Privacy preservation for linear and learning based inference systems

Wang, Chong Xiao

2021

Wang, C. X. (2021). Privacy preservation for linear and learning based inference systems.
Doctoral thesis, Nanyang Technological University, Singapore.
<https://hdl.handle.net/10356/153057>

<https://hdl.handle.net/10356/153057>

<https://doi.org/10.32657/10356/153057>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0
International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 10:21:53 SGT

Privacy Preservation for Linear and Learning Based Inference Systems

Wang Chong Xiao

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

Feb. 21, 2021

.....

Date

.....

Wang Chong Xiao

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

Feb. 21, 2021

.....

Date

.....

Assoc. Prof. Tay Wee Peng

Authorship Attribution Statement

This thesis contains material from 2 papers published, 1 paper submitted in the following peer-reviewed journals in which I am listed as an author.

Chapter 3 is published as **C. X. Wang**, Y. Song, W. P. Tay, “Arbitrarily strong utility-privacy tradeoff in multi-agent systems,” IEEE Trans. Inf. Forensics Security, pp. 1-1, 2020.

The contributions of the co-authors are as follows:

- I proposed the idea, designed the algorithms, provided the theoretical results and prepared the manuscript drafts.
- Prof. Tay Wee Peng provided guidance and discussed the problem formulation and theoretical results with me.
- Dr. Song Yang reviewed the manuscript and Prof. Tay Wee Peng revised the manuscript.

Chapter 4 is published as Y. Song, **C. X. Wang**, and W. P. Tay, “Compressive privacy for a linear dynamical system,” IEEE Trans. Inf. Forensics Security, vol. 15, no. 1, pp. 895-910, Dec. 2020.

The contributions of the co-authors are as follows:

- I proposed the idea. Dr. Song Yang designed the algorithms and prepared the manuscript drafts.
- Prof. Tay Wee Peng provided guidance and discussed the problem formulation and algorithms with Dr. Song Yang and me.
- Prof. Tay Wee Peng revised the manuscript.

Chapter 5 is submitted as **C. X. Wang**, W. P. Tay, “Data-driven regularized inference privacy,” to IEEE Trans. Dependable Secure Comput.

The contributions of the co-authors are as follows:

- I designed the algorithms, provided the theoretical results and prepared the manuscript drafts.
- Prof. Tay Wee Peng provided guidance and discussed the problem formulation and algorithms with me.
- Prof. Tay Wee Peng revised the manuscript.

Feb. 21, 2021

.....

Date

.....

Wang Chong Xiao

Acknowledgements

Undertaking this PhD has been a tremendously rewarding experience for me. This journey would not have been possible to do without the support and guidance that I received from many people.

First and foremost, I am extremely grateful to my supervisor Prof. Tay Wee Peng for giving me the opportunity to work in his group. I received invaluable advice and guidance from the regular meetings with him where I was taught to analyze problems using mathematical thinking and was provided with fresh perspectives. I am thankful for his efforts on reviewing my works with patience and providing precious feedbacks. I appreciate for the group study session he organized despite his busy schedule, with special thanks to the tutorials on probability theory and real analysis. His rigorous attitude towards research and upholding high standards of work have influenced me a lot and would benefit me all my life.

I would also like to thank the members of Distributed and Adaptive algorithms in Sensor Networks (DASN) group. My sincere appreciation goes to Prof. Zhong Xionghu, Dr. Song Yang, Dr. Ji Feng, Prof. Yan Yong Sheng, Dr. Xu Yi, Dr. Wang Cheng Cheng, Dr. Jin Cheng Ming, Dr. Bajaj Ian, Dr. Sun Meng, Dr. Kang Qiyu, Dr. Yang Jielong for the insightful discussions and valuable inputs. I would like to thank all the friends I met in NTU for making this journey colorful and memorable.

Last but not the least, I would like to thank my parents who have always been giving me unconditional love and support even though we disagree on many things.

"If I had only one hour to save the world, I would spend 55 minutes defining the problem and only 5 minutes finding the solution."

—Einstein, Albert

To my dear family

Abstract

In the activities of data sharing and decentralized processing, data belonging to a user need to be transmitted to a third-party data processor or aggregator. However, the third-party unit may be curious about some private information hidden in the unprocessed data sent by the user, which poses tremendous threat to the data owner's privacy. Therefore, it becomes imperative for the data owner to sanitize its raw data to limit the amount of private information before disclosing it. In the meantime, the data owner hopes to maximally preserve the utility of the sanitized data in terms of a legitimate task that she wishes to perform on the third-party unit. In this thesis, we investigate the problem of protecting parameter estimation in a decentralized linear system and a linear dynamical system with known system models. Apart from that, we explore data-driven approaches to protect the private information carried by raw data that serve as an input to a learning-based inference system.

In the first place, we consider a multi-agent system where each agent in a network makes a local observation that is linearly related to a set of public and private parameters. The agents send their observations to a fusion center to allow it to estimate the public parameters. To prevent leakage of the private parameters, each agent first sanitizes its local observation using a local privacy mechanism before transmitting it to the fusion center. We study the utility-privacy tradeoff in terms of the Cramér-Rao lower bounds for estimating the public and private parameters, and compare the class of privacy mechanisms given by linear compression and noise perturbation. We derive necessary and sufficient conditions for achieving arbitrarily strong privacy without compromising utility, and provide a method to maximize privacy while keeping utility intact. Finally, we propose an algorithm to optimize the utility-privacy tradeoff for satisfying a maximum allowable privacy leakage.

As a further development of the first case, we consider a linear dynamical system in which the state vector is composed of public and private parts and progresses

according to a known transition model. Multiple sensors make a series of measurements of the state vector and send the data to a fusion center who is authorized to perform state estimation for the public states. To prevent the fusion center from estimating the private states with good accuracy, each sensor's measurements at each time step are linearly compressed into a lower dimensional space before being sent to the fusion center. We take account of the fact that the public states at one time step may provide statistical information about the private states in a future time step, and propose a formulation that can ensure the same level of privacy in all future time steps. We develop online algorithms to find the optimal compression matrix for the utility-privacy tradeoff.

Finally, we consider a practical setup where a precise system model cannot be derived analytically due to the complexity of the underlying data. Thus, we need to learn from data to capture the essence of reality. Data are used by service providers as input to inference systems to perform decision making for authorized tasks. The raw data however allows a service provider to infer other sensitive information it has not been authorized for. We explore data driven approaches to sanitize data so as to prevent leakage of sensitive information that is present in the raw data. We deploy mutual information and maximal correlation as privacy leakage measures and reveal that their practical implementation requires different configurations. We present empirical estimators of the privacy metrics, for which asymptotic analysis is carried out. We propose to regularize the domain of the sanitized data to ensure that the sanitized data is still compatible with the service provider's legacy inference system. We develop a deep learning model as an example of the proposed privacy framework.

Contents

Acknowledgements	ix
Abstract	xiii
List of Figures	xix
List of Tables	xxi
Symbols and Acronyms	xxiii
1 Introduction	1
1.1 Background	1
1.2 Motivation	4
1.3 Related Work	7
1.3.1 Inference Privacy, Data Privacy and Data Security	7
1.3.2 Information Theoretic Privacy	10
1.3.3 Learning Based Privacy	12
1.4 Thesis Contribution and Outline	14
2 Preliminaries	17
2.1 Cramér–Rao Lower Bound	17
2.2 Kalman Filter	19
2.3 Information Theory	21
2.4 Kernel Methods	22
3 Privacy-Preserving Decentralized Linear Estimation	25
3.1 Problem formulation	26
3.2 Privacy-Preserving Transformation	30
3.3 Arbitrarily Strong Privacy with Perfect Utility	36
3.3.1 ASUP without prior information	37
3.3.2 ASUP with prior information	41
3.4 Maximum Privacy Under Perfect Utility	44
3.5 Alternating Optimization	46
3.6 Numerical results	49

3.6.1	Maximum privacy under perfect utility	50
3.6.2	Alternating optimization	52
3.7	Conclusion	52
3.A	Derivation of (3.20), (3.23) and (3.24)	54
3.B	Proof of Lemma 2	55
3.C	Proof of Lemma 3	56
3.D	Derivation of (3.62) and (3.63)	57
4	Privacy Preservation in Linear Dynamical System	59
4.1	Problem Formulation	60
4.2	A Single Sensor	69
4.3	Decentralized Sensors	73
4.3.1	Sequential message broadcasts	75
4.4	Numerical Results	79
4.4.1	Simulation for decentralized sensors	79
4.4.2	Privacy-Aware localization	81
4.5	Comparisons with Affine Transformation	83
4.5.1	Affine Transformation	84
4.5.2	Simulation Results	87
4.6	Conclusion	89
4.A	Proof of Lemma 5	89
4.B	Proof of Lemma 7	89
4.C	Derivation of (4.48)	90
5	Learning-Based Privacy Protection	93
5.1	Motivations	94
5.2	Problem Formulation	96
5.3	Utility Metric and Regularization	99
5.3.1	Utility Function	100
5.3.2	Domain Regularization	102
5.3.2.1	Maximum mean discrepancy (MMD)	103
5.3.2.2	Domain adaptation	104
5.4	Privacy Metrics	105
5.4.1	Variational Method: DRIP-Var	105
5.4.2	Maximal Correlation: DRIP-Max	109
5.4.3	Convergence Analysis	114
5.4.4	Average Information Leakage	114
5.4.5	Maximal Correlation	116
5.5	DRIP Framework Optimization	119
5.5.1	Overall DRIP Framework	119
5.5.2	Alternating Optimization Algorithm	120
5.6	Deep Learning Architecture for Images	121
5.6.1	Sanitizer	121

5.6.2	Utility Function $\mathcal{L}(\phi; \theta)$	122
5.6.3	Privacy Function $\mathcal{P}(\tau; \theta)$	123
5.6.4	Regularization $\mathcal{R}(D_X; \theta)$	123
5.7	Numerical Experiments	124
5.7.1	DRIP-Var	124
5.7.1.1	Setup	124
5.7.1.2	Results	125
5.7.2	DRIP-Max	128
5.7.2.1	Setup	128
5.7.2.2	Results	129
5.8	Conclusion	132
5.A	Proof of Proposition 4	132
5.B	Proof of Proposition 5	134
5.C	Kernel Method for Maximal Correlation	135
6	Conclusion and Future Work	139
	List of Author's Publications	142
	Bibliography	145

List of Figures

1.1	Relationship between the accuracy of house occupancy inference and the level of ℓ_4 distortion on the measurement data (blue dotted line).	3
1.2	Data security, data privacy and inference privacy.	9
3.1	Microphone array system.	28
3.2	The maximum privacy under perfect utility when no prior information is available. Because the maximum privacy can be unbounded, it is truncated when exceeding 100.	51
3.3	The maximum privacy under perfect utility when prior information is available. The maximum privacy threshold is 74.1.	51
3.4	The privacy and utility tradeoff when $S = 9, 12, 18$, respectively, obtained by using Algorithm 3. The corresponding maximum privacy under perfect utility obtained by solving (3.54) is indicated.	53
3.5	The utility at each iteration by using Algorithm 3, where the privacy threshold is set to be 10.	53
4.1	Relationships between the problem formulations (P4.0)-(P4.3).	79
4.2	The performance of the sequential scheme where Num-iter = 10 and $N_1, \dots, N_3 = 10$, in terms of public error trace τ_k and the minimum private error variance $\min(\eta_k)$.	80
4.3	The performance of the sequential scheme where Num-iter = 10 and $N_1, \dots, N_5 = 6$, in terms of public error trace τ_k and the minimum private error variance $\min(\eta_k)$.	81
4.4	Convergence of τ_k and $\min(\eta_k(\mathbf{C}_k))$ using Algorithm 5, where $S = 3$, $N_1, \dots, N_3 = 10$ and $\delta = 3$. The results shown are for time step $k = 20$.	81
4.5	Estimation of the mobile node's locations (private states).	83
4.6	Estimation of the mobile node's speed (public states).	84
4.7	The optimal estimation error of the public states at each step.	88
5.1	Privacy framework.	97
5.2	Machine learning DRIP architecture.	99
5.3	Neural network based maximal correlation estimator, where $\mu(\cdot)$ denotes the batch mean operator.	113
5.4	Probabilistic sanitizer $p_{\tilde{X} X}^\theta$.	122

5.5	The true labels are displayed on top, where M, F, Y, and E denote male, female, youngster, and elder, respectively. The 1st row shows the original images. Gender is the private variable and age group classification is the utility.	127
5.6	DRIP-Max configurations.	129

List of Tables

3.1	Summary of commonly-used symbols	30
5.1	DRIP-Var deep learning model	126
5.2	Image classification results. Gender is the private variable.	128
5.3	German credit data results	131

Symbols and Acronyms

Acronyms

PUT	Privacy-Utility Tradeoff
MLE	Maximum Likelihood Estimation
MAP	Maximum A Posteriori Estimation
MMSE	Mimumum Mean Square Error
MVUE	Minimum-Variance Unbiased Estimator
pdf	probability density function
pmf	probability mass function
CRLB	Cramér-Rao Lower Bound
FIM	Fisher Information Matrix
LDS	Linear Dynamical System
KL	Kullback–Leibler
RKHS	Reproducing Kernel Hilbert Space
PCA	Principle component analysis
CP	Compressive privacy
GAP	Generative adversarial privacy
DRIP	Data-driven Regularized Inference Privacy
ASUP	Arbitrarily strong utility-privacy tradeoff
EKF	Extended Kalman filter
w.r.t.	with respect to
i.i.d.	independent and identically distributed
a.s.	almost sure convergence of a random sequence
s.t.	subject to

Symbols

$\mathbb{R}^n$	The n -dimensional Euclidean space
$\mathbb{S}^n$	The set of $n \times n$ positive semi-definite matrices
$\mathbb{S}_+^n$	The set of $n \times n$ positive definite matrices
$(\cdot)^\top$	The transpose operator
$\langle \cdot, \cdot \rangle$	The inner product of two vectors
$\ \cdot\ $	The Euclidean norm
$\ \cdot\ _{\mathcal{G}}$	The induced norm of a vector in $\mathcal{G}$ -space
$\mathbf{A} \succeq \mathbf{B}$	$\mathbf{A} - \mathbf{B}$ is positive semi-definite
$\mathbf{A} \succ \mathbf{0}$	$\mathbf{A}$ is positive definite
$\text{diag}(\cdot)$	The block diagonal operation
$\text{Tr}(\cdot)$	The trace operation
$\mathcal{N}(\mu, \sigma^2)$	The Gaussian distribution with mean μ and variance σ^2
$\text{Null}(\mathbf{A})$	The null space of matrix $\mathbf{A}$
$\text{rank}(\mathbf{A})$	The rank of matrix $\mathbf{A}$
$\mathbf{A}^\perp$	The basis matrix of the null space of $\mathbf{A}$
$\mathbf{0}_N$	The $N \times N$ zero matrix
$\mathbf{1}_N$	The vector of all 1's of length N
$\mathbf{I}_N$	The $N \times N$ identity matrix
$\mathbf{I}$	The identity matrix with proper size
$[\mathbf{A}]_{i,j}$	The entry in the i -th row and j -th column of matrix $\mathbf{A}$
$[\mathbf{A}]_{\mathcal{P},\mathcal{Q}}$	The submatrix of $\mathbf{A}$ consisting of $[\mathbf{A}]_{i,j}$ for $(i,j) \in \mathcal{P} \times \mathcal{Q}$
$\mathbf{A}_{i:j}$	The matrix product $\mathbf{A}_j \dots \mathbf{A}_i$ if $i \leq j$ and $\mathbf{A}_{i:j} = \mathbf{I}$ otherwise.
$\text{vecdiag}(\mathbf{A})$	A column vector consisting of diagonal entries of $\mathbf{A}$
$\otimes$	The Kronecker product
$\circ$	The composition of functions
$p_X(\cdot)$	The pdf of X
$p_{X Y}(\cdot \cdot)$	The conditional pdf of X given Y
$X \sim p_X$	Random variable X follows a pdf p_X
$\log$	The binary logarithm
$I(A, B)$	The mutual information between A and B
$H(A)$	The entropy of A
$D(A \ B)$	The Kullback–Leibler divergence between A and B
$\mathbb{P}$	The probability measure defined on an underlying probability space

$\mathbb{E}_X$ The expectation [with respect to \(w.r.t.\)](#) random variable X

Chapter 1

Introduction

1.1 Background

Privacy has become a major concern when individuals need to share personal data on social networks [1] and send data to service providers [2] in the hope of receiving certain utilities such as storage services and application services. The risk of exposing sensitive information to potential adversary (including the service provider itself) might arouse data owner's antipathy against third-party's access to their personal data. The most recent example of this is the announcement of a new privacy policy by WhatsApp in 2021 [3] declaring that it would share information across the Facebook network of companies, which sparked a massive exodus from the messaging applications to rival messaging applications like Signal, Telegram. Nevertheless, in order to customize their products and services to improve the user experience, companies need to obtain information from the overall user population to gain insight into what many of their users are doing. Individual data to be collected is susceptible to various privacy attacks such as membership inference attack [4], identification attack [5] and reconstruction attack [6].

In addressing the general public's privacy concerns, more comprehensive and stricter privacy and security regulations have been put into effect. For example,

the enforcement of the General Data Protection Regulation (GDPR) law in 2018 imposes obligations onto organizations who target or collect data related to people in the European Union. The GDPR requires organizations to collect and process only as much data as absolutely necessary for the legitimate purposes specified in such a way as to ensure appropriate security, integrity, and confidentiality. Hence data privatization is a desideratum for data owners to limit detailed exposure of their individual data while enabling a learner to make optimal use of the overall user data at scale. To cope with the paradox that learning as much as possible about a group while learning as little as possible about a single individual within it, big tech companies such as Google and Apple have adopted *differential privacy* [7, 8] to privatize data on the user’s device to maintain the privacy of individual users when collecting and sharing aggregate statistics about user habits.

This thesis gives a reconsideration of privacy preservation in the sense that we are concerned with protecting the inference of private information that is statistically dependent on raw data. This is called *inference privacy*, which should be distinguished from the well-known notions of *data security* and *data privacy*. Data privacy is an area of data protection that concerns the proper handling of sensitive data and data security protects data from compromise by external attackers, both of which are tied to an individual data straightforwardly. The similarities and differences among them are highlighted in Section 1.3.1. As an illustration of inference privacy, we consider the data collected by health apps. It can include both statistical information about our gender and health status. However, we may perceive the gender information as confidential and want to hide it. Traditional privacy protection techniques such as anonymization have been shown to be a failure as a mean to protect this type of privacy. For example, researchers have successfully launched a re-identification attack on a set of de-identified hospital records where anonymized individual patients were re-identified using simple demographic information [9]. Thankfully, the brewing anxiety of privacy concern can be alleviated by employing the notion of inference privacy to sanitize raw data via randomization at user’s discretion [10]. A key challenge in inference privacy

research is how to value the utility of data with privacy consideration as sanitizing data can jeopardize the usefulness of data. Despite the fact that there are natural tensions between utility and privacy that arise whenever a data holder must share data containing sensitive information with others, it is of great interest to characterize such tensions by quantifying the utility-privacy tradeoff that can be employed as controllable degrees of freedom when such a system is deployed. To demonstrate this, we exemplify the utility-privacy tradeoff with an application in smart meters excerpted from the work [11].

Example 1 (Real-time privacy-preserving data release for smart meters). Smart Meters (SMs) share the power consumption of users with utility providers almost in real-time. These fine-grained signals however is susceptible to real-time privacy threats. E.g., utility providers can learn sensitive information like house occupancy from SMs data in an online fashion. To limit the performance of such attacks, the paper [11] adopted an information-theoretic privacy measure and designed a privatization mechanism that can provide a target level of privacy by adding a minimal amount of distortion to the SMs measurements. Fig. 1.1 depicts the utility-privacy tradeoff between the accuracy of adversarial inference on house occupancy and the level of data distortion.

FIG. 1.1: Relationship between the accuracy of house occupancy inference and the level of ℓ_2 distortion on the measurement data (blue dotted line).

The gist of this thesis is the investigation of privacy preservation on two scenarios: (1) Raw data to be sanitized is linearly related to private data to be protected and the model parameters are known in advance; (2) Samples of raw data and private data are available, and the intrinsic relationship between raw data and private data is unspecified and can hardly be characterized explicitly.

1.2 Motivation

The increasing number of multifarious sensing and monitoring applications installed in mobile phones, offices and public facilities has led to the proliferation of various services based on network data analytics [12–15]. Due to the limited data that a single sensor or agent can observe because of its geographical placement or location, agent type and computational capabilities, decentralized agent networks [16–24] are often deployed to overcome the insufficiency of data retrieved from a single agent. Agent fusion or sense-making is then used to combine the sensory data derived from disparate sources to reduce the inference uncertainty. However, aggregation of data poses a higher risk of privacy leakage. For example, social network users in a community may share their personal opinions or experiences about different products over time. While the aggregated data can be useful feedback for a company to improve its own product, with a database recording the preferences of each user for multiple products over a period of time, one can infer personal traits and other sensitive attributes like the gender and income level of a user. Another example is in the deployment of various body sensors (agents) for evaluating a user’s health condition [25]. With the raw observations sent from the sensors, a service provider cannot only analyze the user’s health condition but can also infer some sensitive information about the user such as her location and personal preferences. Therefore, preserving the privacy of certain sensitive parameters associated with the sensor data is an important requirement.

In Chapter 3, we consider a decentralized network where each source, node or agent makes a noisy observation, which is linearly related to a set of system parameters. A set of public and private parameters are defined to be linear maps of the system parameters. We assume a decentralized estimation scheme is used where the agents send their observations to a fusion center to allow it to infer the public parameter with high fidelity [26–29]. (Following most of the literature, the term “decentralized” refers to the agents sending summary information instead of their raw observations to a fusion center.) The agents however want to keep the private parameters secret. To achieve this, each agent sanitizes its observation before sending to the fusion center. We consider two sanitization methods: linear compression and noise perturbation. Since a trusted third-party who can help perturb the agent observations in a centralized manner does not exist or is impractical in many applications, a decentralized sanitization scheme is considered where each agent performs its local sanitization independently.

While sensors are assumed to make single-shot measurements in Chapter 3, sensors in a plethora of applications also make correlated measurements across time. For example, sensors that are deployed in modern infrastructures like smart grids and traffic monitoring can make continuous observations of a state vector and send the observations in real-time to a fusion center to perform state estimation [30,31]. The state evolving over time could contain private information. For example, in a smart grid infrastructure, customers could be offered better rates if they continuously send their instantaneous power consumption to the utility company, which helps to improve the demand forecasting mechanism. However, from the fine-grained smart meter readings, the customers’ private activities such as when and for how long appliances are used can be inferred [32]. This inspired us to consider the problem of privacy preservation in a linear dynamical system (LDS) in Chapter 4 where the system states are partitioned into private states and public states. Our goal is to allow the fusion center to estimate the public states while making it difficult for it to estimate the private states. To achieve this, we let each sensor linearly compress its measurements to remove statistical information about

the private states before sending it to the fusion center. It is worth noting that in a LDS, the public states at one time step may provide statistical information about the private states in a future time step. There is therefore a need to consider the tradeoff between preserving the public states and protecting the private states over all time steps.

Concurrently, the information explosion has resulted in the collection of massive amounts of unstructured data such as images, videos and audios on a daily basis. It becomes inadequate for signal processing tools (in the classical sense, including linear estimators, adaptive filters, wavelet transforms, to mention a few) to handle such data efficiently. In the meantime, artificial intelligence and machine learning coupled with the rapid development of processor power and computer parallelization have enjoyed tremendous success in complex tasks such as object recognition, speech recognition and language translation. These results are not entirely surprising as neural networks are well-known universal function approximators. Big tech companies such as Google, Microsoft and Amazon have provided easy access to efficient and sophisticated information processing supported by large data storage, high computing power and intelligent software. While this technology benefits humanity, it also comes with privacy leakage risks that is gaining public attention. When a user sends data to a service provider for it to perform inference for an authorized task, potential adversaries (including the service provider) may exploit the raw data to glean sensitive information. Studies [1, 33, 34] have shown that users' personal information such as sexual orientation or political affiliation can be accurately inferred from their activities on social networking platforms. The situation is further aggravated by data breach incidents. E.g., 1.5 million members of Singapore's largest healthcare group had their personal data compromised in 2018 [35]. Due to the privacy concerns, data owners may feel reluctant to share their data with others in collaborative learning where data from various data owners are collected and analyzed. Therefore, data sanitization becomes essential as a tool to reduce the amount of private information before sharing data or sending data for cloud computing. Resting on the assumption that only data samples will

be available in the big data era, we rely on machine learning techniques to learn data sanitization mechanisms from data. We take care of the circumstance that implementing privacy-preserving frameworks usually requires changing the service provider’s backend analytics systems that work with raw data to be compatible with the sanitized data. In Chapter 5, we propose a data-driven inference privacy preserving framework to sanitize data so as to prevent leakage of sensitive information that is present in the raw data, while ensuring that the sanitized data is still compatible with the service provider’s legacy inference system.

1.3 Related Work

In this section, we draw a distinction between data privacy, data security and inference privacy and highlight the related works for inference privacy from different perspectives.

1.3.1 Inference Privacy, Data Privacy and Data Security

In general, privacy can be categorized into *data privacy* and *inference privacy*. They can also be easily confused with *data security*. Data security refers to the practice of protecting data from unauthorized access and corruption by external attackers and malicious insiders. As an example of this, secure communication [36] entitles two entities to communicate in a way unsusceptible to eavesdropping or interception by using techniques such as homomorphic encryption [37, 38]. Other examples include privacy-preserving outsourced storage and computation [2]. Data privacy is the term that concerns the proper collection and dissemination of sensitive data including, notably, personal data but also other confidential data. It often refers to the action of protecting original data or data source from being recognized by an untrusted data curator. For instance, privacy-preserving machine learning [39] prevents a learner from directly accessing the private data from multiple clients by adopting federated learning (training the model in a decentralized

manner) [40]. Within this regard, data anonymization [41, 42] appears as a compelling approach to guarantee that the individuals who are the subjects of the data cannot be re-identified while the data remain practically useful. In addition, differential privacy [7, 8] proposed by Dwork has emerged as an advanced privacy-preserving techniques within the privacy research community. It considers the scenario where statistical information is to be learned from a statistical database and guarantees that the presence or absence of any individual in the data set does not significantly influence the results of analyses on the data set. A typical method of guaranteeing differential privacy is to add Laplacian noise to the released statistical information so that the change of one data entry has limited impact on the changes of the released information.

Inference privacy that we consider in this thesis refers to the procedure of preventing some sensitive information correlated with raw data from being inferred by a data recipient (e.g., service provider, fusion center) [43–46]. To achieve this, raw data are distorted intelligently prior to being revealed in such a way that it reduces possible inferences about the private information and keeps the useful (non-private) information. As distinct from data security, inference privacy assumes the adversarial attacker and the legitimate user of data are the same entity. Unlike data privacy, inference privacy focuses on statistical private information hidden in original data. This is illustrated in Fig. 1.2. It should be pointed out that the approaches for data privacy like encryption are unable to eliminate the sensitive information contained in the privatized data. To measure the exposure of private information retained in the sanitized data, the authors of [43] introduced three inference privacy metrics, namely average information leakage (mutual information), maximum information leakage and information privacy. Specifically, the mutual information between a pair of random variables (X, Y) is defined as $I(X; Y) = D(P_{X,Y} \parallel P_X \otimes P_Y)$, where $P_{X,Y}$ is their joint distribution, P_X and P_Y are the marginal distributions and $D(\cdot \parallel \cdot)$ is the Kullback–Leibler divergence. It quantifies the amount of information obtained about X by observing Y . It was proven in the paper that any fixed level of differential privacy or privacy in terms

FIG. 1.2: Data security, data privacy and inference privacy.

of average or maximum information leakage can be guaranteed by achieving certain level of information privacy. Meanwhile, achieving differential privacy does not necessarily bound the amount of average or maximum information leakage in general. The other privacy metrics that have been extensively used in privacy regime include maximal correlation [47–50] and Fisher information [51–54]. The relationships among different privacy metrics were studied in [55–57].

The reference [58] provided operational interpretation of commonly used privacy leakage measures, including mutual information, maximal correlation, and local differential privacy. We briefly elucidate the intuition behind the privacy metrics used in this thesis. Let $S \in \mathcal{S}$ and Z be a private variable and a sanitized variable that are jointly distributed. A small mutual information between S and Z ensures that only limited bits of S are revealed when releasing Z . It can also be justified using Fano’s inequality: $I(S; Z) \leq \epsilon$ implies that $\mathbb{P}(\hat{S}(Z) \neq S) \geq (H(S) - 1 - \epsilon)/\log|\mathcal{S}|$ for any estimator $\hat{S}$ w.r.t. S and any positive constant ϵ , where $I(\cdot; \cdot)$ and $H(\cdot)$ are the mutual information and entropy functions, respectively. Hence, the probability of an adversary correctly guessing S is upper-bounded. The use of maximal correlation as a privacy metric gives an assurance that an adversary who observes Z cannot efficiently estimate $f(S)$ for any real-valued function f . The

inverse of the Fisher information (the Cramér-Rao bound) gives a lower bound on the variance of any unbiased estimates of S based on observed data Z .

Since data sanitization is traditionally the process of distorting data to achieve some level of privacy, it is a problem of mapping a data to a different one subject to specific utility and privacy requirements. This process is comparable to lossy compression or irreversible compression [59], that is the class of data encoding methods that uses inexact approximations and partial data discarding to represent the content either via dimension reduction or randomizing mechanism. This entails two commonly used data sanitization methods: (1) compressive privacy [60–64] that projects raw measurements to a lower dimensional space, and (2) probabilistic mapping [10, 43, 65] that adds randomness to perturb raw measurements.

1.3.2 Information Theoretic Privacy

Privacy-preserving estimation for a single data source has been well-studied from an information-theoretic point of view with the pioneering works presented in [10, 42, 43]. The work [42] initiated an information-theoretic formulation of the privacy-distortion tradeoff by means of perturbation of the original data. The paper [10] gave a precise quantification of the utility-privacy tradeoff for a sequence of i.i.d. data entries in a database with public and private attributes indicated, where the raw data sequence is encoded into a sanitized sequence from which the public data is expected to be reconstructed. Privacy is quantified by the conditional entropy of the private data given the encoded output and utility is measured by the average distortion of the reconstructed public data.

In general, inference privacy framework can be captured by a statistical model [43]. Given a pair of random variables (S, X) assigned with a known joint probability distribution, a user observing X wishes to reveal to another user as much information about X as possible without disclosing too much knowledge of the private variable S . This is achieved by randomizing X to produce a sanitized data Z according to a probabilistic mapping $p_{Z|X}$. The authors of [43] measured privacy

leakage and utility cost by the mutual information $I(S; Z)$ and average distortion $\mathbb{E}[d(X, Z)]$, respectively, and hence cast the utility-privacy tradeoff as a modified rate-distortion problem [59]. The reference [66] introduced a setup called *privacy funnel* by redefining the utility function to be the mutual information $I(X; Z)$ and hence the overall formulation is the opposite of the information bottleneck problem [67]. Variants of these works are discovered in [68, 69] in which a public variable correlated with X was defined and a case of two-agent interactions was studied. From an estimation-theoretic viewpoint, references [70, 71] relied on the techniques developed in [72] to allow a service provider to reconstruct (in a mean-squared error sense) certain functions of the raw data X , while other functions of the private variable S cannot be reconstructed with distortion below a certain threshold. The papers [73, 74] formulated the utility-privacy tradeoff in terms of the minimum mean-squared error (MMSE). Moreover, the paper [75] presented necessary and sufficient conditions for disclosing a non-trivial information about X under perfect privacy (i.e., Z and S are independent). More studies related to perfect privacy are found in [76–78]. The authors of [79] studied a lossy source coding problem in which a source encoder communicates with two decoders, one with and one without correlated side information with an additional constraint on the privacy of the side information at the uninformed decoder.

All the above-mentioned works suppose a single data source with a source alphabet with finite support and hence do not generalize immediately to a multi-agent system with continuous measurement data and private data. Meanwhile, our works [80, 81] (Chapter 3) studied the problem of preserving the estimation privacy of a set of private parameters when data is distributed among multiple users. Our works [82, 83] (Chapter 4) further extended the privacy framework to a linear dynamical system. Similar problems with a single data source were studied by the following papers. The paper [84] addressed the problem of protecting the private state of a stochastic dynamical system whose sensor measurements are quantized and transmitted to a remote station. In [85–88], the authors investigated

the utility-privacy tradeoff quantified by mutual information in smart metering using a rate-distortion approach.

The following works have addressed the issue of privacy protection under the *hypothesis testing* framework. To determine the true distribution among a set of possible distributions for a given data set with privacy restrictions, the paper [65] applied a randomizing privacy mechanism to maximize the relative entropy between pairs of distributions of the randomized output while requiring that the mutual information between the randomized output and raw data is small. The problem of hypothesis testing with data privacy constraints was revisited by [89–92] where the receiver is provided with side information. The work [93] manipulated the original data sequence in an online fashion to maximize the error probability of an adversarial Bayesian detector while satisfying an average distortion constraint. The works [94, 95] investigated the privacy leakage problem in an eavesdropped distributed hypothesis test network from the Bayesian detection perspective. Under a similar decentralized detection framework, the papers [45, 96] proposed a nonparametric learning approach to design local privacy mappings to distort each agent’s observation, thus preventing the fusion center from using its received information to accurately infer the private hypothesis. The paper [46] considered ways to achieve robust information privacy for a set of private hypotheses.

In contrast to these works that focus on the protection of a private hypothesis in a *hypothesis testing* framework, we consider the privacy protection of a set of parameters in a parameter *estimation* framework in Chapters 3 and 4.

1.3.3 Learning Based Privacy

The aforementioned works are premised on known data distributions or system models. However, for data like images, speeches and natural languages, determining the data distribution or characterize it using an exact model is implausible due to their complex nature. To address this, several works have explored data-driven approaches for preserving inference privacy from different perspectives.

Focusing on protecting the private information of raw data while maximally preserving non-private information, the paper [60] introduced a dimension-reduction subspace approach called compressive privacy to sanitize a feature space by preserving information *w.r.t.* a utility space and reducing information *w.r.t.* a privacy space. This work was extended by [97] through a combination of compressive privacy and multi-kernel method. The works [45,98] enabled the detection of a public hypothesis while protecting a private hypothesis by learning mechanisms to distort sensors' measurements. The references [99–101] used variational auto-encoders to learn representations of input data that are invariant to certain sensitive factors while retaining as much of the non-sensitive information as possible. There are several obstacles when generalizing these approaches to broader privacy problems: (1) The methods in [60,97] are restricted to deal with linear utility and privacy subspaces and hence are incapable of managing nonlinear relationships; (2) Choosing a standard isotropic Gaussian prior and a Gaussian posterior for encoded data and decoded data in [99,100] is too simple to model complex data. Inspired by the generative adversarial networks [102], the state-of-the-art generative adversarial privacy (GAP) method proposed by [103–105] formulated the utility-privacy tradeoff as a competing game between a privatizer and an adversary. Based on adversarial training, the paper [106] adopted the notion of compressive privacy [60] to generate representations that can defend against reconstruction attacks. It takes into account multiple adversarial strategies for more robust evaluation of adversarial reconstruction attacks.

However, all the above-mentioned works perform data sanitization by mapping the raw data into a different space, which prevents legacy inference systems from working with the sanitized data. In addition, using adversarial strategies to compute privacy leakage does not protect privacy against an arbitrary attacker in practice. In Chapter 5, we develop practical implementation of mutual information and maximal correlation for an information or estimation-theoretic privacy metric. Thus we do not restrict the adversary to any particular attack model. Moreover, we propose the use of regularization techniques to minimize the disparity between

the distributions of the sanitized data and raw data, making our privacy preserving framework compatible with a service provider's legacy learning systems. Our proposed approach thus incurs no additional cost for the service provider.

1.4 Thesis Contribution and Outline

In Chapter 3, we consider the case where agents in a network make observations that are linearly related to a set of public parameters and a set of private parameters. Agents send sanitized observations to a fusion center to allow it to infer the public parameters, while preventing it from estimating the private parameters better than a predefined accuracy. Our main contributions are the following:

1. We make explicit the relationship between additive random noise and linear compression as privacy mechanisms under the [Cramér-Rao lower bound \(CRLB\)](#) privacy framework. We show that the [CRLBs](#) for estimating the private parameters under the linear compression mechanism form the boundary of the set of [CRLBs](#) under noise perturbation.
2. We introduce the notion of arbitrarily strong utility-privacy tradeoff (ASUP), and derived necessary and sufficient conditions under which this is achievable. In addition, we propose a method to find the maximum privacy that can be attained while maintaining perfect utility under a constraint on the noise perturbation power.
3. In the case where ASUP is not achievable, we propose an alternating optimization algorithm to find a sanitization to achieve an optimal utility-privacy tradeoff.

For a further development of the topic of protecting parameter estimation in a linear system, we consider a linear dynamical system in Chapter 4, where multiple sensors make observations about a set of public states and private states. We

linearly compress the measurements at each sensor to prevent the fusion center from accurately estimating the private states while still allowing it to estimate the public states with low error. Our main contributions are as follows.

1. We formulate a utility-privacy tradeoff optimization problem for a LDS involving privacy constraints on the predicted estimation error of the private states multiple steps ahead. We show how to find the dimension of the compressive map and the map itself and propose an algorithm to achieve this. We provide a bound for the number of steps to look ahead to achieve the same privacy level in all future time steps.
2. We consider the case where there are multiple decentralized sensors that optimize their own local compressive map. We allow each sensor to take turns to transmit messages to the other sensors and propose optimization methods to find the optimal local compressive map.

Due to the complex nature of unstructured data collected by modern computing infrastructure, characterizing precise data relationship using an analytical model becomes infeasible for such data. In Chapter 5, we formulate a data-driven inference privacy preserving framework that sanitizes data to prevent leakage of sensitive information present in the raw data while ensuring that the sanitized data is still compatible with the service provider's legacy inference system and provides the maximal utility. In the case where the conditional probability distribution of the private data is available, we develop a practical variation method implementation of the average information leakage metric as a privacy metric. In the case where the conditional distribution is unknown, we utilize maximal correlation as a privacy metric. We derive sufficient conditions under which our proposed data-driven implementations converge to their respective expected privacy metrics in the large training sample size regime. Furthermore, we adopt maximum mean discrepancy and domain adaption as techniques to regularize the domain of the sanitized data to ensure its legacy compatibility. Finally, we develop a deep learning model as an

example of the proposed inference privacy framework. Our main contributions are the following:

1. We develop a data-driven privacy sanitization framework DRIP to preserve inference privacy. We introduce domain regularization techniques to avoid over-fitting and to ensure that the sanitized data is compatible with a legacy inference system by forcing the marginal distributions between the sanitized data and the raw data to be similar.
2. We propose data-driven versions of average information leakage and maximal correlation as privacy leakage measures and identify their respective advantages and limitations. The DRIP frameworks are called DRIP-Var and DRIP-Max, respectively. We show that under mild technical conditions and with the training sample size growing large, the proposed data-driven implementations converge to the expected privacy metrics, respectively.
3. We illustrate our proposed frameworks using a deep learning architecture and verify its performance on an image dataset [107] and a diabetes dataset [108]. We compare our approach with the existing state-of-the-art methods.

Chapter 2

Preliminaries

In this chapter, we introduce the preliminary concepts that are necessary for the rest of this thesis. We will briefly discuss the Cramér–Rao bound, the Kalman filter, the fundamental concepts in information theory and the reproducing kernel Hilbert space. For a comprehensive treatment of these topics, we refer the interested reader to [59, 109–111].

2.1 Cramér–Rao Lower Bound

In estimation theory and statistics, the [CRLB](#) expresses a lower bound on the variance of unbiased estimators of multivariate deterministic parameters, stating that the variance of any such estimator is at least as high as the inverse of the Fisher information matrix under some regularity conditions. An unbiased estimator which achieves this lower bound is said to be efficient and is therefore the minimum variance unbiased estimator (MVUE). Moreover, the variances of the maximum likelihood estimates (MLE) achieve this lower bound asymptotically (when the sample size tends to infinity).

Suppose an unknown fixed parameter vector

$$\boldsymbol{\theta} = [\theta_1, \dots, \theta_M] \quad (2.1)$$

is to be estimated from a set of measured data

$$\mathbf{x} = [x_1, \dots, x_N], \quad (2.2)$$

according to some probability density function $f(\mathbf{x}; \boldsymbol{\theta})$. Let $T(\mathbf{x})$ be an estimator of $\boldsymbol{\theta}$. We make two regularity assumptions on $f(\mathbf{x}; \boldsymbol{\theta})$:

- The Fisher information is always defined, i.e., for all $\mathbf{x}$ such that $f(\mathbf{x}; \boldsymbol{\theta}) > 0$,

$$\frac{\partial \log f(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad (2.3)$$

exists, and is finite.

- The operations of integration *w.r.t.* $\mathbf{x}$ and differentiation *w.r.t.* $\boldsymbol{\theta}$ can be interchanged in the expectation of $T(\mathbf{x})$, i.e.,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left[\int T(\mathbf{x}) f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x} \right] = \int T(\mathbf{x}) \frac{\partial}{\partial \boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta}) d\mathbf{x}. \quad (2.4)$$

For all $\boldsymbol{\theta}$, we have

$$\text{cov}(T(\mathbf{x})) \succeq \frac{\partial \mathbb{E}_{\mathbf{x}}[T(\mathbf{x})]}{\partial \boldsymbol{\theta}} [\mathbf{I}(\boldsymbol{\theta})]^{-1} \left(\frac{\partial \mathbb{E}_{\mathbf{x}}[T(\mathbf{x})]}{\partial \boldsymbol{\theta}} \right)^{\top}, \quad (2.5)$$

where $\frac{\partial \mathbb{E}_{\mathbf{x}}[T(\mathbf{x})]}{\partial \boldsymbol{\theta}}$ is a Jacobian matrix of size $M \times M$ and $\mathbf{I}(\boldsymbol{\theta})$ is called the **Fisher Information Matrix (FIM)**, with its element in the i -th row and j -th column defined as

$$[\mathbf{I}(\boldsymbol{\theta})]_{i,j} = -\mathbb{E}_{\mathbf{x}} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(\mathbf{x}; \boldsymbol{\theta}) \right]. \quad (2.6)$$

If $T(\mathbf{x})$ is an unbiased estimator of $\boldsymbol{\theta}$, i.e., $\mathbb{E}_{\mathbf{x}}[T(\mathbf{x})] = \boldsymbol{\theta}$, then (2.5) reduces to

$$\text{cov}(T(\mathbf{x})) \succeq [\mathbf{I}(\boldsymbol{\theta})]^{-1}. \quad (2.7)$$

If we place a prior distribution on the parameter $\boldsymbol{\theta}$ with the probability density function denoted as $\lambda(\boldsymbol{\theta})$, we have the Bayesian Cramér-Rao inequality:

$$\mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}} [(T(\mathbf{x}) - \boldsymbol{\theta})^2] \succeq (\mathbb{E}_{\boldsymbol{\theta}} [\mathbf{I}(\boldsymbol{\theta})] + \mathbf{I}(\lambda))^{-1}, \quad (2.8)$$

where

$$\mathbf{I}(\lambda) = \int \left(\frac{\partial \lambda(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^2 \lambda(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (2.9)$$

2.2 Kalman Filter

In statistics and control theory, Kalman filter [112] has long been regarded as the optimal solution to many tracking and data prediction tasks. It is an algorithm that uses a series of measurements observed over time containing statistical noise and other inaccuracies to produce estimates of unknown state that tend to be more accurate than those based on a single measurement alone.

The standard Kalman filter is based on linear dynamical systems discretized in the time domain, where the true state $\mathbf{x}_k \in \mathbb{R}^L$ at time k is evolved from the state $\mathbf{x}_{k-1} \in \mathbb{R}^L$ at time $k-1$ according to

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_k, \quad (2.10)$$

where $\mathbf{F}_k \in \mathbb{R}^{L \times L}$ is the state transition model which is applied to the previous state $\mathbf{x}_{k-1}$ and $\mathbf{w}_k \in \mathbb{R}^L$ is the process noise, which is assumed to be drawn from a zero mean multivariate normal distribution with covariance $\mathbf{Q}_k$: $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$.

At time k an observation $\mathbf{z}_k$ of the true state $\mathbf{x}_k$ is made according to

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k, \quad (2.11)$$

where $\mathbf{H}_k \in \mathbb{R}^{N_k \times L}$ is the observation model, which maps the true state space into the observed space, and $\mathbf{n}_k \in \mathbb{R}^{N_k}$ is the observation noise, which is assumed to be zero mean Gaussian white noise with covariance $\mathbf{R}_k$: $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$. The initial state $\mathbf{x}_0$ and the noise vectors $\{\mathbf{w}_i, \mathbf{n}_i\}_{i \geq 1}$ are all assumed to be mutually independent.

We are interested in estimating $\mathbf{x}_k$ given the observations up to and including time k . It is well known [112] that the Kalman filter is optimal in terms of MMSE. The state of the filter is represented by two variables:

- $\hat{\mathbf{x}}_k$, the *a posteriori* state estimate at time k given $\{\mathbf{z}_i\}_{i=1}^k$;
- $\mathbf{P}_{k|k}$, the covariance matrix of the *a posteriori* estimate $\hat{\mathbf{x}}_k$.

