Summer 8-2022

Inexact Fixed-Point Proximity Algorithms for Nonsmooth Convex Optimization

Jin Ren
Old Dominion University, jren007@odu.edu

INEXACT FIXED-POINT PROXIMITY ALGORITHMS FOR NONSMOOTH CONVEX OPTIMIZATION

by

Jin Ren
B.S. June 2016, Sun Yat-Sen University, China

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

COMPUTATIONAL AND APPLIED MATHEMATICS

OLD DOMINION UNIVERSITY
August 2022

Approved by

Yuesheng Xu (Director)

Yaohang Li (Member)

Ke Shi (Member)

Guohui Song (Member)

Lucia Tabacu (Member)

ABSTRACT

INEXACT FIXED-POINT PROXIMITY ALGORITHMS FOR NONSMOOTH CONVEX OPTIMIZATION

Jin Ren
Old Dominion University, 2022
Director · Dr. Yuesheng Xu

The aim of this dissertation is to develop efficient inexact fixed-point proximity algorithms with convergence guaranteed for nonsmooth convex optimization problems encountered in data science. Nonsmooth convex optimization is one of the core methodologies in data science to acquire knowledge from real-world data and has wide applications in various fields, including signal/image processing, machine learning and distributed computing. In particular, in the context of image reconstruction, compressed sensing and sparse machine learning, either the objective functions or the constraints of the modeling optimization problems are nondifferentiable. Hence, traditional methods such as the gradient descent method and the Newton method are not applicable since gradients of the objective functions or the constraints do not exist. Fixed-point proximity algorithms were developed via subdifferentials of the objective function to address the challenges. The theory of nonexpansive averaged operators was successfully employed in the existing analysis of exact/inexact fixed-point proximity algorithms for nonsmooth convex optimization. However, this framework has imposed restricted constraints on the algorithm formulation, which slows down the convergence and conceals relations between different algorithms.

In this work, we characterize the solutions of convex optimization as fixed-points of certain operators, and then adopt the matrix splitting technique to obtain a framework of fully implicit fixed-point proximity algorithms. This results in a new class of quasiaveraged operators, which extends the class of nonexpansive averaged operators. Such framework covers and generalizes most of the existing popular algorithms for nonsmooth convex optimization. To deal with the implicitness of this framework, we follow the inspiration of the Schur's lemma on the uniform boundedness of infinite matrices and propose a framework of inexact fixed-point iterations of quasiaveraged operators. This framework generalizes the inexact iterations of nonexpansive averaged operators. A combination of the frameworks of inexact fixed-point iterations and the implicit fixed-point proximity algorithms leads to the framework of inexact fixed-point proximity algorithms, which further extends existing methods for nonsmooth convex optimization. Numerical experiments on image deblurring problems demonstrate the advantages of inexact fixed-point proximity algorithms over existing explicit algorithms.

To my beloved family
with whom we once together live
with whom we are doomed to never meet

ACKNOWLEDGMENTS

This work is accomplished during the coronavirus pandemic. It is a global pandemic of COVID-19 caused by coronavirus SARS-CoV-2, and has caused more than 596 million cases and 6.45 million confirmed deaths as of August 2022 since 2019, making it one of the deadliest in history. Specifically, it has caused massive cases and deaths in USA and extremely strict entry policies in China, which brings tremendous challenges to exchange students in both countries.

During this hard time, I would like to express my sincere thanks and deepest gratitude to my supervisor Professor Yuesheng Xu. Since we first met in Sun-Yat Sen University in 2013 when I was a sophomore, he has been a constant source of inspiration, support and discipline in both academic study and daily life. For ten years he witnesses my growth and never hesitates to help me, even though I'm still far away from being a mature man. His adherence to classical academism and enthusiasm to discover unknowns will be kept as precious and valuable treasure in my future career and life. Without his help I could have never attended the university he graduated from and accomplished this very work.

I'm also sincerely grateful to my committee members, Professors Yaohang Li, Ke Shi, Guohui Song and Lucia Tabacu for their altruistic help in reviewing the dissertation and valuable suggestions. There are also special thanks to former/current department chairs, Professors Hideaki Kaneko and Gordon Melrose and former/current graduate program directors, Professors Raymond Cheng and Ruhai Zhou, for their careful and patient guidance and assistance during my postgraduate study.

My frankest and greatest appreciation belongs to my parents, Shumei Xie and Shaowu Ren. You bring the whole world to me and accompany me with selfless love and full faith. I could have been nowhere without your support, and during such a difficult period, your warm and supportive affection from the opposite side of the globe keeps encouraging and lightening me in the long dark night. Glory to thees.

Finally, I'm thankful to my colleagues and friends. Studying in different countries with distinct cultures is always challenging, and their help is certainly of paramount importance. Special thanks are extended to them with best wishes of bright future and happy life.

NOMENCLATURE

Sets

$\mathbb{N}$	§2.1	Set of non-negative integers
$\mathbb{N}^+$	§3.1	Set of positive integers
$\mathbb{N}^n$	§3.1	Set of positive integers $\{1, 2, \ldots, n\}$ for $n \in \mathbb{N}^+$
$\mathbb{R}$	§3.1	Set of real numbers
$\overline{\mathbb{R}}$	§3.1	Set of positive-extended real numbers
$\mathbb{R}^n$	§2.1	The standard n-dimensional Euclidean space
$\mathbb{R}^{n\times m}$	§1.1	Set of n-by-m real matrices
$\mathbb{P}^n_+$	§3.4	Set of strictly positive-definite n-by-n matrices
$\mathbb{S}^n$	§2.2	Set of symmetric positive-definite n-by-n matrices
$\mathbb{S}^n_+$	§3.2	Set of symmetric strictly positive-definite n-by-n matrices
GL^n	§3.4	Group of invertible n-by-n matrices
$\mathcal{B}(\mathcal{X}, \mathcal{Y})$	§3.1	Set of bounded linear operators from Banach space $\mathcal{X}$ to Banach space $\mathcal{Y}$
$\mathcal{B}(\mathcal{X})$	§3.1	Set of bounded linear operators on from Banach space $\mathcal{X}$ to itself
$\bigtimes_{i=1}^N \mathbb{R}^{n_i}$	§3.1	Hilbert direct sum of spaces $\{\mathbb{R}^{n_i}\}_{i=1}^N$
ℓ^p	§2.1	Sequence ℓ^p space with $1 \leq p \leq \infty$
$B_p(x; \rho)$	§3.2	Closed ball with center x and radius ρ in ℓ^p norm, $1 \leq p \leq \infty$
$\operatorname{argmin} f$	§3.1	Set of global minimizers of a function f
$\operatorname{Ran} f$	§2.4	Range of a function f
$\operatorname{Zer} f$	§3.2	Set of zeros of a function f
$\operatorname{Fix} f$	§2.1	Set of fixed points of a function f
$\operatorname{Gra} f$	§3.2	Graph of a function f
$\operatorname{Dom} f$	§3.2	Domain of a function f
$\Gamma_0(\mathbb{R}^n)$	§3.1	Set of proper, convex and lower semicontinuous functions mapping from $\mathbb{R}^n$ to $\overline{\mathbb{R}}$
$C^1_L(\mathbb{R}^n)$	§3.1	Set of differentiable functions mapping $\mathbb{R}^n$ to $\mathbb{R}$ with L-Lipschitz continuous gradient
$\operatorname{Ker} T$	§2.4	Kernel of an operator T

Linear Operators

$\langle \cdot, \cdot \rangle$	§3.2	Euclidean product in $\mathbb{R}^n$
$\| \cdot \|_p$	§3.1	ℓ^p vector norm for $1 \leq p \leq \infty$
$\| \cdot \|_{\mathsf{TV}}$	§3.1	TV norm
$\vert\!\vert\!\vert \cdot \vert\!\vert\!\vert$	§3.4	Induced operator norm
$L^\top$	§3.1	Adjoint of a bounded linear operator L
$L^\dagger$	§2.4	Moore–Penrose pseudoinverse of a bounded linear operator L
$\sqrt{L}$	§2.2	Square root of a matrix L in $\mathbb{S}^n$
I_n	§2.1	Identity operator in $\mathbb{R}^n$
Diag	§3.2	Create a diagonal matrix

Functions

$(\cdot)_+$	§2.4	Rectified linear unit activation function in $\mathbb{R}$
$f \circ g$	§3.2	Composition of functions f and g
$f \oplus g$	§3.1	Separable sum of functions f and g
$f \otimes g$	§3.2	Cartesian product of functions f and g
f^*	§3.2	Convex conjugate of a function f
∇_f	§2.1	Gradient operator of a function f
∂_f	§3.2	Subdifferential of a function f
prox_f	§3.2	Proximity operator of a function f
ι_C	§3.1	Indicator function of a set C
$f \,\square\, g$	§3.2	Infimal convolution of functions f and g
$f \,\boxdot\, g$	§3.2	Exact infimal convolution of functions f and g
$L \triangleright f$	§3.2	Infimal postcomposition of a function f and a linear mapping L
$L \,\underset{\cdot}{\triangleright}\, f$	§3.2	Exact infimal postcomposition of a function f and a linear mapping L

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This dissertation proposes efficient inexact fixed-point proximity algorithms for nonsmooth convex optimization in data science with convergence analysis. Specifically, we characterize the solutions of convex optimization as fixed-point of proximity operators and then adopt the matrix splitting technique to obtain a general framework of fully implicit fixed-point proximity algorithms with convergence analysis. To overcome the difficulty brought by the implicitness of this framework, we follow the inspiration of the Schur's lemma on the uniform boundedness of infinite matrices to propose a general framework of inexact fixed-point iterations. These two general frameworks cover and generalize existing analyses of exact/inexact fixed-point iterations for nonsmooth convex optimization. Finally we assemble these results to establish the framework of inexact fixed-point proximity algorithms. Numerical experiments on image deblurring problems show advantages of the proposed inexact fixed-point proximity algorithms over existing explicit algorithms.

1.1 NONSMOOTH CONVEX OPTIMIZATION

It is the core essence of the data science that learning information from observations gathered from the real world and then applying those knowledge back to the world for the sake of all well-being. The attempts to apply mathematical methods to extract information from data have been carried out since ancient times, of which the earliest studies in modern form might be traced back to the groundbreaking work *Philosophiæ Naturalis Principia Mathematica* by Sir Issac Newton in 1687 [76] and the works in following centuries, *e.g.*, [9, 43, 53, 54].

As human entered digital era after the profound invention and extensive usage of electronic computer, huge amount of data has been generated and gathered, challenging data scientists for developed and functional tools capable of extracting information with accuracy and efficiency. Since then, optimization, especially general nonsmooth convex optimization, has been proven to be a successful approach, and eventually permeated substantially into all branches of data science. Followings are some representative areas and optimization models.

Image Processing. The observation process of an image $z \in \mathbb{R}^{n\times m}$ can be generally modeled by $z = Kx + \varepsilon$, where $x \in \mathbb{R}^{n\times m}$ stands for the clear image, $K : \mathbb{R}^{n\times m} \to \mathbb{R}^{n\times m}$ is the linear measurement operator, and $\varepsilon \in \mathbb{R}^{n\times m}$ is some randomly-distributed additive noise. Here $\mathbb{R}^{n\times m}$ denotes the set of all real n-by-m matrices. Different models are applied according to different kinds of measurement and noise. Emblematic applications are, Rudin-Osher-Fatemi model [89] for restoring images polluted by Gaussian noise, L1-TV denoising model [21, 23, 77] for uniform impulse noise, L1-TV deblurring model [2, 27, 49, 74] and L2-TV deblurring model [8, 22, 78] for blurred images with corresponding kinds of noise, fast high-resolution image reconstruction models [64, 65, 66, 67] and their applications in medical imaging [44, 48, 50, 60, 69].

Machine Learning. Various optimization problems arising from machine learning always seek to minimize a loss function with regularization. The loss function models the expected cost or the degree of under-fitting with respect to training data, and the regularization term restricts the range of solutions in order to reduce ill-posedness of the problem and avoid over-fitting. Emblematic applications are, the well-known least absolute shrinkage and selection operator method [95], which is also known as basis pursuit problem in compressed sensing [16, 25, 35], the L1-regularized classification model [32, 58, 93, 99], the L1-regularized regression model [38, 57] and the reproducing kernel Banach/Hilbert spaces for machine learning [101, 102, 105, 110].

Distributed Computing. Single computer or processor is increasingly insufficient in big data analysis due to its limited computation capability. In distributed computing, a problem is divided into many tasks, each of which is solved by one or more processors communicating with each other via message passing [68, 94]. Emblematic applications are consensus problems [96, 97] and exchange problems [98].

1.2 FIXED-POINT PROXIMITY ALGORITHMS

As an algorithmic method to solve optimization problems, the gradient method invented by Augustin-Louis Cauchy [17] for a data fitting problem in astronomy, might serve as one of the most successful iterative methods with convergence in optimization. Historically, the first application of fixed-point iterations could be found in the work of solving initial value problem of ordinary differential equations by Charles Émile Picard [80]. Such iteration method is therefore named as *Picard iteration.* In this work Picard proved the convergence of the Picard iteration of a special class of integral operators, which was later extended to the class of contraction mappings, *i.e.*, the class of operators with Lipschitz constant $L < 1$, by Stefan Banach [4]. Then the class of nonexpansive operators, *i.e.*, the class of operators with Lipschitz constant $L = 1$, got noticed by mathematicians and finally resulted in a second extension of Picard iteration. For fixed-point problems of nonexpansive operators, Mark Aleksandrovich Krasnosel'skiĭ [52] developed a method of successive approximations based on the work of mean value method by William Robert Mann [70], which is therefore named as *KM iteration.* This work initiated the study of the class of *nonexpansive averaged operators.* Since then many problems in data science were modeled and studied under the framework of fixed-point problems with KM iterations; see *e.g.*, [28, 73, 82, 88, 103, 108, 109]. This framework also has profound applications in nonsmooth convex optimization.

There are plenty of practical algorithms for nonsmooth convex optimization, popular ones among which could be considered as, gradient descent method [17], proximal point method [71, 72], Douglas-Rachford splitting algorithm [37, 62], first-order primal-dual algorithm [19, 41], primal-dual hybrid gradient method [46], fixed-point proximity algorithm [56, 74], alternating direction method of multipliers [10, 40, 45], split Bregman iteration [47], linearized alternating direction method of multipliers [61, 106] and inexact Uzawa method [11, 111]. These methods were once proposed from different perspectives, *e.g.*, Fenchel-Rockafellar duality theory (first-order primal-dual algorithm, primal-dual hybrid gradient method, *etc.*) and augmented Lagrangian technique (Douglas-Rachford splitting algorithm, alternating direction method of multipliers, *etc.*). However they more or less rely on the notation of *proximity operators.* This class of operators was introduced early in [75, 86] and later was found out to be *firmly-nonexpansive*, *i.e.*, a special case of nonexpansive averaged operators. That was the very time the fixed-point theory for nonexpansive/nonexpansive averaged operators started being applied to this field [6, 30, 56, 74] and eventually forged such a prosperous family of algorithms for nonsmooth convex optimization. In particular, [56, 74] characterized the solutions of nonsmooth convex optimization problems into fixed-points of proximity equations, and then they adopted the *matrix splitting technique* to develop a class of explicit fixed-point proximity algorithms. This analysis has been proven to cover several popular methods, *e.g.*, proximal point method, Douglas-Rachford splitting algorithm, first-order primal-dual algorithm and alternating direction method of multipliers and so on.

However such analysis of fixed-point proximity algorithms highly depends on the framework of iterations of nonexpansive averaged operators. This framework has rather restrict constraint on the iteration formulation and slows down the convergence. With a close study of fixed-point proximity equations, we observe that a full application of matrix splitting technique will lead to a general framework of implicit fixed-point proximity algorithms. This framework covers most of the existing algorithms, but its convergence analysis falls out of the current studies of nonexpansive/nonexpansive averaged operators. This eventually results in the class of *quasi-nonexpansive/quasiaveraged operators* proposed in this work, which covers and generalizes the class of nonexpansive averaged operators and governs the convergence analysis of the general framework of implicit fixed-point proximity algorithms. Part of results from this topic is going to be present in [84].

1.3 INEXACT FIXED-POINT PROXIMITY ALGORITHMS

Besides the development of fixed-point iterations, the *inexact fixed-point iterations* have also been attracting much attention in numerical analysis [1, 33, 79]. Although the theoretical base of fixed-point iterations was growing increasingly solid, data scientists noticed that it would be advantageous to execute iterations with low precision and therefore at a high speed in first few steps [1]. On the other hand, as we embark on a new big-data era, the situation that the iteration scheme can only be carried out approximately has become more and more prevalent over time [26, 85]. Also, explicit iterations always has narrower range of parameters, which conceals the relations between different explicit algorithms and sometimes drags down the convergence speed [63]. These observations drove interest of data scientists into the study of inexact fixed-point iterations. Parallel to the study of exact fixed-point iterations, the inexact fixed-point iterations were first proven to converge with contraction mappings [1, 33], then with nonexpansive averaged operators [29, 59, 63]. Directly inherit from the study of inexact fixed-point iterations of nonexpansive averaged operators, inexact algorithms for nonsmooth convex optimization also access to their development opportunities [29, 59, 63].

Just analogous to general framework of implicit fixed-point proximity algorithms proposed in this dissertation, the convergence of the corresponding inexact implicit fixed-point proximity algorithms is as well not guaranteed by any of the existing literature. This limits the application of the newly proposed implicit algorithms, especially when the iteration is fully implicit and therefore cannot be exactly updated. Following the inspiration of Schur's lemma on the uniform boundedness of infinite matrices [91] (also see *e.g.*, [26, Lemma 6.21]), we propose a generalization of the existing analysis of inexact iterations of nonexpansive averaged operators, which covers the newly proposed quasinonexpansive/quasiaveraged operators. Then a direct application of this framework to implicit fixed-point proximity algorithms establishes the convergence analysis of inexact implicit fixed-point proximity algorithms. Part of results from this topic is going to be present in [84].

1.4 DISSERTATION CONTRIBUTIONS

This dissertation mainly contributes generalizations to inexact fixed-point iterations, implicit fixed-point proximity algorithms for nonsmooth convex optimization, and inexact implicit fixed-point proximity algorithms for nonsmooth convex optimization.

Inexact Fixed-Point Iterations. Inspired by the Schur's lemma [91] (also see *e.g.*, [26, Lemma 6.21]) which considers the uniform boundedness of an infinite matrix, we propose a general framework of inexact fixed-point iterations of Lipschitz operators. Such framework covers existing analysis of inexact fixed-point iterations of Picard/KM iterations for nonexpansive/nonexpansive averaged operators. As a compelling application, we propose new classes of operators, quasinonexpansive and quasiaveraged operators, generalizing the classes of nonexpansive/nonexpansive averaged operators respectively, and prove the convergence of their inexact Picard/KM iterations via the proposed framework of inexact fixed-point iterations.

Implicit Fixed-Point Proximity Algorithms. Inspired by the work of fixed-point proximity algorithms [56, 74], we characterize the solutions of the nonsmooth convex optimization problems to be fixed-points of a proximity operator coupling with a linear mapping. Although proximity operator has novel convergence property, the linear mapping is expanding and therefore invalidates the naïve Picard/KM iteration for such fixed-point problem. Following the matrix splitting technique in [56, 74], we then split all the matrices in the fixed-point equation, which results in a fully implicit fixed-point iteration scheme. Although the convergence of such general scheme could not be proven via the results for nonexpansive/nonexpansive averaged operators, it turns out that the iterating operator is quasiaveraged and therefore the framework of exact fixed-point iteration of quasiaveraged operators could be applied here. Such framework generalizes most of the existing analysis of numerous explicit schemes and could serve as a unified methodology of nonsmooth convex optimization, covering and generalizing gradient descent method, proximal point method, Douglas-Rachford splitting algorithm, first-order primal-dual algorithm, primal-dual hybrid gradient method, fixed-point proximity algorithm, alternating direction method of multipliers, split Bregman iteration, linearized alternating direction method of multipliers and inexact Uzawa method.

Inexact Implicit Fixed-Point Proximity Algorithms. Framework of inexact implicit fixed-point proximity algorithms follows from a direct combination of frameworks of inexact fixed-point iteration and implicit fixed-point proximity algorithms. To have an executable scheme of the fully implicit fixed-point proximity algorithms, we again apply the matrix splitting technique to the implicit scheme, introducing an inner loop in every step of the algorithm. With a careful choice of parameters, the inexact implicit fixed-point proximity algorithms fall into the intersection of the frameworks of inexact fixed-point iteration and implicit fixed-point proximity algorithms. The convergence therefore follows immediately. Such framework covers most of the analysis on inexact iterations of nonexpansive averaged operators, *e.g.*, [29, 59, 63]. As a concrete application, we propose the class of θ-inexact block-separable fixed-point proximity algorithms with convergence analysis, covering and extending the explicit fixed-point proximity algorithms.

1.5 DISSERTATION OUTLINE

This section outlines the main structure of this dissertation. Chapter 2 proposes the general framework of inexact fixed-point iterations, which is one of the main pillars of this dissertation. We introduce the fixed-point problems and review the development of exact/inexact methods with stationary/nonstationary iterations for fixed-point problems. Then we propose the frameworks of inexact stationary/nonstationary iterations, which then are proven to cover most of the typical iterative methods nowadays. Chapter 3 then switches to the nonsmooth convex optimization problem, which is the other main topic of this dissertation. We introduce the nonsmooth convex optimization problems and review the latest iterative methods for this kind of problems. In this chapter we propose the general framework of implicit fixed-point proximity algorithms, which then is proven to cover most of the popular proximity algorithms. Chapter 4 thereafter combines the proposed frameworks of inexact fixed-point iterations and implicit fixed-point proximity algorithms, leading us to a comprehensive framework of inexact implicit fixed-point proximity algorithms. To validate these new frameworks in this dissertation, Chapter 5 applies the inexact implicit fixed-point proximity algorithms to image processing problems, with comparison to popular existing explicit methods. The numerical experiments approve the superiority of the general framework of inexact implicit fixed-point proximity algorithms over explicit ones. As a conclusion, Chapter 6 summarizes results proposed in this dissertation and discusses possible direction of further studies.

CHAPTER 2

INEXACT FIXED-POINT ITERATIONS

We in this chapter review several classical exact/inexact fixed-point methods for fixed-point problems, and propose generalized framework of inexact fixed-point iterations. Specifically, we review exact/inexact Picard/Krasnosel'skiĭ-Mann iterations of nonexpansive/nonexpansive averaged operators. Then following the inspiration of Schur's lemma on uniform boundedness of infinite matrices, we propose a general analysis of inexact fixed-point iterations of Lipschitz operators. As applications, we define the class of *quasinonexpansive*/*quasiaveraged operators* and consider the exact/inexact Picard/Krasnosel'skiĭ-Mann iterations of them. These new results cover and extend existing analysis for nonexpansive/nonexpansive averaged operators accordingly. It is worthy mentioning that this framework provides new understandings and extensions for inexact Picard/Krasnosel'skiĭ-Mann iterations of widely-used firmly quasinonexpansive operators.

2.1 FIXED-POINT PROBLEM AND ITERATIVE METHODS

In this section we clarify the fixed-point problem, and review several universal iterative methods with their existing convergence theorems.

The fixed-point problem is, for an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ having a nonempty fixed-point set

$$\operatorname{Fix} T := \{z \in \mathbb{R}^n : Tz = z\},$$

we need to find a point in $\operatorname{Fix} T$. With $\mathbb{R}^n$ we denote the n-dimensional Euclidean space. Fixed points can describe solutions and many fundamental concepts of equilibria or stability in various fields. Mentioned below are several typical applications.

Ordinary Differential Equations. The initial value problem considers differential equation of $x : \mathbb{R} \to \mathbb{R}^n$ such that

$$x'(t) = f(t, x(t)), \qquad x(t_0) = x_0,$$

with $f : \Omega \to \mathbb{R}^n$ and $(t_0, x_0) \in \Omega$ where $\Omega \subset \mathbb{R} \times \mathbb{R}^n$ is an open set. Then the solution of initial value problem can be characterized [80] as the fixed-point of the operator mapping function $x = x(t)$ to

$$x(t) \mapsto x_0 + \int_{t_0}^{t} f(\tau, x(\tau))\mathrm{d}\tau.$$

Zeros of Operators. For a function $f : \mathbb{R}^n \to \mathbb{R}^n$, the zeros of f can be characterized as fixed points of $I - f$, *i.e.*, $\mathrm{Fix}(I - f)$. Here I is the identity operator, and I_n is the identity operator in $\mathbb{R}^n$ if the dimension is clearly needed.

Smooth Convex Optimization. For a differential convex function $f : \mathbb{R}^n \to \mathbb{R}$, Fermat's rule [42] claims that all global minimizers of f can be characterized as zeros of ∇_f, and therefore can be modeled by $\mathrm{Fix}(I - \nabla_f)$. Here ∇_f is the gradient operator of f.

For general fixed-point problems, there are briefly two types of methods, *i.e.*, *direct method*, which attempts to solve the problem by a finite sequence of operations, and *iterative method*, which uses an initial value to generate a sequence of improving approximate fixed-points, in which the next approximation is derived from the previous ones. Although direct method theoretically gives an exact solution at most of the time, it still suffers from several drawbacks, *e.g.*, theoretical difficulty of convergence analysis with nonlinear problems, and high time/space complexity with problems of large size. As problems and data sets arise from data science are getting more and more complicated and large, iterative method has been increasingly attractive and practical.

In following subsections, we review exact iterative methods and inexact iterative methods for fixed-point problems. Several typical iterative methods are described with existing convergence theorems respectively.

2.1.1 EXACT ITERATIVE METHODS

This subsection reviews general exact iterative methods for fixed-point problems, namely Picard fixed-point iteration for contraction mappings and nonexpansive averaged operators, and Krasnosel'skiĭ-Mann iteration for nonexpansive averaged operators.

Plenty of methods are proposed to solve the fixed-point problems throughout the history of computational mathematics. Charles Émile Picard firstly initiated the fundamental method of successive approximations which then was named after him, the Picard iteration [80], as

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \lfloor\; z_{k+1} \leftarrow T z_k\end{array} \tag{1}$$

for a special kind of contraction operator $T: \mathbb{R}^n \to \mathbb{R}^n$ with $\operatorname{Fix} T \neq \varnothing$ raising from differential equations. Here $\mathbb{N}$ denotes the set of non-positive integers, and '$\leftarrow$' assigns the value on the right-hand side to the variable on the left-hand side. Then the convergence theorem for general contraction mappings was formally stated later by Stefan Banach [4] as follows. A mapping $T: \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous if there exists $L \in [0, +\infty)$ such that

$$\|Tx - Ty\| \leq L\|x - y\|, \qquad \text{for all } x, y \in \mathbb{R}^n,$$

where norm $\|\cdot\| := \|\cdot\|_2$ is the Euclidean norm of $\mathbb{R}^n$. An operator $T: \mathbb{R}^n \to \mathbb{R}^n$ is a contraction mapping if it is L-Lipschitz with $L \in [0, 1)$. The following theorem was once proved in [4].

Theorem 2.1. *If $T: \mathbb{R}^n \to \mathbb{R}^n$ is a contraction mapping, then* $\operatorname{Fix} T$ *is a singleton and for any $z_0 \in \mathbb{R}^n$, the sequence $\{z_k\}_{k \in \mathbb{N}}$ generated by Picard iteration* (1) *converges to the unique point of* $\operatorname{Fix} T$.

However contraction mappings are limited in application due to their rather strong and *ad hoc* hypothesis. Originating in the study [52] by Mark Krasnosel'skiĭ, a broader class of operators that extends contraction mappings, the so-called nonexpansive averaged operators, then gradually took its form and was finally proven to be converging towards fixed-points under Picard iterations as well. An L-Lipschitz continuous operator with $L = 1$ is also called nonexpansive operator. An operator $T: \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive averaged [3, 29, 73, 90], if there exists $\kappa \in (0, 1)$ and a nonexpansive operator $R: \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T = (1 - \kappa)I + \kappa R. \tag{2}$$

In particular, nonexpansive 1/2-averaged operators are called *firmly nonexpansive* [13, 14]. Notice that all nonexpansive averaged operators are contraction mappings, however identity operator I serves as a trivial nonexpansive averaged example that is not contraction mapping. This argument implies that the following theorem, which could be found in [12, 90], essentially extends Theorem 2.1.

Theorem 2.2. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive κ-averaged with $\kappa \in (0, 1)$. If $\operatorname{Fix} T \neq \emptyset$, then for any $z_0 \in \mathbb{R}^n$, the sequence $\{z_k\}_{k \in \mathbb{N}}$ generated by Picard iteration* (1) *converges to a point in* $\operatorname{Fix} T$.

As Picard iteration (1) may fail with nonexpansive operator T, Theorem 2.2 can be considered as a generalized version of Picard iteration for nonexpansive operator T. To see this, for $T : \mathbb{R}^n \to \mathbb{R}^n$ we consider the following algorithm

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \lfloor\ z_{k+1} \leftarrow (1-\lambda) z_k + \lambda T z_k \end{array} \tag{3}$$

where $\lambda \in \mathbb{R}$. If T is a contraction mapping, then Theorem 2.1 claims the convergence of iteration (3) towards $\operatorname{Fix} T$ when $\lambda \in (0, 1]$. If T is merely nonexpansive with $\operatorname{Fix} T \neq \emptyset$ and $\lambda \in (0, 1)$, then $(1 - \lambda)I + \lambda T$ is nonexpansive averaged, therefore Theorem 2.2 ensures that (3) converges to a point in

$$\operatorname{Fix}((1 - \lambda)I + \lambda T) = \operatorname{Fix} T.$$

The argument above proves the following corollary.

Corollary 2.3. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive with $\operatorname{Fix} T \neq \emptyset$. If $\lambda \in (0, 1)$, then for any $z_0 \in \mathbb{R}^n$, the sequence $\{z_k\}_{k \in \mathbb{N}}$ generated by iteration* (3) *converges to a point in* $\operatorname{Fix} T$.

The algorithm (3) also leads to the following well-known corollary.

Corollary 2.4. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive κ-averaged with $\kappa \in (0, 1)$ and $\operatorname{Fix} T \neq \emptyset$. If $\lambda \in (0, 1/\kappa)$, then for any $z_0 \in \mathbb{R}^n$, the sequence $\{z_k\}_{k \in \mathbb{N}}$ generated by iteration* (3) *converges to a point in* $\operatorname{Fix} T$.

Proof. Observe that if T is nonexpansive κ-averaged, then there exists a nonexpansive operator $R : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T = (1 - \kappa)I + \kappa R.$$

Therefore for $\lambda \in \mathbb{R}$ we have

$$\begin{aligned} (1 - \lambda)I + \lambda T &= (1 - \lambda)I + \lambda((1 - \kappa)I + \kappa R) \\ &= (1 - \lambda\kappa)I + \lambda\kappa R, \end{aligned}$$

which means that for any $\lambda \in (0, 1/\kappa)$, operator $(1 - \lambda)I + \lambda T$ is nonexpansive $\lambda\kappa$-averaged. Then a direct application of Theorem 2.2 proves the convergence of iteration (3). □

After combined with the earlier idea of nonstationary iteration by W. Robert Mann [70], (3) was finally extended to the Krasnosel'skiĭ-Mann (KM) iteration for nonexpansive operators. For nonexpansive operator $T : \mathbb{R}^n \to \mathbb{R}^n$, the KM iteration [83] updates

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \lfloor \; z_{k+1} \leftarrow (1-\lambda_k) z_k + \lambda_k T z_k \end{array} \tag{4}$$

where $\{\lambda_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$. Such iteration is no longer a Picard iteration. Actually it is a *nonstationary iteration*, since update rule is changing in each step. Notice that (3) is a special case of iteration (4) with $\lambda_k = \lambda$ for $k \in \mathbb{N}$ and some $\lambda \in \mathbb{R}$. The following theorem was once established in [83].

Theorem 2.5. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive with $\operatorname{Fix} T \neq \varnothing$ and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,1]$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k(1-\lambda_k) = \infty,$$

then for any $z_0 \in \mathbb{R}^n$, the sequence $\{z_k\}_{k\in\mathbb{N}}$ generated by KM iteration (4) converges to a point in $\operatorname{Fix} T$.

Theorem 2.5 clearly covers Theorem 2.2. Analogous to Theorem 2.2, Theorem 2.5 has its corresponding corollary on nonexpansive averaged operators.

Corollary 2.6. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive κ-averaged with $\kappa \in (0,1)$, $\operatorname{Fix} T \neq \varnothing$, and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 1/\kappa]$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k \left(\frac{1}{\kappa} - \lambda_k\right) = \infty,$$

then for any $z_0 \in \mathbb{R}^n$, the sequence $\{z_k\}_{k\in\mathbb{N}}$ generated by KM iteration (4) converges to a point in $\operatorname{Fix} T$.

Proof. Set $R := (T - (1-\kappa)I)/\kappa$ and $\alpha_k := \lambda_k \kappa$ for $k \in \mathbb{N}$. Then by conditions we have that R is nonexpansive, $\operatorname{Fix} T = \operatorname{Fix} R$, $\{\alpha_k\}_{k\in\mathbb{N}} \subset [0,1]$ and

$$\sum_{k\in\mathbb{N}} \alpha_k(1-\alpha_k) = \infty.$$

Notice that (4) is equivalently updating via

$$z_{k+1} = (1-\alpha_k) z_k + \alpha_k R z_k,$$

therefore the convergence of iteration (4) directly follows Theorem 2.5. □

KM iteration nowadays still attracts wide interest among computational mathematics community and gets development continuously, *e.g.*, the variable KM iteration [104, 107] was proposed as

$$\begin{aligned} &\textbf{For } k \in \mathbb{N} \\ &\lfloor\; z_{k+1} \leftarrow (1-\lambda_k)z_k + \lambda_k T_k z_k \end{aligned} \tag{5}$$

where $\{T_k\}_{k\in\mathbb{N}}$ are group of nonexpansive operators satisfying certain conditions. To generally handle such kind of iterations, the nonstationary iteration is modeled by the general form

$$\begin{aligned} &\textbf{For } k \in \mathbb{N} \\ &\lfloor\; z_{k+1} \leftarrow T_k z_k \end{aligned} \tag{6}$$

Notice that Picard iterations (1), KM iteration (4) and variable KM iteration (5) are special cases of nonstationary iterations (6).

2.1.2 INEXACT ITERATIVE METHODS

We in this subsection review inexact iterative methods for fixed-point problems, which generalizes exact methods introduced in Section 2.1.1.

Iterative methods (1), (3) and (4) are practical and useful in finding fixed-points of an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ under different conditions, but such novel properties are theoretically guaranteed only if the value Tz can be exactly evaluated for all $z \in \mathbb{R}^n$. In numerical analysis it is common that the evaluation of T could only be carried out approximately [1, 33, 79], or it is beneficial to approximate the evaluation of T instead of solve it exactly [1, 63]. That is, instead of the exact value Tz, the actual value available is the inexact value $Tz + \varepsilon$, where limited information of ε is known (*e.g.*, $\|\varepsilon\|$ could be obtained or bounded).

Plenty of remarkable works were undertook to generalize convergence results of exact iterative methods to inexact settings. The inexact Picard iteration for a contraction mapping was once systematically formalized and studied in [1] as the following. Consider the inexact Picard iteration for $T : \mathbb{R}^n \to \mathbb{R}^n$,

$$\begin{aligned} &\textbf{For } k \in \mathbb{N} \\ &\lfloor\; \tilde{z}_{k+1} \leftarrow T\tilde{z}_k + \varepsilon_k \end{aligned} \tag{7}$$

where $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ is the inexact iteration sequence and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ are evaluation errors that can not be exactly obtained. The following theorem was once proven in [1].

Theorem 2.7. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction mapping. If $\lim_{k\in\mathbb{N}} \varepsilon_k = 0$, then for any $z_0 \in \mathbb{R}^n$, the sequence $\{z_k\}_{k\in\mathbb{N}}$ generated by inexact Picard iteration* (7) *converges to the unique point of* $\operatorname{Fix} T$.

Theorem 2.1 is apparently a special case of Theorem 2.7 where $\varepsilon_k = 0$ for $k \in \mathbb{N}$. Similarly the inexact version of KM iterations (4) was studied [29, 59]. Consider

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \lfloor\ \tilde{z}_{k+1} \leftarrow (1-\lambda_k)\tilde{z}_k + \lambda_k(T\tilde{z}_k + \varepsilon_k)\end{array} \tag{8}$$

where $\{\lambda_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$, $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ is the inexact iteration sequence and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ are evaluation errors of $\{T\tilde{z}_k\}_{k\in\mathbb{N}}$. Here by $\ell^p(\mathbb{R}^n)$ we denote the collection of all sequences $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ with the property that

$$\sum_{k\in\mathbb{N}} \|x_k\|^p < \infty,$$

and by ℓ^p when the domain is clear in context. The following theorem could be found in [29].

Theorem 2.8. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be nonexpansive with $\operatorname{Fix} T \neq \varnothing$, $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,1]$ and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$. If*

$$\{\lambda_k\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1 \quad \textit{and} \quad \sum_{k\in\mathbb{N}} \lambda_k(1-\lambda_k) = \infty,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact KM iteration (8) *converges to a point in* $\operatorname{Fix} T$.

Theorem 2.8 can be easily extended to inexact iteration of nonexpansive averaged operators similar as Theorem 2.5.

Corollary 2.9. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be nonexpansive κ-averaged with $\kappa \in (0,1)$, $\operatorname{Fix} T \neq \varnothing$, $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 1/\kappa]$ and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$. If*

$$\{\lambda_k\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1 \quad \textit{and} \quad \sum_{k\in\mathbb{N}} \lambda_k\left(\frac{1}{\kappa} - \lambda_k\right) = \infty,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact KM iteration (8) *converges to a point in* $\operatorname{Fix} T$.

Proof. This proof basically follows Corollary 2.6. Set $\alpha_k := \lambda_k \kappa$ for $k \in \mathbb{N}$ and

$$R := \frac{1}{\kappa} T + \left(1 - \frac{1}{\kappa}\right) I.$$

Then by conditions we have that R is nonexpansive, $\operatorname{Fix} T = \operatorname{Fix} R$, $\{\alpha_k\}_{k\in\mathbb{N}} \subset [0,1]$ and

$$\sum_{k\in\mathbb{N}} \alpha_k(1-\alpha_k) = \infty.$$

Notice that (8) is equivalently updating via

$$\tilde{z}_{k+1} = (1-\alpha_k)\tilde{z}_k + \alpha_k R \tilde{z}_k + \lambda_k \varepsilon_k,$$

therefore the convergence of iteration (8) directly follows Theorem 2.8. □

Corollary 2.9 has a trivial corollary on inexact Picard iteration of nonexpansive averaged operators.

Corollary 2.10. *Let* $T : \mathbb{R}^n \to \mathbb{R}^n$ *be nonexpansive* κ*-averaged with* $\kappa \in (0,1)$, $\operatorname{Fix} T \neq \emptyset$ *and* $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$. *If*

$$\{\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1 \quad \text{and} \quad \lambda_k = \lambda \in (0, 1/\kappa) \text{ for all } k \in \mathbb{N},$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, *the inexact sequence* $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ *generated by inexact Picard iteration* (8) *converges to a point in* $\operatorname{Fix} T$.

Finally, for nonstationary iteration (6), [59] considered the following inexact nonstationary KM iteration. Let $T_\Gamma : \mathbb{R}^n \to \mathbb{R}^n$ be a nonexpansive operator depending on parameter Γ and operators $T_{\Gamma_k} : \mathbb{R}^n \to \mathbb{R}^n$ for $k \in \mathbb{N}$. Then the inexact nonstationary KM iteration is read

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \lfloor\ \tilde{z}_{k+1} \leftarrow (1-\lambda_k)\tilde{z}_k + \lambda_k\big(T_{\Gamma_k}\tilde{z}_k + \varepsilon_k\big) \end{array} \tag{9}$$

where $\{\lambda_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$, $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ is the inexact iteration sequence and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ are evaluation errors. Define

$$T_{\Gamma,\lambda_k} := (1-\lambda_k)\, I + \lambda_k T_\Gamma, \qquad T_{\Gamma_k,\lambda_k} := (1-\lambda_k)\, I + \lambda_k T_{\Gamma_k},$$

and for all $\rho \geq 0$, denote

$$\Delta^T_{k,\rho} := \sup_{\|z\|\leq\rho} \big\| T_{\Gamma_k} z - T_\Gamma z \big\|.$$

Readers are referred to [59] for the following theorem.

Theorem 2.11. *Assume that $T_\Gamma : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive with $\operatorname{Fix} T_\Gamma \neq \emptyset$, and followings hold*

(i) T_{Γ_k,λ_k} *is $(1+\beta_k)$-Lipschitz with $\beta_k \geq 0$ for $k \in \mathbb{N}$ and $\{\beta_k\}_{k\in\mathbb{N}} \in \ell^1$,*

(ii) $\{\lambda_k\}_{k\in\mathbb{N}} \subset (0,1)$ *with* $\inf_{k\in\mathbb{N}} \lambda_k(1-\lambda_k) > 0$,

(iii) $\{\lambda_k \varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$,

(iv) $\{\lambda_k \Delta^T_{k,\rho}\}_{k\in\mathbb{N}} \in \ell^1$ *for all* $\rho \geq 0$.

Then for any $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact nonstationary KM iteration (9) *converges to a point in* $\operatorname{Fix} T_\Gamma$.

Theorem 2.11 has a direct corollary for nonexpansive averaged operators similar as Theorem 2.5, proof of which is quite straightforward and was once discussed in [59].

Corollary 2.12. *Assume that R_Γ and $\{R_{\Gamma_k}\}_{k\in\mathbb{N}}$ are nonexpansive operators in $\mathbb{R}^n$ with $\operatorname{Fix} T_\Gamma \neq \emptyset$, and followings hold*

(i) $T_{\Gamma_k} = (1-\kappa_k)I + \kappa_k R_\Gamma$ *is nonexpansive κ_k-averaged with $\kappa_k \in (0,1]$ for $k \in \mathbb{N}$,*

(ii) $\lambda_k \subset (0, 1/\kappa_k)$ *with* $\inf_{k\in\mathbb{N}} \lambda_k\kappa_k(1-\lambda_k\kappa_k) > 0$,

(iii) $\{\lambda_k \kappa_k \varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$,

(iv) $\{\lambda_k \kappa_k \Delta^R_{k,\rho}\}_{k\in\mathbb{N}} \in \ell^1$ *for all* $\rho \geq 0$.

Then for any $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact nonstationary KM iteration (9) *converges to a point in* $\operatorname{Fix} R_\Gamma$.

In this section we review the origin of fixed-point problem and several classical exact/inexact iterative methods for fixed-point problem, *e.g.*, exact/inexact Picard iterations, exact/inexact KM iterations and exact/inexact nonstationary KM iterations, with existing convergence theorems respectively. In the following sections, we propose the general frameworks of inexact Picard fixed-point iterations and inexact nonstationary iterations, which covers and extends results in this section.

2.2 FRAMEWORK OF INEXACT PICARD FIXED-POINT ITERATIONS

In the following two sections we propose the general framework of inexact fixed-point iteration, which is one of the main pillars in this dissertation. First we consider the inexact Picard fixed-point iteration of quasinonexpansive operators. Then we introduce a generalized class of quasinonexpansive operators, to which we extend the proposed inexact fixed-point iteration framework.

