

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS

# **ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ**

**ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ**

**ΜΠΕΪΖΙΑΝΗ ΑΝΑΛΥΣΗ ΙΑΤΡΙΚΩΝ ΕΙΚΟΝΩΝ**

**ΜΠΑΚΑΣ ΚΩΝΣΤΑΝΤΙΝΟΣ**

**Διπλωματική Εργασία**

που υποβλήθηκε στο Τμήμα Στατιστικής  
του Οικονομικού Πανεπιστημίου Αθηνών στο πλαίσιο του  
Προπτυχιακού Προγράμματος Σπουδών

Αθήνα

Σεπτέμβριος 2021



# ΑΦΙΕΡΩΣΗ



## ΕΥΧΑΡΙΣΤΙΕΣ



## **ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ**

Γεννήθηκα στην Αθήνα τον Μάρτιο του έτους 1999. Τον Οκτώβριο του 2017 άρχισα να φοιτώ στο Τμήμα Στατιστικής του Οικονομικού Πανεπιστημίου Αθηνών. Ολοκλήρωσα με επιτυχία τις σπουδές μου μετά από 4 χρόνια και εκπόνησα την προπτυχιακή διπλωματική μου εργασία το εαρινό εξάμηνο του 2021. Τώρα, εργάζομαι ως αναλυτής δεδομένων στην Business-to-Business βιομηχανία της Sony Europe.





# ΠΕΡΙΛΗΨΗ

Μπάκας Κωνσταντίνος

## ΜΠΕΥΪΖΙΑΝΗ ΑΝΑΛΥΣΗ ΙΑΤΡΙΚΩΝ ΕΙΚΟΝΩΝ

Σεπτέμβριος 2021

Παρουσίαση και περιγραφή διαφόρων στατιστικών τεχνικών για ανάλυση ιατρικών εικόνων. Ειδικότερα, θα επικεντρωθούμε σε Μπεϋζιανές μεθόδους βελτιστοποίησης προτύπων για ιατρικές απεικονίσεις. Θα προσπαθήσουμε να αναδείξουμε μια νέα υποσχόμενη μέθοδο για την αποθρομβοποίηση εικόνων με εφαρμογή στην ιατρική.



# ΠΕΡΙΕΧΟΜΕΝΑ

<b>1. Εισαγωγή</b>	<b>1</b>
<b>2. Τεχνικές Βαθιάς Μάθησης για αναγνώριση εικόνας</b>	<b>5</b>
2.1 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα	6
2.2 Ταξινόμηση εικόνας και ανίχνευση αντικειμένων	10
2.2.1 Συνελκτικά Νευρωνικά Δίκτυα	10
2.3 Ανάλυση εικόνας και ανοικοδόμηση	14
2.3.1 Αυτόματοι κωδικοποιητές	15
2.3.2 Variational Autoencoders	18
2.4 Συμπεράσματα και συζήτηση	24
<b>3. Μπεϋζιανές τεχνικές ανάλυσης εικόνας</b>	<b>27</b>
3.1 Μπεϋζιανή συμπερασματολογία	27
3.1.1 Εκ-των-προτέρων κατανομή	28
3.1.2 Πιθανοφάνεια	31
3.1.3 Αξιόπιστα σύνολα	32
3.1.4 Μπεϋζιανοί έλεγχοι υποθέσεων	32
3.1.5 Επιλογή μοντέλου	33
3.1.6 Δειγματολήπτης Gibbs	34
3.2 Μπεϋζιανή ανάλυση εικόνας	36
3.2.1 Τυχαία Πεδία Markov και εκ-των-προτέρων κατανομή	37
3.2.2 Υπολογιστική εκτίμηση Μπεϋζιανής εικόνας	39
3.2.3 Κανονικοποίηση Tikhonov	40
3.3 Ανάλυση Fourier	43
3.3.1 Μιγαδικοί αριθμοί	44
3.3.2 Διακριτός Μετασχηματισμός Fourier	45

3.3.3 Fast Fourier Transform . . . . .	48
3.4 Μπεϋζιανή ανάλυση εικόνας στον χώρο Fourier . . . . .	51
3.4.1 Μοντελοποίηση . . . . .	52
3.4.2 Εκ-των-προτέρων κατανομή . . . . .	53
3.4.3 Εκ-των-προτέρων κατανομή καθοδηγούμενη από βάση δεδομένων (Data-Driven approach) . . . . .	54
3.4.4 Εκ-των-προτέρων κατανομή βασισμένη σε παραμετρικές συναρτήσεις . . . . .	54
3.4.5 Πιθανοφάνεια . . . . .	56
3.4.6 Εκ-των-υστέρων εκτίμηση . . . . .	57
3.4.7 Αλγόριθμος BIFS . . . . .	59
3.5 Συμπεράσματα-Σύνοψη κεφαλαίου-Συζήτηση . . . . .	59
<b>4. Περιγραφή δεδομένων</b>	<b>61</b>
<b>5. Εφαρμογή και αποτελέσματα</b>	<b>63</b>
5.1 Κανονική κατανομή για την πιθανοφάνεια του θορύβου στον χώρο Fourier . . . . .	64
5.2 Κατανομή Rice για την πιθανοφάνεια του θορύβου στον χώρο Fourier . . . . .	64
5.3 Robustness . . . . .	65
5.4 Συμπεριφορά παραμετρικών συναρτήσεων . . . . .	67
5.5 Σύγκριση του αλγορίθμου BIFS με την κανονικοποίηση Τίχονοφ . . . . .	69
5.6 Συμπεράσματα / Επίλογος Κεφαλαίου . . . . .	75
<b>6. Συμπεράσματα και μελλοντική έρευνα</b>	<b>77</b>
<b>Αναφορές</b>	<b>78</b>

# Κατάλογος Πινάκων

5.1	Χρόνοι υλοποίησης του αλγορίμου BIFS για τις διαφορετικές κατανομές πιθανοφανειών του θορύβου στον χώρο Fourier . . . . .	67
-----	---	----



# Κατάλογος Σχημάτων

2.1	Δομή ενός βιολογικού και ενός τεχνητού νευρώνα (perceptron)	6
2.2	Αρχιτεκτονική δομή ενός Τεχνητού Νευρωνικού Δικτύου	9
2.3	Αρχιτεκτονική δομή ενός Συνελικτικού Νευρωνικού Δικτύου (CNN)	14
2.4	Αρχιτεκτονική δομή ενός αυτόματου κωδικοποιητή (Autoencoder)	17
2.5	Κατευθυνόμενο διάγραμμα δύο τυχαίων μεταβλητών	21
2.6	Παραλλαγή κατευθυνόμενου γραφήματος δύο τυχαίων μεταβλητών	22
2.7	Αρχιτεκτονική δομή ενός Variational Autoencoder (VAE)	24
2.8	Κατευθυνόμενο γραφήμα δύο τυχαίων μεταβλητών με την συνεισφορά μιας τυχαίας μεταβλητής	24
3.9	Απεικόνιση ενός εγκεφάλου στον χώρο εικόνας και στον χώρο συχνοτήτων έπειτα από την εφαρμογή του διακριτού μετασχηματισμού Fourier	51
4.10	Αναπαράσταση της φαιάς ουσίας, της λευκής ουσίας, του εγκεφαλονοτιαίου υγρού και μιας μέσης πιθανοθεωρητική απεικόνιση του ιστού του εγκεφάλου	62
5.11	Εφαρμογή του αλγορίθμου BIFS και σύγκριση μεταξύ της κανονική κατανομής και της κατανομή Rice ως πιθανοφάνεια του θορύβου στον χώρο Fourier	66
5.12	Εφαρμογή του αλγορίθμου BIFS με συνάρτησεις παραμέτρων $f_{\mu}( k ) = \frac{a}{\sqrt{ k }}, \frac{a}{ k }, \frac{a}{k^2}$ για διάφορες τιμές των τιμών $c, a$	70
5.13	Διάγραμμα των γειτονικών πίξελ του $x_i$ , με τα χρώματα των τεσσάρων γειτόνων του να αντιστοιχούν στις διαγωνίους του πίνακα $\Gamma$	72
5.14	Εφαρμογή της μεθόδου κανονικοποίησης Tikhonov	73
5.15	Σύγκριση του αλγορίθμου BIFS με την μέθοδο της κανονικοποίησης Tikhonov	74





# 1. Εισαγωγή

Μια εικόνα αποτελεί σημαντικό εργαλείο για την αποτύπωση και την κατανόηση μιας δεδομένης χρονικής στιγμής. Η δυνατότητα να σταματάει τον χρόνο μας δίνει πληροφορίες που δεν θα μπορούσαν να ληφθούν υπόψη σε διαφορετική περίπτωση, μιας και ο χρόνος δεν γυρίζει πίσω. Έτσι, έχουμε την ευκαιρία να αναλύσουμε στιγμιότυπα και να βγάλουμε τα απαραίτητα συμπεράσματα γι' αυτά. Στο σημείο αυτό αξίζει να αναφέρουμε την γνωστή παροιμία «μια εικόνα ισούται με χίλιες λέξεις». Πολύπλοκες (ακόμα και πολλαπλές) έννοιες μπορούν να εξηγηθούν και να αναπαρασταθούν μέσω μονάχα μιας εικόνας. Συνειδητοποιούμε λοιπόν, ότι δύναμη της πληροφορία που μας παρέχει ακόμα και μια εικόνα είναι ανεκτίμητη. Υπάρχει μεγάλη ποικιλία μεθόδων που έχουν αναπτυχθεί για την ανάλυση εικόνων, αλλά στη παρούσα εργασία θα παρουσιάσουμε και θα αναπτύξουμε ένα υποσύνολο αυτών. Εν συνεχεία, μπορούμε να προσαρμόσουμε μεθόδους ανάλυσης εικόνων ακόμα και σε βίντεο, αφού στην πραγματικότητα τα τελευταία αποτελούν μια αλληλουχία συνεχών στιγμιότυπων.

Μια πληθώρα τεχνικών ανάλυσης εικόνων που ανθοφορούν στις μέρες μας στηρίζονται στην Βαθιά Μάθηση (περιοχή της επιστήμης υπολογιστών, που θα μπορούσε να χαρακτηριστεί ως υποσύνολο του τομέα της Τεχνητής Νοημοσύνης). Οι ιδέες και τα θεμέλια των τεχνικών αυτών είχαν τεθεί από τα μέσα του 20ού αιώνα, αλλά βρήκαν την αίγλη τους τα τελευταία χρόνια στον τομέα της επιστήμης των υπολογιστών. Βασική τους λειτουργία αποτελεί η αναζήτηση και η αποστήθιση μοτίβων, προερχόμενων από παρατηρήσεις (στην περίπτωσή μας εικόνες) μιας βάσης δεδομένων, ώστε να μπορούν να τα αναγνωρίζουν σε παρατηρήσεις εκτός αυτής της βάσης. Η ιδέα και η αρχιτεκτονική των Τεχνητών Νευρωνικών Δικτύων (Artificial Neural Networks) καθιστούν τον πυρήνα της Βαθιάς Μάθησης. Φυσικά, η πολυπλοκότητα των εννοιών αυτών είναι ιδιαίτερα αυξημένη και θα φροντίσουμε να θέσουμε το κατάλληλο υπόβαθρο για την κατανόηση τους. Αρχικά, περιγράφουμε κάποιες κλασικές μεθόδους της Βαθιάς Μάθησης και στη συνέχεια επεκτεινόμαστε σε ορισμένες που βρίσκουν εφαρμογές στην αναγνώριση προτύπων (ή μοτίβων)

των εικόνων.

Κύριο μέρος της εργασίας μας αποτελεί η ανάδειξη Μπεϋζιανών μεθόδων ανάλυσης για εικόνες. Η Μπεϋζιανή στατιστική, γεννήθηκε και αυτή στα μέσα του 20ού αιώνα και πιο συγκεκριμένα την δεκαετία του 1950. Αξίζει να αναφέρουμε, ότι τα θεμέλια για την γέννηση αυτής της περιοχής των μαθηματικών έθεσε ο Thomas Bayes τον 18ο αιώνα. Οι επιστήμονες που υποστήριζαν τον τρόπο σκέψης του Thomas Bayes διαχωρίστηκαν από τους υπόλοιπους «κλασικούς στατιστικούς» (όπως ονομάστηκαν έπειτα από τον διαχωρισμό) και ονομάστηκαν Μπεϋζιανοί στατιστικοί. Η διάδοση των αρχών της Μπεϋζιανής θεωρίας είχε σημαντικό αντίκτυπο την δεκαετία του 1950, αλλά η εφαρμογή τους ήταν απογοητευτική. Η ορθότητα των θεωρήματων και των αποτελεσμάτων ήταν αξιοσημείωτη, με δυστυχώς την υπολογιστική τους εκτίμηση να κρίνεται αδύνατη. Από τις αρχές του 21ού αιώνα έως και σήμερα, η υπολογιστική δύναμη (hardware) και οι υπολογιστικές τεχνικές αναπτύσσονται ασταμάτητα, με αποτέλεσμα την πραγματοποίηση των υπολογισμών αυτών. Έτσι, το έδαφος για την μοντελοποίηση προβλημάτων μέσω της Μπεϋζιανής στατιστικής μετατρέπεται σε εξαιρετικά πρόσφορο για τους ερευνητές.

Το σχεδιάγραμμα της παρούσας εργασίας είναι το εξής: Το Κεφάλαιο 2 περιγράφονται διάφορες τεχνικές Βαθιάς Μάθησης. Θα αναπτύξουμε την θεωρία, την αρχιτεκτονική και την λειτουργία ενός απλού Τεχνητού Νευρωνικού Δικτύου ( § 2.1) και στη συνέχεια θα επικεντρωθούμε σε συγκεκριμένες τεχνικές για αναγνώριση προτύπων στις εικόνες, όπως τα Συνελικτικά Νευρωνικά Δίκτυα ( § 2.2.1), τους Autoencoders ( § 2.3.1) και τους Variational Autoencoders ( § 2.3.2). Έπειτα, στο Κεφάλαιο 3, θα εστιάσουμε το ενδιαφέρον μας σε συγκεκριμένες τεχνικές που εφαρμόζονται για την αποθρομβοποίηση και την ανάδειξη χαρακτηριστικών σε εικόνες, καθώς και το απαραίτητο θεωρητικό υπόβαθρο που τις αποτελούν. Σε αντίθεση με τις μεθόδους του Κεφαλαίου 2, αυτές δεν προϋποθέτουν την χρήση μεγάλων βάσεων δεδομένων. Επιγραμματικά θα αναλύσουμε τις έννοιες των τυχαίων πεδίων Markov ( § 3.2.1), της κανονικοποίησης Tikhonov ( § 3.2.3), και της μεθόδου Μπεϋζιανής ανάλυσης εικόνας στον χώρο Fourier (Bayesian Image analysis in Fourier Space, BIFS) με την οποία θα ασχοληθούμε περισσότερο σε αυτή τη διπλωματική ( § 3.4). Φυσικά θα συμπεριλάβουμε και το απαραίτητο θεωρητικό υπόβαθρο που απαρτίζουν αυτές τις τεχνικές, όπως αυτό της Μπεϋζιανής Συμπερασματολογίας ( § 3.1) και της ανάλυσης Fourier ( § 3.3). Στο Κεφάλαιο 4, θα δώσουμε περιγραφή των δεδομένων (ιατρικές εικόνες) που θα χρησιμοποιήσουμε για την εφαρμογή ορισμένων προαναφερθέντων τεχνικών ανάλυσης εικόνας. Εν συνεχεία, στο Κεφάλαιο 5, θα παρουσιάσουμε και θα αναλύσουμε τα

αποτελέσματα που προκύπτουν έπειτα από την εφαρμογή των τεχνικών στα δεδομένα. Τέλος, στο Κεφάλαιο 6, θα συνοψίσουμε τα συμπεράσματα της παρούσας εργασίας και θα δώσουμε κάποιες περαιτέρω ιδέες για μελλοντική έρευνα πάνω στον αντικείμενο της Μπεϋζιανής ανάλυσης εικόνας.

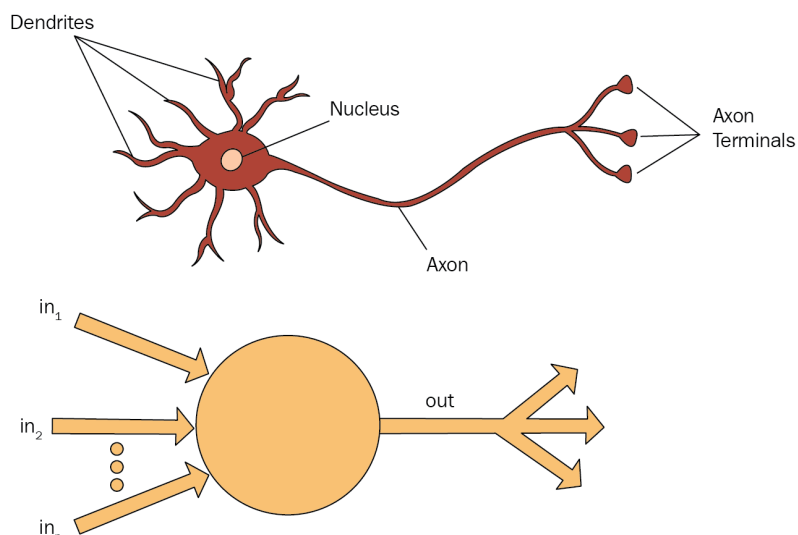


## 2. Τεχνικές Βαθιάς Μάθησης για αναγνώριση εικόνας

Η πληροφορία που περιέχει μια εικόνα είναι εξαιρετικά μεγάλη. Αν αντιληφθούμε την διαδρομή των ματιών μας βλέποντας μια εικόνα, καθώς και την άμεση ανίχνευση των αντικειμένων ή ακόμα και των συναισθημάτων που ενδέχεται να περιέχει, συνειδητοποιούμε πως είναι αρκετά δύσκολο για έναν υπολογιστή να αποκτήσει αυτές τις ικανότητες. Η επικρατέστερη ιδέα είναι να εμφυσύσουμε στον υπολογιστή τον τρόπο που οι άνθρωποι σκέφτονται. Η έννοια που θα περιγράψουμε είναι αυτή του Τεχνητού Νευρωνικού Δικτύου (Artificial Neural Network ή ANN). Η ονομασία του μοντέλου αυτού προέρχεται από τους ανθρώπινους νευρώνες του εγκεφάλου, με τους οποίους παρουσιάζουν αρκετά κοινά χαρακτηριστικά (βλ. Σχήμα 2.1). Η δομή ενός νευρώνα, του ανθρώπινου εγκεφάλου, αποτελείται από το κυτταρικό σώμα, που περιέχει τον πυρήνα, και από έναν πεπερασμένο αριθμό δενδριτών. Οι δενδρίτες τροφοδοτούν τον νευρώνα με ηλεκτρικά σήματα, και αυτός με τη σειρά του τα μεταδίδει στους άξονες. Ένα τεχνητό νευρωνικό δίκτυο αποτελείται από κόμβους (νευρώνες ή perceptrons) και ακμές. Οι κόμβοι συλλέγουν κάποια πληροφορία και μέσω μιας συνάρτησης ενεργοποίησης την στέλνουν σε άλλους κόμβους. Ένα τέτοιο δίκτυο αποτελείται από τρία βασικά επίπεδα (ή στρώματα, layers): το επίπεδο εισόδου, το κρυφό επίπεδο (μπορεί να είναι περισσότερα από ένα) και το επίπεδο εξόδου. Οι λειτουργίες των διαφόρων επιπέδων του νευρωνικού δικτύου, αλλά και ο τρόπος που αυτά επικοινωνούν, θα παρουσιαστούν αναλυτικά στη συνέχεια.

---

<sup>1</sup>Πηγή:<https://www.packtpub.com/product/mobile-artificial-intelligence-projects/9781789344073>



Σχήμα 2.1: Δομή ενός βιολογικού και ενός τεχνητού νευρώνα (perceptron) <sup>1</sup>

## 2.1 Εισαγωγή στα Τεχνητά Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα επιλύουν προβλήματα παλινδρόμησης και ταξινόμησης. Η γενική μορφή αυτών μπορεί να περιγραφεί από την εξίσωση

$$Y = \phi(X^T W). \quad (2.1)$$

Το κρυφό επίπεδο είναι ένας αφηρημένος πολυδιάστατος χώρος, ο οποίος περιέχει μεταβλητές που δεν μπορούν να ερμηνευθούν απευθείας. Ουσιαστικά, το επίπεδο αυτό αποτελεί μια διαφορετική αναπαράσταση των δεδομένων σε συμπιεσμένη μορφή. Έστω ένας νευρώνας  $j$  του κρυφού επιπέδου προκύπτει από τον γραμμικό συνδυασμό όλων των εισόδων  $x_i, i = 1, \dots, n$  (χαρακτηριστικά μιας παρατήρησης), με κάποιους συντελεστές (βάρη ή weights)  $w_{i,j}$ . Τότε, το δίκτυο ονομάζεται πλήρως συνδεδεμένο. Αναφέρουμε ότι, ο πίνακας  $W$  περιέχει τα βάρη (συντελεστές) για την συνεισφορά των νευρώνων ενός επιπέδου στο ακριβώς επόμενο. Στόχος είναι η εκπαίδευση του τεχνητού νευρωνικού δικτύου, ώστε να μάθει τον πίνακα βαρών  $W$ . Αρχικά στον πίνακα αυτό εισάγονται κάποιες αρχικές τυχαίες τιμές. Ο τρόπος εκμάθησης του παρουσιάζεται παρακάτω. Η συνάρτηση  $\phi$  ονομάζεται συνάρτηση ενεργοποίησης (activation function), και συνήθως στην θέση της χρησιμοποιούμε την λογιστική συνάρτηση. Υπάρχουν και άλλες δυνατές επιλογές όπως η συνάρτηση της εφαπτομένης και η ReLU (Rectified Linear Unit). Μπορούμε να μετατοπίσουμε την συνάρτηση ενεργοποίησης προσθέτοντας στο μοντέλο της

<sup>1</sup>Πηγή:<https://www.packtpub.com/product/mobile-artificial-intelligence-projects/9781789344073>

σχέσης (2.1) ένα συντελεστή μεροληψίας, που θα τον συμβολίζουμε με  $b$ . Η μετατόπιση αυτή επηρεάζει το αποτέλεσμα του επιπέδου εξόδου συμβάλλοντας στην καλύτερη προσαρμογή του μοντέλου στα δεδομένα. Επιπρόσθετα, με την παρουσία ενός συντελεστή μεροληψίας  $b$ , οι στήλες του πίνακα  $W$  που περιέχει τα βάρη για όλους τους νευρώνες ενός κρυφού επιπέδου, αυξάνονται κατά μια. Με την προσθήκη του παραπάνω συντελεστή, ουσιαστικά προσθέτουμε μια ακόμα τυχαία μεταβλητή στο μοντέλο μας όπου το βάρος της είναι ίσο με τη μονάδα. Τέλος, κανονικοποιώντας τις τιμές που πήραμε από την προαναφερθείσα συνάρτηση παίρνουμε ένα διάνυσμα μήκους  $K$  και οι δυνατές κατηγορίες (εκτελούμε συνάρτηση softmax). Έτσι, οι τιμές του διανύσματος ανήκουν στο σύνολο  $[0, 1]$  και εκφράζουν τις πιθανότητες, η παρατήρηση να προέρχεται από κάθε ετικέτα (βλ. Σχήμα 2.2). Στόχος της συνάρτησης αυτής, είναι η τοποθέτηση της παρατήρησης που πήραμε ως είσοδο σε μια από τις προκαθορισμένες κατηγορίες. Η συνάρτηση softmax που χρησιμοποιούμε, σε συνδιασμό με την λογιστική συνάρτηση ως συνάρτηση ενεργοποίησης ( $\phi$ ), για την εύρεση της πιθανότητας η μεταβλητή απόκρισης  $Y$  να είναι ίση με  $k$  (σε προβλήματα ταξινόμησης  $k \in \mathbb{Z}$ ) είναι η εξής

$$P(Y = k|X; W) = \frac{\exp(X^T W_k)}{\sum_{j=1}^K \exp(X^T W_j)}. \quad (2.2)$$

Εύκολα παρατηρούμε ότι το αποτέλεσμα τις παραπάνω ποσότητας είναι ένα διάνυσμα με  $K$  στοιχεία.

Αξίζει να αναφέρουμε ότι, εάν τα ενδιάμεσα επίπεδα του δικτύου είναι περισσότερα από ένα, τότε το νευρωνικό δίκτυο ονομάζεται πολυεπίπεδο (Multilayer Perceptron ή MLP). Σε αυτή την περίπτωση η διαδικασία μπορεί να εκφραστεί από την ακόλουθη σχέση

$$Y = \phi(W^{(N)})\phi(W^{(N-1)}) \dots \phi(W^{(0)} + b^{(0)}) \dots b^{(N-1)} + b^{(N)}, \quad (2.3)$$

όπου  $N$  είναι το τελευταίο επίπεδο του νευρωνικού δικτύου. Με την μεταβλητή  $n$  αναφερόμαστε μόνο στα ενδιάμεσα επίπεδα, εξαιρώντας το επίπεδο εισόδου και εξόδου. Για παράδειγμα με  $N = 1$  αναφερόμαστε στην έξοδο του πρώτου επιπέδου, με  $N = 2$  στον δεύτερο κ.ο.κ. Επιπλέον, υπάρχει περίπτωση να χρησιμοποιούμε διαφορετικές συναρτήσεις ενεργοποίησης σε κάθε επίπεδο του νευρωνικού δικτύου. Η ανάμειξη τους πρέπει να γίνεται με προσοχή, ώστε το αποτέλεσμα του επιπέδου εξόδου να έχει νόημα. Για λόγους απλότητας εδώ θεωρούμε πως εφαρμόζουμε την ίδια συνάρτηση ενεργοποίησης  $\phi$ . Για την εκπαίδευση του νευρωνικού δικτύου

πρέπει να γνωρίζουμε εκ-των-προτέρων τις πραγματικές αποκρίσεις  $Y$ . Θα δίνονται σε ένα διάνυσμα με την τιμή 1 στην θέση της κατηγορίας που όντως ανήκει η παρατήρηση που βάλαμε ως είσοδο, και 0 στα υπόλοιπα στοιχεία. Εδώ έγγυται η χρήση του  $k$  ως ακέραιο, αφού κάθε  $k$  ισοδυναμεί με μια κατηγορία. Τώρα, συγκρίνουμε τις πιθανότητες που προκύπτουν, από την διαδικασία που περιγράψαμε, με το one-hot διάνυσμα. Άμεση απόρροια είναι η σύγκριση της πρόβλεψης  $o^N$  που έκανε το δίκτυο, με την πραγματική απόκριση  $Y$ , μέσω μιας συνάρτησης κόστους. Μια ενδεικτική τέτοια συνάρτηση είναι το μέσο τετραγωνικό σφάλμα (MSE) (βλ. Ενότητα ( § 2.3.1)). Βέβαια, υπάρχουν και άλλες, μεταξύ των οποίων χρησιμοποιείται περισσότερο, η διασταυρωμένη εντροπία (Cross Entropy Loss). Η τελευταία είναι της μορφής

$$\mathcal{L} = - \sum_{j=1}^K Y_j \log(o_j^{(n)}), \quad (2.4)$$

όπου  $j$  είναι ο  $j$ -οστός νευρώνας στο  $n$ -οστό κρυφό επίπεδο. Προφανώς, η διάσταση του διανύσματος της πραγματικής απόκρισης  $Y$  και της πρόβλεψης  $o^{(n)}$  είναι ίσες με  $K$ .

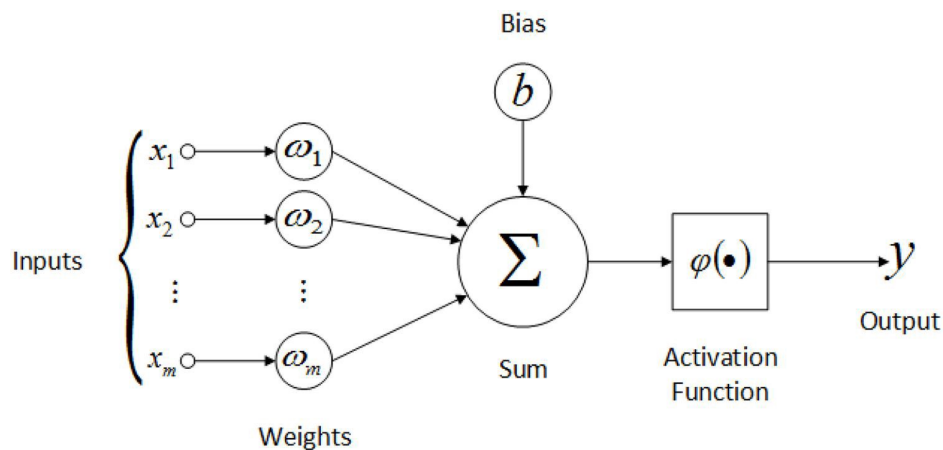
Στη συνέχεια εκτελώντας μια διαδικασία που ονομάζεται ανάστροφη μετάδοση (backpropagation), ενημερώνουμε τον πίνακα βαρών  $W$ . Συνήθως, για την ενημέρωση των βαρών χρησιμοποιούμε τον αλγόριθμο στοχαστικής κατάβασης κλίσης (ή Stochastic Gradient Descent). Εάν έχουμε μεγάλο όγκο δεδομένων ενδείκνυται η στοχαστική κατάβαση κλίσης σε μικρές υποομάδες (mini-batches). Αν αναλογιστούμε τις πράξεις πινάκων που είναι απαραίτητες για την εφαρμογή ενός νευρωνικού δικτύου, εύκολα συμπεραίνουμε πως εάν έχουμε πολλά δεδομένα, ο αλγόριθμος θα αργήσει να εφαρμοστεί. Σε αυτή την περίπτωση ενημερώνουμε ένα μικρό μέρος (mini-batch) του πίνακα, αντί για ολόκληρο. Έτσι, η υπολογιστική πολυπλοκότητα, όπως και η ταχύτητα εφαρμογής του αλγορίθμου μειώνονται σημαντικά. Ο πίνακας  $W$  ενημερώνεται με βάση τον παρακάτω κανόνα

$$W \leftarrow W - \eta \nabla_W \mathcal{L}, \quad (2.5)$$

όπου  $\eta$  το βήμα μάθησης (learning rate) και  $\mathcal{L}$  η συνάρτηση κόστους. Με το σύμβολο  $\nabla$  (ή ανάδελτα) συμβολίζουμε τον διανυσματικό διαφορικό τελεστή μερικών παραγώγων της συνάρτησης  $\mathcal{L}$ . Ο αλγορίθμος στοχαστικής κατάβασης κλίσης είναι ισοδύναμος με το πρόβλημα ελαχιστοποίησης της συνάρτησης κόστους  $\mathcal{L}$ . Στόχος του αλγορίθμου στοχαστικής κατάβασης κλίσης είναι να βρεθεί εκείνος ο πίνακας  $W$  που ελαχιστοποιεί την συνάρτηση κόστους. Για την



αποφυγή υπερπροσαρμογής του μοντέλου (όπου το νευρωνικό δίκτυο μαθαίνει ακριβώς τα δεδομένα εισόδου), συνηθίζεται η χρήση ενός όρου ποινικοποίησης (περισσότερα για όρους ποινικοποίησης θα δούμε στην Ενότητα § 3.2.3). Ο αλγόριθμος επαναλαμβάνεται για έναν συγκεκριμένο αριθμό επαναλήψεων ή εποχών (epochs). Τόσο η παράμετρος  $\eta$  όσο και ο αριθμός εποχών είναι υπερπαράμετροι (καθορίζουν την ταχύτητα και την ποιότητα του στάδιου μάθησης). Επίσης, σημαντικού υπερπαράμετροι είναι ο αριθμός των κρυφών επιπέδων και ο αριθμός νευρώνων σε κάθε επίπεδο. Για την βέλτιστη εκτίμηση αυτών κρίνεται αναγκαία η χρήση διαδικασιών διασταυρωμένης επικύρωσης (Cross Validation). Σκοπός του αλγορίθμου που περιγράψαμε είναι η εύρεση του πίνακα  $W$ , μέσω της μεθόδου ελαχιστοποίησης του  $\mathcal{L}$ . Έτσι, για μια νέα παρατήρηση που εισάγεται στο νευρωνικό δίκτυο (που δεν έχει εκπαιδευθεί έως τώρα ο αλγόριθμος σε αυτή), να προβλέπεται σωστά η κατηγορία που ανήκει.



