Dense Orthogonal Initialization for Deterministic PSO: `ORTHOinit+`

Matteo Diez[1], Andrea Serani[1] and Cecilia Leotardi[1], Emilio Fortunato Campana[1], Giovanni Fasano[2], and Riccardo Gusso[2]

[1] National Research Council–Marine Technology Research Institute (CNR-INSEAN),
Via di Vallerano, 139, 00128 Rome, Italy
(`matteo.diez,emiliofortunato.campana`)`@cnr.it`;
(`andrea.serani,cecilia.leotardi`)`@insean.cnr.it`
[2] University Ca'Foscari of Venice, Department of Management,
S.Giobbe, Cannaregio 873, 30121 Venice, Italy
(`fasano,rgusso`)`@unive.it`

Abstract. This paper describes a novel initialization for Deterministic Particle Swarm Optimization (DPSO), based on choosing specific dense initial positions and velocities for particles. This choice tends to induce orthogonality of particles' trajectories, in the early iterations, in order to better explore the search space. Our proposal represents an improvement, by the same authors, of the theoretical analysis on a previously proposed PSO reformulation, namely the initialization `ORTHOinit`. A preliminary experience on constrained Portfolio Selection problems confirms our expectations.

Keywords: Global Optimization, Deterministic PSO, Particles Initial Position and Velocity

1 Introduction

Particle Swarm Optimization (PSO) is a stochastic approach for the solution of global optimization problems [1]. A population of particles is initialized in PSO, and their trajectories in the search space explore potential solutions, in order to approach a global minimum. There are several convergence studies on PSO in the literature, focusing on the role of PSO control parameters, as for instance swarm size, inertia weight, acceleration coefficients, velocity clamping, as well as particles' initialization. All these studies reveal that PSO parameters must be confined to specific subsets of values, in order to avoid diverging trajectories, and some of them provide necessary conditions for the trajectories convergence (see for instance [2–6], along with [7–10]). Improper initializations of PSO parameters may yield divergent or even cyclic behavior. The initial position and velocity of particles may have dramatic consequences on convergence. Moreover, effective particles initializations are often dependent on the current problem structure (see for instance [4]).

Finally, we consider the deterministic version of PSO, namely DPSO. Earlier studies (see for instance [7–9]) show that coupling proper control parameters

with an efficient particles initialization may give strong synergies and improve the algorithm effectiveness. On this guideline, here we want to propose a novel particles initialization, for DPSO, showing the following two features:

(i) we can scatter particles trajectories, at least in the early iterations, in order to better explore the search space
(ii) the resulting initialization is *dense*, i.e. a large portion of the entries of particles position and velocity is nonzero. On some problems (i.e. the portfolio selection problem we report) the latter event partially allows particles trajectory to avoid too sparse approximate solutions.

2 Preliminaries on DPSO iteration

Let us consider the following PSO iteration

$$
\begin{cases}
v_j^{k+1} = \chi \left[w^k v_j^k + c_j^k r_j^k (p_j^k - x_j^k) + c_g^k r_g^k (p_g^k - x_j^k) \right], & k \geq 0, \\
x_j^{k+1} = x_j^k + v_j^{k+1}, & k \geq 0.
\end{cases}
\tag{1}
$$

Here $j = 1, ..., P$ represents the j-th particle of the swarm, $v_j^k \in \mathbb{R}^n$ and $x_j^k \in \mathbb{R}^n$ are respectively the vector of *velocity* and the vector of *position* of particle j at step k, whereas $\chi, w^k, c_j^k, r_j^k, c_g^k, r_g^k$ are suitable positive bounded coefficients. Finally, p_j^k is the best position outreached by particle j up to step k, and p_g^k represents the best particle position of the overall swarm, up to step k.

