Abstract We discuss the convergence of line search methods for minimization. We explain how Newton’s method and the BFGS method can fail even if the restrictions of the objective function to the search lines are strictly convex functions, the line searches are exact and the Armijo condition is satisfied. This explanation illustrates a new way to combine general mathematical concepts and symbolic computation to analyze the convergence of line search methods.

Mathematics Subject Classification (2000) 20E28 · 20G40 · 20C20

1 Introduction

Line search methods are fundamental algorithms in nonlinear programming. Their theory started with Cauchy [3] and they were implemented in the first electronic computers in the late 1940’s and early 1950’s. They have been intensively studied since then and today they are widely used by scientists and engineers. However, their theory is incomplete and M. Powell writes in [11]:

Moreover, theoretical studies have suggested several improvements to algorithms, and they provide a broad view of the subject that is very helpful to research. However, because of the difficulty of analyzing nonlinear calculations, the vast majority of theoretical questions that are important to the performance of optimization algorithms in practice are unanswered...

We address the difficulty of analyzing nonlinear calculations. We focus on the mathematics underlying the convergence of line search methods to solve the un-
constrained minimization problem for a smooth function $F$:

$$\text{minimize } F(x) \text{ for } x \in \mathbb{R}^n. \quad (1)$$

Line search methods are discrete dynamical systems of the form

$$x_{k+1} = x_k + \alpha_k d_k, \quad d_{k+1} = D(F, x_k, \alpha_k, d_k, e_k), \quad e_{k+1} = E(F, x_k, \alpha_k, d_k, e_k), \quad (2)$$

where $\{x_k\} \subset \mathbb{R}^n$ is the sequence we expect to converge to the solution of problem (1) and $e_k$ contains auxiliary information specific to each method. At the $k$th step we choose a search direction $d_k$ and analyze $F$ along the line $\{x_k + wd_k, w \in \mathbb{R}\}$. We search for a step size $\alpha_k \in \mathbb{R}$ such that the sequence $x_k$ satisfies constraints

$$C(F, x_k, \alpha_k, d_k) \leq 0 \quad (3)$$

that are simple and try to force the $x_k$ to converge to a local minimizer of $F$.

The expressions (2)-(3) for important line search methods have symmetries that allow their study in particular but illustrative situations. Since the work of S. Lie in the late 1800’s, symmetries have lead to remarkable solutions of nonlinear differential equations [2]. We explore similar ideas to produce examples with search lines as in Figure 1. The line search methods we discuss are invariant with respect to orthogonal changes of variables and scaling, in the sense that if they assign a step $s_k = x_{k+1} - x_k$ to the point $x_k$ and objective function $F$, $Q$ is an orthogonal matrix and $\lambda \in \mathbb{R}$ then the step $s_k$ corresponding to the objective function $F(x) = F(\lambda^{-1}Qx)$ at the point $x_k = \lambda Qx_k$ is $s_k = \lambda Qs_k$. We argue that in relevant cases these symmetries lead to iterates as in Figure 1.

Besides symmetries, this work is based on a theorem proved by H. Whitney in 1934 [12]. Whitney’s theorem regards the extension of $C^m$ functions from subsets of $\mathbb{R}^n$ to $\mathbb{R}^n$. It says that if a function $F$ and its partial derivatives up to order $m$ are defined in a subset $E$ of $\mathbb{R}^n$ and $F$’s Taylor series up to order $m$ behave properly in $E$ then $F$ can be extended to a $C^m$ function in $\mathbb{R}^n$. Whitney’s theorem is a handy tool to highlight the weak points of nonlinear programming algorithms.

The possibility of cyclic behavior for line search methods was already mentioned by Curry in 1944 [4]. It was also discussed in [5] [8] [9] [10]. Here we go one step further and present a systematic way to build examples that display this behavior. The examples in sections 2 (Newton’s method) and 7 (the BFGS
method) perform exact line searches, satisfy the first Wolfe condition and the restrictions of their objective functions to the search lines are strictly convex, but their iterates have the cyclic asymptotic behavior illustrated in Figure 1. The BFGS and Newton’s methods are among the most important line search methods and our examples refute the following conjecture:

If when applying the BFGS or Newton’s methods we choose the first local minimizer along the search line then the iterates converge to a local minimizer of the objective function.

Section 2 gives a clue to analyze nonlinear line search methods in particular situations as if they were linear. This analysis is not adversely affected by the number of dimensions. To the contrary, as we go to higher dimensions the number of free parameters at our disposal increases. We are then able to observe phenomena contrary to our 2 or 3 dimensional intuition. However, as the experience with Lax Pairs has shown [2], “exploiting symmetry” is easier said than done. The algebraic manipulations necessary to implement our ideas can be overwhelming. Although the example for Newton’s method presented in section 2 is a direct consequence of symmetry, we would not be able to build the example for the BFGS method in section 7 without the software Mathematica. Fortunately, today we have the luxury of tools like Mathematica and can focus on the fundamental geometrical aspects of the line search methods.

Our arguments can be adapted to objective functions with \( m \)th order Lipschitz continuous derivatives or to the more general class \( C^{m,\omega}_{loc}({\mathbb{R}}^n) \) of functions discussed by C. Fefferman in [6], but we do not aim for utmost generality and restrict ourselves to objective functions with Lipschitz continuous second order derivatives, so that we can speak in terms of gradients and Hessians and avoid the use of higher order multilinear forms. On the other hand, [1] indicates that things are different for analytic objective functions, mainly because these functions are “rigid” and it is not possible to change them only locally, or more technically, due to the lack of analytic partitions of the unity.

This work has seven more sections. Section 2 motivates our approach by using it to analyze the convergence of Newton’s method. The technical concepts that formalize our arguments are presented in sections 3 and 4. Section 5 discusses the Armijo and Wolfe conditions and section 6 explains how to build examples in which the objective function is convex along the search lines. In section 7 we combine the results from the previous sections to build an example of divergence for the BFGS method. Finally, in section 8 we prove our claims.

2 Newton’s method

We now describe a family examples of divergence for Newton’s method for minimization. The examples are parameterized by the step size \( \alpha \): given \( \alpha > 0 \) we build an example in which all step-sizes \( \alpha_k \) are equal to \( \alpha \). This section motivates the theory presented later on. Although the geometry underlying the examples is accurately described by figure 1, the algebraic details make they look more complex than they really are. Thus, we suggest that you pay little attention to the formulae and focus on the structure of our argument, which can be summarized as follows:
(a) we guess general expressions for the iterates $x_k$, function values $f_k$, gradients $g_k$ and Hessians $h_k$ which we believe to be compatible with the symmetries in Newton’s method and the theory presented below.

(b) we plug these expressions into the formula that define Newton’s method and obtain equations relating our guesses in item (a).

(c) we solve these equations and the next sections guarantee the existence of an objective function $F$ such that $F(x_k) = f_k$, $\nabla F(x_k) = g_k$, $\nabla^2 F(x_k) = h_k$ and $s_k \nabla^2 F(x_k + w s_k) x_k > 0$ for $w \in \mathbb{R}$ and $k$ big enough.

Following this recipe, we decomposed $\mathbb{R}^6$ as a direct sum of a three dimensional “horizontal” subspace and a three dimensional “vertical” subspace and tried iterates $x_k$, function values $f_k$, gradients $g_k$ and Hessians $h_k$ of the form:

$$
x_k = Q^k D(\lambda^k) x, \quad f_k = \lambda^k \tilde{f}, \quad g_k = \lambda^k Q^k D(\lambda^{-k}) \tilde{g}, \quad h_k = \lambda^k Q^k D(\lambda^{-k}) \tilde{h} D(\lambda^{-k}) Q^{-k},$$

for $\lambda \in (0, 1)$, $x = (\bar{x}^h, \bar{x}^v)$ and $\tilde{g} = (\bar{g}^h, \bar{g}^v)$ with $\bar{x}^h, \bar{x}^v, \bar{g}^h$ and $\bar{v}^v \in \mathbb{R}^3$ and

$$Q = \begin{pmatrix} Q_h & 0 \\ 0 & Q_v \end{pmatrix}, \quad D(\lambda) = \begin{pmatrix} I & 0 \\ 0 & \lambda I \end{pmatrix}, \quad \tilde{h} = \begin{pmatrix} A & C \\ C & 0 \end{pmatrix},$$

where $I$ is the $3 \times 3$ identity matrix, $Q_h$ and $Q_v$ are $3 \times 3$ orthogonal matrices, $A$ is a symmetric $3 \times 3$ matrix and $C$ is a $3 \times 3$ matrix. We then concluded that

$$\lambda = \frac{1}{2}, \quad \bar{x}^h = \bar{x}^v = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad Q_h = \frac{1}{2} \begin{pmatrix} \sqrt{3} & -1 & 0 \\ 1 & \sqrt{3} & 0 \\ 0 & 0 & -2 \end{pmatrix}, \quad Q_v = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

are convenient: they are simple and after picking them we still have the freedom to choose $\tilde{g}, A$ and $C$ in order to satisfy the hypothesis of the theory presented in the next sections and obtain iterates $x_k$ which are consistent with the formula

$$\nabla^2 F(x_k) s_k + \alpha \nabla F(x_k) = 0$$

that defines Newton’s method with step-size $\alpha$. If we replace $\nabla F(x_k)$ and $\nabla^2 F(x_k)$ in (8) by $g_k$ and $h_k$ in (5) then $\lambda$, $Q_h$ and $Q_v$ cancel out and we obtain the equations

$$A s^h + C s^v + \alpha \tilde{g}^h = 0, \quad \text{and} \quad C^t s^h + \alpha \tilde{g}^v = 0,$$

where $s^h = (Q_h - I) x^h$ and $s^v = (\lambda Q_v - I) x^v$.

Notice that, due to the invariance of Newton’s method with respect to orthogonal changes of variables and scaling, there is no “$k$” in (9)–(10). Equations (9) yields

$$\tilde{g}^v = -Cs^h/\alpha$$

Equations (4) and (5) show that $s^h_k g_k = 2^{-k} \bar{g}^h$, $s^h_{k+1} g_{k+1} = 2^{-(k+1)} \bar{g}^h D(2) \bar{g}$, and $f_{k+1} - f_k = -2^{-(k+1)} \tilde{f}$. Therefore, if

$$\tilde{g} \bar{g} < 0, \quad \tilde{g} D(2) \bar{g} = 0 \quad \text{and} \quad \tilde{f} > 0$$
then \( s_k = x_{k+1} - x_k \) is a descent direction, the line searches are exact \( (s_k^T g_{k+1} = 0) \) and the Armijo condition \( f_{k+1} - f_k \leq \sigma s_k^T g_k \) holds for \( 0 < \sigma < \min \{1, -\overline{f}/(2\overline{\sigma} \overline{s})\} \).

