



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

— ΙΔΡΥΘΕΝ ΤΟ 1837 —

Σχολή Θετικών Επιστήμων

Τμήμα Μαθηματικών

Μέθοδοι Προσομοίωσης σε Προβλήματα Βελτιστοποίησης

Διπλωματική Εργασία

Μεταπτυχιακό Πρόγραμμα Σπουδών
Στατιστική και Επιχειρησιακή Έρευνα

Συγγραφέας:

Τσουτσουλοπούλου Δήμητρα

Επιβλέπων Καθηγητής:

Μπουρνέτας Απόστολος

Αθήνα

Σεπτέμβριος 2021

Περίληψη

Η βελτιστοποίηση αποτελεί ένα κλάδο των μαθηματικών με πολλές πρακτικές εφαρμογές σε πλήθος διαφορετικών επιστημονικών περιοχών. Οι τεχνολογικές εξελίξεις κατέστησαν εφικτή την επιστράτευση της προσομοίωσης ως εργαλείου στη βελτιστοποίηση οδηγώντας στην ανάπτυξη των μεθόδων βελτιστοποίησης μέσω προσομοίωσης. Η βελτιστοποίηση μέσω προσομοίωσης περιλαμβάνει μεθόδους που μπορούν να εφαρμοστούν σε μεγάλο εύρος προβλημάτων και να αντιμετωπίσουν με επιτυχία προβλήματα στα οποία οι κλασσικές μέθοδοι συναντούν δυσκολίες.

Η παρούσα εργασία χωρίζεται σε δύο μέρη. Στο πρώτο μέρος παρουσιάζεται μία ανασκόπηση των βασικών μεθόδων βελτιστοποίησης μέσω προσομοίωσης. Οι διάφορες μέθοδοι κατηγοριοποιούνται με βάση τη μορφή της εφικτής περιοχής του προς επίλυση προβλήματος. Διακρίνονται οι περιπτώσεις με συνεχή ή διακριτή περιοχή, ενώ στη δεύτερη περίπτωση γίνεται περαιτέρω διάκριση ανάλογα με το μέγεθος της εφικτής περιοχής. Παρουσιάζονται σύντομα τα βασικά σημεία μιας επιλογής των μεθόδων αυτών.

Στο δεύτερο μέρος της εργασίας εξετάζεται η εφαρμογή της μεθόδου βελτιστοποίησης Εμφωλευμένων Διαμερίσεων στο πρόβλημα κατανομής θέσεων στους χώρους αναμονής ενός δικτύου, γνωστό ως Buffer Allocation Problem.

Abstract

Optimization is a branch of mathematics with many practical applications in numerous scientific fields. Technological advances enabled the use of simulation in optimization, leading to the development of simulation optimization methods. Simulation optimization comprises various methods that cover a great range of applications and successfully address problems for which classic approaches fall short.

This thesis consists of two parts. In the first part a brief literature review of the basic simulation optimization methods is presented. These methods are categorized according to the form of the problem's feasible region which can be either continuous or discrete. The latter methods can be further classified according to the size of their feasible region. The basic points of a selection of these methods are demonstrated.

In the second part of the thesis, the Nested Partitions method is briefly explained and applied to the Buffer Allocation Problem.

Ευχαριστίες

Πρώτα από όλα θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Απόστολο Μπουρνέτα για την υπομονή και καθοδήγησή του, καθώς κάθε του επέμβαση ήταν πάντα καταλυτική για την πορεία της εργασίας. Τους καθηγητές κ. Αντώνη Οικονόμου και κ. Σάμη Τρέβεζα για τη συμμετοχή τους στην τριμελή επιτροπή και το χρόνο που αφιέρωσαν, καθώς επίσης και όσους ακόμα καθηγητές του τμήματος είχα την τιμή να γνωρίσω στα χρόνια των προπτυχιακών αλλά και μεταπτυχιακών σπουδών μου για όσα έχουν προσφέρει.

Ακόμα θα ήθελα να εκφράσω την ευγνωμοσύνη μου στους γονείς μου και τις αδερφές μου για τη στήριξη και τη βοήθεια τους σε όλα τα χρόνια των σπουδών μου μέχρι τώρα, τους φίλους μου για τη συμπαράστασή τους και φυσικά στο σύντροφό μου που είναι πάντα δίπλα μου και με βοηθάει με κάθε τρόπο.

Περιεχόμενα

1	Εισαγωγή	7
1.1	Βελτιστοποίηση	7
1.2	Προσομοίωση	8
1.3	Βελτιστοποίηση Μέσω Προσομοίωσης	9
1.3.1	Δυσκολίες	9
1.3.2	Πλεονεκτήματα	10
1.3.3	Μορφοποίηση	10
1.4	Κατηγοριοποίηση	11
2	Συνεχής Βελτιστοποίηση Μέσω Προσομοίωσης	14
2.1	Στοχαστική Προσέγγιση	14
2.2	Στοχαστική Εκτίμηση Παραγώγων	15
2.2.1	Εκτίμηση Πεπερασμένων Διαφορών	15
2.2.2	Στοχαστική Προσέγγιση με Ταυτόχρονες Διαταραχές	16
2.2.3	Ανάλυση Διαταραχής	16
2.2.4	Εκτιμήτριες Λόγου Πιθανοφανειών	17
2.3	Μεθοδολογία Επιφάνειας Απόκρισης	18
2.4	Τυχαία Αναζήτηση	19
3	Διακριτή Βελτιστοποίηση Μέσω Προσομοίωσης	21
3.1	Μικρό Πλήθος Εφικτών Λύσεων	21
3.1.1	Κατάταξη και Επιλογή	21
3.1.2	Διαδικασίες Πολλαπλών Συγκρίσεων	22
3.1.3	Βέλτιστη Κατανομή Υπολογιστικών Πόρων	22
3.2	Μεγάλο Πλήθος Εφικτών Λύσεων	23
3.2.1	Προσέγγιση Δειγματικού Μέσου	23
3.2.2	Διατακτική Βελτιστοποίηση	29
3.2.3	Άμεση Αναζήτηση	29
3.2.4	Τυχαία Αναζήτηση	30
3.2.5	Εμφωλευμένες Διαμερίσεις	32

4	Εφαρμογή της Μεθόδου Εμφωλευμένων Διαμερίσεων	42
4.1	Θεμελίωση του Προβλήματος	42
4.2	Ανάπτυξη Εφαρμογής	42
4.3	Παρουσίαση Αποτελεσμάτων	46
5	Συμπεράσματα	54
	Βιβλιογραφία	55

1 Εισαγωγή

1.1 Βελτιστοποίηση

Ο όρος βελτιστοποίηση μπορεί, πολύ γενικά, να οριστεί ως η διερεύνηση και επιλογή της βέλτιστης δυνατής απόφασης κάτω από μια δεδομένη κατάσταση. Τέτοιες αποφάσεις έχουν συνήθως ως κύριο στόχο τη μεγιστοποίηση ή ελαχιστοποίηση ενός κέρδους ή μίας απώλειας, αντίστοιχα, που υπεισέρχεται σε μια συγκεκριμένη διαδικασία. Προκειμένου να υλοποιηθεί αυτός ο στόχος, αρχικά, απαιτείται η έκφραση του επιθυμητού κέρδους, των απαραίτητων πόρων και άλλων σχετικών πληροφοριών του προβλήματος ως μια συνάρτηση συγκεκριμένων μεταβλητών, οι οποίες ονομάζονται συνήθως μεταβλητές απόφασης. Έτσι, ως βελτιστοποίηση ορίζεται η μεγιστοποίηση ή ελαχιστοποίηση μίας συνάρτησης αρκετών μεταβλητών. Η συνάρτηση αυτή, η αντικειμενική συνάρτηση, όπως καλείται, ή μέτρο απόδοσης, μπορεί να είναι ελεύθερη περιορισμών ή να υπόκειται σε περιορισμούς, ανισοτικούς ή εξισωτικούς, για τις μεταβλητές [1].

Η προέλευση προβλημάτων βελτιστοποίησης εντοπίζεται στα μέσα του δεκάτου ογδόου αιώνα, με τις πρώτες εφαρμογές στη γεωμετρία και τη φυσική. Πρωτοποριακό ήταν το έργο των Newton, Lagrange και Cauchy σχετικά με την επίλυση τέτοιων προβλημάτων με τη χρήση διαφορικού λογισμού και λογισμού μεταβολών. Ανέπτυξαν αναλυτικές μεθόδους που αποσκοπούν στην εύρεση ακροτάτων για συνεχείς και διαφορίσιμες συναρτήσεις. Οι προσεγγίσεις αυτές είναι ευρύτερα γνωστές ως κλασσικές μέθοδοι βελτιστοποίησης. Δεν μπορούν, όμως, να εφαρμοστούν επιτυχώς σε οποιοδήποτε πρόβλημα βελτιστοποίησης [1].

Τα προβλήματα που συνήθως δεν είναι επιλύσιμα με τη χρήση μεθόδων κλασσικής βελτιστοποίησης είναι γνωστά ως προβλήματα μαθηματικού προγραμματισμού. Τις τελευταίες πέντε δεκαετίες έχει σημειωθεί πρωτοφανής πρόοδος ως προς την ανάπτυξη θεωρίας και αλγορίθμων για την επίλυση ποικίλων τύπων προβλημάτων μαθηματικού προγραμματισμού [1].

Πιο ειδικά, ο μαθηματικός προγραμματισμός ή αλλιώς μαθηματική βελτιστοποίηση μπορεί να περιγραφεί ως η επιστήμη που προσδιορίζει τις καλύτερες λύσεις για προβλήματα που έχουν οριστεί και μοντελοποιηθεί μαθηματικά. Η ιστορία της μαθηματικής βελτιστοποίησης, όσον αφορά συναρτήσεις πολλών μεταβλητών είναι σχετικά σύντομη. Ξεκινάει από τα τέλη του 1940 με την πολύ σημαντική μέθοδο simplex για την επίλυση προβλημάτων γραμμικού προγραμματισμού. Από τότε έχει αναπτυχθεί πλήθος μεθόδων για την επίλυση γενικότερων προβλημάτων μαθηματικού προγραμματισμού. Σημαντικό άλμα στην ανάπτυξη των μεθόδων και αλγορίθμων βελτιστοποίησης ακολούθησε την εμφάνιση του ηλεκτρονικού υπολογιστή.

Βέβαια, παρά την ταχεία ανάπτυξη που ακολούθησε τις τεχνολογικές αυτές εξελίξεις, δεν υπάρχει κάποια μέθοδος επίλυσης που να απευθύνεται σε όλα τα προβλήματα βελτιστοποίησης καθολικά. Τίθεται, επομένως, το ζήτημα της επιλογής κατάλληλου αλγορίθμου από τον ενδιαφερόμενο για την επίλυση ενός προβλήματος βελτιστοποίησης. Αυτή η επιλογή αποτελεί πολύ σημαντικό παράγοντα για την ταχύτητα εύρεσης βέλτιστης λύσης αλλά ακόμα και για το αν αυτή τελικά θα βρεθεί [2].

1.2 Προσομοίωση

Προσομοίωση καλείται η μίμηση της λειτουργίας μιας διαδικασίας ή ενός συστήματος εξελικτικά στο χρόνο. Δίνει τη δυνατότητα δημιουργίας μιας τεχνητής προϊστορίας του υπό μελέτη συστήματος και την παρατήρηση αυτής για την διεξαγωγή συμπερασματολογίας σχετικά με συγκεκριμένα χαρακτηριστικά ενδιαφέροντος του συστήματος που αναπαριστάται [3, 4].

Ένα μοντελοποιημένο σύστημα μπορεί να υποστεί εύκολα τροποποιήσεις, επαναπροσδιορισμούς και πειραματισμούς κάτι που μπορεί να είναι ακριβό ή μη πρακτικό για ένα πραγματικό σύστημα. Έτσι η προσομοίωση καθίσταται αναγκαία μεθοδολογία επίλυσης για πολλά πραγματικά προβλήματα [3, 4].

Η προσομοίωση χρησιμοποιείται για περιγραφή και ανάλυση της συμπεριφοράς ενός συστήματος κάτω από διαφορετικές παραμέτρους, συμβάλει στην απάντηση υποθετικών ερωτήσεων για το πραγματικό σύστημα και βοηθάει στο σχεδιασμό πραγματικών συστημάτων. Το τελευταίο μπορεί να αφορά στην τροποποίηση ενός υπάρχοντος συστήματος ή την κατασκευή ενός νέου. Πιο συγκεκριμένα η προσομοίωση συμβάλλει στο να ελεγχθούν ζητήματα ελαχιστοποίησης της πιθανότητας αποτυχίας, ικανοποίησης καθορισμένων προδιαγραφών, τήρησης συγκεκριμένου προϋπολογισμού και βελτιστοποίησης της απόδοσης του συστήματος που μελετάται [4].

Οι μέθοδοι προσομοίωσης μπορούν να χωριστούν σε δύο βασικές κατηγορίες, την προσομοίωση διακριτών γεγονότων και την προσομοίωση συνεχούς χρόνου. Στην πρώτη, κεντρική υπόθεση αποτελεί η μεταβολή του συστήματος στιγμιαία, με την πραγματοποίηση καθορισμένων διακριτών γεγονότων. Στη δεύτερη, επιχειρείται η ποσοτικοποίηση των αλλαγών που υφίσταται το σύστημα συνεχώς στο χρόνο κατά τη μεταβολή των παραμέτρων ελέγχου που το επηρεάζουν. Η προσομοίωση διακριτών γεγονότων μπορεί να είναι λιγότερο πληροφοριακή σε σχέση με το συνεχή χρόνο, ακριβώς γιατί «κοιτάει» το σύστημα σε διακριτές χρονικές στιγμές, είναι όμως πολύ πιο εύκολη στην εφαρμογή της [4].

1.3 Βελτιστοποίηση Μέσω Προσομοίωσης

Ιδιαίτερο ενδιαφέρον παρουσιάζει ο συνδυασμός βελτιστοποίησης και προσομοίωσης, καθώς ανοίγει το δρόμο για την επίλυση προβλημάτων βελτιστοποίησης που διαφορετικά μπορεί να ήταν υπερβολικά κοστοβόρα ή ακόμα και ανέφικτα. Ο συνδυασμός αυτός γέννησε τις μεθόδους βελτιστοποίησης μέσω προσομοίωσης (Simulation Optimization) που αποτελούν και το κεντρικό θέμα της παρούσας εργασίας.

Ο όρος βελτιστοποίηση μέσω προσομοίωσης αναφέρεται στη βελτιστοποίηση μίας αντικειμενικής συνάρτησης κάτω από συγκεκριμένους περιορισμούς, όπου τόσο η αντικειμενική συνάρτηση όσο και οι περιορισμοί μπορούν να υπολογιστούν μέσω στοχαστικής προσομοίωσης. Περιλαμβάνει την αναζήτηση των τιμών των παραμέτρων εισόδου μίας στοχαστικής προσομοίωσης, για τις οποίες η αντικειμενική συνάρτηση, που, χωρίς βλάβη της γενικότητας, είναι συνάρτηση των παραμέτρων εξόδου, ελαχιστοποιείται. Κομβικό σημείο για την ανάπτυξη και περαιτέρω μελέτη της βελτιστοποίησης μέσω προσομοίωσης αποτέλεσαν οι εξελίξεις στο πεδίο της μοντελοποίησης και η διαθεσιμότητα οικονομικής υπολογιστικής ισχύος [5].

Σε αντίθεση με τις μεθόδους μαθηματικού προγραμματισμού που βασίζονται σε αλγεβρικά μοντέλα, η βελτιστοποίηση μέσω προσομοίωσης δεν προϋποθέτει την ύπαρξη αλγεβρικής περιγραφής της προσομοίωσης. Η τελευταία μπορεί να εμφανίζεται σαν ένα μαύρο κουτί στην όλη διαδικασία, και να σχετίζεται με αυτή μόνο μέσω των δεδομένων εισόδου-εξόδου. Έτσι, η βελτιστοποίηση μέσω προσομοίωσης εξαλείφει τα μειονεκτήματα της προσομοίωσης, ενώ τη χρησιμοποιεί για την εύρεση της καλύτερης δυνατής λύσης [5, 6].

1.3.1 Δυσκολίες

Η κεντρική δυσκολία στη βελτιστοποίηση μέσω προσομοίωσης αφορά στο συμβιβασμό ανάμεσα στην κατανομή των υπολογιστικών πόρων για την αναζήτηση του χώρου λύσεων και τη διεξαγωγή επιπλέον επαναλήψεων προσομοίωσης για καλύτερη εκτίμηση της απόδοσης των υποψήφιων λύσεων. Ο συμβιβασμός αυτός γίνεται ιδιαίτερα εμφανής καθώς αυξάνεται το κόστος της προσομοίωσης. Έτσι, λόγω ενδεχόμενου υψηλού κόστους, τόσο από υπολογιστικής άποψης όσο και οικονομικής είναι συχνά απαραίτητη η χρήση μικρού σχετικά αριθμού επαναλήψεων προσομοίωσης κατά την αναζήτηση των βέλτιστων παραμέτρων [5, 7].

Ένα ακόμα ζήτημα που προκύπτει κατά τη βελτιστοποίηση προέρχεται από τη στοχαστική φύση της προσομοίωσης, καθώς τα εξαγόμενα δεν είναι ντετερμινιστικά, αλλά ακολουθούν κάποια κατανομή, η οποία ενδέχεται, μάλιστα, να διαφέρει για διαφορετικά σημεία του παρα-

μετρικού χώρου. Η έξοδος της προσομοίωσης, επομένως, χαρακτηρίζεται από αβεβαιότητα, γεγονός που καθιστά δυσκολότερη την αξιολόγηση της ποιότητας των παραμέτρων εξόδου [5].

Τέλος, όταν δεν είναι διαθέσιμη κάποια αλγεβρική περιγραφή της προσομοίωσης, όπως συμβαίνει στην πλειοψηφία των περιπτώσεων, τότε συνήθως δεν είναι διαθέσιμες ούτε πληροφορίες παραγώγων, στις περιπτώσεις, φυσικά, που αυτές υπάρχουν, και επιπλέον, ο θόρυβος που εντοπίζεται στην έξοδο σε συνδυασμό με το υψηλό κόστος προσομοίωσης, καθιστούν δύσκολη την εκτίμηση τους, μέσω πεπερασμένων διαφορών [5].

1.3.2 Πλεονεκτήματα

Ένα από τα βασικά πλεονεκτήματα των μεθόδων βελτιστοποίησης μέσω προσομοίωσης αποτελεί το γεγονός πως η πολυπλοκότητα του συστήματος που μοντελοποιείται δεν επηρεάζει σημαντικά την απόδοση της διαδικασίας. Ακόμα ιδιαίτερα χρήσιμη είναι η δυνατότητα ελέγχου της διασποράς για στοχαστικά συστήματα μέσω ποικίλων τεχνικών ανάλυσης των αποτελεσμάτων. Τέλος, ένα βασικό πλεονέκτημα που προσφέρει η προσομοίωση σχετίζεται με τη δυνατότητα τροποποίησης της αντικειμενικής συνάρτησης ή/και των περιορισμών ανάμεσα στις επαναλήψεις ώστε να αναπαριστώνται οι διαφορετικές μορφές ενός συστήματος, κάτι που δεν μπορεί να υλοποιηθεί από τις κλασσικές διαδικασίες βελτιστοποίησης που απαιτούν αυστηρή μαθηματική μοντελοποίηση [8].

1.3.3 Μορφοποίηση

Μια πολύ γενική μορφή των προβλημάτων στα οποία απευθύνονται οι μέθοδοι βελτιστοποίησης μέσω προσομοίωσης είναι η ακόλουθη [5]:

$$\begin{aligned}
 & \min \mathbb{E}_\omega[f(x, y, \omega)] \\
 & \text{s.t. } \mathbb{E}_\omega[g(x, y, \omega)] \leq 0 \\
 & \quad h(x, y) \leq 0 \\
 & \quad x_l \leq x \leq x_u \\
 & \quad x \in \mathbb{R}^n, y \in \mathbb{D}^m.
 \end{aligned} \tag{1}$$

Σε αυτή τη μορφοποίηση η αντικειμενική συνάρτηση αποτελεί τη μέση τιμή μίας συνάρτησης f , και οι τιμές της μπορούν να εκτιμηθούν μέσω προσομοίωσης για συγκεκριμένες συνεχείς μεταβλητές εισόδου x , διακριτές μεταβλητές εισόδου y και για μια πραγματοποίηση

του τυχαίου διανύσματος ω , που μπορεί να είναι ή όχι συνάρτηση των μεταβλητών εισόδου x και y . Η τυχειότητα, επομένως, εισάγεται στο πρόβλημα μέσω του τυχαίου διανύσματος ω . Το πρόβλημα περιέχει ντετερμινιστικούς και στοχαστικούς περιορισμούς. Όπως και στην περίπτωση της αντικειμενικής συνάρτησης, οι στοχαστικοί περιορισμοί που ορίζονται από τη διανυσματική συνάρτηση g μπορούν επίσης να εκτιμηθούν σε κάθε επανάληψη προσομοίωσης [5].

Η χαλάρωση κάποιων εκ των παραπάνω συνθηκών οδηγεί σε ένα πιο ειδικό πρόβλημα που εντάσσεται, όμως, στα πλαίσια της βελτιστοποίησης μέσω προσομοίωσης. Οι περισσότεροι αλγόριθμοι εστιάζουν σε προβλήματα που διαχειρίζονται αποκλειστικά είτε διακριτές είτε συνεχείς μεταβλητές αποφάσεων. Λίγοι από τους υπάρχοντες αλγόριθμους επιχειρούν να απευθυνθούν και σε συνεχείς και σε διακριτές μεταβλητές απόφασης ταυτόχρονα. Βέβαια υπάρχουν αρκετές προσεγγίσεις που, εκ φύσεως, με κατάλληλες προσαρμογές, είναι εφαρμόσιμες και στις δύο περιπτώσεις [5].