Assumption 1 (DPSO) *Let us be given the iteration (1) and the positive constant values c, r, $\bar{c}$, $\bar{r}$, w. Then, for any $k \geq 0$ and $j = 1, ..., P$, we assume that $c_j^k = c$, $r_j^k = r = 1$, $c_g^k = \bar{c}$, $r_g^k = \bar{r} = 1$ and $w^k = w$.*

As it was proved in [7], setting $a = \chi w$ and $\omega = \chi(cr + \bar{c}\bar{r})$, with $\omega \neq (1 \pm \sqrt{a})^2$, we can consider the distinct real eigenvalues λ_1 and λ_2 of matrix

$$
\Phi(k) = \begin{pmatrix} aI & -\omega I \\ aI & (1-\omega)I \end{pmatrix}.
$$

Then, we can set at step k of DPSO iteration the two nonzero parameters

$$
\gamma_1(k) = \frac{\lambda_1^k(a - \lambda_2) - \lambda_2^k(a - \lambda_1)}{\lambda_1 - \lambda_2} \qquad \gamma_2(k) = \frac{\omega(\lambda_1^k - \lambda_2^k)}{\lambda_1 - \lambda_2},
\tag{2}
$$

along with the $2n$ vectors

$$
z_i(k) = \begin{pmatrix} \frac{\gamma_2(k)}{\gamma_1(k)} e_i \\ e_i \end{pmatrix} \in \mathbb{R}^{2n}, \qquad i = 1, \ldots, n,
\tag{3}
$$

$$
z_{n+i}(k) = \begin{pmatrix} -\frac{\gamma_1(k)}{\gamma_2(k)} e_i \\ e_i \end{pmatrix} \in \mathbb{R}^{2n}, \qquad i = 1, \ldots, n,
\tag{4}
$$

where $e_i \in \mathbb{R}^n$ is the i-th unit vector of $\mathbb{R}^n$. When $k = 0$ the vectors defined in (3)-(4) can be used to fruitfully set the initial particles position and velocity (respectively of the i-th and $(n+i)$-th particle), according with the initializations

$$\begin{pmatrix} v_i^0 \\ \\ x_i^0 \end{pmatrix} = \rho_i^1 z_i(0), \qquad \rho_i^1 \in \mathbb{R} \setminus \{0\}, \ i = 1, \ldots, n \tag{5}$$

$$\begin{pmatrix} v_{n+i}^0 \\ \\ x_{n+i}^0 \end{pmatrix} = \rho_i^2 z_{n+i}(0), \qquad \rho_i^2 \in \mathbb{R} \setminus \{0\}, \ i = 1, \ldots, n. \tag{6}$$

The latter initializations reveal interesting properties, and proved to be effective on several practical problems. As shown in [7], in case Assumption 1 holds (which also implies that a DPSO iteration is considered), then

1. as long as $P \leq 2n$, setting the sequences $\{v_i^0\}$ and $\{x_i^0\}$ as in (5)-(6), guarantees that the particles trajectories are nearly orthogonal at step 0 (see [7] for details), and the latter property is in practice likely maintained also in the subsequent early iterations;
2. in case $P > 2n$ (i.e. more than $2n$ particles are considered), the initializations (5)-(6) can be adopted for the first $2n$ particles; then, for the remaining $(P - 2n)$ particles an arbitrary initialization can be chosen by the user.

3 A possible drawback of the choice (5)-(6)

Here we detail why, on some real problems, the setting (5)-(6) for the initial DPSO population might be still inadequate. As a preliminary consideration, observe that the proposal (5)-(6) matches item (i) of Section 1, but possibly it does not match also (ii). In the following we report a numerical experience on real problems, in order to highlight the latter fact.

We experienced (5)-(6) for the solution of a tough reformulation of a nondifferentiable constrained portfolio selection problem, proposed in [11]. This model uses a coherent risk measure, based on the combination of lower and upper moments of different orders of the portfolio return distribution. Such a measure can manage non-Gaussian distributions of asset returns, to reflect different investors' risk attitudes. The model includes also *cardinality constraints* (minimum and maximum number of assets to trade), along with constraints on the minimum and the maximum capital percentage to invest in each asset. The model uses the following parameters:

- N, number of possible assets;
- r_e, minimum expected return of the portfolio;
- K_d and K_u, minimum and maximum number of assets to trade;
- d and u, minimum and maximum budget (percentage) to invest in each asset;
- r_i, random variable indicating the return of the i-th asset, for $i = 1, \ldots, N$;
- p, index of the norm used in the risk measure of the portfolio, with $p \geq 1$;

— a, parameter of the risk measure, with $0 \leq a \leq 1$.

