On the divergence of line search methods

WALTER F. MASCARENHAS

Instituto de Matemática e Estatística, Universidade de São Paulo,
Rua do Matão 1010, Cidade Universitária, 05508-090 São Paulo, SP, Brazil

E-mail: walterfm@ime.usp.br

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 level sets of the objective functions are compact, the line searches are exact and the Wolfe conditions are satisfied. This explanation illustrates a new way to combine general mathematical concepts and symbolic computation to analyze the convergence of line search methods. It also illustrate the limitations of the asymptotic analysis of the iterates of nonlinear programming algorithms.

Mathematical subject classification: 20E28, 20G40, 20C20.

Key words: line search methods, convergence.

1 Introduction

Line search methods are fundamental algorithms in nonlinear programming. Their theory started with Cauchy [4] 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. Their convergence theory is well developed and is described at length in many good surveys, as [11], and even in text books, like [2] and [12].

Line search methods, as discussed in this work, are used to solve the unconstrained minimization problem for a smooth function f:

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

We see them as discrete dynamical systems of the form

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

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 kth step we choose a search direction d_k and analyze f along the line $\{x_k + w d_k, w \in \mathbb{R}\}$. We search for a step size $\alpha_k \in \mathbb{R}$ such that the sequence x_k satisfies constraints

$$C(f, x_k, \alpha_k, d_k) \geq 0 \tag{3}$$

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

For example, Newton's method for minimizing a function can be written in the framework (2) by taking

$$D(f, x_k, \alpha_k, d_k, e_k) = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

and $E = 0$. The BFGS method could be considered by taking $e_k \in \mathbb{R}^{n \times n}$ and setting

$$D(f, x_k, \alpha_k, d_k, e_k) = -e_k^{-1} g_k, \tag{4}$$
$$E(f, x_k, \alpha_k, d_k, e_k) = e_k + \frac{\alpha_k}{s_k^t g_k} g_k g_k^t - \frac{1}{s_k^t g_k}(g_{k+1} - g_k)(g_{k+1} - g_k)^t, \tag{5}$$

for $g_k = \nabla f(x_k)$, $g_{k+1} = \nabla f(x_k + \alpha_k d_k)$ and $s_k = \alpha_k d_k$. A typical example of constraints C on the stepsize α_k in (3) are the Wolfe conditions:

$$\sigma \alpha_k \nabla f(x_k)^t d_k + f(x_k) - f(x_k + \alpha_k d_k) \geq 0, \tag{6}$$
$$d_k^t (\nabla f(x_k + \alpha_k d_k) - \beta \nabla f(x_k)) \geq 0, \tag{7}$$

where $0 < \sigma < \beta < 1$. Usually, condition (6) enforces a sufficient decay in the value of f from step k to $k + 1$ and condition (7) leads to steps $s_k = x_{k+1} - x_k$ which are not too short.

A typical theorem about the convergence of line search methods look like this one adapted from page 212 in Nocedal and Wright's book [12]:

Theorem 1. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ has continuous second order derivatives,*

$$\Omega = \{x \in \mathbb{R} \ such \ that \ f(x) \leq f(x_0)\} \tag{8}$$

is bounded and there exist $C > 0$ such that

$$y^t \nabla^2 f(x) y \geq C \|y\|^2 \quad for \ all \ x, y \in \mathbb{R}. \tag{9}$$

If the $n \times n$ matrix e_0 is symmetric positive definite and the iterates x_k generated by the BFGS method, as described in (2)–(5), satisfy the Wolfe conditions (6)–(7) then the sequence $\{x_k\}$ converges to a minimizer of f (which is unique since f is strictly convex by (9)). □

Starting in the 1960's, similar theorems have been proved for several line search methods. Since that time people have tried to find more general results regarding the convergence of line search methods. Most researchers were happy with constraints like the Wolfe conditions and acknowledged their need, because it is easy to enforce them in practice and it is also easy to build examples in which results like theorem 1 are false if similar constraints are not imposed. However, convexity constraints like (9) were regarded as too strong and undesirable. Much effort was devoted to eliminating them but progress was slow and frustrating due to the nonlinear nature of expressions like (5). As a consequence, M. Powell, one of the leading researchers in this area, wrote in [16] that:

> 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...

A final answer regarding the need of the convexity hypothesis (9) in theorem 1 was first published in 2002 by Y. Dai [6] and it was somewhat surprising: if f is not convex then the iterates generated by the BFGS method may never approach any point z such that $\nabla f(z) = 0$.

We were unaware of Dai's work and in the year his result was published we found a similar answer regarding the convergence of the BFGS method [13].

Our approach, however, was quite different from his. Our work was based on the observation that equations (4)–(7) have symmetries which can be exploited to build a counterexample for theorem 1 without the hypothesis (9). After the publication of [13] we generalized the argument used in that paper and our purpose in this work is to present this generalization.

Our approach is motivated by the work started by S. Lie in the late 1800's, in which symmetries have lead to remarkable solutions of nonlinear differential equations [3]. We developed a technique to produce examples with search lines as in Figure 1.
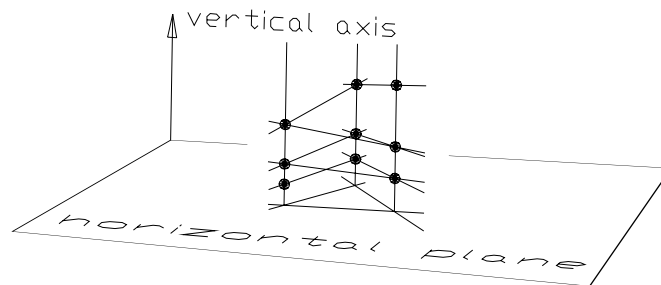

Figure 1 – An example in which the iterates (black dots) approach a cycle with period 3. The inclined lines represent the search lines and the horizontal lines are the limiting search directions. The vertical lines indicate that the projection of the iterates in the horizontal plane are the vertices of the limit cycle.

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 $\overline{s}_k$ corresponding to the objective function $\overline{F}(x) = F(\lambda^{-1}Q^t x)$ at the point $\overline{x}_k = \lambda Q x_k$ is $\overline{s}_k = \lambda Q s_k$. We argue that in relevant cases these symmetries lead to iterates as in Figure 1.

Actually, the possibility of cyclic behavior for line search methods was already mentioned by Curry in 1944 [5]. It was also discussed in [6, 9, 13, 15]. Here we go one step further and present a systematic way to build examples that display this behavior. The qualitative behavior of the iterates in our examples is captured by the concepts of *flower* and *dandelion* described in section 3. In intuitive terms the iterates can be seen as defining the petals of a flower and their accumulation

points lie in a cycle that define the flower's core. When the iterates approach the core along well defined directions we say that the flower is a dandelion.
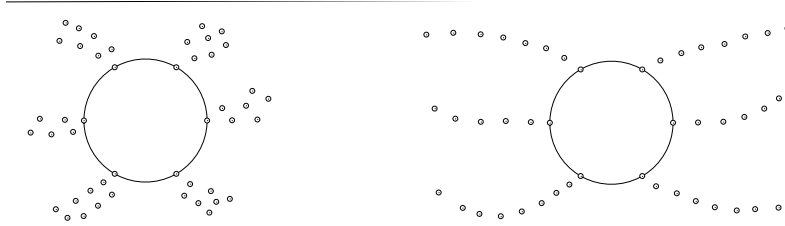

Figure 2 – A flower and a dandelion with six petals. The dots outside the circles represent the iterates. They accumulate at the points at the circles, which are not critical points of f.

In sections 2 and 7 we present concrete examples of dandelions, for Newton's method and the BFGS method, respectively. In these examples the line searches are exact, the first Wolfe condition is satisfied and the restrictions of the objective functions to the search lines are strictly convex, the level sets of f are compact, but yet the 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.

Besides symmetries, this work is based on a theorem proved by H. Whitney in 1934 [17]. 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.

In section 2 we illustrate how symmetries and Whitney's theorem can be combined 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 [3], "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 mth order Lipschitz continuous derivatives or to the more general class $C_{loc}^{m,\omega}(\mathbb{R}^n)$ of functions discussed by C. Fefferman in [7], 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 six more sections and an appendix. 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 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. In the appendix we prove our claims.

Finally, we would like to emphasize that it is important to look at the results in this work from a broad perspective. The examples presented here **should not** be taken as evidence against the use of Newton's method or the BFGS method. To the contrary, these methods perform quite well in practice. In real life numerical algorithms are implemented in floating point arithmetic and rounding errors would break our examples apart (and introduce other subtle problems). This work highlights the limitations of the asymptotic analysis of these algorithms. Our examples show that, even if taken to extremes, complex nonlinear calculations

may not be able to explain the practical behavior of nonlinear programming algorithms. The main advice one can extract from this work is the following rule of thumb:

> If you find that it is difficult to prove that a line search method converges under certain conditions, and other people have tried the same for a couple of decades and did not succeed, then consider the possibility that in theory the method may actually fail under these conditions, even if all your numerical experiments indicate that it always converge.

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 α: given $\alpha > 0$ we build an example in which all step-sizes α_k are equal to α. 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 $s_k^t \nabla^2 F(x_k + w s_k) s_k > 0$ 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 [1]:

$$x_k = Q^k D(\lambda^k)\overline{x}, \qquad\qquad f_k = \lambda^k \overline{f}, \qquad\qquad (10)$$

$$g_k = \lambda^k Q^k D(\lambda^{-k})\overline{g}, \qquad h_k = \lambda^k Q^k D(\lambda^{-k})\overline{h} D(\lambda^{-k}) Q^{-k}, \quad (11)$$

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

$$Q = \begin{pmatrix} Q_h & 0 \\ 0 & Q_v \end{pmatrix}, \quad D(\lambda) = \begin{pmatrix} I & 0 \\ 0 & \lambda I \end{pmatrix}, \quad \overline{h} = \begin{pmatrix} A & C \\ C^t & 0 \end{pmatrix}, \quad (12)$$

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

$$\lambda = \frac{1}{2}, \qquad \overline{x}^h = \overline{x}^v = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix},$$

$$Q_h = \frac{1}{2}\begin{pmatrix} \sqrt{3} & -1 & 0 \\ 1 & \sqrt{3} & 0 \\ 0 & 0 & -2 \end{pmatrix}, \qquad Q_v = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad (13)$$

are convenient: they are simple and after picking them we still have the freedom to choose $\overline{g}$, 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

$$\nabla^2 F(x_k)s_k + \alpha \nabla F(x_k) = 0 \qquad (14)$$

that defines Newton's method with step-size α. If we replace $\nabla F(x_k)$ and $\nabla^2 F(x_x)$ in (14) by g_k and h_k in (11) then λ, Q_h and Q_v cancel out and we obtain the equations

$$A\overline{s}^h + C\overline{s}^v + \alpha\overline{g}^h = 0, \qquad \text{and} \qquad C^t\overline{s}^h + \alpha\overline{g}^v = 0, \qquad (15)$$