Σχετικά με τη φύση της αντικειμενικής συνάρτησης, το πιο συχνό ζητούμενο στη βελτιστοποίηση μέσω προσομοίωσης είναι η βελτιστοποίηση μιας αναμενόμενης τιμής κάποιου μέτρου απόδοσης. Βέβαια και άλλες αντικειμενικές συναρτήσεις μπορεί να είναι κατάλληλες, ανάλογα με την εκάστοτε εφαρμογή. Ένα παράδειγμα τέτοιας εναλλακτικής θα μπορούσε να αποτελεί η ελαχιστοποίηση κάποιου κινδύνου [5].

Στη συνέχεια της εργασίας η μορφή του προβλήματος που θα εξετάζεται θα είναι η ακόλουθη,

$$\min_{\theta \in \Theta} J(\theta)$$

όπου $J(\theta) = \mathbb{E}_{\omega}[f(\theta, \omega)]$ για πραγματική συνάρτηση f και τυχαίο διάνυσμα ω . Το σύνολο επιτρεπόμενων λύσεων Θ θα ορίζεται ανάλογα με το πρόβλημα και μπορεί να είναι διακριτό ή συνεχές. Η φύση των περιορισμών δεν θα μας απασχολήσει για την κατηγοριοποίηση και παρουσίαση των αλγορίθμων που θα ακολουθήσει, έτσι αυτοί δεν θα διατυπώνονται. Σε κάθε περίπτωση θα εντάσσονται στα πλαίσια της γενικής μορφοποίησης που αναφέρθηκε παραπάνω.

1.4 Κατηγοριοποίηση

Διαφορετικές προσεγγίσεις έχουν αποδειχτεί καταλληλότερες για τα διάφορα προβλήματα βελτιστοποίησης ανάλογα με το κατά πόσο είναι διαθέσιμη η αλγεβρική δομή του προβλήματος ή όχι αλλά και ανάλογα με το αν υπεισέρχεται τυχειότητα στο πρόβλημα. Έτσι, για προβλήματα

τα με γνωστό αλγεβρικό μοντέλο επιστρατεύονται τεχνικές παραδοσιακού προγραμματισμού (γραμμικός, ακέραιος και μη γραμμικός προγραμματισμός) αν το πλαίσιο είναι ντετερμινιστικό, ενώ εφαρμόζεται στοχαστικός προγραμματισμός ή ευσταθής βελτιστοποίηση (robust optimization) αν υπάρχει αβεβαιότητα. Τέλος, για προβλήματα με άγνωστη ή σύνθετη δομή εφαρμόζονται μέθοδοι βελτιστοποίησης ελεύθερες παραγώγων (derivative-free optimization) στο ντετερμινιστικό πλαίσιο και βελτιστοποίησης μέσω προσομοίωσης στο στοχαστικό πλαίσιο [5]. Στη συνέχεια παρουσιάζεται μία κατηγοριοποίηση των μεθόδων βελτιστοποίησης μέσω προσομοίωσης και μία σύντομη ανάλυση μερικών από αυτές.

Αρχικά, αξίζει να σημειωθεί πως καθώς η βελτιστοποίηση μέσω προσομοίωσης εμπεριέχει τη χρήση αλγορίθμων από πολλά διαφορετικά πεδία, σχετίζεται με πολλές διαφορετικές περιοχές και έχει πληθώρα πρακτικών εφαρμογών από τη βιολογία μέχρι τη μηχανική και τη λογιστική, δεν αποτελεί έκπληξη ότι απαντάται με πολλές διαφορετικές ονομασίες. Μερικές από αυτές είναι: βελτιστοποίηση βασισμένη στην προσομοίωση, στοχαστική βελτιστοποίηση, παραμετρική βελτιστοποίηση, βελτιστοποίηση μαύρου κουτιού, ενώ οι συνεχείς και διακριτές εκδοχές είναι γνωστές ως: συνεχής βελτιστοποίηση μέσω προσομοίωσης και διακριτή βελτιστοποίηση μέσω προσομοίωσης αντίστοιχα. Ανάλογα, και κάθε αλγοριθμική προσέγγιση απαντάται με πολλά διαφορετικά ονόματα [5].

Σημειώνεται ότι κατά την εφαρμογή μεθόδων βελτιστοποίησης μέσω προσομοίωσης είναι ιδιαίτερα σημαντική η επιστράτευση κάθε διαθέσιμης γνώσης πάνω στον τομέα εφαρμογής. Η γνώση αυτή μπορεί να βρει εφαρμογή στη διαλογή των μεταβλητών εισόδου, τον καθορισμό κατάλληλης κλίμακας και κατάλληλου εύρους των μεταβλητών απόφασης, την παροχή καλών τιμών αρχικοποίησης για τον αλγόριθμο και την άντληση άλλων χρήσιμων πληροφοριών από τη γνωστή δομή του προβλήματος [5].

Οι δυνατότητες στο συνδυασμό προσομοίωσης και βελτιστοποίησης είναι αχανείς. Έχει αναπτυχθεί μεγάλο πλήθος αλγορίθμων βελτιστοποίησης με χρήση προσομοίωσης των οποίων η εφαρμοσιμότητα είναι στενά συνδεδεμένη με τα χαρακτηριστικά της εκάστοτε εφαρμογής. Αυτή η ποικιλομορφία, μάλιστα, καθιστά δύσκολο έργο τη σύγκριση ανάμεσα στους διάφορους αλγορίθμους και τις εφαρμογές τους [5].

Υπάρχουν αρκετές κατηγοριοποιήσεις των προβλημάτων βελτιστοποίησης μέσω προσομοίωσης. Μία από αυτές, διαχωρίζει τα προβλήματα ανάλογα με τη μορφή της αντικειμενικής συνάρτησης, ως προς το αν είναι μονοδιάστατη ή πολυδιάστατη, άλλη ανάλογα με τον τρόπο έκφρασης του Θ , αν δηλαδή αυτό ορίζεται άμεσα ή έμμεσα, ντετερμινιστικά ή πιθανοτικά ενώ μία τρίτη αφορά στο είδος των μεταβλητών απόφασης, αν είναι συνεχείς ή διακριτές, ποσοτικές

ή ποιοτικές. Η κατηγοριοποίηση που θα ακολουθηθεί στη συνέχεια γίνεται με βάση τη μορφή του συνόλου της εφικτής περιοχής Θ , ως προς το είδος, δηλαδή, των μεταβλητών απόφασης. Πιο αναλυτικά, διακρίνονται οι περιπτώσεις που οι μεταβλητές απόφασης είναι συνεχείς ή διακριτές, ενώ στη τελευταία περίπτωση γίνεται περαιτέρω διάκριση ανάλογα με το αν ο χώρος λύσεων είναι πεπερασμένος και μικρός, πεπερασμένος και μεγάλος ή άπειρος [7, 8].

2 Συνεχής Βελτιστοποίηση Μέσω Προσομοίωσης

Οι μέθοδοι στις οποίες θα αναφερθούμε στο παρόν κεφάλαιο απευθύνονται σε προβλήματα με συνεχείς μεταβλητές απόφασης. Πρόκειται για προβλήματα που έχουν απασχολήσει σε μεγάλο βαθμό τους ερευνητές. Στη συνέχεια παρουσιάζεται μία επιλογή των μεθόδων αυτών από την βιβλιογραφία.

2.1 Στοχαστική Προσέγγιση

Οι μέθοδοι στοχαστικής προσέγγισης (stochastic approximation) αποτελούν μεθόδους συνεχούς βελτιστοποίησης μέσω προσομοίωσης βασισμένες στο πρωτότυπο έργο των Robbins και Monro (1951) και Kiefer και Wolfowitz (1952). Πρόκειται για επαναληπτικούς αλγορίθμους που μιμούνται ντετερμινιστικές μεθόδους μη γραμμικής βελτιστοποίησης που κάνουν χρήση παραγώγων. Η στοχαστική προσέγγιση αποτελεί μια από τις πλέον εφαρμόσιμες και χρήσιμες μεθόδους της βελτιστοποίησης μέσω προσομοίωσης. Για τέτοιους αλγορίθμους έχουν αποδειχθεί ιδιότητες ασυμπτωτικής σύγκλισης κάτω από κατάλληλες συνθήκες [7, 8, 9].

Πιο ειδικά, ο όρος στοχαστική προσέγγιση αναφέρεται σε μια οικογένεια διαδικασιών επαναληπτικού χαρακτήρα που προσεγγίζουν το βέλτιστο της θεωρητικής συνάρτησης παλινδρόμησης μίας στοχαστικής επιφάνειας απόκρισης με τη χρήση προσομοιωμένων παρατηρήσεων και των τιμών της συνάρτησης σε αυτές. Η επαναληπτική σχέση για συνάρτηση μίας μεταβλητής έχει τη μορφή:

$$X_{n+1} = X_n + (a_n/2c_n)[f(X_n + c_n) - f(X_n - c_n)],$$

όπου a_n και c_n είναι δύο ακολουθίες πραγματικών αριθμών που ικανοποιούν τις ακόλουθες συνθήκες:

$$\sum a_n < \infty, \lim_{n \rightarrow \infty} (c_n) = 0, \lim_{n \rightarrow \infty} (a_n/c_n)^2 < \infty.$$

Έχει αποδειχθεί ότι καθώς το n τείνει στο άπειρο, η X_n συγκλίνει σε μία λύση τέτοια ώστε η θεωρητική συνάρτηση παλινδρόμησης της στοχαστικής απόκρισης μεγιστοποιείται ή ελαχιστοποιείται. Η απόδειξη αυτή έχει επεκταθεί και στην περίπτωση των πολυδιάστατων μεταβλητών απόφασης [8].

2.2 Στοχαστική Εκτίμηση Παραγώγων

Για προβλήματα βελτιστοποίησης με συνεχείς μεταβλητές απόφασης, η ύπαρξη των παραγώγων μπορεί να βελτιώσει δραστικά την αποδοτικότητα των αλγορίθμων επίλυσης. Οι μέθοδοι στοχαστικής εκτίμησης παραγώγων (stochastic gradient estimation) αποσκοπούν στην εκτίμηση των παραγώγων της συνάρτησης απόδοσης, $\nabla J(\theta)$, για τις διάφορες τιμές των παραμέτρων. Η παρούσα ανάλυση προϋποθέτει ότι η συνάρτηση J είναι διαφορίσιμη στο θ . Με τη γνώση του $\nabla J(\theta)$ μπορεί να διεξαχθεί τοπική αναζήτηση παραγώγων (local gradient search). Η ύπαρξη θορύβου στην εκτίμηση των παραγώγων οδηγεί σε μια στοχαστική μορφή της αναζήτησης με χρήση παραγώγων (gradient search), γνωστή ως στοχαστική προσέγγιση. Κάτω από το στοχαστικό πλαίσιο, όμως, όπου τα εξαγόμενα είναι τα ίδια τυχαία, η εκτίμηση στοχαστικών παραγώγων αποτελεί από μόνη της ένα απαιτητικό πρόβλημα. Σε αυτή την κατεύθυνση ακριβώς, εντοπίζεται και η σημαντικότερη συμβολή των ερευνητών που δραστηριοποιούνται στον τομέα της βελτιστοποίησης μέσω προσομοίωσης [7, 8, 9].

Δύο βασικοί παράγοντες που καθορίζουν την επιτυχία των συγκεκριμένων μεθόδων είναι η αξιοπιστία και η αποδοτικότητα. Η σημαντικότητα της αξιοπιστίας έγκειται στη στοχαστική φύση των εξαγόμενων της προσομοίωσης, καθώς ένα μεγάλο σφάλμα στην εκτίμηση των παραγώγων μπορεί να οδηγήσει στην κίνηση του αλγορίθμου σε μια εντελώς λανθασμένη κατεύθυνση. Η αποδοτικότητα έχει ιδιαίτερη σημασία διότι τα πειράματα της προσομοίωσης είναι κοστοβόρα και για αυτό το λόγο είναι επιθυμητή η εκτίμηση των παραγώγων με τον ελάχιστο αριθμό πραγματοποιήσεων της συνάρτησης [8].

Στη συνέχεια εξετάζονται κάποιες από τις μεθόδους στοχαστικής εκτίμησης παραγώγων.

2.2.1 Εκτίμηση Πεπερασμένων Διαφορών

Η πιο άμεση προσέγγιση στην εκτίμηση παραγώγων περιλαμβάνει την εκτίμηση των παραγώγων ενδιαφέροντος μέσω πεπερασμένων διαφορών (finite differences). Ο αλγόριθμος έχει την παρακάτω επαναληπτική μορφή:

$$\theta_{n+1} = \theta_n - \alpha_n \nabla J(\theta_n),$$

όπου με $J(\cdot)$ συμβολίζεται η αντικειμενική συνάρτηση του προβλήματος, που έχει οριστεί ως $J(\theta) = \mathbb{E}_\omega[f(\theta, \omega)]$, με θ_n η τιμή του διανύσματος παραμέτρων κατά την n -οστή επανάληψη και με $\{\alpha_n\}$ μία ακολουθία συντελεστών μεγέθους βήματος, που καλείται συχνά ακολουθία κέρδους, και πρέπει να συγκλίνει στο μηδέν με κατάλληλο ρυθμό προκειμένου να εξασφαλιστεί

η σχεδόν βέβαιη σύγκλιση του αλγορίθμου.

Η i -οστή συνιστώσα του $\nabla J(\theta)$ μπορεί να εκτιμηθεί ως

$$\left(\hat{\nabla} J(\theta)\right)_i = \frac{f(\theta + h_i e_i, \omega_A) - f(\theta - h_i e_i, \omega_B)}{2h_i},$$

όπου $h_i \in \mathbb{R}$ είναι η διαταραχή της i -οστής συνιστώσας, e_i το μοναδιαίο διάνυσμα στην i -οστή διεύθυνση, και τα ω_A και ω_B αναπαριστούν δύο διαφορετικές επαναλήψεις προσομοίωσης. Επιλέγοντας $\omega_A = \omega_B$ μπορεί να μειωθεί σημαντικά η διασπορά, όμως η δυνατότητα αυτή της επιλογής εξαρτάται από το εκάστοτε πρόβλημα. Το βασικό μειονέκτημα της συγκεκριμένης μεθόδου είναι πως το υπολογιστικό κόστος αυξάνει γραμμικά με τη διάσταση p του διανύσματος θ [7].

2.2.2 Στοχαστική Προσέγγιση με Ταυτόχρονες Διαταραχές

Μια άλλη, αποδοτικότερη προσέγγιση, από την τελευταία, αποτελεί η στοχαστική προσέγγιση με ταυτόχρονες διαταραχές (simultaneous perturbation stochastic approximation) η οποία απαιτεί μόνο δύο επαναλήψεις προσομοίωσης για κάθε εκτίμηση διανύσματος παραγώγων, ανεξάρτητα από τη διάσταση του διανύσματος θ . Η i -οστή συνιστώσα του $\nabla J(\theta)$ με βάση αυτή την προσέγγιση προκύπτει ως:

$$\left(\hat{\nabla} J(\theta)\right)_i = \frac{f(\theta + \Delta, \omega_A) - f(\theta - \Delta, \omega_B)}{2\Delta_i},$$

όπου Δ είναι ένα διάνυσμα η i -οστή συνιστώσα, Δ_i , του οποίου παίρνει τις τιμές $\pm h_i$ με πιθανότητα 0.5 [7].

Και οι δύο προαναφερθείσες προσεγγίσεις χαρακτηρίζονται ως μέθοδοι μαύρου κουτιού καθώς δεν χρησιμοποιείται καμία γνώση του μοντέλου προσομοίωσης. Μάλιστα, δεν υπάρχει κάτι που υποδηλώνει την ύπαρξη στοχαστικότητας στο μοντέλο, καθώς οι ίδιες εκτιμήτριες μπορούν να χρησιμοποιηθούν και σε ντετερμινιστικά μοντέλα. Βέβαια στα πλαίσια της προσομοίωσης μέσω βελτιστοποίησης συχνά υπάρχει διαθέσιμη γνώση σχετικά με το σύστημα που προσομοιώνεται, όπως για παράδειγμα οι κατανομές που γεννούν τις μεταβλητές εισόδου [7].

2.2.3 Ανάλυση Διαταραχής

Η ανάλυση διαταραχής (perturbation analysis) όταν εφαρμοστεί κατάλληλα και σε μοντέλα που ικανοποιούν συγκεκριμένες συνθήκες, εκτιμά όλες τις παραγώγους της αντικειμενικής συνάρτησης από ένα μόνο πείραμα προσομοίωσης. Η βασική αρχή της μεθόδου είναι πως

αν μία μεταβλητή απόφασης διαταράσσεται κατά μία απειροελάχιστη ποσότητα, η ευαισθησία στην απόκριση του συστήματος μπορεί να εκτιμηθεί εντοπίζοντας τον τρόπο διάδοσης της διαταραχής αυτής στο σύστημα. Η ευαισθησία αυτή αποτελεί μία συνάρτηση του ποσοστού των διαταραχών που αποσβένονται προτού ασκήσουν σημαντική επίδραση στην ποσότητα ενδιαφέροντος [8].

Υπάρχουν, όμως, κάποιες αρκετά περιοριστικές συνθήκες που πρέπει να ικανοποιούνται προκειμένου η μέθοδος απειροελάχιστης διαταραχής να είναι εφαρμόσιμη. Για παράδειγμα, αν ως αποτέλεσμα της διαταραχής μιας δοθείσας μεταβλητής η ακολουθία των γεγονότων που καθορίζουν τη συμπεριφορά του συστήματος αλλάζει, τότε τα αποτελέσματα της εν λόγω ανάλυσης μπορεί να μην είναι αξιόπιστα. Η συγκεκριμένη συνθήκη είναι πολύ πιθανό να παραβιάζεται αν ληφθεί υπόψη η πολύπλοκη φύση των προβλημάτων προσομοίωσης [8].

Μια ακόμα δυσκολία της εφαρμογής της ανάλυσης απειροελάχιστης διαταραχής σε προβλήματα βελτιστοποίησης μέσω προσομοίωσης, αποτελεί η ανάγκη καλής γνώσης του μοντέλου της προσομοίωσης και του τρόπου με τον οποίο αυτό λειτουργεί ή συχνά ακόμα και δημιουργίας του μοντέλου από το μηδέν, προκειμένου να ενταχθούν στα κατάλληλα σημεία επιπρόσθετες δυνατότητες εντοπισμού, απαραίτητες για την εφαρμογή της μεθόδου [8].

2.2.4 Εκτιμητρίες Λόγου Πιθανοφαινειών

Η μέθοδος λόγου πιθανοφαινειών (likelihood ratio method) κάνει χρήση προσομοίωσης για να εκτιμήσει την κλίση $g(\theta) = \nabla J(\theta)$ της αντικειμενικής συνάρτησης J . Θυμίζουμε ότι η J ορίζεται ως $J(\theta) = \mathbb{E}_\omega[f(\theta, \omega)]$. Αν θεωρήσουμε επιπλέον F_θ τη συνάρτηση κατανομής και h_θ την συνάρτηση πυκνότητας της $f(\theta, \omega)$ και υποθέσουμε ότι είναι δυνατή η εναλλαγή του ολοκληρώματος με τον τελεστή της κλίσης, προκύπτει

$$g(\theta) = \nabla \int \omega h_\theta(\omega) d\omega = \int \omega \nabla h_\theta(\omega) d\omega.$$

Τότε, πολλαπλασιάζοντας και διαιρώντας με $h_\theta(\omega)$ στο παραπάνω ολοκλήρωμα έχουμε

$$g(\theta) = \int \omega \left(\frac{\nabla h_\theta(\omega)}{h_\theta(\omega)} \right) h_\theta(\omega) d\omega = \mathbb{E} \left\{ f(\theta, \omega) \left(\frac{\nabla h_\theta(f(\theta, \omega))}{h_\theta(f(\theta, \omega))} \right) \right\}.$$

Υποθετείται η σύμβαση ότι $0/0 = 1$. Εύκολα αποδεικνύεται ότι δεν θα προκύψει διαίρεση με το μηδέν, καθώς αν $h_\theta(\omega) = 0$ τότε $\nabla h_\theta(\omega) = 0$. Μπορούμε επομένως να παράγουμε ανεξάρτητες παρατηρήσεις $\omega_1(\theta), \dots, \omega_N(\theta)$ από την κατανομή F_θ και να εκτιμήσουμε την $g(\theta)$ ως [10]:

$$\frac{1}{N} \sum_{i=1}^N \omega_i(\theta) \frac{\nabla h_{\theta}(\omega)}{h_{\theta}(\omega)}.$$

2.3 Μεθοδολογία Επιφάνειας Απόκρισης

Η μεθοδολογία επιφάνειας απόκρισης (Response Surface Methodology-RSM) αναπτύχθηκε από τους Box and Wilson το 1951 και απευθύνεται κυρίως σε προβλήματα συνεχούς βελτιστοποίησης. Πρόκειται για επαναληπτικούς αλγορίθμους που χρησιμοποιούν στατιστικές μεθόδους, κυρίως παλινδρόμηση, με σκοπό την βελτίωση των υποψήφιων λύσεων διεξάγοντας αναζήτηση στο χώρο εφικτών λύσεων. Εστιάζει στην εκμάθηση των σχέσεων εισόδου-εξόδου ώστε να προσεγγίσει την υποκείμενη προσομοίωση μέσω μιας επιφάνειας, γνωστής και ως μετα-μοντέλο ή υποκατάστατη επιφάνεια, για την οποία προσδιορίζεται μια συναρτησιακή μορφή. Αυτή η συναρτησιακή μορφή μπορεί να χρησιμοποιηθεί για την αξιοποίηση ισχυρών τεχνικών βελτιστοποίησης που βασίζονται στις παραγώγους. Τεχνικές που υπάγονται σε αυτή την κατηγορία είναι ιδιαίτερα δημοφιλείς για την αντιμετώπιση προβλημάτων βελτιστοποίησης μέσω προσομοίωσης και η αντίστοιχη βιβλιογραφία είναι αχανής [5, 7].

