Lecture Notes in Networks and Systems 363

Hoai An Le Thi
Tao Pham Dinh
Hoai Minh Le *Editors*

Modelling, Computation and Optimization in Information Systems and Management Sciences

Proceedings of the 4th International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences - MCO 2021

Springer

Lecture Notes in Networks and Systems

Volume 363

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at https://link.springer.com/bookseries/15179

Hoai An Le Thi · Tao Pham Dinh ·
Hoai Minh Le
Editors

Modelling, Computation and Optimization in Information Systems and Management Sciences

Proceedings of the 4th International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences - MCO 2021

Springer

Editors
Hoai An Le Thi
Computer science and Applications
Department, LGIPM
University of Lorraine
Metz Cedex, France

Institut Universitaire de France (IUF)
Paris, France

Tao Pham Dinh
Laboratory of Mathematics
National Institute for Applied
Sciences - Rouen
Saint-Etienne-du-Rouvray Cedex, France

Hoai Minh Le
Computer science and Applications
Department, LGIPM
University of Lorraine
Metz Cedex, France

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-030-92665-6 ISBN 978-3-030-92666-3 (eBook)
https://doi.org/10.1007/978-3-030-92666-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume contains 34 selected full papers (from 70 submitted ones) presented at the MCO 2021 conference, held on December 11–13, 2021, at Hanoi, Vietnam.

MCO 2021 was the fourth event in the series of conferences on Modelling, Computation and Optimization in Information Systems and Management Sciences, traditionally organized by LITA, the Laboratory of Theoretical and Applied Computer Science (LITA has become now the Computer Science and Applications Department of LGIPM), University of Lorraine, in Metz, France. Exceptionally, MCO 2021 was co-organized by the Computer Science and Applications Department, LGIPM, University of Lorraine, France, and Academy of Cryptography Techniques, Vietnam, in collaboration with the Data Science and Optimization of Complex Systems (DataOpt) Laboratory, International School, Vietnam National University, Hanoi, Vietnam.

The first conference, MCO 2004, brought together 100 scientists from 21 countries. It included 8 invited plenary speakers, 70 papers presented and published in the proceedings, "Modelling, Computation and Optimization in Information Systems and Management Sciences," edited by Le Thi Hoai An and Pham Dinh Tao, Hermes Sciences Publishing, June 2004, 668 pages. Two special issues including 22 papers were published in the European Journal of Operational Research and in the Journal of Global Optimization. The second event, MCO 2008, gathered 6 invited plenary speakers and more than 120 scientists from 27 countries. The scientific program consisted of 6 plenary lectures and the oral presentations of 68 selected full papers as well as 34 selected abstracts covering all main topic areas. Its proceedings were edited by Le Thi Hoai An, Pascal Bouvry and Pham Dinh Tao in Communications in Computer and Information Science, Springer (618 pages). Two special issues were published in Journal of Computational, Optimization & Application and Advance on Data Analysis and Classification. The third edition, MCO 2015, was attended by more than 130 scientists from 35 countries. The scientific program includes 5 plenary lectures and the oral presentation of 86 selected full papers and several selected abstracts. The proceedings was edited by Le Thi Hoai An, Pham Dinh Tao and Nguyen Ngoc Thanh in Advances in Intelligent Systems and Computing, Springer (2 volumes for a total of 1000 pages).

MCO 2015, the biggest MCO edition, was marked by the celebration of the 30th birthday of DC programming and DCA, an efficient approach in nonconvex programming framework. One special issue in Mathematical Programming Series B was dedicated to DC programming and DCA, and the second special issue was published in Computers and Operations Research.

MCO 2021 covers, traditionally, several fields of management science and information systems: computer sciences, information technology, mathematical programming, optimization and operations research and related areas. It allows researchers and practitioners to clarify the recent developments in models and solutions for decision making in engineering and information systems and to interact and discuss how to reinforce the role of these fields in potential applications of great impact.

The conference program includes 3 plenary lectures of world-class speakers and the oral presentation of 34 selected papers as well as several selected abstracts.

This book covers theoretical and algorithmic aspects as well as practical issues connected with modeling, computation and optimization in information systems and management science. Each paper was peer-reviewed by at least two members of the International Program Committee and the International Reviewer Board. The book is composed of 3 parts: optimization of complex systems - models and methods, machine learning - algorithms and applications, and cryptography. We hope that researchers and practitioners can find here many inspiring ideas and useful tools and techniques for their works.

We would like to thank all those who contributed to the success of the conference and to this book of proceedings. In particular, we would like to express our gratitude to the members of International Program Committee as well as the reviewers for their hard work in the review process, which helped us to guarantee the highest quality of the selected papers for the conference. Thanks are also due to the plenary lecturers for their interesting and informative talks of a world-class standard. We wish to especially thank all members of the Organizing Committee for their excellent work to make the conference a success.

Our special thanks go to all the authors for their valuable contributions and to the other participants who enriched the conference success.

Finally, we cordially thank Springer, especially Prof. Janusz Kacprzyk and Dr. Thomas Ditzinger, for their supports in publishing this book.

October 2021

Hoai An Le Thi
Tao Pham Dinh
Hoai Minh Le

Organization

MCO 2021 is co-organized by the Computer Science and Applications Department, LGIPM, University of Lorraine, France, and Academy of Cryptography Techniques, Vietnam, in collaboration with the Data Science and Optimization of Complex Systems (DataOpt) Laboratory, International School, Vietnam National University, Hanoi, Vietnam.

Conference Chairs

Hoai An Le Thi	University of Lorraine, France
Huu Hung Nguyen	Academy of Cryptography Techniques, Vietnam

Program Chairs

Hoai An Le Thi	University of Lorraine, France
Tao Pham Dinh	National Institute of Applied Sciences of Rouen, France

Honorary Chair

Nam Hai Nguyen	Vietnam Government Information Security Commission, Vietnam

Organizing Chairs

Hoai Minh Le	University of Lorraine, France
Thi Dao Vu	Academy of Cryptography Techniques, Vietnam

Publicity Chair

Vinh Thanh Ho	University of Lorraine, France

Sponsoring Institutions

DCA Solutions (Vietnam)
Springer

Contents

Cryptography

Optimization of Complex Systems - Models and Methods

An Interior Proximal Method with Proximal Distances for Quasimonotone Equilibrium Problems

Erik Alex Papa Quiroz[1,2(✉)]

[1] Universidad Nacional Mayor de San Marcos, Lima, Peru
epapaq@unmsm.edu.pe, erick.papa@upn.edu.pe
[2] Universidad Privada del Norte, Lima, Peru

Abstract. We introduce an interior proximal point algorithm with proximal distances to solve quasimonotone Equilibrium problems defined on convex sets. Under adequate assumptions, we prove that the sequence generated by the algorithm converges to a solution of the problem and for a broad class of proximal distances the rate of convergence of the sequence is linear or superlinear.