---

[1]The matrices h_k are not positive definite and in practice one would take another search direction if, for example, this fact was detected during a Cholesky factorization of h_k. However, to keep the algebra as simples as possible, in this work we do not enforce the condition that the Hessians $\nabla^2 f(x_k)$ are s.p.d. In [14] we show that Newton's method may fail even $\nabla^2 f(x_k)$ is s.p.d. for all k and the Wolfe conditions are satisfied.

where

$$\overline{s}^h = (Q_h - I)\overline{x}^h \qquad \text{and} \qquad \overline{s}^v = (\lambda Q_v - I)\overline{x}^v. \qquad (16)$$

Notice that, due to the invariance of Newton's method with respect to orthogonal changes of variables and scaling, there is no "k" in (15)–(16). Equations (15) yields

$$\overline{g}^v = -C^t \overline{s}^h / \alpha \qquad (17)$$

Equations (10) and (11) show that $s_k^t g_k = 2^{-k} \overline{s}^t \overline{g}$, $s_k^t g_{k+1} = 2^{-(k+1)} \overline{s}^t D(2) Q \overline{g}$, and $f_{k+1} - f_k = -2^{-(k+1)} \overline{f}$. Therefore, if

$$\overline{s}^t \overline{g} < 0, \qquad \overline{s}^t D(2) Q \overline{g} = 0 \qquad \text{and} \qquad \overline{f} > 0 \qquad (18)$$

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 first Wolfe condition

$$f_{k+1} - f_k \le \sigma s_k^t g_k$$

holds for

$$0 < \sigma < \min\left\{1, \ -\overline{f}/(2\overline{s}^t \overline{g})\right\}.$$

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

$$\overline{g}^h = C \overline{x}^v \qquad \text{and} \qquad \overline{f} = (\overline{g}^v)^t \overline{x}^v \qquad (19)$$

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

$$2\overline{s}^t \overline{g} < -\overline{f} < 0, \qquad \overline{s}^t \overline{h} \overline{s} > 0 \qquad \text{and} \qquad \overline{s}^t Q D(2) \overline{h} D(2) Q^t \overline{s} > 0. \qquad (20)$$

To complete the specification of the terms in (10)–(11) we chose the 9 entries of C and the 6 independent entries of A in order to satisfy (15)–(20). These equations and inequalities are linear in A and C and the following matrices satisfy them:

$$A = (2 + \sqrt{3}) \begin{pmatrix} (2 + \sqrt{3})(17 - 22\alpha + 16\alpha^2) & 2\alpha & 2(\alpha - 1) \\ 2\alpha & (2 + \sqrt{3})(2\alpha - 1) & 0 \\ 2(\alpha - 1) & 0 & 0 \end{pmatrix},$$

$$C = (2 + \sqrt{3}) \begin{pmatrix} 3\alpha - 5\alpha^2 - 4 & 1 - \alpha + 3\alpha^2 & \alpha(1 + 5\alpha) \\ 0 & 2 - \sqrt{3} & 0 \\ 2 - \sqrt{3} & 0 & 0 \end{pmatrix}. \qquad (21)$$

Equations (10)–(13), (16)–(19) and (21) define iterates and function values, gradients and Hessians of the objective function at them. The next sections guarante 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 $D'(0)\overline{x}$ and $D'(0)\overline{s}$ nor the vectors $D(0)\overline{s}$ and $D(0)D(2)\overline{s}$ are aligned [2] and the lines $\mathfrak{L}_k = \{D(0)x_k + wD(0)s_k, w \in \mathbb{R}\}$ are such that $\mathfrak{L}_r \cap \mathfrak{L}_k = \emptyset$ for all $r + 1 < k < r + 5$ and theorem 3 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 $\{x_k + ws_k, w \in [0, 1]\}$: if the function $\psi : [0, 1] \to \mathbb{R}$ has Lipschitz continuous second order derivatives and

$$\psi(0) = \overline{f}, \qquad \psi(1) = \overline{f}/2, \qquad \psi'(0) = \overline{g}^t \overline{s}, \qquad \psi'(1) = 0,$$

$$\psi''(0) = \overline{s}^t \overline{h}\overline{s}, \qquad \text{and} \qquad \psi''(1) = \overline{s}^t Q D(2)\overline{h} D(2) Q^t \overline{s}/2$$

then F can be chosen so that $F(x_k + ws_k) = (1/2)^k \psi(w)$ for $w \in [0, 1]$ and k large. In fact, condition (20) and theorem 4 in section 6 show that F can be chosen so that $s_k^t F(x_k + ws_k)s_k > 0$ for $w \in \mathbb{R}$ and k large and the level sets

$$\Omega(f, z) = \left\{ x \in \mathbb{R}^n \text{ such that } f(x) \leq z \right\}$$

are bounded.

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* (see Figure 2). The limits of these subsequences $\{x_k\}$, $\{f_k\}$, $\{g_k\}$ and $\{h_k\}$ are the members of periodic sequences $\{\chi_k\}$, $\{\varphi_k\}$, $\{\gamma_k\}$ and $\{\theta_k\}$, so that $\lim_{q\to\infty} x_{pq+r} = \chi_r$ for all r and $\chi_{r+p} = \chi_r$, $\lim_{q\to\infty} f_{pq+r} = \varphi_r$ for all r and $\varphi_{r+p} = \varphi_r$, $\gamma_{r+p} = \gamma_r$ and $\theta_{r+p} = \theta_r$. In formal terms:

---

[2] $D'(\lambda)$ here is the derivative of $D(\lambda)$ with respect to λ.

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* $\{\chi_k\}$ *in* $\mathbb{R}^n$ *and a constant* $M > 0$ *such that*

    (a) $x_i = x_j \Leftrightarrow i = j.$

    (b) $\chi_j = \chi_k \Leftrightarrow j \equiv k \mod p.$

    (c) $\lambda^k \leq M\|x_k - \chi_k\| \leq M^2\lambda^k.$

3. *Sequences* $\{f_k\}, \{\varphi_k\} \subset \mathbb{R}, \{g_k\}, \{\gamma_k\} \subset \mathbb{R}^n$ *and* $\{h_k\}, \{\theta_k\} \subset \mathbb{H}^n$ *such that*

$$\varphi_{k+p} = \varphi_k, \qquad \gamma_{k+p} = \gamma_k, \qquad \theta_{k+p} = \theta_k, \tag{22}$$

$$\|h_k - \theta_k\| \leq M\lambda^k, \tag{23}$$

$$\|g_k - \gamma_k - \theta_k(x_k - \chi_k)\| \leq M\lambda^{2k}, \tag{24}$$

$$\|f_k - \varphi_k - \gamma_k^t(x_k - \chi_k) - \frac{1}{2}(x_k - \chi_k)^t h_k (x_k - \chi_k)\| \leq M\lambda^{3k} \tag{25}$$

   where $\mathbb{H}^n$ *is the set of* $n \times n$ *symmetric matrices.* □

Notice that the mod in item 2.b and equation (22) in the definition above imply that the sequences $\{\chi_k\}, \{\gamma_k\}$ and $\{\theta_k\}$ have period p and item 2.c implies that as $k \to \infty$ the sequence x_k accumulates at the limit cycle defined by the χ_k. The next definitions and theorems relates 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* $\mathrm{Lip}^m(U, V)$ *as the space of functions* $F : U \to V$ *with Lipschitz continuous mth derivatives. If* $V = \mathbb{R}$ *then we call this space simply by* $\mathrm{Lip}^m(U)$. □

Definition 3.   *We define* $\mathrm{LC}^2(\mathbb{R}^n)$ *as the set of functions* f *in* $\mathrm{Lip}^2(\mathbb{R}^n)$ *for which there exists constants* C_f *and* $R_f \in \mathbb{R}$, *which depend on* f, *such that if* $\|x\| \geq R_f$ *then* $\nabla^2 f(x)$ *is positive definite and* $\|\nabla^2 f(x)^{-1}\| \leq C_f$. □

The class $\mathrm{LC}^2(\mathbb{R}^n)$ is interesting because if $f \in \mathrm{LC}^2(\mathbb{R}^n)$ then there exists a constant D_f such that

$$f(x) \geq \frac{1}{2C_f}\|x\|^2 - D_f \qquad \text{and} \qquad \|\nabla f(x)\| \geq \frac{1}{C_f}\|x\| - D_f.$$

As a consequence, the level sets

$$\Omega(f, z) = \left\{ x \in \mathbb{R}^n \text{ such that } f(x) \leq z \right\}$$

are compact, as required by many theorems regarding the convergence of line search methods, and all points x with $\nabla f(x) = 0$ are such that $\|x\| \leq D_f C_f$. Thus, the elements of $\mathrm{LC}^2(\mathbb{R}^n)$ have a compact set of critical points and if an algorithm fails to find one of them then the algorithm is to be blamed, not the objective function. The next theorem shows that flowers can be interpolated by functions in $\mathrm{LC}^2(\mathbb{R}^n)$:

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

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

Theorem 3. *If the functions $\{F_k\}$, $\{G_k\}$ and $\{H_k\}$ and the intervals $\{[a_k, b_k]\}$ are compatible with the dandelion $\mathcal{D}(m, n, p, \lambda, x_k, f_k, g_k, h_k)$ then there exists $k_0 \in \mathbb{N}$ and $F \in \mathrm{LC}^2(\mathbb{R}^n)$ such that $F(x_k + w s_k) = F_k(w, \lambda^k)$, $\nabla F(x_k + w s_k) = G_k(w, \lambda^k)$ and $\nabla^2 F(x_k + w s_k) = H_k(w, \lambda^k)$ for $k > k_0$ and $w \in [a_k, b_k]$.* $\square$

This theorem will make sense after you read the following definitions:

Definition 4. *A flower $\mathcal{F}(n, p, \lambda, x_k, f_k, g_k, h_k)$ is a dandelion if there exist functions $X_k \in \mathrm{Lip}^2([0, 1], \mathbb{R}^n)$ 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 $X'_{k+1}(0)$ are linearly independent.*

(c) *The vectors $S_k(0)$, $S_{k+1}(0)$ and $X'_{k+1}(0)$ are linearly independent.* $\square$

Definition 5.   *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* $b_{k+p} = b_k \geq 1$ *for* $k \in \mathbb{N}$.