We now apply the results from the next sections: items 4c and 4d in the definition of seed in section 4 require that

\[
\overline{g}^0 = C\overline{x}^0 \quad \text{and} \quad \overline{f} = (\overline{g}^0)^T \overline{x}^0
\]

and section 6 says that to guarantee the convexity of the objective function along the search lines we should ask for

\[
2\overline{s}^T \overline{g} < -\overline{f} < 0, \quad \overline{s}^T \overline{h} > 0 \quad \text{and} \quad \overline{s}^T QD(2)\overline{h}D(2)Q^T \overline{s} > 0. \quad (14)
\]

To complete the specification of the terms in (4)–(5) we chose the 9 entries of \( C \) and the 6 independent entries of \( A \) in order to satisfy (9)–(14). These equations and inequalities are linear in \( A \) and \( C \) and the following matrices satisfy them:

\[
A = (2 + \sqrt{3}) \begin{pmatrix}
2 - 2\alpha & 2\alpha & 2(\alpha - 1) \\
2\alpha & 2(\alpha - 1) & 0 \\
0 & 0 & 0
\end{pmatrix},
\]

\[
C = (2 + \sqrt{3}) \begin{pmatrix}
3\alpha - 5\alpha^2 - 4 & 1 - \alpha + 3\alpha^2 & \alpha(1 + 5\alpha) \\
0 & 2 - \sqrt{3} & 0 \\
2 - \sqrt{3} & 0 & 0
\end{pmatrix}. \quad (15)
\]

Equations (4)–(7), (10)–(13) and (15) define iterates and function values, gradients and Hessians of the objective function at them. The next sections guarantee the existence of an objective function \( F \) with Lipschitz continuous second order derivatives such that \( F(x_k) = f_k \), \( \nabla F(x_k) = g_k \) and \( \nabla^2 F(x_k) = h_k \). Moreover, neither the vectors \( D'(0)\overline{s} \) and \( D'(0)\overline{\alpha} \) nor the vectors \( D(0)\overline{s} \) and \( D(0)D(2)\overline{s} \) are aligned and the lines \( L_k = \{ D(0)x_k + wD(0)s_k, w \in \mathbb{R} \} \) are such that \( L_k \cap L_{k'} = \emptyset \) for all \( r + 1 < k < r + 5 \) and theorem 2 and lemmas 1 and 3 in the following sections show that we have much freedom to chose the value of \( F(x) \) along the search segments \( \{ x_k + wsk, w \in [0, 1] \} \); if the function \( \psi : [0, 1] \to \mathbb{R} \) has Lipschitz continuous second order derivatives and

\[
\psi(0) = \overline{f}, \quad \psi(1) = \overline{f}/2, \quad \psi'(0) = \overline{g}^0, \quad \psi'(1) = 0, \quad \psi''(0) = \overline{s}^T \overline{h}, \quad \text{and} \quad \psi''(1) = \overline{s}^T QD(2)\overline{h}D(2)Q^T \overline{s}/2
\]

then \( F \) can be chosen so that \( F(x_k + wsk) = (1/2)^k \psi(w) \) for \( w \in [0, 1] \) and \( k \) large. In fact, condition (14) and theorem 3 in section 6 show that \( F \) can be chosen so that \( s_k^T F(x_k + wsk)x_k > 0 \) for \( w \in \mathbb{R} \) and \( k \) large. Therefore, the recipe above leads to the example mentioned in the introduction.

### 3 Flowers and Dandelions

We now present a framework to apply Whitney’s theorem to study the convergence of line search methods. We describe examples in which the iterates \( x_k \) and the function values \( f_k \) the gradients \( g_k \) and the Hessians \( h_k \) of the objective function are grouped into \( p \) converging subsequences, which we call petals. The limits
of these subsequences \( \{x_k\} \), \( \{f_k\} \), \( \{g_k\} \) and \( \{h_k\} \) are the members of periodic sequences \( \{x_k\} \), \( \{\phi_k\} \), \( \{\gamma_k\} \) and \( \{\theta_k\} \), so that \( \lim_{n \to \infty} x_{pq+r} = x_r \) for all \( r \) and \( x_{r+p} = x_r \). \( \lim_{n \to \infty} f_{pq+r} = \phi_r \) for all \( r \) and \( \phi_{r+p} = \phi_r \). \( \gamma_{r+p} = \gamma_r \) and \( \theta_{r+p} = \theta_r \). When the iterates \( x_k \) in each petal approach their limits along well defined directions we say that the flower is a dandelion. In formal terms:

\[
\frac{\nabla}{\nabla} \text{and compatible with the dandelion}
\]

Fig. 2 A flower and a dandelion with \( p = 6 \) petals

**Definition 1** A flower \( \mathcal{F}(n, p, \lambda, x_k, f_k, g_k, h_k) \) is a collection formed by

1. \( \lambda \in (0, 1) \) and positive integers \( n \) and \( p \).
2. Sequences \( \{x_k\} \) and \( \{\gamma_k\} \) in \( \mathbb{R}^n \) and a constant \( M > 0 \) such that
   (a) \( x_i = x_j \Leftrightarrow i = j \).
   (b) \( \gamma_i = \gamma_j \Leftrightarrow i \equiv j \) mod \( p \).
   (c) \( \lambda^k \leq M|x_k - \gamma_k| \leq \lambda^{2k} \).
3. Sequences \( \{f_k\}, \{\phi_k\} \subset \mathbb{R}, \{g_k\}, \{\gamma_k\} \subset \mathbb{R}^n \) and \( \{h_k\}, \{\theta_k\} \subset \mathbb{H}^n \) such that
   \[
   \phi_{k+p} = \phi_k, \quad \gamma_{k+p} = \gamma_k, \quad \theta_{k+p} = \theta_k, 
   \]
   \[
   \|h_k - \theta_k\| \leq M\lambda^k, 
   \]
   \[
   g_k - \gamma_k - \theta_k(x_k - \gamma_k) \| \leq M\lambda^{2k}, 
   \]
   \[
   f_k - \phi_k - \gamma_k(x_k - \gamma_k) - \frac{1}{2}(x_k - \gamma_k)^t h_k(x_k - \gamma_k) \| \leq M\lambda^{3k} 
   \]
   where \( \mathbb{H}^n \) is the set of \( n \times n \) symmetric matrices

The next definitions and theorems relate the \( f_k \), \( g_k \) and \( h_k \) in a flower to an objective function \( F \) with Lipschitz continuous second derivatives:

**Definition 2** Suppose \( U \subset \mathbb{R}^n \) and \( V \subset \mathbb{R}^p \). We define \( \operatorname{Lip}^m(U, V) \) as the space of functions \( F : U \to V \) with Lipschitz continuous \( m \)th derivatives. If \( V = \mathbb{R} \) then we call this space simply by \( \operatorname{Lip}^m(U) \)

**Theorem 1** Given a flower \( \mathcal{F}(n, p, \lambda, x_k, f_k, g_k, h_k) \) there exists \( F \in \operatorname{Lip}^2(\mathbb{R}^n) \) such that \( F(x_k) = f_k, \nabla F(x_k) = g_k \) and \( \nabla^2 F(x_k) = h_k \) for all \( k \)

If the flower is a dandelion then we can improve this result and specify the objective function and its derivatives along the segments \( [x_k + w_{x_k}, w \in [0, 1]] \):

**Theorem 2** If the functions \( \{F_k\}, \{G_k\} \) and \( \{H_k\} \) and the intervals \( \{[a_k, b_k]\} \) are compatible with the dandelion \( \mathcal{F}(m, n, p, \lambda, x_k, f_k, g_k, h_k) \) then there exists \( k_0 \in \mathbb{N} \) and \( F \in \operatorname{Lip}^2(\mathbb{R}^n) \) such that \( F(x_k + w_{x_k}) = F_k(w, \lambda^k), \nabla F(x_k + w_{x_k}) = G_k(w, \lambda^k) \) and \( \nabla^2 F(x_k + w_{x_k}) = H_k(w, \lambda^k) \) for \( k > k_0 \) and \( w \in [a_k, b_k] \)
This theorem will make sense after you read the following definitions:

**Definition 3** A flower \( \mathcal{F}(n, p, \lambda, x_k, f_k, g_k, h_k) \) is a dandelion if there exist functions \( X_k \in \text{Lip}^2([0, 1], \mathbb{R}^p) \) such that, for all \( k \) and \( S_k(z) = X_{k+1}(z) - X_k(z) \),

(a) \( X_{k+p} = X_k \) and \( x_k = X_k(\lambda^k) \).
(b) The vectors \( S_k(0), X_k'(0) \) and \( X_k'(0) \) are linearly independent.
(c) The vectors \( S_k(0), S_k(0) \) and \( X_k'(0) \) are linearly independent.

**Definition 4** The intervals \([a_k, b_k]\), defined for \( k \in \mathbb{N} \), are compatible with a dandelion if

(a) \( a_k = a_{k+p} \leq 0 \) and \( b_k = b_{k+p} = b_k \geq 1 \) for \( k \in \mathbb{N} \).
(b) The segments \( \{X_k(0) + wS_k(0), w \in [a_k, b_k]\} \) and \( \{X_k(0) + wS_k(0), w \in [a_k, b_k]\} \)

are disjoint if \( r + 1 < k < r + p - 1 \)

**Definition 5** For all \( k \in \mathbb{N} \), consider functions \( F_k \in \text{Lip}^2([0, 1], \mathbb{R}^n) \) and \( H_k \in \text{Lip}^1([0, 1], \mathbb{R}^n) \) and \( G_k \in \text{Lip}^1([0, 1], \mathbb{R}^n) \) and \( X_k \in \text{Lip}^0([0, 1], \mathbb{R}^n) \). We say that \( \{F_k\}, \{G_k\} \) and \( \{H_k\} \) are compatible with a dandelion \( \mathcal{D} \) if \( F_{k+p} = F_k, G_{k+p} = G_k \) and \( H_{k+p} = H_k \) and there exists \( k_0 \) such that if \( k > k_0, i \in \{0, 1\}, w \in \mathbb{R} \) and \( z = \lambda^k \) then

\[
F_k(i, z) = f_k,i, \quad G_k(i, z) = g_k,i, \quad H_k(i, z) = h_k,i ,
\]

\[
\frac{\partial F_k}{\partial w}(w, z) = G_k(w, z)'s_k, \quad \frac{\partial F_k}{\partial z}(w, 0) = G_k(w, 0)'V_k(w),
\]

\[
\frac{\partial^2 F_k}{\partial z^2}(w, 0) = \frac{\partial G_k}{\partial w}(w, 0)'V_k(w) + G_k(w, 0)'U_k(w),
\]

\[
\frac{\partial G_k}{\partial w}(w, z) = H_k(w, z)s_k, \quad \frac{\partial G_k}{\partial z}(w, 0) = H_k(w, 0)V_k(w)
\]

for \( V_k(w) = (1 - w)X_k'(0) + wX_{k+1}'(0) \) and \( U_k(w) = (1 - w)X_k''(0) + wX''_{k+1}(0) \)

If we do not care about the derivatives of \( F \) in directions normal to the segments \( \{x_k + ws_k, w \in [0, 1]\} \) then the search for compatible functions \( F_k, G_k \) and \( H_k \) is simplified by the following lemma

**Lemma 1** Let \( \mathcal{D}(m, n, p, \lambda, x_k, f_k, g_k, h_k) \) be a dandelion and, for \( k \in \mathbb{N} \), functions \( F_k \in \text{Lip}^2([0, 1], \mathbb{R}^n) \) for which \( F_{k+p} = F_k \). If there exist \( k_0 \in \mathbb{N} \) such that

\[
F_k(i, \lambda^k) = f_k,i, \quad \frac{\partial F_k}{\partial w}(i, \lambda^k) = g_k,i, \quad \frac{\partial^2 F_k}{\partial w^2}(i, \lambda^k) = h_k,i
\]

for \( i \in \{0, 1\} \) and \( k > k_0 \) then there exist functions \( G_k \in \text{Lip}^1([0, 1], \mathbb{R}^n) \) and \( H_k \in \text{Lip}^0([0, 1], \mathbb{R}^n) \) such that \( \mathcal{D} \) and \( \{F_k\}, \{G_k\} \) and \( \{H_k\} \) are compatible

