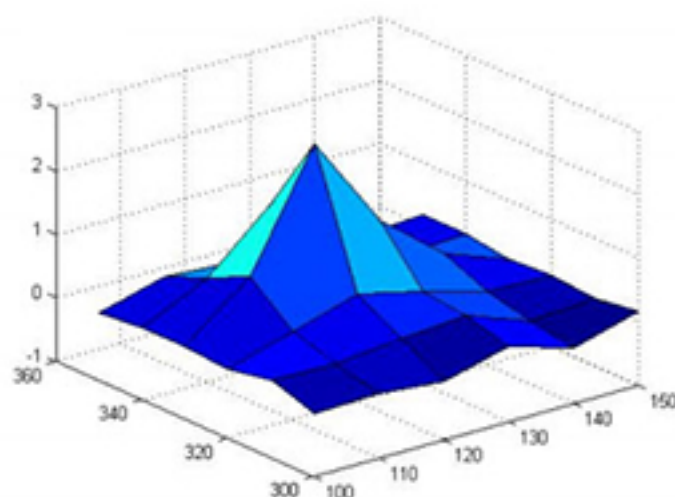




GEOFIZIKAI INVERZIÓ

MOLNÁR GÁBOR

Geofizikai inverzió

Molnár Gábor

Geofizikai inverzió

Molnár Gábor

Szerzői jog © 2013 Eötvös Loránd Tudományegyetem

E könyv kutatási és oktatási célokra szabadon használható. Bármilyen formában való sokszorosítása a jogtulajdonos írásos engedélyéhez kötött.

Készült a TÁMOP-4.1.2.A/1-11/1-2011-0073 számú, „E-learning természettudományos tartalomfejlesztés az ELTE TTK-n” című projekt keretében. Konzorciumvezető: Eötvös Loránd Tudományegyetem, konzorciumi tagok: ELTE TTK Hallgatói Alapítvány, ITStudy Hungary Számítástechnikai Oktató- és Kutatóközpont Kft.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638



A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.



Tartalom

1. Előszó	1
2. Lineáris algebra összefoglaló	2
2.1. Mátrixok és vektorok	2
2.2. Amátrix rangja	2
2.3. Sajátérték, sajátvektor, karakterisztikus egyenlet	3
2.4. Az általánosított inverz	4
2.5. Mátrixok SVD felbontása	5
3. A valószínűségszámítás alapjai	8
3.1. A valószínűség	8
3.2. Valószínűségi változók eloszlásfüggvénye és sűrűségfüggvénye	9
3.3. A valószínűségi változók jellemzői	11
A várható érték	11
A medián	11
A szórás	11
Momentumok és centrális momentumok	12
Két vagy több változó együttes eloszlását jellemző mennyiségek	12
3.4. Nevezetes eloszlások	13
Egyenletes eloszlás	13
Laplace (kétoldali exponenciális) eloszlás	13
Normális eloszlás	14
Többdimenziós (többváltozós) normális eloszlás	14
χ^2 eloszlás	15
Student eloszlás	15
3.5. Határértéktételek, nagy számok törvényei	16
A centrális határeloszlás tétele	16
Előírt pontosságú közelítéshez szükséges kísérletszám meghatározása	17
3.6. Statisztikai sokaság	18
A statisztikai becslések	19
Konfidencia intervallumok	20
3.7. Maximum likelihood elv	20
3.8. Valószínűségi változók függvényének eloszlását jellemző mennyiségek	23
4. A geofizikai inverzió általános megfogalmazása	25
5. A geofizikai inverzió statisztikai megközelítése	29
5.1. A geofizikai inverz feladat megfogalmazása	29
A kísérleti anyag additív modellje	30
A normális eloszlású additív modell	32
Lineáris elméleti terű additív modell	33
5.2. Statisztikai becslési eljárások a geofizikai inverz feladat megoldására	34
A maximum likelihood elv	34
A maximum likelihood elv alkalmazása minőségi interpretációra	35
A maximum likelihood elv alkalmazása mennyiségi interpretációra	35
A maximum likelihood elv alkalmazása lineáris elméleti terű modellek minőségi interpretációjára	36
A maximum likelihood elv alkalmazása minőségi interpretációra a priori valószínűségek felhasználásával	37
A maximum likelihood elv alkalmazása összetett interpretáció esetén	37
Az additív modellek véletlen eltérés komponensének analízise	37
5.3. A kritériumfüggvény lehetséges alakjai	39
5.4. Minimum kereső eljárások	42
Lineáris egyenletek megoldására visszavezethető eljárások	42
Gradiens módszer	43
Konjugált vektor módszer	43
Konjugált gradiens módszer	44
Intervallum keresés	45
Szimplex módszer	45

A szimulált hűtés (Simulated Annealing)	47
Az genetikus algoritmus	48
5.5. A statisztikai becslések minőségét ellenőrző eljárások	49
6. Lineáris inverzió	52
Lineáris inverzió túlhatározott probléma esetén	53
Lineáris inverzió alulhatározott probléma esetén	54
Lineáris inverzió egyidejűleg túl- és alulhatározott problémák esetén	54
7. Az inverzió megoldása a legkisebb négyzetek módszer segítségével	56
8. A legkisebb négyzetek módszerének általános felírása	59
8.1. A legkisebb négyzetek módszer – alapfogalmak	59
8.2. Egyenes illesztése	61
8.3. A legkisebb négyzetek módszerében szereplő mennyiségek kovarianciamátrixai	62
8.4. Másodfokú függvény illesztése	63
8.5. Legkisebb négyzetes becslés eltérő súlyú mérések esetén	64
8.6. A mérések szórásának becslése a mérési anyagból	66
9. A nem lineáris direkt feladat esete	70
9.1. Gravitációs mérés	72
10. Kényszerfeltételek alkalmazása	77
10.1. A feltételes szélsőérték keresés	77
10.2. Egyenes illesztése – lineáris kényszerfeltétel	81
10.3. Egyenes illesztése – nem lineáris kényszerfeltétel	82
11. Kiegyenlítés csak mért mennyiségeket tartalmazó feltételi egyenletekkel	84
11.1. A feladat módosítása: több mérés két álláspont között	86
11.2. A feladat újabb módosítása: több álláspont közötti vonal	87
11.3. Eredmények	88
12. Csillapított (regularizált) legkisebb négyzetes becslés	90
13. Osztályozás	94
13.1. Egyszerű példa	94
13.2. Valódi műholdkép vizsgálata	99
13.3. Maximum likelihood osztályozás	102
13.4. Mahalanobis osztályozás	104
13.5. Minimális távolság (minimum distance) osztályozás	104
13.6. Paralelepipedon osztályozás	105
14. Összefoglalás	106
15. Irodalomjegyzék	107
16. Köszönetnyilvánítás	108

1. fejezet - Előszó

Ez a jegyzet elsősorban geofizikus hallgatók számára íródott. A célja, hogy bevezesse a hallgatókat az inverzióelméletbe, és néhány gyakorlati példa segítségével felkészítse őket, hogy egyszerűbb inverziós példákat maguk is meg tudjanak oldani, a megoldás és az eredmények elemzése során pedig ők maguk is közelebb kerüljenek az elmülethez.

Az inverzióelmélet alapjainak megértésével az a célunk, hogy a leendő geofizikusok később képesek legyenek a szakirodalom olvasása során megérteni az összetettebb inverziós problémákat és megoldásukat.

Az inverzióelmélettel kapcsolatban több – köztük sok angol nyelvű – tankönyv létezik. A tankönyvek az elméletet számos esetben példák segítségével próbálják érthetőbbé tenni. Az inverzióelmélet – mivel általánosságban a mérésekből történő információ kinyerésével foglalkozik, – valamennyi geofizikai módszerrel kapcsolatos. Emiatt az inverzió bemutatásához felhasznált példák a geofizika több területéről merítenek.

A geofizikai inverzióval foglalkozó könyvekben számos nem klasszikus geofizikai példát találunk. Ennek az oka, hogy az inverzióhoz felhasználjuk az ún. „direkt feladat” megoldását, ami sok esetben bonyolult (egy-egy geofizikai témakör általában több féléves tantárgy). Emiatt egy inverziós jegyzet kompromisszumot kell, hogy találjon aközött, hogy egy-egy példát annak szinte teljes geofizikai háttérrel együtt bemutasson, illetve aközött, hogy a direkt feladatok levezetését ismertnek tételezi fel, és csak az inverz feladatot tárgyalja. Meggyőződésünk, hogy a lehető legegyszerűbb direkt feladatokat érdemes alkalmazni az inverzió bemutatásához azért, hogy a direkt feladat bonyolultsága ne vonja el a figyelmet magáról az inverzióról.

Ebben a könyvben a példaként felhozott problémák egy része geofizikai, de tekintettel az űrtudományi képzésre részletesen tárgyalunk egy távérzékelési-képfeldolgozási problémát is. Azok a példák, amelyeket számolással kidolgozva szerepeltetünk a jegyzetben, többnyire egyszerű geometria feladatok, mivel ezeknek a problémáknak a megértése és a számolások reprodukálása a koordináta-geometria alapszintű ismeretét tételezi fel.

Az inverzióelmélet megértése a lineáris algebra és a valószínűség számítás ismeretét igényli. Ezeket a témaköröket a jegyzet elején vázlatosan tárgyaljuk.

Egy ilyen széles területet átfogó jegyzet több forrás anyagát próbálja meg egy egységes műbe összedolgozni. Az egyes területeket, illetve az inverzió különféle megközelítéseit olyan mélységig kell bemutatnia a jegyzetnek, hogy az olvasó számára megérthető és elsajátítható legyen. A mennyiségi korlát miatt a források szövegeit rövidíteni kellett. Abban az esetben, ha valamelyik olvasó egy adott részterülettel komolyabban akar foglalkozni, akkor kézbe kell vennie az adott témakörrel foglalkozó forrásokat.

Mivel különböző „iskolák” eljárásait próbáltuk egy egységes műbe egybedolgozni, nem lehetett a jegyzetben a képletekben alkalmazott jelölésrendszert teljesen egységesíteni: emiatt az egyes fejezetekben nagyrészt megtartottuk az eredeti forrásmű jelölésrendszerét. Ez megkönnyíti az érdeklődő olvasó számára az eredeti forrásmű használatát.

2. fejezet - Lineáris algebra összefoglaló

Ebben a fejezetben a mátrixokkal kapcsolatos – az inverzióelmélet megértéséhez szükséges – ismereteket tekintjük át. A mátrixokkal kapcsolatos alapvető műveleteket ismertnek tételezzük fel. Ilyen műveletek például: mátrixok összeadása, szorzása, mátrix szorzása vektorral, egységmátrix fogalma stb.

2.1. Mátrixok és vektorok

A vizsgálataink során $N \times M$ méretű, valós számokból álló mátrixokkal foglalkozunk, ahol N a sorok, M pedig az oszlopok száma. Eszerint egy $N \times M$ méretű $\mathbf{A}$ mátrix:

$$\mathbf{A} \in \mathbb{R}^{N \times M} \quad (2.1)$$

Egy $\mathbf{A}$ mátrix elemei A_{ij} valós számok (az első index a sorok, a második az oszlop számát jelenti.)

Egy mátrix sorát **sorvektornak**, egy oszlopát **oszlopvektornak** tekintjük. Amennyiben egy mátrix oszlopainak és sorainak a száma megegyezik, a mátrixot **négyzetes mátrixnak** nevezzük ($\mathbf{A} \in \mathbb{R}^{N \times N}$).

Egy $N \times M$ méretű mátrix $\mathbf{A}$ mátrix **transzponáltján** azt a $M \times N$ méretű mátrixot értjük, amit úgy kapunk, hogy a mátrixot a főátlójára tükrözzük. Jele $\mathbf{A}^T$. Ha egy mátrix megegyezik a transzponáltjával akkor **szimmetrikus**, ($\mathbf{A} = \mathbf{A}^T$), ha a transzponáltja mínusz egyszeresével, akkor **antiszimmetrikus** ($\mathbf{A} = -\mathbf{A}^T$). (Ekkor a mátrix oszlopainak és sorainak száma megegyezik: $M = N$.)

A jegyzetben egy vektort általában **oszlopvektornak** gondolunk. Ezt úgy is értelmezhetjük, mint egy $N \times 1$ méretű mátrixot, és kiterjeszthetjük rá a mátrixok szorzására bevezetett műveletünket. Ennek segítségével egy $\mathbf{A} \in \mathbb{R}^{N \times M}$ mátrix és egy $\mathbf{x} \in \mathbb{R}^M$ vektor szorzata az alábbi $\mathbf{y} \in \mathbb{R}^N$ vektort jelenti: $\mathbf{y} = \mathbf{A} \cdot \mathbf{x}$, aminek elemeit az alábbi formulával

$$y_i = \sum_{j=1}^M A_{ij} \cdot x_j$$

kapjuk: (az eredmény szintén egy oszlopvektor!)

Egy mátrix előállítható két vektor **diadikus szorzataként**. Tekintsünk egy $\mathbf{x} \in \mathbb{R}^M$ oszlopvektort és $\mathbf{y} \in \mathbb{R}^N$ oszlopvektort. Képezzük $\mathbf{y}$ -ból transzponálással az $\mathbf{y}^T$ sorvektort. Ekkor egy $\mathbf{C}$ mátrix előállítható (a szorzás művelet érvényben tartásával):

$$\mathbf{C} = \mathbf{x} \cdot \mathbf{y}^T \quad \text{ahol a } \mathbf{C} \text{ mátrix mérete: } M \times N \text{ (a mátrix elemei: } C_{ij} = x_i \cdot y_j \text{)}.$$

Egy $\mathbf{Q} \in \mathbb{R}^{N \times N}$ mátrix **ortogonális**, ha $\mathbf{Q}^T \cdot \mathbf{Q} = \mathbf{I}$, vagyis az **egységmátrix**. Ebben az esetben az oszlopvektorok merőlegesek egymásra, és a normájuk egységnyi. (A merőlegesség a skaláris szorzat segítségével definiálható: két, nem nulla hosszúságú vektor merőleges egymásra, ha a skaláris szorzatuk nulla. A vektor normáján itt a hosszát értjük.) Az ortogonális mátrixok forgatást valósítanak meg, a transzformált vektor hossza (normája) nem változik meg.

2.2. Amátrix rangja

Tegyük fel, hogy van N darab vektorunk ($\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$) amelyeknek azonos a dimenziója (de a dimenzió nem kell pontosan N legyen!), és vizsgáljuk meg, hogy ezek **lineárkombinációja** milyen együttthatók mellett adja ki a null-

$$\sum_{i=1}^N \alpha_i \cdot \mathbf{x}_i = \mathbf{0}$$

vektort. Ha a

fennáll úgy, hogy legalább egy együttható nem nulla, akkor a vektoraink **lineárisan összefüggő**ek. Amennyiben az egyenlőség csak akkor áll fenn, ha az összes együttható zérus, akkor a vektoraink **lineárisan független**ek. Egy vektortér dimenziója megegyezik a lineárisan független vektorok számával, amelyek lineáris kombinációjaként a vektortér bármely eleme előállítható. Ezek a vektorok **bázis**t alkotnak.

Egy vektortér **altere** az a nem üres részhalmaz, amire igaz hogy:

- Bármely két elemének összege a részhalmaz eleme.
- Bármely elemének egy skalárral vett szorzata a részhalmaz eleme marad.

Egy $A \in \mathbb{R}^{N \times M}$ mátrix oszlopvektorai tehát $\mathbb{R}^N$ egy alterét feszítik ki. Ennek az alternek a dimenzióját az adja, hogy ezek közül hány vektor lineárisan független. Az oszlopvektorok által kifeszített alteret $R(A)$ -val jelöljük, dimenziója pedig a mátrix **rangja**.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$$

Példaként tekintsük az A mátrixot. Ennek a két oszlopvektora nem lineárisan független, emiatt csak egyetlen lineárisan független vektor van, emiatt a dimenziója, vagyis a mátrix rangja 1.

Az A mátrixhoz kapcsolódó másik alapvető alteret azok a vektorok alkotják, melyek kielégítik az $A \cdot x = 0$ homogén egyenletet. Ezt nevezzük a mátrix **null-terének**.

Például az előbbi A mátrixhoz az $x = (-2, 1)$ vektor és annak többszöröse a mátrix null-terét alkotja.

A mátrixok rangjának központi szerepe van az inverzióelméletben szereplő egyenletek megoldhatóságának vizsgálatában. A mátrixok rangjának egy másik definíciójának bevezetéséhez fel fogjuk használni a négyzetes mátrixokhoz definiált **determináns** fogalmát. Bizonyítás nélkül közöljük az $N \times N$ -es mátrix determinánsának kiszámoló képletét, ami:

$$\det(A) = \sum_{i_1, i_2, \dots, i_N} (-1)^{I(i_1, i_2, \dots, i_N)} \cdot a_{1i_1} \cdot \dots \cdot a_{Ni_N} \quad (2.2)$$

ahol az összegzés az $(1, 2, \dots, N)$ összes **permutációjára** történik, és $I(i_1, i_2, \dots, i_N)$ jelöli az $(i_1, i_2, \dots, i_N)$ permutációban lévő inverziók számát. Egy $N \times N$ -es mátrix egyik eleméhez tartozó $(N-1)$ -ed rendű aldeterminánsa, a mátrixból az adott elem oszlopában és sorában levő elemek elhagyásával kapott $(N-1) \times (N-1)$ -es mátrix determinánsa.

A mátrix **rangja** vonatkozó második definíció szerint a mátrix rangja k , ha k -ad rendű a mátrix legnagyobb el nem tűnő (nem nulla értékű) aldeterminánsa.

2.3. Sajátérték, sajátvektor, karakterisztikus egyenlet

A mátrixok **lineáris operátorok** reprezentációi, amik általánosságban a vektorok mátrixokkal történő szorzása során egy vektort egy másik vektorra transzformálnak.

A **transzformálás** (a mátrixszal való szorzás) során az eredményként létrejövő vektor iránya eltér a kiinduló vektortól. Találhatók azonban olyan vektorok, amiknek a mátrixszal való szorzás után nem változik az irányuk, ezek a **sajátvektorok**.

A fenti megfogalmazás a sajátvektorokra az alábbi egyenlettel fejezhető ki:

$$A \cdot s = \lambda \cdot s \quad (2.3)$$

Ahol $\mathbf{s}$ a sajátvektor, λ pedig a **sajátérték**.

Az egyenlet átrendezésével kapjuk a karakterisztikus (vagy szekuláris) egyenletet:

$$\det(\mathbf{A} - \lambda \mathbf{I}) \cdot \mathbf{s} = P_N(\lambda) = 0 \quad (2.4)$$

Ahol $\mathbf{I}$ az egységmátrix. Ez az egyenletrendszer csak akkor oldható meg, ha az egyenletben szereplő determináns értéke nulla. Ha az $\mathbf{A}$ mátrix $N \times N$ -es, akkor az egyenlet, tulajdonképpen λ -nak egy N -ed fokú polinomja: $P_N(\lambda)$, amelynek N darab gyöke van.

Definiálnunk kell még a **minimálpolinom**ot: Ez a legalacsonyabb fokszámú olyan polinom, amelybe az $\mathbf{A}$ mátrixot behelyettesítve 0-t kapunk. A minimálpolinom gyökei megegyeznek a karakterisztikus polinom gyökeivel, legfeljebb multiplicitásuk különböző. (A karakterisztikus polinomban többszörös gyökként jelennek meg.) A minimálpolinom általában megegyezik a karakterisztikus polinommal.

Ha minimálpolinom gyökei egyszeresek, akkor a mátrix **diagonizálható**. Egy mátrix diagonalizálta:

$$\mathbf{A} = [\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_N] \cdot \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N \end{bmatrix} \cdot \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_N^T \end{bmatrix} = \mathbf{U} \cdot \text{diag}(\lambda) \cdot \mathbf{V}^T \quad (2.5)$$

Ahol $\mathbf{U}^T \cdot \mathbf{V} = \mathbf{U} \cdot \mathbf{V}^T = \mathbf{I}$, vagyis $\mathbf{U}$ oszlopai a baloldali, $\mathbf{V}$ sorai pedig a jobboldali sajátvektorok.

Az $\mathbf{A}$ mátrix fenti formában történő előállítását **spektrálfelbontásnak** nevezzük.

A sajátvektorok **biortonormált rendszert** alkotnak. (Normáltak, vagyis az egységvektorok normája egységnyi. Ortonormáltak, vagyis két, nem ugyanahhoz a sajátértékhez tartozó egységvektor skaláris szorzata nulla.) A biortonormáltság definíciója szerint a skaláris szorzataikra:

$$\mathbf{v}_i^T \cdot \mathbf{u}_j = \delta_{ij} \text{ ahol } \delta \text{ a Kronecker féle szimbólum.}$$

A polinom gyökeinek vizsgálata nélkül is megállapíthatjuk, hogy ha egy mátrix felcserélhető az adjungáltjával, akkor az diagonalizálható. (A mátrix **adjungáltja**, a mátrix előjeles aldeterminánsaiból képzett mátrix transzponáltja.) Ilyenek a szimmetrikus mátrixok. (Az inverzió során sokszor ilyen mátrixok szerepelnek.) Ezeknek a mátrixoknak a jobb és baloldali sajátvektorai megegyeznek: $\mathbf{V} = \mathbf{U}$, ahol $\mathbf{U}$ **unitér** mátrix, amelynek inverze megegyezik a transzponáltjával.

Ezen tételek segítségével megfogalmazható a mátrix **rangjára** egy harmadik definíció is: a mátrix rangja a 0-tól különböző λ_i sajátértékek száma.

2.4. Az általánosított inverz

Az inverzió alkalmazása során – számos esetben – lineáris egyenletrendszereket oldunk meg, amelyek általános alakban így néznek ki:

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} \quad (2.6)$$

ennek megoldása:

$$\mathbf{x} = \mathbf{A}^{-1} \cdot \mathbf{y} \quad (2.7)$$

ahol $\mathbf{A}^{-1}$ a mátrix **inverze**. Egy mátrix inverzét az alábbi egyenlet definiálja:

$$\mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{I}, \quad (2.8)$$

vagyis egy mátrix inverzének saját magával vett szorzata az egységmátrix.

A mátrixszorzás nem kommutatív művelet, vagyis ha egy mátrixot balról vagy jobbról szorzunk meg egy mátrixszal nem ugyanazt az eredményt kapjuk. Emiatt definiálható **bal oldali inverz** és **jobb oldali inverz**.

1. Egy $\mathbf{R}_{(z,t)}$ es mátrix általános inverze az $\mathbf{Y} = \mathbf{R}^{-}$ mátrix, amire igaz, hogy:

$$\mathbf{R}_{(z,t)} \mathbf{R}^{-}_{(t,z)} \mathbf{R}_{(z,t)} = \mathbf{R}_{(z,t)} \quad (2.9)$$

2. Az $\mathbf{R}_{(z,t)}$ mátrixnak több általánosított inverze létezik, ezek egy csoportját alkotják a **reflexív általános inverzek**, amikre érvényes a fenti definíció mellett, hogy:

$$\mathbf{R}^{-}_{(t,z)} \mathbf{R}_{(z,t)} \mathbf{R}^{-}_{(t,z)} = \mathbf{R}^{-}_{(t,z)} \quad (2.10)$$

3. Az általános inverzek további tulajdonsággal is rendelkezhetnek:

$$\left(\mathbf{R}^{-}_{(t,z)} \mathbf{R}_{(z,t)} \right)^* = \mathbf{R}^{-}_{(t,z)} \mathbf{R}_{(z,t)} \quad (2.11)$$

vagyis hogy a szorzat **Hermitikus** legyen. Ez akkor teljesül, ha a zárójeles kifejezésben a szorzatmátrix komplex konjugáltja megegyezzen a szorzatmátrix transzponáltjával. Az ilyen típusú mátrixok sajátértékei valósak. Az inverzióelméletben valós elemű, általában szimmetrikus mátrixok szerepelnek, amelyekre ez a feltétel teljesül. Az ennek a feltételnek megfelelő $\mathbf{R}^{-}_{\mathbf{m}}$ mátrixokat **balról gyengén általánosított inverzeknek** nevezzük.

4. Emellé definiálható a **jobbról gyengén általánosított inverz**:

$$\left(\mathbf{R}_{(z,t)} \mathbf{R}^{-}_{(t,z)} \right)^* = \mathbf{R}_{(z,t)} \mathbf{R}^{-}_{(t,z)} \quad (2.12)$$

Egy $\mathbf{R}_{(z,t)}$ mátrixhoz található olyan mátrix, amelyik a fenti négy feltételnek együttesen megfelel. Ez a mátrix egyértelmű, és **Moore-Penrose féle pszeudoinverznek** nevezzük. Jelölése: $\mathbf{R}^{+}_{(z,t)}$.

A $\mathbf{R}^{+}_{(z,t)}$ pszeudoinverz mátrixot a Lános-felbontás (szinguláris érték dekompozíció, *Singular Value Decomposition*) felhasználásával valósíthatjuk meg.

2.5. Mátrixok SVD felbontása

A felbontás alaptétele, hogy bármilyen $\mathbf{A} \in \mathbb{R}^{N \times M}$ mátrix felírható az alábbi alakban:

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T \quad (2.13)$$

ahol az $\mathbf{U}$ mátrix $N \times N$ -es, és oszlopvektorai az $\mathbf{A}^T \cdot \mathbf{A}$ sajátvektorai, a $\mathbf{V}$ mátrix $M \times M$ -es és sajátvektorai az $\mathbf{A} \cdot \mathbf{A}^T$ mátrix sajátvektoraival egyeznek meg. A $\mathbf{\Sigma}$ mátrix ($N \times M$ -es), főátlójában a $\mathbf{A}^T \cdot \mathbf{A}$ sajátértékeinek pozitív négyzetgyökei – az ún. **szinguláris értékek** – állnak. A többi elem nulla.

Az $\mathbf{A}^T \cdot \mathbf{A}$ sajátértékei megegyeznek $\mathbf{A} \cdot \mathbf{A}^T$ nem nulla sajátértékeivel. A nem nulla sajátértékek száma megegyezik az $\mathbf{A}$ mátrix **rang**jával.

Az $\mathbf{A}$ mátrixból konstruáljunk egy $\mathbf{S}$ mátrixot az alábbi módon:

$$\mathbf{S} = \begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \quad (2.14)$$

Az $\mathbf{S}$ mátrix mérete: $(N+M) \times (N+M)$. Az $\mathbf{S}$ mátrix szimmetrikus, így sajátértékei valósak. A sajátértékek legyenek σ_i , a hozzájuk tartozó sajátvektorok $\mathbf{w}_i$.

A sajátértékekre vonatkozó egyenlet:

$$\mathbf{S} \cdot \mathbf{w}_i = \sigma_i \cdot \mathbf{w}_i \quad (2.15)$$

Az $(N+M)$ elemű sajátvektort osszuk két részre: egy N elemű $\mathbf{u}$ és egy M elemű $\mathbf{v}$ vektorra.

Ekkor a sajátértékekre vonatkozó egyenlet:

$$\begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} = \sigma_i \cdot \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} \quad (2.16)$$

Ami a hipermátrixban szereplő nullmátrixok miatt, az alábbi két egyenletre esik szét:

$$\begin{aligned} \mathbf{A}^T \cdot \mathbf{u}_i &= \sigma_i \cdot \mathbf{v}_i \\ \mathbf{A} \cdot \mathbf{v}_i &= \sigma_i \cdot \mathbf{u}_i \end{aligned} \quad (2.17-2.18)$$

(A második egyenletet átírhatjuk $\mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma}$ formába, amit később még felhasználunk.)

Az ebben a két egyenletben szereplő vektorok tehát a sajátvektor σ_i –szeresei. Ezt a vektort beírva a sajátérték egyenletbe, kapjuk hogy:

$$\begin{bmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{A} \cdot \mathbf{v}_i \\ \mathbf{A}^T \cdot \mathbf{u}_i \end{bmatrix} = \sigma_i \cdot \sigma_i \cdot \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} \quad (2.19)$$

Vagyis:

$$\begin{aligned} \mathbf{A} \cdot \mathbf{A}^T \cdot \mathbf{u}_i &= \sigma_i^2 \cdot \mathbf{u}_i \\ \mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{v}_i &= \sigma_i^2 \cdot \mathbf{v}_i \end{aligned} \quad (2.20 - 2.21)$$

Az $\mathbf{u}_i$ vektorok tehát az $\mathbf{A} \cdot \mathbf{A}^T$ mátrix sajátvektorai, az $\mathbf{v}_i$ vektorok viszont az $\mathbf{A}^T \cdot \mathbf{A}$ mátrix sajátvektorai. A σ_i –sajátértékek így mindkét mátrixhoz megegyeznek.

Az $\mathbf{u}_i$ vektorokból, mint oszlopvektorokból alkotott $\mathbf{U}$ és az $\mathbf{v}_i$ vektorokból alkotott $\mathbf{V}$ mátrixok ortogonális négyzetmátrixok, amikre fennáll, hogy:

$$\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I}_N \quad (2.22)$$

és

$$\mathbf{V}^T \mathbf{V} = \mathbf{V} \mathbf{V}^T = \mathbf{I}_M \quad (2.23)$$

A $\mathbf{V}$ mátrix ortogonalitása miatt a korábban kiszámolt $\mathbf{A} \mathbf{V} = \mathbf{U} \mathbf{\Sigma}$ egyenlet (2.18) a

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T \quad (2.24)$$

Alakban írható fel.

Ha az $\mathbf{A}$ mátrix rangja r , akkor a felbontás így is felírható:

$$\mathbf{A} = [\mathbf{U}_r, \mathbf{U}_0] \cdot \begin{bmatrix} \mathbf{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{V}_r^T \\ \mathbf{V}_0^T \end{bmatrix} = \mathbf{U}_r \cdot \mathbf{\Sigma}_r \cdot \mathbf{V}_r^T \quad (2.25)$$

Ahol $\mathbf{U}_r$ és $\mathbf{V}_r$ mátrixok az $\mathbf{U}$ és $\mathbf{V}$ mátrixok első r oszlopát tartalmazzák, és a $\mathbf{\Sigma}_r$ pedig egy $r \times r$ méretű, a nullától különböző szinguláris értékeket tartalmazó diagonális négyzetmátrix.

Az általánosított inverz tehát a Láncczos-féle felbontás alkalmazásával:

$$\mathbf{A}^{-\varepsilon} = (\mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T)^{-1} = \mathbf{V} \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{U}^T \quad (2.26)$$

Ahol a $\mathbf{\Sigma}^{-1}$ főátlójában az $1/\sigma_i$ szinguláris értékek reciprokai szerepelnek ($\sigma_i \neq 0$ esetekben).

Vizsgáljuk meg, hogy az SVD eljárással nyert inverzet alkalmazva, visszakapjuk-e az inverzmátrix (2.8) egyenlettel adott definíciójában szereplő egyenlőséget, vagyis az inverz mátrix és az eredeti mátrix szorzata az egységmátrixot eredményezi!

$$\mathbf{A}^{-\varepsilon} \cdot \mathbf{A} = \mathbf{V} \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T = \mathbf{V} \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T \quad (2.27)$$

Ez a kifejezés csak akkor ad egységmátrixot, ha a $\mathbf{\Sigma}^{-1} \cdot \mathbf{\Sigma} = \mathbf{I}$. Ezt a gyakorlatban ellenőrizve azt kapjuk, hogy $M \geq N$ esetben, (amikor a mátrix oszlopainak száma nagyobb vagy egyenlő az sorok számánál), akkor $\mathbf{A}^{-\varepsilon}$ jobboldali inverze $\mathbf{A}$ -nak, ha $N \geq M$ akkor a baloldali inverze. $\mathbf{A} \mathbf{A}^{-\varepsilon}$ csak akkor kétoldali inverze $\mathbf{A}$ -nak, ha $N = M = r$. Ekkor visszakaptuk a négyzetes mátrixokra érvényes inverz definíciójában szereplő mátrixot.

Ha az $\mathbf{A}$ mátrix felbontásában szerepelnek 0 szinguláris értékek, akkor az inverzet így definiálhatjuk:

$$\mathbf{A}^{-\varepsilon} = \mathbf{V}_r \cdot \mathbf{\Sigma}_r^{-1} \cdot \mathbf{U}_r^T \quad (2.28)$$

Ekkor $\mathbf{\Sigma}$ egy diagonális négyzetmátrix, amellyel fennáll a két egyenlőség:

$$\begin{aligned} \mathbf{A}^{-\varepsilon} \cdot \mathbf{A} &= \mathbf{V}_r \cdot \mathbf{V}_r^T \\ \mathbf{A} \cdot \mathbf{A}^{-\varepsilon} &= \mathbf{U}_r \cdot \mathbf{U}_r^T \end{aligned} \quad (2.29 - 2.30)$$

Az első egyenlet akkor egyenlő az egységmátrixszal, ha $M = r$, a második egyenlet pedig ha $N = m$.

3. fejezet - A valószínűesszámítás alapjai

A környezetünkről alkotott kép pontosításához méréseket végzünk. A méréseinkkel kapcsolatos számos tapasztalatot és jelenséget értelmezni tudunk a valószínűesszámítás és a matematikai statisztika segítségével. Tekintsük át röviden a valószínűesszámítás alapjait!

3.1. A valószínűség

A valószínűesszámítás **véletlen tömegjelenségekkel** foglalkozik. Elvileg eszerint egy mérés (kísérlet) végtelen sokszor elvégezhető.

Egy **kísérletnek** – mérésnek – több kimenetele lehetséges. Egy adott kimenet **elemi eseménynek** nevezzük, és ω -val jelöljük. Az elemi események összessége az Ω **eseménytér**. Az eseménytér valamely részhalmazát **összetett eseménynek** nevezzük, és a latin abc betűivel $(A, B, \dots)$ jelöljük.

Tegyük fel, hogy n -szer elvégezve a kísérletet k_A -szor következik be A esemény. Ekkor a k_A/n hányados az A esemény **relatív gyakoriságának** nevezzük. Többször is megismételve a kísérletsorozatot, mindig n -szer elvégezve a kísérletet, azt tapasztaljuk, hogy a kapott k_A/n hányadosok egy elméleti érték körül ingadoznak. Ezt az értéket $P(A)$ -val jelöljük, és az A **esemény valószínűségének** nevezzük.

A valószínűségekre fennáll:

$$0 \leq P(A) \leq 1 \quad (3.1)$$

Egy A esemény valószínűsége nulla, ha az esemény soha nem következik be, és 1 ha biztosan bekövetkezik.

Tekintsünk az A és B eseményeket. (Ezek az elemi események egy-egy részhalmazai.)

A két esemény összegén $(A+B)$ azt értjük, hogy legalább az egyik bekövetkezik. Két esemény szorzata $(A \cdot B)$ azt jelenti, hogy mindkettő bekövetkezik. Egy A esemény B (nem nulla valószínűségű) eseményre vonatkozó **feltételes valószínűségét** úgy értelmezzük, hogy annak a valószínűsége, hogy A esemény bekövetkezik, ha B esemény bekövetkezett. Ennek értéke kiszámítható az $A \cdot B$ és a B esemény valószínűségéből:

$$P(A|B) = \frac{P(A \cdot B)}{P(B)} \quad (3.2)$$

A feltételes valószínűsége igazak a következők, hogy a feltételes valószínűség is a $[0, 1]$ zárt intervallumon vehet fel értékeket, és egy esemény saját magára vett feltételes valószínűsége egységnyi: $P(B|B) = 1$.

Egy véges vagy megszámlálható $\{A_n\}$ ($n=1, 2, \dots$) eseményekből álló eseményrendszert **teljesnek** nevezzük, ha $i \neq j$ esetén $A_i \cdot A_j = 0$ (vagyis az események diszjunktak) és az A_n események közül pontosan egy biztosan (1 valószínűséggel) bekövetkezik. (A kapcsos zárójel a „halmaz” jele.)

$$P\left(\sum_n A_n\right) = \sum_n (P(A_n)) = 1 \quad (3.3)$$

(Vagyis a teljes eseményrendszer összetett eseményeinek összegének valószínűsége megegyezik az események valószínűségének összegével.)

Az Ω eseménytéren (az elemi események halmazán) értelmezett bármely függvényt **valószínűségi változónak** nevezzük. A méréseinket is úgy tekinthetjük, mint egy valószínűségi változót, vagyis egy elemi esemény realizációját, függvényét, amit a görög ξ kisbetűjével jelölünk.

$$\xi = \xi(\omega) \quad \text{ahol } \omega \in \Omega \quad (3.4)$$

Célszerűségi okokból több mérést együtt kezelhetünk, ekkor a méréseinket egy **valószínűségi vektorváltozó** elemeinek tekintjük: $\xi = (\xi_1, \xi_2, \dots, \xi_n)$

A valószínűségi vektorváltozót nevezik **többdimenziós valószínűségi változónak** is.

A valószínűségi változó lehet **folytonos** vagy **diszkrét**. Egy távolságmérés eredményét tekinthetjük folytonos változónak, bár egy mérőszalagot mm-es pontossággal tudunk leolvasni.

3.2. Valószínűségi változók eloszlásfüggvénye és sűrűségfüggvénye

Definiáljuk egy valószínűségi változó **eloszlásfüggvényét**!

Valamely ξ valószínűségi változó $F(x)$ eloszlásfüggvény az $(\xi \leq x)$ esemény valószínűségét leíró függvényt értjük:

$$F(x) = P(\xi \leq x) \quad (3.5)$$

A valószínűségi változókról és a relatív gyakoriságról mondottak alapján, ha egy mérést n -szer elvégzünk, akkor közelítőleg $n \cdot F(x)$ esetben fogunk x -nél kisebb, és $n \cdot [1 - F(x)]$ esetben x -nél nagyobb értéket mérni.

A későbbiekben meg fogunk vizsgálni néhány fontos eloszlásfüggvényt. Ezek az eloszlásfüggvények matematikailag jól kezelhető függvények, és – később bemutatott módon – kapcsolatba hozhatók a méréseinkkel. Ezeket eloszlásnak nevezzük, (pl. egyenletes-, normális-, log-normális-eloszlás).

Az eloszlásfüggvény tulajdonságai:

Az F eloszlásfüggvény értéke mindig 0 és 1 zárt intervallumba esik: $0 \leq F \leq 1$

Az F eloszlásfüggvény monoton nem csökkenő: $x_1 < x_2$ esetén $F(x_1) \leq F(x_2)$

Az eloszlásfüggvényre mindig teljesül: $\lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0$ és $\lim_{x \rightarrow \infty} F(x) = F(\infty) = 1$

Ha a ξ valószínűségi változó eloszlásfüggvénye F , akkor annak valószínűsége, hogy ξ a (c, d) intervallumba esik: $P(c \leq \xi \leq d) = F(d) - F(c)$

Ha az F függvény az $x = c$ helyen folytonos, akkor annak valószínűsége, hogy $\xi = c$, éppen: $P(\xi = c) = 0$

Diszkrét valószínűségi változók esetén az F függvény szakaszonként konstans függvény.

Folytonos valószínűségi változó sűrűségfüggvényét az alábbi módon definiáljuk:

Ha egy valószínűségi változóhoz tartozó F eloszlásfüggvény valamely intervallumban differenciálható, akkor ebben az intervallumban a sűrűségfüggvény az eloszlásfüggvény deriváltja, vagyis:

$$f = F' \quad (3.6)$$

A fenti definícióból adódik:

$$F(x) = \int_{-\infty}^x f(t) dt \quad (3.7)$$

A sűrűségfüggvény tulajdonságai:

- $f \geq 0$
- $\int_{-\infty}^{\infty} f(t) dt = 1$
- $P(c \leq \xi \leq d) = \int_c^d f(t) dt$

Definiálhatjuk egy ξ valószínűségi változó A eseményre vonatkozó feltételes valószínűségét leíró **feltételes eloszlásfüggvényen** az x változónak az alábbi függvényét értjük: $F(x|A) = P(\xi \leq x|A)$ feltéve hogy $P(A) > 0$.

Ha az $F(x|A)$ függvény x szerint differenciálható, akkor az $f(x|A) = F'(x|A)$ függvény a ξ változó A eseményre vonatkoztatott **feltételes sűrűség függvénye**.

Folytonos eloszlás esetén, eléggé kis Δx -et választva kapjuk, hogy:

$$P(x \leq \xi \leq x + \Delta x) = \int_x^{x+\Delta x} f(t) dt \approx f(x) \Delta x \quad (3.8)$$

Készítsünk egy diagramot, amely a sűrűségfüggvény empirikus közelítésére szolgál. Ehhez a diagram vízszintes tengelyén osszuk fel a ξ valószínűségi változó által felvehető értékek tartományát Δx intervallumokra. Végezzük el n -szer a kísérletet, és jegyezzük fel a kísérletek eredményeit. Jelölje k_x azoknak a kísérleteknek a számát, amikre

x -nél kisebb eredmény született. Az $(x, x+\Delta x)$ intervallum fölé rajzoljunk fel a $\frac{k_{x+\Delta x} - k_x}{n \cdot \Delta x}$ magasságú téglalapot. Ez a mennyiség a relatív gyakoriság osztva Δx -szel. Ezt a diagramot **hisztogramnak** nevezzük.

A ξ valószínűségi vektorváltozó eloszlásfüggvénye a

$$F(\mathbf{x}) = P(\xi \leq \mathbf{x}) \quad (3.9)$$

függvény, amely részletesen kiírva:

$$F(x_1, x_2, \dots, x_n) = P(\xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_n \leq x_n) \quad (3.10)$$

A $\xi_1, \xi_2, \dots, \xi_n$ valószínűségi vektorváltozó elemek közül kiválasztott $m < n$ számú valószínűségi változó együttes eloszlását m -dimenziós **peremeloszlásnak** nevezzük. A peremeloszlás az eredeti eloszlásfüggvényből úgy kapható meg, hogy a kiválasztottak között nem szereplő $n-m$ számú valószínűségi változó argumentuma helyére $+\infty$ kerül. A peremeloszlásokat a feltételes valószínűségek felírásához használhatjuk fel.

Ha a $\xi_1, \xi_2, \dots, \xi_n$ valószínűségi változók folytonosak, akkor értelmezhető az **valószínűségi vektorváltozó sűrűségfüggvénye** az alábbi módon:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n} \quad (3.11)$$

3.3. A valószínűségi változók jellemzői

Egy valószínűségi változó eloszlás vagy sűrűségfüggvénye pontosan jellemzi a változó viselkedését. Képezhünk azonban olyan mennyiségeket, amelyek segítségével egyszerűbben és szemléletes módon jellemezhetünk egy valószínűségi változót.

A várható érték

Egy valószínűségi változó várható értéke egy skalár mennyiség, ami a ξ valószínűségi változó függvénye. A várható érték képzés jele: $M(\xi)$

Folytonos változó esetén:

$$M(\xi) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (3.12)$$

Integrál adja meg az eloszlás várható értékét, diszkrét változó esetén, amikor a ξ változó az $x_1, x_2, \dots, x_n$ értékeket veheti fel, az egyes értékekhez $p_1, p_2, \dots, p_n$ valószínűségek rendelhetők, akkor a várható érték:

$$M(\xi) = \sum_{i=1}^n p_i \cdot x_i \quad (3.13)$$

A medián

A medián az az érték, amelynél a valószínűségi változó 0,5 valószínűséggel vesz fel kisebb értéket. Egy valószínűségi változó mediánjára igaz, hogy

$$F(m_e) = 0,5 \quad (3.14)$$

Ahol m_e a medián.

A szórás

A másik fontos mennyiség, amivel egy eloszlást jellemezhetünk a szórás és annak négyzete a **variancia**.

A szórásnégyzet definíció szerint a várható értéktől való eltérés négyzeteinek várható értéke, képlettel:

$$\sigma^2(\xi) = M\left(\left(\xi - M(\xi)\right)^2\right) \quad (3.15)$$

Ez kifejezhető:

$$\sigma^2(\xi) = M(\xi^2) - (M(\xi))^2 \quad (3.16)$$

Vagyis a változók négyzeteinek várható értékéből kivonva a várható érték négyzetét. Ez a mennyiség mindig pozitív.

Momentumok és centrális momentumok

A ξ valószínűségi változó k -adik momentuma:

$$\alpha_k = M(\xi^k) \quad (3.17)$$

A k -adik centrális momentum:

$$\alpha_k = M\left(\left(\xi - M(\xi)\right)^k\right) \quad (3.18)$$

A **második centrális momentum** a szórásnégyzet.

Két vagy több változó együttes eloszlását jellemző mennyiségek

Két valószínűségi változó, vagyis egy kételemű valószínűségi vektorváltozó várható értéke az egyes változó-elemek várható értéke:

$$M([\xi, \eta]) = [M(\xi), M(\eta)] \quad (3.19)$$

Ahol a szögletes zárójel a vektorképzés jele.

A fenti képlet szerint, egy valószínűségi vektorváltozó esetén a elemek várható értékei a **várható érték vektor** elemei.

Valószínűségi vektorváltozókra a vektoriális felírást alkalmazva, írhatjuk, hogy:

$$\mathbf{a} = M(\xi) = \int \mathbf{x} \cdot f(\mathbf{x}) d\mathbf{x} \quad (3.20)$$

Ahol $\mathbf{a}$ a várható érték vektor, $f(\mathbf{x})$ az együttes sűrűségfüggvény.

A szórást (3.16) egyenletben szereplő definícióját általánosíthatjuk két változóra, vagy egy **vektorváltozóra** is, úgy hogy a várható értéktől való eltérés négyzetét a vektorváltozó két elemének a saját várható értéküktől való eltérésének szorzatával helyettesítjük:

$$c_{ij} = M\left(\left(\xi_i - M(\xi_i)\right) \cdot \left(\xi_j - M(\xi_j)\right)\right) \quad (3.21)$$

a c_{ij} mennyiségek egy mátrix, a **kovariancia-mátrix** ($\mathbf{C}$) elemei. A mátrix főátlójában az egyes változók szórásnégyzetei (varianciái) állnak.

$$\sigma^2(\xi_i) = c_{ii} \quad (3.22)$$

A kovariancia-mátrix mindig szimmetrikus ($c_{ij} = c_{ji}$). A kovariancia-mátrix egy nem diagonális eleme (c_{ij}) a valószínűségi vektorváltozó vektor két elemének (ξ_i, ξ_j) az együttes varianciáját jellemzik. A mátrix valamelyik sora tehát az adott indexű változónak a többi változóhoz való viszonyát jellemzik.