Οι διάφοροι αλγόριθμοι επιφάνειας απόκρισης διαφέρουν ως προς την επιλογή ανάμεσα σε παλινδρόμηση και παρεμβολή, τη φύση της συναρτησιακής μορφής που χρησιμοποιείται για την προσέγγιση (πολυώνυμα, splines, Kriging, νευρωνικά δίκτυα), την επιλογή του πλήθους των δειγμάτων αλλά και της περιοχής από την οποία αυτά πρέπει να ληφθούν καθώς και στον τρόπο με τον οποίο ενημερώνουν την απαντητική επιφάνεια. Οι RSM προσεγγίσεις έχουν τη δυνατότητα είτε να χτίζουν υποκατάστατα μοντέλα που είναι αποτελεσματικά σε τοπικές περιοχές, και να χρησιμοποιούν ακολουθιακά αυτά τα μοντέλα για να καθοδηγήσουν την αναζήτηση, είτε να κατασκευάζουν υποκατάστατα μοντέλα για ολόκληρο τον παραμετρικό χώρο και στη συνέχεια να τα χρησιμοποιούν για την επιλογή δειγμάτων από τις περιοχές ενδιαφέροντος [5].

Αρχικά η μέθοδος RSM αποτελούνταν από δύο φάσεις οι οποίες περιγράφονται ως εξής: κατά τη Φάση I, χτίζονται μοντέλα πρώτης τάξης χρησιμοποιώντας δείγματα που προκύπτουν από το σχεδιασμό πειραμάτων. Χρησιμοποιείται ένας κανόνας για την μετακίνηση προς συγκεκριμένη κατεύθυνση, κάτι που συνεχίζεται επαναληπτικά μέχρι η εκτιμώμενη παράγωγος να πλησιάζει το μηδέν. Ακολουθεί η Φάση II κατά την οποία χτίζεται ένα πιο λεπτομερές τετραγωνικό μοντέλο που χρησιμοποιείται για την επαλήθευση της βελτιστότητας του πειραματικού σχεδιασμού [5].

2.4 Τυχαία Αναζήτηση

Οι μέθοδοι τυχαίας αναζήτησης (random search) αποτελούν ένα ιδιαίτερα χρήσιμο εργαλείο της βελτιστοποίησης μέσω προσομοίωσης καθώς είναι σχετικά εύκολες στην εφαρμογή και μπορούν να βρίσκουν «καλές» λύσεις σε σύντομο χρόνο. Συνδυάζουν επαναλαμβανόμενη δειγματοληψία και αξιολόγηση σημείων της εφικτής περιοχής κατά την αναζήτηση της βέλτιστης εφικτής λύσης. Η αξιολόγηση των υποψήφιων λύσεων γίνεται σύμφωνα με την παρατηρούμενη ιστορία, δηλαδή τις λύσεις που έχουν επιλεγεί κατά τη δειγματοληψία μέχρι την τρέχουσα χρονική στιγμή και την εκτιμημένη απόδοση στις λύσεις αυτές. Αντιμετωπίζουν με ιδιαίτερη επιτυχία προβλήματα βελτιστοποίησης μέσω προσομοίωσης για τα οποία η δομή της αντικειμενικής συνάρτησης είναι σε μεγάλο βαθμό άγνωστη και η διαδικασία της βελτιστοποίησης πρέπει επομένως να καθοδηγηθεί από εκτιμήσεις της συνάρτησης απόδοσης. Καλύπτουν ένα ευρύ πεδίο εφαρμογών που περιλαμβάνει προβλήματα βελτιστοποίησης τόσο ντετερμινιστικά όσο και στοχαστικά με διακριτές ή/και συνεχείς παραμέτρους. Ένα βασικό μειονέκτημά τους, σχετίζεται με την ισχυρή σύγκλιση σε ολικό βέλτιστο, καθώς για να εξασφαλιστεί απαιτούνται υποθέσεις πάνω στη δομή του προβλήματος [9].

Η πιο απλή μορφή τυχαίας αναζήτησης είναι η γνήσια τυχαία αναζήτηση (pure random search) ή τυφλή αναζήτηση (blind search). Σύμφωνα με αυτή τη μέθοδο γίνεται επαναλαμβανόμενα δειγματοληψία από την εφικτή περιοχή, χωρίς τη χρήση πληροφορίας που μπορεί να προκύψει κατά την πορεία του αλγορίθμου σχετικά με τα ιδιαίτερα χαρακτηριστικά του προβλήματος. Υπάρχουν διάφορες προσεγγίσεις σχετικά με τον τρόπο εκτίμησης της βέλτιστης λύσης μετά τη δειγματοληψία. Μια από αυτές, για παράδειγμα, λειτουργεί εκτιμώντας σε κάθε σημείο θ , που επιλέγεται κατά τη διαδικασία της δειγματοληψίας, την τιμή της αντικειμενικής συνάρτησης, ως το δειγματικό μέσο των παρατηρήσεων που βρίσκονται εντός μιας καθορισμένης απόστασης από το θ . Το σημείο με την καλύτερη εκτιμημένη τιμή της αντικειμενικής συνάρτησης επιλέγεται ως εκτίμηση της βέλτιστης λύσης [9].

Μπορεί ναδειχθεί ότι η γνήσια τυχαία αναζήτηση συγκλίνει σε μία λύση που βρίσκεται ε -κοντά στο ολικό βέλτιστο, με πιθανότητα ένα. Για πεπερασμένα προβλήματα αυτό το ε ισούται με μηδέν, ενώ για συνεχή προβλήματα το ε είναι μια μικρή θετική τιμή. Βέβαια κάτι τέτοιο μπορεί να απαιτεί μεγάλο αριθμό επαναλήψεων για να επιτευχθεί. Το αναμενόμενο πλήθος σημείων που πρέπει να εξεταστούν ώστε να φτάσουμε ε -κοντά στο βέλτιστο αυξάνεται εκθετικά με τη διάσταση του προβλήματος [9].

Έχουν αναπτυχθεί πολλοί αλγόριθμοι με αφετηρία τη γνήσια τυχαία αναζήτηση. Μια

τέτοια κατηγορία μεθόδων, που έχει προκύψει ως παραλλαγή της τελευταίας, αποτελούν οι μέθοδοι γνήσιας προσαρμοστικής αναζήτησης (pure adaptive search) [9].

Η γνήσια προσαρμοστική αναζήτηση αρχικά προτάθηκε για κυρτό προγραμματισμό και στη συνέχεια για προβλήματα με Lipschitz συνεχείς αντικειμενικές συναρτήσεις. Σε αντίθεση με τις παραπάνω μεθόδους όπου η δειγματοληψία πραγματοποιείται ανεξάρτητα και ισόνομα, κατά τη γνήσια προσαρμοστική αναζήτηση κάθε δείγμα εξαρτάται από το αμέσως προηγούμενό του, με την έννοια ότι κάθε νέα λύση που παράγεται πρέπει να είναι αυστηρά καλύτερη από την προηγούμενη. Η δειγματοληψία στην γνήσια προσαρμοστική αναζήτηση είναι πιο δύσκολη από την απλή τυχαία δειγματοληψία που εφαρμόζεται στην γνήσια τυχαία αναζήτηση, όμως η βελτίωση των επόμενων λύσεων οδηγεί σε γραμμική σύγκλιση του αλγορίθμου σε ϵ -βέλτιστη λύση [9].

3 Διακριτή Βελτιστοποίηση Μέσω Προσομοίωσης

Η διακριτή βελτιστοποίηση μέσω προσομοίωσης αναφέρεται σε προβλήματα με μεταβλητές απόφασης που παίρνουν μόνο διακριτές τιμές. Τα προβλήματα που εμπίπτουν στην κατηγορία της διακριτής βελτιστοποίησης μέσω προσομοίωσης χωρίζονται σε δύο βασικές υποκατηγορίες. Στα προβλήματα εκείνα των οποίων ο χώρος των εφικτών λύσεων είναι πεπερασμένος και μικρός, με την έννοια ότι περιέχει διαχειρίσιμο πλήθος στοιχείων, από υπολογιστικής άποψης, και σε εκείνα τα προβλήματα όπου ο χώρος των εφικτών λύσεων είναι ή άπειρος είτε μεγάλος [5].

Στην πρώτη περίπτωση απευθύνονται αλγόριθμοι που μπορεί να περιλαμβάνουν επιλογή της καλύτερης λύσης από ένα πεπερασμένο σύνολο εναλλακτικών, τη σύγκριση μέτρων απόδοσης της προσομοίωσης για κάθε μία από τις εναλλακτικές με μια γνωστή σταθερή τιμή ή την κατά ζεύγη σύγκριση όλων των υποψήφιων λύσεων. Πρωταρχικός στόχος αποτελεί η κατανομή των επαναλήψεων προσομοίωσης ανάμεσα στις εναλλακτικές. Καθώς το πλήθος των δυνατών λύσεων είναι μικρό δεν δίνεται ιδιαίτερη έμφαση στην αναζήτηση. Σκοπός είναι να προσδιοριστούν σε κάθε επανάληψη οι καλύτερες, με κάποια στατιστική έννοια, παράμετροι για την επόμενη προσομοίωση που θα εκτελεστεί [5].

Για τις δύο τελευταίες περιπτώσεις είναι απαραίτητοι αλγόριθμοι που περιλαμβάνουν συνιστώσα αναζήτησης. Πολλοί από αυτούς προκύπτουν, με κατάλληλες τροποποιήσεις, από αλγορίθμους που εφαρμόζονται στη συνεχή περίπτωση, όπως η διατακτική βελτιστοποίηση, οι μέθοδοι τυχαίας αναζήτησης και οι μέθοδοι άμεσης αναζήτησης [5].

3.1 Μικρό Πλήθος Εφικτών Λύσεων

3.1.1 Κατάταξη και Επιλογή

Η μέθοδος κατάταξης και επιλογής (ranking and selection) απευθύνεται σε προβλήματα με μικρό πλήθος εφικτών λύσεων, έστω k , όπου ισχύει για παράδειγμα $2 \leq k \leq 20$. Με τη χρήση προσομοίωσης καθορίζεται ποιες από αυτές τις k λύσεις είναι πιο πιθανό να βελτιστοποιούν την αντικειμενική συνάρτηση του προβλήματος.

Κατά την παραδοσιακή εφαρμογή της μεθόδου, σκοπός είναι η ελαχιστοποίηση του πλήθους των επαναλήψεων προσομοίωσης ενώ ταυτόχρονα εξασφαλίζεται μια συγκεκριμένη πιθανότητα σωστής επιλογής ανάμεσα στις εναλλακτικές. Οι περισσότερες διαδικασίες επιχειρούν να εξασφαλίσουν ότι το σημείο που επιλέγεται τελικά είναι καλύτερο από όλες τις εναλλακτικές κατά μία ποσότητα δ με πιθανότητα τουλάχιστον $1 - \alpha$. Το δ αποκαλείται ζώνη αδιαφορίας

και είναι η τιμή που θεωρείται επαρκής για να διαχωρίσει την αναμενόμενη απόδοση των υποψήφιων λύσεων. Γενικά καθώς αυξάνεται ο αριθμός των επαναλήψεων προσομοίωσης, η πιθανότητα σωστής επιλογής προσεγγίζει τη μονάδα. Έτσι, επιλέγοντας κατάλληλα το πλήθος των προσομοιώσεων που θα διεξαχθούν μπορούμε να εξασφαλίσουμε μια αποδεκτή πιθανότητα σωστής επιλογής. Σημειώνεται ότι για την εφαρμογή της μεθόδου οι επαναλήψεις προσομοίωσης πρέπει να είναι ανεξάρτητες [5, 11].

Μια εναλλακτική μορφοποίηση του προβλήματος κατάταξης και επιλογής αποτελεί η αναζήτηση της καλύτερης δυνατής λύσης, χωρίς όμως να ξεπεραστεί κάποιο προκαθορισμένο υπολογιστικό κόστος. Η προσέγγιση αυτή καλείται βέλτιστη κατανομή υπολογιστικών πόρων (Optimal Computing Budget Allocation-OCBA) και θα παρουσιαστεί στη συνέχεια.

3.1.2 Διαδικασίες Πολλαπλών Συγκρίσεων

Οι διαδικασίες πολλαπλών συγκρίσεων (multiple comparison procedures) προσεγγίζουν το πρόβλημα της βελτιστοποίησης ως πρόβλημα στατιστικής συμπερασματολογίας και δεν εγγυώνται την εύρεση βέλτιστης λύσης, σε αντίθεση με τις μεθόδους κατάταξης και επιλογής. Εδώ, για κάθε δυνατή λύση, πραγματοποιείται ένα πλήθος επαναλήψεων προσομοίωσης και η συμπερασματολογία βασίζεται σε διαστήματα εμπιστοσύνης της συνάρτησης απόδοσης. Στην πράξη χρησιμοποιούνται τρεις κλάσεις μεθόδων αυτής της κατηγορίας. Η πρώτη συγκρίνει ανά δύο μεταξύ τους όλες τις εναλλακτικές, η δεύτερη συγκρίνει όλες τις υποψήφιες λύσεις με την καλύτερη, ενώ η τρίτη με μία καθορισμένη τιμή ελέγχου [5, 11].

3.1.3 Βέλτιστη Κατανομή Υπολογιστικών Πόρων

Όπως έχει ήδη αναφερθεί, στην περίπτωση των προβλημάτων με πεπερασμένο και σχετικά μικρό σύνολο Θ απευθύνονται προσεγγίσεις που στοχεύουν στην επιλογή της καλύτερης λύσης ανάμεσα σε ένα σταθερό σύνολο εναλλακτικών, με τον έλεγχο κάθε μίας από αυτές. Η πρόταση αυτή είναι εφικτή ακριβώς λόγω του περιορισμένου πλήθους εφικτών λύσεων. Ένα βασικό ζήτημα αποτελεί η αποδοτική κατανομή των επαναλήψεων προσομοίωσης ανάμεσα στις διάφορες υποψήφιες λύσεις, έτσι ώστε να μην σπαταληθούν άσκοπα υπολογιστικοί πόροι και αυτό ακριβώς πραγματεύεται η βέλτιστη κατανομή υπολογιστικών πόρων. Η πιο απλή κατανομή είναι εκείνη που αποδίδει ίσο αριθμό επαναλήψεων προσομοίωσης σε όλες τις υποψήφιες λύσεις, μπορεί, όμως, σε αρκετά προβλήματα να αποδειχθεί αναποτελεσματική. Στη θέση της, η βέλτιστη κατανομή υπολογιστικών πόρων προτείνει την αξιοποίηση πληροφορίας από το μέσο και τη διασπορά κάθε υποψήφιας λύσης και μάλιστα συνδυάζοντας τις πληροφορίες αυτές,

έτσι ώστε σε λύσεις με μεγαλύτερες διασπορές ή/και μικρότερους μέσους να αντιστοιχούν περισσότερες επαναλήψεις. Έτσι για παράδειγμα αν μία υποψήφια λύση διακρίνεται από μικρή διασπορά τότε μπορεί να είναι επαρκής ένας μικρός αριθμός επαναλήψεων προσομοίωσης για την εκτίμηση της τιμής της αντικειμενικής συνάρτησης στη λύση αυτή. Από την άλλη αν κάποια εναλλακτική έχει πολύ μεγάλο μέσο (για ένα πρόβλημα ελαχιστοποίησης) τότε μπορεί να θεωρηθεί μη ανταγωνιστική κι επομένως να μην σπαταληθούν υπολογιστικοί πόροι για αυτή. Η συγκεκριμένη μέθοδος λειτουργεί μεγιστοποιώντας μια προσέγγιση της πιθανότητας σωστής επιλογής της βέλτιστης λύσης, οδηγώντας σε έναν αποδοτικό αλγόριθμο που συμπεριλαμβάνει τόσο μέσους όσο και διασπορές [7].

3.2 Μεγάλο Πλήθος Εφικτών Λύσεων

3.2.1 Προσέγγιση Δειγματικού Μέσου

Οι μέθοδοι προσέγγισης δειγματικού μέσου (sample average approximation) στοχεύουν στη βελτιστοποίηση μιας εκτίμησης της υποκειμενικής αντικειμενικής συνάρτησης, αντί να επιχειρούν τη βελτιστοποίηση της συνάρτησης αυτής καθεαυτήν. Η εκτίμηση γίνεται συνήθως μέσω κάποιας συνεπούς εκτιμήτριας, όπως ο δειγματικός μέσος ανεξάρτητων πραγματοποιήσεων της συνάρτησης σε συγκεκριμένο σημείο. Για παράδειγμα, αντί για την συνάρτηση ενδιαφέροντος $J(\theta) = \mathbb{E}[f(\theta, \omega)]$, χρησιμοποιείται η εκτίμησή της, $\hat{J}_N(\theta) = \frac{1}{N} \sum_{i=1}^N f(\theta, \omega^i)$. Έτσι, καθώς επιχειρείται βελτιστοποίηση της εκτιμημένης συνάρτησης, επιστρατεύονται εργαλεία παραδοσιακού μαθηματικού προγραμματισμού για την επίλυση του προβλήματος [5].

Στη συνέχεια γίνεται παρουσίαση της μεθόδου προσέγγισης μέσω δειγματικού μέσου για προβλήματα στοχαστικής διακριτής βελτιστοποίησης όπως αυτή περιγράφεται στο [12]. Αρχικά το πρόβλημα ενδιαφέροντος, και εδώ, είναι το ακόλουθο:

$$\min_{\theta \in \Theta} \{J(\theta) := \mathbb{E}_\omega[f(\theta, \omega)]\} \quad (2)$$

όπου ω τυχαίο διάνυσμα με κατανομή P , Θ ένα πεπερασμένο σύνολο, $f(\theta, \omega)$ μια πραγματική συνάρτηση και $\mathbb{E}_\omega[f(\theta, \omega)]$ η αντίστοιχη μέση τιμή. Η $J(\theta)$ θεωρείται ότι είναι καλά ορισμένη, δηλαδή $\forall \theta \in \Theta$ η συνάρτηση $f(\theta, \cdot)$ είναι μετρήσιμη και $\mathbb{E}_\omega\{|f(\theta, \omega)|\} < \infty$. Τα προβλήματα στα οποία απευθύνεται η παρούσα μέθοδος είναι προβλήματα για τα οποία η $J(\theta)$ δεν μπορεί να γραφεί σε κλειστή μορφή ή/και παρουσιάζεται δυσκολία στον υπολογισμό των τιμών της, η $f(\theta, \omega)$ μπορεί να υπολογιστεί εύκολα για γνωστά θ και ω και το πλήθος των εφικτών λύσεων, αν και πεπερασμένο, είναι μεγάλο. Η κεντρική ιδέα, που βασίζεται στην προσομοίωση Monte

Carlo, είναι η γέννηση τυχαίου δείγματος από την ω και η προσέγγιση της μέσης τιμής που εμφανίζεται στο πρόβλημα από την συνάρτηση του δειγματικού μέσου

$$\hat{J}_N(\theta) := \frac{1}{N} \sum_{j=1}^N f(\theta, \omega^j).$$

Έστω ανεξάρτητο και ισόνομο τυχαίο δείγμα $\omega^1, \omega^2, \dots, \omega^N$, N πραγματοποιήσεων του τυχαίου διανύσματος ω . Θεωρείται το πρόβλημα

$$\min_{\theta \in \Theta} \hat{J}_N(\theta). \quad (3)$$

Το πρόβλημα (2) θα αναφέρεται ως το πραγματικό πρόβλημα ενώ το πρόβλημα (3) ως προσέγγιση δειγματικού μέσου ή εν συντομία SAA (Sample Average Approximation). Με Θ^* και $\hat{\Theta}_N$ συμβολίζονται τα σύνολα βέλτιστων λύσεων για τα (2) και (3), αντίστοιχα, ενώ με

$$v^* := \min_{\theta \in \Theta} J(\theta) \quad \text{και} \quad \hat{v}_N := \min_{\theta \in \Theta} \hat{J}_N(\theta)$$

οι αντίστοιχες βέλτιστες τιμές. Ακόμα, για $\varepsilon > 0$ λέμε ότι το $\bar{\theta}$ είναι ε -βέλτιστη λύση του (2) αν $\bar{\theta} \in \Theta$ και $g(\bar{\theta}) \leq v^* + \varepsilon$. Ανάλογα ορίζεται η ε -βέλτιστη λύση του (3), ενώ με Θ^ε και $\hat{\Theta}_N^\varepsilon$ συμβολίζονται τα αντίστοιχα σύνολα ε -βέλτιστων λύσεων [12].

Οι Klewegt, Shapiro, and Homem-de Mello [12] έχουν αποδείξει τη σύγκλιση της βέλτιστης τιμής του προβλήματος SAA σε αυτή του πραγματικού με πιθανότητα 1 καθώς $N \rightarrow \infty$, όπως και ότι το ενδεχόμενο $\{\hat{\Theta}_N^\varepsilon \subset \Theta^\varepsilon\}$ πραγματοποιείται με πιθανότητα 1 $\forall \varepsilon \geq 0$ όταν $N \rightarrow \infty$.

Επιπλέον, άμεσα αποδεικνύεται ότι $\forall \delta \in [0, \varepsilon]$ το ενδεχόμενο $\{\hat{\Theta}_N^\delta \subset \Theta^\varepsilon\}$ συμβαίνει με πιθανότητα 1 καθώς $N \rightarrow \infty$. Από το τελευταίο συμπεραίνουμε ότι αν το πραγματικό πρόβλημα έχει μοναδική βέλτιστη λύση θ^* τότε με πιθανότητα 1 το SAA έχει μοναδική βέλτιστη λύση $\hat{\theta}_N$, με $\hat{\theta}_N = \theta^*$, για $N \rightarrow \infty$ [12].

