



STUDYING THE BASIN OF CONVERGENCE OF METHODS FOR COMPUTING PERIODIC ORBITS

MICHAEL G. EPITROPAKIS* and MICHAEL N. VRAHATIS†
*Computational Intelligence Laboratory, Department of Mathematics,
 University of Patras Artificial Intelligence Research Center,
 University of Patras, GR-26110 Patras, Greece*
 *mikeagn@math.upatras.gr
 †vrahatis@math.upatras.gr

Received January 15, 2011; Revised April 10, 2011

Starting from the well-known Newton's fractal which is formed by the basin of convergence of Newton's method applied to a cubic equation in one variable in the field \mathbb{C} , we were able to find methods for which the corresponding basins of convergence do not exhibit a fractal-like structure. Using this approach we are able to distinguish reliable and robust methods for tackling a specific problem. Also, our approach is illustrated here for methods for computing periodic orbits of nonlinear mappings as well as for fixed points of the Poincaré map on a surface of section.

Keywords: Structure of fractals; fractals; nonlinear dynamics; numerical methods (mathematics); periodic orbits of nonlinear mappings; fixed points of the Poincaré map.

1. Introduction

Many problems in different areas of science and technology can be studied using periodic orbits of nonlinear mappings or dynamical systems. For example, such problems appear in Quantum Mechanics where a weighted sum over unstable periodic orbits yields quantum mechanical energy level spacings as well as in Statistical Mechanics where a weighted, according to the values of their Liapunov exponents, sum over unstable periodic orbits can be used to calculate thermodynamic averages [Gutzwiller, 1990; Morriss & Rondoni, 1996; Henry *et al.*, 2000]. Furthermore, periodic orbits play a major role in assigning the vibrational levels of highly excited polyatomic molecules [Vrahatis *et al.*, 2001a; Perdiou *et al.*, 2002] as well as in Celestial Mechanics and Galactic Dynamics (see for example

[Kalantonis *et al.*, 2001; Vrahatis *et al.*, 2001b; Perdios *et al.*, 2002; Kalantonis *et al.*, 2003a, 2003b; Katsanikas & Patsis, 2011; Katsanikas *et al.*, 2011a; Katsanikas *et al.*, 2011b]). Also, periodic orbits can be used in the study of the structure and breakdown properties of invariant tori in the case of symplectic mappings of direct relevance of the beam stability problem in circular accelerators like the Large Hadron Collider (LHC) machine at the European Organization for Nuclear Research (CERN). Such a 4-D symplectic mapping can be found in [Vrahatis *et al.*, 1996; Vrahatis *et al.*, 1997] which describes the (instantaneous) effect experienced by a hadronic particle as it passes through a magnetic focusing element of the FODO cell type [Bountis & Tompaids, 1991]. Recently periodic orbits have been used in cryptography

† Author for correspondence

for the construction of novel cryptosystems [Vrahatis *et al.*, 2010].

In general, analytic expressions for evaluating periodic orbits of nonlinear mappings or fixed points of the Poincaré map on a surface of section are not available. A well-known and widely used efficient approach to compute periodic orbits or fixed points is to solve a system of nonlinear equations by using Newton's method. On the other hand, Newton's method as well as Newton-like methods [Dennis & Schnabel, 1996; Ortega & Rheinboldt, 2000], often converge to a solution almost independently, of the starting guess of the considered solution. This is obviously a disadvantage when various other solutions exist nearby, all of which are needed in the application under consideration. This can be easily verified by studying the basins of convergence of these methods [Androulakis & Vrahatis, 1996]. It may also happen that these methods fail due to the nonexistence of derivatives or poorly behaved partial derivatives.

In general, the performance of a method for computing a specific solution can be studied by the basin of convergence of the considered method at the solution. Without loss of generality, if we consider the method at hand as Newton's method, we can draw its basin of convergence as follows: In a given domain, each picture's element (pixel) corresponds to an initial guess of Newton's method and takes the color of the solution to which the method converges, while, if for this initial guess the method does not converge, after a specific number of iterations, this element is colored by using an additional color. As we shall see later, Newton's method is *unstable* since it is sensitive to small perturbations of the starting points, and its basin of convergence has a *fractal structure* [Barnsley, 1988; Bountis, 2004].

In 1879, *Arthur Cayley* asked the following question:

Given an initial input to which root will Newton's method converge?

Thus, Cayley was the first who noticed the difficulties in generalizing Newton's method to complex roots of polynomials with degree greater than two, and complex initial values [Cayley, 1879]. The answer was only fully understood recently and leads to a beautiful *fractal pattern* — the so-called *Newton's fractal*.

Definition 1.1. *Newton's fractal* is a boundary set in the complex plane which is characterized by *Newton's method* applied to a fixed polynomial $p(z) \in \mathbb{C}[z]$.

In this contribution, we present methods that can be used for the computation of periodic orbits and we study their basins of convergence. In the next section, we outline the problem formulation, while in Sec. 3, we briefly present methods for computing periodic orbits. In Secs. 4 and 5 we provide some applications, a discussion for extracting information from a picture and some concluding remarks.

2. Problem Formulation

A large variety of methods can be applied to compute periodic orbits of nonlinear mappings, or fixed points of the Poincaré map on a surface of section. The three principal categories of these methods are:

- (1) *Methods for computing fixed points*
- (2) *Methods for computing solutions of systems of nonlinear equations*
- (3) *Methods for computing global minimizers of objective functions.*

2.1. Fixed points approach

The problem of computing periodic orbits of nonlinear mappings (or fixed points of the Poincaré map on a surface of section) can be tackled by using *fixed points*. More specifically, the problem of finding *periodic orbits of nonlinear mappings*

$$\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n)^\top : \mathcal{D}_n \subset \mathbb{R}^n \rightarrow \mathbb{R}^n,$$

of period p amounts to finding *fixed points* $x^* = (x_1^*, x_2^*, \dots, x_n^*)^\top \in \mathcal{D}_n$ of period p which satisfy the following equation:

$$\Phi^p(x^*) = \underbrace{\Phi(\Phi(\dots \Phi(\Phi(x^*)) \dots))}_{p \text{ times}} = x^*. \quad (1)$$

2.2. Systems of nonlinear equations approach

This problem can also be tackled using *systems of nonlinear equations*. Finding such a periodic orbit (or fixed points of the Poincaré map on a surface

of section) is equivalent to solving the following equation:

$$F_n(x) = \mathcal{O}_n,$$

with

$$F_n = (f_1, f_2, \dots, f_n)^\top = \Phi^p - I_n : \mathcal{D}_n \subset \mathbb{R}^n,$$

where Φ^p is given in Eq. (1), I_n is the $n \times n$ identity matrix and where $\mathcal{O}_n = (0, 0, \dots, 0)^\top$ is the origin of \mathbb{R}^n . Obviously, the above equation is equivalent to the following system of equations:

$$\begin{cases} \Phi_1^p(x_1, x_2, \dots, x_n) - x_1 = 0, \\ \Phi_2^p(x_1, x_2, \dots, x_n) - x_2 = 0, \\ \vdots \\ \Phi_n^p(x_1, x_2, \dots, x_n) - x_n = 0. \end{cases} \quad (2)$$

2.3. Global optimization approach

The same problem can be addressed through the *global minimization of an objective function*. Thus, finding such a periodic orbit (or fixed points of the Poincaré map on a surface of section) is equivalent to computing the *global minimizer* $x^* = (x_1^*, x_2^*, \dots, x_n^*)^\top \in \mathcal{D}_n$ of the following objective function $f : \mathcal{D}_n \subset \mathbb{R}^n \rightarrow \mathbb{R}$,

$$f(x) = \sum_{i=1}^n (\Phi_i^p(x) - x_i)^2,$$

or, more general

$$f(x) = \|(\Phi_1^p(x) - x_1, \Phi_2^p(x) - x_2, \dots, \Phi_n^p(x) - x_n)^\top\|,$$

that is,

$$f(x^*) = 0, \quad \text{for } x^* \in \arg \min_x f(x). \quad (3)$$

2.4. Convergence characteristics of methods

As we have already mentioned previously, there is a large variety of methods that can be used to solve Problems (1)–(3). These methods possess advantages and disadvantages and it is not always evident which one is more *efficient*, *effective* or *robust* for a given class of nonlinear mappings. In principle, these methods are formulated as iterative processes

and three main problems may be associated with any one method for solving a specific problem, namely:

- (a) Is the method well-defined, that is, can it be continued to a satisfactory end?
- (b) Do the iterates converge to a solution?
- (c) How economical is the entire operation?

In most instances only partial answers can be given to the above questions. Question (b) consists of three parts:

- (b1) The question of whether the iterates converge to a limit (convergence or divergence).
- (b2) The question of whether this limit is in fact a solution.
- (b3) The question of whether for a given initial guess the algorithm converges to the closest solution.

Question (c) concerns the computational complexity of the algorithm and it includes the following sub-questions:

- (c1) What is the cost of any one step of the iterative process?
- (c2) How fast does the sequence converge?
- (c3) How sensitive (stable) is the process to changes of the function or the initial data?
- (c4) How is the performance of the method in the presence of noise (impact of imprecise information regarding the values of the function)?

2.5. Imprecise information — presence of noise

To study the impact of imprecise information (regarding the values of the function), we simulate imprecisions through the following approach. Information about the values of the function $f(x)$ is obtained in the form of $f^\eta(x)$, where $f^\eta(x)$ is an approximation to $f(x)$, contaminated by a small amount of noise η . Thus, the function values are obtained, for the *additive noise* case, as [Elster & Neumaier, 1997]:

$$f^\eta(x) = f(x) + \eta,$$

and for the *multiplicative noise* case, as [Parsopoulos & Vrahatis, 2002, 2010]:

$$f^\eta(x) = f(x)(1 + \eta), \quad 1 + \eta > 0,$$

where η is a Gaussian noise term with zero mean and standard deviation σ

$$\eta \sim \mathcal{N}(0, \sigma^2),$$

i.e. relative stochastic errors are used. Suppose that a normal noise distribution may be regarded as an approximation of reality based on the maximum entropy principle [Jaynes, 1979; Beyer, 2000], the probability density function of the noise reads

$$p(\eta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\eta}{\sigma}\right)^2\right].$$

Here, to obtain η , we use the method of Box and Muller [1958], using various *noise strengths* (standard deviations) σ . In our approach, the assumption of the normal distribution is not mandatory. However, the measurement errors in nature and technology are very often modeled using this distribution. Moreover, most of the mean values of stochastic processes have this distribution (central limit theorem of statistics) [Beyer, 2001].

2.6. Aim

In this contribution, we try to answer the above addressed questions relative to the convergence characteristics of methods by providing experimental data concerning (b3), (c2), (c3) and (c4). The other questions can usually be addressed more satisfactorily by theory.

Our approach consists of analyzing and visualizing the convergence behavior of the considered method. Specifically, using different colors, we exhibit the regions of convergence to a solution of a given function. Thus, we are able to study through images, the advantages and disadvantages of any method as well as to compare various methods in order to choose the proper one for a given problem.

Remark 2.1. If an image has for instance a *fractal-like structure*, then the considered method is *unstable* since it is sensitive to small perturbations of the starting points.

3. Methods

In this section, we briefly describe a few methods that can be used for computing periodic orbits of nonlinear mappings or fixed points of the Poincaré map on a surface of section.

3.1. Methods for computing fixed points

One of the most famous and powerful theorems of Mathematics is the Banach fixed point theorem [Banach, 1922].

Theorem 1 (Banach's Fixed Point Theorem [1922]). *Let $F_n = (f_1, f_2, \dots, f_n)^\top : \mathcal{D}_n \subset \mathbb{B} \rightarrow \mathcal{D}_n$, be defined on a closed subset of a Banach space \mathbb{B} . Assume that F_n is contractive with a factor $L < 1$, (Lipschitz constant L such that $\|F_n(x) - F_n(y)\| \leq L\|x - y\|$, $\forall x, y \in \mathcal{D}_n$). Then there exists a unique $x^* \in \mathcal{D}_n$ such that x^* is a fixed point of F_n , i.e. $x^* = F_n(x^*)$.*

The proof of Banach's theorem is based on a simple *iteration algorithm* defined as follows:

Banach's Iteration Algorithm:

Choose any starting point x_0 from \mathcal{D}_n and let

$$x_{k+1} = F_n(x_k), \quad \text{for } k = 0, 1, \dots \quad (4)$$

Then

$$\|x_k - x^*\| \leq L^k \|x_0 - x^*\|.$$

Assume that $x_0 = 0 \in \mathcal{D}_n$, then the *relative error* of the k th approximation x_k , is estimated by

$$\frac{\|x_k - x^*\|}{x^*} \leq L^k.$$

This bound is sharp. Thus to guarantee that the *relative error* is at most ε , $\varepsilon > 0$, the following number of *iterations* (*function evaluations*) are required:

$$\#S = \#S(\varepsilon, L) = \left\lceil \frac{\log_2 \varepsilon^{-1}}{\log_2 L^{-1}} \right\rceil.$$

Fixed point problems with $L = 1$ and $L < 1$ appear in the study of *nonlinear dynamics*. They model *conservative* or *dissipative* systems depending on whether the mapping is *area-preserving* or *area-contracting*, respectively.

Remark 3.1. Some complexity issues follows:

- (a) For a Lipschitz function that is contractive $L < 1$ and large dimension n the *Banach Simple Iteration Algorithm* (4) is *optimal* [Sikorski, 2001].
- (b) For moderate n and $L = 1$, the *Interior Ellipsoid Algorithm* is *optimal* [Sikorski, 2001]. It requires $cn \log_2 \varepsilon^{-1}$ function evaluations to

compute an ε -residual approximation $\tilde{x} : \|\tilde{x} - F_n(\tilde{x})\| \leq \varepsilon$.

- (c) The *worst-case complexity* of computing an ε -absolute approximation $\|\tilde{x} - x^*\| \leq \varepsilon$ for $L = 1$ is *infinite*.
- (d) For Lipschitz functions with constant $L > 1$ with respect to the infinity norm, *Homotopy Continuation, Simplicial and Newton-type Methods* for ε -residual solutions exhibit *exponential complexity in the worst case* and that the lower bound on the complexity is also exponential.

3.2. Methods for computing solutions of systems of nonlinear equations

The well known and widely used *Newton's method* computes a root of continuously differentiable functions $F_n = (f_1, f_2, \dots, f_n)^\top : \mathcal{D}_n \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$. The corresponding iterative scheme can be briefly described as follows:

Newton's iterative scheme:

Choose a *starting point* x_0 . The iterations for $k = 0, 1, \dots$ are given by:

Solve the *system of linear equations*: $J_{F_n}(x_k)s_k = -F_n(x_k)$, for s_k .

Set: $x_{k+1} = x_k + s_k$,

where $J_{F_n}(x_k)$ is the $n \times n$ Jacobian matrix:

$$J_{F_n}(x_k) = \begin{bmatrix} \frac{\partial f_1(x_k)}{\partial x_1} & \frac{\partial f_1(x_k)}{\partial x_2} & \dots & \frac{\partial f_1(x_k)}{\partial x_n} \\ \frac{\partial f_2(x_k)}{\partial x_1} & \frac{\partial f_2(x_k)}{\partial x_2} & \dots & \frac{\partial f_2(x_k)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(x_k)}{\partial x_1} & \frac{\partial f_n(x_k)}{\partial x_2} & \dots & \frac{\partial f_n(x_k)}{\partial x_n} \end{bmatrix}.$$

Newton's method computes the Jacobian matrix at every iteration which costs n^2 function evaluations per iteration.

An alternative approach to Newton's method is *Broyden's method* [Broyden, 1965]. Broyden's method belongs to the wide class of quasi-Newton methods utilized for the numerical solution of n nonlinear equations in n variables. The class of quasi-Newton methods, in general does not require

calculations of the true Jacobian matrix $J_{F_n}(x)$. Instead they iteratively construct their own approximation A_k which tends to mimic the behavior of the true Jacobian matrix as the algorithm proceeds. Broyden's method initially computes an estimation A_0 of the Jacobian matrix $J_{F_n}(x_0)$ at the first iteration and then performs an update at the remaining iterations. The algorithmic scheme of the Broyden's method can be briefly described as follows:

Broyden's iterative scheme:

Choose a *starting point* x_0 and an *estimate* A_0 of the Jacobian matrix $J_{F_n}(x_0)$. The iterations for $k = 0, 1, \dots$ are given by:

Solve the *system of linear equations*: $A_k s_k = -F_n(x_k)$, for s_k .

