



Πανεπιστήμιο Κρήτης

Τμήμα Μαθηματικών

Διακριτά Αντίστροφα Προβλήματα

Σημειώσεις του Μαθήματος

βασισμένες κυρίως στο Βιβλίο :

Geophysical Data Analysis : Discrete Inverse Theory

του William Menke

Μιχάλης Ταρουδάκης

Καθηγητής

Εαρινό εξάμηνο 2012-2013

Πρόλογος

Το παρόν τεύχος των σημειώσεων έχει ως στόχο να βοηθήσει τους φοιτητές του μαθήματος «Διακριτά Αντίστροφα Προβλήματα», να το παρακολουθήσουν με τον καλύτερο δυνατό τρόπο. Βασίζονται στο Βιβλίο: Geophysical Data Analysis : Discrete Inverse Theory, του William Menke το οποίο αποτελεί και το κύριο διδακτικό βιβλίο του μαθήματος.

Οι σημειώσεις στην παρούσα φάση είναι πρόχειρες και περιληπτικές. Θα συμπληρωθούν στο μέλλον για να αποτελέσουν αυτοτελές εγχειρίδιο για τους φοιτητές που ενδιαφέρονται να παρακολουθήσουν το εν λόγω μάθημα.

Για την προσωρινή και περιληπτική τους μορφή ζητώ την κατανόηση των φοιτητών.

Ηράκλειο, Φεβρουάριος 2013.

Περιεχόμενα

1.	ΕΙΣΑΓΩΓΗ	9
1.1	Ευθύ και Αντίστροφο Πρόβλημα	9
1.2	Διατύπωση ενός αντίστροφου προβλήματος.....	10
1.3	Παραδείγματα διατύπωσης απλών αντιστρόφων προβλημάτων.....	12
1.3.1	Υπολογίζοντας παραμέτρους μίας ευθείας.	12
1.3.2	Υπολογίζοντας παραμέτρους μιας παραβολής.	12
1.3.3	Μη καταστροφικός έλεγχος υλικών με ακουστικά κύματα.....	13
1.3.4	Ένα απλό πρόβλημα αξονικής τομογραφίας	14
1.4	Λύσεις αντιστρόφων προβλημάτων.	17
2.	ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ	18
2.1	Κατανομή πιθανότητας	18
2.2	Συσχέτιση δεδομένων	19
2.3	Συναρτήσεις τυχαίων μεταβλητών	21
2.4	Κανονικές κατανομές	23
2.5	Διαστήματα εμπιστοσύνης	24
3.	ΕΠΙΛΥΣΗ ΓΡΑΜΜΙΚΟΥ ΑΝΤΙΣΤΡΟΦΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΜΕ ΤΗ ΜΕΘΟΔΟ ΜΗΚΩΝ.....	25
3.1	Διαφορά μετρήσεων από εκτιμήσεις μετρήσεων.	25
3.2	Νόρμες διανυσμάτων.....	26
3.3	Η νόρμα της διαφοράς μετρήσεων και εκτιμήσεων.....	26
3.4	Η λύση ελαχίστων τετραγώνων για το γενικό γραμμικό αντίστροφο πρόβλημα.	28
3.4.1	Αναλυτικός υπολογισμός των παραμέτρων	28
3.4.2	Η λύση ελαχίστων τετραγώνων από την πλευρά της Γραμμικής Άλγεβρας... ..	29

3.5	Παραδείγματα επίλυσης γραμμικών αντιστρόφων προβλημάτων με την αρχή των ελαχίστων τετραγώνων.	30
3.5.1	Υπολογισμός ευθείας γραμμής	30
3.5.2	Υπολογισμός παραβολής	31
3.5.3	Υπολογισμός εξίσωσης επιπέδου στο χώρο.	32
3.6	Υπο-ορισμένα προβλήματα	34
3.7	Μεικτά ορισμένα προβλήματα	36
3.8	Λύσεις αντιστρόφων προβλημάτων με χρήση ζυγισμένων μετρικών μήκους.....	37
3.9	Άλλοι τύποι εκ προοιμίου (a-priori) πληροφορίας.	40
3.10	Πίνακας συνδιακύμανσης των παραμέτρων	42
4	ΓΕΝΙΚΕΥΜΕΝΟΙ ΑΝΤΙΣΤΡΟΦΟΙ ΠΙΝΑΚΕΣ	44
4.1	Πίνακας ανάλυσης δεδομένων	44
4.2	Πίνακας ανάλυσης παραμέτρων.....	46
4.3	Ο Μοναδιαίος πίνακας συνδιακύμανσης	47
4.4	Ανάλυση και συνδιακύμανση για ορισμένες περιπτώσεις γενικευμένων αντιστρόφων	48
4.5	Μετρήσεις ποιότητας της ανάλυσης και της συνδιακύμανσης.....	49
4.6	Μία εφαρμογή: Ανάκτηση της εξίσωσης ευθείας	49
5	ΜΕΘΟΔΟΙ ΜΕΓΙΣΤΗΣ ΠΙΘΑΝΟΦΑΝΕΙΑΣ	53
5.1	Η συνάρτηση μέγιστης πιθανοφάνειας	53
5.2	Εκτιμήσεις μέγιστης πιθανοφάνειας για το γραμμικό αντίστροφο πρόβλημα.....	55
5.3	Εκ προοιμίου κατανομές	55
5.4	Εκτιμήσεις μέγιστης πιθανοφάνειας για ακριβή θεωρία.	56
5.5	Εκτιμήσεις μέγιστης πιθανοφάνειας για μη ακριβή θεωρία.....	57
5.6	Η απλή περίπτωση με κανονική κατανομή και γραμμική θεωρία	58
5.7	Η γενική γραμμική περίπτωση με κανονικές κατανομές.....	59
5.8	Ειδικές περιπτώσεις	60
5.8.1	Ακριβή δεδομένα και ακριβές μοντέλο (θεωρία)	60

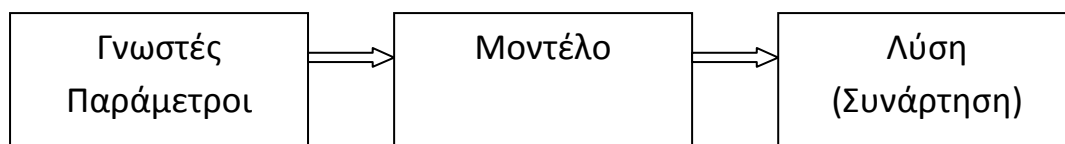
5.8.2	Μη ακριβές μοντέλο και μη ακριβή δεδομένα.....	60
5.8.3	Μη ακριβής αρχική πληροφορία για τις παραμέτρους.....	60
6	ΔΙΑΝΥΣΜΑΤΙΚΟΙ ΧΩΡΟΙ ΚΑΙ ΑΝΤΙΣΤΡΟΦΑ ΠΡΟΒΛΗΜΑΤΑ.....	61
6.1	Διανυσματικοί χώροι παραμέτρων και μετρήσεων.....	61
6.2	Γραμμικοί μετασχηματισμοί	63
6.2.1	Υπο-ορισμένο πρόβλημα.....	64
6.2.2.	Υπερ-ορισμένο πρόβλημα.....	65
6.2.3	Το μεικτά ορισμένο πρόβλημα.	66
6.3	Ανάλυση Ιδιαζουσών Τιμών (Singular Value Decomposition)	67
6.3.1	Η ανάλυση	67
6.3.2	Ανάλυση ιδιαζουσών τιμών και γενικευμένος αντίστροφος.....	68
7	ΜΗ ΓΡΑΜΜΙΚΑ ΑΝΤΙΣΤΡΟΦΑ ΠΡΟΒΛΗΜΑΤΑ	71
7.1	Παραμετροποίηση αντιστρόφων προβλημάτων	71
7.2	Μη γραμμικά αντίστροφα προβλήματα.....	74
7.2.1	Εισαγωγή	74
7.2.2	Μία επαναληπτική μέθοδος για την επίλυση του μη γραμμικού προβλήματος.....	75

1. ΕΙΣΑΓΩΓΗ

1.1 Ευθύ και Αντίστροφο Πρόβλημα

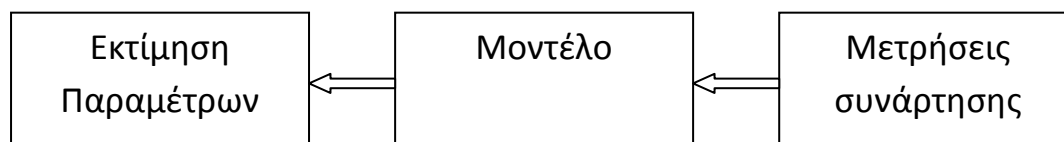
Στην μαθηματική προσομοίωση **ευθύ** ονομάζεται ένα πρόβλημα στο οποίο ζητείται ο προσδιορισμός ενός μεγέθους, συνήθως με τη μορφή μιας συνάρτησης, όταν είναι γνωστές τόσο οι παράμετροι που διέπουν το πρόβλημα, όσο και η εξίσωση ή οι εξισώσεις που περιγράφουν τις μεταβολές του μεγέθους. Στα πλαίσια της ντετερμινιστικής θεώρησης, τα προβλήματα αυτά είναι «Καλώς Τεθειμένα» (Well Posed) με την έννοια ότι θα πρέπει, όταν επιλυθούν, να δίνουν μοναδική λύση της άγνωστης συνάρτησης.

Στο επόμενο σχήμα περιγράφονται τα ευθέα προβλήματα. Με τον όρο «Μοντέλο» υπονοούμε την εξίσωση ή τις εξισώσεις που διέπουν το πρόβλημα, καθώς και τη διαδικασία επίλυσής τους.



Σχήμα 1.1 : Το ευθύ πρόβλημα.

Στο **αντίστροφο** πρόβλημα, ζητάμε συνήθως την εκτίμηση των παραμέτρων ενός φυσικού φαινομένου ή ενός μαθηματικού προβλήματος, όταν είναι γνωστές μετρήσεις μια συνάρτησης ή ενός κατάλληλου μεγέθους. Οι μετρήσεις και οι προς ανάκτηση παράμετροι συνδέονται με το ίδιο μοντέλο που ορίζεται στο ευθύ πρόβλημα, αλλά η διαδικασία επίλυσης ξεκινά αντίστροφα (Σχήμα 1.2). Τα αντίστροφα προβλήματα σπάνια είναι καλώς τεθειμένα, συνήθως είναι «Κακώς Τεθειμένα» (Ill Posed), και επιδέχονται από καμία έως πολλές λύσεις. Στις φυσικές επιστήμες γνωρίζουμε βέβαια ότι οι παράμετροι ενός φυσικού φαινομένου υπάρχουν, συνεπώς μέσω της επίλυσης ενός αντιστρόφου προβλήματος πρέπει να βρεθούν κατάλληλες εκτιμήσεις τους. Αυτό είναι και το αντικείμενο του μαθήματος.



Σχήμα 1.2 : Το αντίστροφο πρόβλημα.

1.2 Διατύπωση ενός αντίστροφου προβλήματος

Οι παράμετροι που πρέπει να υπολογιστούν σε ένα αντίστροφο πρόβλημα χαρακτηρίζονται ως «model parameters» και συμβολίζονται με το σύμβολο m . Αντίστοιχα, οι μετρήσεις χαρακτηρίζονται ως δεδομένα (“data”) και συμβολίζονται με το γράμμα d . Οι παράμετροι και τα δεδομένα μπορεί να είναι συνεχείς συναρτήσεις $m(x)$, $d(y)$ όπου με x, y συμβολίζουμε γενικά τις μεταβλητές από τις οποίες εξαρτώνται. Εάν μπορέσουμε να διαχωρίσουμε δεδομένα και παραμέτρους μέσω κάποιου «πυρήνα» $G(x, y)$ που θα μας υποδειχθεί από το Μοντέλο, μπορούμε να διατυπώσουμε ενδεχομένως μία ολοκληρωτική εξίσωση της μορφής :

$$d(y) = \int G(x, y)m(x)dx \quad (1.1)$$

η οποία επιλυόμενη ως προς $m(x)$ θα μας δώσει τις παραμέτρους.

Εάν τα δεδομένα (μετρήσεις) έχουν γίνει, όπως συνήθως, σε διακριτά σημεία ή χρονικές στιγμές, ορίζεται ένα διάνυσμα από τιμές $\mathbf{d} = [d_1, d_2, \dots, d_N]^T$ που αντιπροσωπεύει τις μετρήσεις. Στην περίπτωση αυτή η εξίσωση (1.1) θα γραφεί για κάθε διακριτή τιμή μετρήσεων και ορίζει αυτό που ονομάζουμε «Συνεχές Αντίστροφο Πρόβλημα»

$$d_i = d_i = \int G(x)m(x)dx, \quad i = 1, \dots, N \quad (1.2)$$

Τέλος μπορεί και οι παράμετροι να είναι διακριτές $\mathbf{m} = [m_1, m_2, \dots, m_M]^T$, οπότε έχουμε την γενική περίπτωση ενός **διακριτού αντίστροφου προβλήματος**.

Θα συνεχίσουμε θεωρώντας ότι δεδομένα και παράμετροι είναι διακριτά μεγέθη και θα δούμε πως αυτά συνδέονται μεταξύ τους.

Η γενική περίπτωση είναι να συνδέονται μέσω εξισώσεων της μορφής

$$f_j(\mathbf{d}, \mathbf{m}) = 0, \quad j = 1, \dots, L \quad (1.3)$$

που σε διανυσματική μορφή γράφονται ως $\mathbf{f}(\mathbf{d}, \mathbf{m}) = 0$. Οι ανωτέρω εκφράσεις μπορεί να είναι περίπλοκες. Στη συνέχεια πάντως θα αναφερθούμε σε ειδικές περιπτώσεις που μας δίνουν τη δυνατότητα αντιμετώπισης του αντίστροφου προβλήματος με λογικής μορφής δυσκολία.

- Έμμεση Γραμμική Μορφή

Εάν η συνάρτηση $\mathbf{f}$ είναι γραμμική ως προς δεδομένα και παραμέτρους, παίρνουμε μία έκφραση της μορφής :

$$\mathbf{f}(\mathbf{d}, \mathbf{m}) = 0 = \mathbf{F} \begin{bmatrix} \mathbf{d} \\ \mathbf{m} \end{bmatrix} \quad (1.4)$$

Όπου $\mathbf{F}$ είναι ένας πίνακας διαστάσεων $L \times (M+N)$.

- Άμεση Μορφή

Ορισμένες φορές, μπορούμε να διαχωρίσουμε τα δεδομένα από τις παραμέτρους και να διατυπώσουμε $L=N$ που είναι γραμμικές μεν ως προς τα δεδομένα, μη γραμμικές όμως ως προς τις παραμέτρους μέσω μιας διανυσματικής συνάρτησης $\mathbf{g}$.

$$\mathbf{f}(\mathbf{d}, \mathbf{m}) = 0 = \mathbf{d} - \mathbf{g}(\mathbf{m}) \quad (1.5)$$

Μπορούμε δηλαδή να γράψουμε :

$$d_i = g_i(\mathbf{m}), \quad i = 1, \dots, N \quad (1.6)$$

όπου η συνάρτηση $\mathbf{g}$ είναι εν γένει μη γραμμική. Η ανωτέρω περίπτωση συναντάται συχνότατα σε αντίστροφα προβλήματα από τις φυσικές επιστήμες, όπου η συνάρτηση $\mathbf{g}$ εξαρτάται εκτός από το φυσικό μοντέλο που διέπει τη σχέση μετρήσεων-παραμέτρων και από τις συνθήκες της μέτρησης

- Άμεση Γραμμική Μορφή

Εάν και η συνάρτηση $\mathbf{g}$ είναι γραμμική, ως προς τις παραμέτρους, παίρνουμε την απλούστερη μορφή για τα αντίστροφα προβλήματα που είναι :

$$\mathbf{f}(\mathbf{d}, \mathbf{m}) = 0 = \mathbf{d} - \mathbf{G}\mathbf{m} \quad (1.7)$$

όπου $L=N$ ξανά, όπως και στην προηγούμενη περίπτωση, αλλά η συνάρτηση $\mathbf{g}$ εκφράζεται μέσω του πίνακα $\mathbf{G}$ ο οποίος είναι διαστάσεων $N \times M$. Με άλλα λόγια, εκφράζουμε το αντίστροφο πρόβλημα ως ένα γραμμικό σύστημα N εξισώσεων με M αγνώστους :

$$d_i = \sum_{j=1}^M G_{ij} m_j, \quad i = 1, \dots, N \quad (1.8)$$

Για την επιλυσιμότητα του ανωτέρω προβλήματος θα μιλήσουμε σε άλλο σημείο του μαθήματος. Στο σημείο αυτό να επισημάνουμε ότι ο πίνακας $\mathbf{G}$ είναι γενικά παραλληλόγραμμος αφού δεν έχουμε κατ' ανάγκη ίσο αριθμό δεδομένων και παραμέτρων.

Αρκετά αντίστροφα προβλήματα μπορούν να αναχθούν σε προβλήματα της μορφής 1.8.

1.3 Παραδείγματα διατύπωσης απλών αντιστρόφων προβλημάτων.

1.3.1 Υπολογίζοντας παραμέτρους μίας ευθείας.

Ας υποθέσουμε ότι πραγματοποιούμε N μετρήσεις της θερμοκρασίας (Δεδομένα) σε διαφορετικά σημεία μίας ράβδου που γνωρίζουμε από την θερμοδυναμική ότι θα πρέπει να έχει θερμοκρασία γραμμικά μεταβαλλόμενη συναρτήσει του μήκους της (Μοντέλο). Ζητούμε να υπολογίσουμε τις παραμέτρους της ευθείας που παριστάνει τη θερμοκρασία T (παράμετροι) συναρτήσει του μήκους x στη ράβδο. Επομένως το διάνυσμα των δεδομένων θα είναι $\mathbf{d}=[T_1, T_2, \dots, T_N]^T$ και το διάνυσμα των παραμέτρων θα είναι $\mathbf{m}=[a, b]^T$ όπου a, b θα είναι οι παράμετροι της εξίσωσης της ευθείας $T = a + bx$ που αντιπροσωπεύει το μοντέλο μας.

Θα παρατηρήσει κανείς ότι για να χαράξουμε μία ευθεία χρειαζόμαστε μόνο δύο μετρήσεις. Επειδή όμως οι μετρήσεις σε ένα πραγματικό πρόβλημα δεν γίνονται με ακρίβεια, αλλά υπάρχουν λάθη που προέρχονται από διάφορες αιτίες, πραγματοποιούμε περισσότερες μετρήσεις. Έτσι καταστρώνουμε το επόμενο σύστημα εξισώσεων :

$$T_i = a + bx_i, \quad i = 1, \dots, N \quad (1.9)$$

που γράφεται διαφορετικά ως

$$\begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad (1.10)$$

Το πρόβλημα προφανώς (για $N > 2$) είναι «υπερορισμένο». Έχουμε περισσότερες εξισώσεις σε σχέση με τους αγνώστους.

1.3.2 Υπολογίζοντας παραμέτρους μιας παραβολής.

Ας υποθέσουμε τώρα ότι το φυσικό μοντέλο για το παραπάνω πρόβλημα υποδεικνύει μεταβολή της θερμοκρασίας με το μήκος που περιγράφεται ως καμπύλη 2^{ου} βαθμού. Τότε το μοντέλο απαιτεί εξίσωση της μορφής : $T = a + bx + cx^2$ και επομένως για τον ίδιο αριθμό μετρήσεων (δεδομένων) έχουμε τώρα να υπολογίσουμε ένα διάνυσμα τριών στοιχείων $\mathbf{m}=[a, b, c]^T$ και οι εξισώσεις μας γράφονται

$$T_i = a + bx_i + cx_i^2, \quad i = 1, \dots, N \quad (1.11)$$

και το σύστημα με τη μορφή πινάκων ως :

$$\begin{bmatrix} T_1 \\ T_2 \\ \cdot \\ \cdot \\ T_N \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_N & x_N^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}. \quad (1.12)$$

Και το πρόβλημα αυτό είναι υπερορισμένο για $N > 3$.

1.3.3 Μη καταστροφικός έλεγχος υλικών με ακουστικά κύματα.

Ένα απλό αντίστροφο πρόβλημα μηχανικής προέρχεται από την ανάγκη να υπολογιστούν ιδιότητες ενός υλικού χωρίς αυτό να σπάσει ή να ληφθούν δείγματα από τη δομή του. Αυτό μπορεί να γίνει με χρήση ήχων που διαπερνούν το υλικό και καταγράφονται στην έξοδό του. Επειδή μία χαρακτηριστική ιδιότητα του υλικού που εν πολλοίς καθορίζει τη σύνθεσή του είναι η ταχύτητα διάδοσης του ήχου, θα μπορούσε να υπολογιστεί για το υλικό το μέγεθος αυτό, και μέσω αυτού να καθοριστεί η ποιότητά του. Το πείραμα χαρακτηρίζεται ως πείραμα **ακουστικής τομογραφίας** και είναι βέβαια ορίζει ένα αντίστροφο πρόβλημα.. Από τη φυσική γνωρίζουμε ότι η ταχύτητα διάδοσης του ήχου c και η διανυθείσα απόσταση σε μέσες τιμές δίδονται από την απλή σχέση $c = h / t$ όπου h είναι η διανυθείσα απόσταση και t είναι ο χρόνος, η μέτρηση του χρόνου με γνωστό το μήκος διάδοσης μπορεί να μας δώσει την ταχύτητα (μοντέλο).

Θα θεωρήσουμε λοιπόν για το παράδειγμά μας ότι έχουμε να υπολογίσουμε τις ιδιότητες 16 κυβικών τούβλων που έχουν διάσταση πλευράς h το καθένα και τα διατάσσουμε σε τέσσερις ομάδες των τεσσάρων (Σχήμα 1.3). Σε κάθε οριζόντια γραμμή και κατακόρυφη στήλη κάνουμε μία μέτρηση ακουστικής διάδοσης ήχου στέλνοντας μία στενή δέσμη (που την περιγράφουμε ως ακτίνα) ήχου που διαδίδεται σε ευθεία γραμμή και μετρώντας το χρόνο που πέρασε. Θεωρώντας το αντίστροφο της ταχύτητας (slowness) $s = 1 / c$ και αποδίδοντας σε κάθε τούβλο το δείκτη i όπως στο σχήμα, οι 8 συνολικά μετρήσεις (4 οριζόντιες και 4 κατακόρυφες) μας δίνουν τις εξισώσεις που περιγράφονται συνοπτικά ως :

$$\begin{aligned} T_1 &= hs_1 + hs_2 + hs_3 + hs_4 \\ T_2 &= hs_5 + hs_6 + hs_7 + hs_8 \\ &\cdot \\ &\cdot \\ &\cdot \\ T_8 &= hs_4 + hs_8 + hs_{12} + hs_{16} \end{aligned} \quad (1.13)$$

και σε μορφή εξίσωσης πινάκων ως

$$\begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_8 \end{bmatrix} = h \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{16} \end{bmatrix} \quad (1.14)$$

Προσέξτε ότι τώρα το γραμμικό πρόβλημα είναι υπο-ορισμένο σε αντίθεση με τα προβλήματα των περιπτώσεων 1.3.1 και 1.3.2.



Σχήμα 1.3 Διάταξη πειράματος ακουστικής τομογραφίας. Ο πομπός S και ο δέκτης R διατάσσονται έτσι ώστε η ακουστική ακτίνα να σαρώνει μία γραμμή ή στήλη.

1.3.4 Ένα απλό πρόβλημα αξονικής τομογραφίας

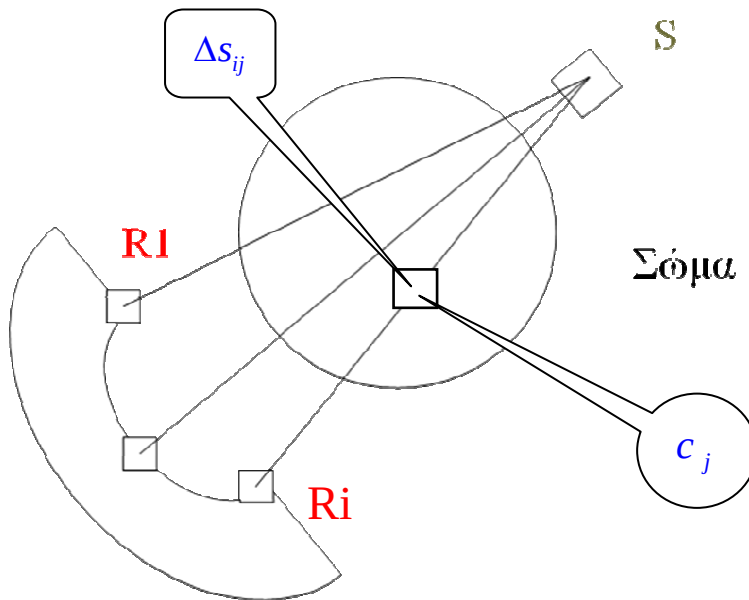
Ευρύτατη εφαρμογή έχουν τα αντίστροφα προβλήματα στην ιατρική διαγνωστική. Η αξονική ή μαγνητική τομογραφία βασίζουν τα αποτελέσματά τους στην επίλυση αντιστρόφων προβλημάτων από την ηλεκτρομαγνητική κυματική διάδοση και την θεωρία μαγνητικών πεδίων αντίστοιχα και βέβαια βοηθούνται αποτελεσματικά από την απεικόνιση και την επεξεργασία εικόνας. Ως παράδειγμα εδώ θα δούμε πως μπορεί κατ' αρχήν να διατυπωθεί ένα αντίστροφο πρόβλημα που σχετίζεται με την αξονική τομογραφία και πως αυτό μπορεί να γραμμικοποιηθεί.

Ο διθηνής όρος για την διαγνωστική τεχνική που βασίζεται στην χρήση ηλεκτρομαγνητικών κυμάτων που διαπερνούν ένα σώμα και μεταφέρουν πληροφορίες για τη σύνθεσή του είναι Computerized Axial Tomography (CAT). Το απλό μοντέλο στο οποίο βασίζεται συσχετίζει την μετρούμενη ένταση του ηλεκτρομαγνητικού πεδίου που παράγεται από ένα κύμα (X ray) γνωστής αρχικής έντασης όταν αυτό διαπεράσει ένα σώμα, με το συντελεστή απορρόφησης της ακτινοβολίας που είναι μία χαρακτηριστική ιδιότητα των ιστών του σώματος. Το μοντέλο βέβαια στην πράξη είναι πιο σύνθετο αλλά η απλοποιημένη του εκδοχή μπορεί να μας δώσει την ιδέα της διατύπωσης του αντίστροφου προβλήματος.

Σύμφωνα λοιπόν με τα παραπάνω, η μεταβολή της έντασης της ακτινοβολίας καθώς διαπερνά ένα σώμα είναι αντίστροφα ανάλογη της έντασης της ακτινοβολίας με το συντελεστή αναλογίας να αντιπροσωπεύει το συντελεστή απορρόφησης, μέσω του οποίου μοντελοποιείται το σώμα.

$$dI / ds = -c(x, y)I \quad (1.15)$$

Όπου I είναι η ένταση s είναι το στοιχειώδες μήκος διάδοσης σε ευθεία, $c(x, y)$ είναι ο συντελεστής απορρόφησης που υπολογίζεται ως συνάρτηση των χωρικών μεταβλητών σε ένα επίπεδο. Για να μπορέσει η ακτινολογία να δώσει μια εικόνα του τρισδιάστατου σώματος, η ακτινοβολήση γίνεται σε συνδυασμό πηγής και πολλών δεκτών που συνήθως βρίσκονται στην περιφέρεια ενός κύκλου με κέντρο την πηγή. Το ακτινοβολούμενο σώμα βρίσκεται ενδιάμεσα ενώ κάθε συνδυασμός πηγής και δέκτη ορίζει ένα επίπεδο στο οποίο απεικονίζεται η ενδεχόμενη ανομοιογένεια. Στο απλοποιημένο παράδειγμα που θα ακολουθήσει θα θεωρήσουμε ότι το επίπεδο εκφυλίζεται σε ευθεία και επομένως θα δεχτούμε ότι η εφαρμογή της αξονικής τομογραφίας γίνεται σε πολλές ευθείες που ορίζονται από δύο σημεία (πηγής και δέκτη) σύμφωνα με το σχήμα 1.4. Ωστόσο το σώμα θα θεωρηθεί ότι βρίσκεται σε ένα επίπεδο. Επομένως έχουμε περιορίσει τη διάσταση του πραγματικού προβλήματος κατά ένα.