A fenti képlet a vektoriális felírás alkalmazásával:

$$\mathbf{C} = M\left((\xi - \mathbf{a}) \cdot (\xi - \mathbf{a})^T\right) \quad (3.23)$$

A kovariancia-mátrix egyes sorait osszuk végig az sor i indexével azonos indexű változó szórásértékeivel, $\sigma(\xi_i)$ -vel! A kapott mátrix oszlopait osszuk végig az adott oszlop j indexének megfelelő szórásértékkel, $\sigma(\xi_j)$ -vel :

$$r_{ij} = \frac{c_{ij}}{\sigma_i \cdot \sigma_j} = \frac{c_{ij}}{\sqrt{c_{ii}} \cdot \sqrt{c_{jj}}} \quad (3.24)$$

Az így kapott mátrixelemek az ún. **korrelációs mátrix (R)** elemei. Ennek főátlójában – a definícióból könnyen levezethetően – 1-ek állnak, a nem-diagonális elemekre pedig érvényes hogy: $-1 \leq r_{ij} \leq 1$

A korrelációs mátrix elemei azt fejezik ki, hogy két valószínűségi változó mennyire „korrelál”, vagyis mennyire erős kapcsolatban vannak egymással.

A főátlóban szereplő 1 értékek azt fejezik ki, hogy egy mennyiség saját magával tökéletesen korrelál. A nem diagonális helyen levő, 1-hez közeli érték azt fejezi ki, hogy a két változó (amelynek indexeit a sor és az oszlop sorszáma adja meg) erősen **korrelál**, a -1 közeli érték erős **negatív korrelációt** jelent.

Ha két valószínűségi változó **független**, (vagyis egymásra vonatkozó feltételes valószínűségük 0) akkor a köztük levő korrelációs együttható értéke 0. (Fordítva azonban nem igaz, két változó 0 korrelációja nem jelenti a két változó függetlenségét.)

3.4. Nevezetes eloszlások

Nagyon sokféle eloszlás létezik, ezek közül csak néhány – az inverzióelmélet szempontjából – kiemelkedően fontos eloszlást mutatunk be.

Egyenletes eloszlás

Egy (a, b) intervallumon egyenletes eloszlás eloszlásfüggvénye:

$$F(x) = \frac{x-a}{b-a} \quad (3.25)$$

sűrűségfüggvénye pedig

$$\begin{aligned} f(x) &= \frac{1}{b-a} \quad \text{ha } (a \leq x \leq b) \\ f(x) &= 0 \quad \text{egyébként} \end{aligned} \quad (3.26)$$

Várható értéke $(a+b)/2$, szórása $(b-a)/\sqrt{12}$

Laplace (kétoldali exponenciális) eloszlás

A földtudományokban gyakran alkalmazott eloszlás, mert gyorsabban cseng le, mint a normális eloszlás. A standard alak sűrűségfüggvénye a következő:

$$f(x) = \frac{1}{2} e^{-|x|} \quad (3.27)$$

Az általános alak:

$$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \quad (3.28)$$

Ahol a $b(>0)$ az eloszlás skála paramétere és μ a helyparaméter. Az eloszlás várható értéke μ a szórása pedig $\sqrt{2} \cdot b$.

Normális eloszlás

A normális eloszlás a valószínűségszámításban és statisztikában központi jelentőségű.

Egy ξ valószínűségi változó normális eloszlású, ha eloszlásfüggvénye a következő:

$$F(x) = P(\xi < x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt \quad (3.29)$$

Az egyváltozós normális eloszlást szokásos jelölése $N(a, \sigma)$, ahol a az eloszlás várható értéke, σ pedig a szórása. Az eloszlást a várható érték és a szórás egyértelműen megadja. A normális eloszlás másik neve **Gauss eloszlás**.

A normális eloszlás sűrűségfüggvénye az eloszlásfüggvény deriváltja:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad (3.30)$$

Az $N(a, \sigma)$ eloszlás sűrűségfüggvényének képe a haranggörbe (Gauss-görbe). A haranggörbe szimmetriatengelye az a várható értékhez esik, és alakját a σ szórás adja meg. A görbe inflexió pontja σ távolságra van a görbe tengelyétől.

A normális eloszlású ξ valószínűségi változó **standardizált**ja a

$$\xi^* = \frac{\xi - a}{\sigma} \quad (3.31)$$

Valószínűségi változó, amelynek várható értéke 0, és szórása 1. Ennek eloszlás- és sűrűségfüggvénye:

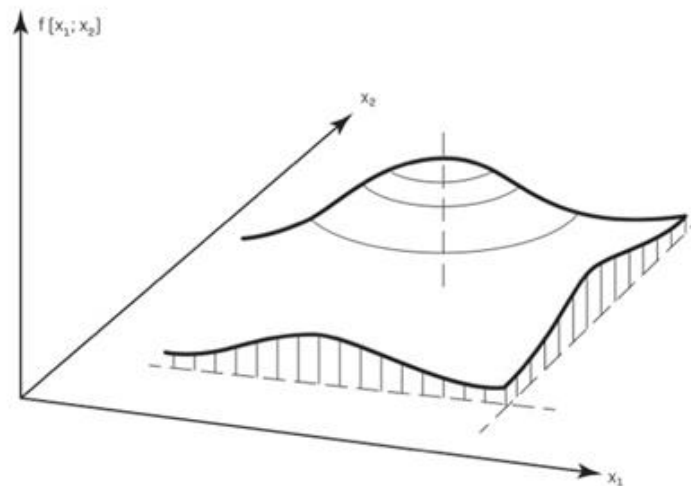
$$\begin{aligned} \Phi(u) &= P(\xi^* < u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt \\ \phi(u) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \end{aligned} \quad (3.32-3.33)$$

Többdimenziós (többváltozós) normális eloszlás

A $\xi^* = (\xi_1^*, \xi_2^*, \dots, \xi_n^*)$ valószínűségi vektorváltozó együttes eloszlása akkor n -dimenziós normális eloszlás, ha folytonos, és sűrűségfüggvénye a következő alakú:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \mathbf{C}}} \cdot e^{-\frac{1}{2} \{ \mathbf{x} - \mathbf{a} \}^T \mathbf{C}^{-1} \{ \mathbf{x} - \mathbf{a} \}} \quad (3.34)$$

Ahol $\mathbf{a} = (a_1, a_2, \dots, a_n)$ a várható értékek vektora, $\mathbf{C}$ pedig a (3.21) képlettel definiált kovariancia mátrix ($\det \mathbf{C} \neq 0$). A $\mathbf{C}^{-1}$ jelölés a kovariancia-mátrix inverzét jelenti.



3.1. ábra. Kétváltozós normális eloszlás sűrűségfüggvénye

χ^2 eloszlás

Legyenek $(\xi_1, \xi_2, \dots, \xi_n)$ független normális eloszlású, 0 várható értékű és 1 szórású valószínűségi változók. Képezzük egy új valószínűségi változót az alábbi módon:

$$\chi_n^2 = \sum_{i=1}^n \xi_i^2 \quad (3.35)$$

Ennek a valószínűségi változónak a sűrűségfüggvénye:

$$k_n(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (3.36)$$

Ahol $\Gamma(x)$ az ún. gammafüggvény:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (3.37)$$

Az összegzésben megjelenő n számot az eloszlás **szabadsági fokának** nevezzük. χ^2 várható értéke n , szórásnégyzete pedig $2n$.

Student eloszlás

Legyenek $\xi_1, \xi_2, \dots, \xi_n$ független, normális eloszlású, 0 várható értékű és 1 szórású valószínűségi változók, az η pedig egy 0 várható értékű, $\sqrt{n}$ szórású normális eloszlású valószínűségi változó. Képezzük az alábbi valószínűségi változót:

$$t_n = \frac{\sqrt{n} \cdot \eta}{\sqrt{\xi_1^2 + \xi_2^2 + \dots + \xi_n^2}} \quad (3.38)$$

A t_x valószínűségi változót **Student-törtnék** nevezzük. Ennek sűrűségfüggvénye a

$$s_n = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (3.39)$$

Várható értéke 0, szórása pedig: $n/(n-2)$. (Itt n értékét ugyancsak szabadsági foknak nevezzük.) Ebből következik, hogy $n = 1$ és $n = 2$ esetben, bár a sűrűségfüggvény létezik, nem létezik szórás.

3.5. Határértéktételek, nagy számok törvényei

Az előző fejezetben szó volt arról, hogy ha egy kísérletet egymástól függetlenül elég sokszor elvégzünk, és egy A eseményt kiválasztunk, akkor a k_A/n relatív gyakoriság stabilitást mutat. Ezt a tapasztalati adatokkal alátámasztott tényt a nagy számok törvényének nevezzük, és ez teszi lehetővé a valószínűségelmélet gyakorlati alkalmazását.

A valószínűségelmélet azonban a gyakorlattól függetlenül is létezik. Az eddig definiált fogalmakat felhasználva nagy számok törvényének nevezünk minden olyan tételt, ami a $\xi_1, \xi_2, \dots, \xi_n$ valószínűségi változók számtani átlagaiból alkotott sorozat:

$$\eta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} \quad (3.40)$$

valamilyen konvergenciáját állítja adott feltételek mellett.

A centrális határeloszlás tétele.

A centrális (központi) határeloszlás tétele azt mondja ki, hogyha nagyszámú, független valószínűségi változót összeadunk, amelyeknek a szórásnégyzete véges, akkor összegük a normális eloszláshoz tart, függetlenül attól, hogy az egyes valószínűségi változók milyen eloszlásúak. (Szemléletes példa erre, több $(-0,5 - 0,5)$ intervallumon egyenletes eloszlású valószínűségi változó összegéből képzett valószínűségi változó. Ennek sűrűségfüggvénye a Gauss görbéhez tart. Gauss ennek segítségével vezette le a normális eloszlás sűrűségfüggvényét.)

Vizsgáljuk most az $\xi_1, \xi_2, \dots, \xi_n$ valószínűségi változók összegéből alkotott $\eta_n = \xi_1 + \xi_2 + \dots + \xi_n$ sorozatot. Jegyezzük meg, hogy amennyiben a $M(\xi_1) \neq 0$, akkor határeloszlást sem határozhatunk meg $n \rightarrow \infty$ esetén, mivel η egyre jobban elkenődik, azonban vegyük észre, hogy:

$$\begin{aligned} M(\eta_n) &= nM(\xi_1) \\ \sigma^2(\eta_n) &= n\sigma^2(\xi_1) \end{aligned} \quad (3.41 - 3.42)$$

Valamint tetszőleges véges (a, b) intervallumra igaz hogy:

$$\lim_{n \rightarrow \infty} P(a \leq \eta_n \leq b) = 0 \quad (3.43)$$

Ha azonban η_n helyett ennek standardizáltját vizsgáljuk:

$$\eta_n^* = \frac{\eta_n - n \cdot m}{\sigma \cdot \sqrt{n}} \quad (3.44)$$

Ahol $m = M(\xi_k)$ és $\sigma = \sigma(\xi_k)$, $k = 1, 2, \dots$ akkor erre bebizonyítható, hogy

$$\lim_{n \rightarrow \infty} P(a \leq \eta_n^* \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \quad (3.45)$$

Vagyis elég nagy n esetén tehát η_n^* standard normális eloszlású ($N(0,1)$). Általában η_n^* közelítően normális eloszlásúvá válik aránylag kis n esetén.

Itt jegyezzük meg, hogy véges szórású valószínűségi változók **szorzatából** képzett valószínűségi változó ($\eta_n = \xi_1 \cdot \xi_2 \cdot \dots \cdot \xi_n$) eloszlása a **log-normális eloszláshoz** tart, amelynek eloszlásfüggvénye:

$$F(x) = \Phi\left(\frac{\ln(x) - a}{\sigma}\right) \quad (3.46)$$

sűrűségfüggvénye:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-a)^2}{2\sigma^2}}, \quad (x > 0) \quad (3.47)$$

ahol a a várható érték és σ a szórás.

Előírt pontosságú közelítéshez szükséges kísérletszám meghatározása

A centrális határeloszlás tétel segítségével kiszámolhatjuk, hogy hány mérés szükséges ahhoz, hogy ismert (vagy meghatározható) szórású eloszlásból származó valószínűségi változók átlaga legfeljebb $(1-p_0)$ valószínűséggel térjen el az eloszlás tényleges várható értékétől.

Tekintsünk egy ξ_1, ξ_2 , azonos eloszlású, véges szórású és független valószínűségi változókból álló sorozatot. Jelölje m ezek közös várható értékét. A standardizáltjukra (3.31) a (3.45) egyenlet segítségével felírható:

$$P\left(\left|\frac{\xi_1 + \xi_2 + \dots + \xi_n - nm}{\sigma\sqrt{n}}\right| < \lambda\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-\frac{x^2}{2}} dx = 2 \cdot \Phi(\lambda) - 1 \quad (3.48)$$

Tehát elég nagy n esetén:

$$P\left(\left|\frac{\xi_1 + \xi_2 + \dots + \xi_n - nm}{\sigma\sqrt{n}}\right| < \lambda\right) \approx 2 \cdot \Phi(\lambda) - 1 \quad (3.49)$$

Ennek átrendezésével adódik:

$$P\left(\left|\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} - m\right| < \lambda \cdot \frac{\sigma}{\sqrt{n}}\right) \approx 2 \cdot \Phi(\lambda) - 1 \quad (3.50)$$

Ha tehát előírt p_0 esetén λ -t meghatározzuk oly módon, hogy $1 - p_0 = 2 \cdot \Phi(\lambda) - 1$, majd az előírt ε -hoz n értékét

olyan nagyra választjuk, hogy $\lambda \cdot \frac{\sigma}{\sqrt{n}} \leq \varepsilon$, akkor azt kapjuk, hogy

$$P\left(\left|\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} - m\right| < \varepsilon\right) \geq 1 - p_0 \quad (3.51)$$

Ehhez tehát n -nek ki kell elégítenie az alábbi feltételt:

$$n \geq \lambda^2 \cdot \frac{\sigma^2}{\varepsilon^2} \quad (3.52)$$

Az eredmény felhasználása során az előírt p_0 valószínűséghez kiszámítjuk λ -t a fenti módon, majd az előírt ε pontosság és a változó σ szórása segítségével kiszámoljuk n -t a mérések minimális számát.

3.6. Statisztikai sokaság

A statisztikai vizsgálat tárgyát képező egyedek összességét a hozzájuk rendelhető számértékekkel együtt **statisztikai sokaságnak** nevezzük. A statisztikai sokaság elemeit egy halmaz elemeinek tekinthetjük, és a hozzájuk rendelhető számértékeket pedig ezen elemeken értelmezett függvényként. Egy sokaság lehet véges, vagy végtelen számú. A későbbiekben egy adott mennyiségre többször elvégzett mérést egy elvileg végtelen sok elemű sokaság egy véges számú elemet tartalmazó részhalmazának tekinthetünk.

A mérést tehát egy elem kiválasztásának tekinthetjük egy statisztikai sokaságból. Egy elem kiválasztása egy halmazból matematikai szempontból a halmaz részhalmazain egy valószínűség eloszlás értelmezését jelenti. Eszerint egy statisztikai sokaság valószínűségi mezővé, az egyeden értelmezett számértékekből felépülő függvény pedig egy ξ valószínűségi változóvá lesz. Ennek a ξ valószínűségi változónak az eloszlását a **statisztikai sokaság eloszlásának** nevezzük.

A statisztikai sokaság eloszlásának, vagy egyes **paramétereinek** (pl. várható érték, szórás) meghatározására statisztikai vizsgálatot végzünk. Ehhez a sokaságból mintát veszünk. A **mintavétel** a sokaság n elemének véletlenszerű kiválasztásából áll.

Jelöljük $x_1, x_2, \dots, x_n$ a kiválasztott elemekhez tartozó számértékeket a kiválasztás sorrendjében. Kiválaszthatunk visszatevéses és visszatevés nélküli módon, minket csak a visszatevéses mintavétel érdekel, mivel a geofizikai méréseink eleget tesznek ennek a kritériumnak.

Az egymástól független és ξ -vel megegyező eloszlású $x_1, x_2, \dots, x_n$ valószínűségi változók összességét **n -elemű mintának** nevezzük. Ha ξ eloszlásfüggvénye $F(x)$ akkor azt mondjuk, hogy $x_1, x_2, \dots, x_n$ egy, az $F(x)$ eloszlású sokaságból vett (n -elemű véletlen) minta. Az x_i valószínűségi változókat **mintaelemeknek** nevezzük.

A mintavételezés – mérés – célja, hogy valamit megismerjünk. A **statisztikai következtetések** alapelve megegyezik a logikai következtetésekkel, azzal a különbséggel, hogy a következményt nem logikai bizonyossággal állítjuk, hanem csak valamilyen – általában 1-hez közeli – valószínűséggel. (Ebből következik, hogy adott esetben tévedhetünk is.)

Egy $x_1, x_2, \dots, x_n$ elemekből álló mintára meghatározhatjuk a **tapasztalati (empirikus) eloszlást**, (ezt a mintaelemekből képzett hisztogram adja meg) valamint az empirikus jellemző adatokat, amelyek közül a legfontosabbakat megadjuk.

Az **empirikus várható érték**, $\bar{x}$ (amit a mintaelemek átlagaként kapunk meg):

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (3.53)$$

Az **empirikus szórásnégyzetet** (s^2 -t):

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} \quad (3.54)$$

Az empirikus mennyiségek bevezetésével értelmet nyer, hogy bevezessük az **elméleti jellemző adatokat**, amik az adott statisztikai sokaságot jellemzik, és amiknek a meghatározása a célunk.

A tapasztalat szerint statisztikai sokaságból vett kis számú minta esetén az empirikus szórásnégyzet várható értéke nem egyezik meg a statisztikai sokaság elméleti szórásnégyzetével. Ezért az elméleti szórásnégyzet közelítésére a **korrigált tapasztalati szórásnégyzetet** használjuk:

$$s^{*2} = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3.55)$$

Amennyiben a minta **valószínűségi vektorváltozókat** tartalmaz, és a vektorváltozó k elemű, a minta pedig n -elemű, akkor a vektorváltozó egyes elemeiből is képezhetjük az egyes elemek empirikus várható értékét, amit egy k -elemű vektorba rendezhetünk:

$$\bar{\mathbf{x}} = (\bar{x}^{(1)}, \bar{x}^{(2)}, \dots, \bar{x}^{(k)}) \quad (3.56)$$

Ennek segítségével a korrigált tapasztalati szórásnégyzet mintájára képezhetjük a $k \times k$ méretű **korrigált tapasztalati kovarianci mátrixot**, amelynek elemeit a

$$c_{ij}^* = \frac{1}{n-1} \sum_{i=1}^n (x_i^{(i)} - \bar{x}^{(i)}) \cdot (x_i^{(j)} - \bar{x}^{(j)}) \quad (3.57)$$

képlet segítségével definiálhatjuk.

A statisztikai becslések

A statisztikai minta alapján a statisztikai sokaság eloszlását akarjuk meghatározni. Az eloszlás, vagy valamilyen jellemző mennyiségének (paraméterének) meghatározási eljárását **becslésnek** nevezzük. Választva egy ismert eloszlású statisztikai sokaságot, amelynek valamely paraméterére kíváncsiak vagyunk, a sokaságból vett minta elemein értelmezett

$$\hat{a} = \hat{a}(x_1, x_2, \dots, x_n) \quad (3.58)$$

függvényt használjuk az a paraméter „valódi” értékének becslésére. Ezt a tetszőleges függvényt **statisztikai függvénynek**, vagy **statisztikának** nevezzük. Minden becslés valószínűségi változót eredményez, amelynek van eloszlása. Az a paraméter $\hat{a}$ becslése annál jobb, minél inkább koncentrálódik $\hat{a}$ eloszlása az a paraméter valódi értéke körül. Ezt konkretizálva, a statisztikákkal szemben az alábbi elvárásokat fogalmazhatjuk meg:

Az a paraméter $\hat{a}$ becslését **torzítatlannak** nevezzük, ha az $\hat{a}$ várható értéke a -val egyenlő: $M(\hat{a}) = a$. Eszerint például egy statisztikus sokaságból vett mintán, a korrigált empirikus szórásnégyzet kiszámítása a statisztikai sokaság szórásnégyzetének torzítatlan becslését szolgáltatja.

Ha $\hat{a}_1$ és $\hat{a}_2$ az a paraméter torzítatlan becslése, és $\sigma^2(\hat{a}_1) < \sigma^2(\hat{a}_2)$, akkor az $\hat{a}_1$ becslést az $\hat{a}_2$ becslésnél **hatásosabbnak** (effektívebbnek) nevezzük.

Ha van olyan $\hat{a}_0$ torzítatlan becslés, amelynek szórása minimális az a paraméter összes torzítatlan becslése körében akkor ezt **hatásos becslésnek** nevezzük. (A becslés paraméter szórása általában nem csökkenthető minden határon túl.)

Mivel egy adott becslés különböző elemszámokra is alkalmazható, a minta n elemszámának növelésével nem egyetlen becslésünk, hanem egy **becslés-sorozatunk** van. Az a paraméter egy $\hat{a}_1, \hat{a}_2, \hat{a}_3, \dots$ becsléssorozatát **aszimptotikusan torzítatlannak** nevezzük, ha

$$\lim_{n \rightarrow \infty} M(\hat{a}_n) = a \quad (3.59)$$

Példa erre az empirikus szórásnégyzetekből alkotott becsléssorozat, ami aszimptotikusan torzítatlan becslése az elméleti szórásnégyzetnek.

Egy becslési eljárást **konzisztensnek** nevezünk, ha a mérések n számának növekedésével a paraméterek becslött értékei a valódi értékekhez tartanak (sztochasztikus értelemben):

$$\lim_{n \rightarrow \infty} P(|\hat{a}_n - a| \geq \varepsilon) = 0 \quad (3.60)$$

minden n -re és pozitív ε -ra.

Konfidencia intervallumok

Említettük, hogy a statisztikai következtetések esetén a következményt csak valamilyen (1-hez közeli) valószínűséggel állíthatjuk.

Általában lehetőségünk van az $x_1, x_2, \dots, x_n$ elemekből álló mintára támaszkodva olyan $\hat{a}_1$ és $\hat{a}_2$ statisztikák konstruálására, amelyre teljesül, hogy:

$$P(\hat{a}_1 \leq a \leq \hat{a}_2) = 1 - p \quad (3.61)$$

Ahol p egy általunk megválasztott pozitív szám, amelytől az $\hat{a}_1$ és $\hat{a}_2$ statisztikák függenek. Jelen esetben az a paramétert egy intervallummal becsüljük, ezt **intervallumbecslésnek** nevezzük. Az $(\hat{a}_1, \hat{a}_2)$ véletlen helyzetű intervallumot **konfidencia-(megbízhatósági) intervallumnak**, az $(1-p) \cdot 100\%$ -ot megbízhatóság szintjének, az intervallum kezdő- és végpontját pedig **konfidencia határoknak** nevezzük.

3.7. Maximum likelihood elv

A becslések kapcsán felvetődik a kérdés, hogy hogyan lehet olyan becsléseket konstruálni, amelyek a becslésekkel szemben megfogalmazott fenti kívánságaink közül a legtöbbet teljesíti. Általánosságban azt fogalmazhatjuk meg, hogy adott kísérleti anyag (minta, vagyis mérések) esetén a paraméterek azon értékeit fogadjuk el legjobb becslésnek, amelyeknek a valószínűsége a maximális.

Ezt egy diszkrét valószínűségi eloszlás esetén, annak valószínűsége, hogy egy a paraméterű statisztikai sokaságból származó minta pont x értéket vesz fel:

$$P(\xi = x) = p(x, a) \quad (3.62)$$

Egy konkrét $x_1, x_2, \dots, x_n$ értékekből álló minta esetén az adott minta és az a paraméter együttes valószínűsége:

$$L = p(x_1, a) \cdot p(x_2, a) \cdot \dots \cdot p(x_n, a) \quad (3.63)$$

Az egyes valószínűségek szorzatából álló L függvényt **likelihood függvénynek** nevezzük. Ez tehát annak a valószínűsége, hogy éppen az adott mintát kapjuk véletlen mintavétel során. Ez az adott mintaértékek és az a paraméter együttes valószínűsége. Egy konkrét minta esetén, ez az a paraméter valószínűsége. Konkrét minta esetén tehát az a paraméter különböző értékeihez különböző valószínűség tartozik.

Ezt felhasználhatjuk az a paraméter becslésére: legyen az a paraméter becslése az az érték, ahol a fenti függvénynek maximuma van.

A **maximum likelihood** becslés során az a paraméter valódi értékét azzal a speciális $\hat{a}$ értékkel becsüljük, amely – ha a paraméter valódi értéke volna – akkor éppen az adott minta bekövetkezése volna a legvalószínűbb az összes lehetséges n -elemű minták közül.

Folytonos eloszlású sokaság esetén a

$$L = f(x_1, a) \cdot f(x_2, a) \cdot \dots \cdot f(x_n, a) \quad (3.64)$$

Függvény maximumát keressük, ahol az $f(x_i, a)$ függvény a sokaság sűrűségfüggvénye. (A sűrűségfüggvény definíciójából fakadóan egy konkrét x_i értékű minta bekövetkezésének a valószínűsége nulla, azonban rögzített $\Delta x_1, \Delta x_2, \dots, \Delta x_n$ esetén beszélhetünk annak valószínűségéről, hogy a minta elemei rendre a $(x_1, x_1 + \Delta x_1), (x_2, x_2 + \Delta x_2), \dots, (x_n, x_n + \Delta x_n)$ intervallumokba essenek. Ez a valószínűség kis Δx -ek esetén

$$f(x_1, a) \cdot f(x_2, a) \cdot \dots \cdot f(x_n, a) \cdot \Delta x_1 \cdot \Delta x_2 \cdot \dots \cdot \Delta x_n \quad (3.65)$$

Valószínűséggel egyenlő. Ekkor az a paraméter valódi értéke becsléseként azt az értéket fogadjuk el, melyet a helyébe téve, a fenti szorzat maximális.

A fenti szorzatok maximumhelyének meghatározásához felhasználhatjuk, hogy a függvények az a szerint differenciálhatók. Elsősorban célszerűségi okokból, a fenti szorzatfüggvénynek a logaritmusát véve (a folytonos esetben):

$$\ln L = \sum_{i=1}^n \ln f(x_i, a) \quad (3.66)$$

Látjuk, hogy a szorzat helyett a fenti képletben a sűrűségfüggvények logaritmusának összege áll. Mivel a logaritmusfüggvények monoton függvények, a fenti (3.66) függvény maximumhelye megegyezik a megfelelő likelihood-függvény (3.65) maximumhelyével.

Ekkor az a paraméter becslését az

$$\frac{\partial \ln L}{\partial a} = 0 \quad (3.67)$$

ún. Likelihood-egyenlet a -ra való megoldásával határozhatjuk meg. (Feltéve, hogy az $\ln L$ az a változó szerint differenciálható, és a fenti egyenlet a maximumhelyet szolgáltatja, vagyis

$$\frac{\partial^2 \ln L}{\partial a^2} < 0 \quad (3.68)$$

feltétel is teljesül a minimumhelyen.)

Ha létezik az a paraméternek egy $\hat{a}$ minimális szórású vagyis hatásos (effektív) becslése, akkor a likelihood egyenletnek egy megoldása van, és az egyenlő $\hat{a}$ -val.

A maximum likelihood elv alkalmazását nézzük meg ismeretlen m várható érték és σ^2 szórásnégyzet paraméterekkel jellemezhető normális eloszlásból vett minta paramétereinek becslésére. A sűrűségfüggvény a mostani jelölésekkel:

$$f(x, m, a) = \frac{1}{\sqrt{2\pi a}} \cdot e^{-\frac{(x-m)^2}{2a}} \quad (3.69)$$

Amiből a likelihood függvény logaritmusa:

$$\ln L = \ln \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi a}} \cdot e^{-\frac{(x_i-m)^2}{2a}} \right) = -\frac{n}{2} (\ln 2\pi + \ln a) - \sum_{i=1}^n \frac{(x_i-m)^2}{2a} \quad (3.70)$$

Ebből m és a szerinti parciális differenciálással kapjuk a likelihood-egyenletet:

$$\frac{\partial \ln L}{\partial m} = \sum_{i=1}^n \frac{x_i - m}{a} = 0 \quad \text{és} \quad (3.71)$$

$$\frac{\partial \ln L}{\partial a} = -\frac{n}{2a} + \sum_{i=1}^n \frac{(x_i - m)^2}{2a^2} = 0 \quad (3.72)$$

Ebből a két egyenletből:

$$\hat{m} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \quad (3.73)$$

valamint

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{m})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2 \quad (3.74)$$

Vagyis ha $a=\sigma^2$ ismert, és csak m -et becsüljük, akkor az $\bar{x}$ (empirikus várható érték) adódik, ha viszont m -et ismerjük, és csak a -t becsüljük, akkor

$$\hat{a} = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2 \quad (3.75)$$

Ami torzítatlan becslése σ^2 -nak.

Vizsgáljuk meg azt az esetet, amikor a mintáról azt feltételezzük, hogy azonos várható értékű, de különböző szórású normális eloszlásból származnak. Ekkor az i -ik valószínűségi változó várható értéke a , szórása pedig σ_i . Ekkor a likelihood függvény az alábbi alakú:

$$L(x_1, x_2, \dots, x_n) = \frac{1}{\prod_{i=1}^n \sqrt{2\pi\sigma_i^2}} \cdot e^{-\sum_{i=1}^n \frac{(x_i-a)^2}{2\sigma_i^2}} \quad (3.76)$$

Aminek a logaritmusa:

$$\ln L(x_1, x_2, \dots, x_n) = -\sum_{i=1}^n \ln(\sqrt{2\pi\sigma_i^2}) - \sum_{i=1}^n \frac{x_i^2}{2\sigma_i^2} + a \cdot \sum_{i=1}^n \frac{x_i}{\sigma_i^2} - \frac{a^2}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \quad (3.77)$$

Az a paraméter szerinti derivált

$$\frac{\partial \ln L}{\partial a} = \sum_{i=1}^n \frac{x_i}{\sigma_i^2} - a \cdot \sum_{i=1}^n \frac{1}{\sigma_i^2} = 0 \quad (3.78)$$

amiből következik:

$$\hat{a} = \frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}} \quad (3.79)$$

Vagyis a várható érték becslésére a **súlyozott átlagot** használjuk.

A XIX. század elején ismert volt Gauss tétele, hogy a szórásnégyzetek reciprokával súlyozott átlag a közös várható érték minimális szórású becslése a lineáris becslések között. A maximum likelihood elvből levezethettük, hogy ennek szórása az összes becslések közül is a minimális.

A maximum likelihood elvet alkalmazhatjuk más (nem normális) eloszlásokra is. Nézzünk erre két példát!

Laplace (kétoldali exponenciális) **eloszlásból** vett mintán, az eloszlás μ paraméterének – a várható értéknek – véges minta esetén effektív, torzítatlan becslése a **tapasztalati medián**: A minta elemeit növekvő sorba rendezzük, és a középső (páros mintaszám esetén valamelyik mellette levő) elem a becsült érték.

Egyenletes eloszlás esetén a várható érték effektív, torzítatlan becslését úgy kapjuk, ha a mintaelemeket növekvő sorrendbe rendezzük, és a legnagyobb és legkisebb mintaelem számtani átlagát tekintjük a várható érték becslésének.

3.8. Valószínűségi változók függvényének eloszlását jellemző mennyiségek

Tekintsünk egyetlen mérést. Ezt a valószínűségelméleti bevezető alapján egy valószínűségi változó realizációjának tekinthetjük. Ebből a mérésből számoljunk ki egy minket érdeklő mennyiséget. Ez a mennyiség ugyancsak valószínűségi változó lesz. Eszerint a minket érdeklő mennyiség, egy valószínűségi változó függvénye.

Ahogy korábban láttuk, egy valószínűségi változót legteljesebb mértékben az eloszlásfüggvénye jellemez, de lehet, hogy minket csak valamely jellemzője, pl. várható értéke, vagy szórása érdekel.

A Gauss féle hibaterjedési törvény azt mondja ki, hogy amennyiben van egy ξ valószínűségi változónk, ami egy ismert eloszlásból származik (ismert az $f(x)$ sűrűségfüggvény), és adott a változó $y = \Psi(x)$ függvénye, amely monoton csökkenő vagy növekvő, de mindenképpen differenciálható, amelynek inverze: $x = \Psi^{-1}(y) = h(y)$ akkor az $\eta = \Psi(\xi)$ valószínűségi változó is folytonos eloszlású, amelynek a sűrűségfüggvénye:

$$g(y) = f[h(y)] \cdot \left| \frac{dh(y)}{dy} \right| \quad (3.80)$$

A fenti tételt alkalmazzuk egy ξ standardizált normális eloszlású változóra, amelynek sűrűségfüggvénye:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (3.81)$$

És határozzuk meg az $y = x^2$ függvénnyel megadható valószínűségi változó sűrűségfüggvényét!

Ennek a függvénynek az inverz függvénye a következő:

$$x = h(y) = y^{\frac{1}{2}} \quad (3.82)$$

(a négyzetgyököknek csak a pozitív értékét vesszük figyelembe a továbbiakban).

A h inverz függvény segítségével írjuk fel a sűrűségfüggvényt, mint az y változó függvényét!

$$f[h(y)] = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \quad (3.83)$$

A h inverz függvény deriváltjának az abszolút értéke:

$$\left| \frac{dh(y)}{dy} \right| = \frac{1}{2} \cdot \frac{1}{\sqrt{y}} \quad (3.84)$$

A két függvény szorzata adja a keresett sűrűségfüggvényt:

$$g(y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y}{2}} \cdot \frac{1}{2} \cdot \frac{1}{\sqrt{y}} = \frac{1}{2\sqrt{2\pi y}} \cdot e^{-\frac{y}{2}} \quad (3.85)$$

Vizsgáljuk meg, hogy hogyan alakul valamely valószínűségi vektorváltozók függvényének a várható értéke! Ehhez válasszunk egy $\xi = (\xi_1, \xi_2, \dots, \xi_N)$ valószínűségi vektorváltozót, amelynek tekintsük egy lineáris függvényét (transzformáltját):

$$\eta = A \cdot \xi \quad (3.86)$$

Ahol az η ugyancsak valószínűségi vektorváltozó. Az A mátrix a lineáris transzformációt leíró mátrix. Az η vektorváltozó várható értékének képlete:

$$\mathbf{b} = M(\eta) = M(A \cdot \xi) = \int A \cdot \mathbf{x} \cdot f(\mathbf{x}) d\mathbf{x} = A \cdot \int \mathbf{x} \cdot f(\mathbf{x}) d\mathbf{x} = A \cdot \mathbf{a} \quad (3.87)$$

Ahol $\mathbf{a}$ a ξ vektorváltozó elemeinek várható értékeinek vektora.

Vizsgáljuk meg, hogy hogyan alakul valamely valószínűségi vektorváltozó függvényének a kovariancia mátrixa! Használjuk fel a kovarianciamátrix vektoros felírását!

$$\begin{aligned} \mathbf{C}_\eta &= M((\eta - \mathbf{b}) \cdot (\eta - \mathbf{b})^T) = M(A \cdot (\xi - \mathbf{a}) \cdot (\xi - \mathbf{a})^T \cdot A^T) = A \cdot M((\xi - \mathbf{a}) \cdot (\xi - \mathbf{a})^T) \cdot A^T = \\ &= A \cdot \mathbf{C}_\xi \cdot A^T \end{aligned} \quad (3.88)$$

Ha az A mátrix, egy olyan mátrix, amelyik a (2.5) képletben szerepelő U mátrixnak felel meg, akkor megkapjuk az ξ vektorváltozó kovarianciamátrixának spektrálfelbontását. Ezt az eredményünket a geofizikai inverz feladat megoldása után, a paraméterek becsült hibáinak (szórásának, kovarianciáinak és korrelációjának) a meghatározására fogjuk használni.

4. fejezet - A geofizikai inverzió általános megfogalmazása

A hétköznapi életben megszoktuk, hogy egy méréssel általában meg tudunk határozni egy minket érdeklő mennyiséget. Például a mérlegre állva megmérjük a testtömegünket, vagy egy mérőszalag segítségével az asztal szélességét.

A fizikában, a csillagászatban, a geofizikában és a távérzékelésben meg kell barátkoznunk a gondolattal, hogy a minket érdeklő mennyiségeket nem tudjuk közvetlenül megmérni!

Nézzünk néhány, a fenti állítást alátámasztó geofizikai példát! A minket érdeklő földtani környezet megismerése céljából végzett mérések lehetnek például:

- A nehézségi gyorsulás mérése a földfelszín egy területén egy szelvény mentén
- Mágneses indukció (télerősség) mérése a földfelszínen egy egyenközű rácson
- Egyenáramú multielektrodás mérés egy szelvény mentén Wenner-Schlumberger elrendezésben.
- Reflexiós Szeizmikus mérés egy szelvény mentén.

Ezekben a mérésekben a közös, hogy a mérések eredményeként **adatok** nyerünk. Az adataink különféle: nehézségi gyorsulás értékek a szelvény menti hosszúság függvényében, mágneses indukció értékek egy rácson pontjain, látszólagos fajlagos ellenállás értékek a szelvény menti hosszúság és az elektródátávolság függvényében, vagy a robbantás után egy időintervallumban a geofonokon a mért kitérések hullámképe.

Általában valamilyen módon vizualizálhatjuk ezeket az adatokat, és ez valamilyen előzetes képet ad a vizsgált terület geológiai- geofizikai viszonyairól. Egy eddig még nem kutatott terület első vizsgálata során a méréseinktől azt várjuk, hogy „leljen meg” minket, a mérési adatokban valami „szokatlan”, nem várt jelentkezzen. Ezt úgy foglalthatjuk össze, hogy a méréseinkben a háttér értékeitől (valamilyen rendezett formában) eltérő mennyiségeket mérünk. Ilyen jelenség lehet, hogy

- a mért nehézségi gyorsulás térjen el a háttérértékektől a szelvény pontjaiban, de lehetőleg úgy, hogy egy szakaszon nagyobb, egy szakaszon kisebb legyen
- a mágneses indukció térjen el úgy a háttérértékektől a rácson, hogy valahol egy területen a háttérnél nagyobb, annak közelében, attól (közel) északra egy kisebb, a háttérértékeknél kisebb indukcióértékekkel jellemezhető terület legyen.
- az elektromos szelvény mentén a szelvény egy szakaszán nagy ellenállások legyenek, úgy, hogy a szelvény többi részén
- szeizmikus hullámképen jól követhető, változó „mélységű” reflexió jelenjen meg.

Általában nem elégszünk meg a mért értékekből előállított vizuális termékek szemlélésével, ugyanis minket számos alkalommal nem közvetlenül a mért értékek érdekelnek, hanem azok a **hatók**, amelyek valamilyen fizikai jelenség segítségével létrehozták a mérési adatokat. A hatók valamilyen geológia, földtani (esetenként ember alkotta, pl. régészeti) objektumok. A hatókat, mivel valamilyen fizikai teret befolyásolnak **forrásoknak** is nevezzük. Amennyiben nem teljesen ismeretlen területen mérünk, a kutatási területről meglevő **előzetes (a priori) ismereteink** lehetnek a hatókról. A fenti példánál maradva:

- Egy eltemetett vető, amelynek a két oldalán azonos mélységben levő különböző sűrűségű kőzetek különböző gravitációs teret hoznak létre
- Az egykori kráter kürtőjét kitöltő kihűlt lávatest, amelyik saját mágneses térrel rendelkezik

- Az eltemetett egykori kavicsos-homokos folyómeder, aminek a fajlagos ellenállása eltér az egykori hullámtér agyagjának fajlagos elektromos ellenállásától
- A váltakozó homokos-agyagos üledékes rétegsor felboltozódása, ami a visszavert hullámképben jelenik meg.

Ezekről a minket érdeklő hatókról van valamilyen előzetes szemléletes képünk. Láttunk vár vetőt egy hegylábi területen, kipreparálódott vulkáni tömzsöt, recens folyómedret és erőziósan feltárt gyűrt rétegsort. Közelről szemlélve látjuk ezeknek a szerkezeteknek a bonyolultságát. Távolabbról szemlélve – a részletek elhagyásával – **egyszerűsíthetjük** ezekről a hatókról alkotott képünket, ezt az egyszerűsített képet nevezzük **modellnek**:

- A vetőt egy térképre berajzolható csapásvonalú lépcsőfüggvénnyel helyettesítjük, amelynek magassága a vető elvetési magassága, és amely fölötti és alatti térrészekre különböző (de térrészenként állandó) sűrűségű kőzeteket gondolunk.
- A kihűlt lávatestet egy függőleges, téglalap alapú hasábbal helyettesítjük, amelynek állandó mágnesezettsége van
- A folyómedret az elektromos szelvény vonalára merőleges irányba végtelen, a szelvény hossza és mélysége irányában rögzített méretű derékszögű hasábnak gondoljuk, amelynek fajlagos ellenállása eltér annak a feltérnek az ellenállásától, amely magába foglalja
- A szeizmikus szelvény nyomvonalára merőleges csapású, közel haranggörbe alakú felboltozódás, antiklinális forma

A modell, tehát valamilyen formában egyszerűsített képe a valóságnak. Az egyszerűsítés lehet **geometriai** jellegű, vagyis olyan geometriai forma, ami matematikailag jól kezelhető:

- Az eltemetett vető vonalát egyenesnek gondoljuk, az eltemetett kőzettestek felszínét vízszintesnek és a vetősíkot függőlegesnek
- A vulkáni tömzs egy függőlegesen álló négyzetes hasáb
- A folyómeder a szelvényre merőleges irányban végtelen, a szelvény irányában és függőlegesen lefele véges kiterjedésű hasáb.
- A réteg teteje egy másodfokú függvénnyel leírható felület.

Alkalmazhatunk anyagi jellegű egyszerűsítéseket is: eltekintünk az **inhomogenitásoktól**, a sűrűség, a mágnesezettség, a fajlagos ellenállás és a szeizmikus sebességek helyfüggésétől, helyettük térrészenként **homogén** anyagi jellemzőket tételezünk fel. Nem vesszük figyelembe az elektromos ellenállás és a szeizmikus sebességek irányfüggését, így a közegmodellünk **izotróp**. (Bizonyos geofizikai feladatokban az anyagi jellemzők irányfüggőek, pl. vékony homok és agyagrétegek váltakozás esetén az egész rétegösszletre számolt fajlagos elektromos ellenállás függőleges irányban más, mint vízszintes irányban. Ilyen rétegösszletekre a rugalmas paraméterek és így a szeizmikus sebességek is irányfüggőek lehetnek. Az ilyen közettulajdonságot **anizotrópiának** nevezzük.)

A modellünkkel szemben elvárás, hogy segítségével meg tudjuk jósolni, a méréseink eredményeit. Például:

- Ki tudjuk számolni egy adott csapásvonalon érintkező, különböző sűrűségű lemezek gravitációs hatását a felszín bármelyik pontjában (Azokban a pontokban is, ahol a gravitációs méréseink történtek.)
- Ki tudjuk számolni a földi mágneses tér és egy mágnesezett hasáb együttes terét (A tér bármely pontjában, akár ott is ahol a mágneses méréseink történtek)
- Ki tudjuk számolni egy a szelvényre merőleges irányban végtelen, a szelvény mentén és függőlegesen véges kiterjedésű, érintkező hasábokból álló, hasábokként adott fajlagos ellenállású feltér felszínén, a szelvény különböző pontjain elhelyezett áram és mérőelektrodák esetén a látszólagos fajlagos ellenállásokat.
- A sugárkövetés módszerével meg tudjuk adni, hogy egy ismert geometriájú reflektáló felületről visszaverődve, adott szeizmikus sebességtér esetén mikor érkezik be egy adott geofonhoz a szeizmikus forrásból származó jel.

A fenti példákban, a mérési eredmények „jólását” lehetővé tevő számolást **direkt feladat**nak nevezzük. (Megkülönböztetendő a később tárgyalt inverz feladattól.) A direkt feladat további nevei: elméleti (theoretikus) mennyiségek, szintetikus adatok (mivel mi számoltuk ki), modell tér (mivel a modell hozza létre, vagy befolyásolja a fizikai teret) illetve elméleti válaszfüggvény.

Ahhoz, hogy a direkt feladatot ki tudjuk számolni (vagyis olyan mennyiségeket kapjunk, amiket össze tudunk hasonlítani a méréseink eredményeivel) ismernünk kell a direkt feladat **paramétereit**. Paramétereknek nevezzük a modell jellemzőit, amik szerepelnek a direkt feladat egyenleteiben. (A paraméterek további nevei: ismeretlenek, együttthatók, a ható jellemzői.) A példánknál maradva:

- A vető két oldalán levő közzettestek sűrűsége, és a két közzettest felszín alatti mélysége.
- Az álló négyzetes hasáb magassága, alaplapjának oldalhossza, a fedlapjának a felszín alatti mélysége és a mágnesezettsége.
- A beágyazó közet fajlagos ellenállása, a téglalap alakú szelvényre merőlegesen végtelen hasáb szelvény menti szélessége és magassága, valamint a hasáb középvonalának szelvény menti koordinátája és mélysége.
- A reflektáló felületet másodfokú függvénnyel közelítő felület egyenletében szereplő együttthatók és a felülettől a felszínig tartó térrész szeizmikus sebessége.

A direkt feladatnak tehát a paraméterek (és a mérési pontok koordinátái) a változói, és visszaadott értékei azok a fizikai mennyiségek, amelyek a mérési eredményekkel közvetlenül összehasonlíthatók.

Általában a méréseink száma nagyobb, mint a direkt feladatban szereplő paraméterek száma. Tételezzük fel, hogy valamilyen forrásból tudjuk a paramétereink „jó” értékeit, és ezeket az értékeket behelyettesítve a direkt feladatba, a mérések helyén megkapjuk a direkt feladat segítségével az elméleti tér értékeinket. Ezek az értékek általában eltérnek a mért értékektől. Ezeket az eltéréseket összefoglalóan **hibáknak** nevezzük.

A hibák természetével kapcsolatban többféle feltételezéssel élünk, ezeket a „A valószínűségszámítás alapjai” c. fejezetben részletesen tárgyaljuk.

Vizsgáljuk meg, hogy hogyan tudnánk a méréseinkből a direkt feladatban szereplő paramétereket meghatározni! Nagyon elegáns lenne, ha az egyes direkt feladatokban szereplő matematikai egyenleteket meg tudnánk fordítani (invertálni), majd ezeket az egyenleteket a mérési mennyiségekre alkalmaznánk és megkapnánk a paramétereket.

Ezt bizonyos esetekben meg is tudjuk tenni. Erre épül a **lineáris inverzió** elmélete, amit a „Lineáris inverzió” c. fejezetben tárgyalunk.

Sajnos a megoldás – a valódi geofizikai problémák esetén – nem ilyen egyszerű, ezért a matematikai statisztika eszközkészletéhez kell visszanyúlnunk („A valószínűségszámítás alapjai” c. fejezet).

A matematikai statisztika problémafelvetése szerint valamilyen eloszlású statisztikai sokaságból származó valószínűségi változóból mintát veszünk, és ennek a mintának a segítségével becsüljük meg a statisztikai sokaság valamely jellemző paraméterét.