(b)   *The segments $\{X_k(0) + w S_k(0), w \in [a_k, b_k]\}$ and $\{X_r(0) + w S_r(0), w \in [a_r, b_r]\}$ are disjoint if $r + 1 < k < r + p - 1$.*   □

Definition 6.   *For all $k \in \mathbb{N}$, consider functions $F_k \in \mathrm{Lip}^2(\mathbb{R}^2)$, $G_k \in \mathrm{Lip}^1(\mathbb{R}^2, \mathbb{R}^n)$ and $H_k \in \mathrm{Lip}^0(\mathbb{R}^2, \mathbb{H}^n)$. 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*

$$F_k(i, z) = f_{k+i}, \qquad G_k(i, z) = g_{k+i}, \qquad H_k(i, z) = h_{k+i}, \qquad (26)$$

$$\frac{\partial F_k}{\partial w}(w, z) = G_k(w, z)^t s_k, \qquad \frac{\partial F_k}{\partial z}(w, 0) = G_k(w, 0)^t V_k(w), \qquad (27)$$

$$\frac{\partial^2 F_k}{\partial z^2}(w, 0) = \frac{\partial G_k}{\partial z}(w, 0)^t V_k(w) + G_k(w, 0)^t U_k(w), \qquad (28)$$

$$\frac{\partial G_k}{\partial w}(w, z) = H_k(w, z) s_k, \qquad \frac{\partial G_k}{\partial z}(w, 0) = H_k(w, 0) V_k(w) \qquad (29)$$

for $V_k(w) = (1 - w) X_k'(0) + w X_{k+1}'(0)$ and $U_k(w) = (1 - w) X_k''(0) + w X_{k+1}''(0)$.   □

If we do not care about the derivatives of F in directions normal to the segments $\{x_k + w s_k, w \in [0, 1]\}$ 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 $F_k \in \mathrm{Lip}^2(\mathbb{R}^2)$ for which $F_{k+p} = F_k$. If there exist $k_0 \in \mathbb{N}$ such that*

$$F_k(i, \lambda^k) = f_{k+i}, \qquad \frac{\partial F_k}{\partial w}(i, \lambda^k) = g_{k+i}^t s_k, \qquad \frac{\partial^2 F_k}{\partial w^2}(i, \lambda^k) = s_k^t h_{k+i} s_k \qquad (30)$$

for $i \in \{0, 1\}$ and $k > k_0$ then there exist functions $G_k \in \mathrm{Lip}^1(\mathbb{R}^2, \mathbb{R}^n)$ and $H_k \in \mathrm{Lip}^0(\mathbb{R}^2, \mathbb{H}^n)$ 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 $F_k(w, z) = z^{d_n} \psi_k(w)$.

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 7. *A seed* $S(n, p, d, Q, \overline{x}_k, \overline{f}_k, \overline{g}_k, \overline{h}_k)$ *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* $Q^p = I$ *and the diagonal matrix* $D(z)$ *with ith diagonal entry equal to* z^{d_i} *commutes with* Q *for* $z \in \mathbb{R}$.

3. *A sequence* $\{\overline{x}_k\} \subset \mathbb{R}^n$ *such that* $\overline{x}_{k+p} = \overline{x}_k$, *the points* $\chi_r = D(0)Q^r \overline{x}_r$ *are distinct for* $0 \leq r < p$ *and the vectors*

$$\overline{s}_k = D(\lambda)Q\overline{x}_{k+1} - \overline{x}_k \tag{31}$$

are such that, for $0 \leq r < p$, *the vectors* $D'(0)\overline{x}_r$ *and* $D'(0)\overline{s}_r$ *are not aligned and neither are the vectors* $D(0)\overline{s}_r$ *and* $D(0)\overline{s}_{r+1}$.

4. *Sequences* $\{\overline{f}_k\} \subset \mathbb{R}$, $\{\overline{g}_k\} \subset \mathbb{R}^n$ *and* $\{\overline{h}_k\} \subset \mathbb{H}^n$ *such that*

   (a) $\overline{f}_{k+p} = \overline{f}_k$, $\overline{g}_{k+p} = \overline{g}_k$ *and* $\overline{h}_{k+p} = \overline{h}_k$ *for all* k.

   (b) $(\overline{h}_k)_{ij} = 0$ *if* $d_i + d_j > d_n$.

   (c) *If* $d_l = d_n - 1$ *and* $J = \{j \mid d_j > 0\}$ *then*

   $$(\overline{g}_k)_l = \sum_{j \in J} (\overline{h}_k)_{lj} (\overline{x}_k)_j. \tag{32}$$

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

   $$\overline{f}_k = \sum_{l \in L} (\overline{g}_k)_l (\overline{x}_k)_l + \frac{1}{2} \sum_{(i,j) \in S} (\overline{h}_k)_{ij} (\overline{x}_k)_i (\overline{x}_k)_j. \qquad \Box \tag{33}$$

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

$$X_k(z) = Q^k D(z) \overline{x}_k, \qquad x_k = X_k(\lambda^k), \qquad \chi_k = X_k(0), \qquad (34)$$

$$f_k = \lambda^{kd_n} \overline{f}_k, \qquad \varphi_k = 0, \qquad (35)$$

$$g_k = \lambda^{kd_n} Q^k D(\lambda^{-k}) \overline{g}_k, \qquad \gamma_k = Q^k g_k^\infty, \qquad (36)$$

$$h_k = \lambda^{kd_n} D(\lambda^{-k}) Q^k \overline{h}_k Q^{-k} D(\lambda^{-k}), \qquad \theta_k = Q^k h_k^\infty Q^{-k}, \qquad (37)$$

where

$$(g_k^\infty)_i = (\overline{g}_k)_i \quad \text{if} \quad d_i = d_n \quad \text{and} \quad (g_k^\infty)_i = 0 \quad \text{otherwise}, \quad (38)$$

$$(h_k^\infty)_{ij} = (\overline{h}_k)_{ij} \quad \text{if} \quad d_i + d_j = d_n \quad \text{and} \quad (h_k^\infty)_{ij} = 0 \quad \text{otherwise}. \quad (39)$$

This result is formalized by the following lemma:

Lemma 2. *If* $S(n, p, d, Q, \overline{x}_k, \overline{f}_k, \overline{g}_k, \overline{h}_k)$ *is a seed and* $\lambda \in (0, 1)$ *then there exists a unique dandelion* $\mathcal{D}(S)$ *with* X_k, f_k, g_k, h_k, φ_k, γ_k *and* θ_k *as in* (34)–(39)*. Moreover,* $\mathcal{D}(S)$*'s steps* $s_k = x_{k+1} - x_k$ *are given by*

$$s_k = Q^k D(\lambda^k) \overline{s}_k \qquad (40)$$

for $\overline{s}_k$ *defined in* (31)*.* $\qquad \square$

It is easy to apply lemma 1 and theorem 3 to $\mathcal{D}(S)$:

Lemma 3. *If* $S(n, p, d, Q, \overline{x}_k, \overline{f}_k, \overline{g}_k, \overline{h}_k)$ *is a seed,* $\lambda \in (0, 1)$*,* x_k*,* f_k*,* g_k *and* h_k *are given by* (34)–(37) *and the functions* $\psi_r \in \mathrm{Lip}^2(\mathbb{R})$*, defined for* $0 \leq r < p$*, satisfy*

$$\psi_r(i) = f_{r+i}, \qquad \psi'_r(i) = s_r^t g_{r+i}, \qquad \psi''_r(i) = s_r^t h_{r+i} s_r \qquad (41)$$

for $i \in \{0, 1\}$ *then the functions* $F_k(w, z) = z^{d_n} \psi_{k \bmod p}(w)$ *satisfy* (30)*.* $\qquad \square$

To build a symmetric dandelion and specify the value of the objective function along the search segments $\{x_k + w s_k, w \in [0, 1]\}$ it is enough to find the right seed and functions ψ_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 $F(x_k + w s_k) = \lambda^{k d_n} \psi_{k \bmod p}(w)$ for x_k, f_k, g_k and h_k in (34)–(37) is guaranteed by lemma 3 and theorem 3. To find a seed we proceed as in section 2: we plug (34)–(37) into the expressions that define

(a)   the method of our interest,

(b)   the compatibility conditions in the definition of seed and theorem 3,

(c)   additional constraints, like the Wolfe conditions in section 5 and the convexity conditions in section 6.

and analyze the result. If equations (34)–(37) 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 constraints and may even find closed form solutions for them. If equations (34)–(37) are not related to the method's symmetries then the dandelion will not bloom.

5   The Wolfe conditions

In this section we show that the Wolfe conditions

$$f_{k+1} - f_k \leq \sigma g_k^t s_k \tag{42}$$

and

$$s_k^t g_{k+1} \geq \beta g_k^t s_k \tag{43}$$

may not prevent the cyclic behavior illustrated in figures 1 and 2. In fact, we can have cyclic behavior even if we replace the second Wolfe condition (43) by the stronger requirement $s_k^t g_{k+1} = 0$, 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* $S(n, p, d, Q, \overline{x}_k, \overline{f}_k, \overline{g}_k, \overline{h}_k)$ *be a seed and* $\mathcal{D}(m, n, p, \lambda, x_k, f_k, g_k, h_k)$ *the corresponding dandelion.*

(a)   *If* $\overline{s}_r^t \overline{g}_{r+1} = 0$ *for* $0 \leq r < p$ *then* $s_k^t g_{k+1} = 0$ *for all* k.

(b) *If* $\overline{s}_r^t \overline{g}_r < 0$ *for* $0 \le r < p$ *then the first Wolfe condition* (42) *holds for*

$$\sigma < \sigma_0 = \min_{0 \le r < p} \left\{ \frac{\lambda^{d_n} \overline{f}_{r+1} - \overline{f}_r}{\overline{s}_r^t \overline{g}_r} \right\}.$$

(c) *If* $\overline{s}_r^t \overline{g}_r < 0$ *for* $0 \le r < p$ *then the second Wolfe condition* (43) *is verified for*

$$\beta > \beta_0 = \max_{0 \le r < p} \frac{\overline{s}_r^t \overline{g}_{r+1}}{\overline{s}_r^t \overline{g}_r}. \qquad \square$$

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 [3]. 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 4. *Let* $S(n, p, d, Q, \overline{x}_k, \overline{f}_k, \overline{g}_k, \overline{h}_k)$ *be a seed,* $\lambda \in (0, 1)$ *and* x_k, χ_k, f_k, g_k *and* h_k *be given by* (34)–(37). *Let* $\mathfrak{R}$ *be the convex hull of* $\{\chi_k, 0 \le k \le p\}$. *Consider the lines* $\mathfrak{L}_k = \{\chi_k + w(\chi_{k+1} - \chi_k), w \in \mathbb{R}\}$ *and assume that* $\mathfrak{L}_r \cap \mathfrak{L}_k \cap \mathfrak{R} = \emptyset$ *for* $r + 1 < k < r + p - 1$. *If, for* $0 \le r < p$,

$$s_r^t g_r < f_{r+1} - f_r < s_r^t g_{r+1}, \quad s_r^t h_r s_r > 0 \quad and \quad s_r^t h_{r+1} s_r > 0 \qquad (44)$$

then there exist a function $F \in \mathrm{LC}^2(\mathbb{R}^n)$ *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* $s_k^t \nabla^2 F(x_k + ws_k)s_k > 0$ *for all* $w \in \mathbb{R}$. $\square$

If the vector d that defines the seed S has more than two entries equal to 0 then the lines $\mathfrak{L}_r$ and $\mathfrak{L}_k$ and $r + 1 < k < r + p - 1$ are disjoint for almost all choices of the points $\overline{x}_k$. Therefore, to build examples of divergence in which

---

[3]By strict convexity along the search line we mean that the directional second derivatives $s_k^t \nabla^2 F(x_k + ws_k)s_k$ are positive.