Therefore, we can focus on \( F_k(w, z) \) and neglect \( H_k(w, z) \) and \( G_k(w, z) \) for \( w \notin \{0, 1\} \). In the next section we describe a class of dandelions for which we can restrict ourselves to functions \( F_k \) of the form \( F_k(w, z) = e^{\alpha_k} \psi_k(w) \).
4 Symmetric dandelions and their seeds

In this section we present a family of dandelions that includes the examples for the BFGS and Newton’s methods mentioned in the introduction. These dandelions combine symmetries imposed by an orthogonal matrix \( Q \) with contractions dictated by a diagonal matrix \( D \). They are defined by their seeds:

**Definition 6** A seed \( \mathcal{S}(n, p, d, Q, x_k, T_k, \bar{x}_k, \bar{y}_k) \) is a collection formed by

1. \( n, p \in \mathbb{N} \) and \( d = (d_1, \ldots, d_n) \in \mathbb{N}^n \) such that \( p \geq 3 \) and \( d_n = \max d_i \).
2. A \( n \times n \) orthogonal matrix \( Q \) such that \( Q^T = I \) and the diagonal matrix \( D(z) \) with its diagonal entry equal to \( z^j \) commutes with \( Q \) for \( z \in \mathbb{R} \).
3. A sequence \( \{x_k\} \subset \mathbb{R}^n \) such that \( x_{k+p} = x_k \), the points \( x_r = D(0)Q^r x_0 \) are distinct for \( 0 \leq r < p \) and the vectors

\[
\bar{x}_k = D(\lambda)Q_k{x}_{k+1} - x_k
\]  

(25)

are such that, for \( 0 \leq r < p \), the vectors \( D(0)x_r \) and \( D'(0)x_r \) are not aligned and neither are the vectors \( D(0)x_r \) and \( D'(0)x_{r+1} \).
4. Sequences \( \{T_k\} \subset \mathbb{R}, \{\bar{g}_k\} \subset \mathbb{R}^n \text{ and } \{\bar{h}_k\} \subset \mathbb{H}^n \) such that

(a) \( T_{k+p} = T_k, \bar{g}_{k+p} = \bar{g}_k \text{ and } \bar{h}_{k+p} = \bar{h}_k \) for all \( k \).
(b) \( (\bar{h}_k)_{ij} = 0 \) if \( d_i + d_j > d_n \).
(c) If \( d_i = d_n - 1 \) and \( J = \{j | d_j > 0\} \) then

\[
(\bar{g}_k)_i = \sum_{j \in J} (\bar{h}_k)_{ij}(\bar{x}_k)_j.
\]  

(26)

(d) If \( d_n \leq 2 \), \( L = \{l | d_l = d_n \} \) and \( S = \{(i, j) | d_i + d_j = d_n, d_i > 0, d_j > 0\} \) then

\[
\bar{T}_k = \sum_{l \in L} (\bar{g}_k)_l(\bar{x}_k)_l + \frac{1}{2} \sum_{(i,j) \in S} (\bar{h}_k)_{ij}(\bar{x}_k)_i(\bar{x}_k)_j.
\]  

(27)

A seed and \( \lambda \in (0, 1) \) generate a dandelion through the formulae

\[
X_k(z) = Q^k D(z)x_k, \quad x_k = X_k(\lambda^k), \quad \chi_k = X_k(0),
\]  

(28)

\[
f_k = \lambda^{kd_n} T_k, \quad g_k = \lambda^{kd_n} Q^k D(\lambda^{-k})\bar{x}_k, \quad \gamma_k = Q^k \bar{g}_k,
\]  

(29)

\[
h_k = \lambda^{kd_n} D(\lambda^{-k})Q^k \bar{h}_k Q^{-k} D(\lambda^{-k}), \quad \theta_k = Q^k \bar{h}_k Q^{-k},
\]  

(30)

where

\[
(g^w_k)_i = (\bar{g}_k)_i \quad \text{if } d_i = d_n \quad \text{and} \quad (g^w_k)_i = 0 \quad \text{otherwise},
\]  

(32)

\[
(h^w_k)_{ij} = (\bar{h}_k)_{ij} \quad \text{if } d_i + d_j = d_n \quad \text{and} \quad (h^w_k)_{ij} = 0 \quad \text{otherwise}.
\]  

(33)

This result is formalized by the following lemma:

**Lemma 2** If \( \mathcal{S}(n, p, d, Q, x_k, T_k, \bar{x}_k, \bar{y}_k) \) is a seed and \( \lambda \in (0, 1) \) then there exists a unique dandelion \( \mathcal{D}(\mathcal{S}) \) with \( x_k, f_k, g_k, h_k, \phi_k, \gamma_k \text{ and } \theta_k \) as in (28)–(33). Moreover, \( \mathcal{D}(\mathcal{S}) \)’s steps \( s_k = x_{k+1} - x_k \) are given by

\[
s_k = Q^k D(\lambda^k)\bar{x}_k.
\]  

(34)

for \( \bar{x}_k \) defined in (25) \( \square \)
It is easy to apply lemma 1 and theorem 2 to $\mathcal{D}(\mathcal{A})$:

**Lemma 3** If $\mathcal{A}(n, p, d, Q, \lambda, f_k, g_k, h_k)$ is a seed, $\lambda \in (0, 1)$, $x_k$, $f_k$, $g_k$ and $h_k$ are given by (28)–(31) and the functions $\psi_r \in \text{Lip}^2(\mathbb{R})$, defined for $0 \leq r < p$, satisfy

$$
\psi_r(i) = f_{r+i}, \quad \psi_r'(i) = s'_r g_{r+i}, \quad \psi_r''(i) = s''_r h_{r+i}, \quad (35)
$$

for $i \in \{0, 1\}$, then the functions $F_k(w, z) = \varepsilon^d w \psi_k \text{mod} p(w)$ satisfy (24).

To build a symmetric dandelion and specify the value of the objective function along the search segments $\{x_k + ws_k, w \in [0, 1]\}$, it is enough to find the right seed and functions $\psi_r$. Once we find them, the existence of an objective function $F$ with $\nabla F(x_k) = g_k$, $\nabla^2 F(x_k) = h_k$ and $F(x_k + ws_k) = \lambda^{kw} \psi_k \text{mod} p(w)$ for $x_k$, $f_k$, $g_k$ and $h_k$ in (28)–(31) is guaranteed by lemma 3 and theorem 2. To find a seed we proceed as in section 2: we plug (28)–(31) into the expressions that define

(a) the method of our interest,
(b) the compatibility conditions in the definition of seed and theorem 2,
(c) additional constrains, like the Armijo and Wolfe conditions in section 5 and the convexity conditions in section 6.

and analyze the result. If equations (28)–(31) are compatible with the method’s symmetries, as they are for the BFGS and Newton’s methods, then we have a chance of handling these constrains and may even find closed form solutions for them. If equations (28)–(31) are not related to the method’s symmetries then the dandelion will not bloom.

## 5 The Armijo and Wolfe conditions

In this section we show that the Armijo condition

$$
f_{k+1} - f_k < \sigma s'_k g_k \tag{36}
$$

and the Wolfe condition

$$
s'_k g_{k+1} > \beta s'_k g_k \tag{37}
$$

may not prevent the cyclic behavior illustrated in figures 1 and 2. In fact, we can have cyclic behavior even if we replace the Wolfe condition by the stronger requirement $s'_k g_{k+1} = 0$, which is called exact line search condition. These conditions are invariant with respect to orthogonal changes of variables and scaling and can be easily checked for the dandelions coming from a seed:

**Lemma 4** Let $\mathcal{A}(n, p, d, Q, x_k, \bar{T}_k, \bar{g}_k, \bar{h}_k)$ be a seed and $\mathcal{D}(m, n, \lambda, x_k, f_k, g_k, h_k)$ the corresponding dandelion.

(a) If $\bar{\sigma}_{r+1} = 0$ for $0 \leq r < p$ then $s'_r g_{k+1} = 0$ for all $k$.

(b) If $\bar{\sigma}_r < 0$ for $0 \leq r < p$ then the Armijo condition holds for

$$
\sigma < \sigma_0 = \min_{0 \leq r < p} \frac{\bar{T}_{r+1} - \bar{T}_r}{\bar{s}^2_{r+1}}.
$$

(c) If $\bar{\sigma}_r < 0$ for $0 \leq r < p$ then the Wolfe condition is verified for

$$
\beta > \beta_0 = \max_{0 \leq r < p} \frac{\bar{\sigma}_{r+1}}{\bar{\sigma}_{r}} \square.
6 Convexity along the search lines

Convexity is an important simplifying assumption in optimization. It is tempting to conjecture that strict convexity along the search lines guarantees the convergence of line search methods \(^1\). Sometimes even stronger conjectures are made, like having convergence if we choose a global minimizer or the first local minimizer along the search line. We now show that strict convexity along the search lines does not rule out the cyclic behavior depicted in figures 1 and 2. To do that, we present a theorem that yields an objective function which is strictly convex along the search lines of a dandelion that comes from a seed:

**Theorem 3** Let \(\mathcal{S}(n, p, d, Q, x_0, f_k, g_k, h_k)\) be a seed, \(\lambda \in (0, 1)\) and \(x_k, f_k, g_k\) and \(h_k\) be given by (28)-(31). Let \(\mathcal{R}\) be the convex hull of \(\{x_k, 0 \leq k \leq p\}\). Consider the lines \(\Sigma_k = \{x_k + w(x_{k+1} - x_k), w \in \mathcal{R}\}\) and assume that \(\Sigma_r \cap \Sigma_k \cap \mathcal{R} = \emptyset\) for \(r + 1 < k < r + p - 1\). If, for \(0 \leq r < p\),

\[ s^r_k < s^r_{r+1} - f_r < s^r_{r+1}g^r_{r+1}, \quad s^r_h_r - s_r > 0 \quad \text{and} \quad s^r_h_{r+1} - s_r > 0 \quad (38) \]

then there exist a function \(F \in \text{Lip}^2(\mathbb{R}^n)\) and \(k_0\) such that if \(k > k_0\) then \(F(x_k) = f_k\), \(\nabla F(x_k) = g_k\), \(\nabla^2 F(x_k) = h_k\) and \(s^r_k\nabla^2 F(x_k + w(x_k))s_k > 0\) for all \(w \in \mathcal{R}\). \(\square\)

If the vector \(d\) that defines the seed \(\mathcal{S}\) has more than two entries equal to 0 then the lines \(\Sigma_r\) and \(\Sigma_k\) and \(r + 1 < k < r + p - 1\) are disjoint for almost all choices of the points \(x_k\). Therefore, to build examples of divergence in which the objective function is strictly convex along the search lines we can focus on (38) and check the intersections later, just to make sure we were not unlucky. This is what we did in section 2 and will do again in the next section.

7 The BFGS method

In this section we present an example of divergence for the BFGS method in which the objective function is strictly convex along the search lines. The essence of this example is already present in [9] and we suggest that you read this reference as an introduction to this section. Our purpose here is to show that the concepts of flower, dandelion and seed reduce the construction of examples like the one in [9] to the solution of algebraic problems. Software like Mathematica or Mapple can decide if such algebraic problems have solutions and find accurate approximation to these solutions. The validation of the example becomes then a question of using these approximate solutions wisely to check the requirements in the definitions of flower, dandelion and seed and the hypothesis of the lemmas and theorems in the previous sections.

