Αλγοριθμικές Μέθοδοι Βελτιστοποίησης με Έμφαση σε Κατανεμημένα Προβλήματα

Βασικές Τεχνικές Κατανεμημένης Βελτιστοποίησης

Δημήτρης Αμπελιώτης

Επίκουρος Καθηγητής, Ιόνιο Πανεπιστήμιο

Περιεχόμενα

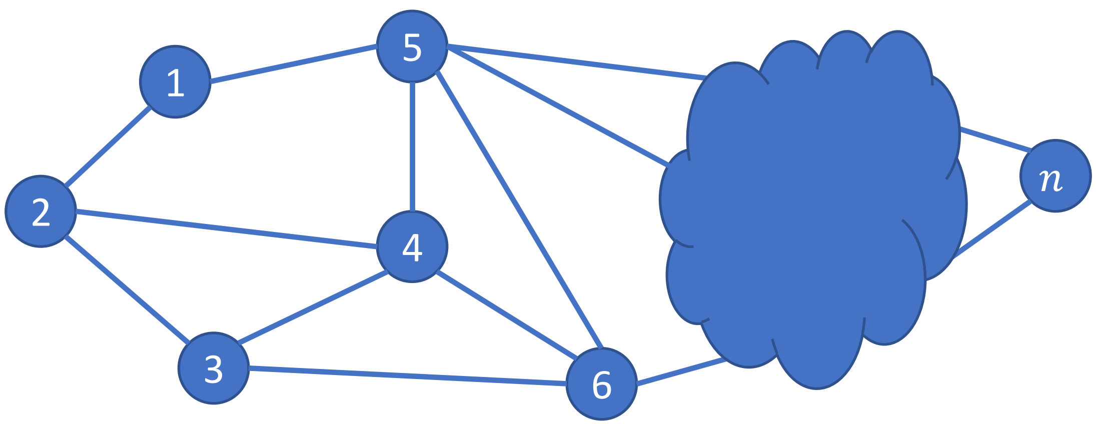
- Εισαγωγή
- Κατανεμημένη Βελτιστοποίηση
  - Αναπαράσταση γραφημάτων με πίνακα
  - Χαλάρωση των περιορισμών συναίνεσης
  - Παράδειγμα
- Κατανεμημένος Αλγόριθμος Κλίσεων
- Αλγόριθμος Δυικής Ανόδου (Dual Ascent)

Εισαγωγή

(Αντιστοιχεί στην Παράγραφο 5.1 στο “Cooperative and Graph Signal Processing: Principles and Applications”)

Εισαγωγή

- Κατανεμημένη βελτιστοποίηση πολλαπλών agents
  - Κάθε agent (κόμβος/node) έχει μια **τοπική** (local) συνάρτηση κόστους
  - Όλοι οι agents έχουν **κοινό** σκοπό την ελαχιστοποίηση του αθροίσματος (ή ισοδύναμα, του μέσου όρου) των συναρτήσεων κόστους
  - Για τον κοινό αυτό σκοπό, οι agents μπορούν να επικοινωνούν χρησιμοποιώντας ένα **δίκτυο** μέσω του οποίου διασυνδέονται και το οποίο περιορίζει την επικοινωνία μόνο ανάμεσα σε **γειτονικούς** κόμβους



Εισαγωγή

- Τρεις κατηγορίες εφαρμογών:
  a) Οι κατανεμημένες τεχνικές βελτιστοποίησης ταιριάζουν σε εφαρμογές όπου **η συλλογή των δεδομένων γίνεται εγγενώς με κατανεμημένο τρόπο**, στους κόμβους ενός δικτύου. Για παράδειγμα:
     - Σε δίκτυα αισθητήρων
     - Σε κατανεμημένα συστήματα ελέγχου, π.χ. στην οργάνωση ενός σμήνους από drones
  b) Σε εφαρμογές **βέλτιστης κατανομής πόρων** (resource allocation) σε (ενσύρματα ή ασύρματα) δίκτυα. Για παράδειγμα:
     - Εύρεση μιας συντομότερης διαδρομής για δρομολόγηση
  c) Σε εφαρμογές μηχανικής μάθησης/εξόρυξης πληροφορίας από **μεγάλο όγκο δεδομένων**

