Capabilities and Prospects of Inductive Modeling

Volodymyr STEPASHKO

Prof., Dr. Sci., Head of Department
INFORMATION TECHNOLOGIES
FOR INDUCTIVE MODELING
International Research and Training Centre
of the Academy of Sciences of Ukraine

Layout

1. Historical aspects of IM
2. International events on IM
3. Attempt to define IM: what is it?
4. IM destination: what is this for?
5. IM explanation: basic algorithms and tools
6. Basic Theoretical Results
7. IM compared to ANN and CI
8. Real-world applications of IM
9. Main centers of IM research
10. IM development prospects

1. Historical aspects of IM

1968 First publication on GMDH:

- Івахненко О.Г. Метод групового урахування аргументів – конкурент методу стохастичної апроксимації // Автоматика. – 1968. – № 3. – С. 58-72.

- *Terminology evolution*:

  - heuristic self-organization of models (1970s)
  - inductive method of model building (1980s)
  - inductive learning algorithms for modeling (1992)
  - inductive modeling (1998)

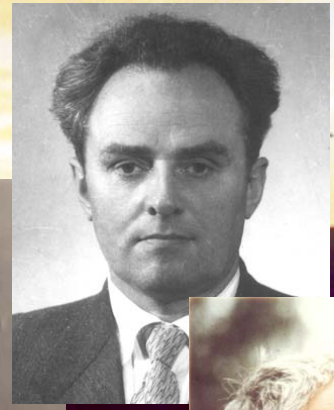
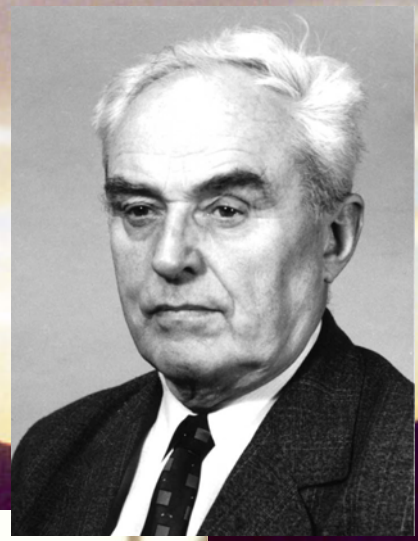
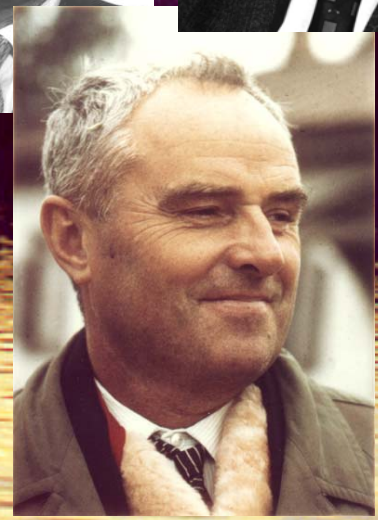
GMDH: Group Method of Data Handling

MGUA: Method of Group Using of Arguments

A.G.Ivakhnenko: GMDH originator





Main scientific results in inductive modelling theory:

- Foundations of cybernetic forecasting device construction
- Theory of models self-organization by experimental data
- Group method of data handling (GMDH) for automatic construction (self-organization) of model for complex systems
- Method of control with optimization of forecast
- Principles of noise-immunity modelling from noisy data
- Principles of polynomial networks construction
- Principle of neural networks construction with active neurons

Academician Ivakhnenko

- *Originator of the scientific school of inductive modelling*
- *Author of 44 monographs and numerous articls*
- *Prepared more than 200 Cand. Sci (Ph.D.) and 27 Doct. Sci*