Σχήμα 2.2: Αρχιτεκτονική δομή ενός Τεχνητού Νευρωνικού Δικτύου<sup>2</sup>

Επόμενο βήμα είναι να δούμε πώς μπορούμε να εισάγουμε μια εικόνα στην αρχιτεκτονική δομή που αναφέραμε. Είναι γνωστό ότι μια εικόνα στην οθόνη ενός υπολογιστή αναπαριστάτε από ενωμένα εικονοκύτταρα ή πίξελ (pixels), που παίρνουν την μορφή ψηφιδωτού. Ας υποθέσουμε, χάρην απλότητας, μια ασπρόμαυρη εικόνα. Κάθε πίξελ αυτής, μπορούμε να το περιγράψουμε με έναν αριθμό από το μηδέν ως το ένα, ανάλογα με το χρώμα του. Δίνουμε στο μαύρο χρώμα την τιμή 0, στο άσπρο την τιμή 1 και στις διάφορες αποχρώσεις του γκρι τις ενδιάμεσες τιμές, ανάλογα με την ένταση τους. Άρα, μια εικόνα μπορεί πλέον να αναπαρασταθεί ως ένας πίνακας που το κάθε στοιχείο του περιέχει τιμές στο διάστημα  $[0, 1]$ . Στην περίπτωση μιας έγχρωμης εικόνας η διάσταση του προβλήματος αυξάνεται. Πλέον, υπάρχουν τρία κανάλια (channels) για τα ισάριθμα κύρια χρώματα: το κόκκινο, το πράσινο και το μπλέ (Red, Green, Blue, RGB). Συνεπώς για κάθε ένα κανάλι αντιστοιχεί και ένας πίνακας όπως αυτός που περιγράψαμε. Όμως τώρα οι

<sup>2</sup> Πηγή: <https://ryash1921.medium.com/introduction-neural-networks-single-layer-neural-network-b80ddaa943c7>

αποχρώσεις των χρωμάτων αλλάζουν σε κάθε κανάλι. Μια ιδέα για την εισαγωγή εικόνας σε ένα νευρωνικό δίκτυο είναι η επιπεδοποίηση της (flatten). Δηλαδή, ο μετασχηματισμός της εικόνας σε ένα διάνυσμα μιας διάστασης, έχοντας ως στοιχεία  $x_i$  τις αντίστοιχες τιμές σε όλα τα πίξελ αυτής. Το μειονέκτημα αυτής της μεθόδου είναι πως χάνεται σημαντική πληροφορία του χώρου, και ο αλγόριθμος αναγκάζεται να την ανακτήσει. Στην επόμενη ενότητα θα δούμε πώς μπορούμε να εξαλείψουμε το πρόβλημα αυτό με την εισαγωγή ενός νέου είδους νευρωνικού δικτύου.

## 2.2 Ταξινόμηση εικόνας και ανίχνευση αντικειμένων

Μια σημαντική δυνατότητα που θέλουμε να έχει ο υπολογιστής είναι η ταξινόμηση μιας εικόνας σε ένα σύνολο από ετικέτες. Για παράδειγμα, τροφοδοτώντας ένα νευρωνικό δίκτυο με μια εικόνα ενός αυτοκινήτου, θέλουμε αυτό να προβλέπει σωστά πως όντως απεικονίζεται ένα αυτοκίνητο. Για να γίνει αυτό, αναγκαία χρίζεται η εκπαίδευση του σε πολλές εικόνες, ώστε να ξεχωρίζει τα αυτοκίνητα από άλλα αντικείμενα σε μια φωτογραφία. Όπως γίνεται αντιληπτό, το πρόβλημα αυτό ανήκει στην κατηγορία μάθησης με επίβλεψη (supervised learning), αφού οι ετικέτες δίνονται ταυτόχρονα με τις εικόνες κατά την εκπαίδευση. Ως βάση θα χρησιμοποιήσουμε την δομή των ANN, που αναπτύξαμε προηγουμένως. Επίσης, αρκετές φορές θέλουμε το δίκτυο να βρίσκει τη θέση ορισμένων αντικειμένων στην εικόνα, και έμμεσα να τα κατηγοριοποιεί.

### 2.2.1 Συνελικτικά Νευρωνικά Δίκτυα

Αρχικά, παρουσιάζουμε το θεμελιώδες νευρωνικό δίκτυο για την ταξινόμηση εικόνων, ανάλογα με το περιεχόμενό τους. Για παράδειγμα, έχουμε αρκετές φωτογραφίες που απεικονίζουν είτε αυτόκινητα είτε ανθρώπους και σκοπός μας είναι η εκπαίδευση ενός νευρωνικού δικτύου να αναγνωρίζει το περιεχόμενό τους. Ο τρόπος εκπαίδευσης για την λειτουργία αυτή γίνεται μέσω Συνελικτικών Νευρωνικών Δικτύων (Convolutional Neural Networks ή CNN). Οι περισσότεροι μεταγενέστεροι αλγόριθμοι μεθόδους ταξινόμησης μιας εικόνας βασίζονται στην αρχιτεκτονική και τον τρόπο λειτουργίας των εν λόγω δικτύων.

Η πληροφορία του χώρου μιας εικόνας διαδραματίζει πολύ σημαντικό ρόλο, όπως επισημάναμε νωρίτερα. Γειτονικά πίξελ ενδέχεται να είναι ισχυρά συσχετισμένα μεταξύ τους. Προτείναμε να εισάγουμε στο πρώτο επιπέδο του νευρωνικού δικτύου την επιπεδοποιημένη εικόνα, και να τη συνδέσουμε πλήρως με του νευρώνες του επόμενου κρυφού επιπέδου. Αντί αυτού, θα

τροφοδοτούμε σε έναν νευρώνα  $j$  μόνο εκείνα τα  $x_i$  που αποτελούν ένα μικρότερο στιγμιότυπο της συνολικής εικόνας. Συνήθως ξεκινάμε παίρνοντας τα στιγμιότυπα από την αριστερή πάνω γωνία της εικόνας και με κάποιο βήμα  $k$  μετατοπίζουμε το τελευταίο τόσα πίξελ και παίρνουμε το επόμενο (τόσο στον οριζόντιο όσο και στον κάθετο άξονα). Τα βήματα μετάβασης από ένα στιγμιότυπο σε ένα άλλο, ενδέχεται να διαφέρουν ως προς τον άξονα. Ο κάθε νευρώνας απαρτίζεται από τα πίξελ του εκάστοτε μέρους της εικόνας. Η διαδικασία ολοκληρώνεται έως ότου να έχει ληφθεί υπόψη η ολόκληρη η εικόνα.

Ένα από τα κυριότερα κομμάτια στην εφαρμογή των CNN, είναι τα φίλτρα (filters) ή πυρήνες (kernels). Αν υποθέσουμε, πως γνωρίζουμε εκ-των-προτέρων κάποια από τα γεωμετρικά στοιχεία του αντικειμένου που θέλουμε να αναγνωρίσουμε, μπορούμε να τα λάβουμε υπόψη μέσω των φίλτρων. Αποτελούν πίνακες μικρότερης διάστασης από αυτή της εικόνας, και αναφέρονται στην γνώση που έχουμε για την αναπαράσταση του ειδώλου που αναζητάμε. Στιγμιότυπα ονομάζουμε μικρότερα κομμάτια της εικόνας. Υπογραμμίζουμε πως τα στοιχεία του φίλτρου μπορούν να είναι τυχαία επιλεγμένα, αλλά να διατηρούν τα γεωμετρικά στοιχεία του προς εξέταση αντικειμένου. Έχοντας πλέον διαχωρίσει την εικόνα σε στιγμιότυπα ψάχνουμε να βρούμε κάποια συγκεκριμένα χαρακτηριστικά του αντικειμένου αυτού. Η διάσταση των φίλτρων τις περισσότερες φορές είναι ίση με αυτή των στιγμιότυπων. Θέλουμε με κάποιον τρόπο να ελέγχουμε, διατρέχοντας το φίλτρο σε όλη την εικόνα, εάν το συγκεκριμένο αντικείμενο του φίλτρου υπάρχει στην καθολική εικόνα. Έτσι, εφαρμόζοντας το γινόμενο Hadamard (δηλαδή πολλαπλασιασμό ανά στοιχείο) μεταξύ των δύο πινάκων (του στιγμιότυπου και του φίλτρου) και έπειτα αθροίζοντας τα στοιχεία του τελικού προκύπτει ένας φυσικός αριθμός. Έστω  $S_i$  το  $i$ -οστό στιγμιότυπο της εικόνας και  $F$  ο πίνακας του φίλτρου. Οι διαστάσεις των δύο πινάκων είναι ίσες με  $h \times v$ . Τότε, το  $i$ -οστό αποτέλεσμα μετά την εφαρμογή του φίλτρου θα είναι ίσο με

$$r_i = \sum_{n=1}^h \sum_{m=1}^v \left( S_i \odot F \right)_{n,m}, \quad (2.6)$$

όπου με  $\odot$  συμβολίζουμε το γινόμενο Hadamard. Αποθηκεύουμε κάθε τέτοιο αριθμό  $r_i$  που προκύπτει καθόλη την διάρκεια ελέγχου του φίλτρου πάνω στην εικόνα, σε έναν νέο πίνακα. Διαπιστώνουμε ότι ο τελικός πίνακας είναι μικρότερης διάστασης από την αρχική εικόνα. Ανάλογα με το φίλτρο που έχουμε επιλέξει, αφού μετασχηματίσουμε την αρχική εικόνα, μπορούμε να εξάγουμε από αυτές πολύ σημαντικές πληροφορίες. Στην περίπτωση που θέλουμε να εξετάσουμε ένα σύνολο από φίλτρα, τότε οι πίνακες τους ενώνονται σε έναν, διάστασης

$h \times v \times f$ , όπου  $h \times v$  η διάσταση του φίλτρου και  $f$  το πλήθος τους. Φυσικά η ταυτόχρονη διενέργεια φίλτρων αυξάνει την πολυπλοκότητα του αλγορίθμου. Υπογραμμίζουμε πως το επίπεδο του νευρωνικού δικτύου που προκύπτει από την εφαρμογή των φίλτρων, ονομάζεται επίπεδο συνέλιξης (convolutional layer). Πρόκειται ουσιαστικά για την διαδικασία που το νευρωνικό δίκτυο που εξετάζουμε οφείλει το όνομά του.

Επιπρόσθετα, θέλουμε να αναδείξουμε την στενή σχέση της συνάρτησης ενεργοποίησης ReLU που αναφέραμε στην αρχή του κεφαλαίου, με τα CNN και τα φίλτρα [Kuo \(2016\)](#). Η συνάρτηση είναι η εξής

$$ReLU(x) = \begin{cases} x, & \text{αν } x > 0 \\ 0, & \text{αν } x \leq 0 \end{cases}. \quad (2.7)$$

Εφαρμόζεται κυρίως μετά την εφαρμογή των φίλτρων. Οι νευρώνες του κρυφού επιπέδου μετά την ομαδοποίηση, απεικονίζουν μια πιο αυθαίρετη αναπαράσταση της εικόνας που έχει εισαχθεί αρχικά στο νευρωνικό δίκτυο. Προφανώς, δεν είναι γνωστό στο υπολογιστή ποιές αναπαραστάσεις είναι πληροφοριακές, και κατά συνέπεια ποίοι νευρώνες είναι χρήσιμοι. Αναφέρουμε και πάλι πως οι αναπαραστάσεις είναι ανεξάρτητες μεταξύ τους. Η ανεξάρτησία διατηρείται και στους νευρώνες του κρυφού επιπέδου που μας ενδιαφέρει. Κάποιοι νευρώνες ενδέχεται να είναι ασήμαντοι, ακόμα και παραπλανητικοί, σε σχέση με την ταξινόμηση που θέλουμε να υλοποιήσουμε. Διαδεδομένες συναρτήσεις ενεργοποίησης όπως η λογιστική ή σιγμοειδής (logistic ή sigmoid function), όπως περιγράφεται στην εξίσωση (2.8), και η εφαπτομένη (tangent function), η οποία δίνεται στη σχέση (2.9), μπορούν να εφαρμοστούν στο εν λόγω σημείο, όμως όχι τόσο αποτελεσματικά. Οι δύο τελευταίες συναρτήσεις ορίζονται ως

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (2.8)$$

$$\tan(x) = \frac{\sin(x)}{\cos(x)}, \quad (2.9)$$

όπου  $x \in \mathbb{R}$ . Με την εφαρμογή των συναρτήσεων αυτών στους συγκεκριμένους κόμβους προκύπτει ένας αρνητικός αριθμός κοντά στο μηδέν, αλλά ποτέ ίσος με μηδέν. Αντίθετα, η

συνάρτηση ενεργοποίησης ReLU μηδενίζει οποιοδήποτε αρνητική τιμή που εμφανίζεται στους νευρώνες. Ουσιαστικά, απενεργοποιεί τους κόμβους που λειτουργούν ανασταλτικά στην διαδικασία εκπαίδευσης του νευρωνικού δικτύου. Σημειώνουμε πως έχουν αναπτυχθεί και διάφορες παραλλαγές της, οι οποίες επιτρέπουν αρνητικές τιμές πολύ κοντά στο μηδέν. Οι τελευταίες δηλαδή, δέχονται την λειτουργία νευρώνων που οι αναπαραστάσεις τους ίσως περιέχουν μη σημαντική πληροφορία. Πιο συγκεκριμένα ο κλάδος της σχέσης (2.7) για  $x \leq 0$ , δέχεται κάποια παραμετροποίηση. Για περισσότερες λεπτομέρειες συστήνουμε στον αναγνώστη την δημοσίευση των [He et al. \(2015\)](#)

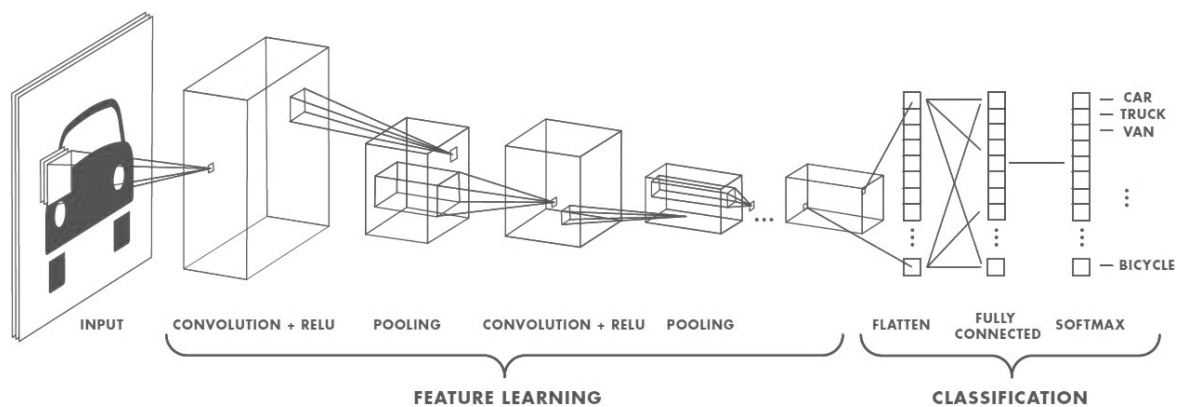
Στη συνέχεια θα εισάγουμε την έννοια της ομαδοποίησης (pooling) και θα εξηγήσουμε τον λόγο για τον οποίο είναι χρήσιμη. Έχοντας την πραγματική εικόνα θέλουμε να μειώσουμε την διάσταση της, παράλληλα όμως διατηρώντας την πληροφορία την πληροφορία των πίξελ που την δημιουργούν. Ο λόγος αυτής της ενέργειας έγγειται στην ευκολότερη διαχείριση προβλημάτων μικρότερης διάστασης, τόσο υπολογιστικά όσο και ερμηνευτικά (η ερμηνεία δεν αποτελεί κύριο ζήτημα στην συγκεκριμένη περίπτωση). Υπάρχουν δύο διαδεδομένες μέθοδοι ομαδοποίησης των πίξελ, η average-pooling και η max-pooling. Οι μέθοδοι αυτοί, αντί να παίρνουν αυτούσιους τους πίνακες που προκύπτουν μετά από την εφαρμογή των φίλτρων, παίρνουν την μέση και την μεγαλύτερη τιμή αυτών των πινάκων αντίστοιχα. Έτσι, εκτός από την μείωση της υπολογιστικής πολυπλοκότητας, εξάγουμε τα κυρίαρχα χαρακτηριστικά μιας εικόνας, διατηρώντας την αποτελεσματικότητα εκπαίδευσης του μοντέλου. Επίσης, η μέθοδος max-pooling λειτουργεί και ως κατασταλτικό θορύβου, σε αντίθεση με την average-pooling που δεν έχει αυτή την δυνατότητα. Συνεπώς, μπορούμε να διαπιστώσουμε την χρησιμότητα της πρώτης μεθόδου και τον λόγο που χρησιμοποιείται στα Συνελικτικά Νευρωνικά Δίκτυα.

Ο συνδιασμός των μεθόδων που εξηγήσαμε και η μετέπειτα χρήση ενός ANN, μας οδηγεί στην ταξινόμηση της εικόνας σε μια από τις κατηγορίες που εξετάζουμε. Γνωρίζουμε ότι τα ANN παίρνουν ως είσοδο χαρακτηριστικά  $x_i$  και όχι πίνακες. Με την επανάληψη των μεθόδων που αναλύσαμε, μπορούμε να μειώσουμε ακόμα περισσότερο την διάσταση της εικόνας, επιτυγχώντας γρηγορότερη υπολογιστική ικανότητα. Εύκολα διαπιστώνουμε πως ο τελικός πίνακας που θα προκύψει θα έχει μικρότερη διάσταση από την αρχική, συγκεντρώνοντας παράλληλα την σημαντικότερη πληροφορία της. Τελικά, επιπεδοποιούμε τον εν λόγω πίνακα που προκύπτει και τον εισάγουμε σε ένα Τεχνητό Νευρωνικό Δίκτυο. Η μέθοδος ταξόμησης που ακολουθείται στη συνέχεια είναι ίδια με αυτή που περιγράψαμε στην πρώτη ενότητα του

κεφαλαίου (Ενότητα § 2.1).

Οι υπερπαραμέτροι του προβλήματος είναι αρκετές. Η διάσταση του φίλτρου που θα χρησιμοποιήσουμε, η διάσταση του τελικού πίνακα που θα προκύψει, ο αριθμός των νευρώνων στο κρυφό επίπεδο του ANN καθώς και το πλήθος των κρυφών επιπέδων αποτελούν υπερπαραμέτρους που καλούμαστε να καθορίσουμε. Επίσης, σημαντικές υπερπαραμέτροι είναι το βήμα μάθησης και ο αριθμός των εποχών για τον αλγόριθμο κατάβασης κλίσης στην εφαρμογή του ANN. Η επιλογή τους γίνεται συνήθως με μεθόδους διασταυρωμένης επικύρωσης, όπως ήδη έχουμε εξηγήσει, και αποτελούν μέρος της εκπαίδευσης του Συνελκτικού Νευρωνικού Δικτύου.

Ας υποθέσουμε πως θέλουμε ο υπολογιστής λαμβάνοντας ως είσοδο μια εικόνα να αναγνωρίζει τι είδους όχημα περιέχεται σε αυτή [Jain et al. \(2019\)](#). Έστω ότι η εικόνα που παίρνει ως είσοδο είναι ένα αυτοκίνητο και οι κατηγορίες είναι μεταξύ άλλων: αυτοκίνητο, φορτηγό, ποδήλατο κ.ο.κ. Η διαδικασία που ακολουθείται περιγράφεται κατατοπιστικά στο Σχήμα 2.3. Το σχήμα αυτό αποτελεί μια πολύ καλή αναπαράσταση της λειτουργίας καθώς και της σειράς εκτέλεσης των μεθόδων που περιγράψαμε.



Σχήμα 2.3: Αρχιτεκτονική δομή ενός Συνελκτικού Νευρωνικού Δικτύου (CNN)<sup>3</sup>.

## 2.3 Ανάλυση εικόνας και ανοικοδόμηση

Στην ενότητα αυτή θα ασχοληθούμε με τεχνικές ανάλυσης και ανοικοδόμησης εικόνας. Αποσκοπούμε στην εύρεση της υποκείμενης δομής και των χαρακτηριστικών μιας εικόνας. Πιο

<sup>3</sup>Πηγή: [Jain et al. \(2019\)](#)

συγκεκριμένα στην κατανομή που την γεννάει. Σε αντίθεση με τις μεθόδους των νευρωνικών δικτύων που εξετάσαμε, τώρα δεν θα μας δίνονται οι ετικέτες, αλλά μόνο τα δεδομένα (εικόνες). Πρόκειται για μια νέα κλάση εκμάθησης προβλημάτων, τη λεγόμενη μη επιβλεπόμενη μάθηση. Θα επικεντρωθούμε κυρίως σε αλγόριθμους γεννητικής μοντελοποίησης. Στόχος αυτής είναι η εκμάθηση ενός μοντέλου που αντιπροσωπεύει την κατανομή των εικόνων που παίρνονται ως είσοδοι. Ας συμβολίζουμε την κατανομή των δεδομένων εκπαίδευσης και την κατανομή του παραγόμενου μοντέλου ως  $P_{data}$  και  $P_{model}$  αντίστοιχα. Η βασική ιδέα για την επίλυση του προβλήματος είναι ο τρόπος εκπαίδευσης της  $P_{data}$  ώστε να μοιάζει με την  $P_{model}$ . Οι γεννητικοί αλγόριθμοι αποτελούν το κλειδί για την αυτόματη εκμάθηση της κατανομής του κάδρου μιας εικόνας. Ο λόγος που εξετάζουμε τέτοιου είδους αλγόριθμους είναι η χρήση της πληροφορίας που προκύπτει για την δημιουργία ενός αμερόληπτου συνόλου δεδομένων. Έτσι, το μοντέλο που θα καταλήξουμε να είναι και αυτό αμερόληπτο. Σε κάποια σύνολα δεδομένων ενδέχεται οι παρατηρήσεις να μοιάζουν μεταξύ τους. Για παράδειγμα, είτε ο τόνος των χρωμάτων να είναι παρόμοιος είτε το αντικείμενο ενδιαφέροντος να παρουσιάζεται από συγκεκριμένη οπτική γωνία. Εύκολα αντιλαμβάνεται κανείς πως με τέτοιου τύπου δεδομένα θα καταφέρουμε να μάθουμε εύκολα την κατανομή  $P_{model}$  σε ένα μικρό φάσμα από δεδομένα. Δηλαδή, η ανάλυση μας θα βασίζεται σε μια συγκεκριμένη περιοχή όλων των δυνατών παραδειγμάτων που θα μπορούσαμε να έχουμε. Σε αντίθεση με εικόνες αρκετά διαφορετικές μεταξύ τους από άποψη αντίθεσης, φωτεινότητας, σκιών, ακόμα και οπτικής γωνίας, η πολυπλοκότητα της κατανομής αυτής αυξάνεται. Και στις δύο αυτές περιπτώσεις μπορούμε να αντιμετωπίσουμε τα αντίστοιχα προβλήματα με την βοήθεια ορισμένων αλγορίθμων. Επίσης, ένα άλλο παράδειγμα χρήσης του μοντέλων αυτών είναι η ανίχνευση ανωμαλιών (anomaly detection) σε μια εικόνα. Σε αυτή την περίπτωση θέλουμε να λαμβάνουμε υπόψη στο μοντέλο, οποιαδήποτε ασυνήθιστη αλλαγή στις εικόνες. Με αυτόν τον τρόπο θα αποφεύγουμε απρόβλεπτες συμπεριφορές κατά την διάρκεια της εκπαίδευσης. Αρχικά λοιπόν, θα εισάγουμε την έννοια του αυτόματου κωδικοποιητή (Autoencoder), καθώς και τον τρόπο λειτουργίας του. Στη συνέχεια θα αναλύσουμε εις βάθος μια πολύ σημαντική παραλλαγή του, τους Variational Autoencoders (VAEs). Για την πληρέστερη κατανόηση των τελευταίων, θα εισάγουμε κάποιες θεμελιώδεις έννοιες, όπως την εντροπία και την απόκλιση Kullback-Leibler.

### 2.3.1 Αυτόματοι κωδικοποιητές

Οι αυτόματοι κωδικοποιητές (Autoencoders) αποτελούν ένα είδος τεχνητού νευρωνικού δικτύου που μαθαίνει αποτελεσματικά μοτίβα σε δεδομένα. Ουσιαστικά, αποβλέπουν σε μια μικρότερης

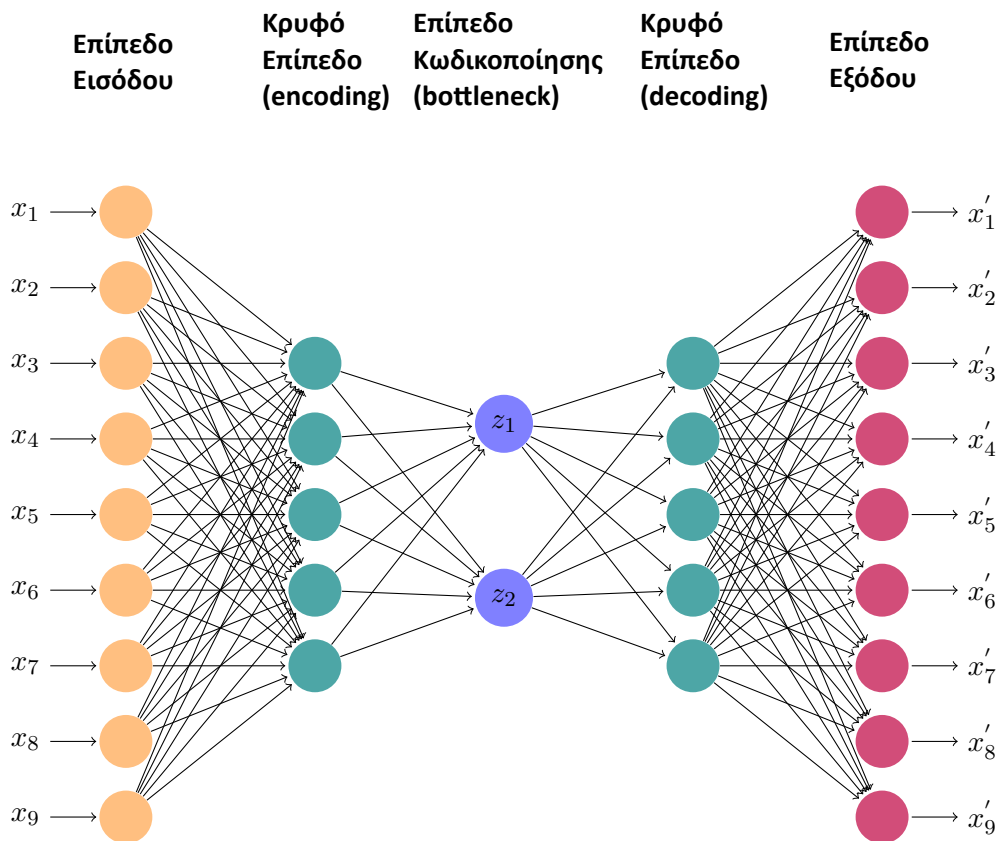
διάστασης αναπαράσταση των αρχικών ανεπεξέργαστων δεδομένων.

Η αρχιτεκτονική του δομή χωρίζεται σε δύο μέρη. Τον κωδικοποιητή (Encoder) και τον αποκωδικοποιητή (Decoder). Θεωρούμε τον κωδικοποιητή ως μια συνάρτηση  $\phi$  που απεικονίζει τα ακατέργαστα δεδομένα  $\mathcal{X}$ , σε έναν κρυφό χώρο (latent space)  $\mathcal{F}$  μικρότερης διάστασης. Η συνάρτηση του αποκωδικοποιητή απεικονίζει τον προηγούμενο χώρο  $\mathcal{F}$  σε μια ανοικοδομημένη αναπαράσταση των δεδομένων  $\mathcal{X}$ . Στην πραγματικότητα, πρόκειται για ένα τεχνητό νευρωνικό δίκτυο, όπου τα επίπεδα εισόδου και εξόδου αποτελούνται από τον ίδιο αριθμό κόμβων. Μεταξύ αυτών, παρεμβάλλονται ένα ή περισσότερα κρυφά επίπεδα. Στόχος είναι η ανακατασκευή των δεδομένων εισόδου, ελαχιστοποιώντας την διαφορά μεταξύ των τελευταίων με τα δεδομένα εξόδου. Στην παρούσα εργασία, τα δεδομένα εισόδου και εξόδου θα είναι οι ανεπεξέργαστες (RAW images) και ανοικοδομημένες εικόνες αντίστοιχα. Θα συμβολίζουμε με  $x$  τις εικόνες εισόδου και με  $x'$  τις εικόνες εξόδου. Προφανώς, τα  $x, x'$  θα είναι οι επιτεδοποιημένοι πίνακες των εικόνων και άρα διανύσματα ίδιας διάστασης. Επίσης, είναι σημαντικό να αναφέρουμε ότι, πριν την αποκωδικοποίηση υπάρχει ένα κρυφό επίπεδο  $z$ , το οποίο περιέχει την συμπίεση της εικόνας εισόδου σε μια πολύ μικρότερης διάστασης.

Ας υποθέσουμε ότι έχουμε μια ασπρόμαυρη εικόνα διάστασης  $3 \times 3 \times 1$ . Τότε η επιτεδοποιημένη μορφή της θα είναι η  $x = (x_1, x_2, \dots, x_9)^T$ . Η συμπιεσμένη εικόνα στο μεσαίο κρυφό επίπεδο δίνεται ως  $z = (z_1, z_2)^T$  και έχει διάσταση μικρότερη της αρχικής (όχι απαραίτητα διάνυσμα δύο στοιχείων όπως στο Σχήμα 2.4, στην Αγγλική βιβλιογραφία αναφέρεται ως bottleneck). Αντίστοιχα, η έξοδος του δικτύου θα είναι μια ανακατασκευασμένη εικόνα ίδιας διάστασης με την αρχική, η οποία θα είναι της μορφής  $x' = (x'_1, x'_2, \dots, x'_9)^T$ . Μια χρήσιμη και αντιπροσωπευτική αναπαράσταση της αρχιτεκτονικής των αυτόματων κωδικοποιητών δίνεται στο Σχήμα 2.4. Ακολούθως θα εμβαθύνουμε στις λειτουργίες των κωδικοποιητών και των αποκωδικοποιητών.

Ένας κωδικοποιητής αποσκοπεί στην εύρεση και αποθήκευση μιας όσο το δυνατόν μικρότερης αναπαράστασης μιας εικόνας εισόδου, ώστε στη συνέχεια να την μεταφέρει στον αποκωδικοποιητή. Αντί λοιπόν να αποστηθίσει την αρχική εικόνα, προσπαθεί να την συμπίεσει σε μικρότερη μορφή κρατώντας τα κύρια χαρακτηριστικά της. Η συμπιεσμένη εικόνα που προκύπτει ενδέχεται να μην είναι πληροφοριακή για τον άνθρωπο, αλλά αποτελεί εξαιρετικό εργαλείο για τον αποκωδικοποιητή.





Σχήμα 2.4: Αρχιτεκτονική δομή ενός αυτόματου κωδικοποιητή (Autoencoder)

Ο αποκωδικοποιητής δουλεύει πανομοιότυπα με τον κωδικοποιητή αλλά αντίστροφα. Πιο συγκεκριμένα, μαθαίνει να διαβάζει την συμπιεσμένη αναπαράσταση της εικόνας και στη συνέχεια να γεννάει μια νέα βασισμένος σε αυτήν και μόνο την πληροφορία. Η νέα εικόνα που θα προκύψει αξιολογείται συγκρίνοντας την με την αρχική που δίνεται ως είσοδος. Η σύγκριση αυτή επιτυγχάνεται συνήθως μέσω του μέσου τετραγωνικού σφάλματος (MSE). Η συνάρτηση κόστους περιγράφεται ως εξής

$$\mathcal{L}(x, \hat{x}) = \|x - \hat{x}\|^2. \quad (2.10)$$

Το κυριότερο μέρος ενός αυτόματου κωδικοποιητή είναι το στάδιο της κωδικοποίησης (encoding). Η συμπίεση της εικόνας, δηλαδή η μείωση διάστασης του προβλήματος, κρίνεται αξιοσημείωτη λόγω της σημαντικότητας διατήρησης των κύριων χαρακτηριστικών της. Επίσης, όπως αναφέραμε προηγουμένως, παίζει εξίσου βασικό ρόλο στην γέννηση της νέας εικόνας που θα έχει ως έξοδο το συγκεκριμένο δίκτυο. Για την διαφύλλαξη της πληροφορίας, θα εισάγουμε τις έννοιες των συνελκτικών κωδικοποιητών και αποκωδικοποιητών. Οι τεχνικές αυτές βασίζονται σε εκείνες που επεξηγήσαμε στην προηγούμενη ενότητα για Συνελκτικά Νευρωνικά Δίκτυα (CNN). Κατά την

διάρκεια της κωδικοποίησης, η μείωση της διάστασης της εικόνας γίνεται μέσω των γνωστών μεθόδων του φίλτρου, της εφαρμογής της συνάρτησης ενεργοποίησης (ReLU) και της ομαδοποίησης (max-pooling). Αξίζει να υπογραμμίσουμε πως εάν η εικόνα εισόδου είναι μικρής διάστασης, τότε η μέθοδος της ομαδοποίησης καλύτερα να αποφεύγεται. Εξαιτίας της είδη μικρής διάστασης της εικόνας, η μέθοδος αυτή συμπιέζοντάς την αρκετά, ενδέχεται να αποβάλλει σημαντικές λεπτομέρειές της. Στο κομμάτι της αποκωδικοποίησης, λόγω της συμμετρίας του δικτύου, προβλέπονται οι ίδιες ενέργειες. Πάντα μετά την συνέλιξη της εικόνας, ενδεχομένως μέσω ενός φίλτρου, προβλέπεται η εφαρμογή μιας συνάρτησης ενεργοποίησης. Δεν αναφερθήκαμε στην μέθοδο της ομαδοποίησης, χάρην στην δυσκολία που αντιμετωπίζει στις εικόνες μικρής διάστασης.

