The divergence of the BFGS and Gauss Newton Methods

Walter F. Mascarenhas*

September 30, 2013

Abstract

We present examples of divergence for the BFGS and Gauss Newton methods. These examples have objective functions with bounded level sets and other properties concerning the examples published recently in this journal, like unit steps and convexity along the search lines. As these other examples, the iterates, function values and gradients in the new examples fit into the general formulation in our previous work *On the divergence of line search methods, Comput. Appl. Math. vol.26 no.1 (2007)*, which also presents an example of divergence for Newton's method.

1 Introduction

In the past ten years a few articles have been published, in this journal and the one mentioned in the abstract, presenting elaborate theoretical examples of divergence of line search methods. These methods start from a point $x_0 \in \mathbb{R}^n$ and iterate according to

$$x_{k+1} := x_k + \alpha_k d_k, \tag{1}$$

with search directions $d_k \in \mathbb{R}^n$ and parameters $\alpha_k \in \mathbb{R}$ chosen with the intention that x_k converge to a local minimizer of a function $f : \mathbb{R}^n \mapsto \mathbb{R}$. Textbooks [2, 14] present popular choices for the directions d_k and the parameters α_k and explain why they work in usual circumstances. The articles [4, 5, 12, 13] analyze d_k's given by BFGS and conclude that these methods may not succeed in extreme situations. Our article [13] presents a similar result for the d_k's corresponding to Newton's method for minimization and hints that the techniques it describes could be also applied to other methods, but it does not elaborate on these possible extensions. Finally, our article [11] also offers examples of unexpected behavior for Newton's method.

These examples have subtle points, but none of the them is really perfect. For instance, the objective functions in our examples in [13] are not explicit and have only Lipschitz continuous second derivatives. The objective function in [5] is an explicit polynomial, but it has no local minimizers, its degree is high and its coefficients are not simple.

The examples are concerned with the behavior of line search methods in situations that are not considered in the hypothesis found in textbooks, or even in the majority of

*Instituto de Matemática e Estatística, Universidade de São Paulo, Cidade Universitária, Rua do Matão 1010, São Paulo SP, Brazil. CEP 05508-090. Tel.: +55-11-3091 5411, Fax: +55-11-3091 6134, walter.mascarenhas@gmail.com

research papers. Usually, textbooks and articles impose reasonable hypothesis, which are frequently met in practice. Their spirit is similar to the following theorem [1]:

Theorem 1 *Consider $n \times n$ matrices $\{M_k,\ k \in \mathbb{N}\}$, positive numbers $\{\alpha_k,\ k \in \mathbb{N}\}$ and a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ with continuous first order derivatives and bounded level sets. Suppose the matrices $M_k M_k^t$ are non singular and the iterates $x_k \in \mathbb{R}^n$ are defined by (1) with*

$$d_k := -\left(M_k M_k^t\right)^{-1} \nabla f(x_k) \tag{2}$$

and satisfy the first Wolfe condition. If there exists $\overline{\alpha} > 0$ such that $\alpha_k \geq \overline{\alpha}$ for all k and the matrices M_k are bounded then $\lim_{k \to \infty} \nabla f(x_k) = 0$ for every starting point x_0.

The level sets of f are of the form $\{x \in \mathbb{R}^n$ with $f(x) \leq a\}$ and the first Wolfe condition is the requirement that there exists $\sigma \in (0,1)$ such that

$$f(x_{k+1}) \leq f(x_k) + \sigma \nabla f(x_k)^t (x_{k+1} - x_k) \tag{3}$$

for all k. Equation (2) describes important nonlinear programming methods, such as steepest descent, BFGS and Gauss Newton. It also applies to an adaptation of Newton's method in which we take the steepest descent direction when we detect that the Hessian $\nabla^2 f(x_k)$ is not positive definite or it is almost singular and we would need a very small α_k in order to obtain a step $s_k = \alpha_k d_k = -\alpha_k \nabla^2 f(x_k)^{-1} \nabla f(x_k)$ of reasonable size.

The bounded level sets hypothesis in Theorem 1 implies that the sequence x_k has a converging subsequence for every x_0. As a result, if the level sets are bounded and we enforce (3) then for every x_0 we either are fortunate and have a subsequence converging to x_∞ with $\nabla f(x_\infty) = 0$ or we suffer from one of these two pathologies:

(i) The parameters α_k in (1) get too small, i.e., a subsequence α_{n_k} converges to 0.

(ii) The matrices M_k in (2) are unbounded.

As a consequence, for methods like Gauss Newton in which the matrices M_k are continuous functions of x_k, we can only have divergence if the parameters α_k have a subsequence converging to 0. Theorem 1 also explains why the α_k's in [11] converge to 0.

There are other hypothesis that guarantee the convergence of line search methods. Some require that the search directions do not get almost orthogonal to the gradient or that Hessians are well conditioned. Others ask for an analytic objective functions or apply to more elaborate classes of objective functions, as in [1, 9]. Textbooks rely on a combination of these hypothesis to prove the convergence of the methods they are concerned with, as we did in Theorem 1. They sacrifice generality for a cleaner exposition of the most common situations, and are quite right in doing so. We go in the opposite direction: we explore the consequences of violating the usual conditions. In theory, we conclude that methods like BFGS and Gauss Newton may fail if the parameters α_k get too small or matrices like the M_k above get too large. This theoretical conclusion disregards rounding errors and the precautions taken in practice. In fact, items (i) and (ii) above show that the examples will not work in practice. We cannot handle an unbounded sequence of matrices or arbitrarily small $\alpha_k > 0$ on a real computer. We would also have trouble working with the inverses of such matrices.

More than presenting particular examples, we expose the neat analytical, algebraic and geometric concepts underlying them. There is a subtle relation between the examples presented here and some techniques to find closed form solutions of nonlinear

[1] For a proof of this theorem, look at equation (67) in the appendix.

differential equations: they can be both explained in terms of symmetry groups. By adding a simple term to a nonlinear differential equation with closed form solutions we may destroy its symmetries and turn it into an equation for which one can prove that there are no convenient closed form solutions. Examples of divergence are similar. We did build examples for Newton's method, BFGS and Gauss Newton using our tools. However, each example relies on specific features of the method it considers.

This article has four more sections and one appendix. Section 2 overviews the previous examples. It shows how they fit in the framework in [13]. Section 3 explains the analytical and algebraic underpinnings of the examples. Section 4 builds an example of divergence for the Gauss Newton method. Section 5 is about the divergence of BFGS. The appendix contains proofs and corroborates our claim that the examples in [5, 12, 13] have similar iterates, function values and gradients. The supplementary material aims to facilitate the reader in using the software Mathematica to verify the algebra in the examples.

2 Overview of the examples in [5, 12, 13]

The examples in [5, 12, 13] are based on classical mathematical ideas. From the numerical point of view, Powell's work [18] already presents an interesting analysis of divergence in the same context of functions with second order derivatives and bounded level sets that we consider here. From a broader perspective, our ideas and Powell's are just a natural extension of the mathematical techniques used to analyze periodic orbits in celestial mechanics in the late 1800s.

The basic ideas behind our work and Powell's are present in the first volume of Poincaré's masterpiece [17], which was published in 1892. In that book Poincaré uses power series to analyze the convergence of solutions of differential equations to periodic orbits; Powell uses a similar technique to analyze the convergence of the iterates of a version of the conjugate gradient method to a limit cycle. In 1901, Hadarmard [8] looked at the same problem from a perspective that is quite similar to the one we present in this article. He was then followed by Cotton [3] and Perron [16].

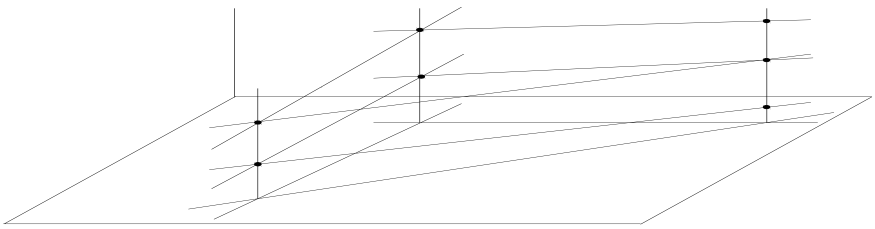

Figure 1: The geometry of divergence: the iterates (black dots) converge to the vertices of a polygon. The search lines connect consecutive iterates. They converge to the horizontal lines containing the sides of the polygon.

From Hadamard's perspective, the dynamics of all the examples of divergence in [5, 12, 13] is described by Figure 1. In this figure at each step the vertical coordinate is contracted by a factor $\lambda \in (0,1)$ and the horizontal coordinates are rotated. The iterates converge to a cycle in the horizontal subspace, which contains no critical points. In the examples the vertical and horizontal subspaces may have higher dimensions, the limiting polygon has more vertices and the rotations are replaced by orthogonal transformations, but the qualitative picture is the same.

