REPORT DOCUMENTATION PAGE	Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
11-08-2021	Final Report	19-Jul-2017 - 30-Jul-2021

4. TITLE AND SUBTITLE	
Final Report: Ordering for Hypothesis Testing and Beyond	5a. CONTRACT NUMBER W911NF-17-1-0331
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER 611102
6. AUTHORS	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Lehigh University Research & Sponsored Programs 526 Brodhead Avenue Bethlehem, PA 18015 -3008	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)	10. SPONSOR/MONITOR'S ACRONYM(S) ARO
U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	11. SPONSOR/MONITOR'S REPORT NUMBER(S) 69870-NS.36

12. DISTRIBUTION AVAILIBILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Rick Blum
UU	UU	UU	UU		19b. TELEPHONE NUMBER 610-758-3459

Standard Form 298 (Rev 8/98)
Prescribed by ANSI Std. Z39.18

RPPR Final Report
as of 19-Oct-2021

Agency Code: 21XD

Proposal Number: 69870NS **Agreement Number: W911NF-17-1-0331**

INVESTIGATOR(S):

Name: Ph.D. Rick Blum
Email: rblum@lehigh.edu
Phone Number: 6107583459
Principal: Y

Organization: **Lehigh University**
Address: Research & Sponsored Programs, Bethlehem, PA 180153008
Country: USA

DUNS Number: 808264444 EIN: 240795445
Report Date: 30-Oct-2021 Date Received: 11-Aug-2021
Final Report for Period Beginning 19-Jul-2017 and Ending 30-Jul-2021
Title: Ordering for Hypothesis Testing and Beyond
Begin Performance Period: 19-Jul-2017 **End Performance Period:** 30-Jul-2021
Report Term: 0-Other
Submitted By: Ph.D. Rick Blum Email: rblum@lehigh.edu
Phone: (610) 758-3459
Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 2 **STEM Participants:** 2

Major Goals: Major Goals:

This effort has developed the general theoretical underpinnings needed to design and justify a different, more energy efficient, distributed sensor processing approach which can significantly outperform existing approaches. Such developments should have major impact on many military and commercial applications, including those involving sensor systems. The ideas have been applied to as many applications as possible, especially applications of great current interest.

Motivations:

In this project, we have developed novel distributed processing approaches which can reduce the number of required communications in many applications. The distributed processing approaches that are popular with the research community require a very large number of communications and this has a number of disadvantages. The Army has many applications which employ wireless communications between battery powered nodes. In these applications, communications are typically one of the largest sources of energy usage. Thus, these excessive communications tend to drain the batteries and this can have serious negative impact on battlefield superiority. On the other hand, the latency for typical computing units to send data over a standard wired network connection is around 2500 times larger than that for accessing data in its own main memory. Thus reducing the number of communications can provide much faster results. This project seeks to develop methods which would allow the number of communications to be dramatically reduced with no significant loss in performance when compared to the performance of the best full-communication approaches.

Hypothesis testing problems (hypothesis H0 versus H1) which attempt to decide which of two known probability density functions produced a given set of observations are of considerable interest in many Army applications like communications and sensor networking. One example of such a problem attempts to determine if an enemy vehicle is located at a particular location in space based on a set of distributed sensor observations. Our team had previously developed an impressive new approach to reduce communications in distributed processing focused on solving such hypothesis testing problems. Our team originally developed this approach, called ordering, for the case where the observations were statistically independent from sensor to sensor conditioned on the hypothesis. A major objective for this project was to extend our approach using ordering, combined with an alternative approach called censoring, to cases with dependent observations. We have successfully done this as described next, while going a bit farther.

RPPR Final Report
as of 19-Oct-2021

Accomplishments: We were first able to extend the results for independent observations to a very important hypothesis testing problem which tests for a shift in the mean of a Gaussian random variable whose probability density function can be represented by a decomposable Gaussian graphical model. Gaussian graphical models allow one to use a graph to represent the dependency structure implied by the covariance matrix of the Gaussian vector. Using the graphical model, one can exactly describe the optimum distributed processing required to solve the shift in mean hypothesis testing problem. This processing reveals how to best break up the processing into distributed parts, called cliques, which can be calculated in parallel. Thus, the cliques denote clusters of nodes who need to collect all their data at a single node, call it the cluster head, in order to calculate a cluster test statistic, which can be ultimately used to calculate the optimum communication-unconstrained test statistic. To compute the optimum communication-unconstrained test statistic, the cluster test statistics only need to be summed together, just as in the independent sensor observation case. Thus, after forming the clusters (cliques), we can perform ordering over the clusters to attempt to reduce the number of needed cluster head communications with no performance loss. The nodes in a cluster have highly correlated data and they are usually physically close together. Thus, collecting their data at a cluster head only requires short distance communications whose energy can be ignored. We have shown that we can provably save more than half the cluster head transmissions while still achieving optimum full-communication performance in the case where the mean shift is sufficiently large. We next extended our results to the next most studied hypothesis testing problem, testing the covariance matrix of a Gaussian vector, assumed to follow a decomposable Gaussian graphical model. Initially, we assumed the graph structure was the same under the two hypothesis (same connections). We were able to develop a similar ordering over the cliques. For a sufficiently large value of a particular quantity (a distance measure), we have shown that the approach can provably save more than half the cluster head transmissions with no loss in performance. This quantity (distance measure) is the smallest eigenvalue of the covariance under H1 after first whitening using the covariance matrix under H0. Very recently we have shown that we can extend these results to any hypothesis testing problem (H0 vs H1) between any two known probability density functions (not just a shift in mean or a change on covariance matrix) where these two probability density functions follow any graphical models, even if one or both of the two distributions are not Gaussian or decomposable. These results apply to arbitrary graphs under each of the two hypotheses. Given the existence of a distance measure (between the two probability density functions corresponding to the two hypotheses) that satisfies some mild conditions, then for a sufficiently large distance we have proven at least half of cluster head transmissions can be saved while still achieving identical performance to the optimum communication-unconstrained approach.

The quickest change detection problem is a fundamental model for many very important practical Army problems which does not fit into the class of hypothesis testing problems just discussed. This problem is focused on detecting a change in a probability density function as quickly as possible. Very recently, we developed the first method we are aware of to apply our ordering ideas to a quickest change detection problem where we wish to find the exact time when a set of distributed sensor data changes from one probability density function to another. We assume the observations are independent over time but are statistically dependent from sensor to sensor and allow arbitrary nonGaussian graphical models.

Recently, we have shown that these same approaches be modified for use in other distributed processing problems. We first considered one of the most popular learning and optimization algorithms in use today, the gradient descent algorithm. Reducing communications for gradient descent has been identified as a very important problem. To distribute the processing, we considered the commonly studied worker-server distributed architecture where a server wants to optimize a function and each worker will be assigned computational work to help distribute the processing. We applied ordering to this scenario. Our ordered gradient approach provably achieves the same order of convergence rate as the standard distributed gradient descent approach for nonconvex smooth loss functions while the standard distributed gradient descent always requires more worker communications. Experiments show significant worker communication savings compared to the best existing approaches in some important cases. In particular, we compare to another popular approach for reducing communications which is based on censoring and we find very significant gains in some cases. The results also indicate that one can incorporate both ordering and censoring to obtain even better performance in some cases. The heavy ball algorithm is a modified gradient descent algorithm which involves a momentum term. This momentum term can significantly improve optimization/learning performance. We have recently investigated communications savings in the heavy ball algorithm. We have shown that both censoring and ordering approaches can save considerable worker transmissions in a worker-server architecture using extensive numerical and analytical results.

In very recent work, we have applied our communication reduction methods for new architectures beyond the worker-server architecture. For example, we have developed approaches for a communication efficient approach building on the popular alternating direction method of multipliers (ADMM) algorithm where each node communicates only with its immediate neighbors. We have also applied our communication reduction ideas to the normalized gradient descent (coordinate descent) algorithm.

Training Opportunities: PhD students and postdocs have been trained in how to properly perform research by our team. This includes how to write papers and proposals; how to present results; how to choose topics; and how to focus on important work. The students and postdocs were able to present their work at conferences, in venues at Lehigh University and to our ARL collaborators. One of the students was able to work at ARL over the last four summers. This student is currently working at ARL, having joined ARL after graduation based on the great experiences he had working with our ARL collaborators. This grant had a tremendous impact on this student ultimately joining ARL and I am sure ARL would like to thank ARO for this. The interaction with ARL has been great and we will continue to work with Brian Sadler and Paul Yu. My student was able to do some realistic testing at ARL that we could not do at Lehigh. We have published several excellent journal (4) and conference papers (4) with our ARL collaborators and we were also awarded a patent as a team. We feel that our ARL collaborators have guided us to important topics and helped us find excellent solutions. They have been great role models and collaborators for my graduate students, even coauthoring many journal and conference papers. Our team graduated one PhD student December 2019 who joined Qualcomm. We graduated two more PhD students August 2020. One joined Apple and the other ARL. We graduated two more PhD students in May 2021. One joined a company working on connected vehicle networks and the other has not yet accepted any offers. He has offers from US universities and companies.

Results Dissemination: Our research results have been published in journal papers in the best journals and we have presented our results at the best conferences. The PI has given talks on the research at the best universities. He was also an IEEE Signal Processing Society Distinguished Lecturer and he frequently presented this work to broad audiences under those lectures. All papers are first put on the archive ArXiv before publication and then the reference for the accepted paper is put on our lab website.

Honors and Awards: The PI completed his term as an IEEE Signal Processing Society Distinguished Lecturer during this grant. Being named an IEEE Signal Processing Society Distinguished Lecturer is a great honor that implies our group is producing important research. This gave our group an outstanding opportunity to present our work all over the world in the form of 3 or 4 talks each month at the top universities and companies, helping to spread our ideas very broadly.

Protocol Activity Status:

Technology Transfer: We were issued a patent on work we completed on authentication with our collaborators at ARL. The team included Brian Sadler, Paul Yu, a Lehigh PhD student (Jake Perazzone) and the PI. ARL filed for the patent and it seems the Army is very interested in this technology. The PI has developed very close collaboration with ARL, meeting each week during the project for collaboration and joint papers. One graduate student working on the project (Jake Perazzone) is now full time at ARL. The PI is continuing collaroration and planning for more in the future.