Ennek során a direkt feladatban szereplő modellparamétereket valószínűségi változóknak tekintjük. A direkt feladat megoldást a valószínűségi változók egy függvényének, a mérési eredményeket, pedig ennek a függő valószínűségi változó realizációjának, vagyis ebből a statisztikai sokaságból vett mintának tekintjük.

Eszerint a geofizikai inverzió, vagyis a modell paraméterek meghatározása a mérési adatokból, statisztikai értelemben véve **becslésnek** tekinthető. Emiatt az inverzióval szemben ugyanazokat a feltételeket támaszthatjuk, mint a statisztikai becslésekkel kapcsolatban: a torzítatlanságot és a hatásosságot.

A fenti modellekkel kapcsolatban összefoglalhatjuk, hogy modellnek nevezzük azt az elvi vagy anyagi objektumot, amelyik visszatükrözve a geofizikai kutatás tárgyát és módszereit, képes azt úgy helyettesíteni, hogy vizsgálata új információt nyújtson az objektumról. A modellek megalkotása és használata során az alábbi típusú egyszerűsítéseket alkalmazzuk:

1. Reális források ideális forrásokkal történő felcserélése

2. Számításba nem vett ismeretlen források hatása
3. Természeti törvények nem teljes és pontos ismerete
4. A számítások során (tudatosan) megengedett közelítések

5. fejezet - A geofizikai inverzió statisztikai megközelítése

Az előző fejezetben bevezettük a modell fogalmát. Megállapítottuk, hogy egy modellhez mindig kapcsolni tudunk egy direkt feladatot. A direkt feladat alkalmas arra, hogy segítségével a „megjósoljuk” a mérési eredményeinket. Ebben a fejezetben megvizsgáljuk azt, hogy hogyan kell felépíteni a matematikai statisztikai alapelveinek a figyelembevételével az inverz feladatot megoldó algoritmusainkat.

A fejezet első részében bemutatjuk, hogy a valószínűségelmélet segítségével hogyan fogalmazhatjuk meg a geofizikai inverz feladatot.

A fejezet második részében bemutatjuk az inverz feladat megoldási menetét.

A harmadik részben a „kritériumfüggvény” lehetséges alakjait vizsgáljuk.

A negyedik részben a paraméterek becsléséhez használt minimumkereső eljárásokat mutatjuk be.

A fejezet utolsó részében mutatjuk be becsült paraméterek statisztikai elemzését szolgáló algoritmusokat, ami az eredmények minőségellenőrzését szolgálja.

5.1. A geofizikai inverz feladat megfogalmazása

A valószínűségelmélet diszkrét és folytonos változókra felírt összefüggéseiből kiindulva a három modell típust különböztetünk meg. Ezeket, mivel a mérési anyagból kiindulva von le következtetést a geológiai-geofizikai valóságra, a mérési anyag **interpretációjának** nevezzük.

Az első típus a tiszta **minőségi interpretáció** (diszkrét eset). Ekkor a mérések segítségével valamilyen diszkrét értéket akarunk az adott mérési adatrendszerhez hozzárendelni. Erre példa az osztályozás: Egy műholdkép egyetlen képpontjáról el akarjuk dönteni, hogy az növényzetet, csupasz talajt vagy vizet ábrázol.

A második típus a tiszta **mennyiségi interpretáció** (folytonos eset): a geológia-geofizikai modellt definiáljuk, ebből következik a direkt feladat, és a méréseink segítségével csak a modellben szereplő paraméterek becsült értékeire vagyunk kíváncsiak. Példa erre a gravitációs kutatási példa, amelyben csak a vető elvetésének nagyságát és térképi nyomvonalát akarjuk meghatározni.

A harmadik típusba az **összetett (mennyiségi-minőségi) interpretáció** tartozik. A méréseinkre többféle modell is illeszkedik, többféle modell alapján is el tudnánk végezni a mennyiségi interpretációt. A kitűzött cél, hogy miután elvégeztük a szóba jöhető modellekre a mennyiségi interpretációt (meghatároztuk a paramétereket) megpróbáljuk eldönteni, hogy melyik modell a legvalószínűbb. Erre példa lehet a vulkáni kürtő kitöltését vizsgáló mágneses kutatás: a kürtőt kitöltő vulkáni test geometriai modellje lehet az általunk eddig használt négyzetes hasáb, de lehet álló henger vagy csonka-kúp. A különböző modellekhez így más paraméterek tartoznak, és modellenként más és más lehet a paraméterek száma is. Összefoglalva tehát ez esetben az interpretáció azt jelenti, hogy a méréseket feldolgozom a különböző modellekkel, és a feldolgozási eredmények alapján döntöm el, hogy melyik modellt fogadom el a valóság megfelelő reprezentációjának.

A geofizikai kutatás elengedhetetlen része a vizsgált objektumról előzetesen – *a priori* – rendelkezésre álló valamennyi ismeret összegyűjtése. Ez nem csak a vizsgált területre, a modell-objektumra, a környezetében levő hatókra, hanem az alkalmazott módszerre a hatók és a mérési eredmények közötti kapcsolatra is vonatkozik. Ezeket az *a priori* információkat a legteljesebb mértékben figyelembe kell venni a sikeres kiértékelés érdekében. Ezeknek az ismereteknek az elhagyása esetén – szélsőséges esetben – nem tudunk a méréseinkből semmilyen következtetést levonni. Az előzetes feltételezések tehát nemcsak térben, de gondolati szinten is lehatárolják a vizsgálatainkat. Ez utóbbit a **kísérleti anyag matematikai modelljének** tekintjük.

A kísérleti anyag additív modellje

Vizsgáljuk a Földfelszín egy darabját, – olyat, amelyet geológiai (pl. fúrásokkal) elfogadhatóan ismerünk – valamilyen felszíni geofizikai mérési módszerrel. Megalkotva a terület geológia-geofizikai modelljét, jóslásokat tehetünk a mérési értékekre a direkt feladat segítségével. A mért értékeink – szinte biztosan – **eltérnek** majd ezektől az elméletileg meghatározott értékektől. Ezt az eltérést **hibának** tekintjük. Az eltérés oka lehet – ahogy az előző fejezet végén láttuk – a nem modellezett hatók hatása, a hatók egyszerűsítése, a pontatlanul ismert vagy megadott direkt feladat. Ezeket a hibaforrásokat együttesen **modellhibának** nevezzük. A mérőműszer pontatlansága, vagy egyéb a mérés során fellépő zavaró tényezőket összefoglalva **mérési hibának** nevezzük. Előfordulhatnak olyan hibák is, amelyek (általában) ritkán fordulnak elő, és hatásuk nagyon megváltoztatja a mérési eredményt. Ezeket **durva hibáknak** nevezzük. (Nevezhetjük baklövésnek, az angol nyelvű szakirodalomban *blunder* vagy *outlier*.) Ilyen hibák lehetnek például a műszer rossz leolvasásából, a jegyzőkönyvben hibásan felírt adatból, vagy a mérési tartomány túllépéséből eredő hibák. Ezek a hibák, mivel ritkán fordulnak elő, általában nem kezelhetők a statisztika segítségével. A durva hibákat általában az inverzió első lépésében, sokszor manuálisan vagy félig automatizált módon, de el kell távolítani az adatrendszerből, ezeket a méréseket ki kell hagyni a további feldolgozásból.

A mérési mennyiségeket eszerint – meghagyva az általános kezelés lehetőségét – az alábbi formában írhatjuk:

$$\mathbf{u} = \mathbf{f} + \mathbf{n} \quad (5.1)$$

(a vastagon szedett betűk itt is a vektorokat jelölik). Az $\mathbf{u}$ vektor jelöli a kísérleti anyagot, vagyis a méréseket, ennek elemei az egyes mérések.

Az $\mathbf{f}$ vektorba vannak összefoglalva a *direkt feladat* megoldásai, vagyis minden mérési adathoz van egy elméleti megoldásunk. Ahogy korábban említettük, ez mennyiségi interpretációnál a paraméterek függvénye. Ennek alakját úgy választják meg, hogy ne legyen túlságosan bonyolult (lehetőleg kevés paramétertől függjön) de a reális objektumnak lehetőleg minden – a mérési eredményeket befolyásoló – tulajdonságát vegye figyelembe.

Az $\mathbf{n}$ vektor a **hibavektor**, ami a mért és az ideális tér eltérését jellemzi a megfigyelési pontokban. Ezt a továbbiakban **véletlen komponensnek** (összetevőnek) nevezzük.

Az $\mathbf{u}$ vektor, vagyis a méréseink véletlen jellegét az adja, hogy elkerülhetetlenül léteznek mérési hibák. A felírt egyenlőségben vizsgáljuk meg az ismert és ismeretlen mennyiségek számát! Mivel a mérést elvégeztük, az $\mathbf{u}$ vektor elemei ismertek, ezek együttesen N darab mérés. Az $\mathbf{f}$ vektort, vagyis a direkt feladat (N darab, $\mathbf{u}$ -val megegyező darabszámú) elemét általában néhány (S darab) **paraméter** segítségével ki tudjuk számolni. A paraméterek S száma általában kisebb N -nél. Ha az $\mathbf{n}$ elemekre semmilyen előzetes ismeretünk nincs, akkor a feladat alulhatározott, mert összességében több ismeretlenünk van ($N+S$), mint mérési eredményünk (N).

Mivel a modellünk a valóság egyszerűsítéséből adódott, ezért a modell alkalmazásával tudatosan kizártunk bizonyos **nem véletlen jellegű hatásokat** is. Ezeknek a nem véletlen jellegű hatások – terek – figyelembe vételével a mérések modellje

$$\mathbf{u} = \mathbf{f} + \mathbf{a} + \mathbf{n} \quad (5.2)$$

Alakú lesz, ahol $\mathbf{a}$ vektor jelöli a terek nem véletlen jellegű eltérését. Egy ilyen modell esetén a méréseinkkel azonos számú ismeretlenünk lesz az $\mathbf{a}$ vektorban és – hasonlóan az előző esethez – a feladat alulhatározott lesz. Emiatt az egyik lehetőség, hogy az $\mathbf{a}$ nem véletlen komponens elemeit valamilyen egyszerű függvénnyel közelítjük – ekkor tulajdonképpen a direkt feladatot bonyolítjuk el. A másik lehetőség, hogy az $\mathbf{n}$ véletlen komponens bővítjük ki, úgy, hogy tartalmazza az nem véletlen komponens hatását is, például a modellhibákat. Ez, ahogy később látni fogjuk, torzíthatja a meghatározott paramétereinket.

Ez utóbbi esetben $\mathbf{f}$ vektort **hasznos jelnek**, az $\mathbf{a} + \mathbf{n}$ elemeit összevonva tartalmazó új $\mathbf{n}$ vektort **zajnak** nevezzük.

Írjuk fel a méréseink „hasznos jel + zaj” szerinti felbontásával adódó modelljét!

$$\mathbf{u} = \mathbf{f} + \mathbf{n} \quad (5.3)$$

A mérési adatok ilyen modelljét **additív** modellnek nevezzük.

Általában a mérési anyagban a méréseinket csoportosítani tudjuk. Az adataink átláthatóbb kezelése érdekében képezzünk összesen $2K+1$ csoportot, és a csoportjainkat számozzuk meg: a k index $-K$ -tól $+K$ -ig fusson. A csoportosítást úgy végezzük el, hogy két mérési adat, amit különböző k indexű csoportból vettünk statisztikailag független legyen. Ugyanakkor az ugyanazon k indexű csoportból származó mérések korrelációs kapcsolatban legyenek egymással. A k -adik csoportba tartozó méréseket számozzuk be az i indexszel, amelynek értéke a k -tól függő $-I_k$ -tól $+I_k$ -ig fusson, így i összesen $2I_k+1$ különböző értéket vehet fel. Ez a felírást Wiener (1949) alkalmazta, aki az információ átvitel kapcsán, időben változó, egymástól független drótvégeken mérhető, egyenletesen mintavételezett elektromos jeleket vizsgált. Itt a nagyszámú drótvégek közül kiválasztott egyetlen drótvég környezetében levő drótvégeket jelöljük a k indexszel (így a $k = 0$ a kiválasztott drótvég), az i index pedig egy kiválasztott pillanat ($i = 0$) előtti és utáni időpillanatokban mérhető feszültségek indexe. A felírást megtartását az indokolja, hogy a mérések ilyen csoportosításával alkalmazhatjuk a matematikai modellt több mérési módszerrel végzett mérés csoport együttes, vagy különböző körülmények között (eltérő időpontokban, más pontosságú műszerrel) végzett mérések kiértékelésére. A következőkben bemutatott esetekben általában megtehetjük, hogy egyetlen mérés csoportot alakítunk ki a méréseinkből, ekkor $k = 0$. Az ettől eltérő eseteket külön jelöljük.

Vizsgáljuk meg az egyes modell típusainkra, hogy az additív modell milyen konkrét alakoknak felel meg!

Minőségi interpretáció esetén azt keressük, hogy a mérési anyag az N darab lehetséges modell közül melyiket valószínűsíti leginkább. Ekkor:

$$\mathbf{u}_k = \begin{cases} \text{vagy } \mathbf{f}_{1k} + \mathbf{n}_{1k} \\ \text{vagy } \mathbf{f}_{2k} + \mathbf{n}_{2k} \\ \vdots \\ \text{vagy } \mathbf{f}_{Nk} + \mathbf{n}_{Nk} \end{cases} \quad (5.4)$$

A $v=1,2,\dots,N$ index az $\mathbf{f}_{vk}$ vektor jelölésében a modell-objektum lehetséges állapotától függő funkcionális összefüggést jelöli. Az inverz feladat megfogalmazása ezen állapotok közötti választás optimális eljárásának a meghatározása. (A minőségi interpretációra példaként a műholdkép osztályozását hoztuk fel. Ebben az esetben egy $\mathbf{f}_{vk}$ vektor a v -edik osztályhoz tartozó, osztályközpontokból képzett vektor. A véletlen eltérés komponens pedig megadja, a vizsgált képpont ($\mathbf{u}$) és az adott osztályközpont ($\mathbf{f}$) különbségvektorát.)

A **menyiségi interpretáció** esetén a modell alakja:

$$\mathbf{u}_k = \mathbf{f}_k(\mathbf{p}) + \mathbf{n}_k \quad (5.5)$$

Itt $\mathbf{f}_k(\mathbf{p})$ vektor a direkt feladat megoldás, adott alakú az alkalmazott geológiai geofizikai modell által meghatározott alakú függvény, a $\mathbf{p}$ vektor pedig a p_s elemekből álló paramétervektor, amelynek $s=1,2,\dots,S$ darab eleme van.

Az összetett (menyiségi-minőségi) interpretáció esetén a modell alakja:

$$\mathbf{u}_k = \begin{cases} \text{vagy } \mathbf{f}_{1k}(\mathbf{p}_1) + \mathbf{n}_{1k} \\ \text{vagy } \mathbf{f}_{2k}(\mathbf{p}_2) + \mathbf{n}_{2k} \\ \vdots \\ \text{vagy } \mathbf{f}_{Nk}(\mathbf{p}_N) + \mathbf{n}_{Nk} \end{cases} \quad (5.6)$$

Ahol $\mathbf{f}_{vk}(\mathbf{p}_v)$ a $v = 1,2,\dots,N$ lehetséges modellekhez tartozó direkt feladat megoldások, amelyek mindegyikéhez tartozik egy $\mathbf{p}_v$ paramétervektor, amelynek elemei p_{vs} elemek ($s = 1,2,\dots,S_v$). Az S_v elemek száma függ az adott modelltől: a különböző modellekhez különböző számú paraméter tartozhat.

Itt $\mathbf{f}_{vk}$ a v -edik osztály osztályközéppontjának csatornánkénti értékeiből képzett vektor, $\mathbf{R}_{vk}$ pedig a v -edik modellállapothoz tartozó k -adik méréscsoport kovariancia mátrixa.

A gondolatmenetet a mennyiségi interpretációra alkalmazva, azt kapjuk, hogy a bal oldalon az a feltételes valószínűség áll, hogy mennyi a valószínűsége annak, hogy pont az $\mathbf{u}$ vektor elemeivel megegyező mennyiségeket mérünk egy konkrét rögzített $\mathbf{p}$ modellparaméter vektor esetén. Ez az $f(\mathbf{u}|\mathbf{p})$ feltételes valószínűsége!

$$f(\mathbf{u}|\mathbf{p}) = \prod_{k=1}^K \left(\frac{1}{(2\pi)^{\frac{1}{2}(2I_k+1)} \cdot \sqrt{\det(\mathbf{R}_k)}} \right) \cdot e^{-\frac{1}{2} \sum_{k=1}^K ((\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^T \mathbf{R}_k^{-1} (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})))} \quad (5.9)$$

Az összetett interpretációnál egy valószínűség sűrűség függvényekből álló sorozatot kapunk, amelynek elemei egy adott v -edik modell és a hozzá tartozó rögzített paramétervektor esetén azt adják meg, hogy mi annak a valószínűsége, hogy pont az $\mathbf{u}$ mérési eredmény valósul meg.

$$f(\mathbf{u}|\mathbf{p}_v) = \prod_{k=1}^K \left(\frac{1}{(2\pi)^{\frac{1}{2}(2I_k+1)} \cdot \sqrt{\det(\mathbf{R}_k)}} \right) \cdot e^{-\frac{1}{2} \sum_{k=1}^K ((\mathbf{u}_k - \mathbf{f}_{vk}(\mathbf{p}_v))^T \mathbf{R}_k^{-1} (\mathbf{u}_k - \mathbf{f}_{vk}(\mathbf{p}_v)))} \quad (5.10)$$

A fenti esetekben teljesen általános függvényt tételztünk fel ($\mathbf{f}_v$, és $\mathbf{f}_{vk}$).

Lineáris elméleti terű additív modell

A mennyiségi interpretáció egy – a gyakorlatban gyakran előforduló – speciális esetét vizsgáljuk meg, ahol a direkt feladat ($\mathbf{f}_k$) a paraméterek lineáris függvénye. Ezeket **lineáris elméleti terű additív modellek**nek nevezzük. Ebben az esetben a mérési eredményekre vonatkozó általános formula (egyetlen mérési csoportot vizsgálva, és elhagyva a k indexet):

$$\mathbf{u} = \mathbf{f}(\mathbf{p}) + \mathbf{n} \quad (5.11)$$

ahol

$$\mathbf{f}(\mathbf{p}) = \mathbf{A} \cdot \mathbf{p} \quad (5.12)$$

Az $\mathbf{A}$ mátrix neve: strukturális vagy konstruktív mátrix. Ha N darab mérésünk van, akkor a mérések – és így a direkt feladat megoldás – vektora egy N elemű oszlopvektor, amelynek elemei u_i -k ($i=1,2,\dots,N$). Ha a paramétereink száma M , akkor a $\mathbf{p}$ vektor egy M elemű oszlopvektor, amelynek elemei p_j -k ($j=1,2,\dots,M$). Ekkor az $\mathbf{A}$ mátrix egy N sorból, és M oszlopból álló mátrix, amelynek elemei A_{ij} -k.

A modellre vonatkozó feltevésünk, hogy az N elemű hibavektor elemei azonos eloszlásúak és nulla várható értékűek:

$$M(n_i) = 0 \quad (5.13)$$

Ekkor a méréseink várható értéke:

$$M(\mathbf{u}) = \mathbf{f}(\mathbf{p}) = \mathbf{A} \cdot \mathbf{p} \quad (5.14)$$

5.2. Statisztikai becslési eljárások a geofizikai inverz feladat megoldására

Ebben a fejezetben bemutatjuk, hogy a geofizikai inverz feladatot hogyan tudjuk megfogalmazni a matematikai statisztikában nagy sikerrel alkalmazott alapelvek segítségével.

A maximum likelihood elv

Az előző fejezetben láttuk, hogy a meghatározandó paraméterek bonyolult függvénykapcsolatban állnak a méréseinkkel, amik valószínűségi változóknak tekinthetők. Minthogy a direkt feladatban szereplő paramétereinket a méréseinkből akarjuk meghatározni, ezért a paramétereink is valószínűségi változóknak tekinthetők.

A matematikai statisztikában egy minta alapján a statisztikai sokaság eloszlásfüggvényében szereplő paramétereket a maximum likelihood elv segítségével becsüljük. Alkalmazzuk ezt az elvet a geofizikai inverz feladat megoldására!

A **maximum likelihood elv** azt mondja ki, hogy a paraméterek azon értékeit fogadjuk el a paraméterek becslésének, amelyek a megvalósuló mérések (adott minta) esetén a legvalószínűbbek. Ez a maximumfeltétel a minőségi, mennyiségi és az összetett interpretáció esetén az alábbi **együttes valószínűségekkel** fejezhető ki:

$$\begin{aligned} P(\nu, \mathbf{u}) &:= \max \\ f(\mathbf{p}, \mathbf{u}) &:= \max \\ f(\mathbf{p}_\nu, \mathbf{u}) &:= \max \end{aligned} \quad (5.15, 5.16, 5.17)$$

(Mivel a minőségi interpretáció során diszkrét értékekhez rendelünk valószínűségeket, ezért itt a valószínűség szerepel, míg a mennyiségi és összetett interpretációnál a paraméterek folytonos függvénye, a valószínűség sűrűsége szerepel.) A likelihood függvény (3.65) ezeknek a valószínűség- és sűrűségfüggvények logaritmusából áll:

$$\begin{aligned} L(\nu) &= \ln(P(\nu, \mathbf{u})) \\ L(\mathbf{p}) &= \ln(f(\mathbf{p}, \mathbf{u})) \\ L(\mathbf{p}_\nu) &= \ln(f(\mathbf{p}_\nu, \mathbf{u})) \end{aligned} \quad (5.18, 5.19, 5.20)$$

Ezeket a valószínűség- és sűrűségfüggvényeket Bayes (1701-1761) tétele segítségével fejezhetjük ki:

$$\begin{aligned} P(\nu, \mathbf{u}) &= P(\nu | \mathbf{u}) \cdot f(\mathbf{u}) = P(\nu) \cdot f(\mathbf{u} | \nu) \\ f(\mathbf{p}, \mathbf{u}) &= f(\mathbf{p} | \mathbf{u}) \cdot f(\mathbf{u}) = f(\mathbf{p}) \cdot f(\mathbf{u} | \mathbf{p}) \\ f(\mathbf{p}_\nu, \mathbf{u}) &= f(\mathbf{p}_\nu | \mathbf{u}) \cdot f(\mathbf{u}) = f(\mathbf{p}_\nu) \cdot f(\mathbf{u} | \mathbf{p}_\nu) \end{aligned} \quad (5.21, 5.22, 5.23)$$

Ezek a fenti valószínűség- és sűrűségfüggvények tartalmazzák a modellre vonatkozó, a méréssel megszerzett információt is, így ezeket **a posteriori** valószínűség- és sűrűségfüggvényeknek nevezhetjük. Az egyenletek bal szélén álló mennyiségeket a jobb szélén felírt képlet segítségével tudjuk kiszámolni. Elemezzük ezeket a mennyiségeket!

Az jobboldali egyenletek jobb oldalán a $f(\mathbf{u} | \nu)$, $f(\mathbf{u} | \mathbf{p})$ és $f(\mathbf{u} | \mathbf{p}_\nu)$ feltételes valószínűsége-sűrűségek állnak, amelyeknek az alakját az előző fejezetben (5.8, 5.9, 5.10) már meghatároztuk!

A jobboldali egyenletek első tagja rendre a $P(\nu)$ valószínűség- és $f(\mathbf{p})$, $f(\mathbf{p}_\nu)$ sűrűségfüggvények. Ezek rendre a lehetséges ν állapotok, valamint a $\mathbf{p}$ és $\mathbf{p}_\nu$ paraméterértékek előzetes – **a priori** – valószínűségei illetve valószínűség sűrűségei. Ezek a mennyiségek nem függenek a mérésektől. (Ezért nevezzük előzetes mennyiségeknek.) Ezek a mennyiségek tartalmazhatják az objektum állapotára vonatkozó – már a mérés elvégzése előtt meglevő – ismereteket.

Ilyen előzetes információ minőségi interpretáció esetén $P(\nu)$ a ν -edik modellállapot előzetes valószínűsége. (A műholdkép osztályozási példa esetén, $P(\nu)$ a ν -edik osztály valószínűsége a képen, például ha a kép felét erdő borítja, akkor az erdő osztály *a priori* valószínűsége 0,5.)

Mennyiségi interpretáció esetén az $f(\mathbf{p})$ *a priori* valószínűségek az egyes paraméterekre vonatkozó előzetes ismereteket foglalja magába. Ilyen feltétel lehet, hogy egy p_j paraméter csak egy tartományban vehet fel értékeket. (Példa lehet erre, a mélyfúrási adatok inverziója során a kőzet porozitása: csak 0 és 30% közötti porozitás-értékeket fogadunk el reálisnak. Ekkor a porozitást leíró paraméter valószínűség-sűrűségfüggvénye egy 0 és 30% közötti egyenletes eloszlás sűrűségfüggvényével adható meg.) Később látni fogjuk, hogy az *a priori* valószínűségek segítségével a paraméterek közötti függvénykapcsolatokat is leírhatjuk. (Amennyiben ezek determinisztikus függvénykapcsolatot jelentenek, akkor az ilyen függvénykapcsolatot **kényszerfeltételeknek** nevezzük.)

Az összetett interpretáció esetén az $f(\mathbf{p}_\nu)$ valószínűségi sűrűség a ν -edik modellhez tartozó paraméterek sűrűségfüggvénye. Ez a sűrűségfüggvény tartalmazhatja a ν -edik modell előzetes valószínűségét is.

A három interpretációs folyamat mindegyikében tehát a **döntési kritérium**, az *a posteriori* valószínűségből és sűrűségfüggvényekből képzett likelihood függvény maximalizálása. Vizsgáljuk meg a likelihood függvény alakját a három interpretáció során!

A maximum likelihood elv alkalmazása minőségi interpretációra

Minőségi interpretáció esetén a likelihood függvény minden lehetséges modell objektumhoz rendel egy számot (valószínűséget).

Ekkor az egyes modellekhez tartozó

$$\lambda_\nu = \ln(P(\nu) \cdot f(\mathbf{u}|\nu)) \quad (5.24)$$

Értékek sorozatát **kritériumfüggvénynek** nevezzük. A **becslési algoritmus** tehát abból áll, hogy **megkeressük**, melyik ν modellállapothoz tartozik a **függvény maximuma**.

Tegyük most fel, hogy az egyes modellek előzetes *a priori* valószínűsége megegyezik, ekkor a likelihood függvényben csak a (5.8) alakú $f(\mathbf{u}|\nu)$ feltételes valószínűségek szerepelnek. Ekkor a korábban bemutatott **normális eloszlású additív modell** segítségével a kritériumfüggvény

$$\lambda_\nu = -\frac{1}{2} \sum_{k=-K}^K ((\mathbf{u}_k - \mathbf{f}_{\nu k})^\top \cdot \mathbf{R}_{\nu k}^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_{\nu k})) \quad (5.25)$$

alakú lesz. (Ez a konkrét példa világít rá arra, hogy miért célszerű a likelihood függvényt, mint a valószínűségek logaritmusát definiálni. Ekkor ugyanis a normális eloszlás sűrűségfüggvényében szereplő exponenciális függvényt logaritmálva, csak a kitevőben szereplő összegzés jelenik meg a kritériumfüggvényben, ami így matematikailag jól kezelhető függvénné válik.)

A maximum likelihood elv alkalmazása mennyiségi interpretációra

Vizsgáljuk meg részletesen a mennyiségi interpretáció esetét. Írjuk fel a likelihood függvény konkrét alakját az additív, normális hibaeloszlású modell esetére azzal a feltételezéssel, hogy az *a priori* valószínűségi sűrűségek valamennyi paraméterre megegyeznek. Ekkor a (5.9) egyenlet logaritmálásával az alábbi képlet adódik:

$$\lambda(\mathbf{p}) = -\frac{1}{2} \sum_{k=1}^K \left\{ \ln(\det(\mathbf{R}_k)) + (2I_k + 1) \cdot \ln(2\pi) + \left[(\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^T \cdot \mathbf{R}_k^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})) \right] \right\} \quad (5.26)$$

Ennek a függvénynek a maximumához tartozó $\hat{\mathbf{p}}$ argumentum jelenti a keresett függvényt.

A maximum likelihood becslés azért nevezhetjük optimálisnak, mert a független kísérleti eredmények (a méréseink) számának határtalan megnövelése esetén ezek a becslések torzítatlanok (vagyis szisztematikusan hibától mentesek), effektívek (vagyis a paraméterek szórása minimális) és normális eloszlásúak.

A mennyiségi interpretáció esetében tehát a likelihood függvény maximumát akarjuk meghatározni a paraméterek függvényében. Ez egy többváltozós függvény szélsőértékének meghatározását jelenti. (A függvény annál bonyolultabb lehet, minél több elemű a paramétervektor.) Egy többváltozós függvény szélsőértéke az argumentumában szereplő paramétervektor elemei által kifeszített **paraméterter** több pontján is felveheti a maximumát. Ezt a jelenséget **ekvivalenciának** nevezzük, és azt fejezi ki, hogy a kritériumfüggvény alapján nem tudunk a több különböző **paraméterkombináció** között dönteni. A feladat teljes megoldásához meg kell találnunk valamennyi ilyen maximumhelyet. A függvény szélsőértékének megkeresését általában **iterációs algoritmusok** segítségével végezzük. Az iterációt valamilyen kezdő paraméterkombinációból indítjuk el, és az eljárás valamilyen **lokális maximum**helyre konvergál. Nagyon sok kezdő paraméterkombinációból elindítva az eljárást, csaknem biztosan meg tudunk találni valamennyi maximumhelyet.

A maximum likelihood elv alkalmazása lineáris elméleti terű modellek minőségi interpretációjára

A $\lambda(\mathbf{p})$ függvény alakja, és a maximumhely keresése jelentősen leegyszerűsödik, ha a direkt feladat megoldást, mint a paraméterek lineáris függvényét tudjuk felírni, vagyis **lineáris elméleti terű** modellünk van (lásd fentebb).

Ekkor – emlékezzünk vissza – a direkt feladat az alábbi formában írható fel:

$$\mathbf{f}(\mathbf{p}) = \mathbf{A} \cdot \mathbf{p} \quad (5.27)$$

Ennek felhasználásával a (5.9) egyenlettel definiált függvény maximalizálása, egy mérési csoport esetén az alábbi függvény minimalizálásával egyenértékű:

$$\lambda^*(\mathbf{p}) = (\mathbf{u} - \mathbf{A} \cdot \mathbf{p})^T \cdot \mathbf{R}^{-1} \cdot (\mathbf{u} - \mathbf{A} \cdot \mathbf{p}) \quad (5.28)$$

Ennek az egyenletnek a szélsőértéke – vagyis a $\mathbf{p}$ vektor $\hat{\mathbf{p}}$ becslése – analitikusan is megadható:

$$\hat{\mathbf{p}} = (\mathbf{A}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{u}) \quad (5.29)$$

Ahol $\mathbf{A}$ mátrix a struktúra mátrix, $\mathbf{R}$ mátrix a mérések véletlen hibakomponens vektorából ($\mathbf{n}$) készített kovariancia mátrix, $\mathbf{u}$ pedig a méréseinkből kialakított vektor.

Módosítsuk annyiban ezt a kifejezést, hogy a méréseinket egyforma súlyúnak tekintjük, ez annyit jelent, hogy a véletlen eltérés vektor elemekből képzett $\mathbf{R}$ kovarianciamátrix egy egységmátrix ($\mathbf{I}$) lesz.

Ekkor a (5.28) egyenletet az alábbi formában is felírhatjuk:

$$\lambda^*(\mathbf{p}) = \mathbf{n}^T \cdot \mathbf{n} = (\mathbf{u} - \mathbf{A} \cdot \mathbf{p})^T \cdot (\mathbf{u} - \mathbf{A} \cdot \mathbf{p}) \quad (5.30)$$

Az egyenletben szerepel a véletlen eltérés komponens vektor ($\mathbf{n}$) saját transzponáltjával vett skalárszorzata. Ez a vektorelemekre felírva:

$$\lambda^*(\mathbf{p}) = \sum_{i=1}^N (n_i)^2 \quad (5.31)$$

Vagyis a véletlen komponens vektorban szereplő eltérések négyzetösszege.

Ennek a függvénynek a minimalizálásából adódó feltételt a **legkisebb négyzetek elvének** nevezzük. Komoly tudományos jelentősége van ennek az eredménynek, ugyanis a fenti becslési forma már a XIX. század elején ismert volt. Mivel a módszer segítségével a becsült paraméterek egy lineáris egyenletrendszer megoldásával megkaphatók voltak, (ahogyan a 5.29 egyenletből látszik) a számítógépek elterjedése előtti időben ez volt a paraméterbecslések egyetlen gazdaságosan végrehajtható módja.

A legkisebb négyzetes becslés csak abban az esetben adja ugyanazt a becslési eredményt, mint a maximum likelihood becslés, amikor a véletlen komponens normális eloszlású, centírozott (nulla várható értékű) és korrelálatlan a mérési pontokban. (A fenti levezetésben ennek megfelelő feltételeket vezettünk be.)

A legkisebb négyzetes módszert (és annak változatait) – egyszerűségük, és könnyen programozhatóságuk miatt – részletesen ismertetjük jelen jegyzet későbbi fejezeteiben.

A maximum likelihood elv alkalmazása minőségi interpretációra a priori valószínűségek felhasználásával

Vizsgáljuk meg a maximum likelihood módszert a minőségi interpretáció esetén, ha nem tételezzük fel, hogy az egyes becsült paraméterek egyforma valószínűségűek, vagyis a becslésben fel szeretnénk használni a paraméterekre vonatkozó előzetes (*a priori*) ismereteinket! Ekkor – felhasználva az *a posteriori* valószínűség (5.22) alakját – a maximalizálandó likelihood függvény logaritmusát:

$$\lambda(\mathbf{p}) = \ln(f(\mathbf{p})) + \ln(f(\mathbf{u} | \mathbf{p})) \quad (5.32)$$

Ahogy említettük, az $f(\mathbf{p})$ a paraméterekre vonatkozó előzetes információ alapján megkonstruált valószínűsége-sűrűség-függvény.

A maximum likelihood elv alkalmazása összetett interpretáció esetén

Az összetett (minőségi-mennyiségi) interpretáció során a lehetséges modellekhez ki kell választani a $\mathbf{f}_v(\mathbf{p}_v)$ direkt feladat összefüggések típusát, és a mérési anyagból meg kell határozni minden típusra a p_{vs} ($s = 1, 2, \dots, S$) paraméterek becsült értékeit.

Ennek során minden modellre a mennyiségi interpretációnál bemutatott folyamatot alkalmazzuk a p_{vs} paraméterek becslésére. Ennek eredményeképpen minden modellre rendelkezésre állnak a paraméterek becsült értékeiből álló

$\hat{\mathbf{p}}_v$ vektorok. Ezeket az egyes modellekhez tartozó, a minőségi interpretációhoz felhasznált likelihood függvényekbe visszahelyettesítve, megkapjuk a likelihood függvények maximum értékeiből álló számsorozatot λ_v elemeit. A becslés ezek közül kiválasztani a maximális értékűt. Ezáltal az összetett interpretációt visszavezettük egymás után elvégzett minőségi és mennyiségi interpretációra.

Az additív modellek véletlen eltérés komponensének analízise

Vizsgáljuk meg a mennyiségi interpretációnál a maximum likelihood feltételből levezetett maximalizálandó függvényt:

$$\lambda(\mathbf{p}) = -\frac{1}{2} \sum_{k=-K}^K \left\{ \ln(\det(\mathbf{R}_k)) + (2I_k + 1) \cdot \ln(2\pi) + \left[(\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^\top \cdot \mathbf{R}_k^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})) \right] \right\} \quad (5.33)$$

Ennek a függvénynek keressük a maximumát a paraméterek függvényében, ám utaltunk arra, hogy $\mathbf{p}$ paraméterek megváltozása esetén változik a direkt feladat megoldás, ezzel változnak a véletlen eltérés komponens vektor elemei, ami visszahat az ezekből képzett $\mathbf{R}_k$ kovarianciamátrixokra is.

A megfigyeléseink $2K+1$ csoportra osztásánál feltételeztük, hogy az véletlen eltérés komponens vektor n_{ik} elemeire igaz, hogy normális eloszlásúak, centírozottak, k szerint függetlenek és i szerint stacionáriusak. Ekkor a k -adik mérés-csoport-hoz egyetlen σ_k^2 szórásérték tartozik, ami az k -adik csoportra állandó. Ennek segítségével az $\mathbf{R}_k$ kovariancia mátrixot felírhatjuk a szórás és a korrelációs mátrix (3.24) szorzataként: (Figyeljünk az eltérő jelölésekre!)

$$\mathbf{r}_k = \frac{\mathbf{R}_k}{\sigma_k^2} \quad (5.34)$$

Ennek felhasználásával írjuk fel újra, a mennyiségi interpretáció esetén maximalizálandó függvényt (5.22):

$$\lambda(\mathbf{p}, \sigma_k^2, \mathbf{r}_k) = -\frac{1}{2} \sum_{k=-K}^K \left\{ (2I_k + 1) \cdot \ln(\sigma_k^2) + \ln(\det(\mathbf{r}_k)) + (2I_k + 1) \cdot \ln(2\pi) + \frac{1}{\sigma_k^2} \left[(\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^\top \cdot \mathbf{r}_k^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})) \right] \right\} \quad (5.35)$$

Vizsgáljuk először a legegyszerűbb esetet, amikor a σ_k^2 szórásnégyzetek és a $\mathbf{p}$ paramétervektor elemei ismeretlenek, és az $\mathbf{r}$ korrelációs mátrix elemeit ismertnek tételezzük fel. Ekkor a fenti függvény σ_k^2 -ek szerinti maximuma adja a σ_k^2 -ek becslését. (A fenti függvényt σ_k^2 -ek szerint lederiváljuk, és ahol a derivált nulla, az a szélsőérték.) A kapott egyenlet átrendezésével kapjuk a σ_k^2 -k becslését:

$$\hat{\sigma}_k^2 = \frac{1}{2I_k + 1} \cdot \left[(\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^\top \cdot \mathbf{r}_k^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})) \right] \quad (5.36)$$

A kapott becslt szórásokat visszairva a (5.35) egyenletbe, és az egyszerűsítések elvégzése után:

$$\begin{aligned} \lambda(\mathbf{p}, \sigma_k^2) &= -\sum_{k=-K}^K (2I_k + 1) \cdot \ln(\sigma_k^2) = \\ &= -\sum_{k=-K}^K (2I_k + 1) \cdot \ln \left\{ \frac{1}{(2I_k + 1)} \left[(\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^\top \cdot \mathbf{r}_k^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})) \right] \right\} \end{aligned} \quad (5.37)$$

Ez a formula csak az ismeretlen $\mathbf{p}$ vektortól függ.

Először tehát ennek a formulának a segítségével meghatározzuk a $\hat{\mathbf{p}}$ becslést, és azt visszahelyettesítve (5.36) egyenletbe számítjuk ki a σ_k^2 -eknek a $\hat{\sigma}_k^2$ becslését.

Vizsgáljuk meg azt a speciális esetet, amikor a méréseinkből alkotott k -darab csoportban az elemek száma (I_k) mindegy csoportban azonos (I -vel egyenlő) és az n_{ik} véletlen komponens pedig azonos szórású: $\sigma_k^2 = \sigma^2$. Ekkor az egyszerűsítések után adódik:

$$\hat{\sigma}^2 = \frac{1}{(2I+1) \cdot (2K+1)} \cdot \sum_{k=-K}^K \left[(\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^\top \cdot \mathbf{r}_k^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})) \right] \quad (5.38)$$

És ezzel a becsléshez felhasznált függvény, amelynek a maximumát keressük:

$$\lambda(\mathbf{p}, \hat{\sigma}^2) = -\ln \left\{ \sum_{k=1}^K \left[(\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^T \cdot \mathbf{r}_k^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})) \right] \right\} \quad (5.39)$$

Ebből, felhasználva, hogy a logaritmusfüggvény az argumentumának monoton függvénye, kapjuk, hogy:

$$\lambda(\mathbf{p}, \hat{\sigma}^2) = - \sum_{k=1}^K \left[(\mathbf{u}_k - \mathbf{f}_k(\mathbf{p}))^T \cdot \mathbf{r}_k^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k(\mathbf{p})) \right] \quad (5.40)$$

Ez az eredmény azt fejezi ki, hogy ismeretlen szórások esetén is ugyanolyan alakú a maximalizálandó függvény, mint ismert szórások esetén. Megjegyezzük, hogy a méréseink korreláltságát jellemző korrelációs mátrixot itt is ismertnek tételeztük fel.

Ezt az eredményt alkalmazhatjuk a lineáris additív terű modelljeinkre is. Minthogy az itt levezetésre kerülő mennyiségek már átvezetnek minket a legkisebb négyzetes becslések témakörébe, vezessük be az ott alkalmazásra kerülő **súlymátrixot**:

$$\mathbf{W} = \sigma_0^2 \cdot \mathbf{R}^{-1} \quad (5.41)$$

Ahol a $\mathbf{W}$ súlymátrix a kovariancia-mátrix inverzének konstansszorosa. A σ_0^2 szorzót az **egységnyi súlyú méréshez tartozó szórásnégyzetnek** nevezzük.

Ekkor – ismét felhasználva, hogy a direkt feladat megoldás a paraméterek lineáris függvénye (5.27 egyenlet) – kapjuk, hogy :

$$\hat{\mathbf{p}} = (\mathbf{A}^T \cdot \mathbf{W} \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{W} \cdot \mathbf{u}) \quad (5.42)$$

Az egységnyi súlyú méréshez tartozó szórásnégyzet (σ_0^2) torzítatlan becslése:

$$\hat{\sigma}_0^2 = \frac{(\mathbf{u} - \mathbf{A} \cdot \hat{\mathbf{p}})^T \cdot \mathbf{W} \cdot (\mathbf{u} - \mathbf{A} \cdot \hat{\mathbf{p}})}{K - S} \quad (5.43)$$

A nevezőben a mérések számának (K) és a paraméterek számának (S) különbsége, vagyis a szabadsági fokok száma áll.

5.3. A kritériumfüggvény lehetséges alakjai

Ebben a fejezetben láttuk, hogy hogyan lehet a maximum likelihood elv segítségével egy olyan függvényt előállítani, amelynek szélsőértéke a keresett paraméterek **optimális becslését** szolgáltatja. (Optimálisnak nevezünk egy becslést, ami aszimptotikusan torzítatlan és effektív.) Láttuk, hogy normális hibaeloszlás esetén a maximum likelihood módszer a legkisebb négyzetek módszerére vezet. A legkisebb négyzetek módszerénél is egy függvény szélsőértékének – minimumhelyének – meghatározása segítségével szolgáltatja a paraméterek becsült helyét. (Ebben az esetben analitikus formában tudjuk kifejezni a paramétereket.) Általánosságban elmondhatjuk, hogy a becslési eljárás valamilyen függvény szélső értékének megtalálásából áll.

Általánosítsuk a fenti meghatározást! **Kritériumfüggvénynek** nevezzük a mérési adatok (az $\mathbf{u}$ mérések), a $\mathbf{p}$ paraméterek és a rendelkezésre álló információk olyan függvényét, amelynek a $\mathbf{p}$ paraméterek szerinti szélsőértéke a paraméterek becslését szolgáltatja. Ezek a kritériumfüggvények lehetnek **szuboptimálisak** is, ha a használatuk valamilyen szempontból célszerű.

Vizsgáljuk meg még egyszer a legkisebb négyzetek módszerben szereplő kritériumfüggvényt:

$$\hat{\lambda}^*(\mathbf{p}) = \sum_{i=1}^N (n_i)^2 \quad (5.44)$$

Ebben a függvényben az eltérésvektor ($\mathbf{n}$) elemeinek a négyzeteinek az összege szerepel. Az eltérésvektor elemei pedig a mérési pontokban a mért érték és az elméleti modell (direkt feladat) által mondott értékek különbsége.

A kritériumfüggvény azt fejezi ki, hogy a modell mennyire tér el a mért értékektől, milyen „távol” van tőle.

Ha N darab mérésünk van, akkor ezek a mérések egy N -dimenziós teret feszítenek ki. Az N darab mérési adatból képzett vektor ($\mathbf{u}$) egy vektor ebben a térben, amelynek elemei: $u_1, u_2, \dots, u_N$. A mérési helyekre valamilyen paraméterértékekkel kiszámolva a direktfeladatot ugyancsak egy N dimenziós vektort kapunk ($\mathbf{f}$), amelynek elemei: $f_1, f_2, \dots, f_N$. A (5.44) képlettel kifejezett mennyiség, e két vektor elemeinek különbségeiből képzett négyzetek összege. Ez pontosan az N -dimenziós térben a Pitagorasz-tétel általánosításával nyert **Euklédesszi távolság** négyzete. A következőkben megmutatjuk, hogy a fenti módon képzett mennyiség, a vektortereken általánosan bevezetett távolságfogalom egy speciális esete.

A távolság, és hosszúság fogalmának általánosítása véges dimenziós vektortereken a **norma** fogalma. A norma egy vektortéren értelmezett leképezés, ami a nullvektor kivételével minden $\mathbf{x}$ vektorhoz egy pozitív számot rendel. Érvényesek rá, az abszolútértékhez hasonló tulajdonságok:

$$\begin{aligned} d(\mathbf{x}) &\geq 0 \\ d(\mathbf{x}) = 0 &\Leftrightarrow \mathbf{x} = \mathbf{0} \\ d(\mathbf{x} + \mathbf{y}) &\leq d(\mathbf{x}) + d(\mathbf{y}) \\ d(\lambda \cdot \mathbf{x}) &= |\lambda| \cdot d(\mathbf{x}) \end{aligned} \quad \begin{array}{l} (5.45, \quad 5.46, \quad 5.47, \\ 5.48) \end{array}$$

Ekkor a $d(\mathbf{x})$ -et az $\mathbf{x}$ vektor normájának nevezzük.

A véges (n -dimenziós) valós vektortereken a p -normákat használják:

$$L_p(\mathbf{x}) = \left(\sum_{i=1}^N |x_i|^p \right)^{\frac{1}{p}} \quad (5.49)$$

A (5.44) és a (5.49) egyenletek összehasonlításával látszik, hogy a legkisebb négyzetes becslés kritériumfüggvényében a szereplő mennyiség a $p = 2$ értékhez tartozó, L_2 norma négyzete. Az L_2 norma tehát megegyezik az Euklédesszi terekben definiálható távolsággal. A 2-dimenziós síkon felvett koordináta-rendszer középpontjától egységnyi távolságra levő pontok halmaza az origó köré rajzolt egység sugarú körön van.

