ATTRACTION OF NEWTON METHOD TO CRITICAL LAGRANGE MULTIPLIERS: FULLY QUADRATIC CASE*

A. F. Izmailov† and E. I. Uskov‡

December 28, 2012

ABSTRACT

All previously known results concerned with attraction of Newton-type iterations for optimality systems to critical Lagrange multipliers were a posteriori by nature: they were showing that in case of convergence, the dual limit is in a sense unlikely to be noncritical. This paper suggests the first a priori result in this direction, showing that critical multipliers actually serve as attractors: for a fully quadratic optimization problem with equality constraints, under certain reasonable assumptions we establish actual local convergence to a critical multiplier starting from a "dense" set around the given critical multiplier. This is an important step forward in understanding the attraction phenomenon.

Key words: quadratic optimization problem, Lagrange optimality system, critical Lagrange multiplier, Newton method.
AMS subject classifications. 49M05, 49M15.

* This research is supported by the Russian Foundation for Basic Research Grants 10-01-00251, 12-01-31025, and 12-01-33023.
† Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, IO Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: `izmaf@ccas.ru`
‡ Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, IO Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: `ydoom@narod.ru`

1 Introduction

Critical multipliers, first introduced in [3], are special Lagrange multipliers usually forming a thin subset in the set of all multipliers when the latter set is not a singleton. In particular, such multipliers necessarily violate the second-order sufficient optimality conditions. By now, there exists a convincing theoretical and numerical evidence of the following striking phenomena: dual sequences generated by Newton-type methods for optimality systems have a strong tendency to converge to critical multipliers when the latter exist, and moreover, this is precisely the reason for the lack of superlinear convergence rate, which is typical for problems with degenerate constraints; see [4, 5, 6].

However, the existing theoretical results of this kind are far from giving a complete picture. First, all these results are "negative" by nature: they attempt to give a characterization of what would have happened in the case of convergence to a noncritical multiplier, showing that this scenario is in a sense unlikely. Clearly, this analysis must be complemented by results of a "positive" nature, demonstrating that the set of critical multipliers is indeed an attractor in some sense. Second, the exiting results rely on some questionable assumptions, and perhaps the most questionable one is asymptotic stabilization of the primal directions generated by Newtonian subproblems; see [4, Proposition 1], [6, Remark 1]. Obtaining the first result on actual local convergence to a critical multiplier, and avoiding undesirable assumptions, are the main goals of this work.

2 Preliminaries

In this paper we confine ourselves to the following fully quadratic equality-constrained problem, serving as a model setting capturing the most intrinsic consequences of constraints degeneracy:

$$\text{minimize} \quad \frac{1}{2}\langle Ax,\, x\rangle \quad \text{subject to} \quad \frac{1}{2}B[x,\, x] = 0, \tag{2.1}$$

where $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix and $B\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^l$ is a symmetric bilinear mapping, that is, $B[x,\, x] = (\langle B_1x,\, x\rangle,\, \ldots,\, \langle B_lx,\, x\rangle)$, where $B_1,\, \ldots,\, B_l \in \mathbb{R}^{n\times n}$ are symmetric matrices. Extension of the results presented below to more general settings (with linear terms and/or higher-order terms) might be possible using the Liapunov–Schmidt procedure (see, e.g., [1, Ch. VII]), and this will be the subject of our future research. We note, however, that this extension will certainly be concerned with technical complications, and even in the purely quadratic setting adopted in this paper, the required analysis is already quite involved.

For any $\lambda \in \mathbb{R}^l$ define the symmetric matrix $H(\lambda) = A + \lambda B$, where $\lambda B = \sum_{i=1}^{l} \lambda_i B_i$ Stationary points of problem (2.1) and associated Lagrange multipliers are characterized by the Lagrange optimality system

$$H(\lambda)x = 0, \quad \frac{1}{2}B[x,\, x] = 0 \tag{2.2}$$

with respect to $(x,\, \lambda) \in \mathbb{R}^m \times \mathbb{R}^l$. Our stationary point of interest is $\bar{x} = 0$; this point is indeed always stationary, and the set of associated Lagrange multipliers is the entire $\mathbb{R}^l$.

According to the general definition in [3], in this case critical multipliers are those satisfying $\det H(\lambda) = 0$.

Iteration of the Newton method for system (2.2) is defined as follows: for a given primal-dual iterate $(x^k,\, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^l$, the next iterate $(x^{k+1},\, \lambda^{k+1})$ is supposed to satisfy

$$\begin{pmatrix} H(\lambda^k) & (B[x^k])^{\mathrm{T}} \\ B[x^k] & 0 \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} = - \begin{pmatrix} H(\lambda^k) x^k \\ \frac{1}{2} B[x^k,\, x^k] \end{pmatrix}, \tag{2.3}$$

where for a given $\xi \in \mathbb{R}^n$, the matrix $B[\xi] \in \mathbb{R}^{l \times n}$ is defined by $B[\xi]x = B[\xi,\, x]$.

We start with three examples demonstrating possible scenarios of dual behavior of the specified iterations when critical multipliers do exist. The first simple example demonstrates the behavior which is typical: dual sequences steadily converge to a critical multiplier.

Example 2.1 Consider the case when $n = l = 1$:

$$\text{minimize} \quad ax^2 \quad \text{subject to} \quad bx^2 = 0,$$

where $a, b \in \mathbb{R} \setminus \{0\}$. For this problem $H(\lambda) = 2(a + \lambda b)$, and hence, $\bar{\lambda} = -a/b$ is the unique critical multiplier. For each k, equality in (2.3) takes the form:

$$(a + \lambda^k b)(x^{k+1} - x^k) + bx^k(\lambda^{k+1} - \lambda^k) = -(a + \lambda^k b)x^k, \quad 2bx^k(x^{k+1} - x^k) = -b(x^k)^2$$

Assuming that $x^k \neq 0$, we obtain from the second equality that $x^{k+1} = x^k/2$, and then the first equality gives $\lambda^{k+1} = (\lambda^k - a/b)/2$. Hence, $\lambda^{k+1} + a/b = (\lambda^k + a/b)/2$. Therefore, if $x^0 \neq 0$, then $x^k \neq 0$ for all k, and the sequence $\{(x^k,\, \lambda^k)\}$ is well-defined by (2.3) and converges linearly to $(0,\, \bar{\lambda})$. Moreover, if $\lambda^0 \neq \bar{\lambda}$, then the convergence rates of both $\{x^k\}$ and $\{\lambda^k\}$ are linear.

According to the existing numerical experience, dual behavior different from convergence to a critical multiplier is completely atypical, but still possible, as we demonstrate by the next (much more involved) example.

Example 2.2 Consider the problem

$$\text{minimize} \quad 2x_1^2 + 2x_2^2 + 5x_2x_3 \quad \text{subject to} \quad x_1x_2 + x_1x_3 + x_2x_3 = 0.$$

For this problem

$$A = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 5 \\ 0 & 5 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad H(\lambda) = \begin{pmatrix} 4 & \lambda & \lambda \\ \lambda & 4 & 5 + \lambda \\ \lambda & 5 + \lambda & 0 \end{pmatrix}$$

(here by B we mean the symmetric matrix of the corresponding quadratic form). In particular, $\det H(\lambda) = 2(\lambda - 5)(\lambda^2 + 6\lambda + 10)$, and the unique critical multiplier is $\bar{\lambda} = 5$.

It can be easily verified that if $\lambda^k \neq 5$, then the matrix in the left-hand side of (2.3) is nonsingular if, and only if,

$$\langle Bx^k,\, \hat{H}(\lambda^k) Bx^k \rangle \neq 0. \tag{2.4}$$

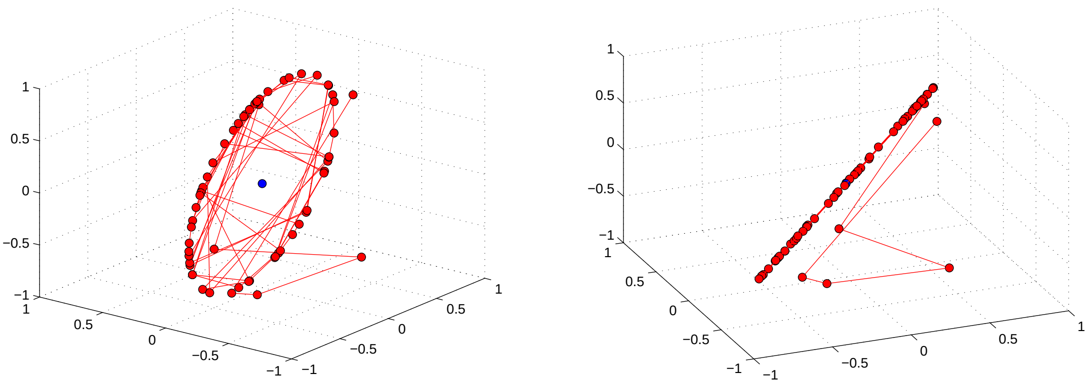


Figure 1: Sequence $\{x^k/\|x^k\|\}$ for Example 2.2, for $x^0 = (1,\, -1,\, 2)$, $\lambda^0 = 1$.

Here and throughout by $\hat{H}$ we denote the cofactor matrix of a square matrix H. Furthermore, if $\lambda_k \neq 5$ and (2.4) holds, then (2.3) gives

$$x^{k+1} = \frac{1}{2}\,\frac{\langle Bx^k,\, x^k\rangle}{\langle Bx^k,\, \hat{H}(\lambda_k)Bx^k\rangle}\hat{H}(\lambda_k)Bx^k, \quad \lambda^{k+1} = \lambda^k - \frac{\det H(\lambda^k)}{2}\,\frac{\langle Bx^k,\, x^k\rangle}{\langle Bx^k,\, \hat{H}(\lambda^k)Bx^k\rangle}. \quad (2.5)$$

Now consider the linear subspace $\mathcal{L} = \ker \hat{H}(\bar{\lambda})B \subset \mathbb{R}^n$. One can easily derive that $\mathcal{L}$ is spanned by $a^1 = (1,\, -1,\, 3)$ and $a^2 = (3,\, 2,\, 2)$, and hence $\dim \mathcal{L} = 2$. Observe that for any $\lambda \neq 5$ it holds that $\hat{H}(\bar{\lambda})B\hat{H}(\lambda)B = 2(\lambda^2 + 6\lambda + 10)\hat{H}(\bar{\lambda})B$, and hence, $\hat{H}(\bar{\lambda})B\hat{H}(\lambda)Ba^1 = 0$, $\hat{H}(\bar{\lambda})B\hat{H}(\lambda)Ba^2 = 0$, which further implies $\hat{H}(\lambda)B\mathcal{L} \subset \mathcal{L}$. Finally, observe that if $x^k \in \mathcal{L}$ and $\lambda^k = 5$, then the matrix in the left-hand side of (2.3) is always singular.

By the above we conclude that if $x^k \in \mathcal{L}$, $\lambda^k \neq 5$, and (2.4) holds, then x^{k+1} and λ^{k+1} are well-defined by (2.5), and $x^{k+1} \in \mathcal{L}$. Therefore, if $x^0 \in \mathcal{L}$, $\lambda^0 \neq 5$, and (2.4) holds for $k = 0$, then either $\lambda^k = 5$ or $\langle Bx^k,\, \hat{H}(\lambda^k)Bx^k\rangle = 0$ for some k (which means that the next iteration would not be well-defined), or $x^k \in \mathcal{L}$, $\lambda^k \neq 5$, and (2.4) holds (which further implies that $x^k \neq 0$) for all k.

In the latter case, define the matrix $U \in \mathbb{R}^{3\times 2}$ with columns a^1, a^2, and consider the (unique) sequence $\{z^k\} \subset \mathbb{R}^2 \setminus \{0\}$ such that $x^k = Uz^k$ for all k. It can be directly verified that

$$\hat{H}(\lambda)BU = 2(\lambda - 5)UM_1(\lambda), \quad U^{\mathrm{T}}B\hat{H}(\lambda)BU = 4(\lambda - 5)M_2(\lambda),$$

where

$$M_1(\lambda) = \begin{pmatrix} \lambda + 2 & -2 \\ 1 & \lambda + 4 \end{pmatrix}, \quad M_2(\lambda) = \begin{pmatrix} 5 - \lambda & 7\lambda + 30 \\ 7\lambda + 30 & 16\lambda + 50 \end{pmatrix}.$$

Therefore, from (2.5) we derive

$$x^{k+1} = \frac{\langle Mz^k,\, z^k\rangle}{2\langle M_2(\lambda_k)z^k,\, z^k\rangle}UM_1(\lambda_k)z^k, \quad \lambda^k - \lambda^{k+1} = ((\lambda^k)^2 + 6\lambda^k + 10)\frac{\langle Mz^k,\, z^k\rangle}{2\langle M_2(\lambda^k)z^k,\, z^k\rangle},$$

where

$$M = \frac{1}{2}U^{\mathrm{T}}BU = \begin{pmatrix} -1 & 7 \\ 7 & 16 \end{pmatrix},$$

and hence

$$z^{k+1} = \frac{\langle Mz^k,\, z^k\rangle}{2\langle M_2(\lambda_k)z^k,\, z^k\rangle} M_1(\lambda_k)z^k.$$

The above equalities imply

$$\zeta^{k+1} = \frac{M_1(\lambda^k)\zeta^k}{\|M_1(\lambda^k)\zeta^k\|}, \quad \lambda^k - \lambda^{k+1} = ((\lambda^k)^2 + 6\lambda^k + 10)\frac{\langle M\zeta^k,\, \zeta^k\rangle}{2\langle M_2(\lambda^k)\zeta^k,\, \zeta^k\rangle}, \tag{2.6}$$

where $\zeta^k = z^k/\|z^k\|$ (recall that $z^k \neq 0$ for all k).

Suppose that $\{\lambda^k\}$ converges to some $\lambda^* \in \mathbb{R}$. The second equality in (2.6) and convergence of $\{\lambda^k\}$ evidently imply that $\langle M\zeta^k,\, \zeta^k\rangle$ tends to 0. Hence, each limit point $\bar{\zeta}$ of the sequence $\{\zeta^k\}$ must satisfy $\langle M\bar{\zeta},\, \bar{\zeta}\rangle = 0$. However, the first equality in (2.6) implies that $\hat{\zeta} = M_1(\lambda^*)\bar{\zeta}/\|M_1(\lambda^*)\bar{\zeta}\|$ is also a limit point of $\{\zeta^k\}$, and therefore $\langle M\hat{\zeta},\, \hat{\zeta}\rangle = 0$, which implies $\langle M_1(\lambda^*)^{\mathrm{T}}MM_1(\lambda^*)\bar{\zeta},\, \bar{\zeta}\rangle = 0$. Furthermore, it can be verified that $M_1(\lambda)^{\mathrm{T}}MM_1(\lambda) = 130(\lambda+3)\hat{M} + (\lambda^2 - 10\lambda - 40)M$, where

$$\hat{M} = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix}.$$

Hence, if $\lambda^* \neq -3$, then $\bar{\zeta}$ must also satisfy $\langle \hat{M}\bar{\zeta},\, \bar{\zeta}\rangle = 0$, which cannot hold simultaneously with $\langle M\bar{\zeta},\, \bar{\zeta}\rangle = 0$. Suppose that $\lambda^* = -3$. From (2.6) and from the equality $M_2(\lambda) = MM_1(\lambda)$ we get

$$\lambda^{k+1} - \lambda^{k+2} = ((\lambda^{k+1})^2 + 6\lambda^{k+1} + 10)\frac{\langle MM_1(\lambda^k)\zeta^k,\, M_1(\lambda^k)\zeta^k\rangle}{2\langle MM_1(\lambda^k)\zeta^k,\, M_1(\lambda^{k+1})M_1(\lambda^k)\zeta^k\rangle}.$$

From the second equality in (2.6) we also derive

$$\lambda^{k+1} + 3 = \frac{\langle MM_1(\lambda^k)\zeta^k,\, M_1(\lambda^k)\zeta^k\rangle}{2\langle M\zeta^k,\, M_1(\lambda^k)\zeta^k\rangle}.$$

Dividing the next-to-the-last equality by the last one, we now obtain

$$1 - \frac{\lambda^{k+2}+3}{\lambda^{k+1}+3} = ((\lambda^{k+1})^2 + 6\lambda^{k+1} + 10)\frac{\langle M\zeta^k,\, M_1(\lambda^k)\zeta^k\rangle}{\langle MM_1(\lambda^k)\zeta^k,\, M_1(\lambda^{k+1})M_1(\lambda^k)\zeta^k\rangle}.$$

Observe that there can be at most four different accumulation points of the sequence $\{\zeta^k\}$ (since these points must satisfy $\langle M\bar{\zeta},\, \bar{\zeta}\rangle = 0$), and it can be directly verified that for any accumulation point $\bar{\zeta}$ the value $\langle M\bar{\zeta},\, M_1(\lambda^*)\bar{\zeta}\rangle / \langle MM_1(\lambda^*)\bar{\zeta},\, M_1(\lambda^*)M_1(\lambda^*)\bar{\zeta}\rangle$ is negative, which implies that $(\lambda^{k+2}+3)/(\lambda^{k+1}+3) > 1 + \varepsilon$ for some $\varepsilon > 0$ for all k large enough. Evidently, the latter implies that λ^{k+2} will be further from -3 than λ^{k+1}, and hence, $\{\lambda^k\}$ cannot converge to -3.

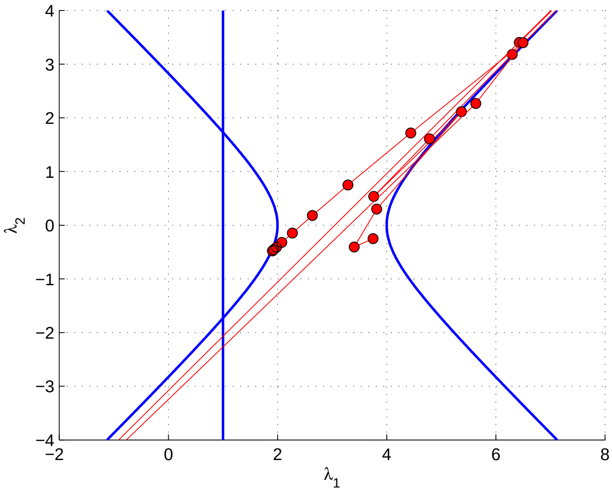


Figure 2: Dual sequence for Example 2.3.

