Exercises

1. Check your understanding of the PSO algorithm:
   a. Why do we call the terms with α_1, α_2 “forces” and ω itself “inertia”?
   Answer: This relates to a suggestive physical interpretation: Inertia is the tendency to retain a certain movement, and force is proportional to acceleration, i.e., the change of the velocity. We don’t care about units [such as seconds] here, but we have the parameters α_1, α_2 and ω in place to account for the fact that there is a proportionality rather than an equality of the terms and the respective physical quantities. Besides, inertia in the physical sense would imply ω=1, so the use of arbitrary values of ω could be seen as “degree of inertia”, it is still often called “inertia” in the context of PSO.
   b. Would the algorithm work with negative values for α_1 and/or α_2 ?
   Answer: For negative α’s, the particles are repelled from the current best solutions, and the further the particles are away from the bests the stronger the repulsion gets such that the swarm will diverge even for small negative values of $\alpha_1 < 0$ and $\alpha_2 < 0$. If $\alpha_1 < 0 < \alpha_2$ and $|\alpha_1| < |\alpha_2|$, i.e., the personal best is repulsive, but the global best is attractive, the algorithm can still work, namely, if we assume that **p**=**g,** then the combined effect is still attractive for $\alpha_2 + \alpha_1 > 0$. If the bests are not the same, they will more and more appear to lie in the same direction, if the particle moves further outwards, so the combined force from that region will attract the particle. If the particle returns, it will again be repelled from one of the bests, such that it is not very likely to exploit the good regions. In the opposite case, similarly to above, the algorithm could still work, but its performance might be even worse than in the previous case.
   c. How well would the algorithm work for $\alpha_1 >> \alpha_2 > 0$ or for $0 < \alpha_1 << \alpha_2$?
   Answer: This is not dynamical problem, but the performance will probably not be great, because the stronger global best seems preferable to a stronger personal best for relatively simple problems. For complex problems it could be useful to maintain a certain diversity which is achieved best by the personal bests. In simulations, it often seems to be the case that the sum of the two (or perhaps their quadratic sum) is the more essential parameter as compared to their individual values. However, if one of the α’s is close to zero, then the performance usually drops.
   d. What is the benefit from using ω close to 1? What is the downside of this?
   Answer: We assume there that $|\omega|<1$, then the particles continue for a long time and perform large excursions, before being retracted by the distance-dependent forces. The benefit is more exploration, the downside is less exploitation. In addition, a modulus of ω close to 1 leads to increased instability. Namely, if the forces draw in the same direction as the velocity, then the particle is accelerated by the combined effect. Due to the inertia it will pass the bests and will be slowed down, but for slightly stronger forces this slow-down will not be enough, such that the particle diverges. Weak forces on the other hand don’t allow the particle to come often near the promising regions such that the performance will not be good.
   e. Would the algorithm necessarily diverge if $\omega \geq 1$?
   Answer: Yes. It may seem that this is not necessary as the forces retract the particles, but if there is no damping in the system the oscillation about the bests built up and divergence is seen as a spiralling-away. For $|\omega|=1$ the system is in a limit case, so for non-zero forces it will still diverge, as randomness adds up and there is no damping. For zero forces, the particle simply follows its initial velocity, if this is non-zero.

f. Would it work with negative values for ω?

Answer: This is no problem, unless $|\omega| \geq 1$, see above. Negative inertia does not occur in standard physical systems, as it means that the velocity flips-over in every time step, but this is no problem for PSO which is a purely computational system. The strong velocity changes at every time step do in fact not make much of a difference, except that they interfere differently with the force terms, so the situation in not completely symmetric with respect to positive or negative ω.

g. Discuss how diversity can be maintained in a particle swarm.

Answer: The discussions above should give some hints: values of $|\omega|$ close to 1 help, also overshooting forces ($\alpha_1 + \alpha_2 > 4$). The multiplicative randomness doesn't help much towards diversity, as we have discussed in relation to "biases in PSO".

2. Consider a particle "swarm" consisting of a single particle with only a personal best.

a. How does a deterministic PSO particle move in a one-dimensional search space? Assume the random factors are constant and equal to 1, and that the personal best never changes. Try to solve the problem analytically.
Hint: Consider a matrix equation for the 2D vector $(v,x)^T$!

