1 Actual state of the problem

One of possible approaches to *knowledge integration* is based on application of iterative procedures. The best-known example is the *Iterative Proportional Fitting Procedure* (IPFP). It was proved by Csizsár [2] (for the case of discrete probability distributions) that IPFP converges if the *input set* of low-dimensional probability distributions is *consistent*, i.e. if there exists a probability distribution whose marginals equal to the probability distributions from the *input set*. If the initial distribution is decomposable and all the input distributions corresponds to this decomposable model then the distributions computed during IPFP process are *decomposable* as well, so that they can be effectively stored and manipulated. *Decomposability* makes possible the fast computation of conditional probabilities. Csiszár showed that IPFP converges to the distribution that is I_1-projection of initial distribution to the set of distributions having their marginals equal to probability distributions from *input set*. Furthermore, if the initial distribution is uniform (or a product of positive one-dimensional probability distributions), then the resulting distribution maximizes *Shannon entropy* within the set of distributions having its marginals equal to the probability distributions from *input set*. Arguments in favor of *maximum entropy principle* are based on the fact that the distribution maximizing *Shannon entropy* contains least additional information[7, 8].

Another theoretical problem, we are dealing with is the behavior of IPFP in the case of *input set* generated by *decomposable generating class*. For this case, it was proved by S. Haberman [6] that there exists an ordering of elements of the *input set* such that IPFP converges within one *cycle*. Furthermore, for certain decomposable generating classes IPFP converges within one cycle regardless the ordering used. S. Haberman [6] showed that *strongly (totally) decomposable* generating classes meet this property. *Strong decomposability* is a sufficient but not a necessary condition.

2 Objective of the thesis

The main purpose of the thesis is a design and characterization of methods applicable to the process called *knowledge integration*. It is supposed that *knowledge* is represented by a set of probability distributions. Probabilistic models have the ability to handle uncertain information that systems used in artificial intelligence ought to posses. Within the probabilistic framework, *knowledge integration* is understood as a process of building a *joint probability distribution* from an *input set* of low-dimensional distributions. Finally, the process should yield the joint probability distribution embodying knowledge about a chosen domain. In the case of an *inconsistent input set* of probability distributions, the IPFP procedure tends to come in cycles. Nevertheless, we wish to get some representative of *input knowledge* even in the case when it is inconsistent. We endeavor to find *iterative procedures* that converge even for *inconsistent input sets*, the marginals of the resulting distribution should be somehow close to the input distributions and independent of the ordering of the input low-dimensional distributions. Application to *consistent input sets* should lead to the same distributions as classical IPFP.

Another objective of the thesis is to prove a necessary and sufficient condition for the *family of generating classes* that guarantees IPFP convergence within one cycle regardless the ordering used. Further we want to prove or reject Haberman's conjecture given in [6] that for all *decomposable generating classes* IPFP converges within two *cycles* regardless from the order of *generating class*.

3 Applied methods

I-divergence geometry, introduced by Csiszár [2] (I_1*-projection*) and Čencov [1] (I_2*-projection*), extended by Matúš [9, 10] is an effective instrument enabling us to study *iterative procedures* since these procedures are based on the operation of I-projection to a set of probability distributions. We introduce the *additively constrained sets* and the *multiplicatively constrained sets*. I-projections to both the *additively* and *multiplicatively constrained sets* undergo an extensive exploration.

We have exploited properties of β-acyclicity and α-acyclicity of *hypergraphs* defined by Ronald Fagin [5] and the relation between chordal graphs and acyclic hypergraphs proved in [4]. The graph methodology together with properties of probability distributions factorizing with respect to *decomposable generating classes* enabled us to verify results concerning IPFP convergence in the case of *decomposable generating classes*.

4 Results

The first task was to prepare the theory that enable us to design and study iterative methods for integration of *knowledge* represented by a *set of probability distributions*. We presented self-contained theory of two basic types of probability distributions' sets: *additively constrained sets* and *multiplicatively constrained sets*. Afterwards, we applied the theory to the study of iterative methods. Its application was based on the fact that all of proposed methods iterate operations of *I-projections* to *additively constrained sets*. We have done original proofs of some theorems and lemmata of this part, let us make reference at least to the theorem which states that the distribution from intersection of *additively constrained set* $\mathcal{S}_\mathcal{E}$ and *multiplicatively constrained set* $\bar{\mathcal{R}}_\mathcal{E}$ minimizes the *I-divergence* with respect to a distribution from $\mathcal{S}_\mathcal{E}$ within the set $\bar{\mathcal{R}}_\mathcal{E}$.

