UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
Departamento de Matemática
Santiago-Chile

# Alternating and randomized projections on convex optimization

Memoria presentada por:
**Cristian Jesús Vega Cereño**

*Como requisito parcial para optar al título profesional Ingeniero Civil Matemático*

*Profesor Guía:*
**Luis Briceño Arias**

Febrero 2020

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

Departamento de Matemática

Santiago-Chile


# Alternating and randomized projections on convex optimization

Memoria presentada por:


**Cristian Jesús Vega Cereño**


*Como requisito parcial para optar al título profesional Ingeniero Civil Matemático*

*Profesor Guía:*
Luis Briceño Arias

*Examinadores:*
P. Gajardo
J. Deride
E. Cerpa
J. Peypouquet.

Febrero, 2020

TÍTULO DE LA MEMORIA:
Alternating and randomized projections on convex optimization.

AUTOR: Cristian Jesús Vega Cereño.

TRABAJO DE MEMORIA, presentado como requisito parcial para optar al título profesional Ingeniero Civil Matemático de la Universidad Técnica Federico Santa María.

COMISIÓN EVALUADORA:

| Integrantes | Firma |
| --- | --- |
| Luis Briceño Arias<br>Universidad Técnica Federico Santa María, Chile. | ______ |
| Pedro Gajardo<br>Universidad Técnica Federico Santa María, Chile. | ______ |
| Julio Deride<br>Universidad Técnica Federico Santa María, Chile. | ______ |
| Eduardo Cerpa<br>Pontificia Universidad Católica de Chile, Chile. | ______ |
| Juan Peypouquet<br>Universidad de Chile, Chile. | ______ |

Santiago, Febrero 2020.

**Resumen**

En este trabajo, proponemos dos enfoques numéricos para resolver problemas primales-duales de optimización convexa con restricciones. Las restricciones del problema están representadas por la intersección de un número finito de conjuntos convexos cerrados sobre los cuales los algoritmos propuestos proyectan de manera alternada y/o aleatoria. El primer algoritmo incluye un paso de activación aleatorio sobre un esquema de proyección cíclico, mientras que el segundo elige un elemento aleatorio del conjunto de operadores de proyección. La convergencia casi segura de ambos algoritmos se deriva de las propiedades de las sucesiones estocásticas Quasi-Fejér.

Como casos especiales de los algoritmos propuestos, recuperamos varios algoritmos primales-duales en la literatura y algoritmos clásicos para resolver problemas de factibilidad de conjuntos convexos, como proyecciones cíclicas, Kaczmarz y Kaczmarz aleatorio. Finalmente, probamos ambos algoritmos en un problema de expansión de capacidad de arco en una red de transporte. El problema puede formularse como un problema primal-dual de optimización convexa con restricciones. Luego comparamos la eficiencia de diferentes esquemas de proyección alternada / aleatoria propuestos en este trabajo con el algoritmo primal-dual sin ninguna proyección. Todos los algoritmos propuestos que incluyen una proyección mejoran considerablemente el tiempo de ejecución y el número de iteraciones. En el caso de los algoritmos que incluyen proyecciones aleatorias y alternada obtenemos hasta un 31% y 35% de mejora en el tiempo de ejecución promedio, respectivamente, en los ejemplos de dimensiones superiores.

**Palabras clave:** Optimización convexa con restricciones, Algoritmo de minimización, Algoritmo Primal-dual, Sucesiones estocásticas Quasi-Fejér, Algoritmo de Kaczmarz aleatorio.

# Agradecimientos

En primer lugar, me gustaría expresar mi más sincero agradecimiento a mi profesor guía, Luis Briceño Arias, por el continuo apoyo de mi trabajo, por su paciencia, motivación e inmenso conocimiento. Su guía me ayudó en todo el tiempo de investigación y en la redacción de este trabajo. También me gustaría reconocer a Julio Deride, el segundo lector de esta memoria, del que estoy agradecido por sus valiosos comentarios sobre este trabajo.

También me gustaría agradecer al Departamento de Matemáticas de la Universidad Técnica Federico Santa María, especialmente a Erwin Hernández e Isabel Flores, su estilo de enseñanza y su entusiasmo me impresionaron mucho y siempre he tenido recuerdos positivos de sus clases conmigo. Deseo expresar mi más profundo agradecimiento a mis compañeros, especialmente Sergio López, Fernando Roldán, Gonzalo Arias, Gustavo Ulloa y Hugo Parada quienes han sido un apoyo personal y profesional durante el tiempo que pasé en la Universidad.

Me gustaría agradecer a mi padre, abuela y a mi novia Valeska por brindarme un apoyo incondicional durante mis años de estudio y el proceso de investigación y redacción de este trabajo. También estoy agradecido con mis otros familiares y amigos, Omar y Jeremías, que me han apoyado en el camino.

Finalmente, aprovecho esta oportunidad para expresar mi gratitud a todos los que me apoyaron a lo largo de la carrera. Este logro no hubiera sido posible sin ellos.

# Contents

# Chapter 1

# Introduction

## 1.1 State of the art and context

The main goal of this work is study and propose an efficient algorithm to solve the following convex minimization problem.

**Problem 1.1** Let $\mathcal{H}$ and $\mathcal{G}$ be separable real Hilbert spaces, endowed by the scalar product $\langle \cdot \mid \cdot \rangle$ and the associated norm $\| \cdot \|$. Let $f : \mathcal{H} \mapsto ]-\infty, +\infty]$ and $g : \mathcal{G} \mapsto ]-\infty, +\infty]$ be proper lower semicontinuous convex functions, let $h : \mathcal{H} \to \mathbb{R}$ be convex and differentiable with $\mu^{-1}$- Lipschitzian gradient, for some $\mu \in ]0, +\infty[$, and let $L : \mathcal{H} \to \mathcal{G}$ be a nonzero bounded linear operator. Let $C := \bigcap_{i=1}^{m} C_i \neq \emptyset$, where, for every $i$ in $\{1, ..., m\}$, $C_i$ is a nonempty closed convex subset of $\mathcal{H}$. Consider the primal problem

$$\text{find } x \in C \cap \operatorname{argmin}_{x \in \mathcal{H}}(f(x) + g(Lx) + h(x)) \tag{1.1}$$

and the dual problem

$$\text{find } u \in \operatorname{argmin}_{u \in \mathcal{G}}(g^*(u) + (f + h)^*(-L^* u)), \tag{1.2}$$

where $G^*$ denote the conjugate function of $G$ and $L^*$ is the adjoint of $L$.

Problem 1.1 arises in several areas such as image recovery [6, 10, 20], partial diferential equations [1, 17], signal processing [11, 14] and arc capacity expansion over a directed graph in a stochastic context [9]. In [8] a primal-dual algorithm for solving Problem 1.1 is proposed, in the case when $h = 0$ and $C = \mathcal{H}$. In [16, 24] the previous algorithm is extended to the case $h \neq 0$, while in [7], Problem 1.1 is solved in its all generality, by including a deterministic projection onto $C$. In several cases the projection operator is not easy to compute, and the main goal of this work is to provide eficcient algorithms which implement projections onto $C_1, \ldots, C_m$ and generate sequences converging to a solution to Problem 1.1.