The Kalman filter is a recursive estimator. This means that only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state. The Kalman filter contains two distinct phases: “predict” and “update”. In the “predict” phase, the predicted state estimate (*a priori* state estimate) and error covariance are predicted, respectively, by

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1} &= \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}, \\ \mathbf{P}_{k|k-1} &= \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k. \end{aligned} \quad (2.12)$$

In the “update” phase, the current *a priori* prediction is combined with current observation information to refine the state estimate. This improved estimate is termed the *a posteriori* state estimate. The updated state (*a posteriori*) estimate

and error covariance are updated, respectively, through

$$\begin{aligned}\hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k-1|k-1}), \\ \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1},\end{aligned}\tag{2.13}$$

where $\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^\top \mathbf{S}_k^{-1}$ denotes the Kalman gain with $\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k$ being the covariance of the innovation term $\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k-1|k-1}$.

2.3 Information Theory

In information theory, the entropy of a random variable measures the average level of uncertainty inherent in the variable's possible outcomes.

Definition 1 (Entropy). Let $X \in \mathcal{X}$ be a random variable associated with probability mass function P_X . The entropy of X , denoted as $H(X)$, is defined as

$$H(X) = \sum_{x \in \mathcal{X}} P_X(x) \log P_X(x).\tag{2.14}$$

Kullback–Leibler divergence (also called relative entropy) measures the closeness of two probability distributions.

Definition 2 (Kullback–Leibler divergence). For discrete probability distributions P and Q defined on the same probability space $\mathcal{X}$, the relative entropy from Q to P is defined to be

$$D(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.\tag{2.15}$$

It can be easily verified by Jensen's inequality that $D(P \parallel Q) \geq 0$ and is equal to 0 if and only if $P = Q$:

$$D(P \parallel Q) = - \sum_{x \in \mathcal{X}} P(x) \log \frac{Q(x)}{P(x)} \geq \log \left(\sum_{x \in \mathcal{X}} Q(x) \right) = 0,\tag{2.16}$$

where the equality holds true if and only if $P = Q$.

The mutual information of two random variables quantifies the amount of information obtained about one random variable through observing the other random variable.

Definition 3 (Mutual information). Let (X, Y) be a pair of random variables with joint distribution denoted as $P_{X,Y}$ defined over the space $\mathcal{X} \times \mathcal{Y}$. The marginal distributions of X and Y are denoted as P_X and P_Y , respectively. The mutual information is defined as

$$I(X; Y) = D(P_{X,Y} \parallel P_X \otimes P_Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P_{X,Y}(x, y) \frac{P_{X,Y}(x, y)}{P_X(x)P_Y(y)}. \quad (2.17)$$

2.4 Kernel Methods

Kernel methods have received considerable attention in machine learning community, mainly due to their capability of working with linear inference models, allowing at the same time to identify nonlinear relationships among input patterns [113].

We will make use of several terms from functional analysis to introduce the [reproducing kernel Hilbert space \(RKHS\)](#).

Definition 4 (Inner product). Let $\mathcal{H}$ be a vector space over $\mathbb{R}$. A function $\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \mapsto \mathbb{R}$ is an inner product if for all $f, g, h \in \mathcal{H}$ and $\alpha \in \mathbb{R}$, we have

- Linearity: $\langle f + h, g \rangle_{\mathcal{H}} = \langle f, g \rangle_{\mathcal{H}} + \langle h, g \rangle_{\mathcal{H}}$ and $\langle \alpha f, g \rangle_{\mathcal{H}} = \alpha \langle f, g \rangle_{\mathcal{H}}$.
- Hermitian symmetry: $\langle f, g \rangle_{\mathcal{H}} = \langle g, f \rangle_{\mathcal{H}}$.
- Positive-definite: $\langle f, f \rangle_{\mathcal{H}} > 0$ if $f \neq 0$.

Definition 5 (Hilbert space). A Hilbert space $\mathcal{H}$ is a complete metric space [w.r.t.](#) the distance function induced by the inner product.

Definition 6 (Evaluation functional). Let $\mathcal{X}$ be an arbitrary set and $\mathcal{H}$ a Hilbert space of real-valued functions on $\mathcal{X}$. The evaluation functional over the Hilbert space of functions $\mathcal{H}$ is a linear functional $F_x : \mathcal{H} \mapsto \mathbb{R}$ that evaluates each function at a point x :

$$F_x[f] = f(x), \forall f \in \mathcal{H}. \quad (2.18)$$

Definition 7 (Reproducing kernel Hilbert space). A Hilbert space $\mathcal{H}$ is a [RKHS](#) if the evaluation functionals are bounded, i.e., if for every $x \in \mathcal{X}$, there exists a constant M_x such that

$$|F_x[f]| = |f(x)| \leq M_x \|f\|_{\mathcal{H}}, \forall f \in \mathcal{H}. \quad (2.19)$$

The Riesz representation theorem implies the following result.

Theorem 1. If $\mathcal{H}$ is a [RKHS](#), then for all $x \in \mathcal{X}$, there exists a unique element $K_x \in \mathcal{H}$ with the reproducing property:

$$f(x) = F_x[f] = \langle f, K_x \rangle_{\mathcal{H}}, \forall f \in \mathcal{H}. \quad (2.20)$$

This allows us to represent our linear evaluation functional by taking the inner product with an element of $\mathcal{H}$. Since K_x is a function in $\mathcal{H}$, by the reproducing property, for each $y \in \mathcal{X}$ we can write

$$K_x(y) = \langle K_x, K_y \rangle_{\mathcal{H}}. \quad (2.21)$$

This induces the definition of reproducing kernel in $\mathcal{H}$.

Definition 8. The reproducing kernel of $\mathcal{H}$ is a function $K : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, defined by

$$K(x, y) := K_x(y). \quad (2.22)$$

Kernel method maps data points in the original space to an implicit higher-dimensional feature space and the kernel trick enables us to operate in the implicit higher-dimensional feature space without computing the coordinates of the data in the space. As an illustration, denoting K_x as the feature mapping of a point x induced by a reproducing kernel $K(\cdot, \cdot)$, the kernel trick leads to $\langle K_x, K_y \rangle = K(x, y)$.

Chapter 3

Privacy-Preserving Decentralized Linear Estimation

Privacy-preserving data aggregation in wireless sensor networks has become a research hotspot. While a fusion center's role is to perform inference on a particular task of interest, there is nothing stopping it from using the received sensor information to learn other sensitive information. An example is in the deployment of home security system comprising surveillance cameras, motion sensors, door access, etc. If these devices transmit the raw data to a fusion center, the fusion center cannot only use the collected data for a legitimate task, but also has the potential to intrude on the privacy of the home inhabitants. The devices therefore need to perform data sanitization to limit the amount and quality of information they send to the fusion center. In this chapter, we investigate the problem of protecting parameter estimation in a decentralized system, where each agent in a network makes a local observation that is linearly related to a set of public and private parameters. The agents send their observations to a fusion center to allow it to estimate the public parameters. To prevent leakage of the private parameters, each agent first sanitizes its local observation using a local privacy mechanism before transmitting it to the fusion center. We investigate the utility-privacy tradeoff in terms of the Cramér-Rao lower bounds for estimating the public and private parameters. We

study the class of privacy mechanisms given by linear compression and noise perturbation, and derive necessary and sufficient conditions for achieving arbitrarily strong utility-privacy tradeoff in a multi-agent system for both the cases where prior information is available and unavailable, respectively. We also provide a method to find the maximum estimation privacy achievable without compromising the utility and propose an alternating algorithm to optimize the utility-privacy tradeoff in the case where arbitrarily strong utility-privacy tradeoff is not achievable.

3.1 Problem formulation

In this section, we present our system model and assumptions. Consider a multi-agent network consisting of the agents $i = 1, \dots, S$ and a fusion center. Each agent i makes a noisy observation $\mathbf{y}_i \in \mathbb{R}^{N_i}$ about a system parameter $\mathbf{x} \in \mathbb{R}^L$. The observation model for agent i is given by

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x} + \mathbf{n}_i, \quad (3.1)$$

where $\mathbf{H}_i \in \mathbb{R}^{N_i \times L}$ is the observation model matrix, and $\mathbf{n}_i \sim \mathcal{N}(\mathbf{0}_{N_i}, \mathbf{R}_i)$ is the measurement noise. Here, N_i may be smaller than L , hence each agent i may not be able to infer the system parameter $\mathbf{x}$ based on its own observation $\mathbf{y}_i$. Furthermore, even if an agent i is able to estimate $\mathbf{x}$ from its local observation $\mathbf{y}_i$, the fusion center with access to all agents' observations achieves a higher accuracy than each individual agent.

In this chapter, we consider the case where the fusion center aims to estimate $\mathbf{u} = \mathbf{U}\mathbf{x} \in \mathbb{R}^U$, where $\mathbf{U} \in \mathbb{R}^{U \times L}$ and $U > 0$, based on information it receives from all agents in the network. However, the agents also want to prevent the fusion center from inferring a set of private parameters $\mathbf{g}_i = \mathbf{G}_i \mathbf{x} \in \mathbb{R}^{G_i}$, where $\mathbf{G}_i \in \mathbb{R}^{G_i \times L}$, $i = 1, \dots, S$. We assume that at least one $\mathbf{G}_i$, $i = 1, \dots, S$, is non-zero. Otherwise our problem formulation reduces to the case without privacy consideration.

Stacking up all agents' measurements, we have

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (3.2)$$

where $\mathbf{y} = [\mathbf{y}_1^\top, \dots, \mathbf{y}_S^\top]^\top \in \mathbb{R}^N$, $\mathbf{H} = [\mathbf{H}_1^\top, \dots, \mathbf{H}_S^\top]^\top \in \mathbb{R}^{N \times L}$, $\mathbf{n} = [\mathbf{n}_1^\top, \dots, \mathbf{n}_S^\top]^\top \in \mathbb{R}^N$, and $N = \sum_{i=1}^S N_i$. Let $\mathcal{S}_i = \{\sum_{k=1}^{i-1} N_k + 1, \dots, \sum_{k=1}^i N_k\}$ be the index set corresponding to agent i . Note that the measurement noise $\mathbf{n}_i$ and $\mathbf{n}_j$ for any two different agents i and j are not necessarily independent of each other. We assume that $\mathbf{n} \sim \mathcal{N}(\mathbf{0}_N, \mathbf{R})$, where the block diagonal part of $\mathbf{R}$ is equal to $\text{diag}(\mathbf{R}_1, \dots, \mathbf{R}_S)$ and the off-diagonal entries are the noise correlations between agents.

Example 2. Consider an audio system consisting of S microphone arrays placed at different spatial locations as illustrated in Fig. 3.1. Let $[\mathbf{x}]_k$ represent the narrow band signal of the k -th person's speech in a group, and $\mathbf{y}_i = \mathbf{H}_i \mathbf{x} + \mathbf{n}_i$ be the mixed signals received by the i -th microphone array, where $\mathbf{n}_i$ is a complex Gaussian noise, and $\mathbf{H}_i$ is a mixing matrix, where $[\mathbf{H}_i]_{j,k} = \alpha_i e^{i\theta_{ijk}}$ with $i = \sqrt{-1}$, α_i being the gain and θ_{ijk} being the phase shift of the received signal of the k -th person's speech at the j -th sensor of the i -th microphone array. The microphone arrays send the received signal to a cloud service to decode the speech of a subgroup of persons in the index set $\mathcal{P}$. Meanwhile, we wish to protect the speech signals of people in another distinct group indexed by $\mathcal{Q}$ from the cloud service. Accordingly, $\mathbf{U}$ is diagonal matrix with 1's on the diagonal at the row indices in the index set $\mathcal{P}$ and zero everywhere else, and for each $i = 1, \dots, S$, $\mathbf{G}_i$ is a diagonal matrix with 1's on the diagonal at the row indices in the index set $\mathcal{Q}$ and zero everywhere else.

We assume all system parameters and model (including $\mathbf{H}$, $\mathbf{R}$, $\mathbf{U}$, $\mathbf{G}$, etc.) are known to the fusion center. To protect the private parameters $\{\mathbf{g}_i : i = 1, \dots, S\}$ from being inferred by the fusion center based on the collective measurements, each agent sanitizes its local observation before transmitting to the fusion center. Let $\mathcal{T}(\mathbf{y}) = [\mathcal{T}_1(\mathbf{y}_1)^\top, \dots, \mathcal{T}_S(\mathbf{y}_S)^\top]^\top$ denote the sanitized information received at the

FIG. 3.1: Microphone array system.

fusion center, where $\mathcal{T}$ belongs to a predefined class of sanitization mechanisms. We measure the utility and privacy by the [CRLBs](#) for estimating the public parameter $\mathbf{u}$ and private parameters $\{\mathbf{g}_i : i = 1, \dots, S\}$, respectively. Recall that the [CRLB](#) [114, 115] is a lower bound on the variance of an unbiased estimator. Suppose $p_{\mathbf{y}|\mathbf{x}}$ is a probability density function of a random variable $\mathbf{y}$ conditioned on $\mathbf{x}$, and $\mathbf{x} \sim p_{\mathbf{x}}$ (prior information). For any unbiased estimator $\hat{\mathbf{x}}(\mathbf{y})$ of $\mathbf{x}$ based on $\mathbf{y}$, we have $\mathbb{E}[(\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x})(\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x})^\top] \succeq (\mathbf{J}_{\mathbf{x}} + \mathbf{J}_0)^{-1}$, where

$$\mathbf{J}_{\mathbf{x}} = -\mathbb{E} \left[\frac{\partial^2 \log(p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}))}{\partial \mathbf{x}^2} \right], \quad \mathbf{J}_0 = -\mathbb{E} \left[\frac{\partial^2 \log(p(\mathbf{x}))}{\partial \mathbf{x}^2} \right] \quad (3.3)$$

with $(\mathbf{J}_{\mathbf{x}} + \mathbf{J}_0)^{-1}$ known as the [CRLB](#) for estimating $\mathbf{x}$. If $\mathbf{x}$ is a deterministic parameter ($\mathbf{J}_0 = \mathbf{0}$), we have $\text{cov}(\hat{\mathbf{x}}(\mathbf{y})) \succeq \mathbf{J}_{\mathbf{x}}^{-1}$ for any unbiased estimator $\hat{\mathbf{x}}(\mathbf{y})$. In other words, no unbiased estimator can outperform the [CRLB](#) in terms of error variance.

Denoting the [CRLB](#) of the system variable $\mathbf{x}$ before sanitization as $\mathbf{P}_{\mathbf{x}}$ and after sanitization as $\tilde{\mathbf{P}}_{\mathbf{x}}$, the [CRLBs](#) for the public parameter $\mathbf{u}$ before and after sanitization at the agents are

$$\begin{aligned} \mathbf{P}_{\mathbf{u}} &= \mathbf{U} \mathbf{P}_{\mathbf{x}} \mathbf{U}^\top, \\ \tilde{\mathbf{P}}_{\mathbf{u}} &= \mathbf{U} \tilde{\mathbf{P}}_{\mathbf{x}} \mathbf{U}^\top, \end{aligned} \quad (3.4)$$

respectively. Similarly, the CRLBs for the private parameter $\mathbf{g}_i$ before and after sanitization at the agents are

$$\begin{aligned}\mathbf{P}_{\mathbf{g}_i} &= \mathbf{G}_i \mathbf{P}_{\mathbf{x}} \mathbf{G}_i^\top, \\ \tilde{\mathbf{P}}_{\mathbf{g}_i} &= \mathbf{G}_i \tilde{\mathbf{P}}_{\mathbf{x}} \mathbf{G}_i^\top,\end{aligned}\tag{3.5}$$

respectively.

The quantity $\text{Tr}(\mathbf{P}_{\mathbf{u}})$ gives a lower bound for the sum error variance for estimating every component of $\mathbf{u}$. For each sanitization function $\mathcal{T}$, we define the system utility function for the public parameter $\mathbf{u}$ as

$$u(\mathcal{T}) = 1 - \frac{\text{Tr}(\tilde{\mathbf{P}}_{\mathbf{u}})}{\text{Tr}(\mathbf{P}_{\mathbf{u}})} \leq 0,\tag{3.6}$$

which is the negative of the percentage increase in sum error variance lower bound for estimating $\mathbf{u}$ due to privacy sanitization. We define the system privacy function for the private parameter $\mathbf{g}_i$, $i = 1, \dots, S$, as

$$p_i(\mathcal{T}) = \frac{\text{Tr}(\tilde{\mathbf{P}}_{\mathbf{g}_i})}{\text{Tr}(\mathbf{P}_{\mathbf{g}_i})} - 1,\tag{3.7}$$

which is the percentage increase in sum error variance lower bound for estimating $\mathbf{g}_i$ due to privacy sanitization. Since $\mathcal{T}$ can be viewed as part of the estimation procedure, we have $\tilde{\mathbf{P}}_{\mathbf{x}} \succ \mathbf{P}_{\mathbf{x}}$. Therefore, $u(\mathcal{T}) \leq 0$ and $p_i(\mathcal{T}) \geq 0$ for all $\mathcal{T}$. The utility and privacy functions as defined are thus intuitive: perturbation on the measurement decreases the utility and increases the privacy. Our goal is to

$$\begin{aligned}\max_{\mathcal{T}} \quad & u(\mathcal{T}), \\ \text{s.t.} \quad & p_i(\mathcal{T}) \geq \epsilon_i, \quad i = 1, \dots, S,\end{aligned}\tag{P3.0}$$

where $\epsilon_i \geq 0$ is called the *privacy threshold* for agent i . We summarize the notations introduced so far in Table 3.1 for the reader's convenience.

TABLE 3.1: Summary of commonly-used symbols

Notation	Definition
$\mathbf{y}_i$	The agent i 's measurements: $\mathbf{y}_i = \mathbf{H}_i \mathbf{x} + \mathbf{n}_i \in \mathbb{R}^{N_i}$, where $\mathbf{x} \in \mathbb{R}^L$ and $\mathbf{n}_i \sim \mathcal{N}(\mathbf{0}_{N_i}, \mathbf{R}_i)$.
$\mathbf{y}, \mathcal{T}(\mathbf{y})$	The raw measurements and sanitized measurements from all agents $1, \dots, S$.
$\mathbf{R}, \mathbf{H}$	The agents' measurements noise covariance matrix $\mathbf{R} \in \mathbb{S}^N$ and observation matrix $\mathbf{H} \in \mathbb{R}^{N \times L}$.
$\mathcal{S}_i$	The index set corresponding to agent i , i.e., $[\mathbf{R}]_{\mathcal{S}_i, \mathcal{S}_i} = \mathbf{R}_i$ and $[\mathbf{H}]_{\mathcal{S}_i, :} = \mathbf{H}_i$.
$\mathbf{u}, \mathbf{g}_i, \mathbf{U}, \mathbf{G}_i$	The public parameter $\mathbf{u} = \mathbf{U} \mathbf{x} \in \mathbb{R}^U$ and the i -th private parameter $\mathbf{g}_i = \mathbf{G}_i \mathbf{x} \in \mathbb{R}^{G_i}$.
$\mathbf{P}_{\mathbf{x}}$	The CRLB for estimating $\mathbf{x}$ from the raw measurements $\mathbf{y}$.
$\tilde{\mathbf{P}}_{\mathbf{x}}, \tilde{\mathbf{P}}_{\mathbf{u}}, \tilde{\mathbf{P}}_{\mathbf{g}_i}$	The CRLB for estimating $\mathbf{x}$, $\mathbf{u}$ and $\mathbf{g}_i$, respectively, from sanitized data $\mathcal{T}(\mathbf{y})$.
$u(\mathcal{T}), p_i(\mathcal{T})$	The utility function for $\mathbf{u}$ and privacy function for $\mathbf{g}_i$, respectively, using sanitization function $\mathcal{T}$.
ϵ_i	The privacy threshold assigned to agent i .
Φ	If $\mathbf{J}_0 = \mathbf{0}_L$, $\Phi = \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{H} \mathbf{P}_{\mathbf{x}} \mathbf{H}^T \mathbf{R}^{-1}$. Otherwise, $\Phi = (\mathbf{H} \mathbf{P}_0 \mathbf{H}^T + \mathbf{R})^{-1}$. Cf. (3.21) and (3.25).
Ψ	If $\mathbf{J}_0 = \mathbf{0}_L$, $\Psi = \mathbf{P}_0 \mathbf{H}^T \Phi$. Otherwise, $\Psi = \mathbf{P}_{\mathbf{x}} \mathbf{H}^T \mathbf{R}^{-1}$. Cf. (3.22) and (3.26).

3.2 Privacy-Preserving Transformation

In this section, we study two sanitization schemes $\mathcal{T}$ that are widely used in the privacy literature: 1) linear compression that reduces the dimension of the measurements [60], and 2) noise perturbation that adds random noise to the measurements [87]. We derive the relationship between these two schemes under CRLB-based privacy criteria like that in (3.7). We show that the noise perturbation scheme is equivalent to linear compression in an asymptotic sense.

To prevent the fusion center or potential adversarial agents from inferring the set of private parameters $\mathbf{g}_1, \dots, \mathbf{g}_S$ through the measurement $\mathbf{y}$, we consider an affine transformation $\mathcal{T}_i : \mathbb{R}^{N_i} \mapsto \mathbb{R}^{M_i}$ where $M_i \leq N_i$ to sanitize the measurement

of agent i to obtain:

$$\mathcal{T}_i(\mathbf{y}_i) = \mathbf{C}_i \mathbf{y}_i + \boldsymbol{\xi}_i = \tilde{\mathbf{H}}_i \mathbf{x}_i + \tilde{\mathbf{n}}_i + \boldsymbol{\xi}_i, \quad (3.8)$$

where $\mathbf{C}_i \in \mathbb{R}^{M_i \times N_i}$ is called the compression matrix and $\boldsymbol{\xi}_i \in \mathbb{R}^{M_i}$ is a perturbation noise with $\boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0}_{N_i}, \boldsymbol{\Theta}_i)$. We also have $\tilde{\mathbf{H}}_i = \mathbf{C}_i \mathbf{H}_i \in \mathbb{R}^{M_i \times L}$, $\tilde{\mathbf{n}}_i = \mathbf{C}_i \mathbf{n}_i \in \mathbb{R}^{M_i}$ and $\tilde{\mathbf{n}}_i \sim \mathcal{N}(\mathbf{0}_{N_i}, \tilde{\mathbf{R}}_i)$ with $\tilde{\mathbf{R}}_i = \mathbf{C}_i \mathbf{R}_i \mathbf{C}_i^\top$. By collecting all the perturbed measurements, the measurement model at the fusion center can be summarized as follows:

$$\mathcal{T}(\mathbf{y}) = \mathbf{C} \mathbf{y} + \boldsymbol{\xi} = \tilde{\mathbf{H}} \mathbf{x} + \tilde{\mathbf{n}} + \boldsymbol{\xi}, \quad (3.9)$$

where $\mathcal{T}(\mathbf{y}) = [\mathcal{T}_1(\mathbf{y}_1)^\top, \dots, \mathcal{T}_S(\mathbf{y}_S)^\top]^\top \in \mathbb{R}^M$, $\tilde{\mathbf{H}} = \mathbf{C} \mathbf{H}$, $M = \sum_{i=1}^S M_i$, $\tilde{\mathbf{n}} \sim \mathcal{N}(\mathbf{0}_N, \mathbf{C} \mathbf{R} \mathbf{C}^\top)$, $\boldsymbol{\xi}^\top = [\boldsymbol{\xi}_1^\top, \dots, \boldsymbol{\xi}_S^\top] \in \mathbb{R}^M$, $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}_N, \boldsymbol{\Theta})$, $\mathbf{C} = \text{diag}(\mathbf{C}_1, \dots, \mathbf{C}_S)$, $\boldsymbol{\Theta} = \text{diag}(\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_S)$.

Note that the global sanitization function $\mathcal{T}$ is composed of independent local sanitization functions $\mathcal{T}_1, \dots, \mathcal{T}_S$, the i -th of which applies a local linear compression matrix $\mathbf{C}_i$ and additive noise $\boldsymbol{\xi}_i$ with covariance $\boldsymbol{\Theta}_i$ to the i -th agent's measurement $\mathbf{y}_i$. Both the global compression matrix $\mathbf{C}$ and additive noise covariance matrix $\boldsymbol{\Theta}$ have block diagonal form. There is no message exchange between agents during the sanitization process as this is prone to privacy attacks.

From (3.6) and (3.7), we can express the utility and privacy function w.r.t. $\mathbf{C}$ and $\boldsymbol{\Theta}$ as

$$u(\mathbf{C}, \boldsymbol{\Theta}) = 1 - \frac{\text{Tr}(\mathbf{U} \tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta}) \mathbf{U}^\top)}{\text{Tr}(\mathbf{U} \mathbf{P}_{\mathbf{x}} \mathbf{U}^\top)}, \quad (3.10)$$

$$p_i(\mathbf{C}, \boldsymbol{\Theta}) = \frac{\text{Tr}(\mathbf{G}_i \tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta}) \mathbf{G}_i^\top)}{\text{Tr}(\mathbf{G}_i \mathbf{P}_{\mathbf{x}} \mathbf{G}_i^\top)} - 1, \quad (3.11)$$

where

$$\begin{aligned}\tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta}) &= \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta})^{-1} \\ &= (\mathbf{J}_0 + \mathbf{H}^\top \mathbf{C}^\top (\mathbf{C} \mathbf{R} \mathbf{C}^\top + \boldsymbol{\Theta})^{-1} \mathbf{C} \mathbf{H})^{-1}.\end{aligned}\quad (3.12)$$

Here, $\mathbf{J}_0$ is the [FIM](#) of any prior information. We let $\mathbf{J}_0 = \mathbf{0}_L$ if no prior information is available. There is no loss in generality if we restrict compression matrices $\mathbf{C}_i$ s to be square matrices, i.e., we consider $(\mathbf{C}, \boldsymbol{\Theta})$ to be chosen from the following sanitization parameter set

$$\begin{aligned}\mathcal{C} = \left\{ (\mathbf{C}, \boldsymbol{\Theta}) \mid \mathbf{C} = \text{diag}(\mathbf{C}_1, \dots, \mathbf{C}_S) \in \mathbb{R}^{N \times N}, \right. \\ \boldsymbol{\Theta} = \text{diag}(\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_S), \\ \left. \mathbf{C}_i \in \mathbb{R}^{N_i \times N_i}, \boldsymbol{\Theta}_i \in \mathbb{S}^{N_i}, i = 1, \dots, S \right\}.\end{aligned}\quad (3.13)$$

This is because the perturbed [CRLB](#) remains unaltered by padding a non-square $\mathbf{C}$ and $\boldsymbol{\Theta}$ with zeroes: $\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta}) = \tilde{\mathbf{J}}_{\mathbf{x}}(\bar{\mathbf{C}}, \bar{\boldsymbol{\Theta}})$, where $\bar{\mathbf{C}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{0} \end{bmatrix}$, $\bar{\boldsymbol{\Theta}} = \begin{bmatrix} \boldsymbol{\Theta} & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, and $\mathbf{0}$ and $\mathbf{0}'$ are $(N - M) \times N$ and $M \times (N - M)$ zero matrices respectively.

In the following, we prove some properties of $\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta})$ with the domain $\mathcal{C}$. We consider the metric spaces $\mathbb{R}^{N \times N} \times \mathbb{S}^N$ and $\mathbb{S}^L$ endowed with the Frobenius norm so that $\mathcal{C}$ is a subset of $\mathbb{R}^{N \times N} \times \mathbb{S}^N$ and $\tilde{\mathbf{P}}_{\mathbf{x}}(\cdot, \cdot)$ is a mapping from $\mathbb{R}^{N \times N} \times \mathbb{S}^N$ to $\mathbb{S}^L$. Notions of metric properties like continuity are defined [w.r.t.](#) these metric spaces.

Lemma 1. Let $(\mathbf{C}, \boldsymbol{\Theta}) \in \mathcal{C}$, where $\mathbf{C} \in \mathbb{R}^{N \times N}$ and $\boldsymbol{\Theta} \in \mathbb{S}^{N \times N}$.

- (i) $\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta})$ is a continuous function on $\mathcal{C}$.
- (ii) Suppose $\mathbf{A}$ is either a unitary matrix or an invertible block diagonal matrix. Then, we have $\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{A}\mathbf{C}, \boldsymbol{\Theta}) = \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \mathbf{A}^\top \boldsymbol{\Theta} \mathbf{A})$.

Proof. The first claim follows immediately from (3.12). For the second claim, suppose first that $\mathbf{A}$ is unitary. We then have

$$\begin{aligned}\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{A}\mathbf{C}, \boldsymbol{\Theta}) &= \mathbf{J}_0 + \mathbf{H}^T \mathbf{C}^T \mathbf{A}^T (\mathbf{A}\mathbf{C}\mathbf{R}\mathbf{C}^T \mathbf{A}^T + \boldsymbol{\Theta})^{-1} \mathbf{A}\mathbf{C}\mathbf{H} \\ &= \mathbf{J}_0 + \mathbf{H}^T \mathbf{C}^T (\mathbf{C}\mathbf{R}\mathbf{C}^T + \mathbf{A}^T \boldsymbol{\Theta} \mathbf{A})^{-1} \mathbf{C}\mathbf{H} \\ &= \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \mathbf{A}^T \boldsymbol{\Theta} \mathbf{A}).\end{aligned}\tag{3.14}$$

The proof for the case where $\mathbf{A}$ is an invertible block diagonal matrix is similar and the proof is complete. $\square$

Proposition 1. (i) Let

$$\mathcal{A} = \left\{ \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{I}_N, \boldsymbol{\Theta}) \mid (\mathbf{I}_N, \boldsymbol{\Theta}) \in \mathcal{C} \right\} \text{ and } \mathcal{B} = \left\{ \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta}) \mid (\mathbf{C}, \boldsymbol{\Theta}) \in \mathcal{C} \right\}.\tag{3.15}$$

Then, $\mathcal{B} \setminus \mathcal{A}$ is the boundary of $\mathcal{A}$.

(ii) For any $(\mathbf{C}, \boldsymbol{\Theta}) \in \mathcal{C}$, there exists $(\mathbf{C}', \boldsymbol{\Lambda}_b) \in \mathcal{C}$, where $\boldsymbol{\Lambda}_b$ is a diagonal matrix with $[\boldsymbol{\Lambda}_b]_{i,i} \in \{0, 1\}$, for all $i \geq 1$, such that

$$\tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{C}', \boldsymbol{\Lambda}_b) = \tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta}).\tag{3.16}$$

Proof. (i) Since $\mathcal{A} \subset \mathcal{B}$, it suffices to show that for any $\mathbf{B} \in \mathcal{B}$, there exists a sequence in $\mathcal{A}$ that converges to $\mathbf{B}$. From Lemma 1(ii), we have $\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta}) = \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{I}_N, \mathbf{C}^T \boldsymbol{\Theta} \mathbf{C})$ with $\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{I}_N, \mathbf{C}^T \boldsymbol{\Theta} \mathbf{C}) \in \mathcal{A}$ if $\mathbf{C}$ is invertible. On the other hand, if $\mathbf{C}$ is a singular matrix, we let the eigendecomposition of $\mathbf{C} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}$, where $\mathbf{Q}$ is the square matrix whose columns are the eigenvector, and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. Denote $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}', \mathbf{0}_A)$ with $\boldsymbol{\Lambda}'$ containing all the non-zero eigenvalues and an appropriate A . From Lemma 1(ii), we have

$$\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \boldsymbol{\Theta}) = \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}, \boldsymbol{\Theta}) = \tilde{\mathbf{J}}_{\mathbf{x}}(\boldsymbol{\Lambda}\mathbf{Q}^{-1}, \boldsymbol{\Theta}'),\tag{3.17}$$

where $\Theta' = \mathbf{Q}^\top \Theta \mathbf{Q}$. For any $\lambda > 0$, let $\Lambda_\lambda = \text{diag}(\Lambda', \lambda \mathbf{I}_A)$, which is an invertible diagonal matrix. From Lemma 1(ii), we have

$$\tilde{\mathbf{J}}_{\mathbf{x}}(\Lambda_\lambda \mathbf{Q}^{-1}, \Theta') = \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{I}_N, (\Lambda_\lambda \mathbf{Q}^{-1})^\top \Theta' \Lambda_\lambda \mathbf{Q}^{-1}) \in \mathcal{A}, \quad (3.18)$$

whose left-hand side converges to the right-hand side of (3.17) as $\lambda \rightarrow 0$ since $\tilde{\mathbf{J}}_{\mathbf{x}}$ is continuous by Lemma 1(i). The proof is now complete.

- (ii) Let the eigendecomposition of $\Theta = \mathbf{Q} \Lambda \mathbf{Q}^\top$. The diagonal matrix Λ can be written as $\Lambda_+^{1/2} \Lambda_b \Lambda_+^{1/2}$, where Λ_+ and Λ_b are diagonal matrices with $[\Lambda_+]_{i,i} = [\Lambda]_{i,i}$, $[\Lambda_b]_{i,i} = 1$ when $[\Lambda]_{i,i} > 0$, and $[\Lambda_+]_{i,i} = 1$, $[\Lambda_b]_{i,i} = 0$ when $[\Lambda]_{i,i} = 0$. Two applications of Lemma 1(ii) gives us $\tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}, \Theta) = \tilde{\mathbf{J}}_{\mathbf{x}}(\mathbf{C}', \Lambda_b)$, where $\mathbf{C}' = \Lambda_+^{-1/2} \mathbf{Q}^\top \mathbf{C}$ is a block diagonal matrix since $\mathbf{Q}$, Λ_+ and $\mathbf{C}$ are all block diagonal matrices. Thus $(\mathbf{C}', \Lambda_b) \in \mathcal{C}$ and the claim follows.

□

Proposition 1(ii) shows that the power of the additive noise can be normalized by the compression matrix when sanitizing the data. On the other hand, Proposition 1(i) shows that adding only noise can approximate linear compression arbitrarily well in terms of the estimation error covariance, but this requires arbitrarily large noise power. Therefore, perturbation noise by itself cannot replace linear compression in a practical system. Without loss of generality, we set the compression matrix $\mathbf{C} = \mathbf{I}_N$ in the sequel as we can always normalize the noise by a compression matrix to obtain the same utility and privacy tradeoff. We now rewrite the optimization problem (P3.0) as

$$\begin{aligned} & \max_{(\mathbf{I}_N, \Theta) \in \mathcal{C}} u(\mathbf{I}_N, \Theta), \\ & \text{s.t. } p_i(\mathbf{I}_N, \Theta) \geq \epsilon_i, i = 1, \dots, S. \end{aligned} \quad (\text{P3.1})$$

Note that for (P3.1) to be feasible when $\mathbf{J}_0 \neq \mathbf{0}_L$, the privacy threshold ϵ_i should satisfy $\epsilon_i \leq \epsilon_i^{\max}$, where

$$\epsilon_i^{\max} = \text{Tr}(\mathbf{G}_i \mathbf{J}_0^{-1} \mathbf{G}_i^\top) / \text{Tr}(\mathbf{G}_i \mathbf{P}_x \mathbf{G}_i^\top) - 1, \quad (3.19)$$

for $\mathbf{G}_i \neq \mathbf{0}$, and $\epsilon_i^{\max} = 0$ if $\mathbf{G}_i = \mathbf{0}$. This is because the prior information (as quantified by its FIM $\mathbf{J}_0$) already leaks privacy and any sanitization cannot achieve a higher level of privacy than this. When $\mathbf{J}_0 = \mathbf{0}_L$, we have $\epsilon_i^{\max} = \infty$ if $\mathbf{G}_i \neq \mathbf{0}$ and 0 otherwise.

The following useful expressions can be derived by using the binomial inverse theorem and Woodbury matrix identity (see Appendix 3.A for details).

1. No prior information, i.e., $\mathbf{J}_0 = \mathbf{0}_L$, we have

$$\tilde{\mathbf{P}}_x(\mathbf{I}_N, \boldsymbol{\Theta}) = \mathbf{P}_x + \boldsymbol{\Psi} (\mathbf{I}_N + \boldsymbol{\Theta} \boldsymbol{\Phi})^{-1} \boldsymbol{\Theta} \boldsymbol{\Psi}^\top, \quad (3.20)$$

where

$$\boldsymbol{\Phi} = \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{H} \mathbf{P}_x \mathbf{H}^\top \mathbf{R}^{-1} \quad (3.21)$$

is a degenerate matrix and

$$\boldsymbol{\Psi} = \mathbf{P}_x \mathbf{H}^\top \mathbf{R}^{-1}. \quad (3.22)$$

2. With prior information, i.e., $\mathbf{J}_0 \succ \mathbf{0}_L$, we have

$$\tilde{\mathbf{P}}_x(\mathbf{I}_N, \boldsymbol{\Theta}) = \mathbf{P}_0 - \mathbf{P}_0 \mathbf{H}^\top (\boldsymbol{\Phi}^{-1} + \boldsymbol{\Theta})^{-1} \mathbf{H} \mathbf{P}_0 \quad (3.23)$$

$$= \mathbf{P}_x + \boldsymbol{\Psi} (\mathbf{I}_N + \boldsymbol{\Theta} \boldsymbol{\Phi})^{-1} \boldsymbol{\Theta} \boldsymbol{\Psi}^\top, \quad (3.24)$$

where $\mathbf{P}_0 = \mathbf{J}_0^{-1}$,

$$\boldsymbol{\Phi} = (\mathbf{H} \mathbf{P}_0 \mathbf{H}^\top + \mathbf{R})^{-1}, \quad (3.25)$$

and

$$\Psi = \mathbf{P}_0 \mathbf{H}^\top \Phi. \quad (3.26)$$

By considering the observation model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, Ψ in (3.20) is the same as the minimum mean square error estimation matrix. By regarding $\mathbf{P}_0$ as the predicted estimate covariance, Ψ and Φ^{-1} in (3.24) are equivalent to the Kalman gain and innovation covariance, respectively. The equations (3.20) and (3.24) under both the cases where prior information is available and unavailable have meaningful interpretations: the first term of both equations is the CRLB without perturbation while the second term is the increased CRLB caused by noise perturbation.

3.3 Arbitrarily Strong Privacy with Perfect Utility

In this section, we derive necessary and sufficient conditions for achieving perfect utility and arbitrarily strong privacy when prior information is available and unavailable, respectively.

Definition 9. We say that arbitrarily strong utility-privacy tradeoff (ASUP) is achievable if for any non-negative $\epsilon_i < \epsilon_i^{\max}$, $i = 1, \dots, S$, there exists $(\mathbf{C}, \Theta) \in \mathcal{C}$ such that $u(\mathbf{C}, \Theta) = 0$ and $p_i(\mathbf{C}, \Theta) \geq \epsilon_i$ for all $i = 1, \dots, S$.

While achieving perfect utility (by taking $\mathcal{T}$ to be $(\mathbf{I}_N, \mathbf{0}_N)$) and arbitrarily strong privacy (by taking $\mathcal{T}$ to be $(\mathbf{0}_N, \mathbf{0}_N)$) are easy, achieving both at the same time is difficult or infeasible. Therefore, it is of interest to know when ASUP is achievable. In the following, we derive necessary and sufficient conditions for ASUP under both the cases where prior information is available and unavailable. Moreover, we provide a method to construct the noise covariance to achieve ASUP inside the proofs (if ASUP is feasible). We start with the following two preliminary lemmas.

Lemma 2. For matrices $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\mathbf{B} \in \mathbb{S}^N$ and a sequence of pairwise commuting matrices $(\mathbf{S}_n)_{n \geq 1}$ where each $\mathbf{S}_n \in \mathbb{S}_+^N$, we have

$$\lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A}(\mathbf{S}_n + \mathbf{B})^{-1} \mathbf{A}^\top) = 0, \quad (3.27)$$

if and only if

$$\lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} \mathbf{S}_n^{-1} \mathbf{A}^\top) = 0. \quad (3.28)$$

Proof. See Appendix 3.B. □

Lemma 3. For matrices $\mathbf{A} \in \mathbb{R}^{M_1 \times N}$ and $\mathbf{B} \in \mathbb{R}^{M_2 \times N}$ such that $M_1, M_2 \leq N$, there exists $\mathbf{S} \in \mathbb{S}^N$ such that $\text{Tr}(\mathbf{A} \mathbf{S} \mathbf{A}^\top) = 0$ and

$$\lim_{n \rightarrow \infty} \text{Tr} \left(\mathbf{B} \left(n \mathbf{S} + \frac{1}{n} \mathbf{I}_N \right)^{-1} \mathbf{B}^\top \right) = 0, \quad (3.29)$$

if and only if $\mathbf{A} \mathbf{B}^\top = \mathbf{0}$.

Proof. See Appendix 3.C. □

3.3.1 ASUP without prior information

In this subsection, we consider the case $\mathbf{J}_0 = \mathbf{0}_L$.

Theorem 2. Suppose $\mathbf{J}_0 = \mathbf{0}_L$. Recall Φ and Ψ as defined in (3.21) and (3.22), respectively. For $i \in \{1, \dots, S\}$, let $\Xi_i = \begin{bmatrix} \mathbf{U}[\Psi]_{:,S_i} \\ [\Phi]_{S_i,S_i} \end{bmatrix}$. There exists $\Theta \in \mathbb{S}^N$ such that ASUP is achievable if and only if for each $\mathbf{G}_j \neq \mathbf{0}$ where $j \in \{1, \dots, S\}$, there exists at least an agent $i \in \{1, \dots, S\}$ such that

- (i) $N_i > \text{rank}(\Xi_i)$, and
- (ii) there does not exist any matrix $\mathbf{P}$ such that $\mathbf{P} \Xi_i = \mathbf{G}_j [\Psi]_{:,S_i}$.

Proof. Note that condition (i) is equivalent to

$$\text{Null} \left([\Phi]_{\mathcal{S}_i, \mathcal{S}_i} \right) \cap \text{Null} \left(\mathbf{U} [\Psi]_{:, \mathcal{S}_i} \right) \neq \{\mathbf{0}\}, \quad (3.30)$$

and condition (ii) is equivalent to

$$\begin{aligned} & \left(\text{Null} \left([\Phi]_{\mathcal{S}_i, \mathcal{S}_i} \right) \cap \text{Null} \left(\mathbf{U} [\Psi]_{:, \mathcal{S}_i} \right) \right) \\ & \not\subseteq \text{Null} \left(\mathbf{G}_j [\Psi]_{:, \mathcal{S}_i} \right). \end{aligned} \quad (3.31)$$

We first show the necessity of conditions (i) and (ii). From (3.6) and (3.20), perfect utility $u(\mathbf{I}_N, \Theta) = 0$ is obtained only if

$$\text{Tr} \left(\mathbf{U} \Psi (\mathbf{I}_N + \Theta \Phi)^{-1} \Theta \Psi^\top \mathbf{U}^\top \right) = 0. \quad (3.32)$$

Since the left-hand side (L.H.S.) of the above equation is a continuous function w.r.t. Θ , the above equation can be expressed as

$$\lim_{n \rightarrow \infty} \text{Tr} \left(\mathbf{U} \Psi \left(\left(\Theta + \frac{1}{n} \mathbf{I}_N \right)^{-1} + \Phi \right)^{-1} \Psi^\top \mathbf{U}^\top \right) = 0. \quad (3.33)$$

From Lemma 2, this is equivalent to

$$\lim_{n \rightarrow \infty} \text{Tr} \left(\mathbf{U} \Psi \left(\Theta + \frac{1}{n} \mathbf{I}_N \right) \Psi^\top \mathbf{U}^\top \right) = 0, \quad (3.34)$$

or

$$\text{Tr} (\mathbf{U} \Psi \Theta \Psi^\top \mathbf{U}^\top) = 0. \quad (3.35)$$

Therefore, $u(\mathbf{I}_N, \Theta) = 0$ only if $\mathbf{U} \Psi \Theta = \mathbf{0}$. Following that, $\mathbf{U} [\Psi]_{:, \mathcal{S}_k} \Theta_k = \mathbf{0}$, for all $k = 1, \dots, S$ since $\Theta = \text{diag}(\Theta_1, \dots, \Theta_S)$. From the rank-nullity theorem, this implies that $\Theta_k = \mathbf{0}_{N_k}$ for an agent k with $\text{rank}(\mathbf{U} [\Psi]_{:, \mathcal{S}_k}) = N_k$. Since $\mathbf{G}_j \neq \mathbf{0}$, to achieve any level of privacy, there must exist at least an agent i with $\text{rank}(\mathbf{U} [\Psi]_{:, \mathcal{S}_i}) < N_i$ with $\mathbf{U} [\Psi]_{:, \mathcal{S}_i}$ having a non-trivial null space. Let Θ be such

that $\Theta_i \neq \mathbf{0}_{N_i}$ and

$$\mathbf{U}[\Psi]_{:,S_i} \Theta_i = \mathbf{0}. \quad (3.36)$$

Since $p_j(\mathbf{I}_N, \Theta)$ is a non-decreasing function w.r.t. Θ , arbitrarily strong privacy for the private parameter $\mathbf{g}_j$ is achieved only if

$$\lim_{n \rightarrow \infty} p_j(\mathbf{I}_N, n\Theta) = \infty. \quad (3.37)$$

From (3.7) and (3.20), the L.H.S. of the above statement (3.37) is equivalent to

$$\begin{aligned} & \lim_{n \rightarrow \infty} \text{Tr} (\mathbf{G}_j \Psi (\mathbf{I}_N + n\Theta \Phi)^{-1} n\Theta \Psi^\top \mathbf{G}_j^\top) \\ &= \lim_{n \rightarrow \infty} \text{Tr} \left(\mathbf{G}_j \Psi \left(\left(n\Theta + \frac{1}{n} \mathbf{I}_N \right)^{-1} + \Phi \right)^{-1} \Psi^\top \mathbf{G}_j^\top \right) \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^N \alpha_k(n, j) \lambda_k(n), \end{aligned} \quad (3.38)$$

where $\mathbf{T}(n) = \left(\left(n\Theta + \frac{1}{n} \mathbf{I}_N \right)^{-1} + \Phi \right)^{-1}$, $\lambda_k(n)$ is the k -th eigenvalue of $\mathbf{T}(n)$ with $\mathbf{v}_k(n)$ being the corresponding unit eigenvector, and $\alpha_k(n, j) = \|\mathbf{G}_j \Psi \mathbf{v}_k(n)\|_2^2$. Therefore, (3.37) holds only if there exists an index k_j such that $\lim_{n \rightarrow \infty} \lambda_{k_j}(n) = \infty$ with corresponding eigenvectors $\mathbf{v}_{k_j}(n) \notin \text{Null}(\mathbf{G}_j \Psi)$. Since $\mathbf{G}_j \Psi$ has finite rank, by passing to a subsequence if necessary, there exists a unit vector $\mathbf{w}_j \notin \text{Null}(\mathbf{G}_j \Psi)$ such that

$$\lim_{n \rightarrow \infty} \mathbf{w}_j^\top (\mathbf{T}(n))^{-1} \mathbf{w}_j = 0. \quad (3.39)$$

We then have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbf{w}_j^\top \left(n\Theta + \frac{1}{n} \mathbf{I}_N \right)^{-1} \mathbf{w}_j = 0, \text{ and} \\ & \mathbf{w}_j^\top \Phi \mathbf{w}_j = 0. \end{aligned} \quad (3.40)$$

Since $\Theta_i \neq \mathbf{0}_{N_i}$, (3.40) holds only if $\mathbf{w}'_j = [\mathbf{w}_j]_{\mathcal{S}_i} \notin \text{Null}(\mathbf{G}_j[\Psi]_{:, \mathcal{S}_i})$ and $\mathbf{w}'_j \neq \mathbf{0}$, such that

$$\lim_{n \rightarrow \infty} \mathbf{w}'_j{}^\top \left(n\Theta_i + \frac{1}{n}\mathbf{I}_{N_i} \right)^{-1} \mathbf{w}'_j = 0, \text{ and} \quad (3.41)$$

$$[\Phi]_{\mathcal{S}_i, \mathcal{S}_i} \mathbf{w}'_j = 0. \quad (3.42)$$

From Lemma 3, both (3.36) and (3.41) hold only if $\mathbf{w}'_j \in \text{Null}(\mathbf{U}[\Psi]_{:, \mathcal{S}_i})$. Therefore, conditions (i) and (ii) are necessary for ASUP.

