



Università
Ca' Foscari
Venezia

**Department
of Management**

Working Paper Series

E.F. Campana, M. Diez, G. Fasano and D. Peri

**Initial particles position for
PSO, in Bound Constrained
Optimization**

**Working Paper n. 6/2013
June 2013**

ISSN: 2239-2734



This Working Paper is published under the auspices of the Department of Management at Università Ca' Foscari Venezia. Opinions expressed herein are those of the authors and not those of the Department or the University. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional nature.

Initial particles position for PSO, in Bound Constrained Optimization

Emilio F. Campana¹, Matteo Diez¹, Giovanni Fasano², and Daniele Peri¹

¹ National Research Council-Maritime Research Centre (CNR-INSEAN),
Via di Vallerano, 139, 00128 Rome, Italy

(emiliofortunato.campana,matteo.diez,daniele.peri)@cnr.it,

² University Ca'Foscari of Venice, Department of Management,
fasano@unive.it

Abstract. We consider the solution of bound constrained optimization problems, where we assume that the evaluation of the objective function is costly, its derivatives are unavailable and the use of exact derivative-free algorithms may imply a too large computational burden. There is plenty of real applications, e.g. several design optimization problems [1, 2], belonging to the latter class, where the objective function must be treated as a 'black-box' and automatic differentiation turns to be unsuitable. Since the objective function is often obtained as the result of a simulation, it might be affected also by noise, so that the use of finite differences may be definitely harmful.

In this paper we consider the use of the evolutionary Particle Swarm Optimization (PSO) algorithm, where the choice of the parameters is inspired by [4], in order to avoid diverging trajectories of the particles, and help the exploration of the feasible set. Moreover, we extend the ideas in [4] and propose a specific set of initial particles position for the bound constrained problem.

Keywords: Bound Constrained Optimization, Discrete Dynamic Linear Systems, Free and Forced Responses, Particles Initial Position.

1 Introduction

Applied sciences offer several challenging applications of bound constrained optimization, where the computational cost of the objective function is remarkably large. In this regard, optimization tools combining the theoretical properties of exact methods and the fast progress of heuristics represent an active research area. Furthermore, on large scale real problems, which are typically more difficult and require correspondingly larger computational resources, both practitioners and theoreticians claim for robust methods, often endowed also with theoretical properties. Moreover, in many cases the derivatives are unavailable. In the latter case, the use of 'black-box' simulations for computing the objective function makes the adoption of automatic differentiation impossible, due to the unavailability of the source code. In addition, simulations represent an essential tool,

but often introduce an unexpected artificial noise, which unavoidably imposes strong care when adopting finite differences.

This paper considers PSO [6], with a specific choice of the parameters, for the solution of the bound constrained *global optimization problem*

$$\min_{x \in \mathcal{F}} f(x), \quad f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad (1)$$

where $\mathcal{F} = \{x \in \mathbb{R}^n : l \leq x \leq u\}$, $l, u \in \mathbb{R}^n$ and without loss of generality $l < u$. Obviously, in case $l_i = -\infty$, $i \in \{1, \dots, n\}$ and $u_i = +\infty$, $i \in \{1, \dots, n\}$ problem (1) reduces to an unconstrained optimization problem. At present $f(x)$ is assumed to be a nonlinear and non-convex *continuous function*.

This paper has a twofold purpose. First we propose some novel rules for the selection of parameters in PSO, using the reformulation of PSO iteration described in [4]. Then, we suitably adapt the choice of particles position/velocity studied in [4] for the unconstrained case, to the feasible set \mathcal{F} of (1). The latter adaptation requires some geometric insight and involves a negligibly small algebra, even when the scale n is large.

As regards the symbols used in this paper, the subscripts identify the particles in a PSO scheme, while we use the superscript to indicate the iteration. I_k is the identity matrix of order k . If σ is a real random unknown and $u \in \mathbb{R}^n$, the symbol $\sigma \otimes u$ indicates an n -real vector, whose j -th and i -th entries are respectively given by $\sigma_j u_j$ and $\sigma_i u_i$, where σ_j and σ_i are different occurrences of σ . Finally, $\|A\|_F$ indicates the Frobenius norm of matrix A , i.e. $\|A\|_F = \text{tr}(A^T A)^{1/2}$, where $\text{tr}(\cdot)$ indicates the *trace* of a matrix.