Therefore, we conclude that if $x^k \in \mathcal{L}$ for some k, then the sequence $\{\lambda^k\}$ cannot converge (and in particular, cannot converge to the critical multiplier $\bar{\lambda} = 5$). At the same time, according to our numerical experience, this sequence usually has noncritical accumulation points, though it is not clear whether the latter can be formally proven. We further emphasize that this behavior is rather persistent in this example: even if $x^0 \notin \mathcal{L}$, the directions $x^k/\|x^k\|$ quite often tend to the subspace $\mathcal{L}$. One run of this kind is shown in Figure 1, where $x^0 = (1, -1, 2)$, $\lambda^0 = 1$. Moreover, small perturbations of the starting point do not destroy this behavior. On the other hand, convergence to critical multiplier is also a common behavior in this example, especially if λ^0 is close to $\bar{\lambda}$ (of course, in this case $x^k \notin \mathcal{L}$ for all k).

Our third example demonstrates the "mixed" behavior, which is also quite common: the dual sequence moves chaotically in the beginning but eventually the iterates enter the region of convergence to a critical multiplier.

Example 2.3 ([5], Example 2.3) Consider the problem

$$\text{minimize} \quad x_1^2 - x_2^2 + 2x_3^2 \quad \text{subject to} \quad -\frac{1}{2}x_1^2 + x_2^2 - \frac{1}{2}x_3^2 = 0, \; x_1 x_3 = 0.$$

For this problem critical multipliers are those $\lambda \in \mathbb{R}^2$ satisfying $\lambda_1 = 1$ or $(\lambda_1 - 3)^2 - \lambda_2^2 = 1$ (vertical line and hyperbola in Figure 4).

Dual sequence for $x^0 = (3, 2, 1)$, $\lambda^0 = (3.75, -0.25)$ is shown in Figure 2. (In [5, Example 2.3], convergence of this sequence has not been detected due to too early termination of the process.)

Section 3 contains the main result of this work and the analysis needed to justify it: under certain reasonable assumptions we establish local convergence to a critical multiplier when the process is initialized in a "dense" domain around the given critical multiplier. Section 4 gives some new insight in repulsion of Newton iterates from noncritical multipliers.

Below we use the convention that the determinant of an empty matrix, and hence the cofactor matrix of 1×1 matrix, are both equal to 1; this is needed in order to avoid the necessity of a special treatment of the case when $n = 1$. On the other hand, note that the complete result for this trivial case is obtained in Example 2.1.

For a given matrix M, by M_I we denote the submatrix with rows indexed by $i \in I$, and by $M_{I,\,J}$ we denote the submatrix with rows indexed by $i \in I$ and columns indexed by $j \in J$. E stands for the identity matrix of an appropriate size. All vector norms are Euclidian; all matrix norms are spectral (i.e., induced by Euclidian). By $B(x,\, \rho)$ we denote the closed ball centered at x of radius ρ.

3 Attraction domains of critical multipliers

Equality (2.3) is further equivalent to

$$H(\lambda^k)x^{k+1} + (B[x^k])^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) = 0, \quad B[x^k,\, x^{k+1}] = \frac{1}{2}B[x^k,\, x^k]. \tag{3.1}$$

Assuming that $H(\lambda^k)$ is nonsingular, from the first equality we derive that

$$x^{k+1} = -(H(\lambda^k))^{-1}(B[x^k])^{\mathrm{T}}(\lambda^{k+1} - \lambda^k). \tag{3.2}$$

Substituting this into the second equality in (3.1), we obtain

$$-B[x^k](H(\lambda^k))^{-1}(B[x^k])^{\mathrm{T}}(\lambda^{k+1} - \lambda^k) = \frac{1}{2}B[x^k,\, x^k],$$

and if the matrix $B[x^k](H(\lambda^k))^{-1}(B[x^k])^{\mathrm{T}}$ is nonsingular (subsuming that $x^k \neq 0$), this further implies that

$$\lambda^{k+1} - \lambda^k = -\frac{1}{2}\left(B[x^k](H(\lambda^k))^{-1}(B[x^k])^{\mathrm{T}}\right)^{-1} B[x^k,\, x^k]. \tag{3.3}$$

Throughout the paper we use the following definitions. For any $\xi \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^l$ set

$$\delta(\xi,\, \lambda) = \left(B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}\right)^{-1} B[\xi,\, \xi], \quad P(\xi,\, \lambda) = (H(\lambda))^{-1}(\delta(\xi,\, \lambda)B). \tag{3.4}$$

Then (3.2) and (3.3) imply that

$$x^{k+1} = \frac{1}{2}P(\xi^k,\, \lambda^k)x^k, \quad \lambda^{k+1} = \lambda^k - \frac{\delta(\xi^k,\, \lambda^k)}{2}, \tag{3.5}$$

where $\xi^k = x^k/\|x^k\|$.

Let $\mathcal{M}_0$ stand for the set of all critical multipliers:

$$\mathcal{M}_0 = \{\lambda \in \mathbb{R}^l \mid \Delta(\lambda) = 0\},$$

where $\Delta\colon \mathbb{R}^l \to \mathbb{R}$, $\Delta(\lambda) = \det H(\lambda)$. In this section, we consider the behavior of the Newton iteration (2.3) near some $\bar{\lambda} \in \mathcal{M}_0$ such that

$$\operatorname{corank} H(\bar{\lambda}) = 1. \tag{3.6}$$

We start with the following simple fact, which is probably rather standard, but we provide the proof for completeness.

Lemma 3.1 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$*, there exists* $\rho > 0$ *and an analytic (hence, infinitely differentiable) mapping* $\xi\colon B(\bar{\lambda}, \rho) \to \mathbb{R}^n$ *such that* $\ker H(\lambda)$ *is spanned by* $\xi(\lambda)$ *for all* $\lambda \in \mathcal{M}_0 \cap B(\bar{\lambda}, \rho)$*, and* $\|\xi(\lambda)\| = 1$ *for all* $\lambda \in B(\bar{\lambda}, \rho)$*.*

Proof. Fix any $\bar{\xi} \in \ker H(\bar{\lambda})$ such that $\|\bar{\xi}\| = 1$. Consider the parametric linear system

$$H(\lambda)\xi - \tau\bar{\xi} = 0, \quad \langle \bar{\xi}, \xi \rangle = 1 \tag{3.7}$$

with respect to $(\xi, \tau) \in \mathbb{R}^n \times \mathbb{R}$, where $\lambda \in \mathbb{R}^l$ is playing the role of a parameter. Obviously, if $\lambda = \bar{\lambda}$, then this system has a solution $(\bar{\xi}, 0)$, and the matrix of this system is nonsingular. Indeed, if

$$H(\bar{\lambda})\xi - \tau\bar{\xi} = 0, \quad \langle \bar{\xi}, \xi \rangle = 0 \tag{3.8}$$

holds for some (ξ, τ), then, multiplying the first equality by $\bar{\xi}$, we obtain

$$0 = \langle H(\bar{\lambda})\xi, \bar{\xi} \rangle - \tau\|\bar{\xi}\|^2 = \langle \xi, H(\bar{\lambda})\bar{\xi} \rangle - \tau = -\tau,$$

and hence, $\tau = 0$. Therefore, by the first equality in (3.8), $\xi \in \ker H(\bar{\lambda})$, and according to (3.6), the second equality in (3.8) may hold only provided $\xi = 0$.

For any $\lambda \in \mathbb{R}^l$ close enough to $\bar{\lambda}$, the matrix of system (3.7) is nonsingular as well, and hence, this system has the unique solution $(\xi(\lambda), \tau(\lambda))$.

Furthermore, by the definition of $\mathcal{M}_0$, for any $\lambda \in \mathcal{M}_0$, the first equation in (3.7) has a solution of the form $(\xi, 0)$ with some $\xi \in \mathbb{R}^n \setminus \{0\}$. Observe that if λ is close enough to $\bar{\lambda}$, then $\langle \bar{\xi}, \xi \rangle \neq 0$. Indeed, the contrary would mean the existence of sequences $\{\lambda^k\} \subset \mathcal{M}_0$ and $\{\xi^k\} \subset \mathbb{R}^l \setminus \{0\}$ such that $\{\lambda^k\} \to \bar{\lambda}$, $\xi^k \in \ker H(\lambda^k)$, and $\langle \bar{\xi}, \xi^k \rangle = 0$, for all k. Then for any accumulation point of $\tilde{\xi} \in \mathbb{R}^n$ of the sequence $\{\xi^k/\|\xi^k\|\}$ it holds that $\tilde{\xi} \in \ker H(\bar{\lambda})$, $\|\tilde{\xi}\| = 1$, and $\langle \bar{\xi}, \tilde{\xi} \rangle = 0$, which contradicts (3.6). We conclude that $(\xi/\langle \bar{\xi}, \xi \rangle, 0)$ satisfies both equations in (3.7).

Therefore, for all $\lambda \in \mathcal{M}_0$ close enough to $\bar{\lambda}$, for the unique solution $(\xi(\lambda), \tau(\lambda))$ of system (3.7) it necessarily holds that $\tau(\lambda) = 0$, which implies that $\xi(\lambda) \in \ker H(\lambda)$.

Finally, by the classical implicit function theorem applied to system (3.7) at $(\xi, \tau) = (\bar{\xi}, 0)$ for the base parameter value $\bar{\lambda}$ it follows that $(\xi(\cdot), \tau(\cdot))$ is analytic in some neighborhood of $\bar{\lambda}$. In particular, $\|\xi(\lambda)\| \to \|\bar{\xi}\| = 1$ as $\lambda \to \bar{\lambda}$. Therefore, replacing $\xi(\cdot)$ by $\xi(\cdot)/\|\xi(\cdot)\|$, we obtain the mapping which is well-defined and analytic near $\bar{\lambda}$, and satisfying $H(\lambda)\xi(\lambda) = 0$ for all $\lambda \in \mathcal{M}_0$ close enough to $\bar{\lambda}$, and $\|\xi(\lambda)\| = 1$ for all λ close enough to $\bar{\lambda}$. It remains to observe that $\dim\ker H(\lambda) = 1$ for all such λ (e.g., by the continuity of eigenvalues of symmetric matrices). Therefore, the constructed mapping $\xi(\cdot)$ possesses all the needed properties. ∎

Lemma 3.2 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$*, there exist* $\rho > 0$ *and an analytic function* $t\colon B(\bar{\lambda}, \rho) \to \mathbb{R} \setminus \{0\}$ *such that for the mapping* $\xi(\cdot)$ *defined according to Lemma 3.1 it holds that*

$$\hat{H}(\lambda) = t(\lambda)\xi(\lambda)(\xi(\lambda))^{\mathrm{T}} \quad \forall\, \lambda \in \mathcal{M}_0 \cap B(\bar{\lambda}, \rho), \tag{3.9}$$

where $\xi(\lambda)$ *is considered as a column vector.*

Proof. For any square matrix H it holds that $\hat{H}H = H\hat{H} = E \det H$. Hence, for all $\lambda \in \mathcal{M}_0$ we have equality $H(\lambda)\hat{H}(\lambda) = 0$, and therefore,

$$\operatorname{im} \hat{H}(\lambda) \subset \ker H(\lambda), \tag{3.10}$$

further implying that $\operatorname{rank} \hat{H}(\lambda) \leq 1$ provided λ is close enough to $\bar{\lambda}$, since $\dim \ker H(\lambda) = 1$ for such λ. Moreover, the latter implies that $H(\lambda)$ has some nonzero $(n-1) \times (n-1)$ minor. Therefore, $\hat{H}(\lambda)$ has a nonzero entry, implying that $\operatorname{rank} \hat{H}(\lambda) = 1$. Combining the latter with (3.10), we conclude that for all $\lambda \in \mathcal{M}_0$ close enough to $\bar{\lambda}$ it holds that

$$\operatorname{im} \hat{H}(\lambda) = \ker H(\lambda), \tag{3.11}$$

and hence, taking into account the definition of $\xi(\cdot)$, we conclude that $\hat{H}(\lambda) = \xi(\lambda)(\zeta(\lambda))^{\mathrm{T}}$ with some $\zeta(\lambda) \in \mathbb{R}^n$. Since $\hat{H}(\lambda)$ is symmetric, the latter implies the equality $\hat{H}(\lambda) = \zeta(\lambda)(\xi(\lambda))^{\mathrm{T}}$, and hence, according to (3.11), $\zeta(\lambda) \in \ker H(\lambda)$. Therefore, $\zeta(\lambda) = t(\lambda)\xi(\lambda)$, with some real $t(\lambda)$, further implying that $\hat{H}(\lambda) = t(\lambda)\xi(\lambda)(\xi(\lambda))^{\mathrm{T}}$, and hence,

$$t(\lambda) = t(\lambda)\|\xi(\lambda)\|^2 = \langle \hat{H}(\lambda)\xi(\lambda),\, \xi(\lambda)\rangle,$$

and (3.9) holds with the function $t\colon B(\bar{\lambda}, \rho) \to \mathbb{R}$ defined by the last relation for a sufficiently small $\rho > 0$.

To complete the proof, it remains to observe that t is analytic in $B(\bar{\lambda}, \rho)$ (since $\xi(\cdot)$ possesses this property, and $\hat{H}(\cdot)$ has polynomial entries), and $t(\bar{\lambda}) \neq 0$ (since otherwise, $\hat{H}(\bar{\lambda}) = 0$, which would be in a contradiction with (3.6)). ∎

Lemma 3.3 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$*, there exists* $\rho > 0$ *such that for the mapping* $\xi(\cdot)$ *defined according to Lemma 3.1, and for the function* $t(\cdot)$ *defined according to Lemma 3.2, it holds that*

$$\Delta'(\lambda) = t(\lambda)B[\xi(\lambda),\, \xi(\lambda)] \quad \forall\, \lambda \in \mathcal{M}_0 \cap B(\bar{\lambda}, \rho). \tag{3.12}$$

Proof. Again employing the equality $H\hat{H} = E \det H$ (which holds for any square matrix H), we observe that for any $\lambda \in \mathcal{M}_0 \cap B(\bar{\lambda}, \rho)$ with a sufficiently small $\rho > 0$, and for any $\delta \in \mathbb{R}^l$ it holds that

$$\begin{aligned}
\Delta(\lambda + \delta) &= \|\xi(\lambda)\|^2 \Delta(\lambda + \delta) \\
&= \big\langle \hat{H}(\lambda + \delta)\xi(\lambda),\, H(\lambda + \delta)\xi(\lambda)\big\rangle \\
&= \big\langle \hat{H}(\lambda + \delta)\xi(\lambda),\, H(\lambda)\xi(\lambda) + (H(\lambda + \delta) - H(\lambda))\xi(\lambda)\big\rangle \\
&= \big\langle \hat{H}(\lambda + \delta)\xi(\lambda),\, (B[\xi(\lambda)])^{\mathrm{T}}\delta\big\rangle,
\end{aligned}$$

where the definitions of $H(\cdot)$ and $\xi(\cdot)$ were employed. Therefore,

$$\begin{aligned} &\Delta(\lambda+\delta)-\Delta(\lambda) \\ -\big\langle t(\lambda)B[\xi(\lambda),\,\xi(\lambda)],\,\delta\big\rangle &= \big\langle \hat{H}(\lambda+\delta)\xi(\lambda),\,(B[\xi(\lambda)])^{\mathrm{T}}\delta\big\rangle-\big\langle t(\lambda)B[\xi(\lambda),\,\xi(\lambda)],\,\delta\big\rangle \\ &= \langle B[\xi(\lambda),\,\hat{H}(\lambda+\delta)\xi(\lambda)-t(\lambda)\xi(\lambda)],\,\delta\rangle. \end{aligned} \tag{3.13}$$

According to (3.9),

$$\hat{H}(\lambda+\delta)\xi(\lambda)-t(\lambda)\xi(\lambda)\to t(\lambda)(\xi(\lambda)\|\xi(\lambda))\|^2-\xi(\lambda))=0 \text{ as } \delta\to 0,$$

and hence, the right-hand side in (3.13) is $o(\|\delta\|)$, implying (3.12). ∎

Lemma 3.4 *For any matrices $P,\,Q\in\mathbb{R}^{n\times s}$ and $H\in\mathbb{R}^{n\times n}$, if $s\le n$ and $\det H\neq 0$, then*

$$\det(P^{\mathrm{T}}H^{-1}Q)\det H=\sum_{I\in\mathcal{I}_s}\det S(P,\,Q,\,H;\,I), \tag{3.14}$$

where $\mathcal{I}_s$ is the set of all subsets of $\{1,\,\ldots,\,n\}$ whose cardinality is equal to s, and for any $I\in\mathcal{I}_s$, the matrix $S(P,\,Q,\,H;\,I)\in\mathbb{R}^{n\times n}$ has the rows

$$(S(P,\,Q,\,H;\,I))_i=\begin{cases}(PQ^{\mathrm{T}})_i & \text{if } i\in I,\\ H_i & \text{if } i\in\{1,\,\ldots,\,n\}\setminus I.\end{cases}$$

Proof. If $s=n$, equality (3.14) is evidently satisfied since

$$\det(P^{\mathrm{T}}H^{-1}Q)\det H=\det P\det Q=\det(PQ^{\mathrm{T}}),$$

and the only set in $\mathcal{I}_s$ is the entire $\{1,\,\ldots,\,n\}$.

