J Glob Optim (2008) 41:335–350
DOI 10.1007/s10898-007-9231-4

Smoothing by mollifiers. Part II: nonlinear optimization

Hubertus Th. Jongen · Oliver Stein

Received: 7 December 2006 / Accepted: 5 September 2007 / Published online: 10 October 2007

Abstract This article complements the paper (Jongen, Stein, Smoothing by mollifers part I: semi-infinite optimization J Glob Optim doi:10.1007/s10898-007-9232-3), where we showed 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. In the special case of nonlinear programming this function is constructed as the mollification of the finite min-function which describes the feasible set. In the present article we treat the correspondences between Karush–Kuhn–Tucker points of the original and the smoothed problem, and between their associated Morse indices.

Keywords Nonlinear optimization · Smoothing · Mollifier · Stationarity · Morse index

AMS subject Classifications 2000 90C31 · 90C30 · 57R12

1 Introduction

We consider the constrained nonlinear programming problem

$$P: \qquad \min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad g_i(x) \geq 0, \quad i \in I,$$

with objective function $f \in C^2(\mathbb{R}^n, \mathbb{R})$, constraint functions $g_i \in C^2(\mathbb{R}^n, \mathbb{R})$, $i \in I$, and a finite index set $I = \{1, \ldots, p\}$ with $p \in \mathbb{N}$. We denote the feasible set of P by $M = \{x \in \mathbb{R}^n |\, g_i(x) \geq 0,\ i \in I\}$.

In [7] we show that a nonempty and compact feasible set M of a *semi-infinite* program can be approximated arbitrarily well by a level set of a single smooth function with certain

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

regularity properties. This function is constructed by using a so-called mollifier. Moreover, a correspondence between Karush–Kuhn–Tucker points of the original and the smoothed problem, along with their Morse indices, enables us to prove the connectedness of the so-called min-max digraph for semi-infinite problems.

On the one hand the result about Morse indices has a very elaborate proof, and moreover it is basically related to finite, and not semi-infinite programming. Thus, we treat it in the present, separate article. We also reformulate the results from [7] about smoothing by mollifiers for the present setting of nonlinear programming.

A different smoothing procedure for nonlinear programming problems is given in [6]. There the main idea is to use the logarithmic barrier approach to approximate the finitely many inequality constraints $g_i(x) \geq 0,\ i \in I$, by one smooth and nondegenerate constraint $\sum_{i\in I} \ln(g_i(x)) \geq \ln(\varepsilon)$ for $\varepsilon > 0$. A similar approach is taken in [4] to smooth finite maximum functions. In [7] we explain why several obvious generalizations of this approach to semi-infinite programming are not successful. This motivates our investigation of smoothing by mollifiers for semi-infinite as well as finite optimization problems.

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 [2].

2 Preliminaries

2.1 Regularity concepts

At a feasible point $\bar{x} \in M$ the *Mangasarian-Fromovitz Constraint Qualification* (*MFCQ*, [8]) is said to hold if there exists some vector $d \in \mathbb{R}^n$ with

$$Dg_i(\bar{x})\,d \;>\; 0, \quad i \in I_0(\bar{x}). \tag{2.1}$$

Here $I_0(\bar{x}) = \{i \in I |\ g_i(\bar{x}) = 0\}$ is the active index set at $\bar{x}$, and $Dg_i(\bar{x})$ denotes the row vector of partial derivatives of the function g_i at $\bar{x}$. We will abbreviate the corresponding column vector of partial derivatives with $\nabla g_i(\bar{x})$. Furthermore, sometimes we will write $g_{I_0(\bar{x})}$ for the column vector of functions $g_i,\ i \in I_0(\bar{x})$.

The (stronger) *Linear Independence Constraint Qualification (LICQ)* is satisfied at $\bar{x} \in M$ if the vectors $Dg_i(\bar{x}),\ i \in I_0(\bar{x})$, are linearly independent. Note that LICQ implies $|I_0(\bar{x})| \leq n$.

A point $\bar{x} \in M$ with LICQ is called a *critical point* for P if there exist real numbers $\bar{\lambda}_i,\ i \in I_0(\bar{x})$, (*Lagrange multipliers*) such that

$$Df(\bar{x}) \;=\; \sum_{i\in I_0(\bar{x})} \bar{\lambda}_i\, Dg_i(\bar{x}). \tag{2.2}$$

A critical point is called *Karush–Kuhn–Tucker point (KKT-point)* if all multipliers in (2.2) are nonnegative, $\bar{\lambda}_i \geq 0,\ i \in I_0(\bar{x})$, and a KKT point is called *nondegenerate* if the following two conditions hold:

(SCS) : $\bar{\lambda}_i > 0,\ i \in I_0(\bar{x})$ (*strict complementary slackness*),
(SOSC) : $D^2L(\bar{x})|_{T_{\bar{x}}M}$ is nonsingular (*second order sufficiency condition*).

The matrix D^2L stands for the Hessian of the Lagrange function L,

$$L(x) \;=\; f(x) \;-\; \sum_{i\in I_0(\bar{x})} \bar{\lambda}_i\, g_i(x), \tag{2.3}$$

and $T_{\bar{x}}M$ denotes the tangent space of M at $\bar{x}$,

$$T_{\bar{x}}M \;=\; \{d \in \mathbb{R}^n \,|\; Dg_i(\bar{x})d = 0,\; i \in I_0(\bar{x})\}. \tag{2.4}$$

Condition SOSC means that the matrix $V^\top D^2L(\bar{x})V$ is nonsingular, where V is some $(n, n - |I_0(\bar{x})|)$—matrix whose columns form a basis for the tangent space $T_{\bar{x}}M$. The number of negative eigenvalues of $V^\top D^2L(\bar{x})V$ is called the *Morse index* of $\bar{x}$. In particular, a nondegenerate KKT point is a local minimizer for P iff its Morse index vanishes.

2.2 Mollifiers

With the Euclidean norm $||\cdot||_2$ on $\mathbb{R}^n$, the *standard mollifier* (cf., e.g., [3]) is the $C^\infty-$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 set

$$\eta_\varepsilon(x) \;=\; \frac{1}{\varepsilon^n}\,\eta\left(\frac{x}{\varepsilon}\right). \tag{2.5}$$

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 $\overline{A}$ denotes the topological closure of a set A.

Definition 2.1 *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 2.2 **([3])**

(a) For all $\varepsilon > 0$, the $\varepsilon-$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 [3].

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 P is nonempty and compact.*

Assumption 3.2 *The MFCQ holds everywhere in M.*

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

With

$$G(x) = \min_{i \in I} g_i(x)$$

the feasible set M can obviously be described by the single but arguably nonsmooth constraint $G(x) \geq 0$. Our smoothing approach is based on the mollification of G,

$$G^\varepsilon = \eta_\varepsilon * G = \eta_\varepsilon * \min_{i \in I} g_i(\cdot).$$

In view of Theorem 2.2 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

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

should be strongly related to M and P, 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 $\varepsilon > 0$, MFCQ 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 P 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 P^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 P^ε for sufficiently small $\varepsilon > 0$. Moreover, x^ε is nondegenerate, and the Morse index of $\bar{x}$ in P and the Morse index of x^ε in P^ε coincide.