The examples discussed here have features summarized in the following table:

Feature	Article and year of publication		
	[5] – 2012	[13] – 2007	This article – 2014
Example for Newton's method	No	Yes	No
Example for BFGS	Yes	Yes	Yes
Example for Gauss Newton	No	No	Yes
Bounded level sets	No	Yes	Yes
Smoothness of the Objective function	Explicit Polynomial	Lipschitz continuous second derivatives	Lipschitz continuous second derivatives
step size (α_k)	One	Any for Newton, different from one for BFGS	One for BFGS Converges to 0 for Gauss Newton
Convexity along search lines	Yes	Yes	Yes
Armijo and Wolfe conditions	Yes	Yes	Yes
Goldstein condition	Yes	No	Yes
Exact line searches	Yes	Yes	Yes

The examples have the same kind of iterates x_k, function values $f_k = f(x_k)$ and gradients $g_k = g(x_k)$, which can be written as

$$x_k = Q^k D(\lambda)^k \overline{x}_k, \qquad f_k = \lambda^{kd_n} \overline{f}_k \qquad \text{and} \qquad g_k = \lambda^{kd_n} Q^k D(\lambda)^{-k} \overline{g}_k. \quad (4)$$

The matrix Q in (4) is orthogonal and $Q^p = I$ for a period $p \in \mathbb{N}$. The parameters $\overline{x}_k \in \mathbb{R}^n$, $\overline{f}_k \in \mathbb{R}$ and $\overline{g}_k \in \mathbb{R}^n$ also have period p, in the sense that $\overline{x}_{k+p} = \overline{x}_k$, $\overline{f}_{k+p} = \overline{f}_k$ and $\overline{g}_{k+p} = \overline{g}_k$. The matrices $D(\lambda)$ are diagonal. They commute with Q and their diagonal entries are powers of the parameter $\lambda \in (0,1)$. The constant d_n is equal to the biggest exponent of λ in the diagonal of $D(\lambda)$ (The appendix explains how the equations in [5, 12, 13] fit into (4).)

The parameters in (4) mix well with Hessians of the form

$$h_k = \lambda^{kd_n} Q^k D(\lambda)^{-k} \overline{h}_k D(\lambda)^{-k} Q^{-k} \quad (5)$$

and, for the BFGS method, with Hessian approximations of the form

$$B_k = -\sum_{i=0}^{n-1} \frac{\alpha_{k+i}}{g_{k+i}^t s_{k+i}} g_{k+i} g_{k+i}^t, \quad (6)$$

where the α_k are the parameters in (1) and also satisfy $\alpha_{k+p} = \alpha_k$.

To build an example of divergence, we express the formulae that define the method we are concerned with, the Armijo, Goldstein and Wolfe conditions and equation (4) as a system of equations and inequalities in D, Q, λ, $\overline{x}_k$, $\overline{f}_k$ $\overline{g}_k$, and B_k. We then solve this system of equations and inequalities and interpolate an appropriate objective function at the x_k. Due to the periodicity of α_k, $\overline{x}_k$, $\overline{f}_k$, $\overline{g}_k$ and $\overline{h}_k$ this system consists of a finite number of equations and inequalities. It is important to realize that we do not need to solve them in closed form. We can use interval arithmetic and the following version of Moore's Theorem [19] to prove that the equations can be solved and get accurate estimates of their solutions

Lemma 1 *Consider* $\overline{x} \in \mathbb{R}^n$, $r > 0$ *and* $D = \{x \in \mathbb{R}^n \text{ with } \|x - \overline{x}\|_1 < r\}$. *If* $f : D \mapsto \mathbb{R}^n$ *has continuous first derivatives and the* $n \times n$ *matrix* A *and* $a > 0$ *are such that*

$$\sup_{1 \le i \le n, x \in D} \left\| A^t \nabla f_i(x) - e_i \right\|_1 \|f(\overline{x})\|_\infty \le a < 1$$

and

$$b := \sup_{1\le i\le n} \left\|A^t e_i\right\|_1 \|f(\overline{x})\|_\infty < r(1-a) \tag{7}$$

then there exists $x^* \in D$ *with* $\|x^* - \overline{x}\|_\infty \le b/(1-a)$ *such that* $f(x^*) = 0$.

The proof of Lemma 1 starts at equation (66) in the appendix. Once we obtain estimates for the solution of the equations we can use interval arithmetic to verify the inequalities, as exemplified in the supplementary material. Note that all we need to use Lemma 1 is a good preconditioner A for the Jacobians of f in D. We do not need to estimate Lipschitz constants for these Jacobians as we would if we were to apply Kantorovich's Theorem [15].

The interpolation processes in [5] and [13] are quite different. In [13] we interpolate by extending cubic splines defined along the search lines to the whole space via Whitney's Extension Theorem, obtaining an objective function with Lipschitz continuous second derivatives. The article [5] uses polynomial interpolation. Since the requirement of Lipschitz continuity of the second derivative is weaker than polynomiality, our objective functions are more flexible and we can enforce the fundamental condition of bounded level sets for them. By choosing polynomial interpolation, [5] is constrained by the lack of flexibility of analytic functions and, as a consequence, its objective function does not have local minimizers.

The choice of $d_n = 1$ and appropriate D, Q, λ, $\overline{x}_k$, $\overline{f}_k$, $\overline{g}_k$ and $\overline{h}_k$ in (4) are basically all we need to build an example of divergence for Newton's method for minimization with any constant positive α_k's. Unfortunately, things are more complicated in the BFGS method, because we must also handle the matrices B_k. Therefore, the merits of the examples of divergence for the BFGS method in [5] and [13] go beyond their common use of formula (4) and the interpolation processes mentioned above. In [5] you will find the end result of skillful and hard work. In the next sections we present examples based on the framework developed in [13]. None of these examples is trivial. On the contrary, they are steps towards the noble goal of excellence.

3 The essence: geometry, algebra and analysis

This section outlines how we can build examples of divergence by combining Whitney's Extension Theorem with the algebra of matrices, if we are guided by the geometry of Figure 1. We summarize previous results so that the reader can have a self contained view of this process. The methodical construction of an example of divergence involves two tasks:

(a) Choosing convenient iterates x_k, function values f_k, gradients g_k and Hessians h_k compatible with the method we are concerned with.

(b) Finding an objective function compatible with the x_k, f_k, g_k and h_k above.

Whitney's Extension Theorem [6, 7, 20] is the key ingredient to impose conditions in the x_k, f_k, g_k and h_k in item (a) so that we can perform the interpolation step (b). It is our opinion that this deep theorem exposes as no other the relation between the nature of functions with Lipschitz continuous derivatives, the conditions by Armijo, Goldstein and Wolfe and the theoretical limitations of line searches in spaces of high dimension. We need the following definition to use Whitney's Extension Theorem for building examples of divergence:

Definition 1 *We define* $\mathrm{LC}^2(\mathbb{R}^n)$ *as the class of functions* $f : \mathbb{R}^n \mapsto \mathbb{R}$ *with Lipschitz continuous second derivatives for which there exists constants C and R, which depend on f, such that if* $\|x\| \geq R$ *then* $\nabla^2 f(x)$ *is positive definite and* $\left\| \nabla^2 f(x)^{-1} \right\| \leq C$.

We also consider the space $\mathbb{H}_n$ of $n \times n$ symmetric matrices. Using these concepts we can state the following corollary of Whitney's Extension Theorem [2]:

Theorem 2 *Let E be a bounded subset of* $\mathbb{R}^n$ *and consider functions* $f : E \mapsto \mathbb{R}$, $g : E \mapsto \mathbb{R}^n$ *and* $h : E \mapsto \mathbb{H}_n$. *If there exists a constant* $M \in \mathbb{R}$ *such that*

$$\|h(x) - h(y)\| \leq M \|x - y\|, \tag{8}$$

$$\|g(y) - g(x) - h(x)(y - x)\| \leq M \|x - y\|^2, \tag{9}$$

$$\left\| f(y) - f(x) - g(x)^t (y - x) - \frac{1}{2} (y - x)^t h(x) (y - x) \right\| \leq M \|x - y\|^3, \tag{10}$$

then there exists $F \in \mathrm{LC}^2(\mathbb{R}^n)$ *with* $F(x) = f(x)$, $\nabla F(x) = g(x)$ *and* $\nabla^2 F(x) = h(x)$ *for* $x \in E$.

We could state similar theorems for higher order derivatives, but the algebra needed to express and handle the consistency conditions analogous to (8) – (10), and the inequalities they lead to proves to be too complicated. We believe that the work required to build examples with higher order derivatives would not justify the insights they would bring. This is why the objective functions in our examples have only Lipschitz continuous second order derivatives.

Whitney's Extension Theorem exposes a fundamental difference between our objective functions and the polynomials (or analytic functions) in examples like [5]. Analytic functions do not have the extension property described in Theorem 2. They are rigid and cannot be modified locally. Analytic functions satisfy Łojasiewicz's inequality [10], as do the more general families of functions described in [9]. On the other hand, functions in $\mathrm{LC}^2(\mathbb{R}^n)$ have bounded level sets and are easy to work with. By targeting objective functions in this class we do not need to worry about large x when building examples, because with a little work we can modify a function with Lipschitz continuous second derivatives to turn it into an element of $\mathrm{LC}^2(\mathbb{R}^n)$. Doing the same for an analytic function would be a very delicate process, if feasible at all. This is a fundamental reason why we prefer functions in $\mathrm{LC}^2(\mathbb{R}^n)$ instead of polynomials or analytic functions. Such choice is also justified as it is common to find, in textbooks and research papers, theorems in which the hypothesis asks for a function with Lipschitz continuous second derivatives and bounded level sets.

