Goal oriented communications : the quantization problem

Hang Zou

université
PARIS-SACLAY

Communications Orientées Objectifs : Le Problème de Quantification

Goal Oriented Communications : The Quantization Problem

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 580 Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat : Réseaux, Information et Communications
Graduate school : Informatique et science du numérique
Référent : CentraleSupélec, Université Paris-Saclay

Thèse préparée dans Laboratoire des Signaux et Systèmes, **CentraleSupélec** sous la direction de **Samson LASAULCE**, directeur de recherche, centre de Recherche en Automatique de Nancy, CNRS, France.

Thèse soutenue à GIF-SUR-YVETTE, le 06 avril 2022, par

Hang ZOU

Composition du jury

Michèle Wigger Professeur, Télécom Paris, France	Présidente
Walid Saad Professeur, Virginia Tech, USA	Rapporteur & Examinateur
Marios Kountouris Professeur, Eurecom, Sophia Antipolis, France	Rapporteur & Examinateur
Mehdi Bennis Professeur, University of Oulu, Finland	Examinateur
Sorin Olaru Professeur, CentraleSupélec, France	Examinateur
Mohamad Assaad Professeur, CentraleSupélec, France	Examinateur
Lucas Saludjian Ingénieur de recherche, RTE, France	Examinateur
Samson Lasaulce Directeur de recherche, CNRS, France	Directeur de thèse

THESE DE DOCTORAT
NNT : 2022UPASG021

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my advisor Prof. Samson Lasaulce for the continuous support of my Ph.D study and related researches. His patience, motivation, and immense knowledge always attracts me deeply since our first meet in his optimization course in the second year of my master. His guidance helped me in all the time of research and writing of this thesis.

In addition, I would like to thank Dr. Chao Zhang who played the role of my co-advisor. When I encountered various difficulties, he always came up with feasible solutions. I hope I can maintain close cooperation with him in the future.

I also would like to thank Prof. Walid Saad and Prof. Marios Kountouris to be my reviewers. Same gratitude should be dedicated to my examiners as well : Prof. Mehdi Bennis, Prof. Michèle Wigger, Prof. Sorin Olaru, Prof. Mohamad Assaad and Dr. Lucas Saludjian for their insightful comments and encouragement on my research works.

Thanks for RTE Chairs for their continuous support for me and the funding for my PhD.

My sincere appreciation also goes to all my friends and colleagues from CentraleSupelec for various interesting discussions we had during the last three years : Dr. Zhenyu Liao, Dr. Jian Song, Dr. Xiaojun Xi, Dr. Weichao Liang, Dr. Chen Kang and Dr. Xuewen Qian.

All my Chinese friends also relieved me a lot : Dr. Bozhang Dong, Dr. Ming Chen, Dr. Yang Xiao, Miaobing Chen, Yifan Ding, Zheng Zhou, Xiao Li. They have always encouraged me to support me from a variety of different ideas. Hope our friendship will last forever.

Last, but not the least important, I would like to thank my parents Xiaoyan Fan and Huaqun Zou. They are all very experienced primary and secondary school teachers. The habits they cultivated since my childhood support me all the way to where I am today. If it were not for their encouragement and encouragement, it would be difficult for me to decide on the successful path of studying in France.

For the future, I will always remember the first one of the Delphi maxims : “Know Thyself” to keep self motivated.

Table des matières

Table des figures

Liste des tableaux

1

French Summary

Le paradigme classique pour concevoir un émetteur (codeur) et un récepteur (décodeur) est de concevoir ces éléments en assurant que l'information reconstruite par le récepteur soit suffisamment proche de l'information que l'émetteur a mis en forme pour l'envoyer sur le médium de communication ; on parle de critère de fidélité ou de qualité de reconstruction (mesurée par exemple en termes de distorsion, de taux d'erreur binaire, de taux d'erreur paquet ou de probabilité de coupure de la communication). Le problème du paradigme classique est qu'il peut conduire à un investissement injustifié en termes de ressources de communication (surdimensionnement de l'espace de stockage de données, médium de communication à très haut débit et onéreux, composants très rapides, etc.) et même à rendre les échanges plus vulnérables aux attaques. La raison à cela est que l'exploitation de l'approche classique (fondée sur le critère de fidélité de l'information) dans les réseaux sans fil conduira typiquement à des échanges excessivement riches en information, trop riches au regard de la décision que devra prendre le destinataire de l'information ; dans le cas plus simple, cette décision peut même être binaire, indiquant qu'en théorie un seul bit d'information pourrait suffire. Il s'avère qu'actuellement, l'ingénieur n'a pas à sa disposition une méthodologie lui permettant de concevoir une telle paire émetteur-récepteur qui serait adaptée à l'utilisation (ou les utilisations) du destinataire.

Une première étape consistera à obtenir la structure optimale de l'étage de conversion analogique-numérique (incluant typiquement la quantification et l'échantillonnage), optimale au sens d'un critère de performance propre à l'organe de prise de décision ciblé. Dans un deuxième temps, nous étudierons le scénario où plusieurs objectifs sont visés par le récepteur, soit simultanément (par exemple avoir une estimation suffisamment bonne de la grandeur mesurée par l'émetteur tout en garantissant un niveau de confidentialité), soit séquentiellement (garantir une confidentialité forte en mode de marche normale puis garantir une réactivité maximale du récepteur en mode défaut). Une troisième étape consiste en l'étude du cas distribué, c'est-à-dire lorsqu'il y a plusieurs émetteurs qui mesurent chacun une partie de l'information dont a besoin le récepteur. L'étude de ce cas permettra par exemple de savoir dans quelle mesure des erreurs de synchronisation entre

les capteurs de mesures affecte la décision du récepteur. Une dernière étape clairement identifiée à ce jour sera d'étudier le cas où les fonctions qui représentent les critères de performances du récepteur ne sont pas connus. Seules les réalisations de ces fonctions seront supposées connues.

Par conséquent, un nouveau paradigme de communication appelé la communication orientée objectif est proposé pour résoudre le problème des communications classiques. Le but ultime des communications orientées objectifs est d'accomplir certaines tâches ou certains objectifs au lieu de viser un critère de reconstruction du signal source. Les tâches sont généralement caractérisées par des fonctions d'utilité ou des fonctions de coût à optimiser.

Dans le chapitre 3, le problème de quantification orientée objectif est formellement formulé et comparé à la quantification conventionnelle. Tout d'abord, l'algorithme basique de quantification orientée objectif imitant le célèbre algorithme Lloyd-Max est proposé et divise le puzzle original en deux sous-problèmes étant relativement plus faciles à résoudre pour le scénario où l'on dispose d'une parfaite connaissance de la fonction d'utilité : trouver un ensemble de décision optimal pour des régions de quantification données ; trouver des régions de quantification pour un ensemble de décision donné.

Ensuite, le problème de quantification orientée objectif est abordé pour le cas particulier où l'espace de décision de la fonction de coût est polyédrique et convexe. Supposant que la fonction d'utilité est concave par rapport à la variable de décision, on peut obtenir une borne supérieure de la perte d'optimalité empirique en appliquant l'inégalité de Jensen. De plus, pour un ensemble d'échantillons de paramètres donné, les ensembles de décisions sont divisés en classes équivalentes en fonction de l'étiquette de décision optimale. Un algorithme qui améliore itérativement l'ensemble de décisions de manière avide au sein d'une classe équivalente est proposé pour trouver le meilleur ensemble de décisions qui minimise la borne supérieure de la perte d'optimalité. Ci-après, nous introduisons l'inégalité de Jensen généralisée pour les fonctions d'utilité dites faiblement concaves. En remplaçant l'inégalité de Jensen par la généralisée, une méthode analogue pourrait également être appliquée aux fonctions d'utilité faiblement concaves. Une version améliorée de cet algorithme est proposée en étendant l'ensemble de décision de la manière que nous avons décrite avant. L'avantage de l'algorithme proposé est double : la connaissance complète de la fonction de décision optimale n'est pas nécessaire pour la mise en œuvre de l'algorithme ; au lieu de résoudre le problème d'optimisation compliqué d'origine pour le quantificateur orienté objectif, seul des basique computations matricielles et la comparaison répétée de la fonction de l'utilité sont nécessaires pour trouver l'ensemble de décision souhaité. Nous appliquons l'algorithme proposé à la fonction de capacité somme-taux qui est bien connue pour être convexe par rapport à la puissance. Les résultats numériques montrent que l'algorithme proposé surpasse l'approche conventionnelle. En attendant, il est important pour souligner que la méthode proposée pourrait être redondante si l'ensemble de décision optimal se situe sur les sommets du polyèdre de décision, par exemple, le contrôle de puissance binaire.

Enfin, nous essayons de résoudre le problème de quantification orientée objectif lorsque seules les réalisations des fonctions d'utilité sont supposées connues. Le problème est divisé en deux étapes. Pour trouver les régions de quantification fixant l'ensemble de décision, un réseau de neurones à propagation avant est appliquée au problème de l'allocation de puissance, la quantification très grossière des gains du canal n'induit qu'une

très faible perte d'optimalité par rapport au cas où les gains sont parfaitement connus de l'émetteur lorsque l'utilité est le débit de transmission. Cependant, pour la fonction d'efficacité énergétique, les gains de canal doivent être quantifiés plus précisément. L'utilisation d'un schéma de quantification classique basé sur la distorsion (quantification k-moyennes) conduit à une perte de performance assez importante (environ 30%), montrant le potentiel de notre approche. En outre, cela permet de reconsidérer l'hypothèse globale faite dans les problèmes d'allocation des ressources, à savoir que la politique d'allocation des ressources est conçue en supposant une connaissance parfaite. Mathématiquement, des recherches supplémentaires devraient être développées pour identifier les propriétés de la fonction d'utilité qui représente sa sensibilité à être maximisée sous une connaissance imparfaite de ses paramètres. Pour trouver un ensemble de décision optimal, un algorithme évolutif appelé Invasive Weeds Optimization - Differential Evolution (IWO-DE) qui combine deux algorithmes évolutionnaires classiques est proposé. Un problème de recherche conjointe d'un ensemble de vecteurs de niveau de puissance et de formation de faisceaux pour des communications efficaces en énergie est pris comme exemple de notre méthode proposée. Alors que la version continue du problème pourrait être résolue facilement, le problème doit être formulé correctement lorsque l'ensemble de décision est imposé pour être fini. Notre approche est montrée pour surpasser les techniques de pointe telles que l'algorithme Lloyd-Max et la quantification vectorielle aléatoire. Lorsque les dimensions du système augmentent, des problèmes de complexité doivent être pris en compte. Lorsqu'il y a interférence, le cadre proposé doit être étendu. D'autres problèmes tels que le paradoxe de Braess peuvent survenir et rendre le problème encore plus difficile. Enfin, le codage de source orienté objectif et le codage de canal orienté objectif restent l'extension difficile du codage actuel scénario.

Dans le chapitre 4, nous analysons le problème de quantification orientée objectif en régime haute résolution. Le cas scalaire et le cas vectoriel sont traités séparément. Pour le cas scalaire, la nouvelle formule approximative proposée de la perte d'optimalité conduit à une nouvelle qualité définie comme la densité de valeurs représentant l'importance du paramètre. Nous introduisons une nouvelle qualité appelée perte d'optimalité normalisée lors de la comparaison de la dureté de la quantification pour différentes fonctions de coût. En rapprochant simplement cette qualité en régime haute résolution, nous sommes capables de déterminer la dureté de la quantification pour différentes fonctions de coût sans effectuer de véritables simulations. Pour le cas vectoriel, une formule approximative indépendante de la cellule n'est plus disponible pour la perte d'optimalité puisque la forme optimale des cellules de pavage est inconnue. Néanmoins, en admettant la conjecture de Gersho, une borne supérieure et une borne inférieure sont dérivées pour la perte d'optimalité lorsque la dimension du paramètre est plus petite que la dimension de la variable de décision. De plus, nous proposons un nouvel algorithme qui met à jour itérativement les représentants en utilisant l'approximation des valeurs propres sur la perte d'optimalié. A chaque itération, on essaie de trouver le pire échantillon de paramètres dans le sens d'introduire la plus grande perte d'optimalité individuelle. Ensuite, son représentant correspondant est révisé de sorte que la perte d'optimalité moyenne pourrait probablement être diminuée. L'algorithme proposé pourrait également être étendu à la fonction de coût avec des contraintes. Les résultats de la simulation montrent que le quantificateur orienté objectif proposé surpasse largement le quantificateur Lloyd-Max pour un nombre quelconque de bits de quantification. L'algorithme avec mise à jour gourmande domine légèrement celui avec mise à jour satisfaisante tandis que le dernier prend beaucoup moins de temps à fonctionner. Cependant, étendre notre méthode proposée au cas vectoriel reste

fastidieux pour des scénarios de grande dimension, par exemple des images ou des vidéos. En outre, la sélection de l'ensemble d'échantillons de paramètres pourrait également être essentielle en raison à la fois des échantillons eux-mêmes et du nombre d'échantillons de paramètres qui ont un impact sur la complexité de l'algorithme proposé.

Dans le chapitre 5, au lieu de se concentrer sur la structure optimale de la communication orientée objectif, nous commençons à aborder le problème de quantification orientée objectif lorsque plusieurs fonctions d'utilité corrélatives sont ciblées par différents utilisateurs du système. En d'autres termes, le problème de quantification orientée objectif est développé dans le cadre des jeux. Plus précisément, nous nous limitons à l'étude des jeux potentiels avec l'ensemble des actions identique. En prenant le bien-être social comme critère de performance, nous avons prouvé que le bien-être social optimal est une fonction sous-modulaire de l'ensemble d'action avec l'équilibre de Nash raffiné dans l'ensemble argmax du potentiel. De plus, le fameux paradoxe de Braess est lié à la monotonie de cette fonction. Sur la base de ces propriétés, nous concevons un algorithme pour trouver un ensemble d'actions fini visant à maximiser le bien-être social moyen sous l'équilibre de Nash du système. Nous prenons le jeu de canaux d'accès multiples MIMO multi-utilisateurs avec efficacité spectrale comme l'utilité individuelle dans laquelle l'existence du paradoxe de Braess est déjà confirmée comme l'application de notre théorie. Pour le scénario à 2 utilisateurs et 2 bandes, nous avons prouvé que l'ensemble de sélection de canaux est l'ensemble d'actions optimal pour maximiser le bien-être social. L'ensemble de type Telatar est supposé être l'ensemble d'action optimal pour maximiser le bien-être social sous l'équilibre de Nash dans les cas généraux. L'existence du paradoxe de Braess n'est pas garantie pour la fonction d'utilité générale du jeu arbitraire. La faisabilité des méthodes proposées doit être vérifiée pour d'autres applications.

Dans le chapitre 6, un jeu où la fonction d'utilité individuelle est l'efficacité énergétique dans un système à canaux d'accès multiples MIMO est considéré. L'existence et l'unicité de l'équilibre de Nash est prouvée, et un algorithme exact et un algorithme sous-optimal sont proposés pour approcher le NE de ce jeu. Les résultats de simulation montrent que si le nombre d'antennes d'émission et le nombre d'antennes de réception sont les mêmes, les performances sous NE trouvées par les algorithmes proposés sont toujours meilleures qu'une politique d'allocation de puissance uniforme à la fois à l'intérieur et à l'extérieur de la plage couverte par la proposition principale du chapitre. Lorsque la condition pour les antennes n'est pas remplie, notre algorithme proposé déploie une meilleure réponse approximative ε qui ne conduira pas à un pur équilibre de Nash. Etonnamment, la solution trouvée par notre algorithme sous-optimal domine légèrement le NE exact du jeu. Cette observation montre que les performances de l'algorithme proposé sont acceptables alors qu'il est relativement facile à mettre en œuvre. La situation où chaque utilisateur est autorisé à choisir librement sa matrice de covariance simplement contrainte à la puissance maximale est le prolongement naturel de ce chapitre. Les résultats de ce chapitre doivent être considérés comme l'exploitation de base pour la quantification orientée but dans un jeu général ou un système décentralisé. La discrétisation de l'espace d'action va fortement influencer la détermination de NE puisqu'elle transforme la nature du jeu. Cela pourrait être le principal défi des futurs travaux.

2 Introduction

In this thesis, the main problem is to investigate the so-called goal-oriented quantization (GOQ) problem which is how to design an efficient quantizer when a specific goal is given. The goal could be minimizing some cost functions (or maximizing some utility functions).

2.1 Context of The Thesis

Since the groundbreaking seminal work [3] of Shannon on information theory and communication system, the fundamental problem of communication is defined as "that of reproducing at one point either exactly or approximately a message selected at another point". Shannon further argued that the semantic aspects of communication should be considered as irrelevant to the engineering problem. Researchers are working diligently on improving the accuracy of decoded (reconstructed) signal, i.e., to minimize the the distortion introduced during coding, compression or quantization. Meanwhile, most existing communication technologies are developed to maximize data-oriented performance metrics such as communication data rate, while ignoring the service/content/semantic-related information and the ultimate goal of the entire communication system.

However, if the communication system is assumed to be goal-oriented, i.e., achieve a specific goal, e.g., help users to make critical decisions, optimize some critical performance under limited resources, intuitive idea of naively increasing the accuracy of reconstructed signal could be wasteful or useless. Thus we refer this conventional methodology of designing communication as as goal-ignorant. Before our framework of the goal-oriented communication, there exist already several works on the goal-oriented communication in the sense of information and message, or in recent semantic communications : [5, 7, 8]. In those works, the authors addressed the problem of potential "misunderstanding" among parties involved in a communication, where the misunderstanding arises from lack of initial agreement on what protocol and/or language is being used in communication. No-

wadays, semantic aspect of communication foreseen by Shannon and Weaver in [4] is attracting more and more attentions for its potential applications in future 6G networks [6, 9, 10, 11, 12, 13]. Semantic communication goes beyond the common Shannon paradigm of guaranteeing the correct reception of each single transmitted bit, irrespective of the meaning conveyed by the transmitted bits. The idea is that, whenever communication occurs to convey meaning or to accomplish a goal, what really matters is the impact that the received bits have on the interpretation of the meaning intended by the transmitter or on the accomplishment of a common goal. Some recent research works are briefly introduced in a non-exhaustive way. In [12], it is shown semantics-empowered policies could reduce real-time construction and the cost of actuation errors in an autonomous system tasked with real-time source reconstruction for remote actuation. A novel stochastic model of semantics-native communication (SNC) for generic tasks is proposed in [10]. Simulation results reveal significant reduction of the semantic representation length without compromising communication reliability is possible. A deep learning-enabled semantic communication system for speech recognition by designing the transceiver as an end-to-end (E2E) system is studied in [6]. The simulation results demonstrate that our proposed system outperforms the traditional communication systems for character-error-rate and word-error-rate while and is robust to channel variations.

In contrast with this line of research works, we introduce a general framework for GOQ. The task or goal of the receiver is chosen to be modeled by a generic optimization problem (OP) which contains both decision variables and parameters. The goal function of the OP is generic function $f(\boldsymbol{x};\boldsymbol{g}) : \mathcal{X}\times\mathcal{G} \to \mathbb{R}$ with $\boldsymbol{x}$ being the **decision** (action) variable and $\boldsymbol{g}$ being the **parameter** variable, referred as the **utility function** (**cost function**). Proposed paradigm of GOQ is illustrated in Fig. The ultimate goal is to minimize the **optimality loss** (OL) introduced by our communication paradigm optimizing the utility function. To the best knowledge of authors, the goal-oriented conception is rarely considered in following highly connected domains : quantization, classification or compression. Here we list some of them in related works. Goal-oriented communication in our definition is first formulated and studied in [15]. To effectively determine a good goal-oriented quantizer in the vector case, some sufficient but reasonable sufficient conditions on the utility function are assumed (such as the decomposability assumption) and a suboptimal iterative algorithm is proposed to solve the problem of energy-efficient and spectral efficient power control problem. Significant gains can be obtained in terms of payoff especially when the number of quantization bits decreases. A data-driven goal-oriented quantization systems with scalar analog-to-digital converters (ADCs), which determine how to map an analog signal into its digital representation using deep learning tools in [34]. The performance of large scale inputs of goal-oriented quantizer is studied in [35]. It is demonstrated that the minimal achievable average mean squared error (MSE) in massive multiple-input multiple-out (MIMO) channel estimation can be approached by properly designed quantization systems utilizing scalar low-resolution ADCs, and that the proposed approach outperforms previous channel estimators operating only in the digital domain. In [36], the problem of compressing band-limited graph signals into a finite-length sequence of bits by joint sampling and quantization is considered. The graph signal compression is shown to be similar with task-based quantization. Other applications using the conceptions but not the goal-oriented quantizer directly. In [37], a collaborative task is assigned to a multi-agent system in which agents are allowed to communicate however under limited communication rate. this problem is equivalent to a form of rate-distortion problem called the task-based information compression.

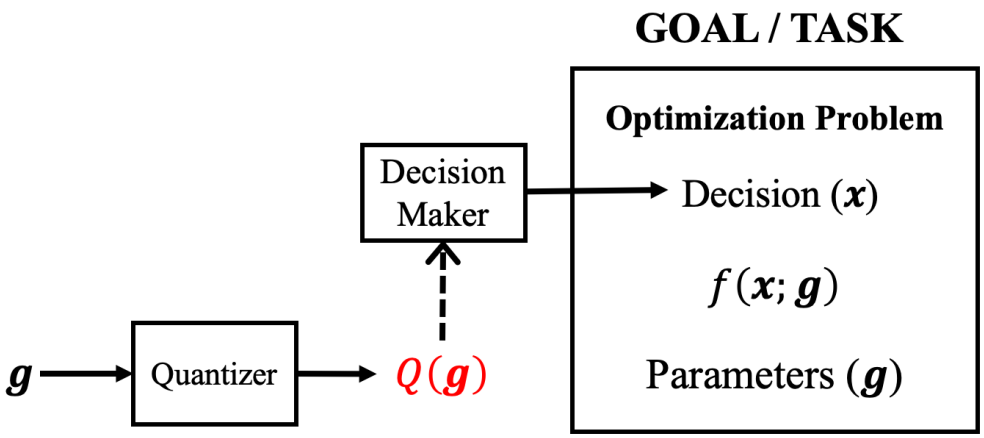


FIGURE 2.1 – Proposed definition for the goal-oriented quantization paradigm

2.2 Contributions

The contribution of this manuscript can be summarized based on four main aspects : 1) formulate the one-shot single-objective goal-oriented quantization problem ; provide a method to find the goal-oriented quantizer for (weakly) concave utility functions with convex polyhedral decision space ; propose an approach to find the goal-oriented quantizer when only the realizations of utility function are allowed to use. 2) extend the high-resolution quantization theory to goal-oriented quantization. 3) goal-oriented quantization problem is developed to the scenario where multiple correlated utility functions exist, i.e., in games ; 4) Nash equilibrium of a multi-user MIMO multiple access channel game with spectral efficiency as the individual utility function is studied and achieved in different proposed algorithms.

In Chapter 3, the one-shot quantization problem for goal-oriented quantization for a single utillity function is formulated. A basic algorithm to deal with goal-oriented quantization problem under perfect knowledge of the utility functions is proposed by mimicking the Lloyd-Max algorithm. Then we study the special case where utility function is concave and the decision space is a convex polyhedron. An algorithm based on minimizing the upper bound obtained by Jensen's inequality is proposed to find the goal-oriented quantizer. Finally we propose an learning based approach to solve the goal-oriented quantization problem when only the realizations of utility function are available. The main contributions of Chapter 3 are :

▶ For concave utility functions with convex polyhedral decision space, we use Jensen's inequality to upper bound the empirical optimality loss introduced by a goal-oriented quantizer. Then we introduce an equivalent class based on the decision label vector for a given parameter sample set. By iteratively optimizing the optimality loss within an equivalent class, we propose an algorithm called improve and branch algorithm which aims at finding a decision set so that the upper bound of optimality loss is minimized.

▶ Then we extend improve and branch algorithm designed for concave utility functions to weakly concave utility functions. By using generalized Jensen's equality for weakly concave utility functions, a modified version of improve and branch algorithm is thus applicable to weakly concave utility functions.

▶ An enhanced version of improve and branch algorithm is proposed by extending the

decision set according to improve and branch algorithm is proposed. Numeric results entails promising gain could be obtained compared to conventional quantizer
▶ We adapt the basic goal-oriented quantization algorithm to the scenario where only realizations of utility function is allowed to use.
▶ To find the quantization regions for fixed decision set, we use a feed-forward network to do so. A feed-forward neural network is proposed to find the quantization regions by learning from a training set. By comparing the sum-rate capacity function and an energy efficiency function, it is observed that the hardness of quantizing utility function could be largely different. Sum-rate capacity function could be a representative of nontrivial utility functions which requires almost no quantization bits to achieve tiny optimality loss.
▶ To find optimal decision set, IWO-DE algorithm combining the Invasive Weeds Occupation (IWO) and Differential Evolution (DE) is proposed. This approach is hereafter applied to a joint power and beamforming quantization problem. A reduction of a half quantization bits could be achieved by our proposed algorithm compared to conventional approaches.

We apply high resolution quantization theory to our goal-oriented quantization in Chapter 4. Different from existing research, we find a systematic way of finding good goal-oriented quantizer for general utility functions provided with some easily satisfied assumptions. Our analysis on cost function are divided into scalar case and vector case separately. The main contributions of Chapter are as follows :

▶ For scalar case, we have found a new high-resolution approximation for optimality loss and the corresponding optimal density of quantization interval. This new approximation formula entails that we could introduce a so-called value density (VD) $p(\boldsymbol{g})$ which could represent the contribution to the optimality loss for a given parameter. Replacing the conventional probability distribution by VD, one could easily find a well-performed goal-oriented quantizer by applying merely Lloy-Max Algorithm. Moreover, this approximation formula allows us to find the cost function which could introduce most (least) optimality loss while maintaining the optimal representative density. Finally, a new quantity called normalized optimality loss (NOL) is introduced to characterize the hardness of quantization comparing different cost functions.
▶ For vector case, we have extended the results in scalar case. Different from scalar case, The existence of value density is impossible due to a lack of universal (cell-independent) approximation formula for optimility loss. Therefore, we have found an upper-bound and lower bound for opitimality loss by admitting the correctness of Gersho's conjecture.
▶ Two algorithms are proposed based on eigenvalue approximation and iterative update of representatives to find a better good goal-oriented quantizer than the original one. Simulation results entail that proposed algorithm could reduce the optimality loss by replacing the current goal-ignorant quantizer.

For chapter 5, the goal-oriented quantization is extended to the scenario where multiple utility functions exist in a communication system and these utility functions are correlated through resources competition. The goal-oriented quantization is thus developed to the framework of games. Potential game with identical action set for different players is studied. Our main contributions are :
▶ First of all, we confirm that the use of goal-oriented quantization could be beneficial

for improving the social welfare of the system due to the existence of Braess's Paradox in a potential game.
► Secondly, viewing the maximum social welfare as a function of action set, we have shown that the existence of Braess's Paradox is connected to the monotonicity of this function. Besides, we have proven that the social welfare function is submodular with respect to the action set with a refinement of Nash equilibrium on the argmax set of the potential function of a potential game. Based on this property, an algorithm is proposed to find the optimal finite action set maximizing the social welfare of the game.
► Finally, we have proven that channel selection set is the optimal action set for 2-user 2-band multiple access channel (MAC) if spectral efficiency is taken as the individual utility function of each user. We also conjecture that the optimal action set for general scenario should be a Telatar-type set.

For chapter 6, we focus on a game where user's utility function is the energy efficiency (EE) in a MIMO multiple access channel system. Our main contributions are :
► The existence of Nash equilibrium is proven. The uniqueness of Nash equilibrium is confirmed by showing the standard property of MIMO-EE game.
► An algorithm by applying the approximate best response is proposed to approach the unique Nash equilibrium of the game. For 2-user 2-band scenario, our proposed algorithm Pareto-dominates the pure Nash equilibrium of the game.

2.3 Publications during Ph.D

Journal Papers :

- **H. Zou** , C. Zhang, S. Lasaulce, L. Saludjian and V. Poor, "Goal-oriented Quantization :
  High-resolution Analysis, Algorithms, and Performance Analysis", (Submitted to) *IEEE Transactions on Signal Processing.*

Conference Papers :

- **H. Zou**, C. Zhang, S. Lasaulce, L. Saludjian and P. Panciatici, "A Game-theoretical Approach For Energy Efficiency In Multiuser MIMO System", 2021 International Conference on Network Games, Control and Optimization (**NETGCOOP'21**), Corsica, France.
- C. Zhang, **H. Zou**, S. Lasaulce, Vineeth S. Varma, L. Saludjian and P. Panciatici, "Optimal Pricing Approach Based on Expected Utility Maximization with Partial Information", 2021 International Conference on Network Games, Control and Optimization (**NETGCOOP'21**), Corsica, France.
- Y. Sun, **H. Zou**, M. Kieffer, L. Saludjian and P. Panciatici, "A New Approach of Data Pre-processing for Data Compression in Smart Grids", The 7th International Conference on Wireless Networks and Mobile Communications (**WINCOM'19**), Fez, Morocco.

- **H. Zou**, C. Zhang, S. Lasaulce, L. Saludjian and P. Panciatici, "Energy-Efficient MIMO Multiuser Systems: Nash Equilibrium Analysis", International Symposium on Ubiquitous Networking (**UNET'19**), Limoges, France.

- **H. Zou**, C. Zhang, S. Lasaulce, L. Saludjian and P. Panciatici, "The Use of Machine Learning in Goal-oriented Communication", 1st DigiCosme Junior Conference on Wireless and Optical Communications (**JWOC'19**), Paris-Saclay University, France.

- **H. Zou**, C. Zhang, S. Lasaulce, L. Saludjian and P. Panciatici, "Decision Set Optimization and Energy-Efficient MIMO Communications", 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (**PIMRC'19**), Istanbul, Turkey.

- **H. Zou**, Y. Nait-Belaid, M. Kieffer, "Optimal Opponent Selection for Distributed Multi-Agent Self-Classification", IEEE Global Communications Conference (**GLOBECOM'18**), Abu Dhabi, UAE.

- **H. Zou**, C. Zhang, S. Lasaulce, L. Saludjian and P. Panciatici, "Decision-Oriented Communications: Application to Energy-Efficient Resource Allocation", The 6th International Conference on Wireless Networks and Mobile Communications (**WINCOM'18**), Marrakesh, Morocco.

- **H. Zou**, C. Zhang, S. Lasaulce, L. Saludjian and P. Panciatici, "Task Oriented Channel State Information Quantization", 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (**PIMRC'18**), Bologna, Italy.

3

Goal-Oriented Quantization for Single Utility Function

In this chapter, we first formulate the goal-oriented quantization problem tasked with a single utility for single-realization of parameter, i.e., one-shot quantization in Sec. 3.1. This problem is the basis of goal-oriented communication and could be extended to more complicated situations treated later in this manuscript. Secondly, in Sec. 3.2, we move to the ideal scenario where full information of utility function and p.d.f. of parameter are known. We propose a standard methodology to solve the goal-oriented quantization problem by splitting the goal-oriented quantization problem into two sub-problems : finding optimal quantization region for fixed decision set and finding optimal decision set for fixed quantization regions. Then in Sec. 3.3, we focus on the situation where utility functions involve specific structure or possess appropriate regularity properties. Finally the goal-oriented quantization is tackled under the assumption that only realizations of utility function is known. A feed-forward neural network is proposed to find the quantization region and an evolutionary algorithm is used to find the optimal decision set for the goal-orieted quantizer. Compared to conventional quantization which is goal-ignorant, the peformance of the communication system could be largely improved by implementing our proposed methods for respective scenarios.

3.1 Problem Formulation

To formulate the problem properly, we first briefly introduce some underlying notations of quantization and quantizer.

Definition 3.1.1. *A quantizer with $R = \log_2 M$ bits, input space $\mathcal{G}$, output alphabet $\widetilde{\mathcal{G}}$, consists of : 1) An encoding function $f_e : \mathcal{G} \to \mathcal{M} = \{1, 2, \ldots, M\}$ that maps the input into a positive integer representing its index; 2) A decoding function $f_d : \mathcal{M} \to \widetilde{\mathcal{G}}$ which maps the index $m \in \mathcal{M}$ into a codeword $\boldsymbol{z}_m \in \widetilde{\mathcal{G}}$.*

The objective of the quantization is to use a quantized version of the signal (referred to as representative) to represent the original signal, which could convey the information as much as possible. Therefore, the design of a quantizer consists in exploring the relationship between the representative and the original signal. For notational convenience, we could combine the encoding and decoding function of the standard quantization as a joint mapping :

$$\begin{aligned} \mathcal{Q} :& \mathcal{G} \to \widetilde{\mathcal{G}} \\ & \boldsymbol{g} \mapsto \boldsymbol{z}_m, \end{aligned} \tag{3.1}$$

i.e., $\boldsymbol{z}_m = f_d(f_e(\boldsymbol{g})) = \mathcal{Q}(\boldsymbol{g})$. In this manuscript, it is always assumed that $\widetilde{\mathcal{G}} \subseteq \mathcal{G}$ and $\mathcal{G}$ will be referred as the **parameter space** which can be $\mathbb{R}^{d_2}$ or $\mathbb{C}^{d_2}$ depending on the situation with d_2 the dimension of parameter variable g. For a quantizer with M representatives $\mathcal{R} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M\} \subset \widetilde{\mathcal{G}}$, the quantization regions can be denoted as

$$\mathcal{C}_m = \{\boldsymbol{g} \in \mathcal{G} : \mathcal{Q}(\boldsymbol{g}) = \boldsymbol{z}_m\}, \quad m \in \{1, \ldots, M\} \tag{3.2}$$

Quantization regions $\mathcal{C} = \{\mathcal{C}_m\}_{m=1}^M$ are disjoint and exhaustive, i.e.,

$$\bigcup_{m=1}^{M} \mathcal{C}_m = \mathcal{G}, \ \mathcal{C}_i \bigcap \mathcal{C}_j = \varnothing, \ \forall i \neq j. \tag{3.3}$$

Therefore, a quantizer $\mathcal{Q}$ can be fully characterized by its quantization regions $\mathcal{C}$ and corresponding representatives $\mathcal{R}$. For distortion-oriented quantizer, the m-th quantization region is defined as :

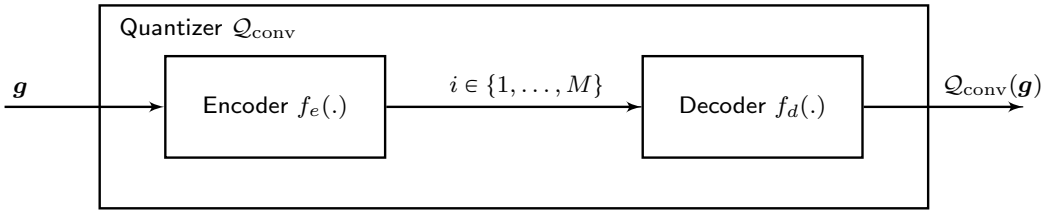
$$\mathcal{C}_m = \{\boldsymbol{g} \in \mathcal{G} : \|\boldsymbol{g} - \boldsymbol{z}_m\| \leq \|\boldsymbol{g} - \boldsymbol{z}_n\|, \ \forall n \neq m\} \tag{3.4}$$

The most conventional approach is to find $\mathcal{Q}$ that minimizes the mean square error (MSE) between the original signal and its quantized version :

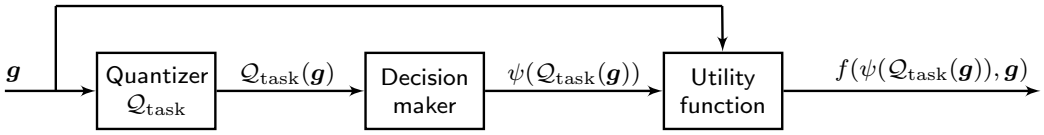
$$\mathcal{Q}_{\text{conv}} \in \arg\min_{\mathcal{Q}} \quad \mathbb{E}_{\boldsymbol{g}} \left[\|\mathcal{Q}(\boldsymbol{g}) - \boldsymbol{g}\|^2 \right] \tag{3.5}$$

A well-known method of solving the above minimization problem is to use alternating optimization algorithms to sequentially update the quantization regions $\{\mathcal{C}_m\}_{m=1}^M$ and the representative set $\{\boldsymbol{z}_m\}_{m=1}^M$, such as Lloyd-Max algorithm [31]. Due to its simplicity of implementation and fast convergence time, this approach has been widely applied in quantization and clustering problems. However, using predefined distance metric (e.g., Euclidean distance) to obtain the partitions and representatives is sub-optimal since the final use of the quantized signal are not taken into account. For instance, concerning a communication system consists of transmitters and receivers, receivers aim to send the quantized channel state information (CSI) to the transmitters such that the transmit power (or beamforming vector) can be better chosen at the transmitter side or vice versa. Obviously, the objective of the quantization here is not to reconstruct the CSI, but to convey the informative messages of CSI related to the subsequent decision-making process as much as possible. When the final use of the quantized parameter is known, the way to partition the parameter space can be made according to relevant features of the parameter and thus improved. This is precisely the quantization approach studied in this manuscript.

More precisely, the goal-oriented quantization consists in assuming that the task to be performed by the decision-making entity (e.g., the transmitter) can be represented



(a) Illustration of the conventional quantizer



(b) Illustration of the goal-oriented quantizer

FIGURE 3.1 – Comparison between conventional quantizer and goal-oriented quantizer.

by a standard optimization problem (OP), that is, a given function has to be minimized under some constraints. Therefore, the objective is to maximize a certain function or performance metric $f(\boldsymbol{x}; \boldsymbol{g})$ (e.g., some cost or expense function) with respect to the decision variable $x \in \mathcal{X} \subset \mathbb{R}^{d_1}$ with $\mathcal{X}$ being the **decision space** which is generally $\mathbb{R}^{d_1}$ with d_1 the dimension of decision space. This mathematically writes as the following standard form OP :

$$\underset{\boldsymbol{x} \in \mathcal{X}}{\text{maximize}} \quad f(\boldsymbol{x}; \boldsymbol{g}) \tag{3.6}$$

Denote $\psi(\boldsymbol{g})$ an optimal solution of the above OP, i.e.,

$$\psi(\boldsymbol{g}) \in \arg\max_{\boldsymbol{x} \in \mathcal{X}} \quad f(\boldsymbol{x}; \boldsymbol{g}), \tag{3.7}$$

and the optimal value of utility function (optimal value function) :

$$f^{\star}(\boldsymbol{g}) = \max_{x \in \mathcal{X}} f(\boldsymbol{x}; \boldsymbol{g}) \tag{3.8}$$

It is worth mentioning that $\psi(g)$ is generally not unique for utility functions, especially for vector case ($d_2 \geq 2$). Function $\psi(\boldsymbol{g})$ will be referred as the optimal decision function (ODF) in this manuscript. For example, if the utility function is symmetric w.r.t. the decision variable $\boldsymbol{x}$, one can has multiple optimal decision functions. For the sake of simplicity, we assume that the optimal decision function is unique or choose the one satisfying some assumptions later imposed in this manuscript. This assumption is reasonable since the the performance of the designed quantizer is the central interest of our goal-oriented quantization problem. Therefore two optimal decision functions could be treated equivalently for corresponding to an unique optimal value function. The problem of finding a goal-oriented quantization scheme therefore amounts to solving the following problem

$$\mathcal{Q}_{\text{task}} \in \arg\max_{\mathcal{Q}} \mathbb{E}_{\boldsymbol{g}}[f(\psi(\mathcal{Q}(\boldsymbol{g})); \boldsymbol{g})]. \tag{3.9}$$

For a fixed probability density function (p.d.f.) $\phi(\boldsymbol{g})$ of the parameter $\boldsymbol{g}$, to evaluate the performance of the goal-oriented quantization, we can assess the absolute optimality

loss induced by quantization error of quantizer $\mathcal{Q}$ with $R = \log_2 M$ bits for utility function f as follows :

$$\begin{aligned} &\mathrm{L}\left(\mathcal{Q}_{\text{task}}; f, R, \phi\right) \\ &= \mathbb{E}_{\boldsymbol{g}}\left[f\left(\psi\left(\boldsymbol{g}\right); \boldsymbol{g}\right) - f\left(\psi\left(\mathcal{Q}_{\text{task}}\left(\boldsymbol{g}\right)\right); \boldsymbol{g}\right)\right] \\ &= \int_{\boldsymbol{g}\in\mathcal{G}} \left[f\left(\psi\left(\boldsymbol{g}\right); \boldsymbol{g}\right) - f\left(\psi\left(\mathcal{Q}_{\text{task}}\left(\boldsymbol{g}\right)\right); \boldsymbol{g}\right)\right] \phi\left(\boldsymbol{g}\right) \mathrm{d}\boldsymbol{g} \end{aligned} \tag{3.10}$$

Note that $\mathbb{E}_{\boldsymbol{g}}\left[f\left(\psi\left(\boldsymbol{g}\right); \boldsymbol{g}\right)\right]$ is independent of the quantizer, thus the problem defined by (3.9) can be equivalently treated as minimizing the OL $\mathrm{L}\left(\mathcal{Q}; f, R, \phi\right)$. The OP to be solved can be equivalently written as :

$$\min_{\{\boldsymbol{z}_m\}, \{\mathcal{C}_m\}} \sum_{m=1}^{M} \int_{\boldsymbol{g}\in\mathcal{C}_m} \left[f\left(\psi\left(\boldsymbol{g}\right); \boldsymbol{g}\right) - f\left(\psi\left(\boldsymbol{z}_m\right); \boldsymbol{g}\right)\right] \phi(\boldsymbol{g}) \mathrm{d}\boldsymbol{g}, \tag{3.11}$$

where the m-th quantization region is defined as :

$$\mathcal{C}_m = \left\{\boldsymbol{g} \in \mathbb{R}^{d_2} : f\left(\psi\left(\boldsymbol{z}_m\right); \boldsymbol{g}\right) \geq f\left(\psi\left(\boldsymbol{z}_n\right); \boldsymbol{g}\right),\ \forall n \neq m\right\} \tag{3.12}$$

Interestingly, it can be checked that the conventional quantization approach can be treated as a special case of the OP defined by (3.11) by choosing the function as $f(\boldsymbol{x}; \boldsymbol{g}) = (\boldsymbol{x} - \boldsymbol{g})^2$.

For each representative $\boldsymbol{z}_m \in \mathcal{R}_M$, the ODF actually fix a so-called decision $\boldsymbol{d}_m = \psi\left(\boldsymbol{z}_m\right)$ and all these decisions form a **decision set** $\mathcal{D} \triangleq \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M\}$. Then OP in (3.11) can be rewritten as

$$\min_{\{\boldsymbol{d}_m\}, \{\mathcal{C}_m\}} \sum_{m=1}^{M} \int_{\boldsymbol{g}\in\mathcal{C}_m} \left[f\left(\psi\left(\boldsymbol{g}\right); \boldsymbol{g}\right) - f\left(\boldsymbol{d}_m; \boldsymbol{g}\right)\right] \phi(\boldsymbol{g}) \mathrm{d}\boldsymbol{g} \tag{3.13}$$

OP in (3.13) is more general since it is normal and frequent that the knowledge of ODF is limited or missing. We denote the solution of OP (3.11) as $(\mathcal{D}^\star, \mathcal{C}^\star)$.

For a given quantizer $\mathcal{Q}$ and utility function f, we define the relative optimality loss (ROL) as :

$$\sigma(\%) = \frac{\mathrm{L}\left(\mathcal{Q}; f, R, \phi\right)}{\mathbb{E} f^\star\left(g\right)} \times 100\% \tag{3.14}$$

Then it is naturally reasonable to ask following questions : i) for a given ROL ratio (σ), could we determine the minimum number of cells (M_σ) to achieve a certain ROL ? ii) how could we find such a quantizer $\mathcal{Q}$ if perfect knowledge of the utility function lacks ? iii) for different utility functions, could M_σ be tremendously distinct ? To answer these questions, we will start with GOQ with perfect knowledge of utility function.

3.2 Goal-Oriented Quantization with Perfect Knowledge of Utility Function

We first tackle the GOQ problem in the following scenario : perfect knowledge of utility function is available. One can split the GOQ problem into 2 sub-problems like classical approach such as Lloyd-Max algorithm :

1. The representative-to-cell step which is essentially finding the quantization region (cells) given the concrete decision set $\mathcal{D}$ (or set of representatives) :

$$\mathcal{C}_m^\star = \left\{ \boldsymbol{g} \in \mathcal{G} \,\middle|\, f\left(\boldsymbol{d}_m; \boldsymbol{g}\right) = \max_l f\left(\boldsymbol{d}_l; \boldsymbol{g}\right) \right\} \tag{3.15}$$

2. The cell-to-representative step which is essentially finding the optimal decision (representatives) given the concrete quantization regions (cells) $\mathcal{C}$:

$$\boldsymbol{d}_m^\star \in \arg\max_{\boldsymbol{x} \in \mathcal{X}} \int_{\boldsymbol{g} \in \mathcal{C}_m} f\left(\boldsymbol{x}; \boldsymbol{g}\right) \phi\left(\boldsymbol{g}\right) \mathrm{d}\boldsymbol{g}, \ 1 \leq m \leq M \tag{3.16}$$

If we run step 1) and 2) iteratively, we obtain the basic GOQ algorithm which is actually an alternating descent algorithm summarized in alg.

Inputs : error tolerance ε and max iteration T
Inputs : initial decision set $\mathcal{D}^{(0)} = \left\{ \boldsymbol{d}_1^{(0)}, \ldots, \boldsymbol{d}_M^{(0)} \right\}$;
Inputs : initial quantization region $\mathcal{C}^{(0)} = \left\{ \mathcal{C}_1^{(0)}, \ldots, \mathcal{C}_M^{(0)} \right\}$;
for $i = 1$ **to** T **do**
  for $m = 1$ **to** M **do**
    Update $\mathcal{C}_m^{(i)}$ by $\mathcal{C}_m^\star = \{ \boldsymbol{g} \in \mathcal{G} \,|\, f\left(\boldsymbol{d}_m; \boldsymbol{g}\right) = \max_l f\left(\boldsymbol{d}_l; \boldsymbol{g}\right) \}$;
    Update $\boldsymbol{d}_m^{(i)}$ by $\boldsymbol{d}_m^\star \in \arg\max_{\boldsymbol{x} \in \mathcal{X}} \int_{\boldsymbol{g} \in \mathcal{C}_m} f\left(\boldsymbol{x}; \boldsymbol{g}\right) \phi\left(\boldsymbol{g}\right) \mathrm{d}\boldsymbol{g}$;
  end
  if $\sum_{m=1}^{M} \left\| \boldsymbol{d}_m^{(i)} - \boldsymbol{d}_m^{(i-1)} \right\|^2 < \varepsilon$ **then**
    Break;
  end
end
Outputs : $\mathcal{D}^\star = \mathcal{D}^{(i)}$ and $\mathcal{C}^\star = \mathcal{C}^{(i)}$.

Algorithm 1: Basic goal-oriented quantization algorithm

There are two major difficulties in applying the basic GOQ algorithm. The first one is that, in many practical applications, only the realizations of utility function instead of its explicit expression is known. Thus the opitimization problem in Eq. 3.16 is cumbersome to be solve directly. The second one is that finding decision region is equivalent to solve an infinite dimensional problem even it is trivial for a given parameter. Therefore the integral in Eq. 3.16 is generally difficult to evaluate which make alg. 1 a highly abstract algorithm to be implemented. To begin with, we will first consider a special case where extra regularity properties are assumed for utility functions.