2. International events on IM

- ***2002 Lviv, Ukraine***
  1st International Conference on Inductive Modelling ICIM'2002
- ***2005 Kyiv, Ukraine***
  1st International Workshop on Inductive Modelling IWIM'2002
- ***2007 Prague, Czech Republic***
  2nd International Workshop on Inductive Modelling IWIM'2007
- ***2008 Kyiv, Ukraine***
  2nd International Conference on Inductive Modelling ICIM'2008
- ***2009 Krynica, Poland***
  3rd International Workshop on Inductive Modelling IWIM'2009
- ***2010 Yevpatoria, Crimea, Ukraine***
  3rd International Conference on Inductive Modelling ICIM'2009
- *Zhukyn (near Kyiv, Ukraine)*
  Annual International Summer School on Inductive Modelling

3. Attempt to define IM: what is it?

- IM is MGUA / GMDH
- IM is a technique for model self-organization
- IM is a technology for building models from noisy data
- IM is the technology of inductive transition from data to models under uncertainty conditions:
  - small volume of noisy data
  - unknown character and level of noise
  - inexact composition of relevant arguments (factors)
  - unknown structure of relationships in an object

4. IM destination: what is this for?

IM is used for solving the following problems:

- ✓ Modelling from experimental data
- ✓ Forecasting of complex processes
- ✓ Structure and parametric identification
- ✓ Classification and pattern recognition
- ✓ Data clasterization
- ✓ Machine learning
- ✓ Data Mining
- ✓ Knowledge Discovery

5. IM explanation: algorithms and tools

Basic Principles of GMDH as an Inductive Method

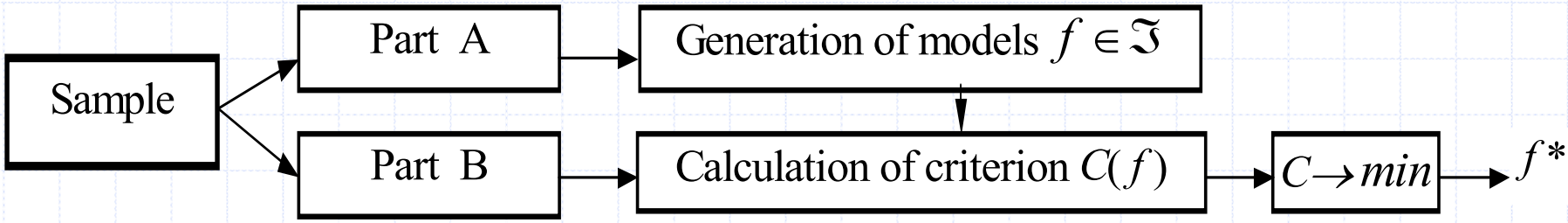
Given: data sample of n observations after m input $x_1, x_2, \ldots, x_m$ and output y variables

Find: model $y = f(x_1, x_2, \ldots, x_m, \theta)$ with *minimum variance of prediction error*