Corollary 3.8 *Let $\varepsilon > 0$ be sufficiently small. Then Assumptions 3.1, 3.2 and 3.3 do not only hold for P, but also for P^ε, and the local minimizers of P^ε are located arbitrarily close to those of P. Moreover, if M is connected, so is M^ε.*

Theorems 3.4, 3.5, 3.6, and parts (a) and (b) of Theorem 3.7 are shown in the separate article [7], even for the more general setting of semi-infinite programming. Corollary 3.8 is an immediate consequence of Theorems 3.4, 3.5, 3.6 and 3.7.

Thus, here we can concentrate on the proof of Theorem 3.7(c). We break down the proof into several steps which are treated in the following subsections.

3.2 Formulae for G^ε, its gradient, and its Hessian

From now on we fix some nondegenerate KKT point $\bar{x}$ of P. For $\bar{x}$ from the interior of M, Theorem 3.4 implies that $\bar{x}$ is also an interior point of M^ε for sufficiently small $\varepsilon > 0$, so that the assertion of Theorem 3.7(c) trivially holds for $\bar{x}$ with $x^\varepsilon \equiv \bar{x}$. Note that in this case $\nabla f(\bar{x})$ vanishes.

Hence, in the following let $\bar{x}$ be a boundary point of M, so that the active index set $I_0(\bar{x})$ of $\bar{x}$ is necessarily nonempty. Since all our considerations will concern M only locally around $\bar{x}$, without loss of generality we may assume that $I_0(\bar{x})$ coincides with I, that is, we have $g(\bar{x}) = 0$ for the vector function $g := g_I = g_{I_0(\bar{x})}$. Furthermore, we will abbreviate $I_{(i)} := I \setminus \{i\}$ for $i \in I$. Note that the nondegeneracy of the KKT point implies $\nabla f(\bar{x}) \neq 0$.

The following sets will play a major role in the sequel:

$$S^i = \{x \in \mathbb{R}^n |\ g_i(x) < g_j(x),\ j \in I_{(i)}\}, \quad i \in I, \tag{3.1}$$

$$\Sigma^{ij} = \{x \in \mathbb{R}^n |\ g_i(x) = g_j(x)\},\ i, j \in I, \quad i \neq j. \tag{3.2}$$

In fact, to take first and second derivatives of

$$G^\varepsilon(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x - z) \min_{i \in I} g_i(z)\, dz$$

at $\bar{x}$, we rewrite $G^\varepsilon(x)$ as a sum of integrals with smooth integrands on the subdomains S^i, $i \in I$:

Lemma 3.9

(a) For all $i \in I$ and all $x \in S^i$ we have $G(x) = g_i(x)$.

(b) There exists some $R > 0$ with $B(\bar{x}, R) \subset \bigcup_{i \in I} \overline{S^i}$, where $\overline{S^i}$ denotes the topological closure of S^i, $i \in I$.

(c) For all $x \in \mathbb{R}^n$ and $\varepsilon > 0$ with $||x - \bar{x}|| + \varepsilon < R$ we have

$$G^\varepsilon(x) = \sum_{i \in I} \int_{S^i} \eta_\varepsilon(x - z) g_i(z)\, dz.$$

Proof The assertion of part (a) is clear from the definitions of S^i, $i \in I$, and G.

To see part (b), consider the sets

$$\widetilde{S^i} = \{x \in \mathbb{R}^n |\ g_i(x) \leq g_j(x),\ j \in I_{(i)}\}, \quad i \in I,$$

which all contain $\bar{x}$. It is not hard to see that LICQ at $\bar{x}$ in M implies LICQ at $\bar{x}$ in $\widetilde{S^i}$ for each $i \in I$. By continuity, for each $i \in I$ there exists some $R_i > 0$ such that LICQ holds everywhere in $\widetilde{S^i} \cap B(\bar{x}, R_i)$. Under LICQ it is well known that $\widetilde{S^i}$ and $\overline{S^i}$ coincide, that is, we have $\widetilde{S^i} \cap B(\bar{x}, R_i) = \overline{S^i} \cap B(\bar{x}, R_i)$ for each $i \in I$. With $R = \min_{i \in I} R_i$ consider any point $x \in B(\bar{x}, R)$. Then for each $i \in I$ with $g_i(x) = G(x)$ we have $x \in \overline{S^i}$, that is, $x \in \bigcup_{i \in I} \overline{S^i}$.

For part (c) recall that the support of η_ε is $\overline{B(0, \varepsilon)}$, so that we may write

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

Since the assumption $||x - \bar{x}|| + \varepsilon < R$ implies $B(x, \varepsilon) \subset B(\bar{x}, R)$, and we have $B(\bar{x}, R) \subset \bigcup_{i \in I} \overline{S^i}$ by part b), we may replace the domain of integration by

$$B(x, \varepsilon) \cap \bigcup_{i \in I} \overline{S^i} = \bigcup_{i \in I} \left(B(x, \varepsilon) \cap \overline{S^i} \right)$$

to obtain

$$G^\varepsilon(x) \;=\; \sum_{i\in I}\int_{B(x,\varepsilon)\cap\overline{S^i}} \eta_\varepsilon(x-z)G(z)\,dz.$$

As the integrand vanishes outside of $B(x,\varepsilon)$, and in view of LICQ in S^i, we may further replace the domain of integration in each summand by S^i, so that part a) yields the assertion. □

In the following we put

$$\Delta_{ij}(x) \;=\; \nabla g_i(x) - \nabla g_j(x) \tag{3.3}$$

for $i,\, j \in I,\ i \neq j$, and $x \in \mathbb{R}^n$.

Proposition 3.10 *For all $x \in \mathbb{R}^n$ and $\varepsilon > 0$ with $\|x-\bar{x}\| + \varepsilon < R$ the gradient and Hessian of G^ε at x satisfy*

$$DG^\varepsilon(x) \;=\; \sum_{i\in I}\int_{S^i} \eta_\varepsilon(x-z)Dg_i(z)\,dz$$

and

$$D^2G^\varepsilon(x) \;=\; \sum_{i\in I}\int_{S^i} \eta_\varepsilon(x-z)D^2g_i(z)\,dz - \sum_{i\neq j\in I}\int_{\Sigma^{ij}} \eta_\varepsilon(x-z)\,\frac{\Delta_{ij}(z)\Delta_{ij}(z)^\top}{\|\Delta_{ij}(z)\|}\,dz,$$

where the last term is a boundary integral.

