New hybrid algorithm based on nonmonotone spectral gradient and simultaneous perturbation

Zineb Tabbakh, Rachid Ellaia, Abderrahmane Habbal

▶ **To cite this version:**

Zineb Tabbakh, Rachid Ellaia, Abderrahmane Habbal. New hybrid algorithm based on nonmonotone spectral gradient and simultaneous perturbation. International Journal of Mathematical Modelling and Numerical Optimisation, 2019, 9 (1), pp.1-23. 10.1504/IJMMNO.2019.096911 . hal-01944548

HAL Id: hal-01944548
https://inria.hal.science/hal-01944548

Submitted on 4 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Int. J. of Mathematical Modelling and Numerical Optimisation

New Hybrid Algorithm based on Nonmonotone Spectral Gradient and Simultaneous Perturbation

Tabbakh Zineb

LERMA-EMI,
Mohammed V University in Rabat,
Mohammadia School of Engineers, BP. 765, Ibn Sina Avenue, Agdal, Rabat, Morocco
E-mail: tabbakh.zineb@gmail.com

Ellaia Rachid

LERMA-EMI,
Mohammed V University in Rabat,
Mohammadia School of Engineers, BP. 765, Ibn Sina Avenue, Agdal, Rabat, Morocco
E-mail: ellaia@emi.ac.ma

Habbal Abderrahmane

Université Côte d'Azur, Inria, CNRS, LJAD, UMR 7351, Parc Valrose, 06108 Nice, France.
E-mail: habbal@unice.fr

Abstract: In this paper, we introduce a new hybrid method called Nonmonotone Spectral Gradient and Simultaneous Perturbation (NSGSP). It combines the advantages of nonmonotone spectral gradient (NSG), and simultaneous perturbation (SP) methods. The main idea of our approach is to use the simultaneous perturbation (SP) method in order to get a non expensive estimate of the gradient, and exploit the good properties of the nonmonotone spectral gradient (NSG) method in order to compute an efficient line search. Several numerical experiments are provided. The results indicate that the new method is effective and outperforms most of other popular methods.

Keywords: Nonmonotone line search; spectral gradient method; Simultaneous Perturbation Stochastic Approximation

Biographical notes: **Tabbakh Zineb** is PhD student at Mohammed V University in Rabat, Mohammadia School of Engineers, Morocco.

Ellaia Rachid is full Professor at Mohammed V University in Rabat, Mohammadia School of Engineers, Morocco. He received his Doctoral Thesis from Toulouse University in Applied Mathematics and Optimization, and Ph.D. degree in 1992 in Applied and Computational Mathematics. He has held visiting professor positions at Toulouse university, Chile university, INSA of Rouen , Nice Sophia Antipolis University, Haute Alsace University and INRIA Lille. His interests are in Metaheuristics and Nature Inspired Computing, Optimization and Uncertainties, Finance, Global and Stochastic Optimization. Presently, he

Copyright © 201X Inderscience Enterprises Ltd.

is President of the Moroccan Society of Nature-Inspired Computing and Metaheuristics (METAH).

Habbal Abderrahmane is Associate Professor at Nice Sophia-Antipolis University, and INRIA Sophia-Antipolis, France. Research activities of A. Habbal concern analysis and control of systems governed by partial differential equations (PDEs), optimization theory and algorithms and PDE-constrained games. Application fields are related to (nonlinear) mechanics, image processing, data recovering and cell dynamics.

1 Introduction

We consider minimization problems of the general form

$$\begin{cases} \text{minimize} \quad f(x) \\ \text{subject to} \ x \in S \end{cases} \tag{1}$$

where f is a smooth cost function and $S \subset \mathbf{R}^n$ defines the optimization constraints.

In the vast nonlinear optimization literature, many deterministic and stochastic methods have been developed to solve (1). Among the most popular are the steepest descent algorithms (It is out of the scope of the present paper to provide a review of these methods). From the latter family, spectral gradient methods, originated in Barzilai and Borwein (1988) have been applied successfully to find local minimizers of minimization problems for which computational issues arise such as cost and gradient evaluation is expensive, the Hessian is either not available or requires a prohibitive amount of storage. Notice that these methods are essentially steepest descent method, where an exact computation of the optimal step size along the direction of negative gradient is approximated by different effective choices, yielding as many various methods.

Barzilai and Borwein (1988) step size is computed from a two-point approximation to the secant equation required in quasi-Newton methods. Raydan (1997) presented the global Barzilai-Borwein (GBB) method based on the technique of nonmonotone line search, which was first introduced in Grippo et al. (1986). This technique allows the convergence to a minimum of a broader class of cost functions f, starting from any initial point x_0, than the usual well behaving case (strongly convex to say it roughly). Its main characteristic is the non-descent property.

Therefore, Birgin et al. (2000) have improved the work of Raydan (1997) and introduced the nonmonotone spectral projected gradient (NSPG) method. This approach was developed for the minimization of smooth functions on a closed convex set and has been successfully applied, see e.g. Lakhbab and Bernoussi (2016). On the other hand, simultaneous perturbation stochastic approximation method (SPSA) has been developed by Spall (2000) and used to solve the problem (1). In this method, the gradient is estimated using only two cost evaluations regardless of the dimension n of the optimization space. In simultaneous perturbation stochastic approximation (SPSA), all component directions are perturbed simultaneously using perturbations which are vectors of independent random variables assumed to be symmetric, zero-mean, ± 1 valued, and Bernoulli distributed.

Unlike standard finite difference approximations which require $2n$ cost evaluations to estimate the gradient, Kiefer and Wolfowitz (1952), SPSA methods are very attractive as

soon as the cost evaluation is computationally expensive. Indeed, it is the case for most engineering optimization problems, where cost functions are obtained through the solution of very large scale linear and nonlinear systems, and where closed or implementable forms of the gradient are generally not available.

The main objective of this paper is to propose a method to solve problem (1) combining the nonmonotone spectral gradient (NSG) with the simultaneous perturbation (SP) method. The descent direction is a nearly unbiased estime of the gradient provided by the simultaneous perturbation (SP) method. Along this direction, we use a spectral choice of the step, combined with a nonmonotone line search strategy allowing for the cost function to increase for a limited number of iterations. The advantages of the proposed algorithm are assessed on numerous benchmark functions and the results are compared with many of the existing methods. The obtained results illustrate the efficiency and robustness of our method to solve large scale problems for a broad class of smooth objective functions.
The remaining of the paper is organized in the following way. In Section 2, the new hybrid method is described. In Section 3, numerical results are presented for the solution of various benchmark problems. Finally, in Sect.4 the paper is concluded.

2 NSG, SPSA and hybrid NSGSP methods

We recall in this section the basic features of the Nonmonotone Spectral Gradient (NSG) and Simultaneous Perturbation (SP) methods ; then, we describe with more details our proposed Nonmonotone Spectral Gradient and Simultaneous Perturbation hybrid method (NSGSP), and finally shortly recall the penalization method which we use in (NSGSP) to deal with constrained optimization.

2.1 Nonmonotone Spectral Gradient NSG method

The Problem (1) is frequently numerically solved, seeking a local minimizer $\mathbf{x}^*$, by using iterative methods according to the following generic scheme

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k g_k(\mathbf{x}_k) \tag{2}$$

where g_k is the gradient of f at $\mathbf{x}_k$, and $\alpha_k \in \mathbf{R}^+$ the step length, given initial $\mathbf{x}_0 \in \mathbf{R^n}$.

Instead of an exact and expensive computation of the optimal line search

$$\alpha_k = \mathrm{argmin}_\alpha f(\mathbf{x}_k + \alpha g_k(\mathbf{x}_k))$$

the so-called spectral gradient methods aim at providing cheap yet effective approximations of the step length. By combining the nonmonotone line search of Grippo et al. (1986) with the spectral gradient choice of the step length of Barzilai and Borwein (1988), Raydan (1997) proposed the following choice

$$\alpha_k = \frac{\mathbf{s}_{k-1}^T \mathbf{s}_{k-1}}{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}} \tag{3}$$

where $\mathbf{s}_{k-1} = \mathbf{x}_k - \mathbf{x}_{k-1}$ and $\mathbf{y}_{k-1} = g_k - g_{k-1}$.

This line search essentially enforces the following condition:

$$f(\mathbf{x}_{k+1}) \leq \max_{0\leq j\leq \min\{k,M\}} f(\mathbf{x}_{k-j}) + \gamma g_k^T(\mathbf{x}_{k+1} - \mathbf{x}_k), \tag{4}$$

where M is a nonnegative integer and γ is a small positive number.