Az eltérések négyzeteinek összegét tartalmazó kritériumfüggvényekkel történő becslést, **L_2 norma szerinti becslésnek** nevezzük. Ehhez abból indultunk ki, hogy a hibák normális eloszlásúak, ennek segítségével írtuk fel a likelihood függvényt, amelynek a szélsőértékét (maximumát) meghatározva kapjuk a legkisebb négyzetek formuláját.

Említettük a „A valószínűség-számítás alapjai” c. fejezetben, hogy a hibák normális eloszlásának feltételezését a centrális határeloszlás tétel indokolja, másrészt amennyiben a hibakomponens eloszlása nem ismert, rögzített első két momentum esetén a legnagyobb szabadsági fokkal (más megfogalmazás szerint a maximális entrópiával) rendelkező eloszlást, a **normális eloszlás** választjuk a maximum likelihood függvényben.

Vizsgáljuk meg a $p = 1$ esetet, vagyis az L_1 normát!

$$L_1 = \sum_{i=1}^N |x_i| \quad (5.50)$$

Ebben az esetben az eltéréseink abszolút értékeinek összegét fejezi ki a fenti norma. Az így képzett távolság a koordinátatengelyek mentén mért távolságok összege. Ez 2 dimenzióban olyan, mintha egy derékszögű utcahálózattal rendelkező városban mérnénk a két pont közötti eljutáshoz szükséges távolságot, úgy, hogy csak az utcákon tudunk haladni. Emiatt ez az angol nyelvű szakirodalomban **taxicab** vagy **Manhattan** távolságnak is nevezik. Ennek a távolságdefiníciónak a használatával a síkon felvett 2-dimenziós koordinátarendszer középpontjától egységnyi távolságra levő pontok egy esetén csúcsára állított négyzet oldalainak pontja, amely négyzet csúcsai rendre a (0,1), (1,0) (0,-1) (-1,0) pontok. Természetesen a fenti szemléletes képet több dimenzióra is általánosíthatjuk.

Ennek a mennyiségnek a minimalizálásával történő becslést **L_1 norma szerinti becslésnek** nevezzük.

Amennyiben előzetes ismereteink alapján feltételezhetjük, hogy a hibakomponens Laplace (kétoldali exponenciális) eloszlású, akkor a hibakomponens vektor elemeiből képzett likelihood függvényben a Laplace eloszlás (3.28) összefüggéssel megadott képlete szerepel. Továbbra is feltételezzük, hogy a méréseink függetlenek, a méréseinkből egyetlen csoportot képezünk, amelyben a mérések változékonyságát jellemző b (skála paraméter) valamennyi mérésre azonos. Tételezzük fel továbbá, hogy nincsenek előzetes *a priori* ismereteink a paramétervektor elemeivel

kapcsolatban, így a lehetséges paraméterkombinációk valószínűsége egyforma ($f(\mathbf{p}) = \text{áll.}$). Irjuk fel a mért ($\mathbf{u}$) és a direkt feladatból számolt ($\mathbf{f}(\mathbf{p})$) elméleti értékek különbségeivel a likelihood függvényt:

$$L(\mathbf{p}) = f(\mathbf{u} | \mathbf{p}) = \frac{1}{2 \cdot b} e^{-\frac{|\mathbf{u} - \mathbf{f}(\mathbf{p})|}{b}} \quad (5.51)$$

A függvény maximuma (a vektorelemekkel kiírva) megegyezik az alábbi függvény minimumával:

$$\sum_{i=1}^N |u_i - f_i(\mathbf{p})| = \sum_{i=1}^N |n_i| := \min \quad (5.52)$$

Láthatjuk tehát a (5.50) és (5.52) egyenletek összehasonlításával, hogy az L_1 norma szerinti becslés a Laplace eloszlás esetén megegyezik a maximum likelihood becsléssel.

Vizsgáljuk meg a $\mathbf{p} \rightarrow \infty$ határátmenet képzésével kapott eredményt! Ezt L_∞ normának nevezzük. Ekkor a vektor elemek abszolút értékeinek egyre növekvő kitevőjű hatványait képezzük. A kitevő növelésével a legnagyobb abszolút értékű vektorelemből képzett hatványhoz képest kevésbé növekszik a többi vektorelem, és kimutatható, hogy a hatványok összegéből a p -edik gyökvonás hatására csak a legnagyobb abszolút értékű vektorelem marad meg:

$$L_\infty = \max_{i=1}^N |x_i| \quad (5.53)$$

Ennek a távolságdefiníciónak a használatával a 2-dimenziós esetben, a síkon felvett koordinátarendszer középpontjától egységnyi távolságra levő pontok egy olyan négyzet oldalainak pontja, amely négyzet csúcsai a (1,1) (1,-1) (-1,-1) (-1,1) pontok.

Ezt a normát nevezik angolul **chessboard** (=sakktábla) normának, mivel azt adja meg két mező távolságaként, hogy az egyik mezőre helyezett király hány lépésben tud eljutni a másik mezőre. Az első definiálójáról **Csebishev normának** is nevezik.

Az L_∞ normát használó becsléseket **minimax becslésnek** is nevezzük. Ezt a becslést először Laplace alkalmazta a XVIII. század végén, amikor a különböző földrajzi szélességeken mért meridián ívhosszokból a föld alakját közelítő ellipszoid paramétereit számolta ki.

Amennyiben a hibakomponens vektor elemeiről előzetes ismeretek alapján tudjuk, hogy azok egy egyenletes eloszlásból származnak, akkor a hibakomponens vektor elemei segítségével felírt likelihood függvény szélsőértékének meghatározása, azonos az alábbi feltétellel:

$$\max \{ |u_1 - f_1(\mathbf{p})|, |u_2 - f_2(\mathbf{p})|, \dots, |u_N - f_N(\mathbf{p})| \} := \min \quad (5.54)$$

Az L_1 és L_∞ becslések előnye az L_2 norma szerinti becslésekkel szemben, hogy sokkal kevésbé érzékenyek a durva hibákra, általában függetlenek a véletlen komponens eloszlásától. Az ilyen becsléseket **robosztusnak** nevezzük.

5.4. Minimum kereső eljárások

Az előző fejezetben láttuk, hogy az inverzió egyik fontos lépése a direktfeladatban szereplő paraméterek meghatározása a mérési anyag és az *a priori* ismeretek alapján. Láttuk, hogy ez általánosságban valamilyen kritériumfüggvény minimumának (szélső értékének) a meghatározását jelenti. Az alábbiakban röviden áttekintjük a többváltozós függvények szélsőértékét meghatározó numerikus eljárásokat. Az eljárásokban a levezetéseket mellőzzük.

Lineáris egyenletek megoldására visszavezethető eljárások

Láttuk, hogy a lineáris elméleti terű modelleknél kiemelt szerepe van az L_2 norma szerinti becslésnek, ami a legkisebb négyzetek módszeréhez vezet. A módszer alkalmazása során általában egy

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{h} \quad (5.55)$$

Alakú lineáris egyenlet $\mathbf{x}$ megoldásvektorát keressük, ahol $\mathbf{A}$ egy $N \times N$ méretű invertálható mátrix.

Az egyenlet megoldható a lineáris algebrai bevezetőben bemutatott **általános inverz** segítségével. Az eljárás azonban nem mindig célravezető nagy dimenziószámú vektorok esetén, ezért az adatok és a paraméterek nagy száma esetén **iterációs** eljárásokat alkalmazunk. Ezek lényege, hogy a paramétervektor kiinduló értékét ($\mathbf{x}^{(0)}$ -t) önkényesen megválasztjuk, és valamilyen eljárás segítségével kiszámolunk egy olyan $\Delta \mathbf{x}^{(0)}$ vektort, amit a paramétervektorhoz hozzáadva a kapott vektor ($\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \Delta \mathbf{x}^{(0)}$) már közelebb van az egyenletrendszer tényleges megoldásához. Ezt az eljárást addig ismételjük, amíg elegendően közel nem kerültünk az egyenlet megoldásához.

A kiinduló vektorhoz hozzáadandó $\Delta \mathbf{x}$ vektor kiszámítására többféle eljárás létezik (pl. gradiens módszer, konjugált gradiens módszer, stb.). Ezeknek az eljárásoknak a megértéséhez először meg kell ismerkednünk a **kvadratikus függvény** fogalmával.

Definiáljuk az $\mathbf{x} \in \mathbb{R}^N$ vektorhoz tartozó kvadratikus függvényt:

$$f(\mathbf{x}) = \frac{1}{2}(\mathbf{x}^T \cdot \mathbf{A} \cdot \mathbf{x}) - (\mathbf{h}^T \cdot \mathbf{x}) + c \quad (5.56)$$

Ahol $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{h}, \mathbf{x} \in \mathbb{R}^N$, c pedig egy konstans. A kvadratikus függvény tulajdonságai megegyeznek az $\mathbf{A}$ mátrix tulajdonságaival, tehát lehet szimmetrikus, pozitív definit, stb. A fenti egyenlet $\mathbf{x}$ szerinti szélsőértéke ott van, ahol az egyenlet $\mathbf{x}$ szerinti deriváltja nulla. A függvény deriváltja:

$$f'(\mathbf{x}) = \mathbf{A} \cdot \mathbf{x} - \mathbf{h} \quad (5.57)$$

Vagyis egy kvadratikus függvény szélsőérték helyének meghatározása egyenértékű a $\mathbf{A} \cdot \mathbf{x} = \mathbf{h}$ egyenlet megoldásával.

Ha az $\mathbf{A}$ mátrix szimmetrikus és pozitív definit, valamint egy $\mathbf{z}$ vektor kielégíti a $\mathbf{A} \cdot \mathbf{z} = \mathbf{h}$ egyenletet, akkor a kvadratikus függvény az alábbi alakba írható:

$$f(\mathbf{x}) = f(\mathbf{z}) + \frac{1}{2}((\mathbf{x} - \mathbf{z})^T \cdot \mathbf{A} \cdot (\mathbf{x} - \mathbf{z})) \quad (5.58)$$

A jobboldali tag nem negatív, ezért a $\mathbf{z}$ vektor a kvadratikus függvény minimumhelye. Kimutatható, hogy a fenti esetben a $\mathbf{z}$ vektor az egyetlen minimumhely.

Gradiens módszer

A fentiek értelmében a $\mathbf{A} \cdot \mathbf{x} = \mathbf{h}$ egyenlet megoldása azonos a kvadrátikus függvény minimumhelyének meghatározásával. Jelöljük a kvadrátikus függvény negatív gradiensét egy $\mathbf{r}$ vektorral:

$$\mathbf{r} = -f'(\mathbf{x}) = \mathbf{h} - \mathbf{A} \cdot \mathbf{x} = \mathbf{A} \cdot (\mathbf{z} - \mathbf{x}) \quad (5.59)$$

Az $\mathbf{r}$ vektor a korábban megismert **reziduálvektor**.

Találhatunk olyan α számot, amelyre teljesül az alábbi két feltétel:

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{x} + 2 \cdot \alpha \cdot \mathbf{r}) \\ f(\mathbf{x} + \alpha \cdot \mathbf{r}) - f(\mathbf{x}) &\leq 0 \end{aligned} \quad (5.60, 5.61)$$

Egy kvadrátikus függvény szintfelületei általánosított ellipszoidok, 2-dimenzióban ellipszisek. A felső egyenlet azt fejezi ki, hogy a paraméterter egyik ($\mathbf{x}$) pontjáról a gradiens irányában $2 \cdot \alpha \cdot \mathbf{r}$ vektorral arrébb mozdulva ugyanarra a szintfelületre (2-dimenzióban szintvonalra) jutunk. Az alsó egyenlet azt fejezi ki, hogy ha ennek a távolságnak a felével megyünk csak arrébb, akkor közeledünk a minimumhoz (vagy legalábbis nem távolodunk tőle).

A részletek mellőzésével közöljük, hogy az α skalár az alábbi módon számolható:

$$\alpha = \frac{(\mathbf{r}^T \cdot \mathbf{r})}{(\mathbf{r}^T \cdot \mathbf{A} \cdot \mathbf{r})} \quad (5.62)$$

Látszik, hogy ennek kiszámításához egy önkényesen választott $\mathbf{x}$ hely esetén csak az $\mathbf{A}$ és a $\mathbf{r} = \mathbf{h} - \mathbf{A} \cdot \mathbf{x}$ mennyiségek kelljenek.

Összefoglalásként megadható a gradiens módszer algoritmus:

Válasszunk egy tetszőleges $\mathbf{x}^{(0)}$ vektort. Számítsuk ki hozzá a $\mathbf{r}^{(0)} = \mathbf{h} - \mathbf{A} \cdot \mathbf{x}^{(0)}$ gradiens vektort. Számítsuk ki az alábbi mennyiségeket ($k=1$):

$$\begin{aligned} \alpha^{(k)} &= \frac{\left(\left(\mathbf{r}^{(k-1)} \right)^T \cdot \mathbf{r}^{(k-1)} \right)}{\left(\left(\mathbf{r}^{(k-1)} \right)^T \cdot \mathbf{A} \cdot \mathbf{r}^{(k-1)} \right)} \\ \mathbf{x}^{(k)} &= \mathbf{x}^{(k-1)} + \alpha^{(k)} \cdot \mathbf{r}^{(k-1)} \\ \mathbf{r}^k &= \mathbf{h} - \mathbf{A} \cdot \mathbf{x}^{(k)} \end{aligned} \quad (5.63, 5.64, 5.65)$$

A fenti számítást ismételjük meg $k=2,3,4,\dots$ számokra. Az $\mathbf{x}^{(k)}$ vektorok sorozata az $\mathbf{A} \cdot \mathbf{x} = \mathbf{h}$ egyenletrendszer $\mathbf{x}$ megoldásához tart.

Az eljárás hátránya, hogy amennyiben a kvadrátikus függvény szintvonalai nagyon elnyúlt ellipszisek, akkor a konvergencia nagyon lassú.

Konjugált vektor módszer

Említettük, hogy a gradiens módszer hátránya, hogy rosszul kondicionált mátrixok esetén az $\mathbf{r}$ reziduálvektor irányába tett lépések csak rosszul konvergálnak a minimumhoz. Válasszunk az $\mathbf{r}$ vektor helyett egy (később meghatározandó) $\mathbf{p}$ vektort, amit a (5.60) egyenletbe betéve a minimum helyé felé konvergáló megoldást kapnánk.

Eszerint a (5.64) egyenletben az $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \alpha^{(k)} \cdot \mathbf{r}^{(k-1)}$ helyett $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \alpha^{(k)} \cdot \mathbf{p}^{(k-1)}$ áll. Eszerint szükségünk lenne a $\mathbf{p}^{(k)}$ vektorok sorozatára.

Válasszuk úgy a $\mathbf{p}^{(k)}$ vektorokat, hogy azok a pozitív definit $\mathbf{A}$ mátrix konjugáltjai legyenek, vagyis $\mathbf{p}^{(i)\top} \cdot \mathbf{A} \cdot \mathbf{p}^{(j)} = 0$ ha $i \neq j$. Ekkor ezek a vektorok egy **bázis** alkotnak és ezek a vektorok lineárisan függetlenek. Ebből következik, hogy egy tetszőleges $\mathbf{y}$ vektor felírható az N darab $\mathbf{p}$ vektor lineárkombinációjaként (N a vektor dimenziója).

Írjuk fel a kiinduló $\mathbf{x}^{(0)}$ vektor és az egyenlet megoldását jelentő $\mathbf{z}$ vektor különbségét, mint a $\mathbf{p}^{(k)}$ vektorok lineárkombinációját!

$$\mathbf{z} - \mathbf{x}^{(0)} = \sum_{i=0}^{N-1} \alpha_{i+1} \cdot \mathbf{p}^{(i)} \quad (5.66)$$

ahol

$$\alpha_{i+1} = \frac{\left(\left(\mathbf{p}^{(i)} \right)^\top \cdot \mathbf{r}^{(i)} \right)}{\left(\left(\mathbf{p}^{(i)} \right)^\top \cdot \mathbf{A} \cdot \mathbf{p}^{(i)} \right)} \quad (5.67)$$

(Ez az eredmény a kvadratikus függvény $\mathbf{x}_k + \alpha \cdot \mathbf{p}_k$ helyen felvett értékének a Taylor-sorfejtésével kapható meg.)

Eszerint ha a $\mathbf{p}^{(k)}$ vektorok az $\mathbf{A}$ mátrixnak konjugált vektorai akkor a gradiens módszer módosítása ($\mathbf{r}$ helyett $\mathbf{p}$ vektorral számolva) legfeljebb N lépésben az egzakt megoldáshoz konvergál, ahol N az $\mathbf{A}$ mátrix dimenziója.

Konjugált gradiens módszer

A konjugált gradiens módszer lényege, hogy a konjugált vektor módszerben szereplő $\mathbf{p}^{(k)}$ vektorokat iteratív módon állítjuk elő, úgy, hogy az iteráció első lépésében $\mathbf{p}^{(0)}$ vektor az $\mathbf{r}^{(0)}$ (gradiens)vektor.

$$\mathbf{p}^{(k)} = \mathbf{r}^{(k)} + \sum_{j=1}^{k-1} \beta_{k,j} \cdot \mathbf{p}^{(j)} \quad (5.68)$$

Vagyis az aktuális k -adik iterációs lépésben felhasznált vektort állítsuk elő a gradiens vektor ($\mathbf{r}$) és a korábbi $\mathbf{p}^{(0)}$, $\mathbf{p}^{(1)}$, ..., $\mathbf{p}^{(k-1)}$ vektorok lineárkombinációjaként. Ehhez a $\beta_{k,j}$ együtthatók meghatározása szükséges. (Ezek a $\beta_{k,j}$ együtthatók biztosítják, hogy a konjugáltság fennálljon.)

A számításokat mellőzve kimutatható, hogy a $\beta_{k,j}$ együtthatókra, adott k esetén az egyetlen nem nulla együttható a $\beta_{k,k-1}$.

Ekkor a konjugált gradiens módszer szerint, egy tetszőleges $\mathbf{x}^{(0)}$ vektort választva, és kiszámolva hozzá a $\mathbf{p}^{(0)} = \mathbf{r}^{(0)} = \mathbf{h} - \mathbf{A} \cdot \mathbf{x}^{(0)}$ vektorokat, $k=1, 2, \dots, N$ esetén a kiszámolva az alábbi mennyiségeket:

$$\begin{aligned}
\alpha_{k+1} &= \frac{\left(\left(\mathbf{r}^{(k)} \right)^T \cdot \mathbf{r}^{(k)} \right)}{\left(\left(\mathbf{p}^{(k)} \right)^T \cdot \mathbf{A} \cdot \mathbf{p}^{(k)} \right)} \\
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} + \alpha_{k+1} \cdot \mathbf{p}^{(k)} \\
\mathbf{r}^{(k+1)} &= \mathbf{r}^{(k)} - \alpha_{k+1} \cdot \mathbf{A} \cdot \mathbf{p}^{(k)} \\
\beta_{k+1} &= \frac{\left(\left(\mathbf{r}^{(k+1)} \right)^T \cdot \mathbf{r}^{(k+1)} \right)}{\left(\left(\mathbf{r}^{(k)} \right)^T \cdot \mathbf{r}^{(k)} \right)} \\
\mathbf{p}^{(k+1)} &= \mathbf{r}^{(k+1)} + \beta_{k+1} \cdot \mathbf{p}^{(k)}
\end{aligned}
\tag{5.69, 5.70, 5.71, 5.72, 5.73}$$

Az $k+1=N$ esetén az $\mathbf{x}^{(N)}$ vektor az egyenlet keresett megoldását szolgáltatja. Ez az eljárás viszonylag egyszerű, könnyen programozható, ezért nagy népszerűségnek örvend.

Intervallum keresés

Ennél a módszernél előzetes ismeretek alapján meghatározzuk, hogy a paramétervektor egyes elemei milyen intervallumban vehetnek fel értékeket. (Egy egyváltozós függvény esetén ez a számegegyenesen egy szakasz, kétváltozós függvény esetén ez a síkon egy téglalap.) Ezután a paraméterternek ezt a részét minden paraméter-tengely mentén felosztjuk kis szakaszokra. (Ez a számegegyenesen a szakasz felosztását, a síkon a téglalapra berácsosozását jelenti.) Ezután a felosztás határpontjain, (a szakaszok végpontjain illetve a rácpontokban) kiszámoljuk a kritériumfüggvény értékét. Ahol a kritériumfüggvény legkisebb, ott annak a paramétervektornak a környezetében – kisebb szakaszokat választva – finomítjuk a felbontást és ezekben a pontokon újra kiszámoljuk a kritériumfüggvény értékeit. Ezt az eljárást többször megismételve, elegendő pontossággal megkaphatjuk egy többváltozós függvény minimumát. Az eljárás során nem szükséges az egyes paraméterekre megadható intervallumokat azonos darabszámú szakaszokra osztani, illetve egy-egy paraméter esetén nem szükséges, hogy szakaszok egyforma hosszúságúak legyenek.

Az eljárás előnye, hogy egyszerűen programozható, és a direkt feladat számítását igényli csak, nem szükséges az elméletei válaszfüggvény deriváltjainak számítása.

Hátránya, hogy sokelemű paramétervektor esetén nagyon megnő a számításigény: ha az N elemű (dimenziójú) paramétervektor mindegyik elemét $M-1$ szakaszra (intervallumra) osztjuk fel, akkor M^N alkalommal kell kiszámolnunk a függvény értékét egyetlen iterációs lépésben.

Szimplex módszer

A **Nelder-Mead féle szimplex módszer** (más néven amőba módszert) 1947-ben javasolták lineáris programozási problémák megoldására. (A lineáris programozás általános feladata többváltozós lineáris függvény szélsőértékének keresése adott feltételek mellett. A feltételek lineáris egyenletek, vagy egyenlőtlenségek alakban adóttak. A lineáris programozást eleinte gazdasági és katonai logisztikai problémák megoldására használták.)

A szimplex módszer a szélsőérték meghatározásához nem igényli a függvény deriváltjainak számítását, ezért már a számítógépek használatának kezdetén alkalmazható volt, amikor még nem voltak kidolgozva a numerikus deriváltakat számító eljárások.

A **szimplex** egy N dimenziós térben egy $N+1$ csúcsponttal rendelkező idom. (A 2-dimenziós síkon ilyen idom a háromszög, a 3-dimenziós térben a tetraéder.)

A szimplex módszer egy iterációs eljárás, amely képes több lépésben megtalálni a függvény minimumhelyét, ha a függvény „sima” és egyetlen minimumhelye van.

Az inverzió során a paraméterek kiinduló értéke a paraméterek terének egy pontja. Ez lesz a $\mathbf{x}^{(0)}$ vektor. A módszer inicializálása során megadunk egy $\Delta \mathbf{x}^{(0)}$ vektort, aminek i -edik eleme azt adja meg, hogy az i -edik paramétert

mekkora lépéssel léptetjük arrébb a paraméterterben. A kiinduló szimplexet úgy képezzük, hogy a paramétervektor kiinduló értéke ($\mathbf{x}^{(0)}$) lesz a szimplex egyik csúcsa, az $(i+1)$ -edik csúcsot pedig úgy kapjuk, hogy ennek a $\mathbf{x}^{(0)}$ vektornak i -edik elemét $\Delta \mathbf{x}^{(0)}$ i -edik elemével megnöveljük. Ekkor a változatlanul hagyott paramétervektor és az összesen N darab, egy-egy elemében megváltoztatott vektor adja a szimplex $N+1$ darab csúcsát.

Az eljárás röviden úgy írható le, hogy egy kezdetben felvett szimplex csúcsaira kiszámoljuk a függvény értékeket. Azt a csúcsot, amelynél a függvény értéke a legnagyobb, áttükrözzük a szimplex súlypontjára. Ekkor egy új szimplexet kapunk, úgy, hogy elhagyjuk a legnagyobb függvényértékű csúcsot és helyére a szimplexbe a tükrözéssel kapott csúcsot tesszük. Ennek az eljárásnak az ismétlésével a szimplex „lépeget”, a szimplex a függvény minimumához egyre közeledik. Ha a tükrözéssel újonnan kapott csúcsban a függvény értéke kisebb, mint a szimplex bármelyik csúcsában, akkor egy nyújtást végzünk a minimum irányában. Ha a tükrözéssel kapott ponton a függvény értéke nem csökken, akkor a minimumhelyet körülveszi a szimplex, ezért a szimplexet összehúzzuk. Ha ennek hatására sem találunk olyan csúcsot, amelyiken a függvény értéke csökkenne, akkor a legjobb csúcsra „ráhúzzuk” a szimplexet.

A következőkben bemutatjuk az eljárás részletes leírását.

A módszer inicializálásakor vegyünk fel az N -dimenziós térben egy $N+1$ csúcsú szimplexet. Ennek csúcsait jelöljék $(\mathbf{x}_1^{(0)}, \mathbf{x}_2^{(0)}, \dots, \mathbf{x}_{N+1}^{(0)})$ vektorok. (a felső index az iterációs lépések számát jelenti, az alsó index a csúcsok indexe. Minden vektor N elemű.)

1. Számítsuk ki ebben az $N+1$ csúcsban a függvény értékét! Ez után rakjuk sorba a vektorokat úgy, hogy a hozzájuk számított függvényértékek nagyság szerint növekedjenek: $f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \dots \leq f(\mathbf{x}_{N+1})$
2. Számítsuk ki az $\mathbf{x}$ vektorok $\mathbf{x}_0$ súlypontját, úgy, hogy elhagyjuk az $\mathbf{x}_{N+1}$ vektort!
3. Tükrözés: Tükrözzük a $\mathbf{x}_{N+1}$ pontot a súlypontra! Ez a tükrözött pont: $\mathbf{x}_r = \mathbf{x}_0 + \alpha \cdot (\mathbf{x}_0 - \mathbf{x}_{N+1})$ Ha a tükrözött pont jobb, mint a második legrosszabb pont, de nem jobb, mint a legjobb pont (vagyis $f(\mathbf{x}_1) \leq f(\mathbf{x}_r) < f(\mathbf{x}_N)$), akkor helyettesítsük az $\mathbf{x}_{N+1}$ pontot a $\mathbf{x}_r$ ponttal, és ismételjük meg az 1. lépést.
4. Ha a tükrözött pont jobb, mint az eddigi legjobb pont, (vagyis $f(\mathbf{x}_r) < f(\mathbf{x}_1)$) akkor számoljuk ki az $\mathbf{x}_e = \mathbf{x}_0 + \gamma \cdot (\mathbf{x}_0 - \mathbf{x}_{N+1})$ elnyújtott pontot. Ha ez az elnyújtott pont jobb, mint a tükrözött pont (vagyis $f(\mathbf{x}_e) < f(\mathbf{x}_r)$), akkor helyettesítsük az $\mathbf{x}_{N+1}$ legrosszabb pontot az $\mathbf{x}_e$ elnyújtott ponttal és ismételjük meg az 1. lépést. Ha ez az elnyújtott pont nem jobb, mint a tükrözött pont akkor helyettesítsük az a $\mathbf{x}_{N+1}$ legrosszabb pontot a $\mathbf{x}_r$ tükrözött ponttal és ismételjük meg az 1. lépést.
5. Ha a tükrözött pont rosszabb, mint az eddigi második legrosszabb pont, (vagyis $f(\mathbf{x}_r) \geq f(\mathbf{x}_N)$) akkor számoljuk ki az $\mathbf{x}_c = \mathbf{x}_0 + \rho \cdot (\mathbf{x}_0 - \mathbf{x}_{N+1})$ összehúzott pontot. Ha ez az összehúzott pont, jobb mint a legrosszabb pont (vagyis $f(\mathbf{x}_c) < f(\mathbf{x}_{N+1})$) akkor helyettesítsük az a $\mathbf{x}_{N+1}$ legrosszabb pontot a $\mathbf{x}_c$ összehúzott ponttal és ismételjük meg az 1. lépést. Egyébként lépünk tovább a 6. pontra.
6. A legjobb pont kivételével az összes pontot helyettesítsük az alábbi mennyiséggel: $\mathbf{x}_i = \mathbf{x}_1 + \sigma \cdot (\mathbf{x}_i - \mathbf{x}_1)$ ahol $i=2,3,\dots,N$. Ezután ismételjük meg az 1. lépést.

Az α a tükrözési, γ a nyújtási, ρ az összehúzási, σ pedig a zsugorítási együttható. Ezek szokásos értékei: $\alpha=1$, $\gamma=2$, $\rho=-1/2$, $\sigma=1/2$. A 6. pont abban az esetben fordul elő, ha a minimumhely felé közeledve valamilyen irányból a függvény még növekszik, ekkor a szimplex összehúzása az 5. lépésben nem csökkenti a függvény értékét.

Több változó és összetett direkt feladat esetén a függvénynek több lokális minimumhelye lehet. A globális minimumhely megtalálásához lehetséges, hogy a szimplexet a paramétertér több pontjáról kell elindítani, ezzel feltérképezve az összes lokális minimumot.

A szimulált hűtés (Simulated Annealing)

A szimulált hűtés (Simulated Annealing; SA) módszert, arról a folyamatról nevezzük el, amelyben az olvadt anyag megszilárdulva más, jellemzően alacsonyabb energiaszintű kristályállapotba kerül azt követően, hogy lassan vagy épp gyorsan lehűlt. A fizikai rendszerek mindig a minimális energiájú állapot felé törekszenek, és e rendszerek matematikai modellezése a tetszőleges függvényalakú, ún. energiafüggvények minimalizálására is alkalmas.

Az olvadt anyagok atomszerkezete és – ennek következtében – energiaállapota is állandóan változik. Amennyiben az olvadt anyag termodinamikai szempontból egyensúlyban van, az energia-eloszlását a Boltzmann-féle valószínűségi függvénnyel adhatjuk meg:

$$P(E) \propto e^{\left(-\frac{E}{kT}\right)} \quad (5.74)$$

ahol E jelöli az energiát, k a Boltzmann-állandót, T pedig a hőmérsékletet. Az energiaállapot valószínűségi jellege miatt az energiaszint akár igen kis hőmérséklet esetén is lehet viszonylag magas. Ez számunkra azért fontos, mert azáltal, hogy az olvadt anyag energiaszintje alacsonyabból magasabbra vált, előfordulhat, hogy a rendszer kilép egy lokális energiaminimum-helyzetből. Valójában nem a rendszer energiaszintje függ a hőmérséklettől, hanem annak növekedési valószínűsége:

$$P(E_1 \rightarrow E_2) = e^{\left(-\frac{E_2 - E_1}{kT}\right)} \quad \text{ha } E_1 < E_2 \quad (5.75)$$

Így minél magasabb a rendszer hőmérséklete, annál valószínűbb, hogy energiaszintje is növekedni fog. Lassú hűtés esetén az olvadt anyag energiaszintje csökkenni fog, a hűtés befejeztével minimális értéket vesz fel, „nulla hőmérsékleten” pedig ez az állapot rögzül, az anyag ebben az állapotban fagy meg.

Az SA inverziós módszer alkalmazásával a folyamat modellezését használjuk fel a célfüggvény globális minimumának meghatározására. Ebben a modellben az olvadt anyag atomszerkezetének különböző értékeit a saját modellünk paramétereivel fogjuk megfeleltetni. A fizikai rendszer energiaszintjének megfelelője a mi eljárásunkban a minimalizálandó célfüggvény lesz. A módszer alkalmazása során a modellünk terében hajtunk végre irányított véletlen lépéseket, és a minimalizálandó $\Phi(\mathbf{m})$ célfüggvény értékeit pedig a lépéseket követően feljegyezzük. A kiinduló modellünket jelöljük $\mathbf{m}_0$ -lal, $\Phi_0 = \Phi(\mathbf{m}_0)$ pedig jelölje a célfüggvény ehhez tartozó értékét. Az irányított véletlen elmozdulás során a modellterben egy $\Delta\mathbf{m}$ értékkel elmozdulunk és megállapítjuk a célfüggvény Φ_1 értékét ebben az $\mathbf{m}_1 = \mathbf{m}_0 + \Delta\mathbf{m}$ pontban. A $\Delta\mathbf{m}$ lépés hosszát az adott feladathoz kiválasztva azonos értéként kezeljük, irányát pedig véletlenszerűen adjuk meg. A létrejövő új $\mathbf{m}_1$ modell elfogadásának valószínűsége pedig a következőképpen adható meg:

$$p = 1 \quad \text{ha } \Delta\Phi = \Phi_1 - \Phi_0 \leq 0 \quad (5.76)$$

$$p = e^{\left(-\frac{\Delta\Phi}{T}\right)} \quad \text{ha } \Delta\Phi > 0 \quad (5.77)$$

amelyben T egy pozitív paraméter, amelyet a (5.74) egyenlettel analóg módon kT -vel azonosítunk. A (5.76) egyenlet azt mondja ki, hogy a létrejövő új $\mathbf{m}_1$ modellt akkor fogadjuk el feltétel nélkül, ha a célfüggvény értéke csökkent ($\Phi_1 \leq \Phi_0$). Az $\mathbf{m}_1$ modell elfogadási valószínűsége viszont akkor sem nulla, ha a célfüggvény értéke növekszik; ez teszi lehetővé, hogy az eljárás ne ragadjon meg a paramétertér egy lokális minimumában. Minél kevésbé növekszik a célfüggvény értéke, annál inkább elfogadjuk ekkor is az új paraméter-halmazt.

Az elfogadás után az $\mathbf{m}_1$ modellt már $\mathbf{m}_0$ -nak tekintve kezdjük előlről a véletlen ugrást, míg ha nem fogadjuk el az új paramétersort, az $\mathbf{m}_0$ értéke nem változik, és e régi helyről kísérünk meg újabb véletlen ugrást.

A hőmérséklet-analóg T paraméter értékétől függ, hogy egy lokális minimumból hány lépéssel sikerül kitörni. Itt nem tárgyalt elméleti okokból T értékének megválasztása az alábbi korlátokkal történik:

$$0,5 < e^{\left(-\frac{\Delta\Phi}{T}\right)} < 0,9 \quad (5.78)$$

Erre azért van szükség, mert ha az elfogadási valószínűség értéke 1-hez közeli, akkor bármilyen lépést elfogadjunk, ha pedig túl alacsony, akkor túl sok lépéssel tudunk csak kitörni egy-egy lokális minimumból. A lépések abszolútértékét is úgy kell megválasszuk, hogy a lokális minimumból néhány lépéssel ki lehessen törni.

Az algoritmus ismételt végrehajtása során a T paramétert – akárcsak a fizikai rendszerbeli hűtés során a hőmérsékletet – fokozatosan csökkentjük. Ahogy a globális minimumhoz közelítünk, a rendszer mintegy „megfagy”, és a túl kicsi lépések miatt abból már nem tud kitörni. Az egyik legegyszerűbb hűtési modellt a

$$T_{k+1} = \gamma \cdot T_k \quad (5.79)$$

összefüggés írja le. Itt k az aktuális modell-átmenetet jelző sorszámozott index, γ pedig egy 1-nél kisebb, de ahhoz közeli állandó érték.

A fentiek összefoglalva megadhatjuk az SA eljárást algoritmikus formában is:

Legyen $\mathbf{m}_0$ a modell kiinduló állapota, $\Phi(\mathbf{m})$ az ehhez tartozó, minimalizálandó célfüggvény, T_0 pedig egy kellően magas kiinduló hőmérséklet. Az algoritmus lépései a következők:

1. Legyen $\Phi_0 = \Phi(\mathbf{m}_0)$ és $T_0 = \gamma \cdot T_0$, ahol $\gamma < 1$, de ≈ 1 (Pl. $\gamma = 0.98$).
2. Véletlenszerűen adjunk meg egy előre rögzített hosszúságú $\Delta\mathbf{m}$ elmozdulásvektort
3. Legyen $\mathbf{m}_1 = \mathbf{m}_0 + \Delta\mathbf{m}$ és $\Phi_1 = \Phi(\mathbf{m}_1)$.
4. Ha $\Phi_1 \leq \Phi_0$, legyen $\mathbf{m}_0 = \mathbf{m}_1$. Folytatás az 1. lépéstől.
5. Ha $\Phi_1 \geq \Phi_0$, legyen $\Delta\Phi = \Phi_1 - \Phi_0$ és $p = e^{\left(-\frac{\Delta\Phi}{T}\right)}$.
 - (a) Generáljunk egy q véletlen számot a $[0,1]$ intervallumból.
 - (b) Ha $q < p$, legyen $\mathbf{m}_0 = \mathbf{m}_1$. Folytatás az 1. lépéstől.
 - (c) Ha $q > p$, folytatás a 2. lépéstől.

Az genetikusan algoritmus

A genetikusan algoritmus (Genetic Algorithm; GA) — az előző pontban tárgyalt szimulált hűtéshez hasonlóan — egy valódi, a természetben is végbemenő folyamatot, az evolúciós kiválasztódást modellezi. A természetben azok az egyedek és fajok, amelyek egy külső követelményrendszerhez jól alkalmazkodnak, túlélnek és szaporodni fognak, a többiek pedig kihalnak. A modell-térnek a paraméter-halmazzal jellemzett pontjait esetünkben a biológiai egyedekkel azonosítjuk. A túlélés feltételének azt tekintjük, hogy a minimalizálandó célfüggvény hozzájuk rendelt értéke minél kisebb legyen. A természetes kiválasztódás folyamatának analógiájára végül egy olyan populációt kapunk, melynek egyedei a célfüggvény globális minimuma közelében „élnek”.

Adjunk meg az M -dimenziós modellterben egy alteret, megadva az $\mathbf{m} = (m_1, m_2, \dots, m_M)$ modellparaméterek lehetséges minimális és maximális értékeit megadó $\mathbf{a} = (a_1, a_2, \dots, a_M)^T$ és $\mathbf{b} = (b_1, b_2, \dots, b_M)^T$ vektorokat! Ez az alter legyen diszkrét: az m_i modellparaméterhez tartozó $[a_i, b_i]$ intervallumot 2^{N_i} szintre osztjuk. Ezáltal az m_i diszkrét értékei egy N_i hosszúságú bináris adatlánc elemeiként állnak elő. Amikor az M modellparaméterhez tartozó bináris adatlánccokat összefűzzük, létrejön egy $\sum N_i$ bitből álló ún. „**kromoszóma**”, amely esetünkben a modellter egy pontja lesz. Ez a diszkrét modellter $2^{\sum N_i}$ modellt tartalmaz, a **populáció** pedig n darab kromoszómából áll. A kiinduló populáció kromoszómaát a modellterből véletlenszerűen választjuk ki. A természetes kiválasztódást pedig az alábbi három folyamattal modellezzük:

Kiválasztás: Válasszunk ki az a meglévő populációból úgy n darab modellt, hogy az m_k modell kiválasztási valószínűsége a $\Phi(\mathbf{m}_k)$ célfüggvény értékétől függjön úgy, hogy minél kisebb a célfüggvény $\Phi(\mathbf{m}_k)$ értéke, annál valószínűbb, hogy a modellt ki is választjuk. E kiválasztási valószínűség egy lehetséges módja például:

$$P_s(\mathbf{m}_k) = \frac{\Phi_{\max} - \Phi(\mathbf{m}_k)}{\sum_{i=1}^n (\Phi_{\max} - \Phi(\mathbf{m}_i))} \quad (5.80)$$

ahol $\Phi_{\max}$ a populáció maximális célfüggvény-értéke. Ez az eljárás biztosítja, hogy a legkisebb célfüggvény-értékeket produkáló modellek megmaradnak, míg a nagyobb értékekkel jellemzettek „kipusztulnak”, eltűnnek a populációból.

Kereszteződés: A kiválasztást követően a „kromoszómáinkat” véletlen párosítással $n/2$ párba csoportosítjuk és a párok tagjait (a „szülő kromoszómákat”) keresztezzük egymással. A kereszteződés bekövetkezési valószínűsége legyen P_c ! A kereszteződéskor véletlenszerűen kiválasztunk egy bitet a kromoszómák bináris adatláncában, és az ettől jobbra lévő biteket a szülő-kromoszómák közt felcseréljük, két új kromoszómát hozva létre ezzel. Amennyiben nem történik kereszteződés, akkor az utódok kromoszómái megegyeznek a szülőkével.

Mutáció: Az így kapott új populáció kromoszómaát tovább változtathatjuk véletlenszerű mutáció által. Ennek bekövetkezési valószínűsége legyen P_m . Amikor mutáció történik, a kromoszóma egyik bitjét véletlenszerűen kiválasztjuk és értékét az ellenkezőjére módosítjuk (0 helyett 1-et, 1 helyett pedig 0-t írunk bele). Ezzel a véletlenszerű beavatkozással válik lehetővé, hogy a rendszerben új információ jelenhessen meg. A mutáció teszi lehetővé, hogy új információkat hordozó kromoszómák jöwhessenek létre. Mindazonáltal a P_m valószínűség értéke nem lehet akármilyen nagy, célszerű azt alacsonyan, 1-2% körül tartani, egyébként az eljárás inkább a tisztán véletlen lépéseket tartalmazó Monte-Carlo-módszerhez lesz hasonlatos.

A fenti kiválasztás-keresztezés-mutáció műveleti hármast addig ismételjük egymás után, amíg a populáció nem lesz homogén, vagyis az egyes „kromoszómák” (modellparaméter-halmazok) nem válnak egymáshoz nagyon hasonlóvá, vagy amíg a létrejött egyedszám nem ér el egy előre beállított értéket. Arra azonban nincs garancia, hogy az eljárás eredményeként a „kromoszómák” tartalma valóban a globális minimum közelében lesz. A sok különböző helyen elvégzett célfüggvény-értékelés mellett mindazonáltal egy lokális minimum-gödörben való „beragadás” esélye meglehetősen kicsi, a végső helyzetet elérő, „túlélő” populáció egyedeinek „kromoszómái” általában a globális minimum közelébe esnek, és a vizsgált inverziós probléma megoldását jelentik.

5.5. A statisztikai becslések minőségét ellenőrző eljárások

Az eddig elmondottak összefoglalásából kiemelhetjük, hogy a méréseink összessége, (a kísérleti anyag) valószínűségi változók realizációjának tekinthető, így az ebből levezett minőségi vagy mennyiségi jellemzők (a paraméterek) maguk is valószínűségi változók.

Az előző fejezetben láttuk, hogy ezeknek a jellemzőknek a meghatározásához olyan becslési eljárást választottunk (a maximum likelihood becslést), amely ezeknek a jellemzőknek, a paramétereknek (legalább aszimptotikusan) torzítatlan és legkisebb szórású (effektív) becsléseit adják. Ebben a fejezetben, a mennyiségi interpretáción keresztül bemutatjuk a becsült paraméterek statisztikai jellemzőinek a kiszámítási módját.

A statisztikus megközelítés szemszögéből a becsült paraméterek statisztikus jellemzőinek meghatározása ugyanolyan fontos eleme az inverzióknak, mint a maguknak a paramétereknek a meghatározása. Ebből a szempontból az inverzió nem teljes addig, amíg ezeket a statisztikai jellemzőket is meg nem határoztuk.

Ebben a fejezetben csak a mennyiségi interpretáció esetét tárgyaljuk. A minőségi interpretáció esete nagy hasonlóságot mutat a statisztikai próbák elméletével, amit ez a jegyzet nem tárgyal. Az összetett (mennyiségi-minőségi) interpretáció esetét a mennyiségi interpretáció után, csak vázlatosan tárgyaljuk.

A mennyiségi interpretáció során, az alkalmazott maximum likelihood becslés segítségével meghatároztuk a modellünkben szereplő paraméterek becsült értékeit ($\hat{\mathbf{p}}$). Ahogy említettük, ezek maguk is valószínűségi változókból (a mérésekből) származtatott mennyiségek. Az ezekre a valószínűségi változókra vonatkozó legteljesebb információt ezeknek a mennyiségeknek (a paramétereknek) az eloszlásfüggvénye szolgáltatja. A teljes eloszlásfüggvényt általában nem tudjuk meghatározni. Feltételezve, hogy a paraméterek normális eloszlásúak, elegendő az első két momentumukat ismerni az eloszlás pontos leírásához.

Az első momentum a várható érték. Egy becslés torzítatlansága azt jelenti, hogy a becsült paraméter várható értéke megegyezik a paraméter elméleti értékével. Ez a feltétel általában nem teljesül, hanem csak az aszimptotikus torzítatlanság. A gyakorlatban sok mérés és kevés meghatározandó paraméter esetén ez általában teljesül, és egy paraméter becsült várható értéke megegyezik az adott paraméter elméleti értékével.

A becsléseinkkel szemben a torzítatlanság mellett a másik elvárásunk az effektivitás volt, vagyis hogy a becsült paraméterek szórása minimális legyen. Optimálisnak tekintettük azokat a becsléseket, amelyek ezt (a torzítatlanság mellett) megvalósították. Minthogy ilyen becsléseket nehéz konstruálni, általában normális eloszlásúnak tekinthető véletlen komponens modell esetén a maximum likelihood becslést használjuk, tudva, hogy az általa szolgáltatott becslés csak **szuboptimális**. Az optimális (effektív) becslés tehát az elérhető legkisebb szórással becsli a paramétereket, a gyakorlatban használt maximum likelihood becslés viszont ennél nagyobb szórással.

Vizsgáljuk meg a paraméterek szórását leíró mennyiségeket! Említettük, hogy egy többváltozós normális eloszlást egyértelműen megad a várható értékek vektora és a centrális momentumokból képzett kovariancia mátrix. Eszerint

a **paraméterek kovariancia mátrixa**, $\mathbf{D}(\hat{\mathbf{p}})$ az a mennyiség, amit meg kell határoznunk az effektivitás vizsgálatához.