Σχήμα 1.4 Σχηματική διάταξη αξονικής τομογραφίας. Η πηγή S στέλνει ηλεκτρομαγνητικά κύματα γνωστής αρχικής έντασης I_0 που σαρώνουν το σώμα και καταγράφονται στους δέκτες R_i . Για τη διακριτοποίηση-γραμμικοποίηση του προβλήματος το τετράγωνο στο σώμα αντιπροσωπεύει το στοιχείο j

Εάν λοιπόν η αρχική ένταση της δέσμης είναι I_0 , η εξίσωση 1.15 επιλυόμενη δίνει για κάθε δέσμη i μετρούμενη ισχύ πεδίου :

$$I_i = I_0 \exp\left(- \int_{beam} c(x, y) ds\right), \quad i = 1, \dots, N \quad (1.16)$$

Με βάση όσα έχουμε πει παραπάνω, το πρόβλημα θυμίζει αυτό που περιγράψαμε στην εξίσωση 1.2 και είναι ένα διακριτό ως προς τα δεδομένα, αλλά συνεχές ως προς τις παραμέτρους ($c(x, y)$) άμεσο αλλά μη γραμμικό πρόβλημα. Η επίλυσή του μπορεί να γίνει με τεχνικές που θα αναφερθούν σε ανάλογο κεφάλαιο των σημειώσεων, ωστόσο εδώ έχει σημασία να δούμε πως μπορούμε να το απλοποιήσουμε γραμμικοποιώντας το.

Αυτό μπορεί να γίνει εάν υποθέσουμε ότι ο συντελεστής απορρόφησης είναι μικρός και συνεπώς το ολοκλήρωμα στην 1.16 είναι επίσης μικρό.. Επίσης γνωρίζουμε ότι εκθετική συνάρτηση $\exp(-x)$ μπορεί να προσεγγιστεί με τους δύο πρώτους όρους του αναπτύγματος Taylor και να πάρουμε : $\exp(-x) \approx 1 - x$. Με βάση τα παραπάνω, η 1.16 μπορεί να γραφεί ως :

$$I_i = I_0 \left(1 - \int_{beam} c(x, y) ds\right), \quad i = 1, \dots, N \quad (1.17)$$

Έτσι παίρνουμε :

$$\Delta I_i = \frac{I_0 - I_i}{I_0} = \int_{beam} c(x, y) ds, \quad i = 1, \dots, N \quad (1.18)$$

Ένα ακόμη βήμα θα μας γραμμικοποιήσει πλήρως το πρόβλημα. Η διακριτοποίηση του σώματος σε τετραγωνικά στοιχεία σταθερού συντελεστή απορρόφησης c_j $j = 1, \dots, M$ το καθένα, και η προσέγγιση του ολοκληρώματος με άθροισμα (Αριθμητική Ανάλυση) εφ' όσον θεωρήσουμε ότι σε κάθε στοιχείο j και κάθε δέσμη i , το διανυόμενο μήκος είναι Δs_{ij} (Σχήμα 1.4). Η υιοθέτηση της αντιστοίχισης με δείκτες που ακολουθήσαμε μας επιτρέπει να γράψουμε την εξίσωση (1.18) με την προσέγγιση του ολοκληρώματος με άθροισμα ως :

$$\Delta I_i = \frac{I_0 - I_i}{I_0} = \sum_{j=1}^M \Delta s_{ij} c_j, \quad i = 1, \dots, N \quad (1.19)$$

Προσέξτε ότι η ανωτέρω διακριτοποίηση μας οδηγεί σε ένα πρόβλημα ανάλογο με εκείνο της ακουστικής τομογραφίας : Κάθε δέσμη δεν διαπερνά όλα τα στοιχειώδη τετραγωνικά στοιχεία του σώματος, αλλά μόνο αυτά που βρίσκονται στο πέρασμά της. Καταλήξαμε λοιπόν σε ένα σύστημα N εξισώσεων με M αγνώστους όπως και στην περίπτωση της ακουστικής τομογραφίας. Ο πίνακας $\mathbf{G}$ του γραμμικού συστήματος έχει αρκετά μηδενικά στοιχεία και αποτελείται από τα μήκη των διαδρομών των ακτίνων στο σώμα.

1.4 Λύσεις αντιστρόφων προβλημάτων.

Ο τρόπος παρουσίασης της λύσης ενός αντίστροφου προβλήματος μπορεί να διαφέρει ανάλογα με τη «φιλοσοφία» επίλυσης που ακολουθείται και τη δυνατότητα παροχής πρόσθετης πληροφορίας σχετικά με τη λύση.

- Εκτιμήσεις των παραμέτρων.

Η απλούστερη μορφή που μπορεί να πάρει η λύση ενός αντίστροφου προβλήματος είναι με συγκεκριμένα αριθμητικά στοιχεία να αποτελούν το διάνυσμα $\mathbf{m}$ των προς ανάκτηση παραμέτρων. Για παράδειγμα $\mathbf{m}^{est} = [1.2, 3, 4, \dots, 1.5]^T$. Συνήθως οι εκτιμήσεις αυτής της μορφής είναι οι πλέον χρήσιμες ως λύσεις ενός αντίστροφου προβλήματος. Δεν δίνουν όμως πρόσθετη πληροφορία για το ενδεχόμενο οι εκτιμήσεις αυτές να είναι προσεγγιστικές ή να έχουν μεγάλο περιθώριο ασάφειας.

- Οριακές τιμές

Εάν είναι δυνατή η εκτίμηση ακραίων τιμών είτε χρησιμοποιώντας απόλυτη είτε πιθανοθεωρητική έννοια, τότε μπορεί η λύση του αντίστροφου προβλήματος να δίδεται στη μορφή π.χ. $1.1 \leq m_1 \leq 1.3$. Εάν η θεώρηση είναι πιθανοθεωρητική μπορεί η ανωτέρω έκφραση να σημαίνει ότι η πιθανότητα να βρίσκεται η παράμετρος m_1 ανάμεσα στις δύο παρατιθέμενες τιμές είναι συγκεκριμένη. Έτσι μπορεί να γράφουμε: $m_1^{est} = 1.2 \pm 0.1$ και να εννοούμε ότι η πιθανότητα η παράμετρος m_1 να βρίσκεται ανάμεσα στο 1.1 και στο 1.3 είναι 95 %.

- Κατανομές Πιθανοτήτων

Εάν θεωρήσουμε ότι οι παράμετροι είναι τυχαίες μεταβλητές και έχουμε τη δυνατότητα να υπολογίσουμε τη συνάρτηση πυκνότητας πιθανότητας (δείτε επόμενο κεφάλαιο), για κάθε παράμετρο, μπορεί να δώσουμε τις σχετικές συναρτήσεις ως λύσεις του αντίστροφου προβλήματος. Στην πράξη αυτό δεν είναι πολύ βολικό γιατί με εξαίρεση κατανομές που παρουσιάζουν χαρακτηριστικές μοναδικές κορυφές η πληροφορία που δίνει η κατανομή δεν είναι άμεσα αξιοποιήσιμη στις εφαρμογές.

- Σταθμισμένες μέσες τιμές παραμέτρων

Σε πολλές περιπτώσεις, ένα αντίστροφο πρόβλημα μπορεί να δώσει καλύτερες απαντήσεις για συνδυασμούς (π.χ. γραμμικούς) παραμέτρων, οι οποίες μπορεί να υπολογιστούν ευκολότερα αλλά και να έχουν νόημα για τις εφαρμογές. Μπορεί για παράδειγμα εάν το διάνυσμα των παραμέτρων είναι το $\mathbf{m} = [m_1, m_2]^T$, αντί να υπολογιστεί χωριστά κάθε παράμετρος, να είναι ευκολότερο να υπολογιστεί μια μέση τιμή, για παράδειγμα της παράστασης $\langle m \rangle = 0.2m_1 + 0.8m_2$. Το αν αυτή είναι χρήσιμη πληροφορία ή όχι εξαρτάται από την εφαρμογή. Συνήθως αυτή την περίπτωση την αντιμετωπίζουμε όταν οι παράμετροι αφορούν διακριτοποιήσεις συνεχών παραμέτρων προκειμένου να διατυπωθεί ένα διακριτό αντίστροφο πρόβλημα.

2. ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΗ ΘΕΩΡΙΑ ΠΙΘΑΝΟΤΗΤΩΝ

Στη συνέχεια να αναφερθούμε σε βασικές έννοιες από τη θεωρία πιθανοτήτων που μας είναι ιδιαίτερα χρήσιμες στην περίπτωση που τα δεδομένα μας και συνακόλουθα και οι προς ανάκτηση παράμετροι θεωρηθούν τυχαίες μεταβλητές. Υπάρχει συγκεκριμένος λόγος γι αυτή τη θεώρηση που έχει να κάνει με το γεγονός ότι σε μία πραγματική εφαρμογή, οι μετρήσεις (που δίδουν τα δεδομένα σε ένα αντίστροφο πρόβλημα) γίνονται με λάθη ή σε ένα περιβάλλον θορύβου, με αποτέλεσμα η τιμή της μέτρησης να μπορεί να είναι διαφορετική εάν το πείραμα επαναληφθεί αμέσως. Επομένως η τιμή της μέτρησης μπορεί να θεωρηθεί τυχαία μεταβλητή (*random variable*). Οι ιδιότητες της τυχαίας μεταβλητής συνήθως είναι γνωστές ποιοτικά, αλλά οι συγκεκριμένες μετρήσεις που προκύπτουν θεωρούνται ως «πραγματοποιήσεις» (realizations) της τυχαίας μεταβλητής.

2.1 Κατανομή πιθανότητας

Εάν d είναι συνεχής τυχαία μεταβλητή, η *Συνάρτηση Πυκνότητας Πιθανότητας* $P(d)$ μας δίνει την πιθανότητα, η πραγματοποίηση της τυχαίας μεταβλητής να πάρει μία τιμή στην γειτονιά του ορίσμά της, δηλαδή ανάμεσα στο d και στο $d + \partial d$ (Σχήμα 2.1) μέσω του γινομένου $P(d)\partial d$.

Η πιθανότητα, η μεταβλητή d να πάρει τιμή ανάμεσα στο a και στο b είναι $\int_a^b P(d)\partial d$ ενώ υπάρχει πλήρης βεβαιότητα ότι η μεταβλητή d θα πάρει τιμή ανάμεσα στο $-\infty$ και στο $+\infty$, οπότε έχουμε :

$$\int_{-\infty}^{+\infty} P(d)\partial d = 1 \quad (2.1)$$

Η **μέση τιμή** (ή αναμενόμενη τιμή) για μία τυχαία μεταβλητή d η οποία χαρακτηρίζεται από τη συνάρτηση πυκνότητας πιθανότητας $P(d)$ δίδεται από τη σχέση :

$$E(d) = \langle d \rangle = \int_{-\infty}^{+\infty} dP(d)\partial d \quad (2.2)$$

Εάν η τυχαία μεταβλητή είναι διακριτή με N δυνατές πραγματοποιήσεις, κάθε μία από τις οποίες έχει πιθανότητα P_i , η ανωτέρω μέση τιμή δίδεται από την έκφραση :

$$\mu = \langle \vec{d} \rangle = \sum_{i=1}^N d_i P_i \quad (2.3)$$

Η διασπορά (variance) μια συνεχούς τυχαίας μεταβλητής, μας δίνει ποσοτικά το κατά πόσον η κατανομή πιθανότητας είναι ευρεία ή στενή, γύρω από τη μέση της τιμή. Η σχετική έκφραση είναι :

$$\sigma^2 = \text{Var}(d) = \int_{-\infty}^{+\infty} (d - \langle d \rangle)^2 P(d) \partial d \quad (2.4)$$

Το εύρος της κατανομής γύρω από τη μέση τιμή, δίδεται από το σ .

Για την περίπτωση της διακριτής τυχαίας μεταβλητής, η αντίστοιχη έκφραση είναι :

$$\text{Var}(\vec{d}) = \sum_{i=1}^N (d_i - \langle \vec{d} \rangle)^2 P_i \quad (2.5)$$

2.2 Συσχέτιση δεδομένων

Εάν σε ένα πείραμα έχουμε περισσότερα από ένα δεδομένα τα οποία αντιμετωπίζονται ως τυχαίες μεταβλητές, αποκτά ενδιαφέρον η συσχέτιση των μεταβλητών μεταξύ τους. Ορίζουμε στην περίπτωση αυτή την *από κοινού συνάρτηση πυκνότητας πιθανότητας* (joint distribution) $P(\mathbf{d}) = P(d_1, d_2, \dots, d_N)$, που δίνει τη πιθανότητα η πρώτη μεταβλητή να πάρει τιμή στη γειτονιά του d_1 , η δεύτερη στη γειτονιά του d_2 κ.λ.π.. Εάν τα δεδομένα είναι ανεξάρτητα τότε μόνο μπορούμε να πούμε ότι η από κοινού συνάρτηση πυκνότητας πιθανότητας είναι ίση με το γινόμενο των συναρτήσεων πυκνότητας πιθανότητας των μεταβλητών

$$P(\mathbf{d}) = P(d_1)P(d_2) \cdots P(d_N) \quad (2.6)$$

Σε άλλες περιπτώσεις πάντως τα δεδομένα συσχετίζονται και αποκτά ιδιαίτερη σημασία η έκφραση ενός μέτρου συσχέτισης μεταξύ τους.

Εάν θεωρήσουμε λοιπόν την περίπτωση συσχέτισης δύο τυχαίων μεταβλητών (συνεχών) d_1 και d_2 , ορίζεται η **συνδιακύμανση** (covariance) από τη σχέση :

$$\text{cov}(d_1, d_2) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} [d_1 - \langle d_1 \rangle][d_2 - \langle d_2 \rangle] P(\mathbf{d}) \partial d_1 \partial d_2 \cdots \partial d_N \quad (2.7)$$

Προσέξτε ότι οι τυχαίες μεταβλητές d_1 και d_2 είναι δύο μόνο από τις μεταβλητές (δεδομένα) που αποτελούν το διάνυσμα $\mathbf{d}$.

Για ένα διάνυσμα $\mathbf{d}$, N τυχαίων αλλά συνεχών μεταβλητών, ορίζουμε το διάνυσμα των μέσων τιμών από τη σχέση

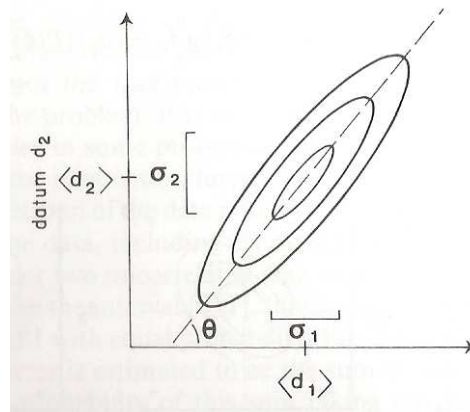
$$\langle \mathbf{d} \rangle_i = \int_{-\infty}^{+\infty} \partial d_1 \int_{-\infty}^{+\infty} \partial d_2 \cdots \int_{-\infty}^{+\infty} \partial d_N d_i P(\mathbf{d}) \quad (2.8)$$

και τον **πίνακα συνδιακύμανσης** από τη σχέση :

$$[\text{cov} \mathbf{d}]_{ij} = \int_{-\infty}^{+\infty} \partial d_1 \int_{-\infty}^{+\infty} \partial d_2 \cdots \int_{-\infty}^{+\infty} \partial d_N [d_i - \langle d \rangle_i][d_j - \langle d \rangle_j] P(\mathbf{d}) \quad (2.9)$$

Τα διαγώνια στοιχεία του πίνακα, αντιπροσωπεύουν το εύρος της κατανομής της κάθε μεταβλητής (διακύμανση), ενώ τα στοιχεία εκτός διαγωνίου αντιπροσωπεύουν το βαθμό συσχέτισης των αντίστοιχων μεταβλητών.

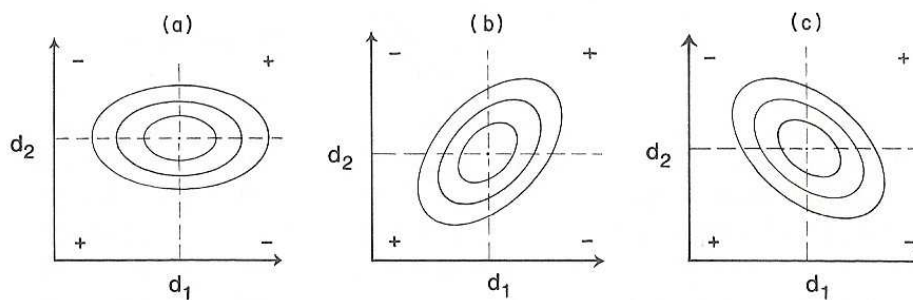
Στο επόμενο σχήμα βλέπουμε εποπτικά καμπύλες ίσης κατανομής πιθανότητας για δύο τυχαίες μεταβλητές d_1 και d_2 . Στο σχήμα φαίνονται οι μέσες τιμές και οι διακυμάνσεις των δύο μεταβλητών. Η γωνία θ είναι ένα μέτρο της συσχέτισης των δύο μεταβλητών και σχετίζεται με τη συνδιακύμανση.



Σχήμα 2.1 Καμπύλες ίσης κατανομής πιθανότητας $P(d_1, d_2)$ για δύο τυχαίες μεταβλητές

Στο επόμενο σχήμα 2.2 βλέπουμε τρία διαγράμματα στα οποία εμφανίζονται και πάλι καμπύλες ίσης κατανομής πιθανότητας για δύο τυχαίες μεταβλητές

Το πρώτο διάγραμμα αντιπροσωπεύει ουσιαστικά μη συσχετιζόμενες μεταβλητές. Μεγάλες τιμές της μιας από αυτές μπορεί να συσχετίζονται με μικρές ή μεγάλες τιμές της άλλης με ίδια πιθανότητα. Αντίθετα το δεύτερο διάγραμμα αντιπροσωπεύει μεταβλητές που συσχετίζονται θετικά. Δηλαδή μεγάλες τιμές της μιας μεταβλητής έχουν μεγάλη πιθανότητα να συσχετίζονται με μεγάλες τιμές της δεύτερης μεταβλητής και μικρές τιμές της μια με μικρές τιμές της άλλης. Με την ίδια λογική το τρίτο διάγραμμα αντιπροσωπεύει αρνητική συσχέτιση.



Σχήμα 2.2 Καμπύλες ίσης ίσης κατανομής πιθανότητας $P(d_1, d_2)$ για δύο τυχαίες μεταβλητές που είναι α) ασυσχέτιστες, β) θετικά και γ) αρνητικά συσχετιζόμενες.

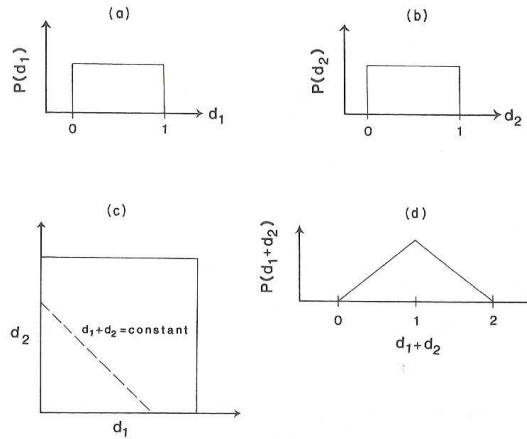
2.3 Συναρτήσεις τυχαίων μεταβλητών

Δεδομένου ότι στα αντίστροφα προβλήματα οι προς εκτίμηση παράμετροι $\mathbf{m}^{est}$ σχετίζονται με τα δεδομένα, όταν τα τελευταία είναι τυχαίες μεταβλητές, αντίστοιχα και οι παράμετροι μπορούν να θεωρηθούν τυχαίες μεταβλητές και συνεπώς μπορεί να οριστούν κατανομές πιθανοτήτων και γι αυτές $P(\mathbf{m}^{est})$. Σημειώνεται εν προκειμένω ότι οι πραγματικές παράμετροι μπορεί να είναι τυχαίες μεταβλητές ή ντετερμινιστικά μεγέθη ανάλογα με το πρόβλημα. Οι εκτιμήσεις τους πάντως είναι τυχαίες μεταβλητές τη στιγμή που έτσι θεωρούνται οι μετρήσεις.

Όταν λοιπόν οι προς ανάκτηση παράμετροι θεωρηθούν συναρτήσεις των δεδομένων, μπορεί από τα στατιστικά χαρακτηριστικά των μετρήσεων, να προκύψουν τα αντίστοιχα χαρακτηριστικά των παραμέτρων. Η διαδικασία μπορεί να μην είναι απλή είναι όμως άμεση.

Θα εξετάσουμε εδώ ένα απλό πρόβλημα. Υποθέτουμε ότι έχουμε δύο μετρήσεις d_1 και d_2 και ότι η προς ανάκτηση παράμετρος είναι το άθροισμά τους. Επομένως ισχύει $m = d_1 + d_2$. Ας υποθέσουμε επίσης ότι οι μετρήσεις μας (δεδομένα), έχουν ίση πιθανότητα να πάρουν τιμές ανάμεσα στο 1 και στο 2 (Σχήμα 2.3).

Για να υπολογίσουμε την πιθανότητα το άθροισμα των δύο μεταβλητών να πάρει κάποια τιμή, παίρνομε στο επίπεδο d_1, d_2 καμπύλες ίσης τιμής του αθροίσματος και ολοκληρώνουμε πάνω στις καμπύλες. Εδώ οι καμπύλες είναι ευθείες (διάγραμμα c) και ο υπολογισμός είναι εύκολος οδηγώντας μας στο διάγραμμα d που αντιπροσωπεύει την κατανομή πιθανότητας του αθροίσματος και βλέπομε ότι είναι ένα τρίγωνο με βάση ανάμεσα στο 1 και το 2.



Σχήμα 2.3 Κατανομές πιθανότητας για δύο μεταβλητές d_1 και d_2 και για το άθροισμά τους $m = d_1 + d_2$

Ο υπολογισμός που κάναμε παραπάνω μπορεί να είναι δύσκολος σε γενικές περιπτώσεις, όμως μπορεί ναδειχθεί ότι εάν δεδομένα και παράμετροι συσχετίζονται γραμμικά μέσω της σχέσης :

$$\mathbf{m} = \mathbf{M}\mathbf{d} + \mathbf{v} \quad (2.10)$$

όπου $\mathbf{M}$ και $\mathbf{v}$ είναι κάποιος πίνακας και διάνυσμα αντίστοιχα, η μέση τιμή και η συνδιακύμανση των παραμέτρων υπολογίζονται άμεσα από τις σχέσεις :

$$\langle \mathbf{m} \rangle = \mathbf{M} \langle \mathbf{d} \rangle + \mathbf{v} \quad (2.11)$$

και

$$[\text{cov} \mathbf{m}] = \mathbf{M}[\text{cov} \mathbf{d}]\mathbf{M}^T \quad (2.12)$$

Άσκηση : Να υπολογιστούν τα στατιστικά χαρακτηριστικά της παραμέτρου m που είναι η μέση τιμή μιας ομάδας δεδομένων : $m = 1/N \sum_{i=1}^N d_i = (1/N)[1,1,\dots,1]\mathbf{d}$ για τα οποία γνωρίζουμε ότι έχουν όλα ίδια μέση τιμή $\langle d \rangle$ και διακύμανση σ_d^2

Λύση : Με βάση τη σχέση 2.10, θα ισχύει : $\mathbf{M} = [1,1,\dots,1]/N$, $\mathbf{v} = 0$. Από τη σχέση 2.11 έχουμε : $\langle m \rangle = \mathbf{M} \langle \mathbf{d} \rangle + \mathbf{v} = \langle d \rangle$ και $\text{var}(m) = \mathbf{M}[\text{cov} \mathbf{d}]\mathbf{M}^T = \sigma_d^2 / N$.

Παρατηρούμε ότι η τετραγωνική ρίζα της διακύμανσης που είναι ένα μέτρο του εύρους της διασποράς των πιθανών τιμών της παραμέτρου m γύρω από τη μέση τιμή, και επομένως της πιθανότητας ένα πείραμα να δώσει τιμή κοντά στη μέση, είναι ανάλογη του $N^{-1/2}$ που σημαίνει ότι η ακρίβεια του υπολογισμού του μέσου μιας ομάδας μετρήσεων, αυξάνει πολύ αργά σε σχέση με την αύξηση του αριθμού των μετρήσεων.

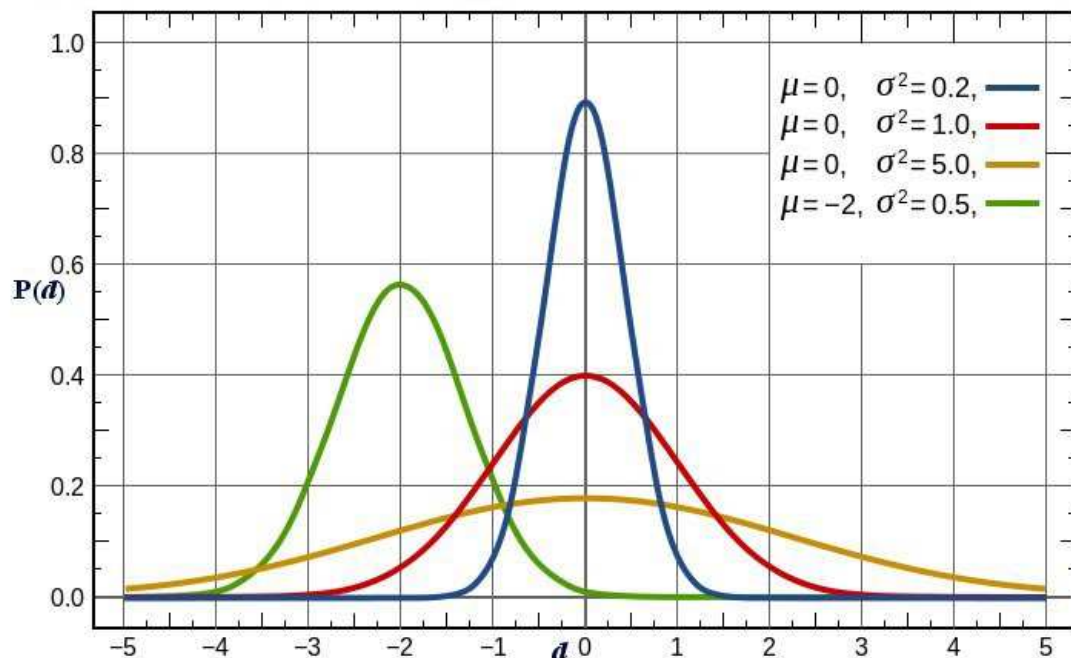
2.4 Κανονικές κατανομές

Μια τυχαία μεταβλητή d θα λέγεται **κανονική** (Normal ή Gaussian) με παραμέτρους μ και σ αν η συνάρτηση πυκνότητας πιθανότητας της δίδεται από τη σχέση :

$$P(d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(d-\mu)^2}{2\sigma^2}\right] \quad (2.13)$$

όπου $\sigma^2 = \text{Var}(d)$ και $\mu = \langle d \rangle$. Η κανονική κατανομή εμφανίζεται στις περισσότερες εφαρμογές καθώς είναι και η οριακή κατανομή για ένα άθροισμα τυχαίων μεταβλητών, βάσει του κεντρικού οριακού θεωρήματος. Για προβλήματα ανάκτησης παραμέτρων από μετρήσεις, όταν οι μετρήσεις γίνονται σε περιβάλλον θορύβου που προέρχεται από πολλές πηγές, η κατανομή τους τείνει στην μορφή της κανονικής κατανομής.

Στο Σχήμα 2.4 φαίνεται το διάγραμμα για διαφορετικές κανονικές κατανομές. Για τρεις από αυτές η μέση τιμή είναι 0 ενώ για την Τρίτη είναι -2. Θα πρέπει να σημειωθεί ότι η επιφάνεια κάτω από την καμπύλη με οριακές τιμές $-1 < \sigma < 1$ είναι 0.68 ενώ η επιφάνεια κάτω από την καμπύλη με οριακές τιμές $-2 < \sigma < 2$ είναι 0.95. Με άλλα λόγια η πιθανότητα μια τυχαία μεταβλητή που ακολουθεί κανονική κατανομή να πάρει τιμή $\mu \pm \sigma$ είναι 68 %, ενώ η αντίστοιχη πιθανότητα να πάρει τιμή $\mu \pm 2\sigma$ είναι 95 %.



Σχήμα 2.4 Κανονική Κατανομή

Η από κοινού συνάρτηση πυκνότητας πιθανότητας δύο ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν κανονική κατανομή είναι το γινόμενο των συναρτήσεων πυκνότητας πιθανότητας τους όπως είδαμε και στη γενική περίπτωση. Εάν όμως οι μεταβλητές συσχετίζονται και έχουν μέση τιμή $\langle \mathbf{d} \rangle$ και συνδιακύμανση $[\text{cov} \mathbf{d}]$ η κατανομή είναι περισσότερο περίπλοκη. Η έκφραση :

$$P(\mathbf{d}) = \frac{[[\text{cov} \mathbf{d}]]^{-1/2}}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}[\mathbf{d} - \langle \mathbf{d} \rangle]^T [\text{cov} \mathbf{d}]^{-1} [\mathbf{d} - \langle \mathbf{d} \rangle]\right) \quad (2.14)$$

μας δίνει τη σωστή μέση τιμή και τη σωστή διακύμανση όταν οι μεταβλητές που αποτελούν το διάνυσμα $\mathbf{d}$ είναι ασυσχέτιστες και έχει συνδιακύμανση $[\text{cov} \mathbf{d}]$ για συσχετιζόμενες μεταβλητές.