Among alternative approaches to compute a line search, let us mention the one based on safeguarded quadratic interpolation proposed in Dai et al. (2002). The following quadratic model is used to approximate the objective function at $f(\mathbf{x}_k + \lambda \mathbf{s}_{k-1})$:

$$q_k(\lambda) = f_k + \lambda g_k^T \mathbf{s}_{k-1} + \frac{1}{2}\alpha_k \lambda^2 ||\mathbf{s}_{k-1}||^2, \tag{5}$$

where $\alpha_k = \frac{\mathbf{s}_{k-1}^T \mathbf{y}_{k-1}}{\mathbf{y}_{k-1}^T \mathbf{y}_{k-1}}$, and the above model satisfies two interpolation conditions

$$\nabla q_k(0) = g_k^T \mathbf{s}_{k-1} \quad and \quad q_k(-1) = f_{k-1},$$

which yields the following choice of stepsize formula

$$\lambda_k = -\frac{\mathbf{s}_{k-1}^T \mathbf{s}_{k-1}}{2(f_{k-1} - f_k) + 2\mathbf{s}_{k-1}^T g_k}. \tag{6}$$

After the initial choice of step size, the nonmonotone line search is used to determine the appropriate step length in the appropriate direction, based on the formula (6).

2.2 Simultaneous perturbation

In practice, exact values of the gradient of cost functions, notably those considered in engineering applications, are either not available or very costly to approximate by some finite difference scheme, due to the large number of variable and to expensive evaluation of the costs. In this case, nonmonotone spectral gradient (NSG) cannot be applied directly. To overcome the problem of the possibly prohibitive computational cost in gradient approximation, one may consider the simultaneous perturbation (SP) method due to Spall (2000).

The gradient estimate of $f(.)$ when using simultaneous perturbation (SP) is given by the following centered formula

$$\hat{g}_k(x_k) = \frac{f(x_k + c_k\Delta_k) - f(x_k - c_k\Delta_k)}{2c_k} \left[\Delta_{k1}^{-1}, \Delta_{k2}^{-1}, ..., \Delta_{kn}^{-1}\right]^T \tag{7}$$

where c_k is a small positive number, and the components Δ_{ki} $(i = 1, 2, \cdots, n)$ of the perturbation $\Delta_k = [\Delta_{k1}, \Delta_{k2}, \cdots, \Delta_{kn}]^T$ are independent, symmetrically distributed about zero with finite inverse and second moments, and satisfying $|\Delta_{ki}| < \alpha_0$, and $E(\left|\Delta_{ki}^{-1}\right|) < \alpha_1$ for all indices (k, i) for some finite $\alpha_1 0$ and α_1. A simple and popular choice for each component of Δ_k is a Bernoulli ± 1 distribution with probability $\frac{1}{2}$.
As shown in Spall (1992), simultaneous perturbation (SP) provides an asymptotically unbiased estimate of the exact gradient g_k while requiring only a fraction $\frac{1}{n}$ of the total cost spent in approximating the gradient with a centered finite difference scheme.

We recall hereafter the estimation theorem proved in Spall (1992), as well as a sketched proof (for it is simple enough, despite the striking claim of the theorem).

Theorem 1: *Let be* $\Delta_k = [\Delta_{k1}, \Delta_{k2}, \cdots, \Delta_{kn}]^T$ *that satisfy the above conditions, and suppose that* f *is smooth enough, at least twice continuously differentiable, and noise-free. Then,* $\hat{g}_k(x_k)$ *is a nearly unbiased estimate of the* $g_k(x_k)$ *of* f

$$E\left(\hat{g}_k(x_k)\right) = g_k(x_k) + \mathcal{O}(c_k^2) \;\; almost\ surely \tag{8}$$

where $g_k(x_k)$ *is the true gradient of* f *at the -deterministic- point* x_k*.*

Proof 1: *At some fixed iteration* k*, we take noise-free measurements of* f *at* $x_k + c_k\Delta_k$*, using a standard second order Taylor series expansion,*

$$f(x_k \pm c_k\Delta_k) = f(x_k) \pm c_k g(x_k)^T\Delta_k + \frac{c_k^2}{2}\Delta_k^T.\nabla^2 f(x_k)\Delta_k + \mathcal{O}(c_k^3\|\Delta_k\|^3) \tag{9}$$

Then, for $i = 1 \cdots n$*, one has*

$$E\left[\hat{g}_{ki}(x_k)\right] = E\left[\frac{f(x_k + c_k\Delta_k) - f(x_k - c_k\Delta_k)}{2c_k\Delta_{ki}}\right] \tag{10}$$

From (9) we have

$$E\left[\hat{g}_{ki}(x_k)\right] = E\left[\frac{\left[c_k g(x_k)^T\Delta_k + \mathcal{O}(c_k^3\|\Delta_k\|^3)\right] - \left[-c_k g(x_k)^T\Delta_k + \mathcal{O}(c_k^3\|\Delta_k\|^3)\right]}{2c_k\Delta_{ki}}\right] \tag{11}$$

It follows that:

$$E\left[\hat{g}_{ki}(x_k)\right] = E\left[\frac{2c_k g(x_k)^T\Delta_k + \mathcal{O}(c_k^3\|\Delta_k\|^3)}{2c_k\Delta_{ki}}\right] \tag{12}$$

$$= E\left[\frac{2c_k \sum_{j=1}^{n} g_j(x_k)\Delta_{kj} + \mathcal{O}(c_k^3\|\Delta_k\|^3)}{2c_k\Delta_{ki}}\right] \tag{13}$$

$$= g_i(x_k) + \sum_{j\neq i} g_j(x_k) E\left(\frac{\Delta_{kj}}{\Delta_{ki}}\right) + \mathcal{O}\left(c_k^2 E\left[\|\Delta_k\|^2\right]\right). \tag{14}$$

Hence, under the conditions given above on Δ_k*,*

$$\forall j = 1, \cdots, n \;\; E\left(\Delta_{kj}^2\right) = 1, \;\; \textit{and for all } j \neq i \;\; E\left(\frac{\Delta_{kj}}{\Delta_{ki}}\right) = 0.$$

Consequently,

$$E\left[\hat{g}_{ki}(x_k)\right] = g_i(x_k)(x_k) + \mathcal{O}(c_k^2). \tag{15}$$

The bias of gradient estimation is of order $\mathcal{O}(c_k^2)$ *.*

Remark 1: *Notice that the proof above assumes that the point* x_k *is deterministic, which no more holds when we shall consider iterates* x_k *updated using the stochastic SP approximation. The proof however still holds in this case, by considering expectations conditional to* x_k*, see Spall (1992).*

2.3 The proposed NSGSP method

The nonmonotone spectral gradient and simultaneous perturbation (NSGSP) method combines the local search algorithm nonmonotone spectral gradient (NSG) with the descent method where the gradient is estimated by using the simultaneous perturbation. The recursive procedure (2) becomes:

$$\hat{x}_{k+1} = \hat{x}_k - \alpha_k \hat{g}_k(\hat{x}_k), \tag{16}$$

where $\hat{x}_k$ is the current point, $\hat{g}_k(\hat{x}_k)$ is the estimate gradient of f at $\hat{x}_k$ obtained using simultaneous perturbation (SP), and α_k is the Barzilai-Borwein (BB) step size attained via nonmonotone line search.

2.4 Handling the constraints in NSGSP

To solve constrained optimization problems we have used a classical penalty method, owing to some encouraging properties such as the simplicity of these methods and their ease of implementation. These methods were originally proposed by Courant R. (1943) and developed among others by Carroll (1961) and Fiacco and McCormick (1966).

Applying penalty methods, we solve a sequence of unconstrained problems by adding to the objective function a penalty term for violation of constraints. The penalty function is weighted by an increasing sequence of large positive numbers, forcing the unconstrained solution to converge to a feasible one, that is, to satisfy all constraints. There are two types of methods: interior penalty methods, and exterior penalty methods. Since our problem has both inequality and equality constraints, we used exterior penalty functions and combined it with the NSGSP method to handle the constraints in the problem (1).

Assume that the feasible region S of the minimization problem (1) is given by

$$S = \{x : G_i(x) \leq 0, \quad i = 1, 2, \cdots, m \quad and \quad H_j(x) = 0, \quad j = 1, 2, \cdots, p\}.$$

where m is the number of inequality constraints and p the number of equality constraints. The problem (1) can be replaced by the unconstrained one as follows

$$\min_{x \in \mathbf{R}^n} \; f(x) + \mu P(x), \tag{17}$$

where μ is a positive parameter and P is a penalty function of the form:

$$P(x) = \sum_{i=1}^{m} (\max[0, G_i(x)])^2 + \sum_{j=1}^{p} |H_j(x)|^2$$

P is of course positive and continuous and $P(x) = 0$ if and only if $x \in S$.
However, despite its implementation simplicity, the difficulty of the penalty function approach is to find an appropriate sequence of penalty parameters, that must be chosen large enough to converge to the solution of the constrained problem but not too large, to avoid degeneracy and ill conditioning.