We analyze the BFGS method with exact line searches, i.e., \(s^r_k g_{k+1} = 0\). In this case the Hessian approximations \(B_k\) are updated according to the formula:

\[ B_{k+1} = B_k + \frac{\alpha_k}{s^r_k g_k} g_k s^r_k - \frac{1}{s^r_k g_k} (g_{k+1} - g_k)(g_{k+1} - g_k)^T, \quad (39) \]

\(^1\) by strict convexity along the search line we mean that the directional second derivatives \(s^r_k \nabla^2 F(x_k + w(x_k))s_k\) are positive.
where $g_k = \nabla F(x_k)$. The iterates $x_k$ evolve according to

$$s_k = -\alpha_k B_k^{-1} g_k \quad \text{and} \quad x_{k+1} = x_k + s_k.$$  \hspace{1cm} (40)

Equations (39)–(40) are invariant with respect to orthogonal changes of variables and scaling, in the sense that if $Q$ is an orthogonal matrix, $\lambda \in \mathbb{R}$ and $F, B_k$ and $x_k$ satisfy equations (39)–(40) then $F(x) = F(\lambda^{-1} Q x)$, $B_k = \lambda^{-2} Q B_k Q^T$ and $\tau_k = \lambda Q x_k$ satisfy them too. It is hard to exploit these symmetries because the BFGS method was conceived to correct the matrices $B_k$. However, it has an additional symmetry: if we take $B_k$ of the form

$$B_k = -\sum_{i=0}^{n-1} \alpha_{k+i}^\prime s_{k+i}^T g_k s_{k+i},$$  \hspace{1cm} (41)

combined with the conditions

$$s_{k+j}^T g_k < 0 \quad \text{and} \quad s_{k+j}^T g_k = 0 \quad \text{for} \quad 1 \leq j < n \quad \text{and} \quad k \in \mathbb{N},$$  \hspace{1cm} (42)

then equation (40) is automatically satisfied and (39) holds if

$$g_{k+n} = \rho_k (g_{k+1} - g_k),$$  \hspace{1cm} (43)

for $\rho_k$ such that

$$\alpha_{k+n} = \frac{s_{k+n}^T g_{k+n}}{s_k^T g_k} \rho_k^2.$$  \hspace{1cm} (44)

Notice that if we assume (42) and take $\alpha_k = 1$ for $0 \leq k < n$, $\alpha_k$ in (44) for $k \geq n$ then the $\alpha_k$ are positive and the vectors $g_k, \ldots, g_{k+n-1}$ are linearly independent, because if we suppose that $\sum_{i=0}^{n-1} \mu_j g_{k+j} = 0$ with $\mu_0 = 1$ then (42) leads to the contradiction $0 = s_j^T \sum_{i=0}^{n-1} \mu_j g_{k+j} = s_j^T g_k < 0$. Therefore, under (42)–(44) the matrices $B_k$ defined by (41) are positive definite and to build an example of divergence for the BFGS method we can ignore $\alpha_k$ and $B_k$ and focus on (42)–(43).

To apply the theory above we plug (28)–(31) into (42) and (43) and deduce that a seed $\mathcal{G}(n, p, d, Q, \tau_k, \pi_k, \bar{g}_k, \bar{h}_k)$ is consistent with the BFGS method if

$$\bar{z}_k^T g_k < 0 \quad \text{and} \quad \bar{z}_k^T D(\lambda^{-1}) Q^T \bar{g}_{k+j} = 0 \quad \text{for} \quad 1 \leq j < n,$$  \hspace{1cm} (45)

$$Z_k^T g_{k+n} = \rho_k (Z_k^T g_{k+1} - \bar{g}_k),$$  \hspace{1cm} (46)

where $Z = \lambda^{\psi_k} D(\lambda^{-1}) Q$. If we fix $Z$ then (46) becomes similar to an eigenvalue problem, with the $\rho$’s playing the role of eigenvalues and the $\pi$’s of eigenvectors. We solved equations (46) for $\pi_k$ and $\rho_k$ in the case

$$Q = \begin{pmatrix} \cos \frac{\pi}{3} & -\sin \frac{\pi}{3} & 1 \\ \sin \frac{\pi}{3} & \cos \frac{\pi}{3} & -1 \\ 1 & 1 & 1 \end{pmatrix}, \quad D(\lambda) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \lambda & \lambda^2 \\ 1 & \lambda^2 & \lambda^3 \end{pmatrix},$$

$\lambda = \sqrt[3]{0.9}$ and $n = 6$, as described in the lemma.
Lemma 5 If $Q$ and $D(\lambda)$ are the matrices above, $d_n = 3$, $n = 6$ and $\lambda = \sqrt[3]{0.9}$ then there exist vectors $\overline{g}_k \in \mathbb{R}^6$ and coefficients $\rho_k \in \mathbb{R}$ that satisfy (46) and are such that $\overline{g}_{k+1} = \overline{g}_k$, $\overline{p}_{k+1} = \overline{p}_k$ and the six vectors in the set

$$\left\{ D(\lambda^{-j})Q^j\overline{g}_k, \; j = 0, 1, \ldots, 5 \right\} \quad (47)$$

are linearly independent for each $k \in \mathbb{N}$. □

The linear independence of the vectors $D(\lambda^{-j})Q^j\overline{g}_k$ in lemma 5 implies that we can solve the following linear systems of equations on the vectors $\overline{g}_k$:

$$\overline{g}_k = -1 \quad \text{and} \quad \overline{g}_kD(\lambda^{-j})Q^j\overline{g}_k = 0 \quad \text{for} \; 1 \leq j < 6. \quad (48)$$

The vectors $\overline{g}_k$ obtained from (48) define steps $s_k$ by (34):

$$s_k = X^k\overline{g}_k \quad \text{for} \quad X = D(\lambda)Q$$

and we claim that the points

$$\overline{x}_0 = - (I - X^{11})^{-1} \sum_{j=0}^{10} X^j, \quad (49)$$

$$\overline{x}_k = X^{-k} \left( \overline{x}_0 + \sum_{j=0}^{k-1} X^j \overline{g}_j \right) \quad \text{for} \; k > 0 \quad (50)$$

satisfy

$$\overline{x}_{k+66} = \overline{x}_k \quad (51)$$

and lead to points $x_k = X^k\overline{g}_k$ such that $s_k = x_{k+1} - x_k$. In fact, using (50) it is straightforward to deduce that $x_{k+1} - x_k = X^{k+1}\overline{g}_{k+1} - X^k\overline{g}_k = X^k\overline{g}_k = s_k$. To verify (51), notice that (50) yields

$$\overline{x}_{k+66} = X^{-k} \left( \overline{x}_0 + \sum_{j=66}^{k+65} X^j \overline{g}_j + X^{-66} \Delta \right) \quad (52)$$

where

$$\Delta = (I - X^{66})\overline{x}_0 + \sum_{j=0}^{65} X^j \overline{g}_j. \quad (53)$$

Now, since $\overline{g}_{k+1} = \overline{g}_k$, equation (48) implies that $\overline{x}_{k+1} = \overline{x}_k$ and

$$\sum_{j=0}^{65} X^j \overline{g}_j = \left( \sum_{r=0}^{5} X^{11r} \right) \sum_{j=0}^{10} \overline{g}_j = (I - X^{11})^{-1} (I - X^{66}) \sum_{j=0}^{10} \overline{g}_j. \quad (54)$$

Combining this equation, (49) and (53) we conclude that $\Delta = 0$ and equation (52) and the identity $\overline{x}_{j-66} = \overline{x}_j$ lead to (51).

In the proof of lemma 5 we show that the $\overline{g}_k$ can be chosen so that the vectors $\overline{g}_k$ defined by (48) satisfy the linear independence requirements in the definition of seed. The $\overline{g}_k$ and $\overline{p}_k$ above, $d = (0, 0, 0, 1, 1, 3)$,

$$\overline{f}_k = 1 \quad \text{and} \quad \overline{h}_k = D'(0)\overline{g}_k X' \overline{g}'(0) \quad (54)$$
are also compatible with this definition. As a consequence, \( n = 6, p = 66, d, Q \) \( \eta_k, J_k, g_k \) and \( \bar{F}_k \) define a seed \( \mathcal{D} \). This seed leads to a dandelion \( \mathcal{D}(F) \) with steps \( s_k \) and gradients \( g_k \) compatible with the BFGS method with the matrices \( B_k \) in (39) and step sizes \( \alpha_k \) in (44). Equations (28)–(31) and (48) imply that the function values and gradients associated to \( \mathcal{D}(F) \) satisfy

\[
s_k^t g_k = -\lambda^3 s_k^t - \lambda^{3k+3} = f_{k+1} - f_k < 0 = s_k^t g_{k+1}.
\]

Lemma 4 shows that the iterates \( x_k \) corresponding to \( \mathcal{D} \) satisfy the Armijo condition for \( 0 < \sigma < 1 - \lambda^3 \). Moreover, almost all choices of the vectors \( \bar{x}_k \) in the proof of lemma 5 lead to lines \( L_k = \{ D(0)Q^k \bar{x}_k + wD(0)Q^k, \, w \in \mathbb{R} \} \) such that \( L_k \cap L\_r = \emptyset \) for \( r + 1 < k < r + 65 \) and \( s_k^t g_{k+1} \neq 0 \). Equation (54) shows that \( s_k^t h_k s_k > 0 \) and \( s_k^t h_{k+1} s_k > 0 \). As a result, equation (55) and theorem 3 yield an objective function \( F \in \text{Lip}^2(\mathbb{R}^n) \) such that the iterates \( x_k \) generated by applying the BFGS method to \( F \) with \( x_0 = x_0 \) and \( B_k \) above satisfy \( s_k^t \nabla^2 F(x_k + w_k s_k) s_k > 0 \) for \( k \) large and \( w \in \mathbb{R} \).

This completes the presentation of an example showing that there is not enough strength in the definition of the BFGS method, the exact line search condition and the Armijo condition to prevent the cyclic behavior in figures 1 and 2, even when the objective function is strictly convex along the search lines.

8 Proofs

We now prove the results stated in the previous sections. Our main tool is the following corollary of the Whitney’s extension theorem:

**Lemma 6** Let \( E \) be a bounded subset of \( \mathbb{R}^n \) and suppose \( F: E \to \mathbb{R}, G: E \to \mathbb{R}^n \) and \( H: E \to \mathbb{R}^m \) are functions with domain \( E \) and \( M \) is a constant. If

\[
\| H(x) - H(y) \| \leq M \| y - x \|, \tag{56}
\]

\[
\| G(y) - G(x) - H(x)(y - x) \| \leq M \| y - x \|^2, \tag{57}
\]

\[
| F(y) - F(x) - G(x)^t (y - x) - \frac{1}{2} (y - x)^t H(x)(y - x) | \leq M \| y - x \|^3 \tag{58}
\]

then there exists \( F \in \text{Lip}^2(\mathbb{R}^n) \) such that \( F(x) = F(x), \nabla F(x) = G(x) \) and \( \nabla^2 F(x) = H(x) \) for \( x \in E \). \( \square \)

We can apply Whitney’s theorem to a dandelion \( \mathcal{D} \) with \( x_k = X_k(\lambda^k) \) because if

\[
S_k(z) = X_{k+1}(z) - X_k(z), \quad \overline{X}_k(w, z) = X_k(z) + wS_k(z) \tag{59}
\]

and \( \delta > 0 \) is small and the intervals \( \{ [a_k, b_k] \} \) are compatible with \( \mathcal{D} \) then the distance between points in the 2 dimensional surface

\[
\bigcup_{k=0}^{\infty} \{ \overline{X}_k(w, z), \, w \in [a_k, b_k] \text{ and } z \in [0, \delta] \}
\]

can be estimated in terms of \( w \) and \( z \).
Lemma 7 Given a dandelion with $x_k = X_k(\lambda^k)$, compatible intervals $[a_k, b_k]$ and $\overline{X}_k$ in (59) there exists $\delta > 0$ such that if $w_j, w_k \in [a_k, b_k]$, $z_j, z_k \in [0, \delta]$, $y_j = \overline{X}_j(w_j, z_j)$, $y_k = \overline{X}_k(w_k, z_k)$ and $\|y_j - y_k\| \leq \delta$ then either

(a) $j \equiv k \mod p$,  \hspace{1cm} (b) $j \equiv k + 1 \mod p$ \hspace{1cm} or \hspace{1cm} (c) $k \equiv j + 1 \mod p$ \hspace{1cm} (60)

and in case (a)

$$\|y_j - y_k\| \geq \delta (|w_j - w_k| + |z_j - z_k|),$$ \hspace{1cm} (61)

in case (b)

$$\|y_j - y_k\| \geq \delta (|w_j| + |z_j - z_k| + |1 - w_k|)$$ \hspace{1cm} (62)

and in case (c)

$$\|y_j - y_k\| \geq \delta (|1 - w_j| + |z_j - z_k| + |w_k|) \square$$ \hspace{1cm} (63)

We also use the next lemmas. After the statement of these lemmas we prove the theorems and the paper ends with the proofs of the lemmas.