Keywords: Proximal algorithms · Proximal distances · Quasimonotone bifunctions

1 Introduction

In this paper we consider the well known Equilibrium Problem (EP): find $\bar{x} \in \overline{C}$ such that

$$f(\bar{x}, y) \geq 0, \quad \forall y \in \overline{C}, \tag{1}$$

where $f : \overline{C} \times \overline{C} \to \mathbb{R}$ is a bifunction, C is a nonempty open convex set in the Euclidean space $\mathbb{R}^n$ and $\overline{C}$ is the closure of C.

Equilibrium problem is a general mathematical model which includes as particular cases minimization problems, variational inequalities problems, monotone inclusion problems, sadle point problems, complemetarity problems, vector minimization problems and Nash equilibria problems with noncooperative games, see for example Blum and Oettli [3], Iusem and Sosa [4,5] and references therein.

There are several methods for solving (EP), for example, splitting proximal methods [8], hybrid extragradient methods [1], extragradient methods [9], double projection-type method [15] and proximal point algorithms [6], among many others.

In previous works the standard condition to guarantee the convergence of the algorithms to solve the problem (1) is the monotonicity or the pseudomonotonicity of the bifunction $f(.,.)$. However, to broaden the field of applications of the model, in 2017, Mallma et al. [7] introduced the quasimonotone condition

Supported by Universidad Privada del Norte, Peru.

H. A. Le Thi et al. (Eds.): MCO 2021, LNNS 363, pp. 3–15, 2022.
https://doi.org/10.1007/978-3-030-92666-3_1

in the assumptions of the proposed algorithm. In that paper, for $k = 1, 2, \ldots,$ given $x^{k-1} \in C$, the authors considered the following main step: find $x^k \in C$ and $g^k \in \partial_2 f(x^k, x^k)$ such that

$$g^k + \lambda_k \nabla_1 d(x^k, x^{k-1}) = e^k, \tag{2}$$

where the error criteria satisfies $\sum_{k=1}^{+\infty} \frac{\|e^k\|}{\lambda_k} < +\infty$ and $\sum_{k=1}^{+\infty} \frac{|\langle e^k, x^k\rangle|}{\lambda_k} < +\infty$. They proved that the sequence generated by the method converges weakly to a solution of the (EP). Furthermore, if there exists an accumulation point, of the sequence generated by the method, which belongs to a certain subset of the solution set of (EP), then the sequence converges to a point of the solution subset. However, they did not report any rate of convergence results of the algorithm. This is the motivation of the present paper.

Due to that the above error criteria is not appropriate to obtain some results about the rate of convergence of the algorithm, we consider in the present paper other error criteria, that is, we consider the following criteria:

$$\frac{\|e^k\|}{\lambda_k} \leq \eta_k \sqrt{H(x^k, x^{k-1})} \tag{3}$$

$$\sum_{k=1}^{+\infty} \eta_k < +\infty \tag{4}$$

where H is an induced proximal distance, see Definition 3 of Sect. 2. Observe that (3) and (4) are new in proximal point methods with proximal distances but they are motivated from the work of Papa Quiroz and Cruzado [13].

We prove, under adequate assumptions on the (EP), the same convergence results found in the paper of Mallma et al. [7], but we add the linear or superlinear rate of convergence of the proposed algorithm, we consider this fact the principal contribution of the paper.

The organization of the paper is the following: In Sect. 2 we present the preliminaries, the concept of proximal and induced proximal distances and we introduce the definition of H-linear and H-superlinear convergence. In Sect. 3, we present the interior proximal algorithm, the assumption on the problem and the result of convergence. In Sect. 4, we present the rate of convergence of the algorithm.

2 Preliminaries

Throughout this paper $\mathbb{R}^n$ is the Euclidean space endowed with the inner product $\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i$, where $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$. The norm of x given by $||x|| := \langle x, x\rangle^{1/2}$ and $bd(C)$, $\overline{C}$ denotes the boundary and closure of the subset $C \subset \mathbb{R}^n$ respectively.

A bifunction $f : \overline{C} \times \overline{C} \to \mathbb{R}$ is said to be quasimonotone on $\overline{C}$ if $f(x, y) > 0 \Rightarrow f(y, x) \leq 0$, $\forall x, y \in \overline{C}$.

Definition 1. *Let* $f : \overline{C} \times \overline{C} \to \mathrm{I\!R}$ *be a bifunction. For each fixed* $z \in \overline{C}$, *the diagonal subdifferential of* $f(z, .)$ *at* $x \in \overline{C}$, *denoted by* $\partial_2 f(z, x)$, *is defined and denoted by*

$$\partial_2 f(z, x) = \{g \in \mathrm{I\!R}^n : f(z, y) \geq f(z, x) + \langle g, y - x\rangle , \forall y \in \overline{C}\}$$

Furthermore, if $f(x, x) = 0$, *then*

$$\partial_2 f(x, x) = \{g \in \mathrm{I\!R}^n : f(x, y) \geq \langle g, y - x\rangle , \forall y \in \overline{C}\}.$$

We present the definitions of proximal and induced proximal distances, introduced by Auslender and Teboulle [2]. For applications of these proximal distances for optimization, variational inequality problems and equilibrium problems see for example the references [10–12, 17].

Definition 2. *A function* $d : \mathrm{I\!R}^n \times \mathrm{I\!R}^n \to \mathrm{I\!R}_+ \cup \{+\infty\}$ *is called a proximal distance with respect to an open nonempty convex set* C *if for each* $y \in C$ *it satisfies the following properties:*

i. $d(\cdot, y)$ *is proper, lower semicontinuous, strictly convex and continuously differentiable on* C*;*
ii. $\mathrm{dom}\,(d(\cdot, y)) \subset \overline{C}$ *and* $\mathrm{dom}(\,\partial_1 d(\cdot, y)) = C$, *where* $\partial_1 d(\cdot, y)$ *denotes the classical subgradient map of the function* $d(\cdot, y)$ *with respect to the first variable;*
iii. $d(\cdot, y)$ *is coercive on* $\mathrm{I\!R}^n$ *(i.e.,* $\lim_{\|u\|\to\infty} d(u, y) = +\infty$);
iv. $d(y, y) = 0$.

We denote by $D(C)$ *the family of functions satisfying the above definition.*

Definition 3. *Given* $d \in D(C)$, *a function* $H : \mathrm{I\!R}^n \times \mathrm{I\!R}^n \to \mathrm{I\!R}_+ \cup \{+\infty\}$ *is called the induced proximal distance to* d *if there exists* $\gamma \in (0, 1]$ *with* H *a finite-valued function on* $C \times C$ *and for each* $a, b \in C$ *we have:*

(Ii) $H(a, a) = 0$.
(Iii) $\langle c - b, \nabla_1 d(b, a)\rangle \leq H(c, a) - H(c, b) - \gamma H(b, a), \quad \forall c \in C$; *where the notation* $\nabla_1 d(., .)$ *means the gradient of* d *with respect to the first variable.*

Denote by $(d, H) \in \mathcal{F}(\mathcal{C})$ the proximal distance that satisfies the conditions of Definition 3.
We also denote $(d, H) \in \mathcal{F}(\overline{C})$ if there exists H such that:

(Iiii) H is finite valued on $\overline{C} \times C$ satisfying (**Ii**) and (**Iii**), for each $c \in \overline{C}$.
(Iiv) For each $c \in \overline{C}$, $H(c, \cdot)$ has level bounded sets on C.

