

Διαλέξεις 12 & 13: Βέλτιστα Φίλτρα Σημειώσεις

Κωνσταντίνος Χατζηλυγερούδης
costashatz@upatras.gr

Δήμητρα Κουμούτσου
dkoumoutsou@upatras.gr

21 Δεκεμβρίου, 2025

Contents

1	Βέλτιστα Φίλτρα	3
2	Φιλτράρισμα Wiener: Ο Βέλτιστος Γραμμικός Εκτιμητής MMSE	4
2.1	Υποθέσεις και Μοντελοποίηση	4
2.2	Στόχος Γραμμικού MMSE	5
2.2.1	Εξισώσεις Wiener–Hopf (Κανονικές) και Λύση Κλειστής Μορφής	5
2.3	Αναλυτική Απόδειξη του Φίλτρου Wiener	6
2.3.1	Συνάρτηση Στόχου	6
2.3.2	Ανάπτυξη του Τετραγωνικού Όρου	6
2.3.3	Υπολογισμός Κάθε Όρου	6
2.3.4	Τελική Μορφή της Συνάρτησης Κόστους	7
2.3.5	Παράγωγος και Συνθήκη Βελτιστότητας	7
2.3.6	Υπολογισμός των $\mathbf{R}_{xd}$ και $\mathbf{R}_{xx}$ από το Μοντέλο Παρατήρησης	7
2.4	Πρακτική Χρήση του Φιλτραρίσματος Wiener	8
2.4.1	Περίπτωση 1: Όταν το επιθυμητό σήμα $\mathbf{d}[n]$ είναι διαθέσιμο (επιβλεπόμενη μάθηση / εκπαίδευση)	9
2.4.2	Περίπτωση 2: Όταν είναι γνωστές ή εκτιμήσιμες μόνο οι συσχετίσεις (μη επιβλεπόμενη μάθηση)	9
2.5	Λυμένα Παραδείγματα	11
2.5.1	Παράδειγμα 1: Αποθορυβοποίηση με $\mathbf{A} = \mathbf{I}$ (Γνωστές Συνδιακυμάνσεις)	11
2.5.2	Παράδειγμα 2: Αποθορυβοποίηση όταν το $\mathbf{d}[n]$ είναι Γνωστό στην Εκπαίδευση (Μάθηση του $\mathbf{W}$)	12
2.5.3	Παράδειγμα 3: Αποσυνέλιξη / Ανάμιξη με μη τετριμμένο $\mathbf{A}$	12
2.6	M -tap FIR Φιλτράρισμα Wiener	13
2.6.1	Συσσωρευμένο Διάνυσμα Δεδομένων και Παραμετροποίηση FIR	13
2.6.2	Σχεδίαση του FIR Φίλτρου Wiener από Στατιστικά Δεύτερης Τάξης	14
2.6.3	Πρακτικό Φιλτράρισμα Wiener σε Δεδομένα (Περίπτωση FIR)	14

2.6.4	Αριθμητικό Παράδειγμα (1Δ , $M = 2$)	15
2.6.5	Υλοποίηση σε Python	18
3	Φιλτράρισμα Kalman	23
3.1	Κίνητρο: Γιατί Φιλτράρισμα Kalman;	23
3.2	Μοντελοποίηση στον Χώρο Κατάστασης	23
3.3	Φίλτρο Kalman: Γραμμικές–Γκαουσιανές Υποθέσεις	24
3.4	Φιλτράρισμα Kalman ως Αναδρομική Bayesian Εκτίμηση	24
3.4.1	Αναδρομή Bayesian Φιλτράρισματος	24
3.4.2	Φίλτρο Kalman: Βήμα Πρόβλεψης	25
3.4.3	Φίλτρο Kalman: Βήμα Διόρθωσης	25
3.5	Bayesian Φιλτράρισμα: Το Γενικό Πλαίσιο	26
3.5.1	Κατάσταση, Έλεγχοι, Μετρήσεις και Πεποίθηση	26
3.5.2	Πιθανοτικές Υποθέσεις (Μοντέλο στο Χώρο Κατάστασης)	26
3.5.3	Αναδρομή του Bayes Φίλτρου	27
3.5.4	Πρόβλεψη μέσω του Νόμου Ολικής Πιθανότητας	27
3.5.5	Bayesian Φιλτράρισμα σε Δύο Βήματα	28
3.5.6	Φιλτράρισμα Kalman ως Ειδική Περίπτωση	28
3.6	Το Φίλτρο Kalman ως Αναδρομικό Φίλτρο Wiener	28
3.6.1	Κάθε χρονική στιγμή k λύνουμε ένα πρόβλημα LMMSE	28
3.6.2	Λύση Wiener $\Rightarrow$ Κέρδος Kalman	29
3.6.3	Ερμηνεία: «Αναδρομικό» Φιλτράρισμα Wiener	29
3.7	Παραδείγματα Kalman Φίλτρου	30
3.7.1	Παράδειγμα 1 (1Δ): Random Walk με Θορυβώδεις Μετρήσεις	30
3.7.2	Παράδειγμα 2 (2Δ): Μοντέλο Σταθερής Ταχύτητας με Μετρήσεις Θέσης	31
3.7.3	Υλοποίηση σε Python	33
3.8	Αλγόριθμος και Πρακτικές Παρατηρήσεις	37
3.8.1	Αλγόριθμος Φίλτρου Kalman (Πρόβλεψη–Διόρθωση)	37
3.8.2	Διαίσθηση I: Ο Ρόλος του Κέρδους Kalman	38
3.8.3	Διαίσθηση II: Καινοτομία = Νέα Πληροφορία	38
3.8.4	Διαίσθηση III: Γιατί η Αβεβαιότητα Μειώνεται μετά την Ενημέρωση	39
3.9	Πρακτικές Παρατηρήσεις	39
3.9.1	Όταν Παραβιάζονται οι Υποθέσεις	39
3.9.2	Επίδραση της Ρύθμισης των $\mathbf{Q}_k$ και $\mathbf{R}_k$	40
3.10	Φιλτράρισμα vs. Εξομάλυνση	41
3.10.1	Φιλτράρισμα vs. Εξομάλυνση	41
3.10.2	Εξομάλυνση Kalman: Αλγόριθμος Rauch–Tung–Striebel (RTS)	42
3.10.3	Batch Οπτική: MAP Εκτίμηση Ολόκληρης της Τροχιάς	42

1 Βέλτιστα Φίλτρα

Η κλασική φιλτράριση εισάγεται συχνά μέσω *εμπειρικά σχεδιασμένων* φίλτρων όπως φίλτρα FIR/IIR, κινητοί μέσοι όροι, καθώς και τυπικά φίλτρα ζώνης διέλευσης ή απόρριψης (band-pass / notch). Σε αυτό το πλαίσιο, πρώτα *επιλέγουμε* μια δομή φίλτρου και στη συνέχεια *ρυθμίζουμε* τις παραμέτρους του (π.χ. συχνότητες αποκοπής, τάξη φίλτρου, μήκος παραθύρου) βασιζόμενοι στη διαίσθηση για το φάσμα. Η προσέγγιση αυτή λειτουργεί καλά όταν ο φασματικός διαχωρισμός μεταξύ σήματος και θορύβου είναι σαφής και σταθερός.

Σε πολλά πραγματικά προβλήματα, ωστόσο, οι υποθέσεις πίσω από τη εμπειρική ρύθμιση είναι εύθραυστες. Ο θόρυβος μπορεί να είναι έγχρωμος αντί για λευκός και μπορεί να μεταβάλλεται με τον χρόνο (μη στάσιμη συμπεριφορά). Επιπλέον, τα φάσματα σήματος και θορύβου μπορεί να αλληλεπικαλύπτονται, οπότε σπάνια υπάρχει μια μοναδική «σωστή» συχνότητα αποκοπής. Διαφορετικές σχεδιαστικές επιλογές οδηγούν επομένως σε διαφορετικούς συμβιβασμούς μεταξύ *καταστολής θορύβου και παραμόρφωσης σήματος*, και δεν είναι πάντοτε προφανές πώς να επιλεγεί ο καλύτερος δυνατός συμβιβασμός.

Αυτό οδηγεί φυσικά στο ερώτημα:

Μπορούμε να σχεδιάσουμε ένα φίλτρο αυτόματα, με βέλτιστο τρόπο, εφόσον ορίσουμε τι σημαίνει «βέλτιστο»;

Η οπτική του βέλτιστου *φιλτραρίσματος* απαντά σε αυτό το ερώτημα μετατρέποντας τον σχεδιασμό φίλτρων σε πρόβλημα εκτίμησης. Αντί να επιλέγουμε ένα φίλτρο ευριστικά, καθορίζουμε (i) έναν **στόχο** (π.χ. εκτίμηση ή ανακατασκευή ενός καθαρού σήματος), (ii) ένα **κριτήριο** (συνηθέστερα, την ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος), και (iii) ένα **μοντέλο** για το σήμα και τον θόρυβο (τυπικά μέσω στατιστικών δεύτερης τάξης). Το προκύπτον φίλτρο προκύπτει τότε ως εκείνο που επιτυγχάνει την *καλύτερη δυνατή απόδοση* υπό αυτές τις υποθέσεις.

Στο **φιλτράρισμα Wiener**, εξετάζουμε έναν στατικό (μη αναδρομικό) βέλτιστο γραμμικό εκτιμητή: δεδομένης μιας θορυβώδους διαδικασίας παρατήρησης, αναζητούμε το γραμμικό φίλτρο που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα εκτίμησης. Αυτό οδηγεί στις εξισώσεις Wiener-Hopf και αποδίδει μια κλειστής μορφής βέλτιστη λύση σε όρους συναρτήσεων συσχέτισης (ή, ισοδύναμα, φασματικών πυκνοτήτων ισχύος). Το φιλτράρισμα Wiener είναι ιδιαίτερα χρήσιμο όταν μπορούμε να υποθέσουμε στασιμότητα και όταν ένα batch/offline φίλτρο είναι αποδεκτό.

Στο **φιλτράρισμα Kalman**, περνάμε από τη στατική εκτίμηση στη *δυναμική* εκτίμηση. Εδώ, το μέγεθος ενδιαφέροντος εξελίσσεται με τον χρόνο σύμφωνα με ένα *μοντέλο στο χώρο κατάστασης*, και οι μετρήσεις φτάνουν διαδοχικά. Αντί να υπολογίζουμε μια μοναδική λύση παρτίδας, επιδιώκουμε έναν *αναδρομικό* εκτιμητή που ενημερώνει την πεποίθησή του για την κατάσταση κάθε φορά που γίνεται διαθέσιμη μια νέα μέτρηση. Υπό υποθέσεις γραμμικής δυναμικής, γραμμικών μετρήσεων και Γκαουσιανού θορύβου, το φίλτρο Kalman παρέχει τη βέλτιστη εκτίμηση *ελάχιστου μέσου τετραγωνικού σφάλματος* (MMSE) της κατάστασης και το επιτυγχάνει αποδοτικά, μεταφέροντας μόνο την τρέχουσα εκτίμηση και την αβεβαιότητά της (συνδιακύμανση).

Έτσι, το φιλτράρισμα Wiener και το φιλτράρισμα Kalman μπορούν να θεωρηθούν ως δύο κλασικές περιπτώσεις βέλτιστο φιλτράρισμα: το φιλτράρισμα Wiener αντιμετωπίζει τη βέλτιστη γραμμική εκτίμηση MMSE σε (τυπικά) στάσιμα περιβάλλοντα, ενώ το φιλτράρισμα Kalman γενικεύει την ιδέα σε χρονικά μεταβαλλόμενα συστήματα με μια θεμελιωμένη αναδρομική ενημέρωση. Μαζί, αναδεικνύουν το κεντρικό μήνυμα αυτής της ενότητας: μόλις ορίσουμε ένα μοντέλο και ένα κριτήριο βελτιστότητας, το φίλτρο προκύπτει από τα μαθηματικά και όχι από αποσπασματική ρύθμιση παραμέτρων.

2 Φιλτράρισμα Wiener: Ο Βέλτιστος Γραμμικός Εκτιμητής MMSE

Στη συνέχεια παραθέτουμε το φίλτρο Wiener, δηλαδή τον κλασικό βέλτιστο γραμμικό εκτιμητή που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα υπό τυπικές υποθέσεις δεύτερης τάξης (και, προαιρετικά, Γκαουσιανότητας).

2.1 Υποθέσεις και Μοντελοποίηση

Το φιλτράρισμα Wiener μπορεί να ερμηνευθεί ως η Bayesian λύση MMSE (μέση τιμή της εκ των υστέρων κατανομής) σε ένα γραμμικό–Γκαουσιανό πλαίσιο και, ισοδύναμα, ως η λύση γραμμικού MMSE (LMMSE) όταν υποθέτουμε μόνο στατιστικά δεύτερης τάξης. Συγκεκριμένα, υιοθετούμε το ακόλουθο μοντέλο.

1) Γραμμικό μοντέλο παρατήρησης. Υποθέτουμε ότι οι μετρήσεις $\mathbf{x}[n] \in \mathbb{R}^p$ παράγονται από το άγνωστο (καθαρό) σήμα $\mathbf{d}[n] \in \mathbb{R}^q$ μέσω

$$\mathbf{x}[n] = \mathbf{A} \mathbf{d}[n] + \mathbf{v}[n], \quad (2.1)$$

όπου $\mathbf{A} \in \mathbb{R}^{p \times q}$ είναι ένας γνωστός γραμμικός τελεστής. Στην αποθρομβοποίηση, $\mathbf{A} = \mathbf{I}$, ενώ στη αποσυνέλιξη ή αποθόλωση, η $\mathbf{A}$ είναι συχνά ένας τελεστής συνέλιξης.

2) Μοντέλο σήματος δεύτερης τάξης (WSS). Το καθαρό σήμα $\mathbf{d}[n]$ μοντελοποιείται ως τυχαία διεργασία στάσιμη κατά τη ευρεία έννοια (wide-sense stationary), χαρακτηριζόμενη από τη συνάρτηση αυτοσυσχέτισης $\mathbf{R}_{dd}[k]$ (ή, ισοδύναμα, από τη φασματική πυκνότητα ισχύος $\mathbf{S}_{dd}(\omega)$ στη βαθμωτή/WSS περίπτωση).

3) Προσθετικό μοντέλο θορύβου (WSS) και ασυσχέτιση σήματος–θορύβου. Ο θόρυβος $\mathbf{v}[n]$ μοντελοποιείται ως WSS με αυτοσυσχέτιση $\mathbf{R}_{vv}[k]$ (ή φασματική πυκνότητα ισχύος $\mathbf{S}_{vv}(\omega)$), και υποθέτουμε ότι είναι ασυσχέτιστος με το σήμα:

$$\mathbf{R}_{dv}[k] = \mathbf{0}. \quad (2.2)$$

4) Γκαουσιανότητα (Bayesian θεμελίωση). Αν, επιπλέον, τα $\mathbf{d}$ και $\mathbf{v}$ μοντελοποιηθούν ως Γκαουσιανές διεργασίες, τότε η εκ των υστέρων κατανομή $p(\mathbf{d} | \mathbf{x})$ είναι Γκαουσιανή και η μέση τιμή της είναι γραμμική ως προς το $\mathbf{x}$. Αυτό συνεπάγεται ότι ο εκτιμητής MMSE συμπίπτει με τον βέλτιστο γραμμικό εκτιμητή.

5) Περιορισμός σε γραμμικούς εκτιμητές. Αναζητούμε έναν εκτιμητή εντός της κλάσης των γραμμικών απεικονίσεων, ο οποίος (για σταθερό χρονικό δείκτη n) γράφεται ως

$$\hat{\mathbf{d}}[n] = \mathbf{W}^T \mathbf{x}[n], \quad (2.3)$$

όπου $\mathbf{W} \in \mathbb{R}^{p \times q}$ (ώστε $\mathbf{W}^T \mathbf{x} \in \mathbb{R}^q$). Στο πλαίσιο LTI/WSS, αυτή η γραμμική απεικόνιση αντιστοιχεί σε συνέλιξη με μια χρονστική απόκριση (δηλαδή σε ένα LTI φίλτρο). Το φίλτρο Wiener είναι η συγκεκριμένη επιλογή της $\mathbf{W}$ που ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα.

2.2 Στόχος Γραμμικού MMSE

Αναζητούμε τον γραμμικό εκτιμητή που ελαχιστοποιεί το αναμενόμενο τετραγωνικό σφάλμα ανακατασκευής:

$$\mathbf{W}_\star = \arg \min_{\mathbf{W}} \mathbb{E} \left[\|\mathbf{d}[n] - \mathbf{W}^T \mathbf{x}[n]\|_2^2 \right]. \quad (2.4)$$

Για λόγους συμβολικής απλότητας, παραλείπουμε τον δείκτη $[n]$ και γράφουμε $\mathbf{x}, \mathbf{d}$ όταν δεν υπάρχει κίνδυνος σύγχυσης.