PARTICIPANTS:

Participant Type: PD/PI
Participant: Ricky S Blum
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

Participant Type: Graduate Student (research assistant)
Participant: Yicheng Chen
Person Months Worked: 2.00 **Funding Support:**
Project Contribution:
National Academy Member: N

ARTICLES:

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Signal Processing Letters
Publication Identifier Type: Other Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: 7/28/20 12:00AM Date Published:
Publication Location:
Article Title: Shared channel ordered transmissions for energy-efficient signal detection in sensor networks
Authors: N. Sriranga, K. G. Nagananda, R. S. Blum
Keywords: Distributed detection; energy efficiency; censoring; ordering; collisions
Abstract: An existing approach to energy-efficient signal detection in N?sensor networks is to order the transmissions such that highly informative sensors enjoy higher priority for transmission. Then when sufficient evidence is collected at the fusion center (FC), the transmissions are stopped. This scheme incurs the same error probability as the optimum unconstrained energy approach, but with fewer sensor transmissions leading to significant energy savings. However, it is assumed that each sensor uses an orthogonal channel to transmit to the FC. For a shift-in-mean signal detection problem, the ordering approach is revisited here by considering the shared medium between sensors and the FC which can be more efficient, but gives rise to collisions. After the N sensors make observations they wish to transmit to the FC, time is divided into frames and within each frame, sensors with similar likelihood ratios (LRs) contend for transmission. The contention for transmission is resolved using ALOHA.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Information Forensics and Security
Publication Identifier Type: DOI Publication Identifier: 10.1109/TIFS.2018.2812202
Volume: 13 Issue: 9 First Page #: 2216
Date Submitted: 8/6/19 12:00AM Date Published: 9/1/18 4:00AM
Publication Location: IEEE, nj
Article Title: Cryptographic Side-Channel Signaling and Authentication via Fingerprint Embedding
Authors: Jake B. Perazzone, Paul L. Yu, Brian M. Sadler, Rick S. Blum
Keywords: Authentication, fingerprint embedding, sidechannel signaling, physical layer security
Abstract: Authentication via fingerprint embedding at the physical layer utilizes noise in the wireless channel to attain a certain degree of information theoretic security that traditional HMAC methods cannot provide. Fingerprint embedding refers to a key-aided process of superimposing a low-power tag to the primary message waveform for the purpose of authenticating the transmission. The tag is uniquely created from the message and key and successful authentication is achieved when the correct tag is detected by the receiver. This paper generalizes a framework for embedding physical layer fingerprints to create an authenticated side-channel for minimal cost. Side-channel information is conveyed to the receiver through the transmitter’s choice of tag from a secret codebook generated by the primary message and a shared secret key. Additionally, a new linear coding scheme is introduced which enhances the ability to trade off the performance goals of authentication, side-channel rate, secrecy, ...
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 19-Oct-2021

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Communications Surveys and Tutorials
Publication Identifier Type: DOI Publication Identifier: 10.1109/COMST.2018.2820021
Volume: 20 Issue: 3 First Page #: 1710
Date Submitted: 8/6/19 12:00AM Date Published: 8/1/19 4:00AM
Publication Location: IEEE, NJ
Article Title: A survey of caching techniques in cellular networks: Research issues and challenges in content placement and delivery strategies
Authors: Liying Li, Guodong Zhao, Rick S. Blum
Keywords: Caching, cache-enabled networks, cellular networks, D2D networks, HetNets
Abstract: Mobile data traffic is currently growing exponentially and these rapid increases have caused the backhaul data rate requirements to become the major bottleneck to reducing costs and raising revenue for operators. To address this problem, caching techniques have attracted significant attention since they can effectively reduce the backhaul traffic by eliminating duplicate data transmission that carries popular content. In addition, other system performance metrics can also be improved through caching techniques, e.g., spectrum efficiency, energy efficiency, and transmission delay. In this paper, we provide a systematical survey of the state-of-the-art caching techniques that were recently developed in cellular networks, including macro-cellular networks, heterogeneous networks (HetNets), device-to-device (D2D) networks, cloud-radio access networks (C-RANs), and fog-radio access networks (F-RANs). In particular, we give a tutorial on the fundamental caching techniques and ...
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: DOI Publication Identifier: 10.1109/TCOMM.2019.2900317
Volume: 67 Issue: 5 First Page #: 3624
Date Submitted: 7/28/20 12:00AM Date Published:
Publication Location:
Article Title: Improved Detection Performance of Passive Radars Exploiting Known Communication Signal Form
Authors: Anantha K. Karthik, Rick S. Blum
Keywords: Passive Radar, Generalized Likelihood Ratio Test, Code-Division Multiple Access, Digital Video Broadcasting- Terrestrial standard, Long Term Evaluation.
Abstract: In this paper, we address the problem of target detection in passive multiple-input-multiple-output (MIMO) radar networks. A generalized likelihood ratio test is derived, assuming prior knowledge of the signal format used in the non-cooperative transmit stations. We consider scenarios in which the unknown transmitted signal uses either a linear digital modulation scheme or the Orthogonal Frequency-Division Multiplexing (OFDM) modulation scheme. These digital modulation schemes are used in popular standards including Code-Division Multiple Access (CDMA), Digital Video Broadcasting-Terrestrial (DVB-T) and Long Term Evaluation (LTE). The performance of the generalized likelihood ratio test in the known signal format case is often significantly more favorable when compared to the case that does not exploit this information. Further, the performance improves with increasing number of samples per symbol and, for a sufficiently large number of samples per symbol, the performance closely ....
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: Publication Identifier: 10.1109/TSP.2018.2883922
Volume: Issue: First Page #:
Date Submitted: 7/28/20 12:00AM Date Published:
Publication Location:
Article Title: On the Impact of Unknown Signals in Passive Radar with Direct Path and Reflected Path Observations
Authors: Yicheng Chen, Rick S. Blum
Keywords: Cramer-Rao bound, joint estimation, unknown signals, passive radar.
Abstract: We derive the closed form Cramer-Rao bound (CRB) expressions for joint estimation of time delay and Doppler shift with unknown signals with possibly known structure. The results are especially useful for passive radar where direct path and reflected path signals are present. Time delay and Doppler shift estimation is an important fundamental tool in signal processing which has received extensive study for cases with known transmitted signals, but little study for unknown transmitted signals. The presented results generalize previous results for known transmitted signals and show how many looks from the direct path and the reflected path we need to derive an accurate joint estimation of time delay and Doppler shift. After analysis under a simple common signal-to-clutter-plus-noise ratio (SCNR) model with separated direct and reflected path signals, white clutter-plus-noise and line of sight propagation, extensions to other interesting (practical) cases are provided.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: DOI Publication Identifier: 10.1109/TSP.2018.2802459
Volume: 66 Issue: 8 First Page #: 2070
Date Submitted: 7/13/18 12:00AM Date Published: 4/1/18 4:00AM
Publication Location:
Article Title: Attack Detection in Sensor Network Target Localization Systems With Quantized Data
Authors: Jiangfan Zhang, Xiaodong Wang, Rick S. Blum, Lance M. Kaplan
Keywords: Target localization, attack detection, spoofing attack, man-in-the-middle attack, malfunction, sensor network, large deviations theory
Abstract: We consider a sensor network focused on target localization, where sensors measure the signal strength emitted from the target. Each measurement is quantized to one bit and sent to the fusion center. A general attack is considered at some sensors that attempts to cause the fusion center to produce an inaccurate estimation of the target location. The attack is a combination of man-in-the-middle, hacking, and spoofing attacks that can effectively change both signals going into and coming out of the sensor nodes in a realistic manner. We show that the essential effect of attacks is to alter the naive estimate of the distance between the target and each attacked sensor, which ignores the existence of attacks, to a different extent, giving rise to a geometric inconsistency among the attacked and unattacked sensors. With the help of two secure sensors, a class of detectors are proposed to detect the attacked sensors by scrutinizing the existence of the geometric inconsistency. We show that t
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: Fusion Conference
Publication Identifier Type: Other Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: 7/13/18 12:00AM Date Published: 7/10/18 4:00AM
Publication Location:
Article Title: Energy-efficient Decision Fusion for Distributed Detection in Wireless Sensor Networks
Authors: N. Sriranga, K. G. Nagananda, R. S. Blum, and P. K. Varshney
Keywords: Distributed detection; counting rule; ordering; energy-efficiency.
Abstract: This paper proposes an energy-efficient counting rule for distributed detection by ordering sensor transmissions in wireless sensor networks. In the counting rule-based detection in an N?sensor network, the local sensors transmit binary decisions to the fusion center, where the number of all N local-sensor detections are counted and compared to a threshold. In the ordering scheme, sensors transmit their unquantized statistics to the fusion center in a sequential manner; highly informative sensors enjoy higher priority for transmission. When sufficient evidence is collected at the fusion center for decision making, the transmissions from the sensors are stopped. The ordering scheme achieves the same error probability as the optimum unconstrained energy approach (which requires observations from all the N sensors) with far fewer sensor transmissions. The scheme proposed in this paper improves the energy efficiency of the counting rule detector by ordering the sensor transmissions.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: DOI Publication Identifier: 10.1109/TSP.2018.2849704
Volume: Issue: First Page #: 1
Date Submitted: 7/13/18 12:00AM Date Published: 8/15/18 4:00AM
Publication Location:
Article Title: Signal Amplitude Estimation and Detection from Unlabeled Binary Quantized Samples
Authors: Guanyu Wang, Jiang Zhu, Rick Blum, Peter K. Willett, Stefano Marano, Vincenzo Matta, Paolo Braca
Keywords: Estimation, detection, permutation, unlabeled sensing, quantization, identifiability, alternating maximization.
Abstract: Signal amplitude estimation and detection from unlabeled quantized binary samples are studied, assuming that the order of the time indexes is completely unknown. First, maximum likelihood (ML) estimators are utilized to estimate both the permutation matrix and unknown signal amplitude under arbitrary but known signal shape and quantizer thresholds. Sufficient conditions are provided, under which an ML estimator can be found in polynomial time, and an alternating maximization algorithm is proposed to solve the general problem via good initialization. In addition, the statistical identifiability of the model is studied. Furthermore, an approximation of the generalized likelihood ratio test detector is adopted to detect the presence of the signal. In addition, an accurate approximation of the probability of successful permutation matrix recovery is derived, and explicit expressions are provided to reveal the relationship between the signal length and the number of quantizers.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 19-Oct-2021