We next prove sufficiency by constructing a Θ that yields ASUP. Let a unit vector $\mathbf{v}'$ satisfy

- (a) $\mathbf{v}' \in \text{Null}([\Phi]_{\mathcal{S}_i, \mathcal{S}_i})$, and
- (b) $\mathbf{v}' \in \text{Null}(\mathbf{U}[\Psi]_{:, \mathcal{S}_i})$, and
- (c) $\mathbf{v}' \notin \text{Null}(\mathbf{G}_j[\Psi]_{:, \mathcal{S}_i})$.

The existence of $\mathbf{v}'$ is guaranteed by condition i and ii. Set all the eigenvalues of Θ_i to 0 except for one eigenvalue $\lambda' > 0$ associated with a unit eigenvector $\mathbf{v}'$. Let $\tilde{\Theta}_j = \text{diag}(\dots, \mathbf{0}_{N_{i-1}}, \Theta_i, \mathbf{0}_{N_{i+1}} \dots)$. By reversing the arguments used in the necessity proof, it can be verified that for any $\epsilon_j > 0$, there exists a large enough λ' such that $u(\mathbf{I}_N, \tilde{\Theta}_j) = 0$ and $p_j(\mathbf{I}_N, \tilde{\Theta}_j) \geq \epsilon_j$. Let Θ be the sum of $\tilde{\Theta}_j$ over j and such Θ leads to ASUP. The proof is now complete. $\square$

As an illustration, consider the special case $\mathbf{H} \in \mathbb{R}^{L \times L}$, where Ψ turns out to be $\mathbf{H}^{-1}$ and Φ becomes $\mathbf{0}_N$. Then from (3.20), the perturbed CRLB for estimating $\mathbf{x}$ can be written as

$$\begin{aligned} \tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{I}_N, \Theta) &= \mathbf{P}_{\mathbf{x}} + \mathbf{H}^{-1} \Theta (\mathbf{H}^\top)^{-1} \\ &= \mathbf{P}_{\mathbf{x}} + \sum_{i=1}^S [\mathbf{H}^{-1}]_{:, \mathcal{S}_i} \Theta_i [(\mathbf{H}^\top)^{-1}]_{:, \mathcal{S}_i}^\top, \end{aligned} \quad (3.43)$$

Algorithm 1 ASUP sanitization algorithm ($\mathbf{J}_0 = \mathbf{0}_L$).

Input: For each agent i , Ψ in (3.22), $p_i(\mathbf{I}_N, \Theta)$ in (3.11), and Ξ_i as defined in Theorem 2. Conditions of Theorem 2 are satisfied.

- 1: **for** each agent $i = 1, \dots, S$ **do**
 - 2: Choose an orthonormal basis $\mathbf{W}_i$ of $\text{Null}(\Xi_i)$.
 - 3: Choose a row vector $\mathbf{v}_i$ from $\mathbf{W}_i$ such that $\mathbf{v}_i \notin \text{Null}(\mathbf{G}_i[\Psi]_{:,S_i})$.
 - 4: Choose a matrix $\mathbf{X}_i$ such that $\mathbf{V}_i = [\mathbf{v}_i^\top, \mathbf{X}_i]$ is unitary. Set $\Theta_i = \mathbf{V}_i \text{diag}(\lambda_i, \mathbf{0}_{N_i-1}) \mathbf{V}_i^\top$ with $\lambda_i > 0$ and $\tilde{\Theta}_i = \text{diag}(\dots, \mathbf{0}_{N_i-1}, \Theta_i, \mathbf{0}_{N_{i+1}}, \dots)$. Choose λ_i to be large enough such that $p_i(\mathbf{I}_N, \tilde{\Theta}_i) \geq \epsilon_i$.
 - 5: Agent i generates a noise $\xi_i \sim \mathcal{N}(\mathbf{0}, \Theta_i)$ and adds this to its measurement $\mathbf{y}_i$.
 - 6: **end for**
-

which is linear in the perturbation noise covariance Θ_i of each agent i . Recall that $\mathbf{u} = \mathbf{U}\mathbf{x}$ and $\mathbf{g}_i = \mathbf{G}_i\mathbf{x}$. It can be seen that if there is one row in $\mathbf{U}[\mathbf{H}^{-1}]_{:,S_i}$ not in the row space of $\mathbf{G}_i[\mathbf{H}^{-1}]_{:,S_i}$, agent i can use arbitrarily large noise power in the corresponding row of Θ_i so that the estimation error of $\mathbf{g}_i$ becomes arbitrarily large while the estimation error of $\mathbf{u}$ remains unchanged. Theorem 2 generalizes this result to any $\mathbf{H}$. Based on Theorem 2, we can implement a decentralized sanitization scheme as illustrated in Algorithm 1 if we have $N_i > \text{rank}(\Xi_i)$ and $\text{Null}(\Xi_i) \not\subseteq \text{Null}(\mathbf{G}_i[\Psi]_{:,S_i})$ for each agent $i = 1, \dots, S$.

3.3.2 ASUP with prior information

In this subsection, we consider the case $\mathbf{J}_0 \succ \mathbf{0}_L$.

Theorem 3. Suppose $\mathbf{J}_0 \succ \mathbf{0}_L$. ASUP is achievable if and only if

$$\mathbf{U}[\Psi]_{:,S_i} \mathbf{H}_i \mathbf{P}_0 \mathbf{G}^\top = \mathbf{0}, \quad (3.44)$$

for every agent $i = 1, \dots, S$, where $\mathbf{G} = [\mathbf{G}_1^\top, \dots, \mathbf{G}_S^\top]^\top \in \mathbb{R}^{G \times L}$.

Proof. From (3.24), perfect utility $u(\mathbf{I}_N, \Theta) = 0$ requires

$$\text{Tr}(\mathbf{U}\Psi(\mathbf{I}_N + \Theta\Phi)^{-1}\Theta\Psi^\top\mathbf{U}^\top) = 0. \quad (3.45)$$

Applying Lemma 2 and the same argument in the proof of Theorem 2, the above statement holds if and only if

$$\text{Tr}(\mathbf{U}\Psi\Theta\Psi^\top\mathbf{U}^\top) = 0. \quad (3.46)$$

Arbitrarily strong privacy requires the existence of Θ such that

$$\lim_{n \rightarrow \infty} p_i(\mathbf{I}_N, n\Theta) = \epsilon_i^{\max}, \quad \forall i = 1, \dots, S. \quad (3.47)$$

From (3.23), the above equation can be written as

$$\lim_{n \rightarrow \infty} \text{Tr} \left(\mathbf{G}\mathbf{P}_0\mathbf{H}^\top (\Phi^{-1} + n\Theta)^{-1} \mathbf{H}\mathbf{P}_0\mathbf{G}^\top \right) = 0. \quad (3.48)$$

Since the L.H.S. of the above statement is a continuous function w.r.t. Θ , it is equivalent to

$$\lim_{n \rightarrow \infty} \text{Tr} \left(\mathbf{G}\mathbf{P}_0\mathbf{H}^\top \left(\Phi^{-1} + n\Theta + \frac{1}{n}\mathbf{I}_N \right)^{-1} \mathbf{H}\mathbf{P}_0\mathbf{G}^\top \right) = 0. \quad (3.49)$$

From Lemma 2, we have equivalently,

$$\lim_{n \rightarrow \infty} \text{Tr} \left(\mathbf{G}\mathbf{P}_0\mathbf{H}^\top \left(n\Theta + \frac{1}{n}\mathbf{I}_N \right)^{-1} \mathbf{H}\mathbf{P}_0\mathbf{G}^\top \right) = 0. \quad (3.50)$$

Rewriting the perfect utility condition (3.46) and the arbitrarily strong privacy condition (3.50) in terms of Θ_i , $i = 1, \dots, S$, yields

$$\begin{aligned} \sum_{i=1}^S \text{Tr} \left(\mathbf{U}[\Psi]_{:, \mathcal{S}_i} \Theta_i [\Psi]_{:, \mathcal{S}_i}^\top \mathbf{U}^\top \right) &= 0, \\ \lim_{n \rightarrow \infty} \sum_{i=1}^S \text{Tr} \left(\mathbf{G}\mathbf{P}_0\mathbf{H}_i^\top \left(n\Theta_i + \frac{1}{n}\mathbf{I}_{N_i} \right)^{-1} \mathbf{H}_i\mathbf{P}_0\mathbf{G}^\top \right) &= 0. \end{aligned} \quad (3.51)$$

The proof of necessity now follows by applying Lemma 3 to the above statements. To show sufficiency, we can follow the method in Appendix 3.C to construct Θ as illustrated in Algorithm 2, which guarantees $p_i(\mathbf{I}_N, \Theta) \geq \epsilon_i$, for any given $\epsilon_i < \epsilon_i^{\max}$,

Algorithm 2 ASUP sanitization algorithm ($\mathbf{J}_0 \succ \mathbf{0}_L$).

Input: For each agent i , Ψ in (3.26) and $\Xi_i = \mathbf{G}\mathbf{J}_0^{-1}\mathbf{H}_i^\top$ with $\mathbf{G}$ defined in Theorem 3. Condition in Theorem 3 is satisfied.

- 1: **for** each agent $i = 1, \dots, S$ **do**
 - 2: Choose an orthonormal basis $\mathbf{W}_i \in \mathbb{R}^{W_i \times N_i}$ of the row space of Ξ_i .
 - 3: Choose an orthonormal basis $\mathbf{Y}_i$ of the row space of $\mathbf{U}[\Psi]_{:,S_i}$.
 - 4: Choose a matrix $\mathbf{X}_i$ such that $\mathbf{V}_i^\top = [\mathbf{W}_i^\top, \mathbf{Y}_i^\top, \mathbf{X}_i^\top]$ is unitary. Choose a diagonal matrix $\Lambda_i \succeq \mathbf{0}_{W_i}$ such that $\text{Tr}(\Xi_i \mathbf{W}_i^\top \Lambda_i^{-1} \mathbf{W}_i \Xi_i^\top) \leq \epsilon_i^{\max} - \epsilon_i$. Let $\Theta_i = \mathbf{V}_i^\top \text{diag}(\Lambda_i, \mathbf{0}_{N_i - W_i}) \mathbf{V}_i$.
 - 5: Agent i generates a noise $\xi_i \sim \mathcal{N}(\mathbf{0}, \Theta_i)$ and adds this to its measurement $\mathbf{y}_i$.
 - 6: **end for**
-

$i = 1, \dots, S$. The theorem is now proved. $\square$

Note that $\Psi \mathbf{H} \mathbf{P}_0$ is the reduced estimate covariance for $\mathbf{x}$ w.r.t. the prior estimate covariance $\mathbf{P}_0$ after agents observe $\mathbf{y}$. We may consider $[\Psi]_{:,S_i} \mathbf{H}_i \mathbf{P}_0$ in condition (3.44) as the reduced estimate covariance for agent i to observe $\mathbf{y}_i$. Theorem 3 gives the condition for each agent $i = 1, \dots, S$, to be able to decrease this reduced estimate covariance to its minimum for the private parameter $\mathbf{g}_i = \mathbf{G}_i \mathbf{x}$ without affecting the reduced estimate covariance for the public parameter $\mathbf{u} = \mathbf{U} \mathbf{x}$.

We note that the privacy function is bounded when $\mathbf{J}_0 \succ \mathbf{0}_L$ because $\mathbf{J}_0 \succ \mathbf{0}_L$ provides prior information about every agent's private parameter. The achievable privacy is the cumulative result of the noise perturbation applied by each agent. Since the privacy contributed by each agent is bounded when $\mathbf{J}_0 \succ \mathbf{0}_L$, the maximum privacy is only obtained when all the agents offer their maximum privacy. This is different from the case $\mathbf{J}_0 = \mathbf{0}_L$, where the privacy each agent can contribute is unbounded. Therefore, we can rely on one agent to provide arbitrarily strong privacy in Theorem 2. This explains why ASUP for $\mathbf{J}_0 = \mathbf{0}_L$ requires only one agent while ASUP for $\mathbf{J}_0 \succ \mathbf{0}_L$ requires every agent. In the case where only a set of agents satisfy the condition given in Theorem 3 when $\mathbf{J}_0 \succ \mathbf{0}_L$, we still can apply the sanitization method in Algorithm 2 to these nodes. However, as a consequence, we will obtain perfect utility but not arbitrarily strong privacy. In what follows we give an example in which condition (3.44) in Theorem 3 holds.

Example 3. Suppose $\mathbf{H}_i = \mathbf{H}_0 \in \mathbb{R}^{(N/S) \times L}$, for all $i = 1, \dots, S$, and the measurement noise is white, i.e., $\mathbf{R} = \sigma^2 \mathbf{I}_N$. In this case, we can verify that for some $\Psi_0 \in \mathbb{R}^{N \times (N/S)}$, $[\Psi]_{:,S_i} = \Psi_0$, for all $i = 1, \dots, S$. When $\mathbf{U}\Psi_0\mathbf{H}_0\mathbf{P}_0\mathbf{G}^\top = \mathbf{0}$, condition (3.44) is satisfied. Note that $\Psi_0\mathbf{H}_0\mathbf{P}_0$ is the reduced estimate covariance of $\mathbf{x}$ after agent i observes $\mathbf{y}_i$. Following that, $\mathbf{U}\Psi_0\mathbf{H}_0\mathbf{P}_0\mathbf{G}^\top$ can be interpreted as the reduced estimate cross-covariance between the public parameter $\mathbf{u}$ and the private parameters $\mathbf{g}_1, \dots, \mathbf{g}_S$, which need to be uncorrelated to make ASUP feasible.

3.4 Maximum Privacy Under Perfect Utility

In this section, we formulate the problem of achieving maximal privacy under perfect utility (i.e., $u(\mathbf{C}, \Theta) = 0$) as a linear matrix inequality optimization. When discussing ASUP, we allow the power of the perturbation noise to be arbitrarily large so that the privacy function can exceed arbitrarily large privacy thresholds. In most practical applications, there is a power constraint on the perturbation noise that can be added (if we normalize the noise as discussed immediately before (P3.1), this translates into a constraint on the dynamic range of the compression mechanism.).

We consider the following optimization:

$$\begin{aligned} \max_{(\mathbf{I}_N, \Theta) \in \mathcal{C}} \quad & \sum_{i=1}^S p_i(\mathbf{I}_N, \Theta), \\ \text{s. t.} \quad & u(\mathbf{I}_N, \Theta) = 0, \\ & \text{Tr}(\Theta_i) \leq \delta_i, \text{ for } i = 1, \dots, S, \end{aligned} \tag{3.52}$$

where $\delta_i > 0$ is the maximum total power available to agent i . In the following, we show how to reformulate (3.52) as a convex optimization problem with linear matrix inequality (LMI) constraints.

Recall from Theorems 2 and 3 that $u(\mathbf{I}_N, \boldsymbol{\Theta}) = 0$ under both the cases $\mathbf{J}_0 = \mathbf{0}_L$ and $\mathbf{J}_0 \succ \mathbf{0}_L$ is achievable if and only if

$$\text{Tr}(\mathbf{U}\Psi\boldsymbol{\Theta}\Psi^\top\mathbf{U}^\top) = 0. \quad (3.53)$$

By introducing an auxiliary variable $\mathbf{Z} \in \mathbb{S}^L$, (3.52) can be cast as

$$\begin{aligned} & \max_{(\mathbf{I}_N, \boldsymbol{\Theta}) \in \mathcal{C}, \mathbf{Z} \in \mathbb{S}^L} \sum_{i=1}^S \mathbf{G}_i \mathbf{Z} \mathbf{G}_i^\top, \\ & \text{s.t. } \mathbf{Z} \preceq \tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{I}_N, \boldsymbol{\Theta}), \\ & \text{Tr}(\mathbf{U}\Psi\boldsymbol{\Theta}\Psi^\top\mathbf{U}^\top) = 0, \\ & \text{diag}((\text{Tr}(\boldsymbol{\Theta}_i))_{i=1}^S) \leq \text{diag}((\delta_i)_{i=1}^S). \end{aligned} \quad (3.54)$$

Note that the maximization will force $\mathbf{Z}$ to achieve its upper bound.

In the case $\mathbf{J}_0 = \mathbf{0}_L$, we need the following form of the perturbed CRLB before transforming the first constraint of (3.54) into a LMI. Let the column vectors of $\bar{\mathbf{H}}$ be a basis of $\text{Null}(\mathbf{H}^\top)$ and let $\mathbf{K} = [\mathbf{K}_1^\top, \mathbf{K}_2^\top]^\top = [\mathbf{H}, \bar{\mathbf{H}}]^{-1}$ with $\mathbf{K}_1 \in \mathbb{R}^{L \times N}$. We obtain from (3.12)

$$\begin{aligned} \tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{I}_N, \boldsymbol{\Theta}) &= (\mathbf{H}^\top (\mathbf{R} + \boldsymbol{\Theta})^{-1} \mathbf{H})^{-1} \\ &= \left([(\mathbf{K}^\top)^{-1} (\mathbf{R} + \boldsymbol{\Theta})^{-1} \mathbf{K}^{-1}]_{\mathcal{L}, \mathcal{L}} \right)^{-1} \\ &= \left([(\mathbf{K}(\mathbf{R} + \boldsymbol{\Theta})\mathbf{K}^\top)^{-1}]_{\mathcal{L}, \mathcal{L}} \right)^{-1} \\ &= \mathbf{K}_1 \tilde{\boldsymbol{\Theta}} \mathbf{K}_1^\top - \mathbf{K}_1 \tilde{\boldsymbol{\Theta}} \mathbf{K}_2^\top \left(\mathbf{K}_2 \tilde{\boldsymbol{\Theta}} \mathbf{K}_2^\top \right)^{-1} \mathbf{K}_2 \tilde{\boldsymbol{\Theta}} \mathbf{K}_1^\top, \end{aligned} \quad (3.55)$$

where $\tilde{\boldsymbol{\Theta}} = \mathbf{R} + \boldsymbol{\Theta}$, $\mathcal{L} = \{1, \dots, L\}$ and the last equation is a consequence of the block matrix inversion. Substituting the above expression into the first constraint of (3.54) and using the Schur complement, the constraint can be cast as the following LMI:

$$\begin{bmatrix} \mathbf{K}_1(\boldsymbol{\Theta} + \mathbf{R})\mathbf{K}_1^\top & \mathbf{K}_1(\boldsymbol{\Theta} + \mathbf{R})\mathbf{K}_2^\top \\ \mathbf{K}_2(\boldsymbol{\Theta} + \mathbf{R})\mathbf{K}_1^\top & \mathbf{K}_2(\boldsymbol{\Theta} + \mathbf{R})\mathbf{K}_2^\top \end{bmatrix} - \begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{N-L} \end{bmatrix} \succeq \mathbf{0}_N. \quad (3.56)$$

Similarly, in the case $\mathbf{J}_0 \succ \mathbf{0}_L$, by substituting (3.23) into the first constraint of (3.54), it can be linearized as

$$\begin{bmatrix} \mathbf{P}_0 - \mathbf{Z} & \mathbf{P}_0 \mathbf{H}^\top \\ \mathbf{H} \mathbf{P}_0 & \Phi^{-1} + \Theta \end{bmatrix} \succeq \mathbf{0}_{N+L}. \quad (3.57)$$

The above formulations can then be solved by using standard semi-definite programming techniques [116].

Together with the results in Theorems 2 and 3, one may proceed to design the sanitization mechanism at each agent as follows. One first checks if the necessary and sufficient conditions for achieving ASUP are satisfied. If so and if there is no power constraint, we can use the steps outlined in Algorithms 1 and 2 to choose the sanitization mappings. If there is a power constraint, we solve (3.54). On the other hand, if ASUP is not achievable, we propose an alternating optimization procedure to solve (P3.1) in the following section.

3.5 Alternating Optimization

In this section, we propose an algorithm to solve (P3.1) for both the cases where information is available and unavailable. Because Θ is restricted to be a block diagonal matrix and the utility and privacy functions are non-linear w.r.t. Θ , it is difficult to directly solve the optimization problem (P3.1) using standard optimization tools. We propose to optimize (P3.1) w.r.t. $\Theta_1, \dots, \Theta_S$, sequentially and in an alternating fashion, instead of Θ . We show that optimizing (P3.1) over Θ_i is a convex problem, which can be solved by semi-definite programming [116].

To proceed, we need to do a trivial relaxation by assuming Θ is invertible. Then both the perturbed CRLBs when $\mathbf{J}_0 = \mathbf{0}_L$ and $\mathbf{J}_0 \succ \mathbf{0}_L$ in (3.20) and (3.24) can be written as

$$\tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{I}_N, \Theta) = \mathbf{P}_{\mathbf{x}} + \Psi (\Theta^{-1} + \Phi)^{-1} \Psi^\top, \quad (3.58)$$

where Ψ and Φ refer to their respective definitions when prior information is available or unavailable. Therefore, the algorithm to be proposed when $\mathbf{J}_0 = \mathbf{0}_L$ and $\mathbf{J}_0 \succ \mathbf{0}_L$ can be described as one. For sequential optimization, only Θ_i is updated at the i -th sequence while the rest of the parameters are fixed. By doing this, the utility and privacy tradeoff can be cast as a standard semi-definite programming problem w.r.t. Θ_i . Denote $\bar{\Theta} = \Theta^{-1}$ and for $i = 1, \dots, S$, let $\bar{\Theta}_i = \Theta_i^{-1}$ and

$$\bar{\Theta}_{\setminus i} = \text{diag}(\bar{\Theta}_1, \dots, \bar{\Theta}_{i-1}, \bar{\Theta}_{i+1}, \dots, \bar{\Theta}_S). \quad (3.59)$$

We can permute the rows of $\bar{\Theta}$ to obtain $\text{diag}(\bar{\Theta}_i, \bar{\Theta}_{\setminus i})$. In the same way, we permute the rows of $\mathbf{H}$ and $\mathbf{R}$ to obtain

$$\begin{aligned} & [\mathbf{H}_i^\top, \mathbf{H}_1^\top, \dots, \mathbf{H}_{i-1}^\top, \mathbf{H}_{i+1}^\top, \dots, \mathbf{H}_S^\top]^\top, \\ & \text{diag}(\mathbf{R}_i, \mathbf{R}_1, \dots, \mathbf{R}_{i-1}, \mathbf{R}_{i+1}, \dots, \mathbf{R}_S), \end{aligned} \quad (3.60)$$

respectively. Substitute the above expressions for $\mathbf{H}$ and $\mathbf{R}$, respectively, in Φ and Ψ to obtain Φ_i and Ψ_i . Treating $\bar{\Theta}_{\setminus i}$ as a constant, the perturbed CRLB (3.58) can be written as a function of $\bar{\Theta}_i$:

$$\bar{\mathbf{P}}_{\mathbf{x}}(\bar{\Theta}_i) = \mathbf{P}_{\mathbf{x}} + \Psi_i (\text{diag}(\bar{\Theta}_i, \bar{\Theta}_{\setminus i}) + \Phi_i)^{-1} \Psi_i^\top. \quad (3.61)$$

Note the above expression of the CRLB is equal to the CRLB in (3.58) since the permutation process keeps the CRLB unchanged.

Consequently, the utility and j -th privacy function for $j = 1, \dots, S$, w.r.t. $\bar{\Theta}_i$ can be expressed as (see Appendix 3.D for details)

$$u(\bar{\Theta}_i) = -\text{Tr} \left(\Gamma_u (\bar{\Theta}_i + \Omega)^{-1} \right) / \text{Tr} (\mathbf{U} \mathbf{P}_{\mathbf{x}} \mathbf{U}^\top) - \Delta_u, \quad (3.62)$$

$$p_j(\bar{\Theta}_i) = \text{Tr} \left(\Gamma_{p_j} (\bar{\Theta}_i + \Omega)^{-1} \right) / \text{Tr} (\mathbf{G}_j \mathbf{P}_{\mathbf{x}} \mathbf{G}_j^\top) + \Delta_{p_j}, \quad (3.63)$$

where $\mathbf{\Omega}, \mathbf{\Delta}_{p_j}, \mathbf{\Delta}_u, \mathbf{\Gamma}_u, \mathbf{\Gamma}_p$ are constants independent of $\overline{\mathbf{\Theta}}_i$ given by

$$\mathcal{N}_i = \{1, \dots, N_i\}, \mathcal{N}_{\setminus i} = \{N_i + 1, \dots, N\}, \quad (3.64)$$

$$\mathbf{\Delta}_{p_j} = \text{Tr} \left(\mathbf{G}_j [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}}^\top \mathbf{G}_j^\top \right) / \text{Tr} (\mathbf{G}_j \mathbf{P}_x \mathbf{G}_j^\top), \quad (3.65)$$

$$\mathbf{\Delta}_u = \text{Tr} \left(\mathbf{U} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}}^\top \mathbf{U}^\top \right) / \text{Tr} (\mathbf{U} \mathbf{P}_x \mathbf{U}^\top), \quad (3.66)$$

$$\begin{aligned} \mathbf{\Gamma}_u &= [\mathbf{\Psi}_i]_{:, \mathcal{N}_i}^\top \mathbf{U}^\top \mathbf{U} [\mathbf{\Psi}_i]_{:, \mathcal{N}_i} \\ &\quad - [\mathbf{\Psi}_i]_{:, \mathcal{N}_i}^\top \mathbf{U}^\top \mathbf{U} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Phi}_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i} \\ &\quad - [\mathbf{\Phi}_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}}^\top \mathbf{U}^\top \mathbf{U} [\mathbf{\Psi}_i]_{:, \mathcal{N}_i} \\ &\quad + [\mathbf{\Phi}_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}}^\top \mathbf{U}^\top \mathbf{U} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Phi}_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i}, \end{aligned} \quad (3.67)$$

$$\begin{aligned} \mathbf{\Gamma}_{p_j} &= [\mathbf{\Psi}_i]_{:, \mathcal{N}_i}^\top \mathbf{G}_j^\top \mathbf{G}_j [\mathbf{\Psi}_i]_{:, \mathcal{N}_i} \\ &\quad - [\mathbf{\Psi}_i]_{:, \mathcal{N}_i}^\top \mathbf{G}_j^\top \mathbf{G}_j [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Phi}_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i} \\ &\quad - [\mathbf{\Phi}_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}}^\top \mathbf{G}_j^\top \mathbf{G}_j [\mathbf{\Psi}_i]_{:, \mathcal{N}_i} \\ &\quad + [\mathbf{\Phi}_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}}^\top \mathbf{G}_j^\top \mathbf{G}_j [\mathbf{\Psi}_i]_{:, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Phi}_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i}, \end{aligned} \quad (3.68)$$

$$\mathbf{\Omega} = [\mathbf{\Phi}_i]_{\mathcal{N}_i, \mathcal{N}_i} - [\mathbf{\Phi}_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \mathbf{T} [\mathbf{\Phi}_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i}, \quad (3.69)$$

$$\mathbf{T} = \left(\overline{\mathbf{\Theta}}_{\setminus i} + [\mathbf{\Phi}_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_{\setminus i}} \right)^{-1}. \quad (3.70)$$

At the i -th sequence, we treat $\overline{\mathbf{\Theta}}_{\setminus i}$ as a constant and optimize (P3.1) over $\overline{\mathbf{\Theta}}_i$. This optimization problem can be formulated as

$$\begin{aligned} &\max_{\overline{\mathbf{\Theta}}_i \succeq \mathbf{0}_{N_i}} u(\overline{\mathbf{\Theta}}_i), \\ &\text{s.t. } p_j(\overline{\mathbf{\Theta}}_i) \geq \epsilon_j, \forall j = 1, \dots, S, \end{aligned} \quad (3.71)$$

which can be re-cast as a semidefinite programming problem by introducing $\mathbf{Z} = (\overline{\mathbf{\Theta}}_i + \mathbf{\Omega})^{-1}$ to obtain:

$$\begin{aligned} &\min_{\mathbf{Z}} \text{Tr} (\mathbf{\Gamma}_u \mathbf{Z}), \\ &\text{s.t. } \text{Tr} (\mathbf{\Gamma}_{p_j} \mathbf{Z}) \geq \epsilon'_j, \forall j = 1, \dots, S, \\ &\quad \mathbf{0}_{N_i} \preceq \mathbf{Z} \preceq \mathbf{\Omega}^{-1}, \end{aligned} \quad (\text{P3.2})$$

where $\epsilon'_j = (\epsilon_j - \Delta_{p_j}) \times \text{Tr}(\mathbf{G}_j \mathbf{P}_x \mathbf{G}_j^\top)$.

We propose to optimize (P3.1) over $\bar{\Theta}_1, \dots, \bar{\Theta}_S$ sequentially over multiple iterations. The algorithm is summarized in Algorithm 3.

Algorithm 3 Alternating optimization

Initialize: $\bar{\Theta}_i^0 = \mathbf{0}_{N_i}, i = 1, \dots, S$ at iteration 0.

Output: $\bar{\Theta}_i^k, i = 1, \dots, S$ at iteration k .

```

1: while  $k \leq \text{Max. number of iterations}$  do
2:   for  $i = 1$  to  $S$  do
3:     Compute  $\epsilon'_j, \Gamma_u$  and  $\Gamma_{p_j}$  for all  $j = 1, \dots, S$ .
4:     Solve (P3.2) to obtain  $\bar{\Theta}_i^k$ .
5:     Use  $\bar{\Theta}_i^k$  to update  $\bar{\Theta}_{\setminus i+1}^k, \dots, \bar{\Theta}_{\setminus S}^k$ .
6:   end for
7: end while

```

The interior point methods [116] for solving semidefinite programming problem perform very well in practice and have the worst case polynomial complexity. For example, the computational complexity of the faster interior point method given by [117] for solving (P3.2) is $\mathcal{O}(\sqrt{n}(mn^2) \log(1/\eta))$, where $n = N_i$ is the size of sensor measurements, $m = S + 2$ is the number of constraints and η is the relative accuracy.

3.6 Numerical results

In the section, we investigate the impact of different parameters on the maximum privacy attainable under perfect utility discussed in Section 3.3 and verify the performance of Algorithm 3 proposed in Section 3.5. The settings used for the simulations in this section are given as follows unless otherwise stated:

- The total number of agents' observations, $N = 72$.
- The number of agent i 's observations, $N_i = N/S$ for $i = 1, \dots, S$.
- The dimension of the hidden parameter $\mathbf{x}$, $L = 12$.

- The measurement noise covariance $\mathbf{R} = \mathbf{A}\mathbf{A}^\top$, where the entries of $\mathbf{A} \in \mathbb{R}^{72 \times 72}$ are independent samples drawn from $\text{Unif}(-0.5, 0.5)$.
- The entries of the observation model matrix $\mathbf{H} \in \mathbb{R}^{72 \times 12}$ are independent samples drawn from $\text{Unif}(-0.5, 0.5)$.
- The prior information $\mathbf{P}_0 = \mathbf{A}\mathbf{A}^\top$, where the entries of $\mathbf{A} \in \mathbb{R}^{12 \times 12}$ are independent samples drawn from $\text{Unif}(-10, 10)$.
- The entries of the projection matrix of the public variable $\mathbf{U} \in \mathbb{R}^{U \times 12}$ are independent samples drawn from $\text{Unif}(-0.5, 0.5)$.
- The projection matrix of the i -th private variable, $\mathbf{G}_i = \mathbf{G} \in \mathbb{R}^{3 \times 12}$, where the entries of $\mathbf{G}$ are independent samples drawn from $\text{Unif}(-0.5, 0.5)$.

We do not impose the noise power constraint when retrieving the maximum privacy under perfect utility. Each data point shown in the figures is averaged over 100 independent experiments.

3.6.1 Maximum privacy under perfect utility

We vary S (the number of the agents) and U (the dimension of the public parameter vector) to investigate the impact of these parameters on the maximum privacy attainable under perfect utility by solving (3.54). Recall that $\epsilon_1^{\max} = \infty$ when $\mathbf{J}_0 = \mathbf{0}_L$ while $\epsilon_1^{\max}$ is bounded when $\mathbf{J}_0 \succ \mathbf{0}_L$. From Fig. 3.2 with $\mathbf{J}_0 = \mathbf{0}_L$, it can be seen that the maximum privacy under perfect utility goes to infinity (ASUP is achieved) when $S = 1$. By examining the necessary and sufficient conditions for ASUP in Theorem 2, it is observed that a system with a smaller number of agents S is more likely to fulfill the ASUP conditions than a system with a larger S with all the other settings remaining the same. In Fig. 3.3 with $\mathbf{J}_0 \succ \mathbf{0}_L$, none of the maximum privacy obtained under perfect utility is close to $\epsilon_1^{\max}$ (ASUP is not achieved). This can be elucidated from Theorem 3, where every agent is required to satisfy certain orthogonality conditions for ASUP to be achievable. Thus it

FIG. 3.2: The maximum privacy under perfect utility when no prior information is available. Because the maximum privacy can be unbounded, it is truncated when exceeding 100.

FIG. 3.3: The maximum privacy under perfect utility when prior information is available. The maximum privacy threshold is 74.1.

is unlikely to generate a random system model to achieve ASUP. Furthermore, both Figs. 3.2 and 3.3 demonstrate a descending trend for the maximum privacy achievable as the number of agents S increases or the length of the public parameter vector $\mathbf{u}$ increases. To expound further on this, recall from Theorems 2 and 3 that

perfect utility is achieved if and only if

$$\text{Tr} \left(\sum_{i=1}^S \mathbf{U} [\Psi]_{:,S_i} \Theta_i [\Psi]_{:,S_i}^\top \mathbf{U}^\top \right) = 0. \quad (3.72)$$

Therefore, to obtain $u(\mathbf{I}_N, \Theta) = 0$, a larger S or U is likely to force more agents to have $\Theta_i = \mathbf{0}_{N_i}$. Thus a system with a larger S or U has less degrees-of-freedom for privacy perturbation when maintaining perfect utility compared to a system with a smaller S or U .

3.6.2 Alternating optimization

By varying S and the privacy threshold ϵ_1 , Fig. 3.4 demonstrates the utility and privacy tradeoff by using Algorithm 3 proposed in Section 3.5. It is observed that the utility decreases as the privacy increases. Under the same setting, a larger number of agents deteriorates the utility. This is because each agent contains less measurements, thus giving less degrees-of-freedom for the optimization and leading to smaller utility. Furthermore, it can be seen that the optimal utility obtained by the algorithm is close to perfect utility ($u = 0$) when the privacy threshold ϵ_1 is set to be equal to the maximum privacy obtained under perfect utility. This implies that the algorithm proposed in Section 3.5 approximates the optimal solution well. Fig. 3.5 shows the convergence of the algorithm.

3.7 Conclusion

We defined utility and privacy using the CRLB and investigated the utility-privacy tradeoff in a decentralized setting. We showed that the privacy mechanism of linear compression can be arbitrarily closely approximated by the privacy mechanism using noise perturbation, and we showed how to translate between these two mechanisms. Furthermore, we derived necessary and sufficient conditions for achieving

FIG. 3.4: The privacy and utility tradeoff when $S = 9, 12, 18$, respectively, obtained by using Algorithm 3. The corresponding maximum privacy under perfect utility obtained by solving (3.54) is indicated.

FIG. 3.5: The utility at each iteration by using Algorithm 3, where the privacy threshold is set to be 10.

arbitrarily strong utility-privacy tradeoff, and developed methods to minimize privacy leakage while maintaining perfect utility under these conditions. When arbitrarily strong utility-privacy tradeoff is not achievable, we proposed an alternating optimization approach to find the optimal privacy noise power to add. Simulation results demonstrated the efficacy of our approach.

3.A Derivation of (3.20), (3.23) and (3.24)

If $\mathbf{J}_0 = \mathbf{0}_L$, we have

$$\begin{aligned} \tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{I}_N, \boldsymbol{\Theta}) &= (\mathbf{H}^\top (\mathbf{R} + \boldsymbol{\Theta})^{-1} \mathbf{H})^{-1} \\ &= \left(\mathbf{H}^\top \left(\mathbf{R}^{-1} - \mathbf{R}^{-1} (\mathbf{I}_N + \boldsymbol{\Theta} \mathbf{R}^{-1})^{-1} \boldsymbol{\Theta} \mathbf{R}^{-1} \right) \mathbf{H} \right)^{-1} \end{aligned} \quad (3.73)$$

$$\begin{aligned} &= \left(\mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H} - \mathbf{H}^\top \mathbf{R}^{-1} (\mathbf{I}_N + \boldsymbol{\Theta} \mathbf{R}^{-1})^{-1} \boldsymbol{\Theta} \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \\ &= \mathbf{P}_{\mathbf{x}} - \mathbf{P}_{\mathbf{x}} \mathbf{H}^\top \mathbf{R}^{-1} \left(-\mathbf{I}_N - \boldsymbol{\Theta} \mathbf{R}^{-1} \right. \\ &\quad \left. + \boldsymbol{\Theta} \mathbf{R}^{-1} \mathbf{H} \mathbf{P}_{\mathbf{x}} \mathbf{H}^\top \mathbf{R}^{-1} \right)^{-1} \boldsymbol{\Theta} \mathbf{R}^{-1} \mathbf{H} \mathbf{P}_{\mathbf{x}} \\ &= \mathbf{P}_{\mathbf{x}} + \boldsymbol{\Psi} (\mathbf{I}_N + \boldsymbol{\Theta} \boldsymbol{\Phi})^{-1} \boldsymbol{\Theta} \boldsymbol{\Psi}^\top, \end{aligned} \quad (3.74)$$

where $\boldsymbol{\Psi} = \mathbf{P}_{\mathbf{x}} \mathbf{H}^\top \mathbf{R}^{-1}$ and $\boldsymbol{\Phi} = \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{H} \mathbf{P}_{\mathbf{x}} \mathbf{H}^\top \mathbf{R}^{-1} \in \mathbb{S}^N$, (3.73) follows from the binomial inverse theorem, and (3.74) follows from the Woodbury matrix identity. Note that $\boldsymbol{\Phi}$ is a degenerate matrix since by letting the singular value decomposition of $\mathbf{H}^\top \mathbf{R}^{-1/2} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\top$, we have

$$\begin{aligned} \boldsymbol{\Phi} &= \mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{H} \mathbf{P}_{\mathbf{x}} \mathbf{H}^\top \mathbf{R}^{-1} \\ &= \mathbf{R}^{-1} - \mathbf{R}^{-1/2} \mathbf{R}^{-1/2} \mathbf{H} \\ &\quad \left(\mathbf{H}^\top \mathbf{R}^{-1/2} \mathbf{R}^{-1/2} \mathbf{H} \right)^{-1} \mathbf{H}^\top \mathbf{R}^{-1/2} \mathbf{R}^{-1/2} \\ &= \mathbf{R}^{-1/2} \mathbf{V} \text{diag}(\mathbf{0}_L, \mathbf{I}_{N-L}) \mathbf{V}^\top \mathbf{R}^{-1/2}. \end{aligned} \quad (3.75)$$

On the other hand, if $\mathbf{J}_0 \succ \mathbf{0}_L$, we obtain

$$\begin{aligned} \tilde{\mathbf{P}}_{\mathbf{x}}(\mathbf{I}_N, \boldsymbol{\Theta}) &= (\mathbf{J}_0 + \mathbf{H}^\top (\mathbf{R} + \boldsymbol{\Theta})^{-1} \mathbf{H})^{-1} \\ &= \mathbf{P}_0 - \mathbf{P}_0 \mathbf{H}^\top (\boldsymbol{\Phi}^{-1} + \boldsymbol{\Theta})^{-1} \mathbf{H} \mathbf{P}_0 \end{aligned} \quad (3.76)$$

$$\begin{aligned} &= \mathbf{P}_0 - \mathbf{P}_0 \mathbf{H}^\top (\boldsymbol{\Phi} - \boldsymbol{\Phi} (\mathbf{I}_N + \boldsymbol{\Theta} \boldsymbol{\Phi})^{-1} \boldsymbol{\Theta} \boldsymbol{\Phi}) \mathbf{H} \mathbf{P}_0 \\ &= \mathbf{P}_0 - \mathbf{P}_0 \mathbf{H}^\top \boldsymbol{\Phi} \mathbf{H}^\top \mathbf{P}_0 + \mathbf{P}_0 \mathbf{H}^\top \boldsymbol{\Phi} (\mathbf{I}_N + \boldsymbol{\Theta} \boldsymbol{\Phi})^{-1} \boldsymbol{\Theta} \boldsymbol{\Phi} \mathbf{H} \mathbf{P}_0 \end{aligned} \quad (3.77)$$

$$\begin{aligned}
&= (\mathbf{J}_0 + \mathbf{H}^\top \mathbf{R}^{-1} \mathbf{H})^{-1} + \mathbf{P}_0 \mathbf{H}^\top \Phi (\mathbf{I}_N + \Theta \Phi)^{-1} \Theta \Phi \mathbf{H} \mathbf{P}_0 \\
&= \mathbf{P}_x + \Psi (\mathbf{I}_N + \Theta \Phi)^{-1} \Theta \Psi^\top,
\end{aligned}$$

where $\mathbf{P}_0 = \mathbf{J}_0^{-1}$, $\Phi = (\mathbf{H} \mathbf{P}_0 \mathbf{H}^\top + \mathbf{R})^{-1}$, $\Psi = \mathbf{P}_0 \mathbf{H}^\top \Phi$, (3.76) follows from the Woodbury matrix identity, and (3.77) is a consequence of the binomial inverse theorem.