Finally, denote $(d, H) \in \mathcal{F}_+(\overline{C})$ if

(Iv) $(d, H) \in \mathcal{F}(\overline{C})$.
(Ivi) $\forall y \in \overline{C}$ y $\forall \{y^k\} \subset C$ bounded with $\lim_{k\to+\infty} H(y, y^k) = 0$, then $\lim_{k\to+\infty} y^k = y$.
(Ivii) $\forall y \in \overline{C}$, y $\forall \{y^k\} \subset C$ such that $\lim_{k\to+\infty} y^k = y$, then $\lim_{k\to+\infty} H(y, y^k) = 0$.

Remark 1. Examples of proximal distances which satisfy the above definitions may be seen in Auslender and Teboulle [2], Sect. 3.

Definition 4. *Let $(d, H) \in \mathcal{F}(\mathcal{C})$ and $\{x^k\} \subset \mathbb{R}^n$ be a sequence such that $\{x^k\}$ converges to a point $\overline{x} \in \mathbb{R}^n$. Then, the convergence is said to be:*

1. *$H-$linear, if there exist a constant $0 < \theta < 1$ and $n_0 \in \mathbb{N}$ such that*

$$H(x^k, \overline{x}) \leq \theta H(x^{k-1}, \overline{x}), \quad \forall\, k \geq n_0; \tag{5}$$

2. *$H-$superlinear, if there exist a sequence $\{\beta_k\}$ converging to zero and $\overline{n} \in \mathbb{N}$ such that*

$$H(x^k, \overline{x}) \leq \beta_k H(x^{k-1}, \overline{x}), \quad \forall\, k \geq \overline{n}. \tag{6}$$

In the particular case when the induced proximal distance H is given by $H(x, y) = \bar{\eta}\, \|x - y\|^2$, for some $\bar{\eta} > 0$, we obtain the usual definition of rate of convergence.

Lemma 1 *[14, Lemma 2, pp. 44]. Let $\{v_k\}$, $\{\gamma_k\}$, and $\{\beta_k\}$ be nonnegative sequences of real numbers satisfying $v_{k+1} \leq (1 + \gamma_k)\, v_k + \beta_k$ and such that $\sum_{k=1}^{\infty} \beta_k < \infty$, $\sum_{k=1}^{\infty} \gamma_k < \infty$. Then, the sequence $\{v_k\}$ converges.*

3 Proximal Method

Let C be a nonempty open convex set and $f : \overline{C} \times \overline{C} \to \mathbb{R}$ an equilibrium bifunction, i.e., satisfying $f(x, x) = 0$ for every $x \in \overline{C}$. The equilibrium problem, $EP(f, \overline{C})$ in short, consists to find a point $\overline{x} \in \overline{C}$ such that

$$EP(f, \overline{C}) \quad f(\overline{x}, y) \geq 0 \quad \forall y \in \overline{C}. \tag{7}$$

The solution set of the $EP(f, \overline{C})$, is denoted by $S(f, \overline{C})$. Next, we give the following assumptions on the equilibrium bifunction.

(H1) $f(., y) : \overline{C} \to \mathbb{R}$ is upper semicontinuous for all $y \in \overline{C}$.

(H2) $f(x, .) : \overline{C} \to \mathbb{R}$ is convex, for all $x \in \overline{C}$.

(H3) $f(., .)$ is quasimonotone.

Remark 2. Observe that assumptions **(H1)** and **(H2)** are standard for the study of equilibrium problems, see Remark 3.1 of [7] for a justification of each assumption. We impose the assumption **(H3)** because monotonicty assumption on $f(., .)$ turned out to be too restrictive for many applied problems, especially in Economics and Operation Research.

Inexact Algorithm

Initialization: Let $\{\lambda_k\}$ be a sequence of positive parameters and a starting point:

$$x^0 \in C. \tag{8}$$

Main Steps: For $k = 1, 2, \ldots,$ and $x^{k-1} \in C$, find $x^k \in C$ and $g^k \in \partial_2 f(x^k, x^k)$ such that

$$g^k + \lambda_k \nabla_1 d(x^k, x^{k-1}) = e^k, \tag{9}$$

where d is a proximal distance such that $(d, H) \in \mathcal{F}_+(\bar{C})$ and e^k is an approximation error which satisfies the following conditions:

$$\frac{\|e^k\|}{\lambda_k} \leq \eta_k \sqrt{H(x^k, x^{k-1})} \tag{10}$$

$$\sum_{k=1}^{+\infty} \eta_k < +\infty \tag{11}$$

Stop Criterion: If $x^k = x^{k-1}$ or $e^k \in \partial_2 f(x^k, x^k)$, then finish. Otherwise, to do $k - 1 \leftarrow k$ and return to Main Steps.

Observe that the error e^k is not prescribed before the finding of x^k. We impose the following additional assumption:

(H4) For each $k \in \mathbb{N}$, there exist x^k and g^k satisfying (9).

We are interested in analyzing the iterations when $x^k \neq x^{k-1}$ for each $k = 1, 2, \ldots$ because, otherwise, we obtain $g^k = e^k \in \partial_2 f(x^k, x^k)$ and therefore the algorithm finishes.

Now we define the following particular solution set of $S(f, \overline{C})$ which has been introduced in Mallma et al. [7]:

$$S^*(f, \overline{C}) = \{x \in S(f, \overline{C}) : \; f(x, w) > 0, \; \forall w \in C\}. \tag{12}$$

We will use the following assumption.

(H5) $S^*(f, \overline{C}) \neq \emptyset$.

Proposition 1. *Under the assumptions* (**H2**), (**H3**), (**H4**), (**H5**) *and* $(d, H) \in \mathcal{F}(\overline{\mathcal{C}})$*, we have*

$$H(\overline{x}, x^k) \leq H(\overline{x}, x^{k-1}) - \gamma H(x^k, x^{k-1}) - \frac{1}{\lambda_k} \left\langle e^k, \overline{x} - x^k \right\rangle \quad \forall \overline{x} \in S^*(f, \overline{C}). \tag{13}$$

Proof. Given that $\overline{x} \in S^*(f, \overline{C})$, then $f(\overline{x}, w) > 0$, for all $w \in C$, and as $x^k \in C$ (by assumption (**H4**)), we obtain $f(\overline{x}, x^k) > 0$. Then, as f is quasimonotone, then $f(x^k, \overline{x}) \leq 0$. Due to $g^k \in \partial_2 f(x^k, x^k)$, from (**H2**) and from Definition 1, we have

$$\left\langle g^k, \overline{x} - x^k \right\rangle \leq f(x^k, \overline{x}) \leq 0. \tag{14}$$

Replacing (9) in the previous expression and making use of Definition 3 (**Iii**) we obtain the result. □

We introduce the following extra condition on the induced proximal distance:

(Iviii) There exists $\theta > 0$ such that: $\|x-y\|^2 \leq \theta H(x,y)$, for all $x \in \overline{C}$ and for all $y \in C$.