Egy valószínűségi vektorváltozó függvényeként előállt vektorváltozó kovarianciamátrixát (3.88) a hibaterjedést leíró fejezetben meghatároztuk. Az ottani eredmények a valószínűségi változók lineáris függvényeire vonatkoznak, így azokat a lineáris additív terű modellekre változtatás nélkül alkalmazhatjuk. A (3.88) képletet – a jelen fejezetben használt jelölésekkel – az alábbi alakba írhatjuk:

$$\mathbf{D}(\hat{\mathbf{p}}) = (\mathbf{A}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{A})^{-1} \quad (5.81)$$

Ahol az $\mathbf{A}$ mátrix a (5.27) egyenletben szereplő alak (struktúra) mátrix, $\mathbf{R}$ pedig a mérés kovariancia-mátrixa, a mérési hibákból képezhetünk. Felhasználva az lineáris additív modellt, – amelyben a kovarianciamátrixot az **egységnyi súlyú mérés szórásának** és a **súlymátrixnak** a szorzataként írjuk fel – a paraméterek kovariancia mátrixa tehát:

$$\mathbf{D}(\hat{\mathbf{p}}) = \frac{(\mathbf{u} - \mathbf{A} \cdot \mathbf{p})^T \cdot \mathbf{W} \cdot (\mathbf{u} - \mathbf{A} \cdot \mathbf{p})}{K - S} \cdot (\mathbf{A}^T \cdot \mathbf{W} \cdot \mathbf{A})^{-1} \quad (5.82)$$

A nem lineáris elméleti terű modellekre is felhasználhatjuk a fenti eredményt. Ehhez használjuk fel a direkt feladat Taylor-sorfejtését a paraméterek becsült ($\hat{\mathbf{p}}$) értékénél (az elsőrendű tagig végezve a sorfejtést):

$$\mathbf{f}(\mathbf{p}) = \mathbf{f}(\hat{\mathbf{p}}) + \left. \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \right|_{\mathbf{p}=\hat{\mathbf{p}}} \cdot (\mathbf{p} - \hat{\mathbf{p}}) + \dots \quad (5.83)$$

Mivel az $\mathbf{f}$ függvény egy K elemű vektor (K mérésünk van) és ezt az S darab paraméter szerint deriváljuk, a $\left. \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \right|_{\mathbf{p}=\hat{\mathbf{p}}}$ mennyiség egy $K \times S$ méretű mátrix. Ez a mátrix a (5.81) egyenletben behelyettesíthető az $\mathbf{A}$ mátrix helyére (ami ugyancsak a direkt feladatnak a paraméterek szerinti deriváltja), aminek a segítségével tetszőleges alakú direkt feladat esetén meghatározható a **paraméterek kovariancia mátrixa**:

$$\mathbf{D}(\hat{\mathbf{p}}) = \left[\left(\left. \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \right|_{\mathbf{p}=\hat{\mathbf{p}}} \right)^T \cdot \mathbf{R}^{-1} \cdot \left(\left. \frac{\partial \mathbf{f}}{\partial \mathbf{p}} \right|_{\mathbf{p}=\hat{\mathbf{p}}} \right) \right]^{-1} \quad (5.84)$$

Ez az eredmény központi jelentőségű az inverzióelméletben.

A paraméterek kovariancia mátrixát számos esetben csak közelítőleg tudjuk kiszámolni. A gyakorlatban előforduló esetek közül nagyon sokban, a mérések kovariancia mátrixát ($\mathbf{R}$) csak közelítőleg ismerjük. Sok esetben csak az egyes mérések szórásnégyzeteit ismerjük, a mérések közötti kovarianciákról nincsenek előzetes ismereteink. Ekkor az $\mathbf{R}$ mátrixot egy olyan diagonális (csak főátlóban elemeket tartalmazó) mátrixszal közelítjük, aminek a főátlójában az egyes mérések szórásnégyzetei szerepelnek. Más esetekben az egyes mérések között a mérési pontok fizikai távolsága, vagy a mérések időpontjai között eltelt idő alapján csökkenő korrelációt tételezünk fel. Ebben az esetben az $\mathbf{R}$ mátrix a főátlótól távolodva egyre kisebb abszolútértékű elemeket tartalmaz.

Az egyenletben szereplő deriváltak meghatározása bizonyos feladatok esetén analitikusan is megadható, de összetett direkt feladat esetén numerikus deriválással is ki lehet számolni a deriváltakat. Amennyiben a direkt feladatban szereplő függvény kellően sima, akkor az egyoldali differenciával közelítjük a deriváltat. Bonyolultabb esetben a kétoldali (három pontos) numerikus deriválást kell alkalmazni.

Amennyiben előzetes ismereteink alapján feltételezhetjük, hogy a hibakomponens nem normális eloszlású, és az adott problémára fel tudjuk írni a likelihood függvényt, akkor a likelihood függvény segítségével **általánosságban**

fel tudjuk írni a **paramétereink kovariancia mátrixát**, $\mathbf{D}(\hat{\mathbf{p}})$ -t amelynek elemeit az alábbi egyenlőség definiálja:

$$D_{ij}(\hat{\mathbf{p}}) = \frac{1}{-M \left(\left. \frac{\partial^2 L(\mathbf{u}, \mathbf{p})}{\partial a_i \partial a_j} \right|_{\mathbf{p}=\hat{\mathbf{p}}} \right)} \quad (5.85)$$

ahol az i és j indexek $1, 2, \dots, S$ értékeket vesznek fel (S a modellparaméterek száma). Az $L(\mathbf{u}, \hat{\mathbf{p}})$ függvény a problémához felírt likelihood függvény, aminek a paraméterek szerinti második deriváltjait a paraméterek inverzióval becsült értékénél ($\mathbf{p} = \hat{\mathbf{p}}$) számítjuk. Az $M(\dots)$ függvény itt is a várható érték képzést jelenti.

6. fejezet - Lineáris inverzió

A lineáris inverzió elméletében hasonló alapfogalmakat használunk fel, mint az inverzió statisztikai elméletében.

Jelölje az $\mathbf{m} = (m_1, m_2, \dots, m_M)$ a **modellparamétereket** tartalmazó vektort, amely a modellter egy eleme és $\mathbf{d} = (d_1, d_2, \dots, d_N)$ a mérési adatokat tartalmazó vektort, ami az adattér egy eleme. A lineáris inverzió elméletében abból indulunk ki, hogy az adatok és a modell között lineáris függvénykapcsolat van, amit egy $N \times M$ méretű $\mathbf{A}$ mátrix fejez ki.

Az $\mathbf{A}$ mátrix minden olyan információt tartalmaz, amit a modellparaméterek meghatározása során fel akarunk használni. Az adatokat zaj terheli. A zaj egy $\mathbf{e}$ vektorba foglalható össze, amelynek elemszáma megegyezik a mérések számával, vagyis minden méréshez tartozik a zaj-vektor egy eleme. Összefoglalva:

$$\mathbf{d} = \mathbf{A} \cdot \mathbf{m}_{true} + \mathbf{e} \quad (6.1)$$

A modell felépítésének megválasztása (vagyis hogy milyen alakú egyenleteket, hány paramétert használunk) egy szubjektív döntés, bizonyos ésszerűségi határokon belül. Ugyanazon mérési anyagra többféle modellt is illeszthetünk, ez többféle lineáris direkt feladatot, így többféle $\mathbf{A}$ mátrixot eredményezhet.

Jelölje az inverzió során meghatározott modellparamétereket $\tilde{\mathbf{m}}$. (A hullámvonal itt a „becslés, becült érték” jelentéssel bír.) Ekkor a fenti egyenletrendszer általános megoldását az alábbi formában írhatjuk fel:

$$\tilde{\mathbf{m}} = \mathbf{A}^{-\varepsilon} \cdot \mathbf{d} \quad (6.2)$$

Ahol $\mathbf{A}^{-\varepsilon}$ az $\mathbf{A}$ mátrix **általánosított inverze** (vagy pszeudoinverze). (Ennek meghatározását a „Lineáris algebra összefoglaló” c. fejezetben találhatjuk.) Az előző két egyenletből következik:

$$\tilde{\mathbf{m}} = \mathbf{A}^{-\varepsilon} \cdot \mathbf{A} \cdot \mathbf{m}_{true} + \mathbf{A}^{-\varepsilon} \cdot \mathbf{e} \quad (6.3)$$

A fenti egyenletben az $\mathbf{R} \equiv \mathbf{A}^{-\varepsilon} \cdot \mathbf{A}$ mátrixot **felbontás mátrixnak** nevezzük. Ideális esetben $\tilde{\mathbf{m}} = \mathbf{m}_{true}$. A fenti egyenletben az $\mathbf{R} \cdot \mathbf{m}_{true}$ tag azt fejezi ki, hogy a becült modellparaméterek a valódi modellparaméterek lineáris kombinációjaként adódnak, azaz az ideális mennyiségeknek csak valamilyen átlagát, elkent értékeit kapjuk vissza. Ideális esetben $\mathbf{R} = \mathbf{I}$ lenne ($\mathbf{I}$ az egységmátrix). Ekkor minden paramétert egymástól függetlenül meg tudnánk becsülni, vagyis a felbontás tökéletes lenne. Néhány elemi példa kivételével erre kevés esélyünk van. A később bemutatásra kerülő elemi példa, az egyenes illesztése azonban ilyen, ott a felbontás mátrix megegyezik az egységmátrixszal.

Az egyenlet másik tagja, az $\mathbf{A}^{-\varepsilon} \cdot \mathbf{e}$ tag azt fejezi ki, hogy az inverzió miként képezi le a zajt a modell térbe.

Az inverzió statisztikus elméletében láttuk, hogy a becült modellparaméterek kovarianciáit is meg kell határozzuk.

Ha a mérések $\mathbf{d}$ vektorának a d_j elemét σ_{d_j} szórású korrelálatlan zaj terheli, akkor az $\tilde{m}_i$ modellparaméter hibaterjedésből eredő hibája a lineáris inverzió elmélete szerint:

$$\sigma_{\tilde{m}_i}^2 = \sum_{j=1}^N \left(A_{ij}^{-\varepsilon} \cdot \sigma_{d_j} \right)^2 \quad (6.4)$$

A gyakorlatban nem lehet olyan $\mathbf{A}^{-\varepsilon}$ általános inverz mátrixot generálni, ami tökéletes felbontást és zérus hibaterjedést eredményez.

Lineáris inverzió túlhatározott probléma esetén

Túlhatározott egy probléma, ha $N > M$, vagyis a mérések N száma nagyobb, mint a problémában szereplő paraméterek M száma. Az adathibák miatt itt nem feltétlenül sikerül kielégíteni az $\mathbf{d} = \mathbf{A} \cdot \mathbf{m}$ egyenletet.

Példaként tekintsük, hogy megmérjük két tárgy tömegét külön-külön (1. és 2. mérés), és megmérjük együttesen (3. mérés) is. Ebből a három mérésből akarjuk megbecsülni a két test tömegét. A méréseink eredményei:

$d_1 = 1$ (kg), $d_2 = 2$ (kg), $d_3 = 2$ (kg). A két test tömegét jelölje m_1 és m_2 .

Látható, hogy ellentmondásosak a méréseink és a modellünk, nem tudunk két olyan tömeget meghatározni, hogy mind a három egyenlet igaz legyen.

Ekkor a (6.1) képlet alapján, formalizáljuk a méréseinket. A méréseket a paraméterekkel összekapcsoló $\mathbf{A}$ mátrix alakja:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \quad (6.5)$$

A zaj (hiba-) vektor komponenseit fejezzük ki a (6.1) egyenlet segítségével!

$$\mathbf{e} = \mathbf{d} - \mathbf{A} \cdot \mathbf{m} \quad (6.6)$$

Ahogy láttuk, esetünkben nem tudunk olyan $\mathbf{m}$ paraméterértékeket meghatározni, hogy az $\mathbf{e}$ vektor elemei nullák legyenek. Ezért írjuk elő azt, hogy az $\mathbf{e}$ vektor elemei a „lehető legkisebbek” legyenek. A lehető legkisebb fogalmat definiáljuk úgy, hogy az L_2 normájuk legyen minimális, vagyis az $\mathbf{e}$ vektor elemeinek négyzeteiből képzett összeg legyen minimális. Ez a vektorok skalárszorzásának definícióját felhasználva:

$$\mathbf{e}^T \cdot \mathbf{e} := \min \quad (6.7)$$

A fenti feltétel átírható

$$(\mathbf{d} - \mathbf{A} \cdot \mathbf{m})^T \cdot (\mathbf{d} - \mathbf{A} \cdot \mathbf{m}) := \min \quad (6.8)$$

alakban. Ennek az egyenletnek a szélsőértékénél, az egyenletnek az $\mathbf{m}$ paramétervektor szerinti deriváltja nulla. A deriválást elvégezve, az egyenletrendszer átrendezve kapjuk, az $\tilde{\mathbf{m}}$ becslését:

$$\tilde{\mathbf{m}} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{d} \quad (6.9)$$

Látható, hogy itt az $\mathbf{A}^T \cdot \mathbf{A}$ mátrix inverze szerepel. Ez az alak megegyezik a már levezetett (5.29) egyenlettel, ami a legkisebb négyzetes becslés. A mátrixszorzások kifejtésével megadható a két test becsült tömege:

$$\tilde{\mathbf{m}} = \left(\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{5}{3} \end{bmatrix} \quad (6.10)$$

Vagyis a két test becsült tömege $\frac{2}{3}$ (kg) és $\frac{5}{3}$ (kg).

Határozzuk meg a felbontás mátrixot! Ez a numerikus pontosság értékén belül:

$$\mathbf{R} = \mathbf{A}^{-\text{T}} \cdot \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (6.11)$$

Vagyis mindkét paraméter jól meghatározható.

Lineáris inverzió alulhatározott probléma esetén

Alulhatározott egy probléma, ha a méréseink száma kevesebb, mint az ismeretlen paraméterek száma, vagyis $N < M$. A korábbi példánál maradva, ez azt jelenti, mintha csak a 3. mérést végeztük volna el, vagyis a két test tömegét együtt mértük volna meg. Ekkor egyetlen mért értékünk van, és két meghatározandó paraméterünk. Mellékfeltételként előírhatjuk még, hogy a **becsült paraméterek L_2 normája** a legkisebb legyen. Ekkor (a levezetés mellőzésével) kapjuk, hogy:

$$\tilde{\mathbf{m}} = \mathbf{A}^{\text{T}} \cdot (\mathbf{A} \cdot \mathbf{A}^{\text{T}})^{-1} \cdot \mathbf{d} \quad (6.12)$$

Az $\mathbf{A}$ mátrix jelen esetben:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad (6.13)$$

Ebből a paraméterek becsült értékei:

$$\tilde{\mathbf{m}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \left(\begin{bmatrix} 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)^{-1} \cdot 2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \left[\frac{1}{2} \right] \cdot 2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (6.14)$$

Vagyis a két test tömege megegyezik, értékük 1–1 (kg). Ezt az eljárást **minimum norma** becslésnek nevezzük.

Ugyanezt az eredményt kaptuk volna, ha a (6.2) egyenletnek megfelelően, az általánosított inverzzel szoroztuk volna meg az adatvektort.

A felbontás mátrix a (6.3) képletnek megfelelően:

$$\mathbf{R} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \quad (6.15)$$

Vagyis a felbontás mátrix alapján látszik, hogy a diagonális elemek nem 1 körüli értékek, vagyis a paramétereket nem lehet biztonságosan meghatározni.

Lineáris inverzió egyidejűleg túl- és alulhatározott problémák esetén

(Az egyidejűleg túl- és alulhatározott problémákat **kevert határozottságú problémának** is nevezzük.)

A legkisebb négyzetes becslés során feltételeztük, hogy elegendő információval rendelkezünk ahhoz, hogy az ellentmondó egyenletek ellenére az összes modellparamétert meghatározzuk. Ez a tisztán túlhatározott probléma esete, ahol $\mathbf{A}^{\text{T}} \cdot \mathbf{A}$ egy invertálható mátrix.

A minimum norma becslés során pedig azt tételeztük fel, hogy az egyenletek között semmi ellentmondás sincs, azonban nincs elegendő információnk ahhoz, hogy az összes ismeretlen modellparamétert meghatározzuk. Ez a tisztán alulhatározott probléma esete, ahol $\mathbf{A} \cdot \mathbf{A}^T$ invertálható.

A gyakorlatban a második eset nem szokott előfordulni. Általában mindig vannak olyan modellparaméterek, amelyekre vonatkozólag ellentmondó információink vannak, és egyidejűleg vannak olyan modellparaméterek is, amelyekről semmilyen információt sem adnak a mérési adatok. Emiatt sem $\mathbf{A}^T \cdot \mathbf{A}$, sem $\mathbf{A} \cdot \mathbf{A}^T$ nem invertálható. De ha ezek az inverzek formálisan léteznek is, a probléma **rosszul meghatározott** (ill-conditioned), azaz az adatok kis megváltozása nagy változást okoz a modellparaméterekben. Ez azt jelenti, hogy a megoldás nagyon érzékeny a mérési hibákra. Ennek elkerülésére vezette be Levenberg 1944-ben a **csillapított (regularizált) legkisebb négyzetes** megoldást.

A problémát az okozza, hogy az $\mathbf{A}^T \cdot \mathbf{A}$ mátrix zérus, vagy közel zérus sajátértékekkel rendelkezik. Emiatt a mátrix rangja kisebb mint a mátrix dimenziója, így a hagyományos inverz nem értelmezhető.

Jelöljük az $\mathbf{M}$ mátrix sajátértékeit λ_n -nel, sajátvektorait pedig $\mathbf{v}_n$ -nel:

$$\mathbf{M} \cdot \mathbf{v}_n = \lambda_n \cdot \mathbf{v}_n \quad (6.16)$$

Egy skálázott egységvektor hozzáadásával megnövelhetjük az eredeti mátrix sajátértékeit:

$$(\mathbf{M} + \gamma \cdot \mathbf{I}) \cdot \mathbf{v}_n = (\lambda_n + \gamma) \cdot \mathbf{v}_n \quad (6.17)$$

Ezt az eredményt felhasználhatjuk a lineáris inverziós probléma megoldásához:

$$\tilde{\mathbf{m}} = (\mathbf{A}^T \cdot \mathbf{A} + \gamma \mathbf{I})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{d} \quad (6.18)$$

A kvadratikus formula segítségével levezethető, hogy $\mathbf{A}^T \cdot \mathbf{A}$ mátrix sajátértékei nem negatívak. Ekkor a pozitív γ esetén a zárójeles kifejezésben szereplő mátrix sajátértékei mind különbözni fognak nullától. Ennek következtében a mátrix rangja megegyezik a mátrix dimenziójával, és így invertálhatóvá válik. Ezzel az eredeti célkitűzést teljesítettük.

A fenti eredményre jutunk, a

$$(\mathbf{d} - \mathbf{A} \cdot \mathbf{m})^T \cdot (\mathbf{d} - \mathbf{A} \cdot \mathbf{m}) + \gamma \cdot \mathbf{m}^T \cdot \mathbf{m} := \min \quad (6.19)$$

menyiség (kritériumfüggvény) minimalizálásakor. A függvényen látszik, hogy a bal oldali tag a tisztán túlhatározott problémák esetén fellépő legkisebb négyzetes illesztést jelenti, a jobb oldali tag viszont a tisztán alulhatározott probléma esetén alkalmazott minimum norma becslést a csillapítási tényezővel megszorozva. Eszerint ha $\gamma = 0$, akkor visszakapjuk a legkisebb négyzetes megoldást, ha $\gamma \gg 1$ esetben viszont a jobb oldali tag dominál, így a minimum norma megoldást kapjuk.

Ezt a fenti eredményt a gyakorlatban is alkalmazzuk a „Csillapított (regularizált) legkisebb négyzetes becslés” c. fejezetben.

7. fejezet - Az inverzió megoldása a legkisebb négyzetek módszer segítségével

Oldjuk meg az alábbi feladatot! Egy felfüggesztett rugóra különböző tömegeket akasztunk, és mérjük meg a rugó teljes hosszát a ráakasztott tömegek hatására. Ebből határozzuk meg a rugóállandót és a rugó nyugalmi hosszát! (a nehézségi gyorsulás értékét vegyük 9.81m/s^2 -nek.)

A rugón ébredő rugóerő ($F_{\text{rugó}}$) nagysága a rugóállandó (K) és a rugó megnyúlásának (Δl) szorzata. A rugó megnyúlása (Δl) a rugó hosszának (l) és a nyugalmi hosszának (l_0) a különbsége. A rugóerő tart egyensúlyt a rugóra akasztott tömeg (m) súlyerejével ($m \cdot g$). A két erő egyensúlya az alábbi formában írható:

$$m_i \cdot g = K \cdot (l_i - l_0) \quad (7.1)$$

Ahol az i alsó index az egyes mérések sorszámát jelenti. Az egyenletet átrendezhetjük az alábbi formába:

$$m_i = -\frac{K \cdot l_0}{g} + \frac{K}{g} \cdot l_i \quad (7.2)$$

A képletet megvizsgálva látszik, hogy az m_i tömegek az l_i hosszúságok lineáris függvényének tekinthetők.

A mérési adatokat egy koordináta-rendszerben ábrázolhatjuk, amelynek vízszintes tengelyére az l_i hosszúságokat, a függőleges tengelyre pedig az m_i tömegeket mérjük fel. Ideális esetben, a mért tömegek a rugó hosszának függvényei lennének, vagyis a koordináta-rendszerbe felvitt mérési pontok pontosan egy egyenesre esnének. Az egyenes meredeksége a fenti képlet alapján K/g , tengelymetszete pedig $-K \cdot l_0/g$ lenne. Az egyenes ezen két paraméterét ha leolvassuk a grafikonról, azokból kiszámolhatjuk K és l_0 értékét.

Tényleges mérési adatok esetén a mérési pontok várhatóan nem esnek egy egyenesbe. Ennek egyik oka a mérés pontatlansága. Ilyen például a leolvasott rugóhosszúság hibája, illetve a méréshez használt tömegek nem pontosan egyeznek meg a ráírt értékekkel. Ezeket a hibákat, ebben a helyzetben, összefoglalóan **mérési hibáknak** nevezzük.

A másik ok lehet például, hogy a rugó nem pontosan követi a lineáris megnyúlási törvényt, és nagy ráakasztott tömegek esetén a vártnál jobban megnyúlik. Az ilyen típusú hibákat **modellhibáknak** nevezzük.

Lehetséges, hogy a mérés vagy a jegyzőkönyv elkészítése során egy súlyértéket rosszul olvasunk le, vagy az adatok számítógépes rögzítésénél egy helyiértéket tévedésből rosszul írunk. Ekkor olyan pontot kapunk, ami nem illeszkedik a trendbe, az ilyen típusú hibát **durva hibának** nevezzük. Az ilyen típusú hibákat el kell távolítani az adatrendszerből, mivel a bemutatásra kerülő legkisebb négyzetek elvét használó módszer érzékeny a durva hibákra.

A mérési pontokra grafikusán is illeszthetünk egy egyenest. Ekkor egy olyan egyenest próbálunk rajzolni, amely a pontok „között” húzódik, illetve leginkább a pontokon megy át. A grafikus illesztésnél azt próbáljuk elérni, hogy nagyjából ugyanannyi pont legyen az egyenes alatt, mint fölötté. Az így berajzolt egyenes tengelymetszetéből és meredekségéből számítjuk ki a K és l_0 értékeit. Ezzel a feladatot visszavezettük egy – a mérési pontokra legjobban illeszkedő – egyenes tengelymetszetének és meredekségének meghatározására.

A konkrét mérésünktől elvonatkoztatva, egy olyan adatrendszert vizsgálhatunk, amelyben N darab adatpárunk van és amelynek elemeit x_i -vel és y_i -vel jelöljük, és ezekre az adatokra szeretnénk egyenest illeszteni.

Az egyenes egyenlete:

$$y = a_0 + a_1 \cdot x \quad (7.3)$$

Amelyben a_0 és a_1 az egyenes meghatározandó paraméterei a tengelymetszet és a meredekség.

A problémát Gauss 1808-ban úgy fogalmazta meg, hogy a legjobban illeszkedő egyenesnek azt tekinthetjük, amelyik egyenes esetén a mérési pontokban (az x koordinátákhoz) tartozó y koordináták a lehető legközelebb vannak a pontokra illeszkedő egyeneshez (illetve az egyenes azon pontjához, ami ugyanahhoz az x koordinátához tartozik). A feladat megfogalmazásában szereplő „lehető legközelebb” kitétel tulajdonképpen egy minimumfeltétel. Gauss – elméleti megfontolásokból kiindulva – és ahogy később látjuk majd, a számítási egyszerűség miatt is, ezen **eltérések négyzeteinek összegének a minimumát** írta elő, ezért a módszert a legkisebb négyzetek módszerének nevezzük.

Egyetlen mérési pontban az x_i -hez tartozó mért y_i mennyiség és ugyanahhoz az x_i ponthoz az egyenes egyenletéből számolt elméleti y érték különbsége:

$$v_i = y_i - (a_0 + a_1 \cdot x_i) \quad (7.4)$$

Ahol v_i mennyiségek a reziduálok. (Nevezzük még maradékhibának vagy később részletezett okokból javításoknak is.) A legkisebb négyzetek módszere szerint tehát az

$$\sum_{i=1}^N [y_i - (a_0 + a_1 \cdot x_i)]^2 := \min \quad (7.5)$$

feltétel kell teljesüljön, vagyis a fenti összeg minimális kell legyen az a_0 és a_1 paraméterek függvényében.

A legkisebb négyzetek módszerét elvnek tekintjük, az előző fejezetben láttuk a kapcsolatát a maximum likelihood becsléssel. A későbbiekben látni fogjuk ennek a módszernek a széleskörű alkalmazhatóságát és az alkalmazás korlátait. A legkisebb négyzetek módszere – „A geofizikai inverzió statisztikai megközelítése” c. fejezetben részletesen tárgyalt esetekben – azonos eredményt ad az inverzióelmélet valószínűségelméleti megalapozásából levezethető módszerekkel.

A fenti feltétel, vagyis hogy a négyzetösszeg minimális legyen az a_0 és a_1 paraméterek függvényében, azt jelenti, hogy az összegnek ezen két paraméter szerint deriváltja nulla legyen. (Szükséges továbbá, hogy a második deriváltja negatív legyen de ezt most nem vizsgáljuk.)

$$\begin{aligned} \frac{\partial}{\partial a_0} \sum_{i=1}^N [y_i - (a_0 + a_1 \cdot x_i)]^2 &= 0 \\ \frac{\partial}{\partial a_1} \sum_{i=1}^N [y_i - (a_0 + a_1 \cdot x_i)]^2 &= 0 \end{aligned} \quad (7.6, 7.7)$$

Az összeg deriváltja a deriváltak összegével egyezik meg. A deriválást elvégezve és az összegzés elé kiemelve a bennük szereplő tagokat kapjuk:

$$\begin{aligned} 2 \cdot \sum y_i - 2 \cdot a_0 \cdot \sum 1 - 2 \cdot a_1 \sum x_i &= 0 \\ 2 \cdot \sum x_i \cdot y_i - 2 \cdot a_0 \cdot \sum x_i - 2 \cdot a_1 \sum x_i^2 &= 0 \end{aligned} \quad (7.8, 7.9)$$

Az egyenleteket 2-vel osztva és átrendezve kapjuk:

$$\begin{aligned} a_0 \sum 1 + a_1 \sum x_i &= \sum y_i \\ a_0 \sum x_i + a_1 \sum x_i^2 &= \sum x_i \cdot y_i \end{aligned} \quad (7.10, 7.11)$$

A fenti két egyenlet mátrix alakban írva:

$$\begin{bmatrix} \sum 1 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i \cdot y_i \end{bmatrix} \quad (7.12)$$

A fenti egyenletet normálegyenletnek nevezzük. Ez egy lineáris egyenletrendszer, amelynek megoldása szolgáltatja az a_0 és a_1 elemekből álló „paraméter vektor”-t.

Az egyenletet általános alakja:

$$\mathbf{N} \cdot \mathbf{x} = \mathbf{n} \quad (7.13)$$

Ahol a vastag szedés nagybetű esetén mátrixot, kisbetű esetén vektort jelent.

$\mathbf{N}$ a normálmátrix, és itt $\mathbf{x}$ vektor jelöli a meghatározandó paraméterek vektorát. Az $\mathbf{N}$ mátrix mindig szimmetrikus. A normálegyenletet megoldását az

$$\mathbf{x} = \mathbf{N}^{-1} \cdot \mathbf{n} \quad (7.14)$$

egyenlet szolgáltatja. Az $\mathbf{N}^{-1}$ mátrix az $\mathbf{N}$ mátrix inverze. (Ha $\det(\mathbf{N}) = 0$, akkor az általánosított inverzek segítségével kapunk megoldást.)

8. fejezet - A legkisebb négyzetek módszerének általános felírása

8.1. A legkisebb négyzetek módszer – alapfogalmak

A legkisebb négyzetek módszerét a geodéziában **kiegyenlítő számításnak** nevezik. Ismerkedjünk meg a kiegyenlítő számításokban alkalmazott fogalmakkal.

A méréseinket itt is valószínűségi változók realizációjának tekintjük. Jelentsék a mérendő mennyiségeket a $\xi_1, \xi_2, \dots, \xi_N$ valószínűségi változók, a minta elemeit (vagyis a tényleges mért értékeket) a $L_1, L_2, \dots, L_N$ valószínűségi változók. A mérések feldolgozásának célja a valószínűségi változók jellemzőinek, a várható értékének és a kovarianciamátrixának a meghatározása.

Ha egyetlen mennyiségre végzünk méréseket (például egy távolságot akarunk meghatározni a mérést többször elvégezve) akkor a ξ_i változók várható értéke megegyezik az L_i mintaelemek várható értékével.

$$M(\xi_i) = M(L_i) = a_i \quad (8.1)$$

A várható értékek $\hat{a}_i$ becslése lesz a **kiegyenlített mérési eredmény**: U_i . A legkisebb négyzetek módszerének alapelve, hogy a kiegyenlített mérési eredmények és a mérési eredmények különbségeire, a

$$v_i = \hat{a}_i - L_i = U_i - L_i \quad (8.2)$$

javításokra, fennálljon a következő összefüggés:

$$\mathbf{v}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{v} := \min \quad (8.3)$$

ahol a $\mathbf{v}$ a v_i elemekből képzett vektor. (A „javítás” kifejezés értelmét, később tárgyaljuk.)

A $\mathbf{P}_{LL}$ mátrix a súlymátrix. A súlymátrix – ahogy a statisztikus elméletben láttuk – az alábbi kapcsolatban van a **mérések kovarianciamátrixával**:

$$\mathbf{C}_{LL} = \hat{\mathbf{c}} \cdot \mathbf{P}_{LL}^{-1} \quad (8.4)$$

ahol a $\hat{\mathbf{c}}$ mennyiség az **egységnyi súlyú mérés szórásnégyzete**, amit a **súlyegység középhibájának** négyzetének is neveznek.

Amennyiben **egyforma súlyú mérések**re akarjuk elvégezni a kiegyenlítést, akkor a $\mathbf{P}_{LL}$ mátrix egy **I** egységmátrixszal helyettesíthető, amit behelyettesítve az alábbi egyszerűbb alakot kapjuk,

$$\mathbf{v}^T \cdot \mathbf{v} := \min \quad (8.5)$$

vagyis a kiegyenlítés során a $\mathbf{v}$ vektor elemeinek négyzetösszegét kell minimalizálni. A négyzetösszeg megegyezik a vektor saját magával vett skaláris szorzatával, ami formálisan a sorvektorként felírt, illetve egy oszlopvektorként felírt vektornak a mátrixok és vektorok szorzási szabályai szerint előállított szorzata: egy skalár szám. (A $\mathbf{v}$ vektort oszlopvektornak tekintve a transzponálás segítségével állítottunk elő belőle sorvektort.)

A legkisebb négyzetek módszerének alkalmazásakor a különböző feladatok megoldásakor felhasználunk **nem mért mennyiségeket** is. Ezeket a továbbiakban **paramétereknek** nevezzük. (Ez a megfogalmazás a geofizika inverzió szempontjából triviális, azonban tanulságos példát látunk majd arra, hogy a kiegyenlítő számítások során, csak mért mennyiségek közötti összefüggés segítségével is végezhetünk kiegyenlítést a mért mennyiségekre, amelyben nem szerepelnek paraméterek.)

A paramétereket X_j -vel jelöljük. A j index $1, 2, \dots, P$ értékeket vehet fel, ahol P a paraméterek száma. A paramétereket vektor formában is felírhatjuk: $\mathbf{X}$. Általában a paramétereket is felhasználó kiegyenlítések során a paraméterek X_{j0} előzetes (kiinduló) értékéhez meghatározzák a **paraméterek javításait**, amelyek összege adja a **paramétervektor kiegyenlített értékét**.

$$\mathbf{X} = \mathbf{X}_0 + \mathbf{x} \quad (8.6)$$

A kiegyenlítő számítások elméletében a direkt feladatot **funkcionális modellnek**, ennek tényleges alakját **közvetítő egyenletnek** nevezik.

A legegyszerűbb alakú közvetítő egyenletben (a mért mennyiségeink „jóslását” jelentő) elméleti függvényértékek a paraméterek **lineáris függvényei**. Ekkor a direkt feladat megoldása az $\mathbf{X}$ vektornak egy mátrixszal való szorzását jelenti:

$$\mathbf{U} = \mathbf{A} \cdot \mathbf{X} \quad (8.7)$$

Vizsgáljuk meg először az egyenletünk jobb oldalát. Az $\mathbf{A}$ mátrix a már korábban megismert **alakmátrix**. Az $\mathbf{X}$ vektor a **paramétereink kiegyenlített értéke**.

Az egyenlőség bal oldalán álló $\mathbf{U}$ vektor nem más, mint a **méréseink kiegyenlített értéke**. Ez az egyenlet azt fejezi ki, hogyha nem lenne semmilyen hiba (sem mérési, sem pedig modellhiba) akkor a méréseink a paramétereinkkel szigorú (algebrai) függvénykapcsolatban lennének. Rendezzük át az előbbi egyenletet, úgy, hogy a mért értékek és a javítások segítségével fejezzük ki kiegyenlített mérési eredményeket, a paraméterek előzetes értékének és a paraméterek javításának segítségével fejezzük ki paraméterek kiegyenlített értékeit:

$$\mathbf{L} + \mathbf{v} = \mathbf{A} \cdot (\mathbf{X}_0 + \mathbf{x}) = \mathbf{A} \cdot \mathbf{X}_0 + \mathbf{A} \cdot \mathbf{x} = \mathbf{a}_0 + \mathbf{A} \cdot \mathbf{x} \quad (8.8)$$

Az $\mathbf{a}_0$ vektor a **direkt feladat megoldása a paraméterek kiinduló értéke mellett**.

Fejezzük ki a mérések **javitását** (a $\mathbf{v}$ vektort) a fenti egyenletből.

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{x} - (\mathbf{L} - \mathbf{a}_0) = \mathbf{A} \cdot \mathbf{x} - \mathbf{l} \quad (8.9)$$

Ezt az egyenletet **javitási egyenletnek** nevezzük. Az elnevezés magyarázat az, hogyha a $\mathbf{v}$ vektor elemeivel „megjavítanánk” az $\mathbf{L}$ mérési adatokat (vagyis $\mathbf{v}$ -t hozzáadnánk $\mathbf{L}$ -hez), akkor a megjavított értékekre a modell (a megjavított paraméterekkel számolva) tökéletesen illeszkedne. A $\mathbf{v}$ vektor elemei a **reziduálok**, a vektor mérete N , vagyis megegyezik a mérések számával.

A fenti képletben szereplő $\mathbf{l}$ vektor a tisztatag vektor. Látszik, hogy elemei a mérések ($\mathbf{L}$) és a direkt feladatnak a paraméterek kezdő értéke melletti megoldásának ($\mathbf{a}_0$) a különbsége.

A (8.3) képlettel megadott minimumfeltételbe már be tudjuk írni a javításvektor (8.9) képlettel megadott alakját. Az adott kifejezésnek ott lesz minimuma (szélsőértéke), ahol az $\mathbf{x}$ vektor szerinti deriváltja nulla.

A minimalizálandó függvény felírva a javítások szorzataként:

$$(\mathbf{A}\mathbf{x} - \mathbf{l})^T \cdot \mathbf{P}_{LL} \cdot (\mathbf{A}\mathbf{x} - \mathbf{l}) = \mathbf{A}\mathbf{x}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A}\mathbf{x} - \mathbf{l}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A} \cdot \mathbf{x} - \mathbf{A} \cdot \mathbf{P}_{LL} \cdot \mathbf{x}^T \cdot \mathbf{l} - \mathbf{l}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{l} \quad (8.10)$$

Ennek a mennyiségnek az $\mathbf{x}$ szerinti parciális deriváltja (amely csak az $\mathbf{x}$ -től függő tagokat tartalmazza) nulla, az elégséges feltétel a szélsőértékhez. A deriválást elvégezve, figyelembe véve, hogy a deriváltak összege nulla, és az egyenletet átrendezve kapjuk:

$$\begin{aligned} \mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A} \cdot \mathbf{x} &= \mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{l} \Rightarrow \\ \mathbf{N} \cdot \mathbf{x} &= \mathbf{n} \end{aligned} \quad (8.11 - 8.12)$$

A fenti formulával tehát ismét megkaptuk a normál egyenletet, amiből látszik, hogy az $\mathbf{N}$ normál mátrix megegyezik (egyenlő súlyú mérések esetén) az $\mathbf{A}^T \cdot \mathbf{A}$ mennyiséggel, ami az alakmátrix saját transzponáltjával vett szorzata, valamint az $\mathbf{n}$ vektor előállítható az alakmátrix transzponáltjának a tisztatag vektorral vett szorzataként ($\mathbf{A}^T \cdot \mathbf{l}$). A **paraméterek javítása** kifejezve a normálegyenlet (8.12) segítségével:

$$\mathbf{x} = \mathbf{N}^{-1} \cdot \mathbf{n} = (\mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{l}) \quad (8.13)$$

8.2. Egyenes illesztése

Vizsgáljuk meg ismét az előző fejezetben tárgyalt egyenes illesztést, de használjuk fel az előző alfejezetben bevezetett fogalmakat!

Az egyenes illesztése során x - y pontpárookra határozzuk meg a legjobban illeszkedő egyenes tengelymetszetét és meredekségét. Ugyanazon x - y pontpárookra megpróbálkozhatnánk akár magasabb fokszámú polinom illesztésével is. Az hogy egyenest illesztünk a mi döntésünk. (Döntésünket természetesen befolyásolják korábbi tapasztalataink és ismereteink.) Az hogy egyenest illesztünk ez a **modellválasztás**. Az egyenes lesz tehát a modell, az egyenes egyenlete a **direkt feladat**, és az egyenes egyenletében szereplő tengelymetszet és meredekség a modell paraméterei. A levezetés során az x értékeket **mérési helyeknek**, az y értékeket **mért értékeknek** tekintjük.

Definiáljuk most is a maradék hibát a (8.9) képlethez hasonlóan! Ekkor a maradék hibák (reziduálok):

$$\begin{aligned} v_1 &= a_0 + a_1 \cdot x_1 - y_1 \\ v_2 &= a_0 + a_1 \cdot x_2 - y_2 \\ &\vdots \\ v_i &= a_0 + a_1 \cdot x_i - y_i \\ &\vdots \\ v_N &= a_0 + a_1 \cdot x_N - y_N \end{aligned} \quad (8.14)$$

Ezek a fenti egyenlőségek mátrix-vektor formában is felírhatók:

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{x} - \mathbf{l} \quad (8.15)$$

Ez az előző fejezetben megismert javítási egyenlet. Az $\mathbf{x}$ vektor elemei a meghatározandó paraméterek javításai, esetünkben a_0 és a_1 . (Sajnos a továbbiakban $\mathbf{x}$ nem koordinátákat, hanem a **paraméterek vektorát** jelenti.) A jelen esetben azzal a feltételezéssel élünk, hogy a paraméterek előzetes (kiinduló) értékei ($\mathbf{X}_0$ vektor elemei) nullák. Ekkor a paraméterek javítását tartalmazó $\mathbf{x}$ vektort a kiegyenlítővel meghatározva, egyben a paraméterek kiegyenlített értékeinek $\mathbf{X}$ vektorát kapjuk (8.13) egyenlet!

Az $\mathbf{l}$ vektort a tisztatag. A tisztatag vektort úgy definiáltuk, hogy az a mért értékekből kivonva a kezdő paraméterekkel számított direkt feladat megoldást. Jelen esetben is ellenőrizhetjük a definíció helyességét az adott feladatra. A mért értékek az y értékek. A modell paraméterek előzetes értékeit nullának tekintettük ($a_0 = 0$ és $a_1 = 0$), így azokból az x mérési pontok sorozatára kiszámolva a direkt feladatot, nulla értékeket kapunk. Ezeket a nulla értékeket kivonva az y mérési értékekből visszakapjuk az y értékeket, tehát az $\mathbf{l}$ vektor az y értékekből áll.)

Vizsgáljuk meg az $\mathbf{A}$ alakmátrix alakját!

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \quad (8.16)$$

Vegyük észre, hogy a mátrix egy-egy sorában a direkt feladatnak – az egyenes (7.3) képlettel adott egyenletének – a paraméterek javításai (estünkben a_0 és a_1) szerinti deriváltjai szerepelnek!

Általánosságban elmondhatjuk, hogy az alakmátrix a direkt feladatnak az egyes mérésekre vonatkozóan (a mátrix sorai) az egyes paraméterek szerinti (a mátrix oszlopai) deriváltjait tartalmazza. Mivel ez a mátrix, egy vektort visszaadó többváltozós függvénynek a paraméterek szerinti deriváltjaiból áll elő, a matematikai analízisből ismert néven, **Jacobi-mátrix**nak is nevezzük.

Szükségünk van még a mérések $\mathbf{P}_{LL}$ súlymátrixára. A méréseinket egyforma súlyúnak tételezzük fel, így a súlymátrix egy egységmátrix, ekkor a paraméterek javításának $\mathbf{x}$ vektorára az alábbi egyenletet írhatjuk fel:

$$(\mathbf{A}^T \cdot \mathbf{A}) \cdot \mathbf{x} = \mathbf{A}^T \cdot \mathbf{I} \quad (8.17)$$

amelyből az $\mathbf{x}$ vektort az alábbi egyenlet megoldásával kapjuk:

$$\mathbf{x} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{I} \quad (8.18)$$

8.3. A legkisebb négyzetek módszerében szereplő mennyiségek kovarianciamátrixai

A legkisebb négyzetek módszer nem csak a **kiegyenlítésben szereplő mennyiségek várható értékeinek**, hanem ezeknek a mennyiségeknek a **kovarianciamátrixának** a meghatározására is felhasználhatók.

A kiegyenlítésben szereplő mennyiségek: $\mathbf{X}$ a paraméterek kiegyenlített mennyiségei, $\mathbf{U}$, a mérések kiegyenlített mennyiségei, $\mathbf{v}$ a javítások, és $\mathbf{L}$ a mérések. Ezeknek a mennyiségeknek a kovariancia mátrixát általánosságban a

$$\mathbf{M}_{YY} = c^2 \cdot \mathbf{Q}_{YY}, \text{ ahol } Y = X, U, v, L \quad (8.19)$$

képlettel adhatjuk meg. A $\mathbf{Q}_{YY}$ mátrixot általánosságban **súlykoefficiens mátrix**nak nevezzük.

A c^2 mennyiség becslésre az **egység súlyú mérés szórásnégyzetét** használjuk:

$$\hat{c}^2 = \frac{\mathbf{v}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{v}}{N - R} \quad (8.20)$$

a képletben N a mérések száma, R pedig a direkt feladatban szereplő paraméterek száma. (A kettő különbsége a rendszer szabadsági fokainak száma, amit **fölölmérésnek** neveznek.)

A kiegyenlített **paraméterek (X) súlykoefficiens mátrixa** (a levezetés mellőzésével):

$$\mathbf{Q}_{XX} = (\mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A})^{-1} \quad (8.21)$$

Ennek a mennyiségnek a $\hat{\epsilon}^2$ -tel vett szorzata adja a **paraméterek kovariancia mátrixát**.

A kiegyenlített mérési eredmények (U) súlykoefficiens mátrixa (ugyancsak mellőzve a részleteket):

$$\mathbf{Q}_{UU} = \mathbf{A} \cdot \mathbf{Q}_{XX} \cdot \mathbf{A}^T \quad (8.22)$$

ahol felhasználtuk az előzőleg kiszámított $\mathbf{Q}_{XX}$ mátrixot.

A mérési javítások (v) súlykoefficiens mátrixának általános alakja az alábbi:

$$\mathbf{Q}_w = \mathbf{P}_{LL}^{-1} - \mathbf{A} \cdot (\mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \quad (8.23)$$

Ennek kiszámításához felhasználhatjuk, hogy a **mérések súlykoefficiens mátrixa** ($\mathbf{Q}_{LL}$) a súlymátrix inverze:

$$\mathbf{Q}_{LL} = \mathbf{P}_{LL}^{-1} \quad (8.24)$$

valamint, hogy a javítások súlykoefficiens mátrixa a mérések és a kiegyenlített mérések súlykoefficiens mátrixából az alábbi módon számolható:

$$\mathbf{Q}_w = \mathbf{Q}_{LL} - \mathbf{Q}_{LU} \quad (8.25)$$

A kiegyenlített paraméterek súlykoefficiens mátrixából számolható mátrixot, a **paraméterek kovariancia mátrixát** minden kiegyenlítési feladat során ki kell számolni, az eredmények megbízhatóságának számszerűsítése céljából.

8.4. Másodfokú függvény illesztése

Az előbb bemutatott egyenes illesztési problémával összehasonlítva beláthatjuk, hogy bonyolultabb függvény illesztése esetén több paraméterünk van.

Az előbbi x - y pontpárookra próbáljunk **másodfokú függvényt** illeszteni, írjuk fel az alakmátrixot, és segítségével oldjuk meg a normálegyenletet!

Ebben az esetben a direkt feladat (vagyis az alkalmazott modell) egyenlete a másodfokú függvény általános egyenlete, ami a következő:

$$y = a_0 + a_1 \cdot x + a_2 \cdot x^2 \quad (8.26)$$

A paraméterek vektora itt három elemű:

$$\mathbf{x} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \quad (8.27)$$

A (8.14) képlethez hasonlóan felírhatjuk a reziduálokat, mint a paraméterek függvényét. Az $\mathbf{l}$ tisztag vektor itt is az y mért értékekből áll, az $\mathbf{A}$ alakmátrixot itt is a direkt feladatnak a paraméterek szerinti deriváltjából állíthatjuk elő, amelynek alakja az alábbi:

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_i & x_i^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 \end{bmatrix} \quad (8.28)$$

A paraméterek vektorát itt is az

$$\mathbf{x} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{l}) \quad (8.29)$$

Egyenlet segítségével számolhatjuk ki.

8.5. Legkisebb négyzetes becslés eltérő súlyú mérések esetén

Eddigi példáinkban a méréseinket egyforma súlyúnak tekintettük. Előfordulhat azonban, hogy valamely méréseket nem gondolunk megbízhatónak, azokat **nagyobb zajjal** terheltnek tekintjük. Ezeket a méréseket **kisebb súllyal** szeretnénk szerepeltetni a minimumban, míg a megbízható méréseket nagyobb súllyal.

Tekintsük az alábbi feladatot! Két karó távolságát 4-szer megmértük egy mérőszalaggal. Az alábbi táblázat tartalmazza a mért adatokat.

10,13
9,86
10,04
10,21

8.1. táblázat. Ugyanarra a távolságra vonatkozó négy távolságmérés eredménye

Tekintsük ismeretlen meghatározandó paraméternek a karók távolságát (a_0), eszerint az ismeretlenek vektora egy 1 elemű vektor.

