Chapter 17

Dynamics of a swarm

Maurice Clerc. 2005-02-03. Rough translation into English of the book published by Hermès-Lavoisier "L'optimisation par essaims particulaires"

17.1. Motivations and tools

At the time when these lines are written (during 2004), there is still no satisfactory theoretical analysis of PSO. The reason is that the problem is not simple, because of the interactions between particles. It has been well known, since Poincaré, that the evolution of such systems can lead to a literally indescribable chaos. It would however be quite interesting to have elements more reliable than the simple overall experimentation as guides of improvement of the method. How to go about it ?

We have here mobile particles which influence each other, admittedly in general in a space of size quite higher than that of the spaces defined in physics, but the tools and methods of statistical dynamics are perhaps usable, except for two points :

the field implemented is more complicated than, for example, a single gravitational field, since each particle is influenced only by some other informants, and not by all. If one wants to push the analogy further, it would thus be necessary to consider the simultaneous influence of several fields of various natures;

the size of the swarm is generally low. The variance of statistical quantities defined on such a small population is extremely likely to be very large.

Nevertheless, studies based on such modelings are underway, but without having, for the time being produced results usable in practice.

A less ambitious step consists in considering a swarm reduced to only one particle. That can seem paradoxical, since precisely the interactions play a crucial role, but in fact, as we saw in the chapter on the memory-swarm, when one speaks about only one particle, it is according to the historical terminology. Actually, there are always at least two of them: the explorer and the memory. Mathematical analyses are then possible and have indeed provided, as already mentioned, precise recommendations which are theoretically validated as for the choices of the confidence coefficients, in particular *via* the coefficients of constriction [CLE 02, TRE 03, VAN 02].

We will not study them again here, the more so as they are rather unpleasing (the amatheurs will be able to relish it at the end of this chapter), but rather study in detail a very simple example and the lesson which we can already draw.

17.2. An example with the magnifying glass

Let us consider the function Parabola 1D, defined on $[-20\ 20]$ by the equation. $f(x) = x^2$.We wish a particle to find the minimum (zero, obviously), with a precision at least equal to. $\varepsilon = 0{,}001$. In other words, a particle at least must reach a position located in the interval. $\left[-\sqrt{\varepsilon}\ \sqrt{\varepsilon}\right]$. The theoretical difficulty of this problem is 6,45. To still simplify the analysis, we will use only version OEP 0.

The question which interests us is the influence of the interactions on the effectiveness of the algorithm. That's why we will consider and compare the results obtained with a swarm reduced to only one particle (in fact, as has already been said, an explorer and a memory) and a swarm of two particles (makes some 2 + 2).

17.2.1. *One particle*

With only one particle, the equations of motion can be written in a simplified form:

$$\begin{cases} v \leftarrow c_1 v + c_2 (p - x) \\ x \leftarrow x + v \end{cases}$$

There are primarily two cases: either the initial velocity is such that the first two positions frame that of the minimum, or on the contrary these two positions are on the same side. In the first case the particle oscillates around the optimal position, in the second it tends there directly, at the latest at the very second step of time (see figure 17.1). Let x_2 be the position reached with the second step of time.

The significant point when the first positions are on the same side is that the memorized position p is then always equal to the current position *x*. There is certainly constant improvement, but each displacement is strictly equal to the preceding one multiplied by c_1. There can therefore be convergence towards the optimum only if the infinite sum of successive displacements $\left(\sum_{t=2}^{\infty} c_1^{t-1} \right) |v_2|$ is at least equal to. $|x_2|$. However, in our example we have:

$$\begin{cases} c_1 = 0{,}7 \\ v_2 = 3{,}2 \\ x_2 = -20 \end{cases}$$

Under these conditions the total distance travelled by the particle even at the end of an infinite time cannot exceed approximately 10.7. It is insufficient to reach the optimum. On the other hand if the particle oscillates around this optimum, the things occur completely differently, because the last best position known is not necessarily any more the current position. Velocity still will tend towards zero but not at all any more in a regular way, which prevents a premature convergence.

