Alkalmazott Matematikai Lapok **15** (1990–91) 163–179

PARCIÁLIS FÜGGŐSÉG RELÁCIÓS ADATBÁZISOKBAN*

DEMETROVICS JÁNOS, KATONA GYULA és MIKLÓS DEZSŐ

Budapest

Az *Armstrong féle funkcionális függőség* fogalmának gyengítésével kapjuk a relációs adatbázisokon értelmezett parciális függőség fogalmát. Megmutatjuk, hogy a parciális függőségek jól jellemezhetők az adatbázis attributum-halmazán definiált parciális függvények által alkotott parciálisan rendezett halmaz lezárásaival. Más oldalról, szükséges és elégséges feltételt adunk ahhoz, hogy egy ilyen lezáráshoz találjunk az adott attributum-halmazon olyan adatbázist, ami által definiált parciális függőségek pont az adott lezárást generálják. Megvizsgáljuk azt is, hogy milyen tulajdonságú függőséget definiálnak a parciális függőségek az attributumok részhalmazai által alkotott parciálisan rendezett halmazon és hogy ezek miképpen realizálhatók adatbázisokkal.

1. Bevezetés

Ebben a dolgozatban relációs adatbázisok egy újfajta függőségét, az ún. *parciális függőséget* vizsgáljuk, amely először [4]–ben került bevezetésre. Az adatbázisok modelljének itt a mátrixot tekintjük. A mátrix oszlopai, azaz az egyes lehetséges tulajdonságok nevei az adatbázis *attributumai*, míg a sorok, azaz az egyes elemek tulajdonságainak az összessége az adatbázis *rekordjai.*

Jelöljük Ω–val az attributumok (azaz a mátrix oszlopainak) halmazát. Legyen $A \subseteq \Omega$, $b \in \Omega$. A függőség hagyományos definíciója szerint (ARMSTRONG [1], CODD [2]) akkor mondjuk, hogy b *funkcionálisan függ* A–tól, ha az A–ban álló adatok már egyértelműen meghatározzák a b–ben lévő adatot is. Sok gyakorlati esetben egy ilyen függés csak majdnem teljesül. Pl. egy nem túl nagy adathalmaz esetén (egy iskola tanulóinak adatai) a név rendszerint meghatározza a személyt, azaz minden más adatot, azaz ettől az egyelemű A adathalmaztól minden más b adat funkcionálisan függ. „Csak" néhány kivétel van. A *Kovácsok, Tóthok*, esetleg két unokatestvér azonos névvel. Tehát az A–ban lévő legtöbb adat már egyértelműen meghatározza a többi b oszlop adatait, de néhány más kivételes adat nem. Ezekhez még egy (vagy több) azonosító adat is kell. Ez motiválja a parciális függőség fogalmának bevezetését. Heurisztikusan szólva akkor mondjuk, hogy az A oszlophalmazban álló adatsorozattól b parciálisan függ, ha minden sorban (rekordban), ahol A–ban ezen értékek állnak, b–ben ugyanaz az érték van.

Két indokot tudunk felsorakoztatni e struktúra vizsgálata mellett. Az egyik elméleti. Minden funkcionális függőségi modell mögött ott van a parciális függőségek modellje is. Ez tehát egy finomabb modell, ami azonban még mindig csak modell

*Az 2575–ös Országos Tudományos Kutatási Alap által támogatva

a valódi adatbázis és a funkcionális függőségi modell között. Tehát a rá vonatkozó eredmények más megvilágításba helyezhetik a funkcionális függőségek modelljét.

A másik indok gyakorlati. Tekintsünk egy példát. Tegyük fel, hogy $\Omega = \{x_1, x_2, x_3, x_4\}$, x_2 funkcionálisan függ x_1–től, x_4 pedig x_3–tól. Az egész 4-oszlopú mátrix tárolása helyett elegendő x_1 és x_3 oszlopait tárolni, emellett két kis mátrix tárolására van szükség, amely megadja a két funkcionális függést. Jól látható, hogy ez jelentős memóriamegtakarítást eredményez. Ha x_2–nek x_1–től, illetve x_4–nek x_3–tól való függése csak néhány kivételes érték miatt nem teljesül, akkor még mindig sok tárolóhelyet takaríthatunk meg, ha különválasztjuk a kivételes x_1, x_3 párokat, és mellettük felsoroljuk a hozzájuk tartozó x_2–t és x_4–et, a többivel meg úgy járunk el, mint a „rendesen" függő esetben. Tehát ez a finomabb modell helymegtakarítást tehet lehetővé olyankor, amikor a funkcionális függőség segítségével ezt nem érhetjük el.

A dolgozat 2. fejezetében megadjuk a szükséges definíciókat, a parciális függőség definícióját és különböző modelljeit. A 3. fejezetben megvizsgáljuk a parciális függőségek realizálhatóságának kérdését, illetve azt, hogy a parciális függőségek milyen struktúrát hoznak létre az attributumok halmazán.

2. A parciális függőség definíciója és különböző modelljei

Az adatbázis modellje itt az M mátrix, amelynek a sorai számát m és az oszlopai számát n jelöli. Az oszlopokat más néven attributumoknak nevezzük, ezek halmazát Ω jelöli. Az adatbázis elemein a továbbiakban a mátrix elemeit értjük.

Az adatbázisokon definiált függőségek talán legismertebbike a *funkcionális függőség* (vagy röviden csak függőség), amelyet a következő módon definiálunk: ha $A, B \subseteq \Omega$ két részhalmaza az attributumok halmazának, akkor azt mondjuk, hogy B funkcionálisan függ A–tól, amennyiben az adatbázisnak nincs két olyan sora, melyben az elemek megegyeznek az A halmazon, de legalább egy helyen különböznek a B halmazon. Jelölése $A \longrightarrow B$. Más szavakkal ez azt jelenti, hogy az adatbázis A–beli oszlopaiban lévő adatok meghatározzák a B–beli oszlopokban levő adatokat. Ebben az esetben nem lényeges, hogy mik a konkrét elemek az egyes attributumoknál, csak az, hogy ha azok két különböző sorban megegyeznek az A–beli attributumoknál, akkor meg kell egyezniük a B–belieknél is. Az egyes adatbázisokat nyilván nem különböztetik meg egyértelműen az általuk meghatározott függőségek, de azért sok lényeges információt tartalmaznak. A funkcionális függőségek megadott halmaza felfogható, mint azon (konkrét) mátrixok modellje, amelyek kielégítik ezeket a funkcionális függőségeket.

A funkcionális függőség gyengítésével kaphatjuk a sokkal gyengébb és konkrétabb *parciális függőséget*, amely már figyelembe veszi az egyes konkrét elemeket is. Tegyük fel, hogy egy adatbázisban az a_1 és a_2 oszlopokban előforduló elemek nem határozzák meg mindig egyértelműen a b oszlopban előforduló elemet, ám ha egy sorban az a_1 attributum értéke r_1 és a_2 attributum értéke r_2, akkor a b attributum értéke r_3, azaz egyértelműen meghatározott. Ezt a tényt úgy fogjuk jelölni, hogy $(a_1, a_2; r_1, r_2) \rightarrow (b; r_3)$. Az általános definícióhoz szükségünk van a *parciális függvény* fogalmára. Ha az a_i attributum lehetséges értékeinek a halmaza az adott adatbázis esetén D_i és $r_i \in D_i$, akkor $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$-t parciális függvénynek hívjuk, amennyiben az a_i elemek az Ω különböző elemei. Az $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ parciális függvény *értelmezési tartománya* az $\text{ÉT}(\alpha) = \{a_1, \ldots, a_k\}$ halmaz, *értékkészlete* pedig az $\text{ÉK}(\alpha) = \{r_1, \ldots, r_k\}$ halmaz. Azt mondjuk, hogy a $\beta = (b_1, \ldots b_l; s_1, \ldots, s_l)$ parciális függvény *függ* az α–tól (és úgy jelöljük, hogy $\alpha \rightarrow \beta$), ha minden olyan sorban, amelyben az a_i oszlopokban r_i elemek vannak $(1 \leq i \leq k)$ a b_j oszlopokba s_j elemek kerülnek $(1 \leq j \leq l)$.