the objective function is strictly convex along the search lines we can focus on
(44) 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 [13] 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 [13] 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., $s_k^t g_{k+1} = 0$. In
this case the Hessian approximations B_k are updated according to the formula:

$$B_{k+1} = B_k + \frac{\alpha_k}{s_k^t g_k} g_k g_k^t - \frac{1}{s_k^t g_k}(g_{k+1} - g_k)(g_{k+1} - g_k)^t, \qquad (45)$$

where $g_k = \nabla F(x_k)$. The iterates x_k evolve according to

$$s_k = -\alpha_k B_k^{-1} g_k \qquad \text{and} \qquad x_{k+1} = x_k + s_k. \qquad (46)$$

Equations (45)–(46) 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 (45)–(46) then $\overline{F}(x) = F(\lambda^{-1} Q^t x)$, $\overline{B}_k = \lambda^{-2} Q B_k Q^t$
and $\overline{x}_k = \lambda Q x_k$ 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

$$B_k = -\sum_{i=0}^{n-1} \frac{\alpha_{k+i}}{g_{k+i}^t s_{k+i}} g_{k+i} g_{k+i}^t, \qquad (47)$$

combined with the conditions

$$s_k^t g_k < 0 \quad \text{and} \quad s_k^t g_{k+j} = 0 \qquad \text{for} \qquad 1 \leq j < n \quad \text{and} \quad k \in \mathbb{N}, \quad (48)$$

then equation (46) is automatically satisfied and (45) holds if

$$g_{k+n} = \rho_k (g_{k+1} - g_k), \qquad (49)$$

for ρ_k such that

$$\alpha_{k+n} = \frac{s_{k+n}^t g_{k+n}}{s_k^t g_k \rho_k^2}. \qquad (50)$$

Notice that if we assume (48) and take $\alpha_k = 1$ for $0 \leq k < n$, α_k in (50) for $k \geq n$ then the α_k are positive and the vectors $g_k, \dots g_{k+n-1}$ are linearly independent, because if we suppose that $\sum_{j=0}^{n-1} \mu_j g_{k+j} = 0$ with $\mu_0 = 1$ then (48) leads to the contradiction $0 = s_k^t \sum_{j=0}^{n-1} \mu_j g_{k+j} = s_k^t g_k < 0$. Therefore, under (48)–(50) the matrices B_k defined by (47) are positive definite and to build an example of divergence for the BFGS method we can ignore α_k and B_k and focus on (48)–(49).

To apply the theory above we plug (34)–(37) into (48) and (49) and deduce that a seed $S(n, p, d, Q, \overline{x}_k, \overline{f}_k, \overline{g}_k, \overline{h}_k)$ is consistent with the BFGS method if

$$\overline{s}_k^t \overline{g}_k < 0 \quad \text{and} \quad \overline{s}_k^t D(\lambda^{-j}) Q^j \overline{g}_{k+j} = 0 \quad \text{for} \quad 1 \leq j < n, \qquad (51)$$

$$Z^n \overline{g}_{k+n} = \rho_k \left(Z \overline{g}_{k+1} - \overline{g}_k \right), \qquad (52)$$

where $Z = \lambda^{d_n} D(\lambda^{-1}) Q$. If we fix Z then (52) becomes similar to an eigenvalue problem, with the ρ's playing the role of eigenvalues and the $\overline{g}$'s of eigenvectors. We solved equations (52) for $\overline{g}_k$ and ρ_k in the case

$$Q = \begin{pmatrix} \cos \frac{\pi}{33} & -\sin \frac{\pi}{33} & & & & \\ \sin \frac{\pi}{33} & \cos \frac{\pi}{33} & & & & \\ & & -1 & & & \\ & & & 1 & & \\ & & & & -1 & \\ & & & & & 1 \end{pmatrix}, \quad D(\lambda) = \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \lambda & & \\ & & & & \lambda & \\ & & & & & \lambda^3 \end{pmatrix},$$

$\lambda = \sqrt[11]{0.9}$ 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 $\lambda = \sqrt[11]{0.9}$ then there exist vectors $\overline{g}_k \in \mathbb{R}^6$ and coefficients and $\rho_k \in \mathbb{R}$ that satisfy (52) and are such that $\overline{g}_{k+11} = \overline{g}_k$, $\overline{\rho}_{k+11} = \overline{\rho}_k$ and the six vectors in the set*

$$\left\{ D(\lambda^{-j}) Q^j \overline{g}_{k+j}, \ j = 0, 1, \ldots, 5 \right\} \tag{53}$$

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

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

$$\overline{s}_k^t \overline{g}_k = -1 \quad \text{and} \quad \overline{s}_k^t D(\lambda^{-j}) Q^j \overline{g}_{k+j} = 0 \quad \text{for} \quad 1 \leq j < 6. \tag{54}$$

The vectors $\overline{s}_k$ obtained from (54) define steps s_k by (40):

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

and we claim that the points

$$\overline{x}_0 \ = \ -\left(I - X^{11}\right)^{-1} \sum_{j=0}^{10} X^j \overline{s}_j, \tag{55}$$

$$\overline{x}_k \ = \ X^{-k} \left(\overline{x}_0 + \sum_{j=0}^{k-1} X^j \overline{s}_j \right) \qquad \text{for} \quad k > 0 \tag{56}$$

satisfy

$$\overline{x}_{k+66} = \overline{x}_k \tag{57}$$

and lead to points $x_k = X^k \overline{x}_k$ such that $s_k = x_{k+1} - x_k$. In fact, using (56) it is straightforward to deduce that $x_{k+1} - x_k = X^{k+1} \overline{x}_{k+1} - X^k \overline{x}_k = X^k \overline{s}_k = s_k$. To verify (57), notice that (56) yields

$$\overline{x}_{k+66} = X^{-k} \left(\overline{x}_0 + \sum_{j=66}^{k+65} X^{j-66} \overline{s}_j + X^{-66} \Delta \right) \tag{58}$$

where

$$\Delta = \left(I - X^{66}\right) \overline{x}_0 + \sum_{j=0}^{65} X^j \overline{s}_j. \tag{59}$$

Now, since $\overline{g}_{k+11} = \overline{g}_k$, equation (54) implies that $\overline{s}_{k+11} = \overline{s}_k$ and

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

Combining this equation, (55) and (59) we conclude that $\Delta = 0$ and equation (58) and the identity $\overline{s}_{j-66} = \overline{s}_j$ lead to (57).

In the proof of lemma 5 we show that the $\overline{g}_k$ can be chosen so that the vectors $\overline{s}_k$ defined by (54) satisfy the linear independence requirements in the definition of seed. The $\overline{x}_k$ and $\overline{g}_k$ above, $d = (0, 0, 0, 1, 1, 3)$,

$$\overline{f}_k = 1 \qquad \text{and} \qquad \overline{h}_k = D'(0)\overline{s}_k \overline{s}_k^t D'(0) \qquad (60)$$

are also compatible with this definition. As a consequence, $n = 6$, $p = 66$, d, Q $\overline{x}_k, \overline{f}_k, \overline{g}_k$ and $\overline{h}_k$ define a seed S. This seed leads to a dandelion $\mathcal{D}(S)$ with steps s_k and gradients g_k compatible with the BFGS method with the matrices B_k in (45) and step sizes α_k in (50). Equations (34)–(37) and (54) imply that the function values and gradients associated to $\mathcal{D}(S)$ satisfy

$$s_k^t g_k = -\lambda^{3k} < \lambda^{3k+3} - \lambda^{3k} = f_{k+1} - f_k < 0 = s_k^t g_{k+1}. \qquad (61)$$

Lemma 4 shows that the iterates x_k corresponding to $\mathcal{D}$ satisfy the first Wolfe condition for $0 < \sigma < 1 - \lambda^3$. Moreover, almost all choices of the vectors $\overline{g}_k$ in the proof of lemma 5 lead to lines $\mathcal{L}_k = \{D(0)Q^k \overline{x}_k + wD(0)Q^k \overline{s}_k, \ w \in \mathbb{R}\}$ such that $\mathcal{L}_r \cap \mathcal{L}_k = \emptyset$ for $r + 1 < k < r + 65$ and $s_k^t D'(0)s_{k+1} \neq 0$. Equation (60) shows that $s_k^t h_k s_k > 0$ and $s_k^t h_{k+1} s_k > 0$. As a result, equation (61) and theorem 4 yield an objective function $F \in \mathrm{LC}^2(\mathbb{R}^6)$ such that the iterates x_k generated by applying the BFGS method to F with $x_0 = \overline{x}_0$ and B_k above satisfy $s_k^t \nabla^2 F(x_k + ws_k)s_k > 0$ 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 Wolfe conditions to prevent the cyclic behavior in figures 1 and 2, even when the objective function is strictly convex along the search lines and has compact level sets.

You may now be wondering if we could not find a simpler example, in dimension less than 6. Unfortunately, we were not able to find such example because

the eigenvalue problem (52) does not seem to have real solutions $\rho_k \neq 0$ for $n < 6$. Notice that this is a purely algebraic issue, which has nothing to do with nonlinear programming. Similarly, the minimum dimension at which we could build counter examples with the techniques described here for other methods could be bigger or smaller than 6, depending on the particular symmetries of the method and the existence of solutions for the corresponding algebraic problems.

8 Appendix

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 $H : E \to \mathbb{H}^n$ are functions with domain E and M is a constant. If*

$$\|H(x) - H(y)\| \le M\|y - x\|, \tag{62}$$

$$\|G(y) - G(x) - H(x)(y - x)\| \le M\|y - x\|^2, \tag{63}$$

$$|F(y) - F(x) - G(x)^t(y - x) - \frac{1}{2}(y - x)^t H(x)(y - x))|$$
$$\le M\|y - x\|^3 \tag{64}$$

then there exists $\overline{F} \in \mathrm{Lip}^2(\mathbb{R}^n)$ such that $\overline{F}(x) = F(x)$, $\nabla \overline{F}(x) = G(x)$ and $\nabla^2 \overline{F}(x) = H(x)$ for $x \in E$. Moreover, there exists constants C and R such that if $\|x\| \ge R$ then $\nabla^2 \overline{F}(x)$ is positive definite and $\|\nabla^2 \overline{F}(x)^{-1}\| \le R$. $\qquad\square$

Whitney's theorem is stated in different levels of generality in the pure mathematics literature. The most complete and general approach is due to C. Fefferman [7, 8] but, unfortunately, it is stated in a language that is too abstract for the average non linear programming researcher. In page 48 of [10], L. Hörmander presents a more concrete version of the theorem for functions in C^m. This version of the theorem is stated using n-uples to denote partial derivatives, i.e., if $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, $|\alpha| = \sum \alpha_i$ and f is a function with partial derivatives of order $|\alpha|$ then $\partial^\alpha f(x)$ is defined as

$$\partial^\alpha f(x) = \frac{\partial^{\alpha_1} \partial^{\alpha_2} \ldots \partial^{\alpha_k}}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_n^{\alpha_k}} f(x).$$

Moreover, $\alpha! = \prod \alpha_i$ and $x^\alpha = \prod x^{\alpha_i}$. Using this notation, the arguments used to prove theorem 2.3.6 in [10] can be adapted to prove the following version of Whitney's theorem:

Theorem 5. *Let E be a bounded set in $\mathbb{R}^n$ and consider, for each $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq m$, a Lipschitz continuous function $u_\alpha : E \to \mathbb{R}$. If there exists a constant M such that*

$$\left| u_\alpha(x) - \sum_{|\beta| \leq m-|\alpha|} \frac{1}{\beta!} u_{\alpha+\beta}(y)(x-y)^\beta \right| \leq M \|x-y\|^{m-|\alpha|+1}$$

for all $x, y \in E$ and $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq m$ then there exists $f \in \mathrm{Lip}^m(\mathbb{R}^n)$ such that $\partial^\alpha f(x) = u_\alpha(x)$ for all $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq m$ and $x \in E$. Moreover, there exists a constant C such that $|\partial^\alpha f(x)| \leq C$ for all $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq m$. $\square$

This theorem allows us to extend much of the discussion in this work to $\mathrm{Lip}^m(\mathbb{R}^n)$ for $m > 2$. However, for $m > 2$ we cannot group the partial derivatives in vectors (gradients) and matrices (Hessians) as we did in the statement of lemma 6. As a consequence, the geometry behind the examples gets blurred by the technicalities as m increases and we decided it would be best to focus on the consequences of lemma 6 instead of exploring more general results like theorem 5.

