

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΧΩΡΙΣ ΤΗ ΧΡΗΣΗ ΠΑΡΑΓΩΓΟΥ ΣΕ ΠΡΟΒΛΗΜΑΤΑ ΜΕΓΑΛΗΣ ΚΛΙΜΑΚΑΣ

Καραγιάννης Μηνάς

Αρ. μητρώου: ge16034

Επιβλέπων Καθηγητής: Χρυσάφινος Κωνσταντίνος

Ημερομηνία: 25 Σεπτεμβρίου 2024



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

*Στον Βασίλη - Μιχαήλ,
και στην Ελπίδα*

Ευχαριστίες

Πρωτίστως, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Χρυσάφινο Κωνσταντίνο, για την εμπιστοσύνη που μου έδειξε για την εκπόνηση της παρούσας Διπλωματικής εργασίας, καθώς και για την καθοδήγηση που παρείχε καθ' όλη τη διάρκεια της συνεργασίας μας.

Έπειτα, θα ήθελα να ευχαριστήσω τον Πέτρο. Οι συζητήσεις μας επί του θέματος είχαν καταλυτικό ρόλο για την εκπόνηση της εργασίας. Εύχομαι να είναι πάντα υγιής.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου, τους φίλους μου (ιδιαίτερα τον Γιώργο), και την Ελπίδα. Η στήριξη που έλαβα από όλους, και ειδικά από τους δύο τελευταίους, μου ήταν πολύτιμη.

Περιεχόμενα

Περίληψη	1
Εισαγωγή	3
1 Οικογένειες Μεθόδων Βελτιστοποίησης	5
1.1 Μέθοδοι Αναζήτησης Γραμμής	5
1.2 Μέθοδοι Αναζήτησης Περιοχής Εμπιστοσύνης Περιγραφή της Προσέγγισης για την Περιοχή Εμπιστοσύνης Κλιμακοποίηση της Μεθόδου σε άλλες Νόρμες	9 13 14
1.3 Μέθοδοι Συζυγών Κλίσεων Ρυθμός Σύγκλισης της Μεθόδου Προετοιμασία της Μεθόδου Συζυγείς κλίσεις και Μη Γραμμικά Προβλήματα - Η μέθοδος Polak Ribière	15 21 22 23
1.4 Μέθοδοι Quasi - Newton	23
2 Μέθοδοι Βελτιστοποίησης χωρίς τη χρήση Παραγώγων	27
2.1 Μέθοδος Πεπερασμένων Διαφορών	27
2.2 Μέθοδοι Ανάπτυξης Μοντέλου Παρεμβολή και Πολυωνυμικές Βάσεις Αναβάθμιση του Συνόλου Παρεμβολής Μία Μέθοδος βασισμένη στις Μεθόδους Quasi - Newton	29 30 31 32
2.3 Μέθοδοι Συντεταγμένων και Μέθοδοι Αναζήτησης Μοτίβου Μέθοδοι Συντεταγμένων Μέθοδοι Αναζήτησης Μοτίβου	35 35 35
2.4 Μέθοδοι Συζυγών Κλίσεων	41
2.5 Μέθοδοι Nelder - Mead Σύγκλιση της μεθόδου Επανεκκίνηση της μεθόδου	45 50 53
3 Αριθμητικά Παραδείγματα	56
3.1 Παρουσίαση και περιγραφή των μεθόδων Μέθοδος Pattern Search Μέθοδος Συζυγών Κλίσεων Μέθοδος Nelder-Mead	56 56 58 60
3.2 Ανάλυση των αποτελεσμάτων Μέθοδος Pattern Search Μέθοδος Συζυγών κλίσεων Μέθοδος Nelder-Mead	63 63 66 68

Περίληψη

Στην παρούσα διπλωματική εργασία γίνεται μία μελέτη των πιο διαδεδομένων μεθόδων για την επίλυση προβλημάτων βελτιστοποίησης χωρίς τη χρήση παραγώγων (Μέθοδοι DFO). Αυτές οι μέθοδοι ενδείκνυνται για προβλήματα μεγάλης κλίμακας καθώς ο υπολογισμός των παραγώγων αφενός είναι απαιτητικός, ενώ παράλληλα λόγω των σφαλμάτων που περιέχονται στα δεδομένα είναι βέβαιο πως ούτως ή άλλως θα οδηγηθούμε σε μία προσέγγιση της λύσης.

Στην εισαγωγή αναφέρονται κάποια βασικά στοιχεία της Μαθηματικής Βελτιστοποίησης και των αλγορίθμων που χρησιμοποιούνται για την επίλυση προβλημάτων βελτιστοποίησης πάσης φύσεως.

Στο πρώτο μέρος γίνεται μία ανασκόπηση των μεθόδων βελτιστοποίησης με χρήση παραγώγων. Συγκεκριμένα περιγράφονται οι Μέθοδοι Αναζήτησης Γραμμής (Line Search Methods), οι Μέθοδοι Αναζήτησης Περιοχής Εμπιστοσύνης (Trust Region Methods), οι Μέθοδοι Συζυγών Κλίσεων (Conjugate Gradients Methods) και οι μέθοδοι Quasi - Newton. Δίνονται τα σημαντικά στοιχεία για κάθε μέθοδο, όπως η φιλοσοφία τους, θεωρήματα που αφορούν τη σύγκλιση και σημαντικές ιδιότητες, καθώς και μία σύντομη περιγραφή των αλγορίθμων. Έτσι, προετοιμάζεται το έδαφος για τη μελέτη των Μεθόδων DFO, οι οποίες βασίζονται ακριβώς σε αυτές τις μεθόδους.

Στο δεύτερο μέρος γίνεται μία ενδελεχή μελέτη των πιο διαδεδομένων Μεθόδων DFO. Αρχικά παραθέτουμε την πιο απλή Μέθοδο DFO, τη μέθοδο Πεπερασμένων Διαφορών (Finite Differences Method). Στη συνέχεια μελετάμε τις Μεθόδους Ανάπτυξης Μοντέλου (Model Based Methods), τις Μεθόδους Αναζήτησης Συντεταγμένων (Coordinate Search Methods) και τις Μεθόδους Αναζήτησης Μοτίβου (Pattern Search Methods), τις Μεθόδους Συζυγών Κατευθύνσεων (Conjugate Directions Method), και ολοκληρώνουμε με τη μέθοδο των Nelder και Mead (Nelder-Mead Simplex Method). Όπως και στο πρώτο μέρος, δίνουμε μία σύντομη περιγραφή του αλγορίθμου, και συμπληρώνουμε με τις απαραίτητες θεωρητικές πληροφορίες.

Το τρίτο μέρος αποτελείται από αριθμητικά παραδείγματα, υλοποιώντας κάθε μία από τις μεθόδους του δεύτερου μέρους με δύο δοκιμαστικές συναρτήσεις, οι διαστάσεις των οποίων κυμαίνονται από $n = 20$ έως $n = 200$, ενώ στη Μέθοδο Συζυγών Κλίσεων πραγματοποιήθηκε ένα παράδειγμα διάστασης $n = 10.000$ μεταβλητών. Εξετάζουμε τη συμπεριφορά της κάθε μεθόδου και αναλύουμε τα σχετικά αποτελέσματα.

Εισαγωγή

Με τον όρο Βελτιστοποίηση, εννοούμε την εύρεση της καλύτερης δυνατής λύσης σε ένα δεδομένο πρόβλημα. Αυτή είναι μία διαδικασία που συμβαίνει παντού, από τα καθημερινά ανθρώπινα προβλήματα, μέχρι και στα πιο περίπλοκα φυσικά συστήματα.

Προβλήματα Μαθηματικής Βελτιστοποίησης αποτελούν αναπόσπαστο κομμάτι των φυσικών επιστημών. Η συνολική διαδικασία μπορεί να χωριστεί σε συγκεκριμένα μέρη, όλα τους εξίσου σημαντικά.

Καταρχάς ερχόμαστε αντιμέτωποι με τη μοντελοποίηση του προβλήματος, τον ορισμό δηλαδή των σημαντικότερων παραμέτρων του. Ορίζουμε ως μεταβλητές τους παράγοντες που επηρεάζουν το αποτέλεσμα του προβλήματος καθώς αυτοί μεταβάλλονται. Έπειτα πρέπει να προσδιορίσουμε την αντικειμενική συνάρτηση, τον τρόπο με τον οποίο ποσοτικοποιείται το αποτέλεσμα του προβλήματος συναρτήσει των μεταβλητών. Τέλος, πρέπει να ερευνήσουμε και να διατυπώσουμε τις εξισώσεις ή τις ανισώσεις που περιγράφουν τους περιορισμούς στους οποίους υπόκεινται οι μεταβλητές και η αντικειμενική συνάρτηση. Μόνο εφόσον έχουν υλοποιηθεί τα παραπάνω βήματα επιτυχώς, μπορούμε να προχωρήσουμε στη διαδικασία βελτιστοποίησης του προβλήματος.

Αξίζει να αναφερθεί ότι η μοντελοποίηση ενός προβλήματος μπορεί να γίνει με πολλούς διαφορετικούς τρόπους. Μπορεί να ασχολούμαστε με το ίδιο σύστημα, αλλά να έχουμε κάθε φορά διαφορετικούς στόχους, οπότε να αλλάζει η φύση του προβλήματος. Μπορεί να χρειαστεί να επιλέξουμε μια ακριβέστερη περιγραφή του προβλήματος, ώστε να έχουμε και πιο ακριβή αποτελέσματα που θα παρέχουν βαθύτερη κατανόηση του προβλήματος. Από την άλλη, μπορούμε να 'θυσιάσουμε' την ακρίβεια των αποτελεσμάτων μέσω μίας απλοϊκότερης περιγραφής, εν όψει οικονομίας χώρου και χρόνου, ή απλά επειδή μια πιο εκλεπτυσμένη μοντελοποίηση υπερβαίνει τις εκάστοτε απαιτήσεις μας.

Η σημασία των περιορισμών έγκειται, ως συνήθως, στον προσδιορισμό της εφικτής περιοχής. Η εφικτή περιοχή ορίζει το χώρο των αποδεκτών λύσεων. Η βέλτιστη λύση, οι τιμές των μεταβλητών για τις οποίες η αντικειμενική συνάρτηση του ίδιου προβλήματος παίρνει τη μέγιστη (ή την ελάχιστη) τιμή της μπορεί να διαφέρει ανάλογα με το αν υπάρχουν περιορισμοί και ποιοι είναι κάθε φορά.

Τα προβλήματα βελτιστοποίησης μπορούν να καταταχθούν σε ορισμένες κατηγορίες οι οποίες διαμορφώνουν τη στρατηγική επίλυσης και την επιλογή του αλγορίθμου. Συνοπτικά, υπάρχουν οι εξής κατηγορίες:

- **Συνεχής ή Διακριτή Βελτιστοποίηση**, ανάλογα με το είδος των μεταβλητών, αν είναι δηλαδή συνεχείς ή διακριτές.
- **Βελτιστοποίηση με ή χωρίς περιορισμούς.**
- **Ολική ή Τοπική Βελτιστοποίηση**, αναλόγως αν ψάχνουμε το ολικό ακρότατο της αντικειμενικής συνάρτησης, ή αν ακρούμαστε σε μία τοπική διερεύνηση. Προφανώς το ολικό ακρότατο αποτελεί την ιδανική λύση στο πρόβλημα. Ωστόσο, πολύ συχνά υπάρχουν δυσκολίες στην εύρεση του ολικού ακρότατου, οπότε η εύρεση ενός τοπικού ακρότατου, του ακρότατου εντός μίας προκαθορισμένης περιοχής του χώρου (της εφικτής περιοχής), είναι μία απολύτως αποδεκτή και δικαιολογημένη στρατηγική επίλυσης.
- **Στοχαστική ή Ντετερμινιστική Βελτιστοποίηση.** Σε πολλές περιπτώσεις υπάρχει ελλιπής γνώση - όχι δεδομένα - και αβεβαιότητα εμποδίζοντας έτσι τον άρτιο προσδιορισμό του επίμαχου προβλήματος. Με άλλα λόγια, υπάρχει η πιθανότητα η αντικειμενική συνάρτηση του προβλήματος να επηρεάζεται από μεταβλητές οι οποίες δεν είναι γνωστές κατά τη διάρκεια της μοντελοποίησης. Θα πρέπει λοιπόν να ακολουθήσουμε μία στοχαστική διαδικασία για τη μοντελοποίηση του προβλήματος. Να μην αρχεστούμε δηλαδή σε μία εικασία, αλλά να αξιοποιήσουμε

προϋπάρχοντες πιθανοτικές εκτιμήσεις σχετικές με τις μεταβλητές και την αντικειμενική συνάρτηση.

- Μία ακόμα σημαντική συνθήκη που αλλάζει πλήρως την προσέγγιση ενός προβλήματος είναι η **κυρτότητα** της αντικειμενικής συνάρτησης. Η κυρτότητα συνεπάγεται ότι κάθε τοπικό ακρότατο της αντικειμενικής συνάρτησης είναι ταυτόχρονα και ολικό ακρότατο, επομένως διευκολύνεται σε μεγάλο βαθμό η επίλυση του προβλήματος. Σε επόμενο κεφάλαιο θα γίνει μία αναλυτικότερη παρουσίαση των βασικών εννοιών και θεωρημάτων που αφορούν την κυρτότητα.

Οι αλγόριθμοι βελτιστοποίησης είναι επαναληπτικές διαδικασίες. Κατά την εκκίνηση του αλγορίθμου γίνεται μία αρχική εκτίμηση της λύσης, και στη συνέχεια, κάθε επανάληψη υπολογίζει μία καλύτερη προσέγγιση της λύσης, έως ότου ικανοποιηθεί μία συνθήκη τερματισμού. Τόσο οι πράξεις για την εύρεση της επόμενης προσέγγισης, όσο και η συνθήκη τερματισμού διαφέρουν από μέθοδο σε μέθοδο. Συνήθως οι μέθοδοι αξιοποιούν τιμές της αντικειμενικής συνάρτησης και της πρώτης και δεύτερης παραγώγου της. Ορισμένες μέθοδοι αποθηκεύουν στοιχεία από προηγούμενες επαναλήψεις για να τις αξιοποιήσουν στις επόμενες, ενώ άλλες αξιοποιούν στοιχεία μόνο από την τρέχουσα επανάληψη. Σε κάθε περίπτωση, μία καλή μέθοδος χαρακτηρίζεται από τα εξής χαρακτηριστικά:

- **Προσαρμοστικότητα:** Οι μέθοδοι πρέπει να προσαρμόζονται καλά σε ποικίλα προβλήματα, ώστε να έχουν "αξία χρήσης".
- **Αποδοτικότητα:** Οι μέθοδοι πρέπει να δίνουν ένα ικανοποιητικό αποτέλεσμα όσο είναι δυνατόν πιο γρήγορα (δηλαδή με λίγες επαναλήψεις), χωρίς να καταλαμβάνουν αποθηκευτικό χώρο, και δίχως πολλές πράξεις.
- **Ακρίβεια:** Οι μέθοδοι πρέπει να καταλήγουν σε μία ικανοποιητική λύση, η οποία να είναι ανθεκτική στα σφάλματα, είτε αυτά είναι σφάλματα στρογγυλοποίησης κατά τη διάρκεια των υπολογιστικών πράξεων, είτε είναι προϋπάρχοντα σφάλματα στα δεδομένα μας.

Είναι εξαιρετικά σπάνιο ένας αλγόριθμος να εμφανίζει και τα τρία χαρακτηριστικά σε υψηλό βαθμό. Μάλιστα, τα χαρακτηριστικά είναι συνήθως αντικρουόμενα. Για παράδειγμα, ένας αλγόριθμος μπορεί να είναι πολύ αποδοτικός και ακριβής, αλλά μόνο σε συγκεκριμένο τύπο προβλημάτων. Ένας άλλος μπορεί να έχει ευρύτερη χρήση, αλλά να απαιτούνται πολλές πράξεις για να καταλήξει σε κάποιο ικανοποιητικό αποτέλεσμα. Ενώ ένας τρίτος αλγόριθμος μπορεί να είναι ταχύτατος, αλλά το αποτέλεσμα να είναι ευαίσθητο σε σφάλματα. Οπότε, γίνεται κατανοητό πως η επιλογή του κατάλληλου αλγορίθμου εξαρτάται από τι συνθήκες του προβλήματος και τις απαιτήσεις της εργασίας.

Τέλος, η εις βάθος γνώση της θεωρίας λειτουργεί καταλυτικά σε όλους τους σχετικούς τομείς: τον σχεδιασμό των αλγορίθμων, την επιλογή του κατάλληλου αλγορίθμου, την ανάλυση των αποτελεσμάτων, και ούτω καθεξής. Όσο πληρέστερη είναι η κατανόηση γύρω από τα θεωρητικά μοντέλα, τόσο πιο εμπειριστατωμένα είναι και τα αποτελέσματα της εργασίας. Το πνεύμα της παρούσας εργασίας είναι διαμορφωμένο τοιούτοτρόπως.

1 Οικογένειες Μεθόδων Βελτιστοποίησης

1.1 Μέθοδοι Αναζήτησης Γραμμής

Η παράγραφος αυτή βασίζεται στο ([1], σελ.30-65). Μία από τις πιο γνωστές κατηγορίες των μεθόδων βελτιστοποίησης είναι οι Μέθοδοι Αναζήτησης Γραμμής (Line Search Methods). Σε κάθε επανάληψη των μεθόδων αυτών υπολογίζεται πρώτα μία κατεύθυνση αναζήτησης p_k , και ύστερα ένα βήμα $a_k > 0$ που υποδεικνύει πόσο θα μετακινηθούμε προς την κατεύθυνση του p_k . Έτσι, σε κάθε επανάληψη υπολογίζεται ένα νέο σημείο

$$x_{k+1} = x_k + a_k p_k.$$

Προκειμένου να έχουμε έναν αποδοτικό αλγόριθμο από άποψη χρόνου και ακρίβειας είναι σημαντικό να επιλέγουμε κατάλληλα αυτές τις δύο παραμέτρους.

Ας δούμε αρχικά πώς επιλέγουμε την κατεύθυνση p_k . Μία πρώτη σκέψη είναι να απαιτήσουμε για την p_k να είναι μία κατεύθυνση καθόδου, να εξασφαλίζει δηλαδή ότι η τιμή της συνάρτησης θα μειωθεί ανεξάρτητα από την επιλογή του βήματος a_k . Η παραπάνω απαίτηση μπορεί να περιγραφεί με τη σχέση

$$p_k^T \nabla f_k < 0.$$

Πράγματι, αναπτύσσοντας το $f(x_{k+1})$ με τη χρήση του θεωρήματος Taylor σε όρους μέχρι και πρώτης τάξης έχουμε τα εξής:

$$\begin{aligned} f(x_{k+1}) &= f(x_k + a_k p_k) = f(x_k) + a_k p_k^T \nabla f(x_k) \\ f(x_{k+1}) - f(x_k) &= a_k \cdot (p_k^T \nabla f(x_k)) < 0 \end{aligned}$$

από όπου έπεται και το επιθυμητό αποτέλεσμα, $f(x_{k+1}) < f(x_k)$.

Συνήθως η p_k έχει τη μορφή

$$p_k = -B_k^{-1} \nabla f_k,$$

όπου B_k είναι ένας συμμετρικός και αντιστρέψιμος πίνακας. Ενδεικτικά, στη μέθοδο μέγιστης καθόδου ο B_k είναι ο μοναδιαίος πίνακας I , στις μεθόδους Newton είναι ο εσσιανός πίνακας $\nabla^2 f(x_k)$, ενώ στις μεθόδους Quasi - Newton είναι μία προσέγγιση του εσσιανού. Πράγματι, εάν θεωρήσουμε ότι ο πίνακας B_k είναι θετικά ορισμένος, τότε έχουμε:

$$\begin{aligned} p_k^T \nabla f_k &= -\nabla f_k^T (B_k^{-1})^T \nabla f_k \\ &= -\nabla f_k^T B_k^{-1} \nabla f_k < 0 \end{aligned}$$

οπότε η p_k είναι σίγουρα κατεύθυνση καθόδου.

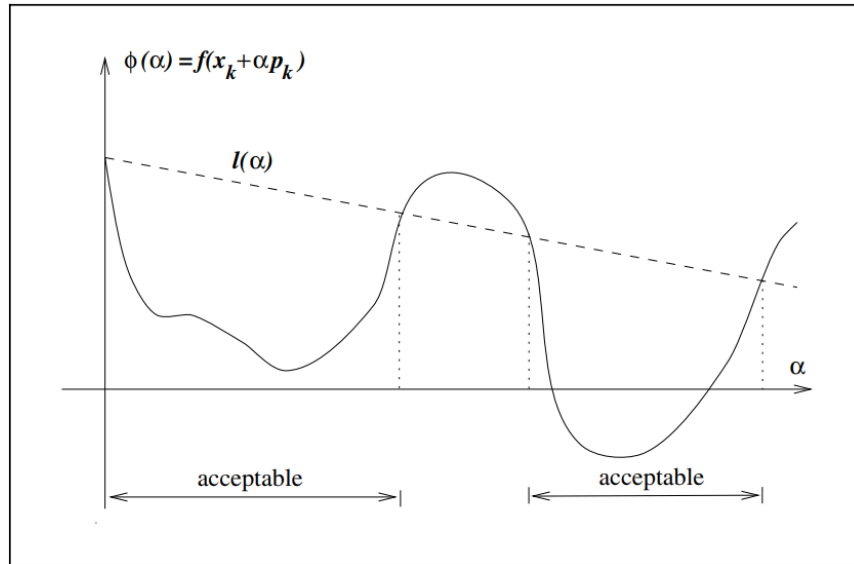
Όσον αφορά το βήμα a_k τώρα, ερχόμαστε αντιμέτωποι με ένα δίλημμα. Όσο πιο ακριβής είναι ο υπολογισμός του καλύτερου βήματος a_k , τόσο περισσότερες είναι και οι πράξεις που θα πρέπει να εκτελέσουμε. Θα πρέπει λοιπόν να βρούμε μία καλή ισορροπία μεταξύ αυτών των δύο επιλογών. Ίδανικά, θα θέλαμε να βρούμε το βήμα που ελαχιστοποιεί τη συνάρτηση κατά μήκος της γραμμής αναζήτησης. Με άλλα λόγια, θα θέλαμε να βρούμε το ολικό ελάχιστο της συνάρτησης

$$\varphi(a) = f(x_k + a p_k), \quad a > 0$$

κάτι που όμως είναι χρονοβόρο όπως είπαμε προηγουμένως. Εξίσου απαγορευτική είναι η προσπάθεια να βρούμε ακόμα και ένα τοπικό ελάχιστο της συνάρτησης φ . Στην πράξη είναι προτιμότερο από θέμα

χρόνου να πραγματοποιήσουμε την αναζήτηση σε μία λιγότερο ακριβή γραμμή, και πάνω σε αυτή να βρούμε ένα βήμα που να επιτυγχάνει μία ικανοποιητική μείωση της f .

Γενικά η αναζήτηση του βήματος χωρίζεται σε δύο στάδια: Στο πρώτο στάδιο βρίσκουμε ένα διάστημα που περιέχει επιθυμητά βήματα. Το δεύτερο στάδιο περιλαμβάνει μία διαδικασία διχοτόμησης ή παρεμβολής, με σκοπό τον υπολογισμό ενός ικανοποιητικού βήματος εντός αυτού του διαστήματος. Είναι εξίσου σημαντικό να ορίσουμε τότε το βήμα μας είναι ικανοποιητικό, ώστε να προχωρήσουμε παρακάτω με τη μέθοδό μας. Οι συνθήκες που πρέπει να ικανοποιούνται ώστε να θεωρηθεί ικανοποιητικό το a_k ώστε να προχωρήσουμε παρακάτω στη μέθοδο ονομάζονται συνθήκες τερματισμού της τεχνικής γραμμικής αναζήτησης.



Σχήμα 1.1: Συνθήκη Επαρκούς Μείωσης (συνθήκη Armijo).

Η πιο εύλογη συνθήκη τερματισμού είναι η απαίτηση η τιμή της συνάρτησης να μειώνεται, δηλαδή $f(x_k + a_k p_k) < f(x_k)$. Αυτή η συνθήκη, παρά την απλότητα της, δυστυχώς είναι αρκετά αναποτελεσματική. Δεν μπορεί να εξασφαλίσει με καμία βεβαιότητα τη σύγκλιση προς κάποιο ελάχιστο, καθώς η μείωση της συνάρτησης μπορεί να συμβαίνει με τρομερά αργούς ρυθμούς. Αμέσως διαπιστώνουμε την ανάγκη να ορίσουμε κάποιες συνθήκες επαρκούς μείωσης της συνάρτησης, για να επιτύχουμε και καλύτερη σύγκλιση της μεθόδου.

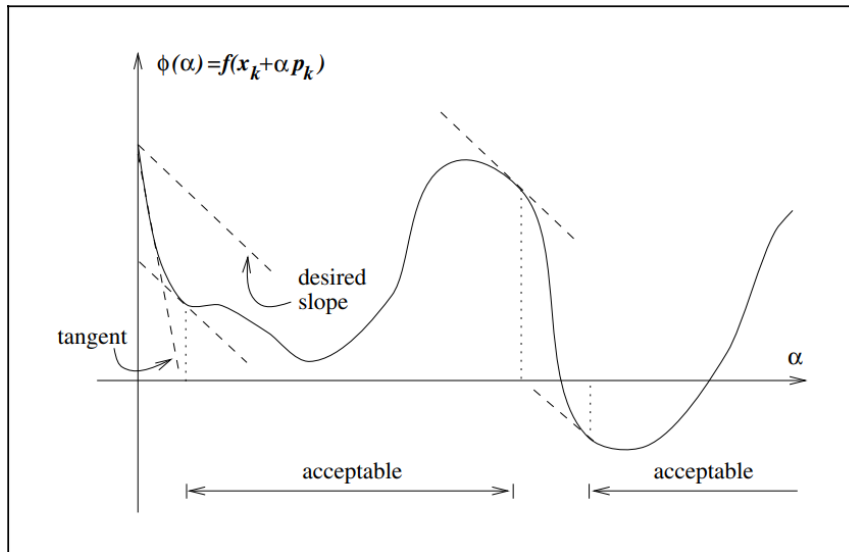
Από τις πιο δημοφιλείς συνθήκες επαρκούς μείωσης είναι οι συνθήκες Wolfe:

$$\begin{aligned} f(x_k + a_k p_k) &\leq f(x_k) + c_1 a_k \nabla f_k^T p_k \\ \nabla f(x_k + a_k p_k)^T p_k &\geq c_2 \nabla f_k^T p_k \\ 0 &< c_1 < c_2 < 1 \end{aligned} \quad (1.1)$$

Η πρώτη από τις συνθήκες (γνωστή και ως συνθήκη Armijo) απαιτεί η μείωση στη συνάρτηση f να είναι ανάλογη τόσο του βήματος a_k όσο και της κατά κατεύθυνση παραγώγου $\nabla f_k^T p_k$. Για τη σταθερά c_1 ισχύει $c_1 \in (0, 1)$. Στο σχήμα 1.1 φαίνεται μία γεωμετρική περιγραφή της συνθήκης. Ορίζουμε το δεύτερο μέλος της πρώτης σχέσης ως $l(a) = f(x_k) + c_1 a \nabla f_k^T p_k$. Όπως φαίνεται και στο σχήμα,

είναι πολύ εύκολο να εντοπίσουμε τα διαστήματα που προφέρουν επαρκή μείωση στην αντικειμενική συνάρτηση. Όμως η πρώτη συνθήκη δεν είναι αρκετή από μόνη της για να εξασφαλίσει την πρόοδο του αλγορίθμου, μιας και συμπεριλαμβάνει τιμές του a_k που βρίσκονται κοντά στο μηδέν (είναι δηλαδή τιμές πλησίον του αρχικού μας σημείου x_k).

Το πρόβλημα αυτό λύνει η δεύτερη συνθήκη, η οποία ονομάζεται συνθήκη καμπυλότητας. Εδώ απαιτείται η κλίση της συνάρτησης $\varphi(a_k) = f(x_k + a_k p_k)$ (η οποία ισούται με το πρώτο μέλος της δεύτερης σχέσης) να είναι μεγαλύτερη της αρχικής κλίσης $\varphi'(0)$ κατά c_2 φορές, όπου $c_2 \in (c_1, 1)$. Η συνθήκη προσφέρει δύο πλεονεκτήματα. Αφενός, όταν η κλίση $\varphi'(a_k)$ είναι 'πολύ' αρνητική, τότε έχουμε μια ένδειξη ότι η συνάρτηση f μπορεί να ελαττωθεί περισσότερο αν προχωρήσουμε ακόμα πιο μακριά στην συγκεκριμένη κατεύθυνση. Αφετέρου, όταν η κλίση $\varphi'(a_k)$ είναι 'λίγο' αρνητική, ή ακόμα και θετική, τότε έχουμε μια ένδειξη ότι η συνάρτηση f θα παρουσιάζει μικρή μείωση, οπότε είναι φρόνιμο να τερματίσουμε την αναζήτηση. Οι τιμές της σταθεράς c_2 εξαρτώνται από τη μέθοδο που ακολουθούμε. Για παράδειγμα, όταν έχουμε βρει τις κατευθύνσεις p_k μέσω μίας μεθόδου Newton ή Quasi - Newton δίνουμε συνήθως στη σταθερά την τιμή 0,9. Το παρακάτω σχήμα 1.2 περιγράφει τη λειτουργία της συνθήκης καμπυλότητας, ανεξάρτητα από τη συνθήκη Armijo. Συναληθεύοντας τις επιτρεπόμενες τιμές για το βήμα a_k καθειμάς από τις συνθήκες, παίρνουμε τις δεκτές τιμές των συνθηκών Wolfe, όπως φαίνεται και στο σχήμα 1.3

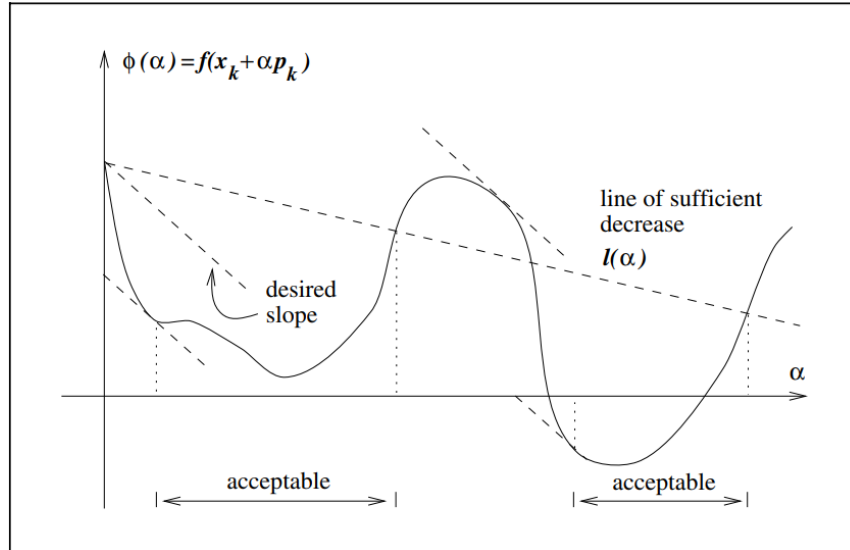


Σχήμα 1.2: Συνθήκη Καμπυλότητας.

Υπάρχουν επίσης και οι αυστηρές συνθήκες Wolfe, στις οποίες ουσιαστικά έχουμε προσαρμόσει της συνθήκη καμπυλότητας, ούτως ώστε το βήμα μας a_k να έγκειται τουλάχιστον σε μία ευρεία περιοχή ενός τοπικού ελαχίστου ή ενός στάσιμου σημείου της συνάρτησης φ . Έτσι, οι ισχυρές συνθήκες Wolfe γράφονται ως εξής:

$$\begin{aligned}
 f(x_k + a_k p_k) &\leq f(x_k) + c_1 a_k \nabla f_k^T p_k \\
 |\nabla f(x_k + a_k p_k)^T p_k| &\geq c_2 |\nabla f_k^T p_k| \\
 0 < c_1 < c_2 < 1
 \end{aligned}
 \tag{1.2}$$

Όπως φαίνεται από τη δεύτερη συνθήκη, η διαφορά που εν τέλει αποκλείει τα σημεία που βρίσκονται μακριά από τα στάσιμα σημεία της φ είναι ότι τώρα δεν επιτρέπουμε στην παράγωγο $\varphi'(a_k)$ να είναι 'πολύ' θετική, με τη χρήση των απολύτων τιμών.



Σχήμα 1.3: Βήματα που ικανοποιούν τις συνθήκες Wolfe.

Προσθέτοντας κάποιες τελευταίες σημειώσεις, αξίζει να αναφέρουμε ότι οι συνθήκες Wolfe είναι αναλλοίωτες σε μετασχηματισμούς συντεταγμένων (scaling), πράγμα που τις καθιστά χρήσιμες σε όλες τις μεθόδους, ειδικά τις Quasi - Newton. Τέλος, αποδεικνύεται ότι για κάθε συνάρτηση f που είναι ομαλή και κάτω φραγμένη υπάρχει πάντα ένα βήμα που ικανοποιεί τις συνθήκες Wolfe, όπως φαίνεται στο παρακάτω Λήμμα ([1], Λήμμα 3.1):

Λήμμα 1.1.

Έστω $f : \mathbb{R}^n \rightarrow \mathbb{R}$ μία συνεχώς διαφορίσιμη συνάρτηση. Έστω p_k μία κατεύθυνση καθόδου στο x_k , και έστω ότι η f είναι κάτω φραγμένη κατά μήκος της ευθείας $\{x_k + \alpha p_k | \alpha > 0\}$. Τότε, αν $0 < c_1 < c_2 < 1$, υπάρχουν διαστήματα για το βήμα a που ικανοποιούν τις συνθήκες Wolfe (1.1) και τις ισχυρές συνθήκες Wolfe (1.2).

