this way we obtain a kind of "bang–bang" path in M which connects certain local minima and local maxima. The main question that arises is whether we can reach all local minima via such a bang–bang strategy. Of course, we have to assume that M is connected, since we only use local information.

Even for finitely many constraints, in general the answer to the latter question is negative. A two-dimensional counterexample was given by H. Zank (Pers. Comm.), and the general mechanism which generates obstructions is presented in [15]. In fact, in [15] it is shown that certain stable absorbing cycles may appear. On the other hand, a special global adaptation of the metric, constructed in [22], gives a positive result. Moreover, [23] presents an automatic adaptation of the metric based on local information which generically gives a positive result.

As our subsequent analysis is based on the results in [23], we briefly recall the main ideas. We slightly strengthen the assumptions of Sect. 3 throughout Sect. 4 as follows.

Assumption 4.1 The feasible set M of *SIP* is nonempty, compact, and connected.

Assumption 4.2 The EMFCQ holds everywhere in M.

Assumption 4.3 All KKT points of $\pm f$ on M are nondegenerate.

A smooth mapping $\mathcal{R}$ from $\mathbb{R}^n$ to the set of all symmetric positive definite (n, n)-matrices is called a variable or Riemannian metric. The gradient $\text{grad}_{\mathcal{R}} f(\bar{x})$ of f at $\bar{x}$ with respect to the metric $\mathcal{R}$ is defined to be the unique vector ξ satisfying the system $v^\top \cdot \mathcal{R}(\bar{x}) \cdot \xi = v^\top \cdot D^\top f(\bar{x}),\ v \in \mathbb{R}^n$. Note that $\text{grad}_{\mathcal{R}} f(\bar{x}) = \mathcal{R}(\bar{x})^{-1} D^\top f(\bar{x})$.

Now we fix some sufficiently small $\varepsilon > 0$ and consider the set

$$M^\varepsilon = \{ x \in \mathbb{R}^n \mid G^\varepsilon(x) \geq 0 \}$$

with the C^∞-function G^ε from Sect. 3. According to Theorems 3.5, 3.6, and 3.7, Assumptions 4.1, 4.2 and 4.3 do not only hold for *SIP*, but with sufficiently small $\varepsilon > 0$ also for SIP^ε.

In [23] the main idea for the construction of an appropriate Riemannian metric is to equalize the inequality constraint by adding a quadratic slack variable. Thus the problem

$$\text{minimize} \;\; f(x) \;\; \text{subject to} \;\; (x, z) \in \widetilde{M}^\varepsilon$$

is considered, where

$$\widetilde{M}^\varepsilon = \left\{ (x, z) \in \mathbb{R}^n \times \mathbb{R} \mid G^\varepsilon(x) - \tfrac{1}{2} z^2 = 0 \right\}.$$

Any Riemannian metric on $\mathbb{R}^n$ induces a Riemannian metric $\widetilde{\mathcal{R}}$ on $\mathbb{R}^{n+1}$ by putting

$$\widetilde{\mathcal{R}}(x, z) = \left(\begin{array}{c|c} \mathcal{R}(x) & 0 \\ \hline 0 & 1 \end{array} \right)$$

for $(x, z) \in \mathbb{R}^{n+1}$. The x-part of the gradient $\text{grad}_{\widetilde{\mathcal{R}}} f$ of f on $\widetilde{M}$ with respect to $\widetilde{\mathcal{R}}$ can be computed to be [23]

$$\text{grad}_{\widetilde{\mathcal{R}},x} f = \text{grad}_{\mathcal{R}} f - \left(\frac{DG^\varepsilon \mathcal{R}^{-1} D^\top f}{2G^\varepsilon + DG^\varepsilon \mathcal{R}^{-1} D^\top G^\varepsilon} \right) \mathcal{R}^{-1} D^\top G^\varepsilon.$$

It is not hard to see that the vector field $x \mapsto \text{grad}_{\widetilde{\mathcal{R}},x} f(x)$ is smooth on M^ε, and that it induces a smooth flow $\Psi : \mathbb{R} \times M^\varepsilon \to M^\varepsilon$ with the interior and the boundary of M^ε as invariant manifolds [23]. Integrating this flow forwards or backwards in time yields ascent and descent flows, respectively.