Αναπτύσσοντας το τετραγωνικό μέλος προκύπτει

$$\begin{aligned} J(\mathbf{W}) &= \mathbb{E}[(\mathbf{d} - \mathbf{W}^T \mathbf{x})^T (\mathbf{d} - \mathbf{W}^T \mathbf{x})] \\ &= \text{tr}(\mathbb{E}\{\mathbf{d}\mathbf{d}^T\}) - 2 \text{tr}(\mathbf{W}^T \mathbb{E}\{\mathbf{x}\mathbf{d}^T\}) + \text{tr}(\mathbf{W}^T \mathbb{E}\{\mathbf{x}\mathbf{x}^T\} \mathbf{W}). \end{aligned} \quad (2.5)$$

Ορίζουμε τους πίνακες συσχέτισης

$$\mathbf{R}_{dd} \triangleq \mathbb{E}\{\mathbf{d}\mathbf{d}^T\}, \quad \mathbf{R}_{xd} \triangleq \mathbb{E}\{\mathbf{x}\mathbf{d}^T\}, \quad \mathbf{R}_{xx} \triangleq \mathbb{E}\{\mathbf{x}\mathbf{x}^T\}. \quad (2.6)$$

Τότε η (2.5) γράφεται συνοπτικά ως

$$J(\mathbf{W}) = \text{tr}(\mathbf{R}_{dd}) - 2 \text{tr}(\mathbf{W}^T \mathbf{R}_{xd}) + \text{tr}(\mathbf{W}^T \mathbf{R}_{xx} \mathbf{W}). \quad (2.7)$$

2.2.1 Εξισώσεις Wiener–Hopf (Κανονικές) και Λύση Κλειστής Μορφής

Παραγωγίζοντας τη $J(\mathbf{W})$ ως προς $\mathbf{W}$ και θέτοντας την παράγωγο ίση με μηδέν, προκύπτουν οι κανονικές εξισώσεις (Wiener–Hopf):

$$\frac{\partial J}{\partial \mathbf{W}} = \mathbf{0} \implies \mathbf{R}_{xx} \mathbf{W}_\star = \mathbf{R}_{xd}. \quad (2.8)$$

Υποθέτοντας ότι ο $\mathbf{R}_{xx}$ είναι αντιστρέψιμος, λαμβάνουμε τη λύση Wiener/LMMSE

$$\boxed{\mathbf{W}_\star = \mathbf{R}_{xx}^{-1} \mathbf{R}_{xd}.} \quad (2.9)$$

Χρησιμοποιώντας το μοντέλο παρατήρησης (2.1) και την υπόθεση ότι $\mathbf{v} \perp \mathbf{d}$, μπορούμε να εκφράσουμε τους $\mathbf{R}_{xd}$ και $\mathbf{R}_{xx}$ σε όρους των συνδιακυμάνσεων του μοντέλου. Πράγματι,

$$\mathbf{R}_{xd} = \mathbb{E}\{(\mathbf{A}\mathbf{d} + \mathbf{v})\mathbf{d}^T\} = \mathbf{A} \mathbb{E}\{\mathbf{d}\mathbf{d}^T\} + \mathbb{E}\{\mathbf{v}\mathbf{d}^T\} = \mathbf{A}\mathbf{R}_{dd}, \quad (2.10)$$

$$\begin{aligned} \mathbf{R}_{xx} &= \mathbb{E}\{(\mathbf{A}\mathbf{d} + \mathbf{v})(\mathbf{A}\mathbf{d} + \mathbf{v})^T\} \\ &= \mathbf{A}\mathbf{R}_{dd}\mathbf{A}^T + \mathbf{R}_{vv}, \end{aligned} \quad (2.11)$$

όπου $\mathbf{R}_{vv} \triangleq \mathbb{E}\{\mathbf{v}\mathbf{v}^T\}$. Αντικαθιστώντας τις (2.10)–(2.11) στην (2.9), προκύπτει η συνήθως χρησιμοποιούμενη λύση κλειστής μορφής:

$$\boxed{\mathbf{W}_\star = (\mathbf{A}\mathbf{R}_{dd}\mathbf{A}^T + \mathbf{R}_{vv})^{-1} \mathbf{A}\mathbf{R}_{dd}.} \quad (2.12)$$

Παρατήρηση (συσχέτιση έναντι συνδιακύμανσης). Υπό τη συνήθη υπόθεση διεργασιών μηδενικού μέσου, $\mathbb{E}\{\mathbf{x}\} = \mathbf{0}$ και $\mathbb{E}\{\mathbf{d}\} = \mathbf{0}$, οι πίνακες συσχέτισης και συνδιακύμανσης συμπίπτουν:

$$\text{Cov}[\mathbf{x}, \mathbf{x}] = \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^T] = \mathbb{E}\{\mathbf{x}\mathbf{x}^T\} = \mathbf{R}_{xx}, \quad (2.13)$$

και ομοίως

$$\text{Cov}[\mathbf{x}, \mathbf{d}] = \mathbb{E}[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{d} - \boldsymbol{\mu}_d)^T] = \mathbb{E}\{\mathbf{x}\mathbf{d}^T\} = \mathbf{R}_{xd}. \quad (2.14)$$

2.3 Αναλυτική Απόδειξη του Φίλτρου Wiener

Παρακάτω δίνουμε μια αναλυτική απόδειξη του φίλτρου Wiener, ελαχιστοποιώντας ρητά το μέσο τετραγωνικό σφάλμα μέσα στην κλάση των γραμμικών εκτιμητών. Σε όλη την ανάπτυξη υποθέτουμε πραγματικές τυχαίες διανυσματικές μεταβλητές και διεργασίες μηδενικού μέσου.

Έστω $\mathbf{x} \in \mathbb{R}^p$ το διάνυσμα παρατήρησης, $\mathbf{d} \in \mathbb{R}^q$ το επιθυμητό (καθαρό) σήμα, και $\mathbf{W} \in \mathbb{R}^{p \times q}$ ο γραμμικός εκτιμητής έτσι ώστε $\hat{\mathbf{d}} = \mathbf{W}^T \mathbf{x} \in \mathbb{R}^q$.

2.3.1 Συνάρτηση Στόχου

Ο στόχος γραμμικού MMSE είναι

$$J(\mathbf{W}) = \mathbb{E}[\|\mathbf{d} - \mathbf{W}^T \mathbf{x}\|_2^2] = \mathbb{E}[(\mathbf{d} - \mathbf{W}^T \mathbf{x})^T (\mathbf{d} - \mathbf{W}^T \mathbf{x})]. \quad (2.15)$$

Χρησιμοποιώντας τον τελεστή ίχνους (trace), ο οποίος επιτρέπει βολική επεξεργασία εκφράσεων με πίνακες, μπορούμε να γράψουμε

$$\begin{aligned} J(\mathbf{W}) &= \mathbb{E}[\text{tr}((\mathbf{d} - \mathbf{W}^T \mathbf{x})^T (\mathbf{d} - \mathbf{W}^T \mathbf{x}))] \\ &= \text{tr}(\mathbb{E}[(\mathbf{d} - \mathbf{W}^T \mathbf{x})^T (\mathbf{d} - \mathbf{W}^T \mathbf{x})]), \end{aligned} \quad (2.16)$$

όπου χρησιμοποιήθηκε η γραμμικότητα του ίχνους και της αναμενόμενης τιμής.

2.3.2 Ανάπτυξη του Τετραγωνικού Όρου

Αρχικά αναπτύσσουμε τον τετραγωνικό όρο:

$$(\mathbf{d} - \mathbf{W}^T \mathbf{x})^T (\mathbf{d} - \mathbf{W}^T \mathbf{x}) = \mathbf{d}^T \mathbf{d} - \mathbf{d}^T \mathbf{W}^T \mathbf{x} - \mathbf{x}^T \mathbf{W} \mathbf{d} + \mathbf{x}^T \mathbf{W} \mathbf{W}^T \mathbf{x}. \quad (2.17)$$

Αντικαθιστώντας αυτή την ανάπτυξη στη συνάρτηση στόχου, παίρνουμε

$$J(\mathbf{W}) = \text{tr}(\mathbb{E}[\mathbf{d}^T \mathbf{d}]) - \text{tr}(\mathbb{E}[\mathbf{d}^T \mathbf{W}^T \mathbf{x}]) - \text{tr}(\mathbb{E}[\mathbf{x}^T \mathbf{W} \mathbf{d}]) + \text{tr}(\mathbb{E}[\mathbf{x}^T \mathbf{W} \mathbf{W}^T \mathbf{x}]). \quad (2.18)$$

2.3.3 Υπολογισμός Κάθε Όρου

Πρώτος όρος. Χρησιμοποιώντας $\mathbf{d}^T \mathbf{d} = \text{tr}(\mathbf{d} \mathbf{d}^T)$, προκύπτει

$$\text{tr}(\mathbb{E}[\mathbf{d}^T \mathbf{d}]) = \text{tr}(\mathbb{E}[\mathbf{d} \mathbf{d}^T]) = \text{tr}(\mathbf{R}_{dd}), \quad (2.19)$$

όπου $\mathbf{R}_{dd} \triangleq \mathbb{E}[\mathbf{d} \mathbf{d}^T]$.

Δεύτερος όρος. Χρησιμοποιώντας την κυκλική μετατόπιση στο ίχνος,

$$\begin{aligned} \text{tr}(\mathbb{E}[\mathbf{d}^T \mathbf{W}^T \mathbf{x}]) &= \mathbb{E}[\text{tr}(\mathbf{d}^T \mathbf{W}^T \mathbf{x})] = \mathbb{E}[\text{tr}(\mathbf{W}^T \mathbf{x} \mathbf{d}^T)] \\ &= \text{tr}(\mathbf{W}^T \mathbb{E}[\mathbf{x} \mathbf{d}^T]) = \text{tr}(\mathbf{W}^T \mathbf{R}_{xd}), \end{aligned} \quad (2.20)$$

όπου $\mathbf{R}_{xd} \triangleq \mathbb{E}[\mathbf{x} \mathbf{d}^T]$.

Τρίτος όρος. Ομοίως,

$$\begin{aligned}\text{tr}(\mathbb{E}[\mathbf{x}^T \mathbf{W} \mathbf{d}]) &= \mathbb{E}[\text{tr}(\mathbf{x}^T \mathbf{W} \mathbf{d})] = \mathbb{E}[\text{tr}(\mathbf{W} \mathbf{d} \mathbf{x}^T)] \\ &= \text{tr}(\mathbf{W} \mathbb{E}[\mathbf{d} \mathbf{x}^T]) = \text{tr}(\mathbf{W} \mathbf{R}_{dx}),\end{aligned}\quad (2.21)$$

με $\mathbf{R}_{dx} = \mathbf{R}_{xd}^T$. Χρησιμοποιώντας $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^T)$, αυτό γίνεται

$$\text{tr}(\mathbf{W} \mathbf{R}_{dx}) = \text{tr}(\mathbf{R}_{xd} \mathbf{W}^T) = \text{tr}(\mathbf{W}^T \mathbf{R}_{xd}). \quad (2.22)$$

Τέταρτος όρος. Τέλος,

$$\begin{aligned}\text{tr}(\mathbb{E}[\mathbf{x}^T \mathbf{W} \mathbf{W}^T \mathbf{x}]) &= \mathbb{E}[\text{tr}(\mathbf{x}^T \mathbf{W} \mathbf{W}^T \mathbf{x})] \\ &= \mathbb{E}[\text{tr}(\mathbf{W}^T \mathbf{x} \mathbf{x}^T \mathbf{W})] \\ &= \text{tr}(\mathbf{W}^T \mathbb{E}[\mathbf{x} \mathbf{x}^T] \mathbf{W}) = \text{tr}(\mathbf{W}^T \mathbf{R}_{xx} \mathbf{W}),\end{aligned}\quad (2.23)$$

όπου $\mathbf{R}_{xx} \triangleq \mathbb{E}[\mathbf{x} \mathbf{x}^T]$.

2.3.4 Τελική Μορφή της Συνάρτησης Κόστους

Συλλέγοντας όλους τους όρους, η συνάρτηση στόχου γράφεται ως

$$\boxed{J(\mathbf{W}) = \text{tr}(\mathbf{R}_{dd}) - 2 \text{tr}(\mathbf{W}^T \mathbf{R}_{xd}) + \text{tr}(\mathbf{W}^T \mathbf{R}_{xx} \mathbf{W})}. \quad (2.24)$$

2.3.5 Παράγωγος και Συνθήκη Βελτιστότητας

Παραγωγίζουμε τώρα τη $J(\mathbf{W})$ ως προς $\mathbf{W}$. Οι ταυτότητες λογισμού πινάκων που χρειαζόμαστε είναι:

$$\frac{\partial}{\partial \mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{A}) = \mathbf{A}, \quad \frac{\partial}{\partial \mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{A} \mathbf{W}) = (\mathbf{A} + \mathbf{A}^T) \mathbf{W}.$$

Εφόσον ο $\mathbf{R}_{xx}$ είναι συμμετρικός, $\mathbf{R}_{xx} = \mathbf{R}_{xx}^T$, παίρνουμε

$$\frac{\partial J}{\partial \mathbf{W}} = -2\mathbf{R}_{xd} + 2\mathbf{R}_{xx} \mathbf{W}. \quad (2.25)$$

Θέτοντας την παράγωγο ίση με μηδέν, προκύπτουν οι εξισώσεις Wiener–Hopf

$$\boxed{\mathbf{R}_{xx} \mathbf{W}_* = \mathbf{R}_{xd}}, \quad (2.26)$$

οι οποίες οδηγούν άμεσα στη λύση του φίλτρου Wiener $\mathbf{W}_* = \mathbf{R}_{xx}^{-1} \mathbf{R}_{xd}$.

2.3.6 Υπολογισμός των $\mathbf{R}_{xd}$ και $\mathbf{R}_{xx}$ από το Μοντέλο Παρατήρησης

Υπολογίζουμε τώρα ρητά τη δια-συνδιακύμανση $\mathbf{R}_{xd}$ και τη συνδιακύμανση των παρατηρήσεων $\mathbf{R}_{xx}$ από το υποτιθέμενο γραμμικό μοντέλο μετρήσεων. Υπενθυμίζουμε ότι οι παρατηρήσεις παράγονται σύμφωνα με

$$\mathbf{x}[n] = \mathbf{A} \mathbf{d}[n] + \mathbf{v}[n], \quad (2.27)$$

όπου $\mathbb{E}[\mathbf{d}] = \mathbf{0}$, $\mathbb{E}[\mathbf{v}] = \mathbf{0}$, και ο θόρυβος είναι ασυσχέτιστος με το σήμα, δηλαδή $\mathbf{v} \perp \mathbf{d}$.

Δια-συνδιακύμανση παρατήρησης και σήματος. Εξ ορισμού,

$$\mathbf{R}_{xd} \triangleq \mathbb{E}\{\mathbf{x}\mathbf{d}^T\}. \quad (2.28)$$

Αντικαθιστώντας το μοντέλο παρατήρησης (2.27), παίρνουμε

$$\begin{aligned} \mathbf{R}_{xd} &= \mathbb{E}\{(\mathbf{A}\mathbf{d} + \mathbf{v})\mathbf{d}^T\} \\ &= \mathbf{A} \mathbb{E}\{\mathbf{d}\mathbf{d}^T\} + \mathbb{E}\{\mathbf{v}\mathbf{d}^T\}. \end{aligned} \quad (2.29)$$

Επειδή $\mathbf{v}$ και $\mathbf{d}$ είναι ασυσχέτιστα, ο δεύτερος όρος μηδενίζεται: $\mathbb{E}\{\mathbf{v}\mathbf{d}^T\} = \mathbf{0}$. Άρα,

$$\boxed{\mathbf{R}_{xd} = \mathbf{A}\mathbf{R}_{dd}}. \quad (2.30)$$

Αυτο-συνδιακύμανση των παρατηρήσεων. Αντίστοιχα, η συνδιακύμανση των παρατηρήσεων είναι

$$\mathbf{R}_{xx} \triangleq \mathbb{E}\{\mathbf{x}\mathbf{x}^T\} = \mathbb{E}\{(\mathbf{A}\mathbf{d} + \mathbf{v})(\mathbf{A}\mathbf{d} + \mathbf{v})^T\}. \quad (2.31)$$

Αναπτύσσοντας το γινόμενο προκύπτει

$$\mathbf{R}_{xx} = \mathbf{A}\mathbb{E}\{\mathbf{d}\mathbf{d}^T\}\mathbf{A}^T + \mathbf{A}\mathbb{E}\{\mathbf{d}\mathbf{v}^T\} + \mathbb{E}\{\mathbf{v}\mathbf{d}^T\}\mathbf{A}^T + \mathbb{E}\{\mathbf{v}\mathbf{v}^T\}. \quad (2.32)$$

Και πάλι, η ασυσχέτιση συνεπάγεται ότι οι μικτοί όροι μηδενίζονται:

$$\mathbb{E}\{\mathbf{d}\mathbf{v}^T\} = \mathbb{E}\{\mathbf{v}\mathbf{d}^T\} = \mathbf{0}.$$

Έτσι, η συνδιακύμανση των παρατηρήσεων ανάγεται σε

$$\boxed{\mathbf{R}_{xx} = \mathbf{A}\mathbf{R}_{dd}\mathbf{A}^T + \mathbf{R}_{vv}}, \quad (2.33)$$

όπου $\mathbf{R}_{vv} \triangleq \mathbb{E}\{\mathbf{v}\mathbf{v}^T\}$.

Παρατήρηση. Οι εξισώσεις (2.30) και (2.33) δείχνουν ότι το φίλτρο Wiener εξαρτάται μόνο από τον γραμμικό τελεστή $\mathbf{A}$ και από στατιστικά δεύτερης τάξης του σήματος και του θορύβου. Δεν απαιτούνται ροπές υψηλότερης τάξης, και η Γκαουσιανότητα είναι απαραίτητη μόνο εάν επιθυμούμε να ερμηνεύσουμε τη λύση ως τον πλήρη Bayesian εκτιμητή MMSE αντί ως τη λύση γραμμικού MMSE.

2.4 Πρακτική Χρήση του Φιλτραρίσματος Wiener

Η λύση Wiener γράφεται σε όρους πινάκων συσχέτισης όπως οι $\mathbf{R}_{xx}$ και $\mathbf{R}_{xd}$, οι οποίοι είναι αναμενόμενες τιμές ως προς τη (άγνωστη) διαδικασία που παράγει τα δεδομένα. Στην πράξη, σπάνια έχουμε άμεση πρόσβαση σε αυτές τις αναμενόμενες τιμές και πρέπει να τις προσεγγίσουμε από δεδομένα. Υπάρχουν δύο συνήθεις περιπτώσεις:

2.4.1 Περίπτωση 1: Όταν το επιθυμητό σήμα $d[n]$ είναι διαθέσιμο (επιβλεπόμενη μάθηση / εκπαίδευση)

Σε ορισμένες εφαρμογές μπορούμε να αποκτήσουμε (τουλάχιστον περιστασιακά) ζεύγη παρατηρήσεων και αληθινές τιμές,

$$(\mathbf{x}[n], \mathbf{d}[n]), \quad n = 1, \dots, N,$$

για παράδειγμα από πειράματα βαθμονόμησης, αισθητήρες υψηλής ποιότητας, offline μεταεπεξεργασία, ή από προσομοιωμένα δεδομένα. Σε αυτή την περίπτωση, το φίλτρο Wiener μπορεί να υπολογιστεί μέσω ελαχιστοποίησης εμπειρικού κινδύνου (empirical risk minimization): ελαχιστοποιούμε το μέσο τετραγωνικό σφάλμα στα δείγματα

$$\widehat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{N} \sum_{n=1}^N \|\mathbf{d}[n] - \mathbf{W}^T \mathbf{x}[n]\|_2^2. \quad (2.34)$$

Αυτό δεν είναι παρά πολυδιάστατα ελάχιστα τετράγωνα (γραμμική παλινδρόμηση) με εισόδους $\mathbf{x}[n]$ και στόχους $\mathbf{d}[n]$.

Ορίζουμε τους πίνακες δεδομένων

$$\mathbf{X} \triangleq [\mathbf{x}[1] \ \dots \ \mathbf{x}[N]] \in \mathbb{R}^{p \times N}, \quad \mathbf{D} \triangleq [\mathbf{d}[1] \ \dots \ \mathbf{d}[N]] \in \mathbb{R}^{q \times N}.$$

Τότε το (2.34) γράφεται ως $\min_{\mathbf{W}} \|\mathbf{D} - \mathbf{W}^T \mathbf{X}\|_F^2$, και οι κανονικές εξισώσεις δίνουν

$$\boxed{\widehat{\mathbf{W}} = (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}\mathbf{D}^T} \quad (\text{υπό την υπόθεση ότι ο } \mathbf{X}\mathbf{X}^T \text{ είναι αντιστρέψιμος}). \quad (2.35)$$

Ισοδύναμα, εισάγοντας τις εκτιμήσεις συσχέτισης από δείγμα

$$\hat{\mathbf{R}}_{xx} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}[n]\mathbf{x}[n]^T = \frac{1}{N} \mathbf{X}\mathbf{X}^T, \quad \hat{\mathbf{R}}_{xd} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}[n]\mathbf{d}[n]^T = \frac{1}{N} \mathbf{X}\mathbf{D}^T,$$

αναχτούμε την οικεία μορφή Wiener

$$\boxed{\widehat{\mathbf{W}} = \hat{\mathbf{R}}_{xx}^{-1} \hat{\mathbf{R}}_{xd}}. \quad (2.36)$$

Online χρήση μετά την εκπαίδευση. Μόλις υπολογιστεί το $\widehat{\mathbf{W}}$, η εφαρμογή είναι πλήρως online και εξαιρετικά γρήγορη: για κάθε νέα παρατήρηση $\mathbf{x}[n]$ υπολογίζουμε

$$\hat{\mathbf{d}}[n] = \widehat{\mathbf{W}}^T \mathbf{x}[n].$$

Αν το περιβάλλον μεταβάλλεται (drift), μπορούμε να ενημερώνουμε το $\widehat{\mathbf{W}}$ online με αναδρομικά ελάχιστα τετράγωνα (RLS) ή με στοχαστική κατάβαση κλίσης στη στιγμιαία τετραγωνική απώλεια $\|\mathbf{d}[n] - \mathbf{W}^T \mathbf{x}[n]\|_2^2$ όταν η αληθής τιμή γίνεται διαθέσιμη.