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Signal Processing Letters
Publication Identifier Type: DOI Publication Identifier: 10.1109/LSP.2017.2769129
Volume: 25 Issue: 1 First Page #: 100
Date Submitted: 7/13/18 12:00AM Date Published: 1/1/18 12:00AM
Publication Location:
Article Title: On the Analysis of the Fisher Information of a Perturbed Linear Model After Random Compression
Authors: Jiang Zhu, Lin Han, Rick S. Blum, Zhiwei Xu
Keywords: Cramer-Rao bound, Fisher information, pertur- ´ bation, random compression.
Abstract: The impact of random compression on the Fisher information matrix (FIM) and the Cramer-Rao bound (CRB) is studied when estimating unknown complex parameters in the perturbed linear model. A random compression matrix is considered whose elements are i.i.d. standard complex normal random variables. The FIM averaged over compression is equal to a scalar of the FIM before compression plus an additional term. The upper and lower bounds of the CRB averaged over the random compression matrix are also given. Finally, numerical results are conducted to verify our theoretical results.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: 2018 IEEE International Symposium on Information Theory
Publication Identifier Type: Other Publication Identifier:
Volume: Issue: First Page #:
Date Submitted: 7/19/18 12:00AM Date Published: 6/17/18 4:00AM
Publication Location: NY, NY
Article Title: Inner Bound for the Capacity Region of Noisy Channels with an Authentication Requirement
Authors: Jake Perazzone, Eric Graves, Paul Yu and Rick Blum
Keywords: Inner Bound, Capacity Region, Authentication
Abstract: The rate regions of many variations of the standard and wire-tap channels have been thoroughly explored. Secrecy capacity characterizes the loss of rate required to ensure that the adversary gains no information about the transmissions. Authentication does not have a standard metric, despite being an important counterpart to secrecy. While some results have taken an information-theoretic approach to the problem of authentication coding, the full rate region and accompanying trade-offs have yet to be characterized. In this paper, we provide an inner bound of achievable rates with an average authentication and reliability constraint. The bound is established by combining and analyzing two existing authentication schemes for both noisy and noiseless channels. We find that our coding scheme improves upon existing schemes.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 19-Oct-2021

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: DOI Publication Identifier: 10.1109/TSP.2018.2883922
Volume: 67 Issue: 2 First Page #: 431
Date Submitted: 8/6/19 12:00AM Date Published: 1/1/19 5:00AM
Publication Location:
Article Title: On the Impact of Unknown Signals on Delay, Doppler, Amplitude, and Phase Parameter Estimation
Authors: Yicheng Chen, Rick S. Blum
Keywords: Cramer-Rao bound, known signal format, unknown signals, parameter estimation
Abstract: The estimation of time delay, Doppler shift, amplitude, and phase is an important fundamental tool in signal processing, which has received extensive study for cases with known transmitted signals, but little study for unknown transmitted signals.Wederivetheclosed-formCram´er–Raobound(CRB)expressions for joint or separate estimation of time delay, Doppler shift, amplitude, and phase with unknown signals with possibly known structure and possible multiple looks at direct path and re?ected path observations. The presented results generalize previous resultsforknowntransmittedsignalsandshowhowmanylooksfrom thedirectpathandthere?ectedpathweneedtoderiveanaccurate estimation of time delay, Doppler shift, amplitude, and phase. The advantages of the known signal format with unknown parameters over totally unknown signals are illustrated. After analysis under asimplewhiteclutter-plus-noisemodel,extensionstothecasewith dependentclutterplusnoisearediscussed.Numericalresultsshow very similar behav
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Signal Processing Letters
Publication Identifier Type: DOI Publication Identifier: 10.1109/LSP.2018.2870341
Volume: 25 Issue: 11 First Page #: 1625
Date Submitted: 8/6/19 12:00AM Date Published: 11/1/18 4:00AM
Publication Location:
Article Title: Improved Detection Performance for Passive Radars Exploiting Known Communication Signal Form
Authors: Anantha K. Karthik, Rick S. Blum
Keywords: Code-division multiple access, digital video broadcastingterrestrialstandard, generalizedlikelihoodratiotest, passive radar.
Abstract: In this letter, we address the problem of target detection in passive multiple-input multiple-output radar networks. A generalized likelihood ratio test is derived, assuming prior knowledge of the signal format used in the noncooperative transmit stations. The performance of the generalized likelihood ratio test in the known signal format case is often signi?cantly more favorable when compared to the case that does not exploit this information. Further, the performance improves with increasing number of samples per symbol and for a suf?ciently large number of samples per symbol, the performance closely approximates that of an active radar with a known transmitted signal.
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

RPPR Final Report
as of 19-Oct-2021

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Information Theory
Publication Identifier Type: DOI Publication Identifier: 10.1109/TIT.2018.2850968
Volume: 64 Issue: 9 First Page #: 6180
Date Submitted: 8/6/19 12:00AM Date Published: 9/1/18 4:00AM
Publication Location:
Article Title: A Fundamental Limitation on Maximum Parameter Dimension for Accurate Estimation With Quantized Data
Authors: Jiangfan Zhang, Rick S. Blum, Lance M. Kaplan, Xuanxuan Lu
Keywords: Distributed sensor parameter estimation, inestimable dimension for quantized data, singular Fisher information matrix, identi?ability, quantization.
Abstract: It is revealed that there is a link between the quantization approach employed and the dimension of the vector parameter which can be accurately estimated by a quantized estimation system. A critical quantity called inestimable dimension for quantized data (IDQD) is introduced, which does not depend on the quantization regions and the statistical models of the observations but instead depends only on the number of sensors and on the precision of the vector quantizers employed by the system. It is shown that the IDQD describes a quantizationinduced fundamental limitation on the estimation capabilities of the system. To be speci?c, if the dimension of the desired vector parameter is larger than the IDQD of the quantized estimation system, then the Fisher information matrix for estimating the desired vector parameter is singular, and, moreover, there exist in?nitely many nonidenti?able vector parameter points in the vector parameter space. Furthermore, it is shown that under some common a
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Communications Letters
Publication Identifier Type: DOI Publication Identifier: 10.1109/LCOMM.2018.2883045
Volume: 23 Issue: 1 First Page #: 96
Date Submitted: 8/6/19 12:00AM Date Published: 1/1/19 5:00AM
Publication Location:
Article Title: Shared Channel Ordered Transmissions for Energy-Efficient Distributed Signal Detection
Authors: N. Sriranga, K. G. Nagananda, R. S. Blum
Keywords: Distributed detection, energy ef?ciency, censoring, ordering, collisions.
Abstract: An existing approach to energy-ef?cient signal detection in sensor networks is to order the sensor transmissions such that highly informative sensors enjoy higher priority for transmission. Then, when suf?cient evidence is collected at the fusion center (FC) for decision making, the transmissions are stopped. This scheme incurs the same error probability as the optimum unconstrained energy approach, but with fewer sensor transmissions leading to signi?cant energy savings. However, it is assumed that each sensor uses an orthogonal channel to transmit to the FC. In this paper, the ordering approach is revisited by considering the shared medium between sensors and the FC. After the N sensors make observations they wish to transmit to theFC, timeis dividedintoframes and withineach frame, sensors with similar likelihood ratios contend for transmission. The contention for transmission is resolved using Slotted ALOHA. Under certain conditions, this new scheme requires less than half the senso
Distribution Statement: 3-Distribution authorized to U.S. Government Agencies and their contractors
Acknowledged Federal Support: **Y**

Publication Type: Journal Article Peer Reviewed: Y **Publication Status:** 1-Published
Journal: IEEE Transactions on Signal Processing
Publication Identifier Type: Publication Identifier: 10.1109/TSP.2019.2940119
Volume: 67 Issue: 20 First Page #: 5391
Date Submitted: 7/28/20 12:00AM Date Published: 10/1/19 4:00AM
Publication Location:
Article Title: Testing the Structure of a Gaussian Graphical Model with Reduced Transmissions in a Distributed Setting
Authors: Yicheng Chen, Rick S. Blum, Brian M. Sadler, and Jiangfan Zhang
Keywords: Covariancematrixtesting, distributeddetection, energy ef?ciency, Gaussian graphical models, ordered transmission.
Abstract: Testing a covariance matrix following a Gaussian graphical model (GGM) is considered in this paper based on observations made at a set of distributed sensors grouped into clusters. Ordered transmissions are proposed to achieve the same Bayes risk as the optimum centralized energy unconstrained approachbutwithfewertransmissionsandacompletelydistributed approach. In this approach, we represent the Bayes optimum test statistic as a sum of local test statistics which can be calculated by only utilizing the observations available at one cluster. We select one sensor to be the cluster head (CH) to collect and summarize the observed data in each cluster and intercluster communications are assumed to be inexpensive. The CHs with more informative observations transmit their data to the fusion center (FC) ?rst. By halting before all transmissions have taken place, transmissions can be saved without performance loss. It is shown that this ordering approach can guarantee a lower bound on savings.
Distribution Statement: 2-Distribution Limited to U.S. Government agencies only; report contains proprietary info
Acknowledged Federal Support: **Y**

CONFERENCE PAPERS:

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2018 International Conference on Information Fusion (FUSION)
Date Received: 06-Aug-2019 Conference Date: 10-Jul-2018 Date Published: 13-Jul-2018
Conference Location: Cambridge, United Kingdom
Paper Title: Energy-efficient Decision Fusion for Distributed Detection in Wireless Sensor Networks
Authors: N. Sriranga, K. G. Nagananda, R. S. Blum, P. K. Varshney
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2018 52nd Asilomar Conference on Signals, Systems, and Computers
Date Received: 06-Aug-2019 Conference Date: 28-Oct-2018 Date Published:
Conference Location: Pacific Grove, CA, USA
Paper Title: Ordered Transmission for Efficient Wireless Autonomy
Authors: Yicheng Chen, Brian M. Sadler, Rick S. Blum
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2019 53rd Annual Conference on Information Sciences and Systems (CISS)
Date Received: 07-Aug-2019 Conference Date: 20-Mar-2019 Date Published:
Conference Location: Baltimore, MD, USA
Paper Title: Physical Layer Authentication via Fingerprint Embedding: Min-Entropy Analysis : Invited Presentation
Authors: Jake Bailey Perazzone, Paul L. Yu, Brian M. Sadler, Rick S. Blum
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: IEEE Conf on Comm and Network Security
Date Received: 07-Aug-2019 Conference Date: 30-Jun-2019 Date Published: 30-Jun-2019
Conference Location: Wash, DC
Paper Title: Artificial noise and physical layer authentication: Miso regime
Authors: J. B. Perazzone, L. Y. Paul, B. M. Sadler, and R. S. Blum
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2020 54th Annual Conference on Information Sciences and Systems (CISS)
Date Received: 28-Jul-2020 Conference Date: 18-Mar-2020 Date Published:
Conference Location: Princeton, NJ, USA
Paper Title: Reduced-rank Least Squares Parameter Estimation in the Presence of Byzantine Sensors
Authors: K. G. Nagananda, Rick. S. Blum, Alec Koppel
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 1-Published
Conference Name: 2020 54th Annual Conference on Information Sciences and Systems (CISS)
Date Received: 29-Jul-2020 Conference Date: 18-Mar-2020 Date Published:
Conference Location: Princeton, NJ, USA
Paper Title: Cybersecurity of Inference in Vehicular Ad-hoc Networks : Invited Presentation
Authors: Z. Wang and R. S. Blum
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 2-Awaiting Publica
Conference Name: Workshop on Signal Processing Advances in Wireless Communications (SPAWC)
Date Received: 29-Jul-2020 Conference Date: 01-May-2020 Date Published:
Conference Location: Atlanta, USA
Paper Title: Ordered gradient approach for Communication-Efficient distributed learning,” in 2020 IEEE 21st International
Authors: Y. Chen, B. Sadler, R. Blum
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 0-Other
Conference Name: Y. Chen, R. Blum, and B. Sadler, “Optimal quickest c2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)
Date Received: 29-Jul-2020 Conference Date: 01-May-2020 Date Published:
Conference Location: Atlanta, USA
Paper Title: Optimal quickest change detection in sensor networks using ordered transmissions
Authors: Y. Chen, R. Blum, B. Sadler
Acknowledged Federal Support: **Y**