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

$$S_k(z) = X_{k+1}(z) - X_k(z), \qquad \overline{X}_k(w, z) = X_k(z) + w S_k(z) \qquad (65)$$

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

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

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 (65) 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 $\|y_j - y_k\| \le \delta$ then either*

$$(a)\ j \equiv k \bmod p, \quad (b)\ j \equiv k+1 \bmod p \quad or \quad (c)\ k \equiv j+1 \bmod p \quad (66)$$

and in case (a)

$$\|y_j - y_k\| \ge \delta \left(|w_j - w_k| + |z_j - z_k| \right), \tag{67}$$

in case (b)

$$\|y_j - y_k\| \ge \delta \left(|w_j| + |z_j - z_k| + |1 - w_k| \right) \tag{68}$$

and in case (c)

$$\|y_j - y_k\| \ge \delta \left(|1 - w_j| + |z_j - z_k| + |w_k| \right). \qquad \square \tag{69}$$

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 $y_{m+1} = z$,

$$\sum_{i=0}^{m} \|y_{i+1} - y_i\| \le K \|x - z\|, \tag{70}$$

$$\|H(y_i) - H(y_{i+1})\| \le K \|y_{i+1} - y_i\|, \tag{71}$$

$$\|G(y_{i+1}) - G(y_i) - \frac{1}{2}(H(y_{i+1}) + H(y_i))(y_{i+1} - y_i)\|$$
$$\le K \|y_{i+1} - y_i\|^2, \tag{72}$$

$$\|F(y_{i+1}) - F(y_i) - \frac{1}{2}(G(y_{i+1}) + G(y_i))(y_{i+1} - y_i)\|$$
$$\le K \|y_{i+1} - y_i\|^3, \tag{73}$$

then all $x, y \in E$ with $\|x - y\| \le \delta$ satisfy (62) − (64) with $M = 3K^4$. $\qquad \square$

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 (26)–(29) 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*

$$\|H_k(u, \lambda^k) - H_k(w, \lambda^k)\| \leq M|u - w|, \tag{74}$$

$$\|G_k(u, \lambda^k) - G_k(w, \lambda^k) - \frac{1}{2}(H_k(u, \lambda^k) + H_k(w, \lambda^k))\Delta h\|$$
$$\leq M|u - w|^2, \tag{75}$$

$$\|F_k(u, \lambda^k) - F_k(w, \lambda^k) - \frac{1}{2}(G_k(u, \lambda^k) + G_k(w, \lambda^k))^t \Delta h\|$$
$$\leq M|u - w|^3, \tag{76}$$

$$\|H_k(w, \lambda^j) - H_k(w, \lambda^k)\| \leq M|\lambda^k - \lambda^j|, \tag{77}$$

$$\|G_k(w, \lambda^j) - G_k(w, \lambda^k) - \frac{1}{2}(H_k(w, \lambda^j) + H_k(w, \lambda^k))\Delta v\|$$
$$\leq M|\lambda^k - \lambda^j|^2, \tag{78}$$

$$\|F_k(w, \lambda^j) - F_k(w, \lambda^k) - \frac{1}{2}(G_k(w, \lambda^j) + G_k(w, \lambda^k))^t \Delta v\|$$
$$\leq M|\lambda^k - \lambda^j|^3, \tag{79}$$

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 (65). $\qquad\square$

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 $\psi \in \mathrm{Lip}^2(\mathbb{R})$ 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 $w \in \mathbb{R}$.* $\qquad\square$

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

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 $i \in \{0, 1\}$,*

$$F(i, 0) = 0, \quad \nabla F(i, 0) = 0 \quad and \quad \nabla^2 F(i, 0) = 0 \tag{80}$$

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

$$G(i, z) = 0, \qquad G(w, z)^t Y_3(z) = \frac{\partial F}{\partial w}(w, z),$$
$$G(w, 0)^t V(w) = \frac{\partial F}{\partial z}(w, 0), \tag{81}$$

$$\frac{\partial G}{\partial z}(w, 0)^t V(w) + G(w, 0)^t U(w) = \frac{\partial^2 F}{\partial z^2}(w, 0), \tag{82}$$

$$H(i, z) = 0, \qquad H(w, z) Y_3(z) = \frac{\partial G}{\partial w}(w, z),$$
$$H(w, 0) V(w) = \frac{\partial G}{\partial z}(w, 0) \tag{83}$$

for

$$V(w) = (1 - w) Y_1'(0) + w Y_2'(0) \quad and$$
$$U(w) = (1 - w) Y_1''(0) + w Y_2''(0). \qquad \square$$

Proof of Theorem 2. 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 $j \equiv k \bmod p$. The periodicity of φ_k, γ_k and θ_k given by (22), item 2 in definition 1 and the bounds (22)–(25) 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 (70)–(73) in lemma 8. Therefore, lemmas 6 and 8 imply that there exists a function F as required by theorem 2. $\square$

Proof of Theorem 3. Consider k_0 and δ obtained from lemma 7 and

$$E = \bigcup_{k=k_0}^{\infty} \left\{ \overline{X}_k(w, \lambda^k), \ w \in [a_k, b_k] \right\}.$$

Lemma 7 and the three equations in (26) imply that the expressions

$$F(\overline{X}_k(w, \lambda^k)) = F_k(w, \lambda^k),$$
$$G(\overline{X}_k(w, \lambda^k)) = G_k(w, \lambda^k),$$
$$H(\overline{X}_k(w, \lambda^k)) = H_k(w, \lambda^k)$$

define functions F, G and H with domain E, i.e., if $\overline{X}_j(w_j, \lambda^j) = \overline{X}_k(w_k, \lambda^k)$ 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 3 follows from lemma 6. $\qquad\square$

Proof of Theorem 4.   For $\rho > 0$, consider the compact convex set

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

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

$$\mathfrak{L}_r \cap \aleph_{2\epsilon} = \{X_r(0) + w S_r(0), \quad w \in [a_r, b_r]\}.$$

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

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

$$\mathfrak{L}_k \cap \aleph_\epsilon = \big\{X_k(0) + w S_k(0), \quad w \in [c_k, d_k]\big\}$$

and vectors U_k, V_k such that $U_{k+p} = U_k$, $V_{k+p} = V_k$, $U_k^t S_k(0) < 0$, $V_k^t S_k(0) > 0$ and

$$U_k^t(x - X_k(0) - c_k S_k(0)) \leq 0 \quad \text{and} \quad V_k^t(x - X_k(0) - d_k S_k(0)) \leq 0 \quad \text{for} \quad x \in \Re_\epsilon.$$

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

$$\mu \sqrt[3]{\|S_k(0)\|} < -U_k^t S_k(0) \quad \text{and} \quad \mu \sqrt[3]{\|S_k(0)\|} < V_k^t S_k(0) \quad \text{for} \quad k \in \mathbb{N}$$

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

$$F(x) = \overline{F}(x) + \frac{L}{\mu^3} \sum_{r=0}^{p} \Big(\tau(U_r^t(x - X_r(0) - c_r S_r(0))) $$
$$+ \tau(V_r^t(x - X_r(0) - d_r S_r(0))) \Big)$$

and $k_0 > k_1$ such that $\mu \sqrt[3]{\|s_k\|} < -3 U_k^t s_k$ and $\mu \sqrt[3]{\|s_k\|} < 3 V_k^t s_k$,

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

for $k > k_0$ are as required by theorem 4. In fact, F coincides with $\overline{F}$ in $\Re_\epsilon$ and if $k > k_0$ and $w > b_k$ then

$$s_k^t \frac{\partial^2 F}{\partial w^2}(x_k + w s_k) s_k \geq s_k^t \frac{\partial^2 \overline{F}}{\partial w^2}(s_k + b_k s_k) s_k - L(w - b_k)\|s_k\|$$
$$+ \frac{6L}{\mu^3}(V_r^t s_k)^2 V_k^t(x_k - b_k s_k - X_k(0) - d_k S_r(0)) \qquad \Box$$
$$+ (w - b_k)\frac{6L}{\mu^3}(V_r^t s_k)^3 > L(w - b_k)\|s_k\|.$$

Proof of Lemma 1. Let F be the function obtained by applying theorem 2 to $\mathcal{D}$. Lemma 12 applied to the functions $\tilde{F} = F_k - F$, $Y_1 = X_k$ and $Y_2 = X_{k+1}$ yield functions $\tilde{G}_k$ and $\tilde{H}_k$ such that $G_k(x) = \nabla F(x) + \tilde{G}_k(x)$ and $H_k(x) = \nabla^2 F(x) + \tilde{H}_k(x)$ are as claimed in lemma 1. $\qquad \Box$

Proof of Lemma 2. Equations (34)–(37) show that

$$X_k(z) = \sum_{j=1}^{n} z^{d_j}(\overline{x}_r)_j e_j \qquad \text{and} \qquad g_k = \sum_{i=1}^{n} \lambda^{k(d_n - d_i)}(\overline{g}_k)_i e_i \qquad (84)$$

where $e_i \in \mathbb{R}^n$ is the vector with $e_{ii} = 1$ and $e_{ij} = 0$ for $i \neq j$, Item 4b in definition 7 implies that if $T = \{(i, j) \text{ such that } d_i + d_j \leq d_n\}$ then

$$h_k = \sum_{(i,j) \in T} \lambda^{k(d_n - d_i - d_j)} (\overline{h}_k)_{ij} e_i e_j^t. \tag{85}$$