Lemma 8 Consider $E \subset \mathbb{R}^n$, constants $K > 1$, $\delta > 0$ and functions $F : E \to \mathbb{R}$, $G : E \to \mathbb{R}^n$ and $H : E \to \mathbb{H}^n$. If for all $x, z \in E$ there exist $m \in \mathbb{N}$ and $y_1, \ldots, y_m \in E$ such that, for $y_0 = x$ and $y_{m+1} = z$,

$$\sum_{i=0}^{m} \|y_{i+1} - y_i\| \leq K\|x - z\|,$$ \hspace{1cm} (64)

$$\|H(y_i) - H(y_{i+1})\| \leq K\|y_{i+1} - y_i\|,$$ \hspace{1cm} (65)

$$\|G(y_{i+1}) - G(y_i) - \frac{1}{2}(H(y_{i+1}) + H(y_i))(y_{i+1} - y_i)\| \leq K\|y_{i+1} - y_i\|^2,$$ \hspace{1cm} (66)

$$\|F(y_{i+1}) - F(y_i) - \frac{1}{2}(G(y_{i+1}) + G(y_i))(y_{i+1} - y_i)\| \leq K\|y_{i+1} - y_i\|^3,$$ \hspace{1cm} (67)

then all $x, y \in E$ with $\|x - y\| \leq \delta$ satisfy (56)–(58) with $M = 3K^4 \square$

Lemma 9 If the dandelion $\mathcal{D}(m, n, p, \lambda, x_k, f_k, g_k, h_k)$ and the functions $\{F_k\}$, $\{G_k\}$ and $\{H_k\}$ are compatible and the equations (20)–(23) hold for $k > k_0$ then there exists $M \in \mathbb{R}$ such that if $j, k > k_0$ and $u, w \in \mathbb{R}$ then

$$\|H_k(u, \lambda^k) - H_k(w, \lambda^k)\| \leq M|u - w|,$$ \hspace{1cm} (68)

$$\|G_k(u, \lambda^k) - G_k(w, \lambda^k) - \frac{1}{2}(H_k(u, \lambda^k) + H_k(w, \lambda^k))\Delta h\| \leq M|u - w|^2,$$ \hspace{1cm} (69)

$$\|F_k(u, \lambda^k) - F_k(w, \lambda^k) - \frac{1}{2}(G_k(u, \lambda^k) + G_k(w, \lambda^k))\Delta h\| \leq M|u - w|^3,$$ \hspace{1cm} (70)

$$\|H_k(w, \lambda^j) - H_k(u, \lambda^j)\| \leq M|\lambda^k - \lambda^j|,$$ \hspace{1cm} (71)

$$\|G_k(w, \lambda^j) - G_k(u, \lambda^k) - \frac{1}{2}(H_k(w, \lambda^j) + H_k(u, \lambda^j))\Delta h\| \leq M|\lambda^k - \lambda^j|^2,$$ \hspace{1cm} (72)

$$\|F_k(w, \lambda^j) - F_k(u, \lambda^k) - \frac{1}{2}(G_k(w, \lambda^j) + G_k(u, \lambda^k))\Delta h\| \leq M|\lambda^k - \lambda^j|^3,$$ \hspace{1cm} (73)

for $\Delta h = \overline{X}_k(u, \lambda^k) - \overline{X}_k(w, \lambda^k)$, $\Delta v = \overline{X}_k(w, \lambda^j) - \overline{X}_k(w, \lambda^k)$ and $\overline{X}$ as in (59) \hspace{1cm} \square
**Lemma 10** If $f_0, f_1, g_0, g_1, h_0$ and $h_1 \in \mathbb{R}$ are such that $g_0 < f_1 - f_0 < g_1$, $h_1 > 0$ and $h_2 > 0$ then there exists $\psi \in \text{Lip}^2(\mathbb{R})$ such that $\psi(0) = f_0$, $\psi(1) = f_1$, $\psi'(0) = g_0$, $\psi'(1) = g_1$, $\psi''(0) = h_0$, $\psi''(1) = h_1$ and $\psi''(w) > 0$ for all $w \in \mathbb{R}$ \hfill \Box

**Lemma 11** Given a function $\psi \in \text{Lip}^1(\mathbb{R}^2)$ such that $\psi(i, 0) = 0$ and $\nabla \psi(i, 0) = 0$ for $i \in \{0, 1\}$ there exists a function $\phi \in \text{Lip}^1(\mathbb{R}^2)$ such that

(a) $\phi(i, z) = 0$ and $\nabla \phi(i, z) = 0$ for $i \in \{0, 1\}$ and $z \in \mathbb{R}$.

(b) $\phi(w, 0) = \psi(w, 0)$ and $\nabla \phi(w, 0) = \nabla \psi(w, 0)$ for $w \in \mathbb{R}$ \hfill \Box

**Lemma 12** Consider $\lambda \in (0, 1)$, a function $F \in \text{Lip}^2(\mathbb{R}^2)$, functions $Y_1$ and $Y_2$ in $\text{Lip}^2([0, 1], \mathbb{R}^n)$ and $Y_3(z) = Y_2(z) - Y_1(z)$. If the vectors $Y'_1(0), Y'_2(0)$ and $Y'_3(0)$ are linearly independent and, for $i \in \{0, 1\}$,

$$F(i, 0) = 0, \quad \nabla F(i, 0) = 0 \quad \text{and} \quad \nabla^2 F(i, 0) = 0$$

(74)

then there exist $\delta > 0$, $G \in \text{Lip}^1(\mathbb{R}^2, \mathbb{R}^n)$ and $H \in \text{Lip}^0(\mathbb{R}^2, \mathbb{R}^n)$ such that if $z \in [0, \delta]$ and $i \in \{0, 1\}$ then

$$G(i, z) = 0, \quad G(w, z)'Y_3(z) = \frac{dF}{dw}(w, z), \quad G(w, 0)'V(w) = \frac{dF}{d\dot{z}}(w, 0),$$

(75)

$$H(i, z) = 0, \quad H(w, z)'Y_3(z) = \frac{dG}{dw}(w, z), \quad H(w, 0)V(w) = \frac{dG}{d\dot{z}}(w, 0)$$

(76)

(77)

for $V(w) = (1 - w)Y'_1(0) + wY'_2(0)$ and $U(w) = (1 - w)Y''_1(0) + wY''_2(0)$ \hfill \Box

**Proof of theorem 1.** Item 2 in the definition of flower in section 3 implies that there exists $\delta > 0$ such that if $\|x_k - x_j\| \leq \delta$ then $j \equiv k \mod p$. The periodicity of $\phi_k$, $\gamma_k$ and $\theta_k$ given by (16), item 2 in definition 1 and the bounds (16)–(19) imply that if $x$ and $z$ belong to the set $E = \{x_k, k \in \mathbb{N}\} \cup \{\chi_k, k \in \mathbb{N}\}$ and $\|x - z\| \leq \delta$ then there exists $k$ such that $m = 1$ and $y_1 = \chi_k$ satisfy the conditions (64)–(67) in lemma 8. Therefore, lemmas 6 and 8 imply that there exists a function $F$ as required by theorem 1 \hfill \Box

**Proof of theorem 2.** Consider $k_0$ and $\delta$ obtained from lemma 7 and

$$E = \bigcup_{k=k_0}^{\infty} \{\chi_k(w, \lambda^k), w \in [a_k, b_k]\}.$$

Lemma 7 and the three equations in (20) imply that the expressions

$$F(\chi_k(w, \lambda^k)) = F_k(w, \lambda^k), \quad G(\chi_k(w, \lambda^k)) = G_k(w, \lambda^k), \quad H(\chi_k(w, \lambda^k)) = H_k(w, \lambda^k)$$

define functions $F$, $G$ and $H$ with domain $E$, i.e., if $\chi_j(w_j, \lambda^j) = \chi_k(w_k, \lambda^k)$ then $F_j(w_j, \lambda^j) = F_k(w_k, \lambda^k)$, $G_j(w_j, \lambda^j) = G_k(w_k, \lambda^k)$, $H_j(w_j, \lambda^j) = H_k(w_k, \lambda^k)$. 


Lemmas 7 and 9 show that the functions $F$, $G$ and $H$ above satisfy the hypothesis of lemma 8 and theorem 2 follows from lemma 6.

**Proof of theorem 3.** For $\rho > 0$, consider the compact convex set

$$\mathcal{R}_\rho = \{ x \in \mathbb{R}^n \mid \sup_{y \in \mathbb{R}} \| x - y \| \leq \rho \}.$$  

Since $\mathcal{L}_k \cap \mathcal{L}_r \cap \mathcal{R} = \emptyset$ for $r + 1 < k < r + p - 1$ there exists $\epsilon > 0$ such that $\mathcal{L}_k \cap \mathcal{L}_r \cap \mathcal{R}_{2k} = \emptyset$ for the same $k$ and $r$. For each $0 \leq r < p$ there exist $a_r < 0$ and $b_r > 1$ such that

$$\mathcal{L}_r \cap \mathcal{R}_{2k} = \{ X_r(0) + w S_r(0), \ w \in [a_r, b_r] \}.$$

Extending the definition of $a_k$ and $b_k$ by periodicity, $a_k = a_{k \text{ mod } p}$ and $b_k = b_{k \text{ mod } p}$, we obtain intervals $[a_k, b_k]$ compatible with $\mathcal{D}$. Combining lemmas 1, 3 and 10 and theorem 2 we obtain $k_1$ and $\mathcal{F} \in \text{Lip}^2(\mathbb{R}^n)$ such that if $k > k_1$ then $\mathcal{F}(x_k) = f_k$, $\nabla \mathcal{F}(x_k) = g_k$, $\nabla^2 \mathcal{F}(x_k) = h_k$ and $s_k \nabla^2 \mathcal{F}(x_k + w S_k) > 0$ for $w \in [a_k, b_k]$.

The points $X_k(0)$ belong to the interior of the compact convex set $\mathcal{R}_{\epsilon}$ and there exist $c_k < 0$ and $d_k > 1$ such that $c_{k+p} = c_k, d_{k+p} = d_k$ and

$$\mathcal{L}_k \cap \mathcal{R}_{\epsilon} = \{ X_k(0) + w S_k(0), \ w \in [c_k, d_k] \}$$

and vectors $U_k, V_k$ such that $U_{k+p} = U_k, V_{k+p} = V_k, U_k S_k(0) < 0, V_k S_k(0) > 0$ and $U_k(x - X_k(0) - c_k S_k(0)) \leq 0$ and $V_k(x - X_k(0) - d_k S_k(0)) \leq 0$ for $x \in \mathcal{R}_{\epsilon}$.