3 Computational Experiments

The proposed Nonmonotone Spectral Gradient and Simultaneous Perturbation (NSGSP) algorithm was implemented in Matlab. We have followed a Taguchi methodology to tune different parameters specific to the code, and then applied our method to a wide range of benchmark problems.

We assessed our algorithm versus a large number of unconstrained and constrained academic problems, and compared our obtained results to numerous popular methods. We thereafter led a comparative study on a worldwide referenced engineering optimization benchmarks : three bar truss, optimal control, pressure vessel design, tension-compression spring design and the speed reducer design.

For all the presented numerical experiments, the same stopping criteria were used : either the gradient norm was below some small tolerance threshold (relevant for unconstrained and penalized problems), or the maximum iterations were used without achieving convergence.

3.1 Preliminary parameter design and analysis

For the robust tuning of parameters specific to our code, we have used a Taguchi methodology. We have paid particular attention to three parameters : γ, c and M. These parameters are not classical in descent methods, and a sensitivity analysis was then mandatory. After a preliminary analysis of the nonmonotone spectral gradient and simultaneous perturbation (NSGSP) algorithm, we have chosen three levels for each parameter, as described in Table 1. The aim is to estimate the parameters variation effects, in order to set up a robust choice of the latter.

The Taguchi method uses a loss function to measure the deviation between an experimental value and a desired one. See e.g. Peace (1993). This loss function is transformed into a signal-to-noise (S/N) ratio, so that a larger signal-to-noise (S/N) ratio corresponds to a better parameter design, that is, the parameter setting with the highest signal-to-noise (S/N) ratio s the best choice for the experiment. In the present study, the lower-the-better-quality characteristic in signal-to-noise ratio is considered

$$\eta = -10 \log \frac{1}{m} \sum_{j=1}^{m} y_j^2 \tag{18}$$

where y_j is the observed data in jth experimentation and m is the number of observations. The $L_{18}(3^3)$ orthogonal array (refer to Peace (1993) for details) shown in Table 2 was used to select the optimal parameters.

After the S/N ratios are measured (Figure 1 and Table 2), the next step is to analyze their significance and effects based on analysis of variance (ANOVA). The results are shown in Table 3.

The results of the variance analysis of the parameters design experiment (Table 3) show that the parameter c (B) is the factor that represents the larger effect on the total variation in the function value, with a contribution amount of $37,91\%$; the second factor is γ with a contribution of 24,50%. Then follows the effect of M with a contribution of 14,62%.

The Taguchi design experiment resulted in the choice of the following parameters for the nonmonotone spectral gradient and simultaneous perturbation (NSGSP) :
$\gamma = \mathbf{10^{-4}}$, **M=10** and $\mathbf{c = 0.01}$.

The simultaneous perturbation (SP) sequence is given by $c_k = \frac{c}{k^\tau}$ where $\tau = 0.101$ and k is the iteration counter.
The setup for other computational parameters is (see the significance in Algorithm 1) $\alpha_{min} = 10^{-30},\ \alpha_{max} = 10^{30}$, $\sigma_1 = 0.1$, $\sigma_2 = 0.9$ and $etol = 10^{-6}$.

Next, we use the following indicators : SD for standard deviation, SR for successful rate. The indicator SR is used to calculate the number of successful runs, that is, when the algorithm generates a desired solution with a required accuracy. We have chosen an accuracy of 10^{-4} for all the tests we ran.

Furthermore, to illustrate the statistical significance of the performance of the proposed method, Cohen's d statistical distance is used, for unconstrained optimization examples, to quantify the difference between two means. Each minimization example was run 50 times for twice, and Cohen's d can then be calculated as:

$$d = \frac{\text{Mean}_1 - \text{Mean}_2}{\sigma_{\text{pooled}}} \qquad \text{where} \qquad \sigma_{\text{pooled}} = \sqrt{\frac{(SD_1^2 + SD_2^2)}{2}} \tag{19}$$

For our minimization benchmarks, the indicators SD and SR, as well as the Cohen's d statistics are reported in Table 7.

3.2 Unconstrained optimization benchmark

We consider a standard, widely used, benchmark list of unconstrained optimization problems. Objective functions, spaces and optima values are listed in Table 4 and in Table 5. These test case functions have different features making them challenging, like non convexity and/or multi-modality. They have known local and global minima.

The NSGSP algorithm introduced in the present paper was ran on these test cases, and the obtained results were compared to numerous other well proved algorithms listed in Table 6. The computational results and statistical metrics are presented in Table 7.

We have plotted in Figure 2 (Eason and Fenton function) and in Figure 3 (Aluffi-Pentini function) the evolution history of the objective functions with respect to the optimization iterations. These two profiles illustrate very clearly the non-descent property of the NSGSP algorithm (inherited from the NSG management of the iterates).

3.3 Constrained optimization benchmark

We have considered five constrained optimization test cases as detailed below. In Table 8, we have reported the results obtained with the NSGSP method. These results are compared to the ones available from the literature.

3.3.1 Constrained problem I:

$$\begin{array}{l} f(x) = (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{s.t,} \\ g_1(x) = x_1 - 2x_2 + 1 = 0, \\ g_2(x) = x_1^2/4 - x_2^2 + 1 \geq 0, \\ \qquad -10 \leq x_1, x_2 \leq 10. \end{array} \tag{20}$$

The global minimum is attained at $x^* = (0.82288, 0.91144)$, and $f(x^*) = 1.3935$.

3.3.2 Constrained problem II:

$$
\begin{aligned}
f(x) &= (x_1^2 + x_2 - 11)^2 + (x_1^2 + x_2^2 - 7)^2 \\
&\text{s.t,} \\
g_1(x) &= 4.84 - (x_1 - 0.05)^2 - (x_2 - 2.5)^2 \geq 0, \\
g_2(x) &= x_1^2 + (x_2 - 2.5)^2 - 4.84 \geq 0, \\
&0 \leq x_1, x_2 \leq 6.
\end{aligned} \tag{21}
$$

The global minimum is attained at $x^* = (2.246826, 2.381865)$, and $f(x^*) = 13.59085$, see Bracken and McCormick (1968).

3.3.3 Constrained problem III:

$$
\begin{aligned}
f(x) &= 5.3578547x_3^2 + 0.8356891x_1x_5 + 37.293239x_1 - 40792.141, \\
&\text{s.t,} \\
g_1(x) &= 85.334407 + 0.0056858x_2x_5 + 0.0006262x_1x_4 - 0.0022053x_3x_5 - 92 \leq 0, \\
g_2(x) &= -85.334407 - 0.0056858x_2x_5 - 0.0006262x_1x_4 + 0.0022053x_3x_5 \leq 0, \\
g_3(x) &= 80.51249 + 0.0071317x_2x_5 + 0.0029955x_1x_2 + 0.0021813x_3^2 - 110 \leq 0, \\
g_4(x) &= -80.51249 - 0.0071317x_2x_5 - 0.0029955x_1x_2 - 0.0021813x_3^2 + 90 \leq 0, \\
g_5(x) &= 9.300961 + 0.0047026x_3x_5 + 0.0012547x_1x_3 + 0.0019085x_3x_4 - 25 \leq 0, \\
g_6(x) &= -9.300961 - 0.0047026x_3x_5 - 0.0012547x_1x_3 - 0.0019085x_3x_4 + 20 \leq 0, \\
&78 \leq x_1 \leq 102, 33 \leq x_2 \leq 45 \text{ and } 27 \leq x_i \leq 45 (i = 3; 4; 5)..
\end{aligned} \tag{22}
$$

3.3.4 Constrained problem IV

$$
\begin{aligned}
f(x) &= (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7, \\
&\text{s.t,} \\
g_1(x) &= 127 - 2x_1^2 - 3x_2^4 - x_3 - 4x_4^2 - 5x_5 \geq 0, \\
g_2(x) &= 282 - 7x_1 - 3x_2 - 10x_3^2 - x_4 + x_5 \geq 0, \\
g_3(x) &= 196 - 23x_1 - x_2^2 - 6x_6^2 + 8x_7 \geq 0, \\
g_4(x) &= -4x_1^2 - x_2^2 + 3x_1x_2 - 2x_3^2 - 5x_6 + 11x_7 \geq 0, \\
&-10 \leq x_i \leq 10 \quad (i = 1 \cdots 7).
\end{aligned} \tag{23}
$$

Optimal solution is $x^* = (2.330499, 1.951372, -0.4775414, 4.365726, -0.6244870, 1.038131, 1.594227)$, and $f(x^*) = 680.6300573$.