The "bang–bang" strategy mentioned in the introduction of this section can now be put in mathematical terms as follows. Let $\bar{x}_1, \ldots, \bar{x}_p$ and $\bar{y}_1, \ldots, \bar{y}_q$ be the local minima and the local maxima of f on M^ε, respectively. Choose arbitrarily small neighborhoods (*germs*) $U_{\bar{x}_1}, \ldots, U_{\bar{y}_q}$ of $\bar{x}_1, \ldots, \bar{y}_q$ in M^ε. We define the so-called (bipartite) *min–max digraph* as follows.

Definition 4.4 The set of nodes of the min–max digraph is partitioned into the set of local minima $\{\bar{x}_1, \ldots, \bar{x}_p\}$ and the set of local maxima $\{\bar{y}_1, \ldots, \bar{y}_q\}$. There exists an arc from $\bar{x}_i$ to $\bar{y}_j$ (from $\bar{y}_j$ to $\bar{x}_i$) iff the ascent flow (descent flow) connects some point from $U_{\bar{x}_i}$ $(U_{\bar{y}_j})$ with a point of $U_{\bar{y}_j}$ $(U_{\bar{x}_i})$.

In terms of the min–max digraph we can reach all local minima via a bang–bang strategy in case that the digraph is strongly connected. The latter is true for generic metrics:

Theorem 4.5 ([23]) *For generic $\mathcal{R}$ the corresponding min–max digraph of SIP^ε is strongly connected.*

We emphasize that so-called *decomposition points* of f on M^ε, that is, certain KKT points with Morse index 1, play a crucial role in the proof of Theorem 4.5. In fact, first the connectivity of the graph with local minimizers and decomposition points as node sets is shown, and then the decomposition points are "lifted" to local maximizers.

In view of Theorem 3.7 the corresponding KKT points (especially the local minima and maxima) of SIP^ε are arbitrarily close to those of the unperturbed problem *SIP*. This shows that *SIP* can be approximated arbitrarily well by a smooth finite problem SIP^ε with strongly connected min–max digraph. We emphasize that, in contrast to the argumentation in [22,23], the min–max digraph of SIP^ε does not induce a min–max digraph of *SIP* in our present approach, since M^ε is not necessarily an *inner* approximation of M.

Acknowledgements We thank the anonymous referees for precise and substantial remarks which lead to a significantly improved version of this manuscript. The second author gratefully acknowledges financial support through a Heisenberg grant of the *Deutsche Forschungsgemeinschaft*.

References

1. Abbe, L.: Two logarithmic barrier methods for convex semi-infinite problems. In: Goberna, M.A., López, M.A. (eds.) Semi-infinite Programming—Recent Advances, pp. 169–195. Kluwer, Dordrecht (2001)
2. Brosowski, B.: Parametric Semi-infinite Optimization. Peter Lang, Frankfurt a.M. (1982)
3. Cheney, E.W.: Introduction to Approximation Theory. McGraw-Hill, New York (1966)
4. Clarke, F.H.: Generalized gradients and applications. Trans. Am. Math. Soc. **205**, 247–262 (1975)
5. Danskin, J.M.: The Theory of Max–Min and its Applications to Weapons Allocation Problems. Springer, New York (1967)
6. Ermoliev, Y.M., Norkin, V.I., Wets, R.J.B.: The minimization of semicontinuous functions: mollifier subgradients. SIAM J. Control Optim. **33**, 149–167 (1995)
7. Evans, L.C.: Partial Differential Equations. American Mathematical Society, Providence (1998)
8. Fiacco, A.V., McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. Wiley, New York (1968)
9. Floudas, C.A., Stein, O.: The adaptive convexification algorithm: a feasible point method for semi-infinite programming SIAM J Optim, to appear.
10. Glashoff, K., Gustafson, S.A.: Linear Optimization and Approximation. Springer, Berlin (1983)
11. Goberna, M.A., López, M.A.: Linear Semi-infinite Optimization. Wiley, Chichester (1998)
12. Guddat, J., Jongen, H.Th., Rückmann, J.-J.: On stability and stationary points in nonlinear optimization. J. Aust. Math. Soc. **B28**, 36–56 (1986)
13. Guerra Vasquez, F., Rückmann, J.-J.: An approximation of feasible sets in semi-infinite optimization. Top **10**, 325–336 (2002)
14. Guerra Vasquez, F., Günzel, H., Jongen, H.Th.: On logarithmic smoothing of the maximum function. Ann. Oper. Res. **101**, 209–220 (2001)
15. Günzel, H., Jongen, H.Th.: On absorbing cycles in min–max digraphs. J. Glob. Optim. **31**, 85–92 (2005)
16. Hettich, R., Jongen, H.Th.: Semi-infinite programming: conditions of optimality and applications. In: Stoer, J. (ed.) Optimization Techniques, Part 2, Lecture Notes in Control and Information Sciences, vol. 7, pp. 1–11. Springer, Berlin (1978)
17. Hettich, R., Kortanek, K.O.: Semi-infinite programming: theory, methods, and applications. SIAM Rev. **35**, 380–429 (1993)
18. Hettich, R., Zencke, P.: Numerische Methoden der Approximation und semi-infiniten Optimierung. Teubner, Stuttgart (1982)
19. Hogan, W.W.: Point-to-set maps in mathematical programming. SIAM Rev. **15**, 591–603 (1973)
20. Jarre, F.: Comparing two interior-point approaches for semi-infinite programs, Preprint, University of Trier (1999)
21. John, F.: Extremum problems with inequalities as subsidiary conditions, In: Studies and Essays, R. Courant Anniversary Volume, pp. 187–204. Interscience, New York (1948)
22. Jongen, H.Th., Ruiz Jhones, A.: Nonlinear optimization: on the min–max digraph and global smoothing. In: Ioffe, A., Reich, S., Shafrir, I. (eds.) Calculus of Variations and Differential Equations. Chapman and Hall/CRC Research Notes in Mathematics Series, vol. 410, pp. 119–135. CRC Press, Boca Raton FL (1999)