3.B Proof of Lemma 2

Since $\text{Tr}(\mathbf{A} \mathbf{S}_n^{-1} \mathbf{A}^\top) \geq \text{Tr}(\mathbf{A} (\mathbf{S}_n + \mathbf{B})^{-1} \mathbf{A}^\top) \geq 0$, we have

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} \mathbf{S}_n^{-1} \mathbf{A}^\top) = 0 \\
&\implies \lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} (\mathbf{S}_n + \mathbf{B})^{-1} \mathbf{A}^\top) = 0.
\end{aligned} \tag{3.78}$$

On the other hand, for μ a positive scalar greater than the largest eigenvalue of $\mathbf{B}$, we have

$$\text{Tr}(\mathbf{A} (\mathbf{S}_n + \mathbf{B})^{-1} \mathbf{A}^\top) \geq \text{Tr}(\mathbf{A} (\mathbf{S}_n + \mu \mathbf{I}_N)^{-1} \mathbf{A}^\top) \geq 0. \tag{3.79}$$

Therefore,

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} (\mathbf{S}_n + \mathbf{B})^{-1} \mathbf{A}^\top) = 0 \\
&\implies \lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} (\mathbf{S}_n + \mu \mathbf{I})^{-1} \mathbf{A}^\top) = 0.
\end{aligned} \tag{3.80}$$

Let $\lambda_i(n) > 0$ be the i -th eigenvalue of $\mathbf{S}_n$, associated with a unit eigenvector $\mathbf{v}_i$ (commuting matrices share the same eigenspaces). Let $\alpha_i = \|\mathbf{A} \mathbf{v}_i\|_2^2$. We obtain

$$0 = \lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} (\mathbf{S}_n + \mu \mathbf{I}_N)^{-1} \mathbf{A}^\top) = \lim_{n \rightarrow \infty} \sum_{i=1}^N \frac{\alpha_i}{\lambda_i(n) + \mu}, \tag{3.81}$$

hence either $\alpha_i = 0$ or $\lambda_i(n) \rightarrow \infty$ as $n \rightarrow \infty$ for each $i = 1, \dots, N$, which implies

$$\lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} \mathbf{S}_n^{-1} \mathbf{A}^\top) = \lim_{n \rightarrow \infty} \sum_{i=1}^N \frac{\alpha_i}{\lambda_i(n)} = 0. \quad (3.82)$$

Combining the above result with (3.80), we obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} (\mathbf{S}_n + \mathbf{B})^{-1} \mathbf{A}^\top) &= 0 \\ \implies \lim_{n \rightarrow \infty} \text{Tr}(\mathbf{A} \mathbf{S}_n^{-1} \mathbf{A}^\top) &= 0. \end{aligned} \quad (3.83)$$

The proof is now complete.

3.C Proof of Lemma 3

Let $\lambda_i \geq 0$ be the i -th eigenvalue of $\mathbf{S}$, associated with a unit eigenvector $\mathbf{v}_i$. We first note that

$$\text{Tr}(\mathbf{A} \mathbf{S} \mathbf{A}^\top) = \sum_{i=1}^N \lambda_i \|\mathbf{A} \mathbf{v}_i\|_2^2, \quad (3.84)$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Tr} \left(\mathbf{B} \left(n \mathbf{S} + \frac{1}{n} \mathbf{I}_N \right)^{-1} \mathbf{B}^\top \right) \\ = \lim_{n \rightarrow \infty} \sum_{i=1}^N \frac{1}{n \lambda_i + 1/n} \|\mathbf{B} \mathbf{v}_i\|_2^2. \end{aligned} \quad (3.85)$$

We first prove necessity. If both (3.84) and (3.85) are 0, $\mathbf{A} \mathbf{v}_i = \mathbf{0}$ if $\lambda_i \neq 0$, and $\mathbf{B} \mathbf{v}_i = \mathbf{0}$ if $\lambda_i = 0$ for each $i = 1, \dots, N$. This implies that $\mathbf{v}_i$ is in the null space of either $\mathbf{A}$, $\mathbf{B}$ or both. Therefore, we have $\mathbf{A} \mathbf{Q} \mathbf{Q}^\top \mathbf{B}^\top = \mathbf{0}$, where $\mathbf{Q} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$ is a unitary matrix. Thus, $\mathbf{A} \mathbf{B}^\top = \mathbf{0}$.

We next assume that $\mathbf{A} \mathbf{B}^\top = \mathbf{0}$. Let $\{\mathbf{v}_i\}_{i=1, \dots, |\mathbf{B}|}$ be a basis of the row space of $\mathbf{B}$ and $\{\mathbf{v}_i\}_{i=|\mathbf{B}|+1, \dots, |\mathbf{B}|+|\mathbf{A}|}$ be a basis of the row space of $\mathbf{A}$, where $|\mathbf{A}| = \text{rank}(\mathbf{A})$ and $|\mathbf{B}| = \text{rank}(\mathbf{B})$. Let $\lambda_i > 0$ if $i = 1, \dots, |\mathbf{B}|$, else $\lambda_i = 0$.

Let $\mathbf{S} = \mathbf{Q} \text{diag}((\lambda_i)_{i=1}^N) \mathbf{Q}^\top$, where $\mathbf{Q} = [\mathbf{v}_1, \dots, \mathbf{v}_N]$ is a unitary matrix. It can be easily verified that such $\mathbf{S}$ satisfies the lemma conditions, and the proof is complete.

3.D Derivation of (3.62) and (3.63)

We show the steps to obtain the utility and privacy function **w.r.t.** $\bar{\Theta}_i$ described in Section 3.5. Partition Φ_i as

$$\Phi_i = \begin{bmatrix} [\Phi_i]_{\mathcal{N}_i, \mathcal{N}_i} & [\Phi_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \\ [\Phi_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i} & [\Phi_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_{\setminus i}} \end{bmatrix}, \quad (3.86)$$

where $[\Phi_i]_{\mathcal{N}_i, \mathcal{N}_i} \in \mathbb{R}^{N_i \times N_i}$, $\mathcal{N}_i = \{1, \dots, N_i\}$, $\mathcal{N}_{\setminus i} = \{N_i + 1, \dots, N\}$. We have

$$\begin{aligned} & (\text{diag}(\bar{\Theta}_i, \bar{\Theta}_{\setminus i}) + \Phi_i)^{-1} \\ &= \begin{bmatrix} \bar{\Theta}_i + [\Phi_i]_{\mathcal{N}_i, \mathcal{N}_i} & [\Phi_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \\ [\Phi_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i} & \bar{\Theta}_{\setminus i} + [\Phi_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_{\setminus i}} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \end{aligned} \quad (3.87)$$

where

$$\begin{aligned} \mathbf{A}_{11} &= (\bar{\Theta}_i + \Omega)^{-1}, \\ \mathbf{A}_{12} &= -\mathbf{A}_{11} [\Phi_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \mathbf{T}, \\ \mathbf{A}_{21} &= -\mathbf{T} [\Phi_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i} \mathbf{A}_{11}, \\ \mathbf{A}_{22} &= \mathbf{T} + [\Phi_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_{\setminus i}} \mathbf{A}_{11} [\Phi_i]_{\mathcal{N}_i, \mathcal{N}_i} \mathbf{T}, \end{aligned} \quad (3.88)$$

where $\Omega = [\Phi_i]_{\mathcal{N}_i, \mathcal{N}_i} - [\Phi_i]_{\mathcal{N}_i, \mathcal{N}_{\setminus i}} \mathbf{T} [\Phi_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_i}$ and $\mathbf{T} = (\bar{\Theta}_{\setminus i} + [\Phi_i]_{\mathcal{N}_{\setminus i}, \mathcal{N}_{\setminus i}})^{-1}$. Therefore, we have

$$\bar{\mathbf{P}}_{\mathbf{x}}(\bar{\Theta}_i) = \mathbf{P}_{\mathbf{x}} + \begin{bmatrix} [\Psi_i]_{:, \mathcal{N}_i} & [\Psi_i]_{:, \mathcal{N}_{\setminus i}} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} [\Psi_i]_{:, \mathcal{N}_i} & [\Psi_i]_{:, \mathcal{N}_{\setminus i}} \end{bmatrix}^\top$$

$$\begin{aligned}
&= \mathbf{P}_{\mathbf{x}} + [\Psi_i]_{:, \mathcal{N}_1} \mathbf{A}_{11} [\Psi_i]_{:, \mathcal{N}_i}^T + [\Psi_i]_{:, \mathcal{N}_{\setminus i}} \mathbf{A}_{21} [\Psi_i]_{:, \mathcal{N}_i}^T \\
&\quad + [\Psi_i]_{:, \mathcal{N}_1} \mathbf{A}_{12} [\Psi_i]_{:, \mathcal{N}_{\setminus i}}^T + [\Psi_i]_{:, \mathcal{N}_{\setminus i}} \mathbf{A}_{22} [\Psi_i]_{:, \mathcal{N}_{\setminus i}}^T.
\end{aligned} \tag{3.89}$$

Therefore, the utility and j -th privacy function can be written as

$$\begin{aligned}
u(\bar{\Theta}_i) &= -\text{Tr} \left(\Gamma_u (\bar{\Theta}_i + \Omega)^{-1} \right) / \text{Tr} (\mathbf{U} \mathbf{P}_{\mathbf{x}} \mathbf{U}^T) - \Delta_u, \\
p_j(\bar{\Theta}_i) &= \text{Tr} \left(\Gamma_{p_j} (\bar{\Theta}_i + \Omega)^{-1} \right) / \text{Tr} (\mathbf{G}_j \mathbf{P}_{\mathbf{x}} \mathbf{G}_j^T) + \Delta_{p_j},
\end{aligned} \tag{3.90}$$

where $\Delta_{p_j}, \Delta_u, \Gamma_u, \Gamma_p$ are defined in Section 3.5.

Chapter 4

Privacy Preservation in Linear Dynamical System

The privacy protection framework developed in Chapter 3 is premised on a static system where the system state at any instant of time is determined by factors at the same time only. Thus the system states at different time steps are temporally independent of each other. However, the system state in a dynamical system depends on present factors as well as past factors. The resulted temporal correlation prevents us from directly applying the techniques developed in Chapter 3 to the privacy problems in a dynamical system. For example, in traffic monitoring systems, individual users send anonymized personal location traces, which can be GPS coordinates measured by their smartphones, continuously to a data aggregator to aid in traffic state estimation. However, an adversary may link an anonymous GPS trace to a particular person provided the knowledge of the person's residence or working location. In this chapter, we consider a linear dynamical system in which the state vector consists of both public and private states. One or more sensors make measurements of the state vector and sends information to a fusion center, which performs the final state estimation. To achieve an optimal tradeoff between the utility of estimating the public states and the privacy of protecting the private states, the measurements at each time step are linearly compressed into a lower

dimensional space. Under the centralized setting where all measurements are collected by a single sensor, we propose an optimization problem and an algorithm to find the best compression matrix. Under the decentralized setting where measurements are made separately at multiple sensors, each sensor optimizes its own local compression matrix. We propose methods to separate the overall optimization problem into multiple sub-problems that can be solved locally at each sensor. We consider the case where each sensor takes turns to transmit messages to the other sensors. Simulations and empirical experiments demonstrate the efficiency of our proposed approach in allowing the fusion center to estimate the public states with good accuracy while preventing it from estimating the private states accurately.

4.1 Problem Formulation

We consider an LDS given by the following state and measurement equations at time step k :

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{v}_k, \quad (4.1a)$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k, \quad (4.1b)$$

where $\mathbf{x}_k \in \mathbb{R}^L$ and $\mathbf{z}_k \in \mathbb{R}^N$ are the system's state and measurement at time k , respectively. The state and process noise $\mathbf{v}_k$ and $\mathbf{n}_k$ are independent, and follow zero-mean Gaussian distributions with positive definite covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively. We assume that the state evolution matrices $\mathbf{F}_k \in \mathbb{R}^{L \times L}$ are known for all k , and the measurement matrices $\{\mathbf{H}_j \in \mathbb{R}^{N \times L} : j \leq k\}$, where $N \geq L$ are known only up to the current time step k . This assumption is made because in many applications like target tracking [118, 119], the measurement matrix $\mathbf{H}_k$ is chosen adaptively at each time step k .

Another example is estimation over lossy networks where the measurement

matrices are time-varying [120]. On their routes to the gateway, sensor packets, possibly aggregated with measurements from several nodes, may become intermittent because of time-varying transmission intervals or delays [121], packet dropouts [122], random message exchanges depending on the availability of appropriate network links [123], fading channels [124], and other communication constraints [125]. The measurement matrix $\mathbf{H}_k$ is thus unknown until the sensor measurements are received at time k .

To obtain the minimum mean square estimate of the system state in (4.1), it is well known [112] that the Kalman filter is optimal. The Kalman filter contains two distinct phases: “predict” and “update”. In the “predict” phase, the state estimate and error covariance are predicted, respectively, by

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}, \quad (4.2)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^\top + \mathbf{Q}_k. \quad (4.3)$$

In the “update” phase, the state estimate and error covariance are updated, respectively, through

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k-1|k-1}), \quad (4.4)$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}, \quad (4.5)$$

where $\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^\top (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)^{-1}$ denotes the Kalman gain.

In this chapter, we consider the case where the state $\mathbf{x}_k$ may be partitioned into two parts as

$$\mathbf{x}_k = \begin{bmatrix} [\mathbf{x}_k]_{\mathcal{P}} \\ [\mathbf{x}_k]_{\mathcal{Q}} \end{bmatrix}, \quad (4.6)$$

where $[\mathbf{x}_k]_{\mathcal{P}} \in \mathbb{R}^{|\mathcal{P}|}$, with $\mathcal{P} = \{1, \dots, |\mathcal{P}|\}$ and $|\mathcal{P}| < L$, contains the public states that are to be estimated, while $[\mathbf{x}_k]_{\mathcal{Q}} \in \mathbb{R}^{|\mathcal{Q}|}$, with $\mathcal{Q} = \{|\mathcal{P}| + 1, \dots, L\}$ and $|\mathcal{Q}| = L - |\mathcal{P}| > 0$, are the private states containing sensitive information that

we wish to protect. Our goal is to minimize the estimation errors of the public states $[\hat{\mathbf{x}}_{k|k}]_{\mathcal{P}}$, while ensuring that the estimation errors of private states $[\hat{\mathbf{x}}_{k|k}]_{\mathcal{Q}}$ are above a predefined threshold.

In [60], it is assumed that the measurement or feature vector $\mathbf{z}_k$ can be mapped as $\mathbf{\Gamma}^{(\mathcal{P})}\mathbf{z}_k$ to a utility subspace and as $\mathbf{\Gamma}^{(\mathcal{Q})}\mathbf{z}_k$ to a privacy subspace. Our formulation is equivalent to a variant of the formulation in [60] if there is no additive noise $\mathbf{n}_k$ in (4.1b), and $[\mathbf{H}_k]_{:, \mathcal{P}}$ is orthogonal to $[\mathbf{H}_k]_{:, \mathcal{Q}}$, but is in general different from [60]. The advantage of our formulation is that in applications with a known LDS model, it is easier to specify which system states are public and private directly instead of through the measurements or observed features. Furthermore, to apply the formulation from [60] to a LDS where we want to protect the privacy of some states over multiple time steps, will require prior knowledge of the measurement matrices $\mathbf{H}_j$ for $j \geq k$, where k is the current time step. In particular, such an approach is impractical if j is large.

To prevent the fusion center from inferring the private states $[\mathbf{x}_k]_{\mathcal{Q}}$, we assume that a linear mapping or compression matrix $\mathbf{C}_k \in \mathbb{R}^{M \times N}$, where $1 \leq M \leq N$, is applied to the measurement to obtain

$$\tilde{\mathbf{z}}_k = \mathbf{C}_k \mathbf{z}_k = \mathbf{C}_k \mathbf{H}_k \mathbf{x}_k + \mathbf{C}_k \mathbf{n}_k. \quad (4.7)$$

Let $\tilde{\mathbf{P}}_{k|k}$ be the state error covariance based on the measurements $\tilde{\mathbf{z}}_{k'}$, where $k' \leq k$.

A smaller $\tilde{\mathbf{P}}_{k|k}(i, i)$ implies that the i -th state can be estimated with lower error on average. Therefore, as the utility, we aim to minimize the sum of the expected estimation errors of the public states or the public error trace at time k defined as

$$\tau_k(\mathbf{C}_k) = \text{Tr} \left([\tilde{\mathbf{P}}_{k|k}]_{\mathcal{P}, \mathcal{P}} \right). \quad (4.8)$$

On the other hand, the private error function at time k is defined as

$$\eta_k(\mathbf{C}_k) = \mathcal{F} \left(\text{vecdiag} \left([\tilde{\mathbf{P}}_{k|k}]_{\mathcal{Q}, \mathcal{Q}} \right) \right), \quad (4.9)$$

where $\mathcal{F}(\mathbf{x}) = \mathbf{A}\mathbf{x} : \mathbb{R}^{|\mathcal{Q}|} \rightarrow \mathbb{R}^{|\mathcal{F}|}$ is a user-defined linear map such that $\mathcal{F}(\mathbf{a}) \leq \mathcal{F}(\mathbf{b})$ if $\mathbf{a} \leq \mathbf{b}$, with $\leq$ here denoting element-wise inequality. The non-decreasing property follows iff $\mathbf{A}$ has non-negative entries.

For example, if $\mathcal{F}(\mathbf{x}) = \mathbf{1}_{|\mathcal{Q}|}^T \mathbf{x}$ with $|\mathcal{F}| = 1$, the private error function is the sum of the private states' error variances and

$$\eta_k(\mathbf{C}_k) = \text{Tr} \left(\left[\tilde{\mathbf{P}}_{k|k} \right]_{\mathcal{Q}, \mathcal{Q}} \right). \quad (4.10)$$

If the privacy of *every* private state is important, we can choose $\mathcal{F}(\mathbf{x}) = \mathbf{I}_{\mathcal{Q}} \mathbf{x}$ and $|\mathcal{F}| = |\mathcal{Q}|$. Then, the private error function becomes

$$\boldsymbol{\eta}_k(\mathbf{C}_k) = \text{vecdiag} \left(\left[\tilde{\mathbf{P}}_{k|k} \right]_{\mathcal{Q}, \mathcal{Q}} \right). \quad (4.11)$$

At each time k , we seek to optimize the utility-privacy tradeoff myopically as follows:

$$\begin{aligned} \min_{\mathbf{C}_k} \quad & \tau_k(\mathbf{C}_k), \\ \text{s.t.} \quad & \eta_k(\mathbf{C}_k) \geq \mathcal{F}(\delta \mathbf{1}_{|\mathcal{Q}|}), \end{aligned} \quad (\text{P4.0})$$

where δ is a predefined threshold, and the minimization is over the set of compression matrices $\{\mathbf{C}_k : \mathbf{C}_k \in \mathbb{R}^M, 1 \leq M \leq N\}$. Note that we are optimizing over $\mathbf{C}_k$ as well as its dimension, i.e., M , at every time step and (P4.0) is solved sequentially for each time step k .

The private error function given in (4.11) is more restrictive than that given in (4.10). In many practical problems, the private information depends on all the private states so that protecting some (but not necessarily all) of the private states may be sufficient to protect the overall private information. For example, to protect someone's geo-location information, which consists of x -, y - and z - coordinates, it may be sufficient in some applications to obfuscate just one or two coordinates instead of all three to achieve reasonable geo-location privacy. This motivates the use of the private error trace in (4.10) as one potential privacy measure.

Problem (P4.0) focuses on the utility-privacy tradeoff at the current time step without taking into account the future time steps. As the predicted error covariance $\tilde{\mathbf{P}}_{k+1|k}$ relates to the error covariance $\tilde{\mathbf{P}}_{k|k}$ through (4.3), i.e.,

$$\tilde{\mathbf{P}}_{k+1|k} = \mathbf{F}_{k+1} \tilde{\mathbf{P}}_{k|k} \mathbf{F}_{k+1}^\top + \mathbf{Q}_{k+1}, \quad (4.12)$$

$\mathbf{C}_k$ affects not just the utility-privacy tradeoff at the current time step k , but also the tradeoff at future time steps. This is illustrated in an example below.

Example 4. Consider the case where $[\mathbf{F}_k]_{\mathcal{P},\mathcal{P}} = 0$ and $[\mathbf{F}_k]_{\mathcal{Q},\mathcal{Q}} = 0$, i.e.,

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{0} & [\mathbf{F}_k]_{\mathcal{P},\mathcal{Q}} \\ [\mathbf{F}_k]_{\mathcal{Q},\mathcal{P}} & \mathbf{0} \end{bmatrix}. \quad (4.13)$$

Recall that the prediction error covariance $\mathbf{P}_{k|k-1}$ is related to $\mathbf{P}_{k-1|k-1}$ via (4.3). Substituting $\mathbf{F}_k$ into (4.2) and (4.3) yields

$$\begin{bmatrix} [\hat{\mathbf{x}}_{k|k-1}]_{\mathcal{P}} \\ [\hat{\mathbf{x}}_{k|k-1}]_{\mathcal{Q}} \end{bmatrix} = \begin{bmatrix} [\mathbf{F}_k]_{\mathcal{P},\mathcal{Q}} [\hat{\mathbf{x}}_{k-1|k-1}]_{\mathcal{Q}} \\ [\mathbf{F}_k]_{\mathcal{Q},\mathcal{P}} [\hat{\mathbf{x}}_{k-1|k-1}]_{\mathcal{P}} \end{bmatrix}, \quad (4.14)$$

$$\mathbf{P}_{k|k-1} = \begin{bmatrix} [\mathbf{F}_k]_{\mathcal{P},\mathcal{Q}} [\mathbf{P}_{k-1|k-1}]_{\mathcal{Q},\mathcal{Q}} [\mathbf{F}_k]_{\mathcal{Q},\mathcal{P}} & * \\ * & [\mathbf{F}_k]_{\mathcal{Q},\mathcal{P}} [\mathbf{P}_{k-1|k-1}]_{\mathcal{P},\mathcal{P}} [\mathbf{F}_k]_{\mathcal{P},\mathcal{Q}} \end{bmatrix} + \mathbf{Q}_k. \quad (4.15)$$

We see that the state evolution matrix $\mathbf{F}_k$ converts the public state at time $k-1$ into a linear function of the private state at time k , and vice versa. This means that if a low public error trace is achieved at time $k-1$, then a low private error function value at time k is inevitable if (P4.0) is used to design the utility-privacy tradeoff. This example shows that it is necessary to incorporate privacy constraints for future time steps into (P4.0).

Since the dynamical model (4.1a) is publicly known, we may predict the system's state n time steps in the future. The n -step prediction error covariance $\tilde{\mathbf{P}}_{k+n|k}$ at

time k is given by

$$\tilde{\mathbf{P}}_{k+n|k} = \mathbf{F}_{k+1:k+n} \tilde{\mathbf{P}}_{k|k} \mathbf{F}_{k+1:k+n}^\top + \sum_{i=1}^n \mathbf{F}_{k+i+1:k+n} \mathbf{Q}_{k+i} \mathbf{F}_{k+i+1:k+n}^\top. \quad (4.16)$$

The *look-ahead* private error function at time $k+n$ can be defined as

$$\eta_{k+n|k}(\mathbf{C}_k) = \mathcal{F} \left(\text{vecdiag} \left(\left[\tilde{\mathbf{P}}_{k+n|k} \right]_{\mathcal{Q}, \mathcal{Q}} \right) \right). \quad (4.17)$$

Incorporating privacy constraints on the prediction error covariance r_k steps ahead, we have the following optimization problem at each time step k :

$$\begin{aligned} \min_{\mathbf{C}_k} \quad & \tau_k(\mathbf{C}_k), \\ \text{s.t.} \quad & \eta_{k+n|k}(\mathbf{C}_k) \geq \mathcal{F}(\delta 1_{|\mathcal{Q}|}), \quad n = 0, \dots, r_k, \end{aligned} \quad (\text{P4.1})$$

where $\eta_{k|k} = \eta_k$ in (4.9). When $r_k = 0$, (P4.1) becomes (P4.0). Note again that (P4.1) is solved sequentially or myopically at each time step k , with the minimization over $\{\mathbf{C}_k : \mathbf{C}_k \in \mathbb{R}^M, 1 \leq M \leq N\}$.

Let

$$\mathbf{T}_k = \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k, \quad (4.18)$$

and for $n \geq 0$, let

$$\mathbf{G}_{k+n|k} = \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1} \mathbf{F}_{k+1:k+n}^\top. \quad (4.19)$$

We define

$$\mathbf{D}_{k+n|k} = \mathbf{F}_{k+1:k+n} \mathbf{D}_{k|k} \mathbf{F}_{k+1:k+n}^\top, \quad (4.20)$$

where

$$\mathbf{D}_{k|k} = \tilde{\mathbf{P}}_{k|k-1} - \tilde{\mathbf{P}}_{k|k}$$

$$= \mathbf{G}_{k|k}^\top \mathbf{C}_k^\top (\mathbf{C}_k \mathbf{T}_k \mathbf{C}_k^\top)^{-1} \mathbf{C}_k \mathbf{G}_{k|k} \quad (4.21)$$

is the reduction of the error covariance due to the measurement made at time k .

From (4.20) and (4.21), we obtain

$$\mathbf{D}_{k+n|k} = \mathbf{G}_{k+n|k}^\top \mathbf{C}_k^\top (\mathbf{C}_k \mathbf{T}_k \mathbf{C}_k^\top)^{-1} \mathbf{C}_k \mathbf{G}_{k+n|k}. \quad (4.22)$$

From (4.16) and (4.20), we have

$$\tilde{\mathbf{P}}_{k+n|k} = \tilde{\mathbf{P}}_{k+n|k-1} - \mathbf{D}_{k+n|k}. \quad (4.23)$$

Replacing n by $n+1$ and k by $k-1$ in (4.16), and applying (4.12), we obtain

$$\tilde{\mathbf{P}}_{k+n|k-1} = \mathbf{F}_{k+1:k+n} \tilde{\mathbf{P}}_{k|k-1} \mathbf{F}_{k+1:k+n}^\top + \sum_{i=1}^n \mathbf{F}_{k+i+1:k+n} \mathbf{Q}_{k+i} \mathbf{F}_{k+i+1:k+n}^\top. \quad (4.24)$$

Therefore, $\tilde{\mathbf{P}}_{k+n|k-1}$ does not depend on $\mathbf{C}_k$. We can now define the *utility* to be

$$u_k(\mathbf{C}_k) = \text{Tr} \left(\left[\tilde{\mathbf{P}}_{k|k-1} \right]_{\mathcal{P}, \mathcal{P}} \right) - \tau_k(\mathbf{C}_k) = \text{Tr} \left(\left[\mathbf{D}_{k|k} \right]_{\mathcal{P}, \mathcal{P}} \right), \quad (4.25)$$

and the n -step look-ahead *privacy loss* function at time k as

$$\begin{aligned} \ell_{k+n|k}(\mathbf{C}_k) &= \mathcal{F} \left(\text{vecdiag} \left(\left[\tilde{\mathbf{P}}_{k+n|k-1} \right]_{\mathcal{Q}, \mathcal{Q}} \right) \right) - \eta_{k+n|k}(\mathbf{C}_k) \\ &= \mathcal{F} \left(\text{vecdiag} \left(\left[\mathbf{D}_{k+n|k} \right]_{\mathcal{Q}, \mathcal{Q}} \right) \right). \end{aligned} \quad (4.26)$$

Problem (P4.1) can then be equivalently recast as

$$\begin{aligned} \max_{\mathbf{C}_k} u_k(\mathbf{C}_k) & \\ \text{s.t. } \ell_{k+n|k}(\mathbf{C}_k) &\leq \mathcal{F}(\bar{\boldsymbol{\delta}}_{k+n}), \quad n = 0, \dots, r_k, \end{aligned} \quad (\text{P4.2})$$

where

$$\bar{\delta}_{k+n} = \text{vecdiag} \left(\left[\tilde{\mathbf{P}}_{k+n|k-1} \right]_{\mathcal{Q}, \mathcal{Q}} \right) - \delta 1_{|\mathcal{Q}|} \quad (4.27)$$

is the n -step *privacy loss threshold*. From Example 4, we see that a sufficiently large r_k at the current time step k is required to ensure that there exists a $\mathbf{C}_t$ such that $\eta_{t|t}(\mathbf{C}_t) \geq \mathcal{F}(\delta 1_{|\mathcal{Q}|})$ for all future time steps $t > k$. In the following, we provide a lower bound for r_k under different assumptions. We start off with an elementary lemma.

Lemma 4. Consider a matrix $\mathbf{A}$ and square matrix $\mathbf{B}$. Suppose $\mathbf{A}\mathbf{A}^\top \succeq \alpha\mathbf{I}$ and $\mathbf{B} \succeq \beta\mathbf{I}$, then $\mathbf{A}\mathbf{B}\mathbf{A}^\top \succeq \alpha\beta\mathbf{I}$, where α and β are non-negative scalars.

Proof. Since $\mathbf{B} - \beta\mathbf{I} \succeq \mathbf{0}$, we have $\mathbf{A}(\mathbf{B} - \beta\mathbf{I})\mathbf{A}^\top \succeq \mathbf{0}$. From $\mathbf{A}\mathbf{A}^\top - \alpha\mathbf{I} \succeq \mathbf{0}$, we obtain $\mathbf{A}\mathbf{B}\mathbf{A}^\top - \alpha\beta\mathbf{I} \succeq \mathbf{0}$, which completes the proof. $\square$

Proposition 2. Suppose that at the current time step k , $\tilde{\mathbf{P}}_{k|k-1} \succeq \nu_k\mathbf{I}$ for some $\nu_k > 0$. Suppose also that for all $t \geq k$, $\mathbf{Q}_t \succeq \epsilon\mathbf{I}$ for some $\epsilon > 0$, and $\mathbf{F}_t\mathbf{F}_t^\top \succeq \xi\mathbf{I}$ for some $\xi > 0$. Then if either

(i) $\xi \geq 1$, or

(ii) $\xi < 1$, $\mathcal{F}(\delta 1_{|\mathcal{Q}|}) \leq \frac{\epsilon}{1-\xi} \mathcal{F}(1_{|\mathcal{Q}|})$, and $\nu_k < \frac{\epsilon}{1-\xi}$,

and ¹

$$\begin{aligned} r_k &\geq r_{k,\star} \\ &\triangleq \min_{r \geq 0} \left\{ r : \mathcal{F}(1_{|\mathcal{Q}|}) \left(\epsilon \frac{1-\xi^r}{1-\xi} + \xi^r \nu_k \right) \geq \mathcal{F}(\delta 1_{|\mathcal{Q}|}) \right\} - 1, \end{aligned} \quad (4.28)$$

there exists $\mathbf{C}_t$ such that $\eta_{t'|t}(\mathbf{C}_t) \geq \mathcal{F}(\delta 1_{|\mathcal{Q}|})$, for all $t' \geq t \geq k$.

¹We define $\frac{1-\xi^r}{1-\xi} \Big|_{\xi=1} = \lim_{\xi \rightarrow 1} \frac{1-\xi^r}{1-\xi} = r$.

Proof. From (4.23), we have for all $t' \geq t \geq k$, $\tilde{\mathbf{P}}_{t'|k-1} \succeq \tilde{\mathbf{P}}_{t'|t}$. If

$$\left[\tilde{\mathbf{P}}_{t'|k-1} \right]_{\mathcal{Q}, \mathcal{Q}} \succeq \delta \mathbf{I}_{|\mathcal{Q}|}, \quad (4.29)$$

we can always choose $\mathbf{C}_t = \mathbf{0}$ so that $\left[\tilde{\mathbf{P}}_{t'|t} \right]_{\mathcal{Q}, \mathcal{Q}} = \left[\tilde{\mathbf{P}}_{t'|k-1} \right]_{\mathcal{Q}, \mathcal{Q}} \succeq \delta \mathbf{I}_{|\mathcal{Q}|}$, which along with the non-decreasing property of $\mathcal{F}$ implies that $\eta_{t'|t}(\mathbf{C}_t) \geq \mathcal{F}(\delta \mathbf{1}_{|\mathcal{Q}|})$. We next show that the conditions given in the proposition statement lead to (4.29).

From (4.16), after some minor manipulations using (4.12), we can rewrite $\tilde{\mathbf{P}}_{t'|k-1}$ as

$$\tilde{\mathbf{P}}_{t'|k-1} = \mathbf{F}_{k+1:t'} \tilde{\mathbf{P}}_{k|k-1} \mathbf{F}_{k+1:t'}^\top + \sum_{i=1}^n \mathbf{F}_{k+i+1:t'} \mathbf{Q}_{k+i} \mathbf{F}_{k+i+1:t'}^\top. \quad (4.30)$$

Applying Lemma 4 to each term on the right hand side of (4.30), we obtain

$$\tilde{\mathbf{P}}_{t'|k-1} \succeq \left(\xi^r \nu_k + \sum_{i=1}^r \xi^{r-i} \epsilon \right) \mathbf{I}, \quad (4.31)$$

where $r = t' - k$. We therefore have

$$\begin{aligned} & \mathcal{F} \left(\text{vecdiag} \left(\left[\tilde{\mathbf{P}}_{t'|k-1} \right]_{\mathcal{Q}, \mathcal{Q}} \right) \right) \\ & \geq f(r) \triangleq \mathcal{F}(\mathbf{1}_{|\mathcal{Q}|}) \left(\epsilon \frac{1 - \xi^r}{1 - \xi} + \xi^r \nu_k \right). \end{aligned} \quad (4.32)$$

If $\xi \geq 1$, $f(r) \uparrow \infty$ as $r \rightarrow \infty$. On the other hand, if $\xi < 1$, we have

$$f(r+1) - f(r) = \mathcal{F}(\mathbf{1}_{|\mathcal{Q}|}) \left(\frac{\epsilon}{1 - \xi} - \nu_k \right) (\xi^r - \xi^{r+1}), \quad (4.33)$$

and $f(r) \uparrow \mathcal{F}(\mathbf{1}_{|\mathcal{Q}|}) \frac{\epsilon}{1 - \xi}$ as $r \rightarrow \infty$ if $\nu_k < \epsilon/(1 - \xi)$. Therefore, under the conditions in the proposition, by choosing the number of look-ahead steps r_k to satisfy $f(r_k + 1) \geq \mathcal{F}(\delta \mathbf{1}_{|\mathcal{Q}|})$, $\eta_{t'|t}(\mathbf{C}_t) \geq \mathcal{F}(\delta \mathbf{1}_{|\mathcal{Q}|})$ holds for all $t' \geq t \geq k$, and the proposition is proved. $\square$

4.2 A Single Sensor

In this section, we consider the case where there is only a single sensor so that (P4.2) can be optimized over all $\mathbf{C}_k \in \mathbb{R}^{M \times N}$ and $1 \leq M \leq N$ in a centralized setting. This gives the best utility-privacy tradeoff and serves as a benchmark for the decentralized case that we consider in Section 4.3. We first prove an elementary lemma.

Lemma 5. Suppose $\mathbf{A} \in \mathbb{R}^{M \times N}$ with $M \leq N$ has full row rank, and $\mathbf{B} \in \mathbb{R}^{N \times N}$ is a positive definite matrix. Then,

$$\mathbf{A}^\top (\mathbf{A}\mathbf{B}\mathbf{A}^\top)^{-1} \mathbf{A} = \mathbf{B}^{-1/2} \mathbf{U} \mathbf{U}^\top \mathbf{B}^{-1/2}, \quad (4.34)$$

where $\mathbf{U} \in \mathbb{R}^{N \times M}$ consists of the M right unit singular vectors associated with the M non-zero singular values of $\mathbf{A}\mathbf{B}^{1/2}$.

Proof. See Appendix 4.A. □

Suppose that $\mathbf{C}_k$ is full row rank (otherwise we can choose a smaller M). From $\mathbf{D}_{k+n|k}$ defined in (4.22) and Lemma 5, we obtain

$$\mathbf{D}_{k+n|k} = \mathbf{G}_{k+n|k}^\top \mathbf{T}_k^{-1/2} \mathbf{U}_k \mathbf{U}_k^\top \mathbf{T}_k^{-1/2} \mathbf{G}_{k+n|k}, \quad (4.35)$$

where $\mathbf{U}_k$ consists of the M right unit singular vectors of $\mathbf{C}_k \mathbf{T}_k^{1/2}$ associated with its non-zero singular values. For any $n \geq 0$ and index set $\mathcal{I}$, let

$$\boldsymbol{\Theta}_{k+n|k}^{(\mathcal{I})} = \left[\mathbf{T}_k^{-1/2} \mathbf{G}_{k+n|k} \right]_{:, \mathcal{I}} \left[\mathbf{T}_k^{-1/2} \mathbf{G}_{k+n|k} \right]_{:, \mathcal{I}}^\top. \quad (4.36)$$

Then, the utility function and privacy loss function in (4.25) and (4.26) can be expressed as functions of $\mathbf{U}_k$ as follows:

$$u_k(\mathbf{U}_k) = \text{Tr} \left(\mathbf{U}_k^\top \boldsymbol{\Theta}_{k|k}^{(\mathcal{P})} \mathbf{U}_k \right), \quad (4.37)$$

$$\ell_{k+n|k}(\mathbf{U}_k) = \mathcal{F} \left(\left[\text{Tr} \left(\mathbf{U}_k^\top \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_1)} \mathbf{U}_k \right), \dots, \text{Tr} \left(\mathbf{U}_k^\top \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_{|\mathcal{Q}|})} \mathbf{U}_k \right) \right]^\top \right). \quad (4.38)$$

where we let $\mathcal{Q}_i = \{|\mathcal{P}| + i\}$ be a set consisting of the index of the i -th private state such that $\mathcal{Q} = \bigcup_{i=1}^{|\mathcal{Q}|} \mathcal{Q}_i$.

Recall that $\mathcal{F}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. The Lagrangian $L(\mathbf{U}_k, \{\gamma_{n,f}\}, \{\lambda_m\})$ of problem (P4.2) is then given by

$$\begin{aligned} & L(\mathbf{U}_k, \{\gamma_{n,f}\}, \{\lambda_m\}) \\ &= \sum_{m=1}^M [\mathbf{U}_k]_{:,m}^\top \boldsymbol{\Theta}_{k|k}^{(\mathcal{P})} [\mathbf{U}_k]_{:,m} \\ &\quad - \sum_{n=0}^{r_k} \sum_{f=1}^{|\mathcal{F}|} \gamma_{n,f} \left(\sum_{m=1}^M [\mathbf{U}_k]_{:,m}^\top \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f,j) \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_j)} [\mathbf{U}_k]_{:,m} - [\mathbf{A}]_{\mathcal{I}_f} \bar{\boldsymbol{\delta}}_{k+n|k} \right) \\ &\quad - \sum_{m=1}^M \lambda_m \left([\mathbf{U}_k]_{:,m}^\top [\mathbf{U}_k]_{:,m} - 1 \right), \end{aligned} \quad (4.39)$$

where $\gamma_{n,f}$, $n = 0, \dots, r_k$, $f = 1, \dots, |\mathcal{F}|$, $\mathcal{I}_f = \{f\}$ and λ_m , $m = 1, \dots, M$, are the Lagrange multipliers. Differentiating $L(\mathbf{U}_k, \{\gamma_{n,f}\}, \{\lambda_m\})$ with respect to $[\mathbf{U}_k]_{:,m}$ and equating to zero leads to

$$\begin{aligned} & \left(\boldsymbol{\Theta}_{k|k}^{(\mathcal{P})} - \sum_{n=0}^{r_k} \sum_{f=1}^{|\mathcal{F}|} \gamma_{n,f} \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f,j) \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_j)} \right) [\mathbf{U}_k]_{:,m} \\ &= \lambda_m [\mathbf{U}_k]_{:,m}, \quad m = 1, \dots, M, \end{aligned} \quad (4.40)$$

which implies the objective function of (P4.2) is maximized when $\mathbf{U}_k$ consists of the M unit eigenvectors of $\boldsymbol{\Theta}_{k|k}^{(\mathcal{P})} - \sum_{n=0}^{r_k} \sum_{f=1}^{|\mathcal{F}|} \gamma_{n,f} \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f,j) \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_j)}$ associated with its M largest *non-zero* eigenvalues. Observe that there is no need to consider zero eigenvalues in our solution as these do not change the Lagrangian in (4.39). We see that both $\{\mathbf{U}_k\}$ and $\{\lambda_m\}$ depend on $\{\gamma_{n,f} : n = 0, \dots, r_k; f = 1, \dots, |\mathcal{F}|\}$ and M .

Problem (P4.2) can be equivalently recast as

$$\max_{\{\gamma_{n,f}\}} u_k(\mathbf{U}_k) \quad (4.41)$$

$$\begin{aligned}
& \text{s.t. } \text{Tr} \left(\mathbf{U}_k^T \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f, j) \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_j)} \mathbf{U}_k \right) \leq [\mathbf{A}]_{\mathcal{I}_f} \bar{\boldsymbol{\delta}}_{k+n|k}, \\
& n = 0, \dots, r_k, \quad f = 1, \dots, |\mathcal{F}|, \\
& \mathbf{U}_k = \text{eig}_{\neq 0} \left(\boldsymbol{\Theta}_{k|k}^{(\mathcal{P})} - \sum_{n=0}^{r_k} \sum_{f=1}^{|\mathcal{F}|} \gamma_{n,f} \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f, j) \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_j)}, M \right),
\end{aligned}$$

where $\text{eig}_{\neq 0}(\mathbf{A}, M)$ denotes the matrix of eigenvectors of $\mathbf{A}$ corresponding to its M largest non-zero eigenvalues. The optimization problem (4.41) can be solved using standard iterative methods such as the interior-point method and sequential quadratic programming. However, due to non-convexity, there is no guarantee of finding the global optimum using such iterative methods.

Comparing (4.41) (or equivalently (P4.2)) with original problem (P4.1), we can see that (P4.1) is optimizing over a matrix $\mathbf{C}_k$ while (4.41) is optimizing over scalars $\{\gamma_{n,f}\}$. The optimal $\mathbf{U}_k$ (or equivalently $\mathbf{C}_k$) is then given by a closed-form expression in terms of $\{\gamma_{n,f}\}$. Hence, solving (4.41) is computationally easier than solving (P4.1) directly.

Since $\sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f, j) \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_j)}$ are positive definite for every $f = 1, \dots, |\mathcal{F}|$, the privacy loss function $\ell_{k+n|k}(\mathbf{U}_k)$ is increasing in M . Hence, the privacy constraint in (P4.2) may not be feasible when M is large. To determine the optimal M , we have the following lemma.

Lemma 6. For two positive integers M^+ and M^- such that $M^+ > M^-$, suppose that $\{\gamma_{n,f}^+\}$ and $\{\gamma_{n,f}^-\}$ are the solutions of (4.41) when $M = M^+$ and $M = M^-$, respectively, and let $\mathbf{U}_k^+$ and $\mathbf{U}_k^-$ be the corresponding $\mathbf{U}_k$. Then, $u_k(\mathbf{U}_k^+) \geq u_k(\mathbf{U}_k^-)$.

Proof. Let

$$\Phi_k^- = \boldsymbol{\Theta}_{k|k}^{(\mathcal{P})} - \sum_{n=0}^{r_k} \sum_{f=1}^{|\mathcal{F}|} \gamma_{n,f}^- \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f, j) \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_j)},$$

$$\Phi_k^+ = \Theta_{k|k}^{(\mathcal{P})} - \sum_{n=0}^{r_k} \sum_{f=1}^{|\mathcal{F}|} \gamma_{n,f}^+ \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f, j) \Theta_{k+n|k}^{(\mathcal{Q}_j)}. \quad (4.42)$$

Then, we have

$$\begin{aligned} u_k(\mathbf{U}_k^-) &= \text{Tr} \left(\text{eig}_{\neq 0}(\Phi_k^-, M^-)^\top \Theta_{k|k}^{(\mathcal{P})} \text{eig}_{\neq 0}(\Phi_k^-, M^-) \right) \\ &\leq \text{Tr} \left(\text{eig}_{\neq 0}(\Phi_k^-, M^+)^\top \Theta_{k|k}^{(\mathcal{P})} \text{eig}_{\neq 0}(\Phi_k^-, M^+) \right), \\ &\leq \text{Tr} \left(\text{eig}_{\neq 0}(\Phi_k^+, M^+)^\top \Theta_{k|k}^{(\mathcal{P})} \text{eig}_{\neq 0}(\Phi_k^+, M^+) \right) \\ &= u_k(\mathbf{U}_k^+), \end{aligned} \quad (4.43)$$

where the first inequality follows because $\Theta_{k|k}^{(\mathcal{P})} \succeq \mathbf{0}, M^+ > M^-$ and the last inequality follows because the maximum of $u_k(\mathbf{U}_k^+)$ is achieved when $\gamma_{n,f} = \gamma_n^+$, for $n = 0, \dots, r_k, f = 1, \dots, |\mathcal{F}|$. The lemma is now proved. $\square$

Lemma 6 shows that as long as the privacy constraints in (P4.2) are satisfied, M should be chosen as large as possible to maximize u_k . Let $\Phi_k = \Theta_{k|k}^{(\mathcal{P})} - \sum_{n=0}^{r_k} \sum_{f=1}^{|\mathcal{F}|} \gamma_{n,f} \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f, j) \Theta_{k+n|k}^{(\mathcal{Q}_j)}$. Due to the fact that $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$, $\text{rank}(\Phi_k) \leq \min(N, |\mathcal{P}| + (r_k + 1)|\mathcal{Q}|)$, thus we have $M \leq \min(N, |\mathcal{P}| + (r_k + 1)|\mathcal{Q}|)$. The distribution of the eigenvalues of Φ_k is revealed by the following lemma.

Lemma 7. Let $\mathbf{A} \succeq \mathbf{0}$ and $\mathbf{B} \succeq \mathbf{0}$ be two $N \times N$ matrices whose ranks are $r_{\mathbf{A}}$ and $r_{\mathbf{B}}$, respectively. If $N > r_{\mathbf{A}} + r_{\mathbf{B}}$, $\mathbf{C} = \mathbf{A} - \mathbf{B}$ has at most $r_{\mathbf{A}}$ positive eigenvalues and at most $r_{\mathbf{B}}$ negative eigenvalues.