In Section 2 we propose a reformulation of PSO iteration, which is essential for our proposal, then Section 3 suggests some basics on the choice of parameters in PSO, and Section 4 proposes some indications to properly choose the initial position/velocity of particles for problem (1).

2 A reformulation of PSO

Consider the trajectory of the j -th PSO iteration ($k \geq 0$)

$$\begin{aligned} v_j^{k+1} &= \chi_j^k [w_j^k v_j^k + c_j r_j \otimes (p_j^k - x_j^k) + c_g r_g \otimes (p_g^k - x_j^k)], \\ x_j^{k+1} &= x_j^k + v_j^{k+1}, \end{aligned} \quad (2)$$

where $j = 1, \dots, P$ indicates the j -th *particle* and P is a positive integer. The vectors v_j^k and x_j^k are n -real vectors representing respectively the *velocity* (i.e. the search direction) and the *position* of the j -th particle at step k . Moreover, the n -real vectors p_j^k and p_g^k satisfy

$$\begin{aligned} f(p_j^k) &\leq f(x_j^\ell), & \text{for any } \ell \leq k, \quad p_j^k \in \{x_j^\ell\}, \\ f(p_g^k) &\leq f(x_j^\ell), & \text{for any } \ell \leq k \text{ and } j = 1, \dots, P, \quad p_g^k \in \{x_j^\ell\}, \end{aligned} \quad (3)$$

while $\chi_j^k, w_j^k, c_j, r_j, c_g, r_g$ are positive bounded coefficients. As well known, p_j^k represents the ‘best position’ in the trajectory of the j -th particle up to step k , while p_g^k is the ‘best position’ among all the particles up to step k . The choice of the coefficients as well as the number of particles P is often problem dependent (see also [7]), and here we consider the choice [4], which is very general. The latter choice also includes the case where both the *inertia* coefficient w_j^k and the *constriction* coefficient χ_j^k are used. Finally, as usually, we can assume without loss of generality that r_j and r_g are uniformly distributed random parameters, with $r_j \in [0, 1]$ and $r_g \in [0, 1]$.

After some simplifications, for each particle j , assuming for brevity that $w_j^k = w_j$ and $\chi_j^k = \chi_j$, for any $k \geq 0$ the iteration (2) is equivalent to the *discrete stationary (time-invariant) system* (see also [4])

$$X_j(k+1) = \begin{bmatrix} \chi_j w_j I_n & -\chi_j (c_j r_j + c_g r_g) I_n \\ \chi_j w_j I_n & [1 - \chi_j (c_j r_j + c_g r_g)] I_n \end{bmatrix} X_j(k) + \begin{bmatrix} \chi_j (c_j r_j p_j^k + c_g r_g p_g^k) \\ \chi_j (c_j r_j p_j^k + c_g r_g p_g^k) \end{bmatrix} \quad (4)$$

where

$$X_j(k) = \begin{pmatrix} v_j^k \\ x_j^k \end{pmatrix} \in \mathbb{R}^{2n}, \quad k \geq 0. \quad (5)$$

From a geometric perspective the sequence $\{X_j(k)\}$ represents the trajectory of the j -th particle in the state space \mathbb{R}^{2n} . Moreover, using a standard notation for linear systems, we can split $X_j(k)$ into the *free response* $X_{jL}(k)$ and the *forced response* $X_{jF}(k)$ (see also [8]). Thus, on summary for any $k \geq 0$ the $2n$ -real vector $X_j(k)$ may be rewritten as

$$X_j(k) = X_{jL}(k) + X_{jF}(k), \quad (6)$$

where

$$X_{jL}(k) = \Phi_j(k) X_j(0), \quad X_{jF}(k) = \sum_{\tau=0}^{k-1} H_j(k-\tau) U_j(\tau), \quad (7)$$

and after some computation we obtain (see also [4])

$$\begin{aligned} \Phi_j(k) &= \begin{pmatrix} \chi_j w_j I_n & -\chi_j (c_j r_j + c_g r_g) I_n \\ \chi_j w_j I_n & [1 - \chi_j (c_j r_j + c_g r_g)] I_n \end{pmatrix}^k \in \mathbb{R}^{2n \times 2n}, \quad (8) \\ H_j(k-\tau) &= \begin{pmatrix} \chi_j w_j I_n & -\chi_j (c_j r_j + c_g r_g) I_n \\ \chi_j w_j I_n & [1 - \chi_j (c_j r_j + c_g r_g)] I_n \end{pmatrix}^{k-\tau-1} \in \mathbb{R}^{2n \times 2n}, \\ U_j(\tau) &= \begin{pmatrix} \chi_j (c_j r_j p_j^\tau + c_g r_g p_g^\tau) \\ \chi_j (c_j r_j p_j^\tau + c_g r_g p_g^\tau) \end{pmatrix} \in \mathbb{R}^{2n}. \end{aligned}$$

