



**Ανάπτυξη μιας βελτιστοποιημένης εκδοχής της μεθόδου
Hamiltonian Markov Chain Monte Carlo και εφαρμογή σε
προβλήματα Μπεϋζιανής Επικαιροποίησης**

Διπλωματική Εργασία

Αντώνιος Λύρας (cn18002)

**Επιβλέποντες: Βησσαρίων Παπαδόπουλος (Καθηγητής)
Ιωάννης Καλογεράς
Στέφανος Πυριαλάκος**

**Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Πολιτικών Μηχανικών
Τομέας Δομοστατικής
Αθήνα, 6 Ιουλίου 2023**

Η εργασία είναι αφιερωμένη στην οικογένειά μου, η οποία είναι ο βράχος και η πυξίδα μου.

Copyright © Λύρας Αντώνιος, 2023.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Σύντομη περιγραφή

Στην παρούσα εργασία παρουσιάζεται η μεθοδολογία Hamiltonian Markov Chain Monte Carlo στα πλαίσια προβλημάτων Μπεϋζιανής Επικαιροποίησης (Bayesian Update), με έμφαση σε μία βελτιστοποιημένη προγραμματιστική εκδοχή της, ενώ παρατίθεται και το ευρύτερο πλαίσιο γύρω από την ανάπτυξή της. Πραγματοποιείται αρχικά μια επισκόπηση του αντικειμένου της Μπεϋζιανής Στατιστικής, με εστίαση στην υποκατηγορία επικαιροποίησης και προσδιορισμού παραμέτρων (parametric identification) μαθηματικών και μηχανικών μοντέλων. Είναι γεγονός πως τόσο η αναλυτική όσο και η αριθμητική επίλυση των πραγματικών εφαρμογών Μπεϋζιανών Επικαιροποιήσεων είναι αδύνατες στις περισσότερες περιπτώσεις. Γι' αυτόν τον λόγο αναπτύχθηκαν οι δειγματοληπτικοί αλγόριθμοι, οι βασικές αρχές λειτουργίας των οποίων παρουσιάζονται εν συντομία. Στα προβλήματα επικαιροποίησης πολλαπλών παραμέτρων η πρώτη κατηγορία άμεσων δειγματοληπτικών αλγόριθμων δεν είναι αποτελεσματική, συνεπώς προκρίνεται η χρήση της δεύτερης, δηλαδή των αλγόριθμων Markov Chain Monte Carlo (MCMC). Οι συγκεκριμένες μεθοδολογίες λειτουργούν αναπτύσσοντας κατάλληλες εργοδικές αλυσίδες Markov και λαμβάνοντας τυχαία δείγματα Monte Carlo ικανού μεγέθους από την ζητούμενη επικαιροποιημένη κατανομή, η οποία ταυτίζεται με την στάσιμη κατανομή των αλυσίδων. Παρουσιάζεται μια ειδική κατηγορία μεθόδου MCMC, η Hamiltonian MCMC, η οποία, χάρη στην αποδοτική και γρήγορη εξερεύνηση του τυπικού συνόλου της επικαιροποιημένης κατανομής, είναι πιο αποτελεσματική από τις κλασικές. Αφού παρουσιαστεί το θεωρητικό της υπόβαθρο και οι αρχές λειτουργίας της, αναπτύσσεται μια βελτιστοποιημένη εκδοχή της Hamiltonian MCMC, η οποία προτάθηκε αρχικά από τους Hoffman & Gelman και η οποία επιλύει αυτόματα τη βαθμονόμηση των δύο παραμέτρων του αλγόριθμου (αριθμός βημάτων L και μήκος βήματος ϵ), απλοποιώντας έτσι την εφαρμογή της. Τέλος, πραγματοποιούνται δύο αριθμητικές εφαρμογές. Η πρώτη είναι μια απλή εφαρμογή επικαιροποίησης μια παραμέτρου, η οποία επιδέχεται αναλυτική λύση και βοηθάει στην επαλήθευση της ακρίβειας της προτεινόμενης μεθοδολογίας και την ορθότητα του κώδικα. Η δεύτερη και πιο σύνθετη εφαρμογή αναφέρεται στην επικαιροποίηση των τριών παραμέτρων ενός μη-γραμμικού νόμου υλικού Von Mises με αξιοποίηση μετρήσεων δύναμης-μετατόπισης σε ένα δομικό πρόβλημα πολιτικού μηχανικού.

Abstract

In the present diploma thesis, the Hamiltonian Markov Chain Monte Carlo methodology, which is used to perform Bayesian Updates, is presented, and emphasis is given on an optimized computational implementation, along with the context of its development. The thesis begins with a brief overview of the field of Bayesian Statistics, focusing on the subcategory of parametric identification in mathematical and mechanical models. It is a fact that the evaluation of the exact analytical or numerical solution is not viable in most real Bayesian Update applications. For this reason, sampling algorithms were developed. The way of operation, types, and main properties of these methodologies are briefly presented, highlighting the fact that in multi-parameter applications the first type of sampling algorithms (direct sampling algorithms) becomes inefficient. Thus, the other category (Markov Chain Monte Carlo (MCMC) algorithms) is preferable. MCMC methodologies work by developing appropriate ergodic Markov chains, constructed in a specific way to ensure that their stationary distribution is the same as the unknown posterior target distribution. Once the algorithm has run long enough to converge, a random Monte Carlo sample is drawn from the chain, approximating the real posterior distribution. A special MCMC methodology, the Hamiltonian MCMC, takes advantage of the properties of Hamiltonian Dynamics to perform a faster and more thorough exploration of the target distribution's typical set than the classic MCMC algorithms. After an overview of its theoretical foundation, an optimized modification of the Hamiltonian MCMC algorithm is presented, originally proposed by Hoffman & Gelman, which can automatically tune the algorithm's two main parameters (number of steps L and step size ϵ), thus rendering its implementation simple. To examine the effectiveness of the hybrid algorithm, finally, two numeric applications are performed: a relatively simple update of a single parameter, and a more complex example where the three parameters of the Von Mises Hardening Law are updated.

Πίνακας περιεχομένων

Σύντομη περιγραφή	2
Abstract	3
Πίνακας Εικόνων	5
1. Εισαγωγή.....	6
1.1. Bayesian Inference	6
1.2. Δειγματοληπτικοί αλγόριθμοι (sampling algorithms)	7
1.2.1. Δειγματοληψία Αποδοχής-Απόρριψης (Acceptance-Rejection sampling (ARS))	8
1.2.2. Markov Chain Monte Carlo	8
2. Η πορεία προς την Hamiltonian MCMC.....	12
2.1. Η σημασία του τυπικού συνόλου	12
2.1.1. Ένα βασικό χαρακτηριστικό των χώρων πολλών διαστάσεων	12
2.1.2. Το τυπικό σύνολο	13
2.2. Η ιδιαιτερότητα της Hamiltonian MCMC.....	15
3. Οικοδομώντας την Hamiltonian MCMC.....	19
3.1. Χαμιλτονιακή Δυναμική	19
3.1.1. Βασικές εξισώσεις.....	19
3.1.2. Βασικές ιδιότητες.....	20
3.2. Υπολογιστική υλοποίηση της Hamiltonian MCMC	22
3.2.1. Μέθοδος Leapfrog	22
3.2.2. Τοπικό και καθολικό σφάλμα των μεθόδων διακριτοποίησης	23
3.2.3. Ο βασικός αλγόριθμος Hamiltonian MCMC	23
3.3. Εργοδικότητα της Hamiltonian MCMC	26
3.4. Η βαθμονόμηση των παραμέτρων ϵ και L στην Hamiltonian MCMC.....	26
3.4.1. Βαθμονόμηση βήματος ϵ	26
3.4.2. Βαθμονόμηση αριθμού βημάτων L	27
4. Βελτιστοποιημένος αλγόριθμος Hamiltonian MCMC.....	28
4.1. Ο αλγόριθμος NUTS	28
4.1.1. Δομή του NUTS	28
4.1.2. Κριτήρια παύσης του NUTS	30
4.1.3. Συνάρτηση Build_Tree	31
4.2. Dual averaging για το ϵ	33
5. Αριθμητικές εφαρμογές.....	34

5.1. Μπεϋζιανή Επικαιροποίηση 1 τυχαίας μεταβλητής.....	34
5.2. Εφαρμογή Μπεϋζιανής Επικαιροποίηση 3 παραμέτρων μοντέλου.....	37
5.3. Συμπεράσματα.....	41
Βιβλιογραφικές αναφορές.....	43
Παράρτημα 1	44
Παράρτημα 2	50

Πίνακας εικόνων

Εικόνα 1: Καταμερισμός του πεδίου τιμών μιας κατανομής σε ορθογώνια κουτιά γύρω από την κυρίαρχη τιμή (Betancourt, 2017).	12
Εικόνα 2: Γραφική αναπαράσταση του τυπικού συνόλου (Betancourt, 2017). Με $\pi(q)$ αναπαρίσταται η πυκνότητα πιθανότητας, με dq ο όγκος ολοκλήρωσης και με $\pi(q)dq$ το τυπικό σύνολο. Ο οριζόντιος άξονας αναπαριστά την απόσταση σε απόλυτες τιμές από την επικρατούσα τιμή.	14
Εικόνα 3: Αλγόριθμος Random Walk Metropolis και τυπικό σύνολο (Betancourt, 2017). Με κόκκινο αναπαρίσταται το τυπικό σύνολο και με πράσινο η κατανομή των προτεινόμενων Μπεϋζιανών επικαιροποιήσεων.	15
Εικόνα 4: Αποτελεσματική εξερεύνηση του τυπικού συνόλου της ζητούμενης επικαιροποιημένης κατανομής (Betancourt, 2017).....	16
Εικόνα 5: Διανυσματικό πεδίο ευθυγραμμισμένο με το τυπικό σύνολο (Betancourt, 2017).	17
Εικόνα 6: Διανυσματικό πεδίο παραγόμενο από την παράγωγο της κατανομής στόχου (Betancourt, 2017).	18
Εικόνα 7: Ιστόγραμμα πρώτης αριθμητικής εφαρμογής.....	35
Εικόνα 8: Συναρτήσεις πυκνότητας πιθανότητας 1ης αριθμητικής εφαρμογής.....	36
Εικόνα 9: Αμφιέριστη δοκός η οποία υποβάλλεται σε συγκεντρωμένο φορτίο στο μέσο της (3-point test).....	37
Εικόνα 10: Ιστόγραμμα παραμέτρου a	40
Εικόνα 11: Ιστόγραμμα παραμέτρου b.....	40
Εικόνα 12: Ιστόγραμμα παραμέτρου c	40

1. Εισαγωγή

1.1. Bayesian Inference

Η Μπεϋζιανή Εξαγωγή Συμπεράσματος ή Μπεϋζιανή Επικαιροποίηση (Bayesian Inference/Bayesian Updating) αποτελεί μια ειδική κατηγορία εφαρμογών των στοχαστικών μεθόδων. Αναπτύχθηκε αρχικά στον επιστημονικό κλάδο της Στατιστικής, όμως αποδείχτηκε ιδιαίτερα χρήσιμη και σε προβλήματα μηχανικών, στους τομείς που άπτονται του περιορισμού της αβεβαιότητας παραμέτρων (parametric uncertainty) και της αβεβαιότητας μοντελοποίησης (modeling uncertainty). Πιο συγκεκριμένα, η Μπεϋζιανή Επικαιροποίηση εφαρμόζεται στις 2 ακόλουθες ειδικές κατηγορίες προβλημάτων:

- I. Προσδιορισμός παραμέτρων (parametric identification) σε ένα δεδομένο μαθηματικό ή μηχανικό μοντέλο, ορισμένες παράμετροι του οποίου είναι αβέβαιες, με βάση τις πραγματικές μετρήσεις του συστήματος (system response measurements).
- II. Επικαιροποίηση του ίδιου του μοντέλου (model updating), δηλαδή επιλογή ενός μαθηματικού μοντέλου από ένα σύνολο υποψήφίων παραμετρικών μοντέλων, με βάση τη βέλτιστη εναρμόνιση με τις πραγματικές μετρήσεις που έχουν καταγραφεί.

Στην παρούσα εργασία παρουσιάζεται μια ειδική προσέγγιση της πρώτης κατηγορίας (parametric identification).

Σύμφωνα με τη θεωρία πιθανοτήτων, εάν υποτεθεί ότι τα A και B είναι δύο γεγονότα, τότε η δεσμευμένη πιθανότητα του γεγονότος A να συμβεί με δεδομένη την πραγματοποίηση του γεγονότος B ισούται με:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \quad (1.1)$$

Κτίζοντας πάνω στον τύπο της δεσμευμένης πιθανότητας, δεδομένου ότι το πεδίο ενδιαφέροντος αφορά τον προσδιορισμό συνεχών παραμέτρων (continuous-valued parameters), παράγεται το ακόλουθο θεώρημα, το οποίο είναι μια ειδική περίπτωση του θεωρήματος του Bayes και εφαρμόζεται στην περίπτωση του προσδιορισμού παραμέτρων:

$$P(x/D) = \frac{P(D/x) * P(x)}{\int P(D/x) * P(x) dx} \quad (1.2)$$

όπου $\mathbf{X} = (X_1, \dots, X_N)$ είναι οι συνεχείς αβέβαιες παράμετροι των οποίων η συνάρτηση κατανομής πιθανότητας (pdf) θα ανανεωθεί, $\mathbf{x} = (x_1, \dots, x_N)$ οι αντίστοιχες τιμές τους, ενώ το D είναι το σύνολο των πραγματικών διακριτών μετρήσεων. Η $P(\mathbf{x}/D)$ ονομάζεται επικαιροποιημένη κατανομή ή κατανομή στόχου (posterior pdf/target distribution) και αποτελεί το ζητούμενο του προβλήματος. Η συνάρτηση $P(D/\mathbf{x})$ ονομάζεται συνάρτηση πιθανοφάνειας (likelihood function) και εισάγει στην εξίσωση τις πραγματικές μετρήσεις. Τέλος, ο παρονομαστής είναι μια σταθερά $k = \int P(D/\mathbf{x}) * P(\mathbf{x}) d\mathbf{x} = P(D)$, που υπολογίζεται μέσω της ολοκλήρωσης στο σύνολο του παραμετρικού χώρου Θ (parametric space) των παραμέτρων, και χρησιμοποιείται προκειμένου η επικαιροποιημένη κατανομή να ολοκληρώνεται στην μονάδα, να αποτελεί δηλαδή συνάρτηση πυκνότητας πιθανότητας.

Η Μπεϋζιανή Επικαιροποίηση θεμελιώνεται πάνω στην παραπάνω εξίσωση. Η αρχική πληροφόρηση για το αντικείμενο ενδιαφέροντος εμπεριέχεται στην αρχική κατανομή της συνάρτησης πιθανότητας (prior distribution), και στην συνέχεια αξιοποιούνται οι πραγματικές μετρήσεις μέσω της συνάρτησης πιθανοφάνειας προκειμένου να εξαχθεί μια νέα, επικαιροποιημένη και περισσότερο ακριβής συνάρτηση κατανομής πιθανότητας (posterior distribution), η οποία θα λαμβάνει υπόψιν (θα δεσμεύεται δηλαδή από) τα πραγματικά καταγεγραμμένα δεδομένα.

Η απευθείας εφαρμογή της εξίσωσης (1.2), όμως, στην πράξη καθίσταται πρακτικά αδύνατη από έναν θεμελιώδη περιορισμό. Πιο συγκεκριμένα, στα περισσότερα πραγματικά προβλήματα το ολοκλήρωμα του παρονομαστή της εξίσωσης είναι αδύνατο να προσδιοριστεί τόσο αναλυτικά όσο και με αριθμητική ολοκλήρωση (σε ρεαλιστικό χρονικό διάστημα). Κατά συνέπεια, δημιουργήθηκε η ανάγκη για μια διαφορετική προσέγγιση του προβλήματος, η οποία οδήγησε στην ανάπτυξη των δειγματοληπτικών αλγορίθμων (sampling algorithms).

1.2. Δειγματοληπτικοί αλγόριθμοι (sampling algorithms)

Οι δειγματοληπτικοί αλγόριθμοι βασίζονται στην αρχή πως, εφόσον δεν είναι δυνατόν να παραχθεί επακριβώς η επικαιροποιημένη κατανομή, επαρκεί να ληφθεί ένα αρκετά μεγάλο τυχαίο δείγμα τιμών από αυτήν (δείγμα Monte Carlo). Σε αυτήν την περίπτωση, μπορεί να υποθεθεί ορθώς, όπως έχει αποδειχτεί, ότι αυτό το τυχαίο δείγμα ισούται με την ακριβή κατανομή της επικαιροποιημένης κατανομής. Ακολούθως παρουσιάζονται (ο πρώτος με συνοπτικό τρόπο καθώς δεν αποτελεί αντικείμενο της παρούσας εργασίας) δύο ευρέως διαδεδομένοι δειγματοληπτικοί αλγόριθμοι, και σημειώνονται οι κύριες αδυναμίες τους.

1.2.1. Δειγματοληψία Αποδοχής-Απόρριψης (Acceptance-Rejection sampling (ARS))

Η βασική εξίσωση της Μπεϋζιανής Επικαιροποίησης, αν και στις περισσότερες περιπτώσεις δεν μπορεί να λυθεί άμεσα, όπως διατυπώθηκε προηγουμένως, εντούτοις επιτρέπει τον προσδιορισμό της ως ένα βαθμό, αναλογικά ως προς έναν συντελεστή σχήματος $1/k$. Ειδικότερα, άμεση απόρριξη της (1.2) αποτελεί το γεγονός ότι η επικαιροποιημένη κατανομή είναι ανάλογη της ακόλουθης ποσότητας η οποία ονομάζεται «μη προσαρμοσμένη κατανομή στόχου» (unscaled target distribution) και η οποία δύναται να υπολογιστεί :

$$P(\mathbf{x}/D) \propto P(D/\mathbf{x}) * P(\mathbf{x}) \quad (1.3)$$

Αξιοποιώντας αυτήν την ιδιότητα, ο αλγόριθμος ARS (όπως και άλλοι αντίστοιχοι δειγματοληπτικοί αλγόριθμοι Monte Carlo) λειτουργεί αναδιαμορφώνοντας ένα τυχαίο δείγμα που λαμβάνεται από μία αρχική υποψήφια κατανομή, αποδεχόμενος ορισμένες μόνο από τις τιμές στο τελικό δείγμα Monte Carlo και απορρίπτοντας τις υπόλοιπες. Κατ' αυτόν τον τρόπο, το τελικό δείγμα αποκτάει το ίδιο σχήμα (scaled up) με την ζητούμενη επικαιροποιημένη κατανομή και στην πράξη εξομοιώνεται με αυτήν.

Το κύριο μειονέκτημα του αλγόριθμου ARS, όμως, το οποίο τον καθιστά ακατάλληλο για τις περισσότερες εφαρμογές μηχανικών, είναι το γεγονός πως στα σχετικά σύνθετα προβλήματα (με μεγάλο πλήθος μετρήσεων) το ποσοστό αποδοχών μειώνεται δραματικά, με αποτέλεσμα να απαιτείται μη βιώσιμος υπολογιστικός χρόνος και ισχύς. Εκτός αυτού, μπορεί να κολλήσει σε περιπτώσεις όπου η μη προσαρμοσμένη κατανομή στόχου χαρακτηρίζεται από απότομες κορυφές (spikes). Κατά συνέπεια, γεννάται η ανάγκη για την ανάπτυξη διαφορετικών, καλύτερων δειγματοληπτικών αλγόριθμων.

1.2.2. Markov Chain Monte Carlo

Οι αλγόριθμοι Markov Chain Monte Carlo (Αλυσίδα Markov Monte Carlo (MCMC εν συντομία)) παρουσιάστηκαν αρχικά στα πλαίσια της πολύ γνωστής δημοσίευσης των Metropolis et al. το 1953. Στη δημοσίευση αυτή εισήχθη στην επιστημονική κοινότητα ο αλγόριθμος Metropolis-Hastings, η πρώτη μορφή αλγόριθμου MCMC που έχει την δυνατότητα να εκτελεί Μπεϋζιανές επικαιροποιήσεις. Εξαρχής κατέστη σαφές ότι οι μεθοδολογίες MCMC παρουσιάζουν σημαντικά συγκριτικά πλεονεκτήματα σχετικά με την μεθοδολογία ARS, καθώς μπορούν να αποφύγουν σε μεγάλο βαθμό τα δύο βασικά κολλήματα της ARS που παρατέθηκαν στην προηγούμενη ενότητα.

Ακολουθως, με δεδομένο ότι ο αλγόριθμος που αξιοποιήθηκε σε αυτήν την εργασία αποτελεί μια ειδική εξελιγμένη μορφή MCMC, παρουσιάζονται οι βασικές αρχές του MCMC και σημειώνονται οι κλασικοί αλγόριθμοι που τον εκτελούν, ενώ ειδική αναφορά γίνεται στα μειονεκτήματά τους.