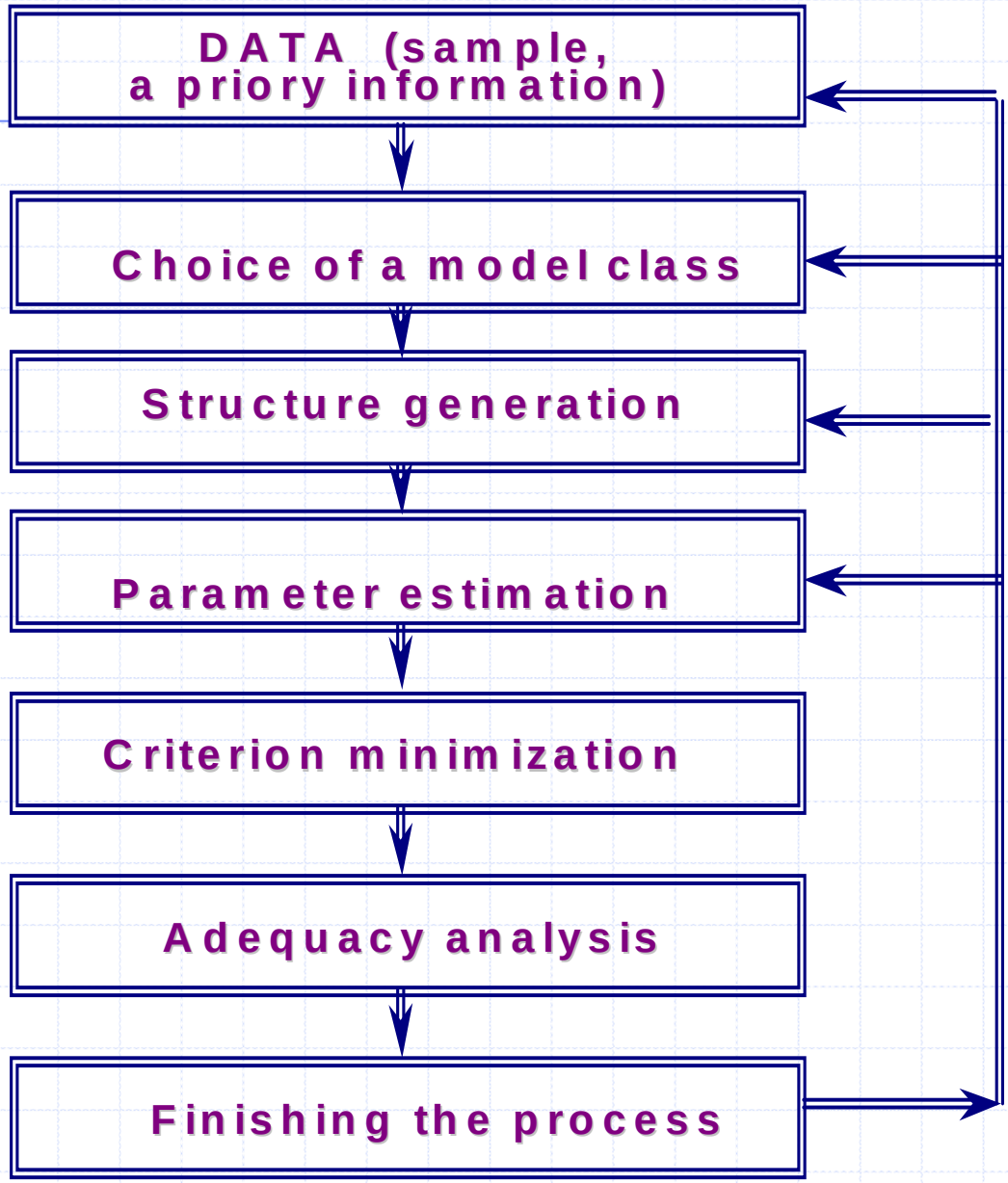
GMDH Task: $f^* = \arg\min_{f \in \Im} C(f)$, $C(f)$ – model quality criterion, $\Im$ – set of models

Basic principles of the GMDH as an ***inductive*** **method**:

1. ***generation of variants*** of the *gradually complicated structures* of models
2. ***successive selection*** of the best variants using the *freedom of decisions choice*
3. ***external addition*** (due to the *sample division*) as the selection criterion