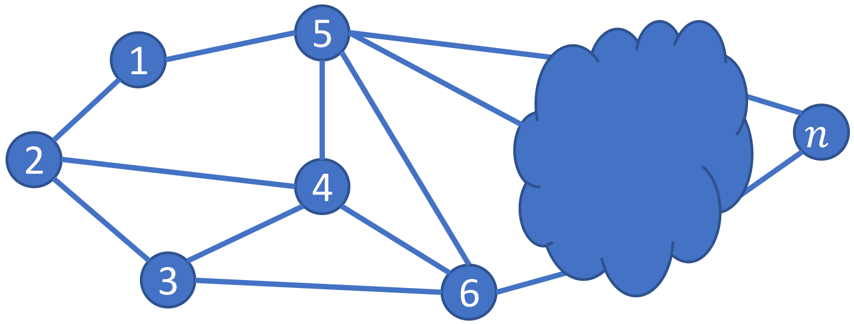
Εισαγωγή

- Οι τεχνικές κατανεμημένης βελτιστοποίησης ακολουθούν, γενικά, μια κοινή μεθοδολογία:
  1. Το αρχικό πρόβλημα βελτιστοποίησης «αποσυνδέεται» σε ανεξάρτητα προβλήματα, επιτρέποντας σε κάθε κόμβο να **διατηρεί ένα δικό του, τοπικό, αντίγραφο** των μεταβλητών ως προς τις οποίες γίνεται η βελτιστοποίηση (από ένα πρόβλημα βελτιστοποίησης p μεταβλητών πάμε σε ένα πρόβλημα βελτιστοποίησης $n \cdot p$ μεταβλητών)
  2. Το πρόβλημα «επανασυνδέεται», εισάγοντας **περιορισμούς ισότητας** των τοπικών μεταβλητών που διατηρούν γειτονικοί κόμβοι (consensus constraints). Αν το δίκτυο/γράφημα είναι συνεκτικό, οι περιορισμοί ικανοποιούνται όταν υπάρχει καθολική συμφωνία
  3. Η τροποποιημένη μορφή του προβλήματος επιτρέπει τον υπολογισμό κατευθύνσεων μείωσης του κόστους (descend directions) με κατανεμημένο τρόπο

Κατανεμημένη Βελτιστοποίηση

(Αντιστοιχεί στην Παράγραφο 5.2 στο “Cooperative and Graph Signal Processing: Principles and Applications”)

Κατανεμημένη Βελτιστοποίηση



- Θεωρούμε n agents
- Κάθε agent έχει την τοπική συνάρτηση κόστους

$$f_i\colon \mathbb{R}^p \to \mathbb{R}$$

- Οι agents ενδιαφέρονται να ελαχιστοποιήσουν τη συνολική συνάρτηση κόστους

$$f(x) = \sum_{i=1}^{n} f_i(x), \qquad x \in \mathbb{R}^p$$

- Το πρόβλημα αυτό είναι «συνδεμένο» καθώς οι μεταβλητές κάθε τοπικής συνάρτησης κόστους είναι κοινές

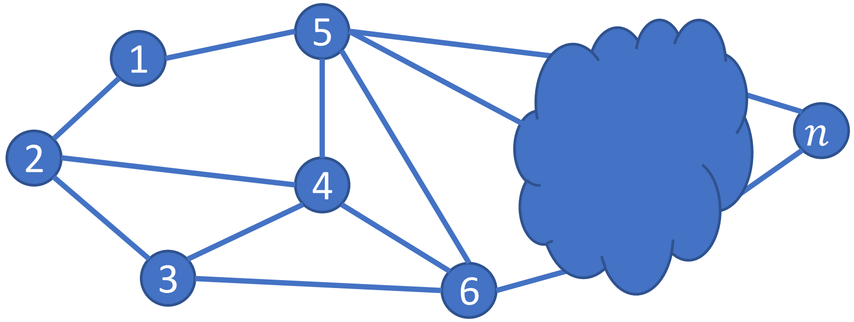
Κατανεμημένη Βελτιστοποίηση

- Οι κόμβοι/agents συνεργάζονται για να λύσουν το πρόβλημα βελτιστοποίησης

$$X^* = \operatorname*{argmin}_{x \in \mathbb{R}^p} f(x) = \operatorname*{argmin}_{x \in \mathbb{R}^p} \sum_{i=1}^{n} f_i(x)$$

- Εδώ, με X^* συμβολίζουμε το *σύνολο* των σημείων στα οποία η συνολική συνάρτηση ελαχιστοποιείται (global minimizers)
- Υποθέτουμε πως το σύνολο X^* δεν είναι κενό και πως επίσης το σύνολο των βέλτιστων τιμών της συνολικής συνάρτησης κόστους δεν είναι κενό
- Η συνολική συνάρτηση είναι κυρτή, ως άθροισμα από κυρτές συναρτήσεις κόστους

Κατανεμημένη Βελτιστοποίηση



- Θεωρούμε το γράφημα $G = (V, E)$ όπου

$$V = \{1, 2, 3, \dots, n\}$$

- Επίσης, το σύνολο των ακμών περιέχει $2m$ στοιχεία, όπου m το πλήθος των ακμών του γραφήματος

agent i can communicate with agent j $\Rightarrow (i, j) \in E$

- Υποθέτουμε μη-κατευθυνόμενα γραφήματα

$$\forall\, (i, j) \in E \Rightarrow (j, i) \in E$$

- Ορίζουμε το σύνολο των γειτονικών κόμβων του κόμβου i

$$n(i) = \{j : (i, j) \in E\}$$

Κατανεμημένη Βελτιστοποίηση

Διασπάμε το πρόβλημα σε ανεξάρτητα προβλήματα

- Εισάγουμε τοπικά «αντίγραφα» των υπό-βελτιστοποίηση μεταβλητών x
- Τα τοπικά αντίγραφα τα συμβολίζουμε ως $x_i \in \mathbb{R}^p$
- Επίσης, ορίζουμε ένα διάνυσμα με όλα τα αντίγραφα των μεταβλητών, σε όλους τους κόμβους του δικτύου, το οποίο συμβολίζουμε ως

$$\boldsymbol{x} = [x_1^T \quad x_2^T \quad \cdots \quad x_n^T]^T \in \mathbb{R}^{n \cdot p}$$

- Με τη βοήθεια αυτού του επεκταμένου διανύσματος μεταβλητών ορίζουμε το «αποσυνδεμένο» πρόβλημα βελτιστοποίησης

$$\widetilde{\boldsymbol{x}} \in \underset{\boldsymbol{x} \in \mathbb{R}^{n \cdot p}}{\operatorname{argmin}} F(x) = \underset{\boldsymbol{x} \in \mathbb{R}^{n \cdot p}}{\operatorname{argmin}} \sum_{i=1}^{n} f_i(x_i)$$