2.4.2 Περίπτωση 2: Όταν είναι γνωστές ή εκτιμήσιμες μόνο οι συσχετίσεις (μη επιβλεπόμενη μάθηση)

Συχνά, το καθαρό σήμα $\mathbf{d}[n]$ δεν παρατηρείται ποτέ άμεσα, άρα δεν μπορούμε να σχηματίσουμε το $\hat{\mathbf{R}}_{xd}$ από ζευγοποιημένα δείγματα. Παρόλα αυτά, το φιλτράρισμα Wiener παραμένει

εφαρμόσιμο εφόσον γνωρίζουμε (ή μπορούμε να εκτιμήσουμε) τα στατιστικά δεύτερης τάξης που υπαγορεύει το μοντέλο.

Μια συνηθισμένη περίπτωση είναι ο προσθετικός θόρυβος με $\mathbf{A} = \mathbf{I}$:

$$\mathbf{x}[n] = \mathbf{d}[n] + \mathbf{v}[n], \quad \mathbf{v} \perp \mathbf{d}.$$

Από τις ταυτότητες που προέκυψαν νωρίτερα,

$$\mathbf{R}_{xx} = \mathbf{R}_{dd} + \mathbf{R}_{vv}, \quad \mathbf{R}_{xd} = \mathbf{R}_{dd}.$$

Άρα η λύση Wiener γίνεται

$$\mathbf{W}_* = (\mathbf{R}_{dd} + \mathbf{R}_{vv})^{-1} \mathbf{R}_{dd}. \quad (2.37)$$

Στη βαθμωτή περίπτωση LTI/WSS, αυτό αντιστοιχεί στο πεδίο της συχνότητας στη γνωστή μορφή PSD

$$H(\omega) = \frac{S_{dd}(\omega)}{S_{dd}(\omega) + S_{vv}(\omega)}. \quad (2.38)$$

Πώς αποκτούμε τα απαιτούμενα στατιστικά; Στην πράξη, ακολουθείται συνήθως μία από τις παρακάτω διαδρομές:

- **Βαθμονόμηση:** εκτίμηση του $\mathbf{R}_{vv}$ (ή του $S_{vv}(\omega)$) από χρονικά διαστήματα όπου το σήμα απουσιάζει.
- **Μοντελοποίηση σήματος:** υπόθεση παραμετρικής πρότερης κατανομής για το $\mathbf{d}$ (π.χ. AR/ARMA) και εκτίμηση των παραμέτρων της από δεδομένα, κάτι που στη συνέχεια καθορίζει το $\mathbf{R}_{dd}$ (ή το $S_{dd}(\omega)$).
- **Γνώση πεδίου:** χρήση γνωστής ή αναμενόμενης μορφής PSD για το σήμα και/ή τον θόρυβο (π.χ. θόρυβος $1/f$, σήματα περιορισμένου εύρους ζώνης).
- **Μη στάσιμες συνθήκες / online προσαρμογή:** εκτίμηση συσχετίσεων σε ολισθαίνον παράθυρο ώστε το $S_{vv}(\omega)$ ή το $S_{xx}(\omega)$ να προσαρμόζεται σε μεταβαλλόμενες συνθήκες.

Οπτική online υλοποίησης. Ακόμη και όταν οι συσχετίσεις εκτιμώνται, η online λειτουργία έχει τυπικά δομή δύο σταδίων: (i) ενημέρωση των στατιστικών δεύτερης τάξης (ή των παραμετρικών μοντέλων τους) από τα εισερχόμενα δεδομένα και (ii) ενημέρωση ή εφαρμογή του φίλτρου ώστε να παραχθεί το $\hat{\mathbf{d}}[n]$. Σε στάσιμες συνθήκες, το φίλτρο μπορεί να σχεδιαστεί μία φορά και στη συνέχεια να εφαρμοστεί online. Σε αργά μεταβαλλόμενες συνθήκες, μπορούμε να επανεκτιμούμε τα στατιστικά και να επανασχεδιάζουμε το φίλτρο Wiener περιοδικά (κατά μπλοκ) ή συνεχώς (προσαρμοστικό φιλτράρισμα).

Σύνοψη. Αν το $\mathbf{d}[n]$ είναι διαθέσιμο (έστω και περιστασιακά), το φίλτρο Wiener μπορεί να εκτιμηθεί από δεδομένα μέσω ελαχίστων τετραγώνων και στη συνέχεια να εφαρμοστεί online. Αν το $\mathbf{d}[n]$ δεν είναι διαθέσιμο, το φιλτράρισμα Wiener παραμένει δυνατό εφόσον τα σχετικά στατιστικά δεύτερης τάξης (συσχετίσεις/PSDs) είναι γνωστά ή μπορούν να εκτιμηθούν από τη ροή μετρήσεων υπό υποθέσεις μοντελοποίησης.

2.5 Λυμένα Παραδείγματα

Για να γίνει το φιλτράρισμα Wiener πιο απτό, δουλεύουμε τώρα μικρά παραδείγματα χαμηλής διάστασης. Χρησιμοποιούμε 2-διάστατα σήματα ώστε κάθε πίνακας και κάθε βήμα να ελέγχεται εύκολα «με το χέρι».

2.5.1 Παράδειγμα 1: Αποθορυβοποίηση με $A = I$ (Γνωστές Συνδιακυμάνσεις)

Διατύπωση. Θεωρούμε το μοντέλο παρατήρησης

$$\mathbf{x} = \mathbf{d} + \mathbf{v}, \quad \mathbb{E}[\mathbf{d}] = \mathbf{0}, \mathbb{E}[\mathbf{v}] = \mathbf{0}, \mathbf{v} \perp \mathbf{d},$$

με

$$\mathbf{R}_{dd} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{R}_{vv} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

Τότε

$$\mathbf{R}_{xx} = \mathbf{R}_{dd} + \mathbf{R}_{vv} = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix}, \quad \mathbf{R}_{xd} = \mathbb{E}[\mathbf{x}\mathbf{d}^T] = \mathbf{R}_{dd}.$$

Λύση Wiener. Ο πίνακας Wiener είναι

$$\mathbf{W}_* = \mathbf{R}_{xx}^{-1} \mathbf{R}_{xd} = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Υπολογίζουμε το αντίστροφο:

$$\det(\mathbf{R}_{xx}) = 3 \cdot 5 - 1 \cdot 1 = 14, \quad \mathbf{R}_{xx}^{-1} = \frac{1}{14} \begin{bmatrix} 5 & -1 \\ -1 & 3 \end{bmatrix}.$$

Άρα,

$$\mathbf{W}_* = \frac{1}{14} \begin{bmatrix} 5 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 9 & 3 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} \frac{9}{14} & \frac{3}{14} \\ \frac{1}{14} & \frac{5}{14} \end{bmatrix}.$$

Εφαρμογή του φίλτρου. Δεδομένου ενός μετρούμενου διανύσματος $\mathbf{x} \in \mathbb{R}^2$, η εκτίμηση είναι

$$\hat{\mathbf{d}} = \mathbf{W}_*^T \mathbf{x}.$$

Για παράδειγμα, αν $\mathbf{x} = [1 \ 2]^T$, τότε

$$\hat{\mathbf{d}} = \begin{bmatrix} \frac{9}{14} & \frac{1}{14} \\ \frac{3}{14} & \frac{5}{14} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{11}{14} \\ \frac{13}{14} \end{bmatrix}.$$

Ερμηνεία. Η δεύτερη συνιστώσα έχει μεγαλύτερη διασπορά θορύβου (3 έναντι 1), άρα το φίλτρο «συρρικνώνει» και αναμειγνύει τις συνιστώσες ασύμμετρα. Η εκτίμηση δεν είναι απλώς $a\mathbf{x}$: εκμεταλλεύεται τη συσχέτιση στο $\mathbf{d}$ (τους εκτός διαγωνίου όρους του $\mathbf{R}_{dd}$) ώστε να «δανείζεται πληροφορία» από το καθαρότερο κανάλι.

2.5.2 Παράδειγμα 2: Αποθορυβοποίηση όταν το $d[n]$ είναι Γνωστό στην Εκπαίδευση (Μάθηση του W)

Διατύπωση. Υποθέτουμε ότι διαθέτουμε N επιβλεπόμενα δείγματα εκπαίδευσης $\{(\mathbf{x}[n], \mathbf{d}[n])\}_{n=1}^N$, με $\mathbf{x}[n], \mathbf{d}[n] \in \mathbb{R}^2$. Ορίζουμε

$$\mathbf{X} = [\mathbf{x}[1] \ \cdots \ \mathbf{x}[N]] \in \mathbb{R}^{2 \times N}, \quad \mathbf{D} = [\mathbf{d}[1] \ \cdots \ \mathbf{d}[N]] \in \mathbb{R}^{2 \times N}.$$

Η εκτίμηση ελαχίστων τετραγώνων για τον πίνακα Wiener είναι

$$\widehat{\mathbf{W}} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{D}^T.$$

Ένα μικρό αριθμητικό παράδειγμα ($N = 3$). Θεωρούμε το σύνολο εκπαίδευσης

$$\mathbf{x}[1] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{d}[1] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{x}[2] = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{d}[2] = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}[3] = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{d}[3] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Τότε

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Υπολογίζουμε

$$\mathbf{X}\mathbf{X}^T = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{X}\mathbf{D}^T = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}.$$

Εφόσον

$$(\mathbf{X}\mathbf{X}^T)^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

προκύπτει

$$\widehat{\mathbf{W}} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 3 & -1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & \frac{2}{3} \end{bmatrix}.$$

Εφαρμογή του φίλτρου. Για μια νέα μέτρηση $\mathbf{x}$, η εκτίμηση είναι $\hat{\mathbf{d}} = \widehat{\mathbf{W}}^T \mathbf{x}$.

2.5.3 Παράδειγμα 3: Αποσυνέλιξη / Ανάμιξη με μη τετριμμένο $\mathbf{A}$

Διατύπωση. Υποθέτουμε έναν γραμμικό τελεστή ανάμιξης («θόλωση»)

$$\mathbf{x} = \mathbf{A}\mathbf{d} + \mathbf{v}, \quad \mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{R}_{dd} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{R}_{vv} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Τότε

$$\mathbf{R}_{xd} = \mathbf{A}\mathbf{R}_{dd} = \mathbf{A}, \quad \mathbf{R}_{xx} = \mathbf{A}\mathbf{R}_{dd}\mathbf{A}^T + \mathbf{R}_{vv} = \mathbf{A}\mathbf{A}^T + \mathbf{I}.$$

Υπολογίζουμε

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{R}_{xx} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}.$$

Αντιστροφή:

$$\det(\mathbf{R}_{xx}) = 3 \cdot 2 - 1 \cdot 1 = 5, \quad \mathbf{R}_{xx}^{-1} = \frac{1}{5} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}.$$

Άρα

$$\mathbf{W}_\star = \mathbf{R}_{xx}^{-1} \mathbf{R}_{xd} = \frac{1}{5} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}.$$

Επομένως $\hat{\mathbf{d}} = \mathbf{W}_\star^T \mathbf{x}$ με

$$\mathbf{W}_\star^T = \frac{1}{5} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}.$$

Ερμηνεία. Επειδή ο $\mathbf{A}$ αναμειγνύει τις δύο συνιστώσες, ο βέλτιστος εκτιμητής πρέπει ταυτόχρονα (i) να «αναιρεί» την ανάμιξη (σαν αντίστροφος) και (ii) να κάνει εξομάλυνση ή κανονικοποίηση έναντι του θορύβου. Το φίλτρο Wiener πραγματοποιεί ακριβώς αυτόν τον συμβιβασμό αυτόματα μέσω των όρων συνδιακύμανσης.

2.6 M -tap FIR Φιλτράρισμα Wiener

Μέχρι τώρα, θεωρήσαμε έναν χωρίς μνήμη γραμμικό εκτιμητή της μορφής

$$\hat{\mathbf{d}}[n] = \mathbf{W}^T \mathbf{x}[n], \quad \mathbf{x}[n] \in \mathbb{R}^p, \quad \mathbf{d}[n] \in \mathbb{R}^q.$$

Αυτό είναι κατάλληλο όταν το $\mathbf{x}[n]$ περιέχει όλη τη σχετική πληροφορία στη χρονική στιγμή n . Ωστόσο, σε πολλά προβλήματα επεξεργασίας σήματος, η επιθυμητή εκτίμηση στη χρονική στιγμή n θα πρέπει να εξαρτάται από ένα παράθυρο προηγούμενων παρατηρήσεων, κάτι που οδηγεί φυσικά σε ένα αιτιατό FIR φίλτρο M συντελεστών (taps).

2.6.1 Συσσωρευμένο Διάνυσμα Δεδομένων και Παραμετροποίηση FIR

Για να προκύψει ένας αιτιατός FIR εκτιμητής M συντελεστών, σχηματίζουμε ένα διευρυμένο διάνυσμα δεδομένων στοιβάζοντας την τρέχουσα και τις προηγούμενες μετρήσεις:

$$\mathbf{x}_n \triangleq \begin{bmatrix} \mathbf{x}[n] \\ \mathbf{x}[n-1] \\ \vdots \\ \mathbf{x}[n-M+1] \end{bmatrix} \in \mathbb{R}^{pM}. \quad (2.39)$$

Στη συνέχεια εφαρμόζουμε έναν γραμμικό εκτιμητή στο $\mathbf{x}_n$:

$$\hat{\mathbf{d}}[n] = \mathbf{W}^T \mathbf{x}_n, \quad \mathbf{W} \in \mathbb{R}^{(pM) \times q}. \quad (2.40)$$

Διαμερίζουμε τον $\mathbf{W}$ σε M μπλοκ,

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_0 \\ \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_{M-1} \end{bmatrix}, \quad \mathbf{W}_k \in \mathbb{R}^{p \times q}.$$

Τότε η (2.40) γράφεται ως

$$\hat{\mathbf{d}}[n] = [\mathbf{W}_0^T \quad \mathbf{W}_1^T \quad \cdots \quad \mathbf{W}_{M-1}^T] \begin{bmatrix} \mathbf{x}[n] \\ \mathbf{x}[n-1] \\ \vdots \\ \mathbf{x}[n-M+1] \end{bmatrix}. \quad (2.41)$$

Ισοδύναμα, αυτό είναι ακριβώς μια αιτιατή FIR συνέλιξη:

$$\hat{\mathbf{d}}[n] = \sum_{k=0}^{M-1} \mathbf{W}_k^T \mathbf{x}[n-k]. \quad (2.42)$$

Άρα, η συνήθης μορφή LMMSE $\hat{\mathbf{d}}[n] = \mathbf{W}^T \mathbf{x}_n$ γίνεται ένα FIR φίλτρο Wiener M συντελεστών απλώς επιλέγοντας το $\mathbf{x}_n$ ώστε να περιλαμβάνει M προηγούμενα δείγματα.

2.6.2 Σχεδίαση του FIR Φίλτρου Wiener από Στατιστικά Δεύτερης Τάξης

Το βέλτιστο FIR φίλτρο Wiener προκύπτει ελαχιστοποιώντας το MSE ως προς τον συσσωρευμένο εκτιμητή (2.40), δηλαδή,

$$\mathbf{W}_* = \arg \min_{\mathbf{W}} \mathbb{E} \left[\|\mathbf{d}[n] - \mathbf{W}^T \mathbf{x}_n\|_2^2 \right].$$

Η λύση έχει την ίδια αλγεβρική μορφή όπως και πριν:

$$\mathbf{W}_* = \mathbf{R}_{x_n x_n}^{-1} \mathbf{R}_{x_n d}, \quad (2.43)$$

όπου

$$\mathbf{R}_{x_n x_n} \triangleq \mathbb{E}\{\mathbf{x}_n \mathbf{x}_n^T\} \in \mathbb{R}^{pM \times pM}, \quad \mathbf{R}_{x_n d} \triangleq \mathbb{E}\{\mathbf{x}_n \mathbf{d}[n]^T\} \in \mathbb{R}^{pM \times q}.$$

Στην περίπτωση στάσιμων κατά τη ευρεία έννοια (WSS) διεργασιών, οι πίνακες αυτοί έχουν δομή μπλοκ-Toeplitz. Ορίζουμε τους πίνακες αυτοσυσχέτισης παρατηρήσεων για υστέρηση k

$$\mathbf{R}_{xx}[k] \triangleq \mathbb{E}\{\mathbf{x}[n] \mathbf{x}[n-k]^T\}, \quad k \in \mathbb{Z},$$

και τους πίνακες συσχέτισης

$$\mathbf{R}_{xd}[k] \triangleq \mathbb{E}\{\mathbf{x}[n] \mathbf{d}[n-k]^T\}.$$

Τότε

$$\mathbf{R}_{x_n x_n} = \begin{bmatrix} \mathbf{R}_{xx}[0] & \mathbf{R}_{xx}[1] & \cdots & \mathbf{R}_{xx}[M-1] \\ \mathbf{R}_{xx}[-1] & \mathbf{R}_{xx}[0] & \cdots & \mathbf{R}_{xx}[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{xx}[-M+1] & \mathbf{R}_{xx}[-M+2] & \cdots & \mathbf{R}_{xx}[0] \end{bmatrix} \in \mathbb{R}^{pM \times pM}, \quad (2.44)$$

και

$$\mathbf{R}_{x_n d} = \begin{bmatrix} \mathbf{R}_{xd}[0] \\ \mathbf{R}_{xd}[1] \\ \vdots \\ \mathbf{R}_{xd}[M-1] \end{bmatrix} \in \mathbb{R}^{pM \times q}. \quad (2.45)$$

Για WSS $\mathbf{x}[n]$, ισχύει επίσης η συμμετρία

$$\mathbf{R}_{xx}[-k] = \mathbf{R}_{xx}[k]^T.$$

2.6.3 Πρακτικό Φιλτράρισμα Wiener σε Δεδομένα (Περίπτωση FIR)

Έστω ότι μας δίνονται μετρήσεις $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ και είτε (i) ένα καθαρό σήμα αναφοράς $\mathbf{d}[n]$ (επιβλεπόμενη μάθηση) είτε (ii) ένα μοντέλο που μας επιτρέπει να σχηματίσουμε τις απαιτούμενες συσχετίσεις (μη-επιβλεπόμενη μάθηση).

Βήμα 1: Επιλογή της μορφής του φίλτρου. Επιλέγουμε μήκος FIR M και χρησιμοποιούμε το συσσωρευμένο διάνυσμα $\mathbf{x}_n$ όπως στο (2.39).

Βήμα 2: Εκτίμηση των απαιτούμενων στατιστικών δεύτερης τάξης. Μια συνηθισμένη αμερόληπτη εκτίμηση δείγματος για τους πίνακες αυτοσυσχέτισης των παρατηρήσεων είναι

$$\hat{\mathbf{R}}_{xx}[k] = \frac{1}{N-k} \sum_{n=k}^{N-1} \mathbf{x}[n]\mathbf{x}[n-k]^T, \quad k = 0, \dots, M-1. \quad (2.46)$$

Αν υπάρχει καθαρό σήμα αναφοράς, η συσχέτιση μπορεί να εκτιμηθεί ως

$$\hat{\mathbf{R}}_{xd}[k] = \frac{1}{N-k} \sum_{n=k}^{N-1} \mathbf{x}[n]\mathbf{d}[n-k]^T, \quad k = 0, \dots, M-1. \quad (2.47)$$

Αν το $\mathbf{d}$ δεν παρατηρείται, χρησιμοποιείται αντ' αυτού ένα μοντέλο (π.χ. $\mathbf{x} = \mathbf{A}\mathbf{d} + \mathbf{v}$), μαζί με εκτιμήσεις των στατιστικών του σήματος και του θορύβου, ώστε να σχηματιστούν οι $\hat{\mathbf{R}}_{xx}[k]$ και $\hat{\mathbf{R}}_{xd}[k]$.