As a particular instances of Problem 1.1, consider the case when $f = g = L = h = 0$, $\mathcal{H} = \mathbb{R}^n$ and $C = \{x \in \mathcal{H} | \ Rx = b\}$, where $R$ is full rank $m \times n$ real matrix, such that $m \leq n$, and $b \in \mathbb{R}^m$. In this case Problem 1.1 reduces to the problem of finding $x \in C$ and one possibility to solve the system is to use the Kaczmarz method [19], which performs cyclic projections onto hyperplanes defined by linear equations $r_i^\top x = b_i$, for all $i \in \{1, ..., m\}$, where $r_i \in \mathbb{R}^m$ is the *ith* line of matrix $R$ . The algorithm converges to a feasible solution of the problem in the consistent case. In the case when the linear system is inconsistent we refer the reader to [18]. In [23] a randomized version of the Kaczmarz method for consistent and overdetermined linear systems with an expected exponential convergence rate is proposed.

The objective of this work is to propose new algorithms with alternating and random projections to solve Problem 1.1 in its whole generality and verify its numerical performance of the algorithms in capacity expansion problems in transport networks.

As a consequence of the results of this work, we obtain generalizations of Kaczmarz method [19], Randomized Kaczmarz [23], and several deterministic methods for the convex feasibility problem [3]. On the other hand, we generalize primal-dual methods [7, 8] by including alternating and randomized projections onto a priori knowledge of the solutions.

This document is organized as follows. In Chapter 1 we introduce some notation and a background in Convex Optimization and Probability in Hilbert spaces. In the next section we present an equivalent formulation and preliminary results. In chapter 2 and 3 we propose a primal-dual method with random binary projection algorithm and randomized Kaczmarz version. We prove convergence results by using Stochastic Quasi-Fejér sequences as in [13]. Finally we verify the performance of the algorithms in the example of arc capacity expansion problem in transport networks.

## 1.2 Notation

The identity operator on $\mathcal{H}$ is denoted by Id and $\rightharpoonup$ and $\to$ denote, weak and strong convergence in $\mathcal{H}$, respectively. The set of weak sequential cluster points of a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ is denoted by $\mathfrak{W}\,(x_n)_{n\in\mathbb{N}}$. The adjoint of a linear bounded operator $L : \mathcal{H} \mapsto \mathcal{G}$ is denoted by $L^*$. The projector operator onto a nonempty closed convex set $C \subset \mathcal{H}$ is denoted by $P_C : x \in \mathcal{H} \to \operatorname{argmin}_{y\in C} \|y - x\|$ and its normal cone operator is defined by

$$N_C : \mathcal{H} \rightrightarrows \mathcal{H} : x \mapsto \begin{cases} \{u \in \mathcal{H} \mid (\forall y \in C)\ \langle u \mid y - x\rangle \leq 0\} & \text{if } x \in C; \\ \varnothing & \text{if } x \notin C. \end{cases} \tag{1.3}$$

Let $C$ be a nonempty convex subset of $\mathcal{H}$. The strong relative interior of $C$ is

$$\operatorname{sri} C = \{x \in C \mid \mathbb{R}_{++}(C - x) = \overline{\operatorname{span}}(C - x)\}\,, \tag{1.4}$$

where $R_{++}C_1 = \{\lambda y \mid (\lambda > 0) \wedge (y \in C_1)\}$ and $\overline{\operatorname{span}}(C_1)$ is the smallest closed linear subspace of $\mathcal{H}$ containing $C_1$.

Given $\alpha \in ]0, 1[$, an operator $T : \mathcal{H} \to \mathcal{H}$ is $\alpha$-averaged nonexpansive iff,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \ \|Tx - Ty\|^2 \leq \|x - y\|^2 - \frac{1-\alpha}{\alpha}\|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2, \tag{1.5}$$

and $T$ is firmly nonexpansive if and only if it is $\frac{1}{2}$-averaged.

Let $M : \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued operator. The domain of $M$ is $\operatorname{dom} M := \{x \in \mathcal{H} \mid Mx \neq \varnothing\}$, we denote by $\operatorname{ran}(M) := \{u \in \mathcal{H} \mid (\exists x \in \mathcal{H})\ u \in Mx\}$ the range of $M$ and the graph of $M$ is $\operatorname{gra} M := \{(x, u) \in \mathcal{H}^2 \mid u \in Mx\}$. The inverse $M^{-1}$ of $M$ is defined via the equivalences $\big(\forall (x, u) \in \mathcal{H}^2\big)\ x \in M^{-1}u \Leftrightarrow u \in Mx$. Given $\rho \geq 0$, $M$ is $\rho$-strongly monotone iff,

$$(\forall (x, u) \in \operatorname{gra}(M))(\forall (y, v) \in \operatorname{gra}(M)) \quad \langle x - y \mid u - v\rangle \geq \rho \|x - y\|^2.$$

$M$ is $\rho$-cocoercive iff $M^{-1}$ is $\rho$-strongly monotone, $M$ is monotone iff it is $\rho$-strongly monotone with $\rho = 0$, and it is maximally monotone iff its graph is maximal, in the sense of inclusions in $\mathcal{H} \times \mathcal{H}$, among the graphs of monotone operators. The resolvent of $M$ is denoted by $J_M = (\operatorname{Id} + M)^{-1}$, where Id is the identity operator. If $M$ is maximally monotone, then $J_M$ is single-valued and firmly nonexpansive operator, with $\operatorname{dom} J_M = \mathcal{H}$. We denote by $\Gamma_0(\mathcal{H})$ the set of proper, lower semicontinuous and convex functions from $\mathcal{H} \mapsto \mathbb{R} \cup \{+\infty\}$. The subdifferential of $f \in \Gamma_0(\mathcal{H})$ is the maximal monotone operator

$$\partial f : \mathcal{H} \rightrightarrows \mathcal{H} : x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ f(x) + \langle y - x \mid u\rangle \leq f(y)\} \tag{1.6}$$

and, if $f$ is Gâteaux differentiable in $x$, then $\partial f(x) = \{\nabla f(x)\}$. We have $(\partial f)^{-1} = \partial f^*$, where $f^* \in \Gamma_0(\mathcal{H})$ is the conjugate function of $f \in \Gamma_0(\mathcal{H})$ defined by $f^* : u \mapsto \sup_{x \in \mathcal{H}}(\langle x \mid u\rangle - f(x))$. The proximal operator of $f \in \Gamma_0(\mathcal{H})$ is

$$\operatorname{prox}_f : \mathcal{H} \to \mathcal{H} : x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} f(y) + \frac{1}{2}\|x - y\|^2 \tag{1.7}$$

and we have $J_{\partial f} = \operatorname{prox}_f$. Moreover, if $C \subset \mathcal{H}$ is a nonempty convex closed subset, then $\delta_C \in \Gamma_0(\mathcal{H})$, $N_C = \partial \delta_C$, and $J_{N_C} = P_C$, where