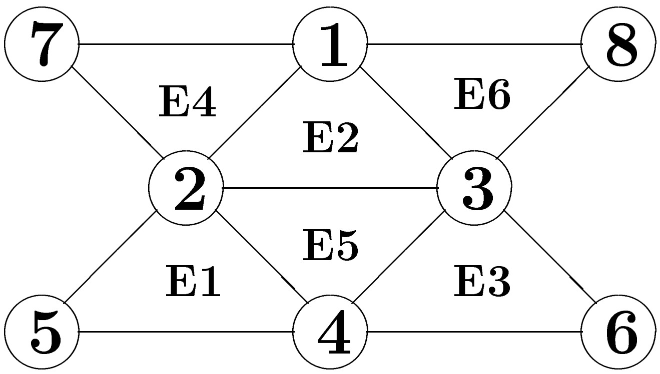
Theorem 1 *Let $\mathcal{S}_\mathcal{E}$ be an additively constrained set. For $\hat{S} \in \bar{\mathcal{R}}_\mathcal{E} \cap \mathcal{S}_\mathcal{E}$, $Q \in \bar{\mathcal{R}}_\mathcal{E}$, and $S \in \mathcal{S}_\mathcal{E}$ it holds that $I(S \parallel Q) \;\geq\; I(S \parallel \hat{S})$.*

Some of the fundamental theorems are rewritten from different papers [2, 3, 9, 10]. We had to adapt these theorems in order to give them in the form appropriate for the context. Some proofs were not available, thus we had to work them out afresh.

The second basic part was devoted to the study of *iterative proportional fitting procedure* in the case of *decomposable generating classes*. We have achieved several original results. We proved a sufficient condition for IPFP convergence within one cycle, which extends the family of *decomposable generating classes* such that for any ordering of the sets in the class IPFP converges within one cycle.

Theorem 2 *Suppose there is a set $E^* \in \mathcal{E} = \{E_i : i = 1, 2, \ldots, s\}$ such that $E_i \cap E_j \subset E^*, \forall i, j = 1, 2, \ldots s, i \neq j$. If $\mathcal{E}$ is a generating class of input set $\mathcal{P}_\mathcal{E}$ and if there exist $P \in \mathcal{S}_\mathcal{E}, P \ll Q_{(0)}$ then IPFP starting with $Q_{(0)} \in \mathcal{R}_\mathcal{F}$ converges within s steps (one cycle) no matter how $\mathcal{E}$ is ordered.*

An important result is the counterexample to S. Haberman conjecture [6], that for all *decomposable generating classes* IPFP always converges within two *cycles*. An example of such ordering is the ordering $\{E_1, E_3, E_4, E_6, E_5, E_2\}$ of the *input set* whose *generating class* is described by the following figure.



It is known that orderings of the *input set* meeting *running intersection property* guarantee IPFP convergence within one cycle We proved a sufficient condition that extends the class of orderings that guarantee IPFP convergence within one cycle.

Theorem 3 *Let* $\{E_1, E_2, \ldots E_s\}$ *be an ordering of generating class* $\mathcal{E}$, *initial probability distribution* $Q_{(0)} \in \mathcal{P}$ *be uniform and there exist a* $Q \in \mathcal{S}_{\mathcal{E}} = \{Q(x) \in \mathcal{P} : Q^{E_i} = P_{E_i}, i = 1, 2, \ldots, s\}$. *If for every iteration* $1 < i \leq s$ *hypergraph*

$$H_{(i)} = (V_{(i)}, \mathcal{E}_{(i)}), \quad V_{(i)} = \cup_{j=1}^{i} E_j, \quad \mathcal{E}_{(i)} = \{E_1, E_2, \ldots E_i\}$$