2.5 Διαστήματα εμπιστοσύνης

Η *εμπιστοσύνη* (confidence) μιας συγκεκριμένης παρατήρησης (μέτρησης) είναι η πιθανότητα μια πραγματοποίηση της μέτρησης να λάβει τιμή μέσα σε ένα προκαθορισμένο εύρος τιμών γύρω από τη μέση της τιμή. Όπως είναι φυσικό, αναφερόμενοι σε μετρήσεις που η τιμή τους είναι τυχαία μεταβλητή, η μεγάλη τιμή της διακύμανσης δίνει μεγάλα *διαστήματα εμπιστοσύνης* (confidence intervals) και αντίστροφα. Όπως αναφέρθηκε ήδη, τυχαίες μεταβλητές κανονικής κατανομής έχουν 68 % διάστημα εμπιστοσύνης για εύρος 1σ και 95 % για εύρος 2σ .

Εάν λοιπόν μία τυχαία μεταβλητή λάβει μία συγκεκριμένη τιμή a (π.χ. στην περίπτωση μιας μέτρησης), στην περίπτωση που αυτή ακολουθεί την κανονική κατανομή, μπορούμε να ισχυριστούμε ότι η πιθανότητα η πραγματική μέση τιμή της μέτρησης να είναι $a \pm 2\sigma$ είναι 95 %.

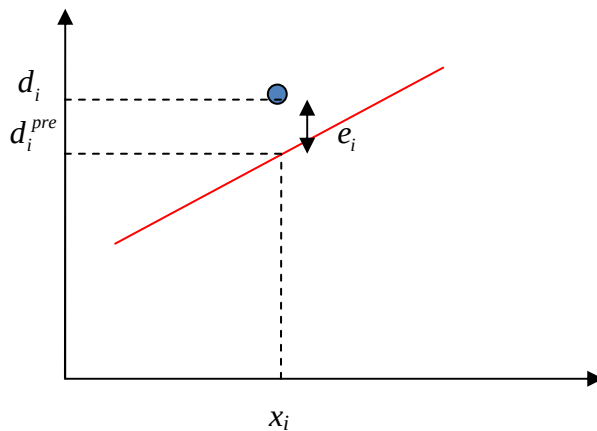
Φυσικά τα διαστήματα εμπιστοσύνης είναι πολύ πιο δύσκολο να υπολογιστούν όταν έχουμε περισσότερες συσχετιζόμενες μεταβλητές.

3. ΕΠΙΛΥΣΗ ΓΡΑΜΜΙΚΟΥ ΑΝΤΙΣΤΡΟΦΟΥ ΠΡΟΒΛΗΜΑΤΟΣ ΜΕ ΤΗ ΜΕΘΟΔΟ ΜΗΚΩΝ

3.1 Διαφορά μετρήσεων από εκτιμήσεις μετρήσεων.

Όταν επιλύουμε ένα αντίστροφο πρόβλημα υπολογίζουμε ένα διάνυσμα παραμέτρων $\mathbf{m}^{est}$ το οποίο αντιπροσωπεύει μία *εκτίμηση* της λύσης του αντίστροφου προβλήματος. Στην περίπτωση του γραμμικού αντίστροφου προβλήματος που παράμετροι και δεδομένα συσχετίζονται μέσω της σχέσης $\mathbf{d} = \mathbf{G}\mathbf{m}$, η αντικατάσταση της λύσης $\mathbf{m}^{est}$ στην προηγούμενη σχέση, δεν είναι απαραίτητο ότι θα μας δώσει ξανά το διάνυσμα $\mathbf{d}$ των παρατηρήσεων, αλλά ένα άλλο διάνυσμα $\mathbf{d}^{pre}$ το οποίο μπορεί να διαφέρει ως προς τα στοιχεία του από το $\mathbf{d}$. Το διάνυσμα των διαφορών (misfit) ανάμεσα στις μετρήσεις και στις εκτιμήσεις των μετρήσεων συμβολίζεται με $\mathbf{e}$ και έχει διάσταση N . Φυσικά $e_i = d_i - d_i^{pre}$.

Στο Σχήμα 3.1 φαίνεται η σχέση μέτρησης και εκτίμησης μέτρησης για την περίπτωση του υπολογισμού των παραμέτρων μιας ευθείας. Για την τυχούσα μέτρηση στο σημείο z_i η ευθεία που χαράζουμε μας δίνει την τιμή d_i^{pre} ενώ η μετρηθείσα τιμή είναι d_i . Η αντίστοιχη διαφορά (misfit) είναι e_i .



Σχήμα 3.1 Διαφορά μέτρησης και εκτίμησης μέτρησης

Στόχος γενικά της διαδικασίας επίλυσης ενός αντίστροφου προβλήματος είναι να ελαχιστοποιήσει τις διαφορές με κάποιο ορθολογικό τρόπο, ώστε το αποτέλεσμα των εκτιμήσεων να είναι πολύ κοντά στην πραγματικότητα, δηλαδή στις πραγματικές μετρήσεις. Για το σκοπό αυτό χρησιμοποιούμε την έννοια της νόρμας διανυσμάτων για να έχουμε ένα μέγεθος προς βελτιστοποίηση.

3.2 Νόρμες διανυσμάτων

Γνωρίζουμε ότι η έννοια της νόρμας $\|x\|$ είναι συνυφασμένη με τις εξής ιδιότητες :

$$\|x\| > 0 \quad (3.1\alpha)$$

$$\|ax\| = |a|\|x\| \quad (3.1\beta)$$

$$\|x + y\| \leq \|x\| + \|y\| \quad (3.1\gamma)$$

Εάν x και y είναι διανύσματα ισχύουν ακριβώς οι ίδιες σχέσεις.

Στην περίπτωση των διανυσμάτων ορίζουμε τις παρακάτω νόρμες που είναι οι διακριτές ανάλογες των νορμών των συνεχών μεταβλητών :

$$L_1 \text{ νόρμα : } \|x\|_1 = \left[\sum_i |x_i|^1 \right] \quad (3.2\alpha)$$

$$L_2 \text{ νόρμα : } \|x\|_2 = \left[\sum_i |x_i|^2 \right]^{1/2} \quad (3.2\beta)$$

.

.

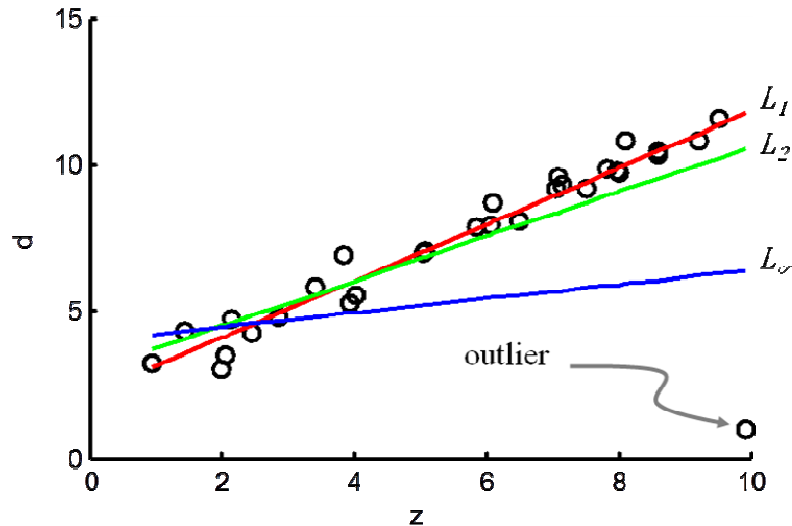
$$L_n \text{ νόρμα : } \|x\|_n = \left[\sum_i |x_i|^{1/n} \right] \quad (3.2\gamma)$$

Επίσης ορίζεται η νόρμα μεγίστου από τη σχέση

$$L_\infty \text{ νόρμα : } \|x\|_\infty = \max_i |x_i| \quad (3.2\delta)$$

3.3 Η νόρμα της διαφοράς μετρήσεων και εκτιμήσεων

Εάν στη θέση του διανύσματος x βάλομε το διάνυσμα e οι νόρμες που ορίσαμε παραπάνω μας δίνουν μία μέτρηση «μήκους» ως προς την διαφορά μετρήσεων και εκτιμήσεων. Ελαχιστοποιώντας μία νόρμα αυτής της μορφής, ελαχιστοποιούμε λοιπόν και μία ποσότητα που σχετίζεται διαφορές μετρήσεων και εκτιμήσεων. Ασφαλώς η ελαχιστοποίηση μιας συγκεκριμένης νόρμας ως προς τις παραμέτρους που προκαλούν τη διαφορά, δεν δίδει το ίδιο αποτέλεσμα με την ελαχιστοποίηση κάποιας άλλης.



Σχήμα 3.2 Πέρασμα ευθείας από σημεία μετρήσεων που περιλαμβάνουν και μία ακραία μέτρηση (outlier). (Σχήμα από παραδόσεις του Menke).

Έχει λοιπόν ιδιαίτερο ενδιαφέρον να δούμε τη σημασία κάθε νόρμας ως προς τα χαρακτηριστικά των στοιχείων που την αποτελούν.

Ας δούμε το Σχήμα 3.2. Αναφέρεται στο πρόβλημα του υπολογισμού παραμέτρων ευθείας όταν υπάρχουν N μετρήσεις σε διάφορα σημεία του άξονα z που συμβολίζονται με μικρούς κύκλους. Το πρόβλημα αυτό το είδαμε στο εδάφιο 1.3.1. Στο σχήμα έχουμε υποθέσει ότι υπάρχει μία μέτρηση (outlier) που διαφέρει πολύ από τις υπόλοιπες. Στόχος μας είναι να ελαχιστοποιήσουμε τις διαφορές μετρήσεων και εκτιμήσεων μετρήσεων και για το σκοπό αυτό υπολογίζουμε τις παραμέτρους της ευθείας ελαχιστοποιώντας κάποια νόρμα. Το πώς θα γίνει αυτό θα το δούμε σε άλλο εδάφιο.

Εάν χρησιμοποιήσουμε την διακριτή L_1 νόρμα, παίρνομε την κόκκινη γραμμή. Εάν χρησιμοποιήσουμε την διακριτή L_2 νόρμα, παίρνομε την πράσινη γραμμή και εάν χρησιμοποιήσουμε την L_∞ νόρμα παίρνομε την μπλε γραμμή. Βλέπομε ότι η μπλε γραμμή είναι πιο κοντά στην ακραία μέτρηση, άρα της δίδει ιδιαίτερη σημασία. Γενικά η νόρμα μεγαλύτερης τάξης δίνει μεγαλύτερη σημασία σε ακραίες τιμές μετρήσεων. Αντίθετα η νόρμα μικρής τάξης δίνει μικρή αξία σε ακραίες τιμές μετρήσεων. Εάν έχουμε λοιπόν λόγους να πιστεύουμε ότι ακραίες τιμές είναι σημαντικές και αξιόπιστες, πρέπει να χρησιμοποιούμε νόρμες μεγαλύτερης τάξης. Εάν όμως από την άλλη μεριά θεωρήσουμε ότι οι μετρήσεις είναι τυχαίες μεταβλητές που ακολουθούν **κανονική κατανομή** τότε μία νόρμα που δίνει καλά αποτελέσματα είναι η διακριτή L_2 νόρμα που καλείται και *Ευκλείδεια Νόρμα*. Για μετρήσεις που ακολουθούν κανονική κατανομή, είναι απίθανο να έχουμε καταστάσεις όπως αυτή του σχήματος 3.2 με την ακραία μέτρηση, αλλά ακόμη και εάν αυτή υπάρχει, μπορούμε να θεωρήσουμε ότι πρόκειται για τυχαίο λάθος μέτρησης.

Εάν λοιπόν χρησιμοποιήσουμε την L_2 νόρμα, μπορούμε να δούμε ότι για το διάνυσμα $\mathbf{e}$, αυτή είναι ίση με το γινόμενο

$$E = \mathbf{e}^T \mathbf{e} = \sum_{i=1}^N e_i^2, \quad (3.3)$$

που χαρακτηρίζεται και ως *ολικό λάθος* (overall error).

Ελαχιστοποιώντας λοιπόν το ολικό λάθος ως προς μία προς μία τις παραμέτρους του γραμμικού αντίστροφου προβλήματος, μπορούμε να πάρουμε εκτιμήσεις της λύσης του που έχουν προκύψει από την αρχή των *ελαχίστων τετραγώνων*.

3.4 Η λύση ελαχίστων τετραγώνων για το γενικό γραμμικό αντίστροφο πρόβλημα.

3.4.1 Αναλυτικός υπολογισμός των παραμέτρων

Ας θεωρήσουμε το γραμμικό αντίστροφο πρόβλημα που ορίζεται από τη σχέση :

$$\mathbf{d} = \mathbf{Gm} \quad (3.4)$$

Εάν λύσουμε το πρόβλημα και πάρουμε λύσεις $\mathbf{m}^{est}$ η ανωτέρω σχέση δίνει :

$$\mathbf{d}^{obs} = \mathbf{Gm}^{est} \quad (3.5)$$

και το ολικό λάθος E είναι :

$$\begin{aligned} E = \mathbf{e}^T \mathbf{e} &= (\mathbf{d} - \mathbf{d}^{obs})^T (\mathbf{d} - \mathbf{d}^{obs}) = (\mathbf{d} - \mathbf{Gm}^{est})^T (\mathbf{d} - \mathbf{Gm}^{est}) = \\ &= \sum_{i=1}^N \left[d_i - \sum_{j=1}^M G_{ij} m_j^{est} \right] \left[d_i - \sum_{k=1}^M G_{ik} m_k^{est} \right] \end{aligned} \quad (3.6)$$

Παίρνοντας παράγωγο ως προς κάθε παράμετρο προς υπολογισμό $m_q, q=1,...,M$ μπορούμε μετά από πράξεις (άσκηση) να καταλήξουμε στο σύστημα :

$$\mathbf{G}^T \mathbf{Gm}^{est} - \mathbf{G}^T \mathbf{d} = 0 \quad (3.7)$$

Ο πίνακας $\mathbf{G}^T \mathbf{G}$ είναι ένας τετραγωνικός πίνακας $M \times M$. Εάν ο πίνακας αυτός αντιστρέφεται, τότε μπορούμε να πάρουμε τη λύση :

$$\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d}, \quad (3.8)$$

που χαρακτηρίζεται ως λύση *ελαχίστων τετραγώνων* για το διακριτό αντίστροφο πρόβλημα (3.4).

3.4.2 Η λύση ελαχίστων τετραγώνων από την πλευρά της Γραμμικής Άλγεβρας.

Θα δούμε τώρα πως μπορούμε να καταλήξουμε στο ίδιο αποτέλεσμα για την λύση ελαχίστων τετραγώνων χρησιμοποιώντας έννοιες της Γραμμικής Άλγεβρας.

Ας θεωρήσουμε πάλι το γραμμικό πρόβλημα που ορίζεται από τη σχέση :

$$\mathbf{d} = \mathbf{G}\mathbf{m} \quad (3.4)$$

Παρατηρούμε ότι ο πίνακας $\mathbf{G}$ είναι παραλληλόγραμμος με διαστάσεις $N \times M$. Το πρόβλημα αυτό γνωρίζουμε πολύ καλά ότι γενικά δεν έχει λύση. Εάν έχουμε $N > M$, και πάρουμε ένα νέο διάνυσμα $\mathbf{d}'$ το οποίο να ανήκει στο χώρο των στηλών του $\mathbf{G}$, τότε το αντίστοιχο πρόβλημα έχει λύση. Ποιο όμως θα είναι αυτό το διάνυσμα ; Ασφαλώς με βάση αυτά που έχουμε πει μέχρι τώρα, θέλουμε το $\mathbf{d}'$ να είναι κατά το δυνατόν πλησιέστερο στο $\mathbf{d}$. Γνωρίζουμε ότι αυτό είναι η ορθογώνια προβολή του $\mathbf{d}$ στο χώρο των στηλών του $\mathbf{G}$, και θα ισχύει :

$$\mathbf{d}' = \mathbf{G}\mathbf{m}^{est} \quad (3.9)$$

Τότε όμως το διάνυσμα της διαφοράς $\mathbf{d} - \mathbf{d}' = \mathbf{d} - \mathbf{G}\mathbf{m}^{est}$ θα είναι ορθογώνιο στο χώρο των στηλών του $\mathbf{G}$ και επομένως θα ανήκει στον αριστερό μηδενόχωρο του $\mathbf{G}$. Επομένως θα ισχύει η σχέση :

$$\mathbf{G}^T(\mathbf{d} - \mathbf{d}') = 0 \Rightarrow \mathbf{G}^T(\mathbf{d} - \mathbf{G}\mathbf{m}^{est}) = 0 \Rightarrow \mathbf{G}^T\mathbf{d} = \mathbf{G}^T\mathbf{G}\mathbf{m}^{est} \quad (3.10)$$

Ανακτήσαμε συνεπώς τη σχέση 3.7.

Ξέρομε τώρα από την Γραμμική Άλγεβρα ότι για να αντιστρέφεται ο πίνακας $\mathbf{G}^T\mathbf{G}$ **θα πρέπει ο $\mathbf{G}$ να έχει γραμμικά ανεξάρτητες στήλες**. Αυτό σημαίνει ότι οι μετρήσεις πρέπει να γίνονται με κατάλληλο τρόπο ώστε να εξασφαλίζεται η εν λόγω προϋπόθεση.

3.5 Παραδείγματα επίλυσης γραμμικών αντιστρόφων προβλημάτων με την αρχή των ελαχίστων τετραγώνων.

3.5.1 Υπολογισμός ευθείας γραμμής

Επανερχόμαστε στο πρόβλημα του υπολογισμού των παραμέτρων μιας ευθείας γραμμής από N μετρήσεις. Είδαμε στην εξίσωση 1.9 το μοντέλο του προβλήματος που εδώ θα το γράψουμε στη μορφή:

$$d_i = m_1 + m_2 z_i \quad (3.11)$$

και στη μορφή της Γραμμικής Εξίσωσης $\mathbf{Gm} = \mathbf{d}$ παίρνει το σχήμα :

$$\begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & z_N \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_N \end{bmatrix} \quad (3.12)$$

οπότε

$$\mathbf{G}^T \mathbf{G} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ z_1 & z_2 & \cdot & \cdot & \cdot & z_N \end{bmatrix} \begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & z_N \end{bmatrix} = \begin{bmatrix} N & \Sigma z_i \\ \Sigma z_i & \Sigma z_i^2 \end{bmatrix} \quad (3.13)$$

και

$$\mathbf{G}^T \mathbf{d} = \dots = \begin{bmatrix} \Sigma d_i \\ \Sigma z_i d_i \end{bmatrix} \quad (3.14)$$

οπότε η λύση ελαχίστων τετραγώνων είναι :

$$\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d} = \begin{bmatrix} N & \Sigma z_i \\ \Sigma z_i & \Sigma z_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma d_i \\ \Sigma z_i d_i \end{bmatrix} \quad (3.15)$$

Εάν η ορίζουσα του πίνακα $\mathbf{G}^T \mathbf{G}$ είναι 0, προφανώς ο πίνακας δεν αντιστρέφεται και δεν έχουμε λύση ελαχίστων τετραγώνων. Από την Γραμμική Άλγεβρα ξέρομε πάντως ότι όταν οι στήλες του πίνακα $\mathbf{G}$, είναι γραμμικά ανεξάρτητες, τότε ο πίνακας $\mathbf{G}^T \mathbf{G}$ αντιστρέφεται. Απλή παρατήρηση στην 3.12 μας υποδηλώνει ότι ειδικά για το συγκεκριμένο πρόβλημα αρκεί οι μετρήσεις να μην έχουν όλες γίνει στο ίδιο σημείο προκειμένου να υπάρξει λύση.

3.5.2 Υπολογισμός παραβολής

Θεωρούμε τώρα το πρόβλημα του υπολογισμού των παραμέτρων μιας παραβολής από N μετρήσεις όπως ακριβώς και στο προηγούμενο πρόβλημα. Εδώ το μοντέλο μας είναι

$$d_i = m_1 + m_2 z_i + m_3 z_i^2 \quad (3.16)$$

και στη μορφή της Γραμμικής Εξίσωσης $\mathbf{Gm} = \mathbf{d}$ παίρνει το σχήμα :

$$\begin{bmatrix} 1 & z_1 & z_1^2 \\ 1 & z_2 & z_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & z_N & z_N^2 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_N \end{bmatrix} \quad (3.17)$$

Οπότε για το γινόμενο $\mathbf{G}^T \mathbf{G}$ παίρνουμε :

$$\mathbf{G}^T \mathbf{G} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ z_1 & z_2 & \cdot & \cdot & \cdot & z_N \\ z_1^2 & z_2^2 & \cdot & \cdot & \cdot & z_N^2 \end{bmatrix} \begin{bmatrix} 1 & z_1 & z_1^2 \\ 1 & z_2 & z_2^2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & z_N & z_N^2 \end{bmatrix} = \begin{bmatrix} N & \Sigma z_i & \Sigma z_i^2 \\ \Sigma z_i & \Sigma z_i^2 & \Sigma z_i^3 \\ \Sigma z_i^2 & \Sigma z_i^3 & \Sigma z_i^4 \end{bmatrix} \quad (3.18)$$

και για τον όρο $\mathbf{G}^T \mathbf{d}$

$$\mathbf{G}^T \mathbf{d} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_N \\ z_1^2 & z_2^2 & \dots & z_N^2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} = \begin{bmatrix} \sum d_i \\ \sum z_i d_i \\ \sum z_i^2 d_i \end{bmatrix} \quad (3.19)$$

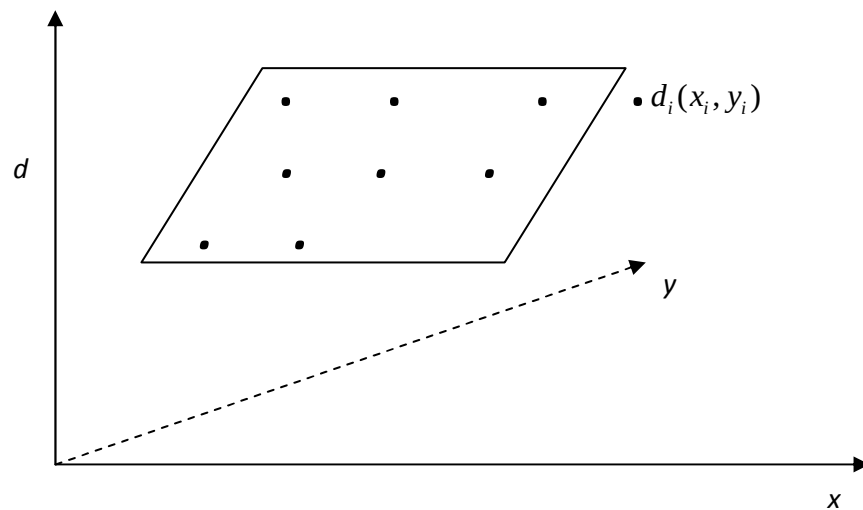
Έτσι η λύση ελαχίστων τετραγώνων είναι η :

$$\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d} = \begin{bmatrix} N & \sum z_i & \sum z_i^2 \\ \sum z_i & \sum z_i^2 & \sum z_i^3 \\ \sum z_i^2 & \sum z_i^3 & \sum z_i^4 \end{bmatrix}^{-1} \begin{bmatrix} \sum d_i \\ \sum z_i d_i \\ \sum z_i^2 d_i \end{bmatrix} \quad (3.20)$$

Ως προς την επιλυσιμότητα του συστήματος 3.19 και την ύπαρξη της λύσης 3.20, ισχύουν τα αναφερθέντα και στο προηγούμενο παράδειγμα σχετικά με τις ιδιότητες του πίνακα $\mathbf{G}$.

3.5.3 Υπολογισμός εξίσωσης επιπέδου στο χώρο.

Ας δούμε ένα ακόμη παράδειγμα επίλυσης αντιστρόφου προβλήματος με την αρχή των ελαχίστων τετραγώνων. Θέλουμε να υπολογίσουμε τις παραμέτρους της εξίσωσης που παριστάνει ένα επίπεδο στο χώρο.



Σχήμα 3.3 Υπολογίζοντας τις παραμέτρους της εξίσωσης επιπέδου

Με αναφορά στο σχήμα 3.3 θα πρέπει να υπολογίσουμε τις παραμέτρους $\mathbf{m} = [m_1, m_2, m_3]^T$ μια και το επίπεδο έχει εξίσωση :

$$d_i = m_1 + m_2 x + m_3 y \quad (3.21)$$

Η εξίσωση $\mathbf{G}\mathbf{m} = \mathbf{d}$ παίρνει τη μορφή :

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_N & y_N \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_N \end{bmatrix} \quad (3.22)$$

Αντίστοιχα έχουμε για το γινόμενο $\mathbf{G}^T \mathbf{G}$:

$$\mathbf{G}^T \mathbf{G} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ x_1 & x_2 & \cdot & \cdot & \cdot & x_N \\ y_1 & y_2 & \cdot & \cdot & \cdot & y_N \end{bmatrix} \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & x_N & y_N \end{bmatrix} = \begin{bmatrix} N & \Sigma x_i & \Sigma y_i \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i y_i \\ \Sigma y_i & \Sigma x_i y_i & \Sigma y_i^2 \end{bmatrix} \quad (3.23)$$

και για τον όρο $\mathbf{G}^T \mathbf{d}$

$$\mathbf{G}^T \mathbf{d} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ x_1 & x_2 & \cdot & \cdot & \cdot & x_N \\ y_1 & y_2 & \cdot & \cdot & \cdot & y_N \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_N \end{bmatrix} = \begin{bmatrix} \Sigma d_i \\ \Sigma x_i d_i \\ \Sigma y_i d_i \end{bmatrix} \quad (3.24)$$

που δίνουν τη λύση ελαχίστων τετραγώνων :

$$\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d} = \begin{bmatrix} N & \Sigma x_i & \Sigma y_i \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i y_i \\ \Sigma y_i & \Sigma x_i y_i & \Sigma y_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma d_i \\ \Sigma x_i d_i \\ \Sigma y_i d_i \end{bmatrix} . \quad (3.25)$$

Επισημαίνεται για μία ακόμη φορά ότι ο αντίστροφος του πίνακα $\mathbf{G}^T \mathbf{G}$ υπάρχει όταν οι στήλες του $\mathbf{G}$ είναι γραμμικά ανεξάρτητες. Δεν μπορεί για παράδειγμα οι

μετρήσεις να γίνονται στο ίδιο σημείο, αλλά ούτε και να γίνονται με τρόπο που $x_i = ky_i$ για όλα τα i .

3.6 Υπο-ορισμένα προβλήματα

Η λύση ελαχίστων τετραγώνων που είδαμε στο προηγούμενο εδάφιο, αναφέρεται σε προβλήματα που έχουν περισσότερη πληροφορία απ' ό,τι χρειάζεται για να δοθεί μοναδική λύση. Τα προβλήματα αυτά αναφέρονται ως *υπερ-ορισμένα* (*overdetermined*). Με την έννοια της Γραμμικής Άλγεβρας αναφερόμαστε σε γραμμικά προβλήματα που ανάγονται σε συστήματα N εξισώσεων με M αγνώστους με $N > M$. Με την ίδια λογική, προβλήματα που έχουν λιγότερες εξισώσεις απ' ό,τι αγνώστους χαρακτηρίζονται *υπο-ορισμένα* (*underdetermined*).

Στη θεωρία πάντως των αντιστρόφων προβλημάτων, ο χαρακτηρισμός σε υπο-ορισμένα ή υπερ-ορισμένα προβλήματα σχετίζεται και με τη διαδικασία λήψης των μετρήσεων, καθώς υπάρχουν προβλήματα στα οποία ανεξάρτητα από τον αριθμό των μετρήσεων και των παραμέτρων, υπάρχει περισσότερη πληροφορία απ' ό,τι χρειάζεται για ορισμένες παραμέτρους και λιγότερη για κάποιες άλλες. Τα προβλήματα αυτά χαρακτηρίζονται γενικά ως *μεικτά*.