Since $Q^p = I$ and $\overline{x}_{k+p} = \overline{x}_k$, the X_k in (34) satisfy item (a) in the definition of dandelion. The relations $\overline{f}_{k+p} = \overline{f}_k$, $\overline{g}_{k+p} = \overline{g}_k$ and $\overline{h}_{k+p} = \overline{h}_k$ imply that f_k, g_k and h_k accumulate at the limits φ_k, γ_k and θ_k in the last column of (34)–(37) and these limits satisfy (22). We now verify (23)–(25). Equation (85) yields (23). The first equation in (84) leads to

$$\Delta_k = x_k - \chi_k = \sum_{j \in J} \lambda^{k d_j} (\overline{x}_k)_j, \tag{86}$$

where $J = \{j \mid d_j > 0\}$. The second equation in (84) and (85) imply that

$$\|g_k - \gamma_k - \theta_k^\infty \Delta_k\| \leq n \lambda^k \sum_{u \in U} |(\overline{g}_k)_u - \sum_{j \in J} (\overline{h}_k)_{uj} (\overline{x}_k)_j| + O(\lambda^{2k}) \tag{87}$$

for $U = \{u \mid d_u = d_n - 1\}$ and equations (87) and (32) show that the g_k satisfy (24). Equations (38), (39) and (86) lead to

$$g_k^\infty \Delta_k = \lambda^{k d_n} \sum_{j \in G} (\overline{g}_k)_j (\overline{x}_k)_j \quad \text{and} \quad \Delta_k^t \theta_k^\infty \Delta_k = \lambda^{k d_n} \sum_{(i,j) \in T} (\overline{h}_k)_{ij} (\overline{x}_k)_i (\overline{x}_k)_j,$$

for L and S in item 4b of definition 7, and (25) follows from (33). Finally, the linear independence requirements in items (b) and (c) of definition 4 follow from item 3 in definition 7. $\qquad \square$

Proof of Lemma 3. Direct computation using (34)–(37) and (41) show that the function F_k satisfies (30). $\qquad \square$

Proof of Lemma 4. To prove this lemma, plug (34)–(37) into the expressions in the hypothesis of lemma 4 and compare the results with (42)–(43). $\qquad \square$

Proof of Lemma 5. The matrix Z in (52) can be written as

$$
Z = \lambda^3 D(\lambda^{-1}) Q =
\begin{pmatrix}
Z_1 & 0 & 0 & 0 & 0 \\
0 & Z_2 & 0 & 0 & 0 \\
0 & 0 & Z_3 & 0 & 0 \\
0 & 0 & 0 & Z_4 & 0 \\
0 & 0 & 0 & 0 & Z_5
\end{pmatrix},
\tag{88}
$$

where the Z_i's are the following 1×1 and 2×2 blocks:

$$
Z_1 = \lambda^3
\begin{pmatrix}
\cos \frac{\pi}{33} & -\sin \frac{\pi}{33} \\
\sin \frac{\pi}{33} & \cos \frac{\pi}{33}
\end{pmatrix},
\quad Z_2 = -\lambda^3, \; Z_3 = \lambda^2, \; Z_4 = -\lambda^2, \; Z_5 = 1.
\tag{89}
$$

If we identify the blocks Z_i with the complex numbers

$$
z_1 = \lambda^3 e^{\frac{\pi i}{33}}, \quad z_2 = -\lambda^3, \quad z_3 = \lambda^2, \quad z_4 = -\lambda^2, \quad z_5 = 1.
\tag{90}
$$

then (52) can be interpreted as the equations

$$
z_k^6 \hat{g}_{k+6} = \rho_r \left(z_k \hat{g}_{j,k+1} - \hat{g}_{j,k} \right),
\tag{91}
$$

in the complex variables $\hat{g}_{j,k}$ given by

$$
\hat{g}_{1,k} = (\overline{g}_k)_1 + i (\overline{g}_k)_2, \qquad \hat{g}_{2,k} = (\overline{g}_k)_3,
\tag{92}
$$

$$
\hat{g}_{3,k} = (\overline{g}_k)_4, \qquad\qquad \hat{g}_{4,k} = (\overline{g}_k)_5, \qquad \hat{g}_{5,k} = (\overline{g}_k)_6.
\tag{93}
$$

In the case $\rho_{k+11} = \rho_k$ and $\overline{g}_{r+11} = \overline{g}_r$ equation (91) is equivalent to

$$
A(1/\rho_0, 1/\rho_2, \ldots, 1/\rho_{10}, -z_j)\hat{g}_j = 0,
\tag{94}
$$

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

$$
A(y, z) =
\begin{pmatrix}
1 & z & 0 & 0 & 0 & 0 & y_0 z^6 & 0 & 0 & 0 & 0 \\
0 & 1 & z & 0 & 0 & 0 & 0 & y_1 z^6 & 0 & 0 & 0 \\
0 & 0 & 1 & z & 0 & 0 & 0 & 0 & y_2 z^6 & 0 & 0 \\
0 & 0 & 0 & 1 & z & 0 & 0 & 0 & 0 & y_3 z^6 & 0 \\
0 & 0 & 0 & 0 & 1 & z & 0 & 0 & 0 & 0 & y_4 z^6 \\
y_5 z^6 & 0 & 0 & 0 & 0 & 1 & z & 0 & 0 & 0 & 0 \\
0 & y_6 z^6 & 0 & 0 & 0 & 0 & 1 & z & 0 & 0 & 0 \\
0 & 0 & y_7 z^6 & 0 & 0 & 0 & 0 & 1 & z & 0 & 0 \\
0 & 0 & 0 & y_8 z^6 & 0 & 0 & 0 & 0 & 1 & z & 0 \\
0 & 0 & 0 & 0 & y_9 z^6 & 0 & 0 & 0 & 0 & 1 & z \\
z & 0 & 0 & 0 & 0 & y_{10} z^6 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}.
\tag{95}
$$

Equation (94) suggests that we take $\rho_0, \rho_1, \ldots, \rho_{10}$ such that the corresponding matrices A in (95) are singular. Given such ρ's we can find vectors $\hat{g}_j \neq 0$ that satisfy (94) and use them to define the vectors $\overline{g}$ taking real and imaginary parts of (92)–(93). This approach leads to the polynomial equations

$$\det(A(y, -z_j)) = 0 \qquad \text{for} \qquad j = 1, 2, 3, 4, 5 \qquad (96)$$

on the $y_r = 1/\rho_r$. To obtain accurate approximations for appropriated ρ_r we take

$$y_6 = -1.9948, \; y_7 = -0.3737, \; y_8 = -1.2355, \; y_9 = 0.9857, \; y_{10} = 0.11717$$

and apply Newton's method to find the remaining y_i. Expression (96) correspond to a system of six real equations and Newton's method starting with

$$y_0 = 2.04831, \qquad y_1 = 3.33798, \qquad y_2 = -1.15867,$$
$$y_3 = 0.300795, \qquad y_4 = -0.634211, \qquad y_5 = -2.44761,$$

converges quickly to a solution of this system. This approach led to rational approximations $\overline{y}'s$ for the y_i's such that $\|\det(A(\overline{y}, -z_j))\| < 10^{-1000}$ for $j = 1, 2, 3, 4$ and 5 and such that the Jacobian of the system (96) is well conditioned at $\overline{y}$. 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 $\overline{y}$ we dropped the last row in each of the matrices $A(y, -z_j)$ and computed highly accurate approximations for the five complex vectors $\hat{g}^j \in \mathbb{R}^{11}$ in (94) normalized by the condition $(\hat{g}^j)_1 = 1$. Using (92)–(93) we obtained accurate approximations to the vectors $\overline{g}_j$ required by lemma 5. Using these approximations we verified that the vectors in (53) are indeed linearly independent and solved the systems (54) for the vectors $\overline{s}_k$. Finally, we computed approximations for the $\overline{x}_k$ in (55)–(56) and verified that the corresponding lines $\mathfrak{L}_k$ and $\mathfrak{L}_r$ in the hypothesis of theorem 4 are at least 10^{-5} apart. Our computations indicate that the exact lines $\mathfrak{L}_k$ and $\mathfrak{L}_r$ 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, $\overline{g}$'s, $\overline{s}$'s and $\overline{x}$'s. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proof of Lemma 6. Applying the version of Whitney's theorem in page 6 of [8] to $\omega(t) = t$ and a properly scaled version of the polynomials

$$P^x(y) = F(x) + G(x)^t(y - x) + \frac{1}{2}(y - x)^t H(x)(y - x)$$

we obtain a function $\tilde{F} \in \text{Lip}^2(\mathbb{R}^n)$ and a constant $C > 0$ such that $\tilde{F}(x) = F(x)$, $\nabla \tilde{F}(x) = \nabla F(x)$ and $\nabla^2 \tilde{F}(x) = \nabla^2 F(x)$ for all $x \in E$ and $\|\nabla^2 \tilde{F}(x)\| \leq 1/C$ for all $x \in \mathbb{R}^n$. Let $R > 0$ be such that $2\|x\| < R$ for all $x \in E$ and consider a polynomial Φ such that

$$\Phi\left(\frac{R^2}{4}\right) = \Phi'\left(\frac{R^2}{4}\right) = \Phi''\left(\frac{R^2}{4}\right) = \Phi(R^2) = \Phi''\left(r^2\right) = 0 \quad \text{and} \quad \Phi'(R^2) = \frac{1}{C}.$$

The constants C and R and the function $\overline{F}$ defined by

$$
\begin{aligned}
\overline{F}(x) &= \tilde{F}(x) & &\text{for } \|x\| \leq \frac{R}{2}, \\
\overline{F}(x) &= \tilde{F}(x) + \Phi(\|x\|^2) & &\text{for } \frac{R}{2} < \|x\| \leq R, \\
\overline{F}(x) &= \tilde{F}(x) + \frac{1}{C}\left(\|x\|^2 - R^2\right) & &\text{for } \|x\| > R,
\end{aligned}
$$

satisfies the requirement of lemma 6. $\qquad\square$

Proof of Lemma 7. By compatibility of $\{[a_k, b_k]\}$ and $\mathcal{D}$ and periodicity $(X_{k+p} = X_k)$, there exists $\epsilon > 0$ for which the segments $\mathfrak{S}_k^\epsilon = \{\overline{X}_k(w, 0), w \in [a_k, b_k]\}$ are such that $\text{dist}(\mathfrak{S}_r^\epsilon, \mathfrak{S}_k^\epsilon) > \epsilon$ if $r + 1 < k < r + p - 1$. If $\delta > 0$ is small enough then the surfaces

$$\mathfrak{P}_k^\delta = \left\{\overline{X}_k(w, z), \ w \in [a_k, b_k], \ z \in [0, \delta]\right\}$$

are such that

$$\text{dist}(\mathfrak{P}_r^\delta, \mathfrak{P}_k^\delta) > \delta \quad \text{for} \quad r + 1 < k < r + p - 1. \tag{97}$$

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

$$
\begin{aligned}
\|a S_k(z) + b X_k'(z) &+ c X_{k+1}'(z)\| \\
&\geq 3\delta(\|a S_k(z)\| + \|b X_k'(z)\| + \|c X_{k+1}'(z)\|),
\end{aligned}
\tag{98}
$$

$$\|aS_k(v) + bX'_k(z) + cS_{k+1}(z)\|$$
$$\geq 3\delta(\|aS_k(v)\| + \|bX'_k(z)\| + \|cS_{k+1}(z)\|). \tag{99}$$

