Részlegesen megfigyelt döntési folyamatok

Szakdolgozat

Készítette:
Maricza István
V. évf. matematikus
hallgató

Témavezető:
Dr. Michaletzky György
Valószínűségelméleti és Statisztika
Tanszék

Eötvös Loránd Tudományegyetem
Természettudományi Kar
1996

Kivonat

E dolgozatban a döntési folyamatok elméletében a részleges megfigyelhetőség kapcsán felmerülő speciális matematikai problémákat tárgyaljuk. Az egyszerűség kedvéért a diszkrét esetre szorítkozunk. A Bevezetésben beillesztjük a témát a matematikai szakterületek közé, majd a második fejezetben részletesen tárgyaljuk a döntési folyamatok matematikai modelljét. A következő két fejezet témája a Dinamikus Programozási módszer, illetve a DP egyenlet egyszerűsítésére szolgáló elégséges statisztikák fogalmának kifejtése. Az ötödik fejezetben alkalmazzuk az általános elveket a véges állapotterű Markov Döntési Folyamatokra, és megadunk egy megoldási algoritmust. Ezt követi annak vizsgálata, hogy hogyan hat a megfigyelés pontossága az irányítás minőségére. Végül az utolsó fejezetben kidolgozzuk a speciális folytonos állapotterű LQG probléma megoldását, kiemelve a Kálmánszűrés tulajdonságaiból adódó egyszerűsítések jelentőségét.

Tartalomjegyzék

1. Bevezetés

Az alkalmazott matematika egyik fontos feladata olyan tudományosan megalapozott módszerek és eljárások kidolgozása, melyek különböző döntések meghozatalában segítik az alkalmazókat. E módszerekre alapvetően két okból van szükség. Egyrészt *mennyiségi* problémát jelent a döntések alapjául szolgáló adatok nagy tömege és esetleges valós idejű (on- line) beérkezése, ami számítógépes feldolgozást tesz szükségessé. Másrészt a döntések szükségképpen szubjektív jellegének csökkentése végett magát a döntési folyamatot *minőségi* elemzésnek kell alávetni. Az adatredukció és a döntésanalízis problémaköreinek határán helyezkedik el a *döntési folyamatok* elmélete. Története *Wald Ábrahám* munkásságáig nyúlik vissza, aki szekvenciális módszerével először vizsgálta összefüggő statisztikai döntések sorozatát.

A döntési folyamatok elmélete interdiszciplináris terület. Több egymástól távol álló tudományág használja a döntési rendszerek paradigmáját, és e területek között kevés a kommunikáció. A közös alapot az *irányított rendszer* kibernetikai fogalma képezi.

Az *operációkutatás* és a *matematikai közgazdaságtan* köznapi értelemben vett (gazdasági) döntések meghozatalában nyújt segítséget. A tipikus feladat abból áll, hogy az erőforrásokra vonatkozó korlátozások és egyéb információk alapján meghatározunk bizonyos mennyiségeket úgy, hogy céljaink a legteljesebb mértékben beteljesüljenek. Több ilyen részfeladat összekapcsolásával eljutunk a szekvenciális döntési feladatokhoz. Ha a valóságot hűen akarjuk modellezni, többnyire nem hanyagolhatjuk el a *véletlen* szerepét a feladatban. E kérdésekbe nyújt betekintést **Hillier** és **Liebermann** könyvének 3. része ([HL86]), illetve **Prékopa András** monográfiájának ([Pré94]) XII. és XIII. fejezete.

A *matematikai statisztikában* a már említett szekvenciális módszerek segítségével pl. minőségellenőrzési feladatokat oldanak meg. A döntések a mintavétel folytatására vagy bizonyos következtetések levonására vonatkoznak. Ezen a téren utalunk **Wald** eredeti könyvére ([Wal50]). Ide sorolható az ún. *hasznossági elmélet* (utility theory — ld. **French**, [Fre88]), melynek alapgondolata a döntéshozó preferenciáit kifejező hasznossági függvény.

A mérnöki alkalmazásokból kifejlődött *irányításelmélet* célkitűzése egy analitikusan leírt fizikai rendszer jellemzőinek befolyásolása. Ezen a területen döntés helyett inkább *szabályozásról* beszélnek, ugyanis a gyakorlatban folyamatos irányításról van szó, ezért elvész a döntések diszkrét szekvenciális jellege. Ez az a terület, ahol nem lehet teljesen figyelmen kívül hagyni a *megfigyelhetőség* problémáját, ami e dolgozat központi kérdése. A témába jó bevezetést nyújt **Bagchi** könyve ([Bag93]).

Ugyancsak nagy jelentősége van a megfigyelésnek az *információelméletben*, bár itt kisebb hangsúly esik a kontrollra. A részlegesen megfigyelt folyamatmodellek nagy gyakorlati haszonnal bírnak, és az állapotbecslésre kidolgozott módszerek sikeresen alkalmazhatóak más területeken is. A rejtett Markov modellek elméleti és gyakorlati kérdéseit (pl. beszédfelismerés) mutatja be a [Rab89] összefoglaló cikk és újszerű elméleti módszerek találhatók az [ELM95] monográfiában.

Újabban érdekes alkalmazásra találtak a *mesterséges intelligencia kutatásában* bizonyos döntési folyamatok, amennyiben autonóm ügynökök (robotok) viselkedésének modellezésére használják őket. A kérdéskör egyik legizgalmasabb matematikai kérdése olyan formalizmus felállítása, mely kifejezi a robot által „kitűzött" absztrakt célokat, és összhangban van azzal az elvárással, hogy a robot birtokában legyen olyan alapvető képességeknek mint pl. a tanulás. A téma irodalmából lásd a [Sze95] cikket.

E dolgozatban megkíséreljük bemutatni azokat a speciális matematikai problémákat[1], amelyek a részleges megfigyelhetőség bevezetésével jelennek meg a döntési folyamatok analízisében. Az így nyert általánosítás nem triviális, és új irányokba viszi tovább az elméletet. A témakör jellegzetessége, hogy aránylag egyszerű elvekből bonyolult egyenletek vezethetők le, melyek ritkán oldhatók meg expliciten. Mindazonáltal arra törekedtünk, hogy rávilágítsunk a fogalmak lényegére és a formulák szerkezetére, és elkerüljük a pusztán „algebrai" számolást.

Végezetül egy gyakorlati példán szemléltetjük, hogy milyen kérdések vetődhetnek fel a döntési rendszerek elemzése során. Tekintsünk egy munkadarabokat előállító gépet. Egy elkészült darabról meg tudjuk állapítani, hogy „jó"-e, vagy „rossz", és célunk minél több jó munkadarab készítése a legkevesebb költséggel, másszóval a nettó nyereséget akarjuk optimalizálni. A mindennapi életben számtalan hasonló feladat merül fel. Tudjuk, hogy a gépben több egymástól függetlenül működő azonos alkatrész meghibásodása bizonyos, tapasztalatból ismert eséllyel rossz munkadarabot eredményezhet, ezért a gép több „belső állapotából" származhat egy rossz munkadarab. Két

[1] A teljesen megfigyelt rendszerek részletes tárgyalása megtalálható pl. a [Put90] és [Bag93] könyvekben.

alkatrész és 0.5 hibaokozási valószínűség esetén pl. a belső állapotok „két jó", „egy jó, egy rossz", „két rossz" és a rossz munkadarab észlelésének valószínűségei rendre $0, 0.5, 0.75$. Minden lépésben az alkatrészek elromolhatnak, ezért a belső állapot változik. Ha az elromlás esélye 0.1, akkor az állapotok közötti átmenetek valószínűségei a következő mátrixban foglalhatók össze:

$$\begin{bmatrix} 0.81 & 0.18 & 0.01 \\ 0 & 0.9 & 0.1 \\ 0 & 0 & 1 \end{bmatrix}$$

A fenti valószínűségek arra az esetre vonatkoznak, amikor minden munkadarab legyártását megfigyelés követi. De a munkafolyamat során nem akarunk minden darabot megvizsgálni, mert az időbe, ezért pénzbe kerül. Viszont ellenőrzés nélkül hagyva a termelést nem számíthatunk túl jó eredményre. Ha időnként szétszedjük a gépet és kicseréljük az elromlott alkatrészeket, javíthatunk a gyártási minőségen, bár ez sok pénzbe kerül. Természetesen ezek a műveletek megváltoztatják a fenti valószínűségeket[2].

Olyan precíz módszert keresünk, mellyel optimálisan egyensúlyozhatunk a minőség és a költségek ellentétes irányú elvárásai között a véletlen ellenében. A módszertől azt várjuk, hogy minden lépésben megmondja, hogy a rendelkezésre álló megfigyelések alapján milyen műveletet kell végrehajtanunk ahhoz, hogy összességében a legnagyobb profitra tegyünk szert.

[2] A részleteket illetően ld. [SS73].

2. Döntési folyamatok matematikai modellje

A továbbiakban felvázoljuk a részlegesen megfigyelt döntési folyamatok általános matematikai modelljét. Felváltva használjuk majd az operációkutatási és a rendszerelméleti terminológiát attól függően, hogy melyik a szemléletesebb. Figyelembe vesszük a különböző véletlen hatásokat, és ezáltal a valóságot jól leíró, gazdag matematikai elméletet kapunk.

Matematikai szempontból a döntési folyamatok modellje *irányított sztochasztikus dinamikai rendszer.* E definíció a következő alapelemeket foglalja magába: *állapot, átmenet, külső hatás* és *megfigyelés.*

Egy absztrakt rendszer *állapotát* egy időnek nevezett változóval paraméterezett matematikai objektumsereggel írjuk le. Az egyes paraméterértékekhez tartozó mennyiségek általában egyszerű (véletlen) változók, de előfordulnak összetettebb objektumok is (ld. a 4 fejezetet). A T paramétertérről feltesszük, hogy az $\Re$ valós számhalmaz rendezett részhalmaza.[3] Az időskála jellege szerint *diszkrét* vagy *folytonos*, nagysága szerint *véges* vagy *végtelen időhorizont*ról szokás beszélni.

Az állapot idő szerinti megváltozását jelenti az *átmenet*. Értelmezünk egy léptetőoperátort, mely az X_t állapotnak megfelelteti a rendszer változását a t időpontban jellemző X_t^+ mennyiséget. Nem egységes a terminológia a diszkrét és a folytonos esetben: az előbbi esetben, mikor $T \subset \mathbf{N}$, $X_t^+ = X_{t+1}$, míg a másodikban X_t^+ valamilyen differenciálást jelent. A továbbiakban csak a diszkrét esetet tárgyaljuk. Ez sokszor nem jelent túl nagy megszorítást, hiszen egy folytonosan változó rendszerre elég sűrűn egyenlő időközönként rátekintve az eredetitől „nem nagyon különböző” diszkrét rendszert látunk.

A rendszer dinamikai egyenletének nevezünk egy olyan összefüggést, mely az $X_t^+ = X_{t+1}$ mennyiséget és a rendszer korábbi, $\mathbf{X}_t = \{X_s : s < t, t \in T\}$ állapotait hozza kapcsolatba. Ha a rendszert befolyásoló külső hatásokkal is számolnunk kell, akkor a dinamikai egyenletnek ezt is tartalmaznia kell. Jelölje a t pillanatban kifejtett hatást U_t, illetve a korábbi hatások összességét $\mathbf{U}_t = \{U_s : s < t, t \in T\}$. *Kontrollról* beszélünk, ha U_t valamilyen módon függ a rendszer előző állapotaitól (a rendszerben *visszacsatolás* van). A dinamikai egyenlet alakja tehát

$$X_t^+ = \mathcal{F}_t(\mathbf{X}_t, \mathbf{U}_t) \tag{1}$$

[3] Szoktak vizsgálni pl. $\Re^2$- vel paraméterezett kontrollált mezőket is, ld. [ELM95]

Az irányítás szempontjából lényeges, hogy a fent említett kapcsolat milyen jellegű. Az az általános elvárás[4], hogy a döntések a rendszer megfigyeléséből származó adatokon alapuljanak. Mindazonáltal többnyire nem biztosítható, hogy a megfigyelések egyértelműen jellemezzék a rendszer belső állapotát. Mint azt az első fejezet példájában láttuk, a megfigyelések általában részlegesek és pontatlanok.

Az operációkutatásban a döntések két fő csoportját különböztetik meg: az egyik fajta döntés a rendszer megismerésére, míg a másik annak megváltoztatására irányul (ld. [Alb79]). Természetesen elképzelhetőek vegyes hatású döntések is. Az első, általánosabb típusú döntés befolyásolja a megfigyelést, ezért a t pillanatban észlelt megfigyelésre a következő összefüggés adódik:

$$Y_t = \mathcal{G}_t(\mathbf{X}_t, \mathbf{U}_t) \tag{2}$$

A fenti két egyenletben szereplő $\mathcal{F}_t$ és $\mathcal{G}_t$ véletlen függvények: minden $\mathbf{X}_t = \mathbf{x}$ és $\mathbf{U}_t = \mathbf{u}$ realizációra $\mathcal{F}_t(\mathbf{x}, \mathbf{u})$ és $\mathcal{G}_t(\mathbf{x}, \mathbf{u})$ valószínűségi változók egy $(\Omega, \mathcal{A}, \mathcal{P})$ mezőn. Megadásuk módja kétféle lehet:

- Implicit – Minden realizációra megadjuk a változók eloszlását.
- Explicit – $\mathcal{F}_t$-t és $\mathcal{G}_t$-t argumentumaik és „külső" valószínűségi változók determinisztikus függvényeként adjuk meg.