A modell szerint, a távolság a paraméter értéke. Eszerint a direktfeladatnak a modell paraméter szerinti deriváltja: 1. Ez minden mérésre igaz, eszerint az $\mathbf{A}$ alakmátrix egy 1 oszlopból és 4 sorból álló, csak 1-et tartalmazó vektor. Az $\mathbf{l}$ tisztag tartalmazza a mért adatok és a modell különbségét, ha a modellparaméter 0, vagyis az $\mathbf{l}$ tisztagvektor a mért távolságokból álló 4 elemű oszlopvektor.

$$\mathbf{A} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{x} = [a_0] \quad \mathbf{l} = \begin{bmatrix} 10,13 \\ 9,86 \\ 10,04 \\ 10,21 \end{bmatrix} \quad (8.30, 8.31, 8.32)$$

A (8.29) egyenlet segítségével határozzuk meg a paraméter értékét, ami 10,06 lesz. (A legkisebb négyzetek módszeréből levezetett megoldás – egyforma súlyú mérések esetén – megegyezik, a mérések átlagával.)

Módosítsuk az előbbi feladatot! Egy újabb mérőcsoport újabb 4 alkalommal megméri a két karó távolságát, de ők pontosabban mértek. A méréseik eredménye az alábbi táblázatban van:

10,02
9,97
10,01
10,00

8.2. táblázat. Ugyanarra a távolságra vonatkozó újabb négy távolságmérés eredménye

Látszik, hogy ezeknek a méréseknek kisebb a szórása. (A mérések átlaga: 10,00) Szeretnénk mind a 8 mérést felhasználni, a paraméter értékének meghatározásához, de figyelembe szeretnénk venni, hogy az első négy mérést „pontatlanabb” az utolsó négy „pontosabb”.

A legkisebb négyzetek elvének levezetésénél abból indultunk ki, hogy minimalizáljuk a rezidualok négyzeteinek összegét:

$$\sum_{i=1}^N (v_i)^2 = \min \quad (8.33)$$

Jelen esetben, az eltérés négyzeteket súlyozni kellene:

$$\sum_{i=1}^N p_i \cdot v_i^2 = \min \quad (8.34)$$

A p_i súlyokat a tapasztalat alapján általában az adott mérés szórásnégyzetének reciprokának választják:

$$p_i = \frac{1}{\sigma_i^2} \quad (8.35)$$

A fenti feladatban ez azt jelenti, hogy az első négy mérésre meghatározzuk a szórásnégyzetet és azt használjuk súlynak az első négy mérés mellé, majd az utolsó 4 mérésből meghatározott szórásnégyzet reciproka lesz az utolsó 4 mérés súlya. A súlyok számításához a **tapasztalati szórásnégyzet** (3.54) képletét használjuk fel.

Eszerint a súlyok:

$$p = \frac{1}{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2} \quad (8.36)$$

A képletben $\bar{y}$ az első négy mérés átlagát jelenti, N pedig a mérések száma, vagyis 4, amiből a szórás: $\sigma_I = 0,150$. Ugyanez a képlet használható a második 4 mérés szórássának kiszámításához, természetesen itt az utóbbi négy mérés és azok átlaga szerepel, aminek a segítségével a szórás: $\sigma_{II} = 0,021$.

A minimumfeltételt mátrix-vektor formában felírva kapjuk:

$$\mathbf{v}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{v} = \min \quad (8.37)$$

Ahol a $\mathbf{P}_{LL}$ mátrix a súlymátrix, amely a mérések kovarianciamátrixának inverze elosztva az egységnyi súlyú mérés szórásnégyzetével (8.4). Abban a speciális esetben, ha a mérések függetlenek, csak a főátlóban lesznek elemek, a szórásnégyzetek reciprokai:

Esetünkben a súlymátrix elemei: $p_I = 44,24$ és $p_{II} = 2142,85$.

$$\mathbf{P}_{LL} = \begin{bmatrix} \frac{1}{\sigma_I^2} & & & & & & & \\ & \frac{1}{\sigma_I^2} & & & & & & \\ & & \frac{1}{\sigma_I^2} & & & & & \\ & & & \frac{1}{\sigma_I^2} & & & & \\ & & & & \frac{1}{\sigma_{II}^2} & & & \\ & & & & & \frac{1}{\sigma_{II}^2} & & \\ & & & & & & \frac{1}{\sigma_{II}^2} & \\ & & & & & & & \frac{1}{\sigma_{II}^2} \end{bmatrix} \quad (8.38)$$

A súlymátrixot is tartalmazó minimumfeltétel kifejtéséhez felhasználjuk a javítási egyenlet (8.9) alakját. Ekkor:

$$(\mathbf{Ax} - \mathbf{l})^T \mathbf{P}_{LL} (\mathbf{Ax} - \mathbf{l}) := \min \quad (8.39)$$

Egyenletnek az $\mathbf{x}$ paramétervektor szerinti deriváltját képezzük, és ahol ez a derivált nulla annál az $\mathbf{x}$ vektor értékénél van a függvénynek szélsőértéke.

Az egyenlet megoldása, a (8.3)-tól (8.13)-ig tartó levezetés analógiájára:

$$\mathbf{x} = (\mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{l}) \quad (8.40)$$

Az adott feladatnál az egyenletben szereplő $\mathbf{A}$ alakmátrix most egy 1-esekből álló 8 elemű oszlopvektor, az $\mathbf{l}$ tisztag vektor tartalmazza a 8 mérést. A $\mathbf{P}_{LL}$ súlymátrix a (8.4) képlet segítségével határozható meg a mérések kovariancia mátrixából. A súlyozott átlag: 10,0012 sokkal közelebb esik a második 4 mérés átlagához.

8.6. A mérések szórásának becslése a mérési anyagból

Vizsgáljuk meg újra a fenti példát! Láttuk, hogy az első négy mérés átlagát képezve, megkaptuk a várható érték becslését, és ennek felhasználásával kaptuk az első négy mérés szórását. Ugyanezt a műveletet megismételve a második négy mérésre, megkaptuk a második négy mérés szórását. Ezeknek a szórásnégyzeteknek a reciprokjait alkalmaztuk a mérések $\mathbf{P}_{LL}$ súlymátrixában, amit a kiegyenlítéshez felhasználtunk.

A fenti esetben szerencsésen felhasználhattuk, hogy az egyes mérési csoportok szórását könnyen kiszámíthattuk, ehhez külön-külön elvégeztük az inverziót, ami jelen esetben az átlagolást jelentette. A gyakorlatban előforduló esetekben azonban nem tudjuk, vagy nem akarjuk szétválasztani az egyes mérési csoportjainkat, és azokra külön-külön elvégezni az inverziót.

Ebben az esetben, az inverzió elvégzése után, lehetőségünk van az egyes mérési csoportokhoz tartozó szórások becslésére, de csak az inverzió elvégzése után.

Ennek levezetéséhez írjuk fel ismét az egység súlyú mérés szórásának képletét:

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{v}}{r} \quad (8.41)$$

ahol r a fölősmérések száma. Az r mennyiség értelmezhető úgy, mint az egyes mérésekre jutó r_i **redundancia komponensek** összege, ahol r_i az egy mérésre jutó redundancia. Ez a mennyiség megkapható mint a $\mathbf{Q}_{rr} \cdot \mathbf{P}_{rr}$, $N \times N$ méretű mátrix főátlóban levő elemeiből képzett vektor. A redundancia komponenseket valamennyi mérésre összegezve kapjuk a fölősmérések számát:

$$r = \sum_{i=1}^N r_i \quad (8.42)$$

A méréseinkből képezzünk K darab csoportot úgy, hogy a k -adik csoportban azonos p_k súlyt tételezhessünk fel a csoportba tartozó mérésekhez. Ennek segítségével felírhatjuk az egységnyi súlyú mérés szórásnégyzetét:

$$\sigma_0^2 = \frac{\sum_{i=1}^{n_k} (v_i^2 \cdot p_{k,a-priori})}{\sum_{i=1}^{n_k} r_i} = \frac{p_{k,a-priori} \cdot \sum_{i=1}^{n_k} v_i^2}{\sum_{i=1}^{n_k} r_i} \quad (8.43)$$

ahol n_k a mérések száma a k -adik csoportban. Ha a $p_{k,a-priori}$ súlyokat jól választjuk meg, akkor a fenti képlettel valóban az egységnyi súlyú mérés szórását kapjuk meg, bármelyik k csoportra számolva.

Az eljárás alkalmazása során az iteráció első lépésében egyforma súlyokat feltételezve elvégezzük a kiegyenlítést.

Ekkor a (8.41) képlet alkalmazásával becslést kapunk $\hat{\sigma}_0^2$ -re. Ennek és a (8.43) egyenlet átrendezése segítségével kifejezzük az *a posteriori* súlyokat:

$$p_{k,a-posteriori} = \sigma_0^2 \cdot \frac{\sum_{i=1}^{n_k} r_i}{\sum_{i=1}^{n_k} v_i^2} \quad (8.44)$$

Ezután a $p_{k,a-posteriori}$ súlyokat $p_{k,a-priori}$ súlyoknak tekintjük a kiegyenlítés újabb elvégzése során. Néhány iterációs lépés után azt kapjuk, hogy sem a súlyok, sem a becsült mennyiségek nem változnak.

A fenti módszer gyakorlati alkalmazására tekintsük a korábban bemutatott 4-4 hosszúságmérés esetét. A módszer alkalmazása során első lépésben a két mérési csoportot azonos súlyúnak tekintjük.

Mérések		javítások $p_1=1$ és $p_2=1$		javítások $p_1=0,374$ $p_2=16,48$	
l_1	l_2	v_1	v_2	v_1	v_2
10,13	10,02	-0,10	0,01	-0,128	-0,018
9,86	9,97	0,17	0,06	0,141	0,031
10,04	10,01	-0,01	0,02	-0,038	-0,008
10,21	10,00	-0,18	0,03	-0,208	0,001
Σ	40,24	40,00	-0,12	0,12	-0,234
					0,005

8.3. táblázat. A két mérési sorozatban mért hosszúságok találhatók a baloldali oszlopokban. A középső oszlopokban az egyforma súllyal végzett kiegyenlítések esetén a maradékhibák, a jobboldali oszlopban a mérési adatokból becsült mérési hibákkal végzett kiegyenlítés maradékhibái láthatók.

Az első iterációs lépéshez a méréseket azonos súlyúaknak tételezzük fel. Ekkor a $\mathbf{P}_{LL}$ mátrixot és annak inverzét (a $\mathbf{Q}_{LL}$ mátrixot) is egységmátrixnak ($\mathbf{I}$) vehetjük. A kiegyenlítés elvégzése után ki kell számolnunk a redundancia együtthatókat, amik a $(\mathbf{Q}_{vv} \cdot \mathbf{P}_{LL})$ szorzatmátrix főátlóban levő elemei. Ehhez felhasználjuk még a $\mathbf{Q}_{vv}$ mátrixot, amit a (8.23) egyenlet segítségével számolunk. Ehhez az

$$\mathbf{A} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{és} \quad \mathbf{Q}_{xx} = (\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A})^{-1} = \frac{1}{8} \quad (8.45, 8.46)$$

mátrixokat használjuk fel. Ennek segítségével:

$$\mathbf{Q}_{vv} = \mathbf{I} - \begin{bmatrix} \frac{1}{8} & & \\ & \ddots & \\ & & \frac{1}{8} \end{bmatrix} = \begin{bmatrix} 0,875 & & \\ & \ddots & \\ & & 0,875 \end{bmatrix} \quad (8.47)$$

A redundancia komponens elemei eszerint 0,875 értékűek. (Ezt elemi megfontolásból is megkaphattuk volna, ugyanis 8 mérésünk van, és 1 meghatározandó paraméterünk, amiből a redundancia komponens a szabadsági fokok és a mérések számának hányadosa vagyis $7/8 = 0,875$. Az újabb iterációs lépések során a redundancia komponensek változni fognak!)

A 8 darab mérésre az egyforma súlyokkal történő kiegyenlítésből megkaptuk, hogy a kiegyenlített távolság ($\mathbf{x}$) 10,03 lesz. Az egységnyi súlyú mérés szórása (8.41) alapján 0,104-nek adódik, szórásnégyzete ($\hat{\sigma}_0^2$) pedig 0,011.

Az egységnyi súlyú mérés szórásnégyzete és a redundancia komponensek segítségével a (8.44) az első iterációs lépés után kiszámolt súlyok:

$$p_{1,a-positív} = 0,011 \cdot \frac{4 \cdot 0,875}{(-0,1)^2 + (0,17)^2 + (-0,01)^2 + (-0,18)^2} = 0,535$$

$$p_{2,a-positív} = 0,011 \cdot \frac{4 \cdot 0,875}{(0,01)^2 + (0,06)^2 + (0,02)^2 + (0,03)^2} = 7,634 \quad (8.48, 8.49)$$

A **második iterációs lépésben** a súlyokat beírjuk a $\mathbf{P}_{LL}$ mátrixba, és ezzel az új mátrixszal elvégezve a kiegyenlítést, megkapjuk az új kiegyenlített távolságot ($\mathbf{x}$). A $\mathbf{P}_{LL}$ mátrix segítségével kiszámoljuk a $\mathbf{Q}_{vv}$ mátrixot, majd az r_i redundancia komponenseket. Kiszámoljuk az egységnyi súlyú mérés szórásnégyzetét ($\hat{\sigma}_0^2$), majd a súlyokat (p_1 és p_2).

A fenti példában a **harmadik iterációs lépés** elvégzésére is szükség van. Ennek során megismételjük a második iterációs lépést.

A bemutatott példa esetén a harmadik iterációs lépésben a kiegyenlített távolság ($\mathbf{x}$) 10,0013 lesz (ami gyakorlatilag megegyezik az ismert súlyokból számolt eredménnyel) a súlyok pedig:

$$p_1 = 0,374$$

$$p_2 = 16,48 \quad (8.50, 8.51)$$

A **további iterációs lépések** során már a súlyok nem változnak. A 8.3. táblázat utolsó két oszlopában láthatók a javítások, amik megegyeznek a maradékhibákkal.

Az inverzió minőségének meghatározásához meg kell határozni a kiegyenlített paraméterek kovariancia mátrixát. Jelen esetben egyetlen ismeretlenünk van: a_0 . A paraméterek kovariancia mátrixa az egységnyi súlyú mérés és a kiegyenlített paraméterek súlykoefficiens mátrixának szorzata:

$$\sigma_0^2 \cdot \mathbf{Q}_{xx} = \sigma_0^2 \cdot (\mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A})^{-1} = 0,000113 \quad (8.52)$$

amiből az $\hat{a}_0$ becslt paramétere szórása: 0,0106.

A fenti módszernek, vagyis a szórások *a posteriori* becslésének a két szélsőséges esetét vizsgáljuk meg!

Az első esetben a valamennyi mérésünknek azonos szórást tételezhetünk fel, ekkor a *p*-*a-priori* szórás 1-nek választható, és visszkapjuk az azonos súlyú mérésekre felírt egyenleteket, amik a fenti példában az $x = 10,03$ eredményre vezet.

A második esetben a minden méréshez külön szórást akarunk meghatározni, vagyis annyi mérési csoportunk van, ahány mérésünk. Ezt a fenti számítás módszer átalakításával meg tudjuk valósítani.

9. fejezet - A nem lineáris direkt feladat esete

Az előző fejezetben tárgyalt példákban (egyenes és másodfokú függvény illesztése) a direkt feladatok **a paraméterek lineáris függvényei** voltak. Ez számos esetben teljesül, például valamely mérési pontokra történő polinomfüggvények illesztésekor. Az általános geofizika területén számos olyan eset van, amikor egy függvényt sorfejtés alakjában írjuk fel, ekkor a függvény a sorfejtés együtthatóinak lineáris függvénye, ezért a sorfejtés együtthatóinak a mérési adatokból történő meghatározása az előző fejezetben tárgyalt módszer segítségével elvégezhető.

Az ilyen típusú direkt feladat nagy előnyt jelentett az **A** alakmátrix számításához, mivel az alakmátrix a direkt feladatnak a paraméterek szerinti deriválásával jön létre, ezért ha a direkt feladat a paraméterek lineáris függvénye, akkor az alakmátrix nem tartalmazza a paramétert. Ez azt jelenti, hogy az alakmátrix független a paramétertől, így annak számításához nem kell ismernünk (vagy feltételeznünk) a paraméter (előzetes) értékét.

Ez a fenti feltétel már néhány viszonylag egyszerű esetben sem teljesül.

Tekintsük az alábbi példát! Illesszünk sinus függvényt, a 9.1. táblázatban található x - y pontpárookra!

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
y	0	2	3	6	9	11	12	11	9	6	2	0	2	5	8	10	11	10

9.1. táblázat. x - y pontpárok.

A direkt feladat az alábbi alakú lehet:

$$y = a_0 + a_1 \cdot \sin(a_2 \cdot x + a_3) \quad (9.1)$$

A paraméterek közül a_0 egy konstans y irányú tolást jelent, a_1 a sinusos jel amplitúdóját, a_2 a körhullámszámot, a_3 pedig a jel fázisát. Képezzük (9.1) egyenletnek a mérési helyeken vett deriválásával az alakmátrixot!

$$\mathbf{A} = \begin{bmatrix} 1 & \sin(a_2 \cdot x_1 + a_3) & a_1 \cdot \cos(a_2 \cdot x_1 + a_3) \cdot x_1 & a_1 \cdot \cos(a_2 \cdot x_1 + a_3) \\ 1 & \sin(a_2 \cdot x_2 + a_3) & a_1 \cdot \cos(a_2 \cdot x_2 + a_3) \cdot x_2 & a_1 \cdot \cos(a_2 \cdot x_2 + a_3) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \sin(a_2 \cdot x_i + a_3) & a_1 \cdot \cos(a_2 \cdot x_i + a_3) \cdot x_i & a_1 \cdot \cos(a_2 \cdot x_i + a_3) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \sin(a_2 \cdot x_N + a_3) & a_1 \cdot \cos(a_2 \cdot x_N + a_3) \cdot x_N & a_1 \cdot \cos(a_2 \cdot x_N + a_3) \end{bmatrix} \quad (9.2)$$

Jól látszik, hogy az **alakmátrixban** szereplő mennyiségek **a paraméterek függvényei**.

A problémát úgy orvosolhatjuk, hogy a paramétereknek kezdőértéket (előzetes értéket vagy kiinduló értéket) adunk. Ezek a kezdőértékek lehetőség szerint a tényleges megoldáshoz közelítő értékek legyenek.

A sinus függvény illesztés esetén, ha a pontokra kézzel sinusgörbét rajzolnánk, meghatározhatnánk a paraméterek kezdőértékeit. Leolvasható lenne az illesztett függvény átlagértéke, vagyis az a_0 paraméter kezdőértéke. A rajzolt görbe amplitúdóját tekinthetjük az a_1 paraméter kezdőértékének. 2π -t elosztva a leolvasott hullámhosszal kapjuk az a_2 paraméter kezdőértékét. Megbecsüljük végül, hogy a függvényt vízszintes irányban a hullámhossz hányad részével kellene eltolni, hogy a sinushullám kezdőpontja (az inflexiós pontja) az $x = 0$ pont fölé essen, és ebből kiszámoljuk az a_3 paraméter kezdőértékét. A továbbiakban a paraméterek kezdőértékeit a felső indexbe, zárójellel írt 0-val jelöljük. Ezzel a jelöléssel a paramétervektor kezdőértéke: $\mathbf{x}^{(0)}$.

Ezeknek a mennyiségeknek a segítségével már ki tudnánk számolni az **A** alakmátrixot.

Az előző fejezetben említettük, hogy az $\mathbf{l}$ tisztatag vektort úgy definiálhatjuk, hogy az a mért értékekből kivonjuk a kezdő paraméterekkel (előzetes értékekkel) számított direkt feladat megoldást. Ezt a definíciót most is alkalmazzuk.

$$\mathbf{l} = \begin{bmatrix} y_1 - \left(a_0^{(0)} + a_1^{(0)} \cdot \sin \left(a_2^{(0)} \cdot x_1 + a_3^{(0)} \right) \right) \\ y_2 - \left(a_0^{(0)} + a_1^{(0)} \cdot \sin \left(a_2^{(0)} \cdot x_{21} + a_3^{(0)} \right) \right) \\ \vdots \\ y_i - \left(a_0^{(0)} + a_1^{(0)} \cdot \sin \left(a_2^{(0)} \cdot x_i + a_3^{(0)} \right) \right) \\ \vdots \\ y_N - \left(a_0^{(0)} + a_1^{(0)} \cdot \sin \left(a_2^{(0)} \cdot x_N + a_3^{(0)} \right) \right) \end{bmatrix} \quad (9.3)$$

Alkalmazzuk ezt a felírást a jelen feladatra! A direkt feladatot egy $\mathbf{f}(\dots)$ függvénnyel jelöljük. Keressük azt az $\mathbf{x}$ értéket (a paraméterek javítás vektorát) amelyre igaz az, hogy az értékeit hozzáadva a paraméterek kezdő értékeihez ($\mathbf{x}^{(0)} + \mathbf{x}$), az $\mathbf{f}(\mathbf{x}^{(0)} + \mathbf{x})$ függvény értékei a méréseink kiegyenlített mennyiségeit szolgáltatják ($\mathbf{U}$). Ekkor a direkt feladat megoldás a legjobban megközelíti a méréseinket ($\mathbf{L}$). A mérésekhez hozzáadva a javítások vektorát ($\mathbf{v}$) kapjuk a mérések kiegyenlített értékeit ($\mathbf{U}$).

$$\mathbf{L} + \mathbf{v} = \mathbf{U} = \mathbf{f}(\mathbf{x}^{(0)} + \mathbf{x}) \quad (9.4)$$

A függvény kiegyenlített értékét írjuk fel a függvény Taylor-sorfejtése segítségével:

$$\mathbf{f}(\mathbf{x}^{(0)} + \mathbf{x}) \approx \mathbf{f}(\mathbf{x}^{(0)}) + \mathbf{f}'(\mathbf{x}^{(0)}) \cdot \mathbf{x} + \dots \quad (9.5)$$

A függvény első deriváltja az $\mathbf{A}$ Jacobi mátrix. Ennek felhasználásával – kifejezve a javítások vektorát – kapjuk, hogy

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{x} - \mathbf{l} \quad (9.6)$$

ahol az $\mathbf{l}$ tisztatag vektor alakja:

$$\mathbf{l} = \mathbf{L} - \mathbf{f}(\mathbf{x}^{(0)}) \quad (9.7)$$

Látható, hogy a kapott (9.6) egyenlet **formailag megegyezik a lineáris** esettel.

Visszatérve a sinusfüggvény illesztési feladathoz, a paraméterek kezdő értékeivel kiszámolt $\mathbf{A}$ alakmátrix és $\mathbf{l}$ tisztatag vektor segítségével a (8.29) képlettel kiszámoljuk az $\mathbf{x}$ vektor. Az $\mathbf{x}$ vektor itt nem a paraméterek konkrét értékét jelenti, hanem a paraméterek javítását. A paraméterek javított értéke az $\mathbf{x}^{(0)}$ vektor és az $\mathbf{x}$ vektorok összege.

Amennyiben tehát a direkt feladat a paramétereknek nem a lineáris függvénye, úgy az Taylor-sor első két tagja nem adja meg pontosan a függvény értékét az ($\mathbf{x}^{(0)} + \mathbf{x}$) helyen, ennek következtében a kapott $\mathbf{x}$ vektorral javított paraméterek még nem adják meg az eltérés négyzetek minimumát.

Ennek a problémának a megoldására iterációt végzünk. Ennek során az első iterációs lépésben a fenti műveletekkel felírjuk az $\mathbf{A}$ alakmátrixot és az $\mathbf{l}$ tisztatag vektort, amiből kiszámoljuk paraméterek $\mathbf{x}$ javításvektorát. A második iterációs lépésben a megjavított ($\mathbf{x}^{(0)} + \mathbf{x}$) paramétervektort tekintjük a paraméterek kezdőértékének, ezzel számoljuk ki az $\mathbf{A}$ alakmátrixot és az $\mathbf{l}$ tisztatag vektort, majd ezekből a paraméterek javítását leíró $\mathbf{x}$ vektort. Ezt a lépés többször megismételve általában azt látjuk, hogy a javítások $\mathbf{x}$ vektora egyre kisebb abszolút értékű számokat tartalmaz, az újabb lépésekkel a kiegyenlített paramétervektor elemei konkrét értékekhez konvergálnak. Az iterációt általában néhány lépés után leállítjuk. A leállítás feltétele általában az, hogy a kapott javításvektor normája az újabb iterációs lépés hatására már kevesebbet változik, mint egy előre megadott kicsiny küszöbérték.

Probléma lehet, hogy a normál egyenlet csak azt írja elő, hogy a függvény **szélsőértékét** keressük, ahol a függvény első deriváltja nulla. Az, hogy egy **minimumhely**, vagyis itt a reziduálok négyzetösszege minimális, a függvény második deriváltjának vizsgálatából derül csak ki.

A minimalizálandó függvényünk tehát a reziduálok négyzetösszege. Ez egy skalár mennyiség. Esetünkben ezt **kritériumfüggvénynek** tekinthetjük. (A kritériumfüggvény általánosabb definícióját az inverzió statisztikus elméleténél tárgyaltuk.) A kritériumfüggvény által visszaadott skalár mennyiség, valóban a paraméterek függvénye, és mi azt a paraméterkombinációt (paraméter vektor elemeinek azon értékeit) keressük, ahol ez a függvény minimális. Ezt a függvényt jelöljük g -vel.

$$g(\mathbf{x}) = g(x_1, \dots, x_p) \quad (9.8)$$

Itt x_i -k a paramétervektor elemeit – nem pedig a javításukat – jelenti. Legyen a $g(\mathbf{x})$ függvény kétszer deriválható! Képezzük a $g(\mathbf{x})$ függvénynek a paraméterek szerinti második deriváltjait tartalmazó mátrixot!

$$\mathbf{H}(g) = \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 g}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 g}{\partial x_1 \partial x_p} \\ \frac{\partial^2 g}{\partial x_2 \partial x_1} & \frac{\partial^2 g}{\partial x_2^2} & \dots & \frac{\partial^2 g}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial x_p \partial x_1} & \frac{\partial^2 g}{\partial x_p \partial x_2} & \dots & \frac{\partial^2 g}{\partial x_p^2} \end{bmatrix} \quad (9.9)$$

A $\mathbf{H}$ mátrix neve **Hesse mátrix**. A parciális deriválás sorrendjének felcserélhetősége miatt a Hesse mátrix mindig **szimmetrikus**. A Hesse mátrix vizsgálata mutatja meg, hogy az adott függvényérték minimum, maximum, vagy nyeregpont.

A **minimum feltétele**, hogy a **Hesse mátrix pozitív definit** legyen.

A pozitív definitésg több megfogalmazása közül kiemelünk két definíciót. Az első szerint egy mátrix pozitív definit, ha valamennyi sajátértéke pozitív. Ezzel egyenértékű definíció, hogy valamennyi, a mátrix bal felső sarkából képzett négyzetes mátrix determinánsa pozitív legyen. (Konkrét számítás során bármelyik definíció felhasználható a pozitív definitésg vizsgálatára.)

A nemlineáris direktfeladat esetén láttuk, hogy a reziduálok négyzetösszegének minimumához tartozó paraméterértékek megtalálása csak több iterációs lépésben lehetséges. Ennek az az oka, hogy a direkt feladat függvényt a kiinduló paraméterek környezetében egy Taylor-sor első tagjáig bezárólag írtuk fel.

A kritériumfüggvényt felírhatjuk a kiinduló paraméterek környezetében a Taylor-sor második tagjáig bezárólag:

$$g(\mathbf{x}^{(0)} + \mathbf{x}) \approx g(\mathbf{x}^{(0)}) + \mathbf{J}(\mathbf{x}^{(0)}) \cdot \mathbf{x} + \frac{1}{2} \cdot \mathbf{x}^T \cdot \mathbf{H}(\mathbf{x}^{(0)}) \cdot \mathbf{x} \quad (9.10)$$

A fenti képletben a $\mathbf{J}$ Jacobi-mátrix a g skalár függvény **gradiense** (jelen esetben egy vektor), a $\mathbf{H}$ mátrix pedig az előbb megismert Hesse mátrix. Mivel ezt az egyenletet a kritériumfüggvényre írtuk fel, ennek a függvénynek a paraméterek (az $\mathbf{x}$ vektor elemei) szerinti minimalizálása adja a keresett paraméterek javításait. Az egyenletben szereplő $\mathbf{J}$ vektor és $\mathbf{H}$ mátrix a gyakorlati problémák esetén általában numerikus deriválással határozhatók meg a paraméterek kiinduló értékeinél ($\mathbf{x}^{(0)}$). Az egyenlet megoldása nagy mátrixok esetén általában a „Minimum kereső eljárások” c. fejezetben bemutatott iterációs eljárások valamelyikével történik.

9.1. Gravitációs mérés

Tekintsük az alábbi gravitációs mérési problémát, amelynek valamilyen változatát több inverziós könyv is felhasznál az inverzió alapelveinek bemutatására.

Egy megadott sík terület egy pontján egy nagytömegű gömb van elásva. Graviméterrel mérjük meg a nehézségi gyorsulást a területen egy (közel) egyenközű rácsban, és ebből határozzuk meg a gömb tömegét, felszín alatti mélységét és az elásás helyének síkkordinátáit!

Tételezzük fel, hogy egy jólképzett geofizikus elvégezte a mérést, tehát nekünk nem kell foglalkoznunk a mérés megtervezésével és végrehajtásával, rendelkezésünkre állnak a mérési helykoordináták és a graviméterrel mért nehézségi gyorsulás értékek.

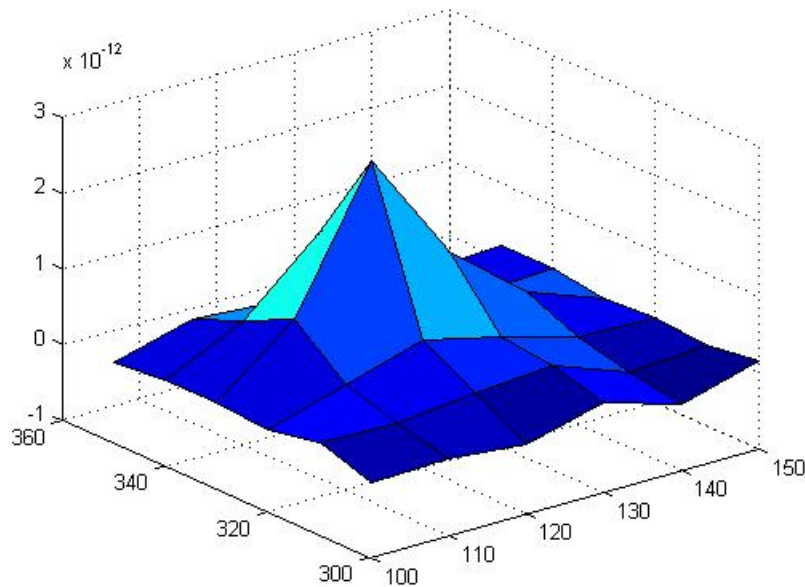
Íranyítsuk a koordinátarendszerünket úgy, hogy az X tengely észak felé, az Y tengely kelet felé, a Z tengely pedig lefelé irányuljon. A felszín a $Z = 0$ magasságban legyen. A felszín egy pontjának koordinátája így: $(x, y, 0)$. A nehézségi gyorsulás háttér értéke – vagyis amit a gömbtől olyan távolságban mérnénk, ahol a gömb hatása már elhanyagolható – legyen g_0 . A területen mért nehézségi gyorsulás értékekből vonjuk ki a nehézségi gyorsulás háttér értékét, az így kapott anomáliaértékeket tekinthetjük a gömb hatásának. A mért adatainkat a (9.2.) táblázat foglalja össze.

Mérés sorszáma	X[m]	Y[m]	Z[m]	$\Delta g[m/s^2]$
1	100	300	0	-1,68922e-014
2	100	310	0	2,08384e-013
3	100	320	0	9,11552e-014
4	100	330	0	1,3398e-013
5	100	340	0	1,35904e-013
6	100	350	0	7,48595e-014
7	110	300	0	1,07663e-014
8	110	310	0	1,71658e-013
9	110	320	0	3,90526e-013
10	110	330	0	9,64405e-013
11	110	340	0	6,26942e-013
12	110	350	0	3,54347e-013
13	120	300	0	-7,3977e-014
14	120	310	0	2,87387e-013
15	120	320	0	6,87852e-013
16	120	330	0	2,7552e-012
17	120	340	0	1,52322e-012
18	120	350	0	2,83205e-013
19	130	300	0	1,87324e-013
20	130	310	0	3,69917e-013
21	130	320	0	4,54991e-013
22	130	330	0	1,26672e-012
23	130	340	0	1,08389e-012
24	130	350	0	4,87435e-013
25	140	300	0	-1,34739e-013
26	140	310	0	-3,68395e-015
27	140	320	0	1,58061e-013
28	140	330	0	4,50285e-013
29	140	340	0	1,37541e-013
30	140	350	0	1,24567e-013

31	150	300	0	1,61628e-013
32	150	310	0	6,26579e-014
33	150	320	0	1,30561e-013
34	150	330	0	7,56786e-014
35	150	340	0	1,80268e-013
36	150	350	0	1,66113e-013

9.2. táblázat. Gravitációs anomáliaértékek (Δg^M). Az anomáliaértékek szimulált adatok, azokat egy ismert helyű és tömegű ható terének számításával, majd a számított anomáliaértékekhez véletlen jellegű zaj hozzáadásával kaptuk.

A fenti táblázat adatait egy kétváltozós függvény értékeinek tekinthetők. Ennek vizualizációját – a mérési pontokra illeszkedő szintfelületet – az alábbi ábrán láthatjuk.



9.1 ábra. A „mért” gravitációs anomáliák (Δg_z^M) megjelenítése szintfelületként.

A 9.1 ábrán egy aszimmetrikus anomália csúcs látszik. Ebből következtethetünk, hogy a ható nem pontosan a rácspontra helyezkedik el.

Az inverzió végrehajtására, először írjuk fel a **direkt feladatot**! Egy tetszőleges felszíni, $(x, y, 0)$ koordinátájú pontban az elásott gömb által okozott nehézségi gyorsulás (az ún. anomália), az alábbi képlettel számítható:

$$\Delta g_z^{\text{TH}} = \frac{G \cdot \rho \cdot \frac{4}{3} \pi \cdot r^3 \cdot z}{\left((x - x_0)^2 + (y - y_0)^2 + z_0^2 \right)^{3/2}} \quad (9.11)$$

A képletben szereplő G , a gravitációs állandó, x és y a mérési pont koordinátái. A képletben szereplő m, x_0, y_0, z_0 értékek azonban a feladat kiírásában szereplő meghatározandó ismeretlenek, vagyis a **paraméterek**.

A fenti képlet tehát azt fejezi ki, hogy ha ismernénk a paraméterek m, x_0, y_0, z_0 értékeit, ki tudnánk számolni, hogy egy tetszőleges $(x, y, 0)$ koordinátájú helyen a gömb mekkora anomáliát okozna.

Az $\mathbf{A}$ alakmátrix meghatározásához tehát a (9.11) egyenletnek a mérési helyeken a paraméterek szerinti deriváltjait kell képezni. Esetünkben ezek a deriváltak analitikusan is számolhatók:

$$\begin{aligned}\frac{\partial(\Delta g_z)}{\partial x_0} &= \frac{3 \cdot (x - x_0)}{\left((x - x_0)^2 + (y - y_0)^2 + z_0^2\right)} \cdot \Delta g_z \\ \frac{\partial(\Delta g_z)}{\partial y_0} &= \frac{3 \cdot (y - y_0)}{\left((x - x_0)^2 + (y - y_0)^2 + z_0^2\right)} \cdot \Delta g_z \\ \frac{\partial(\Delta g_z)}{\partial z_0} &= \frac{-3 \cdot z_0}{\left((x - x_0)^2 + (y - y_0)^2 + z_0^2\right)} \cdot \Delta g_z \\ \frac{\partial(\Delta g_z)}{\partial m} &= \frac{1}{m} \cdot \Delta g_z\end{aligned}\quad (9.12, 9.13, 9.14, 9.15)$$

Látszik, hogy ezekben a deriváltakban szerepelnek a paraméterek, tehát ahhoz hogy ezeket a deriváltakat számítani tudjuk, a paramétereknek valamilyen hihető kiinduló értékeket kell adnunk.

A paraméterek kiinduló értékeit tehát nekünk kell megadnunk. Válasszuk az anomália-térkép maximumhelyét a sikkordináták előzetes értékének: $x_0^{(0)} = 120\text{m}$ és $y_0^{(0)} = 330\text{m}$. A mélység előzetes értékét az anomáliakép félérték-szélességéből becsülhetjük: $z_0^{(0)} = 5\text{m}$. A gömb tömegének kiinduló értékét válasszuk az 1 méter sugarú ólomgömb tömegének: értéke: $m^{(0)} = 37866\text{kg}$. (Az ólomgömb tömegének számításánál az ólom és a felszínközeli kőzetek sűrűségkülönbségét használtuk.)

A felvett előzetes értékek alapján már a mérési helyekre ki tudjuk számolni az $\mathbf{A}$ alakmátrixot. A mátrix 4 oszlopból (a paraméterek száma) és 36 sorból (a méréseink száma) áll. Az $\mathbf{I}$ tiszttag vektort a 36 pontban mért anomáliaértékek (9.2. táblázat) és felvett előzetes paraméterértékek alapján kiszámolt elméleti értékek (9.11 képlet) különbségéből képezzük.

A paraméterek javítását a

$$\mathbf{x} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{I}) \quad (9.16)$$

alapján számoljuk. A számolt javítások:

$$\mathbf{x} = \begin{bmatrix} 0,09 \\ 0,21 \\ 0,21 \\ -33525,10 \end{bmatrix} \quad (9.17)$$

A javításokat hozzáadva az előzetes értékekhez megkapjuk a paraméterek kiegyenlített értékét. Újabb iterációs lépést végezve, a paramétereink tovább javulnak. Néhány iterációs lépés után a paramétereink már keveset változnak, ekkor leállítjuk az iterációt. A paraméterek becsült értékei:

$$\hat{\mathbf{x}} = \begin{bmatrix} 121,54 \\ 332,62 \\ 10,08 \\ 4748,75 \end{bmatrix} \quad (9.18)$$

A paraméterek kovariancia mátrixát a

$$\mathbf{D}_{XX} = \sigma_0^2 \cdot \mathbf{Q}_{XX} = \frac{\mathbf{v}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{v}}{N - P} \cdot (\mathbf{A}^T \cdot \mathbf{P} \cdot \mathbf{A})^{-1} \quad (9.19)$$

képlet segítségével számolhatjuk, ahol N a mérések, P pedig a paraméterek száma. A számításhoz a $\mathbf{P}_{LL}$ mátrixot egységmátrixnak tételeztük fel.

$$\mathbf{D}_{XX} = \begin{bmatrix} 0,137 & 0,005 & -0,044 & -14,152 \\ 0,005 & 0,105 & -0,047 & -15,232 \\ -0,044 & -0,047 & 0,170 & 79,377 \\ -14,152 & -15,232 & 79,377 & 53851,000 \end{bmatrix} \quad (9.20)$$

amiből a becsült paraméterek szórásai:

$$\sigma = \begin{bmatrix} 0,3706 \\ 0,3238 \\ 0,4122 \\ 232,0600 \end{bmatrix} \quad (9.21)$$

A gravitációs mérés feldolgozásához használt adatok nem valódi gravitációs adatok voltak, hanem **szimulált** adatok. Ezeket az adatok úgy álltak elő, hogy egy **rögzített paramétervektor** segítségével kiszámoltuk az anomáliaértékeket a feltételezett mérési pontokban, és az anomáliaértékeket 10^{-10} szórású, normális eloszlású, korrelálatlan zajjal szórtuk meg. A paraméterek rögzített értékei: $x_0 = 121$, $y_0 = 133$, $z_0 = 10$ és $m = 4733$ kg.

Jelen esetben a paraméterek becsült értékei a „valódi” értékeket az egyszeres szórásnál jobban megközelítik. Amennyiben a szimulált mérésekhez nagyobb hibát adtunk volna, úgy a paraméterek szórásai is jelentősen megnövekedtek volna.

Ebben a feladatban az anomáliaértékeket – gravitációs gyakorlattól eltérően – SI mértékegységben adtuk meg. A geofizikában szokásos mértékegységek esetén a maximális hatás $3 \cdot 10^{-9}$ m/s² nagyságrendbe esett, ami közel egyharmad *mikro*gal. Ez a legmodernebb – a szupravezetés elvén működő – graviméterekkel már jól mérhető pontosság.

10. fejezet - Kényszerfeltételek alkalmazása

Az inverzió során, amikor egy modell paramétereit határozzuk meg, lehetséges, hogy vannak előzetes – *a priori* – ismereteink a modellparaméterekről. Ezeket az előzetes ismereteket valamilyen matematika formában kell megfogalmaznunk.

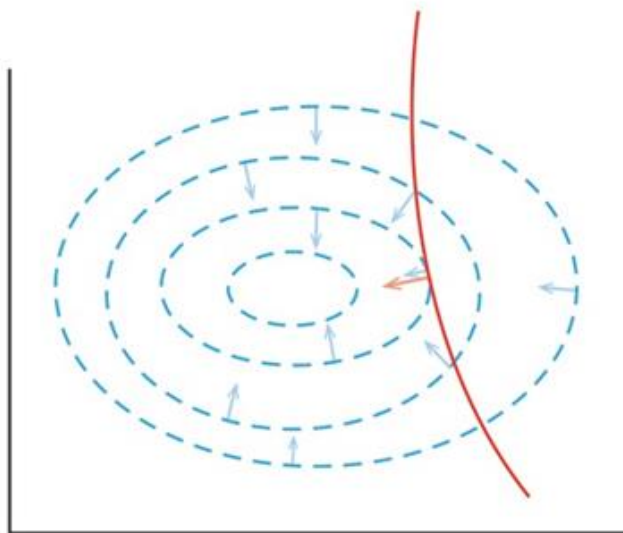
Amennyiben ezek az előzetes ismeretek determinisztikusak, vagyis mindenképpen teljesülniük kell, ezeket összefoglalóan **kényszerfeltételeknek** nevezzük. Van olyan kényszerfeltétel, amit valamilyen **egyenlőtlenség** formájában tudunk megfogalmazni. Ilyen feltétel lehet, hogy egy paraméter csak egy bizonyos tartományba eső értéket vehet fel. Más esetben egy **egyenlőség** formájában tudjuk megadni, a függvénykapcsolatot, vagyis hogy a paraméterek valamilyen függvénye egy konkrét értéket vegyen fel.

Eszerint a kényszerfeltételek alkalmazása az inverzió során azt jelenti, hogy úgy szeretnénk a paramétereket meghatározni, hogy a kritériumfüggvényünk (a reziduálok négyzetösszege) minimális legyen, de közben az vagy több kényszerfeltételi egyenletünk is teljesüljön. Ez egy **feltételes szélsőérték keresési** probléma, amit Lagrange oldott meg a XIX század elején.

10.1. A feltételes szélsőérték keresés

Tekintsünk egy E függvényt, amely két változó (x és y) függvénye: $E(x,y)$. (Egy kétváltozós skalárfüggvény egy felületként ábrázolható, ahol a függvény értéke egy z magasság az x - y sík felett.) Minimalizáljuk úgy az E függvényt,

hogy közben az x és y változókra teljesül, a $\Phi(x,y)=0$ implicit alakban felírt kényszerfeltétel. A kényszerfeltételnek egy görbe felel meg az x - y síkon. Amennyiben ebből a kényszerfeltételből ki tudnánk fejezni akár $y(x)$ -t akár $x(y)$ -t, azt visszahelyettesíthetnénk az E függvénybe, és ezután egyetlen változó szerint kellene E minimumát meghatározni. Sajnos ezt az esetek többségében nem tudjuk megvalósítani, ezért az alább vázolt módszert kell követnünk.



10.1 ábra: Az $E(x,y)$ kétváltozós függvény szintvonalai (kék szaggatott vonalak) és egy, a változók közötti, $\Phi(x,y)=0$ kényszerfeltételnek eleget tevő pontok halmaza (piros görbe). Abban a pontban, ahol a függvény értéke a görbe mentén minimális, a függvény gradiense és a görbe normálisa párhuzamosak.

A 10.1 ábrán látszik, hogy az E függvény minimumánál a függvény **gradiense** nulla.

Az E függvény gradiense az alábbi vektor:

$$\nabla E = \left(\frac{\partial E}{\partial x}, \frac{\partial E}{\partial y} \right) \quad (10.1)$$

Látszik, hogy a gradiens merőleges az E függvény egyenlő értékeket felvevő szintvonalaira (ekvipotenciális görbéire).

Az ábrába berajzoltunk egy kényszerfeltételi egyenletnek megfelelő görbét. A görbe **normálvektora** a görbére az adott pontjában húzott érintőre merőleges vektor, amely a kényszerfeltételi egyenletnek a változók szerinti parciális deriváltjaiból képezhető:

$$\mathbf{n} = \left(\frac{\partial \Phi}{\partial x}, \frac{\partial \Phi}{\partial y} \right) \quad (10.2)$$

Az ábrán látszik, a Φ görbe azon pontja, ahol az E függvénynek minimuma van, ott az E függvény gradiense, és a Φ görbe normálisa párhuzamosak. Ezt úgy fejezhetjük ki, hogy a

$$\nabla E = -\lambda \cdot \mathbf{n} \quad (10.3)$$

ahol λ egy skalár, arányossági tényező. Ezt elemenként kiírva, és nullára rendezve az alábbi két egyenletet kapjuk:

$$\left(\frac{\partial E}{\partial x} + \lambda \cdot \frac{\partial \Phi}{\partial x} \right) = 0 \quad (10.4)$$

$$\left(\frac{\partial E}{\partial y} + \lambda \cdot \frac{\partial \Phi}{\partial y} \right) = 0 \quad (10.5)$$

Ez a két egyenlet annak felel meg, mintha az

$$E(x, y) + \lambda \cdot \Phi(x, y) := \min \quad (10.6)$$

összegfüggvénynek keresnénk a minimumát, ugyanis ennek az összegfüggvénynek az adott pontban a változók szerint parciális deriváltja nullák, tehát ennek az összegfüggvénynek szélső értéke van.

A fenti két egyenlet mellé felhasználva a $\Phi(x, y) = 0$ kényszerfeltételi egyenletet, összesen három egyenletünk van, és három ismeretlenünk (x, y, λ) , amivel a probléma már megoldható.

A fenti eredményeket $p=1, 2, \dots, P$ változóból álló $\mathbf{x}$ paramétervektor, és $r=1, 2, \dots, R$ darab kényszerfeltétel esetére is általánosíthatjuk. Valamennyi p esetére teljesülnie kell az alábbi feltételnek:

$$\frac{\partial E}{\partial x_p} + \sum_{r=1}^R \lambda_r \frac{\partial \Phi_r}{\partial x_p} = 0 \quad (10.7)$$

Valamint valamennyi r esetére:

$$\Phi_r(\mathbf{x}) = 0 \quad (10.8)$$

Ez összesen $P+R$ számú meghatározandó ismeretlent jelent, $P+R$ egyenletből.