- Εύκολα διαπιστώνουμε πως το πρόβλημα αυτό είναι πολύ διαφορετικό από το αρχικό πρόβλημα. Για να λύσουμε το «αποσυνδεμένο» πρόβλημα, κάθε κόμβος ανεξάρτητα αναζητά το ελάχιστο της δικής του συνάρτησης κόστους

Κατανεμημένη Βελτιστοποίηση

- Μπορούμε να «διορθώσουμε» το τελευταίο πρόβλημα ώστε να γίνει ισοδύναμο με το αρχικό πρόβλημα εισάγοντας κατάλληλους περιορισμούς

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\mathbb{R}^{n\cdot p}}{\operatorname{argmin}} \sum_{i=1}^{n} f_i(x_i) \quad subject\ to\ x_i = x_j, \forall\, i = 1,2,\dots,n \quad and \ \ j \in n(i)$$

- Το πρόβλημα αυτό (ας θεωρήσουμε πως μπορούμε να το λύσουμε) είναι ισοδύναμο με το αρχικό πρόβλημα, καθώς εύκολα διαπιστώνουμε πως κάθε βέλτιστη λύση του ενός είναι και βέλτιστη λύση για το άλλο, αρκεί το γράφημα που διασυνδέει τους κόμβους να είναι συνεκτικό.

Κατανεμημένη Βελτιστοποίηση

Παράδειγμα: Κατανεμημένο Πρόβλημα Γραμμικών Ελ. Τετραγώνων

- Υποθέτουμε πως κάθε ένας από τους n κόμβους ενός δικτύου έχει στη διάθεσή του
  - Ένα διάνυσμα $h_i \in \mathbb{R}^p$ (π.χ. είσοδος σε ένα γραμμικό φίλτρο)
  - Μια βαθμωτή τιμή $y_i \in \mathbb{R}$ (π.χ. έξοδος του φίλτρου)
- Υποθέτουμε πως οι μετρήσεις (h_i και y_i) κάθε κόμβου συνδέονται από ένα γραμμικό μοντέλο της μορφής

$$y_i = h_i^T x + w_i$$

  - όπου $x \in \mathbb{R}^p$ είναι ένα άγνωστο διάνυσμα παραμέτρων (κοινό για όλους τους κόμβους) και το w_i συμβολίζει τιμές λευκού (χωρικά) θορύβου
- Κάθε κόμβος ενδιαφέρεται να ελαχιστοποιήσει το τοπικό του σφάλμα, επομένως επιλέγει την τοπική συνάρτηση κόστους

$$f_i(x) = \left(y_i - h_i^T x\right)^2$$

- Έτσι, το (κεντρικοποιημένο) πρόβλημα γράφεται:

$$\min_{x \in \mathbb{R}^p} \sum_{i=1}^{n} \left(y_i - h_i^T x\right)^2 = \|y - Hx\|_2^2$$

Κατανεμημένη Βελτιστοποίηση

Παράδειγμα (συνέχεια)

- Έτσι, το (κεντρικοποιημένο) πρόβλημα γράφεται:

$$\min_{x \in \mathbb{R}^p} \sum_{i=1}^{n} \left(y_i - h_i^T x\right)^2 = \|y - Hx\|_2^2$$

  - όπου έχουμε «στοιβάξει» τις επιμέρους τιμές y_i στο διάνυσμα $y \in \mathbb{R}^n$ και με το σύμβολο H έχουμε συμβολίσει έναν $n \times p$ πίνακα του οποίου η γραμμή i είναι ίση με h_i^T

- Αν θεωρήσουμε το «αποσυνδεμένο» πρόβλημα που αντιστοιχεί στο παραπάνω πρόβλημα, τότε κάθε κόμβος ανεξάρτητα αναζητά ένα διάνυσμα $x_i \in \mathbb{R}^p$ (τοπικό αντίγραφο της κοινής μεταβλητής x) λύνοντας το τοπικό πρόβλημα

$$\min_{x_i \in \mathbb{R}^p} \left(y_i - h_i^T x_i\right)^2$$

  - Παρατηρούμε πως για $p > 1$ όλα τα τοπικά προβλήματα είναι υπό-ορισμένα (έχουν άπειρο πλήθος λύσεων)
  - Λύση έχω με εισαγωγή περιορισμών, $n \geq p$ και γραμ. ανεξ. στήλες για τον H

Αναπαράσταση Γραφημάτων με Πίνακα

(Αντιστοιχεί στην Παράγραφο 5.2.1 στο "Cooperative and Graph Signal Processing: Principles and Applications")

Αναπαράσταση Γραφημάτων με Πίνακα

- Στην ενότητα αυτή, θα δούμε πως μπορούμε να εκφράσουμε τους περιορισμούς συναίνεσης (consensus constraints), που περιγράψαμε στα προηγούμενα, χρησιμοποιώντας αναπαραστάσεις γραφημάτων με πίνακες
- Έστω πως σε κάθε ακμή $e = (i, j) \in E$ αντιστοιχούμε ένα μη αρνητικό βάρος w_{ij}
- Επίσης, θεωρούμε πως τα βάρη αυτά είναι συμμετρικά, δηλαδή $w_{ij} = w_{ji}$
- Ακόμα, θεωρούμε πως, για κάθε κόμβο i, τα βάρη των ακμών που τον συνδέουν με τους γείτονές του έχουν κανονικοποιηθεί ώστε να έχουν άθροισμα μονάδα:

$$\forall\, i, \sum_{j:(i,j)\in E} w_{ij} = 1$$

- Θεωρούμε πως τα βάρη αυτά είναι στοιχεία ενός πίνακα W. Εύκολα διαπιστώνουμε πως
  - Ο πίνακας W είναι συμμετρικός
  - Οι γραμμές και οι στήλες του πίνακα W έχουν άθροισμα 1 (doubly stochastic)