Υπάρχει πάντως μία κατηγορία γραμμικών αντιστρόφων προβλημάτων για τα οποία ισχύει $N < M$ και επί πλέον οι εξισώσεις είναι συνεπείς (consistent). Τα προβλήματα αυτά χαρακτηρίζονται ως *καθαρά υπο-ορισμένα προβλήματα*. Για παράδειγμα θεωρήστε το γραμμικό σύστημα

$$\begin{aligned}x + y + z &= 6 \\ 2x + y - z &= 1\end{aligned}\tag{3.26}$$

Το πρόβλημα υπολογισμού των αγνώστων είναι καθαρά υπο-ορισμένο. Υπάρχουν περισσότερες από μία λύσεις (στην πραγματικότητα άπειρες) που όλες τους ικανοποιούν τις παραπάνω εξισώσεις και δίνουν μηδενικό ολικό λάθος. Δοκιμάστε τις λύσεις $[1, 2, 3]^T$ και $[-3, 8, 1]^T$. Και οι δύο ικανοποιούν την 3.26 ακριβώς. Εάν οι λύσεις αυτές αφορούσαν ένα αντίστροφο πρόβλημα, θα είχαμε κάποια δυνατότητα να αποφασίσουμε ποια είναι η καλύτερη; Με άλλα λόγια, σε ένα καθαρά υπο-ορισμένο πρόβλημα έχουμε κάποιο μηχανισμό για να μας δώσει μία *βέλτιστη* λύση;

Η απάντηση είναι ότι υπάρχουν διάφορα κριτήρια που μας δίνουν τη δυνατότητα να επιλέξουμε από τις δυνατές λύσεις του γραμμικού συστήματος στο οποίο ανάγεται το αντίστροφο πρόβλημα εκείνη που είναι πιο κοντά στην αναμενόμενη πραγματικότητα. Τα κριτήρια αυτά συνοψίζονται στον όρο *εκ προοιμίου* ή *αρχική πληροφορία* (*a-priori information*) που μπορεί να διατυπώνεται με εναλλακτικούς τρόπους. Για παράδειγμα, ας υποθέσουμε ότι μελετάμε ένα αντίστροφο πρόβλημα που έχει ως στόχο τον υπολογισμό της πυκνότητας υλικών. Η πυκνότητα γνωρίζουμε ότι

είναι ένα μέγεθος που εκφράζεται με θετικό αριθμό. Συνεπώς αποκλείονται από τις δυνατές λύσεις εκείνες που μπορεί να δίνουν αρνητικές πυκνότητες. Στο παράδειγμα του συστήματος 3.26 εάν οι άγνωστοι πρέπει να είναι θετικές ποσότητες, θα πρέπει να αποκλείσουμε την δεύτερη από τις αναφερθείσες λύσεις αφού μας δίνει $x < 0$

Ένας περιορισμός που μπορεί σε ορισμένες περιπτώσεις να χρησιμοποιηθεί σε αντίστροφα προβλήματα από φυσικές επιστήμες είναι εκείνος που αναφέρεται στο *ελάχιστο μήκος* της λύσης. Δηλαδή ζητάμε λύσεις που να έχουν όσο το δυνατόν χαμηλές αριθμητικά τιμές. Αυτό μπορεί να εκφραστεί μέσω της ευκλείδειας νόρμας της λύσης που είναι :

$$L = \mathbf{m}^T \mathbf{m} = \sum_{j=1}^M m_j^2 \quad (3.27)$$

Θα δούμε στη συνέχεια περισσότερο σύνθετες περιπτώσεις όπου το μήκος της λύσης συνδυάζεται με άλλα κριτήρια για να μας δώσει βέλτιστη λύση στο αντίστροφο πρόβλημα.

Στην απλή περίπτωση που χρησιμοποιηθεί ο περιορισμός του ελάχιστου μήκους, μπορούμε να διατυπώσουμε ένα γραμμικό αντίστροφο πρόβλημα ως εξής :

- Υπολογίστε το διάνυσμα $\mathbf{m}^{est}$ που ελαχιστοποιεί τη νόρμα $L = \mathbf{m}^T \mathbf{m}$, όταν ισχύει επίσης $\mathbf{e} = \mathbf{d} - \mathbf{Gm} = 0$.

Στην ανωτέρω διατύπωση υπονοούμε ότι έχουμε ένα πρόβλημα καθαρά υπο-ορισμένο και γι αυτό οι λύσεις του προβλήματος ικανοποιούν ακριβώς το γραμμικό σύστημα $\mathbf{d} = \mathbf{Gm}$ που ορίζει το αντίστροφο πρόβλημα.

Ένα πρόβλημα αυτής της μορφής λύνεται με τη χρήση των πολλαπλασιαστών Lagrange. Το πρόβλημα ανάγεται στην ελαχιστοποίηση μιας κατάλληλης *αντικειμενικής συνάρτησης* που στην περίπτωσή μας ορίζεται ως :

$$\Phi(m) = L + \sum_{i=1}^N \lambda_i e_i = \sum_{j=1}^M m_j^2 + \sum_{i=1}^N \lambda_i \left[d_i - \sum_{j=1}^M G_{ij} m_j \right] \quad (3.28)$$

Παραγωγίζοντας τη συνάρτηση ως προς τις παραμέτρους m και ζητώντας την παράγωγο να είναι 0, παίρνομε ένα σύστημα ως προς τους πολλαπλασιαστές που επιλύόμενο μας δίνει ένα νέο σύστημα ως προς τις βέλτιστες λύσεις $\mathbf{m}^{est}$.

Ας δούμε την πορεία αυτή :

1. Παραγωγή ως προς όλα τα m

$$\frac{\partial \Phi}{\partial m_q} = \sum_{j=1}^M 2 \frac{\partial m_j}{\partial m_q} m_j - \sum_{i=1}^N \lambda_i \sum_{j=1}^M G_{ij} \frac{\partial m_j}{\partial m_q} = 2m_q - \sum_{i=1}^N \lambda_i G_{iq} = 0 \quad (3.29)$$

2. Γράφουμε τις ανωτέρω σχέσεις για $q = 1, \dots, M$ σε διανυσματική μορφή :

$$2\mathbf{m} = \mathbf{G}^T \boldsymbol{\lambda} \quad (3.30)$$

3. Επειδή ισχύει παράλληλα και $\mathbf{d} = \mathbf{G}\mathbf{m}$, αντικαθιστώντας το $\mathbf{m}$ από την 3.30 στην αρχική εξίσωση παίρνουμε :

$$\mathbf{d} = \mathbf{G} \left[\mathbf{G}^T \boldsymbol{\lambda} / 2 \right] \quad (3.31)$$

4. Εάν ο πίνακας $\mathbf{G}\mathbf{G}^T$ που έχει διαστάσεις $M \times M$ αντιστρέφεται, μπορούμε να εκφράσουμε μέσω αυτού τους πολλαπλασιαστές Lagrange.

$$\boldsymbol{\lambda} = 2 \left[\mathbf{G}\mathbf{G}^T \right]^{-1} \mathbf{d} \quad (3.32)$$

5. Με αντικατάσταση της 3.32 στην 3.30 παίρνουμε τη λύση (ελαχίστου μήκους)

$$\mathbf{m}^{est} = \mathbf{G}^T \left[\mathbf{G}\mathbf{G}^T \right]^{-1} \mathbf{d} \quad (3.33)$$

Ως προς την αντιστρεψιμότητα του πίνακα $\mathbf{G}\mathbf{G}^T$, από τη Γραμμική Άλγεβρα γνωρίζουμε ότι αρκεί οι γραμμές του $\mathbf{G}$ να είναι γραμμικά ανεξάρτητες. Αυτό όμως εξασφαλίζεται από την υπόθεση για ένα καθαρά υπο-ορισμένο αντίστροφο πρόβλημα χωρίς ασυνέπειες.

ΑΣΚΗΣΗ Επιλύστε το πρόβλημα 3.26 με την αρχή του ελαχίστου μήκους.

3.7 Μεικτά ορισμένα προβλήματα

Τα περισσότερα διακριτά αντίστροφα προβλήματα δεν είναι ούτε καθαρά υπο-ορισμένα, ούτε καθαρά υπερ-ορισμένα. Εάν υπήρχε η δυνατότητα οι άγνωστες παράμετροι να ομαδοποιηθούν έτσι ώστε να οριστούν δύο ομάδες από τις οποίες η πρώτη να έχει τις παραμέτρους που με βάση τις μετρήσεις είναι υπερ-ορισμένες και στην άλλη να έχουμε τις υπο-ορισμένες παραμέτρους, θα ήταν θεωρητικά εφικτό να δημιουργήσουμε ένα νέο γραμμικό σύστημα $\mathbf{G}\mathbf{m} = \mathbf{d} \rightarrow \mathbf{G}'\mathbf{m}' = \mathbf{d}'$ όπου ο νέος πίνακας $\mathbf{G}'$ μπορεί να διαχωριστεί ως :

$$\mathbf{G}' \rightarrow \begin{bmatrix} \mathbf{G}^o & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^u \end{bmatrix} \text{ έτσι ώστε το γραμμικό σύστημα να πάρει τη μορφή :}$$

$$\begin{bmatrix} \mathbf{G}^o & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^u \end{bmatrix} \begin{bmatrix} \mathbf{m}^o \\ \mathbf{m}^u \end{bmatrix} = \begin{bmatrix} \mathbf{d}^o \\ \mathbf{d}^u \end{bmatrix} \quad (3.34)$$

με τους δείκτες o και u να υποδηλώνουν το υπερ-ορισμένο (**overdetermined**) και υπο-ορισμένο (**underdetermined**) τμήμα του προβλήματος. Στην περίπτωση αυτή μπορεί

κανείς να λύσει το πρόβλημά του χωριστά για το υπερορισμένο και υποορισμένο τμήμα με τεχνικές όπως αυτές που αναφέραμε ήδη.

Μια δυνατότητα αυτής της μορφής δεν πάντα εύκολη. Μπορούμε όμως να ορίσουμε μια αντικειμενική συνάρτηση που να λαμβάνει υπ' όψιν της τόσο το κριτήριο των ελαχίστων τετραγώνων όσο και το κριτήριο του ελάχιστου μήκους (ή κάτι ανάλογο). Μια συνάρτηση αυτής της μορφής είναι η :

$$\Phi(\mathbf{m}) = E + \varepsilon^2 L = \mathbf{e}^T \mathbf{e} + \varepsilon^2 \mathbf{m}^T \mathbf{m} \quad (3.35)$$

όπου ο συντελεστής ε^2 ορίζει τη σχετική βαρύτητα που έχουν τα ελάχιστα τετράγωνα σε σχέση με το ελάχιστο μήκος. Εάν ο συντελεστής αυτός είναι μεγάλος, τονίζεται το υποορισμένο τμήμα του προβλήματος αφού η ελαχιστοποίηση της Φ μας οδηγεί σε λύσεις ελαχίστου μήκους. Εάν πάλι είναι μικρός, τονίζεται η λύση ελαχίστων τετραγώνων και εάν τείνει στο 0 δεν λαμβάνεται ουσιαστικά υπ' όψιν ο περιορισμός του μήκους των παραμέτρων. Είναι προφανές ότι από την εμπειρία σε ορισμένα προβλήματα μπορεί να οριστεί κατάλληλο ε ανά περίπτωση.

Παραγωγίζοντας την Φ ως προς τις παραμέτρους $\mathbf{m}$ με διαδικασία ανάλογη με αυτές που παρουσιάσαμε, οδηγούμαστε στην έκφραση της λύσης για τις παραμέτρους $\mathbf{m}$ ως

$$\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G} + \varepsilon^2 \mathbf{I}]^{-1} \mathbf{G}^T \mathbf{d} \quad (3.36)$$

όπου $\mathbf{I}$ είναι ο διαγώνιος μοναδιαίος πίνακας $M \times M$.

Η λύση αυτή ονομάζεται *λύση ελαχίστων τετραγώνων με απόσβεση (damped least square solution)* μια και μέσω του ε , έχει «αποσβεσθεί» η επίδραση του υποορισμένου τμήματος του προβλήματος.

3.8 Λύσεις αντιστρόφων προβλημάτων με χρήση ζυγισμένων μετρικών μήκους

Σε πολλές περιπτώσεις (μάλλον στις περισσότερες) η αρχή του ελαχίστου μήκους δεν είναι ικανοποιητική ακόμη και σε ένα καθαρά υπο-ορισμένο πρόβλημα αφού αναφέρεται σε ειδική περίπτωση φυσικής ερμηνείας των προς ανάκτηση παραμέτρων. Εάν όμως υπάρχει πληροφορία ότι οι προς ανάκτηση παράμετροι εκφράζουν ένα φυσικό μέγεθος του οποίου γνωρίζουμε (για παράδειγμα) τη μέση τιμή, ή μια τιμή κοντά στην πραγματική για όλες τις παραμέτρους, είναι φυσικό να περιμένουμε οι προς ανάκτηση παράμετροι να μην απέχουν πολύ από αυτή. Δημιουργούμε λοιπόν μία καινούργια νόρμα μήκους, L' οριζόμενη από τη σχέση :

$$L' = (\mathbf{m} - \langle \mathbf{m} \rangle)^T (\mathbf{m} - \langle \mathbf{m} \rangle) \quad (3.37)$$

όπου με $\langle \mathbf{m} \rangle$ συμβολίζουμε τις εκ προοιμίου (a priori) γνωστές τιμές της παραμέτρου $\mathbf{m}$. Η ελαχιστοποίηση της νέας αυτής νόρμας θα μας δώσει τις λύσεις που θα απέχουν το ελάχιστο από τις γνωστές τιμές.

Πρόσθετη πληροφορία μπορούμε να έχουμε σε ό,τι αφορά την ομαλότητα των παραμέτρων. Δηλαδή το πόσο μεταβάλλονται οι γειτονικές παράμετροι (m_j σε σχέση με την m_{j+1} αλλά και με την m_{j-1}). Ένα μέτρο για αυτή τη μεταβολή είναι η διαφορά $m_j - m_{j-1}$, Η διαφορά αυτή προκύπτει από την έννοια της σχέσης διαφορών πρώτης τάξης για διακριτές συναρτήσεις. Εάν επιθυμούμε ομαλότητα μεγαλύτερης τάξης, μπορούμε να ανατρέξουμε στις σχέσεις διαφορών δεύτερης τάξης που είναι $m_{j-1} - 2m_j + m_{j+1}$. Οι σχέσεις αυτές προέρχονται από αναπτύγματα Taylor που χρησιμοποιούνται για να προσεγγίσει κανείς πρώτες και δεύτερες παραγώγους συνεχών συναρτήσεων με διακριτό τρόπο. Μικρές τιμές παραγώγων λοιπόν σημαίνουν μικρές διαφορές και ομαλές μεταβολές των τιμών των συναρτήσεων. Στην περίπτωση μας μικρές διαφορές σημαίνουν ομαλές μεταβολές παραμέτρων.

Η ομαλότητα λοιπόν μπορεί να εκφραστεί μέσω σχέσεων της μορφής :

$$\mathbf{I} = \begin{bmatrix} -1 & 1 & 0 & . & . \\ 0 & -1 & 1 & 0 & . \\ & & . & . & . & . \\ & & & . & . & . & . \\ & & & & . & . & . & 0 \\ & & & & . & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ . \\ . \\ . \\ m_M \end{bmatrix} = \mathbf{Dm} \quad (3.38)$$

ή

$$\mathbf{I}' = \begin{bmatrix} 1 & -2 & 1 & 0 & . \\ 0 & 1 & -2 & 1 & 0 \\ & & . & . & . & . \\ & & & . & . & . & . \\ & & & & . & . & . & 0 \\ & & & & 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ . \\ . \\ . \\ m_M \end{bmatrix} = \mathbf{D}'\mathbf{m} \quad (3.39)$$

όπου $\mathbf{D}$ και $\mathbf{D}'$ είναι πίνακες ομαλότητας (flatness matrices) για τις περιπτώσεις που αναφέραμε παραπάνω.

Θεωρώντας τώρα μια καινούργια μέτρηση μήκους από τη σχέση :

$$L^w = \mathbf{I}^T \mathbf{I} = [\mathbf{Dm}]^T [\mathbf{Dm}] = \mathbf{m}^T \mathbf{D}^T \mathbf{Dm} = \mathbf{m}^T \mathbf{W}_m \mathbf{m} , \quad (3.40)$$

μπορούμε να την αντικαταστήσουμε στις προηγούμενες εκφράσεις που είδαμε και στις οποίες υπήρχε μέτρηση μήκους παραμέτρων, προκειμένου να εξασφαλίσουμε και την ισχύ του πρόσθετου περιορισμού της ομαλότητας. Η νέα όμως μέτρηση μήκους δεν είναι κανονική νόρμα, αφού μπορεί να πάρει μηδενική τιμή και για μη μηδενικά διανύσματα $\mathbf{m}$ (π.χ. για διανύσματα ίσων όρων). Ο πίνακας $\mathbf{W}_m = \mathbf{D}^T \mathbf{D}$ είναι ένας όρος ζύγισης (weighting factor). Ο όρος αυτός μπορεί να εισαχθεί και σε άλλες μετρήσεις μήκους όπως η L' (3.37), και να μας οδηγήσει σε λύσεις επαρκούς ομαλότητας κοντά σε αρχικές (γνωστές) τιμές παραμέτρων

$$L'^w = (\mathbf{m} - \langle \mathbf{m} \rangle)^T \mathbf{W}_m (\mathbf{m} - \langle \mathbf{m} \rangle) \quad (3.41)$$

Τέλος να σημειώσουμε ότι ανάλογος όρος ζύγισης μπορεί να εισαχθεί και στη λύση ελαχίστων τετραγώνων όταν υπάρχει πληροφορία π.χ. για αυξημένη ακρίβεια κάποιας μέτρησης. Τότε η μέτρηση αυτή μπορεί να ληφθεί περισσότερο υπ' όψιν με κατάλληλη ζύγιση της σημασίας της. Δημιουργούμε έτσι ένα καινούργιο διαγώνιο πίνακα με μη μηδενικά στοιχεία 1 και μεγαλύτερα (ή μικρότερα) του 1 ανάλογα με την αξιοπιστία των μετρήσεων. Ο όρος ζύγισης στην περίπτωση αυτή συμβολίζεται με $\mathbf{W}_e$ και το λάθος εκτίμησης γίνεται $E^w = \mathbf{e}^T \mathbf{W}_e \mathbf{e}$.

Στα επόμενα πινακοποιούμε τις λύσεις που έχουμε δει μέχρι τώρα όταν χρησιμοποιηθεί ο πίνακας ζύγισης.

<u>Λύση ελαχίστων τετραγώνων</u> $\implies$ Καθαρά υπερ-ορισμένο πρόβλημα Ελαχιστοποίηση της $E^w = \mathbf{e}^T \mathbf{W}_e \mathbf{e}$ $\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{W}_e \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{W}_e \mathbf{d}$
<u>Λύση ελαχίστου μήκους με a-priori πληροφορία</u> $\implies$ Καθαρά υπο-ορισμένο πρόβλημα Ελαχιστοποίηση της $L'^w = (\mathbf{m} - \langle \mathbf{m} \rangle)^T \mathbf{W}_m (\mathbf{m} - \langle \mathbf{m} \rangle)$ $\mathbf{m}^{est} = \langle \mathbf{m} \rangle + \mathbf{W}_m \mathbf{G}^T [\mathbf{G} \mathbf{W}_m \mathbf{G}^T]^{-1} [\mathbf{d} - \mathbf{G} \langle \mathbf{m} \rangle]$
<u>Λύση ελαχίστων τετραγώνων με a-priori πληροφορία και απόσβεση</u> $\implies$ Μεικτό πρόβλημα Ελαχιστοποίηση της $\Phi' = E^w + \varepsilon^2 L'^w$ $\mathbf{m}^{est} = \langle \mathbf{m} \rangle + \mathbf{W}_m^{-1} \mathbf{G}^T [\mathbf{G} \mathbf{W}_m^{-1} \mathbf{G}^T + \varepsilon^2 \mathbf{W}_e^{-1}]^{-1} [\mathbf{d} - \mathbf{G} \langle \mathbf{m} \rangle]$

Φυσικά όλες οι παραπάνω σχέσεις δίδονται με την υπόθεση ότι υπάρχουν οι αναφερόμενοι αντίστροφοι πίνακες.

3.9 Άλλοι τύποι εκ προοιμίου (a-priori) πληροφορίας.

Υπάρχουν περιπτώσεις αντιστρόφων προβλημάτων στις οποίες γνωρίζουμε ότι ισχύει μία σχέση ανάμεσα στις παραμέτρους. Σχέσεις της μορφής $\mathbf{Fm} = \mathbf{h}$ είναι σχετικά εύκολο να εισαχθούν στο πρόβλημα. Εάν για παράδειγμα δεχτούμε ότι ο μέσος όρος των παραμέτρων είναι γνωστός και παίρνει την τιμή h_1 έχουμε :

$$\mathbf{Fm} = \frac{1}{M} \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_M \end{bmatrix} = [h_1] = \mathbf{h} \quad (3.42)$$

Μια άλλη ανάλογη περίπτωση είναι όταν γνωρίζουμε ότι μια συγκεκριμένη παράμετρος (ας πούμε η m_j) παίρνει δεδομένη τιμή. Αυτό το πρόβλημα δεν μπορεί πολλές φορές να αντιμετωπισθεί με εξαίρεση της δεδομένης παραμέτρου από την διατύπωση του αντιστρόφου προβλήματος, όπως θα μπορούσε να φανταστεί κανείς. Στην περίπτωση αυτή, έχουμε μια απλή διατύπωση της γραμμικής σχέσης :

$$\mathbf{Fm} = \begin{bmatrix} 0 & 0 & \dots & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_j \\ \vdots \\ m_M \end{bmatrix} = [h_1] = \mathbf{h} \quad (3.43)$$

Πολλές φορές ζητάμε την επίλυση ενός αντίστροφου γραμμικού προβλήματος με την αρχή των ελαχίστων τετραγώνων, γνωρίζοντας ότι θα πρέπει οι παράμετροι να σχετίζονται με σχέσεις της μορφής $\mathbf{Fm} = \mathbf{h}$. Αυτού του είδους τα προβλήματα μπορούμε να τα αντιμετωπίσουμε είτε προσθέτοντας τις παραπάνω εξισώσεις ως γραμμές στο σύστημα $\mathbf{Gm} = \mathbf{d}$ και ορίζοντας τον πίνακα ζύγισης $\mathbf{W}_e$ ώστε αυτές οι εξισώσεις να έχουν περισσότερο βάρος (ιδανικά άπειρο) σε σχέση με τις υπόλοιπες, είτε με τους πολλαπλασιαστές Lagrange. Θα αναφερθούμε στην τελευταία αυτή περίπτωση :

Ζητάμε την επίλυση ενός αντιστρόφου γραμμικού προβλήματος έτσι ώστε να ελαχιστοποιείται το «λάθος» $E = \mathbf{e}^T \mathbf{e}$ με τον περιορισμό : $\mathbf{Fm} - \mathbf{h} = 0$. Ορίζουμε τη συνάρτηση :

$$\Phi(\mathbf{m}) = \sum_{i=1}^N \left[d_i - \sum_{j=1}^M G_{ij} m_j \right]^2 + 2 \sum_{i=1}^p \lambda_i \left[\sum_{j=1}^M F_{ij} m_j - h \right], \quad (3.44)$$

όπου έχουμε p περιορισμούς και λ_i είναι οι πολλαπλασιαστές Lagrange.

Παραγωγίζοντας ως προς τις παραμέτρους και ζητώντας οι παράγωγοι να είναι 0, παίρνομε :

$$\frac{\partial \Phi(\mathbf{m})}{\partial m_q} = 2 \sum_{i=1}^M m_i \sum_{j=1}^N G_{jq} G_{ji} - 2 \sum_{i=1}^N G_{iq} d_i + 2 \sum_{i=1}^p \lambda_i F_{iq} = 0, \quad q = 1, \dots, M. \quad (3.45)$$

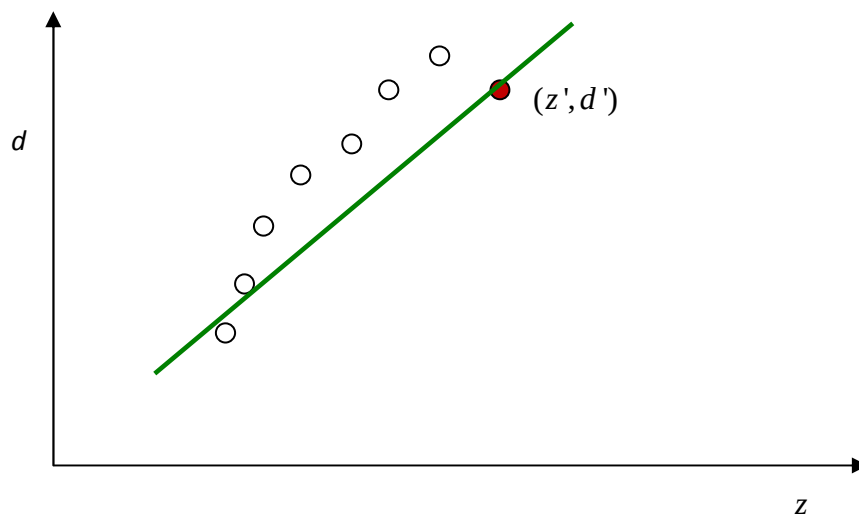
Στις ανωτέρω M εξισώσεις προστίθενται και οι p εξισώσεις των περιορισμών για να πάρουμε το σύστημα :

$$\begin{bmatrix} \mathbf{G}^T \mathbf{G} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{m} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{G}^T \mathbf{d} \\ \mathbf{h} \end{bmatrix} \quad (3.46)$$

που επιλυόμενο μας δίδει τόσο τις προς ανάκτηση παραμέτρους όσο και τους πολλαπλασιαστές Lagrange.

ΠΑΡΑΔΕΙΓΜΑ Υπολογισμός της εξίσωσης ευθείας όταν αυτή περνά από δεδομένο σημείο.

Θεωρείστε το πρόβλημα του υπολογισμού των παραμέτρων της εξίσωσης ευθείας $y = m_1 + m_2 z$ όταν γνωρίζομε ότι αυτή θα πρέπει να περνά από το ζεύγος των σημείων (z', d') , όπως αποτυπώνεται στο σχήμα 3.4



Σχήμα 3.4 Υπολογισμός εξίσωσης ευθείας που περνά από το σημείο (z', d')

Στην περίπτωσή μας έχουμε $p=1$ περιορισμό της μορφής :

$$d' = m_1 + m_2 z' \quad (3.47)$$

που γράφεται ως :

$$\mathbf{Fm} = \begin{bmatrix} 1 & z' \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = [d'] = h. \quad (3.48)$$

Η εξίσωση 3.46 στην περίπτωση αυτή χρησιμοποιώντας τους πίνακες που είδαμε στην παράγραφο 3.4.1 γράφεται ως :

$$\begin{bmatrix} N & \sum z_i & 1 \\ \sum z_i & \sum z_i^2 & z' \\ 1 & z' & 0 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} \sum d_i \\ \sum z_i d_i \\ d' \end{bmatrix} \quad (3.49)$$

και ο υπολογισμός των παραμέτρων και του μοναδικού συντελεστή Lagrange δίδεται από τη σχέση :

$$\begin{bmatrix} m_1^{est} \\ m_2^{est} \\ \lambda \end{bmatrix} = \begin{bmatrix} N & \sum z_i & 1 \\ \sum z_i & \sum z_i^2 & z' \\ 1 & z' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \sum d_i \\ \sum z_i d_i \\ d' \end{bmatrix} \quad (3.50)$$

3.10 Πίνακας συνδιακύμανσης των παραμέτρων

Όταν είναι γνωστά τα στατιστικά χαρακτηριστικά των μετρήσεων, μπορούμε να εκτιμήσουμε ανάλογα χαρακτηριστικά των παραμέτρων παρατηρώντας ότι οι σχέσεις που συνδέουν παραμέτρους και μετρήσεις είναι γραμμικές, της μορφής

$$\mathbf{m}^{est} = \mathbf{M}\mathbf{d} + \mathbf{v} \quad , \quad (3.51)$$

με $\mathbf{M}$ κάποιο πίνακα και $\mathbf{v}$ ένα διάνυσμα. Εάν υποθέσουμε λοιπόν ότι τα δεδομένα έχουν μία κατανομή που χαρακτηρίζεται από ένα πίνακα συνδιακύμανσης $[\text{cov } \mathbf{d}]$ οι εκτιμήσεις των παραμέτρων θα έχουν ένα πίνακα συνδιακύμανσης

$$[\text{cov } \mathbf{m}] = \mathbf{M}[\text{cov } \mathbf{d}]\mathbf{M}^T \quad (3.52)$$

και συνεπώς ανάλογα με την περίπτωση να υπολογίσουμε την συνδιακύμανση των παραμέτρων με άμεσο τρόπο. Η απλούστερη περίπτωση είναι εκείνη για την οποία τα δεδομένα είναι ασυσχέτιστα και έχουν όλα την ίδια διακύμανση σ_d^2 . Τότε $[\text{cov } \mathbf{d}] = \sigma_d^2 \mathbf{I}$ και ανάλογα με την περίπτωση υπολογίζουμε τους πίνακες

συνδιακύμανσης. Θα αναφερθούμε στη συνέχεια ως παραδείγματα σε δύο περιπτώσεις :

Λύση ελαχίστων τετραγώνων :

Ισχύει : $\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d}$ συνεπώς $\mathbf{M} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T$ και από την 3.52 παίρνουμε :

$$[\text{cov} \mathbf{m}] = [[\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T] \sigma_d^2 \mathbf{I} [[\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T]^T = \sigma_d^2 [\mathbf{G}^T \mathbf{G}]^{-1} \quad (3.53)$$

Λύση ελαχίστου μήκους :

Ισχύει : $\mathbf{m}^{est} = \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1} \mathbf{d}$ συνεπώς $\mathbf{M} = \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1}$ και από την 3.52 παίρνουμε :

$$[\text{cov} \mathbf{m}] = [\mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1}] \sigma_d^2 \mathbf{I} [\mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1}]^T = \sigma_d^2 \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1} \mathbf{G} \quad (3.54)$$

4 ΓΕΝΙΚΕΥΜΕΝΟΙ ΑΝΤΙΣΤΡΟΦΟΙ ΠΙΝΑΚΕΣ

Μέχρι τώρα, ασχοληθήκαμε με λύσεις γραμμικών αντιστρόφων προβλημάτων που εκφράζονται με σχέσεις της μορφής : $\mathbf{m}^{est} = \mathbf{M}\mathbf{d} + \mathbf{v}$. Ο πίνακας $\mathbf{M}$ και το διάνυσμα $\mathbf{v}$ δεν εξαρτώνται από τα δεδομένα. Ειδικά για τον πίνακα $\mathbf{M}$ μπορούμε να πούμε ότι λειτουργεί ως πίνακας-τελεστής που πολλαπλασιάζει τα δεδομένα για να προκύψουν οι προς ανάκτηση παράμετροι. Με αυτή την έννοια μπορεί να ονομαστεί *γενικευμένος αντίστροφος* (*gemeralized inverse*) καθώς λειτουργεί με ανάλογο τρόπο που σε ένα καλώς τεθειμένο γραμμικό πρόβλημα της μορφής $\mathbf{A}\mathbf{x} = \mathbf{d}$ πολλαπλασιάζει ο $\mathbf{A}^{-1}$ το διάνυσμα των δεδομένων για να προκύψει το $\mathbf{x}$: $\mathbf{x} = \mathbf{A}^{-1}\mathbf{d}$. Ο συμβολισμός του είναι $\mathbf{M} = \mathbf{G}^{-g}$ αφού αναφέρεται στο γενικό αντίστροφο πρόβλημα $\mathbf{G}\mathbf{m} = \mathbf{d}$. Η μορφή του γενικευμένου αντιστρόφου όμως εξαρτάται από την λύση που έχουμε επιλέξει. Ως εφαρμογές αναφερόμαστε στις περιπτώσεις ελαχίστων τετραγώνων και ελαχίστου μήκους που είδαμε στο ακριβώς προηγούμενο Κεφάλαιο, απλά για να εισαγάγουμε τον συμβολισμό του γενικευμένου αντιστρόφου στον πίνακα $\mathbf{M}$.