Σχετικά με τους ρυθμούς των παραπάνω συγκλίσεων, αποδεικνύεται ότι για κάθε $\varepsilon \geq 0$ και $\delta \in [0, \varepsilon]$ υπάρχει σταθερά $\gamma(\delta, \varepsilon) \geq 0$ τέτοια ώστε

$$\mathbb{P}(\hat{\Theta}_N^\delta \not\subset \Theta^\varepsilon) \leq |\Theta \setminus \Theta^\varepsilon| e^{-N\gamma(\delta, \varepsilon)}.$$

Η τελευταία ανισότητα ισχύει για κάθε τυχαίο δείγμα μεγέθους N και με βάση αυτή αποδει-

κνύεται άμεσα η σχέση

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \left[1 - P(\hat{\Theta}_N^\delta \subset \Theta^\varepsilon) \right] \leq -\gamma(\delta, \varepsilon). \quad (4)$$

Δηλαδή η πιθανότητα του ενδεχομένου $\{\hat{\Theta}_N^\delta \subset \Theta^\varepsilon\}$ προσεγγίζει το 1 με εκθετική ταχύτητα καθώς $N \rightarrow \infty$. Η μέθοδος Monte Carlo, επομένως, σε συνδυασμό με μια αποτελεσματική μέθοδο επίλυσης για το ντετερμινιστικό SAA μπορεί να λύσει τον τύπο των προβλημάτων ενδιαφέροντος, με την προϋπόθεση ότι η σταθερά $\gamma(\delta, \varepsilon)$ δεν είναι πολύ μικρή [12].

Στη συνέχεια εξετάζονται οι συνέπειες του τελευταίου φράγματος σε θέματα πολυπλοκότητας. Από την ανισότητα (4) υπολογίζεται ότι το απαιτούμενο μέγεθος δείγματος N ώστε να ισχύει $\mathbb{P}(\hat{\Theta}_N^\delta \subset \Theta^\varepsilon) \geq 1 - \alpha$ σε επιλεγμένο επίπεδο στατιστικής σημαντικότητας $\alpha \in (0, 1)$ είναι

$$N \geq \frac{1}{\gamma(\delta, \varepsilon)} \log \left(\frac{|\Theta \setminus \Theta^\varepsilon|}{\alpha} \right). \quad (5)$$

[12] Ακόμα, έχει αποδειχτεί ότι $\gamma(\delta, \varepsilon) \geq (\varepsilon - \delta)^2 / (3\sigma_{\max}^2)$, για θετική σταθερά σ_{\max}^2 , επομένως για $\varepsilon \geq 0$ αρκετά μικρό και $\delta \in [0, \varepsilon)$ μια ικανή συνθήκη για την (5) είναι

$$N \geq \frac{3\sigma_{\max}^2}{(\varepsilon - \delta)^2} \log \left(\frac{|\Theta \setminus \Theta^\varepsilon|}{\alpha} \right). \quad (6)$$

Επιπλέον, στην περίπτωση που το σύνολο βέλτιστων λύσεων είναι μονοσύνολο, $\Theta = \{\theta^*\}$, το απαιτούμενο μέγεθος δείγματος για τον ακριβή υπολογισμό της θ^* με μεγάλη πιθανότητα μπορεί να είναι πολύ μεγάλο ακόμα κι αν αυτή είναι μοναδική, γεγονός που καθιστά φανερή την ανάγκη αναζήτησης ε -βέλτιστων λύσεων [12].

Ακολουθεί μια αναφορά στα ασυμπτωτικά αποτελέσματα για τη βέλτιστη τιμή της αντικειμενικής συνάρτησης του SAA, \hat{v}_N . Συγκεκριμένα έχουν αποδειχθεί τα ακόλουθα: υπό την προϋπόθεση ότι οι διασπορές $\sigma^2(\theta) := \text{Var}[f(\theta, \omega)]$ υπάρχουν $\forall \theta \in \Theta$ προκύπτει από το Κεντρικό Οριακό Θεώρημα ότι

$$\sqrt{N}[\hat{J}_N(\theta) - J(\theta)] \xrightarrow{d} Z(\theta),$$

$\forall \theta \in \Theta$, όπου $Z(\theta) \sim \mathcal{N}(0, \sigma^2(\theta))$. Έτσι προκύπτει ότι $\forall \theta \in \Theta^*$

$$\sqrt{N}(\hat{v}_N - v^*) \xrightarrow{d} \min_{\theta \in \Theta^*} Z(\theta),$$

ενώ μπορεί ακόμα να δειχθεί ότι $\sqrt{N}[\mathbb{E}(\hat{v}_N) - v^*] \rightarrow \mathbb{E}[\min_{\theta \in \Theta^*} Z(\theta)]$, κάτω από επιπλέον

προϋποθέσεις [12].

Σχετικά με τη διαδικασία επιλογής κατάλληλου μεγέθους δείγματος N και κατάλληλου αριθμού επαναλήψεων, είναι αρχικά σημαντικό να αναφερθεί ο συμβιβασμός ανάμεσα σε ακρίβεια και υπολογιστική πολυπλοκότητα που λαμβάνει χώρα. Γενικά, η διαδικασία εκτίμησης της τιμής της αντικειμενικής συνάρτησης $J(\theta)$ σε συγκεκριμένο σημείο $\theta \in \Theta$ από το δειγματικό μέσο $\hat{J}_N(\theta)$ είναι λιγότερο απαιτητική από υπολογιστικής άποψης σε σχέση με τη λύση του αντίστοιχου SAA. Αυτή η παρατήρηση δίνει κίνητρο για την επιλογή μικρού μεγέθους δείγματος N για την επίλυση του SAA και μεγαλύτερου μεγέθους δείγματος N' για την εκτίμηση της $J(\hat{\theta}_\Theta)$ από της $\hat{J}_{N'}(\hat{\theta}_N)$ στη βέλτιστη λύση $\hat{\theta}_N$ του SAA. Όσον αφορά στην ακρίβεια της παραγόμενης λύσης, μπορεί να μετρηθεί από την αντίστοιχη δειγματική διακύμανση $S_{N'}^2(\hat{\theta}_N)/N'$ η οποία μπορεί να εκτιμηθεί από το ίδιο δείγμα [12].

Στην περίπτωση, όμως, που η υπολογιστική πολυπλοκότητα επίλυσης του SAA αυξάνει γρηγορότερα από γραμμικά, μπορεί να είναι αποτελεσματικότερη η επιλογή μικρότερου μεγέθους δείγματος N και η προσομοίωση και λύση αρκετών SAA με ανεξάρτητα και ισόνομα δείγματα. Στο σημείο αυτό, βέβαια, τίθενται θέματα καταλληλότητας της παραγόμενης λύσης. Έτσι, αν p η πιθανότητα η βέλτιστη λύση για ένα από τα SAA να είναι και βέλτιστη για το πραγματικό πρόβλημα, η πιθανότητα να παραχθεί η βέλτιστη λύση του πραγματικού τουλάχιστον μία φορά στις M δοκιμές είναι $1 - (1 - p)^M$, που τείνει στο 1 καθώς $M \rightarrow \infty$, με την προϋπόθεση, όμως, ότι $p > 0$. Μέσα από κατάλληλο παράδειγμα, μπορεί ναδειχθεί πως το απαιτούμενο μέγεθος δείγματος N για να εξασφαλιστεί ότι $p > 0$, εξαρτάται από το πρόβλημα, δεν μπορεί να φραχθεί από μία συνάρτηση που εξαρτάται μόνο από το πλήθος των εφικτών λύσεων και μπορεί να είναι αυθαίρετα μεγάλο. Δεν υπάρχει επομένως εγγύηση πως με τη διαδικασία αυτή θα παραχθεί λύση που θα είναι βέλτιστη για το πραγματικό πρόβλημα [12].

Ένα ακόμα θέμα που προκύπτει είναι η επιλογή του πλήθους M των προβλημάτων SAA που θα επιλυθούν, η οποία μπορεί να γίνει και δυναμικά. Ένα κριτήριο τερματισμού, που μπορεί να χρησιμοποιηθεί, βασίζεται σε μία εκτίμηση του χάσματος βελτιστότητας, $J(\hat{\theta}_N^m) - v^*$, όπου $\hat{\theta}_N^m$ η υποψήφια λύση που παράγει το m -οστό SAA. Αν το κριτήριο αυτό ικανοποιηθεί η διαδικασία σταματάει, διαφορετικά ή επιλύονται νέα SAA με το ίδιο μέγεθος δείγματος N , είτε επιλέγεται νέο, μεγαλύτερο μέγεθος δείγματος. Το ποια από τις δύο επιλογές είναι πιθανότερο να δώσει καλύτερη εκτίμηση της ζητούμενης λύσης μπορεί να υπολογιστεί όπως παρουσιάζεται στη συνέχεια. Αρχικά, οι τυχαίες μεταβλητές $J(\hat{\theta}_N^m)$, $m = 1, 2, \dots$, είναι ανεξάρτητες και ισόνομες με πεπερασμένο στήριγμα, αφού το Θ είναι πεπερασμένο. Αν υποθέσουμε ότι έχουν υλοποιηθεί μέχρι στιγμής M προβλήματα SAA με μέγεθος δείγματος N και αν επιπλέον η

κατανομή της $J(\hat{\theta}_N)$ ήταν συνεχής, είναι γνωστό ότι η πιθανότητα η λύση που θα παραχθεί από το $(M+1)$ -οστό SAA με το ίδιο μέγεθος δείγματος, να είναι καλύτερη από τη βέλτιστη μέχρι στιγμής λύση, θα ήταν $1/(M+1)$. Επειδή, όμως η κατανομή της $J(\hat{\theta}_N)$ είναι διακριτή, η πιθανότητα αυτή θα είναι μικρότερη ή ίση του $1/(M+1)$. Επομένως, όταν το $1/(M+1)$ γίνει επαρκώς μικρό επιλέγεται αύξηση μεγέθους δείγματος του επόμενου SAA ή τερματισμός της διαδικασίας [12].

Έχει ήδη γίνει φανερό η ανάγκη εκτίμησης του χάσματος βελτιστότητας $J(\hat{\theta}) - v^*$ για μια συγκεκριμένη λύση $\hat{\theta} \in \Theta$. Η

$$\hat{J}_{N'}(\hat{\theta}) := \frac{1}{N'} \sum_{j=1}^{N'} f(\hat{\theta}, \omega^j) \quad (7)$$

αποτελεί αμερόληπτη εκτιμήτρια του $J(\hat{\theta})$ με εκτιμώμενη διακύμανση $S_{N'}^2(\hat{\theta})/N'$, όπου $S_{N'}^2(\hat{\theta})$ είναι η δειγματική διασπορά του $f(\hat{\theta}, \omega^j)$. Ακόμα, η

$$\bar{v}_N^M := \frac{1}{M} \sum_{m=1}^M \hat{v}_N^m, \quad (8)$$

είναι μια εκτιμήτρια του v^* η οποία μάλιστα έχει την ίδια αρνητική μεροληψία με την \hat{v}_N καθώς $\mathbb{E}[\bar{v}_N^M] = \mathbb{E}[\hat{v}_N]$. Μια εκτιμήτρια του χάσματος βελτιστότητας, επομένως, αποτελεί η $\hat{J}_{N'}(\hat{\theta}) - \bar{v}_N^M$, η οποία το υπερεκτιμά, αφού

$$\mathbb{E}[\hat{J}_{N'}(\hat{\theta}) - \bar{v}_N^M] = J(\hat{\theta}) - \mathbb{E}[\hat{\theta}_N] \geq J(\hat{\theta}) - v^*. \quad (9)$$

Μπορεί να δειχθεί σχετικά ότι η μεροληψία της \hat{v}_N είναι φθίνουσα ως προς το μέγεθος δείγματος N . Η διασπορά της \bar{v}_N^M δίνεται από τη σχέση

$$\frac{S_M^2}{M} = \frac{1}{M(M-1)} \sum_{m=1}^M (\hat{v}_N^m - \bar{v}_N^M)^2 \quad (10)$$

και έτσι αν τα M δείγματα μεγέθους N και το δείγμα εκτίμησης N' είναι ανεξάρτητα τότε η διασπορά της εκτιμήτριας $\hat{J}_{N'}(\hat{\theta}) - \bar{v}_N^M$ μπορεί να εκτιμηθεί από την $S_{N'}^2(\hat{\theta})/N' + S_M^2/M$. Μια άλλη εκτιμήτρια του χάσματος βελτιστότητας, με πιθανώς μικρότερη διασπορά, είναι η $\bar{J}_N^M(\hat{\theta}) - \bar{v}_N^M$, όπου

$$\bar{J}_N^M(\hat{\theta}) := \frac{1}{M} \sum_{m=1}^M \hat{J}_N^m(\hat{\theta}) \quad (11)$$

με

$$\hat{J}_N^m(\hat{\theta}) := \frac{1}{N} \sum_{j=1}^N f(\hat{\theta}, \omega^j), \quad (12)$$

που έχει εκτιμώμενη διασπορά

$$\frac{\bar{S}_M^2}{M} = \frac{1}{M(M-1)} \sum_{m=1}^M [(\hat{J}_N^m(\hat{\theta}) - \hat{v}_N^m) - (\bar{J}_N^m - \bar{v}_N^M)]^2. \quad (13)$$

Το ποια εκτιμήτρια έχει τη μικρότερη διασπορά εξαρτάται από τη συσχέτιση των $\hat{J}_N^m(\hat{\theta})$ και \hat{v}_N^m και τα μεγέθη N , N' , M . Πρέπει, όμως, επιπλέον να ληφθεί υπόψη και η υπολογιστική προσπάθεια που απαιτείται για την αξιολόγηση της κάθε εκτιμήτριας. Σε κάθε περίπτωση η ακρίβεια μιας εκτιμήτριας μπορεί να συνυπολογιστεί με χρήση του Κεντρικού Οριακού Θεωρήματος. Ακόμα, με τη βοήθεια του τελευταίου γίνεται εμφανές ένα μειονέκτημα που ακολουθεί τη χρήση των χασμάτων βελτιστότητας στην αξιολόγηση υποψήφιων λύσεων, ότι η εκτιμήτρια του χάσματος μπορεί να είναι μεγάλη αν κάποιο από τα N , N' , ή M είναι μικρά, ακόμα και αν η εξεταζόμενη λύση είναι βέλτιστη [12].

Αν υποθέσουμε τώρα ότι η παραπάνω διαδικασία έχει τερματιστεί, ως λύση θα μπορούσε να επιλεγεί η υποψήφια λύση $\hat{\theta} \in \Theta$ που αντιστοιχεί στην καλύτερη τιμή της $\hat{J}_{N'}(\hat{\theta})$. Μπορεί, όμως, να έχει νόημα μια περαιτέρω αξιολόγηση της παραγόμενης λύσης για την οποία αρκετές μέθοδοι μπορούν να αναζητηθούν στην υπάρχουσα βιβλιογραφία [12].

Ακολουθεί η παρουσίαση του προτεινόμενου αλγορίθμου.

1. Επιλογή αρχικών μεγεθών δείγματος N και N' , επιλογή ενός κανόνα απόφασης για τον καθορισμό του πλήθους M των προβλημάτων SAA (ενδεχομένως με χρήση ενός μέγιστου πλήθους M' προβλημάτων SAA που θα χρησιμοποιούν το ίδιο μέγεθος δείγματος έτσι ώστε το $1/(M' + 1)$ να είναι επαρκώς μικρό), ενός κανόνα απόφασης για αύξηση των N και N' αν χρειαστεί και μιας επιθυμητής ακρίβειας ε .
2. Για $m = 1, \dots, M$ επανάληψη των υποβλημάτων (i) ως (iii).
 - (i) Δημιουργία δείγματος μεγέθους N και επίλυση του προβλήματος SAA (3) με βέλτιστη τιμή \hat{v}_N^m και ε -βέλτιστη λύση $\hat{\theta}_N^m$.
 - (ii) Εκτίμηση του χάσματος βελτιστότητας $J(\hat{\theta}_N^m) - v^*$ και της διασποράς της εκτιμήτριας του χάσματος.

- (iii) Αν το χάσμα βελτιστότητας ή η διασπορά του είναι επαρκώς μικρά, μετάβαση στο βήμα 4.
- 3. Αν το χάσμα βελτιστότητας ή η διασπορά του είναι πολύ μεγάλα, αύξηση των μεγεθών δείγματος N και/ή N' , και επιστροφή στο βήμα 2.
- 4. Επιλογή της καλύτερης λύσης $\hat{\theta}$ από τις υποψήφιες λύσεις θ_N^m , με χρήση μιας διαδικασίας διαλογής και επιλογής. Τερματισμός.

3.2.2 Διατακτική Βελτιστοποίηση

Οι μέθοδοι διατακτικής βελτιστοποίησης (ordinal optimization) είναι κατάλληλες σε περιπτώσεις που το πλήθος εναλλακτικών είναι πολύ μεγάλο για να βρεθεί το ολικό βέλτιστο κάτω από το πρίσμα της διακριτής βελτιστοποίησης. Σκοπός αυτών των μεθόδων είναι η εύρεση μιας ικανοποιητικής λύσης με κάποια εγγύηση για την ποιότητά της. Η προσπάθεια εδώ εστιάζεται στη δειγματοληψία ενός δεδομένου υποσυνόλου S της εφικτής περιοχής Θ , και ο υπολογισμός των τιμών της αντικειμενικής συνάρτησης στα σημεία που έχουν επιλεγεί προκειμένου να καθοριστεί το καλύτερο από αυτά. Η ποιότητα ή αλλιώς το επίπεδο ικανοποίησης του επιλεγμένου συνόλου μπορεί να ποσοτικοποιηθεί. Συγκεκριμένα ποσότητα ενδιαφέροντος αποτελεί η πιθανότητα $P(|T \cap S| \geq l)$ (alignment probability), όπου T είναι ένα σύνολο επαρκώς καλών λύσεων, που συχνά είναι οι t καλύτερες λύσεις του Θ . Μας ενδιαφέρει δηλαδή η πιθανότητα τουλάχιστον l από τις λύσεις στο T να ανήκουν και στο S . [5, 9].

Η διατακτική βελτιστοποίηση βασίζεται, λοιπόν, σε δύο κεντρικές ιδέες. Πρώτον, ότι η εκτίμηση μιας διάταξης και η σύγκριση ανάμεσα στις υποψήφιες λύσεις είναι πολύ λιγότερο απαιτητική διαδικασία από την εκτίμηση των ίδιων των τιμών της αντικειμενικής συνάρτησης σε κάθε λύση, και δεύτερον, στην ελάττωση του υπολογιστικού φορτίου που συνεπάγεται η χαλάρωση του στόχου βελτιστοποίησης με την αποδοχή επαρκώς καλών λύσεων [9].

3.2.3 Άμεση Αναζήτηση

Ο όρος άμεση αναζήτηση (direct search) εισήχθη για πρώτη φορά το 1961 από τους Hooke and Jeeves. Αναφέρεται σε μια κατηγορία μεθόδων βελτιστοποίησης μέσω προσομοίωσης, ελεύθερων παραγώγων, που βασίζονται στη σχετική διάταξη των τιμών της αντικειμενικής συνάρτησης στις δοκιμαστικές λύσεις και όχι στις ίδιες τις τιμές [13].

Στην αρχική μέθοδο που προτάθηκε από τους Hooke and Jeeves κάθε δοκιμαστική λύση συγκρίνεται με την καλύτερη μέχρι εκείνη τη στιγμή. Αρχικά, θεωρείται το σύνολο σημείων P

που αναπαριστούν τις δυνατές λύσεις αλλά και ένας κανόνας διάκρισης για το πότε μία λύση, P_1 , είναι καλύτερη από μια άλλη, P_2 , (συμβ. $P_1 \subset P_2$) για οποιαδήποτε δύο σημεία του χώρου. Γίνεται ακόμα, η υποθεση πως υπάρχει μοναδικό σημείο P^* με την ιδιότητα $P^* \subset P, \forall P \neq P^*$ [14].

Ακολουθώντας επιλέγεται αυθαίρετα σημείο B_0 ως το πρώτο «σημείο βάσης». Επιλέγεται δεύτερο σημείο P_1 και συγκρίνεται με το B_0 , αν $P_1 \subset B_0$, τότε το P_1 γίνεται το δεύτερο σημείο βάσης, B_1 , διαφορετικά $B_1 = B_0$. Η διαδικασία συνεχίζεται με τη σύγκριση κάθε σημείου με το τρέχον σημείο βάσης. Η στρατηγική επιλογής νέων σημείων καθορίζεται από ένα σύνολο καταστάσεων, πεπερασμένων το πλήθος. Υπάρχει ένα αυθαίρετα ορισμένο αρχικό σύνολο καταστάσεων και μια τελική κατάσταση που τερματίζει την αναζήτηση. Οι υπόλοιπες καταστάσεις αναπαριστούν διάφορες συνθήκες που προκύπτουν κατά την αναζήτηση και εξαρτώνται από τις επιλογές που γίνονται κατά την πορεία της αναζήτησης. Η επιλογή της στρατηγικής υπαγορεύεται από το ίδιο το πρόβλημα, μέσω της ιδιαίτερης δομής του χώρου των λύσεων ή άλλων παραμέτρων που το χαρακτηρίζουν. Κομμάτι της στρατηγικής αυτής αποτελούν, μεταξύ άλλων, η επιλογή των B_0 και S_0 , οι κανόνες μετάβασης ανάμεσα στις καταστάσεις και οι κανόνες που καθορίζουν τα σημεία που θα επιλέγονται κατά την αναζήτηση δεδομένης της τρέχουσας κατάστασης και του σημείου βάσης [14].

Οι μέθοδοι άμεσης αναζήτησης διαθέτουν κάποια βασικά πλεονεκτήματα τα οποία τις κάνουν να ξεχωρίζουν ακόμα και μπροστά σε πιο εξεζητημένες μεθόδους. Πρώτα και κύρια, όπως επισημαίνουν οι Lewis, Torczon, and Trosset [13] οι μέθοδοι άμεσης αναζήτησης είναι ιδιαίτερα δημοφιλείς καθώς δίνουν καλά αποτελέσματα στην πράξη. Για πολλές μεθόδους έχουν αποδειχθεί αποτελέσματα εγγυημένης σύγκλισης σε ολικό βέλτιστο ανάλογα με γνωστά αποτελέσματα για ολικές μεθόδους quasi-Newton. Επιπλέον, είναι εύκολες άμεσες και σαφείς στην εφαρμογή τους, είναι εφαρμόσιμες για μεγάλο πλήθος προβλημάτων με μικρές απαιτήσεις από τον χρήστη και μικρό πλήθος παραμέτρων. Αυτή η απλότητα, υπό μία έννοια, που τις χαρακτηρίζει, καθιστά τις μεθόδους άμεσης αναζήτησης ιδιαίτερα κατάλληλες για την ενσωμάτωσή τους σε υβριδικούς αλγορίθμους, συνήθως για την εύρεση κάποιου καλού σημείου εκκίνησης, αλλά όχι μόνο.