$$\delta_C(x) = \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{if } x \notin C. \end{cases} \tag{1.8}$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The space of all random variables $z$ with values in $\mathcal{H}$ such that $\|z\|$ is integrable is denoted by $L^1(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{H})$. Given a $\sigma$-algebra $\mathcal{E}$ of $\Omega$, $x \in L^1(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{H})$, and $y \in L^1(\Omega, \mathcal{F}, \mathbb{P}; \mathcal{H})$, $y$ is the conditional expectation of $x$ with respect to $\mathcal{E}$ iff $(\forall E \in \mathcal{E})\ \int_E x d\mathbb{P} = \int_E y d\mathbb{P}$, in this case we write $y = \mathbb{E}(x \,|\, \mathcal{E})$. The characteristic function on $D \subset \Omega$ is denote by $\mathbb{1}_D$, which is 1 in $D$ and 0 otherwise. An $\mathcal{H}$-valued random variable is a measurable map $x : (\Omega, \mathcal{F}) \to (\mathcal{H}, \mathcal{B})$, where $\mathcal{B}$ is the Borel $\sigma$-algebra. The $\sigma$-algebra generated by a family $\Phi$ of random variables is denoted by $\sigma(\Phi)$. Let $\mathscr{F} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ be a sequence of sub-sigma algebras of $\mathcal{F}$ such that $(\forall n \in \mathbb{N}) \quad \mathcal{F}_n \subset \mathcal{F}_{n+1}$. We denote by $\ell_+(\mathscr{F})$ the set of sequences of $[0, \infty)$-valued random variables $(\xi_n)_{n \in \mathbb{N}}$ such that, for every $n \in \mathbb{N}$, $\xi_n$ is $\mathcal{F}_n$-measurable. We set

$$(\forall p \in ]0, \infty[) \quad \ell_+^p(\mathscr{F}) := \left\{ (\xi_n)_{n \in \mathbb{N}} \in \ell_+(\mathscr{F}) \,\middle|\, \sum_{n \in \mathbb{N}} \xi_n^p < \infty \quad P - a.s. \right\}. \tag{1.9}$$

## 1.3 Alternative formulation and Preliminaries

Assume that the following qualification condition is satisfied

$$0 \in \operatorname{sri}\left(L\left(\operatorname{dom} f\right) - \operatorname{dom} g\right). \tag{1.10}$$

Then, applying [3, Theorem 16.47], we have the following equivalent formulation for the Problem 1.1.

**Problem 1.2** Consider the setting of Problem 1.1. The problem can be restated as solving the primal-dual inclusions

$$\text{find} \quad x \in C = \bigcap_{i=1}^{m} C_i \quad \text{such that} \quad 0 \in \partial f(x) + L^*\partial g(Lx) + \nabla h(x), \tag{$\mathcal{P}_0$}$$

together with the dual inclusion

$$\text{find} \quad u \in \mathcal{G} \quad \text{such that} \quad (\exists x \in C) \begin{cases} 0 \in \partial f(x) + L^*u + \nabla h(x) \\ \\ 0 \in \partial g^*(u) - Lx, \end{cases} \tag{$\mathcal{D}_0$}$$

under the assumption that solutions exist. We denote by $Z \subset \mathcal{H}\times\mathcal{G}$ the set of primal-dual solutions.

The following proposition give us some technical inequalities and properties used in the convergence of algorithm proposed in the chapters 2 and 3.

**Proposition 1.3** *Consider the setting of Problem 1.2. Let $\tau \in ]0, 2\mu[$, let $\gamma > 0$, and let $(x^0, \overline{x}^0, u^0) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G}$ be such that $x^0 = \overline{x}^0$. Moreover, let $(\epsilon_k)_{k\in\mathbb{N}}$ be a sequence of independent I-valued random variables, where I is a finite set of $\mathbb{N} \cup \{0\}$, and let $(G_k(p^{k+1}, \epsilon_{k+1}))_{k\in\mathbb{N}}$ be a sequence of $\mathcal{H}$-valued random variables. Consider the following routine*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = \mathrm{prox}_{\gamma g^*}(u^k + \gamma L\bar{x}^k) \\ p^{k+1} = \mathrm{prox}_{\tau f}\left(x^k - \tau(L^*u^{k+1} + \nabla h(x^k))\right) \\ x^{k+1} = G_{k+1}(p^{k+1}, \epsilon_{k+1}) \\ \bar{x}^{k+1} = x^{k+1} + p^{k+1} - x^k. \end{array} \right. \tag{1.11}$$

*Then the following hold:*

(i) *For every $k \geq 1$ and $(\hat{x}, \hat{u}) \in Z$, we have*

$$\begin{aligned} \frac{\|x^k - \hat{x}\|^2}{\tau} + \frac{\|u^k - \hat{u}\|^2}{\gamma} \geq \frac{\|p^{k+1} - \hat{x}\|^2}{\tau} &+ \left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|x^k - p^{k+1}\|^2 + \frac{\|u^{k+1} - \hat{u}\|^2}{\gamma} + \frac{\|u^{k+1} - u^k\|^2}{\gamma} \\ &+ 2\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\rangle - 2\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle \\ &- 2\|L\|\|p^k - x^{k-1}\|\|u^{k+1} - u^k\|. \end{aligned} \tag{1.12}$$

(ii) *Suppose that, for every $(\hat{x}, \hat{u}) \in Z$, the sequence $\left(\frac{\|x^k-\hat{x}\|^2}{\tau} + \frac{\|u^k-\hat{u}\|^2}{\gamma}\right)_{k\in\mathbb{N}}$ converges P-a.s. to a $[0, \infty[$-valued random variable. Then there exists $\Omega_Z$ such that $\mathbb{P}(\Omega_Z) = 1$ and, for every $\omega \in \Omega_Z$ and for every $(\hat{x}, \hat{u}) \in Z$, $\left(\frac{\|x^k(w)-\hat{x}\|^2}{\tau} + \frac{\|u^k(w)-\hat{u}\|^2}{\gamma}\right)_{k\in\mathbb{N}}$ converges.*

*Proof.*
(i) Fix $k \in \mathbb{N}$ and $(\hat{x}, \hat{u}) \in Z$. It follows from (1.11) that

$$\begin{aligned} \frac{x^k - p^{k+1}}{\tau} - L^*u^{k+1} - \nabla h(x^k) &\in \partial f(p^{k+1}) \\ \frac{u^k - u^{k+1}}{\gamma} + L(x^k + p^k - x^{k-1}) &\in \partial g^*(u^{k+1}). \end{aligned} \tag{1.13}$$

Since $f \in \Gamma_0(\mathcal{H})$ and $g^* \in \Gamma_0(\mathcal{G})$ [3, Proposition 13.13], $\partial f$ and $\partial g^*$ are maximally monotone operators [3, Theorem 20.25] then we have

$$\left\langle \frac{x^k - p^{k+1}}{\tau} - L^*(u^{k+1} - \hat{u}) \;\middle|\; p^{k+1} - \hat{x} \right\rangle + \left\langle \frac{u^k - u^{k+1}}{\gamma} + L(x^k + p^k - x^{k-1} - \hat{x}) \;\middle|\; u^{k+1} - \hat{u} \right\rangle - \left\langle \nabla h(x^k) - \nabla h(\hat{x}) \;\middle|\; p^{k+1} - \hat{x} \right\rangle \geq 0, \tag{1.14}$$

multiplying (1.14) by 2 and from [3, Lemma 2.12 (i)] we obtain

$$\begin{aligned} \frac{\|x^k - \hat{x}\|^2}{\tau} + \frac{\|u^k - \hat{u}\|^2}{\gamma} \geq & \frac{1}{\tau}\|x^k - p^{k+1}\|^2 + \frac{\|p^{k+1} - \hat{x}\|^2}{\tau} + \frac{\|u^{k+1} - u^k\|^2}{\gamma} + \frac{\|u^{k+1} - \hat{u}\|^2}{\gamma} \\ & + 2\langle L(p^{k+1} - \hat{x}) \mid u^{k+1} - \hat{u}\rangle - 2\langle L(x^k + p^k - x^{k-1} - \hat{x}) \mid u^{k+1} - \hat{u}\rangle \\ & + 2\left\langle \nabla h(x^k) - \nabla h(\hat{x}) \;\middle|\; p^{k+1} - \hat{x} \right\rangle. \end{aligned} \tag{1.15}$$

