- Some criticisms may be addressed to the standard SETAR model:
  - Although there are situations where an abrupt change in regime takes place, nature usually does not produce discontinuties.
  - It may seem a bit limited that the change in regime be linked to only one lagged sample.
- This has triggered the development of variants to the SETAR which incorporate one (or both) remarks.

- A smoother transition between regimes can be obtained by replacing the indicator function by a continuous function $\phi(.)$ changing gradually from 0 to 1.
- One gets a Smooth Transition AR (STAR) model [1]. With z_n the transition variable, a popular choice for $\phi(.)$ is the logistic function:

$$\phi(z_n;\gamma,r)=\frac{1}{1+\exp[-\gamma(z_n-r)]}$$

- A 2-regime model can be described by:

$$x_n = (a_{10} + a_{11}x_{n-1} + \cdots + a_{1p_1}x_{n-p_1})(1 - \phi(x_n;\gamma;r))$$
$$+ (a_{20} + a_{21}x_{n-1} + \cdots + a_{2p_2}x_{n-p_2})\phi(x_n;\gamma;r) + \varepsilon_n$$

with ε_n is an i.i.d. sequence.

- The parameter γ determines the steepness of the logistic function. The larger it is, the closer $\phi(.)$ is to as step function. When $\gamma = 0$, $\phi(.)$ is a constant.

- Since the STAR model is based on a continuous function, least squares estimation is a maximum likelihood one if the residuals are supposed to be Gaussian.
- Estimation of the sub-models coefficients is easy if γ and r are fixed , but full estimation requires either the use of an optimization algorithm or an exhaustive grid search.

- The idea is quite simple: it consists in defining a partition of the state space itself instead of a partition of the real line for a single threshold variable as for the SETAR:

$$x_n = \sum_{k=1}^{K} \left\{ a_{k0} + a_{k1} x_{n-1} + \cdots + a_{kp_k} x_{n-p_k} + \sigma_k \varepsilon_n \right\} I\left(\boldsymbol{x}_{n-1} \in A_k \right)$$

with $\boldsymbol{x}_{n\text{-}1} = [x_{n\text{-}1}, \ldots, x_{n\text{-}p}]^\top$, p usually taken smaller than or equal to $\max[p_k]$.

- The easiest way to define the partition $\{A_k\}$ is through a set of centers $\{\boldsymbol{c}_k\}$ and the use of the Euclidian distance, that is:

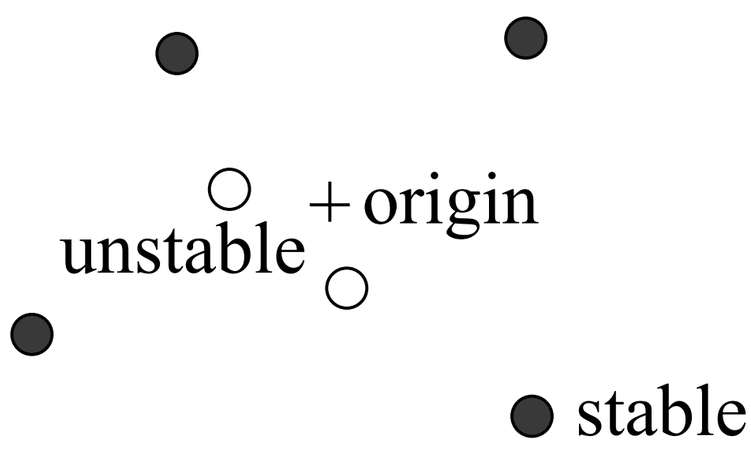
$$\boldsymbol{x}_{n-1} \in A_k \Leftrightarrow \| \boldsymbol{x}_{n-1} - \boldsymbol{c}_k \| \lhd \| \boldsymbol{x}_{n-1} - \boldsymbol{c}_m \|, \; m \neq k$$

- This corresponds to segmenting the state space by hyperplanes. Of course the main question is how to define the centers $\{\boldsymbol{c}_k\}$.

- Three main avenues:
  - The centers are selected to represent the spatial distribution of the $\{\boldsymbol{x}_{n\text{-}1}\}$, with an algorithm such as the Lloyd-Max one.
  - A sequential segmentation of the state space.
  - A heuristic selection (by a genetic algorithm for instance) of the centers among the $\{\boldsymbol{x}_{n\text{-}1}\}$.