Λύση ελαχίστων τετραγώνων :

$$\text{Ισχύει : } \mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d} \text{ συνεπώς } \mathbf{G}^{-g} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T = \mathbf{M}$$

Λύση ελαχίστου μήκους :

$$\text{Ισχύει } \mathbf{m}^{est} = \mathbf{G}^T [\mathbf{G}\mathbf{G}^T]^{-1} \mathbf{d} \text{ συνεπώς } \mathbf{G}^{-g} = \mathbf{G}^T [\mathbf{G}\mathbf{G}^T]^{-1} = \mathbf{M}$$

Αξίζει να παρατηρήσουμε ότι ο γενικευμένος αντίστροφος δεν είναι τετραγωνικός πίνακας (με την εξαίρεση ενός καλώς τεθειμένου προβλήματος) και φυσικά δεν διαθέτει τις ιδιότητες ενός αντίστροφου πίνακα.

4.1 Πίνακας ανάλυσης δεδομένων

Ας υποθέσουμε ότι σε κάποιο γραμμικό αντίστροφο πρόβλημα της μορφής $\mathbf{G}\mathbf{m} = \mathbf{d}$ έχουμε ορίσει ένα γενικευμένο αντίστροφο πίνακα $\mathbf{G}^{-g}$ οπότε οι προς ανάκτηση παράμετροι υπολογίζονται από τη σχέση : $\mathbf{m}^{est} = \mathbf{G}^{-g}\mathbf{d}$. Εάν χρησιμοποιήσουμε το διάνυσμα αυτό στην εξίσωση ορισμού του αντιστρόφου προβλήματος θα πάρουμε ένα διάνυσμα $\mathbf{d}$ το οποίο όπως έχουμε δει το συμβολίζουμε με $\mathbf{d}^{pre}$ και είναι διαφορετικό από το διάνυσμα των μετρήσεων για το οποίο χρησιμοποιούμε το συμβολισμό $\mathbf{d}^{obs}$. Τα στοιχεία του $\mathbf{d}$ θα τα χαρακτηρίσουμε «προβλέψεις» (*predictions*) των μετρήσεων. Το διάνυσμα των προβλέψεων λοιπόν με βάση τα παραπάνω θα δίδεται από τη σχέση :

$$\mathbf{d}^{pre} = \mathbf{G}\mathbf{m}^{est} = \mathbf{G}[\mathbf{G}^{-g}\mathbf{d}^{obs}] = [\mathbf{G}\mathbf{G}^{-g}]\mathbf{d}^{obs} = \mathbf{N}\mathbf{d}^{obs} \quad (4.1)$$

Ο πίνακας $\mathbf{N} = \mathbf{G}\mathbf{G}^{-g}$ είναι τετραγωνικός πίνακας $N \times N$ και ονομάζεται «πίνακας ανάλυσης δεδομένων» (*data resolution matrix*). Ο πίνακας αυτός περιγράφει το πόσο καλά οι παράμετροι που ανακτήθηκαν δίδουν προβλέψεις μετρήσεων που προσεγγίζουν τις πραγματικές μετρήσεις. Προφανώς όταν $\mathbf{N} = \mathbf{I}$ τότε και $\mathbf{d}^{pre} = \mathbf{d}^{obs}$, και το λάθος πρόβλεψης είναι μηδενικό.

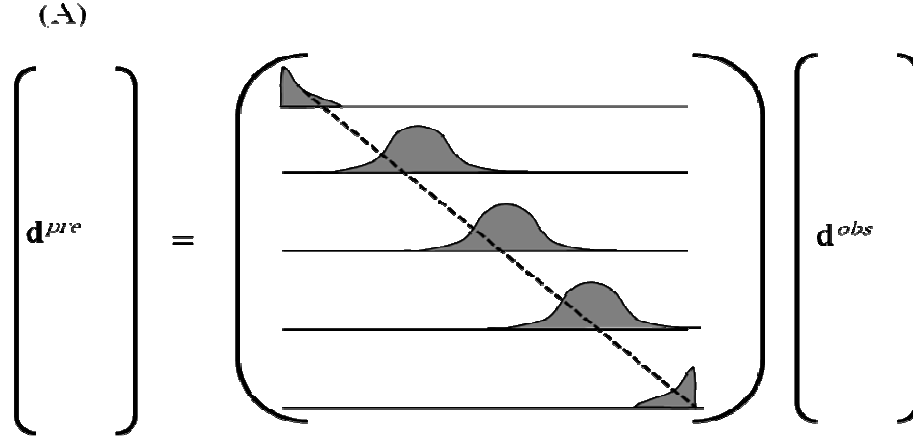
Τα στοιχεία του πίνακα ανάλυσης δεδομένων μας δίδουν σημαντική πληροφορία σχετικά με τη σημασία «γειτονικών» μετρήσεων σε σχέση με την ανάκτησή τους μέσω του διανύσματος των εκτιμηθέντων παραμέτρων. Εάν για παράδειγμα για κάποια μέτρηση με δείκτη i έχουμε το στοιχείο του πίνακα $N_{ii} = 1$, τότε η μέτρηση ανακτάται ακριβώς από τις εκτιμηθείσες παραμέτρους. Σε διαφορετική περίπτωση θα ισχύσει μία σχέση της μορφής :

$$d_i^{pre} = \sum_{j=1}^N N_{ij} d_j^{obs} \quad (4.2)$$

και η εκτίμηση της μέτρησης i θα προκύψει ως ζυγισμένη μέση τιμή όλων των μετρήσεων. Ιδανικά ο πίνακας $\mathbf{N}$ στην περίπτωση που οι μετρήσεις αφορούν διαδοχικές φυσικές μεταβλητές (όπως π.χ. στην περίπτωση του υπολογισμού της εξίσωσης ευθείας όταν οι μετρήσεις γίνονται σε διαδοχικές τετμημένες,) είναι λωριδωτός με κυριαρχική διαγώνιο και μειούμενες τιμές εκτός διαγωνίου (δείτε σχήμα 4.1) οι μετρήσεις ανακτώνται ικανοποιητικά και η ανάλυση θεωρείται καλή. Εάν όμως υπάρχει μεγάλη διασπορά γύρω από την διαγώνιο, η ανάκτηση των μετρήσεων δεν θεωρείται ικανοποιητική.

Μερικές φορές το διάνυσμα της διαγωνίου του πίνακα $\mathbf{N}$, $\mathbf{d} = \text{diag}(\mathbf{N})$ χαρακτηρίζεται ως *σημαντικότητα (importance)* των δεδομένων.

Αξίζει να παρατηρήσουμε ότι ο πίνακας ανάλυσης δεδομένων δεν σχετίζεται με τις πραγματικές τιμές των μετρήσεων, αλλά μόνο με το μοντέλο του προβλήματος, δηλαδή με τον πίνακα $\mathbf{G}$ καθώς και με κάθε περιορισμό η εκ προοιμίου πληροφορία διαθέτουμε για τη λύση του αντιστρόφου προβλήματος. Υπολογίζεται συνεπώς πριν γίνουν οι μετρήσεις και μπορεί να αποδειχθεί ιδιαίτερα χρήσιμος στο σχεδιασμό ενός πειράματος, όπως θα δούμε με ένα παράδειγμα στη συνέχεια.



Σχήμα 4.1 Σχηματική διάταξη των στοιχείων του πίνακα ανάλυσης δεδομένων. Η διακεκομμένη γραμμή αντιστοιχεί στη διαγώνιο του πίνακα. (Σχήμα από παραδόσεις του Menke).

4.2 Πίνακας ανάλυσης παραμέτρων

Εντελώς ανάλογα μπορούμε να δούμε την ποιότητα των εκτιμηθέντων παραμέτρων ως προς τις «πραγματικές» παραμέτρους ενός προβλήματος. Οι πραγματικές παράμετροι σε ένα ντετερμινιστικό φυσικό πρόβλημα ασφαλώς υπάρχουν αλλά δεν μπορούμε να είμαστε σίγουροι ότι ανακτώνται επακριβώς μέσω της επίλυσης ενός αντιστρόφου προβλήματος για το οποίο έχουμε πει ότι γενικά δεν είναι καλώς τεθειμένο. Έχει λοιπόν σημασία να δούμε εκ προοιμίου εάν η ανάκτηση που επιχειρούμε μπορεί να μας δώσει πληροφορία για την ποιότητά της. Με άλλα λόγια να δούμε εάν οι διακριτές παράμετροι ενός αντίστροφου προβλήματος μπορούν να ανακτηθούν επακριβώς.

Η λογική είναι αντίστοιχη με εκείνη που είδαμε στο προηγούμενο κεφάλαιο. Θα ξεκινήσουμε θεωρώντας ότι υπάρχει ένα διάνυσμα $\mathbf{m}^{true}$ πραγματικών παραμέτρων που πρέπει να λύνουν το πρόβλημα $\mathbf{G}\mathbf{m}^{true} = \mathbf{d}^{obs}$. Δεδομένου ότι η λύση του αντίστροφου προβλήματος δίδει εκτιμήσεις των παραμέτρων $\mathbf{m}^{est}$ με κάποιο τρόπο που έχουμε επιλέξει, έχει ενδιαφέρον να δούμε την εκτιμώμενη διαφορά ανάμεσα στις πραγματικές και τις εκτιμώμενες παραμέτρους.

Ακολουθώντας τη λογική που μας οδήγησε στην 4.1 έχουμε :

$$\mathbf{m}^{est} = \mathbf{G}^{-g} \mathbf{d}^{obs} = \mathbf{G}^{-g} [\mathbf{G}\mathbf{m}^{true}] = [\mathbf{G}^{-g}\mathbf{G}]\mathbf{m}^{true} = \mathbf{R}\mathbf{m}^{true} \quad (4.3)$$

Τώρα έχουμε ορίσει ένα πίνακα διαστάσεων $M \times M$ που χαρακτηρίζεται ως «πίνακας ανάλυσης παραμέτρων» (*model resolution matrix*). Εάν $\mathbf{R}=\mathbf{I}$ κάθε παράμετρος υπολογίζεται μονοσήμαντα. Εάν ο $\mathbf{R}$ είναι ένας πυκνός πίνακας, τότε οι μεμονωμένες παράμετροι δεν μπορούν να υπολογιστούν μονοσήμαντα. Τέλος, εάν οι παράμετροι έχουν κατά σειρά κάποια φυσική σημασία (π.χ. εάν πρόκειται για διακριτές τιμές μιας φυσικής παραμέτρου που εκφράζεται ως συνεχής συνάρτηση), διαγράμματα που παριστούν τις μεταβολές τιμών των σειρών του πίνακα $\mathbf{R}$, υποδηλώνουν τη δυνατότητα αξιόπιστης ανάκτησης της τιμής κάποιας παραμέτρου αλλά και τη σχέση της με τις γειτονικές τιμές. Έτσι μπορούμε να φανταστούμε κάποιο σχήμα ανάλογο του 4.1 με τα διανύσματα να αντιστοιχούν στις παραμέτρους.

Επισημαίνουμε ότι και στην παρούσα περίπτωση, ο πίνακας ανάλυσης παραμέτρων δεν εξαρτάται από τις μετρήσεις.

4.3 Ο μοναδιαίος πίνακας συνδιακύμανσης

Εάν εισάγουμε πιθανοθεωρητικούς όρους στο πρόβλημά μας, έχει νόημα να συσχετίσουμε στατιστικά χαρακτηριστικά των μετρήσεων (παρατηρήσεων) με εκείνα των προς ανάκτηση παραμέτρων. Με τον τρόπο αυτό μπορούμε να εκτιμήσουμε το πώς μεταφέρονται λάθη ή αβεβαιότητες στις τιμές των μετρήσεων στις τιμές των προς ανάκτηση παραμέτρων.

Θυμίζουμε τα αναφερθέντα στο κεφάλαιο 2.3 σε ό,τι αφορά τον υπολογισμό του πίνακα συνδιακύμανσης : Όταν παράμετροι και δεδομένα σχετίζονται με γραμμικές σχέσεις της μορφής $\mathbf{m} = \mathbf{M}\mathbf{d} + \mathbf{v}$, ο πίνακας συνδιακύμανσης των παραμέτρων παίρνει τη μορφή : $[\text{cov}\mathbf{m}] = \mathbf{M}[\text{cov}\mathbf{d}]\mathbf{M}^T$. Επιλύοντας ένα γραμμικό αντίστροφο πρόβλημα και χρησιμοποιώντας την έννοια του γενικευμένου αντίστροφου πίνακα, έχουμε μια σχέση της μορφής $\mathbf{m}^{est} = \mathbf{G}^{-g}\mathbf{d}$ και συνεπώς μπορούμε να εκφράσουμε τον πίνακα συνδιακύμανσης των παραμέτρων από τη σχέση $[\text{cov}\mathbf{m}] = \mathbf{G}^{-g}[\text{cov}\mathbf{d}]\mathbf{G}^{-gT}$. Στην πραγματικότητα «κανονικοποιούμε» τον πίνακα αυτό και ορίζουμε τον «μοναδιαίο πίνακα συνδιακύμανσης» διαιρώντας με την διασπορά των μετρήσεων την οποία θεωρούμε ίση για όλες τις μετρήσεις με σ^2 και τις μετρήσεις ασυσχέτιστες μεταξύ τους. Έτσι προκύπτει η σχέση :

$$[\text{cov}_u \mathbf{m}] = \sigma^{-2} \mathbf{G}^{-g} [\text{cov}\mathbf{d}] \mathbf{G}^{-gT} = \mathbf{G}^{-g} \mathbf{G}^{-gT} \quad (4.4)$$

που ορίζει τον μοναδιαίο πίνακα συνδιακύμανσης. Εάν οι μετρήσεις μας παρουσιάζουν συσχέτιση, μπορεί να οριστεί αντίστοιχα ένας *μοναδιαίος πίνακας συνδιακύμανσης μετρήσεων* (*unit data covariance matrix*) $[\text{cov}_u \mathbf{d}]$ και να ορίσουμε τον μοναδιαίο πίνακα συνδιακύμανσης παραμέτρων από τη σχέση :

$$[\text{cov}_u \mathbf{m}] = \mathbf{G}^{-g} [\text{cov}_u \mathbf{d}] \mathbf{G}^{-gT} \quad (4.5)$$

Ο μοναδιαίος πίνακας συνδιακύμανσης παραμέτρων, δεν εξαρτάται φυσικά από τις μετρήσεις αυτές καθ' εαυτές αλλά μόνο από τον γενικευμένο αντίστροφο πίνακα και από τα στατιστικά χαρακτηριστικά των μετρήσεων.

4.4 Ανάλυση και συνδιακύμανση για ορισμένες περιπτώσεις γενικευμένων αντιστροφών

Θα δώσουμε τώρα τη μορφή των πινάκων ανάλυσης και συνδιακύμανσης για τις δύο βασικές λύσης του γραμμικού αντίστροφου προβλήματος που είναι η των ελαχίστων τετραγώνων για προβλήματα καθαρά υπερ-ορισμένα και ελαχίστου μήκους για προβλήματα καθαρά υπο-ορισμένα. Θα υποθέσουμε για λόγους απλότητας ότι $[\text{cov}_u \mathbf{d}] = \mathbf{I}$

Λύση ελαχίστων τετραγώνων :

$$\mathbf{G}^{-g} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T$$

$$\mathbf{N} = \mathbf{G} \mathbf{G}^{-g} = \mathbf{G} [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T$$

$$\mathbf{R} = \mathbf{G}^{-g} \mathbf{G} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{G} = \mathbf{I}$$

$$[\text{cov}_u \mathbf{m}] = \mathbf{G}^{-g} \mathbf{G}^{-gT} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{G} [\mathbf{G}^T \mathbf{G}]^{-1} = [\mathbf{G}^T \mathbf{G}]^{-1}$$

Λύση ελαχίστου μήκους

$$\mathbf{G}^{-g} = \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1}$$

$$\mathbf{N} = \mathbf{G} \mathbf{G}^{-g} = \mathbf{G} \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1} = \mathbf{I}$$

$$\mathbf{R} = \mathbf{G}^{-g} \mathbf{G} = \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1} \mathbf{G}$$

$$[\text{cov}_u \mathbf{m}] = \mathbf{G}^{-g} \mathbf{G}^{-gT} = \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1} [\mathbf{G} \mathbf{G}^T]^{-1} \mathbf{G} = \mathbf{G}^T \left([\mathbf{G} \mathbf{G}^T]^{-1} \right)^2 \mathbf{G}$$

Παρατηρούμε τη συμμετρία ανάμεσα στις δύο περιπτώσεις με τη λύση ελαχίστων τετραγώνων να μας δίνει μοναδιαίο πίνακα ανάλυσης παραμέτρων (η εκτίμηση των παραμέτρων γίνεται με τέλεια ανάλυση – αυτό δεν σημαίνει πάντως ότι ανακτήθηκαν με ακρίβεια οι παράμετροι), ενώ η λύση ελαχίστου μήκους δίνει μοναδιαίο πίνακα ανάλυσης δεδομένων. Ενδιάμεσες περιπτώσεις ασφαλώς θα δίδουν και ενδιάμεσους πίνακες ανάλυσης δεδομένων και παραμέτρων.

4.5 Μετρήσεις ποιότητας της ανάλυσης και της συνδιακύμανσης.

Έχοντας δώσει τους ορισμούς των πινάκων ανάλυσης και συνδιακύμανσης, έχει νόημα να δούμε εάν μπορούμε να τους χαρακτηρίσουμε ως προς την ποιότητά τους σε σχέση με αυτό που εκφράζουν. Ασφαλώς η ποιότητα θεωρείται άριστη όταν οι πίνακες είναι μοναδιαίοι. Κάθε απόκλιση από τους μοναδιαίους θεωρείται ότι εκφράζει μείωση της ποιότητας επίλυσης του αντιστρόφου προβλήματος. Η ποιότητα λοιπόν εκφράζεται μέσω της διαφοράς των στοιχείων από τα αντίστοιχα στοιχεία του μοναδιαίου πίνακα και μας οδηγεί στον ορισμό του «εύρους» του πίνακα ανάλυσης δεδομένων και μετρήσεων που εκφράζεται από την ευκλείδεια νόρμα :

$$\text{spread}(\mathbf{N}) = \|\mathbf{N} - \mathbf{I}\|_2^2 = \sum_{i=1}^N \sum_{j=1}^N [N_{ij} - I_{ij}]^2 \quad (4.6)$$

$$\text{spread}(\mathbf{R}) = \|\mathbf{R} - \mathbf{I}\|_2^2 = \sum_{i=1}^N \sum_{j=1}^N [R_{ij} - I_{ij}]^2 \quad (4.7)$$

Επίσης ορίζεται το μέγεθος του μοναδιαίου πίνακα συνδιακύμανσης από τη σχέση

$$\text{size}([\text{cov}_u \mathbf{m}]) = \|[\text{var}_u \mathbf{m}]^{1/2}\|_2^2 = \sum_{i=1}^M [\text{cov}_u \mathbf{m}]_{ii} \quad (4.8)$$

η οποία δεν λαμβάνει υπ' όψιν της τα στοιχεία εκτός διαγωνίου.

4.6 Μία εφαρμογή: Ανάκτηση της εξίσωσης ευθείας

Ως εφαρμογή των ανωτέρω, ας δούμε ένα απλό παράδειγμα υπολογισμού της ποιότητας της ανάκτησης δεδομένων και παραμέτρων όταν έχουμε ένα καθαρά υπερορισμένο πρόβλημα, όπως αυτό της ανάκτησης της εξίσωσης ευθείας από μετρήσεις. Θα υποθέσουμε ότι οι μετρήσεις έχουν ίση διασπορά και ότι είναι ασυσχέτιστες μεταξύ τους.

Έστω ότι οι μετρήσεις γίνονται σε μια πρώτη περίπτωση στις τετμημένες 1,2,3 ($N=3$) και σε μία δεύτερη περίπτωση στις τετμημένες 1, 2, 4 ($N=3$). Πρόκειται ουσιαστικά για δύο διαφορετικά πειράματα για την εκτίμηση των δύο παραμέτρων που περιγράφουν την ευθεία και θα προσπαθήσουμε να δούμε ποιοτικές διαφορές μεταξύ τους, χωρίς να αναφερθούμε καθόλου σε μεγέθη μετρήσεων.

Το πρόβλημα είναι καθαρά υπερορισμένο και γνωρίζουμε ήδη ότι η λύση ελαχίστων τετραγώνων στη γενική περίπτωση είναι

$$\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d} = \begin{bmatrix} N & \Sigma z_i \\ \Sigma z_i & \Sigma z_i^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_N \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} \quad (4.9)$$

$$\text{Ο γενικευμένος αντίστροφος είναι } \mathbf{G}^{-g} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \quad (4.10)$$

Εξειδικεύοντας τα παραπάνω στην περίπτωση των τριών μετρήσεων, έχουμε :

$$\mathbf{m}^{est} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d} = \begin{bmatrix} 3 & \Sigma z_i \\ \Sigma z_i & \Sigma z_i^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \quad (4.11)$$

και

$$\mathbf{G}^{-g} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T = \frac{1}{3\Sigma z_i^2 - (\Sigma z_i)^2} \begin{bmatrix} \Sigma z_i^2 & -\Sigma z_i \\ -\Sigma z_i & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \end{bmatrix} \quad (4.12)$$

Θα αναλύσουμε την πρώτη περίπτωση :

Έχουμε : $\Sigma z_i = 6$, $\Sigma z_i^2 = 14$, $(\Sigma z_i)^2 = 36$, επομένως

$$\mathbf{G}^{-g} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 8 & 2 & -4 \\ -3 & 0 & 3 \end{bmatrix} \quad (4.13)$$

Ο πίνακας ανάλυσης δεδομένων είναι :

$$\mathbf{N} = \mathbf{G} \mathbf{G}^{-g} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \frac{1}{6} \begin{bmatrix} 8 & 2 & -4 \\ -3 & 0 & 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix} \quad (4.14)$$

και ο πίνακας ανάλυσης μετρήσεων :

$$\mathbf{R} = \mathbf{G}^{-g} \mathbf{G} = \frac{1}{6} \begin{bmatrix} 8 & 2 & -4 \\ -3 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \quad (4.15)$$

Από τον πίνακα ανάλυσης δεδομένων, βλέπουμε ότι η πρώτη και η τρίτη μέτρηση είναι πιο σημαντικές αφού αυτές ουσιαστικά καθορίζουν το αποτέλεσμα. Από τον πίνακα ανάλυσης παραμέτρων βλέπουμε ότι οι δύο άγνωστες παράμετροι ανακτώνται ανεξάρτητα η μία της άλλης.

Υπολογίζοντας τον μοναδιαίο πίνακα συνδιακύμανσης έχουμε ότι

$$[\text{cov}_u \mathbf{m}] = [\mathbf{G}^T \mathbf{G}]^{-1} = \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} = \begin{bmatrix} 14/6 & -1 \\ -1 & 1/2 \end{bmatrix} \quad (4.16)$$

και

$$\text{size}([\text{cov}_u \mathbf{m}]) = \sum_{i=1}^M [\text{cov}_u \mathbf{m}]_{ii} = \frac{17}{6} \quad (4.17)$$

Αναλύοντας την δεύτερη περίπτωση έχουμε αντίστοιχα :

$$\Sigma z_i = 7, \Sigma z_i^2 = 21, (\Sigma z_i)^2 = 49, \text{επομένως}$$

$$\mathbf{G}^{-g} = \frac{1}{14} \begin{bmatrix} 21 & -7 \\ -7 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 14 & 7 & -7 \\ -4 & -1 & 5 \end{bmatrix} \quad (4.18)$$

Ο πίνακας ανάλυσης δεδομένων είναι :

$$\mathbf{N} = \mathbf{G} \mathbf{G}^{-g} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} \frac{1}{14} \begin{bmatrix} 14 & 7 & -7 \\ -4 & -1 & 5 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 10 & 6 & -2 \\ 6 & 5 & 3 \\ -2 & 3 & 13 \end{bmatrix} \quad (4.19)$$

και ο πίνακας ανάλυσης μετρήσεων :

$$\mathbf{R} = \mathbf{G}^{-g} \mathbf{G} = \frac{1}{14} \begin{bmatrix} 14 & 7 & -7 \\ -4 & -1 & -5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 14 & 0 \\ 0 & 14 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \quad (4.20)$$

Από τον πίνακα ανάλυσης δεδομένων, βλέπουμε ότι η πρώτη και η τρίτη μέτρηση είναι επίσης οι πιο σημαντικές και ουσιαστικά καθορίζουν το αποτέλεσμα. Ωστόσο, σε σχέση με την προηγούμενη περίπτωση, η τρίτη μέτρηση που είναι πιο απομακρυσμένη από τις άλλες είναι πιο σημαντική, αφού τώρα $N_{3,3} = 13/14 > 5/6$ ενώ τα άλλα δύο στοιχεία της $3^{\text{ης}}$ γραμμής του πίνακα είναι μικρότερα έναντι της προηγούμενης περίπτωσης. Από τον πίνακα ανάλυσης παραμέτρων βλέπουμε ότι οι

δύο άγνωστες παράμετροι ανακτώνται ανεξάρτητα η μία της άλλης όπως άλλωστε αναμενόταν.

Υπολογίζοντας τον μοναδιαίο πίνακα συνδιακύμανσης έχουμε ότι

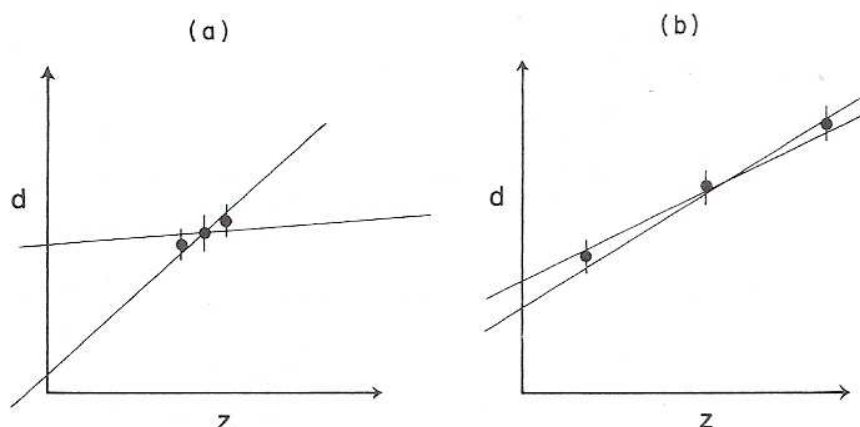
$$[\text{cov}_u \mathbf{m}] = [\mathbf{G}^T \mathbf{G}]^{-1} = \frac{1}{14} \begin{bmatrix} 21 & -7 \\ -7 & 3 \end{bmatrix} = \begin{bmatrix} 21/14 & -(7/14) \\ -(7/14) & 3/14 \end{bmatrix} \quad (4.21)$$

και

$$\text{size}([\text{cov}_u \mathbf{m}]) = \sum_{i=1}^M [\text{cov}_u \mathbf{m}]_{ii} = \frac{24}{14} < \frac{17}{6} \quad (4.22)$$

που σημαίνει ουσιαστικά μικρότερο εύρος διακύμανσης των προς εκτίμηση παραμέτρων σε περίπτωση που υπάρχει κάποιο λάθος στις μετρήσεις. Δηλαδή το λάθος στις μετρήσεις όταν οι μετρήσεις είναι σχετικά απομακρυσμένες μεταξύ τους, δίνει μικρότερη πιθανή διακύμανση παραμέτρων υπολογισμού της ευθείας.