Let $L$ be a Lipschitz constant for the second derivatives of $\mathcal{F}$, $\mu > 0$ such that

$$\mu \sqrt{\| S_k(0) \|} < -U_k S_k(0) \quad \text{and} \quad \mu \sqrt{\| S_k(0) \|} < V_k S_k(0) \quad \text{for} \quad k \in \mathbb{N}$$

and $\tau : \mathbb{R} \to \mathbb{R}$ be the function $\tau(w) = \max(w, 0)^3$. The function

$$F(x) = \mathcal{F}(x) + \frac{L}{\mu^3} \sum_{s=0}^p \left( \tau(U_k(x - X_k(0) - c_k S_k(0))) + \tau(V_k(x - X_k(0) - d_k S_k(0))) \right)$$

and $k_0 > k_1$ such that $\mu \sqrt{\| S_k \|} < -3U_k S_k$ and $\mu \sqrt{\| S_k \|} < 3V_k S_k$.

$$U_k(x_k + b_k S_k - X_k(0) - c_k S_k(0)) > 0 \quad \text{and} \quad V_k(x_k + b_k S_k - X_k(0) - d_k S_k(0)) > 0$$

for $k > k_0$ are as required by theorem 3. In fact, $F$ coincides with $\mathcal{F}$ in $\mathcal{R}_{\epsilon}$ and if $k > k_0$ and $w > b_k$ then

$$s_k \frac{\partial^2 F}{\partial w^2}(x_k + w S_k) s_k \geq s_k \frac{\partial^2 \mathcal{F}}{\partial w^2}(x_k + b_k S_k) s_k - L(w - b_k) \| s_k \| +$$

$$\frac{6L}{\mu^3} V_k^3 (V_k^2 s_k^2) V_k^2(x_k - b_k S_k - X_k(0) - d_k S_k(0)) + (w - b_k) \frac{6L}{\mu^3} V_k^3 s_k^3 \geq L(w - b_k) \| s_k \| \quad \square$$

**Proof of lemma 1.** Let $F$ be the function obtained by applying theorem 1 to $\mathcal{D}$. Lemma 12 applied to the functions $\hat{F} = F_k - F, Y_1 = X_k$ and $Y_2 = X_{k+1}$ yield functions $\hat{G}_k$ and $\hat{H}_k$ such that $\hat{G}_k(x) = \nabla F(x) + \hat{G}_k(x)$ and $\hat{H}_k(x) = \nabla^2 F(x) + \hat{H}_k(x)$ are as claimed in lemma 1. \hfill \square
Proof of lemma 2. Equations (28)–(31) show that
\[ X_k(z) = \sum_{j=1}^{n} z^j (\tau_\gamma)_j e_j \quad \text{and} \quad g_k = \sum_{i=1}^{n} \lambda^{k(d_i-d_j)} (\tau_i)_i e_i \tag{78} \]
where \( e_i \in \mathbb{R}^n \) is the vector with \( e_{ii} = 1 \) and \( e_{ij} = 0 \) for \( i \neq j \). Item 3b in definition 6 implies that if \( T = \{(i, j) \mid d_i + d_j \leq d_n \} \) then
\[ h_k = \sum_{(i, j)\in T} \lambda^{k(d_i-d_j)} (\bar{\tau}_{ij}) e_i e_j. \tag{79} \]
Since \( Q^p = I \) and \( \bar{x}_{k+p} = \bar{x}_k \), the \( X_k \) in (28) satisfy item (a) in the definition of dandelion. The relations \( \bar{T}_{k+p} = \bar{T}_k, \bar{\tau}_{k+p} = \bar{\tau}_k \) and \( \bar{\tau}_{k+p} = \bar{\tau}_k \) imply that \( f_k, g_k \) and \( h_k \) accumulate at the limits \( \bar{\phi}_k, \bar{\gamma}_k \) and \( \bar{\theta}_k \) in the last column of (28)–(31) and these limits satisfy (16). We now verify (17)–(19). Equation (79) yields (17). The first equation in (78) leads to
\[ \Delta_k = x_k - \chi_k = \sum_{j\in J} \lambda^{kd_j} (\tau_\gamma)_j, \tag{80} \]
where \( J = \{ j \mid d_j > 0 \} \). The second equation in (78) and (79) imply that
\[ \| g_k - \theta_k \Delta_k \| \leq n\lambda^k \sum_{u\in U} |(\bar{\tau}_u)_u - \sum_{j\in J} (\bar{\tau}_{ij})_i | + O(\lambda^2) \tag{81} \]
for \( U = \{ u \mid d_u = d_n - 1 \} \) and equations (81) and (26) show that the \( g_k \) satisfy (18). Equations (32), (33) and (80) lead to
\[ g^\omega_k \Delta_k = \lambda^{kd_n} \sum_{j\in G} (\tau_\gamma)_j e_j \quad \text{and} \quad \Delta_k^T \theta^\omega_k \Delta_k = \lambda^{kd_n} \sum_{(i, j) \in T} (\bar{\tau}_{ij}) e_i e_j, \]
for \( L \) and \( S \) in item 4b of definition 6, and (19) follows from (27). Finally, the linear independence requirements in items (b) and (c) of definition 3 follow from item 3 in definition 6.

Proof of lemma 3. Direct computation using (28)–(31) and (35) show that the function \( F_k \) satisfies (24).

Proof of lemma 4. To prove this lemma, plug (28)–(31) into the expressions in the hypothesis of lemma 4 and compare the results with (36)–(37).

Proof of lemma 5. The matrix \( Z \) in (46) can be written as
\[ Z = \lambda^3 D(\lambda^{-1}) Q = \begin{pmatrix} Z_1 & 0 & 0 & 0 \\ 0 & Z_2 & 0 & 0 \\ 0 & 0 & Z_3 & 0 \\ 0 & 0 & 0 & Z_4 \end{pmatrix}, \tag{82} \]
where the \( Z_i \)'s are the following \( 1 \times 1 \) and \( 2 \times 2 \) blocks:
\[ Z_1 = \lambda^3 \begin{pmatrix} \cos \frac{\pi}{3} & -\sin \frac{\pi}{3} \\ \sin \frac{\pi}{3} & \cos \frac{\pi}{3} \end{pmatrix}, \quad Z_2 = -\lambda^3, \quad Z_3 = \lambda^2, \quad Z_4 = -\lambda^2, \quad Z_5 = 1. \tag{83} \]
If we identify the blocks $Z_i$ with the complex numbers
\[ z_1 = \lambda^3 e^{i\theta}, \quad z_2 = -\lambda^3, \quad z_3 = \lambda^2, \quad z_4 = -\lambda^2, \quad z_5 = 1. \] (84)
then (46) can be interpreted as the equations
\[ z_k^6 \hat{g}_{k+6} = \rho \left( z_k \hat{g}_{j,k+1} - \hat{g}_{j,k} \right), \] (85)
in the complex variables $\hat{g}_{j,k}$ given by
\[
\begin{align*}
\hat{g}_{1,k} &= (\bar{g}_k)_1 + i(\bar{g}_k)_2, \\
\hat{g}_{2,k} &= (\bar{g}_k)_3, \\
\hat{g}_{3,k} &= (\bar{g}_k)_4, \\
\hat{g}_{4,k} &= (\bar{g}_k)_5, \\
\hat{g}_{5,k} &= (\bar{g}_k)_6.
\end{align*}
\] (86)
In the case $\rho_{k+1} = \rho$ and $\bar{g}_{r+1} = \bar{g}_r$, equation (85) is equivalent to
\[ A(1/\rho_0, 1/\rho_2, \ldots, 1/\rho_{10}, -z_j)\hat{g}_j = 0, \] (88)
where $\hat{g}_j = (\hat{g}_{j,0}, \hat{g}_{j,1}, \hat{g}_{j,2}, \hat{g}_{j,3}, \ldots, \hat{g}_{j,7}, \hat{g}_{j,10})^T$ and
\[
A(y, z) = \begin{pmatrix}
1 & z & 0 & 0 & 0 & 0 & y_6 z^6 & 0 & 0 & 0 \\
0 & 1 & z & 0 & 0 & 0 & y_1 z^6 & 0 & 0 & 0 \\
0 & 0 & 1 & z & 0 & 0 & 0 & y_2 z^6 & 0 & 0 \\
0 & 0 & 0 & 1 & z & 0 & 0 & 0 & y_3 z^6 & 0 \\
0 & 0 & 0 & 0 & 1 & z & 0 & 0 & 0 & y_4 z^6 \\
0 & y_6 z^6 & 0 & 0 & 0 & 1 & z & 0 & 0 & 0 \\
0 & y_1 z^6 & 0 & 0 & 0 & 1 & z & 0 & 0 & 0 \\
0 & 0 & y_8 z^6 & 0 & 0 & 0 & 1 & z & 0 & 0 \\
0 & 0 & 0 & y_9 z^6 & 0 & 0 & 0 & 1 & z & 0 \\
0 & 0 & 0 & 0 & y_{10} z^6 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}. \] (89)
Equation (88) suggests that we take $\rho_0, \rho_1, \ldots, \rho_{10}$ such that the corresponding matrices $A$ in (89) are singular. Given such $\rho$'s we can find vectors $\hat{g}_j \neq 0$ that satisfy (88) and use them to define the vectors $\bar{g}$ taking real and imaginary parts of (86)–(87). This approach leads to the polynomial equations
\[ \det(A(y, -z_j)) = 0 \] (90)
on the $y_r = 1/\rho_r$. To obtain accurate approximations for appropriated $\rho_r$ we take
\[ y_6 = -1.9948, \quad y_7 = -0.3737, \quad y_8 = -1.2355, \quad y_9 = 0.9857, \quad y_{10} = 0.11717 \]
and apply Newton's method to find the remaining $y_r$. Expression (90) correspond to a system of six real equations and Newton's method starting with
\[ y_0 = 2.04831, \quad y_1 = 3.33798, \quad y_2 = -1.15867, \]
\[ y_3 = 0.300795, \quad y_4 = -0.634211, \quad y_5 = -2.44761, \]
converges quickly to a solution of this system. This approach led to rational approximations $\bar{y}$'s for the $y_r$'s such that $\| \det(A(\bar{y}, -z_j)) \| < 10^{-1000}$ for $j = 1, 2, 3, 4$ and 5 and such that the Jacobian of the system (90) is well conditioned at $\bar{y}$. A
standard argument using Kantorovich’s theorem proves the existence of an exact \( y \) in a neighborhood of radius \( 10^{-500} \) of our rational approximation. Using the approximation \( \tilde{y} \) we dropped the last row in each of the matrices \( A(y, z_j) \) and computed highly accurate approximations for the five complex vectors \( \tilde{y} \) in (88) normalized by the condition \( (\tilde{y}^\dagger)J_1 = 1 \). Using (86)–(87) we obtained accurate approximations to the vectors \( \tilde{y} \) required by lemma 5. Using these approximations we verified that the vectors in (47) are indeed linearly independent and solved the systems (48) for the vectors \( \bar{x}_k \). Finally, we computed approximations for the \( \bar{x}_k \) in (49)–(50) and verified that the corresponding lines \( \Sigma_k \) and \( \Sigma_k \) in the hypothesis of theorem 3 are at least \( 10^{-3} \) apart. Our computations indicate that the exact lines \( \Sigma_k \) and \( \Sigma_k \) do not cross. We did a rigorous sensitivity analysis of the computations above and it indicated that 500 digits is precision enough to guarantee that our conclusions apply to the exact \( y \)'s, \( \bar{y} \)'s, \( \bar{z} \)'s and \( \bar{x} \)'s.