Η αλυσίδα Markov (Markov Chain (MC)) είναι μια ειδική κατηγορία στοχαστικών διαδικασιών, στην οποία η μελλοντική εξέλιξη της διαδικασίας, δεδομένης της τωρινής και όλων των παρελθοντικών καταστάσεων της αλυσίδας, εξαρτάται αποκλειστικά από την πρώτη. Η συγκεκριμένη ιδιότητα στη βιβλιογραφία αναφέρεται ως ιδιότητα Markov. Αναπτύσσοντας την προηγούμενη πρόταση με μαθηματικούς όρους, μπορεί να υποστηριχτεί ότι μια αλυσίδα Markov είναι μια άπειρη ακολουθία εξαρτημένων τυχαίων μεταβλητών $\{X_1, X_2, \dots\}$, η οποία ικανοποιεί την ακόλουθη εξίσωση:

$$P(X_{n+1} \in B / X_1, X_2, \dots, X_n) = P(X_{n+1} \in B / X_n) \quad (1.3)$$

Από την παραπάνω εξίσωση εξάγεται το συμπέρασμα ότι μια αλυσίδα Markov μπορεί να προσδιοριστεί πλήρως μέσω της αρχικής κατάστασης X_1 και της κατανομής μετάβασης (transition distribution) από την κατάσταση X_n στην κατάσταση X_{n+1} .

Ορισμένα βασικά στοιχεία σχετικά με τις αλυσίδες Markov, η κατανόηση των οποίων είναι σημαντική για την συνέχεια, παρατίθενται συνοπτικά ακολούθως:

- Αν η πιθανότητα μετάβασης από μια τυχαία κατάσταση i της αλυσίδας σε μια άλλη τυχαία κατάσταση j παραμένει σταθερή ανεξαρτήτως του χρονικού βήματος n , τότε η αλυσίδα ονομάζεται ομογενής (homogeneous).
- Μια αλυσίδα Markov ονομάζεται ανάγωγη (irreducible) αν ο χώρος καταστάσεων της αποτελεί μια μοναδική κλάση, δηλαδή αν είναι δυνατή η μετάβαση από κάθε κατάσταση της αλυσίδας σε όλες τις υπόλοιπες καταστάσεις της.
- Μια κατάσταση i ονομάζεται παροδική (transient) όταν, ξεκινώντας από την κατάσταση i , υπάρχει μια μη μηδενική πιθανότητα η αλυσίδα να μην επανέλθει ποτέ σε αυτήν την κατάσταση.
- Η κατάσταση i , αντιθέτως, ονομάζεται επαναληπτική (recurrent) όταν δεν είναι παροδική. Μια αλυσίδα Markov της οποίας όλες οι καταστάσεις είναι επαναληπτικές ονομάζεται επαναληπτική.
- Μια αλυσίδα Markov ονομάζεται θετικά επαναληπτική (positive recurrent) όταν όλες οι καταστάσεις αυτής έχουν πεπερασμένο μέσο αναμενόμενο χρόνο επιστροφής.
- Μια αλυσίδα Markov ονομάζεται απεριοδική (aperiodic) όταν σε όλες τις καταστάσεις της επιστροφές μπορούν να συμβούν σε ακανόνιστους χρόνους, και όχι ανά προκαθορισμένα σταθερά χρονικά διαστήματα.

- ο Μια ομογενής, απεριοδική, μη υποβιβάσιμη αλυσίδα Markov, της οποίας όλες οι καταστάσεις είναι θετικά επαναληπτικές, ονομάζεται **εργοδική** (ergodic).

Πέραν αυτών, πρέπει να σημειωθεί ότι μια κατανομή f ονομάζεται στάσιμη (stationary) ως προς μία αλυσίδα Markov, όταν ισχύει πως αν η αρχική κατάσταση X_1 της αλυσίδας ακολουθεί την f , τότε και όλες οι μεταγενέστερες καταστάσεις X_n θα ακολουθούν επίσης την f . Οι εργοδικές αλυσίδες Markov έχει αποδειχτεί ότι έχουν μέχρι μία στάσιμη κατανομή.

Όλοι οι αλγόριθμοι MCMC χρησιμοποιούν εργοδικές αλυσίδες Markov. Ειδικότερα, λειτουργούν κατασκευάζοντας κατάλληλες εργοδικές αλυσίδες, των οποίων η στάσιμη κατανομή ταυτίζεται με την ζητούμενη επικαιροποιημένη κατανομή $P(x/D)$. Η ουσιαστική διαφορά τους από τους απλούς αλγόριθμους Monte Carlo συνίσταται στο γεγονός πως λαμβάνουν τυχαία δείγματα και κατασκευάζουν την επικαιροποιημένη κατανομή από μία εξαρτημένη ακολουθία τυχαίων μεταβλητών μιας αλυσίδας Markov, και όχι ανεξάρτητη. Το παραπάνω χαρακτηριστικό προσφέρει σημαντικά πλεονεκτήματα και έχει αποδειχτεί ότι δεν αναιρεί την αποτελεσματικότητα της μεθόδου Monte Carlo, εφόσον η αλυσίδα Markov που αξιοποιείται είναι εργοδική και έχει τρέξει για αρκετό χρονικό διάστημα (Bolstad, 2009). Η τελευταία συνθήκη είναι απαραίτητη καθώς η μακροχρόνια κατανομή (long-run distribution) μιας εργοδικής αλυσίδας θα ισούται με την στάσιμη κατανομή της. Τότε και λαμβάνοντας ένα επαρκώς μεγάλο τυχαίο δείγμα από την στάσιμη κατανομή, είναι δυνατή η ακριβής εξακρίβωση της επικαιροποιημένης κατανομής. Για τον παραπάνω λόγο, σε όλους τους δειγματοληπτικούς αλγόριθμους MCMC (συμπεριλαμβανομένου του Hamiltonian MCMC που παρουσιάζεται στην παρούσα εργασία), το δείγμα Monte Carlo ξεκινά να λαμβάνεται αφότου ο αλγόριθμος έχει τρέξει για έναν ικανό αριθμό επαναλήψεων (burn-in period), προκειμένου η αλυσίδα Markov να έχει συγκλίνει στην επικαιροποιημένη κατανομή. Στη βιβλιογραφία ως ενδεικτικό burn-in period στις περισσότερες περιπτώσεις ορίζονται οι 10.000 επαναλήψεις.

Αξιοποιώντας την παραπάνω ιδιαίτερα χρήσιμη ιδιότητα, οι κλασικοί αλγόριθμοι MCMC, οι κυριότεροι εκ των οποίων είναι η μέθοδος Metropolis-Hastings (ή Random Walk Metropolis) και η μέθοδος Gibbs Sampler, επιτυγχάνουν να εκτελέσουν Μπεϋζιανές επικαιροποιήσεις με αποδοτικότερο τρόπο συγκριτικά με τους υπόλοιπους δειγματοληπτικούς αλγορίθμους. Η ακριβής μεθοδολογία τους δεν θα αναλυθεί, καθώς δεν αποτελεί αντικείμενο της παρούσας εργασίας. Παρόλα αυτά, στα σύνθετα προβλήματα μηχανικών, η εφαρμογή των συνηθισμένων αλγορίθμων MCMC περιορίζεται από δύο βασικά προβλήματα.

Αρχικά, οι συγκεκριμένοι αλγόριθμοι χαρακτηρίζονται από μία εν γένει αργή εξερεύνηση του παραμετρικού χώρου της ζητούμενης επικαιροποιημένης κατανομής. Αυτό οφείλεται στο γεγονός ότι, ειδικότερα ο αλγόριθμος Metropolis – Hastings, λειτουργεί μέσω ενός τυχαίου περιπάτου (random walk) και δεν είναι προσαρμοσμένος στο τυπικό σύνολο (typical set) της κατανομής στόχου. Με άλλα λόγια, δεν υπάρχει εναρμόνιση μεταξύ του τρόπου λειτουργίας του αλγορίθμου και του τρόπου που κατανέμεται η μάζα της $P(x/D)$. Κατά συνέπεια αυξάνεται ο απαραίτητος αριθμός επαναλήψεων, και άρα ο υπολογιστικός χρόνος και το κόστος των

μοντέλων. Η σημασία του τυπικού συνόλου στην απόδοση των μεθοδολογιών MCMC αναπτύσσεται στο επόμενο κεφάλαιο.

Εκτός αυτού, όπως τονίζεται σε πλήθος βιβλιογραφικών αναφορών με χαρακτηριστική την επιχειρηματολογία του Neal (2011), οι κλασικοί αλγόριθμοι MCMC δεν ανταποκρίνονται ικανοποιητικά όταν αυξάνονται οι διαστάσεις D (dimensionality) των εφαρμογών, δηλαδή το πλήθος των παραμέτρων που επικαιροποιούνται. Πιο συγκεκριμένα, αποδεικνύεται (Neal, 2011) ότι το μέσο ποσοστό αποδοχών του αλγόριθμου Metropolis-Hastings μειώνεται προοδευτικά με την αύξηση των διαστάσεων, γεγονός που έχει ως απόρροια η μεθοδολογία να καθίσταται πρακτικά μη βιώσιμη (λόγω υπερβολικά μικρού αριθμού αποδοχών και άρα υπερβολικά υψηλού απαιτούμενου αριθμού επαναλήψεων του αλγόριθμου) στα προβλήματα που περιλαμβάνουν αριθμό παραμέτρων μεγαλύτερο από ένα όριο.

Υπό το πρίσμα των παραπάνω παρατηρήσεων, καθίσταται σαφής η ανάγκη για την ανάπτυξη μιας πιο αποδοτικής μεθοδολογίας MCMC, οι οποία θα αντιμετωπίζει αποτελεσματικά αυτούς ακριβώς τους περιορισμούς. Μία τέτοιου είδους μεθοδολογία είναι η Hamiltonian Markov Chain Monte Carlo), μια βελτιστοποιημένη μορφή της οποίας εφαρμόστηκε στα πλαίσια της παρούσας εργασίας, και η οποία παρουσιάζεται αναλυτικά στην συνέχεια.

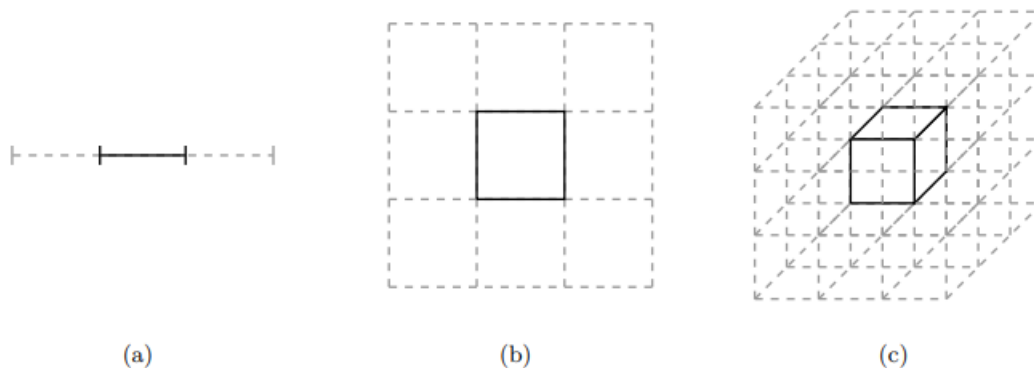
2. Η πορεία προς την Hamiltonian MCMC

2.1. Η σημασία του τυπικού συνόλου

Στο προηγούμενο κεφάλαιο έγινε μια σύντομη αναφορά στην σημασία της αποδοτικής εξερεύνησης του τυπικού συνόλου (typical set) ως προς την αποτελεσματικότητα των αλγορίθμων MCMC. Προκειμένου να καταστεί περισσότερο σαφής η αξία και τα συγκριτικά πλεονεκτήματα της μεθόδου Hamiltonian MCMC, αλλά και για να τεθεί η βάση για την θεμελίωσή της, ειδικότερα στα πλαίσια εφαρμογών πολλών διαστάσεων D (high dimensionality), κρίνεται απαραίτητο να παρουσιαστεί αναλυτικότερα, στην ακόλουθη ενότητα, η ερμηνεία και η βαρύτητα του τυπικού συνόλου.

2.1.1. Ένα βασικό χαρακτηριστικό των χώρων πολλών διαστάσεων

Είναι σημαντικό, αρχικά, να σημειωθεί ένα βασικό χαρακτηριστικό των χώρων πολλών διαστάσεων. Αποτελεί γεγονός πως όσο αυξάνονται οι διαστάσεις ενός χώρου, ο όγκος πέριξ μιας οποιαδήποτε περιοχής (neighborhood), συγκριτικά με τον όγκο της ίδιας της περιοχής, αυξάνεται επίσης και κυριαρχεί ως προς αυτόν. Η παραπάνω πρόταση καταδεικνύεται γραφικά στο ακόλουθο παράδειγμα.



Εικόνα 1: Καταμερισμός του πεδίου τιμών μιας κατανομής σε ορθογώνια κουτιά γύρω από την κυρίαρχη τιμή (Betancourt, 2017).

Στην συγκεκριμένη εικόνα πραγματοποιείται καταμερισμός του παραμετρικού χώρου μιας κατανομής, με την υπόθεση πως αυτό μπορεί να αναλυθεί σε ορθογώνια κουτιά, γύρω από την κυρίαρχη τιμή της (mode), στις περιπτώσεις όπου η εφαρμογή εξετάζεται σε μία, δύο και τρεις διαστάσεις. Ειδικότερα, στην μία διάσταση (a), το σχετικό βάρος της κεντρικής περιοχής όπου

βρίσκεται η κυρίαρχη τιμή ως προς τις γειτονικές της είναι $1/3$. Στις δύο διαστάσεις (b) το αντίστοιχο σχετικό βάρος είναι $1/9$, ενώ στις τρεις διαστάσεις (c), μειώνεται ακόμα περισσότερο, στο $1/27$.

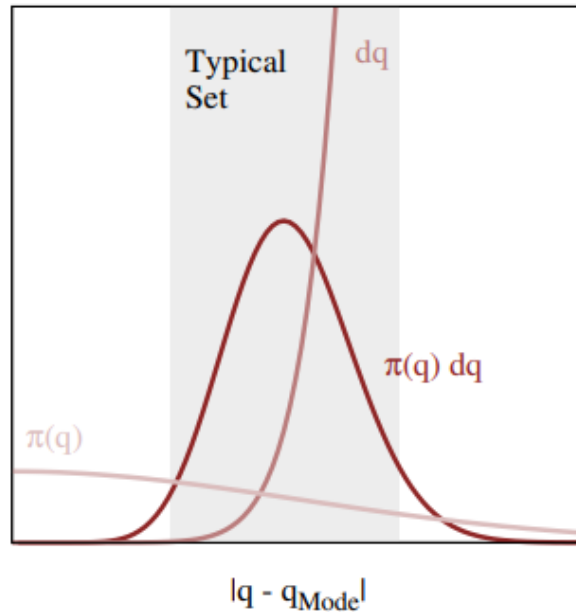
Γενικεύοντας με βάση αυτό το παράδειγμα, ισχύει πως σε ένα χώρο D διαστάσεων, υπάρχουν $3^D - 1$ γειτονικοί καταμερισμοί (neighboring partitions) περίξ μιας οποιασδήποτε δοθείσας περιοχής. Κατά συνέπεια, ακόμα και για σχετικό μικρό αριθμό διαστάσεων D , καθίσταται σαφές πως ο όγκος που συνορεύει με την επικρατούσα τιμή κυριαρχεί επί αυτής, ενώ το φαινόμενο αυτό ενισχύεται περαιτέρω αν ληφθούν υπόψιν και καταμερισμοί ακόμα πιο μακριά από την κυρίαρχη τιμή, οι οποίοι αυξάνονται με ακόμα ταχύτερο ρυθμό με την αύξηση του D .

Ερμηνεύοντας την παραπάνω ιδιότητα στα πλαίσια των εφαρμογών Μπεϋζιανών επικαιροποιήσεων, μπορεί να υποστηριχτεί ότι ο όγκος στις ουρές (tails) της κατανομής στόχου είναι κατά πολύ μεγαλύτερος του όγκου περίξ της επικρατούσας τιμής αυτής. Κατά συνέπεια, παρά το ότι η πυκνότητα της συνάρτησης κατανομής πιθανότητας της κατανομής στόχου είναι συγκριτικά πολύ μικρότερη στις ουρές της, το γεγονός ότι ο όγκος, ως προς τον οποίον γίνεται και η ολοκλήρωση της κατανομής για να εξαχθούν αναμενόμενες τιμές (expectations), σε αυτές τις περιοχές είναι σημαντικά μεγαλύτερος, όπως αποδείχτηκε, συνεπάγεται πως και από τις ουρές της κατανομής πραγματοποιούνται σημαντικές συνεισφορές στην αναμενόμενη επικαιροποιημένη κατανομή. Βάσει αυτού, εξάγεται το συμπέρασμα πως, προκειμένου να προσδιοριστούν και να απομονωθούν οι περιοχές του πεδίου παραμέτρων που συνεισφέρουν περισσότερο στην επικαιροποιημένη κατανομή, πρέπει να πραγματοποιηθεί εξισορρόπηση μεταξύ της πυκνότητας της κατανομής και του όγκου στον οποίον αυτή ολοκληρώνεται.

2.1.2. Το τυπικό σύνολο

Όπως αναφέρθηκε προηγουμένως, η περιοχή γειτονικά της κυρίαρχης τιμής χαρακτηρίζεται από υψηλή πυκνότητα πιθανότητας, όμως στα προβλήματα ενός αριθμού διαστάσεων και πάνω ο μικρός της όγκος την αποτρέπει από το να έχει σημαντική συνεισφορά στις αναμενόμενες τιμές της επικαιροποιημένης κατανομής. Από την άλλη πλευρά, οι συμπληρωματικές περιοχές μακριά από την κυρίαρχη τιμή καταλαμβάνουν πολύ μεγαλύτερο όγκο, όμως οι εξαιρετικά απομειωμένες πυκνότητες πιθανότητας εκεί προκαλούν επίσης αμελητέες συνεισφορές στην επικαιροποιημένη κατανομή. Οι μόνες σημαντικές συνεισφορές προέρχονται από μια περιοχή του παραμετρικού χώρου ανάμεσα στα δύο αυτά άκρα, η οποία ονομάζεται τυπικό σύνολο (Εικόνα 2). Στην πράξη, το τυπικό σύνολο ισούται με το γινόμενο των δύο προαναφερθέντων μεγεθών. Πρέπει να σημειωθεί πως επειδή η πυκνότητα πιθανότητας και ο όγκος μεταβάλλονται αντίστροφα όταν γίνεται οποιαδήποτε νέα παραμετροποίηση (reparameterization), το τυπικό σύνολο παραμένει ένα **αμετάβλητο μέγεθος** (invariant object) το οποίο είναι πλήρως ανεξάρτητο από τον τρόπο με τον οποίον γίνεται η παραμετροποίηση του προβλήματος, ενώ

εξαρτάται αποκλειστικά από την επικαιροποιημένη κατανομή της οποίας αποτελεί ιδιότητα (Betancourt, 2017).



Εικόνα 2: Γραφική αναπαράσταση του τυπικού συνόλου (Betancourt, 2017). Με $\pi(q)$ αναπαρίσταται η πυκνότητα πιθανότητας, με dq ο όγκος ολοκλήρωσης και με $\pi(q)dq$ το τυπικό σύνολο. Ο οριζόντιος άξονας αναπαριστά την απόσταση σε απόλυτες τιμές από την επικρατούσα τιμή.

Όσο αυξάνεται το πλήθος διαστάσεων D του παραμετρικού χώρου, η τριβή ανάμεσα στην πυκνότητα πιθανότητας και τον όγκο ολοένα και εντείνεται, ενώ οι περιοχές όπου και τα δύο μεγέθη είναι ικανά να προσφέρουν υπολογίσιμες συνεισφορές στην ζητούμενη επικαιροποιημένη κατανομή γίνονται αντίστοιχα ολοένα και πιο στενές. Κατά συνέπεια, το τυπικό σύνολο διαφάνεται με πιο μοναδικό (singular) τρόπο, όσο αυξάνεται το D , στα πλαίσια ενός φαινομένου της αρχής συγκέντρωσης μεγέθους (concentration of measure). Έτσι μπορεί να υποστηριχτεί με πολύ μικρό σφάλμα ότι, ειδικά στις εφαρμογές με περισσότερες από μία ή δύο διαστάσεις, οι μοναδικές συνεισφορές στην επικαιροποιημένη κατανομή, στα πλαίσια μιας Μπεϋζιανής επικαιροποίησης, πραγματοποιούνται εντός του τυπικού συνόλου. Ή, εναλλακτικά, μπορούν όλες οι περιοχές του παραμετρικού χώρου εκτός του τυπικού συνόλου να παραληφθούν τελείως από την εφαρμογή ως αμελητέες.