On the other hand, we have

$$\begin{aligned} & \langle L(p^{k+1} - \hat{x}) \mid u^{k+1} - \hat{u}\rangle - \langle L(x^k + p^k - x^{k-1} - \hat{x}) \mid u^{k+1} - \hat{u}\rangle \\ & = \langle L(p^{k+1} - \hat{x}) \mid u^{k+1} - \hat{u}\rangle - \langle L(x^k - \hat{x}) \mid u^{k+1} - \hat{u}\rangle - \langle L(p^k - x^{k-1}) \mid u^{k+1} - \hat{u}\rangle \\ & = \langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\rangle - \langle L(p^k - x^{k-1}) \mid u^{k+1} - \hat{u}\rangle \\ & = \langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\rangle - \langle L(p^k - x^{k-1}) \mid u^{k+1} - u^k\rangle - \langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle \\ & \geq \langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\rangle - \|L\|\|p^k - x^{k-1}\|\|u^{k+1} - u^k\| - \langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle. \end{aligned} \tag{1.16}$$

Finally since $h$ has $\mu^{-1}$- Lipschitzian gradient then $\nabla h$ is $\mu$-cocoercive [3, Corollary 18.17], and $ab \leq \beta a^2 + \frac{b^2}{4\beta}$ yield

$$\begin{aligned} \left\langle \nabla h(x^k) - \nabla h(\hat{x}) \mid p^{k+1} - \hat{x} \right\rangle & = \left\langle \nabla h(x^k) - \nabla h(\hat{x}) \mid p^{k+1} - x^k \right\rangle + \left\langle \nabla h(x^k) - \nabla h(\hat{x}) \mid x^k - \hat{x} \right\rangle \\ & \geq -\|\nabla h(x^k) - \nabla h(\hat{x})\|\|p^{k+1} - x^k\| + \mu\|\nabla h(x^k) - \nabla h(\hat{x})\|^2 \\ & \geq -\frac{\left\|p^{k+1} - x^k\right\|^2}{4\mu}, \end{aligned} \tag{1.17}$$

the result follows replacing (1.16) and (1.17) in (1.15).

(ii) We define a norm $|\!|\!| \cdot |\!|\!|_{\tau,\gamma}$ on $\mathcal{H} \times \mathcal{G}$ by

$$\forall (x,u) \in \mathcal{H} \times \mathcal{G} \quad |\!|\!|(x,u)|\!|\!|_{\tau,\gamma} = \sqrt{\frac{\|x\|^2}{\tau} + \frac{\|u\|^2}{\gamma}}. \tag{1.18}$$

Now since $\mathcal{H}$ and $\mathcal{G}$ are separable then $\mathcal{H} \times \mathcal{G}$ and $Z$ are separable and, by definition, there exists $F \subset \mathcal{H} \times \mathcal{G}$ countable such that $\overline{F} = Z$. Moreover, for every $(x,u) \in Z$, there exists a set $\Omega_{(x,u)} \in \mathcal{F}$ such that $\mathbb{P}(\Omega_{(x,u)}) = 1$ and, $(|\!|\!|(x^k,u^k) - (x,u)|\!|\!|_{\tau,\gamma})_{k\in\mathbb{N}}$ converges to a $[0,\infty[$-valued random variable $\varphi_{(x,u)} : \Omega_{(x,u)} \to [0,\infty[$. If we define $\Omega_Z = \bigcap_{(x,u)\in F} \Omega_{(x,u)}$, since $F$ is countable, we have that $\mathbb{P}(\Omega_Z) = \mathbb{P}(\bigcap_{(x,u)\in F} \Omega_{(x,u)}) = 1 - \mathbb{P}(\bigcup_{(x,u)\in F} \Omega^c_{(x,u)}) \geq 1 - \sum_{(x,u)\in F} \mathbb{P}(\Omega^c_{(x,u)}) = 1$, which yields $\mathbb{P}(\Omega_Z) = 1$.

We want to prove that, for every $(x,u) \in Z$ and $w \in \Omega_Z$, the sequence $(|||(x^k(w), u^k(w)) - (x,u)|||_{\tau,\gamma})_{k\in\mathbb{N}}$ converges. Since $F$ is dense in $Z$, there exists a sequence $((x_n,u_n))_{n\in\mathbb{N}} \subset F$ such that $(x_n,u_n) \to (x,u)$. Now let $w \in \Omega_Z$. For every $k \in \mathbb{N}$ and $n \in \mathbb{N}$, we have

$$\begin{aligned} -|||(x_n,u_n) - (x,u)|||_{\tau,\gamma} &\le |||(x^k(w), u^k(w)) - (x,u)|||_{\tau,\gamma} - |||(x^k(w), u^k(w)) - (x_n,u_n)|||_{\tau,\gamma} \\ &\le |||(x_n,u_n) - (x,u)|||_{\tau,\gamma}. \end{aligned} \tag{1.19}$$

Therefore, for every $n \in \mathbb{N}$, we obtain

$$\begin{aligned} -|||(x_n,u_n) - (x,u)|||_{\tau,\gamma} &\le \liminf_{k\to\infty} |||(x^k(w), u^k(w)) - (x,u)|||_{\tau,\gamma} - \lim_{k\to\infty} |||(x^k(w), u^k(w)) - (x_n,u_n)|||_{\tau,\gamma} \\ &= \liminf_{k\to\infty} |||(x^k(w), u^k(w)) - (x,u)|||_{\tau,\gamma} - \varphi_{(x_n,u_n)}(w) \\ &\le \limsup_{k\to\infty} |||(x^k(w), u^k(w)) - (x,u)|||_{\tau,\gamma} - \varphi_{(x_n,u_n)}(w) \\ &= \limsup_{k\to\infty} |||(x^k(w), u^k(w)) - (x,u)|||_{\tau,\gamma} - \lim_{k\to\infty} |||(x^k(w), u^k(w)) - (x_n,u_n)|||_{\tau,\gamma} \\ &\le |||(x_n,u_n) - (x,u)|||_{\tau,\gamma}. \end{aligned} \tag{1.20}$$

Therefore, taking the limit as $n \to \infty$, we obtain that $(|||(x^k,u^k) - (x,u)|||_{\tau,\gamma})_{k\in\mathbb{N}}$ converges P-a.s. and

$$(\forall w \in \Omega_Z) \quad \lim_{k\to\infty} |||(x^k(w), u^k(w)) - (x,u)|||_{\tau,\gamma} = \lim_{n\to\infty} \varphi_{(x_n,u_n)}(w), \tag{1.21}$$

which yields the results. □

The following lemma is an especial case of [12, Theorem 3.2] and is the main tool to prove the convergence of Stochastic Quasi-Fejér sequences.