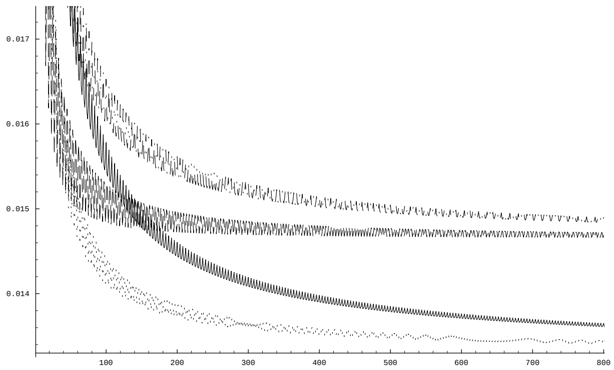
is decomposable then IPFP converges in one cycle.

We have proposed two conjectures. The first one offers a full characterization of chordal graphs corresponding to generating classes that guarantee IPFP convergence within one cycle. The second one is the analogy of the first one for the convergence within two cycles.

The third task that we solved in this text was to find a method that converge to a joint probability distribution even if the *input set* is *inconsistent*. The requirements were that the marginals of the resulting distribution are close to the input distributions and the resulting distribution does not depend on the ordering of the *input set*. An application to the *consistent input sets* should result in the same distribution as *iterative proportional fitting procedure*. The detailed study of five *iterative methods* for the *probabilistic knowledge integration* is presented: *convex* and *log-convex conservative modifications* of IPFP, *methods of iterative geometric* and *arithmetic averages*, and a method that is an instance of

GEM-algorithm. Discussion of their properties and experimental results are given together with several examples.

The following plot shows values of the joint probability distributions for one combination of random variables' values computed using *conservative modifications* on *inconsistent input set* $\{P_{E_1}, P_{E_2}, P_{E_3}, P_{E_4}\}$. Since the oscillations of the joint probability distributions' values within one cycle are significantly high and graphics resolution is limited, the curves are displayed as bands.



A serious difficulty that can complicate practical application of *conservative modifications* is their low convergence rate for both *inconsistent input sets* and *consistent input set* as well.

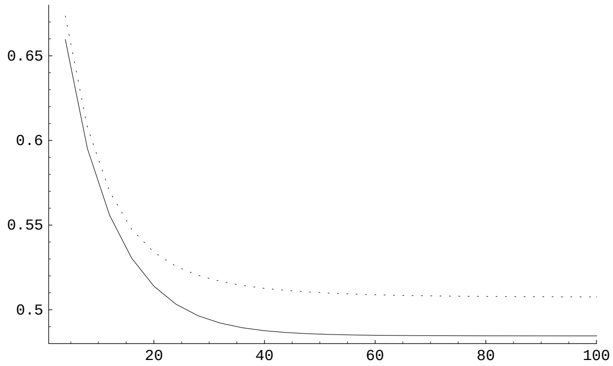
More promising are properties of *method of iterative geometric* and *arithmetic averages.* It is shown that *method of iterative geometric* minimizes

$$\sum_{j=1}^{s} w_j \cdot I(P_{E_j} \parallel Q^{E_j})$$

with respect to given probability distributions $P_{E_j}, j = 1, 2 \ldots s$. On the other side, *method of iterative geometric averages* and our proposed method that is an instance of *GEM-algorithm* minimize

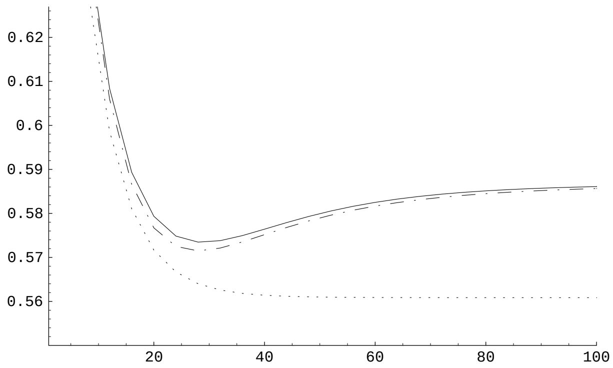
$$\sum_{j=1}^{s} w_j \cdot I(Q^{E_j} \parallel P_{E_j}).$$