Απόδειξη. Έστω η συνάρτηση $\varphi(a) = f(x_k + \alpha p_k)$ που είναι κάτω φραγμένη για κάθε $a > 0$. Αφού $0 < c_1 < 1$, τότε η ευθεία $l(a) = f(x_k) + a c_1 \nabla f_k^T p_k$ δεν είναι κάτω φραγμένη, και άρα θα έχει τουλάχιστον ένα σημείο τομής με τη συνάρτηση φ . Έστω $a' > 0$ η μικρότερη τιμή του a που συμβαίνει αυτό, δηλαδή

$$f(x_k + a' p_k) = f(x_k) + a' c_1 \nabla f_k^T p_k. \tag{1.3}$$

Η συνθήκη επαρκούς μείωσης (1.1) ισχύει για κάθε βήμα μικρότερο του a' .

Άρα, από το Θεώρημα Μέσης Τιμής, υπάρχει ένα $a'' \in (0, a')$ τέτοιο ώστε

$$f(x_k + a' p_k) - f(x_k) = a' \nabla f(x_k + a'' p_k)^T p_k. \tag{1.4}$$

Συνδυάζοντας τις εξισώσεις (1.3) και (1.4) έχουμε

$$\nabla f(x_k + a''p_k)^T p_k = c_1 \nabla f_k^T p_k > c_2 \nabla f_k^T p_k, \quad (1.5)$$

μιας και $c_1 < c_2$ και $\nabla f_k^T p_k < 0$. Έτσι, το a'' ικανοποιεί τις συνθήκες Wolfe (1.1). Οπότε, λόγω της ομαλότητας της f , θα υπάρχει μία περιοχή του a'' για την οποία ικανοποιούνται οι συνθήκες Wolfe. Επιπλέον, επειδή το αριστερό μέλος της (1.5) είναι αρνητικό, η περιοχή αυτή θα ικανοποιεί και τις αυστηρές συνθήκες Wolfe. \square

1.2 Μέθοδοι Αναζήτησης Περιοχής Εμπιστοσύνης

Αυτή η παράγραφος ακολουθεί το ([1], σελ. 66-100). Οι μέθοδοι αναζήτησης περιοχής εμπιστοσύνης κατασκευάζουν μία περιοχή μέσα στην οποία θεωρείται ότι ένα τετραγωνικό μοντέλο προσεγγίζει ικανοποιητικά την αντικειμενική συνάρτηση. Έστερα εντοπίζουν το σημείο που ελαχιστοποιεί την τιμή της συνάρτησης μέσα στην περιοχή αυτή. Αν τις συγκρίνουμε με τις μεθόδους αναζήτησης γραμμής, και οι δύο αξιοποιούν ένα τετραγωνικό μοντέλο, αλλά με ελαφρώς διαφορετικό τρόπο. Επίσης, στις μεθόδους περιοχής εμπιστοσύνης η διαδικασία επιλογής του βήματος και της κατεύθυνσης ελαχιστοποίησης γίνεται κατά κάποιον τρόπο ταυτόχρονα.

Το μέγεθος της περιοχής εμπιστοσύνης είναι το καίριο χαρακτηριστικό για την απόδοση των μεθόδων αυτών. Εάν το μέγεθος είναι πολύ μικρό ενδεχομένως να χαθεί μία ευκαιρία για να γίνουν μεγάλα άλματα προς το ελάχιστο της συνάρτησης, οπότε η πρόοδος της μεθόδου είναι πολύ αργή. Αντιθέτως, εάν το μέγεθος της περιοχής είναι πολύ μεγάλο, δεν μπορούμε να είμαστε εξίσου βέβαιοι για την ποιότητα των αποτελεσμάτων, αν δηλαδή το σημείο που θα επιλεγεί από τον αλγόριθμο είναι πράγματι το ελάχιστο της συνάρτησης, ή αν θα είναι καν πλησίον του πραγματικού ελαχίστου. Στην πράξη, το μέγεθος της περιοχής εμπιστοσύνης που θα επιλέξουμε καθορίζεται σε μεγάλο βαθμό από την απόδοση του αλγορίθμου στις προηγούμενες επαναλήψεις. Όταν έχουμε ενδείξεις για καλή προσαρμογή του μοντέλου, ίσως αξίζει να τολμήσουμε να κατασκευάσουμε περιοχές μεγαλύτερης ακτίνας, ένα ρίσκο που όπως περιγράψαμε πριν μπορεί να γλιτώσει πολύ χρόνο για την ολοκλήρωση της μεθόδου. Όταν όμως παρατηρούμε ότι το μοντέλο δε δίνει πολύ ακριβή αποτελέσματα δεδομένης μιας περιοχής εμπιστοσύνης, μπορούμε να επαναλάβουμε την αναζήτηση, αυτή τη φορά εντός μίας μικρότερης περιοχής.

Το τετραγωνικό μοντέλο βασίζεται στο ανάπτυγμα Taylor με όρους μέχρι και δεύτερης τάξης της συνάρτησης f με κέντρο το x_k ,

$$f(x_k + p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k + tp) p, \quad \mu\epsilon \quad t \in (0, 1).$$

Έτσι, το μοντέλο μας m_k έχει τη μορφή

$$m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p,$$

όπου $g_k = \nabla f(x_k)$ και B_k μία προσέγγιση του εσσιανού πίνακα $\nabla^2 f(x_k)$. Η διαφορά μεταξύ του μοντέλου και της αντικειμενικής συνάρτησης είναι της τάξεως του $O(\|p\|^2)$, οπότε αντιλαμβανόμαστε ότι είναι ένα πολύ καλό μοντέλο όταν το p γίνεται αρκετά μικρό. Ειδικά όταν ο πίνακας B_k είναι ακριβώς ο εσσιανός πίνακας $\nabla^2 f(x_k)$ (στην περίπτωση των μεθόδων Newton), τότε το σφάλμα του μοντέλου γίνεται της τάξης του $O(\|p\|^3)$.

Για να επιλέξουμε το βήμα μας, πρέπει να επιλύσουμε το υποπρόβλημα

$$\min_{p \in \mathbb{R}^n} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p, \text{ τ.ώ. } \|p\| \leq \Delta_k,$$

όπου $\Delta_k > 0$ είναι η ακτίνα της περιοχής εμπιστοσύνης. Ο περιορισμός που θέτουμε στο p μπορεί να γραφεί διαφορετικά ως $p^T p \leq \Delta_k^2$, γράφοντας και αυτόν σε τετραγωνική μορφή. Η λύση του υποπροβλήματος αυτού δεν είναι προφανής, ωστόσο μπορεί να προσεγγιστεί σχετικά γρήγορα ικανοποιώντας τις ανάγκες της μεθόδου για καλή λειτουργία.

Μας αρκεί να βρούμε μία προσεγγιστική λύση του υποπροβλήματος, το οποίο μπορεί να γίνει ακολουθώντας διαφορετικές στρατηγικές. Η πρώτη αξιοποιεί το σημείο Cauchy, που πρακτικά είναι το σημείο που ελαχιστοποιεί την τιμή του μοντέλου m_k στη διεύθυνση της μέγιστης καθόδου (δηλαδή τη διεύθυνση $-g_k$), πάντα εντός του περιορισμού της περιοχής εμπιστοσύνης. Ονομαστικά, δύο από τις δημοφιλέστερες μεθόδους είναι η μέθοδος dogleg και η μέθοδος two-dimensional subspace minimization. Η πρώτη μπορεί να χρησιμοποιηθεί όταν ο Εσσιανός πίνακας του μοντέλου B_k είναι θετικά ορισμένος, ενώ η δεύτερη όποτε ο πίνακας B_k είναι μη αντιστρέψιμος.

Μία τρίτη μέθοδος, η οποία θα μας φανεί χρήσιμη και στο δεύτερο μέρος της εργασίας, δίνει το έναυσμα για τη δημιουργία μίας επαναληπτικής μεθόδου, αξιοποιώντας παράλληλα ορισμένες ιδιότητες που προκύπτουν από τη μορφή του υποπροβλήματος. Συγκεκριμένα, ισχύει το ακόλουθο Λήμμα:

Λήμμα 1.2. Έστω m η τετραγωνική συνάρτηση

$$m(p) = g^T p + \frac{1}{2} p^T B p, \quad (1.6)$$

όπου B συμμετρικός πίνακας. Ισχύουν οι ακόλουθες προτάσεις:

- (i) Η συνάρτηση m παρουσιάζει ελάχιστο αν και μόνο αν ο B είναι θετικά ημιορισμένος και υπάρχει p τέτοιο ώστε $Bp = -g$. Επιπλέον, κάθε p για το οποίο ισχύει $Bp = -g$ είναι σημείο ολικού ελαχίστου της m .
- (ii) Η m έχει μοναδικό ολικό ελάχιστο αν και μόνο αν ο B είναι θετικά ορισμένος.

Απόδειξη. Αποδεικνύουμε κάθε μία από τις προτάσεις.

- (i) Αποδεικνύουμε πρώτα το αντίστροφο. Έστω ότι υπάρχει p τέτοιο ώστε $Bp = -g$. Έτσι, για κάθε $w \in \mathbb{R}^n$, έχουμε:

$$\begin{aligned} m(p+w) &= g^T(p+w) + \frac{1}{2}(p+w)^T B(p+w) \\ &= g^T p + g^T w + \frac{1}{2}(p^T B p + 2p^T B w + w^T B w) \\ &= g^T p + g^T w + \frac{1}{2} p^T B p + p^T B w + \frac{1}{2} w^T B w \\ &= \left(g^T p + \frac{1}{2} p^T B p \right) + g^T w + (Bp)^T w + \frac{1}{2} w^T B w \\ &= m(p) + (g^T + (Bp)^T) w + \frac{1}{2} w^T B w \\ &= m(p) + \frac{1}{2} w^T B w \\ &\geq m(p), \end{aligned} \quad (1.7)$$

διότι B : θετικά ημιορισμένος. Άρα, το p είναι ελάχιστο του m .

Για το ευθύ, υποθέτουμε ότι p είναι ελάχιστο του m . Άρα παίρνουμε

$$\nabla m(p) = Bp + g = 0,$$

οπότε παίρνουμε τη σχέση $Bp = -g$, και

$$\nabla^2 m(p) = B: \text{θετικά ημιορισμένος,}$$

ολοκληρώνοντας της απόδειξη της πρώτης πρότασης.

- (ii) Ξεκινάμε πάλι με το αντίστροφο. Ισχύουν ακριβώς τα ίδια με την απόδειξη του (i), με τη διαφορά ότι τώρα $w^T B w > 0$ για κάθε $w \neq 0$. Οπότε, για κάθε $w \neq 0$ ισχύει $m(p + w) > m(p)$, και άρα το p είναι το μοναδικό σημείο ελαχίστου για τη συνάρτηση m .

Για το ευθύ, έστω p το μοναδικό ελάχιστο του μοντέλου. Από το (i) έχουμε ότι B : θετικά ημιορισμένος. Έστω ότι ο B δεν είναι θετικά ορισμένος. Άρα υπάρχει ένα $w \neq 0$ για το οποίο ισχύει $w^T B w = 0$. Τότε, από τη σχέση (1.7), παίρνουμε ότι $m(p + w) = m(p)$, άτοπο καθώς η συνάρτηση m έχει μοναδικό σημείο ελαχίστου.

□

Εφόσον αποδείξαμε το Λήμμα, είμαστε σε θέση να αποδείξουμε και το παρακάτω Θεώρημα.

Θεώρημα 1.1. Το διάνυσμα p^* είναι ολική λύση του προβλήματος

$$\min_{p \in \mathbb{R}^n} m(p) = f + g^T p + \frac{1}{2} p^T B p, \text{ τ.ώ. } \|p\| \leq \Delta, \quad (1.8)$$

αν και μόνο αν p^* ανήκει στην εφικτή περιοχή, και υπάρχει ένα βαθμωτό $\lambda \geq 0$ τέτοιο ώστε να ικανοποιούνται οι ακόλουθες συνθήκες:

$$(B + \lambda I)p^* = -g, \quad (1.9\alpha')$$

$$\lambda(\Delta - \|p^*\|) = 0, \quad (1.9\beta')$$

$$(B + \lambda I): \text{θετικά ημιορισμένος.} \quad (1.9\gamma')$$

Απόδειξη. Έστω ότι υπάρχει $\lambda \geq 0$ και ότι ισχύουν οι συνθήκες (1.9). Από το Λήμμα 1.2 προκύπτει ότι το p^* είναι ολικό ελάχιστο της τετραγωνικής συνάρτησης

$$\hat{m}(p) = g^T p + \frac{1}{2} p^T (B + \lambda I)p = m(p) + \frac{\lambda}{2} p^T p.$$

Αφού $\hat{m}(p) \geq \hat{m}(p^*)$, έχουμε

$$m(p) \geq m(p^*) + \frac{\lambda}{2} ((p^*)^T p^* - p^T p). \quad (1.10)$$

Αφού $\lambda(\Delta - \|p^*\|) = 0$, άρα $\lambda(\Delta^2 - (p^*)^T p^*) = 0$, έχουμε

$$m(p) \geq m(p^*) + \frac{\lambda}{2} (\Delta - p^T p).$$

Άρα, από το γεγονός ότι $\lambda \geq 0$ παίρνουμε τελικά $m(p) \geq m(p^*)$ για κάθε p με $\|p\| \leq \Delta$, άρα πράγματι το p^* είναι ολικό ελάχιστο του προβλήματος (1.8).

Για το αντίστροφο, έστω p^* ένα ολικό ελάχιστο του προβλήματος (1.8). Θα δείξουμε ότι υπάρχει $\lambda \geq 0$ έτσι ώστε να ισχύουν οι συνθήκες (1.9).

Έστω ότι $\|p^*\| < \Delta$. Το p^* θα αποτελεί ελάχιστο της m χωρίς την ανάγκη του περιορισμού, οπότε θα ισχύει

$$\nabla m(p^*) = Bp^* + g = 0, \quad \nabla^2 m(p^*) = B: \text{ θετικά ημιορισμένος,}$$

οπότε οι συνθήκες (1.9) ισχύουν για $\lambda = 0$.

Έστω τώρα ότι $\|p^*\| = \Delta$. Η (1.9β') έπεται αμέσως. Επίσης, το p^* αποτελεί λύση του προβλήματος

$$\min m(p), \text{ τ.ώ. } \|p\| = \Delta.$$

Για τη Λαγκρανζιανή συνάρτηση του προβλήματος, που ορίζεται ως

$$L(p, \lambda) = m(p) + \frac{\lambda}{2}(p^T p - \Delta),$$

θα ισχύει

$$\begin{aligned} \nabla_p L(p^*, \lambda) &= 0 \\ \nabla_p m(p^*) + \lambda p &= 0 \\ Bp^* + g + \lambda p^* &= 0 \\ (B + \lambda I)p^* &= -g, \end{aligned} \tag{1.11}$$

και καταλήγουμε στην (1.9α'). Επιπλέον, αφού $m(p) \geq m(p^*)$ για κάθε p με $p^T p = (p^*)^T p^* = \Delta^2$, θα ισχύει

$$m(p) \geq m(p^*) + \frac{\lambda}{2}((p^*)^T p^* - p^T p)$$

από όπου παίρνουμε την ισοδύναμη σχέση

$$\frac{1}{2}(p - p^*)^T (B + \lambda I)(p - p^*) \geq 0.$$

Πράγματι, έχουμε

$$\begin{aligned} m(p) + \frac{\lambda}{2}p^T p &\geq m(p^*) + \frac{\lambda}{2}(p^*)^T p^* \\ g^T p + \frac{1}{2}p^T Bp + \frac{\lambda}{2}p^T p &\geq g^T p^* + \frac{1}{2}(p^*)^T Bp^* + \frac{\lambda}{2}(p^*)^T p^* \\ g^T(p - p^*) + \frac{1}{2}p^T(B + \lambda I)p - \frac{1}{2}(p^*)^T(B + \lambda I)p^* &\geq 0 \\ g^T(p - p^*) + \frac{1}{2}p^T(B + \lambda I)p - \frac{1}{2}p^T(B + \lambda I)p^* + \frac{1}{2}p^T(B + \lambda I)p^* - \frac{1}{2}(p^*)^T(B + \lambda I)p^* &\geq 0 \\ g^T(p - p^*) + \frac{1}{2}p^T(B + \lambda I)(p - p^*) + \frac{1}{2}(p - p^*)^T(B + \lambda I)p^* &\geq 0 \end{aligned}$$

Επειδή B : συμμετρικός, τότε και $B + \lambda I$: συμμετρικός. Οπότε παίρνουμε:

$$\begin{aligned} g^T(p - p^*) + \frac{1}{2}(p - p^*)^T(B + \lambda I)p + \frac{1}{2}(p - p^*)^T(B + \lambda I)p^* &\geq 0 \\ (p - p^*)^T g + \frac{1}{2}(p - p^*)^T(B + \lambda I)(p + p^*) &\geq 0, \end{aligned}$$

και αντικαθιστώντας τη σχέση $(B + \lambda I)p^* = -g$, παίρνουμε:

$$\begin{aligned} -(p - p^*)^T(B + \lambda I)p^* + \frac{1}{2}(p - p^*)^T(B + \lambda I)(p + p^*) &\geq 0 \\ -\frac{1}{2}(p - p^*)^T(B + \lambda I) \cdot (2p^*) + \frac{1}{2}(p - p^*)^T(B + \lambda I)(p + p^*) &\geq 0 \\ \frac{1}{2}(p - p^*)^T(B + \lambda I)(p + p^* - 2p^*) &\geq 0 \\ \frac{1}{2}(p - p^*)^T(B + \lambda I)(p - p^*) &\geq 0 \end{aligned}$$

Επειδή το σύνολο των κατευθύνσεων

$$\left\{ w : w = \pm \frac{p - p^*}{\|p - p^*\|}, \text{ για κάποιο } p \text{ με } \|p\| = \Delta \right\}$$

είναι πυκνό στη μοναδιαία σφαίρα, το τελευταίο αποτέλεσμα συνεπάγεται την (1.9γ'), ότι δηλαδή ο πίνακας $B + \lambda I$ είναι θετικά ορισμένος.

Τέλος, πρέπει να αποδείξουμε ότι $\lambda \geq 0$. Από τις σχέσεις (1.9α') και (1.9γ') και από το Λήμμα 1.2, το p^* ελαχιστοποιεί το μοντέλο \hat{m} , οπότε ισχύει η σχέση (1.10). Έστω ότι οι σχέσεις (1.9α') και (1.9β') ικανοποιούνται μόνο για αρνητικά λ . Πάλι, από την (1.10) παίρνουμε $m(p) \geq m(p^*)$ όταν $\|p\| \geq \|p^*\| = \Delta$. Το p^* δηλαδή θα είναι ολικό ελάχιστο του μοντέλου m . Έτσι, από το Λήμμα 4.7 έπεται ότι $Bp = -g$ και B είναι θετικά ημιορισμένος. Τελικά, προκύπτει ότι οι συνθήκες (1.9α') και (1.9γ') ικανοποιούνται για $\lambda = 0$, άτοπο. \square

Στην πιο απλή περίπτωση, όπου ισχύει η σχέση $\|B_k^{-1}g_k\| \leq \Delta_k$ η λύση p_k^* ταυτίζεται με τη λύση του αντίστοιχου προβλήματος χωρίς τον περιορισμό, $p_k^B = -B_k^{-1}g_k$, την οποία ονομάζουμε πλήρες βήμα για το υποπρόβλημα.

Περιγραφή της Προσέγγισης για την Περιοχή Εμπιστοσύνης

Η επιλογή της ακτίνας Δ_k σε κάθε επανάληψη είπαμε πως είναι ένα ιδιαίτερα σημαντικό ζήτημα που καθορίζει ουσιαστικά τη μέθοδο. Ενσωματώνουμε τα αναγκαία χαρακτηριστικά που πρέπει να τη διέπουν ορίζοντας το πηλίκο

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}.$$

Ο αριθμητής είναι η πραγματική μείωση της συνάρτησης, ενώ ο παρονομαστής είναι η προβλεπόμενη μείωση (η μείωση της συνάρτησης f όπως αυτή προσεγγίζεται από το μοντέλο m_k), δεδομένου ενός βήματος p_k . Έχουμε ως απαίτηση η τιμή του μοντέλου να μειώνεται, άρα ο παρονομαστής θα είναι σίγουρα μη αρνητικός. Επομένως, όταν το ρ_k είναι αρνητικό συνεπάγεται ότι η τιμή της αντικειμενικής συνάρτησης έχει αυξηθεί, οπότε οφείλουμε να απορρίψουμε το βήμα. Όταν όμως το πηλίκο πλησιάζει

στο 1, το βήμα μας είναι ιδανικό μιας και οι προβλεπόμενη μείωση σχεδόν συμπίπτει με την πραγματική. Τέλος, όταν το ρ_k είναι μεν θετικό αλλά κοντά στο μηδέν, παρόλο που επιτυγχάνεται μείωση και στην αντικειμενική συνάρτηση, αυτή δεν αντανακλάται επαρκώς στο μοντέλο μας. Για να διορθώσουμε αυτή τη δυσλειτουργία μπορούμε να συρρικνώσουμε την ακτίνα Δ_k και να επαναλάβουμε την παραπάνω διαδικασία. Σημειωτέον, όταν το βήμα μας γίνει όσο η ακτίνα Δ_k (δηλαδή όταν $\|p_k\| = \Delta_k$) πρόκειται για μία ένδειξη ότι η ακτίνα που έχουμε θέσει περιορίζει τον αλγόριθμο και άρα πρέπει να την επεκτείνουμε.

Κλιμακοποίηση τη Μεθόδου σε άλλες Νόρμες

Μία εξαιρετικά σημαντική ιδιότητα των μεθόδων περιοχής εμπιστοσύνης είναι η καλή προσαρμογή που έχουν σε προβλήματα κακής κατάστασης. Ως προβλήματα κακής κατάστασης ορίζουμε τα προβλήματα στα οποία υπάρχει έντονη ανομοιομορφία στην ευαισθησία της συνάρτησης σε αλλαγές κατά των διαφόρων συνιστωσών των μεταβλητών. Ή, με πιο απλά λόγια, όταν η αντικειμενική συνάρτηση f παρουσιάζει πολύ μικρή ευαισθησία σε αλλαγές κάποιων συνιστωσών του διανύσματος x , ενώ παρουσιάζει τρομερή ευαισθησία σε μικρές αλλαγές σε κάποιες άλλες συνιστώσες. Η δυσκολία που αντιμετωπίζουμε σε τέτοια προβλήματα έγκειται στην τοπολογία τους. Οι ισοϋψείς καμπύλες της συνάρτησης f απεικονίζονται ως ελλείψεις με πολύ μεγάλη εκκεντρότητα ειδικά κοντά στο ελάχιστό της, x^* .

Στη συζήτηση που είχαμε πιο πάνω η γεωμετρική αναπαράσταση της περιοχής εμπιστοσύνης είναι μία n -διάστατη σφαίρα με κέντρο το x_k και ακτίνα Δ_k . Όπως γίνεται αντιληπτό, τέτοιες περιοχές εμπιστοσύνης δεν μπορούν να χρησιμεύσουν στην επίλυση των παραπάνω προβλημάτων, ακριβώς λόγω τη γεωμετρίας τους. Ωστόσο, αν μπορούσαμε να μετατρέψουμε το σχήμα της περιοχής εμπιστοσύνης επίσης σε μία έλλειψη, θα έχουμε καταφέρει να γενικεύσουμε τα αποτελέσματά μας. Αυτό γίνεται πολύ εύκολα, ορίζοντας τη νέα περιοχή εμπιστοσύνης ως εξής:

$$\|Dp\| \leq \Delta_k,$$

όπου D ένας διαγώνιος πίνακας με θετικά διαγώνια στοιχεία. Το νέο - κλιμακοποιημένο - πρόβλημα δεν αλλάζει καθόλου, παρά μόνο στον περιορισμό (την περιοχή εμπιστοσύνης):

$$\min_{p_k \in \mathbb{R}^n} m_k(p) \stackrel{\text{def}}{=} f_k + g_k^T p + \frac{1}{2} p^T B_k p, \text{ τ.ώ. } \|Dp\| \leq \Delta_k. \quad (1.12)$$

Ο πίνακας D είναι αυτός που ευθύνεται για την κλιμάκωση τη περιοχής εμπιστοσύνης ώστε να ταιριάζει -κατά προσέγγιση- με την κλιμάκωση του προβλήματος. Σε γενικές γραμμές, όσο πιο ευαίσθητη είναι η συνάρτηση f ως προς τη συνιστώσα x_i , τόσο μεγαλύτερες είναι και οι τιμές των αντίστοιχων στοιχείων d_{ii} του D .

Εναλλακτικά, μία ίσως πιο απλή μέθοδος αντιμετώπισης, είναι να εφαρμόσουμε έναν μετασχηματισμό στις συντεταγμένες p ώστε να βελτιώσουμε την κλίμακα του αρχικού προβλήματος. Ορίζοντας ως

$$\tilde{p} \stackrel{\text{def}}{=} Dp$$

και αντικαθιστώντας στην σχέση (1.12), παίρνουμε στο μετασχηματισμένο υποπρόβλημα

$$\min_{p_k \in \mathbb{R}^n} \tilde{m}_k(\tilde{p}) \stackrel{\text{def}}{=} f_k + g_k^T D^{-1} \tilde{p} + \frac{1}{2} \tilde{p}^T D^{-1} B_k D^{-1} \tilde{p}, \text{ τ.ώ. } \|\tilde{p}\| \leq \Delta_k.$$

Εδώ παρατηρούμε ότι ο περιορισμός, δηλαδή η περιοχή εμπιστοσύνης, έχει διατηρήσει σφαιρικό σχήμα λόγω του μετασχηματισμού στις συντεταγμένες. Για τον ίδιο λόγο όμως έχει αλλάξει ο ορισμός του μοντέλου m_k που χρησιμοποιούμε.

Τέλος, η νόρμα που εμφανίζεται στον περιορισμό του υποπροβλήματος είναι συνήθως η Ευκλείδεια νόρμα, μπορεί όμως να είναι και οποιαδήποτε άλλη, χωρίς να επηρεάζει το πρόβλημα. Στα προβλήματα χωρίς περιορισμούς μικρής και μεσαίας κλίμακας δεν παρατηρείται κανένα ιδιαίτερο πλεονέκτημα. Αντιθέτως, η χρήση μιας διαφορετικής από τη συνήθη νόρμα μπορεί να λειτουργήσει υπέρ μας όταν αντιμετωπίζουμε προβλήματα με περιορισμούς, μιας και κάθε νόρμα φέρνει αντίστοιχες αλλαγές στο σχήμα τόσο της περιοχής εμπιστοσύνης, όσο και της επικτής περιοχής του προβλήματος.

1.3 Μέθοδοι Συζυγών Κλίσεων

Η παράγραφος αυτή βασίζεται στο ([1], σελ. 101-134). Οι μέθοδοι συζυγών κλίσεων πρωτοεμφανίστηκαν ως επαναληπτικές μέθοδοι για την επίλυση γραμμικών συστημάτων με θετικά ορισμένο πίνακα συντελεστών, ως μία εναλλακτική της μεθόδου απαλοιφής Gauss. Αντιμετωπίζουν άψογα προβλήματα μεγάλης κλίμακας, ενώ με κατάλληλες προσαρμογές δύνανται να επιλύουν και μη γραμμικά προβλήματα. Η απόδοσή τους εξαρτάται από την κατανομή των ιδιοτιμών του πίνακα συντελεστών του προβλήματος, την οποία μπορούμε να τη διορθώσουμε προς το συμφέρον μας προρρυθμίζοντας το γραμμικό σύστημα, επιταχύνοντας τελικά την σύγκλιση της μεθόδου. Επιπλέον πλεονεκτήματα έναντι των άλλων μεθόδων είναι επίσης ότι δεν χρειάζεται να αποθηκεύει τον πίνακα συντελεστών του γραμμικού συστήματος, ενώ είναι ταχύτερη από τη μέθοδο μέγιστης καθόδου. Παρακάτω θα αναλύσουμε τη γραμμική μέθοδο συζυγών κλίσεων.

Ας πάρουμε ένα γραμμικό σύστημα που ικανοποιεί την προϋπόθεση της μεθόδου, δηλαδή

$$Ax = b, \quad (1.13)$$

όπου A είναι ένας $n \times n$ συμμετρικός και θετικά ορισμένος πίνακας. Η λύση του συστήματος ισοδυναμεί με αυτή του αντίστοιχου προβλήματος ελαχιστοποίησης

$$\min f(x) \stackrel{\text{def}}{=} \frac{1}{2}x^T Ax - b^T x, \quad x \in \mathbb{R}^n.$$

Η παραπάνω ισοδυναμία συνεπάγεται ότι η μέθοδος των συζυγών κλίσεων μπορεί να αξιοποιηθεί και ως μέθοδος ελαχιστοποίησης κυρτών τετραγωνικών συναρτήσεων. Ορίζουμε την κλίση της συνάρτησης f ως το υπόλοιπο που αφήνει το γραμμικό σύστημα, δηλαδή

$$\nabla f(x) = Ax - b \stackrel{\text{def}}{=} r(x),$$

οπότε για οποιοδήποτε $x = x_k$ έχουμε

$$\nabla f(x_k) = Ax_k - b \stackrel{\text{def}}{=} r(x_k). \quad (1.14)$$

Ο ορισμός των συζυγών κλίσεων είναι ο εξής: Ένα σύνολο διανυσμάτων $\{p_0, p_1, \dots, p_l\}$ λέγεται συζυγές ως προς τον συμμετρικό και θετικά ορισμένο πίνακα A όταν

$$p_i^T A p_j = 0, \quad \text{για κάθε } i \neq j.$$

Για το σύνολο $\{p_0, p_1, \dots, p_l\}$ ισχύει η παρακάτω πρόταση.

Πρόταση 1.1. Έστω ένα σύνολο $\{p_0, p_1, \dots, p_l\}$ συζυγών κλίσεων ως προς έναν συμμετρικό και θετικά ορισμένο πίνακα A . Τότε, τα διανύσματα p_0, p_1, \dots, p_l είναι γραμμικά ανεξάρτητα.

Απόδειξη. Έστω ένα σύνολο p_1, p_2, \dots, p_l συζυγές ως προς τον συμμετρικό και θετικά ορισμένο πίνακα A , και έστω ότι τα διανύσματα είναι γραμμικά εξαρτημένα, δηλαδή

$$\lambda_1 p_1 + \lambda_2 p_2 + \dots + \lambda_l p_l = 0$$

και λ_i , $i = 1, 2, \dots, l$ όχι όλα μηδέν. Υποθέτουμε χωρίς βλάβη της γενικότητας ότι $\lambda_1 \neq 0$. Άρα έχουμε

$$p_1 = -\frac{\lambda_2}{\lambda_1} p_2 - \frac{\lambda_3}{\lambda_1} p_3 - \dots - \frac{\lambda_l}{\lambda_1} p_l,$$

και κατά συνέπεια έχουμε τα εξής:

$$\begin{aligned} p_1^T A p_1 &= p_1^T A \left(-\frac{\lambda_2}{\lambda_1} p_2 - \frac{\lambda_3}{\lambda_1} p_3 - \dots - \frac{\lambda_l}{\lambda_1} p_l \right) \\ p_1^T A p_1 &= p_1^T A \left(-\frac{\lambda_2}{\lambda_1} p_2 \right) + p_1^T A \left(-\frac{\lambda_3}{\lambda_1} p_3 \right) + \dots + p_1^T A \left(-\frac{\lambda_l}{\lambda_1} p_l \right) \\ p_1^T A p_1 &= -\frac{\lambda_2}{\lambda_1} p_1^T A p_2 - \frac{\lambda_3}{\lambda_1} p_1^T A p_3 - \dots - \frac{\lambda_l}{\lambda_1} p_1^T A p_l \\ p_1^T A p_1 &= 0 \end{aligned}$$

το οποίο είναι άτοπο, καθώς ο πίνακας A είναι θετικά ορισμένος. Έτσι, προκύπτει το ζητούμενο ότι τα διανύσματα $\{p_1, p_2, \dots, p_l\}$ είναι γραμμικά ανεξάρτητα. \square

Η ιδιότητα των συζυγών κλίσεων εξασφαλίζει ότι η συνάρτηση $f(\cdot)$ μπορεί να ελαχιστοποιηθεί το πολύ σε n επαναλήψεις, ελαχιστοποιώντας διαδοχικά κατά μήκος της κάθε μίας από τις συζυγείς κατευθύνσεις. Έτσι, διαπιστώνουμε την ταχύτητα της μεθόδου.

Η τελευταία ιδιότητα αποδεικνύεται με το Θεώρημα 1.2. Προτού προχωρήσουμε στο Θεώρημα, πρέπει να γίνουν οι απαραίτητες προετοιμασίες. Δεδομένου ενός αρχικού σημείου $x_0 \in \mathbb{R}^n$ και ενός συνόλου συζυγών κατευθύνσεων $\{p_0, p_1, \dots, p_{n-1}\}$ δημιουργούμε μία ακολουθία

$$x_{k+1} = x_k + a_k p_k, \quad (1.15)$$

όπου το a_k ελαχιστοποιεί τη συνάρτηση $f(\cdot)$ κατά μήκος του $x_k + a p_k$, και ισούται με

$$a_k = -\frac{r_k^T p_k}{p_k^T A p_k}, \quad (1.16)$$

όπως αποδεικνύεται στην επόμενη πρόταση.

Πρόταση 1.2.

Η τιμή του a_k που ελαχιστοποιεί τη συνάρτηση $f(x) = \frac{1}{2} x^T A x - b^T x$, με $x = x_k + a_k p_k$ ισούται με

$$a_k = -\frac{r_k^T p_k}{p_k^T A p_k}.$$

Απόδειξη. Ορίζουμε τη συνάρτηση $\lambda(a) = f(x_k + a p_k)$. Έχουμε τα εξής:

$$\begin{aligned} \lambda(a) &= \frac{1}{2} (x_k + a p_k)^T A (x_k + a p_k) - b^T (x_k + a p_k) \\ \lambda(a) &= \frac{1}{2} [x_k^T A x_k + 2a x_k^T A p_k + a^2 p_k^T A p_k] - b^T x_k - a b^T p_k \\ \lambda(a) &= a^2 \left(\frac{1}{2} p_k^T A p_k \right) + a (x_k^T A p_k - b^T p_k) + \frac{1}{2} x_k^T A x_k - b^T x_k. \end{aligned}$$

Μηδενίζοντας την παράγωγο της συνάρτησης ως προς a έχουμε:

$$\begin{aligned} \lambda'(a) &= 2a \left(\frac{1}{2} p_k^T A p_k \right) + x_k^T A p_k - b^T p_k = 0 \\ a \cdot p_k^T A p_k &= b^T p_k - x_k^T A p_k \\ a &= \frac{(b^T - x_k^T A) p_k}{p_k^T A p_k}, \end{aligned}$$

και από τη σχέση (1.14) παίρνουμε το ελάχιστο

$$a_k = -\frac{r_k^T p_k}{p_k^T A p_k}. \quad (1.17)$$

□

Πλέον είμαστε έτοιμοι να αποδείξουμε το Θεώρημα ([1], Θεώρημα 5.1).

