APPLIED TECHNIQUES OF INTERVAL ANALYSIS FOR ESTIMATION OF EXPERIMENTAL DATA

S. I. Kumkov

Institute of Mathematics and Mechanics
UrB RAS, Ekaterinburg, Russia, kumkov@imm.uran.ru
Ural Federal University, Ekaterinburg, Russia

16th GAMM-IMACS International Symposium
on Scientific Computing, Computer Arithmetic,
and Validated Numerics

SCAN-2014, September 21 – 26, Wurzburg, Germany

The aim of this presentation is to demonstrate one interesting practical problem of estimation of experimental process parameters under uncertainty conditions when components of the parameter vector can be only estimated on the basis of the Interval Analysis approach and available a priori data

Topics of presentation

Experimental process and its model.

Measured information and its uncertainty.

Problem formulation and how to solve it ?

Interval approach and its peculiarities.

Computation results.

Conclusions.

References.

Experimental process

Description of a molten salt conductivity

vs the temperature [1, 2]:

$$\sigma(T, A, B, C, D) = A\text{exp}\Big(-B/T + C/(T * VM(T)) + D/VM(T)\Big),$$

where T is the temperature, the main argument; $VM(T)$ is an auxiliary dependence given tabulated for each value of the temperature; $A > 0$, $B > 0$, $C > 0$, and $D > 0$ are constant parameters to be estimated. The parameters reflect influence (on conductivity) of various internal properties of the melt.

Measured information and its uncertainty

Results of the experiment are presented as the following collection (sample) of conductivity measurements:

$$\{T_n, \sigma_n\}, n = \overline{1, N},$$

where the temperature values T_n are assumed to be known exactly, but the conductivity σ is measured with essential error bounded by the value e_{max}.

The sample is dramatically short: $N \approx 5 \sim 7$ measurements only.

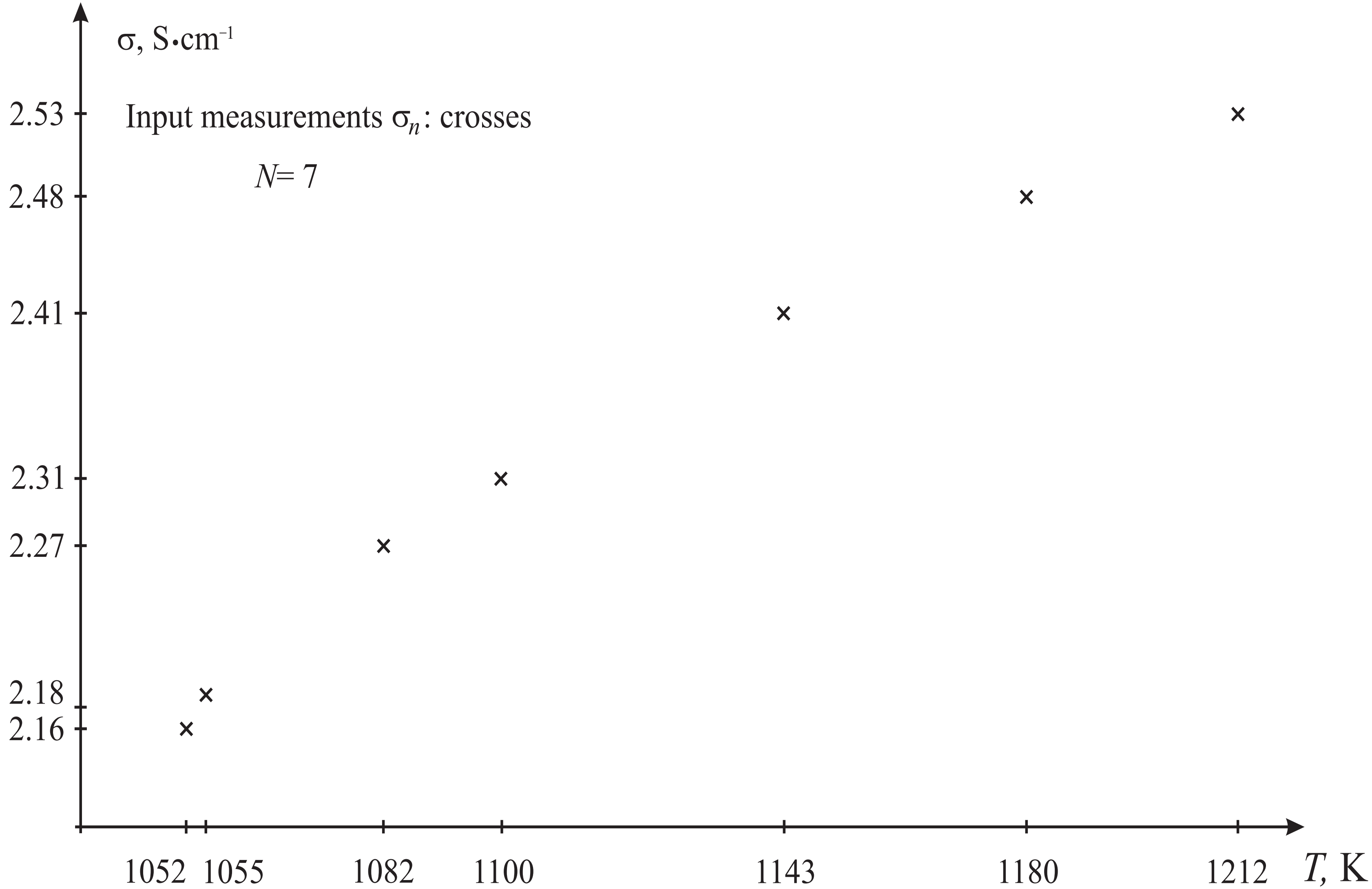
No probabilistic information on errors is known.

From theoretical estimations or previous experience, usually, the following *a priori* constraints on possible values of the coefficients could be known:

$$[A^{\text{apr}}_{\text{min}}, A^{\text{apr}_{\text{max}}}], [B^{\text{apr}}_{\text{min}}, B^{\text{apr}_{\text{max}}}], [C^{\text{apr}}_{\text{min}}, C^{\text{apr}_{\text{max}}}], [D^{\text{apr}}_{\text{min}}, D^{\text{apr}_{\text{max}}}].$$

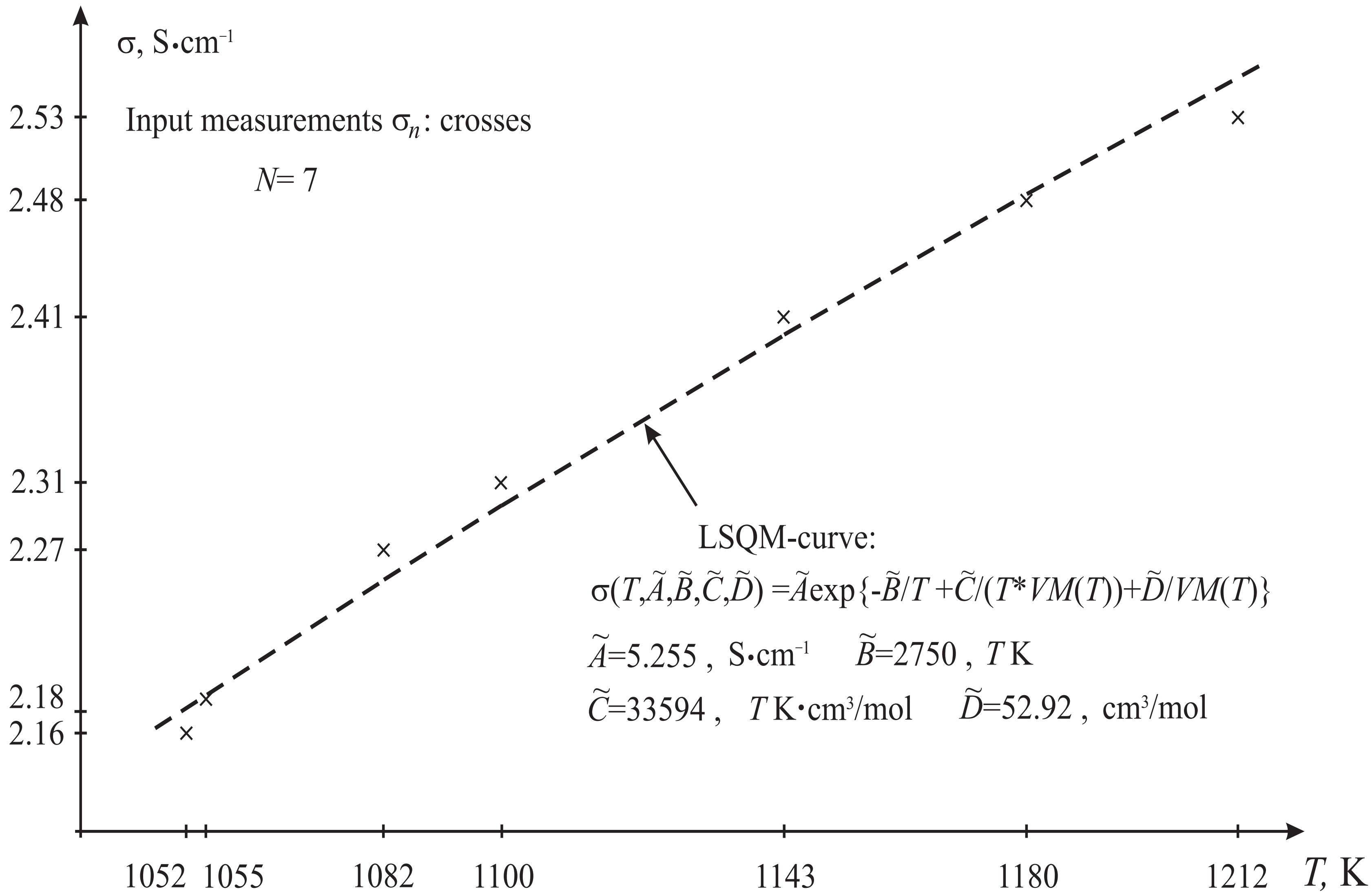
Measured information

The given experimental sample of the potassium-chlorine (K-Cl) melt [1] measurements has the following form.