Remark 3. Some examples of proximal distances which satisfy the above condition are the following:

1. $d(x,y) := \sum_{j=1}^{n} x_j - y_j - y_j \ln \frac{x_j}{y_j} + (\sigma/2)\|x-y\|^2$, with $\theta = \frac{2}{\sigma}$ and
$$H(x,y) = \sum_{j=1}^{n} x_j \ln\left(\frac{x_j}{y_j}\right) + y_j - x_j + \frac{\sigma}{2}||x-y||^2.$$
2. Let $\varphi : \mathbb{R} \to \mathbb{R}\cup\{+\infty\}$ be a closed proper convex function such that $dom\varphi \subset \mathbb{R}_+$ and $dom\partial\varphi = \mathbb{R}_{++}$. We suppose in addition that φ is $C^2(\mathbb{R}_{++})$, strictly convex, and nonnegative on $\mathbb{R}_{++}$ with $\varphi(1) = \varphi'(1) = 0$. We denote by Φ the class of such kernels and by
$$\bar{\Phi} = \left\{\varphi \in \Phi :\ \varphi''(1)\left(1 - \frac{1}{t}\right) \leq \varphi'(t) \leq \varphi''(1)(t-1) \quad \forall t > 0\right\}$$
the subclass of these kernels.

   Let $\varphi(t) = \mu p(t) + \frac{\nu}{2}(t-1)^2$ with $\nu \geq \mu p''(1) > 0,\ p \in \bar{\Phi}$ and let the associated proximal distance be defined by
$$d_\varphi(x,y) = \sum_{j=1}^{n} y_j^2 \varphi\left(\frac{x_j}{y_j}\right).$$
   The use of φ-divergence proximal distances is particularly suitable for handling polyhedral constraints. Let $C = \{x \in \mathbb{R}^n : Ax < b\}$, where A is an (m,n) matrix of full rank m $(m \geq n)$. Particularly important cases include $C = \mathbb{R}^n_{++}$ or $C = \{x \in \mathbb{R}^n_{++} : a_i < x_i < b_i\ \forall i = 1,\dots,n\}$, with $a_i, b_i \in \mathbb{R}$. In [16], example (c) of the appendix section, was showed that for $H(x,y) = \bar{\eta}\|x-y\|^2$ with $\bar{\eta} = 2^{-1}(\nu + \mu p''(1))$, we have $(d_\varphi, H) \in \mathcal{F}_+(\mathbb{R}^n_+)$.

Proposition 2. *Let $(d,H) \in \mathcal{F}(\overline{C})$ and suppose that the assumptions* (**H2**) $-$ (**H5**) *are satisfied. If the proximal distance $H(.,.)$ satisfies the additional condition **(Iviii)**, then*

i) there exists an integer $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ and for all $\overline{x} \in S^*(f,\overline{C})$, we have
$$H(\overline{x},x^k) \leq \left(1 + \frac{\theta\eta_k}{1-\theta\eta_k}\right) H(\overline{x},x^{k-1}) + \left(\frac{\eta_k}{4} - \gamma\right) H(x^k,x^{k-1}); \qquad (15)$$
ii) $\{H(\overline{x},x^k)\}$ converges for all $\overline{x} \in S^*(f,\overline{C})$;

iii) $\{x^k\}$ *is bounded;*
iv) $\lim_{k\to+\infty} H(x^k, x^{k-1}) = 0$.

Proof. **i)** Let $\overline{x} \in S^*(f, \overline{C})$, then

$$0 \leq \left\| \frac{e^k}{\sqrt{2\lambda_k \eta_k}} + \sqrt{2\lambda_k \eta_k}(\overline{x} - x^k) \right\|^2 = \frac{\|e^k\|^2}{2\lambda_k \eta_k} + 2\lambda_k \eta_k \|\overline{x} - x^k\|^2 + 2\langle e^k, \overline{x} - x^k \rangle,$$

thus,

$$-\frac{1}{\lambda_k}\langle e^k, \overline{x} - x^k \rangle \leq \frac{\|e^k\|^2}{4\lambda_k^2 \eta_k} + \eta_k \|\overline{x} - x^k\|^2.$$

Replacing the previous expression in (13) we have

$$H(\overline{x}, x^k) \leq H(\overline{x}, x^{k-1}) - \gamma H(x^k, x^{k-1}) + \frac{\|e^k\|^2}{4\lambda_k^2 \eta_k} + \eta_k \|\overline{x} - x^k\|^2.$$

Taking into account the hypothesis (10) and the condition **(Iviii)**, we have

$$H(\overline{x}, x^k) \leq H(\overline{x}, x^{k-1}) - \gamma H(x^k, x^{k-1}) + \frac{\eta_k}{4} H(x^k, x^{k-1}) + \theta \eta_k H(\overline{x}, x^k),$$

thus,

$$(1 - \theta\eta_k) H(\overline{x}, x^k) \leq H(\overline{x}, x^{k-1}) + \left(\frac{\eta_k}{4} - \gamma\right) H(x^k, x^{k-1}).$$

As $\eta_k \to 0^+$ (this is true from (11)) and $\theta > 0$, then there exists $k_0 \geq 0$ such that $0 < 1 - \theta\eta_k \leq 1$ and $\frac{\eta_k}{4} - \gamma < 0$, for all $k \geq k_0$. So applying this fact in the previous expression we have

$$H(\overline{x}, x^k) \leq \left(\frac{1}{1 - \theta\eta_k}\right) H(\overline{x}, x^{k-1}) + \left(\frac{1}{1 - \theta\eta_k}\right) \left(\frac{\eta_k}{4} - \gamma\right) H(x^k, x^{k-1}).$$

As $1 - \theta\eta_k \leq 1$, then the previous expression becomes

$$H(\overline{x}, x^k) \leq \left(1 + \frac{\theta\eta_k}{1 - \theta\eta_k}\right) H(\overline{x}, x^{k-1}) + \left(\frac{\eta_k}{4} - \gamma\right) H(x^k, x^{k-1}).$$

ii) From (15), it is clear that

$$H(\overline{x}, x^k) \leq \left(1 + \frac{\theta\eta_k}{1 - \theta\eta_k}\right) H(\overline{x}, x^{k-1}), \qquad \forall k \geq k_0. \tag{16}$$

As $\eta_k \to 0^+$ and $\theta > 0$, then for all $0 < \epsilon < 1$ there exists $\widetilde{k}_0 \in \mathbb{N}$, such that $\theta\eta_k < \epsilon$, for all $k \geq \widetilde{k}_0$, then $1 - \epsilon < 1 - \theta\eta_k \leq 1$. So

$$\frac{\theta\eta_k}{1 - \theta\eta_k} < \frac{\theta\eta_k}{1 - \epsilon}, \qquad \forall\, k \geq \widetilde{k}_0.$$