3.3.5 Constrained problem V

$$
\begin{array}{l}
f(x) = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3-10)^2 + 4(x_4-5)^2 + (x_5-3)^2 + 2(x_6-1)^2 + 5x_7^2 \\
\quad +7(x_8-11)^2 + 2(x_9-10)^2 + (x_{10}-7)^2 + 45, \\
\text{s.t,} \\
g_1(x) = 4x_1 + 5x_2 - 3x_7 + 9x_8 - 105 \geq 0, \\
g_2(x) = 10x_1 - 8x_1 - 17x_7 + 2x_8 \geq 0, \\
g_3(x) = -8x_1 + 2x_2 + 5x_9 - 2x_{10} - 12 \geq 0, \\
g_4(x) = 3(x_1-2)^2 + 4(x_2-3)^2 + 2x_3^2 - 7x_4 - 120 \geq 0, \\
g_5(x) = 5x_1^2 + 8x_2 + (x_3-6)^2 - 2x_4 - 40 \geq 0, \\
g_6(x) = x_1^2 + 2(x_2-2)^2 - 2x_1x_2 + 14x_5 - 6x_6 \geq 0, \\
g_7(x) = 0.5(x_1-8)^2 + 2(x_2-4)^2 + 3x_5^2 - x_6 - 30 \geq 0, \\
g_8(x) = -3x_1 + 6x_2 + 12(x_9-8)^2 - 7x_{10} \geq 0, \\
\quad -10 \leq x_i \leq 10 \quad (i = 1 \cdots 10).
\end{array}
\tag{24}
$$

Optimal solution is $x^* = (2.1719, 2.3636, 8.7739, 5.9598, 0.9906, 1.4305, 1.3216, 9.8287, 8.2800, 8.3759)$, and $f(x^*) = 24.3062091$.

3.4 Optimal Control problem

The computational solution of optimal control problems is an important field in optimization theory. It has been studied by many workers, see e.g. Delprat et al. (2004), Delprat et al. (1996) and Mehra and Davis (1972).

We consider the following linear quadratic optimal control

$$
\begin{cases}
\min_u \quad J(x_0,u) = \int_0^\infty [x(t)^T Q x(t) + u(t)^T R u(t)]dt \\
\text{subject to} \qquad \dot{x} = Ax + Bu\,; \qquad x(0) = x_0
\end{cases}
\tag{25}
$$

where Q is a positive semidefinite matrix, R is a positive definite matrix, and x_0 is given. The problem (25) can be solved by using the algebraic Riccati equation (see, for example Broucke (2012)), a powerful tool in linear control theory. Since the algebraic Riccati equation is a nonlinear algebraic equation, solving it analytically in dimensions greater than 3 is intractable in general.

Assumption 1: $(A,\ B)$ is stabilizable.
By stabilizability, there exists a feedback gain K such that the closed loop system

$$\dot{x} = (A - BK)x \tag{26}$$

is stable. The feedback law $u = -Kx$ is clearly admissible. The solution of 26 is given by

$$x(t) = e^{(A-BK)t}x_0 \tag{27}$$

The cost criterion is then given by

$$J(x_0, -Kx) = x_0^T \int_0^\infty \mathbf{e}^{(A-BK)^T t}(Q + K^T R K)\mathbf{e}^{(A-BK)t}dt x_0$$

which is finite. Hence the optimization problem is meaningful.

3.4.1 An NSGSP illustration

Consider the optimal control of a servomotor. Here $u(t)$ is a scalar control and $x \in \mathbf{R}^2$.

$$\begin{cases} \min_u \quad J(x_0,u) = \int_0^\infty [x^T(t)Qx(t) + ru^2(t)]dt \\ \text{subject to} \qquad \frac{dx}{dt} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} x + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \end{cases} \tag{28}$$

where $Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, r > 0.$

We discretize problem (28) as follow

$$\begin{cases} \text{minimize} \qquad h(\sum_1^{n+1} y_i^2 + r\sum_1^n u_i^2) \\ \text{subject to} \qquad y_{i+1} - y_i - dt.z_i = 0 \\ \qquad\qquad z_{i+1} + (-1+dt)z_i - dt.u_i = 0 \end{cases} \tag{29}$$

where $n = 20$, and $x = (y, z)$; the integration step to approximate the cost function is $h = 10^{-3}$ and the ODE time step is $dt = 10^{-4}$.

We choose $r = 0.5, u_0 = -5.\mathbb{1}_n$ and $x_0 = [3.25; 4.95]$. By using nonmonotone spectral gradient and simultaneous perturbation (NSGSP) method in problem (29), we found $K \approx [1.41; 0.96]$ and $J(x_0, u^*) = 4.5712$. The optimal control is given by $u^*(t) = -1.41y(t) - 0.96z(t)$. The same result is obtained by using the Matlab command **lqr**.

3.5 Pressure vessel design problem

The pressure vessel design problem is a well known benchmark problem introduced in Kannan and Kramer (1994). The main objective is to minimise the cost $f(x)$ of the material mass, forming and welding. There are four design variables: $Ts(= x_1)$ is the thickness of the shell, $Th(= x_2)$ is the spherical head thickness, $R(= x_3)$ is the inner radius, and $L(= x_4)$ is the length of the cylindrical section. Length unit is in inches.

In the literature, Ts and Th are integer multiples of 0.0625, and x_3, x_4 are continuous variables, such that $40 \le x_3 \le 80$, and $20 \le x_4 \le 60$.

The formulation of this problem can be written as follows:

$$\begin{cases} \text{minimize} \quad f(x) = 0.6224x_1x_3x_4 + 1.7781x_2x_3^3 + 3.1611x_1^2x_4 + 19.84x_1^2x_3 \\ \text{subject to} \qquad g_1(x) = 0.0193x_3 - x_1 \ge 0 \\ \qquad\qquad g_2(x) = 0.00954x_3 - x_2 \ge 0 \\ \qquad\qquad g_3(x) = 750.0 \times 1728.0 - \pi x_3^2 x_4 - \frac{4}{3}\pi x_3^2 \ge 0 \\ \qquad\qquad g_4(x) = x_4 - 240.0 \ge 0 \\ \qquad\qquad g_5(x) = 1.1 - x_1 \ge 0 \\ \qquad\qquad g_6(x) = 0.6 - x_2 \ge 0 \end{cases} \tag{30}$$

The NSGSP method was applied to the pressure vessel design problem and the obtained results were compared to those of the literature (De KS Lee et al, Wu, Chow and Sandgren) as illustrated in Table 11.

3.6 Tension-compression spring design problem

The tension/compression spring design problem is described in Arora (1989). The objective is to minimize the weight $f(x)$ of the tension compression spring (as shown in Figure 5). The optimization problem involves four nonlinear inequality constraints, namely the minimum deflection, shear stress, surge frequency, outside diameter, and bounds on the design variables. The design variables are wire diameter $d(= x1)$, mean coil diameter $D(= x2)$, and number of active coils $P(= x3)$. The NSGSP obtained solution is given in Table 13 where it is compared to those obtained by other methods.

$$\begin{cases} \text{minimize} & f(x) = (x_3+2)x_2x_1^2, \\ \text{Subject to} & g_1(x) = 1 - \dfrac{x_2^2 x_3}{71785x_1^4} \geq 0, \\ & g_2(x) = \dfrac{4x_2^2 - x_1x_2}{12566(x_2x_1^3 - x_1^4)} + \dfrac{1}{5108x_1^2} - 1 \geq 0, \\ & g_3(x) = 1 - \dfrac{140.45x_1}{x_2^2x_3} \geq 0, \\ & g_4(x) = \dfrac{x_2+x_1}{1.5} - 1 \geq 0, \\ & 0.05 \leq x_1 \leq 2.0,\ 0.25 \leq x_2 \leq 1.3,\ 2 \leq x_3 \leq 15 \end{cases} \tag{31}$$

3.7 Speed reducer design problem

The speed reducer design problem is shown in Figure 6. The objective to minimize is the weight of the speed reducer subject to constraints on on bending stress of the gear teeth, surface stress, transverse deflections of the shafts, stresses in the shaft, and bounds on the design variables. The decision variables are the face width $b(= x_1)$, module of teeth $m(= x_2)$, number of teeth on pinion $z(= x3)$, length of the first shaft between bearings $l_1(= x_4)$, length of the second shaft between bearings $l_2(= x_5)$, diameter of the first shaft $d_1(= x_6)$, and diameter of the second shaft $d_2(= x_7)$.