The following figures compare function history of $\sum_{j=1}^{s} I(P_{E_j} \parallel Q_{(i)}^{E_j})$ and $\sum_{j=1}^{s} I(Q_{(i)}^{E_j} \parallel P_{E_j})$, respectively, during the processes of *arithmetic averages* (full line), *geometric averages* (dotted line) and *GEM-algorithm* (in the first figure equivalent to arithmetic averages, in the second one dash and dot line) applied to an *inconsistent input set.*



We have proposed the instance of the GEM-algorithm, which we claim to be the most appropriate of all the methods described. Results proven for the general GEM-algorithm were exploited to prove convergence properties of the proposed method. This method meets all the demanded properties and since it results in the joint probability distribution that *factorizes* with respect to the *generating class* it can be considered to be an extension of *iterative proportional fitting procedure* to the case of of the not necessarily *consistent input set.* Since *inconsistent input sets* of probability distributions can be derived from *incomplete (missing) data* we did an overview of methods that are used for *missing data* in statistics. A comparison of these methods with the proposed iterative procedures has been carried out on a simple problem.

In spite of our efforts some questions remains still open. They concern two areas. At first, the conjectures concerning full characterization of *decomposable generating classes* that guarantee IPFP convergence within one and two cycles need to be proven. In the thesis, we strove for results that are not restricted to strictly positive probability distribution. However, some assertions concerning convergence properties of iterative



methods were not achieved in this general case due to problems with probability distributions containing zeroes. Thus, convergence proofs of some proposed methods are still missing in this general case.

References

[1] N. N. Čencov. *Statistical Decision Rules and Optimal Inference.* Translations of Mathematical Monographs. American Mathematical Society, Providence, 1982.

[2] I. Csiszár. *I*-divergence geometry of probability distributions and minimization problems. *Ann. Prob.*, 3(1):146–158, 1975.

[3] I. Csizsár and G. Tusnády. Information geometry and alternating minimization procedures. *Statistics & Decisions*, Supplement Issue No. 1:205–237, 1984.

[4] A. D'Atri and M. Moscarini. On hypergraph acyclicity and graph chordality. *Information Processing Letters*, 29(5):271–274, 1988.

[5] R. Fagin. Degrees of acyclicity for hypergraphs and relational database schemes. *Journal of the Association for Computing Machinery*, 30(3):514–550, 1983.

[6] S. Haberman. *The Analysis of Frequency Data.* The University of Chicago Press, Chicago and London, 1974.

[7] P. Hájek, T. Havránek, and R. Jiroušek. *Uncertain Information Processing in Expert Systems.* CRC Press, Boca Raton, Ann Arbor, London, Tokyo, 1992.

[8] R. Jiroušek. *Probabilistic Methods in Artificial Intelligence (DrSc Thesis).* Institute of Information Theory and Automation, Prague, 1993.

[9] F. Matúš. On alternating minimization of i-divergence and averages of i-projections. *(unpublished manuscript)*, 1997.

[10] F. Matúš. On *I*-divergence geometry of affine and log-affine sets. *(unpublished manuscript)*, 1999.

5 List of publications and references

[**1**] R. Jiroušek and J. Vomlel. Inconsistent knowledge integration in a probabilistic model. In *Proceedings of Workshop "Mathematical Models for handling partial knowledge in A.I."*, Erice, Sicily, 1994. Plenum Publ. Corp.

Participation 80% approved by Radim Jiroušek (20%)

Referred in:

Zentralblatt MATH 1999, European Mathematical Society, FIZ Karlsruhe & Springer-Verlag.

[**2**] J. Vomlel. *Probabilistic Models in Artificial Intelligence.* Research Report, Czech Technical University, Prague, Czech Republic, 1995.

[**3**] J. Vomlel. *Dependency Models.* Research Report, Institute of Information Theory and Automation, 1996, Prague.

[**4**] J. Vomlel. Statistical methods for probabilistic model parameter estimation from incomplete data and their application to the marginal problem. In *Proceedings of Workshop on Uncertainty Processing*, Prague, Czech Republic, 1997. University of Economics.