Σε αυτό το πλαίσιο, καθίσταται σαφές ότι, στις υπολογιστικές εφαρμογές Μπεϋζιανών επικαιροποιήσεων, ο ακριβής προσδιορισμός του τυπικού συνόλου της ζητούμενης επικαιροποιημένης κατανομής αποκτά πρωταρχική σημασία. Έχει αποδειχθεί στην πράξη εξάλλου και έπειτα από εμπειρία χρόνων, πως η ικανότητα των αλγορίθμων MCMC να προσδιορίσουν το τυπικό σύνολο και η εν γένει αποτελεσματικότητά τους είναι συνδεδεμένες με υψηλό συντελεστή συσχέτισης (Betancourt, 2017).

2.2. Η ιδιαιτερότητα της Hamiltonian MCMC

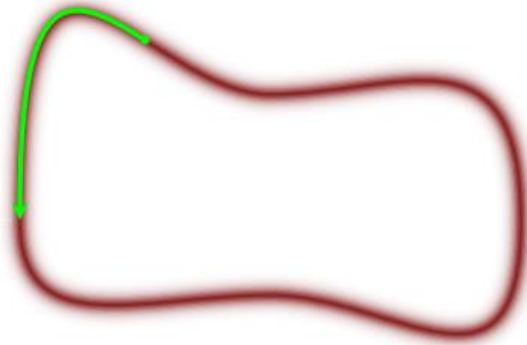
Το βασικό πρόβλημα του κλασικού αλγόριθμου τυχαίου περιπάτου (random walk) Metropolis-Hastings και των αντίστοιχων μεθοδολογιών είναι πως είναι καταδικασμένος να αποτύχει σε πολυδιάστατους χώρους. Αυτό συμβαίνει καθώς υπάρχει ένας εκθετικά αυξανόμενος αριθμός κατευθύνσεων κατά τις οποίες μπορεί να κινηθεί ο αλγόριθμος στα πλαίσια μιας Μπεϋζιανής επικαιροποίησης όσο αυξάνεται το D , όμως ένα πολύ μικρό ποσοστό εξ' αυτών βρίσκονται μέσα στο τυπικό σύνολο. Ειδικότερα, όσο οι διαστάσεις της κατανομής στόχου αυξάνονται, ο όγκος εξωτερικά του τυπικού συνόλου κυριαρχεί επί του όγκου εσωτερικά αυτού, με αποτέλεσμα σχεδόν κάθε προτεινόμενη επικαιροποίηση της μεθοδολογίας να βρίσκεται εκτός του τυπικού συνόλου, προς τις ουρές της κατανομής (Εικόνα 3(a)). Η πυκνότητα πιθανότητας σε αυτά τα σημεία, όμως, είναι τόσο μικρή, ώστε η πιθανότητα αποδοχής των προτάσεων είναι αμελητέα. Τότε και σχεδόν όλες οι προτεινόμενες επικαιροποιήσεις θα απορριφθούν, και η αλυσίδα Markov θα παραμείνει πρακτικά ακίνητη. Ένας τρόπος να αντιμετωπιστεί αυτό το πρόβλημα και να αυξηθεί η πιθανότητα αποδοχής είναι να περιοριστεί η διασπορά των προτάσεων ώστε αυτές να μείνουν εντός του τυπικού συνόλου, όμως τότε αυτά τα πολύ μικρά άλματα θα μετακινήσουν την αλυσίδα Markov εξαιρετικά αργά και αναποτελεσματικά (Εικόνα 3(b)), αδυνατώντας να εξερευνήσει τον παραμετρικό χώρο της επικαιροποιημένης κατανομής.



Εικόνα 3: Αλγόριθμος Random Walk Metropolis και τυπικό σύνολο (Betancourt, 2017). Με κόκκινο αναπαρίσταται το τυπικό σύνολο και με πράσινο η κατανομή των προτεινόμενων Μπεϋζιανών επικαιροποιήσεων.

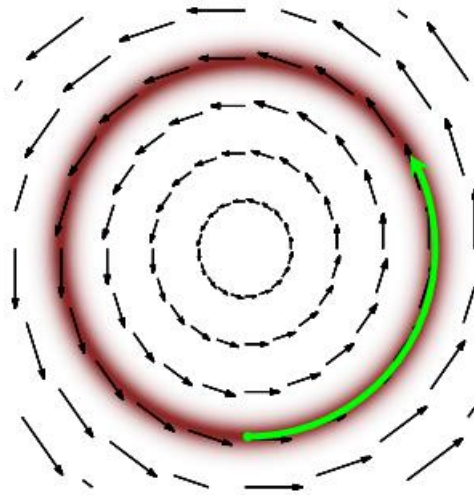
Με βάση τα παραπάνω, δημιουργείται η ανάγκη καλύτερης αξιοποίησης της **γεωμετρίας** του τυπικού συνόλου ούτως ώστε να επιτευχθεί η πραγματοποίηση μεγάλων αλμάτων μακριά από το αρχικό σημείο, εντός όμως του τυπικού συνόλου, τα οποία θα γίνονται αποδεκτά. Προς αυτόν τον στόχο, είναι κρίσιμος ο σχεδιασμός επικαιροποιήσεων οι οποίες μπορούν να ακολουθήσουν

τις διόδους υψηλής μάζας πιθανότητας (contours of high probability mass), προχωρώντας έτσι ομαλά μέσα στο τυπικό σύνολο (Εικόνα 4). Η μεθοδολογία Hamiltonian MCMC είναι μία μοναδική διαδικασία η οποία είναι σε θέση να πραγματοποιεί αυτόματα αλλά και ταυτόχρονα ομαλά αυτήν την εξερεύνηση στις περισσότερες περιπτώσεις των ζητούμενων κατανομών στόχου.



Εικόνα 4: Αποτελεσματική εξερεύνηση του τυπικού συνόλου της ζητούμενης επικαιροποιημένης κατανομής (Betancourt, 2017).

Το βασικό ζήτημα που τίθεται, σε αυτό το σημείο, είναι η «μετάφραση» της γεωμετρίας του τυπικού συνόλου σε εφαρμόσιμες πληροφορίες και μεθόδους για την αποτελεσματική εξερεύνησή του. Ένας τρόπος για να επιτευχθεί αυτό, όταν ο δειγματοληπτικός χώρος είναι συνεχής, είναι μέσω ενός διανυσματικού πεδίου (vector field) ευθυγραμμισμένου με το τυπικό σύνολο (Εικόνα 5). Ένα διανυσματικό πεδίο είναι, στην ουσία, η ανάθεση μιας κατεύθυνσης σε κάθε σημείο του παραμετρικού χώρου. Εάν οι κατευθύνσεις αυτές είναι εναρμονισμένες με το τυπικό σύνολο, τότε λειτουργούν ως οδηγός μέσα στην περιοχή της υψηλότερης μάζας πιθανότητας. Να σημειωθεί πως, στα πλαίσια αυτής της εργασίας, τυπικό σύνολο και περιοχή υψηλής μάζας πιθανότητας είναι συνώνυμες έννοιες.

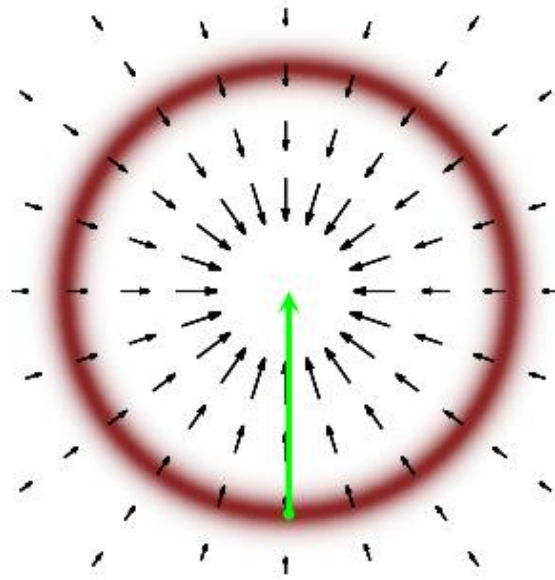


Εικόνα 5: Διανυσματικό πεδίο ευθυγραμμισμένο με το τυπικό σύνολο (Betancourt, 2017).

Με διαφορετική διατύπωση, στα πλαίσια ενός σωστά διαμορφωμένου διανυσματικού πεδίου, αποφεύγεται η τυφλή μετακίνηση μέσα στον παραμετρικό χώρο μέσω τυχαίων ανανεώσεων, ενώ προωθείται η πορεία κατά την κατεύθυνση που επιδεικνύεται στο εκάστοτε σημείο για ένα μικρό διάστημα. Κατ' αυτόν τον τρόπο, η αλυσίδα Markov θα φτάσει σε ένα καινούργιο παραπλήσιο σημείο, από το οποίο λαμβάνει μια νέα κατεύθυνση να ακολουθήσει, και ούτω καθεξής. Επεκτείνοντας αυτή τη διαδικασία, καθίσταται δυνατή μια ομαλή και συνεχής εξερεύνηση του τυπικού συνόλου γρήγορα και χωρίς αποκλίσεις.

Το πρόβλημα που ανακύπτει, εδώ, είναι ο τρόπος με τον οποίον θα κατασκευαστεί το κατάλληλο διανυσματικό πεδίο. Οι πληροφορίες σχετικά με την κατανομή στόχου, οι οποίες δεν έχουν ακόμα αξιοποιηθεί, αφορούν τη διαφορική δομή (differential structure) αυτής και μια μέθοδος για να αποκωδικοποιηθούν είναι μέσω της παραγώγου της επικαιροποιημένης συνάρτησης πυκνότητας πιθανότητας. Ειδικότερα, η παράγωγος ορίζει ένα διανυσματικό πεδίο στον παραμετρικό χώρο το οποίο είναι άμεσα συνδεδεμένο με την κατανομή στόχου (Εικόνα 6).

Το εν λόγω διανυσματικό πεδίο, όμως, δυστυχώς δεν είναι το ζητούμενο, καθώς απλώς η σύνδεσή του με την επικαιροποιημένη κατανομή δεν επαρκεί για να το εναρμονίσει με το τυπικό της σύνολο. Πιο συγκεκριμένα, όπως φαίνεται στην εικόνα 6, η καθοδήγηση της παραγώγου τελικά απομακρύνει την αλυσίδα Markov από το τυπικό σύνολο και την οδηγεί προς την επικρατούσα τιμή. Αυτό συμβαίνει καθώς η κατανομή στόχου εξαρτάται από την παραμετροποίηση, οπότε και η παράγωγος αυτής θα εξαρτάται επίσης από την παραμετροποίηση. Κατά συνέπεια, η παράγωγος είναι σε θέση να καθοδηγήσει μόνο προς περιοχές που είναι ευαίσθητες ως προς την παραμετροποίηση, δηλαδή την επικρατούσα τιμή, και όχι προς περιοχές που είναι ανεξάρτητες από αυτήν, όπως είναι το τυπικό σύνολο.



Εικόνα 6: Διανυσματικό πεδίο παραγόμενο από την παράγωγο της κατανομής στόχου (Betancourt, 2017).

Προκειμένου να επιλυθεί το παραπάνω πρόβλημα, θα πρέπει η πληροφορία για την παράγωγο της επικαιροποιημένης κατανομής να συνδυαστεί με επιπλέον γεωμετρικούς προσδιορισμούς, οι οποίοι θα αφαιρούν την εξάρτηση από την εκάστοτε παραμετροποίηση και θα ευθυγραμμίζουν τις κατευθύνσεις των διανυσμάτων με το τυπικό σύνολο. Ο επιστημονικός κλάδος που ασχολείται με ζητήματα τέτοιου είδους ονομάζεται διαφορική γεωμετρία (differential geometry). Η διαφορική γεωμετρία είναι ένα αρκετά σύνθετο αντικείμενο και στα πλαίσια αυτής της εργασίας δεν θα γίνει διεξοδική ανάλυσή της. Αξιοποιείται, όμως, μια ιδιαίτερως χρήσιμη ιδιότητά της στο συγκεκριμένο πρόβλημα, η οποία ορίζει πως η ίδια ακριβώς γνώση διαφορικής γεωμετρίας που απαιτείται για τη διόρθωση των παραγώγων της κατανομής στόχου συμπίπτει με τα μαθηματικά που περιγράφουν την κλασική φυσική. Με διαφορετική διατύπωση, σε κάθε πιθανοτικό/στοχαστικό (probabilistic) σύστημα αντιστοιχεί ένα μαθηματικά ισοδύναμο φυσικό σύστημα, το οποίο μπορεί να αναλυθεί πιο εύκολα (Betancourt, 2017).

Σε αυτό το πλαίσιο, αποδεικνύεται ότι το κλειδί για τον σωστό μετασχηματισμό του διανυσματικού πεδίου παραμέτρων σε διανυσματικό πεδίο εναρμονισμένο με το τυπικό σύνολο είναι η επέκταση του αρχικού στοχαστικού συστήματος με την εισαγωγή βοηθητικών παραμέτρων ορμής (auxiliary momentum parameters). Οι συγκεκριμένες παράμετροι, προκειμένου να είναι αποτελεσματικές, δεν θα πρέπει να εισαχθούν αυθαίρετα. Αντιθέτως, οφείλουν να είναι σύμφωνες με την πιθανοτική δομή που χαρακτηρίζει τη Δυναμική της κλασικής Φυσικής. Στην μέχρι σήμερα βιβλιογραφία, μόνο μία μεθοδολογία έχει αναπτυχθεί για την εισαγωγή σε αυτού του είδους τα προβλήματα των εν λόγω βοηθητικών παραμέτρων ορμής με την συγκεκριμένη δομή: Η μέθοδος Hamiltonian Markov Chain Monte Carlo.

3. Οικοδομώντας την Hamiltonian MCMC

3.1. Χαμιλτονιακή Δυναμική

Στη Χαμιλτονιακή Δυναμική, κάθε παράμετρος ενδιαφέροντος του εξεταζόμενου συστήματος περιγράφεται πλήρως από την μεταβλητή θέσης της, η οποία είναι ένα διάνυσμα q , και την μεταβλητή ορμής της, η οποία είναι επίσης διάνυσμα και συμβολίζεται με p . Στην θέση q βασίζεται η δυναμική ενέργεια $U(q)$ της παραμέτρου, ενώ πάνω στην ορμή p θεμελιώνεται η κινητική της ενέργεια $K(p)$, με τύπους που θα αναλυθούν ακολούθως. Στις πρακτικές Χαμιλτονιακές εφαρμογές μηχανικού, η μεταβλητή θέσης αντιστοιχεί στην παράμετρο ενδιαφέροντος της οποίας η Μπεϋζιανή επικαιροποίηση είναι ο στόχος, και η δυναμική ενέργεια ορίζεται ως η αρνητική τιμή του λογαρίθμου της επικαιροποιημένης συνάρτησης κατανομής πιθανότητας της παραμέτρου ενδιαφέροντος. Οι μεταβλητές ορμής εισάγονται τεχνητά και αυθαίρετα, με τρόπο κατάλληλο ως προς τις απαιτήσεις του εκάστοτε προβλήματος.

3.1.1. Βασικές εξισώσεις

Με βάση τα παραπάνω, εξάγεται το συμπέρασμα ότι σε μια εφαρμογή D διαστάσεων, η Χαμιλτονιακή Δυναμική θα λειτουργεί μέσω ενός διανύσματος θέσης q D διαστάσεων, το οποίο θα συμπληρώνεται από ένα διάνυσμα ορμής p επίσης D διαστάσεων, με το συνολικό πεδίο τιμών (space state) να έχει $2D$ διαστάσεις. Το σύστημα περιγράφεται μέσω μιας κοινής συνάρτησης των q και p η οποία ονομάζεται Hamiltonian και συμβολίζεται με $H(q,p)$. Παρακάτω παρατίθενται οι καταστάσεις κίνησης της Χαμιλτονιακής Δυναμικής, εκφρασμένες ως μερικές παράγωγοι, οι οποίες υποδεικνύουν πως μεταβάλλονται οι μεταβλητές q και p με το χρόνο t .

$$\frac{\partial q_i}{\partial t} = \frac{\partial H}{\partial p_i} \quad (3.1)$$

$$\frac{\partial p_i}{\partial t} = -\frac{\partial H}{\partial q_i} \quad (3.2)$$

Για $i=1,\dots,D$.

Στις εφαρμογές Hamiltonian MCMC, έχει επικρατήσει στη βιβλιογραφία η ακόλουθη γραφή της συνάρτησης Hamiltonian:

$$H(q, p) = U(q) + K(p) \quad (3.3)$$

Η κινητική ενέργεια έχει επικρατήσει να λαμβάνει την ακόλουθη μορφή, η οποία αντιστοιχεί στην αρνητική τιμή του λογαρίθμου της Κανονικής κατανομής με μηδενικό μέσο όρο και πίνακα συνδιασπορών (covariance matrix) M .

$$K(p) = p^T M^{-1} p / 2 \quad (3.4)$$

Με βάση την εξίσωση (2.4), οι σχέσεις (2.1) και (2.2) μπορούν να γραφτούν και ως εξής:

$$\frac{\partial q_i}{\partial t} = [M^{-1}p]_i \quad (3.5)$$

$$\frac{\partial p_i}{\partial t} = -\frac{\partial U}{\partial q_i} \quad (3.6)$$

3.1.2. Βασικές ιδιότητες

Οι ακόλουθες ιδιότητες της Χαμιλτονιακής Δυναμικής είναι κρίσιμες, μεταξύ άλλων, για την αποτελεσματική πραγματοποίηση Μπεϋζιανών επικαιροποιήσεων MCMC.

- I. **Αντιστρεψιμότητα** (reversibility): Η Χαμιλτονιακή Δυναμική είναι αντιστρέψιμη, δηλαδή η πορεία (mapping) T_s από την κατάσταση $(q(t), p(t))$ τη χρονική στιγμή t στην κατάσταση $(q(t+s), p(t+s))$ τη χρονική στιγμή $t+s$ είναι 1-προς-1. Συνεπώς για κάθε πορεία T_s θα υπάρχει και μια αντίστροφη πορεία T_{-s} , η οποία μπορεί να προσδιοριστεί με την λήψη των αντίθετων τιμών των μερικών παραγώγων ως προς το χρόνο στις εξισώσεις (3.1) και (3.2). Όταν επιπλέον η συνάρτηση Hamiltonian είναι της μορφής (3.3) και η κινητική ενέργεια είναι του τύπου (3.4), όπως είθισται, η αντίστροφη πορεία μπορεί να βρεθεί με ακόμα πιο απλό τρόπο, θέτοντας την αντίθετη τιμή στο p , υπολογίζοντας την κανονική πορεία, και τέλος εισάγοντας πάλι την αντίθετη τιμή του p . Η αντιστρεψιμότητα της Χαμιλτονιακής Δυναμικής είναι ιδιαίτερα σημαντική για τις ανανεώσεις MCMC καθώς εξασφαλίζει ότι η κατανομή στόχου θα παραμείνει αμετάβλητη (invariant) κατά την επικαιροποίηση.

- II. **Διατήρηση της συνάρτησης Hamiltonian** (Conservation of the Hamiltonian): Στη Χαμιλτονιακή Δυναμική η συνάρτηση Hamiltonian παραμένει αμετάβλητη με το χρόνο, όπως φαίνεται ακολούθως μέσω απλής επεξεργασίας των εξισώσεων (3.1) και (3.2).

$$\frac{\partial H}{\partial t} = \sum_{i=1}^D \left[\frac{dq_i}{dt} \frac{\partial H}{\partial q_i} + \frac{dp_i}{dt} \frac{\partial H}{\partial p_i} \right] = \sum_{i=1}^D \left[\frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} \right] = 0 \quad (3.7)$$

Στις επικαιροποιήσεις MCMC η πιθανότητα αποδοχής είναι θεωρητικά ίση με 1 εάν η Hamiltonian διατηρείται ακριβώς αμετάβλητη. Στις πραγματικές εφαρμογές, όμως, κάτι τέτοιο αποδεικνύεται ανεπίτευκτο (Neal, 2011). Συνεπώς, επιδιώκεται η συνάρτηση H να παραμένει **περίπου αμετάβλητη**.