3.2.4 Τυχαία Αναζήτηση

Όπως έχει ήδη αναφερθεί έχουν αναπτυχθεί μέθοδοι τυχαίας αναζήτησης για την αντιμετώπιση προβλημάτων βελτιστοποίησης με συνεχείς αλλά και διακριτές μεταβλητές. Στις μεθόδους τυχαίας αναζήτησης για διακριτή βελτιστοποίηση μέσω προσομοίωσης κατατάσσεται

πλήθος τεχνικών όπως η προσομοιωμένη εξάχνωση, οι γενετικοί αλγόριθμοι, ο στοχαστικός κανόνας (stochastic ruler), η στοχαστική σύγκριση (stochastic comparison), οι εμφωλευμένες διαμερίσεις, η βελτιστοποίηση με αποικίες (ant colony optimization) και η tabu αναζήτηση. Για περισσότερους από τους παραπάνω αλγορίθμους είναι διαθέσιμες αποδείξεις για σύγκλιση σε κάποιο ολικό ή τοπικό βέλτιστο [5].

Γενετικοί Αλγόριθμοι

Οι γενετικοί αλγόριθμοι (genetic algorithms), που αναπτύχθηκαν πρώτη φορά από τον John Holland τη δεκαετία του 1960, είναι αλγόριθμοι τυχαίας αναζήτησης βασισμένοι στη φυσική επιλογή και τους νόμους της γενετικής. Σε γενικές γραμμές, ένας γενετικός αλγόριθμος λειτουργεί δημιουργώντας έναν πληθυσμό από σειρές χαρακτήρων, κάθε μία από τις οποίες καλείται χρωμόσωμα. Κάθε ακολουθία χρωμοσωμάτων είναι στην ουσία ένα διάνυσμα σημείων στο χώρο αναζήτησης. Νέα χρωμοσώματα δημιουργούνται μέσω επιλογής, μετάλλαξης και συναρτήσεων διασταύρωσης. Η διαδικασία επιλογής καθοδηγείται από τον υπολογισμό της αντικειμενικής συνάρτησης για κάθε χρωμόσωμα. Στη συνέχεια νέα χρωμοσώματα δημιουργούνται μέσω των συναρτήσεων διασταύρωσης και μετάλλαξης. Οι συναρτήσεις αυτές εξασφαλίζουν την διατήρηση της ποικιλομορφίας ανάμεσα στις παραγόμενες λύσεις. Οι γενετικοί αλγόριθμοι είναι ιδιαίτερα δημοφιλείς καθώς χαρακτηρίζονται από εύκολη εφαρμογή και χρησιμοποιούνται από αρκετά εμπορικά πακέτα λογισμικού για τη βελτιστοποίηση μέσω προσομοίωσης [5, 15].

Προσομοιωμένη Εξάχνωση

Η προσομοιωμένη εξάχνωση (simulated annealing) υλοποιεί μια πιθανοτική μέθοδο που πηγάζει από την φυσική διαδικασία της εξάχνωσης κατά την οποία ένα υλικό ψύχεται αργά, έτσι ώστε, ενώ η δομή του ψύχεται, εκείνο φτάνει μία κατάσταση ελάχιστης ενέργειας. Ξεκινώντας από ένα συγκεκριμένο σημείο i σε μία κατάσταση j , παράγεται ένα γειτονικό σημείο i' . Ο αλγόριθμος κινείται από το σημείο i στο i' χρησιμοποιώντας ένα πιθανοτικό κριτήριο που εξαρτάται από τη «θερμοκρασία» στην κατάσταση j . Αυτή η θερμοκρασία αποτελεί το ανάλογο της θερμοκρασίας στην φυσική εξάχνωση, και στον αλγόριθμο λειτουργεί ως παράμετρος ελέγχου. Αν η λύση στο i' είναι καλύτερη από την υπάρχουσα λύση, τότε το νέο σημείο γίνεται δεκτό. Αν η νέα λύση είναι χειρότερη από την υπάρχουσα λύση, τότε η πιθανότητα αποδοχής της νέα λύσης ορίζεται ως $\exp\{-(f(i') - f(i))/T(j)\}$, όπου $f(\cdot)$ είναι η τιμή της αντικειμενικής συνάρτησης σε δοθέν σημείο, και $T(j)$ η θερμοκρασία στη κατάσταση j . Αφού

αξιολογηθεί ένα πλήθος γειτονικών σημείων, η θερμοκρασία μειώνεται και δημιουργείται η νέα κατάσταση $j+1$. Λόγω της εκθετικής μορφής, η πιθανότητα αποδοχής ενός γειτονικού σημείου είναι υψηλότερη σε υψηλές θερμοκρασίες και χαμηλότερη καθώς η θερμοκρασία μειώνεται. Με αυτό τον τρόπο ο αλγόριθμος εξερευνά ένα μεγάλο πλήθος γειτονικών σημείων στη αρχή, αλλά καθώς η θερμοκρασία μειώνεται, ελαττώνεται και το πλήθος των σημείων που εξετάζονται [5].

Αναζήτηση Tabu

Η αναζήτηση tabu αξιοποιεί ειδικές δομές μνήμης κατά τη διάρκεια της διαδικασίας αναζήτησης που επιτρέπουν στη μέθοδο να ξεφεύγει από τα όρια της τοπικής βελτιστοποίησης και να εξερευνά υποσχόμενες περιοχές του χώρου αναζήτησης. Η βασική μορφή της αναζήτησης tabu αποτελείται από μια τροποποιημένη διαδικασία αναζήτησης γειτόνων που χρησιμοποιεί τη μνήμη προκειμένου να παρακολουθεί τις προηγούμενες λύσεις, σε συνδυασμό με στρατηγικές για την εκμετάλλευση της μνήμης [5].

3.2.5 Εμφωλευμένες Διαμερίσεις

Η μέθοδος εμφωλευμένων διαμερίσεων (Nested Partitions-NP) είναι μία μέθοδος βελτιστοποίησης μέσω προσομοίωσης, εφαρμόσιμη σε προβλήματα διακριτής βελτιστοποίησης στοχαστικά και ντετερμινιστικά. Έχει αποδειχθεί ότι μπορεί να δώσει γρήγορα σχεδόν βέλτιστες λύσεις και ότι συγκλίνει σε ολικό βέλτιστο με πιθανότητα ένα [16].

Η μέθοδος εμφωλευμένων διαμερίσεων συνδυάζει διαμέριση, τυχαία δειγματοληψία, επιλογή ενός «υποσχέσιμου» δείκτη και οπισθοχώρηση για τη δημιουργία μίας Μαρκοβιανής αλυσίδας της οποίας οι καταστάσεις με τις μέγιστες στάσιμες πιθανότητες αντιστοιχούν σε βέλτιστες λύσεις του αρχικού προβλήματος βελτιστοποίησης. Τα ντετερμινιστικά προβλήματα στα οποία επιχειρεί να δώσει λύση ο αλγόριθμος είναι της ακόλουθης μορφής:

$$\min_{\theta \in \Theta} g(\theta), \quad (14)$$

όπου Θ μια πεπερασμένη εφικτή περιοχή και $g : \Theta \rightarrow \mathbb{R}$ ένα μέτρο απόδοσης. Θεωρείται επίσης το πεπερασμένο στοχαστικό πρόβλημα της ακόλουθης μορφής:

$$\min_{\theta \in \Theta} J(\theta) := \mathbb{E}[f(\theta, \omega)], \quad (15)$$

όπου $J(\theta)$ είναι η μέση τιμή του μέτρου απόδοσης ενδιαφέροντος, $f(\theta, \omega)$ είναι η δειγματική απόδοση και το ω αναπαριστά τη στοχαστική επίδραση στο σύστημα. Σημειώνεται ότι το

πρόβλημα (15) είναι πολύ πιο δύσκολο να λυθεί από ότι το (14), καθώς στις περισσότερες περιπτώσεις δεν υπάρχει αναλυτική έκφραση της συνάρτησης απόδοσης $J(\theta)$ σε μία λύση $\theta \in \Theta$. Συχνά είναι απαραίτητη η χρήση παρατηρήσεων πραγματικού χρόνου ή η προσφυγή στην προσομοίωση προκειμένου να υπολογιστούν οι τιμές της συνάρτησης απόδοσης [17].

Στη συνέχεια παρουσιάζεται η ορολογία που θα ακολουθηθεί:

$$\Sigma = \{\sigma \subseteq \Theta \mid \sigma \text{ έγκυρη περιοχή για δοθείσα διαμέριση}\}$$

$$\Sigma_0 = \{\sigma \in \Sigma \mid \sigma \text{ περιοχή μεγίστου βάρους}\}$$

$$\sigma(k) \in \Sigma = \text{Η πιο υποσχόμενη περιοχή στην } k\text{-οστή επανάληψη}$$

$$N_k(\sigma) \in \mathbb{N} = \text{Το πλήθος των επαναλήψεων κατά τις οποίες η περιοχή } \sigma \in \Sigma \text{ έχει επιλεγεί} \\ \text{ως πλέον υποσχόμενη στις } k \text{ πρώτες επαναλήψεις}$$

$$\hat{\sigma}_{opt}(k) \in \Sigma_0 = \text{Η περιοχή μεγίστου βάρους που έχει επιλεγεί περισσότερες φορές ως πλέον} \\ \text{υποσχόμενη στις } k \text{ πρώτες επαναλήψεις}$$

$$d(\sigma) = \text{Το βάθος της } \sigma \in \Sigma$$

$$d^* = \text{Το μέγιστο βάθος}$$

$$s(\sigma) \in \Sigma = \text{Η περιοχή-γονέας της } \sigma \in \Sigma$$

Σημειώνεται ότι μια περιοχή $\sigma \subseteq \Theta$ καλείται έγκυρη αν αποτελεί ένα από τα υποσύνολα του Θ που διαμορφώνουν τη διαμέριση που έχει επιλεγεί [16].

Έχει αποδειχθεί πως αρκετά σημαντικά προβλήματα που εμπίπτουν στην κατηγορία που εξετάζεται εδώ, όπως το πρόβλημα του πλανόδιου πωλητή, είναι δύσκολα στην επίλυσή τους, υπό την έννοια πως η εύρεση ε -βέλτιστης λύσης παραμένει NP πρόβλημα. Επομένως, ούτε ακολουθιακοί ούτε παράλληλοι αλγόριθμοι μπορούν να χρησιμοποιηθούν για την αποτελεσματική επίλυση αυτών των προβλημάτων [17].

Στον τομέα των τυχαιοποιημένων αλγορίθμων έχουν προταθεί μέθοδοι βασισμένες σε Μαρκοβιανές αλυσίδες όπως η μέθοδος Metropolis (Hastings 1970) και η προσομοιωμένη εξάχνωση (Kirkpatrick, Helatt, Vecchi 1983) που αποσκοπούν στην επίλυση συνδυαστικών προβλημάτων βελτιστοποίησης. Ενώ, από την άλλη πλευρά, μέθοδοι όπως αυτές που έχουν προταθεί από τους Yan και Mukai (1992) και Andradottir (1995) στοχεύουν στην επίλυση διακριτών στοχαστικών προβλημάτων βελτιστοποίησης. Η μέθοδος που προτείνεται από τους συγγραφείς, η μέθοδος των εμφωλευμένων διαμερίσεων, (Shi και Olafsson 1996a) μπορεί να εφαρμοστεί και στις δύο κατηγορίες προβλημάτων βελτιστοποίησης. Είναι, επίσης, κατάλληλη,

για παράλληλες υπολογιστικές δομές [17].

Δοθέντων ενός πεπερασμένου συνόλου Θ , και μίας συνάρτησης απόδοσης που συμβολίζεται είτε με $g(\theta)$ ή $J(\theta)$, για τα προβλήματα που περιγράφηκαν παραπάνω, η μέθοδος εμφωλευμένων διαμερίσεων μπορεί να περιγραφεί από τα παρακάτω βήματα. Σε κάθε επανάληψη του αλγορίθμου υποθέτουμε ότι έχουμε μια περιοχή, ένα υποσύνολο του Θ , η οποία θεωρείται πως είναι η πλέον υποσχόμενη. Στη συνέχεια διαμερίζουμε αυτή την περιοχή σε M υποπεριοχές και συνενώνουμε την περιβάλλουσα περιοχή σε μία. Το πλήθος M των υποπεριοχών μπορεί να διαφέρει ανάλογα με την τρέχουσα πλέον υποσχόμενη περιοχή, δηλαδή δεν διαμερίζονται όλες οι περιοχές απαραίτητα στο ίδιο πλήθος υποσυνόλων. Σε αυτή την περίπτωση χρησιμοποιούμε το συμβολισμό $M_{\sigma(k)}$ για να δηλώσουμε το πλήθος των περιοχών στις οποίες διαμερίζεται η πλέον υποσχόμενη περιοχή στη k -οστή επανάληψη, $\sigma(k)$. Σε κάθε επανάληψη, επομένως, έχουμε στη διάθεσή μας $M + 1$ ξένα ανά δύο υποσύνολα του Θ σε καθένα από τα οποία πραγματοποιείται τυχαία δειγματοληψία και υπολογίζεται η τιμή του δείκτη με βάση την οποία θα κριθεί ποιο από αυτά αποτελεί την πλέον υποσχόμενη περιοχή στην επόμενη επανάληψη. Αν προκύψει ότι η περιβάλλουσα περιοχή είναι η πιο υποσχόμενη, τότε ο αλγόριθμος οπισθοχωρεί σε μία μεγαλύτερη περιοχή, που περιέχει την τρέχουσα περισσότερο υποσχόμενη και ορίζει εκείνη ως την νέα πλέον υποσχόμενη περιοχή. Η νέα πλέον υποσχόμενη περιοχή διαμερίζεται και λαμβάνονται νέα δείγματα με παρόμοιο τρόπο [17].

Στην πρώτη επανάληψη συνήθως επιλέγεται ολόκληρη η εφικτή περιοχή Θ ως περισσότερο υποσχόμενη, και αφού η περιβάλλουσα περιοχή είναι κενή γίνεται δειγματοληψία μόνο από τις M υποπεριοχές στις οποίες έχει διαμεριστεί το Θ . Αυτό ισχύει τόσο για την πρώτη επανάληψη, όσο και για κάθε επανάληψη στην οποία το Θ θεωρείται η πλέον υποσχόμενη περιοχή. Είναι φανερό πως, εφόσον το Θ είναι πεπερασμένο, η διαμέριση μπορεί να συνεχιστεί μέχρι που όλες οι περιοχές θα είναι μονοσύνολα και δεν θα είναι δυνατόν να διαμεριστούν περαιτέρω. Αυτές οι περιοχές καλούνται περιοχές μέγιστου βάθους. Η έννοια του βάθους γενικεύεται για οποιαδήποτε περιοχή, με το Θ να έχει βάθος 0 κ.ο.κ. [17].

Στη συνέχεια παρουσιάζονται αναλυτικά τα βήματα του αλγορίθμου των εμφωλευμένων διαμερίσεων.

1. Διαμέριση

Διαμέριση της πλέον υποσχόμενης περιοχής $\sigma(k)$, σε $M_{\sigma(k)}$ υποπεριοχές $\sigma_1(k)$, $\sigma_2(k)$, ..., $\sigma_{M_{\sigma(k)}}(k)$, και συνένωση της περιβάλλουσας περιοχής, $\Theta \setminus \sigma(k)$, σε μία περιοχή, που θα συμβολίζεται με $\sigma_{M_{\sigma(k)}+1}(k)$.

Πρέπει να σημειωθεί ότι το σχήμα της διαμέρισης καθορίζει το χώρο καταστάσεων την Μαρκοβιανής αλυσίδας και πρέπει να είναι σταθερό κατά την εφαρμογή της μεθόδου. Δηλαδή, αν μία περιοχή επιλέγεται επανειλημμένα ως η πλέον υποσχόμενη περιοχή, τότε ο ίδιος κανόνας διαμέρισης πρέπει να εφαρμόζεται στην περιοχή κάθε φορά.

Επιπλέον, καθώς η διαμέριση επιβάλλει μία δομή στην εφικτή περιοχή, η επιλογή της παίζει σημαντικό ρόλο στην ταχύτητα σύγκλισης του αλγορίθμου. Αν η διαμέριση είναι τέτοια ώστε καλές λύσεις να ομαδοποιούνται στην ίδια υποπεριοχή, είναι πιο πιθανό ο αλγόριθμος να επικεντρώνει γρήγορα την αναζήτηση σε αυτές ακριβώς τις υποπεριοχές της εφικτής περιοχής [18, 17].

2. Δειγματοληψία

Λήψη τυχαίου δείγματος N_j σημείων από κάθε μία από τις περιοχές $\sigma_j(k)$, $j = 1, 2, \dots, M_{\sigma(k)} + 1$,

$$\theta^{j_1}, \theta^{j_2}, \dots, \theta^{j_{N_j}}, \quad j = 1, 2, \dots, M_{\sigma(k)} + 1,$$

και υπολογισμός των αντίστοιχων τιμών της συνάρτησης απόδοσης $g(\theta)$ (ή $f(\theta, \omega)$ στη στοχαστική περίπτωση),

$$g(\theta^{j_1}), g(\theta^{j_2}), \dots, g(\theta^{j_{N_j}}), \quad j = 1, 2, \dots, M_{\sigma(k)} + 1.$$

Υπάρχει μεγάλη ευελιξία ως προς την στρατηγική δειγματοληψίας που θα ακολουθηθεί. Ο μοναδικός περιορισμός είναι πως κάθε σημείο στην περιοχή δειγματοληψίας πρέπει να έχει θετική πιθανότητα να επιλεγεί [17].

3. Υπολογισμός του υποσχέσιμου δείκτη

Για κάθε περιοχή σ_j , $j = 1, 2, \dots, M + 1$, ορίζεται μια συνάρτηση δείκτης, $I(\sigma_j)$, που θα καθορίσει την περισσότερο υποσχόμενη περιοχή και υπολογίζεται η τιμή του δείκτη για κάθε περιοχή. Ο δείκτης αυτός πρέπει να βασίζεται σε δειγματικές πληροφορίες που προκύπτουν από τις προσομοιωμένες εφικτές λύσεις από κάθε περιοχή. Βέβαια, μπορούν να ενσωματωθούν και άλλες πληροφορίες. Μία συχνή επιλογή υποσχέσιμου δείκτη, $I(\sigma_j)$, αποτελεί η βέλτιστη τιμή της συνάρτησης

απόδοσης στην περιοχή

$$I(\sigma_j) := \min_{\theta \in \sigma_j} g(\theta), \quad j = 1, 2, \dots, M_{\sigma(k)} + 1$$

που μπορεί να εκτιμηθεί από την

$$\hat{I}(\sigma_j) := \min_{i=1,2,\dots,N_j} g(\theta^i), \quad j = 1, 2, \dots, M_{\sigma(k)} + 1.$$

Στην στοχαστική περίπτωση ο αντίστοιχος δείκτης είναι

$$I(\sigma_j) = \min_{\theta \in \sigma_j} J(\theta), \quad j = 1, 2, \dots, M_{\sigma(k)} + 1.$$

Θεωρούμε $\hat{I}_t(\sigma_j, \omega)$ μια εκτίμηση του $I(\sigma_j)$, όπου t η μεταβλητή του χρόνου στην προσομοίωση. Τότε η $\hat{I}_t(\sigma_j, \omega)$ μπορεί να οριστεί ως

$$\hat{I}_t(\sigma_j, \omega) := \min_{i=1,2,\dots,N_j} f_t(\theta_i^j, \omega), \quad j = 1, 2, \dots, M_{\sigma_j} + 1.$$

Οι παραπάνω δείκτες αναφέρονται ως διατακτικοί υποσχέσιμοι δείκτες. Υπάρχει μεγάλη ευελιξία στην επιλογή του δείκτη. Ο μόνος περιορισμός είναι πως οι τιμές του δείκτη πρέπει να συμφωνούν, ως προς τη διάταξη, με τις τιμές της αντικειμενικής συνάρτησης στις περιοχές μεγίστου βάθους (μονοσύνολα) [17, 18, 19].

4. Οπισθοχώρηση

Καθορίζεται η πλέον υποσχόμενη περιοχή σ_{j_k} .

$$j_k \in \arg \min_{j=1,\dots,M_{\sigma(k)}+1} \hat{I}(\sigma_j).$$

Αν περισσότερες από μία περιοχές δώσουν ίδια τιμή στον υποσχέσιμο δείκτη, τότε επιλέγεται μία αυθαίρετα. Αν ο δείκτης υποδεικνύει μια υποπεριοχή της $\sigma(k)$ τότε αυτή ορίζεται ως πιο υποσχόμενη στην επόμενη επανάληψη. Διαφορετικά, αν ο δείκτης υποδεικνύει την περιβάλλουσα περιοχή, γίνεται οπισθοχώρηση στην περιοχή που αποτελεί το γεννήτορα της τρέχουσας πλέον υποσχόμενης περιοχής. Το βάθος του γεννήτορα αποτελεί επιλογή του εκάστοτε μελετητή. Μια συχνή επιλογή είναι η οπισθοχώρηση σε περιοχή με βάθος κατά ένα μικρότερο από αυτό της τρέχουσας πιο υποσχόμενης περιοχής, ενώ μία άλλη υπαγορεύει οπισθοχώρηση σε

βάθος μηδέν, σε ολόκληρη, δηλαδή, την εφικτή περιοχή [17].