Publication Type: Conference Paper or Presentation **Publication Status:** 4-Under Review
Conference Name: Neural Information Processing System. NeurIPS
Date Received: 29-Jul-2020 Conference Date: 06-Dec-2020 Date Published:
Conference Location: Virtual
Paper Title: Communication-efficient heavy ball distributed learning using censoring
Authors: Y. Chen, R. S. Blum, B. M. Sadler
Acknowledged Federal Support: **Y**

Partners

,

Jake Perazzone - Army Research Lab - was supported in previous years of grant Zisheng Wang - Lehigh University

I certify that the information in the report is complete and accurate:
Signature: Ricky S. Blum
Signature Date: 8/11/21 10:03AM

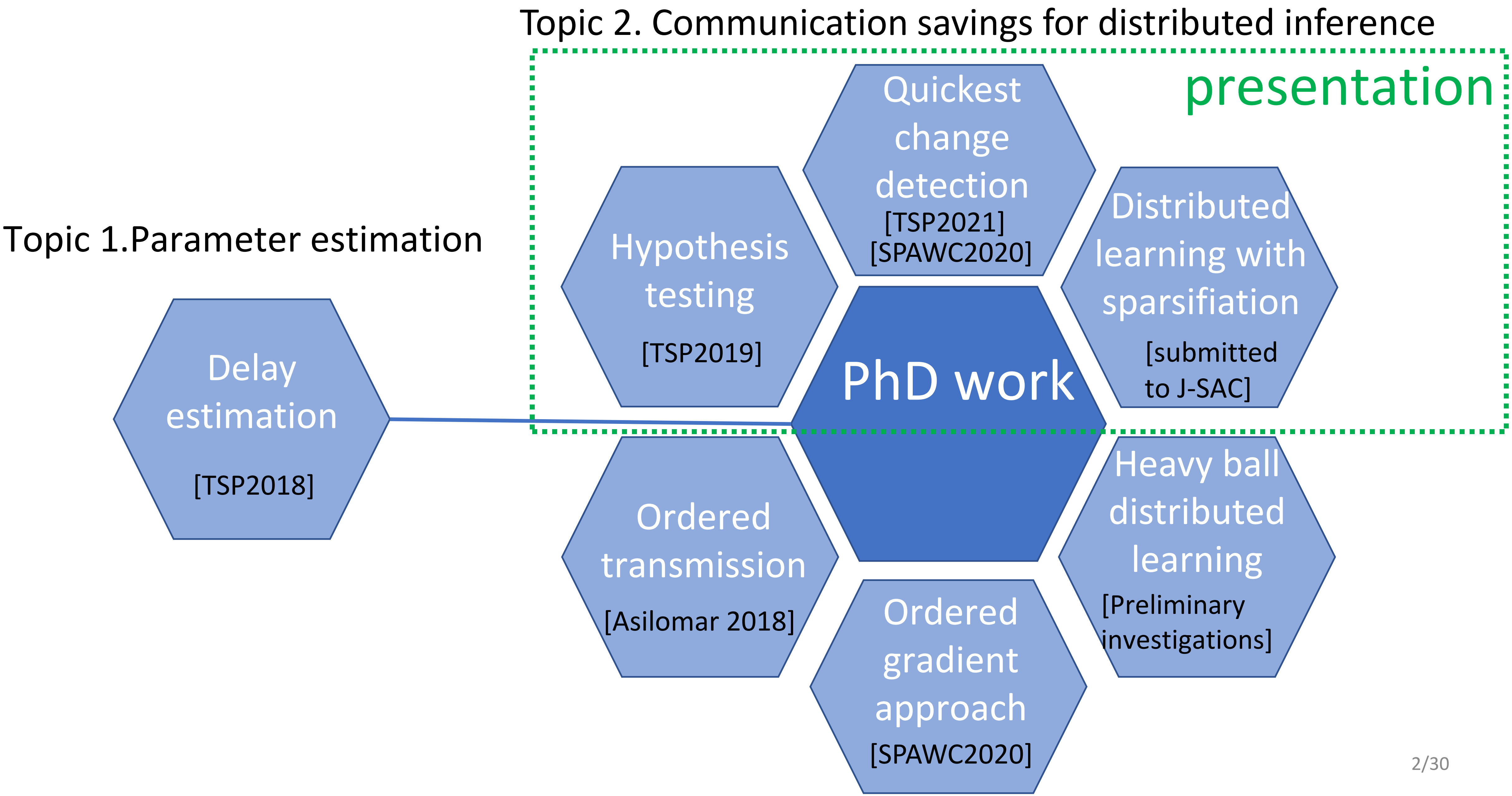
Ordered Transmissions, Estimation, and Parameter Learning

Yicheng Chen

Advisor: Prof. Rick S. Blum

PhD Defense
April 2021

➢ PhD Work Summary



- Communication savings for distributed inference

    - ➢ Testing the Structure of a Gaussian Graphical Model

First work on saving communications without loss in a covariance change detection problem

    - ➢ Optimal Quickest Change Detection in Sensor Networks

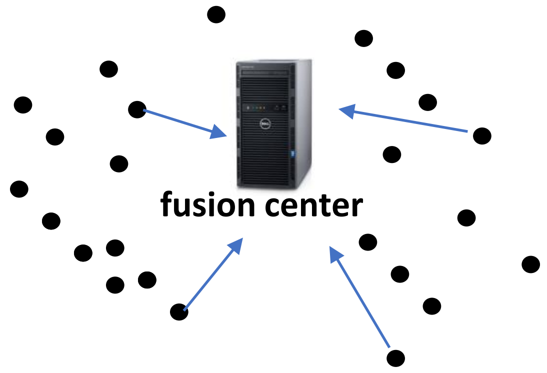
    - ➢ Distributed Learning with Sparsified Gradient Differences

- Conclusion

Y. Chen, R. S. Blum, B. M. Sadler, and J. Zhang, "Testing the structure of a gaussian graphical model with reduced transmissions in a distributed setting," IEEE Transactions on Signal Processing (TSP), vol. 67, no. 20, pp. 5391–5401, 2019.111

LEHIGH UNIVERSITY

Problem formulation



- Gaussian covariance matrix testing problem

$$H_0 : \mathbf{x} \sim \mathcal{N}(0, \mathbf{I})$$
$$H_1 : \mathbf{x} \sim \mathcal{N}(0, \mathbf{\Sigma}) \quad (1)$$

Sensor data

- (1) is very general

$$H_0 : \mathbf{x} \sim \mathcal{N}(0, \mathbf{\Sigma_0}) \qquad \Longrightarrow \qquad H_0 : \mathbf{x} \sim \mathcal{N}(0, \mathbf{I})$$
$$H_1 : \mathbf{x} \sim \mathcal{N}(0, \mathbf{\Sigma_1}) \qquad\qquad\quad H_1 : \mathbf{x} \sim \mathcal{N}(0, \mathbf{\Sigma})$$

known ($\mathbf{\Sigma_0}$), known ($\mathbf{\Sigma_1}$)

- Goal: solve (1) in a distributed setting with communication savings

LEHIGH UNIVERSITY

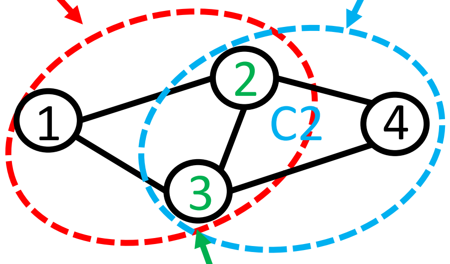
clustering

threshold

$$T(\mathbf{x}) = \mathbf{x}^{\mathbf{T}}\mathbf{I}^{-1}\mathbf{x} - \underbrace{\mathbf{x}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}} - \ln\det(\boldsymbol{\Sigma}) \gtrless \tau \qquad (1)$$

log-likelihood ratio

involve all sensor data



- In decomposable GGMs,

$$\boldsymbol{\Sigma}^{-1} = \sum_{k=1}^{K}[(\boldsymbol{\Sigma}_{\mathcal{C}_k})^{-1}]^{\mathcal{V}} - \sum_{k=2}^{K}[(\boldsymbol{\Sigma}_{\mathcal{S}_k})^{-1}]^{\mathcal{V}}$$

Example:

$$\boldsymbol{\Sigma}^{-1} = [(\boldsymbol{\Sigma}_{\mathcal{C}_1})^{-1}]^{\mathcal{V}} + [(\boldsymbol{\Sigma}_{\mathcal{C}_2})^{-1}]^{\mathcal{V}} - [(\boldsymbol{\Sigma}_{\mathcal{S}_2})^{-1}]^{\mathcal{V}}$$

$$= \underbrace{\left([(\boldsymbol{\Sigma}_{\mathcal{C}_1})^{-1}]^{\mathcal{V}} - \beta[(\boldsymbol{\Sigma}_{\mathcal{S}_2})^{-1}]^{\mathcal{V}}\right)}_{\text{compute at Cluster 1}} + \underbrace{\left([(\boldsymbol{\Sigma}_{\mathcal{C}_2})^{-1}]^{\mathcal{V}} - (1-\beta)[(\boldsymbol{\Sigma}_{\mathcal{S}_2})^{-1}]^{\mathcal{V}}\right)}_{\text{compute at Cluster 2}} \qquad (2)$$

group the separator terms into the associated cluster terms

- Plugging (2) into (1), $T(\mathbf{x}) = L_1(\mathbf{x}_{\mathcal{C}_1}) + \underbrace{L_2(\mathbf{x}_{\mathcal{C}_2})}$

Compute based only on cluster data

- **Distributed processing**: $T(\mathbf{x}) = \sum_{k=1}^{K} L_k(\mathbf{x}_{\mathcal{C}_k})$