Τα παραπάνω αποτυπώνονται σχηματικά στο Σχήμα 4.2 όπου φαίνονται οι ακραίες ευθείες που μπορεί να σχεδιαστούν με λύση ελαχίστων τετραγώνων, όταν οι μετρήσεις έχουν γίνει με ίδια διασπορά αλλά σε τετμημένες που στην μία περίπτωση είναι κοντά μεταξύ τους ενώ στην δεύτερη απέχουν περισσότερο. Η διασπορά των επαγόμενων παραμέτρων στην δεύτερη περίπτωση είναι αισθητά μικρότερη.



Σχήμα 4.2. Σχεδιασμός ευθείας με βάση τρεις μετρήσεις που έχουν γίνει σε διαφορετικά σημεία. Στην περίπτωση (α) οι μετρήσεις έγιναν σε κοντινά σημεία και η διασπορά των παραμέτρων που ορίζουν τις ευθείες είναι μεγάλη, ενώ στην περίπτωση (β) η αντίστοιχη διασπορά είναι μικρή.

5 ΜΕΘΟΔΟΙ ΜΕΓΙΣΤΗΣ ΠΙΘΑΝΟΦΑΝΕΙΑΣ

5.1 Η συνάρτηση μέγιστης πιθανοφάνειας

Έστω μία τυχαία μεταβλητή d η οποία αντιπροσωπεύει την μέτρηση κάποιας συγκεκριμένης ποσότητας με πραγματική αλλά άγνωστη τιμή θ σε ένα πείραμα που εκτελείται N φορές. Οι μετρήσεις που λαμβάνονται d_i αποτελούν τα στοιχεία ενός διανύσματος $\vec{d}$. Εάν οι μετρήσεις γίνονται σε περιβάλλον θορύβου ή με όργανα που ενδέχεται να εμφανίζουν λάθος μέτρηση, τότε οι τιμές d_i δεν θα είναι κατ' ανάγκη ίσες με την πραγματική τιμή της ποσότητας. Εάν η πιθανότητα η μέτρηση d_i να πάρει την τιμή θ είναι $P_d(d_i; \theta)$, η πιθανότητα όλες οι μετρήσεις να πάρουν την τιμή θ , με δεδομένο ότι η πιθανότητα κάθε μέτρηση να πάρει την εν λόγω τιμή είναι ανεξάρτητη οποιασδήποτε άλλης είναι :

$$P_d(d_1, d_2, \dots, d_N; \theta) = P_d(d_1; \theta) \cdots P_d(d_N; \theta) \quad (5.1)$$

Ορίζουμε ως *συνάρτηση πιθανοφάνειας (likelihood function)* $L(\theta; \vec{d})$ την ανωτέρω έκφραση της από κοινού πιθανότητας των μετρήσεων $\vec{d}$.

$$L(\theta; \vec{d}) = P_d(d_1, d_2, \dots, d_N; \theta) = P_d(d_1; \theta) \cdots P_d(d_N; \theta) \quad (5.2)$$

Η τιμή της θ που μεγιστοποιεί την συνάρτηση πιθανοφάνειας, ονομάζεται *εκτίμηση μέγιστης πιθανοφάνειας*; (*maximum likelihood estimate*) μια μπορεί να βρεθεί παραγωγίζοντας την L ως προς θ και ελέγχοντας ότι η δεύτερη παράγωγος είναι μικρότερη του 0 :

$$\frac{dL}{d\theta} = 0, \frac{d^2L}{d\theta^2} < 0 \quad (5.3)$$

Εάν η $\vec{d}$ ακολουθεί κανονική κατανομή με μέση τιμή m και διασπορά σ^2 , η έκφραση της συνάρτησης πιθανοφάνειας γίνεται :

$$L(m; \sigma^2 / \vec{d}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_1 - m)^2}{2\sigma^2}\right) \right) \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_2 - m)^2}{2\sigma^2}\right) \right) \cdots \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_N - m)^2}{2\sigma^2}\right) \right) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^N (d_i - m)^2\right] \quad (5.4)$$

Η εκτίμηση μέγιστης πιθανοφάνειας για την μέση τιμή m προκύπτει από την 5.3. Παράλληλα μπορούμε να ζητήσουμε και την εκτίμηση μέγιστης πιθανοφάνειας για την διασπορά, από τον ίδιο τύπο.

Η εκτίμηση μέγιστης πιθανοφάνειας εκφράζει την μέση τιμή της κατανομής πιθανότητας των μετρήσεων που δικαιολογούν με τον καλύτερο τρόπο τις μετρήσεις. Με άλλα λόγια, γνωρίζοντας την συνάρτηση πυκνότητας πιθανότητας των μετρήσεων ως προς τη μορφή της (εδώ έχουμε κανονική κατανομή), αναζητούμε τις παραμέτρους της κατανομής που να δίνει μετρήσεις όσο το δυνατόν κοντύτερα σε αυτή που αναμένεται να συγκεντρώνει τη μέγιστη πιθανότητα.

Πολλές φορές χρησιμοποιούμε τον λογάριθμο της συνάρτησης πιθανοφάνειας ως την προς μεγιστοποίηση συνάρτηση, καθώς ο λογάριθμος της L είναι συνάρτηση με ίδια χαρακτηριστικά μονοτονίας σε σχέση με την L . Επομένως μπορούμε να εκφράσουμε τις παραγώγους της 5.3 ως προς $\ln(L)$. Έτσι παίρνουμε :

$$L' = \ln(L) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^N (d_i - m)^2 \quad (5.5)$$

$$\frac{\partial L'}{\partial m} = \frac{1}{\sigma^2} \sum_{i=1}^N (d_i - m) = \frac{1}{\sigma^2} \left(\sum_{i=1}^N d_i - Nm \right) \quad (5.6)$$

$$\frac{\partial L'}{\partial \sigma^2} = -\frac{N}{\sigma^2} + \frac{1}{(\sigma^2)^2} \sum_{i=1}^N (d_i - m)^2 = \frac{N}{(\sigma^2)^2} \left(\sigma^2 - \frac{1}{N} \sum_{i=1}^N (d_i - m)^2 \right) \quad (5.7)$$

Θέτοντας τις παραγώγους από τις 5.6 και 5.7 ίσες με 0, παίρνουμε τις εκτιμήσεις μέγιστης πιθανοφάνειας για m και σ^2 που είναι αντίστοιχα :

$$m^{est} = \frac{1}{N} \sum_{i=1}^N d_i \quad (5.8)$$

$$\text{και} \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^N (d_i - m^{est})^2 \quad (5.9)$$

που αποτελούν τις γνωστές εκφράσεις για την μέση αριθμητική τιμή και την τυπική απόκλιση (standard deviation) της δειγματοληψίας. Επισημαίνεται ότι η εκτίμηση μέγιστης πιθανοφάνειας των μετρήσεων συμπίπτει με τον αριθμητικό μέσο, λόγω της υπόθεσης που κάναμε για κανονική κατανομή.

5.2 Εκτιμήσεις μέγιστης πιθανοφάνειας για το γραμμικό αντίστροφο πρόβλημα.

Θα δούμε τώρα πως υλοποιούνται οι ανωτέρω έννοιες στο γραμμικό αντίστροφο πρόβλημα της μορφής $\mathbf{Gm} = \mathbf{d}$. Υποθέτουμε ότι οι μετρήσεις του προβλήματος υπακούουν σε μία κανονική κατανομή της μορφής :

$$P(\mathbf{d}) \propto \exp\left[-\frac{1}{2}(\mathbf{d} - \mathbf{Gm})^T [\text{cov} \mathbf{d}]^{-1} (\mathbf{d} - \mathbf{Gm})\right] \quad (5.10)$$

Ο όρος $\mathbf{Gm}$ μπορεί να θεωρηθεί ότι αντιπροσωπεύει κάτι ανάλογο με τον μέσο των μετρήσεων ($\langle \mathbf{d} \rangle$) (δείτε εξίσωση 2.14). Ωστόσο εδώ έχει τη σημασία των εκτιμήσεων των μετρήσεων για δεδομένο $\mathbf{m}$. Επομένως μπορεί να δει κανείς την διαφορά $\mathbf{d} - \mathbf{Gm} = \mathbf{d}^{obs} - \mathbf{d}^{pre} = \mathbf{e}_i$.

Η μεγιστοποίηση της $P(\mathbf{d})$ αντιστοιχεί σε ελαχιστοποίηση της ποσότητας $(\mathbf{d} - \mathbf{Gm})^T [\text{cov} \mathbf{d}]^{-1} (\mathbf{d} - \mathbf{Gm})$. Η εκτίμηση μέγιστης πιθανοφάνειας λοιπόν στην περίπτωση μας θα δώσει ένα $\mathbf{m}^{est}$ το οποίο επιβεβαιώνει με τον καλύτερο τρόπο τις μετρήσεις. Με άλλα λόγια η βέλτιστη λύση για τις παραμέτρους $\mathbf{m}$, είναι η λύση ελαχίστων τετραγώνων για ένα καθαρά υπερ-ορισμένο πρόβλημα, στην οποία η έκφραση του «λάθους» των μετρήσεων έχει ζυγιστεί με τον αντίστροφο του πίνακα συνδιακύμανσης. Δηλαδή $\mathbf{W}_e = [\text{cov} \mathbf{d}]^{-1}$. Εάν όλες οι μετρήσεις είναι ασυσχέτιστες και έχουν την ίδια διασπορά, τότε $[\text{cov} \mathbf{d}] = \sigma_d^2 \mathbf{I}$ και η λύση μέγιστης πιθανοφάνειας είναι η απλή λύση ελαχίστων τετραγώνων. Εάν έχουμε μετρήσεις ασυσχέτιστες αλλά με διαφορετική διασπορά (σ_{di}^2) τότε το λάθος εκτίμησης είναι

$$E = \sum_{i=1}^N \sigma_{di}^{-2} e_i^2 \quad (5.11)$$

5.3 Εκ προοιμίου κατανομές

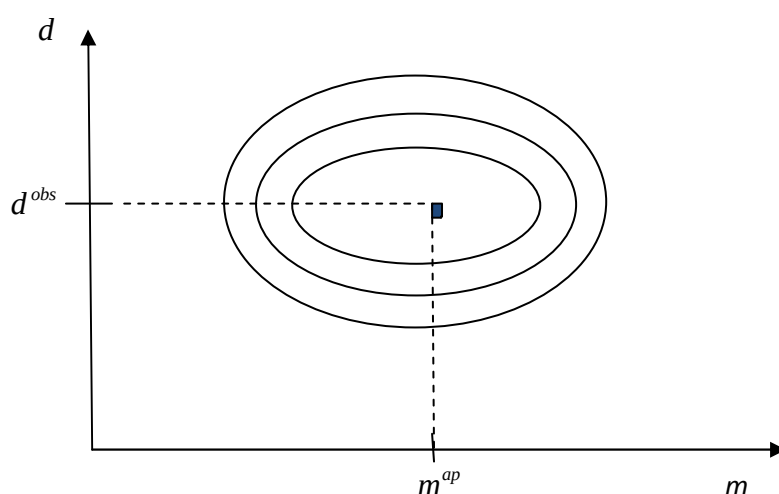
Εάν το πρόβλημα είναι υπο-ορισμένο, δεν υφίσταται η λύση ελαχίστων τετραγώνων. Όπως έχουμε δει, θα πρέπει να εισαχθούν αρχικές συνθήκες ως προς τις παραμέτρους, που εδώ μπορούν να πάρουν τη μορφή πιθανοθεωρητικών κατανομών. Στην περίπτωση αυτή θα πρέπει να αξιοποιήσουμε τη συνάρτηση κατανομής πιθανότητας $P_A(\mathbf{m})$ εφ' όσον είναι γνωστή, οπότε έχουμε πληροφορία για την μέση τιμή των παραμέτρων αλλά και για τη διασπορά τους. Συνδυάζοντας αυτή την πληροφορία με την αντίστοιχη για τις μετρήσεις, μπορούμε να βρούμε την εκτίμηση μέγιστης πιθανοφάνειας που βελτιστοποιεί εκτιμήσεις παραμέτρων και μετρήσεων.

Συνήθως οι κατανομές πιθανότητας παραμέτρων και μετρήσεων είναι ασυσχέτιστες, συνεπώς, η από κοινού συνάρτηση πιθανότητας εκφράζεται μέσω της σχέσης :

$$P_A(\mathbf{m}, \mathbf{d}) = P_A(\mathbf{m})P(\mathbf{d}) \quad (5.12)$$

Εάν εφαρμοστεί μέθοδος μέγιστης πιθανοφάνειας σε συνάρτηση που προκύπτει από την 5.12, σημειώνουμε ότι δεν έχει ληφθεί υπ' όψιν της το «μοντέλο» που χαρακτηρίζει το αντίστροφο πρόβλημα, δηλαδή η σχέση που συνδέει παραμέτρους και μετρήσεις, αλλά μόνο η πληροφορία για τις κατανομές πιθανοτήτων.

Στο σχήμα 5.1 παρουσιάζονται καμπύλες ίσης πιθανότητας για μία παράμετρο και μία μέτρηση σε ένα σχετικό πρόβλημα. Οι καμπύλες εκτείνονται γύρω από το σημείο (d^{obs}, m^{ap}) που δίνει τη μέγιστη πιθανότητα συνδυασμού μέτρησης και παραμέτρου.

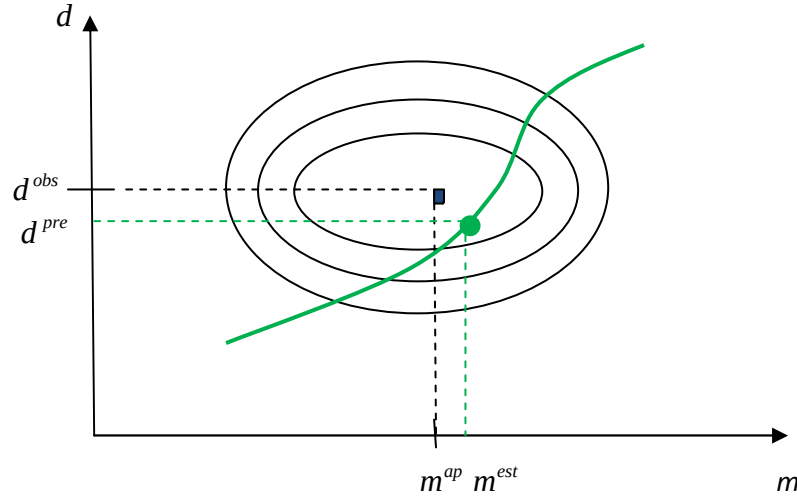


Σχήμα 5.1 Από κοινού συνάρτηση κατανομής πιθανότητας για μετρήσεις και παραμέτρους

5.4 Εκτιμήσεις μέγιστης πιθανοφάνειας για ακριβή θεωρία.

Εάν στην πληροφορία για κατανομές πιθανοτήτων μετρήσεων και παραμέτρων προστεθεί και η σχέση που συνδέει τις δύο ποσότητες (είτε γραμμική είτε μη γραμμική), έχουμε και μία επί πλέον πληροφορία που θα πρέπει να αξιοποιήσουμε. Εάν η θεωρία αυτή είναι ακριβής, γνωρίζουμε με σιγουριά την επιφάνεια (σε ένα πολυεπίπεδο χώρο) στην οποία πρέπει να αναζητηθεί η βέλτιστη λύση. Η θεωρία εκφράζεται στη γενική περίπτωση των διακριτών προβλημάτων από μία σχέση της μορφής : $\mathbf{g}(\mathbf{m}) = \mathbf{d}$ (που δεν είναι κατ' ανάγκην γραμμική). Σχηματικά, βλέπουμε τη διαφοροποίηση σε σχέση με την προηγούμενη περίπτωση στο σχήμα 5.2 όπου στις καμπύλες ίσης πιθανότητας παραμέτρων και δεδομένων όπως αυτές έχουν δοθεί «εκ προοιμίου», έχει προστεθεί και η επιφάνεια (γραμμή σε δύο διαστάσεις) που εκφράζει τη σχέση ανάμεσα σε παραμέτρους και μετρήσεις. Η βέλτιστη λύση πρέπει να

αναζητηθεί πάνω σε αυτή τη γραμμή και μπορεί να μας δώσει βέλτιστο σημείο (d^{pre}, m^{est}) διαφορετικό από το προηγούμενο.



Σχήμα 5.2 Από κοινού συνάρτηση κατανομής πιθανότητας για μετρήσεις και παραμέτρους με ακριβή θεωρία (μοντέλο).

5.5 Εκτιμήσεις μέγιστης πιθανοφάνειας για μη ακριβή θεωρία.

Εάν το μοντέλο που συνδέει παραμέτρους και μετρήσεις δεν είναι ακριβές, τότε μπορεί να εκφραστεί μέσω οικογένειας επιφανειών διαφορετικής πιθανότητας, εφ' όσον βέβαια γνωρίζουμε μία κατανομή πιθανότητας $P_g(\mathbf{m}|\mathbf{d})$ με κέντρο μία σχέση της μορφής $\mathbf{g}(\mathbf{m}) = \mathbf{d}$. Πρέπει να προσέξουμε στην έκφραση της $P_g(\mathbf{m}|\mathbf{d})$ ότι, έχουμε μία πιθανότητα υπό συνθήκη, δηλαδή έχουμε την πιθανότητα η θεωρία να προβλέπει τις μετρήσεις $\mathbf{d}$ όταν δίδονται οι παράμετροι $\mathbf{m}$. Στην περίπτωση αυτή ζητάμε να υπολογίσουμε παραμέτρους και μετρήσεις με την αρχή της μέγιστης πιθανοφάνειας, ορίζοντας μία νέα συνάρτηση πιθανοφάνειας :

$$P_T(\mathbf{m}, \mathbf{d}) = P_g(\mathbf{m}|\mathbf{d})P_A(\mathbf{m}, \mathbf{d}) \quad (5.13)$$

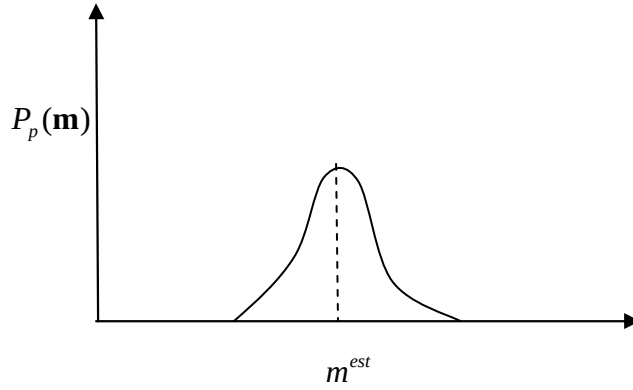
που δικαιολογείται από το γεγονός ότι η εκ προοιμίου κατανομή πιθανότητας παραμέτρων και μετρήσεων είναι ανεξάρτητη από την κατανομή πιθανότητας της θεωρίας.

Αξίζει να προσέξουμε ότι με βάση την αρχή της μέγιστης πιθανοφάνειας, υπολογίζουμε ταυτόχρονα βέλτιστες παραμέτρους και μετρήσεις, και όχι μόνο παρამέτρους όπως κάναμε π.χ. στην περίπτωση των ελαχίστων τετραγώνων. Έτσι οι εκτιμήσεις των παραμέτρων με τις δύο προσεγγίσεις μπορεί να είναι διαφορετικές.

Εάν ολοκληρώσουμε την έκφραση της συνάρτησης πιθανοφάνειας από την 5.13 ως προς τις μετρήσεις $\mathbf{d}$, προβάλλουμε ουσιαστικά την συνάρτηση στο επίπεδο $\mathbf{d}=0$:

$$P_p(\mathbf{m}) = \int P_T(\mathbf{m}, \mathbf{d}) \partial \mathbf{d} \quad (5.14)$$

Σε περίπτωση ακριβούς θεωρίας, η εκτίμηση μέγιστης πιθανοφάνειας που δίδει η 5.14 δεν είναι διαφορετική από αυτή που δίδεται εάν μεγιστοποιήσουμε την 5.12 στην επιφάνεια που ορίζει η ακριβής θεωρία.



Σχήμα 5.3 Η κατανομή $P_p(\bar{m})$. Το μέγιστο της κατανομής ορίζει την εκτίμηση μέγιστης πιθανοφάνειας m^{est} .

5.6 Η απλή περίπτωση με κανονική κατανομή και γραμμική θεωρία

Υποθέτουμε ότι δεν υπάρχει κάποια πληροφορία για την $P_A(\mathbf{m})$ αλλά ότι τα δεδομένα ακολουθούν κανονική κατανομή, που σημαίνει :

$$P_A(\mathbf{d}) \propto \exp[-\frac{1}{2}(\mathbf{d} - \mathbf{d}^{obs})^T [\text{cov } \mathbf{d}]^{-1} (\mathbf{d} - \mathbf{d}^{obs})]. \quad (5.15)$$

Εάν δεν υπάρχει λάθος στο μοντέλο που έχει επιλεγεί και για το μοντέλο ισχύσει η σχέση $\mathbf{Gm} = \mathbf{d}$, η στατιστική κατανομή εκφράζεται με συνάρτηση δέλτα της μορφής :

$$P_g(\mathbf{m}|\mathbf{d}) = \delta[\mathbf{Gm} - \mathbf{d}] \quad (5.16)$$

Η συνάρτηση πιθανοφάνειας θα είναι τότε :

$$P_T(\mathbf{m}, \mathbf{d}) \propto \exp[-\frac{1}{2}(\mathbf{d} - \mathbf{d}^{obs})^T [\text{cov } \mathbf{d}]^{-1} (\mathbf{d} - \mathbf{d}^{obs})] \delta[\mathbf{Gm} - \mathbf{d}] \quad (5.17)$$

Ολοκληρώνοντας ως προς $\mathbf{d}$ παίρνουμε από τις ιδιότητες της συνάρτησης δέλτα :

$$P_T(\mathbf{m}) \propto \exp[-\frac{1}{2}(\mathbf{Gm} - \mathbf{d}^{obs})^T [\text{cov } \mathbf{d}]^{-1} (\mathbf{Gm} - \mathbf{d}^{obs})] \quad (5.18)$$

Η μεγιστοποίηση της συνάρτησης πιθανοφάνειας συνεπώς στην περίπτωση αυτή, αντιστοιχεί σε λύση ελαχίστων τετραγώνων.

5.7 Η γενική γραμμική περίπτωση με κανονικές κατανομές.

Υποθέτουμε ότι για όλες τις κατανομές ισχύουν οι σχέσεις :

$$P_A(\mathbf{m}) \propto \exp[-\frac{1}{2}(\mathbf{m} - \langle \mathbf{m} \rangle)^T [\text{cov } \mathbf{m}]^{-1} (\mathbf{m} - \langle \mathbf{m} \rangle)] \quad (5.19)$$

$$P_A(\mathbf{d}) \propto \exp[-\frac{1}{2}(\mathbf{d} - \mathbf{d}^{obs})^T [\text{cov } \mathbf{d}]^{-1} (\mathbf{d} - \mathbf{d}^{obs})] \quad (5.20)$$

$$P_g(\mathbf{m}|\mathbf{d}) \propto \exp[-\frac{1}{2}(\mathbf{d} - \mathbf{G}\mathbf{m})^T [\text{cov } \mathbf{g}]^{-1} (\mathbf{d} - \mathbf{G}\mathbf{m})] \quad (5.21)$$

Μπορούμε να αποδείξουμε ότι η συνδυασμένη κατανομή που δίνει και τη συνάρτηση πιθανοφάνειας είναι επίσης κανονική :

$$P_T(\mathbf{m}) = P_g(\mathbf{m}|\mathbf{d})P_A(\mathbf{m})P_A(\mathbf{d}) \quad (5.22)$$

Με βάση την αρχή της μέγιστης πιθανοφάνειας, μπορούμε τώρα να δώσουμε λύση για τις εκτιμήσεις των παραμέτρων από τη σχέση :

$$\mathbf{m}^{est} = \langle \mathbf{m} \rangle + \mathbf{G}^{-g} [\mathbf{d}^{obs} - \mathbf{G} \langle \mathbf{m} \rangle] = \mathbf{G}^{-g} \mathbf{d} + [\mathbf{I} - \mathbf{R}] \langle \mathbf{m} \rangle \quad (5.23)$$

όπου

$$\begin{aligned} \mathbf{G}^{-g} &= [\text{cov } \mathbf{m}] \mathbf{G}^T \{ [\text{cov } \mathbf{d}] + [\text{cov } \mathbf{g}] + \mathbf{G} [\text{cov } \mathbf{m}] \mathbf{G}^T \}^{-1} \\ &= \{ \mathbf{G}^T [[\text{cov } \mathbf{d}] + [\text{cov } \mathbf{g}]]^{-1} \mathbf{G} + [\text{cov } \mathbf{m}]^{-1} \}^{-1} \mathbf{G}^T \{ [\text{cov } \mathbf{d}] + [\text{cov } \mathbf{g}] \}^{-1} \end{aligned} \quad (5.24)$$

Σημειώνουμε ότι οι δύο μορφές του γενικευμένου αντίστροφου πίνακα είναι ισοδύναμες.

Ο πίνακας συνδιακύμανσης των παραμέτρων υπολογίζεται από την παρατήρηση ότι οι παράμετροι προκύπτουν ως γραμμικός συνδυασμός παρατηρήσεων και εκ προοιμίου δεδομένων και προκύπτει ως :

$$[\text{cov } \mathbf{m}^{est}] = \mathbf{G}^{-g} [\text{cov } \mathbf{d}] \mathbf{G}^{-gT} + [\mathbf{I} - \mathbf{R}] [\text{cov } \mathbf{m}] [\mathbf{I} - \mathbf{R}]^T \quad (5.25)$$

που διαφέρει από προηγούμενες αντίστοιχες εκφράσεις (βλ. 3.52) ως προς την ύπαρξη της εκ προοιμίου πληροφορίας για τις προς ανάκτηση παραμέτρους.

5.8 Ειδικές περιπτώσεις

Θα εξετάσουμε τώρα τη μορφή που παίρνει η λύση μέγιστης πιθανοφάνειας σε ορισμένες απλές περιπτώσεις :

5.8.1 Ακριβή δεδομένα και ακριβές μοντέλο (θεωρία)

Εάν υποθέσουμε ότι ισχύει $\sigma_d^2 = \sigma_g^2 = 0$ στη γενική περίπτωση που έχουμε εκ προοιμίου πληροφορία για τις παραμέτρους, παίρνουμε λύση της μορφής :

$$\mathbf{m}^{est} = \mathbf{G}^{-g} \mathbf{d}^{obs} + [\mathbf{I} - \mathbf{R}] \langle \mathbf{m} \rangle \quad (5.26)$$

όπου η έκφραση του γενικευμένου αντιστρόφου εξαρτάται από το είδος του προβλήματος (υπερ-ορισμένο ή υπο-ορισμένο) κατά τα γνωστά. Με την ίδια λογική, ο δεύτερος όρος στην έκφραση του 5.26 είναι 0 όταν έχουμε ένα υπερ-ορισμένο πρόβλημα με ακριβείς μετρήσεις που σημαίνει ότι η λύση μέγιστης πιθανοφάνειας δεν λαμβάνει υπ' όψιν της την αρχική πληροφορία για τις παραμέτρους. Φυσικά στην περίπτωση του υπο-ορισμένου προβλήματος η πληροφορία αυτή συμμετέχει στη λύση.

5.8.2 Μη ακριβές μοντέλο και μη ακριβή δεδομένα.

Μια άλλη ακραία περίπτωση (μη ρεαλιστική ωστόσο σε πρακτικές εφαρμογές) είναι να έχουμε τελείως ανακριβές μοντέλο και τελείως ανακριβή δεδομένα. Τότε βέβαια θα ισχύει $\sigma_d^2 \rightarrow \infty, \sigma_g^2 \rightarrow \infty$ και η λύση μέγιστης πιθανοφάνειας προέρχεται αποκλειστικά από την αρχική πληροφορία :

$$\mathbf{m}^{est} = \langle \mathbf{m} \rangle \quad (5.27)$$

5.8.3 Μη ακριβής αρχική πληροφορία για τις παραμέτρους.

Μπορεί να έχουμε και μια άλλη ακραία περίπτωση σύμφωνα με την οποία δεν εμπιστευόμαστε την αρχική πληροφορία για τις παραμέτρους. Αυτό είναι ισοδύναμο με το να θεωρήσουμε ότι ισχύσει μια σχέση της μορφής $\sigma_m^2 \rightarrow \infty$.