Proof. See Appendix 4.B $\square$

Suppose the eigenvalues of Φ_k are sorted in descending order as $\lambda_1, \dots, \lambda_N$. Lemma 7 shows that if $N > |\mathcal{P}| + (r_k + 1)|\mathcal{Q}|$, we should choose $\mathbf{U}_k$ to consist of the M unit eigenvectors of Φ_k associated with its M largest non-zero eigenvalues

in $\{\lambda_1, \dots, \lambda_{|\mathcal{P}|}, \lambda_{N-(r_k+1)|\mathcal{Q}|+1}, \dots, \lambda_N\}$. If $N \gg |\mathcal{P}| + (r_k+1)|\mathcal{Q}|$, the measurement can be significantly compressed without losing utility.

We summarize our solution approach to (P4.1) in Algorithm 4.

Algorithm 4 Centralized solver for (P4.1)

Input: $\mathcal{F}(\cdot), r_k, \delta, \tilde{\mathbf{P}}_{k-1|k-1}, \mathbf{H}_k, \mathbf{R}_k, \mathbf{F}_{k+n}, \mathbf{Q}_{k+n}, n = 0, \dots, r_k$

Output: $M, \mathbf{C}_k = \mathbf{U}_k^\top \mathbf{T}_k^{-1/2}$

- 1: Compute $\{\boldsymbol{\Theta}_{k+n|k}\}$ according to (4.36), and $\{\mathcal{F}(\bar{\boldsymbol{\delta}}_{k+n})\}$ defined in (P4.2)
- 2: Initialize $M = \min(\{N, |\mathcal{P}| + (r_k+1)|\mathcal{Q}|\})$
- 3: **while** $M \geq 1$ **do**
- 4: For each M , find $\{\gamma_{n,f}\}$ by solving (4.41).
- 5: Compute $\mathbf{U}_k$ by

$$\text{eig}_{\neq 0} \left(\boldsymbol{\Theta}_{k|k}^{(\mathcal{P})} - \sum_{n=0}^{r_k} \sum_{f=1}^{|\mathcal{F}|} \gamma_{n,f} \sum_{j=1}^{|\mathcal{Q}|} \mathbf{A}(f, j) \boldsymbol{\Theta}_{k+n|k}^{(\mathcal{Q}_j)} M \right).$$

- 6: **if** $\ell_{k+n|k}(\mathbf{U}_k) \leq \mathcal{F}(\bar{\boldsymbol{\delta}}_{k+n}), n = 0, \dots, r_k$ **then**
 - 7: **break**
 - 8: **end if**
 - 9: $M \leftarrow M - 1$
 - 10: **end while**
-

The computational complexity of both eigenvalue decomposition and quadratic programming for solving (4.41) in Algorithm 4 is around $\mathcal{O}(n^3)$ using the interior point methods [116], where $n = M$ is the size of sensor measurements after compression.

4.3 Decentralized Sensors

Suppose that there are $S > 1$ sensors and each sensor s makes a $N_s \times 1$ measurement. Dividing the measurement model in (4.1b) into S parts, the measurement model at each sensor s is given by

$$[\mathbf{z}_k]_{\mathcal{I}_s} = [\mathbf{H}_k]_{\mathcal{I}_s} \mathbf{x}_k + [\mathbf{n}_k]_{\mathcal{I}_s}, \quad s = 1, \dots, S, \quad (4.44)$$

where $\mathcal{I}_s = \{1 + \sum_{i=1}^{s-1} N_i, \dots, \sum_{i=1}^s N_i\}$ and $\sum_{s=1}^S N_s = N$. The quantities $[\mathbf{z}_k]_{\mathcal{I}_s}$, $[\mathbf{H}_k]_{\mathcal{I}_s} \in \mathbb{R}^{N_s \times L}$ and $[\mathbf{n}_k]_{\mathcal{I}_s} \in \mathbb{R}^{N_s}$ are, respectively, the measurement made by sensor s , the measurement matrix of sensor s , and the measurement noise.

Each sensor s applies its own compression $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} \in \mathbb{R}^{M_s \times N_s}$, which is a linear mapping from a vector space with dimension N_s to one with dimension $M_s < N_s$, and $\mathcal{J}_s = \{1 + \sum_{i=1}^{s-1} M_i, \dots, \sum_{i=1}^s M_i\}$, on its measurement before sending it to the fusion center. We have

$$[\tilde{\mathbf{z}}_k]_{\mathcal{I}_s} = [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} \cdot [\mathbf{z}_k]_{\mathcal{I}_s} = [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{H}_k]_{\mathcal{I}_s} \mathbf{x}_k + [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{n}_k]_{\mathcal{I}_s}. \quad (4.45)$$

Relating the compression map at each sensor in (4.45) to the overall compression map in (4.7), we see that the distributed implementation restricts the structure of the transformation matrix $\mathbf{C}_k$ to be a block diagonal matrix whose diagonal entries are the transformation matrices $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}$, $s = 1, \dots, S$ and the remaining entries are all zeros. The version of (P4.2) for decentralized sensors can then be formulated as follows:

$$\begin{aligned} \max_{\mathbf{C}_k} u_k(\mathbf{C}_k) & \quad (P4.3) \\ \text{s.t. } \ell_{k+n|k}(\mathbf{C}_k) & \leq \mathcal{F}(\bar{\boldsymbol{\delta}}_{k+n}), \quad n = 0, \dots, r_k, \\ \mathbf{C}_k & = \text{diag}([\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}, \dots, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}). \end{aligned}$$

We firstly rewrite (4.22) to express $\mathbf{D}_{k+n|k}$ in terms of $[\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}, \dots, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}$ as follows:

$$\begin{aligned} \mathbf{D}_{k+n|k} & = \mathbf{G}_{k+n|k}^\top \mathbf{C}_k^\top \left(\mathbf{C}_k \mathbf{T}_k \mathbf{C}_k^\top \right)^{-1} \mathbf{C}_k \mathbf{G}_{k+n|k}, \\ & = \left[[\mathbf{G}_{k+n|k}]_{\mathcal{I}_1}^\top [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^\top, \dots, [\mathbf{G}_{k+n|k}]_{\mathcal{I}_S}^\top [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^\top \right] \\ & \quad \Psi_k^{-1} \begin{bmatrix} [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_1} \\ \vdots \\ [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_S} \end{bmatrix}, \end{aligned} \quad (4.46)$$

where Ψ_k is defined in (4.47).

$$\Psi_k = \begin{bmatrix} [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1} [\mathbf{T}_k]_{\mathcal{I}_1, \mathcal{I}_1} [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^\top & \cdots & [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1} [\mathbf{T}_k]_{\mathcal{I}_1, \mathcal{I}_S} [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^\top \\ \vdots & \ddots & \vdots \\ [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S} [\mathbf{T}_k]_{\mathcal{I}_S, \mathcal{I}_1} [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^\top & \cdots & [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S} [\mathbf{T}_k]_{\mathcal{I}_S, \mathcal{I}_S} [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^\top \end{bmatrix}. \quad (4.47)$$

To solve (P4.3) in a decentralized fashion, we aim to separate the objective function into individual local objective functions at each sensor. However, without additional assumptions and reformulation of the objective function, this is not possible due to the inverse of the matrix Ψ_k . In the following, we consider the case where sensors are allowed to broadcast messages sequentially to all other sensors, and we propose new objective functions that are separable.

4.3.1 Sequential message broadcasts

In this subsection, we consider the case where information is exchanged between sensors to facilitate optimization of the compression map at each sensor.

We rewrite $\mathbf{D}_{k+n|k}$ in (4.46) to isolate sensor s 's compression map, $[\mathbf{C}_k]_{\mathcal{I}_s, \mathcal{J}_s}$, from the transformations of the other sensors $[\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}$ (see Appendix 4.C for the derivation):

$$\begin{aligned} \mathbf{D}_{k+n|k} &= [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}^\top [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\Phi_k]_{\mathcal{J}_{\setminus s}, \mathcal{J}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}} \\ &\quad + \left(\mathbf{G}_{k+n|k}^{(s)} \right)^\top \mathbf{U}_k^{(s)} \mathbf{U}_k^{(s)\top} \mathbf{G}_{k+n|k}^{(s)} - \left(\mathbf{G}_{k+n|k}^{(s)} \right)^\top \mathbf{U}_k^{(s)} \mathbf{U}_k^{(s)\top} \mathbf{G}_{k+n|k}^{(\setminus s)} \\ &\quad - \left(\mathbf{G}_{k+n|k}^{(\setminus s)} \right)^\top \mathbf{U}_k^{(s)} \mathbf{U}_k^{(s)\top} \mathbf{G}_{k+n|k}^{(s)} + \left(\mathbf{G}_{k+n|k}^{(\setminus s)} \right)^\top \mathbf{U}_k^{(s)} \mathbf{U}_k^{(s)\top} \mathbf{G}_{k+n|k}^{(\setminus s)}, \end{aligned} \quad (4.48)$$

where $\mathbf{U}_k^{(s)}$ consists of the M_s right unit singular vectors of $[\mathbf{C}_k]_{\mathcal{I}_s, \mathcal{J}_s} \left(\mathbf{T}_k^{(s)} \right)^{1/2}$ associated with its non-zero singular values and

$$\mathbf{T}_k^{(s)} = [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_s} - [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_{\setminus s}} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\Phi_k]_{\mathcal{J}_{\setminus s}, \mathcal{J}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{T}_k]_{\mathcal{I}_{\setminus s}, \mathcal{I}_{\setminus s}}, \quad (4.49)$$

$$\mathbf{G}_{k+n|k}^{(s)} = \left(\mathbf{T}_k^{(s)} \right)^{-1/2} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s}, \quad (4.50)$$

$$\mathbf{G}_{k+n|k}^{(\setminus s)} = \left(\mathbf{T}_k^{(s)} \right)^{-1/2} [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_{\setminus s}} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\mathbf{\Phi}_k]_{\mathcal{J}_{\setminus s}, \mathcal{J}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}, \quad (4.51)$$

$$\mathcal{I}_{\setminus s} = \mathcal{I}_1 \cup \dots \cup \mathcal{I}_{s-1} \cup \mathcal{I}_{s+1} \cup \dots \cup \mathcal{I}_S. \quad (4.52)$$

For any $n \geq 0$ and any index set $\mathcal{I}$, let

$$\begin{aligned} \Xi_{k+n|k}^{(\mathcal{I}, s)} &= [\mathbf{G}_{k+n|k}^{(s)}]_{:, \mathcal{I}} [\mathbf{G}_{k+n|k}^{(s)}]_{:, \mathcal{I}}^\top - [\mathbf{G}_{k+n|k}^{(\setminus s)}]_{:, \mathcal{I}} [\mathbf{G}_{k+n|k}^{(s)}]_{:, \mathcal{I}}^\top \\ &\quad - [\mathbf{G}_{k+n|k}^{(s)}]_{:, \mathcal{I}} [\mathbf{G}_{k+n|k}^{(\setminus s)}]_{:, \mathcal{I}}^\top + [\mathbf{G}_{k+n|k}^{(\setminus s)}]_{:, \mathcal{I}} [\mathbf{G}_{k+n|k}^{(\setminus s)}]_{:, \mathcal{I}}^\top. \end{aligned} \quad (4.53)$$

Then, the utility u_k and privacy loss $\ell_{k+n|k}$ for $n \geq 0$ can be rewritten as

$$\begin{aligned} u_k &= \text{Tr} \left([\mathbf{D}_{k+n|k}]_{\mathcal{P}, \mathcal{P}} \right) \\ &= \text{Tr} \left(\mathbf{U}_k^{(s)\top} \Xi_{k|k}^{(\mathcal{P}, s)} \mathbf{U}_k^{(s)} \right) \\ &\quad + \text{Tr} \left(\left[[\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}^\top [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\mathbf{\Phi}_k]_{\mathcal{J}_{\setminus s}, \mathcal{J}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}} \right]_{\mathcal{P}, \mathcal{P}} \right), \end{aligned} \quad (4.54)$$

and

$$\begin{aligned} \ell_{k+n|k} &= \mathcal{F} \left(\text{vecdiag} \left([\mathbf{D}_{k+n|k}]_{\mathcal{Q}, \mathcal{Q}} \right) \right) \\ &= \mathcal{F} \left(\left[\text{Tr} \left(\mathbf{U}_k^{(s)\top} \Xi_{k+n|k}^{(\mathcal{Q}_1, s)} \mathbf{U}_k^{(s)} \right), \dots, \text{Tr} \left(\mathbf{U}_k^{(s)\top} \Xi_{k+n|k}^{(\mathcal{Q}_Q, s)} \mathbf{U}_k^{(s)} \right) \right]^\top \right) + \\ &\quad \mathcal{F} \left(\text{vecdiag} \left(\left[[\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}^\top [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\mathbf{\Phi}_k]_{\mathcal{J}_{\setminus s}, \mathcal{J}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}} \right]_{\mathcal{Q}, \mathcal{Q}} \right) \right). \end{aligned} \quad (4.55)$$

Since u_k and $\ell_{k+n|k}$ depend on some information that are not available locally at each sensor, the following information need to be shared between the sensors:

1. $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}$, the transformations applied locally at sensor s ;
2. $[\mathbf{H}_k]_{\mathcal{I}_s}$, the local measurement matrix of sensor s ;
3. $[\mathbf{R}_k]_{\mathcal{I}_s, \mathcal{J}_s}$, the covariance matrix of the local measurement noise at sensor s ;

4. $[\mathbf{R}_k]_{\mathcal{I}_s, \mathcal{J}_{\setminus s}}$, the correlation between the local measurement noise at sensor s and that of other sensors.

The first item is used to construct $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_{\setminus s}}$ while the last three items are needed to compute $\mathbf{G}_{k+n|k}^{(\setminus s)}$ at sensor s . However, the correlations between the noise measured at different sensors maybe difficult to know *a priori* in practice if the measurement are made separately. In such a case, we assume the noise measured at different sensors are independent, i.e., $[\mathbf{R}_k]_{\mathcal{I}_i, \mathcal{J}_j} = \mathbf{0}, \forall i \neq j$.

To solve (P4.3), we propose an alternating optimization procedure. We consider a sequential message passing schedule: each sensor in turn transmits messages to all other sensors. Suppose the order of transmission is predefined as $1, 2, \dots, S$. Sensor s finds $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{i\}}$ at iteration i using

$$\left[[\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^{\{i\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_{s-1}, \mathcal{I}_{s-1}}^{\{i\}}, [\mathbf{C}_k]_{\mathcal{J}_{s+1}, \mathcal{I}_{s+1}}^{\{i-1\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^{\{i-1\}} \right] \quad (4.56)$$

as $[\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}$. The details are summarized in Algorithm 5.

Proposition 3. In Algorithm 5, suppose that $u_k^{\{i\}}$ is the utility gain at iteration i . Then $u_k^{\{i\}}$ converges under the sequential schedule.

Proof. For each sensor s and iteration i , let

$$u_k^{(s)}([\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^{\{i\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{i\}}, [\mathbf{C}_k]_{\mathcal{J}_{s+1}, \mathcal{I}_{s+1}}^{\{i-1\}}, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^{\{i-1\}}) \quad (4.57)$$

be the utility at sensor s in iteration i under the sequential schedule. We have

$$\begin{aligned} & u_k^{\{i-1\}} \\ &= u_k^{(S)}([\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^{\{i-1\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^{\{i-1\}}), \text{ iteration } i-1 \text{ at sensor } S \\ &\leq u_k^{(1)}([\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^{\{i\}}, [\mathbf{C}_k]_{\mathcal{J}_2, \mathcal{I}_2}^{\{i-1\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^{\{i-1\}}), \text{ iteration } i \text{ at sensor } 1 \\ &\vdots \\ &\leq u_k^{(s)}([\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^{\{i\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{i\}}, [\mathbf{C}_k]_{\mathcal{J}_{s+1}, \mathcal{I}_{s+1}}^{\{i-1\}}, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^{\{i-1\}}), \text{ iteration } i \text{ at sensor } s \\ &\vdots \end{aligned}$$

$$\begin{aligned}
&\leq u_k^{(S)}([\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^{\{i\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^{\{i\}}), \text{ iteration } i \text{ at sensor } S \\
&= u_k^{\{i\}},
\end{aligned} \tag{4.58}$$

where the above inequalities follow because for each sensor s , $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{i\}}$ is obtained by maximizing the objective function of (P4.3) given

$$[\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^{\{i\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_{s-1}, \mathcal{I}_{s-1}}^{\{i\}}, [\mathbf{C}_k]_{\mathcal{J}_{s+1}, \mathcal{I}_{s+1}}^{\{i-1\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^{\{i-1\}}. \tag{4.59}$$

Since $u_k^{\{i\}} \leq \text{Tr} \left(\left[\tilde{\mathbf{P}}_{k|k-1} \right]_{\mathcal{P}, \mathcal{P}} \right)$ for all $i \geq 1$, the proposition is proved. $\square$

Algorithm 5 Distributed algorithm: sequential update schedule at time step k

Input: $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{0\}} = \mathbf{U}_k^{(s)\top} \left(\mathbf{T}_k^{(s)} \right)^{-1/2}$, $\forall s$ with $M_s^{\{0\}}$, $\mathbf{U}_k^{(s)}$ obtained by solving Algorithm 4 at each sensor individually.

Output: $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{\text{NumIter}\}}$, $s = 1, \dots, S$

- 1: Num-iter $\leftarrow 1$ {iteration index}
- 2: **while** $|u_k^{\{\text{NumIter}\}} - u_k^{\{\text{NumIter}-1\}}| \geq \epsilon$ **do**
- 3: **for** $s = 1$ **to** S **do**
- 4: Broadcast $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{i-1\}}$,
- 5: Broadcast $\left\{ [\mathbf{H}_k]_{\mathcal{I}_s}, [\mathbf{R}_k]_{\mathcal{I}_s, \mathcal{I}_s} \right\}$ once only at $i = 1$
- 6: Receive $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{i-1\}}$,
- 7: Receive $\left\{ [\mathbf{H}_k]_{\mathcal{I}_s}, [\mathbf{R}_k]_{\mathcal{I}_s, \mathcal{I}_s} \right\}$ once only at $i = 1$
- 8: Update $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{i\}}$ by solving (P4.3) where

$$[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^{\{i\}} = \left[[\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^{\{i\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_{s-1}, \mathcal{I}_{s-1}}^{\{i\}}, [\mathbf{C}_k]_{\mathcal{J}_{s+1}, \mathcal{I}_{s+1}}^{\{i-1\}}, \dots, [\mathbf{C}_k]_{\mathcal{J}_S, \mathcal{I}_S}^{\{i-1\}} \right].$$

- 9: **end for**
 - 10: NumIter $\leftarrow$ NumIter + 1
 - 11: **end while**
-

Fig. 4.1 summarizes the relationships between the problem formulations (P4.0)-(P4.3) in Section 4.2 and Section 4.3.

FIG. 4.1: Relationships between the problem formulations (P4.0)-(P4.3).

4.4 Numerical Results

In this section, we present simulation studies to understand the impact of different parameters on the utility-privacy tradeoff in distributed cases. We also verify the performance of our proposed approach on an empirical ultra wideband (UWB) localization system.

4.4.1 Simulation for decentralized sensors

The settings used for the simulations are described as follows. The entries of $\mathbf{H}_k$ at each time step are drawn independently from $\mathcal{N}(0, 1)$, $\mathbf{R}_k = \mathbf{I}_N$, $\mathbf{P}_{0|0} = 0.01\mathbf{I}_L$, $|\mathcal{P}| = 4$, and $|\mathcal{Q}| = 4$. Let $\mathbf{Q}_k = 4\mathbf{I}_L$, $r_k = r_{k,\star}$ in (4.28), $\delta = 3$ and $N = 30$. We randomly generate each $\mathbf{F}_k$ for all $k \geq 1$ so that its singular values are uniformly drawn from $[1, 1.2]$, and use (4.11) as the private error function. Each data point shown in the following figures is averaged over 50 independent experiments. Fig. 4.2 uses $S = 3$ sensors and each sensor s 's measurement has dimension $N_s = N/S = 10$, which is greater than the number of unknowns $L = 8$, whereas Fig. 4.3 uses $S = 5$ sensors with $N_s = N/S = 6$, which is less than L thus making each sensor an under-determined system. Comparing Fig. 4.2 where $N_s = 10$ and Fig. 4.3 where $N_s = 6$, we notice that the compression matrix of $N_s = 6$ is a submatrix of the compression matrix of $N_s = 10$. Thus, τ_k for $N_s = 10$ is less than that for $N_s = 6$.

In Figs. 4.2 and 4.3, while the sequential scheme yields $\min(\eta_k) = \delta$ for all time steps, the "no info. exch." scheme yields $\min(\eta_k) < \delta$ thus no privacy guarantee is ensured. In Fig. 4.3 where each sensor is an under-determined system, the estimation errors of both public and private states are infinitely large at each sensor. Therefore, the no information exchange scheme at each sensor s will choose M_s to be N_s for all $s = 1, \dots, S$, i.e., no compression is used. As a result, after aggregating the measurements from all sensors at fusion center, the estimation errors of both public and private states are as low as the unsanitized ones. This explains why both $\min(\eta_k)$ and τ_k obtained using the no information exchange scheme are very small in Fig. 4.3.

FIG. 4.2: The performance of the sequential scheme where Num-iter = 10 and $N_1, \dots, N_3 = 10$, in terms of public error trace τ_k and the minimum private error variance $\min(\eta_k)$.

Fig. 4.4, which uses the same settings as in Fig. 4.2, shows how τ_k and $\min(\eta_k(\mathbf{C}_k))$ evolve over the time steps k for the sequential scheme. Both τ_k and $\min(\eta_k(\mathbf{C}_k))$ converge after a few iterations.

FIG. 4.3: The performance of the sequential scheme where Num-iter = 10 and $N_1, \dots, N_5 = 6$, in terms of public error trace τ_k and the minimum private error variance $\min(\eta_k)$.

FIG. 4.4: Convergence of τ_k and $\min(\eta_k(\mathbf{C}_k))$ using Algorithm 5, where $S = 3$, $N_1, \dots, N_3 = 10$ and $\delta = 3$. The results shown are for time step $k = 20$.

4.4.2 Privacy-Aware localization

We conduct an experiment using DecaWave UWB sensors for localization. We place 5 anchor nodes at known locations $\mathbf{p}^{(A)} \in \mathbb{R}^{2 \times 5}$ and estimate a mobile node's 2-dimensional (2D) trajectory. The mobile node's state at time k is denoted as

$\mathbf{x}_k = [v_k, \theta_k, p_{x,k}, p_{y,k}]^\top$. The public state $[\mathbf{x}_k]_{\mathcal{P}} = v_k$ is the mobile node's speed while the private state $[\mathbf{x}_k]_{\mathcal{Q}} = [\theta_k, p_{x,k}, p_{y,k}]^\top$ contains the node's heading and 2D location. The measurements are anchor-to-mobile ranges, which are a non-linear function of the state: $\mathbf{z}_k = h_k(\mathbf{x}_k) + \mathbf{n}_k$. We use the extended Kalman filter (EKF) in our estimation procedure. At each time k , we linearize h_k and f_k around the estimate $\hat{\mathbf{x}}_{k|k-1}$ and define the measurement matrix and the transition matrix, respectively, to be $\mathbf{H}_k = \left. \frac{\partial h_k(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{k|k-1}}$ and $\mathbf{F}_k = \left. \frac{\partial f_k(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\hat{\mathbf{x}}_{k-1|k-1}}$, where $f_k(\mathbf{x}_k) = [v_k, \theta_k, p_{x,k} + \Delta v_k \cos \theta_k, p_{y,k} + \Delta v_k \sin \theta_k]^\top$ with $\Delta = 0.1$ being the sampling interval. The transition matrix is approximated by

$$\mathbf{F}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \Delta \cos \hat{\theta}_{k-1} & -\Delta \sin \hat{\theta}_{k-1} \hat{v}_{k-1} & 1 & 0 \\ \Delta \sin \hat{\theta}_{k-1} & \Delta \cos \hat{\theta}_{k-1} \hat{v}_{k-1} & 1 & 0 \end{bmatrix}. \quad (4.60)$$

The measurement $\mathbf{z}_k \in \mathbb{R}^{5 \times 1}$ contains range measurements between the mobile node and anchor nodes with

$$[\mathbf{z}_k]_i = \left\| [\mathbf{p}^{(A)}]_{:,i} - [\mathbf{x}_k]_{\mathcal{Q}} \right\| + [\mathbf{n}_k]_i, i = 1, \dots, 5, \quad (4.61)$$

where $\|\cdot\|$ denotes the Euclidean norm. Under the EKF framework, we let the measurement matrix $\mathbf{H}_k \in \mathbb{R}^{5 \times 4}$ to be given by

$$[\mathbf{H}_k]_i = - \begin{bmatrix} 0, 0, \frac{[\mathbf{p}^{(A)}]_{1,i} - [\hat{\mathbf{x}}_{k|k-1}]_3}{\left\| [\mathbf{p}^{(A)}]_{:,i} - [\hat{\mathbf{x}}_{k|k-1}]_{\mathcal{Q}} \right\|}, \frac{[\mathbf{p}^{(A)}]_{2,i} - [\hat{\mathbf{x}}_{k|k-1}]_4}{\left\| [\mathbf{p}^{(A)}]_{:,i} - [\hat{\mathbf{x}}_{k|k-1}]_{\mathcal{Q}} \right\|} \end{bmatrix}, \quad i = 1, \dots, 5, \quad (4.62)$$

where $\hat{\mathbf{x}}_{k|k-1}$ is the predicted state before measurements at time k . Recall that we want to protect the mobile node's location while estimating its velocity, which however is not directly observed. In Fig. 4.6 and Fig. 4.5, we consider the centralized setting where the mobile node collects the ranging measurements and sanitizes

them before sending to a server. We set $\mathbf{P}_0 = 0.01\mathbf{I}_4$ and $\mathbf{R}_k = 0.04\mathbf{I}_5$, $\mathbf{Q}_k = 0.04\mathbf{I}_4$, and $r_k = 0$ for all time steps k . The private error function used is (4.10).

The mobile node's state is estimated over 318 time steps with or without compressive sanitization. We let $|\mathcal{Q}|\delta = 2$. The trajectory estimated with compressive sanitization (red crosses in Fig. 4.5) is noticeably distorted as compared to the ground true line (black line) thus the private information is protected whereas the private information is exposed in the trajectory estimated without compressive sanitization (blue dashes), which matches the ground truth line well. On the other hand, the public information (red line in Fig. 4.6), i.e., the speed of the mobile node matches fairly well with the unsanitized one (black line). This experiment shows the proposed approach is also applicable to non-linear dynamical systems that can be approximated using the EKF framework.

FIG. 4.5: Estimation of the mobile node's locations (private states).

4.5 Comparisons with Affine Transformation

In Chapter 3, we have shown that it is possible for the privacy to be arbitrarily strong when choosing additive noise perturbation as the privacy mechanism. However, noise perturbation scheme requires hardware random number generator, which is not available on all devices. In this chapter, we adopted linear compression

FIG. 4.6: Estimation of the mobile node's speed (public states).

as the data sanitization scheme due to its simplicity and practicability. For comparative purposes, it is of great interest to consider combining linear compression and noise perturbation for data sanitization. We refer to this combined sanitization scheme as *affine transformation*.

In what follows, we replace the compression mechanism in (4.7) with an affine transformation for sanitizing sensor measurements. Accordingly, we develop a method to optimize the utility-privacy tradeoff given by the formulation (P4.1) over an affine transformation. We conduct simulations to compare the utility-privacy tradeoffs w.r.t. affine transformation and linear compression in the single sensor case.

4.5.1 Affine Transformation

Instead of compressing the measurements as shown in (4.7), we propose to compress the measurements and then add noise perturbation into the compressed measurements to protect the private state $[\mathbf{x}_k]_Q$. We call this an affine transformation. The

perturbed measurement after an affine transformation at time k can be written as:

$$\tilde{\mathbf{z}}_k = \mathbf{C}_k \mathbf{z}_k + \boldsymbol{\xi}_k = \mathbf{C}_k \mathbf{H}_k \mathbf{x}_k + \mathbf{C}_k \mathbf{n}_k + \boldsymbol{\xi}_k, \quad (4.63)$$

where $\mathbf{C}_k \in \mathbb{R}^{M \times N}$ is a compression matrix and $\boldsymbol{\xi}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_k)$ is a perturbation noise. From this perturbed measurement $\tilde{\mathbf{z}}_k$, the state estimate and error covariance as shown in (4.4) and (4.5) become

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{K}_k (\tilde{\mathbf{z}}_k - \mathbf{C}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}), \\ \tilde{\mathbf{P}}_{k|k} &= \tilde{\mathbf{P}}_{k|k-1} - \mathbf{K}_k \mathbf{C}_k \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}, \end{aligned} \quad (4.64)$$

where

$$\begin{aligned} \tilde{\mathbf{P}}_{k|k-1} &= \mathbf{F}_k \tilde{\mathbf{P}}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k, \\ \mathbf{S}_k &= \mathbf{C}_k \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T \mathbf{C}_k^T + \mathbf{C}_k \mathbf{R}_k \mathbf{C}_k^T + \boldsymbol{\Sigma}_k, \\ \mathbf{K}_k &= \tilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T \mathbf{C}_k^T \mathbf{S}_k^{-1}. \end{aligned} \quad (4.65)$$

Recall the notations in (4.18) and (4.19):

$$\begin{aligned} \mathbf{T}_k &= \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k, \\ \mathbf{G}_{k+n|k} &= \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1} \mathbf{F}_{k+1:k+n}^T. \end{aligned} \quad (4.66)$$

The reduction of the error covariance at time k due to the perturbed measurement $\tilde{\mathbf{z}}_k$ as shown in (4.21) can now be rewritten as

$$\mathbf{D}_{k|k}(\mathbf{C}_k, \boldsymbol{\Sigma}_k) = \mathbf{G}_{k|k}^T \mathbf{C}_k^T (\mathbf{C}_k \mathbf{T}_k \mathbf{C}_k^T + \boldsymbol{\Sigma}_k)^{-1} \mathbf{C}_k \mathbf{G}_{k|k}. \quad (4.67)$$

Correspondingly, the look-ahead error reduction $\mathbf{D}_{k+n|k}(\mathbf{C}_k, \boldsymbol{\Sigma}_k)$ defined in (4.20) can be expressed as

$$\mathbf{D}_{k+n|k}(\mathbf{C}_k, \boldsymbol{\Sigma}_k) = \mathbf{G}_{k+n|k}^T \mathbf{C}_k^T (\mathbf{C}_k \mathbf{T}_k \mathbf{C}_k^T + \boldsymbol{\Sigma}_k)^{-1} \mathbf{C}_k \mathbf{G}_{k+n|k}. \quad (4.68)$$

Lemma 8 implies the optimization problem (P4.2) is invariant to noise covariance matrix Σ_k .

Lemma 8. For every pair of $\mathbf{C}_k, \Sigma_k$, there exists $\mathbf{C}'_k$ such that $\mathbf{D}_{k+n|k}(\mathbf{C}_k, \Sigma_k) = \mathbf{D}_{k+n|k}(\mathbf{C}'_k, \mathbf{I}_M)$.

Proof. Let $\mathbf{C}'_k = \Sigma_k^{-1/2} \mathbf{C}_k$, we have

$$\begin{aligned}
& \mathbf{D}_{k+n|k}(\mathbf{C}_k, \Sigma_k), \\
&= \mathbf{G}_{k+n|k}^\top \mathbf{C}_k^\top (\mathbf{C}_k \mathbf{T}_k \mathbf{C}_k^\top + \Sigma_k)^{-1} \mathbf{C}_k \mathbf{G}_{k+n|k}, \\
&= \mathbf{G}_{k+n|k}^\top \mathbf{C}_k^\top \Sigma_k^{-1/2} \left(\Sigma_k^{-1/2} \mathbf{C}_k \mathbf{T}_k \mathbf{C}_k^\top \Sigma_k^{-1/2} + \mathbf{I}_M \right)^{-1} \Sigma_k^{-1/2} \mathbf{C}_k \mathbf{G}_{k+n|k}, \\
&= \mathbf{G}_{k+n|k}^\top \mathbf{C}'_k{}^\top (\mathbf{C}'_k \mathbf{T}_k \mathbf{C}'_k{}^\top + \mathbf{I}_M)^{-1} \mathbf{C}'_k \mathbf{G}_{k+n|k}, \\
&= \mathbf{D}_{k+n|k}(\mathbf{C}'_k, \mathbf{I}_M). \tag{4.69}
\end{aligned}$$

□

Therefore, we can fix $\Sigma_k = \mathbf{I}_M$, and optimize over $\mathbf{C}_k$ instead of optimizing over $\mathbf{C}_k$ and Σ_k for an affine transformation. Then we discuss the solution for (P4.2) under an affine transformation. Replacing $\mathbf{C}_k$ in $\mathbf{D}_{k+n|k}(\mathbf{C}_k, \mathbf{I}_M)$ by $\mathbf{C}'_k \mathbf{T}_k^{-1/2}$, we have

$$\mathbf{D}_{k+n|k} = \mathbf{G}_{k+n|k}^\top \mathbf{T}_k^{-1/2} \mathbf{C}'_k{}^\top (\mathbf{C}'_k \mathbf{C}'_k{}^\top + \mathbf{I}_M)^{-1} \mathbf{C}'_k \mathbf{T}_k^{-1/2} \mathbf{G}_{k+n|k}, \tag{4.70}$$

Let the singular value decomposition of $\mathbf{C}'_k$ be $\mathbf{V}_k \Phi_k \mathbf{U}_k^\top$, where $\mathbf{U}_k \in \mathbb{R}^{N \times M}$ consists of the M right singular vectors associated with the M non-zero singular values, $\Phi_k = \text{diag}(\phi_1, \dots, \phi_M) \in \mathbb{R}^{M \times M}$ is a diagonal matrix with non-zeros singular values being its diagonal elements and $\mathbf{V}_k \in \mathbb{R}^{M \times M}$ contains the M left singular vectors associated with the M non-zero singular values. Then we obtain

$$\mathbf{C}'_k{}^\top (\mathbf{C}'_k \mathbf{C}'_k{}^\top + \mathbf{I}_M)^{-1} \mathbf{C}'_k = \mathbf{U}_k \Lambda_k \mathbf{U}_k^\top, \tag{4.71}$$

where $\mathbf{\Lambda}_k = \text{diag}(\lambda_1, \dots, \lambda_M)$ with $\lambda_i = \frac{\phi_i^2}{\phi_i^2 + 1}$. Substituting the above equation into (4.70), we have

$$\mathbf{D}_{k+n|k} = \mathbf{G}_{k+n|k}^\top \mathbf{T}_k^{-1/2} \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^\top \mathbf{T}_k^{-1/2} \mathbf{G}_{k+n|k}. \quad (4.72)$$

We let $\mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^\top = \mathbf{\Omega}_k$, which is a positive semi-definite matrix with its eigenvalues smaller than 1. This allows us to solve (P4.2) by semidefinite programming (SDP) [116].

Reuse the notation in (4.36):

$$\mathbf{\Theta}_{k+n|k}^{(\mathcal{I})} = \left[\mathbf{T}_k^{-1/2} \mathbf{G}_{k+n|k} \right]_{:, \mathcal{I}} \left[\mathbf{T}_k^{-1/2} \mathbf{G}_{k+n|k} \right]_{:, \mathcal{I}}^\top. \quad (4.73)$$

Consequently, the utility function and privacy function shown in (4.37) and (4.38) can be rewritten as

$$\begin{aligned} u_k(\mathbf{\Omega}_k) &= \text{Tr} \left(\mathbf{\Theta}_k^{(\mathcal{P})} \mathbf{\Omega}_k \right), \\ \ell_{k+n|k}(\mathbf{\Omega}_k) &= \mathcal{F} \left(\left[\text{Tr} \left(\mathbf{\Theta}_{k+n|k}^{(\mathcal{Q}_1)} \mathbf{\Omega}_k \right), \dots, \text{Tr} \left(\mathbf{\Theta}_{k+n|k}^{(\mathcal{Q}_{|\mathcal{Q}|})} \mathbf{\Omega}_k \right) \right]^\top \right). \end{aligned} \quad (4.74)$$

The problem (P4.2) can be cast as a semidefinite programming [116]:

$$\begin{aligned} \max_{\mathbf{\Omega}_k} \quad & u_k(\mathbf{\Omega}_k), \\ \text{s.t.} \quad & \ell_{k+n|k}(\mathbf{\Omega}_k) \leq \bar{\delta}_{k+n|k}, \quad \forall n = 0, \dots, r_k, \\ & \mathbf{I}_N \succeq \mathbf{\Omega}_k \succeq \mathbf{0}. \end{aligned} \quad (4.75)$$

The optimal $\mathbf{C}_k$ can be recovered from the optimal $\mathbf{\Omega}_k$ after solving this optimization. Note that the optimal compression size M is same as the rank of $\mathbf{\Omega}_k$.

4.5.2 Simulation Results

In this section, we carry out numerical experiments to compare the utility-privacy tradeoffs given by linear compression and affine transformation. The settings used

for the simulations are described as follows. We choose $|\mathcal{P}| = |\mathcal{Q}| = 3$, $N = 6$, $N_s = 1$, $r_k = r_{k,\star}$ in (4.28) and $\delta = 3$. For $k = 1, \dots, 20$, the entries of all $\mathbf{F}_k$ and $\mathbf{H}_k$ are drawn independently from $\text{Unif}(-0.6, 0.6)$ and $\text{Unif}(-0.5, 0.5)$, respectively, and the eigenvalues of $\mathbf{R}_k$ and $\mathbf{Q}_k$ are drawn independently from $\text{Unif}(0, 1)$ and $\text{Unif}(0, 0.5)$, respectively, with their eigenvectors are constructed by the orthogonal basis for range of a random matrix whose entries are drawn independently from $\text{Unif}(0, 1)$. Each data point in Fig. 4.7 is averaged over 50 independent experiments.

Fig. 4.7 shows the estimation error of the public states by solving (P4.2) based on Algorithm 4 for linear compression and the optimization (4.75) for affine transformation. It is observed that the optimal utility given by affine transformation is slightly better than the utility provided by linear compression when both methods satisfying the same privacy budget. Therefore, combining noise perturbation and linear compression for data sanitization provides only trivial improvement for the utility-privacy tradeoff.

FIG. 4.7: The optimal estimation error of the public states at each step.

4.6 Conclusion

We have studied the use of compressive privacy schemes in an LDS to prevent the fusion center from estimating a set of private states accurately while still allowing it to estimate a set of public states with good accuracy. We developed an optimization framework to find the optimal compression matrix at each time step to achieve an optimal tradeoff between the utility at the current time step and privacy protection at multiple steps in the future. We showed that this approach allows us to ensure the same level of privacy in all future time steps in an LDS. One empirical experiment was conducted to verify that our proposed approach on real-world time series data.

4.A Proof of Lemma 5

Let $\mathbf{A}' = \mathbf{A}\mathbf{B}^{1/2}$. Let $\mathbf{V}\mathbf{\Lambda}\mathbf{U}^\top$ be the truncated version of the singular value decomposition (SVD) of $\mathbf{A}'$, where $\mathbf{\Lambda} \in \mathbb{R}^{M \times M}$ is a diagonal matrix with non-zero singular values being its diagonal elements, $\mathbf{V} \in \mathbb{R}^{M \times M}$ contains the M left singular vectors, and $\mathbf{U}$ is as defined in the lemma statement. Substituting the SVD of $\mathbf{A}'$ into the left-hand side of (4.34), we obtain the right-hand side. This completes the proof.

4.B Proof of Lemma 7

Let $\{\lambda_{\mathbf{A},i}, i = 1, \dots, N\}$ and $\{\lambda_{-\mathbf{B},i}, i = 1, \dots, N\}$ denote the eigenvalues of $\mathbf{A}$ and $-\mathbf{B}$, respectively, sorted in descending order. The eigenvalues of $\mathbf{C}$, denoted as $\{\lambda_{\mathbf{C},i}, i = 1, \dots, N\}$ sorted in descending order, satisfy Weyl's inequality:

$$\lambda_{\mathbf{A},N} + \lambda_{-\mathbf{B},i} \leq \lambda_{\mathbf{C},i} \leq \lambda_{\mathbf{A},i} + \lambda_{-\mathbf{B},1}, i = 1, \dots, N. \quad (4.76)$$

Since $N > r_{\mathbf{A}} + r_{\mathbf{B}}$, $\lambda_{\mathbf{A},i} = 0$ for $i > r_{\mathbf{A}}$ and $\lambda_{-\mathbf{B},i} = 0$ for $i \leq N - r_{\mathbf{B}}$. Then, we have

$$\begin{aligned} 0 &\leq \lambda_{\mathbf{C},i} \leq \lambda_{\mathbf{A},i}, \quad 1 \leq i \leq r_{\mathbf{A}}, \\ \lambda_{\mathbf{C},i} &= 0, \quad r_{\mathbf{A}} + 1 \leq i \leq N - r_{\mathbf{B}} \\ \lambda_{-\mathbf{B},i} &\leq \lambda_{\mathbf{C},i} \leq 0, \quad N - r_{\mathbf{B}} + 1 \leq i \leq N. \end{aligned} \quad (4.77)$$

The proof is now complete.

4.C Derivation of (4.48)

Let $\mathbf{\Pi}_{i,j}$ be a permutation matrix that swaps the i th and the j th columns of the identity matrix $\mathbf{I}_S$. Then, right multiplying a matrix by

$$\mathbf{\Pi}_{s \rightarrow 1} = \mathbf{\Pi}_{s,s-1} \mathbf{\Pi}_{s-1,s-2} \cdots \mathbf{\Pi}_{2,1} \quad (4.78)$$

ends up moving the s th column of this matrix to its 1st column and it's easy to verify that $\mathbf{\Pi}_{s \rightarrow 1} \mathbf{\Pi}_{s \rightarrow 1}^T = \mathbf{I}_S$. In what follows, we firstly show that $\mathbf{D}_{k+n|k}$ does not depend on the order of sensors and then separate the terms depending on $[\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}$ in $\mathbf{D}_{k+n|k}$ from the terms depending on $[\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}$. From (4.46), we have

$$\begin{aligned} \mathbf{D}_{k+n|k} &= \left[[\mathbf{G}_{k+n|k}]_{\mathcal{I}_1}^T [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^T, \dots, [\mathbf{G}_{k+n|k}]_{\mathcal{I}_S}^T [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^T \right] \mathbf{\Pi}_{s \rightarrow 1} \\ &\quad \mathbf{\Pi}_{s \rightarrow 1}^T \mathbf{\Psi}_k^{-1} \mathbf{\Pi}_{s \rightarrow 1} \mathbf{\Pi}_{s \rightarrow 1}^T \begin{bmatrix} [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_1} \\ \vdots \\ [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s} \end{bmatrix}, \\ &= \left[[\mathbf{G}_{k+n|k}]_{\mathcal{I}_1}^T [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1}^T, \dots, [\mathbf{G}_{k+n|k}]_{\mathcal{I}_S}^T [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^T \right] \mathbf{\Pi}_{s \rightarrow 1} \\ &\quad (\mathbf{\Pi}_{s \rightarrow 1}^T \mathbf{\Psi}_k \mathbf{\Pi}_{s \rightarrow 1})^{-1} \mathbf{\Pi}_{s \rightarrow 1}^T \begin{bmatrix} [\mathbf{C}_k]_{\mathcal{J}_1, \mathcal{I}_1} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_1} \\ \vdots \\ [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s} \end{bmatrix}, \end{aligned}$$

$$\begin{aligned}
&= \begin{bmatrix} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s}^\top [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top, [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}^\top [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top \\ \begin{bmatrix} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_s} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top & [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_{\setminus s}} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top \\ [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_{\setminus s}}^\top [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top & [\mathbf{\Phi}_k]_{\mathcal{I}_{\setminus s}, \mathcal{I}_{\setminus s}} \end{bmatrix} \end{bmatrix}^{-1} \\
&\quad \begin{bmatrix} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s} \\ [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}} \end{bmatrix}, \\
&= [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s}^\top [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top \left([\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} \mathbf{T}_k^{(s)} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top \right)^{-1} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s} \\
&\quad - [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s}^\top [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top \left([\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} \mathbf{T}_k^{(s)} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top \right)^{-1} \\
&\quad [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_{\setminus s}} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\mathbf{\Phi}_k]_{\mathcal{I}_{\setminus s}, \mathcal{I}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}} \\
&\quad - [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}^\top [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\mathbf{\Phi}_k]_{\mathcal{I}_{\setminus s}, \mathcal{I}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_{\setminus s}}^\top \\
&\quad [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top \left([\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} \mathbf{T}_k^{(s)} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top \right)^{-1} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_s} \\
&\quad + [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}^\top [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\mathbf{\Phi}_k]_{\mathcal{I}_{\setminus s}, \mathcal{I}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_{\setminus s}}^\top \\
&\quad [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top \left([\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} \mathbf{T}_k^{(s)} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s}^\top \right)^{-1} [\mathbf{C}_k]_{\mathcal{J}_s, \mathcal{I}_s} [\mathbf{T}_k]_{\mathcal{I}_s, \mathcal{I}_{\setminus s}} \\
&\quad [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\mathbf{\Phi}_k]_{\mathcal{I}_{\setminus s}, \mathcal{I}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}} \\
&\quad + [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}^\top [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}}^\top [\mathbf{\Phi}_k]_{\mathcal{I}_{\setminus s}, \mathcal{I}_{\setminus s}}^{-1} [\mathbf{C}_k]_{\mathcal{J}_{\setminus s}, \mathcal{I}_{\setminus s}} [\mathbf{G}_{k+n|k}]_{\mathcal{I}_{\setminus s}}, \tag{4.79}
\end{aligned}$$

where the last equality follows from the matrix inversion lemma. Applying Lemma 5, we obtain (4.48).