\[ \square \]

**Proof of lemma 6.** Lemma 6 follows from the version of Whitney’s theorem in page 6 of [7] applied to \( \omega(i) = t \) and a properly scaled version of the polynomials

\[ P_i(y) = F(x) + G(x)\phi(y-x) - H(x)(y-x) \]

**Proof of lemma 7.** By compatibility of \( \{a_k, b_k\} \) and \( \mathcal{P} \) and periodicity \( X_k = X_0 \), there exists \( \varepsilon > 0 \) for which the segments \( \Sigma_k = \{X_k(w, 0), w \in [a_k, b_k]\} \) are such that dist\((\Sigma_k, \Sigma_k')\) \( > \varepsilon \) if \( r+1 < k < r + p - 1 \). If \( \delta > 0 \) is small enough then the surfaces

\[ \mathcal{P}_k = \{X_k(w, z), w \in [a_k, b_k], z \in [0, \delta]\} \]

are such that

\[ \text{dist}(\mathcal{P}_k, \mathcal{P}_k') > \delta \quad \text{for} \quad r+1 < k < r + p - 1. \]

The definition of dandelion and the Lipschitz continuity of the derivatives of \( X_k \) imply that for \( \delta > 0 \) small if \( v, z \in [0, \delta], k \in \mathbb{N} \) and \( a, b, c \in \mathbb{R} \) then

\[ \|aS_k(z) + bX_k(z) + cX_{k+1}(z)\| \geq \delta(\|aS_k(z)\| + \|bX_k(z)\| + \|cX_{k+1}(z)\|), \]

\[ \|aS_k(v) + bX_k(z) + cS_{k+1}(z)\| \geq \delta(\|aS_k(v)\| + \|bX_k(z)\| + \|cS_{k+1}(z)\|). \]

We also have that

\[ X_k(v) = X_k(z) + (v - z)X_k'(z) + A_k(v, z) \]

where \( A_k(v, z) \) satisfies the inequality \( \|A_k(v, z)\| < \delta \|v - z\| \max_{z \in [0, 1]} \|X''(z)\| \) when \( v, z \in [0, \delta] \). The linear independence requirements in the definition of dandelion imply that \( X_k(0) \neq 0 \) and (94) implies that if \( \delta > 0 \) is small and \( v, z \in [0, \delta] \) then

\[ 2\|v-z\|X_k'(z) \geq \|X_k(v) - X_k(z)\|. \]

We now show that \( \delta > 0 \) for which the expressions (91)–(95) above are valid fulfills the requirements in lemma 7. Given \( y_j \) and \( y_k \) as in the hypothesis of lemma 7 there exists \( i \in [k, k-p] \) such that \( i \equiv j \mod p \). Since \( \|y_j - y_k\| \leq \delta \) and \( \bar{X}_i = \bar{X}_j \) the surfaces \( \mathcal{P}_k \) and \( \mathcal{P}_k \) are at most \( \delta \) apart and (91) leaves only the three possibilities: (a) \( i = k \), (b) \( i = k+1 \) or (c) \( i = k + p - 1 \). This corresponds to (60) and to complete this proof we now verify the bounds (61)–(63). In case (a) \( \bar{X}_j = \bar{X}_k \) and \( \bar{S}_j = \bar{S}_k \) and the definition of \( y_j \) and \( y_k \) in the hypothesis and (94) lead to

\[ y_j - y_k = (w_j - w_k)S_k(z_k) + (1 - w_j)(X_k(z_j) - X_k(z_k)) + w_j(X_{k+1}(z_j) - X_{k+1}(z_k)) \]
\[ = (w_j - w_k)S_k(z_k) + (1 - w_j)(z_j - z_k)X'_j(z_k) + w_j(z_j - z_k)X'_{k+1}(z_k) + (1 - w_j)A_k(z_j, z_k) + w_jA_{k+1}(z_j, z_k). \]

The bounds (92) and (95) and the abbreviation \((w_j - w_k)S_k(z_k) = \Gamma\) yield
\[
\|y_j - y_k\| \geq 3\delta \left( \|\Gamma\| + \|(1 - w_j)(z_j - z_k)X'_j(z_k)\| + \|w_j(z_j - z_k)X'_{k+1}(z_k)\| \right)
- \|(1 - w_j)A_k(z_j, z_k)\| - \|w_jA_{k+1}(z_j, z_k)\|.
\]
and (61) follows from the equations \(\Gamma = X_j(w_j, z_k) - y_k\) and
\[
y_j - X_j(w_j, z_k) = (1 - w_j)(X_j(z_j) - X_k(z_k)) + w_j(X_{k+1}(z_j) - X_{k+1}(z_k)).
\]
In case (b) \(X_j = X_{k+1}\) and \(S_j = S_{k+1}\) and (94) lead to
\[
y_j - y_k = w_jS_{k+1}(z_j) + (z_j - z_k)X'_j(z_k) + (1 - w_k)S_k(z_k) + \Delta_j(z_k, z_j)
\]
and using (93) and (95) we deduce that
\[
\|y_j - y_k\| \geq 3\delta \left( \|w_jS_j(z_j)\| + \|(z_j - z_k)X'_j(z_k)\| + \|(1 - w_k)S_k(z_k)\| \right)
- \|\Delta_j(z_k, z_j)\| \geq 3\delta \left( \|w_jS_j(z_j)\| + \|(z_j - z_k)X'_j(z_k)\| + \|(1 - w_k)S_k(z_k)\| \right)
\]
and (62) follows at once. Finally, case (c) is analogous to case (b).

**Proof of lemma 8.** Given \(x, z \in E\), let \(y_1, \ldots, y_m\) be as in the hypothesis of lemma 8 and define \(y_0 = x, y_{m+1} = z\) and \(f_j = F(y_j), g_j = G(y_j)\) and \(h_j = H(y_j)\). The bounds (65) and (64) yield
\[
\|h_j - h_0\| \leq \sum_{i=1}^{j} \|h_i - h_{i-1}\| \leq \sum_{i=1}^{j} K\|y_i - y_{i-1}\| \leq K^2\|x - z\| \quad (96)
\]
and (56) is obtained by taking \(j = m + 1\) in (96). The identity
\[
g_j - g_0 - h_0(y_j - y_0) = \sum_{i=1}^{j} g_i - g_{i-1} - \frac{1}{2}(h_i + h_{i-1})(y_i - y_{i-1}) + \frac{1}{2} \sum_{i=1}^{j} (h_i + h_{i-1} - 2h_0)(y_i - y_{i-1})
\]
and the bounds (64), (66) and (96) and imply that
\[
\|g_j - g_0 - h_0(y_j - y_0)\| \leq 2K^3\|x - z\|^2. \quad (97)
\]
Taking \(j = m + 1\) we deduce that \(x\) and \(z\) satisfy (57). Finally, notice that
\[
f_j - f_0 - g_0(y_j - y_0) - \frac{1}{2}(y_j - y_0)'h_0(y_j - y_0) = \sum_{i=1}^{j} \left( f_i - f_{i-1} - \frac{1}{2}(g_i + g_{i-1})'(y_i - y_{i-1}) \right) +
\]
\[
\frac{1}{2} \sum_{i=1}^{j} (g_{i-1} - g_0 - h_0(y_{i-1} - y_0)) f(y_i - y_{i-1}) + \\
\frac{1}{2} \sum_{i=1}^{j} (g_i - g_0 - h_0(y_i - y_0)) f(y_i - y_{i-1}) - \frac{1}{2} (y_j - y_0) f h_0(y_j - y_0) + \frac{1}{2} \Delta \quad (98)
\]

for \( \Delta = \sum_{i=1}^{j} (y_i + y_{i-1} - 2y_0) f h_0(y_i - y_{i-1}) \). The last terms in (98) cancel because

\[
\Delta = \sum_{i=1}^{j} ((y_i - y_0) + (y_{i-1} - y_0)) f h_0((y_i - y_0) - (y_{i-1} - y_0)) = \\
\sum_{i=1}^{j} (y_i - y_0) f h_0(y_i - y_0) - \sum_{i=1}^{j} (y_{i-1} - y_0) f h_0(y_{i-1} - y_0) = (y_j - y_0) f h_0(y_j - y_0).
\]

Thus, if we take \( j = m + 1 \) then (64), (97) and (98) yield (58) for \( M = 3K^4 \)

Proof of lemma 9. The bounds (68) and (71) follow from the Lipschitz continuity of \( H_k \). Equation (69) can be derived from the first equation in (23),

\[
\Delta h = \mathcal{X}_{k_0}(u, \lambda^k) - \mathcal{X}_{k}(w, \lambda^k) = (u - w)s_k, \quad (99)
\]

and the Lipschitz continuity of the first derivatives of \( G_k \). Equation (70) is a consequence of the first equation in (21) and (23), (99) and the Lipschitz continuity of the second derivatives of \( F_k \). The bounds (72) and (73) are clearly satisfied if \( j = k \) and from now on we assume that \( j \neq k \). In this case

\[
\lambda^j + \lambda^k < \frac{1 + \lambda}{1 - \lambda} |\lambda^j - \lambda^k| \quad (100)
\]

and \( X_k(\lambda^j) = X_k(\lambda^k) + X_k'(0)(\lambda^j - \lambda^k) + \frac{1}{2} X_k''(0)(\lambda^{2j} - \lambda^{2k}) + \mu_{j,k} \), where \( \mu_{j,k} = \int_{\lambda^k}^{\lambda^j} f^2(X''(\xi) - X''(0))d\xi d\xi \) satisfies

\[
|\mu_{j,k}| \leq L|\lambda^k - \lambda^j||\lambda^k + \lambda^j|^2 \leq \frac{L(1 + \lambda)^2}{(1 - \lambda)^2} |\lambda^k - \lambda^j|^3,
\]

for a Lipschitz constant \( L \) for \( X_k'' \). The bound (100) and \( X_k \in \text{Lip}^2([0,1], \mathbb{R}^n) \) lead to

\[
\Delta v = (1 - w)(X_k(\lambda^j) - X_k(\lambda^k)) + w(X_{k+1}(\lambda^j) - X_{k+1}(\lambda^k)) = \\
(\lambda^j - \lambda^k)V_k(w) + \frac{1}{2}(\lambda^{2j} - \lambda^{2k})U_k(w) + O(|\lambda^j - \lambda^k|^3) \quad (101)
\]

for

\[
U_k = (1 - w)X'_k(0) + w X'_{k+1}(0) \quad \text{and} \quad V_k = (1 - w)X''_k(0) + w X''_{k+1}(0).
\]
The conditions $F_k \in \text{Lip}^2(\mathbb{R}^2)$ and $G_k \in \text{Lip}^1(\mathbb{R}^2, \mathbb{R}^n)$ imply that

$$G_k(w, \lambda^j) = G_k(w, 0) + \lambda^j \frac{\partial G_k}{\partial z}(w, 0) + O(\lambda^j),$$  \hfill (102)

$$G_k(w, \lambda^j) - G_k(w, \lambda^k) = \frac{\partial G_k}{\partial z}(w, 0)(\lambda^j - \lambda^k) + O(|\lambda^j - \lambda^k|^2),$$  \hfill (103)

$$F_k(w, \lambda^j) - F_k(w, \lambda^k) = \frac{\partial F_k}{\partial z}(w, 0)(\lambda^j - \lambda^k) + \frac{1}{2} \frac{\partial^2 F_k}{\partial z^2}(w, 0)(\lambda^j - \lambda^k)^2$$

$$+ O(|\lambda^j - \lambda^k|^3).$$  \hfill (104)