Formal application of the LSQM-approximation [3 – 5]

The LSQM-curve and pointwise estimation $\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$, and $\widetilde{D}$, and their practically meaningless "cloud-built" intervals are available.



Problem formulation and how to solve it?

Since of very short length of the measurements sample, absence of probabilistic characteristics of the errors, and measurements uncertainty, it is impossible to use (with any good reasoning) the standard statistical methods [3–5]).

It is necessary:
on the basis of the Interval Analysis methods to built the set of admissible values (Informational Set, or the Set-membership) of coefficients A, B, C, D, *consistent* with the described data.

Interval approach and its essence

Ideas and methods of the Interval Analysis Theory and Applications arose from the fundamental, pioneer work by L.V. Kantorovich [6]. Nowadays, very effective developments of the theory and computational methods were created by many researchers both abroad [7-9] and in Russia [10–13].

Special interval algorithms have been elaborated for estimating parameters of experimental chemical processes [14-18].

Remind that essence of this branch of numerical methods theory and application consists in **estimation (or identification) of parameters under bounded errors, noises, or perturbations in the input information to be processed, under total absence of probabilistic characteristics of errors.**

The main definitions

Uncertainty set of each measurement (USM). It is the interval of values of measured process consistent with the measurement and the error bound:

$$H_n = [\underline{h_n}, \overline{h_n}]: \ \underline{h_n} = \sigma_n - e_{\text{max}}, \ \overline{h_n} = \sigma_n + e_{\text{max}}.$$

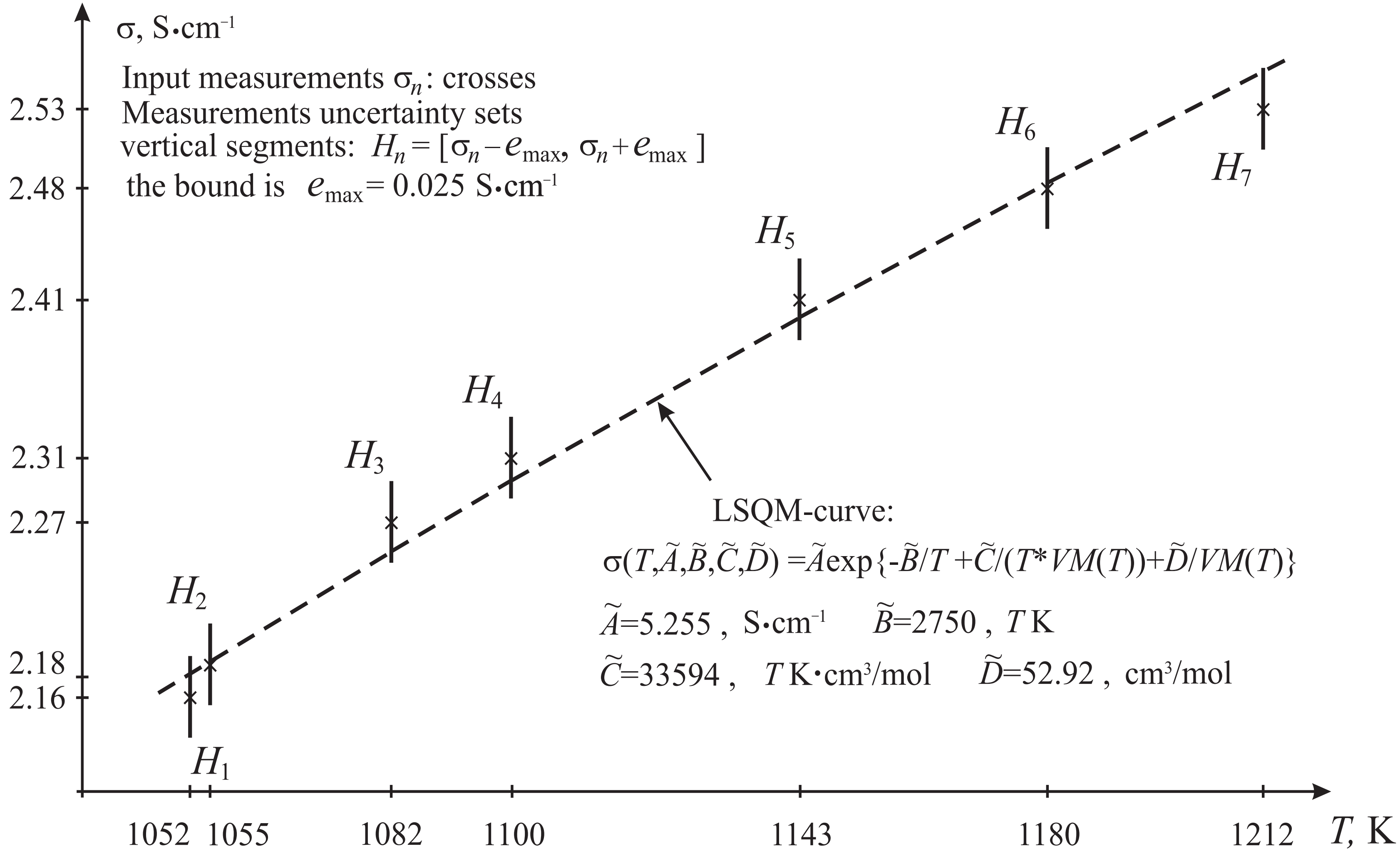
Admissible value of the parameters vector and corresponding admissible curve:

$$(A, B, C, D): \ \sigma(T_n, A, B, C, D) \in H_n, \ \text{for all } n = \overline{1, N}.$$

Informational Set (InfSet) is a totality of admissible values of the parameters vector

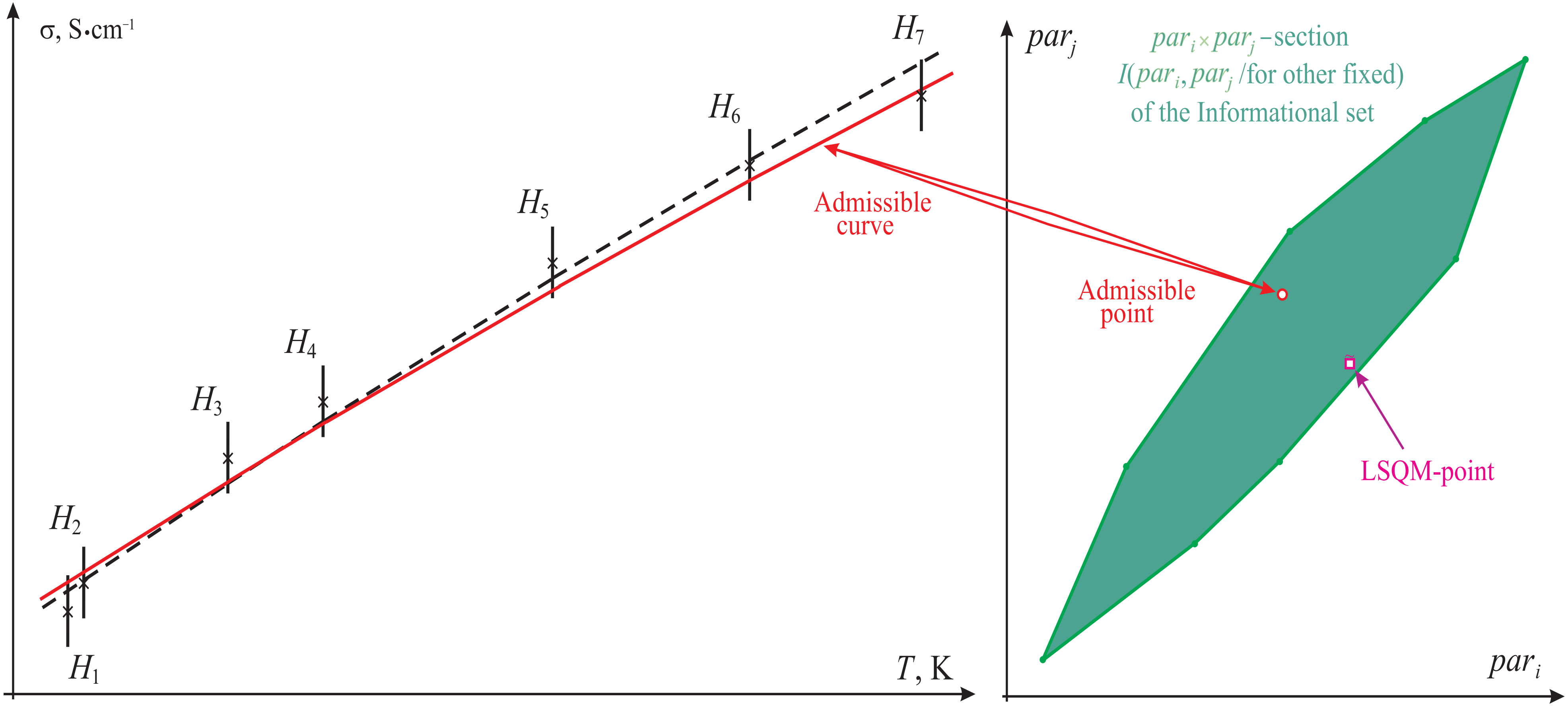
$$I(A, B, C, D) = \{(A, B, C, D): \ \sigma(T_n, A, B, C, D) \in H_n, \ \text{for all } n = \overline{1, N}\}.$$

Measurements and their uncertainty sets (USM)



Remark: if the actual level of errors in the sample is lower than *a priori* bound $e_{\max}$, the LSQM-curve and the point-estimate ($\widetilde{A}$, $\widetilde{B}$, $\widetilde{C}$, $\widetilde{D}$) could be *admissible*.