**Lemma 1.4** *[22, Theorem 1] Let $\mathscr{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub-sigma algebras of $\mathcal{F}$ such that $(\forall n \in \mathbb{N})$ $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Let $(a_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{F})$ and $(b_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{F})$ be such that*

$$(\forall n \in \mathbb{N}) \quad \mathbb{E}(a_{n+1} \mid \mathcal{F}_n) + b_n \le a_n \quad P-a.s. \tag{1.22}$$

*Then $(b_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathscr{F})$ and $(a_n)_{n\in\mathbb{N}}$ converges P-a.s. to a $[0,\infty[$-valued random variable.*

# Chapter 2

# Random binary projections in convex optimization

In this chapter, we explore an extension of primal-dual algorithm proposed in [7], which includes random binary alternating projections, i.e., given an order of the sets $(C_i)_{i=1}^m$ onto which projections take place at each iteration, the method "flips a coin" and decide to project or not.

**Theorem 2.1** *Consider the setting of Problem 1.2. Let* $\tau \in ]0, 2\mu[$, *let* $\gamma > 0$, *let* $(x^0, \overline{x}^0, u^0) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G}$ *be such that* $x^0 = \overline{x}^0$, *and let* $(\epsilon_k)_{k\in\mathbb{N}}$ *be a sequence of independent* $\{0,1\}-$*Bernoulli random variables satisfying* $\mathbb{P}(\epsilon_k^{-1}(\{1\})) = \pi_k$. *Let* $(D_k)_{k=1}^{\infty}$ *be a sequence of nonempty closed convex subsets of* $\mathcal{H}$ *and consider the following routine*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = \mathrm{prox}_{\gamma g^*}(u^k + \gamma L\bar{x}^k) \\ p^{k+1} = \mathrm{prox}_{\tau f}(x^k - \tau(L^* u^{k+1} + \nabla h(x^k))) \\ x^{k+1} = p^{k+1} + \epsilon_{k+1}(P_{D_{k+1}} p^{k+1} - p^{k+1}) \\ \bar{x}^{k+1} = x^{k+1} + p^{k+1} - x^k. \end{array} \right. \tag{2.1}$$

*Moreover, assume that the following hold:*

(i) *For every* $k \geq 1$, $D_k \in \{C_1, ..., C_m\}$ *and* $C = \bigcap_{i=1}^m C_i = \bigcap_{k=1}^{+\infty} D_k$.

(ii) $(\exists N \in \mathbb{N})(\forall i \in \{1, .., m\})(\forall k \in \mathbb{N})(\exists l_k(i) \in \{k, ..., k+N-1\})$ *such that* $D_{l_k(i)} = C_i$.

(iii) $0 < \inf_{k\in\mathbb{N}} \pi_k$ *and* $\|L\|^2 < \frac{1}{\gamma}\left(\frac{1}{\tau} - \frac{1}{2\mu}\right)$.

*Then* $((x^k, u^k))_{k\in\mathbb{N}}$ *converges weakly P-a.s. to a Z-valued random variable* $(x, u)$.

*Proof.*
Let $(\hat{x}, \hat{u}) \in Z$. Let $\mathscr{X} = (\mathcal{X}_k)_{k\in\mathbb{N}}$ and $(\mathcal{E}_k)_{k\in\mathbb{N}}$ be sequences of sub-sigma-algebras of $\mathcal{F}$ such that, for every $k \in \mathbb{N}$, $\mathcal{X}_k = \sigma(x^0, ..., x^k)$ and $\mathcal{E}_k = \sigma(\varepsilon_k)$. It follows from (2.1) that

$$\begin{aligned} &\mathbb{E}(\|x^{k+1} - \hat{x}\|^2 \mid \mathcal{X}_k) \\ &\quad = \mathbb{E}(\langle p^{k+1} - \hat{x} + \epsilon_{k+1}(P_{D_{k+1}} p^{k+1} - p^{k+1}) \mid p^{k+1} - \hat{x} + \epsilon_{k+1}(P_{D_{k+1}} p^{k+1} - p^{k+1})\rangle \mid \mathcal{X}_k) \\ &\quad = \mathbb{E}(\|p^{k+1} - \hat{x}\|^2 \mid \mathcal{X}_k) + \mathbb{E}(2\epsilon_{k+1}\langle p^{k+1} - \hat{x} \mid P_{D_{k+1}} p^{k+1} - p^{k+1}\rangle \mid \mathcal{X}_k) \\ &\qquad + \mathbb{E}(\epsilon_{k+1}^2 \|P_{D_{k+1}} p^{k+1} - p^{k+1}\|^2 \mid \mathcal{X}_k). \end{aligned} \tag{2.2}$$

Since $(\epsilon_k)_{k\in\mathbb{N}}$ are independent, $\mathcal{X}_k$ and $\mathcal{E}_{k+1}$ are independent. Moreover, from the firm- nonexpansiveness of $(P_{D_k})_{k\in\mathbb{N}}$ [3, Proposition 4.16] and $\hat{x} = P_{D_k}\hat{x}$, for every $k \in \mathbb{N}$, we obtain, P-a.s.,

$$\begin{aligned}\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) &= \|p^{k+1}-\hat{x}\|^2 + 2\pi_{k+1}\langle p^{k+1}-\hat{x} \mid P_{D_{k+1}}p^{k+1}-p^{k+1}\rangle + \pi_{k+1}\|P_{D_{k+1}}p^{k+1}-p^{k+1}\|^2 \\ &= (1-\pi_{k+1})\|p^{k+1}-\hat{x}\|^2 + \pi_{k+1}\|P_{D_{k+1}}p^{k+1}-\hat{x}\|^2 \\ &\leq (1-\pi_{k+1})\|p^{k+1}-\hat{x}\|^2 + \pi_{k+1}\|p^{k+1}-\hat{x}\|^2 - \pi_{k+1}\|P_{D_{k+1}}p^{k+1}-p^{k+1}\|^2,\end{aligned} \tag{2.3}$$

which yields, P-a.s.

$$\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) + \pi_{k+1}\|P_{D_{k+1}}p^{k+1}-p^{k+1}\|^2 \leq \|p^{k+1}-\hat{x}\|^2. \tag{2.4}$$

It follows from Proposition 1.3 (i) applied to $G_{k+1}(p^{k+1}, \epsilon_{k+1}) = p^{k+1} + \epsilon_{k+1}(P_{D_{k+1}}p^{k+1} - p^{k+1})$ and (2.4) that

$$\begin{aligned}\frac{\|x^k-\hat{x}\|^2}{\tau} + \frac{\|u^k-\hat{u}\|^2}{\gamma} &\geq \frac{1}{\tau}\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) + \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)\|x^k-p^{k+1}\|^2 + \frac{\|u^{k+1}-\hat{u}\|^2}{\gamma} \\ &\quad + \frac{\|u^{k+1}-u^k\|^2}{\gamma} + 2\langle L(p^{k+1}-x^k) \mid u^{k+1}-\hat{u}\rangle \\ &\quad - 2\langle L(p^k-x^{k-1}) \mid u^k-\hat{u}\rangle - 2\|L\|\cdot\|p^k-x^{k-1}\|\cdot\|u^{k+1}-u^k\| \\ &\quad + \frac{\pi_{k+1}}{\tau}\|P_{D_{k+1}}p^{k+1}-p^{k+1}\|^2 \qquad \text{P-a.s.} \\ &\geq \frac{1}{\tau}\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) + \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)\|x^k-p^{k+1}\|^2 + \frac{\|u^{k+1}-\hat{u}\|^2}{\gamma} \\ &\quad + 2\langle L(p^{k+1}-x^k) \mid u^{k+1}-\hat{u}\rangle - 2\langle L(p^k-x^{k-1}) \mid u^k-\hat{u}\rangle \\ &\quad + \left(\frac{1}{\gamma}-\frac{1}{\nu}\right)\|u^{k+1}-u^k\|^2 - \nu\|L\|^2\|p^k-x^{k-1}\|^2 \\ &\quad + \frac{\pi_{k+1}}{\tau}\|P_{D_{k+1}}p^{k+1}-p^{k+1}\|^2 \qquad \text{P-a.s.,}\end{aligned} \tag{2.5}$$