Természetesen a fenti definíció nem értelmezhető az összes parciális függvényre egy adott adatbázis esetében, csak azokra, amelyek *koherensek*, azaz előfordulnak az adatbázis valamelyik sorában (vagy, kényelmi szempontokból feltehetjük, hogy az „ értelmetlen", nem koherens parciális függvényektől az összes koherens parciális függvény függ). Mindazonáltal a koherens parciális függvények egy szép struktúrát alkotnak. Ennek megadásához szükségünk van egy definícióra. Ha $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ és $\beta = (b_1, \ldots, b_l; s_1, \ldots, s_l)$ két parciális függvény, akkor $\alpha \subset \beta$, azaz α része β-nak, amennyiben $\{a_1, \ldots, a_k\} \subseteq \{b_1, \ldots, b_l\}$ és $a_i = b_j$-ből következik, hogy $r_i = s_j$. Azt mondjuk, hogy a parciális függvények egy P halmaza *leszálló* struktúrájú, ha az alábbi két tulajdonságot kielégíti:

(2.1) ha $\alpha \in P$ és $\beta \subset \alpha$, akkor $\beta \in P$.

(2.2) minden $\alpha \in P$–re létezik egy olyan $\gamma \in P$, amelyre $\alpha \subseteq \gamma$, $\gamma = (c_1, \ldots, c_n; \ldots)$ és $\{c_1, \ldots, c_n\} = \Omega$.

Nyilvánvaló, hogy egy adott adatbázis esetében a koherens parciális függvények halmaza leszálló lesz.

Parciális függvények uniója és metszete is definiálható, (még ha az első csak bizonyos párokra is) habár kicsit más módon, mint az egyszerű halmazok esetében. Legyen $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ és $\beta = (b_1, \ldots, b_l; s_1, \ldots, s_l)$ két parciális függvény. Kettőjük metszete az a γ parciális függvény, amelynek az értelmezési tartománya azon c attributumok halmaza, amelyek α értelmezési tartományának is és β értelmezési tartományának is elemei (mondjuk $c = a_i = b_j$) és amelyekre $r_i = s_j$. A γ parciális függvény természetesen az $r_i = s_j$ értéket veszi fel a c helyen. Az unió értelmezése egy kicsit több figyelmet kíván. Az $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ és $\beta = (b_1, \ldots, b_l; s_1, \ldots, s_l)$ parciális függvényeknek csak akkor értelmezzük az unióját, ha a metszetük értelmezési tartománya megegyezik az értelmezési tartományaik metszetével. Nem nehéz megállapitani, hogy ez pontosan akkor áll fenn, ha a két parciális függvény a mindkettőjük értelmezési tartományában előforduló attributumokon ugyanazt az értéket veszi fel. Ebben az esetben értelmezhetjük (és értelmezzük) az uniójukként azt a γ parciális függvényt, amelynek az értelmezési tartománya a két eredeti függvény értelmezési tartományának uniója, és amely egy c attributumon a következő értéket veszi fel:

$$\begin{aligned} r_i - t, &\quad \text{ha } c = a_i \text{ és } c \notin \{b_1, \ldots, b_l\}, \\ s_j - t, &\quad \text{ha } c = b_j \text{ és } c \notin \{a_1, \ldots, a_k\}, \\ r_i = s_j - t, &\quad \text{ha } c = a_i = b_j. \end{aligned}$$

Az eddig megadott alapvető definíciók segítségével megadhatjuk a parciális függvények lezártját, hasonlóan, mint ahogy a funkcionális függőségek esetében az Ω egy részhalmazának a lezártja adott. A definíció világos megadása érdekében azonban először egy lemmát mondunk ki a parciális függőségek tulajdonságairól.

2.1. LEMMA. *Ha α, β, γ, δ egy adatbázishoz tartozó parciális függvények, akkor*
(2.3) $\alpha \to \alpha$;
(2.4) $\alpha \to \beta$ és $\beta \to \gamma$-ból következik, hogy $\alpha \to \gamma$;
(2.5) $\alpha \subseteq \gamma$, $\delta \subseteq \beta$ és $\alpha \to \beta$-ból következik, hogy $\gamma \to \delta$;
(2.6) ha $\alpha \to \beta$, $\gamma \to \delta$ és $\alpha \cup \gamma$ létezik és koherens, akkor $\beta \cup \delta$ is létezik és koherens, és $\alpha \cup \gamma \to \beta \cup \delta$.

Bizonyítás. (2.3) nyilvánvaló. A továbbiakban legyen $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$, $\beta = (b_1, \ldots, b_l; s_1, \ldots, s_l)$, és $\gamma = (c_1, \ldots, c_p; t_1, \ldots, t_p)$. Ha $\alpha \to \beta$ és $\beta \to \gamma$, akkor az adatbázis minden olyan sorában, ahol az a_i oszlopokban r_i elemek állnak a b_j oszlopokba s_j elemek kerülnek, és így a c_h oszlopokba t_h elemek lesznek, azaz $\alpha \to \gamma$, ami bizonyítja (2.4)-et.

(2.5) bizonyításához elég belátni, hogy $\alpha \subseteq \gamma$-ból következik, hogy $\gamma \to \alpha$, ami éppen úgy nyilvánvaló a definíciókból, mint (2.3). Akkor viszont az itteni feltételekből $\gamma \to \alpha$, $\alpha \to \beta$ és $\beta \to \delta$ következik, amelyek a $\to$ reláció (2.4) szerinti tranzitív volta miatt $\gamma \to \delta$-t adják.

(2.6) bizonyításakor legyen $\delta = (d_1, \ldots, d_q; u_1, \ldots, u_q)$. Ha α és γ két parciális függvény, akkor a „$\alpha \cup \gamma$ létezik és koherens" feltétel azt jelenti, hogy léteznek sorok, amelyek az a_i oszlopokban r_i értékeket vesznek fel és a c_h oszlopokban t_h értékeket, beleértve azt is, hogy amennyiben valamely $a_i = c_h$, akkor r_i megegyezik t_h-val. Ezekben a sorokban viszont a $\alpha \to \beta$ feltételből következően a b_j oszlopokban s_j értékeknek és az d_g oszlopokban u_g értékeknek kell szerepelniük. Ez természetesen maga után vonja, hogy amennyiben valamely $b_j = d_g$, akkor s_j-nek meg kell egyeznie u_g-vel, azaz a $\beta \cup \delta$ parciális függvény létezik (definiált). Koherens is lesz, hiszen minden olyan sor, amely az $\alpha \cup \gamma$ parciális függvényt tartalmazza, tartalmazza $\beta \cup \delta$-t is. Természetesen az eddig elmondottak a $\alpha \cup \gamma \to \beta \cup \delta$ relációt is bizonyítják. □

Ezen lemma birtokában térjünk rá a parciális függvények lezárásának a definíciójára. Adott adatbázis esetén egy α tetszőleges (koherens) parciális függvényhez rendeljük hozzá $L(\alpha)$-t, azt a legnagyobb β parciális függvényt, amelyre $\alpha \to \beta$. Itt a legnagyobb alatt azt a parciális függvényt értjük, amelynek az értelmezési tartománya a legnagyobb. Ez nyilván létezik az adott esetben, mivel ha $\alpha \to \beta_1$ és

$\alpha \to \beta_2$, akkor a β_1 és β_2 parciális függvényeknek a (2.6) pont értelmében létezik az uniója és $\alpha \to \beta_1 \cup \beta_2$. Tehát $L(\alpha)$ nem más, mint

$$L(\alpha) = \bigcup_{\alpha \to \beta} \beta . \tag{2.7}$$

Természetesen a fenti unióban nem kell az összes parciális függvényt szerepeltetni, elég azokat, amelyek egyetlen attributumot tartalmaznak. $L(\alpha)$ ezek birtokában is megkapható:

$$L(\alpha) = \bigcup_{\alpha \to (b;s)} (b;s) . \tag{2.8}$$

Ezen L függvényről az alábbi tulaldonságokat állíthatjuk.