Answer:

The dynamics of this simplified model is

$$v_{t+1} = \omega v_t + \alpha (p - x_t)$$
$$x_{t+1} = x_t + v_{t+1}$$

which can be written also as

$$v_{t+1} = \omega v_t + \alpha (p - x_t)$$
$$x_{t+1} = x_t + \omega v_t + \alpha (p - x_t)$$

where we have inserted the first equation into the second one. This equation

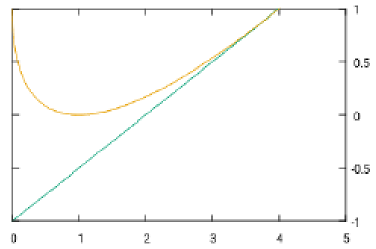
$$\begin{pmatrix} v_{t+1} \\ x_{t+1} \end{pmatrix} = \begin{pmatrix} \omega & -\alpha \\ \omega & 1-\alpha \end{pmatrix} \begin{pmatrix} v_t \\ x_t \end{pmatrix} + \alpha \begin{pmatrix} p \\ p \end{pmatrix}$$

is the same in matrix form, so we need to consider the eigenvalues of $\begin{pmatrix} \omega & -\alpha \\ \omega & 1-\alpha \end{pmatrix}$ which can be found from $\lambda_1 \lambda_2 = \omega (1-\alpha) + \omega\alpha = \omega$ and $\lambda_1 + \lambda_2 = \omega + 1 - \alpha$, which leads to

$$\lambda_1 = -\frac{\sqrt{\omega^2 - (2\alpha+2)\omega + \alpha^2 - 2\alpha + 1} - \omega + \alpha - 1}{2}$$

$$\lambda_2 = \frac{\sqrt{\omega^2 - (2\alpha+2)\omega + \alpha^2 - 2\alpha + 1} + \omega - \alpha + 1}{2}$$

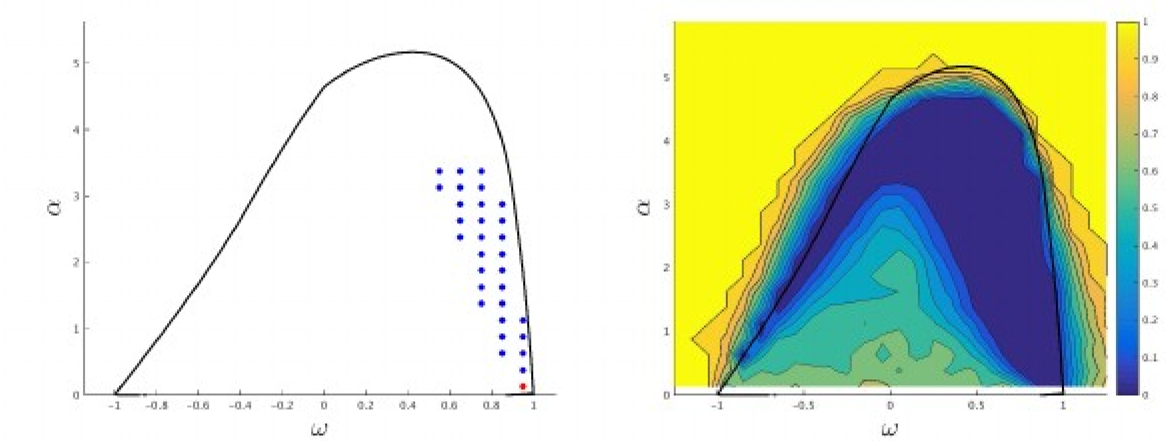
The eigenvalues can now be analysed whether they are real or complex, the latter indicating oscillatory behaviour. More importantly, we should check the real parts of the eigenvalues which indicate stability (if $|\Re(\lambda)| < 1$).



Plot of α (x-axis) vs. ω (y axis): the upper area within the curve has negative discriminant, i.e. oscillatory behaviour. If $\alpha = 1$ and $\omega = 0$ (i.e. at the minimum of the curve) the particle jumps directly to the personal best. In the area left of the straight line both eigenvalues have modulus smaller than one. For example, if $\omega = 0$ and $\alpha < 2$ then the particles in this approaximation are convergent. If you like to check in the literature (Jiang et al., Inf. Proc. Lett. 2007), please note that the noise enters in a different way. The simplicity of the assumptions lead in either case to a different behaviour than in the actual algorithm. E.g. the possibility to use negative ω does not show in the present approximation.