We now present Theorem 3. It is a powerful tool for constructing examples of divergence for line search methods when combined with matrices $D(\lambda)$ and Q of the form

$$D(\lambda) = \begin{pmatrix} I_a & 0 & 0 \\ 0 & \lambda I_b & 0 \\ 0 & 0 & \lambda^{d_n} I_c \end{pmatrix} \qquad \text{and} \qquad Q = \begin{pmatrix} Q_a & 0 & 0 \\ 0 & Q_b & 0 \\ 0 & 0 & Q_c \end{pmatrix}, \tag{11}$$

where I_i stands for the $i \times i$ identity matrix, a, b and c are positive integers and Q_i represents a $i \times i$ orthogonal matrix such that $Q_i^p = I_i$ for some common period $p \in \mathbb{N}$. There are good reasons for considering matrices Q and D with three blocks instead of

[2] Theorem 2 is Lemma 6 in page 150 of [13].

the two blocks used in some examples in [4, 5, 12, 13], and also for considering $a \geq 3$, $b \geq 2$ and $d_n > 2$ in (11). These conditions help to ensure the correct behavior of the limit search lines

$$\mathcal{L}_k = \left\{ D(0)\, Q^k \left(\bar{x}_k + \alpha \bar{s}_k \right), \ \ \alpha \in \mathbb{R} \right\}, \tag{12}$$

where

$$\bar{s}_k = QD(\lambda)\, \bar{x}_{k+1} - \bar{x}_k. \tag{13}$$

With $a \geq 3$ it is quite unlikely that non consecutive limit search lines will cross. The choice $b \geq 2$ helps to control the rate at which the search lines approach the limit hyperplane. Finally, the choice $d_n > 2$ takes care of technical issues regarding the differentiability of the resulting objective function so that we can satisfy the hypothesis of Theorem 2. We can then state the theorem underlying the results in this article [3]:

Theorem 3 *Consider $\lambda \in (0,1)$, the matrices Q and $D(\lambda)$ in (11) and sequences $\overline{g}_k, \bar{x}_k \in \mathbb{R}^n$, $\overline{f}_k \in \mathbb{R}$, $\overline{h}_k \in \mathbb{H}_n$. Suppose that for a period $p \in \mathbb{N}$ we have $Q^p = I_n$, $\overline{g}_{k+p} = \overline{g}_k$, $\bar{x}_{k+p} = \bar{x}_k$, $\overline{f}_{k+p} = \overline{f}_k$ and $\overline{h}_{k+p} = \overline{h}_k$. If $d_n > 2$ and, for all k,*

$$D(0)\, \bar{s}_k \ \text{ and } \ D(0)\, Q\bar{s}_{k+1} \ \text{ are linearly independent,} \tag{14}$$

$$\text{if } j - k \bmod p \notin \{-1,0,1\} \ \text{ then } \ \mathcal{L}_j \cap \mathcal{L}_k = \emptyset, \tag{15}$$

$$D'(0)\, \bar{x}_k \text{ and } D'(0)\, \bar{s}_k \ \text{ are linearly independent,} \tag{16}$$

$$\left(\overline{h}_k \right)_{ij} = 0 \ \text{ for } d_i + d_j > d_n, \tag{17}$$

then there exists $k_0 \in \mathbb{N}$ and $f \in \mathrm{LC}^2(\mathbb{R}^n)$ such that for $k \geq k_0$ and f_k, x_k, g_k and h_k in (4) and (5) we have $f(x_k) = f_k$, $\nabla g(x_k) = g_k$, $\nabla^2 f(x_k) = h_k$. If we also assume that

$$\bar{s}_k^t \overline{g}_k < \lambda^{d_n} \overline{f}_{k+1} - \overline{f}_k < \lambda^{d_n} \bar{s}_k^t D(\lambda)^{-1} Q \overline{g}_{k+1}, \tag{18}$$

$$\bar{s}_k^t \overline{h}_k \bar{s}_k > 0 \ \text{ and } \ \bar{s}_k^t Q D(\lambda)^{-1} \overline{h}_{k+1} Q^t D(\lambda)^{-1} \bar{s}_k > 0, \tag{19}$$

then f can be chosen to be strictly convex along the search lines, i.e., $s_k^t \nabla^2 f(x_k + \alpha s_k) s_k > 0$ for all $\alpha \in \mathbb{R}$.

In words, Theorem 3 says that if $d_n > 2$ and the technical conditions (14)–(19) are satisfied then there exists a function f with bounded level sets and Lipschitz continuous second derivatives which interpolates the f_k, x_k, g_k and h_k in (4)–(5) for k large enough and is strictly convex along the search lines. Theorem 3 simplifies the process of building examples of divergence, because it spares us from the construction of an explicit objective function. It allows us to concentrate in finding x_k, f_k, g_k and h_k compatible with our method. Once we find them we only need to make sure they satisfy the technical conditions (14)–(19). As a consequence, we can explore the iterates in higher dimensions and observe phenomena which do not occur in lower dimensions. We can go beyond the two dimensions in which Powell proved the convergence of the BFGS method and analyze its behavior in situations in which our fallible lower dimensional intuition may mislead us. In this exploration we can, and should, take advantage of the modern symbolic and numerical tools at our disposal, as exemplified in the supplementary material. The perception that we can use Theorem 3 with help of a software allows us to focus on the creative part in the construction of divergence examples: the choice of appropriate forms for λ, $\bar{x}_k$, $\overline{f}_k$, $\overline{g}_k$ and $\overline{h}_k$ in (4) and (5) and Q in (11). This choice is the result of our understanding of the method we are considering and the need

[3] This theorem is proved in the last paragraph of the appendix.

to balance the freedom secured by moving to more dimensions with the complexity of the resulting algebraic problem.

The next sections apply Theorem 3 to build examples for Gauss Newton and BFGS. In order to do that we must look at the particular details that define these methods. We would need to do the same to every method for which we would like to apply Theorem 3 to build an example of divergence. The next three subsections describe technical facts that hold for several methods. They show that enforcing conditions like Armijo's, Goldstein and Wolfe and convexity along the search lines is relatively easy once we get iterates, function values and gradients that satisfy (4) and (5). The reader may prefer to skip these technical details and proceed to the next sections.

3.1 Defining the normalized iterates $\overline{x}_k$ in terms of the normalized steps $\overline{s}_k$

Usually it is more convenient to work with the steps $s_k = x_{k+1} - x_k$, and their normalized version $\overline{s}_k = Q^k D(\lambda)^k \overline{x}_{k+1} - \overline{x}_k$, instead of the iterates x_k and their normalization $\overline{x}_k$. Equation (13) shows how to obtain the $\overline{s}_k$'s from the $\overline{x}_k$'s. However, we must be cautious when expressing the $\overline{x}_k$'s in terms of the $\overline{s}_k$'s, since equation (13) gives rise to a singularity. In fact, by multiplying (13) by $Q^j D(\lambda)^j$ and recalling that $Q^p = I$ and $\overline{x}_{k+p} = \overline{x}_k$ we obtain

$$\sum_{j=0}^{p-1} Q^j D(\lambda)^j \overline{s}_{k+j} = (D(\lambda)^p - I)\overline{x}_k. \tag{20}$$

If we decompose D and Q as in (11), with the corresponding decomposition

$$\overline{x}_k = \left(\overline{x}_{a,k}, \overline{x}_{b,k}, \overline{x}_{c,k}\right)^t \qquad \text{and} \qquad \overline{s}_k = \left(\overline{s}_{a,k}, \overline{s}_{b,k}, \overline{s}_{c,k}\right)^t, \tag{21}$$

then (20) leads to these equations:

$$\sum_{j=0}^{p-1} Q_a^j \overline{s}_{a,k+j} = 0, \tag{22}$$

$$\overline{x}_{b,k} = \frac{1}{\lambda^p - 1} \sum_{j=0}^{p-1} \lambda^j Q_b^j \overline{s}_{b,k+j} \qquad \text{and} \qquad \overline{x}_{c,k} = \frac{1}{\lambda^{pd_n} - 1} \sum_{j=0}^{p-1} \lambda^{jd_n} Q_c^j \overline{s}_{c,k+j}. \tag{23}$$

We cannot derive $\overline{x}_{a,k}$ from (22), because the examples are invariant under translations in the $\overline{x}_{a,k}$'s. However, if (22) holds for $k = 0$ then it holds for all k. Once we enforce (22) for $k = 0$ we can define $\overline{x}_{a,k}$ by

$$\overline{x}_{a,0} = 0 \qquad \text{and} \qquad k > 0 \ \Rightarrow\ \overline{x}_{a,k} = Q_a^{-k} \sum_{j=0}^{k-1} Q_a^j \overline{s}_{a,j}. \tag{24}$$

Using induction we can then derive the $\overline{x}_k$ from the $\overline{s}_k$ and prove the following lemma:

Lemma 2 *If, for $k \in \mathbb{N}$, the $\overline{x}_k, \overline{s}_k \in \mathbb{R}^n$ are decomposed as in (21) and satisfy (22) then the $\overline{x}_k$ in (23) and (24) are compatible with $\overline{s}_k$ defined in (13).*

Using (24) we can write the projection $\mathcal{P}_k$ of the limit search line $\mathcal{L}_k$ in (12) in the subspace corresponding to $\overline{x}_{a,k}$ as

$$\mathcal{P}_k := \left\{ \sum_{j=0}^{k-1} Q_a^j \overline{s}_{a,j} + \alpha Q_a^k \overline{s}_{a,k}, \ \ \alpha \in \mathbb{R} \right\},$$

under the usual convention that $\sum_{j=0}^{-1} Q_a^j \overline{s}_{a,k} = 0$. To verify the hypothesis (15) in Theorem 3 it suffices to show that $\mathcal{P}_k \cap \mathcal{P}_{k+m} = \emptyset$ for $0 \leq k < p-1$ and $1 < m < p-1$. This is equivalent to saying that there exists no $\alpha, \beta \in \mathbb{R}$ such that

$$\sum_{j=0}^{k-1} Q_a^j \overline{s}_{a,j} + \alpha Q_a^k \overline{s}_{a,k} = \sum_{j=0}^{k+m-1} Q_a^j \overline{s}_{a,j} + \beta Q_a^{k+m} \overline{s}_{a,k+m}$$

or, equivalently, that there exists no $\alpha, \beta \in \mathbb{R}$ such that

$$\alpha \overline{s}_{a,k} = \sum_{j=0}^{m-1} Q_a^j \overline{s}_{a,k+j} + \beta Q_a^m \overline{s}_{a,k+m}. \tag{25}$$

In resume, we have proved the following Lemma

Lemma 3 *If the normalized steps $\overline{s}_k$ are such that for every $0 \leq k < p$ and $1 < m < p-1$ there exists no α and β satisfying equation (25) then the corresponding normalized iterates $\overline{x}_k$ satisfy the hypothesis (15) of Theorem 3.*

If $a \geq 3$ the hypothesis of Lemma 3 will be satisfied unless the vectors in equation (25) align in some unexpected way. As a consequence, we do not need to worry about this condition as we explore the parameters that define our example at first. We only need to check (25) after we find them. If, by any chance, the hypothesis of Lemma 3 is not satisfied in the first try, then we should adjust the parameters slightly, so that this hypothesis holds. If we cannot find a suitable modification then maybe it would be advisable to consider whether the method we are considering converges.

3.2 Convexity along the search lines

The conditions (18)–(19) enforce convexity along the search lines. In 2007 we brought up this condition in the abstract of [13] in order to make sure that our examples would choose the only local minimizer along the search line. Under this condition, the simple algebraic condition $s_k^t g_{k+1} = 0$ guarantees that the iterates in our examples would be generated by methods that choose a global minimizer along the search line as well as methods that choose the first local minimizer. In the examples for Gauss Newton and BFGS in the next sections, and for other methods that do not use the Hessian of the objective function, it is easy to enforce (17) and (19) by decomposing $\overline{h}_k$ in (11) in the block diagonal matrix

$$\overline{h}_k = \begin{pmatrix} I_a & 0 & 0 \\ 0 & I_b & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{26}$$

because in this case (19) holds as long as $\overline{s}_{a,k}$ or $\overline{s}_{b,k}$ are not zero. Since (16) implies that $\overline{s}_{b,k} \neq 0$, by assuming (26) we do not need to worry about (19).

3.3 The conditions by Goldstein, Armijo and Wolfe

The Goldstein condition [14] requires that there exists $c \in (0, 1/2)$ such that

$$(1-c)\, s_k^t g_k \leq f_{k+1} - f_k \leq c s_k^t g_k. \tag{27}$$

Using (4) we can reduce it to

$$(1-c)\overline{s}_k^t \overline{g}_k \leq \lambda^{d_n} \overline{f}_{k+1} - \overline{f}_k \leq c \overline{s}_k^t \overline{g}_k. \tag{28}$$

We can enforce the second inequality in (28) by taking $\overline{f}_k = 1$ for all k and choosing a tiny $c > 0$. In some cases we can enforce the first inequality in (28) by scaling the $\overline{g}_k$ we already have by μ in the range

$$\frac{\lambda^{d_n}\overline{f}_{k+1} - \overline{f}_k}{(1-c)\max_{1 \leq k \leq p} \overline{s}_k^t \overline{g}_k} \leq \mu \leq \frac{\lambda^{d_n}\overline{f}_{k+1} - \overline{f}_k}{c \min_{1 \leq k \leq p} \overline{s}_k^t \overline{g}_k}, \tag{29}$$

which is not empty as long as

$$c \leq \frac{\max_{1 \leq k \leq p} \overline{s}_k^t \overline{g}_k}{\max_{1 \leq k \leq p} \overline{s}_k^t \overline{g}_k + \min_{1 \leq k \leq p} \overline{s}_k^t \overline{g}_k}.$$

By imposing the first inequality in the Goldstein condition (28) we also enforce the first condition in (18). The second condition in (18) follows from the exact line search condition ($s_k^t g_{k+1} = 0$) when $\overline{f}_{k+1} = \overline{f}_k$, and (19) follows from the choice of $\overline{h}_k$ in (26). Therefore, imposing convexity along the search line is not more demanding than enforcing the Goldstein condition in examples based on Theorem 3. On the other hand, examples using analytic function must perform extra work to enforce convexity along the search lines, as can be noticed by considering the difference in complexity among the examples with and without this condition present in [5].

The first Wolfe condition (3) is sometimes called Armijo's condition, because Armijo proposed a line search in which we reduce the step $s_k = x_{k+1} - x_k$ until (3) is satisfied. In our examples the first try for a step already satisfies the first Wolfe condition. Therefore, the Armijo condition and the first Wolfe condition are equivalent as far as we we are concerned. Using (4) we can demonstrate that the first Wolfe condition is satisfied for all positive σ smaller than

$$\sigma_0 = \min_{0 \leq k < p} \left\{ \frac{\lambda^{d_n}\overline{f}_{k+1} - \overline{f}_k}{\overline{s}_k^t \overline{g}_k} \right\}.$$

Our examples always have $\overline{s}_k^t \overline{g}_k < 0$. Therefore, in order for them to satisfy the first Wolfe condition and the Armijo condition it suffices that $\lambda^{d_n}\overline{f}_{k+1} - \overline{f}_k < 0$ for all k. This can be enforced by taking $\overline{f}_k = 1$ for all k.

The second Wolfe condition requires that for some $\beta \in (\sigma, 1)$, where σ is the parameter in the first Wolfe condition, we have

$$\nabla f(x_{k+1})^t s_k \geq \beta \nabla f(x_k)^t s_k. \tag{30}$$

In our examples equation (30) follows from the descent condition ($s_k^t g_k < 0$) and from the use of exact line searches. Noticing that (4) yields $s_k^t g_{k+1} = \overline{s}_k^t Q D(\lambda)^{-1} \overline{g}_{k+1}$, we enforce exact line searches by requiring that

$$\overline{s}_k^t Q D(\lambda)^{-1} \overline{g}_{k+1} = 0. \tag{31}$$

4 An example of divergence for the Gauss Newton method

The purpose of the Gauss Newton method is to minimize $f : \mathbb{R}^n \mapsto \mathbb{R}$ given by

$$f(x) = \frac{1}{2} \sum_{j=1}^{m} r_j(x)^2,$$

where each r_j is a function from $\mathbb{R}^n$ to $\mathbb{R}$ and $m \geq n$. The method is defined in terms of the Jacobian matrix J_r of the function $r : \mathbb{R}^n \mapsto \mathbb{R}^m$ given by $r(x) = (r_1(x), \ldots, r_m(x))^t$, whose transpose has the gradients of the r_j's as columns:

$$J_r(x)^t = (\nabla r_1(x) \mid \nabla r_2(x) \mid \cdots \mid \nabla r_m(x)).$$

Defining $G_k := J_r(x_k)^t$, and assuming that $G_k G_k^t$ is not singular, the iterates are defined by

$$x_{k+1} := x_k - \alpha_k \left(G_k G_k^t\right)^{-1} \nabla f(x_k) = x_k - \alpha_k \left(G_k G_k^t\right)^{-1} G_k r(x_k), \tag{32}$$

for appropriate $\alpha_k > 0$. By taking $M_k = G_k$ we see that this method is in the format used in Theorem 1. If f has bounded level sets and the first Wolfe condition is satisfied then the matrices G_k are bounded and Theorem 1 shows that if the parameters α_k do not get too close to zero then $\lim_{k \to \infty} \nabla f(x_k) = 0$ for every starting point x_0. Therefore, to construct an example of divergence for Gauss Newton we must allow arbitrarily small α_k. Once we accept this fact we can use Theorem 3 to build an example of divergence for Gauss Newton.