2.2. LEMMA. *Ha α és β egy adott adatbázishoz tartozó két parciális függvény, akkor*

$$\alpha \subseteq L(\alpha) ; \tag{2.9}$$

$$\alpha \subseteq \beta - \textit{ból következik, hogy } L(\alpha) \subseteq L(\beta) ; \tag{2.10}$$

$$L(L(\alpha)) = L(\alpha) . \tag{2.11}$$

Bizonyítás. $L(\alpha)$ definíciója szerint (2.9) egyszerű következménye (2.3)–nak. (2.5) szerint $\alpha \subseteq \beta$–ból következik, hogy $\alpha \to \gamma$ maga után vonja a $\beta \to \gamma$ relációt is. Így minden olyan parciális függvény, amely $L(\alpha)$ (2.7) szerinti definíciójában az unióban szerepel, szerepel $L(\beta)$ definíciójában is, azaz $L(\alpha) \subseteq L(\beta)$ fennáll.

(2.11) bizonyításához (2.9) szerint elég belátni, hogy $L(L(\alpha)) \subseteq L(\alpha)$. Tegyük fel, hogy a c attributum eleme $L(L(\alpha))$ értelmezési tartományának, azaz valamely $(c;s)$ parciális függvény szerepel $L(L(\alpha))$ (2.8) szerinti előállításában. Ekkor, definíció szerint, $L(\alpha) \to (b;s)$. Ugyanakkor, megint csak definíció szerint $\alpha \to L(\alpha)$, és így a $\to$ reláció 2.1. lemmában bizonyított tranzitivitása miatt $\alpha \to (b;s)$, tehát a $(b;s)$ parciális függvény szerepel $L(\alpha)$ (2.8) tipusú előállításában. □

Egy olyan L függvényt, amely parciális függvények egy leszálló tulajdonságú halmazán van értelmezve és kielégíti a (2.9)–(2.11) tulajdonságokat, *függvény-lezárásnak* [4] nevezünk. Bizonyítottuk tehát, hogy az a $\alpha \to L(\alpha)$ függvény, amely egy adott adatbázis esetén az α parciális függvényhez hozzárendeli annak (2.7) szerinti *lezártját*, egy függvénylezárás. Ezen fejezet további részében a funkcionális függőségből kapott szokásos matematikai lezárás és maga a függőség kapcsolatának [3, 4]–ben elvégzett ill. található vizsgálatait ismételjük meg parciális függőségekre.

Először is megjegyezzük, hogy egy (adott adatbázishoz tartozó) parciális függőségből kapott L függvény-lezárás egyértelműen meghatározza a parciális függőséget.

2.3. LEMMA. *$\alpha \to \beta$ pontosan akkor áll fenn, ha $\beta \subseteq L(\alpha)$.*

Bizonyítás. Nyilvánvaló $L(\alpha)$ (2.7) definíciójából. □

Egy α parciális függvényről azt mondjuk, hogy egy adott adatbázishoz tartozik, ha α koherens az adatbázisban. Legyen T (egy adott adatbázishoz tartozó) parciális függőségekből alkotott rendezett párok családja. Azt mondjuk, hogy a T egy ***függőségi család***, ha a T–ből definiált

$$\alpha \to \beta \text{ akkor és csak akkor, ha } (\alpha, \beta) \in T$$

$\to$ reláció kielégíti a (2.3)–(2.6) tulajdonságokat.

A 2.1. lemmában láttuk, hogy ha a T családot úgy definiáljuk, hogy $(\alpha, \beta) \in T$ pontosan akkor, ha β függ α–tól, akkor így egy függőségi családot kapunk. A 2.2. lemmában a lezárás tulajdonságairól bizonyítottak abból következtek, amit a $\to$ relációról a 2.1. lemmában bebizonyítottunk, tehát egy L függvény-lezárás függvényt minden függőségi családhoz definiálhatunk.

2.4. TÉTEL. *Az ugyanazon alaphalmazon értelmezett T függőségi családok és L függvény-lezárási operációk között egy egyértelmű megfeleltetés adható meg az alábbiak szerint:*

$$T \to_1 L(\alpha) = \bigcup_{(\alpha,(b;s))\in T} (b;s)\,. \tag{2.12}$$

A megfeleltetés inverzét az alábbiak szerint kapjuk.

$$L \to_2 T = \{(\alpha, \beta) : \beta \subseteq L(\alpha)\}\,. \tag{2.13}$$

Bizonyítás. Azt, hogy az $\to_1$ megfeleltetés által adott operáció függvény-lezárás lesz, lényegében a 2.2. lemmában bizonyítottuk. Most belátjuk, hogy a (2.13) által adott családok függőségi családok lesznek, azaz kielégítik a (2.3)–(2.6) feltételeket. (2.3) abból következik, hogy L kielégíti a (2.9) feltételt. Ha $\beta \subseteq L(\alpha)$ és $\gamma \subseteq L(\beta)$, akkor (2.10) és (2.11) szerint $\gamma \subseteq L(\beta) \subseteq L(L(\alpha)) = L(\alpha)$, azaz (2.4) fennál T–re.

(2.5) a definíciókból és a (2.10) tulajdonságból következik. Ha $\alpha \subseteq \gamma$, $\delta \subseteq \beta$ és $\beta \subseteq L(\alpha)$, akkor $\delta \subseteq \beta \subseteq L(\alpha) \subseteq L(\gamma)$, azaz $(\gamma, \delta) \in T$.

Végül (2.6) bizonyítása egy kicsit bonyolultabb gondolatmenetet igényel. Mivel $\alpha \cup \gamma$ létezik és koherens, $L(\alpha \cup \gamma)$ is létezik és koherens. A parciális függvények uniója definíciójából könnyen látható, hogy ha két parciális függvénynek létezik közös majoránsa, akkor az uniójuk is létezik, sőt ha a majoráns koherens, akkor az unió is az lesz. Márpedig ebben az esetben $\beta \subseteq L(\alpha) \subseteq L(\alpha \cup \gamma)$ és $\delta \subseteq L(\gamma) \subseteq L(\alpha \cup \gamma)$, azaz $\beta \cup \delta$ létezik, koherens, és természetesen része $L(\alpha \cup \gamma)$–nak, ami éppen (2.6) konklúziója.

Azt kell még belátnunk, hogy a két megfeleltetés egymás inverze. Tekintsük először a T_1 függőségi családhoz hozzárendelt L függvény-lezárásból kapott T_2 függőségi családot. Ha egy (α, β) pár eleme T_1–nek, akkor β minden attributuma a β

által ott felvett értékkel szerepel $L(\alpha)$ (2.12) előállításában T_1 (2.5) tulajdonsága szerint, azaz $\beta \subseteq L(\alpha)$, ami maga után vonja, hogy $(\alpha,\beta) \in T_2$. Fordítva, ha egy (α,β) pár eleme a T_2 családnak, akkor $\beta \subseteq L(\alpha)$, azaz β minden b attributumára és a β által ott felvett s értékre $\alpha \rightarrow (b;s)$, azaz a T_1 család (2.6) tulajdonsága szerint $\alpha \rightarrow \beta$, $(\alpha,\beta) \in T_1$.

Végül legyen az L_1 lezárásból (2.13) szerint kapott T családhoz rendelt lezárás L_2 és α egy tetszőleges koherens parciális függvény. $L_1(\alpha)$ minden b attributumára és az ott felvett s értékére $(b;s) \subseteq L_1(\alpha)$, így $(\alpha,(b;s)) \in T$. Ebből viszont a (2.12) definíció szerint $L_2(\alpha) = L_1(\alpha)$. □

Az eddigiekben tehát láttuk, hogy a parciális függőségek egyértelműen jellemezhetők a megfelelő függvény-lezárással (2.2. lemma), vagy a (lényegében a függőséggel azonos) megfelelő függőségi családdal (2.4. tétel). A függvény-lezárás nyilván egyszerűbb struktúra, mint a függőségi család, így ennek vizsgálatával foglalkozunk a fejezet hátralevő részében. Ugyanakkor megjegyezzük, hogy nem minden, a parciális függvényeken értelmezett függvény-lezárás (és így az annak megfelelő parciális függőség) áll elő mint valamely adatbázishoz tartozó függvény-lezárás. Ennek a kérdésnek a közelebbi vizsgálatára a 3. fejezetben térünk ki.

Egy α parciális függvény (egy adott adatbázisban, ill. annak L függvény-lezárása szerint) *zárt*, ha $\alpha = L(\alpha)$. A funkcionális függőségek esetében az Ω zárt részhalmazai egyértelműen meghatározzák a lezárást ill. a függőséget [3]. Hasonló állítás igaz a parciális függőségek esetére is.