Az első eset akkor kezelhető jól, amikor ez az eloszlás véges sok pontra koncentrált. Részletesen az 5. fejezetben foglalkozunk vele.

A második esetben az (1) és a (2) egyenletek alakja a következő:

$$\begin{aligned} X_t^+ &= f_t(\mathbf{X}_t, \mathbf{U}_t, W_t) && (3) \\ Y_t &= g_t(\mathbf{X}_t, \mathbf{U}_t, V_t), && (4) \end{aligned}$$

ahol f_t és g_t tetszőleges (nemlineáris) függvények, W_t és V_t pedig valószínűségi változók. A nemlineáris kontrollfeladatok megoldása tipikusan nehéz, ezért bizonyos esetekben (pl. ha a függvények korlátosak) az állapotteret véges sok részre osztva az első típusba tartozó feladatokkal közelítik őket. Egy másik lehetséges szuboptimális megoldás sima függvények esetén a linearizálás. Erre láthatunk példát a 7. fejezetben.

[4]Kivételt képeznek pl. a mesterséges intelligenciában alkalmazott ún. heurisztikus stratégiák.

$$\begin{tikzcd}
\cdots \arrow[r] & \boxed{X_{k-1}} \arrow[r, "U_{k-1}"] \arrow[d] & \boxed{X_k} \arrow[r, "U_k"] \arrow[d] & \boxed{X_{k+1}} \arrow[r, "U_{k+1}"] \arrow[d] & \cdots \\
 & Y_{k-1} & Y_k & Y_{k+1} &
\end{tikzcd}$$

1. ábra: A döntési folyamat

Megjegyezzük, hogy ebben a modellben a véletlen mint „zavaró hatás" kettős szerepet játszik: megfosztja a döntéshozót a mindentudás és a mindenhatóság képességétől. A 6. és a 7. fejezetben látni fogjuk, hogy mi az ára az így nyert általánosságnak.

A döntési folyamat időbeli lefolyását tekintve kétféle modellt alkalmazhatunk, melyek csak az egy döntési lépést alkotó elemek sorrendjében különböznek egymástól:

1. Az **I. modell** esetében a *megfigyelés* automatikus, ezt követi egy *döntés*, ami *átmenetet* eredményez. Ez az irányításelméleti megközelítés.

2. A **II. modell**nél minden lépés elején hozunk egy döntést, ami egy átmenetet, majd egy *megfigyelést* eredményez. Ezzel a modellel az operációkutatásban találkozhatunk.

A továbbiakban az egyszerűség kedvéért csak az I. modell szerinti döntési folyamatokat vizsgálunk, de a képletek apró algebrai átalakítások után a másik esetben is érvényesek lesznek. A (2) egyenlet ezáltal az egyszerűbb

$$Y_t = \mathcal{G}_t(\mathbf{X}_t) \tag{5}$$

alakot kapja. A 2. ábra szemlélteti a folyamatot diszkrét időhorizont esetén.

Az X_t és Y_t mennyiségek akkor lesznek definiáltak az (1) és (8) egyenletek és a kezdeti feltételek által, ha ismert az U_t kontroll változók függése a döntési információtól. A t. pillanatban a döntés meghozása előtt rendelkezésre álló információ az összes eddigi megfigyelésekből álló $\mathbf{Y}_t = \{Y_s : s \leq t, t \in T\}$ vektor, illetve a korábbi döntések összessége: $\mathbf{U}_t = \{U_s : s \leq t, t \in T\}$. Bár ez utóbbi mennyiségek nem függetlenek a megfigyelésektől, mégis belevesszük őket a pillanatnyi döntési információt kifejező $Z_t = (\mathbf{Y}_t, \mathbf{U}_{t-1})$ vektorba. Döntési stratégiának nevezzük függvények egy olyan $\boldsymbol{\phi} = (\phi_t)_{t\in T}$

sorozatát, melynek t. eleme egy konkrét $z_t = (\mathbf{y}_t, \mathbf{u}_{t-1})$ információvektorhoz hozzárendeli a megfelelő u_t döntést. Elvben tekinthetnénk véletlen függvényeket is (mint ahogyan azt a teljesen megfigyelt esetben meg is teszik), de mivel a rendszerben a véletlen már többszörösen jelen van, ettől eltekintünk. Megjegyezzük, hogy a ϕ_k függvények értelmezési tartományai bővülő halmazsorozatot alkotnak: ha $\mathcal{Y}$ a lehetséges megfigyelések, $\mathcal{U}$ pedig a döntések halmaza, akkor

$$\operatorname{dom}\phi_{k+1} = \operatorname{dom}\phi_k \times \mathcal{Y} \times \mathcal{U}, \text{ és } \operatorname{dom}\phi_0 = \mathcal{Y}, \tag{6}$$

hiszen $Z_{k+1} = Z_k \cup \{Y_{k+1}, U_k\}$ és $Z_0 = Y_0$. Ezt a jelenséget az irodalom úgy említi, mint a „dimenzionalitás átka", mivel $\operatorname{dom}\phi_k$ dimenziója lineárisan nő. E problémát és megoldását a 4 fejezetben vizsgáljuk.

Az irányított sztochasztikus dinamikai rendszerek elemeinek ismertetése után rátérünk a megoldandó feladatra. A fentiekből látszik, hogy a modell akkor teljes, ha adott egy $\boldsymbol{\phi}$ döntési stratégia. A feladat az, hogy stratégiák egy megengedett köréből kiválasszuk az(oka)t, mely(ek) céljainknak leginkább megfelel(nek). A megfelelőséget különböző alkalmazásokban sokféleképpen értelmezik. A matematikai közgazdaságtan modelljeiben a kitűzött cél például a maximális nyereség. Az irányításelmélettel foglalkozó mérnök számára az a fontos, hogy a rendszer ne hagyja el a normális működési tartományt. A robotikában az lehet a cél, hogy a robot megtaláljon egy tárgyat egy bizonyos tartományban. Matematikailag ez minden esetben a $\boldsymbol{\phi}$ stratégia egy funkcionálja extrémumának megkeresésére vezethető vissza.

A modell valószínűségi jellege miatt optimalitási kritériumként különböző valószínűségek vagy (feltételes) várható értékek extrémumszámítása jöhet szóba. Egylépéses döntési feladatokra több ilyen kritérium esetén léteznek megoldási módszerek (ld. [Pré94]). A többlépéses döntési feladatokat nagyon speciális funkcionálok esetén vizsgálták, de az így feltárt matematikai struktúrának is gazdag irodalma van. White cikke ([Whi88]) összefoglalja a nemstandard kritériumokkal kapcsolatos eredményeket.

Az irodalomban vizsgált legtöbb döntési folyamat Markov-i jellegű. Ez azt jelenti, hogy az átmenet utáni állapot csak a jelenlegi állapottól *és döntéstől* függ expliciten, a korábbiaktól nem, vagyis az állapotegyenlet alakja

$$X_{t+1} = \mathcal{F}_t(X_t, U_t) \tag{7}$$

Hasonlóképpen a megfigyelési egyenlet

$$Y_t = \mathcal{G}_t(X_t) \tag{8}$$

Ez nem feltétlenül jelenti azt, hogy az (X_t) állapot–folyamat Markov–folyamat, hiszen a döntések figyelembe vehetik a korábbi megfigyeléseket. A továbbiakban csak ezt az esetet vizsgáljuk.

A fenti tulajdonság lehetővé teszi, hogy az egyes átmenetek egyedi *költségeiről* beszéljünk. Legyen $\gamma_t^u(x, y; \theta)$ a költsége a t. időpillanatban az u döntés hatására történt $x \to y$ átmenetnek, ha a megfigyelés θ. Adott x esetén a megfelelő y és θ értéke a véletlentől függ, ezért gyakran az egyszerűbb

$$\gamma_t(x, u) = E\left[\gamma_t^u(X_t, X_{t+1}; Y_t) \mid X_t = x, U_t = u\right]$$

várható (átlag-) *költséget* használják.

Minden $\boldsymbol{\phi}$ stratégia rögzített kezdeti feltételek mellett a $\{\gamma_t(X_t, \phi_t(Z_t))$ véletlen költségsorozatot generálja, ahol X_t az állapotegyenlet rekurziója által adott. Ezekből kell kikeverni egy determinisztikus valós funkcionált, ami a rendszer viselkedését jellemzi.

Véges időhorizont esetén egyszerű feltételek mellet (a költségfüggvény véges és mérhető) a

$$V(\boldsymbol{\phi}) = E\left[\sum_{k=0}^{N} \gamma_k(X_k, U_k)\right] \tag{9}$$

várható összköltség létezik és véges, ezért megfelelő aggregált célfüggvény.[5]

Ha a horizont végtelen, akkor a helyzet bonyolultabb. Alapvető könyvében ([How60]) Howard megmutatta, hogy véges állapotterű teljesen megfigyelt Markov döntési folyamatokra adott stratégia mellett a költségek összege aszimptotikusan lineárisan függ a döntési lépések számától, vagyis végtelen horizontra az aggregált összköltség végtelenhez tart. Helyette az egy lépésre eső költséget lehet használni, ami az előbbiek szerint $V_N(\phi)$ lineáris növekedési sebessége:

$$\bar{V}(\phi) := \lim_{N\to\infty} \frac{1}{N+1} \cdot V_N(\phi) \tag{10}$$

[5] Bizonyos optimalitási feladatokban (pl. a híres Szentpétervár Paradoxonban, ld. [Fel68]) a várható érték használata nem célravezető.

Az átlagköltséges modellek szabatos matematikai elemzése számos technikai problémát vet fel és a kapcsolódó numerikus módszerek is bonyolultabbak. A téma kiváló áttekinését adja az [Ara93]cikk.

Míg az átlagköltséges modellek akkor működnek jól, amikor a rendszer hosszútávú, stacionárius tulajdonságaira vagyunk kíváncsiak, a gazdasági alkalmazásokban, ahol az időben közeli, tranziens jelenségek a fontosak, a *diszkontált* (lecsengetett) modellek bizonyulnak hasznosnak. Kiindulva abból, hogy a „most elköltött pénz többet ér a jövőben elköltendőnél", a $(t+1)$.-ik lépés költségének csupán $0 < \beta < 1$ hányadát vesszük figyelembe az előző lépés költségéhez képest. Így a diszkontált összköltség

$$V_\beta(\phi) = \sum_{k=0}^{\infty} \beta^k E\left[\gamma_k(X_k, U_k)\right]. \tag{11}$$

β értékének rögzítésével megszabjuk, hogy milyen fontosságot tulajdonítunk a jövőbeli eseményeknek.

A diszkontált modell nemcsak szemléletes, hanem matematikailag is sokkal jobban kezelhető, mint az előző. A részleteket mellőzve megjegyezzük, hogy valójában az átlagköltséges kritériumra vonatkozó tételeket is a diszkontált esetből származtatják. A két kritérium cseppet sem nyilvánvaló kapcsolatára a következő Tauber típusú tétel világít rá:

Tétel: Ha $\{a_n\}$ nemnegatív számsorozat és $\beta \in (0,1)$, akkor

$$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} a_k \;&\le\; \liminf_{\beta\nearrow 1}\,(1-\beta)\sum_{n=0}^{\infty}\beta^n a_n \le \\ &\le\; \limsup_{\beta\nearrow 1}\,(1-\beta)\sum_{n=0}^{\infty}\beta^n a_n \le \limsup_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1} a_k \end{aligned}$$

Bizonyos speciális esetekben az átlagköltséges feladatok formálisan diszkontált problémákká alakíthatóak, melyek megoldását „visszaalakítva" megkapjuk az eredeti feladat megoldását (ld. [Ros70]).

$$
\begin{array}{|rcl|}
\hline
X_{k+1}^{\phi} & = & f(k, X_k^{\phi}, U_k^{\phi}, W_k), \quad X_0^{\phi} = X_0 \\
Y_k^{\phi} & = & g(k, X_k^{\phi}, V_k) \\
U_k^{\phi} & = & \phi_k(\mathbf{Y}_k^{\phi}, \mathbf{U}_{k-1}^{\phi}) \\
\hline
\phi^* & \in & \operatorname*{argmin}_{\phi} V_N(\phi) \\
\hline
\end{array}
$$

2. ábra: A nemlineáris döntési feladat

A 2. ábra táblázata összefoglalja a véges idejű nemlineáris döntési feladatot. A következő fejezetekben bevezetjük e témakör alapösszefüggését, a Bellman-egyenletet, vázoljuk az elvi megoldást, majd egy speciális esetben bemutatunk egy algoritmust.

3. A dinamikus programozási elv

Az előző fejezetben megfogalmazott optimális irányítási feladat megoldásának kulcsa az ún. **optimalitási elv.** A *Richard Bellman* által az 50-es években kimondott, széleskörűen alkalmazható megoldási módszer lényege az, hogy egy optimalizálási feladatot alkalmas módon részekre bontunk, a részfeladatokat megoldjuk, majd a részmegoldásokból összeállítjuk az egész feladat optimális megoldását. A számítástudományban számos modell született azon feladatok karakterizálására, melyekre érvényes az optimalitási elv (ld. [Hel86]). Ezek közül az egyik legfontosabb *Karp* és*Held* automataelméleti megközelítése. Ebben a formalizmusban az alkalmazhatóság kritériuma az, hogy a *diszkrét döntési probléma* reprezentálható legyen egy *monoton szekvenciális döntési folyamat*ot megoldó automatával. Ez jelen esetben annak felel meg, hogy a célfüggvényt egyedi költségek aggregációjával állítottuk elő.