- Once the centers are selected, least squares estimation of the AR coefficients for each sub-model is simple.
- Here again, a modified MDL selection criterion can be employed. Quantization of the center coordinates (as for the RBF networks) must be incorporated to prevent selection of a limited number of centers only.

- Concerning stability, the sufficient condition of *drift back to the center* is satisfied if all AR sub-models used when $\|\boldsymbol{x}_n\|$ becomes large are stable. For instance:



- A very interesting alternative has been proposed in [2] under the name *competitive local linear modeling*.
- The opposite course of action is taken: first, the AR sub-models are estimated and second, the partition is performed.
- In this way, it may be that more complex partitions than center-based ones are obtained.

Sub-model estimation

1) First, a set of AR sub-models $\boldsymbol{a}_k = [a_{k0}, \ldots, a_{kp}]^\top$ is initialized randomly.
2) A vector $\boldsymbol{x}_{n\text{-}1} = [1, x_{n\text{-}1}, \ldots, x_{n\text{-}p}]^\top$ is chosen at random in the set of available data.
3) The AR model $\boldsymbol{a}_i$ giving minimum prediction error e_n with:

$$e_n = x_n - \boldsymbol{a}_i{}^\top \boldsymbol{x}_{n\text{-}1}$$

is found.

4) This model is updated using normalized LMS (NLMS):

$$\boldsymbol{a}_i \leftarrow \boldsymbol{a}_i + \beta e_n \boldsymbol{x}_{n-1} \Big/ (\boldsymbol{x}_{n-1}^{\mathsf{T}} \boldsymbol{x}_{n-1})$$

5) Back to 2) until the error variance stabilizes or becomes smaller than some threshold.

Partitioning

Some classifier (MLP, k-nearest neighbors) is trained to establish the correspondence between each $\boldsymbol{x}_{n\text{-}1}$ and the appropriate model.

- The exponential AR (EXPAR) model has been introduced in [3]. It is described by:

$$x_n = \sum_{i=1}^{p} [a_i + b_i \exp(-\alpha x_{n-1}^2)] x_{n-i} + \varepsilon_n$$

- The motivation was to define a continuous NAR model able to produce limit cycles such as observed in the Lynx data.

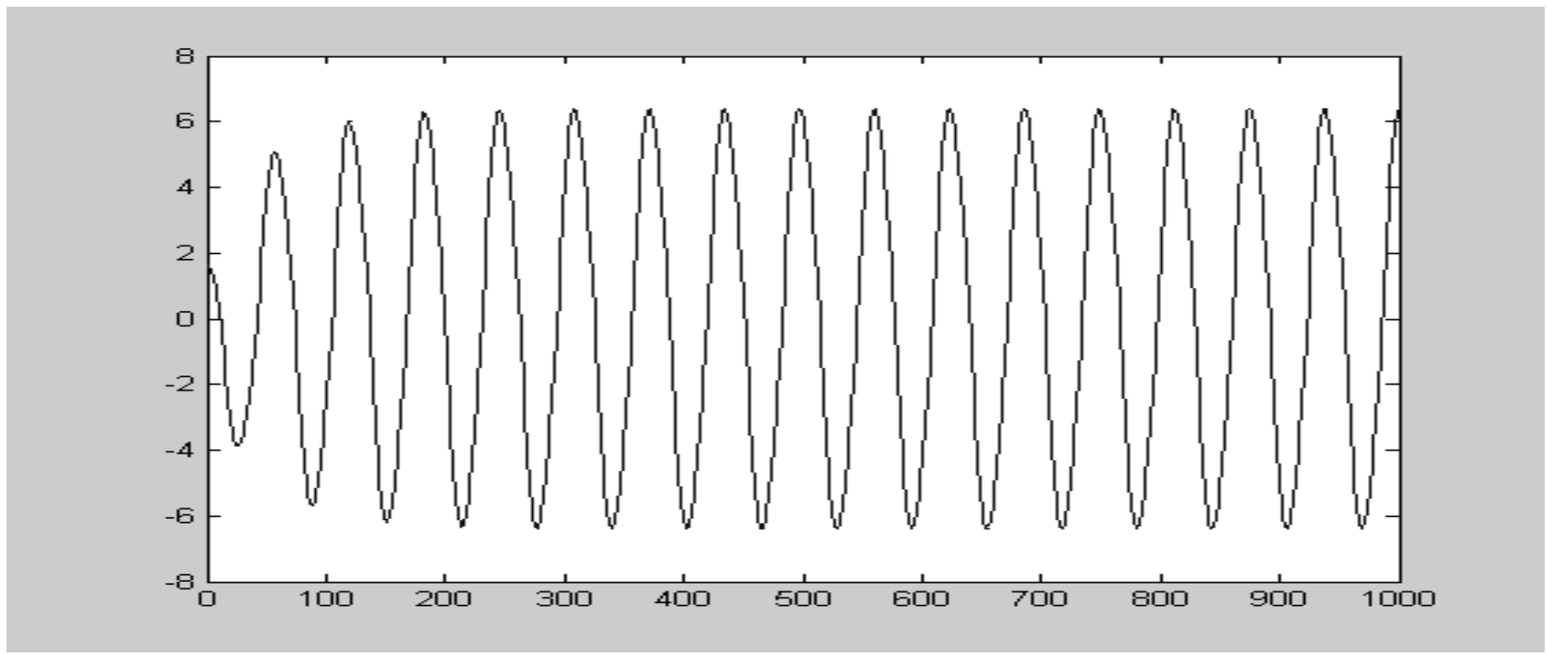
- It is shown in [3] that sufficient conditions for the existence of a limit cycle (no excitation) are:

1. All the poles of the AR model defined by the coefficients $a_1, \ldots, a_p$, i.e. for x_{n-1} very large, should be inside the unit circle.
2. Some of the poles of the AR model defined by the coefficients $a_1+b_1, \ldots, a_p+b_p$, i.e. for $x_{n-1}= 0$, should be outside the unit circle.
3. One of the two conditions below should hold:

$$(1-\Sigma a_i)/\Sigma b_i > 1 \quad \text{or} \quad (1-\Sigma a_i)/\Sigma b_i < 0$$

- Note that condition 1) is a stability condition of *drift back to the center*.
- Example:

$$x_n = [1.95 + 0.23\exp(-x_{n-1}^2)]x_{n-1} - [0.96 + 0.24\exp(-x_{n-1}^2)]x_{n-2}$$



- Least squares estimation of the coefficients $\{a_i,b_i\}$, once the scale parameter α is fixed, is quite easy.
- Concerning α, an optimization procedure can be used. Otherwise, an exhaustive grid search in a suitable range of values can be performed. Typically, $\exp(-\alpha x_{n-1}^2)$ should be in the interval [0.25 0.75] for all values of x_{n-1}, which gives bounds on α.

- As a matter of fact the EXPAR model can be seen as a special (simple) case of the state dependent AR model introduced by Priestley [4]. This model is described by:

$$x_n = a_0(\boldsymbol{x}_{n-1}) + a_1(\boldsymbol{x}_{n-1})x_{n-1} + \cdots + a_p(\boldsymbol{x}_{n-1})x_{n-p} + \varepsilon_n$$

which expresses nonlinearity as a dependence of the AR coefficents on the state vector $\boldsymbol{x}_{n\text{-}1} = [x_{n\text{-}1}, \ldots, x_{n\text{-}p}]^{\mathsf{T}}$.

- But one can observe that the state dependence in the EXPAR model can be seen as a very simple, mono-dimensional RBF network with only one center at 0.
- Also, if there is indeed a direct dependence of the AR coefficients on the state, then it could be well approximated by an RBF network, since it possesses the property of universal approximation.

- These remarks lead to the definition of the RBF-AR model, in which the state dependence of the AR coefficients is expressed by:

$$a_i(\boldsymbol{x}_{n-1}) = w_{i0} + \sum_{k=1}^{K} w_{ik} \exp\left(-\frac{\|\boldsymbol{x}_{n-1}-\boldsymbol{c}_k\|^2}{2\beta^2}\right)$$

- Note that, although it is not mandatory, one uses generally the same centers for all the coefficient expansions.

- The centers and width parameter β can be determined using heuristics as such as the ones presented for the RBF networks (it is also possible to have multiple width parameters).
- Once this is done, estimation of the coefficients $\{w_{ki}\}$ becomes a classical linear least squares problem.
- The MDL approach beased on center quantization can also be employed.