Referred by:

F. Matúš. *On alternating minimization of I-divergence and averages of I-projections*, 1997, (unpublished manuscript submited to Annals of Statistics)

6 Summary

The objective of the thesis is the design and description of methods applicable to the integration of *knowledge* represented by a *set of probability distributions* even in the case of *input set* being inconsistent.

We have presented five new methods that under some assumptions converge even if the *input set* is *inconsistent*. The theory of *additively constrained sets* and *multiplicatively constrained sets* has enabled us to describe and confirm the methods' properties. The fundamental advantage of proposed iterative methods is that, in comparison with conventional optimization methods, they do not require complicated calculations. This makes it possible to apply them to problems of high dimensionality. We have implemented all these methods and performed a number of computational experiments. A systematic summary of the achieved results is given. One of the proposed methods, an instance of the GEM-algorithm, can be considered to be the appropriate method for the task of *inconsistent knowledge integration*.

In the case of *consistent input set iterative proportional fitting procedure* possess the best convergence rate of all presented iterative methods. Moreover, in the case of a *decomposable generating class*, there exist orderings of the *input set* that guarantee IPFP convergence within one cycle. The thesis extends the family of *decomposable generating classes* for which IPFP converges within one cycle regardless the ordering. On the other side, we have shown that there exist *decomposable generating classes* such that (in spite of Haberman's conjecture of 1974) IPFP need not converge even within two cycles. Still, there are several open questions in the study of *decomposable generating classes*.

The theory of *iterative methods* proposed for *probabilistic knowledge integration*, which we have presented, makes a self-contained and closed exposure. Iterative methods can, for instance, play the role of algorithms for learning parameters of probabilistic systems. At the present time, the learning algorithms receive great attention. *Expert* or *knowledge systems* with a probabilistic engine have been predicted to be applied in many domains in the near future.

7 Resumé

Cílem disertační práce je návrh a popis metod, které se hodí pro *spojování znalostí* reprezentovaných pomocí *pravděpodobnostních distribucí* i v případě, kdy tyto distribuce nejsou konzistentní.

V práci je navrženo pět nových metod, které za určitých podmínek konvergují i v případě, kdy *vstupní množina* distribucí není *konzistentí.* Teorie *součtově* a *součinově omezených množin* pravděpodobnostních distribucí nám umožnila efektivní popis vlastností navržených metod. Základní výhodou navržených iterativních metod je, že, v porovnání s běžnými optimalizačními metodami, nevyžadují složité výpočty. Všech pět metod bylo implementováno, což umožnilo provést velké množství experimentů s různými *vstupními množinami* distribucí. Práce systematicky shrnuje dosažené výsledky. Jednu z námi navržených metod, která je instancí GEM algoritmu, považujeme za nejvhodnější metodu pro *spojování nekonzistentních znalostí*

V případě *konzistentní vstupní množiny* distribucí, rychleji než navržené metody, konverguje známá metoda, zkráceně nazývaná IPFP. Navíc v případě, kdy *vstupní množina* distribucí je generována *rozložitelnou třídou* podmnožin *indexové množiny*, existují uspořádání *vstupní množiny*, taková, že IPFP konverguje během jednoho cyklu. Disertace rozšiřuje rodinu *generujících tříd*, pro které IPFP konverguje bez ohledu na uspořádání *vstupní množiny.* Na druhou stranu, je v disertaci ukázáno, že existují *rozložitelné generující třídy*, pro které, navzdory doměnce S. Habermanové z roku 1974, existují uspořádání taková, že IPFP nemusí konvergovat ani během dvou cyklů. V teorii *rozložitelných generujících tříd* zůstává stále ještě několik otevřených otázek.

Prezentovaná teorie *iterativních metod* pro *spojování pravděpodobnostních znalostí* představuje uzavřený a relativně nezávislý celek. *Iterativní metody* mohou být v praxi aplikovány jako učící algoritmy *pravděpodobnostních systémů.* V současné době je studiu učících algoritmů věnována značná pozornost. Očekává se, že expertní nebo znalostní systémy s pravděpodobnostním inferenčním mechanismem, budou v brzké době aplikovány v mnoha oblastech.