Στη συνέχεια θα περιγράψουμε τον τρόπο λειτουργίας του συνελκτικού δικτύου που χρησιμοποιούμε για την κωδικοποίηση και την αποκωδικοποίηση [Mao et al. \(2016\)](#). Αφού τροφοδοτήσουμε μια εικόνα στα επίπεδα συνέλιξης (convolutional layers) που περιγράψαμε, αυτή θα μετατραπεί σε μια λιγότερο "περίπλοκη". Προφανώς, ορισμένες λεπτομέρειες ενδέχεται να εκλείπουν μετά την κωδικοποίηση. Όμως, το ζητούμενο είναι η πληροφορία που μεταφέρουν να είναι όσο το δυνατόν πιο αντιπροσωπευτική της εικόνας εισόδου. Κατά την αποκωδικοποίηση, τα αντίστοιχα κρυφά επίπεδα συνδοιάζονται με τον τρόπο που περιγράψαμε, ώστε να ανακτήσουν οποιαδήποτε πληροφορία περιέχει η αρχική εικόνα. Στη συνέχεια, μέσω της συνάρτησης κόστους  $\mathcal{L}$  θα συγκρίνουμε την ανοικοδομημένη εικόνα με την πραγματική. Τέλος, θα εκπαιδεύσουμε το νευρωνικό δίκτυο ώστε να αναπαράγει εικόνες πανομοιότυπες με των δεδομένων. Φυσικά, η υπερπροσαρμογή (overfitting) του μοντέλου θα πρέπει να λαμβάνεται υπόψη και να αποφεύγεται. Έτσι, εισάγοντας μια νέα παρατήρηση με ελαφρώς διαφορετικά χαρακτηριστικά από αυτές που εκπαιδεύτηκε, το δίκτυο να είναι σε θέση να την ανακατασκευάσει.

### 2.3.2 Variational Autoencoders

Ο Variational Autoencoder (VAE) αποτελεί ένα γεννητικό μοντέλο με πιθανοθεωρητικό υπόβαθρο. Μια χαρακτηριστική του ικανότητα είναι η μοντελοποίηση περίπλοκων κατανομών σε ένα κρυφό χώρο (latent space), ώστε να ανιχνεύει ανωμαλίες σε εικόνες. Δηλαδή, αποσκοπεί στην εξαγωγή μοτίβων που δεν ταιριάζουν με βάση τα αναμενόμενα αποτελέσματα. Λειτουργεί κατασκευάζοντας νέες παρατηρήσεις οι οποίες περιέχουν χαρακτηριστικά που δεν είναι εκ-των-προτέρων γνώριμα. Πιο συγκεκριμένα με βάση τις ήδη υπάρχουσες εικόνες (δεδομένα),

προσπαθώντας να τις ανακατασκευάσει, ανακαλύπτει νέες αναπαραστάσεις τους. Βασική υπόθεση είναι ότι οι εικόνες που δίνονται δεν περιέχουν κάποιου είδους ανεξήγητη συμπεριφορά. Μέσα από την κανονικότητα των δεδομένων, η μέθοδος αυτή στοχεύει στην ανοικοδόμησή τους ώστε να αναγνωρίζει μη αναμενόμενα μοτίβα που ενδέχεται να προκύψουν με την εισαγωγή νέων παρατηρήσεων. Η αρχιτεκτονική του δομή βασίζεται σε αυτή των αυτόματων κωδικοποιητών που περιγράψαμε, αλλά διαφέρει ως προς την ντετερμινιστική συμπεριφορά τους. Όμως, για την ομαλότερη μετάβαση στον τρόπο κατασκευής ενός VAE σημαντική κρίνεται μια μικρή αναφορά στην θεωρία πιθανοτήτων και πληροφορίας [Lupo \(2019\)](#).

### 2.3.2.1 Πληροφορία και εντροπία

Στη θεωρία πιθανοτήτων η πληροφορία κάποιου γεγονότος περιγράφεται από την έκπληξη μας ως προς την εκδήλωση του. Για παράδειγμα, θεωρούμε την πρόταση "τον μήνα Ιούλιο θα έχει πολύ ζέστη στην Αθήνα", της οποίας το περιεχόμενο πρόκειται αρκετά πιθανό να συμβεί. Είναι ευρέως αποδεκτό ότι τον μήνα Ιούλιο στην Ελλάδα και πόσο μάλλον στην Αθήνα, ο καιρός είναι ιδιαίτερα ηλιόλουστος με υψηλές θερμοκρασίες. Άρα, η πρόταση αυτή δεν περιέχει αρκετή πληροφορία. Επίσης, αν υποθέσουμε την πρόταση "τον μήνα Ιούλιο θα χιονίσει στην Αθήνα", η οποία αναφέρεται σε ένα άκρως ασυνήθιστο φαινόμενο. Η πληροφορία που μας δίνεται τώρα είναι πολύ σημαντική, για το συγκεκριμένο καλοκαίρι στην Αθήνα. Έτσι προκύπτει πως οποιαδήποτε πληροφόρηση για ένα μη αναμενόμενο γεγονός, κρίνεται ιδιαίτερα σημαντική. Μπορούμε να ποσοτικοποιήσουμε την έννοια της πληροφορίας ως εξής

$$I = -\log p(x), \quad (2.11)$$

όπου  $x$  είναι κάποιο γεγονός. Έτσι, εάν η πιθανότητα να συμβεί κάποιο γεγονός είναι αρκετά υψηλή, τότε η ποσότητα  $I$  τείνει να μηδενιστεί (συγκεκριμένα  $I \rightarrow 0^-$ ), ενώ στην αντίθετη περίπτωση μειώνεται λογαριθμικά.

Μια επίσης σημαντική έννοια που θα περιγράψουμε είναι η εντροπία. Πρόκειται για την μέση πληροφόρηση που μας δίνεται από μια πρόταση. Συνεπώς, ο τύπος που την περιγράφει, με την βοήθεια τον ορισμό της πληροφορίας που δώσαμε παραπάνω, δίνεται ακολούθως

$$H = -\sum_{x \in \mathbb{X}} p(x) \log p(x), \quad (2.12)$$

όπου  $\mathbb{X}$  θεωρούμε το σύνολο που περιέχει όλα τα δυνατά γεγονότα. Το γεγονός  $x$  είναι τυχαία μεταβλητή σε ένα χώρο πιθανότητας  $(\mathbb{X}, \mathcal{F}, \mathcal{P})$ . Μπορούμε να αποφανθούμε πως η εντροπία ποσοτικοποιεί την αβεβαιότητα που περιλαμβάνεται σε μια τυχαία μεταβλητή ή στο αποτέλεσμα μια τυχαίας διαδικασίας.

### 2.3.2.2 Απόκλιση Kullback-Leibler

Μια έννοια που θα μας φανεί πολύ χρήσιμη στη συνέχεια είναι η απόκλιση Kullback-Leibler (KL-Divergence). Ας θεωρήσουμε δύο τυχαίες μεταβλητές  $X$  και  $Z$ , ορισμένες στον ίδιο χώρο πιθανότητας  $(\mathbb{X}, \mathcal{F}, \mathcal{P})$ , με συναρτήσεις πυκνότητας  $p$  και  $q$  αντίστοιχα. Η απόκλιση Kullback-Leibler εκφράζεται ως

$$\mathcal{D}_{KL}(p||q) = - \sum_{x \in \mathbb{X}} p(x) \log q(x) + \sum_{x \in \mathbb{X}} p(x) \log p(x), \quad (2.13)$$

όπου  $\mathcal{D}_{KL} \geq 0$  και  $\mathcal{D}_{KL} = 0$  εάν και μόνο εάν  $p(x) = q(x)$ . Η παραπάνω ποσότητα αποτελεί ένα είδος ομοιότητας μεταξύ των κατανομών  $p$  και  $q$ . Μια επίσης σημαντική ιδιότητα που την χαρακτηρίζει είναι η ασυμμετρία. Δηλαδή ισχύει ότι  $\mathcal{D}_{KL}(p||q) \neq \mathcal{D}_{KL}(q||p)$ . Άρα δε πρόκειται για κάποιου είδους στατιστικής απόστασης, αφού η έννοια της απόστασης προϋποθέτει την συμμετρία. Για παράδειγμα, η απόσταση από ένα σημείο  $A$  σε ένα σημείο  $B$  είναι ίση με αυτή του σημείου  $B$  ως το  $A$ . Η σχέση (2.13) εκφράζεται με βάση την κατανομή  $p$  ως η διαφορά της εντροπίας της κατανομής  $p$  και της εντροπίας της  $q$  με βάση την κατανομή  $p$  (έτσι εξηγείται το πρώτο συνθετικό δεξιά της ισότητας). Επίσης, η απόκλιση KL περιγράφεται ισοδύναμα και ως

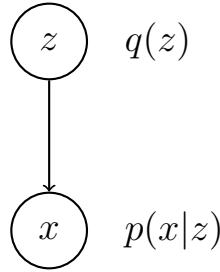
$$\mathcal{D}_{KL}(p||q) = \sum_{x \in \mathbb{X}} p(x) \log \frac{p(x)}{q(x)} = - \sum_{x \in \mathbb{X}} p(x) \log \frac{q(x)}{p(x)}. \quad (2.14)$$

### 2.3.2.2 Πιθανοθεωρητικό μοντέλο

Ένας Variational Autoencoder μπορεί να περιγραφεί ως ένα διάγραμμα πιθανοθεωρητικής προσέγγισης (βλ. Σχήμα 2.5). Θεωρούμε  $x$  μια παρατήρηση και  $z$  μια κρυφή μεταβλητή. Υποθέτουμε ότι και οι δύο μεταβλητές είναι συνεχείς, βασισμένοι στο γεγονός ότι οι τιμές που παίρνουν τα πίξελ είναι αντίστοιχα συνεχείς. Θέλουμε να βρούμε την εκ-των-υστέρων κατανομή  $p(z|x)$ , η οποία με την βοήθεια του Θεωρήματος Bayes (για περισσότερες πληροφορίες βλέπε Ενότητα § 3.1) γράφεται ως

$$p(z|x) = \frac{p(x, z)}{p(x)} = \frac{p(x|z)p(z)}{\int p(x|z)p(z)dz}. \quad (2.15)$$

Ο παρονομαστής της σχέσης (2.15) είναι συνήθως δύσκολο να υπολογιστεί, ειδικά σε μεγάλες διαστάσεις. Υπάρχουν δύο κύριες μέθοδοι για την προσέγγιση της περιθώριας κατανομής των δεδομένων. Η πρώτη είναι οι μέθοδοι προσωμοίωσης Markov Chain Monte Carlo, όπως ο δειγματολήπτης Gibbs ή ο αλγόριθμος Metropolis-Hastings. Η δεύτερη, και αυτή που θα μας απασχολήσει σε αυτή την ενότητα, είναι η Variational συμπερασματολογία (inference) η οποία προσεγγίζει την πραγματική ποσότητα προσεγγιστικά κάνοντας ορισμένες απλουστεύσεις (μερικές φορές μη ρεαλιστικές) που διευκολύνουν τους υπολογισμούς.



Σχήμα 2.5: Κατευθυνόμενο διάγραμμα δύο τυχαίων μεταβλητών  $z, x$

Η βασική ιδέα προσέγγισης της εκ-των-υστέρων κατανομής  $p(z|x)$ , γίνεται μέσω μιας νέας κατανομής  $q(z)$ . Επιλέγουμε η κατανομή αυτή να είναι γνωστή (για παράδειγμα να ανήκει στην εκθετική οικογένεια κατανομών) και εύκολα διαχειρίσιμη. Παραμετροποιούμε κατάλληλα την  $q(z)$ , ώστε να είναι σχεδόν όμοια με την ζητούμενη κατανομή  $p(z|x)$ . Ουσιαστικά αποφεύγουμε την ισότητα (2.15) μέσω της νέας συνάρτησης που θεωρήσαμε. Στην προσπάθεια αυτή, σημαντικό εργαλείο αποδεικνύεται η απόκλιση Kullback-Leibler που συζητήσαμε προηγουμένως. Στόχος είναι η ελαχιστοποίηση της απόκλισης KL μεταξύ των προηγούμενων κατανομών, δηλαδή θέλουμε

$$\min_{p, q} \mathcal{D}_{KL}(q(z)||p(z|x)) = \min_{p, q} \left( - \sum_{z \in \mathcal{Z}} q(z) \log \frac{p(z|x)}{q(z)} \right). \quad (2.16)$$

Αντικαθιστώντας την σχέση (2.15) στην (2.16) και λύνοντας ως προς τον λογάριθμο της  $p(x)$  παίρνουμε

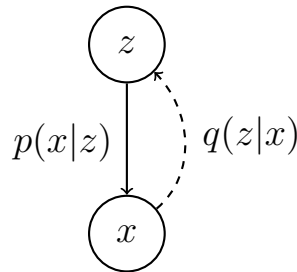
$$\log p(x) = \mathcal{D}_{KL}(q(z)||p(z|x)) + \lambda, \quad (2.17)$$

όπου,

$$\lambda = \mathbb{E}_{q(z)} [\log p(x|z)] - \mathcal{D}_{KL}(q(z)||p(z)). \quad (2.18)$$

Γνωρίζουμε ότι η ποσότητα  $\mathcal{D}_{KL}(q(z)||p(z|x))$  είναι μη αρνητική, άρα μπορούμε να υποθέσουμε πως η ποσότητα  $\lambda$  είναι ένα κάτω όριο του λογαρίθμου του  $p(x)$ . Δηλαδή ισχύει ότι  $\lambda \leq \log p(x)$ . Θέλουμε να μεγιστοποιήσουμε το  $\lambda$ , ώστε να μεγιστοποιήσουμε έμμεσα την  $\log p(x)$ . Το  $\lambda$  καλείται και Variational ή Evidence Lower Bound (ELBO) της περιθώριας πιθανοφάνειας των παρατηρήσεων. Έτσι, η ελαχιστοποίηση της  $\mathcal{D}_{KL}$  ισοδυναμεί με την μεγιστοποίηση της ELBO [Jospin et al. \(2020\)](#). Για να το κάνουμε αυτό, η χρήση του αφελή εκτιμητή κλίσης Monte Carlo (Naive Monte Carlo Gradient Estimator) κρίνεται εξαιρετικά προβληματική. Μέσω του τελευταίου επιτυγχάνεται μεγάλη διακύμανση, γεγονός που αναδεικνύεται στην εργασία των [Paisley et al. \(2012\)](#). Για τον λόγο αυτό μπορούμε να χρησιμοποιήσουμε τον αλγόριθμο του Αυτόματου Κωδικοποιητή Παραλλαγής Bayes (Autoencoding Variational Bayes ή AEVB) με την βοήθεια του γενετικού εκτιμητή Στοχαστικής Κλίσης Παραλλαγής Bayes (Stochastic Gradient Variational Bayes ή SGVB). Ο τρόπος λειτουργίας του περιγράφεται αναλυτικά από τους [Kingma and Welling \(2013\)](#).

Επόμενο βήμα είναι η προσαρμογή του πιθανοθεωρητικού μοντέλου σε έναν αυτόματο κωδικοποιητή. Θεωρούμε μια παραλλαγή του Σχήματος 2.5, σύμφωνα με την οποία υπάρχει μια κατανομή  $q(z|x)$ , η οποία καθορίζει την λανθάνουσα μεταβλητή  $z$  με βάση τη μεταβλητή  $x$  (βλ. Σχήμα 2.6). Όπως είδαμε, ο υπολογισμός της κατανομής  $p(x|z)$  ενδέχεται να είναι δύσκολος. Αποσκοπούμε στον υπολογισμό της κατανομής  $q(z|x)$ , στοχεύοντας στην μεγάλη ομοιότητα της με την δυσεύρετη αυτή κατανομή.



Σχήμα 2.6: Παραλλαγή κατευθυνόμενου γραφήματος δύο τυχαίων μεταβλητών  $z, x$

Τώρα, μπορούμε να θεωρήσουμε πως η  $q$  είναι ένα νευρωνικό δίκτυο (encoder) όπου λαμβάνει δεδομένα  $x$  και τα απεικονίζει σε ένα κρυφό επίπεδο  $z$ . Επίσης υποστηρίζουμε πως και η  $p$  είναι νευρωνικό δίκτυο (decoder) που λαμβάνει τιμές από ένα κρυφό χώρο  $z$  και τις απεικονίζει σε μια

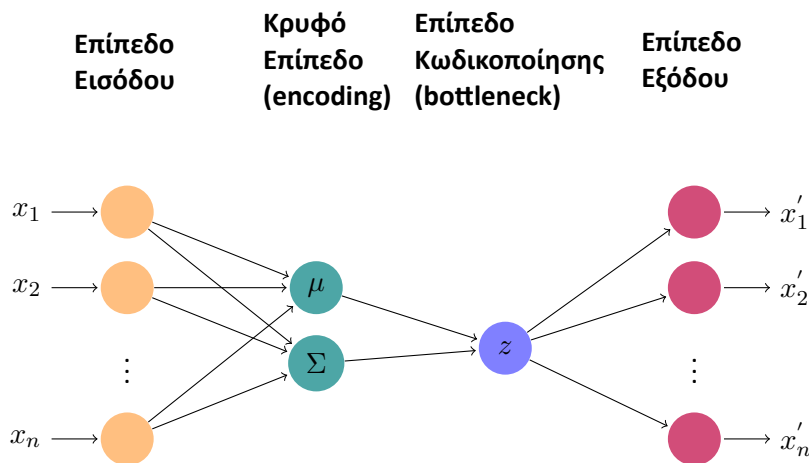
ανακατασκευή των δεδομένων  $\hat{x}$ . Έως τώρα έχουμε περιγράψει μια ντετερμινιστική προσέγγιση, και δεν πρέπει να ξεχνάμε πως το μοντέλο αυτό είναι πιθανοθεωρητικό. Αντί να παίρνουμε την ακριβής κωδικοποιημένη μορφή του  $z$  που προκύπτει από το νευρωνικό δίκτυο του κωδικοποιητή, θεωρούμε την περιθώρια κατανομή του  $z$ ,  $q(z)$ , και θέλουμε να βρούμε τις παραμέτρους της κατανομής που την γεννάει. Έτσι, θα προσομοιώνουμε τιμές από αυτή και θα τις τροφοδοτούμε στον αποκωδικοποιητή.

Συνήθως, παίρνουμε εκ-των-προτέρων την περιθώρια κατανομή του  $z$  ως πολυμεταβλητή κανονική κατανομή με μέση τιμή  $\mu$  διάστασης  $d \times 1$  και πίνακα διακύμανσης συνδιακύμανσης  $\Sigma'$  διάστασης  $d \times d$ . Εξαιτίας της πιθανοθεωρητικής προσέγγισης που υιοθετούμε, μέθοδοι ανάστροφης μετάδοσης και τελικά στοχαστικοί αλγόριθμοι βελτιστοποίησης, είναι ανέφικτοι. Αιτία του προβλήματος είναι ότι τώρα ο κόμβος  $z$  είναι τυχαία μεταβλητή που ακολουθεί την κανονική κατανομή. Με κάποιον τρόπο θέλουμε να μετατρέψουμε τον τρόπο λειτουργίας του VAE, στο κρυφό επίπεδο του bottleneck, σε ντετερμινιστικό. Αυτό μπορεί να επιτευχθεί μέσω μιας παραμετροποίησης στο γραφικό μοντέλο που περιγράψαμε [Kingma and Welling \(2013\)](#). Για λόγους απλότητας θεωρούμε τον πίνακα  $\Sigma'$  διαγώνιο, με αποτέλεσμα να περιγράψουμε την διαγώνιο μέσω ενός νέου διάνυσματος  $\Sigma$  διάστασης  $d \times 1$ . Χωρίζουμε σε δύο ίσα μέρη τις τιμές που παίρνουμε από την έξοδο του κωδικοποιητή, ώστε το καθένα να περιέχει  $d$  στοιχεία. Προφανώς ο αριθμός των κόμβων εξόδου του κωδικοποιητή, όπως και των υπόλοιπων κρυφών επιπέδων, αποτελεί υπερπαραμέτρο προς επιλογή από τον χρήστη. Το ένα μέρος θα αντιστοιχεί στο διάνυσμα της μέσης τιμής και το άλλο στις συνδιακυμάνσεις του πίνακα  $\Sigma'$ . Στη συνέχεια θα παράγουμε τιμές από την κατανομή  $\mathcal{N}(\mu_d, \Sigma_d)$  και θα τις τροφοδοτούμε στον αποκωδικοποιητή, ώστε να γεννήσουμε τις τιμές  $\hat{x}$ . Στο [Σχήμα 2.7](#) βλέπουμε την αρχιτεκτονική δομή ενός VAE σύμφωνα με την προηγούμενη περιγραφή.

Μια ισοδύναμη αναπαράσταση της διαδικασίας που περιγράψαμε δίνεται στο [Σχήμα 2.8](#), όπου  $\epsilon$  τυχαία μεταβλητή που ακολουθεί την κατανομή  $\mathcal{N}(0, 1)$ . Έτσι η τιμή ενός κόμβου  $z_i$  προκύπτει από την σχέση

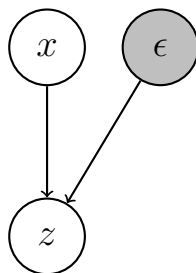
$$z_i = \mu_i + \Sigma_i \odot \epsilon. \quad (2.19)$$

Με την παραμετροποίηση αυτή καταφέρνουμε ο κόμβος  $z$  να είναι ντετερμινιστικός. Οι δειγματοληψίες από την κατανομή που ακολουθεί δεν επηρεάζουν την εκπαίδευση του δικτύου,



Σχήμα 2.7: Αρχιτεκτονική δομή ενός Variational Autoencoder (VAE)

αφού η τυχαία μεταβλητή  $\epsilon$  αποτελεί απομονωμένο κομμάτι της διαδικασίας.



Σχήμα 2.8: Κατευθυνόμενο διάγραμμα δύο τυχαίων μεταβλητών  $z, x$ , με την συνεισφορά μια τυχαίας μεταβλητής  $\epsilon \sim \mathcal{N}(0, 1)$

## 2.4 Συμπεράσματα και συζήτηση

Έως τώρα, οι τεχνικές που αναφέραμε αφορούν την ταξινόμηση των εικόνων ανάλογα με το περιεχόμενό τους, καθώς και την ανοικοδόμησή τους με την βοήθεια τεχνικών βαθιάς μάθησης. Σημαντική προϋπόθεση για την εφαρμογή τέτοιων μεθόδων είναι η διαθεσιμότητα μεγάλων συνόλων δεδομένων από τα οποία θα γίνει η εκμάθηση τους. Οι τελευταίες μπορούν να συγκαταλεχθούν σε μια γενικότερη κατηγορία μεθόδων, αυτή των γενετικών, εξαιτίας της δυνατότητας δημιουργίας νέων πανομοιότυπων εικόνων σύμφωνα με μια βάση δεδομένων.

Αντίθετα, η Μπεϋζιανή ανάλυση εικόνας τα τελευταία χρόνια διαδραματίζει σημαντικό ρόλο στην επίλυση προβλημάτων μείωσης θορύβου, βελτιστοποίησης χαρακτηριστικών και ανίχνευσης αντικειμένων. Μέσω των εκ-των-προτέρων (apriori) πεποιθήσεών μας σχετικά με την δομή μιας εικόνας, η Μπεϋζιανή ανάλυση μπορεί να αυξήσει την ποιότητα της. Βασικό μέρος της ανάλυσης



αυτής αποτελεί το Θεώρημα του Bayes. Η σημαντική διαφορά με τις τεχνικές που αναλύσαμε στο παρών κεφάλαιο είναι ότι δεν απαιτείται μεγάλη βάση δεδομένων για την εκμάθηση.

Οι ιατρικές εικόνες, όπως οι αξονικές (CT), οι μαγνητικές (MRI) και οι τομογραφίες εκπομπής ποζιτρονίων (PET), στηρίζονται στις παραπάνω εφαρμογές. Αναμφισβήτητα αποτελούν ευαίσθητες εικόνες, στις οποίες ένας κλινικός ιατρός καλείται να πάρει πολύ σημαντικές αποφάσεις. Αυτές μπορεί να αναφέρονται σε θεραπεία για την ύπαρξη ή όχι καρκίνου σε κάποιο μέρος του σώματος. Ένα ακόμα παράδειγμα είναι ο εύγλωττος διαχωρισμός όγκου ως καλοήθους ή κακοήθους από τον αρμόδιο ιατρό. Η ποιότητα της εικόνας και η ανάδειξη χαρακτηριστικών που καλύπτονται από θόρυβο, αποτελούν τους πιο χρήσιμους παράγοντες για την λήψη τέτοιων αποφάσεων.



## 3. Μπεϋζιανές τεχνικές ανάλυσης εικόνας

Οι παραδοσιακές Μπεϋζιανές μέθοδοι επικεντρώνονται στον τομέα του πραγματικού χώρου της εικόνας τον οποίο μπορούμε να ερμηνεύσουμε εμείς οι άνθρωποι. Εμπνευσμένοι από την εργασία των Kornak et al. (2020), θέλουμε να αναδείξουμε την μέθοδο Μπεϋζιανής ανάλυσης στον Χώρο Fourier (Bayesian Image analysis in Fourier Space ή BIFS). Σε αντίθεση με τις παραδοσιακές προσεγγίσεις, η τεχνική αυτή μετασχηματίζει την εικόνα στον χώρο συχνοτήτων εκμεταλλευόμενη ορισμένες χρήσιμες ιδιότητες που τον διέπουν. Η εύκολη μοντελοποίηση και η ταχύτητα υπολογισμού είναι μερικά από τα πλεονεκτήματα που την χαρακτηρίζουν.

Αρχικά, στην Ενότητα § 3.1 θα αναφερθούμε στον κεντρικό πυλώνα της Μπεϋζιανής ανάλυσης, την Μπεϋζιανή συμπερασματολογία. Θα αναπτύξουμε σε βάθος το θεωρητικό υπόβαθρο που θα μας φανεί χρήσιμο για την ανάλυση εικόνων. Στην Ενότητα § 3.2 θα αναφέρουμε μια από τις επικρατέστερες μεθόδους Μπεϋζιανής ανάλυσης εικόνων, τα τυχαία πεδία Markov (Random Markov Fields). Θα εξετάσουμε τον τρόπο λειτουργίας τους, καθώς και την υπολογιστική τους ικανότητα. Επίσης θα αναλύσουμε τη μέθοδο της κανονικοποίησης Tikhonov καθώς και το πως ερμηνεύεται με την Μπεϋζιανή ανάλυση. Στην Ενότητα § 3.3, θα κάνουμε μια εισαγωγή στους μιγαδικούς αριθμούς και στη συνέχεια θα επικεντρωθούμε στην ανάλυση Fourier και στις ιδιότητες του χώρου αυτού. Η τελευταία θα αποτελέσει ένα σημαντικό θεωρητικό εργαλείο για την ανάδειξη της μεθόδου BIFS. Τέλος, στην Ενότητα § 3.4 θα εστιάσουμε στον αλγόριθμο BIFS, ο οποίος θα αποτελέσει την κύρια μέθοδο ανάλυσης εικόνων της παρούσας εργασίας.

### 3.1 Μπεϋζιανή συμπερασματολογία

Η Μπεϋζιανή συμπερασματολογία (Bayesian Inference) στηρίζεται στο Θεώρημα του Bayes, το οποίο περιγράφεται στην παρακάτω σχέση

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (3.20)$$

Αντικαθιστώντας τώρα το ενδεχόμενο  $B$  με τα δεδομένα (παρατηρήσεις)  $D$ , το ενδεχόμενο  $A$  με ένα σύνολο παραμέτρων  $\Theta$  και τις πιθανότητες με τις κατανομές  $p$ , προκύπτει το Θεώρημα Bayes όπως χρησιμοποιείται στη σύγχρονη Μπεϋζιανή στατιστική:

$$p(\Theta|D) = \frac{p(D|\Theta)p(\Theta)}{p(D)} = \frac{p(D|\Theta)p(\Theta)}{\int p(D|\Theta)p(\Theta)d\Theta}, \quad (3.21)$$

όπου  $p(\Theta)$  είναι η κατανομή των παραμέτρων  $\Theta$  πριν παρατηρήσουμε τα δεδομένα  $D$  (εκ-των-προτέρων ή *apriori* κατανομή),  $p(D|\Theta)$  η συνάρτηση πιθανοφάνειας των δεδομένων  $D$ ,  $p(\Theta|D)$  η από κοινού εκ-των-υστέρων κατανομή των παραμέτρων  $\Theta$  (posterior) δοθέντων των δεδομένων  $D$  και  $p(D)$  η περιθώρια πιθανοφάνεια των δεδομένων  $D$ . Η τελευταία ονομάζεται και σταθερά κανονικοποίησης (στις περισσότερες περιπτώσεις) της εκ-των-υστέρων κατανομής,  $p(\Theta|D)$ . Ο υπολογισμός της είναι αρκετά δύσκολος, ακόμα και αδύνατος όταν η διάσταση του προβλήματος είναι σημαντικά αυξημένη. Αφαιρώντας την περιθώρια κατανομή των δεδομένων από την σχέση (3.21), η σχέση της ισότητας αντικαθίσταται από τη σχέση αναλογίας ( $\propto$ ). Άρα έχουμε,

$$p(\Theta|D) \propto p(D|\Theta)p(\Theta). \quad (3.22)$$

Έτσι, μπορούμε να πούμε ότι η εκ-των-υστέρων κατανομή του  $\Theta$  είναι ανάλογη του γινομένου της συνάρτησης πιθανοφάνειας με την εκ-των-προτέρων κατανομή. Δηλαδή, προκύπτει συνδυάζοντας τα δεδομένα  $D$  που παρατηρήσαμε και την προσωπική μας άποψη (εκ-των-προτέρων κατανομή). Ολόκληρη η συμπερασματολογία για μια παράμετρο  $\theta \in \Theta$  στηρίζεται στην εκ-των-υστέρων κατανομή της.

### 3.1.1 Εκ-των-προτέρων κατανομή

Σε πολλά προβλήματα που μας ενδιαφέρει η συμπερασματολογία για μια άγνωστη παράμετρο  $\theta \in \Theta$  μπορεί να υπάρχει διαθέσιμη εκ-των-προτέρων πληροφορία, ίσως από προηγούμενη εμπειρία. Η πληροφορία αυτή περιγράφεται μέσω της εκ-των-προτέρων κατανομής  $p(\theta)$ . Θα μπορούσαμε να αποφανθούμε πως η κατανομή αυτή αποτελεί μια έκφραση της αβεβαιότητας

που έχουμε, πρώτου τα δεδομένα ληφθούν υπόψη. Όπως είναι προφανές από την σχέση (3.22), η εκ-των-προτέρων κατανομή επηρεάζει την εκ-των-υστέρων. Συνεπώς, η επιλογή της χρήζει ιδιαίτερης προσοχής. Αξίζει να αναφέρουμε ότι αποτελεί ένα από τα δυσκολότερα σημεία της Μπεϋζιανής ανάλυσης. Αν το σύνολο  $\Theta$  είναι διακριτό τότε σε κάθε τιμή του αντιστοιχίζεται και μια υποκειμενική πιθανότητα. Στην περίπτωση που το σύνολο  $\Theta$  είναι συνεχές, η κατανομή του μπορεί να καθοριστεί μέσω ενός ιστογράμματος, όπου σε κάθε διαμέρισή του αναλογεί μια υποκειμενική πιθανότητα. Επίσης, η επιλογή της εκ-των-προτέρων κατανομής μπορεί να καθοριστεί μέσω μιας συγκεκριμένης παραμετρικής οικογένειας κατανομών.