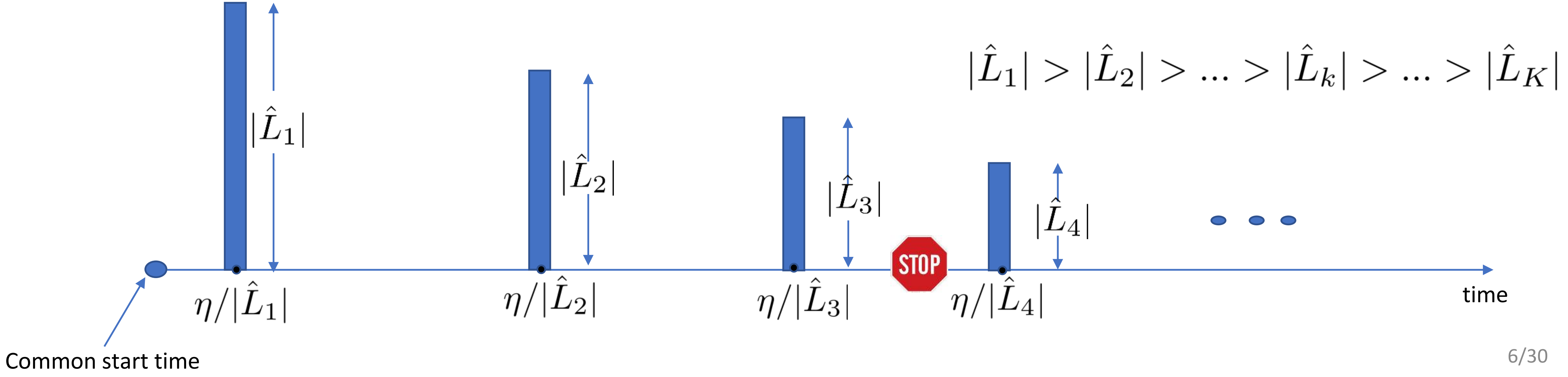
  Log-likelihood ratio

  Local test statistic (LTS): compute at cluster k

- **Ordered transmission**: each cluster decides when to transmit its LTS
  - most informative transmit first

    implement

  - each cluster k transmits LTS $L_k(\mathbf{x}_{\mathcal{C}_k})$ after $\eta/|L_k(\mathbf{x}_{\mathcal{C}_k})|$ seconds



Making optimal decision without seeing all data

$$\left|\hat{L}_1\right| \geq \left|\hat{L}_2\right| \geq \dots \geq \left|\hat{L}_K\right|$$

- FC decides H_1 when

unconstrained opt. threshold — most recent (k'th) cluster transmission

$$\underbrace{\sum_{k=1}^{k'} \hat{L}_k}_{\text{the first } k' \text{ transmissions}} > t_U = \tau + \underbrace{(K - k')\left|\hat{L}_{k'}\right|}_{\text{largest possible contribution from remaining clusters}} \qquad (1)$$

Once (1) is true, sum will be larger than τ, regardless of the data not transmitted.

Without transmitting further, can implement the optimum approach

decides H_0 if

$$\sum_{k=1}^{k'} \hat{L}_k < t_L = \tau - (K - k')\left|\hat{L}_{k'}\right| \qquad (2)$$

LEHIGH UNIVERSITY

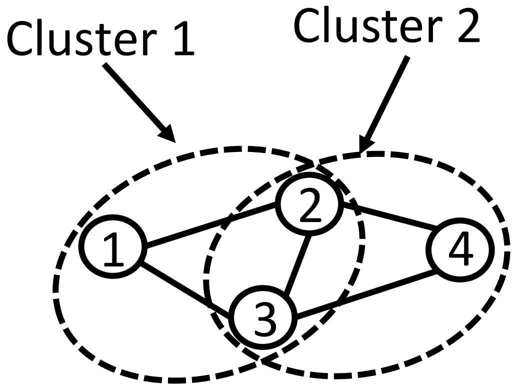
Save energy, but how large are savings?

If $\boldsymbol{\Sigma}_{\mathcal{C}_k}$ is very different from $\mathbf{I}$, easy to detect and can save more!

evaluate difference

- Define $\lambda_{\min} = \min_k \lambda_{\min,k}$

minimum eigenvalue of $\boldsymbol{\Sigma}_{\mathcal{C}_k}$



Theorem: For a sufficiently large λ_{min}, the average number of transmissions saved is strictly larger than $\lceil K/2 \rceil - 1$.

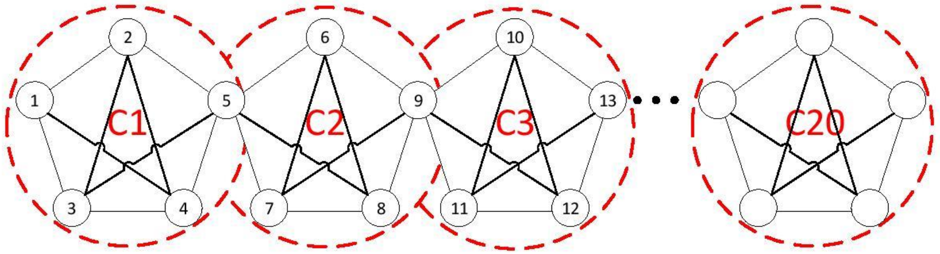
number of clusters

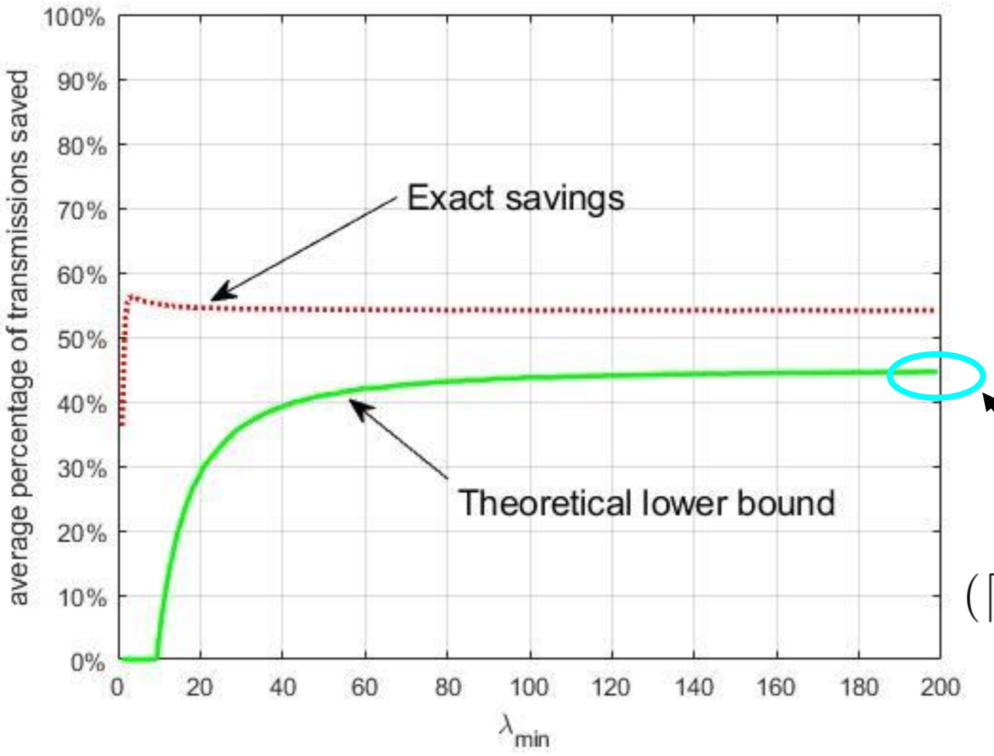
LEHIGH UNIVERSITY

Numerical results

- decomposable GGM with chain structure

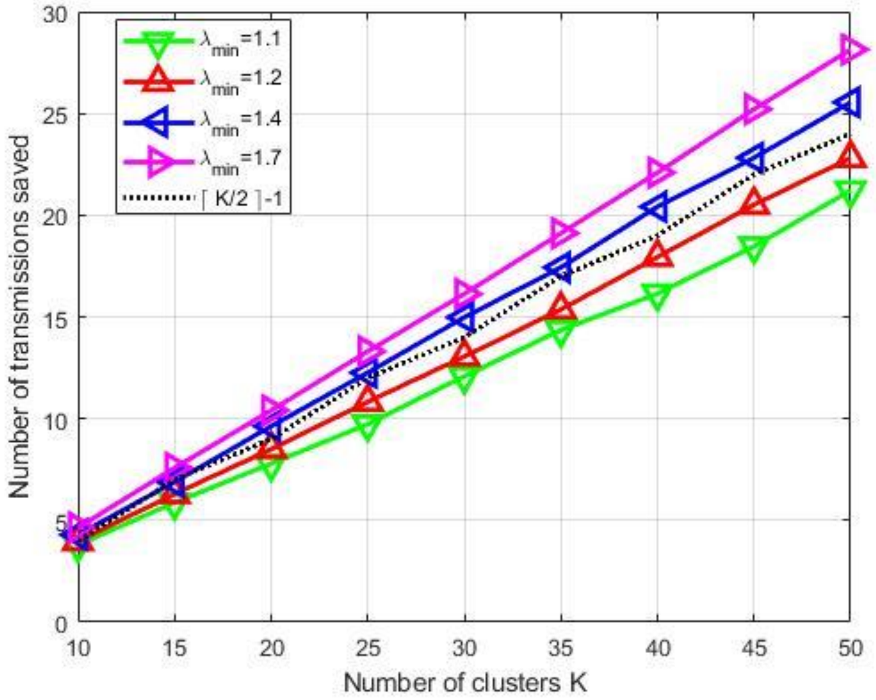
$\lambda_{\min}$ is the minimum eigenvalue of $\mathbf{\Sigma}_{\mathcal{C}_k}$





$(\lceil K/2 \rceil - 1)/K = 0.45$

a valid lower bound



$(\lceil K/2 \rceil - 1)$ serves as a lower bound as $\lambda_{\min} \geq 1.4$

- Communication savings for distributed inference

  - ➢ Testing the Structure of a Gaussian Graphical Model
  - ➢ Optimal Quickest Change Detection in Sensor Networks

    First work on comm. efficient QCD without loss

  - ➢ Distributed Learning with Sparsified Gradient Differences

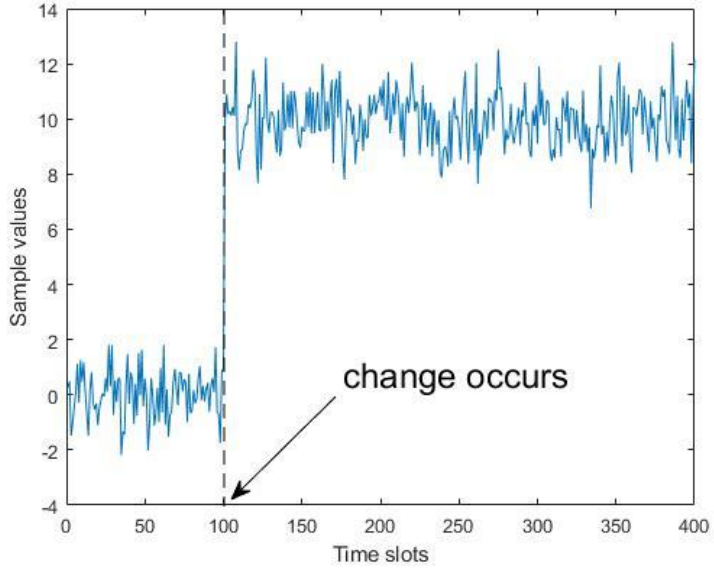
- Conclusion

Yicheng Chen, Rick S Blum, and Brian M Sadler. Ordering for communication-efficient quickest change detection in a decomposable graphical model. IEEE Transactions on Signal Processing, 2021, accepted.