Η βασική ιδέα του αλγορίθμου είναι η μεταφορά της εστίασης από το σύνολο της εφικτής περιοχής σε μια ακολουθία υποσυνόλων αυτής. Εξαιτίας αυτής της μεταφοράς εστίασης, μαζί με το γεγονός πως ο υποσχέσιμος δείκτης ορίζεται σε καθένα από τα υποσύνολα, είναι δυνατή η ενσωμάτωση οποιασδήποτε αποτελεσματικής ευρετικής μεθόδου στον αλγόριθμο με χρήση της για τον ορισμό του δείκτη που θα καθορίσει τις πλέον υποσχόμενες περιοχές στις διάφορες επαναλήψεις. Μια ακόμα σημαντική συνέπεια του τρόπου που λειτουργεί ο συγκεκριμένος αλγόριθμος είναι πως καθιστά τη μέθοδο ιδιαίτερα συμβατή με δομές παράλληλης επεξεργασίας. Αφού κάθε υποπεριοχή μπορεί να εξεταστεί παράλληλα και ανεξάρτητα από τις υπόλοιπες ο αλγόριθμος μπορεί να εκμεταλλευτεί στο μέγιστο τους διαθέσιμους πόρους [17].

Η μέθοδος που περιγράφηκε μπορεί να θεωρηθεί ως ένας γενικός αλγόριθμος ικανός να υποστηρίξει διαφορετικές τεχνικές διαμέρισης και μεθόδους δειγματοληψίας, καθώς και διαφορετικούς δείκτες και κανόνες οπισθοχώρησης [17].

Στη συνέχεια παρουσιάζονται τα βασικά αποτελέσματα σύγκλισης για τον αλγόριθμο εμφωλευμένων διαμερίσεων. Υποθέτουμε ότι ο αλγόριθμος εφαρμόζεται σε συνδυαστικό πρόβλημα βελτιστοποίησης, που περιλαμβάνει, δηλαδή, συνεχείς και διακριτές μεταβλητές απόφασης. Αρχικά διατυπώνουμε αυστηρά τη Μαρκοβιανή ιδιότητα [19].

Πρόταση 1. Έστω ότι η διαμέριση της εφικτής περιοχής είναι σταθερή, και Σ είναι το σύνολο όλων των έγκυρων περιοχών. Η στοχαστική διαδικασία $\{\sigma(k)\}_{k=1}^{\infty}$ που ορίζεται από την πλέον υποσχόμενη περιοχή σε κάθε επανάληψη του αλγορίθμου NP είναι ομογενής Μαρκοβιανή αλυσίδα με χώρο καταστάσεων Σ [19].

Θεώρημα 1. Η στάσιμη κατανομή $\{\pi(\eta)\}_{\eta \in \Sigma}$ της Μαρκοβιανής αλυσίδας $\{\sigma(k)\}_{k=1}^{\infty}$ που δημιουργείται από τον αλγόριθμο εμφωλευμένων διαμερίσεων, με κανόνα οπισθοχώρησης που υπαγορεύει οπισθοχώρηση σε ολόκληρη την εφικτή περιοχή Θ , ικανοποιεί την ακόλουθη εξίσωση,

$$\pi(\sigma) = \frac{P^{d^*}(\Theta, \eta)}{P(\eta, \Theta)} \pi(\Theta), \quad \forall \eta \in \Sigma_0. \quad (16)$$

Για τη συνέχεια υπενθυμίζουμε ότι $I(\eta) = \min_{\theta \in \eta} J(\theta)$ είναι ο υποσχέσιμος δείκτης που έχει επιλεγεί και θεωρούμε $S = \arg \min_{\eta \in \Sigma} I(\eta)$ το σύνολο των σημείων ολικού βέλτιστου του προβλήματος. Υποθέτουμε, ακόμα ότι υπάρχει $\sigma_{opt} \in S$ τέτοιο ώστε

$$P^{d^*}(\Theta, \sigma_{opt}) \geq P^{d^*}(\Theta, \eta), \quad \forall \eta \in \Sigma_0 \quad (17)$$

και επιπλέον ισχύει ότι αν $P^{d^*}(\Theta, \sigma_{opt}) = P^{d^*}(\Theta, \eta)$ τότε $\eta \in S$.

Αν ισχύει η παραπάνω συνθήκη κανονικότητας τότε $\pi(\sigma_{opt}) > \pi(\eta) \forall \eta \in \Sigma_0$ και συνεπώς η μέθοδος εμφωλευμένων διαμερίσεων συγκλίνει σχεδόν βέβαια σε αυτό το ολικό βέλτιστο.

Απόδειξη: Έστω $\eta \in \Sigma \setminus (\Sigma_0 \cup \Theta)$. Οι εξισώσεις ισορροπίας για την η δίνονται από τη σχέση

$$\pi(\eta) = P(s(\eta), \eta)\pi(s(\eta)). \quad (18)$$

Αν οι παραπάνω λυθούν επαναληπτικά, παίρνουμε

$$\pi(\eta) = P^{d(\eta)}(\Theta, \eta)\pi(\Theta) \quad (19)$$

Έστω τώρα $\sigma \in \Sigma_0$. Οι εξισώσεις ισορροπίας εδώ είναι

$$\pi(\sigma)P(\sigma, \Theta) = \pi(s(\sigma))P(s(\sigma), \sigma). \quad (20)$$

Κάνοντας χρήση της (18) η τελευταία σχέση γίνεται

$$\begin{aligned} \pi(\sigma) &= \frac{P(s(\sigma), \sigma)}{P(\sigma, \Theta)}\pi(s(\sigma)) \\ &= \frac{P(s(\sigma), \sigma)}{P(\sigma, \Theta)}P^{d(s(\sigma))}(\Theta, s(\sigma))\pi(\Theta) \\ &= \frac{P^{d^*}(\Theta, \sigma)}{P(\sigma, \Theta)}\pi(\Theta) \end{aligned} \quad (21)$$

Θεωρούμε, στη συνέχεια, σ_{opt} ολικό βέλτιστο και $\eta \in \Sigma_0$. Τότε χρησιμοποιώντας δύο φορές την (21) παίρνουμε ότι

$$\begin{aligned} \pi(\sigma_{opt}) &= \frac{P^{d^*}(\Theta, \sigma_{opt})}{P(\sigma_{opt}, \Theta)}\pi(\Theta) \\ &= \frac{P^{d^*}(\Theta, \sigma_{opt})}{P(\sigma_{opt}, \Theta)} \frac{P(\eta, \Theta)}{P^{d^*}(\Theta, \eta)}\pi(\eta) \\ &\geq \frac{P^{d^*}(\Theta, \sigma_{opt})}{P^{d^*}(\Theta, \eta)}\pi(\eta) \\ &\geq \pi(\eta) \end{aligned}$$

όπου η τελευταία ανισότητα προκύπτει από τη συνθήκη κανονικότητας. Δείξαμε, έτσι, ότι η στάσιμη πιθανότητα του βέλτιστου είναι μεγαλύτερη από αυτή κάθε άλλης περιοχής μεγίστου βάθους. Θυμίζουμε ότι η μέθοδος εκτιμά τη βέλτιστη λύση ως την περιοχή μεγίστου βάθους με τη μέγιστη συχνότητα εμφάνισης, $V_k = \arg \max_{\sigma \in \Sigma_0} N_k(\sigma)$. Έχουμε λοιπόν

$$\begin{aligned}
\lim_{k \rightarrow \infty} V_k &\in \lim_{k \rightarrow \infty} \arg \max_{\sigma \in \Sigma_0} N_k(\sigma) \\
&= \arg \max_{\sigma \in \Sigma_0} \lim_{k \rightarrow \infty} N_k(\sigma) \\
&= \arg \max_{\sigma \in \Sigma_0} \pi(\sigma), \quad a.s.
\end{aligned}$$

Εδώ, η τελευταία εξίσωση προκύπτει από το γνωστό αποτέλεσμα που υπαγορεύει πως αν $N_k(\sigma)$ είναι το πλήθος επισκέψεων σε μία κατάσταση σ μιας αδιαχώριστης Μαρκοβιανής αλυσίδας, τότε

$$\lim_{k \rightarrow \infty} \frac{N_k(\sigma)}{k} = \pi(\sigma), \quad a.s.$$

Το τελευταίο αποδεικνύει ότι η μέθοδος των εμφωλευμένων διαμερίσεων προσδιορίζει την περιοχή μεγίστου βάθους με τη μέγιστη στάσιμη πιθανότητα, και επομένως με αυτό ολοκληρώνεται και η απόδειξη [16].

Είναι προφανές ότι οι πιθανότητες μετάβασης της Μαρκοβιανής αλυσίδας εξαρτώνται από τη διαμέριση, αλλά και από τον τρόπο με τον οποίο προσομοιώνονται οι εφικτές λύσεις. Αυτά στη συνέχεια θα καθορίσουν την ταχύτητα σύγκλισης της Μαρκοβιανής αλυσίδας στην απορροφητική κατάσταση και επομένως καθορίζουν την απόδοση του αλγορίθμου NP. Είναι επομένως ζήτημα μεγάλης πρακτικής σημασίας, η ανάπτυξη μιας ευφυούς διαμερίσης που θα εκμεταλλεύεται τις ιδιαίτερες δομές του προβλήματος ενδιαφέροντος [19].

Μετά τα βασικά αποτελέσματα σύγκλισης που παρουσιάστηκαν παραπάνω, αποτελέσματα μεγάλης σημασίας θεωρούνται όσα σχετίζονται με το χρόνο σύγκλισης του αλγορίθμου. Έτσι, ακολούθως παρουσιάζονται φράγματα για το πλήθος των επαναλήψεων, και συνεπώς των πραγματοποιήσεων της συνάρτησης απόδοσης, που απαιτούνται για την εύρεση του επιθυμητού βέλτιστου. Θα υιοθετηθεί η υπόθεση ότι το Σ είναι πεπερασμένο, κάτι που γενικά ισχύει για τα προβλήματα που μας ενδιαφέρουν, εφόσον η διαμέριση έχει οριστεί κατάλληλα, ενώ χάριν απλότητας, η παρακάτω ανάλυση γίνεται υπό την υπόθεση ότι το ολικό βέλτιστο είναι μοναδικό [19].

Έστω Y το πλήθος επαναλήψεων μέχρι η Μαρκοβιανή αλυσίδα να απορροφηθεί, και έστω Y_η το πλήθος των επαναλήψεων που αναλογούν στην κατάσταση $\eta \in \Sigma$. Επειδή το ολικό μέγιστο είναι μοναδικό, η Μαρκοβιανή αλυσίδα ξοδεύει έναν αρχικό αριθμό επαναλήψεων στις μεταβατικές καταστάσεις και όταν πετυχαίνει για πρώτη φορά τη μοναδική απορροφητική κατάσταση, δεν επισκέπτεται ξανά κάποια άλλη κατάσταση. Επομένως, αρκεί να βρεθεί το αναμενόμενο πλήθος επισκέψεων σε κάθε μία από τις μεταβατικές καταστάσεις. Έστω, ακόμα,

T_η η χρονική στιγμή της πρώτης επίσκεψης στην κατάσταση $\eta \in \Sigma$. Έστω E ένα αυθαίρετο γεγονός και $\eta \in \Sigma$ μια έγκυρη περιοχή της διαμέρισης. Ορίζουμε τότε $P_\eta(E)$ την πιθανότητα του γεγονότος E , δεδομένου ότι η αλυσίδα ξεκινάει από την κατάσταση $\eta \in \Sigma$ [19].

Θα ήταν βολικό να δούμε το χώρο καταστάσεων Σ ως ένωση τριών ξένων ανά δύο υποσυνόλων, ως εξής. Έστω σ^* η περιοχή που αντιστοιχεί στο μοναδικό ολικό βέλτιστο. Ορίζουμε ακόμα $\Sigma_1 = \{\eta \in \Sigma \setminus \{\sigma^*\} \mid \sigma^* \subseteq \eta\}$ και $\Sigma_2 = \{\eta \in \Sigma \mid \sigma^* \not\subseteq \eta\}$. Τότε $\Sigma = \{\sigma^*\} \cup \Sigma_1 \cup \Sigma_2$ όπου τα τρία αυτά σύνολα είναι ξένα ανά δύο. Χρησιμοποιώντας τον παραπάνω συμβολισμό, μπορούμε τώρα να διατυπώσουμε το ακόλουθο αποτέλεσμα, που συνδέει τον αναμενόμενο χρόνο με τις πιθανότητες πρώτης επίσκεψης [19].

Θεώρημα 2. Το αναμενόμενο πλήθος επαναλήψεων μέχρι η Μαρχοβιανή αλυσίδα του αλγορίθμου NP να απορροφηθεί δίνεται από τη σχέση

$$E[Y] = 1 + \sum_{\eta \in \Sigma_1} \frac{1}{P_\eta[T_{\sigma^*} < T_\eta]} + \sum_{\eta \in \Sigma_2} \frac{P_X[T_\eta < \min\{T_X, T_{\sigma^*}\}]}{P_\eta[T_X < T_\eta] \cdot P_X[T_{\sigma^*} < \min\{T_X, T_\eta\}]} \quad (22)$$

Το αναμενόμενο πλήθος επαναλήψεων (22) εξαρτάται τόσο από τη διαμέριση και από τον τρόπο με τον οποίο δημιουργούνται οι δειγματικές λύσεις, όσο και από τη δομή του ίδιου του προβλήματος. Ο ακριβής υπολογισμός αυτής της αναμενόμενης τιμής είναι επομένως περίπλοκος. Βέβαια, κάτω από συγκεκριμένες επιπλέον υποθέσεις είναι δυνατόν να βρεθούν χρήσιμα φράγματα για το αναμενόμενο πλήθος επαναλήψεων [19].

Υποθέτουμε ότι P^* είναι ένα κάτω φράγμα της πιθανότητας επιλογής της σωστής περιοχής. Θα αναφερόμαστε στην πιθανότητα αυτή ως ελάχιστη πιθανότητα επιτυχίας. Το ακόλουθο θεώρημα παρέχει ένα άνω φράγμα για τον αναμενόμενο χρόνο μέχρι ο αλγόριθμος NP να συγκλίνει, το οποίο εξαρτάται από την ελάχιστη πιθανότητα επιτυχίας P^* και το μέγεθος της επικτής περιοχής όπως μετράται από το μέγιστο βάθος, d^* [19].

Θεώρημα 3. Έστω ότι $P^* > 0.5$. Το αναμενόμενο πλήθος επαναλήψεων μέχρι η Μαρχοβιανή αλυσίδα του αλγορίθμου NP να απορροφηθεί φράσσεται ως εξής [19]

$$E[Y] \leq \frac{d^*}{2P^* - 1}. \quad (23)$$

Σημειώνεται ότι το δεξί μέλος της (23) αυξάνεται μόλις γραμμικά ως προς d^* . Επιπλέον, όπως και για οποιοδήποτε άλλο δέντρο, έτσι και το βάθος d^* του δέντρου της διαμέρισης είναι λογαριθμική συνάρτηση των παραμέτρων εισόδου. Με άλλα λόγια, αν υποθέσουμε, χάριν ευκολίας, ότι έχουμε μία μόνο παράμετρο εισόδου n , τότε η (23) παρέχει ένα άνω φράγμα για το αναμενόμενο πλήθος επαναλήψεων που είναι $O(\log n)$. Αυτό δείχνει ότι ο αναμενόμενος

αριθμός επαναλήψεων που απαιτείται από τον αλγόριθμο NP, αυξάνεται πολύ αργά, καθώς αυξάνεται το μέγεθος του προβλήματος, κάτι το οποίο εξηγεί εν μέρει γιατί ο αλγόριθμος NP είναι ιδιαίτερα αποδοτικός στην επίλυση προβλημάτων βελτιστοποίησης μεγάλης κλίμακας. Από την άλλη μεριά, τα φράγματα της (23) αυξάνουν εκθετικά καθώς $P^* \rightarrow \frac{1}{2}$ [19].

Τα παραπάνω αποτελέσματα ενισχύουν την προαναφερθείσα δήλωση ότι η αποδοτικότητα του NP αλγορίθμου βασίζεται στο να γίνεται συχνά η σωστή κίνηση. Αυτή η πιθανότητα επιτυχίας εξαρτάται με τη σειρά της, τόσο από τη διαμέριση όσο και από τον τρόπο με τον οποίο προσομοιώνονται οι εφικτές λύσεις. Είναι επομένως σημαντικό, για κάθε πρακτική εφαρμογή, να αυξάνεται η πιθανότητα επιτυχίας μέσα από την από την ανάπτυξη μεθόδων ευφυούς διαμέρισης, που ενσωματώνουν τις ιδιαιτερότητες στη δομή κάθε προβλήματος μέσω δειγματοληψίας με βάρη, και την εφαρμογή τυχαιοποιημένων ευρετικών μεθόδων για τη δημιουργία ποιοτικών εφικτών λύσεων [19].

4 Εφαρμογή της Μεθόδου Εμφωλευμένων Διαμερίσεων

Στο παρόν κεφάλαιο παρουσιάζεται η εφαρμογή της μεθόδου εμφωλευμένων διαμερίσεων στο πρόβλημα κατανομής θέσεων στους χώρους αναμονής ενός δικτύου, γνωστό ως Buffer Allocation Problem. Η ανάπτυξη του αλγορίθμου γίνεται σε προγραμματιστικό περιβάλλον MATLAB®, έκδοση R2020.

4.1 Θεμελίωση του Προβλήματος

Το συγκεκριμένο πρόβλημα απαντάται σε πολλούς τομείς όπως δίκτυα επικοινωνιών, γραμμές παραγωγής και διαχείριση υλικών. Πρόκειται για ένα πρόβλημα που εμφανίζεται σε συστήματα που έχουν τη μορφή δικτύων ουρών με πεπερασμένους χώρους αναμονής (buffer). Εξαιτίας του κρίσιμου κόστους των θέσεων στους χώρους αναμονής ο καθορισμός της χωρητικότητας των buffers αποτελεί ζήτημα μείζονος σημασίας με σκοπό να εξασφαλιστεί η μέγιστη απόδοση με το ελάχιστο δυνατό κόστος. Για να γίνει περισσότερο κατανοητή η φύση του προβλήματος, θα παρουσιαστούν σύντομα στα πλαίσια της γραμμής παραγωγής τα βασικά του σημεία [20].

Μία γραμμή παραγωγής αποτελείται από τις πρώτες ύλες, το χώρο αποθήκευσης και το χώρο των εργασιών. Στα πλαίσια της ορολογίας της θεωρίας ουρών οι πρώτες ύλες θα αναφέρονται ως εργασίες, οι χώροι αποθήκευσης ως buffers και οι χώροι εργασιών ως εξυπηρετητές (servers). Η πορεία των πρώτων υλών ξεκινάει από το χώρο αποθήκευσης προς το χώρο εργασιών από όπου και εξέρχονται από τη γραμμή παραγωγής με τη μορφή του τελικού προϊόντος αφού έχουν υποστεί την απαραίτητη επεξεργασία. Μέσω της κατάλληλης κατανομής θέσεων στους buffers μπορεί να περιοριστεί ο αντίκτυπος των διακυμάνσεων και να εξασφαλιστεί η ομαλή λειτουργία της γραμμής παραγωγής. Η ανάθεση πλεοναζόντων θέσεων στους χώρους αναμονής αυξάνει ανούσια το κόστος, ενώ από την άλλη ο μη επαρκής αριθμός θέσεων οδηγεί σε παρεμπόδιση (blocking) της παραγωγής. Είναι φανερό, λοιπόν, πως η κατανομή θέσεων στους χώρους αναμονής μπορεί να συμβάλει σημαντικά στην εύρυθμη λειτουργία μιας γραμμής παραγωγής και αξίζει να μελετηθεί περαιτέρω [21].

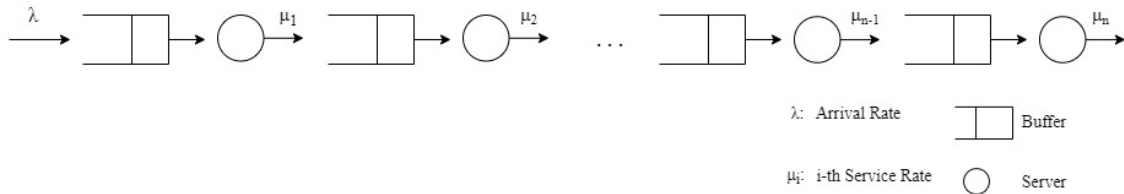
4.2 Ανάπτυξη Εφαρμογής

Για τους σκοπούς της παρούσας εργασίας θα προσομοιώσουμε ένα γραμμικό δίκτυο το οποίο θα αποτελείται από n buffers και n servers, όπως απεικονίζεται στο Σχήμα 1. Θεωρούμε ότι υπάρχουν N διαθέσιμες θέσεις για κατανομή στους χώρους αναμονής. Η ελάχιστη χωρη-

τικότητα που μπορεί να ανατεθεί σε ένα buffer είναι 0 και η μέγιστη N , με την προϋπόθεση, πάντα, ότι η συνολική χωρητικότητα είναι ίση με N . Στόχος είναι η εύρεση της κατανομής των διαθέσιμων θέσεων στους buffers του δικτύου έτσι ώστε να μεγιστοποιείται ο ρυθμός εξόδου (throughput). Ο όρος throughput γενικά χρησιμοποιείται για να δηλώσει το ρυθμό με τον οποίο εξυπηρετούνται οι πελάτες σε ένα δίκτυο. Ορίζεται ως το πηλίκο του πλήθους πελατών (εργασιών) που εξυπηρετούνται (διεκπεραιώνονται) προς τη μονάδα του χρόνου. Το πρόβλημα που καλούμαστε να επιλύσουμε, επομένως, έχει την ακόλουθη μορφή

$$\begin{aligned} \max_b \quad & J(b = (b_1, \dots, b_n)) \\ \text{s.t.} \quad & \sum_{i=1}^n b_i = N, \\ & b_i \geq 0, \quad i = 1, \dots, n \end{aligned}$$

όπου b_i είναι η χωρητικότητα του i -οστού buffer, και $J(b)$ το throughput που προκύπτει από τη λειτουργία του δικτύου με την κατανομή $b = (b_1, \dots, b_n)$. Δεδομένου ότι η αντικειμενική συνάρτηση, $J(b)$, δεν μπορεί να υπολογιστεί αναλυτικά, καταφεύγουμε στην προσομοίωση για τον υπολογισμό των τιμών της για τις διάφορες κατανομές των χωρητικοτήτων στους n buffers.