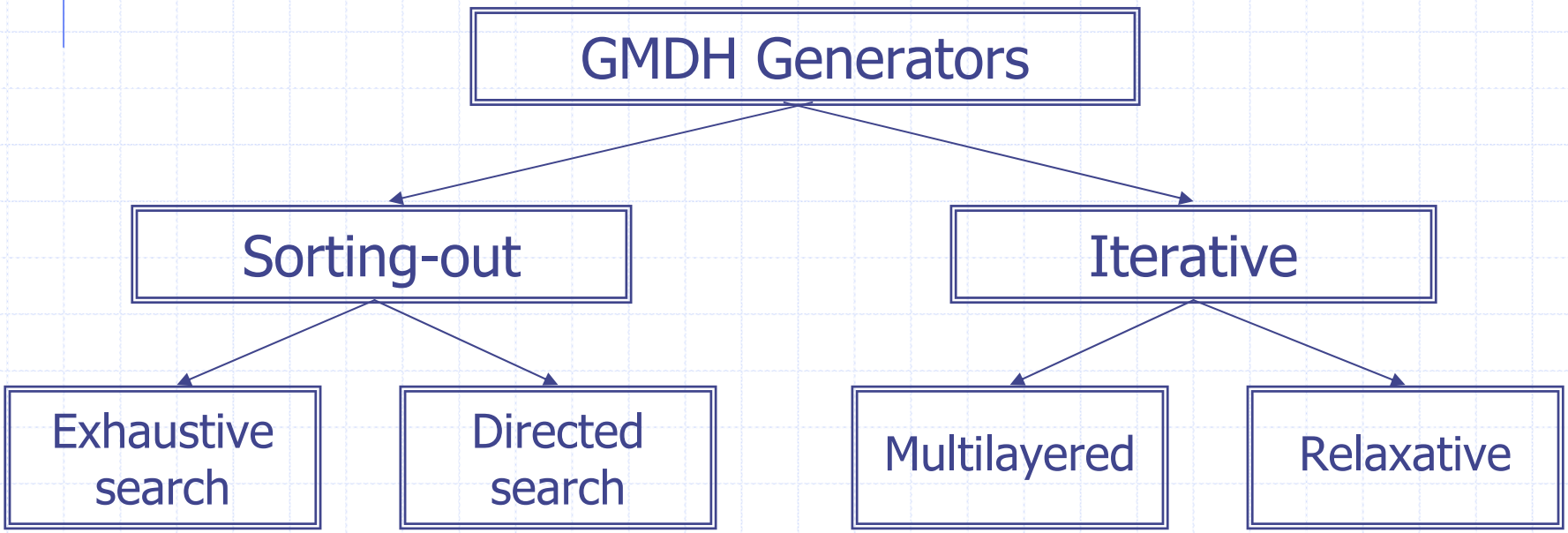


Main stages of the modeling process



GMDH features

- **Model Classes**: linear, polynomial, autoregressive, difference (dynamic), nonlinear of network type etc.
- **Parameter estimation**: Least Squares Method (LSM)
- **Model structure generators**:



Main generators of models structures

1. Combinatorial:

$$y_v = X_v\widehat{\theta}_v,\ v = 1,\ldots,2^m;\quad d = (d_1, d_2,\ldots,d_m)$$

2. Combinatorial-selective:

$$\widehat{y}_s^l = (X_{s-1}^i \mid x_s^j)\widehat{\theta}_s,\ s = \overline{1,m},\quad i,l = \overline{1,F_{s-1}}$$

3. Selective (multilayered iterative):

$$y_l^{r+1} = \vartheta_{1l}y_i^r + \vartheta_{2l}y_j^r + \vartheta_{3l}y_i^r y_j^r + \vartheta_4(y_i^r)^2 + \vartheta_5(y_j^r)^2,$$

$$r = 0,1,\ldots;\ i,j = \overline{1,F};\ l = \overline{1,C_F^2}$$

External Selection Criteria

Given sample: $W = (X \mid y)$, $X[n \times m]$, $y[n \times 1]$

Division into two subsamples:

$$W = \begin{bmatrix} W_A \\ W_B \end{bmatrix}; X = \begin{bmatrix} X_A \\ X_B \end{bmatrix}; y = \begin{bmatrix} y_A \\ y_B \end{bmatrix}; n_A + n_B = n$$

Parameter estimation for a model $y = X\theta$:

$$\widehat{\theta}_G = (X_G^T X_G)^{-1} X_G^T y_G, \quad G = A, B, W,$$

Regularity **criterion:** $$AR_B = \left\| y_B - X_B \widehat{\theta}_A \right\|^2$$