where $\nu > 0$. In particular, if we set $\nu = 2\left(\frac{1}{\gamma} + \frac{2\mu\tau\|L\|^2}{2\mu-\tau}\right)^{-1}$, and $\rho = \frac{1}{2}(\frac{1}{\gamma} - \frac{2\mu\tau\|L\|^2}{2\mu-\tau}) > 0$, we have $\nu\|L\|^2 = \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)(1-\nu\rho)$. Hence, from (2.5) we obtain, P-a.s.

$$\begin{aligned}\frac{\|x^k-\hat{x}\|^2}{\tau} + \frac{\|u^k-\hat{u}\|^2}{\gamma} &\geq \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)\|x^k-p^{k+1}\|^2 + \frac{\|u^{k+1}-\hat{u}\|^2}{\gamma} + \frac{1}{\tau}\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) \\ &\quad + 2\langle L(p^{k+1}-x^k) \mid u^{k+1}-\hat{u}\rangle - 2\langle L(p^k-x^{k-1}) \mid u^k-\hat{u}\rangle \\ &\quad + \rho\|u^{k+1}-u^k\|^2 - \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)(1-\nu\rho)\|p^k-x^{k-1}\|^2 \\ &\quad + \frac{\pi_{k+1}}{\tau}\|P_{D_{k+1}}p^{k+1}-p^{k+1}\|^2.\end{aligned} \tag{2.6}$$

Moreover, since $\mathcal{X}_k = \sigma(x^0, \ldots, x^k)$, we have, P-a.s.,

$$\begin{aligned}&\frac{1}{\tau}\mathbb{E}\left(\|x^{k+1}-\hat{x}\|^2\middle|\mathcal{X}_k\right) + \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)\|p^{k+1}-x^k\|^2 + 2\langle L(p^{k+1}-x^k) \mid u^{k+1}-\hat{u}\rangle + \frac{\|u^{k+1}-\hat{u}\|^2}{\gamma} \\ &\quad = \mathbb{E}\left(\frac{\|x^{k+1}-\hat{x}\|^2}{\tau} + \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)\|p^{k+1}-x^k\|^2 + 2\langle L(p^{k+1}-x^k)|u^{k+1}-\hat{u}\rangle + \frac{\|u^{k+1}-\hat{u}\|^2}{\gamma}\middle|\mathcal{X}_k\right)\end{aligned} \tag{2.7}$$

and, defining

$$a_k = \left(\frac{1}{\tau} - \frac{1}{2\mu}\right) \|p^k - x^{k-1}\|^2 + 2\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle + \frac{\|u^k - \hat{u}\|^2}{\gamma}, \tag{2.8}$$

we deduce from (2.6), (2.7), and (2.8) that, P-a.s.,

$$\begin{aligned} a_k + \frac{\|x^k - \hat{x}\|^2}{\tau} \geq & \mathbb{E}\left(a_{k+1} + \frac{\|x^{k+1} - \hat{x}\|^2}{\tau} \mid \mathcal{X}_k\right) + \rho\|u^{k+1} - u^k\|^2 + \nu\rho\left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|p^k - x^{k-1}\|^2 \\ & + \frac{\pi_{k+1}}{\tau}\|P_{D_{k+1}}p^{k+1} - p^{k+1}\|^2. \end{aligned} \tag{2.9}$$

Note that from (iii) we have, for every $k \in \mathbb{N}$,

$$\begin{aligned} a_k &\geq \gamma\|L\|^2\|p^k - x^{k-1}\|^2 + 2\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle + \frac{\|u^k - \hat{u}\|^2}{\gamma} \\ &\geq \frac{1}{\gamma}(\|\gamma L(p^k - x^{k-1})\|^2 + 2\gamma\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle + \|u^k - \hat{u}\|^2) \\ &\geq \frac{1}{\gamma}(\|\gamma L(p^k - x^{k-1})\|^2 - 2\gamma\|L(p^k - x^{k-1})\|\,\|u^k - \hat{u}\| + \|u^k - \hat{u}\|^2) \\ &= \frac{1}{\gamma}\Big(\|u^k - \hat{u}\| - \|\gamma L(p^k - x^{k-1})\|\Big)^2 \geq 0. \end{aligned} \tag{2.10}$$

Thus, for every $k \in \mathbb{N}$, $a_k + \frac{\|x^k - \hat{x}\|^2}{\tau}$ is a $[0, \infty[$-valued $\mathcal{X}_k$-measurable random variable. On the other hand, we have

$$b_k := \rho\|u^{k+1} - u^k\|^2 + \nu\rho\left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|p^k - x^{k-1}\|^2 + \frac{\pi_{k+1}}{\tau}\|P_{D_{k+1}}p^{k+1} - p^{k+1}\|^2 \in \ell_+(\mathscr{X}) \tag{2.11}$$

and Lemma 1.4 imply that there exists a set $\Omega_1 \in \mathcal{F}$ such that $\mathbb{P}(\Omega_1) = 1$ and for every $w \in \Omega_1$, $\Big(a_k(w) + \frac{\|x^k(w) - \hat{x}(w)\|^2}{\tau}\Big)_{k\in\mathbb{N}}$ converges to a $[0, \infty[$-valued random variable and $(b_k)_{k\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$. Therefore, since $0 < \inf_{k\in\mathbb{N}} \pi_k$ we have that, $P - a.s.$

$$\sum_{i=1}^{\infty}\|u^{k+1} - u^k\|^2 < +\infty, \quad \sum_{i=1}^{\infty}\|p^k - x^{k-1}\|^2 < +\infty, \text{ and } \sum_{i=1}^{\infty}\|P_{D_{k+1}}p^{k+1} - p^{k+1}\|^2 < +\infty. \tag{2.12}$$

It follows from (2.10), (2.12) and the convergence of $(a_k + \frac{\|x^k - \hat{x}\|^2}{\tau})_{k\in\mathbb{N}}$ that $(u^k - \hat{u})_{k\in\mathbb{N}}$ is bounded $P - a.s.$, and, from (2.12), we have that

$$(\forall w \in \Omega_1) \quad \lim_{k\to\infty} \frac{\|x^k(w) - \hat{x}(w)\|^2}{\tau} + \frac{\|u^k(w) - \hat{u}(w)\|^2}{\gamma} = \lim_{k\to\infty} a_k(w) + \frac{\|x^k(w) - \hat{x}(w)\|^2}{\tau}. \tag{2.13}$$

Then $\frac{\|x^k - \hat{x}\|^2}{\tau} + \frac{\|u^k - \hat{u}\|^2}{\gamma}$ converges P-a.s. to a $[0, \infty[$-valued random variable. It follows from Proposition 1.3 (ii) with $G_{k+1}(p^{k+1}, \epsilon_{k+1}) = p^{k+1} + \epsilon_{k+1}(P_{D_{k+1}}\, p^{k+1} - p^{k+1})$ that there exists $\Omega_Z$ such that $\mathbb{P}(\Omega_Z) = 1$ and, for every $\omega \in \Omega_Z$ and for every $(\hat{x}, \hat{u}) \in Z$, $\Big(\frac{\|x^k(w) - \hat{x}\|^2}{\tau} + \frac{\|u^k(w) - \hat{u}\|^2}{\gamma}\Big)_{k\in\mathbb{N}}$ converges.