Recall that inexact Picard iteration designates the following numerical scheme

$$\begin{aligned}&\textbf{For } k \in \mathbb{N}\\ &\big\lfloor\ \tilde{z}_{k+1} \leftarrow T\tilde{z}_k + \varepsilon_k\end{aligned} \tag{7}$$

where $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ are the *local errors*. One of the main purpose of this study is to establish a general theorem which identifies essential conditions on the local error to ensure convergence of the inexact fixed-point iteration for Lipschitz operators. The critical issue for the convergence is to enforce local errors to be distributed in certain uniform manner so that they are under control globally. Inspired by the Schur's lemma [91] (also see *e.g.*, [26, Lemma 6.21]), which considers the uniform boundedness of an infinite matrix, we put forward a general convergence theorem for inexact Picard iterations. We first present a technical lemma.

Lemma 2.13. *Let $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$. If*

$$\sum_{k\in\mathbb{N}} \|x_{k+1} - x_k\| < \infty,$$

then $\{x_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence.

Proof. By hypothesis, for any given $\varepsilon > 0$, there exists $N > 0$ such that for any $p > N$, there holds $\sum_{k=p}^{\infty} \|x_{k+1} - x_k\| < \varepsilon$. This ensures that

$$\|x_p - x_q\| \leq \sum_{k=p}^{q-1} \|x_{k+1} - x_k\| < \varepsilon, \qquad \text{for all } p > q > N,$$

which proves that $\{x_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence. □

We are now ready to state a general convergence theorem for inexact Picard iteration (7). Recall that L-Lipschitz continuous operator with $L = 1$ is called nonexpansive operator, and an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is called *quasinonexpansive* [34, 36], if

$$\|Tx - z\| \leq \|x - z\|, \qquad \text{for all } x \in \mathbb{R}^n, z \in \operatorname{Fix} T. \tag{10}$$

A nonexpansive operator is quasinonexpansive, but a quasinonexpansive operator is not necessarily nonexpansive. The following function serves as a counterexample.

Example 2.14. Define $T : \mathbb{R} \to \mathbb{R}$ as

$$T(x) := \begin{cases} \pi^{-1} x \arctan(x) \sin(\ln|x|), & x \neq 0, \\ 0, & x = 0, \end{cases}$$

then T is quasinonexpansive and Lipschitz continuous, but not nonexpansive.

For quasinonexpansive and Lipschitz operators, we have the following general theorem of inexact iterations.

Theorem 2.15. *Suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$ is quasinonexpansive with $\operatorname{Fix} T \neq \emptyset$, Lipschitz continuous, and for all $z \in \mathbb{R}^n$, the exact Picard sequence $\{T^k z\}_{k\in\mathbb{N}}$ converges to a point in $\operatorname{Fix} T$. If local errors*

$$\{\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact Picard iteration (7) converges to a point in $\operatorname{Fix} T$.

Proof. Let L be the Lipschitz constant of T. If $L < 1$, then T is indeed a contraction mapping, and by Theorem 2.7 the inexact Picard sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by iteration (7) converges to a point in $\operatorname{Fix} T$.

We now consider the case when $L \geq 1$. First we prove the boundedness of $\{\tilde{z}_k\}_{k\in\mathbb{N}}$. For any fixed $z \in \operatorname{Fix} T$ and for all $k \in \mathbb{N}$, using the triangle inequality we have that

$$\|\tilde{z}_k - z\| = \|T\tilde{z}_{k-1} + \varepsilon_{k-1} - z\| \leq \|T\tilde{z}_{k-1} - z\| + \|\varepsilon_{k-1}\|.$$

Applying (10) to the first term on the right hand side of the inequality above yields

$$\|\tilde{z}_k - z\| \leq \|\tilde{z}_{k-1} - z\| + \|\varepsilon_{k-1}\|.$$

Recursive application of the above inequality leads to

$$\|\tilde{z}_k - z\| \leq \|\tilde{z}_0 - z\| + \sum_{l=0}^{k-1} \|\varepsilon_l\| \leq \|\tilde{z}_0 - z\| + \sum_{l\in\mathbb{N}} \|\varepsilon_l\|,$$

which implies that for all $k \in \mathbb{N}$, $\|\tilde{z}_k\| \leq c := \|z\| + \|\tilde{z}_0 - z\| + \sum_{l\in\mathbb{N}} \|\varepsilon_l\|$. The boundedness of $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ ensures that there exists a convergent subsequence $\{\tilde{z}_{k_l}\}_{l\in\mathbb{N}}$. Let $u := \lim_{l\to\infty} \tilde{z}_{k_l}$. By assumptions, $v := \lim_{l\to\infty} T^l u$ exists and $v \in \operatorname{Fix} T$. We next prove that $\lim_{k\to\infty} \tilde{z}_k = v$, which would complete the proof of this theorem.

Suppose $\varepsilon > 0$. Since $v = \lim_{l\to\infty} T^l u$, there exists $N \in \mathbb{N}$ such that

$$\left\|T^N u - v\right\| < \varepsilon. \tag{11}$$

Since $\{\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$, there exists $M_1 \in \mathbb{N}$ such that for all $m \geq M_1$,

$$\sum_{j=k_m}^{\infty} \left\|\varepsilon_j\right\| < L^{-N}\varepsilon. \tag{12}$$

Moreover, because $\lim_{l\to\infty} \tilde{z}_{k_l} = u$, there exists $M_2 \in \mathbb{N}$ such that for all $m \geq M_2$,

$$\left\|\tilde{z}_{k_m} - u\right\| < L^{-N}\varepsilon. \tag{13}$$

Let $M := \max\{M_1, M_2\}$. We next estimate $\|\tilde{z}_{k_M+N} - v\|$ by employing the decomposition

$$\tilde{z}_{k_M+N} - v = \left(\tilde{z}_{k_M+N} - T^N\tilde{z}_{k_M}\right) + \left(T^N\tilde{z}_{k_M} - T^N u\right) + \left(T^N u - v\right). \tag{14}$$

For the first term on the right hand side of (14), by (7) we obtain that

$$\tilde{z}_{k_M+N} - T^N\tilde{z}_{k_M} = T\tilde{z}_{k_M+N-1} + \varepsilon_{k_M+N-1} - T^N\tilde{z}_{k_M}.$$

The triangle inequality combined with the Lipschitz continuity of T implies that

$$\left\|\tilde{z}_{k_M+N} - T^N\tilde{z}_{k_M}\right\| \leq L\left\|\tilde{z}_{k_M+N-1} - T^{N-1}\tilde{z}_{k_M}\right\| + \left\|\varepsilon_{k_M+N-1}\right\|.$$

Repeatedly applying the above inequality, we observe that

$$\left\|\tilde{z}_{k_M+N} - T^N\tilde{z}_{k_M}\right\| \leq L^{N-1}\sum_{j=0}^{N-1} L^{-j}\left\|\varepsilon_{k_M+j}\right\|. \tag{15}$$

The hypothesis $L \geq 1$ ensures that

$$\sum_{j=0}^{N-1} L^{-j} \left\| \varepsilon_{k_M+j} \right\| \leq \sum_{j=0}^{N-1} \left\| \varepsilon_{k_M+j} \right\| \leq \sum_{j=k_M}^{\infty} \left\| \varepsilon_j \right\|.$$

Substituting this inequality into the right hand side of (15) and using (12), we find that

$$\left\| \tilde{z}_{k_M+N} - T^N \tilde{z}_{k_M} \right\| \leq L^{N-1} \sum_{j=k_M}^{\infty} \left\| \varepsilon_j \right\| < L^{-1} \varepsilon. \tag{16}$$

For the second term on the right hand side of (14), by the Lipschitz continuity of T and (13), we derive that

$$\left\| T^N \tilde{z}_{k_M} - T^N u \right\| \leq L^N \left\| \tilde{z}_{k_M} - u \right\| < \varepsilon. \tag{17}$$

The third term on the right hand side of (14) has been estimated by (11). Combining (11), (14), (16) and (17), we get the estimation

$$\left\| \tilde{z}_{k_M+N} - v \right\| < \left(L^{-1} + 2 \right) \varepsilon. \tag{18}$$

Once again, by repeatedly using (10), for $v \in \operatorname{Fix} T$, we have for all $j \in \mathbb{N}$,

$$\begin{aligned} \left\| \tilde{z}_{k_M+N+j} - v \right\| &= \left\| T \tilde{z}_{k_M+N+j-1} + \varepsilon_{k_M+N+j-1} - v \right\| \\ &\leq \left\| \tilde{z}_{k_M+N+j-1} - v \right\| + \left\| \varepsilon_{k_M+N+j-1} \right\| \\ &\leq \left\| \tilde{z}_{k_M+N} - v \right\| + \sum_{p=0}^{j-1} \left\| \varepsilon_{k_M+N+p} \right\|. \end{aligned}$$

Substituting (12) and (18) inter the right hand side of the above inequality gives for all $j \in \mathbb{N}$,

$$\left\| \tilde{z}_{k_M+N+j} - v \right\| < \left(L^{-1} + 3 \right) \varepsilon, \tag{19}$$

which proves that $\lim_{k \to \infty} \tilde{z}_k = v \in \operatorname{Fix} T$. □

Here we further consider inexact Picard iterations for a generalized class of quasinonexpansive operator. We denote the set of symmetric positive definite n-by-n matrices as $\mathbb{S}^n$, and $\| \cdot \|_V := \sqrt{\langle \cdot, V \cdot \rangle}$ is the weighted semi-norm with respect to $V \in \mathbb{S}^n$. An operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is quasinonexpansive with respect to $V \in \mathbb{S}^n$, if

$$\|Tx - z\|_V \leq \|x - z\|_V, \qquad \text{for all } x \in \mathbb{R}^n, z \in \operatorname{Fix} T. \tag{20}$$

Clearly quasinonexpansive operators are special cases of quasinonexpansive operators with respect to I_n. An operator is L-Lipschitz with respect to $V \in \mathbb{S}^n$ if

$$\|Tx - Ty\|_V \leq L\|x - y\|_V, \qquad \text{for all } x, y \in \mathbb{R}^n,$$

and the class of 1-Lipschitz operators with respect to V are also specified as nonexpansive operators with respect to V.

Now we are ready to propose a generalization of Theorem 2.15. For $V \in \mathbb{S}^n$, denote $\sqrt{V} \in \mathbb{S}^n$ as the unique symmetric positive definite matrix such that $V = \sqrt{V}\sqrt{V}$.

Theorem 2.16. *Suppose that $V \in \mathbb{S}^n$, $T : \mathbb{R}^n \to \mathbb{R}^n$ is quasinonexpansive with respect to V and Lipschitz continuous with respect to V, $\operatorname{Fix} T \neq \varnothing$, and for all $z \in \mathbb{R}^n$, the exact Picard sequence $\{T^k z\}_{k \in \mathbb{N}}$ converges to a point in $\operatorname{Fix} T$. If local errors*

$$\left\{ \sqrt{V}\varepsilon_k \right\}_{k \in \mathbb{N}} \in \ell^1,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\sqrt{V}\tilde{z}_k\}_{k \in \mathbb{N}}$ generated by inexact Picard iteration (7) converges to a point in $\sqrt{V} \operatorname{Fix} T$.

Proof. This proof basically follows the proof of Theorem 2.15. By Lipschitz continuity of T, assume that there exists $L \geq 1$ such that

$$\|Tx - Ty\|_V \leq L\|x - y\|_V. \qquad \text{for all } x, y \in \mathbb{R}^n.$$

First notice (20) ensures for $k \in \mathbb{N}$ and $z \in \operatorname{Fix} T$,

$$\|\tilde{z}_k - z\|_V \leq \|T\tilde{z}_{k-1} - z\|_V + \|\varepsilon_{k-1}\|_V \leq \|\tilde{z}_{k-1} - z\|_V + \|\varepsilon_{k-1}\|_V.$$

Recursive application of the above inequality results in

$$\left\|\sqrt{V}\tilde{z}_k\right\| \leq \|z\|_V + \sum_{l \in \mathbb{N}} \|\varepsilon_l\|_V, \qquad \text{for all } k \in \mathbb{N}.$$

This proves that $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ is bounded, then there exists a convergent subsequence $\{\sqrt{V}\tilde{z}_{k_l}\}_{l\in\mathbb{N}}$. Since range of $\sqrt{V}$ is closed, there exists $u \in \mathbb{R}^n$ such that $\sqrt{V}u = \lim_{l\to\infty}\sqrt{V}\tilde{z}_{k_l}$. By assumptions, $v := \lim_{l\to\infty} T^l u$ exists and $v \in \operatorname{Fix} T$. We next prove that $\lim_{k\to\infty}\sqrt{V}\tilde{z}_k = \sqrt{V}v$, which would complete the proof of this theorem.

Suppose $\varepsilon > 0$. Since $v = \lim_{l\to\infty} T^l u$, there exists $N \in \mathbb{N}$ such that

$$\left\|T^N u - v\right\|_V < \varepsilon. \tag{21}$$

Since $\{\sqrt{V}\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$, there exists $M_1 \in \mathbb{N}$ such that for all $m \geq M_1$,

$$\sum_{j=k_m}^{\infty}\left\|\varepsilon_j\right\|_V < L^{-N}\varepsilon. \tag{22}$$

Moreover, because $\sqrt{V}u = \lim_{l\to\infty}\sqrt{V}\tilde{z}_{k_l}$, there exists $M_2 \in \mathbb{N}$ such that for all $m > M_2$,

$$\left\|\tilde{z}_{k_l} - u\right\|_V < L^{-N}\varepsilon. \tag{23}$$

Let $M := \max\{M_1, M_2\}$. We next estimate $\|\tilde{z}_{k_M+N} - v\|_V$ by employing the decomposition

$$\tilde{z}_{k_M+N} - v = \left(\tilde{z}_{k_M+N} - T^N\tilde{z}_{k_M}\right) + \left(T^N\tilde{z}_{k_M} - T^N u\right) + \left(T^N u - v\right). \tag{24}$$

For the first term on the right hand side of (24), by (7)

$$\tilde{z}_{k_M+N} - T^N\tilde{z}_{k_M} = T\tilde{z}_{k_M+N-1} + \varepsilon_{k_M+N-1} - T^N\tilde{z}_{k_M}.$$

The triangle inequality combined with the Lipschitz continuity of T implies that

$$\left\|\tilde{z}_{k_M+N} - T^N\tilde{z}_{k_M}\right\|_V \leq L\left\|\tilde{z}_{k_M+N-1} - T^{N-1}\tilde{z}_{k_M}\right\|_V + \left\|\varepsilon_{k_M+N-1}\right\|_V.$$

Repeatedly applying the above inequality, we observe that

$$\left\|\tilde{z}_{k_M+N} - T^N\tilde{z}_{k_M}\right\|_V \leq L^{N-1}\sum_{j=0}^{N-1} L^{-j}\left\|\varepsilon_{k_M+j}\right\|_V. \tag{25}$$

The hypothesis $L \geq 1$ ensures that

$$\sum_{j=0}^{N-1} L^{-j} \left\|\varepsilon_{k_M+j}\right\|_V \leq \sum_{j=0}^{N-1} \left\|\varepsilon_{k_M+j}\right\|_V \leq \sum_{j=k_M}^{\infty} \left\|\varepsilon_j\right\|_V.$$

Substituting this inequality into the right hand side of (25) and using (22), we find that

$$\left\|\tilde{z}_{k_M+N} - T^N \tilde{z}_{k_M}\right\|_V \leq L^{N-1} \sum_{j=k_M}^{\infty} \left\|\varepsilon_j\right\|_V < L^{-1}\varepsilon. \tag{26}$$

For the second term on the right hand side of (24), by the Lipschitz continuity of T and (23), we derive

$$\left\|T^N \tilde{z}_{k_M} - T^N u\right\|_V \leq L^N \left\|\tilde{z}_{k_M} - u\right\|_V < \varepsilon. \tag{27}$$

The third term on the right hand side of (24) has been estimated by (21). Combining (21), (24), (26) and (27), we get the estimation

$$\left\|\tilde{z}_{k_M+N} - v\right\|_V < \left(L^{-1} + 2\right)\varepsilon. \tag{28}$$

Once again, by repeatedly using (10), for $v \in \operatorname{Fix} T$, we have for all $j \in \mathbb{N}$,

$$\begin{aligned}
\left\|\tilde{z}_{k_M+N+j} - v\right\|_V &= \left\|T\tilde{z}_{k_M+N+j-1} + \varepsilon_{k_M+N+j-1} - v\right\|_V \\
&\leq \left\|\tilde{z}_{k_M+N+j-1} - v\right\|_V + \left\|\varepsilon_{k_M+N+j-1}\right\|_V \\
&\leq \left\|\tilde{z}_{k_M+N} - v\right\|_V + \sum_{p=0}^{j-1} \left\|\varepsilon_{k_M+N+p}\right\|_V.
\end{aligned}$$

Substituting (22) and (28) inter the right hand side of the above inequality gives for all $j \in \mathbb{N}$,

$$\left\|\tilde{z}_{k_M+N+j} - v\right\|_V < \left(L^{-1} + 3\right)\varepsilon, \tag{29}$$

which proves that $\lim_{k\to\infty} \sqrt{V}\tilde{z}_k = \sqrt{V}v \in \sqrt{V} \operatorname{Fix} T$. □

It is easy to check that Theorem 2.15 is a special case of Theorem 2.16, where $V = I_n$.

In this section we propose the framework for inexact Picard fixed-point iterations of quasi-nonexpansive operators Theorems 2.15 and 2.16. In the next section we consider the framework of inexact nonstationary iterations.

2.3 FRAMEWORK OF INEXACT NONSTATIONARY ITERATIONS

In this section we propose the general framework of inexact nonstationary iterations, as an important generalization of inexact fixed-point iterations discussed in Section 2.2. This framework covers the inexact Krasnosel'skiĭ-Mann (KM) iteration (8) and the inexact nonstationary KM iteration (9).

Consider the following inexact nonstationary iteration, designating the following numerical scheme

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \lfloor\ \tilde{z}_{k+1} \leftarrow T_k \tilde{z}_k + \varepsilon_k \end{array} \tag{30}$$

where $T_k : \mathbb{R}^n \to \mathbb{R}^n$ are dependent on iteration number $k \in \mathbb{N}$, $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ is the inexact iteration sequence and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ are local errors. Notice that inexact Picard iterations (7) are special cases of inexact nonstationary iterations (30), where T_k are all identical for $k \in \mathbb{N}$. First we propose a convergence theorem by a generalization of Theorem 2.16. To analyze such inexact nonstationary iterations, we need the following lemma, which could be found in majority of complex analysis textbooks (*e.g.*, [51]).

Lemma 2.17. *Let $\{a_k\}_{k\in\mathbb{N}}$ be a sequence of nonnegative numbers. Then the infinite product $\prod_{k\in\mathbb{N}}(1+a_k) < \infty$ if and only if $\sum_{k\in\mathbb{N}} a_k < \infty$.*

Now we are ready for the following convergence theorem for inexact nonstationary iteration, generalizing inexact Picard fixed-point iteration Theorems 2.15 and 2.16.

Theorem 2.18. *Suppose $C \subseteq \mathbb{R}^n$ is nonempty, $V \in \mathbb{S}^n$, operators $\{T_k\}_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ and $\{\sqrt{V}\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ satisfy*

(i) T_k is L-Lipschitz continuous for $k \in \mathbb{N}$, and there exist $\{L_k\}_{k\in\mathbb{N}} \subset [1, +\infty)$ such that $\sum_{k\in\mathbb{N}}(L_k - 1) < \infty$ and $\|T_k x - z\|_V \le L_k\|x - z\|_V$ for all $x \in \mathbb{R}^n$ and $z \in C$,

(ii) $\{T_k\}_{k\in\mathbb{N}}$ pointwisely converges to an operator $T : \mathbb{R}^n \to \mathbb{R}^n$, and for all $z \in \mathbb{R}^n$, $\{T^k z\}_{k\in\mathbb{N}}$ converges to a point in C,

(iii) $\{\sqrt{V}\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$.

Then for all $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact nonstationary iteration (30) *converges to a point in $\sqrt{V}C$.*

Proof. This proof inherits the one of Theorem 2.16. By Lipschitz continuity of $\{T_k\}_{k\in\mathbb{N}}$, we have that there exists $L > 1$ such that for all $k \in \mathbb{N}$,

$$\|T_k x - T_k y\|_V \le L\|x - y\|_V, \qquad \text{for all } x, y \in \mathbb{R}^n.$$

First we prove that $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ is bounded. Define $K := \prod_{k\in\mathbb{N}} L_k$. By Lemma 2.17 we have $K < \infty$. Then by assumption (i), for all $k \in \mathbb{N}$ and $z \in C$ we have

$$\begin{aligned}\|\tilde{z}_k - z\|_V &\le \|T\tilde{z}_{k-1} - z + \varepsilon_{k-1}\|_V \\ &\le \|T\tilde{z}_{k-1} - z\|_V + \|\varepsilon_k\|_V \\ &\le L_{k-1}\|\tilde{z}_{k-1} - z\|_V + \|\varepsilon_k\|_V.\end{aligned}$$

Recursive application of the above inequality gives us

$$\|\tilde{z}_k - z\|_V \le L_k^{-1}\sum_{l=0}^{k-1}\prod_{m=l+1}^{k} L_m\|\varepsilon_l\|_V + \prod_{l=0}^{k-1} L_l\|\tilde{z}_0 - z\|_V \le K\left(\sum_{l\in\mathbb{N}}\|\varepsilon_l\|_V + \|\tilde{z}_0 - z\|_V\right),$$

which, with assumption (iii), proves the boundedness of $\{\tilde{z}_k\}_{k\in\mathbb{N}}$. Therefore there exists a convergence subsequence $\{\sqrt{V}\tilde{z}_{k_l}\}_{l\in\mathbb{N}}$. Since the range of V is closed, there exists $u \in \mathbb{R}^n$ such that $\sqrt{V}u = \lim_{l\to\infty}\sqrt{V}\tilde{z}_{k_l}$. Then by assumption (ii), we have $v := \lim_{k\to\infty} T^k u$ exist and $v \in C$. Now it is sufficient to prove that $\lim_{k\to\infty}\sqrt{V}\tilde{z}_k = \sqrt{V}v$.

Suppose $\varepsilon > 0$. By definition of v, there exists $N \in \mathbb{N}$ such that

$$\left\|T^N u - v\right\|_V < \varepsilon. \tag{31}$$

Then we prove the desired result by the decomposition

$$\begin{aligned}\tilde{z}_{k_l+N+1} - v = {} & \left(\tilde{z}_{k_l+N+1} - T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}\tilde{z}_l\right) \\ & + \left(T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}\tilde{z}_l - T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}u\right) \\ & + \left(T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}u - T^N u\right) + \left(T^N u - v\right).\end{aligned} \tag{32}$$

For the first term, by Lipschitz continuity of $\{T_k\}_{k\in\mathbb{N}}$, there holds

$$\left\|\tilde{z}_{k_l+N+1} - T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}\tilde{z}_l\right\|_V \le \sum_{m=0}^{N} L^{N-m}\left\|\varepsilon_{k_l+m}\right\|_V,$$

which, together with assumption (iii), means that there exists $M_1 \in \mathbb{N}$ such that for any $l > M_1$ there holds

$$\left\|\tilde{z}_{k_l+N+1} - T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}\tilde{z}_l\right\|_V < \varepsilon. \tag{33}$$

For the second term, notice that $\lim_{l\to\infty}\sqrt{V}\tilde{z}_{k_l} = \sqrt{V}u$, which means that there exists $M_2 \in \mathbb{N}$ such that for all $l > M_2$,

$$\left\|T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}\tilde{z}_l - T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}u\right\|_V \le L^N\left\|\tilde{z}_{k_l} - u\right\|_V < \varepsilon. \tag{34}$$

For the third term, we are going to prove that $T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}u \to T^N u$ as $l \to \infty$. We show this by induction on N. For $N = 1$ it is trivial to see that $\lim_{l\to\infty} T_{k_l}u = Tu$. For $N > 1$, suppose that $\lim_{l\to\infty} T_{k_l+N-1}T_{k_l+N-2}\cdots T_{k_l}u = T^{N-1}u$. Then notice that T_{k_l+N} is Lipschitz continuous, we have

$$\begin{aligned}\left\|T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}u - T^N u\right\|_V &= \left\|T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}u - T_{k_l+N}T^{N-1}u\right\|_V + \left\|T_{k_l+N}T^{N-1}u - T^N u\right\|_V\\ &\le L\left\|T_{k_l+N-1}T_{k_l+N-2}\cdots T_{k_l}u - T^{N-1}u\right\|_V + \left\|T_{k_l+N}T^{N-1}u - T^N u\right\|_V,\end{aligned}$$

whose right-hand side tends to 0 as $l \to \infty$, due to the induction hypothesis and assumption (ii). Therefore there exists $M_3 \in \mathbb{N}$ such that for all $l > M_3$, there holds

$$\left\|T_{k_l+N}T_{k_l+N-1}\cdots T_{k_l}u - T^N u\right\|_V < \varepsilon. \tag{35}$$

Thus, combine (31) to (35), then for all $l > \max\{M_1, M_2, M_3\}$ we have $\|\tilde{z}_{k_l+N+1} - v\| < 4\varepsilon$. Finally, by $v \in C$ and assumption (i), for all $l > \max\{M_1, M_2, M_3\}$ and $m \in \mathbb{N}$ we have

$$\left\|\tilde{z}_{k_l+N+m} - v\right\|_V \le K\left\|\tilde{z}_{k_l+N+1} - v\right\|_V < 4K\varepsilon,$$

which finishes the proof. □

It is clear that Theorem 2.18 covers Theorems 2.15 and 2.16. Conversely, by assumptions (i) and (ii) of Theorem 2.18, we have $\lim_{k\to\infty} L_k = 1$ and $\{T_k\}_{k\in\mathbb{N}}$ is indeed uniformly converging to T, which means that T in Theorem 2.18 actually satisfies conditions of Theorem 2.16.

However there are several important nonstationary iterations that $\{T_k\}_{k\in\mathbb{N}}$ is not converging to any operators, such as KM iterations (8) and nonstationary KM iterations (9). To investigate such kind of inexact nonstationary iterations, we propose another convergence theorem for such kind of inexact nonstationary iterations. In the following theorem, we assume that for any $z \in \mathbb{R}^n$ and $k \in \mathbb{N}$, the exact iteration $\{T_l T_{l-1} \cdots T_k z\}_{l \geq k}$ always converges. In this case, the inexact nonstationary iteration (30) could be illustrated by diagram

$$\begin{array}{ccccccccccccc}
\tilde{z}_0 & \Rightarrow & T_0\tilde{z}_0 & \Rightarrow & T_1T_0\tilde{z}_0 & \Rightarrow & T_2T_1T_0\tilde{z}_0 & \Rightarrow & \cdots & \to & U_0\tilde{z}_0 \\
& & \rightsquigarrow & & & & & & & & \\
& & \tilde{z}_1 & \Rightarrow & T_1\tilde{z}_1 & \Rightarrow & T_2T_1\tilde{z}_1 & \Rightarrow & \cdots & \to & U_1\tilde{z}_1 \\
& & & & \rightsquigarrow & & & & & & \\
& & & & \tilde{z}_2 & \Rightarrow & T_2\tilde{z}_2 & \Rightarrow & \cdots & \to & U_2\tilde{z}_2 \\
& & & & & & \vdots & & & & \vdots
\end{array}$$

where the limit $U_k z := \lim_{l\to\infty} T_l T_{l-1} \cdots T_k z$ for $z \in \mathbb{R}^n$, '$\Rightarrow$' denotes exact iteration, '$\to$' denotes convergence, and '$\rightsquigarrow$' denotes inexact evaluation, with local errors $\{\varepsilon_k\}_{k\in\mathbb{N}}$.

Now we are ready for the convergence theorem of inexact nonstationary iteration (30) where operators $\{T_k\}_{k\in\mathbb{N}}$ are not converging.

Theorem 2.19. *Suppose $C \subset \mathbb{R}^n$ is nonempty and closed, $V \in \mathbb{S}^n$, operators $\{T_k\}_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ satisfy*

(i) T_k is L_k-Lipschitz continuous with respect to V and $\{L_k\}_{k\in\mathbb{N}} \in [1, +\infty)$ such that

$$\sum_{k\in\mathbb{N}} (L_k - 1) < \infty;$$

(ii) For all $k \in \mathbb{N}$ and $z \in \mathbb{R}^n$, $\{T_l T_{l-1} \cdots T_k z\}_{l\geq k}$ converges to a point in C.

(iii) $\{\sqrt{V}\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$.

Then for any $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact nonstationary iteration (30) *converges to a point in $\sqrt{V}C$.*

Proof. Define $L := \prod_{k\in\mathbb{N}} L_k$, and notice that since $\sum_{k\in\mathbb{N}}(L_k - 1) < \infty$, Lemma 2.17 implies that $L < \infty$. Since T_k is L_k-Lipschitz with respect to V, for any $k, m \in \mathbb{N}$ such that $m \geq k$ we have

$$\begin{aligned}
&\|T_m T_{m-1} \cdots T_{k+1}\tilde{z}_{k+1} - T_m T_{m-1} \cdots T_k \tilde{z}_k\|_V \\
&\leq L_m \|T_{m-1} \cdots T_{k+1}\tilde{z}_{k+1} - T_{m-1} \cdots T_k \tilde{z}_k\|_V \\
&\leq \|\tilde{z}_{k+1} - T_k\tilde{z}_k\|_V \prod_{l=k+1}^{m} L_l = \|\varepsilon_k\|_V \prod_{l=k+1}^{m} L_l.
\end{aligned}$$

Therefore, let $m \to \infty$ and recall the notation $U_k z := \lim_{m\to\infty} T_m T_{m-1} \cdots T_k z$, then we have

$$\|U_{k+1}\tilde{z}_{k+1} - U_k \tilde{z}_k\|_V \leq L\|\varepsilon_k\|_V.$$

Then, by assumption $\{\sqrt{V}\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$ and Lemma 2.13, we have that $\{\sqrt{V}U_k\tilde{z}_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence in $\sqrt{V}C$. Since $\sqrt{V}C$ is closed, there exists $z_* \in C$ such that $\lim_{k\to\infty} \sqrt{V}U_k\tilde{z}_k = \sqrt{V}z_*$.

Then it is sufficient to prove $\sqrt{V}\tilde{z}_k \to \sqrt{V}z_*$ as $k \to \infty$. To this end, for $k, N > 0$ we reformulate

$$\begin{aligned}\tilde{z}_{N+k} - z_* = {} & (\tilde{z}_{N+k} - T_{N+k-1}T_{N+k-2}\cdots T_N\tilde{z}_N) \\ & + (T_{N+k-1}T_{N+k-2}\cdots T_N\tilde{z}_N - U_N\tilde{z}_N) + (U_N\tilde{z}_N - z_*),\end{aligned} \tag{36}$$

and we estimate these three terms in right-hand side respectively. For the first term, since $\{\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$, there exists $N_1 \in \mathbb{N}$ such that

$$\sum_{l=N_1}^{\infty} \|\varepsilon_l\|_V < \varepsilon.$$

Therefore for $N > N_1$, by iteration (30), we have the estimation

$$\begin{aligned}&\|\tilde{z}_{N+k} - T_{N+k-1}T_{N+k-2}\cdots T_N\tilde{z}_N\|_V \\ \leq {} & \|\varepsilon_{N+k-1} + T_{N+k-1}\tilde{z}_{N+k-1} - T_{N+k-1}T_{N+k-2}\cdots T_N\tilde{z}_N\|_V \\ \leq {} & \|\varepsilon_{N+k-1}\|_V + L_{N+k-1}\|\tilde{z}_{N+k-1} - T_{N+k-2}\cdots T_N\tilde{z}_N\|_V.\end{aligned}$$

Recursive application of the above inequality gives, for any $N > N_1$ and $k > 0$,

$$\|\tilde{z}_{N+k} - T_{N+k-1}T_{N+k-2}\cdots T_N\tilde{z}_N\|_V \leq L\sum_{l=N}^{N+k-1}\|\varepsilon_l\|_V < L\varepsilon. \tag{37}$$

For the second term, notice that by definition, $\lim_{l\to\infty} T_{N+l}T_{N+l-1}\cdots T_N\tilde{z}_N = U_N\tilde{z}_N$. Then there exists $N_2 > 0$ such that for all $k > N_2$,

$$\|T_{N+k-1}T_{N+k-2}\cdots T_N\tilde{z}_N - U_N\tilde{z}_N\|_V < \varepsilon. \tag{38}$$

For the third term, by the previous assertion, we have $\lim_{N\to\infty} U_N\tilde{z}_N = z_*$. This means that there exists $N_3 > 0$ such that for any $N > N_3$, there holds

$$\|U_N \tilde{z}_N - z_*\|_V < \varepsilon. \tag{39}$$

Finally, combine (36) and estimations (37) to (39), then for all $N > \max\{N_1, N_3\}$ and $k > N_2$, we have

$$\begin{aligned} \|\tilde{z}_{N+k} - z_*\|_V \leq \|\tilde{z}_{N+k} - T_{N+k-1}T_{N+k-2}\cdots T_N\tilde{z}_N\|_V \\ + \|T_{N+k-1}T_{N+k-2}\cdots T_N\tilde{z}_N - U_N\tilde{z}_N\|_V + \|U_N\tilde{z}_N - z_*\|_V < (2+L)\varepsilon, \end{aligned}$$

which proves $\sqrt{V}\tilde{z}_k \to \sqrt{V}z_*$ as $k \to \infty$. □

Notice the difference between Theorems 2.18 and 2.19. Although Theorem 2.19 does not require $\{T_k\}_{k\in\mathbb{N}}$ to be convergent, it requests rather strong Lipschitz properties of operators $\{T_k\}_{k\in\mathbb{N}}$. To clarify this, let's reduce Theorems 2.18 and 2.19 to Picard fixed-point iteration, *i.e.*, $T_k = T$ for $k \in \mathbb{N}$. In this scenario, both frameworks demand $\{T^k z\}_{k\in\mathbb{N}}$ to converge for all $z \in \mathbb{R}^n$, however Theorem 2.18 requires T to be Lipschitz continuous and quasinonexpansive, while Theorem 2.19 requires T to be nonexpansive.

In this section we propose two frameworks for inexact nonstationary iteration (30) under different conditions. In the following section, we consider several applications of the proposed frameworks of inexact iterations.

2.4 APPLICATIONS OF INEXACT METHODS

In this section we apply the frameworks of inexact iterations, Theorems 2.15, 2.16, 2.18 and 2.19, to various fixed-point problems, and discuss the results with the existing inexact methods.

Firstly, guided by the fact that quasinonexpansive operators generalizes nonexpansive operators, we consider the exact/inexact iterations of quasinonexpansive operators parallel to Theorems 2.5 and 2.8. To this end, we need the following propositions. For $V \in \mathbb{S}^n$, denote $V^\dagger$ as the Moore-Penrose inverse of V, *i.e.*, the unique matrix satisfying $VV^\dagger V = V$, $V^\dagger VV^\dagger = V^\dagger$,

$(VV^\dagger)^\top = VV^\dagger$ and $(V^\dagger V)^\top = V^\dagger V$. For $T : \mathbb{R}^n \to \mathbb{R}^n$, define the kernel

$$\operatorname{Ker} T := \{x \in \mathbb{R}^n : T(x+y) = Ty \text{ for all } y \in \mathbb{R}^n\}.$$

When T is linear, $\operatorname{Ker} T$ is identical to the classical definition of the kernel of linear mappings, *i.e.*, $\operatorname{Ker} T = \{x \in \mathbb{R}^n : Tx = 0\}$.

Proposition 2.20. *Suppose $V \in \mathbb{R}^{n\times n}$ and $T : \mathbb{R}^n \to \mathbb{R}^n$. If $\operatorname{Ker} V \subseteq \operatorname{Ker} T$, then $T = TV^\dagger V$, and there hold*

$$\operatorname{Fix} T = TV^\dagger \operatorname{Fix}\big(VTV^\dagger\big) \quad \textit{and} \quad \operatorname{Fix}\big(VTV^\dagger\big) = V \operatorname{Fix} T.$$

Proof. First notice that

$$V\big(I - V^\dagger V\big) = V - VV^\dagger V = V - V = 0,$$

which means $x - V^\dagger V x \in \operatorname{Ker} V \subseteq \operatorname{Ker} T$ for any $x \in \mathbb{R}^n$. This implies $T = TV^\dagger V$.

Suppose $x \in \operatorname{Fix} T$. Define $z := Vx$. Then we have $x = Tx = TV^\dagger z$, and then $z = VTx = VTV^* z$. Therefore we have $z \in \operatorname{Fix}(VTV^\dagger)$. Conversely, suppose $x \in \operatorname{Fix}(VTV^\dagger)$. Denote $z := TV^\dagger x$, then we have $Vz = VTV^\dagger x = x$ and therefore $z = TV^\dagger V z = Tz$, which means $z \in \operatorname{Fix} T$. These arguments proves the identities of fixed-point sets simultaneously. □

Also we need the following proposition of linear combinations of vectors in Hilbert spaces.

Proposition 2.21. *Suppose $V \in \mathbb{S}^n$. For all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, there holds*

$$\|\alpha x + (1-\alpha) y\|_V^2 + \alpha(1-\alpha)\|x - y\|_V^2 = \alpha\|x\|_V^2 + (1-\alpha)\|y\|_V^2.$$

Proof. It is straightforward to check that

$$\begin{aligned}\|\alpha x + (1-\alpha) y\|_V^2 &= \alpha^2\|x\|_V^2 + (1-\alpha)^2\|y\|_V^2 + 2\alpha(1-\alpha)\langle x, y\rangle_V \\ &= \alpha\|x\|_V^2 + (1-\alpha)\|y\|_V^2 - \alpha(1-\alpha)\big(\|x\|_V^2 - 2\langle x, y\rangle_V + \|y\|_V^2\big) \\ &= \alpha\|x\|_V^2 + (1-\alpha)\|y\|_V^2 - \alpha(1-\alpha)\|x - y\|_V^2.\end{aligned}$$ □

The special case of Proposition 2.21 with $V = I$ and $\alpha = 1/2$ is the so-called *Parallelogram identity*, and the special case with $V = I$ and $\alpha = 1/2$, $x = z' - x'$ and $y = z' - y'$ with $x', y', z' \in \mathbb{R}^n$ is the so-called *Apollonius's identity*.

Now we are ready for the convergence theorem of exact KM iteration of quasinonexpansive operators.

Theorem 2.22. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and quasinonexpansive with respect to $V \in \mathbb{S}^n$, $\operatorname{Fix} T \neq \varnothing$, $\operatorname{Ker} V \subseteq \operatorname{Ker} T$ and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,1]$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k(1-\lambda_k) = \infty,$$

then for all $z_0 \in \mathbb{R}^n$, the sequence $\{\sqrt{V} z_k\}_{k\in\mathbb{N}}$ generated by KM iteration (4) converges to a point in $\sqrt{V} \operatorname{Fix} T$.

Proof. Since T is quasinonexpansive with respect to V, by (20) we have

$$\|Tz_k - z\|_V \leq \|z_k - z\|_V, \qquad \text{for all } k \in \mathbb{N}, z \in \operatorname{Fix} T.$$

Then by $z_{k+1} = (1-\lambda_k)z_k + \lambda_k T z_k$ and Proposition 2.21, for $k \in \mathbb{N}$ and $z \in \operatorname{Fix} T$ we have

$$\begin{aligned}
\|z_{k+1} - z\|_V^2 &= \|(1-\lambda_k)(z_k - z) + \lambda_k(Tz_k - z)\|_V^2 \\
&= (1-\lambda_k)\|z_k - z\|_V^2 + \lambda_k\|Tz_k - z\|_V^2 - \lambda_k(1-\lambda_k)\|Tz_k - z_k\|_V^2 \\
&\leq \|z_k - z\|_V^2 - \lambda_k(1-\lambda_k)\|Tz_k - z_k\|_V^2.
\end{aligned}$$

This ensures that $\{\sqrt{V} z_k\}_{k\in\mathbb{N}}$ is bounded and $\sum_{k\in\mathbb{N}} \lambda_k(1-\lambda_k)\|Tz_k - z_k\|_V < \infty$. By assumption $\sum_{k\to\infty} \lambda_k(1-\lambda_k) = \infty$, we have $\liminf_{k\to\infty} \|Tz_k - z_k\|_V = 0$. Thus there exists a convergent subsequence $\{\sqrt{V} z_{k_l}\}_{l\in\mathbb{N}}$ such that $\lim_{l\to\infty} \|Tz_{k_l} - z_{k_l}\|_V = 0$. Define $u := \lim_{l\to\infty} \sqrt{V} z_{k_l}$. Then by $\operatorname{Ker} \sqrt{V} = \operatorname{Ker} V \subseteq \operatorname{Ker} T$, continuity of T and Proposition 2.20, we have

$$\begin{aligned}
0 &= \lim_{l\to\infty}\left(\sqrt{V} T z_{k_l} - \sqrt{V} z_{k_l}\right) \\
&= \lim_{l\to\infty} \sqrt{V} T \sqrt{V}^{\dagger} \sqrt{V} z_{k_l} - u \\
&= \sqrt{V} T \sqrt{V}^{\dagger} u - u,
\end{aligned}$$

which proves that $u \in \operatorname{Fix}(\sqrt{V} T \sqrt{V}^{\dagger})$. Therefore by Proposition 2.20, there exists $z_* \in \operatorname{Fix} T$ such that $u = \sqrt{V} z_*$. This means that $\lim_{l\to\infty} \sqrt{V} z_{k_l} = \sqrt{V} z_*$. Finally, by $\|z_{k+1} - z_*\|_V \leq \|z_k - z_*\|_V$ and $\lim_{l\to\infty} \|z_{k_l} - z_*\|_V = 0$, we have $\lim_{l\to\infty} \sqrt{V} z_k = \sqrt{V} z_*$. □

Theorem 2.22 has a direct corollary in exact Picard iteration (3) of quasinonexpansive operators. To this end we need the following proposition.

Proposition 2.23. *If $\{x_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ converges, then for any $\lambda \in (0,2)$ and $y_0 \in \mathbb{R}^n$, the sequence $\{y_k\}_{k\in\mathbb{N}}$ defined by*

$$y_{k+1} = (1-\lambda)y_k + \lambda x_k, \qquad \text{for all } k \in \mathbb{N}$$

converges to the same limit of $\{x_k\}_{k\in\mathbb{N}}$.