2.5. TÉTEL. *Legyen $\mathcal{G}$ ugyanazon Ω alaphalmazon értelmezett parciális függvények egy családja. Ekkor $\mathcal{G}$ pontosan akkor lesz egy valamely L függvény-lezáráshoz tartozó zárt parciális függvények családja ha*

(2.14) *minden $\alpha \in \mathcal{G}$-re létezik egy olyan $\gamma \in \mathcal{G}$, amelyre $\alpha \subseteq \gamma$, $\gamma = \{c_1,\ldots,c_n;\ldots\}$ és $\{c_1,\ldots,c_n\} = \Omega$;*

(2.15) *$\alpha,\beta \in \mathcal{G}$-ből következik, hogy $\alpha \cap \beta \in \mathcal{G}$.*

Bizonyítás. Először lássuk be, hogy az adott függvény-lezáráshoz tartozó zárt függvények családja kielégíti (2.14) és (2.15)-öt. (2.2) szerint minden α-ra, és persze így minden zárt α-ra is létezik egy (2.14)-et kielégítő γ. (2.9) szerint erre a γ-ra $\gamma \subseteq L(\gamma)$, de persze $L(\gamma) \subseteq \gamma$ is fennáll, így γ zárt.

Legyenek most α és β zárt parciális függvények. Ekkor $\alpha = L(\alpha)$ és $\beta = L(\beta)$. $\alpha\cap\beta \subseteq \alpha$, és így (2.10)-ből következik, hogy $L(\alpha\cap\beta) \subseteq L(\alpha) = \alpha$. Hasonlóan látható be, hogy $L(\alpha \cap \beta) \subseteq L(\beta) = \beta$, és így a parciális függvények metszete definíciójából könnyen láthatóan $L(\alpha \cap \beta) \subseteq \alpha \cap \beta$. Ennek a fordítottja következik (2.9)-ből, így $L(\alpha \cap \beta) = \alpha \cap \beta$, azaz $\alpha \cap \beta$ is zárt parciális függvény az adott lezárásban.

Legyen most $\mathcal{G}$ egy tetszőleges parciális függvénycsalád, amely kielégíti (2.14) és (2.15)-öt. Meg kell adnunk egy olyan L függvény-lezárást, amelyben a zárt függvények pontosan $\mathcal{G}$ elemei. Legyen a függvény-lezárás azokon a parciális függvényeken értelmezve, amelyeket $\mathcal{G}$ valamely maximális tagja (azaz olyan tagja, amelynek az értelmezési tartománya Ω) tartalmaz. Ezekre a parciális függvényekre legyen

$$L(\alpha) = \bigcap_{\alpha \subseteq \beta \in \mathcal{G}} \beta .$$

L értelmezési tartományának választása következtében a fenti metszet nem üres. (2.9) és (2.10) tulajdonságok a definícióból következnek. Ha $\alpha \in \mathcal{G}$, akkor α szerepel a metszetben, és a metszet minden más tagja nagyobb α–nál, tehát $L(\alpha) = \alpha$. (2.15) szerint minden α–ra $L(\alpha) \in \mathcal{G}$, tehát $L(L(\alpha)) = L(\alpha)$, azaz (2.11) is fennál, L egy függvény-lezárás. Azt kell még belátni, hogy egy α parciális függvény pontosan akkor zárt, ha benne van $\mathcal{G}$–ben. Mint láttuk, a $\mathcal{G}$ α elemeire $\alpha = L(\alpha)$ igaz, így ezek zártak. A zárt elemekre viszont $\alpha = L(\alpha)$, és mivel $L(\alpha)$ benne van $\mathcal{G}$–ben, így α is. $\square$

A fenti tételben a (2.15) tulajdonságot úgy is megfogalmazhatjuk, hogy $\mathcal{G}$ zárt a metszetre. A (2.14) és (2.15) tulajdonságokat kielégítő parciális függvény családokat *parciális metszet-félhálóknak* fogjuk nevezni.

Láttuk, hogy a zárt parciális függvények családja mindig parciális metszet-félhálókat alkotnak. A függvény-lezárásból nyilván egyértelműen kapható meg a metszet-félháló. Bebizonyítjuk ennek az ellenkezőjét is.

2.6. TÉTEL. *Az ugyanazon alaphalmazon értelmezett $\mathcal{G}$ parciális metszet-félhálók és L függvény-lezárási operációk között egy egyértelmű megfeleltetés adható meg az alábbiak szerint:*

(2.16)
$$\mathcal{G} \longrightarrow_1 L(\alpha) = \bigcap_{\alpha \subseteq \beta \in \mathcal{G}} \beta \,.$$

A megfeleltetés inverzét az alábbiak szerint kapjuk.

(2.17)
$$L \longrightarrow_2 \mathcal{G} = \{\alpha : L(\alpha) = \alpha\}\,.$$

Bizonyítás. Már beláttuk az előző tételben, hogy a (2.17) definíció egy parciális metszet-félhálót, a (2.16) pedig egy függvény-lezárási operációt hoz létre és (2.17) megadja az összes parciális metszet-félhálót. Lássuk be, hogy $\longrightarrow_2$ injektív. Legyen L_1 és L_2 két különböző függvény-lezárás ugyanazon az alaphalmazon és $\mathcal{G}_1$ és $\mathcal{G}_2$ a megfelelő parciális metszet-félhálók. Akkor létezik egy α parciális függvény, amelyre vagy $L_1(\alpha)$ és $L_2(\alpha)$ különbözők, vagy pedig csak az egyik van definiálva. Ha $L_1(\alpha)$ létezik és $L_2(\alpha)$ nem, akkor az előbbi nem lehet eleme $\mathcal{G}_2$–nek, hisz akkor L_2–ben lenne egy zárt parciális függvény, ami tartalmazza α–t, és így $L_2(\alpha)$ is értelmezve lenne. Tehát ekkor $L_1(\alpha) \in \mathcal{G}_1$ és $L_2(\alpha) \notin \mathcal{G}_2$, így $\mathcal{G}_1 \neq \mathcal{G}_2$. Ha mind $L_1(\alpha)$, mind $L_2(\alpha)$ léteznek, csak különbözőek, akkor létezik egy $(b;s)$ egyetlen helyen értelmezett parciális függvény, amelyet $L_1(\alpha)$ és $L_2(\alpha)$ pontosan egyike tartalmaz. Legyen $(b;s) \subseteq L_1(\alpha)$ és $(b;s) \not\subseteq L_2(\alpha)$. Ekkor (2.10) következményeként, mivel

$\alpha \subseteq L_2(\alpha)$, $L_1(L_2(\alpha))$ tartalmazza $L_1(\alpha)$–t, tehát $L_1(L_2(\alpha))$ tartalmazza $(b; s)$–et. Ugyanakkor $L_2(\alpha)$ létezik, így (2.11) szerint $L_2(L_2(\alpha)) = L_2(\alpha)$ is létezik, de ez már nem tartalmazza $(b; s)$–et. Tehát az $L_2(\alpha)$ parciális függvény lezártja L_1–ben nagyobb $L_2(\alpha)$–nál, azaz $\mathcal{G}_1$ nem tartalmazza $L_2(\alpha)$–t, amit viszont $\mathcal{G}_2$ nyilván tartalmaz.

Már csak azt kell belátni, hogy a (2.16) szerinti operáció a (2.17) szerintinek az inverze. Legyen az L_1 függvény-lezáráshoz rendelt $\mathcal{G}_1$ parciális metszet-félhálóból kapott lezárás L_2 és α egy tetszőleges parciális függvény az adott alaphalmazon. Ha $L_1(\alpha)$ adott, akkor $L_1(\alpha) \in \mathcal{G}$, és így (2.16) szerint $L_2(\alpha)$ is definiálva van, sőt az előző tételben bizonyítottak szerint egyenlő is $L_1(\alpha)$–val. Ha viszont $L_2(\alpha)$ adott, akkor $\mathcal{G}$ tartalmaz α–t tartalmazó parciális függvényeket, azaz lezárása definiált L_1–ben is. Viszont láttuk már, hogy $L_1(\alpha)$ létezéséből következik, hogy $L_1(\alpha) = L_2(\alpha)$.