Θεώρημα 1.2.

Για κάθε $x_0 \in \mathbb{R}^n$ η ακολουθία $\{x_k\}$ που περιγράφεται από τις σχέσεις (1.15), (1.16), συγκλίνει στη λύση x^* του γραμμικού συστήματος (1.13) σε n το πολύ βήματα

Απόδειξη. Αφού οι κατευθύνσεις $\{p_i\}$ είναι γραμμικά ανεξάρτητες, πρέπει να καλύπτουν όλο τον χώρο \mathbb{R}^n ($\text{span}\{p_1, p_2, \dots, p_n\} = \mathbb{R}^n$). Οπότε, μπορούμε να γράψουμε τη διαφορά μεταξύ του x_0 και της λύσης x^* ως:

$$x^* - x_0 = \sigma_0 p_0 + \sigma_1 p_1 + \dots + \sigma_{n-1} p_{n-1}$$

για κάποια βαθμωτά σ_k . Πολλαπλασιάζοντας την εξίσωση με $p_k^T A$ και αξιοποιώντας την ιδιότητα των συζυγών κλίσεων έχουμε

$$\begin{aligned} p_k^T A(x^* - x_0) &= p_k^T A \sigma_0 p_0 + p_k^T A \sigma_1 p_1 + \dots + p_k^T A \sigma_{n-1} p_{n-1} \\ p_k^T A(x^* - x_0) &= \sigma_0 p_k^T A p_0 + \sigma_1 p_k^T A p_1 + \dots + \sigma_{n-1} p_k^T A p_{n-1} \\ p_k^T A(x^* - x_0) &= \sigma_k p_k^T A p_k \\ \frac{p_k^T A(x^* - x_0)}{p_k^T A p_k} &= \sigma_k. \end{aligned} \quad (1.18)$$

Αρκεί να δείξουμε ότι οι συντελεστές σ_k συμπίπτουν με τα βήματα a_k της σχέσης (1.16). Εφόσον το x_k προκύπτει από τη σχέση (1.15), έχουμε

$$x_k = x_0 + a_0 p_0 + a_1 p_1 + \dots + a_{k-1} p_{k-1}.$$

Όπως προηγουμένως, πολλαπλασιάζοντας με $p_k^T A$, και αξιοποιώντας την ιδιότητα των συζυγών κλίσεων, παίρνουμε τελικά

$$\begin{aligned} p_k^T A x_k &= p_k^T A x_0 \\ p_k^T A(x_k - x_0) &= 0 \end{aligned}$$

και άρα

$$\begin{aligned}p_k^T A(x^* - x_0) &= p_k^T A(x^* - x_k + x_k - x_0) \\p_k^T A(x^* - x_0) &= p_k^T (Ax - Ax_k) \\p_k^T A(x^* - x_0) &= p_k^T (b - Ax_k) \\p_k^T A(x^* - x_0) &= -p_k^T r_k.\end{aligned}$$

Συγκρίνοντας το αποτέλεσμα με τις σχέσεις (1.16) και (1.3), βλέπουμε ότι πράγματι $\sigma_k = a_k$, παίρνοντας τελικά το επιθυμητό αποτέλεσμα. \square

Η γεωμετρική ερμηνεία του θεωρήματος είναι πολύ απλή: Όταν ο πίνακας A είναι διαγώνιος, οι ισοϋψείς καμπύλες της συνάρτησης $f(\cdot)$ αναπαρίστανται ως ελλείψεις με άξονες παράλληλους στις κατευθύνσεις των συντεταγμένων. Οπότε, μπορούμε να βρούμε το ελάχιστο της $f(\cdot)$ αρκεί να ελαχιστοποιήσουμε (διαδοχικά πάντα) κατά την κατεύθυνση της κάθε συντεταγμένης, φτάνοντας στη λύση του προβλήματος μετά από n επαναλήψεις (μία για κάθε συντεταγμένη). Όταν ο πίνακας A δεν είναι διαγώνιος αυτή η ιδιότητα παύει να ισχύει, διότι οι ισοϋψείς καμπύλες τις $f(\cdot)$, αν και παραμένουν ελλείψεις, αυτή τη φορά δεν έχουν τους άξονές τους παράλληλους με τις κατευθύνσεις των συντεταγμένων. Ωστόσο, εάν μετασχηματίσουμε ολόκληρο το πρόβλημα ώστε ο πίνακας A να γίνει διαγώνιος, θα μπορούσαμε να αξιοποιήσουμε ξανά την ιδιότητα, ως προς τις συντεταγμένες που θα προκύψουν μετά από τον μετασχηματισμό. Αυτό ακριβώς κάνουν και οι συζυγείς κατευθύνσεις. Συγκεκριμένα, αν θέσουμε τον πίνακα

$$S = [p_0 \ p_1 \ \dots \ p_{n-1}],$$

δηλαδή κάθε στήλη του να είναι και μία από τις συζυγείς κατευθύνσεις, και θέσουμε τον μετασχηματισμό

$$\hat{x} = S^{-1}x,$$

η συνάρτηση $f(\cdot)$ γίνεται

$$\hat{f}(x) \stackrel{\text{def}}{=} f(S\hat{x}) = \frac{1}{2}\hat{x}(S^T AS)\hat{x} - (S^T b)^T \hat{x}.$$

Από την ιδιότητα των συζυγών κατευθύνσεων προκύπτει άμεσα ότι ο πίνακας $S^T AS$ είναι διαγώνιος,

εμφανίζοντας έτσι το επιθυμητό αποτέλεσμα. Πράγματι, αν κάνουμε τις πράξεις έχουμε τα εξής:

$$S^T AS = [p_0 \ p_1 \ \dots \ p_{n-1}]^T A [p_0 \ p_1 \ \dots \ p_{n-1}]$$

$$S^T AS = \begin{bmatrix} p_0^T \\ p_1^T \\ \vdots \\ p_{n-1}^T \end{bmatrix} A [p_0 \ p_1 \ \dots \ p_{n-1}]$$

$$S^T AS = \begin{bmatrix} p_0^T \\ p_1^T \\ \vdots \\ p_{n-1}^T \end{bmatrix} [Ap_0 \ Ap_1 \ \dots \ Ap_{n-1}]$$

$$S^T AS = \begin{bmatrix} p_0^T Ap_0 & p_0^T Ap_1 & \dots & p_0^T Ap_{n-1} \\ p_1^T Ap_0 & p_1^T Ap_1 & \dots & p_1^T Ap_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n-1}^T Ap_0 & p_{n-1}^T Ap_1 & \dots & p_{n-1}^T Ap_{n-1} \end{bmatrix}$$

και αξιοποιώντας την ιδιότητα των συζυγών κλίσεων προκύπτει άμεσα ότι ο πίνακας είναι διαγώνιος.

Επιπλέον, με τον τρόπο που έχουμε ορίσει το \hat{x} διαπιστώνουμε ότι η συντεταγμένη \hat{x}_i στον χώρο των \hat{x} αντιστοιχεί στη συζυγή κατεύθυνση p_i στον χώρο των x . Βρήκαμε δηλαδή ότι οι διαδοχικές ελαχιστοποιήσεις που περιγράψαμε προηγουμένως πρέπει αυτή τη φορά να γίνουν κατά μήκος των συζυγών κατευθύνσεων. Τελικά, και το μετασχηματισμένο πρόβλημα μπορεί να λυθεί το πολύ σε n επαναλήψεις.

Το ζήτημα πλέον είναι πώς θα υπολογιστούν οι συζυγείς κατευθύνσεις. Η απλούστερη προσέγγιση είναι να βρούμε τα ιδιοδιανύσματα του πίνακα A . Πράγματι, τα ιδιοδιανύσματα είναι ορθογώνια μεταξύ τους, και συζυγή ως προς τον πίνακα A , οπότε θα μπορούσαν να σχηματίσουν το σύνολο των συζυγών κατευθύνσεων. Όμως, σε προβλήματα μεγάλης κλίμακας, ο υπολογισμός των ιδιοδιανυσμάτων μπορεί να είναι από επίπονος έως και ακατόρθωτος. Μία εναλλακτική προσέγγιση για την επιλογή των συζυγών κατευθύνσεων είναι η τροποποίηση της ορθογωνιοποίησης Gram - Schmidt ώστε να εξυπηρετεί τον σκοπό μας. Ακόμα και αυτή η διαδικασία μπορεί να έχει αυξημένες απαιτήσεις. Ευτυχώς όμως, έχουμε μία πληθώρα επιλογών στη διάθεσή μας.

Η μέθοδος των συζυγών κλίσεων είναι μία μέθοδος συζυγών κατευθύνσεων, αλλά με μία επιπλέον ιδιαιτερότητα. Κατά τη διαδικασία κατασκευής των συζυγών διανυσμάτων, το νέο συζυγές διάνυσμα p_k μπορεί να κατασκευαστεί γνωρίζοντας μόνο το προηγούμενο διάνυσμα, p_{k-1} . Τα προηγούμενα διανύσματα δε μας απασχολούν, καθώς ο σχεδιασμός της μεθόδου εξασφαλίζει ότι το διάνυσμα p_k θα είναι συζυγές και με όλα τα προηγούμενα, p_0, p_1, \dots, p_{k-2} .

Επιλέγουμε κάθε συζυγής κατεύθυνση να δίνεται από τον τύπο

$$p_k = -r_k + \beta_k p_{k-1}, \quad (1.19)$$

όπου β_k είναι το βαθμωτό μέγεθος το οποίο επιλέγεται ώστε τα δύο διανύσματα p_k, p_{k-1} είναι μεταξύ τους συζυγή, πάντα ως προς τον πίνακα A . Για να υπολογίσουμε το β_k πολλαπλασιάζουμε τα δύο μέλη της σχέσης μας με $p_{k-1}^T A$:

$$\begin{aligned}
p_{k-1}^T A p_k &= p_{k-1}^T A (-r_k + \beta_k p_{k-1}) \\
0 &= -p_{k-1}^T A r_k + \beta_k p_{k-1}^T A p_{k-1} \\
p_{k-1}^T A r_k &= \beta_k p_{k-1}^T A p_{k-1}
\end{aligned}$$

και διαιρώντας με $p_{k-1}^T A p_{k-1}$,

$$\begin{aligned}
\frac{p_{k-1}^T A r_k}{p_{k-1}^T A p_{k-1}} &= \beta_k \\
\frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}} &= \beta_k
\end{aligned}$$

Για να ξεκινήσει η μέθοδος θέτουμε ως p_0 την κατεύθυνση μέγιστης καθόδου. Μπορούμε να βελτιώσουμε τη μέθοδο συνδυάζοντας κάποιες ιδιότητες των συζυγών κλίσεων. Μία από αυτές, όπως περιγράφεται και από το Θεώρημα 1.3 ([1], Θεώρημα 5.2), λέει ότι μετά από k επαναλήψεις, η μέθοδος των συζυγών κλίσεων θα έχει βρει το ελάχιστο στο χώρο που καλύπτουν οι πρώτες k συζυγείς κατευθύνσεις, δηλαδή στο χώρο

$$\text{span}\{p_0, p_1, \dots, p_k\}.$$

Επίσης, συνδυάζοντας τις σχέσεις (1.14) και (1.15) παίρνουμε την πολύ σημαντική σχέση

$$r_{k+1} = r_k + a_k A p_k \quad (1.20)$$

η οποία χρησιμεύει και για την απόδειξη του Θεωρήματος.

Θεώρημα 1.3.

Έστω $x_0 \in \mathbb{R}^n$ ένα τυχαίο σημείο και έστω $\{x_k\}$ μία ακολουθία που προκύπτει από τις σχέσεις (1.15), (1.16). Τότε

$$r_k^T p_i = 0, \text{ για } i = 0, 1, \dots, k-1, \quad (1.21)$$

και x_k είναι το ελάχιστο της συνάρτησης $f(x) = \frac{1}{2} x^T A x - b^T x$ επί του συνόλου

$$\{x \mid x = x_0 + \text{span}\{p_0, p_1, \dots, p_{k-1}\}\}.$$

Οι παρακάτω προτάσεις δίνουν μερικούς πολύ χρήσιμους ορισμούς για τις ποσότητες a_k και β_k .

Πρόταση 1.3.

Για την ποσότητα a_k της μεθόδου ισχύει:

$$a_k = \frac{r_k^T r_k}{p_k^T A p_k}$$

Απόδειξη. Πολλαπλασιάζοντας και τα δύο μέλη της σχέσης (1.19) με r_k^T έχουμε:

$$r_k^T p_k = -r_k^T r_k + \beta_k r_k^T p_{k-1}$$

και από την (1.21) παίρνουμε

$$r_k^T p_k = -r_k^T r_k.$$

Τώρα, αντικαθιστώντας στη σχέση (1.16) παίρνουμε το επιθυμητό αποτέλεσμα. \square

Πρόταση 1.4.

Για την ποσότητα β_k της μεθόδου ισχύει:

$$\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} \quad (1.22)$$

Απόδειξη. Ξεκινώντας από τη σχέση (1.20), πολλαπλασιάζοντας με r_{k+1}^T έχουμε:

$$\begin{aligned} r_{k+1} &= r_k + a_k A p_k \\ r_{k+1}^T r_{k+1} &= r_{k+1}^T r_k + a_k r_{k+1}^T A p_k. \end{aligned}$$

Αντικαθιστώντας το a_k σύμφωνα με την Πρόταση 1.3,

$$\begin{aligned} r_{k+1}^T r_{k+1} - r_{k+1}^T r_k &= \frac{r_k^T r_k}{p_k^T A p_k} \cdot r_{k+1}^T p_k \\ r_{k+1}^T r_{k+1} - r_{k+1}^T r_k &= r_k^T r_k \cdot \frac{r_{k+1}^T p_k}{p_k^T A p_k} \\ r_{k+1}^T r_{k+1} - r_{k+1}^T r_k &= r_k^T r_k \beta_{k+1}, \end{aligned}$$

και διαιρώντας με $r_k^T r_k$, παίρνουμε

$$\frac{r_{k+1}^T r_{k+1}}{r_k^T r_k} - \frac{r_{k+1}^T r_k}{r_k^T r_k} = \beta_{k+1}$$

Αρκεί να δείξουμε ότι $\frac{r_{k+1}^T r_k}{r_k^T r_k} = 0 \Rightarrow r_{k+1}^T r_k = 0$. Η σχέση (1.19) γράφεται εναλλακτικά ως:

$$\begin{aligned} p_k &= -r_k + \beta_k p_{k-1} \\ r_{k+1}^T p_k &= -r_{k+1}^T r_k + \beta_k (r_{k+1}^T p_{k-1}) \end{aligned}$$

και από τη σχέση (1.21) προκύπτει $r_{k+1}^T r_k = 0$. Τελικά,

$$\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}.$$

□

Διαπιστώνουμε λοιπόν ότι ανά πάσα στιγμή ο αλγόριθμος χρειάζεται να αποθηκεύει μόνο τα μεγέθη x_k, r_k, p_k της τρέχουσας και της τελευταίας επανάληψης. Παράλληλα, η μέθοδος δεν απαιτεί δύσκολες και χρονοβόρες πράξεις, οι πιο απαιτητικές πράξεις που θα πρέπει να γίνει είναι ο πολλαπλασιασμός πίνακα με διάνυσμα, $A p_k$ και τα εσωτερικά γινόμενα $p_k^T (A p_k)$ και $r_{k+1}^T r_{k+1}$. Τα παραπάνω επιβεβαιώνουν τους ισχυρισμούς μας για την οικονομία που προσφέρει η μέθοδος τόσο ως προς το χρόνο εκτέλεσης όσο και ως προς τον αποθηκευτικό χώρο. Έτσι, δεν υπάρχει καμία αμφιβολία ότι η μέθοδος ενδείκνυται για προβλήματα μεγάλης κλίμακας. Βέβαια, είναι επιρρεπής σε σφάλματα στρογγυλοποίησης.

Ρυθμός Σύγκλισης της Μεθόδου

Αναφέραμε νωρίτερα ότι υπάρχει μία σύνδεση της μεθόδου συζυγών κλίσεων και με τις ιδιοτιμές του πίνακα A (είπαμε ότι ως συζυγείς κλίσεις μπορούμε να θέσουμε τα ιδιοδιανύσματα του πίνακα). Αναφέραμε επίσης ότι η μέθοδος θα δώσει λύση σε οποιοδήποτε πρόβλημα το πολύ σε n επαναλήψεις. Η αλήθεια είναι ότι η μέθοδος μπορεί να τερματίσει πολύ πιο σύντομα, αν οι συνθήκες το ευνοούν. Συγκεκριμένα, αν οι ιδιοτιμές του πίνακα A είναι κατανομημένες γύρω από r διακριτές τιμές, τότε η μέθοδος θα πλησιάσει αρκετά κοντά στη λύση του προβλήματος έπειτα από r επαναλήψεις. Μάλιστα, αν x^* είναι το πραγματικό ελάχιστο της αντικειμενικής συνάρτησης, θα ισχύει ότι

$$\|x_{k+1} - x^*\|_A^2 \leq \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)^2 \|x_0 - x^*\|_A^2.$$

Εναλλακτικά, μπορούμε να αποφανθούμε για τη σύγκλιση της μεθόδου αξιοποιώντας τον δείκτη κατάστασης $\kappa(A)$ του πίνακα A , όπου για την Ευκλείδεια νόρμα ορίζεται ως

$$\kappa(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\lambda_n}{\lambda_1},$$

όπου λ_1, λ_n η μικρότερη και η μεγαλύτερη ιδιοτιμή του πίνακα A αντίστοιχα. Αποδεικνύεται ότι

$$\|x_k - x^*\|_A \leq 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k \|x_0 - x^*\|_A,$$

το οποίο είναι μία πολύ ελαστική εκτίμηση του σφάλματος, είναι ωστόσο χρήσιμη σε πολλές περιπτώσεις. Επίσης, το φράγμα αυτό είναι άμεσα συγκρίσιμο με το φράγμα της μεθόδου μέγιστης καθόδου, με τη διαφορά ότι εδώ το φράγμα εξαρτάται από την τετραγωνική ρίζα του δείκτη κατάστασης του πίνακα A .

Προετοιμασία της Μεθόδου

Υπάρχει η δυνατότητα να επιταχύνουμε ακόμα περισσότερο τη σύγκλιση της μεθόδου, μετασχηματίζοντας το γραμμικό μας σύστημα ώστε να βελτιώσουμε την κατανομή των ιδιοτιμών του πίνακα A . Εισάγοντας τον μετασχηματισμό

$$\hat{x} = Cx, \tag{1.23}$$

όπου C ένας αντιστρέψιμος πίνακας, η τετραγωνική συνάρτηση f γίνεται

$$\hat{f}(\hat{x}) = \frac{1}{2} \hat{x}^T (C^{-T} A C^{-1}) \hat{x} - (C^{-T} b)^T \hat{x}$$

και το αντίστοιχο γραμμικό σύστημα

$$(C^{-T} A C^{-1}) \hat{x} = C^{-T} b.$$

Στο μετασχηματισμένο πρόβλημα η σύγκλιση της μεθόδου (αλλά και οποιαδήποτε άλλη διεργασία) θα πρέπει να γίνει βάσει του πίνακα $C^{-T} A C^{-1}$. Επομένως, η σωστή επιλογή του πίνακα C θα παράξει έναν μετασχηματισμένο πίνακα με καλύτερες προδιαγραφές από τον αρχικό, είτε με καλύτερη κατανομή των ιδιοτιμών, συσταδοποιημένες γύρω από λίγες διακριτές τιμές, είτε με καλύτερο δείκτη κατάστασης.

Δεν είναι ανάγκη να μεταφέρουμε τον μετασχηματισμό (1.23) καθ' όλη τη διάρκεια της επίλυσης. Αντιθέτως, εκτελούμε τον αλγόριθμο των συζυγών κλίσεων ως προς τις μετασχηματισμένες μεταβλητές \hat{x} , και ύστερα αντιστρέφουμε τον μετασχηματισμό για να βρούμε και τη λύση x του προβλήματος. Επίσης, αξίζει να παρατηρήσουμε ότι ο αλγόριθμος δεν αξιοποιεί ακριβώς τον πίνακα C , αλλά τον πίνακα $M = C^T C$, που εκ των πραγμάτων θα είναι ένας συμμετρικός και θετικά ορισμένος πίνακας.

Στην πράξη δε υπάρχει κάποια "σωστή" επιλογή για τον πίνακα C που να δίνει καλά αποτελέσματα κατ' επανάληψη. Εξαρτάται πάντα από τις συνθήκες του προβλήματος, αλλά και τις προτιμήσεις του χρήστη όσον αφορά το κόστος των πράξεων ή του αποθηκευτικού χώρου.

Συζυγείς κλίσεις και Μη Γραμμικά Προβλήματα - Η μέθοδος Polak Ribière

Όπως έχουμε αναφέρει προηγουμένως, η μέθοδος των συζυγών κλίσεων μπορεί να αντιμετωπίσει εξίσου και προβλήματα στα οποία η αντικειμενική συνάρτηση f είναι μη γραμμική, τροποποιώντας ορισμένα σημεία της. Όπως προτείνουν οι Fletcher και Reeves, οι δύο αλλαγές που πρέπει να γίνουν αφορούν το βήμα a_k και το υπόλοιπο r_k , καθώς αυτά τα δύο στοιχεία αξιοποιούν την πληροφορία για την τετραγωνική συνάρτηση f . Επομένως, πρέπει να προσαρμόσουμε τον συνήθη αλγόριθμο των συζυγών κλίσεων ώστε να λαμβάνουμε υπόψιν ότι η αντικειμενική συνάρτηση πλέον είναι μη γραμμική. Έτσι, η ποσότητα β_{k+1} , αντί για τη σχέση (1.22) θα υπολογίζεται από τον τύπο

$$\beta_{k+1}^{FR} = \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}.$$

Από τις δεκάδες παραλλαγές των μη γραμμικών συζυγών κλίσεων, αυτή που ξεχωρίζει είναι η μέθοδος των Polak και Ribière. Η διαφορά έγκειται στον υπολογισμό του β_{k+1} , το οποίο τώρα θα υπολογίζεται από τον τύπο

$$\beta_{k+1}^{PR} = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\|\nabla f_k\|^2}.$$

Ένα χαρακτηριστικό της μεθόδου Polak-Ribière που πρέπει να λάβουμε σοβαρά υπόψιν είναι ότι οι συνθήκες Wolfe (1.1) δεν μας εξασφαλίζουν ότι το διάνυσμα p_k θα είναι πάντα κατεύθυνση καθόδου. Για να προσπεράσουμε αυτή τη δυσλειτουργία, ορίζουμε την παράμετρο β_{k+1} ως

$$\beta_{k+1}^+ = \max\{\beta_{k+1}^{PR}, 0\},$$

οπότε με μία απλή προσαρμογή των συνθηκών Wolfe καταφέρνουμε τελικά να εξασφαλίζουμε πάντα μία κατεύθυνση καθόδου. Πρακτικά, η παραπάνω σχέση λέει ότι αν η παράμετρος β_{k+1}^{PR} δε δίνει μία κατεύθυνση καθόδου (οπότε και $\beta_{k+1}^{PR} < 0$), επιλέγουμε να κάνουμε ένα βήμα μέγιστης καθόδου.

Η μέθοδος των συζυγών κλίσεων για μη γραμμικά προβλήματα είναι ιδιαίτερα ελκυστική για προβλήματα μεγάλων διαστάσεων λόγω των ελάχιστων υπολογισμών (υπολογίζονται μόνο η αντικειμενική συνάρτηση και η κλίση της) και της μνήμης που απαιτείται για την υλοποίησή της.

1.4 Μέθοδοι Quasi - Newton

Αυτή η παράγραφος βασίζεται στο ([1], σελ. 135-163). Οι μέθοδοι Quasi-Newton αποτελούν μία από τις σημαντικότερες μεθόδους βελτιστοποίησης. Το γεγονός ότι χρησιμοποιούν μία προέγγιση του εσσιανού πίνακα, σε αντίθεση με τις μεθόδους Newton, τις καθιστά πιο γρήγορες από θέμα πράξεων, εμφανίζοντας μάλιστα εξίσου καλά αποτελέσματα. Επιπλέον, η δομή των Quasi-Newton τους επιτρέπει να επιτυγχάνουν ακόμα και υπεργραμμική σύγκλιση, καθιστώντας τις ανώτερες από τη μέθοδο μέγιστης

καθόδου. Ταυτόχρονα, καλύπτουν μία ευρεία γκάμα προβλημάτων με ή χωρίς περιορισμούς, ακόμα και προβλήματα μεγάλης κλίμακας.

Η δημοφιλέστερη μέθοδος Quasi-Newton, μέσω της οποίας θα δούμε και τις ιδιότητες των μεθόδων της οικογενείας, είναι η μέθοδος BFGS (από τα ονόματα των Broyden, Fletcher, Goldfarb και Shanno που την ανακάλυψαν). Στην μέθοδο BFGS ξεκινάμε προσεγγίζοντας την αντικειμενική συνάρτηση με το παρακάτω τετραγωνικό μοντέλο:

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p,$$

όπου B_k είναι ένας $n \times n$ συμμετρικός και θετικά ορισμένος πίνακας. Ως κατεύθυνση αναζήτησης για την επόμενη επανάληψη τη μεθόδου ορίζουμε το

$$x_{k+1} = x_k + a_k p_k,$$

όπου p_k είναι η θέση ελαχίστου για το μοντέλο,

$$\begin{aligned} \nabla m_k(p_k) &= \nabla f_k + B_k p_k = 0 \\ B_k p_k &= -\nabla f_k \\ p_k &= -B_k^{-1} \nabla f_k. \end{aligned}$$

Το βήμα a_k επιλέγεται έτσι ώστε να ικανοποιούνται οι συνθήκες Wolfe. Παρατηρούμε ότι η μεθοδολογία σχεδόν ταυτίζεται με αυτή των μεθόδων Newton αναζήτησης γραμμής. Όπως είπαμε όμως και στην αρχή του Κεφαλαίου, αντί για τον ακριβή Εσσιανό πίνακα $\nabla^2 f_k$ χρησιμοποιούμε την προσέγγισή του B_k .

Ο υπολογισμός του πίνακα B_k είναι γενικά μία χρονοβόρα διαδικασία. Προκειμένου να ελαφρύνουμε τον αλγόριθμο από αυτό το βάρος μπορούμε αντ' αυτού να ανανεώνουμε τον πίνακα, αξιοποιώντας πληροφορίες από την τελευταία επανάληψη. Συγκεκριμένα, έχοντας υπολογίσει το x_{k+1} της επόμενης επανάληψης θέλουμε να βρούμε τον πίνακα B_{k+1} ανανεώνοντας τον B_k , για να δημιουργήσουμε το μοντέλο της επόμενης επανάληψης,

$$m_{k+1}(p) = f_{k+1} + \nabla f_{k+1}^T p + \frac{1}{2} p^T B_{k+1} p.$$

Για να το κάνουμε αυτό, απαιτούμε η κλίση του m_{k+1} να ταιριάζει με την κλίση της αντικειμενικής συνάρτησης f στα δύο τελευταία σημεία, x_k και x_{k+1} . Παρατηρούμε ότι η δεύτερη συνθήκη ισχύει αν θέσουμε $p = 0$,

$$\nabla m_{k+1}(0) = \nabla f_{k+1},$$

και μάλιστα είναι μια συνθήκη που ισχύει σε οποιοδήποτε στάδιο του αλγορίθμου. Για την πρώτη συνθήκη, έχουμε

$$\nabla m_{k+1}(-a_k p_k) = \nabla f_{k+1} - a_k B_{k+1} p_k = \nabla f_k,$$

και αναδιατάσσοντας τους όρους παίρνουμε

$$B_{k+1} a_k p_k = \nabla f_{k+1} - \nabla f_k. \quad (1.24)$$

Αυτή η συνθήκη ονομάζεται εξίσωση τομής (secant equation). Για να απλοποιήσουμε τη μορφή της, ορίζουμε

$$s_k = x_{k+1} - x_k = a_k p_k, \quad y_k = \nabla f_{k+1} - \nabla f_k, \quad (1.25)$$

και άρα η εξίσωση παίρνει τη μορφή

$$B_{k+1}s_k = y_k.$$

Στην πράξη η (1.24) απαιτεί ο πίνακας B_{k+1} να απεικονίζει τη μεταβολή στην μεταβλητή x_k της επανάληψης, s_k με τη μεταβολή στην κλίση, y_k , όπως επί της ουσίας κάνει και ο εσσιανός πίνακας $\nabla^2 f$. Πολλαπλασιάζοντας τώρα με s_k^T , και λαμβάνοντας υπόψιν ότι ο πίνακας B_{k+1} είναι θετικά ορισμένος, παίρνουμε τη συνθήκη καμπυλότητας

$$s_k^T B_{k+1} s_k = s_k^T y_k > 0,$$

που πρέπει να ικανοποιούν τα δύο διανύσματα. Όταν η αντικειμενική συνάρτηση είναι κυρτή, τότε η συνθήκη ισχύει για κάθε δύο σημεία x_k, x_{k+1} . Όταν όμως η συνάρτηση f είναι μη κυρτή, ίσως χρειαστεί να υποβάλλουμε κάποιους επιπλέον περιορισμούς. Για παράδειγμα, αν στη γραμμική αναζήτηση εισάγουμε επιπλέον τις απλές ή τις αυστηρές συνθήκες Wolfe, τότε η συνθήκη καμπυλότητας εξακολουθεί να ισχύει. Πράγματι, συνδυάζοντας τις σχέσεις (1.1) και (1.25) έχουμε:

$$\begin{aligned} \nabla f(x_k + a_k p_k)^T p_k &\geq c_2 \nabla f(x_k)^T p_k \\ \nabla f_{k+1}^T \frac{s_k}{a_k} &\geq c_2 \nabla f_k^T \frac{s_k}{a_k} \\ (\nabla f_{k+1}^T - \nabla f_k^T) s_k &\geq (c_2 - 1) \nabla f_k^T s_k \\ y_k^T s_k &\geq (c_2 - 1) a_k \nabla f_k^T p_k. \end{aligned}$$

Επειδή $c_2 < 1$ και το διάνυσμα p_k είναι κατεύθυνση καθόδου για τη συνάρτηση f , το δεύτερο μέλος είναι τελικά θετικό, επαληθεύοντας έτσι τη συνθήκη καμπυλότητας.

Όταν ικανοποιείται η συνθήκη καμπυλότητας, η σχέση (1.24) δίνει άπειρες λύσεις για τον συμμετρικό πίνακα B_{k+1} , μιας και οι $n(n+1)/2$ μεταβλητές του υπερβαίνουν σε πλήθος τις n εξισώσεις που προκύπτουν από αυτή. Από όλες αυτές τις πιθανές λύσεις, εμείς θα πρέπει να επιλέξουμε αυτή που είναι πλησιέστερη στον πίνακα B_k . Δηλαδή, θα πρέπει να λύσουμε το πρόβλημα

$$\begin{aligned} \min_B \|B - B_k\|, \\ \text{τ.ώ. } B = B^T, \text{ και } B s_k = y_k. \end{aligned}$$

Οι δύο συνθήκες που πρέπει να ικανοποιούνται είναι ακριβώς ότι συζητήσαμε προηγουμένως, η συμμετρία του πίνακα B και η εξίσωση τομής (1.24). Για το πρόβλημα μπορούμε να χρησιμοποιήσουμε οποιαδήποτε νόρμα πινάκων, καθεμιά από τις οποίες δίνει διαφορετική λύση, άρα και μία διαφορετική μέθοδο Quasi-Newton.

Ωστόσο, η μέθοδος BFGS διαφοροποιείται με έναν περίτεχνο τρόπο ο οποίος εν τέλει την κάνει πιο αποδοτική από τις υπόλοιπες μεθόδους της ίδιας οικογένειας. Αντί να απαιτήσει να ικανοποιούνται οι παραπάνω συνθήκες για τον πίνακα B_k , επιβάλλει αντίστοιχες συνθήκες στον αντίστροφο της, $H_k = b_k^{-1}$. Πλέον καλούμαστε να ανανεώσουμε το πίνακα H_k . Το πρόβλημα που καλούμαστε να λύσουμε και οι συνθήκες που επιβάλλουμε, κατ' αναλογία με τα προηγούμενα, είναι οι εξής:

$$\begin{aligned} \min_H \|H - H_k\|, \\ \text{τ.ώ. } H = H^T, \text{ και } H y_k = s_k. \end{aligned} \tag{1.26}$$

Όπως βλέπουμε, ο ανανεωμένος πίνακας H_{k+1} πρέπει να είναι συμμετρικός, και να ικανοποιεί την ανάλογη της σχέσης (1.24), που είναι η δεύτερη συνθήκη. Η νόρμα που χρησιμοποιούμε για το

πρόβλημα (1.26) είναι η νόρμα Frobenius με βάρη,

$$\|A\|_W \equiv \|W^{1/2}AW^{1/2}\|_F,$$

όπου $\|C\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n c_{ij}^2$. Ο πίνακας βαρύτητας W μπορεί να επιλεγεί ως οποιοσδήποτε πίνακας ικανοποιεί τη σχέση $W s_k = y_k$. Η μοναδική λύση δίνεται από τον τύπο

$$H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T,$$

όπου

$$\rho_k = \frac{1}{y_k^T s_k}.$$

Το μοναδικό ερώτημα που μένει να απαντηθεί είναι η επιλογή του πίνακα H_0 . Αν και δεν υπάρχει κάποια καθολική μεθοδολογία, μπορούμε να προσαρμόσουμε την επιλογή μας ανάλογα με τις απαιτήσεις του προβλήματος. Ή πιο απλά, μπορούμε να ξεκινήσουμε θέτοντάς τον ως τον μοναδιαίο πίνακα, ή ένα πολλαπλάσιό του, έτσι ώστε να ταιριάζει με την κλίμακα του προβλήματος.

