Optimizing large scale problems with metaheuristics in a reduced space mapped by autoencoders – application to the wind-hydro coordination

Journal:	*IEEE Transactions on Power Systems*
Manuscript ID:	TPWRS-01482-2013.R3
Manuscript Type:	Transactions
Date Submitted by the Author:	n/a
Complete List of Authors:	Miranda, Vladimiro; INESC TEC, University of Porto da Hora Martins, Joana; INESC TEC, Palma, Vera; INESC TEC,
Technical Topic Area :	Power system computational analysis
Key Words:	Wind-hydro coordination, Large scale optimization, Neural networks, Autoencoders, Evolutionary algorithms, Meta-heuristics

Optimizing large scale problems with metaheuristics in a reduced space mapped by autoencoders – application to the wind-hydro coordination

Vladimiro Miranda, *Fellow IEEE*, Joana da Hora Martins and Vera Palma

Abstract – This paper explores a technique denoted LASCA to solve large scale optimization problems with metaheuristics by reducing the search space dimension with autoassociative neural networks. The technique applies autoencoders as a reversible mapping between the original problem space and a reduced space. A metaheuristic then evolves in the latter, having its objective function assessed in the original space. The technique is illustrated with an application of an EPSO (Evolutionary Particle Swarm Optimization) algorithm to four benchmarking unconstrained optimization functions and to a wind-hydro constrained coordination problem. The new technique allows an improvement in the quality of the solutions attained.

Index Terms — Wind-hydro coordination problem, large scale optimization, neural networks, autoencoders, evolutionary algorithms, metaheuristics .

I. Introduction

SOLVING problems in high dimensional spaces is both demanding in computing resources and difficult in the convergence to satisfactory solutions. One of the major practical problems relates to the curse of dimensionality [1]. This is particularly true in the case of the behavior of metaheuristics . This difficulty usually leads to a termination of runs earlier than necessary, when using metaheuristics such as: genetic algorithms [2], evolutionary algorithms [3] or cooperative coevolution [4], among many. This paper addresses the problem of solving large scale optimization problems with metaheuristics by achieving a reduction in the dimensionality of their search spaces.

The problem of space reduction has been addressed e.g. in clustering and in image processing. One important technique is Principal Component Analysis (PCA) [5], which projects the data into a linear subspace: data are multiplied by the eigenvectors from the sample covariance matrix, each point being represented by its coordinates along the orthogonal directions of the greatest variance in the data set.

One topic explored in this paper and not usually addressed is the adoption of dimensionality reduction techniques as a general optimization tool for large scale problems. It is thus necessary to transfer into a reduced space not only the data but also the constraints.

The technique developed was firstly suggested in [6], and is here denoted as LASCA (LArge SCale optimization with Autoencoders). The main idea is to allow an evolutionary metaheuristic to evolve in a reduced dimension space $\boldsymbol{S}'$, controlling its progression in the original space $\boldsymbol{S}$. The transition between $\boldsymbol{S}$ and $\boldsymbol{S}'$ is made through an autoencoder, applied as a reversible mapping between the two spaces.

The LASCA technique is illustrated with an application of an EPSO (Evolutionary Particle Swarm Optimization) [7] algorithm to 5 case studies in high-dimensional spaces: four mathematic optimization functions suggested in [8] (the Alpine, Shifted Sphere, Shifted Rastrigin and Griewank functions). Then, a power system problem in wind-hydro coordination problem is also tested.

The results obtained show that the LASCA approach may lead to better quality solutions than the attempt to solve them with metaheuristics in the original large dimensional space. Only in one case (Griewank) the process did not return a significant gain in accuracy, in computational effort or in the quality of the solution achieved. But the reasons for this seem to be associated to the fact that the convergence obtained with the metaheuristic in the original space was found to be already sufficient to achieve the optimum within few iterations.

LASCA is, therefore, a process to be taken in account in solving difficult large scale problems. Its contribution to solving the wind-hydro coordination problem is relevant.

II. Why autoencoders

Autoencoders, or autoassociative neural networks (NN), are feedforward neural networks trained to replicate the input data vectors, represented in a space $\boldsymbol{S}$, in their output equally in $\boldsymbol{S}$. In a simple autoencoder (Fig. 1), there is a middle layer with a distinct, usually smaller dimension from input or output. The information flowing through the autoencoder must pass such bottleneck with minimum loss. A one-to-one reversible mapping between points in $\boldsymbol{S}$ of dimension m and a space $\boldsymbol{S}'$ of dimension n (with $n < m$) is established: the first half of the autoencoder projects the input information onto $\boldsymbol{S}'$ and the second half produces the inverse projection. The quality of the process highly depends on the quality of the training.

The adequacy of autoencoders to reduce the dimensionality of data is widely known [9][10]. Autoencoders have been applied to perform signal analysis [11][12] or to reconstruct missing sensor signals [13].

This work is partially funded by the ERDF from the EU through the Programme COMPETE and by the Portuguese Government through FCT - Foundation for Science and Technology, project ref. LASCA PTDC/EEA-EEL/104278/2008..

J. Hora (joana.hora@gmail.com), V. Palma (vera.p.ferreira@inescporto.pt) and V. Miranda (vmiranda@inescporto.pt) are with INESC TEC - INESC Technology and Science, coordinated by INESC Porto, Portugal. V. Miranda is also with FEUP, the Faculty of Engineering of the University of Porto, Portugal.

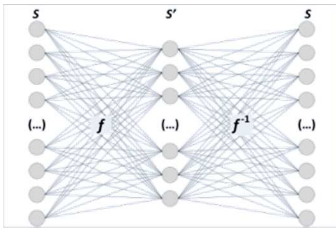

Fig. 1. Schematic of an autoencoder. An instance in space $\boldsymbol{S}$ is encoded into a reduced space $\boldsymbol{S'}$ by f, and is expanded back into $\boldsymbol{S}$ by f^{-1}.

Other applications include the representation of images within a reduced space [14][15], allowing the application of several processing techniques such as pattern recognition. For instance, face images could be identified and clustered according to sex and distinguished from non-faces [16]. Optimizing the weights of multiple inner layer autoencoders with non-linear activation functions is a heavy and difficult task; therefore, new schemes to achieve a more efficient training are being proposed [17][18].