Οι εκ-των-προτέρων κατανομές χωρίζονται σε δύο κατηγορίες, τις πληροφοριακές και τις μη-πληροφοριακές. Εάν η εκ-των-προτέρων γνώση για την παράμετρο  $\theta$  είναι διαθέσιμη, για παράδειγμα από προηγούμενα δεδομένα ή άλλες μελέτες, τότε μπορεί να περιγραφεί από την εκ-των-προτέρων κατανομή  $p(\theta)$ . Τότε η  $p(\theta)$  ονομάζεται πληροφοριακή. Αντίθετα, υπάρχουν περιπτώσεις που δεν έχουμε διαθέσιμη πληροφορία είτε δεν επιθυμούμε να την προσαρμόσουμε στο μοντέλο μας. Τότε, η ενσωμάτωση μιας κατανομής η οποία έχει σταθερή πυκνότητα σε όλο το πεδίο ορισμού της μπορεί να είναι μια καλή ιδέα. Τέτοιες εκ-των-προτέρων κατανομές ονομάζονται μη πληροφοριακές. Κλασικό παράδειγμα αποτελεί η ομοιόμορφη κατανομή  $\mathcal{U}(a, b)$ , η οποία δίνει ίδια πιθανότητα σε κάθε υποδιάστημα ίδιου μήκους του  $[a, b]$ . Είναι προφανές πως μια μη-πληροφορική εκ-των-προτέρων κατανομή, προϋποθέτει το παραμετρικό σύνολο  $\Theta$  να είναι φραγμένο (αφού πρόκειται για κατανομή πιθανότητας). Υπογραμμίζουμε ότι σε μετασχηματισμούς μιας παραμέτρου  $\theta \in \Theta$ , η ιδιότητα της μη-πληροφορίας της εκ-των-προτέρων κατανομής δεν διατηρείται. Παρόλ'αυτά, στη θέση της εκ-των-προτέρων κατανομής μπορούμε να χρησιμοποιήσουμε οποιαδήποτε μη αρνητική συνάρτηση  $p(\theta)$ , για την οποία ισχύει  $\int_{\Theta} p(\theta) d\theta = +\infty$ . Τέτοιας μορφής εκ-των-προτέρων κατανομές ονομάζονται καταχρηστικές. Όμως, για την χρήση της πρέπει η εκ-των-υστέρων κατανομή να είναι καλώς ορισμένη ως κατανομή, ειδάλως δεν μπορούμε να καταλήξουμε σε κάποιο λογικό συμπέρασμα για την παράμετρο ενδιαφέροντος  $\theta$ .

Στη συνέχεια, θα αναφέρουμε την μέθοδο του Jeffreys για την εύρεση μη πληροφοριακών εκ-των-προτέρων κατανομών για μονοδιάστατες παραμέτρους  $\theta$ . Η μέθοδος αυτή στηρίζεται στον αριθμό πληροφορίας του Fisher

$$I(\theta) = E_{\theta}^D \left[ \frac{\partial}{\partial \theta} \log f(D|\theta) \right]. \quad (3.23)$$

Η εκ-των-προτέρων κατανομή του Jeffreys δίνεται ακολούθως.

$$p_j(\theta) \propto \sqrt{I(\theta)} \quad (3.24)$$

Σημειώνουμε πως η  $p_j(\theta)$  παραμένει αναλλοίωτη ως προς ένα προς ένα μετασχηματισμούς της  $\theta$ . Ενδέχεται εκτός από κατανομή πιθανότητας, να είναι και καταχρηστική κατανομή. Σημειώνουμε ότι σε περίπτωση που το  $\theta$  είναι πολυδιάστατο τότε η παραπάνω σχέση δεν ισχύει.

Όταν η εκ-των-υστέρων πιθανότητα  $p(\Theta|D)$  ανήκει στην ίδια οικογένεια με την εκ-των-προτέρων κατανομή, τότε η εκ-των-υστέρων και η εκ-των-προτέρων ονομάζονται συζυγείς κατανομές. Η έννοια της εκ-των-προτέρων συζυγής κατανομής εισήχθει από τους [Raiffa and Schlaifer \(1961\)](#). Μια οικογένεια κατανομών  $\mathcal{H}$  με στήριγμα  $\Theta$  λέμε ότι είναι συζυγής για την συνάρτηση πιθανοφάνειας  $p(D|\theta)$  αν για κάθε εκ-των-προτέρων κατανομή  $p(\theta) \in \mathcal{H}$ , η αντίστοιχη εκ-των-υστέρων κατανομή ανήκει επίσης στην οικογένεια  $\mathcal{H}$ . Ο παραπάνω ορισμός είναι πολύ γενικός, όμως οι πιο ενδιαφέρουσες περιπτώσεις επικεντρώνονται σε συγκεκριμένες οικογένειες κατανομών. Για παράδειγμα, έχοντας κανονική συνάρτηση πιθανοφάνειας, τότε εάν επιλέξουμε μια κανονική εκ-των-προτέρων κατανομή για το  $\mu$  (δηλαδή το μέσο της κατανομής), τότε η εκ-των-υστέρων κατανομή θα είναι και αυτή κανονική. Όταν επιλέγουμε να χρησιμοποιούμε συζυγείς εκ-των-προτέρων κατανομές, τότε δεν είναι απαραίτητος ο υπολογισμός της περιθώριας πιθανοφάνειας των δεδομένων  $p(D)$ . Πράγματι, από την σχέση (3.21), οι όροι που περιέχουν το  $\theta$  στην εκ-των-υστέρων κατανομή, θα είναι παρόμοιοι με αυτούς που περιέχουν το  $\theta$  στην  $p(D|\Theta)p(\Theta)$ .

Κατόπιν, θα εξετάσουμε την ισχύ της προηγούμενης παραγράφου για τις συζυγείς κατανομές μέσω ενός παραδείγματος. Ας θεωρήσουμε  $x = (x_1, x_2, \dots, x_n)$  ένα τυχαίο δείγμα από την κατανομή  $\mathcal{N}(\mu, \sigma^2)$ . Υποθέτουμε πως  $\sigma > 0$  και ότι  $\theta = \mu \in \Theta = \mathbb{R}$  άγνωστη παράμετρος. Εκ-των-προτέρων ορίζουμε ότι  $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$  όπου  $\mu_0 \in \mathbb{R}$  και  $\sigma_0 > 0$  δεδομένες σταθερές. Θέλουμε να υπολογίσουμε την εκ-των-υστέρων κατανομή του  $\mu$ . Παρακάτω δίνουμε την ιδιότητα που θα μας βοηθήσει για τον υπολογισμό της ζητούμενης εκ-των-υστέρων κατανομής

$$\sum_{i=1}^n (x_i - \mu)^2 = (n-1)s^2 + n(\bar{x} - \mu)^2, \quad (3.25)$$

όπου  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  και  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ . Έτσι, η πιθανοφάνεια των δεδομένων είναι

$$p(x|\mu) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{(n-1)s^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right). \quad (3.26)$$

Επίσης, ξέρουμε ότι η εκ-των-προτέρων κατανομή για την παράμετρο  $\mu$  είναι η παρακάτω

$$p(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right). \quad (3.27)$$

Αντικαθιστώντας τις σχέσεις (3.26),(3.27) στην (3.22) και έπειτα από υπολογισμούς προκύπτει ότι η εκ-των-υστέρων κατανομή του  $\mu$  είναι της μορφής

$$p(\mu|x) \propto \exp\left(-a\mu^2 + b\mu\right), \quad (3.28)$$

όπου  $a = \frac{1}{2} \left( \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right)$  και  $b = \left( \frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right)$ . Τώρα εύκολα αποδεικνύεται ότι  $\mu|x \sim \mathcal{N}\left(\frac{b}{2a}, \frac{1}{2a}\right)$ . Το αποτέλεσμα αυτό που αναδείξαμε, επαληθεύει τον ορισμό για τις συζυγείς κατανομές. Η αναφορά σε αυτή τη συγκεκριμένη περίπτωση συζυγών κατανομών έγινε, διότι το αποτέλεσμα της θα αποτελέσει χρήσιμο εργαλείο για την μετέπειτα ανάλυσή μας.

### 3.1.2 Πιθανοφάνεια

Η συνάρτηση πιθανοφάνειας, όπως και η εκ-των-προτέρων κατανομή, πρέπει να είναι πλήρως καθορισμένες. Σε περίπτωση που δεν μπορεί να υπολογιστεί ακριβώς, οφείλουμε να την προσεγγίσουμε. Η πιθανοφάνεια  $p(D|\Theta)$  εμπεριέχει όλη τη διαθέσιμη πληροφορία που μας δίνεται από το δείγμα το παρατηρήσεων μας  $D$ . Συνήθως συμβολίζεται ως  $\mathcal{L}(\Theta|D)$  και δίνεται από την παρακάτω σχέση

$$\mathcal{L}(\Theta|D) = p(D|\Theta) = \prod_{i=1}^n p(d_i|\Theta), \quad (3.29)$$

όπου  $n$  το πλήθος των παρατηρήσεων,  $d_i$  η  $i$ -οστή παρατήρηση και  $d_i \perp\!\!\!\perp d_j|\Theta, \forall i \neq j$ , όπου  $i, j = 1, 2, \dots, n$ . Επίσης, από τη σχέση (3.22), παρατηρούμε ότι τα δεδομένα  $D$  επηρεάζουν την εκ-των-υστέρων κατανομή  $p(\Theta|D)$  αποκλειστικά μέσω της συνάρτησης πιθανοφάνειας (έτσι, η Μπεϋζιανή συμπερασματολογία ανταποκρίνεται πλήρως στην Αρχή της πιθανοφάνειας). Δηλαδή,

η πληροφορία που φέρουν τα δεδομένα περιέχεται εξ'ολοκλήρου στην συνάρτηση πιθανοφάνειας  $\mathcal{L}(\Theta|D)$ . Αν υποθέσουμε  $D_1$  και  $D_2$ , δύο διαφορετικά σύνολα παρατηρήσεων που εξαρτώνται από την ίδια παράμετρο  $\theta$  και υπάρχει θετική σταθερά  $c$  τέτοια ώστε

$$\mathcal{L}(\theta|D_1) = c\mathcal{L}(\theta|D_2), \quad (3.30)$$

για κάθε  $\theta \in \Theta$ , τότε τα  $D_1, D_2$  φέρουν την ίδια πληροφορία για το  $\theta$ .

### 3.1.3 Αξιόπιστα σύνολα

Σε αυτήν την υποενότητα θα κάνουμε μια σύντομη αναφορά στα αξιόπιστα σύνολα για μια παράμετρο  $\theta \in \Theta$  και στα αντίστοιχα ύψιστης εκ-των-υστέρων πυκνότητας. Στην Μπεϋζιανή συμπερασματολογία, αρκετά συχνά εκτιμούμε παραμέτρους με κάποιο διάστημα. Θα αναφερθούμε στο ανάλογο διάστημα εμπιστοσύνης της κλασικής Στατιστικής, που υπό την Μπεϋζιανή ονομάζεται αξιόπιστο σύνολο (credible interval). Ένα υποσύνολο του  $\Theta$  καλείται  $100(1 - a)\%$  αξιόπιστο σύνολο για το  $\theta$  εάν ισχύει η ακόλουθη ανισότητα

$$P(\theta \in C|D) = \int_{\theta \in C} p(\theta|D)d\theta \geq 1 - a. \quad (3.31)$$

Συνεπώς, ένα  $100(1 - a)\%$  αξιόπιστο σύνολο περιέχει την παράμετρο  $\theta$  με πιθανότητα  $1 - a$ . Ο παραπάνω ορισμός έρχεται σε πλήρη αντίθεση με την ερμηνεία ενός διαστήματος εμπιστοσύνης της κλασικής Στατιστικής. Ένα αξιόπιστο σύνολο που χρησιμοποιείται συνήθως στη Μπεϋζιανή Συμπερασματολογία είναι αυτό της ύψιστης εκ-των-υστέρων πυκνότητας (Highest Posterior Density Interval ή HPDI) της  $\theta$ . Το σύνολο αυτό αναδεικνύει το μικρότερο δυνατό διάστημα, το οποίο περιέχει τις τιμές με τις μεγαλύτερες εκ-των-υστέρων πιθανότητες. Επίσης, ένα από τα σημαντικότερα στοιχεία που περιέχει είναι η μέγιστη εκ-των-υστέρων πιθανότητα, για την οποία θα συζητήσουμε στη συνέχεια. Προφανώς, λοιπόν, το αξιόπιστο σύνολο ύψιστης εκ-των-υστέρων πυκνότητας είναι ακριβέστερο και μικρότερου μήκους από ένα οποιοδήποτε αξιόπιστο σύνολο.

### 3.1.4 Μπεϋζιανοί έλεγχοι υποθέσεων

Ένα σημαντικό εργαλείο στην μπεϋζιανή στατιστική είναι οι έλεγχοι υποθέσεων. Η συνεισφορά τους στην λήψη αποφάσεων και ειδικότερα στην επιλογή μοντέλου είναι άμεση. Θεωρούμε μια



παράμετρο  $\theta$  και μας ενδιαφέρει να ελέγξουμε εάν προέρχεται από το σύνολο κατανομών  $\Theta_0$  ή από το  $\Theta_1$  (όπου  $\Theta_0 = \Theta_1^c$ ). Το πρόβλημα αυτό αφορά έναν έλεγχο υπόθεσης και μπορεί να αναπαρασταθεί ως εξής

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1. \quad (3.32)$$

Σύμφωνα με την Θεωρία Αποφάσεων το παραπάνω πρόβλημα είναι ανάλογο με την λήψη μιας απόφασης  $\phi \in \{0, 1\}$ , όπου 1 είναι η απόρριψη της μηδενικής υπόθεσης ( $H_0$ ) και 0 η αποδοχή της. Η συνάρτηση κόστους που χρησιμοποιείται σε τέτοιους ελέγχους ποικίλλει ανάλογα με το πρόβλημα που εξετάζουμε. Η επικρατέστερη συνάρτηση κόστους είναι αυτή του σφάλματος 0-1. Συμβολίζουμε τη συνάρτηση κόστους ως  $L$  και την αναπαριστούμε ως

$$L(\theta, \phi) = \begin{cases} \phi, & \theta \in \Theta_0 \\ 1 - \phi, & \theta \in \Theta_1 \end{cases}. \quad (3.33)$$

Τότε, η απόφαση Bayes για το πρόβλημα δίνεται από την εξής έκφραση

$$\phi(D) = \begin{cases} 1, & P(\theta \in \Theta_0 | D) < \frac{1}{2} \\ 0, & P(\theta \in \Theta_0 | D) \geq \frac{1}{2} \end{cases}. \quad (3.34)$$

Στην προσπάθεια μας να εξηγήσουμε την εκ-των-υστέρων κατανομή  $p(\Theta|D)$ , κάποιου είδους πληροφορία χάνεται. Ένα από τα κύρια σημεία της Μπεϋζιανής προσέγγισης είναι η βελτιστοποίηση της εκ-των-υστέρων κατανομής. Ουσιαστικά, η θεώρηση μια συνάρτησης κόστους είναι άρρηκτα συνεδωμένη με τον υπολογισμό πιθανοτήτων της εκ-των-υστέρων κατανομής και συνεπώς, με την απόφαση που τελικά καλούμαστε να επιλέξουμε.

### 3.1.5 Επιλογή μοντέλου

Με την βοήθεια των ελέγχων υποθέσεων που είδαμε στην προηγούμενη υποενότητα, τώρα θεωρούμε το γενικότερο πρόβλημα επιλογής μοντέλου. Έστω  $M_i$  ένα μοντέλο που προέρχεται από το σύνολο όλων των πιθανών μοντέλων  $\mathcal{M}$ . Η επιλογή μοντέλου σύμφωνα με την Μπεϋζιανή προσέγγιση γίνεται μέσω της εκ-των-υστέρων πιθανότητας κάθε μοντέλου  $M_i$ , η οποία δίνεται ακολούθως

$$P(M_i|D) = \frac{\pi_i p_i(D)}{\sum_{k=1}^K \pi_k p_k(D)}, \quad (3.35)$$

όπου  $\pi_k$  η εκ-των-προτέρων πιθανότητα του μοντέλου  $k$  και  $K$  το πλήθος των μοντέλων που ανήκουν στο σύνολο  $\mathcal{M}$  (ο οποίος ονομάζεται και μοντελο-χώρος). Επίσης,  $p_k(D)$  είναι η περιθωριακή πιθανοφάνεια των δεδομένων  $D$  για το μοντέλο  $M_k$ , όπου  $k = 1, \dots, K$ .

Το βέλτιστο μοντέλο είναι εκείνο με την μεγαλύτερη εκ-των-υστέρων πιθανότητα  $P(M_i|D)$  (Maximum A Posteriori ή MAP). Εάν θεωρήσουμε  $\widehat{M}$  την MAP εκτίμηση του  $M$  τότε ισοδύναμα μπορούμε να την ορίσουμε ως

$$\widehat{M} = \arg \max_{M \in \mathcal{M}} P(M|D). \quad (3.36)$$

Με μια γεωμετρική ματιά μπορούμε να αποφανθούμε ότι η συγκεκριμένη πιθανότητα αντιστοιχεί στην κορυφή της εκ-των-υστέρων κατανομής των μοντέλων. Σε αρκετά σύγχρονα προβλήματα όπου ο όγκος των δεδομένων είναι αρκετά μεγάλος, το πλήθος των συγκρινόμενων μοντέλων, με τη σειρά του, είναι ανάλογα αυξημένο. Σημειώνουμε ότι σε αυτές τις περιπτώσεις ενδέχεται η μέγιστη εκ-των-υστέρων πιθανότητα που περιγράψαμε να είναι αρκετά μικρή. Αυτό συμβαίνει λόγω του φαινομένου της μοντελο-διάχυσης (model splitting) (++βλ. ΑΝΑΦΟΡΑ).

### 3.1.6 Δειγματολήπτης Gibbs

Στα περισσότερα πολυμεταβλητά προβλήματα, η εκ-των-υστέρων πιθανότητα (20) δεν μπορεί να υπολογιστεί αναλυτικά. Έτσι, η απευθείας προσομοίωση από την posterior κατανομή φαντάζει αδύνατη. Οι μέθοδοι Markov Chain Monte Carlo (MCMC) παίζουν κυρίαρχο ρόλο στην αντιμετώπιση τέτοιων δυσκολιών [Dellaportas and Roberts \(2003\)](#). Μέσω αυτών των αλγορίθμων, προσομοιώνουμε τιμές από μια Μαρκοβιανή αλυσίδα (Markov chain), όπου η στάσιμη κατανομή της είναι η ζητούμενη εκ-των-υστέρων κατανομή. Ο δειγματολήπτης Gibbs ανήκει σε αυτή την κατηγορία μεθόδων. Αρχικά θα αναφέρουμε κάποιους ορισμούς, ώστε να γίνει κατανοητή η λειτουργία των μεθόδων MCMC. Μία Μαρκοβιανή αλυσίδα χαρακτηρίζεται από, την κατανομή αρχικής κατάστασης  $p_0$  και τον πυρήνα μετάβασης  $K$ . Αν υποθέσουμε ότι το σύνολο  $\Theta$  είναι διακριτό, τότε ο πυρήνας μετάβασης είναι ένας πίνακας με στοιχεία που καθορίζονται από την εξής σχέση

$$k_{ij} = P(\theta_{t+1} = j | \theta_t = i), \quad (3.37)$$

όπου  $t$  η χρονική στιγμή παραγωγής μιας τιμής  $\theta$ . Πρόκειται για έναν στοχαστικό πίνακα όπου το άθροισμα των γραμμών του ισούται με την μονάδα. Επίσης, η κατανομή  $p$  λέγεται στάσιμη για την Μαρκοβιανή αλυσίδα εάν

$$\theta_t \sim p \Rightarrow \theta_{t+1} \sim p. \quad (3.38)$$

Αναφερόμενοι πάλι στην μεγάλη διάσταση, ο δειγματολήπτης Gibbs καταφέρνει να διαμερίζει ένα πολυδιάστατο πρόβλημα προσομοίωσης σε πολλά προβλήματα μικρότερης διάστασης. Έτσι, ο υπολογισμός των τελευταίων είναι εύκολα διαχειρήσιμος. Έστω, ότι θέλουμε να προσομοιώσουμε τιμές από μια πολυμεταβλητή κατανομή  $p(\theta|D)$ , όπου  $\theta = (\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)})$ . Ο δειγματολήπτης Gibbs, προσομοιώνει διαδοχικά από τις δεσμευμένες κατανομές  $\theta^{(i)}|\theta^{(-i)}$ , όπου  $\theta^{(-i)} = \{\theta^{(j)} : j \neq i\}$  (εάν βέβαια αυτό είναι εφικτό). Τότε, μέσω της διαδικασίας που περιγράψαμε, οι τιμές που θα παραχθούν θα έχουν στάσιμη κατανομή την  $p(\theta|D)$ . Παρακάτω δίνεται ο αλγόριθμος του δειγματολήπτη Gibbs σε μορφή ψευδοκώδικα.

---

**Algorithm 1:** Αλγόριθμος Δειγματολήπτη Gibbs

---

Αρχικοποίησε τιμές  $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)}, \dots, \theta_0^{(n)})$ ;

**for**  $t = 1, 2, 3, \dots, m$  **do**

    Προσομοίωσε  $\theta_t^{(1)} \sim p(\theta^{(1)} | \theta_{t-1}^{(2)}, \dots, \theta_{t-1}^{(n)})$ ;

    Προσομοίωσε  $\theta_t^{(2)} \sim p(\theta^{(2)} | \theta_t^{(1)}, \theta_{t-1}^{(3)}, \dots, \theta_{t-1}^{(n)})$ ;

    ⋮

    Προσομοίωσε  $\theta_t^{(n)} \sim p(\theta^{(n)} | \theta_t^{(1)}, \theta_t^{(2)}, \dots, \theta_t^{(n-1)})$ ;

**end**

---

Όταν το πλήθος των επαναλήψεων  $m \rightarrow \infty$ , ο αλγόριθμος θα συγκλίνει στην κατανομή στόχο (δηλαδή την εκ-των-υστέρων κατανομή της παραμέτρου ενδιαφέροντος) από οποιαδήποτε αρχική τιμή  $\theta_0$  του δώσουμε ως είσοδο. Για την σύγκλιση, η αλυσίδα που παράγεται, αρχικά θα πρέπει να περάσει από περιοχές με χαμηλή εκ-των-υστέρων πιθανότητα για να καταλήξει στη ζητούμενη κατανομή. Για τον λόγο αυτό, συνηθίζεται η εξαίρεση των αρχικών προσομοιωμένων τιμών από την ανάλυσή. Η περίοδος του αλγορίθμου όπου οι τιμές αφαιρούνται ονομάζεται προθέρμανση (burnin period).

Ο δειγματολήπτης Gibbs, μειώνοντας τη διάσταση του προβλήματος, καλείται να αντιμετωπίσει ένα ακόμα πρόβλημα, την αυτοσυσχέτιση. Είναι πιθανό, η αυτοσυσχέτιση της διαδικασίας σε ένα διαδοχικό βήμα να είναι αρκετά μεγάλη (ειδικότερα στην αρχή της διαδικασίας). Προφανώς, όσο μεγαλύτερη η αυτοσυσχέτιση διαδοχικών τιμών, τόσο πιο αργή είναι η εξέλιξη του αλγορίθμου. Σε αυτή την περίπτωση, μια τεχνική που χρησιμοποιείται είναι αυτή της λέπτυνσης της αλυσίδας (thinning). Ουσιαστικά, εξαιρούμε τιμές της προσομοίωσης με κάποιο βήμα λέπτυνσης. Για παράδειγμα εάν το βήμα είναι  $s$ , η τιμές που θα αποθηκεύουμε για κάθε διάσταση θα είναι  $\theta_t, \theta_{t+s}, \theta_{t+2s}$  κ.ο.κ. Ο σκοπός που εφαρμόζουμε thinning στην αλυσίδα είναι η εξοικονόμηση χώρου αποθήκευσης, καθώς και η επίτευξη ανεξαρτησίας μεταξύ των παραγόμενων τιμών. Ένα σημαντικό κομμάτι για την σύγκλιση του αλγορίθμου είναι η επιλογή των παραμέτρων της δεσμευμένης κατανομής, που για περισσότερες πληροφορίες παραπέμπουμε τον αναγνώστη στην εργασία των [Paspiliopoulos et al. \(2003\)](#).

### 3.2 Μπεϋζιανή ανάλυση εικόνας

Η Μπεϋζιανή συμπερασματολογία, στην οποία αναφερθήκαμε στην προηγούμενη ενότητα, αποτελεί την βάση για την Μπεϋζιανή Ανάλυση εικόνων. Αρχικά, θα προσαρμόσουμε την θεωρία που είδαμε στη συγκεκριμένη κατηγορία ανάλυσης. Σε αυτή την περίπτωση, το σύνολο  $\Theta$  αφορά τις παραμέτρους μιας πραγματικής εικόνας, την οποία θέλουμε να ανακτήσουμε. Ο αναγνώστης μπορεί να πληροφορηθεί περισσότερα για τα δεδομένα, καθώς και τα είδη των ιατρικών εικόνων που θα μας απασχολήσουν, στο Κεφάλαιο 4. Επίσης, το σέτ δεδομένων  $D$  αφορά υποβέλτιστες εικόνες, οι οποίες έχουν εκτεθεί σε θόρυβο. Η παρουσία θορύβου οδηγεί σε μείωση της ποιότητας, με αποτέλεσμα ορισμένα χαρακτηριστικά των εικόνων να αλλοιωθούν ή να χαθούν. Γενικά, η πληροφορία που χάνεται ενδέχεται να είναι πολύ σημαντική, το οποίο είναι αδιαμφισβήτητο όταν αναφερόμαστε σε ιατρικές εικόνες. Η καθαρότητα των εικόνων παίζει καθοριστικό ρόλο στην απόφαση ενός ιατρού για τον εντοπισμό επιθυμητών χαρακτηριστικών. Αντιπροσωπευτικό παράδειγμα αποτελεί η ανίχνευση ενός όγκου στον εγκέφαλο. Επίσης, η απόφαση του αρμοδίου εμπειρογνώμονα μπορεί να βασίζεται στην αναγνώριση του όγκου ως καλοήθους ή κακοήθους. Σε τέτοιες περιπτώσεις τα γεωμετρικά χαρακτηριστικά της εικόνας (για παράδειγμα αξονικής τομογραφίας) πρέπει να είναι ευδιάκριτα για τον ιατρό, ώστε να είναι σε θέση να ξεχωρίσει την υποκείμενη πάθηση.

Για να κάνουμε εύγλωττο τον συμβολισμό, θα αναφέρουμε ως  $x$  την αληθινή εικόνα (χωρίς θόρυβο και με τη βέλτιστη ποιότητα) και ως  $y$  την υποβέλτιστη εικόνα. Φυσικά μπορούμε να γενικεύσουμε την μεταβλητή  $y$  ένα σύνολο από υποβέλτιστες εικόνες (δεδομένα). Η Μπεϋζιανή ανάλυση, προσπαθεί μέσω κάποιας εκ-των-προτέρων κατανομής  $p(x)$  να συμπεριλάβει όσο περισσότερη πληροφορία για τα χωρικά χαρακτηριστικά της πραγματικής εικόνας. Η ελάττωση του θορύβου επιτυγχάνεται μέσω της συνάρτησης πιθανοφάνειας  $p(y|x)$ . Ο συγκερασμός των δύο προηγούμενων κατανομών οδηγεί στην εκ-των-υστέρων κατανομή  $p(x|y)$ , όπως είδαμε και στην σχέση (20). Με τον νέο συμβολισμό η σχέση (3.22) γίνεται

$$p(x|y) \propto p(y|x)p(x). \quad (3.39)$$

### 3.2.1 Τυχαία Πεδία Markov και εκ-των-προτέρων κατανομή

Η πιο διαδεδομένη επιλογή εκ-των-προτέρων κατανομής στην Μπεϋζιανή ανάλυση εικόνων είναι ένα τυχαίο πεδίο Markov (Markov Random Field ή MRF) Besag (1974) Geman and Geman (1984). Ένα MRF αποτελείται από ένα σύνολο τυχαίων μεταβλητών, οι οποίες υπακούουν στην Markovιανή Ιδιότητα (3.39) και επικοινωνούν μεταξύ τους μέσω ενός μη κατευθυνόμενου γράφου. Οι αλληλοσυσχετίσεις μεταξύ των πίξελ μιας εικόνας δημιουργούν μια κατανομή για το συγκεκριμένο σύνολο μεταβλητών. Είναι κατανοητό ότι γειτονικά πίξελ σχετίζονται στενότερα μεταξύ τους, παρά με άλλα που βρίσκονται αρκετά μακριά στην εικόνα. Φυσικά το εύρος της περιοχής της γειτονιάς των πίξελ μπορεί να ποικίλλει ανάλογα με τις εικόνες που θέλουμε να αναλύσουμε. Στην πραγματικότητα οι τυχαίες μεταβλητές, ή αλλιώς οι κόμβοι του γράφου, αποτελούν τις γειτονικές περιοχές που απαρτίζουν την συνολική εικόνα.

Ο ορισμός ενός MRF πάνω σε ένα σύνολο  $S$  από τέτοιες περιοχές δίνεται από τις κατανομές πιθανότητας των τιμών του κάθε πίξελ. Θα συμβολίσουμε  $x_s$  μια τυχαία μεταβλητή πάνω στην τοποθεσία  $s \in S$  (συνήθως ως τοποθεσία αναφερόμαστε σε ένα μόνο πίξελ) και  $\partial s$  την γειτονική περιοχή. Τότε, η εκ-των-υστέρων κατανομή της μεταβλητής  $x_s$  δωθέντων των υπόλοιπων πίξελ της εικόνας  $x_{-s}$  και των δεδομένων  $y$  μπορεί να γραφτεί ως

$$p(x_s|y, x_{-s}) \propto p(y_s|x_s)p(x_s|x_{\partial s}). \quad (3.40)$$

Από την παραπάνω σχέση παρατηρούμε ότι η εκ-των-υστερών δεσμευμένη κατανομή του  $x_s$  εξαρτάται από τα γειτονική περιοχή του. Η χρησιμότητα της εκ-των-προτέρων κατανομής ενός MRF, οφείλεται στην ιδέα της γειτονικότητας των πίξελ και κυρίως της αλληλοεξάρτησής τους. Θέλουμε η γειτονιά του  $x_s$  να είναι αρκούντως μικρή ώστε να συμπεριλαμβάνει όλη της απαραίτητη πληροφορία για το ίδιο το  $x_s$ . Τότε με μικρή υπολογιστική ισχύ μπορούμε να ανακτήσουμε χρήσιμες πληροφορίες από μια εικόνα μέσω της εν λόγω εκ-των-προτέρων κατανομής [Cross and Jain \(1983\)](#). Γενικότερα, οι μέθοδοι των τυχαίων πεδίων Markov μας δίνουν την δυνατότητα να βελτιστοποιήσουμε χαρακτηριστικά των εικόνων που ενδέχεται να έχουν αλλοιωθεί. Φυσικά, οι συντελεστές βαρύτητας που θα επιτρέψουμε να έχουν τα κάθε γειτονικά πίξελ  $\partial s$  ως προς το ζητούμενο  $x_s$ , πρέπει να επιλεγθούν προσεκτικά μέσω της εκ-των-προτέρων κατανομής.

Για λόγους πληρότητας ακολουθώντας παραθέτουμε την Μαρκοβιανή Ιδιότητα για την τυχαία μεταβλητή  $X_s$  που έχουμε αναφερθεί έως τώρα

$$P(X_s = x_s | X_i = x_i, \forall i) = P(X_s = x_s | X_i = x_i, i \in \partial s). \quad (3.41)$$

Η παραπάνω σχέση μας δείχνει ότι η πληροφορία της  $X_s$  προέρχεται αποκλειστικά από τα γειτονικά του πίξελ  $\partial s$  και όχι από τα υπόλοιπα της εικόνας.

Υπάρχουν διαφόρων μορφών εκ-των-προτέρων κατανομές, οι οποίες υιοθετούν την έννοια των τυχαίων πεδίων Markov, αλλά στην παρούσα ενότητα δεν έγινε λόγος για αυτές. Ενδεικτικά θα αναφέρουμε κάποιες από αυτές, όπως οι Γκαουσιανές, οι συνελκτικές και οι εκ-των-προτέρων κατανομές για ετικέτες (χρησιμοποιούνται για προβλήματα ταξινόμησης). Στην παρούσα ενότητα παρουσιάσαμε την γενική ιδέα των Τυχαίων Πεδίων Markov, καθώς και την διαισθητική λειτουργία τους. Παροτρύνουμε τον αναγνώστη για περισσότερες λεπτομέρειες στη δημοσίευση του [Besag \(1989\)](#).