Βήμα 3: Κατασκευή του γραμμικού συστήματος Wiener–Hopf. Χρησιμοποιώντας τις (2.44)–(2.45) με τις εκτιμημένες συσχετίσεις, σχηματίζουμε

$$\hat{\mathbf{R}}_{x_n x_n} \in \mathbb{R}^{pM \times pM}, \quad \hat{\mathbf{R}}_{x_n d} \in \mathbb{R}^{pM \times q}.$$

Βήμα 4: Επίλυση για τον βέλτιστο πίνακα συντελεστών.

$$\boxed{\mathbf{W}^* = \hat{\mathbf{R}}_{x_n x_n}^{-1} \hat{\mathbf{R}}_{x_n d}.} \quad (2.48)$$

(Ισοδύναμα, επιλύουμε το γραμμικό σύστημα $\hat{\mathbf{R}}_{x_n x_n} \mathbf{W}^* = \hat{\mathbf{R}}_{x_n d}$.)

Βήμα 5: Φιλτράρισμα των δεδομένων. Τέλος, διαμερίζουμε το $\mathbf{W}^*$ στους συντελεστές $\mathbf{W}_0^*, \dots, \mathbf{W}_{M-1}^*$ και εφαρμόζουμε

$$\boxed{\hat{\mathbf{d}}[n] = \sum_{k=0}^{M-1} \mathbf{W}_k^{*T} \mathbf{x}[n-k].} \quad (2.49)$$

Παρατήρηση (αιτιότητα και καθυστέρηση). Η παραπάνω κατασκευή οδηγεί σε ένα αιτιατό FIR φίλτρο που χρησιμοποιεί μόνο τρέχοντα και παρελθοντικά δείγματα. Αν επιτρέπεται μη αιτιατή εξομάλυνση (π.χ. σε offline επεξεργασία), μπορούμε να συμπεριλάβουμε και μελλοντικά δείγματα, ώστε να προκύψει ένας δίπλευρος Wiener εξομαλυντής, ο οποίος συχνά επιτυγχάνει μικρότερο MSE με τίμημα την εισαγωγή καθυστέρησης.

2.6.4 Αριθμητικό Παράδειγμα (1Δ , $M = 2$)

Παρουσιάζουμε την πλήρη κατασκευή Wiener–Hopf για ένα αιτιατό FIR φίλτρο δύο συντελεστών

$$\hat{\mathbf{d}}[n] = w_0 x[n] + w_1 x[n-1], \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}.$$

Για $M = 2$, το σύστημα Wiener–Hopf γράφεται

$$\underbrace{\begin{bmatrix} r_{xx}[0] & r_{xx}[1] \\ r_{xx}[1] & r_{xx}[0] \end{bmatrix}}_{\mathbf{R}} \mathbf{w} = \underbrace{\begin{bmatrix} r_{xd}[0] \\ r_{xd}[1] \end{bmatrix}}_{\mathbf{p}}, \quad \mathbf{w} = \mathbf{R}^{-1} \mathbf{p}. \quad (2.50)$$

A) Επιβλεπόμενη περίπτωση (συσχετίσεις από δείγματα). Υποθέτουμε ότι διαθέτουμε $N = 5$ ζευγοποιημένα δείγματα $\{x[n], d[n]\}$:

$$d = [1, 0, -1, 0, 1], \quad v = [0.2, -0.1, 0.1, -0.2, 0], \quad x = d+v = [1.2, -0.1, -0.9, -0.2, 1].$$

Βήμα 1: υπολογισμός του $\hat{r}_{xx}[0]$.

$$\hat{r}_{xx}[0] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]^2 = \frac{1}{5} \left(1.2^2 + (-0.1)^2 + (-0.9)^2 + (-0.2)^2 + 1^2 \right).$$

Υπολογίζουμε κάθε όρο:

$$1.2^2 = 1.44, \quad (-0.1)^2 = 0.01, \quad (-0.9)^2 = 0.81, \quad (-0.2)^2 = 0.04, \quad 1^2 = 1.$$

Άθροισμα:

$$1.44 + 0.01 + 0.81 + 0.04 + 1 = 3.30 \quad \Rightarrow \quad \boxed{\hat{r}_{xx}[0] = 3.30/5 = 0.66.}$$

Βήμα 2: υπολογισμός του $\hat{r}_{xx}[1]$.

$$\hat{r}_{xx}[1] = \frac{1}{N-1} \sum_{n=1}^{N-1} x[n]x[n-1] = \frac{1}{4} \left(x[1]x[0] + x[2]x[1] + x[3]x[2] + x[4]x[3] \right).$$

Υπολογίζουμε τα γινόμενα:

$$\begin{aligned} x[1]x[0] &= (-0.1)(1.2) = -0.12, & x[2]x[1] &= (-0.9)(-0.1) = 0.09, \\ x[3]x[2] &= (-0.2)(-0.9) = 0.18, & x[4]x[3] &= (1)(-0.2) = -0.20. \end{aligned}$$

Άθροισμα:

$$-0.12 + 0.09 + 0.18 - 0.20 = -0.05 \quad \Rightarrow \quad \boxed{\hat{r}_{xx}[1] = -0.05/4 = -0.0125.}$$

Βήμα 3: υπολογισμός του $\hat{r}_{xd}[0]$.

$$\hat{r}_{xd}[0] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]d[n] = \frac{1}{5} \left(x[0]d[0] + x[1]d[1] + x[2]d[2] + x[3]d[3] + x[4]d[4] \right).$$

Υπολογίζουμε:

$$\begin{aligned} x[0]d[0] &= 1.2 \cdot 1 = 1.2, & x[1]d[1] &= (-0.1) \cdot 0 = 0, & x[2]d[2] &= (-0.9) \cdot (-1) = 0.9, \\ x[3]d[3] &= (-0.2) \cdot 0 = 0, & x[4]d[4] &= (1) \cdot 1 = 1. \end{aligned}$$

Άθροισμα:

$$1.2 + 0 + 0.9 + 0 + 1 = 3.1 \quad \Rightarrow \quad \boxed{\hat{r}_{xd}[0] = 3.1/5 = 0.62.}$$

Βήμα 4: υπολογισμός του $\hat{r}_{xd}[1] = \mathbb{E}\{x[n]d[n-1]\}$.

$$\hat{r}_{xd}[1] = \frac{1}{N-1} \sum_{n=1}^{N-1} x[n]d[n-1] = \frac{1}{4} \left(x[1]d[0] + x[2]d[1] + x[3]d[2] + x[4]d[3] \right).$$

Υπολογίζουμε:

$$x[1]d[0] = (-0.1) \cdot 1 = -0.1, \quad x[2]d[1] = (-0.9) \cdot 0 = 0,$$

$$x[3]d[2] = (-0.2) \cdot (-1) = 0.2, \quad x[4]d[3] = 1 \cdot 0 = 0.$$

Άθροισμα:

$$-0.1 + 0 + 0.2 + 0 = 0.1 \quad \Rightarrow \quad \boxed{\hat{r}_{xd}[1] = 0.1/4 = 0.025.}$$

Βήμα 5: κατασκευή και επίλυση του συστήματος Wiener-Hopf. Χρησιμοποιώντας την (2.50),

$$\hat{\mathbf{R}} = \begin{bmatrix} 0.66 & -0.0125 \\ -0.0125 & 0.66 \end{bmatrix}, \quad \hat{\mathbf{p}} = \begin{bmatrix} 0.62 \\ 0.025 \end{bmatrix}.$$

Το αντίστροφο ενός πίνακα 2×2 της μορφής $\begin{bmatrix} a & b \\ b & a \end{bmatrix}$ είναι

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix}^{-1} = \frac{1}{a^2 - b^2} \begin{bmatrix} a & -b \\ -b & a \end{bmatrix}.$$

Εδώ $a = 0.66$, $b = -0.0125$, άρα

$$a^2 - b^2 = 0.66^2 - (-0.0125)^2 = 0.4356 - 0.00015625 = 0.43544375.$$

Επομένως,

$$\hat{\mathbf{R}}^{-1} = \frac{1}{0.43544375} \begin{bmatrix} 0.66 & 0.0125 \\ 0.0125 & 0.66 \end{bmatrix}.$$

Υπολογίζουμε $\hat{\mathbf{w}} = \hat{\mathbf{R}}^{-1} \hat{\mathbf{p}}$. Πρώτα ο πολλαπλασιασμός:

$$\begin{bmatrix} 0.66 & 0.0125 \\ 0.0125 & 0.66 \end{bmatrix} \begin{bmatrix} 0.62 \\ 0.025 \end{bmatrix} = \begin{bmatrix} 0.4095125 \\ 0.02425 \end{bmatrix}.$$

Διαίρεση με 0.43544375:

$$\boxed{\hat{w}_0 \approx 0.9404, \quad \hat{w}_1 \approx 0.0557.}$$

Βήμα 6: εφαρμογή του φίλτρου σε δείγμα. Για παράδειγμα, στο $n = 4$ (με $x[4] = 1$ και $x[3] = -0.2$),

$$\hat{d}[4] = \hat{w}_0 x[4] + \hat{w}_1 x[3] \approx 0.9293.$$

B) Prior AR(1) + λευκός θόρυβος. Υποθέτουμε

$$x[n] = d[n] + v[n], \quad v[n] \text{ λευκός με } \text{Var}(v) = \sigma_v^2, \quad v \perp d,$$

και ένα prior AR(1)

$$d[n] = a d[n-1] + e[n], \quad e[n] \sim \mathcal{N}(0, \sigma_e^2), \quad |a| < 1.$$

Για AR(1), οι στάσιμες αυτοσυσχετίσεις είναι

$$r_{dd}[0] = \frac{\sigma_e^2}{1 - a^2}, \quad r_{dd}[1] = a r_{dd}[0].$$

Χρησιμοποιώντας $x = d + v$ με λευκό θόρυβο,

$$r_{xx}[0] = r_{dd}[0] + \sigma_v^2, \quad r_{xx}[1] = r_{dd}[1], \quad r_{xd}[0] = r_{dd}[0], \quad r_{xd}[1] = r_{dd}[1].$$

Ερμηνεία. Το επιβλεπόμενο παράδειγμα χρησιμοποιεί πολύ λίγα δείγματα ($N = 5$), οπότε οι εκτιμήσεις συσχέτισης και κατ' επέκταση οι συντελεστές ($\hat{w}_0, \hat{w}_1$) είναι θορυβώδεις. Οι συντελεστές που βρίσκουμε στη 2η περίπτωση χρησιμοποιούν στατιστικά δεύτερης τάξης από το υποτιθέμενο μοντέλο AR(1)+λευκού θορύβου, άρα είναι «καθαροί» και σταθεροί. Στην πράξη, με αρκετά δεδομένα εκπαίδευσης (ή με κανονικοποίηση), η επιβλεπόμενη προσέγγιση συγκλίνει στη 2η περίπτωση όταν το μοντέλο είναι σωστό.

2.6.5 Υλοποίηση σε Python

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 def make_ar1(N, a, sigma_e, seed=0):
5     """Generate AR(1): d[n] = a d[n-1] + e[n], e~N(0,sigma_e^2)."""
6     rng = np.random.default_rng(seed)
7     e = rng.normal(0.0, sigma_e, size=N)
8     d = np.zeros(N, dtype=float)
9     for n in range(1, N):
10         d[n] = a * d[n-1] + e[n]
11     return d
12
13 def wiener_m2_from_correlations(rxx0, rxx1, rxd0, rxd1):
14     """Solve M=2 Wiener-Hopf system for w=[w0,w1]."""
15     R = np.array([[rxx0, rxx1],
16                  [rxx1, rxx0]], dtype=float)
17     p = np.array([rxd0, rxd1], dtype=float)
18     w = np.linalg.solve(R, p)
19     return w # [w0, w1]
20
21 def estimate_corrs_supervised(x, d):
22     """Estimate rxx[0], rxx[1], rxd[0], rxd[1] from paired data."""
23     N = len(x)
24     rxx0 = np.mean(x * x)
25     rxx1 = np.mean(x[1:] * x[:-1]) # lag 1
26     rxd0 = np.mean(x * d)
27     rxd1 = np.mean(x[1:] * d[:-1]) # E[x[n] d[n-1]]
28     return rxx0, rxx1, rxd0, rxd1
29
```

```

30 def apply_m2_filter(x, w0, w1):
31     """Apply causal 2-tap FIR: d_hat[n] = w0 x[n] + w1 x[n-1]."""
32     d_hat = np.zeros_like(x, dtype=float)
33     d_hat[0] = w0 * x[0] # assume x[-1]=0
34     d_hat[1:] = w0 * x[1:] + w1 * x[:-1]
35     return d_hat
36
37 # -----
38 # 1) Generate synthetic data
39 # -----
40 N = 800
41 a = 0.9
42 sigma_e = 0.5
43 sigma_v = 0.8
44
45 d = make_ar1(N, a=a, sigma_e=sigma_e, seed=1)
46 rng = np.random.default_rng(2)
47 v = rng.normal(0.0, sigma_v, size=N)
48 x = d + v # denoising setting
49
50 # Split into training/test
51 N_train = N // 2
52 x_tr, d_tr = x[:N_train], d[:N_train]
53 x_te, d_te = x[N_train:], d[N_train:]
54
55 # -----
56 # 2) Supervised M=2 Wiener filter (training)
57 # -----
58 rxx0_hat, rxx1_hat, rxd0_hat, rxd1_hat = estimate_corrs_supervised(x_tr
59     , d_tr)
59 w_sup = wiener_m2_from_correlations(rxx0_hat, rxx1_hat, rxd0_hat,
60     rxd1_hat)
60 w0_sup, w1_sup = w_sup
61
62 # -----
63 # 3) Model-based M=2 Wiener filter (AR(1))
64 # -----
65 # For AR(1): r_dd[0] = sigma_e^2 / (1-a^2), r_dd[1] = a r_dd[0]
66 rdd0 = (sigma_e**2) / (1.0 - a**2)
67 rdd1 = a * rdd0
68
69 # For x=d+v with white noise var sigma_v^2:
70 rxx0_model = rdd0 + sigma_v**2
71 rxx1_model = rdd1
72 rxd0_model = rdd0
73 rxd1_model = rdd1
74
75 w_mod = wiener_m2_from_correlations(rxx0_model, rxx1_model, rxd0_model,
76     rxd1_model)
76 w0_mod, w1_mod = w_mod
77
78 # -----
79 # 4) Apply filters on test
80 # -----
81 dhat_sup = apply_m2_filter(x_te, w0_sup, w1_sup)
82 dhat_mod = apply_m2_filter(x_te, w0_mod, w1_mod)
83
84 # -----

```