b. What would happen in this case in higher dimensions?
   Answer: If we ignore the noise, then the picture is not really different in higher dimensions, because the dimensions are essentially not interacting in linear dynamics.
c. Without aiming for the maths, discuss the effect or the noise in the original algorithm.
   Answer: This is when things are getting difficult, as noise has a number of effects. It will blur any expected boundaries of stability, it will lead to long excursions before the swarm converges, even if the parameters are in the stable region. It also turns out that it is actually stabilising the system, i.e., while the dynamics diverges for large parameters from a certain level, the noise is important at the high parameter values. For a more comprehensive understanding, we need to consider that there are two eigenvectors in the system, one can be unstable, while the other is still stable, and their relation is also dependent on the noise. Also, there may be different effects from the attraction to personal and global best. The question is, whether the resulting particle dynamics actually helps for the optimisation tasks or not.
d. If you like, you can again discuss the PSO search biases here.
   Answer: If a particle, its personal best and the global best have one coordinate in common, then there is no escape, unless another particle becomes the new global best and re-introduces activity in this coordinate. This effect may interfere with stability, but one can assume that after some (long) time no new global bests are found, such that the effect is asymptotically negligible.

3. [**Numerical exercise**] Compare your findings (or intuition) from the previous question with a simulation of a PSO algorithm. Try also to solve an actual optimisation problem such as the minimisation of $f(x)=x^2$ or of a more complex function. The algorithm will be available before the tutorial, but it should not be too difficult to try writing (or finding) one yourself.



PSO performance for the minimisation of a sphere function $f(x,y)=x^2+y^2$. The left-hand figure shows where the 5% best performing (ω,α) parameter pairs are located (α is the sum of the two alphas, which are equal here). The right-hand figure provides a contour representation of performance over the whole (ω,α) parameter plane. The solid line (both figures) is shows a theoretical result for a balanced swarm, i.e. a swarm that is neither converging nor diverging). In all cases a 25-particle unconstrained PSO algorithm is used averaged over executions. Note that the larger α are possible here, but do not give the best performance, which would be different for problems with many local optima.

4. We have mentioned adding a repulsion term to the velocity rule of PSO.
   a. What happens when the particle is repelled from the globally best particle by which it is also attracted?
      Answer: Repulsion has a different dependence on distance, attraction increases with distance and repulsion decays, i.e., there must be some distance where the two effects are neutralising each other, and this is where the particle remains (or about with point its oscillates), provided that nothing else (e.g., repulsion from any other particle or change of personal best) happens in the meantime.
   b. What other terms could you add in order to adapt the PSO algorithm better to a particular problem?
      Answer: We have mentioned already speed control, i.e., a term that is proportional to the difference of the length of the velocity vector and a target speed and which enters the update question as a multiplicative factor for the speed, e.g., $-\mathbf{v}|\mathbf{v}-\mathbf{v}_0|/(|\mathbf{v}|+|\mathbf{v}_0|)$. There may be attraction to or repulsion from neighbours, alignment to their velocity vectors (compare Reynolds' rules), electromagnetic forces that make particle spiralling away from pairs or triplets of other particles, or see below for a DE like term.
   c. Each of these terms comes with one or more parameters. How can you use a genetic algorithm to choose for you the parameters for the new PSO algorithm with inertia, attractive forces, repulsion, alignment of velocities, …
      Answer: This is quite easy, for, say 5 terms, we can have a 5-bit string that encodes the presence or absence of the term. We can add one more bit per term to include also sign of each term and more bits for the parameter values. In order to determine the fitness of the GA, we ran the multi-term PSO on many problems and use the average performance over these problems as a fitness for the GA.
5. The particles in PSO interact only via the global best which is determined over all particles, compare this with the interaction of individuals in biological swarms, and discuss resulting options for the design of metaheuristic optimisation algorithms.
   Answer: It could be interesting to discuss in this context the results of Iain Couzin, who showed that in hierarchically organised animals (Macaques) it is typically not the superior male who make the decision, but a well networked group of intermediate member of the

troop decide which seems to work well. In more technical terms, this question aims at the effects of topology, as individuals in big animal swarms mostly follow their neighbours, while a leader may not be visible. There are various variants of PSO that incorporate such observations either by defining network topologies different from the all-to-all topologies that is used for fitness comparisons in the original PSO. There are many variants (but there is no need to discuss this in detail):

(i) The swarm can be organised into groups that are fully connected within, i.e., each group has a winner, and only group winners are connected to other winners, i.e., receive attractive forces from them (the group-best forces are zero for the group winners).

(ii) Each particle is compared to a small number (say 4) of other particles. These "neighbours" are chosen randomly in the beginning, but do not need to be close in the search space.

(iii) The neighbours are determined by distance, i.e., the "global" bests are determined only among the (say 6) local neighbours.

(iv) Hybrid algorithms are discussed later, but it is possible to remark already here that hybrids groups are also of interest.