Αλγοριθμικές Μέθοδοι Βελτιστοποίησης με Έμφαση σε Κατανεμημένα Προβλήματα

Εισαγωγή στην Κατανεμημένη Βελτιστοποίηση &

Διαδικαστικά του Μαθήματος

Δημήτρης Αμπελιώτης

Επίκουρος Καθηγητής, Ιόνιο Πανεπιστήμιο
Περιεχόμενα

• dSPL and Applications
  • Distributed Signal Processing and Learning
  • Some application areas

• Taxonomies of dSPL algorithms
  • Approach (gradient descent/ADMM)

• Recent / Open research issues

• Διαδικαστικά Μαθήματος
dSPL and Applications
Distributed Signal Processing and Learning

• Networks are everywhere *(Generalized notion of networks)*

• The number of networked devices grows fast, and they may employ M2M communication

• Data transmission becomes impractical

  Need for in-network problem solving

• Signal Processing and Learning Problems
  • Parameter estimation (e.g. frequency estimation)
  • Signal estimation (e.g. remove interference)
  • Detection (e.g. detect the presence of a target)
  • Learning (e.g. learn a suitable signal model)
  • Optimization (e.g. fit a model to the measurements)
Distributed Signal Processing and Learning

Centralized solutions:

• All nodes send measurements to a “fusion center”
  • High energy consumption
  • High complexity and increased delays
  • Impractical for “streaming data”
  • Single point of vulnerability
  • Privacy concerns
  • However – Best Achievable Performance
Distributed Signal Processing and Learning

Distributed solutions:
- Nodes communicate locally
- Local computations (in-network processing)
- Mechanisms for proper fusion of the local outcomes

Objectives:
- Achieve/converge to centralized performance
- Do not exchange data, exchange (tentative) outcomes
- Do not require organization (if possible)
- Be resilient to node/network failures
Applications of dSPL

- Cooperative System Identification, Source Localization
- Collecting (correlated) data sensed from nodes (the so-called sensor reachback problem)
- Models for collective motion of animal groups
  - Modeling bird flight formations
  - Study of bacteria motility
  - Modeling bee swarming behavior
  - Cooperative prey herding
- Study of social media / information propagation
  - Opinion Dynamics
Applications of dSPL

• State Estimation in Microgrids
• Distributed Spectrum Sensing
• Cooperative DoA estimation
  • The pursue for increased bandwidth drives the use of higher carrier frequencies
  • At such high frequencies, geometric channel models have been proposed
  • Agents could cooperate for channel estimation

As new approaches that require less tight assumptions are studied, many more applications will appear
Taxonomies of dSPL Algorithms
Taxonomies of dSPL Algorithms

• **Two main** families of algorithms:
  • Gradient-descent based methods (primal domain)
    • Low complexity
    • Under constant step-size $\mu$, learning converges towards a neighborhood of square-error size $O(\mu^2)$ around the true optimizer
  • The main approaches are:
    • Consensus strategy
    • Diffusion strategies (CTA and ATC variants)
  • Distributed ADMM based (primal and dual domains)
    • High complexity
    • They yield the exact minimizer
Taxonomies of dSPL Algorithms

• Most problems require the solution of

\[ \mathbf{w}_o = \arg\min_{\mathbf{w}} \left( \sum_{n=1}^{N} J_n(\mathbf{w}) \right) \]

where \( J_n(\mathbf{w}) \) is a (usually strongly convex and differentiable) cost function, associated with node \( n \), and \( \mathbf{w} \) is some vector of parameters

• Usually,

\[ J_n(\mathbf{w}) = \mathbb{E} \left[ (d_n(i) - \mathbf{u}_{n,i}^T \mathbf{w})^2 \right] \]

where \( i \) is the time index, \( d_n(i) \) is a desired response signal at node \( n \), and \( \mathbf{u}_{n,i} \) are regression vectors
Consensus Strategy

• Consensus: Reaching an agreement across the network

• Initial approaches involved computing averages (Gossip)
  • Each node has some initial value, all nodes must consent on the average of these values
  • Asynchronous – random approach
    • At each phase, a random subset of nodes is selected, their common new estimate is the average of their previous estimates
  • Synchronous – Deterministic approach
    • At each phase, each node cooperates with all its neighbors