Proof. Denote $x := \lim_{k\to\infty} x_k$. Case $\lambda = 1$ is trivial. Suppose $\lambda \neq 1$. Notice that for $k \in \mathbb{N}$, by definition we have

$$y_k = (1-\lambda)^k y_0 + \lambda(1-\lambda)^k \sum_{i=0}^{k-1} \frac{x_i}{(1-\lambda)^{i+1}}.$$

Given $\varepsilon > 0$. Then since $\lim_{k\to\infty} x_k = x$, there exist $M > 0$ and $N \in \mathbb{N}$ such that $\sup_{k\in\mathbb{N}} \|x_k\| < M$ and for all $k > N$ there holds $\|x_k - x\| < \varepsilon$. Then since $|1-\lambda| < 1$, there exists $K \in \mathbb{N}$ such that for all $k > K$, there holds $|1-\lambda|^k \max\{1, (|1-\lambda|^{-N} - 1)\} < \varepsilon$. Therefore, for all $k > \max\{N, K\}$ we have

$$\begin{aligned}
\|y_k - x\| &\le |1-\lambda|^k \|y_0 - x\| + \lambda |1-\lambda|^k \sum_{i=0}^{k-1} \frac{\|x_k - x\|}{|1-\lambda|^{i+1}} \\
&\le \varepsilon \|y_0 - x\| + 2|1-\lambda|^k \left(\sum_{i=0}^{N-1} \frac{M}{|1-\lambda|^{i+1}} + \sum_{j=N}^{k-1} \frac{\varepsilon}{|1-\lambda|^{j+1}} \right) \\
&\le \varepsilon \|y_0 - x\| + 2\frac{1+M}{1-|1-\lambda|}\varepsilon,
\end{aligned}$$

which proves that $\lim_{k\to\infty} y_k = x$. □

Then we are ready for the convergence theorem of exact Picard iteration (3) of quasinonexpansive operators. For $T : \mathbb{R}^n \to \mathbb{R}^n$ define the range of T as $\operatorname{Ran} T := \{Tx : x \in \mathbb{R}^n\}$.

Corollary 2.24. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and quasinonexpansive with respect to $V \in \mathbb{S}^n$, $\operatorname{Fix} T \neq \varnothing$, $\operatorname{Ker} V \subseteq \operatorname{Ker} T$. If $\lambda \in (0,1)$, then for all $z_0 \in \mathbb{R}^n$, the sequence $\{z_k\}_{k\in\mathbb{N}}$ generated by Picard iteration* (3) *converges to a point in* $\operatorname{Fix} T$.

Proof. By Theorem 2.22, we have $\{\sqrt{V} z_k\}_{k\in\mathbb{N}}$ generated by Picard iteration (3) converges to a point in $\sqrt{V}\operatorname{Fix} T$. Let $z_* \in \operatorname{Fix} T$ such that $\sqrt{V} z_* = \lim_{k\to\infty} \sqrt{V} z_k$. Then by continuity of T in $\operatorname{Ran} V$,

$$\lim_{k\to\infty} Tz_k = \lim_{k\to\infty} T\sqrt{V}^\dagger \sqrt{V} z_k = T\sqrt{V}^\dagger \lim_{k\to\infty} \sqrt{V} z_k = T\sqrt{V}^\dagger \sqrt{V} z_* = Tz_* = z_*.$$

Therefore, by Proposition 2.23, we have $\{z_k\}_{k\in\mathbb{N}}$ also converges to $z_* \in \operatorname{Fix} T$. □

The convergence of inexact KM iteration of nonexpansive operators could be obtained as a direct corollary of Theorems 2.19 and 2.22.

Theorem 2.25. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be nonexpansive with respect to $V \in \mathbb{S}^n$, $\operatorname{Ker} V \subseteq \operatorname{Ker} T$ and $\operatorname{Fix} T \neq \varnothing$, $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,1]$ and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$. If*

$$\left\{\sqrt{V}\lambda_k\varepsilon_k\right\}_{k\in\mathbb{N}} \in \ell^1 \quad \text{and} \quad \sum_{k\in\mathbb{N}} \lambda_k(1-\lambda_k) = \infty,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the sequence $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact KM iteration (8) *converges to a point in $\sqrt{V}\operatorname{Fix} T$.*

Proof. Define $\tilde{u}_k := \sqrt{V}\tilde{z}_k$ for $k \in \mathbb{N}$. Then by Proposition 2.20 we have

$$\begin{aligned}\tilde{u}_{k+1} &= \sqrt{V}((1-\lambda_k)\tilde{z}_k + \lambda_k T\tilde{z}_k + \lambda_k\varepsilon_k) \\ &= (1-\lambda_k)\tilde{u}_k + \lambda_k\sqrt{V}T\sqrt{V}^\dagger\tilde{u}_k + \sqrt{V}\lambda_k\varepsilon_k.\end{aligned}$$

Define $T' := \sqrt{V}T\sqrt{V}^\dagger$ and $T'_k := (1-\lambda_k)I + \lambda_k T'$ for $k \in \mathbb{N}$. Notice that by quasinonexpansiveness of T, for any $x, y \in \mathbb{R}^n$ we have

$$\begin{aligned}\|T'x - T'y\| &= \left\|\sqrt{V}T\sqrt{V}^\dagger x - \sqrt{V}T\sqrt{V}^\dagger y\right\| \\ &= \left\|T\sqrt{V}^\dagger x - T\sqrt{V}^\dagger y\right\|_V \\ &\leq \left\|\sqrt{V}^\dagger x - \sqrt{V}^\dagger y\right\|_V = \left\|\sqrt{V}\sqrt{V}^\dagger(x-y)\right\| \leq \|x-y\|,\end{aligned}$$

which proves that T' is nonexpansive. This, with $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,1]$, also proves that $\{T'_k\}_{k\in\mathbb{N}}$ are nonexpansive. Then, by Theorem 2.22, for any $u \in \mathbb{R}^n$ and $k \in \mathbb{N}$, the exact iteration sequence $\{T'_l T'_{l-1} \cdots T'_k u\}_{l\geq k}$ always converges to a point in $\operatorname{Fix} T'$. Also, notice that $\operatorname{Fix} T'$ is closed due to the continuity of T'. Therefore, by assumptions and Theorem 2.19 we have, for any $\tilde{u}_0 \in \mathbb{R}^n$, the inexact sequence $\{\tilde{u}_k\}_{k\in\mathbb{N}}$ always converges to a point in $\operatorname{Fix} T'$. This means that, for any $\tilde{z}_0 \in \mathbb{R}^n$, the sequence $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact KM iteration (8) converges to a point in $\sqrt{V}\operatorname{Fix} T$. □

Notice that Theorem 2.25 covers Theorem 2.8 and therefore Theorems 2.1 and 2.5. Similar to Theorem 2.22 and Corollary 2.24, Theorem 2.25 has its corresponding corollary in the inexact Picard iteration

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \lfloor\ \tilde{z}_{k+1} \leftarrow (1-\lambda)\tilde{z}_k + \lambda(T\tilde{z}_k + \varepsilon_k)\end{array} \tag{40}$$

where $\lambda \in \mathbb{R}$.

Corollary 2.26. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be nonexpansive with respect to $V \in \mathbb{S}^n$, $\operatorname{Ker} V \subseteq \operatorname{Ker} T$ and $\operatorname{Fix} T \neq \varnothing$, $\lambda \subset (0,1)$ and $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$. If*

$$\left\{\sqrt{V}\varepsilon_k\right\}_{k\in\mathbb{N}} \in \ell^1,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the sequence $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact Picard iteration (40) converges to a point in $\sqrt{V}\operatorname{Fix} T$. If further $\lim_{k\to\infty} \varepsilon_k = 0$, then $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ converges to a point in $\operatorname{Fix} T$.

Proof. The convergence of $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ directly follows Theorem 2.25. Let $z_* \in \operatorname{Fix} T$ such that $\lim_{k\to\infty} \sqrt{V}\tilde{z}_k = \sqrt{V}z_*$. If further $\lim_{k\to\infty} \varepsilon_k = 0$, then since

$$\lim_{k\to\infty}(T\tilde{z}_k + \varepsilon_k) = \lim_{k\to\infty} T\sqrt{V}^{\dagger}\sqrt{V}\tilde{z}_k = T\sqrt{V}^{\dagger} \lim_{k\to\infty} \sqrt{V}z_k = T\sqrt{V}^{\dagger}\sqrt{V}z_* = Tz_* = z_*,$$

by Proposition 2.23, we have that $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ converges to $z_* \in \operatorname{Fix} T$. □

Then we consider the inexact Picard/KM iterations for quasinonexpansive operators, which generates the results for nonexpansive operators. First we consider the inexact Picard iteration for quasinonexpansive operators.

Theorem 2.27. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous and quasinonexpansive with respect to $V \in \mathbb{S}^n$, $\operatorname{Fix} T \neq \varnothing$, $\operatorname{Ker} V \subseteq \operatorname{Ker} T$ and $\lambda \in (0,1)$. If*

$$\{\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact Picard iteration (40) converges to a point in $\operatorname{Fix} T$.

Proof. Define $T_\lambda := (1-\lambda)I + \lambda T$. First notice that $\operatorname{Fix} T = \operatorname{Fix} T_\lambda$ and, by Corollary 2.24, for any $k \in \mathbb{N}$ and $z \in \mathbb{R}^n$, the exact Picard sequence $\{T_\lambda^l z\}_{l\geq k}$ converges to a point in $\operatorname{Fix} T$. Also notice that since T is Lipschitz continuous, so does T_λ. Then a direct application of Theorem 2.15 proves the convergence of $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ towards $\operatorname{Fix} T$. □

Also we have the following theorem for inexact KM iteration for quasinonexpansive operators as a corollary of Theorems 2.19 and 2.22.

Theorem 2.28. *Suppose $V \in \mathbb{S}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz continuous with respect to V and quasinonexpansive with respect to V, $\operatorname{Fix} T \neq \varnothing$, $\operatorname{Ker} V \subseteq \operatorname{Ker} T$ and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,1]$. If*

$$\{\lambda_k \varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1 \quad \textit{and} \quad (L-1)_+ \sum_{k\in\mathbb{N}} \lambda_k < \infty,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact KM iteration (8) converges to a point in $\operatorname{Fix} T$.

Proof. Define $T_k := (1-\lambda_k)I + \lambda_k T$ for $k \in \mathbb{N}$. First notice that by Corollary 2.24, for any $k \in \mathbb{N}$ and $z \in \mathbb{R}^n$, the exact Picard sequence $\{T_l' T_{l-1}' \cdots T_k' z\}_{l\geq k}$ converges to a point in $\operatorname{Fix} T$. Also notice that since T is L-Lipschitz continuous with respect to V, we have T_k is $(1+(L-1)\lambda_k)$-Lipschitz with respect to V. Then a direct application of Theorem 2.19 proves the convergence of $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ towards $\operatorname{Fix} T$. □

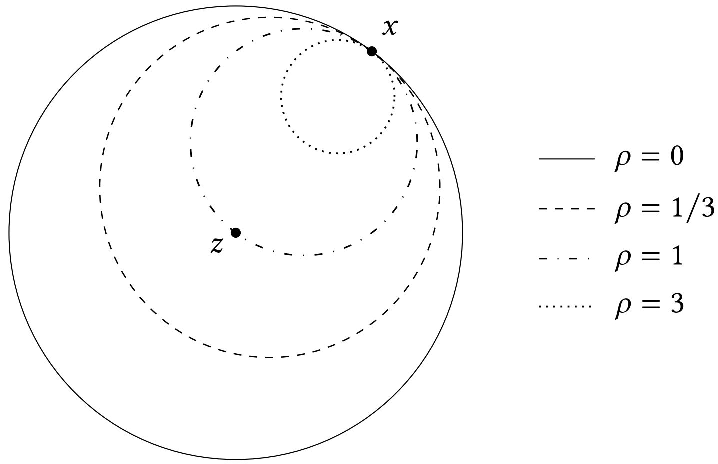
Notice that Theorem 2.28 is not covering Theorem 2.27.

So far we have discussed the convergence of exact/inexact Picard/KM iterations of nonexpansive/quasinonexpansive operators. Besides the above results, the proposed framework also leads to an important generalization of nonexpansive averaged operators and the corresponding inexact iterations. Such generalized class of operators would be of fundamental status in the following discussion of nonsmooth convex optimization. For $V, U \in \mathbb{S}^n$, an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is *quasiaveraged* with respect to (V, U) if

$$\|Tx - z\|_V^2 + \|Tx - x\|_U^2 \leq \|x - z\|_V^2, \qquad \text{for all } x \in \mathbb{R}^n, z \in \operatorname{Fix} T. \tag{41}$$

We further define the class of *firmly quasinonexpansive operators* to be all quasiaveraged operators with respect to (V, V), where $V \in \mathbb{S}^n$. Figure 1 illustrates several cases of quasiaveraged operators in $\mathbb{R}^2$ with respect to $(I_2, \rho I_2)$ for different $\rho \geq 0$.

Figure 1. Possible regions of Tx with $T : \mathbb{R}^2 \to \mathbb{R}^2$ is quasiaveraged with respect to $(I_2, \rho I_2)$ and $z \in \operatorname{Fix} T$



Quasiaveraged operators extends several important class of operators. First notice that nonexpansive averaged operators are special cases of quasiaveraged operators. To see this, observe that an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive κ-averaged with $\kappa \in (0, 1)$ if and only if

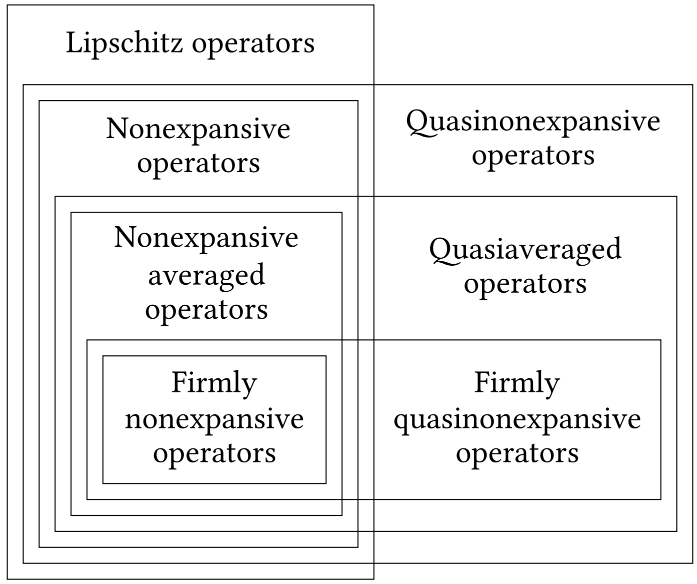
$$\|Tx - Ty\|^2 + \frac{1-\kappa}{\kappa}\|(Tx - Ty) - (x - y)\|^2 \le \|x - y\|^2, \qquad \text{for all } x, y \in \mathbb{R}^n. \tag{42}$$

To prove this, one only need to notice that $(T - (1 - \kappa)I)/\kappa$ is nonexpansive. Restricting $y \in \operatorname{Fix} T$ in (42) clearly shows that all nonexpansive κ-averaged operators are quasiaveraged with respect to $(\kappa I, (1 - \kappa)I)$. Specially, quasiaveraged operators with respect to $(I/2, I/2)$ (equivalent to quasiaveraged with respect to (I, I), also equivalent to firmly quasinonexpansive with respect to I) are also known as *firmly quasinonexpansive* operators [5, 6]. Also, notice that quasiaveraged operators with respect to (V, U) are always quasinonexpansive with respect to V. The relations between classes of operators introduced in this chapter could be summarized as in Figure 2. The following example shows a quasiaveraged operator while not being nonexpansive.

Example 2.29. Define $T : \mathbb{R} \to \mathbb{R}$ as

$$T(x) := \begin{cases} \pi^{-1}|x| \arctan(x)(\sin(\ln|x|) + 1), & x \neq 0, \\ 0, & x = 0, \end{cases}$$

Figure 2. Relations between classes of operators



then T is quasiaveraged with respect to (I, I) and Lipschitz continuous, but not nonexpansive.

Then we propose the convergence theorem for exact/inexact Picard iteration (1) of quasiaveraged operators.

Theorem 2.30. *Suppose* $V, U \in \mathbb{S}^n$, $T : \mathbb{R}^n \to \mathbb{R}^n$ *is continuous, quasiaveraged with respect to* (V, U) *and* $\operatorname{Fix} T \neq \varnothing$. *If*

$$\operatorname{Ker} V \cup \operatorname{Ker} U \subseteq \operatorname{Ker} T,$$

then for all $z_0 \in \mathbb{R}^n$, *the sequence* $\{z_k\}_{k \in \mathbb{N}}$ *generated by Picard iteration* (1) *converges to a point in* $\operatorname{Fix} T$.

Proof. Since T is quasiaveraged with respect to (V, U), by (41) we have

$$\|z_{k+1} - z\|_V^2 + \|z_{k+1} - z_k\|_U^2 \leq \|z_k - z\|_V^2, \qquad \text{for all } k \in \mathbb{N}, z \in \operatorname{Fix} T. \tag{43}$$

This ensures that $\{\sqrt{V} z_k\}_{k \in \mathbb{N}}$ is bounded and $\lim_{k \to \infty} \sqrt{U}(z_{k+1} - z_k) = 0$. Thus there exists a convergent subsequence $\{\sqrt{V} z_{k_l}\}_{l \in \mathbb{N}}$. Since range of $\sqrt{V}$ is closed, there exists $u \in \mathbb{R}^n$ such that $\lim_{l \to \infty} \sqrt{V} z_{k_l} = \sqrt{V} u$. Then by $\operatorname{Ker} \sqrt{V} = \operatorname{Ker} V \subseteq \operatorname{Ker} T$, continuity of T and Proposition 2.20, we have

$$z_{k_l+1} = Tz_{k_l} = T\sqrt{V}^\dagger\sqrt{V}z_{k_l} \to T\sqrt{V}^\dagger\sqrt{V}u = Tu, \qquad \text{as } l \to \infty.$$

Denote $z_* := Tu$. By $\lim_{l\to\infty}\sqrt{U}(z_{k_l+1}-z_{k_l}) = 0$, we have $\lim_{l\to\infty}\sqrt{U}z_{k_l} = \sqrt{U}z_*$. Then by continuity of T and Proposition 2.20, we have

$$z_* = \lim_{l\to\infty} z_{k_l+1} = \lim_{l\to\infty} Tz_{k_l} = \lim_{l\to\infty} T\sqrt{U}^\dagger\sqrt{U}z_{k_l} = T\sqrt{U}^\dagger\sqrt{U}z_* = Tz_*,$$

which proves that $z_* \in \operatorname{Fix} T$. Finally, substituting z by z_* in (43) proves that $\lim_{k\to\infty}\sqrt{V}z_k = \sqrt{V}z_*$, which further implies that

$$z_{k+1} = Tz_k = T\sqrt{V}^\dagger\sqrt{V}z_k \to T\sqrt{V}^\dagger\sqrt{V}z_* = Tz_* = z_*, \qquad \text{as } k \to \infty.$$

This finishes the proof. □

Theorem 2.30 for exact Picard iterations of quasiaveraged operators obviously generalizes Theorem 2.2 for nonexpansive averaged operators. It has the following inexact version via Theorem 2.15.

Corollary 2.31. *Suppose $V, U \in \mathbb{S}^n$, $T : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous, quasiaveraged with respect to (V, U), $\operatorname{Fix} T \neq \varnothing$. If local errors $\{\sqrt{V}\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$ and*

$$\operatorname{Ker} V \cup \operatorname{Ker} U \subseteq \operatorname{Ker} T,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the sequence $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact Picard iteration (40) converges to a point in $\sqrt{V}\operatorname{Fix} T$. If further $\lim_{k\to\infty}\varepsilon_k = 0$, then $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ converges to a point in $\operatorname{Fix} T$.

Proof. This proof direct follows Proposition 2.20 and Theorems 2.15 and 2.30. By Theorem 2.30 we have for all $z \in \mathbb{R}^n$ the exact Picard iteration sequence $\{T^k z\}_{k\in\mathbb{N}}$ converges to a point in $\operatorname{Fix} T$. Then by Theorem 2.15 we have $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ converges to $v \in \sqrt{V}\operatorname{Fix} T$. Finally, if further $\lim_{k\to\infty}\varepsilon_k = 0$, then by continuity of T and Proposition 2.20 we have

$$\tilde{z}_{k+1} = T\tilde{z}_k + \varepsilon_k = T\sqrt{V}^\dagger\sqrt{V}\tilde{z}_k + \varepsilon_k \to T\sqrt{V}^\dagger v, \qquad \text{as } k \to \infty,$$

and by Proposition 2.20, $T\sqrt{V}^\dagger v \in T\sqrt{V}^\dagger \operatorname{Fix}(\sqrt{V}T\sqrt{V}^\dagger) = \operatorname{Fix} T$. □

For exact/inexact KM iterations of quasiaveraged operators, convergence theorems would be rather complicated depending on the relation between V and U. Here we merely consider the KM iterations of quasiaveraged operators with respect to $(V, \rho V)$ for some $\rho > 0$.

Theorem 2.32. *Suppose $V \in \mathbb{S}^n$ and $\rho > 0$, $T : \mathbb{R}^n \to \mathbb{R}^n$ is continuous with respect to V, and quasiaveraged with respect to $(V, \rho V)$, $\operatorname{Fix} T \neq \emptyset$, $\operatorname{Ker} V \subseteq \operatorname{Ker} T$ and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 1+\rho]$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k(\rho + 1 - \lambda_k) = \infty,$$

then for any $z_0 \in \mathbb{R}^n$, the sequence $\{\sqrt{V} z_k\}_{k\in\mathbb{N}}$ generated by exact KM iteration (4) converges to a point in $\sqrt{V} \operatorname{Fix} T$.

Proof. By quasiaveragedness of T, for all $z_* \in \operatorname{Fix} T$ we have

$$\begin{aligned}
\|z_{k+1} - z_*\|_V^2 &= \|(1-\lambda_k)(z_k - z_*) + \lambda_k(Tz_k - z_*)\|_V^2 \\
&= (1-\lambda_k)\|z_k - z_*\|_V^2 + \lambda_k\|Tz_k - z_*\|_V^2 - \lambda_k(1-\lambda_k)\|Tz_k - z_k\|_V^2 \\
&\leq \|z_k - z_*\|_V^2 - \lambda_k(\rho + 1 - \lambda_k)\|Tz_k - z_k\|_V^2.
\end{aligned}$$

Then a similar analysis of Theorem 2.22 will prove that for any $z_0 \in \mathbb{R}^n$, the sequence $\{\sqrt{V} z_k\}_{k\in\mathbb{N}}$ converges to a point of $\sqrt{V} \operatorname{Fix} T$. □

Theorem 2.32 clearly covers and extends Theorems 2.1, 2.5 and 2.22 and Corollaries 2.4 and 2.6. It has the following inexact version via Theorem 2.15, which covers and extends Theorem 2.8 and Corollary 2.9. Define $(\cdot)_+ := \max\{0, \cdot\}$ as the Rectified Linear Unit (ReLU) activation function in $\mathbb{R}$.

Theorem 2.33. *Suppose $V \in \mathbb{S}^n$ and $L, \rho > 0$, $T : \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz continuous with respect to V, and quasiaveraged with respect to $(V, \rho V)$, $\operatorname{Fix} T \neq \emptyset$, $\operatorname{Ker} V \subseteq \operatorname{Ker} T$ and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 1+\rho]$. If*

$$\left\{\sqrt{V}\lambda_k\varepsilon_k\right\}_{k\in\mathbb{N}} \in \ell^1, \quad \sum_{k\in\mathbb{N}} \lambda_k(\rho+1-\lambda_k) = \infty \quad \textit{and} \quad \sum_{k\in\mathbb{N}} ((L-1)\lambda_k + 2(\lambda_k - 1)_+)_+ < \infty,$$

then for any $\tilde{z}_0 \in \mathbb{R}^n$, the sequence $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact KM iteration (8) converges to a point in $\sqrt{V} \operatorname{Fix} T$.

Table 1. Convergence theorems by iteration types and operator classes

Operator Class	Exact Iteration		Inexact Iteration	
	Picard	KM	Picard	KM
Nonexpansive	Corollary 2.3	Theorem 2.5	Corollary 2.26	Theorem 2.25
Nonexp. Averaged	Corollary 2.4	Corollary 2.6	Corollary 2.10	Corollary 2.9
Quasinonexpansive	Corollary 2.24	Theorem 2.22	Theorem 2.27	Theorem 2.28
Quasiaveraged	Theorem 2.30	Theorem 2.32	Corollary 2.31	Theorem 2.33

Proof. This proof is totally analogous to the one of Theorem 2.25. One only need to further notice that by nonexpansiveness of T, for any $\lambda > 0$ and $x, y \in \mathbb{R}^n$,

$$\begin{aligned}\|(1-\lambda)(x-y) + \lambda(Tx - Ty)\|_V &\le (|1-\lambda| + \lambda L)\|x-y\|_V \\ &= (1 + (L-1)\lambda + 2(\lambda-1)_+)\|x-y\|_V,\end{aligned}$$

which proves that $T'_k := (1-\lambda_k)I + \lambda_k T$ is $(1+(L-1)\lambda_k + 2(\lambda_k - 1)_+)$-Lipschitz continuous. Then a direct application of Theorem 2.19 proves the conclusion. □

The convergence theorems of various classes of operators presented in this chapter so far are summarized in Table 1.

To close this section we consider the convergence theorem of inexact nonstationary KM iteration (9) as a direct corollary of Theorem 2.19, which generalizes Theorem 2.11.

Theorem 2.34. *Assume that* $T_\Gamma : \mathbb{R}^n \to \mathbb{R}^n$ *is nonexpansive with* $\operatorname{Fix} T_\Gamma \neq \varnothing$*, and*

(i) T_{Γ_k,λ_k} *is* L_k*-Lipschitz with* $\{L_k\}_{k\in\mathbb{N}} \subset [1,+\infty)$ *and* $\sum_{k\in\mathbb{N}}(L_k - 1) < \infty$*,*

(ii) $\{\lambda_k\}_{k\in\mathbb{N}} \subset (0,1)$ *with* $\sum_{k\in\mathbb{N}} \lambda_k(1-\lambda_k) = \infty$*,*

(iii) $\{\lambda_k \varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$*,*

(iv) $\{\lambda_k \Delta_{k,\rho}\}_{k\in\mathbb{N}} \in \ell^1$ *for all* $\rho \ge 0$*.*

Then for any $\tilde{z}_0 \in \mathbb{R}^n$*, the inexact sequence* $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ *generated by inexact nonstationary KM iteration* (9) *converges to a point in* $\operatorname{Fix} T_\Gamma$*.*

Proof. Rewrite nonstationary KM iteration (9) into stationary form

$$\tilde{z}_{k+1} = T_{\Gamma,\lambda_k}\tilde{z}_k + \lambda_k \eta_k, \qquad \text{for all } k \in \mathbb{N},$$

where $T_{\Gamma,\lambda_k} := (1-\lambda_k)I+\lambda_k T_\Gamma$ and $\eta_k := T_{\Gamma_k}\tilde{z}_k - T_\Gamma \tilde{z}_k + \varepsilon_k$ are local errors. By assumptions and Theorem 2.5 we have, for any $k \in \mathbb{N}$ and $z \in \mathbb{R}^n$, the exact stationary KM iteration $\{T_{\Gamma,\lambda_l} T_{\Gamma,\lambda_{l-1}} \cdots T_{\Gamma,\lambda_k} z\}_{l\geq k}$ converges to a point in $\operatorname{Fix} T_\Gamma$. Also notice that $\operatorname{Fix} T_\Gamma = \operatorname{Fix} T_{\Gamma,\lambda_k}$ for $k \in \mathbb{N}$. Furthermore, by Lipschitz continuity of T_{Γ_k,λ_k} we have, for any $z_* \in \operatorname{Fix} T_\Gamma$,

$$\begin{aligned}\|\tilde{z}_{k+1} - z_*\| &= \left\|T_{\Gamma_k,\lambda_k}\tilde{z}_k + \lambda_k\eta_k - T_{\Gamma,\lambda_k} z_*\right\| \\ &\leq \left\|T_{\Gamma_k,\lambda_k}\tilde{z}_k - T_{\Gamma_k,\lambda_k} z_*\right\| + \left\|T_{\Gamma_k,\lambda_k} z_* - T_{\Gamma,\lambda_k} z_*\right\| + \lambda_k\|\eta_k\| \\ &\leq L_k\|\tilde{z}_k - z_*\| + \lambda_k\Delta_{k,\|z_*\|} + \lambda_k\|\eta_k\|,\end{aligned}$$

which by assumptions, means $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ is bounded. Define $\tilde{\rho} := \sup_{k\in\mathbb{N}}\|\tilde{z}_k\|$, then we have

$$\lambda_k\|\eta_k\| = \lambda_k\left\|T_{\Gamma_k}\tilde{z}_k - T_\Gamma\tilde{z}_k + \varepsilon_k\right\| \leq \lambda_k\left(\Delta_{k,\tilde{\rho}} + \|\varepsilon_k\|\right),$$

right-hand side of which is summable by assumptions, and therefore $\{\lambda_k\eta_k\}_{k\in\mathbb{N}} \in \ell^1$. So far we have proved the conditions of Theorem 2.19 obtain, and the conclusion consequently follows. □

Obviously Theorem 2.34 covers and refines Theorem 2.11, which illustrates the generality of the proposed framework of inexact iterations. The following is the corresponding corollary parallel to Corollary 2.12, proof of which is not repeated herein.

Corollary 2.35. *Assume that R_Γ and $\{R_{\Gamma_k}\}_{k\in\mathbb{N}}$ are nonexpansive operators in $\mathbb{R}^n$ with $\operatorname{Fix} R_\Gamma \neq \varnothing$, and followings hold*

(i) $T_{\Gamma_k} = (1-\kappa_k)I + \kappa_k R_{\Gamma_k}$ *is nonexpansive κ_k-averaged with $\kappa_k \in (0,1]$ for $k \in \mathbb{N}$,*

(ii) $\{\lambda_k\}_{k\in\mathbb{N}} \subset (0, 1/\kappa_k)$ *with* $\sum_{k\in\mathbb{N}} \lambda_k\kappa_k(1-\lambda_k\kappa_k) = \infty$,

(iii) $\{\lambda_k\kappa_k\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$,

(iv) $\{\lambda_k\kappa_k\Delta^R_{k,\rho}\}_{k\in\mathbb{N}} \in \ell^1$ *for all* $\rho \geq 0$.

Then for any $\tilde{z}_0 \in \mathbb{R}^n$, the inexact sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by inexact nonstationary KM iteration (9) *converges to a point in* $\operatorname{Fix} T_\Gamma$.

Theorem 2.34 and Corollary 2.35 could be generalized to quasinonexpansive/quasiaveraged operator respectively via Theorems 2.28 and 2.33 with ease. As an end to the topic of inexact iterations, we propose a convergence theorem of inexact KM iteration of nonexpansive/quasiaveraged operators, which would be found applicable later in the following topic of the dissertation.

Theorem 2.36. *Assume that* $V \in \mathbb{S}^n$, $\rho > 0$ *and* $T : \mathbb{R}^n \to \mathbb{R}^n$ *is nonexpansive with respect to* V *and quasiaveraged with respect to* $(V, \rho V)$ *with closed* $\operatorname{Fix} T \neq \emptyset$*, and followings hold*

(i) $(1-\lambda)I + \lambda T$ *is nonexpansive with respect to* V *with* $\lambda \in (0, 1+\rho]$,

(ii) $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 1+\rho]$ *with* $\sum_{k\in\mathbb{N}} \lambda_k(1+\rho-\lambda_k) = \infty$,

(iii) $\{\sqrt{V}\lambda_k\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$.

Then for any $\tilde{z}_0 \in \mathbb{R}^n$*, the inexact sequence* $\{\sqrt{V}\tilde{z}_k\}_{k\in\mathbb{N}}$ *generated by inexact KM iteration* (8) *converges to a point in* $\sqrt{V}\operatorname{Fix} T$.

Proof. The proof directly follows Theorem 2.33. One only need to notice that the condition $\sum_{k\in\mathbb{N}}((L-1)\lambda_k + 2(\lambda_k - 1)_+)_+ < \infty$ in Theorem 2.33 is replaced by condition (i). □

CHAPTER 3

IMPLICIT FIXED-POINT PROXIMITY ALGORITHMS FOR NONSMOOTH CONVEX OPTIMIZATION

This chapter introduces the nonsmooth convex optimization problem of the main concern in this dissertation and the framework of implicit fixed-point proximity algorithms which solves the designated optimization problem. The main nonsmooth convex optimization problem of this dissertation is firstly clarified with plenty of important applications arising from various scopes in data science, such as inverse problems, image processing, machine learning and distributed computing. Then we review some useful preliminaries in convex analysis and operator theory, of which we make use to characterize solutions of the main model as fixed-points of a proximity equation and further analysis. Due to the expanding nature of the fixed-point equation, the naïve Picard/KM iteration does not necessarily converge. We overcome this difficulty by a full application of the *matrix splitting technique* to the matrices involved in the fixed-point equations, which leads us to a general framework of implicit fixed-point proximity algorithms with convergence theorem. This framework covers and generalizes most of the popular explicit algorithms, including gradient descent method, proximal point method, Douglas-Rachford splitting algorithm, first-order primal-dual algorithm, primal-dual hybrid gradient method, fixed-point proximity algorithm, alternating direction method of multipliers, split Bregman iteration, linearized alternating direction method of multipliers and inexact Uzawa method.

3.1 NONSMOOTH CONVEX OPTIMIZATION PROBLEM

We model the general composed convex optimization problem as

$$\operatorname*{argmin}_{x\in\mathbb{R}^n}\{f_1(A_1x)+f_2(A_2x)\}, \tag{44}$$

where argmin f denotes the set of global minimizers of f, $f_1 \in \Gamma_0(\mathbb{R}^{n_1})$, $f_2 \in \Gamma_0(\mathbb{R}^{n_2}) \cap C_L^1(\mathbb{R}^{n_2})$, $A_1 \in \mathbb{R}^{n_1 \times n}$ and $A_2 \in \mathbb{R}^{n_2 \times n}$. Here $\Gamma_0(\mathbb{R}^n)$ denotes the set of proper, convex and lower-semi continuous functions mapping from $\mathbb{R}^n$ to positive-extended real numbers $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, and $\mathbb{R}$ denotes the set of real numbers. With $C_L^1(\mathbb{R}^n)$ we denote the set of all differentiable functions mapping $\mathbb{R}^n$ to $\mathbb{R}$ with L-Lipschitz gradient. We always assume that mentioned models have at least one solution throughout this dissertation, several sufficient conditions of which could be found in [87].

Model (44) has various equivalent forms. One of the most popular form is the following composite problem

$$\operatorname*{argmin}_{x \in \mathbb{R}^n} \{f(x) + g(Lx) + h(x)\}, \tag{45}$$

with $f \in \Gamma_0(\mathbb{R}^n)$, $g \in \Gamma_0(\mathbb{R}^m)$, $h \in \Gamma_0(\mathbb{R}^n) \cap C_L^1(\mathbb{R}^n)$ and $L \in \mathbb{R}^{m \times n}$. This problem is a special case of model (44) with $f_1 = f \oplus g$, $A_1 = (I_n^\top \oplus L^\top)^\top$, $f_2 = h$ and $A_2 = I_n$. Here $f \oplus g$ denotes the separable sum of functions f and g, mapping $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ to $f(x) + g(y)$, and $A^\top$ denotes the conjugate transpose of $A \in \mathbb{R}^{n \times m}$. Conversely, (44) is a special case of model (45) with $f = 0$, $g = f_1$, $h = f_2 \circ A_2$ and $L = A_1$. In the following discussion we propose the general framework in terms of model (44), results of which are then extended to the composite model (45) as concrete applications.

Model (44) has vast applications in many essential branches of data science throughout history, of which followings are some important examples. The correspondence between (44) and the problems below are straightforward and therefore not further explained.

Multi-Block Convex Optimization. Instead of two functions, multi-block optimization considers

$$\operatorname*{argmin}_{x \in \mathbb{R}^n} \sum_{i=1}^{N} f_i(A_i x), \tag{46}$$

where $f_i \in \Gamma_0(\mathbb{R}^{n_i})$, $A_i \in \mathbb{R}^{n_i \times n}$ with $\{n_i\}_{i=1}^N \subset \mathbb{N}^+$, and $\mathbb{N}^+$ is the set of all positive integers. Model (46) could be recast into the following reduced form

$$\operatorname*{argmin}_{x \in \mathbb{R}^n} f(Ax), \tag{47}$$

where $f \in \Gamma_0(\bigtimes_{i=1}^N \mathbb{R}^{n_i})$ and $A \in \mathcal{B}(\mathbb{R}^n, \bigtimes_{i=1}^N \mathbb{R}^{n_i})$ are defined as

$$f := \bigoplus_{i=1}^{N} f_i, \qquad A := \left(\bigoplus_{i=1}^{N} A_i^\top\right)^\top = \begin{bmatrix} A_1^T & A_2^\top & \cdots & A_N^\top \end{bmatrix}^\top.$$

Here $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ denotes the set of bounded linear operators from Banach space $\mathcal{X}$ to Banach space $\mathcal{Y}$, $\mathcal{B}(\mathcal{X}) := \mathcal{B}(\mathcal{X}, \mathcal{X})$ for abbreviation, and $\bigtimes_{i=1}^N \mathbb{R}^{n_i}$ is the Hilbert direct sum of spaces $\{\mathbb{R}^{n_i}\}_{i=1}^N$ and $(\bigoplus_{i=1}^n f_i)((x_i)_{i=1}^n) = \sum_{i=1}^n f_i(x_i)$ is the separable sum of functions $\{f_i\}_{i=1}^N$. It is worthy to notice that (47) is the very special case of (46) with $N = 1$, which implies the preeminence of the simple model (47).

Multi-Sets Split Feasibility Problems [18]. As a general model for plenty of inverse problems where constraints are imposed on the solutions in the domain of a linear operator as well as in the operator's range, this problem seeks for a point that satisfies finite convex constraints. It could be formulated by

$$\operatorname*{argmin}_{x \in \mathbb{R}^n} \sum_{i=1}^{N} \iota_{C_i}(A_i x),$$

where C_i is a closed convex set in $\mathbb{R}^{n_i}$, $A_i \in \mathbb{R}^{n_i \times n}$ for $i \in \mathbb{N}^N := \{1, 2, \ldots, N\}$, and

$$\iota_C(x) := \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C, \end{cases} \qquad \text{for all } x \in \mathbb{R}^n$$

is the indicator function of set $C \subseteq \mathbb{R}^n$. The special case of $A_i = I$ for $i \in \mathbb{N}^N$ is the so-called convex feasibility problem [109], and the case of $N = 2$ and $A_1 = I$ is known as the split feasibility problem [15].

Image Processing. The observation process of an image $z \in \mathbb{R}^{n \times m}$ can be generally modeled by $z = Kx + \varepsilon$, where $K \in \mathcal{B}(\mathbb{R}^{n \times m})$ is the measurement operator, and $\varepsilon \in \mathbb{R}^{n \times m}$ is some randomly-distributed additive noise. Different models are applied according to different kinds of measurement and noise.

Rudin-Osher-Fatemi (ROF) Model [89]. ROF model is designated for restoration of a polluted image by Gaussian noise. Define $\|x\|_{\mathsf{TV}} := \|Dx\|_1$ as the total variation norm

[20, 89], where D is the first order difference operator and $\|\cdot\|_p$ is the ℓ^p vector norm for $1 \leq p \leq \infty$. Then ROF model reads, with $\lambda > 0$,

$$\underset{x\in\mathbb{R}^{n\times m}}{\operatorname{argmin}} \left\{ \frac{\lambda}{2}\|x - z\|^2 + \|x\|_{\mathsf{TV}} \right\}.$$

L1-TV Denoising Model [21, 77]. This model is designated for restoration of a polluted image by impulsive noise, reading, with $\lambda > 0$,

$$\underset{x\in\mathbb{R}^{n\times m}}{\operatorname{argmin}} \{\lambda\|x - z\|_1 + \|x\|_{\mathsf{TV}}\}.$$

L1-TV Deblurring Model [2, 27, 49, 74]. This model is designated for restoration of a blurred image polluted by impulsive noise. Let $K \in \mathcal{B}(\mathbb{R}^{n\times m})$ be the blurring kernel, then this model reads, with $\lambda > 0$,

$$\underset{x\in\mathbb{R}^{n\times m}}{\operatorname{argmin}} \{\lambda\|Kx - z\|_1 + \|x\|_{\mathsf{TV}}\}.$$

Machine Learning. Various optimization problems arising from machine learning always seek to minimize a loss function with regularization. The loss function models the expected cost or the degree of under-fitting with respect to training data, and the regularization term restricts the range of solutions in order to reduce ill-posedness of the problem and avoid over-fitting.

L1-Regularized Linear Least Squares Problem. This problem is known as Basis Pursuit [16, 25, 35] in compress sensing and Least Absolute Shrinkage and Select Operator (LASSO) [95] in machine learning and statistics field. Let $\Phi \in \mathbb{R}^{n\times m}$ be the sampling matrix, $b \in \mathbb{R}^n$ be observed measurements and $\lambda > 0$, then this model can be written as

$$\underset{x\in\mathbb{R}^{n}}{\operatorname{argmin}} \left\{ \frac{\lambda}{2}\|\Phi x - b\|^2 + \|x\|_1 \right\}.$$

L1-Regularized Classification Model [32, 93, 99]. This model designates a linear classifier for classifying data in machine learning. Suppose $\{(x_i, y_i)\}_{i=1}^N \subset \mathbb{R}^n \times \mathbb{R}$ are the given data points, $\{l_i\}_{i=1}^N$ are loss functions mapping $\mathbb{R} \to \mathbb{R}$ and $\lambda > 0$, then this model can be cast as

$$\operatorname*{argmin}_{w\in\mathbb{R}^n, b\in\mathbb{R}} \left\{ \lambda \sum_{i=1}^{N} l_i(y_i \cdot (\langle w, x_i \rangle + b)) + \|w\|_1 \right\}.$$

Several classification models can be written in the form above with different choice of the loss functions. For example, the Support Vector Machine (SVM) [32, 99] adopts hinge loss function, and logistic regression optimization [93] utilizes the logistic loss function.

Distributed Computing [68, 94]. Single computer or processor is increasingly insufficient in big data analysis due to its limited computation capability. In distributed computing, a problem is divided into many tasks, each of which is solved by one or more processors, which communicate with each other via message passing.

Consensus Problem [96, 97]. The main focus of the consensus problem is to solve

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} \sum_{i=1}^{N} f_i(x),$$

by distributing the vector component $x = (x_j)_{j=1}^n$ among n processors.

Exchange Problem [98]. Exchange problem studies the minimum cost of a system with a fixed amount of resources, *i.e.*,

$$\operatorname*{argmin}_{\{x_i\}_{i=1}^N \subset\mathbb{R}^n} \sum_{i=1}^{N} f_i(x_i), \qquad \text{subject to } \sum_{i=1}^{N} x_i = 0,$$

where $\{x_i\}_{i=1}^N$ can be distributed among different processors. This problem can be equivalently formulated as

$$\operatorname*{argmin}_{\{x_i\}_{i=1}^N \subset\mathbb{R}^n} \left\{ \sum_{i=1}^{N} f_i(x_i) + \iota_{\{0\}}\left(\sum_{i=1}^{N} x_i \right) \right\}.$$

Sharing Problem. Sharing problems serve as extensions of exchange problems, where the strict constraint $\sum_{i=1}^N x_i = 0$ is generalized. Sharing problem solves

$$\operatorname*{argmin}_{\{x_i\}_{i=1}^N \subset \mathbb{R}^n} \left\{ \sum_{i=1}^N f_i(x_i) + g\left(\sum_{i=1}^N x_i \right) \right\},$$

where $g \in \Gamma_0(\mathbb{R}^n)$. Sharing problems reduce to exchange problems when $g = \iota_{\{0\}}$.

In the following sections of this chapter, we propose the general framework of fixed-point proximity algorithms for (44) with convergence theorem, which could be applied to the above problems.