Moreover, the variables in our model are described as follows:

— x_i, percentage of the portfolio invested in the i-th asset, for $i = 1, \ldots, N$;

— $z_i = \begin{cases} 1 \text{ if the } i-th \text{ asset is included in the portfolio, for } i = 1, \ldots, N \\ 0 \text{ otherwise;} \end{cases}$

— r, portfolio return.

In addition, $E[y]$ indicates the expected value of the random argument y, while y^- indicates $\max\{0, -y\}$ and y^+ indicates $(-y)^-$. Finally, we use the symbol $\widehat{r}_i$ for $E[r_i]$. Given the above notation, the expected portfolio return $E[r]$ is equal to $E[r] = \sum_{i=1}^{N} \widehat{r}_i x_i$ so that the overall portfolio selection problem is as follows:

$$\min_{x,z} \quad \rho_{a,p}(r) = a\|(r - E[r])^+\|_1 + (1-a)\|(r - E[r])^-\|_p - E[r] \tag{7}$$

$$\text{s.t.} \quad E[r] \geq r_e \tag{8}$$

$$\sum_{i=1}^{N} x_i = 1 \tag{9}$$

$$K_d \leq \sum_{i=1}^{N} z_i \leq K_u \tag{10}$$

$$z_i d \leq x_i \leq z_i u, \quad i = 1, \ldots, N \tag{11}$$

$$z_i \in \{0, 1\}, \quad i = 1, \ldots, N. \tag{12}$$

The function $\rho_{a,p}(r)$ in (7) is a *coherent* risk measure (i.e. it satisfies some formal properties which are appealing for investors).

In order to solve the constrained nonsmooth mixed integer problem (7)-(12) by DPSO, we considered the unconstrained penalty function reformulation

$$\min_{x \in \mathbb{R}^N, z \in \mathbb{R}^N} \quad P(x, z; \varepsilon), \tag{13}$$

$$P(x, z; \varepsilon) = \rho_{a,p}(r) + \frac{1}{\varepsilon_0} \left[\varepsilon_1 \max\left\{0, r_e - \sum_{i=1}^{N} \widehat{r}_i x_i\right\} + \varepsilon_2 \left|\sum_{i=1}^{N} x_i - 1\right| \right.$$

$$+ \varepsilon_3 \max\left\{0, K_d - \sum_{i=1}^{N} z_i\right\} + \varepsilon_4 \max\left\{0, \sum_{i=1}^{N} z_i - K_u\right\} + \varepsilon_5 \sum_{i=1}^{N} \max\{0, z_i d - x_i\}$$

$$\left. + \varepsilon_6 \sum_{i=1}^{N} \max\{0, x_i - z_i u\} + \varepsilon_7 \sum_{i=1}^{N} |z_i(1 - z_i)| \right] \tag{14}$$

and $\varepsilon = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_7)^T > 0$ is a suitable set of penalty parameters, which are adaptively chosen according with the literature.

Note that the portfolio selection problem (13) is NP-hard and its formulation is nonconvex, nondifferentiable, and mixed-integer, so that DPSO was *specifically adopted to provide fast approximate solutions* on several scenarios.

This is indeed a typical application where tradesmen often claim for a quick solution of different scenarios, rather than an accurate solution to propose to their customers. In particular, for a fair comparison the initialization (5)-(6) was tested vs. a standard random initialization of DPSO, with really effective results in terms of fast minimization of the fitness function $P(x, z; \varepsilon)$ (as Figure 1 shows).
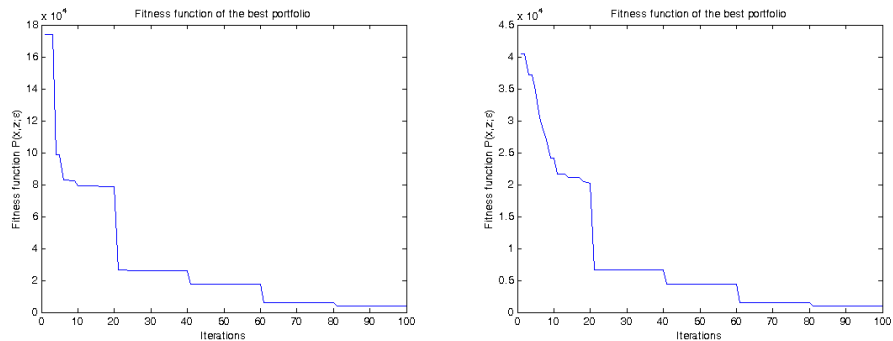

Fig. 1. Fitness function $P(x, z; \varepsilon)$ in (13) vs. the number of iterations: (*left*) when a *random DPSO initialization* is chosen; (*right*) when the *DPSO initialization (5)-(6)* is adopted.