Αξιίζει να παρατηρήσουμε για το πρόβλημα (1.26) ότι ο πίνακας H_{k+1} δεν είναι απαραίτητα θετικά ορισμένος. Όμως αποδεικνύεται εύκολα ότι είναι θετικά ορισμένος αν ο H_k είναι θετικά ορισμένος.

Επίσης, για να είναι η μέθοδος αμετάβλητη σε μετασχηματισμούς συντεταγμένων, θα πρέπει και η λύση του προβλήματος (1.26) να έχει την ίδια ιδιότητα. Αυτό μπορούμε να το εξασφαλίσουμε επιλέγοντας τον κατάλληλο πίνακα βαρύτητας W . Η συνηθέστερη (και όπως επιβεβαιώνεται στην πράξη αποδοτικότερη) επιλογή είναι να θέσουμε $W = \bar{G}_k^{-1}$, όπου \bar{G}_k είναι ο μέσος εσσιανός και ορίζεται ως

$$\bar{G}_k = \left[\int_0^1 \nabla^2 f(x_k + \tau a_k p_k) d\tau \right].$$

Μία σημαντική ιδιότητα που παρουσιάζει η μέθοδος BFGS είναι ότι "διορθώνει" τον εαυτό της. Εξηγώντας, αν σε κάποια επανάληψη γίνει μία ανακριβής εκτίμηση της κυρτότητας της αντικειμενικής συνάρτησης (μέσω του πίνακα H_k), τότε το σφάλμα αυτό θα αμβλυνθεί σε βάθος λίγων επαναλήψεων. Αυτό βέβαια συμβαίνει με την προϋπόθεση ότι εκτελούμε μία καλή γραμμική αναζήτηση, για παράδειγμα μέσω των συνθηκών *Wolfe*. Από την άλλη, όταν η γραμμική αναζήτηση είναι σχετικά ανακριβής, τότε καταφέρνουμε να εξοικονομήσουμε χρόνο ως προς τις πράξεις. Ειδικά για τις συνθήκες *Wolfe*, συνήθως δίνουμε τις τιμές $c_1 = 10^{-4}$ και $c_2 = 0.9$. Πάντα όμως στην αρχή αξίζει να δοκιμαστεί το βήμα $a_k = 1$, καθώς έτσι μπορούμε να επιτύχουμε υπεργραμμική ταχύτητα σύγκλισης στη μέθοδο.

Τέλος, επισημαίνοντας ένα μειονέκτημα της μεθόδου, δεν μπορούμε να επιτύχουμε σύγκλιση σε ολικό ελάχιστο σε μη γραμμικά προβλήματα. Με άλλα λόγια, δεν μπορούμε να αποδείξουμε ότι η μέθοδος θα συγκλίνει σε ένα κρίσιμο σημείο του προβλήματος ξεκινώντας από οποιοδήποτε αρχικό σημείο, εκτός κι αν ικανοποιούνται ορισμένες λογικές προϋποθέσεις. Σε αυτές τις περιπτώσεις, στις οποίες δεν θα επεκταθούμε, παρατηρούμε πάλι μια υπεργραμμική σύγκλιση της μεθόδου.

2 Μέθοδοι Βελτιστοποίησης χωρίς τη χρήση Παραγώγων

Στα προβλήματα βελτιστοποίησης απαιτείται εν γένει κάποια πληροφορία περί της πρώτης και της δεύτερης παραγώγου της αντικειμενικής συνάρτησης του προβλήματος. Ωστόσο, στην πράξη αντιμετωπίζουμε συχνά προβλήματα στα οποία κάτι τέτοιο δεν είναι εφικτό. Μία εύλογη αντιμετώπιση αυτής της δυσκολίας είναι η προσέγγιση της παραγώγου με τα υπάρχοντα στοιχεία, όπως συμβαίνει και στη μέθοδο των Πεπερασμένων Διαφορών. Αυτή η μέθοδος, αν και επιτυχής σε γενικές γραμμές, υστερεί ως προς την ταχύτητα (απαιτείται πλήθος πράξεων για τις προσεγγίσεις) και ως προς την ανεκτικότητα στον θόρυβο. Επιπλέον, σε πολλές περιπτώσεις, ακόμα και όταν υπάρχουν διαθέσιμες πληροφορίες για την παράγωγο της αντικειμενικής συνάρτησης, δεν συμφέρει η υλοποίησή τους σε ένα πρόγραμμα. Για όλους αυτούς του λόγους έχουν εφευρεθεί μέθοδοι που δεν ασχολούνται καθόλου με υπολογισμούς που αφορούν την παράγωγο, αρκούνται μόνο στα δεδομένα - τιμές που αφορούν την συνάρτηση, για αυτό και ονομάζονται Μέθοδοι Βελτιστοποίησης Άνευ Παραγώγου (Derivative Free Optimization Methods - DFO Methods).

Οι Μέθοδοι Βελτιστοποίησης Άνευ Παραγώγου χωρίζονται και αυτές στις κατηγορίες που έχουν αναφερθεί στο πρώτο μέρος της εργασίας: μία σημαντική ομάδα τέτοιων μεθόδων με την οποία θα ασχοληθούμε ιδιαίτερω παρακάτω κατασκευάζει ένα μοντέλο της αντικειμενικής συνάρτησης, γραμμικό ή τετραγωνικό, και εκτελεί έναν αλγόριθμο για την βελτιστοποίηση αυτού του μοντέλου εντός μίας περιοχής εμπιστοσύνης (trust region). Άλλες μέθοδοι αξιοποιούν τη μέθοδο simplex των Nelder και Mead, τη μέθοδο των συζυγών κλίσεων (conjugate gradient), καθώς και μεθόδους αναζήτησης μοτίβου (pattern - search).

Στο τρέχον κεφάλαιο θα ασχοληθούμε με προβλήματα βελτιστοποίησης ομαλών συναρτήσεων οι οποίες δεν υπόκεινται σε περιορισμούς.

2.1 Μέθοδος Πεπερασμένων Διαφορών

Η Μέθοδος Πεπερασμένων Διαφορών ([1], σελ. 194-197, 221-223) είναι μία από τις θεμελιώδεις μεθόδους βελτιστοποίησης γενικά. Η μέθοδος αυτή προέρχεται από το Θεώρημα Taylor. Προκειμένου να υπολογίσουμε το διάνυσμα κλίσης $\nabla f(x)$ αλλά και του εσσιανού πίνακα $\nabla^2 f(x)$ μιας συνάρτησης κατά προσέγγιση, μπορούμε να λάβουμε υπόψιν μας τις τιμές της συνάρτησης σε διάφορα ισαπέχοντα σημεία, ως προς οποιαδήποτε κατεύθυνση που ορίζουν οι μεταβλητές της συνάρτησης. Υπάρχουν πολλοί τρόποι που θα μπορούσε κανείς να συνδυάσει τα δεδομένα του προβλήματος. Εν προκειμένω, θα μελετήσουμε δύο από τους συνηθέστερους τρόπους υπολογισμού, τις προς τα εμπρός και τις κεντρικές διαφορές. Στις προς τα εμπρός πεπερασμένες διαφορές, χρειαζόμαστε $(n + 1)$ σημεία της συνάρτησης f . Από το θεώρημα Taylor, με όρους μέχρι και δεύτερης τάξης, έχουμε

$$f(x + p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x) p, \text{ για κάποιο } t \in (0, 1),$$

$$f(x + p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x) p + O(\|p^3\|),$$

Αν επιλέξουμε το διάνυσμα p ως μία μεταβολή σε μόνο μία από τις συνιστώσες του διανύσματος x , δηλαδή $p = \varepsilon e_i$, και αν θεωρήσουμε το L ως άνω φράγμα της ποσότητας $\|\nabla^2 f(\cdot)\|$ για κάθε $t \in \{0, 1\}$, τότε, με μια αναδιάταξη των όρων παίρνουμε το εξής:

$$f(x + \varepsilon e_i) - f(x) - \nabla f(x)^T \varepsilon e_i = \frac{1}{2} \varepsilon^2 e_i^T \nabla^2 f(x) e_i + O(\varepsilon^3)$$
$$\|f(x + \varepsilon e_i) - f(x) - \nabla f(x)^T \varepsilon e_i\| \leq \frac{L}{2} \varepsilon^2$$

Διαπιστώνουμε ότι καθώς το ε τείνει στο μηδέν, το ίδιο συμβαίνει και στον όρο εντός της νόρμας. Έτσι, για μικρές τιμές του ε , η παράγωγος της συνάρτησης προσεγγίζεται ικανοποιητικά από την ποσότητα

$$\frac{\partial f}{\partial x_i}(x) = \nabla f(x)^T e_i \approx \frac{f(x + \varepsilon e_i) - f(x)}{\varepsilon}, \quad i = 1, \dots, n.$$

Σε πολλές περιπτώσεις μπορεί να χρειαστεί και ο υπολογισμός του Εσσιανού πίνακα. Και αυτός μπορεί να προσεγγιστεί με τις προς τα εμπρός πεπερασμένες διαφορές. Ξανά με τη βοήθεια του θεωρήματος Taylor, αντικαθιστώντας διαδοχικά αυτή τη φορά όπου $p = \varepsilon e_i$, $p = \varepsilon e_j$, και $p = \varepsilon(e_i + e_j)$ παίρνουμε τα εξής αποτελέσματα:

$$f(x + \varepsilon e_i) = f(x) + \nabla f(x)^T \varepsilon e_i + \frac{1}{2}(\varepsilon e_i)^T \nabla^2 f(x) \varepsilon e_i + O(\varepsilon^3),$$

$$f(x + \varepsilon e_j) = f(x) + \nabla f(x)^T \varepsilon e_j + \frac{1}{2}(\varepsilon e_j)^T \nabla^2 f(x) \varepsilon e_j + O(\varepsilon^3),$$

$$f(x + \varepsilon e_i + \varepsilon e_j) = f(x) + \nabla f(x)^T (\varepsilon e_i + \varepsilon e_j) + \frac{1}{2}(\varepsilon e_i + \varepsilon e_j)^T \nabla^2 f(x) (\varepsilon e_i + \varepsilon e_j) + O(\varepsilon^3).$$

και συνδυάζοντας κατάλληλα τις παραπάνω σχέσεις

$$f(x + \varepsilon e_i + \varepsilon e_j) - f(x + \varepsilon e_i) - f(x + \varepsilon e_j) + f(x) \approx \frac{1}{2} \varepsilon^2 \cdot 2 \cdot e_i \nabla^2 f(x) e_j,$$

$$f(x + \varepsilon e_i + \varepsilon e_j) - f(x + \varepsilon e_i) - f(x + \varepsilon e_j) + f(x) \approx \varepsilon^2 \cdot e_i \nabla^2 f(x) e_j,$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = e_i \nabla^2 f(x) e_j \approx \frac{f(x + \varepsilon e_i + \varepsilon e_j) - f(x + \varepsilon e_i) - f(x + \varepsilon e_j) + f(x)}{\varepsilon^2}.$$

Εάν θέλουμε να βρούμε όλα τα στοιχεία του Εσσιανού πίνακα, θα πρέπει να υπολογίσουμε την f στα σημεία $x + \varepsilon(e_i + e_j)$ για όλους τους συνδυασμούς των $i, j = 1, \dots, n$, όπως επίσης και στα σημεία $x + \varepsilon e_i, i = 1, \dots, n$. Λαμβάνοντας υπόψιν και τη συμμετρία του εσσιανού πίνακα, προκύπτει ότι θα πρέπει να υπολογίσουμε τη συνάρτηση συνολικά σε $\frac{n(n+1)}{2} + n = \frac{n^2+n}{2}$ σημεία.

Για τον αντίστοιχο υπολογισμό του διανύσματος $\nabla f(x)$ μέσω των κεντρικών πεπερασμένων διαφορών αξιοποιούμε πάλι το θεώρημα Taylor. Αντικαθιστώντας όπου $p = \varepsilon e_i$ και $p = -\varepsilon e_i, i = 1, \dots, n$ παίρνουμε

$$f(x + \varepsilon e_i) = f(x) + \varepsilon \nabla f(x)^T e_i + \frac{1}{2} \varepsilon^2 \nabla^2 f(x) e_i + O(\varepsilon^3),$$

$$f(x - \varepsilon e_i) = f(x) - \varepsilon \nabla f(x)^T e_i + \frac{1}{2} \varepsilon^2 \nabla^2 f(x) e_i + O(\varepsilon^3),$$

και αφαιρώντας τις δύο σχέσεις κατά μέλη προκύπτει τελικά το αποτέλεσμα:

$$f(x + \varepsilon e_i) - f(x - \varepsilon e_i) = 2\varepsilon \nabla f(x)^T e_i + O(\varepsilon^3)$$

$$\frac{\partial f}{\partial x_i}(x) = \frac{f(x + \varepsilon e_i) - f(x - \varepsilon e_i)}{2\varepsilon} + O(\varepsilon^2).$$

Συγκρίνοντας τα αποτελέσματα με εκείνα των προς τα εμπρός πεπερασμένων διαφορών, διαπιστώνουμε ότι οι κεντρικές διαφορές απαιτούν περισσότερα σημεία ($2n + 1$ για την ακρίβεια) αλλά είναι πιο ακριβής (το αντίστοιχο σφάλμα για τις προς τα εμπρός πεπερασμένες διαφορές ήταν της τάξης του $O(\varepsilon)$).

2.2 Μέθοδοι Ανάπτυξης Μοντέλου

Σε αυτές τις μεθόδους προσεγγίζουμε το πρόβλημα αναπτύσσοντας ένα γραμμικό ή τετραγωνικό μοντέλο που περιγράφει το πρόβλημα και λύνουμε το αντίστοιχο πρόβλημα βελτιστοποίησης, συνήθως με τη βοήθεια μίας μεθόδου trust - region για τον υπολογισμό του βήματος.

Στρέφοντας την προσοχή μας στην ανάπτυξη του τετραγωνικού μοντέλου, υποθέτουμε ότι η συνάρτηση f του προβλήματος προσεγγίζεται από την τετραγωνική συνάρτηση

$$m_k(x_k + p) = c + g^T p + \frac{1}{2} p^T G p. \quad (2.1)$$

Συνήθως το διάνυσμα g και ο πίνακας G είναι το διάνυσμα κλίσης $\nabla f(x)$ και ο εσσιανός πίνακας $\nabla^2 f(x)$ αντίστοιχα, πράγματα που στην παρούσα μελέτη θεωρούμε ότι είναι άγνωστα. Για την κατασκευή του μοντέλου λοιπόν πέρα από τη σταθερά c , θα πρέπει να υπολογίσουμε το διάνυσμα $g \in \mathbb{R}^n$ και τον συμμετρικό πίνακα $G \in \mathbb{R}^{n \times n}$. Ένας κλασικός τρόπος για αυτόν τον υπολογισμό είναι η παρεμβολή του μοντέλου m_k και της συνάρτησης f . Υποθέτουμε δηλαδή ότι

$$m_k(y^l) = f(y^l), \quad l = 1, \dots, q, \quad (2.2)$$

όπου τα $f(y^l)$, $l = 1, \dots, q$ μας είναι γνωστά. Συνήθως το αρχικό σύνολο παρεμβολής $Y = [y_1, y_2, \dots, y_q]$ επιλέγουμε να είναι οι κορυφές και τα μέσα των ακμών ενός μη εκφυλισμένου στοιχείου simplex του \mathbb{R}^n . Κατασκευάζουμε λοιπόν ένα σύστημα εξισώσεων μέσω του οποίου μπορούμε να υπολογίσουμε τα επιθυμητά στοιχεία. Αναλυτικά, χρειάζεται να προσδιορίσουμε το βαθμωτό c (1 στοιχείο), n στοιχεία για το διάνυσμα g , και $\frac{n(n+1)}{2}$ στοιχεία για τον συμμετρικό πίνακα G . Είναι κατανοητό λοιπόν ότι χρειαζόμαστε συνολικά

$$q = 1 + n + \frac{n(n+1)}{2} = \frac{(n+1)(n+2)}{2}$$

σημεία της συνάρτησης για να μπορέσουμε να προσδιορίσουμε πλήρως το μοντέλο, κατά μοναδικό τρόπο. Εξάριση αποτελεί η περίπτωση όπου τα δεδομένα σημεία δημιουργούν ένα αόριστο σύστημα. Κάτι τέτοιο όμως είναι επίσης εύκολα αποτρεψίμο.

Εφόσον έχουμε καταφέρει να αναπτύξουμε το μοντέλο m_k , χρειάζεται να επιλύσουμε το υποπρόβλημα περιοχής εμπιστοσύνης

$$\min_p m_k(x_k + p), \quad \text{με } \|p\|_2 \leq \Delta_k,$$

όπου $\Delta_k \geq 0$ η ακτίνα της περιοχής εμπιστοσύνης. Το πρόβλημα αυτό μπορεί να λυθεί με χρήση της θεωρίας της παραγράφου 1.2. Εάν η τιμή $x_k + p$ παρέχει μία επαρκή μείωση της αντικειμενικής συνάρτησης τότε αυτή ορίζεται ως η τιμή της επόμενης επανάληψης, δηλαδή $x_{k+1} = x_k + p$. Διαφορετικά, θα πρέπει να συρρικνώσουμε την ακτίνα Δ_k ή να επαναπροσδιορίσουμε τα σημεία παρεμβολής y_l . Η ανεπάρκεια των σημείων παρεμβολής y_l εννοείται όταν το σύστημα (2.2) βρίσκεται σε κακή κατάσταση. Σε αυτή την περίπτωση, παρατηρείται οι έξοδοι των επαναλήψεων του αλγορίθμου να περιορίζονται εντός μίας επιφάνειας χαμηλής διάστασης του \mathbb{R}^n που δεν περιέχει τη λύση του προβλήματος. Ο αλγόριθμος ωστόσο παραμένει λειτουργικός, εξακολουθεί να εντοπίζει δηλαδή ένα ακρότατο σε αυτή την επιφάνεια. Για να βελτιώσουμε το σύστημα μπορούμε να αντικαταστήσουμε ένα από τα σημεία παρεμβολής επιδιώκοντας να εξάγουμε ένα σύστημα παρόμοιο με το αρχικό, αλλά με καλύτερο δείκτη κατάστασης. Ωστόσο, αν το σύστημα δεν έχει κάποιο από τα παραπάνω προβλήματα, τότε η μέθοδος διορθώνεται ελαττώνοντας την ακτίνα της περιοχής αναζήτησης Δ_k με έναν από τους τρόπους που περιγράφονται στο Κεφάλαιο 1.2. Ως συνήθως, ο συντελεστής επαρκούς μείωσης μπορεί να ελεγχθεί με το πηλίκο της μεταβολής της πραγματικής συνάρτησης προς τη μεταβολή της συνάρτησης του μοντέλου, δηλαδή

$$\rho = \frac{f(x_k) - f(x_k + p)}{m_k(x_k) - m_k(x_k + p)}.$$

Είναι δυνατό να ελαχιστοποιήσουμε το κόστος της μεθόδου, αν απλώς ανανεώνουμε το υπάρχον μοντέλο m_k σε κάθε επανάληψη αντί να το υπολογίζουμε εκ νέου. Για να είναι εφικτό κάτι τέτοιο όμως, θα πρέπει να έχουμε προνοήσει ώστε η παρεμβολή να γίνεται πάνω σε μία βάση που να διευκολύνει αυτή τη διαδικασία, συνήθως μέσω των πολυωνύμων Lagrange ή Newton.

Η μέθοδος αυτή δυστυχώς είναι πολύ κοστοβόρα, και άρα απαγορευτική για προβλήματα μεγάλων διαστάσεων. Η χρήση των τετραγωνικών μοντέλων απαιτεί πράξεις της τάξης του $O(n^2)$ μόνο για την εκκίνηση του αλγόριθμου, με κάθε επανάληψη να είναι ακόμα πιο απαιτητική σε πράξεις. Ανανεώνοντας το μοντέλο, σε κάθε βήμα το συνολικό κόστος των πράξεων σε κάθε επανάληψη είναι της τάξης του $O(n^4)$. Το κόστος μπορεί να περιοριστεί ως ένα βαθμό χρησιμοποιώντας ένα γραμμικό μοντέλο προσέγγισης, με μικρές αλλαγές στον αλγόριθμο (για την ακρίβεια θέτοντας όπου $G = 0$). Το κόστος εκκίνησης του αλγόριθμου τώρα είναι μόνο $n + 1$ πράξεις με κάθε επανάληψη να απαιτεί μόνο $O(n^3)$ πράξεις. Εντούτοις, παρατηρείται μία δραματική πτώση στην ταχύτητα σύγκλισης του αλγόριθμου, μιας και το γραμμικό μοντέλο δεν περιέχει πληροφορία για την κυρτότητα της συνάρτησης. Ακόμα και αν χρησιμοποιήσουμε ένα υβριδικό μοντέλο, μεταβαίνοντας από ένα γραμμικό σε ένα τετραγωνικό προκειμένου να επωφεληθούμε από τα πλεονεκτήματα και των δύο, ο περιορισμός ως προς τις διαστάσεις του προβλήματος παραμένει.

Παρεμβολή και Πολυωνυμικές Βάσεις

Στην περίπτωση του γραμμικού μοντέλου ($G = 0$)

$$m_k(x_k + p) = f(x_k) + g^T p,$$

χρειάζεται να υπολογίσουμε μόνο το διάνυσμα g . Για αυτό το σκοπό χρειαζόμαστε n σημεία παρεμβολής της μορφής $m_k(y^l) = f(y^l)$. Εξετάζοντας αναλυτικότερα τις εξισώσεις παρεμβολής, και υποθέτοντας ότι $y^l = x_k + s^l$, $l = 1, 2, \dots, n$ έχουμε:

$$\begin{aligned} m_k(y^l) &= f(y^l) \Rightarrow \\ m_k(x_k + s^l) &= f(y^l) \Rightarrow \\ f(x_k) + g^T s^l &= f(y^l) \Rightarrow \\ (s^l)^T g &= f(y^l) - f(x_k), \quad l = 1, 2, \dots, n. \end{aligned} \tag{2.3}$$

Οι εξισώσεις (2.3) σχηματίζουν ένα γραμμικό σύστημα με άγνωστους τις συνιστώσες του διανύσματος g . Οι γραμμές του πίνακα συντελεστών του συστήματος είναι τα διανύσματα $(s^l)^T$. Το σύστημα ως γνωστόν θα έχει μοναδική λύση εφόσον ο πίνακας συντελεστών είναι αντιστρέψιμος. Εν προκειμένω, αυτό συνεπάγεται το σύνολο των διανυσμάτων $\{s^1, s^2, \dots, s^n\}$ να είναι γραμμικά ανεξάρτητα.

Στην περίπτωση του τετραγωνικού μοντέλου τώρα, μπορούμε να γράψουμε το μοντέλο (2.1) λίγο διαφορετικά, ως

$$m_k(x_k + p) = f(x_k) + g^T p + \sum_{i < j} G_{ij} p_i p_j + \frac{1}{2} \sum_{i=1}^n G_{ii} p_i^2 \stackrel{\text{def}}{=} f(x_k) + \hat{g}^T \hat{p}.$$

Το νέο διάνυσμα \hat{g} ισούται με

$$\hat{g} \equiv \left(g^T, \{G_{ij}\}_{i < j}, \left\{ \frac{1}{\sqrt{2}} G_{ii} \right\} \right)^T,$$

ενώ το διάνυσμα \hat{p} με

$$\hat{p} \equiv \left(p^T, \{p_i p_j\}_{i < j}, \left\{ \frac{1}{\sqrt{2}} p_i^2 \right\} \right)^T.$$

Τα δύο διανύσματα αποτελούνται συνολικά από $q-1$ συνιστώσες. Με αυτόν τον formalισμό καταφέρνουμε να φέρουμε το τετραγωνικό μοντέλο σε παρόμοια μορφή με το γραμμικό, και άρα ο προσδιορισμός των αγνώστων g και G μπορεί να πραγματοποιηθεί ανάλογα.

Όλα τα παραπάνω προκύπτουν εύκολα και άμεσα εφαρμόζοντας τη βασική θεωρία της παρεμβολής. Αν και απλό στην υλοποίησή του, στην πράξη μπορεί να οδηγήσει πολλές φορές σε ένα αόριστο σύστημα ή σε ένα σύστημα κακής κατάστασης. Ευτυχώς υπάρχουν κι άλλοι τρόποι για να προσεγγίσουμε την αντικειμενική συνάρτηση μέσω του τετραγωνικού μοντέλου.

Επιλέγοντας μία βάση $\{\varphi(\cdot)\}_{i=1}^q$ για το γραμμικό χώρο των n -διάστατων τετραγωνικών συναρτήσεων, μπορούμε να γράψουμε το μοντέλο (2.1) ως

$$m_k(x) = \sum_{i=1}^q a_i \varphi_i(x).$$

Εδώ, είναι σαφές ότι για να μπορέσουμε να προσδιορίσουμε το τετραγωνικό μοντέλο $m_k(x)$ πρέπει να προσδιορίσουμε τους άγνωστους συντελεστές a_i , $i = 1, 2, \dots, q$. Αξιοποιώντας το σύνολο των σημείων παρεμβολής $Y = \{y^1, y^2, \dots, y^q\}$ θα μπορέσουμε να προσδιορίσουμε τους συντελεστές a_i με μοναδικό τρόπο αν η ορίζουσα

$$\delta(Y) \stackrel{\text{def}}{=} \det \begin{pmatrix} \varphi_1(y^1) & \dots & \varphi_1(y^q) \\ \vdots & & \vdots \\ \varphi_q(y^1) & \dots & \varphi_q(y^q) \end{pmatrix}$$

είναι διάφορη του μηδενός.

Και αυτή η μέθοδος επίσης μπορεί να εμφανίσει ανακρίβειες ή ακόμα και να αποτύχει να δώσει κάποιο αποτέλεσμα, διότι η ορίζουσα μπορεί να πλησιάζει προς το μηδέν καθώς προχωράει ο αλγόριθμος. Πρέπει λοιπόν να εντάξουμε στη συζήτηση κάποιους τρόπους ώστε τα σημεία παρεμβολής να δίνουν ένα αξιόπιστο αποτέλεσμα. Ο κλασικός τρόπος είναι να αφήσουμε τα πράγματα ως έχουν έως ότου η ορίζουσα γίνει μικρότερη από μία δεδομένη τιμή. Ωστόσο υπάρχουν πολύ πιο αποδοτικοί μηχανισμοί από αυτή την απλοϊκή λύση, οι οποίοι περιγράφονται στη συνέχεια.

Αναβάθμιση του Συνόλου Παρεμβολής

Ένας από αυτούς τους μηχανισμούς είναι να ανανεώνουμε το σύνολο των σημείων παρεμβολής. Όταν παρατηρείται ανεπαρκής μείωση στη συνάρτηση f μπορούμε να αλλάξουμε κάποιο από τα σημεία παρεμβολής y^i ώστε η τιμή της ορίζουσας $\delta(Y)$ να αυξηθεί. Βασιζόμαστε σε μία σημαντική ιδιότητα του $\delta(Y)$. Πρώτα όμως χρειάζεται να ορίσουμε τα πολυώνυμα Lagrange: Για κάθε $y_j \in Y$, ορίζουμε τα πολυώνυμα Lagrange $l_j(x)$ να είναι ένα πολυώνυμο το πολύ δευτέρου βαθμού, έτσι ώστε να ισχύει $l_j(y_i) = \delta_{ij}$ για κάθε $y_i \in Y$. Έτσι, τα πολυώνυμα Lagrange έχουν τη μορφή

$$l_j(x) = c_j + g_j^T x + \frac{1}{2} x^T G_j x, \quad j = 1, 2, \dots, q.$$

Ως γνωστόν, για τα πολυώνυμα αυτά ισχύει

$$m_k(x) = \sum_{j=1}^q f(y_j)l_j(y_j),$$

άρα και για του συντελεστές c_j , g_j , και G_j ισχύει

$$c = \sum_{j=1}^q f(y_j)c_j, \quad g = \sum_{j=1}^q f(y_j)g_j \quad \text{και} \quad G = \sum_{j=1}^q f(y_j)G_j.$$

Το αναβαθμισμένο σύνολο Y^+ προκύπτει αν αντικαταστήσουμε ένα στοιχείο y_j από το παλιό σύνολο Y με ένα νέο στοιχείο y_+ . Αποδεικνύεται ότι ισχύει [2]

$$|\delta(Y^+)| \leq |l_j(y^+)| \cdot |\delta(Y)|. \quad (2.4)$$

Διακρίνουμε τις εξής δύο περιπτώσεις: Στην πρώτη περίπτωση το δοκιμαστικό σημείο $y^+ = x_k + p_k$ ελαττώνει σημαντικά την τιμή της συνάρτησης f , οπότε για το πηλίκο ρ ισχύει $\rho \geq \eta$, όπου $\eta \in (0, 1)$, ανάλογα με την ακρίβεια που θέλουμε να έχουμε στη διαφορά του μοντέλου σε σχέση με τη διαφορά της αντικειμενικής συνάρτησης. Σε αυτή την περίπτωση λοιπόν, πρέπει να συμπεριλάβουμε το y^+ στο σύνολο Y αντικαθιστώντας το σε ένα άλλο στοιχείο του, y^- . Συμβουλευόμενοι τη σχέση (2.4), επιλέγουμε το σημείο να είναι το

$$y^- = \arg \max_{y_j \in Y} |l_j(y^+)|, \quad j = 1, 2, \dots, q$$

διότι με αυτή την αντικατάσταση η τιμή της ορίζουσας θα αυξηθεί κατά το μέγιστο δυνατό, αποφέροντας και τα καλύτερα αποτελέσματα στη μέθοδό μας.

Η δεύτερη περίπτωση είναι το y^+ να μην επιτυγχάνει επαρκή μείωση στην f (όταν $\rho < \eta$). Εδώ υπάρχουν δύο τινά. Αρχικά, πρέπει να αποφανθούμε για το αν πρέπει να βελτιώσουμε το σύνολο Y . Το σύνολο Y λέμε ότι είναι επαρκές σε μία δεδομένη επανάληψη x_k όταν για κάθε $y_i \in Y$ ισχύει $\|x_k - y_i\| \leq \Delta$, έχουμε ότι το $|\delta(Y)|$ δεν μπορεί να διπλασιαστεί με την αντικατάσταση οποιουδήποτε από τα σημεία y_i με οποιοδήποτε σημείο y που βρίσκεται εντός της περιοχής εμπιστοσύνης ($\|x_k - y\| \leq \Delta$). Ισοδύναμα, βάσει της σχέσης (2.4) το σύνολο Y είναι επαρκές αν

$$\max_{\|y_j - x_k\| \leq \Delta} |l_j(x_k + p)| < 2, \quad \text{με} \quad \|p\| \leq \Delta,$$

Αν το Y είναι επαρκές αλλά η τιμή της f δε μειώνεται αρκετά, τότε ελαττώνουμε την ακτίνα της περιοχής εμπιστοσύνης και εκτελούμε μία νέα επανάληψη.

Αν όμως το σύνολο Y δεν είναι επαρκές, τότε αξιοποιούμε τον προαναφερθέν μηχανισμό. Επιλέγουμε ένα σημείο $y_- \in Y$ και το αντικαθιστούμε με κάποιο άλλο σημείο y^+ με στόχο να αυξήσουμε την τιμή της ορίζουσας $\delta(Y)$ όσο περισσότερο μπορούμε μέσα στην περιοχή εμπιστοσύνης που βρισκόμαστε. Για κάθε σημείο y_i ορίζουμε τον πιθανό αντικαταστάτη του ως

$$y_i^+ = \arg \max_{\|y - x_k\| \leq \Delta} |l_j(y)|.$$

Η αντικατάσταση που θα πραγματοποιηθεί τελικά είναι αυτή που δίνει τη μεγαλύτερη τιμή $|l_j(y_j^+)|$ για κάθε $y_j \in Y$.

Μία Μέθοδος βασισμένη στις Μεθόδους Quasi - Newton

Μία παραλλαγή της μεθόδου που θυμίζει τις Quasi-Newton βασίζεται στην απαίτηση ο εσσιανός πίνακας, ή μάλλον η προσέγγισή του, να μεταβάλλεται κατά το ελάχιστο δυνατό. Το δεύτερο χαρακτηριστικό της μεθόδου είναι το μοντέλο να παρεμβάλλει την κλίση ∇f στα δύο πιο πρόσφατα σημεία.

Για να κατασκευάσουμε το μοντέλο m_{k+1} , με μορφή όμοια με το μοντέλο (2.1), έχουμε προηγουμένως μεταφερθεί από ένα σημείο x_k στο τρέχον σημείο x_{k+1} . Οι παράμετροι του μοντέλου f_{k+1} , g_{k+1} , G_{k+1} βρίσκονται λύνοντας το πρόβλημα

$$\begin{aligned} \min_{f,g,G} \|G - G_k\|_F^2 \\ G: \text{συμμετρικός} \\ m(y^l) = f(y^l), \quad l = 1, 2, \dots, \hat{q}, \end{aligned} \quad (2.5)$$

όπου $\|\cdot\|_F$ είναι η νόρμα Frobenius, G_k ο εσσιανός του προηγούμενου μοντέλου m_k , και \hat{q} ένας ακέραιος, συνήθως $\hat{q} = 2n + 1$.

Όπως φαίνεται το πρόβλημα είναι ένα τετραγωνικό πρόβλημα με ισοτικούς περιορισμούς οι οποίοι μάλιστα μπορούν να γραφούν με τη μορφή συστήματος. Όπως και προηγουμένως, θέτουμε ορισμένες προϋποθέσεις που πρέπει να πληροί ο αλγόριθμος ώστε να βελτιώσουμε τη λειτουργία του, όπως η επάρκεια του συνόλου των σημείων παρεμβολής Y . Αφενός, το σύνολο πρέπει να είναι τέτοιο ώστε το σύστημα που προκύπτει από τις εξισώσεις παρεμβολής του προβλήματος (2.5) να έχει λύση για οποιοδήποτε δεξί μέλος, δηλαδή για οποιαδήποτε συνάρτηση f . Αφετέρου, θέλουμε τα σημεία y^i να μην βρίσκονται όλα σε ένα υπερεπίπεδο του \mathbb{R}^n . Αν ικανοποιούνται τα παραπάνω, τότε το πρόβλημα λύνεται κατά μοναδικό τρόπο.