The bound (72) follows from the second equation in (23), (101) and (103). Finally, the bound (73) can be deduced from the second equation in (21), equations (22), (100), (101), equation (102) with $l = j$ and $l = k$ and equation (104) \hfill □

**Proof of lemma 10.** Given $\varepsilon > 0$, consider the function

$$F_\varepsilon(w) = f_0 + (f_1 - f_0)w + C_\varepsilon(w),$$

where $C_\varepsilon(w)$ is the piecewise cubic given by

\[
\begin{align*}
 w\overline{g}_0 + w^2\overline{h}_0 & \quad \text{if } w \leq 0, \\
 w\overline{g}_0 + w^2\overline{h}_0 + a_\varepsilon w^3 & \quad \text{if } w \in (0, \varepsilon), \\
 w\overline{g}_0 + w^2\overline{h}_0 + a_\varepsilon w^3 + b_\varepsilon(w - \varepsilon)^3 & \quad \text{if } w \in (\varepsilon, 2\varepsilon), \\
 f_\varepsilon + a_\varepsilon(2w - 1) + \varepsilon(2w - 1)^2 & \quad \text{if } w \in (2\varepsilon, 1 - 2\varepsilon), \\
 (w - 1)\overline{g}_1 + (w - 1)^2\overline{h}_1 + c_\varepsilon(w - 1)^3 & \quad \text{if } w \in (1 - 2\varepsilon, 1 - \varepsilon), \\
 (w - 1)\overline{g}_1 + (w - 1)^2\overline{h}_1 + d_\varepsilon(w - 1)^3 & \quad \text{if } w \in (1 - \varepsilon, 1), \\
 (w - 1)\overline{g}_1 + (w - 1)^2\overline{h}_1 & \quad \text{if } w > 1,
\end{align*}
\]

with $\overline{g}_i = g_i - f_1 + f_0$ and $\overline{h}_i = h_i/2$ for $i \in \{0, 1\}$. The function $F_\varepsilon$ belongs to $\text{Lip}^2(\mathbb{R})$ if and only if it has continuous second order derivatives at $w = 2\varepsilon$ and $w = 1 - 2\varepsilon$. This condition leads to a linear system of six equations on the six variables $a_\varepsilon, b_\varepsilon, c_\varepsilon, d_\varepsilon, f_\varepsilon$ and $g_\varepsilon$. Solving this system we obtain

$$a_\varepsilon = \frac{-\overline{g}_0 + O(\varepsilon)}{6\varepsilon^2(1 - 2\varepsilon)}, \quad d_\varepsilon = \frac{-\overline{g}_1 + O(\varepsilon)}{6\varepsilon^2(1 - 2\varepsilon)}. \hfill (105)$$

The second derivative of $F_\varepsilon$ is a piecewise linear function with values

$$h_0, \quad h_0 + 6\varepsilon a_\varepsilon, \quad \varepsilon, \quad \varepsilon, \quad h_1 - 6\varepsilon d_\varepsilon, \quad h_1 \hfill (106)$$

at the nodes $w = 0, w = \varepsilon, w = 2 \varepsilon, w = 1 - 2\varepsilon, w = 1 - \varepsilon$ and $w = 1$. The hypothesis implies that $\overline{g}_0 < 0$ and $\overline{g}_1 > 0$ and (105) shows that $a_\varepsilon > 0$ and $d_\varepsilon < 0$ if $\varepsilon > 0$ is small. Therefore, (106) implies that $F_\varepsilon''(w) > 0$ for all $w$ if $\varepsilon$ is small enough \hfill □

**Proof of lemma 11.** Applying Whitney’s theorem to the set

$$E = \{(0, y) \mid |y| \leq 3\} \cup \{(1, y) \mid |y| \leq 3\} \cup \{(x, 0) \mid |x| \leq 3\} \subset \mathbb{R}^2$$

and the functions $F : E \to \mathbb{R}$ and $G : E \to \mathbb{R}^2$ given by

$$F(w, 0) = \psi(w, 0), \quad F(i, z) = 0, \quad G(w, 0) = \nabla \psi(w, 0), \quad G(i, z) = 0,$$
for \( i = 0, 1 \) we obtain a function \( \bar{\phi} \in \text{Lip}^1(\mathbb{R}^2) \) such that
\[
\bar{\phi}(w, 0) = \psi(w, 0), \quad \nabla \bar{\phi}(w, 0) = \nabla \psi(w, 0), \quad \bar{\phi}(i, z) = 0, \quad \nabla \bar{\phi}(i, z) = 0
\]
for \( i = 0, 1 \) and \( |w|, |z| \leq 3 \). Let \( \tau : \mathbb{R} \to \mathbb{R} \) be a \( C^\infty \) function such that \( \tau(x) = 1 \) for \( |x| < 2 \) and \( \tau(x) = 0 \) for \( |x| > 3 \). The function
\[
\phi(w, z) = \tau(z) (\tau(w) \bar{\phi}(w, z) + (1 - \tau(w)) \psi(w, z))
\]
satisfies items (a) and (b) in lemma 11.

**Proof of lemma 12.** Let \( Z_1, Z_2 \) and \( Z_3 \) be the functions defined by
\[
Z_1(z) = Y_1(z), \quad Z_2(z) = Y_2(z), \quad Z_3(z) = Y_3(z).
\]
The vectors \( Z_1(0), Z_2(0) \) and \( Z_3(0) \) are linearly independent. Therefore, there exist \( W_1, W_2, W_3 \in \mathbb{R}^n \) such that
\[
W_i^T Z_j(0) = 1 \quad \text{and} \quad W_i^T Z_j(0) = 0 \quad \text{if} \quad i \neq j. \tag{107}
\]
The implicit function theorem guarantees the existence of \( \delta > 0 \) and functions \( a_i \in \text{Lip}^1([0, \delta]) \), defined for \( 1 \leq i, j \leq 3 \), such that
\[
a_i(0) = 1 \quad \text{and} \quad a_{ij}(0) = 0 \quad \text{if} \quad i \neq j \tag{108}
\]
and the vectors
\[
A_i(z) = a_{i1}(z)W_1 + a_{i2}(z)W_2 + a_{i3}(z)W_3 \tag{109}
\]
are such that
\[
A_i(z)'Z_j(z) = 1 \quad \text{and} \quad A_i(z)'Z_j(z) = 0 \quad \text{if} \quad i \neq j \tag{110}
\]
for \( z \in [0, \delta] \). Lemma 11 applied to \( \psi = \frac{\partial \tau}{\partial z} \) yields \( \phi \in \text{Lip}^1(\mathbb{R}^2) \) such that
\[
\phi(i, z) = 0, \quad \nabla \phi(i, z) = 0, \quad \phi(w, 0) = \frac{\partial F}{\partial z}(w, 0), \quad \frac{\partial \phi}{\partial z}(w, 0) = \frac{\partial^2 F}{\partial z^2}(w, 0) \tag{111}
\]
for \( w \in \mathbb{R} \) and \( i \in \{0, 1\} \). We now show that any \( G \in \text{Lip}^1(\mathbb{R}^2, \mathbb{R}^n) \) such that
\[
G(w, z) = \frac{\partial F}{\partial w}(w, z)A_3(z) + \phi(w, z)(A_1(z) + A_2(z)) \tag{112}
\]
for \( z \in [0, \delta] \) satisfies (75) and (76). Equations (74) and (111) imply that \( G(i, z) = 0 \) and second and third equations in (75) follows from (108)–(112) and the definition of \( S(z) \) and \( V(w) \). To verify (76), notice that (111) and (112) lead to
\[
\frac{\partial G}{\partial z}(w, 0) = \frac{\partial^2 F}{\partial w \partial z}(w, 0)W_3 + \frac{\partial F}{\partial w}(w, 0)A_3(0) + \frac{\partial^2 F}{\partial z^2}(w, 0)(W_1 + W_2) + \frac{\partial F}{\partial z}(w, 0)(A_1(0) + A_2(0)) \tag{113}
\]
Equation (110) yields \( A_i(0)'Z_j(0) = 0 \) and (108)–(110) imply that
\[
da_{ij}(0) = -W_i^T Z_j(0). \tag{114}
\]
If $j = 3$ then $Y_1'(0) = Y_2'(0) = Y_3'(0) = Z_2(0) - Z_1(0)$ and (114) leads to

$$a_{13}'(0) = 1, \quad a_{23}'(0) = -1, \quad a_{33}'(0) = 0. \quad (115)$$

Reminding that $Y_i''(0) = Y_j''(0)$ and using (110) and (114)–(115) we obtain

$$A_1'(0) = -W_1'Y_j''(0)W_1 - W_1'Y_j''(0)W_2 + W_3, \quad (116)$$

$$A_2'(0) = -W_2'Y_j''(0)W_1 - W_2'Y_j''(0)W_2 - W_3, \quad (117)$$

$$A_3'(0) = -W_3'Y_j''(0)W_1 - W_3'Y_j''(0)W_2. \quad (118)$$

Equations (111) and (112) show that

$$G(w, 0) = \frac{\partial F}{\partial w}(w, 0)W_3 + \frac{\partial F}{\partial \zeta}(w, 0)(W_1 + W_2) \quad (119)$$

and (76) follows from (110), (112), (116)–(118) and the fact that

$$V(w) = (1 - w)Z_1(0) + wZ_2(0) \quad \text{and} \quad U(w) = (1 - w)Y_1''(0) + wY_2''(0). \quad (120)$$

To complete this proof we define $H$ as

$$H(w, z) = \frac{\partial^2 F}{\partial w^2}(w, z)A_3(z)A_3(z)^2 + \Delta \frac{\partial F}{\partial z}(w, 0) + b(w)A_1(z)A_1(z)^2 + c(w)A_2(z)A_2(z)^2 \quad (121)$$

for

$$\Delta (w, z) = \frac{\partial \phi}{\partial w}(w, z) \left( (A_1(z) + A_2(z))A_3(z)^2 + A_3(z)(A_1(z) + A_2(z))^2 \right), \quad (122)$$

$$b(w) = \frac{1}{1 - w + w^2} (Z_1(0) - wZ_2(0))^2 \frac{\partial G}{\partial z}(w, 0), \quad (123)$$

$$c(w) = \frac{1}{1 - w + w^2} ((w - 1)Z_1(0) + Z_2(0)) \frac{\partial G}{\partial z}(w, 0). \quad (124)$$

Equations (74) and (111) imply that $H(i, z) = 0$ for $i \in \{0, 1\}$ and (112) and (110) imply the second equation in (77). Using (111), (120) and (121)–(124) we get

$$H(w, 0)V(w) = (b(w + wc(w)))W_1 + (1 - w)b(w + c(w))W_2 + \frac{\partial \phi}{\partial w}(w, 0)W_3$$

$$= Z_1(0)\frac{\partial G}{\partial z}(w, 0)W_1 + Z_2(0)\frac{\partial G}{\partial z}(w, 0)W_2 + \frac{\partial^2 F}{\partial w^2}(w, 0)W_3. \quad (125)$$

Equations (107), (113) and (116)–(118) show that

$$Z_3(0)\frac{\partial G}{\partial z}(w, 0) = \frac{\partial^2 F}{\partial w^2}(w, 0).$$

Finally, equations (107) shows that the right hand side of (125) is the expansion of $\frac{\partial G}{\partial \zeta}(w, 0)$ on the basis $\{W_1, W_2, W_3\}$ of a tri-dimensional subspace that contains $\frac{\partial G}{\partial \zeta}(w, 0)$ and we have shown (77). \hfill \Box
References