A szekvenciális döntési feladatok kontextusában az optimalitási elvet szokás *dinamikus programozás*nak is nevezni. A dinamikus programozás (DP) általános elméletét illetően utalunk Bellman eredeti művére ([Bel57]), illetve Bertsekas újabb könyvére ([Ber87]).

A DP elve formálisan a következőképpen fogalmazható meg:

> Legyen $\mathcal{T}$ a feladat időhorizontja és $\mathcal{T} = \mathcal{T}_- \cup \mathcal{T}_+$ a t. pillanat általi diszjunkt felosztás:
>
> $$\mathcal{T}_- = \{s \ : \ s \in \mathcal{T}, s \leq t\} \quad \text{és} \quad \mathcal{T}_+ = \{s \ : \ s \in \mathcal{T}, s > t\}$$
>
> Jelölje $\phi[X_0, \mathcal{T}]$ a $\mathcal{T}$ horizontra vonatkoztatott probléma optimális stratégiáját az X_0 kezdeti feltétel esetén. Ekkor
>
> $$\phi[X_0, \mathcal{T}] = \phi_1 \cdot \phi_2, \text{ ahol } \phi_1 = \phi[X_0, \mathcal{T}_-] \text{ és } \phi_2 = \phi[X_t^{\phi_1}, \mathcal{T}_+]$$
>
> és a $\cdot$ operátor a stratégiák konkatenációját jelenti.

Amennyiben a horizont véges ($\mathcal{T} = \{1, \dots, N\}$), a feladatot elvben véges sok lépésben egylépéses feladatok sorozatává „tördelhetjük”. Vizsgáljuk meg azt a köztes lépést, amikor a $\{k, \dots, N\}$ szakaszból leválasztjuk az első lépést.

Jelölje

$$V_k^*(z_k) = E\left[\sum_{i=k}^{N} \gamma_i(X_i, \phi^*(Z_i)) \mid Z_k = z_k\right] \tag{12}$$

az adott szakaszra eső összköltséget, feltéve, hogy a k. lépésben a rendszer „információs állapota" (a megfigyelésekből és döntésekből álló vektor) éppen z_k és a ϕ^* optimális stratégiát alkalmazzuk, amennyiben ilyen létezik.

Mivel feltettük az optimalitást, a k. lépésben meghozandó $u = \phi_k^*(z)$ döntést úgy kell megválasztanunk, hogy hatására a hátralevő költség a lehető legkisebb legyen:

$$V_k^*(z_k) = \min_u E\left[\gamma_k(X_k,u) + \underbrace{\sum_{i=k+1}^{N} \gamma_i(X_i,\phi_i^*(Z_i))} \mid Z_k = z_k, U_k = u\right] \quad (13)$$

Felhívjuk a figyelmet, hogy e „mohó" döntés létjogosultságát egyelőre csak a szemlélet támasztja alá. Később bizonyítani fogjuk.

A $\{Z_k = z_k, U_k = u\}$ feltétel megfigyelés előtti állapotot ír le. Az Y_{k+1} új érték megfigyelésével az információs állapot Z_{k+1} lesz. Az egyenletben a megjelölt tagot az iterált feltételes várható érték tételének értelmében helyettesíthetjük a bővebb $\{Z_{k+1}\}$ feltételre vett várható értékkel, ami a következőképpen írható:

$$E\left[\sum_{i=k+1}^{N} \gamma_i(X_i,\phi_i^*(Z_i)) \mid Z_{k+1}\right] = V_{k+1}^*(Z_{k+1})$$

Visszahelyettesítve a (13) egyenletbe megkapjuk az ún. **dinamikus programozási** (vagy **Bellman-) egyenletet**:

$$V_k^*(z_k) = \min_u \left\{ E\left[\gamma_k(X_k,u) \mid Z_k = z_k\right] + E\left[V_{k+1}^*(Z_{k+1}) \mid Z_k = z_k, U_k = u\right]\right\} \quad (14)$$

Ez a hátráló rekurzív egyenlet $k = 0, 1, \ldots, N-1$ esetén érvényes. A „végfeltétel":

$$V_N^*(z_N) = \min_u E\left[\gamma_N(X_N,u) \mid Z_N = z_N\right] \quad (15)$$

A (14) egyenlet jobboldalán levő minimalizálandó mennyiséget a következő módon értelmezhetjük. Tegyük fel, hogy a k. lépésben a rendszer információs állapota z és az u döntést hozzuk. Ekkor a rendszer egy lépés utáni várható költsége pont az első tag. Ha most a rendszert az új $Z_{k+1} = z'$ állapotban megállítjuk és kiszabunk rá egy $V_{k+1}^*(z')$ nagyságú „bírságot", akkor az így keletkező várható költséget a második tag fejezi ki. Definiáljuk a

$$(T_k^u g)(z_k) = E\left[\gamma_k(X_k,u) \mid Z_k = z_k\right] + E\left[g(Z_{k+1}) \mid Z_k = z_k, U_k = u\right] \quad (16)$$

formulával az ún. *egylépéses megállítási operátort*, mely minden dom ϕ_{k+1}-en értelmezett korlátos g bírságfüggvényhez hozzárendeli a megfelelő egylépéses költségfüggvényt. Terjesszük ki ezt a definíciót úgy, hogy az u döntés helyett egy ϕ stratégiát írva a képlet minden z_k esetén a ϕ által szolgáltatott döntést használja:

$$(T_k^{\phi}g)(z_k) = (T_k^{\phi_k(z_k)}g)(z_k). \tag{17}$$

Ha $B(S)$ az S-en értelmezett korlátos függvények halmaza, akkor

$$T_k^{\phi} : B(\mathrm{dom}\,\phi_{k+1}) \longrightarrow B(\mathrm{dom}\,\phi_k)$$

Könnyen látható, hogy T_k^{ϕ} *monoton* operátor, vagyis $f \le g$ esetén $T_k^{\phi} f \le$ $\le T_k^{\phi} g$.

A (14) egyenlet alakja ezzel a jelöléssel $V_k^*(z_k) = \min\limits_u (T_k^u V_{k+1}^*)(z_k)$, illetve definiálva a

$$\phi_k'(z_k) = \operatorname*{argmin}_u (T_k^u V_{k+1}^*)(z_k) \tag{18}$$

mohó stratégiát és alkalmazva a $\min(f,g)(x) = \min(f(x), g(x))$ konvenciót, a következő formulákat kapjuk:[6]

$$V_k^* = T_k^{\phi'} V_{k+1}^*, \quad k = 0, 1, \ldots, N-1 \tag{19}$$

$$V_N^* = T_N^{\phi'} \mathbf{0} \tag{20}$$

Az előzőekhez teljesen hasonló módon (a feltételes várható érték kifejtésével) kapjuk, hogy **tetszőleges** ϕ stratégiára

$$V_k^{\phi} = T_k^{\phi} V_{k+1}^{\phi} \tag{21}$$

$$V_N^{\phi} = T_N^{\phi} \mathbf{0} \tag{22}$$

A mohó stratégia optimalitása hátráló indukcióval látható be:

$V_N^{\phi'} = T_N^{\phi'} \mathbf{0} \le T_N^{\phi} \mathbf{0} = V_N^{\phi}$, hiszen ϕ' a minimalizáló argumentumot választja.

Tegyük fel, hogy valamely $i \in \{1, 2, \ldots, N-1\}$-re már igazoltuk, hogy $V_{i+1}^{\phi'} \le V_{i+1}^{\phi}$. Ekkor

$$V_i^{\phi'} = T_i^{\phi'} V_{i+1}^{\phi'} \le T_i^{\phi} V_{i+1}^{\phi'} \le T_i^{\phi} V_{i+1}^{\phi} = V_i^{\phi},$$

ahol felhasználtuk T_i^{ϕ} monotonitását, illetve azt, hogy ϕ' a minimalizáló argumentum.

[6] $\mathbf{0}$ az azonosan nulla függvényt jelenti.

A végtelen időhorizontot vizsgálva kerülünk először szembe a *dimenzionalitás* problémájával: eddig formálisan figyelmen kívül hagyhattuk, hogy dom ϕ_k dimenziója k-ban nő, most azonban, amikor elvileg végtelen sok egyre bővülő értékkészletű operátort kell használnunk, a probléma kezelhetetlenné válik. A következő fejezet egy lehetséges megoldást szolgáltat az ún. *elégséges statisztikák* módszere által. Megelőlegezve az ott elmondottakat, úgy tekintjük, hogy az összes egyenletben Z_k helyett a k-tól független fix dimenziójú S_k mennyiség szerepel. Megvizsgáljuk, hogy ebben az esetben mi lesz az optimális stratégia.

A véges esetben a megoldás kulcsa a hátráló indukció volt. A végtelen esetben nem áll rendelkezésünkre kezdeti érték, de helyette egy „határérték" szolgáltatja majd a kiindulási alapot. Az előző fejezetben említett okok miatt diszkontált költségeket használunk. További egyszerűsítésként a költségfüggvény legyen az időtől független: $\gamma_k(X_k, U_k) = \gamma(X_k, U_k)$.

Látni fogjuk, hogy az S_k statisztika megválasztható úgy, hogy Markov-folyamat legyen. Belátható (ld. [BW90]), hogy ha létezik optimális stratégia, akkor van *stacionárius* optimális stratégia is: $\boldsymbol{\phi} = (\phi, \phi, \ldots)$. Mindezek következménye, hogy az egylépéses megállítási operátort k-tól függetlenül definiálhatjuk:

$$(T^\phi g)(s) = E\left[\gamma(X_k, u) \mid S_k = s\right] + \beta E\left[g(S_{k+1}) \mid S_k = s, U_k = \phi(s)\right] \quad (23)$$

A költségfüggvény most csak a kezdeti értéktől függ, hiszen akárhonnan indítjuk a rendszert, mindig „ugyanakkora" (∞) horizontot kell bejárnunk.

A véges esetben érvényes képletek megfelelői:

1. Ha $u \leq v$, akkor $T^\phi u \leq T^\phi v$.
2. $T^\phi V^\phi = V^\phi$ tetszőleges ϕ stratégiára
3. A DP egyenlet alakja $T^* V^* = V^*$, ahol $T^* = \min T^u$.

A véges eset rekurziói fixpont-egyenletekké alakultak, és mivel a V függvények $B(\mathcal{S})$ tere (a szuprémum-normával) teljes, ezért alkalmazható a Banach-féle fixpont- tétel:

Tétel (Banach): Ha $T : B(\mathcal{S}) \longrightarrow B(\mathcal{S})$ kontrakció, akkor $\exists! g \in B(\mathcal{S})$ úgy, hogy $Tg = g$ és $\forall h \in B(\mathcal{S}) \;:\; T^n h \to g$.

A feltétel igazolásához először belátjuk, hogy T^ϕ kontrakció. Valóban, a (23) egyenlet alakjából látszik, hogy $|T^\phi f - T^\phi g| = \sup_s |(T^\phi f)(s) - (T^\phi g)(s)| \leq \beta |f-g|$. Innen $T^* f - T^* g = \min_u T^u f - \min_u T^u g \leq T^\phi f - T^\phi g \leq \beta |f-g|$, ahol $\phi = \underset{u}{\operatorname{argmin}}\, T^u g$.

Tehát $V^* = \lim_{n\to\infty} (T^*)^n f$ és $V^\phi = \lim_{n\to\infty} (T^\phi)^n g$ tetszőleges $f, g \in B(\mathcal{S})$ függvényekre. Ezen előkészületek után belátjuk, hogy a $\phi' = \operatorname{argmin}_u T^u V^*$ mohó stratégia *optimális*, vagyis $V^{\phi'} = V^*$:

> ϕ' definíciója miatt a DP egyenlet ekvivalens alakja $T^{\phi'} V^* = V^*$. Iterációval kapjuk, hogy $(T^{\phi'})^n V^* = V^*$. De a Banach-tétel szerint a baloldal limesze $V^{\phi'} = V^*$, amit bizonyítani akartunk.

Megjegyzés: Az egylépéses megállítási operátor szemléletes jelentéséből adódik, hogy

$$V^*_{0,N}(s) = \left[(T^*)^N \mathbf{0} \right](s)$$

az optimális diszkontált költség az N hosszúságú véges szakaszon, ami a $V^* = \lim_{n\to\infty} (T^*)^n f$ képlet alapján összhangban van azzal a szemléletes képpel, miszerint a végtelen időhorizont nagy véges horizontok határesete.

Összefoglalva a döntési feladat megoldására kapott dinamikus programozási módszert, a következő két lépést különíthetjük el:

1. Felírunk egy variációs egyenletet az *optimális* költségfüggvényre az egylépéses átmenetek alapján.

2. Az extrémumot szolgáltató értéket tesszük meg optimális kontrollnak. (Mohó stratégia)

A felmerülő gyakorlati számítások kapcsán olyan jól konvergáló módszerekre van szükség, melyek kiértékelik az optimális költségfüggvényt (másszóval *értékfüggvényt*). Ilyen sztochasztikus approximációs módszerek találhatók a [Bra94] disszertációban.