Unbiasedness **criterion:** $$CB = \left\| X_W \widehat{\theta}_A - X_W \widehat{\theta}_B \right\|^2$$

IM tools

- Information Technology ASTRID (Kyiv)
- KnowledgeMiner (Frank Lemke, Berlin)
- FAKE GAME (Pavel Kordik at al., Prague)
- GMDHshell (Oleksiy Koshulko, Kyiv)

6. Basic Theoretical Results

$$f^* = \arg\min_{f \in F} C(f).$$

F – set of model structures

C – criterion of a model quality

Structure of a model:

$$\hat{y}_f = f(X, \hat{\theta}_f)$$

Estimation of parameters:

$$\hat{\theta}_f = \arg\min_{\theta_f \in R^m} Q(\theta_f).$$

Q – criterion of the quality of model parameters estimation

- **Main concept**:

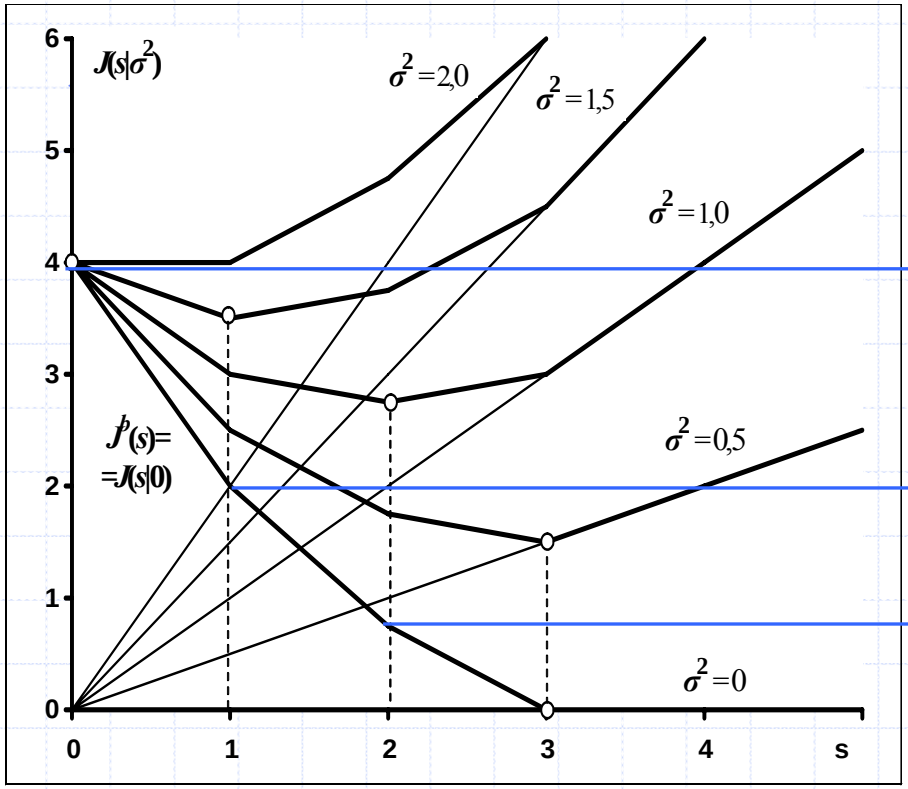
Self-organizing evolution of the model of ***optimal complexity*** under uncertainty conditions

- **Main result:**

Complexity of the optimum forecasting model depends on the level of uncertainty in the data: the higher it is, the simpler (more robust) there must be the optimum model

- **Main conclusion**:

GMDH is the method for construction of models with ***minimum variance*** of forecasting error