Applying summations, and taking into account (11), we have

$$\sum_{k=1}^{+\infty} \frac{\theta\eta_k}{1 - \theta\eta_k} < +\infty. \tag{17}$$

Finally, taking $v_{k+1} = H(\overline{x}, x^k)$, $v_k = H(\overline{x}, x^{k-1})$, $\gamma_k = \frac{\theta\eta_k}{1-\theta\eta_k}$ and $\beta_k = 0$ in Lemma 1 and considering that $\sum_{k=1}^{+\infty} \gamma_k < +\infty$ we obtain that the sequence $\{H(\overline{x}, x^k\}$ converges.
iii) It is immediate from **(ii)** and Definition 3-(Iiv).
iv) It is immediate from **(i)** and **(ii)**. □

Theorem 1. *Let* $(d, H) \in \mathcal{F}_+(\overline{\mathcal{C}})$ *and suppose that the assumptions* $(\mathbf{H1}) - (\mathbf{H5})$, ***(Iviii)*** *are satisfied and* $0 < \lambda_k < \overline{\lambda}$*, then*

i) $\{x^k\}$ *converges weakly to an element of* $S(f, \overline{C})$*, that is,* $Acc(x^k) \neq \emptyset$ *and every element of* $Acc(x^k)$ *is a point of* $S(f, \overline{C})$.
ii) If an accumulation point $\bar{x}$ *belongs to* $S^*(f, \overline{C})$ *then all the sequence* $\{x^k\}$ *converges to* $\bar{x}$.

Proof. (i). From Propositions 2, we have that $\{x^k\}$ is bounded, so there exist a subsequence $\{x^{k_j}\} \subseteq \{x^k\}$ and a point x^* such that $x^{k_j} \to x^*$. Define $L := \{k_1, k_2, ..., k_j, ...\}$, then $\{x^l\}_{l \in L} \to x^*$. We will prove that $x^* \in S(f, \overline{C})$.
From (9) we have that $\forall\, l \in L$ and $\forall\, x \in \overline{C}$:

$$f(x^l, x) \geq \left\langle g^l, x - x^l \right\rangle = \left\langle e^l, x - x^l \right\rangle - \lambda_l \left\langle \nabla_1 d(x^l, x^{l-1}), x - x^l \right\rangle.$$

Using Definition 3-**(Iii)** in the above equality, we obtain

$$\left\langle g^l, x - x^l \right\rangle \geq \left\langle e^l, x - x^l \right\rangle + \lambda_l \left[H(x, x^l) - H(x, x^{l-1}) + \gamma H(x^l, x^{l-1}) \right]. \quad (18)$$

Observe that, from (10) and due that $\{\lambda_k\}$ and $\{x^k\}$ are bounded, $\{\eta_k\}$ and $\{H(x^k, x^{k-1})\}$ converge to zero, then

$$\lim_{l \to \infty} \left\langle e^l, x - x^l \right\rangle = 0. \quad (19)$$

Fix $x \in \overline{C}$, we analyze two cases:

a) If $\{H(x, x^l)\}$ converges, then from Proposition 2-(**iv**), and the fact that $\{\lambda_l\}$ is bounded, we have $\lambda_l \left[H(x, x^l) - H(x, x^{l-1}) + \gamma H(x^l, x^{l-1}) \right] \to 0$. Applying this result and (19) in (18) and from assumption (**H1**) we obtain

$$f(x^*, x) \geq \limsup_{j \to \infty} f(x^l, x) \geq \limsup_{l \to \infty} \left\langle u^l, x - x^l \right\rangle \geq 0.$$

b) If $\{H(x, x^l)\}$ is not convergent, then the sequence is not monotonically decreasing and so there are infinite $l \in L$ such that $H(x, x^l) \geq H(x, x^{l-1})$. Let $\{l_j\} \subset L$, for all $j \in \mathbb{N}$, such that $H(x, x^{l_j}) \geq H(x, x^{l_j - 1})$, then

$$H(x, x^{l_j}) - H(x, x^{l_j-1}) + \gamma H(x^{l_j}, x^{l_j-1}) \geq \gamma H(x^{l_j}, x^{l_j-1}).$$

Taking into account this last result, Proposition 2-(**iv**) and (19), in (18) we have:

$$\limsup_{j \to \infty} \left\langle g^{l_j}, x - x^{l_j} \right\rangle \geq \limsup_{j \to \infty} \lambda_{l_j} \left[H(x, x^{l_j}) - H(x, x^{l_j-1}) + \gamma H(x^{l_j}, x^{l_j-1}) \right] \geq 0,$$

and from assumption (**H1**),

$$f(x^*, x) \geq \limsup_{j \to \infty} f(x^l, x) \geq \limsup_{j \to \infty} \left\langle u^{l_j}, x - x^{l_j} \right\rangle \geq 0.$$

(**ii**). Let $\bar{x}$ such that $x^{k_l} \to \overline{x}$ and $\bar{x} \in S^*(f, \overline{C})$. Then, from Definition 3 **(Ivii)**, $H(\overline{x}, x^{k_l}) \to 0$. Remember that from Proposition 2 (**ii**) we have that $\{H(\overline{x}, x^k)\}$ is convergent and as $H(\overline{x}, x^{k_l}) \to 0$, we obtain that $H(\overline{x}, x^{k_j}) \to 0$; so applying the Definition 3 **(Ivi)** we obtain that $x^{k_j} \to \overline{x}$, and due to the uniqueness of the limit we have $x^* = \overline{x}$. Thus $\{x^k\}$ converges to x^*. □

4 Rate of Convergence

In this section we prove the linear or superlinear rate of convergence of the inexact algorithm. For that, we consider the following additional assumption:
(H6) For $\overline{x} \in S^*(f, \overline{C})$ such that $x^k \to \overline{x}$, there exist $\delta = \delta(\overline{x}) > 0$ and $\tau_k = \tau_k(\overline{x}) > 0$, such that for all $w \in B(0, \delta) \subset \mathbb{R}^n$ and for all x^k with $w \in \partial_2 f(x^k, x^k)$, we have

$$H(\overline{x}, x^k) \leq \tau_k \|w\|^2. \tag{20}$$

Another assumption that we also assume for the proximal distance $(d, H) \in \mathcal{F}_+(\overline{C})$ is the following:
(H7) For all $u \in C$, the function $\bigtriangledown_1 d(., u)$ satisfies the following condition: there exists $L > 0$ such that

$$\| \bigtriangledown_1 d(x, u) - \bigtriangledown_1 d(y, u)\| \leq L\|x - y\|, \qquad \forall x, y \in C.$$