Η επιλογή της νόρμας του Frobenius δίνει μία πολύ ενδιαφέρουσα ιδιότητα όταν η αντικειμενική συνάρτηση f είναι τετραγωνική, η οποία εξηγεί γιατί αυτή η μέθοδος λειτουργεί. Συγκεκριμένα, θέτοντας ως G^* τη λύση του προβλήματος (2.5), έχουμε ότι ([3], σελ. 18-19)

$$\|\nabla^2 f - G^*\|_F^2 = \|\nabla^2 f - G_k\|_F^2 - \|G^* - G_k\|_F^2 \leq \|\nabla^2 f - G_k\|_F^2. \quad (2.6)$$

Γεωμετρικά, η παραπάνω σχέση μας λέει ότι $\nabla^2 f - G^*$ και $G^* - G_k$ είναι ορθογώνια μεταξύ τους. Αν m_{k+1} είναι το μοντέλο που προκύπτει με τον υπολογισμό του G^* , η διαφορά $f - m_{k+1}$ θα είναι ένα τετραγωνικό πολυώνυμο που θα μηδενίζεται στα σημεία παρεμβολής, x_i , $i = 1, 2, \dots, \hat{q}$. Οπότε, η κατασκευή του m_{k+1} συνεπάγεται ότι η ελάχιστη τιμή της συνάρτησης

$$\Phi(\theta) = \|(G^* - G_k) + \theta(\nabla^2 f - G_k)\|_F^2, \quad \theta \in \mathbb{R}$$

προκύπτει όταν $\theta = 0$, δηλαδή όταν [4]

$$\sum_{i=1}^n \sum_{j=1}^n (G^* - G_k)_{ij} (\nabla^2 f - G_k)_{ij} = 0. \quad (2.7)$$

Πράγματι, παραγωγίζοντας τη συνάρτηση Φ έχουμε τα ακόλουθα

$$\begin{aligned}\Phi'(\theta) &= 0 \\ \sum_{i=1}^n \sum_{j=1}^n 2[(G^* - G_k)_{ij} + \theta(\nabla^2 f - G^*)_{ij}](\nabla^2 f - G^*)_{ij} &= 0 \\ 2 \left(\sum_{i=1}^n \sum_{j=1}^n (G^* - G_k)_{ij} (\nabla^2 f - G^*)_{ij} + \sum_{i=1}^n \sum_{j=1}^n \theta (\nabla^2 f - G^*)_{ij} (\nabla^2 f - G^*)_{ij} \right) &= 0 \\ \theta \sum_{i=1}^n \sum_{j=1}^n (\nabla^2 f - G^*)_{ij} (\nabla^2 f - G^*)_{ij} &= - \sum_{i=1}^n \sum_{j=1}^n (G^* - G_k)_{ij} (\nabla^2 f - G^*)_{ij} \\ \theta &= - \frac{\sum_{i=1}^n \sum_{j=1}^n (G^* - G_k)_{ij} (\nabla^2 f - G^*)_{ij}}{\sum_{i=1}^n \sum_{j=1}^n (\nabla^2 f - G^*)_{ij} (\nabla^2 f - G^*)_{ij}},\end{aligned}$$

και από την εξίσωση (2.6) παίρνουμε τη σχέση (2.7), η οποία με τη σειρά της μας δίνει $\theta = 0$.

Όπως είπαμε πριν, η σχέση (2.7) ουσιαστικά λέει ότι οι πίνακες $G^* - G_k$ και $\nabla^2 f - G^*$ είναι "ορθογώνιοι" μεταξύ τους, εξηγώντας έτσι την ισότητα στη σχέση (2.6). Η ανισότητα προκύπτει φυσικά. Από τα παραπάνω συμπεραίνουμε επίσης ότι τα σφάλματα μεταξύ του τετραγωνικού μοντέλου και της αντικειμενικής συνάρτησης μειώνονται σε κάθε επανάληψη.

Το πρόβλημα (2.5) είναι ένα πρόβλημα τετραγωνικού προγραμματισμού με ισοτικούς περιορισμούς, οπότε μπορεί να γραφεί ως ένα γραμμικό σύστημα εξισώσεων. Από τους περιορισμούς του προβλήματος διαπιστώνουμε ότι ([3], σελ. 19)

$$G^* = G_k + \sum_{j=1}^{\hat{q}} \lambda_j x_j x_j^T,$$

όπου οι συντελεστές λ_i , $i = 1, 2, \dots, \hat{q}$ ικανοποιούν τη σχέση

$$X\lambda = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_{\hat{q}} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{\hat{q}} \end{pmatrix} = 0. \quad (2.8)$$

Επομένως, από τις εξισώσεις παρεμβολής, οι παράμετροι c και g που υπολείπονται για τον προσδιορισμό του μοντέλου μας, καθώς και το διάνυσμα λ , μπορούν να υπολογιστούν μέσω της επίλυσης του συστήματος [3],

$$\left(\begin{array}{c|c} A & X^T \\ \hline X & 0 \end{array} \right) \begin{pmatrix} \lambda \\ \hat{g} \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix}$$

όπου για τα στοιχεία του πίνακα $A \in \mathbb{R}^{\hat{q} \times \hat{q}}$ ισχύει

$$A_{ij} = \frac{1}{2} (x_i^T x_j)^2, \quad 1 \leq i, j \leq \hat{q},$$

ο πίνακας X είναι όπως τον ορίσαμε στη σχέση (2.8), το διάνυσμα $\hat{g} \in \mathbb{R}^{n+1}$ έχει ως πρώτο στοιχείο την παράμετρο c και ακολουθούν τα στοιχεία του διανύσματος g , και τα στοιχεία του διανύσματος F είναι οι τιμές της συνάρτησης $f(x_i)$, $i = 1, 2, \dots, \hat{q}$.

Εφόσον έχουμε βρει το μοντέλο m_{k+1} , υπολογίζουμε το νέο βήμα μας λύνοντας ξανά ένα πρόβλημα τύπου περιοχής εμπιστοσύνης. Η μέθοδος που μόλις περιγράψαμε πλεονεκτεί υπέρ της μεθόδου αναβάθμισης του συνόλου παρεμβολής ως προς το πλήθος πράξεων. Αντί για τις $O(n^4)$ πράξεις που απαιτούνται για τη μέθοδο αναβάθμισης του συνόλου παρεμβολής σε κάθε επανάληψη, η παρούσα μέθοδος χρειάζεται $O(n^3)$ πράξεις ανά επανάληψη. Επίσης, χρειάζονται μόνο $O(n)$ σημεία παρεμβολής ώστε η μέθοδος να παράξει καλά αποτελέσματα. Αυτό συμβαίνει διότι η μέθοδος είναι κατασκευασμένη με τέτοιο τρόπο ώστε να απορροφά του εναπομείναντες βαθμούς ελευθερίας με την συνθήκη ελάχιστης αλλαγής του εσσιανού πίνακα. Τέλος, η εμπειρία μάς δείχνει ότι η μέθοδος πλησιάζει σε μία λύση ύστερα από $\frac{1}{2}(n+1)(n+2)$ πράξεις.

2.3 Μέθοδοι Συντεταγμένων και Μέθοδοι Αναζήτησης Μοτίβου

Αυτή η παράγραφος ακολουθεί το ([1], σελ. 229-234). Σε αυτές τις μεθόδους, αντί να κατασκευάσουμε ένα μοντέλο που να περιγράφει το πρόβλημα, κατασκευάζουμε έναν αλγόριθμο ο οποίος αναζητά ένα σημείο στο οποίο η συνάρτηση παίρνει χαμηλότερη τιμή, πάνω σε μία συγκεκριμένη διεύθυνση. Αν βρεθεί ένα τέτοιο σημείο, ο αλγόριθμος μεταβαίνει εκεί και ξεκινάει η επόμενη επανάληψη, ανανεώνοντας ενδεχομένως την κατεύθυνση αναζήτησης. Στην αντίθετη περίπτωση, εκτελείται μία δεύτερη προσπάθεια προσαρμόζοντας πρώτα είτε την κατεύθυνση είτε το βήμα αναζήτησης.

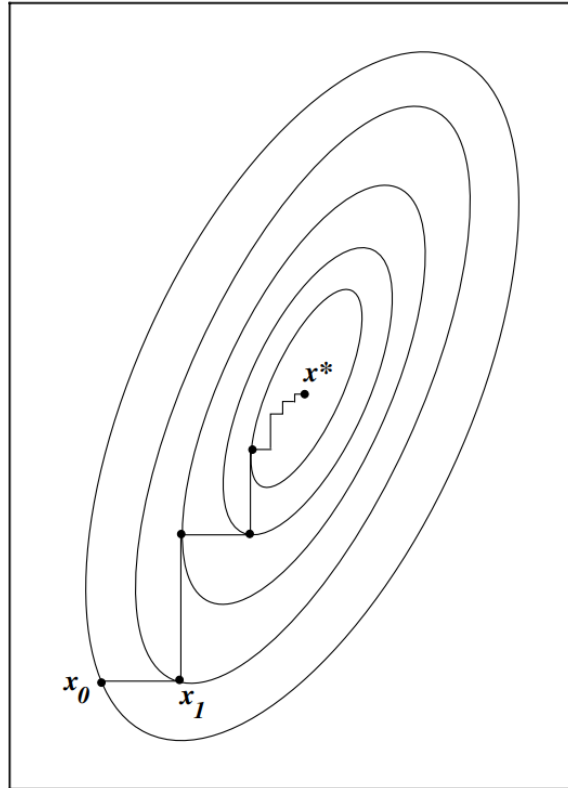
Μέθοδοι Συντεταγμένων

Οι Μέθοδοι Αναζήτησης Συντεταγμένων εκτελούν μία αναζήτηση τύπου Line Search στην κατεύθυνση που ορίζει κάθε μία από τις n συντεταγμένες, e_1, e_2, \dots, e_n περνώντας κυκλικά από όλες, έως ότου καταλήξουμε κοντά στη λύση του προβλήματος ελαχιστοποίησης. Συγκεκριμένα, ξεκινάμε με ένα αρχικό διάνυσμα $x \in \mathbb{R}$ και μεταβάλλουμε μόνο τη συνιστώσα x_1 , με σκοπό να ελαττώσουμε - να ελαχιστοποιήσουμε ιδανικά - την τιμή της αντικειμενικής συνάρτησης. Μόλις βρούμε μία τέτοια τιμή, μεταβάλλουμε μόνο τη συνιστώσα x_2 στο νέο διάνυσμα, και ούτω καθεξής.

Η μέθοδος αυτή είναι μία πολύ απλή διαδικασία που πηγάζει με φυσικό τρόπο από τη διαίσθησή μας. Στην πράξη όμως είναι μια αρκετά αναξιόπιστη μέθοδος. Μετά από μερικές επαναλήψεις, η πρόοδος της μεθόδου θα γίνεται ολοένα και μικρότερη, όπως φαίνεται στο σχήμα 2.1. θεωρητικά μπορεί ακόμα και να μην προσεγγίσει ποτέ ένα σημείο της αντικειμενικής συνάρτησης με μηδενική κλίση. Η εξήγηση αυτού του φαινομένου είναι ότι η κατεύθυνση μέγιστης καθόδου $-\nabla f_k$ γίνεται ολοένα και πιο κάθετη στις κατευθύνσεις που ορίζουν οι συνιστώσες, οπότε ικανοποιείται η συνθήκη Zoutendijk καθώς το $\cos\theta_k$ τείνει στο μηδέν χωρίς να συμβαίνει το ίδιο με το ∇f_k . Ακόμα και όταν η μέθοδος συγκλίνει σε μία λύση, η ταχύτητα σύγκλισης είναι φανερά μικρότερη από αυτή της μεθόδου μέγιστης καθόδου, με τη διαφορά να γίνεται πιο αισθητή όσο περισσότερες μεταβλητές έχουμε. Όμως, όταν οι μεταβλητές της συνάρτησης δεν περιπλέκονται μεταξύ τους, τότε η ταχύτητα σύγκλισης είναι αρκετά ικανοποιητική. Επιπλέον, ένα άλλο πλεονέκτημα της μεθόδου είναι ότι δεν απαιτείται κανένας υπολογισμός για την κλίση ∇f_k , γι αυτό και εντάσσονται στις DFO μεθόδους.

Φυσικά υπάρχουν πολλές παραλλαγές της μεθόδου, κυρίως ως προς τη σειρά αναζήτησης. Ενδεικτικά, σε μία παραλλαγή πραγματοποιείται μία 'παλινδρόμηση' στις συνιστώσες αναζήτησης, δηλαδή η σειρά των αναζητήσεων είναι

$$e_1, e_2, \dots, e_{n-1}, e_n, e_{n-1}, \dots, e_2, e_1, e_2, \dots$$



Σχήμα 2.1: Ενδεικτικά η πρόοδος της Μεθόδου Αναζήτησης Συντεταγμένων για μία συνάρτηση δύο μεταβλητών.

Μέθοδοι Αναζήτησης Μοτίβου

Στις Μεθόδους Αναζήτησης Μοτίβου ξεκινάμε με ένα τυχαίο σημείο ως σημείο εκκίνησης, έστω x_k , και εκτελούμε την εξής διαδικασία. Ορίζουμε ένα σύνολο από διαφορετικές κατευθύνσεις αναζήτησεων, D_k , και βάσει ενός προκαθορισμένου βήματος γ_k υπολογίζουμε την τιμή της συνάρτησης ως προς κάθε κατεύθυνση. Από το σύνολο των παραπάνω σημείων, τα οποία μπορούν να προσδιοριστούν ως $x_k + \gamma_k p_k$, για κάθε $p_k \in D_k$, επιλέγουμε αυτό που επιτυγχάνει τη μεγαλύτερη μείωση στη συνάρτηση και το θέτουμε ως σημείο εκκίνησης της νέας επανάληψης, πάντα με την προϋπόθεση ότι παρατηρείται ικανοποιητική μείωση, ενώ μπορούμε και να αυξήσουμε το βήμα γ_k . Αν η μείωση στη συνάρτηση δεν είναι αρκετή τότε ελαττώνουμε το βήμα, διατηρούμε το ίδιο σημείο εκκίνησης, και επαναλαμβάνουμε την παραπάνω διαδικασία. Σε κάθε περίπτωση μπορούμε να αλλάζουμε το σύνολο D_k πριν την έναρξη της νέας επανάληψης με στόχο να επιτύχουμε μεγαλύτερη μείωση στη συνάρτηση.

Τόσο η σύγκλιση όσο και η γενικότερη απόδοση της μεθόδου επηρεάζεται αρκετά από τον θόρυβο και άλλου είδους ανακρίβειες στις τιμές της συνάρτησης. Ωστόσο, σε αρκετές περιπτώσεις μπορεί να επιτευχθεί ακόμα και ολική σύγκλιση. Η μη ομαλότητα της αντικειμενικής συνάρτησης ενδέχεται να επηρεάσει αρνητικά τη συμπεριφορά της μεθόδου, αλλά υπάρχουν και περιπτώσεις μη ομαλών συναρτήσεων όπου η μέθοδος λειτουργεί χωρίς κανένα πρόβλημα.

Η επιλογή του συνόλου D_k είναι καθοριστικής σημασίας για τη συμπεριφορά της μεθόδου, τόσο σε πρακτικό όσο και σε θεωρητικό επίπεδο. Μία σημαντική συνθήκη που ευνοεί τα αποτελέσματα της μεθόδου είναι το σύνολο να περιέχει τουλάχιστον μία κατεύθυνση καθόδου για την f όταν $\nabla f(x_k) \neq 0$, όταν δηλαδή το x_k δεν είναι στάσιμο σημείο. Υπενθυμίζουμε ότι η γωνία μεταξύ της κλίσης ∇f_k και μιας κατεύθυνσης p ορίζεται ως

$$\cos\theta_k = \frac{-\nabla f_k^T p}{\|\nabla f_k\| \cdot \|p\|}.$$

Επίσης, η ολική σύγκλιση επιτυγχάνεται επιλέγοντας το σύνολο D_k ούτως ώστε για ένα τουλάχιστον $p \in D_k$ να ισχύει $\cos\theta_k \geq \delta$ για κάποιο $\delta > 0$, ανεξάρτητα από την τιμή του ∇f_k . Προσαρμόζοντας τα παραπάνω στα δεδομένα των Μεθόδων Μοτίβου Αναζήτησης, ορίζουμε ως μέτρο συνημιτόνου (cosine measure) την ποσότητα

$$\kappa(D_k) \stackrel{\text{def}}{=} \min_{v \in \mathbb{R}^n} \max_{p \in D_k} \frac{v^T p}{\|v\| \|p\|} \geq \delta.$$

Για να αποφύγουμε την ανεπιθύμητη συμπεριφορά της μεθόδου, θέλουμε η ποσότητα αυτή να μη γίνεται μικρότερη από ένα δεδομένο όριο, δηλαδή

$$\kappa_{\min} \leq \kappa(D_k) \text{ για κάθε } k = 1, 2, \dots.$$

Το μέτρο συνημιτόνου υπολογίζει την μέγιστη απόσταση μεταξύ της κατεύθυνσης μέγιστης καθόδου και του διανύσματος $p \in D_k$, το οποίο σχηματίζει τη μικρότερη γωνία με το διάνυσμα $u = -\nabla f(x)$.

Μία άλλη συνθήκη όσον αφορά το σύνολο D_k είναι τα μήκη των διανυσμάτων του να είναι περίπου ίσα. Αυτή η συνθήκη εξασφαλίζει την ομαλή λειτουργία της καθόδου, εξομαλύνοντας την οποιαδήποτε ανομοιογένεια στα βήματα της μεθόδου. Έτσι, θέλουμε για τα διανύσματα p να ισχύει ότι

$$\beta_{\min} \leq \|p\| \leq \beta_{\max}, \text{ για κάθε } p \in D_k,$$

για κάποιες σταθερές.

Μία από τις δύο συνήθειες επιλογές του συνόλου D_k είναι το σύνολο των κατευθύνσεων των συντεταγμένων, δηλαδή

$$D_k = \{e_1, e_2, \dots, e_n, -e_1, -e_2, \dots, -e_n\}. \quad (2.9)$$

Σε αυτή την περίπτωση η μέθοδος είναι παρόμοια με τη Μέθοδο Αναζήτησης Συντεταγμένων, και της προσδίδεται το όνομα 'Compass Search'. Σημαντική για αυτή τη μέθοδο είναι ότι $\kappa(D_k) = 0$ για όλα τα $k = 1, 2, \dots$. Κατά συνέπεια, σε κάθε επανάληψη το $\cos\theta$ μπορεί να είναι αυθαίρετα κοντά στο μηδέν. Η δεύτερη συνθήκη επιλογή του D_k γίνεται παίρνοντας τα $n + 1$ διανύσματα ως

$$p_i = \frac{1}{2n} e - e_i, \quad i = 1, 2, \dots, n \quad \text{και} \quad p_{n+1} = \frac{1}{2n} e,$$

όπου $e = (1, 1, \dots, 1)^T$.

Οι προαναφερθείσες ιδιότητες δεν ικανοποιούνται αυστηρά από όλες τις κατευθύνσεις του συνόλου D_k . Ορισμένες κατευθύνσεις του συνόλου μπορεί να μην τις ικανοποιούν, ενώ μπορεί να υπάρχουν και κατευθύνσεις που να μην ικανοποιούν τις επιθυμητές ιδιότητες, αλλά να μην περιλαμβάνονται στο σύνολο D_k της εκάστοτε επανάληψης. Οι πρώτες δεν επηρεάζουν με κάποιον τρόπο τον αλγόριθμο, μιας και δεν πρόκειται να επιλεγούν ως κατευθύνσεις επαρκούς μείωσης της συνάρτησης, καταλαβαίνουμε ωστόσο ότι είναι περιττές. Οι δεύτερες από τις κατευθύνσεις πιθανώς να επηρεάζουν σημαντικά την επίλυση του προβλήματος, οπότε καλό θα ήταν να τις λάβουμε υπόψιν μας κατά τη διάρκεια εκτέλεσης

του αλγορίθμου, ενσωματώνοντάς τις στο σύνολο D_k . Αυτό μπορεί να επιτευχθεί με διάφορους τρόπους: είτε ευριστικά, βάσει κάποιας προϋπάρχουσας γνώσης σχετικά με την συνάρτηση f ή σύμφωνα με τις πληροφορίες που μας παρέχουν οι προηγούμενες επαναλήψεις. Πιο απλά, μπορούμε να τις κατασκευάσουμε ως γραμμικούς συνδυασμούς των κατευθύνσεων που βρίσκονται εξαρχής στο σύνολο D_k και ικανοποιούν ταυτόχρονα τις παραπάνω συνθήκες.

Δεν είναι απαραίτητο να επιλέγουμε ως σημείο εκκίνησης της επόμενης επανάληψης το σημείο $x_k + \gamma_k p_k, p_k \in D_k$ που ελαχιστοποιεί τη συνάρτηση. Μία παραλλαγή του αλγορίθμου, αντί να υπολογίζει την τιμή της συνάρτησης για κάθε σημείο $x_k + \gamma_k p_k, p_k \in D_k$ και να επιλέγει τελικά το καλύτερο από αυτά, υπολογίζει την τιμή της συνάρτησης για ένα σημείο τη φορά, και ελέγχει αν η μείωση που παρατηρείται στη συνάρτηση f είναι ικανοποιητική, οπότε ξεκινάει τη νέα επανάληψη με αυτό ως σημείο εκκίνησης, είτε προχωράει στο επόμενο σημείο. Με αυτόν τον τρόπο ελαττώνουμε τις απαιτούμενες πράξεις της μεθόδου, ταυτόχρονα όμως ελαττώνεται και η ταχύτητα σύγκλισης.

Σε κάθε επανάληψη υπάρχει το ενδεχόμενο να αλλάξει το βήμα γ_k . Συγκεκριμένα, σε μία επιτυχημένη επανάληψη το βήμα μπορεί να γίνει $\gamma_{k+1} = \gamma_k \cdot \varphi_k$, με $\varphi_k \geq 1$, ενώ σε μία αποτυχημένη επανάληψη έχουμε $\gamma_{k+1} = \gamma_k \cdot \theta_k$, με $\theta_k \in (0, 1)$. Οι συνθήκες επιλογής για τους συντελεστές αύξησης ή μείωσης του βήματος της μεθόδου, είναι $\varphi_k = 1$ (οπότε το βήμα δεν αυξάνεται αλλά παραμένει σταθερό) και $\theta_k = \frac{1}{2}$ (οπότε το βήμα ελαττώνεται κατά το ήμισυ) αντίστοιχα. Ο συντελεστής θ_k δεν πρέπει να συγγέεται με τη γωνία του συνημιτόνου που ορίσαμε προηγουμένως. Θέτοντας έναν μεγαλύτερο συντελεστή φ_k , παίρνουμε ένα ρίσκο και κάνουμε πιο τολμηρά βήματα προς την κατεύθυνση μείωσης. Αυτό όμως μπορεί να έχει και αντίθετα αποτελέσματα, προσπερνώντας ένα κομβικό σημείο για την εύρεση του ελαχίστου της εκάστοτε συνάρτησης. Ομοίως, αν θέσουμε έναν πολύ μικρό συντελεστή θ_k , έχουμε μία ασφάλεια ως προς την πρόοδο της μεθόδου, ότι δηλαδή δεν προσπερνάμε κάποιο κομβικό σημείο, αλλά η πρόοδος της μεθόδου γίνεται σημαντικά μικρότερη. Οι συντελεστές φ_k και θ_k μπορούν και αυτοί να αλλάξουν σε κάθε επανάληψη.

Αξίζει να αναφερθεί και η σημαντικότητα του συντελεστή επαρκούς μείωσης $\rho(\cdot)$. Η συνάρτηση $\rho(t)$ θέλουμε να ικανοποιεί τις παρακάτω συνθήκες[5]:

$$\begin{aligned} \rho: & \text{συνεχής} \\ \lim_{t \rightarrow 0} \frac{\rho(t)}{t} &= 0 \\ \rho(t_1) &\leq \rho(t_2) \text{ για } t_1 < t_2. \end{aligned} \tag{2.10}$$

Η τελευταία συνθήκη λέει ότι η συνάρτηση $\rho(t)$ πρέπει να είναι μη φθίνουσα. Επίσης, μπορούμε να ορίσουμε $\rho = 0$. Σε αυτή την περίπτωση, οποιαδήποτε μείωση στην αντικειμενική συνάρτηση θεωρείται επαρκής. Αυτή όμως είναι μία πολύ ασθενής συνθήκη που αναστέλλει τη σύγκλιση της μεθόδου, ιδίως την καθολική σύγκλιση. Η επιλογή μίας καλής συνάρτησης $\rho(t)$ θέτει και μία καλή συνθήκη που κατ'επέκταση προσφέρει μία γρήγορη σύγκλιση με καλή ακρίβεια. Συνήθως ορίζεται ως $\rho(t) = Mt^{3/2}$, όπου M είναι μία θετική σταθερά.

Επανερχόμαστε στη μελέτη του συνόλου D_k . Ένα καλό σύνολο αναζήτησης θα πρέπει οπωσδήποτε να περιέχει ένα σύνολο κατευθύνσεων \mathcal{G}_k οι οποίες σε συνδυασμό με το βήμα μας να καλύπτουν όλο το χώρο \mathbb{R}^n . Επειδή όμως το βήμα γ_k είναι θετικό, θα πρέπει το σύνολο \mathcal{G}_k να είναι ένα Positive spanning set του \mathbb{R}^n . Ως Positive spanning set του \mathbb{R}^n ορίζουμε ένα σύνολο διανυσμάτων, έστω G , που καλύπτει τον χώρο \mathbb{R}^n με θετικά βήματα. Δηλαδή, μπορούμε να κατασκευάσουμε κάθε διάνυσμα $v \in \mathbb{R}^n$ ως γραμμικό συνδυασμό διανυσμάτων του G με θετικούς συντελεστές,

$$v = a_1 g_1 + a_2 g_2 + \dots + a_n g_n, \text{ με } a_i \geq 0, i = 1, 2, \dots, n.$$

Για την κατασκευή ενός τέτοιου συνόλου \mathcal{G}_k ισχύει το επόμενο Λήμμα ([5], Λήμμα 3.2)

Λήμμα 2.1.

Το σύνολο κατευθύνσεων $\mathcal{G}_k \subset D_k$ είναι ένα Positive spanning set του \mathbb{R}^n αν και μόνο αν για κάθε διάνυσμα $v \in \mathbb{R}^n$ με $v \neq 0$, υπάρχει $p \in \mathcal{G}_k$ τέτοιο ώστε

$$v^T p > 0.$$

Το γεωμετρικό ισοδύναμο του Λήμματος λέει ότι το σύνολο \mathcal{G}_k είναι ένα Positive spanning set του \mathbb{R}^n αν και μόνο αν το εσωτερικό κάθε ημιχώρου του \mathbb{R}^n περιέχει ένα διάνυσμα του συνόλου \mathcal{G}_k . Πράγματι λοιπόν, αν ισχύει το Λήμμα, τότε είναι βέβαιο ότι σε κάθε επανάληψη k υπάρχει ένα $p \in \mathcal{G}_k$, τέτοιο ώστε

$$-\nabla f(x_k)^T p > 0,$$

πράγμα που σημαίνει ότι υπάρχει μία κατεύθυνση καθόδου για τη συνάρτηση f , όποτε $\nabla f(x_k) \neq 0$. Προφανώς αν το \mathcal{G}_k είναι ένα Positive spanning set του \mathbb{R}^n , τότε $\kappa(\mathcal{G}_k) > 0$. Συνδυάζοντας τα παραπάνω, προκύπτει ότι υπάρχει τουλάχιστον ένα $p \in \mathcal{G}_k$ έτσι ώστε να ισχύει

$$\kappa(\mathcal{G}_k) \|\nabla f(x_k)\| \|p\| \leq -\nabla f(x_k)^T p, \quad k = 1, 2, \dots \quad (2.11)$$

Πρακτικά όμως, το γεγονός ότι δεν έχουμε κάποια πληροφορία για την ποσότητα $\nabla f(x)$, καθιστά αδύνατο να γνωρίζουμε ποιες από τις κατευθύνσεις $p \in D_k$ είναι κατευθύνσεις καθόδου, ακόμα και αν το σύνολο D_k περιέχει ένα Positive spanning set του \mathbb{R}^n . Για αυτό στη γενική μέθοδο ο αλγόριθμος αναζητά κατά μήκος όλων αυτών των διανυσμάτων.

Σε αυτό το σημείο επισημαίνουμε μία παραδοχή, χρήσιμη για την ανάλυση σύγκλισης της μεθόδου. Όταν μία επανάληψη k επιτυγχάνει επαρκή μείωση στη συνάρτηση f , θα λέμε ότι η επανάληψη είναι "επιτυχής", και συμβολίζουμε γράφοντας $k \in \mathcal{S}$. Αν δεν ικανοποιεί τη συνθήκη επαρκούς μείωσης, θα λέμε ότι η επανάληψη είναι "αποτυχημένη", και συμβολίζουμε γράφοντας $k \in \mathcal{U}$.

Το παρακάτω Θεώρημα ([5] Θεώρημα 3.3) δίνει ένα γενικό αποτέλεσμα για την ακολουθία των βημάτων των "αποτυχημένων" επαναλήψεων, $\{\gamma_k\}$, $k \in \mathcal{U}$.

Θεώρημα 2.1.

Έστω $f : \mathbb{R}^n \rightarrow \mathbb{R}$ μία συνεχώς διαφορίσιμη συνάρτηση, και έστω ότι η ∇f είναι Lipschitz συνεχής με σταθερά M . Τότε, η Μέθοδος Μοτίβου Αναζήτησης παράγει επαναλήψεις έτσι ώστε για κάθε $k \in \mathcal{U}$ να ισχύει

$$\|\nabla f(x_k)\| \leq \kappa(\mathcal{G}_k)^{-1} \left[M \gamma_k \beta_{max} + \frac{\rho(\gamma_k)}{\gamma_k \beta_{min}} \right].$$

Απόδειξη. Έστω ένα $d_k \in \mathcal{G}_k \subseteq D_k$ που ικανοποιεί τη σχέση (2.11). Άρα,

$$\kappa(\mathcal{G}_k) \|\nabla f(x_k)\| \|d_k\| \leq -\nabla f(x_k)^T d_k.$$

Από το Θεώρημα Μέσης Τιμής έχουμε ότι για κάποιο $\alpha_k \in [0, 1]$,

$$f(x_k + \gamma_k d_k) - f(x_k) = \gamma_k \nabla f(x_k + \alpha_k \gamma_k d_k)^T d_k.$$

Επειδή η επανάληψη k είναι αποτυχημένη, ισχύει ότι

$$0 \leq f(x_k + \gamma_k d_k) - f(x_k) + \rho(\gamma_k).$$

Συνδυάζοντας τις δύο τελευταίες σχέσεις παίρνουμε

$$\begin{aligned} 0 &\leq \gamma_k \nabla f(x_k + \alpha_k \gamma_k d_k)^T d_k + \rho(\gamma_k) \Rightarrow \\ 0 &\leq \nabla f(x_k + \alpha_k \gamma_k d_k)^T d_k + \frac{\rho(\gamma_k)}{\gamma_k} \Rightarrow \\ -\nabla f(x_k)^T d_k &\leq (\nabla f(x_k + \alpha_k \gamma_k d_k) - \nabla f(x_k))^T d_k + \frac{\rho(\gamma_k)}{\gamma_k} \end{aligned}$$

και άρα

$$\kappa(\mathcal{G}_k) \|\nabla f(x_k)\| \cdot \|d_k\| \leq (\nabla f(x_k + \alpha_k \gamma_k d_k) - \nabla f(x_k))^T d_k + \frac{\rho(\gamma_k)}{\gamma_k}.$$

Επειδή όμως ∇f είναι Lipschitz συνεχής, παίρνουμε

$$\begin{aligned} \kappa(\mathcal{G}_k) \|\nabla f(x_k)\| \cdot \|d_k\| &\leq \|\nabla f(x_k + \alpha_k \gamma_k d_k) - \nabla f(x_k)\| \cdot \|d_k\| + \left\| \frac{\rho(\gamma_k)}{\gamma_k} \right\| \Rightarrow \\ \kappa(\mathcal{G}_k) \|\nabla f(x_k)\| \cdot \|d_k\| &\leq M \|\alpha_k \gamma_k d_k\| \cdot \|d_k\| + \frac{\rho(\gamma_k)}{\gamma_k}. \end{aligned}$$

Επομένως,

$$\begin{aligned} \kappa(\mathcal{G}_k) \|\nabla f(x_k)\| &\leq M \gamma_k \|d_k\| + \frac{\rho(\gamma_k)}{\gamma_k \|d_k\|} \\ \kappa(\mathcal{G}_k) \|\nabla f(x_k)\| &\leq M \gamma_k \beta_{max} + \frac{\rho(\gamma_k)}{\gamma_k \beta_{min}}, \end{aligned}$$

και το ζητούμενο προκύπτει άμεσα. □

Ένα ανάλογο αποτέλεσμα ισχύει ακόμα και αν η ∇f δεν είναι Lipschitz συνεχής.

Για να εξασφαλίσουμε καθολική σύγκλιση στη μέθοδο, θα πρέπει $\nabla f(x_k) \rightarrow 0$ καθώς $k \rightarrow \infty$. Σύμφωνα με το προηγούμενο Θεώρημα, αν δείξουμε ότι $\gamma_k \rightarrow 0$ καθώς $k \rightarrow \infty$, τότε και $\|\nabla f(x_k)\| \rightarrow 0$, λόγω των συνθηκών (2.10). Στο επόμενο Θεώρημα ([5] Θεώρημα 3.4) αποδεικνύεται ακριβώς αυτό.

Θεώρημα 2.2.