Illustration to the main definitions



Applied procedures. Transformed mathematical model

Usually, by standard passage to the natural logarithmic scale [1,2], researchers obtain a simple linear dependency on the parameters $\alpha = \log(A)$, B, C, and D:

$$\log(\sigma(T, \alpha, B, C, D)) = \alpha - B/T + C/(T * VM(T)) + D/VM(T),$$

After the passage, in contrast to standard statistical approaches, the problem formulated is reduced to solving the following system of the interval inequalities:

$$I(\alpha, B, C, D) = \{(\alpha, B, C, D) : \log(\sigma(T_n, \alpha, B, C, D)) \in H_n, \text{ for all } n = \overline{1, N}\}.$$

under *a priori* given additional constraints

$$[A^{\text{apr}}_{\text{min}}, A^{\text{apr}_{\text{max}}}],\ [B^{\text{apr}}_{\text{min}}, B^{\text{apr}_{\text{max}}}],\ [C^{\text{apr}}_{\text{min}}, C^{\text{apr}_{\text{max}}}],\ [D^{\text{apr}}_{\text{min}}, D^{\text{apr}_{\text{max}}}].$$

Estimation of parameters A and B is of the most interest. So, a two dimensional grid in parameters C and D is introduced on their given *a priori* intervals

$[C_{\text{min}}^{\text{apr}}, C_{\text{max}}^{\text{apr}}]$ and $[D_{\text{min}}^{\text{apr}}, D_{\text{max}}^{\text{apr}}]$:

$\{C_k, D_m\}$, $k = \overline{1, K}$, $m = \overline{1, M}$,

with practically acceptable small steps Δ_C and Δ_D.

Applied procedures.
Further transformation of the description dependence

Resolving the transformed dependency w.r.t. parameters A and B and, shifting to the right terms with parameters C_k and D_m, we obtain:

$$\alpha - B/T = \log\Big(\sigma(T_n)\Big) - C_k/(T * VM(T)) - D_m/VM(T),$$

or, shortly, for the whole sample

$$\{\alpha - B/T_n \in W(T_n, H_n, C_k, D_m)\}, \quad n = \overline{1, N}, \quad k = \overline{1, K}, \quad m = \overline{1, M}. \text{ (*)}$$

Note that the right sides are intervals.

This form allows one to build constructively a collection of InfSets

$$\{I(\alpha, B, C_k, D_m)\}, \ k = \overline{1, K}, \quad m = \overline{1, M}. \text{ (**)}$$

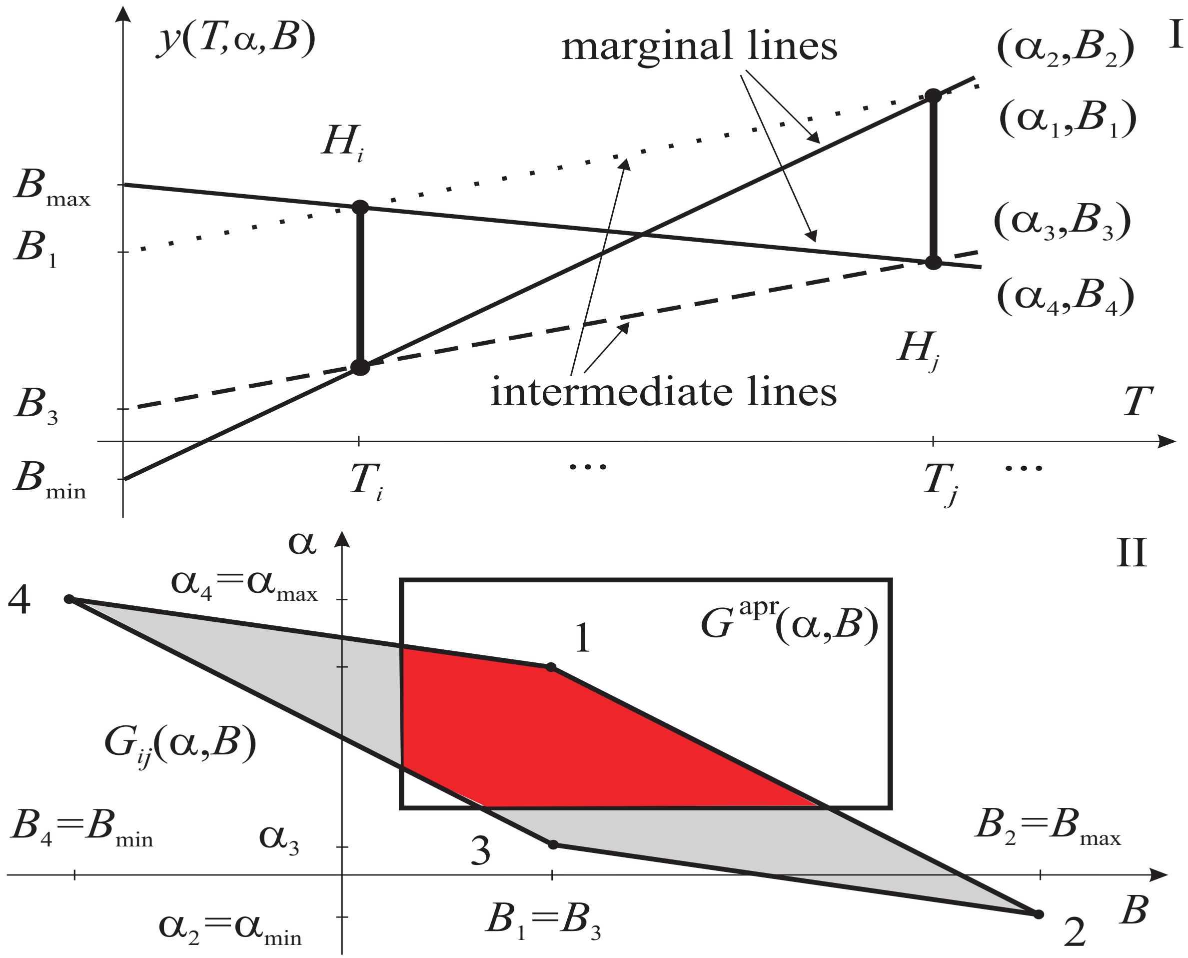
Applied procedures. Direct set-estimation approach

There are several approaches to solve system (*) of the interval linear inequalities
– classic linear programming methods [6, and many others],
– parallelotopes Walter [7], Hansen [8], Fiedler M., *et al* [9], Shary [10],
– by the "stripes" method Shary,Sharaya [11], Sharaya [12], Zhilin [13].

More convenient and faster grid-set method has been elaborated (see, Kumkov and with co-authors [14–18]) that gives *exact* estimations of the Informational sets (**) on part of parameters for each node of the grid on other parameters

Underline that we obtain *exact* estimation of each section
$I(\alpha, B, C_k, D_m)$ of the InfSet $I(\alpha, B, C, D)$ in contrast, for example,
to outer approximation of informational sets
in the parallelotope approaches.

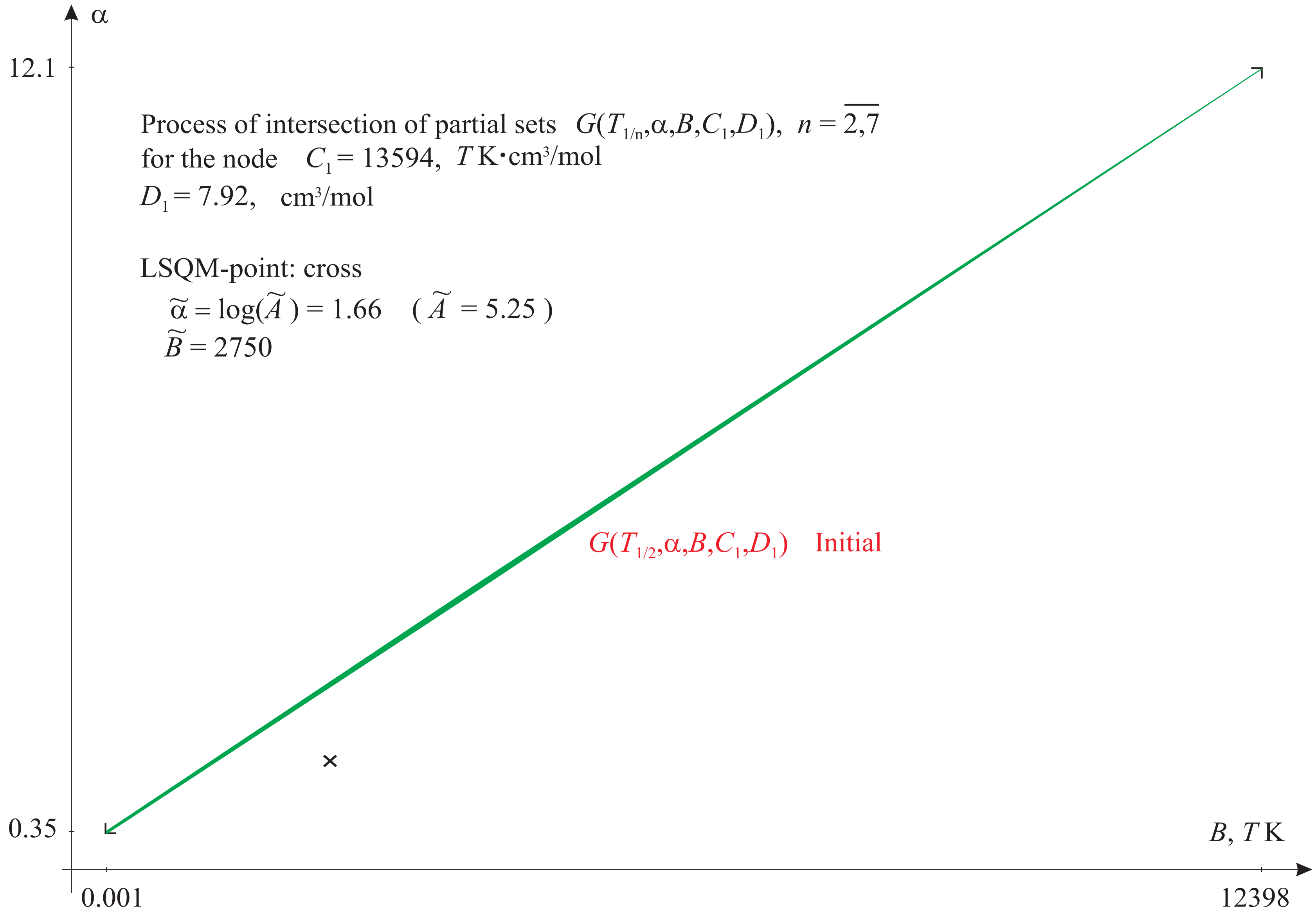
Ideology of the partial informational sets $G_{ij}(\alpha, B)$)