3.2 PRELIMINARIES

In this section we review several useful preliminaries in convex analysis and operator theory for further discussion. Most of the theorems in this section could be found in [6, 87] and references therein.

Assume $f \in \Gamma_0(\mathbb{R}^n)$. The *subdifferential* (or *subderivative*, *subgradient*) of f is defined as

$$\partial_f : \mathbb{R}^n \to 2^{\mathbb{R}^n} : x \mapsto \{u \in \mathbb{R}^n : \langle u, x - y \rangle \geq f(x) - f(y) \text{ for all } y \in \mathbb{R}^n\}.$$

Here $\langle \cdot, \cdot \rangle$ is the Euclidean product of $\mathbb{R}^n$. Subdifferentials play core roles in optimization. Particularly, if f is further differentiable at $x \in \mathbb{R}^n$, then $\partial_f(x) = \{\nabla_f(x)\}$. The *conjugate* (or *Fenchel conjugate*, *Legendre transform*, *Legendre-Fenchel transform*) of f is defined as

$$f^* : \mathbb{R}^n \to \overline{\mathbb{R}} : x \mapsto \sup_{u \in \mathbb{R}^n} \{\langle u, x \rangle - f(u)\}.$$

By Fenchel-Moreau Theorem we have $f^* \in \Gamma_0(\mathbb{R}^n)$ and $f = f^{**}$ if $f \in \Gamma_0(\mathbb{R}^n)$. Here $f^{**} := (f^*)^*$ is the *biconjugate* of function f. Also notice that for separable sum of $f \in \Gamma_0(\mathbb{R}^n)$ and $g \in \Gamma_0(\mathbb{R}^m)$ we have $(f \oplus g)^* = f^* \oplus g^*$ and

$$\partial_{f \oplus g} = \partial_f \otimes \partial_g, \tag{48}$$

where $f \otimes g$ means the Cartesian product of two operators f and g, mapping $(x, y) \mapsto f(x) \times g(y)$. There holds an essential relation of subdifferentials between conjugate functions, that is, for $x, u \in$

$\mathbb{R}^n$,

$$u \in \partial_f(x) \quad \text{if and only if} \quad x \in \partial_{f^*}(u). \tag{49}$$

Define $\mathbb{S}_+^n$ to be the set of symmetric strictly positive-definite matrices in $\mathbb{R}^n$. The *proximity operator* (or *proximal operator*) of f with respect to $P \in \mathbb{S}_+^n$ is

$$\mathrm{prox}_{f,P} \,:\, \mathbb{R}^n \to \mathbb{R}^n \,:\, x \mapsto \operatorname*{argmin}_{u\in\mathbb{R}^n} \left\{ f(u) + \frac{1}{2}\|u - x\|_P^2 \right\}.$$

Specially, denote $\mathrm{prox}_f := \mathrm{prox}_{f,I}$. Notice that for $f_i \in \Gamma_0(\mathbb{R}^{n_i})$ and $P_i \in \mathbb{S}_+^{n_i}$ with $i \in \mathbb{N}^n$, we have

$$\mathrm{prox}_{\bigoplus_{i=1}^n f_i, \bigotimes_{i=1}^n P_i}\big((x_i)_{i=1}^n\big) = \big(\mathrm{prox}_{f_i,P_i}(x_i)\big)_{i=1}^n. \tag{50}$$

Proximity operator closely relates with subdifferential. Specifically, for $x, u \in \mathbb{R}^n$ and $P \in \mathbb{S}_+^n$,

$$u \in \partial_f(x) \quad \text{if and only if} \quad x = \mathrm{prox}_{f,P}\big(x + P^{-1}u\big). \tag{51}$$

With the relation of subdifferentials and conjugate functions, we can easily show the Moreau's identity

$$I = \mathrm{prox}_{f,P} + P^{-1} \circ \mathrm{prox}_{f^*,P^{-1}} \circ P \tag{52}$$

holds for any $P \in \mathbb{S}_+^n$, where '∘' denotes the composition of functions.

Followings are some useful examples of subdifferentials, conjugate functions and proximity operators.

Example 3.1 (Constant function)**.** Let $c \in \mathbb{R}$ and constant function $f(x) = c$ for $x \in \mathbb{R}^n$. Then $\partial_f(x) = 0$ for all $x \in \mathbb{R}^n$, and

$$f^*(x) = \iota_{\{0\}}(x) - c = \begin{cases} -c, & x = 0, \\ +\infty, & x \neq 0, \end{cases} \quad \text{and} \quad \partial_{f^*}(x) = \begin{cases} \mathbb{R}^n, & x = 0, \\ \varnothing, & x \neq 0. \end{cases}$$

Therefore, by Moreau's identity (52), for any $P \in \mathbb{S}_+^n$, we have $\mathrm{prox}_{f,P} = I$ and $\mathrm{prox}_{f^*,P} = 0$.

Example 3.2 (The ℓ^1 norm)**.** For ℓ^1 norm $\|\cdot\|_1$ in $\mathbb{R}^n$, since $\|\cdot\|_1 = \bigoplus_{i=1}^n |\cdot|$, we have $\|\cdot\|_1^* = \iota_{B_\infty(0;1)}$, where $B_\infty(0;1)$ denotes the closed unit ℓ^∞-ball $\{x \in \mathbb{R}^n : \|x\|_\infty \le 1\}$. Then by (48) and (49) we have

$$\partial_{\|\cdot\|_1}(x) = \bigotimes_{i=1}^n \partial_{|\cdot|}(x_i) \quad \text{and} \quad \partial_{\|\cdot\|_1^*}(u) = \bigotimes_{i=1}^n \partial_{|\cdot|^*}(u_i),$$

where

$$\partial_{|\cdot|}(x) = \begin{cases} \{1\}, & x > 0, \\ [-1,1], & x = 0, \\ \{-1\}, & x < 0, \end{cases} \quad \text{and} \quad \partial_{|\cdot|^*}(u) = \begin{cases} [0,+\infty), & u = 1, \\ \{0\}, & u \in (-1,1), \\ (-\infty, 0], & u = -1, \\ \varnothing, & |u| > 1. \end{cases}$$

Therefore, by Moreau's identity (52), for any $P = \mathrm{Diag}(p_i)_{i=1}^n$ with positive $\{p_i\}_{i=1}^n$, we have proximity operators

$$\begin{aligned} \mathrm{prox}_{\|\cdot\|_1, P}(x) &= \left(\mathrm{sgn}(x_i)\left(|x_i| - p_i^{-1}\right)_+\right)_{i=1}^n, \\ \mathrm{prox}_{\|\cdot\|_1^*, P}(u) &= u - P^{-1}\,\mathrm{prox}_{\|\cdot\|_1, P^{-1}}(Pu) \\ &= (\min\{1, \max\{u_i, -1\}\})_{i=1}^n, \end{aligned}$$

where 'Diag' creates diagonal matrix and 'sgn' is the sign function in $\mathbb{R}$. Notice that as the proximity operator of an indicator function of a unit ℓ^∞-ball, $\mathrm{prox}_{\|\cdot\|_1^*, P}$ is independent of P as long as P is diagonal.

Example 3.3 (The ℓ^2 norm)**.** For ℓ^2 norm $\|\cdot\|$ in $\mathbb{R}^n$, we have $\|\cdot\|^* = \iota_{B_2(0;1)}$ and

$$\partial_{\|\cdot\|}(x) = \begin{cases} \left\{\frac{x}{\|x\|}\right\}, & x \ne 0, \\ B_2(0;1), & x = 0, \end{cases} \quad \text{and} \quad \partial_{\|\cdot\|^*}(u) = \begin{cases} \{0\}, & \|u\| < 1, \\ \{\lambda u : \lambda \ge 0\}, & \|u\| = 1, \\ \varnothing, & \|u\| > 1. \end{cases}$$

Therefore we have, for any $\alpha > 0$ we have

$$\mathrm{prox}_{\|\cdot\|, \alpha I}(x) = \begin{cases} x - \frac{1}{\alpha}\frac{x}{\|x\|}, & \alpha\|x\| > 1, \\ 0, & \alpha\|x\| \le 1, \end{cases} \quad \text{and} \quad \mathrm{prox}_{\|\cdot\|^*, \alpha I}(u) = \begin{cases} \frac{u}{\|u\|}, & \|x\| > 1, \\ u, & \|x\| \le 1. \end{cases}$$

Similar to ℓ^1 norm, here $\mathrm{prox}_{\|\cdot\|^*, \alpha I}$ is independent of $\alpha > 0$.

Let us return to the subdifferential for its further properties. First by definition, we can easily get the Fermat's rule as

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} f(x) = \operatorname{Zer}\partial_f, \tag{53}$$

where $\operatorname{Zer}\partial_f := \{x \in \mathbb{R}^n : 0 \in \partial_f(x)\}$ is the zero set of operator ∂_f. Also subdifferentials of functions in $\Gamma_0(\mathbb{R}^n)$ are *monotone operators*, *i.e.*, for

$$\langle x-y, u-v\rangle \geq 0, \quad \text{for all } (x,u),(y,v) \in \operatorname{Gra}\partial_f.$$

Here $\operatorname{Gra}\partial_f := \{(x,u) \in \mathbb{R}^n \times \mathbb{R}^n : u \in \partial_f(x)\}$ denotes the graph of ∂_f. Monotonicity of subdifferential is pivotal in convergence analysis of convex optimization, *e.g.*, we can easily derive from the monotonicity of subdifferentials and (51) that for $f \in \Gamma_0(\mathbb{R}^n)$ and $P \in \mathbb{S}^n_+$, proximity operator $\operatorname{prox}_{f,P}$ is nonexpansive 1/2-average (*i.e.*, firmly nonexpansive) with respect to norm $\|\cdot\|_P$. To handle composite optimization problems, we need the calculus principle of subdifferential with function additions and composition with linear transforms. Define the domain of function f as

$$\operatorname{Dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\},$$

and the *infimal convolution* of $f, g \in \Gamma_0(\mathbb{R}^n)$ as

$$f \,\Box\, g : \mathbb{R}^n \to [-\infty, +\infty] : x \mapsto \inf_{u\in\mathbb{R}^n} \{f(u) + g(x-u)\}.$$

If for $x \in \mathbb{R}^n$ there exists $u \in \mathbb{R}^n$ such that $(f \,\Box\, g)(x) = f(u) + g(x-u)$, then we say that $f \,\Box\, g$ is *exact* at point x. If $f \,\Box\, g$ is exact at every point of its domain, we say $f \,\Box\, g$ is *exact*, in which case it is denoted by $f \boxdot g$. Also we define *infimal postcomposition* of $f \in \Gamma_0(\mathbb{R}^n)$ and $L \in \mathbb{R}^{n\times m}$ as

$$L \triangleright f : \mathbb{R}^n \to [-\infty, +\infty] : x \mapsto \inf_{\substack{u\in\mathbb{R}^m \\ Lu=x}} f(u).$$

Similar to infimal convolution, if for $x \in \mathbb{R}^n$ there exists $u \in \mathbb{R}^m$ such that $(L \triangleright f)(x) = f(u)$, then we say that $L \triangleright f$ is *exact* at point x. If $L \triangleright f$ is exact at every point of its domain, we say $L \triangleright f$ is *exact*, in which case it is denoted by $L \,\dot{\triangleright}\, f$.

We now ready for the calculation rule of subdifferentials. Let $f \in \Gamma_0(\mathbb{R}^n)$, $g \in \Gamma_0(\mathbb{R}^m)$ and $L \in \mathbb{R}^{m\times n}$ be such that $L(\operatorname{Dom} f) \cap \operatorname{Dom} g \neq \varnothing$, and suppose that $(f + g \circ L)^* = f^* \boxdot (L^\top \,\dot{\triangleright}\, g^*)$, then there holds

$$\partial_{f+g\circ L} = \partial_f + L^\top \circ \partial_g \circ L. \tag{54}$$

The above equation is normally mentioned as the *additivity* of subdifferentials when $L = I$, and the *chain rule* of subdifferentials when $f = 0$. Always assume $(f+g\circ L)^* = f^* \boxdot (L^\top \triangleright g^*)$ holds in following discussion, of which several sufficient conditions could be found in [6, Theorem 15.27].

Then we introduce the duality in convex optimization. Suppose $f \in \Gamma_0(\mathbb{R}^n)$, $g \in \Gamma_0(\mathbb{R}^m)$ and $L \in \mathbb{R}^{m\times n}$. The *primal problem* associated with the composite function $f + g \circ L$ is

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} \{f(x) + g(Lx)\}, \tag{55}$$

and its *dual problem* is

$$\operatorname*{argmin}_{u\in\mathbb{R}^m} \left\{ f^*\left(-L^\top u\right) + g^*(u)\right\}. \tag{56}$$

The *primal optimal value* is defined as $\mu := \inf\{f + g \circ L\}$ and the *dual optimal value* as $\mu^* := \inf\{f^* \circ (-L^\top) + g^*\}$. The *dual gap* is then defined as

$$\Delta := \begin{cases} 0, & \mu = -\mu^* \in \{\pm\infty\}, \\ \mu + \mu^*, & \text{otherwise.} \end{cases}$$

Generally we always have $\Delta \geq 0$. Notice that condition $(f + g \circ L)^* = f^* \boxdot (L^\top \triangleright g^*)$ implies that

$$\begin{aligned} \inf_{x\in\mathbb{R}^n} \{f(x) + g(Lx)\} &= (f + g \circ L)^*(0) \\ &= (f^* \boxdot (L^* \triangleright g^*))(0) \\ &= \min_{u\in\mathbb{R}^m} \left\{ f^*\left(-L^\top u\right) + g^*(u)\right\}, \end{aligned}$$

which means that the dual problem has nonempty solution set, and dual gap $\Delta = 0$. Specially, if set $f = f_2 \circ A_2$, $g = f_1$ and $L = A_1$ in (55), then we have the dual problem of model (44) as

$$\operatorname*{argmin}_{u\in\mathbb{R}^{n_1}} \left\{ (f_2 \circ A_2)^*\left(-A_1^\top u\right) + f_1^*(u)\right\}. \tag{57}$$

The primal/dual optimal values are

$$\mu = \min\{f_1 \circ A_1 + f_2 \circ A_2\} \quad \text{and} \quad \mu^* = \min\left\{ (f_2 \circ A_2)^* \circ \left(-A_1^\top\right) + f_1^* \right\}.$$

By assumptions, the dual gap $\Delta = \mu + \mu^* = 0$.

We then introduce several terminology of operator theory for following discussion. An operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is *β-cocoercive* (or *β-inverse strongly monotone*) with $\beta > 0$ if for all $x, y \in \mathbb{R}^n$,

$$\langle Tx - Ty, x - y \rangle \geq \beta \|Tx - Ty\|^2.$$

Notice that $T : \mathbb{R}^n \to \mathbb{R}^n$ is β-cocoercive if and only if βT is nonexpansive $1/2$-averaged (*i.e.*, firmly nonexpansive). A monotone operator $T : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is *maximal*, if there exists no monotone operator $T' : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ such that $\operatorname{Gra} T \subsetneq \operatorname{Gra} T'$. For a monotone operator $T : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, the *Fitzpatrick function* F_T of T is defined by

$$F_T(x, u) := \langle x, u \rangle - \inf_{(y,v) \in \operatorname{Gra} T} \langle x - y, u - v \rangle, \qquad \text{for all } x, u \in \mathbb{R}^n.$$

A monotone operator $T : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is called 3^*-*monotone*, if $\operatorname{Dom} T \times \operatorname{Ran} T \subset \operatorname{Dom} F_T$.

The following identity reveals that Hilbert spaces are uniformly convex and strictly convex Banach spaces, and it is vital in optimization embedded in Hilbert spaces. For all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, we have

$$\|\alpha x + (1 - \alpha) y\|^2 + \alpha(1 - \alpha)\|x - y\|^2 = \alpha\|x\|^2 + (1 - \alpha)\|y\|^2. \tag{58}$$

Identity (58) could be easily extended to weighted semi-norms. Recall that for $W \in \mathbb{S}^n$ we denote the weighted semi-norm $\|\cdot\|_W := \sqrt{\langle \cdot, W \cdot \rangle}$. Then there holds, for all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$,

$$\|\alpha x + (1 - \alpha) y\|_W^2 + \alpha(1 - \alpha)\|x - y\|_W^2 = \alpha\|x\|_W^2 + (1 - \alpha)\|y\|_W^2. \tag{59}$$

Here we propose several useful convergence lemmas. Recall that for $A \in \mathbb{R}^{n \times m}$, we denote its kernel as $\operatorname{Ker} A$.

Lemma 3.4. *Suppose that* $\{x_k\}_{k \in \mathbb{N}} \subset \mathbb{R}^n$ *and* $V, W \in \mathbb{S}^n$ *such that* $\operatorname{Ker} V \subseteq \operatorname{Ker} W$.

(i) If $\{\|x_k\|_V\}_{k \in \mathbb{N}}$ *is bounded, then* $\{W x_k\}_{k \in \mathbb{N}}$ *is bounded.*

(ii) If $\lim_{k \to \infty} \|x_k\|_V = 0$, *then* $\lim_{k \to \infty} W x_k = 0$.

Proof. Notice that for any $V \in \mathbb{S}^n$ there exists $\sqrt{V} \in \mathbb{S}^n$ such that $V = \sqrt{V}\sqrt{V}$, and since $\operatorname{Ker} V \subseteq \operatorname{Ker} W$, there exists $U \in \mathbb{R}^{n \times n}$ such that $W = UV$. Therefore we have $\|x_k\|_V = \|\sqrt{V} x_k\|$.

(i) If $\{\|x_k\|_V\}_{k\in\mathbb{N}}$ is bounded, then $\{\sqrt{V}x_k\}_{k\in\mathbb{N}}$ is bounded, which means $\{Vx_k\}_{k\in\mathbb{N}}$ is bounded, and so does $\{UVx_k\}_{k\in\mathbb{N}}$.

(ii) If $\lim_{k\to\infty}\|x_k\|_V = 0$, then $\lim_{k\to\infty}\sqrt{V}x_k = 0$, which means $\lim_{k\to\infty}Vx_k = 0$. Therefore we have $\lim_{k\to\infty}Wx_k = U\lim_{k\to\infty}Vx_k = 0$. □

3.3 FIXED-POINT CHARACTERIZATION OF MINIMIZERS

In this section we characterize the global minimizers of model (44) into fixed-points of a proximity equation via preliminaries in Section 3.2.

For $f \in \Gamma_0(\mathbb{R}^n)$, Fermat's rule (53) equivalently characterizes the minimizers, argmin f, as Zer ∂_f. However since ∂_f is generally not one-to-one, such characterization is not of much help in algorithm design, not to mention the composite problem (44). Fortunately, with proximity operators, Fermat's characterization could be cast into fixed-points of a proximity equation, which provides a powerful guidance of algorithm framework. For model (44), we define linear transform $S \in \mathcal{B}(\mathbb{R}^{n+n_1})$ such that $S(x,u) = (-A_1^\top u, A_1x)$ for all $(x,u) \in \mathbb{R}^n \times \mathbb{R}^{n_1}$. Notice that S is *skew-symmetric*, that is, for all $z \in \mathbb{R}^{n+n_1}$ we have $\langle z, Sz\rangle = 0$. Also define $\Phi := 0 \oplus f_1^* \in \Gamma_0(\mathbb{R}^{n+n_1})$ and $A := A_2 \oplus 0_{n_2\times n_1} \in \mathbb{R}^{n_2\times(n+n_1)}$.

Theorem 3.5. *Suppose $x \in \mathbb{R}^n$ is a solution of primal problem* (44)*, then there exists $u \in \mathbb{R}^{n_1}$ as a solution of dual problem* (57) *such that for any $R \in \mathbb{S}_+^{n+n_1}$ there holds*

$$z = \operatorname{prox}_{\Phi,R}\left(\left(I + R^{-1}S\right)z - R^{-1}A^\top\nabla_{f_2}(Az)\right), \tag{60}$$

where $z := (x,u)$. Conversely, if there exists $z = (x,u) \in \mathbb{R}^n \times \mathbb{R}^{n_1}$ such that (60) *holds for some $R \in \mathbb{S}_+^{n+n_1}$, then x is a solution of primal problem* (44) *and u is a solution of dual problem* (57)*.*

Proof. Suppose $x \in \mathbb{R}^n$ is a solution of model (44). Then by Fermat's rule (53) and the calculus principle of subdifferentials (54), we have

$$0 \in \partial_{f_1\circ A_1 + f_2\circ A_2}(x) = A_1^\top\partial_{f_1}(A_1x) + A_2^\top\nabla_{f_2}(A_2x),$$

which means, there exists $u \in \partial_{f_1}(A_1x)$ such that $0 \in A_1^\top u + A_2^\top\nabla_{f_2}(A_2x)$. Then by the relation of subdifferentials between conjugate functions (49) and subdifferentials of separable sum of functions (48), we have

$$\begin{cases} -A_1^\top u - A_2^\top \nabla_{f_2}(A_2 x) \in \partial_0(x), \\ \qquad\qquad\qquad A_1 x \in \partial_{f_1^*}(u), \end{cases} \tag{61}$$

where one need to notice that $\partial_0(x) = \{0\}$ for any $x \in \mathbb{R}^n$; see Example 3.1. If define $z := (x, u)$, then by (48), $\Phi = 0 \oplus f_1^*$ and $A = A_2 \oplus 0_{n_2 \times n_1}$, the above inclusions can be cast into a more compact form as

$$Sz - A^\top \nabla_{f_2}(Az) \in \partial_\Phi(z).$$

Then by the relation between subdifferentials and proximity operators (51), for any $R \in \mathbb{S}_+^{n+n_1}$ we have (60) holds. Furthermore, from (61) we have

$$\begin{cases} -A_1 x \in \partial_{(f_2 \circ A_2)^* \circ (-A_1^\top)}(u), \\ \;\; A_1 x \in \partial_{f_1^*}(u), \end{cases}$$

which means

$$0 \in \partial_{(f_2 \circ A_2)^* \circ (-A_1^\top)} + \partial_{f_1^*}(u) = \partial_{(f_2 \circ A_2)^* \circ (-A_1^\top) + f_1^*}(u),$$

showing that u is indeed a solution to the dual problem (57).

Conversely, if (60) holds for $z = (x, u)$ and some $R \in \mathbb{S}_+^{n+n_1}$, then it is direct to check that (61) holds, which finishes the proof with analogous argument above. □

Fixed-point equation (60) not only characterizes the primal solutions, but also the dual solutions integrated in one single equation. It has various forms under different settings. We present several common cases covered by the composite problem (45). By (56) and (57), the dual problem of model (45) is

$$\underset{(u^1, u^2) \in \mathbb{R}^n \times \mathbb{R}^m}{\operatorname{argmin}} \left\{ f^*\left(u^1\right) + g^*\left(u^2\right) + h^*\left(-u^1 - L^\top u^2\right) \right\}. \tag{62}$$

If we set $R = I_n \otimes P^{-1} \otimes Q$ with $P \in \mathbb{S}_+^n$, $Q \in \mathbb{S}_+^m$ and $z = (x, u^1, u^2) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ as a fixed-point of (60), then by (50) we have (60) reduce to

$$\begin{cases} \;x = \operatorname{prox}_0\left(x - u^1 - L^\top u^2 - \nabla_h(x)\right), \\ u^1 = \operatorname{prox}_{f^*, P^{-1}}\left(Px + u^1\right), \\ u^2 = \operatorname{prox}_{g^*, Q}\left(Q^{-1} L x + u^2\right). \end{cases}$$

The above can be simplified via (51) and (52) and Example 3.1 to

$$\begin{cases} x = \operatorname{prox}_{f,P}\big(x - P^{-1}\big(Lu^2 + \nabla_h(x)\big)\big), \\ u^2 = \operatorname{prox}_{g^*,Q}\big(Q^{-1}L^\top x + u^2\big), \\ u^1 = -Lu^2 - \nabla_h(x). \end{cases} \tag{63}$$

By previous discussion, the above characterization describes primal/dual solutions simultaneously, where x is the solution of the primal problem (45) and (u^1, u^2) is the solution of the dual problem (62). Part of result (63) was discussed in [55]. We then discuss possible forms of (63) in every probable cases that some of the three terms f, g, h are missing in (45).

Example 3.6. With $f = 0$, (63) can be simplified to

$$\begin{cases} u^2 = \operatorname{prox}_{g^*,Q}\big(Q^{-1}L^\top x + u^2\big), \\ u^1 = -Lu^2 - \nabla_h(x_*) = 0. \end{cases}$$

With $f = g = 0$, (63) can be further simplified to

$$0 = \nabla_h(x) = u^1 = u^2.$$

Part of this result reverts to Fermat's rule for smooth convex optimization. With $f = h = 0$, (63) can be further simplified to

$$\begin{cases} u^2 = \operatorname{prox}_{g^*,Q}\big(Q^{-1}L^\top x + u^2\big), \\ u^1 = Lu^2 = 0. \end{cases}$$

Part of this result was discussed as dual formulation in [24].

Example 3.7. With $g = 0$, (63) can be simplified to

$$\begin{cases} x = \operatorname{prox}_{f,P}\big(x - P^{-1}\nabla_h(x)\big), \\ u_1 = -\nabla_h(x),\ u^2 = 0. \end{cases}$$

Part of this result was discussed in [31, 73]. With $g = h = 0$, (63) can be further simplified to

$$\begin{cases} x = \operatorname{prox}_{f,P}(x), \\ u^1 = u^2 = 0. \end{cases}$$

Part of this result could be obtained by combining Fermat's rule (53) and relationship (51).

Example 3.8. With $h = 0$, (63) can be simplified to

$$\begin{cases} x = \operatorname{prox}_{f,P}\big(x - P^{-1}Lu^2\big), \\ u^2 = \operatorname{prox}_{g^*,Q}\big(Q^{-1}L^\top x + u^2\big), \\ u^1 = -Lu^2. \end{cases}$$

Notice that in this case the composite problem (45) is identical to (55), but the dual problem (62) is different from (56), implying that dual problems may not unique in form. Part of this result was discussed in [56].

Theorem 3.5 characterizes solution pairs of primal problem (44) and dual problem (57) as fixed-points of the proximity equation (60). Compared with Fermat's characterization, fixed-point equation (60) gives a uniform and comprehensive description of the primal/dual solutions, while being more suggestive for algorithm design. The proposed framework of fixed-point proximity algorithms in this dissertation will based on the fixed-point equation (60).

3.4 IMPLICIT FIXED-POINT PROXIMITY ALGORITHMS

In this section we propose the general framework of implicit fixed-point proximity algorithms for primal/dual problems (44) and (57) based on the fixed-point proximity equation (60). We derive implicit iterations by an equivalent variation of (60) via applying matrix splitting technique, and then consider conditions under which the derived implicit iteration will be well-defined.

Theorem 3.5 characterizes solution pairs of primal/dual problems (44) and (57) to be exactly

$$\operatorname{Fix}\big(\operatorname{prox}_{\Phi,R} \circ \big(I + R^{-1}S - R^{-1}\nabla_{f_2 \circ A}\big)\big)$$

for any $R \in \mathbb{S}_+^{n+n_1}$. The proximity operator $\operatorname{prox}_{\Phi,R}$ is always nonexpansive $1/2$-averaged with respect to $\|\cdot\|_R$, which therefore has novel convergence property by Theorem 2.2. However, the following property shows that $|||I+R^{-1}S|||_R$ is generally greater than 1, which invalidate the general idea of naïve Picard iteration, especially when $A_1 \neq 0$ and $f_2 = 0$. Here $|||\cdot|||_R$ is the induced operator norm with respect to weighted norm $\|\cdot\|_R$, *i.e.*, for $R \in \mathbb{S}_+^n$ and $A \in \mathbb{R}^{n\times n}$, define

$$\|A\| := \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_R}{\|x\|_R}.$$

Part of result for the special case with block-diagonal R was once discussed in [24, 56].

Proposition 3.9. *Suppose $R \in \mathbb{S}_+^{n+n_1}$. There holds*

$$\left\|I + R^{-1}S\right\|_R^2 = 1 + \|S\|_{R^{-1}}^2.$$

In particular, $S \neq 0$ if and only if $\|I + R^{-1}S\|_R > 1$.

Proof. By the skew-symmetry of S, for any $x \in \mathbb{R}^{n+n_1}$ we have

$$\left\|\left(I + R^{-1}S\right)x\right\|_R^2 = \left\langle (R+S)x, \left(I + R^{-1}S\right)x\right\rangle = \|x\|_R^2 + \|Sx\|_{R^{-1}}^2,$$

which proves that

$$\left\|I + R^{-1}S\right\|_R^2 = 1 + \left\|R^{-1}S\right\|_R^2.$$

Thus $\|I + R^{-1}S\|_R^2 \geq 1$, and $\|I + R^{-1}S\|_R^2 = 1$ if and only if $\|R^{-1}S\|_R = 0$, if and only if $S = 0$. □

Proposition 3.9 reveals the expanding nature of the linear transform $I + R^{-1}S$, which makes the naïve Picard iteration of operator $\mathrm{prox}_{\Phi,R} \circ (I + R^{-1}S - R^{-1}\nabla_{f_2\circ A})$ not generally converge. Fortunately, since such defect is brought by the linear transform, the *matrix splitting technique* could be applied to construct a family of implicit algorithms with convergence analysis. This technique was first introduced in [100] and then widely exploited in solving linear systems and differential equations.

Here we apply the matrix splitting technique to the linear transforms I and $I + R^{-1}S$ in both ends of the fixed-point equation (60). Precisely, we split $I = \Lambda + (I - \Lambda)$ in the left-hand side and $I + R^{-1}S = E + M$ in the right-hand side, then (60) becomes

$$\Lambda z + (I - \Lambda)z = \mathrm{prox}_{\Phi,R}\left(Ez + Mz - R^{-1}A^{\top}\nabla_{f_2}(Az)\right),$$

therefore the proposed framework of Implicit Fixed-Point Proximity Algorithms (IFP²A) of this dissertation is proposed as

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} \Lambda z_{k+1} + (I-\Lambda)w_k = \operatorname{prox}_{\Phi,R}\big(Ez_{k+1} + Mw_k - R^{-1}A^\top \nabla_{f_2}(Aw_k)\big) \\ w_{k+1} \leftarrow (1-\lambda)w_k + \lambda z_{k+1} \end{array}\right. \end{array} \tag{64}$$

where $E + M = I + R^{-1}S$ and $\lambda \in \mathbb{R}$. Notice that the first line of IFP²A (64) requires solving an implicit proximity equation, which therefore generates implicit algorithms in general.

There are several fundamental questions following immediately after IFP²A, *i.e.*,

(Q1) The well-definedness of the IFP²A iteration.

(Q2) The convergence of the IFP²A.

(Q3) The feasible computation methods, which might be inexact, as IFP²A is implicit.

(Q4) The convergence of the inexact IFP²A.

In following parts of this dissertation all these questions (Q1) to (Q4) will be answered one after another. In rest of this section we focus on the first question. To answer the first question (Q1) on IFP²A, the following theorem serves as a sufficient condition for (64) to uniquely determine an iteration for all possible inputs. Define GL^n to be the group of invertible matrices in $\mathbb{R}^n$, and $\mathbb{P}^n_+$ to be the set of all strictly-positive definite matrices in $\mathbb{R}^n$. For matrix $A \in \mathcal{B}(\mathbb{R}^{n+n_1})$, define $\overline{A} := (A + A^\top)/2$.

Theorem 3.10. *Suppose $R \in \mathbb{S}^{n+n_1}_+$ and $\Lambda, E \in \mathcal{B}(\mathbb{R}^{n+n_1})$. If $\Lambda \in \mathrm{GL}^{n+n_1}$ and there exists $\Gamma = \gamma_0 I_n \otimes \gamma_1 I_{n_1}$ with $\gamma_0, \gamma_1 > 0$ such that $\Gamma R(I - E\Lambda^{-1}) \in \mathbb{P}^{n+n_1}_+$, then for any $u, v \in \mathbb{R}^{n+n_1}$, the fixed-point equation*

$$\Lambda z + u = \operatorname{prox}_{\Phi,R}(Ez + v) \tag{65}$$

has a unique solution $z \in \mathbb{R}^{n+n_1}$, and function $z = z(u,v)$ is uniformly continuous.

Proof. Let $z = (z_0, z_1) \in \mathbb{R}^n \times \mathbb{R}^{n_1}$ and u, v as so. First we prove that (65) has at least one solution, then we prove the uniqueness. Define $x := \Lambda z + u$ and $P_0 := \Gamma R(I - E\Lambda^{-1}) \in \mathbb{P}^{n+n_1}_+$. Then by (51), we have (65) equivalent to, for any $\Gamma = \gamma_0 I_n \otimes \gamma_1 I_{n_1}$,

$$\Gamma R v - \Gamma R E \Lambda^{-1} u \in \partial_{0 \oplus \gamma_1 f_1^*}(x) + P_0 x. \tag{66}$$

Therefore for (65) to have at least one solution, it is sufficient to prove that $\operatorname{Ran}(\partial_{0\oplus\gamma_1 f_1^*} + P_0) =$

$\mathbb{R}^{n+n_1}$. To this end we need the following key results from operator theory. We can learn that $\partial_{0\oplus\gamma_1 f_1^*}$ is maximally monotone [6, Theorem 20.25] and 3*-monotone [6, Example 25.13], and P_0 is maximally monotone [6, Corollary 20.28]. By strict positive-definiteness of P_0, we have

$$\inf_{x\in\mathbb{R}^{n+n_1}\setminus\{0\}} \frac{\langle x, P_0x\rangle}{\|P_0x\|^2} \geq \left\|P_0^\top P_0\right\|^{-1}\left\|\overline{P_0}^{-1}\right\|^{-1} > 0.$$

It means P_0 is $\|P_0^\top P_0\|^{-1}\|\overline{P_0}^{-1}\|^{-1}$-cocoercive, and thus P_0 is 3*-monotone [6, Proposition 25.16]. Then $\partial_{0\oplus\gamma_1 f_1^*} + P_0$ is maximally monotone [6, Corollary 25.5]. Finally, by non-singularity of P_0 and [6, Corollary 25.27], we have that $\partial_{0\oplus\gamma_1 f_1^*} + P_0$ is surjective on $\mathbb{R}^{n+n_1}$. These prove that (65) has at least one solution.

Then we further prove that (65) has a unique solution. Suppose (65) has solutions z and z'. Define $x := \Lambda z + u$ and $x' := \Lambda z' + u$. Then by (66) and monotonicity of subdifferentials, we have

$$0 \leq \left\langle x - x', P_0\left(x' - x\right)\right\rangle,$$

which, by $P_0 \in \mathbb{P}_+^{n+n_1}$, means $x = x'$ and thus $z = z'$. These prove that (65) has exactly one solution.

Finally we prove the uniform continuity of $z = z(u, v)$. Let $z = z(u, v)$ and $z' = z'(u', v')$ be determined by (65). Define x and x' as before. Then by (66) and monotonicity of subdifferentials, we have

$$\begin{aligned} 0 &\leq \left\langle x - x', \Gamma R\left(v - v'\right) - \Gamma R E \Lambda^{-1}\left(u - u'\right) - P_0\left(x - x'\right)\right\rangle \\ &\leq -\left\|x - x'\right\|_{\overline{P_0}}^2 + \|\Gamma R\|\left\|x - x'\right\|\left(\left\|v - v'\right\| + \left\|E\Lambda^{-1}\right\|\left\|u - u'\right\|\right), \end{aligned}$$

which proves that $z = z(u, v)$ defined by (65) is uniformly continuous. □

Theorem 3.10 can be easily applied to IFP²A (64), resulting in the the following corollary.

Corollary 3.11. *Suppose $R \in \mathbb{S}_+^{n+n_1+n_2}$ and $\Lambda, E \in \mathcal{B}(\mathbb{R}^{n+n_1})$. If $\Lambda \in \mathrm{GL}^{n+n_1}$ and there exists $\Gamma = \gamma_0 I_n \otimes \gamma_1 I_{n_1}$ with $\gamma_0, \gamma_1 > 0$ such that*

$$\Gamma R\left(I - E\Lambda^{-1}\right) \in \mathbb{P}_+^{n+n_1},$$

then for any $\lambda \in \mathbb{R}$ and $w_0 \in \mathbb{R}^{n+n_1}$, IFP²A (64) uniquely determines iteration sequence $\{(z_k, w_k)\}_{k\in\mathbb{N}}$. Moreover, the mapping $w_k \mapsto z_{k+1}$ determined by IFP²A (64) is uniformly continuous.

Proof. Since composition of uniformly continuous functions is uniformly continuous, by Theorem 3.10, it is sufficient to prove that mapping $z \mapsto Mw - R^{-1}A^{\top}\nabla_{f_2}(Az)$ is uniformly continuous. Notice that $f_2 \in C^1_L(\mathbb{R}^{n_2})$, we have that ∇_{f_2} is L-Lipschitz continuous. Then for $z, z' \in \mathbb{R}^{n+n_1}$ we have

$$\left\| M\left(z - z'\right) - R^{-1}A^{\top}\left(\nabla_{f_2}(Az) - \nabla_{f_2}\left(Az'\right)\right)\right\| \le \left(\|M\| + L\left\|R^{-1}A^{\top}\right\|\|A\|\right)\left\|z - z'\right\|,$$

which finishes the proof. □

In the following discussion, for $R \in \mathbb{S}^{n+n_1}_{+}$ and $\Lambda, E \in \mathcal{B}(\mathbb{R}^{n+n_1})$ such that $\Lambda \in \mathrm{GL}^{n+n_1}$ and $\Gamma = \gamma_0 I_n \otimes \gamma_1 I_{n_1}$ with $\gamma_0, \gamma_1 > 0$, we define $P_0 := \Gamma R(I - E\Lambda^{-1})$. Corollary 3.11 provides a sufficient condition for IFP²A (64) to uniquely determine an iteration, covering several results from [56, 63]. In the following corollary we prove that Corollary 3.11 covers a trivial case where IFP²A (64) of the composite problem (45) is actually explicit. To this end we need the following technique lemma.

Lemma 3.12. *Suppose that $P \in \mathcal{B}(\mathbb{R}^{n+m})$ takes block form*

$$P = \begin{bmatrix} A & 0 \\ B & C \end{bmatrix},$$

where $A \in \mathbb{P}^n_+$, $B \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{P}^m_+$. Then there exists $\Gamma = I_n \otimes \gamma I_m$ with $\gamma > 0$ such that $\Gamma P \in \mathbb{P}^{n+m}_+$.

Proof. If $B = 0$, then we have

$$\langle z, Pz\rangle = \|x\|^2_{\bar{A}} + \|u\|^2_{\bar{C}},$$

implying that $P \in \mathbb{P}^{n+m}_+$. Suppose $B \neq 0$. Let $\Gamma = I_n \otimes \gamma I_m$ with $\gamma > 0$. Since $A \in \mathbb{P}^n_+$ and $C \in \mathbb{P}^m_+$, for any $z = (x, u) \in \mathbb{R}^n \times \mathbb{R}^m$ we have

$$\begin{aligned}
\langle z, \Gamma Pz\rangle &= \|x\|^2_{\bar{A}} + \gamma\langle Bx, u\rangle + \gamma\|u\|^2_{\bar{C}} \\
&\ge \left\|\bar{A}^{-1}\right\|^{-1}\|x\|^2 + \gamma\left\|\bar{C}^{-1}\right\|^{-1}\|y\|^2 - \gamma\|B\|\|x\|\|u\| \\
&= \left(\left\|\bar{A}^{-1}\right\|^{-1} - \frac{\gamma}{4}\|B\|^2\left\|\bar{C}^{-1}\right\|\right)\|x\|^2 + \frac{\gamma}{4}\left\|\bar{C}^{-1}\right\|^{-1}\left(2\|u\| - \|B\|\left\|\bar{C}^{-1}\right\|\|x\|\right)^2.
\end{aligned}$$

Then for any $0 < \gamma < 4\|\bar{A}^{-1}\|^{-1}\|B\|^{-2}\|\bar{C}^{-1}\|^{-1}$, we have $\langle z, \Gamma Pz\rangle \ge 0$ and furthermore, $\langle z, \Gamma Pz\rangle = 0$ if and only if $z = 0$. This finishes the proof. □

Now we are ready for the following application of IFP²A (64) to composite problem (45).

Corollary 3.13. *Suppose that* $R = I_n \otimes P^{-1} \otimes Q$ *with* $P \in \mathbb{S}^n_+$ *and* $Q \in \mathbb{S}^m_+$, $\Lambda, E \in \mathcal{B}(\mathbb{R}^{2n+m})$ *take block forms*

$$\Lambda = \begin{bmatrix} 0 & I_n & 0 \\ -P & I_n + PE_1 & 0 \\ 0 & 0 & I_m \end{bmatrix}, \qquad E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I_n + PE_1 & 0 \\ E_2 & E_3 & 0 \end{bmatrix},$$

where $E_1 \in \mathbb{R}^{n\times n}$ *and* $E_2, E_3 \in \mathbb{R}^{m\times n}$, *then* $\Lambda \in \mathrm{GL}^{2n+m}$ *and there exists* $\Gamma = \gamma_0 I_n \otimes \gamma_1 I_n \otimes \gamma_2 I_m$ *with* $\gamma_0, \gamma_1, \gamma_2 > 0$ *such that* $\Gamma P_0 \in \mathbb{P}^{2n+m}_+$; *therefore Corollary 3.11 applies.*

Proof. First notice that

$$\Lambda^{-1} = \begin{bmatrix} P^{-1} + E_1 & -P^{-1} & 0 \\ I_n & 0 & 0 \\ 0 & 0 & I_m \end{bmatrix},$$

which implies

$$R\big(I - E\Lambda^{-1}\big) = \begin{bmatrix} I_n & 0 & 0 \\ -I_n - PE_1 & P^{-1} & 0 \\ -Q\big(E_2\big(P^{-1} + E_1\big) + E_3\big) & QE_2P^{-1} & Q \end{bmatrix}.$$

Then notice that $R(I - E\Lambda^{-1})$ is strictly block-triangular with positive-definite block-diagonal matrices. By a simple induction of Lemma 3.12, there exists $\Gamma = \gamma_0 I_n \otimes \gamma_1 I_n \otimes \gamma_2 I_m$ with $\gamma_0, \gamma_1, \gamma_2 > 0$ such that $\Gamma P_0 \in \mathbb{P}^{2n+m}_+$, which closes the proof. □

Corollary 3.13 corresponds to the explicit IFP²A of composite problem (45), *i.e.*, let $z_k = (x_k, u^1_k, u^2_k)$, then

For $k \in \mathbb{N}$

$$\left\lfloor \begin{aligned} & u^1_{k+1} \leftarrow -L^\top u^2_k - \nabla_h(x_k) \\ & x_{k+1} \leftarrow \mathrm{prox}_{f,P}\big(x_k + P^{-1}u^1_{k+1} + E_1\big(u^1_{k+1} - u^1_k\big)\big) \\ & u^2_{k+1} \leftarrow \mathrm{prox}_{g^*,Q}\big(Q^{-1}Lx_k + E_2(x_{k+1} - x_k) + E_3\big(u^1_{k+1} - u^1_k\big) + u^2_k\big) \end{aligned} \right.$$

The well-definiteness of the above algorithm is self-evident, while Corollary 3.11 in addition provides the analytic property of the iteration. There are many other explicit cases for IFP²A when applied to (45), *e.g.*,

For $k \in \mathbb{N}$

$$\left\lfloor \begin{aligned} & u^1_{k+1} \leftarrow -L^\top u^2_k - \nabla_h(x_k) \\ & u^2_{k+1} \leftarrow \mathrm{prox}_{g^*,Q}\big(Q^{-1}Lx_k + E_1\big(u^1_{k+1} - u^1_k\big) + u^2_k\big) \\ & y_{k+1} \leftarrow \mathrm{prox}_{f,P}\big(x_k + P^{-1}u^1_{k+1} + E_2\big(u^1_{k+1} - u^1_k\big) + E_3\big(u^2_{k+1} - u^2_k\big)\big) \end{aligned} \right.$$

and readers are encouraged to cover these cases via Corollary 3.11 in a similar way.