Now let $s<n$. Applying twice the Cauchy–Binet formula, we then obtain

$$\det(P^{\mathrm{T}}H^{-1}Q)=\sum_{I\in\mathcal{I}_s}\det P_I\det(H^{-1}Q^{\mathrm{T}})_I=\sum_{I,\,J\in\mathcal{I}_s}\det P_I\det(H^{-1})_{I,\,J}\det Q_J.$$

Furthermore, by the Schur complement formula (see, e.g., [7, p. 41]) it can be easily seen that

$$\det(H^{-1})_{I,\,J}=(-1)^{\sigma(I)+\sigma(J)}\frac{\det H_{\hat{I},\,\hat{J}}}{\det H},$$

where

$$\sigma(I)=\sum_{i\in I}i,\quad \hat{I}=\{1,\,\ldots,\,n\}\setminus I,\quad \hat{J}=\{1,\,\ldots,\,n\}\setminus J.$$

Therefore,

$$\begin{aligned}\det(P^{\mathrm{T}}H^{-1}Q)\det H &= \sum_{I,\,J\in\mathcal{I}_s}(-1)^{\sigma(I)+\sigma(J)}\det P_I\det H_{\hat{I},\,\hat{J}}\det Q_J \\ &= \sum_{I,\,J\in\mathcal{I}_s}(-1)^{\sigma(I)+\sigma(J)}\det(PQ^{\mathrm{T}})_{I,\,J}\det H_{\hat{I},\,\hat{J}} \\ &= \sum_{I\in\mathcal{I}_s}\det S(P,\,Q,\,H;\,I),\end{aligned}$$

where the last equality follows from the Laplace expansion of determinant. ■

Using the notation introduced in Lemma 3.4, define the function $R\colon \mathbb{R}^{l\times n} \times \mathbb{R}^{l\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}$,

$$R(M,\, N,\, H) = \sum_{I\in\mathcal{I}_l} \det S(M^{\mathrm{T}},\, N^{\mathrm{T}},\, H;\, I).$$

Evidently, this function is polynomial with respect to entries of M, N and H. For the next lemma, we need the following assumption:

$$R\left(B[\bar{\xi}],\, B[\bar{\xi}],\, H(\bar{\lambda})\right) \neq 0. \tag{3.15}$$

It can be easily seen that this property implies, in particular, that

$$\operatorname{rank} B[\bar{\xi}] = l. \tag{3.16}$$

Lemma 3.5 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$*, let* $\bar{\xi} \in \ker H(\bar{\lambda})$ *be such that* $\|\bar{\xi}\| = 1$*, and assume that* (3.15) *holds.*

Then there exists $\rho > 0$ *such that the matrix* $B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}$ *is nonsingular for all* $\xi \in B(\bar{\xi},\, \rho)$ *and all* $\lambda \in B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0$*, and moreover, there exists an analytic mapping* $M\colon B(\bar{\xi},\, \rho) \times B(\bar{\lambda},\, \rho) \to \mathbb{R}^{l\times l}$ *such that*

$$M(\xi,\, \lambda) = \left(B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}\right)^{-1} \quad \forall\, \xi \in B(\bar{\xi},\, \rho),\ \forall\, \lambda \in B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0.$$

Proof. Assuming that $\rho > 0$ is small enough, for $\xi \in B(\bar{\xi},\, \rho)$ and $\lambda \in B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0$ nonsingularity of $B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}$ readily follows from Lemma 3.4 and assumption (3.15), by continuity of R.

Suppose first that $l \geq 2$. In addition to R, define the function $\tilde{R}\colon \mathbb{R}^{(l-1)\times n} \times \mathbb{R}^{(l-1)\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}$,

$$\tilde{R}(\tilde{M},\, \tilde{N},\, H) = \sum_{I\in\mathcal{I}_{l-1}} \det S(\tilde{M}^{\mathrm{T}},\, \tilde{N}^{\mathrm{T}},\, H;\, I).$$

Evidently, this function is also polynomial with respect to the entries of $\tilde{M}$, $\tilde{N}$ and H.

Set $I = \{1,\, \ldots,\, l\}$. For any indices $i,\, j \in I$, employing the Cramer's rule for the inverse matrix, we obtain

$$\begin{aligned}\left(\left(B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}\right)^{-1}\right)_{ij} &= (-1)^{i+j}\frac{\det\left((B[\xi])_{I\setminus\{j\}}(H(\lambda))^{-1}((B[\xi])_{I\setminus\{i\}})^{\mathrm{T}}\right)}{\det\left(B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}\right)}\\ &= (-1)^{i+j}\frac{\det\left((B[\xi])_{I\setminus\{j\}}(H(\lambda))^{-1}((B[\xi])_{I\setminus\{i\}})^{\mathrm{T}}\right)\Delta(\lambda)}{\det\left(B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}\right)\Delta(\lambda)}\\ &= (-1)^{i+j}\frac{\tilde{R}\left((B[\xi])_{I\setminus\{j\}},\, (B[\xi])_{I\setminus\{j\}},\, H(\lambda)\right)}{R(B[\xi],\, B[\xi],\, H(\lambda))},\end{aligned}$$

where the last equality is by Lemma 3.4 applied to both the numerator and the denominator.

By the assumption (3.15) and by continuity of R, for any $(\xi,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$ close enough to $(\bar{\xi},\, \bar{\lambda})$ it holds that $R(B[\xi],\, B[\xi],\, H(\lambda)) \neq 0$, and we can define the matrix $M(\xi,\, \lambda) \in \mathbb{R}^{l\times l}$ by setting

$$(M(\xi,\, \lambda))_{ij} = (-1)^{i+j} \frac{\tilde{R}\left((B[\xi])_{I\setminus\{j\}},\, (B[\xi])_{I\setminus\{i\}},\, H(\lambda)\right)}{R(B[\xi],\, B[\xi],\, H(\lambda))}, \quad i,\, j \in I.$$

Since R and $\tilde{R}$ are polynomial, we conclude that the mapping M defined this way is analytic near $(\bar{\xi},\, \bar{\lambda})$, and hence, possesses all the needed properties.

Now suppose that $l = 1$. In this case

$$\begin{aligned} \left(B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}\right)^{-1} &= \frac{1}{B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}} \\ &= \frac{\Delta(\lambda)}{\det\left(B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}\right)\Delta(\lambda)} \\ &= \frac{\Delta(\lambda)}{R(B[\xi],\, B[\xi],\, H(\lambda))}, \end{aligned} \tag{3.17}$$

and therefore, defining $M\colon \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$,

$$M(\xi,\, \lambda) = \frac{\Delta(\lambda)}{R(B[\xi],\, B[\xi],\, H(\lambda))}$$

and by the same same reasoning as above, we conclude that M is analytic and possesses all the needed properties. ■

Lemma 3.6 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$, *let* (3.15) *hold for* $\bar{\xi} = \xi(\bar{\lambda})$, *where the mapping* $\xi(\cdot)$ *is defined according to Lemma 3.1.*

Then there exists $\rho > 0$ *such that*

$$\delta(\xi(\tilde{\lambda}),\, \lambda) = \lambda - \tilde{\lambda} \quad \forall\, \lambda \in B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0,\ \forall\, \tilde{\lambda} \in \mathcal{M}_0 \cap B(\bar{\lambda},\, \rho),$$

where the left-hand side is well-defined by the first relation in (3.4).

Proof. Assuming that $\rho > 0$ is small enough, for any $\tilde{\lambda} \in \mathcal{M}_0 \cap B(\bar{\lambda},\, \rho)$ close enough to $\bar{\lambda}$, set $\tilde{\xi} = \xi(\tilde{\lambda})$. Then, by continuity of $\xi(\cdot)$, and by Lemma 3.5, $\delta(\tilde{\xi},\, \lambda)$ is well-defined by the first relation in (3.4), and it is the unique solution of the linear system

$$B[\tilde{\xi}](H(\lambda))^{-1}(\delta B)\tilde{\xi} = B[\tilde{\xi},\, \tilde{\xi}] \tag{3.18}$$

with respect to $\delta \in \mathbb{R}^l$. Note that

$$\tilde{\xi} = (H(\lambda))^{-1}H(\lambda)\tilde{\xi} = (H(\lambda))^{-1}(H(\tilde{\lambda}) + (\lambda - \tilde{\lambda})B)\tilde{\xi} = (H(\lambda))^{-1}((\lambda - \tilde{\lambda})B)\tilde{\xi},$$

immediately implying that $\delta = \lambda - \tilde{\lambda}$ satisfies (3.18). Hence, $\delta(\tilde{\xi},\, \lambda) = \lambda - \tilde{\lambda}$, which is the needed equality. ■

Corollary 3.1 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$*, let* (3.15) *hold for* $\bar{\xi} = \xi(\bar{\lambda})$*, where the mapping* $\xi(\cdot)$ *is defined according to Lemma 3.1.*

Then there exist $\rho > 0$ *and* $\ell > 0$ *such that*

$$\|\delta(\xi, \lambda) - (\lambda - \tilde{\lambda})\| \leq \ell\|\xi - \xi(\tilde{\lambda})\| \quad \forall\, \xi \in B(\bar{\xi}, \rho),\, \forall\, \lambda \in B(\bar{\lambda}, \rho) \setminus \mathcal{M}_0,\, \forall\, \tilde{\lambda} \in \mathcal{M}_0 \cap B(\bar{\lambda}, \rho), \tag{3.19}$$

where $\delta(\xi, \lambda)$ *is well-defined by the first relation in* (3.4).

Proof. According to Lemma 3.5, there exists $\rho > 0$ such that the mapping $\hat{\delta}\colon B(\bar{\xi}, \rho) \times B(\bar{\lambda}, \rho) \to \mathbb{R}^l$, $\hat{\delta}(\xi, \lambda) = M(\xi, \lambda)B[\xi, \xi]$, is well-defined and Lipschitz-continuous with some constant $\ell > 0$, and according to (3.4), $\hat{\delta}(\xi, \lambda) = \delta(\xi, \lambda)$ for all $(\xi, \lambda) \in B(\bar{\xi}, \rho) \times (B(\bar{\lambda}, \rho) \setminus \mathcal{M}_0)$. Therefore, reducing ρ if necessary, for such ξ and λ, and for $\tilde{\lambda} \in \mathcal{M}_0 \cap B(\bar{\lambda}, \rho)$, by Lemma 3.6 we obtain that

$$\|\delta(\xi, \lambda) - (\lambda - \tilde{\lambda})\| = \|\hat{\delta}(\xi, \lambda) - \hat{\delta}(\xi(\tilde{\lambda}), \lambda)\| \leq \ell\|\xi - \xi(\tilde{\lambda})\|.$$

∎

Lemma 3.7 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$*, there exist* $\rho > 0$ *and* $\gamma > 0$ *such that*

$$\|H(\tilde{\lambda})\xi\| \geq \gamma\|\xi - \xi(\tilde{\lambda})\| \quad \forall\, \xi \in B(\bar{\xi}, \rho) \text{ such that } \|\xi\| = 1,\, \forall\, \tilde{\lambda} \in \mathcal{M}_0 \cap B(\bar{\lambda}, \rho),$$

where the mapping $\xi(\cdot)$ *is defined according to Lemma 3.1.*

Proof. Suppose the contrary: there exist sequences $\{\varepsilon_k\}$ of positive reals, $\{\xi^k\} \subset \mathbb{R}^n$, and $\{\tilde{\lambda}^k\} \subset \mathcal{M}_0$, such that $\varepsilon_k \to 0$, $\{\xi^k\} \to \bar{\xi}$, $\{\tilde{\lambda}^k\} \to \bar{\lambda}$, and for all k the element $\xi(\tilde{\lambda}^k)$ is well-defined, $\|\xi^k\| = 1$, and $\|H(\tilde{\lambda}^k)\xi^k\| < \varepsilon_k\|\xi^k - \xi(\tilde{\lambda}^k)\|$. Therefore, $\xi^k \neq \xi(\tilde{\lambda}^k)$ for all k, and $\lim_{k\to\infty} \|H(\tilde{\lambda}^k)\xi^k\|/\|\xi^k - \xi(\tilde{\lambda}^k)\| = 0$, implying

$$\left\{ H(\tilde{\lambda}^k)\frac{\xi^k - \xi(\tilde{\lambda}^k)}{\|\xi^k - \xi(\tilde{\lambda}^k)\|} \right\} \to 0. \tag{3.20}$$

For each k, set $\zeta^k = (\xi^k - \xi(\tilde{\lambda}^k))/\|\xi^k - \xi(\tilde{\lambda}^k)\|$. Then the following chain of equalities is valid:

$$\begin{aligned} \langle \zeta^k, \xi(\tilde{\lambda}^k)\rangle^2 &= \frac{(\langle \xi^k, \xi(\tilde{\lambda}^k)\rangle - 1)^2}{\|\xi^k - \xi(\tilde{\lambda}^k)\|^2} \\ &= \frac{(\langle \xi^k, \xi(\tilde{\lambda}^k)\rangle - 1)^2}{2(1 - \langle \xi^k, \xi(\tilde{\lambda}^k)\rangle)} \\ &= \frac{1}{2}(1 - \langle \xi^k, \xi(\tilde{\lambda}^k)\rangle). \end{aligned}$$

Since both $\{\xi^k\} \to \bar{\xi}$ and $\{\xi(\tilde{\lambda}^k)\} \to \xi(\bar{\lambda}) = \bar{\xi}$, and $\|\bar{\xi}\| = 1$, this further implies that

$$\lim_{k\to\infty} \langle \zeta^k, \xi(\tilde{\lambda}^k)\rangle = 0.$$

Therefore, for any accumulation point $\bar{\zeta}$ of the bounded sequence $\{\zeta^k\}$, we have $\langle \bar{\zeta},\, \xi(\bar{\lambda})\rangle = 0$, or, to put it in other terms, $\bar{\zeta} \in (\ker H(\bar{\lambda}))^{\perp}$. On the other hand, from (3.20) we obtain $\bar{\zeta} \in \ker H(\bar{\lambda})$, which is a contradiction, since $\|\bar{\zeta}\| = 1$. ■

For the following error bound result, we further assume that

$$B[\bar{\xi},\, \bar{\xi}] \neq 0. \tag{3.21}$$

Lemma 3.8 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$, *let* $\bar{\xi} \in \ker H(\bar{\lambda})$ *be such that* $\|\bar{\xi}\| = 1$, *and assume that* (3.21) *holds.*

Then there exist $\rho > 0$ *and* $C_1,\, C_2 > 0$ *such that the following error bound is valid:*

$$C_1|\Delta(\lambda)| \leq \operatorname{dist}(\lambda,\, \mathcal{M}_0) \leq C_2|\Delta(\lambda)| \quad \forall\, \xi \in B(\bar{\xi},\, \rho),\ \forall\, \lambda \in B(\bar{\lambda},\, \rho). \tag{3.22}$$

Proof. The first estimate in (3.22) follows immediately by Lipschitz continuity on bounded sets of the polynomial function Δ. The second estimate in (3.22) follows from the property

$$\Delta'(\bar{\lambda}) \neq 0 \tag{3.23}$$

which is implied by (3.21) and by Lemma 3.3. ■

Define the set

$$\mathcal{M} = \{(\xi,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l \mid H(\lambda)\xi = 0\}.$$

Evidently,

$$\mathcal{M}_0 = \{\lambda \in \mathbb{R}^l \mid \exists\, \xi \in \mathbb{R}^n \setminus \{0\} \text{ such that } (\xi,\, \lambda) \in \mathcal{M}\}.$$

Lemma 3.9 *Assuming that* (3.6) *holds for some* $\bar{\lambda} \in \mathcal{M}_0$, *let* (3.15), (3.21) *hold for* $\bar{\xi} = \xi(\bar{\lambda})$, *where the mapping* $\xi(\cdot)$ *is defined according to Lemma 3.1.*

Then there exists $\rho > 0$ *such that for any* $\ell_1 > 0$ *there exists* $\ell_2 > 0$ *possessing the following properties: for any* $(\xi,\, \lambda) \in B(\bar{\xi},\, \rho) \times (B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0)$ *such that*

$$\frac{\|\delta(\xi,\, \lambda)\|}{|\Delta(\lambda)|} \leq \ell_1, \tag{3.24}$$

where $\delta(\xi,\, \lambda)$ *is well-defined by the first relation in* (3.4), *and for any* $\theta \in [0,\, 1]$, *it holds that*

$$\left\| \xi \cdot (\Delta'(\lambda - \theta\delta(\xi,\, \lambda)))^{\mathrm{T}} - \hat{H}(\lambda)\big(B[\xi]\big)^{\mathrm{T}} \right\| \leq \ell_2 \operatorname{dist}((\xi,\, \lambda),\, \mathcal{M}), \tag{3.25}$$

$$\left| \left\langle \Delta'(\lambda - \theta\delta(\xi,\, \lambda)),\, \frac{\delta(\xi,\, \lambda)}{\Delta(\lambda)} \right\rangle - 1 \right| \leq \ell_2 \operatorname{dist}((\xi,\, \lambda),\, \mathcal{M}). \tag{3.26}$$

Proof. Fix $\rho > 0$ such that the assertions of Lemmas 3.1–3.3, 3.5, and 3.8, and of Corollary 3.1, hold with this ρ, and there exists $\varepsilon > 0$ such that

$$\|B[\xi,\, \xi]\| \geq \varepsilon \quad \forall\, \xi \in B(\bar{\xi},\, \rho) \tag{3.27}$$

(the possibility to choose ρ satisfying the latter property follows from (3.21)). Take any $\tilde{\rho} \in (0,\, \rho)$ such that $\|\delta(\xi,\, \lambda)\| \leq \rho - \tilde{\rho}$ for all $(\xi,\, \lambda) \in B(\bar{\xi},\, \tilde{\rho}) \times (B(\bar{\lambda},\, \tilde{\rho}) \setminus \mathcal{M}_0)$ (the possibility to choose such $\tilde{\rho}$ follows from Corollary 3.1). Next, consider any $(\xi,\, \lambda) \in B(\bar{\xi},\, \tilde{\rho}) \times (B(\bar{\lambda},\, \tilde{\rho}) \setminus \mathcal{M}_0)$ with the properties specified in the statement of the lemma, and any $\theta \in [0,\, 1]$. Let $(\tilde{\xi},\, \tilde{\lambda}) \in \mathcal{M}$ be a projection of $(\xi,\, \lambda)$ onto $\mathcal{M}$. Evidently, reducing $\tilde{\rho}$ if necessary, we can assure that $(\tilde{\xi},\, \tilde{\lambda}) \in B(\bar{\xi},\, \rho) \times (B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0)$, and

$$\xi(\tilde{\lambda}) = \frac{\tilde{\xi}}{\|\tilde{\xi}\|}. \tag{3.28}$$

Set $\delta = \delta(\xi,\, \lambda)$, $\eta = \delta/\Delta(\lambda)$ and $\hat{\lambda} = \lambda - \theta\delta$. Condition (3.24) means that $\|\eta\| \leq \ell_1$. As follows from Lemma 3.8,

$$|\Delta(\lambda)| \leq \frac{1}{C_1}\, \mathrm{dist}(\lambda,\, \mathcal{M}_0) \leq \frac{1}{C_1}\|\lambda - \tilde{\lambda}\|,$$

and therefore, according to (3.24),

$$\|\hat{\lambda} - \tilde{\lambda}\| \leq \|\lambda - \tilde{\lambda}\| + \|\delta\| \leq \ell_3\|\lambda - \tilde{\lambda}\| \tag{3.29}$$

for $\ell_3 = 1 + \ell_1/C_1$. Note that the choice of $\tilde{\rho}$ implies $\hat{\lambda} \in B(\bar{\lambda},\, \rho)$.