Σχήμα 1: Γραφική αναπαράσταση του δικτύου

Ο χώρος των εφικτών λύσεων του προβλήματος, Θ , είναι διακριτός και πεπερασμένος, με μεγάλο όμως πλήθος στοιχείων. Συγκεκριμένα ισχύει $|\Theta| = \binom{N+n-1}{n-1}$. Για παράδειγμα στην περίπτωση που έχουμε $N = 10$ διαθέσιμες θέσεις και $n = 5$ buffers που θα εξετάσουμε παρακάτω, το πλήθος των εφικτών λύσεων ανέρχεται σε $\binom{14}{4} = 1001$ στοιχεία. Τέτοια προβλήματα μπορούν να αντιμετωπιστούν με τη μέθοδο των εμφωλευμένων διαμερίσεων όπως αναπτύχθηκε από τους Shi and Olafsson [16].

Αρχικά, σχεδιάστηκε συνάρτηση για την προσομοίωση της λειτουργίας του δικτύου με συγκεκριμένο χρονικό ορίζοντα. Το δίκτυο δέχεται αφίξεις μέχρι το χρονικό ορίζοντα αλλά

παρατηρείται μέχρι να εξυπηρετηθεί ο τελευταίος πελάτης και να αποχωρήσει από αυτό. Οι χρόνοι αφίξεως των πελατών προσομοιώνονται, και αποθηκεύονται σε ένα διάνυσμα ώστε να χρησιμοποιηθούν στη συνέχεια. Το πρόγραμμα λειτουργεί παρακολουθώντας την πορεία κάθε πελάτη στο δίκτυο. Η πορεία ενός πελάτη, με την προϋπόθεση ότι έχει εισέλθει στο δίκτυο, καθορίζεται από τους χρόνους άφιξης σε κάθε buffer και κάθε server, αλλά και το χρόνο αποχώρησης του από το δίκτυο. Αυτοί οι χρόνοι καταγράφονται και αποθηκεύονται για κάθε πελάτη.

Μοναδική προϋπόθεση για την είσοδο ενός πελάτη στο δίκτυο αποτελεί η εύρεση ελεύθερης θέσης στον πρώτο buffer του δικτύου κατά την άφιξή του. Αν οι πελάτες που βρίσκονται στον πρώτο buffer έχουν εξαντλήσει τη χωρητικότητά του τη στιγμή της άφιξης ενός πελάτη, τότε αυτός δεν εισέρχεται στο δίκτυο και χάνεται. Στη περίπτωση που ο πρώτος buffer του δικτύου είναι μηδενικής χωρητικότητας, η είσοδος στο δίκτυο εξαρτάται από την κατάσταση του πρώτου server κατά την άφιξη. Σε κάθε άλλη περίπτωση ο πελάτης εισέρχεται κανονικά και καταγράφεται η πορεία του στο δίκτυο.

Επιπλέον, στην περίπτωση που ένας πελάτης ολοκληρώσει την εξυπηρέτησή του σε ένα server και οι θέσεις στον επόμενο buffer έχουν εξαντληθεί, τότε παραμένει στο server, μέχρι να ελευθερωθεί θέση στον επόμενο buffer, οπότε και προχωράει αμέσως. Σε αυτή την περίπτωση, όπου ένας πελάτης μένει σε ένα server χωρίς να εξυπηρετείται, λέμε ότι έχουμε το φαινόμενο της παρεμπόδισης (blocking).

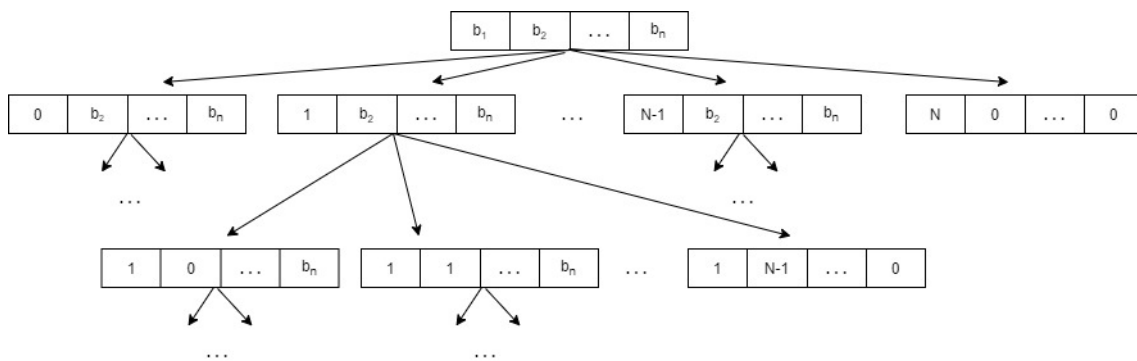
Όσον αφορά σε περισσότερο τεχνικά θέματα σχετικά με την ανάπτυξη της συνάρτησης προσομοίωσης σε MATLAB®, αξίζει να αναφερθεί ότι τα εξαγόμενα της προσομοίωσης συγκεντρώνονται σε μία λίστα αποτελούμενη από αντικείμενα της κλάσης «Customer» και έναν πίνακα με τους χρόνους αφίξεως κάθε πελάτη σε κάθε buffer και server. Η κλάση Customer έχει κατασκευαστεί ώστε να περιγράφει αντικείμενα που χαρακτηρίζονται από τα εξής πεδία:

- ένα διάνυσμα με τους χρόνους αφίξεως σε κάθε buffer,
- ένα διάνυσμα με τους χρόνους αφίξεως σε κάθε server,
- το χρόνο εισόδου στο δίκτυο,
- το χρόνο αποχώρησης από το δίκτυο,
- ένα δείκτη που δηλώνει αν εισήλθε ή όχι ο πελάτης στο δίκτυο και
- μία μεταβλητή που φανερώνει τη θέση του πελάτη στο δίκτυο κάθε στιγμή.

Το δίκτυο προσομοιώνεται με τη χρήση μίας ακόμη κλάσης με το όνομα Network η οποία περιλαμβάνει πεδία που περιγράφουν το δίκτυο, σχετικά με το πλήθος των buffers και servers που διαθέτει και το πώς αυτοί συνδέονται μεταξύ τους, ενώ διαθέτει και πεδία που περιγράφουν την κατάσταση κάθε buffer και server κάθε χρονική στιγμή. Με τον όρο κατάσταση εννοούμε το πλήθος των πελατών που καταλαμβάνουν κάθε θέση αλλά και το αύξοντα αριθμό τους στην λίστα με τα εξαγόμενα της συνάρτησης.

Η αντικειμενική συνάρτηση του προβλήματος βελτιστοποίησης, το throughput όπως αναφέρθηκε, υπολογίζεται μετά την προσομοίωση μίας πλήρους περιόδου λειτουργίας του δικτύου, ως το πηλίκο του πλήθους των πελατών που εισήλθαν στο δίκτυο προς το συνολικό χρόνο λειτουργίας του συστήματος, μέχρι δηλαδή ο τελευταίος πελάτης να αποχωρήσει από αυτό.

Για το παρόν πρόβλημα βελτιστοποίησης εφαρμόσαμε τον αλγόριθμο εμφωλευμένων διαμερίσεων ως εξής: αρχικά, η διαμέριση της εφικτής περιοχής ξεκινάει με τον καθορισμό της χωρητικότητας του πρώτου buffer, ακολούθως του δεύτερου, του τρίτου κ.ο.κ., μέχρις ότου να έχουν ανατεθεί χωρητικότητες σε όλους τους buffers όπως φαίνεται στο Σχήμα 2.



Σχήμα 2: Διαμέριση της εφικτής περιοχής

Σε κάθε επανάληψη, διαμερίζεται η τρέχουσα πλέον υποσχόμενη περιοχή, σύμφωνα με την ανωτέρω διαμέριση και πραγματοποιείται απλή τυχαία δειγματοληψία σε κάθε μία από τις υποπεριοχές της αλλά και από την περιβάλλουσα εφόσον αυτή δεν είναι κενή. Για κάθε σημείο που επιλέγεται από τη δειγματοληψία, προσομοιώνεται μέσω ξεχωριστής συνάρτησης η λειτουργία του δικτύου για χρόνο T και εξάγεται το προσομοιωμένο throughput. Εντοπίζεται η περιοχή στην οποία αντιστοιχεί το μέγιστο throughput. Αν αυτή είναι υποπεριοχή της τρέχουσας πλέον υποσχόμενης, τότε ορίζεται ως περισσότερο υποσχόμενη για την επόμενη επανάληψη. Στην περίπτωση που το μέγιστο throughput βρεθεί να αντιστοιχεί στην περιβάλλουσα περιοχή τότε ο αλγόριθμος οπισθοχωρεί ορίζοντας ως πλέον υποσχόμενη την περιοχή μηδενικού βάθους Θ . Το πρόγραμμα τερματίζει μετά από προκαθορισμένο αριθμό επαναλήψεων. Ως

εκτίμηση της βέλτιστης λύσης του προβλήματος επιλέγεται η περιοχή μεγίστου βάθους με τη μεγαλύτερη συχνότητα επίσκεψης κατά τη λειτουργία του προγράμματος σύμφωνα με το [16].

4.3 Παρουσίαση Αποτελεσμάτων

Στη συνέχεια παρουσιάζεται μία σειρά αποτελεσμάτων από διαφορετικές εκτελέσεις του αλγορίθμου. Οι πελάτες σε όλες τις εφαρμογές που θα ακολουθήσουν θεωρούμε ότι καταφτάνουν στο δίκτυο σύμφωνα με διαδικασία Poisson με ρυθμό λ , ενώ ο χρόνος εξυπηρέτησης στον i -οστό server υποθέτουμε ότι ακολουθεί εκθετική κατανομή με ρυθμό μ_i . Φυσικά, μπορούμε να υποθέσουμε και διαφορετικές κατανομές τόσο για τις αφίξεις όσο και για τους χρόνους εξυπηρέτησης, κάτι που πολύ εύκολα, με μικρές τροποποιήσεις μπορεί να ενταχθεί στον υπάρχοντα κώδικα. Επιπλέον, μπορούμε να υποθέσουμε ότι στο σύστημα φτάνουν πάνω από ένας τύπος πελατών, με διαφορετικές κατανομές άφιξης αλλά και χρόνους εξυπηρέτησης. Πολλές είναι οι παραλλαγές που μπορούν να διατυπωθούν και να αντιμετωπιστούν με μικρότερες ή μεγαλύτερες αλλαγές στον αλγόριθμο που έχει υλοποιηθεί. Το πρόβλημα στη συνέχεια εξετάζεται στην απλή μορφή που αναφέραμε.

Για το πρώτο παράδειγμα θεωρήσαμε δίκτυο με τρεις σταθμούς και πέντε διαθέσιμες θέσεις προς κατανομή στους χώρους αναμονής. Σημειώνουμε ότι ο πληθάριθμος του Θ εδώ είναι ίσος με $\binom{7}{2} = 21$. Υποθέσαμε ρυθμό αφίξεων $\lambda = 0.4$ και ρυθμούς εξυπηρέτησης $\mu_1 = 0.2$, $\mu_2 = 0.5$, και $\mu_3 = 0.4$ για τον πρώτο, το δεύτερο και τον τρίτο server αντίστοιχα. Η λειτουργία του δικτύου παρατηρήθηκε για $T = 200$ χρονικές μονάδες.

Μία παράμετρος του προβλήματος που δεν έχει προκαθορισμένη τιμή είναι το μέγεθος του δείγματος M , που θα λαμβάνεται από τις διάφορες υποπεριοχές κατά τη λειτουργία του αλγορίθμου. Προκειμένου να επιλέξουμε κατάλληλη τιμή για την παράμετρο M εκτελέσαμε τον αλγόριθμο βελτιστοποίησης 10 φορές για κάθε ένα από τα ακόλουθα μεγέθη δείγματος: $M_1 = 100$, $M_2 = 50$, $M_3 = 30$, $M_4 = 20$, $M_5 = 10$, $M_6 = 5$. Από τις 60 εκτελέσεις του αλγορίθμου προέκυψαν 5 διαφορετικές λύσεις. Οι λύσεις αυτές μαζί με τη συχνότητα εμφάνισής τους, αλλά και στατιστικά μεγέθη που προέκυψαν από Monte Carlo προσομοίωση μετά από $K = 1000$ επαναλήψεις παρατίθενται στον Πίνακα 1. Παρατηρούμε ότι οι πιο συχνά εμφανιζόμενες λύσεις έδωσαν και μεγαλύτερη εκτίμηση στην τιμή της αντικειμενικής συνάρτησης. Οι διαφορές βέβαια είναι της τάξης του 10^{-3} ακόμα και για τις λύσεις που εμφανίστηκαν λιγότερες φορές.

Για τη σύγκριση των τιμών του M καταφεύγουμε στον Πίνακα 2. Για να κατασκευάσουμε το συγκεκριμένο Πίνακα εκτιμήσαμε μέσω προσομοίωσης Monte Carlo τις τιμές της αντι-

Πίνακας 1: Βέλτιστες λύσεις για το πρώτο παράδειγμα.

$\lambda = 0.4 \mu = [0.2 \ 0.5 \ 0.4]$						
Allocation	Frequency	Average	Min	Max	CI (95%)	Simulation Time (s)
3 2 0	31	0.1799	0.1036	0.2646	(0.1784, 0.1814)	19.9
3 1 1	21	0.1793	0.0996	0.2640	(0.1778, 0.1808)	20.9
4 1 0	4	0.1758	0.1091	0.2561	(0.1744, 0.1772)	18.9
2 2 1	2	0.1770	0.0923	0.2426	(0.1756, 0.1784)	21.3
2 3 0	2	0.1756	0.1011	0.2583	(0.1741, 0.1770)	18.2

κειμενικής συνάρτησης για κάθε λύση που προέκυψε από τις 60 εκτελέσεις του αλγορίθμου βελτιστοποίησης. Στη στήλη «Average» έχουμε καταγράψει το μέσο όρο των εκτιμημένων throughput για τις δέκα λύσεις που αντιστοιχούν σε κάθε τιμή του M , ενώ στις στήλες «Min» και «Max» το ελάχιστο και το μέγιστο εκτιμημένο throughput αντίστοιχα. Στη στήλη «Backtracking %» καταγράφεται το ποσοστό των φορών που επιλέχθηκε η περιβάλλουσα περιοχή ως πλέον υποσχόμενη και ο αλγόριθμος οπισθοχώρησε στο Θ , ενώ στην τελευταία στήλη έχουμε το μέσο χρόνο εκτέλεσης του αλγορίθμου βελτιστοποίησης. Εύκολα μπορούμε να δούμε ότι η μόνη ποσότητα που παρουσιάζει σημαντική διαφορά καθώς μεταβάλλεται το μέγεθος M είναι οι χρόνοι εκτέλεσης. Οι λύσεις φαίνεται να παρουσιάζουν μικρές διαφορές μεταξύ τους. Δεν έχουμε ενδείξεις, επομένως, ότι κάποια τιμή της παραμέτρου M υπερτερεί των άλλων παρά μόνο από πλευράς χρόνου εκτέλεσης.

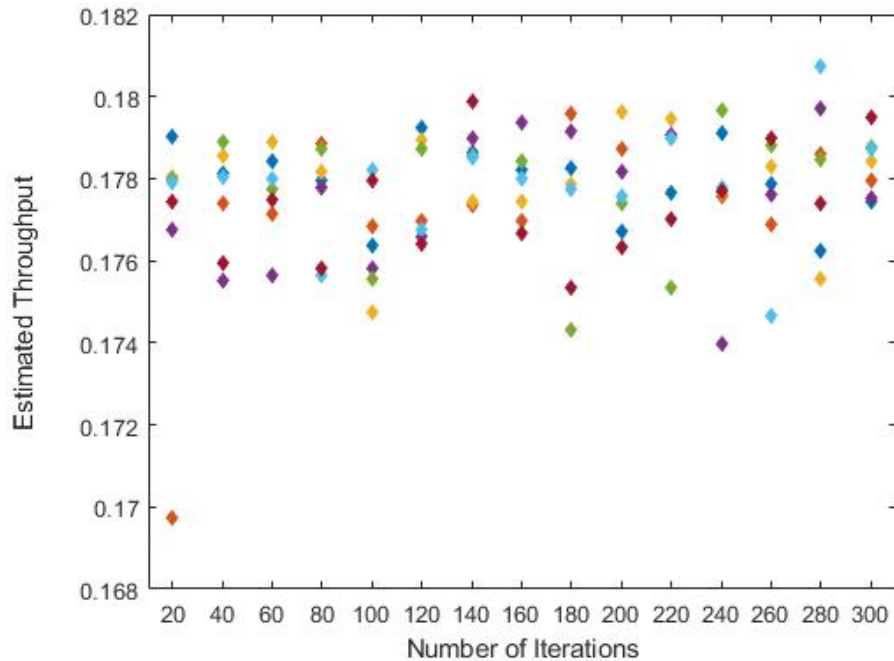
Πίνακας 2: Σύγκριση τιμών της παραμέτρου M .

M	Average Throughput	Min	Max	Bactracking %	Average Time (min)
100	0.1791	0.1758	0.1799	24.2	22.37
50	0.1791	0.1756	0.1799	25.2	11.3
30	0.1798	0.1793	0.1799	25.5	6.92
20	0.1792	0.1758	0.1799	26.3	4.8
10	0.1791	0.1758	0.1799	27.8	2.48
5	0.1790	0.1756	0.1799	26.7	1.26
Random Selection	0.1449	0.1296	0.1694	–	–

Στην τελευταία γραμμή του Πίνακα 2 παρουσιάζονται τα αντίστοιχα μεγέθη για μία τυχαία επιλογή εφικτών λύσεων, διαφορετικών από αυτές του Πίνακα 1, και είναι εμφανές ότι δίνουν μικρότερες εκτιμήσεις στις τιμές της αντικειμενικής συνάρτησης, γεγονός που αποτελεί ένδειξη για την ορθή λειτουργία του αλγορίθμου.

Επισημαίνεται σε αυτό το σημείο ότι για τα ανωτέρω αποτελέσματα ο αλγόριθμος τερματίστηκε μετά από 300 επαναλήψεις. Επιλέξαμε αυτή την τιμή, αρχικά τουλάχιστον, για να εξασφαλίσουμε ότι θα εξερευνηθεί η εφικτή περιοχή σε ένα ικανό ποσοστό και να αυξήσουμε

την πιθανότητα εύρεσης μίας καλής λύσης. Στη συνέχεια, έχοντας εξάγει τα ανωτέρω αποτελέσματα τρέξαμε τον αλγόριθμο με πλήθος επαναλήψεων από από 20 μέχρι 300. Επιλέξαμε $M = 30$ και πήραμε τα εξής:



Σχήμα 3: Σύγκριση αριθμού επαναλήψεων

Όπως μπορούμε να δούμε στο Σχήμα 3 από τις εκτιμημένες τιμές της αντικειμενικής συνάρτησης στις λύσεις που προέκυψαν ως βέλτιστες κατά τις εκτελέσεις του αλγορίθμου, δεν μπορούμε να συμπεράνουμε ότι κάποιος αριθμός επαναλήψεων δίνει συστηματικά καλύτερες λύσεις. Μία λύση φαίνεται να διαφέρει περισσότερο, μάλιστα λαμβάνοντας μικρότερη τιμή από τις υπόλοιπες, και αυτή προέκυψε μετά από 20 επαναλήψεις, το μικρότερο αριθμό επαναλήψεων που ελέγξαμε.

Στη συνέχεια, προσομοιώσαμε ένα παράδειγμα μεγαλύτερου δικτύου με $n = 5$ στάδια και $N = 10$ θέσεις προς κατανομή. Ο ρυθμός αφίξεων για αυτό το παράδειγμα θεωρήθηκε $\lambda = 0.5$ και οι ρυθμοί εξυπηρέτησης $\mu_1 = 1/3$, $\mu_2 = 1/3$, $\mu_3 = 1/3$, $\mu_4 = 1/2$, και $\mu_5 = 1$. Το σύστημα παρατηρήθηκε για $T = 400$ χρονικές μονάδες, ενώ και εδώ δοκιμάσαμε διαφορετικές τιμές για την παράμετρο M . Για κάθε μία από τις 5 τιμές του M εκτελέσαμε 10 φορές τον αλγόριθμο βελτιστοποίησης. Από τις 50 αυτές εκτελέσεις, λάβαμε 17 διαφορετικές κατανομές ως βέλτιστες. Θυμίζουμε, σε αυτό το σημείο ότι το πλήθος των εφικτών λύσεων για το συγκεκριμένο πρόβλημα είναι 1001.