3.5 CONVERGENCE ANALYSIS

This section provides comprehensive convergence analysis of the proposed IFP²A (64), which covers results from [56, 74] as special cases.

In last section we answer the first question (Q1) of IFP²A, that is, propose a sufficient condition for IFP²A (64) to uniquely determine an iteration sequence for all possible initial inputs. Then we consider the second question (Q2) of IFP²A, *i.e.*, the condition for IFP²A (64) to converge towards primal/dual solution pairs of model (44). To this end we first need to investigate the property of IFP²A iteration.

Let $\mathcal{F} : \mathbb{R}^{n+n_1} \to \mathbb{R}^{n+n_1}$ be the mapping satisfying $\mathcal{F}(w_0) = z_1$ for all $w_0 \in \mathbb{R}^{n+n_1}$ in IFP²A (64) throughout discussion, as long as it is well-defined, and define $\mathcal{F}_\lambda := (1-\lambda)I + \lambda\mathcal{F}$ for all $\lambda \in \mathbb{R}$. Then by Theorem 3.5, the primal/dual solution pairs could be characterized as $\operatorname{Fix}\mathcal{F}$, and IFP²A (64) generalizes $\{(z_k, w_k)\}_{k\in\mathbb{N}}$ by $z_{k+1} = \mathcal{F}w_k$ and $w_{k+1} = \mathcal{F}_\lambda w_k$. That means, IFP²A finds points in $\operatorname{Fix}\mathcal{F}$ by Picard iteration of $\mathcal{F}_\lambda$. Therefore we need a close study of the mapping $\mathcal{F}$. First we prove a technical lemma for smooth convex functions with Lipschitz continuous gradient.

Lemma 3.14. *Suppose that $f \in \Gamma_0(\mathbb{R}^n) \cap C_L^1(\mathbb{R}^n)$ with L-Lipschitz continuous gradient. Then for all $x, y, z \in \mathbb{R}^n$ there holds*

$$\left\langle x - z, \nabla_f(y) - \nabla_f(z) \right\rangle \geq -\frac{L}{4}\|x-y\|^2.$$

Proof. It is well-known that for $f \in \Gamma_0(\mathbb{R}^n) \cap C_L^1(\mathbb{R}^n)$ with an L-Lipschitz continuous gradient, ∇_f is L^{-1}-cocoercive, *i.e.*, for $x, z \in \mathbb{R}^n$ there holds

$$\left\langle x - z, \nabla_f(x) - \nabla_f(z) \right\rangle \geq L^{-1}\left\|\nabla_f(x) - \nabla_f(z)\right\|^2.$$

One may find the proof in [6, Corollary 18.14]. Then by completing the square we have

$$\begin{aligned}
\left\langle x - z, \nabla_f(y) - \nabla_f(z) \right\rangle &= \left\langle y - z, \nabla_f(y) - \nabla_f(z) \right\rangle + \left\langle x - y, \nabla_f(y) - \nabla_f(z) \right\rangle \\
&\geq L^{-1}\left\|\nabla_f(y) - \nabla_f(z)\right\|^2 + \left\langle x - y, \nabla_f(y) - \nabla_f(z) \right\rangle \\
&= \left\|\frac{1}{\sqrt{L}}\left(\nabla_f(y) - \nabla_f(z)\right) + \frac{\sqrt{L}}{2}(x-y)\right\|^2 - \frac{L}{4}\|x-y\|^2
\end{aligned}$$

$$\geq -\frac{L}{4}\|x-y\|^2. \qquad \square$$

Now we are ready to state the pivotal quasiaveragedness property of the mapping $\mathcal{F}$. Define

$$V := RM + (R+S)(\Lambda - I), \qquad W := \left(I - \Lambda^\top\right) R(M - I + \Lambda) + L(A\Lambda)^\top A\Lambda/4,$$

where L is the Lipschitz constant of ∇_{f_2}. For $V \in \mathbb{S}^{n+n_1}$, define

$$U_\lambda := \frac{1}{\lambda}((2-\lambda)V - 2\overline{W}),$$

for all $\lambda \in \mathbb{R}$ and $D := \{\lambda \in \mathbb{R} : U_\lambda \in \mathbb{S}^{n+n_1}\}$. Notice that by definition $0 \notin D$.

Theorem 3.15. *Suppose $P_0 \in \mathbb{P}_+^{n+n_1}$ and $R, V \in \mathbb{S}_+^{n+n_1}$. If $\lambda \in D$, then $\mathcal{F}_\lambda$ is nonexpansive with respect to V and quasiaveraged with respect to (V, U_λ). Specially, if $\overline{W} = 0$ then for $\lambda \in (0,2]$, $\mathcal{F}_\lambda$ is nonexpansive with respect to V and quasiaveraged with respect to $(V, (2-\lambda)V/\lambda)$.*

Proof. Suppose $x, y \in \mathbb{R}^{n+n_1}$ and set $u := \mathcal{F}x$, $v := \mathcal{F}y$. Then by (51) and (64), we have

$$\begin{cases} R(E-\Lambda)u + R(M-I+\Lambda)x - A^\top \nabla_{f_2}(Ax) \in \partial_\Phi(\Lambda u + (I-\Lambda)x), \\ R(E-\Lambda)v + R(M-I+\Lambda)y - A^\top \nabla_{f_2}(Ay) \in \partial_\Phi(\Lambda v + (I-\Lambda)y). \end{cases}$$

Then by monotonicity of subdifferentials, $E + M = I + R^{-1}S$ and skew-symmetry of S, we have

$$\begin{aligned} 0 \leq & \Big\langle \Lambda(u-v) + (I-\Lambda)(x-y), \\ & \qquad R(E-\Lambda)(u-v) + R(M-I+\Lambda)(x-y) - A^\top\left(\nabla_{f_2}(Ax) - \nabla_{f_2}(Ay)\right)\Big\rangle \\ = & \langle (I-\Lambda)((x-y)-(u-v)), R(M-I+\Lambda)((x-y)-(u-v))\rangle \\ & + \langle (I-\Lambda)((x-y)-(u-v)), S(u-v)\rangle \\ & + \langle u-v, R(M-I+\Lambda)((x-y)-(u-v))\rangle \\ & - \Big\langle \Lambda(u-v) + (I-\Lambda)(x-y), A^\top\left(\nabla_{f_2}(Ax) - \nabla_{f_2}(Ay)\right)\Big\rangle. \end{aligned}$$

Notice that by Lemma 3.14 we have

$$\begin{aligned} & \Big\langle \Lambda(u-v) + (I-\Lambda)(x-y), A^\top\left(\nabla_{f_2}(Ax) - \nabla_{f_2}(Ay)\right)\Big\rangle \\ = & \Big\langle A\Lambda(u-v) + (A - A\Lambda)(x-y), \left(\nabla_{f_2}(Ax) - \nabla_{f_2}(Ay)\right)\Big\rangle \\ \geq & -\frac{L}{4}\|A\Lambda((x-y)-(u-v))\|^2, \end{aligned}$$

which leads us to

$$\begin{aligned}0 &\le 2\langle (x-y)-(u-v), \overline{W}((x-y)-(u-v))\rangle + 2\langle u-v, V((x-y)-(u-v))\rangle \\ &= \|x-y\|_V^2 - \|u-v\|_V^2 - \langle (u-v)-(x-y), (V-2\overline{W})((u-v)-(x-y))\rangle.\end{aligned}$$

Furthermore, if set $p := \mathcal{F}_\lambda x = (1-\lambda)x + \lambda u$ and $q := \mathcal{F}_\lambda y = (1-\lambda)y + \lambda v$, then by Proposition 2.21 we have

$$\|p-q\|_V^2 = (1-\lambda)\|x-y\|_V^2 + \lambda\|u-v\|_V^2 - \lambda(1-\lambda)\|(x-y)-(u-v)\|_V^2,$$

which implies

$$\|p-q\|_V^2 + \|(x-y)-(p-q)\|_{((2-\lambda)V-2\overline{W})/\lambda}^2 \le \|x-y\|_V^2.$$

This proves that if $U_\lambda \in \mathbb{S}^n$ then $\mathcal{F}_\lambda$ is nonexpansive with respect to V. To prove the quasiaveragedness, set $y \in \operatorname{Fix} \mathcal{F}_\lambda$ then we have $q = \mathcal{F}_\lambda y = y$ and therefore

$$\|\mathcal{F}_\lambda x - y\|_V^2 + \|\mathcal{F}_\lambda x - x\|_{U_\lambda}^2 \le \|x-y\|_V^2,$$

which proves the quasiaveragedness of $\mathcal{F}_\lambda$ with respect to (V, U_λ).

Finally, if $\overline{W} = 0$, then it is clear that for all $\lambda \in (0,2]$, $U_\lambda = (2-\lambda)V/\lambda \in \mathbb{S}^{n+n_1}$, which with preceding arguments proves that $\mathcal{F}_\lambda$ is nonexpansive with respect to V and quasiaveraged with respect to $(V, (2-\lambda)V/\lambda)$. □

With the nonexpansiveness and quasiaveragedness property of $\mathcal{F}_\lambda$, we are now capable to prove the convergence theorem for IFP²A (64) via Theorem 2.30.

Theorem 3.16. *Suppose $P_0 \in \mathbb{P}_+^{n+n_1}$ and $V \in \mathbb{S}^{n+n_1}$. If $\lambda \in D$ and*

$$\operatorname{Ker} V \cup \operatorname{Ker} U_\lambda \subseteq \operatorname{Ker}(I-\Lambda) \cap \operatorname{Ker} M \cap \operatorname{Ker} A,$$

then the sequence $\{z_k\}_{k\in\mathbb{N}}$ and $\{w_k\}_{k\in\mathbb{N}}$ generated by IFP²A (64) *converge to a same fixed-point of $\mathcal{F}$ as a primal-dual solution pair of model* (44).

Proof. This proof mainly follows Theorems 2.30 and 3.15. First notice that for any $x \in \mathbb{R}^{n+n_1}$ and $y \in \operatorname{Ker}(I-\Lambda) \cap \operatorname{Ker} M \cap \operatorname{Ker} A$, we have

$$z = \mathcal{F}(x+y) \quad \Longleftrightarrow \quad \Lambda z + (I-\Lambda)(x+y) = \operatorname{prox}_{\Phi,R}\big(Ez + M(x+y) - R^{-1}A^\top \nabla_{f_2}(A(x+y))\big)$$

$$\begin{aligned} \iff \quad & \Lambda z + (I - \Lambda)x = \operatorname{prox}_{\Phi,R}\big(Ez + Mx - R^{-1}A^\top \nabla_{f_2}(Ax)\big) \\ \iff \quad & z = \mathcal{F}x, \end{aligned}$$

which proves that

$$\operatorname{Ker}(I - \Lambda) \cap \operatorname{Ker} M \cap \operatorname{Ker} A \subseteq \operatorname{Ker} \mathcal{F}$$

and therefore $\operatorname{Ker} V \cup \operatorname{Ker} U_\lambda \subseteq \operatorname{Ker} \mathcal{F}$. Then by Theorems 2.30 and 3.15, if $U_\lambda \in \mathbb{S}^{n+n_1}$ then for all $w_0 \in \mathbb{R}^n$, sequence $\{w_k\}_{k\in\mathbb{N}}$ generated by IFP²A (64) converges to a point in $\operatorname{Fix} \mathcal{F}_\lambda = \operatorname{Fix} \mathcal{F}$. By continuity of $\mathcal{F}$, we have that $z_{k+1} = \mathcal{F}w_k$ converges to a same limit. Finally by Theorem 3.5, $\{z_k\}_{k\in\mathbb{N}}$ and $\{w_k\}_{k\in\mathbb{N}}$ converges to a same primal/dual solution pair of problem (44). □

Notice that $V, U_\lambda \in \mathbb{S}_+^{n+n_1}$ is sufficient for convergence by Theorem 3.16. When $\overline{W} = 0$, Theorem 3.16 could be extended to KM iteration via Theorems 2.32 and 3.15. Consider the following KM IFP²A

$$\begin{aligned} &\textbf{For } k \in \mathbb{N} \\ &\left\lfloor \begin{array}{l} \Lambda z_{k+1} + (I - \Lambda)w_k = \operatorname{prox}_{\Phi,R}\big(Ez_{k+1} + Mw_k - R^{-1}A^\top \nabla_{f_2}(Aw_k)\big) \\ w_{k+1} \leftarrow (1 - \lambda_k)w_k + \lambda_k z_{k+1} \end{array} \right. \end{aligned} \tag{67}$$

where $\{\lambda_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$.

Theorem 3.17. *Suppose $P_0 \in \mathbb{P}_+^{n+n_1}$, $V \in \mathbb{S}^{n+n_1}$, $\overline{W} = 0$ and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 2]$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k(2 - \lambda_k) = \infty \quad \textit{and} \quad \operatorname{Ker} V \subseteq \operatorname{Ker}(I - \Lambda) \cap \operatorname{Ker} M \cap \operatorname{Ker} A,$$

then the sequence $\{z_k\}_{k\in\mathbb{N}}$ generated by KM IFP²A (67) converges to a fixed-point $z_* \in \mathbb{R}^{2n+m}$ of $\mathcal{F}$ as a primal/dual solution pair of model (44), and $\{\sqrt{V}w_k\}_{k\in\mathbb{N}}$ converges to $\sqrt{V}z_*$.

Proof. Theorems 2.32 and 3.15 proves that there exists $z_* \in \operatorname{Fix} T$ such that $\{\sqrt{V}w_k\}_{k\in\mathbb{N}}$ converges to $\sqrt{V}z_*$. Then $\operatorname{Ker}\sqrt{V} \subseteq \operatorname{Ker}\mathcal{F}$ and the continuity of $\mathcal{F}$ show that $z_{k+1} = \mathcal{F}w_k = \mathcal{F}\sqrt{V}^\dagger\sqrt{V}w_k \to \mathcal{F}\sqrt{V}^\dagger\sqrt{V}z_* = Tz_* = z_* \in \operatorname{Fix} T$ as $k \to \infty$. □

Theorems 3.16 and 3.17 provide general convergence theorems for IFP²A (64), which cover various existing results under different settings of model (44). In the following section, we covers several popular existing explicit algorithms as special cases of the proposed framework of IFP²A.

3.6 EXPLICIT FIXED-POINT PROXIMITY ALGORITHMS

In this section we apply convergence analysis Theorem 3.16 to several popular explicit methods of the composite model (45) under various situations, illustrating the generality of the proposed framework of IFP²A.

Recall that the composite problem once discussed in Section 3.1 is

$$\operatorname*{argmin}_{x\in\mathbb{R}^n}\{f(x)+g(Lx)+h(x)\}, \tag{45}$$

where $f\in\Gamma_0(\mathbb{R}^n)$, $g\in\Gamma_0(\mathbb{R}^m)$, $h\in\Gamma_0(\mathbb{R}^n)\cap C^1_L(\mathbb{R}^n)$ and $L\in\mathbb{R}^{m\times n}$. Denote L in this section as the Lipschitz constant of ∇_h. The corresponding IFP²A is

$$\begin{array}{l}\textbf{For } k\in\mathbb{N}\\ \left\lfloor\begin{array}{l}\Lambda z_{k+1}+(I-\Lambda)w_k=\operatorname{prox}_{\Phi,R}\big(Ez_{k+1}+Mw_k-R^{-1}A^\top\nabla_{f_2}(Aw_k)\big)\\ w_{k+1}\leftarrow(1-\lambda)w_k+\lambda z_{k+1}\end{array}\right.\end{array} \tag{68}$$

where $w_k=(u_k,v_k^1,v_k^2), z_k=(x_k,y_k^1,y_k^2)\in\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^m$, $\Phi=0\oplus f^*\oplus g^*$, $R\in\mathbb{S}_+^{(2n+m)\times(2n+m)}$, $\Lambda,E,M\in\mathcal{B}(\mathbb{R}^{2n+m})$ and $A=I_n\oplus 0_{n\times(n+m)}\in\mathbb{R}^{n\times(2n+m)}$. In some cases we further consider the KM IFP²A as

$$\begin{array}{l}\textbf{For } k\in\mathbb{N}\\ \left\lfloor\begin{array}{l}\Lambda z_{k+1}+(I-\Lambda)w_k=\operatorname{prox}_{\Phi,R}\big(Ez_{k+1}+Mw_k-R^{-1}A^\top\nabla_{f_2}(Aw_k)\big)\\ w_{k+1}\leftarrow(1-\lambda_k)w_k+\lambda_k z_{k+1}\end{array}\right.\end{array} \tag{69}$$

where $\{\lambda_k\}_{k\in\mathbb{N}}\subset\mathbb{R}$.

3.6.1 GRADIENT DESCENT METHOD

This subsection identifies gradient descent method [17] as a special case of IFP²A, and then propose the convergence theorem under Theorem 3.16.

Let $f=g=0$ in composite problem (45), which then reduces to

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} h(x). \tag{70}$$

In IFP²A (68) we let $\Lambda=I$, $R=H\otimes I_{n+m}$ with $H\in\mathbb{S}^n_+$, $E=0$, $\lambda=1$ and

$$M=\begin{bmatrix}I_n & 0\\ 0 & 0_{n+m}\end{bmatrix}.$$

Then IFP²A (68) reduces to iteration

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \lfloor\; x_{k+1} \leftarrow x_k - H^{-1}\nabla_h(x_k)\end{array} \tag{71}$$

Gradient descent method (71) serves as a fundamental method for smooth convex optimization. First notice that (71) is self-evidently well-defined, which is covered by Theorem 3.10. To see this, only need to notice that in this case $R(I - E\Lambda^{-1}) = R$. Then we turn to the convergence of iteration (71). Notice that by definition,

$$V = \begin{bmatrix} H & 0 \\ 0 & 0_{n+m} \end{bmatrix}, \qquad W = \frac{L}{4}\begin{bmatrix} I_n & 0 \\ 0 & 0_{n+m} \end{bmatrix}, \qquad U_1 = V - 2\overline{W} = \begin{bmatrix} H - L/2 & 0 \\ 0 & 0_{n+m} \end{bmatrix}.$$

Therefore by Theorem 3.16, we have the convergence theorem of iteration (71).

Theorem 3.18. *If $H - L/2 \in \mathbb{S}_+^n$, then for any $x_0 \in \mathbb{R}^n$, the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by gradient descent method* (71) *converges to a solution of model* (70).

Proof. By Theorem 3.16, it is sufficient to check

$$\operatorname{Ker} V \cup \operatorname{Ker} U_1 \subseteq \operatorname{Ker}(I - \Lambda) \cap \operatorname{Ker} M \cap \operatorname{Ker} A.$$

Notice that $\operatorname{Ker} V = \operatorname{Ker} M = \operatorname{Ker} A = \{0_n\} \times \mathbb{R}^{n+m}$ and $\operatorname{Ker}(I - \Lambda) = \mathbb{R}^{2n+m}$. If $H - L/2 \in \mathbb{S}_+^n$ then we have $\operatorname{Ker} U_1 = \{0_n\} \times \mathbb{R}^{n+m}$, which proves the above inclusion. □

3.6.2 PROXIMAL POINT METHOD

This subsection identifies proximal point method [71, 72] as a special case of IFP²A, and provide a generalization of this method under Theorem 3.17.

Let $g = h = 0$ in composite problem (45), which then reduces to

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} f(x). \tag{72}$$

In IFP²A (68) we let $R = I_n \otimes P^{-1} \otimes I_m$ with $P \in \mathbb{S}_+^n$, $E = I_n \otimes I_n \otimes 0_m$ and

$$\Lambda = \begin{bmatrix} I_n & I_n & 0 \\ -P & I_n & 0 \\ 0 & 0 & I_m \end{bmatrix}, \qquad M = \begin{bmatrix} 0 & -I_n & 0 \\ P & 0 & 0 \\ 0 & 0 & I_m \end{bmatrix}.$$

Then KM IFP²A version of iteration (68) reduces to iteration

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} x_{k+1} \leftarrow \operatorname{prox}_{f,P}(u_k) \\ u_{k+1} \leftarrow (1-\lambda_k)u_k + \lambda_k x_{k+1} \end{array} \right. \end{array} \tag{73}$$

Proximal point method (73) serves as a fundamental method for nonsmooth convex optimization. First notice that (73) is self-evidently well-defined, which is covered by Theorem 3.10. To see this, only need to notice that

$$R\big(I - E\Lambda^{-1}\big) = \begin{bmatrix} P(I+P)^{-1} & (I+P)^{-1} & 0 \\ -(I+P)^{-1} & (I+P)^{-1} & 0 \\ 0 & 0 & I_m \end{bmatrix},$$

therefore setting $\Gamma = I$ gives $P_0 \in \mathbb{P}_+^{2n+m}$. Then we turn to the convergence of iteration (73). Notice that by definition,

$$V = P \otimes I_n \otimes I_m, \qquad W = 0.$$

Therefore by Theorem 3.17, we have the convergence theorem of iteration (73).

Theorem 3.19. *Suppose $P \in \mathbb{S}_+^n$ and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,2]$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = \infty,$$

then for all $u_0 \in \mathbb{R}^n$, sequences $\{u_k\}_{k\in\mathbb{N}}$ and $\{x_k\}_{k\in\mathbb{N}}$ generated by proximal point method (73) *converges to a same solution of model* (72).

Proof. Notice that in this case $\operatorname{Ker} V = \{0\}$ and $\overline{W} = 0$, then a direct application of Theorem 3.17 finishes the proof. □

Here we consider an extension of iteration (73). Consider

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \lfloor\; x_{k+1} \leftarrow (I-Q)x_k + Q\,\mathrm{prox}_f(x_k)\end{array} \tag{74}$$

where $Q \in \mathbb{S}_+^n$. The parameters of iteration (74) are $R = I_n \otimes I_n \otimes I_m$, $E = Q^{-1} \otimes I_n \otimes 0_m$, $\lambda = 1$ and

$$\Lambda = \begin{bmatrix} Q^{-1} & I_n & 0 \\ -Q^{-1} & I_n & 0 \\ 0 & 0 & I_m \end{bmatrix}, \qquad M = \begin{bmatrix} I-Q^{-1} & -I_n & 0 \\ I_n & 0 & 0 \\ 0 & 0 & I_m \end{bmatrix}.$$

Notice that in this case

$$V = Q^{-1} \otimes I_n \otimes I_m, \qquad W = \left(Q^{-1}\left(I - Q^{-1}\right)\right) \otimes 0_n \otimes 0_m.$$

Theorem 3.20. *Suppose $P, Q \in \mathbb{S}_+^n$. If $2I - Q \in \mathbb{S}_+^n$ then for all $x_0 \in \mathbb{R}^n$, sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by generalized proximal point method* (74) *converges to a solution of model* (72).

Proof. Notice that in this case $U_1 = (Q^{-1}(2Q^{-1} - I)) \otimes I_n \otimes I_m$. If $2I - Q \in \mathbb{S}_+^n$, then we have $2Q^{-1} - I \in \mathbb{S}_+^n$ and therefore $U_1 \in \mathbb{S}_+^{2n+m}$. Further notice that $\operatorname{Ker} V = \operatorname{Ker} U_1 = \{0\}$, then Theorem 3.16 closes the proof. □

3.6.3 DOUGLAS-RACHFORD SPLITTING ALGORITHM

This subsection identifies the Douglas-Rachford splitting (DRS) algorithm [37, 62] to be a spacial case of IFP²A (64). Then based on Theorem 3.17 we propose a generalized DRS with convergence theorem.

Let $L = I_n$ and $h = 0$ in composite problem (45), which then reduces to

$$\operatorname*{argmin}_{x\in\mathbb{R}^n} \{f(x) + g(x)\}. \tag{75}$$

DRS designates the following scheme for (75), as

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \lfloor\; t_{k+1} \leftarrow \mathrm{prox}_f\left(2\,\mathrm{prox}_g(t_k) - t_k\right) + t_k - \mathrm{prox}_g(t_k)\end{array} \tag{76}$$

which has been proven to have $\{\mathrm{prox}_g(t_k)\}_{k\in\mathbb{N}}$ converge to a solution of problem (75) with any $t_0 \in \mathbb{R}^n$ in [7, 39].

DRS (76) is equivalent to a special case of IFP²A (67). To see this, set $x_k := t_k - \mathrm{prox}_{g^*}(t_{k-1})$, $y_k^1 := -\mathrm{prox}_{g^*}(t_{k-1})$ and $y_k^2 := \mathrm{prox}_{g^*}(t_{k-1})$, then DRS (76) is identical to KM IFP²A (67) with $R = I_n \otimes I_n \otimes I_n$, $E = 0_n \otimes 2I_n \otimes 0_n$, $\lambda = 1$ and

$$\Lambda = \begin{bmatrix} 0 & I_n & I_n \\ -I_n & 2I_n & 0 \\ 0 & 0 & I_n \end{bmatrix}, \qquad M = \begin{bmatrix} I_n & -I_n & -I_n \\ I_n & -I_n & 0 \\ I_n & 0 & I_n \end{bmatrix}.$$

Notice that in this case

$$V = \begin{bmatrix} I_n & -I_n & 0 \\ -I_n & I_n & I_n \\ 0 & I_n & 2I_n \end{bmatrix}, \qquad W = 0.$$

Although $V \notin \mathbb{S}^{3n}$, notice that $\mathrm{Ran}\,\mathcal{F} \subseteq \mathcal{H} := \{(x, y, -y) \in \mathbb{R}^{3n} : x, y \in \mathbb{R}^n\}$ and $V \in \mathbb{S}(\mathcal{H})$. Here $\mathbb{S}(\mathcal{H})$ denotes all symmetric positive-definite matrices on $\mathcal{H}$. To prove this, for any $z = (x, y, -y) \in \mathcal{H}$, we have

$$\langle z, Vz\rangle = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2 = \|x - y\|^2 \geq 0.$$

Therefore DRS (76) could be covered by KM IFP²A (67) with respect to space $\mathcal{H}$. Here we consider the following generalization of DRS (76) as

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \left| \begin{array}{l} y_{k+1} \leftarrow \mathrm{prox}_{g^*,Q}\big(Q^{-1}u_k + v_k\big) \\ x_{k+1} \leftarrow \mathrm{prox}_{f,P}\big(P^{-1}(v_k - 2y_{k+1}) + u_k\big) \\ u_{k+1} \leftarrow (1 - \lambda_k)u_k + \lambda_k x_{k+1} \\ v_{k+1} \leftarrow (1 - \lambda_k)v_k + \lambda_k y_{k+1} \end{array} \right. \end{array} \tag{77}$$

Clear that DRS (76) is a special case of iteration (77) with $P = Q = I$, $\lambda_k = 1$ for $k \in \mathbb{N}$ and $u_1 = t_1 - \mathrm{prox}_{g^*}(t_0)$, $v_1 = \mathrm{prox}_{g^*}(t_0)$; actually the sequence $\{t_k\}_{k\in\mathbb{N}}$ by (76) is identical to $\{x_k + y_k\}_{k\in\mathbb{N}}$ by (77) under the given parameters. Convergence of iteration (77) is governed by Theorem 3.17.

Theorem 3.21. *Suppose $P, Q \in \mathbb{S}^n$ such that $\|(\sqrt{Q}\sqrt{P})^{-1}\| \leq 1$ and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 2]$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k(2 - \lambda_k) = \infty,$$

then sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}}$ *generated by generalized DRS* (77) *converges to* $(x_*, y_*) \in \mathbb{R}^{2n}$*, with* x_* *is a solution of model* (75) *and* $\{(Pu_k + v_k, u_k + Qv_k)\}_{k\in\mathbb{N}}$ *converges to* $(Px_* + y_*, x_* + Qy_*)$*.*

Proof. First notice that with substitutions $w_k := (u_k, -v_k, v_k)$ and $z_k := (x_k, -y_k, y_k)$ for $k \in \mathbb{N}$, the generalized DRS (77) corresponds to IFP²A (67) with parameters $R = I_n \otimes P^{-1} \otimes Q$, $E = 0_n \otimes 2I_n \otimes 0_n$ and

$$\Lambda = \begin{bmatrix} I_n & 0 & 0 \\ -P & 2I_n & 0 \\ 0 & 0 & I_n \end{bmatrix}, \qquad M = \begin{bmatrix} 0_n & 0 & 0 \\ P & -I_n & 0 \\ Q^{-1} & 0 & I_n \end{bmatrix}.$$

Notice that in this case

$$V = \begin{bmatrix} P & -I_n & 0 \\ 0 & 0_n & 0 \\ I_n & 0 & Q \end{bmatrix}, \qquad W = 0.$$

Although $V \notin \mathbb{S}^n$, notice $\operatorname{Ran} \mathcal{F} \subseteq \mathcal{H} := \{(x, -y, y) \in \mathbb{R}^{3n} : x, y \in \mathbb{R}^n\}$ and for any $z = (x, -y, y) \in \mathcal{H}$ there holds

$$Vz = V'z, \qquad \text{where } V' := \begin{bmatrix} P & 0 & I_n \\ 0 & 0_n & 0 \\ I_n & 0 & Q \end{bmatrix},$$

therefore we study V' within $\mathcal{H}$ instead of V. If $\|(\sqrt{Q}\sqrt{P})^{-1}\| \leq 1$ then for any $z = (x, -y, y) \in \mathcal{H}$,

$$\begin{aligned}
\left\langle z, V'z \right\rangle &= \|x\|_P^2 + 2\langle x, y\rangle + \|y\|_Q^2 \\
&= \left\|\sqrt{P}x\right\|^2 + 2\left\langle \sqrt{P}x, \left(\sqrt{Q}\sqrt{P}\right)^{-1}\sqrt{Q}y \right\rangle + \left\|\sqrt{Q}y\right\|^2 \\
&\geq \left\|\sqrt{P}x\right\|^2 - 2\left\|\left(\sqrt{Q}\sqrt{P}\right)^{-1}\right\|\left\|\sqrt{P}x\right\|\left\|\sqrt{Q}y\right\| + \left\|\sqrt{Q}y\right\|^2 \\
&\geq \left\|\sqrt{P}x\right\|^2 - 2\left\|\sqrt{P}x\right\|\left\|\sqrt{Q}y\right\| + \left\|\sqrt{Q}y\right\|^2 \\
&= \left(\left\|\sqrt{P}x\right\| - \left\|\sqrt{Q}y\right\|\right)^2 \geq 0,
\end{aligned}$$

which proves that $V' \in \mathbb{S}(\mathcal{H})$. Also notice that in space $\mathcal{H}$,

$$\begin{aligned}
\operatorname{Ker} V' &= \left\{(x, -y, y) \in \mathbb{R}^{3n} : Px + y = x + Qy = 0\right\}, \\
\operatorname{Ker}(I - \Lambda) &= \left\{(x, -y, y) \in \mathbb{R}^{3n} : Px + y = 0\right\}, \\
\operatorname{Ker} M &= \left\{(x, -y, y) \in \mathbb{R}^{3n} : Px + y = Q^{-1}x + y = 0\right\},
\end{aligned}$$

which shows $\operatorname{Ker} V' \subseteq \operatorname{Ker}(I - \Lambda) \cap \operatorname{Ker} M$ within $\mathcal{H}$. Then an application of Theorem 3.17 within space $\mathcal{H}$ proves that $\{z_k\}_{k\in\mathbb{N}}$ converges to $z_* = (x_*, -y_*, y_*) \in \mathcal{H}$ as a primal/dual solution pair of problem (75) and $\{V'w_k\}_{k\in\mathbb{N}} = \{(Pu_k + v_k, 0, u_k + Qv_k)\}_{k\in\mathbb{N}}$ converges to $Vz_* = (Px_* + y_*, 0, x_* + Qy_*)$. □

The convergence of DRS (76) is clearly a special case of Theorem 3.21 with $P = Q = I$ and $\lambda_k = 1$ for $k \in \mathbb{N}$. In this case $t_k = x_k + y_k \to x_* + y_*$ and therefore $\mathrm{prox}_g(t_k) = t_k - \mathrm{prox}_{g^*}(t_k) \to x_* + y_* - \mathrm{prox}_{g^*}(x_* + y_*) = x_*$ as $k \to \infty$.

3.6.4 FIRST-ORDER PRIMAL-DUAL ALGORITHM

This subsection identifies the First-Order Primal-Dual algorithm (FOPD) [19, 41], also referred as Primal-Dual Hybrid Gradient method (PDHG) [46], to be a special case of explicit IFP²A (68).

Let $h = 0$ in composite problem (45), which then reduces to

$$\underset{x\in\mathbb{R}^n}{\mathrm{argmin}}\{f(x) + g(Lx)\}. \tag{78}$$

FOPD designates the following scheme for (78), as

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} x_{k+1} \leftarrow \mathrm{prox}_{f,\alpha I}\left(x_k - \alpha^{-1}L^\top y_k\right) \\ y_{k+1} \leftarrow \mathrm{prox}_{g^*,\beta I}\left(\beta^{-1}L(2x_{k+1} - x_k) + y_k\right) \end{array}\right.\end{array} \tag{79}$$

where $\alpha, \beta > 0$. FOPD (79) is equivalent to a special case of IFP²A (67). To see this, set $z_k := (x_k, -L^\top y_k, y_k)$ for $k \in \mathbb{N}$, then FOPD (79) is identical to KM IFP²A (69) with $R = I_n \otimes \alpha^{-1}I_n \otimes \beta I_n$, $\lambda_k = 1$ for $k \in \mathbb{N}$ and

$$\Lambda = \begin{bmatrix} I_n & 0 & 0 \\ -\alpha I_n & 0_n & 0 \\ 0 & 0 & I_m \end{bmatrix}, \qquad E = \begin{bmatrix} I_n & 0 & 0 \\ 0 & 0_n & 0 \\ 2\beta^{-1}L & 0 & 0_m \end{bmatrix}, \qquad M = \begin{bmatrix} 0_n & -I_n & -L^\top \\ \alpha I_n & I_n & 0 \\ -\beta^{-1}L & 0 & I_m \end{bmatrix}.$$

Notice that in this case

$$V = \begin{bmatrix} \alpha I_n & 0 & -L^\top \\ 0 & 0_n & 0 \\ -L & 0 & \beta I_m \end{bmatrix}, \qquad W = 0.$$

Here we consider a generalization of FOPD as

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} x_{k+1} \leftarrow \operatorname{prox}_{f,P}\big(u_k - P^{-1}L^\top v_k\big) \\ y_{k+1} \leftarrow \operatorname{prox}_{g^*,Q}\big(Q^{-1}L(2x_{k+1} - u_k) + v_k\big) \\ u_{k+1} \leftarrow (1-\lambda_k)u_k + \lambda_k x_{k+1} \\ v_{k+1} \leftarrow (1-\lambda_k)v_k + \lambda_k y_{k+1} \end{array}\right.\end{array} \tag{80}$$

where $P \in \mathbb{S}_+^n$ and $Q \in \mathbb{S}_+^m$ are called the *preconditioned* matrices. Clear that FOPD (79) is a special case of generalized FOPD (80) with $P = \alpha I_n$ and $Q = \beta I_m$ and $\lambda_k = 1$ for $k \in \mathbb{N}$.

Theorem 3.22. *Suppose $P \in \mathbb{S}_+^n$ and $Q \in \mathbb{S}_+^m$ such that $\|\sqrt{Q}^{-1} L \sqrt{P}^{-1}\| \leq 1$, and $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,2]$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = \infty,$$

then sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}}$ generated by generalized FOPD (80) converges to $(x_*, y_*) \in \mathbb{R}^n \times \mathbb{R}^m$, with x_* is a solution of model (78) and $\{Pu_k - L^\top v_k\}_{k\in\mathbb{N}}$ converges to $Px_* - L^\top y_*$.

Proof. First notice that with substitutions $w_k := (u_k, -L^\top v_k, v_k)$ and $z_k := (x_k, -L^\top y_k, y_k)$ for $k \in \mathbb{N}$, the generalized FOPD (80) corresponds to IFP²A (67) with parameters $R = I_n \otimes P^{-1} \otimes Q$, and

$$\Lambda = \begin{bmatrix} I_n & 0 & 0 \\ -P & 0_n & 0 \\ 0 & 0 & I_m \end{bmatrix}, \qquad E = \begin{bmatrix} I_n & 0 & 0 \\ 0 & 0_n & 0 \\ 2Q^{-1}L & 0 & 0_m \end{bmatrix}, \qquad M = \begin{bmatrix} 0_n & -I_n & -L^\top \\ P & I_n & 0 \\ -Q^{-1}L & 0 & I_m \end{bmatrix}.$$

Notice that although $\Lambda \notin \mathrm{GL}^{2n+m}$, there still holds $\operatorname{Ran} \mathcal{F} \subseteq \mathcal{H} := \{(x, L^\top y, -y) \in \mathbb{R}^{2n+m} : x \in \mathbb{R}^n, y \in \mathbb{R}^m\}$ and Λ is invertible in $\mathcal{H}$; to see this, for any $z = (x, L^\top y, -y), w = (u, L^\top v, -v) \in \mathcal{H}$, we have

$$\Lambda(z - w) = 0 \quad \iff \quad \begin{cases} x = u \\ y = v \end{cases} \quad \iff \quad z = w,$$

which proves Λ has an inverse in $\mathcal{H}$. Therefore in this case

$$V = \begin{bmatrix} P & 0 & -L^\top \\ 0 & 0_n & 0 \\ -L & 0 & Q \end{bmatrix}, \qquad W = 0.$$

If $\|\sqrt{Q}^{-1}L\sqrt{P}^{-1}\| \leq 1$, then for $z = (x, L^\top y, -y) \in \mathcal{H}$ we have

$$\begin{aligned}\langle z, Vz\rangle &= \|x\|_P^2 + 2\langle Lx, y\rangle + \|y\|_Q^2 \\ &= \left\|\sqrt{P}x\right\|^2 + 2\left\langle \sqrt{Q}^{-1}L\sqrt{P}^{-1}\sqrt{P}x, \sqrt{Q}y\right\rangle + \left\|\sqrt{Q}y\right\|^2 \\ &\geq \left\|\sqrt{P}x\right\|^2 - 2\left\|\sqrt{P}x\right\|\left\|\sqrt{Q}y\right\| + \left\|\sqrt{Q}y\right\|^2 \\ &= \left(\left\|\sqrt{P}x\right\| - \left\|\sqrt{Q}y\right\|\right)^2 \geq 0,\end{aligned}$$

which proves that $V \in \mathbb{S}(\mathcal{H})$. Also notice that

$$\begin{aligned}\operatorname{Ker} V' &= \left\{ \left(x, L^\top y, -y\right) \in \mathcal{H} \;:\; Px + L^\top y = Lx + Qy = 0 \right\}, \\ \operatorname{Ker}(I - \Lambda) &= \left\{ \left(x, L^\top y, -y\right) \in \mathcal{H} \;:\; Px + L^\top y = 0 \right\}, \\ \operatorname{Ker} M &= \left\{ \left(x, L^\top y, -y\right) \in \mathcal{H} \;:\; Px + L^\top y = Lx + Qy = 0 \right\}.\end{aligned}$$

Then an application of Theorem 3.17 within space $\mathcal{H}$ proves that $\{z_k\}_{k\in\mathbb{N}}$ converges to $z_* = (x_*, -L^\top y_*, y_*) \in \mathcal{H}$ as a primal/dual solution pair of problem (75) and $\lim_{k\to\infty} Vw_k = Vz_*$. □

The special case for FOPD (80) with condition $\lambda_k = 1$ for $k \in \mathbb{N}$ and $\|\sqrt{Q}^{-1}L\sqrt{P}^{-1}\| < 1$ was established in [56, 81].

3.6.5 ALTERNATING DIRECTION METHOD OF MULTIPLIERS

This subsection identifies the so-called *Alternating Direction Method of Multipliers* (ADMM) [10, 40, 45], which is identical to the *split Bregman iteration* [47], to be exactly a special case of KM IFP²A (67). ADMM is known to diverge under certain conditions, however it still has some important properties, such as dual convergence. Both of the primal/dual convergence are discussed in this subsection.

Consider the problem

$$\underset{\substack{x^1\in\mathbb{R}^{n_1}, x^2\in\mathbb{R}^{n_2} \\ L_1x^1 + L_2x^2 = c}}{\operatorname{argmin}} \left\{ f_1\left(x^1\right) + f_2\left(x^2\right) \right\}, \tag{81}$$

where $f_1 \in \Gamma_0(\mathbb{R}^{n_1})$, $f_2 \in \Gamma_0(\mathbb{R}^{n_2})$, $L_1 \in \mathbb{R}^{m\times n_1}$, $L_2 \in \mathbb{R}^{m\times n_2}$ and $c \in \mathbb{R}^m$. ADMM designates the following scheme to solve (81).

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} \text{Find } x_{k+1}^2 \in \operatorname{argmin}_{x^2 \in \mathbb{R}^{n_2}} \left\{ f_2\left(x^2\right) + \frac{1}{2}\left\|L_1 x_k^1 + L_2 x^2 + y_k - c\right\|^2 \right\} \\ y_{k+1} \leftarrow y_k + L_1 x_k^1 + L_2 x_{k+1}^2 - c \\ \text{Find } x_{k+1}^1 \in \operatorname{argmin}_{x^1 \in \mathbb{R}^{n_1}} \left\{ f_1\left(x^1\right) + \frac{1}{2}\left\|L_1 x^1 + L_2 x_{k+1}^2 + y_{k+1} - c\right\|^2 \right\} \end{array} \right.\end{array} \tag{82}$$

where the updates are denoted by '∈' because the solutions might not be unique. Problem (81) is a special case of composite problem (45), with $f = f_1 \oplus f_2$, $g = \iota_{\{c\}}$, $h = 0$ and $L = L_1 \oplus L_2$, and ADMM is as well a special case of KM IFP²A (69). To see this, notice that (82) is equivalent to

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} x_{k+1}^1 \in \operatorname{prox}_{f_1}\left(x_{k+1}^1 + L_1^\top L_1\left(x_k^1 - x_{k+1}^1\right) - L_1^\top(2y_{k+1} - y_k)\right) \\ x_{k+1}^2 \in \operatorname{prox}_{f_2}\left(x_{k+1}^2 - L_2^\top y_{k+1}\right) \\ y_{k+1} \leftarrow \operatorname{prox}_{\iota_{\{c\}}^*}\left(L_1 x_k^1 + L_2 x_{k+1}^2 + y_k\right) \end{array} \right.\end{array} \tag{83}$$

On the other hand, if set $z_k = (x_k^1, x_k^2, -L_1^\top y_k, -L_2^\top y_k, y_k)$ and $\lambda_k = 1$ for $k \in \mathbb{N}$, $R = I_{n_1} \otimes I_{n_2} \otimes I_{n_1} \otimes I_{n_2} \otimes I_m$ and

$$\Lambda = \begin{bmatrix} I_{n_1} & 0 & 0 & 0 & 0 \\ 0 & I_{n_2} & 0 & 0 & 0 \\ -L_1^\top L_1 & 0 & 2I_{n_1} & 0 & 0 \\ 0 & 0 & 0 & I_{n_2} & 0 \\ 0 & 0 & 0 & 0 & I_m \end{bmatrix}, \qquad M = \begin{bmatrix} 0_{n_1} & 0 & 0 & 0 & 0 \\ 0 & 0_{n_2} & 0 & 0 & 0 \\ L_1^\top L_1 & 0 & I_{n_1} & 0 & 0 \\ 0 & 0 & 0 & 0_{n_2} & 0 \\ L_1 & 0 & 0 & 0 & I_m \end{bmatrix},$$

then KM IFP²A (69) reduces to (83). It is straightforward to check that (82) does not generally determine an iteration sequence. Here we discuss the convergence of iteration (82) with assumption $L_1^\top L_1, L_2^\top L_2 \in \mathbb{S}_+$.