4. Elégséges statisztikák

Mint azt korábban láttuk, a döntéshozáshoz szükséges információt tartalmazó $Z_k = (\mathbf{Y}_k, \mathbf{U}_{k-1})$ vektor dimenziója frissítéskor nő:

$$Z_{k+1} = (Z_k, U_k, Y_{k+1}) =: \mathcal{R}_k(Z_k, U_k, Y_{k+1})$$

Valójában ez az információ többszörösen redundáns, hiszen $U_l = \phi_l(Z_l)$. Célunk az, hogy a Z_k vektorban rejlő információt egy olyan fix, lehetőleg kis dimenziós mennyiségbe „sűrítsük", melynek ismerete elegendő a feladat megoldásához.

A matematikai statisztikában használt „elégséges statisztika" fogalmának hasonló a szerepe, ezért az elnevezést onnan kölcsönözzük. Eredeti jelentésében az S elégséges statisztika az $(\Omega, \mathcal{A}, P_\theta : \theta \in \Theta)$ statisztikai mezőn a mintának olyan függvénye, melynek ismeretében a minta feltételes eloszlása $(P_\theta(X \in A \mid S = s))$ független az ismeretlen θ paramétertől.[7] Jelen esetben az $\mathcal{S}$ halmazbeli értékű S_k statisztikák *sorozatát* nevezzük elégségesnek, ha teljesülnek a következő feltételek:

1. S_k a Z_k „minta" függvénye (**mérhetőség**)

$$S_k = S_k(Z_k)$$

2. Az új S_{k+1} érték kiszámolható az előző S_k-ból, az U_k döntésből és az Y_{k+1} megfigyelésből (**rekurzivitás**)

$$S_{k+1} = \bar{\mathcal{R}}_k(S_k, U_k, Y_{k+1})$$

3. Az $\{S_k\}_{k=0,1,\ldots}$ sorozat ismeretében meghatározható az optimális stratégia (**ekvivalencia**)

E definícióhoz a következő megjegyzéseket fűzhetjük:

- A 2. feltétel teremt összefüggést a sorozat tagjai között és biztosítja, hogy S_k-t gazdaságosan számítjuk ki az adatokból.

- Ahhoz, hogy a 3. feltételben tett kikötés teljesüljön, szükséges, hogy S_k csak a rendszer megfigyelt adataitól, nem pedig a stratégiafüggvényektől függjön.

[7] ld. [Leh83]

$$
\begin{tikzcd}
\operatorname{dom}\phi_{k+1} \arrow[dd, "S_{k+1}"'] \arrow[dr, "g"] & & \operatorname{dom}\phi_k \arrow[ll, "\mathcal{R}_k(\cdot,u,y)"'] \arrow[dd, "S_k"] \arrow[dl, "T_k^u g"'] \\
& \Re \arrow[u, phantom, "g \overset{T_k^u}{\longmapsto} T_k^u g"] \arrow[d, phantom, "\bar g \dashrightarrow \bar T_k^u \bar g"] & \\
\mathcal{S} \arrow[ur, "\bar g"] & & \mathcal{S} \arrow[ll, "\bar{\mathcal{R}}_k(\cdot,u,y)"] \arrow[ul, "\bar T_k^u \bar g"']
\end{tikzcd}
$$

3. ábra: A szeparációs elv

- A Z_k vektor triviálisan megfelel a követelményeknek.

A harmadik feltétel jelentésének tisztázásához vezessük be a következő fogalmat:

> **Definíció:** A $g \in B(\operatorname{dom}\phi_k)$ függvény *szeparálható* statisztikák egy $\{S_k\}$ elégséges sorozatára nézve, ha létezik egy $g = \bar{g} \circ S_k$ faktorizáció, ahol $\bar{g} \in B(\mathcal{S})$.

Ha a 4. ábra diagrammjára tekintünk, a fenti definíció a $k+1$. lépésre kimondva az ábrán látható bal oldali „háromszög" mint részdiagramm kommutativitásával egyenértékű. Azt mondhatjuk, hogy a mérhető és rekurzív $\{S_k\}$ sorozat akkor elégséges, ha a $V_k(\cdot)$ költség- és $\phi_k(\cdot)$ stratégiafüggvények rá nézve szeparálhatóak. A dinamikus programozási elv értelmében (ld. (21) és (22) egyenletek) ez véges időhorizont esetén azzal ekvivalens, hogy a T_k^u operátor megőrzi a szeparálhatóságot. Valóban, a $\mathbf{0}$ szeparálható függvényből kiindulva hátráló indukcióval belátható a költségfüggvények szétválaszthatósága tetszőleges szeparált stratégia esetén. Az optimális stratégia (ϕ^*) a (18) definíció szerint örökli az értékfüggvénytől ezt a tulajdonságot. Végtelen horizont esetén az előző fejezet végén tett megjegyzés alapján határátmenettel kapjuk, hogy az elégségesség feltétele szintén az, hogy a megállítási operátor megtartsa a szeparálhatóságot.

A T_k^u operátor konkrét alakját vizsgálva könnyen találunk az első feltételt teljesítő mennyiségeket. A (16) egyenletben levő feltételes várható értékeket felírhatjuk az $\mathcal{L}(X_k \mid Z_k)$, illetve $\mathcal{L}(Y_{k+1} \mid Z_k, U_k)$ feltételes eloszlások szerinti integrálként[8]. Az első tagban szereplő $\mathcal{L}(X_k \mid Z_k)$ eloszlás a továbbiakban fontos szerepet játszik. Látni fogjuk, hogy a fenti másik eloszlás kifejezhető belőle, továbbá bebizonyítjuk, hogy teljesíti a másik két feltételt is.

Véges sok – $M < \infty$ számú – elemből álló állapottér esetén az állapot feltételes eloszlásának szemléletes jelentése van. Ábrázoljuk az állapothalmazt egy $M-1$ dimenziós szimplex csúcsainak halmazával. Ekkor a szimplex *belső pontjai* megfeleltethetőek az M ponton értelmezett diszkrét eloszlásoknak. A rekurzivitási feltétel miatt a feltételes eloszlást olyan „információs állapotnak" tekinthetjük, mely X_k-hoz hasonlóan egy megfigyelés és egy döntés után frissítődik. Ezáltal egy kontinuum számosságú állapottéren „bolyongó" új döntési folyamatot kapunk, melyben az állapotátmenetek *determinisztikusak*. A kibővített állapottérben a csúcsokhoz közeli pontok az eredeti rendszer „jól kitalálható" állapotainak felelnek meg, míg a szimplex súlypontjához tartozó belső állapotról „semmit sem tudunk".

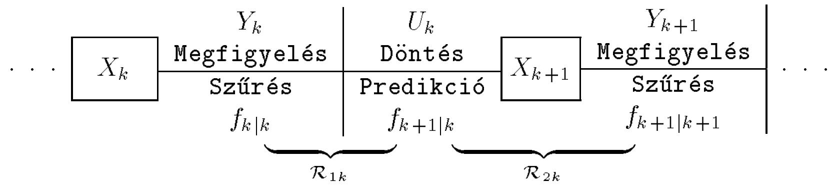
A 2. ábra táblázata alapján megállapíthatjuk, hogy a $P(X_k \in A \mid Z_k)$ feltételes eloszlás formálisan csak a ϕ stratégia ismeretében definiált. A továbbiakban belátjuk, hogy a fenti eloszlás valójában nem függ expliciten a stratégiától. Ehhez egy olyan kettős rekurziót írunk fel, mely figyelembe veszi a döntési folyamat I. Modell szerinti lépéseit (ld. 8. oldal). A bizonyítás tetszőleges állapotterek esetére kiterjed, de az egyszerűség kedvéért feltesszük, hogy az összes használt eloszlásnak van sűrűségfüggvénye. Az összes számolás elvégezhető az általános esetben is, ld. [Str65].

Azt a műveletet, amikor az X_k állapothoz hozzárendeljük az $\mathcal{L}^\phi(X_k \mid Z_k)$ feltételes eloszlást, illetve annak $f_{k|k}^\phi(x_k \mid z_k)$ sűrűségfüggvényét, *szűrésnek* nevezzük.[9] A gyakorlatban ez annak felel meg, hogy egy zajjal terhelt „jelből" kinyerjük a számunkra releváns információt. Jelen esetben a jel a Z_k információs folyamat. A U_k döntés ismerete nem jelent új információt, hiszen azt magunk állítottuk elő Z_k alapján, ezért minden k-ra

$$\mathcal{L}^\phi(X_k \mid Z_k) = \mathcal{L}^\phi(X_k \mid Z_k, U_k).$$

[8] $\mathcal{L}(\xi \mid \eta)$ az $A \mapsto P(\xi \in A \mid \eta)$ véletlen mértéket jelöli.

[9] Szigorúan véve a szűrés ezen eloszlás szerinti várható érték képzését jelenti.



4. ábra: Szűrés és előrejelzés

$\{Z_k, U_k\}$ alapján viszont elő tudjuk állítani a *következő* állapotot jellemző $\mathcal{L}^\phi(X_{k+1} \mid Z_k, U_k)$ eloszlás, illetve annak $f^\phi_{k+1|k}(x_{k+1} \mid z_k, u_k)$ sűrűségfüggvénye. Ez a művelet a *predikció* (ld. 4 ábra).

A szűrési és előrejelzési feltételes sűrűségek segítségével az X_k állapot becsléseit állíthatjuk elő a feltételes várható értékek formájában:

$$\begin{aligned}\hat{X}_k(z_k) = \hat{X}_{k|k}(z_k) &= \int x_k\, f^\phi_{k|k}(x_k \mid z_k)\, dx_k \\ \hat{X}_{k+1|k}(z_k) &= \int x_{k+1}\, f^\phi_{k+1|k}(x_{k+1} \mid z_k, u_k)\, dx_{k+1}.\end{aligned}$$

A 4. ábrán jelölt szűrési és predikciós eloszlások közötti első „átmenet" alatt a rendszer egy lépést tesz, míg az információs státus változatlan marad. A megfelelő sűrűségek ezért egy Kolmogorov-Chapman típusú formulával alakíthatók egymásba:

$$f^\phi_{k+1|k}(x_{k+1} \mid z_k, u_k) = \int p(x_{k+1} \mid x_k, u_k)\, f^\phi_{k|k}(x_k \mid z_k)\, dx_k, \tag{24}$$

ahol a $p(\cdot \mid \cdot, \cdot)$ Markov-mag (átmeneti sűrűség) a (7) egyenlet szerint független ϕ-től. A fenti integráloperátor (jelölje $\mathcal{R}^{u_k}_{1k}$) tehát a szűrési sűrűségfüggvényhez és az u_k értékhez hozzárendeli a predikciós sűrűségfüggvényt:

$$f^\phi_{k+1|k}(\cdot \mid z_k, u_k) = \mathcal{R}^{u_k}_{1k}\, f^\phi_{k|k}(\cdot \mid z_k) \tag{25}$$

Az $\mathcal{L}(X_{k+1} \mid Z_k, U_k) \to \mathcal{L}(X_{k+1} \mid Z_{k+1})$ átalakításnál nincs állapotátmenet, hanem az információmennyiség nő a $Z_{k+1} = \{Z_k, U_k\} \cup \{Y_{k+1}\}$ formulának megfelelően. Az új adat (Y_{k+1}) hatását X_{k+1} eloszlására (az *a posteriori* eloszlást) a $\{Z_k, U_k\}$ feltételes valószínűségi mezőre vonatkoztatott Bayes-tétel alapján határozhatjuk meg a

- a belső állapot (feltételes) eloszlásából, ami maga a predikciós eloszlás és

- a megfigyelés „*a priori*" eloszlásából, mely a (8) egyenlet miatt nem függ expliciten a feltételtől.

Ha ez utóbbi eloszlás sűrűségfüggvényét $q(y \mid x)$- szel jelöljük, akkor a következő összefüggés áll fenn:

$$f^{\phi}_{k+1|k+1}(x_{k+1} \mid z_{k+1}) = \frac{q(y_{k+1} \mid x_{k+1}) f^{\phi}_{k+1|k}(x_{k+1} \mid z_k, u_k)}{\int q(y_{k+1} \mid x_{k+1}) f^{\phi}_{k+1|k}(x_{k+1} \mid z_k, u_k) dx_{k+1}}, \qquad (26)$$

vagy funkcionális alakban

$$f^{\phi}_{k+1|k+1}(\cdot \mid z_{k+1}) = \mathcal{R}^{y_{k+1}}_{2k} f^{\phi}_{k+1|k}(\cdot \mid z_k, u_k), \qquad (27)$$

ahol az $\mathcal{R}_{2k}$ funkcionál független ϕ- től.

A (25)és (27) egyenletek kombinálásával a szűrési sűrűségre kapunk rekurziót:

$$f^{\phi}_{k+1|k+1}(\cdot \mid z_{k+1}) = \mathcal{R}^{y_{k+1}}_{2k} \mathcal{R}^{u_k}_{1k} f^{\phi}_{k|k}(\cdot \mid z_k) \qquad (28)$$

Ha az $f_{0|-1}$ függvényt azonosítjuk az X_0 kezdeti állapot eloszlásának f sűrűségfüggvényével[10], akkor a fenti rekurzió kezdeti értékeként az

$$f_{0|0}(\cdot \mid y_0) = \frac{q(y_0 \mid \cdot) f(\cdot)}{\int q(y_0 \mid u) f(u) du} \qquad (29)$$

függvényt kapjuk, ami független ϕ-től. A (25) és (27) rekurziókból következik, hogy az $f_{k|k}$ és $f_{k+1|k}$ sűrűségfüggvények sem függnek a stratégiától.