3.3 Goal-Oriented Quantization for Utility Function with Convex Polyhedral Decision Space

In this section, we will see the how the regularity properties of the utility function will help us in finding the goal-oriented quantizer. Precisely, we will consider a simple case

of our general goal-oriented quantization problem where the decision space is a convex polyhedron and the utility function is concave w.r.t. decision variable. Moreover, as we will see, the proposed methods in this section could be considered as an example of solving the goal-oriented quantization problem when the knowledge of optimal decision function is missing.

3.3.1 Problem formulation

We assume that **the decision space $\mathcal{X}$ is a convex polyhedron represented by a graph (V, E). $V = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_P\}$ with $\boldsymbol{v}_i$ is called a vertex of the polyhedron. We say that $(\boldsymbol{v}_i, \boldsymbol{v}_j)$ is an edge (face of dimension one) of $\mathcal{X}$ if and only $(\boldsymbol{v}_i, \boldsymbol{v}_j) \in E$.** The scenario where the the decision space of the system is a polyhedral is frequently met in many different domains, e.g., the power allocation problem. Consider a goal-oriented quantizer $\mathcal{Q}$ characterized by its quantization regions $\mathcal{C} = \{\mathcal{C}_m\}_{m=1}^M$ and decision set $\mathcal{D} = \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M\}$. To emphasize the dependence of decision set, we define the sub-optimal solution maximizing the utility function restricted on the decision set $\mathcal{D}$ of $\mathcal{Q}$ instead of $\mathcal{X}$ as

$$\widehat{\psi}\left(\boldsymbol{g}|\mathcal{D}\right) \in \arg\max_{\boldsymbol{x}\in\mathcal{D}} f\left(\boldsymbol{x};\boldsymbol{g}\right) \tag{3.17}$$

The optimality loss introduced by quantizer $\mathcal{Q}$ is thus :

$$\begin{aligned}
&\mathrm{L}\left(\mathcal{Q}; f, R, \phi\right) \\
&= \mathbb{E}_g\left[f\left(\psi\left(\boldsymbol{g}\right);\boldsymbol{g}\right) - f\left(\widehat{\psi}\left(\boldsymbol{g}|\mathcal{D}\right);\boldsymbol{g}\right)\right] \\
&= \sum_{M=1}^{M}\int_{\boldsymbol{g}\in\mathcal{C}_i}\left[f\left(\psi\left(\boldsymbol{g}\right);\boldsymbol{g}\right) - f\left(\boldsymbol{d}_m;\boldsymbol{g}\right)\right]\phi\left(\boldsymbol{g}\right)\mathrm{d}\boldsymbol{g} \\
&= \int_{\boldsymbol{g}\in\mathcal{G}}\sum_{m=1}^{M}\left[f\left(\psi\left(\boldsymbol{g}\right);g\right) - f\left(\boldsymbol{d}_m;\boldsymbol{g}\right)\right]\phi\left(g\right)\mathbb{1}\left\{\boldsymbol{g}\in\mathcal{C}_m\right\}\mathrm{d}\boldsymbol{g},
\end{aligned} \tag{3.18}$$

where $\mathbb{1}\left(\cdot\right)$ is a indicator function. Consider a sufficiently large parameter sample set $\mathcal{T} = \left\{\boldsymbol{g}^{(t)}\right\}_{t=1}^T$, then optimality loss can be approximated by a empirical optimality loss :

$$\overline{\mathrm{L}}\left(\mathcal{Q}; f, R, \mathcal{T}\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\left[f\left(\psi\left(\boldsymbol{g}^{(t)}\right);\boldsymbol{g}^{(t)}\right) - f\left(\boldsymbol{d}_m;\boldsymbol{g}^{(t)}\right)\right]\mathbb{1}\left\{\boldsymbol{g}^{(t)}\in\mathcal{C}_m\right\}. \tag{3.19}$$

Eq. 3.19 actually avoids solving the problem of finding quantization regions in Sec. 3.2. Then it remains how to find the optimal decision set $\mathcal{D}$. To simplify the notation, it is reasonable to take the following abuse of notation $\overline{\mathrm{L}}\left(\mathcal{D}; f, R, \mathcal{T}\right) = \overline{\mathrm{L}}\left(\mathcal{Q}; f, R, \mathcal{T}\right)$. However Eq. 3.19 is difficult to optimize directly under general settings due to the existence of an indicator function. Therefore We assign some extra regularity properties for utility functions.

3.3.2 Analysis of concave utility functions

In this section, we begin with the simplest case where **the utility function is a concave function w.r.t. decision variable $\boldsymbol{x}$**. Since the decision space is a convex

polyhedron, each decision can be expressed as the convex combination of vertices : $\boldsymbol{d}_m = \sum_{i=1}^{P} \mathrm{A}_{im}\boldsymbol{v}_i$ with $\sum_{i=1}^{P} \mathrm{A}_{im} = 1, \forall 1 \leq m \leq M$. Obviously matrix A contains all the information as decision set $\mathcal{D}$. Define vector $\boldsymbol{u}^\star = \left(f\left(\psi\left(\boldsymbol{g}^{(t)}\right);\boldsymbol{g}^{(t)}\right)\right)_{t=1}^{T} \in \mathbb{R}^{T\times 1}$ to store the optimal value of each sample. Introduce a matrix function $\mathrm{B}(\mathrm{A}) = (\mathrm{B}_{tm}(\mathrm{A}))_{t,m} \in \mathbb{R}^{T\times N}$ with $\mathrm{B}_{tm}(\mathrm{A}) = \mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{C}_m\right\}$ for $\forall 1 \leq m \leq M$. Introduce a constant matrix $\mathrm{U} = \left(f\left(\boldsymbol{v}_i;\boldsymbol{g}^{(t)}\right)\right)_{i,t} \in \mathbb{R}^{P\times T}$ and a matrix function $\mathrm{Y}(\mathrm{A}) = \mathrm{UB}(\mathrm{A}) \in \mathbb{R}^{P\times M}$. Furthermore we write $\mathrm{A} = (\mathrm{A}_1, \ldots, \mathrm{A}_M)$, $\mathrm{B} = (\mathrm{B}_1, \ldots, \mathrm{B}_M)$ and $\mathrm{Y} = (\mathrm{Y}_1, \ldots, \mathrm{Y}_M)$, then the empirical optimality loss can be expressed as :

$$
\begin{aligned}
&\overline{\mathrm{L}}(\mathcal{D}; f, R, \mathcal{T}) \\
=&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\left[f\left(\psi\left(\boldsymbol{g}^{(t)}\right);\boldsymbol{g}^{(t)}\right) - f\left(\boldsymbol{d}_m;\boldsymbol{g}^{(t)}\right)\right]\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{C}_m\right\} \\
=&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\left[f\left(\psi\left(\boldsymbol{g}^{(t)}\right);\boldsymbol{g}^{(t)}\right) - f\left(\sum_{i=1}^{P}\mathrm{A}_{im}\boldsymbol{v}_i;\boldsymbol{g}^{(t)}\right)\right]\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{C}_m\right\} \\
\leq&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\left[f\left(\psi\left(\boldsymbol{g}^{(t)}\right);\boldsymbol{g}^{(t)}\right) - \sum_{i=1}^{P}\mathrm{A}_{im}f\left(\boldsymbol{v}_i;\boldsymbol{g}^{(t)}\right)\right]\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{C}_m\right\} \\
=&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}f\left(\psi\left(\boldsymbol{g}^{(t)}\right);\boldsymbol{g}^{(t)}\right)\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{C}_m\right\} - \frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{i=1}^{P}\mathrm{A}_{im}f\left(\boldsymbol{v}_i;\boldsymbol{g}^{(t)}\right)\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{C}_m\right\} \\
=&\underbrace{\frac{1}{T}\sum_{m=1}^{M}\mathrm{B}_m^{\mathrm{T}}(\mathrm{A})\,\boldsymbol{u}^\star - \frac{1}{T}\sum_{m=1}^{M}\mathrm{A}_m^{\mathrm{T}}\mathrm{UB}_m(\mathrm{A})}_{\mathrm{L_{ub}}(\mathcal{D};f,R,\mathcal{T})},
\end{aligned}
\tag{3.20}
$$

where the inequality comes from Jensen' inequality. One thus obtain an upper bound $\mathrm{L_{ub}}(\mathcal{D}; f, R, \mathcal{T})$ for the empirical optimality loss $\overline{\mathrm{L}}(\mathcal{D}; f, R, \mathcal{T})$. If the upper bound could be minimized somehow, then we can ensure that the empirical loss will not exceed this upper bound. Before reveal the details of our method, we need to introduce some extra conceptions. Obviously there is a one-to-one mapping between the decision set $\mathcal{D}$ with cardinality M and the set of N-fold unit simplex $\triangle_P^N$ of dimension P containing all possible setting of matrix A. Therfore we can define the following equivalent relation for matrix A :

Definition 3.3.1. *(Equivalent Relation) For a given sample set $\mathcal{T}$ and $A, A' \in \triangle_P^N$, $A' \sim_{\mathcal{T}} A$ if and only if $B(A) = B(A')$. The equivalent class for A is denoted as $[A]_{\mathcal{T}}$.*

Definition 3.3.2. *(Optimal Improvement) For a given sample set $\mathcal{T}$ and a matrix $A \in \triangle_P^N$, A^+ is said to be an optimal improvement for A if it holds that*

$$
A^+ \in \arg\max_{C\in[A]_{\mathcal{T}}} \sum_m C_m^{\mathrm{T}} UB_m(C) \tag{3.21}
$$

One can easily prove the defined operation $\sim_{\mathcal{T}}$ is an equivalent relation for any sample set $\mathcal{T}$. Two decision sets (represented by matrix A) are equivalent for a given sample set $\mathcal{T}$ means that their images under matrix-valued mapping B are exactly the same. If the sample set is fixed, we will omit it for both equivalent operator and equivalent class to simplify the notation. The meaning for optimal improvement should be further explained :

an optimal improvement A^{+} for a decision set A is the equivalent element of matrix A which minimizes the upper bound obtained in Eq. 3.20. This fact is clearly shown in following equations :

$$\begin{aligned}\mathrm{A}^{+} &\in \arg\min_{\mathrm{C}\in[\mathrm{A}]} \sum_{m} \mathrm{B}_{m}^{\mathrm{T}}(\mathrm{C})\,\boldsymbol{u}^{\star} - \sum_{m} \mathrm{C}_{m}^{\mathrm{T}}\mathrm{UB}_{n}(\mathrm{C}) \\ &\in \arg\max_{\mathrm{C}\sim\mathrm{A}} \sum_{m} \mathrm{C}_{m}^{\mathrm{T}}\mathrm{UB}_{m}(\mathrm{C}) \end{aligned} \tag{3.22}$$

In other words, if one goes further away from A^{+}, one is not sure whether the empirical optimality loss will increase or decrease. However, it is generally difficult to find A^{+} for a given matrix A directly. The reason is that the boundary of each equivalent class is implicitly given and finding A^{+} for each $[\mathrm{A}] \in \mathfrak{A}$ could be costly. Therefore, we will introduce one operator which is more reasonable and easy to operate within an equivalent class.

Definition 3.3.3. *(Greedy Improvement) Define $A^{\dagger}$ for A with $Y = UB(A)$ s.t.*

$$A^{\dagger} - A = \nu E(A)\,, \tag{3.23}$$

with $\nu = \sup\{y \geq 0 \,|\, A \sim A + yE(A)\}$ and an auxiliary matrix $E(A) = (E_{ij}(A))_{i,j} \in \mathbb{R}^{P\times T}$ with

$$E_{ij}(A) = \begin{cases} -1, & i = \ell^{\star}, j = m^{\star} \\ 1, & i = k^{\star}, j = m^{\star} \\ 0, & \textit{otherwise.} \end{cases} \tag{3.24}$$

where

$$m^{\star} \in \arg\max_{m} \left[\max_{i} Y_{im} - \min_{i} Y_{im}\right], \tag{3.25}$$

$$\ell^{\star} \in \arg\min_{1\leq t\leq P} Y_{tm^{\star}}, \tag{3.26}$$

$$k^{\star} \in \arg\max_{1\leq t\leq P} Y_{tm^{\star}}, \tag{3.27}$$

The meaning of greedy improvement is that this matrix minimizes the upper bound of optimality loss along the deepest-gradient-descent direction within one particular equivalent class. Still, it is cumbersome to find the precise value of ν. Fortunately, this fact never stops us to construct an algorithm to find a decision set . The basic idea is that if there $\exists \nu' > \nu$ so that matrix $\mathrm{A}' = \mathrm{A} + \nu\mathrm{E}(\mathrm{A})$ satisfying $\mathrm{L_{ub}}\left(\mathrm{A}'; f, R, \phi\right) < \mathrm{L_{ub}}(\mathrm{A}; f, R, \phi)$ and $\mathrm{A}' \nsim \mathrm{A}$, then we do find a decision set represented by A' strictly better than the decision set A outside of the equivalent class of A which means that one can further improve A' by continuing increase ν' ; Otherwise, the meaning of obtained matrix is not clear, then a new search direction should be created for A'. Based on this idea an algorithm called improvement and branch algorithm is proposed summarized in alg. 2 :

Initialization : choose $\mathrm{A}^{(0)}$ and step size $\varepsilon > 0$, generate sample set $\mathfrak{T}$;
$\mathrm{K}^{(0)} \leftarrow \mathrm{A}^{(0)}$;
$\mathrm{L}^{(0)} \leftarrow \overline{\mathrm{L}}\left(\mathrm{A}^{(0)}; f, R, \mathfrak{T}\right)$;
for $i = 0$ *to ITER* **do**
  $\mathrm{Y}^{(i)} \leftarrow \mathrm{UB}\left(\mathrm{A}^{(i)}\right)$;
  $n^\star \in \arg\max_m \left[\max_t \mathrm{Y}_{tm}^{(i)} - \min_t \mathrm{Y}_{tm}^{(i)}\right]$;
  $\ell^\star \in \arg\min_{1\le t\le P} \mathrm{Y}_{tm^\star}^{(i)}$;
  $k^\star \in \arg\min_{1\le t\le P} \mathrm{Y}_{tm^\star}^{(i)}$;
  $\mathrm{C}^{(0)} \leftarrow \mathrm{A}^{(i)}$;
  for $j = 1$ *to* J **do**
    $\mathrm{C}^{(j)} \leftarrow \mathrm{C}^{(j-1)} + \varepsilon \mathrm{E}\left(\mathrm{A}\right)$;
    if $C^{(j)} \nsim C^{(j-1)}$ **then**
      $\mathrm{A}^{(i+1)} \leftarrow \mathrm{C}^{(j)}$;
      if $\overline{\mathrm{L}}\left(C^{(j-1)}; f, R, \mathfrak{T}\right) < \overline{\mathrm{L}}\left(C^{(j)}; f, R, \mathfrak{T}\right)$ **then**
        $\mathrm{L}^{(i+1)} \leftarrow \overline{\mathrm{L}}\left(\mathrm{C}^{(j-1)}; f, R, \mathfrak{T}\right)$;
        $\mathrm{K}^{(i+1)} \leftarrow \mathrm{C}^{(j-1)}$;
      else
        $\mathrm{L}^{(i+1)} \leftarrow \overline{\mathrm{L}}\left(\mathrm{C}^{(j)}; f, R, \mathfrak{T}\right)$;
        $\mathrm{K}^{(i+1)} \leftarrow \mathrm{C}^{(j)}$;
      end
    end
  end
end
$i^\star \in \arg\min_{0\le i\le \mathrm{ITER}} \mathrm{L}^{(i)}$;
Output: Required decision set is represented by $\mathrm{K}^{(i^\star)}$;

Algorithm 2: Improve and Branch Algorithm

Remark 3.3.4. *It ts obvious that the number of equivalent class matrix vector A depends on the number of parameter samples T. In worst case, we could have $|[A]| = 2^T$ which entails that the direct exhaustive search for matrix B leads to an exponential complexity of $O\left(2^T\right)$. In the other hands, the accuracy of Monte Carlo approximation depends on the number of samples. There is obviously a trade-off between the accuracy of Monte-Carlo approximation and the complexity of the algorithm. Alg. 2 provides a better way than the direct search.*

3.3.3 Analysis of weakly concave utility functions

The basic idea of alg. 2 depends on the concavity of the utility function w.r.t. decision variable . With the help of Jensen's inequality, the original optimization problem is reduced to a family of linear OP labeled by the value of matrix B. However, this method fails for non-concave utility function which is more frequently met in practical applications. In this section, we will show that alg. 2 could be generalized to what we called weakly

concave utility functions. To achieve that, we first introduce the general version of Jensen's inequality. Without loss of generality, we assume that the utility function $f(\boldsymbol{x};\boldsymbol{g})$ is twice-differentiable w.r.t. to decision variable $\boldsymbol{x}$, i.e., $f \in \mathrm{C}^2_{\boldsymbol{x}}[\mathbb{R}]$. We denote the Hessian matrix $\mathrm{H}(\boldsymbol{x}_0;\boldsymbol{g})$ of f w.r.t. $\boldsymbol{x}$ at point $(\boldsymbol{x}_0,\boldsymbol{g})$ and its largest eigen-value given parameter $\boldsymbol{g}$ for all possible $\boldsymbol{x} \in \mathcal{X}$ as $\lambda_{\max}(\boldsymbol{g})$. To generalize Jensen's inequality to non-concave function, we introduce some important conceptions :

Definition 3.3.5. *(r-weakly concave function) Given a continuous function $u : \mathbb{R}^P \to \mathbb{R}$ defined on a convex set S, consider the function $h : \mathbb{R}^{P+1} \to \mathbb{R}$ with $r \in \mathbb{R}$ defined by : $h(\boldsymbol{x},r) = u(\boldsymbol{x}) + \frac{1}{2}r\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$. If function $h(\boldsymbol{x},r)$ is a concave function on S for some $r \in \mathbb{R}$, then $h(\boldsymbol{x},r)$ is called a concavification of u. Function u is said to be r-weakly concave if it has a concavification of weakly concave constant r.*

Proposition 3.3.6. *(Generalized Jensen's Inequality for weakly concave functions) For any r-weakly concave function $u : \mathbb{R}^P \to \mathbb{R}$, for $\forall \boldsymbol{a} \in \triangle_P$ and a series of points $(\boldsymbol{x}^i)_{i=1}^P$ with $\boldsymbol{x}^i \in \mathbb{R}^P$ it holds that*

$$u\left(\sum_{i=1}^{P} \boldsymbol{a}_i \boldsymbol{x}^i\right) \geq \sum_{i=1}^{P} \boldsymbol{a}_i u\left(\boldsymbol{x}^i\right) + \frac{r}{2}\left[\sum_{i,j=1}^{P} \boldsymbol{a}_i \boldsymbol{a}_j \left(\boldsymbol{x}^i - \boldsymbol{x}^j\right)^{\mathrm{T}} \boldsymbol{x}^i\right] \tag{3.28}$$

Equipped with generalized Jensen's inequality, method in above subsection is possible to be applied to weakly concave utility functions. Similar to analysis in previous subsection, empirical loss can be upper bounded as :

$$\overline{\mathrm{L}}(\mathcal{Q};f,R,\mathcal{T}) \leq \frac{1}{T}\sum_{m} \mathrm{B}_m^{\mathrm{T}}\boldsymbol{u}^{\star} - \frac{1}{T}\sum_{m} \mathrm{A}_m^{\mathrm{T}}\mathrm{U}\mathrm{B}_m - \frac{1}{2T}\sum_{t,m,i,j} \mathrm{A}_{im}\mathrm{A}_{jm}\boldsymbol{v}_i^{\mathrm{T}}(\boldsymbol{v}_i - \boldsymbol{v}_j)\mathrm{B}_{tm}\boldsymbol{\rho}_t, \tag{3.29}$$

where $\boldsymbol{\rho}_t$ is a weakly concave constant of function f given $\boldsymbol{g}^{(t)}$ w.r.t. $\boldsymbol{x}$. Further define matrix $\mathrm{V} = (\mathrm{V}_{ij})_{i,j}$ with $\mathrm{V}_{ij} = \boldsymbol{v}_i^{\mathrm{T}}(\boldsymbol{v}_i - \boldsymbol{v}_j)$, Eq. 3.29 can be rewritten as :

$$\overline{\mathrm{L}}(\mathcal{Q};f,R,\mathcal{T}) \leq \frac{1}{T}\sum_{m} \mathrm{B}_m^{\mathrm{T}}\boldsymbol{u}^{\star} - \frac{1}{T}\sum_{m} \mathrm{A}_m^{\mathrm{T}}\mathrm{U}\mathrm{B}_m - \sum_{m} \frac{\mathrm{A}_m^{\mathrm{T}}\mathrm{V}\mathrm{A}_m}{2T}\boldsymbol{\rho}^{\mathrm{T}}\mathrm{B}_m \tag{3.30}$$

Obviously one would like to minimize the optimality loss introduced by $\mathcal{Q}$, then the optimal choice for vector $\boldsymbol{\rho}$ should be $\boldsymbol{\rho}_t = -\lambda_{\max}\left(\boldsymbol{g}^{(t)}\right)$ since only for $r \leq -\lambda_{\max}\left(\boldsymbol{g}^{(t)}\right)$ for $\forall t$, the utility function will have a concavification. Similarly, the concepts of optimal improvement and greedy improvement could be defined for weakly concave utility function to have a similar method to alg. 2. The details are omitted here to avoid duplicated materials.

3.3.4 Enhanced improve and branch algorithm

In this subsection, we will consider the following problem which is the general version of problem in Sec. 3.3.2 : how to extend a decision set $\mathcal{D}_N$ with cardinality N to a new decision set $\mathcal{D}_{N+M}$ with $N \in \mathbb{N}$ efficiently. This problem will finally help us to design an enhanced version of alg. 2. For notation convention, we denote the new extended decision

set as :

$$\begin{aligned}\mathcal{D}_{M+N} &= \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_M\} \\ &= \mathcal{D}_N \bigcup X_M,\end{aligned} \tag{3.31}$$

where $X_M = \{\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_M\}$ is obviously the set of extended decisions. If one set $\mathcal{D}_N = \varnothing$, then one obtains the original problem immediately.

We further define the following partition of the parameter space $\mathcal{G} = \bigcup_{n,m} \mathcal{G}_{mn}$ with $\mathcal{G}_{nm}$ defined as

$$\mathcal{G}_{mn} \triangleq \left\{ \boldsymbol{g} \,\middle|\, \widehat{\psi}\left(\boldsymbol{g}|\mathcal{D}_M\right) = \boldsymbol{d}_n, \widehat{\psi}\left(\boldsymbol{g}|\mathcal{D}_{M+N}\right) = \boldsymbol{\zeta}_m \right\}, \tag{3.32}$$

for $1 \leq n \leq N$ and $1 \leq m \leq M$. The meaning of $\mathcal{G}_{mn}$ is set of all parameter so that the sub-optimal decision switches from $\boldsymbol{d}_n$ to $\boldsymbol{\zeta}_m$ once the decision set $\mathcal{D}_N$ (corresponding quantizer denoted as $\mathcal{Q}_N$) is replaced by new decision set $\mathcal{D}_{N+M}$ (corresponding quantizer denoted as $\mathcal{Q}_{N+M}$). For $\mathcal{D}_{N+M}$, one can easily have :

$$\begin{aligned}&\overline{\mathrm{L}}\left(\mathcal{D}_N; f, \phi\right) - \overline{\mathrm{L}}\left(\mathcal{D}_{N+M}; f, \phi\right) \\ =&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N}\left[f\left(\boldsymbol{d}_n; \boldsymbol{g}^{(t)}\right) - f\left(\boldsymbol{\zeta}_m; \boldsymbol{g}^{(t)}\right)\right]\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{G}_{mn}\right\}\end{aligned} \tag{3.33}$$

Again one has $\boldsymbol{\zeta}_m = \sum_{i=1}^{P} \mathrm{A}_{im}\boldsymbol{v}_i$ for convex polyhedral decision space. The decay of empirical optimality loss by extending $\mathcal{D}_N$ to $\mathcal{D}_{N+M}$ could be expressed as :

$$\begin{aligned}&\overline{\mathrm{L}}\left(\mathcal{Q}_N; f, \phi\right) - \overline{\mathrm{L}}\left(\mathcal{Q}_{N+M}; f, \phi\right) \\ =&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N}\left[f\left(\boldsymbol{\zeta}_m; \boldsymbol{g}^{(t)}\right) - f\left(\boldsymbol{d}_n; \boldsymbol{g}^{(t)}\right)\right]\phi\left(\boldsymbol{g}^{(t)}\right)\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{G}_{mn}\right\} \\ =&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N}\left[f\left(\sum_{i=1}^{P}\mathrm{A}_{im}\boldsymbol{v}_i; \boldsymbol{g}^{(t)}\right) - f\left(\boldsymbol{d}_n; \boldsymbol{g}^{(t)}\right)\right]\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{G}_{mn}\right\} \\ \geq&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N}\left[\sum_{\ell=1}^{P}\mathrm{A}_{im}f\left(\boldsymbol{v}_i; \boldsymbol{g}^{(t)}\right) - f\left(\boldsymbol{d}_n; \boldsymbol{g}^{(t)}\right)\right]\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{G}_{mn}\right\} \\ =&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{i=1}^{P}\mathrm{A}_{im}f\left(\boldsymbol{v}_i; \boldsymbol{g}^{(t)}\right)\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{G}_{mn}\right\} - \frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{n=1}^{N}f\left(\boldsymbol{d}_n; \boldsymbol{g}^{(t)}\right)\mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \mathcal{G}_{mn}\right\}\end{aligned} \tag{3.34}$$

By introduce matrix $\mathrm{W} \in \mathbb{R}^{N\times T}$ with $\mathrm{W}_{nt} = f\left(\boldsymbol{d}_n; \boldsymbol{g}^{(t)}\right)$ and a tensor $\mathrm{B} \in \mathbb{R}^{T\times M\times N}$ with $\mathrm{B}_{tmn} = \mathbb{1}\left\{g_t \in \mathcal{G}_{mn}\right\}$, one finally has

$$\begin{aligned}&\overline{\mathrm{L}}\left(\mathcal{Q}_N; f, \phi\right) - \overline{\mathrm{L}}\left(\mathcal{Q}_{N+M}; f, \phi\right) \\ \geq&\sum_{n,m}\mathrm{W}_n^{\mathrm{T}}\mathrm{B}_{mn} - \sum_{i,m}\mathrm{A}_{im}\sum_{n}\mathrm{U}_i^{\mathrm{T}}\mathrm{B}_{mn}\end{aligned} \tag{3.35}$$

Without loss of generality, we choose $N = 1$, i.e., $\mathcal{D}_N$ will be extended to $\mathcal{D}_{N+1}$, then one has

$$\begin{aligned}&\overline{\mathrm{L}}\left(\mathcal{Q}_N; f, \phi\right) - \overline{\mathrm{L}}\left(\mathcal{Q}_{N+1}; f, \phi\right) \\ \geq&\sum_{n}\mathrm{W}_n^{\mathrm{T}}\mathrm{B}_n - \sum_{i}\mathrm{A}_i\sum_{n}\mathrm{U}_i^{\mathrm{T}}\mathrm{B}_n\end{aligned} \tag{3.36}$$

One can introduce the equivalent relation and two improvement operators as before. To this end, we are able to propose an algorithm summarized in alg. 3 which could help to find decision set with a fixed number of decisions efficiently.

Initialization : Number of decisions M ; set $\mathcal{D}_M^{(0)}$ randomly ; choose the tolerance factor ε
for $t = 1$ *to* T_{ex} **do**
  for $m = M + 1$ *to* K **do**
    Find $\mathcal{D}_m^{(t)}$ by applying alg. 2 for $\mathcal{D}_{m-1}^{(t)}$ for Eq. 3.36 ;
  end
  $\mathcal{D}_M^{(t+1)} \in \arg \min_{\mathcal{D}' \subset \mathcal{D}_K^{(t)}, |\mathcal{D}'|=M} \mathrm{L}\left(\mathcal{D}'; f, \phi\right)$
  if $\mathrm{L}\left(\mathcal{D}_M^{(t)}; f, R, \phi\right) - \mathrm{L}\left(\mathcal{D}_M^{(t+1)}; f, R, \phi\right) < \varepsilon$ **then**
    Break ;
  end
end
Output: Required decision set is $\mathcal{D}_M^{(t)}$;

Algorithm 3: Enhanced Improve and Branch Algorithm

The basic idea of Alg. 3 is that from a decision set $\mathcal{D}_M$ obtained from alg. 2, we first extend it to a decision set $\mathcal{D}_K$ with sufficient large cardinality. Then we select the optimal subset $\mathcal{D}_M^{'}$ of $\mathcal{D}_K$ which introduces the minimum optimality loss. It is obvious that we always have $\mathrm{L}\left(\mathcal{D}_M^{'}; f, R, \phi\right) \leq \mathrm{L}\left(\mathcal{D}_M^{(t)}; f, R, \phi\right)$. Therefore the convergence of this algorithm is guaranteed which is different from the weaker version.

3.3.5 Numerical results

In this section, we aims at showing the benefits of our proposed methods, We consider again the sum-rate capacity function $f^{\mathrm{SL}}(\boldsymbol{x}; \boldsymbol{g}) = -\sum_{i=1}^{S} \log\left(\sigma^2 + \boldsymbol{x}_i \boldsymbol{g}_i\right)$ under maximum power constraint $\sum_{i=1}^{S} \boldsymbol{x}_i \leq P_{\max}$. One can easily verify that its Hessian matrix is $\mathrm{H}_{f^{\mathrm{SL}}}(\boldsymbol{x}; \boldsymbol{g}) = -\mathrm{diag}\left\{\frac{\boldsymbol{g}_i^2}{(\sigma^2 + \boldsymbol{x}_i \boldsymbol{g}_i)^2}\right\}_i$. Therefore the utility function is concave function w.r.t. decision variable. Meanwhile the decision space is a convex polyhedron. For parameter setting, we choose number of bands $S = 6$, power budget $P_{\max} = 4\mathrm{mW}$, variance of noise $\sigma^2 = 1\mathrm{mW}$, number of parameter samples $N_{\mathrm{sample}} = 1000$, iteration number of improve and branch algorithm $\mathrm{ITER} = 1000$; iteration number for decision set extension $T_{\mathrm{ex}} = 10$ and largest cardinality of decision set is chosen as $K = 2M$.

In Fig. 3.2 , the average utility v.s. number of decisions is illustrated for optimal decision set found by fmincon by MATLAB, alg. 3 and Lloyd-Max algorithm. One can observe that the performance of the proposed algorithm always dominates the Lloyd-Max algorithm. Moreover, compared to the optimal discrete action set, the proposed method provides acceptable performance while it requires merely simple linear computation of between matrix. Besides the enhanced improve and branch algorithm requires no knowledge of the optimal decision function.

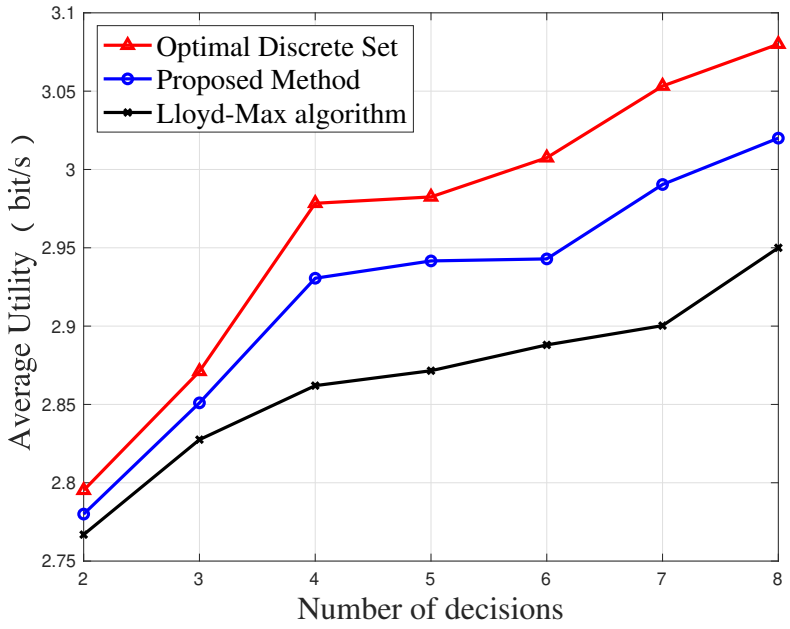


FIGURE 3.2 – Average utility (sum-rate capacity) v.s. number of decisions for Lloyd-Max Algorithm, enhanced improve and branch algorithm and optimal discrete set with given number of decisions. There are 6 bands, maximum power is $P_{\max} = 4\text{mW}$ and variance of noise $\sigma^2 = 1\text{mW}$. Without Knowledge about optimal decision function, proposed approach still yields acceptable performance compared to Lloyd-Max quantizer which entails the knowledge of regularity property assists the design of goal-oriented quantizer.

3.3.6 Conclusions

In this section, the goal-oriented quantization problem for the special case where the the decision space of the utility function is polyhedral and convex. Assuming that the utility function is concave with respect to the decision variable, one can obtain a upper bound of the empirical optimality loss by applying Jensen's inequality. Moreover, for a given parameter sample set, decision sets are divided into equivalent classes based on the optimal decision label. An algorithm which iteratively improves the decision set greedily within a equivalent class is proposed to find the best decision set which minimizes the upper bound of the optimality loss. Hereafter, we introduce generalized Jensen's inequality for so-called weakly concave utility function. By replacing the Jensen's inequality by the generalized one, analogous method could be applied to weakly concave utility function.

The advantage of the proposed algorithm is two-folds : the full knowledge of optimal decision function is unnecessary for the implementation of the algorithm ; instead of solving the original complicated optimization problem for goal-oriented quantizer, merely basic matrix calculation and repeated comparison of utility values are required to find the desired decision set. We apply the proposed algorithm to sum-rate capacity function which is well-known to be concave function w.r.t to power. Numerical results show that proposed algorithm outperforms conventional approach. Finally it is important to point out that proposed method could be redundant if the optimal decision set lies on the vertices of the decision polyhedron, e.g. binary power control in [44].

3.4 Model-Free Goal-Oriented Quantization for Unknown Utility Function

3.4.1 Motivation

In this chapter, we would like to tackle the GOQ problem under the following assumptions : only realizations of utility function are allowed to be known. Except that, one is not allowed to gain any extra information about the utility function. This scenario is frequently met in the practical applications for the two following reasons : i) the entire communication system is too complicated so that it is impossible to have an explicit expression of utility function ; ii) The expression of the utility function is unknown since the communication system is a black box, e.g. for security reasons. Therefore, we would like to design a model-free approach for this scenario so that our goal-oriented quantizer could be implemented to improve the performance of the system. Similar to the basic GOQ algorithm, the we design two approaches to solve two steps separately. To find the quantization regions for fixed decision set, we use a feed-forward network to do so. To find optimal decision set, we implement an algorithm adapted to our GOQ problem which combines two evolutionary algorithms, namely, Invasive Weeds Occupation (IWO) and Differential Evolution (DE) .

3.4.2 Finding quantization regions

The objective of this section is that, for a given decision set $\mathcal{D} = \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M\}$, one need to find the quantization region defined as

$$\mathcal{C}_m = \left\{ \boldsymbol{g} \in \mathcal{G} \,\middle|\, f\left(\boldsymbol{d}_m; \boldsymbol{g}\right) = \max_l f\left(\boldsymbol{d}_l; \boldsymbol{g}\right) \right\},\ 1 \leq m \leq M. \tag{3.37}$$

The quantization region could be regarded as the set of all parameters corresponding to the same optimal decision. For conventional quantization problem, the boundary between two adjacent quantization regions is the hyperplane which is equidistant to two corresponding centroids. For goal-oriented quantization problem, the shape of quantization region could be arbitrary (even disconnected) even if the expression of utility function is known. The only rigorous way of verifying if a parameter $\boldsymbol{g}$ belongs to the a quantization region or not is to determine its optimal decision. However to repeat this check test is not possible in practical applications since parameter space is generally infinite. Therefore, a reasonable approach is to gather some realizations of parameter and its corresponding optimal decision label to form a training set first. Based on this training set, we try to find a predictor or estimator which yields an acceptable estimation of the decision label. One possible way of doing so is to use a simple feed-forward neural network. The principle of neural network for classification is briefly explained here. The basic structure of a feed-forward network is illustrated in Fig. 3.3. $W_{i,j}^{(l)}$ is the weight between the neuron i in the l-th layer and the the neuron j in $(l+1)$-th layer and $b_j^{(l)}$ the bias term for neuron j, the relation between them is given by :

$$o_j^{(l+1)} = f_{\text{act}}\left(b_j^{(l)} + \sum_{i=1}^{N_d} W_{i,j}^{(l)} o_i^{(l)} \right), \tag{3.38}$$

where $o_j^{(l+1)}$ is the output of the neuron j and $o_i^{(l)}$ is the output of the neuron i and the input signal from neuron i to neuron j as well. Without loss of generality, we assume that number of neurons in each hidden layer is the same and denoted as N_d. $f_{\text{act}}(\cdot)$ is the activation function.

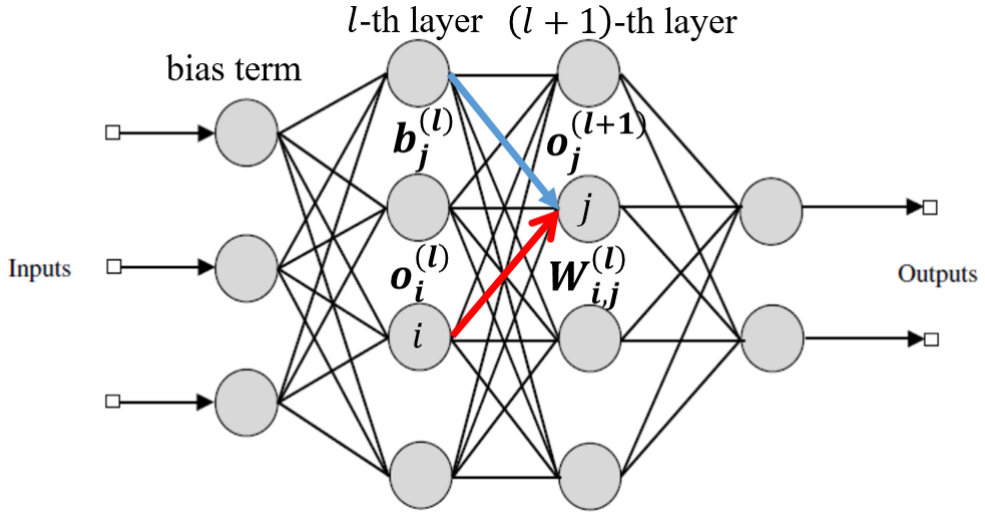


FIGURE 3.3 – Basic structure of an FNN.

We define the training set as $\mathcal{T}^{\text{FNN}} \triangleq \left\{\boldsymbol{g}^{(t)},\ \theta_t^\star\right\}_{t=1}^{N_{\text{train}}}$, where $\theta_t^\star$ is the optimal decision label corresponding to parameter realization $g^{(t)}$ obtained by exhaustive comparison between all possibles decisions :

$$\theta_t^\star \in \arg \max_{\theta \in \{1,\ldots,M\}} f\left(\boldsymbol{d}_\theta; \boldsymbol{g}^{(t)}\right) \tag{3.39}$$

If the error estimation error (test error) is less than some threshold, the FNN trained by this training set can give us a reasonable approximation of the real partition of quantization regions for a goal-oriented quantizer. We will illustrate this procedure to some important utility functions in communication system.

Applications in energy efficient MIMO system

We consider the following single user multiple- input and multiple Output (MIMO) communication system. The receiving signal is modeled by :

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{x}_{\text{SIG}} + \boldsymbol{z} \tag{3.40}$$

where $\mathbf{H}$ is the $N_r \times N_t$ channel transfer matrix with N_t transmit antennas and N_r receive antennas. We assume the entries of $\mathbf{H}$ are i.i.d. zero-mean circularly symmetric complex Gaussian distributed according to $\mathcal{CN}(0,1)$. A vector $\boldsymbol{x}$ is the transmitting symbols vector with dimension N_t and $\boldsymbol{z}$ is the receiving white Gaussian noise vector distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{N_r})$. Moreover $\mathbf{Q} = \mathbb{E}\left[\boldsymbol{x}_{\text{SIG}}\boldsymbol{x}_{\text{SIG}}^{\text{H}}\right]$ denote the covariance matrix of $\boldsymbol{x}_{\text{SIG}}$

which determines the power allocation policy. And we have the common maximum total power constraint :

$$\mathrm{Tr}\left(\mathbf{Q}\right) \leq P_{\max} \tag{3.41}$$

Given this matrix-form of the system, the energy efficiency (EE) can be defined as :

$$f^{\mathrm{MIMO}}\left(\mathbf{Q};\mathbf{H}\right) = \frac{R_0 \log_2 \left|\sigma^2 \mathbf{I}_{N_r} + \mathbf{H}\mathbf{Q}\mathbf{H}^H\right|}{\mathrm{Tr}\left(\mathbf{Q}\right) + P_0} \tag{3.42}$$

where R_0 is the raw data rate (in bits/s) and P_0 represents the power consumed by the transmitter when the radiated power is zero. For instance, in [19] it may represent the computation power or the circuit power.

The existence of P_0 is not only reasonable but also avoids the following fact that the most efficient transmission occurs when $p = \mathrm{Tr}\left(\mathbf{Q}\right) = 0$. The decision set is chosen the Equal Gain Transmission (EGT) with antenna selections. Without loss of generality, we only consider diagonal covariance matrix of the transmission signal, i.e., $\mathbf{Q} = \mathbf{Diag}\left(\boldsymbol{p}\right)$ with $\boldsymbol{p} = \left(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{N_t}\right)$. Where $\mathbf{Diag}\left(\boldsymbol{v}\right)$ generates the diagonal matrix whose diagonal is exactly the vector $\boldsymbol{v}$. The decision set is chosen as following form :

$$\mathcal{D} = \left\{ \mathbf{Q} = \frac{P_{\max}}{l}\mathbf{Diag}\left(\boldsymbol{e}\right) | \boldsymbol{e} \in \mathcal{S}_l, \ \forall l \leq N_t \right\} \tag{3.43}$$

where $\mathcal{S}_l = \left\{ \boldsymbol{e} \in \{0,1\}^{N_t} \left| \sum_{i=1}^{N_t} \boldsymbol{e}_i = l \right. \right\}$ which is the set of N_t -dimensional binary vector summing to l. The decision set $\mathcal{D}_k$ associated to a decisional quantizer with $k \leq 2^{N_t} - 1$ decisions can be constructed as follows iteratively :

$$\mathcal{D}_k = \begin{cases} \{\mathbf{Q}_1\} & \mathbf{Q}_1 \in \mathcal{D}, \ k = 1 \\ \mathcal{D}_{k-1} \cup \{\mathbf{Q}_k\} & \mathbf{Q}_k \in \mathcal{D} \backslash \mathcal{D}_k \end{cases} \tag{3.44}$$

The singleton set is chosen among all possible sets randomly. Consider the optimality of the decision set, we choose the maximum total power $P_{max} = P^*$ s.t.

$$P^* \in \arg \max_{P, \mathbf{Q} \in \mathcal{D}} \mathbb{E}_{\mathbf{H}} \left[f^{\mathrm{MIMO}}\left(\mathbf{Q};\mathbf{H}\right) \right] \tag{3.45}$$

P^* can be found by comparison through Monte-Carlo simulation. One can imagine that finding the analytical decisional quantizer for EGT will be very difficult if the dimension of the system is huge. Thus we propose to use a FNN to mimic the real decisional quantizer.

The simulation results of the MIMO system considered are presented in Fig. 3.5 ($N_t = 4$ and $N_r = 1$ (MISO), $R_0 = 10^6$ bits/s, $\sigma^2 = 5$mW, $P_0 = 10$mW and $P_{max} = 12$mW.) and in Fig. 3.6, ($N_t = 3$ and $N_r = 2$ (MIMO), $R_0 = 10^6$ bits/s, $\sigma^2 = 5$mW, $P_0 = 10$mW and $P_{\max} = 10$mW), respectively. Here, we choose the 3-hidden-layer FNN with fully connected layers comprising 20 neurons each and using the logistic activation function defined as $\mathrm{sig}(x) = \frac{1}{1+\exp(-x)}$. The number of neurons in input layer for Eq. 3.42 is given

by $2N_tN_r$ because the input vector contains the real part and the imaginary part of each entry of the transfer channel matrix. We use the *Levenberg Marquardt Algorithm* in [17] to update the weight matrix. In this FNN model, 100000 Monte-Carlo realizations will be divided into three phases : 70000 realizations for the training phase, 15000 for the validation phase and 15000 realizations for the test phase. The structure of FNN is illustrated in Fig. 3.4.

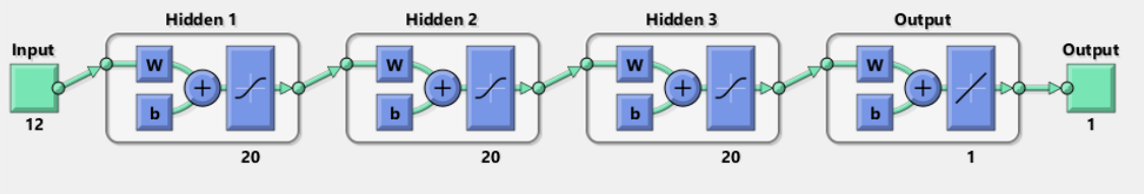


FIGURE 3.4 – Feed-forward neural network model for MIMO system ($N_t = 3$ and $N_r = 2$). Number of neurons in input layer is $2N_tN_r$.

Given the same parameter samples, a k-means quantizer which aims at minimizing the mean square error between the original signal and the quantized signal, is taken as the reference. All the realizations are divided into k regions and each region is assigned with the optimal decision in $\mathcal{D}_k$ found through exhaustive research. It is worth noting that this k-means approach can be seen as a special case implementing the basic GOQ algorithm by taking $f(\boldsymbol{x};\boldsymbol{g}) = -\|\boldsymbol{x}-\boldsymbol{g}\|^2$.

In both two cases, the goal-oriented quantizer outperforms than the k-means quantizer. In MISO scenario, NN can achieve very close performance to the optimal average utility in several decision set ($\mathcal{D}_k$, $k = 2, 5, 6, 7$ and $k \geq 9$) while the average utility found through k-means quantizer is trite. In MIMO scenario, the performance of NN is still better than k-means quantizer. The utility loss introduced by the FNN is perhaps owing to the scarcity of training.