Finally, there is no indication on the adequate ratio between the number of neurons in the middle layer and in the inputs/output layers). This decision has been dictated by trial and error and by the characteristics of the problem, having in mind that the greatest dimension reduction with only negligible loss of quality is almost always desirable.

In summary: the main technical reason to reduce space dimensionality is the ineffectiveness of metaheuristics in solving problems in very large spaces; and the main technical reason to adopt autoencoders is their learning ability: as neural networks, they may learn about the search space from the results of a preliminary search, and if so they may perform a satisfactory space dimensionality reduction.

III. Why EPSO

The optimization metaheuristic adopted was EPSO, for Evolutionary Particle Swarm Optimization. EPSO is a hybrid of the concepts of Evolutionary Algorithms (EA) and Particle Swarm Optimization (PSO) [19], first proposed in [7], improved in [20] and with an enhanced version in [21]. It is an EA with a self-adaptive recombination operator inspired in the "movement rule" of PSO. This rule generates a new individual (chromosome, particle) as a weighted combination of parents: a given individual, its best ancestor and the best ancestor of the swarm. This weighted mix may vary in each space dimension. A mutation operator is applied to the weights, therefore forming a self-adaptive recombination operator.

An EPSO iteration starts with a swarm of p particles. Each particle originates 2 descendants, from which only one will survive. The steps are:

Replication: each particle X_i is replicated. The k^{th} replica of X_i is designated as X_{ik}. In practice, only 2 replications are used.

Mutation: changing the weights of Inertia (wi), Memory (wm) and Cooperation (wc) associated to each replica X_{ik}. The mutation for wm is given in (1), where τ is a learning parameter that can be fixed or be subject to mutation. Similar rules are considered for the wi and wc.

$$wm_{ik}^{t+1} = wm_{ik}^{t} + \tau \cdot N(0,1) \qquad (1)$$

Reproduction: each particle X_{ik} originates a new descendant according to the movement rule

$$X_{ik}^{t+1} = X_{ik}^{t} + wi_{ik}^{t+1} \cdot V_{ik}^{t} + wm_{ik}^{t+1} \cdot (b_i - X_{ik}^{t}) + P\, wc_{ik}^{t+1} \cdot (b_G^{t+1} - X_{ik}^{t}) \qquad (2)$$

where the best known solution achieved by swarm b_G^t is also mutated, inducing agitation in the swarm:

$$b_G^{t+1} = b_G^{t} + \tau' \cdot N(0,1) \qquad (3)$$

P in (3) is a binary variable, taking values {0, 1}, defined by sampling a random uniform number R and comparing it to a threshold *cp*, defined as the *communication probability*.

Evaluation: each descendant fitness is evaluated.

Selection: one particle from each group of *k* descendants is selected to form the new swarm.

EPSO is a metaheuristic very successfully benchmarked against other algorithms, including the very own PSO [7][22][23][24][25]. Yet, it also displays the common difficulty in convergence in large spaces. For these reasons, it has been selected as the test bed for LASCA, as a proxy for any population based metaheuristic.

IV. The LASCA approach

It is known that, in general, population based methods such as EA (Evolutionary Programming or Genetic Algorithms), or non-evolutionary such as PSO or Ant Colony Methods exhibit a growingly slow and inaccurate performance with the increase in dimension of the search space where individuals are defined. This limits their practical application in large scale programming problems. However, these problems are extremely relevant in Power Systems, where one may find planning or operation problems with hundreds to tens of thousands of variables.

The original idea reported in this paper is, therefore, to use autoencoder properties to reduce the dimension of the search space, while keeping the solution evaluation accurate (based on the original space) so that selection may still act and drive the process towards an optimum. This idea can be summarized in the following sequential parts.

Phase A. An EA with individuals (particles) is applied in $\boldsymbol{S}$ – a space with the dimension of the original problem. Distinct solutions obtained over a specified number of iterations are stored. These are used as a dataset to train an autoencoder to encode/decode particles between $\boldsymbol{S}$ and $\boldsymbol{S'}$. This will hopefully capture characteristics of the landscape being searched.

The training and test sets used to generate the autoencoder neural network are not obtained through random sampling. In fact, because the sampling is conducted using an evolutionary optimizing method, it becomes very likely that one will have a denser representation of the solution space in regions closer to the optimum, which is a very desirable trait.

Phase B. The last swarm obtained in Phase A, that was evolving in $\boldsymbol{S}$, will be transferred (projected on) to $\boldsymbol{S'}$. The information transferred includes the particles and the corresponding best positions, velocities and weights of inertia, cooperation, memory and perturbation.

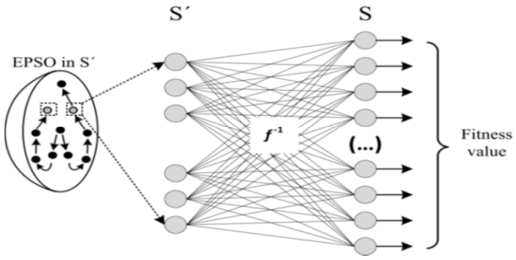

Fig. 2. Phase B: the particles evolve in $\boldsymbol{S'}$ but the fitness evaluation is made in $\boldsymbol{S}$ using the decoding function f^{-1} on the 2nd half of the autoencoder.

The transference is made by applying the encoding function f corresponding to the 1st half of the autoencoder. For each particle, its velocity is calculated in $\boldsymbol{S'}$ as the difference between the corresponding compressed positions in t and in $t-1$. The transference of weights is made directly. Once the swarm is transferred, it is made to evolve in ***S'***. But the evaluation of the fitness function is not possible in ***S'***, since the components of the particles in the reduced space have no direct physical meaning. Therefore, the particles are first decoded onto ***S*** and then evaluated, as represented in Fig. 2. This procedure ensures that the evolution observed in ***S'*** corresponds to an improvement in the problem addressed.

The variables in ***S'*** (the output of the neurons in the middle layer) are a mix with unknown meaning. As constraints associated with limits must be enforced in this space, for the set of values assumed by the variables in ***S'*** limits are defined taking in account the minimum and maximum values registered in the training set. Besides, a classical penalty strategy is applied in ***S*** (output) in the fitness evaluation.