E megállapításból és a (28) egyenletből következik, hogy az $\mathcal{L}(X_k \mid Z_k)$ véletlen eloszlás, illetve annak $\Pi_k(\cdot \mid Z_k)$ sűrűségfüggvénye mérhető és rekurzív a döntési problémára nézve. Hátramaradt annak igazolása, hogy T^u_k megőrzi a $\{\Pi_k\}$-szeparálhatóságot.

A 4. ábra szerint azt akarjuk belátni, hogy a $S_k = \Pi_k$ esetén a szaggatott nyíl behúzásával kommutatív diagrammot kapunk. Másszóval, ha $g \in B(\mathrm{dom}\, \phi_{k+1})$ előáll $g = \bar{g} \circ S_{k+1}$ alakban, akkor a T^u_k operátornak megfeleltethető egy $\bar{T}^u_k : B(\mathcal{S}) \to B(\mathcal{S})$ operátor úgy, hogy fennáll a

$$T^u_k g = (\bar{T}^u_k \bar{g}) \circ S_k \qquad (30)$$

faktorizáció.

[10]Ez lehet egy pontra koncentrált is, amennyiben a folyamat kezdeti állapota ismert.

A (16) egyenletben definiált T_k^u operátor első tagja az $\int \gamma_k(\xi, u)\Pi_k(\xi)d\xi$ integrál, ami a Π_k statisztika mérhető függvénye: $\Psi_0[\Pi_k]$ alakú. A második tag

$$E\left[g(Z_{k+1}) \mid Z_k = z_k, U_k = u\right] = E\left\{g[(z_k, u, Y_{k+1})] \mid Z_k = z_k, U_k = u\right\}, \quad (31)$$

vagyis az $\mathcal{L}(Y_{k+1} \mid Z_k = z_k, U_k = u)$ eloszlás szerinti integrál, melynek sűrűségfüggvénye a teljes várható érték tételéből

$$\int q(y_{k+1} \mid x_{k+1}) f_{k+1|k}(x_{k+1} \mid z_k, u_k) dx_{k+1}.$$

Az integrandus első tényezője a rendszerparaméterek függvénye, a második a predikciós sűrűség, ami (25) szerint Π_k függvénye. Mivel feltettük, hogy g szeparálható, a (31) egyenletben a következő átalakításokat végezhetjük:

$$g[(z_k, u, Y_{k+1})] = \bar{g} \circ S_{k+1}(z_k, u, Y_{k+1}) = \bar{g} \circ \bar{\mathcal{R}}_k(S_k(z_k), u, Y_{k+1}).$$

Vagyis általában az operátor második tagja $\Psi_1^g[S_k(z_k), \Pi_k(z_k)]$ alakú (az adatoktól S_k-n és Π_k-n keresztül függ). Összességében tehát

$$(T_k^u g)(z_k) = \Psi_0[\Pi_k] + \Psi_1^g[S_k(z_k), \Pi_k(z_k)] \quad (32)$$

Ha speciálisan $S_k = \Pi_k$, akkor a jobboldal $h \circ S_k(z_k)$ alakú, ahol a h egyértelműen meghatározott. Ezt a függvényt jelöljük $\bar{T}_k^u \bar{g}$-sal. Ezzel beláttuk, hogy a $T_k^u g$ függvény is szeparálható, vagyis érvényes a (30) egyenlőség.

Az, hogy a Π_k statisztika $\mathcal{X}$-en értelmezett véletlen függvény, látszólag semmit sem könnyít a DP egyenlet megoldásán. Két esetben azonban további egyszerűsítéseket végezhetünk:

1. Ha $\mathcal{X}$ véges halmaz, akkor a Π_k függvény azonosítható egy $|\mathcal{X}|$ dimenziós vektorral (ld. 20. oldal). Ezzel az esettel a következő fejezetben foglalkozunk.

2. Ha az $\mathcal{L}(X_k \mid Z_k)$ eloszlás mint valószínűségi változó jellemezhető bizonyos paraméterekkel (pl. a momentumok sorozatával vagy normális esetben az *első két* momentummal), akkor e paraméterek összessége az eredetivel ekvivalens elégséges statisztikát alkot, és a költségfüggvény, illetve az optimális stratégia felírható e paramétervektor függvényében.[11]

[11]Ezen állítás triviálisan következik a Π_k elégséges voltának bizonyításánál elmondottakból.

5. Markov döntési folyamatok

E fejezetben alkalmazzuk a korábban kidolgozott elméletet a véges állapot- és döntési terű , illetve véges időhorizontú részlegesen megfigyelt Markov döntési folyamatokra. Összefoglaljuk a modell alapelemeit a vizsgálandó speciális esetben.

Legyen egy rendszer lehetséges állapotainak halmaza $\mathcal{X} = \{1, 2, \ldots, M\}$. A rendszer időben homogén Markov-i bolyongást végez az $\mathcal{X}$ halmazon olyan módon, hogy a $(P)_{M\times M}$ átmenetvalószínűség- mátrix minden lépésben egy kívülről megadott paramétertől függ:

$$P_{ij}(u) = P(X_{k+1} = j \mid X_k = i, U_k = u), \tag{33}$$

ahol $i, j \in \mathcal{X}$ és $u \in \mathcal{U}$, $\mathcal{U}$ véges halmaz. Az X_k állapotot nem ismerjük, helyette egy $\mathcal{Y} = \{1, 2, \ldots, K\}$ halmazbeli Y_k áll rendelkezésre. Ez a mennyiség X_k-val valószínűségi kapcsolatban áll:

$$Q_{i\theta} = P(Y_k = \theta \mid X_k = i) \tag{34}$$

tetszőleges $i \in \mathcal{X}, \theta \in \mathcal{Y}$ esetén[12].

A $p_{ij}(u)$ és $q_{i\theta}$ valószínűségeket a $P(u)$ $M \times M$- es, illetve a Q $K \times M$-es *sztochasztikus* mátrixokban foglaljuk össze. Döntések rögzített $u_0, u_1, \ldots$ sorozatára az X_k folyamat inhomogén Markov- lánc, mely a $\pi(0)$ kezdeti eloszlás ismeretében teljesen leírható. A k. pillanatban a döntéshozáshoz felhasználható információ $Z_k = (\theta_1, u_1, \theta_2, u_2, \ldots, u_{k-1}, \theta_k)$.

Az előző fejezetben láttuk, hogy az állapot feltételes eloszlása elégséges statisztika. Mivel az állapottér véges, ezért a rajta értelmezett eloszlások halmaza beágyazható $\Re^M$-be. Legyen

$$\pi_k = (\pi_k(1), \ldots, \pi_k(M))$$

az ún. információvektor, ahol $\pi_k(i) = P(X_k = i \mid Z_k = z_k)$. Az elégséges statisztikák tulajdonságai alapján ha adott egy π realizáció (a továbbiakban mindig megfelelő eseményeken vett realizációkat tekintünk), akkor a „frissített” értéke meghatározható az u döntés és a θ megfigyelés alapján:

$$\pi' = T(\pi \mid u, \theta)$$

[12] A q_{ik} mennyiségek szintén függhetnének u-tól, ld. 8. oldal

A $T(\cdot \mid u, \theta)$ függvény a 4. ábra $\hat{\mathcal{R}}\cdot, u, \theta$ leképezésének felel meg. (Az idő homogenitása miatt az indexet elhagyhatjuk.)

A π_k-ra vonatkozó rekurzió meghatározásához vezessük be a $\pi_{k+1|k}$ jelölést a $[P(X_{k+1} = i \mid Z_k, U_k)]_{i=1,\ldots,M}$ predikciós valószínűség-vektorra. A (25) rekurzió Chapman-Kolmogorov egyenlete

$$\pi_{k+1|k}(i) = \sum_{j=1}^{M} P_{ji}(u)\pi_k(j),$$

vagy mátrix-alakban[13]

$$\pi_{k+1|k} = P(u)'\pi_k. \tag{35}$$

A (26) egyenlet során alkalmazott Bayes-tételben az együttes feltételes $P(X_{k+1} = = i, Y_{k+1} = \theta \mid Z_k, U_k)$ valószínűséget „normáljuk", vagyis leosztjuk az ilyen típusú valószínűségek i szerinti összegével. A számláló könnyen kifejezhető $\pi_{k+1|k}$-val:

$$P(Y_{k+1} = \theta \mid X_{k+1} = i) \cdot P(X_{k+1} = i \mid Z_k, U_k) = Q_{i\theta}\pi_{k+1|k}(i),$$

vagy a $D(\theta) = \mathrm{diag}\,(Q_{1\theta}, \ldots, Q_{M\theta})$ jelöléssel mátrix-alakban $D(\theta)\pi_{k+1|k}$. Egyesítve a (35) egyenlettel és alkalmazva az $A^u_\theta = D(\theta) \cdot P(u)'$ rövidítést, a normálatlan (vagy *Zakai*-féle) $A^u_\theta\,\pi_k$ alakot kapjuk. A normált statisztika

$$\pi_{k+1} = \frac{A^u_\theta\,\pi_k}{||A^u_\theta\,\pi_k||} \tag{36}$$

ahol $||\cdot||$ az argumentuma koordinátáinak összegét jelöli (L^1-norma).

A Zakai-alak közvetlenül is előjön az ún. *referencia-mérték* módszerben (ld. [ELM95]), melynek lényege az, hogy konstruálunk egy olyan valószínűségi mezőt, melyre nézve a megfigyelések független azonos eloszlású valószínűségi változók. Így a problémát Fubini- típusú tételekkel elemi részeire bontjuk, majd visszatérünk az eredeti valószínűségi mezőre.

Megjegyezzük, hogy (36) nevezőjének szemléletes jelentése van: megegyezik a megállítási operátor második tagjának integrálként felírt alakjában szereplő $P(Y_{k+1} = \theta \mid Z_k, U_k)$ valószínűséggel. Ha c^u jelenti az egylépéses költségek $[\gamma(i,u)]_{i=1,\ldots,M}$ vektorát, akkor a T^u operátor alakja

$$(T^u g)(\pi) = \pi' \cdot c^u + \sum_{\theta=1}^{K} ||A^u_\theta \pi||\, g\left(\frac{A^u_\theta \pi}{||A^u_\theta \pi||}\right) \tag{37}$$

[13]A továbbiakban $'$ a transzponálás jele.

Legelőször is belátjuk, hogy T^u megtartja a konvexitást. Mivel az összeadás és kompozíció művelete nem vezet ki a konvex (konkáv) függvények köréből, elég belátni, hogy a g és a $v \mapsto ||v|| \cdot g\left(\frac{v}{||v||}\right)$ függvények egyszerre konvexek vagy konkávak. Ez könnyen látható, ugyanis az L^1-norma additivitása ($||\alpha v_1 + \beta v_2|| = \alpha||v_1|| + \beta||v_2||$) következtében az

$$||\alpha v_1 + \beta v_2|| \cdot g\left(\frac{\alpha||v_1||\frac{v_1}{||v_1||} + \beta||v_2||\frac{v_2}{||v_2||}}{||\alpha v_1 + \beta v_2||}\right)$$

formulában g argumentuma konvex kombináció, ha $\alpha + \beta = 1$ és $\alpha, \beta > 0$. Innen következik az állítás.

Ennél sokkal több is igaz: a $V_k(\cdot)$ függvények *szakaszonként lineáris konkáv* függvények, vagyis

$$V_k(\pi) = \min_{l \in L_k} \pi \cdot \alpha_k^l \tag{38}$$

alakba írhatóak, ahol L_k véges indexhalmaz. Ezen állítás bizonyítása hátráló indukcióval történik.

$V_N(\pi) = \min_u (T^u 0)(\pi) = \min_u \pi' \cdot c^u$, vagyis $c^u = \alpha_N^u, u \in \mathcal{U} = L_N$ és $|L_N| < \infty$.