LEHIGH UNIVERSITY

Results preview

$f_0 = \mathcal{N}(0,1) \qquad f_1 = \mathcal{N}(10,1)$



Quickest change detection (QCD) problem in sensor networks

- detect the change in distribution asap

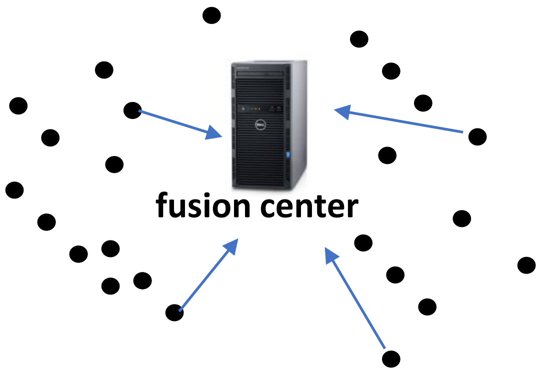
Censoring [1] (only transmit highly informative data)

- save communications
- always lose detection performance

Our contributions:

a new communication-efficient approach:

- achieve the optimal performance (no loss)
- require fewer sensor transmissions
- extensions for dependent observations



[1]. T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in sensor networks," IEEE Transactions on Signal Processing, vol. 63, no. 14, pp. 3727–3735, 2015.

Problem formulation

time slot

sensor index

Assumption 1: observations $X_{n,k}$ are independent.

Assumption 2: distributions at all sensors change simultaneously at time $n = \tau$.

QCD problem is formulated as:

worst case average detection delay

$$\min_{n} \mathrm{WADD}(n)$$
$$s.t.\ \mathbb{E}_{\infty}(n) \geq \gamma.$$

average delay when no change

pre-specified constant

Generalized likelihood ratio test (GLRT) method

QCD can be modeled as a hypothesis testing problem

$$H_0 : \text{no change occurs}$$
$$H_1 : \text{change occurs at some finite time slot } \tau.$$

GLRT procedure (MLE τ) is also called the CUSUM algorithm

Declare H_1 when $\inf\{n \geq 1 : W_n \geq b\}$

CUSUM statistic W_n can be computed recursively as

pdf after change

$$W_n = \max\left\{0, \quad W_{n-1} + \sum_{k=1}^{K} \log \frac{f_1(X_{n,k})}{f_0(X_{n,k})}\right\} \quad \text{with } W_0 = 0. \qquad (1)$$

pdf before change

CUSUM with $W_0 = 0$ implies:

$$\mathrm{WADD}(n) = \mathbb{E}_1[n-1]$$

$\tau = 1$

Ordered-CUSUM transmission

- Classical CUSUM algorithm:

$$W_n = \max\left\{0, \quad W_{n-1} + \sum_{k=1}^{K} \log \frac{f_1\left(X_{n,k}\right)}{f_0\left(X_{n,k}\right)}\right\}$$

$\log L_{n,k}$

- Ordered-CUSUM transmission: each sensor decides when to transmit its LLR
  - most informative transmit first

$$\left|\log \hat{L}_{n,1}\right| \geq \left|\log \hat{L}_{n,2}\right| \geq \ldots \geq \left|\log \hat{L}_{n,K}\right|$$

implement

  - At each time slot n, each sensor k transmits LLR $\log L_{n,k}$ after $\eta / \left|\log L_{n,k}\right|$ seconds

- Claim: Ordered-CUSUM can make optimal decisions without seeing all transmissions

Making optimal decision without seeing all data

- at time slot n, know $W_n = 0$ when

most recent (k'th) sensor transmission

most recent (k'th) sensor transmission

$$W_{n-1} + \underbrace{\sum_{k=1}^{k'} \log \hat{L}_{n,k}}_{\text{the first } k' \text{ transmissions}} < \underbrace{0 - (K - k') \left| \log \hat{L}_{n,k'} \right|}_{\text{largest possible contribution from remaining sensors}} \quad (3)$$

the first k' transmissions

largest possible contribution from remaining sensors

$$W_n = \max \left\{ 0, \quad W_{n-1} + \sum_{k=1}^{K} \log \frac{f_1(X_{n,k})}{f_0(X_{n,k})} \right\}$$

Once (3) is true, W_n will be zero, regardless of the data not transmitted.

Without transmitting further, can implement the optimum approach.

Ordered-CUSUM transmission

Save energy, but how large are savings?

Let s be a distance measure between two distributions f_0 (before change) and f_1 (after change) such that

as $s \to \infty$ for all k, $\Pr\left(\log L_{n,k} < 0 \middle| n < \tau\right) \to 1$ and $\Pr\left(\log L_{n,k} > 0 \middle| n \geq \tau\right) \to 1$.

(τ: real change time)

Theorem: For a sufficiently large s, the average number of transmissions saved increases at least as fast as proportional to K.

LEHIGH UNIVERSITY

Performance comparison with/without ordering

Example: $f_0 = \mathcal{N}(0,1) \xrightarrow{\text{at time slot } \tau = 1} f_1 = \mathcal{N}(s,1)$

false alarm constraint $\mathbb{E}_\infty(n) \geq \gamma = 1000$

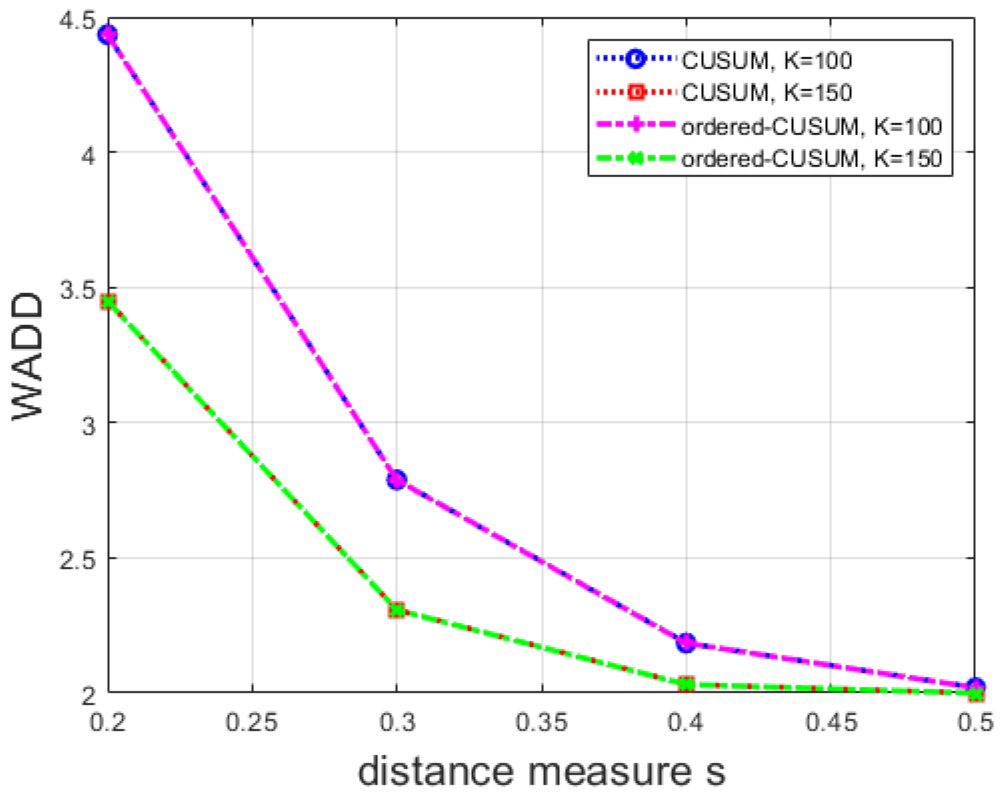


Fig. 1: The worst case average detection delay versus the distance measure for K = 100 and K = 150.

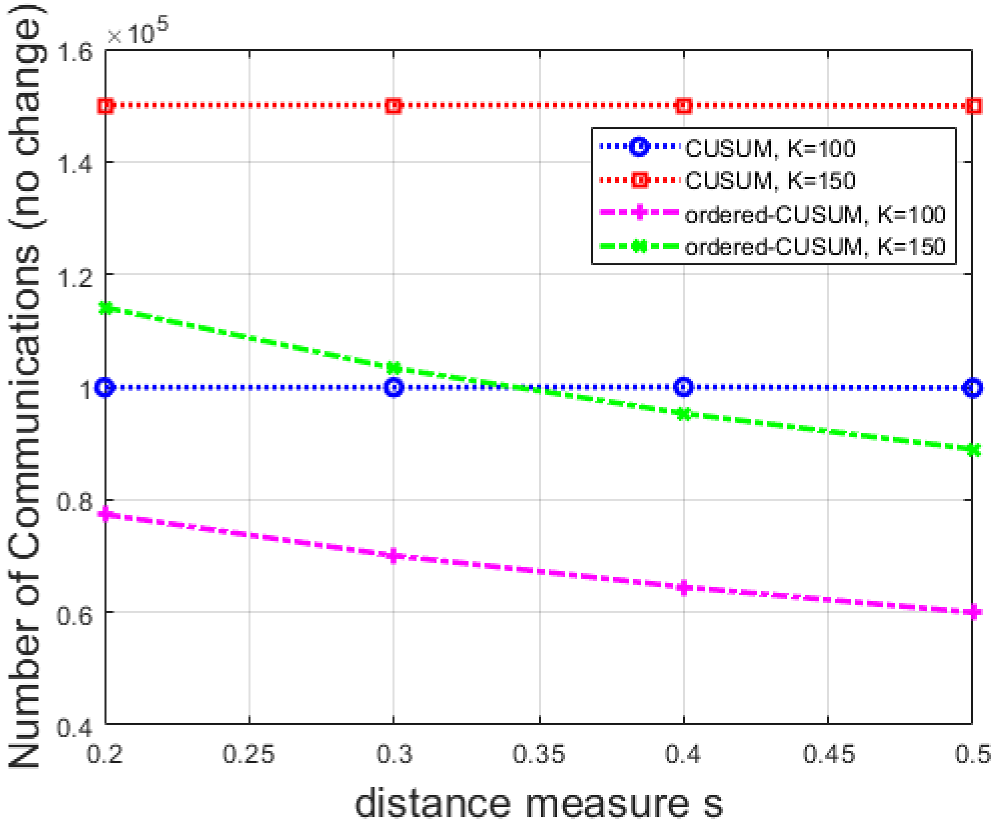
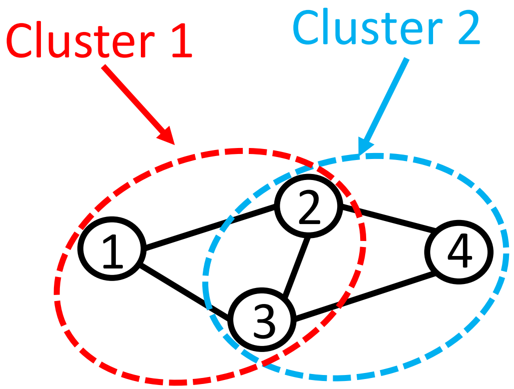


Fig. 2: Number of communications when the change does not occur versus the distance measure for K = 100 and K = 150.