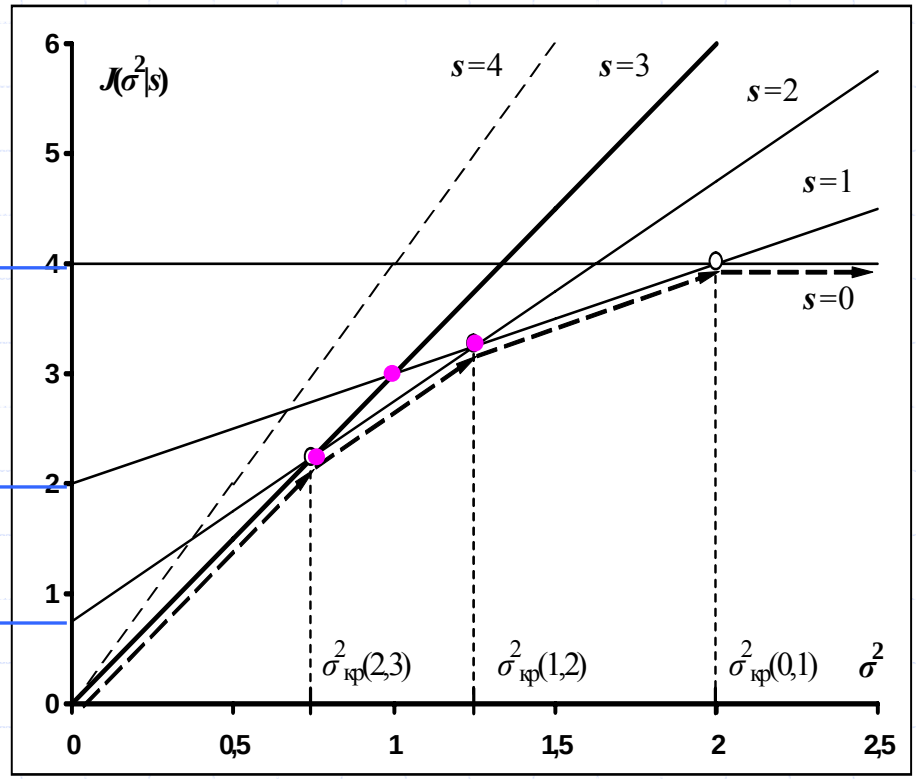
