

## **HPCLab: Εργαστήριο Πληροφοριακών Συστημάτων Υψηλών Επιδόσεων**

### **ΔΙΠΛΩΜΑΤΙΚΕΣ ΕΡΓΑΣΙΕΣ**

#### **Γενική περιγραφή**

- Τα θέματα των Διπλωματικών Εργασιών αφορούν κατά κύριο λόγο στη μελέτη, αξιολόγηση, σχεδίαση και ανάπτυξη μεθόδων και εργαλείων για τον Επιστημονικό Υπολογισμό.
- Το ακαδημαϊκό προσωπικό του HPCLab αποτελείται από τον Δρ. Ιωάννη Βενέτη και τον κ. Ε. Γαλλόπουλο, Διευθυντή του HPCLab. Εκτός αν αναφέρεται διαφορετικά οι Δ.Ε. θα εκπονηθούν υπό την επίβλεψή τους (οι Δ.Ε. 6, 8, 9 θα εκπονηθούν κυρίως υπό την επίβλεψη του κ. Βενέτη). Σε ορισμένες, ενδέχεται να συμμετέχουν επίσης επικουρικά και μεταπτυχιακοί και συνεργάτες από το εξωτερικό.
- Το HPCLab είναι εξοπλισμένο με συστήματα συνεπεξεργασιών (coprocessors) πρόσφατης τεχνολογίας που θα αποτελούν συστατικό στοιχείο σύγχρονων και νέων υπερυπολογιστών, όπως για παράδειγμα της κάρτας K40 με μέγιστη επίδοση περί τα 4.5 (single precision) Tflops.
- Στο HPCLab έχουν αναπτυχθεί αλγόριθμοι, μέθοδοι και περιβάλλοντα για διάφορες εφαρμογές που αφορούν στην ανάλυση δεδομένων και στον παράλληλο υπολογισμό<sup>1</sup>. Ένα από αυτά, το TMG, διανέμεται και χρησιμοποιείται από αρκετούς ερευνητές σε πανεπιστήμια, ερευνητικά κέντρα και εταιρείες καθώς και για τη διδασκαλία Text και Data Mining. Αναφέρουμε επίσης και την πρόσφατη μονογραφία Parallelism in Matrix Computations των E. Gallopoulos, B. Philippe και A. Sameh (Springer, 2015).
- Αφετηρία σε πολλές από τις παρακάτω εκφωνήσεις είναι το γεγονός ότι η Υπολογιστική Υψηλών Επιδόσεων (HPC = High Performance Computing) είναι απαραίτητη υποδομή σε μυριάδες εφαρμογών. Με βάση αυτό και πρόσφατες μελέτες που αναδεικνύουν τη σημασία των επενδύσεων στο αντικείμενο, η απόκτηση ειδικής εμπειρίας σε αυτό μέσω μαθημάτων και Δ.Ε., εφόσον βέβαια συνοδεύεται και από ενδιαφέρον για τα συναφή ζητήματα, μπορεί να αποτελέσει πολύ χρήσιμο εφόδιο.
- Σημαντικό τμήμα κάθε διπλωματικής είναι οι υλοποιήσεις με ανάπτυξη λογισμικού. Το περιβάλλον προγραμματισμού και υλοποίησης εξαρτάται από τη Δ.Ε.
- Οι παρακάτω εκφωνήσεις παρέχουν ένα γενικό πλαίσιο το οποίο όμως μπορεί να τροποποιηθεί ή να αναθεωρηθεί μετά από συνεννόηση. Αν υπάρχει ενδιαφέρον, ορισμένες από τις Δ.Ε. μπορούν να μετεξελιχθούν σε εργασίες μεταπτυχιακού επιπέδου.
- Για περισσότερες πληροφορίες παρακαλείστε να επικοινωνήσετε το συντομότερο με τον κ. Γαλλόπουλο ή τον κ. Βενέτη.

---

<sup>1</sup>Δείτε εδώ για σχετικές εργασίες και λογισμικό.

**Προϋποθέσεις για την εκπόνηση καλής Δ.Ε.:** Η Δ.Ε. είναι συνήθως η πρώτη φορά που ένας φοιτητής αναλαμβάνει να εκπονήσει μια εκτενή εργασία<sup>2</sup> που συμπεριλαμβάνει επισκόπηση και αξιολόγηση βιβλιογραφίας, συστηματική έρευνα γύρω από ένα αντικείμενο, μελέτη θεωρίας και ανάπτυξη συνοδευτικών προγραμμάτων διεξαγωγή πειραμάτων και σύνδεσης με τη θεωρία, αξιολόγηση αποτελεσμάτων, ενίοτε συνεισφορά νέων θεωρητικών αποτελεσμάτων, και εντέλει, συγγραφή συνοδευτικού κειμένου και παρουσίασης των αποτελεσμάτων σας. Επιπλέον, η Δ.Ε. αποτελεί σημαντικό τεκμήριο στην αντιστοίχιση του Διπλώματός σας ως Master's και αποτελεί στο μέλλον βασικό στοιχείο αναφοράς για το βιογραφικό σας σημείωμα. Κατά συνέπεια, εμείς θα σας ενθαρρύνουμε να ετοιμάσετε μία αξιολόγηση Δ.Ε. Επομένως, είναι απαραίτητο να σας ενδιαφέρει η περιοχή που θα επιλέξετε, ώστε να διερευνήσετε με ευχαρίστηση το state-of-the-art και να ασχοληθείτε με αυτό με μελέτη, σχεδιασμό και υλοποίηση αλγορίθμων, κ.λπ. Δεν πρέπει να προκαλεί έκπληξη, βέβαια, το ότι πολλά θέματα αφορούν σε υπολογισμούς με μητρώα!