How to handle dependent data in QCD?

Ideas:

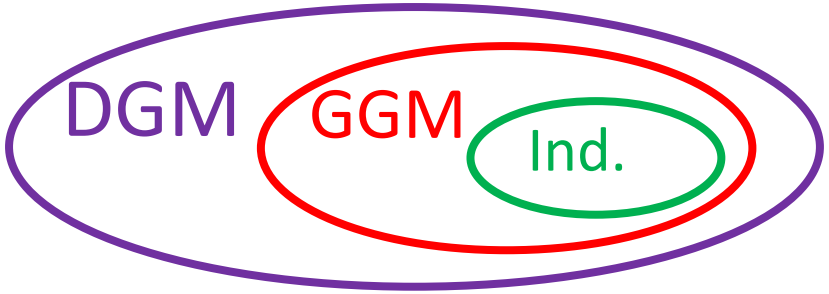
- Step1: Group nodes into cliques/clusters
- Step2: Perform ordering over clusters



Generalize it to nonGaussian case → Decomposable graphical model (DGM)

In the paper, we also consider:

- Graph structure changes
- A nondecomposable graphical model



- Communication savings for distributed inference

  - ➢ Testing the Structure of a Gaussian Graphical Model

  - ➢ Optimal Quickest Change Detection in Sensor Networks

  - ➢ Distributed Learning with Sparsified Gradient Differences

Very little work on comm. efficient optimization with **adaptive sparsification**

- Conclusion

Yicheng Chen, Rick S Blum, Martin Takáč, and Brian M Sadler. Distributed learning with sparsfied gradient differences. Submitted to IEEE Journal on Selected Areas in Communications.

Problem formulation

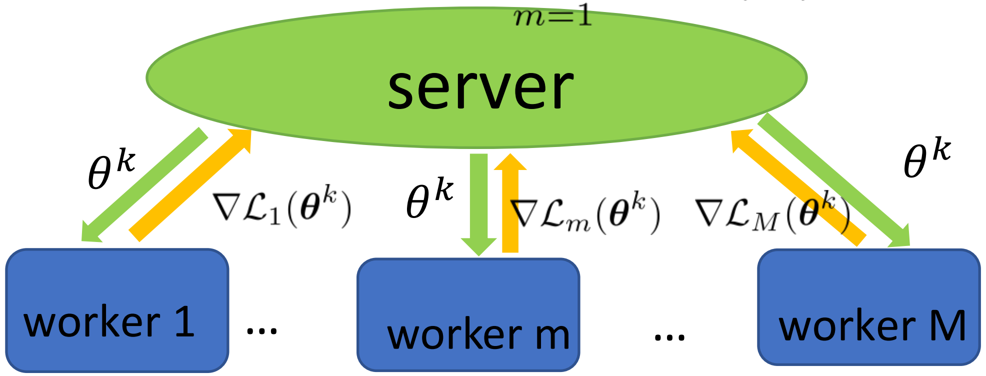
- Formulate the problem as

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{\theta}) \quad \text{with} \quad \mathcal{L}(\boldsymbol{\theta}) \triangleq \sum_{m=1}^{M} \mathcal{L}_m(\boldsymbol{\theta}) \tag{1}$$

- In distributed learning, $\mathcal{L}_m(\boldsymbol{\theta}) \triangleq \sum_{n=1}^{N_m} \ell(\boldsymbol{\theta}; \boldsymbol{x}_n, y_n)$ at worker m is the sum of the loss functions $\ell(\boldsymbol{\theta}; \boldsymbol{x}_n, y_n)$ (feature vector: $\boldsymbol{x}_n$; label: y_n)

- Distributed Gradient descent (GD) method [1]

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \alpha \sum_{m=1}^{M} \nabla \mathcal{L}_m\left(\boldsymbol{\theta}^k\right)$$



$$\text{size}(\boldsymbol{\theta}^k) \approx 200MB$$

Our goal is to find an optimal solution of (1) using limited number of communication bits.

[1]. J. Nocedal and S. Wright, Numerical Optimization, 2nd ed. New York, NY, USA: Springer-Verlag, 2006.

LEHIGH UNIVERSITY

Adaptive Sparsified GD (ASGD) implementation

- Each worker m maintains current gradient $\nabla\mathcal{L}_m(\boldsymbol{\theta}^k)$ and transmitted history $\boldsymbol{h}_m^k$

a linear combination of its past transmissions

locally smooth

- Worker m transmits the sparsified gradient difference
- Not allowed to transmit a given component of gradient difference if its magnitude is not sufficiently large

- Define the difference vector

$$\delta\nabla_m^k \triangleq \nabla\mathcal{L}_m(\boldsymbol{\theta}^k) - \boldsymbol{h}_m^k + \boldsymbol{e}_m^k$$

error correction compensation

- **update rules**
  - Server updates $\boldsymbol{\theta}^k$ via

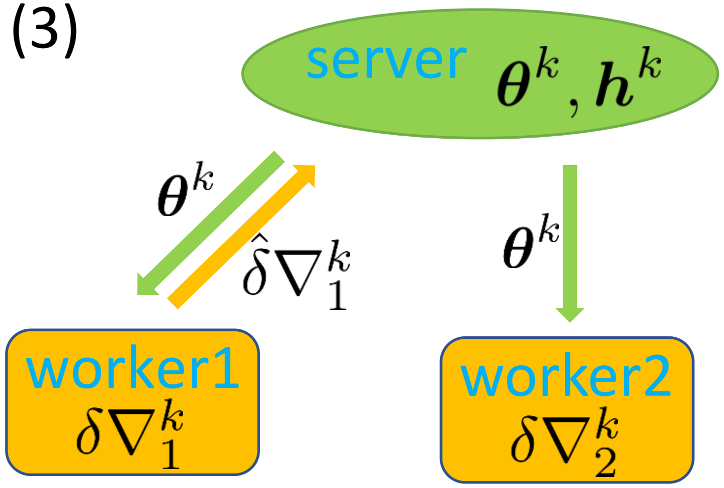
$\delta\nabla_m^k$ → sparsifed → $\hat{\delta}\nabla_m^k$

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \alpha\left(\boldsymbol{h}^k + \sum_{m=1}^{M} \hat{\delta}\nabla_m^k\right) \tag{3}$$

past gradient

  - Server updates $\boldsymbol{h}^k$ via

$$\boldsymbol{h}^{k+1} = \boldsymbol{h}^k + \beta \sum_{m=1}^{M} \hat{\delta}\nabla_m^k$$



$$\delta\nabla_m^k \triangleq \nabla\mathcal{L}_m(\boldsymbol{\theta}^k) - \boldsymbol{h}_m^k + \boldsymbol{e}_m^k$$

  - Each worker m updates $\boldsymbol{h}_m^k$ and $\boldsymbol{e}_m^k$ via

smooth

$$\boldsymbol{h}_m^{k+1} = \boldsymbol{h}_m^k + \beta\hat{\delta}\nabla_m^k$$
$$\boldsymbol{e}_m^{k+1} = \delta\nabla_m^k - \hat{\delta}\nabla_m^k$$

How to sparsify?

LEHIGH UNIVERSITY

- sparsification condition

intuition

Do not transmit if the change is small

- Worker m does not transmit the i-th component of $\delta\nabla_m^k \triangleq \nabla\mathcal{L}_m(\boldsymbol{\theta}^k) - \boldsymbol{h}_m^k + \boldsymbol{e}_m^k$

if

$$|[\delta\nabla_m^k]_i| \leq \underbrace{\frac{\xi_i}{M}|[\boldsymbol{\theta}^k - \boldsymbol{\theta}^{k-1}]_i|}_{\text{previous change (adaptive)}} \qquad (4)$$

- **Advantages of ASGD** :
  - Each worker adaptively decides which component to transmit
  - Reduce the number of bits

Convergence guarantee

Recall

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \mathcal{L}(\boldsymbol{\theta}) \quad \text{with} \quad \mathcal{L}(\boldsymbol{\theta}) \triangleq \sum_{m=1}^{M} \mathcal{L}_m(\boldsymbol{\theta})$$

Assumption 1: each local function $\mathcal{L}_m(\boldsymbol{\theta})$ is coordinate-wise smooth

Three theorems: Under Assumption 1, if we choose $\alpha,\ \beta,\ \{\xi_i\}_{i=1}^{d}$ properly, then ASGD generates K iterates satisfying

strongly convex	$\mathcal{L}(\boldsymbol{\theta}^K) - \mathcal{L}(\boldsymbol{\theta}^*) = \mathcal{O}((1-c)^K)$
convex	$\mathcal{L}(\boldsymbol{\theta}^K) - \mathcal{L}(\boldsymbol{\theta}^*) = \mathcal{O}(1/K)$
nonconvex	$\lim_{K\to\infty} \|\nabla\mathcal{L}(\boldsymbol{\theta}^K)\|^2 \to 0$