- III. **Διατήρηση όγκου** (volume preservation): Μια τρίτη βασική ιδιότητα της Χαμιλτονιακής Δυναμικής είναι πως αυτή διατηρεί τον όγκο στο πεδίο (q,p) , η οποία εμφανίζεται στη βιβλιογραφία και ως Θεώρημα Liouville. Ειδικότερα, εάν πραγματοποιηθεί μια πορεία T_s σε κάποια περιοχή R του πεδίου (q,p) η οποία καταλαμβάνει όγκο V , τότε η εικόνα του R υπό την T_s θα έχει επίσης όγκο V . Η συγκεκριμένη ιδιότητα συνεπάγεται πρακτικά, όσον αφορά τις Μπεϋζιανές επικαιροποιήσεις, πως δεν απαιτείται να προσμετρηθούν μεταβολές στον όγκο κατά τον προσδιορισμό της πιθανότητας αποδοχής του αλγορίθμου Hamiltonian MCMC σε κάθε επανάληψη.
- IV. **Συμπλεκτικότητα** (symplecticness): Μια ακόμα ιδιότητα της Χαμιλτονιακής Δυναμικής, η οποία συνεισφέρει στη διατήρηση όγκου που την χαρακτηρίζει, συνεπώς είναι άξια αναφοράς, είναι η συμπλεκτικότητα. Η συμπλεκτικότητα εκφράζεται μέσω της παρακάτω εξίσωσης (Neal, 2011):

$$B_s^T J^{-1} B_s = J^{-1} \quad (3.8)$$

Όπου ο όρος B_s είναι ο Ιακωβιανός (Jacobian) πίνακας της πορείας T_s και ο όρος J είναι ένας πίνακας $2D \times 2D$ του οποίου το περιεχόμενο παρατίθεται ακολούθως:

$$J = \begin{bmatrix} 0_{D \times D} & I_{D \times D} \\ I_{D \times D} & 0_{D \times D} \end{bmatrix} \quad (3.9)$$

Είναι σημαντικό να σημειωθεί πως η συμπλεκτικότητα συνεχίζει να ισχύει και στις πρακτικές υπολογιστικές εφαρμογές της μεθόδου Hamiltonian MCMC.

3.2. Υπολογιστική υλοποίηση της Hamiltonian MCMC

Κατά τη διαδικασία υπολογιστικής υλοποίησης Μπεϋζιανών Επικαιροποιήσεων της κατηγορίας Markov Chain Monte Carlo βασισμένων στη Χαμιλτονιακή Δυναμική, είναι απαραίτητο να γίνουν ορισμένες προσαρμογές. Πιο συγκεκριμένα, οι βασικές Χαμιλτονιακές εξισώσεις που παρουσιάστηκαν σε προηγούμενη ενότητα προσεγγίζονται μέσω της διακριτοποίησης του χρόνου σε μικρά χρονικά διαστήματα, τα οποία ονομάζονται βήματα (stepsizes) και συμβολίζονται με ϵ . Το ϵ είναι μια από τις παραμέτρους της μεθοδολογίας που ο χρήστης πρέπει να θέσει. Ο αλγόριθμος Hamiltonian MCMC εκκινεί από την αρχική κατάσταση του πεδίου (q_0, p_0) τη χρονική στιγμή $t=0$ και έπειτα υπολογίζει την κατάσταση διαδοχικά κατά τις χρονικές στιγμές $\epsilon, 2*\epsilon, 3*\epsilon$ κ.λπ. μέχρις ότου να του δοθεί η εντολή να σταματήσει.

Στα πλαίσια της εφαρμογής Hamiltonian MCMC της παρούσας εργασίας, γίνεται η παραδοχή πως η συνάρτηση ακολουθεί την μορφή (3.3) και πως η κινητική ενέργεια είναι του τύπου (3.4), δηλαδή αρνητική λογαριθμική Γκαουσιανή κατανομή με μηδενικό μέσο όρο και διασπορές m_i . Αν επιπλέον θεωρηθεί, χωρίς βλάβη της γενικότητας, πως ο πίνακας συνδιασπορών M είναι διαγώνιος με διαγώνια στοιχεία m_1, \dots, m_D , θα ισχύει πως:

$$K(p) = \sum_{i=1}^D \frac{p_i^2}{2m_i} \quad (3.10)$$

Το επόμενο στάδιο στη διαδικασία της υπολογιστικής υλοποίησης είναι η επιλογή του τρόπου με τον οποίο θα προσεγγιστούν οι λύσεις των μερικών διαφορικών εξισώσεων της Χαμιλτονιακής Δυναμικής. Αρκετές μέθοδοι διακριτοποίησης έχουν προταθεί στη βιβλιογραφία, ορισμένες εκ των οποίων είναι η μέθοδος Euler και η προσαρμοσμένη μέθοδος Euler. Όπως όμως έχει αποδεχθεί, η μέθοδος με το μικρότερο σφάλμα και συνεπώς πιο αποτελεσματική στις εφαρμογές Hamiltonian MCMC είναι η μέθοδος Leapfrog.

3.2.1. Μέθοδος Leapfrog

Η μέθοδος Leapfrog αποτυπώνεται στις ακόλουθες 3 σχέσεις όσον αφορά την επεξεργασία της δυναμικής ενέργειας:

$$p_i(t + \epsilon/2) = p_i(t) - (\epsilon/2) \frac{\partial U}{\partial q_i}(q(t)) \quad (3.11)$$

$$q_i(t + \varepsilon) = q_i(t) + \varepsilon \frac{p_i(t + \varepsilon/2)}{m_i} \quad (3.12)$$

$$p_i(t + \varepsilon) = p_i(t + \varepsilon/2) - (\varepsilon/2) \frac{\partial U}{\partial q_i}(q(t + \varepsilon)) \quad (3.13)$$

Στην πρώτη σχέση πραγματοποιείται μισό βήμα για τις μεταβλητές ορμής, έπειτα ένα πλήρες βήμα για τις μεταβλητές θέσης, χρησιμοποιώντας τις νέες τιμές των μεταβλητών ορμής, και τέλος γίνεται ακόμα μισό βήμα για τις μεταβλητές ορμής, με αξιοποίηση των νέων τιμών των μεταβλητών θέσης. Η παραπάνω περιγραφή αποδίδει μία επανάληψη (iteration) του αλγορίθμου. Μια ανάλογη διαδικασία μπορεί να εφαρμοστεί και για την κινητική ενέργεια, μέσω αντικατάστασης στους παραπάνω τύπους του όρου $\frac{\partial U}{\partial q_i}$ με τον όρο $\frac{\partial K}{\partial p_i}$.

Η μέθοδος Leapfrog αποδεικνύεται ότι διατηρεί τον όγκο επακριβώς (Neal, 2011) και επιπλέον, λόγω της συμμετρίας της, είναι εύκολα αναστρέψιμη, μέσω της εισαγωγής αντίθετου p , εφαρμογής των ίδιων βημάτων και τέλος εισαγωγής πάλι αντίθετου p .

3.2.2. Τοπικό και καθολικό σφάλμα των μεθόδων διακριτοποίησης

Στις μεθόδους διακριτοποίησης της Hamiltonian MCMC, το τοπικό σφάλμα (local error) είναι το σφάλμα μετά από ένα βήμα ε , το οποίο μεταφέρεται από τη χρονική στιγμή t στη χρονική στιγμή $t + \varepsilon$. Το καθολικό σφάλμα (global error) είναι το σφάλμα μετά από προσομοίωση για ένα προκαθορισμένο χρονικό διάστημα s , κατά το οποίο έχουν πραγματοποιηθεί s/ε βήματα. Εξυπακούεται ότι το s πρέπει να είναι ακέραιο πολλαπλάσιο του ε . Ισχύει πως αν το τοπικό σφάλμα είναι τάξης ε^p , το καθολικό σφάλμα θα είναι τάξης ε^{p-1} . Αυτό συμβαίνει καθώς το πλήθος των επιμέρους σφαλμάτων ε^p συγκεντρώνεται κατά τα s/ε βήματα, ώστε τελικά να αποδώσει ένα συνολικό σφάλμα τάξης ε^{p-1} .

Έχει αποδειχθεί ότι η Μέθοδος Euler και η Προσαρμοσμένη Μέθοδος Euler έχουν τοπικό σφάλμα ε^2 και ολικό σφάλμα ε . Αντιθέτως, η Μέθοδος Leapfrog χαρακτηρίζεται από τοπικό σφάλμα ε^3 και ολικό σφάλμα ε^2 (Neal, 2011), συνεπώς είναι πιο ακριβής και αποτελεσματική. Το συγκεκριμένο χαρακτηριστικό οφείλεται στο γεγονός ότι, όπως έδειξαν οι Leimkuhler & Reich (2004), η Μέθοδος Leapfrog είναι απολύτως αναστρέψιμη, και ισχύει πως κάθε αναστρέψιμη μέθοδος πρέπει να έχει ολικό σφάλμα του οποίου η τάξη μεγέθους είναι άρτια (σε αυτήν την περίπτωση 2).

3.2.3. Ο βασικός αλγόριθμος Hamiltonian MCMC

Η αξιοποίηση της Χαμιλτονιακής Δυναμικής για τη δειγματοληψία από την κατανομή στόχου, στα πλαίσια ενός αλγορίθμου MCMC, προϋποθέτει αρχικά την μετατροπή της συγκεκριμένης κατανομής σε μια συνάρτηση δυναμικής ενέργειας U . Ακολούθως απαιτείται η εισαγωγή κατάλληλων μεταβλητών ορμής p , οι οποίες θα συμπληρώνουν τις παραμέτρους ενδιαφέροντος (οι οποίες στον αλγόριθμο εκλαμβάνονται ως μεταβλητές θέσης q). Τότε καθίσταται δυνατή η προσομοίωση μιας αλυσίδας Markov στην οποία σε κάθε επανάληψη γίνεται νέα τυχαία δειγματοληψία της ορμής από την κατανομή που έχει οριστεί για αυτήν (Κανονική κατανομή μηδενικού μέσου όρου) και έπειτα επικαιροποιούνται οι τιμές της θέσης, το οποίο είναι το ζητούμενο, μέσω της αντίστοιχης πιθανότητας μετάβασης από μια κατάσταση της αλυσίδας Markov σε μία άλλη.

Ειδικότερα, η δυναμική ενέργεια λαμβάνει την ακόλουθη μορφή (η οποία στην πράξη είναι μη-αρνητική), στην οποία ο όρος $\pi(q)$ είναι η αρχική κατανομή (prior density) και ο όρος $L(q/D)$ είναι η συνάρτηση πιθανοφάνειας (likelihood function), ενώ η κινητική ενέργεια ακολουθεί τον τύπο που παρουσιάστηκε στην εξίσωση (3.10).

$$U(q) = -\log[\pi(q) * L(q / D)] \quad (3.14)$$

Πρέπει να σημειωθεί ότι, προκειμένου να λειτουργήσει σωστά ο αλγόριθμος, απαιτείται να μπορούν να προσδιοριστούν αριθμητικά και οι μερικές παράγωγοι της δυναμικής ενέργειας, πέραν από τις ίδιες τις τιμές της.

Κάθε επανάληψη του βασικού αλγορίθμου Hamiltonian HMC αποτελείται από δύο στάδια. Το πρώτο μεταβάλλει μόνο τις μεταβλητές p , ενώ το δεύτερο ενδέχεται να μεταβάλλει τόσο τις μεταβλητές p όσο και τις μεταβλητές q . Και τα δύο στάδια είναι αντιστρέψιμα, οπότε δεν αλλοιώνεται η ζητούμενη επικαιροποιημένη κατανομή.

Κατά το πρώτο στάδιο, D (όσες οι διαστάσεις της εφαρμογής) νέες τιμές των μεταβλητών ορμής p επιλέγονται τυχαία από την Κανονική κατανομή τους, ανεξάρτητα από τις τιμές των μεταβλητών θέσης q και ανεξάρτητες μεταξύ τους. Έπειτα υπολογίζεται η κινητική ενέργεια βάσει του τύπου της εξίσωσης (3.10), όπου συνήθως γίνεται η παραδοχή πως οι συντελεστές m ισούνται με μονάδα.

Κατά το δεύτερο στάδιο, προτείνεται μία νέα κατάσταση της αλυσίδας Markov του προβλήματος μέσω μιας επικαιροποίησης Metropolis (Metropolis update) με τον ακόλουθο τρόπο. Εκκινώντας από την υφιστάμενη κατάσταση (q,p) , η Χαμιλτονιακή Δυναμική προσομοιώνεται για L βήματα με τη χρήση της μεθόδου Leapfrog, όπου το κάθε βήμα είναι μήκους ϵ . Το L είναι η δεύτερη κρίσιμη παράμετρος για την αποτελεσματικότητα του αλγορίθμου, μετά το ϵ , την οποία θέτει ο χρήστης. Οι μεταβλητές p που λαμβάνονται στο τέλος της προσομοιωμένης τροχιάς L βημάτων πολλαπλασιάζονται επί (-1) , δηλαδή αρνητικοποιούνται, δημιουργώντας έτσι μια νέα

προτεινόμενη κατάσταση (q^*, p^*) . Η κατάσταση αυτή, τέλος, κρίνεται αποδεκτή ως η επόμενη κατάσταση της αλυσίδας Markov με πιθανότητα P , η οποία ισούται με:

$$P = \min[1, \exp(-H(q^*, p^*) + H(q, p))] \quad (3.15)$$

Αν η πρόταση δεν γίνει αποδεκτή, τότε η επόμενη κατάσταση της αλυσίδας Markov ισούται με την αμέσως προηγούμενη. Η αρνητικοποίηση των μεταβλητών ορμής στο τέλος της τροχιάς είναι ιδιαίτερα σημαντική στη θεωρία, καθώς κάνει την προτεινόμενη κατάσταση συμμετρική, ιδιότητα που είναι αναγκαία προκειμένου η πιθανότητα αποδοχής να είναι έγκυρη (Neal, 2011). Στην πράξη όμως μπορεί να παραλειφθεί, καθώς για την συγκεκριμένη μορφή κινητικής ενέργειας που αξιοποιείται στο συγκεκριμένο και στα περισσότερα παραδείγματα στη βιβλιογραφία ισχύει πως $K(p) = K(-p)$, και επιπλέον οι μεταβλητές ορμής αντικαθίστανται πριν επαναχρησιμοποιηθούν, κατά το πρώτο στο στάδιο της επόμενης επανάληψης του αλγορίθμου.

Κατά το δεύτερο στάδιο κάθε επανάληψης, όπου εφαρμόζονται οι αρχές της Χαμιλτονιακής Δυναμικής, η συνάρτηση Hamiltonian παραμένει περίπου σταθερή. Κατά συνέπεια η από κοινού συνάρτηση κατανομής πιθανότητας των μεταβλητών (q, p) επίσης δεν μεταβάλλεται, και άρα η αλυσίδα Markov δεν μετακινείται σε νέα σημεία. Όλη η μετακίνηση και ανανέωση, δηλαδή μεταφορά σε νέα σημεία (q, p) με διαφορετική πυκνότητα πιθανότητας, πραγματοποιείται κατά το πρώτο στάδιο του αλγορίθμου Hamiltonian MCMC, όπου οι μεταβλητές p αντικαθίστανται με νέες. Η συγκεκριμένη ανανέωση της ορμής έχει την ιδιότητα πως μεταβάλλει την κατανομή πιθανότητας κατά μια σημαντική ποσότητα, γεγονός που λειτουργεί προς όφελος της αποτελεσματικότητας του Hamiltonian MCMC, δεδομένου ότι εξερευνάται ταχύτερα το τυπικό σύνολο της κατανομής στόχου.

Εστιάζοντας αποκλειστικά στην συμπεριφορά των μεταβλητών θέσης q , οι οποίες εξάλλου συγκεντρώνουν το ενδιαφέρον του προβλήματος, η Χαμιλτονιακή Δυναμική χαρακτηρίζεται από την ικανότητά της να γεννά νέες τιμές των q (οι οποίες συνεπάγονται και νέες τιμές της δυναμικής ενέργειας $U(q)$) με αξιοσημείωτα διαφορετική πυκνότητα πιθανότητας συγκριτικά με την αμέσως προηγούμενη κατάσταση. Παρόλα αυτά, η ανανέωση των μεταβλητών ορμής κατέχει κρίσιμο ρόλο προκειμένου η αλυσίδα Markov να καταλήξει επιτυχώς στην επικαιροποιημένη κατανομή. Αυτό συμβαίνει καθώς, εάν δεν πραγματοποιούνταν νέα δειγματοληψία της ορμής κατά το πρώτο στάδιο κάθε επανάληψης, η συνάρτηση Hamiltonian $H(q, p) = U(q) + K(p)$ θα παρέμενε σχεδόν σταθερή και, δεδομένου ότι η δυναμική και η κινητική ενέργεια είναι μη αρνητικές, ο όρος της δυναμικής ενέργειας $U(q)$ δεν θα μπορούσε ποτέ να ξεπεράσει την αρχική τιμή $H(q_0, p_0)$ που είχε τεθεί κατά την εκκίνηση του αλγορίθμου. Η παραπάνω ιδιαιτερότητα της Hamiltonian MCMC είναι κατά σημαντικό ποσοστό υπεύθυνη για την ιδιαίτερα αποτελεσματική συμπεριφορά της συγκριτικά με άλλους αλγορίθμους MCMC.

3.3. Εργοδικότητα της Hamiltonian MCMC

Στις περισσότερες περιπτώσεις, ο αλγόριθμος Hamiltonian MCMC είναι εργοδικός. Αυτό σημαίνει ότι δεν παγιδεύεται σε κάποια συγκεκριμένα περιοχή του παραμετρικού χώρου, συνεπώς είναι σίγουρο ότι ασυμπτωματικά θα συγκλίνει στην (μοναδική και αμετάβλητη) επικαιροποιημένη κατανομή. Παρόλα αυτά, η εργοδικότητα ενδέχεται να μην ισχύει εάν εμφανιστεί μια συγκεκριμένη συνθήκη: τα L βήματα Leapfrog μήκους ϵ το καθένα που εκτελούνται σε μια τροχιά να χαρακτηρίζονται από μια περιοδικότητα εναρμονισμένη με κάποια συνάρτηση της εκάστοτε επικαιροποιημένη κατανομής (Neal, 2011).

Για παράδειγμα, εάν οι λύσεις (επικαιροποιημένες τιμές δηλαδή) μιας εφαρμογής Μπεϋζιανής επικαιροποίησης είναι περιοδικές με περίοδο 2π , τότε τροχιές που προσομοιώνονται στα πλαίσια της Hamiltonian MCMC ενδέχεται να επιστρέφουν στο αρχικό τους σημείο, να μην πραγματοποιούν εξερεύνηση του παραμετρικού χώρου δηλαδή, εάν το μήκος τροχιάς $L*\epsilon$ προσεγγίζει την τιμή 2π . Σε αυτήν την περίπτωση, συνεπώς, η Hamiltonian MCMC δεν θα είναι εργοδική, και το γεγονός αυτό θα οφείλεται σε άστοχη επιλογή των παραμέτρων L και ϵ .

3.4. Η βαθμονόμηση των παραμέτρων ϵ και L στην Hamiltonian MCMC

Στην προηγούμενη ενότητα διαφαίνεται ένας από τους λόγους για τους οποίους η ακριβής βαθμονόμηση των παραμέτρων ϵ και L στην Hamiltonian MCMC κατέχει μεγάλη βαρύτητα. Δεν είναι όμως ο μοναδικός. Η επιλογή του βήματος Leapfrog ϵ και του αριθμού βημάτων L διαδραματίζουν κρίσιμο ρόλο όσον αφορά τη γενικότερη συμπεριφορά και αποτελεσματικότητα του αλγορίθμου.

Πέραν αυτού, είναι γεγονός ότι οι περισσότερες μέθοδοι MCMC περιλαμβάνουν παραμέτρους που απαιτούν βαθμονόμηση, εντούτοις η βαθμονόμηση της Hamiltonian MCMC αποδεικνύεται πιο απαιτητική από αυτήν μιας απλής μεθόδου Metropolis-Hastings. Στην συνέχεια παρουσιάζονται με περισσότερη λεπτομέρεια τα ιδιαίτερα χαρακτηριστικά της επιλογής των ϵ και L . Πρέπει να σημειωθεί, σε αυτό το σημείο, πως οι βέλτιστες επιλογές των ϵ και L δεν είναι δεδομένες, αλλά μεταβάλλονται ανάλογα με τα χαρακτηριστικά και τις απαιτήσεις της εκάστοτε εφαρμογής Μπεϋζιανής επικαιροποίησης.

3.4.1. Βαθμονόμηση βήματος ϵ

Το βήμα ϵ της μεθόδου διακριτοποίησης Leapfrog είναι κρίσιμο στην Hamiltonian MCMC. Ένα υπερβολικά μεγάλο ϵ οδηγεί σε πολύ μικρές πιθανότητες αποδοχής για τις καταστάσεις που

προτείνονται μετά από την προσομοίωση των τροχιών. Από την άλλη πλευρά, εάν το ϵ είναι πολύ μικρό, θα υπάρξει σπατάλη υπολογιστικής ισχύος και χρόνου κατά βαθμό ανάλογο με την απόκλιση του ϵ από την ιδεατή τιμή του (Neal, 2011). Ακόμα χειρότερα, μπορεί να εκφυλιστεί η Hamiltonian MCMC σε τυχαίο περίπατο Metropolis, δηλαδή ο αλγόριθμος να εξερευνήσει τον παραμετρικό χώρο με πολύ αργό ρυθμό, εάν το μήκος τροχιάς $\epsilon * L$ είναι αρκετά μικρό.