Tegyük fel, hogy már beláttuk, hogy

$$V_{k+1}(\pi) = \min_{l \in L_{k+1}} \pi' \cdot \alpha_{k+1}^l.$$

Ekkor

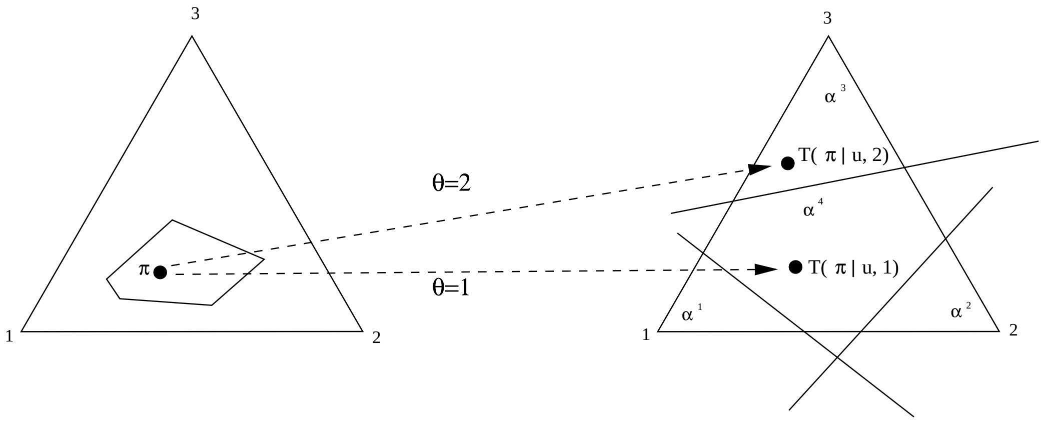
$$V_k(\pi) = \min_u [\pi' \cdot c^u + \sum_{\theta=1}^{K} ||A_\theta^u \pi|| \, V_{k+1}\left(T(\pi \mid u, \theta)\right)] \tag{39}$$

Az indukciós felteveésből tudjuk, hogy az információs vektorok szimplexe $|L_{k+1}|$ számú diszjunkt tartományból áll, melyeken a V_{k+1} függvény lineáris (ld. 5. ábra). Jelölje $l(\pi, u, \theta)$ annak a tartománynak az indexét, amelybe a transzformált $T(\pi \mid u, \theta)$ vektor esik, vagyis

$$V_{k+1}[T(\pi \mid u, \theta)] = T(\pi \mid u, \theta)' \cdot \alpha_{k+1}^{l(\pi,u,\theta)} \tag{40}$$

Behelyettesítve a (39) egyenletbe és elvégezve az egyszerűsítéseket, a

$$V_k(\pi) = \min_u \pi' \cdot [c^u + \sum_{\theta=1}^{K} (A_\theta^u)' \, \alpha_{k+1}^{l(\pi,u,\theta)}] \tag{41}$$



5. ábra: Sondik algoritmusa

formulát kapjuk. Ez már majdnem a kívánt alak, de π együtthatója sajnos $l(\pi, u, \theta)$-n keresztül függ π-től. Ezen a következőképpen segíthetünk. Rögzített u és θ mellett $T(\pi \mid u, \theta)$ folytonossága miatt a véges értékű $l(\pi, u, \theta)$ függvény partíciót generál a szimplexen úgy, hogy az egyes tartományokon $l(\pi, u, \theta)$ állandó. Különböző θ-kra más és más partíciók adódnak. Vegyük ezek *közös finomítását*. Ez a $\mathcal{P}_n$ partíció szintén véges és tartományain $l(\pi, u, \theta)$ állandó. Tehát a (41) egyenletben a „min" argumentuma szakaszonként lineáris, ezért $\min_{A \in \mathcal{P}_n} \pi' \cdot \alpha^A$ alakú. Még egy minimum-képzést alkalmazva

$$V_k(\pi) = \min_{l \in L_k} \pi' \cdot \alpha_k^l,$$

ahol $|L_k| \leq |U| \cdot \max |\mathcal{P}_n|$.

Megjegyezzük, hogy ha ismert az α-vektorok $(\alpha_{k+1}^l)_{l \in L_{k+1}}$ rendszere, akkor (40) és (41) alapján minden π információvektorhoz meghatározhatjuk azt az α-vektort, melyre $V_k(\pi) = \pi' \cdot \alpha_k^l$, illetve az optimális kontroll értékét. Valóban, $l(\pi, u, \theta) = \text{argmin}_{l \in L_{k+1}} T(\pi \mid u, \theta)' \cdot \alpha_{k+1}^l$ minden π, u, θ esetén[14], és (41)-ből az optimális kontroll az u^* minimalizáló argumentum, az α-vektor

[14] Az l függvény tabulálható.

pedig

$$\alpha_k^u = c^u + \sum_{\theta=1}^{K} (A_\theta^{u^*})' \, \alpha_{k+1}^{l(\pi,u^*,\theta)}. \tag{42}$$

Mivel bármely α-vektorhoz tartozik egy kontroll, ezért elég ezt a *párosítást* tárolni a stratégia teljes specifikációja helyett, hiszen a (38) egyenletben elvégezve a minimalizálást megkapjuk a megfelelő α-vektort, és a hozzá tartozó kontroll lesz az optimális.

Az α-vektorok kiszámítása

Az alábbi, **E. Sondik**tól származó algoritmus rekurzívan meghatározza az *összes* α-vektort. A módszer lényegében az $l(\pi, u, \theta)$ által indukált partíció „globális feltérképezésén" alapul.

Kiindulunk a szimplex egy tetszőleges π^0 belső pontjából, majd a fentiek szerint meghatározzuk a megfelelő α_k^* vektort és u^* optimális kontrollt. Ezek után elindulunk π^0 tartományának határa felé. Kétféle „határsértést" követhetünk el: vagy olyan tartományba kerülünk, ahol ugyanaz a kontroll értéke, csak más α-vektorokkal valósul meg, vagy pedig olyanba, ahol a kontroll is különböző.

Vizsgáljuk meg, hogy mikor maradunk a tartományon belül. Rögzített θ-ra akkor nem követünk el elsőfajú határátlépést, ha $l = l(\pi^0, u^*, \theta) = l(\pi, u^*, \theta)$, vagyis a (38) egyenlet szerint

$$T(\pi \mid u^*, \theta)' \cdot \alpha_k^l \leq T(\pi \mid u^*, \theta)' \cdot \alpha_k^n$$

minden $n \in L_k$ esetén. Valójában ezen egyenlőtlenségeket elegendő azon α_k^n-ekre felírni, melyeknek tartománya *szomszédos* α_k^l tartományával.

Másodfajú átlépést akkor nem követünk el, ha a tartományon az u^* által szolgáltatott α-vektor érvényes, vagyis minden a (42) egyenletből π^0-lal és u-val számolt α_k^u vektorra

$$\pi' \cdot \alpha_k^* \leq \pi' \cdot \alpha_k^u$$

Ezen egyenleteket minden θ-ra felírva a következő rendszert kapjuk:

$$\begin{array}{lcll}
\pi' \cdot (A_\theta^{u^*}) \cdot [\alpha_k^n - \alpha_k^l] & \geq & 0 & \text{minden } \theta\text{-ra és szomszédos } n\text{-re} \\
\pi' \cdot [\alpha_k^u - \alpha_k^*] & \geq & 0 & \text{minden } \theta\text{-ra és } u \in U\text{-ra} \\
\pi & \geq & 0 & \\
\pi' \cdot \mathbf{E} & = & 1 &
\end{array}$$

A π^0 tartományának definiálásához általában elegendő ennek a rendszernek egy része is. A következő egyszerű LP-feladat megoldásával meghatározható a minimális definiáló rendszer. Ha a fenti egyenlőtlenségek bármelyikének baloldalát célfüggvényként minimalizáljuk az összes feltétel mellet, akkor az egyenlőtlenség pontosan akkor definiáló, ha a megoldásban a neki megfelelő „slack" változó 0. Sőt, ekkor a kapott π megoldás a szóban forgó határhipersíkon fekszik.

A fenti módszerrel „megnyirbált" feltételek egyúttal újabb α-vektorokat szolgáltatnak, melyek a határon ugyanazt a $V_k(\pi)$ értéket adják, mint α_k^*. Ezeket a vektorokat későbbi feldolgozásra feljegyezzük egy listába a hozzájuk tartozó határponttal és kontroll-értékkel együtt. Így az algoritmus egy lépésben egyrészt feltérképez egy pont körüli tartományt. másrészt új α-vektorokat talál a további lépések számára.

Összefoglalva, az algoritmus felépítése tehát a következő:

1. Kiválasztunk egy kezdeti π^0-t és kiszámoljuk u^*-ot és α_k^*-ot.

2. Felírjuk a teljes feltételrendszert.

3. Kiválasztjuk a definiáló feltételeket és az új α-vektorokat, melyeket egy határponttal és a kontrollal tárolunk.

4. Tároljuk a tartomány szomszédainak indexeit.

5. Ha a listán van α-vektor, akkor vissza a 2. lépéshez, különben STOP.

Ez az algoritmus, mely áttörést hozott a részlegesen megfigyelt Markov döntési folyamatok megoldásában, garantáltan megtalálja az összes α-vektort és ezáltal meghatározza az optimális stratégiát.

A végtelen diszkontált probléma általában nehezebben kezelhető. Teljes megfigyelhetőség esetén *Howard stratégiajavító módszere* véges sok lépésben konvergál [15], de ez általában nincs így. Howard módszere igényel egy a stratégiák kiértékelésére használható eljárást, melyre pontos módszer csak speciális stratégiák esetén létezik, de ismert egy szintén **Sondik** nevéhez fűződő közelítő módszer, ld. [Son78].

[15] ld. [How60]

6. A megfigyelés szerepe

Ezidáig a megfigyelést a döntési folyamat előre adott, szerves részének tekintettük: minden állapotátmenet után vagy automatikusan (I. Modell), vagy pedig döntésünk függvényében (II. Modell) egy *rögzített* szabály szerint észleljük a rendszer kimenetét. A belső állapot és a megfigyelés kapcsolata (véges állapottér esetén a Q mátrix) egyszer és mindenkorra adott volt. Ebben a fejezetben a megfigyelés változtatásának hatását vizsgáljuk.

Teljes megfigyelés esetén az elégséges statisztikák egy pontra koncentrált degenerált eloszlásokká egyszerűsödnek, ezért az előző fejezetek „komplikációi" mind eltűnnek. Azt várnánk, hogy ha esetleg kevésbé drasztikus változtatással csak *javítanánk* a megfigyelés minőségén, akkor már ennek is jelentkeznie kell az eredményben. A kontroll jóságát a költségfüggvény méri, ezért „pontosabb" megfigyelés esetén kevesebb összköltségre számítunk. Be fogjuk látni, hogy bizonyos fajta stratégiák esetén tényleg fennáll egy ilyen adaptivitási tulajdonság. Az egyszerűség kedvéért véges alaptéren dolgozunk, a levezetések általánosabb esetben megtalálhatóak a ([WH80]) cikkben. A 7. fejezetben tárgyalt rendszer az említett tulajdonságon kívül még azt is teljesíti, hogy az optimális kontroll független a megfigyelés minőségétől, különbség csak a költségben jelentkezik.

Először is tisztáznunk kell, hogy mit értünk a megfigyelés pontosságán, illetve elég azt megmondanunk, hogy mikor nevezünk egy megfigyelést a másiknál pontosabbnak.[16] Kézenfekvő a döntések szempontjából megközelíteni a kérdést: a $\hat{Q}$ megfigyelés akkor pontosabb (vagy informatívabb) a Q megfigyelésnél, ha „minden Q alapján hozott döntés $\hat{Q}$ alapján is meghozható". Matematikailag ez a következőt jelenti.

Tegyük fel, hogy a döntési feladat a

$$c^u = [\gamma(i,u)]_{i=1,\dots M}$$

egylépéses költségvektorral jellemezhető. Ha e vektorok $\mathcal{C} = \{c^u : u \in \mathcal{U}\}$ halmazát tekintjük döntési térnek, akkor egy $a = c^u$ „döntés" i. koordinátája egyben a döntés költsége az i belső állapot esetén. A döntési eljárásokat a $\phi : \mathcal{Y} \to \mathcal{C}$ leképezések adják meg, melyek megmondják, hogy egy adott megfigyelés esetén milyen döntést hozzunk.

[16]A kísérletek összehasonlításának statisztikai elméletét illetően ld. [Tor76].

Az eljárás minőségét a

$$v(\phi) = [v_i(\phi)]_{i=1,\dots M}$$

vektor méri, ahol az i. koordináta a ϕ eljárás várható költsége $X = i$ belső állapot esetén:

$$v_i(\phi) = \sum_\theta [\phi(\theta)]_i P(Y = \theta \mid i) = \sum_\theta Q_{i\theta}\Phi_{\theta i}$$

ahol a Φ mátrix *sorai* a különböző θ megfigyelésekhez tartozó döntésvektorok. Mátrix-jelöléssel

$$v(\phi) = \mathrm{diag}(Q \cdot \Phi) \tag{43}$$

A $v(\phi)$ vektorok halmaza legyen $R_1(Q, \mathcal{C})$, ha ϕ befutja a $\mathcal{C}$-beli értékű döntési függvények halmazát. Érdemesebb e halmaz $R(Q, \mathcal{C})$ *konvex burkát* tekinteni, ugyanis ekkor a döntések $\mathcal{C}$ tere helyett is annak konvex burka jön elő, ami szemléletesen azt jelenti, hogy megengedünk véletlenített (ún. kevert) döntéseket is.[17]

Akkor mondjuk a $\hat{Q}$ megfigyelést Q-nál pontosabbnak, ha tetszőleges költségstruktúrára a Q mellett kevert döntéssel elérhető költségvektorok $\hat{Q}$ mellett is elérhetőek:

$$\hat{Q} \supset Q \quad \overset{\text{def}}{\Longleftrightarrow} \quad R(\hat{Q}, \mathcal{C}) \supset R(Q, \mathcal{C}) \text{ minden } \mathcal{C} \text{ konvex halmazra.}$$

Ez a **Bohnenblust**tól származó szemléletes definíció sajnos közvetlenül nem felel meg céljainknak, ugyanis egyrészt konkrét Q és $\hat{Q}$ mátrixokra a feltétel nem ellenőrizhető, másrészt egy „minden konvex halmazra" kimondott tulajdonság nem illeszthető be eddigi technikai apparátusunkba. Az első problémára részleges választ ad **Blackwell** cikke ([Bla51]), melyben egy sor ekvivalens átfogalmazás található a $\supset$ reláció definíciójára. **LeCam** könyvében ([LeC86]) sokkal általánosabb megfontolásokból adódik a következő, aránylag egyszerűbb kritérium:

$$\hat{Q} \supset Q \quad \Longleftrightarrow \quad \sum_i \max_\alpha \sum_j Q_{ij} d_{j\alpha} \le \sum_i \max_\alpha \sum_j \hat{Q}_{ij} d_{j\alpha}$$

a $0 \le d_{\alpha j} \le 1$ együtthatók minden véges rendszerére. A szélsőértékek lineáris programozási módszerekkel kiszámolhatóak.

[17] A konvex kombináció képzése egy teljes várható érték tétel alkalmazásának felel meg.