### 3.2.2 Υπολογιστική εκτίμηση εικόνας κατά Bayes

Για την εκτίμηση της πραγματικής εικόνας  $x$ , οδηγούμαστε σε ένα πρόβλημα βελτιστοποίησης, που με την βοήθεια της Σχέσης (3.36) έχουμε

$$\hat{x} = \arg \max_{x \in X} P(x|y), \quad (3.42)$$

όπου  $X$  είναι το σύνολο με όλες τις δυνατές αναδομήσεις της εικόνας. Η ύπαρξη πολλών τοπικών μεγίστων, κάνει αρκετά δύσκολη την εύρεση του ολικού. Εξαιτίας της διακριτότητας του προβλήματος, το ολικό μέγιστο είναι και το βέλτιστο, αφού από την Θεωρία Αποφάσεων και την συνάρτηση κόστους 0-1, η απόφαση Bayes δίνεται από την σχέση (3.34). Έτσι, είτε η εικόνα θα είναι η αληθινή (πραγματικός στόχος της αναζήτησης) είτε όχι.

Η βασική μέθοδος για την ακριβή εκτίμηση της μέγιστης εκ-των-υστέρων πιθανότητας (MAP) (του ολικού μεγίστου) γίνεται μέσω του αλγορίθμου της προσομοιωμένης απόπτωσης (Simulated Annealing) η οποία πρόταθηκε από τους [Geman and Geman \(1984\)](#). Πρόκειται για ένα στοχαστικό επαναληπτικό αλγόριθμο, ο οποίος με την βοήθεια του δειγματολήπτη Gibbs, μπορεί να προσομοιώσει τιμές οι οποίες καταφέρνουν να ανοικοδομήσουν την υποβέλτιστη εικόνα. Ειδικότερα καταφέρνει να βρεί ένα τοπικό μέγιστο κοντά στο ολικό. Η εύρεση του ολικού είναι πρακτικά αδύνατη σε πεπερασμένο χρόνο, εξαιτίας του ότι η παράμετρος ελέγχου  $T$  πρέπει να μειωθεί λογαριθμικά (βλέπε Σχέση (3.43)). Για να γίνει πιο σαφής η εξάρτηση των δύο αλγορίθμων για το συγκεκριμένο πρόβλημα παραθέτουμε μια διαισθητική αναπαράσταση των λειτουργιών τους. Ο δειγματολήπτης Gibbs παράγει τιμές, ώστε να προσεγγίσει όσο το δυνατόν καλύτερα την κατανομή της εικόνας. Αφού μιλάμε για τοπικά χωρία μιας εικόνας και έχοντας υπόψη την εξάρτηση των γειτονικών στοιχείων με την προς εξέταση περιοχή  $x_s$ , τότε το κάθε βήμα παραγωγής τιμών του δειγματολήπτη Gibbs γίνεται μέσω των πλήρων δεσμευμένων κατανομών των γειτόνων και όχι ολόκληρης της εικόνας. Από την άλλη πλευρά η μέθοδος της προσομοιωμένης απόπτωσης διερευνεί στον χώρο αναζήτησης, την βέλτιστη αναπαράσταση του τοπικού αυτού χωρίου. Σύμφωνα με αυτή την προσέγγιση, αντί να προσομοιώνουμε τιμές από την εκ-των-υστέρων κατανομή  $p(x|y)$ , παίρνουμε δείγματα από μια άλλη συνάρτηση πιθανότητας

$$q_T(x|y) \propto (p(y|x)p(x))^{1/T}, \quad (3.43)$$

όπου  $T$  καλείται παράμετρος ελέγχου (ή θερμοκρασία). Όταν η  $T$  μειώνεται αργά κατά την διάρκεια της προσομοίωσης τότε η πιθανότητα εύρεσης της MAP τείνει στο 1. Στην εργασία των [Geman and Geman \(1984\)](#) μπορεί να βρεθεί μια λεπτομερέστερη ανάλυση της σύγκλισης του αλγορίθμου. Οφείλουμε να αναφέρουμε ότι η μεγάλη υπολογιστική πολυπλοκότητα της μεθόδου, μπορεί να αντιμετωπιστεί με παράλληλους υπολογισμούς στον επεξεργαστή του υπολογιστή που γίνεται η προσομοίωση. Όμως, ακόμα και σήμερα η δυνατότητα γρήγορης εκτίμησης της εκ-των-υστέρων πιθανότητας (3.40), με τη βοήθεια αριθμητικών μεθόδων, αποτελεί ένα δύσκολο πρόβλημα.

### 3.2.3 Κανονικοποίηση Tikhonov

Στην υποενότητα αυτή θα αναφερθούμε σε μια από τις βασικότερες Μπεϋζιανές μεθόδους ανάλυσης εικόνων, την κανονικοποίηση Tikhonov (Tikhonov Regularization). Για αυτή την μέθοδο θα θεωρήσουμε ότι η υποβέλτιστη εικόνα έχει επιπεδοποιηθεί (βλέπε Ενότητα § 2.1), δηλαδή δίνεται σε μορφή διανύσματος. Το πρόβλημα ανακατασκευής της τελευταίας, μπορεί να μοντελοποιηθεί από την παρακάτω εξίσωση

$$y = Hx + \epsilon. \quad (3.44)$$

όπου  $y$  η υποβέλτιστη εικόνα (διάστασης  $n \times 1$ ),  $x$  η πραγματική εικόνα χωρίς θόρυβο (διάστασης  $n \times 1$ ),  $H$  ένας γνωστός πίνακας (διάστασης  $n \times n$ ) και  $\epsilon$  ο προστιθέμενος θόρυβος. Η λειτουργία του πίνακα  $H$  μπορεί να ποικίλλει. Αναπαριστά την ομαλοποίηση, την θόλωση, ακόμα και την περιστροφή που μπορεί να έχει υποστεί η πραγματική εικόνα. Το ζητούμενο του προβλήματος είναι η εκτίμηση της πραγματικής εικόνας  $x$  δοθείσης της υποβέλτιστης εικόνας  $y$ . Σύμφωνα με την μέθοδο της κανονικοποίησης Tikhonov το πρόβλημα της εξίσωσης (3.44) μετατρέπεται στο παρακάτω πρόβλημα ελαχιστοποίησης

$$\arg \min_{x \in \mathbb{R}^{n \times 1}} \mathcal{G}(x) = \arg \min_{x \in \mathbb{R}^{n \times 1}} \left\{ \|y - Hx\|^2 + \lambda \|\Gamma x\|^2 \right\}, \quad (3.45)$$

όπου  $\|\cdot\|$  αναπαριστά την  $L_2$  νόρμα, και  $\Gamma$  γνωστός πίνακας (διάστασης  $n \times n$ ). Στην πραγματικότητα η σχέση ελαχιστοποίησης (3.45) μπορεί ερμηνευθεί ως η MAP εκτίμηση της εικόνας, όπου ο πρώτος όρος,  $\|y - Hx\|^2$ , αναπαριστά την πιθανοφάνεια των δεδομένων  $y$  και ο πίνακας  $\Gamma$  την εκ-των-προτέρων γνώση που έχουμε για την εικόνα  $x$ . Περισσότερες λεπτομέρειες για την σύνδεση της



μεθόδου με την Μπεϋζιανή ανάλυση θα εξετάσουμε στη συνέχεια. Θα μπορούσε κανείς να ορίσει τον  $\Gamma$ , ώστε να ποινικοποιεί τις αποστάσεις που έχει το κάθε πίξελ της εικόνας  $x$  με ένα σύνολο γειτονικών του.

Όπως παρατηρούμε το πρόβλημα ελαχιστοποίησης (3.45) απαρτίζεται από δύο όρους, τον  $\|y - Hx\|^2$  και τον  $\|\Gamma x\|^2$ . Ο πρώτος όρος συγκρίνει την εικόνα  $y$  με την  $x$  και συνεπώς καθορίζει την πιστότητα του μοντέλου με βάση τα διαθέσιμα δεδομένα (εικόνα  $y$ ). Όταν για κάποιο  $x \in \mathbb{R}^{n \times 1}$  η ποσότητα αυτή παίρνει μεγάλες τιμές τότε ίσως το συγκεκριμένο  $x$  να μην είναι αυτό που ψάχνουμε. Όμως, είναι προφανές ότι μπορούν να υπάρξουν άπειρα διανύσματα  $x$  για τα οποία η ποσότητα  $\|y - Hx\|^2$  είναι πολύ κοντά στο μηδέν. Σε τέτοιες περιπτώσεις η εκτίμηση για την εικόνα  $x$  τείνει να είναι ίδια με την υποβέλτιστη εικόνα  $y$ , γεγονός που δεν μας προσφέρει κάποια επιπλέον πληροφορία για την πραγματική εικόνα. Εδώ οφείλει την παρουσία του ο δεύτερος όρος. Ο όρος αυτός βοηθάει στην ομαλοποίηση της εκτίμησης που κάνουμε και ουσιαστικά ποινικοποιεί το αποτέλεσμα του πρώτου όρου, ώστε να μην γίνεται υπερπροσαρμογή του μοντέλου μας. Η παράμετρος  $\lambda$ , που συχνά αναφέρεται ως παράμετρος κανονικοποίησης, παίρνει θετικές τιμές και δίνει την σχετική βαρύτητα των δύο προαναφερθέντων όρων. Η επιλογή της υπερπαραμέτρου αυτής πρέπει να γίνεται με προσοχή, ώστε να αντισταθμίζει κατάλληλα την πιστότητα του μοντέλου στα δεδομένα αλλά και την ομαλότητα της εκτίμησης.

Η εκτίμηση που δίνει το μοντέλο (3.45) για την πραγματική εικόνα δίνεται σε κλειστή μορφή από την παρακάτω κανονική εξίσωση

$$(H^T H + \lambda \Gamma^T \Gamma)x = H^T y. \quad (3.46)$$

Παρατηρούμε ότι η παραπάνω σχέση δεν είναι τίποτα άλλο παρά ένα σύστημα γραμμικών εξισώσεων της μορφής  $Ax = b$ , όπου  $A = (H^T H + \lambda \Gamma^T \Gamma)$  (πίνακας διάστασης  $n \times n$ ) και  $b = H^T y$  (διάνυσμα διάστασης  $n \times 1$ ). Γίνεται αντιληπτό ότι η διάσταση του πίνακα  $A$  είναι ιδιαίτερα μεγάλη αφού ο αριθμός των στηλών και γραμμών του είναι ίσος με το συνολικό αριθμό των πίξελ της προς ανακατασκευή εικόνας. Για την λύση του γραμμικού προβλήματος αρκεί να λύσουμε την εξίσωση (3.46) ως προς  $x$ , αντιστρέφοντας τον πίνακα  $A$ . Αναλογιζόμαστε λοιπόν, ότι η αντιστροφή ενός τόσο μεγάλου πίνακα μπορεί να αυξήσει εκθετικά τον χρόνο υπολογισμού της εκτίμησης. Σε τέτοιες περιπτώσεις ιδανικές υπολογιστικές τεχνικές κρίνονται οι μέθοδοι κλίσεων, όπως ο αλγόριθμος της συζυγούς κλίσης (Conjugate Gradient) στον οποίον θα αναφερθούμε. Η μέθοδος αυτή οδηγεί στο εξής επαναληπτικό σχήμα

$$x_{k+1} \leftarrow x_k + \alpha_k \nabla C(x_k), \quad (3.47)$$

όπου  $C(x_k) = \frac{1}{2} \|y - Hx\|^2 + \frac{\lambda}{2} \|\Gamma x\|^2$  και  $\alpha_k$  μια θετική παράμετρος η οποία καθορίζει το μέγεθος του βήματος για την αναζήτηση του ολικού ελαχίστου. Επίσης, με  $e_k$  συμβολίζουμε την αρνητική κλίση του  $C$  στο σημείο  $x_k$ , δηλαδή την  $-\nabla C(x_k)$ . Σημειώνουμε εδώ ότι πολλαπλασιάσαμε τον πρώτο και τον δεύτερο όρο της σχέσης (3.45) με το  $\frac{1}{2}$  ώστε να μας βοηθήσει με τον υπολογισμό της κλίσης. Η θεώρησή μας αυτή δεν επηρεάζει το αποτέλεσμα του αλγορίθμου. Πρόκειται για μια επαναληπτική μέθοδο η οποία για έναν καθορισμένο αριθμό επαναλήψεων επαναπροσδιορίζει την ανακατασκευασμένη εικόνα έως ότου η διαφορά των εικόνων ανάμεσα σε δύο επαναλήψεις να είναι αρκούντως μικρή (δηλαδή να έχει προσεγγίσει αρκετά καλά το ολικό ελάχιστο). Συνήθως, ο μέγιστος αριθμός επαναλήψεων θεωρείται ίσος με το μήκος του διανύσματος  $x$ , δηλαδή με το πλήθος των πίξελ της εικόνας.

---

**Algorithm 2:** Αλγόριθμος Συζυγούς Βαθμίδας (Conjugate Gradient)

---

Αρχικοποίηση την τιμή  $x_0$  ως μια πρώτη προσέγγιση για την λύση του προβλήματος

$e_0 \leftarrow b - Ax_0;$

$p_0 \leftarrow e_0;$

$k \leftarrow 0;$

**while**  $e_{k+1} < \text{tolerance}$  **do**

$\alpha_k \leftarrow \frac{e_k^T e_k}{p_k^T A p_k};$

$x_{k+1} \leftarrow x_k + \alpha_k p_k;$

$e_{k+1} \leftarrow e_k - \alpha_k A p_k;$

$\beta_k \leftarrow \frac{e_{k+1}^T e_{k+1}}{e_k^T e_k};$

$p_{k+1} \leftarrow e_{k+1} + \beta_k p_k$

$k \leftarrow k + 1$

**end**

---

Στον Αλγόριθμο 2 δίνεται παραπάνω σε μορφή ψευδοκώδικα η μέθοδος συζυγούς κλίσης.

Τέλος, αξίζει να υπογραμμίσουμε ότι η μέθοδος της κανονικοποίησης Tikhonov μπορεί να αναπαρασταθεί ως μια πιθανοθεωρητική προσέγγιση και ειδικότερα σαν μια ειδική περίπτωση ενός τυχαίου πεδίου Markov. Στην πραγματικότητα, η μέθοδος αυτή επιτυγχάνει την εκτίμηση της μέγιστης εκ-των-υστέρων πιθανότητας. Ας αναλογιστούμε ότι η συνάρτηση  $\mathcal{G}(x)$  που θέλουμε να ελαχιστοποιήσουμε (βλέπε Σχέση (3.45)) αποτελεί τον αρνητικό λογάριθμο της εκ-των-υστέρων κατανομής. Για να γίνει αντιληπτή η σύνδεση των δύο τεχνικών, αρκεί να επαναπροσδιορίσουμε

τους ρόλους των δύο όρων της συνάρτησης  $\mathcal{G}$ . Μπορεί να θεωρήσει κανείς ότι ο πρώτος όρος των ελαχίστων τετραγώνων είναι η πιθανοφάνεια κανονικής κατανομής και ο δεύτερος όρος, της κανονικοποίησης, είναι η εκ-των-προτέρων πεποίθηση που έχουμε για την πραγματική εικόνα.

### 3.3 Ανάλυση Fourier

Σε αυτή την ενότητα θα επικεντρωθούμε σε ένα πολύτιμο εργαλείο για την μέθοδο ανάλυσης εικόνας που θα χρησιμοποιήσουμε (βλέπε Ενότητα § 3.4), τον μετασχηματισμό Fourier. Αξίζει να αναφέρουμε ότι αποτέλεσε έναν από τους σημαντικότερους αλγορίθμους του 20ού αιώνα, πάνω στον οποίο, ακόμα και σήμερα, στηρίζεται η τεχνολογία επεξεργασίας ψηφιακών σημάτων. Μετατρέπει τις εικόνες σε συχνότητες, και λειτουργεί ως μέσο για την συμπίεσή τους. Έτσι, η αποθήκευση των εικόνων γίνεται οικονομικότερη και η ανάκτησή τους ευκολότερη. Για την διαδικασία ανάκτησης, χρησιμοποιούμε τον αντίστροφο μετασχηματισμό Fourier, ο οποίος μας δίνει την αρχική αναπαράστασή της στο διακριτό χώρο. Μπορούμε να φανταστούμε μια ασπρόμαυρη εικόνα ως μια συνάρτηση δύο διαστάσεων των συντεταγμένων ενός επιπέδου  $x, y$ , δηλαδή

$$g \rightarrow g(x, y), \quad (3.48)$$

όπου  $g$  περιγράφει μια επιφάνεια, η οποία παίρνει μεγάλες τιμές στα μέρη που η εικόνα πλησιάζει το λευκό χρώμα, και μικρές σε αυτά που πλησιάζει το μαύρο. Στην περίπτωση που η εικόνα είναι έγχρωμη τότε σε κάθε τιμή του επιπέδου η συνάρτηση  $g$  παίρνει τρεις τιμές, μια για κάθε ένα κανάλι των βασικών χρωμάτων (RGB).

Στην περίπτωση της μιας διάστασης του μετασχηματισμού Fourier, τα δεδομένα είναι μια ακολουθία αριθμών  $f_0, f_1, \dots, f_n$ , την οποία μπορούμε να περιγράψουμε με ένα διάνυσμα  $f$ . Το ζητούμενο είναι η μετατροπή του  $f$  σε ένα διάνυσμα ίσου μήκους, το οποίο είναι μιγαδική συνάρτηση συχνότητας. Η διαδικασία αυτή επιτυγχάνεται μέσω ενός πίνακα  $F$  (πίνακας Fourier). Οφείλουμε να αναφέρουμε, ότι υπάρχουν δύο είδη μετασχηματισμών Fourier, ο συνεχής και ο διακριτός. Στην παρούσα ενότητα θα αναφερθούμε μόνο στον δεύτερο, αφού τα δεδομένα μας είναι διακριτές εικόνες.

### 3.3.1 Μιγαδικοί αριθμοί

Η μιγαδική συνάρτηση συχνοτήτων που παίρνουμε με τον μετασχηματισμό Fourier, γεννάει και το κυριότερο πρόβλημα που έχουμε να αντιμετωπίσουμε, την μοντελοποίηση. Για τον λόγο αυτό κρίνεται σημαντική μια αναφορά στους μιγαδικούς αριθμούς [Smith \(2002\)](#). Θα αναλύσουμε την αξιολογική εξίσωση

$$w = \cos \theta + i \sin \theta, \quad (3.49)$$

η οποία αποτελείται από ημίτονα, συνημίτονα και εκθετικές δυνάμεις. Ξεκινώντας θέλουμε να περιγράψουμε τα βασικά χαρακτηριστικά ενός μιγαδικού αριθμού, καθώς και τους τρόπους αναπαράστασής του. Ένας μιγαδικός αριθμός είναι της μορφής

$$z = a + ib, \quad (3.50)$$

όπου  $a$  είναι το πραγματικό μέρος (συμβολίζεται με  $Re(z)$ ),  $b$  είναι το φανταστικό μέρος (συμβολίζεται με  $Im(z)$ ), όπου  $i = \sqrt{-1}$  (θεμελιώδης ιδιότητα των μιγαδικών αριθμών). Ο αριθμός  $z$  μπορεί να αναπαρασταθεί σε μια παραλλαγμένη μορφή του καρτεσιανού επιπέδου (μιγαδικό επίπεδο), με το πραγματικό μέρος να διατρέχει τον οριζόντιο άξονα και το φανταστικό μέρος τον κάθετο. Συμβολίζουμε με  $|z|$  την απόσταση του σημείου  $z$  από την αρχή των αξόνων (modulus). Μπορούμε να θεωρήσουμε ότι το  $\vec{z}$  είναι ένα διάνυσμα με αρχή το  $(0,0)$  και πέρας το εν λόγω σημείο. Τότε, το  $|z|$  είναι το μήκος του διανύσματος  $\vec{z}$ . Επίσης,  $\theta$  είναι η γωνία (στην ορολογία της φυσικής αναγράφεται ως γωνία φάσεως, την οποία θα υιοθετήσουμε) που σχηματίζει το διάνυσμα  $\vec{z}$  με τον πραγματικό άξονα (argument). Τα  $|z|$  και  $\theta$  υπολογίζονται από

$$|z| = \sqrt{a^2 + b^2}, \quad \theta = \arctan\left(\frac{b}{a}\right). \quad (3.51)$$

Η παραπάνω αναπαράσταση του μιγαδικού αριθμού  $z$  είναι με το σύστημα πολικών συντεταγμένων. Αποτελεί μια εύχρηστη αναπαράσταση ενός μιγαδικού αριθμού, η οποία διευκολύνει τους υπολογισμούς σε διάφορα προβλήματα. Σύμφωνα με αυτή, ο  $z$  μπορεί να γραφτεί ως

$$z = |z|(\cos(\theta) + i \sin(\theta)). \quad (3.52)$$

Ουσιαστικά, με τις παραμέτρους  $|z|$  και  $\theta$  μπορούμε να διαγράψουμε ένα κύκλο στο μιγαδικό επίπεδο, όπου  $|z|$  η ακτίνα του. Χάρην απλότητας θα υποθέσουμε ότι  $|z| = 1$ . Δηλαδή θα επικεντρώσουμε την ανάλυσή μας στον μοναδιαίο κύκλο που περιγράφεται από την εξίσωση  $\cos^2(\theta) + \sin^2(\theta) = 1$ .

Επίσης, όπως θα δούμε στην επόμενη υποενότητα, ο πίνακας Fourier έχει ως στοιχεία του δυνάμεις μιγαδικών αριθμών. Συνεπώς, θέλουμε να εξετάσουμε πώς υπολογίζουμε την  $n$ -οστή δύναμη του  $z$ . Το ημίτονο και το συνημίτονο δημιουργούν μια μιγαδική εκθετική δύναμη (γνωστή και ως φόρμουλα του Euler), ακτίνας  $|z| = 1$  και γωνίας φάσεως  $\theta$  ως εξής

$$e^{i\theta} = \cos(\theta) + i \sin(\theta). \quad (3.53)$$

Οι πράξεις του πολλαπλασιασμού και της διαίρεσης δυνάμεων που ισχύουν για τους πραγματικούς αριθμούς, εξακολουθούν να ισχύουν κι όταν οι εκθέτες είναι μιγαδικοί. Άρα, η σχέση (3.49) μπορεί να λυθεί μέσω της

$$z^k = e^{ik\theta}. \quad (3.54)$$

Έτσι, συνδιάζοντας τις σχέσεις (3.53), (3.54) όπου  $k = \frac{2\pi}{n}$ , θεωρούμε την σχέση (3.49) ως

$$e^{i\frac{2\pi}{n}} = \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right). \quad (3.55)$$

### 3.3.2 Διακριτός Μετασχηματισμός Fourier

Ο Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform ή DFT) που θα χρησιμοποιήσουμε είναι αυτός των δύο διαστάσεων. Όπως θα δούμε παρακάτω, η γενίκευση από την μια διάσταση είναι άμεση. Γι' αυτό τον λόγο θα εξετάσουμε λεπτομερώς τον μετασχηματισμό στην μια διάσταση, ώστε να τον κατανοήσουμε σε βάθος [Strang et al. \(1993\)](#), [Smith \(2002\)](#).

Όπως αναφέραμε και στην αρχή της ενότητας, θεωρούμε τα δεδομένα ως ένα διάνυσμα μήκους

$N$ , που θα το συμβολίζουμε με  $f$ . Ο DFT μετατρέπει τα δεδομένα αυτά σε διάνυσμα συχνοτήτων  $\mathcal{F}$  ίσου μήκους, μέσω την εξής σχέσης

$$\mathcal{F}(k) = \sum_{n=0}^{N-1} f(n)e^{-i2\pi nk/N}, \quad (3.56)$$

για  $k = 0, 1, 2, \dots, N-1$ . Ουσιαστικά τα στοιχεία του διανύσματος  $\mathcal{F}$  περιέχουν τα αντίστοιχα των δεδομένων στον χώρο συχνοτήτων και ονομάζονται συντελεστές του DFT. Η σχέση (3.56) μπορεί να γραφτεί και σαν πράξη πινάκων ως εξής,

$$\mathcal{F} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{N-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{N-1} & w^{2(N-1)} & \cdots & w^{(N-1)^2} \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_{N-1} \end{bmatrix} = Wf, \quad (3.57)$$

όπου  $W$ , ο  $N \times N$  πίνακας Fourier με στοιχεία  $W_{n,m} = e^{-i2\pi nm/N}$ , για  $n, m = 0, 1, 2, \dots, N-1$ . Παρατηρούμε ότι ο πίνακας αυτός είναι συμμετρικός αφού  $W = W^T$ .

Τώρα υποθέτοντας ότι έχουμε τους συντελεστές DFT στο χώρο συχνοτήτων και θέλουμε να πάρουμε την αρχική μορφή των δεδομένων (στον διακριτό χώρο), χρησιμοποιούμε τον αντίστροφο μετασχηματισμό Fourier (Inverse Discrete Fourier Transform ή IDFT). Το μόνο που έχουμε να κάνουμε είναι να αντιστρέψουμε τον πίνακα  $W$  και να το διαιρέσουμε με το πλήθος των στοιχείων του διανύσματος  $\mathcal{F}$ , έτσι ώστε να ανακτήσουμε τα δεδομένα  $f^* = \mathcal{F}^{-1}$ . Δηλαδή ο πίνακας  $W^{-1}$  κατασκευάζεται από τις δυνάμεις του  $w^{-1} = 1/w$ , άρα

$$W^{-1} = \frac{1}{N} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w^{-1} & w^{-2} & \cdots & w^{-(N-1)} \\ 1 & w^{-2} & w^{-4} & \cdots & w^{-2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{-(N-1)} & w^{-2(N-1)} & \cdots & w^{-(N-1)^2} \end{bmatrix} \quad (3.58)$$

Συνεπώς η σχέση (3.57) γίνεται

$$\mathcal{F}^{-1} = f^* = \frac{1}{N} W^{-1} \mathcal{F}. \quad (3.59)$$

Από τις σχέσεις (3.57) και (3.59) προκύπτει ότι

$$W^{-1}W = NI, \quad (3.60)$$

όπου  $I$  ο ταυτοτικός  $N \times N$  πίνακας. Άρα, ο πίνακας  $W$  είναι ορθογώνιος, αφού οι στήλες του είναι ορθογώνιες.

Επιπρόσθετα, το κόστος υπολογισμού του μετασχηματισμού Fourier παίζει πρωτεύων ρόλο για την πρακτική υλοποίηση του. Η υπολογιστική πολυπλοκότητα είναι  $O(N^2)$ , εξαιτίας του πολλαπλασιασμού του πίνακα  $W$  με το διάνυσμα  $f$ . Με μια παραλλαγή του DFT, τον ταχύ μετασχηματισμό Fourier, η υπολογιστική πολυπλοκότητα μειώνεται σημαντικά σε  $O(N \log N)$ . Στο θέμα αυτό θα αναφερθούμε λεπτομερώς στον τελευταίο στην Ενότητα § 3.3.3).

Στις δύο διαστάσεις η διαδικασία είναι παρόμοια. Μόνο που αυτή την φορά, τα δεδομένα είναι στον διακριτό χώρο με μορφή πίνακα. Αν υποθέσουμε ότι η είσοδος είναι μια εικόνα  $N \times M$  τότε και ο μετασχηματισμός Fourier θα δώσει ένα πίνακα ίδιας διάστασης. Ο τύπος υπολογισμού του μετασχηματισμού Fourier είναι ανάλογη με αυτόν της μίας διάστασης και δίνεται από την ακόλουθη σχέση

$$\mathcal{F}(u, v) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f(n, m) e^{-i(2\pi un/N + i2\pi vm/M)}, \quad (3.61)$$

όπου  $f(n, m)$  είναι η εικόνα εισόδου (πίνακας διάστασης  $N \times M$ ), για κάθε πίξελ που δίνεται από τις συντεταγμένες  $(n, m)$ . Η μετασχηματισμένη εικόνα είναι η  $\mathcal{F}(u, v)$  (πίνακας διάστασης  $N \times M$ ), για  $u = 0, 1, 2, \dots, N-1$  και  $v = 0, 1, 2, \dots, M-1$ . Ο DFT μπορεί να περιγραφεί ως η διάσπαση μιας εικόνας στα επιμέρους εκθετικά ημίτονα και συνημίτονα. Η εξίσωση (3.61) μπορεί να γραφεί σαν συνδιασμός δύο μονοδιάστατων μετασχηματισμών Fourier, όπου η μία αφορά τις γραμμές και η άλλη τις στήλες του πίνακα εισόδου (των πίξελ). Δηλαδή, η σχέση (3.61) μπορεί να γραφτεί και ως

$$\mathcal{F}(u, v) = \sum_{n=0}^{N-1} \left( \sum_{m=0}^{M-1} f(n, m) e^{-i2\pi um/N} \right) e^{-i2\pi vn/M}. \quad (3.62)$$

Αντίστοιχα, ο αντίστροφος μετασχηματισμός Fourier δίνεται από τον ακόλουθο τύπο.

$$f(n, m) = \frac{1}{MN} \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} \mathcal{F}(u, v) e^{(i2\pi un/N + i2\pi vm/M)} \quad (3.63)$$

Όπως και στον μετασχηματισμό στην μία διάσταση μας ενδιαφέρει και πάλι το κόστος υπολογισμού του αλγορίθμου DFT. Λόγω τις διαχωρισιμότητας δύο μονοδιάστατων DFT σε γραμμές και στήλες, έχουμε ότι η υπολογιστική πολυπλοκότητα είναι  $\mathcal{O}((NM)^2 + (NM)^2)$ , δηλαδή και πάλι  $\mathcal{O}((NM)^2)$ . Στην περίπτωση που η εικόνα εισόδου και εξόδου είναι τετράγωνα, διάστασης  $N \times N$ , τότε η πολυπλοκότητα θα είναι  $\mathcal{O}(N^4)$ .

### 3.3.3 Fast Fourier Transform

Αναλογιζόμενοι τις μεγάλες διαστάσεις που μπορεί να πάρει μια εικόνα, ο άμεσος υπολογισμός του DFT και του αντιστρόφου του, δεν αποτελεί, υπολογιστικά, έναν οικονομικό αλγόριθμο. Έτσι, η ανάγκη για ένα νέο τρόπο υπολογισμού του κρίνεται επιτακτική. Ένας αλγόριθμος που μειώνει αρκετά την πολυπλοκότητα είναι ο ταχύς μετασχηματισμός Fourier (Fast Fourier Transform ή FFT) [Smith \(2019\)](#). Η πολυπλοκότητα που επιτυγχάνεται μέσω αυτού του αλγορίθμου είναι  $\mathcal{O}(2N \log_2 N)$ . Εξαιτίας της διάσπασης της διαδικασίας σε δύο μονοδιάστους μετασχηματισμούς, έναν για τις γραμμές και έναν για τις στήλες του πίνακα εισόδου (αρχική εικόνα), αρκεί να αναλύσουμε τον αλγόριθμο FFT σε μια διάσταση. Η ιδέα προήλθε από τους Cooley και Tukey στα μέσα της δεκαετίας του 1960 [Cooley and Tukey \(1965\)](#), και μερικές φορές αναφέρεται ως Cooley-Tukey αλγόριθμος. Αξίζει να σημειώσουμε ότι έχουν εφευρεθεί νέες πιο προχωρημένες μέθοδοι του αλγορίθμου FFT [Vetterli et al. \(1984\)](#), [Kolba and Parks \(1977\)](#).