Phase C. The autoencoder results in an approximation of the exact mapping $\boldsymbol{S} \leftrightarrow \boldsymbol{S'}$. It is possible that the exact optimum of the original problem may not be found in $\boldsymbol{S'}$, but a near optimal solution or, at least, the location of the optimum will be found. A final search in ***S*** may prove useful. A transfer from ***S'*** to ***S*** is done using f^{-1} (2nd half of the autoencoder) and the EA version from Phase A is re-launched.

V. BENCHMARK OPTIMIZATION FUNCTIONS

This section includes the results obtained with the application of the LASCA approach to four benchmark optimization functions, and the comparison with EPSO running solely in ***S***, in swarms of 400 particles.

Table 1 presents the four problems and variants. Table 2 indicates the architecture of the autoencoders and the training parameters: learning rates τ, communication probabilities *cp* for the EPSO running in spaces ***S*** and ***S'***, as well as the number of EPSO iterations spent in phases A-B-C in each problem.

Table 3 describes the type of activation functions used in the NN neurons, in each experiment. The training was done with two distinct methods. In Table 3, PROP denotes a classical backpropagation training minimizing the MSE (mean square error between input and output). QMI-CS means that the first half of the autoencoder was trained in unsupervised mode [18], maximizing the Quadratic Mutual Information transferred between input and output, estimated using the Cauchy-Schwarz distance, and then the second half was trained independently, using a supervised backpropagation minimizing the MSE. All autoencoders were initialized before training by applying a PCA projection matrix as weight matrix of their first half, as if the activation functions were linear.

Table 4 indicates the location and the value of the optimum in each problem as well as the best results in 10 trials obtained by running EPSO in ***S*** or running LASCA in ***S-S'-S***. It may be seen that in all cases the LASCA approach could beat the EPSO metaheuristic acting solely in the original space ***S***.

TABLE 1. PROBLEM DEFINITION AND DIMENSIONS D OF VARIANTS

Name	$f(x_1,\cdots,x_D) =$	Obj	D
Alpine	$\prod_{i=1}^{D}\sin(x_i)\sqrt{\prod_{i=1}^{D} x_i}$	Max	120 200 300
Shifted Sphere	$\sum_{i=1}^{D} {x_i}^2 + f_{bias}$	min	120 200 300
Shifted Rastrigin	$An + \sum_{i=1}^{D}\left[x_i^2 - A\cos(2\pi x_i)\right] - 330$	min	120 300
Griewank	$1 + \frac{1}{4000}\sum_{i=1}^{D} x_i^2 - \prod_{i=1}^{D} cos\left(\frac{x_i}{\sqrt{i}}\right)$	min	120

TABLE 2. NN ARCHITECTURE, PARAMETERS. ITERATIONS IN PHASES A-B-C

NN input-middle-output	τ_S	cp_S	$\tau_{S'}$	$cp_{S'}$	no. iterations A-B-C	Name
[120-50-120] [200-70-200] [300-150-300]	0.4 0.4 0.4	0.1 0.1 0.1	0.4 0.5 0.1	0.95 0.9 0.9	100-100-100 40-40-20 40-40-20	Alpine
[120-50-120] [200-70-200] [300-150-300]	0.7 0.7 0.7	0.9 0.9 0.9	0.8 0.8 0.8	0.5 0.6 0.6	100-100-100 40-40-20 40-40-20	Shifted Sphere
[120-50-120] [300-150-300]	0.9	0.9	0.8	0.4	100-30-100 100-30-100	Shifted Rastrigin
[120-50-120]	0.6	0.95	0.75	0.5	15-50-100	Griewank

TABLE 3. NN ACTIVATION FUNCTIONS AND TRAINING METHOD

Activation functions			Training method		Name
input	hidden	output	1st half	2nd half	
linear	tanh	linear	PROP		Alpine
linear	tanh	tanh	QMI-CS	PROP	Shifted Sphere
linear	tanh	linear	PROP		Shifted Rastrigin
linear	tanh	linear	QMI-CS	PROP	Griewank

TABLE 4 – EXACT OPTIMUM AND RESULTS BY EPSO AND LASCA

X^{OPT} =	opt	Best Fit EPSO	Best Fit LASCA	D	Name
(7.917,⋯ …,7.917)	6.42E+53 4.78E+89 3.30E+134	1.66E+30 2.63E+48 1.28E+72	1.51E+53 2.63E+88 4.71E+131	120 200 300	Alpine
(0,⋯,0)	-450	-433.05 -332.62 -244.69	-449.94 -416.49 -366.11	120 200 300	Shifted Sphere
(0,⋯,0)	-330	-318.081 -318.264	-318.087 -321.161	120 300	Shifted Rastrigin
(0,⋯,0)	0	0	0	120	Griewank

An examination of the following figures, depicting the average in 10 trials of the evolution of the objective function, allows a deeper understanding of these results

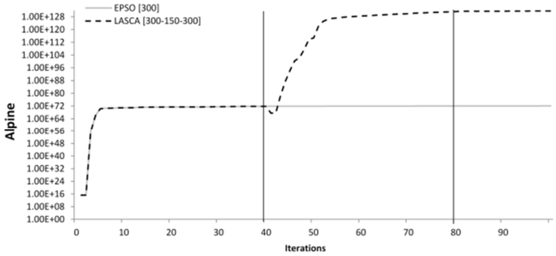

Fig. 3. Results obtained with LASCA[300-150-300] and EPSO[300] for the Alpine function: progress of the objective function. Average of 10 runs.

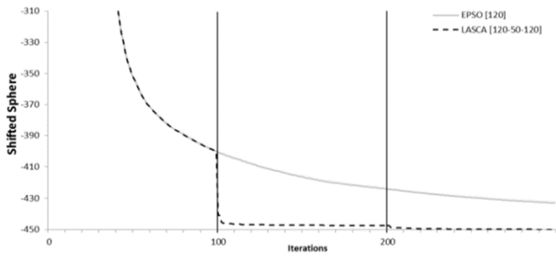

Fig. 4. Progress of the objective function obtained with LASCA[120-50-120] and EPSO[120], for the shifted sphere function. Average of 10 runs.