A második probléma megoldásához tekintsük a megfigyelési pontosság „elromlásának" a következő speciális módját. Ha a rendszer belső állapotát $\hat{Q}$ szerint megfigyeljük, majd az eredményt a *másodlagos* véletlenített F megfigyelésnek vetjük alá, akkor a végeredmény még kevésbé lesz informatív. Ha az összetett megfigyelést Q jelöli, akkor a megfelelő mátrixokra fennáll, hogy

$$Q = \hat{Q} \cdot F$$

Be fogjuk látni, hogy ha $\hat{Q}$ pontosabb Q-nál, akkor előáll egy alkalmas másodlagos megfigyeléssel való kompozícióként[18] (ez utóbbi relációt $\succ$ jelöli).

Tétel (Sherman-Stone): $\hat{Q} \supset Q \quad \Longrightarrow \quad \hat{Q} \succ Q$

A bizonyításhoz először rögzítsük a dimenziókat: $\hat{Q} \in \Re^{M \times L}$ és $Q \in \Re^{M \times K}$. Tekintsünk egy $\Phi \in \Re^{K \times M}$ mátrixával megadott ϕ döntésfüggvényt és sorainak konvex kombinációjából építsük fel az $\mathcal{F}$ halmazt, melynek elemei pont a kevert döntések. $\hat{Q} \supset Q$ esetén speciálisan az is igaz, hogy $R(\hat{Q}, \mathcal{F}) \supset R(Q, \mathcal{F})$, ami a (43) egyenlet alapján azt jelenti, hogy

$$v(\phi) = \mathrm{diag}(Q \cdot \Phi) = \mathrm{diag}(\hat{Q} \cdot \Phi')$$

valamely $\mathcal{F}$-beli értékű ϕ' döntésfüggvényre. De Φ' sorai a Φ sorainak *konvex kombinációi*, ezért létezik egy $F \in \Re^{L \times K}$ sztochasztikus mátrix úgy, hogy $\Phi' = F\Phi$

$$\mathrm{diag}(\hat{Q} F \Phi) = \mathrm{diag}(Q \Phi) \tag{44}$$

Tekintsük a $h(\Phi, F) = \mathrm{Tr}(Q - \hat{Q}F) \cdot \Phi$ bilineáris függvényt, ahol most Φ az 1- nél kisebb pozitív elemű $(K \times M)$-es mátrixokon, F pedig az $L \times K$ dimenziós sztochasztikus mátrixokon fut végig. Mindkét értelmezési tartomány konvex és a mátrix- norma topológiájában kompakt, ezért fennáll a minimax-tétel[19], vagyis h-nak létezik (Φ_0, F_0) *nyeregpontja*:

$$h(\Phi_0, F) \geq h(\Phi_0, F_0) \geq h(\Phi, F_0) \tag{45}$$

A fentiek szerint (44) alapján Φ_0-hoz létezik olyan F, hogy $h(\Phi_0, F) =$ $= \mathrm{Tr}(Q - \hat{Q}F)\Phi_0 = 0$, ezért (45) miatt minden Φ-re

$$h(\Phi, F_0) = \mathrm{Tr}(\underbrace{Q - \hat{Q}F_0}_{U})\Phi \leq 0$$

[18] Valójában a fordított irányú implikáció is igaz: ld. ([Bla51]).
[19] ld. pl. ([BKS50])

Következésképpen a megjelölt mátrix minden eleme ≤ 0. De U két sztochasztikus mátrix különbsége, ezért elemeinek összege 0. Ez csak úgy lehet, ha $U = 0$, vagyis

$$Q = \hat{Q} F_0$$

Ezzel a tétel bizonyítását befejeztük.

Beláttuk tehát, hogy a $\supset$ reláció a mátrixok nyelvén azt jelenti, hogy a rosszabb megfigyelés mátrixa előáll a jobbik mátrixának és egy alkalmas sztochasztikus mátrixnak a szorzataként. Rögtön látszik, hogy e parciális rendezésnek van legjobb és legrosszabb eleme. Intuíciónknak megfelelő módon a legnagyobb elem a *teljes* megfigyelés:

$$\forall Q \in \Re^{M \times K} : Q \prec I, \text{mert } Q = IQ, \text{ahol } I = \mathrm{id}_{M \times M},$$

a legrosszabb elem pedig a nem-informatív megfigyelés, vagyis amikor minden belső állapotot egyforma valószínűséggel láthatunk. A mátrixa $T = \frac{1}{M}\mathbf{E}$, ahol $\mathbf{E}$ a csupa 1-esből álló $M \times M$-es mátrix):

$$\forall Q \in \Re^{M \times K} : T \prec Q, \text{mert } T = Q \cdot \frac{1}{M} H, \text{ahol } H \in \Re^{K \times M} \text{csupa 1-es mátrix.}$$

A megállítási operátort definiáló (16) egyenletből látszik, hogy a megfigyelés mátrixa egyedül az

$$A^u_\theta = \mathrm{diag}(Q_{1\theta}, \ldots, Q_{M\theta}) \cdot P(u)'$$

mátrixban szerepel. Megvizsgáljuk, hogy hogyan változik a π vektorok transzformációs szabálya, ha Q helyett a $\hat{Q}$ megfigyelést használjuk, ahol $\hat{Q} = QF$. $\hat{Q}_{i\eta} = \sum_{\theta=1}^{K} Q_{i\theta} F_{\theta\eta}$, ezért

$$\hat{D}(\eta) = \mathrm{diag}(\hat{Q}_{1\theta}, \ldots, \hat{Q}_{M\theta}) = \sum_{\theta=1}^{K} D(\theta) F_{\theta\eta}$$

vagyis $\hat{A}^u_\eta = \sum_{\theta=1}^{K} A^u_\theta F_{\theta\eta}$. Ezért

$$\hat{T}(\pi \mid u, \eta) = \sum_{\theta=1}^{K} H(\eta \mid \theta, \pi, u)\, T(\pi \mid u, \theta) \tag{46}$$

ahol a $H(\eta \mid \theta, \pi, u) = \frac{||A^u_\theta \pi||}{||\hat{A}^u_\eta \pi||} F_{\theta\eta}$ pozitív együtthatók θ szerinti összege 1.

Tekintsük most a megállítási operátor második tagját:

$$\sum_{\theta=1}^{K} ||A_\theta^u|| \cdot g\left(T(\pi \mid u, \theta)\right)$$

ahol g konkáv függvény. Belátjuk, hogy ez nem lehet nagyobb a $\hat{Q}$-hoz tartozó megfelelő tagnál.

Valóban, a Jensen-egyenlőtlenség szerint

$$g(\hat{T}(\pi \mid u, \eta)) \geq \sum_{\theta=1}^{K} H(\eta \mid \theta, \pi, u) g(T(\pi \mid u, \theta))$$

ezért beszorozva és összegezeve azt kapjuk, hogy

$$\sum_{\eta=1}^{L} ||\hat{A}_\eta^u \pi|| \cdot g(\hat{T}(\pi \mid u, \eta)) \geq \sum_{\eta=1}^{L} ||\hat{A}_\eta^u \pi|| \sum_{\theta=1}^{K} \frac{||A_\theta^u \pi||}{||\hat{A}_\eta^u \pi||} F_{\theta\eta}\, g\left(T(\pi \mid u, \theta)\right).$$

Felcserélve az összegzéseket, a

$$\sum_{\eta=1}^{L} ||\hat{A}_\eta^u \pi|| g(\hat{T}(\pi \mid u, \eta)) \geq \sum_{\theta=1}^{K} ||A_\theta^u \pi|| g\left(T(\pi \mid u, \theta)\right) \tag{47}$$

egyenlőtlenséget kapjuk, amit bizonyítani akartunk.

Jelölje $V_k^{\phi,Q}$ a ϕ stratégiához és a Q megfigyeléshez tartozó költségfüggvényeket ($V := V_0$). A ϕ stratégiát akkor nevezzük Q**- adaptív**nak, ha minden $Q' \succ Q$ megfigyelésre

$$V^{\phi,Q'} \leq V^{\phi,Q}$$

A (47) egyenlőtlenség következtében a ϕ pontosan akkor Q-adaptív, ha a $V_k^{\phi,Q}$ függvények *konkávak*.

A ϕ stratégia legyen Q**-optimális**, ha minden ϕ' stratégiára

$$V^{\phi,Q} \leq V^{\phi',Q}$$

A 26. oldalon belátott tulajdonság alapján e fejezet fő eredményét a következő tételben fogalmazhatjuk meg:

Tétel: A Q-optimális stratégiák Q- adaptívak.

Ezen állításból az is következik, hogy ha ϕ Q-optimális és ϕ' Q'-optimális, akkor az értékfüggvényekre fennáll a

$$V^{\phi,Q} \leq V^{\phi',Q'}$$

egyenlőtlenség.

A Q-adaptivitás fogalma tovább általánosítható. Ha Q' a korábban definiált T nem-informatív megfigyelés és ϕ' a hozzá tartozó optimális stratégia (az ún. OLC—Open Loop Controller), akkor a fenti egyenlőtlenséget teljesítő ϕ stratégiákat nevezzük **adaptív**nak. A gyakorlatban gyakran előforduló szuboptimális megoldás a ϕ' stratégia használata $Q \succ Q'$ megfigyelés mellet (OLFC—Open Loop Feedback Controller). A fentiek szerint ez jobb eredményt ad, mint az OLC („vak") stratégia.

7. Az LQG probléma

Ebben a fejezetben kidolgozzuk az irodalomban LQG (**L**inear–**Q**uadratic–**G**aussian) problémának nevezett feladat megoldását. Ez egyike azon kevés feladatoknak, ahol az értékfüggvény és az optimális kontroll expliciten kiszámolhatók.

Mint azt a 2. fejezetben említettük, a nemlineáris kontrollfeladatok közelítő megoldásának egyik módja a *linearizálás*. Bár az így kapott megoldás szuboptimális, kellően sima függvények esetén az általa nyert számítástechnikai egyszerűség ezt ellensúlyozza. Az állapot-, illetve megfigyelési egyenlet alakja legyen tehát

$$\begin{aligned} X_{k+1} &= A_k X_k + B_k U_k + W_k && (48) \\ Y_k &= C_k X_k + V_k && (49) \end{aligned}$$

ahol W_k és V_k 0 várható értékű Σ_k^W, illetve Σ_k^V szórásmátrixú *normális* eloszlású valószínűségi változók. Az egyszerűség kedvéért tegyük fel, hogy a sztochasztikus tagok *függetlenek* .[20]

A költségfüggvényről feltesszük, hogy X_k-ban és U_k-ban *kvadratikus* tagok összege:

$$V_0(z) = E[\sum_{k=0}^{N} (X_k' Q_k X_k + U_k' R_k U_k) \mid Z_0 = z], \qquad (50)$$

ahol Q_k és R_k megfelelő dimenziós pozitív definit mátrixok. Ez az alak emlékeztet a klasszikus mechanika Hamilton- függvényére, és számos mérnöki alkalmazásban előfordul.

Könnyen látható, hogy az elégséges statisztika $\mathcal{L}(X_k \mid Z_k = z)$ realizációja normális eloszlás, melyet első két momentuma jellemez. A 4. fejezet végén leírtak szerint az $(\hat{X}_k, P_k)$ pár is elégséges statisztika, ahol

$$\hat{X}_k = E(X_k \mid Z_k) \text{ és } P_k = E[(X_k - \hat{X}_k)(X_k - \hat{X}_k)' \mid Z_k].$$

E mennyiségeket széles körben használják a sztochasztikus rendszerek állapotbecslésének elméletében (ld. [ELM95]). Lineáris rendszerekre **Kálmán Rudolf** adta meg az állapot L^2-becslésére vonatkozó rekurziókat. Az úgynevezett *Kálmán- szűrés* egyenleteit a 7. ábra táblázata sorolja fel.

Két megjegyzés az egyenletekhez:

[20] Feltehetnénk azt is, hogy $W_k = V_k$. Az általános esetet illetően ld. [Bag93].

k. lépés	
1. frissítés	2. frissítés
Predikció	Szűrés
$\hat{X}_{k+1\mid k} = A_k \hat{X}_k + B_k U_k$	$\hat{X}_{k+1} = \hat{X}_{k+1\mid k} + K_{k+1}\left[Y_{k+1} - C_{k+1}\hat{X}_{k+1\mid k}\right]$
$P_{k+1\mid k} = A_k P_k A_k' + \Sigma_k^W$	$P_{k+1} = \left[I - K_{k+1} C_{k+1}\right] P_{k+1\mid k}$
ahol $K_{k+1} = P_{k+1\mid k} C_{k+1}' \left[C_{k+1} P_{k+1\mid k} C_{k+1}' + \Sigma_k^V\right]^{-1}$	

6. ábra: A Kálmán-szűrés egyenletei

1. Az $\hat{X}_k$-ra és P_k-ra vonatkozó kettős rekurziók a (24) és (26) egyenletekkel analógok. Az általános formulákból való származtatásuk nehéz; legegyszerűbben a Hilbert-terek geometriai szemléletén alapuló „innovációs" megközelítéssel bizonyíthatóak (ld. [Bag93] és [Mar96]).

2. A normalitás következtében a P_k feltételes szórásmátrixra vonatkozó egyenletek csak a feladat paramétereitől függnek és függetlenek a megfigyelésektől. Ez azt jelenti, hogy a $\{P_k\}$ mátrixsorozat előre („off-line") kiszámolható és elhagyható az elégseges statisztikából. Ez a tulajdonság a kulcsa az LQG-probléma explicit megoldhatóságának.