Η ιδέα των Cooley και Tukey βασίζεται στην κατηγορία αλγορίθμων διαίρει και βασίλευε. Ειδικότερα, αναφέρεται στην αναπαράσταση ενός DFT με διάνυσμα εισόδου μήκους  $N$ , σε δύο αντίστοιχους μετασχηματισμούς με διανύσματα μήκους  $N/2$  ο καθένας. Ας υποθέσουμε ότι το  $N$  είναι δύναμη του 2, δηλαδή  $2^j = N$ , για κάποιο  $j$  ακέραιο. Τώρα, το γενικό πρόβλημα απλοποιείται σε δύο διαφορετικά προβλήματα. Ο πρώτος μετασχηματισμός αφορά τα στοιχεία του διανύσματος  $f$  με άρτιους κάτω δείκτες, και ο δεύτερος με τους περιττούς. Ακολουθως



δίνουμε την μετατροπή του DFT, στηριζόμενοι στη σχέση (3.56)

$$\mathcal{F}(k) = \sum_{n=0}^{N-1} f(n)e^{-i2\pi nk/N} \quad (3.64)$$

$$= \sum_{n=0}^{\frac{N}{2}-1} f(2n)e^{-i2\pi(2n)k/N} + \sum_{n=0}^{\frac{N}{2}-1} f(2n+1)e^{-i2\pi(2n+1)k/N} \quad (3.65)$$

$$= \sum_{n=0}^{\frac{N}{2}-1} f(2n)e^{-i2\pi nk/\frac{N}{2}} + e^{-i2\pi k/N} \sum_{n=0}^{\frac{N}{2}-1} f(2n+1)e^{-i2\pi nk/\frac{N}{2}} \quad (3.66)$$

$$= \mathcal{F}^{\text{even}}(k) + z^k \mathcal{F}^{\text{odd}}(k), \quad (3.67)$$

όπου από την σχέση (3.54) έχουμε ότι  $z^k = e^{-i2\pi k/N}$  (εδώ έχουμε γωνία φάσης  $\theta = -2\pi/N$ ). Επίσης, συμβολίζουμε ως  $\mathcal{F}^{\text{even}}$  τον μετασχηματισμό Fourier με τα στοιχεία των άρτιων υποδεικτών του διανύσματος  $f$  και ως  $\mathcal{F}^{\text{odd}}$  αυτόν με τα στοιχεία των περιττών.

Παρατηρούμε ότι η μετασχηματισμός  $\mathcal{F}$  είναι περιοδικός ως προς το μήκος του. Αυτό μπορεί εύκολα να αποδειχθεί υπολογίζοντας την  $\mathcal{F}(k+N)$ , όπου

$$\mathcal{F}(k+N) = \sum_{n=0}^{N-1} f(n)e^{-i2\pi n(k+N)/N} = \sum_{n=0}^{N-1} f(n)e^{-i2\pi nk/N} e^{-i2\pi k} = \mathcal{F}(k), \quad (3.68)$$

αφού  $e^{-i2\pi k} = 1$  από την σχέση (3.53). Άρα συμπαιρνουμε ότι,

$$\mathcal{F}^{\text{even}}(k) = \mathcal{F}^{\text{even}}(k+N/2) \quad \text{και} \quad \mathcal{F}^{\text{odd}}(k) = \mathcal{F}^{\text{odd}}(k+N/2), \quad (3.69)$$

όπου  $k = 0, 1, 2, \dots, N/2$ . Επίσης,

$$z^{k+N} = -z^k, \quad k = 0, 1, 2, \dots, N/2. \quad (3.70)$$

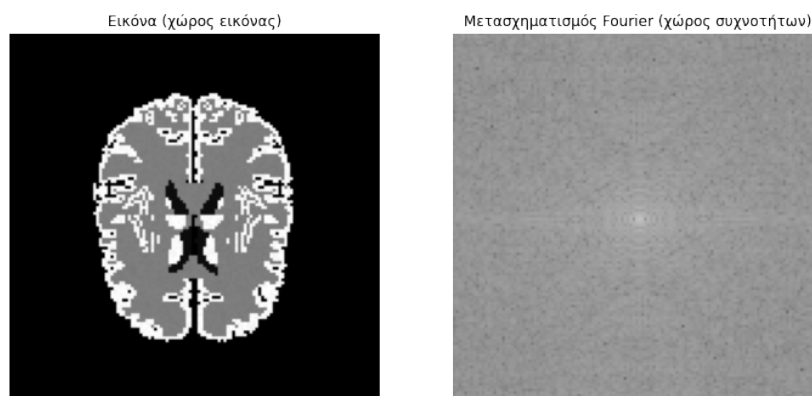
Συνδιάζοντας τις σχέσεις (3.69) και (3.70) παίρνουμε ότι

$$\mathcal{F}(k+N/2) = \sum_{n=0}^{N-1} f(n)e^{-i2\pi nk/N} = \mathcal{F}^{\text{even}}(k) - z^k \mathcal{F}^{\text{odd}}(k). \quad (3.71)$$

Συνεπώς, από τις σχέσεις (3.67) και (3.71) μπορούμε να υπολογίζουμε ταυτόχρονα δύο στοιχεία του  $\mathcal{F}$  μειώνοντας τον χρόνο υπολογισμού στο μισό. Αν συνεχίσουμε να διαιρούμε τους δύο νέους μετασχηματισμούς σε επιμέρους, φτάνουμε να έχουμε  $N$  διαφορετικούς διακριτούς μετασχηματισμούς Fourier. Όπως γίνεται αντιληπτό, η τελευταία πρόταση ισχύει στην περίπτωση που το  $N$  είναι δύναμη του 2. Έτσι, το πρόβλημα μας χωρίστηκε σε  $N$  απλούς υπολογισμούς σημείων. Επιπλέον, διαπιστώνουμε ότι ο μετασχηματισμός ενός μόνο σημείου αντιστοιχεί σε ένα μοναδικό συνδιασμό  $\log_2 N$  άρτιων και περιττών στοιχείων. Έτσι, η πολυπλοκότητα του προβλήματος είναι ίση με  $\mathcal{O}(N \log_2 N)$ . Γενικεύοντας στις δύο διαστάσεις έχουμε προφανώς  $\mathcal{O}(2N \log_2 N)$ , όπως υπογραμμίσαμε στην αρχή της υποενότητας.

Τέλος, θα αναφέρουμε μια σημαντική ιδιότητα η οποία είναι άμεση απόρροια της περιοδικότητας του σήματος που αναφέραμε προηγουμένως. Ο λόγος γίνεται για την ονομαζόμενη ως ιδιότητα της μετατόπισης (shift property). Με βάση την παραπάνω θεώρηση μας όπου  $n = 0, 1, 2, \dots, N - 1$ , τα κύρια χαρακτηριστικά μιας εικόνας συγκεντρώνονται στις γωνίες της μετά τον μετασχηματισμό Fourier. Εφαρμόζοντας την ιδιότητα της μετατόπισης, συγκεντρώνουμε την μεγαλύτερη πληροφορία (μεγάλες συχνότητες) στο κέντρο της εικόνας, με συντεταγμένες  $(0,0)$  στην μετασχηματισμένη εικόνα. Η ιδιότητα αυτή θεωρεί ένα  $\Delta \in \mathbb{Z}$  το οποίο αναφέρεται στην χρονική καθυστέρηση που κάνει ο αλγόριθμος για να πάρει δείγμα από το κανονική εικόνα. Εξαιτίας της περιοδικότητας, το μήκος του δυνύσματος  $z$  (ή modulo, αναφερόμενοι στον συμβολισμό που δώσαμε στην Ενότητα § 3.3.1) παραμένει αναλλοίωτο. Για τον λόγο αυτό, στην ανάλυση για τον δισδιάστατο μετασχηματισμό Fourier, αντί για  $n, m = 0, 1, 2, \dots, N - 1$  (όπου η εικόνα είναι διάστασης  $N \times N$ ), θα θέσουμε μια αλλαγή στις ετικέτες των σημάτων (στον τρόπο με τον οποίο τα λαμβάνουμε υπόψη), όπου  $\Delta = -N/2$ , ώστε  $n, m = -N/2, \dots, 0, 1, \dots, N/2 - 1$ . Συνεπώς πλέον, όσο πιο κοντά βρίσκεται το σήμα  $(n, m)$  στο κέντρο  $(0, 0)$  τόσο μεγαλύτερες είναι οι συχνότητες του σήματος αυτού.

Το Σχήμα 3.9 αποτελεί ένα καλό παράδειγμα απεικόνισης μιας εικόνας στον χώρο Fourier (χώρος συχνοτήτων). Προσομοιώνουμε μια εικόνα, διάστασης  $128 \times 128$ , ενός εγκεφάλου και την μετασχηματίζουμε μέσω του ταχύ μετασχηματισμού Fourier δύο διαστάσεων, στον χώρο συχνοτήτων. Έχοντας εφαρμόσει την ιδιότητα της μετατόπισης, παρατηρούμε ότι στο κέντρο της εικόνας συγκεντρώνονται οι υψηλές συχνότητες. Οι σκούρες περιοχές προσδιορίζουν χαμηλές συχνότητες, δηλαδή μικρές τιμές της ακτίνας του μετασχηματισμού Fourier (και άρα όχι τόσο πληροφορία). Όσο απομακρυνόμαστε από το κέντρο, οι συχνότητες μειώνονται με αποκορύφωμα



Σχήμα 3.9: Απεικόνιση ενός εγκεφάλου στον χώρο εικόνας και στον χώρο συχνοτήτων έπειτα από την εφαρμογή του διακριτού μετασχηματισμού Fourier

τις γωνίες. Στην πραγματικότητα, επειδή το φάσμα συχνοτήτων είναι αρκετά μεγάλο, αυτό που φαίνεται στην εικόνα είναι ο λογάριθμός του. Έτσι, μπορούμε να παρατηρήσουμε περισσότερες λεπτομέρειες, αφού στην πρώτη περίπτωση, θα βλέπαμε ένα υπερβολικά φωτεινό σημείο στο κέντρο και η υπόλοιπη εικόνα θα ήταν σχεδόν μαύρη [Kulkarni \(2000\)](#).

### 3.4 Μπεϋζιανή ανάλυση εικόνας στο χώρο Fourier

Η μέθοδος που θα υλοποιήσουμε για την ανάλυση ιατρικών εικόνων ονομάζεται Μπεϋζιανή ανάλυση εικόνας στον χώρο Fourier (Bayesian Image analysis In Fourier Space ή BIFS), και είναι εμπνευσμένη από την ερευνητική εργασία [Kornak et al. \(2020\)](#). Η κλασικές μέθοδοι που χρησιμοποιούνται για τέτοιου τύπου δεδομένα (βλέπε Κεφάλαιο 4) είναι τα τυχαία πεδία Markov τα οποία εξετάσαμε στην Ενότητα § 3.2. Στην πράξη η υλοποίηση τέτοιων μοντέλων απαιτούν υπολογιστικά ακριβές επαναληπτικές μεθόδους κατά την διαδικασία της εκτίμησης. Επίσης, η μοντελοποίηση τους διαφέρει ανάλογα με την διάσταση των προς ανάλυση εικόνων, γεγονός που μας αναγκάζει να προσαρμόζουμε κάθε φορά διαφορετικό μοντέλο. Έχουμε ήδη αναφέρει ότι η εκ-των-προτέρων κατανομή βάση των MRF εξαρτάται από τα γειτονικά πίξελ, που σχετίζονται με την προς εξέταση περιοχή, μέσω δεσμευμένων κατανομών. Έπεται λοιπόν, ότι και οι εκ-των-προτέρων κατανομές στον χώρο της εικόνας θα είναι μεταξύ τους συσχετισμένες.

Η μέθοδος BIFS αντιπαρέρχεται τέτοιες δυσκολίες, μεταφέροντας την ανάλυση στον χώρο Fourier. Διευκρινιστικά αναφέρουμε ότι θα ονομάζουμε χώρο εικόνας τον διακριτό χώρο των πίξελ και

χώρο Fourier την μετασχηματισμένη εικόνα σε συχνότητες. Το πλεονέκτημα που μας δίνεται στον χώρο Fourier είναι η ανεξαρτησία των εν λόγω κατανομών στις διαφορετικές περιοχές που τον διαμορφώνουν. Πλέον οι εικόνες θα μετατρέπονται σε συχνότητες μέσω του αλγορίθμου FFT όπου αναλύσαμε στην Ενότητα § 3.3. Οι χωρικά συσχετισμένες Γκαουσιανές Διαδικασίες (με τις οποίες συνήθως μοντελοποιούμε τις εκ-των-προτέρων κατανομές) μετατρέπονται σε μη-συσχετισμένες κανονικές τυχαίες μεταβλητές στον χώρο Fourier Zeger (1985)). Βέβαια, η παραπάνω πρόταση γενικεύεται για οποιαδήποτε οικογένεια κατανομών χωρικά συσχετισμένων διαδικασιών Peligrad et al. (2006). Έτσι, το πολυδιάστατο πρόβλημα κατακερματίζεται σε μια σειρά από μονοδιάστατα, που το καθένα αφορά μια περιοχή της εικόνας στον χώρο Fourier. Η μείωση της πολυπλοκότητας του προβλήματος οδηγεί αφενώς σε ταχύτερη υπολογιστική ικανότητα και αφετέρου σε καλύτερη συμπερασματολογία.

### 3.4.1 Μοντελοποίηση

Ο συμβολισμός που θα ακολουθήσουμε είναι ίδιος με αυτόν που επισημάναμε στην αρχή την Ενότητας § 3.2. Δηλαδή, θεωρούμε ως  $y$  το σύνολο υποβέλτιστων εικόνων που θα μας βοηθήσουν να ανακτήσουμε την πραγματική εικόνα  $x$ . Εφαρμόζοντας τον διακριτό μετασχηματισμό Fourier σε όλες τις εικόνες (και στα δεδομένα  $y$  και στην προς ανακατασκευή εικόνα  $x$ ) ο συμβολισμός στον χώρο Fourier αλλάζει σε  $\mathcal{F}y$  και  $\mathcal{F}x$  αντίστοιχα. Υπογραμμίζουμε ότι ο μετασχηματισμός γίνεται με μετατόπιση ώστε η πληροφορία να συγκεντρώνεται στο κέντρο της μετασχηματισμένης εικόνας. Επιπρόσθετα, αντί για τον συμβολισμό των κατανομών πυκνότητας πιθανότητας ως  $p$  τώρα θα χρησιμοποιούμε το  $\pi$ , ώστε να διαχωρίσουμε τους χώρους ανάλυσης ( $p$  στον χώρο εικόνας και  $\pi$  στον χώρο Fourier). Έτσι από την Σχέση 3.22, αφού εφαρμόσουμε το Θεώρημα του Bayes, παίρνουμε

$$\pi(\mathcal{F}x|\mathcal{F}y) \propto \pi(\mathcal{F}x)\pi(\mathcal{F}y|\mathcal{F}x). \quad (3.72)$$

Η εύρεση της εκ-των-υστέρων κατανομής της αληθινής εικόνας προσδιορίζεται απευθείας μέσω της πιθανοφάνειας και την εκ-των-προτέρων κατανομής. Οι τελευταίες ορίζονται στην αρχή της διαδικασίας από ένα σύνολο ανεξάρτητων κατανομών για κάθε μια περιοχή στον χώρο Fourier. Αν θεωρήσουμε μια περιοχή (διακριτό σημείο στον χώρο Fourier)  $k = (k_x, k_y)$  στον χώρο συχνοτήτων τότε η πλήρης δεσμευμένη εκ-των-υστέρων κατανομή της εικόνας στην συγκεκριμένη περιοχή ορίζεται από την σχέση

$$\pi(\mathcal{F}x_k|\mathcal{F}y) \propto \pi(\mathcal{F}x_k)\pi(\mathcal{F}y_k|\mathcal{F}x_k), \quad (3.73)$$

όπου  $x_k$  η περιοχή  $k$  της πραγματικής εικόνας. Σημειώνουμε ότι για κάθε διαφορετική περιοχή  $k$ , επιτρέπουμε την μεταβλητότητα των παραμέτρων της εκ-των-προτέρων κατανομής. Αιτία αποτελεί η συσχέτιση των αντίστοιχων κατανομών στον χώρο εικόνας. Επίσης, η από κοινού εκ-των-υστέρων κατανομή της εικόνας στον χώρο Fourier δίνεται από την

$$\pi(\mathcal{F}x|\mathcal{F}y) \propto \prod_{k \in K} \pi(\mathcal{F}x_k)\pi(\mathcal{F}y_k|\mathcal{F}x_k), \quad (3.74)$$

όπου  $K$  είναι το σύνολο όλων των περιοχών στον χώρο Fourier.

### 3.4.2 Εκ-των-προτέρων κατανομή

Η εκ-των-προτέρων κατανομή, σύμφωνα με τους [Kornak et al. \(2020\)](#) και την μέθοδο BIFS, μπορεί να επιτευχθεί με δύο τρόπους. Για λόγους πληρότητας θα περιγράψουμε και τους δύο. Η πρώτη προσέγγιση γίνεται μέσω μιας βάσης δεδομένων από εικόνες (data-driven approach), η οποία καθορίζει τις τιμές των παραμέτρων της εκ-των-προτέρων κατανομής σε κάθε περιοχή του χώρου Fourier. Η δεύτερη, που θα μας απασχολήσει στην ανάλυσή μας, είναι ο καθορισμός των παραμέτρων της εκ-των-προτέρων κατανομής κάθε περιοχής  $k$  του χώρου Fourier από ένα σύνολο παραμετρικών συναρτήσεων (parameter functions). Αρχικά θα αναφέρουμε το στοιχείο που παραμένει ίδιο και στις δύο προσεγγίσεις. Από την Ενότητα § 3.3, ξέρουμε ότι στον χώρο Fourier η εικόνα δίνεται σε μορφή σήματος με μιγαδικούς αριθμούς. Στην πραγματικότητα, θεωρούμε διαφορετικές εκ-των-προτέρων κατανομές για την ακτίνα και την γωνία, που δημιουργεί στον μιγαδικό κύκλο. Χρησιμοποιώντας τις πολικές συντεταγμένες (σε αντίθεση με τις μιγαδικές καρτεσιανές), υιοθετούμε έναν πιο φυσικό τρόπο αναπαράστασης της εκ-των-προτέρων κατανομής. Επιπλέον, επισημαίνουμε ότι η ακτίνα και η γωνία αντιμετωπίζονται ως ανεξάρτητα μεγέθη. Αφού, χρησιμοποιώντας το φανταστικό και το πραγματικό μέρος, η οποιαδήποτε μετατόπιση στην εικόνα (αλλαγή θέσης αντικειμένου) τα επηρεάζει αμφότερα. Εντούτοις, με τις πολικές συντεταγμένες το μόνο που θα επηρεαστεί είναι η γωνία. Φυσικά, όπως αναφέραμε και στην αρχή, οι εκ-των-προτέρων κατανομές θα διαφέρουν για κάθε περιοχή  $k$  του χώρου Fourier.

### 3.4.3 Εκ-των-προτέρων κατανομή καθοδηγούμενη από βάση δεδομένων (Data-Driven approach)

Η πρώτη προσέγγιση για τον καθορισμό των τιμών παραμέτρων της εκ-των-προτέρων κατανομής, βασίζεται στην εμπειρική μας γνώση από μια βάση εικόνων που διαθέτουμε. Η βάση αυτή θα αποτελείται από μετασχηματισμένες εικόνες μέσω του διακριτού μετασχηματισμού Fourier. Ομολογουμένως, η κατανομή που θα επιλέξουμε μπορεί να ποικίλλει. Οι τιμές των εικόνων της βάσης σε μια περιοχή  $k$ , καθορίζουν τις παραμέτρους της εκ-των-προτέρων κατανομής στην εν λόγω περιοχή. Όπως γίνεται αντιληπτό, οι τιμές που εξάγονται από τις μετασχηματισμένες εικόνες είναι μιγαδικές. Συνεπώς, η κάθε περιοχή  $k$  θα απαρτίζεται από ένα ζεύγος εκ-των-προτέρων κατανομών, με την πρώτη να αναφέρεται στην ακτίνα και την δεύτερη στη γωνία του μιγαδικού αριθμού. Για παράδειγμα, μπορούμε να θεωρήσουμε κανονική εκ-των-προτέρων κατανομή για μια περιοχή  $k$ , με μέση τιμή  $\mu_k$  και διακύμανση  $\sigma_k^2$ . Χάρη στην βάση δεδομένων, μπορούμε εμπειρικά να εκτιμήσουμε τις παραμέτρους  $\mu_k, \sigma_k^2$ , από την αντίστοιχη περιοχή  $k$  των μετασχηματισμένων εικόνων της βάσης.

### 3.4.4 Εκ-των-προτέρων κατανομή βασισμένη σε παραμετρικές συναρτήσεις

Η θεώρηση των παραμετρικών συναρτήσεων προκύπτει από την ανάγκη για την καλύτερη διαχείριση των χαρακτηριστικών του χώρου της εικόνας. Εξοπλισμένοι από την ανεξαρτησία μεταξύ των περιοχών του χώρου Fourier, μπορούμε να καθορίσουμε αντίστοιχα ανεξάρτητες εκ-των-προτέρων κατανομές. Πιο συγκεκριμένα, οι παράμετροι της κάθε κατανομής, σύμφωνα με αυτή την προσέγγιση, καθορίζονται από μια παραμετρική συνάρτηση. Αν υποθέσουμε μια παράμετρο  $\alpha_k$  της εκ-των-προτέρων κατανομής της περιοχής  $k$ , τότε μπορούμε να ορίσουμε μια παραμετρική συνάρτηση  $f_\alpha$  τέτοια ώστε  $f_\alpha(k) = \alpha_k$ . Η συνηθέστερη παραδοχή για την παραμετρική συνάρτηση στην περιοχή  $k$  είναι η απόσταση της περιοχής από την αρχή των αξόνων  $(0, 0)$  του χώρου Fourier. Αναλυτικότερα, μπορούμε να υποθέσουμε ότι  $|k| = \sqrt{k_x^2 + k_y^2}$ , όπου  $k_x, k_y$  οι συντεταγμένες των δύο διαστάσεων της εικόνας στο χώρο Fourier. Όμοια, σε περίπτωση που τα δεδομένα είναι τριών διαστάσεων, η απόσταση  $|k|$  θα είναι της μορφής  $|k| = \sqrt{k_x^2 + k_y^2 + k_z^2}$ , όπου  $k_x, k_y, k_z$  οι συντεταγμένες τριών διαστάσεων στον χώρο Fourier. Όπως ήδη γνωρίζουμε, περιοχές κοντά στο κέντρο του χώρου Fourier περιέχουν περισσότερη πληροφορία για την πρωτότυπη εικόνα, μιας και εκεί οι συχνότητες παίρνουν τις μεγαλύτερες τιμές τους. Είναι επομένως λογικό να υποθέτουμε ότι η παραμετρική συνάρτηση  $f_\alpha$  σε μια περιοχή  $k$  εξαρτάται από την απόσταση της συγκεκριμένης περιοχής από το κέντρο του χώρου Fourier. Στη συνέχεια θα εξετάσουμε τις διαφορετικές μορφές των εκ-των-προτέρων κατανομών,

για την ακτίνα και την γωνία φάσεως.

Η επιλογή της εκ-των-προτέρων κατανομής για την ακτίνα μπορεί να καθοριστεί από οποιαδήποτε συνεχή κατανομή. Αναμφίβολα μπορούμε θεωρήσουμε διαφορετικές κατανομές για κάθε περιοχή  $k$ , γεγονός που προσθέτει αρκετή δυσκολία στην διαδικασία της εκτίμησης. Μια πρακτικότερη λύση είναι να διαφοροποιήσουμε μόνο την εκ-των-προτέρων κατανομή του κέντρου των αξόνων του χώρου Fourier, σε σχέση με τις υπόλοιπες των αντίστοιχων περιοχών. Η διάκρισή τους στηρίζεται στην συσσωρευμένη εκ-των-προτέρων πληροφορία που μας δίνει το σημείο  $(0, 0)$  του χώρου Fourier. Επίσης, μια ακόμα σημαντική ιδιότητα για την εκ-των-προτέρων κατανομή της ακτίνας, είναι το πεδίο ορισμού της να αποτελείται από μη αρνητικές τιμές. Σε περίπτωση που πάρουμε αρνητική τιμή, αυτό θα συνεπάγεται μεταβολή της γωνίας στο μιγαδικό επίπεδο. Το διάνυσμα θα περιστραφεί κατά  $\pi$  στον μιγαδικό κύκλο, κι έτσι η γωνία φάσεως δεν θα προσφέρει την επιθυμητή αναπαράσταση των πεποιθήσεών μας.

Οι παράμετροι  $\alpha_k$  της εκ-των-προτέρων κατανομής για την ακτίνα, αναφέρονται στην μέση τιμή  $\mu(k_x, k_y)$  και την τυπική απόκλιση  $\sigma(k_x, k_y)$  που την απαρτίζουν. Η μέση τιμή έχει νόημα να μειώνεται καθώς απομακρυνόμαστε από την αρχή των αξόνων του χώρου Fourier. Αυτό συμβαίνει επειδή μικρότερη ακτίνα στο χώρο Fourier, συνεπάγεται μικρότερη συχνότητα, δηλαδή λιγότερη πληροφορία. Όσο αφορά την παραμετρική συνάρτηση της τυπικής απόκλισης, θεωρούμε πως είναι ανάλογη με αυτή της μέσης τιμής. Έστω μια θετική σταθερά  $c$ , τέτοια ώστε

$$\sigma(k_x, k_y) = c\mu(k_x, k_y). \quad (3.75)$$

Εύκολα συμπεραίνουμε ότι οι εκ-των-προτέρων πεποιθήσεις που έχουμε για την ακτίνα μας είναι αρκετές. Το ίδιο δεν συμβαίνει και με τον άλλο παράγοντα της ανάλυσης, την γωνία φάσεως. Γνωρίζουμε ότι οποιαδήποτε μετατόπιση αντικειμένου στην εικόνα επηρεάζει σημαντικά τη γωνία αυτή. Είναι αδύνατο να γνωρίζουμε εκ-των-προτέρων την ακριβή θέση των αντικειμένων της προς ανάλυση εικόνας. Για τον λόγο αυτό, θα παίρνουμε μη-πληροφοριακή εκ-των-προτέρων κατανομή για αυτή την μεταβλητή. Ειδικότερα, θεωρούμε ομοιόμορφη εκ-των-προτέρων κατανομή που παίρνει τιμές στο κλειστό διάστημα  $[-\pi, \pi]$ , δηλαδή το πεδίο ορισμού της γωνίας.

### 3.4.5 Πιθανοφάνεια

Όπως η εκ-των-προτέρων κατανομή, έτσι και η πιθανοφάνεια θα μοντελοποιείται ξεχωριστά για την ακτίνα και την γωνία φάσεως, σε κάθε περιοχή  $k$  του χώρου Fourier. Επίσης, αναλογιζόμενοι ότι το σετ δεδομένων περιέχει υποβέλτιστες εικόνες, πρέπει να συμπεριλάβουμε στην πιθανοφάνεια  $\pi(\mathcal{F}y_k|\mathcal{F}x_k)$  τον θόρυβο που τις διαμορφώνουν.

Στις περισσότερες εφαρμογές ανάλυσης εικόνας, οι κατανομές των θορύβων θεωρούνται αναξέρτητες και ισόνομες (i.i.d.) κανονικές. Μετασχηματίζοντας την εικόνα στο χώρο Fourier, ο θόρυβος αλλάζει και γίνεται i.i.d. μιγαδικός. Συγκεκριμένα, πρόκειται για τον θόρυβο που υπόκειται η ακτίνα του σήματος. Υπενθυμίζουμε ότι τα πραγματικά και φανταστικά μέρη του σήματος της εικόνας είναι ανεξάρτητα μεταξύ τους. Όμως, από την στιγμή που δουλεύουμε με τις πολικές συντεταγμένες, η κανονικότητα του θορύβου παύει να ισχύει. Έτσι, θα χρειαστεί να θεωρήσουμε μια νέα κατανομή για τις μιγαδικές πολικές συντεταγμένες στον χώρο Fourier, η οποία να ισοδυναμεί με την κανονική στον χώρο της εικόνας.

Κατάλληλη κατανομή θορύβου για την ακτίνα  $z$  στον χώρο Fourier, κρίνεται η Rice [Nicolas and Tupin \(2019\)](#). Αποτελεί την αντίστοιχη μορφή της κανονικής (ανεξάρτητης και ισόνομης) του χώρου εικόνας. Ας θεωρήσουμε  $z_{\text{obs}}$  την παρατηρούμενη από τα δεδομένα ακτίνα και  $z_{\text{true}}$  την πραγματική (χωρίς την παρουσία θορύβου) [Rowe and Logan \(2004\)](#). Τότε η μορφή της κατανομής Rice για την πιθανοφάνεια της πραγματικής ακτίνας στον χώρο Fourier δίνεται από

$$\pi(z_{\text{obs}}|z_{\text{true}}, \sigma) = \frac{z_{\text{obs}}}{\sigma^2} \exp\left(-\frac{z_{\text{obs}}^2 + z_{\text{true}}^2}{2\sigma^2}\right) I_0\left(\frac{z_{\text{obs}}z_{\text{true}}}{\sigma^2}\right), \quad (3.76)$$

όπου  $\sigma$  η τυπική απόκλιση του πραγματικού και του φανταστικού μέρους, για κάθε σημείο του χώρου Fourier. Ο όρος  $I_0\left(\frac{z_{\text{obs}}z_{\text{true}}}{\sigma^2}\right)$  είναι η πρώτου είδους μηδενικής τάξης συνάρτηση Bessel. Η εν λόγω συνάρτηση πρώτου είδους και n-οστής τάξης δίνεται από τον ακόλουθο τύπο

$$I_n(x) = \frac{1}{\pi} \int_0^\pi e^{x \cos(u)} \cos(nu) du. \quad (3.77)$$

Επιπρόσθετα, όσο αφορά την τυπική απόκλιση  $\sigma = \sigma_{\text{signal}}$  μπορεί να εκτιμηθεί από την αντίστοιχη του χώρου εικόνας ( $\hat{\sigma}_{\text{image}}$ ) σύμφωνα με την παρακάτω σχέση



$$\hat{\sigma}_{\text{signal}} = \frac{\hat{\sigma}_{\text{image}}}{4}. \quad (3.78)$$

Η παραπάνω σχέση αποτελεί μια προσέγγιση της  $\sigma$ . Ένας γενικότερος κανόνας εκτίμησης της τυπικής απόκλισης  $\sigma_{\text{signal}}$  στον χώρο Fourier, είναι να απομονώσουμε τις περιοχές θορύβου στον χώρο εικόνας και να τις επεκτείνουμε σε ένα δείγμα ίδιων διαστάσεων με την κανονική εικόνα (ουσιαστικά γεννάμε νέες τιμές με την μέση τιμή και την τυπική απόκλιση του δείγματος που λάβαμε υπόψη). Στη συνέχεια με την εκτέλεση ενός διακριτού μετασχηματισμού Fourier στην διδιάστατη αναπαράσταση του θορύβου, μπορούμε να εκτιμήσουμε την τυπική απόκλιση του διανύσματος της ακτίνας με βάση την κατανομή του θορύβου στον χώρο συχνοτήτων. Στην πραγματικότητα, το να μην γνωρίζουμε την τυπική απόκλιση του θορύβου στον χώρο Fourier δεν αποτελεί σημαντικό πρόβλημα για την ανακατασκευή της εικόνας. Ο τρόπος καθορισμού της παραμετρικής συνάρτησης της εκ-των-προτέρων κατανομής είναι αυτός που παίζει τον πρωτεύοντα ρόλο για την αποθορυβοποίηση. Έτσι, οποιαδήποτε κακή εκτίμηση της τυπικής απόκλισης μπορεί να διορθωθεί με την κατάλληλη παραμετροποίηση της παραμετρικής συνάρτησης.

Η αντίστοιχη πιθανοφάνεια για την γωνία (argument) [Gudbjartsson and Patz \(1995\)](#) δίνεται από

$$\pi(\psi_{\text{obs}}|z_{\text{true}}, \theta, \sigma) = \frac{\exp\left(-\frac{z_{\text{true}}}{2\sigma^2}\right)}{2\pi} \left[ 1 + \rho\sqrt{2\pi} \exp\left(\int_{z_{\text{obs}}=-\infty}^{\rho} \frac{\rho^2}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt\right) \right], \quad (3.79)$$

όπου  $\rho = \frac{z_{\text{true}} \cos(\psi_{\text{obs}} - \psi_{\text{true}})}{\sigma}$  και  $\psi_{\text{obs}}$  είναι η παρατηρούμενη (από τα δεδομένα) γωνία και  $\psi_{\text{true}}$  η πραγματική (χωρίς την παρουσία θορύβου).