*Αξίζει να σημειώσετε ότι τα θέματα μπορούν να τροποποιηθούν ή να υπάρξει μεγαλύτερη εξειδίκευση εφόσον υπάρχει αμοιβαίο ενδιαφέρον.*

## Μέρος Ι

# Περιβάλλοντα επίλυσης προβλημάτων

Στο εργαστήριο έχει αναπτυχθεί, συντηρείται και επεκτείνεται το περιβάλλον Text to Matrix Generator (TMG) για τη MATLAB. Το περιβάλλον παρέχει τη δυνατότητα κατασκευής μητρώων όρων-κειμένων τη διαχείρισή του για Αναζήτηση, Συσταδοποίηση, Κατηγοριοποίηση, και Διαστατική Μείωση και πρόσφατα για δημιουργία πολυδιάστατων πινάκων με πληροφορίες υψηλότερου επιπέδου για τα κείμενα και την ανάλυσή τους (RandNLA<sup>3</sup>.) Το TMG χρησιμοποιείται από πολλούς ερευνητές παγκοσμίως, και στα πλαίσια μαθημάτων Text και Data Mining. Ορισμένες από τις Δ.Ε. που ανακοινώνονται παρακάτω αφορούν σε περαιτέρω ανάπτυξη του TMG. Εφόσον πληρούνται ορισμένα κριτήρια, το αποτέλεσμα ορισμένων Δ.Ε. θα αναδειχθεί μέσω σύνδεσής του με τη διανομή του TMG.

## 1 Επιτάχυνση του TMG σε συνεπεξεργαστή Intel® Xeon-Phi

Στόχος της Δ.Ε. είναι η αντικατάσταση στοιχείων του TMG στα οποία εντοπίζονται συμφορήσεις κατά την επεξεργασία μεγάλων αρχείων, με συναρτήσεις που αξιοποιούν τον συνεπεξεργαστή Xeon-Phi και ειδικότερα τη δυνατότητα παράλληλης και διανυσματικής επεξεργασίας. Μπορεί να γίνει χρήση του Parallel Computing Toolbox της MATLAB και με όποιον τρόπο προσφέρεται για την διεπαφή με το Xeon-Phi.

## 2 Τανυστές, τυχαιότητα και εφαρμογές στο περιβάλλον TMG: Διεπαφή και εφαρμογές

Στόχος της Δ.Ε. είναι η περαιτέρω ανάπτυξη των μεθόδων που προσφέρονται στο TMG σε συνδυασμό με πρόσφατες εργασίες για την κατασκευή τανυστών που αναδεικνύουν ιεραρχίες στο κείμενο (λέξεις,

<sup>2</sup>Ισοδύναμη με 30 μονάδες ECTS δηλ. αντίστοιχης με 5-6 μαθήματα.

<sup>3</sup>Randomized Numerical Linear Algebra. Δείτε και το θερινό σχολείο G2S3 εδώ και εδώ.

δίλεξα, προτάσεις, κ.λπ) καθώς και με άλλες εργαλειοθήκες MATLAB όπως το Tensor Toolbox καθώς και την παράλληλη βιβλιοθήκη libSkylark<sup>4</sup> που αναπτύχθηκε από την IBM και αφορά σε υλοποιήσεις αλγορίθμων RandNLA<sup>5</sup> και η προετοιμασία φιλικής διεπαφής με το TMG ώστε να καταστεί η βιβλιοθήκη ευπρόσιτη στους χρήστες του TMG.

### **3 Το περιβάλλον Julia: Μελέτη, αξιολόγηση και εφαρμογές**

Το περιβάλλον Julia<sup>6</sup> αναπτύχθηκε πρόσφατα στο MIT και αποσκοπεί να συνδυάσει υψηλές επιδόσεις και ευχρηστία στο πνεύμα της MATLAB, με μεγαλύτερες δυνατότητες όμως για παράλληλο υπολογισμό και για τη διαχείριση δεδομένων μεγάλου όγκου. Στόχος της Διπλωματικής Εργασίας είναι η διεξοδική μελέτη του συστήματος, η αξιολόγηση της επίδοσής του και η ανάπτυξη εργαλείων που αφορούν σε ενδιαφέρουσες εφαρμογές.

#### **Μέρος II**

## **Υπολογιστική Υψηλής Επίδοσης στην Ανάλυση Μεγάλων Δεδομένων (High Performance Data Analytics )**

Τα σύγχρονα υπολογιστικά συστήματα παράγουν και αποθηκεύουν τεράστιες ποσότητες δεδομένων από κάθε τομέα της επιστήμης και της καθημερινότητας. Ωστόσο η ανακάλυψη προτύπων, τάσεων και ανωμαλιών στα τεράστια αυτά σύνολα δεδομένων, καθώς και η σύνοψή τους μέσω απλών και εύχρηστων μοντέλων αποτελεί τεράστια πρόκληση. Η περιοχή των Data Analytics συνδυάζει τεχνολογίες από πολλούς τομείς για την αποδοτική και αποτελεσματική διαχείριση των δεδομένων αυτών καθώς και για την εξαγωγή συμπερασμάτων από αυτά. Σημαντικά εργαλεία που βρίσκονται στο επίκεντρο της προσπάθειας αυτής είναι οι αλγόριθμοι συσταδοποίησης (clustering algorithms). Τυπικά παραδείγματα των αλγορίθμων αυτών αποτελούν οι k-means, hierarchical clustering, k-medoids, Principal Direction Divisive Partitioning (PDDP) και παραλλαγές αυτών. Η υπολογιστική πολυπλοκότητα των αλγορίθμων αυτών τους καθιστά αργούς για το πλήθος των δεδομένων που πρέπει να επεξεργαστούν στα πλαίσια σύγχρονων εφαρμογών. Από την άλλη, υπάρχει πληθώρα διαφορετικών παράλληλων αρχιτεκτονικών που θα μπορούσαν να επιταχύνουν σημαντικά την επεξεργασία, όπως συστάδες υπολογιστών (clusters) ή επιταχυντές (accelerators), όπως κάρτες γραφικών (GPUs) και ο Intel® Xeon Phi.