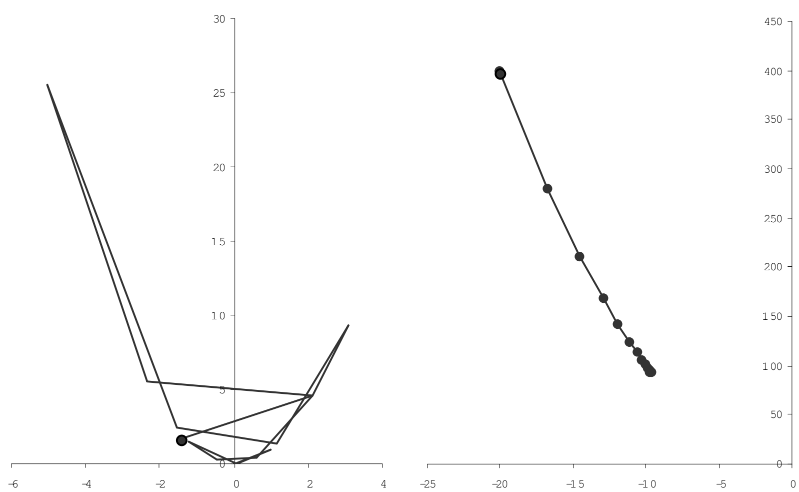
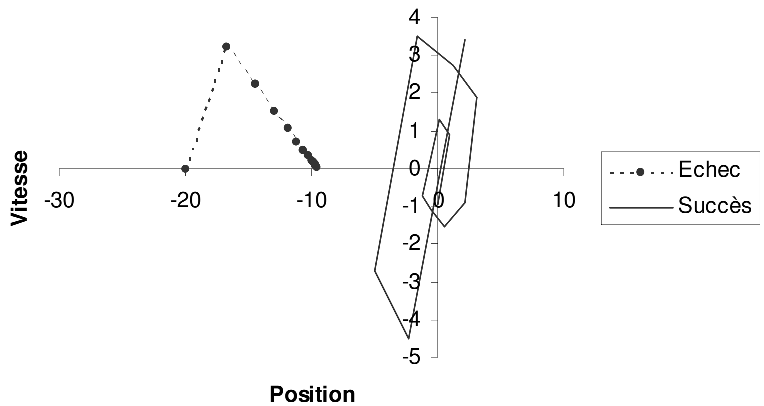


Figure 17. 1. *Parabola problem, a memory and an explorer. The behavior of the system is very different according to the position and the initial velocity of the explorer. On the left the first two positions do not frame that of the minimum. The fast velocity decrease prevents the explorer reaching it. On the right, the first two positions frame that of the minimum. Here using the memory leads the explorer to oscillate around the optimal position and, moreover, the velocity decreases less quickly, which allows convergence*

These two types of behavior are still better highlighted by the representation in the space of the phases in figure 17.2. The traditional fundamental structure which appears almost systematically in the event of convergence towards the solution is that in spiral on the right part of the figure. As we will see it, whether there is one particle or more (in the sense: explorer + memory) does not make any difference. For the algorithm to proceed successfully it is necessary, except in very particular cases, that successive positions are reached on either side of the optimal position: it is necessary that there are oscillations. Mathematically, that is translated in our example by the fact that the scalar product of vectors v and $p - v$ must be negative. In this form, it is a necessary condition which can be generalized whatever be the dimension of the search space and the number of particles.



vitesse= velocity
échec= failure
position= position
succès= success

Figure 17. 2. *Parabola problem, a memory and an explorer, space of the phases. The two cases of figure 16.1 are taken again here, but seen in the plan (position, velocity). The converging oscillatory behavior is represented by a spiral*

17.2.2. *Two particles*

Let us recall that the term "particle" is taken here with the historical meaning, i.e. it covers in fact the double concept of explorer and memory. We thus have now two explorers and two memories. To formally remain identical to the original PSO, the links of information are those represented in figure 17.3. Each memory informs the two explorers but each explorer informs only one memory. Initializations are deliberately defined identical to those, position and velocity, of the single particle of the preceding example, and in both same cases: the first two positions on the same side of the origin or on both sides.