Proof Let us take derivatives on both sides of the formula in Lemma 3.9(c). Since the integrands are smooth and have compact support, we may as well integrate the derivatives of the integrands. Using z as the differentiation variable instead of x, as well as partial integration, we arrive at

$$\begin{aligned}
DG^\varepsilon(x) &= \sum_{i\in I}\int_{S^i} D_x[\eta_\varepsilon(x-z)]g_i(z)\,dz\\
&= \sum_{i\in I}\int_{S^i\cap B(x,\varepsilon)} (-D_z)[\eta_\varepsilon(x-z)]g_i(z)\,dz\\
&= \sum_{i\in I}\int_{S^i} \eta_\varepsilon(x-z)Dg_i(z)\,dz\\
&\quad - \sum_{i\in I}\int_{\partial(S^i\cap B(x,\varepsilon))} \eta_\varepsilon(x-z)g_i(z)n(z)^\top\,dz
\end{aligned}$$

where the last term is a boundary integral, $\partial(S^i\cap B(x,\varepsilon))$ stands for the topological boundary of $S^i\cap B(x,\varepsilon),\ i\in I$, and $n(z)$ denotes the outward pointing normal to $\partial(S^i\cap B(x,\varepsilon))$ at z. To show the assertion for $DG^\varepsilon(x)$ we have to prove that

$$\sum_{i\in I}\int_{\partial(S^i\cap B(x,\varepsilon))} \eta_\varepsilon(x-z)g_i(z)n(z)\,dz \tag{3.4}$$

vanishes. Note that in (3.4) it is sufficient to integrate over the codimension one parts of $\partial(S^i\cap B(x,\varepsilon)),\ i\in I$, since lower dimensional parts do not contribute to the value of the integral. The codimension one parts either belong to $\partial B(x,\varepsilon)$ or to some set Σ^{ij} with $i\neq j$

(compare (3.2)). For all $z \in \partial B(x, \varepsilon)$ the integrand vanishes due to $\eta_\varepsilon(x - z) = 0$. Now consider $z \in \Sigma^{ij}$ for some $i, j \in I$, $i \neq j$. As an element of $\Sigma^{ij} \cap \overline{S^i}$, the outward pointing normal at z is easily seen to be $n^{ij}(z) = \Delta_{ij}(z)/||\Delta_{ij}(z)||$ with $\Delta_{ij}(z)$ from (3.3) (note that $||\Delta_{ij}(z)|| \neq 0$ in view of LICQ at $\bar{x}$ and the choice of ε). At the same time, z is an element of $\Sigma^{ij} \cap \overline{S^j}$ with outward pointing normal $n^{ji}(z) = -n^{ij}(z)$. For the two corresponding summands in (3.4) we obtain

$$\int_{\Sigma^{ij}\cap\overline{S^i}} \eta_\varepsilon(x-z) g_i(z) n^{ij}(z)\, dz + \int_{\Sigma^{ij}\cap\overline{S^j}} \eta_\varepsilon(x-z) g_j(z)(-n^{ij}(z))\, dz$$
$$= \int_{\Sigma^{ij}} \eta_\varepsilon(x-z)(g_i(z) - g_j(z)) n^{ij}(z)\, dz \ = \ 0$$

since $g_i - g_j$ vanishes on Σ^{ij}. In this way, all summands in (3.4) cancel each other, so that the whole term vanishes. This shows the formula for $DG^\varepsilon(x)$.

To derive the formula for the Hessian $D^2G^\varepsilon(x)$ we take the Jacobian of $\nabla G^\varepsilon(x)$ and, as above, arrive at

$$D^2G^\varepsilon(x) = \sum_{i\in I} \int_{S^i} \eta_\varepsilon(x-z) D^2 g_i(z)\, dz$$
$$- \sum_{i\in I} \int_{\partial(S^i\cap B(x,\varepsilon))} \eta_\varepsilon(x-z) \nabla g_i(z) n(z)^\top\, dz.$$

For the second term we can do the same manipulations as above, except that in the end the terms

$$\int_{\Sigma^{ij}} \eta_\varepsilon(x-z)(\nabla g_i(z) - \nabla g_j(z)) n(z)^\top\, dz$$

only necessarily vanish for $\Sigma^{ij} \cap B(x, \varepsilon) = \emptyset$. Because of

$$\int_{\Sigma^{ij}} \eta_\varepsilon(x-z)(\nabla g_i(z) - \nabla g_j(z)) n(z)^\top\, dz \ = \ \int_{\Sigma^{ij}} \eta_\varepsilon(x-z)\, \frac{\Delta_{ij}(z)\Delta_{ij}(z)^\top}{||\Delta_{ij}(z)||}\, dz$$

this shows the assertion for $D^2G^\varepsilon(x)$. □

3.3 A regularization of the KKT system for P^ε and its continuous extension

We can now study the KKT system of P^ε,

$$\nabla f(x) - \mu \nabla G^\varepsilon(x) = 0 \tag{3.5}$$
$$-G^\varepsilon(x) = 0 \tag{3.6}$$

where, in view of $\nabla f(\bar{x}) \neq 0$, the inequality constraint $G^\varepsilon(x) \geq 0$ is treated as binding. Our aim is to show that this system has a unique solution for $\varepsilon > 0$ close to zero and x close to $\bar{x}$. Unfortunately, the implicit function theorem cannot be applied straightforwardly to yield this result, due to problems with the definition of the system and its Jacobian at $\varepsilon = 0$, as well as regularity issues.

It turns out that instead we may study a related system which is 'desingularized' on the normal space to M in $\bar{x}$. In fact, it is not hard to see that there exist neighborhoods U of $\bar{x}$ and V of $0 \in \mathbb{R}^p$ such that for each $c \in V$ the function G is constant on the locally defined manifold $\{x \in U |\ g_I(x) = c\}$. Hence, G^ε behaves smoothly along these manifolds locally around $\bar{x}$. On the other hand, the restriction of G to the normal spaces of these manifolds locally around $\bar{x}$ is nonsmooth, which corresponds to the fact the G^ε has to smooth the nonsmooth fiber of M. This means that singularities of G^ε only occur in normal directions and, moreover, that one has to expect that the approximation of the nonsmooth function implies blow ups in the curvature of the approximating smoothings.

It thus suffices to study how the location of the KKT point in the normal space is affected by small perturbations of ε. For this reason, in the following we will investigate the case $p = n$ without loss of generality, that is, the number of possible active constraints is maximal under LICQ, the tangent space $T_{\bar{x}}M$ becomes trivial, and the normal space is $\mathbb{R}^n$.

With $\mathbb{R}_+$ denoting the set of positive real numbers, we define the function

$$H : \mathbb{R}_+ \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}, \ (\varepsilon, d, \mu) \mapsto \begin{pmatrix} \nabla f(\bar{x} + \varepsilon d) - \mu \nabla G^\varepsilon(\bar{x} + \varepsilon d) \\ -\frac{1}{\varepsilon} G^\varepsilon(\bar{x} + \varepsilon d) \end{pmatrix}$$

and study its zero set. For all $\varepsilon > 0$, since the mapping $x(d) = \bar{x} + \varepsilon d$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ is bijective, the system $H(\varepsilon, d, \mu) = 0$ is equivalent to (3.5), (3.6). In particular, if for given $\varepsilon > 0$ we find d^ε and μ^ε with $H(\varepsilon, d^\varepsilon, \mu^\varepsilon) = 0$, then $(x^\varepsilon, \mu^\varepsilon)$ with $x^\varepsilon = \bar{x} + \varepsilon d^\varepsilon$ solves (3.5), (3.6).