Set: $x_{k+1} = x_k + s_k$,

Set: $y_k = F_n(x_{k+1}) - F_n(x_k)$,

Update A_{k+1} by: $A_{k+1} = A_k + ((y_k - A_k s_k)s_k^\top) / (s_k^\top s_k)$.

It should be noted that special attention needs to be paid to the initial estimation A_0 of the Jacobian matrix. Generally A_k has to be a nonsingular matrix. Additionally, based on the convergence analysis of Broyden's method [Nocedal & Wright, 2006], the initial approximation A_0 of the Jacobian matrix has to be close enough to the true Jacobian matrix at the solution. Thus, a good choice for A_0 can be critical to Broyden's performance. To this end, a good choice for A_0 can be either the true Jacobian matrix at the first iteration, or some finite-difference approximation of the Jacobian matrix at the initial point. The approximation of the Jacobian matrix makes Broyden's method to have a lower computational cost in comparison to Newton's method. Broyden's method requires n^2 functions for the initial computation of the true Jacobian matrix and n function evaluations at each step. Additionally, it has been proved that when Broyden's method is applied to a linear system, it terminates in $2n$ steps.

Remark 3.2. In the case where the partial derivatives of $f(x)$ are not available, we can calculate them using some finite-difference approximation procedure. Thereby, the partial derivative of $f(x)$ with respect to x_i for $i = 1, 2, \dots, n$ can be approximated as follows:

$$\frac{\partial f(x)}{\partial x_i} \approx \frac{f(x + h_i e_i) - f(x)}{h_i},$$

where h_i is a small perturbation of x_i and e_i is the i th column of the $n \times n$ identity matrix I_n . Here, a special attention has to be paid to determine the value of stepsize h_i since a wrong estimation could lead to large round-off errors [Bräuninger, 1981; Dennis & Schnabel, 1996; Press *et al.*, 2007]. A simple rule, when $f(x)$ is assumed to be computed accurately to the machine's precision eps_m , is to perturb roughly half the digits of x_i , with $h_i = \sqrt{\text{eps}_m} \cdot x_i$, [Dennis & Schnabel, 1996]. This well known rule is generally satisfactory in practice.

3.3. Minimization methods

A well-known class of algorithms for unconstrained minimization of functions $f(x)$ in n real variables $f: \mathcal{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}$ having Lipschitz continuous first partial derivatives whose gradient

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^\top,$$

is available, is the *steepest descent methods* [Dennis & Schnabel, 1996; Ortega & Rheinboldt, 2000; Vrahatis *et al.*, 2000] first proposed by Cauchy [1847]. In the general case, a method in this class computes at each iteration a descent search direction, and then decides how far to move along in this direction.

Specifically, *Cauchy's method* utilizes the following iterative scheme:

$$x_{k+1} = x_k - \lambda_k \nabla f(x_k), \quad \text{for } k = 0, 1, 2, \dots \quad (5)$$

where $\lambda_k > 0$. As it can be observed from this iterative scheme, the Cauchy method starting from x_0 at each iteration k moves along the steepest descent direction $-\nabla f(x_k)$ of the problem at hand with a stepsize λ_k . One of its advantages which simultaneously can be a disadvantage in specific cases is that it only requires calculations of the gradient of the problem at hand and not its second derivatives. Additionally, in many cases a wrong estimation of the stepsize λ_k may prevent its convergence, thus several methodologies can be provided to automatically adapt λ_k . One of the well-known and widely used methodologies is the Armijo method.

Armijo provided in 1966 [Armijo, 1966; Vrahatis *et al.*, 2000] a *modification of the steepest descent method* which automatically adapts the stepsize λ_k of the iterative Scheme (5). Specifically, according to Armijo's approach, a stepsize λ_k is acceptable for the iterative Scheme (5) at the

iteration k if the following inequality is fulfilled:

$$f(x_k - \lambda_k \nabla f(x_k)) - f(x_k) \leq -\frac{1}{2} \lambda_k \|\nabla f(x_k)\|^2,$$

otherwise λ_k is repetitively subdivided until the above inequality is fulfilled. In other words, the Armijo's method changes the stepsize length in such a way that it warrants a sufficient decrease in the objective function f , as measured by the aforementioned inequality. The reduction in the objective function f should be proportional to both the utilized stepsize λ_k and the $\|\nabla f(x_k)\|^2$.

A relative to Armijo's step length adaptation strategy consists in accepting a positive stepsize λ_k along the direction d_k if it satisfies the Wolfe conditions [Wolfe, 1969, 1971]. A step length λ_k is said to satisfy the *Wolfe conditions* if the following two inequalities hold:

$$(a) \quad f(x_k + \lambda_k d_k) \leq f(x_k) + c_1 \lambda_k \langle d_k, \nabla f(x_k) \rangle,$$

Armijo Rule

$$(b) \quad \langle d_k, \nabla f(x_k + \lambda_k d_k) \rangle \geq c_2 \langle d_k, \nabla f(x_k) \rangle,$$

Curvature Condition

with $0 < c_1 < c_2 < 1$. The above inequality (a) ensures that the step length λ_k decreases the function f "sufficiently" along the search direction d_k , while inequality (b) ensures that the slope has been reduced sufficiently. Furthermore, inequalities (a) and the following inequality (b) together form the so-called *strong Wolfe conditions*,

$$(\hat{b}) \quad |\langle d_k, \nabla f(x_k + \lambda_k d_k) \rangle| \leq c_2 |\langle d_k, \nabla f(x_k) \rangle|.$$

Based on Cauchy's method various other gradient methods have been proposed. A well known and widely used class is the *conjugate gradient methods*. These methods generate a sequence of points x_k , that tend to approximate a minimizer x^* of the objective function f with the following iterative scheme:

$$\begin{aligned} x_{k+1} &= x_k + \lambda_k d_k, \\ d_k &= \begin{cases} -g_k \equiv -\nabla f(x_k), & k = 0 \\ -g_k + b_k^{\mathbf{x}} d_{k-1}, & k \geq 1 \end{cases} \end{aligned} \quad (6)$$

where d_k is the considered direction along which the function is minimized, $g_k \equiv \nabla f(x_k)$, λ_k is selected to minimize f along the search direction d_k , and $b_k^{\mathbf{x}}$ is a scalar parameter. The iterative process starts with an initial point x_0 and $d_0 = -g_0$. Various versions of conjugate gradient methods have been

proposed. Each method corresponds to a different choice of the scalar parameter $b_k^{\mathbf{X}}$. A few of the best-known versions are given below:

$b_k^{\mathbf{C}} = 0$	Cauchy [Cauchy, 1847]
$b_k^{\mathbf{FR}} = \frac{\ g_k\ ^2}{\ g_{k-1}\ ^2}$	Fletcher–Reeves [Fletcher & Reeves, 1964]
$b_k^{\mathbf{PR}} = \frac{\langle g_k, g_k - g_{k-1} \rangle}{\ g_{k-1}\ ^2}$	Polak–Ribière [Polak & Ribière, 1969]
$b_k^{\mathbf{HS}} = \frac{\langle g_k, g_k - g_{k-1} \rangle}{\langle d_k, g_k - g_{k-1} \rangle}$	Hestenes–Stiefel [Hestenes & Stiefel, 1952]
$b_k^{\mathbf{P}} = \frac{\langle g_k, (x_k - x_{k-1}) - (g_k - g_{k-1}) \rangle}{\langle d_k, g_k - g_{k-1} \rangle}$	Perry [Perry, 1978]

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product.

Furthermore, two additional optimization methods that belong to the wide category of *variable metric methods* or the *quasi-Newton family* methods are described below. In this category, it is assumed that the function to be optimized can be locally approximated by a quadratic function around an optimum. Variable metric methods utilize information from the gradient $\nabla f(x_k)$ and the Hessian matrix:

$$H_f(x_k) = \begin{bmatrix} \frac{\partial^2 f(x_k)}{\partial x_1^2} & \frac{\partial^2 f(x_k)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x_k)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x_k)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x_k)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x_k)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x_k)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x_k)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x_k)}{\partial x_n^2} \end{bmatrix},$$

to find an optimum. They do not compute the Hessian matrix $H_f(x_k)$ but iteratively update it by analyzing successive gradient directions. Strictly speaking, let us consider the aforementioned iterative scheme [Eq. (6)] and generate the direction d_k with the following equation:

$$d_k = -\psi_k g_k + b_k^* d_{k-1},$$

in which ψ_k and b_k^* are parameters which are to be determined. It can be observed that, if $\psi_k = 1$, then we have the classical conjugate gradient methods described above, and it remains to select a proper update rule for the scalar b_k^* . On the other hand, if $b_k^* = 0$, then we obtain a different class of algorithms according to the selection of ψ_k

parameter. In this class, we can let ψ_k either be a positive scalar or a positive definite matrix. If $\psi_k = 1$, we have the traditional Cauchy method. If $\psi_k = H_f(x_k)^{-1}$, or an approximation of it, then we get, respectively, the Newton or the quasi-Newton algorithms applied to $F_n = \nabla f$ since the Jacobian matrix of the gradient is the Hessian matrix, i.e. $J_{\nabla f} \equiv H_f$.

Here, we describe two well-known and widely used methods of the quasi-Newton class, namely the *DFP* (*Davidon–Fletcher–Powell*) *method* and the *BFGS* (*Broyden–Fletcher–Goldfarb–Shanno*) *method* [Nocedal & Wright, 2006]. The iterative scheme of both methods is briefly described below:

DFP's iterative scheme:

Choose a *starting point* x_0 and an *estimate* A_0 of the Hessian matrix $H_f(x_0)$. The iterations for $k = 0, 1, \dots$ are given by:

Compute the correction $s_k : s_k = -\lambda_k A_k^{-1} \nabla f(x_k)$, where λ_k satisfies the Wolfe conditions,

Set: $x_{k+1} = x_k + s_k$,

Set: $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$,

Update the approximate Hessian matrix A_{k+1} by:

$$A_{k+1} = \left(I_n - \frac{y_k s_k^\top}{y_k^\top s_k} \right) A_k \left(I_n - \frac{s_k y_k^\top}{y_k^\top s_k} \right) + \frac{y_k y_k^\top}{y_k^\top s_k}.$$

The algorithmic scheme of the BFGS method is the same as the DFP's method. BFGS differs from DFP only in the update rule of the Hessian matrix,

which is defined by the following equation:

$$A_{k+1} = A_k + \frac{y_k y_k^\top}{y_k^\top s_k} - \frac{A_k s_k s_k^\top A_k}{s_k^\top A_k s_k}.$$

Other popular update formulas, such as the Symmetric Rank 1 (SR1) method, and the Broyden's family, as well as a thorough description of the above methods along with examples, their advantages and disadvantages can be found in [Nocedal & Wright, 2006].

4. Results

As previously mentioned, our approach consists of analyzing and visualizing the convergence behavior of a considered method. Specifically, using different colors, we display the regions of convergence to a solution of a given function, and thus we are able to immediately observe and compare the advantages and disadvantages of the methodologies at hand.

Next we describe how our approach accumulates information in a picture. Firstly, we distinguish the solutions of the considered function by marking them with different colors. Each picture's element (pixel) corresponds to an initial guess of the analyzed method and it is colored according to its convergence. Specifically, each pixel takes the color of the solution to which the corresponding method converges, while, if for this initial guess the considered method does not converge after a specific number of iterations, the pixel is colored white. It is obvious that, if a picture exhibits for instance a fractal-like structure, then the considered method is not stable since it is sensitive to small perturbations of the starting points.

Of course, it is not necessary to know the coordinates of the solutions beforehand, since the considered iterative scheme is applied for any bounded domain of starting points and according to its convergence, it saves the coordinates of the solutions. The first solution can be colored red, the second green, the third blue and so on. Furthermore, in order to display how fast the method converges to a solution for the specific initial point we use color shades. Dark colors indicate rapid convergence while lighter ones indicate slow convergence. To this end, we utilize five different shades per color. Thus, reading a picture, we are able to see the regions of rapid convergence. Moreover, by observing that the colored zones are separated, we can easily answer

the question of whether a solution attracts the initial guesses close to it or not. Also, using the output data of this approach, it is possible to compute the radius of convergence of a method to a specific solution. The *radius of convergence* is the longest possible radius of the circle centered at the solution which contains only starting points which converge to this solution [Androulakis & Vrahatis, 1996; Drossos *et al.*, 1996].

In the paper at hand, we have implemented eleven algorithms with different characteristics from the aforementioned principal categories. From the second category, we have implemented Newton's and Broyden's methods while from the third category, we implemented all the algorithms described above. All the implementations were based on the Numerical Recipes source code with minimal modifications [Press *et al.*, 2007].

First, we employ the well known, $z^3 - 1 = 0$, equation for $z \in \mathbb{C}$ to visualize the convergence characteristics of the aforementioned methods. Specifically considering the corresponding real and imaginary parts yields the following system of equations:

$$\begin{cases} f_1(x_1, x_2) = \Re\{(x_1 + i x_2)^3 - 1\} \\ \quad = x_1^3 - 3x_1 x_2^2 - 1 = 0, \\ f_2(x_1, x_2) = \Im\{(x_1 + i x_2)^3 - 1\} \\ \quad = 3x_1^2 x_2 - x_2^3 = 0, \end{cases} \quad (7)$$

where $i = \sqrt{-1}$. This system has three solutions: $r_1 = (1, 0)^\top$, $r_2 = (-1/2, \sqrt{3}/2)^\top$ and $r_3 = (-1/2, -\sqrt{3}/2)^\top$.

We have generated starting points in the $[-2, 2]^2$ rectangle, with a dense grid of 1000 points per coordinate. For each starting point, we run a method until it either reaches the maximum number of 5000 function evaluations, or the distance of the approximate solution, x^* , from a solution, r , is less than 10^{-4} , i.e. $\|x^* - r\|_2 \leq 10^{-4}$.

Figure 1 illustrates the basins of convergence of Newton's and Broyden's method, left and right respectively, to the solutions r_1 colored in red, r_2 colored in green and r_3 colored in blue. Lighter colors represent more iterations to converge to a solution. We can observe that both methods exhibit large basins of fast convergence for each solution. Newton's method exhibits a fractal structure between solutions, while Broyden's method exhibits a more stable behavior in the corresponding areas.

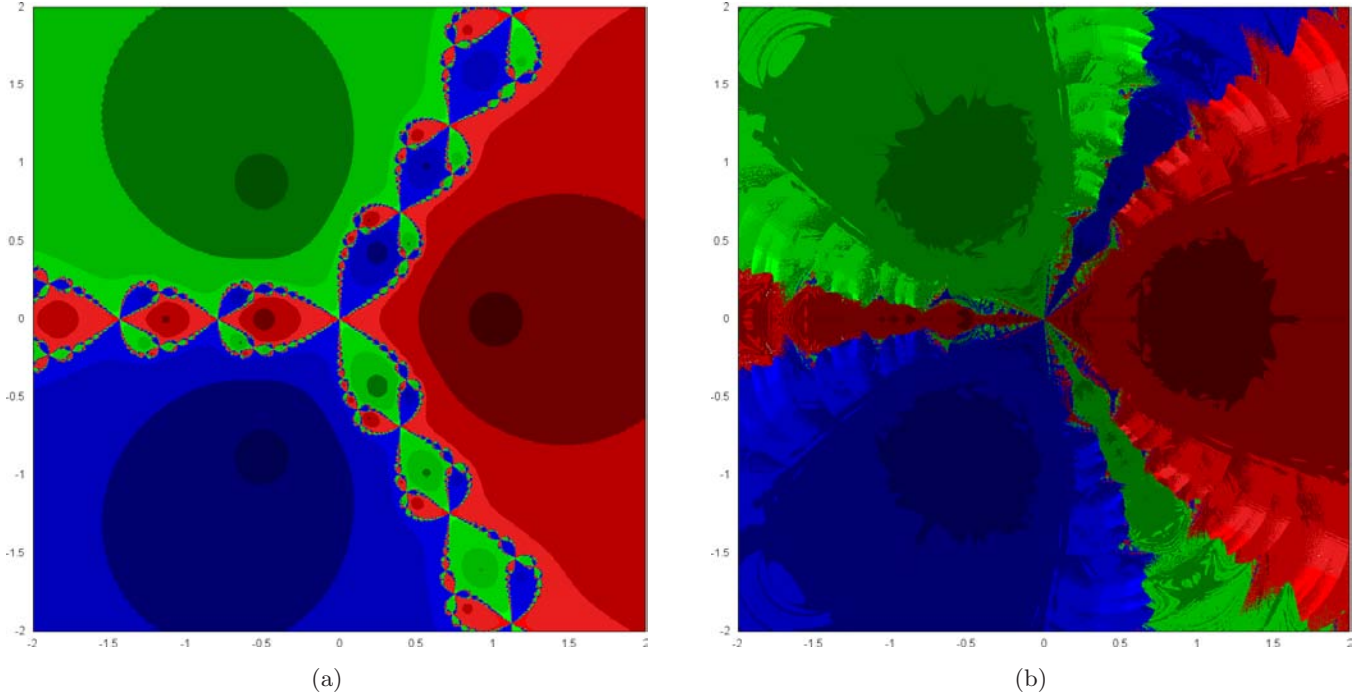


Fig. 1. Basins of convergence of (a) Newton's method and (b) Broyden's method for System (7).