We take $m = n = 7$ and use Theorem 3 to obtain auxiliary functions $\phi_1, \phi_2, \ldots, \phi_7$ with which we define

$$r_j(x) := \sqrt{\kappa + \phi_j(x)}. \tag{33}$$

The parameter κ ensures that $\kappa + \phi_j(x) \geq 1$ for all x and j. The ϕ_j provided by Theorem 3 are bounded below and we take

$$\kappa := 1 - \min_{1 \leq j \leq 7} \inf_{x \in \mathbb{R}^7} \phi_j(x).$$

As a result we obtain an objective function

$$f(x) = \frac{1}{2} \left(7\kappa + \sum_{j=1}^{7} \phi_j(x) \right). \tag{34}$$

The ϕ_j are built from the same set of iterates x_k and use the same search lines as f. The ϕ_j are convex along the search lines. Therefore, f is also convex along these lines. Moreover, f belongs to $\mathrm{LC}^2\left(\mathbb{R}^7\right)$ because the sum of functions in $\mathrm{LC}^2\left(\mathbb{R}^7\right)$ is also in $\mathrm{LC}^2\left(\mathbb{R}^7\right)$. This implies that f has bounded level sets.

We close this section explaining how to use Theorem 3 to find convenient iterates x_k, auxiliary functions ϕ_j and parameters α_k so we can prove that the line search find the only local minimizer along the search line and satisfy the conditions by Armijo, Goldstein and Wolfe. We divide the work in four subsections. The first one defines the matrix Q in (11), the contraction parameter λ, the exponent d_n and the iterates x_k. The next one explains how to use Theorem 3 to obtain the functions ϕ_j. The third subsection defines the parameters α_k so that the iterates x_k are compatible with Gauss Newton. The last subsection shows that the emerging objective function and iterates satisfy the requirements stated in the abstract.

Finally, we emphasize that although our α_k converge to zero, the convexity of the objective function along the search lines and the algebraic condition $s_k^t g_{k+1} = 0$ enforced below imply that these would be the α_k chosen automatically by an algorithm that follows Powell's suggestion of choosing the first minimizer along the search line or if we asked for the global minimizer along the search line.

4.1 Defining λ, the matrix Q and the iterates

We take a constant normalized step

$$\bar{s}_k := \mathbb{1}_7 := (1,1,1,1,1,1,1)^t .$$

The matrices Q and $D(\lambda)$ in (4) are chosen as in (11), with $a=4$, $b=2$, $c=1$ and

$$Q_a := \begin{pmatrix} R_{\pi/3} & 0 \\ 0 & R_{\pi/6} \end{pmatrix}, \qquad Q_b := R_{\pi/2} \qquad \text{and} \qquad Q_c := (-1), \qquad (35)$$

where

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

is the counterclockwise rotation by θ. The parameter λ and the period p are defined as

$$d_n := 3, \qquad p := 12 \qquad \text{and} \qquad \lambda := \sqrt[3]{\frac{1}{1+\sqrt{3}}}.$$

As the reader can verify, $\sum_{j=0}^{11} Q_a^j = 0$. Therefore, Lemma 2 yields these normalized iterates

$$\bar{x}_k = \begin{pmatrix} \bar{x}_{a,k} \\ \bar{x}_{b,k} \\ \bar{x}_{c,k} \end{pmatrix} = \begin{pmatrix} Q_a^{-k} \sum_{j=0}^{k-1} Q_a^j \mathbb{1}_4 \\ \frac{1}{\lambda^{12}-1} \sum_{j=0}^{11} \lambda^j R_{\pi/2}^j \mathbb{1}_2 \\ \frac{1}{\lambda^{36}-1} \sum_{j=0}^{11} \left(-\lambda^3\right)^j \end{pmatrix} = \begin{pmatrix} Q_a^{-k} \sum_{j=0}^{k-1} Q_a^j \mathbb{1}_4 \\ \left(\lambda R_{\pi/2} - I\right)^{-1} \mathbb{1}_2 \\ \frac{-1}{1+\lambda^3} \end{pmatrix}. \qquad (36)$$

The D, Q, $\bar{s}_k = \mathbb{1}_7$ and the $\bar{x}_k$ in the previous equation satisfy the hypothesis (14) and (16) of Theorem 3. We end this subsection using Lemma 3 to verify the hypothesis (15) in Theorem 3. In the present case, equation (25) reduces to

$$\alpha \mathbb{1}_4 = \sum_{j=0}^{m-1} Q_a^j \mathbb{1}_4 + \beta Q_a^m \mathbb{1}_4 = (I - Q_a)^{-1} (I - Q_a^m) \mathbb{1}_4 + \beta Q_a^m \mathbb{1}_4. \qquad (37)$$

To verify (15) it would be enough to show that if $0 < m < 12$ and there are $\alpha, \beta \in \mathbb{R}$ which satisfy (37) then either $m = 1$ or $m = 11$. Let us then prove that these are, indeed, the only two possibilities. Equations (15), (35) and (37) imply that $A\mathbb{1}_2 = B\mathbb{1}_2 = 0$, where

$$A = \alpha \left(I_2 - R_\rho\right) - \left(I_2 - R_\rho^m\right) - \beta \left(I_2 - R_\rho\right) R_\rho^m, \qquad (38)$$

$$B = \alpha \left(I_2 - R_\rho^2\right) - \left(I_2 - R_\rho^{2m}\right) - \beta \left(I_2 - R_\rho^2\right) R_\rho^{2m}, \qquad (39)$$

for $\rho = \pi/6$. The matrices A and B are the sum of a multiple of the identity and an anti-symmetric matrix. As a consequence, the equalities $A\mathbb{1}_2 = B\mathbb{1}_2 = 0$ imply that we actually have $A = B = 0$. Multiplying (38) by $I_2 + R_\rho$ we obtain

$$\alpha \left(I_2 - R_\rho^2\right) - \left(I_2 - R_\rho^m\right) \left(I_2 + R_\rho\right) - \beta \left(I_2 - R_\rho^2\right) R_\rho^m = 0. \qquad (40)$$

Subtracting this from (39) we get

$$-\left(I_2 - R_\rho^m\right) \left(R_\rho^m - R_\rho\right) - \beta \left(I_2 - R_\rho^2\right) R_\rho^m \left(R_\rho^m - I_2\right) = 0.$$

Since $0 < m < 12$ the matrix $R_\rho^m - I_2$ is non singular. Thus, $\beta \left(I_2 - R_\rho^2 \right) R_\rho^m = R_\rho - R_\rho^m$ and

$$\beta \left(I_2 - R_\rho^2 \right) = R_\rho^{1-m} - I_2. \tag{41}$$

Since $\rho = \pi/6$, we have that $\cos 2\rho = 1/2$ and $\sin 2\rho = \sqrt{3}/2$. Equating the entries of the matrix in (41) to zero we get

$$\frac{\sqrt{3}}{2}\beta = \sin\left(\frac{(m-1)\,\pi}{6}\right) \qquad \text{and} \qquad \frac{1}{2}\beta = \cos\left(\frac{(m-1)\,\pi}{6}\right) - 1. \tag{42}$$

It follows that

$$1 = \sin\left(\frac{(m-1)\,\pi}{6}\right)^2 + \cos\left(\frac{(m-1)\,\pi}{6}\right)^2 = \frac{3}{4}\beta^2 + \left(\frac{1}{2}\beta + 1\right)^2 = \beta^2 + \beta + 1.$$

Thus, either $\beta = 0$ or $\beta = -1$. Since $0 < m < 12$, in the case $\beta = 0$ equation (42) implies that $m = 1$. Similarly, if $\beta = -1$ then equation (42) implies that $m = 11$ and we are done.

4.2 Defining the auxiliary functions ϕ_j

This subsection explains how to use Theorem 3 to obtain the functions $\phi_j \in \mathrm{LC}^2\left(\mathbb{R}^7\right)$. The gradients of these function will be expressed in terms of the vectors $e_j \in \mathbb{R}^7$, which satisfy

$$(e_j)_j = 1 \qquad \text{and} \qquad (e_j)_i = 0 \ \text{ for } \ i \neq j.$$

We build ϕ_j such that

$$\phi_j(x_k) = \lambda^{3k} \qquad \text{and} \qquad \nabla\phi_j(x_k) = -\lambda^{3k} Q^k D(\lambda)^{-k} e_j, \tag{43}$$

by applying Theorem 3 with the $\overline{x}_k$, λ and d_n above and

$$\overline{f}_k = 1 \qquad \text{and} \qquad \overline{g}_k = -e_j.$$

The $\overline{x}_k$ in the previous subsection satisfy the hypothesis (14)–(16) of Theorem 3. As explained in subsection 3.2, we can satisfy the hypothesis (17) and (19) by taking $\overline{h}_k$ as in (26). The only hypothesis left in order to apply Theorem 3 to obtain ϕ_j is equation (18). In the present case it reduces to

$$-1 < \lambda^3 - 1 < -\lambda^3 \mathbb{1}_7^t Q^t D(\lambda)^{-1} e_j. \tag{44}$$

We now verify the second inequality in the line above in the three possible cases:

(a) If $1 \leq j \leq 4$ then the second inequality in (44) follows from the observation that $\mathbb{1}_4^t Q_a = \left(1+\sqrt{3}, 1-\sqrt{3}, \sqrt{3}+1, \sqrt{3}-1\right)^t/2$ and

$$\lambda^3 - 1 < -0.6 < \frac{1}{2} = -\frac{\lambda^3}{2} \max\left\{1+\sqrt{3}, 1-\sqrt{3}, 1+\sqrt{3}, \sqrt{3}-1\right\}.$$

(b) If $5 \leq j \leq 6$ then the second inequality in (44) holds because $\lambda^3 - 1 < -0.6 < -\lambda^2$.