Αναπαράσταση Γραφημάτων με Πίνακα

Βεβαρημένος Πίνακας Διασύνδεσης Ακμών-Κόμβων

- Για να αναπαραστήσουμε τις ακμές του γραφήματος, ορίζουμε έναν πίνακα B με διαστάσεις $m \times n$, όπου κάθε γραμμή αντιστοιχεί σε μια ακμή και κάθε στήλη σε έναν κόμβο, ως

$$B_{e,v} = B_{(i,j),v} = \begin{cases} \sqrt{w_{ij}}, & \alpha\nu\ v = i \\ -\sqrt{w_{ij}}, & \alpha\nu\ v = j \\ 0, & \delta\iota\alpha\varphi o\rho\varepsilon\tau\iota\kappa\alpha \end{cases}$$

- Θεωρώ μόνο τις ακμές (i, j) με $i < j$
- Παρατήρηση: Ο πίνακας αυτός είναι αραιός, κάθε γραμμή του έχει μόνο δύο μη-μηδενικά στοιχεία (από τα συνολικά n). Επί της ουσίας, για κάθε ακμή, έχω ένα θετικό βάρος στη στήλη-κόμβο αφετηρία, και ένα αρνητικό βάρος στη στήλη-κόμβο προορισμό
- Χρησιμοποιώντας τον πίνακα B, μπορούμε να κατασκευάσουμε το λεγόμενο Laplacian πίνακα του γραφήματος, ως

$$L = B^T B \in \mathbb{R}^{\mathrm{n} \times n}$$

Αναπαράσταση Γραφημάτων με Πίνακα

- Χρησιμοποιώντας τον πίνακα B, μπορούμε να κατασκευάσουμε το λεγόμενο Laplacian πίνακα του γραφήματος, ως
$$L = B^T B \in \mathbb{R}^{\mathrm{n}\times n}$$
- Εύκολα μπορούμε να δούμε πως
  - Τα διαγώνια στοιχεία του πίνακα L είναι $L_{ii} = 1$
  - Τα μη-διαγώνια στοιχεία είναι $L_{ij} = -w_{ij}$ όταν $(i, j) \in E$ ενώ $L_{ij} = 0$ όταν $(i, j) \notin E$
  - Ο Laplacian πίνακας είναι συμμετρικός
- Στις τεχνικές κατανεμημένης βελτιστοποίησης που εξετάζουν ένα γράφημα/δίκτυο με Laplacian πίνακα L, οι φασματικές ιδιότητες του πίνακα (δηλαδή, οι ιδιοτιμές του) είναι σημαντικές
  - $L \cdot \underline{1} = \underline{0}$ Εύκολα το επαληθεύουμε, καθώς, για κάθε γραμμή i, έχουμε το διαγώνιο στοιχείο L_{ii} που είναι μονάδα και το άθροισμα όλων των $-w_{ij}$ που θα είναι -1
  - $L \cdot \underline{1} = 0 \cdot \underline{1}$ Σημαίνει πως το μηδέν αποτελεί ιδιοτιμή του πίνακα L και αντιστοιχεί στο ιδιοδιάνυσμα $\underline{1}$

Αναπαράσταση Γραφημάτων με Πίνακα

- Αφού βρήκαμε μια μηδενική ιδιοτιμή, καταλαβαίνουμε πως ο πίνακας L δεν είναι αντιστρέψιμος, δηλαδή δεν έχει πλήρη τάξη
- Μπορεί να αποδειχθεί πως, αν το γράφημα μας είναι συνεκτικό, όλες οι άλλες ιδιοτιμές είναι μη-μηδενικές
- Θα συμβολίζουμε την ελάχιστη, μη-μηδενική, ιδιοτιμή του Laplacian πίνακα ως λ και τη μέγιστη με Λ
- Μπορεί να δειχθεί πως $\Lambda \leq 2$
- Ορίζουμε επίσης το «δείκτη κατάστασης» του γραφήματος ως το λόγο

$$\rho = \frac{\Lambda}{\lambda}$$

- Μπορούμε να αλλάξουμε το δείκτη κατάστασης ενός γραφήματος αλλάζοντας τα βάρη των ακμών

Χαλάρωση των Περιορισμών Συναίνεσης

(Αντιστοιχεί στην Παράγραφο 5.2.2 στο “Cooperative and Graph Signal Processing: Principles and Applications”)

Χαλάρωση των Περιορισμών Συναίνεσης

- Ας θυμηθούμε το (ισοδύναμο με το κεντρικοποιημένο) πρόβλημα βελτιστοποίησης

$$\boldsymbol{x}^* \in \operatorname*{argmin}_{\boldsymbol{x} \in \mathbb{R}^{n \cdot p}} \sum_{i=1}^{n} f_i(x_i) \quad subject\ to\ x_i = x_j, \forall\, i = 1,2,\ldots,n \quad and\ \ j \in n(i)$$