For all $\varepsilon > 0$ the function H is obviously continuously differentiable with respect to (d, μ), and its Jacobian is

$$\begin{aligned} J(\varepsilon, d, \mu) &:= D_{(d,\mu)} H(\varepsilon, d, \mu) \\ &= \begin{pmatrix} \varepsilon(D^2 f(\bar{x} + \varepsilon d) - \mu D^2 G^\varepsilon(\bar{x} + \varepsilon d)) & -\nabla G^\varepsilon(\bar{x} + \varepsilon d) \\ -DG^\varepsilon(\bar{x} + \varepsilon d) & 0 \end{pmatrix}. \end{aligned}$$

Since we plan to apply the implicit function theorem to $H(\varepsilon, d, \mu) = 0$ around some point with $\bar{\varepsilon} = 0$, we must first make sure that H and J can be extended to continuous functions on all of $U \times \mathbb{R}^n \times \mathbb{R}$, with some open neighborhood U of 0. For this reason we give alternative formulae for G^ε, DG^ε, and D^2G^ε:

Lemma 3.11 *The following formulae hold for all $d \in \mathbb{R}^n$ and $\varepsilon > 0$ with $\varepsilon(||d|| + 1) < R$:*

(a) $G^\varepsilon(\bar{x} + \varepsilon d) = \sum_{i \in I} \int_{\frac{1}{\varepsilon}(S^i - \bar{x})} \eta(d - z) g_i(\bar{x} + \varepsilon z)\, dz,$

(b) $DG^\varepsilon(\bar{x} + \varepsilon d) = \sum_{i \in I} \int_{\frac{1}{\varepsilon}(S^i - \bar{x})} \eta(d - z) Dg_i(\bar{x} + \varepsilon z)\, dz,$

(c)

$$\begin{aligned} D^2 G^\varepsilon(\bar{x} + \varepsilon d) = &\sum_{i \in I} \int_{\frac{1}{\varepsilon}(S^i - \bar{x})} \eta(d - z) D^2 g_i(\bar{x} + \varepsilon z)\, dz \\ &- \frac{1}{\varepsilon} \sum_{i \neq j \in I} \int_{\frac{1}{\varepsilon}(\Sigma^{ij} - \bar{x})} \eta(d - z) \frac{\Delta_{ij}(\bar{x} + \varepsilon z) \Delta_{ij}(\bar{x} + \varepsilon z)^\top}{||\Delta_{ij}(\bar{x} + \varepsilon z)||}\, dz. \end{aligned}$$

Proof It is easily seen that for all $d \in \mathbb{R}^n$, $\varepsilon > 0$ with $\varepsilon(||d|| + 1) < R$ we may apply Lemma 3.9 (c) and Proposition 3.10 at the point $x = \bar{x} + \varepsilon d$.

To see part (a), we use Lemma 3.9 (c) to write

$$G^\varepsilon(\bar{x} + \varepsilon d) = \sum_{i \in I} \int_{S^i} \eta_\varepsilon(\bar{x} + \varepsilon d - z) g_i(z)\, dz$$

and transform the integrals by defining the new variable $\zeta = (z - \bar{x})/\varepsilon$ or, equivalently, $z = \bar{x} + \varepsilon \zeta$. Under this change of coordinates, for each $i \in I$ the corresponding integral becomes

$$\int_{\frac{1}{\varepsilon}(S^i-\bar{x})} \eta_\varepsilon(\varepsilon d - \varepsilon\zeta) g_i(\bar{x}+\varepsilon\zeta)\,\varepsilon^n d\zeta \;=\; \int_{\frac{1}{\varepsilon}(S^i-\bar{x})} \eta(d-z) g_i(\bar{x}+\varepsilon z)\,dz,$$

where we used the definition of η_ε from (2.5). This proves the assertion of part (a).

Parts (b) and (c) are shown along the same lines, using Proposition 3.10. The factor $1/\varepsilon$ in front of the second term in the right hand side of part (c) is owed to the fact that the second term is a boundary integral, so that the deformation by the change of coordinates is only ε^{n-1}. □

The possibility to continuously extend H and J to arguments with $\varepsilon = 0$ depends on the behavior of

$$\begin{aligned}
\Gamma^1(\varepsilon, d) &= \tfrac{1}{\varepsilon} G^\varepsilon(\bar{x}+\varepsilon d),\\
\Gamma^2(\varepsilon, d) &= DG^\varepsilon(\bar{x}+\varepsilon d),\\
\Gamma^3(\varepsilon, d) &= \varepsilon D^2 G^\varepsilon(\bar{x}+\varepsilon d)
\end{aligned}$$

for $\varepsilon \searrow 0$. For the corresponding results we will need the linearizations of the sets from (3.1) and (3.2),

$$L^i = \{\xi \in \mathbb{R}^n \mid Dg_i(\bar{x})\xi < Dg_j(\bar{x})\xi,\ j \in I_{(i)}\},\ i \in I, \tag{3.7}$$

$$\Lambda^{ij} = \{\xi \in \mathbb{R}^n \mid Dg_i(\bar{x})\xi = Dg_j(\bar{x})\xi\},\ i, j \in I,\ i \neq j. \tag{3.8}$$

Lemma 3.12 *For all $d \in \mathbb{R}^n$ define*

$$\begin{aligned}
\Gamma^1(0, d) &= \sum_{i\in I} \int_{L^i} \eta(d-z) Dg_i(\bar{x}) z\,dz,\\
\Gamma^2(0, d) &= \sum_{i\in I} \int_{L^i} \eta(d-z)\,dz\, Dg_i(\bar{x}),\\
\Gamma^3(0, d) &= -\sum_{i\neq j\in I} \int_{\Lambda^{ij}} \eta(d-z)\,dz\, \frac{\Delta_{ij}(\bar{x})\Delta_{ij}(\bar{x})^\top}{||\Delta_{ij}(\bar{x})||}\,.
\end{aligned}$$

Then we have $\lim_{\varepsilon\searrow 0} \Gamma^k(\varepsilon, d) = \Gamma^k(0, d)$ *for* $1 \le k \le 3$.

Proof Let $d \in \mathbb{R}^n$ be arbitrary. We will use Lemma 3.11 to show the assertions. First we study the behavior of the sets $\frac{1}{\varepsilon}(S^i - \bar{x})$, $i \in I$, for $\varepsilon \searrow 0$.

In fact, let $i \in I$ and $z \in \frac{1}{\varepsilon}(S^i - \bar{x})$. This means

$$g_i(\bar{x}+\varepsilon z) \;<\; g_j(\bar{x}+\varepsilon z),\ \ j \in I_{(i)}$$

or, equivalently,

$$\frac{g_i(\bar{x}+\varepsilon z) - g_i(\bar{x})}{\varepsilon} \;<\; \frac{g_j(\bar{x}+\varepsilon z) - g_j(\bar{x})}{\varepsilon},\quad j \in I_{(i)}\,.$$

Consequently, for $\varepsilon \searrow 0$ the vector z satisfies $Dg_i(\bar{x})z \le Dg_j(\bar{x})z,\ j \in I_{(i)}$, and due to LICQ, $\frac{1}{\varepsilon}(S^i - \bar{x})$ tends to L^i (compare (3.7)), up to a set of codimension one (the boundary of L^i). Analogously it is shown that the sets $\frac{1}{\varepsilon}(\Sigma^{ij} - \bar{x})$ tend to Λ^{ij} for $\varepsilon \searrow 0$.