Στο HPCLAB έχουν αναπτυχθεί μέθοδοι συσταδοποίησης<sup>7</sup>, που συχνά έχουν ως υπολογιστικό πυρήνα μεθόδους της (υπολογιστικής) Γραμμικής Άλγεβρας και που έχουν ενσωματωθεί στο TMG.

<sup>4</sup><http://xdata-skyllark.github.io/libskylark/>

<sup>5</sup>Randomized Numerical Linear Algebra. Δείτε και εδώ.

<sup>6</sup><http://julialang.org/>

<sup>7</sup>Π.χ. <http://goo.gl/sFVLTN> <http://goo.gl/vJQ3Tb> <http://goo.gl/PROD8u>

## 4 Συσταδοποίηση δεδομένων σε συνεπεξεργαστή GPU

Στα πλαίσια της Δ.Ε. ζητείται η παραλληλοποίηση δύο τουλάχιστων αλγορίθμων συσταδοποίησης σε GPUs και σε συστάδες αυτών. Θα γίνει χρήση νέων συστημάτων, όπως της κάρτας K40 με απώτερο στόχο τη δημιουργία παράλληλης βιβλιοθήκης συσταδοποίησης. Ενδιαφέρει επίσης και η συστηματική σύγκριση υλοποιήσεων και αξιολόγηση επιδόσεων αλγορίθμων συσταδοποίησης σε διαφορετικές παράλληλες αρχιτεκτονικές. Απώτερος σκοπός είναι η δημιουργία μιας βιβλιοθήκης μεθόδων συσταδοποίησης υψηλής επίδοσης.

## 5 Συσταδοποίηση δεδομένων σε παράλληλες αρχιτεκτονικές

Στα πλαίσια της Δ.Ε. ζητείται η παραλληλοποίηση δύο τουλάχιστων αλγορίθμων συσταδοποίησης στον Intel<sup>®</sup> Xeon Phi ή στον υπερυπολογιστή ARIS<sup>8</sup> με απώτερο στόχο τη δημιουργία παράλληλης βιβλιοθήκης συσταδοποίησης. Ενδιαφέρει επίσης και η συστηματική σύγκριση υλοποιήσεων και αξιολόγηση επιδόσεων αλγορίθμων συσταδοποίησης σε διαφορετικές παράλληλες αρχιτεκτονικές. Απώτερος σκοπός είναι η δημιουργία μιας βιβλιοθήκης μεθόδων συσταδοποίησης υψηλής επίδοσης.

## 6 Παραλληλοποίηση με χρήση έργων (tasks) του αλγορίθμου συσταδοποίησης PDDP στον συνεπεξεργαστή Intel<sup>®</sup> Xeon Phi

Ο αλγόριθμος PDDP (Principal Direction Divisive Partitioning<sup>9</sup>) αποτελεί παράδειγμα αλγορίθμου συσταδοποίησης που είναι ‘διαιρετικός’ (divisive), αντί των πιο συνηθισμένων ‘συσσωρευτικών’ (agglomerative) αλγορίθμων. Λόγω αυτής της ιδιότητας του, η παραλληλοποίηση του πάσχει από άνιση κατανομή του φόρτου εργασίας μεταξύ επεξεργαστικών μονάδων. Το πρόβλημα εντείνεται όσο περισσότερα είναι τα επεξεργαστικά στοιχεία, όπως στην περίπτωση του συνεπεξεργαστή Intel<sup>®</sup> Xeon Phi. Στα πλαίσια της Δ.Ε. ζητείται η παραλληλοποίηση του αλγορίθμου με χρήση έργων (tasks) στο προγραμματιστικό μοντέλο OpenMP, με σκοπό την καλύτερη εξισορρόπηση του φόρτου εργασίας. Θα χρησιμοποιηθούν πρόσφατες προσθήκες στο πρότυπο του OpenMP, οι οποίες επιτρέπουν τον ορισμό εξαρτήσεων μεταξύ έργων και την αποδοτική, αυτόματη χρονοδρομολόγηση τους από το σύστημα χρόνου εκτέλεσης (run-time system).

## 7 Συσταδοποίηση και προσεγγίσεις μητρώων

Η συσταδοποίηση μπορεί να χρησιμοποιηθεί για την κατασκευή προσεγγίσεων μητρώων. Δείτε για παράδειγμα το φαινομενικά τυχαίο μητρώο της εικόνας 5 που μέσω συσταδοποίησης μπορεί να μετασχηματιστεί στα μητρώα της εικόνας 7. Αυτό μπορεί να βοηθήσει σε πολλές διαδικασίες, π.χ. στην ανάλυση μεγάλων γραφημάτων<sup>10</sup> και σε σχετικές εφαρμογές<sup>11</sup>. Στα πλαίσια της Δ.Ε. θα μελετηθούν και θα αξιολογηθούν διαδικασίες αυτού του τύπου και τα χαρακτηριστικά τους, ενώ θα σχεδιαστούν εναλλακτικές και αποτελεσματικότερες τεχνικές.