Σε κάθε περίπτωση, το βήμα ϵ πρέπει να διατηρείται κάτω από ένα συγκεκριμένο όριο, το οποίο ονομάζεται «όριο σταθερότητας» της Χαμιλτονιακής Δυναμικής. Η συγκεκριμένη τιμή δεν είναι σταθερή, αντιθέτως εξαρτάται από τα χαρακτηριστικά της κατανομής στόχου της εκάστοτε εφαρμογής. Για παράδειγμα, σε απλές περιπτώσεις με Κανονική κατανομή στόχου, το όριο σταθερότητας προκύπτει ίσο με την τετραγωνική ρίζα της μικρότερης τιμής του πίνακα συνδιασπορών των μεταβλητών q (Neal, 2011). Τυχόν χρήση βήματος ϵ μεγαλύτερου από το όριο σταθερότητας, καθιστά την συνάρτηση Hamiltonian ασταθής και εκθετικά αυξανόμενη με τον αριθμό βημάτων L , με αποτέλεσμα η πιθανότητα αποδοχής των νέων καταστάσεων να γίνεται απειροστά μικρή.

3.4.2. Βαθμονόμηση αριθμού βημάτων L

Η επιλογή κατάλληλου L είναι επίσης σημαίνουσα για την Hamiltonian MCMC. Για παράδειγμα, η δειγματοληψία Monte Carlo είναι πολύ απαιτητική σε αρκετές κατανομές, οι οποίες χαρακτηρίζονται από πολλούς περιορισμούς (constraints) σε ορισμένες διευθύνσεις του πεδίου τιμών τους, και καθόλου περιορισμούς σε άλλες. Κατά την εξερεύνηση των συγκεκριμένων κατανομών, η κατασκευή τροχιών με μεγάλο αριθμό βημάτων είναι αναγκαία προκειμένου να προσεγγιστούν σημεία μακριά από το σημείο εκκίνησης. Η συνηθισμένη πρακτική στις περισσότερες εφαρμογές Hamiltonian MCMC υπαγορεύει την χρήση $L = 100$, και εάν αυτή δεν επαρκεί, τότε αύξηση του L σε 1000.

Στην παραδοσιακή βιβλιογραφία δεν υπάρχει ευρέως διαδεδομένη μεθοδολογία για την εύρεση του βέλτιστου ϵ , όπως αντίστοιχα και του L . Προτείνονται, εντούτοις, διάφορες πρακτικές, όπως η τυχαία επιλογή του ϵ από μία προκαθορισμένη κατανομή που βρίσκεται εντός του εύρους σταθερότητας και προς το άνω όριο αυτού και η εκτέλεση του αλγόριθμου ένα πλήθος φορών δοκιμαστικά με διάφορες τιμές των ϵ και L μέχρι να βρεθούν οι βέλτιστες. Καθίσταται σαφές, όμως, πως οι συγκεκριμένες πρακτικές δεν είναι απολύτως ικανοποιητικές. Σε αυτό το πλαίσιο, η ανάπτυξη μιας βέλτιστης μεθόδου Hamiltonian MCMC η οποία κατά την λειτουργία της θα εντοπίζει αυτόματα και θα χρησιμοποιεί τις βέλτιστες τιμές των ϵ και L για την εκάστοτε εφαρμογή, κρίνεται απαραίτητη.

4. Βελτιστοποιημένος αλγόριθμος Hamiltonian MCMC

Όπως διατυπώθηκε στο προηγούμενο κεφάλαιο, η κυριότερη αδυναμία της κλασικής μεθόδου Hamiltonian MCMC έγκειται στο γεγονός ότι εξαρτάται πλήρως από την αυθαίρετη επιλογή των παραμέτρων ϵ και L . Μια τυχόν άστοχη επιλογή αυτών είναι δεδομένο πως θα επιδράσει αρνητικά στην απόδοση του αλγόριθμου. Για παράδειγμα, υποεκτίμηση του L πιθανώς θα προκαλέσει ανεπιθύμητη συμπεριφορά τυχαίου περιπάτου, ενώ μεγάλη υπερεκτίμησή του σπαταλά άσκοπα υπολογιστικό χρόνο. Προκειμένου να αποφευχθούν οι επίπονες δοκιμαστικές εκτελέσεις και να βελτιστοποιηθεί η διαδικασία εφαρμογής της μεθόδου σε πρακτικά προβλήματα, αναπτύχθηκαν στη βιβλιογραφία αρκετές προτάσεις. Ανάμεσα σε αυτές, μία η οποία αποδείχθηκε βάσει αποτελεσμάτων ικανοποιητική και εφαρμόζεται σε αυτήν την εργασία, είναι ο αλγόριθμος No-U-Turn Sampler (NUTS) των Hoffman & Gelman (2014).

Η μέθοδος NUTS επιτυγχάνει να εξαλείψει την ανάγκη της παραδοσιακής βαθμονόμησης του αριθμού βημάτων L της Hamiltonian MCMC. Πιο συγκεκριμένα, χρησιμοποιεί έναν επαναληπτικό αλγόριθμο προκειμένου να κατασκευάσει σταδιακά ένα σύνολο υποψήφιων σημείων τα οποία καλύπτουν μια ευρεία περιοχή της κατανομής στόχου, με την μορφή ενός δυαδικού δέντρου. Η διαδικασία σταματά αυτόματα όταν ανιχνευθεί πως η τροχιά των σημείων πραγματοποίησε αναστροφή και ξαναμπήκε σε περιοχές που είχαν ήδη καλυφθεί. Εμπειρικά και μέσω της εφαρμογής του NUTS σε πλήθος προβλημάτων, έχει αποδειχθεί πως αποδίδει τουλάχιστον το ίδιο αποτελεσματικά με μία καλά βαθμονομημένη κλασική Hamiltonian MCMC, χωρίς όμως να απαιτεί τις κοστοβόρες δοκιμαστικές εκτελέσεις της δεύτερης.

Ο αλγόριθμος NUTS συμπληρώνεται και από ένα επιπλέον εργαλείο, το οποίο προτάθηκε από τον Nesterov το 2009, που έχει την ικανότητα να βαθμονομεί αυτόματα την παράμετρο βήματος ϵ . Το συγκεκριμένο εργαλείο λειτουργεί εντοπίζοντας το κατάλληλο ϵ μέσω μιας διαδικασίας dual averaging, προσαρμοσμένης στην εκάστοτε εφαρμογή και στο επιθυμητό ποσοστό αποδοχής του αλγόριθμου. Κατ' αυτόν τον τρόπο, αναιρείται πλήρως η ανάγκη για βαθμονόμηση της Hamiltonian HMC και καθίσταται δυνατή η αυτόματη εφαρμογή της σε ένα μεγάλο πλήθος προβλημάτων, συμπεριλαμβανομένων φυσικά και των προβλημάτων πολλών διαστάσεων. Ακολουθώς στην παρούσα ενότητα παρουσιάζεται αναλυτικά η μέθοδος NUTS και συνοπτικά το εργαλείο dual averaging, ενώ ο κώδικας σε γλώσσα Matlab που αναπτύχθηκε με βάση τις συγκεκριμένες αρχές παρατίθεται στο πρώτο παράρτημα της εργασίας.

4.1. Ο αλγόριθμος NUTS

4.1.1. Δομή του NUTS

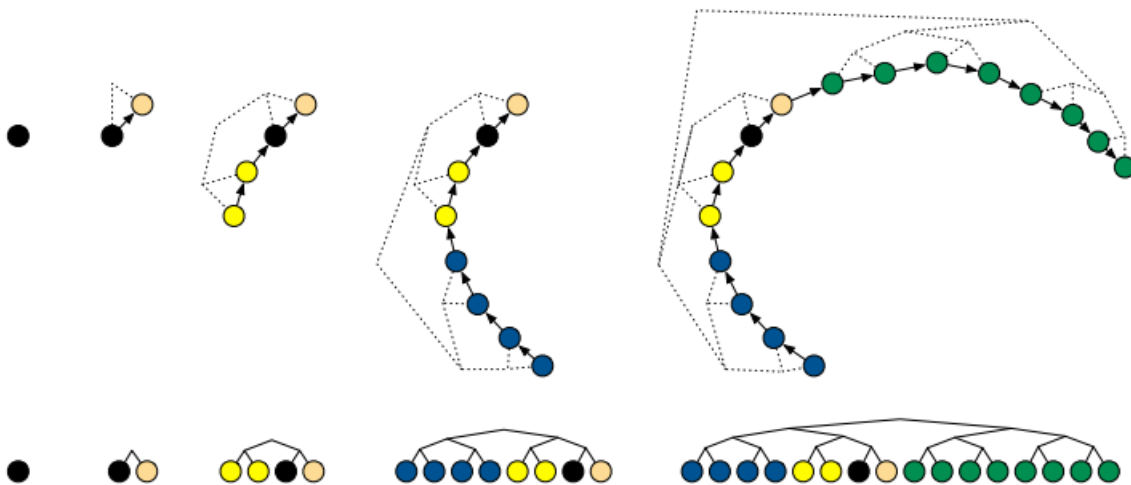
Ο αλγόριθμος NUTS αρχικά εισάγει μία νέα βοηθητική μεταβλητή u , η οποία ονομάζεται *slice variable* και εξάγεται από την ακόλουθη δεσμευμένη κατανομή:

$$P(u / q, p) = \text{Uniform}(u, [0, \exp(U(q) - K(p))]) \quad (4.1)$$

Η δειγματοληψία της u από την παραπάνω κατανομή δεν είναι απολύτως απαραίτητη για την λειτουργία της μεθόδου, όμως συνεισφέρει σημαντικά στην καλύτερη ανάπτυξη και εφαρμογή της. Επιπλέον, κατά τους Hoffman & Gelman, ο ανάλογος αλγόριθμος NUTS ο οποίος παραλείπει την u αποδείχθηκε εμπειρικά λιγότερο αποτελεσματικός από αυτόν που την ενσωματώνει.

Έχοντας πραγματοποιήσει μία νέα ανεξάρτητη δειγματοληψία της u σε κάθε επανάληψη, ο αλγόριθμος στην συνέχεια εφαρμόζει την μέθοδο Leapfrog προκειμένου να χαράξει δοκιμαστικά μια τροχιά τόσο μπροστά όσο και πίσω (με ίσες πιθανότητες για την επιλογή κάθε κατεύθυνσης) σε μη πραγματικό (fictitious) χρόνο. Ειδικότερα, αρχικά πηγαίνει 1 βήμα μπροστά ή πίσω, έπειτα 2, έπειτα 4 και ούτω καθεξής διπλασιάζοντας συνεχώς τον αριθμό βημάτων ϵ που εκτελούνται. Σημειώνεται πως η επιλογή της νέας κατεύθυνσης σε κάθε στάδιο είναι ανεξάρτητη από τις προηγούμενες, ενώ οι όροι «μπροστά» και «πίσω» είναι χρονικοί και αναφέρονται στο πρόσημο του βήματος ϵ . Αν η επιλαχούσα κατεύθυνση είναι «μπροστά» τότε προσομοιώνεται η Χαμιλτονιακή Δυναμική με βήμα $+\epsilon$, ενώ αν είναι «πίσω» τότε εισάγεται το αντίθετο βήμα $-\epsilon$.

Η παραπάνω διαδικασία, η οποία προγραμματιστικά περατώνεται μέσω της συνάρτησης `Build_Tree` που παρουσιάζεται στην συνέχεια, σταδιακά κατασκευάζει ένα ισορροπημένο δυαδικό δέντρο του οποίου τα «φύλλα» αντιπροσωπεύουν υποψήφιες καταστάσεις (q, p) της Hamiltonian MCMC, όπως φαίνεται στο ακόλουθο διάγραμμα.



Εικόνα 7: Παράδειγμα δυαδικού δέντρου, στο οποίο σε κάθε νέο στάδιο διπλασιάζονται τα φύλλα και η επιλογή κατεύθυνσης προς τα μπροστά ή πίσω χρονικά είναι τυχαία (κάθε στάδιο αναπαρίσταται με διαφορετικό χρώμα) (Hoffman & Gelman, 2014).

Η διαδικασία διπλασιασμού σταματά συνήθως επειδή ανιχνεύεται πως η τροχιά από το ένα μέχρι το άλλο άκρο του συνολικού δέντρου αρχίζει να συστρέφεται προς τον εαυτό της, δηλαδή οι υποτροχιές του πραγματοποιούν αναστροφή (U-turn). Όταν συμβεί αυτό, ο αλγόριθμος NUTS παύει την προσομοίωση και λαμβάνει δείγμα καταστάσεων (q,p) από αυτές όλων των πιθανών δυαδικών δέντρων αντίστοιχου μεγέθους που έχουν προσδιοριστεί μέχρι εκείνο το σημείο, προσέχοντας ώστε να μην διαταράξει τις ιδιότητες και την ευστάθεια της αλυσίδας Markov.

4.1.2. Κριτήρια παύσης του NUTS

Αν θεωρηθεί πως το σύνολο όλων των πιθανών σημείων-καταστάσεων (q,p) ενός δυαδικού δέντρου συμβολίζεται ως B , καθίσταται σαφές πως αυτό δεν μπορεί να συνεχίζει να διευρύνεται επ' αόριστον. Αντιθέτως, η ανάπτυξή του οφείλει να παυτεί σε δύο περιπτώσεις προκειμένου να διασφαλιστεί η σωστή λειτουργία του αλγόριθμου NUTS, σύμφωνα με τους Hoffman & Gelman.

Η πρώτη και βασικότερη συνθήκη, όπως αναφέρθηκε και στην προηγούμενη ενότητα, επιβάλλει την παύση όταν η προσομοιωμένη τροχιά πραγματοποιήσει αναστροφή και αρχίσει να διέρχεται από περιοχές τις οποίες έχει ήδη εξερευνήσει. Μια τέτοια συμπεριφορά είναι ανεπιθύμητη για προφανείς λόγους, καθώς ξοδεύεται υπολογιστικός χρόνος χωρίς να παράγεται κάποιο εποικοδομητικό αποτέλεσμα. Προγραμματιστικά, η πρώτη συνθήκη κωδικοποιείται με τον ακόλουθο τρόπο.

Έστω πως το δυαδικό δέντρο βρίσκεται στο στάδιο ανάπτυξης j . Αυτό σημαίνει πως υπάρχουν $2^j - 1$ πιθανά δυαδικά δέντρα τα οποία είναι μη μηδενικά (τετριμμένη περίπτωση). Ο αλγόριθμος εξετάζει ένα προς ένα όλα αυτά τα πιθανά δέντρα και διακόπτει την λειτουργία του μόλις εντοπίσει ένα για το οποίο, αν (q^-, p^-) και (q^+, p^+) είναι τα ακραία σημεία-καταστάσεις του (ή φύλλα όπως επίσης ονομάζονται), ισχύει πως:

$$(q^+ - q^-) \cdot p^- < 0 \text{ ή } (q^+ - q^-) \cdot p^+ < 0 \quad (4.2)$$

Εάν επαληθεύονται οι ανωτέρω εξισώσεις, αυτό σημαίνει πως ο αλγόριθμος, προχωρώντας απειροστά είτε μπροστά είτε πίσω, θα μειώσει την απόσταση μεταξύ των μεταβλητών θέσης (και ενδιαφέροντος) q^- και q^+ . Συνεπώς, το δυαδικό δέντρο έχει συστραφεί προς τον εαυτό του και οι υποτροχιές του έχουν πραγματοποιήσει αναστροφή. Η συγκεκριμένη υπολογιστική διαδικασία απαιτεί την εξέταση $2^{j-2} - 2$ εσωτερικών γινομένων, αριθμός ο οποίος είναι ίδιας τάξης μεγέθους με τα $2^j - 1$ βήματα Leapfrog που πρέπει να γίνουν για να προσδιοριστεί η ακριβής τροχιά κάθε πιθανού δυαδικού δέντρου. Ακόμα σημαντικότερο όμως, όπως

επισημαίνουν οι Hoffman & Gelman, είναι το γεγονός πως, σχεδόν σε όλες τις περιπτώσεις, το υπολογιστικό κόστος πραγματοποίησης αυτών των εσωτερικών γινομένων είναι αμελητέο συγκριτικά με αυτό του υπολογισμού των παραγώγων που πραγματοποιούνται στα πλαίσια της μεθόδου Leapfrog.

Η δεύτερη συνθήκη παύσης του NUTS κωδικοποιεί ένα διαφορετικό προβληματισμό, ο οποίος συνίσταται στο γεγονός πως αν το καθολικό σφάλμα στην προσομοίωση κατασκευής δυαδικού δέντρου στα πλαίσια μιας εφαρμογής Μπεϋζιανής επικαιροποίησης γίνει υπερβολικά μεγάλο, τότε οποιεσδήποτε νέες καταστάσεις αυτή ανακαλύπτει και προσθέτει στο σύνολο B θα χαρακτηρίζονται από εξαιρετικά μικρή πιθανότητα. Κατά συνέπεια, η συνεισφορά τους στην επικαιροποιημένη κατανομή θα είναι αμελητέα, και άρα είναι προτιμότερο να παραλειφθούν.

Το παραπάνω κριτήριο εκφράζεται προγραμματιστικά με τον ακόλουθο τρόπο από τους Hoffman & Gelman: Η διαδικασία διπλασιασμού παύει όταν ανιχνευθεί ένα φύλλο του δέντρου του οποίου η κατάσταση (q,p) σε συνδυασμό με την u ικανοποιούν την συνθήκη:

$$U(q) - K(p) - \log(u) < \Delta_{max} \quad (4.3)$$

Για κάποιο μη αρνητικό Δ_{max} . Προτείνεται ως Δ_{max} να τίθεται μια μεγάλη τιμή, π.χ. 1000, προκειμένου να μην παύεται ο αλγόριθμος εκτός και αν η προσομοίωση δεν έχει υπερβολικά μεγάλο καθολικό σφάλμα. Στην συντριπτική πλειονότητα των περιπτώσεων, συνεπώς, ο NUTS σταματά λόγω του πρώτου κριτηρίου, το οποίο είναι και το βασικότερο.

4.1.3. Συνάρτηση Build_Tree

Η συνάρτηση Build_Tree, εκτός της κατασκευής του δυαδικού δέντρου, είναι επίσης υπεύθυνη και για την επιλογή των κατάλληλων καταστάσεων (q,p) μεταξύ όλων όσων περιλαμβάνονται στο σύνολο B , προκειμένου να πραγματοποιηθεί ανάμεσα σε αυτές τυχαία δειγματοληψία και να οριστεί η επόμενη κατάσταση της αλυσίδας Markov. Αν το συγκεκριμένο σύνολο ονομαστεί C , όπου προφανώς $C \subseteq B$, οι Hoffman & Gelman σημειώνουν ότι οι ακόλουθες 4 συνθήκες πρέπει να ισχύουν για αυτό (με δεδομένες μια κατάσταση (q,p) που ανήκει στο C , μια διαφορετική κατάσταση (q^*,p^*) και τις παραμέτρους u και ε):

- I. Όλα τα στοιχεία του C πρέπει να επιλέγονται με τρόπο που εξασφαλίζει την διατήρηση του όγκου (περισσότερες λεπτομέρειες αναφέρονται στην αντίστοιχη ιδιότητα της Χαμιλτονιακής Δυναμικής στην παράγραφο 3.1.2.).
- II. $P[(q,p) \in C / q,p,u,\varepsilon] = 1$ (4.4)
- III. $P[u \leq \exp(U(q^*) - K(p)) / (q^*,p^*) \in C] = 1$ (4.5)

IV. Αν $(q, p) \in C$ και $(q^*, p^*) \in C$ τότε για οποιοδήποτε B ισχύει πως $P[B, C / q, p, u, \varepsilon] = P[B, C / q^*, p^*, u, \varepsilon]$ (4.6).