(c) If $j = 7$ then second inequality in (44) holds because its right hand side equals one.

Therefore, Theorem 3 shows that there exist functions $\phi_j \in \mathrm{LC}^2\left(\mathbb{R}^2\right)$ which are convex along the lines $x_k + \alpha s_k$ and satisfy (43).

4.3 Defining the α_k for equation (1)

Equations (34) and (43) show that the gradient of f at x_k is

$$g_k := \nabla f(x_k) = \frac{1}{2}\sum_{j=1}^{7} \nabla\phi_j(x) = -\frac{1}{2}\lambda^{3k} Q^k D(\lambda)^{-k}\, \mathbb{1}_7. \tag{45}$$

Equation (33) implies that

$$\nabla r_j(x) = \frac{1}{2r_j(x)} \nabla\phi_j(x)$$

and the matrix $G_k = J_r(x_k)^t$ satisfies

$$G_k = -\frac{\lambda^{3k}}{2\sqrt{\kappa+\lambda^{3k}}} Q^k D(\lambda)^{-k}.$$

Since Q and $D(\lambda)$ commute, $\left(G_k G_k^t\right)^{-1} = 4\lambda^{-6k}\left(\kappa+\lambda^{3k}\right) D(\lambda)^{2k}$. Therefore, the search direction d_k satisfies

$$d_k = -\left(G_k G_k^t\right)^{-1} g_k = 2\lambda^{-3k}\left(\kappa+\lambda^{3k}\right) Q^k D(\lambda)^k\, \mathbb{1}_7.$$

Equations (4), (13) and our choice $\bar{s}_k = \mathbb{1}_7$ yield $s_k = Q^k D(\lambda)^k\, \mathbb{1}_7$. Therefore, if we take

$$\alpha_k := \frac{\lambda^{3k}}{2\left(\kappa+\lambda^{3k}\right)}$$

then $s_k = \alpha_k d_k$. Therefore, our iterates are compatible with Gauss Newton with the α_k above.

4.4 Verifying the line search conditions

Equations (4) and (13) and our choice $\bar{s}_k = \mathbb{1}_7$ yield $s_k = Q^k D(\lambda)^k\, \mathbb{1}_7$. Thus, (45) yields

$$s_k^t g_k = -7\lambda^{3k}/2.$$

Equation (34) and our choice $\phi_j(x_k) = \lambda^{3k}$ lead to $f_k := f(x_k) = 7\left(\kappa+\lambda^{3k}\right)/2$. Thus,

$$f_{k+1} - f_k = \frac{7}{2}\lambda^{3k}\left(\lambda^3 - 1\right) = \left(1-\lambda^3\right) s_k^t g_k.$$

Therefore the first Wolfe condition (3) is satisfied for $\sigma = 1-\lambda^3 \approx 0.6$. The Goldstein condition is satisfied for $c = \lambda^3 \approx 0.4$. Finally, the exact line search condition $s_k^t g_{k+1}$ is satisfied because $\lambda^3 = 1/(1+\sqrt{3})$ and according to (45)

$$-2s_k^t g_{k+1} = \lambda^3 \mathbb{1}_7^t Q^t D(\lambda)^{-1}\, \mathbb{1}_7 = \lambda^3 \mathbb{1}_4^t Q_a \mathbb{1}_4 + \lambda^2 \mathbb{1}_2^t R_{\pi/2} \mathbb{1}_2 - 1 = \lambda^3\left(1+\sqrt{3}\right) - 1 = 0.$$

This completes the construction of the example of divergence for Gauss Newton.

5 The new example of divergence for the BFGS method

This section presents a new example of divergence for the BFGS method. The example shows that this method may fail even under all the conditions described in the table in section 2. In particular, our examples have bounded level sets, a property whose far reaching consequences are illustrated in Theorem 1. Time will tell whether it is possible to build an example similar to ours in which the objective function is a polynomial.

We analyze the BFGS method with exact line searches. In this case $s_k^t g_{k+1} = 0$ and the BFGS iterates are given by

$$s_k = x_{k+1} - x_k = -\alpha_k B_k^{-1} g_k, \tag{46}$$

where $g_k = \nabla f(x_k)$. The positive definite matrices B_k evolve according to

$$B_{k+1} := B_k + \frac{\alpha_k}{s_k^t g_k} g_k g_k^t - \frac{1}{s_k^t g_k} (g_{k+1} - g_k)(g_{k+1} - g_k)^t. \tag{47}$$

It is convenient to work with B_k of the form (6), with the additional requirements that

$$s_k^t g_k < 0 \qquad \text{and} \qquad s_k^t g_{k+j} = 0 \qquad \text{for} \qquad 1 \leq j < n. \tag{48}$$

Equations (6) and (48) show that $B_k s_k = -\alpha_k g_k$. Therefore, we do not need to worry about (46). However we must make sure that the B_k in (6) satisfies (47). We can achieve this goal by imposing yet another set of conditions:

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

for $\rho_k \in \mathbb{R}$ such that

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

Assuming (49) and (50), we can use induction to verify that the matrices (6) satisfy (47). In fact, if B_k is given by (6) then, using (49) and (50) we obtain

$$B_{k+1} = -\sum_{i=0}^{n-1} \frac{\alpha_{k+i}}{g_{k+i}^t s_{k+i}} g_{k+i} g_{k+i}^t + \frac{\alpha_k}{s_k^t g_k} g_k g_k^t - \frac{1}{s_k^t g_k} (g_{k+1} - g_k)(g_{k+1} - g_k)^t$$

$$= -\sum_{i=1}^{n-1} \frac{\alpha_{k+i}}{g_{k+i}^t s_{k+i}} g_{k+i} g_{k+i}^t - \frac{1}{\rho_k^2 s_k^t g_k} g_{k+n} g_{k+n}^t = -\sum_{i=1}^{n} \frac{\alpha_{k+i}}{g_{k+i}^t s_{k+i}} g_{k+i} g_{k+i}^t$$

and B_{k+1} also satisfies (48).

To build an example of divergence for the BFGS method with $\alpha_k = 1$ for all k we only need to find $\bar{s}_k$, $\overline{f}_k$, $\overline{g}_k$ and ρ_k which are compatible with the equations (48)–(50) and the conditions which allow us to use Theorem 3. After the experience gained in [12, 13] we found a better way to parameterize D, Q, $\bar{x}_k$, $\overline{f}_k$, $\overline{g}_k$, $\overline{h}_k$ in (4)–(5). Due to a few subtle algebraic points that we have noticed recently, we now believe it is best to take

$$n = 9, \qquad d_n = 4, \qquad p = 16 \times 36 = 576, \tag{51}$$

$$D(\lambda) = \begin{pmatrix} I_3 & 0 & 0 \\ 0 & \lambda I_2 & 0 \\ 0 & 0 & \lambda^4 I_4 \end{pmatrix} \qquad \text{and} \qquad Q = \begin{pmatrix} I_3 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_4 \end{pmatrix}, \tag{52}$$

where I_n is the $n \times n$ identity matrix. As we explain below, the choice of iterates in $\mathbb{R}^9$ and the large period 576 simplifies the algebra. We define the first 36 normalized iterates, which are then replicated 16 times by symmetry.

Using equation (4), parameterized as in (51) and (52), we reduce the conditions (48)–(50) that ensure the compatibility of our x_k and B_k with the BFGS method with exact line searches and unity steps to these equations:

$$\overline{s}_k^t \overline{g}_k < 0, \tag{53}$$

$$\rho_k^2 \overline{s}_k^t \overline{g}_k = \lambda^{36} \overline{s}_{k+9}^t \overline{g}_{k+9}, \tag{54}$$

$$\overline{s}_k^t Z^j \overline{g}_{k+j} = 0 \text{ for } j = 1, 2, \ldots, 8, \tag{55}$$

$$Z^9 \overline{g}_{k+9} = \rho_k \left(Z \overline{g}_{k+1} - \overline{g}_k \right) \tag{56}$$

for some $\{\rho_k, k \in \mathbb{N}\}$ with $\rho_{k+p} = \rho_p$ and

$$Z = Z(\lambda) = \lambda^4 D(\lambda)^{-1} Q = \begin{pmatrix} \lambda^4 I_3 & 0 & 0 \\ 0 & \lambda^3 I_2 & 0 \\ 0 & 0 & I_4 \end{pmatrix}. \tag{57}$$

A simple inspection of equations (53)–(56) proves the following lemma:

Lemma 4 *Equations (53)–(56) are invariant with respect to scaling in $\{\overline{g}_k, k \in \mathbb{N}\}$, in the sense that if they are part of a solution of these equations and $\mu > 0$ then $\{\mu \overline{g}_k, k \in \mathbb{N}\}$ combined with the same values for the other parameters also satisfy the same equations.*

The existence of an example of divergence for the BFGS method as claimed in the abstract is a consequence of the following lemma:

Lemma 5 *There are numbers $\{\rho_k, k \in \mathbb{N}\}$, vectors $\{\overline{g}_k, k \in \mathbb{N}\} \subset \mathbb{R}^9$ and $\{\overline{x}_k, k \in \mathbb{N}\} \subset \mathbb{R}^9$ such that, for n, p, d_n, D and Q in (52), $\overline{f}_k = 1$, $\overline{h}_k$ in (26), $k \in \mathbb{N}$ and*

$$\lambda := \sqrt[72]{\frac{1}{1 + \sqrt{2 + \sqrt{2}}}} \tag{58}$$

we have $\rho_{k+p} = \rho_k$, $\overline{g}_{k+p} = \overline{g}_k$ and $\overline{x}_{k+p} = \overline{x}_k$ and the vectors $\overline{s}_k$ in (13) satisfy all the conditions (14)–(19) and (53)–(56).