Lemma 2. *Let $(d, H) \in \mathcal{F}_+(\overline{C})$ and suppose that assumptions **(H1)**–**(H7)** and condition **(Iviii)** are satisfied and $0 < \lambda_k < \overline{\lambda}$. Then*

i) there exists $\widetilde{k} \in \mathbb{N}$ such that

$$\|g^k\| < \delta, \quad \forall\, k \geq \widetilde{k}, \tag{21}$$

where g^k is given by (9);
ii) it holds that

$$H(\overline{x}, x^k) \leq \tau_k \lambda_k^2 (\eta_k + L\sqrt{\theta})^2 H(x^k, x^{k-1}), \quad \forall\, k \geq \widetilde{k}. \tag{22}$$

Proof. **i)** Let $\overline{x} = \lim_{k\to+\infty} x^k$, such that $\overline{x} \in S^*(f, \overline{C})$, and thus from assumption (**H7**), there exists $L > 0$ such that

$$\| \bigtriangledown_1 d(x, x^{k-1}) - \bigtriangledown_1 d(y, x^{k-1})\| \leq L\|x - y\|, \qquad \forall x, y \in C.$$

From the above inequality we have

$$\| \bigtriangledown_1 d(x^k, x^{k-1})\| = \| \bigtriangledown_1 d(x^k, x^{k-1}) - \bigtriangledown_1 d(x^{k-1}, x^{k-1})\| \leq L\|x^k - x^{k-1}\|. \tag{23}$$

From (9) we obtain

$$\|g^k\| = \|e^k - \lambda_k \bigtriangledown_1 d(x^k, x^{k-1})\| \leq \|e^k\| + \lambda_k \| \bigtriangledown_1 d(x^k, x^{k-1})\|, \tag{24}$$

so, taking into account (10), (23), the condition (**Iviii**), and the fact that $\lambda_k \leq \overline{\lambda}$, we have that the inequality (24) implies

$$\|g^k\| \leq \lambda_k \eta_k \sqrt{H(x^k, x^{k-1})} + \lambda_k L\sqrt{\theta}\sqrt{H(x^k, x^{k-1})}$$

$$= \lambda_k(\eta_k + L\sqrt{\theta})\sqrt{H(x^k, x^{k-1})} \tag{25}$$

$$\leq \overline{\lambda}(\eta_k + L\sqrt{\theta})\sqrt{H(x^k, x^{k-1})}. \tag{26}$$

Since $\eta_k \to 0$ and $H(x^k, x^{k-1}) \to 0$ (see Proposition 2-(**iv**)), taking $\delta > 0$, there exists $\widetilde{k} \in \mathbb{N}$ such that $\|g^k\| < \delta$ for all $k \geq \widetilde{k}$.

ii) In (20) taking $w = g^k$ for all $k \geq \widetilde{k}$, we have

$$H(\overline{x}, x^k) \leq \tau_k \|g^k\|^2. \tag{27}$$

Therefore, the relation (22) follows from the last inequality combined with (25). □

Theorem 2. *Let $(d, H) \in \mathcal{F}_+(\overline{C})$ and suppose that assumptions **(H1)**-**(H7)** and condition **(Iviii)** are satisfied and $0 < \lambda_k < \overline{\lambda}$. Then,*

$$H(\overline{x}, x^k) \leq r_k H(\overline{x}, x^{k-1}), \tag{28}$$

for k sufficiently large, where

$$r_k = \left(\frac{4\tau_k(\eta_k + L\sqrt{\theta})^2}{4\tau_k(\eta_k + L\sqrt{\theta})^2 + \frac{4\gamma - \eta_k}{\overline{\lambda}^2}}\right)\left(\frac{1}{1 - \theta\eta_k}\right).$$

1. *If $\tau_k = \tau > 0$ then, $\{x^k\}$ converges H-linearly to $\overline{x} \in SOL(T, \overline{C})$.*
2. *If $\{\tau_k\}$ converges to zero then, $\{x^k\}$ converges H-superlinearly to $\overline{x} \in SOL(T, \overline{C})$.*
3. *If $\tau_k = \tau > 0$ and $\lambda_k \searrow 0$ then, $\{x^k\}$ converges H-superlinearly to $\overline{x} \in SOL(T, \overline{C})$.*

Proof. Let $\overline{x} \in S^*(f, \overline{C})$ be the limit point of the sequence $\{x^k\}$ and $g^k \in \partial_2 f(x^k, x^k)$ given by (9). Due to the relationship (21) we have to $\|g^k\| < \delta$ for all $k \geq \widetilde{k}$. So $g^k \in B(0, \delta)$, for all $k \geq \widetilde{k}$.
Considering the inequality (22) in (15) for all $k \geq \max\{k_0, \widetilde{k}\}$, it follows that

$$H(\overline{x}, x^k) \leq \left(1 + \frac{\theta\eta_k}{1 - \theta\eta_k}\right) H(\overline{x}, x^{k-1}) - \left(\gamma - \frac{\eta_k}{4}\right)\left(\frac{1}{\tau_k \lambda_k^2(\eta_k + L\sqrt{\theta})^2}\right) H(\overline{x}, x^k),$$

Thus, we obtain for all $k \geq \max\{k_0, \widetilde{k}\}$:

$$\left(1 + \frac{4\gamma - \eta_k}{4\tau_k \lambda_k^2(\eta_k + L\sqrt{\theta})^2}\right) H(\overline{x}, x^k) \leq \left(\frac{1}{1 - \theta\eta_k}\right) H(\overline{x}, x^{k-1}).$$

As $\tau_k > 0$ and also $(4\gamma - \eta_k) > 0$, then for all $k \geq \max\{k_0, \widetilde{k}\}$ we have

$$H(\overline{x}, x^k) \leq \beta_k H(\overline{x}, x^{k-1}), \tag{29}$$

where

$$\beta_k = \left(\frac{4\tau_k(\eta_k + L\sqrt{\theta})^2}{4\tau_k(\eta_k + L\sqrt{\theta})^2 + \frac{4\gamma - \eta_k}{\lambda_k^2}} \right) \left(\frac{1}{1 - \theta\eta_k} \right). \tag{30}$$

Since that $\lambda_k \leq \overline{\lambda}$ for all $k \in \mathbb{N}$, we obtain

$$\beta_k \leq r_k, \tag{31}$$

where

$$r_k = \left(\frac{4\tau_k(\eta_k + L\sqrt{\theta})^2}{4\tau_k(\eta_k + L\sqrt{\theta})^2 + \frac{4\gamma - \eta_k}{\overline{\lambda}^2}} \right) \left(\frac{1}{1 - \theta\eta_k} \right).$$

Thus we obtain (28).