However, we also observed that the initialization (5)-(6) tends to provide sparse solutions, which reduce diversification and might be therefore of scarce interest for several investors. To better appreciate the latter drawback, we monitored the sparsity of the approximate solution provided by the choice (5)-(6) in DPSO, after a relatively small number of iterations (simulating the time required by a tradesman before yielding a possible scenario to investors). After 100 iterations of DPSO, the approximate solution computed by DPSO with (5)-(6) gave a portfolio including just 2 titles (over 32), which is often too restrictive for many investors.

This was a consequence of the corresponding *sparsity* (i.e. a few nonzero entries) of the initialization (5)-(6). On the other hand, though a standard random initialization of particles' position/velocity had a worse performance in terms of minimization for $P(x, z; \varepsilon)$ (see again Figure 1), nevertheless it yielded a final portfolio with several titles, allowing investors to spread the risk of investment.

4   Dense modification of the choice (5)-(6)

Starting from the conclusions of Section 3, here we want to propose a modification of PSO initialization (5)-(6), in order to possibly pursue a *dense* final solution. On this purpose, let be given the $2n$ vectors $z_i(k)$, $i = 1, \ldots, 2n$, in (3)-(4). After some computation, it can be proved that they coincide with the

orthogonal eigenvectors of the symmetric matrix $\mathcal{A} \in \mathbb{R}^{2n \times 2n}$, being

$$\mathcal{A} = \begin{pmatrix} \sigma_1 I & \sigma_2 I \\ \sigma_2 I & \sigma_3 I \end{pmatrix}$$

and (see also (2)) $\sigma_1 = [\gamma_1(k)]^2$, $\sigma_2 = -\gamma_1(k)\gamma_2(k)$, $\sigma_3 = [\gamma_2(k)]^2$. In particular, the (singular) matrix $\mathcal{A}$ has just the two eigenvalues $\mu_1 = 0$ and $\mu_2 = [\gamma_1(k)]^2 + [\gamma_2(k)]^2 > 0$, such that respectively

$$\begin{cases} \mathcal{A}z_i(k) = \mu_1 z_i(k), & i = 1, \ldots, n, \\ \mathcal{A}z_i(k) = \mu_2 z_i(k), & i = n+1, \ldots, 2n. \end{cases} \tag{15}$$

Now, following the motivations suggested in the end of Section 3, we can consider the set of (*dense*) vectors $\nu_1(k), \ldots, \nu_{2n}(k)$ as in

$$\nu_i(k) = z_i(k) - \alpha \sum_{j=1,\ j\neq i}^{n} z_j(k) - \gamma \sum_{j=n+1}^{2n} z_j(k), \qquad i = 1, \ldots, n \tag{16}$$

$$\nu_t(k) = z_t(k) - \beta \sum_{j=n+1,\ j\neq t}^{2n} z_j(k) - \delta \sum_{j=1}^{n} z_j(k), \qquad t = n+1, \ldots, 2n, \tag{17}$$

which are obtained by linearly combining the eigenvectors $\{z_i(k)\}$ in (15). We want to compute the real parameters α, β, γ and δ such that

$$\begin{cases} \nu_j(k)^T \mathcal{A}\nu_i(k) = 0, & \text{for any } 1 \leq j \neq i \leq n \\ \nu_t^T(k)\mathcal{A}\nu_s(k) = 0, & \text{for any } n+1 \leq t \neq s \leq 2n. \end{cases} \tag{18}$$

It can be proved after a tedious computation that conditions (18) allow to satisfy both the properties (i) and (ii) of Section 1. Moreover, taking the values

$$\alpha \in \mathbb{R} \setminus \{-1, 1/n\}, \qquad \beta = \frac{2}{n-2}, \qquad \gamma = 0, \qquad \delta \in \mathbb{R} \setminus \{0, 1\}, \tag{19}$$

then the vectors $\nu_1(k), \ldots, \nu_{2n}(k)$ in (16)-(17) both satisfy (18) and are *dense* in comparison with the original vectors $z_i(k)$, $i = 1, \ldots, 2n$ (which satisfy a relation similar to (18) but are not dense).