□

Az a tény, hogy a zárt halmazok rendszere egyértelműen leírja a lezárási operációt és így a függőséget, nagy mértékben lecsökkentheti az információ tárolására igénybe vett helyet és egyszerűsítheti az információtárolás struktúráját a funkcionális függőségek esetében.

Ellentétben ezzel, a parciális függőségek esetén a zárt halmazok rendszere egy adott adatbázisra nemhogy egyszerűbb (és így egy kicsit kevesebb, de sok esetben még mindig elegendő információt hordozó) struktúrát alkotnak, hanem bonyolultabbat. Gondoljunk csak arra, hogy az adatbázis minden sora szükség szerint egy zárt parciális függvényt ad (lényegében (2.14)), de ugyanakkor még rengeteg más zárt parciális függvényünk lehet. Tehát egy sokkal bővebb struktúránk van, amely természetesen nem hordozhat több információt, mint az eredeti adatbázis. Ezen redundancia feloldásához még visszatérünk a következő fejezetben is. Most itt azt próbáljuk megvizsgálni, hogy hogyan tudjuk a zárt függvények egy olyan részhalmazát kivenni, amely még mindig hordozza az összes információt, vagy legalábbis annak egy nagy részét.

A (2.15) tulajdondonságból következően a $\mathcal{G}$ parciális metszet-félhálót az elemei számánál sokkal kevesebb tagjával is meghatározhatjuk. Jelölje $\mathcal{M}(\mathcal{G})$ a $\mathcal{G}$ azon γ elemeinek a halmazát, amelyek nem állnak elő mint a $\mathcal{G}$ két másik tagjának a metszete, azaz amelyekre nem léteznek olyan α, β elemei a $\mathcal{G}$–nek amelyek mind különbözőek γ–tól és amelyekre $\gamma = \alpha \cap \beta$. Nyilvánvaló, hogy egy adott adatbázishoz tartozó parciális metszet-félháló esetén az adatbázis sorai, mint parciális függvények elemei lesznek $\mathcal{M}(\mathcal{G})$–nek, de valószínűleg egyéb parciális függvények is.

2.7. LEMMA. *Egy $\mathcal{G}$ parciális metszet-félhaló minden eleme előáll mint néhány (≥ 1) $\mathcal{M}(\mathcal{G})$–beli parciális függvény metszete, de $\mathcal{M}(\mathcal{G})$–nek nincs egyetlen valódi részhalmaza sem, amelyik ezzel a tulajdonsággal bírna.*

Bizonyítás. Bizonyítsuk először az első állítást. Ha van olyan eleme $\mathcal{G}$–nek, ami nem áll elő a megkívánt módon, akkor legyen α egy maximális ilyen. Az $\mathcal{M}(\mathcal{G})$ elemei előállnak, mint egyetlen $\mathcal{M}(\mathcal{G})$–beli elem metszete, így $\alpha \notin \mathcal{M}(\mathcal{G})$. Akkor viszont $\alpha = \beta \cap \gamma$, ahol β és γ α–tól különböző elemek és így annál nagyobbak is, tehát előállnak $\mathcal{M}(\mathcal{G})$-beli elemek metszeteként. β és γ metszetként való előállítását behelyettesítve a $\beta \cap \gamma$ képletbe kapjuk α egy előállítását, mint néhány $\mathcal{M}(\mathcal{G})$-beli halmaz metszete.

Ha viszont egy α parciális függvényt kitörlünk az $\mathcal{M}(\mathcal{G})$-ből, akkor éppen ez az α parciális függvény nem lesz előállítható, mint néhány $\mathcal{M}(\mathcal{G})$-beli elem metszete. Ha ugyanis előállítható lenne, akkor vegyünk egy olyan előállítást, amely minimális számú tagot tartalmaz. A tagok száma nyilván legalább 2, és így $\alpha = \beta \cap (\cap_i \beta_i)$, ahol β és az összes β_i elemek $\mathcal{M}(\mathcal{G})$-ből valók. Tehát $\beta \neq \alpha$ és a minimalitás miatt $\cap_i \beta_i \neq \alpha$, azaz α előállt mint két tőle különböző $\mathcal{G}$-beli elem metszete. Ez viszont a definíció szerint ellentmond annak a ténynek, hogy $\alpha \in \mathcal{M}(\mathcal{G})$. □

A következő két tétel leírja az $\mathcal{M}(\mathcal{G})$ családokat és azok viszonyát a parciális metszet-félhálókhoz.

2.8. TÉTEL. *Parciális függvények egy $\mathcal{Z}$ családja pontosan akkor egyezik meg egy valamely $\mathcal{G}$ parciális metszet-félhálóhoz tartozó $\mathcal{M}(\mathcal{G})$ családdal, ha kielégíti az alábbi feltételeket:*

(2.18) *minden $\alpha \in \mathcal{Z}$-re létezik egy olyan $\gamma \in \mathcal{Z}$, amelyre $\alpha \subseteq \gamma$,*
$\gamma = \{c_1, \ldots, c_n; \ldots\}$ *és* $\{c_1, \ldots, c_n\} = \Omega$;

(2.19) *ha* $\alpha = \bigcap_{i=1}^{r} \alpha_i$, $(r \geq 1)$, $\alpha, \alpha_1, \ldots \alpha_r \in \mathcal{Z}$, *akkor* $\alpha = \alpha_i$ *valamely i-re.*

Bizonyítás. Legyen először $\mathcal{G}$ egy tetszőleges metszet-félháló. Akkor $\mathcal{M}(\mathcal{G})$ (2.18) tulajdonsága következik a (2.14)-ből. (2.19)-hez legyen $\alpha = \bigcap_{i=1}^{r} \alpha_i$, $(r \geq 1)$, $\alpha, \alpha_1, \ldots, \alpha_r \in \mathcal{M}(\mathcal{G})$. Ekkor $\alpha = (\alpha_1 \cap \cdots \cap \alpha_{r-1}) \cap \alpha_r$, és itt $\alpha, \alpha_1, \ldots, \alpha_r$, $\alpha_1 \cap \cdots \cap \alpha_{r-1} \in \mathcal{G}$, valamint a minimalitás miatt $\alpha_1 \cap \cdots \cap \alpha_{r-1} \neq \alpha$. Tehát, mivel $\alpha \in \mathcal{M}(\mathcal{G})$, $\alpha_r = \alpha$.

Tegyük most fel, hogy parciális függvények egy családja kielégíti a (2.18) és (2.19) definíciókat. Konstruáljunk egy olyan $\mathcal{G}$ parciális metszet-félhálót, amelyre $\mathcal{M}(\mathcal{G}) = \mathcal{Z}$. Legyenek $\mathcal{G}$ elemei az összes $\mathcal{Z}$-ből formálható metszetek. Ahhoz, hogy az így kapott struktúra egy parciális metszet-félháló legyen, azt kell belátnunk 2.7 lemma szerint, hogy $\mathcal{Z}$ egyetlen eleme sem áll elő, mint tőle különböző két $\mathcal{G}$-beli elem metszete, de a $\mathcal{G} - \mathcal{Z}$-beli elemeket elő tudjuk állítani így.

Tegyük fel először, hogy $\alpha \in \mathcal{Z}$, $\alpha = \beta \cap \gamma$, $\beta, \gamma \in \mathcal{G}$. A definíció szerint β és γ előállnak, mint $\mathcal{Z}$-beli elemek metszetei, tehát α-ra is kapunk egy ilyen előállítást. Viszont minden ilyen előállításban (2.19) szerint az egyik tag α, tehát β és γ egyike is α kell hogy legyen.

Tegyük most fel, hogy $\alpha \in \mathcal{G} - \mathcal{Z}$. Akkor α-t felírhatjuk, mint $\alpha = \alpha_1 \cap \cdots \cap \alpha_s$; $s \geq 2$, $\alpha_1, \ldots, \alpha_s \neq \alpha$. Tegyük fel hogy a felírásban s-t minimálisnak választottuk. Akkor $\alpha_1 \cap \cdots \cap \alpha_{s-1} \neq \alpha$, és így $\alpha = (\alpha_1 \cap \cdots \cap \alpha_{s-1}) \cap \alpha_s$ az α két tőle különböző $\mathcal{G}$-beli elemként való előállítása. □

A (2.18) és (2.19) tulajdonságokat kielégítő parciális függvény családot *metszet-mentes* családoknak nevezzük. Az alábbiakban bebizonyítjuk, hogy a metszet-mentes családok egyértelműen meghatározzák a parciális metszet-félhálókat.