The problem statement is the following :

$$
\begin{aligned}
\text{minimize}\quad & f(x) = 0.7854x_1x_2^2(3.3333x_3^2 + 14.9334x_3 - 43.0934) - 1.508x_1(x_6^2 + x_7^2) \\
& +7.4777(x_6^3 + x_7^3) \\
\text{Subject to}\quad & g_1(x) = \frac{27}{x_1x_2^2x_3} - 1 \geq 0, \\
& g_2(x) = \frac{397.5}{x_1x_2^2x_3^2} - 1 \geq 0, \\
& g_3(x) = \frac{1.93x_4^3}{x_2x_3x_6^4} - 1 \\
& g_4(x) = \frac{1.93x_5^3}{x_2x_3x_7^4} - 1 \\
& g_5(x) = \frac{((\frac{745x_4}{x_2x_3})^2 + 16.9 \times 10^6)^{1/2}}{110.0x_6^3} - 1 \geq 0, \\
& g_6(x) = \frac{((\frac{745x_4}{x_2x_3}) + 157.5 \times 10^6)^{1/2}}{85.0x_7^3} - 1 \\
& g_7(x) = \frac{x_2x_3}{40} - 1 \\
& g_8(x) = \frac{5x_2}{x_1} - 1 \\
& g_9(x) = \frac{x_1}{12x_2} - 1 \\
& g_{10}(x) = \frac{1.5x_6 + 1.9}{x_5} - 1 \\
& g_{11}(x) = \frac{1.1x_7 + 1.9}{x_5} - 1 \\
& 2.6 \leq x_1 \leq 3.6,\ 0.7 \leq x_2 \leq 0.8,\ 17 \leq x_3 \leq 28,\ 7.3 \leq x_4 \leq 8.3, \\
& 7.8 \leq x_5 \leq 8.3,\ 2.6 \leq x_6 \leq 3.9,\ 5.0 \leq x_7 \leq 5.5
\end{aligned}
\tag{32}
$$

The results obtained by NSGSP are presented in Table 15, where they are compared to the results available in the literature.

3.8 Three bar truss

The three bar truss problem was introduced by Nowcki(1973). It is often used as a benchmark problem in structural optimization. The objective is to minimize the volume of a statically loaded 3-bar truss, subject to stress constraints on each member. The cross sectional areas A_i are within the range $0.1 < A_i < 1$.

The mathematical programming formulation of this problem can be stated as follows:

$$
\begin{cases}
\text{minimize} & f(x) = (2\sqrt{2}A_1 + A_2) \times l \\
\text{subject to} & g_1(x) = \dfrac{\sqrt{2}A_1 + A_2}{\sqrt{2}A_1^2 + 2A_1A_2}P - \sigma \leq 0 \\
& g_2(x) = \dfrac{A_2}{\sqrt{2}A_1^2 + 2A_1A_2}P - \sigma \leq 0 \\
& g_3(x) = \dfrac{1}{A_1 + \sqrt{2}A_2}P - \sigma \leq 0 \\
& 0 \leq A_1 \leq 1,\ 0 \leq A_2 \leq 1
\end{cases}
\tag{33}
$$

where $l = 100cm$, $P = 2KN/cm^2$ and $\sigma = 2KN/cm^2$.
The NSGSP results are presented in Table 16, and compared to other methods in Table 17.

4 Conclusion

The present paper introduces a new minimization method called the nonmonotone spectral gradient and simultaneous perturbation (NSGSP). This method is applicable when the gradient of the objective function is not available or is expensive to compute. The gradient is replaced by a simultaneous perturbation stochastic approximation (SPSA). It then applies the Barzilai-Borwein step size and uses the nonmonotone line search techniques, which greatly improves the efficiency of the method. with the non monotone line search, the algorithm was able to escape the local minimum to better one. In many test cases, the new hybrid method exhibited some nice global optimization abilities.

The proposed method was applied to well-known benchmark functions and different classical problems, including twenty unconstrained, five constrained minimization problems, and structural optimization problems. All obtained results contribute to demonstrate the effectiveness of our method. The results were compared to those of published works found in literature. In all examples, nonmonotone spectral gradient and simultaneous perturbation (NSGSP) gave very satisfactory results, which indicates that the nonmonotone spectral gradient and simultaneous perturbation (NSGSP) is competitive, and can be used as an optimization technique for solving engineering optimization problems in challenging (non convex, multimodal) problems. A next step is to adapt the NSGSP approach in order to tackle multiobjective optimization problems, the aim being to capture the Pareto Frontier for large scale problems (those known as many-objective problems).

References

B. Raphael and I.F.C. Smith.(2003) 'A direct stochastic algorithm for global search.'*Applied Mathematics and Computation*, 146:729–758.

C. Zhang, J. Ning, S. Lu, D. Ouyang, and T. Ding. (2009) 'A novel hybrid differential evolution and particle swarm optimization algorithm for unconstrained optimization',*Operations Research Letters*, 37:117–122.

E.G. Birgin, J.M. Martínez, and M. Raydan.(2000) 'Nonmonotone spectral projected gradient methods on convex sets.' *SIAM Journal on Optimization*, 10:1196–1211.

I.G. Tsoulos.(2008) 'Modifications of real code genetic algorithm for global optimization.'*Applied Mathematics and Computation*, 203:598–607

Jerome Bracken and Garth P McCormick.(1968) 'Selected applications of nonlinear programming.' *Technical report, DTIC Document*.

J. Barzilai and J. M. Borwein.(1988) 'Two point step size gradient methods.' *IMA Journal of Numerical Analysis*, 8:141–148.

J. Kiefer and J. Wolfowitz.(1952) 'Stochastic estimation of the maximum of a regression function.'*Ann. Statist*, 23:462-466.

James C. Spall.(2000) 'Adaptive stochastic approximation by the simultaneous perturbation method.' *IEEE Transactions on Automatic Control*, 45:1839–1853.

Halima Lakhbab and Souad El Bernoussi.(2016) 'Hybrid nonmonotone spectral gradient method for the unconstrained minimization problem.'*Computational and Applied Mathematics*, pages 1–10.

K. Deep and M. Thakur. (2007) 'A new mutation operator for real coded genetic algorithms.' *Applied Mathematics and Computation*, 193:211-230.

L. Wei and M. Zhao.(2005) 'A niche hybrid genetic algorithm for global optimization of continuous multimodal functions.'*Applied Mathematics and Computation*, 160:649–661.

L. Grippo, F. Lampariello, and S. Lucidi.(1986)'A nonmonotone line search technique for newton's method.'*SIAM Journal on Numerical Analysis*, 23:707–716.

M. Broucke.(2012)'Linear geometric control theory.'*Courses Notes, Toronto University*, pages 141–147.

M. Raydan.(1997) 'The barzilai and borwein gradient method for the large scale unconstrained minimization problem.'*SIAM Journal on Optimization*, 7:26–33.

M. Duran Toksary.(2006)'A heuristic approach to find the global optimum of function.'*Journal of Computational and Applied Mathematics*, 209:160–166.

Q. Yuan, Z. He, and H. Leng. (2008) 'A hybrid genetic algorithm for a class of global optimization problems with box constraints.' *Applied Mathematics and Computation*, 197:924–929.

R.K. Mehra and R Davis.(1972) 'A generalized gradient method for optimal control problems with inequality constraints and singular arcs.' *IEEE Transactions on Automatic Control*, AC-17:69–79.

S. Karmakar, S.K. Mahato, and A.K. Bhunia.(2009) 'Interval oriented multi-section techniques for global optimization.'*Journal of Computational and Applied Mathematics*, 224:3476–491.

S. Delprat, J. Lauber, T.M. Guerra, and J. Rimaux.(1996) 'The gradient projection method for solving an optimal control problem.'*Applicationes Mathematica*, 24:141–147.

S. Delprat, J. Lauber, T.M. Guerra, and J. Rimaux.(2004) 'Control of a parallel hybrid powertrain: Optimal control.' *IEEE Transactions on Vehicular Technology*, 53:872–881.

Yi-Tung Kao and Erwie Zahara.(2008) 'A hybrid genetic algorithm and particle swarm optimization for multimodal functions.'*Appl. Soft Comput.*, 8:849–857.

Y.D. Kwon, S.B. Kwon, Seung-Bo J., and J.Y. Kim.(2003) 'Convergence enhanced genetic algorithm with successive zooming method for solving continuous optimization problems.'*Comput. Struct.*, 81:1715–1725.

Y. Dai, J. Yuan, and Y.X. Yuan.(2002) 'Modified two-point stepsize gradient method for unconstrained optimization.'*Comput. Optim. Appl*, 20:103-109.

Z. Xinchao.(2010) 'A perturbed particle swarm algorithm for numerical optimization.' *Applied Soft Computing*, 10:119–124.

Spall, J. C. (1992). Multivariate stochastic approximation using a simultaneous perturbation gradient approximation.*IEEE transactions on automatic control*, 37(3), 332–341.