1. Let $\tau_k = \tau > 0$, then taking into account that $\eta_k \to 0$, then

$$r_k \to \left(\frac{4\tau L^2 \theta}{4\tau L^2\theta + \frac{4\gamma}{\overline{\lambda}^2}} \right).$$

   Thus, there exists a positive number $k_1 \in \mathbb{N}$ with $k \geq k_1$, such that

$$\beta_k \leq r_k < \frac{1}{2}\left(1 + \frac{4\tau L^2 \theta}{4\tau L^2\theta + \frac{4\gamma}{\overline{\lambda}^2}} \right) < 1 \quad \forall\, k \geq k_1.$$

   Then, in (29) we have for all $k \geq \max\{k_0, \widetilde{k}, k_1\}$:

$$H(\overline{x}, x^k) \leq \bar{\theta} H(\overline{x}, x^{k-1}),$$

   where

$$\bar{\theta} = \left(\frac{4\tau L^2 \theta}{4\tau L^2\theta + \frac{4\gamma}{\overline{\lambda}^2}} \right).$$

   Thus, the sequence $\{x^k\}$ converges $H-$linearly to $\overline{x}$.
2. If $\{\tau_k\}$ converges to zero then from (28) we have that $\{r_k\}$ converges to zero and thus we obtain that the sequence $\{x^k\}$ converges $H-$superlinearly to $\overline{x}$.
3. Let $\tau_k = \tau > 0$, we have from (29) and (30)

$$H(\overline{x}, x^k) \leq \beta_k H(\overline{x}, x^{k-1}), \tag{32}$$

   where

$$\beta_k = \left(\frac{4\tau(\eta_k + L\sqrt{\theta})^2}{4\tau(\eta_k + L\sqrt{\theta})^2 + \frac{4\gamma - \eta_k}{\lambda_k^2}} \right) \left(\frac{1}{1 - \theta\eta_k} \right). \tag{33}$$

   As $\lambda_k \searrow 0$ and $\eta_k \to 0$, then sequence $\{x^k\}$ converges $H-$superlinearly to $\overline{x}$. □

Corollary 1. *Under the same assumptions of the previous theorem and suppose that the condition*

$$H(x,y) = \theta\|x-y\|^2 \tag{34}$$

for some $\theta > 0$, is satisfied, then

1. *If $\tau_k = \tau > 0$, then $\{x^k\}$ converges linearly to $\overline{x} \in S^*(f, \overline{C})$*
2. *If $\{\tau_k\}$ converges to zero then $\{x^k\}$ converges superlinearly to $\overline{x} \in S^*(f, \overline{C})$*
3. *If $\tau_k = \tau > 0$, and $\lambda_k \searrow 0$ then $\{x^k\}$ converges superlinearly to $\overline{x} \in S^*(f, \overline{C})$.*

Remark 4. A class of proximal distances which satisfies the above condition (34) is the proximal distance with second order homogeneous distances, see Remark 3.

References

1. Anh, P.N.: A hybrid extragradient method extended to fixed point problems and equilibrium problems. Optimization **62**, 271–283 (2013)
2. Auslender, A., Teboulle, M.: Interior gradient and proximal methods for convex and conic optimization. SIAM J. Optim. **16**, 697–725 (2006)
3. Blum, E., Oettli, W.: From optimization and variational inequalities to equilibrium problems. Math. Student **63**, 123–145 (1994)
4. Iusem, A.N., Sosa, W.: New existence results for equilibrium problems. Nonlinear Anal. **52**, 621–635 (2003)
5. Iusem, A.N., Sosa, W.: On the proximal point method for equilibrium problems in Hilbert spaces. Optimization **59**, 1259–1274 (2010)
6. Khatibzadeh, H., Mohebbi, V., Ranjbar, S.: Convergence analysis of the proximal point algorithm for pseudo-monotone equilibrium problems. Optim. Methods Software **30**, 1146–1163 (2015)
7. Mallma Ramirez, L., Papa Quiroz, E.A., Oliveira, P.R.: An inexact proximal method with proximal distances for quasimonotone equilibrium problems. J. Oper. Res. SOC China **246**(3), 721–729 (2017)
8. Moudafi, A.: On the convergence of splitting proximal methods for equilibrium problems in Hilbert spaces. J. Math. Anal. Appl. **359**, 508–513 (2009)
9. Nguyen, T.T.V., Strodiot, J.J., Nguyen, V.H.: The interior proximal extragradient method for solving equilibrium problems. J. Glob. Optim. **44**, 175–192 (2009)
10. Papa Quiroz, E.A., Oliveira, P.R.: An extension of proximal methods for quasiconvex minimization on the nonnegative orthant. Eur. J. Oper. Res. **216**, 26–32 (2012)
11. Papa Quiroz, E.A., Mallma Ramirez, L., Oliveira. P.R.: An inexact proximal method for quasiconvex minimization. Eur. J. Oper. Res. **246**, 721–729 (2015)
12. Papa Quiroz, E.A., Mallma Ramirez, L., Oliveira, P.R.: An inexact algorithm with proximal distances for variational inequalities. RAIRO Oper. Res. **52**(1), 159–176 (2018)
13. Papa Quiroz, E.A., Cruzado, S.: An inexact scalarization proximal point method for multiobjective quasiconvex minimization. Ann. Oper. Res., 1–26 (2020). https://doi.org/10.1007/s10479-020-03622-8
14. Pedregal, P.: Optimal control. In: Introduction to Optimization. TAM, vol. 46, pp. 195–236. Springer, New York (2004). https://doi.org/10.1007/0-387-21680-4_6
15. Quoc, T.D., Muu, L.D.: Iterative methods for solving monotone equilibrium problems via dual gap functions. Comput. Optim. Appl. **51**, 709–728 (2012)

16. Sarmiento, O., Papa Quiroz, E.A., Oliveira, P.R.: A proximal multiplier method for separable convex minimization. Optimization **65**(2), 501–537 (2016)
17. Villacorta, K.D., Oliveira, P.R.: An interior proximal method in vector optimization. Eur. J. Oper. Res. **214**, 485–492 (2011)

Beyond Pointwise Submodularity: Non-monotone Adaptive Submodular Maximization Subject to Knapsack and k-System Constraints

Shaojie Tang(✉)

Naveen Jindal School of Management, University of Texas at Dallas, Richardson, USA
shaojie.tang@utdallas.edu

Abstract. Although the knapsack-constrained and k-system-constrained non-monotone adaptive submodular maximization have been well studied in the literature, it has only been settled given the additional assumption of pointwise submodularity. In this paper, we remove the common assumption on pointwise submodularity and propose the first approximation solutions for both knapsack and k-system constrained adaptive submodular maximization problems. Inspired by two recent studies on non-monotone adaptive submodular maximization, we develop a sampling-based randomized algorithm that achieves a $\frac{1}{10}$ approximation for the case of a knapsack constraint and that achieves a $\frac{1}{2k+4}$ approximation ratio for the case of a k-system constraint.

Keywords: Adaptive submodularity · Approximation algorithms · Non-monotonicity

1 Introduction

In [5], they extend the study of submodular maximization from the non-adaptive setting [8] to the adaptive setting. They introduce the notions of adaptive monotonicity and submodularity, and show that a simple adaptive greedy policy achieves a $1-1/e$ approximation ratio if the utility function is adaptive submodular and adaptive monotone. Although there have been numerous research studies on adaptive submodular maximization under different settings [2,4,9,11,13], most of them assume adaptive monotonicity. For the case of maximizing a non-monotone adaptive submodular function subject to a cardinality constraint, [10] develops the first constant approximation solution. For the case of maximizing a non-monotone adaptive submodular and pointwise submodular function, [1,3] develop effective solutions for the case of a knapsack and a k-system constraints, respectively. Note that adaptive submodularity does not imply pointwise submodularity and vice versa [5,7], and this raises the following question: Does there exist an approximation solution for maximizing a knapsack-constrained or a k-system constrained non-monotone adaptive submodular function without resorting to pointwise submodularity?