2.9. TÉTEL. *Az ugyanazon alaphalmazon értelmezett $\mathcal{G}$ parciális metszet-félhálók és $\mathcal{Z}$ metszet-mentes parciális függvény családok között egy egyértelmű megfeleltetés adható meg az alábbiak szerint:*

$$\mathcal{G} \to_1 \mathcal{M}(\mathcal{G}). \tag{2.20}$$

A megfeleltetés inverzét az alábbiak szerint kapjuk.

$$\mathcal{Z} \to_2 \mathcal{G} = \{\alpha_1 \cap \dots \cap \alpha_r, \quad r \geq 1, \ \alpha_1, \dots, \alpha_r \in \mathcal{Z}\}. \tag{2.21}$$

Bizonyítás. A 2.6 tétel bizonyításához hasonlóan először azt látjuk be, hogy a $\to_1$ hozzárendelés injektív (azt már láttuk az előző lemmában, hogy az értéke egy metszet-mentes család lesz). Legyenek $\mathcal{G}_1$ és $\mathcal{G}_2$ két különböző parciális metszet-félháló és tegyük fel, hogy $\alpha \in \mathcal{G}_1$ de $\alpha \notin \mathcal{G}_2$. Akkor a 2.7 lemma szerint α előáll, mint néhány $\mathcal{M}(\mathcal{G}_1)$–beli parciális függvény metszete, de nem állhat elő, mint néhány $\mathcal{M}(\mathcal{G}_2)$–beli parciális függvény metszete. Így tehát $\mathcal{M}(\mathcal{G}_1) \neq \mathcal{M}(\mathcal{G}_2)$.

Azt is láttuk már a 2.8 tételben, hogy a $\to_1$ operáció értékkészlete az összes metszet-mentes család. Már csak azt kell belátni, hogy (2.20) inverze (2.21). Rendelje hozzá $\mathcal{M}(\mathcal{G}_1)$–hez $\to_2 \mathcal{G}_2$–t. Ha egy α parciális függvény eleme $\mathcal{G}_1$–nek, akkor az $\mathcal{M}(\mathcal{G})$ definíciója és a 2.7 lemma szerint előáll, mint néhány $\mathcal{M}(\mathcal{G}_1)$–beli parciális függvény metszete, így $\alpha \in \mathcal{G}_2$ is fennáll. Ha viszont $\alpha \in \mathcal{G}_2$, akkor lényegében a 2.8. tétel bizonyításának a második felében leirtak szerint $\alpha \in \mathcal{G}_1$ is, azaz $\mathcal{G}_1 = \mathcal{G}_2$. □

3. A parciális függőségek realizálása

Az eddigiek szerint a parciális függvények metszet-mentes családjai egyértelműen leírják a parciális függvény-lezárásokat ill. a parciális függőségeket. Mást nem is várhatunk, hiszen (2.18) és a 2.9. tétel azt is jelentik, többek között, hogy egy adatbázis minden sora szerepel mint parciális függvény az adatbázis által meghatározott metszet-mentes családban. Ugyanez az állítás viszont már nem igaz az adatbázisok és metszet-mentes családok viszonyára. Noha a 2. fejezetben bizonyítottuk, hogy egy adatbázis egyértelműen meghatározza a metszet-mentes családot, ennek fordítottjából még csak annyit láttunk, hogy a metszet-mentes családhoz egyértelműen tartozik egy parciális függőségi család, vagy még pontosabban egy függvény-lezárás. Azt nem, hogy találunk egy megfelelő adatbázist is. Persze ha találunk egyet, akkor az egyértelmű, hiszen sorai a fentiek szerint adottak. (A továbbiakban is feltesszük, hogy a parciális függvényeink értelmezési tartománya mindig része ugyanannak az Ω alaphalmaznak.)

Azonban nem mindig létezik ilyen adatbázis. Legyenek a parciális függvények metszet-mentes halmazának elemei $(a, b, c; p, q, r)$ és $(a, b; p, q)$. Az ehhez tartozó adatbázis csak az (a, b, c) attributum-halmazú és az egyetlen (p, q, r) sorból álló adatbázis lehetne. Ehhez viszont egy olyan függvény lezárás tartozik, aminek a parciális függvényekből álló metszet-mentes halmaza csak az egyetlen $(a, b, c; p, q, r)$ függvényből áll.

A fentebbi példában az okozta a bajt, hogy volt olyan nem Ω-n értelmezett elemünk, amely nem állt elő metszetként. Belátjuk, hogy ez soha nem fordulhat elő.

3.1. TÉTEL. *Legyen $\mathcal{G}$ egy adatbázis zárt parciális függvényei által alkotott parciális metszet-félháló. Akkor $\mathcal{G}$ minden olyan eleme, amely nem az Ω-n van értelmezve, előáll, mint két, tőle különböző $\mathcal{G}$-beli elem metszete.*

Bizonyítás. Legyen $\{a_1, \ldots, a_k\}$ egy valódi részhalmaza Ω-nak és $\alpha = (a_1, \ldots, a_k;\ r_1, \ldots, r_k)$ egy zárt parciális függvény, azaz $\mathcal{G}$ egy eleme. Először belátjuk, hogy α két tőle különböző $\mathcal{G}$-beli parciális függvénynek is a (valódi) része. Legyen $c \in \Omega$, de $c \notin \{a_1, \ldots, a_k\}$. Az, hogy az $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ parciális függvény zárt azt jelenti, hogy létezik két olyan s, t adat és az adatbázisnak két olyan sora, amelyek a $(a_1, \ldots, a_k, c;\ r_1, \ldots, r_k, s)$ ill. a $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, t)$ parciális függvényeket tartalmazzák. Legyen $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, s)$ lezártja β és $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, t)$ lezártja γ. Ekkor β és γ valóban tartalmazza α-t, tehát $\beta \cap \gamma \supseteq \alpha$.

Legyen most β és γ két olyan, az α-tól különböző zárt (azaz $\mathcal{G}$-beli) parciális függvény, amelyre $\beta \cap \gamma \supseteq \alpha$ és az $\text{ÉT}(\beta \cap \gamma) - \text{ÉT}(\alpha)$ halmaz nagysága minimális. Ekkor persze a (2.15) tulajdonság szerint $\beta \cap \gamma$ is $\mathcal{G}$-beli. Tegyük fel, hogy a $\text{ÉT}(\beta \cap \gamma) - \text{ÉT}(\alpha)$ halmaznak van legalább egy, mondjuk c eleme, és a $\beta \cap \gamma$ parciális függvény s értéket vesz fel a c helyen. Az $\alpha = (a_1, \ldots, a_k; r_1, \ldots, r_k)$ parciális függvény zárt, így az adatbázisnak létezik egy olyan sora, ami tartalmazza α-t, de a c helyen egy más, mondjuk t értéket vesz fel. Legyen δ az $(a_1, \ldots, a_k, c; r_1, \ldots, r_k, t)$ parciális függvény lezártja. Ekkor $\delta \supseteq \alpha$, tehát $(\beta \cap \gamma) \cap \delta \supseteq \alpha$. Ugyanakkor a $(\beta \cap \gamma)$ és a δ parciális függvények különböző értékeket vesznek fel a c helyen, így c már nem eleme $(\beta \cap \gamma) \cap \delta$ értékkészletének, azaz $\text{ÉT}(\beta \cap \gamma) - \text{ÉT}(\alpha)$ szigorúan tartalmazza $\text{ÉT}\{(\beta \cap \gamma) \cap \delta\} - \text{ÉT}(\alpha)$-t ami ellentmondás. Így az $\text{ÉT}(\beta \cap \gamma) - \text{ÉT}(\alpha)$ halmaz üres kell hogy legyen, azaz $\beta \cap \gamma = \alpha$. Tehát α-t valóban előállítottuk, mint két, tőle különböző $\mathcal{G}$-beli parciális függvény metszetét. □

Tehát egy adott adatbázis esetén az $\mathcal{M}(\mathcal{G})$ halmaz minden eleme az Ω-n kell, hogy értelmezett legyen, azaz $\mathcal{M}(\mathcal{G})$ kielégíti az alábbi feltételt:

(3.1) minden $\alpha \in \mathcal{M}(\mathcal{G})$-re $\alpha = \{c_1, \ldots, c_n; \ldots\}$ és $\{c_1, \ldots, c_n\} = \Omega$.