Έστω f μία κάτω φραγμένη συνάρτηση. Έστω επίσης η συνάρτηση $\rho(t)$ που ικανοποιεί τις συνθήκες (2.10). Τότε, η Μέθοδος Μοτίβου Αναζήτησης παράγει επαναλήψεις τέτοιες ώστε

$$\liminf_{k \rightarrow \infty} \gamma_k = 0.$$

Απόδειξη. Έστω ότι αυτό δεν ισχύει. Τότε, υπάρχει ένα $\gamma^* > 0$ τέτοιο ώστε $\gamma^* \geq \gamma_k$ για κάθε k . Έτσι όμως, οι επιτυχημένες επαναλήψεις της μεθόδου, όπως τις ορίσαμε, μπορεί να είναι άπειρες (με άλλα λόγια, το σύνολο \mathcal{S} είναι άπειρο). Επίσης, επειδή η $\rho(t)$ είναι μη φθίνουσα, και από το κάτω φράγμα του γ_k , έχουμε ότι υπάρχει ένα $\rho^* > 0$ τέτοιο ώστε $\rho^* \leq \rho(\gamma_k)$ για κάθε k . Τώρα, για κάθε $k \in \mathcal{S}$, έχουμε

$$f(x_{k+1}) - f(x_k) \leq -\rho(\gamma_k) < -\rho^* < 0,$$

ενώ για κάθε άλλο k , (δηλαδή για κάθε $k \in U$), ισχύει $f(x_{k+1}) - f(x_k) = 0$. Οπότε προκύπτει $f(x_k) \rightarrow -\infty$, Άτοπο διότι η f είναι κάτω φραγμένη. Άρα ισχύει το ζητούμενο. □

Όπως φαίνεται από τις εκφωνήσεις των Θεωρημάτων 2.1 και 2.2 η σύγκλιση επιτυγχάνεται ασυμπτωτικά, δηλαδή όταν $k \rightarrow \infty$. Για να αποφύγουμε την επ' αόριστον επανάληψη της μεθόδου θέτουμε μία πρακτική συνθήκη τερματισμού, κατά την οποία η μέθοδος τερματίζει όταν το βήμα γίνει μικρότερο από ένα επιθυμητό όριο.

Η φιλοσοφία των Μεθόδων Μοτίβου Αναζήτησης μοιάζει αρκετά με αυτή της Μέγιστης Καθόδου, οπότε και οι ομοιότητες ως προς τη θεωρία σύγκλισης των δύο μεθόδων είναι αναμενόμενη. Οι Μέθοδοι Αναζήτησης Μοτίβου πλησιάζουν γρήγορα σε μία περιοχή ενός τοπικού ακρότατου, x^* , και μόλις συμβεί αυτό, είναι βέβαιο ότι όλες οι επόμενες επαναλήψεις θα πλησιάζουν σταδιακά στο x^* , με την προϋπόθεση ότι τα βήματα είναι αρκετά μικρά. Σε γενικές γραμμές θέλουμε να αφήσουμε τη μέθοδο να εξελιχθεί ελεύθερα αλλά όχι ανεξέλεγκτα. Μπορούμε και είναι φρόνιμο να επιτρέψουμε στη μέθοδο να κάνει ενίοτε κάποιο μεγαλύτερο άλμα, ειδικά όταν υπάρχουν και ενθαρρυντικές ενδείξεις, ακόμα κι αν χρειαστεί να παλινδρομήσει σε επόμενες επαναλήψεις. Επίσης, δε θέλουμε να περιορίσουμε τη μέθοδο επιβάλλοντας μία πολύ αυστηρή συνθήκη μείωσης. Τη συμπεριφορά της μεθόδου μπορούμε να την καθορίσουμε, έως ένα βαθμό, θέτοντας τις κατάλληλες παραμέτρους (φ_k , θ_k , και $\rho(t)$).

2.4 Μέθοδοι Συζυγών Κλίσεων

Στο πρώτο μέρος αναφέραμε ότι σε περιπτώσεις κυρτών τετραγωνικών συναρτήσεων,

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad (2.12)$$

ο σχηματισμός των συζυγών κλίσεων γίνεται συνήθως ως ένας γραμμικός συνδυασμός κλίσεων της συνάρτησης του προβλήματος. Εδώ θα δούμε έναν τρόπο σχηματισμού τους αξιοποιώντας μόνο τιμές της συνάρτησης, γεγονός που κατά συνέπεια μας δίνει τη δυνατότητα για τον σχεδιασμό ενός DFO αλγορίθμου.

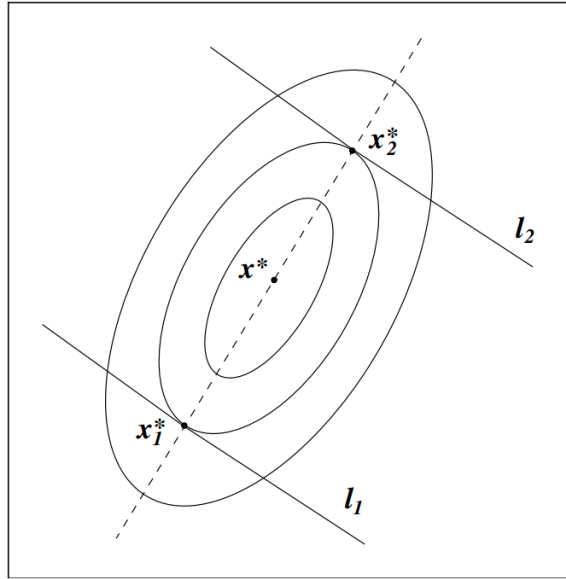
Για τον σχεδιασμό του αλγορίθμου θα πρέπει να αξιοποιήσουμε την πολύ σημαντική ιδιότητα των Παράλληλων Υποχώρων' (parallel subspace property), την οποία θα αποδείξουμε για διδιάστατα προβλήματα ($n = 2$). Έστω λοιπόν δύο παράλληλες γραμμές, $l_1(a) = x_1 + ap$ και $l_2(a) = x_2 + ap$, όπου x_1, x_2 , και p είναι γνωστά διανύσματα του \mathbb{R}^2 και $a \in \mathbb{R}$. Εάν x_1^*, x_2^* είναι τα ελάχιστα της συνάρτησης f επί των ευθειών l_1 και l_2 αντίστοιχα, προκύπτει ότι το διάνυσμα $x_1^* - x_2^*$ είναι συζυγές στο διάνυσμα p . Οπότε, το ελάχιστο x^* της συνάρτησης f θα βρίσκεται πάνω στην ευθεία που ορίζει το $x_1^* - x_2^*$, και είναι πολύ εύκολο να εντοπιστεί. Είμαστε σίγουροι ότι το σημείο x^* είναι το σημείο ελαχίστου της f καθώς έχουμε ελαχιστοποιήσει τη συνάρτηση κατά μήκος των δύο συζυγών κλίσεων (p και $x_1^* - x_2^*$). Μία γεωμετρική αναπαράσταση των παραπάνω φαίνεται και στο σχήμα 2.2.

Ας δούμε τώρα και τη γενική μορφή της ιδιότητας. Έστω x_1, x_2 δύο σημεία του \mathbb{R}^n και $\{p_1, p_2, \dots, p_l\}$ ένα σύνολο γραμμικώς ανεξάρτητων διανυσμάτων του \mathbb{R}^n . Οι δύο παράλληλοι υπόχωροι που χρειαζόμαστε δίνονται ως

$$S_1 = \left\{ x_1 + \sum_{i=1}^l \gamma_i p_i \mid \gamma_i \in \mathbb{R}, i = 1, 2, \dots, l \right\}$$

$$S_2 = \left\{ x_2 + \sum_{i=1}^l \gamma_i p_i \mid \gamma_i \in \mathbb{R}, i = 1, 2, \dots, l \right\}$$

Εάν θέσουμε ως x_1^*, x_2^* τα ελάχιστα της συνάρτησης f κατά μήκος των S_1 και S_2 αντίστοιχα, τότε το



Σχήμα 2.2: Γεωμετρική κατασκευή των συζυγών κατευθύνσεων.

$x_1^* - x_2^*$ είναι συζυγές ως προς όλα τα p_1, p_2, \dots, p_l . Πράγματι, για το x_1^* , όντας ελάχιστο, ισχύει

$$\left. \frac{\partial f(x_1^* + \gamma_i p_i)}{\partial \gamma_i} \right|_{\gamma_i=0} = \nabla f(x_1^*)^T p_i = 0, i = 1, 2, \dots, l,$$

ομοίως για το x_2 . Επίσης, αφού η f είναι τετραγωνική συνάρτηση, ισχύει ότι

$$f(x) = \frac{1}{2} x^T A x - b^T x + c.$$

Συνδυάζοντας τα παραπάνω έχουμε:

$$\begin{aligned} 0 &= (\nabla f(x_1^*) - \nabla f(x_2^*))^T p_i \\ &= (A x_1^* - b - A x_2^* + b)^T p_i \\ &= (x_1^* - x_2^*)^T A p_i, \quad i = 1, 2, \dots, l. \end{aligned}$$

Πρακτικά, για να υπολογίσουμε τις συζυγείς κατευθύνσεις ξεκινάμε με ένα τυχαίο σημείο x_0 και ένα σύνολο $\{p_1, p_2, \dots, p_n\}$ από γραμμικά ανεξάρτητες κατευθύνσεις (συνήθως παίρνουμε το σύνολο $\{e_1, e_2, \dots, e_n\}$), και ακολουθούμε την εξής διαδικασία. Για $k = 1, 2, \dots, n$, βρίσκουμε κάθε φορά το βαθμωτό γ_k για το οποίο η τιμή της συνάρτησης f ελαχιστοποιείται κατά μήκος της ευθείας $x_{k-1} + \gamma p_k$, και θέτουμε $x_k = x_{k-1} + \gamma_k p_k$. Αφού ολοκληρώσουμε αυτές τις επαναλήψεις, θα έχουμε υπολογίσει ένα x_n για το οποίο θα ισχύει

$$x_n = x_0 + \sum_{k=1}^n \gamma_k p_k.$$

Ύστερα, ανανεώνουμε το σύνολο των p_1, p_2, \dots, p_n θέτοντας

$$p_i = p_{i+1} \text{ για } i = 1, 2, \dots, n-1, \text{ και } p_n = x_n - x_0 = \sum_{k=1}^n \gamma_k p_k.$$

Ουσιαστικά, αντικαθιστούμε την κατεύθυνση p_1 με $x_n - x_0$, και μεταθέτουμε τις κατευθύνσεις $\{p_i\}_{i=1}^n$ μία θέση αριστερά. Κατόπιν υπολογίζουμε το ελάχιστο της συνάρτησης f κατά μήκος της ευθείας $x_0 + ap_n$, έστω x^* , και αντικαθιστούμε το x_0 με x^* . Αυτή η διαδικασία εύρεσης συζυγών κλίσεων είναι η διαδικασία Powell (Powell's Procedure[6]).

Μετά από αυτή τη διαδικασία έχουμε αντικαταστήσει την κατεύθυνση p_1 με το διάνυσμα $x_n - x_0 = \sum_{k=1}^n \gamma_k p_k$, και το αρχικό μας σημείο x_0 με το σημείο $x^* = x_0 + ap_n$. Αν επαναλάβουμε αυτή τη διαδικασία έως ότου αντικατασταθούν όλες οι αρχικές κατευθύνσεις p_1, p_2, \dots, p_n θα έχουμε κατασκευάσει και το σύνολο των συζυγών κατευθύνσεων. Με άλλα λόγια η διαδικασία στο σύνολό της αποτελείται από δύο επαναλήψεις, μία εσωτερική και μία εξωτερική. Η εσωτερική επανάληψη είναι αυτή που υπολογίζει τις συζυγείς κατευθύνσεις, ενώ η εξωτερική προσεγγίζει εκ νέου το ελάχιστο της αντικειμενικής συνάρτησης. Ύστερα από k ολοκληρωμένες εξωτερικές επαναλήψεις, οι κατευθύνσεις $p_{n-k}, p_{n-k+1}, \dots, p_n$ θα είναι συζυγείς μεταξύ τους, όπως υποδεικνύει η ιδιότητα parallel subspace property. Αυτό σημαίνει ότι ο αλγόριθμος θα τερματίσει στην ύστερα από $n-1$ επαναλήψεις μιας και θα έχουμε κατασκευάσει το σύνολο των συζυγών κατευθύνσεων. Ο αλγόριθμος θα έχει εκτελέσει $O(n^2)$ υπολογισμούς στη συνάρτηση. Εξαίρεση αποτελεί η περίπτωση όπου κάποια από τις συζυγείς κατευθύνσεις προκύπτει ίση με μηδέν, κάτι το οποίο δεν μπορεί να προβλεφθεί, άρα και να αποκλειστεί μέσω του αλγορίθμου.

Οι γραμμικές (μονοδιάστατες) αναζητήσεις μπορούν να πραγματοποιηθούν με τετραγωνική παρεμβολή αξιοποιώντας τρεις τιμές της συνάρτησης ως προς κάθε κατεύθυνση αναζήτησης. Εφόσον ο περιορισμός της σχέσης (2.12) σε μία γραμμή πρόκειται για ένα αυστηρά τετραγωνικό πρόβλημα, η τετραγωνική παρεμβολή ταιριάζει απόλυτα στο πρόβλημα, και για αυτό η εύρεση του ελαχίστου στη γραμμική αναζήτηση πρόκειται για μία εύκολη υπόθεση.

Όπως αναφέρθηκε προηγουμένως, η μέθοδος των Συζυγών κλίσεων μπορεί να εφαρμοστεί και σε μη τετραγωνικά προβλήματα, διορθώνοντας κατάλληλα τον αλγόριθμο. Η διόρθωση που πρέπει να γίνει αφορά τις γραμμικές αναζητήσεις, οι οποίες θα πρέπει να γίνουν με απλή παρεμβολή, ενώ επίσης ο υπολογισμός θα είναι προσεγγιστικός, σε αντίθεση με όσα αναφέρθηκαν σχετικά με το τετραγωνικό πρόβλημα. Επιπλέον, το πρόβλημα πιθανώς να μην είναι πια κυρτό, οπότε πρέπει να είμαστε ιδιαίτερα προσεκτικοί κατά τη διάρκεια των αναζητήσεων αυτών.

Εμπειρικά, η επέκταση του αλγορίθμου λειτουργεί καλά σε προβλήματα μικρών διαστάσεων, αλλά μπορεί να παρατηρηθεί μία τάση των κατευθύνσεων $\{p_i\}$ να γίνουν γραμμικά εξαρτημένες. Και αυτό μπορεί να διορθωθεί, εισάγοντας τις scaled directions

$$\hat{p}_i = \frac{p_i}{\sqrt{p_i^T A p_i}}, \quad i = 1, 2, \dots, n.$$

Αποδεικνύεται ότι η ποσότητα

$$|\det(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_n)| \quad (2.13)$$

μεγιστοποιείται όταν τα διανύσματα p_i είναι συζυγή ως προς τον πίνακα A , πράγμα που συνιστά ότι η αντικατάσταση μιας υπάρχουσας κατεύθυνσης με την πρόσφατα παραγόμενη συζυγή δε θα πρέπει να πραγματοποιείται όταν μειώνεται η παραπάνω ποσότητα. Η μεγιστοποίηση της ποσότητας είναι και το κριτήριο για το ποια κατεύθυνση p_i θα αντικαταστήσουμε.

Ειδικότερα, αν για τη νέα κατεύθυνση $x_n - x_0 = p_{n+1}$ ισχύει

$$\frac{p_{n+1}}{\sqrt{p_{n+1}^T A p_{n+1}}} = \sum_{i=1}^n a_i \cdot \frac{p_i}{\sqrt{p_i^T A p_i}},$$

το αποτέλεσμα της αντικατάστασης της κατεύθυνσης p_i με την p_{n+1} είναι να πολλαπλασιάσουμε την ποσότητα (2.13) με $|a_i|$ ([7], *The discarding criterion*). Έτσι, η τιμή της ορίζουσας γίνεται μέγιστη όταν το $|a_i|$ γίνεται μέγιστο. Επιπλέον, από τις ελαχιστοποιήσεις της διαδικασίας Powell, αν το βέλτιστο βήμα είναι γ_i , τότε η μείωση Δ_i που παρατηρείται στην αντικειμενική συνάρτηση είναι

$$\Delta_i = \gamma_i^2 p_i^T A p_i.$$

Άρα, η ποσότητα $\sqrt{\Delta_i}/|\gamma_i|$ μπορεί να χρησιμοποιηθεί ως μία εκτίμηση του $\sqrt{p_i^T A p_i}$.

Η παραπάνω στρατηγική μπορεί να ενσωματωθεί στον αλγόριθμο των Συζυγών Κατευθύνσεων. Αποδεικνύεται ότι οι scaled directions \hat{p}_i μπορούν να υπολογιστούν χωρίς τη χρήση του εσσιανού πίνακα A καθώς μπορούμε να αντλήσουμε τους επιθυμητούς όρους $p_i^T A p_i$ από τις γραμμικές αναζητήσεις (κατά μήκος του p_i) στα προηγούμενα βήματα του αλγορίθμου. Και για να βεβαιωθούμε ότι η τιμή της 9.28 δε μειώνεται, αρκεί μονάχα να συγκρίνουμε ορισμένες τιμές της συνάρτησης τις οποίες επίσης έχουμε βρει σε προηγούμενα βήματα του αλγορίθμου.

Ένας εναλλακτικός τρόπος για να αποφύγουμε την γραμμική εξάρτηση των κατευθύνσεων είναι κάθε n ή $n + 1$ επαναλήψεις να επαναφέρουμε τις κατευθύνσεις $\{p_1, p_2, \dots, p_n\}$ στις κατευθύνσεις $\{e_1, e_2, \dots, e_n\}$. Ή για να μη χάσουμε χρήσιμες πληροφορίες για την αντικειμενική συνάρτηση, θα μπορούσαμε να επαναφέρουμε τις κατευθύνσεις $\{p_i\}$ σε οποιοδήποτε σύνολο ορθογώνιων κατευθύνσεων $\{q_1, q_2, \dots, q_n\}$. Ειδικά στην περίπτωση που η συνάρτηση f είναι τετραγωνική, μπορούμε να επιλέξουμε τα ιδιοδιανύσματα του πίνακα A . Αν ο πίνακας A είναι γνωστός, ο υπολογισμός των ιδιοδιανυσμάτων είναι τετριμμένος. Αν όμως ο πίνακας A μας είναι άγνωστος, ο υπολογισμός πρέπει να γίνει με έναν διαφορετικό τρόπο, όπως περιγράφεται παρακάτω ([7], *Finding the principal vectors*).

Ας υποθέσουμε λοιπόν ότι έχουμε μία θετικά ορισμένη τετραγωνική συνάρτηση της μορφής (2.12), με τον πίνακα A να είναι άγνωστος. Σύμφωνα με τα παραπάνω, αν πραγματοποιήσουμε n επαναλήψεις της διαδικασίας Powell με $\gamma_1 \neq 0$ σε κάθε επανάληψη, τότε θα έχουμε και n μη μηδενικές συζυγείς κατευθύνσεις p_1, p_2, \dots, p_n . Λόγω των συζυγών κατευθύνσεων, αν $P = [p_1 \ p_2 \ \dots \ p_n]$ θα έχουμε

$$P^T A P = D,$$

όπου D ένας διαγώνιος πίνακας με θετικά διαγώνια στοιχεία d_i . Τα στοιχεία αυτά μπορούν να υπολογιστούν από τις ελαχιστοποιήσεις (εντός των επαναλήψεων της διαδικασίας Powell) της συνάρτησης

$$\varphi(a) = f(x_{i-1} + a p_i),$$

οι οποίες πραγματοποιούνται όπως είπαμε με τετραγωνική παρεμβολή. Κατά τη διάρκεια της τετραγωνικής παρεμβολής πρέπει να υπολογίσουμε τις δεύτερες διαφορές της φ , δηλαδή την ποσότητα $\varphi[a_0, a_1, a_2] = \varphi(a_0) - 2\varphi(a_1) + \varphi(a_2)$ για τρεις διαφορετικές τιμές a_0, a_1, a_2 . Προκύπτει ότι αν τα a_0, a_1, a_2 είναι ισαπέχοντα, τότε

$$\varphi[a_0, a_1, a_2] = p_i^T A p_i = d_i.$$

Οπότε, αν θέσουμε τους πίνακες $V = P D^{-1/2}$ και $H = A^{-1}$, εύκολα διαπιστώνουμε ότι

$$H = P D^{-1} P^T = V V^T.$$

Ο υπολογισμός του πίνακα V είναι εφικτός, άρα και του πίνακα H . Στόχος μας είναι ο πίνακας $Q = [q_1, q_2, \dots, q_n]$ να είναι ορθογώνιος και να αποτελείται από στήλες που είναι συζυγείς ως προς τον πίνακα A . Άρα και πάλι πρέπει να ισχύει

$$Q^T A Q = \Lambda,$$

όπου Λ ένας διαγώνιος πίνακας. Άρα όντως ο πίνακας Q θα έχει για στήλες τα ιδιοδιανύσματα του πίνακα A , ενώ τα στοιχεία λ_i του πίνακα Λ θα είναι οι αντίστοιχες ιδιοτιμές. Τελικά, για τον υπολογισμό του πίνακα Q , αξιολογούμε τη σχέση

$$Q^T H Q = \Lambda^{-1},$$

όπως αυτή προκύπτει από το γεγονός ότι ο πίνακας A είναι συμμετρικός ενώ ο πίνακας Q είναι ορθογώνιος. Δηλαδή, ο πίνακας Q εκτός από τον πίνακα των ιδιοδιανυσμάτων του A , είναι και ο πίνακας των ιδιοδιανυσμάτων του γνωστού σε εμάς πίνακα H .

Οι παραπάνω υπολογισμοί είναι ακριβείς υπό την προϋπόθεση ότι ο δείκτης κατάστασης του πίνακα H είναι καλός. Αν ο πίνακας έχει κακό δείκτη κατάστασης, τότε η παραπάνω διαδικασία μπορεί να γίνει διαφορετικά, παραδείγματος χάριν μέσω Singular Value Decomposition του πίνακα V .

Η διαδικασία Powell συνεχίζεται μέχρις ότου να ισχύει το κριτήριο ([7], *The stopping criterion*)

$$2\|x_i - x_{i-1}\|_2 \leq \varepsilon^{1/2}\|x_i\|_2 + t,$$

όπου $t > 0$ ένα όριο που επιλέγει ο χρήστης, και ε η ακρίβεια του υπολογιστή.

2.5 Μέθοδοι Nelder - Mead

Η μέθοδος αυτή είναι από τις πιο διαδεδομένες μεθόδους άνευ παραγώγου, από τη στιγμή της εφευρέσής τους [8]. Το πλήρες όνομα της μεθόδου είναι Nelder - Mead simplex - reflexion, ωστόσο δεν έχει καμία σχέση με τη συνηθισμένη μέθοδο simplex. Το προσωνύμιο 'simplex' δίνεται μόνο επειδή σε οποιοδήποτε βήμα του αλγορίθμου μας απασχολούν μόνο $n+1$ στοιχεία του \mathbb{R}^n , των οποίων το κυρτό περίβλημα σχηματίζει ένα στοιχείο simplex. Δεδομένου ενός simplex S με κορυφές $\{x_1, x_2, \dots, x_{n+1}\}$ μπορούμε να σχεδιάσουμε τον πίνακα $V(S)$ που οι στήλες του είναι οι εξισώσεις των 'ακμών' του από μία κορυφή, του x_1 λόγου χάριν. Δηλαδή έχουμε

$$V(S) = \{x_2 - x_1, x_3 - x_1, \dots, x_{n+1} - x_1\}. \quad (2.14)$$

Ο πίνακας $V(S)$ ονομάζεται και πίνακας κατευθύνσεων του στοιχείου simplex. Ορίζουμε επίσης τον n -διάστατο όγκο και τη διάμετρο του simplex ως

$$vol(S) = \frac{|det(V(S))|}{n!}$$

$$diam(S) = \max_{1 \leq i, j \leq n+1} \|x_i - x_j\|$$

αντίστοιχα. Στον ορισμό της διαμέτρου χρησιμοποιούμε την l^2 νόρμα, ή την επαγόμενη νόρμα πίνακα.

Ο πίνακας (2.14) μας δίνει πολλές πληροφορίες για το ίδιο το simplex. Αν ο πίνακας είναι ομαλός, τότε και το simplex ονομάζεται ομαλό ή μη εκφυλισμένο. Με άλλα λόγια, όταν ο πίνακας $V(S)$ είναι μη αντιστρέψιμος (μη ομαλός), τότε το simplex εκφυλίζεται σε ένα μικρότερης διάστασης. Παρατηρούμε επίσης πως όταν το simplex είναι μη ομαλό, τότε ο n -διάστατος όγκος του μηδενίζεται.

Σε μία επανάληψη της μεθόδου, προσπαθούμε να εντοπίσουμε την κορυφή που προσδίδει στη συνάρτηση τη χειρότερη, δηλαδή τη μέγιστη τιμή, ώστε να την αντικαταστήσουμε με μία καλύτερη. Το καινούργιο σημείο πρέπει να κείται στην ευθεία που ενώνει τη χειρότερη κορυφή με το κεντροειδές των υπόλοιπων κορυφών, και το βρίσκουμε ανακλώντας, επεκτείνοντας, ή συμπιέζοντας το simplex πάνω σε αυτή τη γραμμή. Εάν αυτή η τακτική αποτύχει να δώσει κάποιο καλύτερο σημείο, τότε σμικρύνουμε το simplex διατηρώντας μόνο την κορυφή με την μικρότερη τιμή της συνάρτησης.

Προτού μελετήσουμε αναλυτικά τον αλγόριθμο, δίνουμε μερικούς απαραίτητους ορισμούς. Έστω λοιπόν ότι έχουμε ένα στοιχείο simplex με κορυφές τα σημεία $\{x_1, x_2, \dots, x_{n+1}\}$, διατεταγμένα έτσι ώστε να ισχύει

$$f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1}). \quad (2.15)$$

Προφανώς το σημείο που πρέπει να αντικατασταθεί είναι το x_{n+1} . Το κεντροειδές των υπόλοιπων σημείων ορίζεται ως

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Η ευθεία που διέρχεται από το κεντροειδές \bar{x} και το χειρότερο σημείο x_{n+1} περιγράφεται μέσω της εξίσωσης

$$x(t) = \bar{x} + t(x_{n+1} - \bar{x}) = (1-t)\bar{x} + tx_{n+1}. \quad (2.16)$$

Εφόσον έχουν γίνει οι παραπάνω υπολογισμοί, σειρά έχει να αντικαταστήσουμε το σημείο x_{n+1} με κάποιο που να βρίσκεται πάνω στην ευθεία (2.16) (τα σημεία υπολογίζονται αντικαθιστώντας με μία τιμή του t), και που να δίνει ταυτόχρονα μία καλύτερη τιμή της αντικειμενικής συνάρτησης. Τα σημεία που εξετάζουμε ως πιθανούς αντικαταστάτες είναι τα παρακάτω, και μάλιστα τα εξετάζουμε αυστηρά με την αναγραφόμενη σειρά προτεραιότητας:

1. **Σημείο Ανάκλασης (Reflection Point)**, $x(t_r)$: Το σημείο αυτό βρίσκεται στην προέκταση του ευθύγραμμου τμήματος $x_{n+1}\bar{x}$, προς τη μεριά του \bar{x} .
2. **Σημείο Επέκτασης (Expansion Point)**, $x(t_e)$: Το σημείο βρίσκεται επίσης στην προέκταση του ευθύγραμμου τμήματος $x_{n+1}\bar{x}$, προς τη μεριά του \bar{x} , είναι όμως πιο απομακρυσμένο από το σημείο ανάκλασης.
3. **Σημείο Εξωτερικής Συμπίεσης (Outside Contraction Point)**, $x(t_{oc})$: Το σημείο αυτό βρίσκεται μεταξύ του κεντροειδούς \bar{x} και του σημείου ανάκλασης $x(t_r)$.
4. **Σημείο Εσωτερικής Συμπίεσης (Inside Contraction Point)**, $x(t_{ic})$: Το σημείο βρίσκεται μεταξύ του κεντροειδούς \bar{x} και του σημείου x_{n+1} .

Οι τιμές της συνάρτησης που αντιστοιχούν σε καθένα από αυτά τα σημεία, για λόγους απλούστευσης και συντομίας θα συμβολίζονται ως $f(x(t_r)) = f_r$, $f(x(t_e)) = f_e$, $f(x(t_{oc})) = f_{oc}$, και $f(x(t_{ic})) = f_{ic}$.

Αν για το σημείο ανάκλασης ισχύει $f(x_1) \leq f_r < f(x_n)$, τότε έχουμε πετύχει να βρούμε ένα σημείο καλύτερο από το x_{n+1} . Οπότε, κάνουμε την αντικατάσταση και ξεκινάμε μία νέα επανάληψη. Περαιτέρω, αν $f_r < f(x_1)$, αυτό αποτελεί ένα δείγμα ότι η συνάρτηση βελτιώνεται σε εξαιρετικό βαθμό, οπότε αξίζει να επεκτείνουμε την αναζήτησή μας υπολογίζοντας το σημείο $x(t_e)$. Αν και αυτό επιτυγχάνει καλύτερη τιμή από το σημείο ανάκλασης, δηλαδή $f_e < f_r$, τότε το σημείο επέκτασης είναι αυτό που θα αντικαταστήσει το σημείο x_{n+1} . Διαφορετικά, διατηρούμε το σημείο ανάκλασης ως αντικαταστάτη του x_{n+1} .

Αντιθέτως, αν για το σημείο ανάκλασης ισχύει $f(x_n) \leq f_r < f(x_{n+1})$, αυτό υποδηλώνει πως έχουμε επεκταθεί υπερβολικά, πετυχαίνοντας όμως την ίδια στιγμή και μία μείωση στην τιμή της συνάρτησης. Άρα, πρέπει να κάνουμε μία όχι τόσο αλματώδη αναζήτηση, επιμένοντας όμως προς αυτή την κατεύθυνση. Υπολογίζουμε λοιπόν το σημείο $x(t_{oc})$. Αν $f_{oc} \leq f_r$, το σημείο εξωτερικής συμπίεσης είναι προτιμότερη επιλογή ως αντικαταστάτης του x_{n+1} .

Με την ίδια λογική, αν $f_r \geq f(x_{n+1})$, η αναζήτησή μας έχει αποτύχει πλήρως. Οπότε, ως τελευταία επιλογή, κάνουμε την πιο συντηρητική αναζήτηση που μπορούμε, υπολογίζοντας το σημείο εσωτερικής συμπίεσης. Αν για το σημείο ισχύει $f_{ic} < f(x_{n+1})$, το αντικαθιστούμε στη θέση του x_{n+1} και ξεκινάμε την επόμενη επανάληψη.

Για τις τιμές των συντελεστών ισχύει $t_e < t_r < t_{oc} < 0 < t_{ic} < 1$. Επιπλέον, $t_e < -1$. Οι συνηθέστερες, σχεδόν καθολικές τιμές τους είναι ([1], 9.5 Nelder-Mead Method)

$$\{t_r, t_e, t_{oc}, t_{ic}\} = \left\{-1, -2, -\frac{1}{2}, \frac{1}{2}\right\}.$$

Στη συνήθη περίπτωση, η γεωμετρία των σημείων είναι η εξής: Το σημείο ανάκλασης $x(t_t)$ είναι το συμμετρικό του x_{n+1} ως προς το \bar{x} . Το σημείο επέκτασης $x(t_e)$ βρίσκεται από το x_{n+1} σε απόσταση διπλάσια από ότι το $x(t_r)$. Το σημείο εξωτερικής συμπίεσης είναι το μέσο των \bar{x} και $x(t_r)$, ενώ το σημείο εσωτερικής ανάκλασης είναι το μέσο των σημείων \bar{x} και x_{n+1} .

Αν όλα τα παραπάνω σημεία αποτύχουν να μειώσουν την αντικειμενική συνάρτηση, τότε οφείλουμε να σμικρύνουμε το στοιχείο simplex και να προσπαθήσουμε ξανά. Η σμίκρυνση γίνεται αντικαθιστώντας όλες τις κορυφές του simplex, πλην του καλύτερου σημείου x_1 , με σημεία που βρίσκονται πλησιέστερα σε αυτό. Έτσι, τα νέα σημεία υπολογίζονται ως

$$x'_i = x_1 - \frac{x_1 - x_i}{\sigma}, \quad i = 2, 3, \dots, n+1, \quad \text{με } \sigma > 1. \quad (2.17)$$

Συνήθως δίνουμε την τιμή $\sigma = 2$, οπότε τα νέα σημεία είναι τα μέσα των ευθυγράμμων τμημάτων x_1x_i , $i = 2, 3, \dots, n+1$. Ωστόσο, πρέπει να επισημάνουμε ότι κάτι τέτοιο είναι σπάνιο να συμβεί.

Σε κάθε περίπτωση, πριν ξεκινήσει η νέα επανάληψη πρέπει να διατάξουμε τους όρους ώστε να ισχύει η σχέση (2.15). Σε περίπτωση ισοβαθμίας μετά τη σμίκρυνση του στοιχείου simplex, δηλαδή όταν

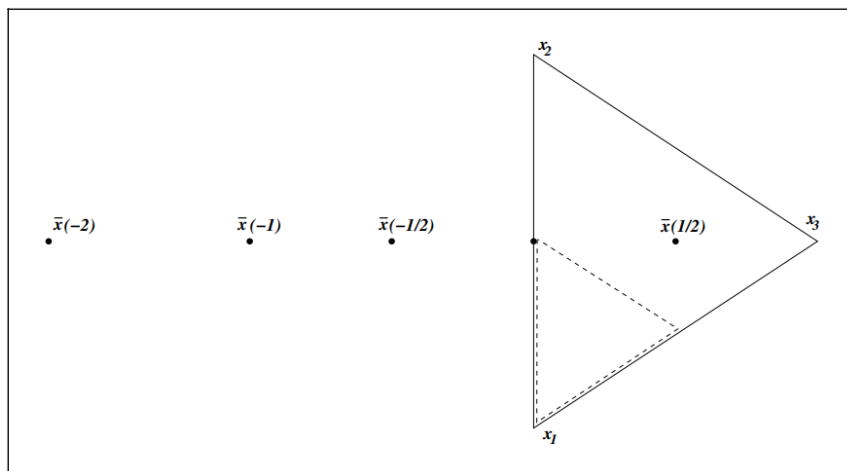
$$\min\{f(x'_2), \dots, f(x'_{n+1})\} = f(x_1),$$

τότε ως καλύτερο σημείο διατηρούμε το αρχικό μας x_1 .