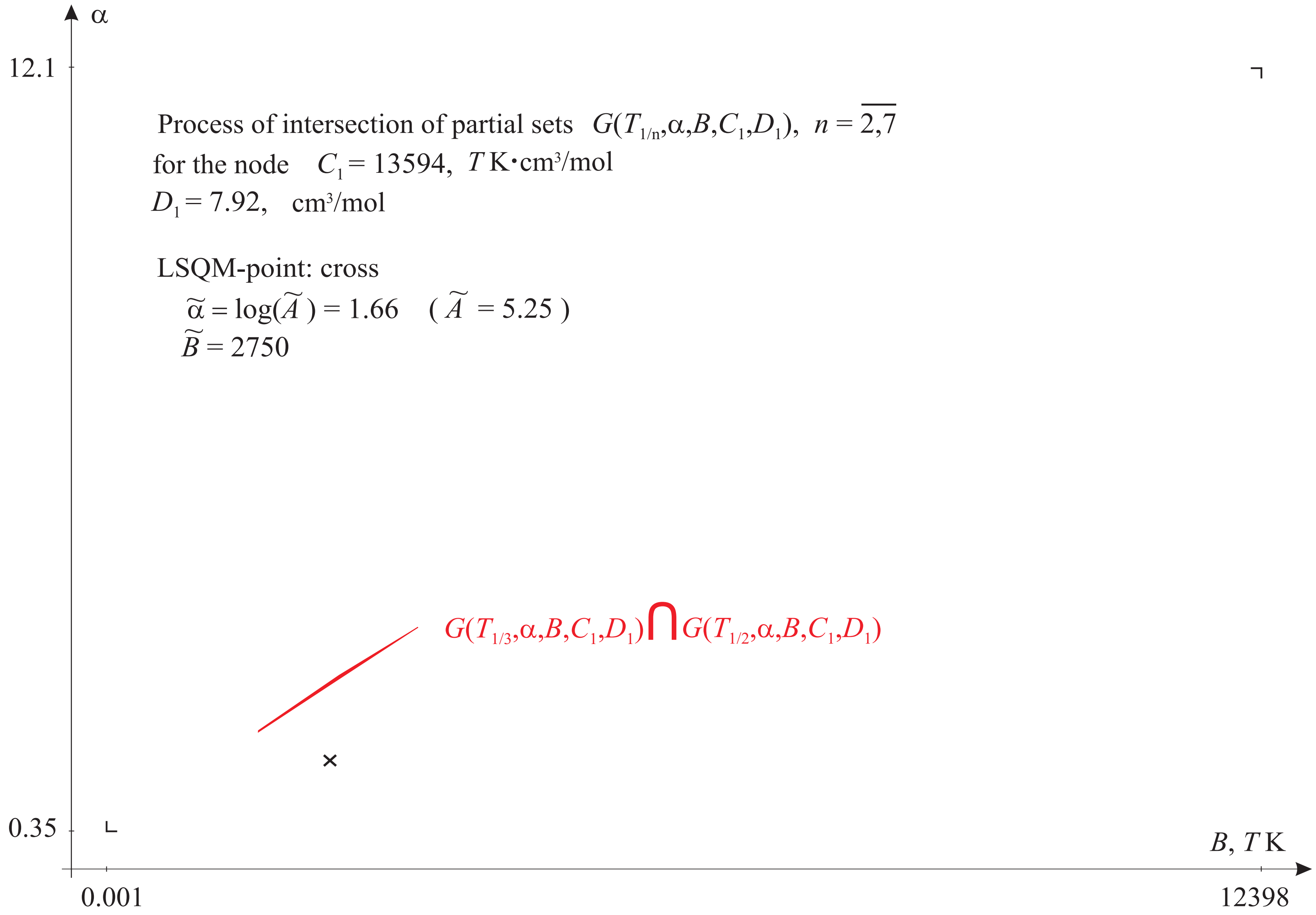
The main intersection procedure: $I(\alpha, B, C_k, D_m) = \cap_{i=1,j=2,N} G_{ij}(\alpha, B, C_k, D_m)$

Remark: Coordinates of apices 1–4 are calculated immediately. Additionally, if given, the *a priory* data could be directly taken into account.

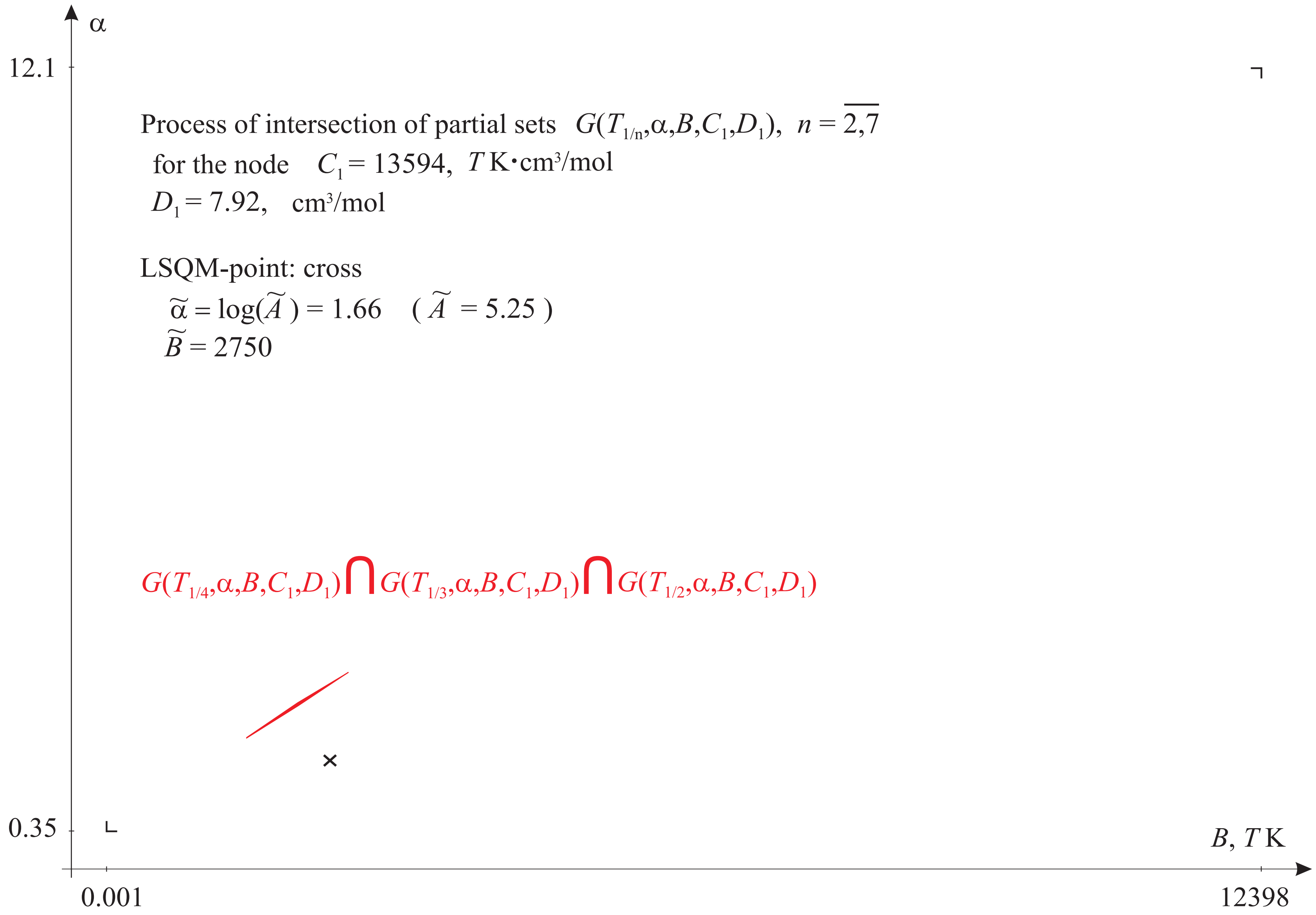
Computation results. Intersection of partial InfSets