### 3.4.6 Εκ-των-υστέρων εκτίμηση

Η εκ-των-υστέρων εκτίμηση σύμφωνα με την μέθοδο BIFS, γίνεται μέσω της μέγιστης εκ-των-υστέρων πιθανότητας (MAP). Για κάθε περιοχή στον χώρο Fourier θα βρούμε την MAP και θα την συμβολίζουμε ως

$$\mathcal{F}_{x_k, \text{MAP}} = \max_{\mathcal{F}_{x_k}} \{\pi(\mathcal{F}_{x_k} | \mathcal{F}_y)\} = \max_{\mathcal{F}_{x_k}} \{\pi(\mathcal{F}_{x_k} | \mathcal{F}_{y_k})\}, \quad (3.80)$$

όπου  $k = 1, 2, \dots, K$ . Παρατηρούμε ότι οι MAP εκτιμήσεις είναι ανεξάρτητες από περιοχή σε

περιοχή (λόγω του ορισμού της ανεξαρτησίας στον χώρο Fourier που έχουμε επισημάνει). Βρίσκουμε την MAP εκτίμηση λόγω της ταχύτερης εύρεσής της. Υπάρχουν δύο παράγοντες που εμπλέκονται στην εκτίμηση της μέγιστης εκ-των-υστέρων πιθανότητας, η ακτίνα (διάνυσμα  $z$ ,  $|z|$ ) και η γωνία του διανύσματος που την δημιουργεί (βλέπε Ενότητα § 3.3). Μπορούμε να θεωρήσουμε ότι οι τιμές των δύο προαναφερθέντων παραγόντων, για μια περιοχή  $k$  του χώρου Fourier σε όλα τα υποβέλτιστα δεδομένα  $y$ , είναι τυχαίες μεταβλητές. Έτσι, η εκτίμηση της εκ-των-υστέρων κατανομής είναι άμεση εξαιτίας του καθορισμού της εκ-των-προτέρων κατανομής και της πιθανοφάνειας. Αξίζει να αναφέρουμε πως μια πολύ σημαντική ιδιότητα που μπορεί να διευκολύνει τους υπολογισμούς είναι η υπόθεση, η εκ-των-προτέρων κατανομή να ανήκει στην ίδια οικογένεια κατανομών με την εκ-των-υστέρων (συζυγής). Φυσικά, όταν οι συνθήκες του προβλήματος το επιτρέπουν. Στη συνέχεια, θέλουμε να δούμε την αντίστοιχη εκτίμηση στον χώρο της εικόνας ώστε να κάνουμε ευδιάκριτο το αποτέλεσμα. Για να επιστρέψουμε στον χώρο εικόνας, αρκεί να εφαρμόσουμε τον αντίστροφο μετασχηματισμό Fourier (βλέπε Ενότητα § 3.3) στην MAP εκτίμηση που πήραμε για κάθε περιοχή  $k$ . Έτσι, η MAP εκτίμηση στον χώρο της εικόνας θα είναι η εξής

$$x_{\text{MAP}} = \mathcal{F}^{-1}(\mathcal{F}x_{\text{MAP}}), \quad (3.81)$$

όπου  $\mathcal{F}x_{\text{MAP}} = \{\mathcal{F}x_{k,\text{MAP}}, k = 1, 2, \dots, K\}$ .

Σύμφωνα με την εκ-των-υστέρων κατανομή που περιγράψαμε, η μέθοδος BIFS μπορεί να χαρακτηριστεί υπολογιστικά οικονομικότερη από τις υπόλοιπες κλασικές που αναλύσαμε. Όπως είδαμε και στην Ενότητα § 3.2, κλασικές μέθοδοι Μπεϋζιανής ανάλυσης εικόνας, όπως τα τυχαία πεδία Markov (MRF) απαιτούν επίπονους επαναληπτικούς αλγορίθμους για την εύρεση της MAP εκτίμησης. Αντίθετα, η πολυπλοκότητα της ανάλυσης με την μέθοδο BIFS μειώνεται σημαντικά. Η ανεξαρτησία που διέπει τον χώρο Fourier, μας επιτρέπει να κατακερματίσουμε το αρχικό πολυδιάστατο πρόβλημα ανάλυσης εικόνας, σε μια σειρά από μονοδιάστατα. Οι εκτιμήσεις των  $K$  περιοχών στον χώρο Fourier γίνονται ξεχωριστά και οι απεικονίσεις τους στον χώρο της εικόνας μια προς μια, δίνουν την τελική εκτίμηση της ανάλυσης.

Φυσικά, θα μπορούσαμε να προσομοιώσουμε δείγμα από την εκ-των-υστέρων κατανομή ώστε να εκτιμήσουμε την μέση τιμή και την τυπική απόκλιση (όπως υπαγορεύουν οι κλασικές μέθοδοι Μπεϋζιανής ανάλυσης εικόνας). Στην περίπτωση αυτή, αναγκαίες κρίνονταν εφαρμογές μεθόδων MCMC που θα μας οδηγούσαν σε ζητήματα σύγκλισης της αλυσίδας, τα οποία θα έπρεπε να

αντιμετωπίσουμε. Παρόλα αυτά, η ανεξαρτησία των εκ-των-υστέρων κατανομών στον χώρο Fourier καθιστά εφικτή την παραγωγή εκ-των-υστέρων δειγμάτων μέσω μιας απλής εφαρμογής Monte Carlo.

### 3.4.7 Αλγόριθμος BIFS

Σε αυτήν την υποενότητα θα δώσουμε την περιγραφή του αλγορίθμου BIFS, με τις παραμέτρους των εκ-των-προτέρων κατανομών να καθορίζονται από παραμετρικές συναρτήσεις. Τα βήματα υλοποίησης του αλγορίθμου είναι τα εξής:

1. **Χώρος εικόνας** → **Χώρος Fourier**: Μετασχηματισμός των δεδομένων, με τον ταχύ μετασχηματισμό Fourier (Fast Fourier Transform).
2. **Εκ-των-προτέρων κατανομή**: Σε κάθε περιοχή του χώρου Fourier, καθορίζουμε την μορφή των εκ-των-προτέρων κατανομών για την ακτίνα και την γωνία του σήματος. Επίσης, ορίζουμε και τις παραμετρικές συναρτήσεις που χαρακτηρίζουν το τρόπο με τον οποίο αυτές μεταβάλλονται στον χώρο Fourier.
3. **Πιθανοφάνεια**: Καθορίζουμε την πιθανοφάνεια στον χώρο Fourier.
4. **Εκ-των-υστέρων κατανομή**: Συνδιασμός των εκ-των-προτέρων κατανομών και της πιθανοφάνειας, σε κάθε περιοχή του χώρου Fourier, για την εύρεση της εκ-των-υστέρων κατανομής (σύμφωνα με το Θεώρημα του Bayes).
5. **Εκτίμηση**: Εκτίμηση της μέγιστης εκ-των-υστέρων πιθανότητας (MAP) σε κάθε περιοχή του χώρου Fourier.
6. **Χώρος Fourier** → **Χώρος εικόνας**: Εφαρμογή αντίστροφου ταχύ μετασχηματισμού Fourier (Inverse FFT), όλων των MAP εκτιμήσεων από το προηγούμενο βήμα, και αναπαράσταση τους στον χώρο εικόνας.

## 3.5 Συμπεράσματα-Σύνοψη κεφαλαίου-Συζήτηση

Στο Κεφάλαιο 3, παρουσιάσαμε ορισμένες βασικές μεθόδους της κλασικής Μπεϋζιανής ανάλυσης εικόνων και εξηγήσαμε την θεωρητική βάση στην οποία στηρίζονται. Πιο συγκεκριμένα,

ασχοληθήκαμε με τα τυχαία πεδία Markov και την κανονικοποίηση Tikhonov. Στη συνέχεια, μεταφερθήκαμε από τον χώρο των πίξελ στον χώρο των συχνοτήτων με τη βοήθεια του διακριτού μετασχηματισμού Fourier. Παρατηρήσαμε ότι ο τελευταίος μας εξοπλίζει με σημαντικές ιδιότητες οι οποίες σε συνδιασμό με την Μπεϋζιανή ανάλυση (BIFS) μας επιτρέπουν να εξετάσουμε από διαφορετική οπτική γωνία το πρόβλημα της αποθρομβοποίησης εικόνων.

Η ποιότητα μιας ιατρικής εικόνας και η ανάδειξη χαρακτηριστικών της που καλύπτονται από θόρυβο, αποτελούν τις κατευθυντήριες γραμμές στις οποίες θα στηριχθεί ένας ιατρός για την λήψη διάγνωσης. Στο Κεφάλαιο 4, θα παρουσιάσουμε και θα περιγράψουμε τα δεδομένα που αργότερα θα χρησιμοποιήσουμε για εφαρμογή των παραπάνω μεθόδων. Στο σημείο αυτό οφείλουμε να αναφέρουμε πως οι μέθοδοι της βαθιάς μάθησης που εξετάσαμε στο Κεφάλαιο 2, δεν θα μπορέσουμε να τις εφαρμόσουμε λόγω έλλειψης μεγάλης βάσης δεδομένων εικόνων. Όπως γίνεται αντιληπτό, τέτοιου είδους εικόνες αποτελούν ευαίσθητα προσωπικά δεδομένα, με αποτέλεσμα η πρόσβαση σε αυτά να μην είναι εύκολη.

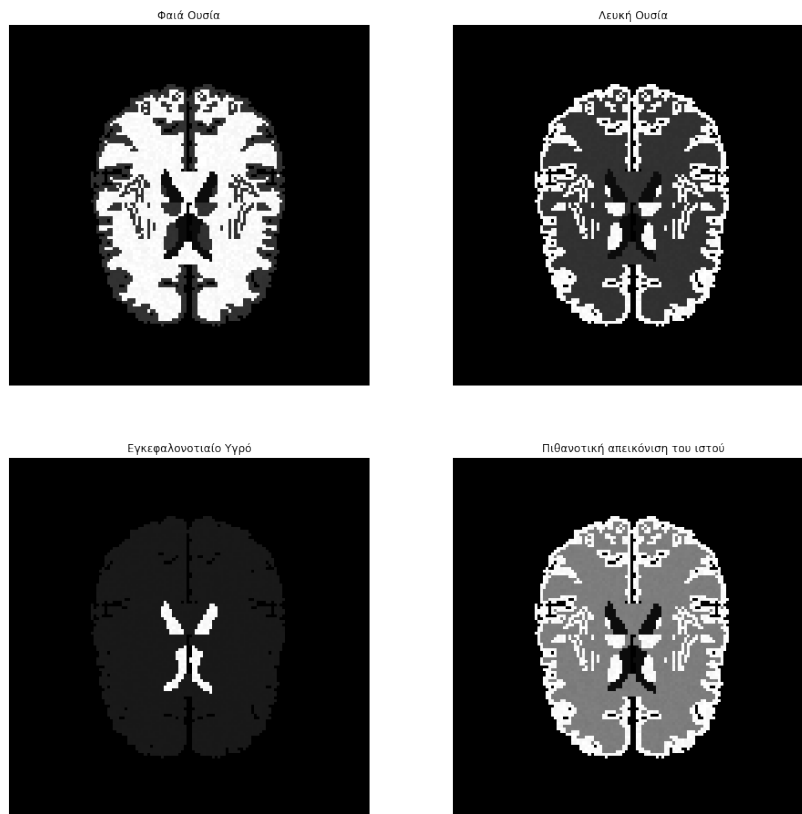
## 4. Περιγραφή δεδομένων

Τα δεδομένα που θα χρησιμοποιήσουμε στην παρούσα εργασία προέρχονται από μια ψευδο-προσομοιωμένη εικόνα (pseudo-simulated image) μαγνητικής τομογραφίας (Magnetic Resonance Imaging ή MRI) διάστασης  $128 \times 128$ , η οποία προέρχεται από πιθανοθεωρητική συνένωση διαφορετικών απεικονίσεων του ιστού του εγκεφάλου (tissue probability maps) ενός ανθρώπου. Σημειώνουμε ότι πρόκειται για ασπρόμαυρες εικόνες, άρα και οι απεικονίσεις θα βρίσκονται στην ίδια κλίμακα χρωμάτων.

Στη συνέχεια θα δώσουμε μια σύντομη αναφορά των μερών του εγκεφάλου που θα μας απασχολήσουν. Ο ανθρώπινος εγκέφαλος εξωτερικά αποτελείται από την φαιά ουσία ή gray matter (σύνολο σωμάτων των νευρικών κυττάρων), την λευκή ουσία ή white matter (συνέχεια της φαιάς ουσίας μέσω νευραξόνων) και το εγκεφαλονοτιαίο υγρό ή cerebrospinal fluid (περιέχεται στο κέντρο και περιμετρικά του εγκεφάλου και καταλήγει την σπονδυλική στήλη).

Η διαθέσιμη πληροφορία του πρώτου σετ δεδομένων προέρχεται, από τις εντάσεις (σε κλίμακα από το 0 έως το 1, όπως περιγράψαμε στην αρχή του Κεφαλαίου 2) των τριών προαναφερθέντων συστατικών του εγκεφάλου από είκοσι επτά MRI εικόνες. Αυτές τελικά συνενώνονται σε μία, παίρνοντας κατά μέσο όρο τις εντάσεις αυτές για κάθε ένα συστατικό. Τα δεδομένα μπορούν να βρεθούν στο <http://nist.mni.mcgill.ca/?p=947>. Στο ακόλουθο σχήμα (βλ. Σχήμα 4.10) αναδεικνύονται με άσπρο (έντονο) χρώμα: πάνω αριστερά η φαιά ουσία, πάνω δεξιά η λευκή ύλη, κάτω αριστερά το εγκεφαλονοτιαίο υγρό και κάτω δεξιά μια μέση απεικόνιση έπειτα από την πιθανοθεωρητική συνένωση των προηγούμενων.

Τα δεδομένα αυτά επεικονίζουν τον εγκέφαλο ενός άνδρα, ο οποίος εισήλθε στον μαγνητικό τομογράφο 27 φορές, ώστε να επιτευχθούν πολύ υψηλής ποιότητας εικόνες, με χαμηλό θόρυβο.



Σχήμα 4.10: Αναπαράσταση της φαιάς ουσίας, της λευκής ουσίας, του εγκεφαλονοτιαίου υγρού και μιας μέσης πιθανοθεωρητική απεικόνιση του ιστού του εγκεφάλου

Υπογραμμίζουμε ότι ο εγκέφαλος του συγκεκριμένου ανθρώπου είναι υγιής. Χειροκίνητα προσθέτουμε έναν υποτιθέμενο όγκο (σε σχήμα κύκλου), αυξάνοντας την ένταση του χρώματος σε μια περιοχή της φαιάς ουσίας. Στη συνέχεια προσθέτουμε κανονικό θόρυβο στην εικόνα, ώστε ο όγκος να πάψει να είναι ευδιάκριτος. Έτσι, έχουμε στην φαρέτρα μας την πραγματική εικόνα (μέγιστης ποιότητας) και την μη βέλτιστη εικόνα (με την προσθήκη του γκαουσιανού θορύβου) την οποία καλούμαστε να ανακατασκευάσουμε.

Επόμενος σταθμός της εργασίας είναι η εφαρμογή των μεθόδων που εξετάσαμε στο Κεφάλαιο 3 σε εικόνες που μόλις περιγράψαμε. Στόχος μας είναι η διερεύνηση και η σύγκριση των υπολογιστικών ικανοτήτων αυτών των μεθόδων σε πραγματικές συνθήκες. Μας ενδιαφέρει τόσο η ποιότητα της ανακατασκευασμένης εικόνας, όσο και ο χρόνος που απαιτείται για την αποθρομβοποίησή της.

## 5. Εφαρμογή και αποτελέσματα

Στο κεφάλαιο αυτό θα εφαρμόσουμε την μέθοδο της Μπεϋζιανής ανάλυσης εικόνας στον χώρο Fourier στο σετ δεδομένων που περιγράψαμε στο Κεφάλαιο 4. Αρχικά, αφού πραγματοποιήσουμε τον ταχύ μετασχηματισμό Fourier στα δεδομένα, θα θεωρήσουμε ότι η πιθανοφάνεια του θορύβου σε κάθε περιοχή του χώρου Fourier είναι κανονική και στη συνέχεια θα εξετάσουμε την περίπτωση που ακολουθεί την κατανομή Rice στον ίδιο χώρο (δηλαδή ακολουθεί την κανονική κατανομή στον χώρο εικόνας). Κατόπιν θα προτείνουμε ένα ζευγάρι εκ-των-προτέρων κατανομών, μια για το διάνυσμα της ακτίνας και μια για την γωνία φάσης. Και στις δύο περιπτώσεις οι εκ-των-προτέρων πεποιθήσεις μας θα στηρίζονται στην ίδια κατανομή για τις δύο τυχαίες μεταβλητές.

Ειδικότερα, το διάνυσμα της ακτίνας, σε κάθε περιοχή  $k$  του χώρου Fourier, θα θεωρήσουμε ότι ακολουθεί εκ-των-προτέρων την κανονική κατανομή με μέση τιμή  $\mu_k$  και διακύμανση  $\sigma_k^2$ , δηλαδή  $\text{Mod}(\mathcal{F}x_k) \sim N(\mu_k, \sigma_k^2)$  (όπου  $\text{Mod}$ , αναφέρεται στο modulus του μιγαδικού αριθμού στην περιοχή  $k$ ). Η γωνία φάσης θα μοντελοποιείται εκ-των-προτέρων από την ομοιόμορφη κατανομή στο διάστημα  $[0, 2\pi]$ , δηλαδή  $\text{Arg}(\mathcal{F}x_k) \sim U(0, 2\pi)$  (όπου  $\text{Arg}$ , αναφέρεται στο argument του μιγαδικού αριθμού στην περιοχή  $k$ ). Η συνάρτηση παραμέτρου της μέσης τιμής, για την εκ-των-προτέρων κατανομή του διανύσματος της ακτίνας, θα θεωρήσουμε ότι είναι της μορφής  $f_\mu(k) = \frac{a}{|k|}$ , όπου  $a$  θετική σταθερά, για κάθε περιοχή  $k$  του χώρου Fourier. Αντίστοιχα, η τυπική απόκλιση μπορεί να βρεθεί από την σχέση αναλογίας (3.75). Το  $a$  και το  $c$  (της Σχέσης 3.75) μπορούν να χαρακτηριστούν ως υπερπαραμέτροι του προβλήματος καθώς η εύρεση των βέλτιστων τιμών τους, μας παρέχει την καλύτερη δυνατή εκ-των-προτέρων αναπαράσταση των χαρακτηριστικών της υποβέλτιστης εικόνας στον χώρο Fourier. Όσο αφορά το κέντρο  $(0,0)$  του χώρου Fourier, θεωρούμε ως εκ-των-υστερών εκτίμηση της ακτίνας την παρατηρούμενη τιμή από τα δεδομένα. Ουσιαστικά είναι σαν να χρησιμοποιούμε μια καταχρηστική εκ-των-προτέρων κατανομή για τον συνολικό μέσο και από αυτή να επιλέγουμε την μέγιστη εκ-των-υστερών

εκτίμηση.

Στο σημείο αυτό, αξίζει να αναφέρουμε τον τρόπο της εκ-των-υστέρων εκτίμησης του argument, που όπως θα δούμε προκύπτει άμεσα. Χάρη στην μη πληροφοριακή κατανομή που θεωρούμε εκ-των-προτέρων για την γωνία, και την πιθανοφάνεια στον χώρο Fourier, συμπεραίνουμε ότι η MAP εκτίμηση της γωνίας σε ένα σημείο  $k$  είναι ακριβώς η παρατηρούμενη τιμή της γωνίας στο σημείο αυτό. Φυσικά, η ομοιόμορφη κατανομή αποτελεί την κατανομή κλειδί για το συμπέρασμα αυτό, μιας και μόνο η πιθανοφάνεια (3.79) καθορίζει το σημείο αυτό, ως εκείνο με την μεγαλύτερη εκ-των-υστέρων πιθανότητα. Η συγκεκριμένη ιδιότητα μας δίνει αφενός την επιθυμητή εκ-των-προτέρων μη πληροφόρηση για την γωνία φάσης και αφεταίρου την άμεση εκ-των-υστέρων εκτίμησης της. Και στις δύο προσεγγίσεις που θα ακολουθήσουμε, η εκ-των-υστέρων εκτίμηση της γωνίας θα παραμείνει η ίδια.

## 5.1 Κανονική κατανομή για την πιθανοφάνεια του θορύβου στον χώρο Fourier

Ξεκινώντας, θα εφαρμόσουμε τον αλγόριθμο BIFS με την υπόθεση ότι η πιθανοφάνεια του θορύβου στον χώρο Fourier είναι κανονική κατανομή με μέση τιμή μηδέν και διακύμανση  $\sigma^2$ , δηλαδή  $\text{Mod}(\mathcal{F}y_k | \mathcal{F}x_k) \sim N(0, \sigma^2)$ . Τότε ωφελούμενοι από την συζυγία των κανονικών, η μέγιστη εκ-των-υστέρων πιθανότητα για το modulus σε κάθε περιοχή  $k$ , όπως αποδείξαμε στην Ενότητα § 3.1 και σύμφωνα με την συμμετρία της κανονικής κατανομής, δίνεται από την σχέση

$$x_{k,\text{MAP}} = \frac{y_k/\sigma^2 + \mu_k/\sigma_k^2}{1/\sigma^2 + 1/\sigma_k^2}. \quad (5.82)$$

## 5.2 Κατανομή Rice για την πιθανοφάνεια του θορύβου στον χώρο Fourier

Στο σημείο αυτό, αξίζει να αναφέρουμε τον τρόπο εκ-των-υστέρων εκτίμησης του modulus, στην περίπτωση που η πιθανοφάνεια ακολουθεί την κατανομή Rice στον χώρο Fourier. Δοθείσης της πιθανοφάνειας της Σχέσης 3.76 και της συνάρτησης παραμέτρου της μέσης τιμής  $\mu_k$ , της κανονικής κατανομής (η συνάρτηση πυκνότητας πιθανότητας δίνεται στην Σχέση 3.27), μπορούμε να υπολογίσουμε την MAP της κατανομής  $\pi(\mathcal{F}x_k | \mathcal{F}y)$ . Έπειτα από ορισμένους υπολογισμούς, η Σχέση 3.73 γίνεται



$$\pi(z_{\text{true}}|z_{\text{obs}}, \sigma) \propto \exp(-\Phi z_{\text{true}}^2 + \Xi z_{\text{true}}) I_0(\Psi z_{\text{true}}), \quad (5.83)$$

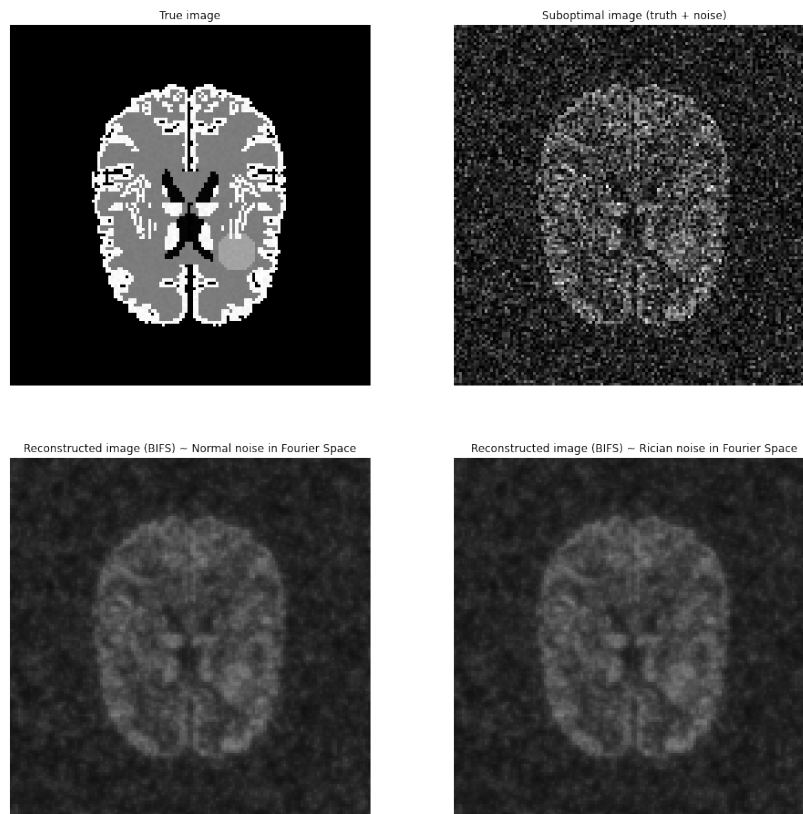
όπου  $\Phi = \frac{\sigma_0^2 + \sigma^2}{2\sigma^2\sigma_0^2}$ ,  $\Xi = \frac{\mu_0}{\sigma_0^2}$  και  $\Psi = \frac{z_{\text{obs}}}{\sigma^2}$ , γνωστές ποσότητες. Παίρνοντας στη συνέχεια την πρώτη παράγωγο της  $\pi(z_{\text{true}}|z_{\text{obs}}, \sigma)$  και λογαριθμίζοντάς την (ο λογάριθμος είναι αύξουσα συνάρτηση η οποία δεν αλλοιώνει τα τοπικά βέλτιστα της αρχικής συνάρτησης), την εξισώνουμε με το μηδέν. Έτσι, μπορούμε να βρούμε το ολικό μέγιστο (την επικρατούσα τιμή ή mode), δηλαδή το  $z_{\text{true}}$  για το οποίο η συνάρτηση πυκνότητας παίρνει τη μέγιστη τιμή της (MAP εκτίμηση). Η εύρεση της τιμής αυτής δεν μπορεί να γίνει άμεσα, αλλά με προσεγγιστικές αριθμητικές μεθόδους. Πιο συγκεκριμένα, η μέθοδος που χρησιμοποιούμε είναι η μέθοδος του Simpson [Süli and Mayers \(2003\)](#). Βασική αιτία εφαρμογής τέτοιων τεχνικών αποτελεί η μη γραμμικότητα της εν λόγω εξίσωσης. Η σχέση που προκύπτει έπειτα από την παραγωγή του λογαρίθμου της εκ-των-υστερών κατανομής είναι η εξής

$$\frac{\partial}{\partial z_{\text{true}}} \log(\pi(z_{\text{true}}|z_{\text{obs}}, \sigma)) \propto -2z_{\text{true}}\Phi + \Xi + \Psi \frac{I_1(\Psi z_{\text{true}})}{I_0(\Psi z_{\text{true}})}. \quad (5.84)$$

Εύκολα επιβεβαιώνουμε ότι πρόκειται για το μέγιστο, βρίσκοντας ότι η δεύτερη παράγωγος είναι αρνητική σε κάθε σημείο του πεδίου ορισμού του  $z_{\text{true}}$ .

### 5.3 Robustness

Εφαρμόζουμε δύο φορές τον αλγόριθμο BIFS, μια θεωρώντας την κανονική κατανομή ως πιθανοφάνεια του θορύβου στον χώρο Fourier, και μια θεωρώντας την κατανομή Rice στον ίδιο χώρο. Η μοντελοποίηση του αλγορίθμου γίνεται σύμφωνα με αυτή που δώσαμε στην Ενότητα § 5.1, για κάθε διαφορετική περίπτωση της πιθανοφάνειας. Η συνάρτηση παραμέτρου που χρησιμοποιήσαμε για την μέση τιμή της εκ-των-προτέρων κατανομής, είναι η  $f_{\mu}(|k|) = \frac{a}{|k|}$  και για τις δύο υλοποιήσεις. Ουσιαστικά, θέλουμε αν εξετάσουμε τον τρόπο με τον οποίο η κατανομή της πιθανοφάνειας επηρεάζει τα αποτελέσματα που παίρνουμε. Στην Εικόνα 5.11, δίνεται μια ολοκληρωμένη αναπαράσταση της εφαρμογής του αλγορίθμου για τις δύο διαφορετικές θεωρήσεις για την πιθανοφάνεια του θορύβου. Να σημειώσουμε ότι η πραγματική εικόνα που δίνεται πάνω αριστερά στο σχήμα δίνεται αποκλειστικά για λόγους σύγκρισης με το αποτέλεσμα του αλγορίθμου. Στην πραγματικότητα δε θα είναι διαθέσιμη.



Σχήμα 5.11: Εφαρμογή του αλγορίθμου BIFS και σύγκριση μεταξύ της κανονικής κατανομής και της κατανομής Rice ως πιθανοφάνεια του θορύβου στον χώρο Fourier

Όπως γίνεται αντιληπτό, τα αποτελέσματα που παίρνουμε και στις δύο περιπτώσεις είναι σχεδόν ίδια, παρόλη την διαφορετική θεώρηση της πιθανοφάνειας στον χώρο Fourier. Η μοντελοποίηση της πιθανοφάνειας με την κανονική κατανομή προσεγγίζει αρκετά καλά αυτήν της κατανομής Rice. Συνεπώς, προκύπτει το συμπέρασμα ότι κυρίαρχο ρόλο για την ανακατασκευή της εικόνας παίζουν οι συναρτήσεις παραμέτρων της εκ-των-προτέρων κατανομής. Με κάθε αλλαγή στη συνάρτηση παραμέτρου δίνεται έμφαση σε διαφορετικά χαρακτηριστικά της εικόνας κατά την ανοικοδόμηση. Οι χρόνοι που χρειάστηκαν για την υλοποίηση της μεθόδου στις δύο περιπτώσεις δίνονται στον Πίνακα [5.1](#)

Παρατηρούμε ότι με την κατανομή Rice να αντιπροσωπεύει την πιθανοφάνεια του θορύβου στον χώρο Fourier απαιτείται σημαντικά περισσότερος χρόνος υπολογισμού για την εκ-των-υστέρων εκτίμηση, από ότι με την κανονική κατανομή. Στην πρώτη περίπτωση χρειάστηκαν λίγο

	Κατανομή Rice	Κανονική κατανομή
Χρόνος υπολογισμού του BIFS σε δευτερόλεπτα (sec)	247.47	0.189

Πίνακας 5.1: Χρόνοι υλοποίησης του αλγορίθμου BIFS για τις διαφορετικές κατανομές πιθανοφανειών του θορύβου στον χώρο Fourier

περισσότερο από τέσσερα λεπτά για την υλοποίηση του αλγορίθμου, ενώ στην δεύτερη μόλις 0.189 δευτερόλεπτα. Καθοριστικός παράγοντας για την καθυστέρηση της πρώτης περίπτωσης, είναι οι μη γραμμικές εξισώσεις (προκύπτουν από την Σχέση 5.84) τις οποίες καλείται να λύσει ο αλγόριθμος για κάθε περιοχή  $k$  του χώρου Fourier. Εξαιτίας της παρόμοιας συμπερασματολογίας (και οπτικού αποτελέσματος), αλλά και της μεγάλης διαφοράς υπολογιστικού χρόνου των δύο περιπτώσεων, μπορούμε δίχως ενδοιασμούς να χρησιμοποιούμε την κανονική κατανομή ως πιθανοφάνεια του θορύβου στον χώρο Fourier, χωρίς να επηρεάζεται σημαντικά το αποτέλεσμα της εκτίμησης.

Επιπροσθέτως, και στις δύο περιπτώσεις παρατηρούμε ότι η ανάλυση της εικόνας με θόρυβο βελτιώθηκε σημαντικά και ορισμένα χαρακτηριστικά έγιναν περισσότερο ευδιάκριτα. Ο τεχνητός όγκος που προσθέσαμε με την παρουσία θορύβου δεν γίνεται εύκολα αντιληπτός, όμως με την εφαρμογή του αλγορίθμου καταφέρνουμε να τον αναδείξουμε (βλέπε Εικόνα 5.11).

## 5.4 Συμπεριφορά παραμετρικών συναρτήσεων

Στην παρούσα ενότητα θα δώσουμε έναν ενδεικτικό οδηγό για την επιλογή των παραμετρικών συναρτήσεων. Έχουμε ήδη αναλύσει, από την Ενότητα § 3.4.4, τον τρόπο λειτουργίας των παραμετρικών συναρτήσεων σε σχέση με το κέντρο των αξόνων του χώρου Fourier. Τώρα, θα επικεντρωθούμε στο πως οι σταθερές αναλογίας  $c$ , για την τυπική απόκλιση, και  $\alpha$  για την μέση τιμή, επηρεάζουν το αποτέλεσμα του αλγορίθμου BIFS.

Στην αρχή του κεφαλαίου θεωρήσαμε ως  $f_{\mu}(k) = \frac{\alpha}{\sqrt{k}}$  την συνάρτηση παραμέτρου της μέσης τιμής σε κάθε περιοχή  $k$  του δισδιάστατου χώρου Fourier. Επιβεβαιώνουμε ότι με την εξέταση συχνοτήτων που βρίσκονται πολύ κοντά στο κέντρο των αξόνων του μιγαδικού επιπέδου, η μέση τιμή που μας δίνει η συνάρτηση παραμέτρου είναι πολύ κοντά στο μηδέν. Όσο μικρότερη θεωρήσουμε την σταθερά  $\alpha$  τόσο πιο κοντά στο μηδέν πλησιάζουμε. Φυσικά δεν επιθυμούμε συχνότητες που μας εφοδιάζουν με χρήσιμη πληροφορία, να μηδενιστούν. Έτσι, ο κύριος λόγος

που χρησιμοποιούμε μικρές τιμές της σταθεράς  $\alpha$  είναι για τις μικρές συχνότητες, αυτές που βρίσκονται μακριά από το κέντρο των αξόνων  $(0,0)$ . Στόχος μας είναι οι συχνότητες αυτές, δεδομένου ότι δεν περιέχουν χρήσιμη πληροφορία για την εικόνα, να γίνονται μικρότερες πλησιάζοντας το μηδέν. Έχει νοήμα λοιπόν να επικεντρωθούμε στις τιμές της  $\alpha$  στο διάστημα  $(0,1)$ . Παρατηρούμε ότι η σταθερά  $\alpha$  με τιμές στο προαναφερθέν διάστημα, επιτυγχάνει μεγαλύτερη μείωση των μη πληροφοριακών συχνοτήτων και μικρότερη των πληροφοριακών. Έτσι, η επιλογή της  $\alpha$  καθορίζει το τμήμα της σημαντικής πληροφορίας που θα πληρώσουμε, σε σύγκριση με την ανεπιθύμητη πληροφορία που θέλουμε να απαλλαχθούμε. Για αυτό το λόγο πρέπει να την επιλέγουμε με προσοχή, και μετά από δοκιμές για διάφορες τιμές της, στον αλγόριθμο BIFS.