For the same problem $z^3 - 1 = 0$, we also consider the following objective function:

$$\begin{aligned} f(x_1, x_2) &= (\Re\{(x_1 + ix_2)^3 - 1\})^2 \\ &\quad + (\Im\{(x_1 + ix_2)^3 - 1\})^2 \\ &= (x_1^3 - 3x_1x_2^2 - 1)^2 + (3x_1^2x_2 - x_2^3)^2. \end{aligned} \quad (8)$$

This function has three global minimizers: $r_1 = (1, 0)^\top$, $r_2 = (-1/2, \sqrt{3}/2)^\top$ and $r_3 = (-1/2, -\sqrt{3}/2)^\top$ with global minimum $f(x_1, x_2) = 0$, which are of course solutions of the equation $z^3 - 1 = 0$. A contour plot of this objective function is illustrated in Fig. 2. The shades of the colors ranging from deep blue to deep red with yellow in between correspond respectively to function values ranging from small to large.

Figure 3 illustrates the basins of convergence to the corresponding global minimizers (solutions) for the (a) Cauchy, (b) Armijo, and (c) Cauchy method with strong Wolfe conditions. It is clear that they exhibit a fast and robust convergence speed with well shaped basins in the whole area of the considered rectangle. The Cauchy method with either constant step (here 0.01) or with strong Wolfe conditions exhibits similar basins of convergence. The latter method converges in the whole rectangle area while the former does not converge close to the rectangle's corners. Furthermore, Armijo method

exhibits well shaped basins of attraction with fast convergence speed. It is worth noticing that the Armijo method converges in the almost whole area of the considered rectangle and does not exhibit any fractal structure.

Furthermore, Fig. 4 demonstrates the behavior of the (a) Fletcher–Reeves, (b) Polak–Ribière, (c) Hestenes–Stiefel and (d) Perry methods respectively. All methods behave in a similar manner and

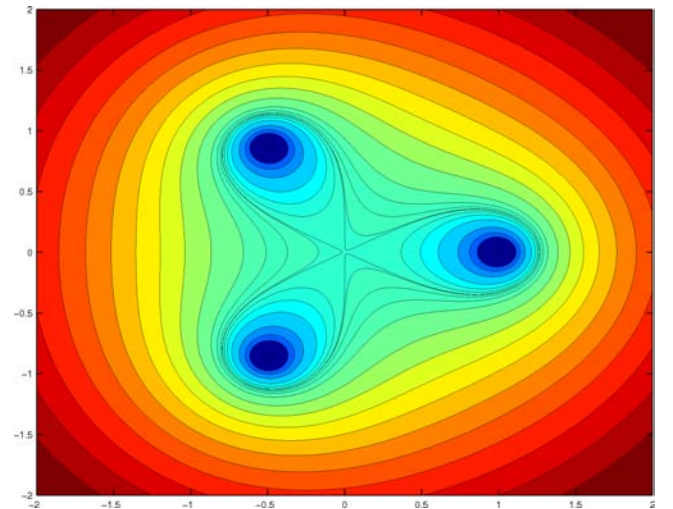


Fig. 2. Contour plot of the objective function of Eq. (8). The shades of the colors ranging from deep blue to deep red with yellow in between correspond respectively to function values ranging from small to large.

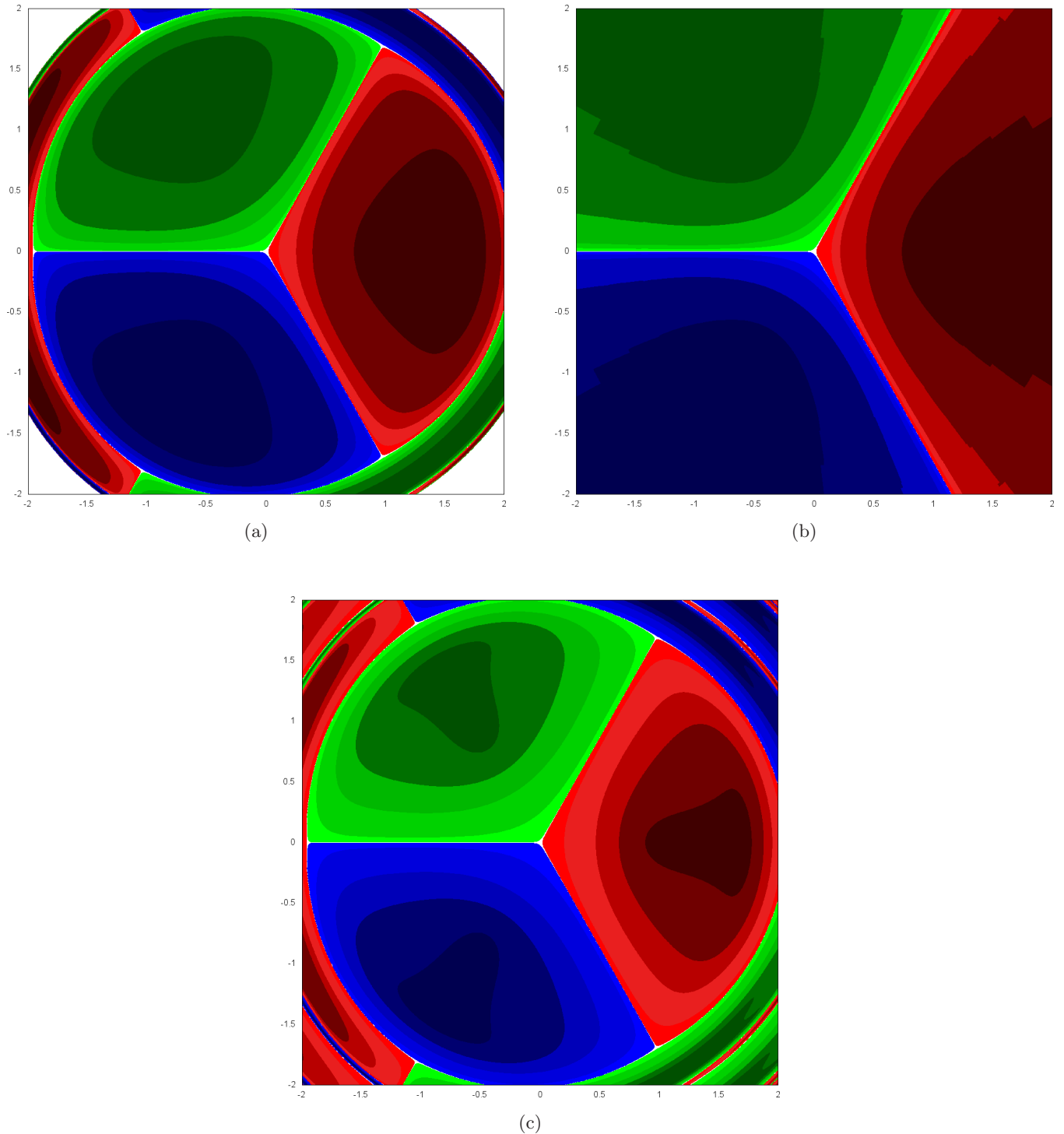


Fig. 3. Basins of convergence of (a) Cauchy's, (b) Armijo's and (c) Cauchy's with strong Wolfe conditions method for System (7).

exhibit basins of attraction with irregular shapes. More specifically, Fletcher–Reeves exhibits fast convergence speed in terms of function evaluations (darker colors), but for many starting points in the center of the rectangle, it does not converge to a solution (points in white). On the contrary,

the remaining three methods exhibit many areas with slower convergence speed but the number of starting points from which they do not converge decreases.

Finally, Fig. 5 exhibits the behavior of the DFP and BFGS methods respectively. Once again

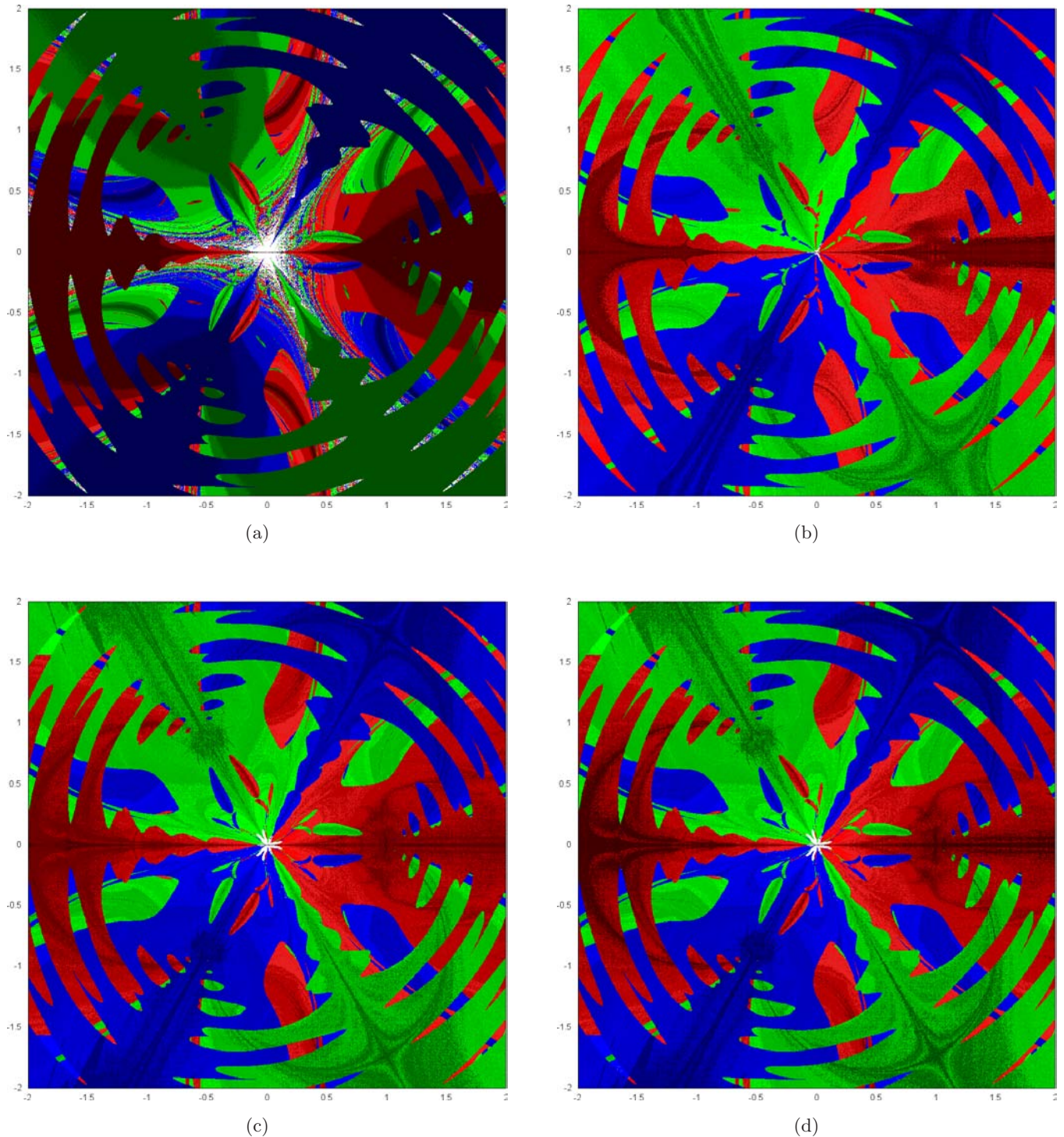


Fig. 4. Basins of convergence of (a) Fletcher-Reeves's, (b) Polak-Ribière's, (c) Hestenes-Stiefel's and (d) Perry's method for System (7).

both methods exhibit irregular shaped basins of convergence with good performance in terms of convergence speed (dark colors). Additionally, they illustrate larger basins of convergence than the latter four methods. BFGS exhibits more well shaped basins of convergence than DFP, while DFP does

not converge to a solution from several starting points.

Additionally, we examine the behavior of all the above methods in the presence of noise. More specifically, at each function evaluation, noise was added to the actual function values, for different

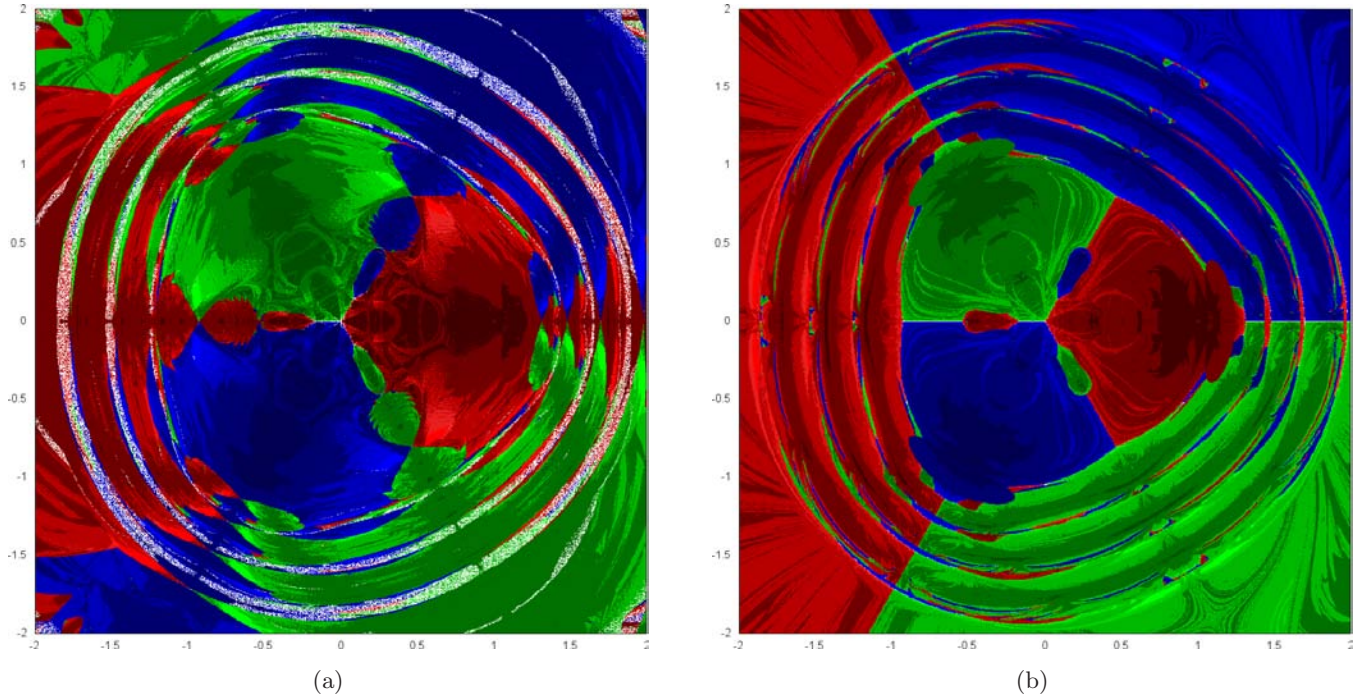


Fig. 5. Basins of convergence of (a) DFP method and (b) BFGS method for System (7).

values of the noise's strength (standard deviation). Here we utilize four different levels of noise strength, $\sigma = 0.04, \sigma = 0.10, \sigma = 0.20$, and $\sigma = 0.40$. Due to space limitations we will exhibit only the multiplicative noise case. In this case, experimental results indicate that in the presence of multiplicative noise, the most robust, stable and efficient method is the Armijo method.