```

85 # 5) Evaluate and plot
86 # -----
87 mse_noisy = np.mean((x_te - d_te)**2)
88 mse_sup = np.mean((dhat_sup - d_te)**2)
89 mse_mod = np.mean((dhat_mod - d_te)**2)
90
91 print("Supervised Wiener (M=2) taps: w0 = %.4f, w1 = %.4f" % (w0_sup,
92     w1_sup))
93 print("Model-based Wiener (M=2) taps: w0 = %.4f, w1 = %.4f" % (w0_mod,
94     w1_mod))
95 print("Test MSE: noisy x vs d: %.4f" % mse_noisy)
96 print("Test MSE: supervised estimate: %.4f" % mse_sup)
97 print("Test MSE: model-based estimate: %.4f" % mse_mod)
98
99 # Plot a short window for readability
100 L = min(250, N - N_train)
101 t = np.arange(L)
102
103 plt.figure()
104 plt.plot(t, d_te[:L], label="true d[n]")
105 plt.plot(t, x_te[:L], label="observed x[n]")
106 plt.plot(t, dhat_sup[:L], label="Wiener M=2 (supervised)")
107 plt.plot(t, dhat_mod[:L], label="Wiener M=2 (model-based)")
108 plt.xlabel("n (test segment)")
109 plt.ylabel("amplitude")
110 plt.title("2-tap FIR Wiener filtering (supervised vs model-based)")
111 plt.legend()
112 plt.grid(True)
113 plt.show()

```

Listing 1: Υλοποίηση φίλτρου Wiener

Η προσομοίωση αυτή υλοποιεί ένα πλήρες, end-to-end παράδειγμα φιλτραρίσματος Wiener για ένα βαθμωτό (1D) σήμα, χρησιμοποιώντας έναν αιτιατό FIR εκτιμητή δύο συντελεστών ($M = 2$). Παρακάτω περιγράφουμε κάθε βήμα του πειράματος με μαθηματικούς όρους.

Μοντέλο σήματος και θορύβου. Το καθαρό σήμα $\{d[n]\}$ παράγεται ως στάσιμη αυτοπαλίνδρομη διεργασία πρώτης τάξης:

$$d[n] = a d[n-1] + e[n], \quad e[n] \sim \mathcal{N}(0, \sigma_e^2), \quad |a| < 1. \quad (2.51)$$

Αυτό ορίζει μια διεργασία μηδενικού μέσου, στάσιμη κατά τη ευρεία έννοια, με αυτοσυσχέτιση

$$r_{dd}[0] = \frac{\sigma_e^2}{1 - a^2}, \quad r_{dd}[1] = a r_{dd}[0]. \quad (2.52)$$

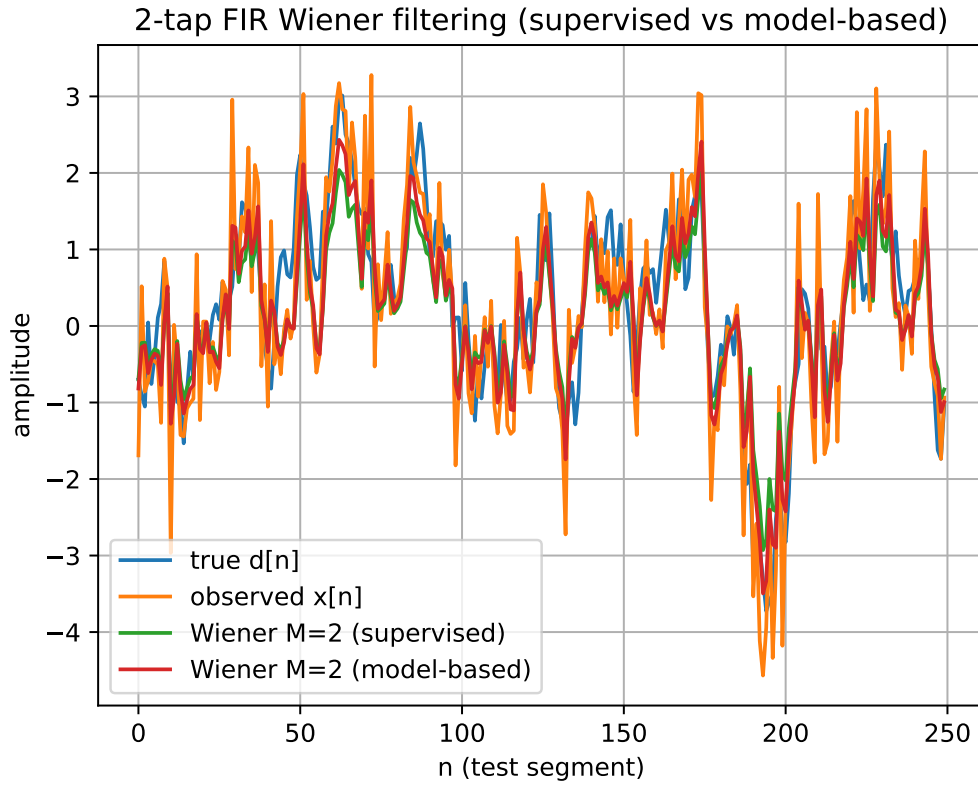
Το παρατηρούμενο σήμα προκύπτει μέσω προσθετικού θορύβου:

$$x[n] = d[n] + v[n], \quad v[n] \sim \mathcal{N}(0, \sigma_v^2), \quad v \perp d. \quad (2.53)$$

Στόχος είναι η εκτίμηση του $d[n]$ από το $x[n]$ με χρήση ενός αιτιατού FIR φίλτρου Wiener.

Δομή εκτιμητή ($M = 2$). Ο εκτιμητής περιορίζεται στην κλάση των αιτιατών FIR φίλτρων δύο συντελεστών:

$$\hat{d}[n] = w_0 x[n] + w_1 x[n-1], \quad (2.54)$$



Σχήμα 2.1: **Δί-συντελεστικό FIR φιλτράρισμα Wiener ($M = 2$) σε 1Δ.** Παρουσιάζουμε το καθαρό σήμα $d[n]$, την θορυβώδη παρατήρηση $x[n] = d[n] + v[n]$, και δύο εκτιμήσεις Wiener $\hat{d}[n] = w_0 x[n] + w_1 x[n-1]$. Το επιβλεπόμενο φίλτρο μαθαίνει τους (w_0, w_1) από ζευγοποιημένα δεδομένα εκπαίδευσης $(x[n], d[n])$ μέσω εμπειρικών συσχετίσεων, ενώ το μοντελο-βασισμένο φίλτρο υπολογίζει τους (w_0, w_1) από υποτιθέμενα στατιστικά δεύτερης τάξης (π.χ. prior AR(1) για το $d[n]$ και γνωστή διασπορά θορύβου).

ή, ισοδύναμα,

$$\hat{d}[n] = \mathbf{w}^T \mathbf{x}_n, \quad \mathbf{x}_n = \begin{bmatrix} x[n] \\ x[n-1] \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}.$$

Κριτήριο βελτιστότητας. Οι συντελεστές (w_0, w_1) επιλέγονται ώστε να ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα

$$\mathbb{E}[(d[n] - \hat{d}[n])^2].$$

Για $M = 2$, οι εξισώσεις Wiener-Hopf ανάγονται σε

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[1] \\ r_{xx}[1] & r_{xx}[0] \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} r_{xd}[0] \\ r_{xd}[1] \end{bmatrix}, \quad (2.55)$$

όπου

$$r_{xx}[k] = \mathbb{E}\{x[n]x[n-k]\}, \quad r_{xd}[k] = \mathbb{E}\{x[n]d[n-k]\}.$$

Επιβλεπόμενη εκτίμηση συσχετίσεων. Στο επιβλεπόμενο μέρος του πειράματος, τα ζευγοποιημένα δείγματα $\{x[n], d[n]\}$ είναι διαθέσιμα σε ένα χρονικό διάστημα εκπαίδευσης.

Οι αναμενόμενες τιμές στην (2.55) προσεγγίζονται με μέσους όρους δείγματος:

$$\hat{r}_{xx}[0] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]^2, \quad (2.56)$$

$$\hat{r}_{xx}[1] = \frac{1}{N-1} \sum_{n=1}^{N-1} x[n]x[n-1], \quad (2.57)$$

$$\hat{r}_{xd}[0] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]d[n], \quad (2.58)$$

$$\hat{r}_{xd}[1] = \frac{1}{N-1} \sum_{n=1}^{N-1} x[n]d[n-1]. \quad (2.59)$$

Αυτές οι εμπειρικές συσχετίσεις εισάγονται στην (2.55), παράγοντας μια εκτίμηση από δεδομένα $\hat{\mathbf{w}}_{\text{sup}}$. Αυτό αντιστοιχεί στην επίλυση ενός προβλήματος ελαχίστων τετραγώνων για την επιλεγμένη δομή FIR.

Υπολογισμός συσχετίσεων με μοντέλο. Σε αυτό το μέρος του πειράματος, το καθαρό σήμα $d[n]$ δεν χρησιμοποιείται άμεσα. Αντίθετα, το πιθανοτικό μοντέλο (2.51)–(2.53) χρησιμοποιείται για τον αναλυτικό υπολογισμό των συσχετίσεων. Χρησιμοποιώντας $x = d + v$ με $v \perp d$, παίρνουμε

$$r_{xx}[0] = r_{dd}[0] + \sigma_v^2, \quad r_{xx}[1] = r_{dd}[1], \quad (2.60)$$

και

$$r_{xd}[0] = r_{dd}[0], \quad r_{xd}[1] = r_{dd}[1]. \quad (2.61)$$

Αντικαθιστώντας τις αυτοσυσχετίσεις AR(1) από την (2.52), προκύπτει ένα σύστημα Wiener–Hopf κλειστής μορφής, του οποίου η λύση $\mathbf{w}_{\text{model}}$ είναι ο βέλτιστος LMMSE εκτιμητής δύο συντελεστών υπό τα υποτιθέμενα στατιστικά.

Φιλτράρισμα και αξιολόγηση. Και τα δύο φίλτρα εφαρμόζονται σε ένα ανεξάρτητο τμήμα ελέγχου (test segment) μέσω της (2.54). Η απόδοση ποσοτικοποιείται με το εμπειρικό μέσο τετραγωνικό σφάλμα

$$\text{MSE} = \frac{1}{N_{\text{test}}} \sum_n (d[n] - \hat{d}[n])^2,$$

και συγκρίνεται με το βασικό MSE της θορυβώδους παρατήρησης $x[n]$.

Κύρια ερμηνεία. Το επιβλεπόμενο φιλτράρισμα Wiener προσεγγίζει τη βέλτιστη λύση μέσω πεπερασμένου πλήθους δειγμάτων για την εκτίμηση των στατιστικών δεύτερης τάξης, και συνεπώς υπόκειται σε διακύμανση εκτίμησης. Το φιλτράρισμα Wiener βασισμένο σε μοντέλο χρησιμοποιεί ακριβή (oracle) στατιστικά δεύτερης τάξης που προκύπτουν από το γενετικό μοντέλο, και συνεπώς βρίσκεται πιο κοντά στην πραγματική λύση LMMSE όταν το μοντέλο είναι σωστό. Καθώς αυξάνεται το πλήθος των δεδομένων εκπαίδευσης, οι δύο περιπτώσεις . Αντίθετα, υπό ασυμφωνία μοντέλου (model mismatch), η επιβλεπόμενη προσέγγιση μπορεί να υπερέχει.

3 Φιλτράρισμα Kalman

3.1 Κίνητρο: Γιατί Φιλτράρισμα Kalman;

Στην προηγούμενη ενότητα μελετήσαμε το *φιλτράρισμα Wiener*, το οποίο παρέχει έναν βέλτιστο γραμμικό εκτιμητή (LMMSE/MMSE υπό Γκαουσιανότητα) για *στάσιμες* καταστάσεις. Σε πολλές εφαρμογές, όμως, το μέγεθος που θέλουμε να εκτιμήσουμε είναι *εγγενώς δυναμικό*: εξελίσσεται στον χρόνο και οι μετρήσεις καταφθάνουν διαδοχικά. Τυπικά παραδείγματα είναι η θέση και η ταχύτητα στη ναυτιλία/πλοήγηση, αργά μεταβαλλόμενα κέρδη καναλιού στις επικοινωνίες, ή η θερμοκρασία σε ένα θερμικό σύστημα.

Αυτό οδηγεί στην ανάγκη για έναν εκτιμητή με τρεις ιδιότητες. Πρώτον, να είναι **αναδρομικός** (online): αντί να επανυπολογίζει μια εκτίμηση από την αρχή όταν φθάνουν νέα δεδομένα, να ενημερώνει αποδοτικά την τρέχουσα εκτίμηση χρησιμοποιώντας τη νέα μέτρηση. Δεύτερον, να είναι **βασισμένος σε μοντέλο**: να εκμεταλλεύεται τη γνώση για το πώς εξελίσσεται η κατάσταση στον χρόνο μέσω ενός δυναμικού μοντέλου. Τρίτον, να είναι **ενήμερος ως προς την αβεβαιότητα**: να παρακολουθεί όχι μόνο μια σημειακή εκτίμηση αλλά και την εμπιστοσύνη της, συνήθως μέσω ενός πίνακα συνδιακύμανσης που ποσοτικοποιεί το σφάλμα εκτίμησης.

Το φίλτρο Kalman καλύπτει ακριβώς αυτή την ανάγκη. Για γραμμικά μοντέλα (στο χώρο κατάστασης) με Γκαουσιανό θόρυβο, δίνει τον **βέλτιστο γραμμικό MMSE** εκτιμητή και, υπό Γκαουσιανές υποθέσεις, συμπίπτει με τον ακριβή Bayesian posterior μέσο. Η εννοιολογική του δομή είναι απλή και ισχυρή: *προβλέπουμε* την κατάσταση προς τα εμπρός χρησιμοποιώντας το μοντέλο και έπειτα *διορθώνουμε* αυτή την πρόβλεψη χρησιμοποιώντας τη νέα μέτρηση.

3.2 Μοντελοποίηση στον Χώρο Κατάστασης

Πολλά δυναμικά συστήματα μπορούν να περιγραφούν μέσω μιας **λανθάνουσας (κρυφής) κατάστασης** που εξελίσσεται στον χρόνο και παράγει θορυβώδεις παρατηρήσεις. Σε μια γενική (συνήθως μη γραμμική) μορφή, αυτό γράφεται ως

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, k-1) + \mathbf{w}_{k-1}, \quad (3.1)$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, k) + \mathbf{v}_k. \quad (3.2)$$

Όπου,

- $\mathbf{x}_k \in \mathbb{R}^n$ είναι η λανθάνουσα/κρυφή κατάσταση (το μέγεθος προς εκτίμηση),
- $\mathbf{y}_k \in \mathbb{R}^m$ είναι η μέτρηση (παρατηρούμενα δεδομένα),
- $\mathbf{u}_k \in \mathbb{R}^r$ είναι μια γνωστή είσοδος/έλεγχος (μπορεί να απουσιάζει),
- $\mathbf{w}_{k-1}$ είναι ο θόρυβος διεργασίας που αποτυπώνει αβεβαιότητα μοντέλου και μη μοντελοποιημένες επιδράσεις,
- $\mathbf{v}_k$ είναι ο θόρυβος μέτρησης που αποτυπώνει αβεβαιότητα αισθητήρα.

Ο στόχος του Bayesian φιλτραρίσματος είναι:

$$\text{να συναχθεί το } \mathbf{x}_k \text{ από τα } \mathbf{y}_{1:k} \triangleq \{\mathbf{y}_1, \dots, \mathbf{y}_k\}.$$

Ισοδύναμα, ζητείται η εκ των υστέρων κατανομή $p(\mathbf{x}_k | \mathbf{y}_{1:k})$, και συνήθως μια σημειακή εκτίμηση.

3.3 Φίλτρο Kalman: Γραμμικές–Γκαουσιανές Υποθέσεις

Το φίλτρο Kalman είναι βέλτιστο στην γραμμική–Γκαουσιανή ειδική περίπτωση των (3.1)–(3.2). Η κατάσταση εξελίσσεται γραμμικά:

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} + \mathbf{w}_{k-1}, \quad (3.3)$$

και οι μετρήσεις είναι γραμμικές συναρτήσεις της τρέχουσας κατάστασης:

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k. \quad (3.4)$$

Οι όροι θορύβου υποτίθεται ότι είναι μηδενικού μέσου, λευκοί και Γκαουσιανοί:

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k), \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k), \quad (3.5)$$

και αμοιβαία ανεξάρτητοι στον χρόνο και μεταξύ διεργασιών, δηλαδή

$$\mathbb{E}[\mathbf{w}_k \mathbf{v}_j^T] = \mathbf{0}, \quad \forall k, j, \quad (\text{και αντίστοιχα για διαφορετικούς χρονικούς δείκτες}). \quad (3.6)$$

Τέλος, η αρχική συνθήκη είναι Γκαουσιανή:

$$\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0). \quad (3.7)$$

Κύρια συνέπεια (κλειστότητα των Γκαουσιανών υπό γραμμικούς μετασχηματισμούς). Υπό τις (3.3)–(3.7), η προβλεπτική κατανομή $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$ είναι Γκαουσιανή εφόσον $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$ είναι Γκαουσιανή. Έπειτα, η ενσωμάτωση του γραμμικού Γκαουσιανού μοντέλου μέτρησης (3.4) διατηρεί τη Γκαουσιανότητα της εκ των υστέρων κατανομής (posterior). Άρα η κατανομή φιλτραρίσματος παραμένει Γκαουσιανή για όλα τα k :

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}),$$

και το φίλτρο Kalman δίνει αναδρομικές εξισώσεις ενημέρωσης για τον posterior μέσο $\hat{\mathbf{x}}_{k|k}$ και τη συνδιακύμανση $\mathbf{P}_{k|k}$.

3.4 Φιλτράρισμα Kalman ως Αναδρομική Bayesian Εκτίμηση

Στόχος του φιλτραρίσματος Kalman είναι ο υπολογισμός της posterior κατανομής της κρυφής κατάστασης στο χρονικό βήμα k , δεδομένων όλων των μετρήσεων μέχρι εκείνη τη στιγμή:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}).$$

Αντί να επανυπολογίζουμε αυτή την posterior από την αρχή σε κάθε χρονικό βήμα, εκμεταλλευόμαστε τη χρονική δομή του προβλήματος ώστε να προκύψει μια αναδρομική Bayesian ενημέρωση.

3.4.1 Αναδρομή Bayesian Φιλτραρίσματος

Το Bayesian φιλτράρισμα αποσυνθέτει τη συμπερασματολογία σε δύο εννοιολογικά διακριτά βήματα: την πρόβλεψη και τη διόρθωση.

Πρόβλεψη. Δεδομένης της posterior στο χρόνο $k-1$, προωθούμε την αβεβαιότητα μέσω της δυναμικής:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}. \quad (3.8)$$

Διόρθωση. Μόλις είναι διαθέσιμη μια νέα μέτρηση $\mathbf{y}_k$, την ενσωματώνουμε με τον κανόνα του Bayes:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}). \quad (3.9)$$

Το βήμα πρόβλεψης προωθεί την προηγούμενη πεποίθησή μας (previous posterior belief) προς τα εμπρός στον χρόνο, ενώ το 2ο βήμα διορθώνει αυτή την πρόβλεψη χρησιμοποιώντας την νέα μέτρηση/πληροφορία. Για γραμμική δυναμική και Γκαουσιανό θόρυβο, και τα δύο βήματα διατηρούν τη Γκαουσιανότητα, επιτρέποντας την πλήρη περιγραφή της posterior μόνο μέσω ενός μέσου και μιας συνδιακύμανσης.

3.4.2 Φίλτρο Kalman: Βήμα Πρόβλεψης

Υποθέτουμε ότι στο χρόνο $k - 1$ η posterior είναι Γκαουσιανή:

$$p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_{k-1}, \Sigma_{k-1}).$$

Το γραμμικό μοντέλο στο χώρο κατάστασης είναι

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} + \mathbf{w}_{k-1}, \quad \mathbf{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}), \quad (3.10)$$

όπου η συνδιακύμανση του θορύβου διεργασίας $\mathbf{Q}_{k-1}$ μπορεί να μεταβάλλεται με τον χρόνο.

Εφαρμόζοντας την (3.8) προκύπτει η προβλεπόμενη (prior) κατανομή:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \mathcal{N}(\hat{\mathbf{x}}_k^-, \Sigma_k^-).$$

Ο προβλεπόμενος μέσος είναι

$$\boxed{\hat{\mathbf{x}}_k^- = \mathbb{E}[\mathbf{x}_k | \mathbf{y}_{1:k-1}] = \mathbf{A}\hat{\mathbf{x}}_{k-1} + \mathbf{B}\mathbf{u}_{k-1}}, \quad (3.11)$$

και η προβλεπόμενη συνδιακύμανση είναι

$$\boxed{\Sigma_k^- = \text{Cov}[\mathbf{x}_k | \mathbf{y}_{1:k-1}] = \mathbf{A}\Sigma_{k-1}\mathbf{A}^\top + \mathbf{Q}_{k-1}}. \quad (3.12)$$

Η συνδιακύμανση αυξάνεται λόγω του θορύβου διεργασίας, αντανακλώντας την αυξανόμενη αβεβαιότητα όταν το σύστημα εξελίσσεται χωρίς ενημερώσεις από μετρήσεις.

3.4.3 Φίλτρο Kalman: Βήμα Διόρθωσης

Μετά το βήμα πρόβλεψης, ενσωματώνουμε τη νέα μέτρηση χρησιμοποιώντας το γραμμικό μοντέλο στο χώρο κατάστασης για τις παρατηρήσεις:

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k), \quad (3.13)$$

όπου $\mathbf{R}_k$ είναι η (ενδεχομένως χρονικά μεταβαλλόμενη) συνδιακύμανση του θορύβου μέτρησης.

Ορίζουμε την **καινοτομία** (υπόλοιπο μέτρησης)

$$\boxed{\mathbf{r}_k = \mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k^-}, \quad (3.14)$$

και τη συνδιακύμανσή της

$$\boxed{\mathbf{S}_k = \mathbf{C}\Sigma_k^-\mathbf{C}^\top + \mathbf{R}_k}. \quad (3.15)$$

Το κέρδος Kalman ορίζεται ως

$$\boxed{\mathbf{K}_k = \Sigma_k^- \mathbf{C}^\top \mathbf{S}_k^{-1}}. \quad (3.16)$$

Ο posterior μέσος και η posterior συνδιακύμανση δίνονται από

$$\boxed{\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \mathbf{r}_k}, \quad (3.17)$$

$$\boxed{\Sigma_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}) \Sigma_k^-}. \quad (3.18)$$

Το κέρδος Kalman εξισορροπεί την εμπιστοσύνη μεταξύ της πρόβλεψης του μοντέλου (κωδικοποιημένης στη Σ_k^-) και της μέτρησης (κωδικοποιημένης στη $\mathbf{R}_k$).

3.5 Bayesian Φιλτράρισμα: Το Γενικό Πλαίσιο

Το Bayesian φιλτράρισμα αποτελεί το γενικό μαθηματικό πλαίσιο για αναδρομική εκτίμηση κατάστασης σε δυναμικά συστήματα. Η βασική ιδέα είναι η διατήρηση μιας πιθανοτικής κατανομής πάνω στην τρέχουσα κατάσταση (η πεποίθηση, belief) και η διαδοχική ενημέρωσή της καθώς καταφθάνουν νέες είσοδοι ελέγχου και νέες μετρήσεις.

3.5.1 Κατάσταση, Έλεγχοι, Μετρήσεις και Πεποίθηση

Εστω

$$\mathbf{x}_k \in \mathbb{R}^n \quad (\text{κατάσταση}), \quad \mathbf{u}_k \in \mathbb{R}^r \quad (\text{έλεγχος/είσοδος}), \quad \mathbf{y}_k \in \mathbb{R}^m \quad (\text{μέτρηση}).$$

Ορίζουμε το ιστορικό μετρήσεων ως $\mathbf{y}_{1:k} = \{\mathbf{y}_1, \dots, \mathbf{y}_k\}$ και το ιστορικό ελέγχων ως $\mathbf{u}_{1:k} = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$.

Η πεποίθηση (belief) στη χρονική στιγμή k ορίζεται ως η κατανομή φιλτραρίσματος

$$\text{bel}(\mathbf{x}_k) \triangleq p(\mathbf{x}_k \mid \mathbf{y}_{1:k}, \mathbf{u}_{1:k}), \quad (3.19)$$

δηλαδή η posterior κατανομή πάνω στην τρέχουσα κατάσταση, δεδομένης όλης της διαθέσιμης πληροφορίας μέχρι τον χρόνο k .

3.5.2 Πιθανοτικές Υποθέσεις (Μοντέλο στο Χώρο Κατάστασης)

Το Bayesian φιλτράρισμα βασίζεται σε δύο βασικές υποθέσεις συνθήκης ανεξαρτησίας.

1) Δυναμική Markov πρώτης τάξης (μοντέλο κίνησης). Η κατάσταση στον χρόνο k εξαρτάται από το παρελθόν μόνο μέσω της προηγούμενης κατάστασης και του τρέχοντος ελέγχου:

$$p(\mathbf{x}_k \mid \mathbf{x}_{0:k-1}, \mathbf{u}_{1:k}) = p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k). \quad (3.20)$$

Η υπό συνθήκη πυκνότητα $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k)$ ονομάζεται **μοντέλο κίνησης** (ή μοντέλο μετάβασης).

2) Υπό συνθήκη ανεξαρτησία των μετρήσεων (μοντέλο παρατήρησης). Η μέτρηση στον χρόνο k εξαρτάται από το παρελθόν μόνο μέσω της τρέχουσας κατάστασης:

$$p(\mathbf{y}_k \mid \mathbf{x}_{0:k}, \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) = p(\mathbf{y}_k \mid \mathbf{x}_k). \quad (3.21)$$

Η υπό συνθήκη πυκνότητα $p(\mathbf{y}_k \mid \mathbf{x}_k)$ ονομάζεται **μοντέλο παρατήρησης** (ή likelihood της μέτρησης).

Οι παραπάνω υποθέσεις ορίζουν ένα κρυφό μοντέλο Markov (Hidden Markov Model, HMM) με εισόδους.

3.5.3 Αναδρομή του Bayes Φίλτρου

Ξεκινώντας από τον ορισμό της πεποίθησης (3.19), εφαρμόζουμε τον κανόνα του Bayes:

$$\begin{aligned} \text{bel}(\mathbf{x}_k) &= p(\mathbf{x}_k \mid \mathbf{y}_{1:k}, \mathbf{u}_{1:k}) \\ &= \frac{p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k})}{p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k})}. \end{aligned} \quad (3.22)$$

Εισάγουμε τη σταθερά κανονικοποίησης

$$\eta \triangleq \frac{1}{p(\mathbf{y}_k \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k})}, \quad (3.23)$$

ώστε

$$\text{bel}(\mathbf{x}_k) = \eta p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}). \quad (3.24)$$

Χρησιμοποιώντας την υπόθεση Markov για το μοντέλο παρατήρησης (3.21),