A fenti eredményeket felhasználjuk a továbbiakban.

A fenti példában egy E skalárfüggvény szélsőértékét kerestük. A legkisebb négyzetek alkalmazásának bevezetőjében láttuk, hogy a minimalizálandó skalárfüggvény a reziduálok négyzetösszege. Ha tehát az E skalárfüggvényt megfeleltetjük a reziduálok négyzetösszegének és a kényszerfeltételeinket fel tudjuk írni (10.8) alakban, akkor ennek segítségével meg tudjuk oldani a kiegyenlítést kényszerfeltételek alkalmazása mellett.

A korábbi fejezetekben láttuk, hogy a reziduálokat (javításokat) fel tudjuk írni, mint a paraméterek lineáris függvényét (8.9 egyenlet).

Szükségünk van még a kényszerfeltételi egyenletek átalakítására, hogy azok is a paraméterek lineáris függvényei legyenek. Láttuk, hogy $r=1, 2, \dots, R$ darab kényszerfeltételi egyenletünk van. Minden egyenlet az alábbi alakú:

$$\Phi_r(\mathbf{x}) = 0 \quad (10.9)$$

A paramétervektor az alábbi alakú:

$$\mathbf{x} = (x_1, x_2, \dots, x_p, \dots, x_p) \quad (10.10)$$

Amennyiben a fenti egyenlet az $\mathbf{x}$ paramétervektor lineáris függvénye, az r -edik lineáris kényszerfeltétel az alábbi alakba írható:

$$c_{r1} \cdot x_1 + c_{r2} \cdot x_2 + \dots + c_{rp} \cdot x_p + \dots + c_{rp} \cdot x_p - w_r = 0 \quad (10.11)$$

Ahol a $c_{r1}, c_{r2}, \dots, c_{rp}$ együtthatók az $\mathbf{x}$ paramétervektor szorzói, a w_r mennyiség pedig a konstans tag.

Az R darab lineáris kényszerfeltételt az alábbi mátrix-vektor formába is átírhatjuk:

$$\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} & \dots & c_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{r1} & c_{r2} & \dots & c_{rp} & \dots & c_{rp} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{R1} & c_{R2} & \dots & c_{Rp} & \dots & c_{Rp} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \\ \vdots \\ x_p \end{bmatrix} - \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_R \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (10.12)$$

Ezzel a kényszerfeltételi egyenletet

$$\mathbf{C} \cdot \mathbf{x} - \mathbf{w} = \mathbf{0} \quad (10.13)$$

alakban sikerült felírunk.

A **nemlineáris kényszerfeltételek** esetében, csakúgy, mint a nemlineáris direktfeladat esetén, fel kell vegyünk a paramétereknek előzetes értéket $\mathbf{x}_0$ -t, és az $\mathbf{x}$ vektor ezeknek az előzetes paramétereknek a javítása.

A kényszerfeltételek linearizálásakor tehát abból indulunk ki, hogy az $\mathbf{x}$ vektor a paraméterek kiinduló értékeiből képzett $\mathbf{x}^{(0)}$ kezdeti paramétervektorhoz hozzáadandó javítás. Ekkor a kényszerfeltételi egyenletet $\mathbf{x}^{(0)}$ körül, $\mathbf{x}$ szerint sorbafejtve, és a sorfejtést csak a lineáris tagig meghagyva kapjuk:

$$\Phi_r(\mathbf{x}^{(0)} + \mathbf{x}) = \Phi_r(\mathbf{x}^{(0)}) + \Phi_r'(\mathbf{x}^{(0)}) \cdot \mathbf{x} + \dots \quad (10.14)$$

A fenti mennyiség a kényszerfeltétel definíciójából adódóan nulla. Az (10.14) egyenletet nullára rendezve, és a lineáris esetben tárgyalta (10.12) mátrix-vektor formára hozva, kapjuk az alábbi alakot:

$$\begin{bmatrix} \frac{\partial \Phi_1}{\partial x_1} & \frac{\partial \Phi_1}{\partial x_2} & \dots & \frac{\partial \Phi_1}{\partial x_p} & \dots & \frac{\partial \Phi_1}{\partial x_p} \\ \frac{\partial \Phi_2}{\partial x_1} & \frac{\partial \Phi_2}{\partial x_2} & \dots & \frac{\partial \Phi_2}{\partial x_p} & \dots & \frac{\partial \Phi_2}{\partial x_p} \\ \vdots & \vdots & & \vdots & & \vdots \\ \frac{\partial \Phi_R}{\partial x_1} & \frac{\partial \Phi_R}{\partial x_2} & \dots & \frac{\partial \Phi_R}{\partial x_p} & \dots & \frac{\partial \Phi_R}{\partial x_p} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \\ \vdots \\ x_p \end{bmatrix} - \begin{bmatrix} -\Phi_1(\mathbf{x}^{(0)}) \\ -\Phi_2(\mathbf{x}^{(0)}) \\ \vdots \\ -\Phi_R(\mathbf{x}^{(0)}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (10.15)$$

A fenti (10.15) és a lineáris esetben kapott (10.12) alakok összehasonlításából látszik, hogy:

$$\mathbf{w}_r = -\Phi_r(\mathbf{x}^{(0)}) \quad (10.16)$$

Visszatérve az összegfüggvényhez, most már a minimalizálandó függvényt felírhatjuk a linearizált javítási- és a feltételi egyenletek segítségével:

$$\Phi = \mathbf{v}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{v} - 2 \cdot \mathbf{k}^T \cdot (\mathbf{C} \cdot \mathbf{x} - \mathbf{w}) := \min \quad (10.17)$$

A minimalizálandó függvényben itt a λ helyett – célszerűségi okokból – a $\mathbf{v}$ tag szerepel. A $\mathbf{k}$ vektort korrelátavektornak nevezzük.

A függvénynek ott lesz szélsőértéke, ahol a teljes differenciál nulla lesz. A teljes differenciál alakja a következő:

$$2\mathbf{v}^T \mathbf{P}_{LL} d\mathbf{v} - 2\mathbf{k}^T \mathbf{C} d\mathbf{x} - 2d\mathbf{k}^T (\mathbf{C}\mathbf{x} - \mathbf{w}) = 0 \quad (10.18)$$

Felhasználva a linearizált javítási egyenletet:

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{x} - \mathbf{l} \quad (10.19)$$

Kapjuk, hogy

$$d\mathbf{v} = \mathbf{A} \cdot d\mathbf{x} \quad (10.20)$$

Amit visszaírva a teljes differenciálba, a következő parciális deriváltakhoz jutunk:

$$\begin{aligned} \mathbf{v}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A} - \mathbf{k}^T \cdot \mathbf{C} &= 0 \\ -\mathbf{C} \cdot \mathbf{x} + \mathbf{w} &= 0 \end{aligned} \quad (10.21 - 10.22)$$

A fenti egyenletbe visszaírva a (10.19) javítási egyenletet, kapjuk:

$$\begin{aligned} (\mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A}) \cdot \mathbf{x} - \mathbf{C}^T \cdot \mathbf{k} - (\mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{l}) &= 0 \\ -\mathbf{C} \cdot \mathbf{x} + \mathbf{w} &= 0 \end{aligned} \quad (10.23 - 10.24)$$

A fenti két egyenlet hipermátrix formában is felírható:

$$\begin{bmatrix} \mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A} & -\mathbf{C}^T \\ -\mathbf{C} & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x} \\ \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{l} \\ -\mathbf{w} \end{bmatrix} \quad (10.25)$$

Ez a feladat megoldásához szükséges normálegyenlet. Amennyiben a normálegyenlet együtthatómátrixa nem szinguláris (invertálható), a fenti egyenletből kifejezhetjük a paraméterek és a korreláták vektorát:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{A} & -\mathbf{C}^T \\ -\mathbf{C} & \mathbf{0} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{A}^T \cdot \mathbf{P}_{LL} \cdot \mathbf{l} \\ -\mathbf{w} \end{bmatrix} \quad (10.26)$$

A feltételes szélsőértékkeresési feladat megoldására itt kapott eredményeket nem csak a kényszerfeltételek, hanem a csak mért mennyiségeket tartalmazó kiegyenlítési feladatban is felhasználjuk.

10.2. Egyenes illesztése – lineáris kényszerfeltétel

Feladat: Illesszünk egyenest az alábbi pontokra, úgy, hogy az egyenes az y tengelyt egy előre megadott pontban metssze:

x	y
1	6
2	7,1
3	8
4	9,1

10.1. táblázat. Négy pont síkkordinátái

Írjuk elő, hogy az egyenes az y tengelyt az $y=5$ pontban metssze!

A direkt feladat megoldásból látszik (8.16), hogy az $\mathbf{A}$ alakmátrix alakja:

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix} \quad (10.27)$$

A tisztatag vektor

$$\mathbf{l} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \quad (10.28)$$

Az ismeretlen paraméterek vektora tartalmazza az egyenes együtthatóit:

$$\mathbf{x} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \quad (10.29)$$

A kényszerfeltételi egyenletünket kell tehát linearizált formában felírni:

$$a_0 = 5 \quad (10.30)$$

ezt a szükséges $\mathbf{C} \cdot \mathbf{x} - \mathbf{w} = 0$ alakba hozva adódik, hogy:

$$\mathbf{C} = [1 \ 0] \text{ és } \mathbf{w} = 5 \quad (10.31, 10.32)$$

Az egyenlet megoldásakor feltételezzük, hogy a mérések egyforma súlyúak, vagyis $\mathbf{P}_{LL} = \mathbf{I}$. A (10.26) egyenlet felhasználásával a paramétervektor javításai, (amik a nulla kiinduló értékek miatt) egyben a kiegyenlített értékei $\mathbf{X} = [5 \ 1,02]$ lesznek, a korreláta $\mathbf{k} = -4.44 \cdot 10^{-15}$ lesz.

10.3. Egyenes illesztése – nem lineáris kényszerfeltétel

Illesszünk ismét egyenest az előbbi 4 pontra. Ebben az esetben az egyenes egyenletét ne a megszokott

$$y = a_0 + a_1 \cdot x \quad (10.33)$$

alakban adjuk meg, hanem használjuk fel a koordináta geometriából ismert

$$n_x \cdot x + n_y \cdot y + c = 0 \quad (10.34)$$

alakot. Ez utóbbi alakban, az egyenes **implicit** megadásában szereplő n_x és n_y mennyiségek az egyenes **normálvektor**ának komponensei, a c konstans abszolút értéke pedig az egyenes és az origó távolságát adja meg. (A c együttható értéke negatív is lehet, ha a normálvektor abba a térfélbe mutat, amelyik nem tartalmazza az origót.) Az itt alkalmazott implicit megadás előnye az, hogy olyan esetben is tudunk egyenest illeszteni a pontokra, ha az illeszkedő egyenes az y -tengellyel (közel) párhuzamos.

Ebben az esetben a modell felírásánál nem tételezhetjük fel a modellparaméterek zéró értékét, mivel a normálvektor egységnyi hosszúságú. Ebben az esetben felvesszünk a paramétereknek valamilyen kiinduló értéket. Válasszunk olyan kiinduló paramétereket, hogy azok egy az x tengelyre eső egyenest írjanak le: $n_x^{(0)} = 0$ és $n_y^{(0)} = 1$, valamint $c^{(0)} = 0$.

Ezeket a paramétereket visszahelyettesíthetjük a direkt feladat (10.34) egyenletébe. Az egyenletet a megadott x - y pontpárookra alkalmazva, kapjuk a **javítási egyenleteket**. Az egyenletek jobb oldalán azonban nem nulla érték szerepel, ezeket a mennyiségeket tekinthetjük a **tisztatag vektor** elemeinek. (Vegyük észre, hogy ezek az egyes pontoknak az egyenestől mért – előjeles – távolságai.)

$$\mathbf{l} = \begin{bmatrix} n_x^{(0)} \cdot x_1 + n_y^{(0)} \cdot y_1 + c^{(0)} \\ n_x^{(0)} \cdot x_2 + n_y^{(0)} \cdot y_2 + c^{(0)} \\ n_x^{(0)} \cdot x_3 + n_y^{(0)} \cdot y_3 + c^{(0)} \\ n_x^{(0)} \cdot x_4 + n_y^{(0)} \cdot y_4 + c^{(0)} \end{bmatrix} \quad (10.35)$$

Szükségünk van az **alakmátrixra**, vagyis az egyes feltételi egyenleteknek a paraméterek szerinti parciális deriváltjaira. Mivel az egyenletek a paraméterek lineáris függvényei, az alakmátrixot könnyen megkapjuk az alábbi formában:

$$\mathbf{A} = \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \\ x_4 & y_4 & 1 \end{bmatrix} \quad (10.36)$$

A **kényszerfeltételünk** az, hogy a normálvektor hossza egységnyi:

$$\sqrt{n_x^2 + n_y^2} - 1 = 0 \quad (10.37)$$

Ennek az egyenletnek a paraméterek kiinduló értékeinél vett deriváltja alkotja a **C** mátrixot:

$$\mathbf{C} = \begin{bmatrix} \frac{n_x^{(0)}}{\sqrt{(n_x^{(0)})^2 + (n_y^{(0)})^2}}; \frac{n_y^{(0)}}{\sqrt{(n_x^{(0)})^2 + (n_y^{(0)})^2}}; 0 \end{bmatrix} \quad (10.38)$$

Látszik, hogy a mátrix harmadik eleme, vagyis a kényszerfeltételi egyenletnek a *c* paraméter szerinti deriváltja nulla, mivel a *c* paraméter nem szerepel a kényszerfeltételi egyenletben.

A – jelen esetben – egy elemű **w** vektor értéke a (10.15) egyenlet alapján:

$$\mathbf{w} = - \left(\sqrt{(n_x^{(0)})^2 + (n_y^{(0)})^2} - 1 \right) \quad (10.39)$$

Ezeket a mennyiségeket felhasználva az **x** vektor számításához (10.26 egyenlet) kapjuk, hogy:

$$n_x = -1,02; n_y = -7,26 \cdot 10^{18}; c = -5 \text{ és a korreláta: } \mathbf{k} = -0,008.$$

Ezeket az értékeket hozzáadva a paraméterek kiinduló értékeihez kapjuk:

$n_x^{(1)} = -1,02; n_y^{(1)} = 1; c^{(1)} = -5$. A javított paraméterértékekből látszik, hogy a normálvektor iránya már közelítőleg jó (merőleges az $y = x$ egyenesre) de a hossza nem egységnyi. A számítást megismételjük ezeket az értékeket felhasználva. Ekkor:

$$n_x = 0,305; n_y = -0,300; c = 1,502 \text{ és a korreláta: } \mathbf{k} = -0,00391.$$

Ezekkel az értékekkel megjavítva a paramétervektort kapjuk:

$n_x^{(2)} = -0,714; n_y^{(2)} = 0,699; c^{(2)} = -3,497$. Ezekkel az értékekkel számolva már az egyenes normálvektora csak a hetedik tizedesjegyben tér el egységtől. Az újabb iterációs lépések már nem javítják számottevően az eredményeket, az egyenes illeszkedése nem lesz jobb.

A fenti példa jól szemlélteti, hogy nemlineáris kényszerfeltételi egyenleteknél több iterációban találjuk meg azt a megoldást, amelynek a pontossága már megfelelő. Az első iterációs lépés a normálvektor irányát korrigálta, a második iteráció pedig a hosszát. Természetesen, ha a kiindulásnál, már egy a pontokra jobban illeszkedő egyenest leíró $(-0,7; 0,7; -3,5)$ paraméterkombinációból indultunk volna ki, akkor egy vagy két iterációs lépés után már nem javult volna számottevően az eredmény.

11. fejezet - Kiegyenlítés csak mért mennyiségeket tartalmazó feltételi egyenletekkel

Az ebben a fejezetben bemutatásra kerülő probléma – mivel **nem tartalmaz meghatározandó paramétereket** – általában nem tárgya a geofizikai inverziót tárgyaló könyveknek. A téma tárgyalását az indokolja, hogy segítségével be tudjuk mutatni az egyes mérések közvetlen súlyozását.

Mérjük meg két pont koordinátáit GPS segítségével (valamely vetületi koordináta rendszerben, amely esetben a vetületi egyenletekből következő hossztorzulást elhanyagolhatónak tekinthetjük). Mérjük meg ugyanezen két terepi pont távolságát mérőszalaggal is!

Az esetek nagy részében, a két pont koordinátakülönbségeiből Pitagorasz-tétellel számított távolság, és a mérőszalaggal mért távolság különbözni fog. Igaz továbbá, hogy ezen 5 mérés (két-két koordináta és a távolság) közül egyet elhagyva a hiányzó ötödik mennyiséget kiszámíthatjuk. (Ez nyilvánvaló a mérőszalaggal mért távolság elhagyása esetén. Valamely koordináta elhagyása esetén nulla, egy, vagy két lehetséges koordinátaértéket kaphatunk a rendelkezésre álló 4 adatból számítással.) Ez azt jelenti, hogy egy darab, **fölösmérés**ünk van.

Az az igényünk, hogy a mért távolság megegyezzen a koordinátakülönbségekből számított távolsággal, egy **feltételi egyenlettel** fejezhető ki. Ez a feltételi egyenlet azonban a mért értékeinkre nem teljesül:

$$\sqrt{(E_2 - E_1)^2 + (N_2 - N_1)^2} - d_{12} \neq 0 \quad (11.1)$$

A fenti egyenlet jobb oldalán, a nullától különböző mennyiséget itt is **tisztatagnak** nevezzük és l -lel jelöljük. Amennyiben több feltételi egyenletet tudunk felírni, a tisztatag mennyiségeket az **$\mathbf{l}$** vektorba rendezhetjük.

Tételezzük fel, hogy valamilyen formában sikerült meghatároznunk azokat a mennyiségeket (a **javitásokat**), amelyeket a mért értékeinkhez hozzáadva olyan javított (kiegyenlített) mennyiségeket kapunk, amelyek már kielégítik a feltételi egyenleteket:

$$\sqrt{((E_2 + v_{E2}) - (E_1 + v_{E1}))^2 + ((N_2 + v_{N2}) - (N_1 + v_{N1}))^2} - (d_{12} + v_{d12}) = 0 \quad (11.2)$$

Az egyenletben szereplő javításokat a – a fenti példában ötelemű – javítás vektor (**$\mathbf{v}$**) elemeinek tekintjük.

A problémát az jelenti, hogy egyetlen feltételi egyenletünk van, azonban 5 meghatározandó mennyiségünk, a javításvektor 5 eleme. Emiatt a feladat alulhatározott. A problémát egy további feltétel bevezetésével orvosolhatjuk, ha előírjuk, hogy a javítások (a **$\mathbf{v}$** vektor elemei) a lehető legkisebbek legyenek, vagyis a négyzetösszegük minimális legyen, úgy, hogy közben a feltételi egyenlet is teljesüljön a javításokkal korrigált mérési eredményekre.

Elfogadjuk-e azonban azt a megoldást, amelyből az jön ki, hogy mind a GPS méréseket, mind a mérőszalagos mérést 2-2 méterrel kell módosítanunk, hogy a feltételi egyenlet teljesüljön? Nyilvánvaló, hogy a mérőszalagos mérésről nehezen fogadjuk el, hogy 10 cm-nél pontatlanabb lenne, a GPS mérésnél akár 10-20 méteres hibát is elhiszünk. Ezt úgy tudjuk figyelembe venni, hogy az egyes méréseket súlyozzuk, és a javítások **súlyozott négyzetösszegét** minimalizáljuk.

A súlyozáshoz általában használhatjuk az egyes mért mennyiségek szórásnégyzetének a reciprokját. Egy nagy szórású mérést – a szórás négyzetekkel osztva – tehát kis súllyal vesszük figyelembe, egy kis szórású mérést pedig nagy súllyal. A súlyokat egy négyzetes mátrix főátlójába is elhelyezhetjük, ekkor a diagonális mátrix főátlójában az egyes mérések szórásnégyzeteinek reciprokai állnak. Ez abban az esetben közelíti jól az adott problémát, ha az egyes mérések korrelálatlanok. Korrelált mérések esetén a **mérések kovarianciamátrixának inverzét** használjuk **súlymátrixnak**.

A mérések súlyait a már korábban megismert súlymátrix eleme tartalmazza. A súlymátrix segítségével a minimumfeltétel alakja:

$$\mathbf{v}^T \mathbf{P}_{LL} \mathbf{v} := \min \quad (11.3)$$

A példánál maradva a GPS mérések szórásának tekinthetjük a GPS által a kijelzőn megjelenített HDOP értékeket (példánkban 5 méter). A mérőszalagos mérés hibáját 2 cm-nek tekinthetjük. Ebben az esetben a $\mathbf{P}_{LL}$ mátrix alakja:

$$\begin{pmatrix} \frac{1}{\sigma_1^2} & & & & \\ & \ddots & & & \\ & & \frac{1}{\sigma_N^2} & & \end{pmatrix} = \begin{pmatrix} 0,04 & & & & \\ & 0,04 & & & \\ & & 2500 & & \\ & & & 0,04 & \\ & & & & 0,04 \end{pmatrix} \quad (11.4)$$

Ahol a mérések sorrendje: $E_1, N_1, d_{12}, E_2, N_2$, emiatt a mátrixban is ebben a sorrendben tüntetjük fel az egyes mérések szórását.

Fontos megjegyezni, hogy mivel a $\mathbf{P}_{LL}$ mátrixot diagonális mátrixként írtuk fel, ezzel azt tételezzük fel, hogy az egyes méréseink függetlenek egymástól!

Ahogy korábban megfogalmaztuk, az (11.3) egyenlet minimumát úgy keressük, hogy közben az (11.2) egyenlettel kifejezett feltétel is teljesüljön. Ehhez az (11.2) egyenletet olyan formába hozunk, hogy az a javítások lineáris függvénye legyen. Ehhez az (11.2) egyenletet **Taylor-sorba fejtjük a javítások szerint az aktuális mérési értékek körül**, és csak a nullad-, és elsőrendű tagokat hagyjuk meg. Egy feltételi egyenlet esetén ez általánosságban az alábbi alakú:

$$f(\mathbf{L} + \mathbf{v}) = f(\mathbf{L}) + f'(\mathbf{L}) \cdot \mathbf{v} + \dots \quad (11.5)$$

ahol az egyenlet baloldala nulla (a nullára rendezett feltételi egyenlet miatt). Az egyenlet jobb oldalán az $f(\mathbf{L})$ mennyiséggel már találkoztunk, ez a mennyiség áll a (11.1) egyenlet jobb oldalán. Az egyenlet jobb oldalán az $f'(\mathbf{L})$ mennyiség a feltételi egyenletnek a mérések szerinti parciális deriváltja. Amennyiben több feltételi egyenletet tudunk felírni, akkor a feltételi egyenletek egy vektorfüggvény elemeit képezik, amelynek elemeire a Taylor-sorfejtés:

$$\mathbf{f}(\mathbf{L} + \mathbf{v}) = \mathbf{f}(\mathbf{L}) + \mathbf{f}'(\mathbf{L}) \cdot \mathbf{v} + \dots \quad (11.6)$$

ahol $\mathbf{f}'(\mathbf{L})$ a feltételi egyenleteknek a mérések szerint parciális deriváltjaiból képzett Jacobi mátrix, amit $\mathbf{B}$ -vel jelölünk. A $\mathbf{v}$ vektor a méréseink javításai, amit meg akarunk határozni. Az $f(\mathbf{L})$ mennyiség mínusz egyszeresét nevezzük tisztatagnak, $\mathbf{l} = -\mathbf{f}(\mathbf{L})$, ugyanis formálisan megegyezik a paraméterek javításainak meghatározására felírt egyenletekben szereplő tisztataggal. Felhasználva az $\mathbf{l}$ tisztatag vektor (11.2) egyenletét, a feltételi egyenletet az alábbi linearizált alakba írhatjuk át:

$$\mathbf{B}\mathbf{v} - \mathbf{l} = \mathbf{0} \quad (11.7)$$

A jelenlegi példánkban a $\mathbf{B}$ mátrix 1×5 elemű, a $\mathbf{v}$ vektor 5×1 elemű (oszlopvektor), az $\mathbf{l}$ vektor 1 elemű.)

A **feltételes szélsőérték kereséséhez** az alábbi egyenletet írhatjuk fel (hasonlóan a 10.17 egyenlethez):

$$\mathbf{v}^T \mathbf{P}_{LL} \mathbf{v} - 2 \cdot \mathbf{k}^T (\mathbf{B}\mathbf{v} - \mathbf{l}) := \min \quad (11.8)$$

Ahol $\mathbf{k}$ vektor az ún. Lagrange-multiplikátor (korreláta).

Míthogy a (11.8) egyenleteknek $\mathbf{v}$ illetve $\mathbf{k}$ szerint szélsőértéke van, ezért az egyenlet $\mathbf{v}$ illetve $\mathbf{k}$ szerint deriválja nulla kell legyen. A deriválást elvégezve majd az első egyenletet átrendezve az alábbi egyenleteket kapjuk:

$$\begin{aligned} \mathbf{P}_{LL} \mathbf{v} - \mathbf{B}^T \mathbf{k} &= 0 \\ \mathbf{B}\mathbf{v} - \mathbf{l} &= 0 \end{aligned} \quad (11.9-11.10)$$

A (11.9-11.10) egyenletek hipermátrix formába rendezhetők:

$$\begin{pmatrix} \mathbf{P}_{LL} & -\mathbf{B}^T \\ -\mathbf{B} & \mathbf{0} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{v} \\ \mathbf{k} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ -\mathbf{l} \end{pmatrix} \quad (11.11)$$

A javítások $\mathbf{v}$ vektorát megkaphatjuk az (11.11) egyenlet megoldásával, vagy akár az (11.9) képletben szereplő felső egyenletből kifejezzük a $\mathbf{v}$ vektort az alábbi alakban:

$$\mathbf{v} = \mathbf{P}_{LL}^{-1} \mathbf{B}^T \mathbf{k} \quad (11.12)$$

Majd azt a (11.10) képlet alsó egyenletébe helyettesítjük:

$$(\mathbf{B} \mathbf{P}_{LL}^{-1} \mathbf{B}^T) \cdot \mathbf{k} - \mathbf{l} = \mathbf{0} \quad (11.13)$$

A (11.13) egyenlet a megoldandó normálegyenlet, amelyből kifejezve $\mathbf{k}$ vektort, majd visszaírva a (11.12) egyenletbe kapjuk a keresett javításokat.

11.1. A feladat módosítása: több mérés két álláspont között

Módosítsuk úgy a feladatunkat, hogy a mérőszalagos mérés mellett, kompasz segítségével meghatározzuk az álláspontok közötti **irányszöget** is!

Ezek alapján két mérési pont között a pontok különböző módszerekkel mért távolságára felírt feltételi egyenlet (11.1) mellé felírhatunk még egy feltételt: az álláspontok GPS-sel mért koordinátáinak különbségéből származó irányszög, és a kompasszal mért irányszög (azimut) különbségére:

$$\arctan((N_2 - N_1)/(E_2 - E_1)) - \varphi_{12} \neq 0 \quad (11.14)$$

Az egyenletbe a mért mennyiségeket behelyettesítve az egyenlet jobb oldalán nullától különböző értéket kapunk. Ennek mínusz egyszerűsét tekintjük a tisztatagnak.

A mért szögek előzetes hibáját 5° -nak tekintjük, és a mágneses északi irány valamint a vetületi rendszer északi iránya közötti különbséget hanyagoljuk el. A gyakorlati számítások során az $\text{atan}()$ függvény helyett az $\text{atan2}()$ függvényt alkalmaztuk, argumentuma nem az $(N_2 - N_1)/(E_2 - E_1)$ hányados, hanem az $(N_2 - N_1)$ és az $(E_2 - E_1)$ koordinátakülönbségek. Ennek segítségével az $\text{atan2}()$ függvény automatikusan kezeli a ténnyedek problémáját. A szögek különbség értékkészlete (a tisztatag) -180° és 180° közötti értékeket vehet fel.

Ekkor tehát a két álláspont 2-2 mért koordinátája mellé jön a pontok mérőszalaggal mért távolsága és a kompasszal mért irányszög, összesen tehát 6 mérésünk van. Mivel ennyi mérésünk van, amit meg akarunk javítani, tehát a javítások vektora 6 elemű lesz. Ehhez 2 feltételi egyenletet írunk fel. A **B** mátrix tehát 2×6 méretű.

11.2. A feladat újabb módosítása: több álláspont közötti vonal

Módosítsuk most azzal a feladattal, hogy a 2. álláspont után tovább folytatjuk a mérést, és újabb álláspontok koordinátáit mérjük meg GPS segítségével, és egy-egy újabb állásponthoz mérőszalaggal távolságot, kompasszal irányszöget mérünk az előző állásponttól.

A kiegyenlítés során a mért mennyiségeket az alábbi sorrendbe rendezzük:

$$E_1, N_1, d_{12}, \varphi_{12}, E_2, N_2, d_{23}, \varphi_{23}, E_3, N_3, d_{34}, \varphi_{34}, \dots, E_N, N_N \quad (11.15)$$

Ez a felírási sorrend megkönnyíti a számítások tetszőlegesen hosszú adatsorra történő alkalmazását.

Minden két álláspont között 2 darab feltételi egyenletet tudunk felírni: egyet a távolságra, egyet pedig a szögre. Ez N álláspont esetén $2N-2$ darab feltételi egyenletet jelentett. A feltételi egyenletekbe behelyettesítve a mért koordinátá, távolság és szögértékeket számítjuk ki a javítások $\mathbf{v}$ vektorát. A feltételi egyenleteknek a – bennük szereplő mérések szerinti – deriváltjait (vagyis a **B** mátrixot) akár analitikusan, akár numerikusan is számíthatjuk.

A mérőszalaggal mért és a GPS mérésekből számolt távolságokra vonatkozó (11.1) feltételi egyenletnek analitikusan is megadhatók az egyes mért mennyiségek szerinti deriváltjai:

$$\begin{aligned} \frac{\partial l_{d_{12}}}{\partial E_1} &= \frac{-E_1}{\sqrt{(E_2 - E_1)^2 + (N_2 - N_1)^2}} \\ \frac{\partial l_{d_{12}}}{\partial E_2} &= \frac{E_2}{\sqrt{(E_2 - E_1)^2 + (N_2 - N_1)^2}} \\ \frac{\partial l_{d_{12}}}{\partial N_1} &= \frac{-N_1}{\sqrt{(E_2 - E_1)^2 + (N_2 - N_1)^2}} \\ \frac{\partial l_{d_{12}}}{\partial N_2} &= \frac{N_2}{\sqrt{(E_2 - E_1)^2 + (N_2 - N_1)^2}} \\ \frac{\partial l_{d_{12}}}{\partial d_{12}} &= -1 \end{aligned} \quad (11.16)$$

A kompasszal mért illetve a GPS-sel mért koordinátákból számolt azimutszögekre vonatkozó feltételi egyenlet (11.14) az egyes mérések szerinti parciális deriváltjai –felhasználva hogy a (11.14) egyenlet helyett $\text{atan2}()$ függvényt használunk –:

$$\begin{aligned}
 \frac{\partial l_{\phi_{12}}}{\partial E_1} &= \frac{E_1}{(E_2 - E_1)^2 + (N_2 - N_1)^2} \\
 \frac{\partial l_{\phi_{12}}}{\partial E_2} &= \frac{-E_2}{(E_2 - E_1)^2 + (N_2 - N_1)^2} \\
 \frac{\partial l_{\phi_{12}}}{\partial N_1} &= \frac{N_1}{(E_2 - E_1)^2 + (N_2 - N_1)^2} \\
 \frac{\partial l_{\phi_{12}}}{\partial N_2} &= \frac{-N_2}{(E_2 - E_1)^2 + (N_2 - N_1)^2} \\
 \frac{\partial l_{\phi_{12}}}{\partial \phi_{12}} &= -1
 \end{aligned} \tag{11.17}$$

A **B** mátrix alakja:

$$\mathbf{B} = \begin{bmatrix}
 \frac{\partial l_{d_{12}}}{\partial E_1} & \frac{\partial l_{d_{12}}}{\partial N_1} & \frac{\partial l_{d_{12}}}{\partial d_{12}} & 0 & \frac{\partial l_{d_{12}}}{\partial E_2} & \frac{\partial l_{d_{12}}}{\partial N_2} & 0 & 0 & 0 & 0 & \dots \\
 \frac{\partial l_{\phi_{12}}}{\partial E_1} & \frac{\partial l_{\phi_{12}}}{\partial N_1} & 0 & \frac{\partial l_{\phi_{12}}}{\partial \phi_{12}} & \frac{\partial l_{\phi_{12}}}{\partial E_2} & \frac{\partial l_{\phi_{12}}}{\partial N_2} & 0 & 0 & 0 & 0 & \dots \\
 0 & 0 & 0 & 0 & \frac{\partial l_{d_{23}}}{\partial E_2} & \frac{\partial l_{d_{23}}}{\partial N_2} & \frac{\partial l_{d_{23}}}{\partial d_{23}} & 0 & \frac{\partial l_{d_{23}}}{\partial E_3} & \frac{\partial l_{d_{23}}}{\partial N_3} & \dots \\
 0 & 0 & 0 & 0 & \frac{\partial l_{\phi_{23}}}{\partial E_2} & \frac{\partial l_{\phi_{23}}}{\partial N_2} & 0 & \frac{\partial l_{\phi_{23}}}{\partial \phi_{23}} & \frac{\partial l_{\phi_{23}}}{\partial E_3} & \frac{\partial l_{\phi_{23}}}{\partial N_3} & \dots \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
 \end{bmatrix} \tag{11.18}$$

A (11.4) egyenlet mutatja a $\mathbf{P}_{LL}$ mátrix alakját. Aktuális mérete (N álláspont esetén): $(4N-2) \times (4N-2)$. A kizárólag diagonális elemeket tartalmazó mátrix inverze az elemek reciproka.

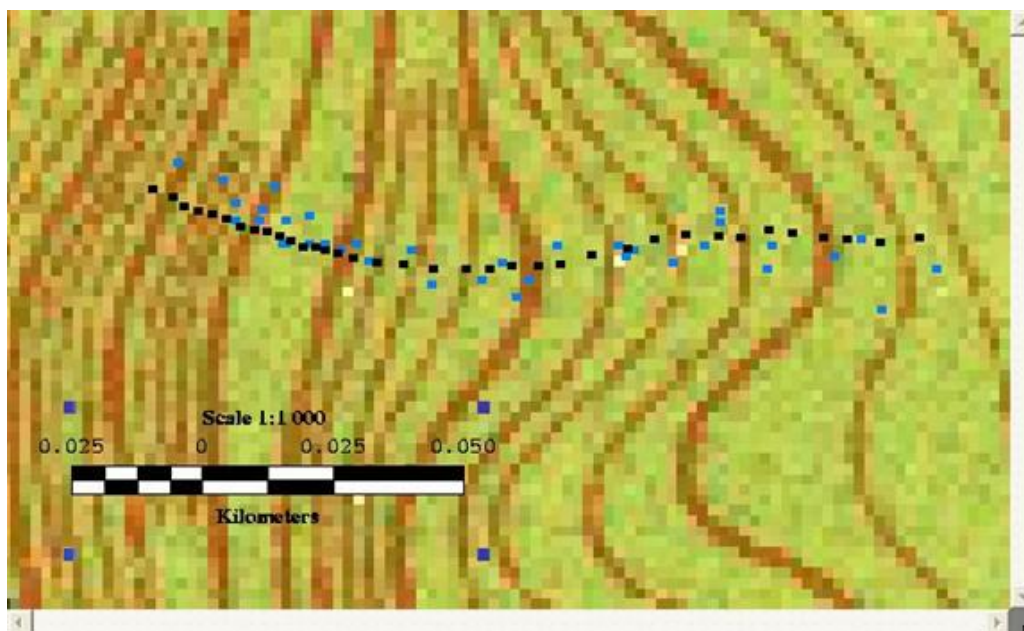
Az így kiszámított mennyiségeket az (11.13) képletbe beírva számíthatjuk a **k** korreláta vektort, majd ennek értékét behelyettesítve az (11.12) képletbe kapjuk a javítások $\mathbf{v}$ vektorát.

11.3. Eredmények

A hagyományos geodéziai mérések kiegyenlítése során a fent bemutatott kiegyenlítési eljárás általában kielégítő pontossággal megadja a javított (kiegyenlített) értékeket, mivel a méréseink eleve közel vannak a kiegyenlített értékekhez, és a (11.7) képletben alkalmazott lineáris összefüggés érvényesnek tekinthető.

A tapasztalatok szerint az eljárás egyszeri alkalmazása egy valódi adatsoron, nem hozza meg a kívánt eredményt. A kapott javításokkal korrigálva a méréseket, egyetlen iterációs lépésben csak egyetlen mérési pont koordinátái javulnak számottevően. Emiatt az eljárás ismétlésével, megközelítőleg a mérési pontok számával azonos számú iterációs lépést kell alkalmaznunk. Az utolsó lépések során, – mint az a nemlineáris kiegyenlítések során megszokott – az eredmény nem javul számottevően, és ilyenkor megtaláltuk a minimalizálandó függvény minimumhelyét. A kiinduló és a kiegyenlített koordináták térképi ábrázolása az 11.1 ábrán látható.

Az eredményeket vizsgálva elmondható, hogy a kapott nyomvonal – a térképi pontosság határain belül – illeszkedik a topográfiai térkép nyomvonalára. A kiegyenlítési folyamat során a GPS mérések gondoskodnak a nyomvonal **globális** elhelyezéséről, a távolság és szögmérések pedig biztosítják az álláspontok **lokálisan** jó pozícióját.



11.1 ábra: A Piliscsév település határában húzódó völgyben végzett szintezés álláspontjainak GPS-sel mért (kék) és kiegyenlített (fekete) síkkordinátái az EOTR térképszelvényen megjelenítve.

12. fejezet - Csillapított (regularizált) legkisebb négyzetes becslés

Vizsgáljuk meg az alábbi kiegyenlítési feladatot, amelyben mérésekkel azonos számú paraméter értékét akarjuk meghatározni a legkisebb négyzetek módszer segítségével.

A Balatonon egycsatornás vízi szeizmikus mérést végzünk több időpontban. A különböző időpontokban mért mérések nyomvonalai több helyen elmetszik egymást. Az egymást metsző szelvényeken egy jól azonosítható réteghatár mélysége nem egyezik meg a különböző szelvényeken a szelvények metszéspontjában, aminek a különböző időpontokban eltérő vízállás az oka. Korrigáljuk a 3 szelvényt egy-egy, szelvényenként konstans magasságtólással, hogy a szelvények metszéspontjában a korrigált mélységek lehetőleg megegyezzenek!

Szelvények metszéspontja	Szelvény száma	Magasság	Szelvény száma	Magasság
I	1	101,57	2	101,83
II	1	100,23	3	100,39
III	2	99,82	3	99,71

12.1. táblázat. Három szelvény, összesen három metszéspontjában az egyes szelvénytárcsák azonosítható szintfelületek magasságai

Határozzuk meg először a **direkt feladatot** és a **javítási egyenletet**!

A három meghatározandó paraméter az egyes szelvények magassági korrekció értékei $\mathbf{x}=(x_1, x_2, x_3)$. Ezeket a korrekciókat hozzáadva az adott szelvényen mért magasságokhoz, akkor a korrigált magasságokat kapjuk.

A korrigált magasságokra írhatjuk fel a feltételi egyenleteket: A szelvények metszéspontjában a korrigált magasságok eltérése minimális kell legyen.

$$\begin{aligned}(101,57 + x_1) - (101,83 + x_2) &= l_1 \\ (100,23 + x_1) - (100,39 + x_3) &= l_2 \\ (99,82 + x_2) - (99,71 + x_3) &= l_3\end{aligned}\tag{12.1}$$

A feltételi egyenletekbe behelyettesítve a paraméterek kiinduló értékeit (nullákat) az egyenlet jobb oldalán áll a tisztatag vektort (**I**) kapjuk (amelynek elemei: $l_1 = -0,26$ $l_2 = -0,16$ $l_3 = 0,11$) Elemezzük a tisztatag vektort! Az első mennyiség azt fejezi ki, hogy a 1. szelvény és az 2. szelvény között mennyi a magasságkülönbség, a második tag az 1. és 3. szelvény, a harmadik tag pedig a 2. és 3. szelvények közötti magasságkülönbséget fejezi ki. Ideális esetben a három tag összege nulla lenne, vagyis az egyes szelvényekhez lehet találni olyan eltolást, hogy azok hatására a magasságkülönbségek eltűnnek. Ezzel ellentétben látni fogjuk, hogy közel 1 cm-es, maradék ellentmondást nem lehet feloldani. A mérési és leolvasási hibák miatt nem meglepő, hogy a kiegyenlítés ellenére is marad ellentmondás a magasságokban.

Készítsük el az **A** alakmátrixot, a feltételi egyenleteknek a paraméterek szerinti deriváltját!

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}\tag{12.2}$$

Próbáljuk meg meghatározni a paraméterek **x** vektorát a (8.29) egyenlet segítségével!

$$\mathbf{x} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{l})\tag{12.3}$$

A problémát az jelenti, hogy az $\mathbf{A}^T \cdot \mathbf{A}$ mátrix determinánsa nulla, így nem invertálható.

Próbáljuk meg elemi megfontolások alapján meghatározni, hogy mi lehet a probléma. Tételezzük fel, hogy ha az első szelvényhez 0,1 méter, a második szelvényhez 0,2 méter, a harmadikhoz 0,3 méter korrekciót adnánk hozzá, és ezeket a korrekciókat a (12.1) egyenletbe beírva kiszámolhatnánk a tisztatag vektor. Ha a szelvényeket ehelyett rendre 1,1 1,2 és 1,3 méterrel korrigálnánk, akkor ugyanazokat a tisztatag vektor elemeket kapnánk az egyenletek alakjából következően. Ez azt jelenti, hogy több különböző paraméterkombináció esetén is a (12.1) feltételi egyenlet jobb oldalán ugyanazok a mennyiségek szerepelnek, ami egyenértékű azzal, hogy a fentebbi (12.1) egyenletnek nincs egyértelmű megoldása.

Az adott probléma esetén két lehetséges kezelési mód rögtön kínálkozik.

Az **első lehetőség**, ha rögzítjük az első szelvény magasságát, vagyis $x_1 = 0$.

Innentől megint két irányba mehetünk tovább: vagy átfogalmazzuk a feladatot, hogy csak 2 paramétert becslünk, vagy pedig kényszerfeltételként előírjuk, hogy $x_1 = 0$, ekkor a „Kényszerfeltételek alkalmazása” c. fejezetben leírt utat követjük.

A **második lehetőség**, hogy valamilyen **további feltételt** vezetünk be, például, hogy a paraméterekre előírjuk, hogy a négyzetösszegük legyen minimális. (Ezt általánosságban nevezhetjük a **paraméterek normájára** vonatkozó feltételnek.) Ez a konkrét példa esetén azt jelenti, hogy mind a három szelvényhez meghatározunk korrekció értékeket, úgy, hogy ezeknek a négyzetösszege a lehető legkisebb legyen. Ekkor az

$$\mathbf{v}^T \mathbf{v} + \mathbf{x}^T \mathbf{x} := \min \quad (12.4)$$

feltételnek kell teljesülni.

A gyakorlatban ilyen esetekben lehetséges, hogy fontosabbnak találjuk a kis reziduálértékeket elfogadva a paraméterek nagyobb javításértékeit (és ezzel együtt nagyobb változékonyságukat) de az is lehet, hogy nagyobb reziduálértékeket is elfogadunk a paraméterek kis (vagy kevésbé változékony) javításértékeiért cserébe.

Ezt a két véglet közötti kompromisszumot egy γ **csillapítási tényező** (trade-off paraméter) bevezetésével érhetjük el. Ekkor a felírható feltétel:

$$\mathbf{v}^T \mathbf{v} + \gamma \cdot \mathbf{x}^T \mathbf{x} := \min \quad (12.5)$$

Ebből a feltételből vezethetjük le (6.17-6.18) egyenletekhez hasonlóan, hogy a paraméterek az alábbi módon számíthatók:

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A} + \gamma \cdot \mathbf{E})^{-1} \cdot (\mathbf{A}^T \mathbf{I}) \quad (12.6)$$

Oldjuk meg az egyenletet $\gamma = 1$ esetben, és vizsgáljuk meg a kapott megoldást!

Ekkor a paraméterek javításai:

$$\mathbf{x}_{\gamma=1} = \begin{bmatrix} 0,1200 \\ -0,0925 \\ -0,0375 \end{bmatrix} \quad (12.7)$$

A paraméterek javításaival kifejezett eltolások hatására a metszéspontok magassága az egyes szelvényeken:

	magasság	magasság
I	101,675	101,738

II	100,335	100,378
III	99,728	99,698

12.2. táblázat. A három metszéspontban az egyes szelvénypárokon azonosítható szintfelületek korrigált magasságai $\gamma=1$ esetén.

A metszéspontokban a maradékhibák:

$$\mathbf{v}_{\gamma=1} = \begin{bmatrix} -0,0625 \\ -0,0425 \\ 0,0300 \end{bmatrix} \quad (12.8)$$

Vagyis mind a paraméterek javításai, mind a maradékhibák a centiméteres nagyságrendbe esnek.

Ha gamma értékét nagyra választom (pl. $\gamma = 10000$), akkor a paraméterek javításai:

$$\mathbf{x}_{\gamma=10000} = \begin{bmatrix} -4,197 \cdot 10^{-5} \\ 3,697 \cdot 10^{-5} \\ 4,997 \cdot 10^{-6} \end{bmatrix} \quad (12.9)$$

vagyis a milliméter tört része. Ennek következtében, a korrigált magasságok is csak kicsit módosulnak a kiinduló magasságadatokhoz képest:

	magasság	magasság
I	101,570	101,830
II	100,230	100,390
III	99,820	99,710

12.3. táblázat. A három metszéspontban az egyes szelvénypárokon azonosítható szintfelületek korrigált magasságai $\gamma=10000$ esetén.

Ennek hatására a maradékhibák kevéssé térnek el a kiinduló tisztagvektor értékeitől:

$$\mathbf{v}_{\gamma=10000} = \begin{bmatrix} -0,260 \\ -0,160 \\ 0,110 \end{bmatrix} \quad (12.10)$$

Ha gamma értékét kicsinek választom (pl. $\gamma = 0.00001$), akkor a paraméterek javításai:

$$\mathbf{x}_{\gamma=0.00001} = \begin{bmatrix} -0,140 \\ 0,123 \\ 0,017 \end{bmatrix} \quad (12.11)$$

Aminék a hatására a korrigált magasságok:

	magasság	magasság
I	101,710	101,707
II	100,370	100,373
III	99,697	99,693

12.4. táblázat. A három metszéspontban az egyes szelvénypárokon azonosítható szintfelületek korrigált magasságai $\gamma=0,00001$ esetén.