The right part of figure 17.4 represents the course of particle 2 in this second case. It is absolutely identical to that on the right part of figure 17.1 for at any moment memory 2 is in better position than memory 1. Thus the existence of particle 1 does not bring anything to particle 2. On the other hand, as the comparison between the left parts of the two figures shows it, particle 1 benefits from the existence of particle 2. The information which is provided to it *via* memory 2 makes it possible for it to also enter an oscillatory process which would ensure to it convergence if the iterations were prolonged beyond the success of particle 2.

Starting from a more unspecified initial configuration, we obtain the courses illustrated by figure 17.5. Each particle benefits from the information provided by the other to manage to oscillate around the optimal position. Thus, on the one hand the number of evaluations is doubled with each step of times, since there are two particles, but, on the other hand, on a simple case like this one, the probability of reaching the solution (with the precision required) is itself roughly doubled. It results from it that, *roughly speaking*, the effectiveness is the same, about thirty evaluations are needed to reach the solution.

But then, what is the point in using several particles? It is that the case of two particles is precisely the limit beyond which the increase in the size of the swarm will become interesting. The force of PSO lies in the fact that the probability of reaching the solution with the step of time T increases more or less as $N(N-1)$, where N is the size of the swarm, whereas the number of evaluations carried out on the whole until the step of time T is only proportional to N (reasoning with N constant during optimization). This is yet only one empirical conclusion, which moreover is nolonger valid for greater values of *N*, but it provides a beginning of explanation as for the increase in the effectiveness with that of *N*. Studies in progress, in particular using statistical mechanics, try to validate this observation more firmly.

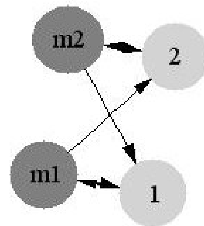


Figure 17. 3. *Graph of information 2 + 2. Each explorer receives the information from the two memories, but informs only one, to conform to the diagram of the traditional PSO, where explorer and memory are merged in only one particle. That is enough for any improvement found by an explorer to be known to the other one.*

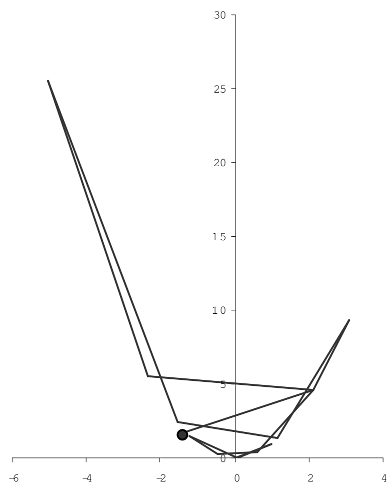


Figure 17. 4. *Two explorers and two memories. The starting points are the same as in the preceding example. But now the particles work together. However, here, memory 2 is always better than memory 1: the course of the explorer 2 is exactly the same as seen previously in the event of convergence (figure on right-hand side). On the other hand, explorer 1 will benefit from the information provided by memory 2: it will end up converging if the iterations are continued (figure on the left).*

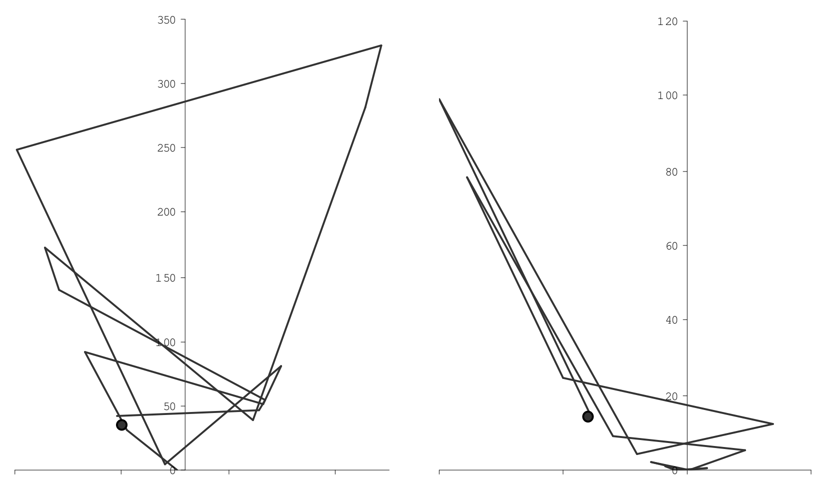
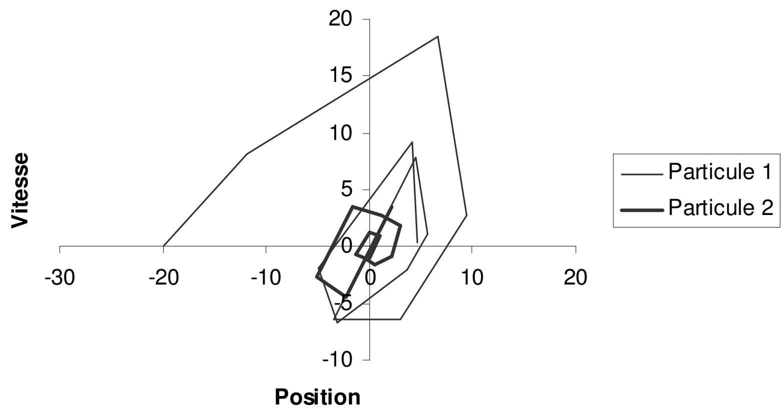