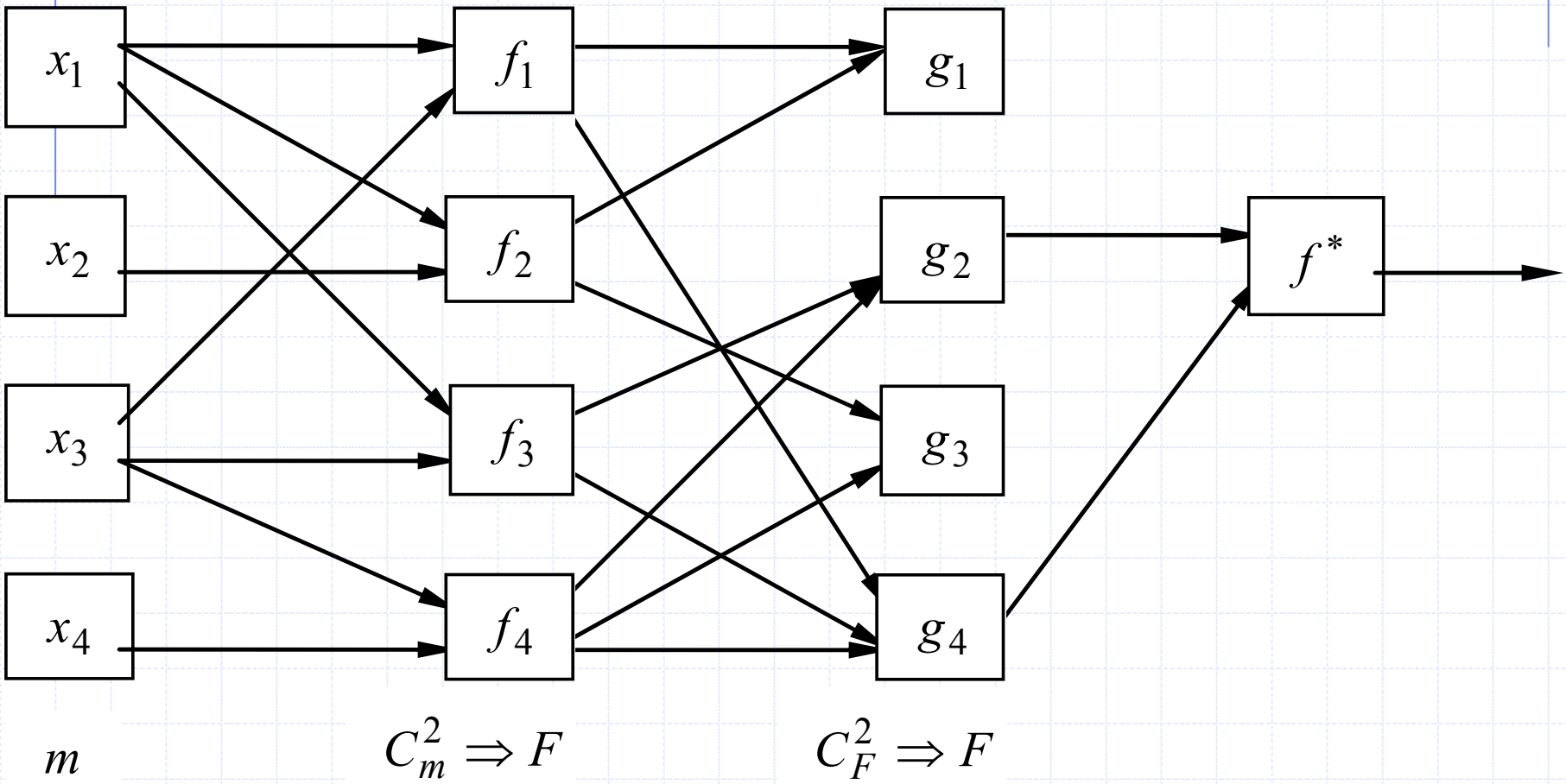