Η συνάρτηση `Build_Tree` επιτυγχάνει να προσδιορίσει το C ενώ ταυτόχρονα προσομοιώνει το B . Αφού πρώτα πραγματοποιήσει τυχαία δειγματοληψία της αρχικής ορμής p και της u από τις αντίστοιχες μεταβλητές τους, εφαρμόζει μια επαναληπτική διαδικασία που προσομοιάζει μία έρευνα βάθους (*depth-first search*), η οποία αναιρεί την ανάγκη να αποθηκεύει πλήρως το δέντρο που προσομοιώνεται από τη διαδικασία διπλασιασμού. Τα δεδομένα εισόδου στην `Build_Tree` σε κάθε επανάληψη είναι η αρχική κατάσταση (q_0, p_0) , η μεταβλητή u , η μεταβλητή κατεύθυνσης $v \in \{-1, 1\}$, το μέγιστο μέγεθος (ή βάθος) δέντρου j και η παράμετρος βήματος ε . Έπειτα η συνάρτηση προσομοιώνει 2^j βήματα `Leapfrog` μεγέθους $v \cdot \varepsilon$ το καθένα (δηλαδή «μπροστά» στο χρόνο αν $v=1$ και «πίσω» στο χρόνο αν $v=-1$, όπου το v επιλέγεται τυχαία από ομοιόμορφη κατανομή μεταξύ των 2 τιμών πριν από κάθε επανάληψη). Η παραπάνω διαδικασία επαναλαμβάνεται από τον αλγόριθμο σε κάθε επανάληψή του για αυξανόμενες ακέραιες τιμές του j , ξεκινώντας από το 1 και μέχρι να ικανοποιηθεί ένα κριτήριο παύσης. Σε αυτό το σημείο, διακρίνονται δύο περιπτώσεις αναφορικά με τον τρόπο που ικανοποιούνται τα κριτήρια παύσης:

- Η διαδικασία διπλασιασμού παύεται επειδή είτε η εξίσωση (4.2) είτε η εξίσωση (4.3) ικανοποιήθηκαν από μια κατάσταση ή ένα υποδέντρο που προστέθηκε κατά τον πιο πρόσφατο διπλασιασμό. Σε αυτήν την περίπτωση πρέπει να αποκλειστούν από το σύνολο B όλες οι καταστάσεις (q, p) που προσομοιώθηκαν κατά τον τελευταίο διπλασιασμό, καθώς αν ξεκινήσει εκ νέου μια διαδικασία διπλασιασμού από οποιοδήποτε από αυτές θα οδηγήσει σε ενεργοποίηση κριτηρίου παύσης προτού το πλήρες δέντρο που αντιστοιχεί στο B κατασκευαστεί.
- Η διαδικασία διπλασιασμού παύεται επειδή η εξίσωση (4.2) ικανοποιείται από τις δύο ακραίες καταστάσεις (q^+, p^+) και (q^-, p^-) του B . Σε αυτήν την περίπτωση κανένα κριτήριο παύσης δεν ενεργοποιήθηκε παρά μόνο αφότου το σύνολο B διαμορφώθηκε πλήρως, και συνεπώς όλα τα στοιχεία αυτού είναι έγκυρα και μπορούν να συμπεριληφθούν στο C .

Αν ονομαστεί C^* το κάθε υποσύνολο των έγκυρων καταστάσεων το οποίο αντιστοιχεί σε μία επανάληψη του αλγόριθμου `Build_Tree` (δηλαδή για ένα συγκεκριμένο μέγεθος δέντρου j), τότε ισχύει πως η έξοδος (output) του `Build_Tree` σε κάθε επανάληψή του περιλαμβάνει: τις ακραίες χρονικά καταστάσεις (q^+, p^+) και (q^-, p^-) ανάμεσα στις 2^j που προσομοιώθηκαν, το σύνολο των έγκυρων καταστάσεων C^* για τις οποίες ισχύει πως $\exp[U(q) - K(p)] > u$ και μια μεταβλητή κατάστασης s . Αν επιστραφεί τιμή $s=0$ αυτό σημαίνει ότι ενεργοποιήθηκε η πρώτη περίπτωση επαλήθευσης των κριτηρίων παύσης που αναφέρθηκε προηγουμένως, και συνεπώς πρέπει όλα τα στοιχεία του C^* να αποκλειστούν.

Σε ανώτερο προγραμματιστικό επίπεδο, ο NUTS καλεί την συνάρτηση `Build_Tree` για αυξανόμενες τιμές του j μέχρις ότου είτε αυτή να επιστρέψει $s=0$ είτε να ισχύσει ο δεύτερος τρόπος ενεργοποίησης των κριτηρίων παύσης (όπου τότε το πιο πρόσφατο C^* λαμβάνεται κανονικά υπόψιν). Έπειτα γίνεται σύνθεση του συνόλου C ως η ένωση όλων των επιμέρους C^* , και τέλος επιλέγεται η επόμενη κατάσταση (q,p) της αλυσίδας Markov τυχαία μεταξύ όλων των στοιχείων του C (στην πράξη συγκρατείται μόνο η παράμετρος ενδιαφέροντος q καθώς η παράμετρος ορμής p θα επιλεγεί εκ νέου με δειγματοληψία από την κατανομή της κατά την επόμενη επανάληψη του NUTS).

Ο βέλτιστος αριθμός βημάτων L , τελικά, ο οποίος ήταν και το έναυσμα για την ανάπτυξη του NUTS, θα ισούται με 2^j .

4.2. Dual averaging για το ϵ

Η τεχνική που αξιοποιείται για την προσαρμοστική βαθμονόμηση (adaptive tuning) του βήματος ϵ στα πλαίσια της βελτιστοποιημένης Hamiltonian MCMC ονομάζεται dual averaging και προτάθηκε από τον Nesterov το 2009 στην δημοσίευσή του «Primal-dual subgradient methods for convex problems». Σε αυτήν την ενότητα πραγματοποιείται μια μικρή αναφορά σε αυτήν, καθώς ο τρόπος που εξάχθηκε δεν αποτελεί άμεσο ενδιαφέρον της παρούσας εργασίας. Ο κώδικας που την εφαρμόζει και παρατίθεται στο πρώτο παράρτημα, λήφθηκε από τη δημοσίευση των Hoffman & Gelman και προσαρμόστηκε έπειτα κατάλληλα.

Πρέπει να σημειωθεί, αρχικά, πως η διαδικασία βαθμονόμησης του ϵ συνίσταται να εκτελείται κατά την περίοδο burn-in της αλυσίδας Markov, πριν δηλαδή αυτή συγκλίνει στην επικαιροποιημένη κατανομή. Στις αριθμητικές εφαρμογές που παρουσιάζονται ακολούθως αποφασίστηκε αυτό το διάστημα να ισούται με το μισό των συνολικών επαναλήψεων του αλγόριθμου NUTS της Hamiltonian MCMC. Η μέθοδος dual averaging που εφαρμόζεται είναι μια υποκατηγορία στοχαστικής βελτιστοποίησης με vanishing adaptation. Στην πράξη λειτουργεί μέσω της ασυμπτωματικής σύγκλισης μιας συνάρτησης που περιλαμβάνει την παράμετρο που βαθμονομείται στην επιθυμητή τιμή παράλληλα με την εκτέλεση της Hamiltonian MCMC κατά την περίοδο burn-in. Οι τιμές των παραμέτρων t_0 , κ και γ της μεθόδου καθορίστηκαν σύμφωνα με τις συστάσεις των Hoffman & Gelman για βέλτιστη συμπεριφορά.

Η μέθοδος dual averaging λειτουργεί με δεδομένο εισόδου το επιθυμητό μέσο ποσοστό αποδοχής Metropolis δ των προτεινόμενων καταστάσεων και μια αρχική τιμή ϵ_0 . Στην παρούσα εργασία αποφασίστηκε το δ να οριστεί στο 0.65, τιμή η οποία σύμφωνα με τους Beskos et al. (2010) είναι ιδανική για τις περισσότερες εφαρμογές Μπεϋζιανών επικαιροποιήσεων. Η αρχική τιμή ϵ_0 προσδιορίζεται με βάση έναν απλό heuristic αλγόριθμο που πρότειναν οι Hoffman & Gelman.

5. Αριθμητικές εφαρμογές

5.1. Μπεϋζιανή επικαιροποίηση 1 τυχαίας μεταβλητής

Η πρώτη αριθμητική εφαρμογή είναι μια απλή περίπτωση επικαιροποίησης μίας παραμέτρου ενός μηχανικού μοντέλου, η οποία θεωρείται πως είναι τυχαία μεταβλητή, με χρήση μίας πραγματικής μέτρησης. Σε αυτό το πλαίσιο, η επικαιροποίηση υλοποιείται με την εύρεση μιας νέας μέσης τιμής για την παράμετρο. Το συγκεκριμένο παράδειγμα αναπτύσσεται για τον έλεγχο της σωστής βασικής λειτουργίας της Hamiltonian MCMC. Ως αρχική (prior) κατανομή της παραμέτρου, η οποία επιλέγεται αυθαίρετα, ορίζεται η ακόλουθη απλή (τυποποιημένη) περίπτωση Γκαουσιανής κατανομής με μηδενική μέση τιμή και τυπική απόκλιση μονάδα:

$$P(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \quad (5.1)$$

Έστω πως η μέτρηση που πραγματοποιείται ισούται με $D = 2$. Γίνεται η παραδοχή πως η μέτρηση προέρχεται από την Γκαουσιανή κατανομή με άγνωστο μέσο όρο (αυτή είναι η προς επικαιροποίηση μεταβλητή) και γνωστή τυπική απόκλιση, έστω $\sigma = 0.5$. Τότε και η συνάρτηση πιθανοφάνειας (likelihood function) θα λαμβάνει την ακόλουθη μορφή:

$$P(D/x) = \text{Normal}(D / x, 0.5) = \frac{1}{0.5\sqrt{2\pi}} \exp\left[-\frac{1}{2}(2 - x)^2/0.5^2\right] \quad (5.2)$$

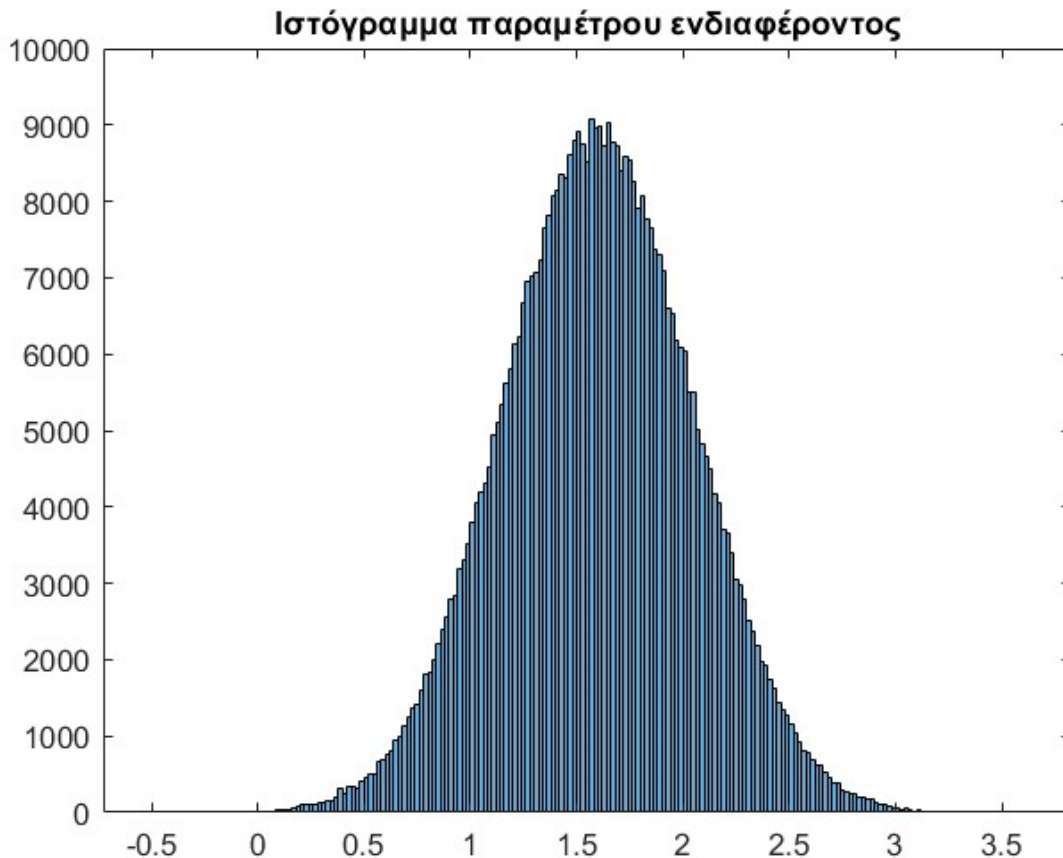
Σημειώνεται πως σε μια πραγματική εφαρμογή είναι διαθέσιμος ένας ικανός αριθμός μετρήσεων, μεταξύ των οποίων συνήθως γίνεται η υπόθεση πως υπάρχει ανεξαρτησία. Τότε και η συνάρτηση πιθανοφάνειας αποτελεί γινόμενο ξεχωριστών κανονικών κατανομών του τύπου που παρατίθεται παραπάνω, ίδιου πλήθους με αυτό των μετρήσεων.

Το συγκεκριμένο παράδειγμα αριθμητικής εφαρμογής επιλέχτηκε καθώς αποτελεί μία ειδική περίπτωση Μπεϋζιανής επικαιροποίησης και μπορεί να επιλυθεί και αναλυτικά. Κατ' αυτόν τον τρόπο, μπορεί να γίνει έλεγχος της ακρίβειας του αλγόριθμου Hamiltonian MCMC που αναπτύχθηκε. Πιο συγκεκριμένα, η επικαιροποιημένη κατανομή $P(x/D)$ όπως προκύπτει αναλυτικά είναι η Γκαουσιανή κατανομή με μέση τιμή 1.6 και τυπική απόκλιση 0.45.

Ο αλγόριθμος έτρεξε για 100.000 επαναλήψεις, εκ των οποίων οι πρώτες 50.000 αξιοποιήθηκαν ως burn-in period (κατά την οποία δεν λήφθηκαν δείγματα) και για να γίνει η προσαρμογή της παραμέτρου βήματος ϵ στην επιθυμητή τιμή μέσω dual averaging. Η προσαρμογή έγινε με στόχο

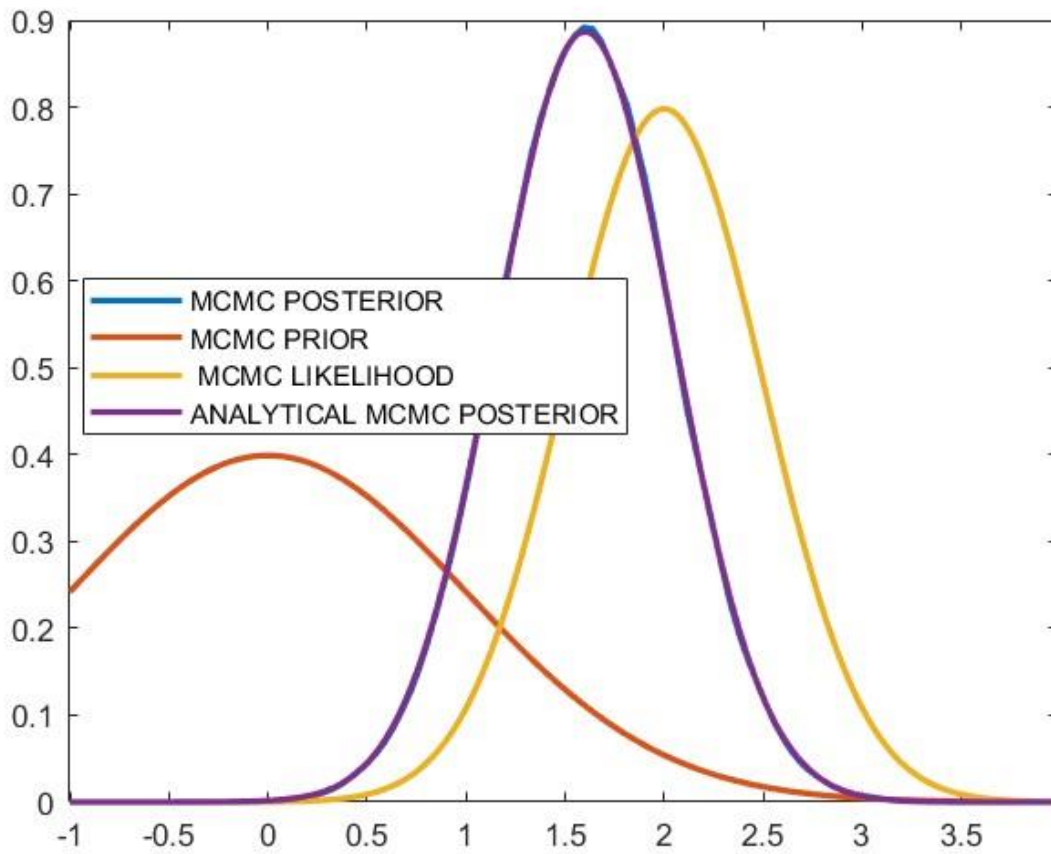
να επιτευχθεί μέσο ποσοστό αποδοχής 0.65 κατά τις τελευταίες 50.000 επαναλήψεις του NUTS όπου λήφθηκαν δείγματα, τιμή η οποία συνίσταται βιβλιογραφικά.

Το παραγόμενο από τον αλγόριθμο ιστόγραμμα παρατίθεται ακολούθως. Η μέση τιμή του δείγματος Monte Carlo που συγκεντρώθηκε ισούται με 1.6004, τιμή ικανοποιητικά κοντά στη θεωρητική (απόκλιση 0.025%). Κατά συνέπεια, φαίνεται ότι ο κώδικας που αναπτύχθηκε λειτουργεί αποτελεσματικά, τουλάχιστον στα πλαίσια ενός απλού παραδείγματος.



Εικόνα 7: Ιστόγραμμα πρώτης αριθμητικής εφαρμογής

Στο ακόλουθο διάγραμμα τέλος, αποτυπώνονται όλες οι συναρτήσεις πυκνότητας πιθανότητας της συγκεκριμένης εφαρμογής. Διαφαίνεται ότι η πραγματική αναλυτική επικαιροποιημένη κατανομή πρακτικά ταυτίζεται με την παραγόμενη από το μοντέλο επικαιροποιημένη κατανομή.

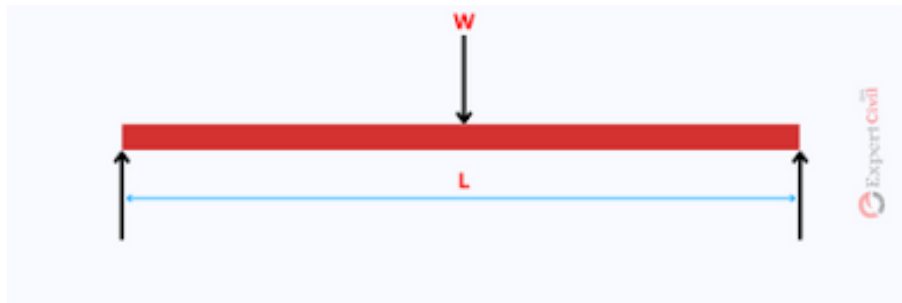


Εικόνα 8: Συναρτήσεις πυκνότητας πιθανότητας 1ης αριθμητικής εφαρμογής

5.2. Εφαρμογή Μπεϋζιανής Επικαιροποίηση 3 παραμέτρων μοντέλου

Στα πλαίσια μίας πιο απτής στο αντικείμενο του πολιτικού μηχανικού εφαρμογής, ο βελτιστοποιημένος κώδικας Hamiltonian MCMC αξιοποιήθηκε για την Μπεϋζιανή επικαιροποίηση των παραμέτρων του πλαστικού νόμου υλικών Von Mises με μη γραμμική σκλήρυνση. Το πρόβλημα διατυπώνεται και διαμορφώνεται ως ακολούθως.

Έστω αμφιέριστη δοκός, μήκους $L = 2$ m και ύψους $H = 0.4$ m, η οποία υπόκειται σε συγκεντρωμένο φορτίο το οποίο ασκείται στο μέσο της. Όσον αφορά τις ελαστικές της ιδιότητες, γίνεται η παραδοχή πως το μέτρο ελαστικότητας E (Young Modulus) ισούται με 3.5 GPa, ενώ ο λόγος Poisson ν είναι ίσος με 0.4.



Εικόνα 9: Αμφιέριστη δοκός η οποία υποβάλλεται σε συγκεντρωμένο φορτίο στο μέσο της (3-point test)

Σύμφωνα με τον υπο διερεύνηση νόμο υλικών Von Mises, η επιφάνεια διαρροής που περιγραφεί το σύνολο των παραμορφωσιακών καταστάσεων για τις οποίες ενεργοποιείται η πλαστικοποίηση του υλικού, χαρακτηρίζεται από την εξής σχέση:

$$\Phi = \sqrt{3J_2} - \sigma_y(\varepsilon_{eq}^{pl}) \quad (5.3)$$

Όπου J_2 είναι η δεύτερη αναλλοίωτη του αποκλίνοντα τανυστή τάσεων, ενώ σ_y είναι η τάση διαρροής που ορίζεται ως συνάρτηση της ισοδύναμης πλαστικής παραμόρφωσης ε_{eq}^{pl} .