Figure 17. 5. *Parabola. Two explorers and two memories. We are here in the more general case where each explorer is from time to time influenced by the memory of the other, when it becomes better than its own. Convergence, when it takes place, is not necessarily faster (here 30 iterations instead of 28), but it is more probable.*



vitesse= velocity
position= position
particule 1= particle 1
particule2= particle 2

Figure 17. 6. *Parabola. Two explorers and two memories. Representation of the courses in the space of the phases. The particles help each other to enter and remain in the oscillatory process which allows convergence towards the solution*

17.3. Energies

17.3.1. *Definitions*

We benefitted from the preceding examples to underline the interest of the representations by trajectories in spaces of phases. Another traditional method in dynamics is to consider the evolution of global variables, such as potential energy, kinetic energy or entropy. We will be satisfied here with the first two, and even especially with the kinetic energy, but first of all it is necessary to give precise definitions of them. We will admit that the search space is provided with a distance and that the function to minimize f is numerical.

For the kinetic energy of a particle, whether it is explorer or memory, it is very simple, it is enough to consider two successive positions x_t and x_{t+1} and to calculate the size. $e = \|x_t - x_{t+1}\|^2 / 2$. It is deliberately that velocity has not been explicitly called upon, for, as we saw for example with the method of the pivots, this variable is not used in all the versions of PSO. Naturally, the total kinetic energy of a swarm is the sum of those of the particles which are its components.

For the potential energy, it is necessary to take into account the required precision ε as for the desired solution. In addition we can benefit from the fact that this type of energy can be known only except for one additive constant, which avoids us having to use the value of f in its minimum, in general unknown. Then the potential energy u of a position x can be defined by the formula. $u = f(x)/\varepsilon$. It can be interpreted as the number of "steps" of the height ε which the particle should still scale down to reach the minimum if its value were zero. Here still, the total potential energy of a swarm is the sum of those of its components.

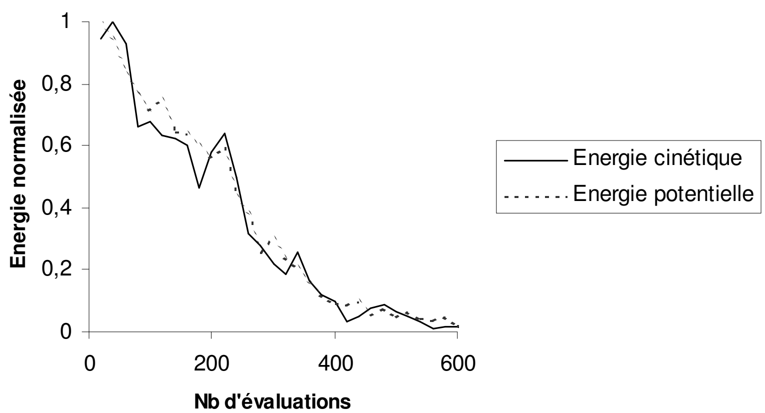
17.3.2. *Evolutions*

We already know that, approximately, velocities tend to be cancelled during the iterative process and that the swarm ends up converging somewhere, even if it is not the desired solution. We can thus expect that the kinetic energy tends on average towards zero and potential energy towards a constant. But what is particularly instructive to observe using these quantities, is the difference in behavior between a version of traditional PSO (N particles given at the beginning once and for all) and an adaptive version for which the size of the swarm is modified as suppressions and generations dictate.