H. A. Le Thi et al. (Eds.): MCO 2021, LNNS 363, pp. 16–27, 2022.
https://doi.org/10.1007/978-3-030-92666-3_2

Table 1. Approximation for non-monotone adaptive submodular function maximization

Source	Ratio	Constraint	Require pointwise submodularity?
[6]	$\frac{1}{e}$	Cardinality constraint	Yes
[1]	$\frac{1}{9}$	Knapsack constraint	Yes
[3]	$\frac{1}{k+2\sqrt{k+1}+2}$	k-system constraint	Yes
[10]	$\frac{1}{e}$	Cardinality constraint	No
This work	$\frac{1}{10}$	Knapsack constraint	No
This work	$\frac{1}{2k+4}$	k-system constraint	No

In this paper, we answer the above question affirmatively by proposing the first approximation solutions for both knapsack or a k-system constraints. Note that many practical constraints, including cardinality, matroid, intersection of k matroids, k-matchoid and k-extendible constraints, all belong to the family of k-system constraints. In particular, we develop a $\frac{1}{10}$ approximate solution for maximizing a knapsack-constrained non-monotone adaptive submodular function. Technically speaking, our design is an extension of the classic modified density greedy algorithm [1,12]. In particular, their design is required to maintain two candidate policies, i.e., one is to choose a best singleton and the other one is to choose items in a density-greedy manner, while our design maintains three candidate policies in order to drop the common assumption about pointwise submodularity. For the case of a k-system constraint, we are inspired by the sampling based policy proposed in [3] and develop a similar policy that achieves a $\frac{1}{2k+4}$ approximation ratio without resorting to pointwise submodularity. We list the performance bounds of the closely related studies in Table 1.

2 Preliminaries

We first introduce some important notations. In the rest of this paper, we use $[m]$ to denote the set $\{0, 1, 2, \cdots, m\}$, and we use $|S|$ to denote the cardinality of a set S.

2.1 Items and States

We consider a set E of n items, where each item $e \in E$ is in a particular state from O. We use $\phi : E \to O$ to denote a *realization*, where $\phi(e)$ represents the state of $e \in E$. Let $\Phi = \{\Phi(e) \mid e \in E\}$ denote a random realization, where $\Phi(e) \in O$ is a random realization of the state of $e \in E$. The state of each item is unknown initially, one must pick an item $e \in E$ before observing the value of $\Phi(e)$. We assume there is a known prior probability distribution $p(\phi) = \{\Pr[\Phi = \phi] : \phi \in U\}$ over realizations U. For any subset of items $S \subseteq E$, we use $\psi : S \to O$ to denote a *partial realization* and $\mathrm{dom}(\psi) = S$ is called the *domain* of ψ. Consider any realization ϕ and any partial realization ψ, we say that ψ is consistent with ϕ, i.e., $\psi \prec \phi$, if they are equal everywhere in the domain of ψ. We say that ψ is a *subrealization* of ψ', i.e., $\psi \subseteq \psi'$, if $\mathrm{dom}(\psi) \subseteq \mathrm{dom}(\psi')$ and they are equal everywhere in $\mathrm{dom}(\psi)$. Moreover, we use $p(\phi \mid \psi)$ to denote the conditional distribution over realizations conditioned on a partial realization ψ: $p(\phi \mid \psi) = \Pr[\Phi = \phi \mid \psi \prec \Phi]$. There is a non-negative utility function f that is defined over items and their states: $f : 2^E \times O^E \rightarrow \mathbb{R}_{\geq 0}$.

2.2 Policies and Problem Formulation

A typical adaptive policy works as follows: select the first item and observe its state, then continue to select the next item based on the observations collected so far, and so on. After each selection, we observe some *partial realization* ψ of the states of some subset of E, for example, we are able to observe the partial realization of the states of those items which have been selected. Formally, any adaptive policy can be represented as a function π that maps a set of observations to a distribution $\mathcal{P}(E)$ of E: $\pi : 2^E \times O^E \rightarrow \mathcal{P}(E)$, specifying which item to pick next based on the current observation.

Definition 1 (Policy Concatenation). *Given two policies π and π', let $\pi@\pi'$ denote a policy that runs π first, and then runs π', ignoring the observation obtained from running π.*

Let the random variable $E(\pi, \phi)$ denote the subset of items selected by π under a realization ϕ. The expected utility $f_{avg}(\pi)$ of a policy π is

$$f_{avg}(\pi) = \mathbb{E}_{\Phi \sim p(\phi), \Pi} f(E(\pi, \Phi), \Phi) \tag{1}$$

where the expectation is taken over Φ with respect to $p(\phi)$ and the random output of π. For ease of presentation, let $f(e) = \mathbb{E}_{\Phi \sim p(\phi)} f(\{e\}, \Phi)$.

Definition 2 (Independence System). *Given a ground set E and a collection of sets $\mathcal{I} \subseteq 2^E$, the pair $(E, \mathcal{I})$ is an* independence system *if*

1. $\emptyset \in \mathcal{I}$;
2. $\mathcal{I}$, *which is called the* independent sets, *is downward-closed, that is, $A \in \mathcal{I}$ and $B \subseteq A$ implies that $B \in \mathcal{I}$.*

A set $B \in \mathcal{I}$ is called a *base* if $A \in \mathcal{I}$ and $B \subseteq A$ imply that $B = A$. A set $B \in \mathcal{I}$ is called a base of R if $B \subseteq R$ and B is a base of the independence system $(R, 2^R \cap \mathcal{I})$.

Definition 3 (k-System). *An independence system $(E, \mathcal{I})$ is a k-system for an integer $k \geq 1$ if for every set $R \subseteq E$, the ratio between the sizes of the largest and smallest bases of R is upper bounded by k.*

Let Ω denote the set of feasible policies and let $U^+ = \{\phi \in U \mid p(\phi) > 0\}$. For the case of knapsack constraint, define $\Omega = \{\pi | \forall \phi \in U^+, \sum_{e \in E(\pi,\phi)} c_e \leq b\}$ where c_e is the cost of e, which is fixed and pre-known, and b is the budget constraint. For the case of k-system constraint, define $\Omega = \{\pi | \forall \phi \in U^+, E(\pi, \phi) \in \mathcal{I}\}$ where $(E, \mathcal{I})$ is a k-system. Our goal is to find a feasible policy π^{opt} that maximizes the expected utility, i.e., $\pi^{opt} \in \arg\max_{\pi \in \Omega} f_{avg}(\pi)$.