

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.

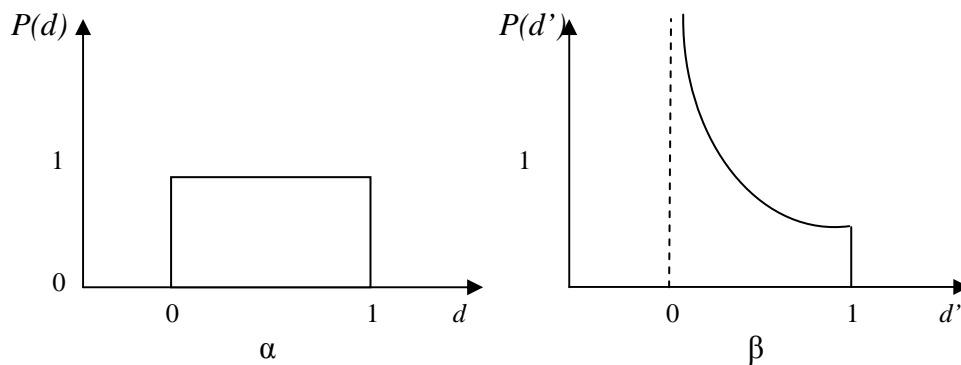
Για παράδειγμα ας δούμε την περίπτωση του υπολογισμού της εξίσωσης ευθείας που πρέπει να «περνά» από N σημεία που αντιστοιχούν σε ισάριθμα ζεύγη (z_i, d_i) . Εάν θεωρήσουμε ότι τα $z_i, i=1, \dots, N$ υπολογίζονται ακριβώς αλλά όχι και τα d_i και μάλιστα ότι η μέτρηση γίνεται με πιθανότητα λάθους που 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 :

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

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

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 τετιμένα. 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

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

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

Από τη σχέση (7.8α) παίρνομε διαδοχικά :

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

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

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

απ' όπου έχουμε :

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

Με αντικατάσταση στην 7.9α παίρνομε :

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

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 συζητήσομε 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 :

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

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 :

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

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).