23. Jongen, H.Th., Stein, O.: Constrained global optimization: adaptive gradient flows. In: Floudas, C.A., Pardalos, P.M. (eds.) Frontiers in Global Optimization, pp. 223–236. Kluwer Academic Publishers, Boston (2004)
24. Jongen, H.Th., Stein, O.: Smoothing by mollifiers. Part II: Nonlinear optimization. J Glob Optim doi: 10.1007/s10898-007-9231-4.
25. Jongen, H.Th., Jonker, P., Twilt, F.: Critical sets in parametric optimization. Math. Program. **34**, 333–353 (1986)
26. Jongen, H.Th., Twilt, F., Weber, G.-W.: Semi-infinite optimization: structure and stability of the feasible set. J. Optim. Theory Appl. **72**, 529–552 (1992)
27. Kaplan, A., Tichatschke, R.: Proximal interior point method for convex semi-infinite programming. Optim. Methods Softw. **15**, 87–119 (2001)
28. Lim, A.E.B., Moore, J.B.: A path following algorithm for infinite quadratic programming on a Hilbert space. Discret. Contin. Dyn. Syst. **4**, 653–670 (1998)
29. Mangasarian, O.L., Fromovitz, S.: The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37–47 (1967)
30. Polak, E.: On the mathematical foundation of nondifferentiable optimization in engineering design. SIAM Rev. **29**, 21–89 (1987)
31. Polak, E.: Optimization. Algorithms and Consistent Approximations. Springer, Berlin (1997)
32. Reemtsen, R., Görner, S.: Numerical methods for semi-infinite programming: a survey. In: Reemtsen, R., Rückmann, J.-J. (eds.) Semi-Infinite Programming, pp. 195–275. Kluwer, Boston (1998)
33. Rückmann, J.-J.: Topological stability of feasible sets in semi-infinite optimization: a tutorial, In: Fiacco, A. (ed.) Mathematical programming with data perturbations. 17th symposium, George Washington University, Washington, DC, USA, May 1995. Lect. Notes Pure Appl. Math. vol. 195, pp. 339–361. Marcel Dekker, New York (1997)
34. Stein, O.: On parametric semi-infinite optimization. Thesis, Shaker, Aachen (1997)
35. Stein, O.: On constraint qualifications in non-smooth optimization. J. Optim. Theory Appl. **121**, 647–671 (2004)
36. Wetterling, W.: Definitheitsbedingungen für relative Extrema bei Optimierungs- und Approximationsaufgaben. Numerische Mathematik **15**, 122–136 (1970)
37. Zwier, G.: Structural analysis in semi-infinite programming. Thesis, University of Twente (1987)