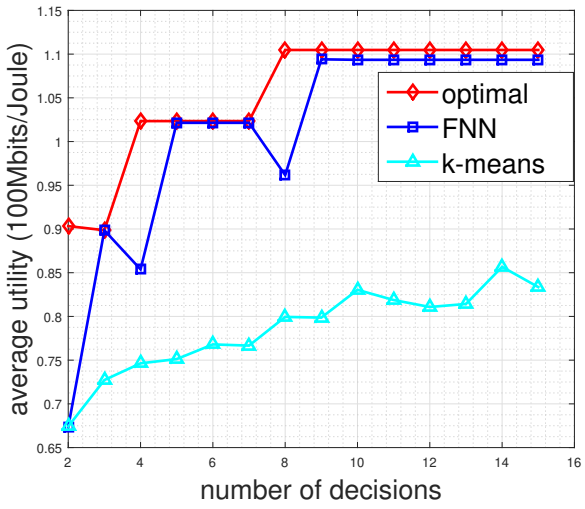


FIGURE 3.5 – Average utility v.s. number of decisions for $N_t = 4$ and $N_r = 1$ (MISO). Here $\sigma^2 = 5$mW, $P_0 = 10$mW and $P_{\max} = 12$mW. FNN is better than k-means quantizer and close to theoretical optimum.

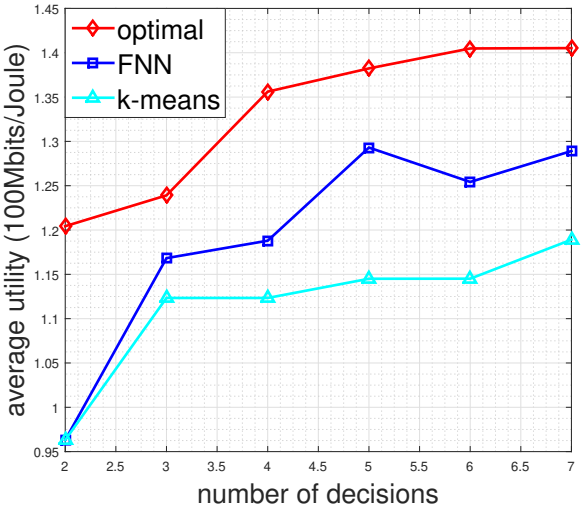


FIGURE 3.6 – Average utility vs. number of decisions for $N_t = 3$ and $N_r = 2$ (MIMO) , $\sigma^2 = 5\text{mW}$, $P_0 = 10\text{mW}$ and $P_{\max} = 10\text{mW}$. FNN is better than k-means quantizer and close to theoretical optimal utility.

Comparison between energy efficiency and sum-rate capacity

We consider the following EE function :

$$f^{\text{MB}}(\boldsymbol{p};\boldsymbol{g}) = \frac{\sum_{i=1}^{N} \exp\left(-\frac{c\sigma^2}{\boldsymbol{p}_i \boldsymbol{g}_i}\right)}{\sum_{i=1}^{N} \boldsymbol{p}_i} \tag{3.46}$$

where $i \in \{1, 2, \ldots, N\}$ is an index which might represent the band, channel, or user index; $g_i > 0$ is the channel gain of i-th channel, $\boldsymbol{p} = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N)$ is the power allocation vector; $\boldsymbol{g} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_N)$ is the vector channels used by transmitter i; there is no interference appears between bands, where σ^2 is the received noise variance and $c \geq 0$ is a parameter related to spectral efficiency (see[71]). Apart from the EE function, we consider the sum-rate capacity as follows :

$$f^{\text{SR}}(\boldsymbol{p};\boldsymbol{g}) = \sum_{i=1}^{N} \log(1 + \frac{\boldsymbol{p}_i \boldsymbol{g}_i}{\sigma^2}) \tag{3.47}$$

For EE defined in Eq. 3.46, the number of input neurons for Eq. 3.46 is obviously the number of bands N. Fig. 3.7 illustrates the decision (quanization) regions for the following simulation configuration : there are two bands in the system ($N = 2$), every band has only two choices to choose : $P_{min} = 2\text{mW}$, $P_{max} = 3\text{mW}$. The noisy level is set to be $\sigma^2 = 10\text{mW}$ and the constant is assumed to be $c = 1$. The channel gain g_i in band i is assumed to be exponentially distributed, i.e., its p.d.f. is $\phi(\boldsymbol{g}_i) = \exp(-\boldsymbol{g}_i)$. There follows our intuitive explanation. Let us take the orange region (P_{min}, P_{max}) as an example. In this region, channel gain $\boldsymbol{g}_1$ is smaller than $\boldsymbol{g}_2$ which means transmission in band 1 is less efficient than band 2, therefore the transmitter chooses the policy (P_{min}, P_{max}). Same principle can be applied to the 3 remaining regions.

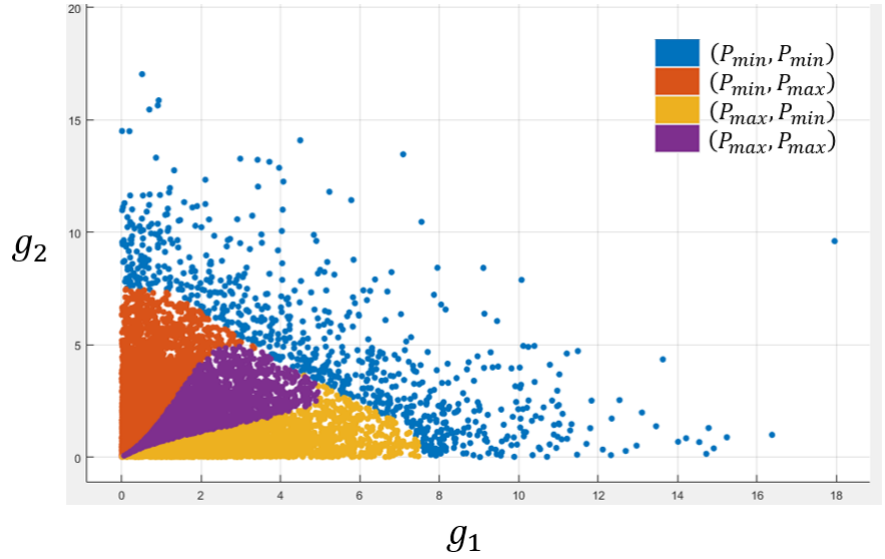


FIGURE 3.7 – Quantization regions of goal-oriented quantizer for 2-band energy efficiency problem. When one channel is dominant, it is better to transmit with higher power levels in that dominant channel. Otherwise, both transmitters choose the same transmit power.

To compare the performance of goal-oriented quantizer by fuond by FNN, we define the relative optimality loss introduced by quantization as following :

$$\sigma\,(\%) = \mathbb{E}_{\boldsymbol{g}}\left[\left|\frac{f^{\star}\left(\boldsymbol{g}\right) - f^{\mathrm{NN}}\left(\boldsymbol{g}\right)}{f^{\star}\left(\boldsymbol{g}\right)}\right|\right] \times 100 \tag{3.48}$$

where $f^{\mathrm{NN}}\left(g\right)$ is the performance achieved by our learning approach. Besides, to compare the influence of the compression between the system with different objectives, Define the compression rate $\gamma\left(\sigma\right)$ of a given relative optimality loss σ as $\gamma\left(\sigma\right) := \frac{M(1\%)}{M(\sigma)}$, where $M\left(\sigma\right)$ is the required number of decisions such that the relative optimality loss σ can be satisfied.

Fig. 3.8 illustrates the compression rate γ in function of optimality loss for two bands in two cases. With the two different utilities, it can be seen that the compression rate increases as the optimality loss grows. For the energy efficiency problem, the compression rate decreases slowly while the optimality loss decreases and the loss is always greater than 1%. As for the sum-rate capacity, the compression rate declines rapidly while the optimality loss reduces and the optimality loss is always less than 1%. It can be observed that it is easier to compress the parameter $\boldsymbol{g}$ for the sum-rate problem than the energy efficiency in two-band scenario, i.e., the energy efficient function is more sensitive to the variable $\boldsymbol{g}$. This can be explained by the fact that the explicit optimal decision function of sum-rate, well known as the water-filling solution, is more concise than the solution of the energy efficiency problem, which is inversely proportional to parameter $\boldsymbol{g}$. The difference in compression difficulty between sum-rate capacity function and energy efficiency confirms that fine quantization could be extravagant for some scenario to achieve a certain optimality loss.

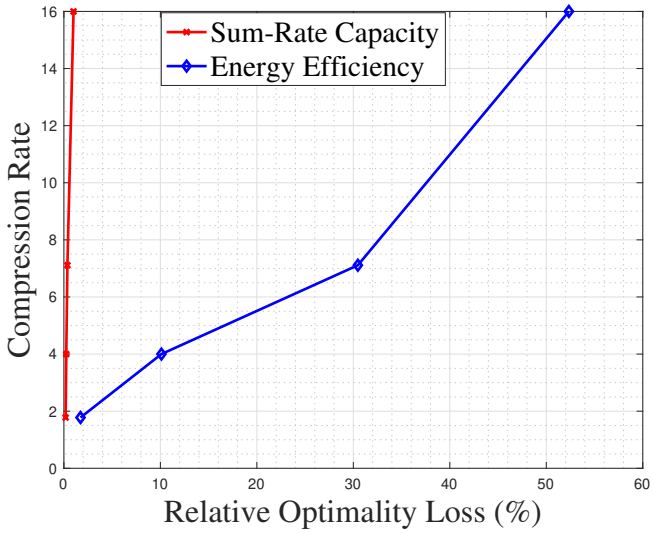


FIGURE 3.8 – The compression rate as a function of the relative optimality loss (%)for single user 2-band scenario for energy efficiency and sum-rate capacity. Compressing the channel gain for sum-rate capacity function is easier than compressing the channel gain for energy-efficiency function.

3.4.3 Finding optimal decision set

The objective of this subsection is to find the optimal decision set $\mathcal{D}^\star = \{\boldsymbol{d}_1^\star, \ldots, \boldsymbol{d}_M^\star\}$ for a goal-oriented quantizer. The basic GOQ algorithm suggests that we should solve M Equations separately :

$$\boldsymbol{d}_m^\star \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} \int_{g\in\mathcal{C}_m} f(\boldsymbol{x};\boldsymbol{g})\,\phi(\boldsymbol{g})\,\mathrm{d}\boldsymbol{g}, \text{ for } \forall 1 \leq m \leq M. \tag{3.49}$$

As we have explained before, the boundary of quantization regions $\{\mathcal{C}\}_{m=1}^{M}$ of a goal-oriented is hard to determine precisely. In other words, to solve eq. 3.49 would be extremely difficult separately. Therefore, it would be reasonable to find the optimal decision set jointly since the average optimality loss of a goal-oriented quantizer is connected to the decision set directly.

To make the computation tasks more convenient to express, we introduce the matrix notation $\mathrm{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M]$ constructed from the decision set $\mathcal{D} = \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M\}$. We will use an evolutionary algorithm called IWO-DE to tackle the puzzle by using merely the realizations of utility function. IWO-DE algorithm is firstly proposed in [67] by combining Invasive Weeds Occupation (IWO) in [65] and Differential Evolution (DE) in [66] which are essentially two evolutionary algorithms. IWO algorithm are known to be very efficient when the search space is relatively large. Evolutionary algorithms have been widely used in many areas with its benefits such as simple computation, robustness and etc (see [61]). This algorithm comprises the following steps :

- **Initialization :** randomly choose W decision sets in the search space : $\mathrm{D}_1^{(0)}, \ldots, \mathrm{D}_W^{(0)}$ as the primitive population. W is called the population size. $\mathrm{D}_k^{(t)}$ denotes the k-th individual of the t-th generation.
- **Reproduction** : the k-th individual reproduces its offspring according to its em-

pirical optimality loss at t-th generation $\overline{\mathrm{L}}\left(\mathrm{D}_k^{(t)}\right)$ by Monte-Carlo simulation. The number of offspring for k-th individual at $(t+1)$-th generation $S_k^{(t+1)}$ is given by :

$$S_k^{(t+1)} = \upsilon\left(\mathrm{D}_k^{(t)}\right)\left[S_M - S_m\right] + S_m \tag{3.50}$$

where

$$\upsilon\left(\mathrm{D}_k^{(t)}\right) = \frac{\overline{\mathrm{L}}\left(\mathrm{D}_k^{(t)}\right) - \max_i \overline{\mathrm{L}}\left(\mathrm{D}_i^{(t)}\right)}{\min_i \overline{\mathrm{L}}\left(\mathrm{D}_i^{(t)}\right) - \max_i \overline{\mathrm{L}}\left(\mathrm{D}_i^{(t)}\right)}, \tag{3.51}$$

S_M and S_m are respectively the maximum and minimum numbers of offspring that an individual is allowed to reproduce. Eq. 3.50 and Eq. 3.51 assure that solution candidates with smaller OL are encouraging to produce more offspring than others.

- **Spatial Dispersion :** for k-th individual, its offspring obey a Gaussian distribution $\mathcal{O}_k^{(t)} \sim \mathcal{N}\left(\mathrm{D}_k^{(t)}, [\mu^{(t)}]^2\right)$. Every individual reproduces its offspring in the feasible set till it achieves the number given by Eq. 3.50. $\mu^{(t)}$ is the standard deviation for every entry of $\mathrm{D}_k^{(t)}$ controlling the divergence of the dispersion. The evolution of $\mu^{(t)}$ through the generations is given by :

$$\mu^{(t)} = \left(\frac{T-t}{T}\right)^{\rho}\left[\mu^{ini} - \mu^{end}\right] + \mu^{end} \tag{3.52}$$

where ρ is called the nonlinear index and μ^{ini} and μ^{end} stands for the initial and final standard derivation, respectively. In general, we should have $\mu^{ini} \gg \mu^{end}$ in order to avoid dropping into a local maximum and $\mu^{end} \to 0$ to increase the accuracy near the potential global optimum.

- **Competitive Exclusion :** sort all the offspring together with their parental individuals in ascending order according to their empirical loss. Then select the W first offspring as the original material for next generation : $\Phi_1^{(t)}, \ldots, \Phi_W^{(t)}$ s.t. $\overline{\mathrm{L}}\left(\Phi_1^{(t)}\right) \leq \cdots \leq \overline{\mathrm{L}}\left(\Phi_W^{(t)}\right)$.
- **Mutation :** there are many different differential evolutionary strategies for creating mutations. For example, for the k-th potential individual, we create its possible mutant by : $\Psi_k^{(t)} = \Phi_{\mathrm{idx}_1}^{(t)} + F_0\left(\Phi_{\mathrm{idx}_2}^{(t)} - \Phi_{\mathrm{idx}_3}^{(t)}\right)$, where F_0 is called the scaling factor. And we further choose $\mathrm{idx}_1 = 1$ (the best one), $\mathrm{idx}_2 = rand\,(2, W)$ and $\mathrm{idx}_3 = rand\,(2, W)$ with $\mathrm{idx}_2 \neq \mathrm{idx}_3$ and $\mathrm{idx}_2, \mathrm{idx}_3 \neq k$.
- **Crossover :** for the l-th decision of the k-th individual at next generation $d_{k,l}^{(t+1)}$, we let

$$d_{k,l}^{(t+1)} = \begin{cases} \psi_{k,l}^{(t)}, & y_l \leq C_r \text{ or } l = I_r \\ \phi_{k,l}^{(t)}, & \text{otherwise} \end{cases}$$

where $d_{k,l}^{(t+1)}, \psi_{k,l}^{(t)}$ and $\phi_{k,l}^{(t)}$ is the l-th decision of $\mathrm{D}_k^{(t+1)}, \Psi_k^{(t)}$ and $\Phi_k^{(t)}$ respectively, y_l is a random variable uniformly distributed over $[0, 1]$, C_r is called the crossover probability and I_r is a randomly chosen index so that the mutant decision set can't be identical to the original one.

Selection Operation : only mutant which reduces empirical loss, i.e., $\overline{\mathrm{L}}\left(\mathrm{D}_k^{(t+1)}\right) < \overline{\mathrm{L}}\left(\Phi_k^{(t)}\right)$ will be conserved. Otherwise, $\mathrm{D}_k^{(t+1)} = \Phi_k^{(t)}$. If the initial population is well selected, the population of decision sets will converge to the optimal direction set $\mathrm{D}^\star$ for a sufficiently large number of generations.

Applications in an energy-efficient MIMO systems

The considered communication scenario comprises a multi-antenna transmitter which has to adapt the transmit power $p \in [0, P_{\max}]$ and its unit beamforming vector $\boldsymbol{\omega} \in \mathbb{C}^{N_t \times 1}$ ($\|\boldsymbol{\omega}\| = 1$) to the realization of the $N_r \times N_t$ channel transfer matrix $\mathbf{H}$, N_t and N_r being respectively the number of transmit antennas and receive antennas. The action or decision of the transmitter is thus given by the pair $\boldsymbol{x} \triangleq (p, \boldsymbol{\omega})$. The objective of the transmitter is to maximize its energy-efficiency by adapting its decision to the channel. A very common measure of energy-efficiency is given by the ratio of a benefit function (e.g., the packet success rate or a measure of the transmission rate) to a cost power (e.g., an increasing function of the radiated power). The assumed utility function has the following form :

$$f^i(\boldsymbol{x}; \mathbf{H}) := \frac{V^i(\boldsymbol{x}; \mathbf{H})}{C(\boldsymbol{x})} \tag{3.53}$$

where $V^i(\boldsymbol{x}; \mathbf{H})$ is the transmission benefit obtained from choosing decision x over a channel matrix $\mathbf{H}$ and $C(\boldsymbol{x})$ the transmission cost of using decision $\boldsymbol{x}$; i stands for the considered case index, the two cases being defined just next. Indeed, for the *benefit function*, we will use one of the following functions :

- *Case I benefit function (channel capacity) :* $V^{\mathrm{I}}(p, \boldsymbol{\omega}; \mathbf{H}) = \log\left(1 + \frac{p\|\mathbf{H}\boldsymbol{\omega}\|^2}{\sigma^2}\right)$ (see e.g., [30][58]).
- *Case II benefit function (packet success transmission rate)* : $V^{\mathrm{II}}(p, \boldsymbol{\omega}; \mathbf{H}) = R_0 \exp\left(-\frac{c\sigma^2}{p\|\mathbf{H}\boldsymbol{\omega}\|^2}\right)$ introduced in [30], where $c > 0$ is a constant related to the spectral efficiency of the system and R_0 the raw transmission rate.

A well-admitted transmission *cost function* is as follows [63] :

$$C(\boldsymbol{x}) = C(p, \boldsymbol{\omega}) = p + P_0 \tag{3.54}$$

where P_0 represents a static cost such as the circuit power or the computation power.

Let respectively denote by M_1 and M_2 the cardinalities of the power level set and the beamforming vector set. These sets are denoted by : $\mathcal{P} = \{p_1, \ldots, p_{M_1}\}$ and $\Omega = \{\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_{M_2}\}$. We define the required *amount of feedback information* to take a decision by $B_i = \log_2 M_i$, which expresses in bit per decision.

In 5G networks, one desirable scenario will be to be able to maximize energy-efficiency under some QoS constraints e.g., for URLLC [55, 56]. Obviously, the choice of the transmission decision set can have an impact on the QoS. This is the reason why we should introduce a transmission reliability constraint for the forward communication link (transmitter $\rightarrow$ receiver) and a delay constraint for the reverse of feedback communication link (receiver $\rightarrow$ transmitter). If the data rate from the transmitter to the receiver has to exceed the minimum rate r_0, this induces a constraint on the benefit function V^i. Equally, if the maximum delay to transfer the channel state information from the receiver to the transmitter is t_0, the sum information-rate therefore has to meet the constraint $B_1 + B_2 \leq Rt_0$, R being the available feedback channel rate. Having introduced these notations and made these observations, the decision set OP writes in the case of energy-efficient power control

and beamforming as :

$$\begin{aligned}\max_{B_1,B_2,\mathcal{P},\Omega}\ & \mathbb{E}_{\mathbf{H}}\left[\frac{V^i\left(\widehat{p}^{\star}_{\mathcal{P}}\left(\mathbf{H}\right),\widehat{\boldsymbol{\omega}}^{\star}_{\Omega}\left(\mathbf{H}\right);\mathbf{H}\right)}{\widehat{p}^{\star}_{\mathcal{P}}\left(\mathbf{H}\right)+P_0}\right]\\ \text{s.t.}\ & -\mathbb{E}_{\mathbf{H}}\left[V^{\mathrm{II}}\left(\widehat{p}^{\star}_{\mathcal{P}}\left(\mathbf{H}\right),\widehat{\boldsymbol{\omega}}^{\star}_{\Omega}\left(\mathbf{H}\right);\mathbf{H}\right)\right]+r_0\leq 0\\ & B_1+B_2-Rt_0\leq 0\end{aligned} \tag{3.55}$$

where

$$\widehat{\boldsymbol{\omega}}^{\star}_{\Omega}\left(\mathbf{H}\right)\in\arg\max_{\boldsymbol{\omega}\in\Omega}\ \left\|\mathbf{H}\boldsymbol{\omega}\right\|^2 \tag{3.56}$$

and

$$\widehat{p}^{\star}_{\mathcal{P}}\left(\mathbf{H}\right)\in\arg\max_{p\in\mathcal{P}}\ \frac{V^i\left(p,\widehat{\boldsymbol{\omega}}^{\star}_{\Omega}\left(\mathbf{H}\right);\mathbf{H}\right)}{p+P_0}. \tag{3.57}$$

The conventional approach consists in quantizing the channel state and reporting the corresponding information to the transmitter. In most real systems and existing standards, uniform quantization is implemented. Here, we consider a more advanced quantizer namely, the Lloyd-Max (LM) quantizer in [68]. Essentially, the LM quantizer consists in determining the quantization cells and representatives in an iterative manner to minimize the distortion $\mathbb{E}\left[\|\boldsymbol{g}-\widehat{\boldsymbol{g}}\|^2\right]$, $\widehat{\boldsymbol{g}}$ being the quantized channel. This quantized information is then used by the transmitter to maximize its utility function $f^i(\boldsymbol{x};\boldsymbol{g})$. We will refer to this algorithm as the *"best conventional approach in SOTA"*. Moreover, the random vector quantization (RVQ) scheme should be taken as reference as well which is proved to be near-optimal for moderate information feedback of capacity maximization problem in [48]. For the simulation setting, we will consider a typical scenario defined by : $N_t=4$; $N_r=1$; $r_0=3\times10^5 bps$, $t_0=0.01s$; $R_0=10^6 bps$; $c=0.1$; $P_0=0.5\text{mW}$; $P_{\max}=1\text{mW}$; $\sigma^2=1\text{mW}$. Similarly, for the the IWO-DE algorithm, a typical setting (in coherence with related evolutionary algorithms) will be assumed as in Table 3.1.

Parameters	Value
Population size W	10
number of generations T	400
max number of offspring S_M	20
min number of offspring S_m	10
Non-linear index γ	2.5
Initial standard derivation μ^{ini}	$\frac{1}{N_t}$
Final standard derivation μ^{end}	$\frac{1}{200N_t}$
Scaling factor F_0	0.9
Crossover probability C_r	0.9

TABLE 3.1 – Parameter setting for IWO-DE algorithm

In order to clarify the impact of the power and beamforming separately, we consider two following different situations :

1. When the influence of B_i is assessed, B_j $(j\neq i)$ is fixed.
2. We fix the total number of quantization bits $B=B_1+B_2$.

First of all, we fix $B_1=4$ bits and analyze the influence of B_2 to the relative optimality loss. To compare our approach with the conventional approach, Fig. 3.9 illustrates the

required amount of information for beamforming quantization to achieve a given relative optimality loss of case I utility function. For a given same optimality loss, remarkably, one can observe that with our approach one can *reduce by a factor 2 the amount of beamforming bits* with respect to the LM quantizer and random vector quantization. In addition, if the number of bits allocated to beamforming quantization is quite small, the relative optimality loss remains acceptable for the proposed approach while it is large for other existing solutions. Moreover, to explore the impact of utility function on beamforming quantization. Fig. 3.10 compares the required B_2 to achieve a given optimality loss between the EE for channel capacity and the EE for packet success transmission rate (PSTR). By implementing the proposed quantization scheme, the minimum number of bits for EE of PSTR is larger than the EE for channel capacity for achieving the same performance which may suggests that the EE for PSTR is slightly sensitive to the quality of quantization than EE of channel capacity and thus worth more feedback bits and better beamforming code book design techniques.

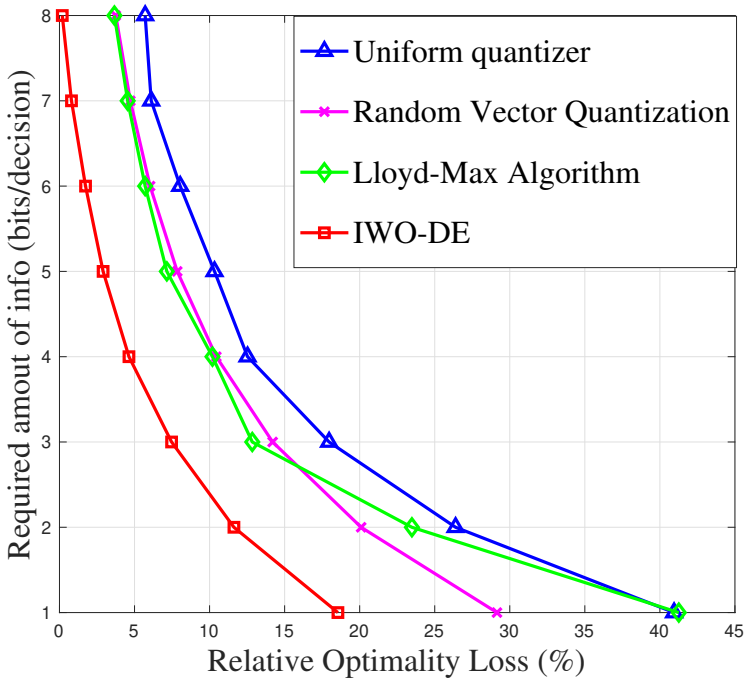


FIGURE 3.9 – Required bits of quantization v.s. relative optimality loss for Case II (packet success transmission rate as benefit of energy efficiency) with $B_1 = 4$ bits/decision. The benefits from using our algorithm is very apparent on this figure. For example, for an optimality loss of 5% between the perfect CSI case and the finite-rate feedback case, the amount of information needed to perform beamforming can be reduced by around 2 by moving from the best state-of-the-art approach to the proposed approach.

To see the influence of the power level quantization, we fix the bits of beamforming quantization as $B_2 = 6$ and vary the bits for power level feedback from 1 to 8. Fig. 3.11 shows the evolution of required amount of information for power quantization as function of relative optimality loss. For EE of channel capacity , different from EE of case II, increasing the number of bits for power quantization have less important impact on

performance of the system. This improvement is always modest while the improvement is firstly sharp when few bits are available but then becomes modest with enough number of bits provided for EE of PSTR. Thus if the bits for beamforming quantization are sufficient even one bit feedback information about power provides acceptable performance for EE of case I. Up to now, We can conclude that EE of PSTR is sensitive to both the quality of beamforming quantization and power quantization combing the observation in Fig. 3.9 and Fig. 3.10. We need to further find the optimal bits allocation policy for EE of PSTR.

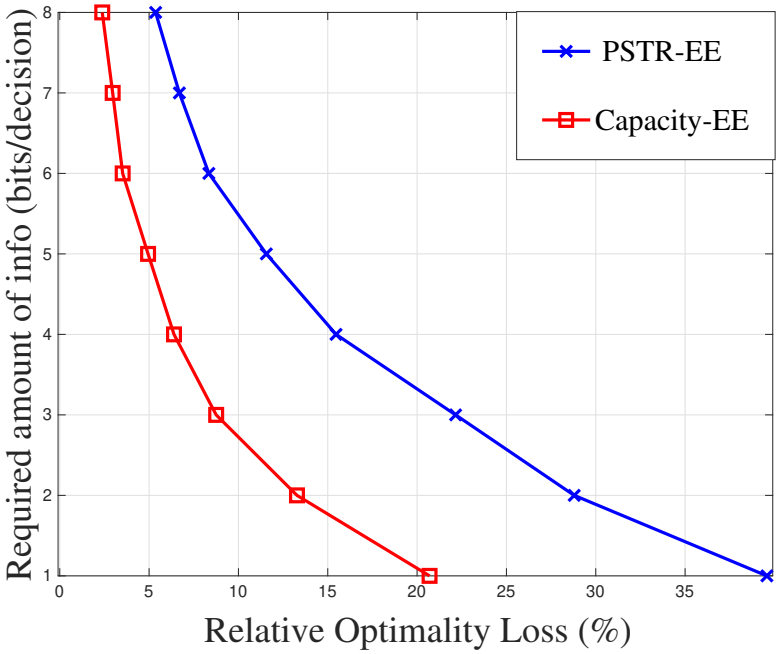


FIGURE 3.10 – Required bits of quantization v.s. relative optimality loss for utility function of case I (capacity as benefit function of EE) and case II (packet success transmission rate as benefit function of EE) with $B_1 = 4$ bits/decision. Here, it is seen that considering the packet success rate as benefit function of EE requires more feedback resources than using the capacity function. Remarkably, it is possible to quantify this extra amount of resources..

According to the precedent observations, finding a proper allocation policy between B_1 and B_2 is necessary. In order to determine the optimal allocation of bits for EE with PSTR as the benefit function, we assume that the total quantization bits are fixed so that the transmitter merely seeds the essential information back to the receiver. We fix the total number of bits for quantization as $B = 8$ (exactly one byte). Fig. 3.12 shows the evolution of energy efficiency of case II as function of quantization bits used for beamforming. To achieve the best performance, among 8 total quantization bits, we should allocate 3 bits for beamforming quantization and 5 bits for power quantization. Moreover, for all methods, sufficient number of bits should be conserved to beamforming quantization by observing the sharp decay of average utility for $1 \leq B_2 \leq 3$. Finally, even no information provided for power level ($B_2 = 8$), the energy efficiency achieved by our proposed approach and RVQ is acceptable which shows the importance of quantizing directly on the decision itself instead of quantizing the CSI in the conventional approach.

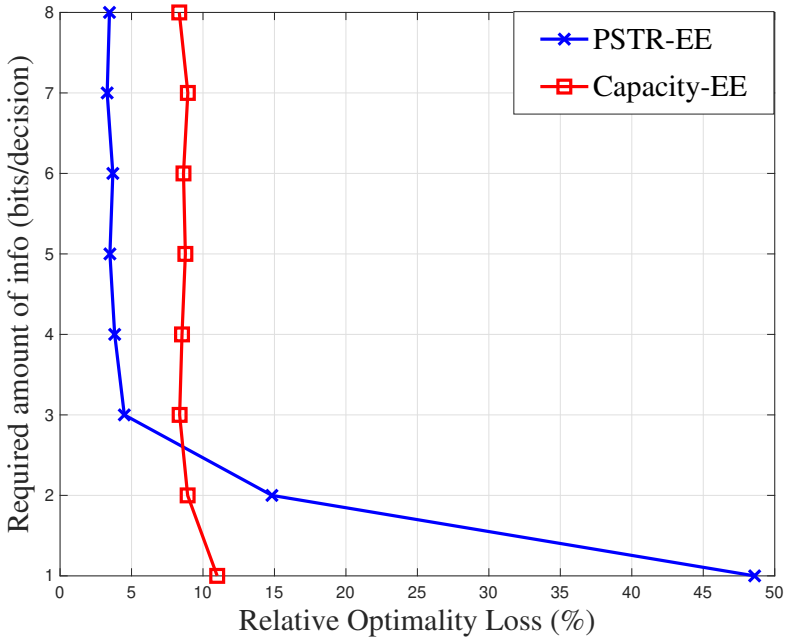


FIGURE 3.11 – Required bits of quantization v.s. relative optimality loss for utility function of case I (capacity as benefit function of EE) and case II (packet success transmission rate as benefit function of EE). Here, with $B_2 = 6$ bits/decision, the curves are much steeper, indicating that the choice of the number of feedback rate is more critical in this regime as soon as small optimality losses are desired.

3.4.4 Conclusions

In this section, we try to solve the goal-oriented quantization problem when only the realizations of cost (utility) functions are allowed to use. The problem is divided into two steps. To find optimal quantization region fixing the decision set, a feed-forward neural network is proposed to do so. When applied to the problem of power allocation, it is seen that quantizing the channel gains very roughly only induces a very small optimality loss w.r.t. the case where the gains are perfectly known to the transmitter when the utility is the transmission rate. However, for energy-efficiency, channel gains need to quantized more accurately. Using a classical distortion-based quantization scheme (k-means quantization) for this is shown to lead to a quite significant performance loss (about 30%), showing the potential of our approach. To better assess the potential of the proposed approach, it should be generalized to goal-oriented source coding and goal-oriented channel coding. Also, it allows one to reconsider the overarching assumption made in resource allocation problem, that is the resource allocation policy is designed by assuming perfect knowledge of the parameters. Mathematically, a deep study should be developed to identifying the properties of the utility function which represents its sensitivity to being maximized under imperfect knowledge of its parameters.

To find optimal decision set, an evolutionary algorithm called IWO-DE algorithm which combines two classic evolutionary algorithms is used. A problem of finding jointly the optiaml decision set of power level and beamforming vectors for energy-efficient communications is taken as an example of our proposed method. While the problem is relatively easy to solve when decisions can be continuous, the problem needs to be formu-

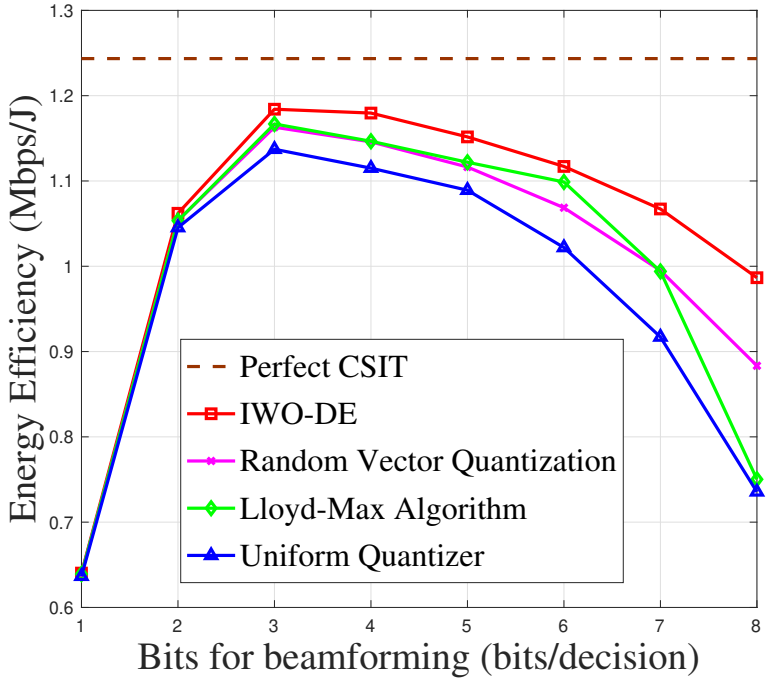


FIGURE 3.12 – Energy efficiency v.s. bits of quantization for beamforming (B_2) for case II (packet success transmission rate as benefit function of EE). The conventional approach is sensitive to the available amount of feedback information for beamforming when power level quantization is rough while the proposed approach offers good performance for a large range of feedback rates.

lated properly when the decision set is imposed to be finite. Our approach is shown to outperform the best state-of-the-art techniques such as Lloyd-Max algorithm and RVQ. Obviously, our approach needs to be explored and developed further. In particular, when the system dimensions increase, complexity issues need to be considered. When there is interference, the proposed framework needs to be extended. In the presence of interactions between the decision-makers, other issues such as Braess's paradox may arise and make the problem even more challenging.

4

High Resolution Analysis of Goal-Oriented Quantization

In this chapter, the goal-oriented qunatization is studied under high-resolution regime. Our approach is to use high-resolution quantization theory in a large rate case, assuming that the probability density of the input is approximately constant across any particular input bin. This approximation provides tractable equations for the performance, and could facilitate the characterization of the relationship between the performance and the quantization. The scalar case and the vector case are treated separately. For scalar case, the proposed new approximate formula of optimality loss leads to a new quality defined as value density representing the importance of parameter. We introduce a new quality called normalized optimality loss when comparing the hardness of quantization for different cost functions. By merely approximating this quality in high resolution regime, we are capable to determine the hardness of quantization for different utility functions without performing real simulations. For vector case, a cell-independent approximated formula for optimality loss is no longer possible since the optimal tessellating cell shape is unknown. Nevertheless, by admitting the Gersho's conjecture, an upper bound and lower bound are derived for optimality loss when the dimension of parameter is smaller than the dimension of decision variable. Moreover, we propose a new algorithm which iteratively updates the representatives based on the eigenvalue approximation of the optimality loss. Proposed algorithms could be extended to cost function with constraints as well.

4.1 Motivation and Related Works

In previous chapter, we have provided a model-free approach for goal-oriented quantization where only the value of utility function is known while other information is missing. The advantage of this approach is its independence of choice of utility function and its ability of gradual learning on the utility function. However, this advantage leads to a

serious drawback as well. The learning essence of this approach prevents us from deeply understanding the difference between utility function since no explanation is given for energy efficiency being much harder to be quantized than sum-rate capacity. Moreover, this example shows that the regularity properties of utility function have at least huge impact on the compressibility of cost function even if not decisive. Therefore, it is reasonable to take the regularity property into account to design the optimal goal-oriented quantizer. Fortunately as we will show in this chapter the impact of regularity properties of utility function can be easily characterized in high resolution regime. Another fact motivating us to start our theoretical analysis in high-resolution regime is inspired from conventional quantization minimizing the squared-error measure. Before going into details, we briefly recall some basic discovery for high-resolution quantization theory. In [20], Bennett first applied this approximation in developing a system performance formula for scalar quantizers, referred to as Bennett's integral. Specially for vector quantization, Gersho's paper [22] extended Bennett's work to the vector quantization and introduced lattice vector quantization to achieve the asymptotically optimal quantizer point density for entropy-constrained vector quantization for a random vector with bounded support. In that paper, Gersho also made the famous conjecture on tessellation, which we presume its correctness in this manuscript. Recently, this approximation is also used in several signal processing applications. [38] considers the development of a general framework for the analysis of transmit beamforming methods in multiple-antenna systems with finite-rate feedback. Tight lower and upper bounds of the average asymptotic distortion are derived by extending the vector version of the Bennett's integral. A characterization of the optimal quantizer through its interval density and an analytical expression for the Fisher information are obtained in [39]. Inspired by the achievement in these works, we resort to high-resolution quantization theory to characterize the optimality loss induced by GOQ, and exploit these results obtained in the high resolution regime to understand the relationship between the goal and the goal-oriented quantizer in general cases. Through out this chapter, **cost functions** are assumed as the objective of the GOQ for notation conventions.

4.2 Scalar High-Resolution Quantization

To start with, we consider the scalar case where $d_1 = d_2 = 1$. We denote $\mu\{\cdot\}$ the Lebesgue measure. The interior and the boundary of a set is denoted as $\mathrm{int}(\cdot)$ and $\mathrm{bd}(\cdot)$ respectively. To make our problem traceable, the following assumptions are made :

1. Cost function $f(x;g)$ has partially derivative w.r.t. x for order $K \geq 2$.
2. $\psi(g)$ is differentiable and $\mu\left\{g : \frac{\mathrm{d}\psi(g)}{\mathrm{d}g} = 0\right\} = 0$.
3. $\forall g \in \mathcal{G}$, $\psi(g) \in \mathrm{int}(\mathcal{X})$.
4. For $\forall i \leq K$, $\int_{g\in\mathcal{G}} \left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^i \frac{\partial^i f(\psi(g);g)}{\partial x^i}\phi(g)\mathrm{d}g < +\infty$.

Assumption (2) actually excludes all cost functions with independent decision of parameter or finite optimal decision space. Assumption (3) is to limit our discussion in unconstrained case in the first time. The extension to constrained case will be discussed in the end of this section. For any point g, define its distance to the closest quantization point

by $\Delta(g)$:

$$\Delta(g) = \min_{1\leq m\leq M} \|g - z_m\| \tag{4.1}$$

At the high resolution regime, i.e., the number of representatives, M, tends to infinity ($M \to \infty$) we are able to introduce a density function $\rho(g)$, which represents the density of the representatives :

$$\rho(g) = \lim_{M\to+\infty} \frac{1}{M\Delta(g)} \tag{4.2}$$

As a consequence, the number of representatives in any interval $[a, b]$ can be approximated by $M\int_a^b \rho(g)\mathrm{d}g$. We further define $C(k) = \frac{1}{k!(k+1)2^k}$, optimality loss in high-resolution regime can be thus approximated by

$$\begin{aligned}
&\mathrm{L}(\mathcal{Q}; f, R, \phi)\\
=&\sum_{m=1}^{M}\int_{g\in\mathcal{G}_m} [f(\psi(z_m); g) - f(\psi(g); g)]\,\phi(g)\,\mathrm{d}g\\
\overset{(a)}{=}&\sum_{m=1}^{M}\int_{g\in\mathcal{G}_m} (\psi(z_m) - \psi(g))^k \frac{1}{k!}\frac{\partial^k f(x; g)}{\partial x^k}|_{x=\psi(g)}\phi(g)\mathrm{d}g + o\left(M^{-k}\right)\\
\overset{(b)}{=}&\sum_{m=1}^{M}\int_{g\in\mathcal{G}_m} (z_m - g)^k \left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{1}{k!}\frac{\partial^k f(\psi(g); g)}{\partial x^k}\phi(g)\mathrm{d}g + o\left(M^{-k}\right)\\
\overset{(c)}{=}&\int_{g\in\mathcal{G}} \frac{\Delta^k(g)}{(k+1)2^k}\left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{1}{k!}\frac{\partial^k f(\psi(g); g)}{\partial x^k}\phi(g)\mathrm{d}g + o\left(M^{-k}\right)\\
\overset{(d)}{=}&\frac{C(k)}{M^k}\int_{g\in\mathcal{G}} \rho^{-k}(g)\left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{\partial^k f(\psi(g); g)}{\partial x^k}\phi(g)\mathrm{d}g + o\left(M^{-k}\right)
\end{aligned} \tag{4.3}$$

where k is defined as

$$k \triangleq \min\left\{ i \in \mathbb{N} \,\middle|\, \forall g,\ \frac{\partial^i f(x; g)}{\partial x^i}|_{x=\psi(g)} \neq 0 \text{ a.s.} \right\} \tag{4.4}$$

(a) follows from the fact that the higher order terms in the Taylor expansion of $(f(\psi(z_m); g) - f(\psi(g); g))$ are negligible to the k-th order term with $o(\cdot)$ representing the infinitesimal; (b) follows from the fact that the higher order terms in the Taylor expansion of $(\psi(z_m) - \psi(g))$ are negligible to first-order term. (c) follows from the fact $\mathbb{E}[(g - z_m)^k | g \in \mathcal{G}_m]$ can be approximated by $\frac{\Delta^k}{(k+1)2^k}$ (see [20][21]) and the sum is approximately equal to the integral when M tends to infinite due to the definition of a Riemann integral; (d) follows results of high resolution quantization.

Remark 4.2.1. *It is generally assumed that integer k be even. In fact, the first reason is that most cost functions have even k since we are solving a minimization optimization problem. The second reason is that odd k has a very smaller influence to $\mathrm{L}(\mathcal{Q}; f, R, \phi)$ compared to even k.*

4.2.1 Minimal optimality loss in high resolution regime

To find the optimal density $\rho^\star$ minimizing the optimality loss for given $\phi(g)$ and cost function $f(x;g)$, we introduce a new function called value density (VD) :

$$p_f(g) \triangleq \left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{\partial^k f(x;g)}{\partial x^k}|_{x=\psi(g)}\phi(g) \geq 0, \tag{4.5}$$

and normalized value density (NVD) due to assumption 4) :

$$\overline{p}_f(g) \triangleq \frac{p_f(g)}{\int_{g\in\mathcal{G}} p_f(g)\,\mathrm{d}g} \tag{4.6}$$

Then we resort to the Hölder's inequality :

$$\int p_f^{\frac{1}{k+1}} \leq \left(\int p_f \rho^{-k}\right)^{\frac{1}{k+1}} \left(\int \rho\right)^{\frac{k}{k+1}} \tag{4.7}$$

knowing $\left(\int \rho\right)^{\frac{k}{k+1}} = 1$, it can be inferred that $\int p_f\rho^{-k} \geq \left(\int p^{\frac{1}{k+1}}\right)^{k+1}$, with the equality if and only if $p_f\rho^{-k} = C_1\rho$ with $C_1 > 0$. Hence the optimal density function of representatives can be written as :

$$\rho^\star(g) = \frac{\left[\left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{\partial^k f(\psi(g);g)}{\partial x^k}\phi(g)\right]^{\frac{1}{k+1}}}{\int_{g\in\mathcal{G}}\left[\left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{\partial^k f(\psi(g);g)}{\partial x^k}\phi(g)\right]^{\frac{1}{k+1}}\mathrm{d}g} \tag{4.8}$$

Therefore, suppose $M = 2^R$, when R is large, the approximate optimality loss $\mathrm{L}(\mathcal{Q}; f, R, \phi)$ can be written as :

$$\begin{aligned}&\widehat{\mathrm{L}}(\mathcal{Q}; f, R, \phi)\\ =&\frac{C(k)}{2^{kR}}\left(\int_{g\in\mathcal{G}}\left[\left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{\partial^k f(\psi(g);g)}{\partial x^k}\phi(g)\right]^{\frac{1}{k+1}}\mathrm{d}g\right)^{k+1}\end{aligned} \tag{4.9}$$

Note that $\frac{\partial f(x;g)}{\partial x}|_{x=\psi(g)} = 0$ (without considering the constraints on decision variable x), so we should always have $k \geq 2$. For the sake of clarity, it is assumed that $k = 2$ in the rest of section except otherwise stated. The form of Eq. 4.8 is similar to the high-resolution approximation of MSE distortion whose optimal density is proportional to $\phi^{\frac{1}{3}}(g)$ while the regularity of cost function also impacts the optimal density of representatives. For example, if there exist two parameters g_1, g_2 s.t. $p_f(g_1) \geq p_f(g_2)$ for cost function f , then more quantization bits should be allocated to the neighbourhood of g_1 than g_2. This is the reason for which the function $p_f(g)$ is called value density. Define $q_f(g) = \left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{\partial^k f(\psi(g);g)}{\partial x^k}$, then one has $p_f(g) = q_f(g)\,\phi(g)$. Obviously if $q_f(g) > 1$, then density of representatives should be denser compared to the case of distortion-like

cost functions. Moreover, the function $q_f(g)$ could represent the intensity of fluctuation of cost function $f(x;g)$ at point g. And the optimal density of representatives found in Eq. 4.8 could be seen as the twisted version of usual density of representatives for distortion-oriented quantization. We take the following family of cost functions as an example to help understand some conceptions above : $h(x;g) = -\frac{\exp\left(-\frac{c}{xg}\right)}{x^\eta}$ with $c > 0$ and $\eta \geq 2$. Assume that g is exponential distributed with p.d.f., i.e., $\phi(g) = \frac{1}{v}\exp\left(-\frac{g}{v}\right)$ with $v > 0$, one can easily verify that $\psi(g) = \frac{c}{\eta g}$, $h^\star(g) = -\left(\frac{\eta g}{ce}\right)^\eta$ and $k = 2$ as expected, then one has

$$\begin{aligned} p_h(g) &= \left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^2 \frac{\partial^2 f(x;g)}{\partial x^2}|_{x=\psi(g)}\phi(g) \\ &= -\left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^2 \frac{\eta}{\psi^2(g)} h^\star(g) \\ &= \frac{\eta^{\eta+1}}{c^\eta e^\eta} g^{\eta-2}\phi(g) \end{aligned} \tag{4.10}$$

Notice that if $\eta = 2$, the obtained NVD is exactly the p.d.f., i.e., $\overline{p}_h(g) = \phi(g)$. This coincidence entails that even MSE quantizer could be optimal in high-regime for non-trivial cost function. If one take $\eta = 3$ for example, then one has $g^\star = \nu$ maximizing $p_h(g)$. And $p(g)$ is increasing for $0 \leq g \leq g^\star$ then decreasing for $g \geq g^\star$. Therefore the density of representatives will be completely different from the situation where larger parameter g leads to fewer representatives for distortion-like cost functions. For $\eta \geq 2$ approximated optimality loss for this family of cost functions can be expressed as :

$$\widehat{\mathrm{L}}(\mathcal{Q}; h, R, \phi) = \frac{(3v)^{\eta-2}}{24}\frac{\eta^{\eta+1}}{e^\eta c^\eta}\Gamma^3\left(\frac{\eta+1}{3}\right)2^{-2R} \tag{4.11}$$

where $\Gamma(\cdot)$ is the famous Gamma function. Back to the approximated optimality loss in Eq. 4.9, one can observe that, in high-resolution regime, the scale of approximated optimality loss $\widehat{\mathrm{L}}(\mathcal{Q}; h, R, \phi)$ is 2^{-kR} independent of probability distribution. Therefore if one wishes to compare the hardness of quantization for different cost functions, functions with larger exponent k should be harder to quantizer than the one with smaller exponent. Moreover, if two different cost functions have the same exponent k, their behaviors in high-resolution regime are basically the same.