Since the gradient of Δ is Lipschitz-continuous on $B(\bar{\lambda},\, \rho)$ with some constant $\ell_4 > 0$, from (3.29) we get

$$\left\|\Delta'(\hat{\lambda}) - \Delta'(\tilde{\lambda})\right\| \leq \ell_4\|\hat{\lambda} - \tilde{\lambda}\| \leq \ell_3\ell_4\|\lambda - \tilde{\lambda}\|. \tag{3.30}$$

Similarly, since $\hat{H}(\cdot)$ is Lipschitz-continuous on $B(\bar{\lambda},\, \rho)$ with some constant $\ell_5 > 0$, it holds that

$$\|\hat{H}(\lambda) - \hat{H}(\tilde{\lambda})\| \leq \ell_5\|\lambda - \tilde{\lambda}\|. \tag{3.31}$$

Also, evidently,

$$\left\|\Delta'(\hat{\lambda})\right\| \leq \ell_6, \quad \|\hat{H}(\lambda)\| \leq \ell_6, \quad \|\tilde{\xi}\| \leq \ell_6, \tag{3.32}$$

$$\|B[\xi]\| \leq \ell_6, \quad \|B[\tilde{\xi}]\| \leq \ell_6, \quad \|B[\xi - \tilde{\xi}]\| \leq \ell_6\|\xi - \tilde{\xi}\| \tag{3.33}$$

for some $\ell_6 > 0$. We emphasize that for any $\rho > 0$, constants ℓ_4, ℓ_5 and ℓ_6 above can be chosen independent of $(\xi,\, \lambda)$ and $(\tilde{\xi},\, \tilde{\lambda})$. Furthermore, it follows from Lemmas 3.2 and 3.3, and from (3.28), that

$$\begin{aligned}
\tilde{\xi} \cdot (\Delta'(\tilde{\lambda}))^{\mathrm{T}} &= t(\tilde{\lambda})\tilde{\xi} \cdot (\xi(\tilde{\lambda}))^{\mathrm{T}}(B[\xi(\tilde{\lambda})])^{\mathrm{T}} \\
&= t(\tilde{\lambda})\frac{\tilde{\xi}}{\|\tilde{\xi}\|}(\xi(\tilde{\lambda}))^{\mathrm{T}}(B[\|\tilde{\xi}\|\xi(\tilde{\lambda})])^{\mathrm{T}} \\
&= t(\tilde{\lambda})\xi(\tilde{\lambda})(\xi(\tilde{\lambda}))^{\mathrm{T}}(B[\tilde{\xi}])^{\mathrm{T}} \\
&= \hat{H}(\tilde{\lambda})(B[\tilde{\xi}])^{\mathrm{T}}.
\end{aligned}$$

Combining this equality with (3.30)–(3.33), we obtain

$$\begin{aligned}\|\xi\cdot(\Delta'(\hat{\lambda}))^{\mathrm{T}}-\hat{H}(\lambda)(B[\xi])^{\mathrm{T}}\| &\leq \|\xi-\tilde{\xi}\|\|\Delta'(\hat{\lambda})\|+\|\tilde{\xi}\|\|\Delta'(\hat{\lambda})-\Delta'(\tilde{\lambda})\| \\ &\quad +\|\hat{H}(\tilde{\lambda})-\hat{H}(\lambda)\|\|(B[\tilde{\xi}])^{\mathrm{T}}\| \\ &\quad +\|\hat{H}(\lambda)\|\|(B[\tilde{\xi}-\xi])^{\mathrm{T}}\| \\ &\leq (\ell_6+\ell_6^2)\|\xi-\tilde{\xi}\|+(\ell_3\ell_4+\ell_5)\ell_6\|\lambda-\tilde{\lambda}\|.\end{aligned}$$

(Recall that the matrix norm induced by the Euclidian norm coincides with the Euclidian norm when the matrix in question has just one row or one column, and that the spectral norms of a matrix and its transposed coincide.) Therefore, setting $\ell_2=\sqrt{2}\max\{\ell_6+\ell_6^2,\,(\ell_3\ell_4+\ell_5)\ell_6\}$, we obtain the inequality (3.25). Note that ℓ_2 only depends on ρ and ℓ_1 and does not depend on $(\xi,\,\lambda)$ and $(\tilde{\xi},\,\tilde{\lambda})$.

We now prove inequality (3.26). It follows from the first relation in (3.4) and from Lemma 3.5 that δ is well-defined by the equality $B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}\delta=B[\xi,\,\xi]$, implying $B[\xi]\hat{H}(\lambda)(B[\xi])^{\mathrm{T}}\eta=B[\xi,\,\xi]$, and therefore,

$$B[\xi,\,\xi](\langle\Delta'(\hat{\lambda}),\,\eta\rangle-1)=B[\xi](\xi\cdot(\Delta'(\hat{\lambda}))^{\mathrm{T}}-\hat{H}(\lambda)(B[\xi])^{\mathrm{T}})\eta.$$

From this equality and from (3.25), (3.27), and the first relation in (3.33), we obtain

$$\begin{aligned}|\langle\Delta'(\hat{\lambda}),\,\eta\rangle-1| &= \frac{\|B[\xi](\xi\cdot(\Delta'(\hat{\lambda}))^{\mathrm{T}}-\hat{H}(\lambda)(B[\xi])^{\mathrm{T}})\eta\|}{\|B[\xi,\,\xi]\|} \\ &\leq \frac{\ell_2\ell_6}{\varepsilon}\|\eta\|\,\mathrm{dist}((\xi,\,\lambda),\,\mathcal{M}) \\ &\leq \frac{\ell_1\ell_2\ell_6}{\varepsilon}\,\mathrm{dist}((\xi,\,\lambda),\,\mathcal{M}),\end{aligned}$$

where the last inequality is by the definition of η and by (3.24).

To complete the proof, it remains to re-define ρ as $\tilde{\rho}$, and to replace ℓ_2 by $\ell_1\ell_2\ell_6/\varepsilon$ if the latter is bigger than ℓ_2. ∎

Lemma 3.10 *Assuming that* (3.6) *holds for some* $\bar{\lambda}\in\mathcal{M}_0$, *let* (3.15), (3.21) *hold for* $\bar{\xi}=\xi(\bar{\lambda})$, *where the mapping* $\xi(\cdot)$ *is defined according to Lemma 3.1.*

Then there exists $\rho>0$ *such that for any* $\ell_1>0$ *and any sequences* $\{\xi^k\}\subset B(\bar{\xi},\,\rho)$ *and* $\{\lambda^k\}\subset B(\bar{\lambda},\,\rho)\setminus\mathcal{M}_0$ *such that* $\{\xi^k\}\to\xi(\tilde{\lambda})$ *and* $\{\lambda^k\}\to\tilde{\lambda}$ *for some* $\tilde{\lambda}\in\mathcal{M}_0$, *and for all* k

$$\frac{\|\delta^k\|}{|\Delta(\lambda^k)|}\leq\ell_1, \tag{3.34}$$

where $\delta^k=\delta(\xi^k,\,\lambda^k)$ *is well-defined by the first relation in* (3.4), *the following conditions hold with* $\hat{\lambda}^k=\lambda^k-\delta^k/2$:

$$\lim_{k\to\infty}\frac{\Delta(\hat{\lambda}^k)}{\Delta(\lambda^k)}=\frac{1}{2}, \tag{3.35}$$

$$\left\{P(\xi^k,\,\lambda^k)\xi^k\right\}\to\xi(\tilde{\lambda}). \tag{3.36}$$

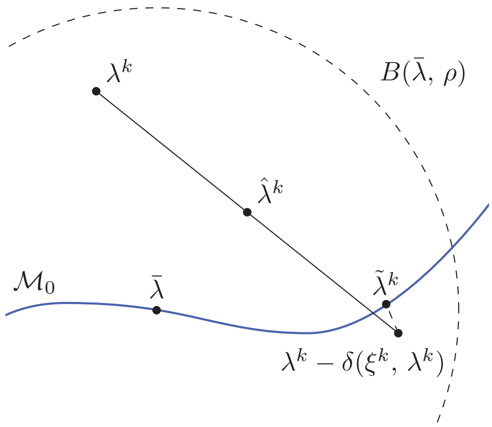


Figure 3: Illustration of Lemma 3.10.

Moreover, there exists $\ell_2 > 0$ *and a sequence* $\{\tilde{\lambda}^k\} \subset \mathcal{M}_0$ *such that* $\{\tilde{\lambda}^k\} \to \tilde{\lambda}$ *and the following estimates hold for all* k *large enough:*

$$\frac{\|\lambda^k - \delta^k - \tilde{\lambda}^k\|}{|\Delta(\lambda^k)|} \leq \ell_2 \operatorname{dist}((\xi^k, \lambda^k), \mathcal{M}), \tag{3.37}$$

$$\frac{\|\hat{\xi}^k - \xi(\tilde{\lambda}^k)\|}{|\Delta(\lambda^k)|} \leq \ell_2 \operatorname{dist}((\xi^k, \lambda^k), \mathcal{M}), \tag{3.38}$$

$$\|\eta^k - \hat{\eta}^k\| \leq \ell_2 \operatorname{dist}((\xi^k, \lambda^k), \mathcal{M}), \tag{3.39}$$

where

$$\hat{\xi}^k = \frac{P(\xi^k, \lambda^k)\xi^k}{\|P(\xi^k, \lambda^k)\xi^k\|}, \quad \eta^k = \frac{\delta^k}{\Delta(\lambda^k)}, \quad \hat{\eta}^k = \frac{\delta(\hat{\xi}^k, \hat{\lambda}^k)}{\Delta(\hat{\lambda}^k)}$$

with $P(\xi^k, \lambda^k)\xi^k \neq 0$ *well-defined by the second relation in* (3.4)*, and with* $\Delta(\hat{\lambda}^k) \neq 0$.

Proof. Fix $\rho > 0$ such that the assertions of Lemmas 3.7–3.9, and of Corollary 3.1 hold with this ρ. Take any $\tilde{\rho} \in (0, \rho)$, and consider any sequences $\{\xi^k\} \subset B(\bar{\xi}, \tilde{\rho})$ and $\{\lambda^k\} \subset B(\bar{\lambda}, \tilde{\rho}) \setminus \mathcal{M}_0$ with the properties specified in the statement above. Condition (3.34) means that the sequence $\{\eta^k\}$ is bounded, and hence $\{\delta^k\} \to 0$ and $\{\hat{\lambda}^k\} \to \tilde{\lambda}$.

We first prove that there exists $\ell_3 > 0$ such that for all k

$$\left| \frac{\Delta(\hat{\lambda}^k)}{\Delta(\lambda^k)} - \frac{1}{2} \right| \leq \ell_3 \operatorname{dist}((\xi^k, \lambda^k), \mathcal{M}). \tag{3.40}$$

Indeed, by the mean-value theorem, there exists a sequence $\{\theta_k\} \subset [0, 1]$ such that

$$1 - \frac{\Delta(\hat{\lambda}^k)}{\Delta(\lambda^k)} = \frac{\Delta(\lambda^k) - \Delta(\hat{\lambda}^k)}{\Delta(\lambda^k)} = \frac{\left\langle \Delta'\left(\lambda^k - \theta_k \frac{\delta^k}{2}\right), \frac{\delta^k}{2} \right\rangle}{\Delta(\lambda^k)} = \frac{1}{2}\left\langle \Delta'\left(\lambda^k - \theta_k \frac{\delta^k}{2}\right), \eta^k \right\rangle.$$

The needed property (3.40) with $\ell_3 = \ell_2/2$ now follows immediately by applying (3.26) in Lemma 3.9. Observe that (3.40) readily implies (3.35).

Furthermore, by the mean-value theorem, there exists a sequence $\{\theta_k\} \subset [0, 1]$ such that for all k

$$1 - \frac{\Delta(\lambda^k - \delta^k)}{\Delta(\lambda^k)} = \langle \Delta'(\lambda^k - \theta_k \delta^k), \eta^k \rangle.$$

Applying again (3.26) in Lemma 3.9, we obtain

$$\left| \frac{\Delta(\lambda^k - \delta^k)}{\Delta(\lambda^k)} \right| \leq \ell_2 \operatorname{dist}((\xi^k, \lambda^k), \mathcal{M}).$$

Therefore, according to the second inequality in (3.22) in Lemma 3.8, for all k large enough

$$\left| \frac{\operatorname{dist}(\lambda^k - \delta^k, \mathcal{M}_0)}{\Delta(\lambda^k)} \right| \leq C_2 \ell_2 \operatorname{dist}((\xi^k, \lambda^k), \mathcal{M})$$

(since $\lambda^k - \delta^k \in B(\bar{\lambda}, \rho)$ for such k), which immediately implies the existence of a sequence $\{\tilde{\lambda}^k\} \subset \mathcal{M}_0$ satisfying (3.37) with ℓ_2 re-defined as $C_2 \ell_2$ if the latter is larger than ℓ_2. Since $\{\lambda^k\} \to \tilde{\lambda}$ and $\{\delta^k\} \to 0$, (3.37) evidently implies that $\{\tilde{\lambda}^k\} \to \tilde{\lambda}$.

For brevity, for all k set $H_k = H(\lambda^k)$, $\hat{H}_k = \hat{H}(\lambda^k)$, $P_k = P(\xi^k, \lambda^k)$. We further show that there exists $\ell_4 > 0$ such that

$$\|P_k \xi^k - \xi^k\| \leq \ell_4 \operatorname{dist}((\xi^k, \lambda^k), \mathcal{M}). \tag{3.41}$$

Indeed, according to the second relation in (3.4), and to the definition of η^k,

$$P_k \xi^k = H_k^{-1}(\delta^k B)\xi^k = \hat{H}_k(\eta^k B)\xi^k = \hat{H}_k(B[\xi^k])^{\mathrm{T}} \eta^k,$$

and therefore

$$P_k \xi^k - \xi^k = (\hat{H}_k(B[\xi^k])^{\mathrm{T}} - \xi^k(\Delta'(\lambda^k))^{\mathrm{T}})\eta^k + \xi^k(\langle \Delta'(\lambda^k), \eta^k \rangle - 1).$$

Since both sequences $\{\xi^k\}$ and $\{\eta^k\}$ are bounded, estimate (3.41) with some $\ell_4 > 0$ follows now immediately from Lemma 3.9. Observe that estimate (3.41) further implies (3.36), and, in particular, $P_k \xi^k \neq 0$ for all k large enough.

We further prove that the estimate (3.38) holds for the sequence $\{\tilde{\lambda}^k\}$ specified above. Observe first that $H_k P_k = \delta^k B$, and hence, setting $\Delta_k = \Delta(\lambda^k) (= \det H_k)$,

$$\frac{(H_k - \delta^k B) P_k \xi^k}{\Delta_k} = \frac{(\delta^k B)(\xi^k - P_k \xi^k)}{\Delta_k} = (\eta^k B)(\xi^k - P_k \xi^k).$$

Therefore, setting $\tilde{H}_k = H(\tilde{\lambda}^k)$, by the definition of $\hat{\xi}^k$ we get

$$\begin{aligned}\frac{\tilde{H}_k\hat{\xi}^k}{\Delta_k} &= \frac{\tilde{H}_k P_k\xi^k}{\Delta_k\|P_k\xi^k\|}\\ &= \frac{(H_k-(\lambda^k-\tilde{\lambda}^k)B)P_k\xi^k}{\Delta_k\|P_k\xi^k\|}\\ &= \frac{1}{\|P_k\xi^k\|}\left(\frac{(H_k-\delta^k B)P_k\xi^k}{\Delta_k}-\frac{((\lambda^k-\delta^k-\tilde{\lambda}^k)B)P_k\xi^k}{\Delta_k}\right)\\ &= \frac{\eta^k B}{\|P_k\xi^k\|}(\xi^k-P_k\xi^k)-\frac{(B[P_k\xi^k])^{\mathrm{T}}(\lambda^k-\delta^k-\tilde{\lambda}^k)}{\|P_k\xi^k\|\Delta_k}. \end{aligned} \tag{3.42}$$

Since $\{\eta^k\}$ is bounded, (3.36) implies the existence of $\ell_5 > 0$ such that

$$\frac{\|\eta^k B\|}{\|P_k\xi^k\|} \le \ell_5, \quad \frac{\|(B[P_k\xi^k])^{\mathrm{T}}\|}{\|P_k\xi^k\|} \le \ell_5$$

for all k large enough. Therefore, combining (3.42) with (3.37) and (3.41), we obtain

$$\frac{\|\tilde{H}_k\hat{\xi}^k\|}{\Delta_k} \le \ell_6\, \mathrm{dist}((\xi^k,\, \lambda^k),\, \mathcal{M}) \tag{3.43}$$

with $\ell_6 = (\ell_2 + \ell_4)\ell_5$. Finally, according to (3.36) and to convergence of $\{\tilde{\lambda}^k\}$ to $\tilde{\lambda}$, for all k large enough it holds that $\hat{\xi}^k \in B(\bar{\xi},\, \rho)$ and $\tilde{\lambda}^k \in B(\bar{\lambda},\, \rho)$, and therefore, by Lemma 3.7, $\|\tilde{H}_k\hat{\xi}^k\| \ge \gamma\|\hat{\xi}^k - \xi(\tilde{\lambda}^k)\|$. Combining this with (3.43), we conclude that

$$\frac{\|\hat{\xi}^k - \xi(\tilde{\lambda}^k)\|}{\Delta_k} \le \frac{1}{\gamma}\frac{\|\tilde{H}_k\hat{\xi}^k\|}{\Delta_k} \le \frac{\ell_6}{\gamma}\, \mathrm{dist}((\xi^k,\, \lambda^k),\, \mathcal{M}),$$

which proves (3.38) with ℓ_2 re-defined as ℓ_6/γ.