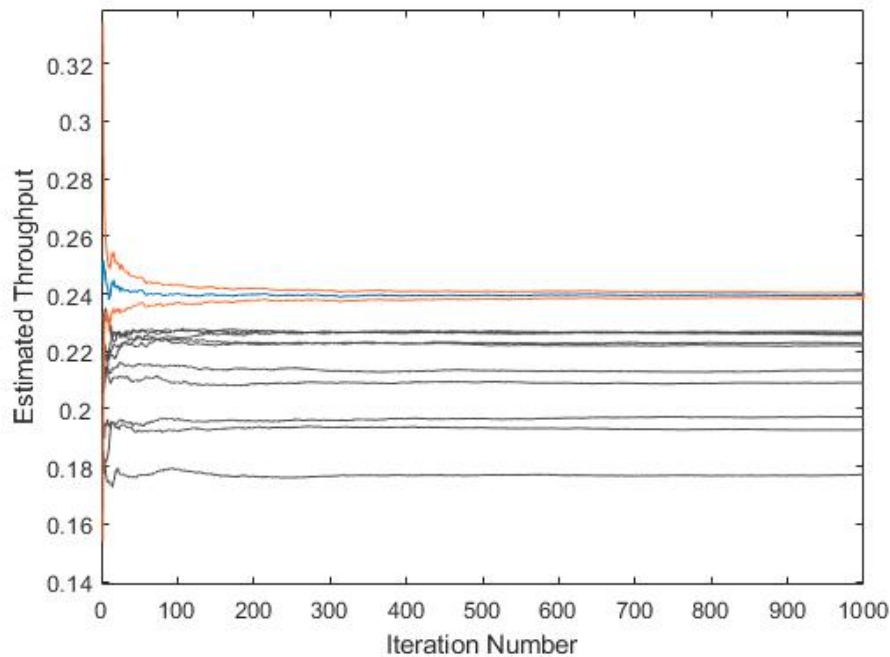
Ακολούθως, για να ελέγξουμε την ποιότητα των παραγόμενων λύσεων, εκτελέσαμε προσομοιώσεις Monte Carlo για την εκτίμηση των τιμών της αντικειμενικής συνάρτησης σε κάθε μία από αυτές. Στον Πίνακα 3 παρουσιάζονται τα αποτελέσματα της Monte Carlo προσομοίωσης. Παρατηρούμε ότι οι διαφορές στις εκτιμώμενες τιμές της αντικειμενικής συνάρτησης για τις παραγόμενες λύσεις είναι σχετικά μικρές, της τάξης του 10^{-3} για τις περισσότερες.

Πίνακας 3: Βέλτιστες λύσεις για το δεύτερο παράδειγμα.

$\lambda = \frac{1}{2}, \mu = [\frac{1}{3} \frac{1}{3} \frac{1}{3} \frac{1}{2} 1]$					
b	Frequency	Average	Min	Max	CI (95%)
2 4 3 1 0	10	0.2395	0.1859	0.2925	(0.2385, 0.2405)
3 3 3 1 0	9	0.2392	0.1922	0.2838	(0.2383, 0.2402)
2 3 4 1 0	6	0.2393	0.1876	0.2947	(0.2384, 0.2403)
3 2 4 1 0	3	0.2383	0.1866	0.2974	(0.2373, 0.2392)
2 5 2 1 0	3	0.2367	0.1897	0.2838	(0.2358, 0.2376)
2 3 3 2 0	3	0.2396	0.1917	0.2889	(0.2387, 0.2406)
2 4 2 2 0	3	0.2384	0.1932	0.2902	(0.2374, 0.2393)
3 2 3 2 0	2	0.2387	0.1887	0.2893	(0.2378, 0.2397)
2 2 5 1 0	2	0.2360	0.1792	0.2810	(0.2351, 0.2370)
2 4 4 0 0	2	0.2319	0.1896	0.2751	(0.2310, 0.2328)
3 1 4 1 1	1	0.2291	0.1800	0.2756	(0.2281, 0.2301)
3 3 2 1 1	1	0.2339	0.1793	0.2752	(0.2330, 0.2349)
3 4 2 1 0	1	0.2368	0.1937	0.2920	(0.2358, 0.2378)
2 5 3 0 0	1	0.2302	0.1830	0.2746	(0.2293, 0.2311)
2 2 3 3 0	1	0.2352	0.1864	0.2857	(0.2343, 0.2362)
2 2 3 1 2	1	0.2322	0.1912	0.2764	(0.2313, 0.2331)
2 2 4 2 0	1	0.2369	0.1914	0.2844	(0.2360, 0.2379)

Ακολούθως, εκτιμήθηκε ο ρυθμός εξόδου για διαφορετικές κατανομές χωρητικότητας, που επιλέχθηκαν μέσα από απλή τυχαία δειγματοληψία, με σκοπό την αξιολόγηση της ποιότητας των παραπάνω λύσεων. Στο Σχήμα 4 απεικονίζονται με γκρι χρώμα οι Monte Carlo μέσοι των τυχαία επιλεγμένων κατανομών, με μπλε χρώμα η εκτίμηση της λύσης που εμφανίστηκε συχνότερα ως βέλτιστη, ενώ με πορτοκαλί χρώμα έχουν σημειωθεί τα αντίστοιχα διαστήματα εμπιστοσύνης για την τελευταία λύση. Είναι εμφανές πως οι κατανομές που επιλέχθηκαν τυχαία διαφέρουν σημαντικά από τη λύση που έδωσε ο αλγόριθμος, γεγονός που αποτελεί ένδειξη της ορθής λειτουργίας του.

Για να διερευνήσουμε περαιτέρω την τιμή της παραμέτρου M εκτελέσαμε το πρόγραμμα για κάθε μέγεθος δείγματος δέκα φορές και εκτιμήσαμε για κάθε λύση που πήραμε το μέγιστο, το ελάχιστο και το μέσο throughput. Τα αποτελέσματα παρουσιάζονται τον Πίνακα 4. Αυτό που παρατηρούμε είναι ότι οι μέσοι κινούνται πολύ κοντά για όλα τα μεγέθη M . Ενώ ούτε από τα μέγιστα ή τα ελάχιστα μπορούμε να αποφανθούμε για την καταλληλότητα κάποιου



Σχήμα 4: Εκτίμηση του ρυθμού εξόδου μέσω προσομοίωσης Monte Carlo.

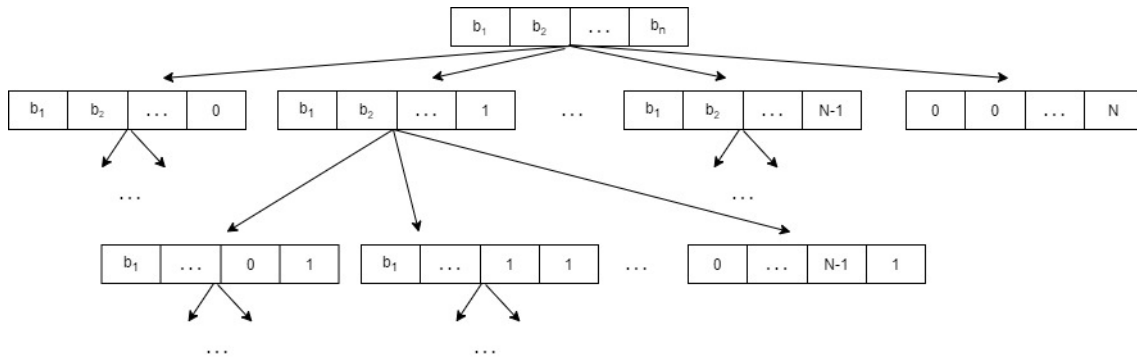
μεγέθους έναντι των υπολοίπων. Έτσι, για λόγους εξοικονόμησης χρόνου μια λογική επιλογή θα αποτελούσε το μικρότερο μέγεθος που δοκιμάσαμε, $M = 10$. Σημειώνουμε ότι στον Πίνακα 6 έχουν καταγραφεί όλες οι λύσεις που προέκυψαν για κάθε μέγεθος δείγματος.

Πίνακας 4: Σύγκριση τιμών της παραμέτρου M .

M	Average Throughput	Min	Max	Bactracking %	Average Time (min)
50	0.2376	0.2291	0.2395	29.8	62.32
40	0.2378	0.2291	0.2396	31.2	47.73
30	0.2380	0.2322	0.2395	18	32.34
20	0.2381	0.2339	0.2395	17	21
10	0.2374	0.2319	0.2396	18	10.65

Έχει αναφερθεί κατά την περιγραφή του αλγορίθμου ότι η διαμέριση κατέχει σημαντικό ρόλο στην απόδοσή του. Στόχος μίας αποδοτικής διαμέρισης είναι να ομαδοποιεί με κάποιον τρόπο τις λύσεις ώστε σημεία με κοντινές τιμές στην αντικειμενική συνάρτηση να κατηγοριοποιούνται στην ίδια υποπεριοχή. Για το παρόν πρόβλημα το κλειδί για την κατασκευή μίας αποδοτικής διαμέρισης βρίσκεται στον όρο «bottleneck». Αυτό τον όρο χρησιμοποιούμε για να αναφερθούμε στον server του δικτύου με τον μικρότερο ρυθμό εξυπηρέτησης. Ο συγκεκριμένος server μας ενδιαφέρει καθώς προκαλεί ή εντείνει φαινόμενα συμφόρησης στο δίκτυο. Προκειμένου να αποφευχθούν τέτοια φαινόμενα θα ήταν αναμενόμενο να ανατεθεί ένας ικανός

αριθμός θέσεων στον αντίστοιχο buffer.



Σχήμα 5: Εναλλακτική διαμέριση της εφικτής περιοχής.

Μία διαμέριση, για το παρόν πρόβλημα, είναι αποδοτική όταν ξεκινάει καθορίζοντας την χωρητικότητα του συγκεκριμένου buffer, όπως, άλλωστε και η διαμέριση που επιλέξαμε για τα ανωτέρω παραδείγματα. Αυτό συμβαίνει γιατί ο αλγόριθμος, έχοντας αναθέσει ένα σημαντικό ποσοστό των θέσεων στον συγκεκριμένο χώρο αναμονής, θα έχει λιγότερες διαθέσιμες θέσεις προς κατανομή για τους υπόλοιπους χώρους. Αυτό οδηγεί σε μικρότερο πλήθος υποπεριοχών από τις οποίες θα καλείται να πραγματοποιήσει δειγματοληψία και τελικά θα καλέσει λιγότερες φορές τη συνάρτηση προσομοίωσης γεγονός που μπορεί να μειώσει σημαντικά το χρόνο εκτέλεσης του αλγορίθμου.

Πίνακας 5: Αποτελέσματα τρίτου παραδείγματος

	$\lambda = \frac{1}{2}, \mu = [\frac{1}{3} \frac{1}{3} \frac{1}{3} \frac{1}{2} 1]$			
	Allocation	Average	Bactracking %	Average Time (min)
$M = 50$	3 3 3 1 0	0.2392	65.6	85.7
	2 4 3 1 0	0.2395	84.7	106.2
	2 2 6 0 0	0.2303	95	107.9
$M = 40$	2 3 4 1 0	0.2393	69.3	82.5
	2 4 3 1 0	0.2395	83.7	82.3
	4 2 3 1 0	0.2349	68.7	78.8
$M = 30$	2 4 3 1 0	0.2395	51.3	52.3
	3 3 3 1 0	0.2392	70.7	63.7
	2 4 4 0 0	0.2311	73.3	65.4
$M = 20$	2 3 4 1 0	0.2393	41.3	35.2
	3 2 2 3 0	0.2334	46.6	38.6
	2 3 2 3 0	0.2351	51	40
$M = 10$	3 3 3 1 0	0.2392	36	18
	1 4 4 1 0	0.2332	36.3	18.7
	1 4 4 1 0	0.2332	22.6	14.6

Εκτελέσαμε λοιπόν, τον αλγόριθμο βελτιστοποίησης για το δεύτερο παράδειγμα, αλλάζο-

ντας όμως αυτή τη φορά τη διαμέριση. Χρησιμοποιήθηκε η διαμέριση που απεικονίζεται στο Σχήμα 5, κατά την οποία ο πρώτος buffer που λαμβάνει τιμή είναι ο b_n , που είναι και ο πιο γρήγορος για το συγκεκριμένο παράδειγμα, καταλήγοντας στους πιο αργούς. Αυτό που θεωρητικά αναμένουμε είναι να πάρουμε μεν αντίστοιχα αποτελέσματα με αυτά των πρώτων εκτελέσεων σε μεγαλύτερο όμως υπολογιστικό χρόνο.

Πράγματι, αυτό που αναμέναμε αποτυπώθηκε στα αποτελέσματα όπως προκύπτει από τον Πίνακα 5. Κάτι ακόμα, αξιοσημείωτο, που προκύπτει από τον παραπάνω Πίνακα είναι τα ποσοστά οπισθοχώρησης. Παρατηρούμε ότι λαμβάνουν τιμές κατά πολύ μεγαλύτερες από αυτές που σημειώθηκαν με την πρώτη διαμέριση, που ήταν της τάξης του 20-30% για όλες τις τιμές του M . Αυτό οφείλεται στο γεγονός ότι ο αλγόριθμος κάνει πολύ συχνότερα διορθωτικές κινήσεις.

Πίνακας 6: Αναλυτικά αποτελέσματα του πρώτου παραδείγματος

$\lambda = \frac{1}{2}, \mu = [\frac{1}{3} \frac{1}{3} \frac{1}{3} \frac{1}{2} 1]$			
	b	Frequency	Average
$M = 50$	3 3 3 1 0	1	0.2392
	3 2 4 1 0	1	0.2383
	3 1 4 1 1	1	0.2291
	2 3 4 1 0	3	0.2393
	2 4 3 1 0	2	0.2395
	2 2 5 1 0	1	0.2360
	2 5 2 1 0	1	0.2367
$M = 40$	3 3 3 1 0	3	0.2392
	3 2 3 2 0	1	0.2387
	2 4 3 1 0	2	0.2395
	2 3 3 2 0	2	0.2396
	2 5 3 0 0	1	0.2302
	2 2 3 3 0	1	0.2352
$M = 30$	3 3 3 1 0	1	0.2392
	2 4 3 1 0	4	0.2395
	2 3 4 1 0	1	0.2393
	2 5 2 1 0	1	0.2367
	2 4 2 2 0	1	0.2384
	2 2 5 1 0	1	0.2360
	2 2 3 1 2	1	0.2322
$M = 20$	3 3 3 1 0	3	0.2392
	3 2 4 1 0	1	0.2383
	3 3 2 1 1	1	0.2339
	3 4 2 1 0	1	0.2368
	2 5 2 1 0	1	0.2367
	2 4 3 1 0	1	0.2395
	2 3 4 1 0	1	0.2393
	2 4 2 2 0	1	0.2384
$M = 10$	3 2 3 2 0	1	0.2387
	3 2 4 1 0	1	0.2383
	3 3 3 1 0	1	0.2392
	2 4 4 0 0	2	0.2319
	2 2 4 2 0	1	0.2369
	2 4 2 2 0	1	0.2384
	2 4 3 1 0	1	0.2395
	2 3 3 2 0	1	0.2396
	2 3 4 1 0	1	0.2393

5 Συμπεράσματα

Η βελτιστοποίηση μέσω προσομοίωσης παρέχει ισχυρά εργαλεία για την αντιμετώπιση μεγάλου πλήθους προβλημάτων βελτιστοποίησης. Συμπεριλαμβάνει μεθόδους και τεχνικές που μπορούν να απευθυνθούν σε πληθώρα προβλημάτων και προσαρμοζόμενες στις εκάστοτε χαρακτηριστικές δομές του προβλήματος μπορούν να δώσουν καλές λύσεις σε σύντομο χρονικό διάστημα.

Μέσα από τα παραδείγματα που προηγήθηκαν παρουσιάστηκε εκτενώς η μέθοδος εμφωλευμένων διαμερίσεων και έγινε φανερό ότι μπορεί να αντιμετωπίσει με σχετική επιτυχία ένα δύσκολο πρόβλημα διακριτής βελτιστοποίησης με μεγάλο αριθμό εφικτών λύσεων. Ο αλγόριθμος που αναπτύχθηκε δίνει αν όχι βέλτιστες, καλές λύσεις στο πρόβλημα που εξετάστηκε. Με κατάλληλη προσαρμογή των παραμέτρων ο αλγόριθμος μπορεί να ανταποκριθεί σε διαφορετικές ανάγκες σε ταχύτητα ή ακρίβεια της εκτιμώμενης λύσης.

Εκτός από το μεγάλο πλήθος ευρετικών μεθόδων που μπορούν να ενταχθούν στον αλγόριθμο για την βελτίωσή του, στην αποδοτικότερη λειτουργία του θα μπορούσαν να συνεισφέρουν σημαντικά θεωρητικά αποτελέσματα σχετικά με τους ρυθμούς σύγκλισης, την αξιολόγηση της παραγόμενης λύσης, αλλά και τα κριτήρια τερματισμού του αλγορίθμου. Τέτοια αποτελέσματα, για μία συγκεκριμένη μορφή της μεθόδου εμφωλευμένων διαμερίσεων, έχουν εξεταστεί από τους Shi and Olafsson [16]. Αυτά θα μπορούσαν να επεκταθούν για τις διάφορες παραλλαγές του αλγορίθμου εξασφαλίζοντας την αποδοτικότερη λειτουργία του.

Βιβλιογραφία

- [1] N Sofi, Dr Showkat Maqbool, and Aquil Ahmad. Mathematical programming: A brief historical sketch and its applications. *Mathematical Theory and Modeling*, 09 2020.
- [2] Jan A Snyman and Daniel N Wilke. *Practical Mathematical Optimization*. Number 978-3-319-77586-9 in Springer Optimization and Its Applications. Springer, May 2018. ISBN ARRAY(0x4e6dc3f0). doi: 10.1007/978-3-319-77586-9. URL <https://ideas.repec.org/b/spr/spopap/978-3-319-77586-9.html>.
- [3] J. Banks. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. A Wiley-Interscience publication. Wiley, 1998. ISBN 9780471134039. URL <https://books.google.gr/books?id=dMZ1Zj3TBgAC>.
- [4] Anu Maria. Introduction to modeling and simulation. In *Proceedings of the 29th conference on Winter simulation*, pages 7–13, 1997.
- [5] Satyajith Amaran, Nikolaos V. Sahinidis, Bikram Sharda, and Scott J. Bury. Simulation optimization: A review of algorithms and applications. *Annals of Operations Research*, 240(1):351–380, 2016. doi: 10.1007/s10479-015-2019-x. This is an updated version of the paper that appeared in 4OR, 12(4), 301–333 (2014).
- [6] Hrčka Lukáš, Važan Pavel, and Šutová Zuzana. Basic Overview of Simulation Optimization. *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, 22(341):1–6, December 2014. URL <https://ideas.repec.org/a/vrs/repfms/v22y2014i341p6n1.html>.
- [7] M. C. Fu, C. Chen, and L. Shi. Some topics for simulation optimization. In *2008 Winter Simulation Conference*, pages 27–38, 2008. doi: 10.1109/WSC.2008.4736053.
- [8] Farhad Azadivar. Simulation optimization methodologies. In *Proceedings of the 31st Conference on Winter Simulation: Simulation—a Bridge to the Future - Volume 1*, WSC '99, page 93–100, New York, NY, USA, 1999. Association for Computing Machinery. ISBN 0780357809. doi: 10.1145/324138.324168. URL <https://doi.org/10.1145/324138.324168>.

- [9] Michael C. Fu, editor. *Handbook of Simulation Optimization*. Springer, New York, NY, 2015. doi: <https://doi.org/10.1007/978-1-4939-1384-8>.
- [10] S. Andradottir. A review of simulation optimization techniques. In *1998 Winter Simulation Conference. Proceedings (Cat. No.98CH36274)*, volume 1, pages 151–158 vol.1, 1998. doi: 10.1109/WSC.1998.744910.
- [11] Jeffrey Joines, R. Barton, Keebom Kang, Paul Fishwick, James Swisher, Sheldon Jacobson, Paul Hyden, and Lee Schruben. A survey of simulation optimization techniques and procedures. 01 2001.
- [12] Anton J. Kleywegt, Alexander Shapiro, and Tito Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002. doi: 10.1137/S1052623499363220. URL <https://doi.org/10.1137/S1052623499363220>.
- [13] Robert Michael Lewis, Virginia Torczon, and Michael W. Trosset. Direct search methods: then and now. *Journal of Computational and Applied Mathematics*, 124(1):191–207, 2000. ISSN 0377-0427. doi: [https://doi.org/10.1016/S0377-0427\(00\)00423-4](https://doi.org/10.1016/S0377-0427(00)00423-4). URL <https://www.sciencedirect.com/science/article/pii/S0377042700004234>. Numerical Analysis 2000. Vol. IV: Optimization and Nonlinear Equations.
- [14] Robert Hooke and T. A. Jeeves. “direct search” solution of numerical and statistical problems. *J. ACM*, 8(2):212–229, April 1961. ISSN 0004-5411. doi: 10.1145/321062.321069. URL <https://doi.org/10.1145/321062.321069>.
- [15] J M García, C A Acosta, and M J Mesa. Genetic algorithms for mathematical optimization. *Journal of Physics: Conference Series*, 1448:012020, jan 2020. doi: 10.1088/1742-6596/1448/1/012020. URL <https://doi.org/10.1088/1742-6596/1448/1/012020>.
- [16] Leyuan Shi and Sigurdur Olafsson. Convergence rate of the nested partitions method for stochastic optimization. 04 1997. doi: 10.1007/978-0-387-71909-2_2.
- [17] Leyuan Shi and Sigurdur Ólafsson. An integrated framework for deterministic and stochastic optimization. In *Proceedings of the 29th Conference on Winter Simulation*,

WSC '97, page 358–365, USA, 1997. IEEE Computer Society. ISBN 078034278X. doi: 10.1145/268437.268508. URL <https://doi.org/10.1145/268437.268508>.

- [18] Leyuan Shi. A new algorithm for stochastic discrete resource allocation optimization. *Discrete Event Dynamic Systems*, 10(3):271–294, 2000. doi: 10.1023/A:1017214011352.
- [19] Leyuan Shi and Sigurdur Ólafsson. Nested partitions optimization. *INFORMS Tutorials in Operations Research*, pages 1–22, 10 2014. doi: 10.1287/educ.1073.0033.
- [20] Frederico Cruz, Anderson Duarte, and Clecio Klechen. Buffer allocation in general queueing networks. 10 2005.
- [21] Sixiao Gao, Toshimitsu Higashi, Toyokazu Kobayashi, Kosuke Taneda, Jose I. U. Rubrico, and Jun Ota. Buffer allocation via bottleneck-based variable neighborhood search. *Applied Sciences*, 10(23), 2020. ISSN 2076-3417. doi: 10.3390/app10238569. URL <https://www.mdpi.com/2076-3417/10/23/8569>.