$$p(\mathbf{y}_k \mid \mathbf{x}_k, \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) = p(\mathbf{y}_k \mid \mathbf{x}_k),$$

παίρνουμε

$$\text{bel}(\mathbf{x}_k) = \eta p(\mathbf{y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}). \quad (3.25)$$

3.5.4 Πρόβλεψη μέσω του Νόμου Ολικής Πιθανότητας

Η προβλεπτική κατανομή $p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k})$ προκύπτει με περιθωριοποίηση ως προς την προηγούμενη κατάσταση:

$$\begin{aligned} p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) &= \int p(\mathbf{x}_k, \mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) d\mathbf{x}_{k-1} \\ &= \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) d\mathbf{x}_{k-1}. \end{aligned} \quad (3.26)$$

Χρησιμοποιώντας την υπόθεση Markov για το μοντέλο κίνησης (3.20) και παρατηρώντας ότι

$$p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) = p(\mathbf{x}_{k-1} \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k-1}) = \text{bel}(\mathbf{x}_{k-1}),$$

καταλήγουμε στο

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k-1}, \mathbf{u}_{1:k}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) \text{bel}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1}. \quad (3.27)$$

Αντικαθιστώντας την (3.27) στην (3.25), προκύπτει η αναδρομή του φίλτρου Bayes:

$$\text{bel}(\mathbf{x}_k) = \eta p(\mathbf{y}_k \mid \mathbf{x}_k) \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{u}_k) \text{bel}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1}. \quad (3.28)$$

3.5.5 Bayesian Φιλτράρισμα σε Δύο Βήματα

Η αναδρομή (3.28) αποσυντίθεται φυσικά σε δύο βήματα.

Πρόβλεψη.

$$\overline{\text{bel}}(\mathbf{x}_k) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_k) \text{bel}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1}. \quad (3.29)$$

Διόρθωση.

$$\text{bel}(\mathbf{x}_k) = \eta p(\mathbf{y}_k | \mathbf{x}_k) \overline{\text{bel}}(\mathbf{x}_k), \quad (3.30)$$

όπου η η επιλέγεται ώστε $\int \text{bel}(\mathbf{x}_k) d\mathbf{x}_k = 1$.

3.5.6 Φιλτράρισμα Kalman ως Ειδική Περίπτωση

Το Bayesian φιλτράρισμα είναι ακριβές αλλά γενικά μη υπολογίσιμο στην πράξη, επειδή η πεποίθηση $\text{bel}(\mathbf{x}_k)$ είναι μια αυθαίρετη κατανομή στο $\mathbb{R}^n$. Το φιλτράρισμα Kalman προκύπτει όταν:

- το μοντέλο κίνησης είναι γραμμικό και Γκαουσιανό:

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \mathbf{w}_k, \quad \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k),$$

- το μοντέλο παρατήρησης είναι γραμμικό και Γκαουσιανό:

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k).$$

Υπό αυτές τις υποθέσεις, η πεποίθηση παραμένει Γκαουσιανή για όλα τα k και η αναδρομή του φίλτρου Bayes (3.29)–(3.30) ανάγεται στις εξισώσεις ενημέρωσης του φίλτρου Kalman για τον μέσο και τη συνδιακύμανση.

3.6 Το Φίλτρο Kalman ως Αναδρομικό Φίλτρο Wiener

Ένας χρήσιμος τρόπος ερμηνείας του φίλτρου Kalman είναι ως μια *χρονικά μεταβαλλόμενη* (αναδρομική) εκδοχή της εκτίμησης Wiener / LMMSE. Στο φιλτράρισμα Wiener κατασκευάζουμε έναν βέλτιστο γραμμικό εκτιμητή με βάση δευτεροβάθμια στατιστικά μεγέθη. Στο φιλτράρισμα Kalman, επιλύουμε ένα στενά συγγενές πρόβλημα LMMSE σε *κάθε χρονική στιγμή*, χρησιμοποιώντας τη *τρέχουσα* προβλεπόμενη συνδιακύμανση, γεγονός που οδηγεί σε ένα κέρδος που μεταβάλλεται με τον χρόνο.

3.6.1 Κάθε χρονική στιγμή k λύνουμε ένα πρόβλημα LMMSE

Θεωρούμε το γραμμικό–Γκαουσιανό μοντέλο μέτρησης στη χρονική στιγμή k :

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k). \quad (3.31)$$

Υποθέτουμε ότι διαθέτουμε ήδη μια *πρότερη* (prior) κατανομή:

$$\mathbf{x}_k | \mathbf{y}_{1:k-1} \sim \mathcal{N}(\hat{\mathbf{x}}_k^-, \Sigma_k^-). \quad (3.32)$$

Ορίζουμε το σφάλμα εκτίμησης και περιοριζόμαστε σε εκτιμητές της αφινικής μορφής

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^-), \quad (3.33)$$

όπου $\mathbf{K}_k$ είναι ένας πίνακας προς προσδιορισμό. Αυτή είναι ακριβώς η μορφή ενημέρωσης του φίλτρου Kalman: ξεκινάμε από τον πρότερο μέσο και προσθέτουμε μια διόρθωση ανάλογη της καινοτομίας/σφάλματος.

Το βασικό σημείο είναι ότι η (3.33) προκύπτει ως λύση ενός προβλήματος LMMSE. Ορίζουμε

$$\tilde{\mathbf{x}}_k \triangleq \mathbf{x}_k - \hat{\mathbf{x}}_k^-, \quad \tilde{\mathbf{y}}_k \triangleq \mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^-.$$

Από τις (3.31)–(3.32) έχουμε

$$\tilde{\mathbf{y}}_k = \mathbf{C}_k \tilde{\mathbf{x}}_k + \mathbf{v}_k, \quad \mathbb{E}[\tilde{\mathbf{x}}_k] = \mathbf{0}, \quad \text{Cov}(\tilde{\mathbf{x}}_k) = \Sigma_k^-. \quad (3.34)$$

Περιοριζόμαστε τώρα σε γραμμικούς εκτιμητές της $\tilde{\mathbf{x}}_k$ από την $\tilde{\mathbf{y}}_k$:

$$\hat{\tilde{\mathbf{x}}}_k = \mathbf{K}_k \tilde{\mathbf{y}}_k.$$

Επιλέγουμε $\mathbf{K}_k$ ώστε να ελαχιστοποιεί το υπό συνθήκη MSE:

$$\mathbf{K}_k^* = \arg \min_{\mathbf{K}} \mathbb{E}[\|\tilde{\mathbf{x}}_k - \mathbf{K} \tilde{\mathbf{y}}_k\|_2^2 \mid \mathbf{y}_{1:k-1}]. \quad (3.35)$$

Αυτό είναι ένα τυπικό πρόβλημα Wiener / LMMSE με «επιθυμητό σήμα» την $\tilde{\mathbf{x}}_k$ και «παρατήρηση» την $\tilde{\mathbf{y}}_k$.

3.6.2 Λύση Wiener $\Rightarrow$ Κέρδος Kalman

Η λύση Wiener / LMMSE έχει τη μορφή

$$\boxed{\mathbf{K}_k^* = \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}} \mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}}^{-1}}, \quad (3.36)$$

όπου (η υπό συνθήκη ως προς $\mathbf{y}_{1:k-1}$ είναι έμμεση)

$$\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}} \triangleq \mathbb{E}\{\tilde{\mathbf{x}}_k \tilde{\mathbf{y}}_k^T\}, \quad \mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} \triangleq \mathbb{E}\{\tilde{\mathbf{y}}_k \tilde{\mathbf{y}}_k^T\}.$$

Χρησιμοποιώντας την (3.34) και την ανεξαρτησία $\tilde{\mathbf{x}}_k \perp \mathbf{v}_k$, προκύπτει

$$\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{y}}} = \Sigma_k^- \mathbf{C}_k^T,$$

και

$$\mathbf{R}_{\tilde{\mathbf{y}}\tilde{\mathbf{y}}} = \mathbf{C}_k \Sigma_k^- \mathbf{C}_k^T + \mathbf{R}_k.$$

Άρα

$$\boxed{\mathbf{K}_k = \Sigma_k^- \mathbf{C}_k^T (\mathbf{C}_k \Sigma_k^- \mathbf{C}_k^T + \mathbf{R}_k)^{-1}}, \quad (3.37)$$

που είναι ακριβώς το κέρδος Kalman.

3.6.3 Ερμηνεία: «Αναδρομικό» Φιλτράρισμα Wiener

Η (3.37) δείχνει ότι το κέρδος Kalman είναι ένα φίλτρο Wiener / LMMSE, υπολογισμένο *online* με χρονικά μεταβαλλόμενα στατιστικά:

- Στο φιλτράρισμα Wiener, το κέρδος εξαρτάται από (συνήθως σταθερά) δευτεροβάθμια στατιστικά όπως τα $\mathbf{R}_{xx}$ και $\mathbf{R}_{xd}$.
- Στο φιλτράρισμα Kalman, τα αντίστοιχα στατιστικά είναι υπό συνθήκη ως προς τα προηγούμενα δεδομένα και εξελίσσονται με τον χρόνο μέσω της πρόβλεψης, δηλαδή μέσω της Σ_k^- .

Τι το καθιστά αναδρομικό; Η βασική αναδρομή είναι ότι η πρότερη συνδιακύμανση Σ_k^- προκύπτει από τη διάδοση της προηγούμενης οπισθίας συνδιακύμανσης μέσω του δυναμικού μοντέλου:

$$\Sigma_k^- = A_k \Sigma_{k-1} A_k^T + Q_{k-1}.$$

Τι το καθιστά χρονικά μεταβαλλόμενο; Ακόμη και αν οι A_k, C_k είναι σταθερές, η υπό συνθήκη συνδιακύμανση Σ_k^- μεταβάλλεται με τον χρόνο (ιδίως στα μεταβατικά στάδια), ενώ και οι θόρυβοι Q_{k-1}, R_k μπορεί να είναι χρονικά μεταβαλλόμενοι. Συνεπώς, το κέρδος K_k είναι γενικά χρονικά μεταβαλλόμενο, σε αντίθεση με το κλασικό στάσιμο φίλτρο Wiener.

Συμπέρασμα. Το φιλτράρισμα Kalman μπορεί να ιδωθεί ως εφαρμογή ενός σχεδιασμού Wiener / LMMSE σε κάθε χρονική στιγμή:

«κέρδος Kalman» = «φίλτρο Wiener υπολογισμένο από τις τρέχουσες υπό συνθήκη συνδιακυμάνσεις».

Υπό γραμμικές-Γκαουσιανές υποθέσεις, αυτή η αναδρομική οπτική Wiener είναι πλήρως ισοδύναμη με την αναδρομική οπτική Bayesian.

3.7 Παραδείγματα Kalman Φίλτρου

Σε αυτή την ενότητα θα δούμε παραδείγματα φίλτρων Kalman και θα τα λύσουμε «με το χέρι». Τονίζουμε τα δύο βήματα σε κάθε χρόνο k :

$$\text{Πρόβλεψη: } (\hat{x}_{k-1}, \Sigma_{k-1}) \mapsto (\hat{x}_k^-, \Sigma_k^-), \quad \text{Διόρθωση: } (\hat{x}_k^-, \Sigma_k^-, y_k) \mapsto (\hat{x}_k, \Sigma_k).$$

Επιτρέπουμε χρονικά μεταβαλλόμενες συνδιακυμάνσεις θορύβου Q_{k-1} και R_k .

3.7.1 Παράδειγμα 1 (1Δ): Random Walk με Θορυβώδεις Μετρήσεις

Μοντέλο. Εκτιμούμε μια βαθμωτή κατάσταση $x_k \in \mathbb{R}$ από βαθμωτές μετρήσεις $y_k \in \mathbb{R}$ με

$$x_k = x_{k-1} + w_{k-1}, \quad w_{k-1} \sim \mathcal{N}(0, Q_{k-1}), \quad (3.38)$$

$$y_k = x_k + v_k, \quad v_k \sim \mathcal{N}(0, R_k). \quad (3.39)$$

Άρα $A = 1$, $C = 1$, και δεν υπάρχει είσοδος ελέγχου.

Εξισώσεις φίλτρου Kalman (βαθμωτή μορφή). Πρόβλεψη:

$$\hat{x}_k^- = \hat{x}_{k-1}, \quad \Sigma_k^- = \Sigma_{k-1} + Q_{k-1}. \quad (3.40)$$

Διόρθωση:

$$S_k = \Sigma_k^- + R_k, \quad K_k = \frac{\Sigma_k^-}{S_k}, \quad (3.41)$$

$$\hat{x}_k = \hat{x}_k^- + K_k (y_k - \hat{x}_k^-), \quad \Sigma_k = (1 - K_k) \Sigma_k^-. \quad (3.42)$$

Αριθμητικές τιμές. Υποθέτουμε prior

$$\hat{x}_0 = 0, \quad \Sigma_0 = 1.$$

Αφήνουμε τις διακυμάνσεις θορύβου να μεταβάλλονται με τον χρόνο:

$$Q_0 = 0.10, \quad R_1 = 0.40, \quad Q_1 = 0.20, \quad R_2 = 0.10.$$

Υποθέτουμε ότι λαμβάνουμε μετρήσεις

$$y_1 = 1.20, \quad y_2 = 0.90.$$

Βήμα $k = 1$. Πρόβλεψη:

$$\hat{x}_1^- = \hat{x}_0 = 0, \quad \Sigma_1^- = \Sigma_0 + Q_0 = 1 + 0.10 = 1.10.$$

Καινοτομία και κέρδος:

$$S_1 = \Sigma_1^- + R_1 = 1.10 + 0.40 = 1.50, \quad K_1 = \frac{\Sigma_1^-}{S_1} = \frac{1.10}{1.50} \approx 0.7333.$$

Διόρθωση:

$$\begin{aligned} \hat{x}_1 &= \hat{x}_1^- + K_1(y_1 - \hat{x}_1^-) = 0 + 0.7333(1.20 - 0) \approx 0.8800, \\ \Sigma_1 &= (1 - K_1)\Sigma_1^- = (1 - 0.7333) \cdot 1.10 \approx 0.2933. \end{aligned}$$

Βήμα $k = 2$. Πρόβλεψη:

$$\hat{x}_2^- = \hat{x}_1 \approx 0.8800, \quad \Sigma_2^- = \Sigma_1 + Q_1 \approx 0.2933 + 0.20 = 0.4933.$$

Καινοτομία και κέρδος:

$$S_2 = \Sigma_2^- + R_2 \approx 0.4933 + 0.10 = 0.5933, \quad K_2 = \frac{\Sigma_2^-}{S_2} \approx \frac{0.4933}{0.5933} \approx 0.8316.$$

Διόρθωση:

$$\begin{aligned} \hat{x}_2 &= \hat{x}_2^- + K_2(y_2 - \hat{x}_2^-) \approx 0.8800 + 0.8316(0.90 - 0.8800) \approx 0.8966, \\ \Sigma_2 &= (1 - K_2)\Sigma_2^- \approx (1 - 0.8316) \cdot 0.4933 \approx 0.0831. \end{aligned}$$

Ερμηνεία. Στο $k = 2$, ο αισθητήρας είναι πιο αξιόπιστος (το R_2 είναι μικρότερο), οπότε το K_2 αυξάνεται και το φίλτρο εμπιστεύεται περισσότερο το y_2 . Η posterior αβεβαιότητα Σ_2 μειώνεται αντίστοιχα.

3.7.2 Παράδειγμα 2 (2Δ): Μοντέλο Σταθερής Ταχύτητας με Μετρήσεις Θέσης

Μοντέλο. Τώρα η κατάσταση είναι 2Δ:

$$\mathbf{x}_k = \begin{bmatrix} p_k \\ v_k \end{bmatrix},$$

όπου p_k είναι η θέση και v_k η ταχύτητα. Για περίοδο δειγματοληψίας $\Delta t = 1$, η δυναμική σταθερής ταχύτητας είναι

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \quad \mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}). \quad (3.43)$$

Υποθέτουμε ότι μετράμε μόνο τη θέση:

$$y_k = \mathbf{C}\mathbf{x}_k + v_k^{(m)}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad v_k^{(m)} \sim \mathcal{N}(0, R_k). \quad (3.44)$$

Εξισώσεις Kalman. Πρόβλεψη:

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1}, \quad \Sigma_k^- = \mathbf{A}\Sigma_{k-1}\mathbf{A}^T + \mathbf{Q}_{k-1}.$$

Διόρθωση:

$$r_k = y_k - \mathbf{C}\hat{\mathbf{x}}_k^-, \quad S_k = \mathbf{C}\Sigma_k^-\mathbf{C}^T + R_k, \quad \mathbf{K}_k = \Sigma_k^-\mathbf{C}^T S_k^{-1},$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k r_k, \quad \Sigma_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C})\Sigma_k^-.$$

Αριθμητικές τιμές (για $k = 1$). Υποθέτουμε αρχική posterior:

$$\hat{\mathbf{x}}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \Sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Επιλέγουμε χρονικά μεταβαλλόμενο θόρυβο διεργασίας (εδώ, στο $k = 0$):

$$\mathbf{Q}_0 = \begin{bmatrix} 0.10 & 0 \\ 0 & 0.20 \end{bmatrix}, \quad R_1 = 0.50.$$

Υποθέτουμε ότι η πρώτη μέτρηση είναι $y_1 = 0.70$.

Πρόβλεψη μέχρι $k = 1$.

$$\hat{\mathbf{x}}_1^- = \mathbf{A}\hat{\mathbf{x}}_0 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Για τη συνδιακύμανση, πρώτα υπολογίζουμε

$$\mathbf{A}\Sigma_0\mathbf{A}^T = \mathbf{A}\mathbf{A}^T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Έπειτα προσθέτουμε $\mathbf{Q}_0$:

$$\Sigma_1^- = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0.10 & 0 \\ 0 & 0.20 \end{bmatrix} = \begin{bmatrix} 2.10 & 1 \\ 1 & 1.20 \end{bmatrix}.$$

Διόρθωση στο $k = 1$. Καινοτομία:

$$r_1 = y_1 - \mathbf{C}\hat{\mathbf{x}}_1^- = 0.70 - [1 \ 0] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0.70 - 1 = -0.30.$$

Συνδιακύμανση καινοτομίας (βαθμωτή):

$$S_1 = \mathbf{C}\Sigma_1^-\mathbf{C}^T + R_1 = \Sigma_{1,11}^- + R_1 = 2.10 + 0.50 = 2.60.$$

Κέρδος Kalman:

$$\mathbf{K}_1 = \Sigma_1^-\mathbf{C}^T S_1^{-1} = \begin{bmatrix} 2.10 & 1 \\ 1 & 1.20 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \frac{1}{2.60} = \begin{bmatrix} 2.10 \\ 1 \end{bmatrix} \frac{1}{2.60} = \begin{bmatrix} 0.8077 \\ 0.3846 \end{bmatrix}.$$

Posterior μέσος:

$$\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_1^- + \mathbf{K}_1 r_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0.8077 \\ 0.3846 \end{bmatrix} (-0.30) \approx \begin{bmatrix} 0.7577 \\ 0.8846 \end{bmatrix}.$$

Posterior συνδιακύμανση:

$$\Sigma_1 = (\mathbf{I} - \mathbf{K}_1 \mathbf{C}) \Sigma_1^-.$$

Εδώ

$$\mathbf{K}_1 \mathbf{C} = \begin{bmatrix} 0.8077 \\ 0.3846 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 0.8077 & 0 \\ 0.3846 & 0 \end{bmatrix}, \quad \mathbf{I} - \mathbf{K}_1 \mathbf{C} = \begin{bmatrix} 0.1923 & 0 \\ -0.3846 & 1 \end{bmatrix}.$$

Άρα

$$\Sigma_1 = \begin{bmatrix} 0.1923 & 0 \\ -0.3846 & 1 \end{bmatrix} \begin{bmatrix} 2.10 & 1 \\ 1 & 1.20 \end{bmatrix} \approx \begin{bmatrix} 0.4038 & 0.1923 \\ 0.1923 & 0.8154 \end{bmatrix}.$$

Ερμηνεία. Παρότι μετράμε μόνο τη θέση, η διόρθωση μεταβάλλει τόσο τη θέση όσο και την ταχύτητα: το υπόλοιπο θέσης r_1 αποδίδεται εν μέρει στην ταχύτητα μέσω της συνδιακύμανσης κατάστασης (και της σύζευξης στη δυναμική). Το διάνυσμα κέρδους $\mathbf{K}_1$ δείχνει πώς μια βαθμωτή μέτρηση ενημερώνει κάθε συνιστώσα της κατάστασης.

3.7.3 Υλοποίηση σε Python

Θεωρούμε ένα πρόβλημα 1Δ παρακολούθησης με 2Δ κατάσταση:

$$\mathbf{x}_k = \begin{bmatrix} p_k \\ v_k \end{bmatrix},$$

όπου p_k είναι η θέση και v_k η ταχύτητα στον διακριτό χρόνο k .

Μοντέλο κίνησης σταθερής ταχύτητας. Με περίοδο δειγματοληψίας $\Delta t > 0$, ένα τυπικό γραμμικό μοντέλο CV είναι

$$\mathbf{x}_k = \mathbf{A} \mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \quad \mathbf{A} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \quad \mathbf{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1}). \quad (3.45)$$

Ο θόρυβος διεργασίας $\mathbf{w}_{k-1}$ αποτυπώνει μη μοντελοποιημένες επιδράσεις (π.χ. μικρές επιταχύνσεις). Γενικά επιτρέπουμε χρονικά μεταβαλλόμενη αβεβαιότητα:

$$\mathbf{Q}_{k-1} \succeq \mathbf{0}.$$

Μοντέλο μέτρησης θέσης. Υποθέτουμε ότι μετράμε μόνο τη θέση:

$$\mathbf{y}_k = \mathbf{C} \mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k), \quad (3.46)$$

όπου $\mathbf{y}_k \in \mathbb{R}$ είναι βαθμωτή παρατήρηση και $\mathbf{R}_k \in \mathbb{R}^{1 \times 1}$ είναι η (χρονικά μεταβαλλόμενη) συνδιακύμανση του θορύβου μέτρησης (δηλ. η διακύμανση).

Υποθέσεις ανεξαρτησίας. Υποθέτουμε τις τυπικές συνθήκες Kalman:

$$w_k \perp v_j \quad \forall k, j, \quad w_k \text{ και } v_k \text{ είναι λευκοί και μηδενικού μέσου.}$$

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 # -----
5 # 2D Kalman Filter (CV model)
6 # State:  $x_k = [p_k, v_k]^T$ 
7 # Measurement:  $y_k = p_k + \text{noise}$ 
8 # Time-varying  $Q_k, R_k$ 
9 # -----
10
11 def simulate_cv_1d(T=80, dt=1.0, q_base=0.05, r_base=0.4, seed=0):
12     """
13     Simulate constant-velocity motion with random acceleration-like
14     process noise.
15     Returns: true states X (T+1,2), measurements y (T+1,), Q_k list,
16             R_k list
17     """
18     rng = np.random.default_rng(seed)
19
20     A = np.array([[1.0, dt],
21                  [0.0, 1.0]])
22     C = np.array([[1.0, 0.0]]) # measure position only
23
24     # True initial state
25     x = np.array([0.0, 1.0]) # p0, v0
26
27     X = np.zeros((T + 1, 2), dtype=float)
28     y = np.zeros(T + 1, dtype=float)
29
30     # Time-varying noise levels
31     Q_list = []
32     R_list = []
33
34     # Make time-varying R_k (sensor gets worse in the middle)
35     for k in range(T + 1):
36         bump = 1.0 + 2.5 * np.exp(-0.5 * ((k - 0.6 * T) / (0.12 * T))
37                                     ** 2)
38         Rk = np.array([[r_base * bump]], dtype=float) # scalar
39               measurement variance
40         R_list.append(Rk)
41
42     # Make time-varying Q_k (process more uncertain near the end)
43     for k in range(T):
44         ramp = 1.0 + 1.5 * (k / max(T - 1, 1))
45         # Simple diagonal Q_k on [p, v]; you can also use a physically-
46           motivated accel model.
47         Qk = np.array([[q_base * ramp, 0.0],
48                        [0.0, 2.0 * q_base * ramp]], dtype=
49                        float)
50         Q_list.append(Qk)
51
52     # Generate trajectory + measurements
53     X[0] = x