To compare Armijo with the classical Newton method we illustrate in Figs. 6 and 7 their basins of convergence in the presence of the aforementioned noise strength levels. Clearly one can observe that the Armijo method can efficiently and rapidly converge to its closest solution in the presence of high noise strength levels e.g. when $\sigma = 0.40$. In contrast, at the same noise strength level Newton's efficiency declines. It either does not converge to a solution or converges very slowly (lighter colors with many white points).

From the other implemented algorithms we have to distinguish Cauchy with constant step size and Cauchy with the strong Wolfe conditions. Both exhibit a behavior similar to that of the Armijo method, and they can efficiently converge to solutions in the presence of high noise strength levels. The remaining methods exhibit good behavior only for low noise strength levels, $\sigma = 0.04$ and $\sigma = 0.10$. For higher noise strength levels, most of them do not manage to converge to a solution within the

prespecified maximum number of function evaluations. In particular, the Broyden method does not converge to solutions even for low noise strength values e.g. $\sigma = 0.04$. Indicatively, Figs. 8 and 9 exhibit their convergence basins for multiplicative noise with $\sigma = 0.04$.

As a second test case, we consider the quadratic area-preserving two-dimensional Hénon's mapping, and calculate for each method the basins of convergence for two prespecified periodic orbits. The two-dimensional Hénon's mapping can be defined as:

$$\Phi : \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \mathcal{R}(\omega) \begin{pmatrix} x_1 \\ x_2 + g(x_1) \end{pmatrix},$$

where $(x_1, x_2)^\top \in \mathbb{R}^2$ and

$$\mathcal{R}(\omega) = \begin{pmatrix} \cos \omega & -\sin \omega \\ \sin \omega & \cos \omega \end{pmatrix},$$

where $\omega \in [0, \pi]$ is the *rotation angle*. Here, we consider the Hénon mapping for $\cos \omega = 0.24$ and $g(x_1) = -x_1^2$. For this example, we can observe in the Hénon's mapping phase plot, illustrated in Fig. 10(a), that there is a chain of five islands around the center of the rectangle. The center points of each island contain a stable elliptic periodic orbit of period five ($p = 5$). Additionally, the five points where the islands connect consist

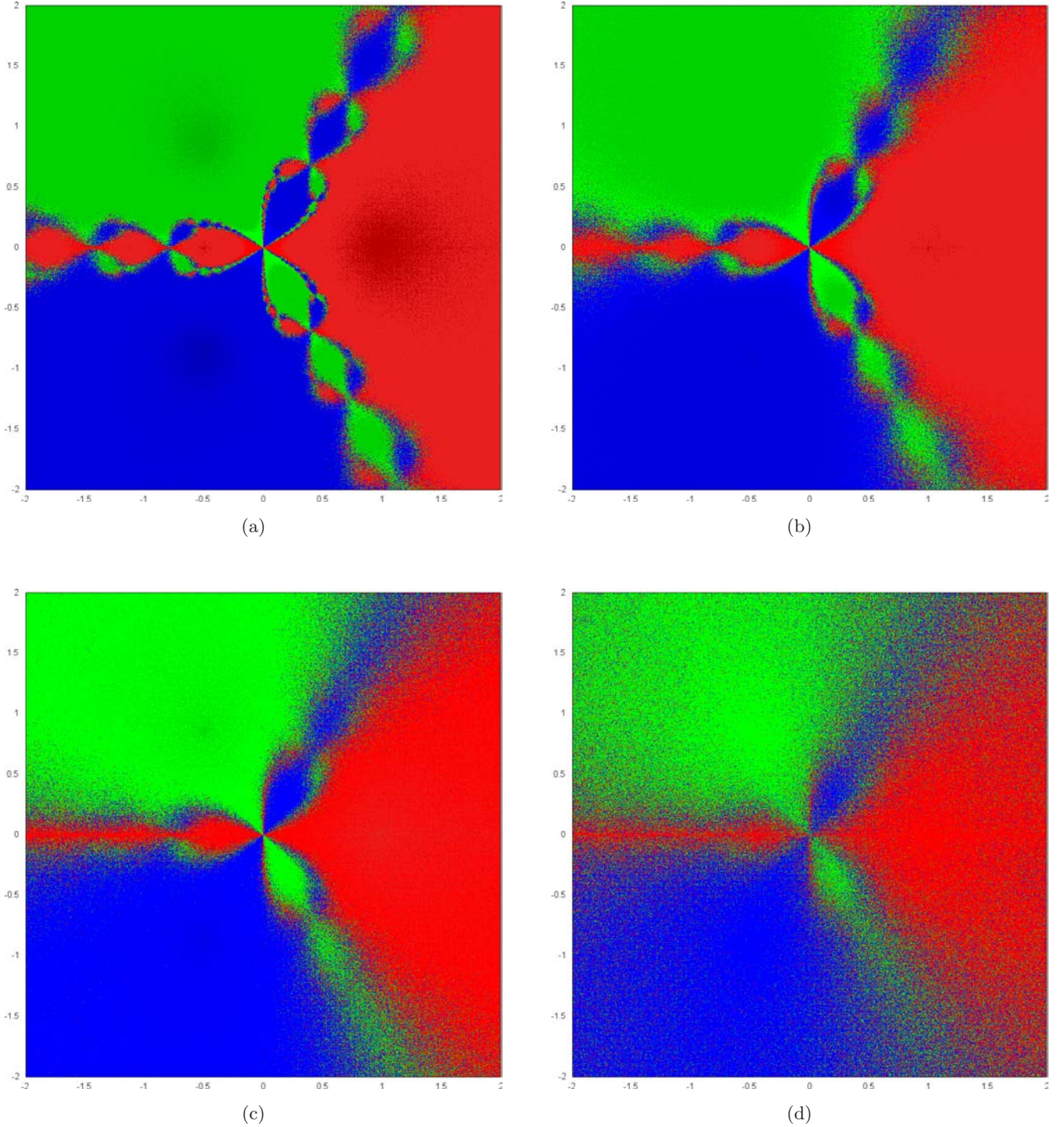


Fig. 6. Basins of convergence of Newton's method in the presence of noise at strength levels of (a) $\sigma = 0.04$, (b) $\sigma = 0.10$, (c) $\sigma = 0.20$ and (d) $\sigma = 0.40$ for System (7).

of an unstable hyperbolic periodic orbit of period five [Vrahatis, 1995]. As discussed previously, one can compute these points by applying a proper method from the aforementioned three categories. When a minimization method is applied, we have to create an objective function and calculate its

global minimizers [Parsopoulos & Vrahatis, 2004]. Thereby, in Fig. 10(b) we illustrate a contour plot of the Hénon mapping ($\cos \omega = 0.24$ and $g(x_1) = -x_1^2$) objective function for calculating periodic orbits of period $p = 5$. The shades of the colors ranging from deep blue to deep red with yellow in between

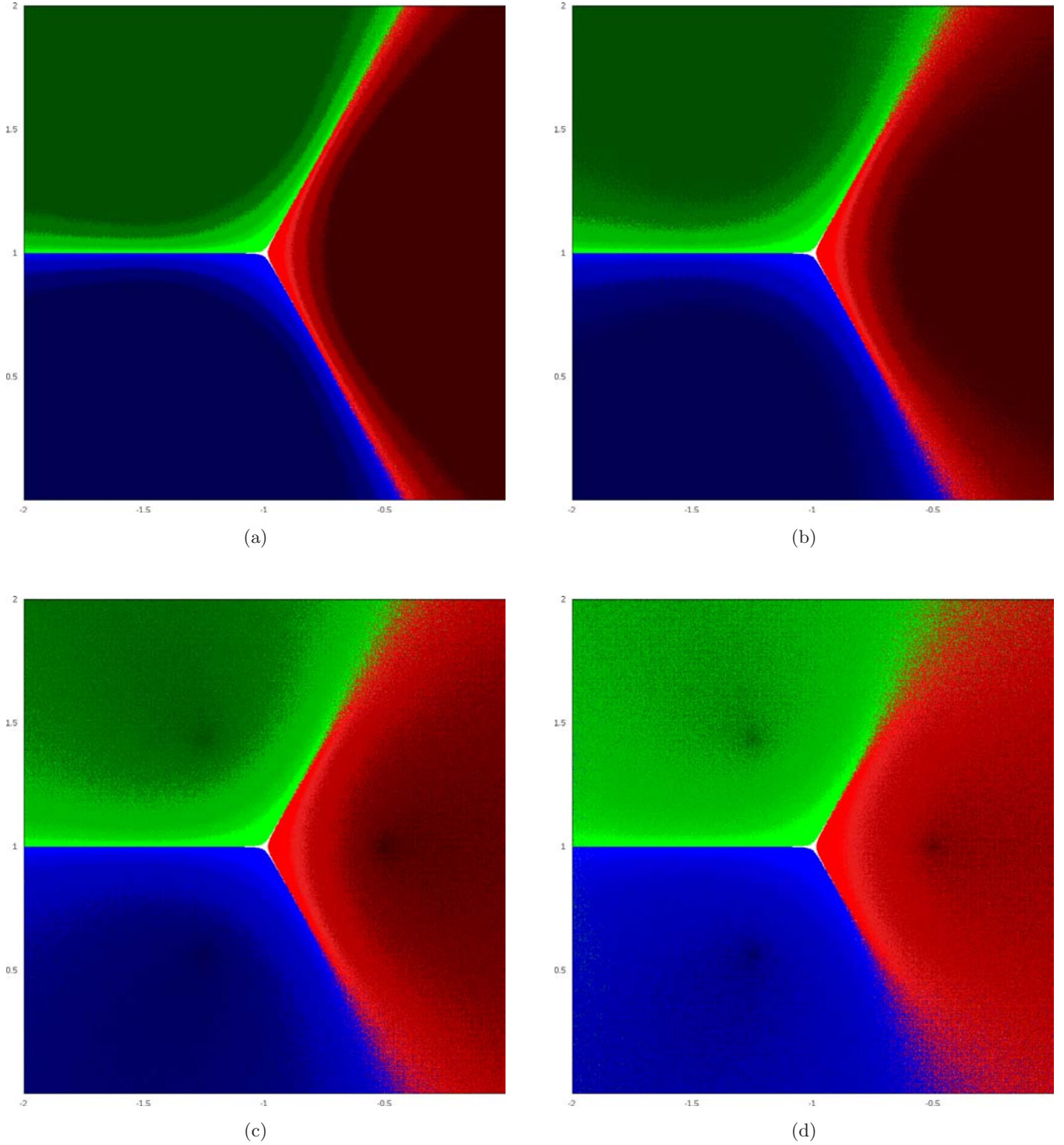


Fig. 7. Basins of convergence of Armijo's method in the presence of noise at strength levels of (a) $\sigma = 0.04$, (b) $\sigma = 0.10$, (c) $\sigma = 0.20$ and (d) $\sigma = 0.40$ for System (7).

correspond respectively to function values ranging from small to large. When we have computed one of these points, we can either subsequently apply the same algorithm with different starting conditions and find another point of the periodic orbit or we can iterate the mapping using one of

the computed points as starting point. To produce the stable periodic orbit, we can iterate the mapping using the following starting point: $(x_1, x_2)^\top = (0.5672405470221847, -0.1223202134278941)^\top$. The rotation number of this orbit is $\sigma = m_1/m_2 = 1/5$. It produces $m_2 = 5$ points by rotating around

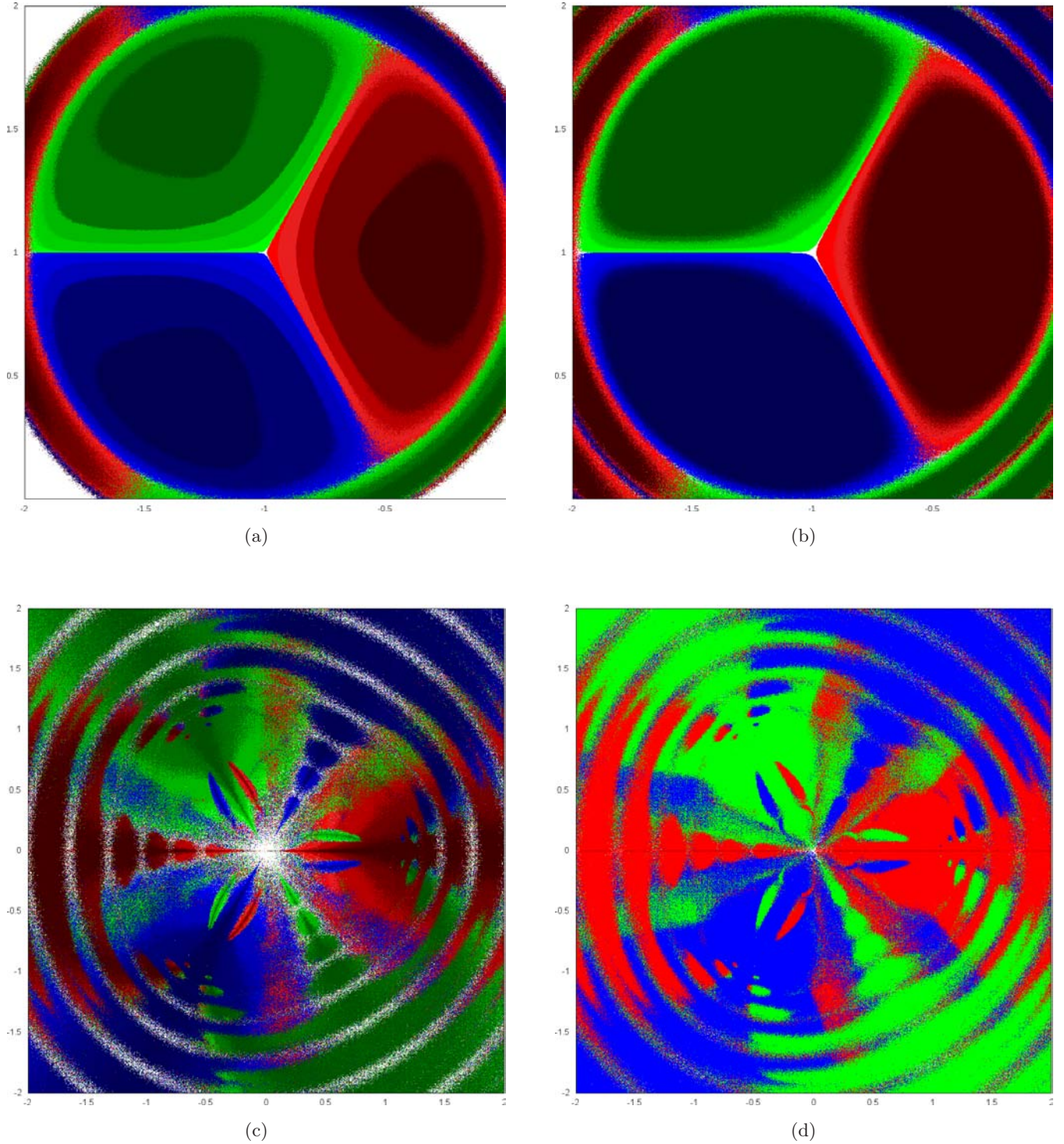


Fig. 8. Basins of convergence of (a) Cauchy's, (b) Cauchy's with strong Wolfe conditions, (c) Fletcher-Reeves's and (d) Polak-Ribière's method in the presence of noise at the strength level of $\sigma = 0.04$ for System (7).

the origin $m_1 = 1$ times. Additionally, to compute the unstable periodic orbit, one can iterate the mapping using at starting point $(x_1, x_2)^\top = (0.2942106885737921, -0.4274862418615337)^\top$ [Vrahatis, 1995].