We end this section proving Lemma 5. This demonstration involves the verification of algebraic identities involving matrices. The supplementary material verifies them using the software Mathematica. The idea of the proof is to write the $\overline{g}_k$ and $\overline{s}_k$ as

$$\overline{g}_k := Z^{-k} \Gamma_k e_1^9 \qquad \text{and} \qquad \overline{s}_k := -\sigma_k \left(\Gamma_k^{-1} Z^k \right)^t e_1^9, \tag{59}$$

where Z is defined in (57), the Γ_k are convenient 9×9 matrices. The vector $e_i^n \in \mathbb{R}^n$ has ith entry equal to one and the others equal to 0 and the $\sigma_k \in \mathbb{R}^9$ are appropriate positive numbers. We use Lemma 2 to obtain normalized iterates $\overline{x}_k$ using (4) and normalized steps $\overline{s}_k$. The resulting $\overline{x}_k$ are described in equations (23) and (24) (We can ignore the matrices Q_a, Q_b and Q_c in these equations because in the present case they are equal to the identity matrix of the corresponding dimension.)

We consider

$$u := \lambda^{36} = \sqrt{\frac{1}{1 + \sqrt{2 + \sqrt{2}}}}$$

and look at $\rho_0, \rho_1, \ldots \rho_9$ and ρ_{18} as free parameters. The other ρ's are defined as

$$\rho_k := \rho_9 \ \text{ for } \ k = 10, \ldots, 17, \qquad \rho_k := \rho_{18} \ \text{ for } \ k = 19, \ldots, 26, \tag{60}$$

$$\rho_k := \frac{u^2}{\rho_{k-27}\alpha_1\alpha_2} \ \text{ for } k = 27, \ldots, 30, \qquad \rho_k := -\frac{u^2}{\rho_{k-27}\alpha_1\alpha_2} \ \text{ for } k = 31, \ldots, 35 \tag{61}$$

and $\rho_k := \rho_{(k \bmod 36)}$ for $k \geq 36$. We then define the 9×9 matrices

$$\Phi(\rho) := \begin{pmatrix} 0 & -\rho \\ I_8 & \rho\, e_1^8 \end{pmatrix} \qquad \text{and} \qquad \Phi_k := \Phi(\rho_k). \tag{62}$$

These matrices Φ_k are the KEY part of our arguments, because the vectors $\overline{g}_k$ satisfy (56) if and only if the 9×9 matrices A_k with columns $Z^{-(k+i)}\overline{g}_{k+i}$, for $i = 0, \ldots 8$, are such that

$$A_{k+1} = A_k \Phi_k.$$

Once we grasp how the matrices A_k and Φ_k are related it is natural to define

$$\Psi(\rho) := \prod_{k=0}^{35} \Phi_k(\rho) \tag{63}$$

and search for a vector ρ such that the matrix $\Psi(\rho)^t$ has eigenvalues $\xi_0, \xi_1, \ldots, \xi_8$ given, respectively, by

$$-u^4, \ u^4 e^{7\mathrm{i}\pi/8}, \ u^4 e^{-7\mathrm{i}\pi/8}, \ u^3\mathrm{i}, \ -u^3\mathrm{i}, \ e^{\mathrm{i}\pi/4}, \ e^{-\mathrm{i}\pi/4}, \ -e^{\mathrm{i}\pi/4} \ \text{and} \ -e^{-\mathrm{i}\pi/4}, \tag{64}$$

where i is the imaginary unit. Once we find ρ, we can use the respective eigenvectors $v_0, v_1, \ldots, v_8 \in \mathbb{C}^9$ to define

$$\gamma_{2k} := \mathrm{Re}(v_{2k}) \in \mathbb{R}^9 \qquad \text{and} \qquad \gamma_{2k+1} := \mathrm{Im}(v_k) \in \mathbb{R}^9$$

for $k = 9, \ldots, 4$. We then consider the 9×9 matrix Γ_0 with rows $\gamma_0, \ldots, \gamma_8$ and define

$$\Gamma_k := \Gamma_0 \prod_{i=0}^{k-1} \Phi_k$$

(We use the convention that a product of the form $\prod_{i=a}^{b} M_i$ with $b < a$ equals the identity and a sum $\sum_{i=a}^{b} v_i$ with $b < a$ is equal to 0.) Finally we define

$$\overline{\sigma}_k := u^{-\lfloor k/9 \rfloor} \left(\prod_{i=0}^{\lfloor k/9 \rfloor - 1} \rho_{9i+(k \bmod 9)}^2 \right).$$

We now have all the ingredients of equation (59) and the iterates in (23) and (24). If we look at the eigenvalues ξ in (64) and take the 9×9 block diagonal matrix

$$\Theta(\lambda) := \begin{pmatrix} -u^4 & & & & \\ & M(\xi_1) & & & \\ & & M(\xi_3) & & \\ & & & M(\xi_5) & \\ & & & & M(\xi_7) \end{pmatrix}$$

for

$$M\left(re^{\mathrm{i}\theta}\right) := r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

then we can derive the relation

$$\Gamma_0\Psi(\rho) = \Theta(\lambda)\Gamma_0. \tag{65}$$

The identities

$$\Theta(\lambda) = \Theta(1)Z^{36}, \qquad \sum_{m=0}^{15}\Theta(1)^m = 0 \qquad \text{and} \qquad \prod_{i=0}^{3}\rho_{9i+k}^2 = u^4 \ \text{ for } k \in \mathbb{N}$$

and definition (59) yield

$$\overline{g}_{36m+k} = \Theta(1)^m\overline{g}_k \qquad \text{and} \qquad \overline{s}_{36m+k} = \Theta(1)^m\overline{s}_k.$$

It follows that $\sum_{k=0}^{p-1}\overline{s}_k^h = 0$ and we can apply Lemma 2. This lemma and equations (23) and (24) lead to

$$\overline{x}_{36m+k} = \Theta(1)^m\overline{x}_k.$$

Combining this with $\Theta(1)^{16} = I_9$ and $p = 16 \times 36$ we conclude that

$$\overline{x}_{k+p} = \overline{x}_k, \qquad \overline{g}_{k+p} = \overline{g}_k \qquad \text{and} \qquad \overline{s}_{k+p} = \overline{s}_k.$$

The matrices Φ_k are such that

$$\left(\prod_{i=1}^{j}\Phi_{k+i}\right)e_1^9 = e_{j+1}^9$$

for all k and $1 \leq j \leq 8$, because $\Phi_k e_i^9 = e_{i+1}^9$ for $1 \leq i \leq 8$. This implies (55) and, similarly, the remaining conditions in (53)–(56) can be verified by plugging (59) into them.

Therefore, all we need to produce $\overline{g}_k$, $\overline{s}_k$ and $\overline{x}_k$ that would satisfy the compatibility conditions (53)–(56) is to find $\rho_0, \ldots, \rho_9$ and ρ_{18} in such way that the matrix $\Psi(\rho)$ in (63) has the eigenvalues ξ in (64). The Mathematica script in the supplementary material proves that these ρ_k's exist using Lemma 1. It also finds bounds on them, computes the corresponding eigenvectors, $\overline{g}_k$, $\overline{s}_k$ and $\overline{x}_k$ using interval arithmetic and shows that these parameters satisfy the geometric constraints (14)–(16). The proof of Lemma 5 is thus completed.

A Technicalities

This appendix begins with an explanation as to why the form of the iterates, function values and gradients in reference [5] was already described in [13]. We then prove Lemma 1 and, finally, the theorems.