With Lemma 3.11(a) and using the fact that the supports of the integrands are contained in the compact set $\overline{B(d,1)}$ it follows that

$$\begin{aligned}\Gamma^1(\varepsilon,d) &= \frac{1}{\varepsilon}\sum_{i\in I}\int_{\frac{1}{\varepsilon}(S^i-\bar{x})}\eta(d-z)g_i(\bar{x}+\varepsilon z)\,dz,\\ &= \sum_{i\in I}\int_{\frac{1}{\varepsilon}(S^i-\bar{x})}\eta(d-z)\frac{g_i(\bar{x}+\varepsilon z)-g_i(\bar{x})}{\varepsilon}\,dz\\ &\to \sum_{i\in I}\int_{L^i}\eta(d-z)Dg_i(\bar{x})z\,dz \;=\; \Gamma^1(0,d)\quad(\varepsilon\searrow 0),\end{aligned}$$

and with Lemma 3.11(b) that

$$\Gamma^2(\varepsilon,d) \;=\; \sum_{i\in I}\int_{\frac{1}{\varepsilon}(S^i-\bar{x})}\eta(d-z)Dg_i(\bar{x}+\varepsilon z)\,dz \;\to\; \Gamma^2(0,d)\quad(\varepsilon\searrow 0).$$

From Lemma 3.11(c) we obtain

$$\begin{aligned}\Gamma^3(\varepsilon,d) &= \varepsilon\sum_{i\in I}\int_{\frac{1}{\varepsilon}(S^i-\bar{x})}\eta(d-z)D^2g_i(\bar{x}+\varepsilon z)\,dz\\ &\quad-\sum_{i\neq j\in I}\int_{\frac{1}{\varepsilon}(\Sigma^{ij}-\bar{x})}\eta(d-z)\,\frac{\Delta_{ij}(\bar{x}+\varepsilon z)\Delta_{ij}(\bar{x}+\varepsilon z)^\top}{\|\Delta_{ij}(\bar{x}+\varepsilon z)\|}\,dz\\ &\to -\sum_{i\neq j\in I}\int_{\Lambda^{ij}}\eta(d-z)\,dz\cdot\frac{\Delta_{ij}(\bar{x})\Delta_{ij}(\bar{x})^\top}{\|\Delta_{ij}(\bar{x})\|}\quad(\varepsilon\searrow 0).\end{aligned}$$

□

Lemma 3.12 shows that the functions $\Gamma^k(|\varepsilon|,d)$, $1\leq k\leq 3$, are continuous on $U\times\mathbb{R}^n$ for some neighborhood U of 0. We may thus define the following continuous extensions of H and J on $U\times\mathbb{R}^n\times\mathbb{R}$,

$$\begin{aligned}\overline{H}(\varepsilon,d,\mu) &= \begin{pmatrix}\nabla f(\bar{x}+|\varepsilon|d)-\mu\Gamma^2(|\varepsilon|,d)^\top\\ -\Gamma^1(|\varepsilon|,d)\end{pmatrix},\\ \overline{J}(\varepsilon,d,\mu) &= \begin{pmatrix}|\varepsilon|D^2f(\bar{x}+|\varepsilon|d)-\mu\Gamma^3(|\varepsilon|,d) & -\Gamma^2(|\varepsilon|,d)^\top\\ -\Gamma^2(|\varepsilon|,d) & 0\end{pmatrix},\end{aligned}$$

with

$$\begin{aligned}\overline{H}(0,d,\mu) &= \begin{pmatrix}\nabla f(\bar{x})-\mu\Gamma^2(0,d)^\top\\ -\Gamma^1(0,d)\end{pmatrix},\\ \overline{J}(0,d,\mu) &= \begin{pmatrix}-\mu\Gamma^3(0,d) & -\Gamma^2(0,d)^\top\\ -\Gamma^2(0,d) & 0\end{pmatrix}.\end{aligned}$$

With the same techniques as above one can also show $D_d\Gamma^1(0,d) = \Gamma^2(0,d)$ and $D_d\Gamma^2(0,d)^\top = \Gamma^3(0,d)$ for all $d\in\mathbb{R}^n$, so that we have

$$D_{(d,\mu)}\overline{H}(0,d,\mu) \;=\; \overline{J}(0,d,\mu)$$

for all $(d,\mu)\in\mathbb{R}^n\times\mathbb{R}$. This means that the function $\overline{H}$ is continuous and partially continuously differentiable with respect to (d,μ) on $U\times\mathbb{R}^n\times\mathbb{R}$.

To apply the implicit function theorem, we need to find a solution $(\bar{d},\bar{\mu})$ of $\overline{H}(0,d,\mu)=0$ with nonsingular matrix $\overline{J}(0,\bar{d},\bar{\mu})$.

3.4 A solution of the regularized system

Let us start by finding a zero of the first component of $\overline{H}(0, d, \mu)$, that is, to solve

$$0 = \nabla f(\bar{x}) - \mu \Gamma^2(0, d)^\top = \nabla f(\bar{x}) - \mu \sum_{i \in I} \int_{L^i} \eta(d - z)\, dz\, \nabla g_i(\bar{x}). \tag{3.9}$$

In view of (2.2) the appearing integrals should be related to the Lagrange multipliers $\bar{\lambda}_i$, $i \in I$. In fact, due to strict complementary slackness we may set

$$\bar{\mu} = \sum_{i \in I} \bar{\lambda}_i$$

and consider the normalized vector $\bar{\lambda}/\bar{\mu}$. This vector is contained in the $(n-1)-$dimensional standard simplex

$$\mathcal{S}^{n-1} = \left\{ \lambda \in \mathbb{R}^n \,\middle|\, \lambda \geq 0,\ \sum_{i=1}^n \lambda_i = 1 \right\}.$$

The mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ with components $\int_{L^i} \eta(d - z)\, dz, i \in I$, is easily seen to satisfy $T(d) \in \mathcal{S}^{n-1}$ for all $d \in \mathbb{R}^n$. Hence, if we can show *surjectivity* of T from $\mathbb{R}^n$ to $\mathcal{S}^{n-1}$, we may choose some $\tilde{d}$ with $T(\tilde{d}) = \bar{\lambda}/\bar{\mu}$ and solve (3.9) with $(\tilde{d}, \bar{\mu})$.

Proposition 3.13 *The mapping T is surjective from $\mathbb{R}^n$ to $\mathcal{S}^{n-1}$.*

Proof We show that T is surjective from $\mathbb{R}^n$ to each stratum of $\mathcal{S}^{n-1}$, where F is a stratum of dimension $k \in \{0, \ldots, n-1\}$ iff $F = \{\lambda \in \mathcal{S}^{n-1} |\ \lambda_{K^c} > 0,\ \lambda_K = 0\}$ for some $K \subset I$ with $|K| = n - 1 - k$. Note that the topological closures of the strata are just the facets of the polyhedron $\mathcal{S}^{n-1}$.

Let F be a stratum of dimension zero, that is, F is a singleton with some unit vector e^i, $i \in I$, as the only element. Due to LICQ, the sets L^i, $i \in I$, are full-dimensional cones, so that for each $i \in I$ we can find some $d^i \in \mathbb{R}^n$ with $B(d^i, 1) \subset L^i$. It follows $T(d^i) = e^i$, $i \in I$, that is, T is surjective from $\mathbb{R}^n$ to each zero dimensional stratum of $\mathcal{S}^{n-1}$.