```

```

48     y[0] = (C @ x.reshape(-1, 1)).item() + rng.normal(0.0, np.sqrt(
        R_list[0].item()))
49
50     for k in range(1, T + 1):
51         # Process noise
52         wk = rng.multivariate_normal(mean=np.zeros(2), cov=Q_list[k -
            1])
53         x = A @ x + wk
54         X[k] = x
55
56         # Measurement noise
57         vk = rng.normal(0.0, np.sqrt(R_list[k].item()))
58         y[k] = (C @ x.reshape(-1, 1)).item() + vk
59
60     return A, C, X, y, Q_list, R_list
61
62
63 def kf_filter(A, C, y, Q_list, R_list, x0_hat, Sigma0, u=None, B=None):
64     """
65     Kalman filter for:
66         x_k = A x_{k-1} + B u_{k-1} + w_{k-1}, w ~ N(0, Q_{k-1})
67         y_k = C x_k + v_k, v ~ N(0, R_k)
68     with time-varying Q_{k-1}, R_k.
69
70     Returns:
71         x_hat (T+1,2), Sigma (T+1,2,2),
72         x_hat_minus (T+1,2), Sigma_minus (T+1,2,2),
73         K (T+1,2,1), innovations r (T+1,)
74     """
75     T = len(y) - 1
76     n = A.shape[0]
77     m = C.shape[0]
78
79     if u is None:
80         u = np.zeros((T, 1))
81     if B is None:
82         B = np.zeros((n, 1))
83
84     x_hat = np.zeros((T + 1, n), dtype=float)
85     x_hat_minus = np.zeros((T + 1, n), dtype=float)
86     Sigma = np.zeros((T + 1, n, n), dtype=float)
87     Sigma_minus = np.zeros((T + 1, n, n), dtype=float)
88     K = np.zeros((T + 1, n, m), dtype=float)
89     r = np.zeros(T + 1, dtype=float)
90
91     # Initialize
92     x_hat[0] = x0_hat
93     Sigma[0] = Sigma0
94
95     # (Optional) incorporate y0 as an update; here we do a standard
96     # update at k=0
97     # Prediction at k=0 is trivial (use initial prior as predicted),
98     # then update with y0
99     x_hat_minus[0] = x_hat[0]
100    Sigma_minus[0] = Sigma[0]
101    S0 = (C @ Sigma_minus[0] @ C.T + R_list[0])
102    K[0] = (Sigma_minus[0] @ C.T) @ np.linalg.inv(S0)
103    r[0] = y[0] - (C @ x_hat_minus[0].reshape(-1, 1)).item()

```

```

102     x_hat[0] = x_hat_minus[0] + (K[0] * r[0]).reshape(-1)
103     Sigma[0] = (np.eye(n) - K[0] @ C) @ Sigma_minus[0]
104
105     # Main loop
106     for k in range(1, T + 1):
107         # Predict
108         x_hat_minus[k] = (A @ x_hat[k - 1].reshape(-1, 1) + B @ u[k -
109             1].reshape(-1, 1)).reshape(-1)
110         Sigma_minus[k] = A @ Sigma[k - 1] @ A.T + Q_list[k - 1]
111
112         # Update
113         Sk = C @ Sigma_minus[k] @ C.T + R_list[k]
114         K[k] = (Sigma_minus[k] @ C.T) @ np.linalg.inv(Sk)
115         r[k] = y[k] - (C @ x_hat_minus[k].reshape(-1, 1)).item()
116         x_hat[k] = x_hat_minus[k] + (K[k] * r[k]).reshape(-1)
117
118         # Covariance update (simple form; Joseph form is more
119             numerically robust)
120         Sigma[k] = (np.eye(n) - K[k] @ C) @ Sigma_minus[k]
121
122     return x_hat, Sigma, x_hat_minus, Sigma_minus, K, r
123
124 # -----
125 # Run example
126 # -----
127 A, C, X_true, y, Q_list, R_list = simulate_cv_1d(
128     T=120, dt=1.0,
129     q_base=0.04, r_base=0.25,
130     seed=3
131 )
132
133 # Prior (intentionally imperfect)
134 x0_hat = np.array([0.0, 0.0])
135 Sigma0 = np.array([[2.0, 0.0],
136     [0.0, 2.0]])
137
138 x_hat, Sigma, x_hat_minus, Sigma_minus, K, r = kf_filter(
139     A=A, C=C, y=y,
140     Q_list=Q_list, R_list=R_list,
141     x0_hat=x0_hat, Sigma0=Sigma0
142 )
143
144 # -----
145 # Plot
146 # -----
147 t = np.arange(len(y))
148
149 plt.figure()
150 plt.plot(t, X_true[:, 0], label="true position $p_k$")
151 plt.plot(t, y, label="measured position $y_k$")
152 plt.plot(t, x_hat[:, 0], label="KF estimate $\hat{p}_k$")
153 plt.xlabel("k")
154 plt.ylabel("position")
155 plt.title("2D Kalman filter (state=[position, velocity], measurement=
156     position)")
157 plt.legend()
158 plt.grid(True)

```

```

157
158 plt.figure()
159 plt.plot(t, X_true[:, 1], label="true velocity $v_k$")
160 plt.plot(t, x_hat[:, 1], label="KF estimate $\hat{v}_k$")
161 plt.xlabel("k")
162 plt.ylabel("velocity")
163 plt.title("Velocity is inferred from position-only measurements")
164 plt.legend()
165 plt.grid(True)
166
167 plt.figure()
168 plt.plot(t, Sigma[:, 0, 0], label="posterior var(p): $\Sigma_k[0,0]$")
169 plt.plot(t, Sigma[:, 1, 1], label="posterior var(v): $\Sigma_k[1,1]$")
170 plt.xlabel("k")
171 plt.ylabel("variance")
172 plt.title("Posterior uncertainties shrink/grow with Q_k and R_k")
173 plt.legend()
174 plt.grid(True)
175
176 plt.show()

```

Listing 2: Kalman Filter implementation

3.8 Αλγόριθμος και Πρακτικές Παρατηρήσεις

3.8.1 Αλγόριθμος Φίλτρου Kalman (Πρόβλεψη–Διόρθωση)

Στη συνέχεια συνοψίζουμε το φίλτρο Kalman ως μια απλή αναδρομή πρόβλεψης–ενημέρωσης. Επιτρέπουμε χρονικά μεταβαλλόμενους πίνακες συστήματος και συνδιακυμάνσεις θορύβου.

Input: $\{A_k, B_k, C_k\}$; $\{Q_k, R_k\}$; αρχικό $(\hat{x}_0, \Sigma_0)$; μετρήσεις $\{y_k\}_{k \geq 1}$; είσοδοι $\{u_k\}_{k \geq 0}$

Output: Φιλτραρισμένες εκτιμήσεις $\{(\hat{x}_k, \Sigma_k)\}_{k \geq 1}$

```

1 for  $k = 1, 2, \dots$  do
    // Πρόβλεψη
2    $\hat{x}_k^- \leftarrow A_k \hat{x}_{k-1} + B_k u_{k-1}$ 
3    $\Sigma_k^- \leftarrow A_k \Sigma_{k-1} A_k^\top + Q_{k-1}$ 

    // Διόρθωση
4    $r_k \leftarrow y_k - C_k \hat{x}_k^-$                                      // καινοτομία
5    $S_k \leftarrow C_k \Sigma_k^- C_k^\top + R_k$                          // συνδ. καινοτομίας
6    $K_k \leftarrow \Sigma_k^- C_k^\top S_k^{-1}$                            // κέρδος Kalman
7    $\hat{x}_k \leftarrow \hat{x}_k^- + K_k r_k$ 
8    $\Sigma_k \leftarrow (I - K_k C_k) \Sigma_k^-$ 

                                     // ή μορφή Joseph (3.51)

```

Εννοιολογική σύνοψη. Το φίλτρο Kalman εναλλάσσει:

- **Πρόβλεψη:** χρήση του μοντέλου για προώθηση του μέσου και της συνδιακύμανσης.
- **Διόρθωση:** χρήση της μέτρησης για μείωση της αβεβαιότητας και βελτίωση της εκτίμησης.

Αυτή η αναδρομή είναι η κλειστής μορφής Γκαουσιανή λύση του φίλτρου Bayes για γραμμικά συστήματα.

3.8.2 Διαίσθηση I: Ο Ρόλος του Κέρδους Kalman

Στον χρόνο k , το κέρδος Kalman είναι

$$\mathbf{K}_k = \Sigma_k^- \mathbf{C}_k^\top (\mathbf{C}_k \Sigma_k^- \mathbf{C}_k^\top + \mathbf{R}_k)^{-1} = \Sigma_k^- \mathbf{C}_k^\top \mathbf{S}_k^{-1}, \quad (3.47)$$

όπου

$$\mathbf{S}_k = \mathbf{C}_k \Sigma_k^- \mathbf{C}_k^\top + \mathbf{R}_k \quad (3.48)$$

είναι η συνδιακύμανση της καινοτομίας.

Η (3.47) δείχνει καθαρά ότι το $\mathbf{K}_k$ είναι μια αυτόματη στάθμιση μεταξύ της πρόβλεψης του μοντέλου και της μέτρησης:

- **Μεγάλος θόρυβος μέτρησης $\mathbf{R}_k \Rightarrow$ μεγάλο $\mathbf{S}_k \Rightarrow$ μικρό $\mathbf{K}_k$.** Ο όρος ενημέρωσης $\mathbf{K}_k \mathbf{r}_k$ αποσβένεται και το φίλτρο εμπιστεύεται περισσότερο το μοντέλο (την πρόβλεψη).
- **Μεγάλη prior αβεβαιότητα Σ_k^-** κάνει μεγάλο τον όρο $\Sigma_k^- \mathbf{C}_k^\top$ και αυξάνει το $\mathbf{K}_k$. Το φίλτρο εμπιστεύεται περισσότερο τη μέτρηση επειδή η πρόβλεψη είναι αβέβαιη.

Έτσι, το $\mathbf{K}_k$ παίζει τον ρόλο ενός συντελεστή μίξης που καθορίζεται από τα δεδομένα, ο οποίος εξισορροπεί την αξιοπιστία του μοντέλου έναντι της αξιοπιστίας του αισθητήρα σε κάθε βήμα χρόνου.

3.8.3 Διαίσθηση II: Καινοτομία = Νέα Πληροφορία

Η καινοτομία (υπόλοιπο μέτρησης) ορίζεται ως

$$\mathbf{r}_k = \mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^- \quad (3.49)$$

Ο όρος $\mathbf{C}_k \hat{\mathbf{x}}_k^-$ είναι αυτό που το μοντέλο προβλέπει ότι θα έπρεπε να είναι η μέτρηση. Το υπόλοιπο $\mathbf{r}_k$ αποτυπώνει το τμήμα της μέτρησης που δεν εξηγείται από την πρόβλεψη.

Δύο οριακές περιπτώσεις είναι διαφωτιστικές:

- Αν $\mathbf{r}_k \approx \mathbf{0}$, η μέτρηση συμφωνεί με την πρόβλεψη, οπότε η ενημέρωση $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \mathbf{r}_k$ κάνει μόνο μικρή διόρθωση.
- Αν $\|\mathbf{r}_k\|$ είναι μεγάλη, η μέτρηση διαφωνεί με την πρόβλεψη. Η διόρθωση είναι ισχυρή, αλλά εξακολουθεί να «ζυγίζεται» από το κέρδος $\mathbf{K}_k$ (δηλ. από την αβεβαιότητα).

Με αυτή την έννοια, το $\mathbf{r}_k$ είναι η νέα πληροφορία που περιέχεται στο $\mathbf{y}_k$ σε σχέση με όσα υπονοούν ήδη οι προηγούμενες μετρήσεις και το μοντέλο συστήματος.

3.8.4 Διαίσθηση III: Γιατί η Αβεβαιότητα Μειώνεται μετά την Ενημέρωση

Η ενημέρωση της posterior συνδιακύμανσης (βασική μορφή) είναι

$$\Sigma_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}_k) \Sigma_k^- \quad (3.50)$$

Αυτό δείχνει ότι η μέτρηση μειώνει την αβεβαιότητα μόνο κατά μήκος διευθύνσεων που «βλέπει» ο αισθητήρας. Πράγματι, το μοντέλο μέτρησης χαρτογραφεί την κατάσταση στον χώρο μέτρησης μέσω του $\mathbf{C}_k$: άρα, μόνο οι συνιστώσες του $\mathbf{x}_k$ που επηρεάζουν το $\mathbf{y}_k$ μπορούν να διορθωθούν.