Next, we demonstrate the basins of convergence for both stable and unstable periodic orbits. We have generated starting points in the $[-1, 1]^2$ rectangle, with a dense grid of 1000 points per coordinate. For each starting point, we execute a method

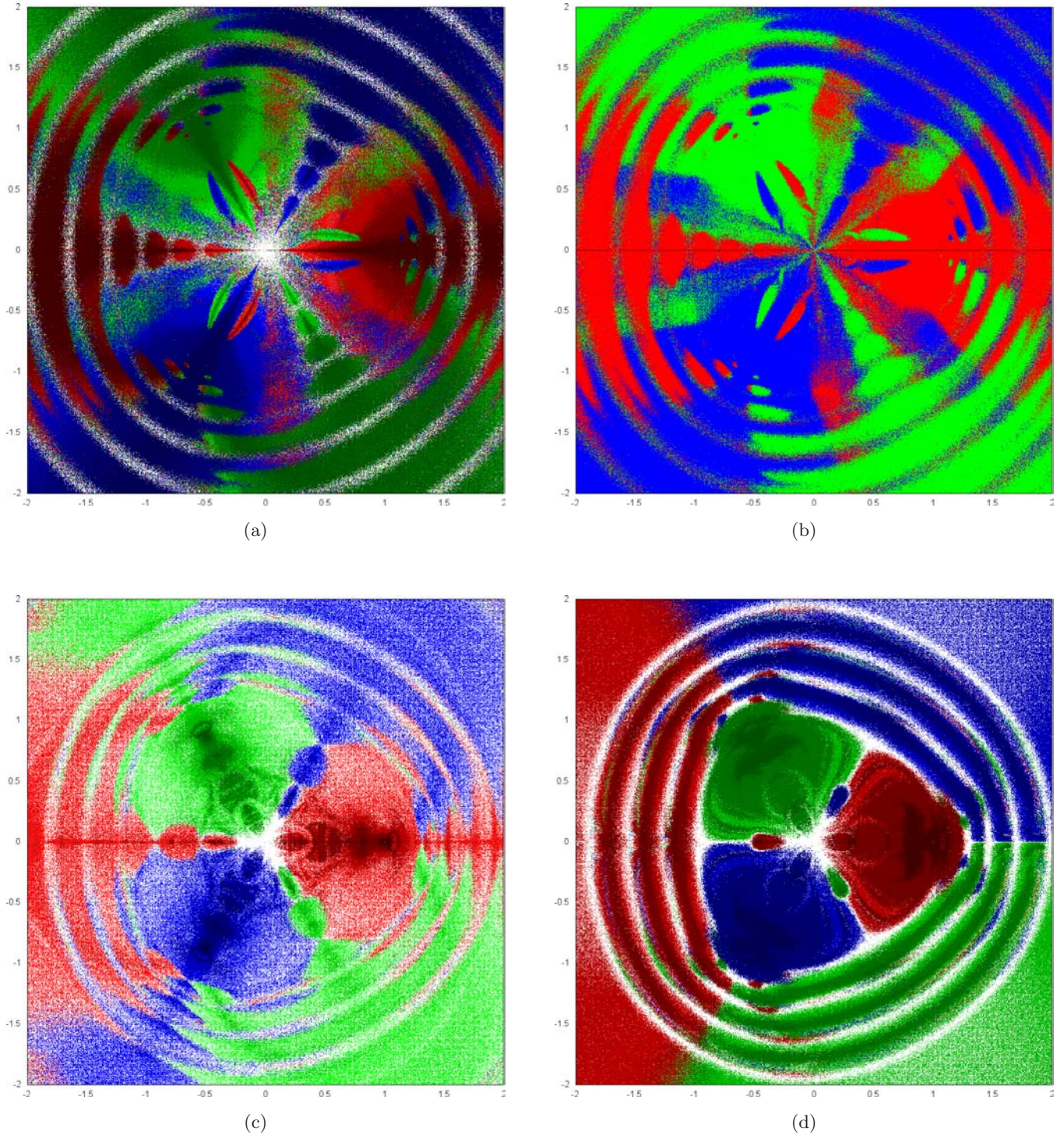


Fig. 9. Basins of convergence of (a) Hestenes–Stiefel's, (b) Perry's, (c) DFP method and (d) BFGS method in the presence of noise at the strength level of $\sigma = 0.04$ for System (7).

until it either reaches the maximum number of 5000 function evaluations, or the distance between it and the approximate solution, x^* , is less than 10^{-4} , i.e. $\|x^* - x^p\|_\infty \leq 10^{-4}$. Each periodic orbit consists of five different points, thus each figure below will consist of ten different colors one for each

point. Additionally, the white color indicates that the corresponding starting point does not converge to any of the ten points within the aforementioned criteria.

Figure 11 illustrates the basin of convergence for Newton's and Broyden's methods. It is clear

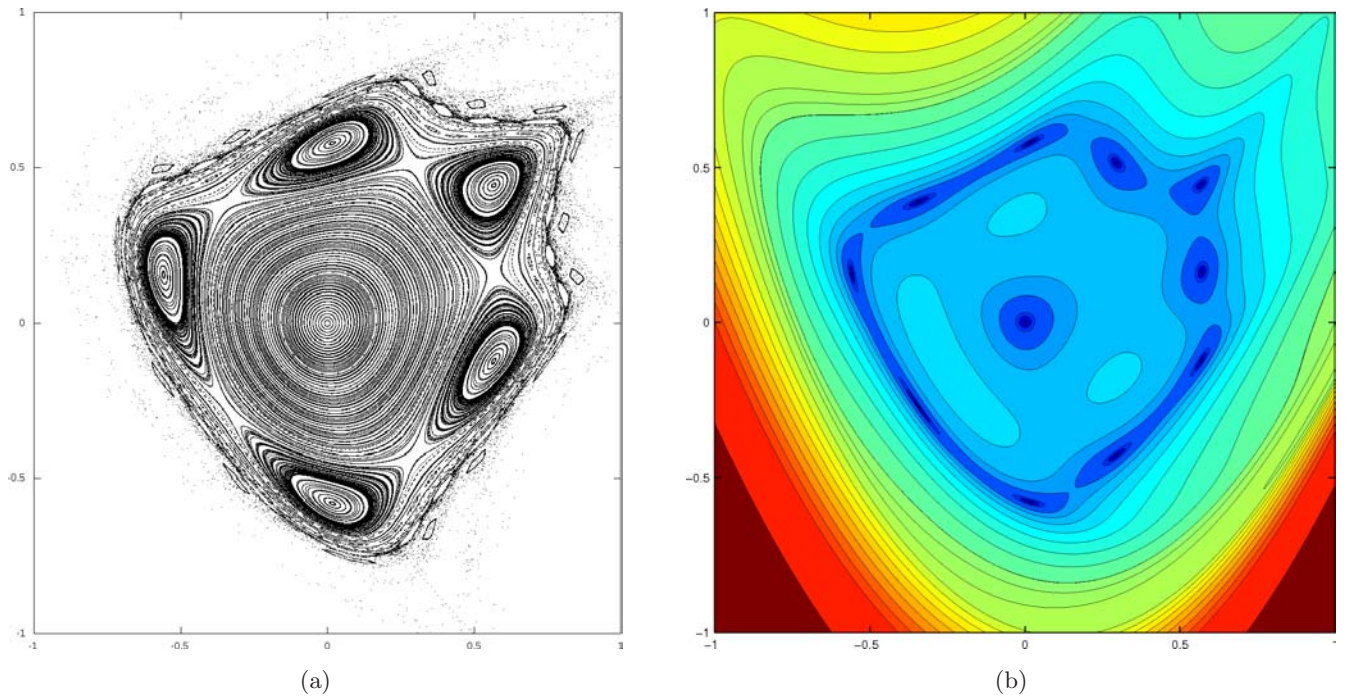


Fig. 10. (a) Hénon mapping for $\cos \omega = 0.24$ and $g(x_1) = -x_1^2$. (b) Contour plot of the Hénon mapping objective function for calculating periodic orbits with period $p = 5$, ($\cos \omega = 0.24$ and $g(x_1) = -x_1^2$).

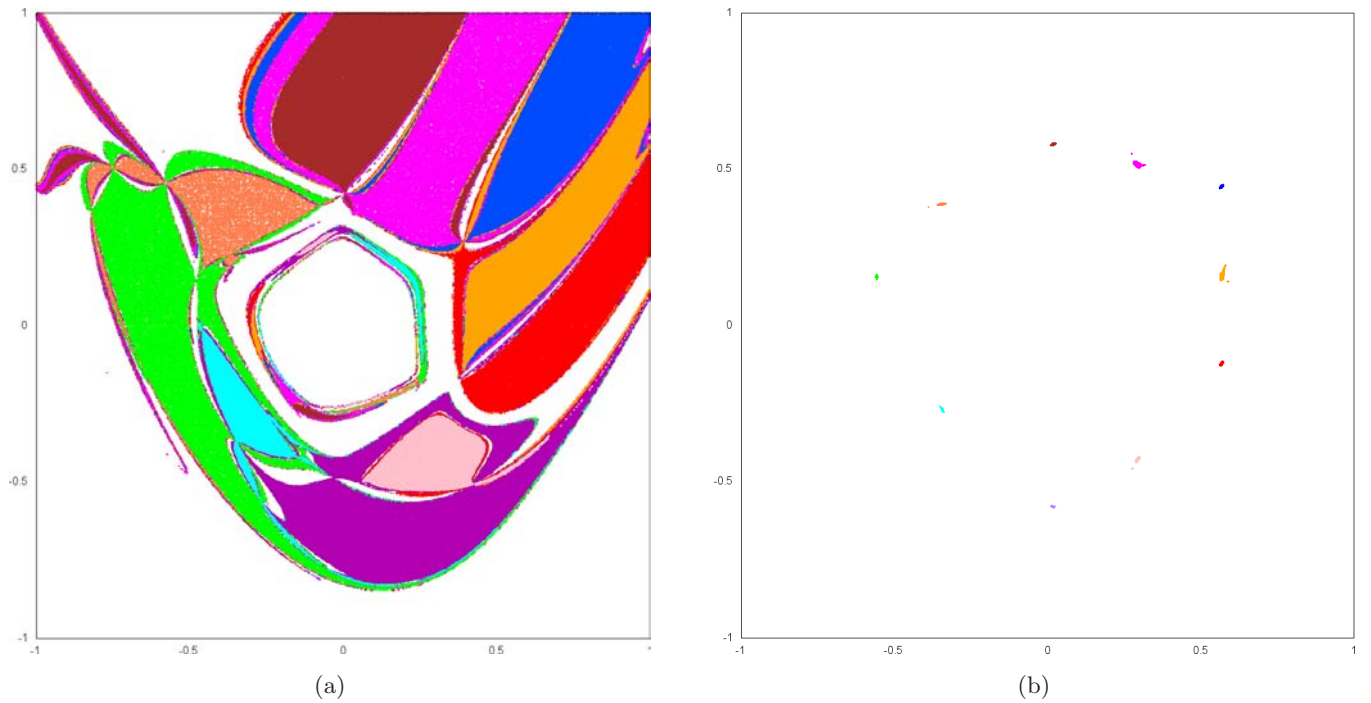


Fig. 11. Basins of convergence of (a) Newton's method and (b) Broyden's method for computing the stable (sp) and unstable (up) periodic orbits. Fixed points colors: sp (0.5672, -0.1223), sp (0.5672, 0.4440), sp (0.0173, 0.5800), sp (-0.5585, 0.1560), sp (0.0173, -0.5797), up (0.2942, -0.4274), up (0.5696, 0.1622), up (0.2942, 0.5140), up (-0.3443, 0.3882), up (-0.3443, -0.2696).

that Newton's method produces several wide and irregular shaped basins of attraction for each point of the periodic orbits, while Broyden's method converges only in small areas around each point of the periodic orbits.

Figure 12 exhibits the basin of convergence for the Cauchy with constant step size (a), the Armijo (b), and the Cauchy with strong Wolfe conditions method (c). Here we can observe that when the methods converge to a point, they produce

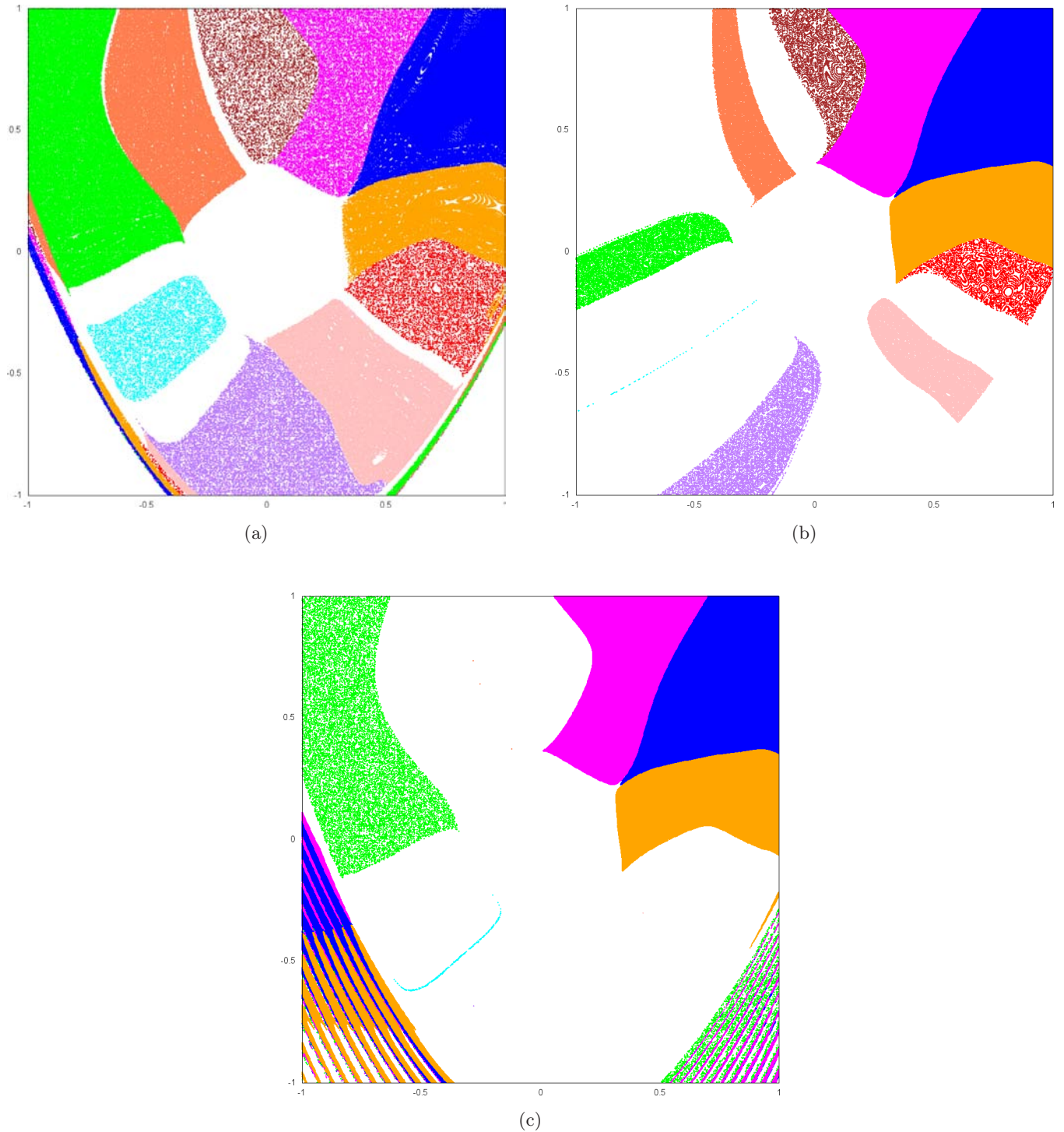


Fig. 12. Basins of convergence of (a) Cauchy's, (b) Armijo's and (c) Cauchy's with strong Wolfe conditions method for computing the stable (sp) and unstable (up) periodic orbits. Fixed points colors: sp (0.5672, -0.1223), sp (0.5672, 0.4440), sp (0.0173, 0.5800), sp (-0.5585, 0.1560), sp (0.0173, -0.5797), up (0.2942, -0.4274), up (0.5696, 0.1622), up (0.2942, 0.5140), up (-0.3443, 0.3882), up (-0.3443, -0.2696).

wide basins of convergence. By comparing the three methods, one could suggest that Cauchy's method with constant step size exhibits the most robust performance, since it converges to a periodic point from most of the starting points of the considered

area. Although, there are several starting points within Cauchy's basins from which the method does not converge to any point of the considered periodic orbits. This may be due to a low budget of function evaluations. By increasing the budget of