Courant, R. (1943). Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American mathematical Society*, 49(1), 1-23.

Fiacco, A. V., & McCormick, G. P. (1966). Extensions of SUMT for nonlinear programming: equality constraints and extrapolation. *Management Science*, 12(11), 816-828.

Carroll, C. W. (1961). The created response surface technique for optimizing nonlinear, restrained systems. *Operations research*, 9(2), 169-184.

Peace, G. S. (1993). Taguchi methods: a hands-on approach. Addison Wesley Publishing Company.

Algorithm 1 Description of the NSG-SP method

Require: $\mathbf{x}_0 \in \mathbf{R}^n$, an integer $M \geq 1$, a small parameter $\alpha_{min} > 0$, a large parameter α_{max}, a sufficiently decreasing parameter $\gamma \in (0,1)$ and safeguarding parameters $0 < \sigma_1 < \sigma_2 < 1$. Initially, a, c, and γ, $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$ is arbitrary, a small number ϵ, a constant parameter β, and the initial and maximal value of the penalty parameter, respectively r; r_{max}.
Set $x_k = x_0,\ f = F_r(x_k),\ r = r_0,$ set $k = 1$

Apply SP procedure to obtain estimate gradient $\hat{g}_k$
Set $\alpha_k = 1/\|\hat{g}(\mathbf{x}_0)\|_\infty$.
while $\|\hat{g}_k\| \neq 0$ **do**
  Compute $d_k = -\alpha_k \hat{g}_k$.
  Set $\tilde{x} = \mathbf{x}_k + \lambda_k d_k$.
  if $P_r(\tilde{x}) \leq \epsilon$ and $f(\tilde{x}) < f(x_k)$ **then**
    if $F_r(\tilde{x}) < F_r(x_k)$ **then**
      $x_k = \tilde{x}$
    else
      $r = \min\{r_{\max}, \beta r\}$;
    end if;
  end if;
  if

$$f(\tilde{x}) \leq \max_{0 \leq j \leq \min\{k,M\}} f(\mathbf{x}_{k-j}) + \gamma \lambda g_k^T d_k \tag{34}$$

  then
    define $\lambda_k = \lambda$, $\mathbf{x}_{k+1} = \tilde{x}$, $s_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, $y_k = \hat{g}_{k+1} - \hat{g}_k$ and go to Step 18.
  else

$$\lambda_{new} \in [\sigma_1, \sigma_2 \lambda] \tag{35}$$

    set $\lambda = \lambda_{new}$ and go to Step 5.
  end if
  compute $b_k = s_k^T y_k$.
  if $b_k \leq 0$ **then**
    set $\alpha_{k+1} = \alpha_{max}$
  else
    compute $a_k = s_k^T s_k$ and $\alpha_{k+1} = \min\{\alpha_{\max}, \max\{\alpha_{\min}, a_k/b_k\}\}$
  end if
end while

Parameters/Level	1	2	3
M	5	10	20
c	0.0001	0.01	1
γ	0.0001	0.01	1

Table 1 NSGSP parameters and their levels

M	c	γ	value of $f(x)$	S/N ratio
5	0,0001	0,0001	0,3979	8,00452
5	0,0100	0,0100	0,3979	8,00452
5	1,0000	1,0000	0,6016	4,41384
10	0,0001	0,0100	0,3979	8,00452
10	0,0100	1,0000	0,4386	7,15863
10	1,0000	0,0001	0,4679	6,59694
20	0,0001	1,0000	0,3979	8,00452
20	0,0100	0,0001	0,3979	8,00452
20	1,0000	0,0100	0,4031	7,89174

Table 2 Tuning NSGSP parameters : S/N ratios values of experiments.

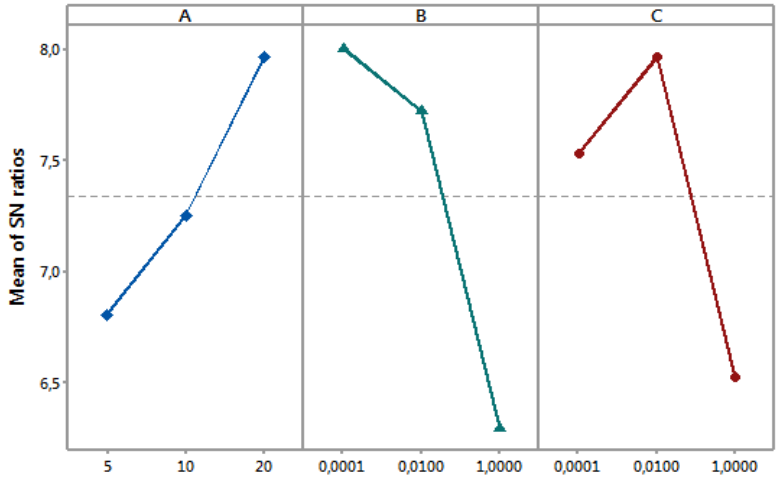

Figure 1: Main Effects Plot for SN ratios for the ($M,\ c,\ \gamma$) NSGSP parameters

Variance source	Degree of freedom (DF)	adjusted sum square	adjusted mean square	F-value	P-value	Contribution rate(%)
M	2	0,006576	0,003288	1,36	0,423	14,62
c	2	0,015131	0,007566	3,14	0,241	37,91
γ	2	0,010203	0,005102	2,12	0,321	24,50
Error	2	0,004817	0,002409			
Total	8	0,036728				

Table 3 Results of ANOVA : analysis of variance on the tuned parameters.

Table 4 Test functions

Test/Benchmark functions	Search region	Optimal value
Griewank's function		
$f_1 = 1 + \sum_{i=1}^{N} \frac{x_i^2}{4000} - \prod_{i=1}^{N} (\cos(x_i/\sqrt{i}))$	$x \in \mathbf{R}^N$	0
Rastrigin's function		
$f_2 = 10N + \sum_{i=1}^{N} [x_i^2 - 10\cos(2\pi x_i)]$	$x \in [-5,12; 5,12]^N$	0
Alluffi-Pentiny's function		
$f_3 = \frac{1}{4}x_1^4 - \frac{1}{2}x_1^2 + \frac{1}{10}x_1 + \frac{1}{2}x_2^2$	$x \in [-10; 10]^2$	-0.352386
Bohachevsky's function		
$f_4 = x_1^2 + 2x_2^2 - \frac{3}{10}\cos(3\pi x_1) - \frac{4}{10}\cos(4\pi x_2) + \frac{7}{10}$	$x \in [-100; 100]^2$	0
Branin function		
$f_5 = (x_2 - \frac{5.1}{4\pi}x_1^2 - \frac{5}{\pi}x_1 - 6) + 10(1 - \frac{1}{8\pi})\cos(x_1) + 10$	$x_1 \in [-5; 10], x_2 \in [0; 15]$	0.397887
Camel's function		
$f_6 = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^5 + x_1x_2 - 4x_2^2 + 4x_2^4$	$x \in [-5; 5]^2$	-1.0316
Shubert function		
$f_9 = -\sum_{i=1}^{2}\sum_{j=1}^{5} j\sin(x_i(j+1)) + 1$	$x \in [-10; 10]^2$	-24.06249
Pen Holder Function		
$f_{10} = -\exp\{-\mid \cos(x_1)\cos(x_2)e^{\mid 1-[(x_1^2+x_2^2)^{0.5}/\pi]\mid}\mid^{-1}\}$	$x \in [-11; 11]^2$	-0.963540
$f_{11} = \sum_{i=1}^{n}(x_i^4 - 16x_i^2 + 5x_i)$	$x \in [-4; 4]^n$	-78.33236
Cosine Mixture Function		
$f_{12} = \sum_{i=1}^{n} x_i^2 - 0.1\sum_{i=1}^{n}\cos(5\pi x_i)$	$x \in [-1; 1]^n$	-0.1n
Rosenbrock Function		
$f_{13} = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7$	$x \in \mathbf{R}^2$	0
Bird Function		
$f_{16} = \sin(x_1)\exp[(1-\cos(x_2))^2] + \cos(x_2)\exp[(1-\sin(x_1))^2] + (x_1 - x_2)^2$	$x \in [-2\pi, 2\pi]^2$	-106.764537
Sphere Function		
$f_{17} = \sum_{i=1}^{n} x_i^2$	$x \in [-5.12, 5.12]^n$	0
Zakharov Function		
$f_{18} = \sum_{i=1}^{n} x_i^2 + \left(\sum_{i=1}^{n} 0.5\ i\ x_i\right)^2 + \left(\sum_{i=1}^{n} 0.5\ i\ x_i\right)^4$	$x \in [-100, 100]^n$	0
Six Hump Camel Back Function		
$f_{19} = (4 - 2.1x_1^2 + \frac{x_1^4}{3})x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$	$x_1 \in [-3, 3], x_2 \in [-2, 2]$	-1.031600
Exponential function		
$f_{20} = \exp(0.5\sum_{i=1}^{n} x_i^2)$	$x \in [-1, 1]^n$	1.00000
Eason and Fenton Function		
$f_{21} = \frac{1}{10}\left(12 + x_1^2 + \frac{1+x_2^2}{x_1^2} + \frac{x_1^2x_2^2 + 100}{(x_1x_2)^4}\right)$	$x \in [0, 10]^2$	1.74