Η σκλήρυνση $\sigma_y(\varepsilon_{eq}^{pl})$ είναι η συνάρτηση που καθορίζει την σχέση μεταξύ της τάσης διαρροής και των πλαστικών παραμορφώσεων και στην συγκεκριμένη εφαρμογή λαμβάνεται ως εκθετικού τύπου που εκφράζεται από την εξίσωση:

$$\sigma_y = a + b * [1 - \exp(-c * \varepsilon_{eq}^{pl})] \quad (5.4)$$

Η τελική μορφή της παραπάνω συνάρτησης σκλήρυνσης επηρεάζεται από τις τρεις παραμέτρους a , b και c . Η παράμετρος a χαρακτηρίζει την αρχική τάση διαρροής, ενώ οι άλλες δύο παράμετροι b και c δεν έχουν κάποια φυσική σημασία και χρησιμοποιούνται για να διαμορφώσουν το σχήμα της εκθετικής καμπύλης.

Στόχος του συγκεκριμένου αριθμητικού παραδείγματος Μπεϋζιανής επικαιροποίησης είναι η ενημέρωση των πιθανοτικών κατανομών των παραμέτρων a , b και c , για τις οποίες γίνεται η παραδοχή ότι αποτελούν τυχαίες μεταβλητές.

Για την υλοποίηση της εφαρμογής απαιτήθηκε η προσομοίωση μέσω ενός μοντέλου πεπερασμένων στοιχείων το οποίο, δεχόμενο ως input την τιμή του συγκεντρωμένου φορτίου στο μέσο των δοκών και τις τιμές των 3 παραμέτρων του μη γραμμικού νόμου, υπολογίζει και δίνει ως output την μέγιστη μετατόπιση (response) που παρατηρείται επίσης στο μέσο της δοκού. Το μοντέλο διακριτοποιήθηκε με τετραπλευρικά πεπερασμένα στοιχεία επίπεδης έντασης, των οποίων το πάχος ορίστηκε ως 0.05 m. Αποφασίστηκε να γίνει καταμερισμός του μοντέλου σε 100 πεπερασμένα στοιχεία κατά x και 20 κατά y , με γνώμονα την επίτευξη ενός βιώσιμου υπολογιστικού κόστους, διατηρώντας παράλληλα και την αναλογία των αντίστοιχων διαστάσεων της δοκού. Ο υπολογιστικός κώδικας σε γλώσσα Matlab που εκτελεί την παραπάνω διαδικασία παρατίθεται στο δεύτερο παράρτημα της εργασίας.

Για να καταρτιστεί η συνάρτηση πιθανοφάνειας (likelihood) του παραδείγματος, αποφασίστηκε να παραχθούν 5 συνθετικά ζευγάρια παρατηρήσεων, της μορφής (φορτίο w – μετατόπιση D). Αυτά προσδιορίστηκαν από το μοντέλο για τιμές παραμέτρων 0.1, 0.02 και 30 αντίστοιχα, οι οποίες είναι εμπειρικά εύλογες, και για 5 φορτία, από 50 έως 250 kN ανά 50 kN. Επιπλέον, θεωρήθηκε ότι κατά την λειτουργία του Hamiltonian MCMC, κάθε νέο σετ (a, b, c) που γίνεται δεκτό από τον αλγόριθμο είναι τέτοιο ώστε η τιμή της μετατόπισης y που παράγεται από το μοντέλο για αυτές και για καθένα από τα 5 φορτία κατανέμεται με Κανονική κατανομή γύρω από την παρατηρημένη συνθετική μετατόπιση D που προέκυψε για το ίδιο φορτίο, με τυπική απόκλιση σ ίση με το 3% κάθε παρατήρησης. Έτσι η συνάρτηση likelihood αποτελείται από 5 ξεχωριστούς όρους, οι οποίοι θεωρούνται ανεξάρτητοι μεταξύ τους, και συνεπώς συνδυάζονται με την μορφή γινομένου 5 Γκαουσιανών κατανομών.

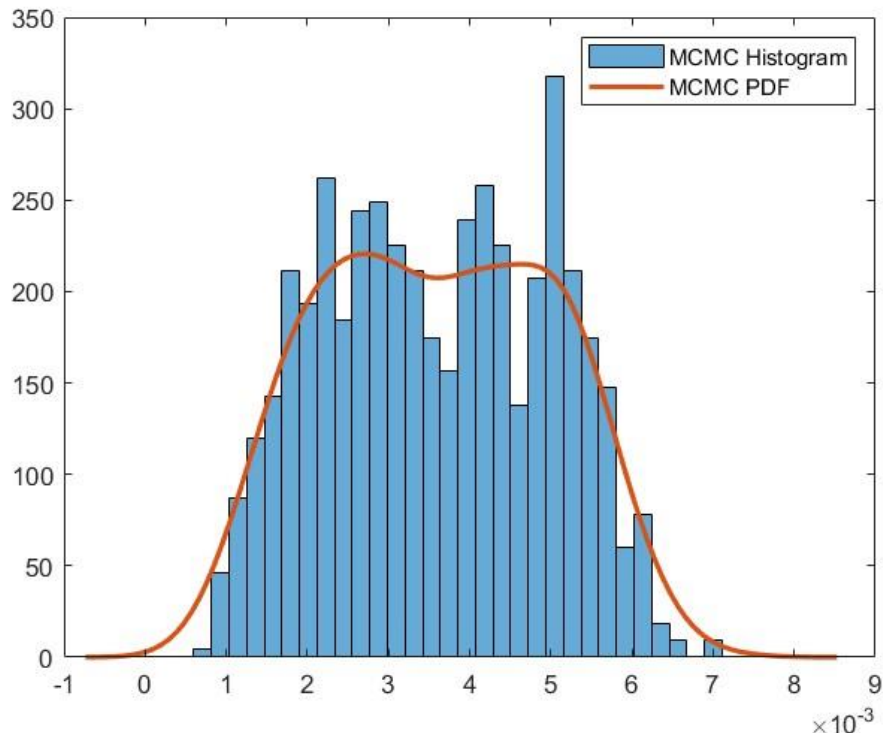
Απαραίτητη προϋπόθεση για την ομαλή εφαρμογή του κώδικα Hamiltonian MCMC υπήρξε και η θέσπιση ορίων για κάθε μία παράμετρο, τα οποία οριοθετούν τον παραμετρικό χώρο και εντός των οποίων κινείται η αλυσίδα Markov του αλγόριθμου. Ειδικότερα, για την a τέθηκε το εύρος [0.00001, 0.1], για την b το εύρος [0.0001, 0.4] και για την c το εύρος [0.001, 100]. Σε αυτό το πλαίσιο, ως prior για κάθε παράμετρο ορίστηκε η ομοιόμορφη κατανομή στα προαναφερθέντα εύρη τιμών, προκειμένου να αποφευχθεί η οποιαδήποτε προκατάληψη.

Δεδομένου ότι σε κάθε επανάληψη του αλγορίθμου NUTS η συνάρτηση δυναμικής ενέργειας, όπως και η συνάρτηση της αριθμητικής παραγωγού της δυναμικής ενέργειας η οποία καλεί την

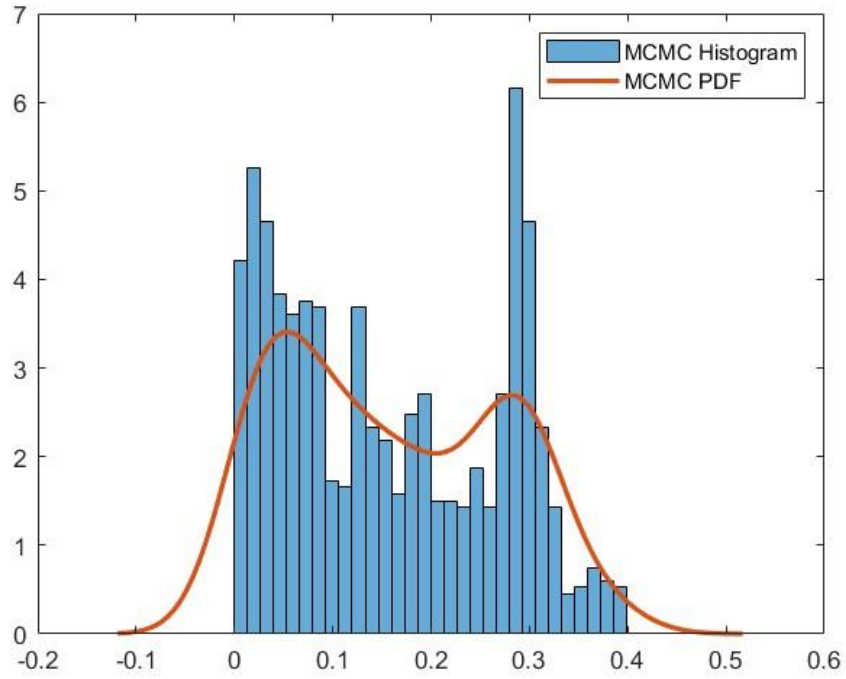
πρώτη, καλούνται αρκετές φορές προκειμένου να κατασκευαστεί το δυαδικό δέντρο, και λαμβάνοντας υπόψιν πως κάθε προσδιορισμός της δυναμικής ενέργειας απαιτεί 5 κλήσεις του μοντέλου πεπερασμένων στοιχείων, καθίσταται σαφές πως η συγκεκριμένη υπολογιστική εφαρμογή έχει σημαντικό υπολογιστικό βάρος. Σε αυτό το πλαίσιο, προκειμένου να είναι χρονικά βιώσιμη αλλά και να έχει τρέξει αρκετά ώστε να συγκλίνει στην επικαιροποιημένη κατανομή, αποφασίστηκε να εκτελεστεί για 1000 δειγματοληπτικές επαναλήψεις NUTS, κατόπιν αρχικών δοκιμαστικών ελέγχων. Επιπλέον, αποφασίστηκε να τεθεί υψηλότερο μέσο ποσοστό αποδοχής, συγκεκριμένα 0.85, προκειμένου το δείγμα να είναι κατά το δυνατόν μεγαλύτερο. Για να επιτευχθεί το παραπάνω μέσο ποσοστό αποδοχής απαιτήθηκε ϵ psilon, σύμφωνα με τον αλγόριθμο dual averaging, ίσο με 0.1. Οι επικαιροποιημένες μέσες τιμές των 3 παραμέτρων είναι οι ακόλουθες:

- Παράμετρος a: **0.0036**
- Παράμετρος b: **0.1586**
- Παράμετρος c: **36.5622**

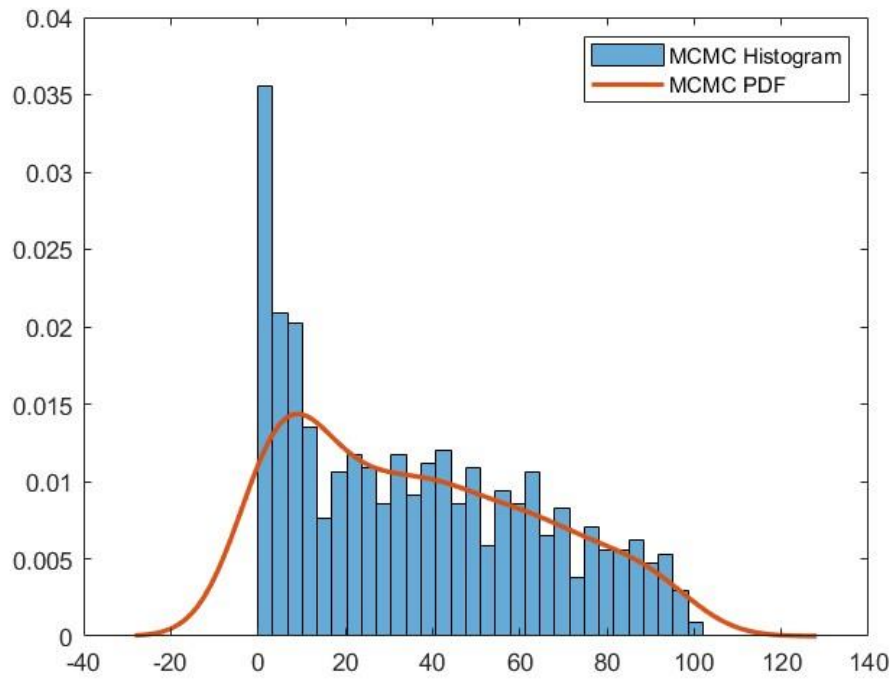
Ακολουθώς παρατίθενται τα ιστογράμματα των παραμέτρων που επικαιροποιήθηκαν και οι επικαιροποιημένες κατανομές τους. Είναι γεγονός ότι και στις 3 περιπτώσεις ο αλγόριθμος Hamiltonian MCMC εξερεύνησε όλο το εύρος του παραμετρικού χώρου και συνέκλινε στις επικρατούσες τιμές, παρά το σχετικά μικρό αριθμό επαναλήψεων, γεγονότα που ενισχύει την αξιοπιστία του.



Εικόνα 10: Ιστόγραμμα παραμέτρου a



Εικόνα 11: Ιστόγραμμα παραμέτρου b



Εικόνα 12: Ιστόγραμμα παραμέτρου c

5.3. Συμπεράσματα

Μέσα από την ενεργοποίηση του βελτιστοποιημένου υβριδικού αλγορίθμου Hamiltonian MCMC σε δύο πραγματικές αριθμητικές εφαρμογές, γίνεται δυνατός ένας ρεαλιστικός έλεγχος της λειτουργίας και της απόδοσής του. Είναι γεγονός πως η μεθοδολογία κατάφερε να ανταποκριθεί ικανοποιητικά και στις δύο περιπτώσεις, αποδεικνύοντας, έστω σε ένα πρώτο στάδιο, τη χρησιμότητά της.

Η πρώτη αριθμητική εφαρμογή, ιδιαίτερα απλή στη φύση της καθώς περιελάμβανε την επικαιροποίηση μίας μόνο παραμέτρου στα πλαίσια ενός παραδείγματος που έχει και αναλυτική λύση, χρησίμευσε κυρίως ως ένα τεστ πως ο αλγόριθμος επιτελεί τις βασικές του λειτουργίες σωστά. Πράγματι, τα στοιχεία της απόδοσής του υπήρξαν εντυπωσιακά. Ο βελτιστοποιημένος Hamiltonian MCMC κατάφερε να εκτελέσει 100.000 επαναλήψεις του NUTS σε πολύ λίγο χρόνο (<2 λεπτά), επιδεικνύοντας έτσι και την σωστή κατασκευή και έγκαιρη παύση των δυαδικών δέντρων, και συνέκλινε στην θεωρητική επικαιροποιημένη κατανομή με ακρίβεια 99.975 %. Σε αυτό το πλαίσιο, ο αλγόριθμος φαίνεται να αποδίδει άρτια, με την προϋπόθεση ότι του παρέχονται οι κατάλληλες συναρτήσεις likelihood και priors.

Το δεύτερο αριθμητικό παράδειγμα αποτέλεσε ένα πιο δύσκολο crash test, βρισκόμενο πιο κοντά στην πραγματικότητα των αληθινών εφαρμογών μηχανικού. Και πάλι, ο βελτιστοποιημένος Hamiltonian MCMC φάνηκε να αποδίδει ικανοποιητικά, εξερευνώντας ολόκληρο τον παραμετρικό χώρο και των τριών παραμέτρων που επικαιροποιήθηκαν, παρά τον συγκριτικά μικρό αριθμό επαναλήψεων που εκτελέστηκαν, ενώ συνέκλινε και στις επικρατούσες τιμές των επικαιροποιημένων παραμέτρων. Η συγκεκριμένη εφαρμογή, όμως, ανέδειξε και δύο ζητήματα σχετικά με την αποτελεσματική εφαρμογή της μεθοδολογίας, τα οποία πρέπει να σημειωθούν.

Πρώτον, στο πλαίσιο σύνθετων εφαρμογών οι οποίες συνεπάγονται πολύπλοκα υπολογιστικά μοντέλα τύπου πεπερασμένων στοιχείων, πεπερασμένων διαφορών κ.λπ., η απαιτούμενη υπολογιστική ισχύς και το υπολογιστικό κόστος εφαρμογής του αλγορίθμου αυξάνονται δραματικά. Αυτό συμβαίνει καθώς τα προαναφερθέντα υπολογιστικά μοντέλα υπεισέρχονται στον υπολογισμό της δυναμικής ενέργειας U της συνάρτησης Hamiltonian, η οποία καλείται επανειλημμένα κατά την κατασκευή των δυαδικών δέντρων του NUTS και επίσης κατά τον προσδιορισμό της αριθμητικής της παραγώγου grad_U . Κατά συνέπεια, ο αριθμός επαναλήψεων του NUTS πρέπει να περιορίζεται σημαντικά προκειμένου ο κώδικας να μπορεί να εκτελείται σε βιώσιμο χρονικό διάστημα, ενώ έμφαση πρέπει να δίνεται και στην «ελάφρυνση» κατά το δυνατόν των συνοδών μοντέλων της εφαρμογής (π.χ. με αραίωση της διακριτοποίησης των πεπερασμένων στοιχείων σε ένα τέτοιου είδους μοντέλου) προκειμένου να μειώνεται το υπολογιστικό κόστος. Χαρακτηριστικά, στην δεύτερη εφαρμογή που παρουσιάστηκε προηγουμένως, η εκτέλεση 1.000 επαναλήψεων του κώδικα Matlab απαίτησε 5 μέρες, και αυτό μετά από μείωση των πεπερασμένων στοιχείων κατά x από 250 σε 100 και κατά

γ από 50 σε 20. Με τις αρχικές ρυθμίσεις η εκτέλεση 1.000 επαναλήψεων θα απαιτούσε περισσότερες από 12.5 ημέρες.

Δεύτερον, αναδεικνύεται η καθοριστική σημασία της επιλογής κατάλληλων συναρτήσεων likelihood και priors. Είναι γεγονός, πως στα περισσότερα πραγματικά προβλήματα Μπεϋζιανών επικαιροποιήσεων δεν είναι δυνατός ο θεωρητικός προσδιορισμός της συνάρτησης likelihood, συνεπώς η επιλογή της πρέπει να γίνει αυθαίρετα. Αποτελεί κοινή πρακτική στη βιβλιογραφία αυτή να διαμορφώνεται μέσω της παραδοχής πως κάθε νέα τιμή των παραμέτρων ενδιαφέροντος που γίνεται αποδεκτή στην αλυσίδα Markov του αλγόριθμου κατανέμεται με Γκαουσιανή Κατανομή γύρω από τις παρατηρήσεις, με γνωστή ή άγνωστη μέση τιμή και τυπική απόκλιση ανάλογα με την περίπτωση, ενώ οι όροι της κατανομής γύρω από κάθε παρατήρηση θεωρούνται ανεξάρτητοι μεταξύ τους. Έτσι η συνάρτηση likelihood λαμβάνει την μορφή γινομένου ίδιου πλήθους όρων με αυτό των παρατηρήσεων. Η παραπάνω μορφή της συνάρτησης πιθανοφάνειας φάνηκε να λειτουργεί αποτελεσματικά στο δεύτερο αριθμητικό παράδειγμα, όμως ενδέχεται σε άλλες εφαρμογές να ταιριάζει καλύτερα μία άλλη μορφή. Εκτός αυτού, όταν δεν υπάρχει καμία προηγούμενη πληροφόρηση για τις παραμέτρους προς επικαιροποίηση, όπως στο δεύτερο αριθμητικό παράδειγμα της παρούσας εργασίας, προκρίνεται η μορφοποίηση των priors ως ομοιόμορφων (uniform) κατανομών σε προκαθορισμένα εύρη, δίνοντας έτσι ίδια βαρύτητα σε κάθε πιθανή τιμή μέσα στον παραμετρικό χώρο. Οι συνθήκες διαφοροποιούνται, όμως, όταν υπάρχει πιο συγκεκριμένη προηγούμενη επικαιροποίηση για τις παραμέτρους ενδιαφέροντος. Σε αυτές τις περιπτώσεις, η χρήση διαφορετικών κατανομών ως priors (π.χ. Κανονικής) ενδέχεται να είναι σωστότερη.

Συνολικά, η βελτιστοποιημένη υβριδική έκδοση της μεθοδολογίας Hamiltonian Markov Chain Monte Carlo για την εκτέλεση Μπεϋζιανής επικαιροποίηση που παρουσιάζεται στην παρούσα εργασία επιδεικνύει μια γενικά άρτια συμπεριφορά στις αριθμητικές εφαρμογές που εξετάστηκε. Χαρακτηρίζεται από σημαντικά συγκριτικά πλεονεκτήματα τα οποία την καθιστούν ελκυστική, με κύρια εξ αυτών την ταχεία εξερεύνηση του παραμετρικού χώρου και τη δυνατότητα επικαιροποίησης πολλών παραμέτρων ταυτόχρονα χωρίς κολλήματα. Παράλληλα, η αυτόματη βαθμονόμηση των βασικών παραμέτρων ϵ και L του κλασικού αλγόριθμου Hamiltonian MCMC επιτρέπει στον χρήστη να ρυθμίζει εύκολα και άμεσα τον κώδικα σύμφωνα με τις ανάγκες της εφαρμογής που αντιμετωπίζει. Προσοχή, όμως, πρέπει να δίνεται στα πλαίσια σύνθετων εφαρμογών που εισάγουν επιπλέον υπολογιστικά μοντέλα, αλλά και στην επιλογή των συναρτήσεων likelihood και των priors, ώστε να εξασφαλίζεται η ομαλή και αποτελεσματική εκτέλεση της μεθόδου.