- Μπορούμε να εκφράσουμε το σύνολο των περιορισμών με μια πιο «συνεκτική» μορφή, χρησιμοποιώντας τον πίνακα B που ορίσαμε πριν

- Πιο συγκεκριμένα,
  - Έστω $\mathbf{I} \in \mathbb{R}^{p \times p}$ ο μοναδιαίος $p \times p$ πίνακας
  - Ορίζουμε ως $\mathbf{B} = B \otimes \mathbf{I}$, το γινόμενο Kronecker ανάμεσα στους πίνακες B και $\mathbf{I}$. Ο πίνακας αυτός έχει διαστάσεις $mp \times np$
  - Μπορούμε να εκφράσουμε το σύνολο των περιορισμών συναίνεσης μέσω της σχέσης $\mathbf{Bx} = \mathbf{0}$, όπου για κάθε ακμή $(i, j) \in E$ η τελευταία σχέση πινάκων έχει την εξίσωση $w_{ij}(x_i - x_j) = 0$, η οποία είναι «ίδια» με την $x_i = x_j$ (ουσιαστικά, έχουμε περισσότερη ευελιξία να χρησιμοποιήσουμε τα βάρη)

Χαλάρωση των Περιορισμών Συναίνεσης

- Χρησιμοποιώντας τα προηγούμενα, εκφράζουμε το πρόβλημα βελτιστοποίησης ως

$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\mathbb{R}^{n\cdot p}}{\operatorname{argmin}} F(\boldsymbol{x}) = \sum_{i=1}^{n} f_i(x_i) \quad subject\ to \ \ \mathbf{Bx} = \mathbf{0}$$

- Η δυσκολία στη βελτιστοποίηση του προβλήματος αυτού έγκειται στην ύπαρξη των περιορισμών
- **Μια επιλογή** είναι να «χαλαρώσουμε» τους περιορισμούς συναίνεσης, αλλά να προσπαθήσουμε να «τιμωρήσουμε» μεγάλες αποκλίσεις από τους περιορισμούς αυτούς
- Κάτι τέτοιο μπορεί να γίνει θεωρώντας το ακόλουθο πρόβλημα χωρίς περιορισμούς

$$\boldsymbol{x}_a^* = \underset{\boldsymbol{x}\in\mathbb{R}^{n\cdot p}}{\operatorname{argmin}}\, a \cdot F(\boldsymbol{x}) + \frac{1}{2}\|\mathbf{Bx}\|^2$$

- Τα δύο προβλήματα **δεν** είναι ισοδύναμα, ωστόσο θα μπορούσαν να είναι «κοντά» για μικρό a, το οποίο ελέγχει τη σχετική σημαντικότητα που δίνουμε στο κόστος και την ικανοποίηση των περιορισμών

Χαλάρωση των Περιορισμών Συναίνεσης

- Μπορεί να δειχθεί πως για την προσέγγιση αυτή έχουμε
$$\|\boldsymbol{x}_a^* - \boldsymbol{x}^*\| = O(a)$$
- Η παραπάνω μέθοδος αντιμετώπισης των περιορισμών συναίνεσης αποκαλείται «χαλάρωση στον πρωτεύοντα χώρο» (primal relaxation) και οδηγεί στον κατανεμημένο αλγόριθμο κλίσεων (distributed gradient descend)

Χαλάρωση των Περιορισμών Συναίνεσης

- **Μια άλλη επιλογή** είναι η εισαγωγή «χαλάρωσης» στο λεγόμενο «δυικό» χώρο (dual constraint relaxation)
- Για να το κάνουμε αυτό, ορίζουμε ένα διάνυσμα από μη-αρνητικούς πολλαπλασιαστές Lagrange, το οποίο συμβολίζουμε με $\boldsymbol{v} \in \mathbb{R}_{+}^{m \cdot p}$ και το οποίο αντιστοιχίζεται με την εξίσωση περιορισμών $\mathbf{Bx} = \mathbf{0}$
- Έτσι, θεωρούμε τη λεγόμενη Lagrangian συνάρτηση για το πρόβλημα που μελετάμε, και η οποία θα είναι
$$\mathcal{L}(\mathbf{x}, \boldsymbol{v}) = F(\mathbf{x}) + \boldsymbol{v}^T \mathbf{Bx}$$
- Έστω πως οι βέλτιστες τιμές του $\mathbf{x}$ για δοσμένες τιμές του $\boldsymbol{v}$ δίνονται από τη σχέση
$$\mathbf{x}(\boldsymbol{v}) = \underset{\mathbf{x} \in \mathbb{R}^{n \cdot p}}{\operatorname{argmin}} \mathcal{L}(\mathbf{x}, \boldsymbol{v})$$
- και οι αντίστοιχες τιμές της Lagrangian δίνονται ως
$$g(\boldsymbol{v}) = \min_{\mathbf{x} \in \mathbb{R}^{n \cdot p}} \mathcal{L}(\mathbf{x}, \boldsymbol{v}) = \mathcal{L}(\mathbf{x}(\boldsymbol{v}), \boldsymbol{v}) = F\big(\mathbf{x}(\boldsymbol{v})\big) + \boldsymbol{v}^T \mathbf{Bx}(\boldsymbol{v})$$