<sup>8</sup><https://www.grnet.gr/en/PR-ARIS-top500-July15>

<sup>9</sup><http://www-users.cs.umn.edu/boley/publications/papers/PDDP.pdf>

<sup>10</sup><https://goo.gl/GLQnrd>

<sup>11</sup><https://goo.gl/6qIN6f>

## **8 Μελέτη μεθόδων Deep Learning και παραλληλοποίηση σε συνεπεξεργαστή (coprocessor)**

Το Deep Learning αποτελεί παρακλάδι της Μηχανικής Μάθησης και βασίζεται σε μια σειρά αλγορίθμων που προσπαθούν να μοντελοποιήσουν αφαιρετικές έννοιες υψηλού επιπέδου χρησιμοποιώντας γραφήματα με πολλαπλά επίπεδα επεξεργασίας, τα οποία με την σειρά τους αποτελούνται από πολλαπλούς γραμμικούς και μη γραμμικούς μετασχηματισμούς. Υπάρχει μεγάλη ποικιλία σε ότι αφορά την αρχιτεκτονική των αλγορίθμων deep learning, π.χ. Convolutional networks, Long short-term memory, Deep Boltzman machines, κλπ.

Στα πλαίσια της Δ.Ε. ζητείται να αναλυθούν κάποιες από τις αρχιτεκτονικές deep learning κυρίως ως προς την δυνατότητα παραλληλοποίησης τους και η παράλληλη υλοποίηση επιλεγμένων αρχιτεκτονικών σε ένα παράλληλο σύστημα που θα επιλεγεί με βάση τα χαρακτηριστικά των αρχιτεκτονικών αυτών (με ιδιαίτερη προτίμηση σε συνεπεξεργαστές GPU ή Intel® Xeon Phi) καθώς και η μελέτη των δυνατοτήτων νέων παράλληλων συστημάτων γι' αυτές τις μεθόδους.

## **9 Παραλληλοποίηση unbalanced εφαρμογών με προγραμματιστικό μοντέλο μαζικού παραλληλισμού**

Ένα πλήθος σημαντικών εφαρμογών παρουσιάζει ανομοιομορφία ως προς τον χρόνο επεξεργασίας κάθε στοιχείου δεδομένων (unbalanced applications). Η αποδοτική παραλληλοποίηση τέτοιων εφαρμογών αποτελεί πρόκληση για τα ευρέως χρησιμοποιούμενα, σύγχρονα προγραμματιστικά μοντέλα, όπως το OpenMP. Για αυτές τις εφαρμογές, για κάθε στοιχείο που πρέπει να επεξεργαστεί η ιδανική λύση θα ήταν η δημιουργία ενός ξεχωριστού νήματος. Με την ολοκλήρωση της εκτέλεσης κάθε νήματος ο αντίστοιχος επεξεργαστής αναλαμβάνει άμεσα την εκτέλεση του επόμενου νήματος. Αυτό οδηγεί στην αυτόματη εξισορρόπηση του φόρτου εργασίας μεταξύ επεξεργαστών. Ωστόσο, η διαχείριση ενός τέτοιου πλήθους νημάτων επιφέρει πολύ μεγάλες επιβαρύνσεις στην διαχείριση του παραλληλισμού και τελικά σε μεγαλύτερο χρόνο εκτέλεσης. Στα πλαίσια της Δ.Ε. θα χρησιμοποιηθεί μια υπάρχουσα βιβλιοθήκη χρόνου εκτέλεσης (run-time library), στην οποία έχουν υλοποιηθεί πρωτοπόρες τεχνικές μείωσης του χρόνου διαχείρισης του παραλληλισμού για το συγκεκριμένο είδος εφαρμογών. Σκοπός είναι η επιλογή ενός πλήθους κατάλληλων εφαρμογών, η παραλληλοποίηση τους με χρήση της προαναφερθείσας βιβλιοθήκης και η αξιολόγηση της αποτελεσματικότητας της βιβλιοθήκης.

## **10 Τοπολογική αντιστροφή για την μοντελοποίηση των παραμορφώσεων εδάφους από σεισμό**

Ο καθορισμός του επίκεντρου ενός σεισμού αποτελεί σημαντική παράμετρο για την μελέτη των ρηγμάτων που προκαλούν σεισμούς. Καθώς ο άμεσος καθορισμός τους είναι αδύνατος, αυτό θα πρέπει να γίνει έμμεσα από τις παραμορφώσεις του εδάφους που μετρώνται μετά την εμφάνιση ενός σεισμού. Η μοντελοποίηση των παραμορφώσεων εδάφους από σεισμό γίνεται τυπικά από ένα σύστημα εξαιρετικά μη γραμμικών εξισώσεων, όπως για παράδειγμα στο μοντέλο που έχει προτείνει ο Y. Okada<sup>12</sup>. Αν και η επίλυση των εξισώσεων αυτών είναι δυνατή και έτσι μπορεί να καθοριστεί ένα θεωρητικό επίκεντρο του σεισμού, στην πράξη αυτό αποδεικνύεται άχρηστο. Ο λόγος είναι πως οι μετρήσεις για την

<sup>12</sup>Surface deformation due to shear and tensile faults in a half space. Y. Okada, Bull. Seismol. Soc. Am., 75, no. 4, 1135-1154, 1985.

παραμόρφωση του εδάφους βασίζονται κυρίως σε δεδομένα GPS και έτσι υπεισέρχονται σφάλματα. Λόγω της μη γραμμικότητας των εξισώσεων το θεωρούμενο επίκεντρο από την ακριβή επίλυση των εξισώσεων με χρήση των μη ακριβών μετρήσεων παραμόρφωσης μπορεί να οδηγήσει στον υπολογισμό ενός επίκεντρου που απέχει πολύ από το πραγματικό.