Table 5 Test functions

Test/Benchmark functions	Search region	Optimal value
Shekel 5 function		
$f_7 = \sum_{i=1}^{5} \frac{1}{(x-a_i)(x-a_i)^T+c_i}$ with $a = \begin{pmatrix} 4&4&4&4 \\ 1&1&1&1 \\ 8&8&8&8 \\ 6&6&6&6 \\ 3&7&3&7 \end{pmatrix}$ and $c = \begin{pmatrix} 0.1 \\ 0.2 \\ 0.2 \\ 0.4 \\ 0.4 \end{pmatrix}$	$x \in [0;10]^4$	-10.107749
Shekel 7 function		
$f_8 = \sum_{i=1}^{7} \frac{1}{(x-a_i)(x-a_i)^T+c_i}$ with $a = \begin{pmatrix} 4&4&4&4 \\ 1&1&1&1 \\ 8&8&8&8 \\ 6&6&6&6 \\ 3&7&3&7 \\ 2&9&2&9 \\ 5&3&5&3 \end{pmatrix}$ and $c = \begin{pmatrix} 0.1 \\ 0.2 \\ 0.2 \\ 0.4 \\ 0.4 \\ 0.6 \\ 0.3 \end{pmatrix}$	$x \in [0;10]^4$	-10.342378
Hartman$(H-3,4)$ **Function**		
$f_{14} = -\sum_{i=1}^{4} c_i \exp[-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2]$ with $a = \begin{pmatrix} 3.0&10.0&30.0 \\ 0.1&10.0&35.0 \\ 3.0&10.0&30.0 \\ 0.1&10.0&35.0 \end{pmatrix}$, $c = \begin{pmatrix} 0.1 \\ 1.2 \\ 3.0 \\ 3.2 \end{pmatrix}$ and $p = \begin{pmatrix} 0.3689&0.1170&0.2673 \\ 0.4699&0.4387&0.7470 \\ 0.1091&0.8732&0.5547 \\ 0.0381&0.5743&0.8827 \end{pmatrix}$	$x \in]0,1[^3$	-3.86278
Hartman$(H-6,4)$ **Function**		
$f_{15} = -\sum_{i=1}^{4} c_i \exp[-\sum_{j=1}^{6} a_{ij}(x_j - p_{ij})^2]$ with $a = \begin{pmatrix} 10.0&3.00&17.0&3.50&1.70&8.00 \\ 0.05&10.0&17.0&0.10&8.00&14.0 \\ 3.00&3.50&1.70&10.0&17.0&8.00 \\ 17.0&8.00&0.05&10.0&0.10&14.0 \end{pmatrix}$, $c = \begin{pmatrix} 0.1 \\ 1.2 \\ 3.0 \\ 3.2 \end{pmatrix}$ and $p = \begin{pmatrix} 0.1312&0.1696&0.5569&0.0124&0.8283&0.5886 \\ 0.2329&0.4135&0.8307&0.3736&0.1004&0.9991 \\ 0.2348&0.1451&0.3522&0.2883&0.3047&0.6650 \\ 0.4047&0.8828&0.8732&0.5743&0.1091&0.0381 \end{pmatrix}$	$x \in]0,1[^6$	-3.32237

Table 6 Optimization methods used for performance analysis

Method	Name (Reference)	Used test function
PGSL	Probabilistic global search lausanneRaphael and Smith (2003)	f_1
DE-PSO	Differential Evolution-Particle Swarm OptimizationZhang et al. (2009)	f_{18}
NHGA	Niche Hybrid Genetic AlgorithmWei and Zhao (2005)	f_{18}
ESGAT	simple genetic algorithmRaphael and Smith (2003)	f_1
Genitor	steady state genetic algorithmRaphael and Smith (2003)	f_1
pPSA	perturbed Particle Swarm AlgorithmXinchao (2010)	f_{17}
GAT	The Genetic Algorithm Toolbox in MATLABYuan et al. (2008)	f_2
HGA	Hybrid Genetic AlgorithmYuan et al. (2008)	f_2
MGA	Modified Genetic AlgorithmTsoulos (2008)	f_3,f_4,f_5,f_6,f_7,f_8,f_9
MACO	The modified ant colony optimizationToksary (2006)	f_{13}
SZGA	Successive zooming genetic algorithmKwon et al. (2003)	f_{13}
LX-PM	Laplace crossover with Power MutationDeep and Thakur (2007)	f_{17}
CTSS	Continuous Tabu Simplex Search	f_{18}
GA-PSO	Genetic algorithms and particle swarm optimizationKao and Zahara (2008)	f_{14},f_{15}

Table 7 Computational results of test functions

Problem no.	NSGSP method								Other Methods	
	Dimension	The Best	The Worst	Mean	SD	Median	SR	Cohen d	Method	Obtained result
1	10	0	2.2204e-15	6.8834e-17	9.7074e-32	0	100%	0.14584	ESGAT	0.0515
									Genitor	0.00496
									PGSL	0.0007
	20	0	4.4409e-16	9.1038e-17	1.3652e-32	0	100%	0.057556	ESGAT	0.0622
									Genitor	0.0240
									PGSL	0.0002
	50	1.1102e-16	1.8874e-15	5.9286e-16	1.5314e-31	4.4409e-16	100%	0	ESGAT	0.0990
									Genitor	0.0170
									PGSL	0.0
2	10	0	2.8422e-14	5.6843e-16	1.5833e-29	0	100%	0	HGA	1.29e-12
									GAT	1.37
	20	0	2.8422e-14	3.4106e-15	8.5303e-29	0	100%	0.29515	HGA	3.5e-12
									GAT	20.65
	50	0	5.6843e-13	1.2164e-13	1.5123e-26	1.1369e-13	100%	-0.10723	HGA	9.09e-12
3	2	-0.35239	-0.35239	-0.35239	1.4828e-31	-0.35239	100%	0	MGA	-0.352386
4	2	0	2.2204e-16	8.8818e-18	1.8933e-33	0	100%	0.28577	MGA	0
5	2	0.39789	0.39789	0.39789	3.469e-21	0.39789	100%	0.2092	MGA	0.397887
6	2	-1.0316	-1.0316	-1.0316	3.7964e-31	-1.0316	100%	0	MGA	-1.0316
7	4	-10.1532	-10.1532	-10.1532	3.96e-24	-10.1532	100%	0.1009	MGA	-10.107749
8	4	-10.4029	-10.4029	-10.4029	2.743e-24	-10.4029	100%	0.0118	MGA	-10.342378
9	2	-186.7309	-186.7309	-186.7309	1.5304e-13	-186.73	100%	0.245	MGA	-24.06249
10	2	-1	-0.99998	-1	3.6028e-11	-1	100%	-0.13288	Karmakar et al. (2009)	-0.963540
11	2	-78.3323	-78.332	-78.332	4.2005e-28	-78.332	100%	0	Karmakar et al. (2009)	-78.3323
12	6	-0.6	-0.6	-0.6	4.3289e-31	-0.6	100%	0	Karmakar et al. (2009)	-0.2
13	2	3.6778e-16	3.3111e-11	9.3907e-13	2.403e-23	4.8213e-16	100%	0.19809	SZGA	2.98e-8
								MACO	0	
14	3	-3.8634	-3.8634	-3.8634	4.5517e-30	-3.8634	100%	0	GA-PSO	-3.86278
15	6	-3.3224	-3.3224	-3.3224	9.4663e-32	-3.3224	100%	0	GA-PSO	-3.32237
16	2	-106.764	-106.76	-106.76	6.7855e-28	-106.76			(Karmakar et al. (2009))	-106.764537
17	30	6.035e-31	2.35e-26	1.08e-27	1.239e-53	5.493e-29	100%	0.1186	pPSA	5.6e-06
									LX-PM	4.75e-11
18	30	6.9125e-34	1.0671e-28	4.2324e-30	2.788e-58	7.9539e-32	100%	-0.17629	CTSS	3e-07
19	2	-1.0316	-1.0316	-1.0316	4.1908e-31	-1.0316	100%	-0.39693	(Karmakar et al. (2009))	-1.031628
20	10	1	1	1	4.1612e-31	1	100%	0	(Karmakar et al. (2009))	1.00000