Now, observe that $z_1(k), \ldots, z_{2n}(k)$ are mutually orthogonal in the extended space $\mathbb{R}^{2n}$, while the latter property in general does not hold also for the vectors $\nu_1(k), \ldots, \nu_{2n}(k)$. Nevertheless, we can prove in the next proposition that $\nu_1(k), \ldots, \nu_{2n}(k)$ are still enough *scattered* in $\mathbb{R}^{2n}$, possibly being uniformly *linearly independent*. The latter fact is of great relevance and is not immediately evident, since the singular matrix $\mathcal{A}$ is only positive semidefinite.

Proposition 1. *Given the vectors $\nu_1(k), \ldots, \nu_{2n}(k)$ in (16)-(17) and (19), then*

$$\det[\nu_1(k) \vdots \cdots \vdots \nu_{2n}(k)] = \frac{(3n-2)n^n}{(n-2)^{n+1}} [(1+\alpha)^n(1-n\alpha)] \cdot \det[z_1(k) \vdots \cdots \vdots z_{2n}(k)].$$

As long as $\alpha \notin \{-1, 1/n\}$, then $\det[\nu_1(k) \vdots \cdots \vdots \nu_{2n}(k)] \neq 0$ and the vectors $\nu_1(k), \ldots, \nu_{2n}(k)$ are possibly uniformly linearly independent. Observe that the last limitation on α is definitely unrestrictive, considering that our proposal is expected to be effective when the size n of the problem increases.

As regards the results on the portfolio selection problem analyzed in Section 3, we compared the deterministic initialization (5)-(6), with the one obtained replacing $z_1(k), \ldots, z_{2n}(k)$ by $\nu_1(k), \ldots, \nu_{2n}(k)$ (and using the parameters $\alpha = 0.75$, $\beta = 2/(n-2)$, $\gamma = 0$, $\delta = 0.25$). We obtained similar results (see Figure 2) in terms of decrease of the fitness function $P(x, z; \varepsilon)$ in (13). Moreover, adopting $\nu_1(k), \ldots, \nu_{2n}(k)$, both we improved also the minimization of the risk measure $\rho_{a,p}(r)$ (see Figure 3), and we drastically improved the density of the final approximate solution, as expected.
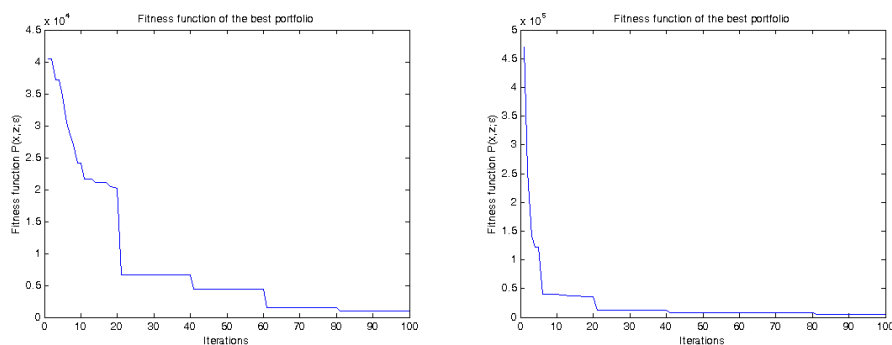