We urge to recall that from the expressions (6)-(7), unlike the vector $X_{jF}(k)$, the free response $X_{jL}(k)$ only depends on the initial point $X_j(0)$, and not on the vectors p_j^τ, p_g^τ , with $\tau \geq 0$. As described in the next section, the latter observation plays a keynote role, in order to design efficient PSO schemes for solving (1).

3 Issues on parameters assessment in PSO

Observe from (8) that $\Phi_j(k) = \Phi_j(1)^k$, for any $k \geq 0$, and the $2n$ eigenvalues of the unsymmetric matrix $\Phi_j(1)$ are real (see also [4]). Setting for simplicity in (8)

$$a_j = \chi_j w_j, \quad \omega_j = \chi_j(c_j r_j + c_g r_g), \quad j = 1, \dots, P, \quad (9)$$

after some computation we see that the matrix $\Phi_j(1)$ has only the two distinct eigenvalues λ_{j1} and λ_{j2} given by

$$\begin{aligned} \lambda_{j1} &= \frac{1 - \omega_j + a_j - [(1 - \omega_j + a_j)^2 - 4a_j]^{1/2}}{2} \\ \lambda_{j2} &= \frac{1 - \omega_j + a_j + [(1 - \omega_j + a_j)^2 - 4a_j]^{1/2}}{2}, \end{aligned} \quad (10)$$

each of them having algebraic multiplicity n . A necessary (but possibly not sufficient) condition for $\{X_j(k)\}$ to be non-diverging (which implies that also $\{x_j^k\}$ and $\{v_j^k\}$ in (2) are non-diverging), is

$$|\lambda_{j1}| < 1, \quad |\lambda_{j2}| < 1, \quad (11)$$

which affect the choice of PSO parameters as described in the next proposition (the next conditions are simplified with respect to [4]).

Proposition 1. *Consider the position (9) in (2), with $\chi_j^k = \chi_j$ and $w_j^k = w_j$, $j = 1, \dots, P$. Suppose for $k \geq 0$*

$$0 < a_j < 1, \quad 0 < \omega_j < 2(a_j + 1), \quad j = 1, \dots, P, \quad (12)$$

with $\omega_j^k \neq (1 \pm a_j^{1/2})^2$. Then, for any $k \geq 0$ and $j = 1, \dots, P$, conditions (11) are fulfilled. \diamond

Observe that conditions (11) imply $\lim_{k \rightarrow \infty} X_{jL}(k) = 0$, $j = 1, \dots, P$, and most of the typical settings for PSO parameters proposed in the literature (see e.g. [7, 9]) satisfy (12). Moreover, from relations (7), (8), (9) and considering that $\Phi_j(1)$ is unsymmetric, we have also that for any j

$$\|\Phi_j(k)\|_F \leq \|\Phi_j(1)\|_F^k = \text{tr} [\Phi_j(1)^T \Phi_j(1)]^{\frac{k}{2}}, \quad (13)$$

and

$$\begin{aligned} \text{tr} [\Phi_j(1)^T \Phi_j(1)]^{\frac{1}{2}} &= \text{tr} \begin{bmatrix} 2a_j^2 I_n & a_j(1 - 2\omega_j)I_n \\ a_j(1 - 2\omega_j)I_n & [\omega_j^2 + (1 - \omega_j)^2]I_n \end{bmatrix}^{\frac{1}{2}} \\ &= [2a_j^2 + \omega_j^2 + (1 - \omega_j)^2]^{\frac{1}{2}}. \end{aligned}$$