• It was later used to compute also other (separable) functions

• Furthermore, it was extended to perform optimization:
  \[ \mathbf{w}_{n,i} = \sum_{l \in \mathcal{N}_n} a_{ln} \mathbf{w}_{l,i-1} - \mu_n(i) [\nabla J_n(\mathbf{w}_{n,i-1})]^T \]
Taxonomies of dSPL Algorithms

Diffusion Strategies (evolved from incremental strategies)

• Combine Then Adapt (CTA)

\[ \psi_{n,i-1} = \sum_{l \in \mathcal{N}_n} a_{ln} \mathbf{w}_{l,i-1} \]

\[ \mathbf{w}_{n,i} = \psi_{n,i-1} - \mu \left[ \nabla \mathbf{J}_n (\psi_{n,i-1}) \right]^T \]

• Adapt Then Combine (ATC)

\[ \psi_{n,i} = \mathbf{w}_{n,i-1} - \mu \left[ \nabla \mathbf{J}_n (\mathbf{w}_{n,i-1}) \right]^T \]

\[ \mathbf{w}_{n,i} = \sum_{l \in \mathcal{N}_n} a_{ln} \mathbf{\psi}_{l,i} \]
Taxonomies of dSPL Algorithms

Consensus Vs Diffusion

• Mean-Square-Deviation at node $n$

$$MSD_n = \lim_{i \to \infty} \left( \mathbb{E} \left[ \left\| \mathbf{w}_o - \mathbf{w}_{n,i} \right\|^2 \right] \right)$$

• Network $MSD$

$$MSD^{net} = \frac{1}{N} \sum_{n=1}^{N} MSD_n$$

• Network $MSD$ for no cooperation

$$MSD_{ncoop}^{net} \approx \frac{\mu M}{2} \cdot \left( \frac{1}{N} \sum_{n=1}^{N} \sigma^2_{v,n} \right)$$

• While, diffusion attains centralized performance

$$MSD_{cent} \approx \frac{\mu M}{2} \cdot \frac{1}{N} \cdot \left( \frac{1}{N} \sum_{n=1}^{N} \sigma^2_{v,n} \right) \approx MSD_{diff}^{net}$$
Taxonomies of dSPL Algorithms

• Comparison between ATC and CTA diffusion strategies in terms of network $MSD$

\[
MSD_{net}^{ATC} \leq MSD_{net}^{CTA}
\]

\[
MSD_{net}^{ATC} \leq MSD_{net}^{consensus}
\]

• Diffusion networks have been shown to converge faster and reach lower mean-square deviation than consensus networks, and their mean-square stability is insensitive to the choice of the combination weights. (Sheng-Yuan Tu, Ali H. Sayed, “Diffusion Strategies Outperform Consensus Strategies for Distributed Estimation Over Adaptive Networks”, IEEE TSP, December 2012)
Taxonomies of dSPL Algorithms

• Distributed Alternating Direction Method of Multipliers (ADMM)
  • Distributed solution to convex optimization problems
  • Start from an initial problem expressed as the minimization of a sum of $N$ cost functions of a common vector of variables
  • Each node uses a local version of the common vector of variables, thus, the optimization must now be performed for a much larger set of variables
  • Constraints are added to force all copies of variables agree (node connectivity is taken into account)
  • The augmented Langrangian function is derived, having a Langrange multiplier per constraint
  • ITERATE
    • Each node minimizes a (modified) local cost function with respect to the local copy of the variables vector
    • Each node updates the multipliers relative to them, to promote consensus (i.e. gradually enforce consensus)
Taxonomies of dSPL Algorithms

• Constrained distributed optimization

\[ w_o = \arg \min_w \left( \sum_{n=1}^{N} J_n(w) \right) \]

Subject to

\[ w \in \bigcap_{n=1}^{N} C_n \]

• Can be solved via
  • The *Projected sub-gradient method* (Nedic et al. 2010)
  • In effect, it constitutes a “Combine-Adapt-Project” approach

• A Combine-Project-Adapt approach exists, but for the unconstrained problem (Chouvardas et al. 2011), after properly formulating it in a set-theoretic setting.
Recent / Open research issues
Recent / Open research issues