Γεωμετρικά:

- Το βήμα πρόβλεψης προωθεί την αβεβαιότητα μέσω της δυναμικής και εισάγει θόρυβο διεργασίας (μέσω $\mathbf{Q}_{k-1}$), τυπικά διογκώνοντας το ελλειψοειδές συνδιακύμανσης.
- Το βήμα διόρθωσης χρησιμοποιεί τη μέτρηση για να συστέλλει την αβεβαιότητα στον παρατηρούμενο υπόχωρο, πολλαπλασιάζοντας με $(\mathbf{I} - \mathbf{K}_k \mathbf{C}_k)$.

Οι μη παρατηρούμενες διευθύνσεις παραμένουν αβέβαιες (ή μπορεί να συνεχίσουν να αυξάνονται κατά την πρόβλεψη).

Σχόλιο (αριθμητικά ανθεκτική ενημέρωση συνδιακύμανσης). Σε υλοποιήσεις συχνά χρησιμοποιείται η μορφή *Joseph*, η οποία διατηρεί καλύτερα τη συμμετρία και τη θετική ημιοριστικότητα σε πεπερασμένη ακρίβεια:

$$\Sigma_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}_k) \Sigma_k^- (\mathbf{I} - \mathbf{K}_k \mathbf{C}_k)^\top + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^\top \quad (3.51)$$

3.9 Πρακτικές Παρατηρήσεις

3.9.1 Όταν Παραβιάζονται οι Υποθέσεις

Το κλασικό φίλτρο Kalman είναι ακριβές (και βέλτιστο με την έννοια MMSE) για γραμμική δυναμική, γραμμικές παρατηρήσεις και Γκαουσιανό θόρυβο. Όταν αυτές οι υποθέσεις δεν ισχύουν, ο αλγόριθμος μπορεί να χρησιμοποιηθεί ως προσέγγιση, αλλά η απόδοση μπορεί να υποβαθμιστεί και, σε ακραίες περιπτώσεις, το φίλτρο μπορεί να γίνει ασυνεπές (πολύ μικρή συνδιακύμανση) ή ακόμη και να αποκλίνει.

Μη γραμμική δυναμική και/ή μη γραμμικές παρατηρήσεις. Σε πολλά συστήματα τα μοντέλα κατάστασης και μέτρησης έχουν τη μη γραμμική μορφή

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1}, \quad \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k, \quad (3.52)$$

με $\mathbf{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ και $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$. Σε αυτή την περίπτωση, η αναδρομή του φίλτρου Bayes παραμένει έγκυρη, αλλά η posterior είναι γενικά μη Γκαουσιανή, άρα δεν μπορεί να παρασταθεί ακριβώς μόνο από μέσο και συνδιακύμανση.

Συνηθισμένες προσεγγίσεις είναι:

- **Εκτεταμένο Φίλτρο Kalman (EKF):** γραμμικοποιεί τα $\mathbf{f}$ και $\mathbf{h}$ γύρω από την τρέχουσα εκτίμηση με χρήση Ιακωβιανών, και εφαρμόζει ενημέρωση τύπου Kalman στο τοπικά γραμμικοποιημένο μοντέλο.

- **Ασύμμετρο/Unscented Φίλτρο Kalman (UKF):** προωθεί ένα ντετερμινιστικό σύνολο σημείων σ μέσα από τα f και h για να προσεγγίσει μέσο/συνδιακύμανση χωρίς ρητές Ιακωβιανές.
- **Φίλτρα σωματιδίων (particle filters):** προσεγγίζουν την πεποίθηση με ένα σταθμισμένο σύνολο δειγμάτων, επιτρέποντας μη γραμμική και μη Γκαουσιανή συμπερασματολογία με υψηλότερο υπολογιστικό κόστος.

Μη Γκαουσιανός θόρυβος, βαριές ουρές και ακραίες τιμές. Αν ο θόρυβος έχει βαριές ουρές (π.χ. περιέχει outliers), η Γκαουσιανή πιθανοφάνεια (likelihood) που υπονοείται από την ενημέρωση Kalman μπορεί να δώσει υπερβολικό βάρος σε αλλοιωμένες μετρήσεις. Στην πράξη αυτό μπορεί να οδηγήσει σε μεγάλα υπόλοιπα και λανθασμένες διορθώσεις. Υπάρχει εκτενής βιβλιογραφία για **ανθεκτικές** παραλλαγές, π.χ. αντικατάσταση της τετραγωνικής απώλειας (που υποστηρίζει το MMSE) με ανθεκτικές απώλειες (Huber/-Tukey), ή μοντελοποίηση του θορύβου με κατανομές βαριών ουρών (π.χ. Student- t), που ουσιαστικά υποβαθμίζουν τα outliers.

Αναντιστοιχία μοντέλου και λανθασμένη προδιαγραφή αβεβαιότητας. Ακόμη και όταν το μοντέλο είναι γραμμικό, η απόδοση εξαρτάται κρίσιμα από το πόσο καλά οι $\mathbf{Q}_{k-1}$ (αβεβαιότητα διεργασίας) και $\mathbf{R}_k$ (αβεβαιότητα μέτρησης) αποτυπώνουν την πραγματικότητα. Αν ρυθμιστούν λάθος, το φίλτρο μπορεί να γίνει:

- **Αργό** (υπερβολική εμπιστοσύνη στο μοντέλο),
- **Θορυβώδες** (υπερβολική εμπιστοσύνη στις μετρήσεις),
- **Ασυνεπές** (η συνδιακύμανση υποεκτιμά το πραγματικό σφάλμα),
- **Αποκλίνον** (τα σφάλματα αυξάνονται και η εκτίμηση γίνεται ασταθής).

Συμπέρασμα. Το φίλτρο Kalman είναι τόσο καλό όσο το μοντέλο και η περιγραφή της αβεβαιότητάς του: οι $\mathbf{Q}_{k-1}$, $\mathbf{R}_k$ κωδικοποιούν αυτό που δεν γνωρίζουμε.

3.9.2 Επίδραση της Ρύθμισης των $\mathbf{Q}_k$ και $\mathbf{R}_k$

Η διαίσθηση της ρύθμισης φαίνεται απευθείας από την έκφραση του κέρδους

$$\mathbf{K}_k = \mathbf{\Sigma}_k^- \mathbf{C}_k^\top (\mathbf{C}_k \mathbf{\Sigma}_k^- \mathbf{C}_k^\top + \mathbf{R}_k)^{-1}. \quad (3.53)$$

Θυμίζουμε επίσης ότι η prior συνδιακύμανση παράγεται από το βήμα πρόβλεψης:

$$\mathbf{\Sigma}_k^- = \mathbf{A}_k \mathbf{\Sigma}_{k-1} \mathbf{A}_k^\top + \mathbf{Q}_{k-1}. \quad (3.54)$$

Αύξηση του $\mathbf{Q}_{k-1}$ (περισσότερος θόρυβος διεργασίας). Από την (3.54), η αύξηση του $\mathbf{Q}_{k-1}$ αυξάνει το $\mathbf{\Sigma}_k^-$, κάτι που τυπικά αυξάνει το $\mathbf{K}_k$ στην (3.53). Άρα ο όρος ενημέρωσης $\mathbf{K}_k \mathbf{r}_k$ γίνεται μεγαλύτερος.

Συμπεριφορά: το φίλτρο εμπιστεύεται περισσότερο τις μετρήσεις. Αυτό κάνει την εκτίμηση πιο «αντιδραστική» σε αλλαγές, αλλά συνήθως πιο θορυβώδη.

Αύξηση του R_k (περισσότερος θόρυβος μέτρησης). Από την (3.53), η αύξηση του R_k αυξάνει τον παρονομαστή $C_k \Sigma_k^- C_k^\top + R_k$, άρα μειώνει το K_k .

Συμπεριφορά: το φίλτρο εμπιστεύεται περισσότερο την πρόβλεψη του μοντέλου. Αυτό κάνει την εκτίμηση πιο ομαλή, αλλά συνήθως πιο αργή στο να αντιδρά σε πραγματικές αλλαγές.

Μια χρήσιμη βαθμωτή διαίσθηση (σε 1Δ). Αν $C_k = 1$ και όλα είναι βαθμωτά,

$$K_k = \frac{\Sigma_k^-}{\Sigma_k^- + R_k}, \quad \Sigma_k^- = \Sigma_{k-1} + Q_{k-1}.$$

Άρα:

$$Q_{k-1} \uparrow \Rightarrow \Sigma_k^- \uparrow \Rightarrow K_k \uparrow, \quad R_k \uparrow \Rightarrow K_k \downarrow.$$

Αυτό δείχνει ρητά πώς τα Q και R ελέγχουν τη στάθμιση μοντέλου-μέτρησης.

Πρακτική συνέπεια. Η ρύθμιση των Q_{k-1} και R_k δεν είναι απλώς ένα «αριθμητικό trick»: κωδικοποιεί την πεποίθηση του σχεδιαστή για την αβεβαιότητα και καθορίζει την ποιοτική συμπεριφορά του φίλτρου.

3.10 Φιλτράρισμα vs. Εξομάλυνση

3.10.1 Φιλτράρισμα vs. Εξομάλυνση

Θεωρούμε ένα γραμμικό-Γκαουσιανό μοντέλο στο χώρο κατάστασης και μια πεπερασμένη ακολουθία μετρήσεων $\mathbf{y}_{1:T}$.

Φιλτράρισμα (online / αιτιακή συμπερασματολογία). Το φίλτρο Kalman υπολογίζει την κατανομή φιλτραρίσματος

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:k}), \quad (3.55)$$

η οποία εξαρτάται μόνο από τις μετρήσεις που είναι διαθέσιμες μέχρι τη χρονική στιγμή k . Η έξοδος του φίλτρου είναι ο posterior μέσος και η συνδιακύμανση

$$(\hat{\mathbf{x}}_k, \Sigma_k),$$

και η αναδρομή είναι αιτιακή και κατάλληλη για λειτουργία σε πραγματικό χρόνο.

Εξομάλυνση (offline / μη-αιτιακή συμπερασματολογία). Αν όλες οι μετρήσεις $\mathbf{y}_{1:T}$ είναι διαθέσιμες, μπορούμε αντίθετα να υπολογίσουμε την κατανομή εξομάλυνσης

$$p(\mathbf{x}_k \mid \mathbf{y}_{1:T}), \quad (3.56)$$

η οποία χρησιμοποιεί τόσο παρελθοντικές όσο και μελλοντικές μετρήσεις ως προς τον χρόνο k . Αυτό δίνει τις εξομαλυσμένες εκτιμήσεις

$$(\hat{\mathbf{x}}_k^s, \Sigma_k^s).$$

Βασική ιδιότητα. Η εξομάλυνση δεν μπορεί ποτέ να είναι χειρότερη από το φιλτράρισμα ως προς την αβεβαιότητα:

$$\boxed{\Sigma_k^s \preceq \Sigma_k \quad \forall k,} \quad (3.57)$$

όπου $\preceq$ δηλώνει τη διάταξη Loewner (θετικά ημιορισμένη). Διαισθητικά, οι μελλοντικές μετρήσεις περιέχουν πληροφορία για παρελθοντικές καταστάσεις και μπορούν να χρησιμοποιηθούν ώστε να διορθώσουν αναδρομικά παλαιότερες εκτιμήσεις.

Εννοιολογική σύνοψη.

- Φιλτράρισμα: πραγματικού χρόνου, αιτιακό, χρησιμοποιεί $\mathbf{y}_{1:k}$.
- Εξομάλυνση: offline, μη-αιτιακό, χρησιμοποιεί $\mathbf{y}_{1:T}$.
- Μελλοντικά δεδομένα πάντα βελτιώνουν (ή αφήνουν αμετάβλητες) τις εκτιμήσεις παρελθοντικών καταστάσεων.

3.10.2 Εξομάλυνση Kalman: Αλγόριθμος Rauch–Tung–Striebel (RTS)

Για γραμμικά–Γκαουσιανά συστήματα, η εξομάλυνση μπορεί να γίνει αποδοτικά με ένα backward πέρασμα μετά το φιλτράρισμα Kalman. Ο πιο συνηθισμένος εξομαλυντής είναι ο **εξομαλυντής Rauch–Tung–Striebel (RTS)**.

Forward πέρασμα (φίλτρο Kalman). Τρέχουμε το τυπικό φίλτρο Kalman για $k = 0:T$ και αποθηκεύουμε:

$$\hat{\mathbf{x}}_k, \quad \hat{\mathbf{x}}_{k+1}^-, \quad \Sigma_k, \quad \Sigma_{k+1}^-.$$

Backward πέρασμα (αναδρομή εξομάλυνσης). Για $k = T - 1, \dots, 0$, ορίζουμε το **κέρδος εξομάλυνσης**

$$\boxed{\mathbf{G}_k = \Sigma_k \mathbf{A}_k^\top (\Sigma_{k+1}^-)^{-1}.} \quad (3.58)$$

Ο εξομαλυμένος μέσος ενημερώνεται ως

$$\boxed{\hat{\mathbf{x}}_k^s = \hat{\mathbf{x}}_k + \mathbf{G}_k (\hat{\mathbf{x}}_{k+1}^s - \hat{\mathbf{x}}_{k+1}^-),} \quad (3.59)$$

και η εξομαλυμένη συνδιακύμανση ως

$$\boxed{\Sigma_k^s = \Sigma_k + \mathbf{G}_k (\Sigma_{k+1}^s - \Sigma_{k+1}^-) \mathbf{G}_k^\top.} \quad (3.60)$$

Ερμηνεία. Ο όρος backward διόρθωσης $\hat{\mathbf{x}}_{k+1}^s - \hat{\mathbf{x}}_{k+1}^-$ αναπαριστά πληροφορία που αποκαλύπτεται από μελλοντικές μετρήσεις. Το κέρδος εξομάλυνσης $\mathbf{G}_k$ προωθεί αυτή την πληροφορία προς τα πίσω μέσω της δυναμικής.

3.10.3 Batch Οπτική: MAP Εκτίμηση Ολόκληρης της Τροχιάς

Το φιλτράρισμα και η εξομάλυνση μπορούν επίσης να ερμηνευθούν από τη σκοπιά της **batch βελτιστοποίησης**. Υπό το γραμμικό–Γκαουσιανό μοντέλο, η εκτίμηση ολόκληρης της τροχιάς κατάστασης $\mathbf{x}_{0:T}$ είναι ισοδύναμη με ένα πρόβλημα μέγιστης εκ των υστέρων πιθανότητας (MAP).

Διατύπωση batch MAP. Η posterior κατανομή της πλήρους τροχιάς είναι Γκαουσιανή και η MAP εκτίμηση προκύπτει από τη λύση

$$\mathbf{x}_{0:T}^* = \arg \min_{\mathbf{x}_{0:T}} \left[\sum_{k=1}^T \|\mathbf{y}_k - \mathbf{C}_k \mathbf{x}_k\|_{\mathbf{R}_k^{-1}}^2 + \sum_{k=1}^T \|\mathbf{x}_k - \mathbf{A}_k \mathbf{x}_{k-1} - \mathbf{B}_k \mathbf{u}_{k-1}\|_{\mathbf{Q}_{k-1}^{-1}}^2 + \|\mathbf{x}_0 - \hat{\mathbf{x}}_0\|_{\mathbf{\Sigma}_0^{-1}}^2 \right]. \quad (3.61)$$

Πρόκειται για ένα τετραγωνικό πρόβλημα ελαχίστων τετραγώνων με block-τριδιαγώνια δομή ως προς τον χρόνο.

Σχέση με το φιλτράρισμα και την εξομάλυνση Kalman.

- Το φιλτράρισμα Kalman υπολογίζει τη MAP εκτίμηση *αναδρομικά*, χρησιμοποιώντας μόνο παρελθοντικά δεδομένα.
- Η εξομάλυνση Kalman υπολογίζει την ίδια λύση με την (3.61), αλλά αποδοτικά, χωρίς να σχηματίζει ή να επιλύει ρητά το πλήρες batch πρόβλημα.
- Και τα δύο αποτελούν αλγοριθμικές υλοποιήσεις της Γκαουσιανής Bayesian συμπερασματολογίας.

Η μεγάλη εικόνα.

- **Φιλτράρισμα:** online λύση ενός προβλήματος συμπερασματολογίας που μεγαλώνει με τον χρόνο.
- **Εξομάλυνση:** offline λύση που χρησιμοποιεί όλα τα δεδομένα.
- **Batch MAP:** καθολική οπτική βελτιστοποίησης.

Το φιλτράρισμα και η εξομάλυνση Kalman δεν είναι ad hoc αλγόριθμοι: είναι αποδοτικοί αναδρομικοί επιλύτες για δομημένα Γκαουσιανά προβλήματα ελαχίστων τετραγώνων.

Glossary (English – Greek)

State Κατάσταση

State-space model Μοντέλο στο χώρο κατάστασης

Hidden state / Latent state Κρυφή κατάσταση / Λανθάνουσα κατάσταση

Measurement / Observation Μέτρηση / Παρατήρηση

Control / Input Έλεγχος / Είσοδος

Process model / Motion model Μοντέλο διεργασίας / Μοντέλο κίνησης

Observation model / Measurement model Μοντέλο παρατήρησης / Μοντέλο μέτρησης

Process noise Θόρυβος διεργασίας

Measurement noise Θόρυβος μέτρησης

Covariance matrix Πίνακας συνδιακύμανσης

Prior Πρότερη κατανομή

Posterior Εκ των υστέρων κατανομή

Belief Πεποίθηση (κατανομή κατάστασης)

Prediction / Time update Πρόβλεψη / Ενημέρωση χρόνου

Correction / Measurement update Διόρθωση / Ενημέρωση μέτρησης

Innovation / Residual Καινοτομία / Υπόλοιπο / Σφάλμα

Innovation covariance Συνδιακύμανση καινοτομίας

Kalman gain Κέρδος Kalman

Filtering Φιλτράρισμα

Smoothing Εξομάλυνση

Kalman filter Φίλτρο Kalman

Extended Kalman Filter (EKF) Εκτεταμένο φίλτρο Kalman

Unscented Kalman Filter (UKF) Unscented φίλτρο Kalman

Particle filter Φίλτρο σωματιδίων

Wiener filter Φίλτρο Wiener

Linear Minimum Mean Square Error (LMMSE) Γραμμικό ελάχιστο μέσο τετραγωνικό σφάλμα

Mean squared error (MSE) Μέσο τετραγωνικό σφάλμα

Recursive estimation Αναδρομική εκτίμηση

Batch estimation Batch εκτίμηση

Maximum a Posteriori (MAP) Μέγιστη εκ των υστέρων πιθανότητα (MAP)

Gaussian distribution Γκαουσιανή κατανομή

White noise Λευκός θόρυβος

Stationary / Wide-sense stationary (WSS) Στάσιμη / Ευρέως στάσιμη

FIR filter FIR φίλτρο (πεπερασμένης χρονικής απόκρισης)

Time-varying Χρονικά μεταβαλλόμενο

Model-based Βασισμένο σε μοντέλο

Supervised (training) Επιβλεπόμενη (εκπαίδευση)

Unsupervised Μη επιβλεπόμενη

Second-order statistics Στατιστικά δεύτερης τάξης

Autocorrelation Αυτοσυσχέτιση

Cross-correlation Διασταυρούμενη/Ετερό συσχέτιση

Power spectral density (PSD) Φασματική πυκνότητα ισχύος