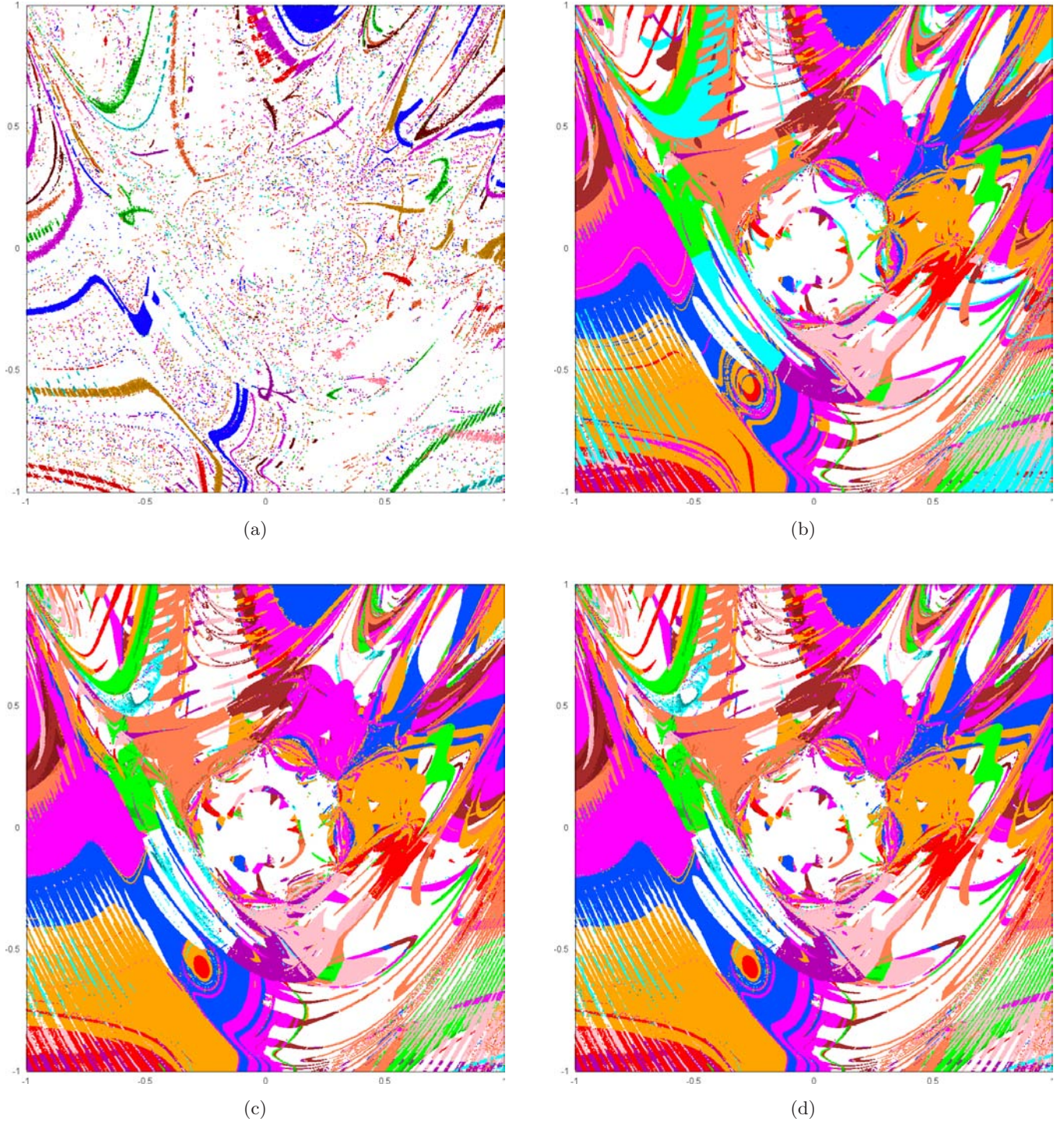


Fig. 13. Basins of convergence of (a) Fletcher-Reeves's, (b) Polak-Ribière's, (c) Hestenes-Stiefel's and (d) Perry's method for computing the stable (sp) and unstable (up) periodic orbits. Fixed points colors: sp (0.5672, -0.1223), sp (0.5672, 0.4440), sp (0.0173, 0.5800), sp (-0.5585, 0.1560), sp (0.0173, -0.5797), up (0.2942, -0.4274), up (0.5696, 0.1622), up (0.2942, 0.5140), up (-0.3443, 0.3882), up (-0.3443, -0.2696).

function evaluations and selecting a proper step size for the problem at hand it may produce more dense and wide basins of convergence. The same configuration settings can be employed to both Armijo's and Cauchy's methods with strong Wolfe conditions to enhance their convergence performance. Both methods produce more robust and dense convergence basins but they do not converge from all starting points to the considered periodic orbits. A much larger budget of function evaluations should enhance their convergence performance and produce wide regions of convergence from all starting points, since both the Armijo rule and the strong Wolfe conditions require a high number of function evaluations.

Figures 13 and 14 illustrate the convergence characteristics of the remaining six optimization algorithms. As it can be observed, they produce very unstable basins of convergence with many irregular shapes without any distinguishing pattern. Specifically, Fletcher-Reeves method does not converge to a periodic orbit point from the majority of the starting points. On the contrary, the rest of the algorithms converge fast to a periodic orbit point in a very unstable manner. From all considered algorithms in these two figures, the BFGS method exhibits the most promising

performance, since it produces the most wide basins of convergence.

In the presence of noise, almost all methods in this example either do not converge to a periodic orbit point, or do not exhibit a robust basin of convergence for any one of the points. Indicatively, here we provide convergence plots for the four best methods and only for the strength noise level of $\sigma = 0.04$. Figure 15 illustrates the basins of convergence in the presence of noise for (a) Newton, (b) Armijo, (c) Hestenes-Stiefel and (d) Fletcher-Reeves method. As it can be observed, Newton maintains most of its initial basin's shapes, Armijo can accurately locate the fixed points only in a small basin around the corresponding periodic orbit point, while Hestenes-Stiefel and Fletcher-Reeves converge to a periodic orbit point without any robust, or stable manner. In the presence of a higher strength noise level, the majority of the considered algorithms do not exhibit any well shaped basin of convergence.

Remark 4.1. Although we have used here the two-dimensional Hénon map for illustration, a large variety of mappings can be used, including the *Beam-Beam Map*, the *Standard Map*, the *Gingerbreadman Map*, the *Predator-Prey*, as well as, higher dimensional maps including the *Lorenz Map*,

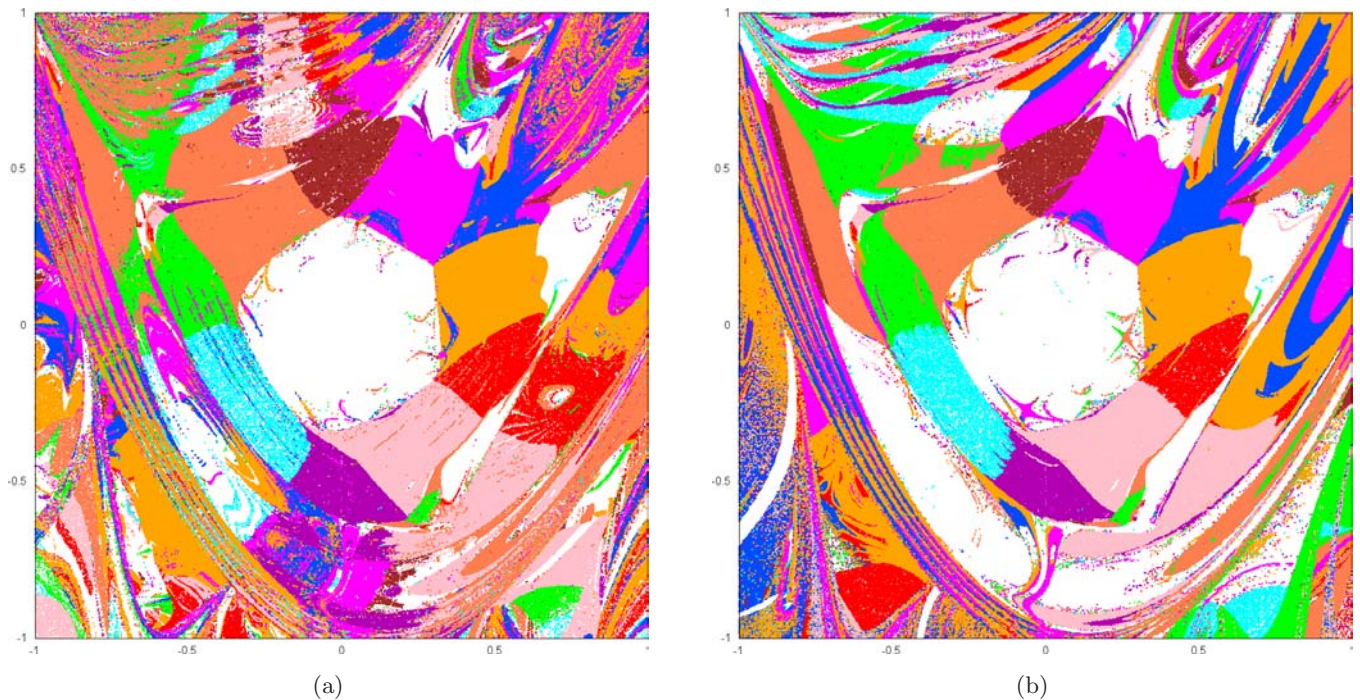


Fig. 14. Basins of convergence of (a) DFP method and (b) BFGS method for computing the stable (sp) and unstable (up) periodic orbits. Fixed points colors: sp (0.5672, -0.1223), sp (0.5672, 0.4440), sp (0.0173, 0.5800), sp (-0.5585, 0.1560), sp (0.0173, -0.5797), up (0.2942, -0.4274), up (0.5696, 0.1622), up (0.2942, 0.5140), up (-0.3443, 0.3882), up (-0.3443, -0.2696).

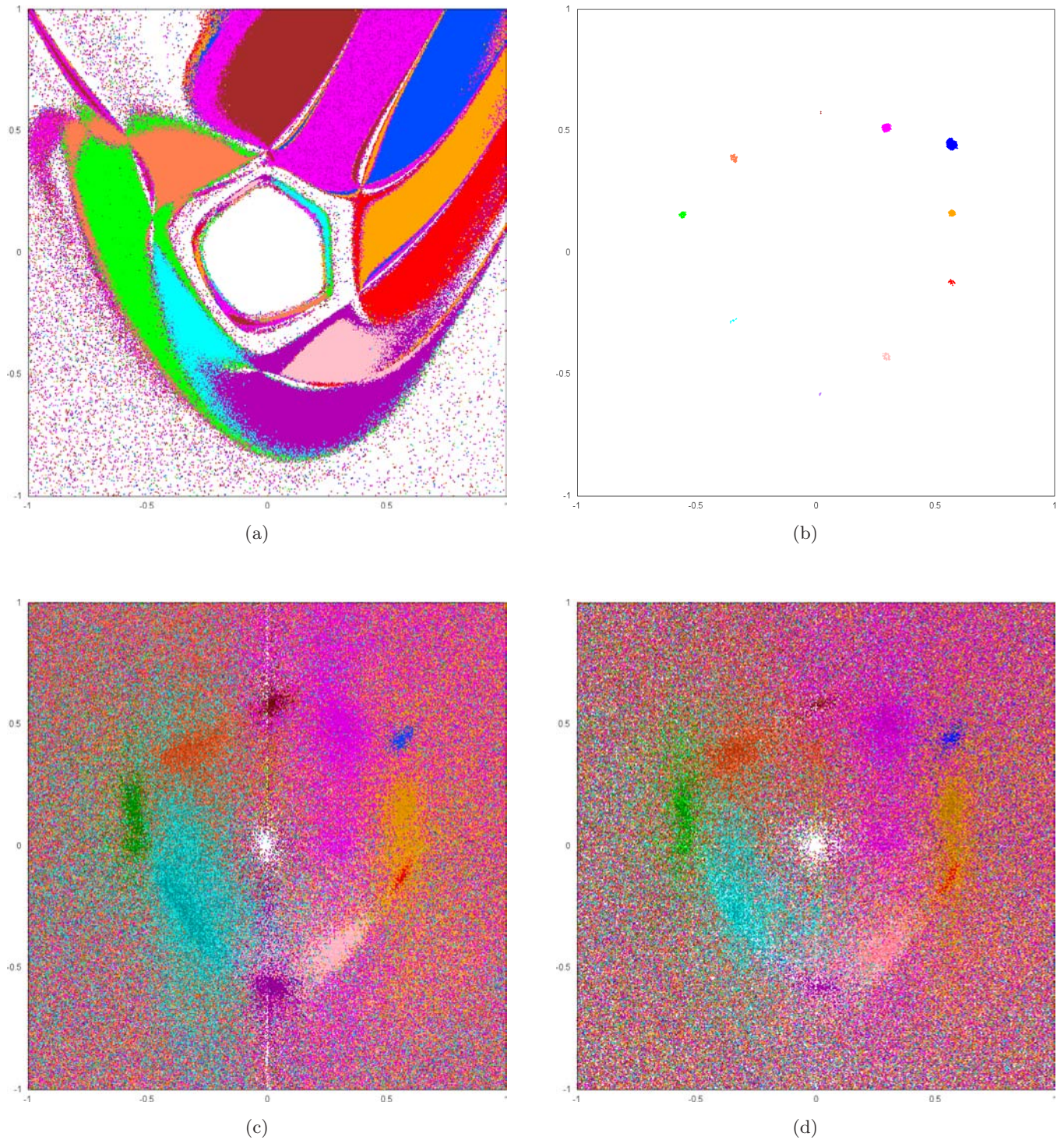


Fig. 15. Basins of convergence of (a) Newton's, (b) Armijo's, (c) Hestenes-Stiefel's and (d) Fletcher-Reeves's method for computing the stable (sp) and unstable (up) periodic orbits in the presence of noise with $\sigma = 0.04$. Fixed points colors: sp (0.5672, -0.1223), sp (0.5672, 0.4440), sp (0.0173, 0.5800), sp (-0.5585, 0.1560), sp (0.0173, -0.5797), up (0.2942, -0.4274), up (0.5696, 0.1622), up (0.2942, 0.5140), up (-0.3443, 0.3882), up (-0.3443, -0.2696).

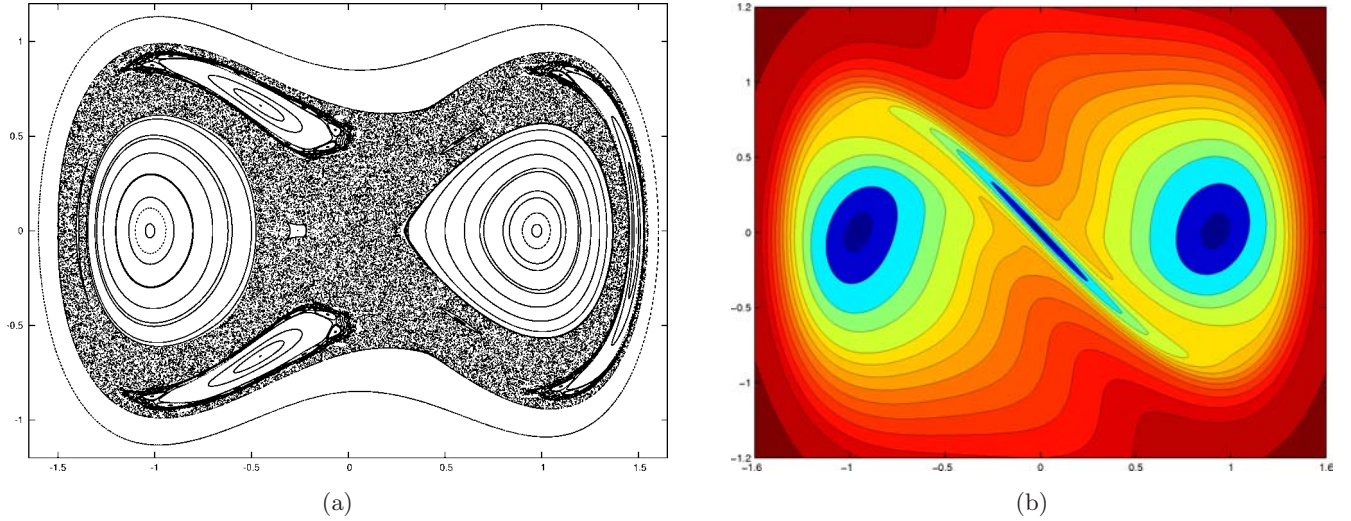


Fig. 16. (a) A Poincaré surface section of Duffing's oscillator for $\alpha = 0.05$ and $\beta = 2$. (b) Contour plot of the Poincaré surface section of Duffing's oscillator objective function for calculating periodic orbits with period $p = 1$, ($\alpha = 0.05, \beta = 2$).

the *Rössler Map* and *Hénon's 4-Dimensional Symplectic Map*, among others.