Χαλάρωση των Περιορισμών Συναίνεσης

- Τέλος, το πρόβλημα λύνεται από τη μεγιστοποίηση της λεγόμενης δυικής συνάρτησης. Η βέλτιστες τιμές για τους πολλαπλασιαστές Lagrange είναι το όρισμα που μεγιστοποιεί τη δυική συνάρτηση

$$\boldsymbol{v}^* = \underset{\boldsymbol{v}\in\mathbb{R}^{m\cdot p}}{\operatorname{argmax}}\, g(\boldsymbol{v})$$

- Σημαντικά σημεία της μεθόδου:
  - Σε αντιδιαστολή με τη μέθοδο χαλάρωσης στον πρωτεύοντα χώρο (primal relaxation), η οποία εισάγει μια τετραγωνική ποινή για τον περιορισμό, εδώ έχουμε μια γραμμική ποινή για τον περιορισμό
  - Το αρχικό πρόβλημα, στο οποίο εμπλεκόταν το διάνυσμα μεταβλητών $\mathbf{x}$, έχει μετασχηματιστεί σε ένα άλλο πρόβλημα στο οποίο εμπλέκονται μόνο οι πολλαπλασιαστές στο διάνυσμα $\boldsymbol{v}$
  - Όταν το αρχικό πρόβλημα είναι κυρτό, κάτω από κάποιες συνθήκες, μπορούμε να υπολογίσουμε τη βέλτιστη τιμή $\mathbf{x}^*$ του αρχικού προβλήματος

$$\mathbf{x}^* \subseteq \mathbf{x}(\boldsymbol{v}^*) = \underset{x\in\mathbb{R}^{n\cdot p}}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{x}, \boldsymbol{v}^*)$$

Παράδειγμα

(Θα δούμε τα προηγούμενα μέσα από ένα απλό παράδειγμα)

Παράδειγμα

- Έστω ένα απλό δίκτυο στο οποίο συμμετέχουν $n = 2$ κόμβοι:

- Θεωρούμε πως οι δύο αυτοί κόμβοι έχουν τις τοπικές συναρτήσεις κόστους
  - $f_1(x) = (x-3)^2,\ x \in \mathbb{R}$ και
  - $f_2(x) = (x-5)^2,\ x \in \mathbb{R}$
- Έτσι, το αρχικό, κεντρικοποιημένο πρόβλημα βελτιστοποίησης μπορεί να εκφραστεί ως

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}} \sum_{i=1}^{2} f_i(x) = (x-3)^2 + (x-5)^2$$

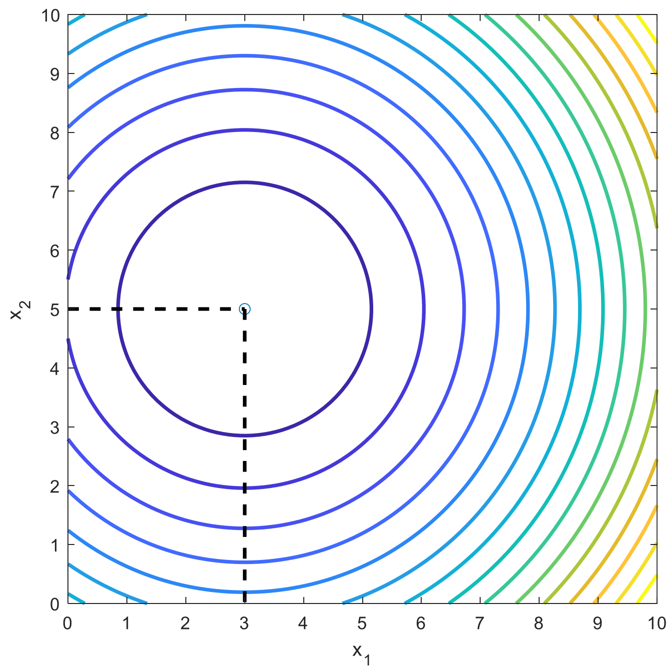
- Το πρόβλημα αυτό λύνεται εύκολα από την ελαχιστοποίηση της συνάρτησης $2x^2 - 16x + 34$, και η βέλτιστη τιμή της μεταβλητής x δίνεται ως $x^* = 4$

Παράδειγμα

- Ας δούμε τώρα το «αποσυνδεμένο» (decoupled) πρόβλημα

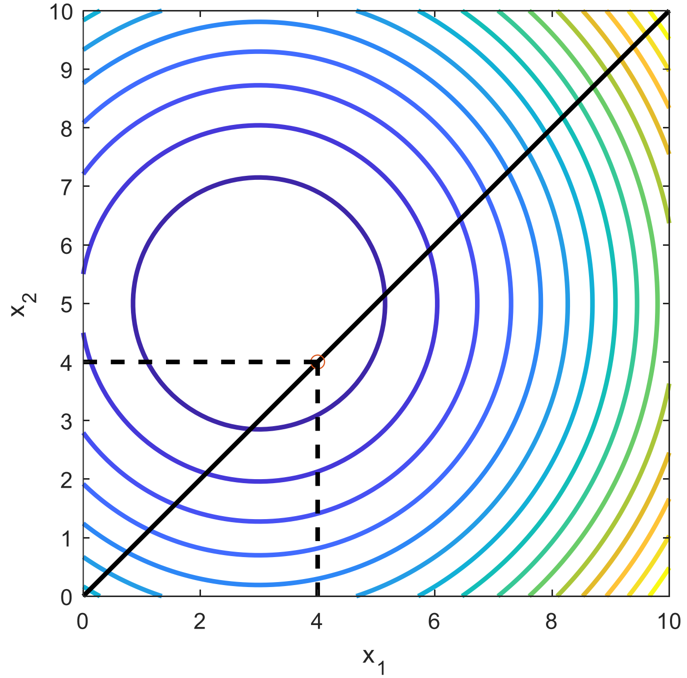
$$\tilde{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^2}{\operatorname{argmin}} F(x) = \underset{\boldsymbol{x}\in\mathbb{R}^{n\cdot p}}{\operatorname{argmin}} \sum_{i=1}^{n} f_i(x_i) = (x_1 - 3)^2 + (x_2 - 5)^2$$