It remains to establish (3.39). Observe that $\Delta(\hat{\lambda}^k) \ne 0$ for all k large enough, since otherwise (3.35) would not be possible. From the equality

$$\frac{\hat{\lambda}^k - \tilde{\lambda}^k}{\Delta_k} = \frac{\eta^k}{2} + \frac{\lambda^k - \delta^k - \tilde{\lambda}^k}{\Delta_k}, \tag{3.44}$$

and from (3.37) we readily obtain that

$$\left\|\eta^k - 2\frac{\hat{\lambda}^k - \tilde{\lambda}^k}{\Delta_k}\right\| \le 2\ell_2\, \mathrm{dist}((\xi^k,\, \lambda^k),\, \mathcal{M}). \tag{3.45}$$

Equality (3.44) and conditions (3.34), (3.37) imply that for some $\varepsilon > 0$ and for all k large enough $\|\hat{\lambda}^k - \tilde{\lambda}^k\|/|\Delta_k| \le \ell_1/2 + \varepsilon$. Furthermore, setting $\hat{\Delta}_k = \Delta(\hat{\lambda}^k)(= \det H(\hat{\lambda}^k))$, we easily get from (3.35) that for all k large enough

$$\left|\frac{\Delta_k}{\hat{\Delta}_k}\right| \le 2 + \varepsilon. \tag{3.46}$$

Combining the last two estimates with (3.40), we get

$$\left\| \frac{\hat{\lambda}^k - \tilde{\lambda}^k}{\Delta_k} - \frac{\hat{\lambda}^k - \tilde{\lambda}^k}{2\hat{\Delta}_k} \right\| = \left| \frac{\hat{\Delta}_k}{\Delta_k} - \frac{1}{2} \right| \left| \frac{\Delta_k}{\hat{\Delta}_k} \right| \frac{\|\hat{\lambda}^k - \tilde{\lambda}^k\|}{|\Delta_k|} \leq \ell_7 \operatorname{dist}((\xi^k,\, \lambda^k),\, \mathcal{M}) \tag{3.47}$$

with $\ell_7 = (\ell_1/2 + \varepsilon)(2 + \varepsilon)\ell_3$. Also, from Corollary 3.1 we have that for all k large enough

$$\|\delta(\hat{\xi}^k,\, \hat{\lambda}^k) - (\hat{\lambda}^k - \tilde{\lambda}^k)\| \leq \ell \|\hat{\xi}^k - \xi(\tilde{\lambda}^k)\|,$$

and hence, by the definition of $\hat{\eta}^k$, and by (3.38), (3.46) we derive the estimate

$$\left\| \hat{\eta}^k - \frac{\hat{\lambda}^k - \tilde{\lambda}^k}{\hat{\Delta}_k} \right\| \leq \ell \frac{\|\hat{\xi}^k - \xi(\tilde{\lambda}^k)\|}{|\hat{\Delta}_k|} = \ell \left| \frac{\Delta_k}{\hat{\Delta}_k} \right| \frac{\|\hat{\xi}^k - \xi(\tilde{\lambda}^k)\|}{|\Delta_k|} \leq \ell_8 \operatorname{dist}((\xi^k,\, \lambda^k),\, \mathcal{M}) \tag{3.48}$$

with $\ell_8 = \ell\ell_2(2 + \varepsilon)$. From (3.45), (3.47), and (3.48) we now obtain

$$\begin{aligned} \|\eta^k - \hat{\eta}^k\| &\leq \left\| \eta^k - 2\frac{\hat{\lambda}^k - \tilde{\lambda}^k}{\Delta_k} \right\| + \left\| 2\frac{\hat{\lambda}^k - \tilde{\lambda}^k}{\Delta_k} - \frac{\hat{\lambda}^k - \tilde{\lambda}^k}{\hat{\Delta}_k} \right\| + \left\| \frac{\hat{\lambda}^k - \tilde{\lambda}^k}{\hat{\Delta}_k} - \hat{\eta}^k \right\| \\ &\leq (2\ell_2 + 2\ell_7 + \ell_8) \operatorname{dist}((\xi^k,\, \lambda^k),\, \mathcal{M}) \end{aligned}$$

for all k large enough. Re-defining ℓ_2 as $2\ell_2 + 2\ell_7 + \ell_8$, we get (3.39).

To complete the proof, re-define ρ as $\tilde{\rho}$. ∎

The constant ℓ_2 in this lemma does not need to be the same for different sequences $\{\xi^k\}$, $\{\lambda^k\}$. However, the next corollary shows that in fact this constant can be chosen independently of the sequences.

Corollary 3.2 *Under the assumptions of Lemma 3.9, for any* $\ell_1 > 0$ *and any* $\varepsilon > 0$*, there exist* $\rho > 0$ *and* $\ell_2 > 0$ *such that for all* $(\xi,\, \lambda) \in B(\bar{\xi},\, \rho) \times (B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0)$ *satisfying* (3.24) *with* $\delta(\xi,\, \lambda)$ *well-defined by the first relation in* (3.4)*, the following conditions hold:*

$$\left| \frac{\Delta(\hat{\lambda})}{\Delta(\lambda)} - \frac{1}{2} \right| \leq \varepsilon, \tag{3.49}$$

$$\left\| \frac{\delta(\xi,\, \lambda)}{\Delta(\lambda)} - \frac{\delta(\hat{\lambda},\, \hat{\xi})}{\Delta(\hat{\lambda})} \right\| \leq \ell_2 \operatorname{dist}((\xi,\, \lambda),\, \mathcal{M}), \tag{3.50}$$

$$\operatorname{dist}((\hat{\xi},\, \hat{\lambda}),\, \mathcal{M}) \leq \ell_2 |\Delta(\lambda)|, \tag{3.51}$$

$$\|\hat{\xi} - \bar{\xi}\| \leq \ell_2 \|\lambda - \bar{\lambda}\|, \tag{3.52}$$

where

$$\hat{\xi} = \frac{P(\xi,\, \lambda)\xi}{\|P(\xi,\, \lambda)\xi\|}, \quad \hat{\lambda} = \lambda - \frac{\delta(\xi,\, \lambda)}{2} \tag{3.53}$$

with $P(\xi,\, \lambda)\xi \neq 0$ *well-defined by the second relation in* (3.4).

Proof. Fix any $\ell_1 > 0$ and any $\varepsilon > 0$. The existence of $\rho > 0$ and $\ell_2 > 0$ such that (3.49) and (3.50) and the condition $P(\xi,\, \lambda)\xi \neq 0$ hold for all $\xi \in B(\bar{\xi},\, \rho)$ and $\lambda \in B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0$ follows readily from (3.35), (3.36), and (3.39) in Lemma 3.10, arguing by contradiction. It remains to establish (3.51) and (3.52).

Fix any $\varepsilon_1 > 0$, $\varepsilon_2 > 0$. Employing (3.37) and (3.38) in Lemma 3.10, by further reducing $\rho > 0$ if necessary, one can assure (again arguing by contradiction) that for all $\xi \in B(\bar{\xi},\, \rho)$ and $\lambda \in B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0$ satisfying (3.24), there exists $\tilde{\lambda} \in \mathcal{M}_0$ such that

$$\frac{\|(\lambda - \tilde{\lambda}) - \delta(\xi,\, \lambda)\|}{|\Delta(\lambda)|} \leq \varepsilon_1, \quad \frac{\|\hat{\xi} - \xi(\tilde{\lambda})\|}{|\Delta(\lambda)|} \leq \varepsilon_2. \tag{3.54}$$

From the first inequality, from the definition of $\hat{\lambda}$, and from (3.24), we then obtain

$$\|\hat{\lambda} - \tilde{\lambda}\| \leq \|(\lambda - \tilde{\lambda}) - \delta(\xi,\, \lambda)\| + \frac{\|\delta(\xi,\, \lambda)\|}{2} \leq \left(\varepsilon_1 + \frac{\ell_1}{2}\right) |\Delta(\lambda)|.$$

Therefore, by the second inequality in (3.54) we get

$$\mathrm{dist}((\hat{\xi},\, \hat{\lambda}),\, \mathcal{M}) \leq \|\hat{\xi} - \xi(\tilde{\lambda})\| + \|\hat{\lambda} - \tilde{\lambda}\| \leq \left(\varepsilon_1 + \varepsilon_2 + \frac{\ell_1}{2}\right) |\Delta(\lambda)|,$$

which proves (3.51) with $\ell_2 = \varepsilon_1 + \varepsilon_2 + \ell_1/2$.

Finally, from (3.24) and the first condition in (3.54) we obtain

$$\frac{\|\lambda - \tilde{\lambda}\|}{|\Delta(\lambda)|} \leq \frac{\|(\lambda - \tilde{\lambda}) - \delta(\xi,\, \lambda)\|}{|\Delta(\lambda)|} + \frac{\|\delta(\xi,\, \lambda)\|}{|\Delta(\lambda)|} \leq \varepsilon_1 + \ell_1,$$

and hence $\|\lambda - \tilde{\lambda}\| \leq (\varepsilon_1 + \ell_1)|\Delta(\lambda)|$, which further implies

$$\|\tilde{\lambda} - \bar{\lambda}\| \leq \|\lambda - \tilde{\lambda}\| + \|\lambda - \bar{\lambda}\| \leq (\varepsilon_1 + \ell_1)|\Delta(\lambda)| + \|\lambda - \bar{\lambda}\|. \tag{3.55}$$

By this estimate we conclude that $\tilde{\lambda} \to \bar{\lambda}$ as $\lambda \to \bar{\lambda}$, and hence, for any $\tilde{\rho} > 0$ we can assure that $\tilde{\lambda} \in B(\bar{\lambda},\, \tilde{\rho})$, reducing ρ if necessary. Since $\xi(\cdot)$ is analytic, there exists $\ell_3 > 0$ such that

$$\|\xi(\tilde{\lambda}) - \bar{\xi}\| \leq \ell_3\|\tilde{\lambda} - \bar{\lambda}\| \quad \forall\, \tilde{\lambda} \in \mathcal{M}_0 \cap B(\bar{\lambda},\, \tilde{\rho}),$$

assuming that $\tilde{\rho} > 0$ is small enough, so that $\xi(\tilde{\lambda})$ is well-defined. Combining this inequality with the second inequality in (3.54) and with (3.55), we further obtain

$$\begin{aligned} \|\hat{\xi} - \bar{\xi}\| &\leq \|\hat{\xi} - \xi(\tilde{\lambda})\| + \|\xi(\tilde{\lambda}) - \bar{\xi}\| \\ &\leq \varepsilon_2|\Delta(\lambda)| + \ell_3\|\tilde{\lambda} - \bar{\lambda}\| \\ &\leq ((\varepsilon_1 + \ell_1)\ell_3 + \varepsilon_2)|\Delta(\lambda)| + \ell_3\|\lambda - \bar{\lambda}\| \end{aligned} \tag{3.56}$$

According to Lemma 3.8, if $\rho > 0$ is small enough, then

$$|\Delta(\lambda)| \leq \frac{1}{C_1}\,\mathrm{dist}(\lambda,\, \mathcal{M}_0) \leq \frac{1}{C_1}\|\lambda - \bar{\lambda}\|,$$

and hence, (3.56) implies (3.52) with $\ell_2 = ((\varepsilon_1 + \ell_1)\ell_3 + \varepsilon_2)/C_1 + \ell_3$. ∎

Proposition 3.1 *Under the assumptions of Lemma 3.9, for any $\ell_0 > 0$ and $\ell_1 > 0$, $\ell_0 < \ell_1$, there exist $\rho_0 > 0$ and $\rho > 0$, $\rho_0 < \rho$, such that for any starting point $(\xi^0,\, \lambda^0) \in B(\bar{\xi},\, \rho_0) \times (B(\bar{\lambda},\, \rho_0) \setminus \mathcal{M}_0)$ satisfying*

$$\frac{\|\delta(\xi^0,\, \lambda^0)\|}{|\Delta(\lambda^0)|} \leq \ell_0 \tag{3.57}$$

with $\delta(\xi^0,\, \lambda^0)$ well-defined by the first relation in (3.4), the iterative scheme

$$\xi^{k+1} = \frac{P(\xi^k,\, \lambda^k)\xi^k}{\|P(\xi^k,\, \lambda^k)\xi^k\|}, \quad \lambda^{k+1} = \lambda^k - \frac{\delta(\xi^k,\, \lambda^k)}{2} \tag{3.58}$$

with $\delta(\xi^k,\, \lambda^k)$ and $P(\xi^k,\, \lambda^k)$ introduced according to (3.4), for $k = 0,\, 1,\, \ldots$, generates the well-defined sequence $\{(\xi^k,\, \lambda^k)\} \subset B(\bar{\xi},\, \rho) \times (B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0)$; this sequence converges to $(\xi(\lambda^*),\, \lambda^*)$ for some $\lambda^* \in \mathcal{M}_0$, and for all k satisfies

$$\frac{\|\delta(\xi^k,\, \lambda^k)\|}{|\Delta(\lambda^k)|} \leq \ell_1, \tag{3.59}$$

and

$$\lim_{k\to\infty} \frac{\|\lambda^{k+1} - \lambda^*\|}{\|\lambda^k - \lambda^*\|} = \frac{1}{2}, \tag{3.60}$$

$$\left\{ \frac{\lambda^k - \lambda^*}{\|\lambda^k - \lambda^*\|} \right\} \to d \tag{3.61}$$

for some $d \in \mathbb{R}^l$ satisfying

$$\langle \Delta'(\lambda^*),\, d\rangle \geq \frac{1}{\ell_1}. \tag{3.62}$$

Proof. Fix any $\ell_0 > 0$, $\ell_1 > 0$, $\ell_0 < \ell_1$, and any $\varepsilon \in (0,\, 1/2)$, and set $q_1 = 1/2 - \varepsilon$, $q_2 = 1/2 + \varepsilon$. According to Corollary 3.2, there exist $\rho > 0$ and $\ell_2 > 0$ such that for all $\xi \in B(\bar{\xi},\, \rho)$ and $\lambda \in B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0$ satisfying (3.24), conditions (3.49)–(3.52) hold with $\hat{\xi}$ and $\hat{\lambda}$ well-defined by (3.53), and in particular, $P(\xi,\, \lambda)\xi \neq 0$. Furthermore, we reduce $\rho > 0$ and increase $\ell_2 > 0$ if necessary in order to assure that $\rho < 1$, the assertion of Lemma 3.9 holds, $\dim \ker H(\tilde{\lambda}) = 1$, $\xi(\tilde{\lambda})$ is well-defined by Lemma 3.1, and $\|\xi(\tilde{\lambda}) - \bar{\xi}\| < 1$ for all $\tilde{\lambda} \in \mathcal{M}_0 \cap B(\bar{\lambda},\, \rho)$. Finally, we take any $\rho_0 \in (0,\, \rho)$ such that the following conditions are satisfied for any $\lambda^0 \in B(\bar{\lambda},\, \rho_0) \setminus \mathcal{M}_0$:

$$\rho_0 + \frac{\ell_1|\Delta(\lambda^0)|}{2(1 - q_2)} \leq \frac{\rho}{\max\{1,\, \ell_2\}}, \tag{3.63}$$

$$\ell_0 + \sqrt{2}\ell_2\rho_0 + \frac{\ell_2^2|\Delta(\lambda^0)|}{1 - q_2} \leq \ell_1. \tag{3.64}$$

We argue by induction. Suppose that $(\xi^0,\, \lambda^0) \in B(\bar{\xi},\, \rho_0) \times (B(\bar{\lambda},\, \rho_0) \setminus \mathcal{M}_0) \subset B(\bar{\xi},\, \rho) \times (B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0)$ satisfies (3.57), and that for some nonnegative integer k, the pairs $(\xi^j,\, \lambda^j)$ are well-defined by

$$\xi^{j+1} = \frac{P(\xi^j,\, \lambda^j)\xi^j}{\|P(\xi^j,\, \lambda^j)\xi^j\|}, \quad \lambda^{j+1} = \lambda^j - \frac{\delta(\xi^j,\, \lambda^j)}{2} \tag{3.65}$$

for all $j = 0, 1, \ldots, k-1$, and satisfy $(\xi^j, \lambda^j) \in B(\bar{\xi}, \rho) \times (B(\bar{\lambda}, \rho) \setminus \mathcal{M}_0)$ and

$$\frac{\|\delta(\xi^j, \lambda^j)\|}{|\Delta(\lambda^j)|} \leq \ell_1 \tag{3.66}$$

for all $j = 0, 1, \ldots, k$. Then according to the comment above regarding the applicability of Corollary 3.2, the pair $(\xi^{k+1}, \lambda^{k+1})$ (coinciding with $(\hat{\xi}, \hat{\lambda})$ for $\xi = \xi^k$ and $\lambda = \lambda^k$) is well-defined by (3.58). Moreover, according to (3.49)–(3.52), we have that

$$\left| \frac{\Delta(\lambda^{j+1})}{\Delta(\lambda^j)} - \frac{1}{2} \right| \leq \varepsilon, \tag{3.67}$$

$$\|\eta^j - \eta^{j+1}\| \leq \ell_2 \operatorname{dist}((\xi^j, \lambda^j), \mathcal{M}), \tag{3.68}$$

$$\operatorname{dist}((\xi^{j+1}, \lambda^{j+1}), \mathcal{M}) \leq \ell_2 |\Delta(\lambda^j)|, \tag{3.69}$$

$$\|\xi^{j+1} - \bar{\xi}\| \leq \ell_2 \|\lambda^j - \bar{\lambda}\| \tag{3.70}$$

for all $j = 0, 1, \ldots, k$, where we have set $\eta^j = \delta(\xi^j, \lambda^j)/\Delta(\lambda^j)$.