Next consider a stratum F of dimension one. Then its closure $\overline{F}$ coincides with the convex hull $E^{ij} := \text{conv}(e^i, e^j)$ of two different unit vectors e^i and e^j, $i, j \in I$. We denote the convex hull of their preimages d^i and d^j in $\mathbb{R}^n$ by $D^{ij} := \text{conv}(d^i, d^j)$. By moving d^i and d^j sufficiently far away from the origin in their corresponding cones L^i and L^j, it is possible to guarantee $B(\delta, 1) \subset L^i \cup L^j$ for all $\delta \in D^{ij}$. Then the continuous function T maps the line segment D^{ij} to the line segment E^{ij}, that is, $T(D^{ij}) \subset E^{ij}$. Moreover, T maps the endpoints of D^{ij} to the endpoints of E^{ij}. The intermediate value theorem now guarantees $F \subset \overline{F} = E^{ij} \subset T(D^{ij})$. This entails that T is surjective from $\mathbb{R}^n$ to each one dimensional stratum of $\mathcal{S}^{n-1}$.

For any stratum F of dimension two or higher, we cannot use the intermediate value theorem in the previous argument. Instead, we will use an argument from algebraic topology and explain the idea in detail for a two-dimensional stratum F.

In fact, let F be a stratum of dimension two. Then $\overline{F}$ coincides with the convex hull $E^{ijk} := \text{conv}(e^i, e^j, e^k)$ of three pairwise different unit vectors e^i, e^j, e^k, $i, j, k \in I$. As above we consider the convex hull $D^{ijk} := \text{conv}(d^i, d^j, d^k)$ of their preimages, and move the points d^i, d^j, d^k sufficiently far away from the origin in their corresponding cones L^i, L^j, L^k, such that $B(\delta, 1) \subset L^i \cup L^j \cup L^k$ holds for all $\delta \in D^{ijk}$. Then the continuous function T maps the 'triangle' D^{ijk} to the 'triangle' E^{ijk}.

From a topological point of view, we may identify D^{ijk} with E^{ijk} without loss of generality, so that T becomes a continuous mapping from E^{ijk} to itself that is, $T(E^{ijk}) \subset E^{ijk}$. Note that T is constant on the vertices of E^{ijk}, and that T maps each edge of E^{ijk} surjectively to itself. In particular, we have $\partial E^{ijk} \subset T(E^{ijk})$. Moreover, with the linear homotopy $h : [0,1] \times E^{ij} \to E^{ij}$, $(t,x) \mapsto tx + (1-t)T(x)$ the restriction of T to the edge E^{ij} is seen to be homotopic to the identity function. As the same holds for the other edges, the restriction of T to ∂E^{ijk} is homotopic to the identity function.

Assume that $T : E^{ijk} \to E^{ijk}$ is not surjective. Then some point λ from the interior $\mathrm{int}(E^{ijk})$ of E^{ijk} does not belong to $T(E^{ijk})$. The latter set is compact as the continuous image of a compact set, so that a whole open neighborhood $\Omega \subset \mathrm{int}(E^{ijk})$ of λ does not belong to $T(E^{ijk})$. Due to $\partial E^{ijk} \subset T(E^{ijk})$, this means that there exists a continuous mapping φ from $T(E^{ijk})$ to ∂E^{ijk} which coincides with the identity mapping on ∂E^{ijk}, that is, ∂E^{ijk} is a retract of $T(E^{ijk})$.

As a consequence, the restriction of the composition $\varphi \circ T : E^{ijk} \to \partial E^{ijk}$ to ∂E^{ijk} is homotopic to the identity function, that is, it is a weak retraction. However, if ∂E^{ijk} was a weak retract of E^{ijk}, then the one-dimensional sphere would also be a weak retract of the two-dimensional ball. This is not the case by Corollary 4 in Sect. 4.7 of [9] and its preceding remark on weak retracts.

Hence T is also surjective from D^{ijk} to E^{ijk}, so that we obtain $F \subset \overline{F} = E^{ijk} \subset T(D^{ijk})$. This completes the proof for surjectivity of T from $\mathbb{R}^n$ to each two-dimensional stratum of $\mathcal{S}^{n-1}$. Surjectivity of T from $\mathbb{R}^n$ to each higher dimensional stratum of $\mathcal{S}^{n-1}$ is shown along the same lines, using that the $(k-1)-$dimensional sphere is not a weak retract of the $k-$dimensional ball for any $k \geq 1$ ([9]). □

We choose some $\tilde{d}$ with $T(\tilde{d}) = \bar{\lambda}/\bar{\mu}$. In more explicit terms this means

$$\frac{\bar{\lambda}_i}{\sum_{i\in I} \bar{\lambda}_i} = \int_{L^i} \eta(\tilde{d} - z)\, dz\,, \quad i \in I. \tag{3.10}$$

In the next step we show that we can change $\tilde{d}$ along a certain direction so that *both* components of $\overline{H}$ vanish.

Lemma 3.14

(a) For each $i \in I$ let K_i denote the common null space of the vectors $Dg_i(\bar{x}) - Dg_j(\bar{x})$, $j \in I_{(i)}$. Then there exists some vector $w \in \mathbb{R}^n \setminus \{0\}$ such that w spans all spaces K_i, $i \in I$.

(b) With w from part (a) consider the functions

$$\gamma^1 : \mathbb{R} \to \mathbb{R}, \quad t \mapsto \Gamma^1(0, \tilde{d} + tw),$$
$$\gamma^2 : \mathbb{R} \to \mathbb{R}^n, \; t \mapsto \Gamma^2(0, \tilde{d} + tw).$$

Then γ^1 is linear and not constant, whereas γ^2 is constant.

(c) There exists some $\bar{t} \in \mathbb{R}$ such that $\bar{d} := \tilde{d} + \bar{t}w$ satisfies $\Gamma^1(0,\bar{d}) = 0$ and $\Gamma^2(0,\bar{d}) = \Gamma^2(0,\tilde{d})$.

Proof For any $i \in I$, due to LICQ the $n-1$ vectors $Dg_i(\bar{x}) - Dg_j(\bar{x})$, $j \in I_{(i)}$, are linearly independent and have, thus, a one dimensional common null space K^i. Let w^i be a basis vector of K^i. Then, up to scaling, w^i is determined by the equations

$$Dg_1(\bar{x})w = \cdots = Dg_n(\bar{x})w \tag{3.11}$$

which do not depend on i. Hence w^i can be chosen independently of $i \in I$, which shows part (a).