4.2.2 Extreme cost function for fixed optimal density

We have considered the case with a given distribution $\phi(g)$ and find the the optimal density $\rho^\star(g)$ for a given cost function and the approximate optimality loss. Then a problem arises naturally, that is, what will be the best (one minimizes the OL) and the worst (one maximizes the OL) cost function if one always maintains the optimal density $\rho^\star(g)$. Before properly define what is the extreme cost function, we first interpret the value density function $p_f(g)$. To have a fair comparison of cost function, we define the following set of cost $\mathcal{F}_C$ which contains all cost function with $C > 0$ as the average fluctuation :

$$\mathcal{F}_C = \left\{ f : \int_{g\in\mathcal{G}} p_f(g)\,\mathrm{d}g = C \right\} \tag{4.12}$$

Therefore the best cost function and the worst function are defined as :

$$f_{\text{worst}} \in \arg\max_{f\in\mathcal{F}_C} \widehat{\mathrm{L}}\left(\mathcal{Q}; f, R, \phi\right) \tag{4.13}$$

and

$$f_{\text{best}} \in \arg\min_{f\in\mathcal{F}_C} \widehat{\mathrm{L}}\left(\mathcal{Q}; f, R, \phi\right) \tag{4.14}$$

According to these notations, the following propositions can be made.

Proposition 4.2.2. *The worst function satisfies* $q_f(g) = C_w\phi^{\frac{1}{k}}(g)$ *; the best functions satisfies* $q_f(g) = C_b\phi^{-1}(g)$ *with* $C_w = \dfrac{1}{\displaystyle\int_{g\in\mathcal{G}} \phi^{\frac{k+1}{k}}\mathrm{d}g}$ *and* $C_b = \frac{C}{|\mathcal{G}|}$.

Proof : See Appendix A. ■

Prop. 4.2.2 actually tell us the following fact. If the average fluctuation of cost functions are fixed in the sense of 4.12, then the value density of worst cost function should be proportional to $\phi^{\frac{1}{k}}(g)$ while the one of best cost function should be inversely proportional to $\phi(g)$.

It is worth mentioning that all discussion about the existence of worst (best) cost function is given in a constructed way, i.e., if there exists a cost function f with its corresponding optimal decision function $\psi(g)$ satisfying those condition, then one does find the desired cost function. In other words, above conclusions are made on the ODF instead on cost function itself. Generally, it is hard to find all functions satisfying those conditions while the existence of such function is easy to prove. For example, we define the following polynomial function of x :

$$f(x; g) = \sum_{i=1}^{k} a_i \left(x - \Psi(g)\right)^i, \tag{4.15}$$

where $a_1 = k!$ and $a_i \in \mathbb{R}$, for $2 \leq i \leq k$. Function $\Psi(g)$ is defined as :

$$\Psi(g) = \left(\int_{\mathcal{G}} \phi^{\frac{k+1}{k}}\mathrm{d}g\right)^{-\frac{1}{k}} \Phi(g) \tag{4.16}$$

with $\Phi(g)$ the primitive function of $\phi^{\frac{1}{k^2}}(g)$. Obviously function in Eq. 4.15 belongs to the category of the worst cost function. For the best cost function, similar analysis can be done as well.

4.2.3 A simple classification of cost functions

In Sec. 4.2.1, we have roughly compared the hardness of quantization with same average fluctuation for different cost functions by their behaviors in high-resolution regime. However it could be not rigorous enough since an universal metric for comparing different cost functions is missing. Our claim is based on the defect of two widely-used quantity of optimality loss : absolute OL and relative OL defined as the ratio of absolute OL

over the optimum.[1] Therefore, to investigate which kind of cost functions are easily to be compressed, we first introduce an universal metric named normalized optimality loss (NOL) for different cost functions[2]. The proposed metric can be seen as the ratio between the optimality loss induced by R quantization bits and the optimality loss without using quantization resource (0 bit), and can be expressed as follows :

$$\mathrm{S}\left(\mathcal{Q}; f, R, \phi\right) = \frac{\mathrm{L}\left(\mathcal{Q}; f, R, \phi\right))}{\mathbb{E}_g\left[f\left(\overline{x}; g\right) - f\left(\psi\left(g\right); g\right)\right]} \tag{4.17}$$

where $\overline{x}$ is the optimal decision minimizing the absolute optimality loss without additional instantaneous information of g, i.e.,

$$\overline{x} \in \arg\min_{x \in \mathcal{X}} \mathbb{E}_g\left[f\left(x; g\right) - f^{\star}\left(g\right)\right] \tag{4.18}$$

A higher value of NOL $\mathrm{S}\left(\mathcal{Q}; f, R, \phi\right)$ indicates that the function f needs more quantization bits to achieve the same optimality loss, and thus harder to be quantized. For two cost functions f and h and fixed rate R, if $\mathrm{S}\left(\mathcal{Q}; f, R, \phi\right) < \mathrm{S}\left(\mathcal{Q}; h, R, \phi\right)$, then we say f is easier to compress than h for rate R and p.d.f. $\phi\left(g\right)$; If that holds for any rate R, then we say f is easier to compress than h for p.d.f. $\phi\left(g\right)$. Equipped with NOL, one is able to fairly compare two cost functions. One could easily verify that NOL is invariant for linear transformation of cost function, i.e., for $\forall a, b \in \mathbb{R}$ with $a \neq 0$, one has

$$\mathrm{S}\left(\mathcal{Q}; af + b, R, \phi\right) = \mathrm{S}\left(\mathcal{Q}; f, R, \phi\right) \tag{4.19}$$

Moreover, assume two functions f, h and a constant a s.t. $\mathcal{X}_f = \mathcal{X}_h - a$ and $h(x + a; g) = f(x; g)$ then we have $\mathrm{S}\left(\mathcal{Q}; f, R, \phi\right) = \mathrm{S}\left(\mathcal{Q}; h, R, \phi\right)$. In scalar high resolution case (R large), NOL can be approximated by :

$$\begin{aligned} &\widehat{\mathrm{S}}\left(\mathcal{Q}; f, R, \phi\right) \\ =&\frac{C\left(k\right)}{2^{kR}} \frac{\left(\int_{g \in \mathcal{G}} \left(\left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^{k} \frac{\partial^k f(\psi\left(g\right); g)}{\partial x^k} \phi(g)\right)^{\frac{1}{k+1}} \mathrm{d}g\right)^{k+1}}{\int_{g \in \mathcal{G}} \left[f\left(\overline{x}; g\right) - f\left(\psi\left(g\right); g\right)\right] \phi\left(g\right) \mathrm{d}g} \end{aligned} \tag{4.20}$$

Using Eq. 4.20, we can compare the hardness of compression for different cost functions without performing simulations for them.

Example 1. *Consider three energy efficiency cost function* $f^{PSTR}(x; g) = \frac{\exp\left(-\frac{c}{gx}\right)}{x}$, $f^{BER}\left(x; g\right) = \frac{(1-\exp(-gx))^N}{x}$ *and* $f^{SUB}\left(x; g\right) = \log(1 + \frac{gx}{\sigma^2}) - bx$, *where PSTR means packet success transmission rate; BER stands for bit error rate and SUB means that the energy efficiency is defined as the form of subtraction. Constant* c, N, σ^2, b *represents the spectre efficiency, number of packets, variance of channel noise and a factor showing the importance of energy consumption.*

1. The first issue of these two metrics is that they are not invariant for linear transformation of the cost function. Second reason for abandoning relative optimality loss is that it can not be applied to distortion-like cost function, e.g., $f\left(x; g\right) = \left(x - g\right)^2$ with $f^{\star}\left(g\right) \equiv 0$ which leads to a infinity.

2. Obviously NOL can not be applied to cost functions with independent decision of parameter. We treat their NOL as zero to show that quantization for those functions are useless.

Assuming the power budget is sufficiently large and the distribution ϕ follows a uniform distribution over $[\varepsilon, 1]$ with $\varepsilon \to 0$, the comparison can be shown in the following table : According our approximation, NOL in high resolution scalar case could be ranked

Cost function	**ODF** $\psi(g)$	**NVD** $\overline{p}(g)$	**NOL**
$\log(1+10gx)-x$	$[1-\frac{1}{10g}]^{+}$	$\frac{1}{333g^4}$	$0.44\frac{C(2)}{2^{2R}}$
$(x-g)^2$	g	$\frac{2g}{1-\varepsilon^2}$	$24\frac{C(2)}{2^{2R}}$
$\frac{\exp(-\frac{1}{gx})}{x}$	$\frac{1}{g}$	$\frac{1}{g\ln(1/\varepsilon)}$	$68.22\frac{C(2)}{2^{2R}}$
$\frac{\exp(-\frac{5}{gx})}{x}$	$\frac{5}{g}$	$\frac{1}{g\ln(1/\varepsilon)}$	$68.22\frac{C(2)}{2^{2R}}$
$\frac{\exp(-\frac{10}{gx})}{x}$	$\frac{10}{g}$	$\frac{1}{g\ln(1/\varepsilon)}$	$68.22\frac{C(2)}{2^{2R}}$
$\frac{(1-\exp(-gx))^{10}}{x}$	$\frac{3.6150}{g}$	$\frac{1}{g\ln(1/\varepsilon)}$	$101.54\frac{C(2)}{2^{2R}}$
$\frac{(1-\exp(-gx))^{50}}{x}$	$\frac{5.6466}{g}$	$\frac{1}{g\ln(1/\varepsilon)}$	$125.63\frac{C(2)}{2^{2R}}$
$\frac{(1-\exp(-gx))^{100}}{x}$	$\frac{6.6746}{g}$	$\frac{1}{g\ln(1/\varepsilon)}$	$136.08\frac{C(2)}{2^{2R}}$

TABLE 4.1 – Comparison of different cost functions

as EE-BER (large number of packets) > EE-BER (small number of packets) > EE-PSTR > MSE > EE-SUB. Specifically, EE in form of subtraction is much easier to be quantized than other cost functions. It is worth mentioning that above conclusion only holds for uniform distribution. For other probability distribution $\phi(g)$, one could even have contrary conclusion. Before ending this section, we discuss a bit about how to extend our current framework to compact constrained decision space.

Remark 4.2.3. *We assume* $\mathcal{X}_c = \left[\underline{X}, \overline{X}\right]$, *then define* $\mathcal{G}_{M-1} = \{g \in \mathcal{G} \text{ s.t. } \psi(g) = \underline{X}\}$ *and* $\mathcal{G}_M = \left\{g \in \mathcal{G} \text{ s.t. } \psi(g) = \overline{X}\right\}$. *Since* $\frac{\partial^k f(\psi(g);g)}{\partial x^k} \neq 0$ *on the boundary, there is another second-order term if one approximate the OL :*

$$\begin{aligned}
&\widehat{\mathrm{L}}(\mathcal{Q}; f, R, \phi) \\
&= \sum_{m=1}^{M-2} \frac{C(k)}{M^k} \int_{g\in\mathcal{G}_m} \rho^{-k}(g) \left[\left(\frac{\mathrm{d}\psi(g)}{\mathrm{d}g}\right)^k \frac{\partial^k f(\psi(g);g)}{\partial x^k}\right] \phi(g)\mathrm{d}g \\
&\quad + \sum_{m=M-1}^{M} \frac{C(2)}{M^2} \int_{g\in\mathcal{G}_m} \rho^{-2}(g) \frac{\mathrm{d}\psi(g)}{\mathrm{d}g} \frac{\partial^2 f(\psi(g);g)}{\partial x^2} \phi(g)\mathrm{d}g
\end{aligned} \tag{4.21}$$

However for $g \in \mathcal{G}_{M-1} \bigcup \mathcal{G}_M$, *the optimal decision is always unique, then one always has* $\frac{\mathrm{d}\psi}{\mathrm{d}g} = 0$. *Therefore, all previous results remain true except one only needs* $(M-2)$ *cells instead of* M *cells.*

4.3 Vector High-Resolution Quantization

In this section, we consider a more general case, both the parameter $\boldsymbol{g}$ to be quantized and the decision $\boldsymbol{x}$ are vectors. We first introduce some notation to facilitate the expression. The optimal decision function should be also in vector form : $\boldsymbol{\kappa}(\boldsymbol{g}) = (\boldsymbol{\kappa}_1(\boldsymbol{g}), \ldots, \boldsymbol{\kappa}_{d_1}(\boldsymbol{g}))$ as previous section. Moreover, we introduce the notation of multi

index in order to represent partial derivative of cost functions easier : $\boldsymbol{n} = (\boldsymbol{n}_1, \ldots, \boldsymbol{n}_{d_1})$ with $\boldsymbol{n}_t \in \{1, \ldots, d_1\}$ for $\forall t$. We further define $|\boldsymbol{n}| \triangleq \sum_{t=1}^{d_1} \boldsymbol{n}_t$ and $\boldsymbol{n}! \triangleq \prod_{t=1}^{d_1} n_t!$. With the multi-index notation, one associate a particular partial derivative to $\boldsymbol{n}$ w.r.t. decision variable $\boldsymbol{x}$: $\mathfrak{D}_{\boldsymbol{x}}^{\boldsymbol{n}} f = \frac{\partial^{|\boldsymbol{n}|} f}{\partial \boldsymbol{x}_1^{\boldsymbol{n}_1} \ldots \partial \boldsymbol{x}_{d_1}^{\boldsymbol{n}_{d_1}}}$. Similar definition can be done for parameter variable $\boldsymbol{g}$ as well. For a vector $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{d_1})$, we introduce multi-index power of $\boldsymbol{n}$ for a vector $\boldsymbol{x}$: $\boldsymbol{x}^{\boldsymbol{n}} = \prod_{i=1}^{d_1} \boldsymbol{x}_i^{\boldsymbol{n}_i}$. Matrix $\mathrm{H}_f(\boldsymbol{x}; \boldsymbol{g})$ represents the Hessian matrix of $f(\boldsymbol{x}; \boldsymbol{g})$ w.r.t. $\boldsymbol{x}$ for given $\boldsymbol{g}$; $\mathrm{J}_{\boldsymbol{\kappa}}(\boldsymbol{g})$ is the Jacobian matrix of $f(\boldsymbol{x}; \boldsymbol{g})$ for optimal decision function $\boldsymbol{\kappa}(\boldsymbol{g})$. Similar to scalar case, we make the following assumptions :

1. Cost function $f(x; g)$ has all K-th order partial derivative w.r.t. $\boldsymbol{x}$ with $K \geq 2$, i.e., $\forall \boldsymbol{n}$ s.t. $|\boldsymbol{n}| \leq K$, $\mathfrak{D}_x^{\boldsymbol{n}} f$ exists.

2. Define $\mathcal{K} \triangleq \{i \in \mathbb{N} | \mathfrak{D}_{\boldsymbol{x}}^{\boldsymbol{n}} f(\boldsymbol{x}; \boldsymbol{g}) \neq 0 \text{ a.s., } \forall \boldsymbol{n} \text{ s.t. } |\boldsymbol{n}| = i\}$, we assume $\mathcal{K}$ is nonempty and define $k \triangleq \min \mathcal{K}$.

3. Jacobian satisfies $\mu\{\boldsymbol{g} : \mathrm{J}_{\boldsymbol{\kappa}}(\boldsymbol{g}) = \mathbf{0}_{d_1 \times d_2}\} = 0$ with $\mathbf{0}_{d_1 \times d_2}$ is the $d_1 \times d_2$ matrix containing only 0.

4. For $\forall \boldsymbol{g} \in \mathcal{G}$, $\boldsymbol{\kappa}(\boldsymbol{g}) \in \mathrm{int}(\mathcal{X})$.

Similarly, assumption (2) and (3) exclude all cost functions whose decision is independent of parameter. Assumption (4) still limits our discussion in unconstrained case. By using the Taylor expansion for multivariate functions, the optimality loss can be rewritten as :

$$
\begin{aligned}
&\mathrm{L}(\mathcal{Q}; f, R, \phi) \\
=&\sum_{m=1}^{M} \int_{\boldsymbol{g} \in \mathcal{G}_m} \left[f(\boldsymbol{\kappa}(\boldsymbol{z}_m); \boldsymbol{g}) - f(\boldsymbol{\kappa}(\boldsymbol{g}); \boldsymbol{g}) \right] \phi(\boldsymbol{g}) \mathrm{d}\boldsymbol{g} \\
=&\sum_{m=1}^{M} \left[\sum_{\boldsymbol{n}: |\boldsymbol{n}| \leq k} \int_{\boldsymbol{g} \in \mathcal{G}_m} \frac{\mathfrak{D}_{\boldsymbol{x}}^{\boldsymbol{n}} f(\boldsymbol{\psi}(\boldsymbol{g}); \boldsymbol{g})}{\boldsymbol{n}!} (\boldsymbol{\psi}(\boldsymbol{z}_m) - \boldsymbol{\psi}(\boldsymbol{g}))^{\boldsymbol{n}} \phi(\boldsymbol{g}) \mathrm{d}\boldsymbol{g} \right. \\
&\left. + \sum_{\widehat{\boldsymbol{n}}: |\widehat{\boldsymbol{n}}| = k+1} \int_{\boldsymbol{g} \in \mathcal{G}_m} O\left((\boldsymbol{\psi}(\boldsymbol{z}_m) - \boldsymbol{\psi}(\boldsymbol{g}))^{\widehat{\boldsymbol{n}}} \right) \phi(\boldsymbol{g}) \mathrm{d}\boldsymbol{g} \right]
\end{aligned}
\tag{4.22}
$$

For $k > 2$, it is difficult to obtain further information from Eq. 4.22. For the sake of simplicity, we assume $k = 2$ as before, the optimality loss in high resolution case can be approximated alternatively by Hessian matrix and Jacobian matrix instead of using Eq.

4.22 :

$$
\begin{aligned}
&\mathrm{L}\left(\mathcal{Q}; f, R, \phi\right) \\
=&\sum_{m=1}^{M} \int_{\boldsymbol{g}\in\mathcal{G}_m} \left[f(\boldsymbol{\kappa}(\boldsymbol{z}_m); \boldsymbol{g}) - f(\boldsymbol{\kappa}\left(\boldsymbol{g}\right); \boldsymbol{g})\right] \phi(\boldsymbol{g})\mathrm{d}\boldsymbol{g} \\
\overset{(a)}{\approx}&\sum_{m=1}^{M} \int_{\boldsymbol{g}\in\mathcal{G}_m} \frac{1}{2}(\boldsymbol{\kappa}\left(\boldsymbol{z}_m\right) - \boldsymbol{\kappa}\left(\boldsymbol{g}\right))^{\mathrm{T}}\mathrm{H}_f(\boldsymbol{\kappa}\left(\boldsymbol{g}\right); \boldsymbol{g})(\boldsymbol{\kappa}\left(\boldsymbol{z}_m\right) - \boldsymbol{\kappa}\left(\boldsymbol{g}\right))\phi(\boldsymbol{g})\mathrm{d}\boldsymbol{g} \\
\overset{(b)}{\approx}&\sum_{m=1}^{M} \int_{\boldsymbol{g}\in\mathcal{G}_m} \frac{1}{2}(\mathrm{J}_{\boldsymbol{\kappa}}\left(\boldsymbol{g}\right)\left(\boldsymbol{z}_m - \boldsymbol{g}\right))^{\mathrm{T}}\mathrm{H}_f(\boldsymbol{\kappa}\left(\boldsymbol{g}\right); \boldsymbol{g})(\mathrm{J}_{\boldsymbol{\kappa}}\left(\boldsymbol{g}\right)\left(\boldsymbol{z}_m - \boldsymbol{g}\right))\phi(\boldsymbol{g})\mathrm{d}g \\
\overset{(c)}{=}&\underbrace{\sum_{m=1}^{M} \int_{\boldsymbol{g}\in\mathcal{G}_m} \frac{1}{2}\|\boldsymbol{g} - \boldsymbol{z}_m\|_2^2 \boldsymbol{e}_m^{\mathrm{T}}\mathrm{J}_{\boldsymbol{\kappa}}^{\mathrm{T}}(\boldsymbol{g})\mathrm{H}_f(\boldsymbol{\kappa}\left(\boldsymbol{g}\right); \boldsymbol{g})\mathrm{J}_{\boldsymbol{\kappa}}\left(\boldsymbol{g}\right)\boldsymbol{e}_m\phi(\boldsymbol{g})\mathrm{d}\boldsymbol{g}}_{\widehat{\mathrm{L}}(\mathcal{Q};f,R,\phi)}
\end{aligned}
\tag{4.23}
$$

where and $\boldsymbol{e}_m$ is defined as the normalized vector of the difference, i.e., $\boldsymbol{e}_m = \frac{\boldsymbol{g}-\boldsymbol{z}_m}{\|\boldsymbol{g}-\boldsymbol{z}_m\|_2}$. (a) follows from the fact that the higher order term in the Taylor expansion of $(f(\boldsymbol{\kappa}(\boldsymbol{z}_m; \boldsymbol{g}) - f(\boldsymbol{\kappa}(\boldsymbol{g}); \boldsymbol{g}))$ are negligible to the second order term ; (b) follows from the fact that the higher order term in the Taylor expansion of $(\boldsymbol{\kappa}(\boldsymbol{g}) - \boldsymbol{\kappa}(\boldsymbol{g}_m))$ are negligible to the first order term ; (c) can be verified by defining $\boldsymbol{e}_m$. It is worth noting that this expression is similar to the classical vector quantization while the p.d.f. of $\boldsymbol{g}$ is weighted by a new coefficient related to the Hessian and Jacobian of the cost function and the normalized vector $\boldsymbol{e}_m$. To simplify the formula, we denote $\mathrm{A}_{f,\boldsymbol{\kappa}}\left(\boldsymbol{g}\right) = \mathrm{J}_{\boldsymbol{\kappa}}^{\mathrm{T}}(\boldsymbol{g})\mathrm{H}_f(\boldsymbol{\kappa}\left(\boldsymbol{g}\right); \boldsymbol{g})\mathrm{J}_{\boldsymbol{\kappa}}\left(\boldsymbol{g}\right)$, then one has :

$$
\widehat{\mathrm{L}}\left(\mathcal{Q}; f, R, \phi\right) = \sum_{m=1}^{M} \int_{\boldsymbol{g}\in\mathcal{C}_m} \frac{1}{2}\|\boldsymbol{g} - \boldsymbol{z}_m\|_2^2 \boldsymbol{e}_m^{\mathrm{T}}\mathrm{A}_{f,\boldsymbol{\kappa}}\left(\boldsymbol{g}\right)\boldsymbol{e}_m\phi(\boldsymbol{g})\mathrm{d}\boldsymbol{g} \tag{4.24}
$$

One can immediately notice that the function $q\left(g\right)$ is the degeneration of matrix $\mathrm{A}_{f,\boldsymbol{\kappa}}\left(\boldsymbol{g}\right)$ in scalar case. However, different from scalar case, we are not able to define VD adapted to specific cost function due to the missing of an cell-invariant integral formula which entails the fundamental difference between scalar case and vector case. Before explain how to use Eq. 4.23 to construct a goal-oriented quantizer, we point out another issue in vector case of GOQ.

For scalar case, we have concentrated on finding extreme cost function for given distribution. For scalar function, even the best cost function could only leads to minimal optimality loss being strict positive. This is not always in vector case as we will show. Obviously if $\mathrm{A}_{f,\boldsymbol{\kappa}}\left(\boldsymbol{g}\right) = \mathbf{0}_{d_2\times d_2}$ almost surely, then the optimality loss could be considered as approximately null. However, as we will show in the following, This condition is not sufficient for having a real loss-less goal-oriented quantizer.

Consider the following cost function $f\left(\boldsymbol{x}; \boldsymbol{g}\right) = \boldsymbol{x}_1^2\boldsymbol{x}_2^2\left[\boldsymbol{x}_1^2 + 4\boldsymbol{x}_2^2 - 3F^2\left(\boldsymbol{g}_1, \boldsymbol{g}_2\right)\right]$ with $d_1 = d_2 = 2$ and F being any scalar function s.t. $|F| \leq \frac{1}{3}$. Then $\boldsymbol{\kappa}\left(\boldsymbol{g}\right) = \left[F\left(\boldsymbol{g}\right), -\frac{F(\boldsymbol{g})}{2}\right]^{\mathrm{T}}$ is one of ODF satisfying all assumptions. One can verify that

$$
\mathrm{H}_f(\boldsymbol{\kappa}(\boldsymbol{g}); \boldsymbol{g}) = 2F^4\begin{bmatrix}1 & 2\\ 2 & 4\end{bmatrix}, \tag{4.25}
$$

and

$$\mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g}) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.26}$$

finally one should have $\widehat{\mathrm{L}}(\mathcal{Q}; f, R, \phi) = 0$ for $\forall \boldsymbol{g} \in \mathcal{G}$. However if one chose $\boldsymbol{\xi}(\boldsymbol{g}) = \left[F(\boldsymbol{g}), \frac{F(\boldsymbol{g})}{2}\right]^{\mathrm{T}}$ as the ODF, then $\mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g})$ is no longer all null. Besides, by the symmetry of the cost function, optimality loss introduced by $\boldsymbol{\kappa}(\boldsymbol{g})$ and $\boldsymbol{\xi}(\boldsymbol{g})$ should be the same, so $\mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g}) = \mathbf{0}_{d_2 \times d_2}$ is only a necessary condition. The choice for ODF should be careful to avoid having a fake lossless cost function like the given example. The existence of real lossless cost functions is out of the scope of this manuscript and we omit it here.

We turn back to the problem of approximating optimality loss. The main difficulty is that the normalized vector $\boldsymbol{e}_m$ depends both on $\boldsymbol{g}$ and the representative $\boldsymbol{z}_m$. Therefore the vector case can not be tackled as the scalar case where one is able to define the value density. Nevertheless, we will show similar properties could be found in vector case. To directly approximate optimality loss defined in (4.23) is complicated, we thus resort to some matrix properties to bound it. The accuracy of our approximation depends on how we approximate the the term $\boldsymbol{e}_m^{\mathrm{T}} \mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g}) \boldsymbol{e}_m$. For a given parameter $\boldsymbol{g}$, eigenvalues of matrix $\mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g})$ are denoted by $0 \leq \lambda_1(\boldsymbol{g}; f) \leq \cdots \leq \lambda_{d_2}(\boldsymbol{g}; f)$ and its **normalized** eigenvectors are defined as $\boldsymbol{\nu}_1(\boldsymbol{g}; f), \ldots, \boldsymbol{\nu}_{d_2}(\boldsymbol{g}; f)$. Since the Hessian matrix $\mathrm{H}_f(\boldsymbol{\kappa}(\boldsymbol{g}); \boldsymbol{g})$ is non negative definite due to optimum. Therefore, the term $\boldsymbol{e}_m^{\mathrm{T}} \mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g}) \boldsymbol{e}_m$ can be upper bounded by maximum eigenvalue as $\lambda_{d_2}(\boldsymbol{g}; f)$ of $\mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g})$ and lower bounded by its minimum eigenvalue $\lambda_1(\boldsymbol{g}; f)$, which yields the following proposition.

Proposition 4.3.1. *When $M = 2^R$ is very large and $d_1 \geq d_2$, assuming that Gersho's conjecture is correct, the approximate optimality loss $\widehat{\mathrm{L}}(\mathcal{Q}; f, R, \phi)$ in (4.24) can be upper bounded by*

$$\widehat{\mathrm{L}}_{\mathrm{sup}}(\mathcal{Q}; f, R, \phi) = \frac{d_2 \mathsf{M}_{d_2}}{2} 2^{\frac{-2R}{d_2}} \left(\int_{\boldsymbol{g} \in \mathcal{G}} (\lambda_{d_2}(\boldsymbol{g}; f) \phi(\boldsymbol{g}))^{\frac{d_2}{d_2+2}} \, \mathrm{d}\boldsymbol{g} \right)^{\frac{d_2+2}{d_2}} \tag{4.27}$$

and lower bounded by

$$\widehat{\mathrm{L}}_{\mathrm{inf}}(\mathcal{Q}; f, R, \phi) = \frac{d_2 \mathsf{M}_{d_2}}{2} 2^{\frac{-2R}{d_2}} \left(\int_{\boldsymbol{g} \in \mathcal{G}} (\lambda_1(\boldsymbol{g}; f) \phi(\boldsymbol{g}))^{\frac{d_2}{d_2+2}} \, \mathrm{d}\boldsymbol{g} \right)^{\frac{d_2+2}{d_2}} \tag{4.28}$$

where M_{d_2} is the least normalized moment of inertia of d_2-dimensional tessellating polytopes.

Proof : See Appendix B. ■

Here, we briefly recall Gersho's conjecture [22] on the optimal block quantization problem or optimal centroidal Voronoi tessellations (CVT). For a collection of points $\boldsymbol{z}_m \in \mathcal{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_M\} \subset \mathcal{G}$, we define the associated the Voronoi cell (comprosing a Voronoi tessellation of $\mathcal{G}$) of point $\boldsymbol{z}_m$ as :

$$\mathcal{C}_m = \{\boldsymbol{g} \in \mathcal{G} : \|\boldsymbol{g} - \boldsymbol{z}_m\| \leq \|\boldsymbol{g} - \boldsymbol{z}_n\|, \ \forall n \neq m\} \tag{4.29}$$

A centroidal Voronoi tessellation means all points y_k are exactly centroids of their asscociated Voronoi cell. Here all parameters are set to be uniformly distributed in parameter

space $\mathcal{G}$. Then quantization loss can be expressed as

$$\mathrm{L}\left(\mathcal{Z}\right)=\sum_{m=1}^{M}\int_{\boldsymbol{g}\in\mathcal{C}_m}\|\boldsymbol{g}-\boldsymbol{z}_m\|^2\mathrm{d}\boldsymbol{g} \tag{4.30}$$

The Gersho's famous conjecture is about the shape of optimal polytope which minimizes the quantization loss. The conjecture could be formally stated as :

Conjecture 4.3.2. *[Gersho,1979] There exists a polytope $\mathcal{C}$ with unit volume which tiles the space with congruent copies such that the following holds : let $\left(\mathcal{Z}_n\right)_n$ be a sequence of minimizers, with $\mathcal{Z}_n\in\arg\min_{|\mathcal{Z}|=n}\mathrm{L}\left(\mathcal{Z}\right)$, then the Voronoi cells of $\left(\mathcal{Z}_n\right)$ are asymptotically congruent to $n^{-\frac{1}{n}}\mathcal{C}$ as $n\to+\infty$.*

Back to our discussion on Prop. 4.3.1, we can check that the two bounds coincide when the cost function degenerates to the MSE, namely, $f^{\mathrm{dis}}(\boldsymbol{x};\boldsymbol{g})=\|\boldsymbol{x}-\boldsymbol{g}\|$ ($\lambda_1=1$ with this special cost function). Moreover, if the influence of each component of $\boldsymbol{g}$ is more comparable or the dimension d_2 is smaller, the difference between $\lambda_1(\boldsymbol{g};f)$ and $\lambda_{d_2}(\boldsymbol{g};f)$ can be predicted to be smaller, resulting in a smaller gap between the proposed upper bound and the lower bound.

As for the inertial profile, when $d_2=1$, the optimum inertial profile is $\mathfrak{m}(\boldsymbol{g})=\frac{1}{12}$ and both the upper bound and the lower bound reduces to what we found in the scalar case. When $d_2\geq 2$, as shown in [22], M_{d_2} can be bounded as

$$\frac{d_2}{d_2+2}V_{d_2}^{-2/d_2}\leq\mathsf{M}_{d_2}\leq\frac{d_2}{12} \tag{4.31}$$

where V_{d_2} is the volume of the unit radius sphere with dimension d_2. Moreover, high resolution theory need not to count solely on Gersho's conjecture, since it has been shown in [24][25] that the distortion can be written in the form

$$\begin{aligned}&\sum_m\int_{\boldsymbol{g}\in\mathcal{C}_m}\frac{1}{2}\|\boldsymbol{g}-\boldsymbol{z}_m\|_2^2\phi(\boldsymbol{g})\mathrm{d}\boldsymbol{g}\\&=b_{d_2}\left(\int_{\boldsymbol{g}\in\mathcal{G}}\left(\phi(\boldsymbol{g})\right)^{d_2/(d_2+2)}\mathrm{d}\boldsymbol{g}\right)^{(d_2+2)/d_2}2^{\frac{-2R}{d_2}}\end{aligned} \tag{4.32}$$

where $b_{d_2}>0$ is independent of $\phi(\boldsymbol{g})$. Therefore, the Gersho's conjecture can be seen a special conjecture about b_{d_2}. Actually, in realistic applications, above discussion could be unsuitable due to the fact that the dimension of decision d_1 is smaller than the dimension of parameter d_2 in clustering and classification problems. For example, if one wish to cluster some graphs into different classes, then one should generally have d_2 much larger than $d_1=1$. Therefore, to extend the conclusions in Prop. 4.3.1, we have the following remark.

Remark 4.3.3. *($d_2\geq d_1$ scenario) The proposed bounds suit well when $d_2\leq d_1$. However, if $d_2\geq d_1$, it can be seen that $\lambda_1(\boldsymbol{g};f)\equiv 0$ since the matrix $A_{f,\boldsymbol{\kappa}}(\boldsymbol{g})$ can be proved to be not full ranked. As a consequence, the lower bound derived in (4.28) is not tight anymore. Hence, it is necessary to find a new tight lower bound in this scenario. To this end, we can treat $J_{\boldsymbol{\kappa}}(\boldsymbol{g})\boldsymbol{e}_m$ as a vector and thus $\left(\boldsymbol{e}_m^{\mathrm{T}}A_{f,\boldsymbol{\kappa}}(\boldsymbol{g})\,\boldsymbol{e}_m\right)$ can be minimized if*

and only if $J_{\boldsymbol{\kappa}}(\boldsymbol{g})\boldsymbol{e}_m$ aligns with the eigenvector corresponding to the smallest eigenvalue of $H_f(\boldsymbol{\kappa}(\boldsymbol{g});\boldsymbol{g})$. Define the smallest eigenvalue of $H_f(\boldsymbol{\kappa}(\boldsymbol{g});\boldsymbol{g})$ as $\Lambda_{\min}(\boldsymbol{g};f)$, the term $\left(\boldsymbol{e}_m^{\mathrm{T}} A_{f,\boldsymbol{\kappa}}(\boldsymbol{g})\boldsymbol{e}_m\right)$ can be lower bounded by $\Lambda_{\min}(\boldsymbol{g};f)\mathsf{a}(J_{\boldsymbol{\kappa}}(\boldsymbol{g}))$, where $\mathsf{a}(J_{\boldsymbol{\kappa}}(\boldsymbol{g}))$ is the amplifying factor between $J_{\boldsymbol{\kappa}}(\boldsymbol{g})\boldsymbol{e}_m$ and the least eigenvector of $H_f(\boldsymbol{\kappa}(\boldsymbol{g});\boldsymbol{g})$. Replace $\lambda_1(\boldsymbol{g},f)$ by $\Lambda_{\min}(\boldsymbol{g};f)\mathsf{a}(J_{\boldsymbol{\kappa}}(\boldsymbol{g}))$, the new lower bound can be derived in a similar way when $d_2 \geq d_1$. They upper bound is not largely affected by the dimension and the proposed approach to derive $\widehat{\mathrm{L}}_{\mathrm{sup}}$ can be implemented either $d_2 \geq d_1$ or $d_2 \leq d_1$.

Before going to next section about how to use approximate optimality loss in Eq. 4.24 to find a goal-oriented quantizer. We end this section by explain how our framework can be extended to compact constrained decision space. Consider the following family of optimization problem parameterized by $\boldsymbol{g}$:

$$\begin{aligned} &\min_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x};\boldsymbol{g}) \\ &\text{s.t. } h_i(\boldsymbol{x}) \leq 0,\ \forall 1 \leq i \leq N_1 \\ &\qquad t_j(\boldsymbol{x}) = 0,\ \forall 1 \leq j \leq N_2 \end{aligned} \tag{4.33}$$

The feasible decision space $\mathcal{X}_c = \{\boldsymbol{x} \in \mathcal{X} \text{ s.t. } h_i(\boldsymbol{x}) \leq 0, t_j(\boldsymbol{x}) = 0, \forall i,j\}$ is formed by constraint functions. Without loss of generality, we assume that $\mathcal{X}_c$ is compact.

Remark 4.3.4. *For the sake of simplicity, we still assume the existence of a smooth optimal decision function $\overline{\boldsymbol{\kappa}}(\boldsymbol{g})$:*

$$\overline{\boldsymbol{\kappa}}(\boldsymbol{g}) \in \arg\min_{\boldsymbol{x}\in\mathcal{X}_c} f(\boldsymbol{x};\boldsymbol{g}) \tag{4.34}$$

In this situation, all representatives can be classified to two sorts $\{\boldsymbol{z}_i\}_{i=1}^M = \{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_N,\ldots,\boldsymbol{z}_{N+1},\ldots,\boldsymbol{z}_M\}$ with $\overline{\boldsymbol{\kappa}}(\boldsymbol{z}_i) = \boldsymbol{\kappa}(\boldsymbol{z}_i)$ for $1 \leq i \leq N$ and $\overline{\boldsymbol{\kappa}}(\boldsymbol{z}_i) \in bd(\mathcal{X}_c)$ for $N+1 \leq i \leq M$. Then the optimality loss can be expressed as $\mathrm{L} = \mathrm{L}_{int} + \mathrm{L}_{bd}$. Our approximation remains valid for L_{int}, while for L_{bd} is a bit different, both gradient and Hessian terms count :

$$\begin{aligned} &\widehat{\mathrm{L}}_{bd}(\mathcal{Q};f,R,\phi) \\ &= \sum_{i=N+1}^{M} \int_{\mathcal{C}_i} \|\boldsymbol{g}-\boldsymbol{z}_m\|_2 \nabla f(\boldsymbol{\kappa}_c(\boldsymbol{g});\boldsymbol{g})^T J_{\boldsymbol{\kappa}}(\boldsymbol{g})\boldsymbol{e}_m\phi(\boldsymbol{g})\mathrm{d}\boldsymbol{g} \\ &\quad + \sum_{i=N+1}^{M} \int_{\mathcal{C}_i} \frac{1}{2}\|\boldsymbol{g}-\boldsymbol{z}_m\|_2^2 \boldsymbol{e}_m^T \sum_{j=1}^{d_1} \frac{\partial f(\overline{\boldsymbol{\kappa}}(\boldsymbol{g});\boldsymbol{g})}{\partial \boldsymbol{x}_j} H_{\overline{\boldsymbol{\kappa}}_j}(\boldsymbol{g})\boldsymbol{e}_m\phi(\boldsymbol{g})\mathrm{d}\boldsymbol{g}, \end{aligned} \tag{4.35}$$

where $\left(\mathbf{H}_{\overline{\boldsymbol{\kappa}}_j}(\boldsymbol{g})\right)_{\ell,k} = \frac{\partial^2 \overline{\boldsymbol{\kappa}}_j(\boldsymbol{g})}{\partial \boldsymbol{g}_\ell \partial \boldsymbol{g}_k}$ for $1 \leq \ell,k \leq d_2$ and $1 \leq j \leq d_1$. Except from the freedom of choosing representatives, for compact constrained decision space, we have another freedom according to Eq. 4.35 : allocation of number of representatives to the boundary and the interior of constraint space.

4.4 Implementable Quantization Schemes

The knowledge of high-resolution optimality loss is very useful since it allows to quantify the hardness of quantization different cost functions. But the problem of designing

efficient object-based quantization scheme remains open : there is no general recipe to find the optimal quantization scheme to minimize the optimality loss. Recently, one efficient method is introduced in [28][29][1] where the optimality loss of the objective function can be minimized by implementing an iterative algorithm. However, that algorithm is kind of complicated, and might be prohibitive with a large size of data. Therefore, take the complexity into account, a novel and simpler algorithm can be proposed here. Therefore, to minimize the optimality loss, a Lloyd-Max-like algorithm can be implemented by solely adapting their distribution based on what we derived in Prop. 4.3.1 .

However this way of applying results does not take full advantage of information provided in Eq. 4.24. Without loss of generality, throughout this section, we assume that $d_1 \leq d_2$. This approach actually corresponds to the simplest situation where $|\lambda_1(\boldsymbol{g}; f) - \lambda_{d_2}(\boldsymbol{g}; f)| \to 0,\ \forall \boldsymbol{g}$. To find the optimal quantizer in the sense of Eq. 4.23, it is sufficient to use Lloyd-Max algorithm and take $\lambda_1(\boldsymbol{g}; f)\,\phi(\boldsymbol{g})$ as value density. However, if the difference between $\lambda_1(\boldsymbol{g}; f)$ and $\lambda_{d_2}(\boldsymbol{g}; f)$ is consistently large for some parameters $\boldsymbol{g}$, both upper bound and lower-bound could be inaccurate. Nevertheless, one can design an algorithm by making full knowledge of information of cost function provided by Eq. 4.24. The idea comes from the fact that eigenvalues and eigenvectors of $\mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g})$ could provide a relatively accurate approximation of the term $\boldsymbol{e}_m^T \mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g})\,\boldsymbol{e}_m$ and thus simplify the formula a lot. We introduce the approximated individual optimality loss for a parameter $\boldsymbol{g}$ with a representative $\boldsymbol{z}$:

$$\begin{aligned} d_f(\boldsymbol{g}, \boldsymbol{z}) =& \frac{1}{2}(\boldsymbol{g}-\boldsymbol{z})^T \mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g})(\boldsymbol{g}-\boldsymbol{z}) \\ =& \frac{1}{2}\|\boldsymbol{g}-\boldsymbol{z}\|^2 \boldsymbol{e}^T \mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g})\,\boldsymbol{e}, \end{aligned} \tag{4.36}$$

where $\boldsymbol{e} = \frac{\boldsymbol{g}-\boldsymbol{z}}{\|\boldsymbol{g}-\boldsymbol{z}\|}$. Therefore, for a given parameter sample set $\mathcal{T} = \left\{\boldsymbol{g}^{(t)}\right\}_{t=1}^{T}$, a quantizer $\mathcal{Q}$ characterized by its representatives $\{\boldsymbol{z}_m\}_{m=1}^{M}$ and quantization regions $\{\mathcal{C}_m\}_{m=1}^{M}$, the approximate optimality loss $\widehat{\mathrm{L}}(\mathcal{Q}; f, R, \phi) = \sum_{m=1}^{M} \int_{\boldsymbol{g}\in\mathcal{C}_m} d_f(\boldsymbol{g}, \boldsymbol{z}_m)\,\phi(\boldsymbol{g})\mathrm{d}\boldsymbol{g}$ can be Monte-Carlo simulated by a empirical optimality loss :

$$\overline{\mathrm{L}}(\mathcal{Q}; f, R, \mathcal{T}) = \frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M} d_f\left(\boldsymbol{g}^{(t)}, \boldsymbol{z}_m\right) \mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \overline{\mathcal{C}}_m\right\} \tag{4.37}$$

It is crucial to point out that the quantization region $\overline{\mathcal{C}}_m$ is defined in goal-oriented way for function d_f :

$$\overline{\mathcal{C}}_m \triangleq \{\boldsymbol{g} \in \mathcal{G} \text{ s.t. } d_f(\boldsymbol{z}_m; \boldsymbol{g}) \leq d_f(\boldsymbol{z}_n; \boldsymbol{g}),\ \forall n \neq m\} \tag{4.38}$$

For a multi-index $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_T)$ with $\boldsymbol{\alpha}_t \in \{1, \ldots, d_2\}$, we introduce the following function :

$$\widetilde{\mathrm{L}}(\mathcal{Q}, \boldsymbol{\alpha}; f, R, \mathcal{T}) = \frac{1}{T}\sum_{t,m} \lambda_{\boldsymbol{\alpha}_t}\left(\boldsymbol{g}^{(t)}; f\right) \|\boldsymbol{g}^{(t)} - \boldsymbol{z}_m\|^2 \mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \overline{\mathcal{C}}_m\right\} \tag{4.39}$$

To this end, we are able to explain the basic idea of our approach. First of all, we would like use Eq. 4.39 to approximate Eq. 4.37, if the number of representatives is sufficiently large, intuitively the norm $\|\boldsymbol{g}_t - \boldsymbol{z}_m\|$ being relatively small, the dominant difference of

d_f should depends on $\lambda_{\boldsymbol{\alpha}_t}\left(\boldsymbol{g}^{(t)}; f\right)$ especially when the matrix $\mathrm{A}_{f,\boldsymbol{\kappa}}\left(\boldsymbol{g}\right)$ is ill-conditioned. Therefore α_t could somehow represent the contribution of $\boldsymbol{g}_t$ to empirical loss. Obviously the "best" quantizer $\mathcal{Q}^\star$ in this sense corresponds to the setting where all sample achieve their minimal index jointly. It is also the "best" approximation that one can achieve by using Eq. 4.39. For n-th iteration with $\boldsymbol{\alpha}(n)$ and $\left\{\boldsymbol{z}_m^{(n)}\right\}_{m=1}^M$, one wishes to update current quantizer to find a new one which could probably introduce less optimality loss. Each iteration contains two steps which are not strictly independent one from each other :

1. Update multi-index $\alpha\left(n+1\right)$ from representatives $\left\{\boldsymbol{z}_m^{(n)}\right\}_{m=1}^M$.
2. Update representatives $\left\{\boldsymbol{z}_m^{(n)}\right\}_{m=1}^M$ from multi-index $\alpha\left(n\right)$.

To make our approach easier to be understand, when we discuss about how to update multi-index, we assume the approach for updating representatives is known and vice versa.