Chapter 5

Learning-Based Privacy Protection

The privacy-preserving frameworks delineated in Chapters 3 and 4 require prior knowledge of the system model (model-dependent). Hence they are incapable of coping with complex data who lack of an exact descriptive model and the application is limited to the field of classical signal processing. Concurrently, machine learning has achieved tremendous success in handling data of complex nature and thus prevails in challenging tasks ranging from object recognition to language translation. When a client sends data to a service provider for it to perform inference for an authorized task, the service provider may exploit the raw data to extract other sensitive information it has not been authorized for. For example, users' personal information such as sexual orientation and political affiliation can be accurately inferred from their activities on social networking platforms. In this chapter, we propose a Data-driven Regularized Inference Privacy (DRIP) framework for meaningfully sanitizing raw data to prevent the exposure of private information when sharing data with an untrusted service provider. We develop practical implementation of mutual information and maximal correlation to quantify privacy leakage, and conduct asymptotic convergence analysis. We ensure that the sanitized data produced by our framework is compatible with a service provider's legacy inference system.

5.1 Motivations

We propose a Data-driven Regularized Inference Privacy (DRIP) framework for meaningfully sanitizing raw data to prevent the exposure of private information when sharing data with an untrusted service provider. We ensure that the sanitized data produced by our framework is compatible with a service provider's legacy inference system. With the goal of developing a practical approach to information or estimation-theoretic privacy in mind, our proposed framework consists of three components:

- The use of mutual information as a *utility metric*.
- The use of either average information leakage or maximal correlation as a *privacy metric*.
- The use of *domain regularization* to minimize the disparity between the distributions of the sanitized data and raw data.

In the following, we discuss the motivations for the choices we have adopted in DRIP.

We quantify privacy leakage by either mutual information or maximal correlation, each of which measures statistical dependence between the sanitized data and the private variable. Both average information leakage (mutual information) and maximal correlation are extensively employed as measures of privacy leakage in the *inference privacy* regime with known underlying data distributions [47–49]. However, utilizing them for data-driven privacy had not been explored. The connections between mutual information and maximal correlation as privacy metrics are studied in [57, 76]. We refer the reader to details in [50, 58]. Let $S \in \mathcal{S}$ and Z be a private variable and a sanitized variable that are jointly distributed. A small mutual information between S and Z ensures that only limited bits of S are revealed when releasing Z . It can also be justified using Fano's inequality: $I(S; Z) \leq \epsilon$ implies that $\mathbb{P}(\hat{S}(Z) \neq S) \geq (H(S) - 1 - \epsilon)/\log|\mathcal{S}|$, where $I(\cdot; \cdot)$

and $H(\cdot)$ are the mutual information and entropy functions, respectively. Hence, the probability of the adversary correctly guessing S is upper-bounded. On the other hand, a small maximal correlation between S and Z gives an assurance that an adversary observing Z cannot efficiently estimate $f(S)$ for any real-valued function f . It is shown in [50] that mutual information bounds maximal correlation under some technical conditions (e.g., when (S, Z) are jointly Gaussian).

Although comparing the two privacy metrics is beyond the scope of this paper, we discuss the strengths and shortcomings of their data-driven implementations in Section 5.4, which reveals that the feasibility of the two metrics entails different configurations. In summary, adopting mutual information as a privacy metric using the variational method is only suited for the case where the probability distribution of the private variable conditioned on each raw data sample is known and an effective variational distribution is available. In such cases and under mild technical conditions [50], mutual information is preferred as it is a stronger privacy metric compared to maximal correlation. For other cases where the conditional distribution is unknown, deploying maximal correlation as an estimation-theoretic privacy metric is applicable.

Akin to the rate-distortion problem [59], the utility cost of sanitizing raw data is typically measured by a distortion function like Euclidean distance in information or estimation-theoretic privacy frameworks. However, for unstructured data such as images and texts, a greater amount of such distortion does not imply less utility, e.g., shifting and rotating images (which keep intact statistical information) can incur a large per-pixel loss. Inspired by the papers [126, 127], we propose the use of mutual information between raw and sanitized data as a utility metric because it can capture salient public information. Moreover, we will be able to embed other sophisticated metrics when building the variational posterior distribution function (cf. Section 5.6.2). For example, the reference [128] has shown that the use of perceptual loss based on high-level features extracted from pre-trained networks for measuring image similarity gives visually pleasing results for image transformation.

To determine the domain of the sanitized output is another challenge for data-driven privacy. The output from the sanitizer can be arbitrary without proper domain regularization. When the data distribution is known and the source alphabet has finite support, the problem reduces to choosing a proper size for the alphabet of the sanitized output, which is a relatively simple task [76, 78, 129]. In an attempt to discipline the domain of the sanitized data, we include a regularization term into the privacy-utility tradeoff that penalizes high distribution disparity between the sanitized data and raw data. We propose to use maximum mean discrepancy and domain adaptation to realize this. Domain regularization helps to lessen the amount of over-fitting when training privacy and utility models as sanitizing raw data is a highly under-determined process. Furthermore, mapping raw data into a similar space makes sanitized data compatible with a service provider's legacy learning systems that are purportedly designed for or trained on the raw data only. Our proposed approach thus incurs no additional cost for the service provider.

It is worth mentioning that utility quantified by mutual information is also a measure of distribution disparity in terms of Kullback–Leibler divergence, which ostensibly has the same effect as domain regularization. However, mutual information and domain regularization achieve completely different purposes. Let X and Z represent raw data and sanitized data. The mutual information $I(X; Z)$ measures the divergence between the conditional distribution $p_{X|Z}$ and marginal distribution p_X averaged over Z , while regularization aims to reduce the divergence between marginal distributions p_X and p_Z . More intuitively, $I(X; Z) = I(X; f(Z))$ for any bijective function f and thus the use of mutual information as a utility metric has no control over the domain of the sanitized data.

5.2 Problem Formulation

We consider a random element or variable $X \in \mathcal{X}$ on which a private element or variable $S \in \mathcal{S}$ is statistically dependent. (In general, both X and S can be

random vectors. For simplicity, we refer to these as random variables throughout.) A client has an instance of X and intends to send it to a service provider (e.g., a cloud service) who can help her to infer some public information from X . However, she is also concerned that the service provider can exploit the raw data X to infer the private variable S . For example, users may wish not to reveal their ethnicity or gender when uploading their facial images to a server for verification purposes. To protect S , the data curator sanitizes the raw data X to reduce the amount of private information, while maximally retaining the rest of the information to preserve utility. As sketched in Fig. 5.1, the data curator sends a sanitized version of data $\tilde{X} \in \tilde{\mathcal{X}}$ to the service provider.

FIG. 5.1: Privacy framework.

The joint pdf $p_{X,S}$ of (X, S) is unknown and is difficult to be estimated when X is high-dimensional or unstructured data. This prevents us from applying a theoretical privacy-preserving framework like those in [43, 47, 48, 66, 70, 71]. Instead, we assume the availability of a training dataset $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$, where each $(\mathbf{x}_i, \mathbf{s}_i)$ is an independent instance of the pair of random variables (X, S) . Our task is to learn a privacy-preserving probabilistic function that maps the raw data X to sanitized data $\tilde{X} \in \tilde{\mathcal{X}}$:

$$p_{\tilde{X}|X} : \mathcal{X} \times \tilde{\mathcal{X}} \mapsto [0, 1]. \quad (5.1)$$

In theory, the search space of the probabilistic mapping $p_{\tilde{X}|X}$ is not constrained, which makes finding a mapping that optimizes the tradeoff between utility and privacy computationally expensive in some cases. To make the problem more tractable, we restrict to a rich class of mappings by relating the mappings to

a parametrized differentiable function via *reparameterization*: Under certain mild conditions (cf. Section 2.4 in [130]) for a chosen posterior $p_{\tilde{X}|X}$, we can reparametrize the random variable $\tilde{X}$ using a differentiable transformation of an auxiliary continuous random variable Ξ independent of X :

$$\tilde{X} = f^{\boldsymbol{\theta}}(X, \Xi), \quad \Xi \sim p_{\Xi}, \quad (5.2)$$

where $f^{\boldsymbol{\theta}}$ is a function parametrized by $\boldsymbol{\theta}$ and differentiable w.r.t. $\boldsymbol{\theta}$. In practice, we let $f^{\boldsymbol{\theta}}$ be represented by a neural network because of its ability to approximate a rich class of functions efficiently. The reparameterization allows us to apply the stochastic gradient-descent method to optimize over the sanitization parameter $\boldsymbol{\theta}$. For the sake of notational clarity, we rewrite $p_{\tilde{X}|X}$ as $p_{\tilde{X}|X}^{\boldsymbol{\theta}}$.

We seek to optimize over $\boldsymbol{\theta}$ to find an optimal sanitizer that maximizes utility with a penalty term for domain regularization, subject to a constraint on the privacy leakage:

$$\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) - \alpha \mathcal{R}(\boldsymbol{\theta}), \quad \text{s. t. } \mathcal{P}(\boldsymbol{\theta}) \leq \epsilon, \quad (5.3)$$

where α is a positive constant and ϵ is the privacy budget, the utility function $\mathcal{L}(\boldsymbol{\theta})$ quantifies the amount of information about X preserved by the sanitized data $\tilde{X}$, the privacy function $\mathcal{P}(\boldsymbol{\theta})$ measures the statistical dependence between the private variable S and sanitized data $\tilde{X}$, and the regularization function $\mathcal{R}(\boldsymbol{\theta})$ measures divergence between the marginal distributions of the sanitized data and raw data. As alluded to in Section 5.1, $\mathcal{R}(\boldsymbol{\theta})$ used for regularizing the domain of $\tilde{X}$ is essential for data-driven privacy as the sanitized output in (5.2) can be arbitrary and thus cause over-fitting for the privacy and utility models. Apart from that, domain regularization makes the sanitized data $\tilde{X}$ compatible with a legacy inference system that is designed to work with the raw unsanitized data X . This allows the service provider to save costs and time as it can reuse the legacy system for the sanitized data. We call this framework DRIP. This formulation is

comparable to the privacy funnel problem [66] if mutual information serves as the utility and privacy measure.

The constrained optimization (5.3) can be solved via the penalty method [131]:

$$\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) - \alpha \mathcal{R}(\boldsymbol{\theta}) - \lambda g(\mathcal{P}(\boldsymbol{\theta})), \quad (5.4)$$

where $g(\mathcal{P}(\boldsymbol{\theta})) = \max\{0, \mathcal{P}(\boldsymbol{\theta}) - \epsilon\}^2$ is the exterior penalty function and λ is the penalty coefficient.

We build the sanitizer, privacy and utility functions, and regularizer in our DRIP framework using machine learning models, each with their own learnable parameters. This allows us to back-propagate the gradient through the sanitizer network (see Fig. 5.2) to update the parameter $\boldsymbol{\theta}$ and learn the optimal sanitizer. We discuss these models and their data-driven implementations in the following sections.

FIG. 5.2: Machine learning DRIP architecture.

5.3 Utility Metric and Regularization

In this section, we present the formulation of the utility function $\mathcal{L}(\boldsymbol{\theta})$ using the variational method [130]. We discuss the use of maximum mean discrepancy and domain adaptation methods for constructing the domain regularization function $\mathcal{R}(\boldsymbol{\theta})$.

5.3.1 Utility Function

When sanitizing the raw data X to reduce the private information leakage, we also seek to retain as much of the non-private information as possible to maintain the fidelity of the sanitized data so as to preserve utility. As previously mentioned, the use of distortion as a utility metric makes little sense for unstructured data while mutual information, measuring statistical dependencies, can capture implicit non-private information from the raw data. While the expression of mutual information is concise and elegant in theory, its computation is challenging as it requires the data distribution to be known, which is often not possible in practice. To overcome this, we propose to use the variational method [130] to approximate the mutual information between X and $\tilde{X}$.

Consider the privacy-preserving probabilistic mapping $p_{\tilde{X}|X}^{\boldsymbol{\theta}}$ with a fixed parameter $\boldsymbol{\theta}$. The mutual information between X and $\tilde{X}$ is

$$I(X; \tilde{X}) = H(X) - H(X|\tilde{X}). \quad (5.5)$$

Since the raw data distribution p_X is implicitly specified, $H(X)$ is a constant. Therefore, we are only interested in estimating

$$H(X|\tilde{X}) = -\mathbb{E}_{(X, \tilde{X}) \sim p_{X, \tilde{X}}^{\boldsymbol{\theta}}} \left[\log p_{X|\tilde{X}}(X|\tilde{X}) \right], \quad (5.6)$$

where $p_{X, \tilde{X}}^{\boldsymbol{\theta}}(\mathbf{x}, \tilde{\mathbf{x}}) = p_{\tilde{X}|X}^{\boldsymbol{\theta}}(\tilde{\mathbf{x}}|\mathbf{x})p_X(\mathbf{x})$. Here $p_{\tilde{X}|X}^{\boldsymbol{\theta}}$ and $p_{X|\tilde{X}}$ bear some resemblance to the notion of the encoder and decoder of an autoencoder, respectively. While an autoencoder encodes X to a latent representation, we wish to render $\tilde{X}$ as an approximation of X with private information removed. To avoid confusion, we refer to $p_{X|\tilde{X}}^{\boldsymbol{\theta}}$ as the posterior of the raw data X . Note that computing $p_{X|\tilde{X}}$ is intractable since p_X is unknown to us. Even if p_X is given, it is not practical to derive $p_{X|\tilde{X}}$ via Bayes' theorem because the computation of the evidence $p_{\tilde{X}}$ involves an integral over the sample space of X , which can be a high dimensional

space. To overcome this difficulty, we employ the variational method to obtain an upper bound for $H(X|\tilde{X})$.

Let $q_{X|\tilde{X}}^\phi(\cdot|\cdot)$ be a proposed variational distribution function parametrized by ϕ , representing a set of candidate functions used for approximating the true distribution $p_{X|\tilde{X}}$. We have

$$\begin{aligned} & -H(X|\tilde{X}) \\ &= \mathbb{E}_{(X,\tilde{X}) \sim p_{X,\tilde{X}}^\theta} \left[\log p_{X|\tilde{X}}(X|\tilde{X}) \right] \\ &= \mathbb{E}_{(X,\tilde{X}) \sim p_{X,\tilde{X}}^\theta} \left[\log q_{X|\tilde{X}}^\phi(X|\tilde{X}) \right] + \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \left[D \left(p_{X|\tilde{X}}(\cdot|\tilde{X}) \| q_{X|\tilde{X}}^\phi(\cdot|\tilde{X}) \right) \right], \quad (5.7) \end{aligned}$$

where $D(\cdot, \cdot)$ denotes the Kullback–Leibler divergence. Because of the non-negativity of the Kullback–Leibler divergence, we have

$$\mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \left[D \left(p_{X|\tilde{X}}(\cdot|\tilde{X}) \| q_{X|\tilde{X}}^\phi(\cdot|\tilde{X}) \right) \right] \geq 0. \quad (5.8)$$

Therefore, $\mathbb{E}_{(X,\tilde{X}) \sim p_{X,\tilde{X}}^\theta} \left[\log q_{X|\tilde{X}}^\phi(X|\tilde{X}) \right]$ forms a lower bound for $-H(X|\tilde{X})$. The equality holds if and only if $q_{X|\tilde{X}}^\phi(X|\tilde{X}) = p_{X|\tilde{X}}(X|\tilde{X})$ almost everywhere. We maximize this lower bound over the variational parameter ϕ to approximate $-H(X|\tilde{X})$. Recall that $\tilde{X}$ is determined by the sanitization parameter θ . Let the utility function be

$$\mathcal{L}(\theta) = \max_{\phi} \mathcal{L}(\phi; \theta), \quad (5.9)$$

where

$$\mathcal{L}(\phi; \theta) = \mathbb{E}_{(X,\tilde{X}) \sim p_{X,\tilde{X}}^\theta} \left[\log q_{X|\tilde{X}}^\phi(X|\tilde{X}) \right]. \quad (5.10)$$

The *reparameterization* in (5.2) is useful for our formulation because it can be used to rewrite the expectation w.r.t. $p_{X,\tilde{X}}^\theta$ as an expectation w.r.t. p_X, p_Ξ . Then,

$\mathcal{L}(\phi; \theta)$ can be rewritten as

$$\mathcal{L}(\phi; \theta) = \mathbb{E}_{X \sim p_X, \Xi \sim p_\Xi} \left[\log q_{X|\tilde{X}}^\phi(X|\tilde{X}) \right], \quad (5.11)$$

where $\tilde{X} = f^\theta(X, \Xi)$. We can use Monte Carlo methods to estimate (5.11) by drawing samples from p_X and p_Ξ independently. Samples of p_X are given as training data and p_Ξ is often chosen to be a common probability distribution such as the Gaussian or uniform distribution. The asymptotic convergence analysis for the empirical estimate of the variational utility function (5.9) based on samples of $(X, \tilde{X})$ is analogous to that for the estimate of the variational privacy metric presented in Section 5.4.3. It is thus omitted in this chapter.

5.3.2 Domain Regularization

The main idea is to penalize high divergence between the marginal distributions of X and $\tilde{X}$ and hence encourage $p_{\tilde{X}}$ to be similar to p_X . Also note that $\tilde{X}$ produced by maximizing (5.9) subject to a privacy constraint can be very different from X because mutual information is insensitive to the symbols used to represent the random variables. In many applications, we desire the sanitized data $\tilde{X}$ to be similar to X . For example, when sanitizing face images to remove private information such as gender and ethnicity, we expect the sanitized data to still be face images. In this chapter, we propose two domain regularization techniques: (1) bounding the maximum mean discrepancy [132], which is a metric on the space of probability measures, and (2) domain adaptation [133, 134] based on deep learning, which learns the mapping between a source space and a target space using a training set. The former enjoys theoretical soundness but can be compute-intensive while the latter is efficacious in processing complex data.

5.3.2.1 Maximum mean discrepancy (MMD)

The MMD that measures the disparity between distributions [135] is defined as:

$$\text{MMD}(\mathcal{F}, p_X, p_{\tilde{X}}) = \sup_{f \in \mathcal{F}} \left\{ \mathbb{E}_{X \sim p_X} f(X) - \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} f(\tilde{X}) \right\}, \quad (5.12)$$

where $\mathcal{F}$ is a class of witness functions $f : \mathcal{X} \mapsto \mathbb{R}$ that determines the probability metric. In particular, the paper [136] sets the witness function class $\mathcal{F}$ to be the unit ball in a reproducing kernel Hilbert space (RKHS). Let $\mathcal{H}$ be a RKHS with kernel $K(\cdot, \cdot)$ and $\Phi(\mathbf{x}) = K(\cdot, \mathbf{x})$ be the feature mapping that maps $\mathbf{x} \in \mathcal{X}$ to elements in $\mathcal{H}$. Denote $\|\cdot\|_{\mathcal{H}}$ as the norm induced by the kernel inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. A RKHS has the reproducing property $f(\mathbf{x}) = \langle \Phi(\mathbf{x}), f \rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$ [137]. This leads to an analytical expression of squared MMD under RKHSs [138]:

$$\begin{aligned} & \text{MMD}^2(\mathcal{H}, p_X, p_{\tilde{X}}) \\ &= \sup_{\|f\|_{\mathcal{H}} \leq 1} \left\{ \mathbb{E}_{X \sim p_X} f(X) - \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} f(\tilde{X}) \right\}^2 \\ &= \sup_{\|f\|_{\mathcal{H}} \leq 1} \left\{ \langle \mathbb{E}_{X \sim p_X} \Phi(X) - \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \Phi(\tilde{X}), f \rangle_{\mathcal{H}} \right\}^2 \\ &= \|\mathbb{E}_{X \sim p_X} \Phi(X) - \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \Phi(\tilde{X})\|_{\mathcal{H}}^2 \\ &= \mathbb{E}_{X, X' \sim p_X} [K(X, X')] - 2\mathbb{E}_{X \sim p_X, \tilde{X} \sim p_{\tilde{X}}} [K(X, \tilde{X})] + \mathbb{E}_{\tilde{X}, \tilde{X}' \sim p_{\tilde{X}}} [K(\tilde{X}, \tilde{X}')], \end{aligned} \quad (5.13)$$

where we have assumed that $\mathbb{E}_{X \sim p_X} \Phi(X)$ and $\mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \Phi(\tilde{X})$ exist. This embedding encodes a probability measure as a mean element in a RKHS and minimizing the MMD can be viewed as matching all the moments of p_X and $p_{\tilde{X}}$. For a characteristic (universal) kernel such as the Gaussian kernel, MMD is zero if and only if $p_X = p_{\tilde{X}}$ [138]. By using two sets of training data $\{\mathbf{x}_i\}_{i=1}^n$ and $\{\tilde{\mathbf{x}}_i\}_{i=1}^n$ sampled from the distributions p_X and $p_{\tilde{X}}$, respectively, the expression (5.13) yields an unbiased

estimator of the squared MMD:

$$\begin{aligned} \widehat{\text{MMD}}_K^2(p_X, p_{\tilde{X}}) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K(\mathbf{x}_i, \mathbf{x}_j) \\ &\quad - \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n K(\tilde{\mathbf{x}}_i, \mathbf{x}_j) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n K(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j). \end{aligned} \quad (5.14)$$

The value of (5.14) decreases as the distribution divergence between p_X and $p_{\tilde{X}}$ becomes smaller. Therefore, we may write the regularization function as

$$\mathcal{R}(\boldsymbol{\theta}) = \widehat{\text{MMD}}_K^2(p_X, p_{\tilde{X}}), \quad (5.15)$$

which is differentiable [w.r.t.](#) the sanitization parameter $\boldsymbol{\theta}$ if the kernel K is differentiable. The paper [139] considers a composition characteristic kernel $K \circ f = K(f(\cdot), f(\cdot))$, where $K(\cdot, \cdot)$ is a pre-specified kernel and f is an injective function. To add robustness to the estimator, it optimizes over a set of possible characteristic kernels: $\max_f \widehat{\text{MMD}}_{K \circ f}^2(p_X, p_{\tilde{X}})$. A similar approach can be found in [140]. As the amount of data required to produce a reliable MMD estimator grows with the dimensionality of the data, using f to encode data into lower spaces can be beneficial for high-dimensional data that exist on a low-dimensional manifold such as visual data.

5.3.2.2 Domain adaptation

Let random variables $A \in \mathcal{A}$ and $B \in \mathcal{B}$ be associated with probability distributions p_A and p_B , respectively. Domain adaptation aims to learn a function G that transforms a data in the source domain $\mathcal{A}$ to the target domain $\mathcal{B}$ such that the distribution of $G(A)$ is indistinguishable from p_B . The objective function can be written as

$$\min_G \max_{D_B} \mathbb{E}_{B \sim p_B} [\log(D_B(B))] + \mathbb{E}_{A \sim p_A} [\log(1 - D_B(G(A)))] , \quad (5.16)$$

where $D_B : \mathcal{B} \cup G(\mathcal{A}) \mapsto [0, 1]$ is an adversarial discriminator seeking to differentiate the transformed data $G(A)$ from B while the function G aims to make the distribution of $G(A)$ resemble that of B . The discriminator and transformation function are trained against each other until an equilibrium is reached. Similarly, we introduce a loss function for domain regularization:

$$\mathcal{R}(D_X; \boldsymbol{\theta}) = \mathbb{E}_{X \sim p_X} [\log(D_X(X))] + \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \left[\log(1 - D_X(\tilde{X})) \right], \quad (5.17)$$

where $D_X : \mathcal{X} \cup \tilde{\mathcal{X}} \mapsto [0, 1]$ is a domain discriminator yielding the likelihood that an input is sampled from p_X . We maximize (5.17) over D_X to obtain the regularization function:

$$\mathcal{R}(\boldsymbol{\theta}) = \max_{D_X} \mathcal{R}(D_X; \boldsymbol{\theta}), \quad (5.18)$$

which we then minimize over $\boldsymbol{\theta}$ in an effort to force $p_{\tilde{X}}$ to be close to p_X .

5.4 Privacy Metrics

In this section, we formulate the privacy function $\mathcal{P}(\boldsymbol{\theta})$ from an information or estimation-theoretic perspective using the concept of average information leakage (mutual information) and maximal correlation. We use the variational method to approximate mutual information and develop methods for estimating maximal correlation. We discuss the advantages and disadvantages of each approach. We refer to the DRIP framework using mutual information or maximal correlation for the privacy metric as DRIP-Var and DRIP-Max, respectively.

5.4.1 Variational Method: DRIP-Var

To prevent an adversarial service provider from obtaining detailed information about the private variable S after it acquires the sanitized data $\tilde{X}$ from the client,

we adopt the notion of average information leakage (mutual information) [43] to form the privacy function. The mutual information $I(\tilde{X}; S)$ quantifies the amount of information about S revealed by observing $\tilde{X}$, and it can be expressed as

$$\mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \left[D \left(p_{S|\tilde{X}}(\cdot|\tilde{X}), p_S(\cdot) \right) \right] = H(S) - H(S|\tilde{X}), \quad (5.19)$$

which attains the minimum when S and $\tilde{X}$ are independent. Because (5.19) can not be computed analytically without knowledge of the joint distribution $p_{S,\tilde{X}}$, we use the variational method [130] to obtain an approximation. Let a variational distribution function $q_{S|\tilde{X}}^{\boldsymbol{\tau}}(\cdot|\cdot)$ parametrized by $\boldsymbol{\tau}$ act as a surrogate for the true distribution $p_{S|\tilde{X}}$. We have

$$\begin{aligned} & -H(S|\tilde{X}) \\ &= \mathbb{E}_{(S,\tilde{X}) \sim p_{S,\tilde{X}}} \left[\log p_{S|\tilde{X}}(S|\tilde{X}) \right] \\ &= \mathbb{E}_{(S,\tilde{X}) \sim p_{S,\tilde{X}}} \left[\log q_{S|\tilde{X}}^{\boldsymbol{\tau}}(S|\tilde{X}) \right] + \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \left[D(p_{S|\tilde{X}}(\cdot|\tilde{X}) \| q_{S|\tilde{X}}^{\boldsymbol{\tau}}(\cdot|\tilde{X})) \right] \\ &\geq \mathbb{E}_{(S,\tilde{X}) \sim p_{S,\tilde{X}}} \left[\log q_{S|\tilde{X}}^{\boldsymbol{\tau}}(S|\tilde{X}) \right], \end{aligned} \quad (5.20)$$

where equality holds if and only if $q_{S|\tilde{X}}^{\boldsymbol{\tau}} = p_{S|\tilde{X}}$ almost everywhere. Since $\tilde{X} - X - S$ forms a Markov chain, we have

$$p_{S,X,\tilde{X}}(\mathbf{s}, \mathbf{x}, \tilde{\mathbf{x}}) = p_{S|X}(\mathbf{s}|\mathbf{x})p_{X,\tilde{X}}(\mathbf{x}, \tilde{\mathbf{x}}).$$

Therefore, we have

$$\begin{aligned} & \mathbb{E}_{(S,\tilde{X}) \sim p_{S,\tilde{X}}} \left[\log q_{S|\tilde{X}}^{\boldsymbol{\tau}}(S|\tilde{X}) \right] \\ &= \mathbb{E}_{(S,\tilde{X}) \sim p_{S,\tilde{X}}} \left[\mathbb{E}_{X \sim p_{X|S,\tilde{X}}} \left[\log q_{S|\tilde{X}}^{\boldsymbol{\tau}}(S|\tilde{X}) \right] \right] \\ &= \mathbb{E}_{(S,X,\tilde{X}) \sim p_{S,X,\tilde{X}}} \left[\log q_{S|\tilde{X}}^{\boldsymbol{\tau}}(S|\tilde{X}) \right] \\ &= -\mathbb{E}_{(X,\tilde{X}) \sim p_{X,\tilde{X}}^{\boldsymbol{\theta}}} \left[H \left(p_{S|X}(\cdot|X), q_{S|\tilde{X}}^{\boldsymbol{\tau}}(\cdot|\tilde{X}) \right) \right], \end{aligned} \quad (5.21)$$

where $H(p, q) = -\mathbb{E}_{Z \sim p} [\log q(Z)]$ is the cross-entropy function. From (5.19) and (5.20), we obtain

$$I(S; \tilde{X}) \geq H(S) + \mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta}), \quad (5.22)$$

where

$$\mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta}) = -\mathbb{E}_{X \sim p_X, \Xi \sim p_\Xi} \left[H \left(p_{S|X}(\cdot | X), q_{S|\tilde{X}}^\tau(\cdot | \tilde{X}) \right) \right] \quad (5.23)$$

is (5.21) after applying the reparametrization trick and $\tilde{X} = f^\theta(X, \Xi)$. Using the lower bound (5.22) and ignoring the constant $H(S)$, we obtain an approximated privacy function as

$$\mathcal{P}(\boldsymbol{\theta}) = \max_{\boldsymbol{\tau}} \mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta}). \quad (5.24)$$

The approximation error for the average information leakage using the lower-bound in (5.22) is given by

$$H(S|\tilde{X}) - \mathcal{P}(\boldsymbol{\theta}) = \min_{\boldsymbol{\tau}} \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \left[D(q_{S|\tilde{X}}^\tau(\cdot | \tilde{X}) || p_{S|\tilde{X}}(\cdot | \tilde{X})) \right] \geq 0. \quad (5.25)$$

From (5.25), it becomes obvious that only if the proposed variational distribution $q_{S|\tilde{X}}^\tau$ is close to the true distribution $p_{S|\tilde{X}}$, can we maintain a small margin of error for the privacy leakage. Otherwise, the variational method introduces a large uncertainty for the amount of privacy leakage. While there are several choices for the variational probability mass function for categorical private variables (e.g., softmax activation for deep-learning classification), proposing an effective variational pdf for continuous private variables is difficult in practice due to a lack of prior knowledge and domain expertise associated with the raw data and private data.

Practically, by rewriting $\mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta})$ as a sample mean, the privacy function becomes

$$\mathcal{P}(\boldsymbol{\theta}) = \max_{\boldsymbol{\tau}} - \sum_{i=1}^n H \left(p_{S|X}(\cdot | \mathbf{x}_i), q_{S|\tilde{X}}^{\boldsymbol{\tau}}(\cdot | \tilde{\mathbf{x}}_i) \right), \quad (5.26)$$

where $\{\mathbf{x}_i, \tilde{\mathbf{x}}_i\}_{i=1}^n$ are samples of $(X, \tilde{X})$. The calculation of (5.26) presumes that soft or hard labels w.r.t. the private variable S that are assigned to the training raw data $\{\mathbf{x}_i\}_{i=1}^n$, i.e., $p_{S|X}(\cdot | \mathbf{x}_i)$ for $i = 1, \dots, n$ are known.

The use of the variational method is however not applicable to all cases and is subject to the following limitations:

- The probability distribution of the private variable S conditioned on each realization of the raw variable X needs to be known.
- There exists a family of viable variational distributions that can be represented as parametrized functions.
- The practical computation of (5.26) requires the private variable to have finite support.

These constraints prohibit a wider application of the variational method in representing the privacy metric. For example, in applications, users may have access only to data samples of (S, X) , instead of the knowledge of the probability distribution of S conditioned on each sample of X . Furthermore, computing the privacy metric using the variational method is intractable for continuous private variables due to the explicit dependence on data distribution. To provide a complement to the variational privacy metric, we propose the use of *maximal correlation* as a privacy metric in Section 5.4.2, whose practical implementation is not contingent on the above-mentioned constraints. It is worth noting that, under some technical conditions, [57] showed that mutual information bounds maximal correlation. Therefore, mutual information is a stronger privacy metric in such cases and hence preferred over maximal correlation if it can be implemented efficiently using the

variational method presented in this section. In cases where this is not possible, we propose the use of maximal correlation as the privacy metric.

5.4.2 Maximal Correlation: DRIP-Max

Maximal correlation was introduced by Hirschfeld [141] and Gebelein [142], and later studied by Rényi [143]. It measures the statistical dependency between two random variables by using two appropriate functions that map these random variables to real-valued spaces. Given two jointly distributed random variables $Y \in \mathcal{Y}$ and $Z \in \mathcal{Z}$, the maximal correlation of Y and Z is defined as follows.

Definition 10 (The Hirschfeld-Gebelin-Renyi Maximal Correlation). For two random variables $Y \in \mathcal{Y}$ and $Z \in \mathcal{Z}$, the maximal correlation between them is given by

$$\rho(Y, Z) = \sup_{f, g} \mathbb{E} [f(Y)g(Z)], \quad (5.27)$$

where the functions $f : \mathcal{Y} \mapsto \mathbb{R}$ and $g : \mathcal{Z} \mapsto \mathbb{R}$ are such that

$$\begin{aligned} \mathbb{E} [f(Y)] &= 0, \quad \mathbb{E} [g(Z)] = 0, \\ \mathbb{E} [f^2(Y)] &= 1, \quad \mathbb{E} [g^2(Z)] = 1. \end{aligned}$$

If the supremum in (5.27) is attainable, the optimal $f(\cdot)$ and $g(\cdot)$ in (5.27) are called the maximal correlation functions of Y and Z , respectively.

Some of the well-known properties of maximal correlation are [143]: (i) $0 \leq \rho(Y, Z) \leq 1$; (ii) $\rho(Y, Z) = 0$ if and only if Y and Z are independent; (iii) $\rho(Y, Z) = 1$ if and only there exists functions $f(Y) = g(Z)$ almost everywhere; and (iv) $\rho(Y, Z)$ is equal to the second largest singular value of the divergence transition matrix Q with $Q(y, z) = \frac{p_{Y,Z}(y,z)}{\sqrt{p_Y(y)p_Z(z)}}$ when $|\mathcal{Y}|$ and $|\mathcal{Z}| < \infty$, indicating that maximal correlation is unrelated to the symbols used for representing the random variables [144].

Remark 1. Given that the maximal correlation of two random variables is zero if and only if these two random variables are independent, the use of maximal correlation as a privacy function to be minimized is well motivated. However, using maximal correlation as the utility function to be maximized is questionable. For example, if the sanitized data $\tilde{X}$ is a compressed version of the raw data X , the maximal correlation of $\tilde{X}$ and X is always equal to one (the maximum attainable). Meanwhile, it can be the case that $\tilde{X}$ leads to little or no utility.

For two random variables $Y \in \mathcal{Y}$ and $Z \in \mathcal{Z}$ with pdfs p_Y and p_Z respectively, consider the following Hilbert spaces:

$$\mathcal{Y} = \{f : \mathcal{Y} \mapsto \mathbb{R} \mid \mathbb{E}[f(Y)^2] < \infty\}, \quad (5.28)$$

$$\mathcal{Z} = \{g : \mathcal{Z} \mapsto \mathbb{R} \mid \mathbb{E}[g(Z)^2] < \infty\}, \quad (5.29)$$

where the inner product is defined as $\langle f_1, f_2 \rangle_{\mathcal{Y}} = \mathbb{E}[f_1(Y)f_2(Y)]$ for $f_1, f_2 \in \mathcal{Y}$ and $\langle g_1, g_2 \rangle_{\mathcal{Z}} = \mathbb{E}[g_1(Z)g_2(Z)]$ for $g_1, g_2 \in \mathcal{Z}$. The respective induced norms are denoted as $\|\cdot\|_{\mathcal{Y}}$ and $\|\cdot\|_{\mathcal{Z}}$.

Define a linear operator $T : \mathcal{Y} \mapsto \mathcal{Z}$ by the conditional expectation $Tf(Z) = \mathbb{E}[f(Y) \mid Z]$ for each $f \in \mathcal{Y}$. Assuming that

$$\int |k(y, z)|^2 p_Y(y) p_Z(z) dy dz < \infty, \quad (5.30)$$

where $k(y, z) = \frac{p_{Y,Z}(y,z)}{p_Y(y)p_Z(z)}$, we conclude from [111, Proposition 4.7] that T is compact. From [143], we have

$$\rho(Y, Z) = \max_{\substack{f \in \mathcal{Y} \\ \mathbb{E}[f(Y)] = 0}} \frac{\|Tf\|_{\mathcal{Z}}}{\|f\|_{\mathcal{Y}}}, \quad (5.31)$$

where we have used max instead of sup on the right-hand side to indicate that the maximal correlation is attainable. From Jensen's inequality, we have

$$\|Tf\|_{\mathcal{Z}}^2 = \mathbb{E}[\mathbb{E}[f(Y) \mid Z]^2] \leq \mathbb{E}[\mathbb{E}[f^2(Y) \mid Z]] = \|f\|_{\mathcal{Y}}^2, \quad (5.32)$$

hence the largest singular value of T , $\|T\| = 1$ is attained by $f(y) = 1$ for all $y \in \mathcal{Y}$. Since $f \in \mathcal{Y}$ such that $\mathbb{E}[f(Y)] = 0$ is orthogonal to the span of constant functions in $\mathcal{Y}$, we have shown that $\rho(Y, Z)$ is the second largest singular value of T .

In Proposition 4, we show that maximal correlation given by Definition 10 can be computed by solving an unconstrained optimization problem.

Proposition 4. Assuming (5.30) and the L^2 spaces $\mathcal{Y}$ and $\mathcal{Z}$ defined in (5.28) and (5.29), we have

$$\rho^2(Y, Z) = \max_{f \in \mathcal{Y}, g \in \mathcal{Z}} \frac{2\mathbb{E}[f(Y)g(Z)] - \mathbb{E}[g^2(Z)] - \mathbb{E}[f^2(Y)]}{\mathbb{E}[f^2(Y)]}. \quad (5.33)$$

The optimal f for (5.33) is the maximal correlation function of Y , and the optimal g is given by $g(Z) = \mathbb{E}[f(Y) | Z]$.

Proof. See Appendix 5.A. □

In practice, it is infeasible to compute maximal correlation by searching over the whole L^2 function space. To provide a computationally tractable implementation, we restrict maximal correlation functions to be from a rich class represented by neural networks and derive an estimator based on (5.33). Alternatively, we may also work with the kernel method by letting the search space of the maximal correlation functions be an RKHS. For the interested reader, we provide details on the kernel method for computing maximal correlation as a privacy metric in Appendix 5.C. Both approaches allow us to update the sanitizer via backpropagation. In what follows, we present the neural network method of estimating maximal correlation for a privacy metric.

Neural network implementation

Given a set of data samples $\{(\mathbf{s}_i, \tilde{\mathbf{x}}_i)\}_{i=1}^n$ that are realizations of $(S, \tilde{X})$, we search over the finite variance functions of $\tilde{X}$ and S represented by neural networks, to approximate the maximal correlation between S and $\tilde{X}$. This configuration is

distinguished from the setup of the variational privacy metric approach in Section 5.4.1, for which $p_{S|X}(\cdot|\mathbf{x})$ is required to be known for each sample $\mathbf{x}$.

Let bounded real-valued functions $f_{\tau_1}(\cdot)$ and $g_{\tau_2}(\cdot)$ be formed by neural networks, with $\tau_1 \in \Omega_1$ and $\tau_2 \in \Omega_2$ denoting their respective trainable parameters. First, we form the following empirical estimates:

$$\begin{aligned}\mathbb{E}[f_{\tau_1}(\tilde{X})] &\approx \mu_n(f_{\tau_1}) := \frac{1}{n} \sum_{i=1}^n f_{\tau_1}(\tilde{\mathbf{x}}_i), \\ \mathbb{E}[f_{\tau_1}^2(\tilde{X})] &\approx \mu_n(f_{\tau_1}^2) := \frac{1}{n} \sum_{i=1}^n f_{\tau_1}^2(\tilde{\mathbf{x}}_i), \\ \mathbb{E}[g_{\tau_2}^2(S)] &\approx \mu_n(g_{\tau_2}^2) := \frac{1}{n} \sum_{i=1}^n g_{\tau_2}^2(\mathbf{s}_i), \\ \mathbb{E}[f_{\tau_1}(\tilde{X})g_{\tau_2}(S)] &\approx \mu_n(f_{\tau_1} \otimes g_{\tau_2}) := \frac{1}{n} \sum_{i=1}^n f_{\tau_1}(\tilde{\mathbf{x}}_i)g_{\tau_2}(\mathbf{s}_i).\end{aligned}\quad (5.34)$$

Substituting the above empirical estimates into (5.33) with $Y = \tilde{X}$ and $Z = S$, and letting $\tau = (\tau_1, \tau_2)$, we perform the maximization in (5.33) over $f \in \{f_{\tau_1} : \tau_1 \in \Omega_1\}$ and $g \in \{g_{\tau_2} : \tau_2 \in \Omega_2\}$ to estimate the maximal correlation as a privacy metric:

$$\mathcal{P}(\theta) = \max_{\tau} \mathcal{P}(\tau; \theta), \quad (5.35)$$

where

$$\mathcal{P}(\tau; \theta) = \frac{2\mu_n(f_{\tau_1} \otimes g_{\tau_2}) - \mu_n(g_{\tau_2}^2) - (\mu_n(f_{\tau_1}))^2}{\mu_n(f_{\tau_1}^2)}. \quad (5.36)$$

This formulation is consistent with the variational privacy metric in (5.24). As illustrated in Fig. 5.3, we can build a machine learning model and apply gradient descent to optimize (5.35) over the joint parameter τ .

In particular, if S is a categorical private variable, i.e., $|\mathcal{S}| < \infty$, we need not propose a search space of functions of S such as g_{τ_2} to compute maximal correlation.

FIG. 5.3: Neural network based maximal correlation estimator, where $\mu(\cdot)$ denotes the batch mean operator.

Instead, we can invoke a nonparametric estimation in terms of the optimal function g in (5.33) based on the aforementioned property that $g^*(S) = \mathbb{E} [f^*(\tilde{X}) \mid S]$ if f^* and g^* optimize the objective function of (5.33) with $Y = \tilde{X}$ and $Z = S$. Let δ be the Dirac delta function. For a fixed function f_{τ_1} , the function $g_{\tau_2}(\mathbf{s}) = \frac{1}{n} \sum_{i=1}^n f_{\tau_1}(\tilde{\mathbf{x}}_i) \delta(\mathbf{s}_i - \mathbf{s})$ which is the sample average estimate of $\mathbb{E} [f_{\tau_1}(\tilde{X}) \mid S = \mathbf{s}]$ maximizes the objective function (5.36). Instead of optimizing the objective function of (5.35) w.r.t. the joint parameters of f_{τ_1} and g_{τ_2} , we optimize over f_{τ_1} and g_{τ_2} in an iterative procedure. This yields an alternating algorithm for solving (5.35) summarized as follows:

- a) Fix g . Update f_{τ_1} by applying gradient descent to optimize

$$\max_{\tau_1} \frac{2\mu_n(f_{\tau_1} \otimes g) - (\mu_n(f_{\tau_1}))^2}{\mu_n(f_{\tau_1}^2)}. \quad (5.37)$$

- b) Fix f_{τ_1} . For each $\mathbf{s} \in \mathcal{S}$, update g according to

$$g(\mathbf{s}) = \frac{1}{M} \sum_{i=1}^M f_{\tau_1}(\tilde{\mathbf{x}}_i) \delta(\mathbf{s}_i - \mathbf{s}). \quad (5.38)$$

- c) Iterate the above two steps until a convergence criterion is reached.

In this way, we remove the approximation error brought about by using the neural network version of the function g_{τ_2} .

5.4.3 Convergence Analysis

In this section, we present asymptotic convergence analysis for the empirical estimates of the privacy metrics outlined in Sections 5.4.1 and 5.4.2. We show that under mild conditions, the empirical estimates (5.26) and (5.35) converge to their corresponding theoretical privacy metrics almost surely ($\xrightarrow{\text{a.s.}}$) or in probability ($\xrightarrow{\text{P}}$) as the number of training samples $n \rightarrow \infty$. As mentioned earlier, the same reasoning applies to the convergence analysis of the estimate of the utility metric (5.9), which is thus omitted in this chapter.

5.4.4 Average Information Leakage

First, we present the conditions that allow the sample average estimator (5.26) to converge to the true average information leakage as the sample size increases. To start with, we need the following notations:

$$\begin{aligned}\lambda(\boldsymbol{\tau}; \mathbf{x}, \tilde{\mathbf{x}}) &= -H\left(p_{S|X}(\cdot|\mathbf{x}), q_{S|\tilde{X}}^{\boldsymbol{\tau}}(\cdot|\tilde{\mathbf{x}})\right), \\ \lambda_n(\boldsymbol{\tau}) &= \frac{1}{n} \sum_{i=1}^n \lambda(\boldsymbol{\tau}; \mathbf{x}_i, \tilde{\mathbf{x}}_i), \\ \bar{\lambda}(\boldsymbol{\tau}) &= \mathbb{E}_{X \sim p_X, \tilde{X} \sim p_{\tilde{X}}} \left[\lambda(\boldsymbol{\tau}; X, \tilde{X}) \right],\end{aligned}\tag{5.39}$$

where $\{\mathbf{x}_i, \tilde{\mathbf{x}}_i\}_{i=1}^n$ are n independent and identically distributed (i.i.d.) samples of the pair of random variables $(X, \tilde{X})$. Note that $\lambda_n(\boldsymbol{\tau})$ is the sample average unbiased estimate of $\bar{\lambda}(\boldsymbol{\tau})$ and $\mathbb{E}[\lambda_n(\boldsymbol{\tau})] = \bar{\lambda}(\boldsymbol{\tau})$. Recall from (5.26) that $\max_{\boldsymbol{\tau}} \lambda_n(\boldsymbol{\tau})$ is our empirical estimator for the theoretical privacy metric $-H(S|\tilde{X})$. From the derivation of (5.21), it can be verified that the theoretical privacy metric can be

written as

$$-H(S|\tilde{X}) = -\mathbb{E}_{X \sim p_X, \Xi \sim p_\Xi} \left[H \left(p_{S|X}(\cdot|X), p_{S|\tilde{X}}(\cdot|\tilde{X}) \right) \right]. \quad (5.40)$$

Thereafter, we show $\max_{\boldsymbol{\tau}} \lambda_n(\boldsymbol{\tau}) \xrightarrow{\text{a.s.}} -H(S|\tilde{X})$ as $n \rightarrow \infty$ if the following conditions are satisfied:

(V-i) $\boldsymbol{\tau} \in \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is a compact set.