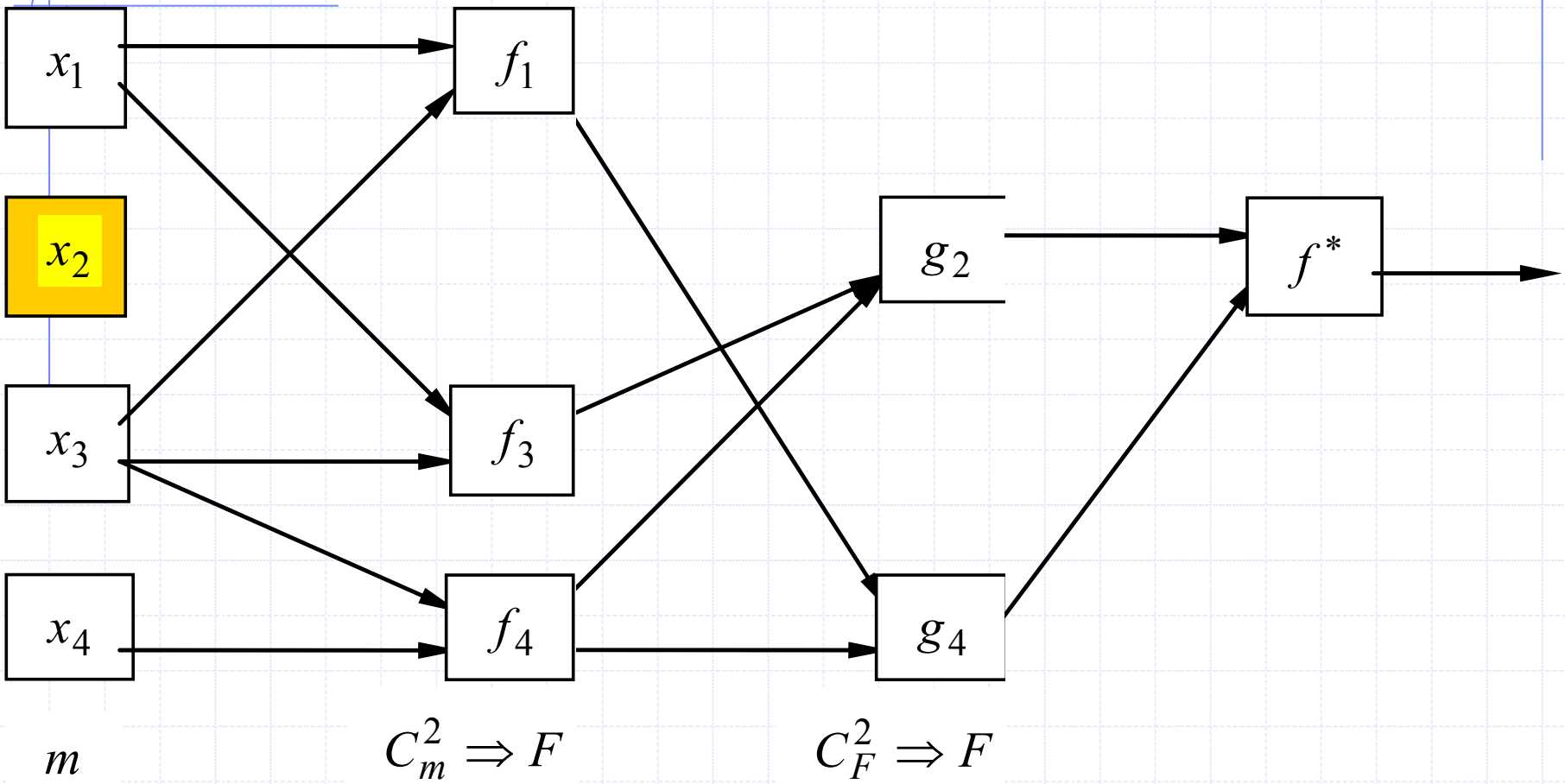
Illustration to the GMDH theory

7. IM compared to ANN and CI

Selective (multilayered) GMDH algorithm:



Optimal structure of the multilayered net



8. Real-world applications of IM

1. Prediction of tax revenues and inflation
2. Modelling of ecological processes
   - activity of microorganisms in soil under influence of heavy metals
   - irrigation of trees by processed wastewaters
   - water ecology
3. System prediction of power indicators
4. Integral evaluation of the state of the complex multidimensional systems
   - economic safety
   - investment activity
   - ecological state of water reservoirs
   - power safety
5. Technology of informative-analytical support of operative management decisions

9. Main centers of IM research

- IRTC ITS NANU, Kyiv, Ukraine
- NTUU “KPI”, Kyiv, Ukraine
- KnowledgeMiner, Berlin, Germany
- CTU in Prague, Czech
- Sichuan University, Chengdu, China

10. IM development prospects

The most promising directions:

1. Theoretical investigations
2. Integration of best developments of IM, NN and CI
3. Paralleling
4. Preprocessing
5. Ensembling
6. Intellectual interface
7. Case studes

THANK YOU!

Volodymyr STEPASHKO

Address: Prof. Volodymyr Stepashko,
International Centre of ITS,
Akademik Glushkov Prospekt 40,
Kyiv, MSP, 03680, Ukraine.

Phone: +38 (044) 526-30-28
Fax: +38 (044) 526-15-70
E-mail: stepashko@irtc.org.ua
Web: www.mgua.irtc.org.ua