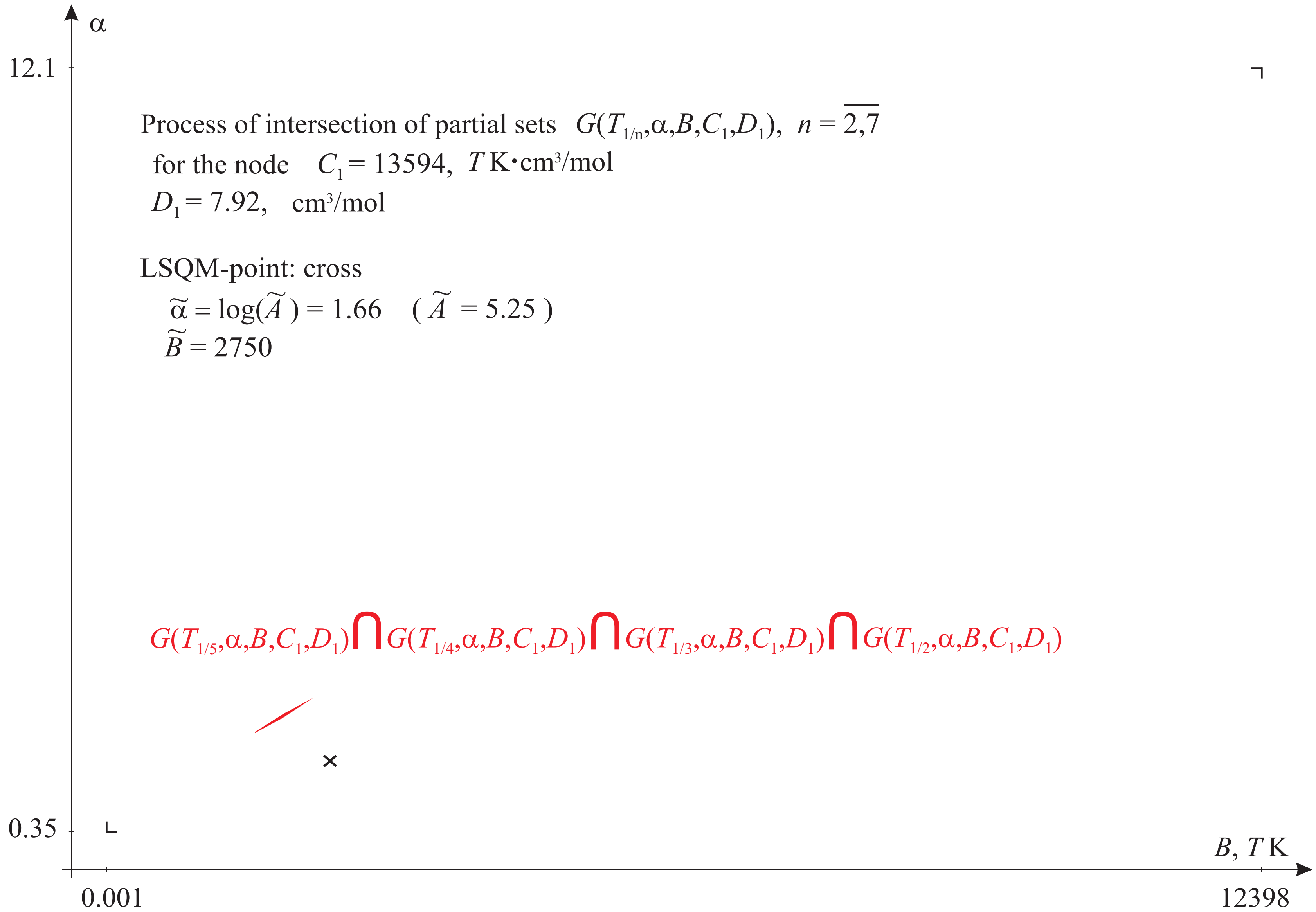
Computation results. Intersection of partial InfSets



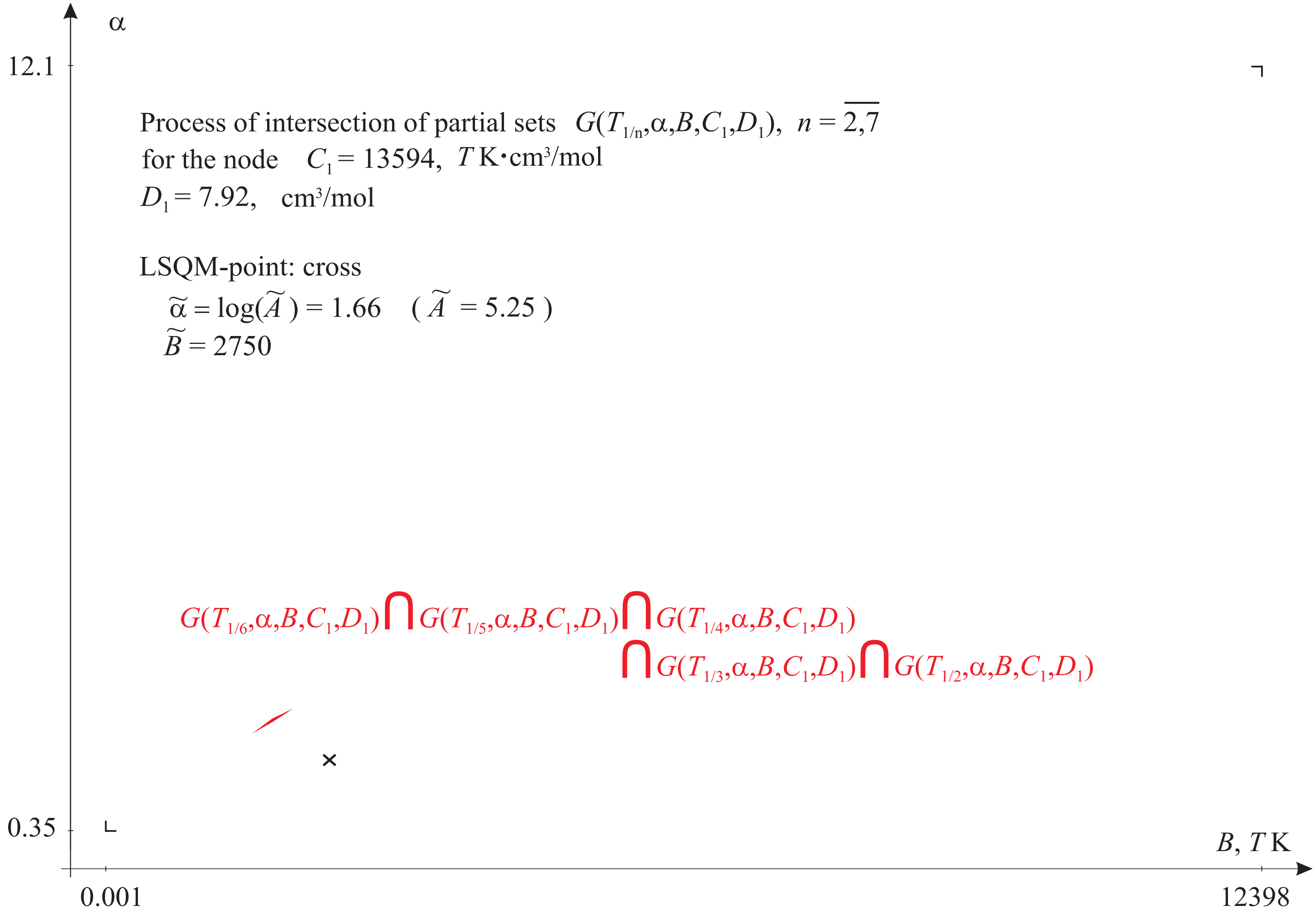
Computation results. Intersection of partial InfSets



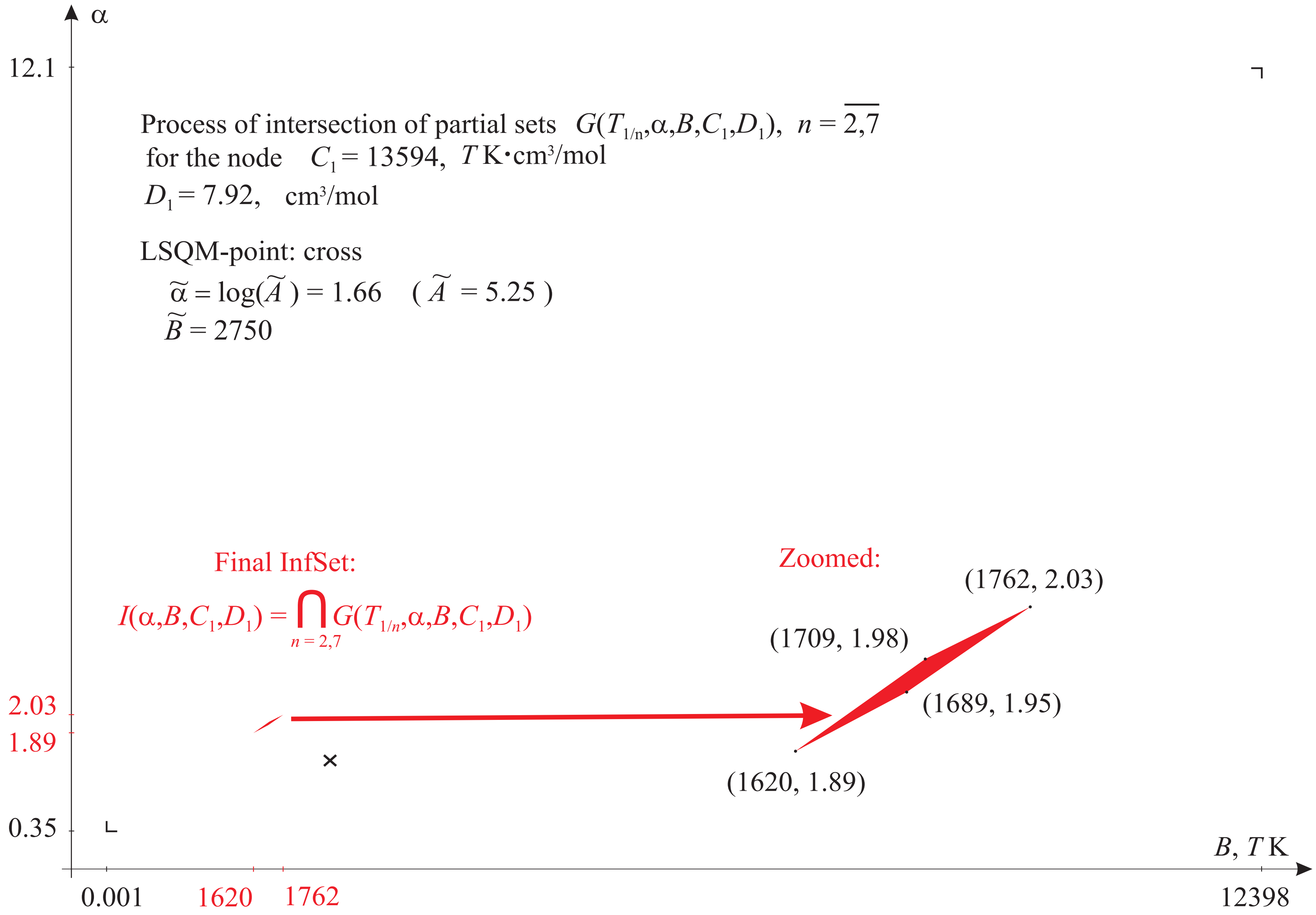
Computation results. Intersection of partial InfSets