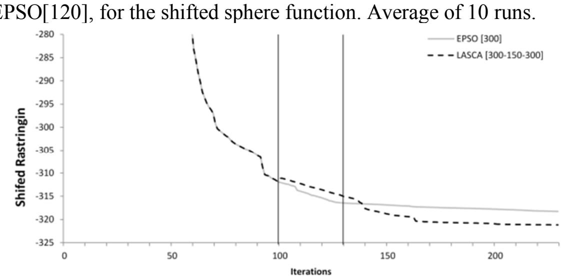

Fig. 5. Progress of the objective function obtained with LASCA[300-150-300] and EPSO[300], for the Shifted Rastrigin function. Average of 10 runs.

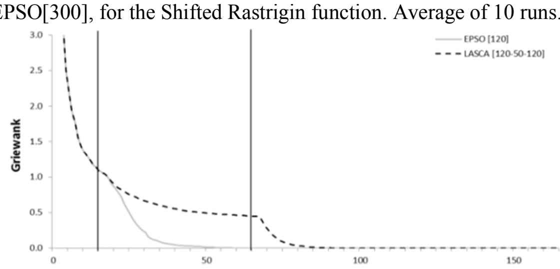

Fig. 6. Progress of the objective function obtained with LASCA[120-50-120] and EPSO[120], for the Griewank function. Average of 10 runs.

A. *Alpine function*

Fig. 3 illustrates the remarkable advantage given by LASCA in solving for the Alpine function in a space of 300 dimensions. It is obvious that the metaheuristic alone gets stuck and does not show progress after some time. The LASCA process, with the search in the reduced space ***S'*** allowed by the autoencoder, gave an impressive boost and the end result, after Phase 3, is remarkably superior. The transition into a reduced space clearly allowed the swarm to find a better search zone.

B. *Shifted Sphere function*

It is very obvious, in Fig. 4, the net advantage given by LASCA in searching for a solution close to the optimum, in a space of 120 dimensions in this case.

C. *Shifted Rastrigin function*

In this case the advantage given by LASCA, although real, is less pronounced – see Fig. 5 for a case of 300 dimensions.

D. *Griewank function*

The results obtained in a space of 120 dimensions are illustrated in Fig. 6. In this problem, even at high dimensionality, EPSO had a particularly fast convergence to the global optimum. Because of the particularly fast convergence of EPSO, a very low number of iterations in Phase A were taken – but this prevented the storage of a sufficient number of different particles to accurately train the NN. Lesson: there is no point in adopting the LASCA approach if the convergence is fast in the original space ***S***.

E. *Computing time*

The LASCA approach involves training an autoencoder, sometimes a heavy computing requirement. Yet, the cost paid for adopting LASCA must be balanced against a very positive trade-off in reaching much better solutions. However, the additional computing effort seems to be manageable, especially if dealing with planning problems.

In Table 5 one may find an estimation of computing times for all the cases reported above, on an average of 10 trials in each case, run in a normal laptop computer, with non-optimized software written to test the methodology. The criterion to stop the training of the autoencoder was: difference between two consecutive MSE (test dataset) lower than 0.00001.

TABLE 5. COMPUTING TIMES ESTIMATED, IN SECONDS

EPSO in S	Epochs NN	Time to train NN	LASCA total	D	Name
1.8	15	4.9	6.7	120	Alpine
1	18	14.7	15.7	200	
1.4	19	46.0	47.4	300	
1.2	10	38.7	39.9	120	Shifted Sphere
0.6	25	64.7	65.3	200	
2	19	64.5	66.5	300	
1.5	46	25.9	27.4	120	Shifted Rastrigin
1.6	55	167.0	168.6	300	
1.2	48	81.8	83.0	120	Griewank

VI. WIND-HYDRO COORDINATION PROBLEM

A. *General description of the problem*

An experimental confirmation of the potential of LASCA is obtained in a wind-hydro coordination (WHC) problem built expressly to serve as test bed. The presentation does not aim at describing a full-fledged model – but rather to show that the technique finds immediate application in Power Systems. Compared with the benchmarking functions, this problem is more complex: it has a set of linear and non-linear constraints while the other were non-linear functions with no constraints.

The WHC problem, from the point of view of an owner of both wind and hydro power stations, aims at maximizing the joint operation profit of a power system composed by several hydro and wind farms. The maximization is made by changing the water volumes to be pumped and released, given a set of specifications defining each scenario. These specifications include the wind forecast and the water inflow to the system. The operation planning is normally made with multiple approaches for different horizons, from the short (days) to the long term (years); a medium term (months, year) operation planning is considered in the following exercise. This problem has obvious similarities with hydro-thermal coordination with pumping storage facilities and has a complex time dependent formulation if cascading river dams are present.

The hydro-thermal coordination is a large scale difficult problem. Several techniques have been used to deal with it such as Lagrangian relaxation [28], Stochastic Dynamic Programing [29], Dual Dynamic Programming [30] or Genetic Algorithms and Evolutionary Programming [31], to refer to early models. Wind-hydro coordination models have also been proposed but mostly dealing with short term [32][33][34][35]. There is surprisingly little material on medium term operation planning in systems with complex cascading hydro power stations. An insightful review concerning applications and methods, but without wind, is provided in [36]. The test problem formulated in this work follows lines similar to many deterministic models proposed in the past for hydro system modeling.

A medium term operation planning requires an evaluation of the operation for a period up to 1 year. The planning period is divided in sub-periods corresponding to different months and distinct load levels with different estimated energy costs. The dimension of the problem may be very large and it would be even larger if a stochastic optimization would be considered, taking in account scenarios describing the uncertainty in the renewable resources.

In general terms, the test of the LASCA approach considers the optimal operation in a deterministic context, meaning that future inflows (water, wind) are taken as known data. This problem may be seen as a sub-problem of a stochastic optimization formulation. It is composed of an independent energy producer that operates a number of cascading hydro power plants and manages also wind power plants, treated as a single source (energy supplied through the transmission grid). Peak and off-peak periods are distinguished. Water can be stored in reservoirs during off-peak periods, and used to produce energy during the peak periods. The decision to store must be weighed against the price of selling directly at the moment it is produced in the wind farms.