Let us take for example the Alpine function. An execution with good parameters easily gives evolutions of energies such those of figure 17.7. The fact that the kinetic energy tends towards zero tell us that, overall, the swarm ceases moving, therefore that each particle converges towards a fixed position. The fact that, at the same time, the potential energy also tends towards zero means that, for all the particles, this position is indeed that of the sought minimum. Here, the number of explorers is equal to 20, just as the number of memories. The light fluctuations of energies, and in particular of the kinetic energy, are due only to the share of the chance in the equations of motion.

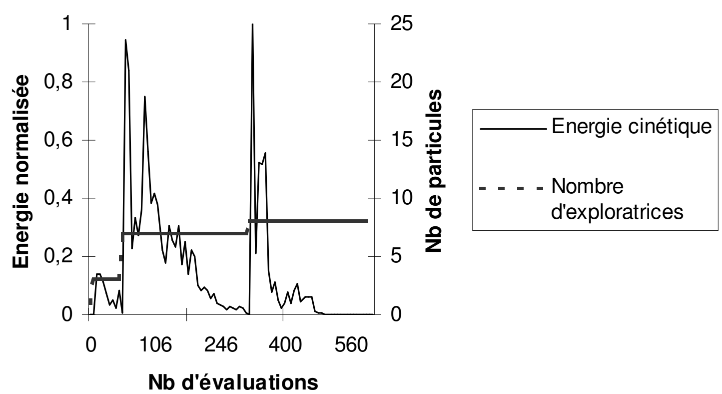
With an adaptive PSO like TRIBES, each particle is at the same time explorer and memory, but their number is modified during the process, by suppressions and generations. Figure 17.8 then indicates the clear evolution of the size of the swarm as well as that of the kinetic energy. When this one tends towards zero and that the objective therefore is not achieved (what is not represented on the figure, for reason of clarity), then there are appreciably more generations than suppressions, which creates a peak of energy, that we can interpret as a reaugmentation of the search power of the swarm.

As we have seen in the chapter on the optimal parameters, it is certainly possible, on this example, to find a solution more quickly than with TRIBES, but, on the another hand, it is also possible to be much less effective if the parameters are not properly selected. An adaptive version, because it starts again exploration judiciously, is much more sound.



Energie normalisée= normalized energy
Nb d’évaluations= number of evaluations
Energie cinetique= kinetic energy
Energie potentielle= potential energy

Figure 17. 7. *Alpine, treated by parametric PSO. The size of the swarm is constant (20) here. The small variations of energy are only fluctuations due to the partially random character of the equations of motion*



Energie normalisée= normalized energy
Nb d’évaluations= number of evaluations
Energie cinetique= kinetic energy
Nb de particules= number of particles
Nombre d’exploratrices= number of explorers

Figure 17. 8. *Alpine, treated by adaptive PSO (TRIBES). During the process, particles are removed and others added. The net assessment is a constant increase in the swarm (it is not always the case) but, especially, of the significant peaks of energy relaunching the exploration when that becomes necessary.*

17.4. For experienced amatheurs: convergence and constriction

The complete analytical study was made for the moment only in the case of only one particle and with constant confidence coefficients (non random). Some of its elements are given below. For more details, see [CLE 02, TRE 03, VAN 02].

17.4.1. *Criterion of convergence*

In the case of only one particle, the equations of motion can be written:

$$\begin{cases} v_{t+1} = c_1 v_t + c_2 (p_t - x_t) + c_3 (g_t - x_t) \\ x_{t+1} = x_t + v_{t+1} \end{cases}$$

where the indices t and $t + 1$ correspond to two steps of successive times. Laying down:

$$\begin{cases} c = c_2 + c_3 \\ p = \dfrac{c_2 p_t + c_3 g_t}{c_2 + c_3} \end{cases}$$

we obtain the canonical system:

$$\begin{cases} v_{t+1} = c_1 v_t + c(p - x_t) \\ x_{t+1} = x_t + v_{t+1} \end{cases}$$