(V-ii) $\lambda(\boldsymbol{\tau}; \mathbf{x}, \tilde{\mathbf{x}})$ is a continuous function **w.r.t.** parameter $\boldsymbol{\tau}$ for almost all $\mathbf{x} \in \mathcal{X}$ and $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$.

(V-iii) There exists a dominating function $g(X, \tilde{X})$ such that $\mathbb{E}_{X \sim p_X, \Xi \sim p_\Xi} [g(X, \tilde{X})] < \infty$ and for all $\boldsymbol{\tau} \in \boldsymbol{\Omega}$, $\lambda(\boldsymbol{\tau}; X, \tilde{X}) \leq g(X, \tilde{X})$.

(V-iv) There exists $\boldsymbol{\tau}^* \in \boldsymbol{\Omega}$ such that $q_{S|\tilde{X}}^{\boldsymbol{\tau}^*} = p_{S|\tilde{X}}$.

From the uniform law of large numbers [145], the conditions (V-i)-(V-iii) guarantee

$$\max_{\boldsymbol{\tau}} |\lambda_n(\boldsymbol{\tau}) - \bar{\lambda}(\boldsymbol{\tau})| \xrightarrow{\text{a.s.}} 0. \quad (5.41)$$

As alluded to by (5.20) and (5.21), condition (V-iv) allows us to write $H(S|\tilde{X}) = -\max_{\boldsymbol{\tau} \in \boldsymbol{\Omega}} \bar{\lambda}(\boldsymbol{\tau})$. Subsequently, we have

$$\begin{aligned} \left| \max_{\boldsymbol{\tau} \in \boldsymbol{\Omega}} \lambda_n(\boldsymbol{\tau}) + H(S|\tilde{X}) \right| &= \left| \max_{\boldsymbol{\tau} \in \boldsymbol{\Omega}} \lambda_n(\boldsymbol{\tau}) - \max_{\boldsymbol{\tau} \in \boldsymbol{\Omega}} \bar{\lambda}(\boldsymbol{\tau}) \right| \\ &\leq \max_{\boldsymbol{\tau} \in \boldsymbol{\Omega}} |\lambda_n(\boldsymbol{\tau}) - \bar{\lambda}(\boldsymbol{\tau})|. \end{aligned} \quad (5.42)$$

Finally, combining (5.41) and (5.42), we have

$$\left| \max_{\boldsymbol{\tau} \in \boldsymbol{\Omega}} \lambda_n(\boldsymbol{\tau}) + H(S|\tilde{X}) \right| \xrightarrow{\text{a.s.}} 0.$$

Therefore, we can conclude that the empirical estimator $\max_{\boldsymbol{\tau} \in \boldsymbol{\Omega}} \lambda_n(\boldsymbol{\tau})$ converges to the theoretical privacy metric $-H(S|\tilde{X})$ almost surely if the conditions (V-i)-(V-iv) are satisfied.

In a neural network representation of $q_{S|\tilde{X}}^\tau(\cdot|\cdot)$, only trivial assumptions are required for conditions (V-i)-(V-iv) to hold. Since τ is a collection of trainable weights that has finite elements, it is an element of the Euclidean space. Thus, the compactness of Ω in (V-i) can be assured by restricting each entry of τ to be from a closed interval. In practice, the range of the parameters in neural networks is limited by a specified data type. Since the layer functions of neural networks are always differentiable w.r.t. their respective trainable parameters, $q_{S|\tilde{X}}^\tau(\cdot|\cdot)$, which is the composition of the layer functions, is continuous. Thus, condition (V-ii) is met. From the cross-entropy minimization:

$$\begin{aligned}\lambda(\tau; \mathbf{x}, \tilde{\mathbf{x}}) &= -H\left(p_{S|X}(\cdot|\mathbf{x}), q_{S|\tilde{X}}^\tau(\cdot|\tilde{\mathbf{x}})\right) \\ &\leq -H\left(p_{S|X}(\cdot|\mathbf{x}), p_{S|X}(\cdot|\mathbf{x})\right),\end{aligned}\tag{5.43}$$

and the inequality:

$$\begin{aligned}\mathbb{E}_{X \sim p_X} \left[-H\left(p_{S|X}(\cdot|\mathbf{x}), p_{S|X}(\cdot|\mathbf{x})\right) \right] \\ = H(S|X) \leq H(S) \leq \log|\mathcal{S}| < \infty,\end{aligned}\tag{5.44}$$

it is obvious that condition (V-iii) is guaranteed. Finally, condition (V-iv) is implied by the universal approximation property of neural networks [146] for τ in a sufficiently high dimensional convex set Ω .

5.4.5 Maximal Correlation

First, we discuss the conditions for the convergence of the estimate of maximal correlation (5.35). Here the uniform law of large numbers does not apply because (5.35) is not in the direct form of a sample average. The analysis entails the following notations:

$$\lambda_n(\tau) = \frac{2\mu_n(f_{\tau_1} \otimes g_{\tau_2}) - \mu_n(g_{\tau_2}^2) - (\mu_n(f_{\tau_1}))^2}{\mu_n(f_{\tau_1}^2)},$$

$$\bar{\lambda}(\boldsymbol{\tau}) = \frac{2\mathbb{E} \left[f_{\tau_1}(\tilde{X}) g_{\tau_2}(S) \right] - \mathbb{E} \left[g_{\tau_2}^2(S) \right] - \mathbb{E} \left[f_{\tau_1}(\tilde{X}) \right]^2}{\mathbb{E} \left[f_{\tau_1}^2(\tilde{X}) \right]}, \quad (5.45)$$

where $\{\mathbf{s}_i, \tilde{\mathbf{x}}_i\}_{i=1}^n$ are n **i.i.d.** samples of the pair of random variables $(S, \tilde{X})$. Recall from (5.35) that $\max_{\boldsymbol{\tau}} \lambda_n(\boldsymbol{\tau})$ is our empirical estimate of maximal correlation $\rho(S, \tilde{X})$. We show $\max_{\boldsymbol{\tau}} \lambda_n(\boldsymbol{\tau}) \xrightarrow{P} \rho^2(S, \tilde{X})$ as $n \rightarrow \infty$ if the following conditions are met:

(M-i) $\boldsymbol{\tau} = (\tau_1, \tau_2) \in \boldsymbol{\Omega} = \boldsymbol{\Omega}_1 \times \boldsymbol{\Omega}_2$, where $\boldsymbol{\Omega}$ is a compact set.

(M-ii) $\lambda_n(\boldsymbol{\tau}) \xrightarrow{P} \bar{\lambda}(\boldsymbol{\tau})$.

(M-iii) The function $\lambda_n(\boldsymbol{\tau})$ is *stochastically equicontinuous* [147], i.e., for any $\eta > 0$ and $\epsilon > 0$, there exists $\delta > 0$ such that

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\|\boldsymbol{\tau} - \boldsymbol{\tau}'\| < \delta} |\lambda_n(\boldsymbol{\tau}) - \lambda_n(\boldsymbol{\tau}')| > \eta \right) < \epsilon.$$

(M-iv) There exists $\boldsymbol{\tau}^* \in \boldsymbol{\Omega}$ such that $\bar{\lambda}(\boldsymbol{\tau}^*) = \rho^2(S, \tilde{X})$.

First we note condition (M-ii) naturally holds since the numerator and denominator of $\lambda_n(\boldsymbol{\tau})$ divided by n converge to the numerator and denominator of $\bar{\lambda}(\boldsymbol{\tau})$ in probability, respectively. From the generic uniform convergence result of [147, Theorem 1], conditions (M-i)-(M-iii) ensure that

$$\max_{\boldsymbol{\tau}} |\lambda_n(\boldsymbol{\tau}) - \bar{\lambda}(\boldsymbol{\tau})| \xrightarrow{P} 0. \quad (5.46)$$

From condition (M-iv) and Proposition 4, we further have

$$\begin{aligned} \left| \max_{\boldsymbol{\tau}} \lambda_n(\boldsymbol{\tau}) - \rho^2(S, \tilde{X}) \right| &= \left| \max_{\boldsymbol{\tau}} \lambda_n(\boldsymbol{\tau}) - \max_{\boldsymbol{\tau}} \bar{\lambda}(\boldsymbol{\tau}) \right| \\ &\leq \max_{\boldsymbol{\tau}} |\lambda_n(\boldsymbol{\tau}) - \bar{\lambda}(\boldsymbol{\tau})|. \end{aligned} \quad (5.47)$$

Finally, combining (5.46) and (5.47), we have $\max_{\boldsymbol{\tau}} \lambda_n(\boldsymbol{\tau}) \xrightarrow{P} \rho^2(S, \tilde{X})$ as $n \rightarrow \infty$ if conditions (M-i), (M-iii) and (M-iv) are satisfied.

In what follows, we discuss some assumptions for satisfying conditions (M-i)-(M-iv) when the real-valued functions $f_{\tau_1}(\cdot)$ and $g_{\tau_2}(\cdot)$ are formed by neural networks. The same argument for the compactness of Ω in Section 5.4.4 applies to condition (M-i). In Proposition 5, we provide conditions that satisfy (M-iii).

Proposition 5. Suppose Ω , $\mathcal{S}$ and $\tilde{\mathcal{X}}$ are compact. If (1) neural network functions $f_{\tau_1}(\tilde{\mathbf{x}})$ and $g_{\tau_2}(\mathbf{s})$ are smooth w.r.t. $\tau_1 \in \Omega_1$ and $\tau_2 \in \Omega_2$, and continuous w.r.t. $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$ and $\mathbf{s} \in \mathcal{S}$, respectively; (2) $\min_{\tau_1 \in \Omega_1} \mathbb{E}[\mu_n(f_{\tau_1}^2)] > 0$, then there exists a sequence of positive random variables (B_n) and a constant $b < \infty$ such that $B_n \xrightarrow{\text{a.s.}} b$ as $n \rightarrow \infty$ and

$$|\lambda_n(\tau) - \lambda_n(\tau')| \leq B_n \|\tau - \tau'\|, \quad \forall \tau, \tau' \in \Omega.$$

Proof. See Appendix 5.B. □

Supposing the assumptions in Proposition 5 hold, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\|\tau - \tau'\| < \delta} |\lambda_n(\tau) - \lambda_n(\tau')| > \eta \right) &\leq \lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\|\tau - \tau'\| < \delta} B_n \|\tau - \tau'\| > \eta \right) \\ &\leq \lim_{n \rightarrow \infty} \mathbb{P}(B_n \delta > \eta). \end{aligned} \quad (5.48)$$

Choosing $\delta = \frac{\eta}{b+\epsilon}$ with $\epsilon > 0$ in (5.48), we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(B_n \delta > \eta) &= \lim_{n \rightarrow \infty} \mathbb{P}(B_n > b + \epsilon) \\ &\leq \lim_{n \rightarrow \infty} \mathbb{P}(|B_n - b| > \epsilon) = 0, \end{aligned} \quad (5.49)$$

where the last equality follows from $B_n \xrightarrow{\text{a.s.}} b$. Thus condition (M-iii) is met.

Finally, similar to the discussion in Section 5.4.4, condition (M-iv) is implied by the universal approximation property of neural networks [146] for τ in a sufficiently high dimensional convex set Ω .

In summary, to satisfy the convergence conditions (M-i)-(M-iv), we constrain the input space and parameter space of the neural networks f_{τ_1} and g_{τ_2} to be

compact, with the parameter space being convex as well. Furthermore, we choose appropriate activation functions (e.g., sigmoid) so that the layer functions of the neural networks are smooth *w.r.t.* their trainable parameters and continuous *w.r.t.* their inputs. The neural networks need to be sufficiently deep so that the dimension of the parameter space is high. Finally, $\min_{\tau_1 \in \Omega_1} \mathbb{E} [\mu_n(f_{\tau_1}^2)] > 0$ implies that we should prohibit the neural network $f_{\tau_1}(\cdot)$ from being a constant zero function.

5.5 DRIP Framework Optimization

In this section, we present the overall DRIP-Max and DRIP-Var frameworks using the utility, regularization and privacy functions described in Sections 5.3 and 5.4 and propose an alternating optimization algorithm.

5.5.1 Overall DRIP Framework

We describe the formulation of the DRIP-Var framework explicitly. The DRIP-Max can be obtained by simply replacing $\mathcal{P}(\tau; \theta)$ by one of the privacy functions given in Section 5.4.2. Using the domain adaptation of (5.18) and substituting (5.9) and (5.24) into (5.4), the overall objective of DRIP-Var can be expressed as an alternating optimization:

$$\max_{\theta, \phi} \mathcal{L}(\phi; \theta) - \alpha \mathcal{R}(D_X^*; \theta) - \lambda g(\mathcal{P}(\tau^*; \theta)) \quad (5.50a)$$

$$\text{s. t. } D_X^* = \arg \max_{D_X} \mathcal{R}(D_X; \theta), \quad (5.50b)$$

$$\tau^* = \arg \max_{\tau} \mathcal{P}(\tau; \theta). \quad (5.50c)$$

The regularization term $\mathcal{R}(D_X^*; \theta)$ should be replaced by (5.14) and (5.50b) should be removed if the MMD is used as the regularizer. Given a maximum allowable privacy leakage, we maximize utility with a penalty term forcing the sanitized data to be in a sensible domain. In order to optimize (5.50), the proposed

sanitizer $p_{\tilde{X}|X}^\theta$ and variational distributions $q_{X|\tilde{X}}^\phi$ and $q_{S|\tilde{X}}^\tau$ are required to take certain forms and constraints, which are summarized as follows:

- (i) The sanitizer $p_{\tilde{X}|X}^\theta$ in (5.1) is determined by $\tilde{X} = f^\theta(X, \Xi)$, where $\Xi \sim p_\Xi$, and $f^\theta(X, \Xi)$ is differentiable *w.r.t.* θ . $\tilde{X}$ has the same dimension as X .
- (ii) The function $q_{X|\tilde{X}}^\phi : \mathcal{X} \times \tilde{\mathcal{X}} \mapsto [0, 1]$ used in the utility function (5.9) is such that $\int_{\mathbf{x} \in \mathcal{X}} q_{X|\tilde{X}}^\phi(\mathbf{x}|\tilde{\mathbf{x}}) = 1$ for any $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$, and $q_{X|\tilde{X}}^\phi(\mathbf{x}|\tilde{\mathbf{x}})$ is differentiable *w.r.t.* ϕ and $\tilde{\mathbf{x}}$.
- (iii) The function $q_{S|\tilde{X}}^\tau : \mathcal{S} \times \tilde{\mathcal{X}} \mapsto [0, 1]$ used in (5.24) is such that $\sum_{\mathbf{s} \in \mathcal{S}} q_{S|\tilde{X}}^\tau(\mathbf{s}|\tilde{\mathbf{x}}) = 1$ for any $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$, and $q_{S|\tilde{X}}^\tau(\mathbf{s}|\tilde{\mathbf{x}})$ is differentiable *w.r.t.* τ and $\tilde{\mathbf{x}}$.

Remark 2. In supervised learning, clients would focus on a specific task related to a public variable U (dependent on X). The utility function $I(X; \tilde{X})$ could yield a degenerate sanitized output *w.r.t.* the public variable U . In this case, we can augment the objective function in (5.50a) by a cost function that is the negative minimized loss of a state-of-the-art inference model for U . For example, if U can be estimated accurately by using a regression model $h(\cdot)$, we then add the cost function $-\min_h \mathbb{E} \left[\left(h(\tilde{X}) - U \right)^2 \right]$. In this manner, as demonstrated in our experiments in Section 5.7.2, we can achieve utility of the sanitized data comparable to that of the unsanitized data.

5.5.2 Alternating Optimization Algorithm

We propose an alternating optimization method for optimizing (5.50). Recall that a training set $\{(\mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ is assumed to be available. We first sample an initial minibatch data $\{(\mathbf{x}_{j_i}, \mathbf{s}_{j_i})\}_{i=1}^M$ from the training set. We draw M *i.i.d.* noise samples from the standard normal distribution and feed the noise samples and $\{\mathbf{x}_{j_i}\}_{i=1}^M$ into the sanitizer function (5.2) to obtain $\{\tilde{\mathbf{x}}_{j_i}\}_{i=1}^M$. We use $\{(\tilde{\mathbf{x}}_{j_i}, \mathbf{x}_{j_i}, \mathbf{s}_{j_i})\}_{i=1}^M$ to form Monte Carlo estimates of (5.11), (5.17) and (5.23), respectively, since they are written as an expectation *w.r.t.* X , $\tilde{X}$ and S . Next, we freeze the sanitizer's parameter

θ and solve (5.18) and (5.24) separately. We update θ and ϕ by maximizing (5.50a) with τ and D_X being fixed. We repeat these three steps until an equilibrium is reached. A minibatch stochastic descent algorithm is outlined in Algorithm 6 for training the DRIP-Var framework. Note that Algorithm 6 can also be applied for optimizing (5.50) using the maximal correlation privacy functions under the DRIP-Max framework.

Algorithm 6 Minibatch stochastic gradient algorithm

- 1: Initialize θ, ϕ, τ, D_X .
 - 2: **repeat**
 - 3: Draw a mini-batch data points $\{(\mathbf{x}_{j_i}, \mathbf{s}_{j_i})\}_{i=1}^M$ from the training dataset and M i.i.d. noise samples $\{\xi_i\}_{i=1}^M$ from $\mathcal{N}(\mathbf{0}, \mathbf{I})$. Compute $\tilde{\mathbf{x}}_{j_i} = h_d(h_e(\mathbf{x}_{j_i}) + h_\sigma(\xi_i))$ to obtain $\{(\tilde{\mathbf{x}}_{j_i}, \mathbf{x}_{j_i}, \mathbf{s}_{j_i})\}_{i=1}^M$.
 - 4: Estimate (5.11), (5.23) and (5.51) by using the M data points $\{(\tilde{\mathbf{x}}_{j_i}, \mathbf{x}_{j_i}, \mathbf{s}_{j_i})\}_{i=1}^M$.
 - 5: Update θ, ϕ by applying stochastic gradient descent to optimize (5.50a).
 - 6: Update τ, D_X by applying stochastic gradient descent to optimize (5.50b) and (5.50c), respectively.
 - 7: **until** θ converges
-

5.6 Deep Learning Architecture for Images

In this section, we present a deep learning architecture for the proposed DRIP-Var framework for use with images, with the additional assumption that $\mathcal{S}$ is a discrete set.

5.6.1 Sanitizer

We let the probabilistic sanitizer $p_{\tilde{X}|X}^\theta$ be a combination of a fully-convolutional network (FCN) [148] encoder and decoder denoted as $h_e(\cdot)$ and $h_d(\cdot)$, respectively, with random noise injected in the bottleneck layers as illustrated in Fig. 5.4. The sanitizer compresses the raw data X to a hidden representation and then perturbs it by adding noise obtained through applying a linear transformation function $h_\sigma(\cdot)$

FIG. 5.4: Probabilistic sanitizer $p_{\tilde{X}|X}^\theta$.

on an input Gaussian noise Ξ (the covariance of the transformed additive noise is denoted as Σ). The decoder h_d up-samples the perturbed hidden representation Z . This process is exactly a reparameterization trick as we introduce an auxiliary random variable Ξ independent of X and let $\tilde{X}$ be a function of X and Ξ :

$$p_{Z|X}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(h_e(\mathbf{x}), \Sigma), \quad \tilde{X} = h_d(Z).$$

Here θ denotes the collection of FCN weights of the encoder and decoder and the noise transformation network.

5.6.2 Utility Function $\mathcal{L}(\phi; \theta)$

We let function $g^\phi(\cdot)$ represent a FCN whose output shape is same as the input shape and denote ϕ as the collection of trainable weights. We build the utility function $\mathcal{L}(\phi; \theta)$ in (5.11) by letting the composition of a radial basis function and the function $g^\phi(\cdot)$ be the variational posterior distribution of X :

$$q_{X|\tilde{X}}^\phi(\mathbf{x}|\tilde{\mathbf{x}}) = C \exp(\|\gamma(\mathbf{x}) - \gamma(g^\phi(\tilde{\mathbf{x}}))\|^2),$$

where C is a normalization constant and $\gamma(\cdot)$ is a fixed high-level feature extraction function obtained from pre-trained networks (cf. perceptual loss in [128]). We let $\gamma(\cdot)$ be an identity function if no feature map function is used.

5.6.3 Privacy Function $\mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta})$

To build the privacy function $\mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta})$, we let the variational distribution $q_{S|\tilde{X}}^{\boldsymbol{\tau}}(\cdot|\cdot)$ be formed by a convolutional neural network (CNN) classifier that relies on the softmax activation to yield the likelihood of $S = \mathbf{s}$ for each $\mathbf{s} \in \mathcal{S}$. Denote $\boldsymbol{\tau}$ as the collection of layer weights of the classifier. We rewrite (5.23) as

$$\mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta}) = \mathbb{E}_{(X, \tilde{X}) \sim p_{X, \tilde{X}}^{\boldsymbol{\theta}}} \left[\sum_{\mathbf{s} \in \mathcal{S}} p(\mathbf{s}|X) \log q_{S|\tilde{X}}^{\boldsymbol{\tau}}(\mathbf{s}|\tilde{X}) \right].$$

In order to use the training data $\{(\mathbf{x}_i, \mathbf{s}_i)\}_{i=1}^N$ to optimize the privacy function, we let $p(\mathbf{s}_i|\mathbf{x}_i) = 1$ and $p(\mathbf{s}|\mathbf{x}_i) = 0$ if $\mathbf{s} \neq \mathbf{s}_i$, for all $i = 1, \dots, N$ (hard labels). Minimizing $\mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta})$ over $\boldsymbol{\tau}$ and $\boldsymbol{\theta}$ is analogous to the adversarial training in GAN [102].

5.6.4 Regularization $\mathcal{R}(D_X; \boldsymbol{\theta})$

We use domain adaptation as the regularizer $\mathcal{R}(D_X; \boldsymbol{\theta})$ since it has been demonstrated to perform well in image to image translation [133]. Denote D_X as a discriminator that seeks to distinguish X from $\tilde{X}$. Let D_X be a FCN with $M \times M$ outputs where the sigmoid activation is applied on the last layer to convert numerical values to probabilities. Each output represents how likely the corresponding patch of an input image is sampled from p_X . This structure is originated from the patch discriminator [149]. Compared to a normal discriminator with one output only, the patch discriminator is capable of capturing both the high and low frequency features of the input image and is thus more robust. Denote $D_X^{(i)}$ as the output corresponding to the i -th patch. By averaging the discriminator loss over all the patches, we obtain a modified version of (5.17) as follows:

$$\mathcal{R}(D_X; \boldsymbol{\theta}) = \sum_{i=1}^{M^2} \mathbb{E}_{\tilde{X} \sim p_{\tilde{X}}} \left[\log \left(1 - D_X^{(i)}(\tilde{X}) \right) \right] + \mathbb{E}_{X \sim p_X} \left[\log \left(D_X^{(i)}(X) \right) \right]. \quad (5.51)$$

5.7 Numerical Experiments

In this section, we evaluate our proposed DRIP frameworks on various data sets and compare our method with state-of-the-art benchmark algorithms.

5.7.1 DRIP-Var

We evaluate the DRIP-Var framework by testing it on the UTKFace dataset [107] using the deep learning architecture sketched in Section 5.6. The UTKFace dataset contains over 20,000 face images of size 64×64 with annotations of gender, age and ethnicity. We select a subset of 7,307 images, consisting of youngsters in the age group of 0 to 20, and elders in the age group of 50 to 70. Each age generation accounts for around half of the population and among each generation, around half of the population is female and the other half is male. Among these 7,307 images, 5,846 images are used for the training and 1,461 images are used for evaluation. We convert the RGB images to grayscale. We let gender be the private variable S and test the age group prediction accuracy of the sanitized images on a pretrained model (legacy system).

5.7.1.1 Setup

The setup of the deep learning model in Section 5.5 is given in Table 5.1, where the residual block (Res-BLK) [150] is made of 2 layers of Conv2D with kernel size 4×4 , stride 1 and the number of filters same with the layer it connected to. We apply batch normalization and LeakyReLU activation with the negative slope coefficient being 0.2 on all the convolutional layers but the last layer of each model where Sigmoid, Sigmoid, Softmax and Sigmoid activation are used respectively, set $\alpha = 1, \lambda = 1, \epsilon = 0.7$ and use the Adam optimizer with learning rate 0.0001 and momentum 0.5. We train the model by following the steps depicted in Algorithm 6 until an equilibrium point is obtained. Moreover, we train an age group classifier on the raw images using the ResNet-18 [150], which serves as the legacy system

and is used to test the utility of the sanitized images. After obtaining the sanitized images, we train a gender classifier on the sanitized images using the ResNet-18 acting as the adversary. We use GeForce RTX 2080 TI Graphics Card for training and evaluation. The average running time is 0.67 seconds for each step of the gradient update and is 0.012 seconds for evaluation when batch size is set to be 60.

5.7.1.2 Results

After the training is done, we feed the raw images to the sanitizer to obtain sanitized images. A few samples of the raw images and sanitized images are shown in the first two rows of Fig. 5.5, respectively. It can be observed that the gender information is obfuscated while other facial features are preserved (e.g., it is easy to match each sanitized image to its corresponding raw image without any prior knowledge). The gender accuracy of the sanitized images for the adversary is 61% (compared to 50% for perfect privacy). To test the utility, we input the sanitized images to the pre-trained ResNet-18 age group classifier. We obtained an accuracy of 89%. This accuracy can be further improved to 95% if we include the age group classifier loss on the sanitized images into our objective function in training (cf. Remark 2).

We compare the proposed DRIP-Var with the state-of-the-art GAP [103], invariant representations learning (IRL) [100], variational fair autoencoder (VFAE) [99], and ϵ -differential privacy (DP) mechanisms [7]. Given raw data $X = \mathbf{x}$ and private variable $S = \mathbf{s}$, both IRL and VFAE learn an encoder and decoder pair $q(\mathbf{z}|\mathbf{x})$ and $p(\mathbf{x}|\mathbf{s}, \mathbf{z})$ to resolve the dependencies between the encoded hidden vector $\mathbf{z}$ and the sensitive variable $\mathbf{s}$. We reconstruct the sanitized image $\tilde{\mathbf{x}}$ from $\mathbf{z}$ using a decoder by minimizing the pixel distortion between $\tilde{\mathbf{x}}$ and the raw image $\mathbf{x}$. Some samples are shown in Fig. 5.5. We adopt the celebrated Laplace mechanism for DP. For comparative purposes, we add a Laplace noise that results in a classification accuracy comparable to our approach for the private variable.

As shown in Table 5.2, DIRP-Var achieves the best utility-privacy tradeoff: it achieves higher age group classification accuracy but lower gender classification

TABLE 5.1: DRIP-Var deep learning model
(A) Sanitizer $p_{\tilde{X}|X}^{\theta}$

Layer	Output	Kernel size	Stride	Filters
Input	$64 \times 64 \times 1$			
Conv2D	$64 \times 64 \times 32$	6×6	1	32
Conv2D	$32 \times 32 \times 64$	3×3	2	64
Conv2D	$16 \times 16 \times 128$	3×3	2	128
Res-BLK	$16 \times 16 \times 128$			
Input	$16 \times 16 \times 1$			
Conv2D	$16 \times 16 \times 64$	4×4	1	64
Conv2D	$16 \times 16 \times 128$	4×4	1	128
Add	$16 \times 16 \times 128$			
Res-BLK	$16 \times 16 \times 128$			
Res-BLK	$16 \times 16 \times 128$			
Conv2DTrans	$32 \times 32 \times 64$	3×3	2	64
Conv2DTrans	$64 \times 64 \times 1$	3×3	2	1

(B) Domain adaptation discriminator D_X

Layer	Output	Kernel size	Stride	Filters
Input	$64 \times 64 \times 1$			
Conv2D	$32 \times 32 \times 32$	4×4	2	32
Conv2D	$16 \times 16 \times 64$	4×4	2	64
Conv2D	$8 \times 8 \times 128$	4×4	2	128
Conv2D	$8 \times 8 \times 32$	4×4	1	32
Conv2D	$8 \times 8 \times 1$	4×4	1	1

(C) Variational distribution $q_{S|\tilde{X}}^{\tau}$

Layer	Output	Kernel size	Stride	Filters
Input	$64 \times 64 \times 1$			
Conv2D	$32 \times 32 \times 32$	4×4	2	32
Conv2D	$16 \times 16 \times 64$	4×4	2	64
Conv2D	$8 \times 8 \times 128$	4×4	2	128
Conv2D	$8 \times 8 \times 32$	4×4	1	32
Dense	2			

(D) g^{ϕ} for variational posterior $q_{X|\tilde{X}}^{\phi}$

Layer	Output	Kernel size	Stride	Filters
Input	$64 \times 64 \times 1$			
Conv2D	$64 \times 64 \times 32$	6×6	1	32
Conv2D	$64 \times 64 \times 64$	3×3	1	64
Conv2D	$64 \times 64 \times 32$	3×3	1	32
Conv2D	$64 \times 64 \times 1$	3×3	1	1

accuracy than the other sanitization methods. In addition, we observe more visual distortions in the sanitized images produced by GAP. This is due to the domain adaptation technique incorporated in DRIP-Var while GAP only considers the pixel distortion as the utility. It can also be observed from Fig. 5.5 that IRL is better than VFAE in terms of removing the gender information. However, both IRL and VFAE do not retain much details of the raw images. This may be because the normal distribution of the hidden vector $\mathbf{z}$ imposed by these methods forces the encoder to extract only the general features of the images. We also observe that DP adds excessive noise to the raw images and leads to greater utility loss.

FIG. 5.5: The true labels are displayed on top, where M, F, Y, and E denote male, female, youngster, and elder, respectively. The 1st row shows the original images. Gender is the private variable and age group classification is the utility.

TABLE 5.2: Image classification results. Gender is the private variable.

Method	Gender accuracy	Age group accuracy
ResNet-18 (raw images)	99.6%	99.1%
DRIP-Var	61.0%	89.0%
GAP	61.5%	70.0%
IRL	76.0%	57.0%
VFAE	78.0%	58.0%
DP	63.0%	58.0%

5.7.2 DRIP-Max

We evaluate the DRIP-Max framework on the German credit dataset obtained from the UCI machine learning repository [101, 151]. The German credit dataset contains 1000 samples of personal financial data and the objective is to predict whether a person has a good credit score based on 20 qualitative and quantitative observations such as credit amount and employment status. For each numerical variable, the data are normalized in the range of 0 to 1. The whole training set is split into 800 training data and 200 testing data. We choose the age (continuous variable) to be the private variable and the credit score (binary variable) to be public variable in Case 1, and reverse the choices in Case 2. The rest of the 19 observations are treated as raw data.

5.7.2.1 Setup

In Case 1, we construct a privacy function using the neural network method depicted in Section 5.4.2. We redefine the utility function $\mathcal{L}(\phi; \theta)$ in (5.50) to be the negative cross-entropy loss of a gender classifier where ϕ denotes the trainable weights of the classifier.

In Case 2, we construct a privacy function using the kernel method depicted in Appendix 5.C where we set $\eta = 0.01$ and $K_{\tilde{X}}$ and K_S to be the RBF kernel with bandwidth $\sigma = 1$. We redefine the utility function $\mathcal{L}(\phi; \theta)$ in (5.50) to be the negative mean absolute error (MAE) loss of an age regressor where ϕ denotes the trainable weights of the regressor.

The configurations of the sanitizer, maximal correlation function networks (in reference to f' and g in Fig. 5.3), age regressor and credit score classifier are illustrated in Fig. 5.6, where dense layers with 20 units and batch normalization are employed followed by LeakyReLU activation with the negative slope coefficient being 0.1, and Sigmoid activation, Linear activation, Softmax activation, Sigmoid activation, are used for each model respectively. The age regressor and credit score classifier serve as the adversary and utility models.

We use MMD as the regularization term in (5.50) and set $\alpha = 0.1$ and $\lambda = 1$ in this experiment. We use the Adam optimizer with learning rate 0.0001 to train the model following Algorithm 6.

FIG. 5.6: DRIP-Max configurations.

5.7.2.2 Results

To evaluate the sanitizer, we train an age regressor and credit score classifier on the sanitized data to obtain the adversarial loss for Cases 1 and 2, respectively. We compare against two prominent tools in compressive privacy: the discriminant component analysis (DCA) for Case 1 and the desensitized principal component analysis (DPCA) for Case 2, as well as the aforementioned VFAE method for both the cases. DCA and DPCA proposed by [152] perform data compression using the concept of the different utility and privacy signal subspaces.

In Table 5.3a, the 2nd column shows the maximal correlation between the age and sanitized data, the 3rd column shows the adversarial loss on the age regression, and the 4th and 5th columns show the cross-entropy loss and accuracy of the credit score classification (utility model). A down arrow $\downarrow$ indicates that a smaller value is better for utility-privacy tradeoff. An up arrow $\uparrow$ indicates the opposite. Same notations are applied to Table 5.3b. It can be observed that privacy can be better protected at the cost of lower utility. A lower maximal correlation leads to a larger adversarial loss. Thus, the proposed maximal correlation estimators seem to be a good privacy leakage measure. DRIP-Max outshines DPCA and VFAE, while DCA is comparable with DRIP-Max in terms of utility-privacy tradeoff. However, unlike DRIP-Max, DCA and DPCA cannot adjust the utility-privacy tradeoff and can only handle categorical utility and private variables, thus have only limited applications.

TABLE 5.3: German credit data results
(A) Case 1: Age is private

Method	Age max. corr. ($\downarrow$)	Age MAE ($\uparrow$)	Credit score loss ($\downarrow$)	Credit score accuracy ($\uparrow$)
Raw data	0.85	0.14	0.52	78%
DRIP-Max($\epsilon = 0.79$)	0.79	0.21	0.56	77.2%
DRIP-Max($\epsilon = 0.58$)	0.58	0.33	0.61	71.1%
VFAE	0.81	0.18	0.62	72%
DCA	0.56	0.34	0.65	69.9%

(B) Case 2: Credit score is private

Method	Credit score loss ($\uparrow$)	Credit score accuracy ($\downarrow$)	Age max. corr. ($\uparrow$)	Age MAE ($\downarrow$)
Raw data	0.52	78%	0.85	0.14
DRIP-Max($\epsilon = 0.59$)	0.61	73.2%	0.83	0.15
DRIP-Max($\epsilon = 0.53$)	0.63	69.7%	0.82	0.16
VFAE	0.60	74.6%	0.82	0.15
DPCA	0.59	76%	0.81	0.16

5.8 Conclusion

We have proposed a data-driven privacy-preserving framework to prevent leakage of private information. We approximated mutual information for quantifying the utility of the sanitized data and introduced domain regularization techniques to regulate the domain of the sanitized data. We developed practical implementation of average information leakage and maximal correlation for privacy leakage measures and carried out asymptotic convergence analysis. Numerical results demonstrated that our approach achieves better privacy-utility tradeoff than current state-of-the-art methods.

5.A Proof of Proposition 4

We need Lemma 5.A.1 to convert (5.31) to an unconstrained optimization. Note $\mathcal{Y}$ and $\mathcal{Z}$ are defined in (5.28) and (5.29).

Lemma 5.A.1. Assuming (5.30), we have

$$\rho^2(Y, Z) = \max_{f \in \mathcal{Z}} \frac{\|Tf\|_{\mathcal{Z}}^2 - \mathbb{E}[f(Y)]^2}{\|f\|_{\mathcal{Y}}^2},$$

Proof. Consider a function $f \in \mathcal{Y}$. We de-mean f to obtain $f' = f - \mathbb{E}[f(Y)]$. Then, we have

$$\begin{aligned} \|Tf'\|_{\mathcal{Z}}^2 &= \|Tf\|_{\mathcal{Z}}^2 - \mathbb{E}[f(Y)]^2, \\ \|f'\|_{\mathcal{Y}}^2 &= \|f\|_{\mathcal{Y}}^2 - \mathbb{E}[f(Y)]^2. \end{aligned}$$

As a result, we have

$$\begin{aligned} \frac{\|Tf'\|_{\mathcal{Z}}^2 - \mathbb{E}[f'(Y)]^2}{\|f'\|_{\mathcal{Y}}^2} &= \frac{\|Tf\|_{\mathcal{Z}}^2 - \mathbb{E}[f(Y)]^2}{\|f\|_{\mathcal{Y}}^2 - \mathbb{E}[f(Y)]^2} \\ &\geq \frac{\|Tf\|_{\mathcal{Z}}^2 - \mathbb{E}[f(Y)]^2}{\|f\|_{\mathcal{Y}}^2}. \end{aligned}$$

This implies that it is sufficient to consider the zero mean functions in $\mathcal{Y}$. Then, we have

$$\begin{aligned} \max_{f \in \mathcal{Y}} \frac{\|Tf\|_{\mathcal{Z}}^2 - \mathbb{E}[f(Y)]^2}{\|f\|_{\mathcal{Y}}^2} &= \max_{\substack{f \in \mathcal{Y} \\ \mathbb{E}[f(Y)] = 0}} \frac{\|Tf\|_{\mathcal{Z}}^2 - \mathbb{E}[f(Y)]^2}{\|f\|_{\mathcal{Y}}^2} \\ &= \max_{\substack{f \in \mathcal{Y} \\ \mathbb{E}[f(Y)] = 0}} \frac{\|Tf\|_{\mathcal{Z}}^2}{\|f\|_{\mathcal{Y}}^2} = \rho^2(Y, Z), \end{aligned}$$

where the last equality is given by (5.31). $\square$

Lemma 5.A.2. Assuming (5.30), for each $f \in \mathcal{Y}$, we have

$$\|Tf\|_{\mathcal{Z}}^2 = \max_{g \in \mathcal{Z}} \{2\mathbb{E}[f(Y)g(Z)] - \mathbb{E}[g^2(Z)]\}.$$

Proof. For a given $f \in \mathcal{Y}$, since $g(Z) = \mathbb{E}[f(Y) | Z]$ minimizes $\mathbb{E}[(f(Y) - g(Z))^2]$, we have

$$\begin{aligned} &\min_{g \in \mathcal{Z}} \mathbb{E}[(f(Y) - g(Z))^2] \\ &= \mathbb{E}[(f(Y) - \mathbb{E}[f(Y) | Z])^2] \\ &= \mathbb{E}[f^2(Y)] - \mathbb{E}[\mathbb{E}[f(Y) | Z]^2] \\ &= \mathbb{E}[f^2(Y)] - \|Tf\|_{\mathcal{Z}}^2. \end{aligned}$$

The proof is complete by expanding the left-hand side of the above equation. $\square$

Lemma 5.A.1 and Lemma 5.A.2 immediately give us the result in Proposition 4 that maximal correlation can be obtained by solving

$$\max_{f \in \mathcal{Y}, g \in \mathcal{Z}} \frac{2\mathbb{E}[f(Y)g(Z)] - \mathbb{E}[g^2(Z)] - \mathbb{E}[f(Y)]^2}{\mathbb{E}[f^2(Y)]}. \quad (5.52)$$

Suppose (5.52) is optimized at $f = f^*$ and $g = g^*$. From the Cauchy-Schwarz inequality, we have

$$\mathbb{E}[f(Y)g(Z)] = \mathbb{E}[g(Z)\mathbb{E}[f(Y) | Z]]$$

$$\leq \sqrt{\mathbb{E}[g^2(Z)] \mathbb{E}[\mathbb{E}[f(Y)|Z]^2]},$$

where the equality holds if and only if $g(Z)$ is proportional to $\mathbb{E}[f(Y)|Z]$. Therefore, $g^*(Z) = \mathbb{E}[f^*(Y) | Z]$ and $\frac{g^*(Z)}{\sqrt{\mathbb{E}[g^{*2}(Z)]}}$ is the maximal correlation function of Z .

5.B Proof of Proposition 5

Denote the numerator and denominator of $\lambda_n(\boldsymbol{\tau})$ respectively as

$$v_n(\boldsymbol{\tau}) = 2\mu_n(f_{\boldsymbol{\tau}_1} \otimes g_{\boldsymbol{\tau}_2}) - \mu_n(g_{\boldsymbol{\tau}_2}^2) - (\mu_n(f_{\boldsymbol{\tau}_1}))^2, \quad (5.53)$$

$$d_n(\boldsymbol{\tau}_1) = \mu_n(f_{\boldsymbol{\tau}_1}^2). \quad (5.54)$$

Consider function $\ell(\mathbf{x}, \mathbf{y}) = v_n(\mathbf{x})d_n(\mathbf{y})$ for $\mathbf{x} \in \Omega$ and $\mathbf{y} \in \Omega_1$. Letting $\nabla v_n(\boldsymbol{\tau})$ be the gradient of $v_n(\boldsymbol{\tau})$, we have

$$\begin{aligned} & |\ell(\mathbf{x}, \mathbf{y}) - \ell(\mathbf{x}', \mathbf{y}')| \\ & \leq \max_{\gamma \in \Omega, \alpha \in \Omega_1} |d_n(\alpha) \nabla v_n(\gamma)^\top (\mathbf{x} - \mathbf{x}') + v_n(\gamma) \nabla d_n(\alpha)^\top (\mathbf{y} - \mathbf{y}')| \\ & \leq \max_{\gamma \in \Omega, \alpha \in \Omega_1} |d_n(\alpha) \nabla v_n(\gamma)^\top (\mathbf{x} - \mathbf{x}')| + \max_{\gamma \in \Omega, \alpha \in \Omega_1} |v_n(\gamma) \nabla d_n(\alpha)^\top (\mathbf{y} - \mathbf{y}')| \\ & \leq \max_{\alpha \in \Omega_1} |d_n(\alpha)| \max_{\gamma \in \Omega} \|\nabla v_n(\gamma)\| \|\mathbf{x} - \mathbf{x}'\| + \max_{\gamma \in \Omega} |v_n(\gamma)| \max_{\alpha \in \Omega_1} \|\nabla d_n(\alpha)\| \|\mathbf{y} - \mathbf{y}'\|, \end{aligned} \quad (5.55)$$

where the last inequality follows from the Cauchy-Schwarz inequality. From (5.55), we have

$$\begin{aligned} |\lambda_n(\boldsymbol{\tau}) - \lambda_n(\boldsymbol{\tau}')| &= \left| \frac{v_n(\boldsymbol{\tau})}{d_n(\boldsymbol{\tau}_1)} - \frac{v_n(\boldsymbol{\tau}')}{d_n(\boldsymbol{\tau}'_1)} \right| \\ &= \frac{|v_n(\boldsymbol{\tau})d_n(\boldsymbol{\tau}'_1) - v_n(\boldsymbol{\tau}')d_n(\boldsymbol{\tau}_1)|}{d_n(\boldsymbol{\tau}_1)d_n(\boldsymbol{\tau}'_1)} \\ &\leq B_n \|\boldsymbol{\tau} - \boldsymbol{\tau}'\|, \end{aligned}$$

where

$$B_n = \left(\max_{\alpha \in \Omega_1} |d_n(\alpha)| \max_{\gamma \in \Omega} \|\nabla v_n(\gamma)\| + \max_{\gamma \in \Omega} |v_n(\gamma)| \max_{\alpha \in \Omega_1} \|\nabla d_n(\alpha)\| \right) \left(\min_{\alpha \in \Omega_1} d_n(\alpha) \right)^{-2}. \quad (5.56)$$

Since $f_{\tau_1}(\tilde{\mathbf{x}})$ is smooth **w.r.t.** $\tau_1 \in \Omega_1$ and continuous **w.r.t.** $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$, its gradient **w.r.t.** τ_1 is (uniformly) continuous on $\Omega_1 \times \tilde{\mathcal{X}}$ in the norm metric and hence bounded due to the compactness of Ω_1 and $\tilde{\mathcal{X}}$. The same property applies to $g_{\tau_2}(\cdot)$. Therefore, we can apply the uniform law of large numbers [145] to verify that each max or min term in (5.56) converges almost surely to some constant, e.g.,

$$\min_{\tau_1 \in \Omega_1} d_n(\tau_1) \xrightarrow{\text{a.s.}} \min_{\tau_1 \in \Omega_1} \mathbb{E} [\mu_n(f_{\tau_1}^2)] > 0.$$

Now we can conclude that there is a scalar $b < \infty$ such that $B_n \xrightarrow{\text{a.s.}} b$ as $n \rightarrow \infty$.