- Εύκολα βρίσκουμε πως για το πρόβλημα αυτό η βέλτιστη λύση είναι $x_1 = 3$ και $x_2 = 5$, όπου, όπως αναμέναμε, οι κόμβοι δεν συμφωνούν ως προς τη λύση



Παράδειγμα

- Το «διορθωμένο» πρόβλημα, το οποίο επί της ουσίας είναι ισοδύναμο με το αρχικό κεντρικοποιημένο πρόβλημα, θα είναι
$$\boldsymbol{x}^* \in \underset{\boldsymbol{x}\in\mathbb{R}^2}{\operatorname{argmin}} F(\boldsymbol{x}) = (x_1 - 3)^2 + (x_2 - 5)^2 \;\; subject\ to\ x_1 = x_2$$
- Όπως περιμέναμε, η εισαγωγή του περιορισμού θα κάνει το πρόβλημα ισοδύναμο με το αρχικό, δίνοντας $x_1 = 4$ και $x_2 = 4$ (συμφωνία)



Παράδειγμα

«Χαλάρωση» του Προβλήματος στον Πρωτεύοντα Χώρο

- Στην περίπτωσή μας, θα έχουμε το ακόλουθο πρόβλημα βελτιστοποίησης, χωρίς περιορισμούς:

$$\boldsymbol{x}_a^* = \underset{\boldsymbol{x}\in\mathbb{R}^2}{\operatorname{argmin}}\, a \cdot \left((x_1 - 3)^2 + (x_2 - 5)^2\right) + \frac{1}{2}(x_1 - x_2)^2$$

- Για να λύσουμε αυτό το πρόβλημα, υπολογίζουμε τις μερικές παραγώγους ως προς x_1 και x_2

$$\frac{\partial g}{\partial x_1} = 2a(x_1 - 3) + (x_1 - x_2)$$

$$\frac{\partial g}{\partial x_2} = 2a(x_2 - 5) - (x_1 - x_2)$$

- Θέτοντας τις μερικές παραγώγους ίσες με το μηδέν, καταλήγουμε σε ένα γραμμικό 2×2 σύστημα εξισώσεων

$$\begin{aligned} (2a + 1)x_1 \qquad\qquad -x_2 &= 6\alpha \\ -x_1 + (2\alpha + 1)x_2 &= 10\alpha \end{aligned}$$

Παράδειγμα

- Η λύση του προηγούμενου συστήματος θα είναι (για $\alpha \neq 0$):

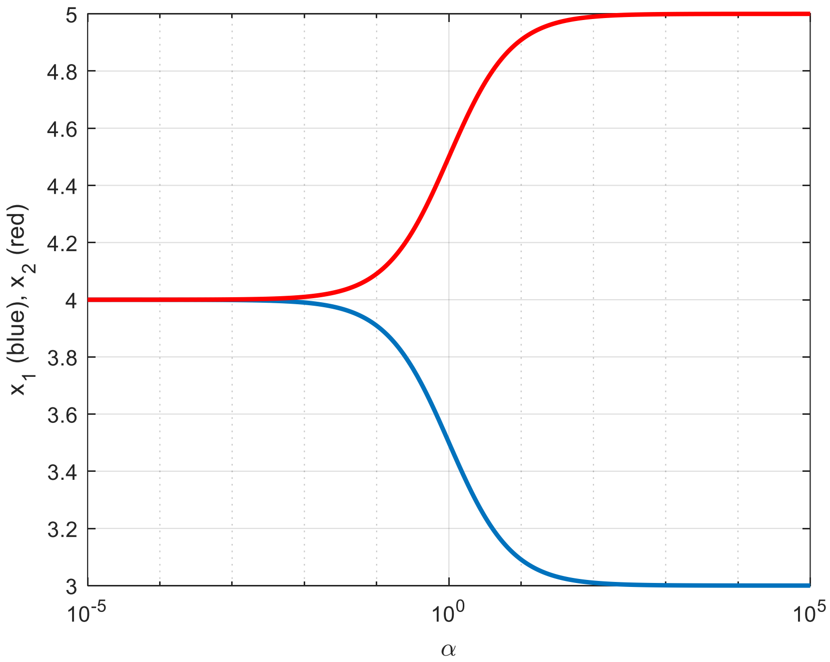
$$x_1 = \frac{3a+4}{\alpha+1} \quad \text{και} \quad x_2 = \frac{5a+4}{\alpha+1}$$

- Παρατηρούμε πως
  - Όταν το $\alpha \to 0$, έχω $x_1 \to 4$ και $x_2 \to 4$ (ο όρος ποινής έχει πολύ μεγαλύτερη επιρροή σε σχέση με τις συναρτήσεις κόστους, ωστόσο, πρέπει να τους δώσω έστω μια μικρή συνεισφορά α γιατί διαφορετικά το πρόβλημα έχει άπειρες λύσεις