Remark 4.2. For a visualization in higher dimensions, we can consider a bounded domain of starting points of the two-dimensional subspace \mathcal{E}^2 of \mathbb{R}^n spanned by the following two vectors: $\{e^{\max}, e^{\min}\}$, where e^{\max}, e^{\min} are the eigenvectors corresponding to the extreme eigenvalues of the Hessian of the objective function f at a minimum $x^*, \nabla^2 f(x^*)$. More specifically, we can apply the aforementioned algorithms for points $x = x^* + y \in \mathbb{R}^n$, with $y = c_1 e^{\max} + c_2 e^{\min}$, $c_1, c_2 \in \mathbb{R}$, by taking a grid of the values of c_1 and c_2 which determine the coordinates

of a pixel for illustration. Since $\nabla^2 f$ is a real and symmetric matrix, all eigenvalues and eigenvectors are real and the eigenvectors corresponding to distinct eigenvalues are orthogonal. Thereby the values of c_1 and c_2 taken by a grid into an orthogonal parallelepiped represent an orthogonal parallelepiped subset of \mathcal{E}^2 . The main reason for the choice of this two-dimensional subspace is that it reveals useful information. It is well known that, studying the sensitivity of the solution to small changes in a sufficiently small neighborhood of x^* , the directions of the principal axes of the elliptical contours (n -dimensional ellipsoids) will be given by the eigenvectors of $\nabla^2 f$, while the lengths of the axes will be

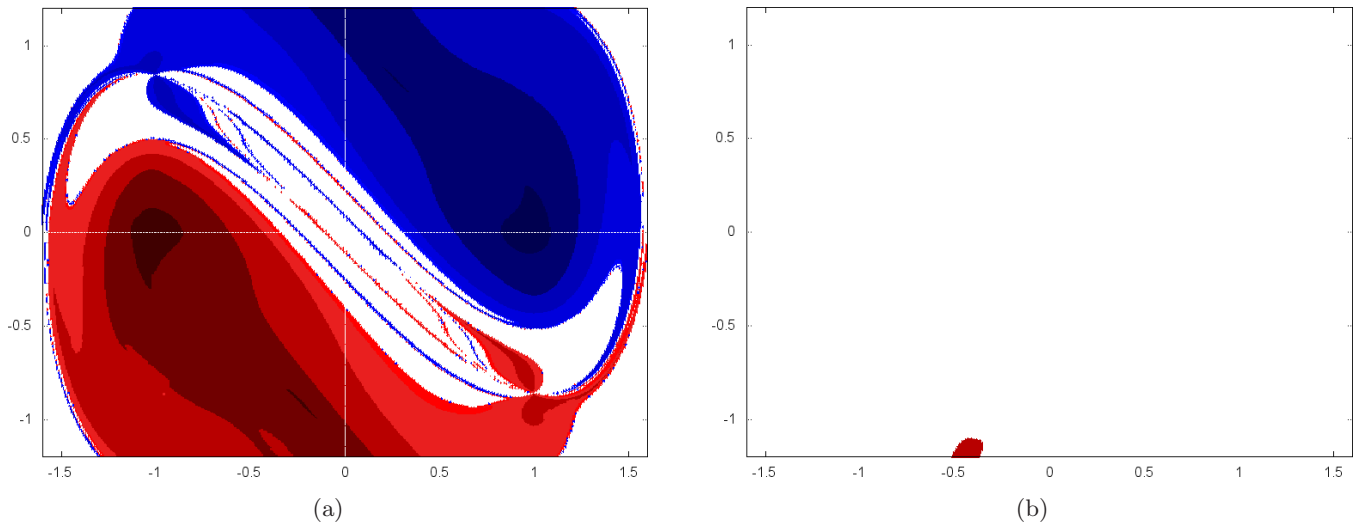


Fig. 17. Basins of convergence of (a) Newton's method and (b) Broyden's method for computing periodic orbits with period $p = 1$, of the Poincaré surface section of Duffing's oscillator ($\alpha = 0.05, \beta = 2$). Fixed points colors: red color for $(-1.0245, 0.0)$ fixed point, blue color for $(0.9746, 0.0)$ fixed point, white color for no convergence.

inversely proportional to the square roots of the corresponding eigenvalues. Hence, a variation along the eigenvector corresponding to the maximum eigenvalue will cause the largest change in f , while the eigenvector corresponding to the minimum eigenvalue yields the least sensitive direction. Thus, we can apply the aforementioned algorithms to this subspace and study their behavior for various directions, including the “extreme ones”. For more information about the visualization procedure in higher dimensions we refer the interested reader to [Androulakis & Vrahatis, 1996; Androulakis *et al.*, 1997; Parsopoulos *et al.*, 2008].

Remark 4.3. In the case of the nonexistence of derivatives or poorly behaved partial derivatives, various derivative-free methods can be used (see

for example [Parsopoulos & Vrahatis, 2003; Giorgi *et al.*, 2004; Conn *et al.*, 2009]).

The experimental analysis continues with the third and final test case. In this case, we consider a Poincaré surface of section for the conservative Duffing’s oscillator and calculate for each method the basins of convergence for two prespecified periodic orbits. More specifically, the conservative Duffing’s oscillator [Drossos *et al.*, 1996] can be described by the following equation:

$$\ddot{x} = x - x^3 + \alpha \cos \beta t,$$

which can be written as:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_1 - x_1^3 + \alpha \cos \beta t. \end{cases}$$

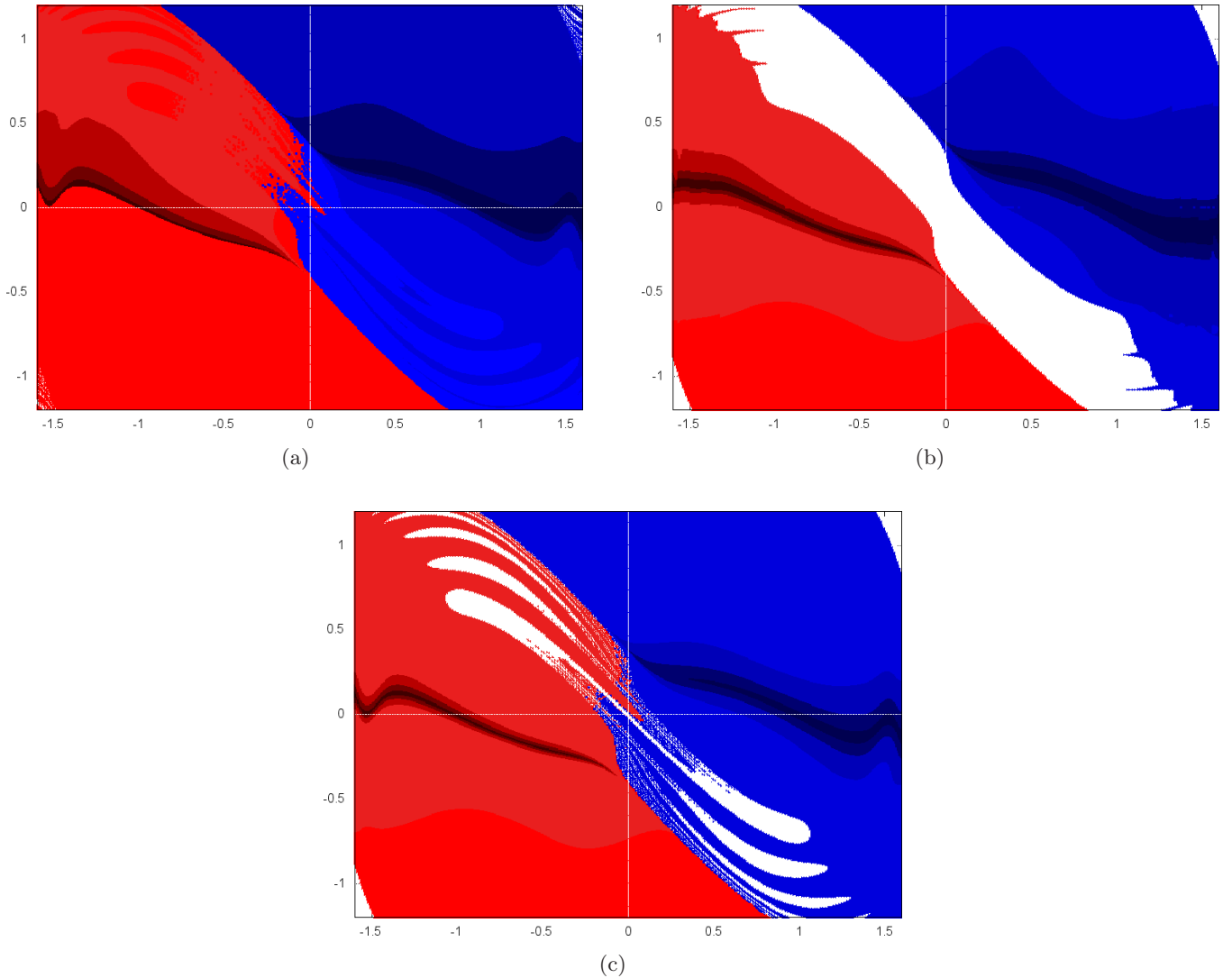


Fig. 18. Basins of convergence of (a) Cauchy’s, (b) Armijo’s, and (c) Cauchy’s with strong Wolfe conditions method for computing periodic orbits with period $p = 1$, of the Poincaré surface section of Duffing’s oscillator ($\alpha = 0.05, \beta = 2$). Fixed points colors: red color for $(-1.0245, 0.0)$ fixed point, blue color for $(0.9746, 0.0)$ fixed point, white color for no convergence.

For the aforementioned dynamical system, we consider the Poincaré surface of section for the parameter values of $\alpha = 0.05$ and $\beta = 2$. Figure 16(a) illustrates the phase plot of this surface, in the $[-1.6, 1.6] \times [-1.2, 1.2]$ rectangle, while Fig. 16(b) illustrates a contour plot of the Poincaré surface of section of Duffing's oscillator objective function for calculating periodic orbits with period $p = 1$. The shades of the colors ranging from deep blue to deep red with yellow in between correspond respectively to function values ranging from small to large. For this example, we can observe two distinct islands along the $x_2 = 0$ axis. The center points of each island correspond to fixed points of period one ($p = 1$). Once again we can easily compute these two points by applying

a proper method from the aforementioned three categories. The two center points correspond to $r_1 = (x_1, x_2)^\top = (-1.024572461190486, 0.0)^\top$, and $r_2 = (x_1, x_2)^\top = (0.9746253482044169, 0.0)^\top$. To this end, we exhibit the basins of convergence for these two fixed points in the $[-1.6, 1.6] \times [-1.2, 1.2]$ rectangle. In this example, the utilized grid of starting points consists of 500 points per coordinate. For each starting point, we run a method until it either reaches the maximum number of 5000 function evaluations, or the distance between a fixed point r and the approximate solution, x^* , is less than 10^{-3} , i.e. $\|x^* - r\|_2 \leq 10^{-3}$. Here, the white color depicting that for the corresponding starting point the method did not converge to any of the two fixed points within the aforementioned criteria.

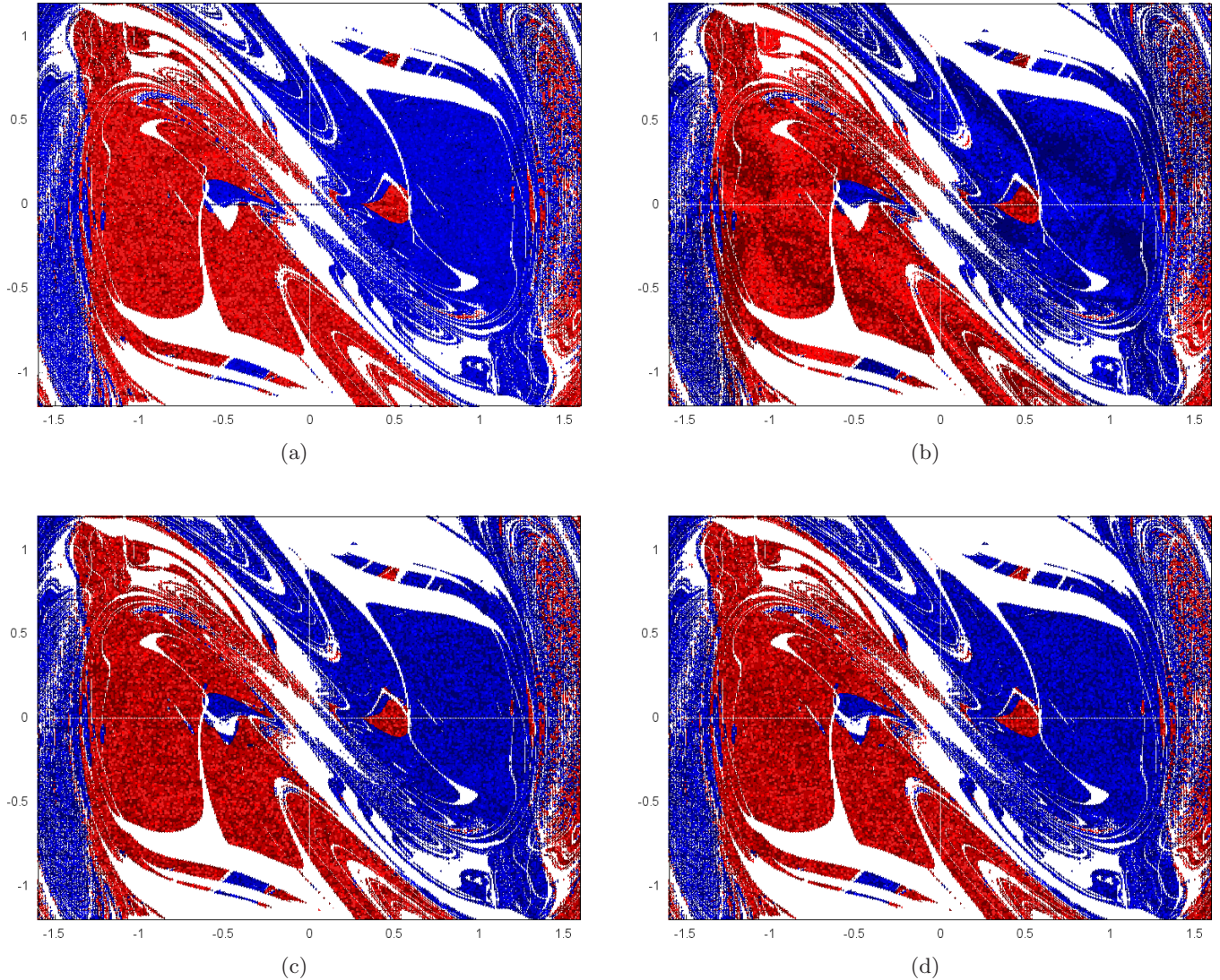


Fig. 19. Basins of convergence of (a) Fletcher–Reeves's, (b) Polak–Ribière's, (c) Hestenes–Stiefel's, and (d) Perry's method for computing periodic orbits with period $p = 1$, of the Poincaré surface section of Duffing's oscillator ($\alpha = 0.05, \beta = 2$). Fixed points colors: red color for $(-1.0245, 0.0)$ fixed point, blue color for $(0.9746, 0.0)$ fixed point, white color for no convergence.