It remains to prove that, for every $w \in \widetilde{\Omega} := \Omega_1 \cap \Omega_Z$, we have that $\mathfrak{W}(x^k(w), u^k(w))_{n\in\mathbb{N}} \subset Z$.

Let $(x(w),u(w)) \in \mathfrak{W}(x^k(w),u^k(w))_{n\in\mathbb{N}}$, say $(x^{k_n}(w),u^{k_n}(w)) \rightharpoonup (x(w),u(w))$. Note that, for every $w \in \Omega_1$ and $(x^{j_n})_{n\in\mathbb{N}}$ such that $k_n+1 \leq j_n \leq k_n+N$, $\sum_{n=1}^{\infty}\|x^{j_n}(w)-x^{k_n}(w)\|^2 < +\infty$. Indeed,

$$\begin{aligned}
\|x^{j_n}(w)-x^{k_n}(w)\|^2 &\leq N\sum_{i=k_n}^{j_n-1}\|x^{i+1}(w)-x^i(w)\|^2\\
&\leq N\sum_{i=k_n}^{j_n-1}\left(\|p^{i+1}(w)+\epsilon_{i+1}(w)\left(P_{D_{i+1}}p^{i+1}(w)-p^{i+1}(w)\right)-x^i(w)\|^2\right)\\
&\leq 2N\sum_{i=k_n}^{k_n+N-1}\left(\|p^{i+1}(w)-x^i(w)\|^2+\|P_{D_{i+1}}p^{i+1}(w)-p^{i+1}(w)\|^2\right) \qquad (2.14)
\end{aligned}$$

and, by suming (2.14) from $n=1$ to infinity, it follows from (2.12) that

$$\begin{aligned}
\sum_{n=1}^{\infty}\|x^{j_n}(w)-x^{k_n}(w)\|^2 &\leq 2N\sum_{n=1}^{\infty}\sum_{i=k_n}^{k_n+N-1}\left(\|p^{i+1}(w)-x^i(w)\|^2+\|P_{D_{i+1}}p^{i+1}(w)-p^{i+1}(w)\|^2\right)\\
&\leq 2N\sum_{n=1}^{\infty}\sum_{i=n}^{n+N-1}\left(\|p^{i+1}(w)-x^i(w)\|^2+\|P_{D_{i+1}}p^{i+1}(w)-p^{i+1}(w)\|^2\right)\\
&\leq 2N^2\sum_{i=1}^{\infty}\left(\|p^{i+1}(w)-x^i(w)\|^2+\|P_{D_{i+1}}p^{i+1}(w)-p^{i+1}(w)\|^2\right)\\
&< +\infty \qquad (2.15)
\end{aligned}$$

and the result follows.

The assumption (ii) imply that, for every $i \in \{1,...,m\}$ and $n \in \mathbb{N}$, there exists $j_n(i) \in \{k_n+1,...,k_n+N\}$ such that $D_{j_n(i)} = C_i$. Since $P_{C_i}$ is nonexpansive, $Id - P_{C_i}$ is maximally monotone operator [3, Example 20.29] and therefore, it has weak-strong closed graph [3, Proposition 20.38]. Hence, it follows from (2.12), (2.15), and $(x^{k_n}(w),u^{k_n}(w)) \rightharpoonup (x(w),u(w))$ that $(Id-P_{D_{j_n(i)}})p^{j_n(i)} = (Id-P_{C_i})p^{j_n(i)} \to 0$ and $p^{j_n(i)}(w) \rightharpoonup x(w)$ and, hence, $x(w) \in Fix(P_{C_i}) = C_i$, for every $i \in \{1,...,m\}$, which yield $x(w) \in \bigcap_{i=1}^m C_i = C$.

Note that (1.13) can be written equivalently as:

$$(y^k,v^k) \in (M+Q)(p^{k+1},u^{k+1}), \qquad (2.16)$$

where

$$\begin{aligned}
M &: (p,\eta) \mapsto (\partial f(p)+L^*\eta)\times(\partial g^*(\eta)-Lp)\\
Q &: (p,\eta) \mapsto (\nabla h(p),0) \qquad (2.17)
\end{aligned}$$

.

are maximally monotone [5, Proposition 2.7(i)] and

$$\begin{aligned}
y^k &:= \frac{x^k-p^{k+1}}{\tau}+\nabla h(p^{k+1})-\nabla h(x^k)\\
v^k &:= \frac{u^k-u^{k+1}}{\gamma}+L(x^k-p^{k+1}+p^k-x^{k-1}). \qquad (2.18)
\end{aligned}$$

It follows from [3, Corollary25.5] that $M + Q$ is maximally monotone. Since (2.12) and $(x^{k_n}(w), u^{k_n}(w)) \rightharpoonup (x(w), u(w))$ yields, $(y^{k_n}(w), v^{k_n}(w)) \to 0$ and $(p^{k_n+1}(w), u^{k_n+1}(w)) \rightharpoonup (x(w), u(w))$, from the weak-strong closedness of the graph of $M+Q$ we deduce that $(x(w), u(w)) \in Z$. The weak convergence and the measurability results follows from follows from [3, Lemma 2.47] and [21, Corollary 1.13], respectively. □

**Remark 2.2** If $\operatorname{argmin}_{x\in\mathcal{H}}(f(x)+g(Lx)+h(x)) \subset C$ we have that $\mathfrak{W}(x^k, u^k)_{n\in\mathbb{N}} \subset C \times \mathcal{G}$ P-a.s., therefore, it is not necessary to prove that $x \in C$ P-a.s. and the conditions (i) and (ii) in Theorem 2.1 are not needed.

**Remark 2.3** From theorem 2.1, we obtain several methods in the literature, as we detail below:

1. **Primal-dual with random cyclic projections**: Conditions (i) and (ii) holds if $N = m$ and, for every $k \in \mathbb{N}$, $D_k := C_{i(k)}$, where $i(k) = (k \bmod m) + 1$. This case corresponds to the primal-dual method with binary random Cyclic projections.

2. **Primal-dual with deterministic cyclic projections** In the case when, for every $k \in \mathbb{N}$, $D_k := C_1$ and $\epsilon_k^{-1}(\{1\}) = \Omega$, the algorithm reduces to the method proposed in [7]. We can also allow for deterministic cyclic projections by setting $m > 1$ and, for every $k \in \mathbb{N}$, $\epsilon_k^{-1}(\{1\}) = \Omega$, $D_k := C_{i(k)}$, where $i(k) = (k \bmod m) + 1$.

3. **Projections onto convex sets**: In the case when $f = g = h = L = 0 \in \Gamma_0(\mathcal{H})$, and $C := \bigcap_{i=1}^m C_i \neq \emptyset$, where $C_i$ is a nonempty closed convex subset of $\mathcal{H}$, for every $i = 1, ..., m$. If we define for every $k \geq 1$, $D_k := C_{i(k)}$, where $i(k) = (k \bmod m) + 1$, and set $\epsilon_k^{-1}(\{1\}) = \Omega$, then (i), (ii), and (iii) holds.