Στη συνέχεια θα εξετάσουμε τον τρόπο επιλογής της σταθεράς  $c$  για την συνάρτηση της τυπικής απόκλισης. Όπως έχουμε αναφέρει η τελευταία θα δίνεται από την  $\sigma(k_x, k_y) = c\mu(k_x, k_y)$  (βλέπε Σχέση 3.75). Παρατηρούμε πως η τυπική απόκλιση δίνεται εξ ορισμού συναρτήσει της μέσης τιμής.

Έχοντας λοιπόν επιλέξει την μέση τιμή, μέσω της σταθεράς  $\alpha$ , που εξηγήσαμε στην προηγούμενη παράγραφο, μπορούμε τώρα να συνεχίσουμε στην αναζήτηση της σταθεράς  $c$ . Γνωρίζουμε ότι η μεγάλη διακύμανση προσδίδει αρκετή αβεβαιότητα στην περίπτωση που γεννάμε τιμές από μια κατανομή. Γίνεται αντιληπτό πως κάτι τέτοιο δεν πρέπει να συμβαίνει σε περιοχές συχνοτήτων της εικόνας μακριά από την αρχή των αξόνων του μιγαδικού επιπέδου. Σκοπός είναι να απαλλαχθούμε όσο το δυνατόν από τις μη πληροφοριακές συχνότητες. Για παράδειγμα, εάν έχουμε μεγάλη διακύμανση τότε υπάρχει πιθανότητα να δώσουμε ισχύ σε συχνότητα οι οποία δεν προσφέρει τίποτα, παρά μόνο θόρυβο, στην ανακτασκευασμένη εικόνα. Επίσης, αν αναλογιστούμε ότι αυτές συχνότητες είναι περισσότερες σε πλήθος από τις πληροφοριακές, τότε η μεγάλη διακύμανση ενδέχεται να αυξήσει σημαντικά τον συνολικό θόρυβο της εικόνας. Άρα σε τέτοιες περιοχές προτιμούμε να επιλέγουμε μεγάλες τιμές της σταθεράς  $c$ , ώστε να κρατάμε χαμηλές τις συχνότητες.

Όπως είδαμε, η μέση τιμή είναι ένας θετικός αριθμός μικρότερος της μονάδας και συνεπώς ο πολλαπλασιασμός της με έναν μεγάλο θετικό αριθμό μας δίνει μικρή διακύμανση. Σε περιπτώσεις όπου η συχνότητα που εξετάζουμε βρίσκεται κοντά στο κέντρο των αξόνων, ιδανικά επιθυμούμε την γέννηση τιμών μεγαλύτερων της μέσης τιμής της κατανομής. Με τον τρόπο αυτό θα

επανακτούμε την πληροφορία που χάνουμε κατά την επιλογή της μέσης τιμής. Έχει νόημα λοιπόν να επιλέγουμε υψηλή τιμή της σταθεράς  $c$  για όλες τις περιοχές. Εάν η διακύμανση είναι αρκετά μικρή, δηλαδή η σταθερά  $c$  να παίρνει μεγάλες τιμές, τότε στην ανακατασκευασμένη εικόνα θα εκκλείπει αφενώς ο θόρυβος, αφαιτέρου όμως δεν θα είναι εφικτή η ολική ανάδειξη των επιθυμητών χαρακτηριστικών της (λόγω αυξημένης ομαλοποίησης της εικόνας).

Ακολούθως, στο Σχήμα 5.12 παραθέτουμε ενδεικτικά τα αποτελέσματα της εφαρμογής του αλγορίθμου BIFS για διάφορες θεωρήσεις της συναρτήσεων παραμέτρων, αλλά και των τιμών  $c, \alpha$  που τις αποτελούν.

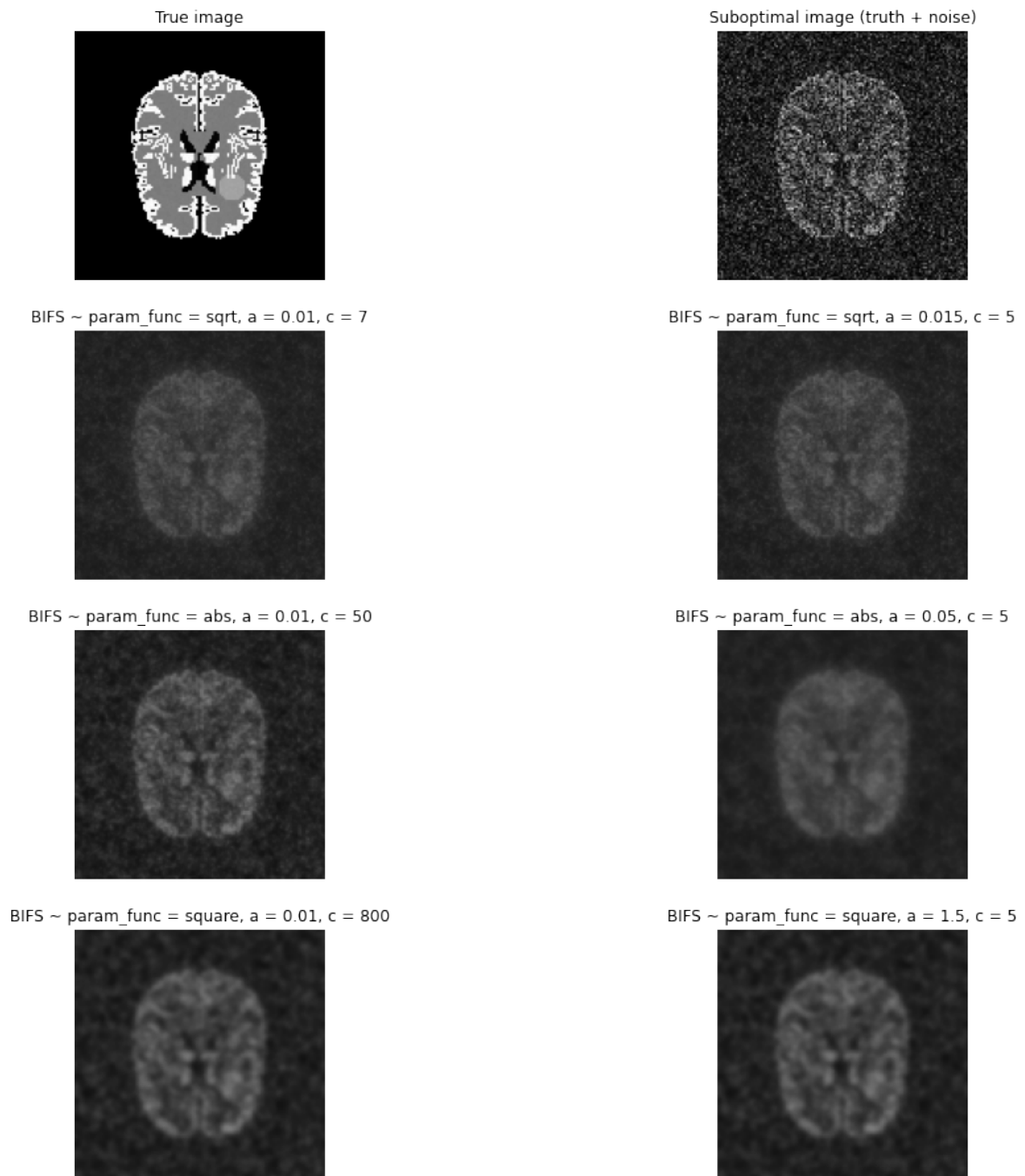
## 5.5 Σύγκριση του αλγορίθμου BIFS με την κανονικοποίηση Tikhonov

Στη συνέχεια θα συγκρίνουμε τα αποτελέσματα που παράγονται από την μέθοδο BIFS και την κανονικοποίηση Tikhonov. Μας ενδιαφέρει να εξετάσουμε πως συμπεριφέρεται ο αλγόριθμος BIFS σε αντίθεση με μια κλασσική μέθοδο ανακατασκευής εικόνας. Προφανώς, για την πρώτη μέθοδο θα θεωρήσουμε ως πιθανοφάνεια την κανονική κατανομή, που όπως είδαμε στην Ενότητα § 5.1, τα αποτελέσματα που παράγει προσεγγίζουν αρκετά αυτά που μας δίνει η κατανομή Rice, αλλά σε σημαντικά μικρότερο υπολογιστικό χρόνο.

Η μοντελοποίηση που θεωρούμε για το πρόβλημα ελαχιστοποίησης της μεθόδου Tikhonov (βλέπε (3.45)) είναι η ακόλουθη. Υποθέτουμε ότι ο πίνακας  $H$  είναι ο ταυτοτικός, έτσι ώστε ο πρώτος όρος της Σχέσης 3.45 να είναι  $\|y - x\|_2^2$ . Δηλαδή παίρνουμε ακριβώς τα πίκσελ της ανακατασκευασμένης εικόνας  $x$ . Επίσης, ορίζουμε τον όρο κανονικοποίησης  $\|\Gamma x\|_2^2$  ώστε να αποτελεί ένα διάνυσμα  $u$  διάστασης  $n \times 1$  και στοιχεία  $u_i$ , όπου

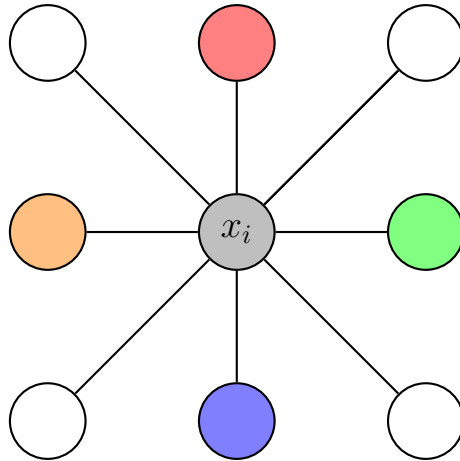
$$u_i = \left( \text{card}(\mathcal{J}_i)x_i - \sum_{x \in \mathcal{J}_i} x \right)^2 = \left( \sum_{x \in \mathcal{J}_i} (x_i - x) \right)^2, \quad (5.85)$$

όπου  $i = 1, 2, \dots, n$ .  $\mathcal{J}_i$  είναι το σύνολο που περιέχει τα γειτονικά πίκσελ του  $i$ -οστού πίκσελ, και  $\text{card}(\mathcal{J}_i)$  είναι το πλήθος των στοιχείων του συνόλου αυτού. Έτσι, τα πίκσελ που βρίσκονται στην περίμετρο της εικόνας θα έχουν δύο (τα τέσσερα γωνιακά πίκσελ) ή τρεις γείτονες. Όλα τα υπόλοιπα πίκσελ θα έχουν τέσσερις γείτονες. Να σημειώσουμε εδώ ότι η έννοια της γειτονικότητας είναι υποκειμενικός, με την έννοια του ότι ο πληθάριθμος του  $\mathcal{J}_i$  θα μπορούσε να έχει θεωρηθεί



Σχήμα 5.12: Εφαρμογή του αλγορίθμου BIFS με συνάρτησεις παραμέτρων  $f_{\mu}(|k|) = \frac{a}{\sqrt{|k|}}, \frac{a}{|k|}, \frac{a}{|k|^2}$  για διάφορες τιμές των τιμών  $c, a$





Σχήμα 5.13: Διάγραμμα των γειτονικών πίξελ του  $x_i$ , με τα χρώματα των τεσσάρων γειτόνων του να αντιστοιχούν στις διαγωνίους του πίνακα  $\Gamma$

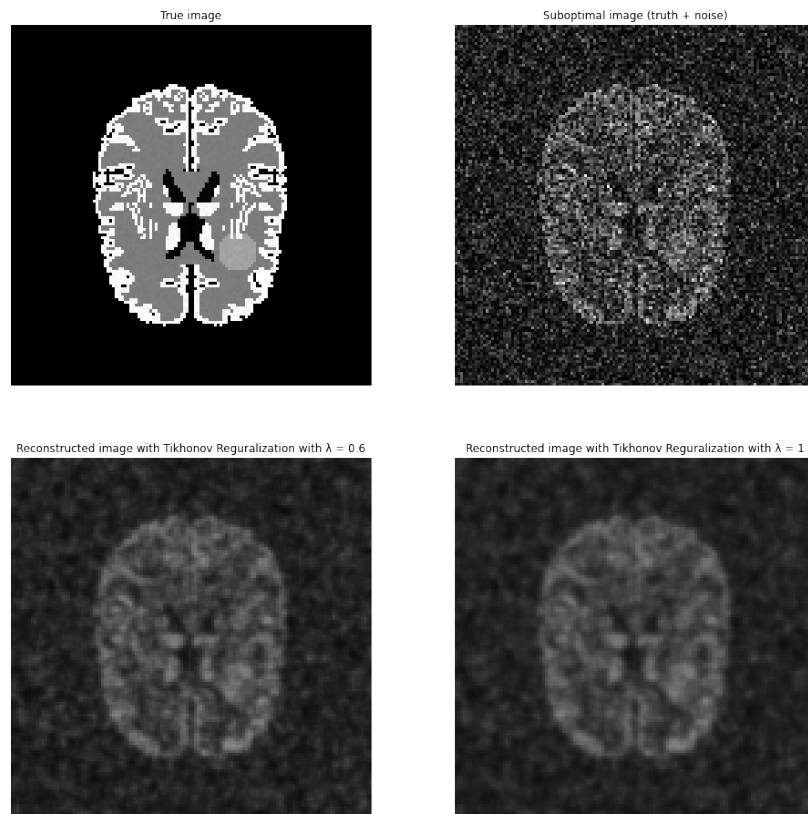
Τώρα, με τη βοήθεια του αλγορίθμου της συζυγούς κλίσης (Αλγόριθμος 2) επιλύουμε το γραμμικό σύστημα εξισώσεων, εκτιμώντας προσεγγιστικά το διάνυσμα  $x$ . Στο Σχήμα 5.14 παρουσιάζουμε την ανακατασκευή της υποβέλτιστης εικόνας του πιθανοθεωρητικού ιστού του εγκεφάλου, που επιτυγχάνει η μέθοδος της κανονικοποίησης Tikhonov. Το σύνολο των τιμών της υπερπαραμέτρου  $\lambda$  που θεωρήσαμε για την εφαρμογή της μεθόδου είναι το  $\{1, 0.6\}$ .

Βλέπουμε ότι η μέθοδος κανονικοποίησης Tikhonov με τον πίνακα  $\Gamma$  που θεωρήσαμε εξαλείφει τον θόρυβο και αναδεικνύει χαρακτηριστικά του εγκεφάλου που πριν δεν ήταν ορατά. Καθώς η τιμή της υπερπαραμέτρου  $\lambda$  αυξάνεται, φαίνεται ότι η ανακατασκευασμένη εικόνα τείνει να ομαλοποιείται. Όμως για αρκετά μεγάλες τιμές θα δημιουργηθεί έντονη θολούρα που ενδεχομένως θα καλύπτει χρήσιμα χαρακτηριστικά του εγκεφάλου κατά την συμπερασματολογία ενός ειδικού.

Εν συνεχεία, θέλουμε να εξετάσουμε την συμπεριφορά της κανονικοποίησης Tikhonov με τον αλγόριθμο BIFS. Για την τελευταία μέθοδο θα θεωρήσουμε την κανονική πιθανοφάνεια για την κατανομή του θορύβου (στον χώρο εικόνας), αφού όπως ήδη έχουμε επισημάνει (Ενότητα § 5.3) η συμπεριφορά της προσεγγίζει αρκετά αυτήν της πιθανοφάνειας Rice στον χώρο συχνοτήτων. Για την κανονικοποίηση Tikhonov θα θέσουμε την υπερπαραμέτρο  $\lambda$  ίση με 1. Τα αποτελέσματα των δύο μεθόδων δίνονται στο Σχήμα 5.15.

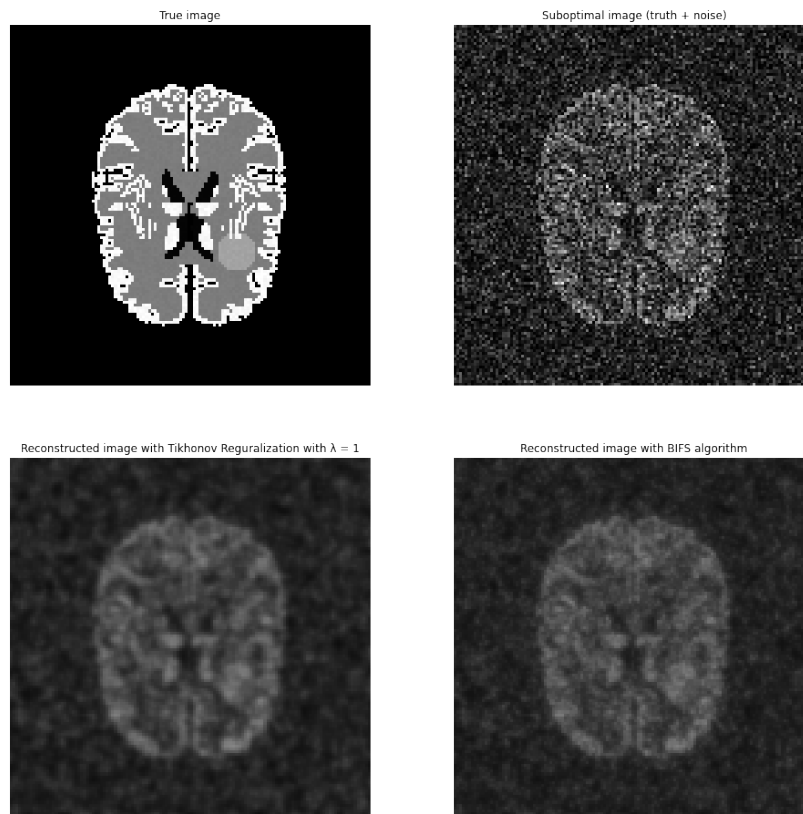
Γίνεται αντιληπτό, ότι και οι δύο μέθοδοι εξομαλύνουν τον θόρυβο και ανακατασκευάζουν αποτελεσματικά την υποβέλτιστη εικόνα. Βέβαια, η μέθοδος BIFS φαίνεται να μας εφοδιάζει με





Σχήμα 5.14: Εφαρμογή της μεθόδου κανονικοποίησης Tikhonov

περισσότερη πληροφορία, με αρκετές λεπτομέρειες της περιοχής του εγκεφάλου να είναι περισσότερο ευδιάκριτες. Αδιαμφισβήτητα, και οι δύο αλγόριθμοι ενισχύουν επιδέξια την ένταση του όγκου, ώστε ο εντοπισμός του να γίνεται εύκολος. Ένα ακόμα σημαντικό μέρος της σύγκρισης των μεθόδων είναι και ο χρόνος υπολογισμού της εκτίμησης εικόνας. Η κανονικοποίηση Tikhonov χρειάστηκε 96.37 δευτερόλεπτα για την ανακατασκευή. Ο χρόνος αυτός είναι σημαντικά καλύτερος από αυτόν της μεθόδου BIFS, όπου η κατανομή της πιθανοφάνειας του θορύβου είναι η Rice (όπως είδαμε στον Πίνακα 5.1). Το ίδιο όμως δεν συμβαίνει στην περίπτωση που θεωρήσουμε την κανονική πιθανοφάνεια του θορύβου στον χώρο Fourier. Η εκτίμηση της εικόνας με την παραπάνω μοντελοποίηση επιτυγχάνεται σε λιγότερο από δύο δέκατα του δευτερολέπτου. Η χρονική πολυπλοκότητα, καθώς και το λεπτομερές αποτέλεσμα της μεθόδου BIFS μας επιβεβαιώνουν ότι κυριαρχεί έναντι της κανονικοποίησης Tikhonov.



Σχήμα 5.15: Σύγκριση του αλγορίθμου BIFS με την μέθοδο της κανονικοποίησης Tikhonov

Τέλος, ένα επίσης σημαντικό πλεονέκτημα της προσέγγισης BIFS είναι η συνάρτηση παραμέτρων που ορίζουμε. Όπως είδαμε στην κανονικοποίηση Tikhonov καθοριστικό ρόλο για την μοντελοποίηση παίζουν οι πίνακες  $H, \Gamma$ , καθώς και η υπερπαραμέτρος  $\lambda$ . Ανάλογα με τα χαρακτηριστικά της εικόνας που θέλουμε να αναλύσουμε, οι δύο προηγούμενοι πίνακες και η παράμετρος κανονικοποίησης οφείλουν να αναπροσαρμόζονται με βάση τις εκάστοτε συνθήκες. Αντίθετα, το αποτέλεσμα της μεθόδου BIFS εξαρτάται μονάχα από ένα παράγοντα, την συνάρτηση παραμέτρων. Το γεγονός αυτό δίνει επιπλέον ευκολία στον χρήστη να προσαρμόζει την ανακατασκευή της εικόνας, απαλλάσσοντας τον από την εύρεση κατάλληλης τιμής για το  $\lambda$  και εκ νέου μοντελοποίησης του προβλήματος.

## 5.6 Συμπεράσματα / Επίλογος Κεφαλαίου

Στο κεφάλαιο αυτό εφαρμόσαμε τον αλγόριθμο BIFS σε ιατρική εικόνα εγκεφάλου που είχε υποστεί θόρυβο. Παρουσιάσαμε την παρόμοια συμπεριφορά που επιδέχεται οπτικά η εκ των υστέρων εκτίμηση της εικόνας, με τη χρήση της κανονικής και της Rice κατανομής ως τις πιθανοφάνειες της στο χώρο συχνοτήτων. Το προνόμιο του τελευταίου αποτελέσματος είναι η εξοικονόμηση σημαντικού υπολογιστικού χρόνου με την κανονική κατανομή ως την πιθανοφάνεια του θορύβου. Επιπλέον, δώσαμε έναν ενδεικτικό οδηγό χρήσης της παραμετρικής συνάρτησης, καθώς και της επίδρασης των σταθερών  $c$  και  $\alpha$  στην αναδόμηση της εικόνας. Στη συνέχεια, εφαρμόσαμε την μέθοδο της κανονικοποίησης Tikhonov στην ίδια θορυβώδη εικόνα. Περιγράψαμε τον τρόπο με τον οποίο επιλέξαμε τον πίνακα  $\Gamma$  του προβλήματος ελαχιστοποίησης που η τελευταία μέθοδος ισοδυναμεί και προσπαθήσαμε να τη συγκρίνουμε με τα αποτελέσματα του αλγορίθμου BIFS. Το συμπέρασμα στο οποίο καταλήξαμε είναι ότι, η μέθοδος BIFS υπερτερεί της μεθόδου κανονικοποίησης Tikhonov εξαιτίας της καλύτερης συνολικά ανάδειξης των χαρακτηριστικών της εικόνας, αλλά και συνάμα της αμεσότητας της μοντελοποίησης.

Όλα τα πειράματα που παρουσιάστηκαν στην παρούσα εργασία έγιναν σε υπολογιστή με λογισμικό Windows, 6 3.40GHz CPU πυρήνες και 16GB RAM μνήμη. Η Python είναι η γλώσσα προγραμματισμού που χρησιμοποιήσαμε για την υλοποίηση των μεθόδων.



## 6. Συμπεράσματα και μελλοντική έρευνα

Από τα αποτελέσματα που πήραμε δοκιμάζοντας και συγκρίνοντας την μέθοδο της Μπεϋζιανής ανάλυσης Εικόνων στον Χώρο Fourier (BIFS), καταλήγουμε στο συμπέρασμα ότι πρόκειται για μια αποτελεσματική και υπολογιστικά γρήγορη μέθοδο. Η ευκολία και η αμεσότητα της μοντελοποίησης αποτελεί ένα ακόμα πλεονέκτημα έναντι άλλων τεχνικών Μπεϋζιανής ανάλυσης εικόνας, όπως τα τυχαία πεδία Markov και την κανονικοποίηση Tikhonov. Ένας ιατρός καλείται να λάβει σημαντικές αποφάσεις από την ανακατασκευή μιας εικόνας που είναι εκτεθειμένη στον θόρυβο. Η επιλογή της συνάρτησης παραμέτρων, και ειδικότερα των σταθερών  $c$  (θετική σταθερά αναλογίας για την τυπική απόκλιση της συνάρτησης παραμέτρων) και  $a$  (σταθερά για την μέση τιμή της συνάρτησης παραμέτρων), είναι αυτή που καθορίζει την αναδόμιση και την ανάδειξη χαρακτηριστικών μιας εικόνας. Το κλειδί για την εύρεση της κατάλληλης συνάρτησης παραμέτρων βρίσκεται στο μοτίβο των εντάσεων της εικόνας σχετικά με τον θόρυβο. Εφαρμόζοντας τον αλγόριθμο BIFS για διάφορες συναρτήσεις παραμέτρων και τιμές των σταθερών  $c, a$ , ο ιατρός έχει την δυνατότητα να βαθμολογεί τις συναρτήσεις παραμέτρου που χρησιμοποιεί για τις προς εξέταση εικόνες, ανάλογα με το πόσο καλά αυτές αναδεικνύουν τα επιθυμητά χαρακτηριστικά. Μια ενδεικτική καθοδηγητική γραμμή για τον τρόπο επιλογής των συναρτήσεων παραμέτρου παρατέθηκε στην Ενότητα § 5.4.

Η μέθοδος BIFS υπερτερεί από άποψη μοντελοποίησης, ακόμα και από γνωστούς αλγορίθμους της βαθιάς μάθησης, όπως είναι οι Autoencoders και οι Variational Autoencoders. Οι δύο τελευταίοι ουσιαστικά αποτελούν μια επέκταση των νευρωνικών δικτύων και λειτουργούν αποτελεσματικά με μεγάλο όγκο δεδομένων προς εκπαίδευση. Και σε αυτό το κομμάτι η Μπεϋζιανή ανάλυση εικόνας και ειδικότερα η μέθοδος BIFS αναδεικνύονται κυρίαρχες στην τομέα της ιατρικής, εξαιτίας των περιορισμένης διαθεσιμότητας τέτοιου τύπου δεδομένων. Επιπλέον, η αρχιτεκτονική κατασκευή τους είναι περίπλοκη και ο αριθμός των υπερπαραμέτρων που στηρίζεται είναι αρκετά μεγάλος. Συνεπώς, αναγκάιες κρίνονται πολλές προσομοιώσεις για τον

καθορισμό τους ξεχωριστά για κάθε πρόβλημα. Όπως γίνεται αντιληπτό, ο αλγόριθμος BIFS πρόκειται για ένα πολύ ισχυρό εργαλείο το οποίο έρχεται να προστεθεί στις έως τώρα γνωστές μεθόδους ανάλυσης εικόνας. Η αποτελεσματική ανακατασκευή μιας εικόνας (ανάδειξη χρήσιμων χαρακτηριστικών) σε συνδιασμό με την μικρή υπολογιστική πολυπλοκότητα είναι μερικά από τα οφέλη που καθιστούν την συγκεκριμένη μέθοδο ένα σθεναρό συναγωνιστή των υπόλοιπων κλασικών.

Επιπρόσθετα, η Μπεϋζιανή μέθοδος ανάλυσης εικόνων BIFS, μπορεί να επεκταθεί εύκολα και σε τρεις διαστάσεις, χρησιμοποιώντας την θεωρία της Ενότητας 3.4 και τον αντίστοιχο μετασχηματισμό Fourier. Μια εφαρμογή που πιστεύουμε θα ενίσχυε την επιστημονική κοινότητα είναι αυτή της συνδεσιμότητας του εγκεφάλου μέσω εικόνων αιμάτωσης ASL (perfusion Arterial Spin Labeling). Βασική τους ιδιότητα είναι να εντοπίζουν και να αναδεικνύουν το αίμα που βρίσκεται στον εγκέφαλο, χρησιμοποιώντας ως ενδογενή ιχνηλάτη προτόνια νερού. Πρόκειται για συνεχόμενες εικόνες μαγνητικής τομογραφίας, μικρής ευκρίνειας και αυξημένου θορύβου (λόγου του μικρού διαστήματος που έχει διαθέσιμος ο τομογράφος για την λήψη της εικόνας). Με την ανακατασκευή και την ανάδειξη χαρακτηριστικών τέτοιων ακολουθιακών εικόνων θα μπορούσαμε να αποφανθούμε με συνέπεια για τον τρόπο με τον οποίο επικοινωνούν τα διάφορα μέρη του εγκεφάλου, μέσω της ροής του αίματος.

# Αναφορές

- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.
- Besag, J. (1989). Digital image processing: Towards bayesian image analysis. *Journal of Applied statistics*, 16(3):395–407.
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301.
- Cross, G. R. and Jain, A. K. (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):25–39.
- Dellaportas, P. and Roberts, G. O. (2003). An introduction to mcmc. In *Spatial statistics and computational methods*, pages 1–41. Springer.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741.
- Gudbjartsson, H. and Patz, S. (1995). The rician distribution of noisy mri data. *Magnetic resonance in medicine*, 34(6):910–914.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034.
- Jain, N., Yerragolla, S., Guha, T., and Mohana (2019). Performance analysis of object detection and tracking algorithms for traffic surveillance applications using neural networks. *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pages 690–696.

- Jospin, L. V., Buntine, W., Boussaid, F., Laga, H., and Bannamoun, M. (2020). Hands-on bayesian neural networks—a tutorial for deep learning users. *arXiv preprint arXiv:2007.06823*.
- Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Kolba, D. and Parks, T. (1977). A prime factor fft algorithm using high-speed convolution. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(4):281–294.
- Kornak, J., Boylan, R., Young, K., Wolf, A., Cobigo, Y., and Rosen, H. (2020). Bayesian image analysis in fourier space using data-driven priors (dd-bifs). In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 380–390. Springer.
- Kulkarni, R. (2000). Chapter 4: Frequency domain and fourier transforms. *Lecture Notes for ELE201 Introduction to Electrical Signals and Systems, Princeton University*, 2001:2002.
- Kuo, C.-C. J. (2016). Understanding convolutional neural networks with a mathematical model. *Journal of Visual Communication and Image Representation*, 41:406–413.
- Lupo, F. (2019). *Variational Autoencoder for unsupervised anomaly detection*. PhD thesis, Politecnico di Torino.
- Mao, X.-J., Shen, C., and Yang, Y.-B. (2016). Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv:1606.08921*.
- Nicolas, J.-M. and Tupin, F. (2019). A new parameterization for the rician distribution. *IEEE Geoscience and Remote Sensing Letters*, 17(11):2011–2015.
- Paisley, J., Blei, D., and Jordan, M. (2012). Variational bayesian inference with stochastic search. *arXiv preprint arXiv:1206.6430*.
- Papaspiliopoulos, O., Roberts, G. O., and Sköld, M. (2003). Non-centered parameterisations for hierarchical models and data augmentation. *Bayesian statistics*, 7:307–326.
- Peligrad, M., Utev, S., et al. (2006). Central limit theorem for stationary linear processes. *Annals of Probability*, 34(4):1608–1622.
- Raiffa, H. and Schlaifer, R. (1961). Applied statistical decision theory. Technical report.
- Rowe, D. B. and Logan, B. R. (2004). A complex way to compute fmri activation. *Neuroimage*, 23(3):1078–1092.



- Smith, G. (2019). The fast fourier transform and its applications.
- Smith, J. O. (2002). Mathematics of the discrete fourier transform (dft). *Center for Computer Research in Music and Acoustics (CCRMA), Department of Music, Stanford University, Stanford, California.*
- Strang, G., Strang, G., Strang, G., and Strang, G. (1993). *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA.
- Süli, E. and Mayers, D. F. (2003). *An introduction to numerical analysis*. Cambridge university press.
- Vetterli, M., Nussbaumer, H. J., et al. (1984). Simple fft and dct algorithms with reduced number of operations. *Signal processing*, 6(4):267–278.
- Zeger, S. L. (1985). Exploring an ozone spatial time series in the frequency domain. *Journal of the American Statistical Association*, 80(390):323–331.