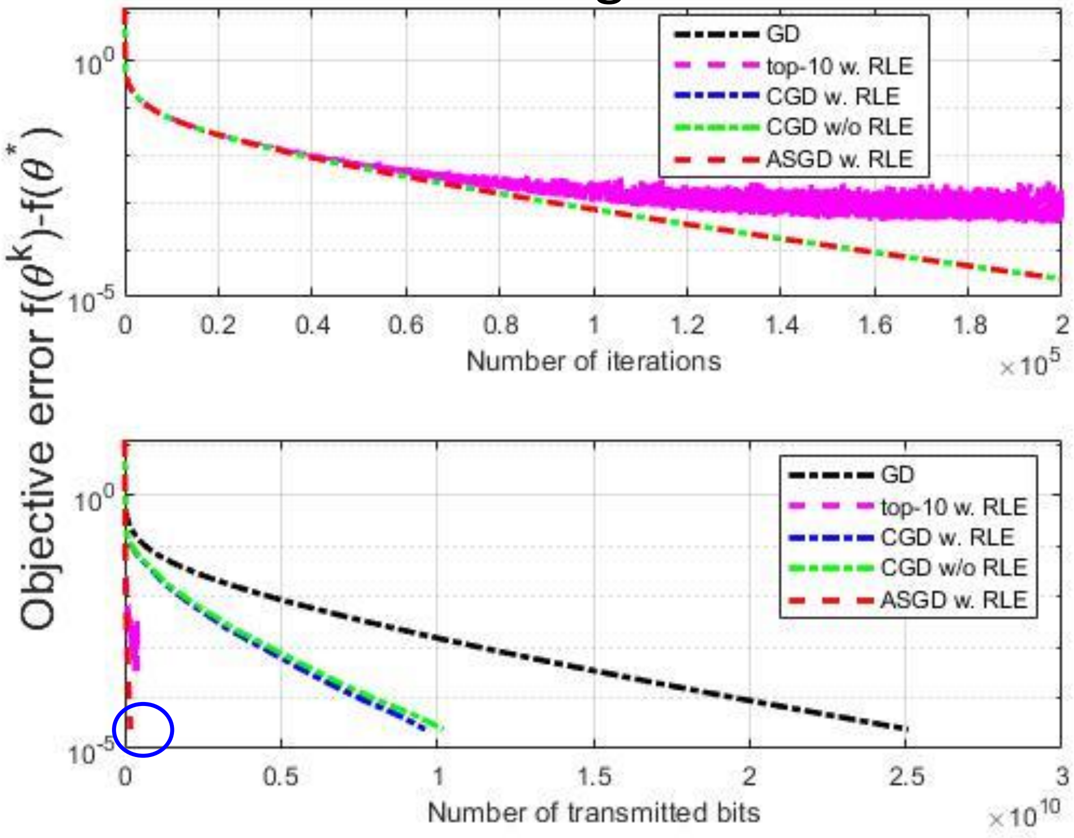
LEHIGH UNIVERSITY

➢ Numerical results

- One central server, and M = 5 distributed workers
- Benchmarks: GD, top-10, censoring-based GD [Chen, Tianyi, et al. 2018]
- Run-length encoding (RLE)

MNIST dataset

Linear regression

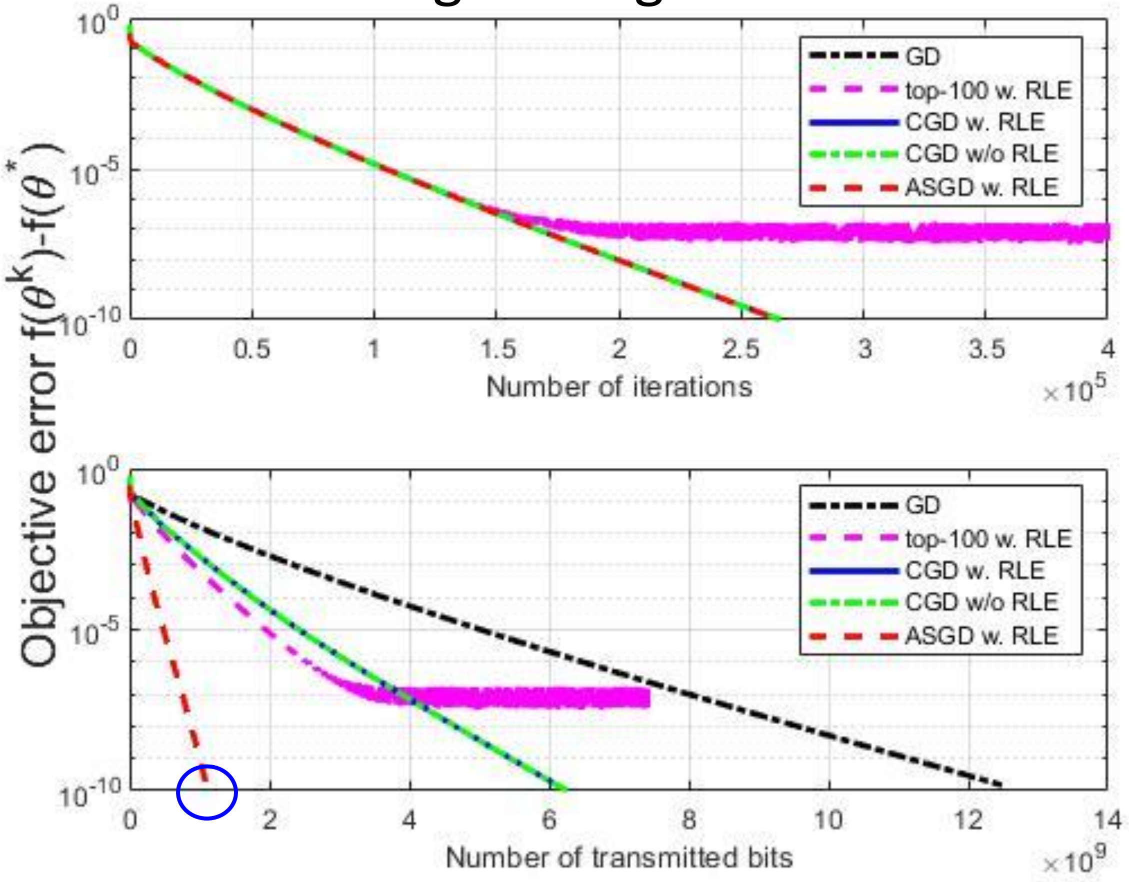


How about other tasks?

Numerical results

Synthetic dataset

Logistic regression



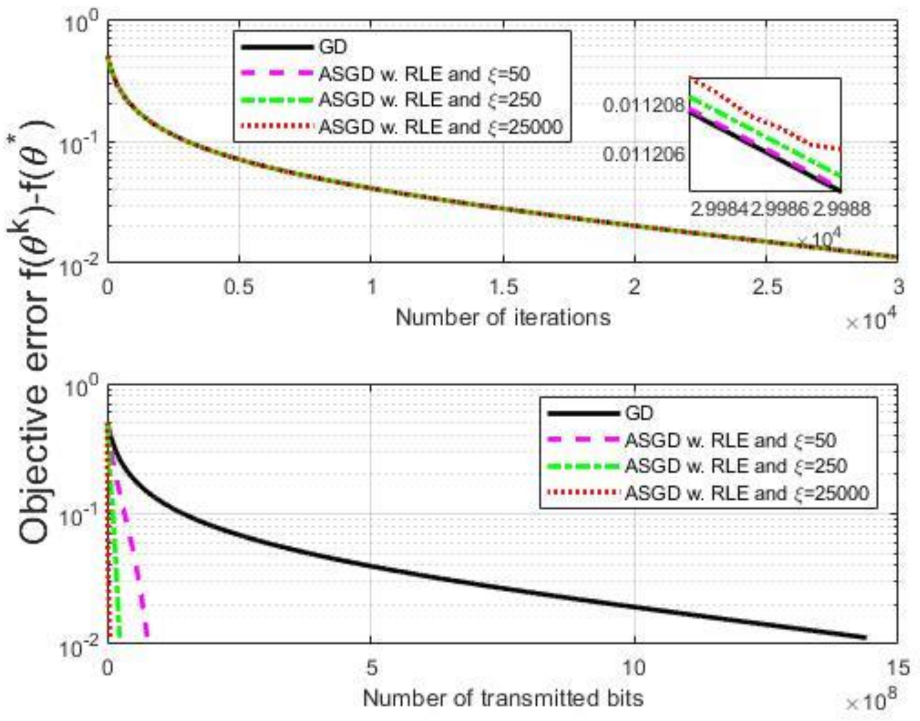
Compared to GD, ASGD can significantly reduce the total transmitted bits.

LEHIGH UNIVERSITY

➢ Numerical results

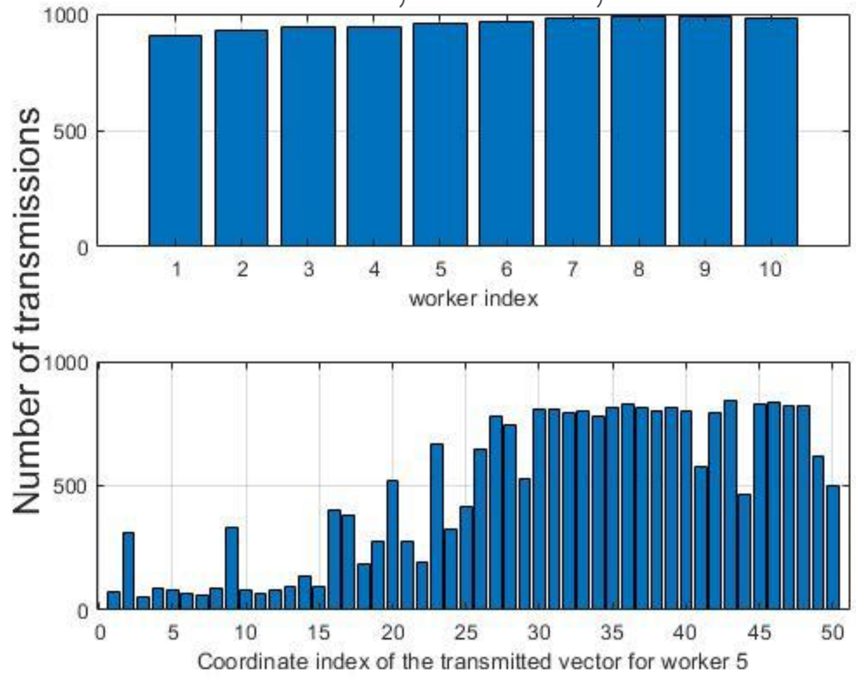
- One central server, and M = 10 distributed workers

● Nonlinear least square (nonconvex)



● linear regression

$$L_1 < L_2 < \ldots < L_{10}$$
$$L_{1,1} < L_{1,2} < \ldots < L_{1,50}$$



ASGD significantly reduces the total transmitted bits for a nonconvex case

Lipschitz constant is smaller (≈small change), the number of transmissions is less

- Communication savings for distributed inference
  - Testing the Structure of a Gaussian Graphical Model
  - Optimal Quickest Change Detection in Sensor Networks
  - Distributed Learning with Sparsified Gradient Differences
- Conclusion

- Jointly optimize communications, decision-making, and distributed learning for battery powered devices

- Open questions:
  - Is the ordered transmission approach optimal?
  - More exact categorization of ordering
  - Stochastic generalizations of distributed learning

LEHIGH UNIVERSITY

Thank you

Questions ?