A szeparációs elv értelmében a DP egyenlet a következő alakot ölti:

$$V_k^*(\hat{x}) = \min_u T_k^u V_{k+1}^*(\hat{x}),$$

ahol $k = 0, \ldots, N$ és $V_{N+1}^* \equiv 0$, a megállítási oprátor pedig

$$(T_k^u g)(\hat{x}) = E[X_k' Q_k X_k + u' R_k u \mid \hat{X}_k = \hat{x}] + E[g(\hat{X}_{k+1}) \mid \hat{X}_k = \hat{x}, U_k = u]. \tag{51}$$

A hátráló indukciós megoldáshoz tekintsük a $k = N$ utolsó lépést. A minimalizáló argumentum $u = 0$, ezért $V_N^* = E[X_N' Q_N X_N \mid \hat{X}_k = \hat{x}]$. Ha X_k-t az $\hat{X}_k$ becslés és a becslési hiba összegeként írjuk fel, akkor az $X_N' Q_N X_N$ kvadratikus alak kifejtésekor a vegyes tagok feltételes várható értéke 0, vagyis V_N^*-ban csak a „négyzetek" jelennek meg („Pithagorasz-tétel"). Felhasználva, hogy $y'Ax = \mathrm{Tr}[yx'A]$, egyszerű átalakítások után kapjuk, hogy

$$V_N^*(\hat{x}) = \hat{x}' Q_N \hat{x} + \mathrm{Tr}[P_N Q_N]. \tag{52}$$

A DP egyenlet alapján várható, hogy a többi értékfüggvény is kvadratikus $\hat{x}$- ben. Ennek belátására írjuk V_k^*-ot a fenti egyenlet mintájára

$$V_k^*(\hat{x}) = \hat{x}' P_k^* \hat{x} + q_k^*, \tag{53}$$

alakba, ahol a P_k^* és q_k^* mennyiségeket próbáljuk azonosítani.

Ha behelyettesítünk a megállítási operátor fenti alakjába, két feltételes várható értéket kell kiértékelnünk. Az elsőről könnyen látható, hogy

$$E[X_k' Q_k X_k \mid \hat{X}_k = \hat{x}] = \hat{x}' Q_k \hat{x} + \mathrm{Tr}[P_k Q_k]$$

a fentiekhez hasonló módon. Valamivel bonyolultabb az

$$N_k(\hat{X}_{k+1}) := E[\hat{X}_{k+1}{}' P_{k+1}^* \hat{X}_{k+1} \mid \hat{X}_k = \hat{x}, U_k = u]$$

kiszámolása. $N_k(\cdot)$ kvadratikus és belátható, hogy argumentumának az $(X \mid Y) =$ $= \mathrm{cov}(X, Y)$ skalárszorzat szerint ortogonális felbontásaira szintén érvényes lesz a Pithagorasz-azonosság. Írjuk fel X_{k+1}-et az $\{\hat{X}_k = \hat{x}, U_k = u\}$ eseményre vonatkoztatott szűrési, illetve predikciós becslési segítségével (a megfelelő becslési hibákat $\tilde{}$ jelzi):

$$X_{k+1} = \hat{X}_{k+1} + \tilde{X}_{k+1} = \hat{X}_{k+1|k} + \tilde{X}_{k+1|k}$$

Az állapotegyenletből $\hat{X}_{k+1|k} = A_k \hat{x} + B_k u$ és $E[\tilde{X}_{k+1|k} \tilde{X}_{k+1|k}{}'] = P_{k+1|k}$ (a predikciós becslés kovariancia-mátrixa független a megfigyelésektől). Mindkét felbontás ortogonális, ezért

$$N_k(X_{k+1}) = N_k(\hat{X}_{k+1}) + N_k(\tilde{X}_{k+1}) = N_k(\hat{X}_{k+1|k}) + N_k(\tilde{X}_{k+1|k}).$$

A mátrixszorzások sorrendjét felcserélve és azonosítva a P_k és $P_{k+1|k}$ mátrixokat, az

$$N_k(\hat{X}_{k+1}) + \mathrm{Tr}[P_{k+1} P_{k+1}^*] = (A_k \hat{x} + B_k u)' P_{k+1}^* (A_k \hat{x} + B_k u) + \mathrm{Tr}[P_{k+1|k} P_{k+1}^*]$$

azonosságot kapjuk, ahonnan adódik $N_k(\hat{X}_{k+1})$ értéke.

Visszahelyettesítve a DP egyenletbe az

$$\hat{x}' P_k^* \hat{x} = \min_u \left\{ u' R_k u + (A_k \hat{x} + B_k u)' P_{k+1}^* (A_k \hat{x} + B_k u) + L_k \right\} \tag{54}$$

alakot kapjuk, ahol $L_k := \hat{x}'Q_k\hat{x} + \mathrm{Tr}[P_kQ_k] + \mathrm{Tr}[(P_{k+1|k} - P_k)\,P^*_{k+1}] + q^*_{k+1}$ független u-tól. A (54) egyenlet jobboldala az u vektor kvadratikus függvényének minimuma, ami differenciálással kiszámolható. A minimalizáló argumentum a

$$\frac{\partial}{\partial u}\left\{u'R_ku + (A_k\hat{x} + B_ku)'P^*_{k+1}(A_k\hat{x} + B_ku)\right\} = 0.$$

egyenlet megoldása. Egy skalár értékű függvény vektor szerinti deriváltját sorvektornak tekintve a $\partial_x(x'Ax) = 2x'A$ és $\partial_x(Bx) = B$ formulák alapján a $2u'R_k + 2(A_k\hat{x} + B_ku)'P^*_{k+1}B_k = 0$ egyenletet kapjuk, amiből

$$u = -G_k\,\hat{x} \quad \text{ahol} \quad G_k = -\left[R_k + B_k'P^*_{k+1}B_k\right]^{-1} B_k'P^*_{k+1}A_k. \tag{55}$$

(Az inverz létezése következik R_k pozitív definit voltából.) Visszahelyettesítve a (54) egyenletbe azonosíthatjuk az ezidáig ismeretlen P^*_k és q^*_k mennyiségeket:

$$P^*_k = A_k'P^*_{k+1}(A_k - B_kG_k) + Q_k \tag{56}$$

$$q^*_k = \mathrm{Tr}[P_kQ_k] + \mathrm{Tr}\left[(P_{k+1|k} - P_k)P^*_{k+1}\right] + q^*_{k+1} \tag{57}$$

Ez az egyenletrendszer a

$$P^*_N = Q_N, \quad q^*_N = \mathrm{Tr}[P_NQ_N]$$

kezdeti feltételekből kiindulva rekurzívan megoldható.

Összefoglalva az eredményeket, az LQG- probléma megoldása a következő:

Optimális kontroll	$U_k = -G_k\hat{X}_k$
Optimális költségfüggvény	$V^*_k(\hat{x}) = \hat{x}'P^*_k\hat{x} + q^*_k$

Az optimális kontrollt vizsgálva látjuk, hogy az lineáris és az (55) egyenletben definiált G_k együttható *teljesen független* a rendszer megfigyelését leíró (49) egyenlettől. A megfigyelés hatása tehát csak az $\hat{X}_k$ becsült állapoton keresztül jelenik meg. Ha a megfigyelések pontosak, akkor az (55) egyenlet triviális és a Kálmán-szűrés becslése X_k-ra egyszerűsödik. Így megkaptuk a teljesen megfigyelt lineáris rendszerek optimális kontroll-stratégiáját: $U_k = -G_kX_k$. A költségfüggvény $V^*_k(x) = \hat{x}'P^*_k\hat{x} + q_k$, ahol $q_N = 0$ és $q_k = \mathrm{Tr}\left[\Sigma^W_kP^*_{k+1}\right] + q_{k+1}$ mivel $P_k \equiv 0$ és $P_{k+1|k} = \Sigma^W_k$.

A szeparációs elv következtében az LQG- probléma kontrollja a következő két lépésre bomlik:

1. először az állapotot a Kálmán-szűrővel becsüljük

2. a rendszert úgy irányítjuk, mintha a becsült érték teljes megfigyelésből származna.

A modell tovább egyszerűsödik, ha az állapotátmenetből is kihagyjuk a véletlent. Formálisan ezt a $\Sigma^W \equiv 0$ helyettesítéssel érjük el. Mivel a P_k^* mátrixok függetlenek W_k-tól, ez a változtatás csak q_k-t érinti, ami most azonosan 0. Az optimális stratégia tehát ugyanaz marad, csak az optimális költség lesz kisebb. Ezt a jelenséget az irodalomban a „bizonyossággal való egyenértékűség elvének" (*certainty equivalence principle*) nevezik.

Irodalom

[Alb79] S. Albright. Structural results for partially observable Markov decision processes. *Operations Res*, 27:1041–1053, 1979.

[Ara93] A. Arapostathis et al. Discrete-time controlled Markov processes with average cost criterion: A survey. *SIAM J. Control Optim*, 31(2):282–344, 1993.

[Åst65] K. J. Åström. Optimal control of Markov processes with incomplete state information. *J. Math. Anal. Appl*, 10:174–205, 1965.

[Åst69] K. J. Åström. Optimal control of Markov processes with incomplete state-information ii. the convexity of the lossfunction. *J. Math. Anal. Appl*, 26:403–406, 1969.

[Bag93] A. Bagchi. *Optimal Control of Stochastic Systems*. Prentice Hall, 1993.

[Bel57] R. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, N. J., 1957.

[Ber87] D. Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models*. Prentice Hall, 1987.

[BKS50] H. F. Bohnenblust, S. Karlin, and L. S. Shapley. *Games with continuous convex payoff*, pages 181–192. Princeton University Press, 1950.

[Bla51] D. Blackwell. Comparison of experiments. In J. Neyman, editor, *Proc. Second Berkeley Sympos. Math. Statist. Probab*, pages 93–102, 1951.

[Bla53] D. Blackwell. Equivalent comparison of experiments. *Ann. Math. Stat*, 24:265–272, 1953.

[Bra94] S. J. Bradtke. *Incremental Dynamic Programming for on-line adaptive optimal control*. PhD thesis, University of Massachusetts Amherst, September 1994.

[BW90] R. Bhattacharya and E. Waymire. *Stochastic processes with applications*. John Wiley & Sons, Inc, 1990.

[Dyn65] E. B. Dynkin. Controlled random sequences. *Theory Probab. Appl*, 10(1):1–14, 1965.

[ELM95] R. J. Elliott, A. Lakhdar, and J B. Moore. *Hidden Markov Models. Estimation and Control*. Springer-Verlag, 1995.

[Fel68] W. Feller. *An Introduction to Probability Theory and Applications*, volume 1. John Wiley & Sons, Inc, 3 edition, 1968.

[Fre88] S. French. *Decision theory: an introduction to the mathematics of rationality*. John Wiley & Sons, Inc, 1988.

[Hel86] P. Helman. The Principle of Optimality in the design of efficient algorithms. *J. Math. Anal. Appl*, 119:97–127, 1986.

[HL86] F. S. Hillier and G. J. Lieberman. *Introduction to operations research*. McGraw Hill, 1986.

[How60] R. A Howard. *Dynamic Programming and Markov Processes*. The Technology Press of MIT and John Wiley & Sons, Inc., New York, London, 1960. Joint Publication.

[Kum85] P. R. Kumar. A survey of some results in stochastic adaptive control. *SIAM J. Control Optim*, 23(3):329–380, May 1985.

[LeC86] L. LeCam. *Asymptotic methods in statistical theory*. Springer, 1986.

[Leh83] E. L. Lehmann. *Theory of Point Estimation*. John Wiley & Sons, Inc, 1983.

[Mar96] I. Maricza. Optimal control of partially observable stochastic systems, January 1996. Lecture Note.

[Mon82] G. E. Monahan. A survey of partially observable Markov decision processes: Theory, models and algorithms. *Management Sci*, 28(1):1–16, 1982.

[Pré94] A. Prékopa. *Stochastic Programming*. Akadémiai Kiadó, 1994.

[Put90] M. L. Puterman. *Markov Decision Processes*, volume 2, pages 331–434. North-Holland, Amsterdam, 1990.

[Rab89] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–286, 1989.

[Rhe74] D. Rhenius. Incomplete information in Markovian decision models. *Ann. Statist*, 2:1327–1334, 1974.

[Ros70] S. M. Ross. *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco, California, 1970.

[Son78] E. J. Sondik. The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs. *Operations Res*, 26(2):282–304, 1978.

[SS73] R. D. Smallwood and E. J. Sondik. The optimal control of partially observable Markov processes over a finite horizon. *Operations Res*, 19:1071–1088, 1973.

[Str65] C. Striebel. Sufficient statistics in the optimum control of stochastic systems. *J. Math. Anal. Appl*, 12:576–592, 1965.

[Sze95] Cs. Szepesvári. Perfect dynamics for neural networks, July 1995.

[Tor76] E. Torgersen. Comparison of statistical experiments. *Scand. J. Statist*, 3:186–208, 1976.

[Wal50] A. Wald. *Statistical decision functions.* John Wiley & Sons, Inc, 1950.

[WH80] C. C. White and D. P. Harrington. Application of Jensen's inequality to adaptive suboptimal design. *J. Optim. Theory and Appl*, 32(1):89–99, 1980.

[Whi88] D. J. White. Mean, variance and probabilistic criteria in finite Markov decision processes: A review. *J. Optim. Theory and Appl*, 56(1):1–27, 1988.