Let us then see how several equations in [5] and [13] correspond to (4). In [5] and the example for Newton's method in [13] the parameter d_n is equal to 1. In the example for the BFGS method in [13] it is equal to 3. Equation (4) corresponds to equations (10)–(12) in [13]. It is generalized in equations (34)–(37) of [13]. [5] adds a constant f^* to f, but this is irrelevant. It calls λ by t and thinks in terms of the steps $\delta_k = x_{k+1} - x_k$, so that $\delta_k = Q^kD(\lambda)^k\overline{\delta}$ in [5]'s equation (2.1). This is just an affine change of coordinates of (4) and only a different notation for the corresponding s_k's used in [13]. The matrix M in [5]'s equation (2.2) is equal to the matrix $QD(\lambda)$, whose kth power multiplies $\overline{x}_k$ in equation (4). The matrix P in [5]'s equation (2.5) is equal to the matrix $\lambda QD(\lambda)^{-1}$ whose kth power multiplies $\overline{g}_k$ in equation (4). Therefore, the iterates, the function values and the gradients are described in essentially the same way in the examples in [5, 12, 13], in terms of orthogonal matrices Q scaled by powers of λ (or t) via the matrices $D(\lambda)$ or their inverses.

Proof of Lemma 1. Let us define $\delta_0 = 0$ and $x_0 = \overline{x}$ and consider δ_k and x_k defined inductively by $\delta_k = Af(x_{k-1})$ and $x_k = x_{k-1} - \delta_k$. The lemma will be proved if we show that

$$\|\delta_k\|_\infty \leq a^{k-1}b, \qquad \|x_{k+1} - \overline{x}\| \leq \frac{1-a^k}{1-a}b \quad \text{and} \quad \|f(x_k)\|_\infty \leq a^k \|f(\overline{x})\|_\infty, \quad (66)$$

because these bounds imply that x_k converges to x_∞ with $\|x_\infty - \overline{x}\|_\infty \leq b/(1-a)$ and $f(x_\infty) = 0$. Let us then prove (66) by induction. Equation (66) certainly holds for $k = 0$. Assuming that (66) holds for k, we conclude from (7) and definition of δ_k that

$$\|\delta_{k+1}\|_\infty \leq \sup_{1 \leq i \leq n} \left\|A^t e_i\right\|_1 \|f(x_k)\|_\infty \leq a^k \left\|A^t e_i\right\|_1 \|f(\overline{x})\|_\infty \leq a^k b$$

and the first bound in (66) holds for $k+1$. The second bound on (66) follows from the analogous bound for $\|x_k - \overline{x}\|_\infty$ and the bound in δ_{k+1} above. It shows that $x_{k+1} \in D$ and the segment S connecting x_k to x_{k+1} is contained in D. As a result, the Mean Value Theorem for the function $f_i : D \mapsto \mathbb{R}$ implies that, for some $\xi \in S$,

$$|f_i(x_{k+1})| = |f_i(x_k - Af(x_k))| = \left|f_i(x_k) - \nabla f_i(\xi)^t Af(x_k)\right|$$

$$= \left|\left(A^t \nabla f_i(\xi) - e_i\right)^t f(x_k)\right| \leq \left|A^t \nabla f_i(\xi) - e_i\right|_1 \|f(x_k)\|_\infty \leq a \|f(x_k)\|_\infty.$$

Thus, $|f(x_{k+1})|_\infty \leq a \|f(x_k)\|_\infty \leq a^{k+1} \|f(\overline{x})\|_\infty$ and we are done. □

Proof of Theorem 1. We start by rewriting (1) and (2) as

$$M_k M_k^t s_k = -\alpha_k g_k, \quad (67)$$

for $s_k := x_{k+1} - x_k$ and $g_k := \nabla f(x_k)$. Equation (67) shows that $s_k^t g_k = -\left\|M_k^t s_k\right\|^2 / \alpha_k \leq 0$ and the first Wolfe condition (3) yields

$$f_k := f(x_k) \geq f_{k+1}. \quad (68)$$

Since f has bounded level sets this implies that

$$x_k \in K := \{x \in \mathbb{R}^n \text{ with } f(x) \leq f(x_0)\}.$$

The set K is compact and the matrices M_k are bounded. Since f has continuous first derivatives there exists a constant κ such that $\|g\|_k \leq \kappa$, $\|M_k\| \leq \kappa$, $\|x_k\| \leq \kappa$ and $\|s_k\| \leq \kappa$.

To prove that $\lim_{k \to \infty} g_k = 0$ we use the well known result that if a bounded sequence $\{u_k,\ k \in \mathbb{N}\} \subset \mathbb{R}$ is such that all its converging subsequences u_{n_k} converge to zero then u_k itself converges to 0. Let us then consider a subsequence g_{n_k} such that $\lim_{k \to \infty} \left\|g_{n_k}\right\| = L$ and show that $L = 0$. Equation (67) leads to $\|g_k\| \leq \kappa^3 / \alpha_k$. Thus, if some sub subsequence of $\alpha_{n_{k_j}}$ converges to $+\infty$ then $L = \lim_{j \to \infty} \left\|g_{n_{k_j}}\right\| = 0$ and we are done. We can then assume that there exists A such that $\alpha_{n_k} \leq A$ for all k. Equations (3), (67) and (68) yield

$$\left\|M_{n_k}^t s_{n_k}\right\|^2 \leq -\alpha_{n_k} s_{n_k}^t g_{n_k} \leq \alpha_{n_k} \left(f\left(x_{n_k}\right) - f\left(x_{n_k+1}\right)\right) / \sigma \leq \alpha_{n_k} \left(f\left(x_{n_k}\right) - f\left(x_{n_{k+1}}\right)\right) / \sigma$$

$$\leq A\left(f\left(x_{n_k}\right) - f\left(x_{n_{k+1}}\right)\right)/\sigma$$

and

$$\sum_{k=0}^{\infty} \left\| M_{n_k}^t s_{n_k} \right\|^2 \leq \frac{A}{\sigma}\left(f\left(x_{n_0}\right) - \inf_{x \in K} f(x) \right).$$

Therefore, $\lim_{k\to\infty} M_{n_k}^t s_{n_k} = 0$. If follows that $g_{n_k} = -M_{n_k} M_{n_k}^t s_{n_k}/\alpha_{n_k}$ converges to 0, because the M_{n_k} are bounded and $\alpha_{n_k} \geq \overline{\alpha}$. □

Proof of Theorem 3. This theorem is an specialization of Theorem 4 in page 145 of [13]. The reader can prove Theorem 3 by using Theorem 4 in [13] by realizing that the hypothesis of Theorem 3 corresponds to a simplified version of the more general concept of seed, which is described in definition 7 in page 142 of [13] (There is a typo in the end of the item 3 in this definition. It should read like "...$D(0)\,\overline{s}_r$ and $D(0)\,Q\overline{s}_{r+1}$.", with a Q between $D(0)$ and $\overline{s}_{r+1}$.) The reader will note that our choice for the diagonal exponents in (11) (0 in the first block, 1 in the second and $d_n > 2$) leads to vacuous sub itens (c) and (d) in the fouth item in the definition of seed. As a result, we do not need to worry about these items. Morevoer, we only need to worry about item (b) in the case $i = j = n$. This is why we have hypothesis (17) in Theorem 3. The reader can now read the many details and involved proving Theorem [13]. This proof is very technical and there is no point repeating it here.

References

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

[2] D. Bertsekas, Nonlinear Programming, Athena Scientific, 2nd Edition, (1999).

[3] Cotton, É: Sur les solutions asymptotiques des équations différentielles. Annales scientifiques de l'École Normale Supérieure, Sér. 3, 28, 473-521, (1911).

[4] Dai, Y.–H.: Convergence properties of the BFGS algorithm. SIAM J. Optim. 13(3), 693–701, (2002).

[5] Dai, Y.–H.: A perfect example for the BFGS method, Published online March (2012).

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

[7] Fefferman, C., 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.

[8] Hadamard, J., Sur l'itération et les solutions asymptotiques des équations différentielles, Bull. Soc. Math. France 29, 224–228, (1901).

[9] Kurdyka, K., On gradients of functions definable in o-minimal structures, Ann. Inst. Fourier 48, 769–783, (1998).

[10] Lojasiewicz, S. Ensemble semi-analytiques. IHES Lecture notes, (1965).

[11] Mascarenhas, W.F., Newton's iterates can converge to non-stationary points, Mathematical Programming, Volume 112(2), 327-334, (2008).

[12] Mascarenhas,W.F.: The BFGS algorithm with exact line searches fails for non-convex functions. Math. Program. 99(1), 49–61 (2004).

[13] Mascarenhas,W.F.: On the divergence of line search methods, Comput. Appl. Math. vol.26 no.1 (2007).

[14] Nocedal J., Wright, S.J., Numerical Optimization, Springer series in operations research, Springer Verlag (1999).

[15] Ortega, J. M.: The Newton-Kantorovich Theorem. Amer. Math. Monthly 75 (6): 658–660 (1968).

[16] Perron, O. : Ueber staebilitat und Asymptotisches Verhalten der Loesung eines Systeme endlicher Differetialgleichungen J. Reine. Angew. Math. 161, 41-64, (1929).

[17] Poincaré, H.: *Les méthodes nouvelles de la mécanique céleste*, Gauthier-Villars et fils (1892).

[18] Powell, M. J. D., Nonconvex minimization calculations and the conjugate gradient method, Lecture Notes in Mathematics Volume 1066, 122-141 (1984).

[19] Rally, L. B.: A comparison of the existence theorems of Kantorovich and Moore, Siam J. Numer. Anal., vol. 17, no 1, Society for Industrial and Applied Mathematics Feb. (1980).

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