Για την αντιμετώπιση του προβλήματος έχει προταθεί ο αλγόριθμος της Τοπολογικής αντιστροφής (TOPOlogical INVersion ή TOPINV). Ο αλγόριθμος αυτός λειτουργεί αντίστροφα. Δημιουργεί ένα πλέγμα γύρω από το πιθανό σημείο του επίκεντρου και για κάθε σημείο του πλέγματος υπολογίζει τις παραμορφώσεις που θα παρατηρούσαμε στο έδαφος αν το επίκεντρο ήταν το σημείο αυτό. Αν οι παραμορφώσεις αυτές συμφωνούν με τις μετρήσεις (επιτρέποντας και ένα ποσοστό σφάλματος) τότε το σημείο αυτό θεωρείται πιθανό επίκεντρο του σεισμού. Αν και ο αλγόριθμος TOPINV έχει αποδειχθεί πως δίνει εξαιρετικά αποτελέσματα σε πολλές περιπτώσεις, εν τούτοις αντιμετωπίζει ένα σημαντικό πρόβλημα. Τα σημεία του πλέγματος για τα οποία πρέπει να πραγματοποιηθούν υπολογισμοί είναι πολλές φορές της τάξης του  $10^{15}$ , ενώ μπορεί να φτάσουν ακόμα και τα  $10^{22}$ , απαιτώντας τεράστιο υπολογιστικό χρόνο.

Στα πλαίσια της Δ.Ε. ζητείται η ταυτόχρονη παραλληλοποίηση του αλγορίθμου TOPINV σε τρία επίπεδα: Με χρήση MPI, για την κατανομή του φόρτου μεταξύ υπολογιστικών κόμβων σε ένα σύστημα κατανεμημένης μνήμης, με χρήση OpenMP για την κατανομή του φόρτου μεταξύ περισσότερων πυρήνων σε κάθε υπολογιστικό κόμβο και τέλος με χρήση διανυσματοποίησης στα πλαίσια κάθε πυρήνα. Αξίζει να σημειωθεί πως υπάρχει ήδη κώδικας που υλοποιεί τον αλγόριθμο TOPINV, ενώ και οι υπολογισμοί για κάθε σημείο του πλέγματος είναι ανεξάρτητοι (δεν υπάρχουν εξαρτήσεις δεδομένων), γεγονός που απλοποιεί αρκετά την υλοποίηση. Κατά την διάρκεια ανάπτυξης της Δ.Ε. θα διερευνηθούν επίσης μέθοδοι προσαρμογής του πλέγματος (grid adaptation methods) για την επιπλέον μείωση του υπολογιστικού χρόνου.

## **11 Πολλαπλασιασμός μητρώου σε αραιή μορφή με διάνυσμα σε GPU και άλλες αρχιτεκτονικές υψηλής επίδοσης: Από τη μοντελοποίηση στην υλοποίηση**

Τα (πολύ μεγάλα) μητρώα που προκύπτουν στις κλασικές εφαρμογές (πεπερασμένες διαφορές, πεπερασμένα στοιχεία), καθώς και σε εφαρμογές που αφορούν στην ανάλυση δεδομένων<sup>13</sup> και στην αναπαράσταση δικτύων είναι συνήθως αραιά και συχνά μπορεί να έχουν επιπλέον δομή (π.χ. συμμετρία τιμών ή συμμετρία δομής, δηλ. ως προς τη θέση των μη μηδενικών στοιχείων τους.) Διάφορες μεθοδολογίες, όπως η Blocked Compressed Sparse Row (BCSR), χρησιμοποιούνται για την αποθήκευση τέτοιων μητρώων ώστε να περιοριστεί η χρήση της μνήμης. Στόχος της Δ.Ε. είναι η μελέτη διαφορετικών μεθόδων αποθήκευσης αραιών μητρώων (όπως η BCSR) και η υλοποίηση κατάλληλου αλγόριθμου για τον πολλαπλασιασμό της ειδικής μορφής με διάνυσμα ή δέσμες πυκνών διανυσμάτων σε συνεπεξεργαστές όπως NVIDIA GPU), καθώς η συγκεκριμένη πράξη αποτελεί δομικό στοιχείο πολλών αλγορίθμων επίλυσης συστημάτων.

<sup>13</sup>[http://link.springer.com/10.1007/978-3-319-44881-7\\_17](http://link.springer.com/10.1007/978-3-319-44881-7_17)

## 12 Νέα μετροπρογράμματα στην υπολογιστική υψηλής επίδοσης (HPC)

Η Δ.Ε. αυτή θα εξετάσει σύγχρονα μετροπρογράμματα για την διερεύνηση επίδοσης υπολογιστικών συστημάτων υψηλής επίδοσης. Μέχρι πρόσφατα, η “κατάταξη αναφοράς”<sup>14</sup> των υπολογιστικών συστημάτων όσον αφορά στις επιδόσεις τους σε επιστημονικούς υπολογισμούς βασιζόταν στο Linpack benchmark ή HPL<sup>15</sup>. Η μεθοδολογία αυτή έχει αμφισβητηθεί για πολλούς λόγους και ως εκ τούτου, έχουν υπάρξει πολλές προτάσεις σχετικά με το καταλληλότερο μετροπρόγραμμα, κ.λπ. Πρόσφατα, έγινε η πρόταση είναι να χρησιμοποιηθεί ο επαναληπτικός επιλυτής Conjugate Gradient (CG) (συζυγών κλίσεων) με όνομα HPCG<sup>16</sup>. Στόχος της Δ.Ε. είναι να εξετάσει τις προτάσεις που έχουν υπάρξει για τη δημιουργία μετροπρογραμμάτων, την αξιολόγηση των μεθόδων αυτών, καθώς επίσης να εξετάσει την πρόταση HPCG και να κατασκευάσει μία εναλλακτική και πιο εξελιγμένη πρόταση για μετροπρόγραμμα βασισμένη σε “πλοκαδοποιημένες” (ή μπλοκ) επαναληπτικές μεθόδους<sup>17</sup>.