To see the linearity of γ^1 in part (b), note that for any $t \in \mathbb{R}$ a simple change of coordinates yields

$$\begin{aligned}\gamma^1(t) = \Gamma^1(0, \tilde{d} + tw) &= \sum_{i \in I} \int_{L^i} \eta(\tilde{d} + tw - z) Dg_i(\bar{x}) z \, dz, \\ &= \sum_{i \in I} \int_{L^i - tw} \eta(\tilde{d} - z) Dg_i(\bar{x})(z + tw) \, dz.\end{aligned}$$

From the definition of w we immediately conclude $L^i - tw = L^i$ for all $i \in I$ and arrive at

$$\begin{aligned}\gamma^1(t) &= \sum_{i \in I} \int_{L^i} \eta(\tilde{d} - z) Dg_i(\bar{x})(z + tw) \, dz \\ &= \Gamma^1(0, \tilde{d}) + t \sum_{i \in I} \int_{L^i} \eta(\tilde{d} - z) \, dz \, Dg_i(\bar{x}) w \\ &= \Gamma^1(0, \tilde{d}) + t \, Dg_1(\bar{x}) w,\end{aligned}$$

where (3.11) allows us to replace $Dg_i(\bar{x})w$ by $Dg_1(\bar{x})w$ for all $i \in I$. This shows the linearity of γ^1. If γ^1 was a constant function, we would have $Dg_1(\bar{x})w = 0$ and, in view of (3.11) also $Dg_i(\bar{x})w = 0$ for all $i \in I$. However, due to LICQ w then has to vanish, in contradiction to its definition.

With the same technique we can show

$$\gamma^2(t) = \Gamma^2(0, \tilde{d} + tw) = \Gamma^2(0, \tilde{d}) = \gamma^2(0)$$

for all $t \in \mathbb{R}$, which completes the proof of part (b).
Part (c) immediately follows from part (b) with $\bar{t} = -\Gamma^1(0, \tilde{d})/(Dg_1(\bar{x})w)$. □

With $\bar{d}$ from Lemma 3.14 (c) we finally have $\overline{H}(0, \bar{d}, \bar{\mu}) = 0$.

3.5 The implicit function

To apply the implicit function theorem, it remains to be shown that the Jacobian $\overline{J}(0, \bar{d}, \bar{\mu})$ is nonsingular.

Recall that, due to strict complementary slackness, all $\bar{\lambda}_i = \bar{\mu} \int_{L^i} \eta(\bar{d} - z) \, dz, i \in I$ are strictly positive (compare (3.10)). Consequently, the intersection of the support $\overline{B(\bar{d}, 1)}$ of $\eta(\bar{d} - z)$ with each set L^i has positive measure. Then also the intersections of $\overline{B(\bar{d}, 1)}$ with the mutual boundaries parts of L^i and L^j in Λ^{ij} have positive $(n-1)$–dimensional measure for all $i, j \in I, \; i \neq j$. This means

$$c_{ij}(\bar{d}) := \int_{\Lambda^{ij}} \eta(\bar{d} - z) \, dz > 0 \quad \text{for all} \quad i, j \in I, \; i \neq j. \tag{3.12}$$

Lemma 3.15 *The matrix* $\overline{J}(0, \bar{d}, \bar{\mu})$ *is nonsingular.*

Proof We have

$$\overline{J}(0, \bar{d}, \bar{\mu}) = \begin{pmatrix} -\bar{\mu}\Gamma^3(0, \bar{d}) & -\Gamma^2(0, \bar{d})^\top \\ -\Gamma^2(0, \bar{d}) & 0 \end{pmatrix}$$

with

$$-\bar{\mu}\Gamma^3(0, \bar{d}) = \bar{\mu} \sum_{i \neq j \in I} c_{ij}(\bar{d}) \frac{\Delta_{ij}(\bar{x})\Delta_{ij}(\bar{x})^\top}{||\Delta_{ij}(\bar{x})||}$$

and

$$\Gamma^2(0, \bar{d})^\top = \sum_{i \in I} \int_{L^i} \eta(\bar{d} - z)\, dz\, \nabla g_i(\bar{x}) = \sum_{i \in I} \bar{\lambda}_i \nabla g_i(\bar{x}) = \nabla f(\bar{x}).$$

Hence by the structure lemma (see, e.g. [5, Theorem 2.3.2]), if Y is some $(n, n-1)-$matrix whose columns form a basis of $\ker(Df(\bar{x}))$, then $\overline{J}(0, \bar{d}, \bar{\mu})$ is nonsingular if and only if

$$\sum_{i \neq j \in I} c_{ij}(\bar{d}) \frac{Y^\top \Delta_{ij}(\bar{x})\, \Delta_{ij}(\bar{x})^\top Y}{||\Delta_{ij}(\bar{x})||}$$

is nonsingular (where we also used $\bar{\mu} > 0$). In view of (3.12) the latter matrix clearly is positive semi-definite. In the following we will show that it is actually positive definite. In fact, let $\theta \in \mathbb{R}^{n-1}$ with

$$0 = \theta^\top \left(\sum_{i \neq j \in I} c_{ij}(\bar{d}) \frac{Y^\top \Delta_{ij}(\bar{x})\, \Delta_{ij}(\bar{x})^\top Y}{||\Delta_{ij}(\bar{x})||} \right) \theta = \sum_{i \neq j \in I} c_{ij}(\bar{d}) \frac{(\Delta_{ij}(\bar{x})^\top Y\theta)^2}{||\Delta_{ij}(\bar{x})||}.$$

Due to (3.12), for all $i, j \in I$, $i \neq j$, we obtain $\Delta_{ij}(\bar{x})^\top Y\theta = 0$, that is

$$Dg_1(\bar{x})Y\theta = \cdots = Dg_n(\bar{x})Y\theta. \tag{3.13}$$

Thus, by the definition of Y and (2.2) we also have

$$0 = Df(\bar{x})Y\theta = \sum_{i \in I} \bar{\lambda}_i Dg_i(\bar{x})Y\theta = \bar{\mu} Dg_j(\bar{x})Y\theta$$

for all $j \in I$. It follows $Y\theta \in \ker(Dg(\bar{x})) = \{0\}$ and, hence, $\theta = 0$. This shows the assertion. □

We may now apply the version of the implicit function theorem from [1] to obtain a continuous function $(d(\varepsilon), \mu(\varepsilon))$ with $(d(0), \mu(0)) = (\bar{d}, \bar{\mu})$, defined locally around $\bar{\varepsilon} = 0$, such that $(d(\varepsilon), \mu(\varepsilon))$ is the locally unique solution of $\overline{H}(\varepsilon, d, \mu) = 0$. As explained at the beginning of Section 3.3, the continuous function $(x(\varepsilon), \mu(\varepsilon))$ with $x(\varepsilon) = \bar{x} + \varepsilon d(\varepsilon)$ satisfies $(x(0), \mu(0)) = (\bar{x}, \bar{\mu})$, and it is the locally unique solution of (3.5), (3.6).

In other words, for sufficiently small $\varepsilon > 0$, around $\bar{x}$ the point $x^\varepsilon := x(\varepsilon)$ is the locally unique KKT point of P^ε, and the corresponding Lagrange multiplier is $\mu(\varepsilon)$. This shows the first assertion of Theorem 3.7 (c).

3.6 The Morse index

Finally we prove the second assertion of Theorem 3.7 (c). Note that in the case $p = n$ the Morse index of $\bar{x}$ vanishes. For continuity reasons, LICQ holds at $x(\varepsilon)$ for sufficiently small $\varepsilon > 0$. Due to $\bar{\mu} = \sum_{i \in I} \bar{\lambda}_i > 0$ and continuity of the function μ, for sufficiently small $\varepsilon > 0$ we also have $\mu(\varepsilon) > 0$, that is, strict complementary slackness is also satisfied at $x(\varepsilon)$.