Parciális függvények ilyen tulajdonságú halmazát a továbbiakban parciális függvények *nagy* halmazának fogjuk nevezni. A nagy halmazok már kielégítik azt az elvárást, hogy realizálhatók lesznek (természetesen egyértelműen) adatbázisok parciális függőségi relációival.

3.2. LEMMA. *Legyen M a parciális függvények egy adott $\mathcal{Z}$ nagy halmazához tartozó azon adatbázis (mátrix) amelynek attributumai (oszlopai) az Ω elemei és amelynek sorai a $\mathcal{Z}$-beli (és így tehát a teljes Ω halmazon értelmezett) parciális függvények. Akkor az M adatbázishoz a 2. fejezetben leírtak szerint hozzárendelt* *metszet-mentes család éppen $\mathcal{Z}$; vagy ami ezzel ekvivalens, az M által meghatározott parciális függőségek rendszere (függőségi család) megegyezik a $\mathcal{Z}$-hez egyértelműen tartozó függőségi családdal.*

Bizonyítás. A bizonyításhoz meg kell gondolni, hogy mit jelent az $\alpha \longrightarrow \beta$ reláció M-ben, ill. a parciális függvények $\mathcal{Z}$ nagy halmaza szerint. A parciális függőség definíciójából jól látható, hogy M-ben $\alpha \longrightarrow \beta$ pontosan akkor áll fenn, ha M minden olyan sora, amely tartalmazza α-t, tartalmazza β-t is.

Tegyük most fel, hogy valamely lezárási operáció (és így a hozzá tartozó parciális metszet-félháló és a parciális halmazok nagy halmaza — mivel már csak olyan lezárási operációkat tekintünk, amelyekhez tartozó metszet-mentes halmazok „nagyok" is egyben) szerint $\alpha \longrightarrow \beta$, azaz $\beta \subseteq L(\alpha)$. Ez a félhálóban éppen azt jelenti, hogy minden olyan félhálóbeli (azaz zárt) parciális függvény, amely tartalmazza α-t, tartalmazza β-t is. A parciális függvények $\mathcal{Z}$ nagy halmazára pedig ez éppen azt jelenti, hogy minden $\mathcal{Z}$-beli parciális függvény, ami tartalmazza α-t, tartalmazza β-t is. Viszont a $\mathcal{Z}$ parciális függvényei éppen az adott adatbázis sorai.

A fentebb leírtak tehát azt jelentik, hogy az így definiált M mátrixból kapott parciális függőség megegyezik a $\mathcal{Z}$ nagy halmazhoz tartozó parciális függőséggel. □

Az eddigiekben a parciális függőségek által meghatározott olyan következményekkel ill. struktúrákkal foglalkoztunk, amelyek maguk a parciális függőséggel ill. az azt megadó adatbázissal ekvivalensek voltak. A továbbiakban azt vizsgáljuk meg néhány esetben, hogy mi történik akkor, ha csak gyengébb struktúrákat veszünk figyelembe.

A 2. fejezet végén vagy az ezen fejezet elején említett redundanciát elkerülhetnénk, ha $\mathcal{M}(\mathcal{G})$-ből kitörölnénk a maximális elemeket (azaz azokat, amelyek az Ω alaphalmazon vannak értelmezve). Ekkor azonban egy üres parciális függvény halmazt kapnánk.

Ezen probléma kiküszöbölésére egy $\mathcal{G}$ parciális metszet-félhálóhoz megadhatunk egy másik struktúrát is oly módon, hogy először töröljük el $\mathcal{G}$-nek az Ω-n értelmezett függvényeit (nevezzük az így kapott parciális függvény halmazt $\mathcal{G}'$-nek), és utána a maradékból vesszük azokat, amelyek nem állnak elő mint két másik, tőle különböző $\mathcal{G}'$-beli függvény metszete. Nevezzük az így kapott parciális függvény családot $\mathcal{M}_1(\mathcal{G})$-nek. Könnyen látható, hogy $\mathcal{M}_1(\mathcal{G})$-ből egyértelműen megkapható az összes, nem Ω-n értelmezett $\mathcal{G}$-beli parciális függvény. Sőt, nem nehéz meggondolni, hogy a $\mathcal{G}'$ és az $\mathcal{M}_1(\mathcal{G})$ viszonyára szinte szóról szóra megismételhetők a 2.8. és 2.9. tételek, azzal a különbséggel, hogy 2.8.-ból törölni kell a (2.18) feltételt.

Tehát egy adott adatbázishoz újabb struktúrát rendeltünk, amely esetenként jól használható az adatbázis leírásához. Láttuk, hogy adott $\mathcal{G}$ metszet-félháló esetén az $\mathcal{M}_1(\mathcal{G})$ parciális függvény halmaz kielégíti a (2.19) feltételt. Azonban az $\mathcal{M}_1(\mathcal{G})$ parciális függvény halmazra már nem lesz igaz az a tulajdonság, hogy $\mathcal{M}_1(\mathcal{G})$ minden (a tartalmazásra nézve) nem maximális halmaza előáll, mint két tőle különböző

$\mathcal{M}_1(\mathcal{G})$-beli parciális függvény metszete. Ezt a következő adatbázis mutatja:

attributumok:	a	b	c
első sor	p	r	s
második sor	p	q	t
harmadik sor	p	q	u

Itt a zárt parciális függvények a sorokon kivül $(a, b; p, q)$ és $(a; p)$, ahol az utóbbi nem maximális, de mégsem metszete két másiknak.

Persze az $\mathcal{M}_1(\mathcal{G})$ halmaz kielégít még egy triviális feltételt, nevezetesen a (3.1) feltétel ellentetjét, azaz egyetlen $\mathcal{M}_1(\mathcal{G})$-beli parciális függvény sem lehet az Ω halmazon értelmezve:

(3.2) minden $(a_1, \ldots, a_k; \ldots) \in \mathcal{M}_1(\mathcal{G})$-re $\{a_1, \ldots, a_k\}$ valódi része Ω-nak.

Kérdés, hogy a (2.19) és (3.2) feltételek jól jellemzik-e ezeket a halmazokat?

3.3. TÉTEL. *Minden olyan $\mathcal{Z}$ parciális függvény halmazhoz, amely kielégíti a (2.19) és (3.2) feltételeket, létezik egy adatbázis, amelyhez tartozó $\mathcal{G}$ parciális metszet-félhálóra $\mathcal{M}_1(\mathcal{G}) = \mathcal{Z}$.*

Bizonyítás. Azt bizonyítjuk be, hogy létezik parciális függvények egy olyan $\mathcal{G}$ metszet-félhálója, amelyre $\mathcal{M}(\mathcal{G})$ nemcsak metszet-mentes, hanem „nagy" is, és $\mathcal{M}_1(\mathcal{G}) = \mathcal{Z}$. Ehhez akkor 3.2. lemma szerint létezik egy őt realizáló adatbázis, amely az itteni céloknak is meg fog felelni.

A $\mathcal{Z}$ minden egyes α parciális függvényéhez felveszünk a leendő adatbázisban két, addig még egyáltalán nem használt elemet, mondjuk a–t és b–t. Ezek után α–hoz definiálunk két, a teljes Ω halmazon értelmezett parciális függvényt: mindkettő tartalmazza α–t, és az α–n kivüli helyeken az egyik a, a másik pedig b értékeket vesz fel. Tekintsük azt az $\mathcal{N}$ „nagy" halmazt, amelyet a $\mathcal{Z}$ összes eleméhez ily módon felvett parciális függvények (és csak azok) alkotnak. Az α–hoz felvett párok metszete így α lesz, viszont könnyen láthatóan akármely más, legalább kéttagú $\mathcal{N}$–beli metszet előáll, mint néhány $\mathcal{Z}$–beli parciális függvény metszete. Így az $\mathcal{N}$–ből előállított $\mathcal{G}$ metszet-félhálóhoz tartozó $\mathcal{M}_1(\mathcal{G})$ halmaz valóban $\mathcal{Z}$ lesz. □

A dolgozat hátralévő részében még egyetlen problémát fogunk megvizsgálni. Számos dolgozatban (pl. [3,4]) megtalálható a 2.1. lemma változata funkcionális függőségekre:

3.4 LEMMA. *Ha egy Ω attributum-halmazon értelmezett adatbázisban $A, B, C, D \subseteq \Omega$ és $\rightarrow$ jelöli a funkcionális függőséget, akkor*

(3.3) $$A \rightarrow A;$$

(3.4) $A \rightarrow B$ és $B \rightarrow C$-ből következik, hogy $A \rightarrow C$;

(3.5) $A \subseteq C$, $D \subseteq B$ és $A \rightarrow B$-ből következik, hogy $C \rightarrow D$;

(3.6) $A \to B$ és $C \to D$-ből következik, hogy $A \cup C \to B \cup D$. □

Ellentétben a parciális függőségek esetével, ennek a lemmának a fordítottja is igaz. Az Ω részhalmazaiból képzett párok egy halmazát *teljes családnak* nevezzük, ha az kielégíti a (3.3)–(3.6) feltételeket.