We also have that

$$X_k(v) = X_k(z) + (v - z)X'_k(z) + \Delta_k(v, z) \tag{100}$$

where $\Delta_k(v, z)$ satisfies the inequality

$$\|\Delta_k(v, z)\| < \delta|v - z| \max_{z \in [0,1]} \|X''(z)\|$$

when $v, z \in [0, \delta]$. The linear independence requirements in the definition of dandelion imply that $X'_k(0) \neq 0$ and (100) implies that if $\delta > 0$ is small and $v, z \in [0, \delta]$ then

$$2\|(v - z)X'_k(z)\| \geq \|X_k(v) - X_k(z)\|. \tag{101}$$

We now show that $\delta > 0$ for which the expressions (97)–(101) above are valid fulfils 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 $\overline{X}_i = \overline{X}_j$ the surfaces $\mathfrak{P}_i^\delta$ and $\mathfrak{P}_k^\delta$ are at most δ apart and (97) leaves only the three possibilities: (a) $i = k$, (b) $i = k+1$ or (c) $i = k+p-1$. This corresponds to (66) and to complete this proof we now verify the bounds (67)–(69). In case (a) $X_j = X_k$ and $S_j = S_k$ and the definition of y_j and y_k in the hypothesis and (100) lead to

$$\begin{aligned}
y_j - y_k &= \big(w_j - w_k\big) S_k(z_k) + (1 - w_j)\big(X_k(z_j) - X_k(z_k)\big) \\
&\quad + w_j\big(X_{k+1}(z_j) - X_{k+1}(z_k)\big) \\
&= (w_j - w_k)S_k(z_k) + (1 - w_j)(z_j - z_k)X'_k(z_k) \\
&\quad + w_j(z_j - z_k)X'_{k+1}(z_k) + (1 - w_j)\Delta_k(z_j, z_k) + w_j\Delta_{k+1}(z_j, z_k).
\end{aligned}$$

The bounds (98) and (101) and the abbreviation $(w_j - w_k)S_k(z_k) = \Gamma$ yield

$$\begin{aligned}
\|y_j - y_k\| &\geq 3\delta\big(\|\Gamma\| + \|(1 - w_j)(z_j - z_k)X'_k(z_k)\| + \|w_j(z_j - z_k)X'_{k+1}(z_k)\|\big) \\
&\quad - \|(1 - w_j)\Delta_k(z_k, z_j)\| - \|w_j\Delta_{k+1}(z_k, z_j)\|. \\
&\geq \delta\left(\|\Gamma\| + \|(1 - w_j)(X_k(z_j) - X_k(z_k)) + w_j(X_{k+1}(z_j) - X_{k+1}(z_k)\|\right)
\end{aligned}$$

and (67) follows from the equations $\Gamma = \overline{X}_j(w_j, z_k) - y_k$ and

$$y_j - \overline{X}_j(w_j, z_k) = (1 - w_j)\left(X_k(z_j) - X_k(z_k)\right) + w_j\left(X_{k+1}(z_j) - X_{k+1}(z_k)\right).$$

In case (b) $X_j = X_{k+1}$ and $S_j = S_{k+1}$ and (100) lead to

$$y_j - y_k = w_j S_{k+1}(z_j) + (z_j - z_k)X'_j(z_k) + (1 - w_k)S_k(z_k) + \Delta_j(z_k, z_j)$$

and using (99) and (101) we deduce that

$$\|y_j - y_k\| \geq 3\delta\left(\|w_j S_j(z_j)\| + \|(z_j - z_k)X'_j(z_k)\| + \|(1 - w_k)S_k(z_k)\|\right)$$
$$-\|\Delta_j(z_k, z_j)\| \geq \delta\left(\|w_j S_j(z_j)\| + \|X_j(z_j) - X_j(z_k)\| + \|(1 - w_k)S_k(z_k)\|\right)$$

and (68) 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

$$y_0 = x, \; y_{m+1} = z \quad \text{and} \quad f_j = F(y_j), \; g_j = G(y_j) \quad \text{and} \quad h_j = H(y_j).$$

The bounds (71) and (70) yield

$$\|h_j - h_0\| \leq \sum_{i=1}^{j} \|h_i - h_{i-1}\| \tag{102}$$
$$\leq \sum_{i=1}^{j} K\|y_i - y_{i-1}\| \leq K^2\|x - z\|$$

and (62) is obtained by taking $j = m + 1$ in (102). 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 (70), (72) and (102) and imply that

$$\|g_j - g_0 - h_0(y_j - y_0)\| \leq 2K^3\|x - z\|^2. \tag{103}$$

Taking $j = m + 1$ we deduce that x and z satisfy (63). Finally, notice that

$$f_j - f_0 - g_0^t(y_j - y_0) - \frac{1}{2}(y_j - y_0)^t h_0(y_j - y_0)$$

$$= \sum_{i=1}^{j} \left(f_i - f_{i-1} - \frac{1}{2}(g_i + g_{i-1})^t(y_i - y_{i-1}) \right)$$

$$+ \frac{1}{2} \sum_{i=1}^{j} (g_{i-1} - g_0 - h_0(y_{i-1} - y_0))^t(y_i - y_{i-1}) \qquad (104)$$

$$+ \frac{1}{2} \sum_{i=1}^{j} (g_i - g_0 - h_0(y_i - y_0))^t(y_i - y_{i-1})$$

$$- \frac{1}{2}(y_j - y_0)^t h_0(y_j - y_0) + \frac{1}{2}\Delta$$

for $\Delta = \sum_{i=1}^{j}(y_i + y_{i-1} - 2y_0)^t h_0(y_i - y_{i-1})$. The last terms in (104) cancel because

$$\Delta = \sum_{i=1}^{j} ((y_i - y_0) + (y_{i-1} - y_0))^t h_0 ((y_i - y_0) - (y_{i-1} - y_0))$$

$$= \sum_{i=1}^{j}(y_i - y_0)^t h_0(y_i - y_0) - \sum_{i=1}^{j}(y_{i-1} - y_0)^t h_0(y_{i-1} - y_0)$$

$$= (y_j - y_0)^t h_0(y_j - y_0).$$

Thus, if we take $j = m + 1$ then (70), (103) and (104) yield (64) for $M = 3K^4$. $\square$

Proof of Lemma 9.   The bounds (74) and (77) follow from the Lipschitz continuity of H_k. Equation (75) can be derived from the first equation in (29),

$$\Delta h = \overline{X}_k(u, \lambda^k) - \overline{X}_k(w, \lambda^k) = (u - w)s_k. \qquad (105)$$

and the Lipschitz continuity of the first derivatives of G_k. Equation (76) is a consequence of the first equation in (27) and (29), (105) and the Lipschitz continuity of the second derivatives of F_k. The bounds (78) and (79) are clearly satisfied if $j = k$ and from now on we assume that $j \neq k$. In this case

$$\lambda^j + \lambda^k < \frac{1 + \lambda}{1 - \lambda}|\lambda^j - \lambda^k| \qquad (106)$$

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 $\mu_{j,k} = \int_{\lambda^j}^{\lambda^k} \int_0^\xi (X''(\zeta) - X''(0))d\zeta\,d\xi$ satisfies

$$|\mu_{j,k}| \le L|\lambda^k - \lambda^j||\lambda^k + \lambda^j|^2 \le \frac{L(1+\lambda)^2}{(1-\lambda)^2}|\lambda^k - \lambda^j|^3,$$

for a Lipschitz constant L for X_k''. The bound (106) and $X_k \in \mathrm{Lip}^2([0, 1], \mathbb{R}^n)$ lead to

$$\begin{aligned}
\Delta v &= (1-w)(X_k(\lambda^j) - X_k(\lambda^k)) + w(X_{k+1}(\lambda^j) - X_{k+1}(\lambda^k)) \\
&= (\lambda^j - \lambda^k)V_k(w) + \frac{1}{2}(\lambda^{2j} - \lambda^{2k})U_k(w) + O(|\lambda^j - \lambda^k|^3)
\end{aligned} \tag{107}$$

for

$$U_k = (1-w)X_k'(0) + wX_{k+1}'(0) \quad \text{and} \quad V_k = (1-w)X_k''(0) + wX_{k+1}''(0).$$

The conditions $F_k \in \mathrm{Lip}^2(\mathbb{R}^2)$ and $G_k \in \mathrm{Lip}^1(\mathbb{R}^2, \mathbb{R}^n)$ imply that

$$G_k(w, \lambda^l) = G_k(w, 0) + \lambda^l \frac{\partial G_k}{\partial z}(w, 0) + O(\lambda^{2l}), \tag{108}$$

$$G_k(w, \lambda^j) - G_k(w, \lambda^k) = \frac{\partial G}{\partial z}(w, 0)(\lambda^j - \lambda^k) + O(|\lambda^j - \lambda^k|^2), \tag{109}$$

$$\begin{aligned}
F_k(w, \lambda^j) - F_k(w, \lambda^k) &= \frac{\partial F}{\partial z}(w, 0)(\lambda^j - \lambda^k) + \frac{1}{2}\frac{\partial^2 F}{\partial z^2}(w, 0)(\lambda^{2j} - \lambda^{2k}) \\
&\quad + O(|\lambda^j - \lambda^k|^3).
\end{aligned} \tag{110}$$

The bound (78) follows from the second equation in (29), (107) and (109). Finally, the bound (79) can be deduced from the second equation in (27), equations (28), (106), (107), equation (108) with $l = j$ and $l = k$ and equation (110). $\quad\square$

Proof of Lemma 10.   Given $\epsilon > 0$, consider the function

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

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

$w\overline{g}_0 + w^2\overline{h}_0$	if $w \le 0$,
$w\overline{g}_0 + w^2\overline{h}_0 + a_\epsilon w^3$	if $w \in (0, \epsilon]$,
$w\overline{g}_0 + w^2\overline{h}_0 + a_\epsilon w^3 + b_\epsilon(w - \epsilon)^3$	if $w \in (\epsilon, 2\epsilon]$,
$f_\epsilon + g_\epsilon(2w - 1) + \epsilon(2w - 1)^2$	if $w \in (2\epsilon, 1 - 2\epsilon]$,
$(w - 1)\overline{g}_1 + (w - 1)^2\overline{h}_1 + d_\epsilon(w - 1)^3 + c_\epsilon(w - 1 - \epsilon)^3$	if $w \in (1 - 2\epsilon, 1 - \epsilon]$,
$(w - 1)\overline{g}_1 + (w - 1)^2\overline{h}_1 + d_\epsilon(w - 1)^3$	if $w \in (1 - \epsilon, 1]$,
$(w - 1)\overline{g}_1 + (w - 1)^2\overline{h}_1$	if $w > 1$,

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_ϵ belongs to $\text{Lip}^2(\mathbb{R})$ if and only if it has continuous second order derivatives at $w = 2\epsilon$ and $w = 1 - 2\epsilon$. This condition leads to a linear system of six equations on the six variables a_ϵ, b_ϵ, c_ϵ, d_ϵ, f_ϵ and g_ϵ. Solving this system we obtain

$$a_\epsilon = -\frac{\overline{g}_0 + O(\epsilon)}{6\epsilon^2(1 - 2\epsilon)}, \qquad\qquad d_\epsilon = -\frac{\overline{g}_1 + O(\epsilon)}{6\epsilon^2(1 - 2\epsilon)}. \qquad (111)$$