4.4.1 Multi-index update

Here the multi-index $\boldsymbol{\alpha}(n)$ characterizes the optimality loss that we expect to achieve by quantizer $\mathcal{Q}^{(n)}$. In the same time, one could find another multi-index $\boldsymbol{\beta}\left(n\right)$ which introduces the minimal loss when replacing $\overline{\mathrm{L}}$ by $\widetilde{\mathrm{L}}$:

$$\boldsymbol{\beta}\left(n\right) \in \arg\min_{\boldsymbol{\gamma}} \ \|\widetilde{\mathrm{L}}\left(\mathcal{Q}^{(n)}, \boldsymbol{\gamma}; f, R, \mathcal{T}\right) - \overline{\mathrm{L}}\left(\mathcal{Q}^{(n)}; f, R, \mathcal{T}\right)\| \tag{4.40}$$

Therefore, there is a deviation between what we expected and the actual situation :

$$\begin{aligned}
&\widetilde{\mathrm{L}}\left(\mathcal{Q}^{(n)}, \boldsymbol{\alpha}\left(n\right); f, R, \mathcal{T}\right) - \widetilde{\mathrm{L}}\left(\mathcal{Q}^{(n)}, \boldsymbol{\beta}\left(n\right); f, R, \mathcal{T}\right) \\
=&\frac{1}{T}\sum_{t=1}^{T}\sum_{m=1}^{M}\left(\lambda_{\boldsymbol{\alpha}_t(n)}\left(\boldsymbol{g}^{(t)}; f\right) - \lambda_{\boldsymbol{\beta}_t(n)}\left(\boldsymbol{g}^{(t)}; f\right)\right)\|\boldsymbol{g}^{(t)} - \boldsymbol{z}_m^{(n)}\|^2 \mathbb{1}\left\{\boldsymbol{g}^{(t)} \in \overline{\mathcal{C}}_m^{(n)}\right\}
\end{aligned}$$

We further introduce the individual deviation function $\mu_t(\boldsymbol{\alpha}(n), \beta(n), \mathcal{Q}^{(n)}; f, R, \mathcal{T})$:

$$\begin{aligned}
&\mu_t(\boldsymbol{\alpha}(n), \boldsymbol{\beta}(n), \mathcal{Q}^{(n)}; f, R, \mathcal{T}) \\
=&\left(\lambda_{\boldsymbol{\alpha}_t(n)}\left(\boldsymbol{g}^{(t)}; f\right) - \lambda_{\boldsymbol{\beta}_t(n)}\left(\boldsymbol{g}^{(t)}; f\right)\right)\|\boldsymbol{g}^{(t)} - \boldsymbol{z}_{m_t}^{(n)}\|^2,
\end{aligned} \tag{4.41}$$

where m_t is the label of corresponding representative of sample $\boldsymbol{g}^{(t)}$. Obviously, deviation in Eq. 4.41 is average of all individual deviation :

$$\begin{aligned}
&\widetilde{\mathrm{L}}\left(\mathcal{Q}^{(n)}, \boldsymbol{\alpha}\left(n\right); f, R, \mathcal{T}\right) - \widetilde{\mathrm{L}}\left(\mathcal{Q}^{(n)}, \boldsymbol{\beta}\left(n\right); f, R, \mathcal{T}\right) \\
=&\frac{1}{T}\sum_{t=1}^{T}\mu_t(\boldsymbol{\alpha}(n), \boldsymbol{\beta}(n), \mathcal{Q}^{(n)}; f, R, \mathcal{T})
\end{aligned} \tag{4.42}$$

For the sake of simple implementation and the accuracy of approximation, we only allow one-shot update, i.e., single replacement for a particular index w.r.t. $\boldsymbol{\beta}(n)$ is allowed :

$$\boldsymbol{\alpha}_t(n+1) = \begin{cases} s, & \text{if } t = \tau \\ \boldsymbol{\beta}_t(n), & \text{if } t \neq \tau \end{cases} \tag{4.43}$$

where s is the index of eigenvalue that we wish to achieve for next iteration for τ-th sample. Naturally one should have $s < \boldsymbol{\beta}_\tau(n)$ since one wishes to have small eigenvalue for $\boldsymbol{g}_\tau$. And we choose $\boldsymbol{\alpha}(0) = (d_2, \ldots, d_2)$ as the initial point which corresponds to the all maximum eigenvalue configuration. The procedure for finding τ is explained as follows. We sort all samples $\{\theta_i\}_{i=1}^T$ based on the its individual deviation μ_t in descending order. Starting from $\tau = \theta_1$, Eq. 4.41 and Eq. 4.43 mean that we try to update the multi-index only based on the local information provided by the parameter sample $\boldsymbol{g}^{(\theta_1)}$ whose deviation is the worst from our expectation. Therefore, if one could amend this deviation, one should obtain a better quantizer intuitively. That is, if there exists s s.t. $\overline{\mathrm{L}}\left(\mathcal{Q}^{(n+1)}; f, R, \mathcal{T}\right) \leq \overline{\mathrm{L}}\left(\mathcal{Q}^{(n)}; f, R, \mathcal{T}\right)$, then we do find a better quantizer. Otherwise one set $\tau = \theta_2$ and so on until one find a such pair (τ, s). Then as iteration goes, one could obtain a decreasing sequence of $\overline{\mathrm{L}}\left(\mathcal{Q}^{(n)}; f, R, \mathcal{T}\right)$ providing such (τ, s) always exists. Our algorithm halts if no such pair exists. In other words, Our algorithm converges to a solution where our expectation coincides with the real approximation using eigen values, i.e., $\boldsymbol{\alpha}(n) = \boldsymbol{\beta}(n)$ or we could no longer find such improvement. This method of updating multi-index will be referred as satisfactory goal-oriented quantization algorithm (SGOQ) summarized in alg. 4. Another way of updating representatives is the greedy method. Inspecting all the possibility combination of pair (τ, s) directly, we choose the one which minimize the empirical loss. Greedy version of the algorithm is resumed in alg. 5. This algorithm will be referred as greedy goal-oriented quantization algorithm (GGOQ).

4.4.2 Representatives update

Without loss of generality, we assume that τ-th sample is what we choose in multi-index update step and $\boldsymbol{g}^{(\tau)} \in \overline{\mathcal{C}}_m^{(n)}$ with m being the label of its corresponding representative. To find a better quantizer, one wish to decrease $\boldsymbol{\beta}_\tau(n)$ to s if such operation is possible $(s < \boldsymbol{\beta}_\tau(n))$. If $\boldsymbol{g}^{(\tau)}$ satisfies :

$$\frac{\boldsymbol{g}^{(\tau)} - \boldsymbol{z}_m^{(n+1)}}{\|\boldsymbol{g}^{(\tau)} - \boldsymbol{z}_m^{(n+1)}\|_2} = \boldsymbol{e}_m = \boldsymbol{\nu}_s(\boldsymbol{g}^{(\tau)}; f), \tag{4.44}$$

then one has

$$d_f(\boldsymbol{g}^{(\tau)}, \boldsymbol{z}_m^{(n+1)}) = \lambda_s(\boldsymbol{g}^{(\tau)}; f)\|\boldsymbol{g}^{(\tau)} - \boldsymbol{z}_m^{(n+1)}\|^2 \tag{4.45}$$

which corresponds perfectly one term of the sum in Eq. 4.39. This operation allows us to find a new representative $\boldsymbol{z}_m^{(n+1)}$ based on the local information of $\boldsymbol{g}^{(\tau)}$. Moreover, for any $r(n) = \|\boldsymbol{g}^{(\tau)} - \boldsymbol{z}_m^{(n+1)}\| > 0$, we could always find such $\boldsymbol{z}_m^{(n+1)}$. In other words, Eq. 4.44 is equivalent to the following rule :

$$\boldsymbol{g}^{(\tau)} - \boldsymbol{z}_m^{(n+1)} = r(n)\boldsymbol{\nu}_s(\boldsymbol{g}^{(\tau)}; f) \tag{4.46}$$

where $r(n)$ is a coefficient for $(n+1)$-th iteration . For example, this coefficient $r(n)$ could be chosen so that $\|\boldsymbol{z}_m^{(n+1)} - \boldsymbol{z}_m^{(n)}\|$ is minimized. Of course, one should always guarantee that $\boldsymbol{z}_m^{(n+1)} \in \mathcal{G}$,i.e., the new representative is still in the parameter space.

In the view of GGOQ, each improvement in SGOQ is satisfied with existence of improvement merely. For both two algorithms, the final performance criterion is chosen as the empirical optimality loss $\overline{\mathrm{L}}\left(\mathcal{Q}; f, R, \mathcal{T}\right)$. Therefore it is natural for both SGOQ and GGOQ converge to a sub-optimum in the sense of empirical optimality loss. If the

number of samples is sufficiently large, two proposed algorithms should both achieve the sub-optimality of the approximate optimality loss.

Inputs : Parameter distribution $\phi(\boldsymbol{g})$, number of samples T, number of regions M and maximum number of iterations $N_{\max}$
Outputs : $\mathcal{Q}^{\mathrm{OPT}} = \{\boldsymbol{z}_1^{\mathrm{OPT}}, ..., \boldsymbol{z}_M^{\mathrm{OPT}}\}$
Initialization : Generate a parameter sample set $\mathcal{T} = \left\{\boldsymbol{g}^{(t)}\right\}_{t=1}^{T}$ according to $\phi(\boldsymbol{g})$;
Generate $\mathcal{Q}^{(0)} = \left\{\boldsymbol{z}_m^{(0)}\right\}_{m=1}^{M}$ by using LM algorithm with $\lambda_{d_2}(\boldsymbol{g}; f)\phi(\boldsymbol{g})$ as p.d.f.
$\boldsymbol{\alpha}_t(0) \leftarrow d_2$, $\forall t = 1$ to T.
for $n = 1$ **to** $N_{\max}$ **do**
  Let $\boldsymbol{\beta}(n) \in \arg\min_{\boldsymbol{\gamma}} \left\| \widetilde{\mathrm{L}}\left(\mathcal{Q}^{(n-1)}, \boldsymbol{\gamma}; f, R, \mathcal{T}\right) - \overline{\mathrm{L}}\left(\mathcal{Q}^{(n-1)}; f, R, \mathcal{T}\right) \right\|$;
  if $\boldsymbol{\beta}(n) = \boldsymbol{\alpha}(n-1)$ **then**
    Return $\mathcal{Q}^{\mathbf{OPT}} \leftarrow \mathcal{Q}^{(n-1)}$ **;**
  end
  sort $\left\{\boldsymbol{g}^{(t)}\right\}_{t=1}^{T}$ by $\mu_t(\boldsymbol{\alpha}(n-1), \boldsymbol{\beta}(n), \mathcal{T}, \mathcal{Q}^{(n-1)})$ in descending order : $\{\theta_i\}_{i=1}^{T}$;
  $\boldsymbol{\alpha}(n) \leftarrow \boldsymbol{\beta}(n)$;
  for $i = 1$ **to** T **do**
    $\tau \leftarrow \theta_i$;
    $m_\tau \leftarrow \arg\min_{1 \le m \le d_2} d_f\left(\boldsymbol{g}^{(\tau)}, \boldsymbol{z}_m^{(n)}\right)$;
    for $s = 1$ **to** $\boldsymbol{\beta}_i(n)$ **do**
      $r_{\tau,s} \in \arg\min_{y>0:\ \boldsymbol{g}^{(\tau)} - y\boldsymbol{\nu}_s(\boldsymbol{g}^{(\tau)}; f) \in \mathcal{G}} \left\| \boldsymbol{g}^{(\tau)} - y\boldsymbol{\nu}_s(\boldsymbol{g}^{(\tau)}; f) - \boldsymbol{z}_{m_\tau}^{(n)} \right\|$;
      $\boldsymbol{\xi} \leftarrow \boldsymbol{g}^{(\tau)} - r_{\tau,s}\boldsymbol{\nu}_s(\boldsymbol{g}^{(\tau)}; f)$;
      $\widetilde{\mathcal{Q}}_{\tau,s}$ is obtained by only exchanging $\boldsymbol{\xi}$ and $\boldsymbol{z}_{m_\tau}^{(n)}$ in $\mathcal{Q}^{(n-1)}$;
      Update quantization regions according to $\widetilde{\mathcal{Q}}_{\tau,s}$;
      if $\overline{\mathrm{L}}\left(\widetilde{\mathcal{Q}}_{\tau,s}; f, R, \mathcal{T}\right) < \overline{\mathrm{L}}\left(\mathcal{Q}^{(n-1)}; f, R, \mathcal{T}\right)$ **then**
        $\boldsymbol{\alpha}_\tau(n) \leftarrow s$;
        $\mathcal{Q}^{(n)} \leftarrow \widetilde{\mathcal{Q}}_{\tau,s}$;
        Goto Step 7 **;**
      end
    end
  end
  Return $\mathcal{Q}^{\mathbf{OPT}} \leftarrow \mathcal{Q}^{(n-1)}$ **;**
end
$\mathcal{Q}^{\mathbf{OPT}} \leftarrow \mathcal{Q}^{(N_{\max})}$;

Algorithm 4: Satisfactory Goal-Oriented Quantization Algorithm

4.5 Numerical Results

4.5.1 Scalar case

We first verify the conclusion concerning the NOL estimation in scalar case. The NOL of EE-BER with $N = 10$, its approximate NOL, EE-PSTR with $c = 1$ and its approximate

Inputs : Parameter distribution $\phi(\boldsymbol{g})$, number of samples T, number of regions M and maximum number of iterations $N_{\max}$
Outputs : $\mathcal{Q}^{\mathrm{OPT}} = \{\boldsymbol{z}_1^{\mathrm{OPT}}, ..., \boldsymbol{z}_M^{\mathrm{OPT}}\}$
Initialization : Generate a sample $\mathcal{T} = \left\{\boldsymbol{g}^{(t)}\right\}_{t=1}^{T}$ according to $\phi(\boldsymbol{g})$; Initialize $\mathcal{Q}^{(0)} = \left\{\boldsymbol{z}_m^{(0)}\right\}_{m=1}^{M}$;
for $n = 0$ **to** $N_{\max}$ **do**
  Let $\boldsymbol{\beta}(n) \in \arg\min_{\boldsymbol{\gamma}} \ \|\widetilde{\mathrm{L}}\left(\mathcal{Q}^{(n)}, \boldsymbol{\gamma}; f, R, \mathcal{T}\right) - \overline{\mathrm{L}}\left(\mathcal{Q}^{(n)}; f, R, \mathcal{T}\right)\|$;
  for $i = 1$ **to** T **do**
    for $s = 1$ **to** $\boldsymbol{\beta}_i(n)$ **do**
      $r_{i,s} \in \arg \min_{y>0:\ \boldsymbol{g}^{(\tau)} - y\boldsymbol{\nu}_s(\boldsymbol{g}^{(\tau)}; f) \in \mathcal{G}} \|\boldsymbol{g}^{(\tau)} - y\boldsymbol{\nu}_s(\boldsymbol{g}^{(\tau)}; f) - \boldsymbol{z}_{m_\tau}^{(n)}\|$;
      $\boldsymbol{\xi} \leftarrow \boldsymbol{g}_i - r_{i,s}\boldsymbol{\nu}_s(\boldsymbol{g}^{(i)}; f)$;
      $\widetilde{\mathcal{Q}}_{i,s}$ is obtained by exchanging $\boldsymbol{\xi}$ and $\boldsymbol{z}_{m_i}^{(n)}$ in $\mathcal{Q}^{(n)}$;
      Update the region according to $\widetilde{\mathcal{Q}}_{i,s}$;
    end
  end
  $(i^\star, s^\star) \in \arg\min_{i,s} \overline{\mathrm{L}}\left(\widetilde{\mathcal{Q}}_{i,s}; f, R, \mathcal{T}\right)$;
  if $\overline{\mathrm{L}}\left(\widetilde{\mathcal{Q}}_{i^\star,s^\star}; f, R, \mathcal{T}\right) < \overline{\mathrm{L}}\left(\mathcal{Q}^{(n)}; f, R, \mathcal{T}\right)$ **then**
    $\mathcal{Q}^{(n+1)} \leftarrow \widetilde{\mathcal{Q}}_{i^\star,s^\star}$;
  else
    Return $\mathcal{Q}^{\mathbf{OPT}} \leftarrow \mathcal{Q}^{(n)}$ **;**
  end
end
$\mathcal{Q}^{\mathbf{OPT}} \leftarrow \mathcal{Q}^{(N_{\max})}$;

Algorithm 5: Greedy Goal-Oriented Quantization Algorithm

NOL as a function of number of cells are illustrated in Fig. 4.1. The approximate NOL is relatively accurate in high-resolution regime for both EE-PSTR and EE-BER while the approximate NOL for EE-PSTR slightly diverge from the real NOL in low-resolution regime for EE-PSTR. Interestingly, the conclusion that EE-BER is hard to compress than EE-PSTR still holds even for non high-resolution regime. It shows our approximation on NOL could effectively represent the hardness of compression without performing simulations. Nevertheless, this results could merely remains valid for probability density being uniform on $[0, 1]$. It is highly possible that one obtains the contrary result if the probability density is different.

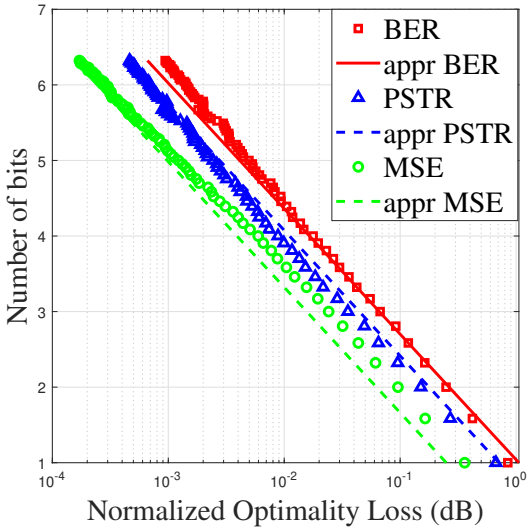


FIGURE 4.1 – Number of bits as a function of the normalized optimality loss (NOL) of EE (bit error rate) with $N = 10$ provided by enhanced LM algorithm (probability distribution is replaced by normalized value density), its approximate NOL , EE (packet success transmission rate) with $c = 1$ and its approximate NOL and MSE and its approximate NOL in dB. This figure reveals the accuracy of our high-resolution approximation. Our approximation in high-resolution regime could explain easily the hardness of quantization for different cost functions in scalar case.

Fig. 4.2 illustrates the relative optimality loss as a function of quantization bits for Lloyd-Max algorithm, enhanced Llyod-Max algorithm (taking normalized value density as the p.d.f. of parameter), IWO-DE algorithm for EE-BER cost function $f(x; g) = -\frac{(1-\exp(-gx))^N}{x}$ with $N = 10$ used in [2]. In low-resolution regime, IWO-DE algorithm has better performance than Lloyd-Max algorithm which dominates slightly the enhanced Lloyd-Max algorithm. However, starting from the moderate regime, the enhanced Lloyd-Max algorithm obviously outperforms than other two approaches and this dominance is even larger in high resolution regime. Take $M = 25$ as an example, our proposed approach could provide a reduction of half optimality loss. This results entails our analysis on high resolution regime is useful. Besides, the running time for IWO-DE algorithm and enhanced Lloyd-Max algorithm as a function of number of bits are listed in Tab II. One can easily observe that the running time for enhanced Lloyd-Max algorithm is almost a constant while the the running time grows tremendously as the number of bits increases for IWO-DE algorithm. This comparison entails enhanced Llyod-Max algorithm is more efficient than IWO-DE algorithm.

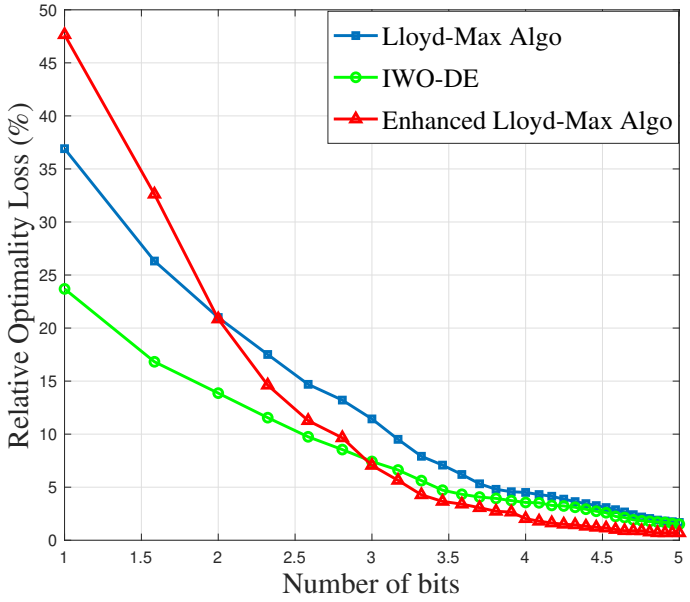


FIGURE 4.2 – Relative optimality loss as a function of number of bits of quantization for LM algorithm, IWO-DE algorithm, enhanced Lloyd-Max algorithm(using value density instead of the original p.d.f.) for EE (bit error rate) cost function $f(x;g) = -\frac{(1-\exp(-gx))^N}{x}$ with $N = 10$. the optimality loss reduction brought by enhanced Lloyd-Max algorithm demonstrates the usefulness of value density which characterizes the contribution of a parameter for the optimality loss much better than the p.d.f. of parameter.

Quantization Bits	3bits	4bits	5bits	6bits
Enhanced LM Algo	1.084s	1.132s	1.085s	1.120s
IWO-DE Algo	3.671s	4.282s	6.541s	10.530s

TABLE 4.2 – Table of running time v.s. quantization bits for enhanced Lloyd-Max algorithm and IWO-DE algorithm

4.5.2 Vector case

We start with a quadratic cost function $d_1 = d_2 = 2$. Consider the following type of function :

$$f^{\mathrm{QUA}}(\boldsymbol{x};\boldsymbol{g}) = (\boldsymbol{x}_1 - h_1(\boldsymbol{g}))^2 + (\boldsymbol{x}_2 - h_2(\boldsymbol{g}))^2 + (\boldsymbol{x}_1 - \boldsymbol{x}_2)^2 \tag{4.47}$$

with $h_1(\boldsymbol{g}) = 2\boldsymbol{g}_1\boldsymbol{g}_2 - \frac{1}{2}\boldsymbol{g}_1^2\boldsymbol{g}_2^2$ and $h_2(\boldsymbol{g}) = \boldsymbol{g}_1^2\boldsymbol{g}_2^2 - \boldsymbol{g}_1\boldsymbol{g}_2$. One has $\boldsymbol{\kappa}(\boldsymbol{g}) = [\boldsymbol{g}_1\boldsymbol{g}_2, \frac{1}{2}\boldsymbol{g}_1^2\boldsymbol{g}_2^2]$

$$\mathrm{H}_{f^{\mathrm{QUA}}}(\boldsymbol{\kappa}(\boldsymbol{g});\boldsymbol{g}) = \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}, \tag{4.48}$$

and we denote

$$\mathrm{J}_{\boldsymbol{\kappa}}(\boldsymbol{g}) = \begin{bmatrix} \frac{\partial \boldsymbol{\kappa}_1}{\partial \boldsymbol{g}_1} & \frac{\partial \boldsymbol{\kappa}_1}{\partial \boldsymbol{g}_2} \\ \frac{\partial \boldsymbol{\kappa}_2}{\partial \boldsymbol{g}_1} & \frac{\partial \boldsymbol{\kappa}_2}{\partial \boldsymbol{g}_2} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{4.49}$$

then one has

$$\begin{aligned} &\mathrm{A}_{\boldsymbol{\kappa},f^{\mathrm{QUA}}}(\boldsymbol{g}) \\ &= \begin{bmatrix} 4a^2 - 4ac + 4c^2 & 4ab + 4cd - 2bc - 2ad \\ 4ab + 4cd - 2bc - 2ad & 4b^2 - 4bd + 4d^2 \end{bmatrix} \end{aligned} \tag{4.50}$$

.

By introducing $A = 4a^2 - 4ac + 4c^2$, $B = 4b^2 - 4bd + 4d^2$ and $C = 4ab + 4cd - 2bc - 2ad$, one finally has $\lambda_2(\boldsymbol{g}; f^{\mathrm{QUA}}) = \frac{1}{2}(A + B + \sqrt{(A-B)^2 + 4C^2})$ and $\lambda_1(\boldsymbol{g}; f^{\mathrm{QUA}}) = \frac{1}{2}(A + B - \sqrt{(A-B)^2 + 4C^2})$. Fig. 4.3 illustrates the optimality loss as a function of number of cells for Lloyd-Max algorithm, approximate upper bound in Eq. 4.27, enhanced Lloyd-Max algorithm (using $\lambda_{d_2}\left(\boldsymbol{g}; f^{\mathrm{QUA}}\right)\phi\left(\boldsymbol{g}; f^{\mathrm{QUA}}\right)$ instead of $\phi(\boldsymbol{g})$ as parameter p.d.f.), SGOQ and GGOQ . One could observe that approximate upper bound is always below the Lloyd-Max algorithm. GGOQ always dominates SGOQ while both of them could reduce the optimality loss up to 12 dB than Lloyd-Max algorithm for moderate regime. Moreover, we notice that enhanced Lloyd-Max algorithm and SGOQ are quite closed to approximate upper bound (Eq. 4.27) which entails $\lambda_{d_2}(\boldsymbol{g}; f^{\mathrm{QUA}})\phi(\boldsymbol{g})$ could represent the contribution of paramter to optimality loss roughly. Therefore, in Fig. 4.4, the original p.d.f. of parameter and the new "density" weighted by the maximum eigenvalue function $\lambda_2(\boldsymbol{g}; f^{\mathrm{QUA}})\phi(\boldsymbol{g})$ is illustrated . One could observe that, different from probability distribution $\phi(\boldsymbol{g})$, $\lambda_2(\boldsymbol{g}; f^{\mathrm{QUA}})\phi(\boldsymbol{g})$ assigns almost inverse weight to parameter $(\boldsymbol{g}_1, \boldsymbol{g}_2)$ which entails the impact of cost function on parameter space in goal-oriented quantization.

We then verify the result of our constrained extension. We consider another widely-known cost function $f^{\mathrm{SL}}(\boldsymbol{x};\boldsymbol{g}) = -\sum_{i=1}^{S} \log(1 + \boldsymbol{x}_i\boldsymbol{g}_i)$ under maximum power constraint $\sum_{i=1}^{S} \boldsymbol{x}_i \leq P_{\max}$. The optimal decision function for this function is the famous water

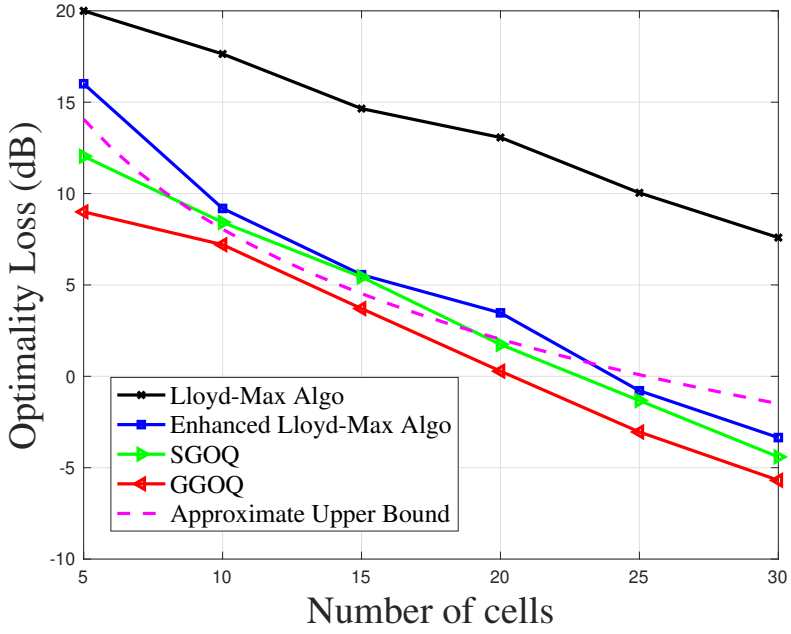
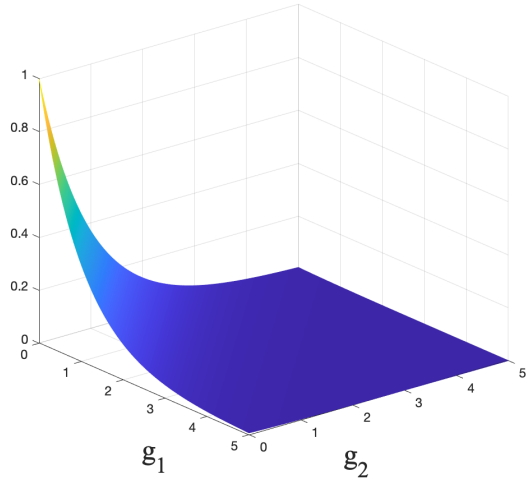


FIGURE 4.3 – Optimality loss (dB) as a function of number of cells for Lloyd-Max algorithm, approximate upper bound in Eq. 4.27, enhanced Lloyd-Max algorithm, SGOQ and GGOQ.

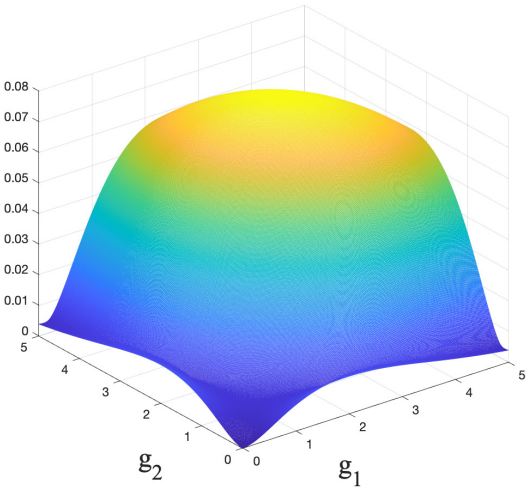
filling solution : $\boldsymbol{\kappa}_i(\boldsymbol{g}) = \left(\chi_{\boldsymbol{g}} - \frac{1}{\boldsymbol{g}_i}\right)_+$ with $\chi_{\boldsymbol{g}}$ s.t. $\sum_i \left(\chi_{\boldsymbol{g}} - \frac{1}{\boldsymbol{g}_i}\right)_+ = P_{\max}$. One could verify that $J_{\boldsymbol{\kappa}}(\boldsymbol{g}) = \operatorname{diag}\left\{\frac{\mathbb{1}\{\boldsymbol{g}_i\chi_{\boldsymbol{g}}-1>0\}}{\boldsymbol{g}_i^2}\right\}_i$, $H_{f^{\mathrm{SL}}}(\boldsymbol{\kappa}(\boldsymbol{g});\boldsymbol{g}) = \operatorname{diag}\left\{\frac{\boldsymbol{g}_i^2}{\left(1+(\boldsymbol{g}_i\chi_{\boldsymbol{g}}-1)^+\right)^2}\right\}_i$ and $\mathrm{A}_{f,\boldsymbol{\kappa}}(\boldsymbol{g}) = \operatorname{diag}\left\{\frac{\mathbb{1}\{\boldsymbol{g}_i\chi_{\boldsymbol{g}}-1>0\}}{\boldsymbol{g}_i^2\left(1+(\boldsymbol{g}_i\chi_{\boldsymbol{g}}-1)_+\right)^2}\right\}_i$. Notice that for water filling solution, the maximum power constraint is always activated, here we lost the freedom of choosing the number of representatives in boundary.

Fig. 4.5 illustrates the relative optimality loss as a function of number of cells for our proposed SGOQ and GGOQ algorithm, Llyod-Max algorithm and enhanced Lloyd-Max algorithm (using $\lambda_1\left(\boldsymbol{g}; f^{\mathrm{SL}}\right)\phi(\boldsymbol{g})$ instead of $\phi(\boldsymbol{g})$) for $d_1 = d_2 = 4$ and $P_{\max} = 20$mW. Both upper-bound (Eq. 4.27) and lower bound (Eq. 4.28) are not illustrated in Fig. 4.5 for the following reasoning. One could easily find that $\lambda_1(\boldsymbol{g}; f^{\mathrm{SL}}) \propto \frac{1}{\min_i \boldsymbol{g}_i^4}$ and $\lambda_{d_2}(\boldsymbol{g}; f^{\mathrm{SL}}) \propto \frac{1}{\max_i \boldsymbol{g}_i^4}$. Therefore the matrix $\mathrm{A}_{\boldsymbol{\kappa},f^{\mathrm{SL}}}(\boldsymbol{g})$ is stronly ill-conditioned which means these two bounds are useless in this scenario. One could observe that both two proposed algorithm and enhanced Lloyd-Max algorithm outperform Lloyd-Max algorithm algorithm in whole range. Besides GGOQ and SGOQ behave better than Enhanced Lloyd-Max algorithm while GGOQ always outperforms SGOQ for around 4 dB. If one takes the complexity into account, SGOQ could be better since it does not require a full check of Td_2 combinations.

Finally, we would like study the relative optimality loss as a function of the dimension of the problem. We still choose the sum-rate capacity function as our utility fuctnion. Fig. 4.6 illustrates relative optimality as a function of the dimension of problem $d_1 = d_2$ for SGOQ and GGOQ. We choose the following setting : number of cells $M = 32$, maximum power $P_{\max} = 20$mW. For both two algorithms, the relative optimality loss grows as the dimension increases. This could be due to the curse of dimension. The generated samples



(a) Original probability distribution $\boldsymbol{\phi}(\boldsymbol{g})$



(b) New density weighted by maximum eigenvalue function $\lambda_2(\boldsymbol{g}; f^{\mathrm{QUA}})$

FIGURE 4.4 – Comparison between a two-dimensional exponential distribution $\boldsymbol{\phi}(\boldsymbol{g})$ and its new density weighted by maximum eigenvalue function $\lambda_2(\boldsymbol{g}; f^{\mathrm{QUA}})$ of cost function f^{QUA}. This figure shows that the contribution of parameter point to the optimality loss could be completely contrary for goal-oriented quantization and conventional distortion-oriented quantization.

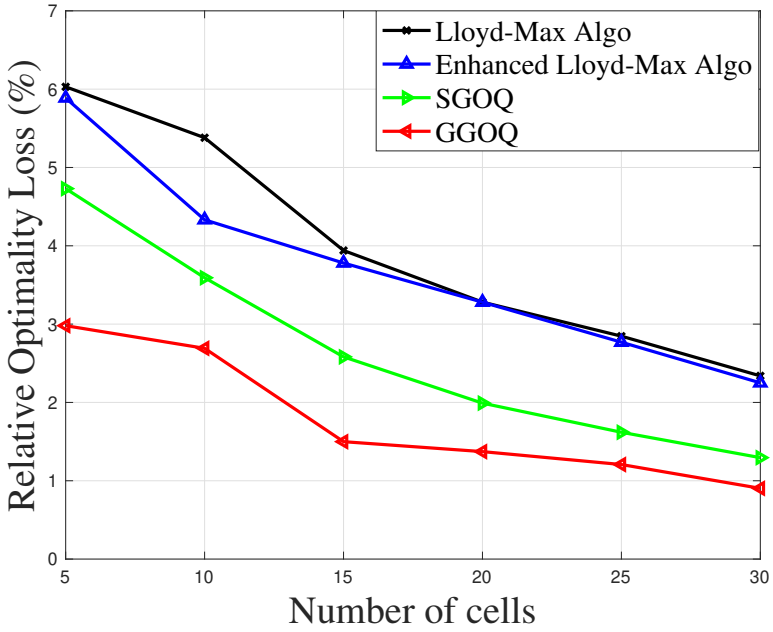


FIGURE 4.5 – Relative optimality loss v.s. number of cells for Lloyd-Max algorithm, enhanced Lloyd-Max algorithm (using weighted parameter distribution $\lambda_1\left(\boldsymbol{g}; f^{\mathrm{SL}}\right)\phi\left(\boldsymbol{g}\right)$ instead of $\phi\left(\boldsymbol{g}\right)$), SGOQ and GGOQ. Here $d_2 = d_1 = 4$ and $P_{\max} = 20\mathrm{mW}$. Proposed two algorithms could largely reduce the relative optimality loss compared to Lloyd-Max algorithms (enhanced version includes).

set is hard to cover the entire parameter space. In small-scale system, GGOQ slightly dominates SGOQ. However, the performance of two version of algorithm are really close in large dimension regime. Taking the complexity into account. Thus SGOQ is encouraging to use in practice than GGOQ in such regime.

4.6 Conclusions

In this chapter, we analyzes the goal-oriented quantization problem in high-resolution regime and discuss how to solve the problem for scalar case and vector case separately. For scalar case, the proposed new approximate formula of optimality loss leads to a new quality defined as value density representing the importance of parameter. We introduce a new quality called normalized optimality loss when comparing the hardness of quantization for different cost functions. By merely approximating this quality in high resolution regime, we are capable to determine the hardness of quantization for different cost functions without performing real simulations. For vector case, an cell-independent approximated formula for optimality loss is no longer possible for optimality loss due to the unknown of tessellating cell shape. Nevertheless, by admitting the Gersho's conjecture, an upper bound and lower bound are derived for optimality loss. Moreover, we propose a new algorithm by iteratively update the representatives based on the eigenvalue approximation to design a goal-oriented quantizer. In each iteration, one tries to find the worst parameter sample in the sense of introducing the largest individual optimality loss. Then its corresponding representative is revised so that the average optimality loss could be de-

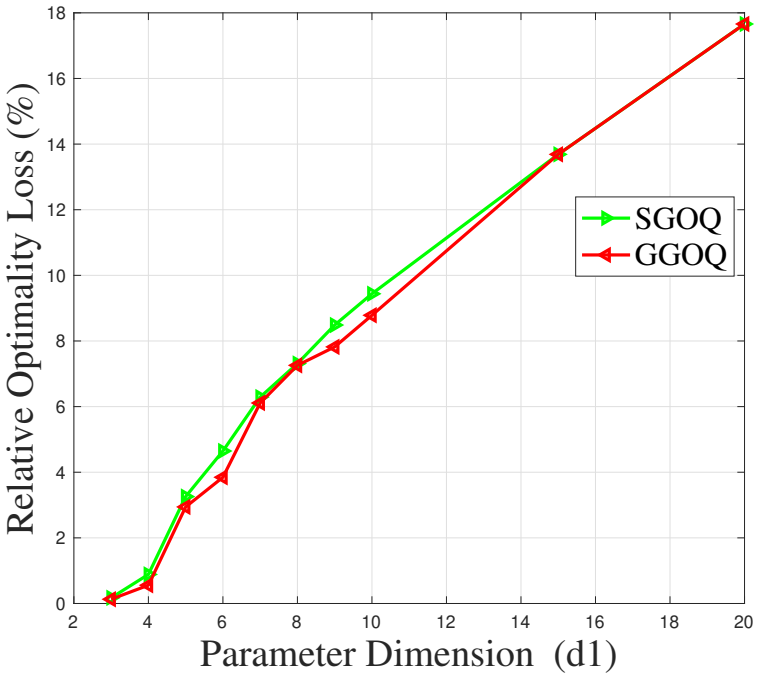


FIGURE 4.6 – Relative optimality loss v.s. dimension of the problem $d_1 = d_2$ for Lloyd-Max algorithm, SGOQ and GGOQ. Number of cells are $M = 32$ and maximum power is $P_{\max} = 20$mW. The performance of GGOQ and SGOQ are close to each other. The performance of both algorithms worsen if the dimeson of parameter increases.

creased probably. Proposed algorithm could be extended to cost function with constraints as well. Simulation results show that proposed goal-oriented quantizer outperforms the Lloyd-Max quantizer for any number of quantization bits largely. Algorithm with greedy update slightly dominates the one with satisfying update while the later takes much less time to operate. However, extending our proposed method in vector case remains cumbersome for high-dimensional scenario of the problem. Besides, the selection of parameter sample set could also be essential owing to both the samples themselves and the number of parameter sample which impacts the complexity of the proposed algorithm.

5

Goal-Oriented Quantization in Potential Games

In this chapter, we start to tackle goal-oriented quantization with multiple correlative utility functions targeted by different users of the system modeled as a game in stragetic form. More specifically, we restrict ourselves in the study of potential games with identical action space. Taking the social welfare as our performance criterion, we have proven that the maximum social welfare under refined Nash equilibrium could be a submodular function of the action set for some conditions. Based on this property, we design an algorithm to find an action set aiming at maximizing the social welfare. We apply our framework to a multiple access channel game where the spectral efficiency taken as the individual utility of user. Tremendous optimality loss reduction is confirmed compared to the conventional quantization paradigm.

5.1 Motivation

In previous chapters, we always focus on how to find a goal-oriented quantizer for a **single** utility (cost) function. This setting actually corresponds to the scenario where a transmitter-receiver pair in a communication system aims at maximizing its proper utility function serving its corresponding user. However, it is more reasonable to assume that users in a system have different objective, i.e., different utility functions. Moreover, these different utility function could be correlated, for example, through the interference between channels. This thus motivates us to explore the goal-oriented quantization problem for multiple utility functions, for example, to develop it in a framework of strategic game. Moving from single-objective optimization problem to a game, of course, introduces other difficulties, e.g., existence of solution concept such as Nash equilibrium (NE), correlated equilibrium and achievability of solution concepts as well. Meanwhile the famous Braess's Paradox (BP) [99, 100] could be extremely beneficial for the goal-oriented quantization

problem with multiple objectives. BP could be interpreted as adding one or more roads to a road network can slow down overall traffic flow through it in. This paradox which is widely observed and proved in many domains, e.g., wireless communication and traffic network [90, 103] could be the essential difference between optimization problems and games. For our goal-oriented quantization, BP entails the possibility that the restriction on the action set of player due to quantization in a game could improve the overall performance of the communication system, for example, the social welfare. This result is fundamentally different from single-objective goal-oriented quantization problem where loss always exists after quantization.

5.2 Game Theory Basics and Problem Formulation

In this section, we will firstly give some basic concepts of any game-theoretic analysis before formulate the goal-oriented quantization problem. First of all, we introduce the strategic form of a game :

Definition 5.2.1. *(Strategic Game) A game is a triplet* $\mathbb{G} = \left(\mathcal{K}, (\mathcal{A}_k)_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$*, where* $\mathcal{K} = \{1, \ldots, K\}$ *is the set of players,* $\mathcal{A}_k$ *is the action space of* k*-th player and* f_k *is the utility function*[1] *of* k*-th player .*

The central concept of game-theoretic analysis is Nash Equilibrium (NE) defined as :

Definition 5.2.2. *(Nash Equilibrium) For game* $\mathbb{G} = \left(\mathcal{K}, (\mathcal{A}_k)_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$*, an action profile* $\boldsymbol{a} = (\boldsymbol{a}_k, \boldsymbol{a}_{-k})$ *is called a Nash equilibrium if for* $\forall k \in \mathcal{K}$ *and* $\forall \boldsymbol{a}' = (\boldsymbol{a}'_k, \boldsymbol{a}_{-k})$ *:*

$$f_k\left(\boldsymbol{a}_k, \boldsymbol{a}_{-k}; \boldsymbol{g}\right) \geq f_k\left(\boldsymbol{a}'_k, \boldsymbol{a}_{-k}; \boldsymbol{g}\right) \tag{5.1}$$

where $\boldsymbol{a} = (\boldsymbol{a}_k, \boldsymbol{a}_{-k})$ with $\boldsymbol{a}_{-k} \triangleq (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k-1}, \boldsymbol{a}_{k+1} \ldots, \boldsymbol{a}_K) \in \mathcal{A}_{-k}$ and $\mathcal{A}_{-k} \triangleq \mathcal{A}_1 \times \cdots \times \mathcal{A}_{k-1} \times \mathcal{A}_{k+1} \times \cdots \times \mathcal{A}_K$ are standard notations of game theory. Similarly the vector $\boldsymbol{g} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_K)$ contains the parameter information of each player. The meaning of NE is that any unilateral change of action at this point won't lead to an increase of individual benefit. Furthermore, we introduce an important conception in game-theoretic analysis known as best response (BR).

Definition 5.2.3. *(Best Response) : In a non-cooperative game* $\mathbb{G}$*, the correspondence* $\mathrm{BR}_k\left(\boldsymbol{a}_{-k}; ; \boldsymbol{g}\right) : \mathcal{A}_{-k} \to \mathcal{A}_k$ *s.t.*

$$\mathrm{BR}_k\left(\boldsymbol{a}_{-k}; \boldsymbol{g}\right) \triangleq \arg\max_{\boldsymbol{a}_k \in \mathcal{A}_k} f_k\left(\boldsymbol{a}_k, \boldsymbol{a}_{-k}; \boldsymbol{g}\right) \tag{5.2}$$

is called the best response (BR) of player $k \in \mathcal{K}$ *given the action profile of other players* $\boldsymbol{a}_{-k}$*. From the definition of best response, one has immediately the following characterization for NE :*

Proposition 5.2.4. *[Nash,1950] An action profile* $\boldsymbol{a}^\star$ *is an NE if and only if :* $\forall k \in \mathcal{K}$*,* $\boldsymbol{a}^\star_k \in \mathrm{BR}_k\left(\boldsymbol{a}^\star_{-k}\right)$*.*

1. More generally, one can use preference order to replace utility function since the existence of utility function is not guaranteed for free. However, for goal-oriented quantization, the existence of utility function is for sure or predetermined.

We try to tackle the goal-oriented quantization problem in the following scenario. In a communication system with K transmitter-receiver pairs. Each Tx-Rx pair serves a user labeled k (player) and aims at maximizing this user's utility function $f_k(\boldsymbol{a}_k, \boldsymbol{a}_{-k}; \boldsymbol{g})$ by choosing its action $\boldsymbol{a}_k$ amongst its action space $\mathcal{A}_k$ depending on $\boldsymbol{a}_{-k}$ for parameter $\boldsymbol{g}$. Utility functions are correlated through the interference of channel between users. Moreover, we assume the decision (action) of each users is made simultaneously. Obviously this scenario could be regarded as the generalization of the single-goal goal-oriented communication system. There the instantaneous objective of the communication system (with parameter $\boldsymbol{g}$) in this scenario could be formulated as a couple of optimization problem whose solution is exactly the Nash equilibrium of the game defined as $\mathbb{G} = \left(\mathcal{K}, (\mathcal{A}_k)_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$.