Theorem 3.23. *If $L_1^\top L_1 \in \mathbb{S}_+^{n_1}$ and $L_2^\top L_2 \in \mathbb{S}_+^{n_2}$, then for any $(x_0^1, x_0^2, y_0) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$, the sequence $\{(x_k^1, x_k^2, y_k)\}_{k\in\mathbb{N}}$ uniquely determined by ADMM* (82) *converges, with $\{(x_k^1, x_k^2)\}_{k\in\mathbb{N}}$ converges to a solution of model* (81).

Proof. It is straightforward to check the conditions of Theorem 3.10 holds under assumptions, which implies that $\{(x_k^1, x_k^2, y_k)\}_{k\in\mathbb{N}}$ is well-defined. Then notice that in this case

$$V = \begin{bmatrix} L_1^\top L_1 & 0 & -I_{n_1} & 0 & 0 \\ 0 & 0_{n_2} & 0 & 0 & 0 \\ 0 & 0 & 0_{n_1} & 0 & 0 \\ 0 & 0 & 0 & 0_{n_2} & 0 \\ L_1 & 0 & 0 & 0 & I_m \end{bmatrix}, \qquad W = 0.$$

Then notice that $\operatorname{Ran} \mathcal{F} \subseteq \mathcal{H} := \{(x^1, x^2, -L_1^\top y, -L_2^\top y, y) \in \mathbb{R}^{2n_1+2n_2+m} : x^1 \in \mathbb{R}^{n_1}, x^2 \in \mathbb{R}^{n_2}, y \in \mathbb{R}^m\}$ and for any $z \in \mathcal{H}$,

$$Vz = V'z, \qquad \text{where } V' := \begin{bmatrix} L_1^\top L_1 & 0 & 0 & 0 & L_1^\top \\ 0 & 0_{n_2} & 0 & 0 & 0 \\ 0 & 0 & 0_{n_1} & 0 & 0 \\ 0 & 0 & 0 & 0_{n_2} & 0 \\ L_1 & 0 & 0 & 0 & I_m \end{bmatrix}.$$

Then since for any $z = (x^1, x^2, -L_1^\top y, -L_2^\top y, y) \in \mathcal{H}$,

$$\left\langle z, V'z \right\rangle = \left\|L_1 x^1\right\|^2 + 2\left\langle L_1 x^1, u \right\rangle + \|u\|^2 = \left\|L_1 x^1 + u\right\|^2 \geq 0,$$

we have $V' \in \mathbb{S}(\mathcal{H})$. Also notice that

$$\begin{aligned} \operatorname{Ker} V' &= \left\{ \left(x^1, x^2, -L_1^\top y, -L_2^\top y, y\right) \in \mathcal{H} : L_1 x^1 + u = 0 \right\}, \\ \operatorname{Ker}(I - \Lambda) &= \left\{ \left(x^1, x^2, -L_1^\top y, -L_2^\top y, y\right) \in \mathcal{H} : L_1^\top L_1 x^1 + L_1^\top u = 0 \right\}, \\ \operatorname{Ker} M &= \left\{ \left(x^1, x^2, -L_1^\top y, -L_2^\top y, y\right) \in \mathcal{H} : L_1 x^1 + u = 0 \right\}, \end{aligned}$$

which proves that $\operatorname{Ker} V' = \operatorname{Ker}(I - \Lambda) \cap \operatorname{Ker} M$. Then an application of Theorem 3.17 within space $\mathcal{H}$ proves that $\{z_k\}_{k\in\mathbb{N}}$ converges to $z_* = (x_*^1, x_*^2, -L_1^\top y_*, -L_2^\top y_*, y_*) \in \mathcal{H}$ as a primal/dual solution pair of problem (81). □

ADMM is also known to have the *dual convergence*, *i.e.*, the convergence of $\{(L_1 x^1, L_2 x^2)\}_{k\in\mathbb{N}}$ towards the dual problem of (81). Here we review it under the framework of IFP²A. Notice that by substitutions $u_k := y_k$, $v_k^1 := -L_1 x_k^1$, $v_k^2 := -L_2 x_k^2$ and $v_k^3 := -v_k^1 - v_k^2$ for $k \in \mathbb{N}$, ADMM (83) leads us to

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} v_{k+1}^1 = \mathrm{prox}_{(f_1^* \circ (-L_1^\top))^*}\left(v_k^1 + 2u_{k+1} - u_k\right) \\ v_{k+1}^2 = \mathrm{prox}_{(f_2^* \circ (-L_2^\top))^*}\left(v_{k+1}^2 + u_{k+1}\right) \\ v_{k+1}^3 + v_{k+1}^1 - v_k^1 + u_k - u_{k+1} = \mathrm{prox}_{\iota_{\{c\}}}\left(v_{k+1}^3 + v_{k+1}^1 - v_k^1 + u_k\right) \end{array}\right. \end{array} \tag{84}$$

which is identical to KM IFP²A (69) with $R = I_{4m}$, $\lambda_k = 1$ for $k \in \mathbb{N}$ and

$$\Lambda = \begin{bmatrix} I_m & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 \\ 0 & 0 & I_m & 0 \\ -I_m & I_m & 0 & I_m \end{bmatrix}, \quad E = \begin{bmatrix} I_m & -I_m & -I_m & -I_m \\ 2I_m & 0_m & 0 & 0 \\ I_m & 0 & I_m & 0 \\ 0 & I_m & 0 & I_m \end{bmatrix}, \quad M = \begin{bmatrix} 0_m & 0 & 0 & 0 \\ -I_m & I_m & 0 & 0 \\ 0 & 0 & 0_m & 0 \\ I_m & -I_m & 0 & 0_m \end{bmatrix}.$$

It is straightforward to check that

$$V = \begin{bmatrix} I_m & -I_m & 0 & 0 \\ -I_m & I_m & 0 & 0 \\ 0 & 0 & 0_m & 0 \\ 0 & 0 & 0 & 0_m \end{bmatrix}, \qquad W = 0,$$

and $V \in \mathbb{S}^{4m}$, $\mathrm{Ker}\, V = \mathrm{Ker}\,(I - \Lambda) = \mathrm{Ker}\, M$. Therefore by Theorem 3.16 we directly have the following convergence theorem for the dual sequence of iteration (82). Notice that by (62), the dual problem of primal problem (81) could be expressed as

$$\operatorname*{argmin}_{u \in \mathbb{R}^m} \left\{ \left(f_1^* \circ \left(-L_1^\top\right)\right)(u) + \left(f_2^* \circ \left(-L_2^\top\right)\right)(u) + \iota_{\{c\}}^*(u) \right\}. \tag{85}$$

Theorem 3.24. *For any* $(x_0^1, x_0^2, y_0) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$*, the sequence* $\{(y_k, -L_1 x_k^1, -L_2 x_k^2, L_1 x_k^1 + L_2 x_k^2)\}_{k \in \mathbb{N}}$ *uniquely generated by ADMM* (82) *converges to a primal/dual solution of the dual problem* (85).

The special case of $L_2 = I$ was once discussed in [56, 92].

3.6.6 INEXACT UZAWA METHOD

This subsection identifies Inexact Uzawa (IU) method [11, 111] as a special case of IFP²A (68), which includes ADMM and the linearized ADMM as special cases. Then we propose a generalized convergence condition for inexact Uzawa method.

Consider the problem (81). IU designates the following explicit scheme for this problem.

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} x^2_{k+1} \leftarrow \operatorname{argmin}_{x^2\in\mathbb{R}^{n_2}}\left\{ f_2\left(x^2\right) + \frac{1}{2}\left\|L_1x^1_k + L_2x^2 + y_k - c\right\|^2 + \frac{1}{2}\left\|x^2 - x^2_k\right\|^2_{P_2} \right\} \\ y_{k+1} \leftarrow y_k + P_3^{-1}\left(L_1x^1_k + L_2x^2_{k+1} - c\right) \\ x^1_{k+1} \leftarrow \operatorname{argmin}_{x^1\in\mathbb{R}^{n_1}}\left\{ f_1\left(x^1\right) + \frac{1}{2}\left\|L_1x^1 + L_2x^2_{k+1} + y_{k+1} - c\right\|^2 + \frac{1}{2}\left\|x^1 - x^1_k\right\|^2_{P_1} \right\} \end{array}\right.\end{array} \tag{86}$$

where $P_1 \in \mathbb{S}^{n_1}$, $P_2 \in \mathbb{S}^{n_2}$ and $P_3 \in \mathbb{S}^m$. Notice that with $P_1 = P_2 = 0$ and $P_3 = I$, IU (86) reduces to ADMM (82); with $P_1 = \alpha I_{n_1} - L_1^\top L_1$, $P_2 = \beta I_{n_2} - L_2^\top L_2$ and $P_3 = I_m$, we have the so-called *linearized ADMM* [61, 106]. Other special cases of iteration (86) were discussed in existing literature, *e.g.*, the special case $L_2 = -I$ and $c = 0$ is considered in [111], and the case of scalar P_3 is proposed in [24] with convergence analysis.

Not surprisingly, IU is another special case of IFP²A (68). Notice that IU (86) is equivalent to

$$\begin{array}{l}\textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} x^1_{k+1} \leftarrow \operatorname{prox}_{f_1,Q_1}\left(x^1_k - Q_1^{-1}L_1^\top(y_{k+1} + P_3(y_{k+1} - y_k))\right) \\ x^2_{k+1} \leftarrow \operatorname{prox}_{f_2,Q_2}\left(Q_2^{-1}P_2x^2_k + Q_2^{-1}L_2^\top\left(L_2x^2_{k+1} - y_k - P_3(y_{k+1} - y_k)\right)\right) \\ y_{k+1} \leftarrow \operatorname{prox}_{\iota^*_{\{c\}},P_3}\left(y_k + P_3^{-1}\left(L_1x^1_k + L_2x^2_{k+1}\right)\right) \end{array}\right.\end{array} \tag{87}$$

where $Q_1 := P_1 + L_1^\top L_1 \in \mathbb{S}^{n_1}_+$ and $Q_2 := P_2 + L_2^\top L_2 \in \mathbb{S}^{n_2}_+$. Set $z_k := (x^1_k, x^2_k, , y_k)$ for $k \in \mathbb{N}$, then the IFP²A can be recast into

$$\Lambda z_{k+1} + (I - \Lambda)z_k = \operatorname{prox}_{\Phi,R}(Ez_{k+1} + Mz_k),$$

where $\Phi = f_1 \oplus f^2 \oplus \iota^*_{\{c\}}$, $R = Q_1 \otimes Q_2 \otimes P_3$ and

$$\Lambda = \begin{bmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & P_3 \end{bmatrix},$$

and

$$E = \begin{bmatrix} 0_{n_1} & 0 & -Q_1^{-1}L_1^\top(I + P_3) \\ 0 & Q_2^{-1}L_2^\top L_2 & -Q_2^{-1}L_2^\top P_3 \\ 0 & P_3^{-1}L_2 & P_3 - I_m \end{bmatrix}, \qquad M = \begin{bmatrix} I_{n_1} & 0 & Q_1^{-1}L_1^\top P_3 \\ 0 & Q_2^{-1}P_2 & Q_2^{-1}L_2^\top(P_3 - I) \\ P_3^{-1}L_1 & 0 & 2I_m - P_3 \end{bmatrix}.$$

Then notice that in this case

$$V = \begin{bmatrix} Q_1 & 0 & L_1^\top \\ 0 & P_2 & 0 \\ L_1 & 0 & P_3 \end{bmatrix}, \qquad U_1 = \begin{bmatrix} Q_1 & 0 & L_1^\top P_3 \\ 0 & P_2 & 0 \\ P_3 L_1 & 0 & 2P_3^2 - P_3 \end{bmatrix}.$$

Assume $P_1 \in \mathbb{S}^{n_1}$, $P_2 \in \mathbb{S}^{n_2}$, $P_3 - I \in \mathbb{S}^m$. Then it is straightforward to check that

$$\begin{aligned}
\operatorname{Ker} V &= \left\{ \left(x^1, x^2, y\right) \in \mathbb{R}^{n_1+n_2+m} \;:\; P_1 x^1 = P_2 x^2 = (P_3 - I)y = L_1 x^1 + y = 0 \right\}, \\
\operatorname{Ker} U_1 &= \left\{ \left(x^1, x^2, y\right) \in \mathbb{R}^{n_1+n_2+m} \;:\; P_1 x^1 = P_2 x^2 = \left(P_3^2 - P_3\right)y = L_1 x^1 + P_3 y = 0 \right\}, \\
\operatorname{Ker}(I - \Lambda) &= \left\{ \left(x^1, x^2, y\right) \in \mathbb{R}^{n_1+n_2+m} \;:\; (P_3 - I)y = 0 \right\}, \\
\operatorname{Ker} M &= \left\{ \left(x^1, x^2, y\right) \in \mathbb{R}^{n_1+n_2+m} \;:\; Q_1 x + L_1^\top P_3 y = P_2 x + L_2^\top (P_3 - I) y \right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. = L_1 x^1 + \left(2P_3^2 - P_3\right) y = 0 \right\},
\end{aligned}$$

therefore we have

$$\begin{aligned}
\operatorname{Ker} V = \operatorname{Ker} U_1 &= \left\{ \left(x^1, x^2, y\right) \in \mathbb{R}^{n_1+n_2+m} \;:\; P_1 x^1 = P_2 x^2 = (P_3 - I)y = L_1 x^1 + y = 0 \right\} \\
&\subseteq \operatorname{Ker}(I - \Lambda) \cap \operatorname{Ker} M.
\end{aligned}$$

Then a direct application of Theorem 3.16 proves the following general convergence theorem of IU.

Theorem 3.25. *Suppose $P_1 \in \mathbb{S}^{n_1}$, $P_2 \in \mathbb{S}^{n_2}$ and $P_3 - I \in \mathbb{S}^m$. If*

$$P_1 + L_1^\top L_1 \in \mathbb{S}_+^{n_1} \quad \textit{and} \quad P_2 + L_2^\top L_2 \in \mathbb{S}_+^{n_1},$$

then for any $(x_0^1, x_0^2, y_0) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$, the sequence $\{(x_k^1, x_k^2, y_k)\}_{k \in \mathbb{N}}$ generated by IU (86) converges to a primal/dual solution pair of model (81).

Theorem 3.25 has a direct corollary for linearized ADMM. Linearized ADMM explicitly reads

$$\begin{array}{l}
\textbf{For } k \in \mathbb{N} \\
\left\lfloor \begin{array}{l}
x_{k+1}^2 \leftarrow \operatorname{prox}_{f_2, \beta I_{n_2}} \left(x_k^2 - \beta^{-1} L_2^\top \left(L_1 x_k^1 + L_2 x_k^2 + y_k - c\right)\right) \\
y_{k+1} \leftarrow y_k + L_1 x_k^1 + L_2 x_{k+1}^2 - c \\
x_{k+1}^1 \leftarrow \operatorname{prox}_{f_1, \alpha I_{n_1}} \left(x_k^1 - \alpha^{-1} L_1^\top (2y_{k+1} - y_k)\right)
\end{array} \right.
\end{array} \tag{88}$$

Theorem 3.26. *Suppose $\alpha, \beta > 0$. If*

$$\alpha \geq \|L_1\|^2 \quad \textit{and} \quad \beta \geq \|L_2\|^2,$$

then for any $(x_0^1, x_0^2, y_0) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^m$, the sequence $\{(x_k^1, x_k^2, y_k)\}_{k \in \mathbb{N}}$ generated by linearized ADMM (88) *converges to a primal/dual solution pair of model* (81).

Proof. Put $P_1 = \alpha I_{n_1} - L_1^\top L_1$,$P_2 = \alpha I_{n_2} - L_2^\top L_2$ and $P_3 = I_m$ in Theorem 3.25 and the result directly follows. □

Convergence theorem with conditions $\alpha > \|L_1\|^2$ and $\beta > \|L_2\|^2$ for linearized ADMM was proposed in [61].

Table 2. Methods and convergence conditions for various cases of nonsmooth convex optimization problems

Model	Method	Iteration	Convergence Condition
$\operatorname*{argmin}_{x\in\mathbb{R}^n} h(x)$ where $h \in \Gamma_0(\mathbb{R}^n) \cap C_L^1(\mathbb{R}^n)$	Gradient Descent Method	**For** $k \in \mathbb{N}$ $x_{k+1} \leftarrow x_k - H^{-1}\nabla_h(x_k)$	$H - L/2 \in \mathbb{S}_+^n$
$\operatorname*{argmin}_{x\in\mathbb{R}^n} f(x)$ where $f \in \Gamma_0(\mathbb{R}^n)$	Proximal Point Method	**For** $k \in \mathbb{N}$ $x_{k+1} \leftarrow \operatorname{prox}_{f,P}(u_k)$ $u_{k+1} \leftarrow (1-\lambda_k)u_k + \lambda_k x_{k+1}$	$P \in \mathbb{S}_+^n$, $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,2]$, $\sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = \infty$
	Generalized Proximal Point Method	**For** $k \in \mathbb{N}$ $x_{k+1} \leftarrow (I-Q)x_k + Q\operatorname{prox}_f(x_k)$	$P, Q \in \mathbb{S}_+^n$, $2I - Q \in \mathbb{S}_+^n$
$\operatorname*{argmin}_{x\in\mathbb{R}^n} \{f(x) + g(x)\}$ where $f, g \in \Gamma_0(\mathbb{R}^n)$	Douglas-Rachford Splitting Algorithm	**For** $k \in \mathbb{N}$ $t_{k+1} \leftarrow \operatorname{prox}_f(2\operatorname{prox}_g(t_k) - t_k) + t_k - \operatorname{prox}_g(t_k)$	Always converges.
	Generalized Douglas-Rachford Splitting Algorithm	**For** $k \in \mathbb{N}$ $y_{k+1} \leftarrow \operatorname{prox}_{g^*,Q}(Q^{-1}u_k + v_k)$ $x_{k+1} \leftarrow \operatorname{prox}_{f,P}(P^{-1}(v_k - 2y_{k+1}) + u_k)$ $u_{k+1} \leftarrow (1-\lambda_k)u_k + \lambda_k x_{k+1}$ $v_{k+1} \leftarrow (1-\lambda_k)v_k + \lambda_k y_{k+1}$	$P, Q \in \mathbb{S}^n$, $\lVert(\sqrt{Q}\sqrt{P})^{-1}\rVert \leq 1$, $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,2]$, $\sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = \infty$
$\operatorname*{argmin}_{x\in\mathbb{R}^n} \{f(x) + g(Lx)\}$ where $f \in \Gamma_0(\mathbb{R}^n)$, $g \in \Gamma_0(\mathbb{R}^n)$, $L \in \mathbb{R}^{m\times n}$	First-Order Primal-Dual Algorithm	**For** $k \in \mathbb{N}$ $x_{k+1} \leftarrow \operatorname{prox}_{f,P}(u_k - P^{-1}L^\top v_k)$ $y_{k+1} \leftarrow \operatorname{prox}_{g^*,Q}(Q^{-1}L(2x_{k+1} - u_k) + v_k)$ $u_{k+1} \leftarrow (1-\lambda_k)u_k + \lambda_k x_{k+1}$ $v_{k+1} \leftarrow (1-\lambda_k)v_k + \lambda_k y_{k+1}$	$P \in \mathbb{S}_+^n$, $Q \in \mathbb{S}_+^m$, $\lVert\sqrt{Q}^{-1}L\sqrt{P}^{-1}\rVert \leq 1$, $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0,2]$, $\sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = \infty$

(Continued on next page)

Table 2. Methods and convergence conditions for various cases of nonsmooth convex optimization problems (Continued)

Model	Method	Iteration	Convergence Condition
$\operatorname*{argmin}_{\substack{x^1\in\mathbb{R}^{n_1},x^2\in\mathbb{R}^{n_2}\\ L_1x^1+L_2x^2=c}} \{f_1(x^1)+f_2(x^2)\}$ where $f_1\in\Gamma_0(\mathbb{R}^{n_1})$, $f_2\in\Gamma_0(\mathbb{R}^{n_2})$ $L_1\in\mathbb{R}^{m\times n_1}$, $L_2\in\mathbb{R}^{m\times n_2}$, $c\in\mathbb{R}^m$	Alternating Direction Method of Multipliers	**For** $k\in\mathbb{N}$ Find $x^2_{k+1}\in\operatorname{argmin}_{x^2\in\mathbb{R}^{n_2}}\left\{f_2(x^2)+\frac{1}{2}\\|L_1x^1_k+L_2x^2+y_k-c\\|^2\right\}$ $y_{k+1}\leftarrow y_k+L_1x^1_k+L_2x^2_{k+1}-c$ Find $x^1_{k+1}\in\operatorname{argmin}_{x^1\in\mathbb{R}^{n_1}}\left\{f_1(x^1)+\frac{1}{2}\\|L_1x^1+L_2x^2_{k+1}+y_{k+1}-c\\|^2\right\}$	(Primal) $L_1^\top L_1\in\mathbb{S}^{n_1}_+$, $L_2^\top L_2\in\mathbb{S}^{n_2}_+$ (Dual) Always converges.
	Linearized Alternating Direction Method of Multipliers	**For** $k\in\mathbb{N}$ $x^2_{k+1}\leftarrow\operatorname{prox}_{f_2,\beta I_{n_2}}\left(x^2_k-\beta^{-1}L_2^\top\times(L_1x^1_k+L_2x^2_k+y_k-c)\right)$ $y_{k+1}\leftarrow y_k+L_1x^1_k+L_2x^2_{k+1}-c$ $x^1_{k+1}\leftarrow\operatorname{prox}_{f_1,\alpha I_{n_1}}\left(x^1_k-\alpha^{-1}L_1^\top\times(2y_{k+1}-y_k)\right)$	$\alpha,\beta>0$, $\alpha\ge\\|L_1\\|^2$, $\beta\ge\\|L_2\\|^2$
	Inexact Uzawa Method	**For** $k\in\mathbb{N}$ $x^2_{k+1}\leftarrow\operatorname{argmin}_{x^2\in\mathbb{R}^{n_2}}\left\{f_2(x^2)+\frac{1}{2}\\|x^2-x^2_k\\|^2_{P_2}+\frac{1}{2}\\|L_1x^1_k+L_2x^2+y_k-c\\|^2\right\}$ $y_{k+1}\leftarrow y_k+P_3^{-1}(L_1x^1_k+L_2x^2_{k+1}-c)$ $x^1_{k+1}\leftarrow\operatorname{argmin}_{x^1\in\mathbb{R}^{n_1}}\left\{f_1(x^1)+\frac{1}{2}\\|x^1-x^1_k\\|^2_{P_1}+\frac{1}{2}\\|L_1x^1+L_2x^2_{k+1}+y_{k+1}-c\\|^2\right\}$	$P_1\in\mathbb{S}^{n_1}$, $P_2\in\mathbb{S}^{n_2}$, $P_3-I\in\mathbb{S}^m$, $P_1+L_1^\top L_1\in\mathbb{S}^{n_1}_+$, $P_2+L_2^\top L_2\in\mathbb{S}^{n_1}_+$
$\operatorname*{argmin}_{x\in\mathbb{R}^n}\{f_1(A_1x)+f_2(A_2x)\}$ where $A_1\in\mathbb{R}^{n_1\times n}$, $A_2\in\mathbb{R}^{n_2\times n}$, $f_1\in\Gamma_0(\mathbb{R}^{n_1})$, $f_2\in\Gamma_0(\mathbb{R}^{n_2})\cap C^1_L(\mathbb{R}^{n_2})$	Implicit Fixed-Point Proximity Algorithm	**For** $k\in\mathbb{N}$ $\Lambda z_{k+1}+(I-\Lambda)w_k=\operatorname{prox}_{\Phi,R}\left(Ez_{k+1}+Mw_k-R^{-1}A^\top\nabla_{f_2}(Aw_k)\right)$ $w_{k+1}\leftarrow(1-\lambda)w_k+\lambda z_{k+1}$	$P_0\in\mathbb{P}^{n+n_1}_+$, $V\in\mathbb{S}^{n+n_1}$, $\lambda\in D$, $\operatorname{Ker}V\cup\operatorname{Ker}U_\lambda\subseteq\operatorname{Ker}(I-\Lambda)\cap\operatorname{Ker}M\cap\operatorname{Ker}A$

CHAPTER 4

INEXACT IMPLICIT FIXED-POINT PROXIMITY ALGORITHMS FOR NONSMOOTH CONVEX OPTIMIZATION

In this chapter we propose the framework of Inexact Implicit Fixed-Point Proximity Algorithms (IIFP²A) by a combination of frameworks of inexact fixed-point iterations proposed in Chapter 2 and implicit fixed-point proximity algorithms proposed in Chapter 3. This framework systematically generalizes explicit FP²As discussed in Section 3.6. To deal with the fully implicit iterations of IFP²A, we again apply the matrix splitting technique to IFP²A, resulting in a class of inexact fixed-point proximity algorithms with inner loop, *i.e.*, IIFP²A. Then a careful choice of parameters ensures that IIFP²A falls into the intersection of the frameworks of inexact fixed-point iterations and implicit fixed-point proximity algorithms, which therefore establishes the framework of IIFP²A. As concrete applications of IIFP²A, we propose the classes of inexact block-separable fixed-point proximity algorithms and θ-inexact block-separable fixed-point proximity algorithms with convergence analysis.

In Chapter 3 we answer the first two questions of IFP²A, *i.e.*,

(Q1) The well-definedness of the IFP²A iteration.

(Q2) The convergence of the IFP²A.

by Theorems 3.10, 3.16 and 3.17, and cover plenty of existing explicit FP²As with generalizations. However, this framework of IFP²A fails when the iterations in (64) are implicit and cannot be exactly updated. In this chapter we answer the remaining two questions for IFP²A, *i.e.*,

(Q3) The feasible computation methods, which might be inexact, as IFP²A is implicit.

(Q4) The convergence of the inexact IFP²A.

Although the main model (44) has more applications in vast kinds of problems, it is far more difficult to propose a comprehensive framework of inexact IFP²A due to its complexity. To have a more concise and condensed analysis, we in this chapter mainly focus in the following simplified model

$$\operatorname*{argmin}_{x\in\mathbb{R}^n}\{f(x)+g(Lx)\}, \tag{89}$$

where $f\in\Gamma_0(\mathbb{R}^n)$, $g\in\Gamma_0(\mathbb{R}^m)$ and $L\in\mathbb{R}^{m\times n}$. Such framework could be easily extended to the full model (44) and the corresponding IFP²A. By Theorem 3.5, all the primal/dual pairs of solutions for (89) could be characterized as fixed points of

$$z_*=\operatorname{prox}_{\Phi,R}\big(\big(I+R^{-1}S\big)z_*\big), \tag{90}$$

where $\Phi=f\oplus g^*$, $R\in\mathbb{S}^{n+m}$ and $S\in\mathbb{R}^{(n+m)\times(n+m)}$ such that $S(x,y)=(-L^\top y,Lx)$ for $(x,y)\in\mathbb{R}^n\times\mathbb{R}^m$. Then by (67), KM IFP²A for (89) is read

$$\begin{array}{l}\textbf{For } k\in\mathbb{N}\\ \left\lfloor\begin{array}{l}\Lambda z_{k+1}+(I-\Lambda)w_k=\operatorname{prox}_{\Phi,R}(Ez_{k+1}+Mw_k)\\ w_{k+1}\leftarrow(1-\lambda_k)w_k+\lambda_k z_{k+1}\end{array}\right.\end{array} \tag{91}$$

Two common explicit FP²As, namely Douglas-Rachford splitting algorithm and first-order primal-dual algorithm, and their generalizations are discussed in Sections 3.6.3 and 3.6.4. These two explicit algorithms (77) and (80) are corresponding to (91) with parameters $R=P\otimes Q$ with $P\in\mathbb{S}^n_+$ and $Q\in\mathbb{S}^m_+$, $\Lambda=I_{n+m}$ and

$$E=\begin{bmatrix}0_n & -2P^{-1}L^\top\\ 0 & 0_m\end{bmatrix}\quad\text{or}\quad\begin{bmatrix}0_n & 0\\ -2Q^{-1}L & 0_m\end{bmatrix}.$$

Due to the strictly block-triangular structures of E, iterations (77) and (80) are indeed explicit algorithms. In this chapter we deal with the general choice of $E\in\mathbb{R}^{(n+m)\times(n+m)}$ with $R=P\otimes Q$ with $P\in\mathbb{S}^n_+$ and $Q\in\mathbb{S}^m_+$, $\Lambda=I_{n+m}$.

4.1 INEXACT BLOCK-SEPARABLE FIXED-POINT PROXIMITY ALGORITHM

This section proposes the Inexact IFP²A (IIFP²A) for (91) with general choice of E, in which case IFP²A (91) is generally implicit and cannot be exactly updated. To deal with the implicitness in iteration, we adopt the matrix splitting technique again to have an explicit but inexact scheme. As applications of IIFP²A, we propose the Inexact Block-Separable FP²As (IBSFP²As) and θ-IBSFP²A.

Recall that to handle the expanding property of $I+R^{-1}S$ in the derivation of IFP²A, we adopt the *matrix splitting technique*, which then provides us IFP²A with novel analytic properties. Here the same idea is applied again to handle the implicitness of IFP²A with general E. Specifically, we split $E = A + B$ with $A, B \in \mathbb{R}^{(n+m)\times(n+m)}$ and introduce the inner loop

$$\begin{array}{l} \textbf{For } l \in \mathbb{N} \\ \left\lfloor\, z_k^{l+1} = \operatorname{prox}_{\Phi,R}\left(Az_k^{l+1} + Bz_k^l + Mw_k\right) \right. \end{array} \tag{92}$$

to proximate the unknown evaluation of $\mathcal{F}w_k$, which leads us to the Inexact IFP²A (IIFP²A)

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} \textbf{Do } l \in \mathbb{N} \\ \quad \left| \; z_k^{l+1} = \operatorname{prox}_{\Phi,R}\left(Az_k^{l+1} + Bz_k^l + Mw_k\right) \right. \\ \textbf{While } \|z_k^{l+1} - z_k^l\| > \delta_k \\ \tilde{z}_{k+1} \leftarrow z_k^{l+1} \\ \tilde{w}_{k+1} \leftarrow (1-\lambda_k)\tilde{w}_k + \lambda_k \tilde{z}_k^{l+1} \end{array} \right. \end{array} \tag{93}$$

where $\{z_k^0\}_{k\in\mathbb{N}} \subset \mathbb{R}^{n+m}$ are initial inputs for inner loops, $\{\lambda_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ and $\{\delta_k\}_{k\in\mathbb{N}} \subset (0, +\infty)$ controls the successive errors in inner loops. To obtain an explicit iterative method, here A is chosen to be strictly block-triangular, *i.e.*,

$$A = \begin{bmatrix} 0_n & 0 \\ A_{21} & 0_m \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0_n & A_{12} \\ 0 & 0_m \end{bmatrix},$$

where $A_{21} \in \mathbb{R}^{m\times n}$ and $A_{12} \in \mathbb{R}^{n\times m}$. This results in the following two Inexact Block-Separable Fixed-Point Proximity Algorithms (IBSFP²As), as

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} \textbf{Do } l \in \mathbb{N} \\ \quad \left| \begin{array}{l} x_k^{l+1} \leftarrow \operatorname{prox}_{f,P}((I - M_{11})x_k^l - (P^{-1}L^* + M_{12})y_k^l \\ \qquad\qquad\qquad + M_{11}\tilde{u}_k + M_{12}\tilde{v}_k) \\ y_k^{l+1} \leftarrow \operatorname{prox}_{g^*,Q}(A_{21}(x_k^{l+1} - x_k^l) + (Q^{-1}L - M_{21})x_k^l \\ \qquad\qquad + (I - M_{22})y_k^l + M_{21}\tilde{u}_k + M_{22}\tilde{v}_k) \end{array} \right. \\ \textbf{While } \|x_k^{l+1} - x_k^l\|^2 + \|y_k^{l+1} - y_k^l\|^2 > \delta_k^2 \\ \tilde{x}_{k+1} \leftarrow x_k^{l+1},\ \tilde{y}_{k+1} \leftarrow y_k^{l+1} \\ \tilde{u}_{k+1} \leftarrow \lambda\tilde{u}_k + (1-\lambda)\tilde{x}_{k+1},\ \tilde{v}_{k+1} \leftarrow \lambda_k\tilde{v}_k + (1-\lambda_k)\tilde{y}_{k+1} \end{array} \right. \end{array} \tag{94}$$

and

$$
\begin{array}{l}
\textbf{For } k \in \mathbb{N} \\
\left| \begin{array}{l}
\textbf{Do } l \in \mathbb{N} \\
\left| \begin{array}{l}
y_k^{l+1} \leftarrow \mathrm{prox}_{g^*,Q}((Q^{-1}L - M_{21})x_k^l + (I - M_{22})y_k^l \\
\qquad\qquad\qquad\qquad\qquad\qquad + M_{21}\tilde{u}_k + M_{22}\tilde{v}_k) \\
x_k^{l+1} \leftarrow \mathrm{prox}_{f,P}((I - M_{11})x_k^l + A_{12}(y_k^{l+1} - y_k^l) \\
\qquad\qquad\qquad -(P^{-1}L^* + M_{12})\tilde{y}_k^l + M_{11}u_k + M_{12}\tilde{v}_k)
\end{array} \right. \\
\textbf{While } \|x_k^{l+1} - x_k^l\|^2 + \|y_k^{l+1} - y_k^l\|^2 > \delta_k^2 \\
\tilde{x}_{k+1} \leftarrow x_k^{l+1},\ \tilde{y}_{k+1} \leftarrow y_k^{l+1} \\
\tilde{u}_{k+1} \leftarrow \lambda\tilde{u}_k + (1-\lambda)\tilde{x}_{k+1},\ \tilde{v}_{k+1} \leftarrow \lambda_k\tilde{v}_k + (1-\lambda_k)\tilde{y}_{k+1}
\end{array} \right.
\end{array}
\tag{95}
$$

where $\{(x_k^0, y_k^0)\}_{k\in\mathbb{N}} \subset \mathbb{R}^{n+m}$ are initial inputs for each inner loop, and $\{\delta_k\}_{k\in\mathbb{N}} \subset (0, +\infty)$ are the controller of the successive error in inner loop.

IBSFP²As (94) and (95) serve as feasible computation methods for IFP²A and therefore answer the question (Q3). For concrete applications, we propose the following θ-IBSFP²A as a special cases of IBSFP²A. Let

$$
A = \begin{bmatrix} 0_n & -(1+\theta)P^{-1}L^\top \\ 0 & 0_m \end{bmatrix}, \quad B = \begin{bmatrix} 0_n & 0 \\ (1-\theta)Q^{-1}L & 0_m \end{bmatrix}, \quad M = \begin{bmatrix} I_n & \theta P^{-1}L^\top \\ \theta Q^{-1}L & I_m \end{bmatrix}, \tag{96}
$$

then IBSFP²A (95) reduces to

$$
\begin{array}{l}
\textbf{For } k \in \mathbb{N} \\
\left| \begin{array}{l}
\textbf{Do } l \in \mathbb{N} \\
\left| \begin{array}{l}
y_k^{l+1} \leftarrow \mathrm{prox}_{g^*,Q}\big(Q^{-1}L\big((1-\theta)x_k^l + \theta\tilde{u}_k\big) + \tilde{v}_k\big) \\
x_k^{l+1} \leftarrow \mathrm{prox}_{f,P}\big(P^{-1}L^\top\big(\theta\tilde{v}_k - (1+\theta)y_k^{l+1}\big) + \tilde{u}_k\big)
\end{array} \right. \\
\textbf{While } \|x_k^{l+1} - x_k^l\|^2 + \|y_k^{l+1} - y_k^l\|^2 > \delta_k^2 \\
\tilde{x}_{k+1} \leftarrow x_k^{l+1},\ \tilde{y}_{k+1} \leftarrow y_k^{l+1} \\
\tilde{u}_{k+1} \leftarrow \lambda_k\tilde{u}_k + (1-\lambda_k)\tilde{x}_{k+1},\ \tilde{v}_{k+1} \leftarrow \lambda_k\tilde{v}_k + (1-\lambda_k)\tilde{y}_{k+1}
\end{array} \right.
\end{array}
\tag{97}
$$

Also notice that θ-IBSFP²A (97) is identical to

$$
\begin{array}{l}
\textbf{For } k \in \mathbb{N} \\
\left| \begin{array}{l}
\textbf{Do } l \in \mathbb{N} \\
\left| \begin{array}{l}
x_k^{l+1} \leftarrow \mathrm{prox}_{f,P}\big(P^{-1}L^\top\big(\theta\tilde{v}_k - (1+\theta)y_k^l\big) + \tilde{u}_k\big) \\
y_k^{l+1} \leftarrow \mathrm{prox}_{g^*,Q}\big(Q^{-1}L\big((1-\theta)x_k^{l+1} + \theta\tilde{u}_k\big) + \tilde{v}_k\big)
\end{array} \right. \\
\textbf{While } \|x_k^{l+1} - x_k^l\|^2 + \|y_k^{l+1} - y_k^l\|^2 > \delta_k^2 \\
\tilde{x}_{k+1} \leftarrow x_k^{l+1},\ \tilde{y}_{k+1} \leftarrow y_k^{l+1} \\
\tilde{u}_{k+1} \leftarrow \lambda\tilde{u}_k + (1-\lambda)\tilde{x}_{k+1},\ \tilde{v}_{k+1} \leftarrow \lambda\tilde{v}_k + (1-\lambda)\tilde{y}_{k+1}
\end{array} \right.
\end{array}
\tag{98}
$$

which corresponds to the parameter set

$$A' = \begin{bmatrix} 0_n & 0 \\ (1-\theta)Q^{-1}L & 0_m \end{bmatrix}, \quad B' = \begin{bmatrix} 0_n & -(1+\theta)P^{-1}L^\top \\ 0 & 0_m \end{bmatrix}, \quad M' = \begin{bmatrix} I_n & \theta P^{-1}L^\top \\ \theta Q^{-1}L & I_m \end{bmatrix}. \tag{99}$$

The correspondence between (97) and (98) will contribute to the final convergence theorem of θ-IBSFP²A.

In this section we apply matrix splitting technique to IFP²A when the updates are implicit, which results in the IIFP²A (91). To have explicit schemes, IBSFP²As (94) and (95), as special cases of IIFP²A, are proposed as answers to (Q3). As a concrete application of IIFP²A, we propose the θ-IBSFP²A (97). In the following sections, we consider the framework of convergence analysis of IIFP²A.

4.2 INNER CONVERGENCE ANALYSIS

In this section we consider the convergence of the inner loop (92) embedded in IIFP²A. Corresponding corollary for the inner loop of θ-IBSFP²A (97) is proposed.

IIFP²A (93) provides a class of algorithms serve as answers to (Q3). However at the same time IIFP²A raises two other questions, *i.e.*,

(Q3.1) The convergence of the inner loop.

(Q3.2) Algorithm is *executable*, *i.e.*, algorithm quits inner loop in finite steps.

In the following theorem we answer (Q3.1) and (Q3.2) via the convergence theorem and the *a posteriori* error estimation for inner loop of IIFP²A (92). Recall that in this case $P_0 = \Gamma R(I-E) = \Gamma(RM - S)$, and define $U := \Gamma R(I-A) = \Gamma RB + P_0$. For $A \in \mathbb{S}^n_+$, denote $A^{-1/2} := \sqrt{A}^{-1}$ for clearness.

Theorem 4.1. *Suppose that* $R \in \mathbb{S}^{n+m}_+$, $P_0, U \in \mathbb{P}^{n+m}_+$ *and* $\{\delta_k\}_{k\in\mathbb{N}} \subset (0, +\infty)$. *If*

$$\left\| \overline{U}^{-1/2} \Gamma R B \overline{U}^{-1/2} \right\| < 1, \tag{100}$$

then for any $(\tilde{u}_0, \tilde{v}_0) \in \mathbb{R}^{n+m}$ *and* $\{(x^0_k, y^0_k)\}_{k\in\mathbb{N}} \subset \mathbb{R}^{n+m}$, *the IIFP²A (93) is executable. Moreover, there exists* $c > 0$ *such that*

$$\|\tilde{z}_{k+1} - \mathcal{F}\tilde{w}_k\| \le c \cdot \delta_k, \qquad \textit{for all } k \in \mathbb{N}.$$

Proof. Denote $\tilde{w}_k := (\tilde{u}_k, \tilde{v}_k)$, $z^l_k := (x^l_k, y^l_k)$ and $\tilde{z}_k := (\tilde{x}_k, \tilde{y}_k)$ for $k, l \in \mathbb{N}$. By Theorem 3.10, $\mathcal{F}\tilde{w}_k$

uniquely exists for any $\tilde{w}_k \in \mathbb{R}^{n+m}$, which satisfies

$$-P_0\mathcal{F}\tilde{w}_k + \Gamma RM\tilde{w}_k \in \partial_{\Phi_\Gamma}(\mathcal{F}\tilde{w}_k).$$

On the other hand, by Theorem 3.10 and (92) we have that z_k^{l+1} exists and uniquely satisfies

$$\Gamma R(A-I)z_k^{l+1} + \Gamma RBz_k^l + \Gamma RM\tilde{w}_k \in \partial_{\Phi_\Gamma}\left(z_k^{l+1}\right).$$

Then, by monotonicity of subdifferential, we have

$$\begin{aligned} 0 &\le \left\langle z_k^{l+1} - \mathcal{F}\tilde{w}_k, \Gamma R(A-I)\left(z_k^{l+1} - \mathcal{F}\tilde{w}_k\right) + \Gamma RB\left(z_k^l - \mathcal{F}\tilde{w}_k\right)\right\rangle \\ &= -\left\|z_k^{l+1} - \mathcal{F}\tilde{w}_k\right\|_{\overline{U}}^2 + \left\langle z_k^{l+1} - \mathcal{F}\tilde{w}_k, \Gamma RB\left(z_k^l - \mathcal{F}\tilde{w}_k\right)\right\rangle \\ &\le -\left\|z_k^{l+1} - \mathcal{F}\tilde{w}_k\right\|_{\overline{U}}^2 + \left\|\overline{U}^{-1/2}\Gamma RB\overline{U}^{-1/2}\right\|\left\|z_k^{l+1} - \mathcal{F}\tilde{w}_k\right\|_{\overline{U}}\left\|z_k^l - \mathcal{F}\tilde{w}_k\right\|_{\overline{U}}. \end{aligned}$$

Thus, if $\rho := \|\overline{U}^{-1/2}\Gamma RB\overline{U}^{-1/2}\| < 1$, then we would have $\|z_k^{l+1} - \mathcal{F}\tilde{w}_k\|_{\overline{U}} \le \rho\|z_k^l - \mathcal{F}\tilde{w}_k\|_{\overline{U}}$, which proves the linear convergence of the sequence $\{z_k^l\}_{l\in\mathbb{N}}$ towards $\mathcal{F}\tilde{w}_k$. Then by $\|z_k^{l+1} - z_k^l\|_{\overline{U}} \le (1+\rho)\|z_k^l - \mathcal{F}\tilde{w}_k\|_{\overline{U}}$, IBSFP²As quit inner loop in finite steps. Finally, again by the monotonicity of subdifferential, we have

$$\begin{aligned} 0 &\le \left\langle z_k^{l+1} - \mathcal{F}\tilde{w}_k, -P_0\left(z_k^{l+1} - \mathcal{F}\tilde{w}_k\right) + \Gamma RB\left(z_k^{l+1} - z_k^l\right)\right\rangle \\ &\le -\left\|z_k^{l+1} - \mathcal{F}\tilde{w}_k\right\|_{\overline{P_0}}^2 + \left\|\overline{P_0}^{-1/2}\Gamma RB\overline{P_0}^{-1/2}\right\|\left\|z_k^{l+1} - \mathcal{F}\tilde{w}_k\right\|_{\overline{P_0}}\left\|z_k^{l+1} - z_k^l\right\|_{\overline{P_0}}, \end{aligned}$$

which along with the quit condition of inner loop, proves the declared *a posteriori* error estimation. □

Theorem 4.1 ensures the IBSFP²As are executable and have their inner loop converge. Here we apply Theorem 4.1 to θ-IBSFP²A. To this end we need the following lemmas.