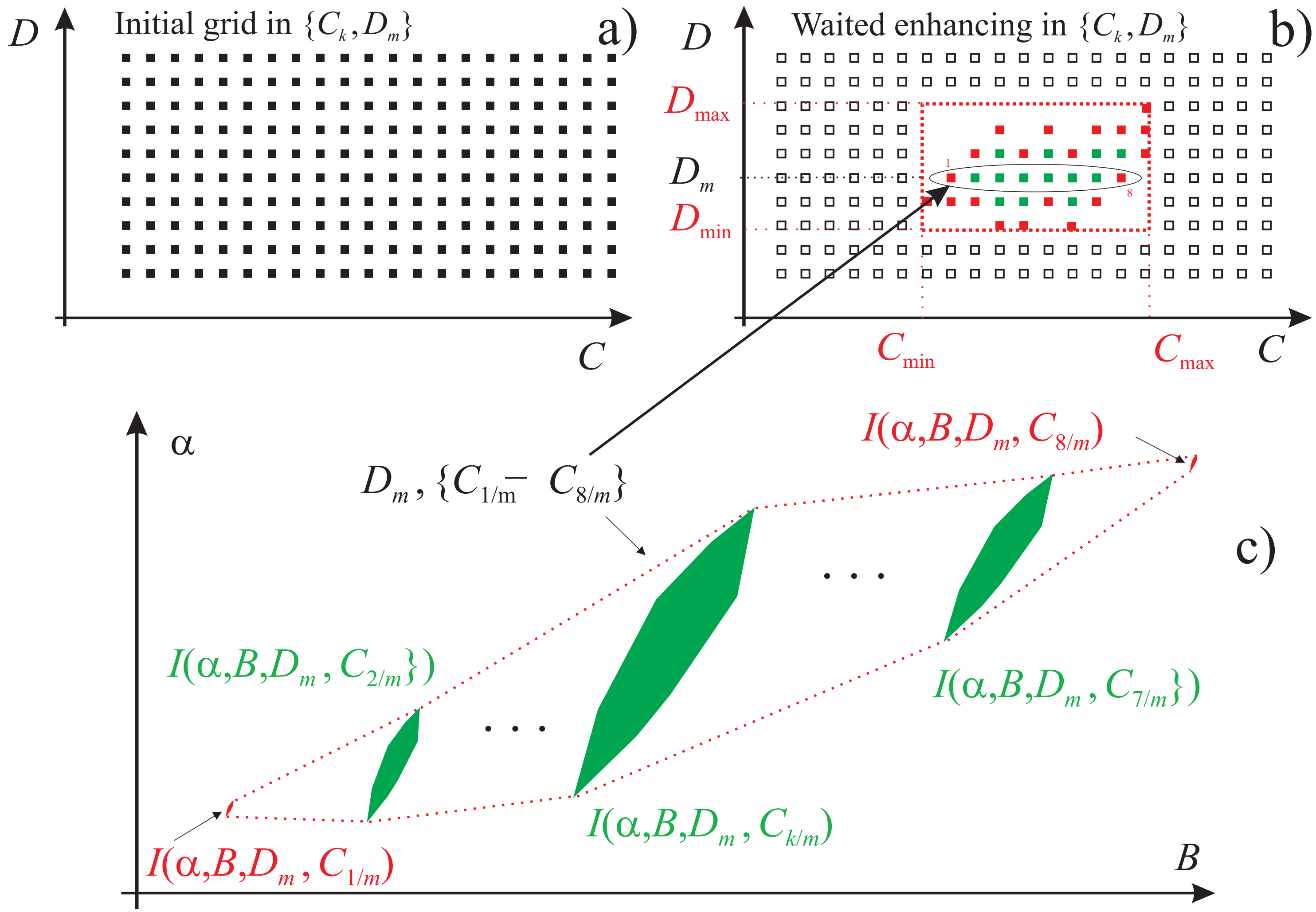
Computation results. Intersection of partial InfSets



Computation results. Intersection of partial InfSets

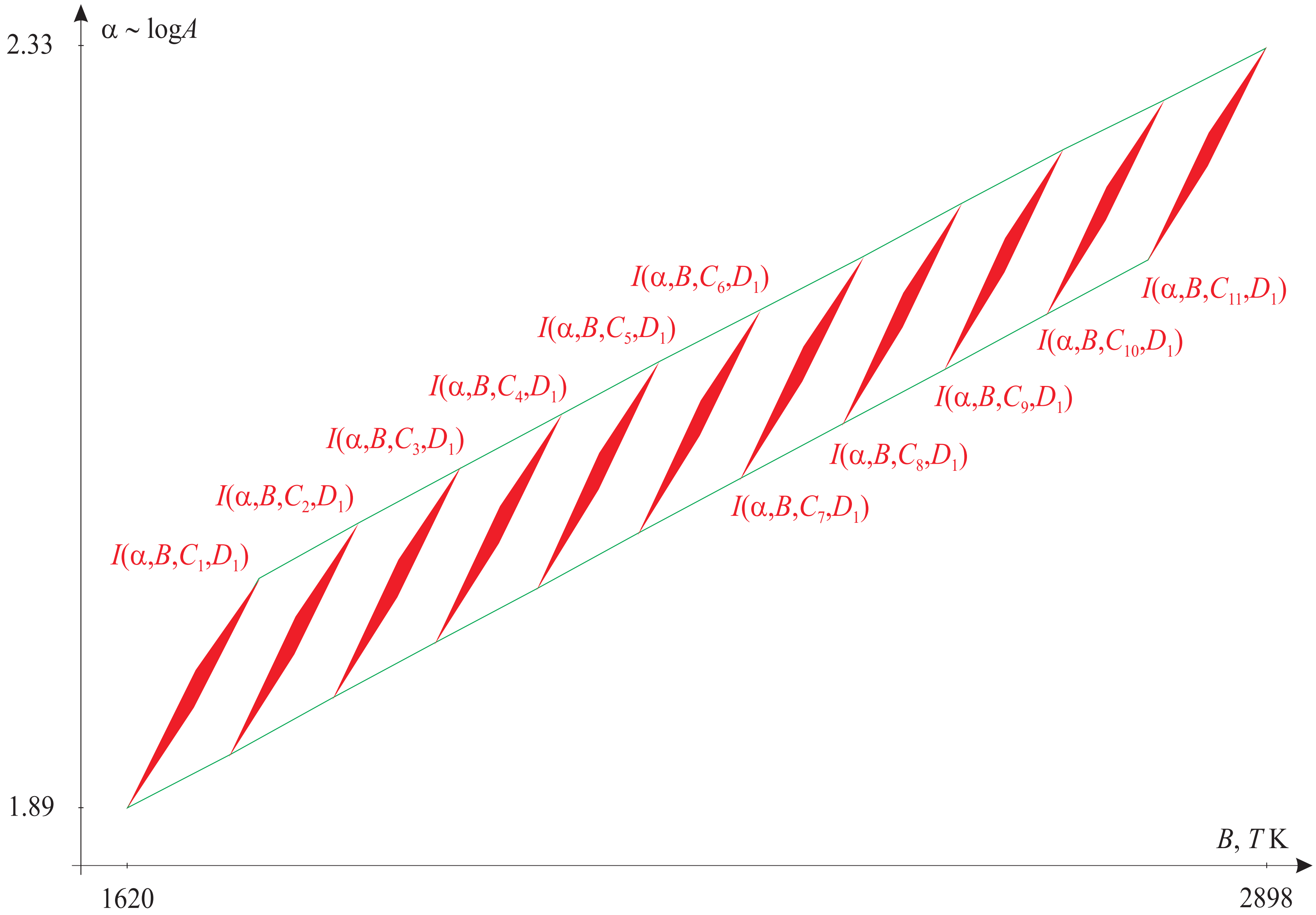


What do we usually obtain by such a hybrid "grid-set" approach?