Using Fact 9.12.1 in [10] (where $B = I_n$ and $\|B\|_F = \sqrt{n}$) we have that

$$\frac{1}{\sqrt{n}} |\text{tr} [\Phi_j(1)]| \leq \|\Phi_j(1)\|_F \quad (14)$$

where

$$\text{tr} [\Phi_j(1)] = a_j + (1 - \omega_j).$$

Now, from (7) and (13)

$$\|X_{jL}(k)\|_F \leq \|\Phi_j(1)\|_F^k \cdot \|X_j(0)\|_F,$$

and though $\lim_{k \rightarrow \infty} X_{jL}(k) = 0$, $j = 1, \dots, P$, we would like $\|X_{jL}(k)\|_F$ not to be attenuated when the index k is still relatively small. On this purpose, given the coefficients c_j , $j = 1, \dots, P$ and c_g , we propose to set χ_j and w_j by solving for each $j = 1, \dots, P$ one of the following two programs, inspired by Proposition 1 and, respectively, relation (13) and relation (14):

$$\begin{aligned} \max_{\chi_j, w_j} & 2a_j^2 + \omega_j^2 + (1 - \omega_j)^2 \\ & 0 < a_j < 1, \\ & 0 < \omega_j < 2(a_j + 1), \end{aligned} \quad (15)$$

$$\begin{aligned} \max_{\chi_j, w_j} & |a_j + (1 - \omega_j)| \\ & 0 < a_j < 1, \\ & 0 < \omega_j < 2(a_j + 1). \end{aligned} \quad (16)$$

The programs (15)-(16) attempt to possibly force larger values of $\|X_{jL}(k)\|_F$ for k small. In Section 4 we give more motivations about the latter issue.

Now, in the light of (7), (11) and the results in Proposition 1, we think that the following question still deserves special consideration: *can we properly choose the initial points $X_j(0)$, $j = 1, \dots, P$, for problem (1), so that the trajectories $\{x_j^k\}$ span as much as possible the feasible set \mathcal{F} ?* Section 4 addresses the latter issue, in order to give indications on the choice of the initial point and velocity of particles.

4 Initial particles position and velocity in PSO, for bound constrained optimization

In this section we study some proposals of initial particles position and velocity, for the bound constrained optimization problem (1). To this aim let us consider

the feasible set \mathcal{F} in (1); we remind that possibly we allow $l_i = -\infty$ and/or $u_i = +\infty$ for some indices $i \in \{1, \dots, n\}$. In the previous section we studied settings for PSO parameters, such that the free response $X_{jL}(k)$ associated to particle j is possibly not attenuated too early, i.e. when k is still relatively small. In this section we show a method to exploit the latter property, in order to possibly improve the overall performance of PSO on bound constrained optimization. In particular, we want to give indications for the choice of the vectors $X_j(0)$, so that possibly the orthogonality conditions (or similar properties)

$$X_{j_iL}(k)^T X_{j_hL}(k) = 0, \quad 1 \leq i \neq h \leq m, \quad (17)$$

among the free responses of the first m particles (with $m \leq n$), are satisfied. Observe that conditions (17) do not impose the trajectories of PSO particles to be orthogonal; however, they guarantee that *part of particles trajectories* (i.e. the free responses in \mathbb{R}^{2n}) are orthogonal, as long as they do not fade. This explains why in Section 3 we studied conditions on PSO parameters, in order to prevent a premature extinction of $X_{jL}(k)$ when k increases.

In particular, our first proposal for the choice of $X_j(0)$, $j = 1, \dots, P$, is the following:

1. If $l < 0 < u$ then set $X_j(0)$ such that $x_j^0 \in \mathcal{F}$, randomly for $j = n+1, \dots, P$, and $v_j^0 \in \mathbb{R}^n$ for $j = n+1, \dots, P$. On the other hand, for $j = 1, \dots, n$ set

$$t_j = \left[\frac{\sqrt{n}}{n} \sum_{i=1}^n -\frac{\sqrt{n}}{2} e_j \right] \in \mathbb{R}^n, \quad X_j(0) = \begin{pmatrix} \alpha_j t_j \\ \beta_j t_j \end{pmatrix} \in \mathbb{R}^{2n}, \quad (18)$$

where α_j is any real value such that $\alpha_j t_j \in \mathcal{F}$, $j = 1, \dots, n$.