## 13 Παράλληλοι αλγόριθμοι προσεγγιστικής παραγοντοποίησης μεγάλων μη αρνητικών μητρώων

Το αντικείμενο αφορά στην προσέγγιση μητρώου σε γινόμενο μη αρνητικών παραγόντων (μητρώων), ένα πρόβλημα που στη βιβλιογραφία συμβολίζεται NMF<sup>18</sup>. Οι παραγοντοποιήσεις αυτές χρησιμοποιούνται στην ανάλυση δεδομένων και σε άλλες εφαρμογές. Για παράδειγμα, στην ανάλυση δεδομένων, αν  $A = WH \in \mathbb{R}^{m \times n}$  όπου όλοι οι παράγοντες είναι μη αρνητικοί και  $k \ll \min\{m, n\}$ , τότε οι  $k$  στήλες του  $W$  μπορούν να θεωρηθούν ως ένα “λεξικό” που μπορεί να χρησιμοποιηθεί για την (προσεγγιστική) ανακατασκευή των στοιχείων (στηλών) του  $A$ . Οι αλγόριθμοι κατασκευής του NMF είναι κατ’ ανάγκην προσεγγιστικοί, καθώς το πρόβλημα είναι πολύ δύσκολο (μη κυρτό πρόβλημα βελτιστοποίησης). Ανάλογα προβλήματα, μεγαλύτερης ακόμα δυσκολίας, αφορούν στην παραγοντοποίηση μη αρνητικών μητρώων σε παράγοντες οι οποίοι αποτελούνται από στήλες και γραμμές του  $A$  (η λεγόμενη παραγοντοποίηση CUR).

Στη Δ.Ε. θα μελετηθούν αλγόριθμοι<sup>19</sup> για το NMF και το συμμετρικό NMF, κ.ά. και η παραλληλοποίηση τους για συνεπεξεργαστές όπως οι κάρτες γραφικών της NVIDIA και/ή το Xeon Phi.

## 14 Αλγόριθμοι υπολογισμών συναρτήσεων μητρώων με εφαρμογές σε βαθμολόγηση γραφημάτων και δικτύων και στην ανάκτηση πληροφορίας

Μας ενδιαφέρουν υπολογιστικές διαδικασίες για προβλήματα μεγάλου μεγέθους από εφαρμογές (κυρίως από την ανάλυση δεδομένων) τα οποία ανάγονται σε υπολογισμούς μητρώων ή τανυστών και για τα οποία καθίσταται απαραίτητη (λόγω μεγέθους και πολυπλοκότητας) η χρήση μεθόδων που συνδυάζουν

<sup>14</sup><https://www.top500.org/lists/2016/06/>

<sup>15</sup>Πρόκειται για τη χρονομέτρηση και το profiling των υπό ανάλυση H/Y στην επίλυση μεγάλων, πυκνών και τυχαίων γραμμικών συστημάτων μέσω παραγοντοποίησης LU με χρήση των βιβλιοθηκών LAPACK.

<sup>16</sup><http://www.hpcg-benchmark.org/>

<sup>17</sup>Η Δ.Ε. είναι συναφής με ζητήματα που αφορούν στην Δ.Ε. 15

<sup>18</sup>Nonnegative Matrix Factorization. Δείτε π.χ. τις <https://goo.gl/DzFK5o>, <https://goo.gl/A3MiGn>, <https://goo.gl/zfc6NN> που εκπονήθηκαν στο HPCLab.

<sup>19</sup>Δείτε για παράδειγμα <https://arxiv.org/abs/1509.09313>

ορισμένα από τα παρακάτω: α) επαναληπτικές υπολογιστικές μεθόδους, β) παράλληλη ή και κατανεμημένη επεξεργασία, γ) τεχνικές τυχαιότητας. Ειδικότερα ενδιαφέρουν (οι πολλές!) περιπτώσεις που οι κεντρικοί υπολογισμοί (υπολογιστικοί πυρήνες) μπορούν να διαμορφωθούν σε υπολογισμούς του τύπου  $C^T \Phi(A^{(k)})D$  όπου  $A^{(k)}$ ,  $k = 1, 2, \dots$  είναι μία συλλογή τετραγωνικών μητρώων,  $\Phi$  είναι κάποια συνάρτηση μητρώου και  $C, D$  διανύσματα ή πλοκάδες διανυσμάτων κατάλληλου μεγέθους. Επίσης σε υπολογισμούς χαρακτηριστικών τιμών (π.χ. της διαγωνίου, του ίχνους, επιλεγμένου ιδιοδιανύσματος) συνάρτησης του τύπου  $\Phi(A)$ . Τυπικά παραδείγματα είναι ορισμένοι δημοφιλείς δείκτες κεντρικότητας γραφημάτων (graph centrality), οι δείκτες statistical leverage, κ.λπ.<sup>20</sup>

Για παράδειγμα, ο PageRank που ήταν ο βασικός αλγόριθμος που χρησιμοποιούσε η Google για την «βαθμολόγηση» ιστοσελίδων καθώς και γενικεύσεις ή και παραλλαγές του βασίζεται στις τιμές του κυρίαρχου ιδιοδιανύσματος του μητρώου  $\alpha S + (1 - \alpha)H$  όπου το μητρώο  $H$  είναι πρώτης τάξης και το  $S$  προέρχεται από το μητρώο γειτνίασης του γραφήματος. Δείτε επίσης και άλλα παραδείγματα

Στόχος της Δ.Ε. είναι η μελέτη πρόσφατων αλγορίθμων υπολογισμού και μεθόδων υλοποίησης για μεγάλα δίκτυα και υλοποιήσεις, όπως π.χ. σε υπολογιστικές δομές που παρέχονται στα παράλληλα συστήματα του εργαστηρίου (Xeon Phi και GPUs) και με βάση σύγχρονες βιβλιοθήκες και περιβάλλοντα (όπως Julia και Elemental).