The idea is to look at what occurs as long as p is constant. We can then pose $y = p - x$ and the system becomes, written in matrix form:

$$\begin{bmatrix} v_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} c_1 & c \\ -c_1 & 1-c \end{bmatrix} \begin{bmatrix} v_t \\ y_t \end{bmatrix} = C \begin{bmatrix} v_t \\ y_t \end{bmatrix}$$

We now have a traditional dynamic system, whose behavior is entirely dependent on the eigenvalues of the matrix C. In particular, a condition of convergence is that these eigenvalues are two conjugated complex numbers of module lower than 1 or two real numbers of absolute values lower than 1. They are solutions of the equation:

$$\begin{vmatrix} c_1 - \lambda & c \\ -c_1 & 1 - c - \lambda \end{vmatrix} = \lambda^2 + (c - c_1 - 1)\lambda + c_1 = 0$$

whose discriminant is $\Delta = (c - c_1 - 1)^2 - 4c_1$

Let us recall that here convergence means simply that the particle tends towards a stable position (velocity tends towards zero). Nothing guarantees that this position is the sought optimum. It is only the interactions between particles which considerably increase the chances that may be the case.

17.4.2. *Coefficients of constriction*

In traditional PSO, it can happen that the swarm "explodes" (divergence) and this is why certain authors add a constraint of maximum velocity, which makes an additional parameter. It was proven [CLE 02] that was not necessary, with the proviso of using one or more coefficients of constriction, calculated starting from the confidence coefficients. To determine them, there are primarily two steps:

to make sure that the eigenvalues of C are true complex numbers (negative discriminant);

or to weight the confidence coefficients judiciously when the eigenvalues are real (positive or null discriminant).

There is an infinity of possibilities. Let us examine simply one of each type, which gives place to relatively simple formulas. The general idea is to pass by an intermediate parameter φ, according to which one expresses c and c_1, so as to respect the criterion of convergence.

Negative discriminant

A simple form corresponds to the following relations:

$$\begin{cases} c_1 = \chi(\varphi) \\ c = \chi\varphi \end{cases}$$

The matrix of the system is then:

$$C = \begin{bmatrix} \chi & \chi\varphi \\ -\chi & 1-\chi\varphi \end{bmatrix}$$

We will seek the coefficient of constriction χ function of φ as near to 1 as possible, while guaranteeing that the discriminant of the equation whose solutions are the eigenvalues of the system remains negative. The condition "negative discriminant " is written here:

$$\chi^2(1-\varphi)^2 - 2(1+\varphi) + 1 < 0$$

It is satisfying only if χ is between the two roots:

$$\chi_{\min} = \frac{1+\varphi-2\sqrt{\varphi}}{(1-\varphi)^2} \text{ and } \chi_{\max} = \frac{1+\varphi+2\sqrt{\varphi}}{(1-\varphi)^2}$$

It is easy to see that $\chi_{\min}$ is always lower than 1 (for φ positive). For $\varphi \leq 4$ the second root $\chi_{\max}$ is higher or equal to 1. We can thus take a coefficient of constriction equal to 1, i.e. in fact no constriction at all. On the other hand, for $\varphi > 4$, the coefficient nearest to 1 that we can take is $\chi_{\max}$ itself. Constriction is thus summarized by the following formulas:

$$\begin{cases} c_1 = \chi(\varphi) = \begin{cases} 1 \text{ si } \varphi \leq 4 \\ \dfrac{1+\varphi+2\sqrt{\varphi}}{(1-\varphi)^2} \text{ si } \varphi > 4 \end{cases} \\ c = c_1\varphi \end{cases}$$

The common value of the module of the complex eigenvalues is then simply $\sqrt{c_1}$ which is at the most equal to 1: the criterion of convergence is satisfied. The interested reader will be able to find similar formulas (and even simpler ones) starting for example from the relations:

$$\begin{cases} c_1 = \chi(\varphi) \\ c = 1+\chi-\chi\varphi \end{cases} \text{ or } \begin{cases} c_1 = \chi(\varphi) \\ c = 1-\chi+\chi\varphi \end{cases}$$

Finally, the equations of motion are written:

$$\begin{cases} v_{t+1} = \chi v_t + \chi\varphi(p - x_t) \\ x_{t+1} = x_t + v_{t+1} \end{cases}$$