  - Όταν το $\alpha \to +\infty$, έχω $x_1 \to 3$ και $x_2 \to 5$ (η συνάρτηση κόστους έχει μεγαλύτερη επιρροή σε σχέση με τον όρο ποινής και πρακτικά το πρόβλημα τείνει προς το «αποσυνδεμένο» πρόβλημα)

Παράδειγμα

- Οι λύσεις που λαμβάνουμε, για διάφορες τιμές του α
- Παρατηρούμε πως καθώς το α μειώνεται, προσεγγίζουμε συμφωνία ανάμεσα στους κόμβους του δικτύου



Παράδειγμα

«Χαλάρωση» του Προβλήματος στο Δυικό Χώρο

- Στην περίπτωσή μας, θα έχουμε τη Lagrangian συνάρτηση
$$\mathcal{L}(\mathbf{x},\lambda) = (x_1 - 3)^2 + (x_2 - 5)^2 + \lambda(x_1 - x_2)$$
- Θεωρούμε (προς το παρόν) πως η παράμετρος λ είναι μια σταθερά, και υπολογίζουμε τις μερικές παραγώγους της Lagrangian ως προς x_1 και x_2
- Λύνουμε το γραμμικό 2×2 σύστημα που προκύπτει με αυτόν τον τρόπο, και βρίσκουμε την παραμετρική λύση
$$x_1 = \frac{6-\lambda}{2} \quad \text{και} \quad x_2 = \frac{10+\lambda}{2}$$
- Έτσι λοιπόν, για δοσμένο λ, η Lagrangian συνάρτηση ελαχιστοποιείται στο σημείο
$$\boldsymbol{x}(\lambda) = \begin{bmatrix} \dfrac{6-\lambda}{2} \\ \dfrac{10+\lambda}{2} \end{bmatrix}$$

Παράδειγμα

- Αντικαθιστώντας την τιμή αυτή στη Lagrangian συνάρτηση, βρίσκουμε

$$\mathcal{L}(\boldsymbol{x}(\lambda),\lambda) = \left(\frac{6-\lambda}{2} - 3\right)^2 + \left(\frac{10+\lambda}{2} - 5\right)^2 - \lambda^2 - 2\lambda$$

- Σύμφωνα με τη θεωρία, θεωρούμε τώρα πως έχουμε μια συνάρτηση του λ (γενικότερα, των πολλαπλασιαστών Lagrange) και αναζητούμε το σημείο στο οποίο αυτή αποκτά τη μέγιστη τιμή της
- Υπολογίζουμε την παράγωγο, θέτουμε ίση με το μηδέν, και τελικά βρίσκουμε πως $\lambda = -2$ (επίσης, δεύτερη παράγωγος αρνητική)
- Τελικά, υπολογίζουμε

$$x_1 = \frac{6-\lambda}{2} = 4 \quad \text{και} \quad x_2 = \frac{10+\lambda}{2} = 4$$

- Δηλαδή, όπως περιμέναμε και από τη θεωρία, βρήκαμε τη βέλτιστη λύση!

Κατανεμημένος Αλγόριθμος Κλίσεων

(Αντιστοιχεί στην Παράγραφο 5.3 στο “Cooperative and Graph Signal Processing: Principles and Applications”)

Κατανεμημένος Αλγόριθμος Κλίσεων

- Βασίζεται στην τεχνική της «χαλάρωσης» στον πρωτεύοντα χώρο

Algorithm 5.1 DISTRIBUTED GRADIENT DESCENT (DGD) METHOD AT NODE *I*

Require: Initial iterate $x_{i,0}$, weights w_{ij}, penalty coefficient α, step sizes ϵ_k.

1: **for** $k = 0, 1, 2, \ldots$ **do**

2: Exchange iterates $x_{i,k}$ with neighbors $j \in n(i)$.

3: Evaluate gradient $g_{i,k} = \alpha \nabla f_i(x_{i,k}) + x_{i,k} - \sum_{j \in n(i)} w_{ij} x_{j,k}$.

4: Update local iterate: $x_{i,k+1} = x_{i,k} - \epsilon_k \, g_{i,k}$.

5: **end for**

Αλγόριθμος Δυικής Ανόδου

(Αντιστοιχεί στην Παράγραφο 5.4.1 στο “Cooperative and Graph Signal Processing: Principles and Applications”)

Αλγόριθμος Δυικής Ανόδου

- Βασίζεται στην τεχνική της «χαλάρωσης» στον δυικό χώρο

$$\boldsymbol{x}^k = \underset{\boldsymbol{x} \in \mathbb{R}^{np}}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{v}^{k-1}),$$

$$\boldsymbol{v}^k = \boldsymbol{v}^{k-1} - \eta_k \boldsymbol{B} \boldsymbol{x}^k,$$

Algorithm 5.2 DISTRIBUTED DUAL ASCENT

Require: Initialize all Lagrange multipliers v_l^0 such that if $e_l = (i_l, j_l)$, then both nodes i_l and j_l know the value of v_l^0.
1: **for** $k = 1, 2, \ldots,$ at each node i in parallel **do**
2: Locally solve for x_i^k as in Eq. (5.21);
3: Send the new value x_i^k to all neighbors and receive values x_j^k from each of its neighbors;
4: Compute new multipliers v_l^k according to Eq. (5.20).
5: **end for**

Ερωτήσεις