5.C Kernel Method for Maximal Correlation

To provide a complement to the neural network method for estimating maximal correlation, we work with functions in RKHSs and make use of the kernel trick to search over this space efficiently [153]. For $Y = \tilde{X}$ and S , let $\mathcal{H}_Y$ be a RHKS of real-valued functions of Y , associated with the Mercer kernel $K_Y(\cdot, \cdot)$. Let $\Phi_Y(y) = K_Y(y, \cdot)$ and $\langle \cdot, \cdot \rangle_Y$ be feature map and inner product, respectively. The point evaluation of $f \in \mathcal{H}_Y$ can be written as $f(y) = \langle \Phi_Y(y), f \rangle$. The maximal correlation of $\tilde{X}$ and S restricted to functions in $\mathcal{H}_{\tilde{X}}, \mathcal{H}_S$ is

$$\begin{aligned} \rho_K(\tilde{X}, S) &= \sup_{f \in \mathcal{H}_{\tilde{X}}, g \in \mathcal{H}_S} \mathbb{E} [f(\tilde{X})g(S)] \\ &= \sup_{f \in \mathcal{H}_{\tilde{X}}, g \in \mathcal{H}_S} \mathbb{E} [\langle \Phi_{\tilde{X}}(\tilde{X}), f \rangle \langle \Phi_S(S), g \rangle], \end{aligned}$$

with constraints $\mathbb{E} [f(\tilde{X})] = 0$, $\mathbb{E} [g(S)] = 0$, $\mathbb{E} [f^2(\tilde{X})] = 1$ and $\mathbb{E} [g^2(S)] = 1$.

It has been shown in [154] that if $\mathcal{H}_{\tilde{X}}$ and $\mathcal{H}_S$ are the RKHSs corresponding to a Gaussian kernel on $\mathbb{R}$, $\rho_K(\tilde{X}, S) = 0$ if and only if the random variables $\tilde{X}$ and S are independent of each other. For $\tilde{X}$ and S with finite support, Lemma 5.C.1 below provides a sufficient condition for $\rho_K(\tilde{X}, S) = \rho(\tilde{X}, S)$ under which $\tilde{X}$ and S are independent if $\rho_K(\tilde{X}, S) = 0$.

Lemma 5.C.1. Suppose $\tilde{\mathcal{X}} = \{\chi_1, \dots, \chi_L\}$ and $\mathcal{S} = \{\nu_1, \dots, \nu_N\}$. Let $\mathbf{G}_{\tilde{X}}$ and $\mathbf{G}_S$ be the Gram matrices associated with $\tilde{\mathcal{X}}$ and $\mathcal{S}$, respectively. If $\text{rank}(\mathbf{G}_{\tilde{X}}) = M$ and $\text{rank}(\mathbf{G}_S) = N$, $\rho_K(\tilde{X}, S) = \rho(\tilde{X}, S)$.

Proof. Let $\mathbf{a} = [a_1, \dots, a_L] \in \mathbb{R}^L$ and $\mathbf{b} = [b_1, \dots, b_N] \in \mathbb{R}^N$. Let $f = \sum_{i=1}^L a_i \Phi_{\tilde{X}}(\chi_i)$ and $g = \sum_{i=1}^N b_i \Phi_S(\nu_i)$. Then, $f \in \mathcal{H}_{\tilde{X}}$ and $g \in \mathcal{H}_S$. Since $f(\chi_i) = \langle \Phi_{\tilde{X}}(\chi_i), f \rangle$ and $g(\nu_i) = \langle \Phi_S(\nu_i), g \rangle$, we have $[f(\chi_1), \dots, f(\chi_L)] = \mathbf{a} \mathbf{G}_{\tilde{X}}$ and $[g(\nu_1), \dots, g(\nu_N)] = \mathbf{b} \mathbf{G}_S$. Since $\mathbf{G}_{\tilde{X}}$ and $\mathbf{G}_S$ are full-rank, there exist $\mathbf{a}$ and $\mathbf{b}$ so that f and g are the maximal correlation functions of $\tilde{X}$ and S given by Definition 10. The proof is completed by noting that $\rho_K(\tilde{X}, S) \leq \rho(\tilde{X}, S)$. $\square$

Using the data samples $\{(\tilde{\mathbf{x}}_i, \mathbf{s}_i)\}_{i=1}^n$, the empirical estimate of the maximal correlation of $\tilde{X}$ and S is:

$$\hat{\rho}_K(\tilde{X}, S) = \frac{1}{n} \max_{f \in \mathcal{H}_{\tilde{X}}, g \in \mathcal{H}_S} \sum_{i=1}^n \langle \bar{\Phi}_{\tilde{X}}(\tilde{\mathbf{x}}_i), f \rangle \langle \bar{\Phi}_S(\mathbf{s}_i), g \rangle, \quad (5.57)$$

where $\bar{\Phi}_{\tilde{X}}(\tilde{\mathbf{x}}_i) = \Phi_{\tilde{X}}(\tilde{\mathbf{x}}_i) - \frac{1}{n} \sum_{i=1}^M \Phi_{\tilde{X}}(\tilde{\mathbf{x}}_i)$ and $\bar{\Phi}_S(\mathbf{s}_i) = \Phi_S(\mathbf{s}_i) - \frac{1}{n} \sum_{i=1}^n \Phi_S(\mathbf{s}_i)$ are called the centered feature maps [155] ensuring the zero mean constraint is satisfied. Since the optimal f and g lie in the linear spaces spanned by the feature maps of the data samples, we write

$$\begin{aligned} f &= \sum_{i=1}^n a_i \Phi_{\tilde{X}}(\tilde{\mathbf{x}}_i), \\ g &= \sum_{i=1}^n b_i \Phi_S(\mathbf{s}_i), \end{aligned}$$

where $[a_1, \dots, a_n] = \mathbf{a}^\top$ and $[b_1, \dots, b_n] = \mathbf{b}^\top \in \mathbb{R}^n$. Let $\mathbf{K}_{\tilde{X}}$ and $\mathbf{K}_S$ be the Gram matrices associated with the data points $\{\tilde{\mathbf{x}}_i\}_{i=1}^n$ and $\{\mathbf{s}_i\}_{i=1}^n$, respectively. Let

$\bar{\mathbf{K}}_{\tilde{X}} = \mathbf{H}\mathbf{K}_{\tilde{X}}\mathbf{H}$ and $\bar{\mathbf{K}}_S = \mathbf{H}\mathbf{K}_S\mathbf{H}$ where $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}$ is a constant centering matrix ($\mathbf{I} \in \mathbb{R}^{n \times n}$ is an identity matrix and $\mathbf{1} \in \mathbb{R}^{n \times n}$ is an all-1s matrix). The problem becomes [154]: $\hat{\rho}_K(\tilde{X}, S) =$

$$\max_{\mathbf{a}, \mathbf{b}} \frac{\mathbf{a}^\top \bar{\mathbf{K}}_{\tilde{X}} \bar{\mathbf{K}}_S \mathbf{b}}{\sqrt{\mathbf{a}^\top (\bar{\mathbf{K}}_{\tilde{X}} + \eta \mathbf{I})^2 \mathbf{a}} \sqrt{\mathbf{b}^\top (\bar{\mathbf{K}}_S + \eta \mathbf{I})^2 \mathbf{b}}}, \quad (5.58)$$

where η is a small positive constant for regularization. This is equivalent to a generalized eigenvalue problem:

$$\begin{bmatrix} \mathbf{0} & \bar{\mathbf{K}}_{\tilde{X}} \bar{\mathbf{K}}_S \\ \bar{\mathbf{K}}_S \bar{\mathbf{K}}_{\tilde{X}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \rho \begin{bmatrix} (\bar{\mathbf{K}}_{\tilde{X}} + \eta \mathbf{I})^2 & \mathbf{0} \\ \mathbf{0} & (\bar{\mathbf{K}}_S + \eta \mathbf{I})^2 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}. \quad (5.59)$$

Let $\mathbf{a} = \mathbf{a}^*$ and $\mathbf{b} = \mathbf{b}^*$ be the solution of (5.59). Then, (5.57) can be written as

$$\hat{\rho}_K(\tilde{X}, S) = \frac{1}{n} \mathbf{a}^{*\top} \bar{\mathbf{K}}_{\tilde{X}} \bar{\mathbf{K}}_S \mathbf{b}^* = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \alpha_{i,j} K_{\tilde{X}}(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j), \quad (5.60)$$

where $\alpha_{ij} = [\mathbf{H}\mathbf{a}^*]_i [\mathbf{H}^2 \mathbf{K}_S \mathbf{H}\mathbf{b}^*]_j$ with $[\mathbf{a}^*]_i$ being the i -th element of $\mathbf{a}^*$.

To make the form of the privacy function here consistent with that of (5.24), we let $\boldsymbol{\tau} = \{\mathbf{a}, \mathbf{b}\}$ and denote $\mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta})$ as the function on the right-hand side of (5.58) to be maximized. Then, we can write $\mathcal{P}(\boldsymbol{\theta}) = \max_{\boldsymbol{\tau}} \mathcal{P}(\boldsymbol{\tau}; \boldsymbol{\theta}) = \hat{\rho}_K(\tilde{X}, S)$. The privacy function $\mathcal{P}(\boldsymbol{\theta})$ is differentiable when the kernel $K_{\tilde{X}}$ is differentiable. For example, for any shift-invariant kernel $K_{\tilde{X}}(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j) = K_{\tilde{X}}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)$ such as the radial basis function kernel, the gradient of the privacy function has a concise expression:

$$\frac{\partial \mathcal{P}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \tilde{\mathbf{x}}_i}{\partial \boldsymbol{\theta}} \bigg|_{\mathbf{x}_i, \boldsymbol{\xi}_i} \left(\sum_{j=1}^n (\alpha_{ij} - \alpha_{ji}) \nabla K_{\tilde{X}}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j) \right), \quad (5.61)$$

where $\nabla K_{\tilde{X}}(\tilde{\mathbf{x}})$ denotes the gradient of the kernel $K_{\tilde{X}}$ at $\tilde{\mathbf{x}}$ and $\frac{\partial \tilde{\mathbf{x}}_i}{\partial \boldsymbol{\theta}} \big|_{\mathbf{x}_i, \boldsymbol{\xi}_i}$ is a Jacobian matrix derived from the sanitizer network at the fixed input $(\mathbf{x}_i, \boldsymbol{\xi}_i)$ with $\tilde{\mathbf{x}}_i = f^\theta(\mathbf{x}_i, \boldsymbol{\xi}_i)$. Thus, the gradient can be backpropagated through the sanitizer network to update the sanitization parameter $\boldsymbol{\theta}$.

Chapter 6

Conclusion and Future Work

At the core of this thesis is developing a methodology for preventing a data aggregator or service provider from obtaining accurate inference of some private information when collecting data from sensors or users. We built the mathematical formulations of our privacy-preserving frameworks using techniques from different fields including information theory, convex optimization, estimation and approximation. To manage the irreconcilable tensions between privacy and utility in privacy preserving data releasing, we provided tools to optimally randomize raw data before transmission for achieving a desired level of privacy. Specifically, given a maximum permitted level of privacy leakage tolerated by a particular application, we gave privacy mechanisms that maximize utility while satisfying the privacy constraint. Several metrics including [CRLB](#), mutual information and maximal correlation were employed to quantify privacy leakage. To overcome the numerical challenges of computing a probabilistic privacy mapping for continuous data, we adopted noise perturbation and compression as data sanitization mechanisms in this thesis. In [Chapter 3](#), we considered protecting the privacy of parameter estimation in a decentralized linear system where system parameters are composed of useful and sensitive parts. It was revealed that it is possible to perturb the sensor measurements in such a way that the estimation error of the public variables remains unchanged while the estimation error of the private variables can become

arbitrarily large. This is called arbitrarily strong utility-privacy tradeoff, for which to be attainable the necessary and sufficient conditions were derived and sanitization methods were developed accordingly. This notion is analogous to the concept of perfect privacy that guarantees that no information is revealed about a private variable [75–78]. Chapter 4 extended the framework in Chapter 3 by considering a dynamic system in which temporal correlation exists among a sequence of measurement data made by sensors over time. Focusing on online privacy mechanisms, we observed that governing the utility-privacy tradeoff using the present information only at each time step may compromise the privacy of a future state. Then, we proposed a look-ahead framework to ensure the same level of privacy in all future time steps in a LDS. We noted many literature on privacy protection including the works in Chapters 3 and 4 require prior knowledge of the system model or the statistics of the underlying data, which is however unknown and cannot be captured explicitly by any analytical model in many real scenarios. In view of the situation, we proposed a data-driven approach in Chapter 5 to learn sanitization mechanisms from data to identify complex patterns. We adopted mutual information and maximal correlation as privacy metrics and proposed to data-driven versions of these metrics. The derived asymptotic convergence analysis provided theoretical guarantees for their practical implementation. Moreover, we introduced domain regularization techniques to avoid over-fitting and to ensure that the sanitized data is compatible with a legacy inference system by forcing the marginal distributions between the sanitized data and the raw data to be similar.

This study leaves a number open issues and areas for future work. First, it would be interesting to enhance the practicability of the frameworks depicted in Chapters 3 and 4 by incorporating additional factors other than privacy and utility constraints. For example, sensor nodes in energy harvesting wireless sensor networks harvest energy from the environment to improve the sensor’ lifetime, sensing reliability and transmission coverage [156, 157]. In such case, there is a need to consider the impact of energy efficiency on the design of sanitization mechanisms.

Furthermore, it would be of interest to adapt the data-driven framework in Chapter 5 (which focuses on a single data point) to deal with temporal correlations in time-series data. This is challenging as we have shown in Chapter 4 that methods that preserve the privacy for the current time may leak significant amount of information at the trace level as the adversary can exploit temporal correlations in a trace. Relevant studies can be found in [158] in which the authors studied the utility-privacy tradeoff in time-series data sharing by solving the problem of minimizing the mutual information between a true data sequence and a shared version while satisfying a distortion constraint based on deep reinforcement learning. However, unlike the framework proposed in Chapter 5, no specific public or private variables dependent on raw data are defined in this case. Recently, graph signal processing and graph neural networks are gaining more attention in the research community. Exploring the privacy issue for data defined on irregular graph domains will be an interesting new research direction. Prior works can be found in [159, 160], where the reference [159] presented the first differentially private clustering method for arbitrary shaped node clusters in a graph and the paper [160] developed a privacy-preserving learning algorithm with local differential privacy guarantee for graph convolutional networks to protect sensitive node attributes. These works fall into the category of *data privacy*. To the best of our knowledge, *inference privacy* in this regard has not been investigated.

List of Author's Publications

Journal Articles

- **C. X. Wang**, Y. Song, W. P. Tay, “Arbitrarily strong utility-privacy tradeoff in multi-agent systems,” IEEE Trans. Inf. Forensics Security, pp. 1-1, 2020.
- Y. Song, **C. X. Wang**, and W. P. Tay, “Compressive privacy for a linear dynamical system,” IEEE Trans. Inf. Forensics Security, vol. 15, no. 1, pp. 895-910, Dec. 2020.

Conference Proceedings

- **C. X. Wang** and W. P. Tay, “Data-driven privacy with domain regularization,” in Proc. IEEE Global Telecomm. Conf., Taipei, Taiwan, Dec. 2020.
- **C. X. Wang**, W. P. Tay, and Y. Song, “Maximum privacy under perfect utility in sensor networks,” in Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop, Hangzhou, China, Jun. 2020.
- **C. X. Wang**, Y. Song, and W. P. Tay, “Preserving parameter privacy in sensor networks,” in Proc. IEEE Global Conf. on Signal and Information Processing, Anaheim, USA, Nov. 2018.
- Y. Song, **C. X. Wang**, and W. P. Tay, “Privacy-aware Kalman filtering,” in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Calgary, Canada, Apr. 2018.

- Y. Song, **C. X. Wang**, W. P. Tay, and C. L. Law, “Grid-based belief propagation,” in Int. Conf. on Indoor Positioning and Indoor Navigation, Sapporo, Japan, Sep. 2017.

Journals under Submission/Revision

- **C. X. Wang**, W. P. Tay, “Data-driven regularized inference privacy,” to IEEE Trans. Dependable Secure Comput.

Bibliography

- [1] J. H. Abawajy, M. I. H. Ninggal, and T. Herawan, “Privacy preserving social network data publication,” *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1974–1997, Mar. 2016. [1](#), [6](#)
- [2] J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, “Security and privacy for cloud-based IoT: Challenges,” *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 26–33, Jan. 2017. [1](#), [7](#)
- [3] Wikipedia. Whatsapp security and privacy. [Online]. Available: https://en.wikipedia.org/wiki/WhatsApp#Security_and_privacy [1](#)
- [4] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, “Membership inference attacks against machine learning models,” in *Proc. IEEE Symp. on Security and Privacy*, San Jose, CA, USA, May 2017. [1](#)
- [5] F. M. Naini, J. Unnikrishnan, P. Thiran, and M. Vetterli, “Where you are is who you are: User identification by matching statistics,” *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 2, pp. 358–372, Feb. 2016. [1](#)
- [6] I. Dinur and K. Nissim, “Revealing information while preserving privacy,” in *Proc. ACM Int. Conf. on Management of Data*, San Diego, California, Jun. 2003. [1](#)
- [7] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, Aug. 2014. [2](#), [8](#), [125](#)

- [8] S. Chawla, C. Dwork, F. McSherry, A. Smith, and H. Wee, “Toward privacy in public databases,” in *Theory of Cryptography*. Berlin, Heidelberg: Springer, 2005, pp. 363–385. [2](#), [8](#)
- [9] K. E. Emam, E. Jonker, L. Arbuckle, and B. Malin, “A systematic review of re-identification attacks on health data,” *PLOS One*, vol. 6, no. 12, Dec. 2011. [2](#)
- [10] L. Sankar, S. R. Rajagopalan, and H. V. Poor, “Utility-Privacy tradeoffs in databases: An information-theoretic approach,” *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 838–852, Jun. 2013. [2](#), [10](#)
- [11] M. Shateri, F. Messina, P. Piantanida, and F. Labeau, “Real-time privacy-preserving data release for smart meters,” *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5174–5183, Nov. 2020. [3](#)
- [12] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A survey on sensor networks,” *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002. [4](#)
- [13] B. C. Csaji, Z. Kemeny, G. Pedone, A. Kuti, and J. Vancza, “Wireless multi-sensor networks for smart cities: A prototype system with statistical data analysis,” *IEEE Sensors J.*, vol. 17, no. 23, pp. 7667–7676, Dec. 2017. [4](#)
- [14] W. Tang, F. Ji, and W. P. Tay, “Estimating infection sources in networks using partial timestamps,” *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 12, pp. 3035–3049, Dec. 2018. [4](#)
- [15] J. Yang, X. Zhong, and W. P. Tay, “A dynamic Bayesian nonparametric model for blind calibration of sensor networks,” *IEEE Internet of Things J.*, vol. 5, no. 5, pp. 3942–3953, Oct. 2018. [4](#)
- [16] J.-F. Chamberland and V. V. Veeravalli, “Decentralized detection in sensor networks,” *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003. [4](#)

- [17] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Asymptotic performance of a censoring sensor network," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4191–4209, Nov. 2007. [4](#)
- [18] ———, "Data fusion trees for detection: Does architecture matter?" *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4155–4168, Sep. 2008. [4](#)
- [19] O. P. Kreidl, J. N. Tsitsiklis, and S. I. Zoumpoulis, "On decentralized detection with partial information sharing among sensors," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1759–1765, Apr. 2011. [4](#)
- [20] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Bayesian detection in bounded height tree networks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4042–4051, Oct. 2009. [4](#)
- [21] W. P. Tay, "The value of feedback in decentralized detection," *IEEE Trans. Inf. Theory*, vol. 58, no. 12, pp. 7226–7239, Dec. 2012. [4](#)
- [22] Z. Zhang, E. K. P. Chong, A. Pezeshki, W. Moran, and S. D. Howard, "Learning in hierarchical social networks," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 2, pp. 305–317, Apr. 2013. [4](#)
- [23] W. P. Tay, "Whose opinion to follow in multihypothesis social learning? A large deviations perspective," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 2, pp. 344–359, Mar. 2015. [4](#)
- [24] J. Ho, W. P. Tay, T. Q. S. Quek, and E. K. P. Chong, "Robust decentralized detection and social learning in tandem networks," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5019–5032, Oct. 2015. [4](#)
- [25] H. Alemdar and C. Ersoy, "Wireless sensor networks for healthcare: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2688–2710, Oct. 2010. [4](#)
- [26] A. S. Behbahani, A. M. Eltawil, and H. Jafarkhani, "Decentralized estimation under correlated noise," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5603–5614, Nov. 2014. [5](#)

- [27] J. Xiao, S. Cui, Z. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 757–770, Feb. 2008. 5
- [28] A. Speranzon, C. Fischione, K. H. Johansson, and A. Sangiovanni-Vincentelli, "A distributed minimum variance estimator for sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 609–621, May 2008. 5
- [29] L. Sankar, "Competitive privacy: Distributed computation with privacy guarantees," in *Proc. IEEE Global Conf. on Signal and Information Processing*, Austin, Texas, Dec. 2013. 5
- [30] V. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *IEEE Trans. Ind. Electron.*, vol. 57, no. 10, pp. 3557–3564, Oct. 2010. 5
- [31] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-Time urban monitoring using cell phones: A case study in Rome," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011. 5
- [32] Y. Hong, W. M. Liu, and L. Wang, "Privacy preserving smart meter streaming against information leakage of appliance status," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 9, pp. 2227–2241, Sep. 2017. 5
- [33] M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: Threats and solutions," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2019–2036, May 2014. 6
- [34] D. Irani, S. Webb, K. Li, and C. Pu, "Modeling unintended personal-information leakage from multiple online social networks," *IEEE Internet Comput.*, vol. 15, no. 3, pp. 13–19, May 2011. 6
- [35] Wikipedia. 2018 SingHealth data breach. [Online]. Available: https://en.wikipedia.org/wiki/2018_SingHealth_data_breach 6

- [36] P. Papadimitratos and Z. J. Haas, “Secure data communication in mobile ad hoc networks,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 2, pp. 343–356, Feb. 2006. [7](#)
- [37] R. L. Lagendijk, Z. Erkin, and M. Barni, “Encrypted signal processing for privacy protection: Conveying the utility of homomorphic encryption and multiparty computation,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 82–105, Jan. 2013. [7](#)
- [38] T. Plantard, W. Susilo, and Z. Zhang, “Fully homomorphic encryption using hidden ideal lattice,” *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 12, pp. 2127–2137, Dec. 2013. [7](#)
- [39] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, “Privacy aware learning,” *J. of ACM*, vol. 61, no. 6, Dec. 2014. [7](#)
- [40] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, “Adaptive federated learning in resource constrained edge computing systems,” *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, Jun. 2019. [8](#)
- [41] P. Samarati, “Protecting respondents identities in microdata release,” *IEEE Trans. Knowl. Data Eng.*, vol. 13, no. 6, pp. 1010–1027, Dec. 2001. [8](#)
- [42] D. Rebollo-Monedero, J. Forné, and J. Domingo-Ferrer, “From t-closeness-like privacy to postrandomization via information theory,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 11, pp. 1623–1636, Nov. 2010. [8](#), [10](#)
- [43] F. P. Calmon and N. Fawaz, “Privacy against statistical inference,” in *Proc. Allerton Conf. on Commun., Control and Computing*, Monticello, IL, Oct. 2012. [8](#), [10](#), [97](#), [106](#)
- [44] M. Sun and W. P. Tay, “Inference and data privacy in IoT networks,” in *Proc. IEEE Workshop on Signal Proc. Advances in Wireless Commun.*, Hokkaido, Japan, Jul. 2017. [8](#)

- [45] M. Sun, W. P. Tay, and X. He, “Toward information privacy for the Internet of things: A nonparametric learning approach,” *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1734–1747, Apr. 2018. [8](#), [12](#), [13](#)
- [46] M. Sun and W. P. Tay, “Decentralized detection with robust information privacy protection,” *IEEE Trans. Inf. Forensics Security*, vol. 15, no. 1, pp. 85–99, Nov. 2020. [8](#), [12](#)
- [47] A. Makhdoumi and N. Fawaz, “Privacy-Utility tradeoff under statistical uncertainty,” in *Proc. Allerton Conf. on Commun., Control and Computing*, Monticello, IL, USA, Oct. 2013. [9](#), [94](#), [97](#)
- [48] F. P. Calmon, M. Varia, M. Médard, M. M. Christiansen, K. R. Duffy, and S. Tessaro, “Bounds on inference,” in *Proc. Allerton Conf. on Commun., Control and Computing*, Monticello, IL, USA, Oct. 2013. [9](#), [94](#), [97](#)
- [49] C. T. Li and A. E. Gamal, “Maximal correlation secrecy,” *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3916–3926, May 2018. [9](#), [94](#)
- [50] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, “Information extraction under privacy constraints,” *Information*, vol. 7, no. 1, Mar. 2016. [9](#), [94](#), [95](#)
- [51] F. Farokhi and H. Sandberg, “Optimal privacy preserving policy using constrained additive noise to minimize the Fisher information,” in *Proc. IEEE Conf. on Decision and Control*, Melbourne, Australia, Dec. 2017. [9](#)
- [52] —, “Fisher information as a measure of privacy: Preserving privacy of households with smart meters using batteries,” *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 4726–4734, Sep. 2018. [9](#)
- [53] —, “Ensuring privacy with constrained additive noise by minimizing Fisher information,” *Automatica*, vol. 99, pp. 275–288, 2019. [9](#)
- [54] F. Farokhi, “Privacy-Preserving constrained quadratic optimization with fisher information,” *IEEE Signal Process. Lett.*, vol. 27, pp. 545–549, Mar. 2020. [9](#)

- [55] W. Wang, L. Ying, and J. Zhang, “On the relation between identifiability, differential privacy, and mutual-information privacy,” *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5018–5029, Sep. 2016. [9](#)
- [56] M. Sun and W. P. Tay, “On the relationship between inference and data privacy in decentralized IoT networks,” *IEEE Trans. Inf. Forensics Security*, vol. 15, no. 1, pp. 852–866, Dec. 2020. [9](#)
- [57] S. Asoodeh, F. Alajaji, and T. Linder, “On maximal correlation, mutual information and data privacy,” in *Canadian Workshop on Inform. Theory*, St. John’s, NL, Canada, Jul. 2015, pp. 27–31. [9](#), [94](#), [108](#)
- [58] I. Issa, A. B. Wagner, and S. Kamath, “An operational approach to information leakage,” *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1625–1657, Mar. 2020. [9](#), [94](#)
- [59] T. A. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: John Wiley & Sons, Inc., 2005. [10](#), [11](#), [17](#), [95](#)
- [60] S. Y. Kung, “Compressive privacy: From information/estimation theory to machine learning,” *IEEE Signal Process. Mag.*, vol. 34, no. 1, pp. 94–112, Jan. 2017. [10](#), [13](#), [30](#), [62](#)
- [61] ———, “A compressive privacy approach to generalized information bottleneck and privacy funnel problems,” *J. the Franklin Institute*, vol. 355, no. 4, pp. 1846–1872, Jul 2017. [10](#)
- [62] M. Al, S. Wan, and S. Kung, “Ratio utility and cost analysis for privacy preserving subspace projection,” *arXiv preprint arXiv:1702.07976*, 2017. [10](#)
- [63] A. Emad and O. Milenkovic, “Compression of noisy signals with information bottlenecks,” in *Proc. IEEE Inform. Theory Workshop (ITW)*, Sevilla, Spain, Sep. 2013. [10](#)
- [64] K. Diamantaras and S. Kung, “Data privacy protection by kernel subspace projection and generalized eigenvalue decomposition,” in *IEEE Int. Workshop*

on Machine Learning for Signal Processing, Vietri sul Mare, Italy, Sep. 2016.

[10](#)

- [65] J. C. Liao, L. Sankar, V. Y. F. Tan, and F. P. Calmon, “Hypothesis testing under mutual information privacy constraints in the high privacy regime,” *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 4, pp. 1058–1071, Dec. 2017. [10](#), [12](#)
- [66] A. Makhdoumi, S. Salamatian, N. Fawaz, and M. Médard, “From the information bottleneck to the privacy funnel,” in *Proc. IEEE Inform. Theory Workshop*, Hobart, TAS, Australia, Nov. 2014. [11](#), [97](#), [99](#)
- [67] T. Naftali, F. C. Pereira, and B. William, “The information bottleneck method,” in *Proc. Allerton Conf. on Commun., Control and Computing*, Monticello, Illinois, Sep. 1999. [11](#)
- [68] Y. O. Basciftci, Y. Wang, and P. Ishwar, “On privacy-utility tradeoffs for constrained data release mechanisms,” in *Inform. Theory and Applications Workshop*, La Jolla, CA, USA, Feb. 2016. [11](#)
- [69] B. Moraffah and L. Sankar, “Privacy-Guaranteed two-agent interactions using information-theoretic mechanisms,” *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 9, pp. 2168–2183, May 2017. [11](#)
- [70] H. Wang and F. P. Calmon, “An estimation-theoretic view of privacy,” in *Proc. Allerton Conf. on Commun., Control and Computing*, Monticello, IL, USA, Oct. 2017. [11](#), [97](#)
- [71] H. Wang, L. Vo, F. P. Calmon, M. Médard, K. R. Duffy, and M. Varia, “Privacy with estimation guarantees,” *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8025–8042, Aug. 2019. [11](#), [97](#)
- [72] F. P. Calmon, A. Makhdoumi, M. Médard, M. Varia, M. Christiansen, and K. R. Duffy, “Principal inertia components and applications,” *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5011–5038, Aug. 2017. [11](#)

- [73] S. Asoodeh, F. Alajaji, and T. Linder, “Privacy-Aware MMSE estimation,” in *Proc. IEEE Int. Symp. on Inform. Theory*, Barcelona, Spain, Jul. 2016. [11](#)
- [74] S. Asoodeh, M. Diaz, F. Alajaji, and T. Linder, “Estimation efficiency under privacy constraints,” *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1512–1534, Mar. 2019. [11](#)
- [75] F. P. Calmon, A. Makhdoumi, and M. Médard, “Fundamental limits of perfect privacy,” in *Proc. IEEE Int. Symp. on Inform. Theory*, Hong Kong, China, Jun. 2015. [11](#), [140](#)
- [76] B. Rassouli and D. Gündüz, “On perfect privacy,” *IEEE J. Sel. Areas Inf. Theory*, pp. 1–1, Jan. 2021. [11](#), [94](#), [96](#), [140](#)
- [77] B. Rassouli, F. Rosas, and D. Gündüz, “Latent feature disclosure under perfect sample privacy,” in *Proc. IEEE Int. Workshop on Inf. Forensics and Security*, Hong Kong, Dec. 2018. [11](#), [140](#)
- [78] B. Rassouli, F. E. Rosas, and D. Gündüz, “Data disclosure under perfect sample privacy,” *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2012–2025, Nov. 2019. [11](#), [96](#), [140](#)
- [79] R. Tandon, L. Sankar, and H. V. Poor, “Discriminatory lossy source coding: Side information privacy,” *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5665–5677, 2013. [11](#)
- [80] C. X. Wang, Y. Song, and W. P. Tay, “Preserving parameter privacy in sensor networks,” in *Proc. IEEE Global Conf. on Signal and Information Processing*, Anaheim, USA, Nov. 2018. [11](#)
- [81] —, “Arbitrarily strong utility-privacy tradeoff in multi-agent systems,” *IEEE Trans. Inf. Forensics Security*, pp. 1–1, 2020. [11](#)

- [82] Y. Song, C. X. Wang, and W. P. Tay, “Privacy-Aware Kalman filtering,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Calgary, Canada, Apr. 2018. [11](#)
- [83] —, “Compressive privacy for a linear dynamical system,” *IEEE Trans. Inf. Forensics Security*, vol. 15, no. 1, pp. 895–910, Dec. 2020. [11](#)
- [84] C. Murguia, I. Shames, F. Farokhi, D. Nesic, and V. Poor, “On privacy of dynamical systems: An optimal probabilistic mapping approach,” *arXiv preprint arXiv:1910.13559*, Jan. 2021. [11](#)
- [85] Y. Liu, A. Khisti, and A. Mahajan, “On privacy in smart metering systems with periodically time-varying input distribution,” in *Proc. IEEE Global Conf. on Signal and Information Processing*, Montreal, Canada, Nov. 2017. [11](#)
- [86] Y. H. Liu, S. H. Lee, and A. Khisti, “Information-Theoretic privacy in smart metering systems using cascaded rechargeable batteries,” *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 314–318, Mar. 2017. [11](#)
- [87] S. Li, A. Khisti, and A. Mahajan, “Information-Theoretic privacy for smart metering systems with a rechargeable battery,” *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3679–3695, May 2018. [11](#), [30](#)
- [88] G. Giacon, D. Gündüz, and H. V. Poor, “Smart meter privacy with an energy harvesting device and instantaneous power constraints,” in *Proc. IEEE Int. Conf. on Commun.*, London, UK, Jun. 2015. [11](#)
- [89] A. Gilani, S. B. Amor, S. Salehkalaibar, and V. T. F. Tan, “Distributed hypothesis testing with privacy constraints,” *Entropy*, vol. 21, no. 5, p. 478, May 2019. [12](#)
- [90] S. Sreekumar, D. Gündüz, and A. Cohen, “Distributed hypothesis testing under privacy constraints,” in *Inform. Theory Workshop*, Guangzhou, China, Nov. 2018. [12](#)

- [91] S. Sreekumar, A. Cohen, and D. Gündüz, “Privacy-aware distributed hypothesis testing,” *Entropy*, vol. 22, no. 6, Jun. 2020. [12](#)
- [92] S. Salehkalaibar, F. Aminifar, and M. Shahidehpour, “Hypothesis testing for privacy of smart meters with side information,” *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2059–2067, Dec. 2017. [12](#)
- [93] Z. Li, T. J. Oechtering, and D. Gündüz, “Privacy against a hypothesis testing adversary,” *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1567–1581, Nov. 2018. [12](#)
- [94] Z. Li and T. J. Oechtering, “Privacy-Aware distributed bayesian detection,” *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1345–1357, Oct. 2015. [12](#)
- [95] ———, “Privacy-Constrained parallel distributed Neyman-Pearson test,” *IEEE Trans. Signal and Inf. Process. over Networks*, vol. 3, no. 1, pp. 77–90, Mar. 2017. [12](#)
- [96] M. Sun and W. P. Tay, “Privacy-Preserving nonparametric decentralized detection,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016. [12](#)
- [97] T. Chanyaswad, J. M. Chang, and S. Y. Kung, “A compressive multi-kernel method for privacy preserving machine learning,” in *Proc. Int. Joint Conf. Neural Networks*, Anchorage, Alaska, May 2017. [13](#)
- [98] X. He, W. P. Tay, L. Huang, M. Sun, and Y. Gong, “Privacy-Aware sensor network via multilayer nonlinear processing,” *IEEE Internet Things J.*, vol. 6, pp. 10 834–10 845, Dec. 2019. [13](#)
- [99] C. Louizos, K. Swersky, Y. J. Li, M. Welling, and R. Zemel, “The variational fair autoencoder,” in *Proc. Int. Conf. on Learning Representations*, San Juan, Puerto Rico, May 2016. [13](#), [125](#)

- [100] D. Moyer, S. Y. Gao, R. Brekelmans, G. V. Steeg, and A. Galstyan, “Invariant representations without adversarial training,” in *Proc. Int. Conf. on Neural Information Processing Systems*, Montreal, Canada, Dec. 2018. [13](#), [125](#)
- [101] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork, “Learning fair representations,” in *Proc. Int. Conf. on Machine Learning*, Atlanta, USA, Jun. 2013. [13](#), [128](#)
- [102] I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Proc. Int. Conf. on Neural Information Processing Systems*, Montreal, Canada, Dec. 2014. [13](#), [123](#)
- [103] C. Huang, P. Kairouz, and L. Sankar, “Generative adversarial privacy: A data-driven approach to information-theoretic privacy,” in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA, USA, Oct. 2018. [13](#), [125](#)
- [104] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, “Context-Aware generative adversarial privacy,” *Entropy*, vol. 19, no. 12, Dec. 2017. [13](#)
- [105] A. Tripathy, Y. Wang, and P. Ishwar, “Privacy-Preserving adversarial networks,” in *Proc. Allerton Conf. on Commun., Control and Computing*, Monticello, IL, USA, Sep. 2019. [13](#)
- [106] B. O. Tseng and P. Y. Wu, “Compressive privacy generative adversarial network,” *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2499–2513, Jan. 2020. [13](#)
- [107] Z. F. Zhang, Y. Song, and H. R. Qi, “Age progression/regression by conditional adversarial autoencoder,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Hawaii, United States, Jul. 2017. [16](#), [124](#)

- [108] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *Annals of Statistics*, vol. 32, pp. 407–499, Apr. 2004. [16](#)
- [109] R. V. Hogg and A. T. Craig, *Introduction to Mathematical Statistics*, 4th ed. New York, NY: Macmillan Publishing Co., Inc., 1978. [17](#)
- [110] H. M. P. Zarchan, *Fundamentals of Kalman Filtering: A Practical Approach*, 4th ed. American Institute of Aeronautics and Astronautics, 2013. [17](#)
- [111] J. B. Conway, *A Course in Functional Analysis*, 2nd ed. New York, NY: Springer-Verlag, 1990. [17](#), [110](#)
- [112] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *J. of Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar 1960. [19](#), [20](#), [61](#)
- [113] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT press, 2001. [22](#)
- [114] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ, USA: Princeton Univ. Press, 1945. [28](#)
- [115] P. J. Bickel and K. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, 1st ed. Oakland, CA: Holden-Day, Inc., 1977. [28](#)
- [116] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004. [46](#), [49](#), [73](#), [87](#)
- [117] H. Jiang, T. Kathuria, Y. T. Lee, S. Padmanabhan, and Z. Song, “A faster interior point method for semidefinite programming,” in *Proc. Conf. Found. Comput. Sci.*, Durham, NC, USA, Nov. 2020. [49](#)
- [118] D. Reid, “An algorithm for tracking multiple targets,” *IEEE Trans. Autom. Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979. [60](#)
- [119] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004. [60](#)

- [120] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, “Kalman filtering with intermittent observations,” *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1453–1464, Sep. 2004. [61](#)
- [121] M. C. F. Donkers, W. P. M. H. Heemels, N. V. D. Wouw, and L. Hetel, “Stability analysis of networked control systems using a switched linear systems approach,” *IEEE Trans. Autom. Control*, vol. 56, no. 9, pp. 2101–2115, Sep. 2011. [61](#)
- [122] L. Shi, M. Epstein, and R. M. Murray, “Kalman filtering over a packet-dropping network: A probabilistic perspective,” *IEEE Trans. Autom. Control*, vol. 55, no. 3, pp. 594–604, Mar. 2010. [61](#)
- [123] S. Kar and J. M. F. Moura, “Gossip and distributed Kalman filtering: Weak consensus under weak detectability,” *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1766–1784, Apr. 2011. [61](#)
- [124] D. E. Quevedo, A. Ahlén, and K. H. Johansson, “State estimation over sensor networks with correlated wireless fading channels,” *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 581–593, Mar. 2013. [61](#)
- [125] E. R. Rohr, D. Marelli, and M. Fu, “Kalman filtering with intermittent observations: On the boundedness of the expected error covariance,” *IEEE Trans. Autom. Control*, vol. 59, no. 10, pp. 2724–2738, Oct. 2014. [61](#)
- [126] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer, “Image similarity using mutual information of regions,” in *Proc. European Conf. Computer Vision*, Prague, Czech Republic, May 2004. [95](#)
- [127] H. B. Mitchell, “Image similarity measures,” in *Image Fusion*. Berlin, Heidelberg: Springer, 2010. [95](#)
- [128] J. Johnson, A. Alahi, and F. F. Li, “Perceptual losses for real-time style transfer and super-resolution,” in *Proc. European Conf. on Computer Vision*, Amsterdam, The Netherlands, Oct. 2016. [95](#), [122](#)

- [129] B. Rassouli and D. Gündüz, “Optimal utility-privacy trade-off with total variation distance as a privacy measure,” *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 594–603, Dec. 2020. [96](#)
- [130] D. P. Kingma and M. Welling, “Auto-Encoding variational bayes,” in *Proc. Int. Conf. on Learning Representations*, Banff, Canada, Apr. 2014. [98](#), [99](#), [100](#), [106](#)
- [131] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Society for Industrial and Applied Mathematics, 1990. [99](#)
- [132] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schoelkopf, and A. J. Smola, “A kernel method for the two-sample-problem,” *Advances in neural information processing systems*, pp. 513–520, 2006. [102](#)
- [133] J. Y. Zhu, T. Park, P. Lsola, and A. A. Efros, “Unpaired image-to-image translation using adversarial consistency loss,” in *Proc. IEEE Int. Conf. on Computer Vision*, Venice, Italy, Oct. 2017. [102](#), [123](#)
- [134] Z. Murez, S. Kolouri, D. Kriegman, R. Ramamoorthi, and K. Kim, “Image to image translation for domain adaptation,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, Jun. 2018. [102](#)
- [135] A. Müller, “Integral probability metrics and their generating classes of functions,” *Advances in Applied Probability*, vol. 29, no. 2, pp. 429–443, Jun. 1997. [103](#)
- [136] B. K. Sriperumbudur, K. Fukumizu, and G. R. G. Lanckriet, “Universality, characteristic kernels and RKHS embedding of measures,” *J. Machine Learning Research*, vol. 12, no. 70, pp. 2389–2410, Jul. 2011. [103](#)
- [137] J. H. Manton and P. O. Amblard, “A primer on reproducing kernel Hilbert spaces,” *Foundations and Trends® in Signal Processing*, vol. 8, no. 1-2, pp. 1–126, Dec. 2015. [103](#)

- [138] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, “A kernel two-sample test,” *J. Machine Learning Research*, vol. 13, no. 25, pp. 723–773, Mar. 2012. [103](#)
- [139] C. L. Li, W. C. Chang, Y. Cheng, Y. M. Yang, and B. Poczos, “MMD GAN: Towards deeper understanding of moment matching network,” in *Proc. Int. Conf. on Neural Inform. Processing Systems*, Long Beach, CA, USA, Dec. 2017. [104](#)
- [140] Y. Li, K. Swersk, and R. Zeme, “Generative moment matching networks,” in *Proc. Int. Conf. on Machine Learning*, Lille, France, Jul. 2015. [104](#)
- [141] H. O. Hirschfeld, “A connection between correlation and contingency,” in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31. Cambridge University Press, 1935, pp. 520–524. [109](#)
- [142] H. Gebelein, “Das statistische problem der korrelation als variations-und eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung,” *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 21, no. 6, pp. 364–379, 1941. [109](#)
- [143] A. Rényi, “On measures of dependence,” *Acta Math. Hung.*, vol. 10, pp. 441–451, Sep. 1959. [109](#), [110](#)
- [144] S. Huang and L. Zheng, “Linear information coupling problems,” in *Proc. IEEE Int. Symp. on Inform. Theory*, Cambridge, MA, USA, Jul. 2012. [109](#)
- [145] R. I. Jennrich, “Asymptotic properties of non-linear least squares estimators,” *Ann. Math. Statist.*, vol. 40, no. 2, pp. 633–643, Apr. 1969. [115](#), [135](#)
- [146] K. Hornik, M. Tinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Networks*, vol. 2, no. 5, pp. 359–366, Jul. 1989. [116](#), [118](#)
- [147] W. K. Newey, “Uniform convergence in probability and stochastic equicontinuity,” *Econometrica*, vol. 59, no. 4, pp. 1161–1167, Jul. 1991. [117](#)

- [148] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, USA, Jun. 2015. [121](#)
- [149] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image translation with conditional adversarial networks,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, United States, Jul. 2017. [123](#)
- [150] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, Jun. 2016. [124](#)
- [151] D. Dua and C. Graff. (2019) UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. [Online]. Available: <http://archive.ics.uci.edu/ml> [128](#)
- [152] S. Y. Kung, T. Chanyaswad, J. M. Chang, and P. Wu, “Collaborative PCA/DCA learning methods for compressive privacy,” *ACM Trans. Embed. Comput. Syst.*, vol. 16, no. 3, Jul. 2017. [129](#)
- [153] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, “Canonical correlation analysis: An overview with application to learning methods,” *Neural computation*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004. [135](#)
- [154] F. R. Bach and M. I. Jordan, “Kernel independent component analysis,” *J. Machine Learning Research*, vol. 3, pp. 1–48, Jul. 2002. [136](#), [137](#)
- [155] B. Schölkopf, A. Smola, and K. R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998. [136](#)
- [156] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, “Wireless networks with RF energy harvesting: A contemporary survey,” *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 757–789, May 2015. [140](#)

-
- [157] S. S. Gupta and N. B. Mehta, “Revisiting effectiveness of energy conserving opportunistic transmission schemes in energy harvesting wireless sensor networks,” *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2968–2980, Apr. 2019. [140](#)
- [158] E. Erdemir, P. L. Dragotti, and D. Gündüz, “Privacy-Aware time-series data sharing with deep reinforcement learning,” *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 389–401, Jul. 2020. [141](#)
- [159] R. Pinot, A. Morvan, F. Yger, C. G. Pailler, and J. Atif, “Graph-Based clustering under differential privacy,” in *Proc. ACM Conf. Uncertainty in Artificial Intelligence*, Monterey, California, USA, Aug. 2018. [141](#)
- [160] S. Sajadmanesh and D. G. Perez, “Locally private graph neural networks,” *arXiv arXiv:2006.05535v6*, Jan. 2021. [141](#)