Βιβλιογραφικές αναφορές

- 1 Beskos, Alexandros, et al. "The Acceptance Probability of the Hybrid Monte Carlo Method in High-Dimensional Problems." AIP Conference Proceedings. Vol. 1281. No. 1. American Institute of Physics, 2010.
- 2 Betancourt, Michael. "A conceptual introduction to Hamiltonian Monte Carlo." *arXiv preprint arXiv:1701.02434* (2017).
- 3 Bolstad, William M. *Understanding computational Bayesian statistics*. Vol. 644. John Wiley & Sons, 2009.
- 4 Hoffman, Matthew D., and Andrew Gelman. "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.* 15.1 (2014): 1593-1623.
- 5 Kalogeris, Y., & Pyrialakos, S. (2022). Stochastic Finite Element Methods & Data-driven models in Engineering Applications. School of Civil Engineering, National Technical University of Athens.
- 6 Neal, Radford M. "MCMC using Hamiltonian dynamics." *Handbook of markov chain monte carlo* 2.11 (2011): 2.
- 7 (Three) 3-Point Bend Test of Concrete and Beam. "(Three) 3-Point Bend Test of Concrete and Beam." Expert Civil, 19 June 2022, expertcivil.com/three-3-point-bend-test/.
- 8 Nesterov, Yurii. "Primal-dual subgradient methods for convex problems." *Mathematical programming* 120.1 (2009): 221-259.
- 9 Straub, Daniel, and Iason Papaioannou. "Bayesian updating with structural reliability methods." *Journal of Engineering Mechanics* 141.3 (2015): 04014134.

Παράρτημα 1

Ακολούθως παρατίθεται ο πλήρης κώδικας σε γλώσσα Matlab της βελτιστοποιημένης μεθόδου Hamiltonian MCMC που ενσωματώνει No-U-Turn Sampler και Dual Averaging. Ο συγκεκριμένος κώδικας είναι προσαρμοσμένος στην πρώτη αριθμητική εφαρμογή και στην αντίστοιχη δυναμική ενέργεια. Η τροποποίηση του όμως για μια διαφορετική εφαρμογή Μπεϋζιανής επικαιροποίησης είναι εύκολη, καθώς απαιτείται μόνο αλλαγή της δυναμικής ενέργειας και προσαρμογή του αριθμού των διαστάσεων D . Ο κώδικας θεμελιώνεται πάνω σε αυτόν που δημοσίευσε ο Hoffman το 2011, συνεπώς τοποθετείται και η κατάλληλη ένδειξη copyright στο τέλος του.

```
function HMC_with_NUTS_and_Dual_Averaging

clear;clc;close all;

%Κύριος αλγόριθμος
D =1; %διαστάσεις του προβλήματος
q_initial = zeros(D,1); %Διάνυσμα στήλη - ενδεικτικό αρχικό διάνυσμα
delta = 0.65; %Πιθανότητα αποδοχής στην οποία στοχεύουμε και για την οποία αναζητούμε
το κατάλληλο  $\epsilon$  μέσω της διαδικασίας dual averaging (προτείνεται το 0.65 στη
βιβλιογραφία)
n_warmup = 50000; %Αριθμός των επαναλήψεων NUTS για τις οποίες θα τρέξει ο αλγόριθμος
dual averaging προσαρμόζοντας το  $\epsilon$  και επίσης θα λειτουργήσει ως burn-in period
% Αξιοποιούνται επίσης και ως burn-in samples (από εκεί και μετά γίνεται sampling
κανονικά)

[q0,epsilon_goal] = dual_averaging(q_initial,delta,n_warmup); %το q0 είναι το αρχικό
q με το οποίο θα τρέξουμε τον NUTS στην αμέσως επόμενη φάση (sampling)
% και το epsilon_goal είναι το  $\epsilon$  με το οποίο θα τρέξουμε τον NUTS
% προκειμένου να έχουμε το ποσοστό αποδοχής (delta) που επιθυμούμε

potential_q_initial = potential(q0);
grad_potential_initial = grad_potential(q0);
M = 1000000; %θα γίνουν 1000000 επαναλήψεις της συνάρτησης NUTS, αλλά δειγματοληψία
μόνο από τις τελευταίες 50.000 κατά τις οποίες η αλυσίδα Markov θα έχει συγκλίνει
q_sampling = zeros(D,M/2);
potential_q_sampling = zeros(D,M/2);
grad_potential_sampling = zeros(D,M/2);
q_sampling(:,1) = q0;
potential_q_sampling(:,1) = potential_q_initial;
grad_potential_sampling(:,1) = grad_potential_initial;
max_tree_depth_sampling = 12; %προτεινόμενη τιμή από Hoffman & Gelman

for j = 1:M/2
    [q_sampling(:,j+1),~,potential_q_sampling(:,j+1), grad_potential_sampling(:,j+1)]
    =...
```

```

NUTS(epsilon_goal,q_sampling(:,j),potential_q_sampling(:,j),grad_potential_sampling(:,j),max_tree_depth_sampling);
end

histogram(q_sampling);
title('Ιστογράμμο παραμέτρον ενδιαφέροντοσ')
disp(mean(q_sampling));
disp(std(q_sampling))

function [q, epsilonbar] = dual_averaging(q_initial, delta, n_warmup)

potentialq= potential(q_initial);
grad_potentialq = grad_potential(q_initial);

max_tree_depth = 12; %μέγιστο βάθοσ δέντρον (μέγιστοσ αριθμόσ φύλλων του δέντρον
δηλαδή βημάτων L: 2^12 - προτεινόμενη τιμή από Hoffman & Gelman)

% επιλογή ενόσ εύλογον αρχικό ε για τη διαδικασία dual averaging μέσω ενόσ απλόσ
heuristic
epsilon = find_reasonable_epsilon(q_initial, grad_potentialq, potentialq);

% Παράμετροι του αλγόριθμο dual averaging όπωσ προτείνονται από τουσ Hoffman &
Gelman
gamma = .05;
t0 = 10;
kappa = 0.75;
mu = log(10 * epsilon);

% Έναρξη αλγόριθμο dual averaging.
epsilonbar = 1;
H_ = 0;

q = q_initial;
for i = 1:n_warmup

    [q, ave_alpha, potentialq, grad_potentialq] = NUTS(epsilon, q, potentialq,
grad_potentialq, max_tree_depth);
    eta = 1 / (i + t0);
    H_ = (1 - eta) * H_ + eta * (delta - ave_alpha);
    epsilon = exp(mu - sqrt(i) / gamma * H_);
    eta = i^-kappa;
    epsilonbar = exp((1 - eta) * log(epsilonbar) + eta * log(epsilon));

end

end

function [q, acceptance_ave, potential_q, grad_potential] = NUTS(epsilon, q_initial,
potential_q_initial, grad_potential_initial, max_tree_depth)

p0 = randn(size(q_initial));

joint = potential_q_initial - kinetic(p0);

```

```

%u ~ uniform([0, exp(joint)]) ισοδύναμο με (log(u) - joint) ~ exponential(1).
logu = joint - exprnd(1);

% Έναρξη κατασκευής δυαδικού δέντρου
qminus = q_initial;
qplus = q_initial;
pminus = p0;
pplus = p0;
gradminus = grad_potential_initial;
gradplus = grad_potential_initial;

depth = 0;
% Αν δεν γίνει δεκτό το συγκεκριμένο update του NUTS συγκρατούνται οι τιμές
% της προηγούμενης επανάληψης
q = q_initial;
grad_potential = grad_potential_initial;
potential_q = potential_q_initial;

n = 1;

% Κύριο loop αλγόριθμου NUTS
s = false;
while ~s

    dir = 2 * (rand() < 0.5) - 1; % Επιλογή κατεύθυνσης: -1=πίσω, 1=μπροστά

    if (dir == -1)
        [qminus, pminus, gradminus, ~, ~, ~, q_, grad_, potential_, n_, s_, a, na] =
...
        build_tree(qminus, pminus, gradminus, logu, dir, depth, epsilon,joint);
    else
        [~, ~, ~, qplus, pplus, gradplus, q_, grad_, potential_, n_, s_, a, na] = ...
        build_tree(qplus, pplus, gradplus, logu, dir, depth, epsilon,joint);
    end

    if (~ s_ && (rand() < n_/n))
        q = q_;
        potential_q = potential_;
        grad_potential = grad_;
    end

    n = n + n_;

    s = s_ || stop_criterion(qminus, qplus, pminus, pplus);

    depth = depth + 1;

    if depth > max_tree_depth
        break
    end

end
acceptance_ave = a / na;

```

```

end

function criterion = stop_criterion(qminus, qplus, pminus, pplus)

thetavec = qplus - qminus;
criterion = (thetavec' * pminus < 0) || (thetavec' * pplus < 0);

end

% Το βασικό recursion.
function [qminus, pminus, gradminus, qplus, pplus, gradplus, q_, grad_, potential_,
n_, s_, a_, na_] = ...
    build_tree(q, p, grad, potential, dir, depth, epsilon, joint0)

if (depth == 0)
    %Βασική περίπτωση: 1 βήμα κατά τη διεύθυνση dir
    [q_, p_, grad_, potential_] = leapfrog(q, p, grad, dir*epsilon);
    joint = potential_ - kinetic(p_);

    n_ = potential < joint;

    s_ = potential - 100 >= joint;

    qminus = q_;
    qplus = q_;
    pminus = p_;
    pplus = p_;
    gradminus = grad_;
    gradplus = grad_;
    % Υπολογισμός acceptance probability.
    a_ = exp(potential_ - kinetic(p_) - joint0);
    if isnan(a_)
        a_ = 0;
    else
        a_ = min(1, a_);
    end
    na_ = 1;
else
    % Recursion
    [qminus, pminus, gradminus, qplus, pplus, gradplus, q_, grad_, potential_, n_,
s_, a_, na_] = ...
        build_tree(q, p, grad, potential, dir, depth-1, epsilon, joint0);

    if ~s_
        if (dir == -1)
            [qminus, pminus, gradminus, ~, ~, ~, q_, grad_, potential_, n_, s_,
a_, na_] = ...
                build_tree(qminus, pminus, gradminus, potential, dir, depth-1,
epsilon, joint0);
            else
                [~, ~, ~, qplus, pplus, gradplus, q_, grad_, potential_, n_, s_,
a_, na_] = ...
                    build_tree(qplus, pplus, gradplus, potential, dir, depth-1, epsilon,
joint0);
        end
    end
end

```



```

end
% Επιλέγουμε από ποιο υποδέντρο θα λάβουμε δείγμα
if (rand() < n__ / (n_ + n__))
    q_ = q__;
    grad_ = grad__;
    potential_ = potential__;
end

n_ = n_ + n__;

s_ = s_ || s__ || stop_criterion(qminus, qplus, pminus, pplus);

a_ = a_ + a__;
na_ = na_ + na__;
end
end
end

function epsilon = find_reasonable_epsilon(q_initial, grad_initial,
potentialq_initial)

epsilon = 1;
p = randn(length(q_initial), 1);

[~, p_, ~, potentialq_new] = leapfrog(q_initial, p, grad_initial, epsilon);
acceptprob = exp(-potentialq_new + potentialq_initial + kinetic(p_) - kinetic(p));

a = 2 * (acceptprob > 0.5) - 1;
% Το ε κινείται σε αυτήν την κατεύθυνση μέχρι η μεταβλητή acceptprob να
% ξεπεράσει το 0.5
while (acceptprob^a > 2^(-a))
    epsilon = epsilon * 2^a;
    [~, p_, ~, potentialq_new] = leapfrog(q_initial, p, grad_initial, epsilon);
    acceptprob = exp(-potentialq_new + potentialq_initial + kinetic(p_) - kinetic(p)
);
end

end

function [qnew, pnew, grad_, potential_] = leapfrog(q, p, grad, epsilon)

pnew = p + 0.5 * epsilon * grad;
qnew = q + epsilon * pnew;
potential_ = potential(qnew);
grad_ = grad_potential(qnew);
pnew = pnew + 0.5 * epsilon * grad_;

end

function U = potential(q) % ο αλγόριθμος NUTS προγραμματιστικά απαιτεί να λαμβάνεται
θετική τιμή ως δεδομένο εισόδου, όχι η τυπική -log

```

```
U = log(1/(sqrt(2*pi))*exp(-q^2/2)/(0.5*sqrt(2*pi))*exp(-0.5*(q-2)^2/0.5^2)); %Τύπος  
δυναμικής ενέργειας της 1ης αριθμητικής εφαρμογής
```

```
end
```

```
function grad_U = grad_potential(q)
```

```
dq = 1e-6;
```

```
q_plus_dq = q + dq;
```

```
U_q = log(1/(sqrt(2*pi))*exp(-q^2/2)/(0.5*sqrt(2*pi))*exp(-0.5*(q-2)^2/0.5^2));
```

```
%Τύπος δυναμικής ενέργειας της 1ης αριθμητικής εφαρμογής
```

```
U_q_plus_dq = log(1/(sqrt(2*pi))*exp(-q_plus_dq^2/2)/(0.5*sqrt(2*pi))*exp(-  
0.5*(q_plus_dq-2)^2/0.5^2));
```

```
grad_U = (U_q_plus_dq - U_q)/dq;
```

```
end
```

```
function K = kinetic (p)
```

```
K = sum(p^2)/2;
```

```
end
```

```
end
```

Copyright (c) 2011, Matthew D. Hoffman All rights reserved

Παράρτημα 2

Ακολούθως παρατίθεται ο κώδικας του μοντέλου πεπερασμένων στοιχείων σε γλώσσα Matlab που χρησιμοποιήθηκε στη δεύτερη αριθμητική εφαρμογή. Ο όρος input «increments» αναφέρεται στον αριθμό των υποδιαίρεσεων του φορτίου load για τις οποίες τρέχει η συνάρτηση, οι οποίες προσαρμόζονται ανάλογα με τις απαιτήσεις της εκάστοτε εφαρμογής. Για παράδειγμα, για increments = 2, η συνάρτηση υπολογίζει δύο responses, ένα για φορτίο $x/2$ και ένα για φορτίο x . Οι συνοδές συναρτήσεις που καλούνται, καθώς και τυχόν διευκρινίσεις, μπορούν να παρασχεθούν από τον συγγραφέα του κώδικα, υποψήφιο διδάκτορα του Τομέα Δομοστατικής της Σχολής Πολιτικών Μηχανικών ΕΜΠ Στέφανο Πυριαλάκο, μέσω του email του stefpyr@gmail.com. Τον ευχαριστώ θερμά για την συμβολή του στην παρούσα εργασία.

```
function [response] = model (increments, load, parameters)

%parameters --> constitutive parameters of the hardening law of Von Mises
%plasticity material. The hardening law is of the form:
%?(x)=[ parameters(1) + parameters(2)*(1-exp(-parameters(3)*x))

%response --> the midspan deflection of a beam subjected to a 3-point test

[ModelProperties,InclusionProperties,CohesiveProperties,MatrixElementProperties,MatrixMaterialProperties] = define_properties(parameters);
%ModelProperties.seed = seed;

[Nodes, ElementConnectivity, NodalLoads, NodalDisplacements, num_free_dofs,
num_constrained_dofs, global_dof_order, element_to_global_dofs, InclusionProperties,
...
InclusionCoord, InclusionMap, EmbeddingProjectionMatrix, EmbeddingMap,
D]=create_model(ModelProperties, MatrixElementProperties, InclusionProperties, load);

%load(['CNT_RVE_' num2str(6)
'.mat'],'InclusionProperties','InclusionMap','InclusionCoord','EmbeddingMap','EmbeddingProjectionMatrix','CohesiveProperties');

%Newton-Raphson Properties
ModelProperties.increments=increments;
ModelProperties.iterations=50;
ModelProperties.tol=1e-2;

[MatrixElementProperties, InclusionProperties, N, B, ~, detJ, w,
Nex]=elastic_stiffness_matrix(ModelProperties, MatrixElementProperties,
MatrixMaterialProperties, InclusionProperties, Nodes, ElementConnectivity);

% [Ksolid_ff, Ksolid_fc]=assemble_global_stiffness_matrix(ModelProperties,
MatrixElementProperties, InclusionProperties, ElementConnectivity ...
% , num_free_dofs, num_constrained_dofs, element_to_global_dofs,
EmbeddingProjectionMatrix, EmbeddingMap);
```

```

[Ksolid_ff, Ksolid_fc]=assemble_global_stiffness_matrix_upper(ModelProperties,
MatrixElementProperties, InclusionProperties, ElementConnectivity ...
, element_to_global_dofs, EmbeddingProjectionMatrix, EmbeddingMap);

U = zeros(num_free_dofs, 1);
dU = zeros(num_free_dofs, 1);
U_prev = zeros(num_free_dofs, 1);

residual_prev = zeros(size(Ksolid_ff,1),1);
IncrDisplacements = NodalDisplacements;
IncrDisplacements(:,3) = NodalDisplacements(:,3) / ModelProperties.increments;
tic
for i=1:ModelProperties.increments
%   %DISPL CONTROL
%   dU = -Ksolid_ff \ (Ksolid_fc * IncrDisplacements(:,3));
%   U_it = dU;
%LOAD CONTROL
U_it = zeros(num_free_dofs, 1);
IncrLoads = NodalLoads;
IncrLoads(:,3) = NodalLoads(:,3) * i / ModelProperties.increments ;
Fext_f =
assemble_global_external_forces(IncrLoads,num_free_dofs,global_dof_order);
for j=1:ModelProperties.iterations

    [MatrixMaterialProperties] = calculate_strain(ModelProperties,
MatrixElementProperties, MatrixMaterialProperties, element_to_global_dofs,
IncrDisplacements, U_it, B);

    [MatrixMaterialProperties] =
update_material_total(ModelProperties,MatrixMaterialProperties);

    %update_cohesive_zone(CohesiveProperties, dE, E_prev);

    [MatrixElementProperties] = update_stiffness_matrix(ModelProperties,
MatrixElementProperties, MatrixMaterialProperties, B, w, detJ);

    [Ksolid_ff,
Ksolid_fc]=assemble_global_stiffness_matrix_upper(ModelProperties,
MatrixElementProperties, InclusionProperties, ElementConnectivity ...
, element_to_global_dofs, EmbeddingProjectionMatrix, EmbeddingMap);

    [MatrixElementProperties] = update_internal_forces(ModelProperties,
MatrixElementProperties, MatrixMaterialProperties, B, w, detJ);

    [InclusionProperties] = update_beam_forces(ModelProperties,
InclusionProperties, EmbeddingProjectionMatrix, EmbeddingMap, IncrDisplacements, i,
U, U_it);

    [Fint, Fint_f, Fint_c] = assemble_global_internal_forces(ModelProperties,
MatrixElementProperties, InclusionProperties, ElementConnectivity ...
, num_free_dofs, num_constrained_dofs,
element_to_global_dofs, EmbeddingProjectionMatrix, EmbeddingMap);

    %residual = residual_prev-Fint_f;
    residual = (Fext_f - Fint_f);

```

```

    if j==1
        residual1 = norm(residual);
    end
    dU = Ksolid_ff \ residual;

    criterion = norm(residual)/residual1;
    U_it = U_it + dU;
    disp(criterion);
    if criterion < ModelProperties.tol
        break;
    end
end
if criterion < ModelProperties.tol
    U_prev = U;
    U = U + U_it;
    [MatrixMaterialProperties] =
save_material_state(ModelProperties,MatrixMaterialProperties);
    residual_prev = residual;

    fprintf('iterations : %i',j)

    if ModelProperties.RVE == true
        [Ksolid_cc,
Ksolid_cf]=assemble_global_stiffness_matrix_lower(ModelProperties,
MatrixElementProperties, InclusionProperties, ElementConnectivity ...
, element_to_global_dofs, EmbeddingProjectionMatrix, EmbeddingMap);
        Kbb_dash=Ksolid_cc-Ksolid_cf*(full(Ksolid_ff)\Ksolid_fc);
        MacroStress=1/ModelProperties.Volume*D*Fint_c;
        C=1/ModelProperties.Volume*D*Kbb_dash*D';

        dimension = 1;
    end

PostProcess.plotF(i)=Fext_f(((ModelProperties.NumberElementsX+1)*ModelProperties.Numb
erElementsY+round(ModelProperties.NumberElementsX/2))*2-4); %42 %1478 %4815
    PostProcess.plotU(i)=U(ModelProperties.NumberElementsX-2); %42 %3387
else
    warning('The Newton solver did not converge at %i increment.',i)
    break;
end
end
response = PostProcess.plotU;
%post_process(ModelProperties,PostProcess);
end

```

Copyright © Πυριαλάκος Στέφανος, 2023.