17.4.3. *Positive discriminant*

A canonical system canonical even simpler than that we have seen can be written:

$$\begin{cases} v_{t+1} = v_t + \varphi(p - x_t) \\ x_{t+1} = x_t + v_{t+1} \end{cases}$$

By posing $y = p - x$, the system becomes:

$$\begin{cases} v_{t+1} = v_t + \varphi y_t \\ y_{t+1} = -v_t + (1 - \varphi) y_{t+1} \end{cases}$$

Its matrix is:

$$C = \begin{bmatrix} 1 & \varphi \\ -1 & 1 - \varphi \end{bmatrix}$$

A possible method of constriction consists in multiplying the whole of the matrix by a coefficient. χ' . The eigenvalues are then solutions of the equation:

$$\begin{vmatrix} \chi' - \lambda & \chi'\varphi \\ -\chi' & \chi'(1 - \varphi) - \lambda \end{vmatrix} = \lambda^2 + \chi'(\varphi - 2)\lambda + \chi'^2 = 0$$

We find then:

$$\lambda = \chi' \left(1 - \frac{\varphi}{2} \pm \frac{\sqrt{\varphi^2 - 4\varphi}}{2} \right)$$

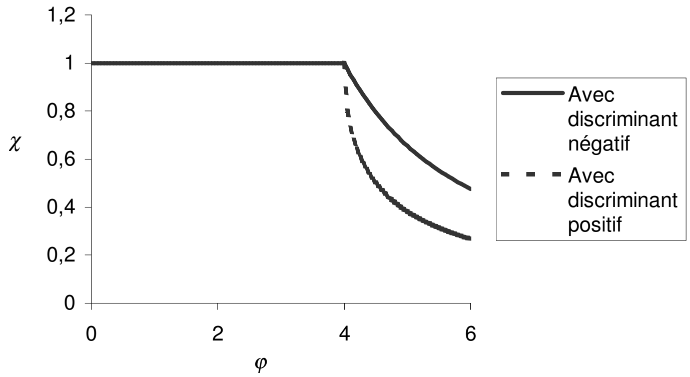
These values are real only if one has $\varphi \geq 4$. So that their absolute values are at the most equal to 1, it is necessary and it is enough that the largest is, which gives us directly:

$$\chi' = \frac{2}{\varphi - 2 + \sqrt{\varphi^2 - 4\varphi}}$$

According to the way in which it was found, this coefficient is applicable to the equations of motion:

$$\begin{cases} v_{t+1} = \chi' v_t + \chi'\varphi(p - x_t) \\ x_{t+1} = x_t + v_{t+1} + (p - x_t)(1 - \chi' + \chi'\varphi) \end{cases}$$

whose physical interpretation is far from being obvious, owing to the fact that a corrective term is applied to displacement due only to velocity. However it is easy to check that one always has $\chi' \leq \chi$ (see figure 17.9). Thus, systematic use of χ' whatever be the scenario (positive or negative discriminant) is mathematically acceptable. In the case of a negative discriminant, constriction is certainly a little too strong, but, in practice, the coefficient φ is taken very slightly higher than 4, which reduces the risk.



Avec discriminant négatif= With negative discriminant
Avec discriminant positif= With positive discriminant

Figure 17.9. *Coefficients of constriction. The two methods indicated in the text lead to different formulas. However the second can be used whether the discriminant of the system is positive or negative since the value obtained is in any case lower than the greatest acceptable value calculated by the first. Nevertheless, it is better then to take the values of φ just a bit higher than 4, to avoid a too strong constriction and a premature convergence.*

17.5. Summary

The dynamics of a swarm can be considered on the level of each particle, a privileged tool being the space of the phases. Convergence results then in trajectories in the shape of spiral. It can also be studied on the whole, *via* variables such as the kinetic energy and the potential energy. The examination of the evolution of the kinetic energy shows indeed the difference in behavior between the traditional PSO and an adaptive version in which the size of the swarm varies according to the progression of search: the peaks of energy announce the revival of exploration when convergence seems to become too slow.

In this chapter, the mathematical part, rather long, summarizes the calculation of some coefficients of constriction, the use of which is of great practical importance.