Όπως μπορεί να δειχθεί στην περίπτωση αυτή παίρνουμε λύσεις μέγιστης πιθανοφάνειας της μορφής :

$$\mathbf{m}^{est} = \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1} \mathbf{d}^{obs} + \{\mathbf{I} - \mathbf{G}^T [\mathbf{G} \mathbf{G}^T]^{-1} \mathbf{G}\} \langle \mathbf{m} \rangle \quad (5.28)$$

που αντιστοιχεί σε λύση ελαχίστου μήκους και έχει τη μορφή της 5.26. Μπορεί να δείχτεί ότι για υπερορισμένο πρόβλημα με τις ίδιες υποθέσεις παίρνουμε πάλι λύση της μορφής 5.26 με αλλαγή της μορφής του γενικευμένου αντίστροφου πίνακα.

6 ΔΙΑΝΥΣΜΑΤΙΚΟΙ ΧΩΡΟΙ ΚΑΙ ΑΝΤΙΣΤΡΟΦΑ ΠΡΟΒΛΗΜΑΤΑ

6.1 Διανυσματικοί χώροι παραμέτρων και μετρήσεων.

Θα δανειστούμε για μία ακόμη φορά έννοιες της Γραμμικής Άλγεβρας προκειμένου να δούμε πως μπορούμε να χειριστούμε τη λύση γραμμικών αντιστρόφων προβλημάτων με μεγαλύτερη ευελιξία.

Θεωρούμε το γραμμικό αντίστροφο πρόβλημα που ορίζεται από τη σχέση $\mathbf{G}\mathbf{m} = \mathbf{d}$. Ο διανυσματικός χώρος των δεδομένων (μετρήσεων) συμβολίζεται με $S(\mathbf{d})$ και κάθε διάνυσμα μετρήσεων $\mathbf{d}$ περιλαμβάνεται σε αυτόν. Αντίστοιχα, ο διανυσματικός χώρος των παραμέτρων συμβολίζεται με $S(\mathbf{m})$ και κάθε διάνυσμα παραμέτρων $\mathbf{m}$ περιέχεται σε αυτόν.

Το μοντέλο που διέπει το γραμμικό αντίστροφο πρόβλημα και εκφράζεται μέσω της εξίσωσης $\mathbf{G}\mathbf{m} = \mathbf{d}$ μπορεί να θεωρηθεί ότι εκφράζει γραμμικό μετασχηματισμό από το χώρο $S(\mathbf{m})$ στον $S(\mathbf{d})$, ενώ η λύση $\mathbf{m}^{est} = \mathbf{G}^{-g}\mathbf{d}$ εκφράζει ένα μετασχηματισμό από το χώρο $S(\mathbf{d})$ στον $S(\mathbf{m})$. με τους πίνακες $\mathbf{G}$ και $\mathbf{G}^{-g}$ να αντιπροσωπεύουν τους πίνακες των αντίστοιχων γραμμικών μετασχηματισμών.

Γνωρίζουμε επίσης ότι κάθε διάνυσμα που περιέχεται σε ένα διανυσματικό χώρο εκφράζεται ως γραμμικός συνδυασμός των διανυσμάτων μιας βάσης του χώρου. Έτσι για παράδειγμα ο διανυσματικός χώρος των παραμέτρων που έχει διάσταση M παράγεται από ισάριθμα διανύσματα βάσης $\mathbf{m}_j$ και κάθε διάνυσμα $\mathbf{m}$ εκφράζεται μέσω της βάσης από μία σχέση της μορφής :

$$\mathbf{m} = \sum_{j=1}^M a_j \mathbf{m}_j \quad (6.1)$$

Ας θεωρήσουμε τώρα δύο διαφορετικά διανύσματα $\mathbf{m}_1$ και $\mathbf{m}_2$ που αποτελούν λύση του γραμμικού αντιστρόφου προβλήματος $\mathbf{G}\mathbf{m} = \mathbf{d}$.

Τότε θα ισχύει $\mathbf{G}\mathbf{m}_1 = \mathbf{d}$ και $\mathbf{G}\mathbf{m}_2 = \mathbf{d}$, οπότε

$$\mathbf{G}(\mathbf{m}_1 - \mathbf{m}_2) = 0 \quad (6.2)$$

Η διαφορά $\mathbf{m}_1 - \mathbf{m}_2$ ανήκει συνεπώς στο μηδενόχωρο του πίνακα $\mathbf{G}$ και η υπόθεση για μηδενόχωρο διάφορο του κενού, μας παραπέμπει αμέσως στην ανυπαρξία μοναδικής λύσης στο διακριτό αντίστροφο πρόβλημα.

Εάν q είναι η τάξη του μηδενόχωρου, $\mathbf{m}^{par}$ είναι μία μη μηδενική λύση του αντιστρόφου προβλήματος $\mathbf{G}\mathbf{m} = \mathbf{d}$ και $\mathbf{m}_i^{null}$ είναι ένα διάνυσμα του μηδενόχωρου του $\mathbf{G}$, μπορούμε να γράψουμε τη γενική λύση του αντίστροφου προβλήματος μέσω της σχέσης

$$\mathbf{m}^{gen} = \mathbf{m}^{par} + \sum_{i=1}^q a_i \mathbf{m}_i^{null} \quad (6.3)$$

Προφανώς για ένα πίνακα $\mathbf{G}$, διαστάσεων $N \times M$ θα έχουμε $0 \leq q \leq M$.

Παράδειγμα

Ας θεωρήσουμε την απλή εξίσωση

$$\mathbf{G}\mathbf{m} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} = [d] \quad (6.4)$$

που σημαίνει ότι μετρήσαμε τη μέση τιμή της παραμέτρου $\mathbf{m}$. Ο πίνακας $\mathbf{G}$ έχει διαστάσεις 1×4 και η τάξη του είναι 1.

Από την άλλη μεριά το ομογενές πρόβλημα $\mathbf{G}\mathbf{m} = 0$, θα πρέπει να έχει 3 ($=N-M$) γραμμικά ανεξάρτητες λύσεις που παράγουν το μηδενόχωρο. Κατά τα γνωστά, μπορούμε να βρούμε τα διανύσματα που αποτελούν βάση του μηδενόχωρου, βάζοντας τη τιμή 1 κατά σειρά στις εξαρτημένες μεταβλητές και μηδενίζοντας τις υπόλοιπες, ώστε να παραχθούν τα διανύσματα

$$\mathbf{m}_1^{null} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{m}_2^{null} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{m}_3^{null} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (6.5)$$

Η γενική λύση του γραμμικού αντίστροφου προβλήματος θα είναι τότε

$$\mathbf{m}^{gen} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} + \sum_{i=1}^3 a_i \mathbf{m}_i^{null} \quad (6.6)$$

Ο υπολογισμός κάποιας συγκεκριμένης λύσης στο αντίστροφο πρόβλημα, περνά από την απόδοση τιμών στις παραμέτρους a_i . Επιλέγοντας την ελαχιστοποίηση της

νόρμας $\|\mathbf{m}\|_2$ παίρνουμε λύση ελαχίστου μήκους και εύκολα διαπιστώνουμε ότι αυτή δίνει $a_i = 0, \quad i = 1, 2, 3$

Συνεπώς η λύση ελαχίστου μήκους δεν περιλαμβάνει διανύσματα του μηδενόχωρου του πίνακα $\mathbf{G}$.

6.2 Γραμμικοί μετασχηματισμοί

Συνεχίζοντας την αναφορά μας στους διανυσματικούς χώρους, θα δούμε ορισμένα θέματα που αφορούν γραμμικούς μετασχηματισμούς που επιβάλλονται στα συστατικά ενός αντιστρόφου προβλήματος προκειμένου να βελτιώσουν τη διαδικασία παραγωγής της λύσης του.

Θεωρούμε τον πίνακα $\mathbf{T}$ ενός γραμμικού μετασχηματισμού, που μετασχηματίζει τα διανύσματα των παραμέτρων ενός προβλήματος αλλάζοντας ουσιαστικά τους άξονες αναφοράς (βάσεις του αντίστοιχου διανυσματικού χώρου).

$$\mathbf{m}' = \mathbf{T}\mathbf{m} \text{ και } \mathbf{m} = \mathbf{T}^{-1}\mathbf{m}' \quad (6.7)$$

Ένα γραμμικό αντίστροφο πρόβλημα μπορεί να αλλάξει λοιπόν βάση αναφοράς μέσω της διαδικασίας :

$$\mathbf{d} = \mathbf{G}\mathbf{m} = \mathbf{G}\mathbf{I}\mathbf{m} = \mathbf{G}\mathbf{T}^{-1}\mathbf{T}\mathbf{m} = \{\mathbf{G}\mathbf{T}^{-1}\}\{\mathbf{T}\mathbf{m}\} = \mathbf{G}'\mathbf{m}' \quad (6.8)$$

όπου $\mathbf{G}'$ είναι ο πίνακας του γραμμικού αντίστροφου προβλήματος που ορίζεται για μία καινούργια βάση του $\mathbf{m}$.

Εάν υποθέσουμε ότι το πρόβλημα είναι υπο-ορισμένο και θέλουμε να βρούμε λύση ελαχίστου μήκους, θα αναζητήσουμε την ελαχιστοποίηση του μήκους $L = \mathbf{m}^T \mathbf{m}$

Το μήκος L εκφράζεται στο καινούργιο σύστημα ως :

$$L = \mathbf{m}^T \mathbf{m} = [\mathbf{T}^{-1}\mathbf{m}']^T [\mathbf{T}^{-1}\mathbf{m}'] = \mathbf{m}'^T [\mathbf{T}^{-1T} \mathbf{T}^{-1}] \mathbf{m}' \quad (6.9)$$

Εάν μπορέσουμε να βρούμε πίνακα $\mathbf{T}$ ώστε να ισχύει $[\mathbf{T}^{-1T} \mathbf{T}^{-1}] = \mathbf{I}$, η λύση ελαχίστου μήκους στα δύο συστήματα θα έχει την ίδια μορφή και συνεπώς η ελαχιστοποίηση του $\mathbf{m}^T \mathbf{m}$ είναι ισοδύναμη με την ελαχιστοποίηση του $\mathbf{m}'^T \mathbf{m}'$. Οι μετασχηματισμοί αυτού του είδους που δεν αλλάζουν το μήκος των διανυσμάτων των παραμέτρων ονομάζονται *μοναδιαίοι (unitary)* και είναι ιδιαίτερα χρήσιμοι σε επιλύσεις αντιστρόφων προβλημάτων.

Ας δούμε δύο παραδείγματα :

6.2.1 Υπο-ορισμένο πρόβλημα

Υποθέτουμε ότι έχουμε ένα υπο-ορισμένο πρόβλημα της μορφής $\mathbf{G}\mathbf{m} = \mathbf{d}$ με τον πίνακα $\mathbf{G}$ διαστάσεων $N \times M$ και $N < M$. Εάν με το μετασχηματισμό οδηγηθούμε σε πίνακα $\mathbf{G}'$ τέτοιο ώστε να μπορούμε να γράψουμε αναλυτικά το μετασχηματισμένο πρόβλημα στη μορφή :

$$\begin{bmatrix} G'_{11} & 0 & 0 & 0 & \dots & 0 & \dots & \dots & 0 \\ G'_{21} & G'_{22} & 0 & 0 & \dots & 0 & \dots & \dots & 0 \\ G'_{31} & G'_{32} & G'_{33} & 0 & \dots & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ G'_{N1} & G'_{N2} & G'_{N3} & G'_{N4} & \dots & G'_{NN} & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} m'_1 \\ m'_2 \\ m'_3 \\ \vdots \\ \vdots \\ \vdots \\ m'_M \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ \vdots \\ \vdots \\ d_N \end{bmatrix}, \quad (6.10)$$

παρατηρούμε κατ' αρχήν ότι μπορούμε να λύσουμε την εξίσωση ως προς τις πρώτες N τιμές του m , $m_1, m_2, \dots, m_N$ μέσω των άμεσων σχέσεων :

$$m'_1{}^{est} = d_1 / G'_{11}$$

$$m'_2{}^{est} = [d_2 - G'_{21}m'_1{}^{est}] / G'_{22}$$

$$m'_3{}^{est} = [d_3 - G'_{31}m'_1{}^{est} - G'_{32}m'_2{}^{est}] / G'_{33}$$

Η διαδικασία χαρακτηρίζεται ως *ευθεία λύση* (forward solving).

Εάν επιθυμούμε να παράγουμε λύση ελαχίστου μήκους, μπορούμε να θέσουμε τις υπόλοιπες τιμές του m ίσες με 0 ($m'_i{}^{est} = 0, i = N+1, M$), αφού από τη μορφή του πίνακα $\mathbf{G}'$ είναι φανερό ότι δεν επηρεάζουν τις τιμές του $\mathbf{d}$.

Η λύση μας στο αρχικό σύστημα συντεταγμένων θα προκύψει από τη σχέση $\mathbf{m}^{est} = \mathbf{T}^{-1}\mathbf{m}'^{est}$, που αποτελεί λύση ελαχίστου μήκους.

Ένας μετασχηματισμός αυτού του είδους που «τριγωνοποιεί» τον πίνακα $\mathbf{G}$ χαρακτηρίζεται μετασχηματισμός *Householder*¹. Δεν θα αναφερθούμε εδώ σε τρόπο υπολογισμού του.

¹ Δεν γνωρίζω κατάλληλο ελληνικό όρο για τον εν λόγω μετασχηματισμό.

6.2.2. Υπερ-ορισμένο πρόβλημα

Ας θεωρήσουμε τώρα της μορφής $\mathbf{Gm} = \mathbf{d}$ με τον πίνακα $\mathbf{G}$ διαστάσεων $N \times M$ και $M < N$. Ζητάμε λύση που να ελαχιστοποιεί το «λάθος» $E = \mathbf{e}^T \mathbf{e}$. Εάν βρούμε ένα μετασχηματισμό τέτοιο ώστε επιδρώντας στο λάθος πρόβλεψης $\mathbf{e}$, ($\mathbf{e}' = \mathbf{T}\mathbf{e}$) να οδηγεί σε ελαχιστοποίηση του $E' = \mathbf{e}'^T \mathbf{e}'$ όταν ελαχιστοποιείται και το E , και να μετασχηματίζει τον πίνακα $\mathbf{G}$ στον πίνακα $\mathbf{G}'$ που να έχει άνω τριγωνική μορφή, θα έχουμε

$$\mathbf{e}' = \mathbf{T}\mathbf{e} = \mathbf{T}(\mathbf{d} - \mathbf{Gm}) = \mathbf{Td} - \mathbf{TGm} = \mathbf{d}' - \mathbf{G}'\mathbf{m} \quad (6.11)$$

που αναλυτικά γράφεται ως

$$\begin{bmatrix} e'_1 \\ e'_2 \\ e'_3 \\ \vdots \\ e'_M \\ \vdots \\ e'_N \end{bmatrix} = - \begin{bmatrix} G'_{11} & G'_{12} & G'_{13} & G'_{14} & \cdot & G'_{1M} \\ 0 & G'_{22} & G'_{23} & G'_{24} & \cdot & G'_{2M} \\ 0 & 0 & G'_{33} & G'_{34} & \cdot & G'_{3M} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & G'_{MM} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} m'_1 \\ m'_2 \\ m'_3 \\ \vdots \\ m'_M \end{bmatrix} + \begin{bmatrix} d'_1 \\ d'_2 \\ d'_3 \\ \vdots \\ d'_M \\ \vdots \\ d'_N \end{bmatrix} \quad (6.12)$$

Στην περίπτωση αυτή βλέπουμε ότι οποιαδήποτε επιλογή των m_i^{est} δεν επηρεάζει τις τελευταίες $N-M$ εξισώσεις. Επομένως μπορούμε να ζητήσουμε να μηδενίζονται οι πρώτες M τιμές του $\mathbf{e}'$, οπότε ικανοποιούνται ακριβώς οι πρώτες M εξισώσεις του συστήματος που παίρνουν τη μορφή $\mathbf{d}' = \mathbf{G}'\mathbf{m}'$. Επί πλέον λόγω της μορφής του πίνακα $\mathbf{G}'$ μπορούμε να λύσουμε το σύστημα με αντίστροφη διαδικασία. Το συνολικό λάθος λοιπόν είναι :

$$E' = \sum_{i=M+1}^N e_i'^2 = \sum_{i=M+1}^N d_i'^2. \quad (6.13)$$

Με το μετασχηματισμό, διαχωρίσαμε το αντίστροφο πρόβλημα σε ένα μέρος το οποίο αφορά μετρήσεις που ικανοποιούνται επακριβώς και σε ένα άλλο μέρος που αφορά μετρήσεις που δεν μπορούν να ικανοποιηθούν. Η λύση επιλέγεται έτσι ώστε να ελαχιστοποιείται το λάθος πρόβλεψης και συνεπώς είναι λύση ελαχίστων τετραγώνων.

6.2.3 Το μεικτά ορισμένο πρόβλημα.

Σε ένα μεικτά ορισμένο πρόβλημα της μορφής $\mathbf{Gm}=\mathbf{d}$ κάποιοι γραμμικοί συνδυασμοί παραμέτρων είναι υπερ-ορισμένοι και κάποιοι υπο-ορισμένοι. Εάν ένα πρόβλημα είναι σε κάποιο βαθμό υπο-ορισμένο, η εξίσωση $\mathbf{Gm}=\mathbf{d}$ περιέχει πληροφορία για μερικές μόνο από τις παραμέτρους. Οι παράμετροι αυτές θεωρούμε ότι περιλαμβάνονται σε ένα υπόχωρο $S_p(\mathbf{m})$ του χώρου των παραμέτρων. Για το συμπλήρωμα του υπόχωρου αυτού ως προς το χώρο των παραμέτρων, ο πίνακας $\mathbf{G}$ δεν παρέχει καμία πληροφορία. Ο υπόχωρος αυτός χαρακτηρίζεται ως $S_0(\mathbf{m})$.

Αντίστοιχα σε ένα υπερ-ορισμένο πρόβλημα το γινόμενο $\mathbf{Gm}$ μπορεί να μην μπορεί να παράγει το χώρο $S(\mathbf{d})$ ανεξάρτητα από την επιλογή των παραμέτρων. Το πολύ να μπορεί να ικανοποιηθεί ένα μέρος των μετρήσεων που χαρακτηρίζεται ως $S_p(\mathbf{d})$ σε αντίθεση με τον υπόχωρο $S_0(\mathbf{d})$ που περιλαμβάνει τις μετρήσεις που δεν μπορούν να ικανοποιηθούν.

Θεωρώντας τώρα παραμέτρους και μετρήσεις που ανήκουν στους αντίστοιχους υπόχωρους με τους ανωτέρω συμβολισμούς, μπορούμε να γράψουμε στη γενική περίπτωση :

$$\mathbf{G}[\mathbf{m}_p + \mathbf{m}_0] = [\mathbf{d}_p + \mathbf{d}_0] \quad (6.14)$$

Το μήκος της λύσης είναι

$$L = \mathbf{m}^T \mathbf{m} = [\mathbf{m}_p + \mathbf{m}_0]^T [\mathbf{m}_p + \mathbf{m}_0] = \mathbf{m}_p^T \mathbf{m}_p + \mathbf{m}_0^T \mathbf{m}_0 \quad (6.15)$$

Το λάθος πρόβλεψης είναι

$$E = [\mathbf{d}_p + \mathbf{d}_0 - \mathbf{Gm}_p]^T [\mathbf{d}_p + \mathbf{d}_0 - \mathbf{Gm}_p] = [\mathbf{d}_p - \mathbf{Gm}_p]^T [\mathbf{d}_p - \mathbf{Gm}_p] + \mathbf{d}_0^T \mathbf{d}_0 \quad (6.16)$$

Όπου στους υπολογισμούς έχει ληφθεί υπ' όψιν ότι γινόμενα διανυσμάτων που ανήκουν σε διαφορετικούς υπόχωρους είναι μηδενικά.

Σε ένα μεικτά ορισμένο πρόβλημα, εκ προοιμίου πληροφορία εισάγεται για να ορίσει παραμέτρους που περιέχονται στο χώρο $S_0(\mathbf{m})$ και το λάθος πρόβλεψης περιορίζεται στο χώρο $S_0(\mathbf{d})$ ικανοποιώντας τη σχέση $\mathbf{e}_p = [\mathbf{d}_p - \mathbf{Gm}_p] = 0$ επακριβώς.

Όταν το πρόβλημα είναι καθαρά υπο-ορισμένο, επιλέγοντας $m_0^{est} = 0$ (που την ονομάζουμε φυσική λύση – natural solution) οδηγούμαστε σε λύση ελαχίστου μήκους ενώ σε ένα καθαρά υπερ-ορισμένο πρόβλημα οδηγούμαστε σε λύση ελαχίστων τετραγώνων.

6.3 Ανάλυση Ιδιαζουσών Τιμών (Singular Value Decomposition)

Στο κεφάλαιο αυτό θα αναφερθούμε σε μία διαδικασία που βοηθά στον καθορισμό του μηδενόχωρου και συνεπώς της λύσης ενός γραμμικού αντίστροφου προβλήματος όπως αυτή εκφράζεται από τη σχέση 6.3, μέσω κατάλληλης παραγοντοποίησης του πίνακα $\mathbf{G}$.

6.3.1 Η ανάλυση

Θεωρούμε τον τετραγωνικό συμμετρικό πίνακα $\mathbf{S}$ διαστάσεων $(N+M) \times (N+M)$ που προκύπτει από τον πίνακα $\mathbf{G}$ του γραμμικού αντίστροφου προβλήματος ως εξής :

$$\mathbf{S} = \begin{bmatrix} \mathbf{0} & \mathbf{G} \\ \mathbf{G}^T & \mathbf{0} \end{bmatrix} \quad (6.17)$$

Από τη Γραμμική Άλγεβρα γνωρίζουμε ότι ο πίνακας αυτός έχει $N+M$ πραγματικές ιδιοτιμές λ_i και μία πλήρη ομάδα ιδιοδιανυσμάτων $\mathbf{w}_i$ που επιλύουν το πρόβλημα

$$\mathbf{S}\mathbf{w}_i = \lambda_i \mathbf{w}_i. \quad (6.18)$$

Διαχωρίζοντας το διάνυσμα $\mathbf{w}_i$ σε ένα μέρος διάστασης N και ένα άλλο διάστασης M , γράφουμε την 6.18 αναλυτικά ως

$$\mathbf{S}\mathbf{w} = \begin{bmatrix} \mathbf{0} & \mathbf{G} \\ \mathbf{G}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} = \lambda_i \begin{bmatrix} \mathbf{u}_i \\ \mathbf{v}_i \end{bmatrix} \quad (6.19)$$

Από την παραπάνω σχέση διαπιστώνουμε ότι ισχύουν ταυτόχρονα :

$$\mathbf{G}\mathbf{v}_i = \lambda_i \mathbf{u}_i \quad \text{και} \quad \mathbf{G}^T \mathbf{u}_i = \lambda_i \mathbf{v}_i. \quad (6.20)$$

Εάν υποθέσουμε ότι έχουμε μία θετική ιδιοτιμή λ_i τότε και $-\lambda_i$ είναι επίσης ιδιοτιμή με αντίστοιχα ιδιοδιανύσματα $[\mathbf{u}_i, \mathbf{v}_i]^T$ και $[-\mathbf{u}_i, \mathbf{v}_i]^T$. Εάν έχουμε p θετικές ιδιοτιμές, θα πρέπει να έχουμε συνολικά $N+M-2p$ μηδενικές ιδιοτιμές. Φυσικά θα πρέπει να ισχύει ότι $p \leq \min(N, M)$ αφού δεν μπορούμε να έχουμε περισσότερες ιδιοτιμές σε σχέση με τη διάσταση του πίνακα.

Από τις 6.20 προκύπτουν εύκολα οι σχέσεις

$$\mathbf{G}^T \mathbf{G} \mathbf{v}_i = \lambda_i^2 \mathbf{v}_i \quad \text{και} \quad \mathbf{G} \mathbf{G}^T \mathbf{u}_i = \lambda_i^2 \mathbf{u}_i \quad (6.21)$$

Εξετάζοντας τους πίνακες $\mathbf{G}^T\mathbf{G}$ και $\mathbf{G}\mathbf{G}^T$ βλέπουμε ότι είναι συμμετρικοί και διαθέτουν αντίστοιχα M και N ιδιοτιμές (με αντίστοιχα ιδιοδιανύσματα $\mathbf{v}_i$ και $\mathbf{u}_i$ που ορίζουν τα σύνολα $\mathbf{V}$ και $\mathbf{U}$ που έχουν τη μορφή πινάκων. Ο πίνακας $\mathbf{U}$ είναι :

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N] \quad (6.22)$$

και ο πίνακας $\mathbf{V}$ είναι

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M] \quad (6.23)$$

Σε κάθε σύνολο τα ιδιοδιανύσματα είναι ορθογώνια και συνεπώς παράγουν τους χώρους $S(\mathbf{m})$ και $S(\mathbf{d})$ αντίστοιχα. Επίσης μπορούν να κανονικοποιηθούν ώστε να ισχύει :

$$\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}_N \text{ και } \mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_M \quad (6.24)$$

Με τους πίνακες $\mathbf{I}_N$ και $\mathbf{I}_M$ να είναι οι μοναδιαίοι πίνακες διαστάσεων N και M αντίστοιχα.

Ορίζοντας τώρα ως $\mathbf{\Lambda}$, τον διαγώνιο πίνακα διαστάσεων $N \times M$ με διαγώνια στοιχεία τις μη αρνητικές ιδιοτιμές (τις χαρακτηρίζουμε *ιδιάζουσες τιμές*) και παρατηρώντας ότι $\mathbf{V}$ και $\mathbf{U}$ είναι ορθογώνιοι, παίρνομε από την πρώτη από τις σχέσεις 6.20 :

$$\mathbf{G}\mathbf{V} = \mathbf{U}\mathbf{\Lambda} \quad (6.25)$$

και πολλαπλασιάζοντας από δεξιά με $\mathbf{V}^T$ παίρνομε :

$$\mathbf{G} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \quad (6.26)$$

που χαρακτηρίζεται ως παραγοντοποίηση ή *ανάλυση ιδιαζουσών τιμών* (Singular Value Decomposition- SVD).

6.3.2 Ανάλυση ιδιαζουσών τιμών και γενικευμένος αντίστροφος.

Όπως είπαμε πιο πάνω, οι ιδιάζουσες τιμές του πίνακα $\mathbf{G}$ δεν μπορεί να είναι περισσότερες από το $\min(N, M)$ συνεπώς στον πίνακα $\mathbf{\Lambda}$ μπορεί να έχουμε γραμμές ή στήλες μηδενικές. Εάν $N > M$ θα έχουμε μία γραμμή τουλάχιστον με μηδενικά στοιχεία και εάν $N < M$ θα έχουμε τουλάχιστον μία στήλη με μηδενικά στοιχεία. Συνεπώς στη γενική περίπτωση θα μπορούσαμε να γράψομε τον πίνακα $\mathbf{\Lambda}$ στη μορφή

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (6.27)$$

Με τον πίνακα $\mathbf{\Lambda}_p$ να είναι ένας $p \times p$ διαγώνιος πίνακας με μη μηδενικά στοιχεία στη διαγώνιο.

Θα πρέπει να σημειώσουμε ότι η διάταξη των ιδιοτιμών στον πίνακα Λ_p γίνεται κατά φθίνουσα σειρά. Λαμβάνοντας υπόψη τη μορφή του Λ , η ανάλυση ιδιαζουσών τιμών (6.26) παίρνει τη μορφή :

$$\mathbf{G} = \mathbf{U}_p \Lambda_p \mathbf{V}_p^T \quad (6.28)$$

με τους πίνακες $\mathbf{U}_p$ και $\mathbf{V}_p$ Να αποτελούνται από τις p πρώτες στήλες των $\mathbf{U}$ και $\mathbf{V}$ αντίστοιχα². Ο πίνακας $\mathbf{G}$ δεν περιλαμβάνει συνεπώς πληροφορία για τους υπόχωρους που παράγονται από τις υπόλοιπες στήλες των $\mathbf{U}$ και $\mathbf{V}$ οι οποίες παράγουν τους πίνακες $\mathbf{U}_0$ και $\mathbf{V}_0$ αντίστοιχα.

Με άλλα λόγια καταφέραμε να παράγουμε το μηδενόχωρο $S_0(\mathbf{m})$ τα διανύσματα του οποίου είδαμε να συμμετέχουν στη γενική λύση του αντιστρόφου προβλήματος με την έννοια της (6.3) αλλά παράλληλα και τους υπόχωρους $S_p(\mathbf{m})$ και $S_p(\mathbf{d})$ που παράγονται από τα διανύσματα $\mathbf{V}_p$ και $\mathbf{U}_p$ αντίστοιχα.

Η φυσική λύση στο αντίστροφο πρόβλημα μπορεί τώρα να κατασκευαστεί μέσω της ανάλυσης ιδιαζουσών τιμών. Η λύση θα πρέπει να αποτελείται από ένα διάνυσμα $\mathbf{m}^{est}$ που δεν έχει στοιχεία στον $S_0(\mathbf{m})$ και ένα λάθος πρόβλεψης που δεν έχει στοιχεία στο $S_p(\mathbf{d})$.