Fig. 2. Fitness function $P(x, z; \varepsilon)$ in (13) vs. the number of iterations: (*left*) when the *DPSO initialization (5)-(6)* is adopted; (*right*) replacing $z_1(k), \ldots, z_{2n}(k)$ by $\nu_1(k), \ldots, \nu_{2n}(k)$ (setting the parameters $\alpha = 0.75$, $\beta = 2/(n-2)$, $\gamma = 0$ and $\delta = 0.25$).
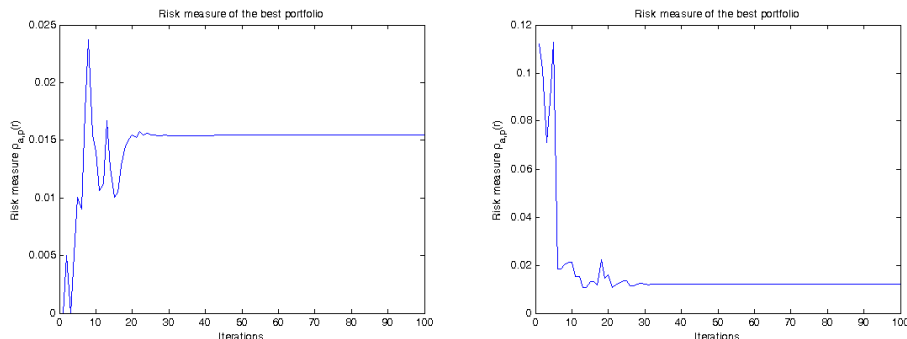

Fig. 3. Risk measure $\rho_{a,p}(r)$ vs. the number of iterations: (*left*) when the initialization (5)-(6) is adopted in DPSO; (*right*) replacing $z_1(k), \ldots, z_{2n}(k)$ by $\nu_1(k), \ldots, \nu_{2n}(k)$ (setting the parameters $\alpha = 0.75$, $\beta = 2/(n-2)$, $\gamma = 0$ and $\delta = 0.25$).

5 Conclusions

In this paper we have described a novel *dense* initialization for DPSO, which is based on a reformulation of PSO iteration as a dynamic linear system. Our

proposal was tested on tough portfolio selection problems, confirming its effectiveness on large scale problems.

Acknowledgments

References

1. Kennedy, J., Eberhart, R.C.: Particle swarm optimization, Proceedings of the 1995 IEEE International Conference on Neural Networks (Perth, Australia), IEEE Service Center, Piscataway, NJ, IV: pp. 1942–1948 (1995)
2. Clerc, M., Kennedy, J.: The Particle Swarm - Explosion, Stability, and Convergence in a Multidimensional Complex Space. IEEE Transactions on Evolutionary Computation 6(1), pp. 58–73 (2002)
3. Ozcan, E., Mohan, C.K.: Particle Swarm Optimization: Surfing the Waves. In: Proceedings of the 1999 IEEE Congress on Evolutionary Computation, pp. 1939-1944. IEEE Service Center, Piscataway (1999)
4. Poli, R.: The Sampling Distribution of Particle Swarm Optimisers and their Stability. Technical Report CSM-465, University of Essex (2007)
5. Trelea, I.C.: The particle swarm optimization algorithm: convergence analysis and parameter selection. Information Processing Letters 85, pp. 317–325 (2003)
6. Van den Berg, F., Engelbrecht, F.: A Study of Particle Swarm Optimization Particle Trajectories. Information Sciences 176, pp. 937–971 (2006)
7. Campana, E.F., Diez, M., Fasano, G., Peri, D.: Improving the initial particles position and parameters selection for PSO, in Bound Constrained Optimization problems, In Y. Tan, Y. Shi, and H. Mo (Eds.): Advances in Swarm Intelligence, Part I, Springer Lecture Notes in Computer Science 7928, pp. 112–119 (2013)
8. Diez, M., Serani, A., Leotardi C., Campana, E.F., Peri, D., Iemma, U., Fasano, G., Giove, S.: A Proposal of PSO Particles Initialization for Costly Unconstrained Optimization Problems: ORTHOinit, In Y. Tan, Y. Shi, and C.A. Coello Coello (Eds.): Advances in Swarm Intelligence, Proceedings of IC-SI 2014, Part I, Springer Lecture Notes in Computer Science 8794, pp. 126–133 (2014)
9. 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, vol. 48, pp. 347–397 (2010)
10. Zheng, Y., Ma L., Zhang, L., Qian J.: On the convergence analysis and parameter selection in particle swarm optimization, in: Proceedings of the International Conference on Machine Learning and Cybernetics, pp. 1802–1807 (2003)
11. Corazza, M., Fasano, G., Gusso, R.: Particle Swarm Optimization with non-smooth penalty reformulation, for a complex portfolio selection problem, Applied Mathematics and Computation, vol. 224, pp. 611–624 (2013)