3.5. TÉTEL. [3] *Legyen $\mathcal{T}$ egy az Ω-n értelmezett teljes család. Akkor létezik egy adatbázis, amelynek az Ω az attributum halmaza és amelyben a (funkcionális) függősége rendszere megegyezik $\mathcal{T}$-val.* □

Mi most itt azt vizsgáljuk meg, hogy a parciális függőségek rendszere milyen struktúrát generál az attributumok halmazán. A továbbiakban egy az Ω attributum halmazon adott adatbázis esetén ha $A, B \subseteq \Omega$, akkor $A \to B$ alatt azt értjük, hogy létezik az adatbázishoz tartozó két olyan α és β parciális függvény, hogy α értelmezési tartománya A, β értelmezési tartománya B és $\alpha \to \beta$. $A \to B$-t úgy mondjuk, hogy B parciálisan függ A-tól.

3.6. LEMMA. *Ha egy adott adatbázis esetén A, B, C, D az Ω négy részhalmaza, akkor (3.3) és (3.5) fennáll rájuk.*

Bizonyítás. (3.3) nyilvánvalóan következik (2.3)-ból. (3.5) pedig (2.5)-nek és annak a ténynek a következménye, hogy ha α, β két parciális függvény és $\alpha \subseteq \beta$, akkor $\text{ÉT}(\alpha) \subseteq \text{ÉT}(\beta)$. □

Az Ω részhalmazaiból képzett párok egy halmazát *parciálisan teljes családnak* nevezzük, ha az kielégíti a (3.3) és (3.5) feltételeket. A dolgozat utolsó eredményeként azt látjuk be, hogy a parciálisan teljes családok reprezentálhatók adatbázisokkal.

3.7. TÉTEL. *Minden, az Ω-n értelmezett parciálisan teljes családhoz létezik egy olyan adatbázis, amely a részhalmazain adott parciális függőségként az adott parciálisan teljes családot adja.*

Bizonyítás. Tekintsük Ω minden A részhalmazára azokat a tartalmazásra nézve maximális B halmazokat, amelyekre $A \to B$. Minden ilyen A, B párra vezessünk be három új értéket (mondjuk a, b, c-t) a konstruálandó adatbázisban, és konstruáljunk néhány új parciális függvényt. Legyen α értelmezési tartománya A és vegyen fel minden A-beli helyen a-t. Ha A és B különböznek, akkor legyen β értelmezési tartománya B és ugyancsak vegyen fel minden helyen a-t. Legyen végül γ és δ értelmezési tartománya Ω, vegyen fel mindkettő a B-beli helyeken a-t, a B-n kivüli helyeken pedig γ b-t és δ c-t. Ekkor $\alpha \to \beta(\alpha)$, $\gamma \to \beta$ és $\delta \to \beta$. Természetesen ezekkel a parciális függvényekkel együtt fel kell venni az általuk tartalmazottakat is. Ha ezt a műveletet elvégezzük Ω minden részhalmazára, és egy adott részhalmaz esetén az összes őt tartalmazó olyan maximális részhalmazra, amely parciálisan függ tőle, kapunk egy parciális függvényekből álló halmazt. Nem nehéz ellenőrizni, hogy ez a halmaz kielégíti a (2.1) és (2.2) feltételeket, azaz leszálló lesz. Definiáljunk ezen parciális függvények között parciális függőségeket. Ha $\alpha, \beta, \gamma, \delta$ egy a fentebb értelmezett $A \to B$ párhoz értelmezett parciális függvények, akkor legyen $\alpha \to \beta$ (ill. α, ha $A = B$), $\gamma \to \beta(\alpha)$ és $\delta \to \beta(\alpha)$. Definiáljuk még az összes olyan parciális függőséget is, amelyet definiálni kell a (2.5) szerint.

Nem nehéz látni, hogy a parciális függvényeken a fentebb értelmezett parciális függőségek által az Ω részhalmazain definiált parciális függőségek éppen az adott parciálisan teljes családnak megfelelőek lesznek. Ugyancsak rutin, habár hosszadalmas feladat annak ellenőrzése, hogy a parciális függőségek itt adott rendszere kielégíti a (2.3)–(2.6) feltételeket. Itt megjegyezzük, hogy a (2.4) és (2.6) feltételek lényegében azért lesznek igazak, mert a feltételeik az adott esetben üres állítások.

Végül az eddigiek fényében azt kell csak belátnunk, hogy az így adott parciális függőség szerint zárt parciális függvények közül azok, amelyek nem az Ω–n vannak értelmezve, előállnak, mint két másik zárt metszete. Ezt viszont a konstrukció biztosítja. Az Ω–nál kisebb halmazokon értelmezett zárt parciális függvények éppen a fenti konsrtukció β–i lesznek (ill. α–i), amik viszont minden egyes esetben előállnak, mint $\gamma \cap \delta$. □

A fentiekben megadott eredményeken kivül még számos dolgot kérdezhetük a parciális függőségről. Néhány ezek közül:

— Az előzőekben több „realizálhatósági" eredményt láttunk. Tudjuk-e ezekben valahogy a realizáló adatbázist minimalizálni?

— Mit tudunk azokról a struktúrákról mondani, amelyeket úgy kapunk, hogy a parciális függőség függvény-lezárási operációjának, vagy a zárt halmazok metszet-félhálójának ill. egy $\mathcal{M}(\mathcal{G})$ halmaznak a „vázát" vesszük, azaz tekintjük a bennük található parciális függvények értékkészleteinek a halmazát?

— Hogyan változnak a parciális függőséghez tartozó struktúrák, ha az adatbázist változtatjuk, azaz pl. eltöröljük egy sorát?

IRODALOM

[1] ARMSTRONG, W. W., "Dependency structures of data base relationships ", *Information Processing 74'* (North Holland, Amsterdam, 1974), 580–583.

[2] CODD, E. F., "A relational model of data for large shared data banks ", *Com. ACM* **13** (1970), 377–387.

[3] DEMETROVICS, J and KATONA, G. O. H., "Combinatorial problems of database models ", Coll. Math. Soc. János Bolyai, 42. Algebra, Combinatorics and Logic in Computer Science, Győr (Hungary), 1983, pp. 331–353.

[4] DEMETROVICS, J. and KATONA, G. O. H., "Extremal Combinatorial problems of database models ", Kézirat.

(Beérkezett: 1988. november 28.)
(Végleges változat: 1991. szeptember 15.)

DEMETROVICS JÁNOS
MTA SZÁMÍTÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZET
1132 BUDAPEST, VICTOR HUGO U. 18

KATONA GYULA ÉS MIKLÓS DEZSŐ
MTA MATEMATIKAI KUTATÓ INTÉZET
1053 BUDAPEST, REÁLTANODA U. 13–15

PARTIAL DEPENDENCIES IN RELATIONAL DATABASES

J. Demetrovics, G. O. H. Katona, D. Miklós

Weakening the functional dependencies defined by Amstrong we get the notion of partial dependencies defined the relational databases. We show that partial dependencies can be characterized by closure operations of the poset formed by the partial functions on the attributes of the databases. On the other hand, we give necessary and sufficient conditions so that for such a closure operation one can find on the given set of attributes a database whose partial dependencies generate the given closure operation. We characterize the dependencies defined by the set of partial dependencies on the poset of the subsets of the attributes and describe how we can realize such a dependency.