Jól látszik, hogy a korrigált magasságok csak néhány milliméterrel térnek el egymástól, amit a feltételi egyenletekből számított reziduál is kifejez:

$$\mathbf{v}_{\gamma=0,00001} = \begin{bmatrix} 0,0033 \\ -0,0033 \\ 0,0033 \end{bmatrix} \quad (12.12)$$

A reziduálértékeket vizsgálva jól látszik, hogy a közel centiméteres hiba, ami nem feloldható ellentmondásként van jelen az adatokban, a három feltételi egyenlet között egyenletesen oszlik el.

Ez a tapasztalat összhangban van tehát a feladat elején megfogalmazott elvárásainkkal, vagyis hogy találunk majd olyan megoldást, ahol a feltételi egyenletekkel kifejezett maradékhibák kicsik. Ez annak a következménye, hogy kis γ értéket választottunk, emiatt a reziduálokra vonatkozó feltételt nagy súllyal, a paraméterek normájára vonatkozó feltételt pedig kis súllyal vettük figyelembe.

Az itt bemutatott példánál az volt a feltételünk, hogy a modellparaméterek javításvektorának elemeiből képzett négyzetösszeg a minimális legyen, úgy, hogy a modellparaméterek kiinduló értékei nullák voltak.

A gyakorlatban előfordulnak olyan esetek, amikor azt a feltételt tesszük, hogy a modellparaméterek egy előzetesen megadott $\mathbf{x}^{(0)}$ paramétervektortól való eltérése legyen minimális. Ilyen eset fordul elő **multielektrodás szelvények inverziójánál**, ahol a a szelvényre merőlegesen elhelyezkedő, rögzített geometriájú félvégtelen hasábok fajlagos ellenállásai a modellparaméterek. A modellparaméterek száma összemérhető a mérések számával, ezért az itt bemutatott regularizált megoldást alkalmazzák, úgy, hogy a hasábok vezetőképességének kiinduló értékéül a mért látszólagos fajlagos ellenállás értékeket választják.

A regularizált megoldást nagy sikerrel alkalmazták **szeizmikus tomográfiai problémák** esetén. A tomográfiai problémák esetében a rugalmas hullámokkal átjárt térrész rácsfelbontásával nyert cellák szeizmikus sebessége a meghatározandó paraméterek együttese, a mérések pedig az ismert keltési és észlelési hely közötti futásidők. A korlátozott mérésszám miatt lehetséges, hogy vannak olyan cellák, amelyeken nem halad át hullám, így nem befolyásolják egyetlen hullám terjedési idejét sem. További probléma, hogy olyan mérési geometria is lehetséges, amelyben egy cellán áthaladó hullám mindig áthalad a mellette levő cellán is, így nincs lehetőség a két sebesség külön-külön meghatározására. A probléma kezelésére alkalmazott regularizált megoldás azokra a cellákra, amelyeken nem halad át hullám, a területre vonatkozó átlagos sebességet rendeli, a nem elkülöníthető cella párokhoz viszont az adott útszakaszhoz tartozó átlagos sebességet.

13. fejezet - Osztályozás

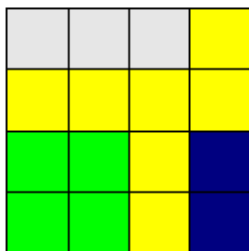
Az osztályozás azon eljárások összefoglaló neve, amely nagyszámú, többcsatornás (ugyanarra az objektumra vonatkozó több független) mérési adatból olyan csoportokat képez, amelyre igaz, hogy az ugyanazon csoportba tartozó adatok „hasznosak” egymáshoz, a különböző csoportba tartozó adatok pedig „különböznek” egymástól.

Az osztályozási eljárásokat az **inverzió statisztikus elmélete** alapján tárgyalhatjuk. Az inverzió statisztikus elméletében az osztályozást **minőségi interpretációnak** nevezik, és azt fejezi ki, hogy a mérési adatok alapján, az objektum lehetséges állapotai között döntünk. Ez a műholdkép osztályozása esetén annak felel meg, hogy egy képpontot csatornánkénti intenzitás (fényesség, radiancia, reflektancia vagy digitális jelszint –CN–) értékei alapján valamilyen felszíntípusként határozzuk meg.

A **hasznosság** definiálásához felhasználhatjuk a korábban megismert távolság definíciók (normák) valamelyikét, vagy a valószínűségek definícióját.

13.1. Egyszerű példa

Tekintsünk egy három csatornás, valódi színes műholdképet! Egy ilyen felvétel készítése során minden egyes képponthoz megméri, hogy mennyi sugárzás érkezik a kék, a zöld és a piros hullámhossztartományban. A műholdképünk legyen 4x4 képpont (pixel) méretű! A képen négy „felszíntípus” legyen látható: a kép bal felső sarkát egy felhő foglalja el, a kép közepén homok látszódjon, a kép jobb alsó sarkában egy mélyvízű tó, a kép bal alsó sarkában pedig egy erdő látszódjon. A kép valahogy így néz ki:



13.1. ábra. Egy 4x4 képpontból álló „műholdkép”. A bal felső sarokban felhő, a bal alsó sarokban erdő, a jobb alsó sarokban egy tó, a kép közepén homokos talaj látszik.

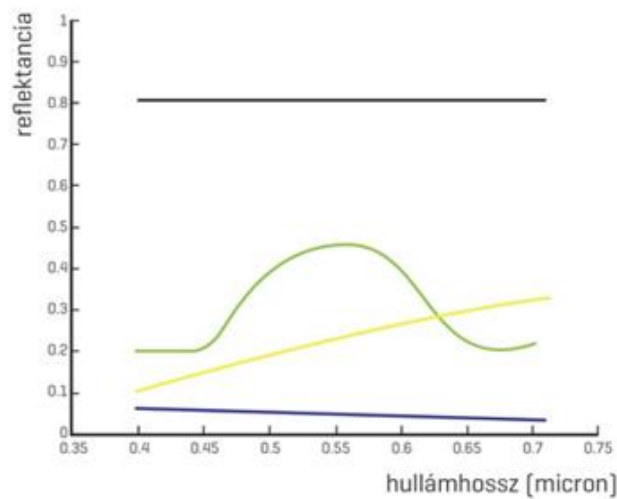
A kép egyes képpontjain látjuk, hogy mit ábrázolnak, ezért ebben az esetben szemre el tudjuk végezni az osztályozást. Az alábbi ábrán az osztályozás eredménye látható

f	f	f	h
h	h	h	h
e	e	h	v
e	e	h	v

13.2. ábra. A 13.1. ábrán látható műholdkép osztályozásának eredménye. Az egyes képpontok különböző osztályba sorolódtak.

A képen az 'f' a felhőt, a 'h' a homokot, az 'e' az erdőt, a 'v' a vizet jelöli. Egy színes képen a felhő fehérnek, a homok sárgának, az erdő zöldnek, a víz pedig feketébe hajló sötétkék színűnek látszik.

Ezek a hétköznapi tapasztalatok összhangban vannak a felszínek visszaverő képességének (reflektanciájának) hullámhossz szerinti menetével. A következő ábrán látjuk a képen látható felszínek reflektanciáját a látható hullámhossztartományban.



13.3. ábra. A víz (kék), homokos talaj (sárga), vegetáció (zöld) és a felhőzet (fekete) reflektanciája a hullámhossz függvényében a látható tartományban. A 0,4-0,5 μm a kék, a 0,5-0,6 μm a zöld, 0,6-0,7 μm a piros színnek felel meg.

Készítsünk egy táblázatot, amely az egyes felszíntípusoknak az egyes csatornákon jellemző reflektanciáját tartalmazza! A reflektanciaértékeket skálázzuk át, úgy hogy a 0-10% reflektanciáértékhez a 0 nulla, a 10-20% reflektanciához 1, stb. értékek tartozzanak. Ekkor a táblázat az alábbi alakot ölti:

	Kék 0,4-0,5 μm	Zöld 0,5-0,6 μm	Piros 0,6-0,7 μm
Felhő	8	8	8
Erdő	2	4	2
Homok	1	2	3
Víz	0	0	0

13.1. táblázat. A táblázatban a függőleges oszlopokban az egyes csatornákon (színeknek megfeleltethető hullámhosszakon), az egyes felszíntípusokhoz rendelhető átskálázott elméleti intenzitásértékek találhatók.

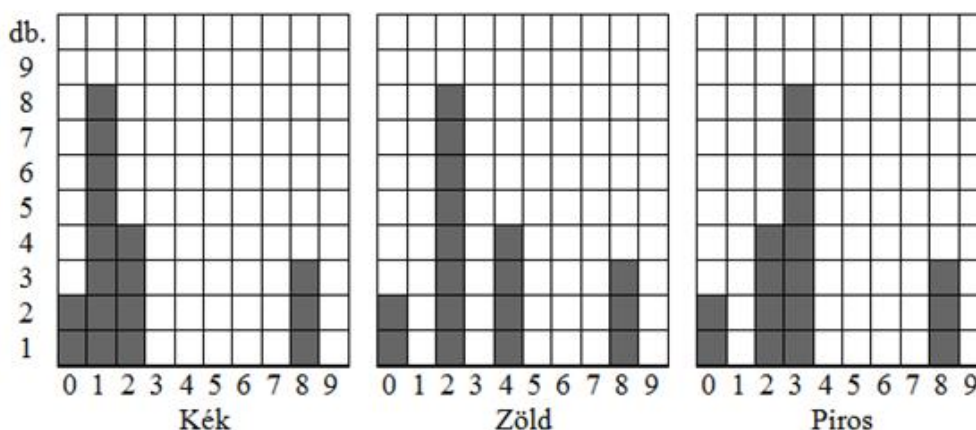
A táblázat alapján szintetikususan is elő tudjuk állítani, hogy mekkora reflektancia értékeket mérnénk az egyes csatornákon.

8	8	8	1	8	8	8	2	8	8	8	3
1	1	1	1	2	2	2	2	3	3	3	3
2	2	1	0	4	4	2	0	2	2	3	0
2	2	1	0	4	4	2	0	2	2	3	0
Kék				Zöld				Piros			

13.4. ábra. A „műholdkép” egyes csatornái, a felszín típusoknak megfelelő átskálázott intenzitásértékekkel.

Hogyan tudnánk egy egyszerű eljárást mondani, ami az egyes pixeleket csoportosítani tudná az intenzitásértékek alapján? Hogyan lehetne algoritmizálni azt az eljárást, amit mi szemmel meg tudunk csinálni, hogy egy képpontra nézve el tudjuk dönteni a színe alapján, hogy az melyik felszín típusba (osztályba) tartozik?

Ehhez végezzük el az adatok (képpontok) statisztikai elemzését! Készítsük el az egyes csatornák **hisztogramját**. A hisztogram – a valószínűség számítását összefoglaló fejezetben leírt definíciónak megfelelően – egy gyakorisági diagram, amely megmondja, hogy egy adatsokaságban a mérési értékek egy-egy meghatározott tartományába hány mért érték esik. A hisztogram vízszintes tengelyén a lehetséges intenzitás értékek, a függőleges tengelyen pedig az adott intenzitásértékű képpontok darabszáma van feltüntetve. Ezen definíció alapján az egyes csatornákra számított hisztogramok a következők:

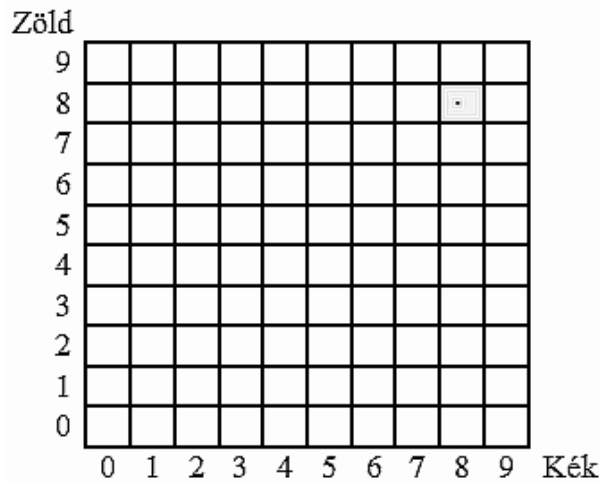


13.5. ábra. Az egyes csatornák hisztogramjai.

A hisztogramok vizsgálatával látszik, hogy a Zöld csatornán válnak el legjobban egymásból az egyes felszín típusok, itt ugyanis annyi csúcs van a hisztogramon, ahány osztályunk van. A zöld csatorna alapján **küszöbértékek** megadásával egyszerűen el tudnánk határolni az egyes felszín típusokat: Az „1” reflektanciaértéknél kisebb reflektanciaértékű képpontok a víz, az „1” és „3” reflektanciaérték közötti pontok a talaj, „3” és „6” közötti értékek a növényzet és „6”-nál nagyobb értékek a felhők. A növényzet és a felhőzet elhatárolására itt a két csúcs (növényzet a „4”-nél és felhőzet a „8”-nál) közötti „távolság” feléhez húztuk a határt.

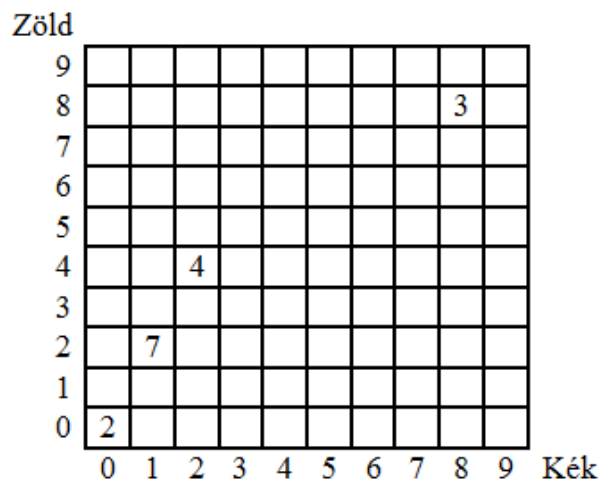
Vizsgáljuk meg, hogy milyen lehetőségünk van arra, hogy a többi csatornában levő információt is felhasználjuk az osztályozáshoz!

Ehhez készítsük el két-két csatorna reflektanciaértékeinek együttes gyakorisági diagramját a **szkattergramot**! A szkattergram két tengelyére felvesszük az egyes csatornákon előforduló intenzitás (reflektancia) értékeket. A szkattergramot úgy készítjük el, hogy végigmegyünk a kiválasztott két csatorna képén a bal felső saroktól kezdődően, és kiolvassuk az egyes képponthoz tartozó intenzitás érték párt. Ez esetünkben a Kék és a Zöld csatorna esetén, a bal felső képpontra: (8; 8). A szkattergram (8;8) cellájában (a példában egy ponttal) jelezzük, hogy találtunk a képen egy ilyen intenzitás érték párral rendelkező pontot.



13.6. ábra. A Kék és a Zöld csatorna adatainak hisztogramja. A hisztogramba még csak a műholdkép bal felső képpontja került be a (8,8) intenzitásértékekkel jellemzett mezőbe.

A műholdképen végighaladva, a szkattegramba bejelöljük minden képpont intenzitás érték párt. Miután ezzel elkészültünk a szkattegram egyes celláiban a pontokat megszámlálva kapjuk a Kék és Zöld csatorna szkattegrammját:

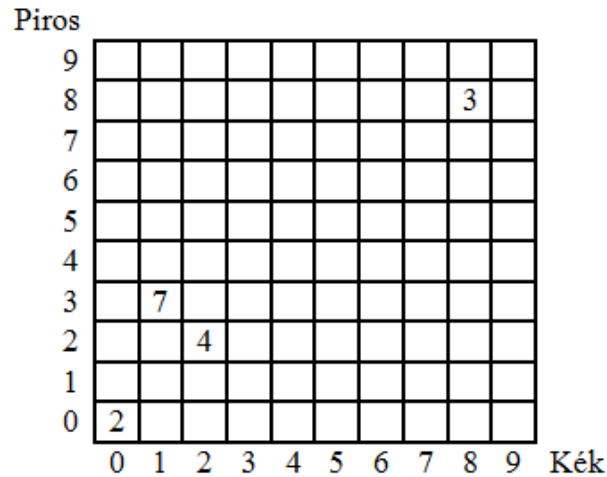


13.7. ábra. A Kék és a Zöld csatorna szkattegramja.

Az elkészült szkattegramba beírt számok összege megadja kép képpontjainak számát ($2+7+4+3=16$). A fenti módszer segítségével el tudjuk készíteni, a Kék-Piros és a Zöld-Piros csatornák szkattegramjait is.

A szkattegram segítségével elő tudjuk állítani a szkattegram elkészítéséhez felhasznált csatornák hisztogramját. A szkattegramban szereplő számok oszloponkénti összegzésével, és az oszloponkénti összegeket a vízszintes tengelyen levő intenzitásértékek rendelve megkapjuk a Kék csatorna hisztogramját. A szkattegramban szereplő gyakoriságok soronkénti összegzésével, és az összegeknek a függőleges tengelyen felvett intenzitásértékekhez rendelésével kaphatjuk a Zöld csatorna hisztogramját. A művelet nem fordítható meg: a hisztogramokból nem állítható elő a szkattegram.

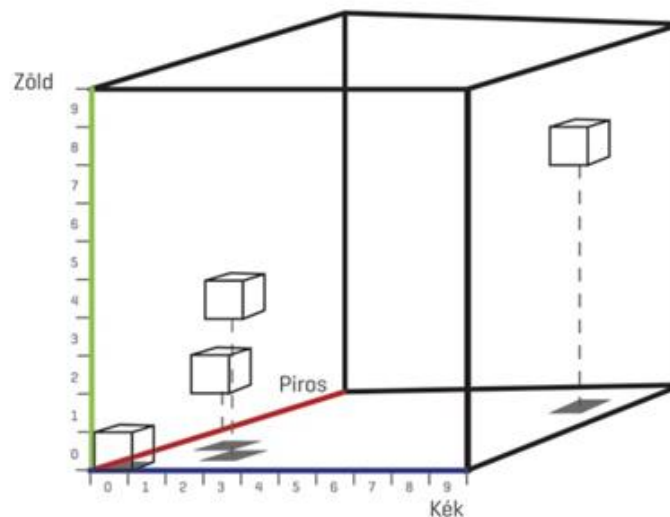
Az elkészített Kék-Zöld szkattegramon látszik, hogy a négy osztály jól elkülöníthető egymástól. Vizsgáljuk meg, hogy el tudnánk-e különíteni egymástól a felszínosztályokat, ha nem a Kék-Zöld, hanem a Kék-Piros csatornák hisztogramját vizsgálnánk!



13.8. ábra. A Kék és a Piros csatornák szattergramja.

A Kék-Piros csatornák hisztogramján látszik, hogy a (2,2) intenzitásértékkel jellemzett növényzet és az (1,3) intenzitásértékű homokos talaj egymáshoz „közelebb” van, mint a Kék-Zöld hisztogram esetében. A tapasztalat szerint a különböző hisztogramok segítségével eltérő pontossággal tudjuk elkülöníteni a felszíntípusokat.

Vizsgáljuk meg, hogy hogyan tudnánk valamennyi csatorna adatát felhasználni egy gyakoriság diagram elkészítéséhez! Az eddig eredményeket általánosíthatjuk, úgy, hogy tovább növeljük a gyakoriság diagramon együttesen feldolgozott csatornák számát, ezzel növelve a diagram **dimenzióját**. Általánosítva tehát a diagram a **csatornák absztrakt terét** mutatja. Három dimenzió (három csatorna) estén még tudjuk az eredményeinket grafikusán is ábrázolni:



13.9. ábra. A csatornák absztrakt tere, benne a négy felszínelem intenzitásértékeiből képzett 3-dimenziós vektorokhoz tartozó cellák jelölésével.

Az ábrán látszik, hogy a négy felszíntípust jól el tudjuk különíteni egymástól. Mivel valódi 3-dimenziós adatrendszer nem tudunk 2-dimenzióban ábrázolni, a „Piros” tengelyen elfoglalt helyzet érzékeltetésére az egyes cellákat levetítettük a Kék-Zöld csatornaadatok síkjára.

Az eddigi eredményeket további csatornák adataira is általánosíthatjuk. A műholdfelvételeink ugyanis nem csak a látható tartományban, hanem a közeli és távoli infravörös hullámhosszoknál is készülhetnek, így egyetlen képponthoz több (<10 , multispektrális) vagy akár nagyszámú (>20 , hiperspektrális) független csatorna tartozik.

13.2. Valódi műholdkép vizsgálata

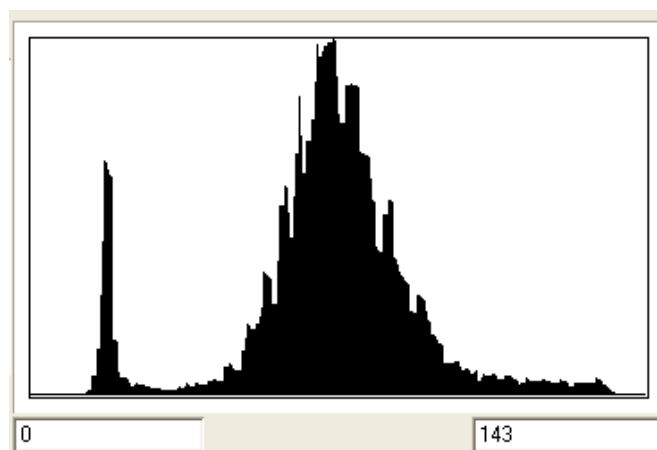
Valódi műholdkép vizsgálata során felhasználjuk az előbbi egyszerűsített műholdkép vizsgálatával szerzett ismereteinket.



13.10. ábra. Landsat TM műholdkép 1,2 és 3. csatornájából készített valódi színes kép részlete

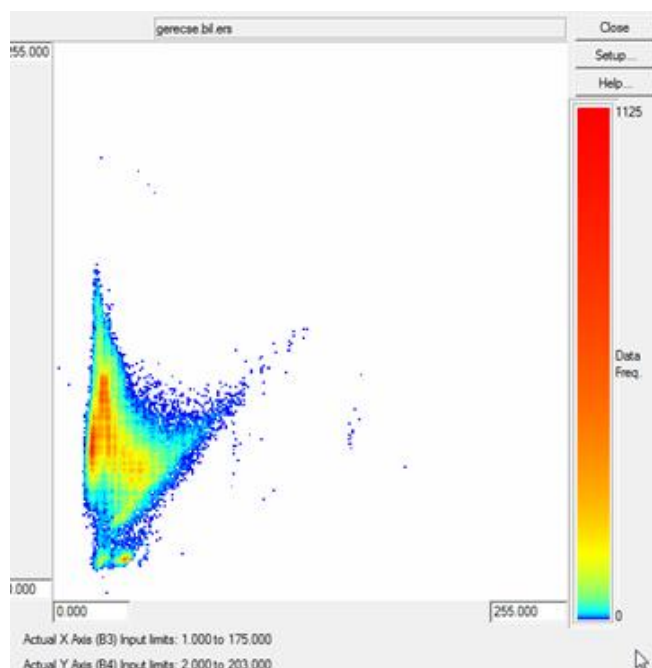


13.11. ábra. Landsat TM műholdkép részlete a 3. csatorna intenzitásértékeinek szürke skálás megjelenítésével.



13.12. ábra. A 13.11 ábra adataiból készített hisztogram

Az (13.12) ábra a (13.11) ábrán megjelenített műholdkép csatorna hisztogramját ábrázolja. A műholdképen látszódik víz, betakarított mezőgazdasági terület és erdő. Azonban a hisztogramon csak a víz jelentkezik önálló csúcsként, a különböző mezőgazdasági területek és az erdők a hisztogram alapján nem különíthetők el egymástól.



13.13. ábra. A Landsat TM kép 3. és 4. csatornáiból képzett szkattergram.

A valódi műholdkép szkattergramján már nem különülnek el jól látható módon az osztályok. A bemutatott szkattergrammon az egyes intenzitásértékekhez tartozó gyakoriságokat színekkel ábrázoljuk. A piros szín a nagy gyakoriságot jelenti. A szkattergramon több ilyen sűrűsödési pontot találunk, ezek az egyes felszíntípusok jellemző pontjai. A szkattergram alsó felén látszó két kis sűrűsödési pontot a tavak és a folyók képpontjai alkotják. A felt baloldalán látszó több lokális csúcsot tartalmazó piros terület az erdők és a mezőgazdasági területek képe. Ennek a feltnek a 45°-os egyeneshez elnyúló része az előbb említett felszínnek feletti légköri pára, felhőzet és füst hatásának tulajdonítható.

A műholdképen – felhasználva a térinformatikai rendszer vektorizáló (rajzoló) funkcióját – rajzoljunk körül **egy olyan felszíndarabot** (objektumot) amelyről biztosan **tudjuk, hogy melyik felszíntípushoz tartozik**. Egy körülrajzolt, sokszögvonallal határolt felszínelemet **poligon**nak nevezzük.

A térinformatikai rendszerrel készítettünk el az **ebbe a poligonba eső képpontok statisztikai vizsgálatát**. A program kiszámolja a pontok **számát**, az egyes csatornákon az előforduló **minimális és maximális értékeket**, az egyes csatornákon az **átlagot**, és az egyes csatornákon a **tapasztalati szórást**, és valamennyi csatorna-párra a **tapasztalati kovariancia mátrixot**. Egy körülrajzolt felszínarabot, melynek neve (egyedi azonosítója) a felszín típusra utal, **tanulóterületnek** nevezzük.

Egy felszín típushoz tartozó tanulóterület statisztikai vizsgálatából nyert adatokat vizualizálhatunk is. Elkészítve a tanulóterület poligonjába eső képpontok csatornánkénti hisztogramját, általában egy egy csúcsú hisztogramot kapunk. A tanulóterület pontjaiból kiválasztott két csatorna szkatogramját elkészítve látjuk, hogy a sűrűsödési pont körül ellipszis alakban helyezkednek el az adott tanulóterülethez tartozó pontok, a sűrűsödési pont körül nagyobb számban, attól távolodva egyre csökkenő számban.

A hisztogram és a szkatogram az 1- és 2-dimenziós **tapasztalati valószínűség-sűrűségfüggvényeknek** felelnek meg. A hisztogram és szkatogram segítségével meg tudjuk becsülni egy **eloszlás paramétereit**. Vizsgáljuk meg részletesen, hogy minek az eloszlását tudjuk meghatározni a hisztogramból és a szkatogramból!

Nézzünk egy példát! Egy három csatornás műholdképen kijelölt 10 képpontból álló tanulóterület csatornánkénti értékei láthatók a 13.2 táblázatban.

	1. csatorna $x^{(1)}$	2. csatorna $x^{(2)}$	3. csatorna $x^{(3)}$	$x^{(1)} - \bar{x}^{(1)}$	$x^{(2)} - \bar{x}^{(2)}$	$x^{(3)} - \bar{x}^{(3)}$
	12,32	28,97	14,79	-0,7348	-0,0756	-3,7796
	12,36	28,09	15,62	-0,6923	-0,9529	-2,9482
	13,52	29,91	22,32	0,4669	0,8661	3,7450
	15,61	29,55	22,26	2,5520	0,5034	3,6851
	13,05	28,65	18,40	-0,0088	-0,3919	-0,1738
	12,47	28,51	19,06	-0,5822	-0,5307	0,4929
	14,45	30,45	19,42	1,3934	1,4086	0,8433
	11,96	28,92	19,20	-1,0937	-0,1203	0,6231
	13,36	29,64	18,85	0,3097	0,5937	0,2750
	11,44	27,74	15,81	-1,6101	-1,3004	-2,7628
átlag	13,06	29,05	18,57			
négyzetösszeg				13,9150	6,3945	59,6631

13.2. táblázat. Egy 3 csatornás műholdképen kijelölt, 10 képpontból álló tanulóterület képpont-adatai, csatornánkénti átlagai, és az átlagtól való eltérései.

A táblázatban a csatornánkénti adatok alatt szerepelnek a csatornánkénti **átlagértékek** is:

$$\begin{bmatrix} \bar{x}^{(1)} \\ \bar{x}^{(2)} \\ \bar{x}^{(3)} \end{bmatrix} = \begin{bmatrix} 13,06 \\ 29,05 \\ 18,57 \end{bmatrix} \quad (13.1)$$

Az átlagok és az átlagoktól való eltérések segítségével ki tudjuk számolni a **mérések kovarianciamátrixát**, (R_v) amit kis mérésszám esetén a (3.57) képlettel adott korrigált tapasztalati kovarianciamátrixszal helyettesítünk:

$$c_{ij}^* = \frac{1}{n-1} \sum_{i=1}^n (x_i^{(i)} - \bar{x}^{(i)}) \cdot (x_i^{(j)} - \bar{x}^{(j)}) \quad (13.2)$$

A táblázatban szereplő, csatornánkénti átlagtól való eltérésekből a (13.2) képlet segítségével számolt kovariancia mátrix:

$$\mathbf{R}_v = \begin{bmatrix} 1,54 & 0,78 & 2,30 \\ 0,78 & 0,71 & 1,43 \\ 2,30 & 1,43 & 6,63 \end{bmatrix} \quad (13.3)$$

Az osztályozási eljárások során annyi tanulóterületet rajzolunk körül a műholdfelvételünkön, ahány osztályba a képpontjainkat sorolni szeretnénk, külön poligonba rajzolva az erdőt, vizet, felhőt stb.

Ezeknek minden tanulóterületnek a fent bemutatott módon külön-külön kiszámoljuk a statisztikai jellemzőiket.

A következőkben bemutatjuk azokat a különféle eljárásokat, amelyek segítségével a tanulóterületekre kiszámolt statisztikai jellemzők felhasználásával a képpontjainkat osztályokba sorolhatjuk. Ezeket a módszereket – mivel az egyes osztályokat a tanulóterületek megadásával mi magunk definiáltuk – összefoglalóan **irányított osztályozásnak** (*supervised classification*) nevezzük.

13.3. Maximum likelihood osztályozás

Azzal hogy egy általunk ismert felszíntípusú terület képpontjainak a statisztikáját készítettük el, az erre a területre eső képpontokból készült hisztogram azt jellemzi, hogy mi a valószínűsége annak, hogy egy képpont pont egy adott intenzitásértéket vesz fel, ha a felszín egy adott típusú. Ez pontosan az a **feltételes valószínűség**, amit a normális hibaeloszlású, lineáris additív terű modelleknél az inverzió statisztikai megközelítését tárgyaló fejezetben már felírtunk (5.8):

$$f(\mathbf{u} | v) = \prod_{k=1}^K \left(\frac{1}{(2\pi)^{\frac{1}{2}(2I_k+1)} \cdot \sqrt{\det(\mathbf{R}_{vk})}} \right) \cdot e^{-\frac{1}{2} \sum_{k=1}^K ((\mathbf{u}_k - \mathbf{f}_k)^T \cdot \mathbf{R}_{vk}^{-1} \cdot (\mathbf{u}_k - \mathbf{f}_k))} \quad (5.8)$$

A fenti képletet egyetlen **mérési csoportra** vonatkoztatva kapjuk:

$$f(\mathbf{u} | v) = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot \sqrt{\det(\mathbf{R}_v)}} \cdot e^{-\frac{1}{2} ((\mathbf{u} - \mathbf{f}_v)^T \cdot \mathbf{R}_v^{-1} \cdot (\mathbf{u} - \mathbf{f}_v))} \quad (13.4)$$

A fenti képletben n a dimenzió, vagyis a csatornák száma, az $\mathbf{u}$ vektor pedig a mérésünk, vagyis egy képpont csatornánkénti intenzitásértékeiből képzett vektor. Az $\mathbf{f}_v$ vektor, az adott v -edik osztályhoz tartozó „elméleti” értékekből képzett vektor. Ez nem más, mint az adott felszíntípus tanulóterület poligonjába eső pontok **csatornánkénti átlagai** (13.1.). A kettő különbsége a hibavektor, amelyről feltételeztük, hogy normális eloszlású.

A $\mathbf{R}_v$ mátrix az adott osztályt jellemző **kovarianciamátrix**, amit a tanulóterület poligonjába eső képpontokból a (13.2) képlet segítségével számíthatunk.

Ez a feltételes valószínűség tehát megadja, hogy egy adott csatornánkénti intenzitásértékekkel jellemzett képpont mekkora valószínűséggel tartozik az adott osztályba. A csatornák absztrakt terében vizsgálva, az osztályközéppont közelében levő pontok nagy valószínűséggel, az attól távolabbi pontok kisebb valószínűséggel tartoznak az adott osztályhoz. Az azonos valószínűségeket **n -dimenziós ellipszoid felületek** jellemzik. (Két csatorna esetén ezek ellipszisek). Az ellipszoidfelület középpontjában van az **osztályközéppont**. Az ellipszoid csak akkor kanonikus helyzetű, ha a kovarianciamátrix inverze diagonális, különben az ellipszoid tengelyei szöget zárnak be a koordinátatengelyekkel. A kisebb valószínűségekhez nagyobb tengelyekkel jellemezhető ellipszoid felületek tartoznak, amelyek koncentrikusak, és tengelyirányaik megegyeznek. Ha valamelyik csatornán nagyon nagy szórása van az adatoknak, akkor annak a tengelynek az irányába elnyúlik az ellipszis. Ha csatornánként azonos a szórás, és a kovariancia mátrix nem diagonális elemek kicsik, akkor az ellipszoid egy **n -dimenziós gömbbé** fajul.

A fenti **feltételes valószínűség** (13.4) segítségével tehát ki tudjuk számolni, hogy egy adott képpont ebbe az osztályba tartozik (v), akkor mennyi a valószínűsége, hogy az egyes csatornákon pont a mért intenzitásértéket mérjük (u).

Ahogy korábban láttuk, a Bayes-tétel alapján ezzel a feltételes valószínűség segítségével ki tudjuk fejezni, hogy mi a valószínűsége annak, hogyha az adott képpont esetén ezeket az intenzitásértékeket mértük a csatornákon, akkor a képpont éppen ebbe az osztályba tartozik:

$$P(v|u) = \frac{P(v) \cdot f(u|v)}{f(u)} \quad (13.5)$$

A képlet szerint annak a valószínűsége, hogy adott intenzitásértékekkel jellemzett képpont az adott osztályba tartozik, az adott osztályba tartozás előzetes (*a priori*) valószínűségének és az adott osztály esetén a mért értékek megvalósulását leíró feltételes valószínűségekre szorzata.

(Korábban említettük, hogy a nevezőben szereplő valószínűség, vagyis hogy egy mérés során pont a mérési eredményeket mérjük, minden v osztály feltételes valószínűségében szerepel, ezért ezt egységnyiinek választhatjuk.)

A $P(v)$ valószínűség az előzetes ismereteinket fejezi ki. A műholdkép esetén, ilyen előzetes ismeret, ha tudjuk, hogy az adott osztály a kép által lefedett terület hányad részét foglalja el. Például ha tudjuk, hogy a terület 50%-a erdő, 30%-a talaj, és 20%-a víz, akkor az ezekhez az osztályokhoz tartozó a priori valószínűségek rendre: 0,50 0,30 0,20.

A **maximum likelihood osztályozás** során a tanulótérületeinkre elvégzett statisztikai számításból meghatározzuk **minden osztályhoz** az osztályközepek f_v vektorát és az R_v kovariancia mátrixot.

Ezután minden képpontra kiszámoljuk, hogy mekkora a valószínűség, hogy az adott képpont az egyes osztályokba tartozzon. Ezt, egy adott osztályra az adott osztály előzetes valószínűsége és a (13.4) képlettel definiált feltételes valószínűség szorzataként állítjuk elő. Az adott képpontra megvizsgáljuk, hogy melyik az az osztály, amelyikhez a legnagyobb valószínűséget számoltunk, és a képpontot abba az osztályba soroljuk.

Vizsgáljuk meg, hogy mekkora a valószínűsége, hogy a (13, 28, 18) intenzitásértékekkel jellemzett pont, a 13.2 táblázatban található pontokkal jellemzett osztályhoz tartozik, ha tudjuk, hogy az adott osztály előzetes (*a priori*) valószínűsége 0,4.

A feltételes valószínűség (13.4) képletébe behelyettesítve a mért pont és az osztályközep pont távolságvektorára:

$$u - f_v = \begin{bmatrix} 13 \\ 28 \\ 18 \end{bmatrix} - \begin{bmatrix} 13,06 \\ 29,05 \\ 18,57 \end{bmatrix} = \begin{bmatrix} -0,06 \\ -1,05 \\ -0,57 \end{bmatrix} \quad (13.6)$$

Szükségünk van az R_v mátrix (13.3) determinánsára, ami:

$$\det(R_v) = 1,43 \quad (13.7)$$

Az R_v mátrix inverze:

$$R_v^{-1} = \begin{bmatrix} 1,6 & -1,35 & -0,35 \\ -1,35 & 3,46 & -0,27 \\ -0,35 & -0,27 & 0,33 \end{bmatrix} \quad (13.8)$$

Ezeket a mennyiségeket, valamint a dimenziószámot ($n = 3$) visszahelyettesítve a feltételes valószínűség képletébe, kapjuk:

$$f(\mathbf{u} | \nu) = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot \sqrt{\det(\mathbf{R}_\nu)}} \cdot e^{-\frac{1}{2}(\mathbf{u}-\mathbf{f}_\nu)^T \mathbf{R}_\nu^{-1} (\mathbf{u}-\mathbf{f}_\nu)} = 0,0097 \quad (13.9)$$

A fenti eredmény és az előzetes valószínűség segítségével tehát:

$$P(\nu | \mathbf{u}) = P(\nu) \cdot f(\mathbf{u} | \nu) = 0,4 \cdot 0,0097 = 0,0039 \quad (13.10)$$

A gyakorlati alkalmazás során az így kapott valószínűséget kell összehasonlítani az adott képpont (13, 28, 18) más osztályokra vonatkozó valószínűségével, és abba az osztályba besorolni, amelyik osztályba tartozás valószínűsége a maximális.

13.4. Mahalanobis osztályozás

A feltételes valószínűség (13.4) képletében, az exponenciális függvény argumentumában az alábbi mennyiség szerepel:

$$d^{Mahalanobis} = \left[(\mathbf{u} - \mathbf{f}_\nu)^T \cdot \mathbf{R}^{-1} \cdot (\mathbf{u} - \mathbf{f}_\nu) \right]^{\frac{1}{2}} \quad (13.11)$$

Ezt a mennyiséget egy irányok szerint súlyozott távolságnak, a **Mahalanobis-féle távolságnak** nevezzük. Ennek a távolságdefiníciónak a segítségével is meg lehet megfogalmazni egy osztályozási algoritmust:

A **Mahalanobis osztályozás** során, a képpontot abba az osztályba soroljuk, amelyik osztály osztályközéppontjától ($\mathbf{f}_\nu$) a fenti képlettel (13.11) definiált távolsága a legkisebb.

Ez az osztályozás – hasonlóan a maximum likelihood osztályozáshoz – figyelembe veszi az egyes osztályok alakját csatornák absztrakt terében. A távolságdefiníció következtében egy osztályközépponttól egyenlő távolságra levő pontok itt is egy ellipszoidfelszínre esnek.

Ez az osztályozás – eltérően a maximum likelihood osztályozástól – nem veszi figyelembe sem az egyes osztályok előzetes valószínűségeit, sem pedig az egyes osztályok jellemző kiterjedését a csatornák absztrakt terében, amit a

mennyiség fejez ki.

13.5. Minimális távolság (minimum distance) osztályozás

Egy intenzitásértékeinek vektorával ($\mathbf{u}$) jellemezhető képpont (Euklédesszi, L_2 norma szerinti) távolságát egy adott $\mathbf{f}_\nu$ osztályközépponttól a

$$d^{Euklédesszi} = \left[(\mathbf{u} - \mathbf{f}_\nu)^T \cdot (\mathbf{u} - \mathbf{f}_\nu) \right]^{\frac{1}{2}} \quad (13.12)$$

távolsággal jellemzünk.

A **minimális távolság osztályozás** során egy képpontot abba az osztályba sorolunk be, amelyik osztály középpontjához a fenti (13.12) távolságdefiníció szerint a legközelebb esik.

Két csatorna és két osztály esetén a szemléletes határ, amely elválasztja egymástól a két osztályt, a két osztály középpontját összekötő szakasz felezőmerőlegese. 3- dimenzióban a határfelületek az osztályközéppontokat összekötő szakaszok felezőjében a szakaszra merőleges síkok.

A minimum távolság osztályozás előnye a Maximum likelihood és a Mahalanobis osztályozással szemben, hogy nem igényli a tanulóterületek statisztikaszámítása során a kovariancia mátrixoknak a számítását, hanem elegendő az osztályközéppontok számítása, valamint a képpontok osztályozása során kevesebb szorzási műveletet igényel.

13.6. Paralelepipedon osztályozás

Láttuk, hogy a csatornák absztrakt terében az egyes felszíntípusok jellemző térrészeket foglalnak el. A paralelepipedon osztályozás esetén, a tanulóterületek statisztikai vizsgálata során a térinformatikai program kikeresi az adott felszíntípushoz, valamennyi csatornára a minimum és maximum értékeket. Ezek az értékek egy felszíntípusokhoz egy csatorna esetén az egyes felszíntípusokhoz egy intervallumot, két csatorna esetén egy téglalapot, három csatorna esetén egy téglatestet jelölnek ki. (Általánosítva n -dimenziós, nem derékszögű koordináta-rendszerre ezeket az n -dimenziós paralellogrammák által határolt idomokat paralelepipedonnak nevezzük.) Az osztályozás során a kép egy pontját abba az osztályba soroljuk, amelyik osztály paralelepipedonjába esik.

14. fejezet - Összefoglalás

A „Lineáris algebra összefoglaló” c. fejezetben összefoglaltuk a mátrixszámítást, és megismertedtünk a mátrixok általánosított inverzével, amit a „Lineáris inverzió” és a „Csillapított (regularizált) legkisebb négyzetes becslés” c. fejezetben alkalmazunk.

„A valószínűségszámítás alapjai” c. fejezetben bevezetjük a mintavétel fogalmát, és bemutatjuk a maximum likelihood becslési eljárást, amit az „A geofizikai inverzió statisztikai megközelítése” c. fejezetben alkalmazunk.

„A geofizikai inverzió általános megfogalmazása” c. fejezetben néhány geofizikai példa segítségével megismertedünk a mérés, mérési hiba, modell, paraméter és direkt feladat fogalmakkal.

„A geofizikai inverzió statisztikai megközelítése” c. fejezetben abból a feltevésből kiindulva, hogy a méréseinket normális eloszlású hibák terhelik, a maximum likelihood elv felhasználásával levezetjük a minket érdeklő modellparamétereket becslő eljárást. Más hibaeloszlásokra példaként bemutatunk más becslési eljárásokat, amelyek mindegyike a mért, és a modellből számolt mennyiségek – valamilyen norma szerinti – eltérésének minimalizálását jelentik. A megoldáshoz felhasználható minimumkereső eljárásokat is (jobb híján) ebben a fejezetben mutatjuk be. A becslések jóságát a becsült modellparaméterek szórásai és kovarianciái írják le: ezek kiszámítási módjával zárul a fejezet.

A „Lineáris inverzió” c. fejezetben egy nagyon tanulságos formalizmust mutatunk be, ami az inverziót, mint egyszerű algebrai problémát tekinti, és a mátrixok általános inverze segítségével oldja meg.

„Az inverzió megoldása a legkisebb négyzetek módszer segítségével” c. fejezetben, majd az következő fejezetekben a Gauss által kitalált, és számos probléma során sikerrel alkalmazott módszert vizsgáljuk egyszerű példák segítségével. Bemutatjuk a módszer különféle változatait, és azok alkalmazási korlátait.

Az „Osztályozás” c. fejezetben a statisztikus elméletben tárgyalt fogalmak segítségével mutatjuk be a mérések statisztikai vizsgálatán alapuló csoportokra bontási eljárást.

15. fejezet - Irodalomjegyzék

- Prékopa, András. *Valószínűségelmélet műszaki alkalmazásokkal*. Műszaki Könyvkiadó, Budapest. 1962.
- Cserepes, László. *Numerikus matematika geofizikus hallgatóknak (kézirat)*. Tankönyvkiadó, Budapest. 1989.
- Detrekői, Ákos. *Kiegyenlítő számítások*. Tankönyvkiadó, Budapest. 1991.
- Menke, William. *Geophysical data analysis: discrete inverse theory*. Academic Press, San Diego. 1989.
- Tarantola, Albert. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM Society for Industrial and Applied Mathematics, Philadelphia. 2005.
- Weber, Zoltán. *Geofizikai Inverzió (Kézirat)*. 2006 Budapest.
- Helmert, F. R.. *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate mit Anwendungen auf die Geodäsie, die Physik und die Theorie der Messinstrumente*. Verlag von B. G. Teubener, Leipzig. 1924.
- Salát, Péter, Tarcsai, György, Cserepes, László, Vermes, Mátyás, és Drahos, Dezső. *A geofizikai interpretáció információs-statisztikus módszerei*. Tankönyvkiadó, Budapest. 1982.
- Steiner, Ferenc. *A geostatisztika alapjai*. Tankönyvkiadó, Budapest. 1990.
- Wiener, N.. *Extrapolation, interpolation and smoothing of stationary time series*. J. Wiley, New York. 1949.
- Scales, J., A., Smith, M., L., és Treitel, S.. *Introductory Geophysical Inverse Theory*. Samizdat Press, Golden – White River Junction. 2001.

16. fejezet - Köszönetnyilvánítás

A jegyzet megírásának ürügyén szeretném kifejezni köszönetemet valamennyi egykori tanáromnak, de szeretném kiemelni Salát Pétert, aki geofizikus nemzedékekben tudatosította az inverzió fontosságát. Hálával tartozom én is Cserepes Lászlónak †, aki a numerikus módszerekbe bevezetett és Drahos Dezsőnek, akinek a szakdolgozati témavezetésével először kerültem egy jól körüljárható inverziós probléma közvetlen közelébe.

Köszönettel tartozom Györffy Jánosnak, aki felhívta a figyelmemet a polinomegyütthatók meghatározására, ezzel a bonyolultabb minimumkeresési eljárásoktól visszatért a Gauss-féle legkisebb négyzetes becslésekhez.

Köszönetemet szeretném kifejezni Timár Gábornak és Székely Baláznak, akikkel együtt dolgozva számos gyakorlati problémán kamatoztathattam a témában elsajátított elméleti ismereteimet.

Hálával tartozom Norbert Pfeifernek a Bécsi Műszaki Egyetemen, aki újabb feladatok kitűzésével lendített tovább, valamint kollegáinak Camillo Resslernek, és Helmut Kagernek, akik a feladatok megoldása felé vezető rögzös úton segítettek. Ebben nagy segítségemre volt Detrekői Ákos könyve (Detrekői, 1990).

Külön köszönöm Balázs László kollégámnak a segítő megjegyzéseket.