Evidently, (3.67) implies the condition

$$q_1 \leq \left| \frac{\Delta(\lambda^{j+1})}{\Delta(\lambda^j)} \right| \leq q_2. \tag{3.71}$$

In particular, $\Delta(\lambda^{k+1}) \neq 0$, which means that $\lambda^{k+1} \notin \mathcal{M}_0$. Combining (3.71) with the second relation in (3.65) and with (3.66), and taking into account condition (3.63), we obtain

$$\begin{aligned}
\|\lambda^j - \bar{\lambda}\| &= \left\| \lambda^0 - \frac{1}{2} \sum_{i=0}^{j-1} \delta(\xi^i, \lambda^i) - \bar{\lambda} \right\| \\
&\leq \|\lambda^0 - \bar{\lambda}\| + \frac{1}{2} \sum_{i=0}^{j-1} \|\delta(\xi^i, \lambda^i)\| \\
&\leq \|\lambda^0 - \bar{\lambda}\| + \frac{\ell_1}{2} \sum_{i=0}^{j-1} |\Delta(\lambda^i)| \\
&\leq \|\lambda^0 - \bar{\lambda}\| + \frac{\ell_1 |\Delta(\lambda^0)|}{2} \sum_{i=0}^{j-1} q_2^i \\
&\leq \rho_0 + \frac{\ell_1 |\Delta(\lambda^0)|}{2(1 - q_2)} \\
&\leq \rho
\end{aligned} \tag{3.72}$$

for all $j = 0, 1, \ldots, k+1$, and in particular,

$$\|\lambda^{k+1} - \bar{\lambda}\| \leq \rho. \tag{3.73}$$

The next-to-the-last inequality in (3.72) and (3.63), (3.70) further imply that

$$\|\xi^{k+1} - \bar{\xi}\| \leq \ell_2 \|\lambda^k - \bar{\lambda}\| \leq \ell_2 \left(\rho_0 + \frac{\ell_1 |\Delta(\lambda^0)|}{2(1 - q_2)} \right) \leq \rho. \tag{3.74}$$

Finally, from conditions (3.68), (3.69) and (3.71) we derive

$$\begin{aligned}\|\eta^{k+1}\| &\leq \|\eta^0\| + \sum_{j=0}^{k} \|\eta^{j+1} - \eta^j\| \\ &\leq \|\eta^0\| + \ell_2 \sum_{j=0}^{k} \operatorname{dist}((\xi^j,\, \lambda^j),\, \mathcal{M}) \\ &\leq \|\eta^0\| + \ell_2 \operatorname{dist}((\xi^0,\, \lambda^0),\, \mathcal{M}) + \ell_2^2 \sum_{j=0}^{k-1} |\Delta(\lambda^j)| \\ &\leq \|\eta^0\| + \ell_2 \operatorname{dist}((\xi^0,\, \lambda^0),\, \mathcal{M}) + \ell_2^2 |\Delta(\lambda^0)| \sum_{j=0}^{k-1} q_2^j.\end{aligned}$$

Since $\operatorname{dist}((\xi^0,\, \lambda^0),\, \mathcal{M}) \leq \|(\xi^0 - \bar{\xi},\, \lambda^0 - \bar{\lambda})\| \leq \sqrt{2}\rho_0$, the above relation and (3.57), (3.64) further imply

$$\|\eta^{k+1}\| \leq \ell_0 + \sqrt{2}\ell_2\rho_0 + \frac{\ell_2^2 |\Delta(\lambda^0)|}{1 - q_2} \leq \ell_1. \tag{3.75}$$

Combining (3.73)–(3.75), and recalling the definition of η^{k+1}, we conclude that $(\xi^{k+1},\, \lambda^{k+1}) \in B(\bar{\xi},\, \rho) \times (B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0)$, and (3.66) holds for $j = k + 1$.

The argument above demonstrates that the iterative scheme (3.58) generates the well-defined sequence $\{(\xi^k,\, \lambda^k)\} \subset B(\bar{\xi},\, \rho) \times (B(\bar{\lambda},\, \rho) \setminus \mathcal{M}_0)$ satisfying (3.59) (for all k), and (3.67)–(3.70) (and hence, (3.71), for all j). We now prove that this sequence converges to $(\xi(\lambda^*),\, \lambda^*)$ for some $\lambda^* \in \mathcal{M}_0$. Indeed, for any nonnegative integers k and j, employing (3.63), the second relation in (3.65), (3.66), (3.71), we obtain

$$\begin{aligned}\|\lambda^{k+j} - \lambda^k\| &\leq \frac{1}{2} \sum_{i=k}^{k+j-1} \|\delta(\xi^i,\, \lambda^i)\| \\ &\leq \frac{\ell_1}{2} \sum_{i=k}^{k+j-1} |\Delta(\lambda^i)| \\ &\leq \frac{\ell_1 |\Delta(\lambda^0)|}{2} \sum_{i=k}^{k+j-1} q_2^i \\ &\leq \frac{\ell_1 |\Delta(\lambda^0)|}{2(1 - q_2)} q_2^k \\ &\leq \rho q_2^k.\end{aligned}$$

Therefore, $\{\lambda^k\} \subset B(\bar{\lambda},\, \rho)$ is a Cauchy sequence, and it converges to some $\lambda^* \in B(\bar{\lambda},\, \rho)$.

Moreover, according to the first relation in (3.58), it holds that $\|\xi^k\| = 1$ for all $k = 1, \ldots$, and hence, the sequence $\{\xi^k\} \subset B(\bar{\xi},\, \rho)$ has a limit point $\tilde{\xi} \in B(\bar{\xi},\, \rho)$, $\|\tilde{\xi}\| = 1$. Property (3.71) implies that $\lim_{k\to\infty} \Delta(\lambda^k) = 0$, and therefore, passing onto the limit in (3.69), we conclude that $(\tilde{\xi},\, \lambda^*) \in \mathcal{M}$, which means that $\lambda^* \in \mathcal{M}_0$ and $\tilde{\xi} \in \ker H(\lambda^*)$.

Since $\dim \ker H(\lambda^*) = 1$, it follows that $\tilde{\xi}$ equals either $\xi(\lambda^*)$ or $-\xi(\lambda^*)$. However, from the inequality $\|\xi(\lambda^*) - \bar{\xi}\| < 1$ it follows that

$$\|\bar{\xi} - (-\xi(\lambda^*))\| = \|2\bar{\xi} - (\bar{\xi} - \xi(\lambda^*))\| \geq 2\|\bar{\xi}\| - \|\bar{\xi} - \xi(\lambda^*)\| > 1,$$

and since $\rho < 1$, we conclude that $\tilde{\xi} \in B(\bar{\xi}, \rho)$ cannot coincide with $-\xi(\lambda^*)$. Therefore, $\tilde{\xi} = \xi(\lambda^*)$. In particular, $\xi(\lambda^*)$ is the unique limit point of the bounded sequence $\{\xi^k\}$, and hence, $\{\xi^k\} \to \xi(\lambda^*)$.

Next we show that $\{\eta^k\}$ converges. Indeed, from (3.68), (3.69), (3.71), we get that for any positive integers k and j

$$\begin{aligned} \|\eta^{k+j} - \eta^k\| &\leq \ell_2 \sum_{i=k}^{k+j-1} \mathrm{dist}((\xi^i, \lambda^i), \mathcal{M}) \\ &\leq \ell_2^2 \sum_{i=k-1}^{k+j-2} |\Delta(\lambda^i)| \\ &\leq \ell_2^2 |\Delta(\lambda^0)| \sum_{i=k-1}^{k+j-2} q_2^i \\ &\leq \frac{\ell_2^2 |\Delta(\lambda^0)|}{1-q_2} q_2^{k-1} \\ &\leq \ell_1 q_2^{k-1}. \end{aligned}$$

where the last inequality is by (3.64). Therefore, $\{\eta^k\}$ is a Cauchy sequence, and it converges to some $\tilde{\eta} \in \mathbb{R}^l$. Clearly, (3.59) implies $\|\tilde{\eta}\| \leq \ell_1$. Moreover, using the estimate (3.26) in Lemma 3.9, we obtain

$$\langle \Delta'(\lambda^*), \tilde{\eta} \rangle = 1, \tag{3.76}$$

and, in particular, $\tilde{\eta} \neq 0$.

We are now in a position to prove relations (3.60) and (3.61). According to condition (3.35) in Lemma 3.10, we conclude that for any $\varepsilon \in (0, 1/2)$ it holds that

$$\left| \frac{\Delta(\lambda^{k+1})}{\Delta(\lambda^k)} - \frac{1}{2} \right| \leq \varepsilon \tag{3.77}$$

for all k large enough. Again we set $q_1 = 1/2 - \varepsilon$, $q_2 = 1/2 + \varepsilon$. Then (3.77) implies that for any such k, and any nonnegative integer j,

$$q_1^j \leq \frac{\Delta(\lambda^{k+j})}{\Delta(\lambda^k)} \leq q_2^j. \tag{3.78}$$

From this expression we further obtain

$$\frac{1-q_1^j}{1-q_1} \leq \sum_{i=k}^{k+j-1} \frac{\Delta(\lambda^i)}{\Delta(\lambda^k)} \leq \frac{1-q_2^j}{1-q_2}. \tag{3.79}$$

Observe that according to the second relation in (3.58),

$$\lambda^k - \lambda^{k+j} = \frac{1}{2} \sum_{i=k}^{k+j-1} \delta(\xi^i,\, \lambda^i) = \frac{1}{2} \sum_{i=k}^{k+j-1} \Delta(\lambda^i)\eta^i,$$

and hence,

$$\frac{\lambda^k - \lambda^{k+j}}{\Delta(\lambda^k)} - \tilde{\eta} = \frac{1}{2} \sum_{i=k}^{k+j-1} \frac{\Delta(\lambda^i)}{\Delta(\lambda^k)}(\eta^i - \tilde{\eta}) + \frac{1}{2} \left(\sum_{i=k}^{k+j-1} \frac{\Delta(\lambda^i)}{\Delta(\lambda^k)} - 2 \right) \tilde{\eta}.$$

Therefore, since $\|\eta^k - \tilde{\eta}\| \leq \varepsilon$ for all k large enough, from (3.78), (3.79) we get

$$\begin{aligned}
\left\| \frac{\lambda^k - \lambda^{k+j}}{\Delta(\lambda^k)} - \tilde{\eta} \right\| &\leq \frac{1}{2} \sum_{i=k}^{k+j-1} \left| \frac{\Delta(\lambda^i)}{\Delta(\lambda^k)} \right| \|\eta^i - \tilde{\eta}\| + \frac{\|\tilde{\eta}\|}{2} \left| \sum_{i=k}^{k+j-1} \frac{\Delta(\lambda^i)}{\Delta(\lambda^k)} - 2 \right| \\
&\leq \frac{\varepsilon}{2} \sum_{i=k}^{k+j-1} q_2^{i-k} + \frac{\|\tilde{\eta}\|}{2} \max \left\{ \left| \frac{1 - q_1^j}{1 - q_1} - 2 \right|,\, \left| \frac{1 - q_2^j}{1 - q_2} - 2 \right| \right\} \\
&\leq \frac{\varepsilon}{2(1 - q_2)} + \frac{\|\tilde{\eta}\|}{2} \max \left\{ \left| \frac{1 - q_1^j}{1 - q_1} - 2 \right|,\, \left| \frac{1 - q_2^j}{1 - q_2} - 2 \right| \right\}.
\end{aligned}$$

Passing onto the limit as $j \to \infty$, we obtain that for all k large enough

$$\left\| \frac{\lambda^k - \lambda^*}{\Delta(\lambda^k)} - \tilde{\eta} \right\| \leq \frac{\varepsilon}{2(1 - q_2)} + \frac{\|\tilde{\eta}\|}{2} \max \left\{ \left| \frac{1}{1 - q_1} - 2 \right|,\, \left| \frac{1}{1 - q_2} - 2 \right| \right\} \leq \frac{\varepsilon}{1 - 2\varepsilon}(1 + 2\|\tilde{\eta}\|),$$

where the definitions of q_1 and q_2 were taken into account. Since the choice of ε is arbitrary, we conclude that

$$\left\{ \frac{\lambda^k - \lambda^*}{\Delta(\lambda^k)} \right\} \to \tilde{\eta}. \tag{3.80}$$

This relation evidently implies that

$$\lim_{k \to \infty} \frac{\|\lambda^{k+1} - \lambda^*\|}{\|\lambda^k - \lambda^*\|} \left| \frac{\Delta(\lambda^k)}{\Delta(\lambda^{k+1})} \right| = 1,$$

and therefore, (3.60) follows by (3.77) (recall that the latter holds for an arbitrarily small $\varepsilon > 0$ for k large enough).

Finally, relation (3.80) implies (3.61) with $d = \tilde{\eta}/\|\tilde{\eta}\|$, and since $\|\tilde{\eta}\| \leq \ell_1$, from (3.76) we obtain $\langle \Delta'(\lambda^*),\, d \rangle = 1/\|\tilde{\eta}\| \geq 1/\ell_1$, giving (3.62). ∎

We can now state the main result of this section, following readily from Proposition 3.1, from the first relation in (3.5) implying the equality

$$\frac{x^{k+1}}{\|x^k\|} = \frac{1}{2} P(\xi^k,\, \lambda^k)\xi^k,$$

and from $\|P(\xi^k,\, \lambda^k)\xi^k\| \to 1$ (which is implied by (3.36) in Lemma 3.10).

Theorem 3.1 *Under the assumptions of Lemma 3.9, for any $\ell_0 > 0$ and $\ell_1 > 0$, $\ell_0 < \ell_1$, there exist $\rho_0 > 0$ and $\rho > 0$, $\rho_0 < \rho$, such that for any starting point $(x^0, \lambda^0) \subset (\mathbb{R}^n \setminus \{0\}) \times (B(\bar{\lambda}, \rho_0) \setminus \mathcal{M}_0)$ satisfying $x^0/\|x^0\| \in B(\bar{\xi}, \rho_0)$ and*

$$\frac{\|\delta(x^0/\|x^0\|, \lambda^0)\|}{|\Delta(\lambda^0)|} \leq \ell_0 \tag{3.81}$$

with $\delta(x^0/\|x^0\|, \lambda^0)$ well-defined by the first relation in (3.4), there exists the unique sequence $\{(x^k, \lambda^k)\} \subset (\mathbb{R}^n \setminus \{0\}) \times (B(\bar{\lambda}, \rho) \setminus \mathcal{M}_0)$ satisfying (2.3) for all $k = 0, 1, \ldots$; this sequence converges to $(0, \lambda^*)$ for some $\lambda^* \in \mathcal{M}_0$, the sequence $\{x^k/\|x^k\|\} \subset B(\bar{\xi}, \rho)$ converges to $\xi(\lambda^*)$,

$$\lim_{k\to\infty} \frac{\|x^{k+1}\|}{\|x^k\|} = \frac{1}{2},$$

and properties (3.60), (3.61) hold with some for some $d \in \mathbb{R}^l$ satisfying (3.62) (implying, in particular, that d is transversal $\mathcal{M}_0$ at λ^*).

What happens in Example 2.2 (e.g., for the run illustrated in Figure 1) is that the sequence $\{x^k/\|x^k\|\}$ stays separated from $\bar{\xi} = \xi(\bar{\lambda}) = \pm(10, -5, -3)/\sqrt{134}$, preventing the dual sequence from convergence to the unique critical multiplier $\bar{\lambda} = 5$ for which all the assumptions (3.6), (3.15) and (3.21) are satisfied. This happens because $x^0/\|x^0\|$ is not close enough to $\bar{\xi}$.

The initial part of the run in Example 2.3 (see Figure 2) also shows that $x^0/\|x^0\|$ must be close enough to $\bar{\xi}$ for the assertion of Theorem 3.1 to take effect. Generally, in our experience, closeness of $x^0/\|x^0\|$ to $\bar{\xi}$ plays a crucial role for convergence of the dual sequence to a critical multiplier close to $\bar{\lambda}$.

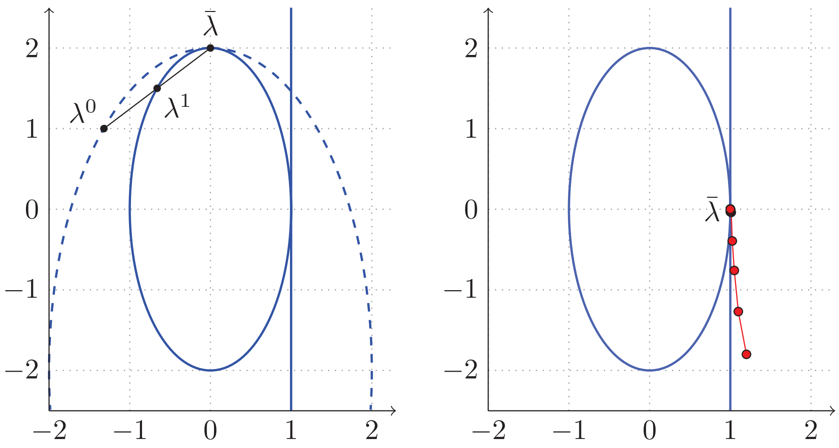
The next example demonstrates that assumption (3.81) on the staring point cannot be dropped as well.