As the purpose of the paper is to demonstrate the potential and usefulness of the new LASCA technique, not much space will be used to describe the subtleties of the real world problem or analyze the effects of uncertainties and will concentrate on the optimization procedure instead. A test system was prepared, inspired in real world cases but not representing any particular case. It integrates N=8 cascading reservoirs, in an arrangement detailed in Fig. 7, built as an expansion from a problem described in [31]. All reservoirs except the lower reservoir (except the most downstream one) are admitted to be equipped with pumps allowing a certain amount of water to be moved up from downstream if convenient.

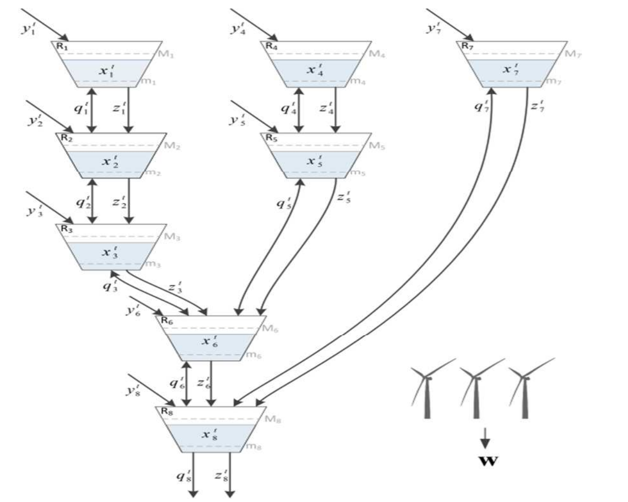

Fig. 7. Scheme for the wind-hydro test power system with a cascading arrangement of reservoirs, different in capacity, constraints and inflows.

The final purpose is to assess the value for a coordinated wind-hydro operation, by deriving an operation plan that maximizes the profit obtained with the operation of the system throughout T time periods with different buying and selling energy prices. The operation plan will determine:

- The water to be released/pumped for each hydro power plant in each period of time and energy sold or used;
- The wind energy to be used for pumping and the wind energy to be sold to the grid in each period.
- The amount of water storage in each reservoir and storage capacity available for each period of time

The T time periods are divided in $T/2$ peak periods and $T/2$ off-peak periods. A horizon of 6 months ($T = 12$) is considered. Six energy prices are defined for each period, also admitting average price forecasting based on market history:

- Hydro energy selling price at peak and off-peak periods;
- Hydro pumping price at peak and off-peak periods;
- Wind energy selling price at peak and off-peak periods.

The variables are defined in terms of water movement for each reservoir in each period. Together, they form the chromosome of each particle. Ecological spills or evaporation are not considered in this example but present no difficulty to be added to the model.

The solution for this problem may be compared to an operation without resorting to pumping (implying the direct sale of wind power to the grid without energy transfer between time periods). This allows the estimation of the added value of the pumping function and may serve as valuable information not only for the operation planning of a system but also for the assessment of the convenience of having a storage function associated to wind power generation. The constraints could have been dealt with penalties but the algorithmic process assured that the solutions achieved were always feasible.

B. The mathematical model

The electric energy of hydro origin generated in moment t by reservoir n is described by

$$H_{n,j}^{t} = K_{n,j}[h_n(x_n^t) - h_n(q_n^t + z_n^t)] \cdot |q_n^t| \quad (8)$$

where:

$H_{n,j}^t$ - the energy generated by the n^{th} reservoir in period t if $j = turbine$, or the energy consumed if $j = pump$ (in Wh);

N – no. of hydro power plants in the system;

$K_{n,j}$ - a specified constant for each reservoir, which agglomerates the gravitational acceleration (g), the efficiency of the turbine (η) and the water density (ρ): $K_{n,j} = \eta \cdot \rho \cdot g$. This constant takes different values for pumping ($j = pump$) and generation ($j = turbine$);

x_n^t - the volume stored in the n^{th} reservoir at the beginning of the period t (in m^3);

q_n^t - volume of water transferred between the n^{th} and the immediately downstream reservoirs, at t: it assumes negative values for pumping and positive for volumes released (in m^3);

z_n^t - volume of water spilled during the period t (in m^3);

$h_n(\cdot)$ – function returning the estimation of the water head (height) given a water volume, for the n^{th}reservoir (in m).

The available water volume for each reservoir is calculated for each period considering all the variables associated to the n^{th} reservoir:

$$x_n^{t+1} = x_n^t + y_n^t + \sum_{k\in\Omega}[q_k^t + z_k^t] - q_n^t - z_n^t \quad (9)$$

where:

y_n^t –natural inflow (in m^3) entering the n^{th} reservoir;

Ω – set of reservoirs immediately upstream of the n^{th}.

The EPSO algorithm is applied to optimize a particle $\boldsymbol{q}$, which includes the volumes q_n^t for each reservoir in each moment, with N referring to the total number of reservoirs and T to the total number of time steps.

$$\boldsymbol{q} = [q_1^1, q_2^1, \cdots, q_N^1, \cdots, q_1^T, q_2^T, \cdots, q_N^T] \quad (10)$$

Constraints ensuring reservoir capacities

The simulation starts with water volumes x_n^t respecting the reservoir minimum and maximum capacity limits m_n and M_n. The model must ensure these limits are satisfied in further temporal moments. Therefore, the volume of the n^{th} reservoir at moment $t+1$ must respect:

$$m_n < x_n^{t+1} < M_n \quad (11)$$

Combining equation (9) and (11), one can obtain the dynamic constraints specified in (12) and (13), which are applied to each position of $\boldsymbol{q}$.

$$x_n^t + y_n^t + \sum_{k\in\Omega}[q_k^t + z_k^t] - z_n^t - M_n < q_n^t \quad (12)$$

$$x_n^t + y_n^t + \sum_{k\in\Omega}[q_k^t + z_k^t] - z_n^t - m_n > q_n^t \quad (13)$$

Constraints ensuring turbine capacities

For each reservoir, the specifications on the turbines installed were considered, which allowed the estimation of maximum and minimum volumes they are able to release or pump. These constraints are considered in the model as represented in

$$q_n^{min} < q_n^t < q_n^{max} \quad (14)$$

Constraints ensuring the available water to pump

When the decision to pump water is made, the maximum value of volume to pump must also be restricted to the available volume in the immediately downstream reservoir (IDR). This constraint is only meaningful to the pumping case since when releasing water, if the IDR exceeds its maximum capacity, the overflow is spilled over. Accordingly, the maximum volume of water to pump into the n^{th} reservoir, γ_n^t, is defined in (15), and constrains q_n^t as defined in (16).