To start with, we assume that all user has **identical action space** : $\mathcal{A}_k = \mathcal{X},\ \forall k \in \mathcal{K}$. The situation where players' action sets are different will be the extension of current framework. This scenario could be applied to domains such as Internet of things (IoT) where tremendous devices, e.g., sensors and actuators are deployed in the networks [85, 86]. Each device could react differently according to its surroundings while the underlying action space is the same. This assumption is useful since we only need to find a single goal-oriented quantizer for the entire system. A goal-oriented quantizer $\mathcal{Q}$ with M quantization regions (cardinality of decision set as well) could be fully characterized by quantization regions $\{\mathcal{C}_m\}_{m=1}^{M}$ and the corresponding decision set $\mathcal{D} \triangleq \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M\}$. Again, details about finding quantization regions are omitted here. We denote the new game where the action set for each user is always $\mathcal{D}$ by $\mathbb{G}^{\mathrm{FA}} \triangleq \left(\mathcal{K}, (\mathcal{D})_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$ where FA means user has merely **finite actions** to choose. The original game is denoted as $\mathbb{G} = \left(\mathcal{K}, (\mathcal{X})_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$. The set of all possible NE of the game $\mathbb{G} \triangleq \left(\mathcal{K}, (\mathcal{D})_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$ for a given parameter $\boldsymbol{g}$ is denoted as $\mathcal{N}[\mathcal{D}; \boldsymbol{g}]$. As the designer of the communication system, it would be reasonable for us to consider the overall performance. One important quality characterizes the overall performance of the system is the social welfare defined as :

$$w(\boldsymbol{a}; \boldsymbol{g}) \triangleq \sum_{k\in\mathcal{K}} f_k(\boldsymbol{a}; \boldsymbol{g}) \tag{5.3}$$

and we define the maximum social welfare under NE for a given action set $\mathcal{D}$:

$$\omega(\mathcal{D}; \boldsymbol{g}) \triangleq \max_{\boldsymbol{a}\in\mathcal{N}[\mathcal{D};\boldsymbol{g}]} w(\boldsymbol{a}; \boldsymbol{g}) \tag{5.4}$$

Furthermore, we define the average maximum optimal social welfare $\Omega(\mathcal{D})$ under NE for a given action set $\mathcal{D}$ as :

$$\Omega(\mathcal{D}) = \mathbb{E}_{\boldsymbol{g}}\left[\omega(\mathcal{D}; \boldsymbol{g})\right] \tag{5.5}$$

To this end, the problem of finding a goal-oriented quantizer (with M quantization regions) with multiple utility functions $(f_k)_{k=1}^{K}$ could be formulated as :

$$\begin{aligned} &\max_{\mathcal{D}\subseteq\mathcal{X}}\ \Omega(\mathcal{D}) \\ &\text{s.t. } |\mathcal{D}| = M. \end{aligned} \tag{5.6}$$

Before ending this section, we will show the connection between the famous Braess's Paradox and our goal-oriented quantization problem with multiple objectives. In our settings, we say there exists a Braess's Paradox for game $\mathbb{G} = \left(\mathcal{K}, (\mathcal{X})_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$, if there exists an action set $\mathcal{D} \subseteq \mathcal{X}$ s.t. $\omega(\mathcal{D}; \boldsymbol{g}) > \omega(\mathcal{X}; \boldsymbol{g})$. The interest of BP for a

game is that when players are restricted on a subset of the original action space, the overall performance of the system is surprisingly improved. Considering our goal-oriented quantization problem, there are two problems could be of great importance for us : i) for a game $\mathbb{G} = \left(\mathcal{K}, (\mathcal{X})_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$, what is condition for the existence of Braess paradox ? ii) if Braess's Paradox does exist, how to find the optimal action set maximizing the function ω :

$$\mathcal{D}^{\mathrm{OPT}}(\boldsymbol{g}) \in \arg\max_{\mathcal{D}\subseteq\mathcal{X}} \omega(\mathcal{D};\boldsymbol{g}) \tag{5.7}$$

Consider firstly the trivial case of the problem. Define $\mathcal{A}^\star(\boldsymbol{g})$ the set of centralized solution for a given parameter $\boldsymbol{g}$:

$$\mathcal{A}^\star(\boldsymbol{g}) \triangleq \arg\max_{\boldsymbol{a}\in\mathcal{X}^K} w(\boldsymbol{a};\boldsymbol{g}), \tag{5.8}$$

and the corresponding maximum of social welfare

$$w^\star(\boldsymbol{g}) = \max_{\boldsymbol{a}\in\mathcal{X}^K} w(\boldsymbol{a};\boldsymbol{g}) \tag{5.9}$$

If there exists $\mathcal{D}_0$ s.t. $\mathcal{N}[\mathcal{D}_0;\boldsymbol{g}]\bigcap\mathcal{A}^\star(\boldsymbol{g}) \neq \varnothing$, then it is obvious that $\omega(\mathcal{D}_0;\boldsymbol{g}) = w^\star(\boldsymbol{g}) > \omega(\mathcal{X};\boldsymbol{g})$. In this trivial case $\boldsymbol{g} \in \mathcal{G}_1 \triangleq \{\boldsymbol{g} : \exists\mathcal{D}\subseteq\mathcal{X} \text{ s.t. } \mathcal{N}[\mathcal{D};\boldsymbol{g}]\bigcap\mathcal{A}^\star(\boldsymbol{g}) \neq \varnothing\}$, the existence of BP is always guaranteed and the optimal action set $\mathcal{D}^\star(\boldsymbol{g})$ an be easily constructed. Therefore the difficulty of verifying the existence of BP lies in the non-trivial case : $\boldsymbol{g} \in \mathcal{G}_2 \triangleq \{\boldsymbol{g} : \forall\mathcal{D}\subseteq\mathcal{X},\ \mathcal{N}[\mathcal{D};\boldsymbol{g}]\bigcap\mathcal{A}^\star(\boldsymbol{g}) = \varnothing\}$. To study the existence of BP in non-trivial case $\mathcal{G}_2$, we need to examine the basic properties of function $\omega(\mathcal{D};\boldsymbol{g})$. For general games, it is hard to give further comments on that. However, as we will show in next section, for a special category of game, namely, deeper analysis is possible.

5.3 Analysis of Potential Games

OP in Eq. 5.6 is generally difficult to tackle since two functions $\omega(\mathcal{D};\boldsymbol{g})$ and $\Omega(\mathcal{D})$ require a full knowledge of information about NE of the game $\mathbb{G}^{\mathrm{FA}}$. Instead of focusing on general game, we will show that our goal-oriented quantization problem is at least solvable for the potential games.

5.3.1 Introduction to potential games

Potential games have been first introduced and studied by Monderer and Shapley in [105]. In wireless communications, various different problems are formulated as potential games for e.g., see [90, 91, 103]. In this section, we limit ourselves in exact potential game defined as :

Definition 5.3.1. *(Exact Potential Game) : Game* $\mathbb{G} = \left(\mathcal{K}, (\mathcal{A}_k)_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$ *is a potential game if there exist a potential function* φ *s.t. for all player* $k \in \mathcal{K}$ *and two action* $\boldsymbol{a}_k$ *and* $\boldsymbol{a}'_k$*, it holds that*

$$f_k(\boldsymbol{a}_k,\boldsymbol{a}_{-k};\boldsymbol{g}) - f_k(\boldsymbol{a}'_k,\boldsymbol{a}_{-k};\boldsymbol{g}) = \varphi(\boldsymbol{a}_k,\boldsymbol{a}_{-k};\boldsymbol{g}) - \varphi(\boldsymbol{a}'_k,\boldsymbol{a}_{-k};\boldsymbol{g}) \tag{5.10}$$

Existence and convergence of Nash equilibrium do not always apply to arbitrary utility functions and strategy sets. However, we have the following proposition assuring the existence of Nash equilibrium under certain mild conditions for potential games :

Proposition 5.3.2. *[Monderer-Shapley,1996] For a potential game* $\mathbb{G}$ *with finite number of players and either non-empty compact strategy sets and continuous utilities or finite non-empty strategy sets, then it has at least one Nash Equilibrium.*

For exact potential game $\mathbb{G}^{\mathrm{FA}} = \left(\mathcal{K}, (\mathcal{D})_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$ with potential function φ, we define the set of all possible NE maximizing the potential function given parameter $\boldsymbol{g}$:

$$\overline{\mathcal{N}}\left[\mathcal{D};\boldsymbol{g}\right] \triangleq \arg\max_{\boldsymbol{a}\in\mathcal{D}^K} \varphi\left(\boldsymbol{a};\boldsymbol{g}\right). \tag{5.11}$$

Above definition is reasonable since the argmax set of potential function is a subset of the set containing all NE of the potential game, i.e., $\overline{\mathcal{N}}\left[\mathcal{D};\boldsymbol{g}\right] \subset \mathcal{N}\left[\mathcal{D};\boldsymbol{g}\right]$. Similarly to functions $\omega\left(\mathcal{D};\boldsymbol{g}\right)$ and $\omega\left(\mathcal{D}\right)$, we introduce

$$\overline{\omega}\left(\mathcal{D};\boldsymbol{g}\right) \triangleq \max_{\boldsymbol{a}\in\overline{\mathcal{N}}[\mathcal{D};\boldsymbol{g}]} w\left(\boldsymbol{a};\boldsymbol{g}\right) \tag{5.12}$$

and

$$\overline{\Omega}(\mathcal{D}) = \mathbb{E}_{\boldsymbol{g}}\left[\overline{\omega}\left(\mathcal{D};\boldsymbol{g}\right)\right] \tag{5.13}$$

The concept of Nash equilibrium refined in the argmax set of potential could be useful for our goal-oriented quantization problem formulated in Eq. 5.6 since it actually simplifies the problem of searching NE of a game a to a simple optimization problem of the potential function, since in many games, it could be hard to determine and achieve all NEs. However this refinement does not conserve the optimality anymore, i.e., $\overline{\omega}\left(\mathcal{D};\boldsymbol{g}\right) \leq \omega\left(\mathcal{D};\boldsymbol{g}\right)$ and $\overline{\Omega}(\mathcal{D}) \leq \Omega(\mathcal{D})$. In next subsection, we will study the basic property for these functions to help us understand our problem better.

5.3.2 Basic property of function ω and $\overline{\omega}$

We first give a necessary condition for the existence of Braess's paradox. To do so, we define the monotonicity for set function, i.e., function $u : 2^{\mathcal{V}} \to \mathbb{R}$ that assign each subset $\mathcal{B} \subset \mathcal{V}$ a real value $u\left(\mathcal{B}\right)$. $2^{\mathcal{V}}$ representing the set of all subsets of ground set $\mathcal{V}$. Usually, we assume $u\left(\emptyset\right) = 0$.

Definition 5.3.3. *(Monotonicity) set function* $u : 2^{\mathcal{V}} \to \mathbb{R}$ *is said to be monotone if for any* $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \mathcal{V}$, $u\left(\mathcal{B}_1\right) \leq u\left(\mathcal{B}_2\right)$.

Obviously, we have the following necessary condition for the existence of BP :

Proposition 5.3.4. *(BP implies non-monotonicity of* ω*) For given parameter* $\boldsymbol{g}$*, if Braess's paradox exists for potential game* $\mathbb{G} = \left(\mathcal{K}, (\mathcal{X})_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$*, then the set function* $\omega\left(\mathcal{D};\boldsymbol{g}\right)$ *is non-monotone.*

Proof : The existence of Braess's Paradox means that there exists a subset $\mathcal{D}_0$ of $\mathcal{X}$ s.t. $\omega\left(\mathcal{D}_0;\boldsymbol{g}\right) > \omega\left(\mathcal{X};\boldsymbol{g}\right)$. ■

We prove then an important property satisfied by function $\overline{\omega}$: submodularity. Submodularity [108, 109, 117] is a functional property with great importance. The submodularity property enables striking algorithm-friendly features and is observed in a good number of application scenarios. These benefits have drawn attention in many different scenarios. For example, sensor selection in [110], detection in [111], resource allocations in [112] and adversarial attacks in [115].

Proposition 5.3.5. *(Submodularity of $\overline{\omega}$) function $\overline{\omega}$ is submodular, i.e., for any $\mathcal{B}_1, \mathcal{B}_2 \subseteq \mathcal{X}$, it holds that :*

$$\overline{\omega}\left(\mathcal{B}_1 \cup \mathcal{B}_2; \boldsymbol{g}\right) + \overline{\omega}\left(\mathcal{B}_1 \cap \mathcal{B}_2; \boldsymbol{g}\right) \leq \overline{\omega}\left(\mathcal{B}_1; \boldsymbol{g}\right) + \overline{\omega}\left(\mathcal{B}_2; \boldsymbol{g}\right). \tag{5.14}$$

Proof : See Appendix C. ■

Remark 5.3.6. *The submodularity of function $\overline{\omega}\left(\mathcal{D}; \boldsymbol{g}\right)$ actually holds for any action set $\mathcal{D}$ regardless $\mathcal{D}$ being finite or not. This conclusion entails the submodularity of function $\overline{\omega}$ comes directly from the property of potential game.*

An intuitive explanation for submodularity is that "diminishing returns", i.e., "smaller" action set tends to have higher revenue. To see this, we need to define the following discrete derivative :

Definition 5.3.7. *(Discrete derivative) For a set function $u : 2^{\mathcal{V}} \to \mathbb{R}$, $\mathcal{B} \subset \mathcal{V}$ and $e \in \mathcal{V}$, let $\Delta_u\left(e \,|\mathcal{B}\right) \triangleq u\left(\mathcal{B} \cup \{e\}\right) - u\left(\mathcal{B}\right)$.*

One has immediately $\Delta_{\overline{\omega}}\left(\boldsymbol{q} \,|\mathcal{D}_1; \boldsymbol{g}\right) \geq \Delta_{\overline{\omega}}\left(\boldsymbol{q} \,|\mathcal{D}_2; \boldsymbol{g}\right)$ for every $\mathcal{D}_1 \subset \mathcal{D}_2 \subset \mathcal{X}$ for $\forall \boldsymbol{q} \in \mathcal{X} \setminus \mathcal{D}_2$ which clearly shows the meaning of "diminishing returns". The immediate corollary of Prop. 5.3.5 for the goal-oriented quantization problem is that the maximum average social welfare function $\overline{\Omega}(\mathcal{D})$ is also submodular (NEs are refined in argmax set of potential function) since Prop. 5.3.5 holds for all possible parameter $\boldsymbol{g}$ and $\overline{\Omega}(\mathcal{D})$ is the expectation of function ω for parameter $\boldsymbol{g}$. However, even we have proven that function $\overline{\Omega}\left(\mathcal{D}\right)$ is submodular, it is still cumbersome and impracticable to find optimal action set with given cardinality since that requires the examination over all finite subsets with same cardinality of the action space $\mathcal{X}$. Optimizing monotone submodular set function be NP-hard in the worst case [117]. If the BP does exist for some parameters $\boldsymbol{g}$, it could be possible for us to deal with a non-monotone submodular optimization problem for $\overline{\Omega}\left(\mathcal{D}\right)$.

5.3.3 Algorithm for finding optimal action set

Knowing that function $\overline{\omega}\left(\mathcal{D}; \boldsymbol{g}\right)$ and $\overline{\Omega}\left(\mathcal{D}\right)$ are both submodular (could be non-monotone if BP exists as expected), one still needs to find efficient method to find such action set. The discrete derivative for function $\Omega\left(\mathcal{D}\right)$ is

$$\Delta_{\overline{\Omega}}\left(\boldsymbol{q} \,|\mathcal{D}\right) \triangleq \overline{\Omega}\left(\mathcal{D} \cup \{\boldsymbol{q}\}\right) - \overline{\Omega}\left(\mathcal{D}\right) \tag{5.15}$$

If discrete derivative $\Delta_{\overline{\Omega}}\left(\boldsymbol{q} \,|\mathcal{D}\right)$ is non-negative for any $\mathcal{D} \subset \mathcal{X}$ and $\boldsymbol{q} \in \mathcal{X}$, then function $\overline{\Omega}\left(\mathcal{D}\right)$ is a monotone set function. In this trivial scenario, replacing the action space $\mathcal{X}$ by any finite action set will reduce the overall performance surely. Therefore it

is reasonable for us to consider only the non-monotone case. Based on this property, we have proposed an algorithm summarized in alg. 6. The first part of algorithm is actually a greedy algorithm to get a initial action set of cardinality M. Then we continue this greedy update until we get an action set $\mathcal{D}^{'}$ such that $\mathcal{D} \subseteq \mathcal{D}^{'}$. Finally one choose the optimal subset of $\mathcal{D}_N$ as the final state of current iteration. The second step are executed repeatedly until convergence or the discrete derivative is always negative for current action set. Notice that one could use $\Delta_{\overline{\Omega}}\left(\boldsymbol{q}\left|\mathcal{D}_{m-1}^{(0)}\right.\right)$ instead of $\Delta_{\Omega}\left(\boldsymbol{q}\left|\mathcal{D}_{m-1}^{(0)}\right.\right)$. The replacement of discrete derivative could lead to a better solution while the complexity of finding all NEs for a game could be higher.

Initialization : Choose largest cardinality of action set N ; set $\mathcal{D}_0^{(0)} = \varnothing$; choose the tolerance error ε
for $m = 1$ *to* M **do**
  Let $\boldsymbol{q}_m \in \underset{\boldsymbol{q}\in\mathcal{X}\setminus\mathcal{D}_{m-1}^{(0)}}{\operatorname{argmax}} \Delta_{\overline{\Omega}}\left(\boldsymbol{q}\left|\mathcal{D}_{m-1}^{(0)}\right.\right)$;
  Let $\mathcal{D}_m^{(0)} \to \mathcal{D}_{m-1}^{(0)} \bigcup \{\boldsymbol{q}_m\}$;
end
for $t = 1$ *to* T **do**
  for $m = M+1$ *to* N **do**
    Let $\boldsymbol{q}_m \in \underset{\boldsymbol{q}\in\mathcal{X}\setminus\mathcal{D}_{m-1}^{(t)}}{\operatorname{argmax}} \Delta_{\overline{\Omega}}\left(\boldsymbol{q}\left|\mathcal{D}_{m-1}^{(t)}\right.\right)$;
    if $\Delta_{\overline{\Omega}}\left(\boldsymbol{q}\left|\mathcal{D}_{m-1}^{(t)}\right.\right) \geq 0$ **then**
      Let $\mathcal{D}_m^{(t)} \leftarrow \mathcal{D}_{m-1}^{(t)} \bigcup \{\boldsymbol{q}_m\}$;
    else
      Let $\mathcal{D}_m^{(t)} \leftarrow \mathcal{D}_{m-1}^{(t)}$;
      Break ;
    end
  end
  Let $\mathcal{D}_M^{(t+1)} \in \underset{\mathcal{D}\subseteq\mathcal{D}_m^{(t)},|\mathcal{D}|=M}{\operatorname{argmax}} \overline{\Omega}(\mathcal{D})$;
  if $\overline{\Omega}\left(\mathcal{D}_M^{(t+1)}\right) - \overline{\Omega}\left(\mathcal{D}_M^{(t)}\right) < \varepsilon$ **then**
    Break ;
  end
end
Output: Required action set is $\mathcal{D}_M^{(t+1)}$;

Algorithm 6: Algorithm for finding optimal action set maximizing average social welfare (with NE refinement)

5.4 Applications in Multiple Access Channel

5.4.1 System model and known results

Consider a parallel multiple access channel (MAC) with K users and S bands where the received signal vector $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_S)$ being written as :

$$\boldsymbol{y} = \sum_{k=1}^{K} \mathbf{H}_k \boldsymbol{x}_k + \boldsymbol{n} \tag{5.16}$$

Here, $\forall k \in \mathcal{K}$, $\mathbf{H}_k$ is the channel transfer matrix from k-th transmitter to the receiver, $\boldsymbol{x}_k$ is the vector of symbols transmitted by k-th transmitter, and vector $\boldsymbol{n}$ represents the noise observed at the receiver. In this scenario, we assume that $\mathbf{P}_k$ is a diagonal matrix with $\mathbf{H}_k = \operatorname{diag}(h_{k,1}, \ldots, h_{k,S})$ and blockin fading channel. Therefore, each entry $h_{k,s}$ for $\forall (k,s) \in \mathcal{K} \times \mathcal{S}$ are time-invariant realizations of a complex circularly symmetric Gaussian random variable with zero mean and unit variance. We denote the covariance matrix of transmitted symbol $\boldsymbol{x}_k$ as $\mathbf{P}_k = \operatorname{diag}(p_{k,1}, \ldots, p_{k,S})$. For notation convention we use the vector $\boldsymbol{p}_k = (p_{k,1}, \ldots, p_{k,S})$ to represent the action set of k-th user. The action set of k-th user is

$$\mathcal{X} = \mathcal{P}_k = \left\{ \boldsymbol{p}_k \,\middle|\, \sum_{s=1}^{S} p_{k,s} \leq P_{\max},\ p_{k,s} \geq 0 \right\} \tag{5.17}$$

Finally, $\boldsymbol{n}$ is the observed noise at receiver side which is circularly symmetric Gaussian distributed according to $\mathcal{CN}(0, \sigma^2 \mathbf{I}_S)$. The utility function of k-th user is the spectral efficiency :

$$f_k^{\mathrm{SE}}\left(\boldsymbol{p}_k, \boldsymbol{p}_{-k}; \mathbf{G}\right) = \sum_{s=1}^{S} \log\left(\frac{\sigma^2 + \sum_{k=1}^{K} p_{k,s} g_{k,s}}{\sigma^2 + \sum_{j \neq k}^{K} p_{j,s} g_{j,s}}\right), \tag{5.18}$$

where $g_{k,s} \triangleq |h_{k,s}|^2$. Parameters are jointly denoted by matrix $\mathbf{G} = (g_{k,s})_{k,s}$. To this end, this game could be formulated as $\mathbb{G}^{\mathrm{MAC}} = \left(\mathcal{K}, (\mathcal{X})_{k \in \mathcal{K}}, \left(f_k^{\mathrm{SE}}\right)_{k \in \mathcal{K}}\right)$. We first show that $\mathbb{G}^{\mathrm{MAC}}$ is really a potential game. One can easily find that utility function can be decomposed into :

$$f_k^{\mathrm{SE}}\left(\boldsymbol{p}_k, \boldsymbol{p}_{-k}; \mathbf{G}\right) = \varphi\left(\boldsymbol{p}; \mathbf{G}\right) - \nu_k\left(\boldsymbol{p}_{-k}; \mathbf{G}\right) \tag{5.19}$$

where

$$\varphi\left(\boldsymbol{p}; \mathbf{G}\right) = \sum_{s=1}^{S} \log\left(\sigma^2 + \sum_{k=1}^{K} p_{k,s} g_{k,s}\right) \tag{5.20}$$

and

$$\nu_k\left(\boldsymbol{p}_{-k}; \mathbf{G}\right) = \sum_{s=1}^{S} \log\left(\sigma^2 + \sum_{j \neq k}^{K} p_{j,s} g_{j,s}\right) \tag{5.21}$$

Obviously function φ is the potential function of game $\mathbb{G}^{\mathrm{MAC}}$. In [104], it is proven that $\mathbb{G}^{\mathrm{MAC}}$ has an unique NE with probability one. Moreover, this NE is given by the famous water-filling solution $\boldsymbol{p}^{\dagger}$:

$$p_{k,s}^{\dagger} = \left[\frac{1}{\lambda_k} - \frac{\sigma^2 + \sum_{j \neq k}^{K} p_{j,s} g_{j,s}}{g_{k,s}}\right]_{+} \tag{5.22}$$

with λ_k the Lagrange multiplier s.t.

$$\sum_{s=1}^{S} p_{k,s}^{\dagger} - P_{\max} = 0, \ \forall k \in \mathcal{K}. \tag{5.23}$$

If we restrict the action space of users to a finite action set $\mathcal{D}$ with $|\mathcal{D}| < \infty$, we obtain the game $\mathbb{G}^{\mathrm{FA}} = \left(\mathcal{K}, (\mathcal{D})_{k\in\mathcal{K}}, \left(f_k^{\mathrm{SE}}\right)_{k\in\mathcal{K}}\right)$. Obviously this game $\mathbb{G}^{\mathrm{FA}}$ is still a potential game. Besides, we present some results about a special case of this game : channel selection game. Define the channel selection set $\mathcal{D}_{\mathrm{CS}} = \{P_{\max}\boldsymbol{e}_s : \forall 1 \leq s \leq S\}$. Vector $\boldsymbol{e}_s$ is defined as $\boldsymbol{e}_s = (\boldsymbol{e}_{s,1}, \ldots, \boldsymbol{e}_{s,S})$ with $\boldsymbol{e}_{r,s} = 0$ for $r \neq s$ and $\boldsymbol{e}_{s,s} = 1$. In [90], this channel selection game $\mathbb{G}^{\mathrm{CS}} = \left(\mathcal{K}, (\mathcal{D}_{\mathrm{CS}})_{k\in\mathcal{K}}, \left(f_k^{\mathrm{SE}}\right)_{k\in\mathcal{K}}\right)$ studied and it is proven that $\mathbb{G}^{\mathrm{CS}}$ could have multiple NEs :

Proposition 5.4.1. *[Perlaza, 2013] Let $\widehat{K} \in \mathbb{N}$ be the highest even number which is less or equal to K. Then $\mathbb{G}^{CS}$ has L Nash equilibria :*

$$1 \leq L \leq 1 + (S-1) \sum_{i\in 2,4,\ldots,\widehat{K}} \binom{K}{i} \tag{5.24}$$

Prop 5.4.1 indicates that there could be multiple NEs for the game $\mathbb{G}^{\mathrm{FA}}$ as well. The existence of multiple NEs will not introduce extra difficulty for our problem since the main concern of the goal-oriented quantizer is the action set. To start with, we consider the simplest case of $\mathbb{G}^{\mathrm{FA}}$.

5.4.2 Case study for 2-user 2-band scenarios

We start our study on game $\mathbb{G}^{\mathrm{FA}}$ by the simplest case where there are only two bands and two users in the system. Since there are only $S = 2$ bands, each action $\boldsymbol{d}_m \in \mathcal{D}$ can be written as

$$\begin{aligned} \boldsymbol{d}_m &= \alpha_m \left[P_{\max}, 0\right]^T + (1-\alpha_m)\left[P_{\max}, 0\right]^T \\ &= \left[\alpha_m P_{\max}, (1-\alpha_m) P_{\max}\right]^T \end{aligned} \tag{5.25}$$

where $\alpha_m \in [0,1]$ can represent action $\boldsymbol{d}_m$. Therefore, the action set $\mathcal{D}$ can be represented by a sequence $\{\alpha_m\}_{m=1}^{M}$. Without loss of generality, we can assume that $0 \leq \alpha_1 < \cdots < \alpha_m < \cdots < \alpha_M \leq 1$. To find the optimal finite action set with cardinality M, the first step is to find the equivalent condition under which an action profile is NE. Without loss of generality, assume that one NE of game is denoted as $\boldsymbol{p}^{\star} = (\boldsymbol{d}_i, \boldsymbol{d}_j)$ or equivalently (α_i, α_j). For the simplicity of notation, we introduce for a multi-index $\boldsymbol{i} \triangleq (i_1, \ldots, i_k, \ldots, i_K)$ with $i_k \in \{1, \ldots, M\}$ and define the the potential function evaluated at the action profile $(\boldsymbol{d}_{i_1}, \ldots, \boldsymbol{d}_{i_k}, \ldots, \boldsymbol{d}_{i_K})$ as :

$$\varphi_{\boldsymbol{i}} \triangleq \varphi\left(\boldsymbol{d}_{i_1}, \ldots, \boldsymbol{d}_{i_k}, \ldots, \boldsymbol{d}_{i_K}\right) \tag{5.26}$$

In 2-user 2-band setting, we have that $\varphi_{i,j} = \psi\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$ is an NE if and only if :

$$\begin{aligned} &\varphi_{i,j} \geq \varphi_{t,j} \text{ for } \ \forall t \in \{1, \ldots, M\} \\ &\varphi_{i,j} \geq \varphi_{i,t} \text{ for } \ \forall t \in \{1, \ldots, M\} \end{aligned} \tag{5.27}$$

And we have

$$\begin{aligned}\varphi_{i,j} &= \varphi\left(\boldsymbol{d}_i, \boldsymbol{d}_j\right) \\ &= \log\left(\sigma^2 + g_{1,1}\alpha_i P_{\max} + g_{2,1}\alpha_j P_{\max}\right) + \log\left(\sigma^2 + g_{1,2}\left(1-\alpha_i\right)P_{\max} + g_{2,2}\left(1-\alpha_j\right)P_{\max}\right) \\ &= \log\left[\left(\sigma^2 + g_{1,1}\alpha_i P_{\max} + g_{2,1}\alpha_j P_{\max}\right)\left(\sigma^2 + g_{1,2}\left(1-\alpha_i\right)P_{\max} + g_{2,2}\left(1-\alpha_j\right)P_{\max}\right)\right]\end{aligned} \tag{5.28}$$

Obviously, the potential function is the composition of the logarithmic function and a quadratic function of action profile. Consequently, conditions in (5.27) are equivalent to following conditions :

$$\begin{aligned}&\varphi_{i,j} \geq \varphi_{i-1,j},\ \varphi_{i,j} \geq \varphi_{i+1,j} \\ &\varphi_{i,j} \geq \psi_{i,j-1},\ \varphi_{i,j} \geq \varphi_{i,j+1}\end{aligned} \tag{5.29}$$

To this end, to find the optimal action set maximizing the social welfare, it is sufficient to solve the following OP :

$$\max_{\{\alpha_m\}_{m=1}^M} \frac{\left(\frac{1}{\gamma} + g_{1,1}\alpha_i + g_{2,1}\alpha_j\right)^2\left(\frac{1}{\gamma} + g_{1,2}\left(1-\alpha_i\right) + g_{2,2}\left(1-\alpha_j\right)\right)^2}{\left(\frac{1}{\gamma} + g_{1,1}\alpha_i\right)\left(\frac{1}{\gamma} + g_{2,1}\alpha_j\right)\left(\frac{1}{\gamma} + g_{1,2}\left(1-\alpha_i\right)\right)\left(\frac{1}{\gamma} + g_{2,2}\left(1-\alpha_j\right)\right)}$$

$$\text{s.t. } \frac{g_{1,2}-g_{1,1}}{\gamma} - g_{1,1}g_{1,2} - g_{1,1}g_{2,2} + g_{1,1}g_{1,2}\left(\alpha_i + \alpha_{i+1}\right) + \left(g_{2,1}g_{1,2} + g_{1,1}g_{1,2}\right)\alpha_j \leq 0 \tag{5.30}$$

$$\frac{g_{1,2}-g_{1,1}}{\gamma} - g_{1,1}g_{1,2} - g_{1,1}g_{2,2} + g_{1,1}g_{1,2}\left(\alpha_i + \alpha_{i-1}\right) + \left(g_{2,1}g_{1,2} + g_{1,1}g_{1,2}\right)\alpha_j \geq 0 \tag{5.31}$$

$$\frac{g_{2,2}-g_{2,1}}{\gamma} - g_{2,1}g_{1,2} - g_{2,1}g_{2,2} + \left(g_{1,1}g_{2,2} + g_{2,1}g_{1,2}\right)\alpha_i + g_{2,1}g_{2,2}\left(\alpha_j + \alpha_{j+1}\right) \leq 0 \tag{5.32}$$

$$\frac{g_{2,2}-g_{2,1}}{\gamma} - g_{2,1}g_{1,2} - g_{2,1}g_{2,2} + \left(g_{1,1}g_{2,2} + g_{2,1}g_{1,2}\right)\alpha_i + g_{2,1}g_{2,2}\left(\alpha_j + \alpha_{j-1}\right) \geq 0 \tag{5.33}$$

where $\gamma = \frac{P_{\max}}{\sigma^2}$ is signal-noise ratio (SNR). Notice that if $M = 2$ and $\mathcal{D} = \{(P_{\max}, 0), (0, P_{\max})\}$ then one obtains the channel selection set. Surprisingly, we have the following proposition :

Proposition 5.4.2. *For 2-user 2-band game $\mathbb{G}^{FA}$, optimal finite action set for maximizing the social welfare for any static channel is the channel selection set.*

Proof : See Appendix D . ■

For 2-user 2-band scenario, we have proven that the channel selection set is the optimal finite action set. If the number of decision M is less or equal to the number of bands S, one can easily find a subset of channel selection set. Nevertheless, if the actual cardinality of optimal action is greater than S, the complexity of finding an acceptable action set could be huge, at least, each evaluation of average social welfare will be cumbersome. Therefore, we would like to find a sequence of finite action set under acceptable complexity. Define

$\mathcal{S}_l = \left\{ \widehat{\boldsymbol{e}} \in \left\{0, \frac{P_{\max}}{l}\right\}^S \middle| \sum_{i=1}^S \widehat{\boldsymbol{e}}_i = P_{\max} \right\}$ which is the set of S-dimensional vector summing to $P_{\max}$ and $\mathcal{S} = \bigcup_{l=1}^S \mathcal{S}_l$. The set $\mathcal{S}$ is usually referred as Telatar set which is conjectured to optimal power allocation policy in [93]. The subset of Telatar set is referred as Telatar-type set in the following of this chapter. Denote the optimal solution of OP in 5.6 for game $\mathrm{G}^{\mathrm{FA}} = \left(\mathcal{K}, (\mathcal{D})_{k\in\mathcal{K}}, \left(f_k^{\mathrm{SE}}\right)_{k\in\mathcal{K}}\right)$ as $\mathcal{D}_M^{\mathrm{SE}} \in \arg\max_{\mathcal{D}:|\mathcal{D}|=M} \Omega(\mathcal{D})$, we have the following conjecture :

Conjecture 5.4.3. *For* $\forall\ 1 \leq M \leq 2^S - 1$*, the optimal finite action set* $\mathcal{D}_M^{SE}$ *maximizing the average social welfare with cardinality* M *is a Telatar-type set, i.e.,* $\mathcal{D}_M^{SE} \subset \mathcal{S}$*.*

Conjecture 5.4.3 is extremely useful if it is true. It basically says that the Telatar-type set could maximize the average social welfare for spectral efficiency of MAC if the cardinality of desired finite action set is less than $2^S - 1$. This condition is generally true if number of bands is legitimately large. However, it seems directly find such Telatar-type set is still costly of complexity $O\left(C_{|\mathcal{S}|}^M\right)$. Generally, we should have $|\mathcal{S}| = 2^S - 1 \gg M$. Therefore, the determination of optimal Telater-type set is of exponential complexity by Stirling's approximation $O\left(C_{|\mathcal{S}|}^M\right) = O\left(|\mathcal{S}|^M\right)$ when the number of bands is relatively large. However, if one apply alg. 6 and choose the Talatar Set $\mathcal{S}$ as underlying action space. For instance, set $N = M + 1$ and choose maximum iteration of alg. 6 as T_e, the complexity of alg. 6 could be reduced to polynomial time $O(T_e |\mathcal{S}|)$ if the conjecture 5.4.3 is correct.

5.4.3 Numerical results

For MAC, we will first show that the optimal action set is non-trivial for static channel. In Fig. 5.1, the average social welfare of iterative water-filling algorithm (IWTA), alg. 6 with Telatar set as the underlying action space, and alg. 6 v.s. the number of actions are depicted for 2-user and 3-user scenarios. We assume that there are $S = 4$ bands and the power budget for user is $P_{\max} = 1\mathrm{mW}$ with noise level $\sigma^2 = 1\mathrm{mW}$. All results are averaged over 1000 randomly generated channel matrix. The reference of optimality loss is the centralized policy which maximizes the scocial welfare. In the rest of section, the relative optimality loss will be calculated in the same way. For 2-user scenario, optimality loss introduced by alg. 6 and alg. 6 with Telatar set as the underlying action space is nearly none and almost independent of the number of decisions. Relative optimality introduced by alg. 6 outperforms IWFA for almost all possible configurations of the system which strongly confirms the existence of Braess paradox. For 3-user scenario, when $M \leq S = 4$, OL introduced by two proposed methods decrease rapidly as the number of decisions grows ; when $M > S = 4$, the decay is not obvious. However, the BP persists for $M \geq K = 3$. These results of could be regarded as the single-sample case of our general conjecture 5.4.3.

Fig. 5.2 shows the relative optimality loss (%) v.s. number of actions for IWFA, alg. 6 with Telatar set and alg. 6 with general action space for 2-user and 3-user scenarios. Number of bands is chosen as $S = 3$. In fig. 5.2, apparently, simulation results confirm our conjecture 5.4.3 again. When the number of actions is less or equal to the number of

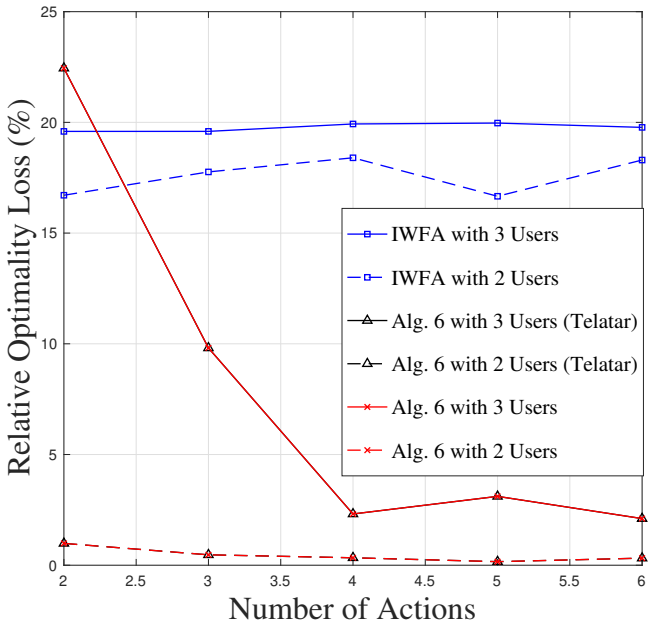


FIGURE 5.1 – Relative optimality loss (%) of social welfare function as a function of number of actions for iterative water-filling algorithm(IWFA), alg. 6 with Telatar set as the underlying action space, and alg. 6for 2-user and 3-user scenarios with $P_{\max} = 1\text{mW}$ and $\sigma^2 = 1\text{mW}$ and number of bands $S = 4$. Braess's Paradox exists in almost all configurations. Optimal action set with given cardinality is always the Telatar-type set. This result verifies the single-sample case of our conjecture.

users (K), there is a clear decay of OL as number of actions M grows. However, if $M > S$, the decay of OL is slow. Compared to the conventional approach, proposed algorithms largely reduce the optimality loss tremendously showing the potential of the goal-oriented quantization.

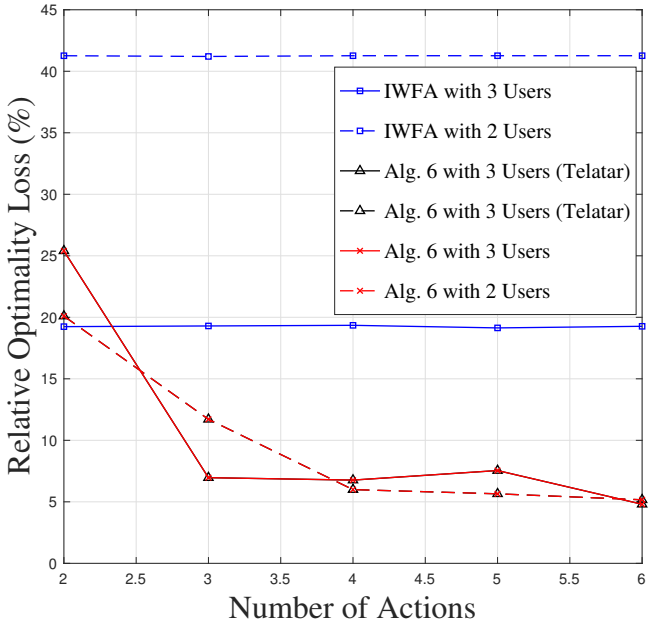


FIGURE 5.2 – Relative optimality loss (%) of average social welfare as a function of number of actions for iterative water-filling algorithm(IWFA), alg. 6 with Telatar set as the underlying action space, and alg. 6 for 2-user and 3-user scenarios with $P_{\max} = 1\text{mW}$ and $\sigma^2 = 1\text{mW}$ and number of bands $S = 3$. Braess's Paradox exists in almost all configurations except for $M = 2$ decisions. Optimal action set maximizing the average social welfare with given cardinality is always the Telatar-type set. Our conjecture is verified for number of actions in $2 \leq M \leq 6$.

In Fig. 5.3, the relative optimality loss v.s. number of bands for fixed $M = 4$ actions are illustrated. relative optimality loss increases as the number of bands grows in tendency. This observation shows that finite action set with more actions should be extended when the dimension of the sytsem grows. Meanwhile, for all methods, the performance of the system worsens if more users enter into the MAC system.

Then, we would like to study the influence of power budget for MAC. The relative optimality loss v.s. the power budget for different methods for 2-user 4-band with 4 actions are illustrated in Fig. 5.4 respectively. The impact of power budget is not obvious for our proposed alg. 6 while the decay of optimality loss introduced by IWFA is obvious when the power budget is relatively small ($P_{\max} \leq 5\text{mW}$). This observation entails that proposed algorithm is still efficient compared to IWFA when the power budget of the device is small which could be useful for scenario such as IoT where energy efficiency of the system is of great importance.

Finally, Fig. 5.5 shows the performance of different approaches as function of number of users in the system. For IWFA, the performance worsen as the number of users increases. For 5-users 3-decisions scenario alg. 6 provides a reduction of relative optimality loss

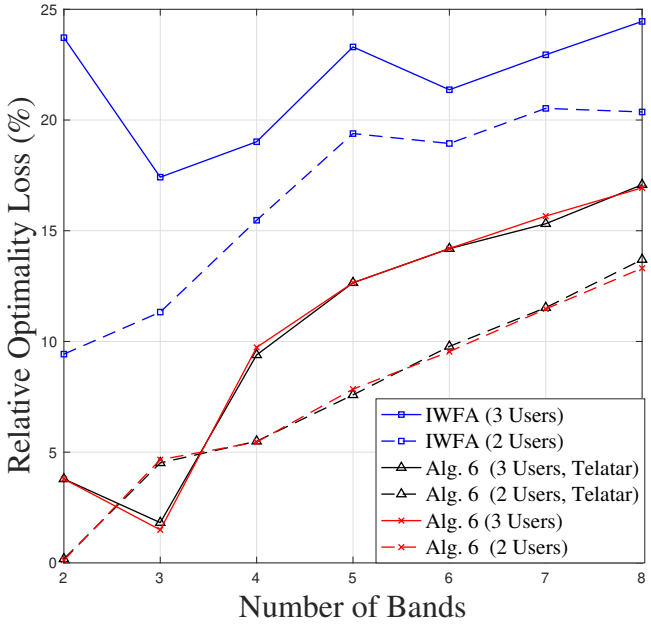


FIGURE 5.3 – Relative optimality loss (%) of average social welfare as a function of number of bands for iterative water-filling algorithm(IWFA), alg. 6 with Telatar set as the underlying action space, and alg. 6 for 2-user and 3-user scenarios with $P_{\max} = 1\text{mW}$ and $\sigma^2 = 1\text{mW}$ and number of actions is fixed as $M = 4$. Braess's Paradox always exists and optimality loss grows for all approaches as the number of bands increases.

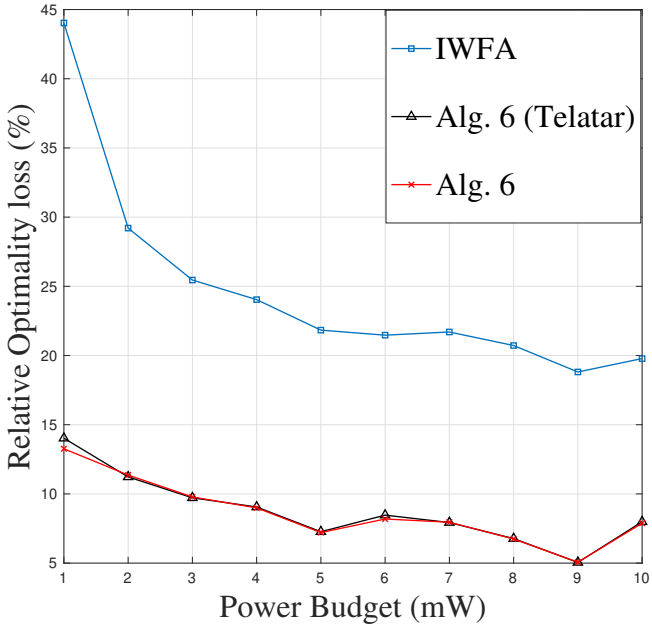


FIGURE 5.4 – Relative optimality loss as a function of number of bands for iterative water-filling algorithm(IWFA), alg. 6 with Telatar set as the underlying action space, and alg. 6 in with $M = 4$ actions. Optimality loss introduced by proposed algorithm is always much smaller than IWFA. When number of actions is fixed ,as the dimesion of system grows, the optimality loss increases as well.

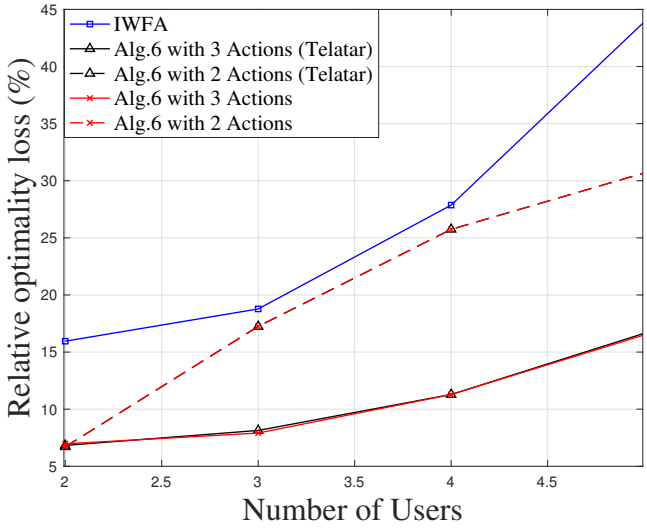


FIGURE 5.5 – Relative optimality loss as a function of number of users for iterative water-filling algorithm(IWFA), alg. 6 with Telatar set as the underlying action space, and alg. 6 for 4-band case with $M = 4$ actions. Optimality loss introduced by proposed algorithm is always largely smaller than IWFA. Increasing power budget has slight impact on the reduction of optimality loss for our methods.

around 30% compared to IWFA revealing the advantage of our framework for potential game.

5.5 Conclusions

In this chapter, instead of focusing on the optimal structure of the goal-oriented communication for single transmitter-receiver pair, we restrict ourselves in the study of potential games with identical action set where the utility function of players could be completely different. The existence of the famous Braess's Paradox makes the goal-oriented quantization problem extremely interesting in this scenario. Taking the social welfare as our performance criterion, we have proven that the maximum average social welfare is a submodular function of the action set if the Nash equilibrium is refined as the maximum of potential function of the game. Based on this property, we design an algorithm to find a goal-oriented quanizer aiming at maximizing average social welfare of the system. Our analysis is applied to a multiple access channel game game with spectral efficiency being the individual utility. Our alysis is applied to a multi-user MAC game with spectral efficiency as the individual utility. For 2-user 2-band scenario, we have proven that the channel selection set is the optimal action set. More generally, Telatar-type set is conjectured to be the optimal action set for maximizing the average social welfare. Besides, the existence of Braess's Paradox is not guaranteed for general utility function of arbitrary game. The feasibility of proposed method should be verified for other applications in the wireless communication system and smart grids as well.

6

Nash Equilibrium Analysis in Multi-User MIMO Energy Efficiency Game

In this chapter, we focus on a game where user's utility function is the energy efficiency in a MIMO multiple access channel system. The existence of Nash equilibrium is proven. The uniqueness of Nash equilibrium is confirmed by showing the standard property of MIMO-EE game. An algorithm by applying the approximate best response is proposed to approach the unique Nash equilibrium of the game. For 2-user 2-band scenario, our proposed algorithm surprisingly Pareto-dominates the pure Nash equilibrium of the game. Nash equilibrium analysis for this type of game could be served as basic results for our goal-oriented quantization framework involving in performance under equilibrium regime in the future.

6.1 Motivation

With the release of first 5G package, it turns out that the number of devices in the upcoming wireless network will increase tremendously, e.g., Internet of Things (IoT). Consequently, classical paradigm which merely aims at optimizing the quantitative performance, e.g., data-rate, bit-error-rate and latency faces extreme difficulty in many domains in both academic research and industrial application. Thus the issue of energy-efficient design of the wireless system tends to be crucial. Different definition of energy efficiency has been proposed in recent years in [78, 79, 80, 81]. Amongst which the most popular one is defined as the total benefit obtained under the unit consumption of energy or power known as global energy efficency (GEE) e.g., in [69, 70, 71, 77]. Taken the bits-per-second type rate function as benefit function, one will obtain the well-known bits-per-joule energy efficiency.