Example 3.1 ([5], Example 2.2) For the problem

$$\text{minimize} \quad x_1^2 - x_2^2 + x_3^2 \quad \text{subject to} \quad x_1^2 + x_2^2 - x_3^2 = 0,\ x_1x_3 = 0,$$

we have $\Delta(\lambda) = -2(1-\lambda_1)(4 - 4\lambda_1^2 - \lambda_2^2)$, and hence, critical multipliers are those $\lambda \in \mathbb{R}^2$ satisfying $\lambda_1 = 1$ or $4\lambda_1^2 + \lambda_2^2 = 4$ (vertical line and oval in Figure 4).

It can be seen that assumptions (3.6), (3.15) and (3.21) do hold for all critical $\bar{\lambda} \in \mathbb{R}^2$ except for those satisfying $\bar{\lambda}_1 = 1$, and for any $\bar{\xi} \in \ker H(\bar{\lambda})$ such that $\|\bar{\xi}\| = 1$. (If $\bar{\lambda} = (1, 0)$, then $\operatorname{corank} H(\bar{\lambda}) = 2$. If $\bar{\lambda}_1 = 1$, $\bar{\lambda}_2 \neq 0$, then $\ker H(\bar{\lambda})$ is spanned by $\bar{\xi} = (0, 1, 0)$, but $\operatorname{rank} B[\bar{\xi}] = 1$, and hence, (3.15) does not hold.) Consider, for instance, $\bar{\lambda} = (0, 2)$ and $\bar{\xi} = (1/\sqrt{2}, 0, -1/\sqrt{2}) \in \ker H(\bar{\lambda})$, and take any $\lambda^0 \in \mathbb{R}^2$ satisfying $4(\lambda_1^0)^2 + (\lambda_2^0 + 2)^2 = 16$. Observe that such λ^0 can be taken arbitrarily close to $\bar{\lambda}$, and that for any $\ell_0 > 0$, inequality (3.57) is violated for $\xi^0 = \bar{\xi}$ provided such λ^0 is close enough to $\bar{\lambda}$. Furthermore, it is easy to verify that $\delta(\bar{\xi}, \lambda^0) = (\lambda_1^0, \lambda_2^0 - 2)$, and hence, $\lambda^1 = \lambda^0 - \delta(\bar{\xi}, \lambda^0)/2 = (\lambda_1^0/2, \lambda_2^0/2 + 1)$. In particular, $4(\lambda_1^1)^2 + (\lambda_2^1)^2 = 4$, which means that $\lambda^1 \in \mathcal{M}_0$, and the assertion of the theorem does not hold (see Figure 4(a)). In particular, $\delta(x^1/\|x^1\|, \lambda^1)$ is not well-defined, and all this line of analysis fails.



(a) (b)

Figure 4: Illustration of Example 3.1.

We now argue that condition (3.81) is not very restrictive. Indeed, the following proposition shows that if ρ_0 is small enough, then (3.57) is satisfied with some $\ell_0 > 0$ for all $\xi \in B(\bar{\xi}, \rho_0)$ and for λ forming, in a sense, a "dense" set in $B(\bar{\lambda}, \rho_0) \setminus \mathcal{M}_0$.

Proposition 3.2 *Under the assumptions of Lemma 3.9 there exists $\rho > 0$ such that for any $\varepsilon \in (0, 1)$ there exists $\ell_0 > 0$ possessing the following properties: for any $\rho_0 \in (0, \rho]$ condition (3.57) holds for all $(\xi^0, \lambda^0) \in B(\bar{\xi}, \rho_0) \times B(\bar{\lambda}, \rho_0)$ satisfying $\operatorname{dist}(\lambda^0, \mathcal{M}_0) \geq \varepsilon\rho_0$.*

Proof. Fix any $\rho > 0$ such that the assertions of Lemma 3.8 and Corollary 3.1 hold with this ρ. It follows from (3.19) that for all $(\xi, \lambda) \in B(\bar{\xi}, \rho) \times (B(\bar{\lambda}, \rho) \setminus \mathcal{M}_0)$

$$\|\delta(\xi, \lambda) - (\lambda - \bar{\lambda})\| \leq \ell\|\xi - \bar{\xi}\|,$$

and hence,

$$\|\delta(\xi, \lambda)\| \leq \|\lambda - \bar{\lambda}\| + \ell\|\xi - \bar{\xi}\| \leq (1 + \ell)\max\{\|\lambda - \bar{\lambda}\|, \|\xi - \bar{\xi}\|\}.$$

Employing the second inequality in (3.22), we further obtain that

$$\left\|\frac{\delta(\xi, \lambda)}{\Delta(\lambda)}\right\| \leq C_2(1 + \ell)\frac{\max\{\|\lambda - \bar{\lambda}\|, \|\xi - \bar{\xi}\|\}}{\operatorname{dist}(\lambda, \mathcal{M}_0)}.$$

Therefore, for any $\varepsilon \in (0, 1)$, if $(\xi^0, \lambda^0) \in B(\bar{\xi}, \rho_0) \times B(\bar{\lambda}, \rho_0)$ and $\operatorname{dist}(\lambda^0, \mathcal{M}_0) \geq \varepsilon\rho_0$ for some $\rho_0 \in (0, \rho]$, then

$$\left\|\frac{\delta(\xi^0, \lambda^0)}{\Delta(\lambda^0)}\right\| \leq C_2(1 + \ell)\frac{\rho_0}{\varepsilon\rho_0} \leq \ell_0$$

with $\ell_0 = C_2(1 + \ell)/\varepsilon$, which does not depend on ρ_0 and (ξ^0, λ^0). ∎

Theorem 3.1 and Proposition 3.2 imply that for any $\varepsilon \in (0, 1)$ there exists $\rho_0 > 0$ such that for any starting point $(x^0, \lambda^0) \subset (\mathbb{R}^n \setminus \{0\}) \times B(\bar{\lambda}, \rho_0)$ satisfying $x^0/\|x^0\| \in B(\bar{\xi}, \rho_0)$ and $\mathrm{dist}(\lambda^0, \mathcal{M}_0) \geq \varepsilon\rho_0$, the method in question converges to a critical multiplier.

Observe further that condition (3.81) can be dropped in Theorem 3.1 in the special case of $l = 1$. Indeed, as follows from the first equality in (3.4) and from (3.17), in this case

$$\frac{\delta(\xi, \lambda)}{\Delta(\lambda)} = \frac{B[\xi, \xi]}{R(B[\xi], B[\xi], H(\lambda))}$$

for all $\xi \in B(\bar{\xi}, \rho)$ and $\lambda \in (B(\bar{\lambda}, \rho) \setminus \mathcal{M}_0$ provided $\rho > 0$ is small enough. Therefore, assumption (3.15) implies the existence of $\rho_0 > 0$ and $\ell_0 > 0$ such that condition (3.57) holds for all $(\xi^0, \lambda^0) \in B(\bar{\xi}, \rho_0) \times (B(\bar{\lambda}, \rho_0) \setminus \mathcal{M}_0)$.

Another interesting observation is that assumption (3.15) can also be dropped in the analysis above in the case of $l = 1$. Indeed, in this case for any $\xi \in \mathbb{R}^n$ and any $\lambda \in \mathbb{R}^l \setminus \{0\}$ it holds that

$$\det(B[\xi](H(\lambda))^{-1}(B[\xi])^{\mathrm{T}}) \det H(\lambda) = B[\xi]\hat{H}(\lambda)(B[\xi])^{\mathrm{T}},$$

and the property that we actually need in Lemma 3.5 is not (3.15) as such but rather that $B[\bar{\xi}]\hat{H}(\bar{\lambda})(B[\bar{\xi}])^{\mathrm{T}} \neq 0$. Since $\hat{H}(\bar{\lambda}) = t(\bar{\lambda})\bar{\xi}\bar{\xi}^{\mathrm{T}}$ (see Lemma 3.2), the needed property reduces to (3.21).

However, assumption (3.15) cannot be removed in the general case: without (3.15), we cannot guarantee that the matrix in the left-hand side of (2.3) remains nonsingular. Indeed, consider any optimization problem with one constraint being identically zero. It is clear that assumptions (3.6) and (3.21) can be easily satisfied, but $\mathrm{corank}\, B[\bar{\xi}] > 0$, and hence, (3.15) is violated. In such example the matrix in the left-hand side of (2.3) always contains the row of zeroes. Hence, the next iterate is not uniquely defined, and the corresponding components of the dual iterates do not necessarily converge. In particular, the assertion of Theorem 3.1 does not hold in this case.

Observe that the pair of assumptions (3.6), (3.21) is in fact equivalent to (3.23). Indeed, if (3.6) holds, then Lemma 3.3 implies that (3.23) is satisfied if, and only if, (3.21) holds. Therefore, we only need to show that (3.23) implies (3.6). To that end, it is sufficient to show that for any matrices $H, B \in \mathbb{R}^{n\times n}$, if $\mathrm{corank}\, H \geq 2$, then

$$\left.\frac{d}{dt}\det(H + tB)\right|_{t=0} = 0. \tag{3.82}$$

If B is nonsingular, then $\mathrm{corank}\, B^{-1}H \geq 2$, and hence, the matrix $B^{-1}H$ has a zero eigenvalue of multiplicity no less than 2. Therefore, $\det(B^{-1}H - tI) = O(t^2)$, which implies

$$\left.\frac{d}{dt}\det(H + tB)\right|_{t=0} = -\det B \left.\frac{d}{dt}\det(B^{-1}H - tI)\right|_{t=0} = 0.$$

If B is singular, consider any sequence $\{B_k\} \subset \mathbb{R}^{n\times n}$ of nonsingular matrices convergent to B (evidently, such sequence always exists). Then, according to argument above,

$$\left.\frac{d}{dt}\det(H + tB_k)\right|_{t=0} = 0$$

for all k. Passing onto the limit, we again obtain (3.82).

Furthermore, let $\mathbf{S}^n$ stand for the space of $n \times n$ symmetric matrices. By the parametric transversality theorem (see, e.g., [2, Theorem 2.7]) it can be derived that for all (A, B) in some massive (and in particular, dense) set in $\mathbf{S}^n \times \bigotimes_{i=1}^{l} \mathbf{S}^n$, for any $\bar{\lambda} \in \mathcal{M}_0$ it holds that

$$\operatorname{corank} H(\bar{\lambda})(\operatorname{corank} H(\bar{\lambda}) + 1) \leq 2l,$$

and if (3.6) is valid, then (3.23) (and hence, (3.21)) holds as well. In particular, if $l \leq 2$, then assumptions (3.6) and (3.21) hold generically for all critical multipliers. If $l \geq 3$, then critical multipliers violating (3.6) or (3.21) may exist in a stable way, but usually these assumptions do hold for "almost all" critical multipliers.

Getting back to Example 3.1, consider critical multipliers satisfying $\lambda_1 = 1$. Assumptions (3.6) and (3.21) (or equivalently (3.23)) are satisfied for all such multipliers except for $(1, 0)$. However, for any $\bar{\lambda} = (1, \bar{\lambda}_2)$ with $\bar{\lambda}_2 \neq 0$ it holds that $\operatorname{rank} B[\bar{\xi}] \leq 1$ for $\bar{\xi} = (0, 1, 0) \in \ker H(\bar{\lambda})$, and hence, (3.15) is violated for such $\bar{\lambda}$. Moreover, assumption (3.6) (and hence, (3.23)) is not satisfied for $\bar{\lambda} = (1, 0)$ since $\operatorname{corank} H(\bar{\lambda}) = 2$. If we take any starting point (x^0, λ^0) such that $(x^0/\|x^0\|, \lambda^0)$ is close enough to $\{(0, 1, 0)\} \times (\{1\} \times \mathbb{R})$, the dual sequences usually converge to $\bar{\lambda} = (1, 0)$ rather than to a critical multiplier close to λ^0. An example of such behavior is shown in Figure 4(b) (where $x^0 = (0.1, 1.1, 0.1)$ and $\lambda^0 = (1.2, -1.8)$).

Therefore, it appears that critical multipliers of higher "multiplicity" (violating (3.6), and hence, (3.23)) seem to be even more attractive for dual sequences than other critical multipliers. However, the analysis of such multipliers is not at issue in this paper.

Finally, the authors are not aware of any examples satisfying (3.6), (3.16), (3.21), but violating (3.15). Therefore, the conjecture is that under (3.6), (3.21), condition (3.15) is actually equivalent to (3.16). At the same time, it can be seen by the transversality theorem that if $n + 2 > 2l$, then for all (A, B) in some massive set in $\mathbf{S}^n \times \bigotimes_{i=1}^{l} \mathbf{S}^n$, for any $\bar{\lambda} \in \mathcal{M}_0$ condition (3.16) holds for all $\bar{\xi} \in \ker H(\bar{\lambda}) \setminus \{0\}$.

4 Repulsion from noncritical multipliers

Along with local attraction to critical multipliers established in Theorem 3.1, it is still quite important to get further understanding of why dual sequences are pushed away from noncritical multipliers and usually do not accumulate around them. An ideal result in this direction would consist of showing that any accumulation point of the dual sequence is necessarily critical. However, Example 2.2 demonstrates that such claim would be too strong: dual sequences in this example do have noncritical accumulation points, at least according to our numerical experience. Note, however, that dual sequences in this example do not converge. This section contains some new observations on why convergence to a noncritical multiplier is at least atypical.

According to [6, Proposition 1], if a sequence $\{(x^k, \lambda^k)\}$ generated by (2.3) converges to $(0, \bar{\lambda})$ with some $\bar{\lambda} \in \mathbb{R}^l$, and if $x^k \neq 0$ for all k, then

$$\langle H(\bar{\lambda})\xi^k, \xi^{k+1} \rangle \to 0, \tag{4.1}$$

and if $\bar{\lambda}$ is noncritical, then, in addition,

$$\{B[\xi^k, \xi^k]\} \to 0. \tag{4.2}$$

Let $\bar{\xi}^1, \bar{\xi}^2 \in \mathbb{R}^n$ be accumulation points of $\{\xi^k\}$ and $\{\xi^{k+1}\}$, respectively. If $\bar{\lambda}$ is noncritical, then from (4.1), (4.2) we have

$$\langle H(\bar{\lambda})\bar{\xi}^1, \bar{\xi}^2\rangle = 0, \quad B[\bar{\xi}^1, \bar{\xi}^1] = 0, \quad B[\bar{\xi}^2, \bar{\xi}^2] = 0. \tag{4.3}$$

Therefore, a noncritical $\bar{\lambda}$ can be a candidate for convergence to it of dual sequences only provided there exist nonzero $\bar{\xi}^1$ and $\bar{\xi}^2$ satisfying (4.3).

In particular, if $\{\xi^k\}$ and $\{\xi^{k+1}\}$ have a common accumulation point $\bar{\xi} \in \mathbb{R}^n$ (or accumulation points $\bar{\xi}$ and $-\bar{\xi}$, respectively), then (4.3) takes the form

$$\langle A\bar{\xi}, \bar{\xi}\rangle = 0, \quad B[\bar{\xi}, \bar{\xi}] = 0,$$

where there is no dependence on $\bar{\lambda}$ anymore. Clearly, for generic A and B such that 0 is a solution of problem (2.1), such nonzero $\bar{\xi}$ does not exist. However, it is questionable how natural in this context is the assumption that $\{\xi^k\}$ and $\{\xi^{k+1}\}$ have a common accumulation point.

Consider the case when $l = 1$. (This special case is actually quite important since it would be of primary interest when more general constraints would come into play: it is only natural to consider first the case when the rank of the constraints Jacobian at the solution is dropped by 1, and then the essential part of the second differential will take values in a 1-dimensional space complementing the range space of the Jacobian.) Then B can be seen as a single symmetric $n \times n$ matrix, and (4.3) takes the form

$$\langle A\bar{\xi}^1, \bar{\xi}^2\rangle + \bar{\lambda}\langle B\bar{\xi}^1, \bar{\xi}^2\rangle = 0, \quad \langle B\bar{\xi}^1, \bar{\xi}^1\rangle = 0, \quad \langle B\bar{\xi}^2, \bar{\xi}^2\rangle = 0. \tag{4.4}$$

This might give a rather strong restriction on possible values of $\bar{\lambda}$. In particular, any pair $(\bar{\xi}^1, \bar{\xi}^2)$ satisfying the last two relations in (4.4), and such that $\langle B\bar{\xi}^1, \bar{\xi}^2\rangle \neq 0$, uniquely defines

$$\bar{\lambda} = -\frac{\langle A\bar{\xi}^1, \bar{\xi}^2\rangle}{\langle B\bar{\xi}^1, \bar{\xi}^2\rangle}, \tag{4.5}$$

according to the first relation in (4.4). Recall Example 2.2, where the “special” multiplier $\bar{\lambda} = -3$ was singled out precisely this way, in fact.