$$\gamma_n^t = x_{IDR}^t - m_{IDR} \quad (15)$$

$$\gamma_n^t < q_n^t \quad (16)$$

The wind energy generated per period is estimated as w^t by an external forecasting procedure – and taken as data in this WHC example. Its value per period is derived from the wind series and each wind farm production characteristic, which can be modeled separately from the optimization procedure. In fact, as there are no "reservoirs for wind", the generation forecast is a direct function of the wind forecast. An auxiliary vector $\boldsymbol{w}$ is considered, where each element w_n^t refers to the available wind energy for the n^{th}reservoir at moment t, and this vector is updated as further described.

Value of the energy produced by the water released

When q_n^t is positive (turbining), the corresponding energy is calculated from equation (8). The value associated to this energy is further calculated by considering the corresponding price, depending if the period type is peak or off-peak

$$V_{n,A}^t = \begin{cases} H_{n,turb}^t \cdot P_{A,1} & \text{,if in a peak period} \\ H_{n,turb}^t \cdot P_{A,0} & \text{,if in an off-peak period} \end{cases} \quad (17)$$

Value of the energy consumed to pump water

When q_n^t is negative (pumping), the energy necessary to pump is calculated using equation (8).

Two possibilities may happen here: there is enough wind energy available w_n^t, to pump the water, or there is not.

When there is enough wind (i.e. $w_n^t \geq H_{n,pump}^t$), w_n^t is used. In this case, the value of the wind energy spent to pump is calculated by

$$V_{n,B}^t = \begin{cases} H_{n,pump}^t \cdot P_{B,1} & \text{, if in a peak period} \\ H_{n,pump}^t \cdot P_{B,0} & \text{, if in an off-peak period} \end{cases} \quad (18)$$

If there is not enough wind energy available (i.e. $w_n^t < H_{n,pump}^t$), then the model spends all the available energy from wind farms and buys the remainder necessary energy from grid (i.e. $G_p = H_{n,pump}^t - w_n^t$). In this situation, equation (18) is used to calculate the value of the energy consumed from wind farms, and equation (19) is considered to calculate the value of the energy bought from grid to pump.

$$V_{n,C}^t = \begin{cases} (H_{n,pump}^t - w_n^t) \cdot P_{C,1} & \text{,if in a peak period} \\ (H_{n,pump}^t - w_n^t) \cdot P_{C,0} & \text{,if in an off-peak period} \end{cases} \quad (19)$$

For both cases, the wind energy available is updated, to give the energy assessment in all reservoirs in the same step.

Value of the wind energy

When all reservoirs are assessed for a specified temporal moment, the model will calculate the monetary value of the available wind energy at that moment, if any is available, which is considered to be sold to the grid. This value is estimated as defined in

$$V_{n,D}^t = \begin{cases} w_n^t \cdot P_{D,1} & \text{, if in a peak period} \\ w_n^t \cdot P_{D,0} & \text{, if in an off-peak period} \end{cases} \quad (20)$$

Revenue

The revenue obtained with each reservoir is defined as

$$R_n^t = \begin{cases} V_{n,A}^t + V_{n,B}^t - V_{n,B}^t - V_{n,C}^t & ,\text{if } n \neq N \\ V_{n,A}^t + V_{n,B}^t - V_{n,B}^t - V_{n,C}^t + V_{n,D}^t & ,\text{if } n = N \end{cases} \quad (21)$$

Once all reservoirs are assessed for all temporal moments, the profit obtained with the entire system is calculated as

$$Profit = \sum_{t=1}^{T} \sum_{n=1}^{N} R_n^t \quad (22)$$

To this *Profit*, a value associated to the water remaining in the reservoirs is added, forming the final fitness function. This avoids the depletion of reservoirs in the end of the process.

C. Results

The wind-hydro coordination problem was simulated with swarms of 50 particles, with EPSO parameters in $\boldsymbol{S}$ being $\tau_S = 0.9$ and $cp_S = 0.7$ and in $\boldsymbol{S'}$ being $\tau_{S'} = 0.8$ and $cp_{S'} = 0.1$. The number of iterations used in Phases A, B and C were 400/400/200. The autoencoder was trained with PROP, with tangent hyperbolic and linear activation functions for the hidden and output layers, respectively. Fig. 8 illustrates the notorious advantage of adopting LASCA. The final profit value obtained by LASCA over 1000 iterations is, for the best solution, 28% better than obtained by EPSO alone. The computing time was of an average of 150.4 seconds, having required 580 NN training epochs to achieve the training criterion of delta MSE < 0.00001.

With LASCA, once in Phase B, an initial deterioration in profit occurs, possibly due to the slight loss of information incurred with the transition to the reduced space. The evolution observed afterwards denotes a progressive increase in the profit value till iteration 600, approximately, from where stabilization is observed till the end of Part B. The transition to Phase C allows a slight new improvement of the profit. It is obvious that the metaheuristic alone had a premature stagnation.

The solutions obtained were examined for quality and were according to sound engineering judgment, taking in account the scenarios built for this experiment. A glimpse may be allowed for a particular optimal solution in Fig. 9 and Fig. 10.

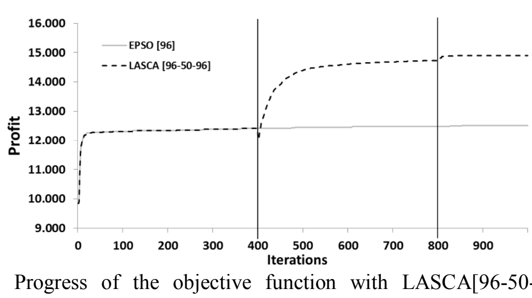

Fig. 8. Progress of the objective function with LASCA[96-50-96] and EPSO[96] for the WHC problem with 8 reservoirs. Average of 10 runs.