One of the pioneer works of studying the maximization of EE in Multiple-Input Multiple-Output (MIMO) system is [71]. In [71], the optimal precoding scheme is stu-

died and divided into different cases with different assumptions on the systems. Till now the optimal precoding matrix for general condition is merely conjectured and unproved. Hereafter, optimal precoding matrix design for single user MIMO system is performed for imperfect channel state information (CSI) scenario in [75]. Then it is later widely realized that the problem of EE maximization actually belongs to the category of fractional programming. Techniques such as Dinkelbach's algorithm (see [74]) is used to solve EE maximization in [75, 76]. These algorithms are generally based on the idea that the optimal solution can be found by solving a sequence of convex optimization problems related to the original one. The main difficulty of EE maximization OP is usually due to the non-convexity of energy efficiency function. Under some assumption on the benefit function, the EE function is well-known as being quasi-concave or even pseudo-concave. However, it is generally difficult to trace the Nash equilibrium of a game where the individual utility function of player is of EE type. In [69], it is shown that there always exists an unique NE for scalar power allocation game in a relay-assisted MIMO systems due to the standard property of the best response dynamics. Similar results in MIMO-MAC system will be given latter in this chapter.

6.2 System Model

Consider a multiple access channel (MAC) with one base station (BS) and K users (players) to be served. BS is equipped with N_r receive antennas and each user terminal is equipped with N_t transmit antennas. We assume a block fading channel where the realization of channel remains a constant during the coherence time of transmission and randomly generated according to some statistical distributions from period to period. The received signal at BS is given by :

$$\boldsymbol{y} = \sum_{k=1}^{K} \mathbf{H}_k \boldsymbol{x}_k^{\mathrm{SIG}} + \boldsymbol{z}, \tag{6.1}$$

where $\mathbf{H}_k \triangleq [\mathbf{H}_{k,i,j}]_{i,j=1}^{N_r,N_t} \in \mathbb{C}^{N_r \times N_t}$ is the channel transmit matrix of k-th user and $\mathbf{H}_{k,i,j}$ is the channel from i-th transmit antenna of k-th user to j-th receive antenna at BS which is assumed to be i.i.d. complex Gaussian distributed according to $\mathcal{CN}(0,1)$. $\boldsymbol{x}_k^{\mathrm{SIG}} = \left(x_{k,1}^{\mathrm{SIG}}, \ldots, x_{k,N_t}\right)^{\mathrm{T}}$ is the transmit symbol of k-th user and $\boldsymbol{z}$ is the noise observed by the receiver with complex Gaussian distribution $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_{N_r})$. For the sake of simplicity, we assume that single user decoding is implemented for each user. Then the capacity the k-user can be achieved is

$$R_k = \log \frac{\det\left(\sigma^2 \mathbf{I}_{N_r} + \sum_{j=1}^{K} \mathbf{H}_j \mathbf{Q}_j \mathbf{H}_j^H\right)}{\det\left(\sigma^2 \mathbf{I}_{N_r} + \sum_{j \neq k}^{K} \mathbf{H}_j \mathbf{Q}_j \mathbf{H}_j^H\right)}, \tag{6.2}$$

where $\mathbf{Q}_k = \mathbb{E}\left[\boldsymbol{x}_k^{\mathrm{SIG}} \left(\boldsymbol{x}_k^{\mathrm{SIG}}\right)^{\mathrm{H}}\right] \in \mathbb{C}^{N_t \times N_t}$ is the covariance matrix of symbol $\boldsymbol{x}_k^{\mathrm{SIG}}$ which determines how power should be allocated for each antenna and $P_c > 0$ is the power dissipated in transmitter circuit to operate the devices. It is reasonable to assume that each user has perfect knowledge about its own channel, e.g., through downlink pilot training. Therefore user k is able to perform the singular value decomposition (SVD) of its own channel $\mathbf{H}_k$ and its covariance matrix $\mathbf{Q}_k$ as well. The SVD of $\mathbf{H}_k$ and $\mathbf{Q}_k$ is given by

$\mathbf{H}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{V}_k^H$ and $\mathbf{Q}_k = \mathbf{W}_k \mathbf{P}_k \mathbf{W}_k^H$ respectively. To simplify the problem, we assume that user k always adapts its covariance matrix to $\mathbf{H}_k$, i.e., choosing $\mathbf{W}_k = \mathbf{V}_k$. $\mathbf{P}_k$ is a diagonal matrix with $\mathbf{P}_k = \operatorname{diag}(\boldsymbol{p}_k) = \operatorname{diag}(p_{k1}, \ldots, p_{kN_t})$ where we use $\operatorname{diag}(\cdot)$ to generate a diagonal matrix from a vector or vice versa. Thus user k's only legal action is represented by $\boldsymbol{p}_k$ or $\mathbf{P}_k$ and the action set of k-th user is

$$\mathcal{P}_k = \left\{ \boldsymbol{p}_k \left| \sum_{i=1}^{N_t} p_{ki} \leq \overline{P}_k,\ p_{ki} \geq 0 \right. \right\} \tag{6.3}$$

where $\overline{P}_k$ is power budget of k-th user. Through out the chapter, we will use the matrix $\mathbf{P}_k$ or its diagonal $\boldsymbol{p}_k$ interchangeably to represent user k's action depending on the context. Further more, we denote $\boldsymbol{p} = (\boldsymbol{p}_k, \boldsymbol{p}_{-k})$ with $\boldsymbol{p}_{-k} \triangleq (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{k-1}, \boldsymbol{p}_{k+1} \ldots, \boldsymbol{p}_K) \in \mathcal{P}_{-k}$ and $\mathcal{P}_{-k} \triangleq \mathcal{P}_1 \times \cdots \times \mathcal{P}_{k-1} \times \mathcal{P}_{k+1} \times \cdots \times \mathcal{P}_K$. In this chapter, energy efficiency is defined as the ratio of a benefit function over the power consumed by producing it can be proven to has the following expression for user k after some simplifications :

$$f_k^{\mathrm{EE}}(\mathbf{P}_k, \mathbf{P}_{-k}) = \frac{\log \frac{\det\left(\sigma^2 \mathbf{I}_{N_r} + \sum_{j=1}^{K} \mathbf{U}_j \mathbf{\Lambda}_j \mathbf{P}_j \mathbf{\Lambda}_j^H \mathbf{U}_j^H\right)}{\det\left(\sigma^2 \mathbf{I}_{N_r} + \sum_{j \neq k}^{K} \mathbf{U}_j \mathbf{\Lambda}_j \mathbf{P}_j \mathbf{\Lambda}_j^H \mathbf{U}_j^H\right)}}{\operatorname{Tr}(\mathbf{P}_k) + P_c} \tag{6.4}$$

To this end, the MIMO MAC EE game is thus given by the following strategic form in triplet :

$$\mathbb{G}^{\mathrm{EE}} = \left(\mathcal{K}, (\mathcal{P}_k)_{k \in \mathcal{K}}, \left(f_k^{\mathrm{EE}}\right)_{k \in \mathcal{K}} \right) \tag{6.5}$$

6.3 Game-Theoretic Analysis

To identify the NE of game in (6.5), the properties of individual utility function should be identified as first step. We define two critical properties satisfied by the individual utility function.

Definition 6.3.1. *(Quasi-concavity) Let $\mathcal{X} \in \mathbb{R}^n$ be a convex set, a function $f : \mathcal{X} \to \mathbb{R}$ is said to be quasi-concave if*

$$u(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \geq \min\{u(\boldsymbol{x}), u(\boldsymbol{y})\} \tag{6.6}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ with $\boldsymbol{x} \neq \boldsymbol{y}$ and $0 < \lambda < 1$.

Definition 6.3.2. *(Pseudo-concavity) Let $\mathcal{X} \in \mathbb{R}^n$ be a convex set, a function $f : \mathcal{X} \to \mathbb{R}$ is said to be quasi-concave if it is differentiable and for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, it holds :*

$$u(\boldsymbol{y}) < u(\boldsymbol{x}) \implies \nabla u(\boldsymbol{y})^T (\boldsymbol{x} - \boldsymbol{y}) > 0 \tag{6.7}$$

With the definition of quasi-concavity and the pseudo-concavity, Prop. 6.3.3 shows that the individual utility function does possess these important properties :

Proposition 6.3.3. *R_k is a concave functions w.r.t. $\boldsymbol{p}_k$ and u_k is a pseudo-concave (quasi-concave) function w.r.t. $\boldsymbol{p}_k$ for $\forall k \in \mathcal{K}$; For any fixed $\boldsymbol{p}_{-k} \in \mathcal{P}_{-k}$ and p_{kj} with $j \neq i$, only one of following statements is true for all $i \in [N_t]$:*

i) $\exists\ p_{ki}^\star > 0$ s.t. f_k^{EE} is an increasing function in $(0, p_{ki}^\star)$ and a decreasing function in $(p_{ki}^\star, +\infty)$ w.r.t. p_{ki}.

ii) f_k^{EE} is a decreasing function in $(0, +\infty)$ w.r.t. p_{ki}.

Proof : It is well-known that R_k is a concave function for $\boldsymbol{p}_k$. Then the pseudo-concavity (quasi-concavity) of f_k^{EE} comes from the fact that it is a ratio of a concave function and an affine function of $\boldsymbol{p}_k$. For more details of the proof, see [59]. Now we prove the second part of this proposition. Rewrite the individual utility function as $f_k^{\mathrm{EE}} = \frac{R_k(\gamma_k)}{\sum_{i=1}^{N_t} p_{ki} + P_c}$ with $R_k(\gamma_k) = \log(1+\gamma_k)$. Then we can prove that $\frac{\partial^2 f_k^{\mathrm{EE}}}{\partial p_{ki}^2} \leq 0$ due to the fact that R_k is an increasing concave function w.r.t. γ_k and γ_k is a also increasing concave function w.r.t. p_{ki}. However we can't conclude directly of the sign of $\lim_{p_{ki}\to+\infty} \frac{\partial^{\mathrm{EE}}}{\partial p_{ki}}$. It can be positive or negative depending on the value of p_{kj} with $j \neq i$. Therefore, if $\lim_{p_{ki}\to+\infty} \frac{\partial f_k^{\mathrm{EE}}}{\partial p_{ki}} \geq 0$ then we are in case ii), otherwise we are in case i). ■

Definition 6.3.4. *(Standard Games) A game* $\mathcal{G} = \left(\mathcal{K}, (\mathcal{P}_k)_{k\in\mathcal{K}}, (f_k)_{k\in\mathcal{K}}\right)$ *is said to be standard if its best response is always standard, i.e.,*

1) Positivity : $\forall\ \mathbf{P}_{-k} \succcurlyeq 0$, $BR_k(\mathbf{P}_{-k}) \succcurlyeq 0$;

2) Monotonicity : if $\mathbf{P}'_{-k} \succcurlyeq \mathbf{P}_{-k}$, *then* $BR_k(\mathbf{P}'_{-k}) \succcurlyeq BR_k(\mathbf{P}_{-k})$;

3) Scalability : $BR_k(\alpha\mathbf{P}_{-k}) \prec \alpha BR_k(\mathbf{P}_{-k})$ *for any* $\alpha > 1$.

Before stating the best response dynamics of the game, we define the following boundary of set $\mathcal{P}_k$ indicated by an index subset $\mathcal{E} \subset [N_t]$:

$$\mathcal{P}_k[\mathcal{E}] \triangleq \{\boldsymbol{p}_k \in \mathcal{P}_k, p_{ki} = 0 \text{ for } i \in \mathcal{E}\} \tag{6.8}$$

and the non-negative index set for a given action $\mathbf{P}_k$:

$$\mathcal{I}(\mathbf{P}_k) \triangleq \{i \in [N_t]\ s.t.\ p_{ki} \geq 0\} \tag{6.9}$$

Proposition 6.3.5. *For any given* $\mathbf{P}_{-k}$ *and provided that the power budget* $\overline{P}_k$ *is sufficiently large, denote the unique solution of the following equation as* $\mathbf{P}_k^*$:

$$diag\left(\mathbf{\Lambda}_k^H \left(\mathbf{\Lambda}_k \mathbf{P}_k \mathbf{\Lambda}_k^H + \mathbf{F}_k + \sigma^2 \mathbf{I}_r\right)^{-1} \mathbf{\Lambda}_k\right) = f_k^{EE}(\mathbf{P}_k, \mathbf{P}_{-k})\, \mathbf{I}_{N_t} \tag{6.10}$$

with $\mathbf{F}_k = \sum_{j\neq k} \mathbf{S}_j \mathbf{P}_j \mathbf{S}_j^H$ *is the interference matrix of k-th user with* $\mathbf{S}_j = \mathbf{U}_k^H \mathbf{U}_j \mathbf{\Lambda}_j$. *Then the BR of* $\mathbf{P}_k$ *w.r.t.* $\mathbf{P}_{-k}$ *is standard and converges to the unique NE admitted by game* $\mathbb{G}^{EE}$ *; The BR is the unique solution of (6.10) restricted to the boundary of* $\mathcal{P}_k$ *indicated by* $\mathcal{I}(\mathbf{P}_k^*)$ *with* $\mathcal{I}(\mathbf{P}_k^*) \neq \emptyset$.

Proof : our proof consists of two parts : i) existence of NE ; ii) uniqueness of NE. For more details of the proof. More details could be found in Appendix E. ■

6.4 Algorithms for Finding NE

Prop. 6.3.5 actually provides an approach for us to find the NE of the game $\mathbb{G}^{\mathrm{EE}}$. One can easily deduce an iterative equation according to (6.10) :

$$\mathrm{diag}\left(\mathbf{\Lambda}_k^H \left(\mathbf{\Lambda}_k \mathbf{P}_k^{(t)} \mathbf{\Lambda}_k^H + \mathbf{F}_k^{(t-1)} + \sigma^2 \mathbf{I}_r\right)^{-1} \mathbf{\Lambda}_k\right) = f_k^{\mathrm{EE}}\left(\mathbf{P}_k^{(t-1)}, \mathbf{P}_{-k}^{(t-1)}\right) \mathbf{I}_{N_t} \tag{6.11}$$

However, due to Prop. 6.3.5, this stationary point might not be in the feasible action set. One can design the following basic algorithm to find NE of the game based on Prop. 6.3.5

Initialization : Number of decisions M ; set $\mathbf{P}_k^{(0)} = \frac{1}{N_t}\mathbf{I}_{N_t}, \forall k$ Choose T and ε ;
for $t = 1$ *to* T **do**
  for $k = 1$ *to* K **do**
    Compute $\mathbf{P}_k^{(t)}$ using (6.11) ;
    if $\mathcal{I}\left(\mathbf{P}_k^{(t)}\right) \neq [N_t]$ **then**
      Compute $\mathbf{P}_k^{(t)}$ using Eq. 6.11 restricted to $\mathcal{I}\left(\mathbf{P}_k^{(t)}\right)$;
    end
  end
  if $\sum_k \left\|\mathbf{P}_k^{(t)} - \mathbf{P}_k^{(t-1)}\right\| < \varepsilon$ **then**
    Break ;
  end
end
Output: $\mathbf{P}_k^{\mathrm{NE}} = \mathbf{P}_k^{(t)}$ for $\forall k$;

Algorithm 7: Basic algorithm for finding NE of MIMO-MAC EE game $\mathbb{G}^{\mathrm{EE}}$

Nevertheless, alg. 7 is not satisfatory way to find the NE of the game. More precisely, to find the BR for given $\mathbf{P}_{-k}$, one actually need to solve an optimization problem. However, if $h = U^\star(\mathbf{P}_{-k}) = \max_{\mathbf{P}_k \in \mathcal{P}_k} f_k^{\mathrm{EE}}(\mathbf{P}_k, \mathbf{P}_{-k})$ is known as *a priori* information, (6.11) can be transformed into following equation which is relatively easy to be solved compared to (6.11) :

$$\mathrm{diag}\left(\mathbf{\Lambda}_k^H \left(\mathbf{\Lambda}_k \mathbf{P}_k^{(t)} \mathbf{\Lambda}_k^H + \mathbf{F}_k^{(t-1)} + \sigma^2 \mathbf{I}_r\right)^{-1} \mathbf{\Lambda}_k\right) = h\mathbf{I}_{N_t} \tag{6.12}$$

Introducing an auxiliary parameter h, one obtains an iterative equation of $\mathbf{P}_k$. Without loss of generality, we assume that the solution of (6.11) belongs to the feasible action set for given $\mathbf{P}_{-k}$. Otherwise, similar analysis can applied for $\mathbf{P}_k$ but restricted on a boundary given by Prop. 6.3.5. For the sake of simplicity, we omit the discussion here and restrict ourselves to the situation where the BR is strictly included in the interior of the feasible action set. Therefore for all $i \in [N_t]$, there exists $p_{ki}^\star$ such that individual utility function $f_k^{\mathrm{EE}}(\mathbf{P}_k, \mathbf{P}_{-k})$ is an increasing function in $(0, p_{ki}^\star)$ and a decreasing function in $(p_{ki}^\star, +\infty)$ with respect to p_{ki}, where $p_{ki}^\star$ is the i-th component of user k's BR for given $\mathbf{P}_{-k}$. Then f_k^{EE} is also an increasing function in $(0, U^\star(\mathbf{P}_{-k}))$ and a decreasing function in $(U^\star(\mathbf{P}_{-k}), +\infty)$ w.r.t. parameter h. In other words, to find $\mathbf{P}_k = \mathrm{BR}(\mathbf{P}_{-k})$, it is sufficient to find $U(\mathbf{P}_{-k})$ by a bisection search due to the special monotonicity of the utility function.

However, it is worth mentioning that it is still difficult to directly find the solution of iterative equation (6.12). because this solution is actually implicitly given. We would like to further simplify (6.12) to facilitate the calculation of BR or NE. To start with, we assume that $N_t = N_r$. Firstly, we remove the diagonal operator of LHS of (6.12). Therefore we have :

$$\mathbf{P}_k^{(t)} = \frac{1}{h}\mathbf{I}_{N_t} - \mathbf{\Lambda}_k^{-1}\left(\mathbf{F}_k^{(t-1)} + \sigma^2 \mathbf{I}_{N_r}\right)\mathbf{\Lambda}_k^{-1} \tag{6.13}$$

If $N_t > N_r$ or $N_t < N_r$ then $\mathbf{\Lambda}_k$ is not directly invertible, then we should consider the

pseudo-inverse matrix of $\mathbf{\Lambda}_k$. Without loss of generality, we assume that $N_t > N_r$, denoting the right pseudo-inverse of $\mathbf{\Lambda}_k$ as $\mathbf{\Lambda}_k^\dagger$ then one has $\mathbf{\Lambda}_k\mathbf{\Lambda}_k^\dagger = \mathbf{I}_{N_r}$ and $\left(\mathbf{\Lambda}_k^\dagger\right)^H \mathbf{\Lambda}_k^H = \mathbf{I}_{N_r}$. Similarly, one has :

$$\begin{aligned}\mathbf{\Lambda}_k^H \left(\mathbf{\Lambda}_k \mathbf{P}_k^{(t)} \mathbf{\Lambda}_k^H + \mathbf{F}_k^{(t-1)} + \sigma^2 \mathbf{I}_r\right)^{-1} \mathbf{\Lambda}_k &= h\mathbf{I}_{N_t} \\ \left(\mathbf{\Lambda}_k \mathbf{P}_k^{(t)} \mathbf{\Lambda}_k^H + \mathbf{F}_k^{(t-1)} + \sigma^2 \mathbf{I}_r\right)^{-1} &= h\left(\mathbf{\Lambda}_k^\dagger\right)^H \mathbf{\Lambda}_k^\dagger \end{aligned} \tag{6.14}$$

However, it is generally impossible to have $\mathbf{\Lambda}_k^\dagger \mathbf{\Lambda}_k = \mathbf{I}_{N_t}$. Thus the equality does not always holds when we multiply $\mathbf{\Lambda}_k^\dagger$ on left and $\left(\mathbf{\Lambda}_k^\dagger\right)^H$ on the right on both sides of the equation. Nevertheless, this operation will yield a linear approximation of the BR dynamics :

$$\widehat{\mathbf{P}}_k^{(t)} = \frac{\mathbf{\Lambda}_k^\dagger \left[\left(\mathbf{\Lambda}_k^\dagger\right)^H \mathbf{\Lambda}_k^\dagger\right]^{-1} \left(\mathbf{\Lambda}_k^\dagger\right)^H}{h} - \mathbf{\Lambda}_k^\dagger \left(\mathbf{F}_k^{(t-1)} + \sigma^2 \mathbf{I}_{N_r}\right) \left(\mathbf{\Lambda}_k^\dagger\right)^H \tag{6.15}$$

Similarly, if $N_t < N_r$ we can obtain exactly the same iterative equation as (6.15). This type of dynamics belongs to the so-called ε-approximate best response. If one deploys (6.15) as the BR dynamics to compute NE according to alg. 8, one may not achieve the NE of the game. However, simulation results will show that the deviation is small. To this end, we obtain a sub-optimal algorithm summarized in alg. 8 by using the iterative equation deduced in (6.15) instead of using (6.11).

6.5 Numeric Results

The goal of this part is to show the performance of the proposed algorithms. Notice if $N_t = N_r$, (6.15) degenerates to (6.13) which conserves the optimality of best response. For this situation, we choose $N_t = N_r = 2$ with $K = 2$ users. A sufficient large power budget is chosen so that the BR is included in the feasible action set $\overline{P}_k = 10\text{mW}$ for $\forall k \in \{1, 2\}$ and the circuit power is $P_c = 1\text{mW}$. The error tolerance for alg. 8 is $\varepsilon_1 = \varepsilon_2 = 0.001$.

In Fig. 6.1, the achievable utility region, the average performance under NE found by alg. 8 and the averaged performance achieved by uniform power allocation (UPA) are depicted. All results are averaged over 1000 randomly generated channel samples. It is observed that the performance achieved by deploying UPA is Pareto-dominated by NE which can be found by alg. 8. Furthermore, the NE found by alg. 8 is closed to the Pareto frontier achieved by some centralized algorithms which suggest the efficiency using alg. 8 is higher than UPA.

Moreover, define the social welfare for a given action profile as $w(\boldsymbol{p}) = \sum_{k\in\mathcal{K}} f_k^{\text{EE}}\left(\boldsymbol{p}_k, \boldsymbol{p}_{-k}\right)$. Then the average social welfare as function of number of number of antennas (still we keep $N_t = N_r$) and the power budget in Fig. 6.2 and Fig. 6.3 respectively. For Fig. 6.2, the averaged social welfare of both UPA policy and our proposed algorithm is increased quasi-linearly as the number of antennas grows. However our proposed algorithm always outperforms the optimal UPA policy which is allowed to tune the power but always equally shared among each transmit antenna. In Fig. 6.3, we would like to show the influence of user's power budget. There are two different regions for social

Initialization : Number of decisions M ; set $\mathbf{P}_k^{(0)} = \frac{1}{N_t}\mathbf{I}_{N_t}, \forall k$. Choose T, ε_1 and ε_2 ;
for $t = 1$ *to* T **do**
  for $k = 1$ *to* K **do**
    Initialization : $\underline{h} = 0$ and $\overline{h} = h_{max}$;
    while $\overline{h} - \underline{h} \geq \varepsilon_1$ **do**
      $h_{\mathrm{M}} = \frac{\underline{h}+\overline{h}}{2}$, $h_{\mathrm{L}} = \max\left(0, h_{\mathrm{M}} - \frac{\varepsilon_1}{2}\right)$ and $h_{\mathrm{R}} = \min\left(h_{max}, h_{\mathrm{M}} + \frac{\varepsilon_1}{2}\right)$;
      Compute $\mathbf{P}_k\left(h_i\right)$ using (6.15), $i \in \{\mathrm{L}, \mathrm{M}, \mathrm{R}\}$;
      $U_i = u_k\left(\mathbf{P}_k(h_i), \mathbf{P}_{-k}^{(t-1)}\right)$, $i \in \{\mathrm{L}, \mathrm{M}, \mathrm{R}\}$;
      if $U_L < U_M < U_R$ **then**
        $\underline{h} = h_{\mathrm{L}}$;
        else if $U_L > U_M > U_R$ **then**
          $\overline{h} = h_{\mathrm{R}}$;
        end
      else
        $\underline{h} = h_{\mathrm{L}}$ and $\overline{h} = h_{\mathrm{R}}$;
      end
    end
    $\mathbf{P}_k^{(t)} = \frac{\mathbf{\Lambda}_k^{\dagger}\left[\left(\mathbf{\Lambda}_k^{\dagger}\right)^H \mathbf{\Lambda}_k^{\dagger}\right]^{-1}\left(\mathbf{\Lambda}_k^{\dagger}\right)^H}{h_{\mathrm{M}}} - \mathbf{\Lambda}_k^{\dagger}\left(\mathbf{F}_k^{(t-1)} + \sigma^2\mathbf{I}_{N_r}\right)\left(\mathbf{\Lambda}_k^{\dagger}\right)^H$;
  end
  if $\sum_k \left\|\mathbf{P}_k^{(t)} - \mathbf{P}_k^{(t-1)}\right\| < \varepsilon_2$ **then**
    Break ;
  end
end
Output: $\mathbf{P}_k^{\mathrm{NE}} = \mathbf{P}_k^{(t)}$ for $\forall k$;

Algorithm 8: Bisection Search Algorithm to approach the NE of the EE game

welfare. In the first region where the power budget is sufficiently large, the NE found by our proposed algorithm is independent of the power budget while the performance of UPA is decreasing with respect to the increase of the power budget. In the second region where the power budget is relatively small, Using proposed algorithm, it is not sure to converge to the NE of the game because Prop. 6.3.5 is no more valid in this region. Nevertheless, the performance achieved by our algorithm is still better than UPA which prove the superiority of our algorithm.

Then a more probable situation is considered where $N_t < N_r$ meaning that the number of antennas in user terminal is less than the one in base station. The discussion in Sec. 6.4 shows that the proposed suboptimal algorithm is actually suboptimal due to the usage of ε-approximate best response. For numeric demonstration, we choose $N_t = 2 < N_r = 4$. The performance of alg. 8 is illustrated in Fig. 6.4. The sub-optimality is clearly demonstrated in this figure. However, the resulted policy actually Pareto-dominates the exact NE found by alg. 7 and the dispersion is relatively small in terms of average performance. This remark entails that even the policy found by alg. 8 is not the NE of the game in its sub-optimal region however its performance does slightly outperforms the exact NE. Moreover the proposed algorithm is easy to implement for using explicit iterative equation even if

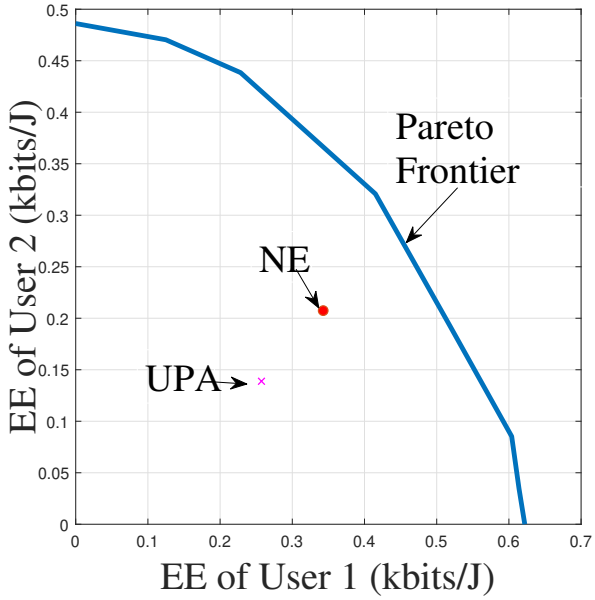


FIGURE 6.1 – Energy Efficiencoes under NE and uniform power allocation (UPA) with $N_t = N_r = 2$ for 2-user situation. NE found by our exact algorithm outperforms Uniform power allocation (UPA) policy.

it is approximated.

6.6 Conclusions

In this chapter, a game where the individual utility function is the energy efficiency in a MIMO multiple access channel system is considered. The existence and the uniqueness of Nash Equilibrium is proved and an exact algorithm and a suboptimal algorithm is proposed to find the NE of this game. Simulation results show that if the the number of transmit antennas and the number of receiving antennas is the same, performance under NE found by proposed algorithms is always better than uniform power allocation policy for both inside or outside the range covered by the main proposition of the chapter. When the condition for antennas is not met, our proposed algorithm actually deploys an ε-approximate best response which will not lead to a pure Nash Equilibrium. Quiet surprisingly the approximate solution found by our sub-optimal algorithm slightly Pareto-dominates the exact NE of the game. This observation shows that the performance of proposed algorithm is acceptable while it is relatively easy to implement. Other techniques such as pricing might be useful to improve the efficiency of the overall system. The situation where each user is allowed to freely choose its covariance matrix merely constrained to the maximum power is the natural extension of this chapter. Moreover, the discussion over the effect of successive interference cancellation and multiple carrier seems to be complicated and serve as the challenge of the future works. Nash Equilibrium analysis fo this game could be served as basic results for our goal-oriented quantization framework involving in performance under equilibrium regime.

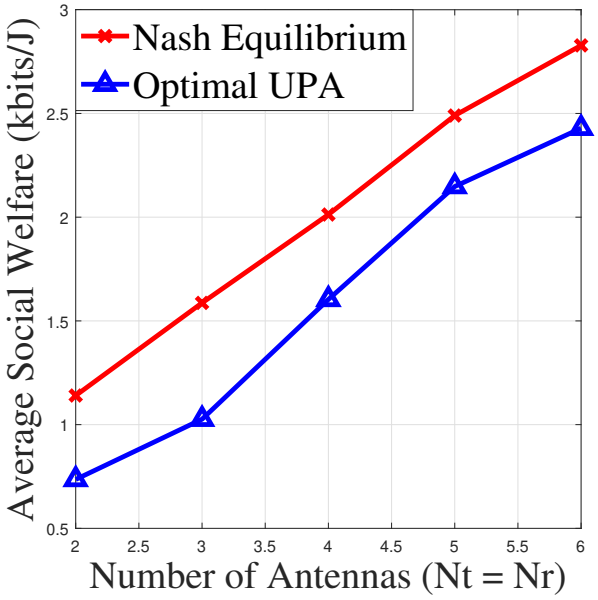


FIGURE 6.2 – Average social welfare under NE and uniform power allocation as function of number of antennas ($N_t = N_r$) with $\overline{P}_k = 10$mW for 2-user situation. NE found by our alg. 7 outperforms UPA

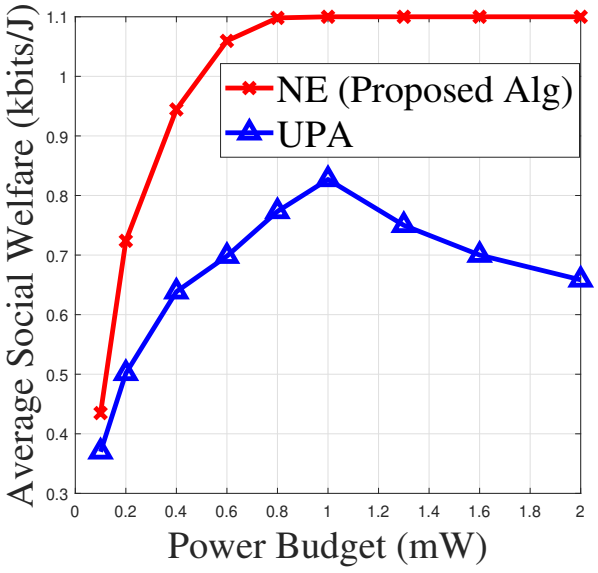


FIGURE 6.3 – Performance under NE and UPA as function of the power budget of user with $N_t = N_r = 2$ for 2-user situation. There are two different regions : one corresponds to Prop. 6.3.5. In the region uncovered by Prop. 6.3.5, alg. 8 still dominates UPA.

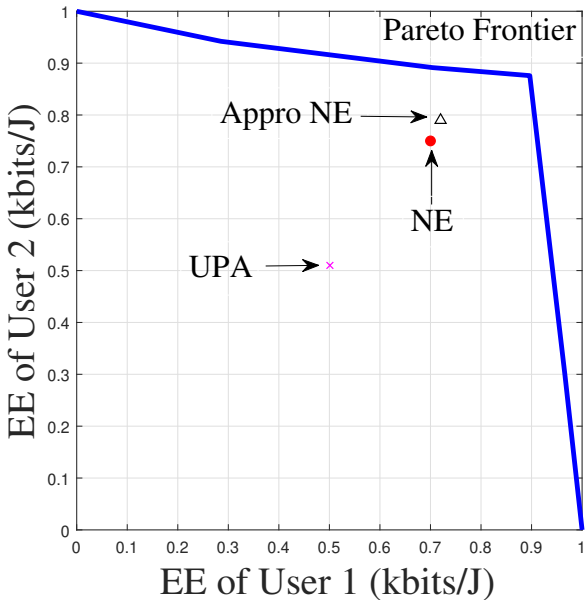


FIGURE 6.4 – Performance achieved by alg. 7 (NE) and alg. 8 and UPA with $N_t = 2$ and $N_r = 4$ for 2-user situation. Policy found by alg. 8 is very near to the exact NE and Pareto-dominates it. Moreover, two policies found by proposed algorithms both outperform UPA.

7

Conclusions and Perspectives

7.1 Conclusions

In this manuscript, the goal-oriented quantization problem is formulated and solved in different scenarios. Conventional quantization scheme is typically designed to minimize some distortion measure between the original signal and its representation, regardless of the system task. However, in many signal processing applications, the goal is not to recover the transmitted signal, but to extract essential information from the quantized signal to help the accomplishment of a goal.

In chapter 3, the goal-oriented quantization problem is first formulated for one-shot quantization scenario for single utility function. The basic goal-oriented quantization algorithm is proposed by mimicking the Lloyd-Max algorithm. The original problem is split into two sub-problems : i) finding optimal decision set for given quantization region ; finding quantization regions for given decision set. Then we consider a special case of the goal-oriented quantization problem for the the decision space of the cost function being polyhedral and convex and concave utility functions. With the help of (generalized) Jensen's inequality, the optimality loss is upper bounded by a linear approximation which is easy to be minimized. Moreover, for a given parameter sample set, decision sets are divided into equivalent classes based on the optimal decision label. We propose an approach named improve and branch algorithm which iteratively improves the decision set in a greedy way within a equivalent class to find the best decision set which minimizes the upper bound of the optimality loss. The proposed method requires no knowledge on the optimal decision function of the utility function and could be extended to the weakly concave utility functions with a little modification. Numerical results show that proposed algorithm outperforms conventional approach. Finally it is important to point out that proposed method could be redundant if the optimal decision set lies on the vertices of the decision polyhedron, e.g. binary power control.

The final part of chapter 3 is dedicated to the goal-oriented quantization problem

when only the realizations of utility functions are known. The problem is divided into two steps. A feed-forward neural network is proposed to learn the quantization regions fixing the decision set based on a trainining set. Numerical results shows that sum-rate capacity requires few quantization bits to achieve a small optimality loss. However, for energy-efficiency, channel gains need to quantized more accurately entails the necessity of further research on how to design a goal-oriented quantizer . To find optimal decision set, an evolutionary algorithm called IWO-DE algorithm which combines two classic evolutionary algorithm is used. IWO-DE mimics the processing of weed occupying the fields within certain generations of breeding. We take the problem of finding jointly set of power level and beamforming vectors to maximize the energy efficiency of the system as our numerical application to show the potential benefits of our approach. Our approach is shown to outperform the best state-of-the-art techniques such as Lloyd-Max algorithm and RVQ. A reduction of 50% quantization bits is observed in this comparison. However the drawback of our proposed method is the rapid increase of computation time as the dimension of problem grows.

In chapter 4, inspired by the high resolution quantization theory for distortion-like quantization, we try to apply this theory to scalar case and vector case of goal-oriented quantization problem separately. For scalar case, the proposed new approximate formula of optimality loss leads to a new quality defined as value density representing the importance of parameter for its contribution to the average optimality loss. We introduce a new quality normalized optimality loss when comparing the hardness of quantization for different cost functions. By merely approximating this quality in high resolution regime, we are capable to determine the hardness of quantization without performing real simulations. For vector case, the similar approximate formula of the opitmality loss is hard to obtain since the characterization matrix is cell-depending. Nevertheless, by admitting the correctness of Gersho's conjecture, Upper bound and lower bound are derived for optimality loss introduced by goal-oriented quantizer. Moreover, we propose a new algorithm by iteratively updating a single representatives in each iteration based on the eigenvalue approximation of the average optimality loss to design a goal-oriented quantizer. In each iteration, one tries to find the worst parameter sample in the sense of introducing the largest individual optimality loss. Then its corresponding representative is improved so that the average optimality loss is reduced. The algorithm stops if such operation is no longer possible. Numerical results shows that the satisfying update is slightly dominated by greedy update which aims at minimizing the average optimality loss in current iterations.

In chapter 5, instead of focusing on the optimal structure of the goal-oriented communication, we start to tackle the goal-oriented quantization problem when several correlative utility functions targeted by different users of the system. In other words, the goal-oriented quantization problem is developed in the framework of games. More specifically, we restrict ourselves in the study of potential games with identical action set. Taking the social welfare as our performance criterion, we have proven that the maximum social welfare is a submodular function of the action set with Nash equilibrium refined in argmax set of potential function. Moreover, the famous Braess's paradox is related to the monotonicity of this function. Based on these properties, we design an algorithm to find a finite action set aiming at maximizing the average social welfare under Nash equilibrium of the system. We take the multi-user MIMO multiple access channel game with spectral efficiency as the individual utility in which the existence of Braess's paradox is already confirmed as

the application of our theory. For 2-user 2-band scenario, we have proven that the channel selection set is the optimal action set for maximizing the social welfare. Telatar-type set is conjectured to be the optimal action set for maximizing the social welfare under Nash equilibrium in general cases. The existence of Braess's paradox is not guaranteed for general utility function of arbitrary game. The feasibility of proposed methods should be verified for other applications.

In chapter 6, a game where the individual utility function is the energy efficiency in a MIMO multiple access channel system is considered. The existence and the uniqueness of Nash Equilibrium is proved by showing the underlying game is a standard game if the total power of the system is less than a threshold. An algorithm is proposed to find the NE of this game by replacing the exact best response dynamics by an approximate one . Simulation results show that if the the number of transmit antennas and the number of receiving antennas is the same, performance of solution found by proposed algorithms is exactly the Nash equilibrium of the game. When the condition for antennas is not met, our proposed algorithm actually deploys an ε-approximate best response which will not lead to a pure Nash Equilibrium surely. Quiet surprisingly the approximate NE found slightly Pareto-dominates the exact NE of the game. The discretization of the action space will severely influence the determination of NE since it transforms the nature of the game. This could be the main challenge of the future works.

7.2 Perspectives

- This manuscript is dedicated to the goal-oriented quantiation problem in on one-shot quantization scenario. It could be more reasonable to quantize a sequence of parameters, i.e., goal-oriented coding. The delay between decision-making and parameter observation could always be troublesome for our goal-oriented communication problem which is ignored in this manuscript.
- In chapter 3, to find the quantization region of a goal-orientd quantizer, a simple feed-forward neural network is proposed to do so. However if the underlying optimization problem is complicated to solved, it is reasonable to apply advanced neural network to find the quantization regions.
- In chapter 3, we consider the speical case where the polyhedral decision space of the goal-oriented quantization problem is convex and the utility function is at least weakly concave. For a given parameter sample set, the sequence of optimal decision label introduces an equivalent relation. Our methods takes good advantage of the convexity of the utility function within an equivalent class. One of future works of this part is the extension to more general utility function with arbitrary decision space. Moreover, when the number of samples tends to large, the complexity of current method could be unacceptable for practical applications.
- For chapter 4, the goal-oriented quantization problem is considered in high-resolution regime. Using Taylor expansion, an approximate formula of the optimality loss is obtained. Higher order term of Taylor expasion will improve the accuracy of the approximate formula but reduce the simplicity of the formula. For vector case, our approximate formula stops for a universal term is missing characterizing the hardness of compression at a particular parameter point. How to find a quantity similar to value density as the scalar case should be further explored in the future.

Finally the proposed algorithm for vector case behaves not as good as expected. This observation will obstacle the possible applications of our approach into high-dimensional scenarios such as images and videos.

- For chapter 5, instead of considering single utility function as the objective of the entire system, utility function of users in the system could be completely different. The action set optimization problem is deeply correlated with the existence of the famous Braess's Paradox in game theory. Our analysis is merely limited in potential games which are a rather easy to solve compared to other types of game. In this chapter, the quantization of user's information are assumed to be simultaneously performed. However if the quantization could be performed sequentially (guaranteeing a high confidentiality in normal operating mode then guarantee maximum receiver reactivity in fault mode), how to design a goal-oriented quantizer could also be challenging in this scenario. Finally, if the accuracy requirement of quantization for each user is different due to the the hardness of quantization for utility function itself or the security reasons, the goal-oriented quantization problem should be re-formulated and revisited.
- In chapter 6, we confirm that the study of goal-oriented communication should not be restricted in quantization problem merely. When user apply approximate best response instead of best response which means user does not insist in optimizing individual interest, the overall performance of the system could be improved compared to the selfish case (Nash Equilibirum). However the university of this phenomenon remains to be examined for general case of other game or decentralized systems.

A

Proof of Proposition 4.2.2

We start by studying f_{worst}. One has :

$$\begin{aligned}
\widehat{\mathrm{L}}\left(\mathcal{Q}; f, R, \phi\right) &= \frac{C\left(k\right)}{2^{Rk}} \left(\int_{\mathcal{G}} (q\phi)^{\frac{1}{k+1}} \mathrm{d}g \right)^{k+1} \\
&= \frac{C\left(k\right)}{2^{Rk}} \left\| q\phi \right\|_{\frac{1}{k+1}} \\
&\overset{(a)}{\leq} \frac{C\left(k\right)}{2^{Rk}} \|q\|_{\frac{1}{k}} \left\| \phi \right\|_{\frac{1}{1}} \\
&\overset{(b)}{=} A_f \frac{C\left(k\right)}{2^{Rk}} \left\| \phi^{\frac{1}{k}} \right\|_{\frac{1}{k}}
\end{aligned} \tag{A.1}$$

(a) is again by Hölder inequality. (b) is for the equality condition for Hölder inequality : $q^k \propto \phi$. Then the worst cost function must satisfy $q\left(g\right) \propto \phi^{\frac{1}{k}}$. (c) A_f is a constant depending on the cost function.

For best cost function, we resort to variational principle. To minimize the optimality loss $\widehat{\mathrm{L}}\left(\mathcal{Q}; f, R, \phi\right)$ with $\int_{g \in \mathcal{G}} q\left(g\right) \phi\left(g\right) \mathrm{d}g = C$, it is equivalent to the following functional optimization problem :

$$\begin{aligned}
\text{minimize } \mathcal{H}\left(\phi, q\right) &= \int_{\mathcal{G}} H\left(\phi\left(g\right), q\left(g\right)\right) \mathrm{d}g \\
\text{s.t. } \mathcal{E}\left(\phi, q\right) &= \int_{\mathcal{G}} E\left(\phi\left(g\right), q\left(g\right)\right) \mathrm{d}g = C
\end{aligned} \tag{A.2}$$

with $H\left(\phi\left(g\right), q\left(g\right)\right) = \left(q\phi\right)^{\frac{1}{k+1}}$ and $H\left(\phi\left(g\right), q\left(g\right)\right) = q\phi$. We introduce the Lagrangian

for OP in (A.2) :

$$\begin{aligned}
&\mathcal{L}(\phi, q; \gamma) \\
=&\mathcal{H}(\phi, q) + \gamma\left(\mathcal{E}(\phi, q) - C\right) \\
=&\int_{\mathcal{G}} H(\phi(g), q(g)) + \gamma\left(E(\phi(g), q(g)) - \frac{C}{|\mathcal{G}|}\right) \mathrm{d}g
\end{aligned} \tag{A.3}$$

We Denote $W = T + \gamma\left(E - \frac{C}{|\mathcal{G}|}\right)$, the necessary condition of optimality for OP in (A.2) is : there exists $\gamma \in \mathbb{R}$ s.t. the well known Euler-Lagrange Equation has solution, i.e.,

$$\frac{\partial W}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}q}\left[\frac{\partial W}{\partial \dot{q}}\right] = 0 \tag{A.4}$$

which implies that

$$\frac{\partial W}{\partial q} = \left((q\phi)^{-\frac{k}{k+1}} + \gamma\right)\left(\phi + q\frac{\mathrm{d}\phi}{\mathrm{d}q}\right) = 0. \tag{A.5}$$

The solution is easily obtained : $q(g) = \frac{C_b}{\phi(g)}$ with $C_b = \frac{C}{|\mathcal{G}|}$.