2. Otherwise, set $X_j(0)$ such that $x_j^0 \in \mathcal{F}$, randomly for $j = n+1, \dots, P$, and $v_j^0 \in \mathbb{R}^n$ for $j = n+1, \dots, P$. Then, for $j = 1, \dots, n$ consider the vertex $\hat{u} \in \mathcal{F}$ which is the closest to the origin; take

$$X_j(0) = \begin{pmatrix} \hat{u}_j \\ z_j \end{pmatrix}, \quad j = 1, \dots, n, \quad (19)$$

\hat{u}_j being the j -th vertex of \mathcal{F} adjacent to \hat{u} (i.e. such that an edge of \mathcal{F} connects \hat{u} and \hat{u}_j), and $z_j \in \mathbb{R}^n$ is randomly chosen.

Observe that while (18) satisfies (17) and α_j is very easy to compute, the choice (19) simply ensures that the vectors $X_j(0)$, $j = 1, \dots, n$, are at least linearly independent (though in general not orthogonal). Now, in order to force condition (17) (or similar conditions) in a more general framework, let us consider the geometry of the feasible set \mathcal{F} (shaded area) in Fig.1. Suppose the point c is the intersection of the diagonals of \mathcal{F} , i.e. $c = (u+l)/2$, and the segment a_i is given by $a_i = (u_i - l_i)/2$, $i = 1, \dots, n$. We want to compute the equations of the dashed hyperellipsoids E_0 , E_1 and E_2 in Fig.1, E_0 being a sphere. It is not difficult to realize that

Table 1. We list the results for 6 test functions from the literature (n is the number of unknowns and f^* is the value of f at a global minimum). The results are over 25 PSO runs, $f_bst/f_wst/f_av$ is the *best/worst/average* value of f over the 25 runs, while *st. dev.* indicates the *standard deviation*. x_{rand} indicates random initial choice for particles position, while x_{orth} indicates initial choice for particles position as in (18).

Function		x_{rand}	x_{orth}	Function		x_{rand}	x_{orth}
Griewank (n=10)	f^*	0.0000		Levy 10 ⁿ loc.min. (n=30)	f^*	0.0000	
	f_bst	0.5562	0.0057		f_bst	53.8192	1.1428
	f_av	0.8485	0.0332		f_av	107.3033	3.4678
	f_wst	1.1650	0.0731		f_wst	299.5744	3.9709
	<i>st. dev.</i>	0.0067	0.0004		<i>st. dev.</i>	0.0001	0.0000
Griewank (n=20)	f^*	0.0000		Levy 15 ⁿ loc.min. (n=30)	f^*	0.0000	
	f_bst	1.2360	0.0016		f_bst	14.4646	3.1471
	f_av	1.3872	0.0022		f_av	31.7934	3.3890
	f_wst	1.7438	0.0653		f_wst	60.5632	3.5046
	<i>st. dev.</i>	0.0001	0.0000		<i>st. dev.</i>	0.0002	0.0000
Levy 5 ⁿ loc.min. (n=30)	f^*	0.0000		Griewank (n=30)	f^*	0.0000	
	f_bst	3.0273	0.0268		f_bst	1.5631	0.0007
	f_av	8.9546	0.0483		f_av	2.1459	0.0389
	f_wst	13.6678	0.0942		f_wst	2.8092	0.0710
	<i>st. dev.</i>	0.0000	0.0000		<i>st. dev.</i>	0.0000	0.0000

$$\begin{aligned}
E_0 & : (x - c)^T A_0 (x - c) = 1, & A_0 & = \text{diag}_{1 \leq i \leq n} \left\{ \left(\sum_{i=1}^n a_i^2 \right)^{-1} \right\}, \\
E_1 & : (x - c)^T A_1 (x - c) = n, & A_1 & = \text{diag}_{1 \leq i \leq n} \left\{ a_i^{-2} \right\}, \\
E_2 & : (x - c)^T A_1 (x - c) = 1;
\end{aligned} \tag{20}$$

indeed, it suffices to consider that E_0 is a sphere, the extreme points v_ℓ , $\ell = 1, \dots, 2^n$, in the corners of \mathcal{F} have coordinates $c_i \pm a_i$, $i = 1, \dots, n$ (which satisfy the first two equations (20)), and the centers of the facets of \mathcal{F} have entries in the sets $\{c_i, l_i, u_i\}$, $i = 1, \dots, n$. We would like to show that for problem (1) it is possible to set $X_j(0)$, $j = 1, \dots, n$ (other than (18)-(19)), so that conditions (17) at least in some cases are satisfied, with $X_j(0)$ such that $x_j^0 \in V$, where V is the region inside either of the hyperellipsoids E_0 , E_1 or E_2 . The importance of the latter property relies on the fact that it *tries to force orthogonality* among particles trajectories, while particles move within \mathcal{F} . Thus, we expect that PSO will be able to explore the feasible region of interest \mathcal{F} , as accurately as possible, while possibly ignoring the exploration in the set $\mathbb{R}^n \setminus V$.