If $l = 1$, then (3.4) takes the form

$$\delta(\xi, \lambda) = \frac{\langle B\xi, \xi\rangle}{\langle B\xi, \Pi(\lambda)\xi\rangle}, \quad P(\xi, \lambda) = \delta(\xi, \lambda)\Pi(\lambda), \tag{4.6}$$

where $\Pi(\lambda) = (H(\lambda))^{-1}B$. Then (3.5) transforms into

$$x^{k+1} = -\eta^k\Pi(\lambda^k)x^k, \quad \lambda^{k+1} = \lambda^k + \eta^k, \tag{4.7}$$

where $\eta^k = -\delta(\xi^k, \lambda^k)/2$. Convergence of $\{\lambda^k\}$ to a noncritical multiplier $\bar{\lambda}$ implies that $\eta^k \to 0$ as $k \to \infty$, $\Pi(\bar{\lambda})$ is well-defined, and $\Pi(\cdot)$ is continuous at $\bar{\lambda}$. Therefore, from the first equality in (4.7) we derive the existence of $C > 0$ such that for all k large enough $\|x^{k+1}\| \leq C|\eta^k|\|x^k\|$. Employing this estimate, for any j we have

$$\|x^{k+j}\| \leq C^j \prod_{i=1}^{j} |\eta^{k+j-i}|\|x^k\|,$$

and hence, by the first equality in (4.7),

$$\begin{aligned}
|\langle B\xi^{k+j},\, \xi^{k+j}\rangle| &= \frac{1}{\|x^{k+j}\|^2}|\langle Bx^{k+j},\, x^{k+j}\rangle| \\
&= \frac{\prod_{i=1}^{j}(\eta^{k+j-i})^2}{\|x^{k+j}\|^2}\left|\left\langle B\prod_{i=1}^{j}\Pi(\lambda^{k+j-i})x^k,\, \prod_{i=1}^{j}\Pi(\lambda^{k+j-i})x^k\right\rangle\right| \\
&\geq \frac{1}{C^{2j}}\left|\left\langle B\prod_{i=1}^{j}\Pi(\lambda^{k+j-i})\xi^k,\, \prod_{i=1}^{j}\Pi(\lambda^{k+j-i})\xi^k\right\rangle\right|.
\end{aligned}$$

According to (4.2), the left-hand side of this expression tends to 0 as $k \to \infty$, and therefore, for any accumulation point $\bar{\xi}$ of $\{\xi^k\}$ we have that not only $\langle B\bar{\xi},\, \bar{\xi}\rangle = 0$ (see (4.4)) but

$$\langle B(\Pi(\bar{\lambda}))^j\bar{\xi},\, (\Pi(\bar{\lambda}))^j\bar{\xi}\rangle = 0 \quad \forall\, j = 0,\, 1,\, \ldots. \tag{4.8}$$

We will now provide some further insight into the effect of repulsion from noncritical multipliers for the even more special case when $n = 2$. We emphasize that even this very special case requires a very subtle analysis. Assuming that B is nonsingular, we may suppose without loss of generality that

$$A = \begin{pmatrix} a_1 & a \\ a & a_2 \end{pmatrix}, \quad B = \begin{pmatrix} b_1 & 0 \\ 0 & -b_2 \end{pmatrix},$$

where both $b_1,\, b_2 > 0$. (Observe that B cannot be positive or negative definite, since otherwise there exist no $\bar{\xi} \in \mathbb{R}^n \setminus \{0\}$ satisfying $\langle B\bar{\xi},\, \bar{\xi}\rangle = 0$, i.e., (4.8) with $j = 0$, and in the orthonormal basis of the eigenvectors B takes the needed form.) Then for $\xi \in \mathbb{R}^n \setminus \{0\}$ the equality $\langle B\xi,\, \xi\rangle = 0$ is equivalent to $|\xi_1/\xi_2| = \bar{\tau}$, where $\bar{\tau} = \sqrt{b_2/b_1}$. Moreover,

$$\begin{aligned}
\Pi(\lambda) &= (B^{-1}A + \lambda E)^{-1} \\
&= \begin{pmatrix} \lambda + a_1/b_1 & a/b_1 \\ -a/b_2 & \lambda - a_2/b_2 \end{pmatrix}^{-1} \\
&= \frac{1}{D(\lambda)}\begin{pmatrix} \lambda - a_2/b_2 & -a/b_1 \\ a/b_2 & \lambda + a_1/b_1 \end{pmatrix},
\end{aligned} \tag{4.9}$$

where

$$D(\lambda) = \det(B^{-1}A + \lambda I) = \left(\lambda + \frac{a_1}{b_1}\right)\left(\lambda - \frac{a_2}{b_2}\right) + \frac{a^2}{b_1 b_2}. \tag{4.10}$$

Consider any $\bar{\xi} \in \mathbb{R}^n \setminus \{0\}$ satisfying $\langle B\bar{\xi},\, \bar{\xi}\rangle = 0$. Then the following two cases are possible.

Case 1. If $\bar{\xi}_1/\bar{\xi}_2 = \bar{\tau}$, then according to (4.9), relation (4.8) with $j = 1$ takes the form

$$\left|\frac{(\bar{\lambda} - a_2/b_2)\bar{\tau} - a/b_1}{a\bar{\tau}/b_2 + (\bar{\lambda} + a_1/b_1)}\right| = \left|\frac{(\Pi(\bar{\lambda})\bar{\xi})_1}{(\Pi(\bar{\lambda})\bar{\xi})_2}\right| = \bar{\tau}. \tag{4.11}$$

If modulus in the left-hand side is removed with "+" sign, then $\bar{\xi}$ is an eigenvector of $\Pi(\bar{\lambda})$, and hence, of $B^{-1}A$. By (4.8) with $j = 1$ we then get that there exists $\tilde{\lambda}$ such that

$$\langle A\bar{\xi},\, \bar{\xi}\rangle = \langle B^{-1}A\bar{\xi},\, B\bar{\xi}\rangle = \tilde{\lambda}\langle \bar{\xi},\, B\bar{\xi}\rangle = 0,$$

i.e., $\bar{\xi}$ satisfies not only the second but the first equation in the system

$$\langle A\xi,\, \xi\rangle = 0, \quad \langle B\xi,\, \xi\rangle = 0. \tag{4.12}$$

For generic A and B there exist no nonzero $\bar{\xi}$ with this properties, and in what follows we assume that $B^{-1}A$ does not have eigenvectors satisfying the second equality in (4.12). (If this assumption does not hold, then (4.8) is satisfied with any $\bar{\lambda}$ and with $\bar{\xi} = (\bar{\tau},\, 1)$.)

It remains to consider the case when modulus in the left-hand side of (4.11) is removed with sign "−", i.e.,

$$\frac{(\bar{\lambda} - a_2/b_2)\bar{\tau} - a/b_1}{a\bar{\tau}/b_2 + (\bar{\lambda} + a_1/b_1)} = -\bar{\tau}.$$

This equation uniquely defines

$$\bar{\lambda} = \frac{1}{2}\left(\frac{a_2}{b_2} - \frac{a_1}{b_1}\right), \tag{4.13}$$

which is in agreement with (4.5). This multiplier $\bar{\lambda}$ is noncritical if

$$D(\bar{\lambda}) = -\frac{1}{4}\left(\frac{a_1}{b_1} + \frac{a_2}{b_2}\right)^2 + \frac{a^2}{b_1 b_2} = -\bar{\lambda}^2 + \frac{\det A}{\det B} \neq 0$$

(where we used (4.10)), and this multiplier is the middle of the segment connecting two critical multipliers

$$\bar{\lambda}^{1,\,2} = \bar{\lambda} \pm \sqrt{\bar{\lambda}^2 - \frac{\det A}{\det B}} = \bar{\lambda} \pm \sqrt{-D(\bar{\lambda})}$$

(if they exist, i.e., if $D(\bar{\lambda}) < 0$).

Case 2. If $\bar{\xi}_1/\bar{\xi}_2 = -\bar{\tau}$, then for the same reasons as above, (4.8) may hold with $j = 1$ only provided

$$\frac{(-\bar{\lambda} - a_2/b_2)\bar{\tau} - a/b_1}{-a\bar{\tau}/b_2 + (\bar{\lambda} + a_1/b_1)} \left(= \frac{(R(\bar{\lambda})\bar{\xi})_1}{(R(\bar{\lambda})\bar{\xi})_2}\right) = \bar{\tau},$$

and this equation is satisfied by the same $\bar{\lambda}$ defined in (4.13).

Observe that in Case 1, $R(\bar{\lambda})\bar{\xi}$ fits Case 2 (being considered instead of $\bar{\xi}$), and vice versa. Putting the two cases together, we conclude that under the stated assumptions condition (4.8) is satisfied for the unique $\bar{\lambda}$ defined in (4.13), and for any $\bar{\xi}$ satisfying $\langle B\bar{\xi},\, \bar{\xi}\rangle = 0$, and therefore, this $\bar{\lambda}$ is the unique remaining noncritical candidate for convergence to it. We now demonstrate that convergence to this $\bar{\lambda}$ is also not possible.

Lemma 4.1 *Under the stated assumptions, define the sets*

$$\Lambda_- = \{(\xi,\, \lambda) \in \mathrm{I\!R}^2 \times \mathrm{I\!R} \mid \lambda + \eta(\xi,\, \lambda) \leq \bar{\lambda}\},$$

$$\Lambda_+ = \{(\xi,\, \lambda) \in \mathrm{I\!R}^2 \times \mathrm{I\!R} \mid \lambda + \eta(\xi,\, \lambda) \geq \bar{\lambda}\},$$

$$\Xi_- = \{(\xi,\, \lambda) \in \mathrm{I\!R}^2 \times \mathrm{I\!R} \mid \langle B\Pi(\lambda)\xi,\, \Pi(\lambda)\xi\rangle \leq 0\},$$

$$\Xi_+ = \{(\xi,\, \lambda) \in \mathrm{I\!R}^2 \times \mathrm{I\!R} \mid \langle B\Pi(\lambda)\xi,\, \Pi(\lambda)\xi\rangle \geq 0\}.$$

Then the following assertions are valid:

(i) If

$$a_1\frac{b_2}{b_1} + 2a\sqrt{\frac{b_2}{b_1}} + a_2 > 0, \tag{4.14}$$

then $\Lambda_- = \Xi_+$, $\Lambda_+ = \Xi_-$, *at least near* $((\bar{\tau},\, 1),\, \bar{\lambda})$.

(ii) If

$$a_1\frac{b_2}{b_1} + 2a\sqrt{\frac{b_2}{b_1}} + a_2 < 0, \tag{4.15}$$

then $\Lambda_- = \Xi_-$, $\Lambda_+ = \Xi_+$, *at least near* $((\bar{\tau},\, 1),\, \bar{\lambda})$.

(iii) If

$$a_1\frac{b_2}{b_1} - 2a\sqrt{\frac{b_2}{b_1}} + a_2 > 0, \tag{4.16}$$

then $\Lambda_- = \Xi_+$, $\Lambda_+ = \Xi_-$, *at least near* $((-\bar{\tau},\, 1),\, \bar{\lambda})$.

(iv) If

$$a_1\frac{b_2}{b_1} - 2a\sqrt{\frac{b_2}{b_1}} + a_2 < 0, \tag{4.17}$$

then $\Lambda_- = \Xi_-$, $\Lambda_+ = \Xi_+$, *at least near* $((-\bar{\tau},\, 1),\, \bar{\lambda})$.

It is subsumed that Ξ_- and Ξ_+ may contain only those $(\xi,\, \lambda)$ for which $\Pi(\lambda)$ is well-defined. Similarly, Λ_- and Λ_+ may contain only those $(\xi,\, \lambda)$ for which $\eta(\xi,\, \lambda)$ is well-defined.

Observe that conditions (4.14) and (4.16) hold simultaneously if, and only if, 0 is a strict solution of problem (2.1). Similarly, (4.15) and (4.17) hold simultaneously if, and only if, 0 is a strict solution of the corresponding maximization problem. Moreover,

$$D(\bar{\lambda}) = -\frac{1}{4b_2^2}\left(a_1\frac{b_2}{b_1} + 2a\sqrt{\frac{b_2}{b_1}} + a_2\right)\left(a_1\frac{b_2}{b_1} - 2a\sqrt{\frac{b_2}{b_1}} + a_2\right),$$

and hence, some pair of conditions (4.14), (4.16) or (4.15), (4.17) holds if, and only if, $D(\bar{\lambda}) < 0$, which, in its turn, is equivalent to the existence of critical multipliers. Furthermore, condition $D(\bar{\lambda}) \neq 0$ (meaning noncriticality of $\bar{\lambda}$) is equivalent to saying that one of the conditions (4.14) or (4.15) is satisfied, and one of the conditions (4.16) or (4.17) is satisfied.

This implies, in particular, that condition $D(\bar{\lambda}) \neq 0$ implies that system (4.12) does not have nonzero solutions.

Proof. Suppose that (4.14) holds. Recall that $\bar{\xi} = (\bar{\tau},\, 1)$ satisfies the equality $\langle B\bar{\xi},\, \bar{\xi}\rangle = 0$.

According to (4.6), (4.9), near $(\bar{\xi},\, \bar{\lambda})$ the set Λ_- consists of pairs of the form $((\tau\xi_2,\, \xi_2),\, \lambda) \in \mathbb{R}^2 \times \mathbb{R}$ satisfying the inequality

$$\lambda - \frac{D(\lambda)(b_1\tau^2 - b_2)}{2(b_1(\lambda - a_2/b_2)\tau^2 - 2a\tau - b_2(\lambda + a_1/b_1))} \leq \bar{\lambda}.$$

Employing (4.10) and (4.13), this inequality takes the form

$$\frac{b_1((\lambda - a_2/b_2)\tau - a/b_1)^2 - b_2(a\tau/b_2 + (\lambda + a_1/b_1))^2}{b_1(\lambda - a_2/b_2)\tau^2 - 2a\tau - b_2(\lambda + a_1/b_1)} \leq 0. \qquad (4.18)$$

Consider the denominator of the left-hand side for $\tau = \bar{\tau}$ and $\lambda = \bar{\lambda}$:

$$\begin{aligned} b_1\left(\bar{\lambda} - \frac{a_2}{b_2}\right)\bar{\tau}^2 - 2a\bar{\tau} - b_2\left(\bar{\lambda} + \frac{a_1}{b_2}\right) &= -\frac{1}{2}b_1\left(\frac{a_1}{b_1} + \frac{a_2}{b_2}\right)\frac{b_2}{b_1} - 2a\bar{\tau} - \frac{1}{2}b_2\left(\frac{a_1}{b_1} + \frac{a_2}{b_2}\right) \\ &= -\left(\frac{a_1}{b_1} + \frac{a_2}{b_2}\right)b_2 - 2a\bar{\tau} \\ &= -a_1\bar{\tau}^2 - 2a\bar{\tau} - a_2 \\ &< 0, \end{aligned}$$

where the last inequality is by (4.14). Therefore, for $(\tau,\, \lambda)$ close to $(\bar{\tau},\, \bar{\lambda})$ inequality (4.18) is equivalent to saying that the numerator is nonnegative, which in it turn, according to (4.9), is equivalent to the inclusion $((\tau\xi_2,\, \xi_2),\, \lambda) \in \Xi_+$. This establishes the equality $\Lambda_- = \Xi_+$ near such $(\bar{\xi},\, \bar{\lambda})$. In order to establish the equality $\Lambda_+ = \Xi_-$ it is sufficient to change the inequality sign in (4.18).

The cases when (4.15), (4.16), or (4.17) hold are analyzed along the same lines. ∎

Proposition 4.1 *Under the stated assumptions, if $D(\bar{\lambda}) \neq 0$, then convergence of $\{\lambda^k\}$ to $\bar{\lambda}$ is not possible.*

Proof. For any k set $\tau_k = |\xi_1^k/\xi_2^k|$. Then (4.2) means that $\tau_k \to \bar{\tau}$ as $k \to \infty$.

Recall that condition $D(\bar{\lambda}) \neq 0$ is equivalent to saying that one of conditions (4.14) or (4.15) is satisfied, and one of conditions (4.16) or (4.17) is satisfied. Let, e.g., (4.14) and (4.16) hold. Consider the case when for some k large enough it holds that $\xi_1^k/\xi_2^k = \tau_k$. Then, according to (4.6), (4.9),

$$\eta(\xi^k,\, \lambda^k) = -\frac{D(\lambda^k)(b_1\tau_k^2 - b_2)}{2(b_1(\lambda^k - a_2/b_2)\tau_k^2 - 2a\tau_k - b_2(\lambda^k + a_1/b_1))}.$$

Similarly to the proof of Lemma 4.1, from (4.14) we derive that the denominator here is negative. Hence, the sign of $\eta(\xi^k,\, \lambda^k)$ is opposite to the sign of $b_1\tau_k^2 - b_2$: if $\tau_k < \bar{\tau}$, then $\eta(\xi^k,\, \lambda^k) > 0$, and if $\tau_k > \bar{\tau}$, then $\eta(\xi^k,\, \lambda^k) < 0$.

Therefore, if $\lambda^k \geq \bar{\lambda}$ and $\tau_k < \bar{\tau}$, then λ^{k+1} will be farther from $\bar{\lambda}$ than λ^k, and, in particular, it will hold that $\lambda^{k+1} = \lambda^k + \eta(\xi^k, \lambda^k) \geq \bar{\lambda}$. Then, according to Lemma 4.1 and (4.7), it holds that $\langle B\xi^{k+1}, \xi^{k+1} \rangle < 0$ (the equality is not possible, since it would imply that $x^{k+2} = 0$), i.e., $\tau_{k+1} < \bar{\tau}$. Hence, λ^{k+2} will be even farther from $\bar{\lambda}$, etc., making convergence of $\{\lambda^k\}$ to $\bar{\lambda}$ impossible.

Similarly, and with the same conclusion, one can consider the case when $\lambda^k \leq \bar{\lambda}$ and $\tau_k > \bar{\tau}$.

For any of the remaining possible locations of λ^k and τ_k, from Lemma 4.1 it follows that either $\lambda^{k+1} \geq \bar{\lambda}$ and $\tau_{k+1} < \bar{\tau}$, or $\lambda^{k+1} \leq \bar{\lambda}$ and $\tau_{k+1} > \bar{\tau}$, and according the cases analyzed above, λ^{k+2}, λ^{k+3}, ... will be getting only farther from $\bar{\lambda}$.

The case $\xi_1^k/\xi_2^k = -\tau_k$ is analyzed along the same lines, but employing (4.16). Moreover, the cases when the pairs of conditions (4.14), (4.17), or (4.15), (4.16), or (4.15), (4.17) do hold are analyzed along the same lines. ■

References

[1] M. Golubitsky and D.G. Schaeffer. Singularities and Groups in Bifurcation Theory. V. 1. Springer, Heidelberg, 1985.

[2] M.W. Hirsh. Differential Topology. Springer, New-York, 1976.

[3] A.F. Izmailov. On the analytical and numerical stability of critical Lagrange multipliers. Comput. Math. Math. Phys. 45 (2005), 930–946.

[4] A.F. Izmailov and M.V. Solodov. On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. Math. Program. 117 (2009), 271–304.

[5] A.F. Izmailov and M.V. Solodov. Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. Comput. Optim. Appl. 42 (2009), 231–264.

[6] A.F. Izmailov and M.V. Solodov. On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. Math. Program. 126 (2011), 231–257.

[7] D. Serre. Matrices: Theory and Applications. Second Edition. Springer, New-York, 2010.