## 15 Παράλληλία στην επίλυση γραμμικών συστημάτων με πολλά δεξιά μέλη και υπολογισμός χαρακτηριστικών τιμών από μητρώα

Πολλές μεγάλες εφαρμογές, από προσομοιώσεις με μαθηματικά μοντέλα ηλεκτρομαγνητισμού ως το ζήτημα της Ποσοτικοποίησης Αβεβαιότητας (Uncertainty Quantification) που απαιτούν την εξαγωγή διαφόρων χαρακτηριστικών<sup>21</sup> από μητρώα και συναρτήσεις τους, βασίζονται σε υπολογιστικούς πυρήνες που χρησιμοποιούν την επίλυση μεγάλων γραμμικών συστημάτων με πολλά δεξιά μέλη. Η παραλληλοποίηση αυτού του προβλήματος είναι φαινομενικά εύκολη καθώς η κάθε επίλυση μπορεί να γίνεται ανεξάρτητα. Στη βιβλιογραφία όμως υπάρχουν αλγόριθμοι που προσπαθούν να ορίσουν και να αξιοποιήσουν την «κοινή πληροφορία» (που οφείλεται στο ότι το μητρώο είναι το ίδιο και το ότι οι λύσεις «ζούν» στον  $\mathbb{R}^n$  ή σε υπόχωρό του, ώστε να επιταχυνθεί η επίλυση. Ανάλογα με την περίπτωση, χρησιμοποιείται επίσης υβριδική αριθμητική (σε συνδυασμό διαφορετικών επιπέδων ακρίβειας) και στατιστικές τεχνικές. Η επιτάχυνση όμως συνεπάγεται αυξημένο κόστος επικοινωνίας. Επομένως προκύπτει ένα σημαντικό ζήτημα διερεύνησης των διαφορετικών επιλογών, των tradeoffs και της αναζήτησης της «χρυσής τομής», ανάλογα με την υπολογιστική πλατφόρμα. Στην περίπτωση μας θα χρησιμοποιήσουμε επεξεργαστές multicore, συνεπεξεργαστές Intel® Xeon Phi και NVIDIA GPU ενώ θα ζητηθεί και χρήση του υπερυπολογιστή ARIS<sup>22</sup>. Σημειώνεται ότι το θέμα αυτό έχει αναδειχθεί ως θέμα ουσίας σε πρόσφατες μελέτες Ευρωπαίων ερευνητών που ασχολούνται με την μελέτες που κρίνονται απαραίτητες για την προετοιμασία της κοινότητας για υπολογιστικά συστήματα κλίμακας Exascale.

<sup>20</sup>π.χ. <http://dl.acm.org/citation.cfm?id=2503352>, <http://arxiv.org/abs/1402.0957v3>, <http://link.springer.com/10.1007/s11075-012-9687-2>, <http://dx.doi.org/10.1016/j.apnum.2007.01.003>, <http://newslab.ece.ohio-state.edu/research/resources/TMC-2015-confidence.pdf>, <http://arxiv.org/abs/1302.6770>, <http://arxiv.org/pdf/1310.3423.pdf>, [http://www.mathcs.emory.edu/~benzi/Web\\_papers/comput\\_kron4.pdf](http://www.mathcs.emory.edu/~benzi/Web_papers/comput_kron4.pdf),

<sup>21</sup>Π.χ. τη διαγώνιο του αντιστρόφου, βλ. <http://goo.gl/ZTOzuz>, <http://goo.gl/4Ekurx>.

<sup>22</sup><https://goo.gl/7demkl>



## 16 Αλγόριθμοι υψηλής επίδοσης μέσω υβριδικής αριθμητικής σε συνεπεξεργαστές και εργαλεία διερεύνησης ακρίβειας υπολογισμών

Όπως φαίνεται από τα τεχνικά δελτία<sup>23</sup> των καρτών γραφικών, οι επιδόσεις των πράξεων σε αριθμητική κινητής υποδιαστολή είναι πολύ καλύτερες όταν αυτές υλοποιούνται με χρήση μονής αντί διπλής ακρίβειας. Επομένως, υπάρχουν πολλές προσπάθειες για την ανεύρεση ευκαιριών χρήσης μειωμένης ακρίβειας σε υπολογισμούς.

Στόχος της Δ.Ε. είναι να διερευνήσει περισσότερο το ζήτημα αυτό και ειδικότερα τις περιπτώσεις χρήσης αριθμητικής μειωμένης ακρίβειας σε υπολογισμούς με στόχο την μεγαλύτερη επίδοση. Ενδιαφέρουν ιδιαίτερα υπολογισμοί που εφαρμόζονται στην ανάλυση δεδομένων, όπως είναι π.χ. η συσταδοποίηση. Σημειώνεται ότι η χρήση μονής ακρίβειας μειώνει και τον συνολικό όγκο των δεδομένων που μεταφέρονται κατά τη διάρκεια της εκτέλεσης προγραμμάτων.