• Single task / Multiple tasks
  • Single task networks: All nodes must estimate a global parameter vector (Lopes et al., 2006)
  • Node-Specific networks: There are global, common (to subsets of nodes) and local (to individual nodes) parameters (Bogdanović et al. 2013, Plata-Chaves et al. 2015)
  • Research related to Node-Specific networks: Also estimate the structure of the interests (Plata-Chaves et al., 2016)
  • Multitask: Parameter vectors at near-by nodes are similar (in the mean square error sense) (Chen et al. 2014)

• Non-smooth cost functions / non-smooth regularizers
  • Proximal diffusion methods (Vlaski et al., 2015, 2016)
Recent / Open research issues

• Distributed sparse diffusion algorithms: The parameter vector is sparse, the support is also being tracked
  • Liu et al. (2012)
  • Di-Lorenzo et al. (2013)

• Consensus and Diffusion approaches to yield the exact minimizer (as ADMM does)
  • Exact consensus algorithm – so called EXTRA (2015)
  • Exact diffusion algorithm (2017)
Recent / Open research issues

• Distributed (event) detection
  • Cattivelli et al. (2011)
  • Matta et al. (2016)

• Distributed signal estimation
  • Bertrand et al. (2011)
  • Node-specific: Szurlev et al. (2017)
  • Coalition formation games for signal enhancement (Ampeliotis et al. 2015)

• Distributed signal detection: Detect the presence of a signal of interest
  • Cavalcante et al. (2012)
Recent / Open research issues

• Distributed Dictionary Learning
  • P. Chainais et al. (2013): An Adapt-Then-Combine scheme is proposed for the Dictionary. Sparse approximation matrices are updated locally using Iterated Soft Thresholding
  • J. Liang et al. (2014): An approach to derive a distributed version of the well-known KSVD algorithm, using ADMM
  • J. Chen et al. (2014, 2015): Proximal algorithms are utilized. The method works in dual space. Nodes have different parts of the dictionary.
  • S. Chouvardas et al. (2015): An online version. Data is encoded using LARS and a distributed RLS algorithm is utilized for dictionary update.
  • H. Raja et al. (2013, 2015): An approach for a cloud version of KSVD.
  • D. Ampeliotis et al. (2016, 2017): Adapt-Align-Combine, POCS based DL
Διαδικαστικά Μαθήματος
Διαδικαστικά Μαθήματος

• Site του μαθήματος:
  • http://xanthippi.ceid.upatras.gr/AMVeKP/
  • Για ανακοινώσεις, βιβλιογραφία, διαλέξεις

• «Εγγραφή» στο μάθημα:
  • https://forms.gle/NdCqWXTejpvX3M4m8
  • Για να γνωρίζουμε ποιοι παρακολουθούν, για να στέλνουμε τις ανακοινώσεις κ.λπ.
Διαδικαστικά Μαθήματος

• Αξιολόγηση
  • Μέσω μιας προόδου που θα γίνει μόλις καλύψουμε το 50% της ύλης του μαθήματος – βάρος 30%
  • Παρουσίαση: Κάθε φοιτήτρια/φοιτητής θα αναλάβει μια εργασία, την οποία και θα παρουσιάσει κατά τη διάρκεια του μαθήματος και θα αξιολογηθεί. Οι παρουσιάσεις θα γίνουν κατά τη διάρκεια των τελευταίων δυο μαθημάτων – βάρος 30%
  • Τελική προφορική εξέταση – βάρος 40%
Διαδικαστικά Μαθήματος

• Πρόγραμμα διαλέξεων
  • Εισαγωγή (Παρασκευή 26/2)
  • Μάθημα 1, Α. Λάλος (Παρασκευή 5/3)
  • Μάθημα 2, Δ. Αμπελιώτης (Τετάρτη 10/3)
  • Μάθημα 3, Α. Λάλος (Παρασκευή 19/3)
  • Μάθημα 4, Δ. Αμπελιώτης (Τετάρτη 24/3)
  • Πρόοδος (Παρασκευή 2/4)
  • Μάθημα 5, Α. Λάλος (Παρασκευή 9/4)
  • Μάθημα 6, Δ. Αμπελιώτης (Τετάρτη 14/4)
  • Μάθημα 7, Α. Λάλος (Παρασκευή 23/4)
  • Μάθημα 8, Δ. Αμπελιώτης (Τετάρτη 12/5)
  • Παρουσιάσεις (α) (Παρασκευή 21/5)
  • Παρουσιάσεις (β) (Παρασκευή 28/5)