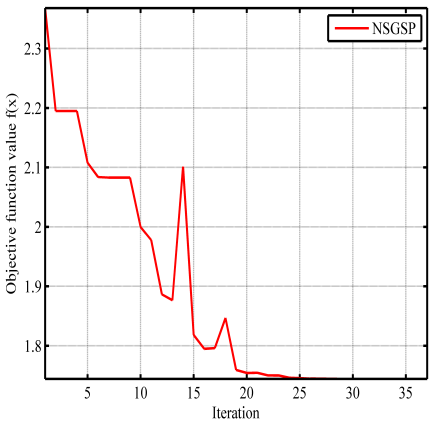

Figure 2: Eason and Fenton gear inertia problem. Minimum is $f(1.74, 2.02) = 1.74$

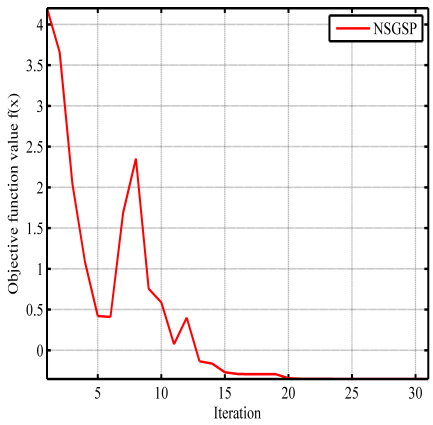

Figure 3: Aluffi-Pentini's function. Minimum is $f(-1.04, 0) = -0.35$

Problem	Ref.	Dimension	Optimal design variables(x)	Objective function value $f(x)$
1	Optimal solution	2	0.82288, 0.91144	1.3935
	Present method		0.82905,0.91255	1.3861
	de KS Lee et al.		0.8343, 0.9121	1.3770
	Homaifar et al.		0.8080, 0.8854	1.4339
	Fogel		0.8350, 0.9125	1.3772
2	Optimal solution	2	2.246826, 2.381865	13.59085
	Present study		2.2472, 2.3467	13.567
	de KS Lee et al.		2.246840, 2.382136	13.590845
	Deb GA with TS-R		Unavailable	13.59085
3	Optimal solution	5	78.0, 33.0, 29.995, 45.0, 36.775	-30665.500
	Present study		80.4786, 35.4786, 27.0000, 37.3972, 42.0962	-30666
	de KS Lee et al.		78.0, 33.0, 29.995, 45.0, 36.776	-30665.500
	Homaifar et al.		80.39, 35.07, 32.05, 40.33, 33.34	-30005.700
	Coello		78.0495, 33.007, 27.081, 45.00, 44.94	-31020.859
4	Optimal solution	7	2.330499, 1.951372,-0.4775414, 4.365726, -0.6244870, 1.038131, 1.594227	680.6300573
	Present study		2.2215, 1.9406, -0.4361, 4.4284, -0.5904, 1.1934, 1.4673	680.64
	de KS Lee et al.		2.32345617, 1.951242, -0.448467, 4.3619199, -0.630075, 1.03866, 1.605384	680.6413574
	Michalewicz		Unavailable	680.642
	Deb		Unavailable	680.63446
5	Optimal solution	10	2.1719, 2.3636, 8.7739, 5.9598, 0.9906, 1.4305, 1.3216, 9.8287, 8.28009, 8.3759	24.30620
	Present study		2.1975, 2.3053, 8.7582, 5.0963, 1.0535, 1.6020, 1.3681, 9.8739, 8.2892, 8.2043	24.314
	de KS Lee et al.		2.1552, 2.4076, 8.7780, 5.1020, 0.9676, 1.3576, 1.2877, 9.8004, 8.1878, 8.2562	24.3667
	Michalewicz		Unavailable	24.690
	Deb		Unavailable	24.37248

Table 8 Computational results of optimal solution and optimal value test problems

Problem no.		NSGSP method				
	Dimension	Optimal solution	The Best	The Worst	Mean	SD
1	10	1.3861	1.3861	1.3861	1.1559e-29	1.3861
2	2	13.592	13.567	13.532	8.32e-04	13.59
3	5	-30665	-30666	-30666	3.0882e-05	-30666
4	7	680.62	680.64	680.68	7.4412e-05	680.64
5	10	24.305	24.314	24.351	1.60e-05	24.307

Table 9 Computational results of constrained test functions

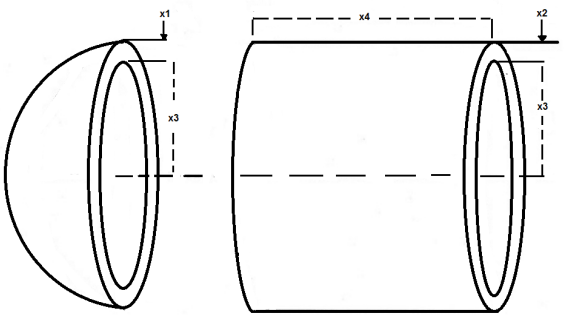

Figure 4: Shematic of the pressure vessel design problem

Pressure vessel design problem			
The best	The worst	Mean	SD
7195.8	7196.0	7195.9	0.0028

Table 10 Optimal solution of Pressure vessel design problem

Design variable	Present method	Sandgren	Wu and Chow	De KS Lee et al.
x_1	1.1332	1.125	1.125	1.125
x_2	0.6250	0.625	0.625	0.625
x_3	58.3267	48.97	58.1978	58.2789
x_4	43.4925	106.72	44.2930	43.7549
Cost	7195.8	7980.894	7207.494	7198.433

Table 11 Comparison results of Pressure vessel design problem

Figure 5: Tension-compression spring design

Tension-compression spring design				
Parameter	x_1	x_2	x_3	$f(x)$
Value	0.058324	0.604	3.7626	0.011868

Table 12 Optimal solution of Tension-compression spring design

Methods	The best	The worst	Mean	SD
NSGSP	0.011868	0.011868	0.011868	3.0093e-36
GWO	0.0126660	0.0122515	0.0121836	1.085e-05
CSA	0.0126652	0.0126701	0.0127690	1.357e-06
CPSO	0.0126747	0.0129240	0.0127330	5.20e-04
HPSO	0.0126652	0.0127190	0.0127072	1.58e-05

Table 13 Comparison results of the Tension-compression spring design problem

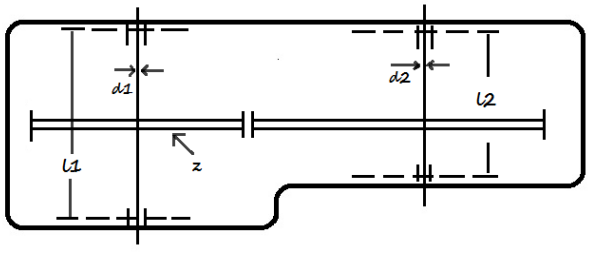

Figure 6: Speed reducer design

Speed reducer design problem								
Parameter	x_1	x_2	x_3	x_4	x_5	x_6	x_7	$f(x)$
Value	3.600	0.700	19.069	7.300	7.800	3.1651	5.000	2995.9

Table 14 Optimal solution of Speed reducer design problem

Methods	The best	The worst	Mean	SD
NSGSP	2995.9	2996.1	2996	0.0028
Baykasoglu and Ozsoydan	2996.37	2996.66	2996.51	0.09
Mezura-Montes et al.	2998.01	3162.88	3056.20	49.40
Aguirre et al.	2996.34	-	2996.40	0.028
Cagnina et al.	2996.34	-	2996.34	0.0

Table 15 Comparison results of Speed reducer design problem

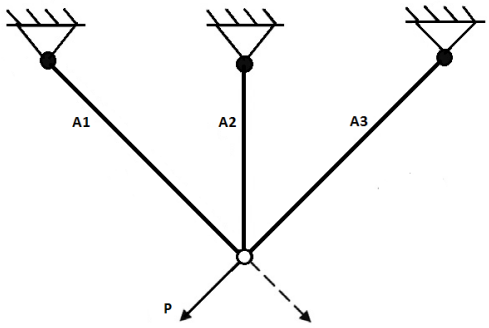

Figure 7: Three bar truss

Three bar Truss			
The best	The worst	Mean	SD
263.6789	263.6789	263.6789	0

Table 16 Optimal solution of Three bar Truss

Design variable	Present method	$\epsilon DE - PCGA$	Ray and Saini	Tsai	Gandomi et al.
A_1	0.7874	0.788675	0.795	0.788	0.78867
A_2	0.4076	0.408248	0.395	0.408	0.40902
Weight	263.6789	263.89584	264.3	263.68	263.9716

Table 17 Comparison results of Three bar Truss