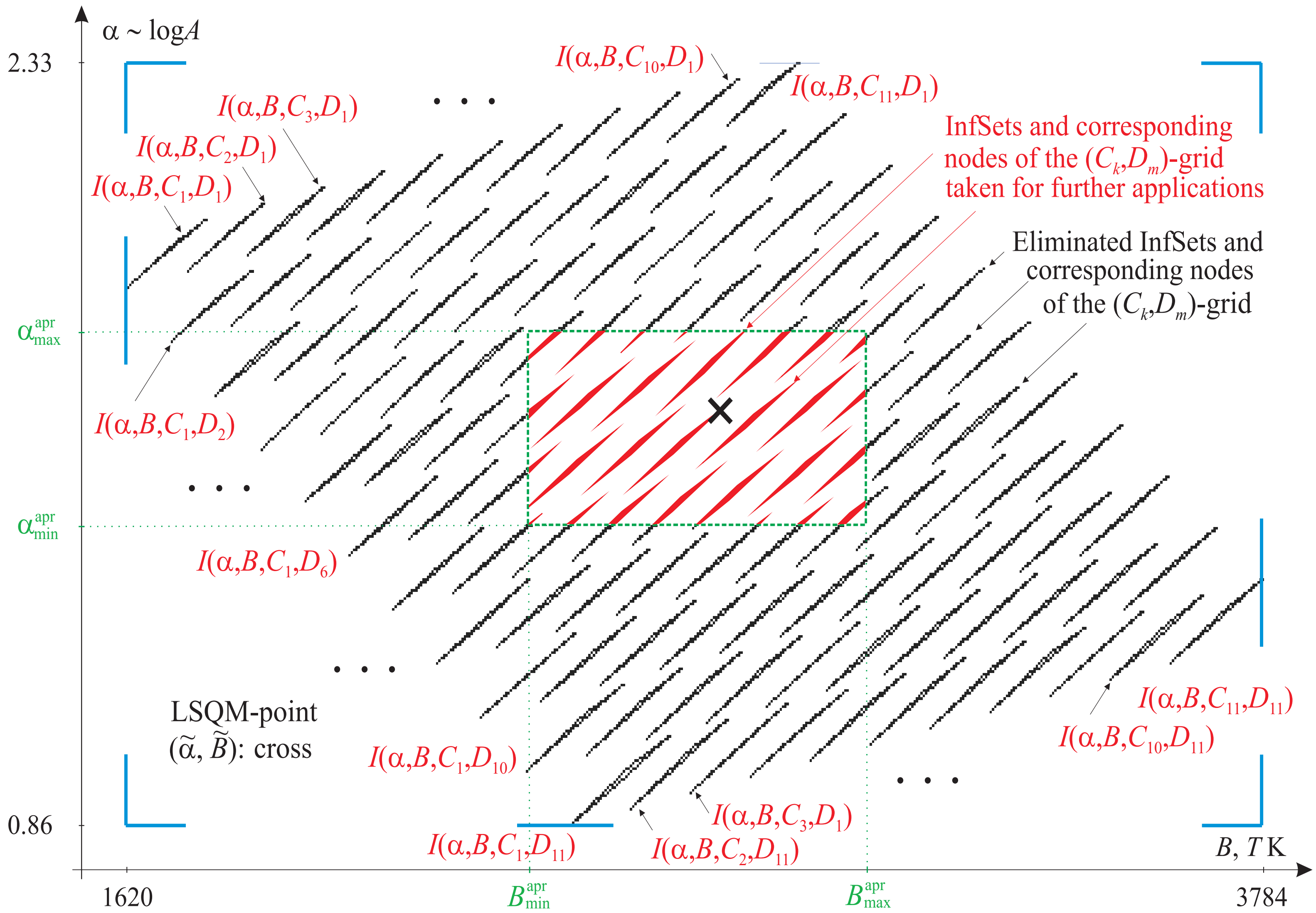


Remark: The result is simultaneous estimation of sets of ALL parameters. But in our case, it is not so.

In our example, collection of $\{I(\alpha, B, C_k, D_1)\}$



Results of constructing the InfSets and taking into account a priori data in parameters A and B



For practical application: procedure of direct finding verified inner estimations of intervals' parameters

1) Calculation of only one InfSet $I(\alpha, B, C, D)$ for *admissible internal point*, in our case, it is the LSQM-point $(\widetilde{C},\ \widetilde{D})$.

2) Calculation the *central* point $\alpha_{\text{cntr}},\ B_{\text{cntr}}$ of the set $I(\alpha, B, \widetilde{C}, \widetilde{D})$.

3) For the fixed point $(\alpha_{\text{cntr}},\ B_{\text{cntr}}, \widetilde{D})$, one implements variation ("up" and "down" from the value $\widetilde{C}$) of the parameter C and determines the corresponding marginal values C_{min} and C_{max} of the verified inner interval $[C_{\text{min}}, C_{\text{max}}]$. For finding these admissible marginal points, direct check of admissibility of the corresponding curve $\sigma(T, \alpha_{\text{cntr}}, B_{\text{cntr}}, C_{\text{var}}, \widetilde{D})$ is performed by the intervals $\{H_n\}$.

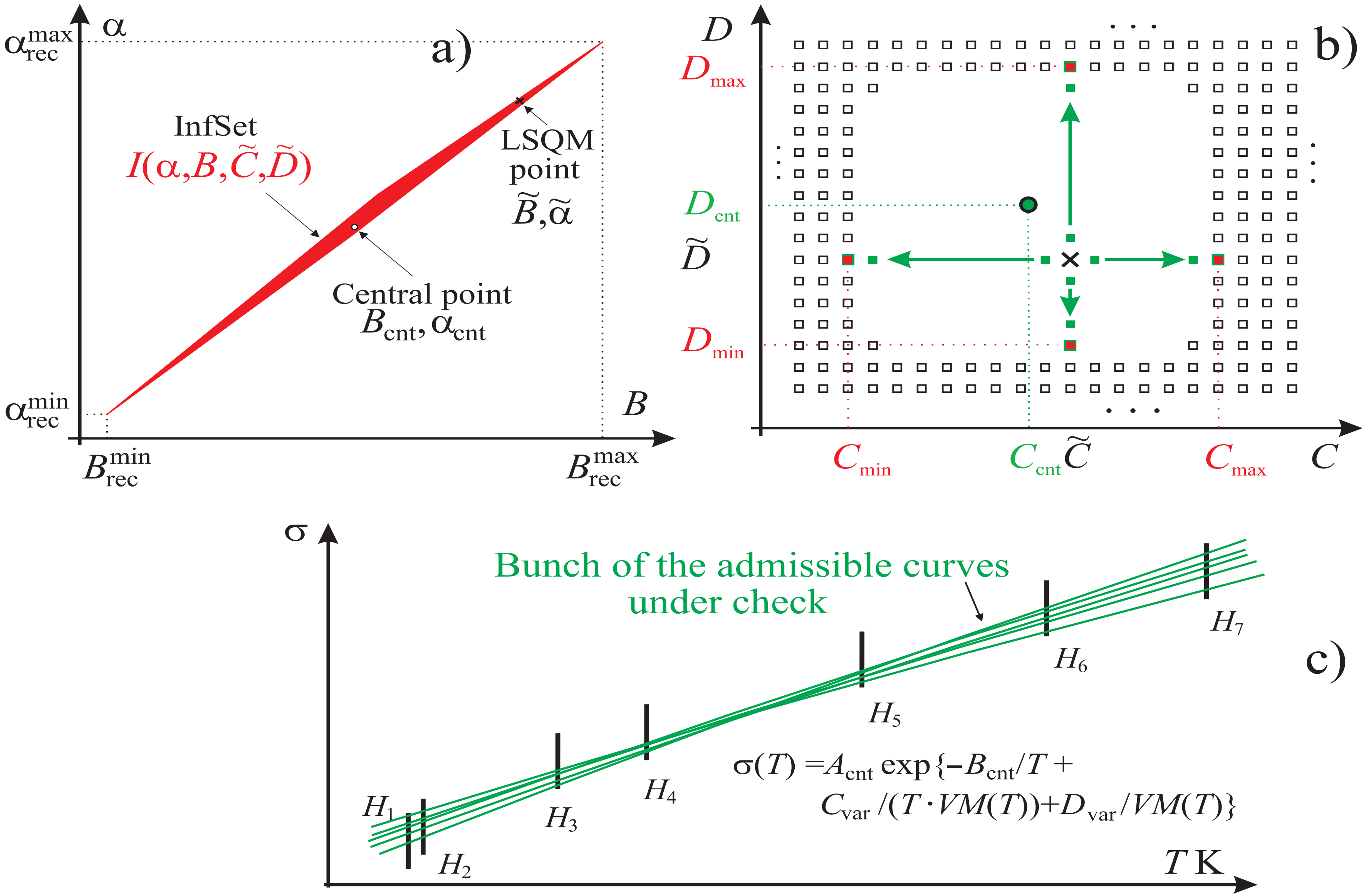
4) Calculation of the central point C_{cntr} of the interval $[C_{\text{min}}, C_{\text{max}}]$.

5) For the fixed point $(\alpha_{\text{cntr}},\ B_{\text{cntr}},\ C_{\text{cntr}})$, one implements variation ("up" and "down" from the value $\widetilde{D}$) of the parameter D and determines the corresponding marginal values D_{min} and D_{max} of the verified inner interval $[D_{\text{min}}, D_{\text{max}}]$. For finding these admissible marginal points, direct check of admissibility of the corresponding curve $\sigma(T, \alpha_{\text{cntr}}, B_{\text{cntr}}, C_{\text{cntr}}, D_{\text{var}})$ is performed by the intervals $\{H_n\}$.

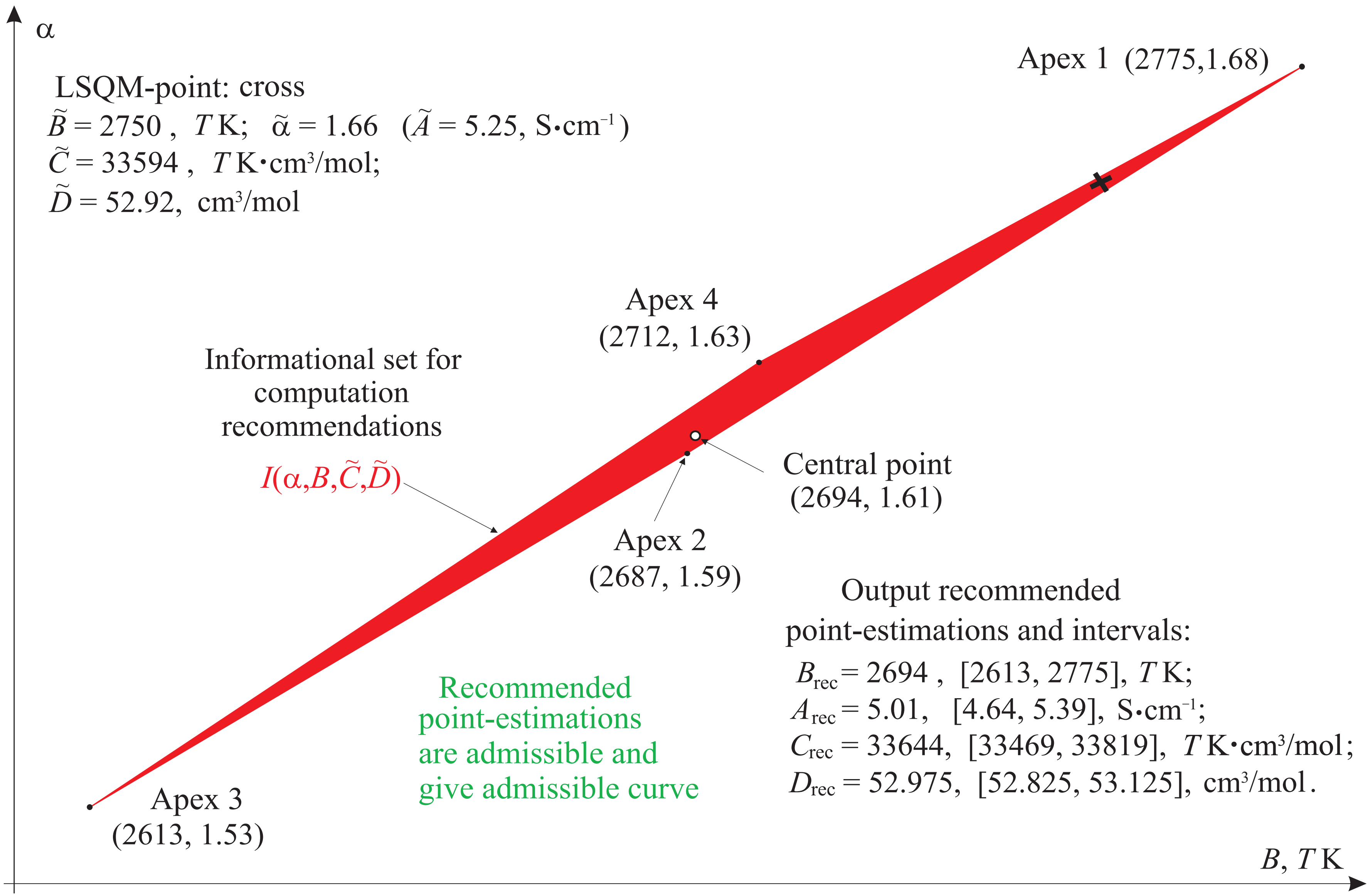
6) Calculation of the central point D_{cntr} of the interval $[D_{\text{min}}, D_{\text{max}}]$.