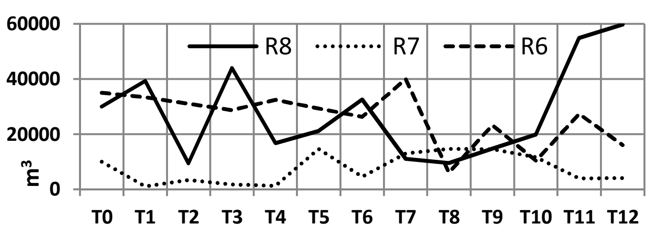

Fig. 9. Storage in 3 reservoirs. T0 is the initial state. Odd labels are for peak periods, even labels are off-peak periods.

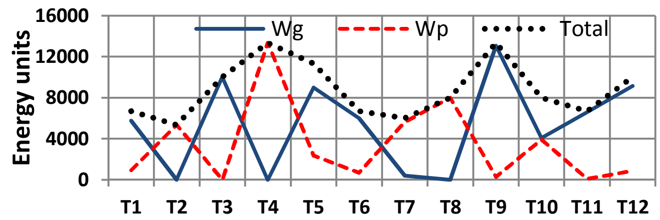

Fig. 10. Wind energy W_g sold to the grid and W_p used to pump.

The solutions are very much conditioned by the costs of buying/selling energy in each period, as well as the wind scenarios and the water inflow regimes. The experiment was conceived to illustrate the power of LASCA and no deep analysis is detailed in its particular results.

VII. Conclusions

Metaheuristics are known to lose efficiency in large scale problems: the convergence becomes slow and the computing effort heavy when the number of variables is large. Eventually, the optimum remains unreachable.

This paper presents, through the testing in benchmark optimization functions and in a practical example on the wind-hydro coordination problem, a novel method to approach the solutions of large scale problems with population-based metaheuristics, by organizing searches in an equivalent reduced dimension search space.

The model was coined as the LArge SCale optimization with Autoencoders approach, or LASCA. The result sought is to obtain a much improved solution, when compared with the action of the metaheuristic alone, even if paying a price in added computing effort. The finesse of the method lies in the fact that the evolutionary process acts upon individuals represented by chromosomes that are not designed ad-hoc by a human; instead, they result from an intelligent coding achieved by a first half of an autoencoder, while the fitness function is evaluated by decoding the intelligent chromosomes with the second half of the same autoencoder. Because the clever chromosomes are represented in a space of reduced dimension, the optimization process is able to move towards different search zones, thus finding better quality solutions.

LASCA worked well in benchmarking unconstrained functions. These served to illustrate how, with LASCA, the iterations were able to escape becoming trapped in local optima at an early stage. To demonstrate its usefulness in the Power Systems domain, a wind-hydro coordination in medium term operation planning was also solved. This is a complex problem with constraints and spatial and temporal dependency, introduced by the cascading hydro power stations and the need to represent a large set of time steps.

The test for LASCA was made with an EPSO algorithm, but there is no loss of generality as any other population-based method could have been used. The results presented fully demonstrate the interest of the technique, which is of general application. In fact, it seems that the search in a compressed space, achieved by LASCA, may allow the discovery of more promising regions than resorting only to the optimization in the original problem space with a traditional metaheuristic.

VIII. References

[1] T. Bengtsson, P. Bickel and B. Li, "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems," *Probability and Statistics: Essays in Honor of David A. Freedman*, vol. 2, pp. 316-334, 2008.

[2] K. Gallagher and M. Sambridge, "Genetic algorithms: a powerful tool for large-scale nonlinear optimization problems," *Computers & Geosciences*, vol. 20, no. 7, pp. 1229-1236, 1994.

[3] T. S. Duque, K. Sastry, A. C. Delbem and D. E. Goldberg, "Evolutionary Algorithm for Large Scale Problems," in *ISDA 2007 - Seventh International Conference on Intelligent Systems Design and Applications*, Rio de Janeiro - Brazil, 2007.

[4] Z. Yang, K. Tang and X. Yao, "Large scale evolutionary optimization using cooperative coevolution," *Information Sciences*, vol. 178, no. 15, pp. 2985-2999, 2008.

[5] I. T. Jolliffe, "Principal Component Analysis", *Springer Series in Statistics*, Springer, 2002.

[6] L. Costa, "Application of Evolutionary Swarms and Autoencoders to Wind-Hydro coordination," M.Sc. Thesis, MIEEC, Faculty of Engineering of the University of Porto, Portugal, 2008.

[7] V. Miranda and N. Fonseca, "EPSO - best-of-two-worlds metaheuristic applied to power system problems," in *CEC '02 - 2002 Congress on Evolutionary Computation*, Honolulu, HI , 2002.

[8] P. N. Suganthan et al., "Problem Definitions and Evaluation Criteria for the Special Session on Real-Parameter Optimization", CEC, Singapore, 2005.

[9] S. Haykin, "Neural Networks: a Comprehensive Foundation", Hamilton, Ontario: Pearson - Prentice Hall, 1999.

[10] M. Daszykowski, B. Walczak and D. L. Massart, "A journey into low-dimensional spaces with autoassociative neural networks," *Talanta*, vol. 59, no. 6, pp. 1095-1105, 2003.

[11] M. Marseguerra and A. Zoia, "The Autoassociative Neural Network in signal analysis: I. The data dimensionality reduction and its geometric interpretation," *Annals of Nuclear Energy*, vol. 32, no. 11, pp. 1191-1206, 2005.

[12] M. Marseguerra and A. Zoia, "The AutoAssociative Neural Network in signal analysis: II. Application to on-line monitoring of a simulated BWR component", *Annals of Nuclear Energy*, vol. 32, no. 11, pp. 1207-1223, 2005.

[13] S. Narayanan, R. J. Marks, J. L. Vian, J. J. Choi, M. A. El-Sharkawi and B. B. Thompson, "Set constraint discovery: missing sensor data restoration using autoassociative regression machines," in IJCNN'02, *Proceedings of the 2002 International Joint Conference on Neural Networks*, Honolulu, Hawaii, 2002.

[14] G. W. Cottrell, P. Munro and D. Zipser, "Learning internal representations from gray-scale images: An example of extensional programming," in *9th Annual Cognitive Science Society Conference*, Seattle (WA), USA, 1987.