Lemma 4.2. *Suppose $R \in \mathbb{S}_+^{n+m}$ and $P_0 \in \mathbb{P}_+^{n+m}$. If*

$$2\|\Gamma RB\|\left\|\overline{P_0}^{-1}\right\| < 1,$$

then $U \in \mathbb{P}_+^{n+m}$ and $\|\overline{U}^{-1/2}\Gamma RB\overline{U}^{-1/2}\| < 1$.

Proof. Set $V := \overline{P_0}$ and $\Delta := \Gamma RB$, then we have $\overline{U} = V + \overline{\Delta}$ and by condition, $\|\Delta\|\|V^{-1}\| < 1/2$. Notice that $\|\overline{\Delta}\| = \|\Delta + \Delta^\mathsf{T}\|/2 \le \|\Delta\|$, we have $\|V^{-1/2}\overline{\Delta}V^{-1/2}\| \le \|\Delta\|\|V^{-1}\| < 1/2$, which means

$U \in \mathbb{P}_+^{n+m}$. Furthermore, since $\|V^{-1}\bar{\Delta}\| < \|V^{-1}\|\|\bar{\Delta}\| < 1$, we have

$$\|\bar{U}^{-1}\| - \|V^{-1}\| \leq \|(V+\bar{\Delta})^{-1} - V^{-1}\| \leq \|V^{-1}\| \left\| I - \left(I + V^{-1}\bar{\Delta}\right)^{-1} \right\|.$$

By Neumann series $(I + V^{-1}\bar{\Delta})^{-1} = \sum_{n=0}^{\infty}(-V^{-1}\bar{\Delta})^n$, we have

$$\begin{aligned} \left\| I - \left(I + V^{-1}\bar{\Delta}\right)^{-1} \right\| &= \left\| I - \sum_{n=0}^{\infty}\left(-V^{-1}\bar{\Delta}\right)^n \right\| \\ &\leq \sum_{n=1}^{\infty} \|V^{-1}\bar{\Delta}\|^n = \frac{\|V^{-1}\bar{\Delta}\|}{1 - \|V^{-1}\bar{\Delta}\|} \\ &\leq \frac{\|V^{-1}\|\|\Delta\|}{1 - \|V^{-1}\|\|\Delta\|}, \end{aligned}$$

which implies

$$\|\bar{U}^{-1/2}\Gamma RB\bar{U}^{-1/2}\| \leq \|\bar{U}^{-1}\|\|\Delta\| \leq \|V^{-1}\|\|\Delta\| + \frac{\|V^{-1}\|^2\|\Delta\|^2}{1 - \|V^{-1}\|\|\Delta\|} = \frac{\|V^{-1}\|\|\Delta\|}{1 - \|V^{-1}\|\|\Delta\|}.$$

The last inequality shows that $\|V^{-1}\|\|\Delta\| < 1/2$ is sufficient for $\|\bar{U}^{-1/2}\Gamma RB\bar{U}^{-1/2}\| < 1$. □

Lemma 4.3. *Suppose $P \in \mathbb{S}_+^n$, $Q \in \mathbb{S}_+^m$, $L \in \mathbb{R}^{n\times m}$, and set*

$$A := \begin{bmatrix} P & L \\ L^\top & Q \end{bmatrix}.$$

Then the following statements hold.

- *(i) $A \in \mathbb{S}^{n+m}$ if and only if $\|P^{-1/2}LQ^{-1/2}\| \leq 1$.*
- *(ii) $A \in \mathbb{S}_+^{n+m}$ if and only if $\|P^{-1/2}LQ^{-1/2}\| < 1$.*
- *(iii) If $A \in \mathbb{S}_+^{n+m}$, then $\|A^{-1}\|^{-1} \geq (1 - \|P^{-1/2}LQ^{-1/2}\|)/\max\{\|P^{-1}\|, \|Q^{-1}\|\}$.*

Proof. Suppose $\|P^{-1/2}LQ^{-1/2}\| \leq 1$. Then for any $(x, y) \in \mathbb{R}^{n+m}$, if define $\tilde{x} := P^{1/2}x$ and $\tilde{y} := Q^{1/2}y$, then we have

$$\begin{aligned} \begin{bmatrix} x^* & y^* \end{bmatrix} \begin{bmatrix} P & L \\ L^* & Q \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= \|\tilde{x}\|^2 + \|\tilde{y}\|^2 + 2\left\langle \tilde{x}, P^{-1/2}LQ^{-1/2}\tilde{y} \right\rangle \\ &\geq \|\tilde{x}\|^2 + \|\tilde{y}\|^2 - 2\left\|P^{-1/2}LQ^{-1/2}\right\|\|\tilde{x}\|\|\tilde{y}\| \\ &\geq (\|\tilde{x}\| - \|\tilde{y}\|)^2 \geq 0, \end{aligned}$$

which proves $A \in \mathbb{S}^{n+m}$. Conversely, if $\|P^{-1/2}LQ^{-1/2}\| > 1$, then there exists $y \in \mathbb{R}^m$ such that $\|P^{-1/2}Ly\| > \|Q^{1/2}y\|$. Set $x = -P^{-1}Ly$ then we have

$$\begin{bmatrix} x^* & y^* \end{bmatrix} \begin{bmatrix} P & L \\ L^* & Q \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \left\|Q^{1/2}y\right\|^2 - \left\|P^{-1/2}Ly\right\|^2 < 0,$$

which means $A \notin \mathbb{S}^{n+m}$. The rest could be proved by analogous argument, or see [56]. □

Combine Lemmas 4.2 and 4.3 and Theorem 4.1, then we have the following convergence theorem for the inner loop of θ-IBSFP²A.

Theorem 4.4. *Suppose $P \in \mathbb{S}^n_+$, $Q \in \mathbb{S}^m_+$ and $\theta \in \mathbb{R}$. If $\theta \neq 0$ and*

$$2 \min\{|1-\theta|, |1+\theta|\} \cdot \max\left\{ \left\|P^{-1}\right\|, \left\|Q^{-1}\right\| \right\} \|L\| \leq 1 - |\theta| \|Q^{-1/2}LP^{-1/2}\|, \tag{101}$$

or, $\theta = 0$ and

$$2\max\left\{ \left\|P^{-1}\right\|, \left\|Q^{-1}\right\| \right\} \|L\| < 1,$$

then for any $k \in \mathbb{N}$ and $(\tilde{u}_k, \tilde{v}_k), (x_k^0, y_k^0) \in \mathbb{R}^{n+m}$, sequence $\{(x_k^l, y_k^l)\}_{l\in\mathbb{N}}$ generated by the inner loop of θ-IBSFP²A (97) converges to $\mathcal{F}(\tilde{u}_k, \tilde{v}_k)$. Moreover, there exists $c > 0$ such that

$$\|(\tilde{x}_{k+1}, \tilde{y}_{k+1}) - \mathcal{F}(\tilde{u}_k, \tilde{v}_k)\| \leq c \cdot \delta_k, \qquad \textit{for all } k \in \mathbb{N}.$$

Proof. If $|\theta| \|Q^{-1/2}LP^{-1/2}\| = 1$, then (101) implies $|\theta| = 1$. This reduces the inner loop of θ-IBSFP²A to an explicit update, which converges and quits immediately after the first inner iteration. Same argument proves the case of $L = 0$. Then we assume $L \neq 0$ and $|\theta| \|Q^{-1/2}LP^{-1/2}\| < 1$.

If $|\theta| \|Q^{-1/2}LP^{-1/2}\| < 1$, then Lemma 4.3 implies $RM \in \mathbb{S}^{n+m}_+$ and therefore, for $\Gamma = \alpha I_n \otimes \beta I_m$ with $\alpha, \beta > 0$ closed enough to each other, we have $P_0 \in \mathbb{P}^{n+m}_+$ and, again by Lemma 4.3, $\|\overline{P_0}^{-1}\|^{-1} \geq (1 - |\alpha(1+\theta) + \beta(\theta - 1)|/(2\sqrt{\alpha\beta}) \|Q^{-1/2}LP^{-1/2}\|)/\max\{\|P^{-1}\|/\alpha, \|Q^{-1}\|/\beta\}$ and $\|\Gamma RB\| \leq \beta |1-\theta| \|L\|$. Therefore we have

$$\|\Gamma RB\| \|\overline{P_0}^{-1}\| \leq \frac{2|1-\theta| \|L\| \max\left\{ \kappa^2 \|P^{-1}\|, \|Q^{-1}\| \right\}}{2 - |\kappa^{-1}(1+\theta) + \kappa(\theta - 1)| \|Q^{-1/2}LP^{-1/2}\|},$$

where $\kappa := \sqrt{\beta/\alpha}$. Define the right-hand side of the above as $h(\kappa)$, and by Theorem 4.1, it is sufficient to show that for any $\rho > 0$, there always exists $\kappa \in (1-\rho, 1+\rho)$ such that $h(\kappa) < 1/2$.

If $\theta = 0$, then by conditions, $h(1) < 1/2$, which finishes the proof. Assume $\theta \neq 0$. First by (101), we assume that

$$2|1-\theta|\max\left\{ \|P^{-1}\|, \|Q^{-1}\| \right\}\|L\| \leq 1 - |\theta|\|Q^{-1/2}LP^{-1/2}\|$$

holds. Then we have $h(1) \leq 1/2$. If $h(1) < 1/2$ then the proof is finished with setting $\alpha = \beta$. Suppose $h(1) = 1/2$ and $\kappa = 1 + \varepsilon$. Then observe that we have

$$\begin{aligned}\left|\kappa^{-1}(1+\theta) + \kappa(\theta-1)\right| &= \left|(1+\varepsilon)^{-1}(1+\theta) + (1+\varepsilon)(\theta-1)\right| \\ &= \left|\left(1 - \varepsilon + O\left(\varepsilon^2\right)\right)(1+\theta) + (1+\varepsilon)(\theta-1)\right| \\ &= 2|\theta - \varepsilon| + O\left(\varepsilon^2\right),\end{aligned}$$

which means, along with the assumption $|\theta| \neq 0$, for sufficiently small $|\varepsilon|$ there holds

$$\left|\kappa^{-1}(1+\theta) + \kappa(\theta-1)\right| = O\left(\varepsilon^2\right) + \begin{cases} 2|\theta| + 2\varepsilon, & \theta < 0, \\ 2|\theta| - 2\varepsilon, & \theta > 0. \end{cases}$$

Then we discuss in cases $\|P^{-1}\| > \|Q^{-1}\|$, $\|P^{-1}\| < \|Q^{-1}\|$ and $\|P^{-1}\| = \|Q^{-1}\|$ and $\theta > 0$, $\theta < 0$. Denote $C := \|Q^{-1/2}LP^{-1/2}\| > 0$ for abbreviation.

Suppose $\|P^{-1}\| > \|Q^{-1}\|$ and $\theta < 0$. Then for sufficiently small $|\varepsilon|$,

$$\begin{aligned}h(\kappa) &= \frac{2\kappa^2|1-\theta|\|L\|\|P^{-1}\|}{2 - |\kappa^{-1}(1+\theta) + \kappa(\theta-1)|C} \\ &= h(1) + \frac{2 - 2|\theta|C + C}{(1-|\theta|C)^2}|1-\theta|\|L\|\|P^{-1}\|\varepsilon + O\left(\varepsilon^2\right),\end{aligned}$$

which, along with $|\theta|C < 1$ and $|\theta| \neq 1$, proves that h has a positive slope at $\kappa = 1$. This proves the existence of $\rho > 0$ such that $h(\kappa) < 1/2$ for all $\kappa \in (1-\rho, 1)$ when $\|P^{-1}\| > \|Q^{-1}\|$ and $\theta < 0$.

Suppose $\|P^{-1}\| < \|Q^{-1}\|$. Then for sufficiently small $|\varepsilon|$,

$$\begin{aligned}h(\kappa) &= \frac{2|1-\theta|\|L\|\|Q^{-1}\|}{2 - |\kappa^{-1}(1+\theta) + \kappa(\theta-1)|C} \\ &= h(1) + O\left(\varepsilon^2\right) + \frac{C}{(1-|\theta|C)^2}|1-\theta|\|L\|\|Q^{-1}\|\varepsilon \times \begin{cases} 1, & \theta < 0, \\ -1, & \theta > 0, \end{cases}\end{aligned}$$

which shows that h has a nonzero slope at $\kappa = 1$. This proves when $\|P^{-1}\| < \|Q^{-1}\|$, for any $\rho > 0$, if $\theta < 0$ then there is always $\kappa \in (1-\rho, 1)$ such that $h(\kappa) < 1/2$; if $\theta > 0$ then there is always

$\kappa \in (1, 1+\rho)$ such that $h(\kappa) < 1/2$.

So far we have proven that if $\|P^{-1}\| > \|Q^{-1}\|$ then there exists $\rho > 0$ such that $h(\kappa) < 1/2$ for all $\kappa \in (1-\rho, 1)$, and if $\|P^{-1}\| < \|Q^{-1}\|$, then for any $\rho > 0$ there is always $\kappa \in (1-\rho, 1+\rho)$ such that $h(\kappa) < 1/2$. For the missing piece of the proof, consider the alternative parameter set (99). Under this set of parameters, we have $\overline{P_0'} = \overline{P_0}$ and $\|\Gamma RB'\| \leq \alpha|1+\theta|\|L\|$. Therefore we have

$$\|\Gamma RB'\|\|\overline{P_0}^{-1}\| \leq \frac{2|1+\theta|\|L\|\max\left\{\|P^{-1}\|, \kappa^{-2}\|Q^{-1}\|\right\}}{2 - |\kappa^{-1}(1+\theta) + \kappa(\theta-1)|\|Q^{-1/2}LP^{-1/2}\|}.$$

Define the right-hand side function as $h'(\kappa)$. Then, if $\|P^{-1}\| > \|Q^{-1}\|$ we would have, for sufficiently small $|\varepsilon|$,

$$\begin{aligned} h'(\kappa) &= \frac{2|1-\theta|\|L\|\|Q^{-1}\|}{2 - |\kappa^{-1}(1+\theta) + \kappa(\theta-1)|C} \\ &= h(1) + O\left(\varepsilon^2\right) + \frac{C}{(1-|\theta|C)^2}|1+\theta|\|L\|\|P^{-1}\|\varepsilon \times \begin{cases} 1, & \theta < 0, \\ -1, & \theta > 0, \end{cases} \end{aligned}$$

which proves that if $\|P^{-1}\| > \|Q^{-1}\|$ then for any $\rho > 0$, if $\theta < 0$ then there is always $\kappa \in (1-\rho, 1)$ such that $h'(\kappa) < 1/2$; if $\theta > 0$ then there is always $\kappa \in (1, 1+\rho)$ such that $h'(\kappa) < 1/2$.

Finally, consider the case $\|P^{-1}\| = \|Q^{-1}\|$. If $\theta > 0$, then we consider $\kappa > 1$ and therefore for any $\rho > 0$ there are always $\kappa \in (1, 1+\rho)$ such that $h'(\kappa) < 1/2$. If $\theta < 0$, then we consider $\kappa < 1$ and therefore there is always $\kappa \in (1-\rho, 1)$ such that $h(\kappa) < 1/2$. This means there is always $\alpha, \beta > 0$ such that

$$\min\left\{\|\Gamma RB\|\|\overline{P_0}\|, \|\Gamma RB'\|\|\overline{P_0'}\|\right\} < \frac{1}{2},$$

where a direct application of Theorem 4.1 and Lemma 4.2 then draws the proof to a close. □

In this section we study the condition for the inner loop of IIFP²A (91) to converge, and propose an *a posteriori* error estimation of the inner loop in Theorem 4.1. Then we apply Theorem 4.1 to the θ-IBSFP²A (97) to have an explicit condition for θ-IBSFP²A to have its inner loop converges and its inner error bounded. In the following section we are about to combine the inner convergence of IIFP²A, the framework of inexact fixed-point iterations in Chapter 2, and framework of IFP²A in Chapter 3, to propose the convergence analysis of IIFP²A.

4.3 CONVERGENCE ANALYSIS OF IIFP²A

In this section we combine the framework of inexact fixed-point iterations in Chapter 2, the framework of IFP²A in Chapter 3, and the convergence analysis of the inner loop of IIFP²A (93) in the preceding section, to propose the general convergence analysis framework of IIFP²A. As a direct application, we apply the proposed framework of IIFP²A to the θ-IBSFPPA (97).

After all the complicated and sophisticated frameworks established in previous chapters and sections, the following convergence theorem for IIFP²A is quite straightforward. Recall that in this case, $V = RM$, $P_0 = \Gamma R(I - E) = \Gamma(RM - S)$ and $U = \Gamma R(I - A) = \Gamma RB + P_0$.

Theorem 4.5. *Suppose $R, V \in \mathbb{S}_+^{n+m}$, $P_0, U \in \mathbb{P}_+^{n+m}$, $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 2]$ and $\{\delta_k\}_{k\in\mathbb{N}} \subset (0, +\infty)$. If*

$$\sum_{k\in\mathbb{N}} \lambda_k(2 - \lambda_k) = \infty, \quad \{\lambda_k\delta_k\}_{k\in\mathbb{N}} \in \ell^1, \quad \textit{and} \quad \left\|\bar{U}^{-1/2}\Gamma RB\bar{U}^{-1/2}\right\| < 1,$$

then for any $\tilde{w}_0 \in \mathbb{R}^{n+m}$ IIFP²A (93) is executable, and the sequence $\{\sqrt{V}\tilde{w}_k\}_{k\in\mathbb{N}}$ generated by IIFP²A (93) converges to a point of $\sqrt{V}\operatorname{Fix}\mathcal{F}$. If further $\lim_{k\to\infty}\delta_k = 0$, then the sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by IIFP²A (93) converges to a primal/dual solution pair of model (89).

Proof. Notice that IIFP²A (93) can be recast into inexact KM iteration (8) as

$$\tilde{w}_{k+1} = (1 - \lambda_k)\tilde{w}_k + \lambda_k(\mathcal{F}\tilde{w}_k + \varepsilon_k),$$

where by Theorem 3.15, $\mathcal{F}$ is nonexpansive with respect to V, quasiaveraged with respect to (V, V), and for all $\lambda \in (0, 2]$ we have $\mathcal{F}_{\lambda_k}$ is nonexpansive with respect to V. Also by Theorem 4.1, there exists $c > 0$ such that

$$\|\varepsilon_k\| = \|\tilde{z}_{k+1} - \mathcal{F}\tilde{w}_k\| \le c \cdot \delta_k, \qquad \text{for all } k \in \mathbb{N}.$$

This ensures that $\{\lambda_k\varepsilon_k\}_{k\in\mathbb{N}} \in \ell^1$. So far conditions of Theorem 2.36 hold, therefore the convergence of $\{\sqrt{V}\tilde{w}_k\}_{k\in\mathbb{N}}$ towards $\sqrt{V}\operatorname{Fix}\mathcal{F}$ is guaranteed. Since $\operatorname{Ran}\sqrt{V}$ is closed, set $w_* \in \operatorname{Fix}\mathcal{F}$ such that $\lim_{k\to\infty}\sqrt{V}\tilde{w}_k = \sqrt{V}\tilde{w}_*$. Finally, if further $\lim_{k\to\infty}\delta_k = 0$ then we have $\lim_{k\to\infty}\varepsilon_k = 0$. Furthermore, by Proposition 2.20 and continuity of $\mathcal{F}$ on $\operatorname{Ran}V$, we have

$$\tilde{z}_{k+1} = \mathcal{F}\tilde{w}_k + \varepsilon_k = \mathcal{F}\sqrt{V}^\dagger\sqrt{V}\tilde{w}_k + \varepsilon_k \to \mathcal{F}\sqrt{V}^\dagger\sqrt{V}\tilde{w}_* = \mathcal{F}w_* = w_*, \qquad \text{as } k \to \infty,$$

which proves the convergence of $\{\tilde{z}_k\}_{k\in\mathbb{N}}$. □

Theorem 4.5 has an immediate corollary for θ-IBSFP²A.

Theorem 4.6. *Suppose $P \in \mathbb{S}^n_+$, $Q \in \mathbb{S}^m_+$, $\theta \in \mathbb{R}$, $\{\delta_k\}_{k\in\mathbb{N}} \subset (0, +\infty)$ and the followings hold.*

(i) Either, $\theta \neq 0$ and

$$2 \min\{|1-\theta|, |1+\theta|\} \max\left\{ \left\|P^{-1}\right\|, \left\|Q^{-1}\right\|\right\} \|L\| \leq 1 - |\theta| \left\|Q^{-1/2} L P^{-1/2}\right\|,$$

or, $\theta = 0$ and

$$2 \max\left\{ \left\|P^{-1}\right\|, \left\|Q^{-1}\right\|\right\} \|L\| < 1,$$

(ii) $\{\lambda_k\}_{k\in\mathbb{N}} \subset [0, 2]$ such that

$$\sum_{k\in\mathbb{N}} \lambda_k (2 - \lambda_k) = \infty,$$

(iii) $\{\lambda_k \delta_k\}_{k\in\mathbb{N}} \in \ell^1$,

then for any $\tilde{w}_0 \in \mathbb{R}^{n+m}$, the θ-IBSFP²A (97) *is executable, and the sequence $\{\sqrt{V}\tilde{w}_k\}_{k\in\mathbb{N}}$ generated by θ-IBSFP²A* (97) *converges to a point of $\sqrt{V}\,\mathrm{Fix}\,\mathcal{F}$. If further $\lim_{k\to\infty} \delta_k = 0$, then the sequence $\{\tilde{z}_k\}_{k\in\mathbb{N}}$ generated by θ-IBSFP²A* (97) *converges to a primal/dual solution pair of model* (89).

Proof. By Theorems 4.4 and 4.5, it is sufficient to check $RM \in \mathbb{S}^{n+m}$. Notice that in this case,

$$RM = \begin{bmatrix} P & \theta L^\top \\ \theta L & Q \end{bmatrix}.$$

If $\theta = 0$ then $RM \in \mathbb{S}^{n+m}_+$. If $\theta \neq 0$, then by assumption (i), we have $|\theta| \|Q^{-1/2} L P^{-1/2}\| \leq 1$. By Lemma 4.3, we have $RM \in \mathbb{S}^{n+m}$. □

Theorem 4.6 extends existing convergence theorems in several aspects. With $|\theta| = 1$ (97) reduces to explicit FP²As Douglas-Rachford splitting algorithm and first-order primal-dual algorithm. With $|\theta| \neq 1$, Theorem 4.6 generalizes explicit FP²As, and refines the analysis of inexact IFP²A cases discussed in [63].

CHAPTER 5

APPLICATIONS OF INEXACT IMPLICIT FIXED-POINT PROXIMITY ALGORITHMS

This chapter applies θ-IBSFP²A and several common explicit FP²As to the problem of recovering images degraded by Gaussian blurring and different kinds of noise. Specially, we propose L2-TV deblurring model and L1-TV deblurring model for Gaussian and uniform impulse noises, then apply ADMM (82), explicit FP²A (77) and the θ-IBSFP²A (97) to each of the deblurring models. Numerical experiments of each method on each model are conducted, and the comparisons between methods are presented thereafter, illustrating the convergence speed advantage of IIFP²A over explicit FP²As.

5.1 IMAGE DEBLURRING PROBLEMS

Assume an 8-bit gray-scale image of size $n \times n$, restored as a vector $u \in \mathbb{R}^{n^2}$, is corrupted by Gaussian blurring kernel and Gaussian/impulse noises. For Gaussian noise, the observed image is modeled by

$$z = Ku + \omega, \tag{102}$$

where $K \in \mathcal{B}(\mathbb{R}^{n^2})$ is a Gaussian blurring kernel, and $\omega \in \mathbb{R}^{n^2}$ represents a Gaussian-distributed noise with 0 mean. For deblurring problem with Gaussian noise, the well-known Rudin-Osher-Fatemi (ROF) image denoising model [89] (or L2-TV model) is applied, namely

$$\min_{x \in \mathbb{R}^{n^2}} \left\{ \frac{\mu}{2} \|Kx - z\|_2^2 + \|x\|_{\mathrm{TV}} \right\}, \tag{103}$$

where $\mu > 0$, $K \in \mathcal{B}(\mathbb{R}^{n^2})$ representing a Gaussian convolution kernel, and the TV semi-norm $\|x\|_{\mathrm{TV}} = \|Lx\|_1$ is defined as

$$L := \begin{bmatrix} I_n \otimes D_n \\ D_n \otimes I_n \end{bmatrix}, \qquad D_n := \begin{bmatrix} 1 & & & -1 \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix}_{n\times n}, \tag{104}$$

in which '⊗', with an abuse of notation, represents the Kronecker product of two matrices.

Then we consider the image deblurring problem with uniform impulse noise. Uniform impulse noise presents itself as sparsely occurring pixels taking random numbers. The observed image $z \in \mathbb{R}^{n\times n}$ therefore can be modeled by

$$z_{ij} := \begin{cases} U[0, 255], & \text{with probability } p, \\ [Ku]_{ij}, & \text{with probability } 1-p, \end{cases} \qquad \text{for all } 1 \le i, j \le n. \tag{105}$$

where $p \in [0, 1]$ and $U[a, b]$ denotes the uniform distribution on $[a, b]$. For the uniform impulse noise the L1-TV model [2, 74, 77] is particularly suggested, which reads

$$\min_{x\in\mathbb{R}^{n^2}} \{\mu\|Kx - z\|_1 + \|x\|_{\mathrm{TV}}\}, \tag{106}$$

where $\mu > 0$. The procedures of models (102) and (105) for Gaussian and impulsive noises are illustrated in Figure 3.

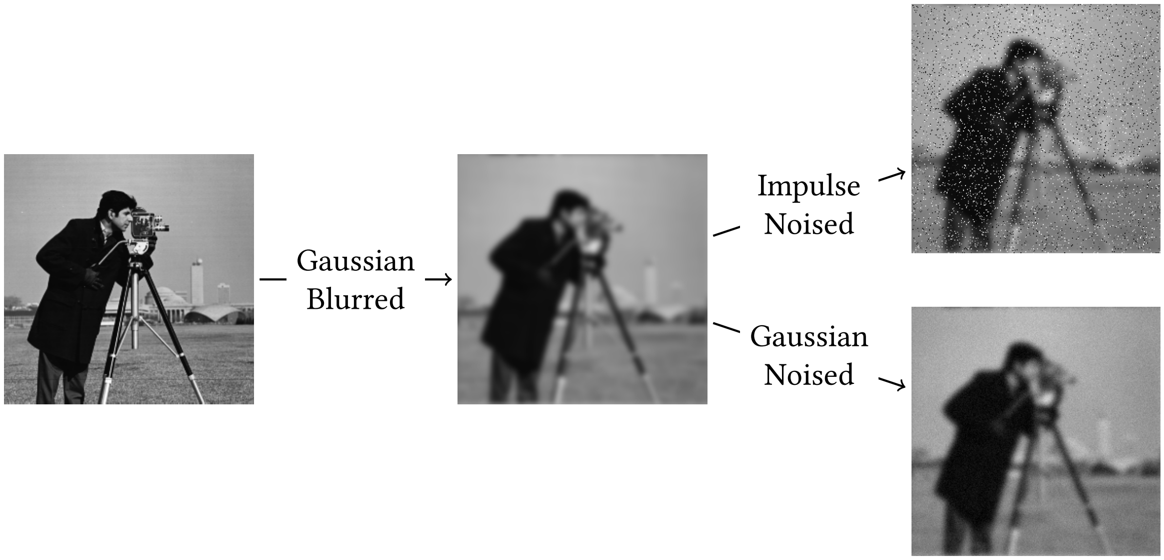
5.2 L2-TV DEBLURRING MODEL

In this section we propose linearized ADMM, explicit FP²A and θ-IBSFP²A for image deblurring problem with Gaussian noise (103).

To propose linearized ADMM, explicit FP²A and inexact implicit FP²A schemes for L2-TV model (103), firstly we need to fit model (103) into the standard form (45) and (81). This could be done by setting $f = \mu\|K\cdot - z\|_2^2/2$ and $g = \|\cdot\|_1$, which gives us

$$\min_{x\in\mathbb{R}^{n^2}} \{f(x) + g(Lx)\}, \tag{107}$$

Figure 3. Procedures of Gaussian blurring and Gaussian/impulse noising of an image



and

$$\min_{x\in\mathbb{R}^{n^2},z\in\mathbb{R}^{2n^2}}\{f(x)+g(z)\},\quad \text{subject to } Lx-z=0. \tag{108}$$

Observe that $\mathrm{prox}_{f,\alpha I_{n^2}}(x)=(\alpha I_{n^2}+\mu K^\top K)^{-1}(\alpha x+\mu K^\top z)$ for $\alpha>0$, and since K is a convolution kernel, $\mathrm{prox}_{f,\alpha I_{n^2}}$ could be efficiently computed by fast Fourier transformation. Then we directly have the linearized ADMM (88) for model (108) as (109).

$$\begin{aligned}&\textbf{For } k\in\mathbb{N}\\ &\left|\begin{array}{l} z_{k+1}\leftarrow \mathrm{prox}_{\|\cdot\|_1,\beta I_{2n^2}}\big(z_k+\beta^{-1}(Lx_k-z_k+y_k)\big)\\ y_{k+1}\leftarrow y_k+Lx_k-z_{k+1}\\ x_{k+1}\leftarrow \big(\alpha I_{n^2}+\mu K^\top K\big)^{-1}\big(\alpha x_k+L^\top(y_k-2y_{k+1})+\mu K^\top z\big)\end{array}\right.\end{aligned} \tag{109}$$

Set $P=\alpha I_{2n^2}$ and $Q=\beta I_{n^2}$ in explicit FP²A (77) on model (107) to get (110).

$$\begin{aligned}&\textbf{For } k\in\mathbb{N}\\ &\left|\begin{array}{l} y_{k+1}\leftarrow \mathrm{prox}_{\|\cdot\|_1^*,\beta I_{2n^2}}\big(\beta^{-1}Lx_k+y_k\big)\\ x_{k+1}\leftarrow \big(\alpha I_{n^2}+\mu K^\top K\big)^{-1}\big(\alpha x_k+L^\top(y_k-2y_{k+1})+\mu K^\top z\big)\end{array}\right.\end{aligned} \tag{110}$$

Furthermore, set $P = \alpha I_{n^2}$, $Q = \beta I_{2n^2}$, $\lambda_k = 1$ for $k \in \mathbb{N}$ and

$$A = \begin{bmatrix} 0_{n^2} & -(1+\theta)\alpha^{-1}L^\top \\ 0 & 0_{2n^2} \end{bmatrix}, \qquad B = \begin{bmatrix} 0_{n^2} & 0 \\ (1-\theta)\beta^{-1}L & 0_{2n^2} \end{bmatrix},$$

in θ-IBSFP²A (97) to get (111).

$$\begin{array}{l}
\textbf{For } k \in \mathbb{N} \\
\quad \textbf{Do } l \in \mathbb{N} \\
\quad\quad y_k^{l+1} \leftarrow \operatorname{prox}_{\|\cdot\|_1^*, \beta I_{2n^2}}\left(\beta^{-1}L\left((1-\theta)x_k^l + \theta x_k\right) + y_k\right) \\
\quad\quad x_k^{l+1} \leftarrow \left(\alpha I_{n^2} + \mu K^\top K\right)^{-1}\left(\alpha x_k + L^\top\left(\theta y_k - (1+\theta)y_k^{l+1}\right) + \mu K^\top z\right) \\
\quad \textbf{While } \|x_k^{l+1} - x_k^l\|^2 + \|y_k^{l+1} - y_k^l\|^2 > \delta_k^2 \\
\quad x_{k+1} \leftarrow x_k^{l+1},\ y_{k+1} \leftarrow y_k^{l+1}
\end{array} \qquad (111)$$

Notice that with $\theta = 1$, (111) reduces to (110), and (110) is identical to (109) when $\beta = 1$. By Theorems 3.21, 3.26 and 4.6, we have the following convergence theorems for (109) to (111).

Corollary 5.1. *If $\alpha \geq \|L\|^2$ and $\beta \geq 1$, then for any $(x_0, y_0, z_0) \in \mathbb{R}^{n^2+2n^2+2n^2}$, the iteration sequence $\{(x_k, y_k, z_k)\}_{k\in\mathbb{N}}$ generated by linearized ADMM for L2-TV model* (109) *converges to a primal/dual solution pair of L2-TV model* (108).

Corollary 5.2. *Suppose $\alpha, \beta > 0$. If $\alpha\beta \geq \|L\|^2$, then for any $(x_0, y_0) \in \mathbb{R}^{n^2+2n^2}$, the iteration sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}}$ generated by explicit FP²A for L2-TV model* (110) *converges to a primal/dual solution pair of L2-TV model* (107).

Corollary 5.3. *Suppose $\alpha, \beta > 0$, $\theta \in \mathbb{R}$ and positive real sequence $\{\delta_k\}_{k\in\mathbb{N}} \in \ell^1$. If, either, $\theta \neq 0$ and*

$$2\min\{|1-\theta|, |1+\theta|\}\max\left\{\sqrt{\alpha/\beta}, \sqrt{\beta/\alpha}\right\}\|L\| \leq \sqrt{\alpha\beta} - |\theta|\|L\|,$$

or, $\theta = 0$ and

$$2(\alpha+\beta)\|L\| < \alpha\beta,$$

then for any $(x_0, y_0) \in \mathbb{R}^{n^2+2n^2}$, the iteration sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}}$ generated by θ-IBSFP²A for L2-TV model (111) *converges to a primal/dual solution pair of L2-TV model* (107).

Notice that Corollary 5.3 covers Corollary 5.2 as a special case of $\theta = 1$.

5.3 L1-TV DEBLURRING MODEL

In this section we propose linearized ADMM, explicit FP²A and θ-IBSFP²A for image deblurring problem with impulse noise (106).

Similar to Section 5.2, to propose linearized ADMM, explicit FP²A and inexact implicit FP²A schemes for L1-TV model (106), firstly we need to fit model (106) into the standard form (45) and (81). This could be done by setting $f = 0$, $g = \|\cdot - b\|_1$ and

$$\mathcal{L} := \begin{bmatrix} \mu K \\ L \end{bmatrix}, \qquad b := \begin{bmatrix} \mu z \\ 0 \end{bmatrix},$$

which gives us

$$\min_{x \in \mathbb{R}^{n^2}} g(\mathcal{L}x). \tag{112}$$

Notice that in this case $\operatorname{prox}_{f,\alpha I_{n^2}} = I_{n^2}$ and $\operatorname{prox}_{g,\beta I_{2n^2}} = \operatorname{prox}_{\|\cdot\|_1,\beta I_{2n^2}}(\cdot - b) + b$ for $\alpha, \beta > 0$. Then we directly have the linearized ADMM (88) for model (112) as (109).

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} z_{k+1} \leftarrow \operatorname{prox}_{\|\cdot - b\|_1, \beta I_{2n^2}}\left(z_k + \beta^{-1}(\mathcal{L}x_k - z_k + y_k)\right) \\ y_{k+1} \leftarrow y_k + \mathcal{L}x_k - z_{k+1} \\ x_{k+1} \leftarrow x_k + \alpha^{-1}\mathcal{L}^\top(y_k - 2y_{k+1}) \end{array} \right. \end{array} \tag{113}$$

Set $P = \alpha I_{2n^2}$ and $Q = \beta I_{n^2}$ in explicit FP²A (77) on model (112) to get (110).

$$\begin{array}{l} \textbf{For } k \in \mathbb{N} \\ \left\lfloor \begin{array}{l} y_{k+1} \leftarrow \operatorname{prox}_{\|\cdot - b\|_1^*, \beta I_{2n^2}}\left(\beta^{-1}Lx_k + y_k\right) \\ x_{k+1} \leftarrow x_k + \alpha^{-1}\mathcal{L}^\top(y_k - 2y_{k+1}) \end{array} \right. \end{array} \tag{114}$$

Furthermore, set $P = \alpha I_{n^2}$, $Q = \beta I_{2n^2}$, $\lambda_k = 1$ for $k \in \mathbb{N}$ and

$$A = \begin{bmatrix} 0_{n^2} & -(1+\theta)\alpha^{-1}L^\top \\ 0 & 0_{2n^2} \end{bmatrix}, \qquad B = \begin{bmatrix} 0_{n^2} & 0 \\ (1-\theta)\beta^{-1}L & 0_{2n^2} \end{bmatrix},$$

in θ-IBSFP²A (97) to get (111).

$$
\begin{array}{l}
\textbf{For } k \in \mathbb{N} \\
\left\lfloor \begin{array}{l}
\textbf{Do } l \in \mathbb{N} \\
\left| \begin{array}{l}
y_k^{l+1} \leftarrow \operatorname{prox}_{\|\cdot-b\|_1^*, \beta I_{2n^2}}\left(\beta^{-1}\mathcal{L}\big((1-\theta)x_k^l + \theta x_k\big) + y_k\right) \\
x_k^{l+1} \leftarrow x_k + \alpha^{-1}\mathcal{L}^\top\big(\theta y_k - (1+\theta)y_k^{l+1}\big)
\end{array}\right. \\
\textbf{While } \|x_k^{l+1} - x_k^l\|^2 + \|y_k^{l+1} - y_k^l\|^2 > \delta_k^2 \\
x_{k+1} \leftarrow x_k^{l+1}, \; y_{k+1} \leftarrow y_k^{l+1}
\end{array}\right.
\end{array}
\tag{115}
$$

Notice that with $\theta = 1$, (115) reduces to (114), and (114) is identical to (113) when $\beta = 1$. By Theorems 3.21, 3.26 and 4.6, we have the following convergence theorems for (113) to (115).

Corollary 5.4. *If $\alpha \geq \|L\|^2$ and $\beta \geq 1$, then for any $(x_0, y_0, z_0) \in \mathbb{R}^{n^2+2n^2+2n^2}$, the iteration sequence $\{(x_k, y_k, z_k)\}_{k\in\mathbb{N}}$ generated by linearized ADMM for L1-TV model* (113) *converges to a primal/dual solution pair of L1-TV model* (112).

Corollary 5.5. *Suppose $\alpha, \beta > 0$. If $\alpha\beta \geq \|L\|^2$, then for any $(x_0, y_0) \in \mathbb{R}^{n^2+2n^2}$, the iteration sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}}$ generated by explicit FP²A for L1-TV model* (114) *converges to a primal/dual solution pair of L1-TV model* (112).

Corollary 5.6. *Suppose $\alpha, \beta > 0$, $\theta \in \mathbb{R}$ and positive real sequence $\{\delta_k\}_{k\in\mathbb{N}} \in \ell^1$. If, either, $\theta \neq 0$ and*

$$2 \min\{|1-\theta|, |1+\theta|\} \max\left\{\sqrt{\alpha/\beta}, \sqrt{\beta/\alpha}\right\}\|L\| \leq \sqrt{\alpha\beta} - |\theta|\|L\|,$$

or, $\theta = 0$ and

$$2(\alpha + \beta)\|L\| < \alpha\beta,$$

then for any $(x_0, y_0) \in \mathbb{R}^{n^2+2n^2}$, the iteration sequence $\{(x_k, y_k)\}_{k\in\mathbb{N}}$ generated by θ-IBSFP²A for L1-TV model (115) *converges to a primal/dual solution pair of L1-TV model* (112).

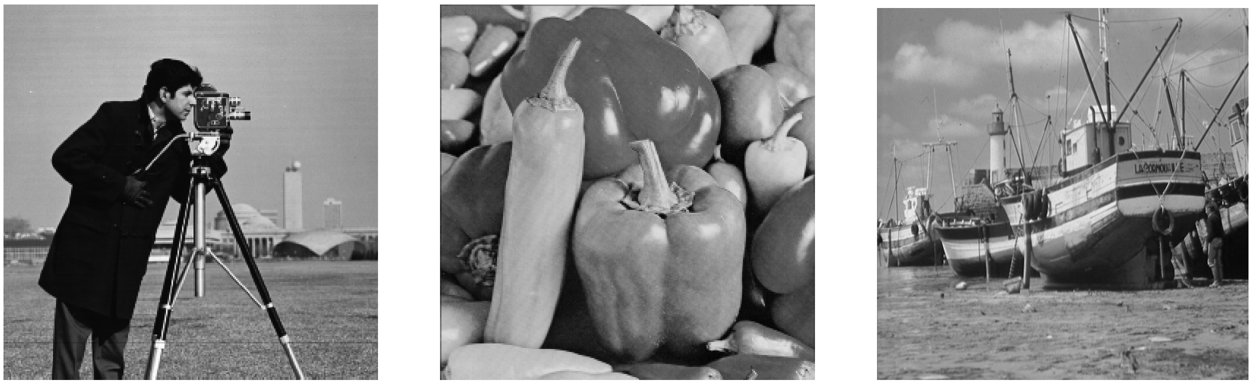
Notice that Corollary 5.6 covers Corollary 5.5 as a special case of $\theta = 1$.

5.4 NUMERICAL EXPERIMENTS

In this section we compare the behavior of linearized ADMM (109) and (113), explicit FP²A (110) and (114) and the proposed θ-IBSFP²A (111) and (115) on image debluring problems L2-TV and L1-TV introduced in previous sections. All numerical experiments are conducted in Matlab

Figure 4. Test images of 8-bit gray-scale with size 256×256

(a) `Cameraman` (b) `Peppers` (c) `Fishingboats`



R2020b, with Windows 10 on an Intel Core i7 CPU @ 3.60GHz and 16GB RAM memory. Three 8-bit gray-scale images, namely `Cameraman`, `Peppers` and `Fishingboats`, are tested in this section, as shown in Figure 4.