In a more general scheme where \mathcal{F} is treated in a penalty framework (i.e. PSO is used for the unconstrained minimization of a penalty function, which is the sum of $f(x)$ and a term penalizing the constraints violation), then we can set $X_j(0)$, $j = 1, \dots, P$, in a different fashion with respect to **1.** and **2.** Indeed, we can consider the choice:

1. If $l < 0 < u$ then set $X_j(0)$ such that $x_j^0 \in \mathcal{F}$, randomly for $j = n + 1, \dots, P$, and $v_j^0 \in \mathbb{R}^n$ for $j = n + 1, \dots, P$. On the other hand, from (18) set

$$X_j(0) = \begin{pmatrix} \alpha_j t_j \\ \beta_j t_j \end{pmatrix} \in \mathbb{R}^{2n}, \quad j = 1, \dots, n,$$

where now α_j is any real value such that $\alpha_j t_j \in V$, $j = 1, \dots, n$, and V is the region inside either of the hyperellipsoids E_0 , E_1 or E_2 in Fig.1.

2. Otherwise, take the choice 2.

We still have to complete in a separate paper a numerical experience, giving full evidence of the effectiveness of the proposals above, in a framework where exact penalty methods are adopted. However, Table 1 summarizes a few preliminary results on six test problems from the literature (the caption describes the setting adopted), indicating that our proposal might be effective and efficient.

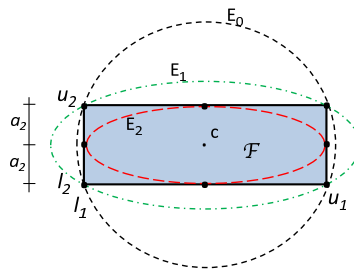


Fig. 1. The feasible set $\mathcal{F} \subset V$ of (1), V is the region inside E_0 or E_1 , and $V \supset E_2$.

References

1. Mohammadi, B., Pironneau, O.: Applied Shape Optimization for Fluids, Clarendon Press, Oxford (2001)
2. Haslinger, J., Mäkinen, R.A.E.: Introduction to Shape Optimization, SIAM Advances in Design and Control, Philadelphia (2003)
3. Pinter, J.D.: Global Optimization in Action. Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications, Kluwer Academic Publishers, The Netherlands (1996)
4. Campana, E.F., Fasano, G., Pinto, A.: Dynamic analysis for the selection of parameters and initial population, in particle swarm optimization, Journal of Global Optimization 48, 347–397 (2010)
5. Campana, E.F., Fasano, G., Peri, D.: Globally Convergent Modifications of Particle Swarm Optimization for Unconstrained Optimization. In: Andrea E. Olsson (ed.) Particle Swarm Optimization: Theory, Techniques and Applications, Nova Publishers Inc., Series: Advances in Engineering Mechanics, pp. 97–118, South Africa (2011)
6. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In Proceedings of the 1995 IEEE International Conference on Neural Networks IV, pp. 1942–1948. IEEE Service Center, Piscataway, Perth, NJ, (1995)

7. Clerc, M., Kennedy, J.: The Particle Swarm - Explosion, Stability, and Convergence in a Multidimensional Complex Space, *IEEE Transactions on Evolutionary Computation* 6, 58–73, (2002)
8. Sarachik, P.E.: Principles of linear systems, Cambridge University Press, Cambridge (1997)
9. Zheng, Y.L., Ma, L.H., Zhang, L.Y., Qian, J.X.: On the convergence analysis and parameter selection in particle swarm optimization. In Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2-5 November (2003)
10. Bernstein, D.S.: Matrix Mathematics: Theory, Facts, and Formulas – Second Edition, Princeton University Press, NJ (2009)