## 17 Αυτόματη διερεύνηση ακρίβειας αριθμητικών διεργασιών

Μία επιπλέον Δ.Ε. στο ζήτημα της αριθμητικής αφορά στη μελέτη και αξιολόγηση μεθόδων για την αυτόματη διερεύνηση της ακρίβειας αριθμητικών διεργασιών. Το ζήτημα αυτό έχει σημασία καθώς είναι πάντα ζητούμενη η αυτόματη εξαγωγή συμπερασμάτων σχετικά με την ακρίβεια των αποτελεσμάτων πολύπλοκων υπολογισμών ώστε να προσφέρεται στο χρήστη η δυνατότητα αξιολόγησης των αποτελεσμάτων του ακόμα και όταν δεν είναι εφικτό να εφαρμοστούν μέθοδοι όπως πίσω ανάλυση σφάλματος κ.λπ. Θα μελετηθούν πρόσφατες εργασίες και λογισμική υποστήριξη για το ζήτημα αυτό<sup>24,25</sup> και θα εξεταστούν υλοποιήσεις σε περιβάλλοντα όπως MATLAB ή Julia καθώς και εφαρμογές σε υπολογιστικούς πυρήνες που ενδιαφέρουν σε σύγχρονες εφαρμογές.

## 18 Νέα αριθμητική σε υπολογιστές: Το unum και εφαρμογές

Πρόσφατες εργασίες του J. Gustafson<sup>26</sup> ανέδειξαν το unum ως έναν ρηξικέλευθο τρόπο αναπαράστασης πραγματικών με στόχο να αντικαταστήσει τους αριθμούς κινητής υποδιαστολής. Η ιδέα είναι ότι πρόκειται για μια αποτελεσματική και ενεργειακά οικονομική αναπαράσταση κατάλληλη για τα σύγχρονα συστήματα υπολογιστών. Η προσπάθεια έχει προκαλέσει το ενδιαφέρον της κοινότητας (αλλά και την δριμεία κριτική του W. Kahan<sup>27</sup>) Στόχος είναι όλες οι αριθμητικές πράξεις να είναι εξίσου γρήγορες, να είναι εύκολο να υλοποιηθεί σε σημερινές τεχνολογίες υλικού, να μην υπάρχουν εξαιρέσεις (υποκανονικοποιημένοι αριθμοί, NaN, αρνητικό μηδέν), και κυρίως, να μην υπάρχουν σφάλματα στρογγύλευσης. Στόχος της Δ.Ε. είναι να μελετήσει αυτή την πρόταση σε σύγκριση με την αριθμητική κινητής υποδιαστολής καθώς επίσης να μελετήσει την υλοποίηση και τη χρήση της (π.χ. μέσω Julia<sup>28</sup>).

<sup>23</sup>π.χ. <http://www.nvidia.com/content/tesla/pdf/nvidia-tesla-kepler-family-datasheet.pdf>

<sup>24</sup><http://dl.acm.org/citation.cfm?id=2665073>

<sup>25</sup><https://arxiv.org/abs/1509.01347>

<sup>26</sup><http://www.johngustafson.net/unums.html>

<sup>27</sup><https://www.youtube.com/watch?v=LZAeZBVzVw&feature=youtu.be>

<sup>28</sup><https://github.com/JuliaComputing/Unums.jl>

## 19 Πολική παραγοντοποίηση, υπολογισμός και τεχνικές ανανέωσης του SVD

Στόχος της Δ.Ε. είναι η διερεύνηση αλγορίθμων για την πολική παραγοντοποίηση μητρώου ( $A = QM$  όπου  $Q$  ορθογώνιο μητρώο και  $M$  συμμετρικό θετικά ημιορισμένο) και αλγορίθμους υλοποίησής της καθώς η διεργασία αυτή βρίσκεται στο κέντρο πρόσφατων μεθόδων για τον υπολογισμό του SVD που προσφέρονται και για παράλληλη επεξεργασία. Ενδιαφέρουν επίσης αλγόριθμοι ανανέωσης των παραγόντων σε περιπτώσεις που ανανεώνεται το μητρώο με στόχο την αποφυγή επανάληψης, από την αρχή, όλης της διαδικασίας παραγοντοποίησης, καθώς είναι πολύ ακριβή. Στόχος είναι η υλοποίηση των μεθόδων που θα προκύψουν σε συστήματα του HPCLAB.

## 20 Εργοδικές ιστορίες και ανάλυση με μεθόδους τύπου PageRank

Το εργαστήριο έχει γίνει γνωστό για την ανάλυση παιδικών ιστοριών<sup>29</sup> με μεθοδολογίες τύπου PageRank έχοντας αποσπάσει πρόσφατα το βραβείο Ted Nelson<sup>30</sup> της ACM<sup>31</sup> για τις αναλύσεις που έγιναν σε παιδικές ιστορίες. Η Διπλωματική Εργασία στοχεύει στη διερεύνηση των μεθόδων αυτών σε ιστορίες που αναρτώνται στο Διαδίκτυο από χρήστες με χρήση νέων μεθόδων<sup>32</sup> και στη δημιουργία σχετικής βάσης δεδομένων με στόχο αποθηκευτήρια όπως το Florida sparse matrix collection. Ενδιαφέρει επίσης η προσθήκη νέων ιστοριών<sup>33</sup> και ανάλυσή τους με τεχνικές που λαμβάνουν υπόψη τους συνδέσμους και τα περιεχόμενα<sup>34</sup>.

## 21 Άλλα θέματα

Αυτή η επιλογή αφορά σε άλλα θέματα που θα προκύψουν μετά από συζήτηση με τους επιβλέποντες.

---

<sup>29</sup><http://www.enet.gr/?i=news.el.article&id=159435>

<sup>30</sup><http://www.sigweb.org/resources/sigweb-awards/43-ted-nelson>

<sup>31</sup><http://dl.acm.org/citation.cfm?doid=2309996.2310018>

<sup>32</sup><http://arxiv.org/abs/1409.1465>

<sup>33</sup>Δείτε και στο <http://arxiv.org/abs/1409.1465>

<sup>34</sup><http://dl.acm.org/citation.cfm?id=2462221>