It remains to study the nonsingularity of the Jacobian of (3.5), (3.6) at $x(\varepsilon)$ and the Morse index of $x(\varepsilon)$. Note that this is *not* the Jacobian $\overline{J}$ of the regularized system. Instead, for

$\varepsilon > 0$ the Jacobian of (3.5), (3.6) at the KKT point is

$$\begin{pmatrix} D^2 f(x(\varepsilon)) - \mu(\varepsilon) D^2 G^\varepsilon(x(\varepsilon)) & -\nabla G^\varepsilon(x(\varepsilon)) \\ -DG^\varepsilon(x(\varepsilon)) & 0 \end{pmatrix}.$$

Due to LICQ at $x(\varepsilon)$ the Jacobian is nonsingular if and only if the restriction of the Hessian to the kernel of $DG^\varepsilon(x(\varepsilon))$,

$$\left(D^2 f(x(\varepsilon)) - \mu(\varepsilon) D^2 G^\varepsilon(x(\varepsilon))\right)|_{\ker(DG^\varepsilon(x(\varepsilon)))}, \tag{3.14}$$

is, and the number of negative eigenvalues of the latter matrix is the Morse index of $x(\varepsilon)$. Since $(x(\varepsilon), \mu(\varepsilon))$ satisfies (3.5), and because of $\mu(\varepsilon) > 0$ for sufficiently small $\varepsilon > 0$, the matrix in (3.14) has the same inertia as

$$\left(D^2 f(x(\varepsilon)) - \mu(\varepsilon) D^2 G^\varepsilon(x(\varepsilon))\right)|_{\ker(Df(x(\varepsilon)))}. \tag{3.15}$$

As ∇f is a continuous function, for sufficiently small $\varepsilon > 0$ the inertia of the matrix in (3.15) also coincides with that of

$$\left(D^2 f(x(\varepsilon)) - \mu(\varepsilon) D^2 G^\varepsilon(x(\varepsilon))\right)|_{\ker(Df(\bar{x}))}, \tag{3.16}$$

provided that we can show nonsingularity of this matrix. Let Y be some $(n, n-1)-$matrix whose columns form a basis of $\ker(Df(\bar{x}))$. Then the matrix in (3.16) can be written as

$$Y^\top \left(D^2 f(x(\varepsilon)) - \mu(\varepsilon) D^2 G^\varepsilon(x(\varepsilon))\right) Y. \tag{3.17}$$

By Lemma 3.11, and using $x(\varepsilon) = \bar{x} + \varepsilon d(\varepsilon)$, we have

$$\begin{aligned} D^2 G^\varepsilon(x(\varepsilon)) = & \sum_{i\in I} \int_{\frac{1}{\varepsilon}(S^i - \bar{x})} \eta(d(\varepsilon) - z) D^2 g_i(\bar{x} + \varepsilon z)\, dz \\ & - \frac{1}{\varepsilon} \sum_{i \neq j \in I} \int_{\frac{1}{\varepsilon}(\Sigma^{ij} - \bar{x})} \eta(d(\varepsilon) - z) \frac{\Delta_{ij}(\bar{x} + \varepsilon z)\Delta_{ij}(\bar{x} + \varepsilon z)^\top}{||\Delta_{ij}(\bar{x} + \varepsilon z)||}\, dz, \end{aligned}$$

so that we may write

$$Y^\top \left(D^2 f(x(\varepsilon)) - \mu(\varepsilon) D^2 G^\varepsilon(x(\varepsilon))\right) Y \;=\; A(\varepsilon) + \frac{1}{\varepsilon} B(\varepsilon)$$

with

$$\begin{aligned} A(\varepsilon) &= Y^\top \left(D^2 f(x(\varepsilon)) - \mu(\varepsilon) \sum_{i\in I} \int_{\frac{1}{\varepsilon}(S^i - \bar{x})} \eta(d(\varepsilon) - z) D^2 g_i(\bar{x} + \varepsilon z)\, dz \right) Y, \\ B(\varepsilon) &= \mu(\varepsilon)\, Y^\top \left(\sum_{i \neq j \in I} \int_{\frac{1}{\varepsilon}(\Sigma^{ij} - \bar{x})} \eta(d(\varepsilon) - z) \frac{\Delta_{ij}(\bar{x} + \varepsilon z)\Delta_{ij}(\bar{x} + \varepsilon z)^\top}{||\Delta_{ij}(\bar{x} + \varepsilon z)||}\, dz \right) Y. \end{aligned}$$

With the techniques from the proof of Lemma 3.12, and using $\lim_{\varepsilon \searrow 0} d(\varepsilon) = \bar{d}$ as well as $T(\bar{d}) = \bar{\lambda}/\bar{\mu}$, we can show

$$\begin{aligned} A(0) \;&:=\; \lim_{\varepsilon\searrow 0} A(\varepsilon) = Y^\top \left(D^2 f(\bar{x}) - \sum_{i\in I} \bar{\lambda}_i D^2 g_i(\bar{x}) \right) Y, \\ B(0) \;&:=\; \lim_{\varepsilon\searrow 0} B(\varepsilon) = \bar{\mu}\, Y^\top \left(\sum_{i\neq j\in I} c_{ij}(\bar{d}) \frac{\Delta_{ij}(\bar{x})\Delta_{ij}(\bar{x})^\top}{||\Delta_{ij}(\bar{x})||} \right) Y \end{aligned}$$

with c_{ij} from (3.12). In the proof of Lemma 3.15 we have seen that $B(0)$ is positive definite, so that for sufficiently small $\varepsilon > 0$ the eigenvalues of $B(\varepsilon)$ are positive and bounded away from zero. At the same time, the eigenvalues of $A(\varepsilon)$ are contained in some bounded set, so that for all sufficiently small $\varepsilon > 0$ the matrix $A(\varepsilon) + \frac{1}{\varepsilon}B(\varepsilon)$ is positive definite.

This means that for all sufficiently small $\varepsilon > 0$ the matrix in (3.14) is positive definite, so that the Jacobian of (3.5), (3.6) at $x(\varepsilon)$ is nonsingular, and the Morse index of $x(\varepsilon)$ vanishes. This completes the proof of Theorem 3.7 (c).

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 support through a Heisenberg grant of the *Deutsche Forschungsgemeinschaft*.

References

1. Deimling, K.: Nichtlineare Gleichungen und Abbildungsgrade. Springer, Heidelberg (1974)
2. 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)
3. Evans L.C.: Partial Differential Equations, American Mathematical Society, Providence, Rhode Island (1998)
4. Guerra Vasquez, F., Günzel, H., Jongen, H.Th.: On logarithmic smoothing of the maximum function. An. Oper. Res. **101**, 209–220 (2001)
5. Jongen, H.Th., Meer, K., Triesch, E.: Optimization Theory. Kluwer, Boston (2004)
6. 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)
7. Jongen, H.Th., Stein O.: Smoothing by Mollifiers. Part I: Semi-infinite Optimization. J Glob Optim. doi:10.1007/s10898-007-9232-3
8. 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)
9. Spanier, E.H.: Algebraic Topology. McGraw Hill, New York (1966)