Το σχήμα 2.3 παρουσιάζει τα βασικά στοιχεία της μεθόδου σε ένα παράδειγμα ενός simplex με τρεις κορυφές. Αρχικά το σημείο x_3 είναι το σημείο προς αντικατάσταση, με το κεντροειδές να είναι ουσιαστικά το μέσο του ευθύγραμμου τμήματος x_1x_2 . Επίσης φαίνονται τα υποψήφια σημεία που αναφέραμε προηγουμένως. Τέλος, αν ελέγξουμε τα σημεία και κανένα από αυτά δεν έχει μικρότερη τιμή από αυτή του x_3 , τότε συμπτύσσουμε το simplex διατηρώντας μόνο την καλύτερη κορυφή x_1 (το διακεκομμένο τρίγωνο στο σχήμα).

Πρακτικά, η μέθοδος Nelder - Mead έχει πολύ ικανοποιητική απόδοση, δεδομένης της απλότητάς της. Από την άλλη, η απλότητά της έχει ως αποτέλεσμα να παρατηρείται συχνά στασιμότητα σε μη βέλτιστα σημεία, ακόμα και σε πολύ απλές, ομαλές, και κυρτές συναρτήσεις. Σε αυτές τις περιπτώσεις, μία επανεκκίνηση της μεθόδου μπορεί να απαλείψει αυτό το πρόβλημα επιτυχώς. Τέλος, αξίζει να σημειωθεί ότι σε οποιαδήποτε περίπτωση, πλην της σύμπτυξης του στοιχείου simplex, η μέση τιμή της συνάρτησης,

$$\underline{f} = \frac{1}{n+1} \sum_{i=1}^{n+1} f(x_i),$$



Σχήμα 2.3: Τα υποψήφια σημεία της μεθόδου Nelder-Mead για τη συνήθη περίπτωση, σε μία συνάρτηση δύο μεταβλητών.

μειώνεται όταν η συνάρτηση f είναι κυρτή.

Μελετάμε τώρα κάποιες από τις σημαντικές ιδιότητες της μεθόδου. Αρχικά πρέπει να επισημάνουμε δύο ιδιαιτερότητες της μεθόδου. Πρώτον, στην περίπτωση της Εσωτερικής Συμπίεσης, μπορεί να προβούμε σε μία σμίκρυνση του simplex, παρόλο που έχουμε βρει ένα καλύτερο σημείο (το σημείο ανάκλασης). Δεύτερον, όταν σμικρύνουμε το simplex η μέση τιμή της συνάρτησης μπορεί τελικά να αυξηθεί, αντίθετα από τον επιθυμητή σε εμάς μείωση. Τέλος, η μη ομαλότητα του αρχικού στοιχείου simplex συνεπάγεται μη ομαλότητα και οποιουδήποτε άλλου παραγόμενου simplex, όπως αποδεικνύεται και στο παρακάτω Λήμμα ([9], Λήμμα 3.1):

Λήμμα 2.2.

1. Αν το αρχικό simplex S_0 είναι μη ομαλό, το ίδιο θα ισχύει και για όλα τα επόμενα στοιχεία simplex.
2. Όταν δεν σμικρύνουμε το simplex, τότε ισχύει $vol(S_{k+1}) = |t|vol(S_k)$.
3. Όταν σμικρύνουμε το simplex, τότε ισχύει $vol(S_{k+1}) = \frac{1}{\sigma^n} \cdot vol(S_k)$.

Απόδειξη. Η απόδειξη της πρότασης (1) προκύπτει άμεσα από τις προτάσεις (2) και (3), λαμβάνοντας υπόψιν ότι για το αρχικό simplex, όντας μη ομαλό, θα ισχύει $vol(S_0) = 0$

Αποδεικνύουμε τώρα τις άλλες δύο προτάσεις. Για την πρόταση (2), όταν δε σμικρύνουμε το simplex, τότε έχουμε απλώς αντικαταστήσει το χειρότερο σημείο x_{n+1} με ένα από τα σημεία της μορφής (2.16) που έχουμε αναφέρει πιο πάνω στην επεξήγηση της μεθόδου. Άρα, η τελευταία στήλη

του πίνακα $V(S)$, που είναι και η μόνη που θα αλλάξει, θα γίνει

$$\begin{aligned}
v_n^{k+1} &= x(t) - x_1 \\
v_n^{k+1} &= \bar{x} + t(x_{n+1} - \bar{x}) - x_1 \\
v_n^{k+1} &= (1-t)\bar{x} + tx_{n+1} - x_1 \\
v_n^{k+1} &= (1-t)\bar{x} + t(x_{n+1} - x_1) - (1-t)x_1 \\
v_n^{k+1} &= (1-t)(\bar{x} - x_1) + t(x_{n+1} - x_1) \\
v_n^{k+1} &= \frac{1-t}{n}(n\bar{x} - nx_1) + t(x_{n+1} - x_1) \\
v_n^{k+1} &= \frac{1-t}{n}(x_1 + x_2 + \dots + x_n - nx_1) + t(x_{n+1} - x_1) \\
v_n^{k+1} &= \frac{1-t}{n}((x_2 - x_1) + (x_3 - x_1) + \dots + (x_n - x_1)) + t(x_{n+1} - x_1) \\
v_n^{k+1} &= \frac{1-t}{n}(v_1^k + v_2^k + \dots + v_{n-1}^k) + tv_n^k.
\end{aligned}$$

Οι υπόλοιπες στήλες, $v_1^{k+1}, \dots, v_{n-1}^{k+1}$ θα παραμείνουν ίδιες, θα ισχύει δηλαδή $v_i^{k+1} = v_i^k, i = 1, 2, \dots, n-1$. Έτσι, για τον νέο πίνακα $V(S_{k+1})$ θα ισχύει

$$\begin{aligned}
|\det(V(S_{k+1}))| &= |\det(v_1^{k+1}, v_2^{k+1}, \dots, v_n^{k+1})| \\
|\det(V(S_{k+1}))| &= |\det(v_1^k, v_2^k, \dots, \frac{1-t}{n}(v_1^k + v_2^k + \dots + v_{n-1}^k) + tv_n^k)|,
\end{aligned}$$

και από τις ιδιότητες των οριζουσών παίρνουμε

$$\begin{aligned}
|\det(V(S_{k+1}))| &= |t \cdot \det(v_1^k, v_2^k, \dots, v_n^k)| \\
|\det(V(S_{k+1}))| &= |t| \cdot |\det(V(S_k))|.
\end{aligned}$$

Άρα,

$$vol(V(S_{k+1})) = |t| \cdot vol(V(S_k)).$$

Για την πρόταση (3) τώρα, κατά τη σύμκρυνση του simplex, αντικαθιστούμε τα σημεία x_i με τα σημεία που υποδεικνύει η σχέση (2.17). Έτσι, για κάθε στήλη v_i^{k+1} έχουμε:

$$v_i^{k+1} = x'_{i+1} - x_1 = x_1 - \frac{x_1 - x_i}{\sigma} - x_1 = \frac{x_1 - x_i}{\sigma} = \frac{v_i^k}{\sigma}, i = 1, 2, \dots, n.$$

Άρα, και για την ορίζουσα του $V(S_{k+1})$ θα ισχύει

$$\begin{aligned}
|\det(V(S_{k+1}))| &= |\det(v_1^{k+1}, v_2^{k+1}, \dots, v_n^{k+1})| \\
|\det(V(S_{k+1}))| &= \left| \det \left(\frac{v_1^k}{\sigma}, \frac{v_2^k}{\sigma}, \dots, \frac{v_n^k}{\sigma} \right) \right| \\
|\det(V(S_{k+1}))| &= \left| \frac{1}{\sigma^n} \cdot \det(v_1^k, v_2^k, \dots, v_n^k) \right|,
\end{aligned}$$

και αφού $\sigma > 0$, παίρνουμε τελικά

$$|\det(V(S_{k+1}))| = \frac{1}{\sigma^n} \cdot |\det(V(S_k))|,$$

και άρα

$$\text{vol}(V(S_{k+1})) = \frac{1}{\sigma^n} \cdot \text{vol}(V(S_k)),$$

ολοκληρώνοντας έτσι την απόδειξη. \square

Σύγκλιση της μεθόδου

Η αστάθεια της μεθόδου Nelder - Mead ακόμα και σε απλά προβλήματα καθιστά πολύ δύσκολη τη συγκρότηση μίας αέραςιας θεωρίας σύγκλισης. Ωστόσο, αν τηρούνται ορισμένες προϋποθέσεις στην αντικειμενική συνάρτηση μπορούμε να εξάγουμε κάποια βασικά συμπεράσματα για τη συμπεριφορά της μεθόδου.

Αρχικά, δίνουμε μερικούς επιπλέον ορισμούς που θα διευκολύνουν τη μελέτη της σύγκλισης. Ξεκινάμε δίνοντας έναν παρόμοιο ορισμό με τον πίνακα $M(S)$, αυτή τη φορά όμως θα υπολογίσουμε τις 'αχμές' του simplex ως προς την καλύτερη κορυφή, x_1 . Υπενθυμίζουμε λοιπόν τον πίνακα κατευθύνσεων του simplex,

$$V(S) = (x_2 - x_1, x_3 - x_1, \dots, x_{n+1} - x_1) = (v_1, v_2, \dots, v_n). \quad (2.18)$$

Πλέον θα αναφερόμαστε στον δείκτη κατάστασης $\kappa(V)$ ως κατάσταση του simplex. Ορίζουμε επίσης το διάνυσμα των διαφορών της αντικειμενικής συνάρτησης,

$$\delta(f : S) = (f(x_2) - f(x_1), f(x_3) - f(x_1), \dots, f(x_{n+1}) - f(x_1))^T,$$

και τα δύο προσανατολισμένα μήκη,

$$\sigma_+(S) = \max_{2 \leq j \leq n+1} \|x_1 - x_j\| \text{ και } \sigma_-(S) = \min_{2 \leq j \leq n+1} \|x_1 - x_j\|,$$

τα οποία θα χρησιμοποιήσουμε αντί της διαμέτρου $\text{diam}(S)$. Προφανώς ισχύει

$$\sigma_+(S) \leq \text{diam}(S) \leq 2\sigma_+(S).$$

Τέλος, δίνουμε τον σημαντικότερο ορισμό για τη σύγκλιση της μεθόδου. Ορίζουμε ως διάνυσμα κλίσης ενός στοιχείου simplex (simplex gradient) με ομαλό πίνακα $V(S)$ την ποσότητα

$$D(f : S) = V^{-T} \delta(f : S). \quad (2.19)$$

Από τα δύο Λήμματα που ακολουθούν ([10], Λήμματα 2.2 και 2.3), το πρώτο χρησιμεύει ως εφελθτήριο για τον ορισμό του διανύσματος κλίσης του simplex, ενώ το δεύτερο αποτελεί ουσιαστικά μία γενίκευση του πρώτου, προσαρμοσμένο στις συναρτήσεις που προκύπτουν σε φυσικά προβλήματα.

Λήμμα 2.3.

Εστω S ένα στοιχείο simplex με διατεταγμένες κορυφές έτσι ώστε να ισχύει η σχέση (2.15). Εστω επίσης ότι η ∇f είναι Lipschitz συνεχής σε μία περιοχή του S με σταθερά Lipschitz $2K_f$. Τότε, υπάρχει $K = K(K_f) > 0$, τέτοιο ώστε

$$\|\nabla f(x_1) - D(f : S)\| \leq K\kappa(V)\sigma_+(S). \quad (2.20)$$

Απόδειξη. Από την ομαλότητα της αντικειμενικής συνάρτησης και από το θεώρημα Taylor έχουμε, για κάθε $1 \leq j \leq n$

$$\begin{aligned} f(x_{j+1}) - f(x_1) - \nabla f(x_1)^T v_j &= \int_0^1 \nabla f(x_1 + tv_j)^T v_j dt - \nabla f(x_1)^T v_j \\ f(x_1) - f(x_{j+1}) + v_j^T \nabla f(x_1) &= - \int_0^1 v_j^T (\nabla f(x_1 + tv_j) - \nabla f(x_1)) dt. \end{aligned}$$

Οπότε, για κάθε $1 \leq j \leq n$,

$$\begin{aligned} \|f(x_1) - f(x_{j+1}) + v_j^T \nabla f(x_1)\| &= \left\| - \int_0^1 v_j^T (\nabla f(x_1 + tv_j) - \nabla f(x_1)) dt \right\| \\ &\leq \int_0^1 \|v_j^T (\nabla f(x_1 + tv_j) - \nabla f(x_1))\| dt \\ &\leq \int_0^1 \|v_j\| \cdot \|\nabla f(x_1 + tv_j) - \nabla f(x_1)\| dt \\ &\leq \int_0^1 2K_f t \|v_j\|^2 dt \\ &\leq K_f \|v_j\|^2 \\ &\leq K_f \sigma_+(S)^2. \end{aligned}$$

Επομένως,

$$\begin{aligned} \|\delta(f : S) - V^T \nabla f(x_1)\| &\leq n^{1/2} K_f \sigma_+(S)^2 \\ \|\nabla f(x_1) - D(f : S)\| &\leq n^{1/2} K_f \sigma_+(S)^2 \|V^{-T}\|. \end{aligned}$$

Το ζητούμενο έπεται θέτοντας $K = n^{1/2} K_f$ και από το γεγονός ότι $\sigma_+(S) \leq \|V\|$. Πράγματι, έχουμε τα εξής

$$\begin{aligned} \sigma_+(S) &= \max_{2 \leq j \leq n+1} \|x_j - x_1\| = \sqrt{\left(\max_{2 \leq j \leq n+1} \|x_j - x_1\| \right)^2} \\ &\leq \sqrt{\sum_{j=2}^{n+1} \|x_j - x_1\|^2} = \|V\|. \end{aligned}$$

□

Για το δεύτερο Λήμμα, θεωρούμε ότι η αντικειμενική συνάρτηση ικανοποιεί τη σχέση

$$f(x) = g(x) + \varphi(x), \quad (2.21)$$

όπου g είναι μία συνάρτηση ομαλή και απλή στη βελτιστοποίησή της, ενώ η φ είναι μία διαταραχή που δέχεται χαμηλές τιμές. Τέτοιες συναρτήσεις συναντώνται πολύ συχνά σε φυσικά προβλήματα. Όταν η Nelder - Mead εφαρμόζεται σε τέτοιες αντικειμενικές συναρτήσεις, χρειάζεται να υπολογίσουμε τη διαταραχή σε κάθε στοιχείο Simplex. Επομένως, ορίζουμε την αντίστοιχη ποσότητα ως

$$\|\varphi\|_S = \max_{1 \leq j \leq n+1} \|\varphi(x_j)\|.$$

Λήμμα 2.4.

Έστω ένα στοιχείο simplex S με διατεταγμένες κορυφές ώστε να ισχύει η σχέση (2.15). Έστω επίσης ότι η ∇g είναι Lipschitz συνεχής σε μία περιοχή του S με σταθερά Lipschitz $2K_g$. Τότε, υπάρχει $K = K(K_g) > 0$, τέτοιο ώστε

$$\|\nabla g(x_1) - D(f : S)\| \leq K\kappa(V) \left(\sigma_+(S) + \frac{\|\varphi\|}{\sigma_+(S)} \right).$$

Απόδειξη. Εφαρμόζοντας το Λήμμα 2.3 στη συνάρτηση g έχουμε

$$\|\nabla g(x_1) - D(g : S)\| \leq K_g n^{1/2} \kappa(V) \sigma_+(S).$$

Επίσης,

$$\begin{aligned} \|\delta(\varphi : S)\| &= \sqrt{\sum_{j=2}^{n+1} \|\varphi(x_j) - \varphi(x_1)\|^2} \leq \sqrt{n \cdot \max_{2 \leq j \leq n+1} \|\varphi(x_j) - \varphi(x_1)\|^2} \\ &\leq n^{1/2} \cdot 2 \max_{2 \leq j \leq n+1} \|\varphi(x_j)\| = 2n^{1/2} \|\varphi\|_S. \end{aligned}$$

και

$$\begin{aligned} D(g : S) - D(f : S) &= V^{-T} \delta(g : S) - V^{-T} \delta(f : S) \\ D(g : S) - D(f : S) &= -V^{-T} (\delta(f : S) - \delta(g : S)) \\ D(g : S) - D(f : S) &= -V^{-T} \delta(f - g : S) \\ D(g : S) - D(f : S) &= -V^{-T} \delta(\varphi : S) \end{aligned}$$

Συνδυάζοντας όλα τα παραπάνω, και λαμβάνοντας επιπλέον υπόψιν ότι $\sigma_+(S) \leq \|V\|$, παίρνουμε

$$\begin{aligned} \|\nabla g(x_1) - D(f : s)\| &\leq \|\nabla g(x_1) - D(g : S)\| + \|D(g : S) - D(f : S)\| \\ &\leq n^{1/2} K_g \kappa(V) \sigma_+(S) + \|-V^{-T} \delta(\varphi : S)\| \\ &\leq n^{1/2} K_g \kappa(V) \sigma_+(S) + \frac{\kappa(V)}{\|V\|} \cdot 2n^{1/2} \|\varphi\|_S \\ &\leq n^{1/2} K_g \kappa(V) \sigma_+(S) + 2n^{1/2} \frac{\kappa(V)}{\sigma_+(S)} \|\varphi\|_S. \end{aligned}$$

Πλέον, το ζητούμενο προκύπτει άμεσα θέτοντας $K = n^{1/2} K_g + 2n^{1/2}$. □

Κατά τη διάρκεια της μεθόδου κατασκευάζεται ουσιαστικά μία ακολουθία των στοιχείων simplex, S^k , και άρα ακολουθίες για όλα τα σχετικά μεγέθη που έχουμε ορίσει,

$$V^k = V(S^k), \quad \delta^k f = \delta(f : S^k), \quad K^k = K(S^k), \quad \text{και} \quad D^k f = D(f : S^k).$$

Αποδεικνύεται εύκολα ότι αν ο πίνακας V^0 είναι αντιστρέψιμος, τότε ο V^k είναι επίσης αντιστρέψιμος για κάθε $k > 0$. Κατά συνέπεια και το διάνυσμα $D^k f$ θα ορίζεται για κάθε $k > 0$.

Για να δώσουμε τελικά κάποια θεωρία σύγκλισης για τη μέθοδο Nelder - Mead, θα πρέπει να δεχτούμε τις παρακάτω προϋποθέσεις ([10], Assumption 2.1):

- ο πίνακας V^0 είναι αντιστρέψιμος.
- οι κορυφές των στοιχείων S^l , $l = 0, 1, \dots, k$ ικανοποιούν τη σχέση (2.15).
- για κάθε k ισχύει $\underline{f}^{k+1} < \underline{f}^k$.

Επίσης, υποθέτουμε ότι η αντικειμενική συνάρτηση είναι ομαλή. Εάν ισχύουν όλα τα παραπάνω, ισχύουν τα εξής δύο Θεωρήματα ([10], Θεωρήματα 3.1 και 3.2).

Θεώρημα 2.3.

Έστω μία ακολουθία από *simplex* που ικανοποιούν τις παραπάνω προϋποθέσεις, και έστω ότι ισχύουν και οι υποθέσεις του Λήμματος 2.3, με τις σταθερές Lipschitz K^k να είναι φραγμένες. Έστω επίσης ότι η ακολουθία $\{\underline{f}^k\}$ είναι κάτω φραγμένη. Τότε, αν ισχύει η σχέση (2.23) για ένα πεπερασμένο πλήθος k , και αν το γινόμενο $\sigma_+(S^k)\kappa(V^k) \rightarrow 0$, τότε κάθε σημείο σημείο συσσώρευσης των *simplex* είναι κρίσιμο σημείο της f .

Απόδειξη. Η υπόθεση ότι η $\{\underline{f}^k\}$ είναι κάτω φραγμένη σε συνδυασμό με τη σχέση (2.23) συνεπάγεται ότι η $\{\underline{f}^k\}$ συγκλίνει σε μία σταθερά. Επίσης, από την τρίτη προϋπόθεση και τη σχέση (2.23) συνεπάγεται ότι $\lim_{k \rightarrow \infty} D^k f = 0$. Άρα, βάζοντας όρια στη σχέση (2.20) του Λήμματος 2.3, έχουμε

$$\lim_{k \rightarrow \infty} \|\nabla f(x_1^k)\| \leq \lim_{k \rightarrow \infty} (K^k \kappa(V^k) \sigma_+(S^k) + \|D^k f\|) = 0$$

Άρα, εάν το σημείο x^* είναι σημείο συσσώρευσης της ακολουθίας $\{x_1^k\}$, τότε $\nabla f(x_k^*) = 0$. Αυτό ολοκληρώνει την απόδειξη, διότι $\sigma_+(S^k) \rightarrow 0$ μιας και $\kappa(V^k) \geq 1$, πράγμα που σημαίνει ότι οι κορυφές έχουν ένα κοινό σημείο συσσώρευσης. \square

Το επόμενο Θεώρημα είναι το ανάλογο του Θεωρήματος 2.3 για τις συναρτήσεις της μορφής (2.21), υποθέτοντας (όπως προκύπτει και από τη σχέση (2.22) του παρακάτω Θεωρήματος) ταυτόχρονα ότι $\|\varphi\|_{S^k} \rightarrow 0$

Θεώρημα 2.4.

Έστω μία ακολουθία από *simplex* που ικανοποιούν τις παραπάνω προϋποθέσεις, και έστω ότι ισχύουν και οι υποθέσεις του Λήμματος 2.4, με τις σταθερές Lipschitz K_g^k να είναι ομοιόμορφα φραγμένες. Έστω επίσης ότι η ακολουθία $\{\underline{f}^k\}$ είναι κάτω φραγμένη. Τότε, αν ισχύει η σχέση (2.23) για ένα πεπερασμένο πλήθος k , και αν

$$\lim_{k \rightarrow \infty} \kappa(V^k) \left(\sigma_+(S^k) + \frac{\|\varphi\|_{S^k}}{\sigma_+(S^k)} \right) = 0, \quad (2.22)$$

τότε κάθε σημείο συσσώρευσης των *simplex* είναι κρίσιμο σημείο της f .

Απόδειξη. Όπως και στο Θεώρημα 2.3 από τις υποθέσεις μας συνεπάγεται ότι $D^k f \rightarrow 0$. Από το Λήμμα 2.4 συνεπάγεται ότι

$$\|D^k g\| \leq \|D^k f\| + K^k \kappa(V^k) \left(\sigma_+(S^k) + \frac{\|\varphi\|_{S^k}}{\sigma_+(S^k)} \right),$$

με την ακολουθία $\{K^k\}$ να είναι φραγμένη. Έτσι, από τη σχέση (2.22) παίρνουμε ότι $D^k g \rightarrow 0$ καθώς $k \rightarrow \infty$. \square

Επανεκκίνηση της μεθόδου

Όπως και στις υπόλοιπες μεθόδους, είναι χρήσιμο να ορίσουμε μία συνθήκη επαρκούς μείωσης της συνάρτησης ως έναν έλεγχο για την απόδοση και την πορεία της μεθόδου. Μάλιστα, όπως είδαμε προηγουμένως, η συνθήκη αυτή παίζει σημαντικό ρόλο και για τη θεωρία σύγκλισης της μεθόδου. Η δομή της Nelder - Mead όμως καλύπει τη διαδικασία, καθώς σε αντίθεση με τις προηγούμενες μεθόδους δεν πραγματοποιείται ούτε καν μία προσέγγιση της παραγώγου της αντικειμενικής συνάρτησης f . Προηγουμένως, δώσαμε τον ορισμό (2.19) ο οποίος, αν και δεν έχει καμία σχέση με την παράγωγο, εξυπηρετεί ικανοποιητικά τις ανάγκες μας. Έτσι, ορμώμενοι από τη συνθήκη επαρκούς μείωσης (2.23) της Μεθόδου Μέγιστης Καθόδου, δίνουμε μία αντίστοιχη συνθήκη και για τη Nelder - Mead. Συγκεκριμένα, απαιτούμε στην $(k + 1)$ -οστή επανάληψη να ισχύει

$$\underline{f}^{k+1} - \underline{f}^k < -a \|D^k f\|^2, \quad (2.23)$$

όπου για την παράμετρο a ισχύει

$$a = a_0 \frac{\sigma_+(S)}{\|D^0 f\|}.$$

Συνήθως θέτουμε $a_0 = 10^{-4}$. Θεωρούμε πως όταν δεν ισχύει η σχέση (2.23) η μέθοδος τελματώνει ή πλησιάζει προς μία στάσιμη κατάσταση καθιστώντας την οποιαδήποτε πρόοδο αδύνατη. Όταν παρατηρείται κάτι τέτοιο, η συνθήκη τακτική είναι να επανεκκινήσουμε τη μέθοδο ([10], 3.2. *Oriented restarts*). Βασική προϋπόθεση είναι, πέρα από την παραβίαση της συνθήκης επαρκούς μείωσης (2.23), να υπάρχει τουλάχιστον μία μείωση στη μέση τιμή της f , δηλαδή $\underline{f}^{k+1} - \underline{f}^k < 0$.

Για την επανεκκίνηση της μεθόδου, βρίσκουμε ένα νέο μικρότερο simplex, αντικαθιστώντας τις κορυφές $\{x_i\}_{i=1}^{n+1}$ με τις νέες $\{y_i\}_{i=1}^{n+1}$, θέτοντας $y_1 = x_1$, και

$$y_i = y_1 + \beta_{i-1} e_{i-1}, \quad \text{για } 2 \leq i \leq n+1,$$

όπου

$$\beta_j = \frac{1}{2} = \begin{cases} \sigma_-(S^k) \text{sign}((D^k f)_j), & \alpha \nu (D^k f)_j \neq 0 \\ \sigma_-(S^k), & \alpha \nu (D^k f)_j = 0 \end{cases}$$

με $(D^k f)_j$ να είναι η συνιστώσα j του διανύσματος $(D^k f)$. Πριν τη διάταξη των σημείων, για το νέο στοιχείο simplex ισχύει $\text{diam}(S^{k+1}) = \frac{\sigma_-(S^k)}{\sqrt{2}}$. Πράγματι, από τον ορισμό της διαμέτρου και των σημείων y_i , έχουμε:

$$\begin{aligned} \text{diam}(S^{k+1}) &= \max_{1 \leq i, j \leq n+1} \|y_i - y_j\| \\ \text{diam}(S^{k+1}) &= \max_{1 \leq i, j \leq n+1} \left\| y_1 \pm \frac{1}{2} \sigma_-(S^k) e_{i-1} - y_1 \pm \frac{1}{2} \sigma_-(S^k) e_{j-1} \right\| \\ \text{diam}(S^{k+1}) &= \max_{1 \leq i, j \leq n+1} \left\| \frac{1}{2} \sigma_-(S^k) (e_{i-1} - e_{j-1}) \right\| \\ \text{diam}(S^{k+1}) &= \frac{1}{2} \sigma_-(S^k) \max_{1 \leq i, j \leq n+1} \|e_{i-1} - e_{j-1}\| = \frac{\sigma_-(S^k) \cdot \sqrt{2}}{2} = \frac{\sigma_-(S^k)}{\sqrt{2}}. \end{aligned}$$

Έτσι, μετά τη διάταξη των σημείων y_i , θα ισχύει $\sigma_+(S^{k+1}) \leq \sigma_-(S^k)$.

Όσον αφορά τον πίνακα V^{k+1} , πριν την αναδιάταξη, έχουμε $\kappa(V^{k+1}) = 1$. Μετά την αναδιάταξη, έχουμε δύο περιπτώσεις. Αν το καλύτερο σημείο παραμένει το x_1 , τότε ο πίνακας θα είναι διαγώνιος

με στοιχεία ± 1 . Αν όμως το καλύτερο στοιχείο αλλάξει, τότε, ύστερα από εναλλαγή γραμμών και πολλαπλασιάζοντας με $\frac{\pm \sigma - (S^k)}{2}$, ο πίνακας παίρνει την άνω τριγωνική μορφή

$$V^{k+1} = (V^{k+1})^{-1} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & -1 & 0 & \dots & 0 \\ \vdots & 0 & -1 & \ddots & \vdots \\ 0 & & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & -1 \end{pmatrix},$$

οπότε και για τον δείκτη κατάστασης θα ισχύει

$$\kappa(V^{k+1}) = \|V^{k+1}\|^2 \leq (1 + \sqrt{n})^2.$$

Σε κάθε περίπτωση, το στοιχείο simplex έχει καλή κατάσταση.

3 Αριθμητικά Παραδείγματα

Από τις μεθόδους που μελετήθηκαν στο δεύτερο μέρος της εργασίας, υλοποιήθηκαν στο λογισμικό Matlab οι εξής μέθοδοι:

- Μία μέθοδος Pattern Search.
- Μία μέθοδος συζυγών κλίσεων με τη στρατηγική 'discarding criterion'.
- Μία μέθοδος Nelder Mead με επανεκκίνηση της μεθόδου (oriented restarts).

Οι μέθοδοι χρησιμοποιήθηκαν για την ελαχιστοποίηση των παρακάτω συναρτήσεων:

1. Μία τετραγωνική συνάρτηση της μορφής

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c,$$

όπου

- A : τριδιαγώνιος με τα στοιχεία της κύριας διαγωνίου ίσα με 5, και τα στοιχεία των άλλων δύο διαγωνίων ίσα με -1 ,
- $b^T = [4 \ 3 \ 3 \ \dots \ 3 \ 3 \ 4]$, ώστε το ελάχιστο της συνάρτησης να είναι το $x_{min}^T = [1 \ 1 \ \dots \ 1]$, και
- c : μία τυχαία σταθερά.

2. Τη συνάρτηση

$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1).$$

Το ελάχιστο αυτής της συνάρτησης είναι $x_{min}^T = [0, 0, \dots, 0]$.

Τέλος, οι συναρτήσεις μελετήθηκαν σε διαστάσεις $n = \{20, 50, 100, 200\}$, ενώ ειδικά για τη μέθοδο συζυγών κλίσεων, μελετήθηκε και η περίπτωση $n = 10^4$. Τα διαγράμματα του Κεφαλαίου 3.2 απεικονίζουν τη διαφορά $f(x_k) - f(x_{min})$, και την ποσότητα $\|x_k - x_{min}\|_\infty$, για κάθε επανάληψη k . Επίσης, καταγράφεται το πλήθος των επαναλήψεων για κάθε μέθοδο, καθώς και ο χρόνος εκτέλεσής τους.