Θεωρώντας τη λύση :

$$\mathbf{m}^{est} = \mathbf{V}_p \Lambda_p^{-1} \mathbf{U}_p^T \mathbf{d} \quad (6.29)$$

μπορούμε να αποδείξουμε ότι ικανοποιεί τις παραπάνω απαιτήσεις και συνεπώς υιοθετείται ως η «φυσική λύση» του γραμμικού αντιστρόφου προβλήματος. Ο πίνακας $\mathbf{G}^{-g} = \mathbf{V}_p \Lambda_p^{-1} \mathbf{U}_p^T$ που προκύπτει από τη σχέση $\mathbf{m}^{est} = \mathbf{G}^{-g} \mathbf{d}$ χαρακτηρίζεται ως «ο φυσικός γενικευμένος αντίστροφος». Ο πίνακας ανάλυσης παραμέτρων για τον εν λόγω γενικευμένο αντίστροφο υπολογίζεται από τη σχέση της μορφής

$$\mathbf{R} = \mathbf{G}^{-g} \mathbf{G} = \{\mathbf{V}_p \Lambda_p^{-1} \mathbf{U}_p^T\} \{\mathbf{U}_p \Lambda_p \mathbf{V}_p^T\} = \mathbf{V}_p \mathbf{V}_p^T \quad (6.30)$$

και ο πίνακας ανάλυσης δεδομένων από τη σχέση :

$$\mathbf{N} = \mathbf{G} \mathbf{G}^{-g} = \{\mathbf{U}_p \Lambda_p \mathbf{V}_p^T\} \{\mathbf{V}_p \Lambda_p^{-1} \mathbf{U}_p^T\} = \mathbf{U}_p \mathbf{U}_p^T \quad (6.31)$$

Ο πίνακας συνδιακύμανσης για ασυσχέτιστα δεδομένα με ίδια διασπορά σ_d^2 είναι :

² Θυμηθείτε ότι υπάρχει διάταξη στις ιδιοτιμές συνεπώς πρέπει να υπάρξει και η αντιστοίχιση των ιδιοδιανυσμάτων.

$$[\text{cov} \mathbf{m}^{est}] = \mathbf{G}^{-g} [\text{cov} \mathbf{d}] \mathbf{G}^{-gT} = \sigma_d^2 \{\mathbf{V}_p \mathbf{\Lambda}_p^{-1} \mathbf{U}_p^T\} \{\mathbf{U}_p \mathbf{\Lambda}_p^{-1} \mathbf{U}_p^T\}^T = \sigma_d^2 \mathbf{V}_p \{\mathbf{\Lambda}_p^{-1}\}^2 \mathbf{V}_p^T \quad (6.32)$$

Είναι προφανές ότι τα δεδομένα (μετρήσεις) παράγονται πλήρως εάν ο υπόχωρος $\mathbf{U}_p$ παράγει το χώρο των μετρήσεων δηλαδή όταν $p=N$.

Εάν υπάρχει εκ προοιμίου πληροφορία για τις παραμέτρους με μέση τιμή $\langle \mathbf{m} \rangle$ και πίνακα συνδιακύμανσης $[\text{cov} \mathbf{m}]$ μπορεί να χρησιμοποιηθεί η σχέση

$$\mathbf{m}^{est} = \mathbf{G}^{-g} \mathbf{d} + [\mathbf{I} - \mathbf{R}] \langle \mathbf{m} \rangle \quad (6.33)$$

Ο πίνακας συνδιακύμανσης παίρνει τη μορφή

$$[\text{cov} \mathbf{m}^{est}] = \mathbf{G}^{-g} [\text{cov} \mathbf{d}] \mathbf{G}^{-gT} + [\mathbf{I} - \mathbf{R}] [\text{cov} \mathbf{m}] [\mathbf{I} - \mathbf{R}]^T \quad (6.34)$$

Επισημαίνεται ότι η εφαρμογή της ανάλυσης ιδιζουσών τιμών βασίζεται στη δυνατότητα ακριβούς υπολογισμού των τιμών αυτών. Σε πολλές περιπτώσεις όταν υπάρχουν μεγάλες διαφορές μεγέθους, που σημαίνει πρακτικά πίνακας $\mathbf{G}$ κακής κατάστασης, η ανάλυση δεν είναι ακριβής με τους υπάρχοντες αλγορίθμους. Σε άλλες περιπτώσεις ιδιάζουσες τιμές κοντά στο μηδέν θεωρούνται μηδενικές, οπότε στην πραγματικότητα επιλύεται ένα διαφορετικό πρόβλημα που προσεγγίζει το κανονικό. Άλλες λύσεις όπως για παράδειγμα η απόσβεση των μικρών ιδιοτιμών είναι επίσης εφαρμόσιμες.

Η γενική λύση που είδαμε με τη σχέση 6.3 μπορεί να θεωρηθεί ότι αποτελείται από το άθροισμα της φυσικής λύσης (μέσω της SVD) και της ειδικής που παράγεται από τα διανύσματα του μηδενόχωρου $S_0(\mathbf{m})$.

7 ΜΗ ΓΡΑΜΜΙΚΑ ΑΝΤΙΣΤΡΟΦΑ ΠΡΟΒΛΗΜΑΤΑ

7.1 Παραμετροποίηση αντιστρόφων προβλημάτων

Τα διακριτά αντίστροφα προβλήματα όπως έχουμε δει αντιμετωπίζουν σχέσεις παραμέτρων ενός φυσικού προβλήματος $\mathbf{m}$ και μετρήσεις κάποιου μεγέθους $\mathbf{d}$ που σχετίζεται με τις παραμέτρους μέσω κάποιας σχέσης $\mathbf{f}(\mathbf{m}, \mathbf{d}) = 0$ που την έχουμε χαρακτηρίσει «μοντέλο».

Σε προηγούμενα κεφάλαια είδαμε πως μπορεί να αντιμετωπισθεί ένα αντίστροφο πρόβλημα όταν το μοντέλο είναι της γραμμικής μορφής $\mathbf{Gm} = \mathbf{d}$. Επίσης είδαμε πως μπορούν να αξιοποιηθούν εκ προοιμίου πληροφορίες για παραμέτρους και μετρήσεις με την υπόθεση πάντα του γραμμικού μοντέλου.

Στο σημείο αυτό να παρατηρήσουμε ότι το είδος των μετρήσεων είναι σημαντικός παράγοντας για την αντιμετώπιση ενός αντιστρόφου προβλήματος. Διαφορετικό είδος μετρήσεων για τις ίδιες παραμέτρους μπορεί να οδηγήσει σε διαφορετικές εκτιμήσεις παραμέτρων.

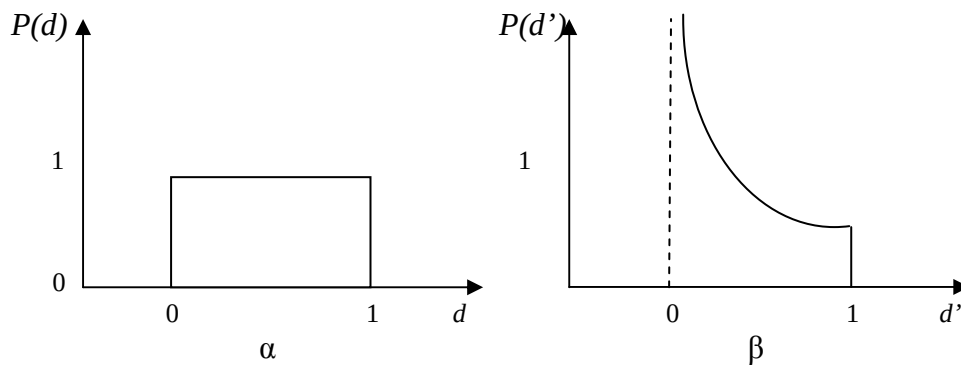
Για παράδειγμα ας δούμε την περίπτωση του υπολογισμού της εξίσωσης ευθείας που πρέπει να «περνά» από N σημεία που αντιστοιχούν σε ισάριθμα ζεύγη (z_i, d_i) . Εάν θεωρήσουμε ότι τα $z_i, i=1, \dots, N$ υπολογίζονται ακριβώς αλλά όχι και τα d_i και μάλιστα ότι η μέτρηση γίνεται με πιθανότητα λάθους που ακολουθεί κανονική κατανομή, παίρνουμε λύση ελαχίστων τετραγώνων. Μπορεί όμως να θεωρήσουμε ότι τα ακριβή στοιχεία είναι τα $d_i, i=1, \dots, N$ ενώ το μετρούμενο με λάθος είναι το $\mathbf{z}$.

Εάν εφαρμόσουμε όσα γνωρίζουμε και αναζητήσουμε λύση ελαχίστων τετραγώνων θα πάρουμε διαφορετική λύση, όπως μπορεί εύκολα κανείς να διαπιστώσει εάν επιχειρήσει να εκτιμήσει τις παραμέτρους της ευθείας που αντιστοιχεί στα ζεύγη (1,1), (2,2), (3,3), (4,5). Εάν θεωρήσουμε το πρώτο στοιχείο των ζευγών ως την ακριβή τετμημένη θα πάρουμε εξίσωση της μορφής $d = -0.5 + 1.3z$. Εάν αντίθετα θεωρήσουμε το δεύτερο στοιχείο των ζευγών ως την ακριβή τετμημένη, η νέα παραμετροποίηση θα αντιστοιχίσει το $z \rightarrow d'$ και το $d \rightarrow z'$ και η λύση ελαχίστων τετραγώνων θα δώσει $d' = -0.309 + 0.743z'$. Με δεδομένη την αντιστοίχιση στη νέα παραμετροποίηση μπορούμε να γράψουμε $z' = -0.416 + 1.34d'$. Υπάρχει συνεπώς σημαντική διαφορά ανάμεσα στις δύο ευθείες

Η αλλαγή στην παραμετροποίηση έχει τις εξής συνέπειες στην αντιμετώπιση ενός αντιστρόφου προβλήματος :

1. Αλλαγή στη συνάρτηση πυκνότητας πιθανότητας για τις νέες παραμέτρους.

Ας υποθέσουμε ότι έχουμε να αντιμετωπίσουμε ένα αντίστροφο πρόβλημα για το οποίο γνωρίζουμε ότι η συνάρτηση πυκνότητας πιθανότητας για τη μέτρηση d ακολουθεί ομοιόμορφη κατανομή στο διάστημα $[0,1]$ (Σχήμα 7.1)



Σχήμα 7.1 Συνάρτηση πυκνότητας πιθανότητας για τις d και d'

Εάν στη νέα παραμετροποίηση θεωρήσουμε ότι παίρνουμε $d' = d^2$ τότε η συνάρτηση πυκνότητας πιθανότητας για την d' υπολογίζεται από τη γνωστή σχέση :

$$P(d') = P(d) \frac{1}{g'(d)} \quad (7.1)$$

για $d' = g(d)$. Έτσι στην περίπτωσή μας έχουμε :

$$P(d') = 1/(2\sqrt{d'}) \quad (7.2)$$

Τα παραπάνω αποτυπώνονται στο σχήμα 7.1β στο οποίο φαίνεται η διαφορά στις δύο συναρτήσεις.

2. Αλλαγή στις εκτιμώμενες τιμές παραμέτρων του αντίστροφου προβλήματος.

Αυτό το είδαμε ήδη στο παράδειγμα της ευθείας γραμμής όταν ακολουθηθεί η μέθοδος επίλυσης με ελάχιστα τετράγωνα. Μπορούμε να το δούμε και στην περίπτωση που ακολουθηθεί η μέθοδος μέγιστης πιθανοφάνειας.

Στο ανωτέρω παράδειγμα εάν θεωρήσουμε ότι η αλλαγή στην παραμετροποίηση αφορά στη παράμετρο m για την οποία ισχύει $m' = m^2$ και αναζητήσουμε λύση μέγιστης πιθανοφάνειας μέσω της συνάρτησης πυκνότητας πιθανότητας, βλέπουμε ότι η παράμετρος m δεν παρουσιάζει μέγιστο ενώ αντίθετα η m' έχει μέγιστο στο

$m'=m=0$. Εάν θελήσουμε να υπολογίσουμε το μέσο για τις δύο κατανομές, θα πάρουμε διαφορετικά αποτελέσματα επίσης :

$$E[m] = \int_0^1 mP(m)dm = \int_0^1 m dm = \frac{1}{2} \quad (7.3)$$

$$E[m'] = \int_0^1 m'P(m')dm' = \int_0^1 \frac{1}{2} \sqrt{m'} dm' = \frac{1}{3} \quad (7.4)$$

Συνεπώς ο αναμενόμενος μέσος του m' δεν αντιστοιχεί στο τετράγωνο του αναμενόμενου μέσου του m .

Βλέπουμε λοιπόν μία ακόμη πτυχή των αντιστρόφων προβλημάτων που σχετίζεται με την ιδιότητά τους να μην είναι καλώς τεθειμένα. Ωστόσο η αλλαγή παραμετροποίησης μπορεί να βοηθήσει πολύ στο να αλλάξουμε τη μορφή ενός αντιστρόφου προβλήματος και για παράδειγμα από μη γραμμικό να γίνει γραμμικό και να λυθεί ευκολότερα. Στα πρώτα παραδείγματα των σημειώσεων είδαμε την περίπτωση της αξονικής τομογραφίας και πως μπορούμε να καταλήξουμε σε ένα προσεγγιστικό γραμμικό αντίστροφο πρόβλημα. Θα δώσουμε εδώ ένα ακόμη παράδειγμα πριν προχωρήσουμε στην ανάλυση των μη γραμμικών αντιστρόφων προβλημάτων.

Θεωρείστε ότι παράμετροι και μετρήσεις συνδέονται μέσω της σχέσης

$$d_i = m_1 \exp(m_2 z_i) \quad (7.5)$$

Η αλλαγή στην παραμετροποίηση $m'_1 = \ln(m_1)$, $m'_2 = m_2$ και $d'_i = \ln(d_i)$ μας μεταφέρει σε ένα αντίστροφο πρόβλημα της μορφής

$$d'_i = m'_1 + m'_2 z_i \quad (7.6)$$

που επιλύεται με τεχνικές που έχουμε ήδη δει, π.χ. με ελάχιστα τετράγωνα στην περίπτωση που θεωρήσουμε ότι οι μετρήσεις των d'_i υπόκεινται σε λάθη που ακολουθούν κανονική κατανομή. Προσοχή όμως. Η τελευταία υπόθεση συνεπάγεται ότι οι αρχικές μετρήσεις d_i δεν πρέπει να ακολουθούν κανονική κατανομή ! Συνεπώς η γραμμικοποίηση πρέπει να εφαρμόζεται με προσοχή. Έχει συνεπώς μεγάλη σημασία να δούμε πως μπορούν να αντιμετωπισθούν προβλήματα στα οποία μετρήσεις και παράμετροι σχετίζονται με μη γραμμικές σχέσεις.

7.2 Μη γραμμικά αντίστροφα προβλήματα.

7.2.1 Εισαγωγή

Ένα γενικό διακριτό αντίστροφο πρόβλημα για δεδομένα $\mathbf{d}$ διάστασης N και παραμέτρους $\mathbf{m}$ διάστασης M είναι της μορφής $\mathbf{f}(\mathbf{m}, \mathbf{d}) = 0$. Η μορφή είναι έμμεση όπως έχουμε πει αφού δεν διαχωρίζονται τα δεδομένα από τις μετρήσεις. Υποθέτουμε ότι $\dim(\mathbf{f}) = p \leq M + N$, υπόθεση που δεν είναι μακριά από την πραγματικότητα στα περισσότερα αντίστροφα προβλήματα που αντιμετωπίζουμε.

Για να δούμε πως μπορούμε να τα αντιμετωπίσουμε στη γενική περίπτωση θα ξεκινήσουμε με κάποιες υποθέσεις : Θεωρούμε ότι τα δεδομένα μας $\mathbf{d}$ ακολουθούν κανονική κατανομή με $[\text{cov} \mathbf{d}]$ γνωστό και αντίστοιχα έχουμε αρχικές πληροφορίες για τις παραμέτρους $\mathbf{m}$ ($\langle \mathbf{m} \rangle$ και $[\text{cov} \mathbf{m}]$) που επίσης ακολουθούν κανονική κατανομή.

Δημιουργούμε ένα καινούργιο διάνυσμα $\mathbf{x} = [\mathbf{d}, \mathbf{m}]^T$ για το οποίο ο πίνακας συνδιακύμανσης περιλαμβάνει τους πίνακες $[\text{cov} \mathbf{d}]$ και $[\text{cov} \mathbf{m}]$ σε διαγώνια μπλόκ. Γενικά τα μη διαγώνια στοιχεία θα μπορούσαν να είναι μη μηδενικά υποδεικνύοντας σχέση ανάμεσα στις παραμέτρους και στις μετρήσεις. Δεν θα μπούμε σε τόσο ειδικές περιπτώσεις πάντως.

Με βάση όσα γνωρίζουμε μέχρι τώρα, μπορούμε από τις υποθέσεις που κάναμε να πάρουμε την κατανομή πιθανότητας του $\mathbf{x}$ από τη σχέση :

$$P_A(\mathbf{x}) \propto \exp\left[-\frac{1}{2}(\mathbf{x} - \langle \mathbf{x} \rangle)^T [\text{cov} \mathbf{x}]^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle)\right] \quad (7.7)$$

όπου $\langle \mathbf{x} \rangle = [\mathbf{d}^{obs}, \langle \mathbf{m} \rangle]^T$.

Χρησιμοποιώντας την έννοια της μέγιστης πιθανοφάνειας, θα μπορούσαμε να θεωρήσουμε τη συνάρτηση $P_A(\mathbf{x})$ ως συνάρτηση πιθανοφάνειας και να αναζητήσουμε λύση στην ελαχιστοποίηση του όρου $\Phi = [(\mathbf{x} - \langle \mathbf{x} \rangle)^T [\text{cov} \mathbf{x}]^{-1} (\mathbf{x} - \langle \mathbf{x} \rangle)]$ κάτω από τον περιορισμό $\mathbf{f}(\mathbf{x}) = 0$.

Χρησιμοποιώντας πολλαπλασιαστές Lagrange, παίρνουμε ένα σύστημα εξισώσεων της μορφής :

$$\partial \Phi / \partial x_i = \sum_{j=1}^p 2\lambda_j \partial f_j / \partial x_i \quad (7.8)$$

που γράφεται στη μορφή

$$[\mathbf{x} - \langle \mathbf{x} \rangle]^T [\text{cov } \mathbf{x}]^{-1} = \boldsymbol{\lambda}^T \mathbf{F} \quad (7.8\alpha)$$

όπου $\boldsymbol{\lambda}$ είναι το διάνυσμα των πολλαπλασιαστών Lagrange και $\mathbf{F}$ είναι ο πίνακας των κλίσεων $\nabla \mathbf{f}$.

Από τη σχέση (7.8α) παίρνουμε διαδοχικά :

$$[\mathbf{x} - \langle \mathbf{x} \rangle]^T = [\text{cov } \mathbf{x}] \boldsymbol{\lambda}^T \mathbf{F} \quad (7.9)$$

$$[\mathbf{x} - \langle \mathbf{x} \rangle] = [\text{cov } \mathbf{x}] \mathbf{F}^T \boldsymbol{\lambda} \quad (7.9\alpha)$$

$$\mathbf{F}[\mathbf{x} - \langle \mathbf{x} \rangle] = \mathbf{F}[\text{cov } \mathbf{x}] \mathbf{F}^T \boldsymbol{\lambda} \quad (7.9\beta)$$

απ' όπου έχουμε :

$$\boldsymbol{\lambda} = (\mathbf{F}[\text{cov } \mathbf{x}] \mathbf{F}^T)^{-1} \mathbf{F}[\mathbf{x} - \langle \mathbf{x} \rangle] \quad (7.10)$$

Με αντικατάσταση στην 7.9α παίρνουμε :

$$[\mathbf{x} - \langle \mathbf{x} \rangle] = [\text{cov } \mathbf{x}] \mathbf{F}^T (\mathbf{F}[\text{cov } \mathbf{x}] \mathbf{F}^T)^{-1} \mathbf{F}[\mathbf{x} - \langle \mathbf{x} \rangle] \quad (7.11)$$

Η εξίσωση αυτή πρέπει να λυθεί ταυτόχρονα με την $\mathbf{f}(\mathbf{x}) = 0$ για να μας δώσει τις τιμές του $\mathbf{x}$ που δικαιολογούν με τον καλύτερο τρόπο τις μετρήσεις $\mathbf{d}^{obs}$ με δεδομένα τα στατιστικά των παραμέτρων.

$$[\mathbf{x} - \langle \mathbf{x} \rangle] = [\text{cov } \mathbf{x}] \mathbf{F}^T (\mathbf{F}[\text{cov } \mathbf{x}] \mathbf{F}^T)^{-1} \mathbf{F}[\mathbf{x} - \langle \mathbf{x} \rangle] - \mathbf{f}(\mathbf{x}) \quad (7.12)$$

Ωστόσο η εξίσωση 7.12 δεν είναι εύκολο να επιλυθεί, αφού τα στοιχεία $\mathbf{x}$ βρίσκονται και στα δύο μέρη και μάλιστα με έμμεση μη γραμμική μορφή.

7.2.2 Μία επαναληπτική μέθοδος για την επίλυση του μη γραμμικού προβλήματος.

Θα αναφερθούμε στη συνέχεια σε μία επαναληπτική μέθοδο για την επίλυση της εξίσωσης 7.12. Δεν θα συζητήσουμε την απόδειξη της σύγκλισης της μεθόδου, αλλά θα αρκεστούμε στην υπόθεση ότι εάν έχουμε μία ένδειξη για ένα αρχικό διάνυσμα $\mathbf{x}_0^{est}$ πολύ κοντά στη λύση μέγιστης πιθανοφάνειας, η μέθοδος θα συγκλίνει στο διάνυσμα μέγιστης πιθανοφάνειας.

Ξεκινώντας λοιπόν από το $\mathbf{x}_0^{est}$, δημιουργούμε διαδοχικές προσεγγίσεις ως εξής :

$$\mathbf{x}_{n+1}^{est} = \langle \mathbf{x} \rangle + [\text{cov } \mathbf{x}] \mathbf{F}_n^T (\mathbf{F}_n [\text{cov } \mathbf{x}] \mathbf{F}_n^T)^{-1} \mathbf{F}_n [\mathbf{x}_n^{est} - \langle \mathbf{x} \rangle] - \mathbf{f}(\mathbf{x}_n^{est}) \quad (7.13)$$

όπου τα μεγέθη με δείκτη n υπολογίζονται για τις τιμές $\mathbf{x}_n^{est}$.

Η διαδικασία απλοποιείται εάν έχουμε άμεση μη γραμμική μορφή για το αντίστροφο πρόβλημα, δηλαδή όταν υπάρχει σχέση της μορφής :

$$\mathbf{d} = \mathbf{g}(\mathbf{m}) \quad (7.14)$$

όπου η μη γραμμικότητα αποδίδεται μέσω της διανυσματικής συνάρτησης $\mathbf{g}$.

Στην περίπτωση αυτή παίρνουμε :

$$\mathbf{m}_{n+1}^{est} = \langle \mathbf{m} \rangle + \mathbf{G}_n^{-g} \left\{ \mathbf{d} - \mathbf{g}(\mathbf{m}_n^{est}) + \mathbf{G}_n [\mathbf{m}_n^{est} - \langle \mathbf{m} \rangle] \right\} \quad (7.15)$$

όπου

$$\begin{aligned} \mathbf{G}_n^{-g} &= [\text{cov} \mathbf{m}] \mathbf{G}_n^T \left\{ [\text{cov} \mathbf{d}] + \mathbf{G}_n [\text{cov} \mathbf{m}] \mathbf{G}_n^T \right\}^{-1} = \\ &= \left\{ [\text{cov} \mathbf{m}]^{-1} + \mathbf{G}_n^T [\text{cov} \mathbf{d}]^{-1} \mathbf{G}_n \right\}^{-1} \mathbf{G}_n^T [\text{cov} \mathbf{d}]^{-1} \end{aligned} \quad (7.16)$$

Τα στοιχεία του πίνακα $\mathbf{G}_n$ αντιστοιχούν στις παραγώγους $[\mathbf{G}_n]_{ij} = \partial g_i / \partial m_j$, $i=1, \dots, N$ και $j=1, \dots, M$ και υπολογίζονται για τα στοιχεία του διανύσματος $\mathbf{m}_n^{est}$.

Σημειώνουμε εδώ ότι τα ανωτέρω ισχύουν για ακριβή θεωρία. Επίσης σημειώνουμε ότι ο υπολογισμός του $[\text{cov} \mathbf{m}]$ δεν είναι εύκολος για μη κανονική κατανομή του $\mathbf{m}_n^{est}$ (η υπόθεση για κανονική κατανομή αφορά στα στατιστικά του $\mathbf{m}$). Ωστόσο με την υπόθεση πάντα για μικρές διαφορές ανάμεσα σε $\mathbf{m}$ και $\mathbf{m}_n^{est}$ μπορούμε να χρησιμοποιήσουμε τη γραμμική σχέση :

$$[\text{cov} \mathbf{m}_n^{est}] \simeq \mathbf{G}_n^{-g} [\text{cov} \mathbf{d}] \mathbf{G}_n^{-gT} + [\mathbf{I} - \mathbf{R}_n] [\text{cov} \mathbf{m}] [\mathbf{I} - \mathbf{R}_n]^T \quad (7.17)$$

όπου $\mathbf{R}_n = \mathbf{G}_n^{-g} \mathbf{G}_n$.

Σημειώνουμε τέλος ότι επαναληπτικές σχέσεις για τον υπολογισμό του $\mathbf{m}^{est}$ μπορούν να προκύψουν εάν θεωρήσουμε ανάπτυγμα Taylor στη μη γραμμική σχέση $\mathbf{d} = \mathbf{g}(\mathbf{m})$ γύρω από μια αρχική τιμή $\mathbf{m}_n^{est}$. Έχουμε αγνοώντας όρους τάξης $[\mathbf{m} - \mathbf{m}_n^{est}]^2 = \Delta \mathbf{m}_{n+1}$ και ανώτερους :

$$\mathbf{g}(\mathbf{m}) \simeq \mathbf{g}(\mathbf{m}_n^{est}) + \nabla \mathbf{g} [\mathbf{m} - \mathbf{m}_n^{est}] = \mathbf{g}(\mathbf{m}_n^{est}) + \mathbf{G}_n [\mathbf{m} - \mathbf{m}_n^{est}] \quad (7.18)$$

Επομένως γράφουμε :

$$\mathbf{G}_n \Delta \mathbf{m}_{n+1} = \mathbf{d} - \mathbf{g}(\mathbf{m}_n^{est}) \quad (7.19)$$

και

$$\mathbf{m}_{n+1}^{est} = \mathbf{m}_n^{est} + \Delta \mathbf{m}_{n+1} \quad (7.20)$$

Ωστόσο η σχέση 7.19 μας παραπέμπει σε γραμμικό αντίστροφο πρόβλημα της μορφής που έχουμε συναντήσει ήδη σε προηγούμενα κεφάλαια. Συνεπώς μπορεί να λυθεί με κάποια από τις μεθόδους που έχουμε επεξεργαστεί στα παρούσες σημειώσεις. Με αρχική υπόθεση για το $\mathbf{m}_0^{est}$ συνεπώς προκύπτουν διαδοχικά οι εκτιμήσεις για τα $\mathbf{m}_n^{est}$.

Ως εφαρμογή ας δούμε την απλή περίπτωση που έχουμε δεδομένα ασυσχέτιστα με ενιαία διασπορά σ_d^2 χωρίς εκ προοιμίου πληροφορία για τις παραμέτρους. Ο συνδυασμός της λύσης της 7.19 με ελάχιστα τετράγωνα και της 7.20 θα μας δώσει λύση της μορφής :

$$\mathbf{m}_{n+1}^{est} = \mathbf{m}_n^{est} + [\mathbf{G}_n^T \mathbf{G}_n]^{-1} \mathbf{G}_n^T [\mathbf{d} - \mathbf{g}(\mathbf{m}_n^{est})] \quad (7.21)$$

Κλείνοντας το εισαγωγικό κεφάλαιο για τα μη γραμμικά αντίστροφα προβλήματα υπογραμμίζουμε ότι οι απλοποιημένες θεωρήσεις που είδαμε μέχρι τώρα αφορούν περιπτώσεις που τα σημεία μέγιστης πιθανοφάνειας είναι καλά ορισμένα και οι αρχικές εκτιμήσεις για τη λύση $\mathbf{m}^{est}$ κοντά στα εν λόγω σημεία. Δυστυχώς στα περισσότερα προβλήματα η συνάρτηση πιθανοφάνειας παρουσιάζει πολλά τοπικά μέγιστα και η επιλογή ανάμεσα σε αυτά της βέλτιστης λύσης δεν είναι πάντα εύκολη. Αντίστοιχα οι αρχικές εκτιμήσεις για το $\mathbf{m}^{est}$ μπορεί να οδηγήσουν σε τοπικά μέγιστα που δεν αντιστοιχούν στη βέλτιστη λύση.

Για όλους αυτούς τους λόγους η συζήτηση για την επίλυση μη γραμμικών αντιστρόφων προβλημάτων είναι στην πραγματικότητα ατέρμων και εξειδικεύεται ανάλογα με το είδος του αντίστροφου προβλήματος, χρησιμοποιώντας την ανάλυση της θεωρίας (μοντέλου).