The second derivative of F_ϵ is a piecewise linear function with values

$$h_0, \quad h_0 + 6\epsilon a_\epsilon, \quad \epsilon, \quad \epsilon, \quad h_1 - 6\epsilon d_\epsilon, \quad h_1 \qquad (112)$$

at the nodes $w = 0$, $w = \epsilon$, $w = 2\epsilon$, $w = 1 - 2\epsilon$, $w = 1 - \epsilon$ and $w = 1$. The hypothesis implies that $\overline{g}_0 < 0$ and $\overline{g}_1 > 0$ and (111) shows that $a_\epsilon > 0$ and $d_\epsilon < 0$ if $\epsilon > 0$ is small. Therefore, (112) implies that $F_\epsilon{}''(w) > 0$ for all w if ϵ is small enough. $\qquad\square$

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 $\overline{\phi} \in \text{Lip}^1(\mathbb{R}^2)$ such that

$$\overline{\phi}(w, 0) = \psi(w, 0), \quad \nabla\overline{\phi}(w, 0) = \nabla\psi(w, 0), \quad \overline{\phi}(i, z) = 0, \quad \nabla\overline{\phi}(i, z) = 0$$

for $i = 0, 1$ and $|w|, |z| \leq 3$. Let $\tau : \mathbb{R} \to \mathbb{R}$ be a C^∞ function such that $\tau(x) = 1$ for $|x| < 2$ and $\tau(x) = 0$ for $|x| > 3$. The function

$$\phi(w, z) = \tau(z)\left(\tau(w)\overline{\phi}(w, z) + (1 - \tau(w))\psi(w, z)\right)$$

satisfies items (a) and (b) in lemma 11. $\qquad\square$

Proof of Lemma 12.   Let Z_1, Z_2 and Z_3 be the functions defined by

$$Z_1(z) = Y_1'(z), \qquad Z_2(z) = Y_2'(z), \qquad Z_3(z) = Y_3(z).$$

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

$$W_i^t Z_i(0) = 1 \qquad \text{and} \qquad W_i^t Z_j(0) = 0 \quad \text{if } i \neq j. \qquad (113)$$

The implicit function theorem guarantees the existence of $\delta > 0$ and functions $a_{ij} \in \text{Lip}^1([0, \delta])$, defined for $1 \leq i, j \leq 3$, such that

$$a_{ii}(0) = 1 \qquad \text{and} \qquad a_{ij}(0) = 0 \quad \text{if } i \neq j \qquad (114)$$

and the vectors

$$A_i(z) = a_{i1}(z) W_1 + a_{i2}(z) W_2 + a_{i3}(z) W_3 \qquad (115)$$

are such that

$$A_i(z)^t Z_i(z) = 1 \qquad \text{and} \qquad A_i(z)^t Z_j(z) = 0 \quad \text{if } i \neq j \qquad (116)$$

for $z \in [0, \delta]$. Lemma 11 applied to $\psi = \frac{\partial F}{\partial z}$ yields $\phi \in \text{Lip}^1(\mathbb{R}^2)$ such that

$$\begin{aligned}
\phi(i, z) &= 0, \\
\nabla\phi(i, z) &= 0, \\
\phi(w, 0) &= \frac{\partial F}{\partial z}(w, 0), \\
\frac{\partial \phi}{\partial z}(w, 0) &= \frac{\partial^2 F}{\partial z^2}(w, 0)
\end{aligned} \qquad (117)$$

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

$$G(w, z) = \frac{\partial F}{\partial w}(w, z) A_3(z) + \phi(w, z)(A_1(z) + A_2(z)) \qquad (118)$$

for $z \in [0, \delta]$ satisfies (81) and (82). Equations (80) and (117) imply that $G(i, z) = 0$ and second and third equations in (81) follows from (114)–(118)

and the definition of $S(z)$ and $V(w)$. To verify (82), notice that (117) and (118) lead to

$$
\begin{aligned}
\frac{\partial G}{\partial z}(w, 0) \; = \; & \frac{\partial^2 F}{\partial w \partial z}(w, 0)W_3 + \frac{\partial F}{\partial w}(w, 0)A_3'(0) \\
& + \frac{\partial^2 F}{\partial z^2}(w, 0)(W_1 + W_2) + \frac{\partial F}{\partial z}(w, 0)(A_1'(0) + A_2'(0)).
\end{aligned}
\tag{119}
$$

Equation (116) yields $A_i'(0)^t Z_j(0) + A_i(0)^t Z_j'(0) = 0$ and (114)–(116) imply that

$$
a_{ij}'(0) = -W_i^t Z_j'(0).
\tag{120}
$$

If $j = 3$ then $Z_j'(0) = Y_3'(0) = Y_2'(0) - Y_1'(0) = Z_2(0) - Z_1(0)$ and (120) leads to

$$
a_{13}'(0) = 1, \qquad a_{23}'(0) = -1, \qquad a_{33}'(0) = 0.
\tag{121}
$$

Reminding that $Z_i'(0) = Y_i''(0)$ and using (116) and (120)–(121) we obtain

$$
A_1'(0) \; = \; -W_1^t Y_1''(0)W_1 - W_1^t Y_2''(0)W_2 + W_3,
\tag{122}
$$

$$
A_2'(0) \; = \; -W_2^t Y_1''(0)W_1 - W_2^t Y_2''(0)W_2 - W_3,
\tag{123}
$$

$$
A_3'(0) \; = \; -W_3^t Y_1''(0)W_1 - W_3^t Y_2''(0)W_2.
\tag{124}
$$

Equations (117) and (118) show that

$$
G(w, 0) = \frac{\partial F}{\partial w}(w, 0)W_3 + \frac{\partial F}{\partial z}(w, 0)(W_1 + W_2)
\tag{125}
$$

and (82) follows from (116), (118), (122)–(124) and the fact that

$$
\begin{aligned}
V(w) &= (1 - w)Z_1(0) + wZ_2(0) \quad \text{and} \\
U(w) &= (1 - w)Y_1''(0) + wY_2''(0).
\end{aligned}
\tag{126}
$$

To complete this proof we define H as

$$
\begin{aligned}
H(w, z) \; = \; & \frac{\partial^2 F}{\partial w^2}(w, z)A_3(z)A_3(z)^t + \Delta(w, z) \\
& + b(w)A_1(z)A_1(z)^t + (b(w) + c(w))(A_2(z)A_1(z)^t \\
& + A_1(z)A_2(z)^t) + c(w)A_2(z)A_2(z)^t
\end{aligned}
\tag{127}
$$

for

$$
\Delta(w, z) = \frac{\partial \phi}{\partial w}(w, z)\left((A_1(z) + A_2(z))A_3(z)^t + A_3(z)(A_1(z) + A_2(z))^t\right), \tag{128}
$$

$$b(w) = \frac{1}{1 - w + w^2} (Z_1(0) - wZ_2(0))^t \frac{\partial G}{\partial z}(w, 0), \qquad (129)$$

$$c(w) = \frac{1}{1 - w + w^2} ((w - 1)Z_1(0) + Z_2(0))^t \frac{\partial G}{\partial z}(w, 0). \qquad (130)$$

Equations (80) and (117) imply that $H(i, z) = 0$ for $i \in \{0, 1\}$ and (118) and (116) imply the second equation in (83). Using (117), (126) and (127)–(130) we get

$$
\begin{aligned}
H(w, 0)V(w) &= ((b(w) + wc(w))W_1 + ((1 - w)b(w) + c(w))W_2 \\
&\quad + \frac{\partial \phi}{\partial w}(w, 0)W_3 \\
&= Z_1(0)^t \frac{\partial G}{\partial z}(w, 0)W_1 + Z_2(0)^t \frac{\partial G}{\partial z}(w, 0)W_2 \\
&\quad + \frac{\partial^2 F}{\partial w \partial z}(w, 0)W_3.
\end{aligned}
\qquad (131)
$$

Equations (113), (119) and (122)–(124) show that

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

Finally, equations (113) shows that the right hand side of (131) is the expansion of $\partial G/\partial z$ $(w, 0)$ on the basis $\{W_1, W_2, W_3\}$ of a tri-dimensional subspace that contains $\partial G/\partial z$ $(w, 0)$ and we have shown (83). $\qquad \square$

REFERENCES

[1] P. Absil, R. Mahony and B. Andrews, Convergence of the iterates of descent methods for analytic functions. Siam Journal of Optimization, **16**(2) (2005), 531–547.

[2] D. Bertsekas, *Nonlinear Programming*, second edition, second printing, Athena Scientific, Belmont, Massachusets, 2003.

[3] B. Cantwell, *Introduction to Symmetry Analysis*, Cambridge University Press, Cambridge UK, (2002).

[4] A. Cauchy, Méthodes générales pour la résolution des systémes d'équations simultanées. C.R. Acad. Sci. Par., **25** (1847), 536–538.

[5] H. Curry, The Method of steepest descent for non-linear minimization problems. Quart. Appl. Math., **2** (1944), 258–261.

[6] Y. Dai, Convergence properties of the BFGS Algorithm. SIAM J. Optim., **13**(3) (2002), 693–701.

[7] C. Fefferman, A generalized sharp Whitney theorem for jets. Revista Matemática Iberoamericana, **21**(2) (2005).

[8] C. Fefferman, Extension of $C^{m,\omega}$-Smooth Functions by Linear Operators, manuscript available at: www.math.princeton.edu/facultypapers/Fefferman. To appear in Revista Matematica Iberoamericana.

[9] C. Gonzaga, Two Facts on the convergence of the Cauchy Algorithm. Journal of Optimization Theory and Applications, **107**(3) (2000), 591–6000.

[10] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*. Springer Verlag, Heidelberg, 1983.

[11] J. Nocedal, Theory of algorithms for unconstrained optimization. Acta numerica, **1** (1992), 199–242.

[12] J. Nocedal and S. Wright, *Numerical Optimization*. Springer series in operations research, Springer, 1999.

[13] W.F. Mascarenhas, The BFGS method with exact line searches fails for non-convex objective functions. Math. Prog. Ser B., **99** (2004), 49–61.

[14] W.F. Mascarenhas, Newton's iterates can converge to non stationary points, Manuscript Submitted to Mathematical Programming. Preliminary version available on line at: www.ime.usp.br/~walterfm/pub/newtonFails.pdf.

[15] M. Powell: Nonconvex minimization calculations and the conjugate gradient method. In: D.F. Griffiths, ed., Numerical Analysis, Lecture Notes in Mathematics 1066, Springer Verlag, Berlin, (1984), 122–141.

[16] M. Powell, Convergence Properties of Algorithms for Nonlinear Optimization SIAM Review, **28**(4) (1986), 487–500.

[17] H. Whitney, Analytic extensions of differentiable functions defined in closed sets. Trans. Amer. Math. Soc., **36** (1934), 63–89.