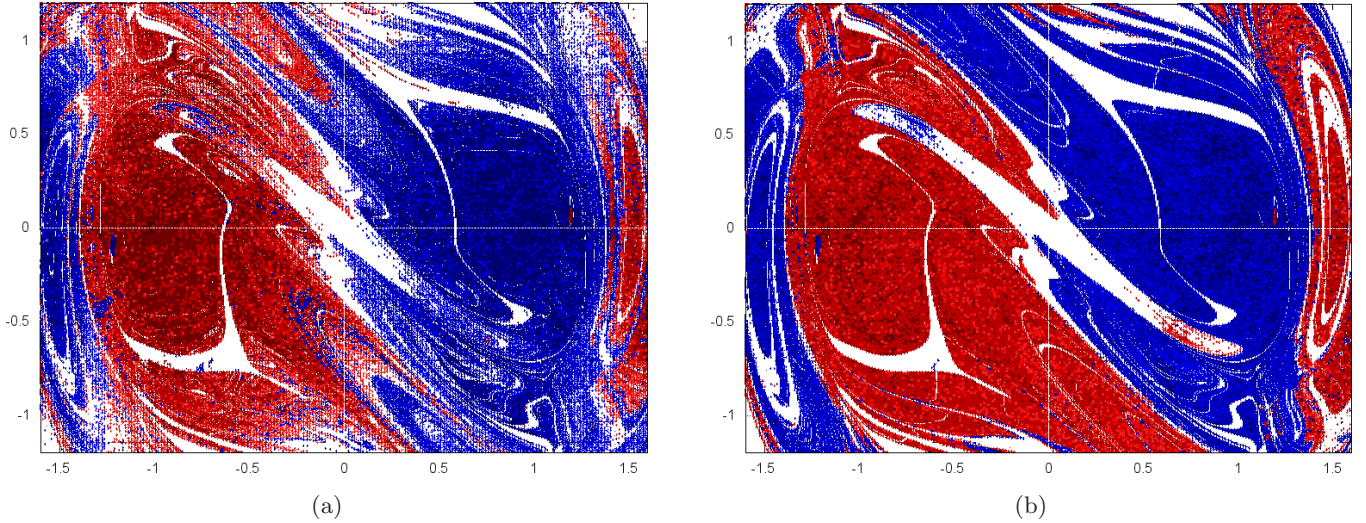


Fig. 20. Basins of convergence of (a) DFP method, and (b) BFGS method for computing periodic orbits with period $p = 1$, of the Poincaré surface section of Duffing's oscillator ($\alpha = 0.05, \beta = 2$). Fixed points colors: red color for $(-1.0245, 0.0)$ fixed point, blue color for $(0.9746, 0.0)$ fixed point, white color for no convergence.

Figures 17–20 demonstrate the basins of convergence to the fixed points for all the considered methods. It can be observed that the Newton, as well as the Cauchy, the Armijo and the Cauchy with strong Wolfe conditions method exhibit stable and robust performance in terms of wide basins of convergence. The Cauchy method with constant step-size, $\lambda_k = 0.01$, yielded the most stable and robust performance, as it converges to the fixed points from almost every starting point in the considered area.

More specifically, Fig. 17 illustrates the basins of convergence of Newton's and Broyden's methods, left and right respectively. Newton's method produces two wide and stable basins of convergence in about two thirds of the examined area, while in the center of the figure there is a wide area where Newton's method does not converge to a fixed point. On the contrary, Broyden's method does not converge to any of the fixed points in the majority of the considered area. In Fig. 18, it can be observed that Armijo's method and Cauchy with strong Wolfe conditions method behave similarly, and produce wide and stable basins of convergence. Additionally, Cauchy's method with constant stepsize (here $\lambda_k = 0.01$) converges to the fixed points in almost all of the considered area. There is a limited number of initial points at the top right and at the bottom left corner from which Cauchy's method cannot converge to any of the two fixed points.

Furthermore, Fig. 19 illustrates the basins of convergence for the Fletcher–Reeves, the Polak–Ribière, the Hestenes–Stiefel, and the Perry

methods. As in the previous examples, all four methods converge to the two fixed points for the majority of the starting points but not in a stable manner, since they produce wide basins of convergence with irregular shapes, that contain many discontinuous areas within their basins. Among the four methods of Fig. 19, we can observe that the Polak–Ribière, the Hestenes–Stiefel, and the Perry method, exhibit more dense convergence areas. Also, the Polak–Ribière method converges faster than the other three methods (many points with darker colors). Figure 20 exhibits the basins of convergence for the DFP and BFGS methods. Once more, although they produce wide basins of convergence, all basins have irregular shapes with many discontinuous areas within their basins. Finally, for this example in the presence of noise the majority of the tested algorithms did not converge to any of the fixed points. We therefore do not provide figures for this case. The only exception was Newton's method which exhibited some convergence areas for the lowest noise strength level, $\sigma = 0.04$.

5. Synopsis

In summary, we have presented an approach to distinguish reliable and robust methods for tackling a specific problem. To this end, we construct and display the “geometry” of the corresponding basins of convergence of the considered methods at the specific solution. By studying the structure of these basins of convergence, we are able to answer

the following questions that, in general, cannot be addressed more satisfactorily by theory, namely:

- (1) The question of whether for a given initial guess the method converges to the closest solution.
- (2) How fast does the method converge to different solutions?
- (3) How sensitive (stable) is the process to changes of the function or the initial data?
- (4) How is the performance of the method affected by the presence of noise (impact of imprecise information regarding the values of the function)?

First, we apply our approach to form the basins of convergence of methods applied to $z^3 - 1 = 0$ for $z \in \mathbb{C}$ and we succeed in finding methods for which the corresponding basins of convergence do not exhibit a fractal-like structure. Also, this approach has been illustrated here for methods for computing periodic orbits of nonlinear mappings as well as for fixed points of the Poincaré map on a surface section.

In general, using our methodology, by visualizing the performance of various methods we are able to trace reliable and robust methods for a given class of problems which possess similar characteristics by analyzing a specific problem of this class.

Acknowledgments

We thank the guest editors and the anonymous reviewers for their valuable comments and suggestions that helped to improve the content as well as the clarity of the manuscript. Also, we gratefully acknowledge many useful discussions with Assistant Professor O. Ragos and Dr. Y. Petalas. This research has been cofinanced by the European Union (European Social Fund — ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) — Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

References

- Androulakis, G. S. & Vrahatis, M. N. [1996] “OPTAC: A portable software package for analyzing and comparing optimization methods by visualization,” *J. Comput. Appl. Math.* **72**, 41–62.
- Androulakis, G. S., Magoulas, G. D. & Vrahatis, M. N. [1997] “Geometry of learning: Visualizing the performance of neural network supervised training methods,” *Nonlin. Anal.: Th. Meth. Appl.* **30**, 4539–4544.
- Armijo, L. [1966] “Minimization of functions having Lipschitz continuous first partial derivatives,” *Pacific J. Math.* **16**, 1–3.
- Banach, S. [1922] “Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales,” *Fund. Math.* **3**, 133–181.
- Barnsley, M. [1988] *Fractals Everywhere* (Academic Press Professional, Inc., San Diego, CA, USA).
- Beyer, H.-G. [2000] “Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice,” *Comput. Meth. Appl. Mech. Engin.* **186**, 239–267.
- Beyer, H.-G. [2001] *The Theory of Evolution Strategies* (Springer-Verlag Inc., NY, USA).
- Bountis, T. C. & Tompaidis, S. [1991] “Strong and weak instabilities in a 4D mapping model of FODO cell dynamics,” *Workshop on Nonlinear Problems in Future Particle Accelerators*, April 19–25, 1990, Capri, Italy, eds. Turchetti, G. & Scandale, W. (World Scientific, Singapore), pp. 112–127.
- Bountis, T. C. [2004] *The Wondrous World of Fractal* (in Greek) (Leader Books, Athens, Greece).
- Box, G. E. P. & Muller, M. E. [1958] “A note on the generation of random normal deviates,” *Ann. Math. Stat.* **29**, 610–611.
- Bräuninger, J. [1981] “A variable metric algorithm for unconstrained minimization without evaluation of derivatives,” *Numer. Math.* **36**, 359–373.
- Broyden, C. G. [1965] “A class of methods for solving nonlinear simultaneous equations,” *Math. Comput.* **19**, 577–593.
- Cauchy, A. [1847] “Méthode générale pour la résolution des systèmes d’équations simultanées,” *Comptes Rendus de l’Académie des Sciences* **25**, 536–538.
- Cayley, A. [1879] “The Newton–Fourier imaginary problem,” *Amer. J. Math.* **2**, 97.
- Conn, A. R., Scheinberg, K. & Vicente, L. N. [2009] *Introduction to Derivative-Free Optimization* (Society for Industrial and Applied Mathematics, SIAM, Philadelphia, PA, U.S.A.).
- Dennis, J. E. & Schnabel, R. B. [1996] *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Classics in Applied Mathematics, Vol. 16 (Society for Industrial and Applied Mathematics, SIAM, Philadelphia, PA, USA).
- Drossos, L., Ragos, O., Vrahatis, M. N. & Bountis, T. C. [1996] “Method for computing long periodic orbits of dynamical systems,” *Phys. Rev. E* **53**, 1206–1211.
- Elster, C. & Neumaier, A. [1997] “A method of trust region type for minimizing noisy functions,” *Computing* **58**, 31–46.

- Fletcher, R. & Reeves, C. M. [1964] "Function minimization by conjugate gradients," *The Computer J.* **7**, 149–154.
- Giorgi, G., Guerraggio, A. & Thierfelder, J. [2004] *Mathematics of Optimization: Smooth and Non-smooth Case* (Elsevier Science, Amsterdam, The Netherlands).
- Gutzwiller, M. C. [1990] *Chaos in Classical and Quantum Mechanics* (Springer-Verlag, NY).
- Henry, B., Watt, S. & Wearne, S. [2000] "A lattice refinement scheme for finding periodic orbits," *Austr. New Zealand Industrial and Appl. Math. J. (ANZIAM J.)* **42**, C735–C751.
- Hestenes, M. R. & Stiefel, E. [1952] "Methods of conjugate gradients for solving linear systems," *J. Res. Nat. Bureau of Standards* **49**, 409–436.
- Jaynes, E. T. [1979] "Where do we stand on maximum entropy?" *The Maximum Entropy Formalism*, eds. Levine, R. D. & Tribus, M. (The MIT Press, Cambridge, U.S.A.), pp. 15–118.
- Kalantonis, V. S., Perdios, E. A., Perdiou, A. E. & Vrahatis, M. N. [2001] "Computing with certainty individual members of families of periodic orbits of a given period," *Celest. Mech. Dyn. Astron.* **80**, 81–96.
- Kalantonis, V. S., Perdios, E. A., Perdiou, A. E., Ragos, O. & Vrahatis, M. N. [2003a] "Deflation techniques for the determination of periodic solutions of a certain period," *Astrophys. Space Sci.* **288**, 489–497.
- Kalantonis, V. S., Perdios, E. A., Perdiou, A. E., Ragos, O. & Vrahatis, M. N. [2003b] "On the application of optimization methods to the determination of members of families of periodic solutions," *Astrophys. Space Sci.* **288**, 479–488.
- Katsanikas, M. & Patsis, P. A. [2011] "The structure of invariant tori in a 3D galactic potential," *Int. J. Bifurcation and Chaos* **21**, 467–496.
- Katsanikas, M., Patsis, P. A. & Contopoulos, G. [2011a] "The structure and evolution of confined tori near a Hamiltonian Hopf bifurcation," *Int. J. Bifurcation and Chaos* **21**, 2321–2330.
- Katsanikas, M., Patsis, P. A. & Pinotsis, A. D. [2011b] "Chains of rotational tori and filamentary structures close to high multiplicity periodic orbits in a 3D galactic potential," *Int. J. Bifurcation and Chaos* **21**, 2331–2342.
- Morriss, G. P. & Rondoni, L. [1996] "Chaos and its impact on the foundations of statistical mechanics," *Austr. J. Phys.* **49**, 51–77.
- Nocedal, J. & Wright, S. J. [2006] *Numerical Optimization* (Springer-Verlag, NY).
- Ortega, J. M. & Rheinboldt, W. C. [2000] *Iterative Solution of Nonlinear Equations in Several Variables* (Society for Industrial and Applied Mathematics, SIAM, Philadelphia, PA, USA).
- Parsopoulos, K. E. & Vrahatis, M. N. [2002] "Recent approaches to global optimization problems through particle swarm optimization," *Natural Comput.* **1**, 235–306.
- Parsopoulos, K. E. & Vrahatis, M. N. [2003] "Computing periodic orbits of nondifferentiable/discontinuous mappings through particle swarm optimization," *Proc. IEEE 2003 Swarm Intelligence Symposium (SIS 2003)*, April 24–26, 2003, Indianapolis, Indiana, U.S.A. (IEEE Press), pp. 34–41.
- Parsopoulos, K. E. & Vrahatis, M. N. [2004] "On the computation of all global minimizers through particle swarm optimization," *IEEE Trans. Evolut. Comput.* **8**, 211–224.
- Parsopoulos, K. E., Georgopoulos, V. C. & Vrahatis, M. N. [2008] "A technique for the visualization of population-based algorithms," *Proc. IEEE 2008 Congress on Evolutionary Computation (CEC 2008)*, June 1–6, 2008, Hong Kong, China (IEEE Press), pp. 1694–1701.
- Parsopoulos, K. E. & Vrahatis, M. N. [2010] *Particle Swarm Optimization and Intelligence: Advances and Applications* (Information Science Reference (IGI Global), Hershey, PA, U.S.A.).
- Perdios, E. A., Kalantonis, V. S. & Vrahatis, M. N. [2002] "Efficient method for computing with certainty periodic orbits on a surface of section," *Celest. Mech. Dyn. Astron.* **84**, 231–244.
- Perdiou, A. E., Kalantonis, V. S., Perdios, E. A. & Vrahatis, M. N. [2002] "Application of efficient composite methods for computing with certainty periodic orbits in molecular systems," *Comput. Phys. Commun.* **148**, 227–235.
- Perry, A. [1978] "A modified conjugate gradient algorithm," *Operat. Res.* **26**, 1073–1078.
- Polak, E. & Ribière, G. [1969] "Note sur la convergence de directions conjuguées," *Revue Française d'Informatique et de Recherche Opérationnelle* **3**, 35–43.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. [2007] *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd edition (Cambridge University Press, New York, NY, USA).
- Sikorski, K. A. [2001] *Optimal Solution of Nonlinear Equations* (Oxford University Press, Inc., NY, USA).
- Vrahatis, M. N. [1995] "An efficient method for locating and computing periodic orbits of nonlinear mappings," *J. Comput. Phys.* **119**, 105–119.
- Vrahatis, M. N., Bountis, T. C. & Kollmann, M. [1996] "Periodic orbits and invariant surfaces of 4-D nonlinear mappings," *Int. J. Bifurcation and Chaos* **6**, 1425–1437.
- Vrahatis, M. N., Isliker, H. & Bountis, T. C. [1997] "Structure and breakdown of invariant tori in a 4-D mapping model of accelerator dynamics," *Int. J. Bifurcation and Chaos* **7**, 2707–2722.
- Vrahatis, M. N., Androulakis, G. S., Lambrinos, J. N. & Magoulas, G. D. [2000] "A class of gradient

- unconstrained minimization algorithms with adaptive stepsize,” *J. Comput. Appl. Math.* **114**, 367–386.
- Vrahatis, M. N., Perdiou, A. E., Kalantonis, V. S., Perdios, E. A., Papadakis, K., Prosmi, R. & Farantos, S. C. [2001a] “Application of the characteristic bisection method for locating and computing periodic orbits in molecular systems,” *Comput. Phys. Commun.* **138**, 53–68.
- Vrahatis, M. N., Ragos, O. & Androulakis, G. S. [2001b] “Computing families of periodic orbits through optimization methods,” *Nonlin. Anal.: Th. Meth. Appl.* **47**, 3449–3454.
- Vrahatis, M. N., Tsirogiannis, G. A. & Laskari, E. C. [2010] “Novel orbit based symmetric cryptosystems,” *Math. Comput. Model.* **51**, 239–246.
- Wolfe, P. [1969] “Convergence conditions for ascent methods,” *SIAM Rev.* **11**, 226–235.
- Wolfe, P. [1971] “Convergence conditions for ascent methods II: Some corrections,” *SIAM Rev.* **13**, 185–188.