[15] M. K. Fleming and G. W. Cottrell, "Categorization of faces using unsupervised feature extraction," in *IJCNN - International Joint Conference on Neural Networks*, San Diego (CA), USA, 1990.

[16] B. Golomb and T. Sejnowski, "Sex Recognition from Faces Using Neural Networks," in *Applications of Neural Networks*, A. Murray, Ed., Kluwer Academic Publishers, 1995, pp. 71-92.

[17] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504-507, 2006.

[18] V. Miranda, J. Krstulovic, J. Hora, V. Palma and J. C. Príncipe, "Breaker status uncovered by autoencoders under unsupervised maximum mutual information training," in *ISAP – Intelligent Systems Applications to Power Systems*, Japan, Jun 2013.

[19] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *IEEE International Conference on Neural Networks*, Perth, WA, 1995.

[20] V. Miranda, H. Keko and A. Jaramillo Duque, "Stochastic star communication topology in Evolutionary Particle Swarms (EPSO)", *Intern. Journal of Computational Intelligence Research*, vol.4, no.2, pp. 105-116, 2008

[21] V. Miranda and R. Alves, "Differential Evolutionary Particle Swarm Optimization (DEEPSO): a successful hybrid", *Procs. of the 1st BRICS-CCI*, Porto de Galinhas (PE), Brazil, 8-11 September 2013

[22] N. W. Oo and V. Miranda, "Multi-energy Retail Market Simulation with Intelligent Agents", *Proceedings of IEEE St. Petersburg Power Tech*, St. Petersburg, Russia, June 2005

[23] H. Mori and Y. Komatsu, "A Hybrid Method of Optimal Data Mining and Artificial Neural Network for Voltage Stability Assessment", *Proceedings of IEEE St. Petersburg PowerTech*, Russia, June 2005

[24] V. Miranda, H. Keko and A. Jaramillo Duque, "EPSO: Evolutionary Particle Swarms", Ch. 6 in "*Advances in Evolutionary Computing for System Design*", Ed. L.C. Jain, V. Palade and S. Srinivasan, Springer Studies in Computational Intelligence, vol.6, pp. 139-168, 2007

[25] Mehdi Eghbal, E.E.El-Araby, Naoto Yorino and Yoshifumi Zoka, "Application of Metaheuristic Methods to Reactive Power Planning: A Comparative Study for GA, PSO and EPSO", *Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, pp. 3755-3760, Montreal (Quebec), Canada, Oct 7-10,2007

[26] D. Midence and A. Vargas, "Comparative study of Evolutionary Computing Algorithms in the Optimization of Reliability in Power Distribution Networks" (in Spanish), *XIII ERIAC – CIGRE*, Puerto Iguazú, Argentina, May 2009

[27] Momin Jamil; Xin-She Yang, "A literature survey of benchmark functions for global optimisation problems", *Int. J. of Mathematical Modelling and Numerical Optimisatio*n, vol.4, no.2, pp.150-194, 2013

[28] H. Yan, P. B. Luh, X. Guan and P. M. Rogan, "Scheduling of hydrothermal power systems," *IEEE Transactions on Power Systems*, vol. 8, no. 3, pp. 1358-1365, 1993.

[29] V. R. Sherkat, R. Campo, K. Moslehi and E. O. Lo, "Stochastic Long-Term Hydrothermal Optimization for a Multireservoir System," *IEEE Transactions on Power Apparatus and Systems*, vol. 104, no. 8, pp. 2040-2050, 1985.

[30] B. G. Gorenstin, N. M. Campodonico, J. P. da Costa and M. V. F. Pereira, "Stochastic optimization of a hydro-thermal system including networkconstraints," *IEEE Transactions on Power Systems*, vol. 7, no. 2, pp. 791-797, 1992.

[31] V. Miranda, D. Srinivasan and L. M. Proença, "Evolutionary computation in power systems," *International Journal of Electrical Power & Energy System*s, vol. 20, no. 2, pp. 89-98, 1998.

[32] C. E. Zoumas, A. G. Bakirtzis, J. B. Theocharis and V. Petridis, "A Genetic Algorithm Solution Approach to the Hydrothermal Coordination Problem," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1356-1364, 2004.

[33] E.D. Castronuovo and J. A. Peças Lopes, "On the Optimization of the Daily Operation of a Wind-Hydro Power Plant", *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1599 - 1606, 2004.

[34] M. Shahidehpour and M. Khodayar, "Coordination of wind and pumped-storage hydro units for managing transmission security", *IEEE PES General Meeting*, San Diego CA, USA, Jul 2012

[35] E. Castronuovo et al., "An Integrated Approach for Optimal Coordination of Wind Power and Hydro Pumping Storage", in *Wind Energy* early view, 2013, DOI : 10.1002/we.1600.

[36] J. W. Labadie, "Optimal Operation of Multireservoir Systems: State-of-the-Art Review," *Journal of Water Resources Planning and Management*, vol. 130, no. 2, pp. 93-111, 2004

IX. Biographies

Vladimiro Miranda (M'90, SM'04, F'06) received his Ph.D. degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 1982 in Electrical Engineering. In 1981 he joined FEUP and currently holds the position of Full Professor.

Prof. Miranda is a researcher at INESC since 1985 and is currently Director at INESC Porto, the leading institution of INESC TEC – INESC Technology and Science, an advanced research network in Portugal. He is also President of INESC P&D Brasil. He has authored many papers and been responsible for many projects in areas related with the application of Computational Intelligence to Power Systems.

Prof. Miranda was the 2013 recipient of the IEEE PES Ramakumar Family Renewable Energy Excellence Award.

Joana da Hora Martins received a M.Sc. degree in Industrial Engineering in 2009 at the Faculty of Engineering of the University of Porto (FEUP), and a M.Sc. degree in Quantitative Methods in 2012 at the Faculty of Economics of the University of Porto (FEP), Portugal. She is currently a PhD student at FEUP. She was a researcher at the department of Industrial Engineering of FEUP (April 2009 - August 2011), and at Power Systems Unit of INESC-TEC (November 2011 - June 2013), Portugal.

Vera Palma graduated from FCUP, in 2007 in Applied Mathematics and is currently a researcher at INESC TEC, Portugal, in the Power Systems Unit.