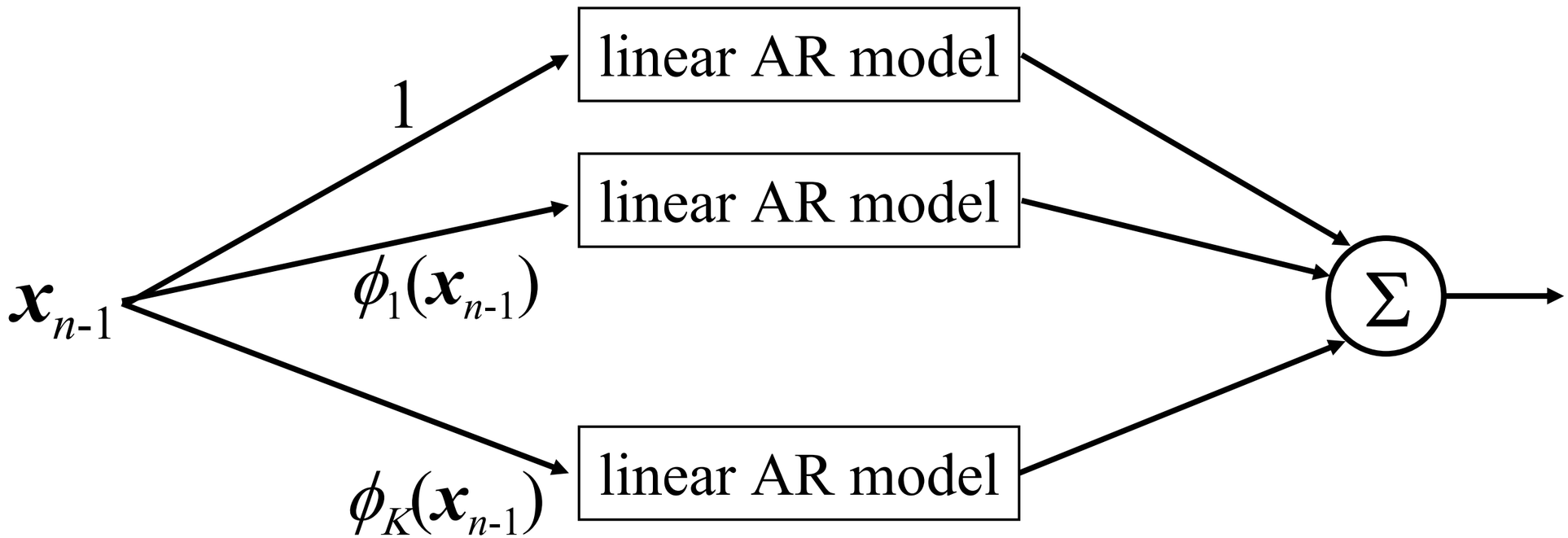
- But the terms in the RBF-AR model can also be regrouped differently:

$$x_n = w_{00} + \sum_{k=1}^{K} w_{0k}\phi_k(\boldsymbol{x}_{n-1}) + \left[w_{10} + \sum_{k=1}^{K} w_{1k}\phi_k(\boldsymbol{x}_{n-1}) \right] x_{n-1}$$

$$+ \cdots + \left[w_{p0} + \sum_{k=1}^{K} w_{pk}\phi(\boldsymbol{x}_{n-1}) \right] x_{n-p}$$

$$= w_{00} + \sum_{i=1}^{p} w_{i0} x_{n-i} + \sum_{k=1}^{K} \phi_k(\boldsymbol{x}_{n-1}) \left[w_{0k} + \sum_{i=1}^{p} w_{ik} x_{n-i} \right]$$

i.e., the RBF-AR model is the sum of a linear AR model plus a sum weighted by RBF of linear AR models.



- The RBF-AR model is thus a soft-transition version of the piecewise AR model. A particular sub-model will have more influence if is closer to the corresponding center.
- Due to the exponential decrease of the RBF with the norm of $\boldsymbol{x}_{n\text{-}1}$, stability of the RBF-AR model depends only on the "background" sub-model (with weight 1).

1. P. H. Franses and D. van Dijk, *Non-Linear Time Series Models in Empirical Finance*, Cambridge Univ. Press, 2000.
2. C. J. Pantaleon-Prieto, I. Santamaria, and A. Figueiras-Vidal, "Competitive local linear modeling," *Signal Processing*, vol. 49, 1996, pp. 73-83.
3. V. Haggan and T. Ozaki, "Modelling non-linear random vibrations using an amplitude- dependent autoregressive time series model," *Biometrika*, vol. 68, 1981, pp. 189-196.
4. M. B. Priestley, "Sate-dependent models: a general approach to non-linear time series analysis," *J. of Time Series Anal.,* vol. 1, 1980, pp. 47-70.
5. H. Kato and T. Ozaki, "Adding data process feedback to the nonlinear autoregressive model," *Signal Processing*, vol. 82, 2002, pp. 1189-1204.