7) For practical application, the determined central points and approximate inner intervals of parameters are given out.

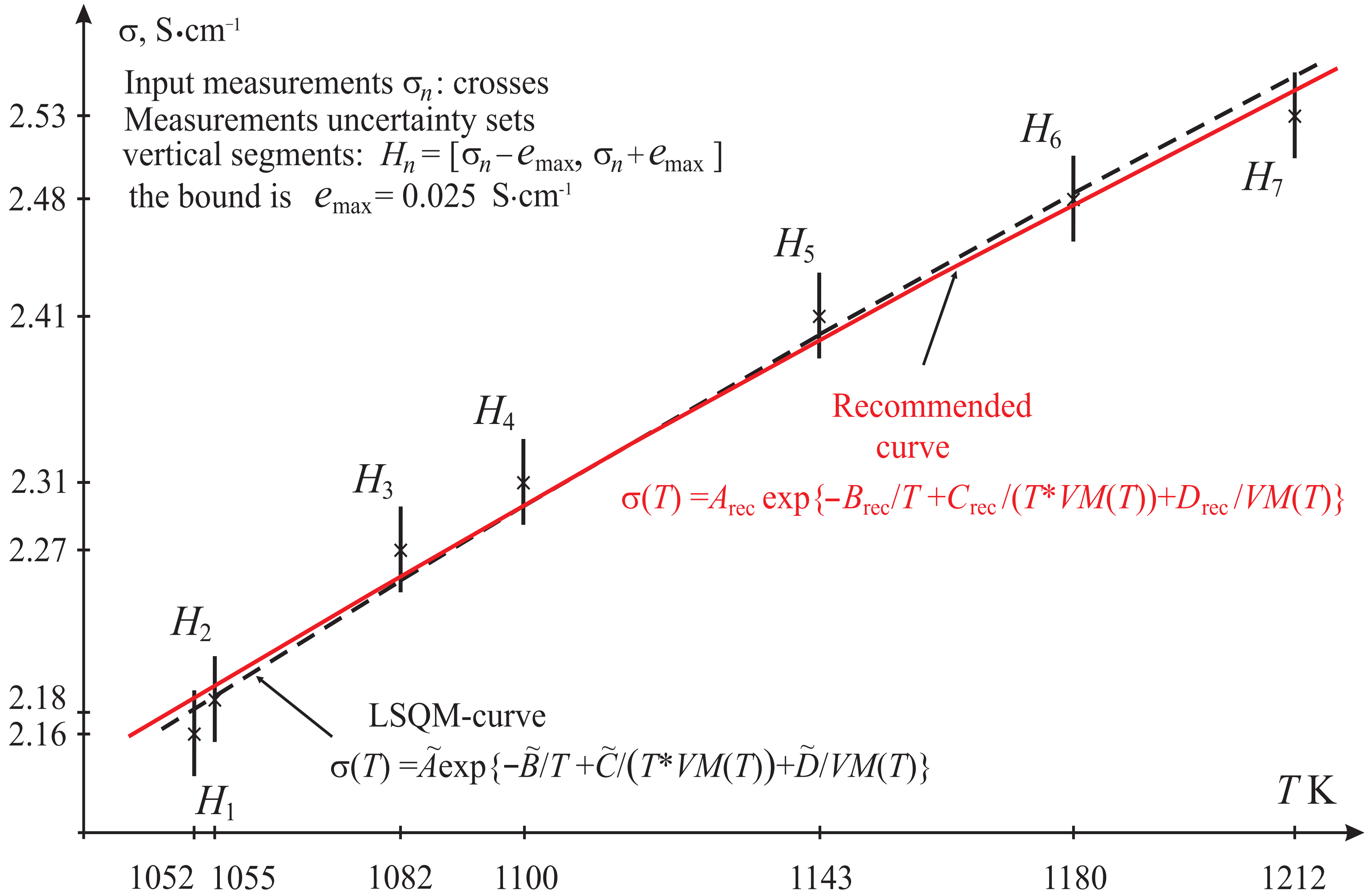
Constructing the recommended output estimations



Recommended output estimations



Recommended output curve



Conclusions

Under mentioned conditions of uncertainty, the interval approach allows one:

– to construct the verified informational set of the process parameters;

– to find the verified estimations of inner intervals for admissible values of the parameters;

– if necessary, using the found InfSet $I(\alpha, B, C, D)$, to construct the tube of admissible process curves and to enhance the uncertainty set of each input measurement.

References

1. Redkin A., Nikolaeva E., Dedyukhin A., and Yu. Zaikov. The electric conductivity of chloride melts // Ionics, 2012, **18**, pp.255–267.

2. Redkin A., Tkacheva O. Electrical Conductivity of Molten Fluoride-Oxide Melts // J. Chem. Eng. Data, 2010, **55**, pp.1930–1939.

3. GOST 8.207-76. The State System for Providing Uniqueness of Measuring. Direct Measuring with Multiple Observation. Methods for Processing the Observation Results. –M.: Goststandart. Official Edition.

4. MI 2083-93. Recommendations. The State System for Providing Uniqueness of Measuring. Indirect Measuring. Determination of the Measuring Results and Estimation of their Errors. –M.: Goststandart. Official Edition.

5. R 40.2.028–2003. Recommendations. The State System for Providing Uniqueness of Measuring. Recommendations on Building the Calibration Characteristics. Estimation of Errors (Uncertainties) of Linear Calibration Characteristics by Application of the Least Square Means Method. –M.: Goststandart. Official Edition.

6. Kantorovich L.V. On new approaches to computational methods and processing the observations // Siberian mathematical journal. 1962, **III**, no. 5, pp. 701-709.

7. Jaulin L., Kieffer M., Didrit O., and E. Walter. Applied Interval Analysis. Springer-Verlag, London. 2001.

8. Hansen E., G.W. Walster. Global Optimization using Interval Analysis. Marcel Dekker, Inc., New York. 2004.

9. Fiedler M., Nedoma J., Ramik J., Rohn J., and K. Zimmermann. Linear optimization problems with inexact data. Springer-Verlag., London. 2006.

References

10. Shary, S.P. Finite–Dimensional Interval Analysis. Electronic Book, 2014, `http://www.nsc.ru/interval/Library/InteBooks`

11. Shary S.P., Sharaya I.A. Raspoznzvaniye razreshimosti interval'nykh uravneniyi i ego prilozheniya k analizu dannykh //Vichslitelnye tekhnologii, 2013, **8**, no. 3, pp.80–109.

12. Sharaya I.A. Dopuskovoye mnozhestvo resheniyi interval'nykh lineyinykh system uravneniyi so svyazannymi coeffitsientami // in Computational Mathematics, Proc. of XIV Baikal International Seminar-School "Methods of Optimization and Applications". Irkutsk, Baikal, Russia, 2 – 8 July, 2008. Irkutsk, ISEM SO RAS. 2008, **3**, pp.196–203.

13. Zhilin S.I. Simple method for outlier detection in fitting experimental data under interval error // Chemometrics and Intelligent Laboratory Systems, 2007, **88**. pp.6–68.

14. Kumkov S.I., Yu.V. Mikushina. Interval Approach to Identification of Catalytic Process Parameters // Reliable Computing. 2014, **19**, issue 2, pp.197–214.

15. Arkhipov P.A., Kumkov S.I., et.al. Estimation of plumbum Activity in Double systems Pb–Sb and Pb–Bi // Rasplavy. 2012, no. 5, pp.43-52.

16. Kumkov S.I. Processing the experimental data on the ion conductivity of molten electrolyte by the interval analysis methods // Rasplavy. 2010, no. 3, pp.86–96.

17. Kumkov S.I. and Yu.V. Mikushina. Interval Estimation of Activity Parameters of Nano-Sized Catalysts // Proceedings of the All-Russian Scientific-Applied Conference "Statistics, Simulation, and Optimization". The Southern-Ural State University, Chelyabinsk, Russia, November 28–December 2, 2011, pp. 141–146.

18. Potapov A.M., Kumkov S.I., and Y. Sato. Procession of Experimental Data on Viscosity under One-Sided Character of Measuring Errors // Rasplavy (2010), no. 3, pp. 55–70.

Thanks for attention