3.1 Παρουσίαση και περιγραφή των μεθόδων

Μέθοδος Pattern Search

Η μέθοδος Pattern Search είναι μία απλή υλοποίηση της κατηγορίας, με ορισμένα διακριτικά στοχαστικά στοιχεία. Το σύνολο κατευθύνσεων D είναι το σύνολο (2.9), το οποίο παραμένει ίδιο καθ' όλη τη διάρκεια της μεθόδου. Η συνάρτηση επαρκούς μείωσης είναι η $\rho(t) = 3t^{3/2}$, έχει δηλαδή μία συνήθη μορφή, όπως αναφέραμε στο Κεφάλαιο 2.3. Όσον αφορά το βήμα a_k , όπου και εμφανίζεται το στοχαστικό στοιχείο της μεθόδου, επιλέγεται αρχικά η τιμή $a_0 = 10$. Ύστερα από τρεις συνεχόμενες επιτυχημένες επαναλήψεις το βήμα αυξάνεται με τυχαίο τρόπο και επιτρέπεται να γίνει μέχρι και διπλάσιο του αρχικού, ενώ σε κάθε αποτυχημένη επανάληψη το βήμα μειώνεται επίσης με τυχαίο τρόπο. Οπότε, για το βήμα της επομένης επανάληψης a_{k+1} , ισχύει αντίστοιχα

$$a_k < a_{k+1} < 2a_k \quad \text{και} \quad 0,6 \cdot a_k < a_{k+1} < 0,9 \cdot a_k.$$

Η μέθοδος τερματίζει όταν $a_k < 10^{-6}$ ή έπειτα από 100.000 επαναλήψεις. Το δεύτερο κριτήριο ουσιαστικά εξαναγκάζει τη μέθοδο να τερματίσει, αντισταθμίζοντας έτσι την ασυμπτωτική σύγκλιση της μεθόδου. Παρακάτω ακολουθεί ο κώδικας της μεθόδου.

```
function [] = Pattern_Search(f,x)

tic;
ITER = 0;

XMIN = A\b;
FMIN = f(XMIN);
FGraph(1) = 0;
XGraph(1) = 0;

M = 3;
atol = 10^(-6);
a = 10;
n = length(x);
F = zeros(n,1);
D = [eye(n), -eye(n)];
k=0; %counter for step increase

while a > atol && ITER < 100000
    %BASIC ITERATION
    for j = 1:length(D)
        F(j) = f(x + a*D(:,j));
    end
    [fmin,i] = min(F);
    if f(x) - fmin > M*a^(3/2)
        x = x + a * D(:,i);
        if k >=3
            a = (1 + rand)*a;
        end
        k = k+1;
    else
        theta = 0.6 + 0.3*rand;
        a = a * theta;
        k = 0;
    end
    %end of 'BASIC ITERATION'
    FGraph(length(FGraph) + 1) = f(x) - FMIN;
    XGraph(length(XGraph) + 1) = norm(x - XMIN,Inf);
    ITER = ITER + 1;
end

%GRAPHS
FGraph(1) = [];
```

```

XGraph(1) = [];
subplot(1,2,1);
plot([1:length(FGraph)],FGraph);
subplot(1,2,2);
plot([1:length(XGraph)],XGraph)
%end of 'GRAPHS'

```

```

FGraph(length(FGraph))
XGraph(length(XGraph))
ITER
toc
end

```

Μέθοδος Συζυγών Κλίσεων

Η μέθοδος που υλοποιήθηκε βασίστηκε στα [7] και [6]. Ακολουθούμε τη διαδικασία του Powell που αναφέρθηκε στο Κεφάλαιο 2.4 για την κατασκευή των συζυγών κλίσεων, και εισάγουμε το discarding criterion, όπως περιγράφεται στις προαναφερθείσες πηγές, αλλά και στο ([1], σελ. 236-237). Η μέθοδος τερματίζει όταν η διαφορά στην τιμή της συνάρτησης μεταξύ των δύο τελευταίων επαναλήψεων γίνει μικρότερη από 10^{-6} . Παρακάτω ακολουθεί ο κώδικας της μεθόδου.

```

function [] = Conjugate_Directions(f,x)

tic;
ITER = 0;
diff = 1;

n = length(x);
Y = zeros(1,n);

PHI = @(a,x,p) (f(x+a*p));
XMIN = A\b;
FMIN = f(XMIN);
FGraph(1) = 0;
XGraph(1) = 0;
p = eye(n);

while (diff >=10^(-6))
    z = x;
    for j=1:n
        %line search: Quadratic interpolation
        phi(1) = PHI(-1,z,p(:,j));
        phi(2) = PHI(0,z,p(:,j));
        phi(3) = PHI(1,z,p(:,j));
        a = (phi(1) - phi(3))/(2*(phi(1) - 2*phi(2) + phi(3)));
        while PHI(a,z,p(:,j)) > PHI(0,z,p(:,j))
            a = a/2;

```

```

end
%end of 'line search: Quadratic interpolation'
Y(j) = f(z) - f(z + a*p(:,j));
z = z + a*p(:,j);
FGraph(length(FGraph) + 1) = f(z) - FMIN;
XGraph(length(XGraph) + 1) = norm(z - XMIN);
end
[y,m] = max(Y);
F(1) = f(x);
F(2) = f(z);
F(3) = f(2*z - x);
if (F(3) >= F(1)) || ((F(1) - 2 * F(2) + F(3))*(F(1) - F(2) - y)
^2 >= y*(F(1) - F(3))^2/2)
diff = f(x) - f(z);
x = z;
else
pp = z - x;
%line search: Quadratic interpolation
phi(1) = PHI(-1,x,pp);
phi(2) = PHI(0,x,pp);
phi(3) = PHI(1,x,pp);
aa = (phi(1) - phi(3))/(2*(phi(1) - 2*phi(2) + phi(3)));
while PHI(aa,x,pp) > PHI(0,x,pp)
aa = aa/2;
end
%end of 'line search: Quadratic interpolation'
p(:,m) = pp;
diff = f(x) - f(x + aa*pp);
x = x + aa*pp;
end

%Graph data
FGraph(length(FGraph) + 1) = f(x) - FMIN;
XGraph(length(XGraph) + 1) = norm(x - XMIN, Inf);
ITER = ITER + 1;
end
%GRAPHS
FGraph(1) = [];
XGraph(1) = [];
subplot(1,2,1);
plot([1:length(FGraph)],FGraph);
subplot(1,2,2);
plot([1:length(XGraph)],XGraph)
%end of 'GRAPHS'
FGraph(length(FGraph))
XGraph(length(XGraph))

```

```

ITER
toc
end

```

Μέθοδος Nelder-Mead

Εδώ ακολουθούμε τη μέθοδο όπως περιγράφεται στο ([1], σελ. 238-239), εμπλουτίζοντάς τη με τη διαδικασία επανεκκίνησης (oriented restart) του [10]. Για τα τέσσερα υποψήφια σημεία (ανάκλασης, επέκτασης, εσωτερικής και εξωτερικής συμπίεσης) κάθε επανάληψης, όπως και για τον συντελεστή σμίκρυνσης σ , έχουν χρησιμοποιηθεί οι συνήθεις τιμές. Το αρχικό simplex επιλέγεται τυχαία, με τις συντεταγμένες των σημείων που το αποτελούν να κυμαίνονται από -5 έως 5 . Η μέθοδος τερματίζει αν στην τελευταία επανάληψη η διαφορά στην τιμή της συνάρτησης μεταξύ του καλύτερου και του χειρότερου σημείου είναι μικρότερη από 10^{-6} , είτε ύστερα από 100.000 επαναλήψεις. Και εδώ το δεύτερο κριτήριο χρησιμοποιείται για να προληφθεί τυχόν ανεπιθύμητη συμπεριφορά της μεθόδου, δηλαδή το τέλος στο οποίο ενδέχεται να καταλήξει η μέθοδος σε ένα μη στάσιμο σημείο. Παρακάτω ακολουθεί ο κώδικας της μεθόδου.

```

function [] = Nelder_Mead(f,S)

tic;
ITER = 0;

XMIN = A\b;
FMIN = f(XMIN);
n = length(S)-1;
F = zeros(n+1,1);
FGraph(1) = 0;
XGraph(1) = 0;

%INITIALIZATION
for j=1:n+1
    F(j)=f(S(:,j));
end

%SORT
[F,I] = sort(F);
S = S(:,I);
%end of 'SORT'
V = S(:,2:n+1) - S(:,1);
delta = F(2:n+1) - F(1);
D = V \ delta;
sm = min(vecnorm(V));
sp = max(vecnorm(V));
a = 10^(-4) * sp / norm(D,2);
%end of 'INITIALIZATION'

%BASIC ITERATION

```

```

while F(n+1) - F(1) > 10(-6) && ITER < 100000

    fbar = sum(F)/(n+1);
    xbar = sum(S(:,1:n),2)/n;
    x = @(t) (xbar + t*(S(:,n+1) - xbar));
    xr = x(-1);
    fr = f(xr);
    if fr < F(1) %Expansion
        xe = x(-2);
        fe = f(xe);
        if fe < fr
            S(:,n+1) = xe;
            F(n+1) = fe;
        else
            S(:,n+1) = xr;
            F(n+1) = fr;
        end
    elseif (F(1) <= fr) && ( fr < F(n)) %Reflection
        S(:,n+1) = xr;
        F(n+1) = fr;
    elseif (F(n) <= fr) && ( fr < F(n+1)) %Out. Contraction
        xoc = x(-1/2);
        foc = f(xoc);
        if foc <= fr
            S(:,n+1) = xoc;
            F(n+1) = foc;
        else %Shrink
            S = (S + S(:,1))/2;
            for j=1:n+1
                F(j)=f(S(:,j));
            end
        end
    end
    else % In. Contraction
        xic = x(1/2);
        fic = f(xic);
        if fic < F(n+1)
            S(:,n+1) = xic;
            F(n+1) = fic;
        else %Shrink
            S = (S + S(:,1))/2;
            for j=1:n+1
                F(j)=f(S(:,j));
            end
        end
    end
end
end

```

```

%SORTING THE ELEMENTS
[F,I] = sort(F);
S = S(:,I);
%end of 'SORTING THE ELEMENTS'
Fdiff = sum(F)/(n+1) - fbar;
%ORIENTED RESTART
if (0 > Fdiff) && ( Fdiff >= -a * norm(D,2)^2)
    b = sm/2 * ones(1,n);
    for j=1:n
        if D(j) < 0
            b(j) = -b(j);
        end
    end
    S(:,2:n+1) = S(:,1) + diag(b);
    %SORTING THE ELEMENTS
    [F,I] = sort(F);
    S = S(:,I);
    %end of 'SORTING THE ELEMENTS'
end
%end of 'ORIENTED RESTART'

%UPDATE STEP
V = S(:,2:n+1) - S(:,1);
delta = F(2:n+1) - F(1);
D = V \ delta;
sm = min(vecnorm(V));
%end of 'UPDATE STEP'
%end of 'BASIC ITERATION'
FGraph(length(FGraph) + 1) = F(1) - FMIN;
XGraph(length(XGraph) + 1) = norm(S(:,1) - XMIND);

ITER = ITER + 1;
end
%GRAPHS
FGraph(1) = [];
XGraph(1) = [];
subplot(1,2,1);
plot([1:length(FGraph)],FGraph);
subplot(1,2,2);
plot([1:length(XGraph)],XGraph)
%end of 'GRAPHS'

FGraph(length(FGraph))
XGraph(length(XGraph))
ITER

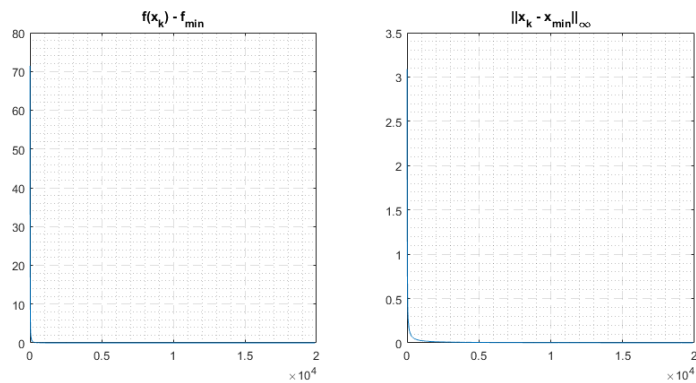
```

```
toc
end
```

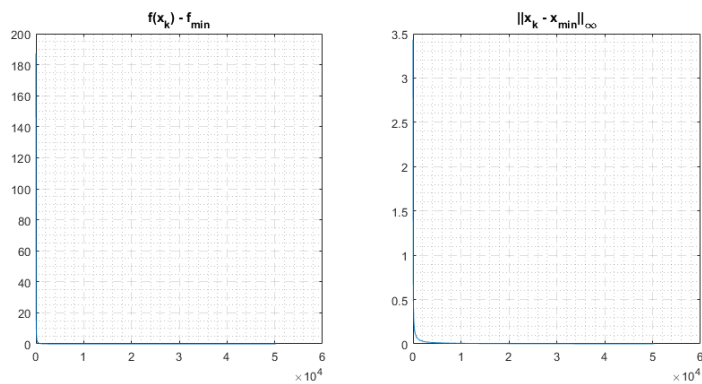
3.2 Ανάλυση των αποτελεσμάτων

Μέθοδος Pattern Search

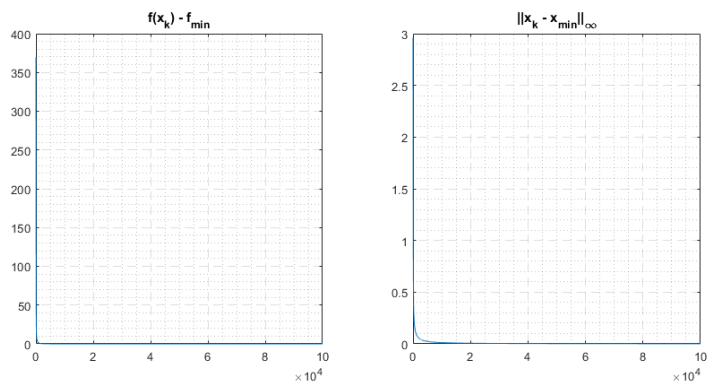
Η μέθοδος Pattern Search εκτελέστηκε για την ελαχιστοποίηση και των δύο συναρτήσεων που αναφέρθηκαν στην αρχή του κεφαλαίου σε διάσταση από $n = 20$ έως και $n = 200$ μεταβλητές. Παρακάτω παρουσιάζουμε τα αντίστοιχα διαγράμματα και τον πίνακα που αναφέρθηκε στην εισαγωγή του Κεφαλαίου.



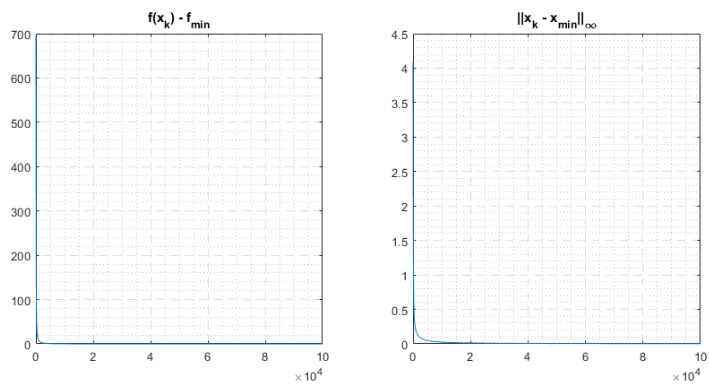
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 20$$



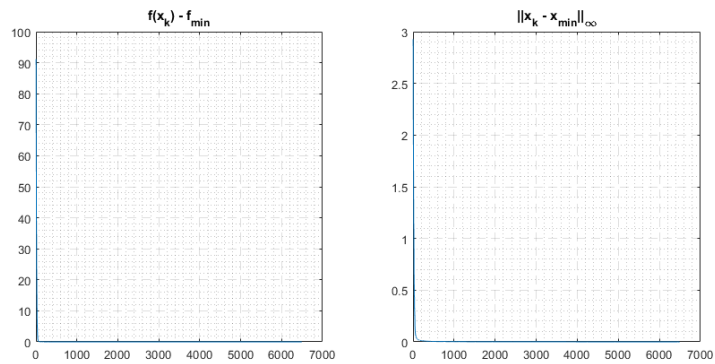
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 50$$



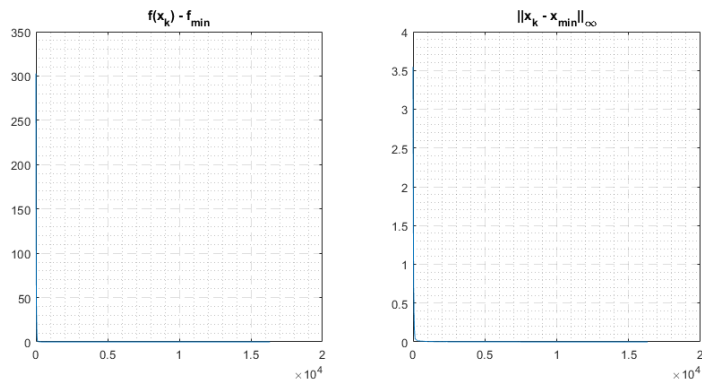
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 100$$



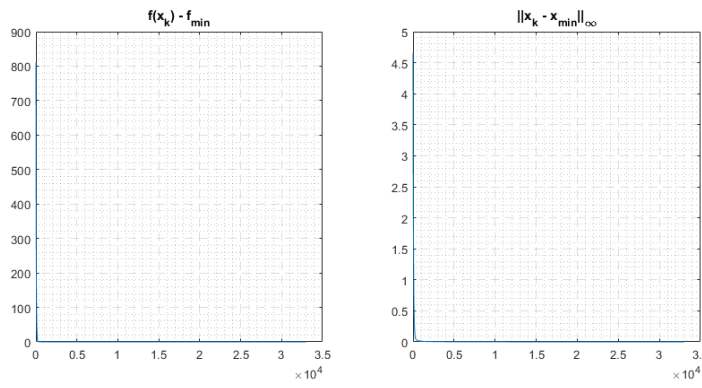
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 200$$



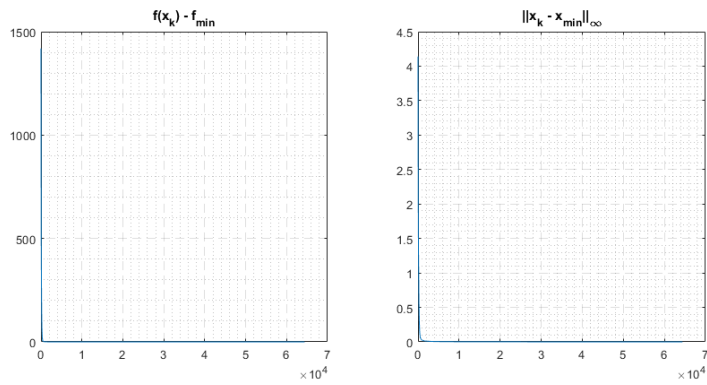
$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1), \quad n = 20$$



$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1), \quad n = 50$$



$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1), \quad n = 100$$



$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1), \quad n = 200$$

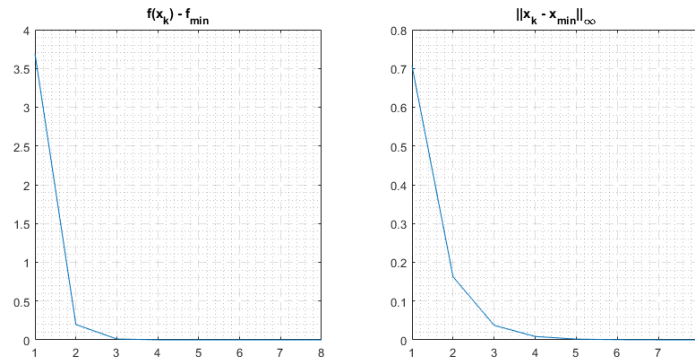
Συνάρτηση	n	Επαναλήψεις	Χρόνος (sec)	$f(x_k) - f_{min}$	$\ x_k - x_{min}\ _{\infty}$
$f(x)$	20	19.888	0,968	$3,73 \cdot 10^{-5}$	0.0013
	50	50.233	5,97	$9,29 \cdot 10^{-5}$	0,0013
	100	10^5 (μέγιστο)	29,19	$1,84 \cdot 10^{-4}$	0,0013
	200	10^5 (μέγιστο)	82,3	0,0018	0,0027
$g(x)$	20	6.491	0,91	$1,21 \cdot 10^{-5}$	$3,18 \cdot 10^{-4}$
	50	16.327	10,99	$2,96 \cdot 10^{-5}$	$3,14 \cdot 10^{-4}$
	100	33.003	84,99	$5,69 \cdot 10^{-5}$	$3,09 \cdot 10^{-4}$
	200	64.394	653,75	$1,23 \cdot 10^{-4}$	$3,20 \cdot 10^{-4}$

Πίνακας 1

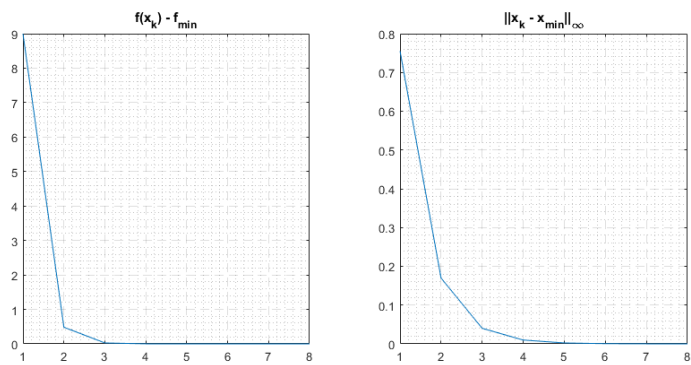
Σε όλες τις περιπτώσεις η ασυμπτωτική σύγκλιση της μεθόδου είναι οφθαλμοφανής. Λόγω αυτής της συμπεριφοράς, στην περίπτωση της τετραγωνικής συνάρτησης, για διάσταση $n = 200$, παρατηρούμε πώς η συνάρτηση δεν συγκλίνει σε μία ικανοποιητική προσέγγιση του ελαχίστου προτού φτάσει το φράγμα των 100.000 επαναλήψεων.

Μέθοδος Συζυγών κλίσεων

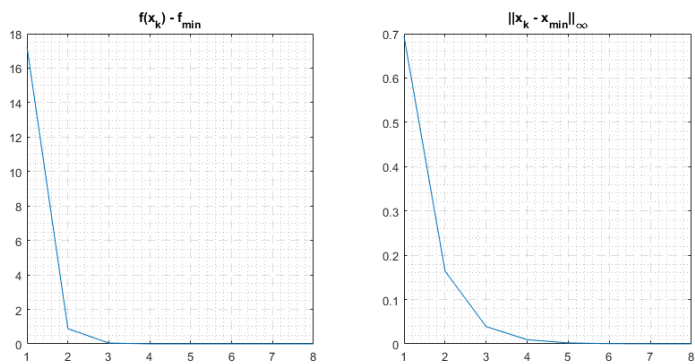
Η μέθοδος Συζυγών Κλίσεων εκτελέστηκε για την ελαχιστοποίηση μόνο της τετραγωνικής συνάρτησης, σε διάσταση από $n = 20$ έως και $n = 200$, αλλά ακόμα και για διάσταση $n = 10^4$. Παρακάτω παρουσιάζουμε τα αντίστοιχα διαγράμματα και τον πίνακα που αναφέρθηκε στην εισαγωγή του Κεφαλαίου.



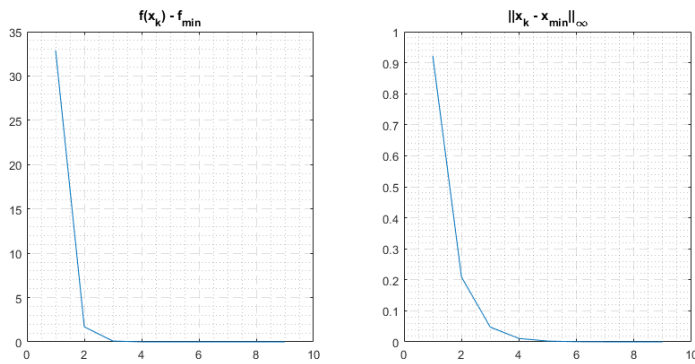
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 20$$



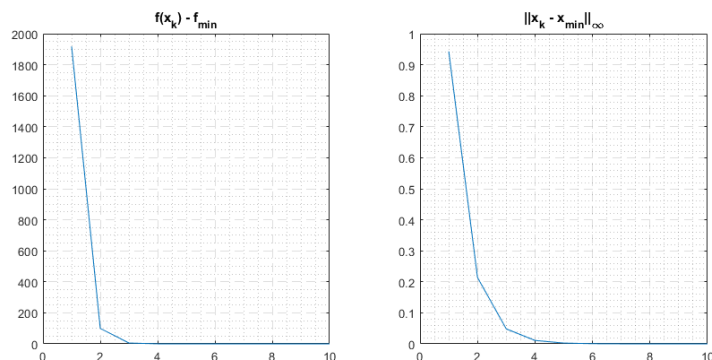
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 50$$



$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 100$$



$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 200$$



$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 10^4$$

Συνάρτηση	n	Επαναλήψεις	Χρόνος (sec)	$f(x_k) - f_{min}$	$\ x_k - x_{min}\ _{\infty}$
$f(x)$	20	8	0,08	$9,25 \cdot 10^{-9}$	$2,90 \cdot 10^{-5}$
	50	8	0,08	$2,1 \cdot 10^{-8}$	$3,37 \cdot 10^{-5}$
	100	8	0,09	$3,23 \cdot 10^{-8}$	$3,14 \cdot 10^{-5}$
	200	9	0,11	$4,29 \cdot 10^{-9}$	$8,90 \cdot 10^{-6}$
	10^4	10	100,13	$1,65 \cdot 10^{-8}$	$2,29 \cdot 10^{-6}$

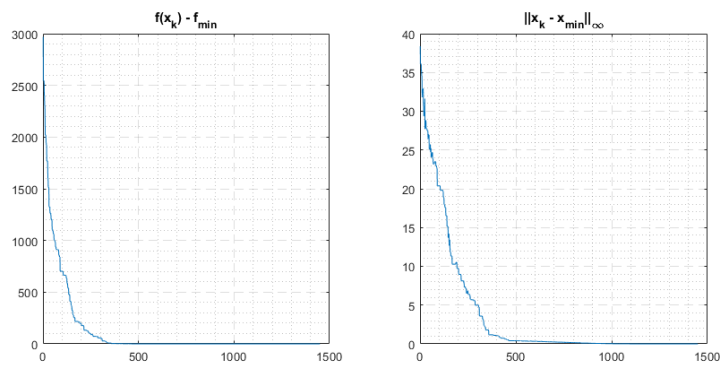
Πίνακας 2

Παρατηρούμε ότι η μέθοδος έχει την πιο ταχεία και ακριβή σύγκλιση σε σχέση με τις υπόλοιπες. Μάλιστα, η σύγκλιση της μεθόδου φαίνεται να είναι ανεξάρτητη από τη διάσταση του προβλήματος, καθώς σε κάθε περίπτωση χρειάστηκε περίπου το ίδιο πλήθος επαναλήψεων. Τέλος, ο χρόνος σύγκλισης της μεθόδου για τη διάσταση $n = 10^4$, είναι συγκρίσιμος με το χρόνο για τη σύγκλιση της μεθόδου Pattern Search για διάσταση $n = 200$. Όλα τα παραπάνω είναι αναμενόμενα, καθώς η μέθοδος συζυγών κλίσεων από την κατασκευή της είναι προορισμένη για τέτοιου είδους προβλήματα.

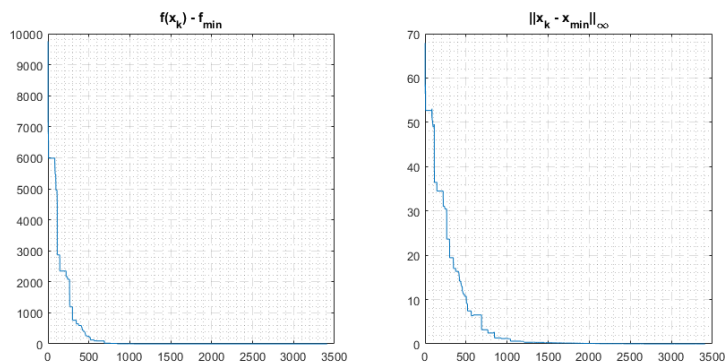
Μέθοδος Nelder-Mead

Η μέθοδος Nelder-Mead εκτελέστηκε για την ελαχιστοποίηση και των δύο συναρτήσεων που αναφέρθηκαν στην εισαγωγή του κεφαλαίου, για διάσταση από $n = 20$ έως και $n = 200$ μεταβλητές. Παρακάτω παρουσιάζουμε τα αντίστοιχα διαγράμματα και τον πίνακα που αναφέρθηκε στην εισαγωγή του Κεφαλαίου.

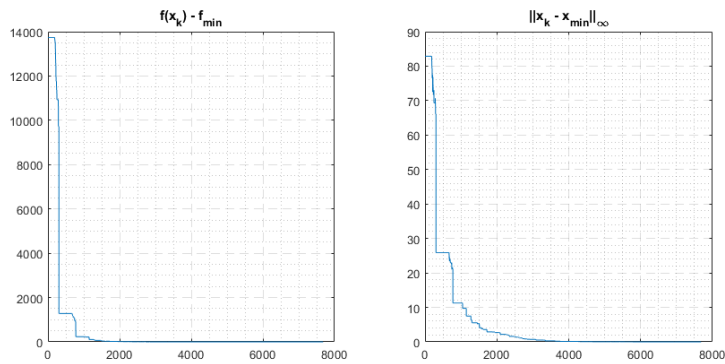
Εδώ αξίζει να σημειωθεί πως ακόμα και στη μικρότερη διάσταση, και για τις δύο συναρτήσεις, χρειάστηκε να κάνουμε αρκετές προσπάθειες για να εκμαιεύσουμε τα παραπάνω αποτελέσματα. Αυτό συνέβη λόγω της τάσης της μεθόδου να τελματώνει (ακόμα και με τον μηχανισμό επανεκκίνησης), σε συνδυασμό με την τυχαία επιλογή του αρχικού simplex. Ωστόσο, όταν η μέθοδος κατέληξε σε ένα καλό αποτέλεσμα, αυτό συνέβη ταχύτερα από ότι στη μέθοδο Pattern Search (ακόμα και αν χρειάστηκαν περισσότερες επαναλήψεις), και μάλιστα με την ίδια ή μεγαλύτερη ακρίβεια.



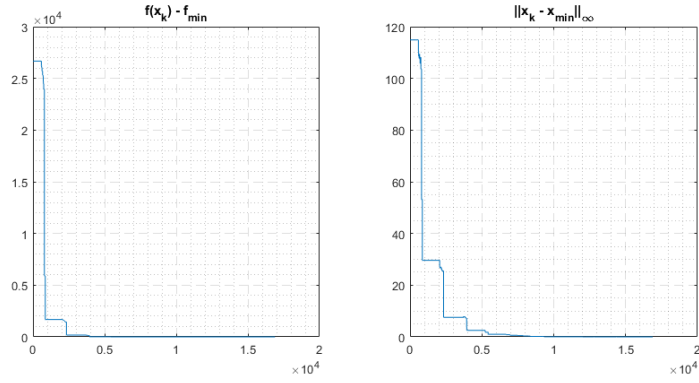
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 20$$



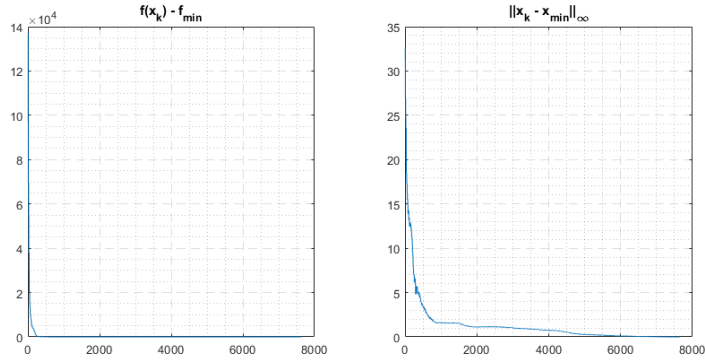
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 50$$



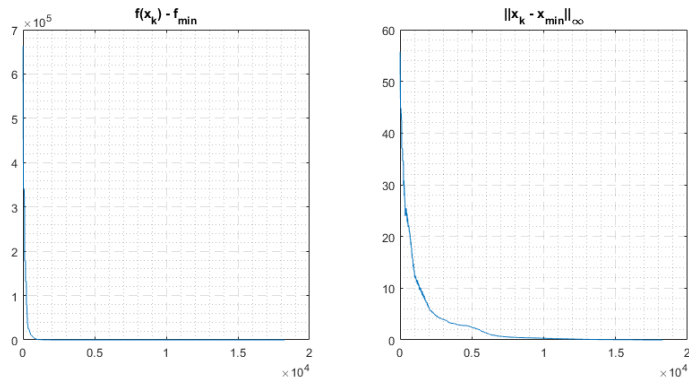
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 100$$



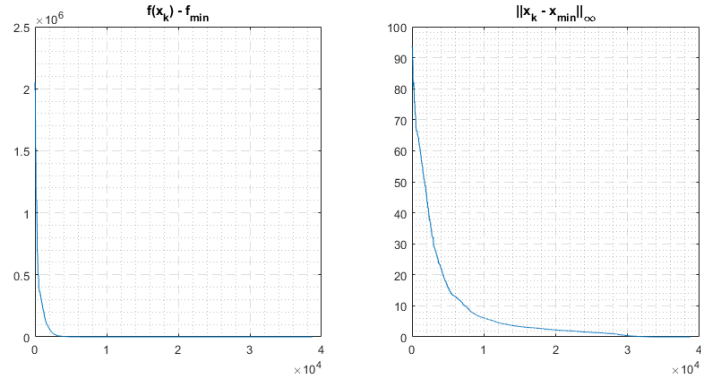
$$f(x) = \frac{1}{2}x^T Ax - b^T x + c, \quad n = 200$$



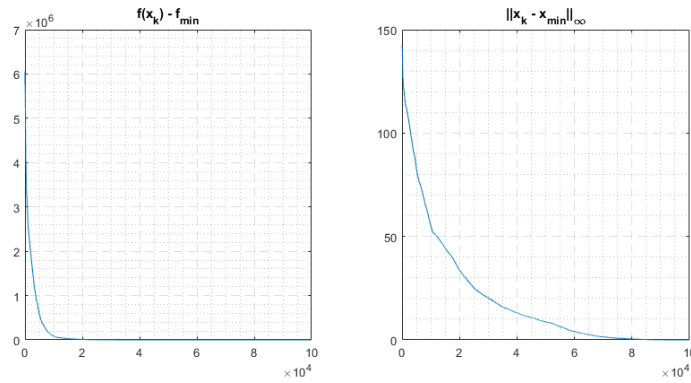
$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1), \quad n = 20$$



$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1), \quad n = 50$$



$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1), \quad n = 100$$



$$g(x) = \sum_{i=1}^n (x_i - 1)^4 - \sum_{i=1}^n (4x_i - 1), \quad n = 200$$

Συνάρτηση	n	Επαναλήψεις	Χρόνος (sec)	$f(x_k) - f_{min}$	$\ x_k - x_{min}\ _{\infty}$
$f(x)$	20	1.449	0,14	$6,6 \cdot 10^{-7}$	$5,47 \cdot 10^{-4}$
	50	3.409	0,37	$1,41 \cdot 10^{-6}$	$8,53 \cdot 10^{-4}$
	100	7.678	1,57	$1,60 \cdot 10^{-6}$	$9,12 \cdot 10^{-4}$
	200	16.882	14,2	$9,27 \cdot 10^{-7}$	$6,53 \cdot 10^{-4}$
$g(x)$	20	7.638	0,34	$3,27 \cdot 10^{-6}$	$7,39 \cdot 10^{-4}$
	50	18.268	1,35	$2,55 \cdot 10^{-6}$	$6,52 \cdot 10^{-4}$
	100	38.757	8,16	$2,24 \cdot 10^{-6}$	$6,11 \cdot 10^{-4}$
	200	10^5 (μέγιστο)	100,3	$3,39 \cdot 10^{-4}$	0,0075

Πίνακας 3

Αναφορές

- [1] J. Nocedal S.J. Wright. *Numerical optimization*. 2006.
- [2] Michael JD Powell. Direct search algorithms for optimization calculations. *Acta numerica*, 7:287–336, 1998.
- [3] Michael JD Powell. On trust region methods for unconstrained minimization without derivatives. *Mathematical programming*, 97:605–623, 2003.
- [4] Michael JD Powell. Least frobenius norm updating of quadratic models that satisfy interpolation conditions. *Mathematical Programming*, 100:183–215, 2004.
- [5] Tamara G Kolda, Robert Michael Lewis, and Virginia Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM review*, 45(3):385–482, 2003.
- [6] Michael JD Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The computer journal*, 7(2):155–162, 1964.
- [7] R Brent. A new algorithm for minimizing a function of several variables without calculating derivatives. *Algorithms for minimization without derivatives*, pages 124–137, 1973.
- [8] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- [9] Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.
- [10] Carl Tim Kelley. Detection and remediation of stagnation in the nelder–mead algorithm using a sufficient decrease condition. *SIAM journal on optimization*, 10(1):43–55, 1999.