As detailed in Section 5.1, images are first blurred using Gaussian kernels with standard deviations $\sigma = 1, 3, 5$, then they are corrupted by two types of noise

(G) Add Gaussian noise with deviation 5; see Figure 5.

(R) Randomly pick 10% pixels on the image and cover them according to the uniform distribution on $[0, 255]$; see Figure 6.

We store $n{\times}n$ images with vectors in $\mathbb{R}^{n^2}$, by stacking each columns of the image. To compare the performance of each algorithms in a level playing field, we unify the regularization parameter of three methods in each scenario. Firstly we execute explicit FP²A (110) and (114) for 10^4 steps, solution of which is treated as the ground truth $x_{\mathrm{gt}} \in \mathbb{R}^{n^2}$ of the corresponding model. Then for each method, we exit the scheme once the iteration satisfies the stop criteria

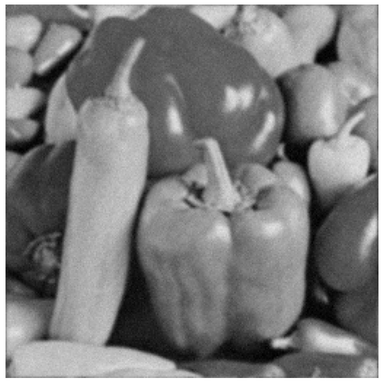
$$\frac{F(x_k) - F\left(x_{\mathrm{gt}}\right)}{\left|F\left(x_{\mathrm{gt}}\right)\right|} < \varepsilon, \tag{116}$$

where F denotes the objective function of the model. Each method is repeated 10 times at the optimal algorithm parameters and the mean CPU time are recorded. We also examine the Peak

Figure 5. Gaussian-blurred images with different standard deviations and additive Gaussian noise of deviation 5

(a) Gaussian blurred with $\sigma = 1$ and noised by (G).

(b) Gaussian blurred with $\sigma = 1$ and noised by (G).



(c) Gaussian blurred with $\sigma = 1$ and noised by (G).

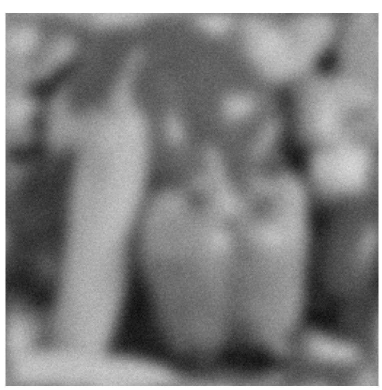
(d) Gaussian blurred with $\sigma = 3$ and noised by (G).

(e) Gaussian blurred with $\sigma = 3$ and noised by (G).

(f) Gaussian blurred with $\sigma = 3$ and noised by (G).

(g) Gaussian blurred with $\sigma = 5$ and noised by (G).

(h) Gaussian blurred with $\sigma = 5$ and noised by (G).



(i) Gaussian blurred with $\sigma = 5$ and noised by (G).

Figure 6. Gaussian-blurred images with different standard deviations and 10% random impulse noise

(a) Gaussian blurred with $\sigma = 1$ and noised by (R).

(b) Gaussian blurred with $\sigma = 1$ and noised by (R).

(c) Gaussian blurred with $\sigma = 1$ and noised by (R).

(d) Gaussian blurred with $\sigma = 3$ and noised by (R).

(e) Gaussian blurred with $\sigma = 3$ and noised by (R).

(f) Gaussian blurred with $\sigma = 3$ and noised by (R).

(g) Gaussian blurred with $\sigma = 5$ and noised by (R).

(h) Gaussian blurred with $\sigma = 5$ and noised by (R).

(i) Gaussian blurred with $\sigma = 5$ and noised by (R).

Signal-to-Noise Ratio (PSNR) measurements to quantify reconstruction qualities for those corrupted images. The PSNR (in dB) for 8-bit gray-scale 256×256 image is defined by

$$\mathrm{PSNR} = 20 \times \log_{10} \frac{255 \times 256}{\|x_{\mathrm{rec}} - x_{\mathrm{ori}}\|_2},$$

where $x_{\mathrm{ori}} \in \mathbb{R}^{n^2}$ is the uncorrupted original image, and $x_{\mathrm{rec}} \in \mathbb{R}^{n^2}$ is the recovered one.

For L2-TV model, Table 3 lists the PSNR and execution time of algorithms on each experiment of L2-TV model and Figures 7 to 9 show the deblurred images of each methods. For L1-TV model, Table 4 lists the PSNR and execution time of algorithms on each experiment of L1-TV model and Figures 10 to 12 show the deblurred images of each methods. Experiments shows that all the methods converges to the solution of the problem with similar PSNR, while θ-IBSFP^{2}A always costs the least of the CPU time, demonstrating the advantages of the proposed inexact implicit fixed-point proximity algorithm over explicit ones.

Table 3. Summary of Linearized ADMM, Explicit FP^2A and θ-IBSFP^{2}A on L2-TV

Blurring Level	Method	Cameraman		Peppers		Fishingboats	
		PSNR	Time(s)	PSNR	Time(s)	PSNR	Time(s)
$\sigma = 1$	LADMM	28.02	1.46	30.52	3.92	28.17	3.78
	Ex-FP^{2}A	28.02	2.21	30.54	3.48	28.18	2.80
	IBSFP^{2}A	28.02	1.14	30.53	1.62	28.18	1.31
$\sigma = 3$	LADMM	22.97	6.34	24.43	6.07	23.29	6.27
	Ex-FP^{2}A	22.94	5.87	24.46	8.02	23.30	7.81
	IBSFP^{2}A	22.95	3.69	24.45	4.49	23.30	4.64
$\sigma = 5$	LADMM	21.27	6.11	22.47	5.37	22.08	5.60
	Ex-FP^{2}A	21.28	6.63	22.50	7.08	22.09	6.92
	IBSFP^{2}A	21.29	4.88	22.49	4.39	22.08	4.53

Table 4. Summary of Linearized ADMM, Explicit FP^2A and θ-IBSFP^{2}A on L1-TV

Blurring Level	Method	Cameraman		Peppers		Fishingboats	
		PSNR	Time(s)	PSNR	Time(s)	PSNR	Time(s)
$\sigma = 1$	LADMM	31.35	109.33	32.40	106.67	33.69	103.13
	Ex-FP^{2}A	31.35	96.85	32.40	101.06	33.69	93.14
	IBSFP^{2}A	32.18	50.59	33.29	51.61	34.80	41.91
$\sigma = 3$	LADMM	25.20	168.42	29.33	158.13	25.95	167.46
	Ex-FP^{2}A	25.20	149.81	29.33	141.80	25.95	165.94
	IBSFP^{2}A	25.14	56.07	29.49	59.29	25.98	58.64
$\sigma = 5$	LADMM	22.81	175.49	26.02	169.66	23.98	164.81
	Ex-FP^{2}A	22.81	166.32	26.02	155.92	23.98	154.70
	IBSFP^{2}A	22.82	89.85	26.01	70.80	23.97	64.23

Figure 7. Recovered images of L2-TV model on `Cameraman`

(a) LADMM for $\sigma = 1$ with PSNR 28.02, CPU time 1.46s.

(b) Ex-FP²A for $\sigma = 1$ with PSNR 28.02, CPU time 2.21s.

(c) IBSFP²A for $\sigma = 1$ with PSNR 28.02, CPU time 1.14s.

(d) LADMM for $\sigma = 3$ with PSNR 22.97, CPU time 6.34s.

(e) Ex-FP²A for $\sigma = 3$ with PSNR 22.94, CPU time 5.87s.

(f) IBSFP²A for $\sigma = 3$ with PSNR 22.95, CPU time 3.69s.

(g) LADMM for $\sigma = 5$ with PSNR 21.27, CPU time 6.11s.

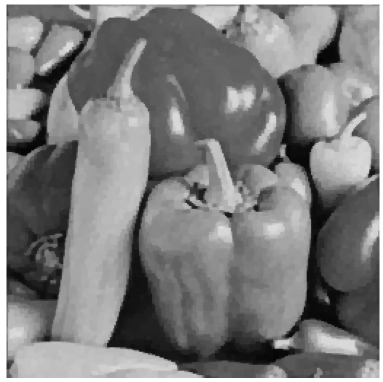
(h) Ex-FP²A for $\sigma = 5$ with PSNR 21.28, CPU time 6.63s.

(i) IBSFP²A for $\sigma = 5$ with PSNR 21.29, CPU time 4.88s.

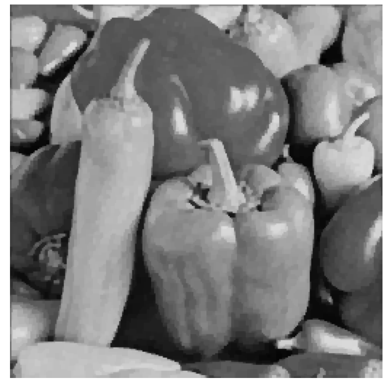
Figure 8. Recovered images of L2-TV model on `Peppers`

(a) LADMM for $\sigma = 1$ with PSNR 30.52, CPU time 3.92s.

(b) Ex-FP²A for $\sigma = 1$ with PSNR 30.54, CPU time 3.48s.



(c) IBSFP²A for $\sigma = 1$ with PSNR 30.53, CPU time 1.62s.



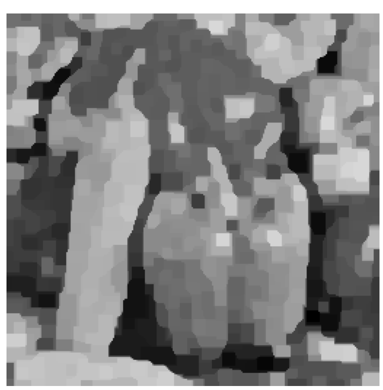
(d) LADMM for $\sigma = 3$ with PSNR 24.43, CPU time 6.07s.

(e) Ex-FP²A for $\sigma = 3$ with PSNR 24.46, CPU time 8.02s.

(f) IBSFP²A for $\sigma = 3$ with PSNR 24.45, CPU time 4.49s.

(g) LADMM for $\sigma = 5$ with PSNR 22.47, CPU time 5.37s.

(h) Ex-FP²A for $\sigma = 5$ with PSNR 22.50, CPU time 7.08s.



(i) IBSFP²A for $\sigma = 5$ with PSNR 22.49, CPU time 4.39s.

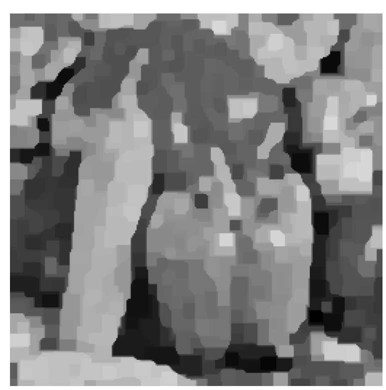


Figure 9. Recovered images of L2-TV model on `Fishingboats`

(a) LADMM for $\sigma = 1$ with PSNR 28.17, CPU time 3.78s.

(b) Ex-FP²A for $\sigma = 1$ with PSNR 28.18, CPU time 2.80s.

(c) IBSFP²A for $\sigma = 1$ with PSNR 28.18, CPU time 1.31s.

(d) LADMM for $\sigma = 3$ with PSNR 23.29, CPU time 6.27s.

(e) Ex-FP²A for $\sigma = 3$ with PSNR 23.30, CPU time 7.81s.

(f) IBSFP²A for $\sigma = 3$ with PSNR 23.30, CPU time 4.64s.

(g) LADMM for $\sigma = 5$ with PSNR 22.08, CPU time 5.60s.

(h) Ex-FP²A for $\sigma = 5$ with PSNR 22.09, CPU time 6.92s.

(i) IBSFP²A for $\sigma = 5$ with PSNR 22.08, CPU time 4.53s.

Figure 10. Recovered images of L1-TV model on `Cameraman`

(a) LADMM for $\sigma = 1$ with PSNR 31.35, CPU time 109.33s.

(b) Ex-FP²A for $\sigma = 1$ with PSNR 31.35, CPU time 96.85s.

(c) IBSFP²A for $\sigma = 1$ with PSNR 32.18, CPU time 50.59s.

(d) LADMM for $\sigma = 3$ with PSNR 25.20, CPU time 168.42s.

(e) Ex-FP²A for $\sigma = 3$ with PSNR 25.20, CPU time 149.81s.

(f) IBSFP²A for $\sigma = 3$ with PSNR 25.14, CPU time 56.07s.

(g) LADMM for $\sigma = 5$ with PSNR 22.81, CPU time 175.49s.

(h) Ex-FP²A for $\sigma = 5$ with PSNR 22.81, CPU time 166.32s.

(i) IBSFP²A for $\sigma = 5$ with PSNR 22.82, CPU time 89.85s.

Figure 11. Recovered images of L1-TV model on Peppers

(a) LADMM for $\sigma = 1$ with PSNR 32.40, CPU time 106.67s.

(b) Ex-FP²A for $\sigma = 1$ with PSNR 32.40, CPU time 101.06s.

(c) IBSFP²A for $\sigma = 1$ with PSNR 33.29, CPU time 51.61s.

(d) LADMM for $\sigma = 3$ with PSNR 29.33, CPU time 158.13s.

(e) Ex-FP²A for $\sigma = 3$ with PSNR 29.33, CPU time 141.80s.

(f) IBSFP²A for $\sigma = 3$ with PSNR 29.49, CPU time 59.29s.

(g) LADMM for $\sigma = 5$ with PSNR 26.02, CPU time 169.66s.

(h) Ex-FP²A for $\sigma = 5$ with PSNR 26.02, CPU time 155.92s.

(i) IBSFP²A for $\sigma = 5$ with PSNR 26.01, CPU time 70.80s.

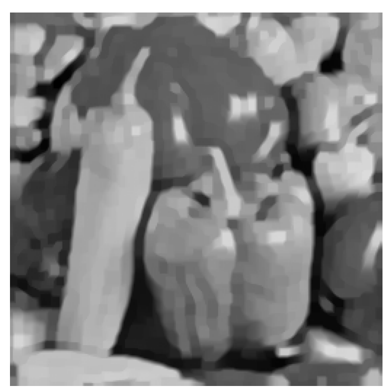


Figure 12. Recovered images of L1-TV model on `Fishingboats`

(a) LADMM for $\sigma = 1$ with PSNR 33.69, CPU time 103.13s.

(b) Ex-FP²A for $\sigma = 1$ with PSNR 33.69, CPU time 93.14s.

(c) IBSFP²A for $\sigma = 1$ with PSNR 34.80, CPU time 41.91s.

(d) LADMM for $\sigma = 3$ with PSNR 25.95, CPU time 167.46s.

(e) Ex-FP²A for $\sigma = 3$ with PSNR 25.95, CPU time 165.94s.

(f) IBSFP²A for $\sigma = 3$ with PSNR 25.98, CPU time 58.64s.

(g) LADMM for $\sigma = 5$ with PSNR 23.98, CPU time 164.81s.

(h) Ex-FP²A for $\sigma = 5$ with PSNR 23.98, CPU time 154.70s.

(i) IBSFP²A for $\sigma = 5$ with PSNR 23.97, CPU time 64.23s.

CHAPTER 6

CONCLUSIONS

This dissertation systematically studies the framework of inexact fixed-point iterations, the framework of implicit fixed-point proximity algorithms for nonsmooth convex optimization, and the framework of inexact implicit fixed-point proximity algorithms as a combination of the above two frameworks. Then numerical experiments on image deblurring problems show the advantages of the proposed framework of inexact implicit fixed-point proximity algorithms when compared with explicit algorithms.

Chapter 2 proposes a framework of inexact fixed-point iterations for Picard iterations and KM iterations, which then is applied to conduct convergence analysis for inexact Picard/KM iterations for quasinonexpansive/quasiaveraged operators. These two classes of operators generalize the classical nonexpansive/nonexpansive averaged operators, and therefore the proposed framework of inexact iterations extends many popular analysis of existing inexact fixed-point iterations.

Chapter 3 proposes a framework of implicit fixed-point proximity algorithms. This framework first characterizes solutions of the nonsmooth convex optimization problems to be fixed-points of a proximity equation composing with an expanding linear transformation. To deal with the expanding matrix, we adopt the matrix splitting technique to have a general formulation of implicit fixed-point proximity algorithm. Then we utilize quasinonexpansive/quasiaveraged operators proposed in Chapter 2 to propose convergence analysis, which then again extends existing convergence analysis of proximity algorithms. This framework covers gradient descent method, proximal point method, Douglas-Rachford splitting algorithm, first-order primal-dual algorithm, alternating direction method of multipliers, linearized alternating direction method of multipliers and inexact Uzawa method as special cases. The proposed framework provides several generalizations to some of the mentioned methods.

Chapter 4 then combines frameworks in Chapters 2 and 3 to propose the inexact implicit fixed-point proximity algorithms when the iteration is essentially implicit and cannot be exactly updated. Here we again apply the matrix splitting technique to introduce an inner loop for the implicit update, and to have *a posteriori* error estimation for the inexactness brought by the inner loop. Then by a direct application of frameworks in Chapters 2 and 3 we have the framework of inexact implicit fixed-point proximity algorithms. As a concrete application, we propose the class of θ-inexact block-separable fixed-point proximity algorithms with convergence analysis.

Finally Chapter 5 applies the θ-inexact block-separable fixed-point proximity algorithms from Chapter 4 to image deblurring problems with different kinds of noise, and compare it with several classical explicit proximity algorithms. Specifically, we test Gaussian and uniform impulse noises with three different level of blurring, and examine linearized alternating direction method of multipliers, explicit fixed-point proximity algorithm and θ-inexact block-separable fixed-point proximity algorithms. Numerical experiments demonstrate the advantage of inexact implicit fixed-point proximity algorithms in convergence speed.

As an end to this dissertation, we discuss possible developments of the framework of inexact fixed-point proximity algorithms in the future. One may notice that IFP²A (64) has a matrix not yet split, *i.e.*, the $R^{-1}A^{\top}$ before $\nabla_{f_2}(Aw_k)$. A decomposition of this matrix will certainly result in a more comprehensive framework. On the other hand, considering full model (44) instead of (89) in IIFP²A will be another possible direction for further study. Moreover, the framework of fixed-point iterations established in Chapter 2 no longer requires nonexpansive operators, which may provide a potential approach for development of algorithms in nonconvex optimization.

REFERENCES

[1] Peter Alfeld. Fixed point iteration with inexact function values. *Math. Comp.*, 38(157): 87–98, 1982. ISSN 0025-5718. doi: 10.2307/2007466.

[2] Stefano Alliney. A property of the minimum vectors of a regularizing functional defined by means of the absolute norm. *IEEE transactions on signal processing*, 45(4):913–917, 1997. doi: 10.1109/78.564179.

[3] J. B. Baillon, R. E. Bruck, and S. Reich. On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces. *Houston J. Math.*, 4(1):1–9, 1978. ISSN 0362-1588. url: `http://hjm.math.uzh.ch/vol04-1.html`.

[4] Stefan Banach. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundamenta Mathematicae*, 3(1):133–181, 1922. url: `http://eudml.org/doc/213289`.

[5] Heinz H. Bauschke and Patrick L. Combettes. A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. *Mathematics of Operations Research*, 26(2): 248–264, 2001. ISSN 0364-765X. doi: 10.1287/moor.26.2.248.10558.

[6] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham, second edition, 2017. ISBN 978-3-319-48310-8; 978-3-319-48311-5. doi: 10.1007/978-3-319-48311-5. With a foreword by Hédy Attouch.

[7] Heinz H. Bauschke, Radu I. Boţ, Warren L. Hare, and Walaa M. Moursi. Attouch-Théra duality revisited: paramonotonicity and operator splitting. *J. Approx. Theory*, 164(8):1065–1084, 2012. ISSN 0021-9045. doi: 10.1016/j.jat.2012.05.008.

[8] Amir Beck and Marc Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, 18(11):2419–2434, 2009. ISSN 1057-7149. doi: 10.1109/TIP.2009.2028250.

[9] Roger Joseph Boscovich. De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impessa. *Bononiensi Scientiarum et Artum Instuto Atque Academia Commentarii*, 4:353–396, 1757. url: `https://archive.org/details/delitterariaexp00boscgoog`.

[10] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. doi: 10.1561/2200000016.

[11] James H. Bramble, Joseph E. Pasciak, and Apostol T. Vassilev. Analysis of the inexact Uzawa algorithm for saddle point problems. *SIAM J. Numer. Anal.*, 34(3):1072–1092, 1997. ISSN 0036-1429. doi: 10.1137/S0036142994273343.

[12] F. E. Browder and W. V. Petryshyn. The solution by iteration of nonlinear functional equations in Banach spaces. *Bull. Amer. Math. Soc.*, 72:571–575, 1966. ISSN 0002-9904. doi: 10.1090/S0002-9904-1966-11544-6.

[13] Felix E Browder. Convergence theorems for sequences of nonlinear operators in banach spaces. *Mathematische Zeitschrift*, 100(3):201–225, 1967. doi: 10.1007/BF01109805.

[14] Ronald Bruck. Nonexpansive projections on subsets of banach spaces. *Pacific Journal of Mathematics*, 47(2):341–355, 1973. doi: 10.2140/pjm.1973.47.341.

[15] Charles Byrne. Iterative oblique projection onto convex sets and the split feasibility problem. *Inverse Problems*, 18(2):441–453, 2002. ISSN 0266-5611. doi: 10.1088/0266-5611/18/2/310.

[16] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006. ISSN 0018-9448. doi: 10.1109/TIT.2005.862083.

[17] Augustin Cauchy et al. Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538, 1847. url: `https://gallica.bnf.fr/ark:/12148/bpt6k2982c`.

[18] Yair Censor, Tommy Elfving, Nirit Kopf, and Thomas Bortfeld. The multiple-sets split feasibility problem and its applications for inverse problems. *Inverse Problems*, 21(6):2071–2084, 2005. ISSN 0266-5611. doi: 10.1088/0266-5611/21/6/017.

[19] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1): 120–145, 2011. ISSN 0924-9907. doi: 10.1007/s10851-010-0251-1.

[20] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 263–340. Walter de Gruyter, Berlin, 2010. doi: 10.1515/9783110226157.263.

[21] Tony F. Chan and Selim Esedoḡlu. Aspects of total variation regularized L^1 function approximation. *SIAM J. Appl. Math.*, 65(5):1817–1837, 2005. ISSN 0036-1399. doi: 10.1137/040604297.

[22] Tony F. Chan, Gene H. Golub, and Pep Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6):1964–1977, 1999. ISSN 1064-8275. doi: 10.1137/S1064827596299767.

[23] Feishe Chen, Lixin Shen, Yuesheng Xu, and Xueying Zeng. The Moreau envelope approach for the L1/TV image denoising model. *Inverse Probl. Imaging*, 8(1):53–77, 2014. ISSN 1930-8337. doi: 10.3934/ipi.2014.8.53.

[24] Feishe Chen, Lixin Shen, Bruce W. Suter, and Yuesheng Xu. Minimizing compositions of functions using proximity algorithms with application in image deblurring. *Frontiers in Applied Mathematics and Statistics*, 2, 2016. ISSN 2297-4687. doi: 10.3389/fams.2016.00012.

[25] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998. ISSN 1064-8275. doi: 10.1137/S1064827596304010.

[26] Zhongying Chen, Charles A. Micchelli, and Yuesheng Xu. *Multiscale methods for Fredholm integral equations*, volume 28 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2015. ISBN 978-1-107-10347-4. doi: 10.1017/CBO9781316216637.

[27] Christian Clason, Bangti Jin, and Karl Kunisch. A duality-based splitting method for ℓ^1-TV image restoration with automatic regularization parameter choice. *SIAM J. Sci. Comput.*, 32(3):1484–1505, 2010. ISSN 1064-8275. doi: 10.1137/090768217.

[28] Patrick L. Combettes. The convex feasibility problem in image recovery. volume 95 of *Advances in Imaging and Electron Physics*, pages 155–270. Elsevier, 1996. doi: 10.1016/S1076-5670(08)70157-5.

[29] Patrick L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization. A Journal of Mathematical Programming and Operations Research*, 53(5-6):475–504, 2004. ISSN 0233-1934. doi: 10.1080/02331930412331327157.

[30] Patrick L. Combettes and Jean-Christophe Pesquet. Fixed point strategies in data science. *IEEE Transactions on Signal Processing*, 69:3878–3905, 2021. doi: 10.1109/TSP.2021.3069677.

[31] Patrick L. Combettes and Valérie R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200, 2005. ISSN 1540-3459. doi: 10.1137/050626090.

[32] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3): 273–297, 1995. doi: 10.1007/BF00994018.

[33] Ron S. Dembo, Stanley C. Eisenstat, and Trond Steihaug. Inexact Newton methods. *SIAM J. Numer. Anal.*, 19(2):400–408, 1982. ISSN 0036-1429. doi: 10.1137/0719025.

[34] J. B. Diaz and F. T. Metcalf. On the structure of the set of subsequential limit points of successive approximations. *Bull. Amer. Math. Soc.*, 73:516–519, 1967. ISSN 0002-9904. doi: 10.1090/S0002-9904-1967-11725-7.

[35] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006. ISSN 0018-9448. doi: 10.1109/TIT.2006.871582.

[36] W. G. Dotson, Jr. Fixed points of quasi-nonexpansive mappings. *J. Austral. Math. Soc.*, 13: 167–170, 1972. ISSN 0263-6115. doi: 10.1017/S144678870001123X.

[37] Jim Douglas and Henry H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439, 1956. url: `https://www.jstor.org/stable/1993056`.

[38] Harris Drucker, Christopher J. C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. In M.C. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9. MIT Press, 1996. url: `https://proceedings.neurips.cc/paper/1996/hash/d38901788c533e8286cb6400b40b386d-Abstract.html`.

[39] Jonathan Eckstein. *Splitting methods for monotone operators with applications to parallel optimization*. Ph.D. thesis, Massachusetts Institute of Technology, 1989. url: `http://hdl.handle.net/1721.1/14356`.

[40] Ernie Esser. Applications of lagrangian-based alternating direction methods and connections to split bregman. *CAM Report*, 9:31, 2009. url: `https://ww3.math.ucla.edu/cam-reports`.

[41] Ernie Esser, Xiaoqun Zhang, and Tony F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.*, 3(4):1015–1046, 2010. doi: 10.1137/09076934X.

[42] Pierre de Fermat, Samuel de Fermat, and Pergaeus Appollonius. *Varia opera mathematica*. apud Johannem Pech, Tolosæ, 1679. doi: 10.5479/sil.128299.39088002705879.

[43] Carl Friedrich Gauss. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss*. sumtibus Frid. Perthes et IH Besser, 1809. url: `https://archive.org/details/bub_gb_ORUOAAAAQAAJ`.

[44] Howard C. Gifford, C. Ross Schmidtlein, Andrzej Krol, and Yuesheng Xu. An assessment of PET dose reduction with penalized likelihood image reconstruction using a computationally efficient model observer. In Guang-Hong Chen and Hilde Bosmans, editors, *Medical Imaging 2020: Physics of Medical Imaging*, volume 11312, pages 192 – 198. International Society for Optics and Photonics, SPIE, 2020. doi: 10.1117/12.2550856.

[45] R. Glowinski and A. Marrocco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér.*, 9(R-2): 41–76, 1975. ISSN 0397-9342. url: `http://www.numdam.org/item/M2AN_1975__9_2_41_0`.

[46] Thomas Goldstein, Min Li, and Xiaoming Yuan. Adaptive primal-dual splitting methods for statistical learning and image processing. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, page 2089–2097, Cambridge, MA, USA, 2015. MIT Press. url: `https://dl.acm.org/doi/10.5555/2969442.2969473`.

[47] Tom Goldstein and Stanley Osher. The split Bregman method for $L1$-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009. doi: 10.1137/080725891.

[48] Jianfeng Guo, C. Ross Schmidtlein, Andrzej Krol, Si Li, Yizun Lin, Sangtae Ahn, Charles Stearns, and Yuesheng Xu. A fast convergent ordered-subsets algorithm with subiteration-dependent preconditioners for PET image reconstruction. *IEEE Transactions on Medical Imaging*, pages 1–1, 2022. doi: 10.1109/TMI.2022.3181813.

[49] Xiaoxia Guo, Fang Li, and Michael K. Ng. A fast $\ell 1$-TV algorithm for image restoration. *SIAM J. Sci. Comput.*, 31(3):2322–2341, 2009. ISSN 1064-8275. doi: 10.1137/080724435.

[50] Ying Jiang, Si Li, and Yuesheng Xu. A higher-order polynomial method for spect reconstruction. *IEEE Transactions on Medical Imaging*, 38(5):1271–1283, 2019. doi: 10.1109/TMI.2018.2881919.

[51] Konrad Knopp. *Theory and application of infinite series.* Courier Corporation, 1944. ISBN 177-288-289-5.

[52] M. A. Krasnosel'skiĭ. Two remarks on the method of successive approximations. *Uspehi Mat. Nauk (N.S.)*, 10(1(63)):123–127, 1955. ISSN 0042-1316. url: http://mi.mathnet.ru/umn7954.

[53] Pierre Simon Laplace. Sur quelques points du système du monde. *Mémoires de l'Académie royale des Sciences de Paris*, page 553, 1789. url: https://gallica.bnf.fr/ark:/12148/bpt6k77599c/f482.

[54] Adrien Marie Legendre. *Nouvelles méthodes pour la détermination des orbites des comètes.* Chez Firmin Didot, 1806. url: https://archive.org/details/nouvellesmthode00legegoog.

[55] Qia Li and Na Zhang. Fast proximity-gradient algorithms for structured convex optimization problems. *Appl. Comput. Harmon. Anal.*, 41(2):491–517, 2016. ISSN 1063-5203. doi: 10.1016/j.acha.2015.11.004.

[56] Qia Li, Lixin Shen, Yuesheng Xu, and Na Zhang. Multi-step fixed-point proximity algorithms for solving a class of optimization problems arising from image processing. *Adv. Comput. Math.*, 41(2):387–422, 2015. ISSN 1019-7168. doi: 10.1007/s10444-014-9363-2.

[57] Zheng Li, Guohui Song, and Yuesheng Xu. A fixed-point proximity approach to solving the support vector regression with the group lasso regularization. *Int. J. Numer. Anal. Model.*, 15(1-2):154–169, 2018. ISSN 1705-5105. url: https://global-sci.org/intro/article_detail/ijnam/10561.html.

[58] Zheng Li, Guohui Song, and Yuesheng Xu. A two-step fixed-point proximity algorithm for a class of non-differentiable optimization models in machine learning. *J. Sci. Comput.*, 81(2):923–940, 2019. ISSN 0885-7474. doi: 10.1007/s10915-019-01045-7.

[59] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Convergence rates with inexact non-expansive operators. *Math. Program.*, 159(1-2, Ser. A):403–434, 2016. ISSN 0025-5610. doi: 10.1007/s10107-015-0964-4.

[60] Yizun Lin, C. Ross Schmidtlein, Qia Li, Si Li, and Yuesheng Xu. A krasnoselskii-mann algorithm with an improved em preconditioner for pet image reconstruction. *IEEE Transactions on Medical Imaging*, 38(9):2114–2126, 2019. doi: 10.1109/TMI.2019.2898271.

[61] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. *Advances in Neural Information Processing Systems*, 24:612–620, 2011. url: https://papers.nips.cc/paper/2011/hash/18997733ec258a9fcaf239cc55d53363-Abstract.html.

[62] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979. ISSN 0036-1429. doi: 10.1137/0716071.

[63] Xiaoxia Liu. *Implicit Fixed-point Proximity Framework for Optimization Problems and Its Applications*. Ph.D. thesis, Syracuse University, 2018. url: https://surface.syr.edu/etd/897.

[64] Yuzhen Liu, Lixin Shen, Yuesheng Xu, and Hongqi Yang. A collocation method solving integral equation models for image restoration. *J. Integral Equations Appl.*, 28(2):263–307, 2016. ISSN 0897-3962. doi: 10.1216/JIE-2016-28-2-263.

[65] Wenting Long, Yao Lu, Lixin Shen, and Yuesheng Xu. High-resolution image reconstruction: an env_{ℓ^1}/TV model and a fixed-point proximity algorithm. *Int. J. Numer. Anal. Model.*, 14(2):255–282, 2017. ISSN 1705-5105. doi: 10.1007/jhep02(2017)117.

[66] Yao Lu, Lixin Shen, and Yuesheng Xu. Multi-parameter regularization methods for high-resolution image reconstruction with displacement errors. *IEEE Trans. Circuits Syst. I. Regul. Pap.*, 54(8):1788–1799, 2007. ISSN 1549-8328. doi: 10.1109/TCSI.2007.902535.

[67] Yao Lu, Lixin Shen, and Yuesheng Xu. Integral equation models for image restoration: high accuracy methods and fast algorithms. *Inverse Problems*, 26(4):045006, 32, 2010. ISSN 0266-5611. doi: 10.1088/0266-5611/26/4/045006.

[68] Nancy A. Lynch. *Distributed algorithms*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, San Francisco, CA, 1996. ISBN 1-55860-348-4. doi: 10.1108/IMDS-01-2014-0013.

[69] Shiqian Ma, Wotao Yin, Yin Zhang, and Amit Chakraborty. An efficient algorithm for compressed mr imaging using total variation and wavelets. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. doi: 10.1109/CVPR.2008.4587391.

[70] W. Robert Mann. Mean value methods in iteration. *Proc. Amer. Math. Soc.*, 4:506–510, 1953. ISSN 0002-9939. doi: 10.2307/2032162.

[71] Bernard Martinet. Brève communication. Régularisation d'inéquations variationnelles par approximations successives. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 4(R3):154–158, 1970. url: http://www.numdam.org/item/M2AN_1970__4_3_154_0/.

[72] Bernard Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox. *C. R. Acad. Sci. Paris Sér. A-B*, 274:A163–A165, 1972. ISSN 0151-0509. url: https://gallica.bnf.fr/ark:/12148/bpt6k5620422z/f37.item.

[73] Charles A. Micchelli, Lixin Shen, and Yuesheng Xu. Proximity algorithms for image models: denoising. *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data*, 27(4):045009, 30, 2011. ISSN 0266-5611. doi: 10.1088/0266-5611/27/4/045009.

[74] Charles A. Micchelli, Lixin Shen, Yuesheng Xu, and Xueying Zeng. Proximity algorithms for the L1/TV image denoising model. *Advances in Computational Mathematics*, 38(2):401–426, 2013. ISSN 1019-7168. doi: 10.1007/s10444-011-9243-y.

[75] Jean Jacques Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965. doi: 10.24033/bsmf.1625.

[76] Isaac Newton. *Philosophiæ Naturalis Principia Mathematica*. Jussu Societatis Regiæ ac Typis Josephi Streater. Prostat apud plures Bibliopolas, 1687. url: `https://archive.org/details/philosophiaenatu28233gut`.

[77] Mila Nikolova. A variational approach to remove outliers and impulse noise. volume 20, pages 99–120. 2004. doi: 10.1023/B:JMIV.0000011920.58935.9c. Special issue on mathematics and image analysis.

[78] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.*, 4(2):460–489, 2005. ISSN 1540-3459. doi: 10.1137/040605412.

[79] Victor Pereyra. Accelerating the convergence of discretization algorithms. *SIAM J. Numer. Anal.*, 4:508–533, 1967. ISSN 0036-1429. doi: 10.1137/0704046.

[80] Emile Picard. Mémoire sur la théorie des équations aux dérivées partielles et la méthode des approximations successives. *Journal de Mathématiques Pures et Appliquées*, 6:145–210, 1890. url: `http://eudml.org/doc/235808`.

[81] Thomas Pock and Antonin Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *2011 International Conference on Computer Vision*, pages 1762–1769, 2011. doi: 10.1109/ICCV.2011.6126441.

[82] Lee C. Potter and K. S. Arun. A dual approach to linear inverse problems with convex constraints. *SIAM Journal on Control and Optimization*, 31(4):1080–1092, 1993. doi: 10.1137/0331049.

[83] Simeon Reich. Weak convergence theorems for nonexpansive mappings in Banach spaces. *J. Math. Anal. Appl.*, 67(2):274–276, 1979. ISSN 0022-247X. doi: 10.1016/0022-247X(79)90024-6.

[84] Jin Ren and Yuesheng Xu. Inexact fixed-point iterations and inexact implicit fixed-point proximity algorithms for convex optimization. preprint.

[85] Herbert Robbins and Sutton Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400 – 407, 1951. doi: 10.1214/aoms/1177729586.

[86] R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976. doi: 10.1137/0314056.

[87] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1998. ISBN 3-540-62772-3. doi: 10.1007/978-3-642-02431-3.

[88] Arie Rond, Raja Giryes, and Michael Elad. Poisson inverse problems by the plug-and-play scheme. *Journal of Visual Communication and Image Representation*, 41:96–108, 2016. ISSN 1047-3203. doi: doi.org/10.1016/j.jvcir.2016.09.009.

[89] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. volume 60, pages 259–268. 1992. doi: 10.1016/0167-2789(92)90242-F. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).

[90] Helmut Schaefer. Über die methode sukzessiver approximationen. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 59:131–140, 1957. url: `http://eudml.org/doc/146424`.

[91] Jssai Schur. Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 140:1–28, 1911. doi: 10.1515/crll.1911.140.1.

[92] Simon Setzer. Split bregman algorithm, douglas-rachford splitting and frame shrinkage. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 464–476. Springer, 2009. doi: 10.1007/978-3-642-02256-2_39.

[93] Jianing Shi, Wotao Yin, Stanley Osher, and Paul Sajda. A fast hybrid algorithm for large-scale ℓ_1-regularized logistic regression. *J. Mach. Learn. Res.*, 11:713–741, 2010. ISSN 1532-4435. url: `https://dl.acm.org/doi/10.5555/1756006.1756029`.

[94] Gerard Tel. *Introduction to distributed algorithms*. Cambridge University Press, Cambridge, second edition, 2000. ISBN 0-521-79483-8. doi: 10.1017/CBO9781139168724.

[95] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996. ISSN 0035-9246. url: `https://www.jstor.org/stable/2346178`.

[96] John N. Tsitsiklis, Dimitri P. Bertsekas, and Michael Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. Automat. Control*, 31(9):803–812, 1986. ISSN 0018-9286. doi: 10.1109/TAC.1986.1104412.

[97] John Nikolas Tsitsiklis. *Problems in decentralized decision making and computation*. PhD thesis, Massachusetts Institute of Technology, 1985. url: `http://hdl.handle.net/1721.1/15254`.

[98] Hirofumi Uzawa. Walras' tâtonnement in the theory of exchange. *The Review of Economic Studies*, 27(3):182–194, 1960. ISSN 00346527, 1467937X. url: `http://www.jstor.org/stable/2296080`.

[99] Vladimir N. Vapnik. *The nature of statistical learning theory.* Springer-Verlag, New York, 1995. ISBN 0-387-94559-8. doi: 10.1007/978-1-4757-2440-0.

[100] Richard S. Varga. Factorization and normalized iterative methods. In *Boundary problems in differential equations*, pages 121–142. Univ. Wisconsin Press, Madison, Wis., 1959. url: https://www.osti.gov/biblio/4231859.

[101] Rui Wang and Yuesheng Xu. Functional reproducing kernel Hilbert spaces for non-point-evaluation functional data. *Appl. Comput. Harmon. Anal.*, 46(3):569–623, 2019. ISSN 1063-5203. doi: 10.1016/j.acha.2017.07.003.

[102] Rui Wang and Yuesheng Xu. Representer theorems in Banach spaces: minimum norm interpolation, regularized learning and semi-discrete inverse problems. *J. Mach. Learn. Res.*, 22:Paper No. 225, 65, 2021. ISSN 1532-4435. url: http://jmlr.org/papers/v22/20-751.html.

[103] Richard G. Wiley. On an iterative technique for recovery of bandlimited signals. *Proceedings of the IEEE*, 66(4):522–523, 1978. doi: 10.1109/PROC.1978.10951.

[104] Hong Xu. A variable krasnosel'skiĭ-mann algorithm and the multiple-set split feasibility problem. *Inverse Problems*, 22:2021, 10 2006. doi: 10.1088/0266-5611/22/6/007.

[105] Yuesheng Xu and Qi Ye. Generalized Mercer kernels and reproducing kernel Banach spaces. *Mem. Amer. Math. Soc.*, 258(1243):vi+122, 2019. ISSN 0065-9266. doi: 10.1090/memo/1243.

[106] Junfeng Yang and Xiaoming Yuan. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. *Math. Comp.*, 82(281):301–329, 2013. ISSN 0025-5718. doi: 10.1090/S0025-5718-2012-02598-1.

[107] Qingzhi Yang and Jinling Zhao. Generalized KM theorems and their applications. *Inverse Problems*, 22(3):833–844, 2006. ISSN 0266-5611. doi: 10.1088/0266-5611/22/3/006.

[108] Dan C. Youla and Heywood Webb. Image restoration by the method of convex projections: Part 1 – theory. *IEEE Transactions on Medical Imaging*, 1(2):81–94, 1982. doi: 10.1109/TMI.1982.4307555.

[109] Dante C. Youla. Generalized image restoration by the method of alternating orthogonal projections. *IEEE Trans. Circuits and Systems*, 25(9):694–702, 1978. ISSN 0098-4094. doi: 10.1109/TCS.1978.1084541. Special issue on the mathematical foundations of system theory.

[110] Haizhang Zhang, Yuesheng Xu, and Jun Zhang. Reproducing kernel Banach spaces for machine learning. *J. Mach. Learn. Res.*, 10:2741–2775, 2009. ISSN 1532-4435. doi: 10.1109/IJCNN.2009.5179093.

[111] Xiaoqun Zhang, Martin Burger, and Stanley Osher. A unified primal-dual algorithm framework based on Bregman iteration. *Journal of Scientific Computing*, 46(1):20–46, 2011. ISSN 0885-7474. doi: 10.1007/s10915-010-9408-8.

VITA

Jin Ren | August 2022
Department of Mathematics & Statistics | jren007@odu.edu
Old Dominion University | Norfolk, VA 23529

Education

- ☐ Ph.D. in Computational and Applied Mathematics, Old Dominion University, Norfolk, VA. Dissertation · *Inexact Fixed-Point Proximity Algorithms for Nonsmooth Convex Optimization.* August 2022 [Anticipated].
- ☐ B.S. in Mathematics and Applied Mathematics, Sun Yat-Sen University, Guangzhou, China. Thesis · *Two Generalizations of KM Iteration and its Application to Parameter-Varying Fixed-Point Proximity Algorithm.* Honors Graduate. Outstanding Graduate Thesis Award. June 2016.

Professional Experience

- ☐ Graduate Research Assistantship, Old Dominion University, 2021–2022.
- ☐ Graduate Teaching Assistantship, Old Dominion University, 2020–2021.
- ☐ Dominion Graduate Scholarship, Old Dominion University, 2018–2020.

Publication

- ☐ Wenyu Hu, Yao Lu, and Jin Ren. A fixed-point proximity algorithm for recovering low-rank components from incomplete observation data with application to motion capture data refinement. *J. Comput. Appl. Math.*, 410:Paper No. 114224, 23, 2022. ISSN 0377-0427. doi: 10.1016/j.cam.2022.114224

Typeset using XƎLaTeX.