B

Proof of Proposition 4.3.1

We first study the lower bound of $\widehat{\mathrm{L}}\left(\mathcal{Q}; f, R, \phi\right)$, i.e., $\widehat{\mathrm{L}}_{\mathrm{inf}}\left(\mathcal{Q}; f, R, \phi\right)$. Similarly, we extend the notation of the point density $\rho(\boldsymbol{g})$ to a vector case which determines the approximate fraction of representatives contained in that region. Define the normalized moment of inertia of the cell $\mathcal{C}_m$ with representative $\boldsymbol{z}_m$ by

$$\mathscr{M}(\mathcal{C}_m, \boldsymbol{z}_m) = \frac{1}{d_2} \frac{1}{\mathrm{vol}(\mathcal{C}_m)^{1+2/d_2}} \int_{\mathcal{G}_m} \|\boldsymbol{g} - \boldsymbol{z}_m\|_2^2 \mathrm{d}\boldsymbol{g}, \tag{B.1}$$

and the inertial profile $\mathfrak{m}(\boldsymbol{g}) = \mathscr{M}(\mathcal{G}_m, \boldsymbol{z}_m)$ when $\boldsymbol{g} \in \mathcal{G}_m$, the OL can be further approximated as [22][23] :

$$\begin{aligned}
&\mathrm{L}\left(\mathcal{Q}; f, R, \phi\right) \\
=&\sum_{m=1}^{M} \int_{\mathcal{G}_m} \left(f(\boldsymbol{\kappa}(\boldsymbol{z}_m); \boldsymbol{g}) - f(\boldsymbol{\kappa}\left(\boldsymbol{g}\right); \boldsymbol{g})\right) \phi(g) \mathrm{d}\boldsymbol{g} \\
\overset{(a)}{\geq}&\sum_{m=1}^{M} \int_{\boldsymbol{g} \in \mathcal{G}_m} \frac{1}{2} \|\boldsymbol{g} - \boldsymbol{z}_m\|_2^2 \lambda_1(\boldsymbol{g}; f) \phi(\boldsymbol{g}) \mathrm{d}\boldsymbol{g} \\
\overset{(b)}{=}&\sum_{m=1}^{M} \frac{d_2}{2M^{2/d_2}} \frac{\mathscr{M}(\mathcal{C}_m, \boldsymbol{z}_m)}{\rho^{2/d_2}(\boldsymbol{z}_m)} \lambda_1(\boldsymbol{z}_m; f) \phi(\boldsymbol{z}_m) \mathrm{vol}(\mathcal{C}_m) + o\left(M\right) \\
\overset{(c)}{=}&\frac{d_2}{2M^{2/d_2}} \int \frac{\mathfrak{m}(\boldsymbol{g})}{\rho^{2/d_2}(\boldsymbol{g})} \lambda_1(\boldsymbol{g}; f) \phi(\boldsymbol{g}) \mathrm{d}\boldsymbol{g} + o\left(M\right)
\end{aligned} \tag{B.2}$$

(a) comes from the fact that $\boldsymbol{e}_m$ is a normalized vector; (b) uses the definition of $\mathscr{M}(\mathcal{C}_m, \boldsymbol{z}_m)$ and the relation $\lim_{M\to\infty} \sum_{m=1}^{M} \mathrm{vol}(\mathcal{C}_m)\rho(\boldsymbol{z}_m) = M$; (c) is still the definition of Riemman integral. This result can be seen as a special case of Bennett's integral (see [20][23]) by replacing $\phi\left(\boldsymbol{g}\right)$ by the product $\lambda_1(\boldsymbol{g}; f)\phi(\boldsymbol{g})$. However, it is unknown how to

find the optimal inertial profile $\mathfrak{m}(\boldsymbol{g})$ and it is not even known what functions are allowable as inertial profiles. To this end, Gersho [22] made the widely accepted conjecture that when R is large, most regions of a d_2-dimensional quantizer aims at minimizing or nearly minimizing the mean square error are approximately congruent to some basic tessellating d_2-dimensional cell shape $\mathbb{T}_{d_2}$. With this conjecture, the optimal inertial profile $\mathfrak{m}(\boldsymbol{g})$ can be seen as a constant M_{d_2} in high resolution case. By using the Hölder's inequality, the optimum density $\rho(\boldsymbol{g})$ to minimize the distortion can be written as

$$\rho^*(\boldsymbol{g}) = \frac{(\lambda_1(\boldsymbol{g}; f)\phi(\boldsymbol{g}))^{d_2/(d_2+2)}}{\displaystyle\int_{\boldsymbol{t}\in\mathcal{G}} (\lambda_1(\boldsymbol{t}; f)\phi(\boldsymbol{t}))^{d_2/(d_2+2)}\,\mathrm{d}\boldsymbol{t}} \tag{B.3}$$

resulting in the low bound of distortion in (4.28). Similar analysis applies to upper bound.

C

Proof of Proposition 5.3.5

The proof is trivial when one has the following relations : For any $\mathcal{B}_1, \mathcal{B}_2 \subseteq \mathcal{X}$, it holds that :

$$\overline{\omega}\left(\mathcal{B}_1 \cup \mathcal{B}_2; \boldsymbol{g}\right) = \max\left\{\overline{\omega}\left(\mathcal{B}_1; \boldsymbol{g}\right), \overline{\omega}\left(\mathcal{B}_2; \boldsymbol{g}\right)\right\}, \tag{C.1}$$

$$\overline{\omega}\left(\mathcal{B}_1 \cap \mathcal{B}_2; \boldsymbol{g}\right) \leq \min\left\{\overline{\omega}\left(\mathcal{B}_1; \boldsymbol{g}\right), \overline{\omega}\left(\mathcal{B}_2; \boldsymbol{g}\right)\right\}. \tag{C.2}$$

We begin by proving the first equality. Take a selection $\boldsymbol{a} \in \overline{\mathcal{N}}\left[\mathcal{B}_1 \cup \mathcal{B}_2; \boldsymbol{g}\right]$, one has :

$$\begin{aligned}
&\omega\left(\mathcal{B}_1 \cup \mathcal{B}_2; \boldsymbol{g}\right) \\
&= \max_{\boldsymbol{a} \in \overline{\mathcal{N}}\left[\mathcal{B}_1 \cup \mathcal{B}_2; \boldsymbol{g}\right]} w\left(\boldsymbol{a}; \boldsymbol{g}\right) \\
&= \max_{\boldsymbol{a} \in \overline{\mathcal{N}}\left[\mathcal{B}_1; \boldsymbol{g}\right] \cup \overline{\mathcal{N}}\left[\mathcal{B}_2; \boldsymbol{g}\right]} w\left(\boldsymbol{a}; \boldsymbol{g}\right) \\
&= \max\left\{\max_{\boldsymbol{a} \in \overline{\mathcal{N}}\left[\mathcal{B}_1; \boldsymbol{g}\right]} w\left(\boldsymbol{a}; \boldsymbol{g}\right), \max_{\boldsymbol{a} \in \overline{\mathcal{N}}\left[\mathcal{B}_2; \boldsymbol{g}\right]} w\left(\boldsymbol{a}; \boldsymbol{g}\right)\right\} \\
&= \max\left\{\overline{\omega}\left(\mathcal{B}_1; \boldsymbol{g}\right), \overline{\omega}\left(\mathcal{B}_2; \boldsymbol{g}\right)\right\}.
\end{aligned} \tag{C.3}$$

For the second inequality, one has :

$$\begin{aligned}
&\overline{\omega}\left(\mathcal{B}_1 \cap \mathcal{B}_2; \boldsymbol{g}\right) \\
&= \max_{\boldsymbol{a} \in \overline{\mathcal{N}}\left[\mathcal{B}_1 \cap \mathcal{B}_2; \boldsymbol{g}\right]} w\left(\boldsymbol{a}; \boldsymbol{g}\right) \\
&\leq \max_{\boldsymbol{a} \in \overline{\mathcal{N}}\left[\mathcal{B}_1; \boldsymbol{g}\right]} w\left(\boldsymbol{a}; \boldsymbol{g}\right) \\
&= \overline{\omega}\left(\mathcal{B}_1; \boldsymbol{g}\right).
\end{aligned} \tag{C.4}$$

Similarly, one has

$$\overline{\omega}\left(\mathcal{B}_1 \cap \mathcal{B}_2; \boldsymbol{g}\right) \leq \overline{\omega}\left(\mathcal{B}_2; \boldsymbol{g}\right), \tag{C.5}$$

which results in

$$\overline{\omega}\left(\mathcal{B}_1 \cap \mathcal{B}_2; \boldsymbol{g}\right) \leq \min\left\{\overline{\omega}\left(\mathcal{B}_1; \boldsymbol{g}\right), \overline{\omega}\left(\mathcal{B}_2; \boldsymbol{g}\right)\right\} \tag{C.6}$$

D

Proof of Proposition 5.4.2

We define two functions for $\alpha_i, \alpha_j \in [0,1]$:

$$u_1\left(\alpha_i, \alpha_j; \mathbf{G}\right) = \frac{\left(\frac{1}{\gamma} + g_{1,1}\alpha_i + g_{2,1}\alpha_j\right)^2}{\left(\frac{1}{\gamma} + g_{1,1}\alpha_i\right)\left(\frac{1}{\gamma} + g_{2,1}\alpha_j\right)} \tag{D.1}$$

$$u_2\left(\alpha_i, \alpha_j; \mathbf{G}\right) = \frac{\left(\frac{1}{\gamma} + g_{1,2}\left(1-\alpha_i\right) + g_{2,2}\left(1-\alpha_j\right)\right)^2}{\left(\frac{1}{\gamma} + g_{1,2}\left(1-\alpha_i\right)\right)\left(\frac{1}{\gamma} + g_{2,2}\left(1-\alpha_j\right)\right)} \tag{D.2}$$

One can find that $\omega\left(\alpha_i, \alpha_j; \mathbf{G}\right) = \log u_1\left(\alpha_i, \alpha_j; \mathbf{G}\right) + \log u_2\left(\alpha_i, \alpha_j; \mathbf{G}\right)$ and we will prove only $(0,1)$ and $(1,0)$ could be the global optimal solution for function u_1, u_2. Without loss of generality, let us take f_1 as an example. set $\frac{\partial f_1}{\partial \alpha_i} \geq 0$, one has

$$g_{1,1}\alpha_i - g_{2,1}\alpha_j + \frac{1}{\gamma} \geq 0 \tag{D.3}$$

By symmetry, one has :

$$g_{1,1}\alpha_i - g_{2,1}\alpha_j - \frac{1}{\gamma} \leq 0 \tag{D.4}$$

Obviously, if $\gamma < \min\left\{\frac{1}{g_{1,1}}, \frac{1}{g_{2,1}}\right\}$, (D.3) and (D.4) are both true for $\forall \alpha_i, \alpha_j \in [0,1]$. Then the global optimal solution could only be $(0,1)$ or $(1,0)$. Otherwise, $\exists\ \alpha_i, \alpha_j$ so that

one of (D.3) and (D.4) is true, one thus obtain

$$\begin{aligned} u_1\left(\alpha_i, \alpha_j; \mathbf{G}\right) &= \frac{\left(\frac{1}{\gamma} + g_{1,1}\alpha_i + g_{2,1}\alpha_j\right)^2}{\left(\frac{1}{\gamma} + g_{1,1}\alpha_i\right)\left(\frac{1}{\gamma} + g_{2,1}\alpha_j\right)} \\ &= \frac{4g_{2,1}^2\alpha_j^2}{g_{2,1}\alpha_j\left(\frac{1}{\gamma} + g_{2,1}\alpha_j\right)} \\ &= \frac{4g_{2,1}\alpha_j}{\frac{1}{\gamma} + g_{2,1}\alpha_j} \end{aligned} \tag{D.5}$$

Obviously $\alpha_j^\star = 1$. Similarly analysis can be apply to $u_2\left(\alpha_i, \alpha_j\right)$ which will result in $\left(\alpha_i^\star, \alpha_j^\star\right) = (0, 1)$ or $(1, 0)$. Then the optimal action set is obviously the channel selection set.

E

Proof of Proposition 6.3.5

Our proof consists of two parts : i) existence of NE ; ii) uniqueness of NE by proving that this game is a standard game. i) Existence of NE : it is easy to prove that the action set $\mathcal{P}_k$ for each player is compact (closed and bounded), combining the quasi-concavity of f_k^{EE} claimed in Prop. 6.3.3, the existence is due to Debreu-Fan-Glicksberg theorem in [73]. Moreover, Prop. 6.3.3 claims that f_k^{EE} is a pseudo-concave function w.r.t. $\mathbf{P}_k$. Due to the property of pseudo-concave function, the unique stationary point (points where derivative vanishes) is the global optimizer of the utility function if the stationary point is in the feasible action set. We first calculate the stationary point of f_k^{EE} for $\forall k \in \mathcal{K}$ using matrix calculus :

$$\frac{\partial f_k^{\mathrm{EE}}}{\partial \mathbf{P}_k} = \mathbf{0}_{N_t \times N_t}, \tag{E.1}$$

Meanwhile, one has :

$$\frac{\partial f_k^{\mathrm{EE}}}{\partial \mathbf{P}_k} = \frac{\frac{\partial R_k}{\partial \mathbf{Q}_k} \left(\mathrm{Tr}\left(\mathbf{P}_k\right) + P_c\right) - R_k \mathbf{I}_{N_t}}{\left(\mathrm{Tr}\left(\mathbf{P}_k\right) + P_c\right)^2}, \tag{E.2}$$

which is equivalent to

$$\frac{\partial R_k}{\partial \mathbf{P}_k} = \frac{R_k \mathbf{I}_{N_t}}{\mathrm{Tr}\left(\mathbf{P}_k\right) + P_c} = f_k^{\mathrm{EE}} \mathbf{I}_{N_t}. \tag{E.3}$$

Further more, we have

$$\begin{aligned}\frac{\partial R_k}{\partial \mathbf{P}_k} &= diag\left(\frac{\partial \log\det\left(\sigma^2\mathbf{I}_{N_r}+\sum_{j=1}^{K}\mathbf{U}_j\mathbf{\Lambda}_j\mathbf{P}_j\mathbf{\Lambda}_j^H\mathbf{U}_j^H\right)}{\partial \mathbf{P}_k}\right)\\ &= diag\left(\frac{\partial \log\det\left(\mathbf{\Lambda}_k\mathbf{P}_k\mathbf{\Lambda}_k^H+\mathbf{F}_k+\sigma^2\mathbf{I}_r\right)}{\partial \mathbf{P}_k}\right)\\ &= diag\left(\mathbf{\Lambda}_k^H\frac{\partial\left[\log\det\left(\mathbf{\Lambda}_k\mathbf{P}_k\mathbf{\Lambda}_k^H+\mathbf{F}_k+\sigma^2\mathbf{I}_r\right)\right]}{\partial\left[\mathbf{\Lambda}_k\mathbf{P}_k\mathbf{\Lambda}_k^H+\mathbf{F}_k\right]}\mathbf{\Lambda}_k\right)\\ &= diag\left(\mathbf{\Lambda}_k^H\left(\mathbf{\Lambda}_k\mathbf{P}_k\mathbf{\Lambda}_k^H+\mathbf{F}_k+\sigma^2\mathbf{I}_r\right)^{-1}\mathbf{\Lambda}_k\right)\end{aligned} \tag{E.4}$$

The reason for which a diagonalized operator is taken is that only the variables lying in the diagonal of $\mathbf{P}_k$ is valid, combining (E.3) and (E.4) yields :

$$diag\left(\mathbf{\Lambda}_k^H\left(\mathbf{\Lambda}_k\mathbf{P}_k\mathbf{\Lambda}_k^H+\mathbf{F}_k+\sigma^2\mathbf{I}_r\right)^{-1}\mathbf{\Lambda}_k\right)=f_k^{\mathrm{EE}}\mathbf{I}_{N_t} \tag{E.5}$$

However, the stationary point might not belong to the feasible action set $\mathcal{P}_k$. Denote $\mathbf{P}_k^*=diag\left(\boldsymbol{p}_k^*\right)$ the unique solution of (6.10) in $\mathbb{R}^{N_t}$. Before stating BR of the game, we need prove some auxiliary results. We first prove that for given $\mathbf{P}_{-k}$ and p_{kj} with $j\neq i$, p_{ki}^* is a decreasing function for $\forall p_{kj}$. If $p_{kj}^{'}\geq p_{kj}$ and suppose $\boldsymbol{p}_{ki}^{'}=\mathrm{BR}_{ki}\left(p_{kj}^{'},\mathbf{P}_{-k}\right)\geq \mathrm{BR}_{ki}\left(p_{kj},\mathbf{P}_{-k}\right)=p_{ki}^*$ then by monotonicity, one has $v\left(p_{ki}^{'},\mathbf{P}_{-k}^{\prime}\right)\geq v\left(p_{ki}^*,\mathbf{P}_{-k}\right)$ which is contradictory to the fact that $v\left(p_{ki}^{'},\mathbf{P}_{-k}^{\prime}\right)=v\left(p_{ki}^*,\mathbf{P}_{-k}\right)=0$. Then we can prove $\mathcal{I}\left(\boldsymbol{p}_k^*\right)\neq\emptyset$ by contradiction as well. We suppose that $\mathcal{I}\left(\boldsymbol{p}_k^*\right)=\emptyset$, i.e., $p_{ki}^*<0$ for $\forall k$. Then the BR is obviously $\mathrm{BR}_k\left(\mathbf{P}_{-k}\right)=\mathbf{0}_{N_t\times N_t}$ with $f_k^{\mathrm{EE}}\left(\mathrm{BR}_k\left(\mathbf{P}_{-k}\right),\mathbf{P}_{-k}\right)=0$. However, we have $f_k^{\mathrm{EE}}\left(\mathbf{P}_k,\mathbf{P}_{-k}\right)>0$ for $\mathbf{P}_k\in\mathcal{P}_k$ and $\mathbf{P}_k\neq\mathbf{0}_{N_t\times N_t}$ which leads to a contradiction. Finally due to this monotonicity of the BR and knowing that the feasible action set $\mathcal{P}_k$ is a polyhedron, then BR must be on the boundary of $\mathcal{P}_k$ except $\mathbf{0}_{N_t\times N_t}$. Since $\mathbf{P}_k=\mathbf{0}_{N_t\times N_t}$ is the only point intersected by all faces of the feasible action polyhedron, so BR can only be on the boundary defined as (6.8) corresponding to $\mathcal{I}\left(\boldsymbol{p}_k^*\right)\neq\emptyset$. This completes the proof for existence.

ii) Now we would like to prove that the BR converges to a point which is the unique NE of the game. We will achieve that by showing that the best response is a standard function.

Positivity is obviously observed in its form given by Prop. 6.3.5. To prove the monotonicity, We firstly prove that $v\left(p_{ki},\mathbf{P}_{-k}\right)$ is a decreasing function of p_{lj} for $\forall j$ and $l\neq k$:

$$\begin{aligned}\frac{\partial v}{\partial p_{lj}} &= \left[R_k^{'}\left(\gamma_k\right)\right]^2\frac{\partial\gamma_k}{\partial p_{lj}}\frac{\partial\gamma_k}{\partial p_{ki}}-R_k^{''}\left(\gamma_k\right)\frac{\partial\gamma_k}{\partial p_{lj}}\frac{\partial\gamma_k}{\partial p_{ki}}R_k\\ &\quad-\frac{\partial^2\gamma_k}{\partial p_{lj}\partial p_{ki}}R_k^{'}\left(\gamma_k\right)R_k\end{aligned} \tag{E.6}$$

One can easily verify that $\frac{\partial^2 \gamma_k}{\partial p_{lj} \partial p_{ki}} = (1+\gamma_k) \frac{\partial \gamma_k}{\partial p_{lj}} \frac{\partial \gamma_k}{\partial p_{ki}} \geq \frac{\partial \gamma_k}{\partial p_{lj}} \frac{\partial \gamma_k}{\partial p_{ki}}$, then one can obtain

$$\begin{aligned}
\frac{\partial v}{\partial p_{lj}} \leq & \left[R_k^{'}(\gamma_k)\right]^2 \frac{\partial \gamma_k}{\partial p_{lj}} \frac{\partial \gamma_k}{\partial p_{ki}} - R_k^{''}(\gamma_k) \frac{\partial \gamma_k}{\partial p_{lj}} \frac{\partial \gamma_k}{\partial p_{ki}} R_k \\
& - \frac{\partial \gamma_k}{\partial p_{lj}} \frac{\partial \gamma_k}{\partial p_{ki}} R_k^{'}(\gamma_k) R_k \\
= & \frac{\partial \gamma_k}{\partial p_{lj}} \frac{\partial \gamma_k}{\partial p_{ki}} \left[\left(R_k^{'}(\gamma_k)\right)^2 - R_k^{''}(\gamma_k) R_k - R_k^{'}(\gamma_k) R_k\right]
\end{aligned} \tag{E.7}$$

One can easily prove that $\frac{\partial \gamma_k}{\partial p_{lj}} \leq 0$ and $\left(R_k^{'}(\gamma_k)\right)^2 - R_k^{''}(\gamma_k) R_k - R_k^{'}(\gamma_k) R_k \geq 0$ resulting $\frac{\partial v}{\partial p_{lj}} \leq 0$. Then using the same argument in the proof of BR, the monotonicity is immediate. Finally, we only need to prove the scalability of BR. Denote the BR for $\mathbf{P}_{-k}$ (which might be different from $\mathbf{P}_k^*$) and $\alpha \mathbf{P}_{-k}$ as $\mathbf{P}_k^\star$ and $\mathbf{P}_{k,\alpha}^\star$ respectively, one has :

$$\begin{aligned}
& \frac{1}{\alpha} \mathbf{P}_{k,\alpha}^\star \\
= & \frac{1}{\alpha} \arg\max_{\mathbf{P}_k} f_k^{\mathrm{EE}}(\mathbf{P}_k, \mathbf{P}_{-k}) \\
= & \frac{1}{\alpha} \arg\max_{\mathbf{P}_k} \frac{\log|\mathbf{P}_k + \alpha\mathbf{F}_k + \sigma^2\mathbf{I}_r| - \log|\alpha\mathbf{F}_k + \sigma^2\mathbf{I}_r|}{\mathrm{Tr}(\mathbf{P}_k) + P_c} \\
= & \arg\max_{\mathbf{P}_k} \frac{\log|\alpha\mathbf{P}_k + \alpha\mathbf{F}_k + \sigma^2\mathbf{I}_r| - \log|\alpha\mathbf{F}_k + \sigma^2\mathbf{I}_r|}{\mathrm{Tr}(\alpha\mathbf{P}_k) + P_c} \\
= & \arg\max_{\mathbf{P}_k} \frac{\log\left|\mathbf{P}_k + \mathbf{F}_k + \frac{\sigma^2}{\alpha}\mathbf{I}_r\right| - \log\left|\mathbf{F}_k + \frac{\sigma^2}{\alpha}\mathbf{I}_r\right|}{\mathrm{Tr}(\alpha\mathbf{P}_k) + P_c} \\
< & \arg\max_{\mathbf{P}_k} \frac{\log|\mathbf{P}_k + \mathbf{F}_k + \sigma^2\mathbf{I}_r| - \log|\mathbf{F}_k + \sigma^2\mathbf{I}_r|}{\mathrm{Tr}(\alpha\mathbf{P}_k) + P_c} \\
= & \arg\max_{\mathbf{P}_k} \frac{\log|\mathbf{P}_k + \mathbf{F}_k + \sigma^2\mathbf{I}_r| - \log|\mathbf{F}_k + \sigma^2\mathbf{I}_r|}{\mathrm{Tr}(\mathbf{P}_k) + \frac{P_c}{\alpha}} \\
< & \arg\max_{\mathbf{P}_k} \frac{\log|\mathbf{P}_k + \mathbf{F}_k + \sigma^2\mathbf{I}_r| - \log|\mathbf{F}_k + \sigma^2\mathbf{I}_r|}{\mathrm{Tr}(\mathbf{P}_k) + P_c} \\
= & \mathbf{P}_k^\star
\end{aligned} \tag{E.8}$$

which completes the proof for scalability.

Bibliographie

[1] H. Zou, C. Zhang, S. Lasaulce, L. Saludjian and P. Panciatici, "Decision-Oriented Communications : Application to Energy-Efficient Resource Allocation", *In proceedings of WINCOM 2018* (invited paper), Marrakesh, Morocco.

[2] H. Zou, C. Zhang, S. Lasaulce, L. Saludjian and P. Panciatici, "Decision Set Optimization and Energy-Efficient MIMO Communications", *30th Annual International Symposium on Personal,Indoor and Mobile Radio Communications (PIMRC'19)*, Istanbul, Turkey.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, Oct. 1948.

[4] W. Weaver, "Recent contributions to the mathematical theory of communication," ETC : a review of general semantics (1953) 261–281.

[5] B. Juba, M. Sudan, "Universal semantic communication ii : A theory of goal-oriented communication," in *Electronic Colloquium on Computational Complexity (ECCC)*, volume 15, 2008.

[6] Z. Weng, Z. Qin and G. Y. Li, "Semantic Communications for Speech Recognition," arXiv :2107.11190.

[7] B. Juba, "Universal semantic communication," *Springer Science and Business Media*, 2011.

[8] O. Goldreich, B. Juba, and M. Sudan, "A Theory of Goal-Oriented Communication", *Journal of the ACM (JACM)*, Vol. 59, No. 2, April 2012.

[9] E. C. Strinati and S. Barbarossa, "6G Networks : Beyond Shannon Towards Semantic and Goal-Oriented Communications", *Computer Networks* Volume 190, 8 May 2021, 107930.

[10] H. Seo, J. Park, M. Bennis and M. Debbah, "Semantics-Native Communication with Contextual Reasoning", arXiv :2108.05681.

[11] W. J. Yun, B. Lim, S. Jung, Y. Ko J. Park, J. Kim and M. Bennis, "Attention-based Reinforcement Learning for Real-Time UAV Semantic Communication," *17th International Symposium on Wireless Communication Systems (ISWCS)*, 2021, pp. 1-6.

[12] N. Pappas and M. Kountouris, "Goal-Oriented Communication For Real-Time Tracking In Autonomous Systems," *2021 IEEE International Conference on Autonomous Systems (ICAS)*, 2021, pp. 1-5.

[13] M. Kountouris and N. Pappas, "Semantics-Empowered Communication for Networked Intelligent Systems," in *IEEE Communications Magazine*, vol. 59, no. 6, pp. 96-102, June 2021.

[14] L. Wang, N. Piotto, and D. Schonfeld, "Boosting Quantization for Lp Norm Distortion Measure", *IEEE Statistical Signal Processing Workshop (SSP)*, 2012.

[15] C. Zhang, N. Khalfet, S. Lasaulce, V. Varma, and S. Tarbouriech, "Payoff-oriented Quantization and Application to Power Control", in *15th IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2016.

[16] E. V. Belmega and S. Lasaulce, "Energy-Efficient Precoding for Multiple-Antenna Terminals", *IEEE Transactions on Signal Processing*, vol. 59, no. 1, January 2011.

[17] Donald W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", *Journal of the Society for Industrial and Applied Mathematics*, 1963, Vol. 11, No.2, 431-441.

[18] B. Müller, J. Reinhardt, and M. T. Strickland, "Neural networks : An introduction", *Springer Science & Business Media*, 2012.

[19] S. M. Betz and H. V. Poor, "Energy Efficient Communications in CDMA Networks : A Game Theoretic Analysis Considering Operating Cost", *IEEE Transactions on Signal Processing*, Vol. 56, No. 10, 5181-5190, 2008.

[20] W. R. Bennett, "Spectra of quantized signals", *Bell Syst. Tech. J.* , vol. 27, pp. 446-472, July 1948.

[21] P. F. Panter and W. Dite, "Quantizing distortion in pulse-count modulation with nonuniform spacing of levels", Proc. IRE, vol. 39, pp. 44-48, Jan. 1951.

[22] A. Gersho, "Asymptotically optimal block quantization", *IEEE Trans. Inform. Theory*, vol. 25, pp. 373-380, July 1979.

[23] R. M. Gray, and D. L. Neuhoff, "Quantization", *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2325-2383,1998.

[24] P. L. Zador, "Development and evaluation of procedures for quantizing multivariate distributions", Ph.D. dissertation, Stanford Univ., 1963, also Stanford Univ. Dept. Statist. Tech. Rep.

[25] P. L. Zador, "Topics in the asymptotic quantization of continuous random variables", *Bell Lab. Tech. Memo.*, 1966.

[26] W. Kreitmeier and T. Linder, "High-Resolution Scalar Quantization With Rényi Entropy Constraint," in *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6837-6859, Oct. 2011.

[27] V. Misra, V. K. Goyal and L. R. Varshney, "Distributed Scalar Quantization for Computing : High-Resolution Analysis and Extensions," in *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5298-5325, Aug. 2011.

[28] C. Zhang, N. Khalfet, S. Lasaulce, V. Varma, and S. Tarbouriech, "Payoff-oriented quantization and application to power control", *15th IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, Paris, France, 2017.

[29] C. Zhang, S. Lasaulce, M. Hennebel, L. Saludjian and H. V. Poor, "Goal-oriented clustering and application to smart grid using smart meter data", to be submitted to TSG.

[30] E. V. Belmega and S. Lasaulce, "Energy-Efficient Precoding for Multiple-Antenna Terminals" *IEEE Transactions on Signal Processing*, vol. 59, no. 1, January 2011.

[31] S. Lloyd, "Least squares quantization in PCM", *IEEE transactions on information theory*, 28(2), 129-137, 1982.

[32] J. Max, "Quantizing for minimum distortion", *IRE Transactions on Information Theory*, 6(1), 7-12, 1960.

[33] P. Fleischer, "Sufficient conditions for achieving minimum distortion in a quantizer," *IEEE Innt. Conv. Rec.*, pp. 104- 111, 1964.

[34] N. Shlezinger and Y. C. Eldar, "Deep Task-Based Quantization", arXiv preprint, arXiv :1908.06845.

[35] N. Shlezinger, Y. C. Eldar, and M. R. Rodrigues, "Asymptotic task-based quantization with application to massive MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 15, pp. 3995-4012, 2019.

[36] P. Li, N. Shlezinger, H. Zhang, B. Wang and Y. C. Eldar, "Graph Signal Compression via Task-Based Quantization," *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP'21), pp. 5514-5518.

[37] A. Mostaani, T. X. Vu, S. Chatzinotas and B. Ottersten, "Task-Based Information Compression for Multi-Agent Communication Problems with Channel Rate Constraints," arXiv :2005.14220.

[38] J. Zheng, E. R. Duni and B. D. Rao, "Analysis of Multiple-Antenna Systems With Finite-Rate Feedback Using High-Resolution Quantization Theory," *in IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1461-1476, April 2007.

[39] R. Cabral Farias and J. Brossier, "Scalar Quantization for Estimation : From An Asymptotic Design to a Practical Solution," *in IEEE Transactions on Signal Processing,* vol. 62, no. 11, pp. 2860-2870, June 1, 2014.

[40] Shashank Kumbhare, Amir Shahmoradi (2020). "MatDRAM : A pure-MATLAB Delayed-Rejection Adaptive Metropolis-Hastings Markov Chain Monte Carlo Sampler", *Journal of Computer Physics Communications* (submitted).

[41] V. Misra, V. K. Goyal, and L. R. Varshney, "Distributed scalar quantization for computing : High-resolution analysis and extensions," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5298–5325, Aug. 2011.

[42] Garey MR, Johnson D, Witsenhausen H., "The complexity of the generalized Lloyd-max problem" (corresp.). IEEE Trans Inf Theory. 1982 Mar ; 28(2) :255-6.

[43] Hanna OA, Ezzeldin YH, Sadjadpour T, Fragouli C, Diggavi S., "On Distributed Quantization for Classification", *IEEE Journal on Selected Areas in Information Theory.* vol. 1, no. 1, pp. 237-249, May 2020.

[44] A. Gjendemsjo, D. Gesbert, G. E. Oien and S. G. Kiani, "Binary Power Control for Sum Rate Maximization over Multiple Interfering Links," in *IEEE Transactions on Wireless Communications* , vol. 7, no. 8, pp. 3164-3173, August 2008.

[45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K. : Cambridge Univ. Press, 2004.

[46] T.K.Y. Lo, "Maximum Ratio Transmission", *IEEE International Conference on Communications*, 6-10 June 1999, Vancouver, Canada.

[47] D. Love, R. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless system", *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735-2747, Oct. 2003.

[48] C. K. Au-yeung and D. Love , "On the performance of random vector quantization limited feedback beamforming in a MISO system", *IEEE Trans. on Signal Processing*, vol. 6, pp. 458-462, Feb. 2007.

[49] N. Jindal, "MIMO Broadcast Channels With Finite-Rate Feedback", *Trans. on Info Theory*, vol. 52, no. 11, pp. 5045-5060, Nov. 2006.

[50] D. Love and R. Heath, "Equal Gain Transmission in Multiple-Input Multiple-Output Wireless Systems", *IEEE Trans. on Communication*, VOL. 51, NO. 7, JULY 2003.

[51] O. Amin, E. Bedeer, M. H. Ahmed and O. A. Dobre, "A Novel Energy Efficient Scheme With a Finite-Rate Feedback Channel", *IEEE Communications Letters*, VOL. 3, NO. 5, October 2014.

[52] G. Caire, N. Jindal and M. Kobayashi, "Achievable rates of MIMO downlink beamforming with non-perfect CSI : a comparison between quantized and analog feedback", *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA.

[53] T.L. Marzetta, B.M. Hochwald, "Fast Transfer of Channel State Information in Wireless Systems", *IEEE Transactions on Signal Processing*, VOL. 54, June 2004.

[54] J.C. Roh, B.D. Rao, "Transmit Beamforming in Multiple-Antenna Systems With Finite Rate Feedback : A VQ-Based Approach", *IEEE Transactions on Information Theory*, VOL. 52, NO.3, 2006.

[55] X. He, H. Xing, Y. Chen and A. Nallanathan, "Energy-Efficient Mobile-Edge Computation Offloading for Applications with Shared Data", *In proceedings of Globecom 2018*, Abu Dhabi, UAE.

[56] Y. Wang, M. Zhou, Z. Tian and W.Tany, "Beamforming and Artificial Noise Design for Energy Efficient Cloud RAN with CSI Uncertainty", *In proceedings of Globecom 2018*, Abu Dhabi, UAE.

[57] R. Radner, "Team decision problems", *The Annals of Mathematical Statistics*, 33(3) :857–881, 1962.

[58] A. Zappone, E. Björnson, L. Sanguinetti and E. Jorswieck, "Globally optimal energy-efficient power control and receiver design in wireless networks", *IEEE Transactions on Signal Processing*, 2017.

[59] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory", *Foundations and Trends® in Communications and Information Theory*, VOL. 11, NO. 3-4, 2015.

[60] A. Zappone, Z. Chong and E. Jorswieck, "Energy-Aware Competitive Power Control in Relay-Assisted Interference Wireless Networks", *IEEE Transactions on Wireless Communication*, 201.

[61] David B. Fogel, "The Advantages of Evolutionary Computation", *In Proceeding of Biocomputing and emergent computation*, page 1-11, 1997.

[62] S. Liu, L. Xie and D.E. Quevedo, "Event-Triggered Quantized Communication-Based Distributed Convex Optimization", *IEEE Trans. on Control of Network Systems*, VOL. 5, NO. 1, MARCH 2018.

[63] F. Richter, A. J. Fehske, and G. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Networks", *IEEE Proceedings of VTC*, Fall'2009.

[64] K. Arulkumaran, M.P. Deisenroth, M. Brundage and A.A. Bharath, "Deep Reinforcement Learning : A brief Survey", *Deep Learning for Visual Understanding*, 13 november 2017.

[65] A. Mehrabian and C. Lucas, "A novel numerical optimization algorithm inspired from weed colonization", *Ecol. Inform.*, vol. 1, no. 4, pp. 355–366, Dec. 2006.

[66] R. Storn and K. Price, "Differential Evolution : A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces", *Journal of Global Optimization*, 11 : 341–359, 1997.

[67] X. Cai, Z. Hu, and Z. Fan, "A novel memetic algorithm based on invasive weed optimization and differential evolution for constrained optimization," *Soft Computing*, vol. 17, no. 10, pp. 1893–1910, Oct. 2013.

[68] S. Lloyd, "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, 28(2), 129-137, 1982.

[69] A. Zappone, Z. Chong and E. Jorswieck, "Energy-Aware Competitive Power Control in Relay-Assisted Interference Wireless Networks", *IEEE Transactions on Wireless Communication*, vol.12, 2013.

[70] C. Zhang, S. Lasaulce, A. Agrawal, and R. Visoz, "Distributed Power Control with Partial Channel State Information : Performance Characterization and Design", IEEE Transactions on Vehicular Technology, vol. 68, no. 9, pp. 8982-8994, 2019.

[71] E. V. Belmega and S. Lasaulce, "Energy-Efficient Precoding for Multiple-Antenna Terminals" *IEEE Transactions on Signal Processing*, vol. 59, no. 1, January 2011.

[72] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman, "Pricing and power control in a multicell wireless data network" . *IEEE Journal on Selected Areas in Communication*, 19(10) :1883-1892, October 2001.

[73] G. Debreu, "A social equilibrium existence theorem", *In National Academy of Sciences*, volume 38, pages 886-893, 1952.

[74] W. Dinkelbach, "On nonlinear fractional programming", *Management Science*, 13(7) :492-498, March 1967.

[75] V. Varma, S. Lasaulce, M. Debbah, and S. Elayoubi, "An energy efficient framework for the analysis of MIMO slow fading channels", *IEEE Transactions on Signal Processing*, 61(10) :2647-2659, May 2013.

[76] P. S. Raghavendra and B. Daneshrad, "An energy-efficient water-filling algorithm for OFDM systems", *In 2010 IEEE International Conference on Communications (ICC)*, pages 1–5, May 2010.

[77] F. Richter, A. J. Fehske, and G. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies for Cellular Network", *IEEE Proceedings of VTC*, Fall'2009.

[78] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity", *IEEE Transactions on Wireless Communications*, 11(10) :3618-3631, October 2012.

[79] S. He, Y. Huang, S. Jin, and L. Yang, "Coordinated beamforming for energy efficient transmission in multicell multiuser systems", *IEEE Transactions on Communications*, 61(12) :4961-4971, December 2013.

[80] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination", *IEEE Transactions on Wireless Communications*, 14(1) :1–14, January 2015.

[81] B. Du, C. Pan, W. Zhang, and M. Chen, "Distributed energy-efficient power optimization for CoMP systems with max-min fairness", *IEEE Communications Letters*, 18(6) :999-1002, 2014.

[82] C. Daskalakis, P.W. Goldberg and C.H. Papadimitriou, "The Complexity of Computing a Nash Equilibrium", *SIAM Journal on Computing*, 39 (3) : 195-259, 2009.

[83] J. Nash, "Equilibrium points in n-person game", *In Proceedings of the National Academy of Sciences*, 36(1) :48-49, 1950.

[84] E.V. Belmega, S. Lasaulce, and M. Debbah, "Power allocation games for MIMO multiple access channels with coordination", IEEE Trans. on Wireless Communications, vol. 8, no. 6, pp. 3182-3192, Jun. 2009.

[85] L. D. Xu, W. He and S. Li, "Internet of Things in Industries : A Survey," in *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233-2243, Nov. 2014.

[86] A. Gupta and R. K. Jha, "A Survey of 5G Network : Architecture and Emerging Technologies," in IEEE Access, vol. 3, pp. 1206-1232, 2015,

[87] S. Zlobec "Jensen's inequality for non-convex functions", *Mathematical Communications*, 9(2004), 119-124.

[88] J. L. W. V. Jensen, Sur les fonctions convexes et les inegalites entre les valeur moyennes, *Acta Math.* 30(1906), 175-193.

[89] E. G. Larsson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[90] S. M. Perlaza, S. Lasaulce, Mérouane Debbah, "Equilibria of Channel Selection Games in Parallel Multiple Access Channels," *EURASIP Journal on Wireless Communications and Networking*, SpringerOpen, 2013, pp.1-23.

[91] N. Nie, C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," Mobile Netw Appl 11(6) :779–797.

[92] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[93] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, Nov/Dec 1999.

[94] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks : How many antennas do we need ?," *IEEE J. Sel. Areas Commun.* , vol. 31, no. 2, pp. 160-171, Feb. 2013.

[95] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Commun. Lett.* , vol. 3, no. 6, pp. 653–656, Dec. 2014.

[96] Z. Ding, and V. H. Poor, "Design of Massive-MIMO-NOMA With Limited Feedback", *IEEE Singal Processing Letters*, VOL. 23, NO. 5, MAY 2016.

[97] O. Amin, E. Bedeer, M. H. Ahmed and O. A. Dobre, "A Novel Energy Efficient Scheme With a Finite-Rate Feedback Channel", *IEEE Communications Letters*, VOL. 3, NO. 5, October 2014.

[98] A. Hyadi, Z. Rezki and M. Alouini, "Secure Multiple-Antenna Block-Fading Wiretap Channels With Limited CSI Feedback", *IEEE Trans. on Wireless Commun.*, VOL. 16, NO. 20, 2017.

[99] E. Altman, V. Kumar, and H. Kameda, "A Braess type paradox in power control over interference channels," in *Proc. 6th Intl. Symp. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*, Berlin, Germany, Apr. 2008.

[100] D. Braess. Uber ein paradoxen der werkehrsplannung. Unternehmenforschung, 12 :256-268, 1968.

[101] J. F. Nash,"Equilibrium points in n-person games," in *Proc. National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 48-49, Jan. 1950.

[102] S. Lasaulce and H. Tembine, "Game Theory and Learning in Wireless Networks : Fundamentals and Applications", Waltham, MA, USA : Elsevier Academic Press, 2011.

[103] Scutari, G., Barbarossa, S., Palomar, D.P., 2006, "Potential games : A framework for vector power control problems with coupled constraints", in *International Conference on Acoustics, Speech and Signal Processing* (ICASSP), May 2006.

[104] P. Mertikopolous, E. V. Belmega, A. Moustakas, and S. Lasaulce, "Dynamic power allocation in parallel multiple access channels," in *5th International ICST Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS)*, Paris, France, May 2011.

[105] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124-143, May 1996.

[106] A. Neyman, "Correlated equilibrium and potential games," in *International Journal of Game Theory* Vol. 26, pages 223–227 (1997).

[107] E. Tohidi, R. Amiri, M. Coutino, D. Gesbert, G. Leus and A. Karbasi, "Submodularity in Action : From Machine Learning to Signal Processing Applications," in *IEEE Signal Processing Magazine*, vol. 37, no. 5, pp. 120-133, Sept. 2020.

[108] A. Krause and D. Golovin, "Submodular function maximization.," 2014.

[109] F. R. Bach, "Learning with submodular functions : A convex optimization perspective," *Foundations and Trends® in Machine Learning*, vol. 6, no. 2-3, pp. 145-373, 2013.

[110] E. Tohidi, M. Coutino, S. P. Chepuri, H. Behroozi, M. M. Nayebi, and G. Leus, "Sparse antenna and pulse placement for colocated MIMO radar," *IEEE Transactions on Signal Processing*, vol. 67, pp. 579-593, Feb 2019.

[111] M. Coutino, S. P. Chepuri, and G. Leus, "Submodular sparse sensing for Gaussian detection with correlated observations," *IEEE Transactions on Signal Processing*, vol. 66, pp. 4025-4039, Aug 2018.

[112] K. Thekumparampil, A. Thangaraj, and R. Vaze, "Combinatorial resource allocation using submodularity of waterfilling," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 206-216, Jan 2016.

[113] D. Golovin and A. Krause, "Adaptive submodularity : Theory and applications in active learning and stochastic optimization," *Journal of Artificial Intelligence Research*, vol. 42, pp. 427-486, 2011.

[114] E. Elenberg, A. G. Dimakis, M. Feldman, and A. Karbasi, "Streaming weak submodularity : Interpreting neural networks on the fly," in *Advances in Neural Information Processing Systems*, pp. 4044-4054, 2017.

[115] V. Tzoumas, K. Gatsis, A. Jadbabaie, and G. J. Pappas, "Resilient monotone submodular function maximization," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 1362-1367, Dec 2017.

[116] U. Feige, V. S. Mirrokni, and J. Vondrak, "Maximizing non-monotone submodular functions," *SIAM Journal on Computing*, vol. 40, no. 4, pp. 1133–1153, 2011.

[117] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions-I," *Mathematical programming*, vol. 14, no. 1, pp. 265-294, 1978.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Communications Orientées Objectifs : Le Problème de Quantification

Mots clés : Quantification, réseaux de neurones artificiels, algorithmes évolutionnaires, optimisation convexe, quantification à haute résolution , théorie des jeux, équilibre de Nash, jeu de potentiel, efficacité énergétique.

Résumé : Le paradigme classique pour concevoir un émetteur (codeur) et un récepteur (décodeur) est de concevoir ces éléments en assurant que l'information reconstruite par le récepteur soit suffisamment proche de l'information que l'émetteur a mis en forme pour l'envoyer sur le médium de communication. On parle de critère de fidélité ou de qualité de reconstruction (mesurée par exemple en termes de distorsion, de taux d'erreur binaire, de taux d'erreur paquet ou de probabilité de coupure de la communication).

Le problème du paradigme classique est qu'il peut conduire à un investissement injustifié en termes de ressources de communication (surdimensionnement de l'espace de stockage de données, médium de communication à très haut débit et onéreux, composants très rapides, etc.) et même à rendre les échanges plus vulnérables aux attaques. La raison à cela est que l'exploitation de l'approche classique (fondée sur le critère de fidélité de l'information) dans les réseaux sans fil conduira typiquement à des échanges excessivement riches en information, trop riches au regard de la décision que devra prendre le destinataire de l'information. Il s'avère qu'actuellement, l'ingénieur n'a pas à sa disposition une méthodologie lui permettant de concevoir une telle paire émetteur-récepteur qui serait adaptée à l'utilisation (ou les utilisations) du destinataire.

Par conséquent, un nouveau paradigme de communication appelé la communication orientée objectif est proposé pour résoudre le problème des communications classiques. Le but ultime des communications orientées objectifs est d'accomplir certaines tâches ou certains objectifs au lieu de viser un critère de reconstruction du signal source. Les tâches sont généralement caractérisées par des fonctions d'utilité ou des fonctions de coût à optimiser.

Dans la présente thèse, nous nous concentrons sur le problème de quantification des communication orientées objectifs, c'est-à-dire la quantification orientée objectif. Nous formulons d'abord formellement le problème de quantification orientée objectif. Deuxièmement, nous proposons une approche pour résoudre le problème lorsque seules des réalisations de fonction d'utilité sont disponibles. Un scénario spécial avec quelques connaissances supplémentaires sur les propriétés de régularité des fonctions d'utilité est également traité. Troisièmement, nous étendons la théorie de la quantification à haute résolution à notre problème de quantification orientée objectif et proposons des schémas implémentables pour concevoir un quantificateur orienté objectif. Quatrièmement, le problème de quantification orientée but est développé dans un cadre de jeux sous forme stratégique. Il est montré que la quantification orientée objectif pourrait améliorer les performances globales du système si le fameux paradoxe de Braess existe. Enfin, l'équilibre de Nash d'un jeu de canaux d'accès multiples à entrées multiples et sorties multiples multi-utilisateurs avec l'efficacité énergétique étant l'utilité est étudié et réalisé selon différentes méthodes.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Title : Goal Oriented Communications : The Quantization Problem
Keywords : Quantization, artificial neural networks, evolutionary algorithm, convex optimization, high-resolution quantization, game theory, Nash equilibrium, potential games, energy efficiency.

Abstract : The classic paradigm for designing a transmitter (encoder) and a receiver (decoder) is to design these elements by ensuring that the information reconstructed by the receiver is sufficiently close to the information that the transmitter has formatted to send it on the communication medium. This is referred to as a criterion of fidelity or of reconstruction quality (measured for example in terms of distortion, binary error rate, packet error rate or communication cut-off probability).

The problem with the classic paradigm is that it can lead to an unjustified investment in terms of communication resources (oversizing of the data storage space, very high speed and expensive communication medium, very fast components, etc.) and even to make exchanges more vulnerable to attacks. The reason for this is that the use of the classic approach (based on the criterion of fidelity of information) in the wireless networks will typically lead to exchanges excessively rich in information, too rich regarding the decision which will have to be taken. the recipient of the information ; in the simpler case, this decision may even be binary, indicating that in theory a single bit of information could be sufficient. As it turns out, the engineer does not currently have at his disposal a methodology to design such a transceiver pair that would be suitable for the intended use (or uses) of the recipient.

Therefore, a new communication paradigm named the goal-oriented communication is proposed to solve the problem of classic communications. The ultimate objective of goal-oriented communications is to achieve some tasks or goals instead of improving the accuracy of reconstructed signal merely. Tasks are generally characterized by some utility functions or cost functions to be optimized.

In the present thesis, we focus on the quantization problem of the goal-oriented communication, i.e., the goal-oriented quantization. We first formulate the goal-oriented quantization problem formally. Secondly, we propose an approach to solve the problem when only realizations of utility function are available. A special scenario with some extra knowledge about regularity properties of the utility functions is treated as well. Thirdly, we extend the high-resolution quantization theory to our goal-oriented quantization problem and propose implementable schemes to design a goal-oriented quantizer. Fourthly, the goal-oriented quantification problem is developed in a framework of games in strategic form. It is shown that goal-oriented quantization could improve the overall performance of the system if the famous Braess paradox exists. Finally, Nash equilibrium of a multi-user multiple-input and multiple output multiple access channel game with energy efficiency being the utility is studied and achieved in different methods.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France