   Let $x^0 = \overline{x}^0 = u^0 \in \mathcal{H}$ and set $\gamma = \tau = 1$. Then, the algorithm (2.1) reduces to

   $$(\forall k \in \mathbb{N}) \;\; x^{k+1} = P_{C_{i(k+1)}} x^k \tag{2.19}$$

   which is the method proposed in [3, Corollary 5.26] and its convergence follows from Theorem 2.1.

4. **Kaczmarz method:** Let $R$ be a full rank $m \times n$ matrix such that $m \leq n$, and $b \in \mathbb{R}^m$. We denote the rows of $R$ by $r_1, ..., r_m$ and let $b = (b_1, ..., b_m)^\top$. A special case of (2.19) is when $\mathcal{H} = \mathbb{R}^n$ and $C = \{x \in \mathcal{H} |\ Rx = b\} = \bigcap_{i=1}^m C_i \neq \emptyset$, where $C_i = \{x \in \mathcal{H} \mid \langle r_i, x\rangle = b_i\}$. It is clear that $C_i$ is a nonempty closed convex subset of $\mathcal{H}$ for every $i = 1, ..., m$. If we define for every $k \geq 1$, $D_k := C_{i(k)}$, where $i(k) = (k \bmod m) + 1$, and set $\epsilon_k^{-1}(\{1\}) = \Omega$, then (i), (ii), and (iii) holds.

   Let $x^0 = \overline{x}^0 = u^0 \in \mathcal{H}$ and let $\gamma = \tau = 1$. It follows from $\operatorname{prox}_f(x) = x$, $\operatorname{prox}_{g^*}(x) = x - \operatorname{prox}_g(x) = 0$ [3, Proposition 24.8(ix)], and (2.1) reduces to

   $$(\forall k \in \mathbb{N}) \;\; x^{k+1} = P_{C_{i(k+1)}} x^k = x^k + \frac{b_{i(k+1)} - \langle r_{i(k+1)} \mid x^k\rangle}{\|r_{i(k+1)}\|^2} r_{i(k+1)} \tag{2.20}$$

   which is the Kaczmarz method proposed in [19] and its convergence follows from Theorem 2.1.

# Chapter 3

# Randomized Kaczmarz projections in convex optimization

In this chapter, we present an extension of the primal-dual algorithm proposed in [7] which, inspired by Randomized Kaczmarz method, at each iteration chooses randomly a set in $\{\mathcal{H}, C_1, \dots, C_m\}$ and projects on it. A difference with respect to the method in Chapter 2 is that in Theorem 3.1 there is no pre-determined order onto which the algorithm projects. Roughly speaking the algorithm "roll a dice" to choose a set onto which it projects at each iteration.

**Theorem 3.1** *Consider the setting of Problem 1.2. Let* $\tau \in ]0, 2\mu[$*, let* $\gamma > 0$*, let* $(x^0, \overline{x}^0, u^0) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G}$*, be such that* $x^0 = \overline{x}^0$ *and, set* $I = \{0, 1, ..., m\}$*. Let* $(\epsilon_k)_{k\in\mathbb{N}}$ *be a sequence of independent* $I$*-valued random variables. Consider the following routine*

$$
(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} u^{k+1} = \operatorname{prox}_{\gamma g^*}(u^k + \gamma L \bar{x}^k) \\ p^{k+1} = \operatorname{prox}_{\tau f}(x^k - \tau(L^* u^{k+1} + \nabla h(x^k))) \\ x^{k+1} = P_{C_{\epsilon_{k+1}}} p^{k+1} \\ \bar{x}^{k+1} = x^{k+1} + p^{k+1} - x^k. \end{array} \right. \tag{3.1}
$$

*and assume that the following hold:*

(i) $C_0 = \mathcal{H}$ *and* $\|L\|^2 < \frac{1}{\gamma}\left(\frac{1}{\tau} - \frac{1}{2\mu}\right)$.

(ii) $(\forall i \in I \backslash \{0\})\ 0 < \inf_{k\in\mathbb{N}} \pi_k^i$*, where* $(\forall i \in I)(\forall k \in \mathbb{N})\ \pi_k^i = \mathbb{P}(\epsilon_k^{-1}(\{i\}))$.

*Then* $((x^k, u^k))_{k\in\mathbb{N}}$ *converges weakly P-a.s. to a Z-valued random variable* $(x, u)$.

*Proof.*
Let $(\hat{x}, \hat{u}) \in Z$ and let $\mathscr{X} = (\mathcal{X}_k)_{k\in\mathbb{N}}$ be a sequence of sub-sigma-algebras of $\mathcal{F}$ such that, for every $k \in \mathbb{N}$, $\mathcal{X}_k = \sigma(x^0, ..., x^k)$. It follows from (3.1), the linearity of conditional expectation, and the

independence of $(\epsilon_k)_{k\in\mathbb{N}}$ that

$$\begin{aligned}\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) &= \mathbb{E}(\sum_{i\in I} 1\!\!1_{\{\epsilon_{k+1}=i\}} \cdot \|P_{C_i}p^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k)\\ &= \sum_{i\in I}\mathbb{E}(1\!\!1_{\{\epsilon_{k+1}=i\}} \cdot \|P_{C_i}p^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k)\\ &= \sum_{i\in I}\mathbb{E}(1\!\!1_{\{\epsilon_{k+1}=i\}} \mid \mathcal{X}_k)\|P_{C_i}p^{k+1}-\hat{x}\|^2\\ &= \sum_{i\in I}\pi^i_{k+1}\|P_{C_i}p^{k+1}-\hat{x}\|^2. \qquad (3.2)\end{aligned}$$

Now set $I_0 := I\backslash\{0\}$, since $\sum_{i\in I}\pi^i_k$, for every $k\in\mathbb{N}$, $\hat{x} = P_{C_i}\hat{x}$, and the firm-nonexpansiveness of $P_{C_i}$, for every $i$ in $I_0$, we obtain, P-a.s.

$$\begin{aligned}\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) &= \pi^0_{k+1}\|p^{k+1}-\hat{x}\|^2 + \sum_{i\in I_0}\pi^i_{k+1}\|P_{C_i}p^{k+1}-\hat{x}\|^2\\ &\le \sum_{i\in I}\pi^i_{k+1}\|p^{k+1}-\hat{x}\|^2 - \sum_{i\in I_0}\pi^i_{k+1}\|P_{C_i}p^{k+1}-p^{k+1}\|^2, \qquad (3.3)\end{aligned}$$

which yields, P-a.s.

$$\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) + \sum_{i\in I_0}\pi^i_{k+1}\|P_{C_i}p^{k+1}-p^{k+1}\|^2 \le \|p^{k+1}-\hat{x}\|^2. \qquad (3.4)$$

It follows from Proposition 1.3 (i) applied to $G_{k+1}(p^{k+1},\epsilon_{k+1}) = P_{C_{\epsilon_{k+1}}}p^{k+1}$ and (3.4) that

$$\begin{aligned}\frac{\|x^k-\hat{x}\|^2}{\tau} + \frac{\|u^k-\hat{u}\|^2}{\gamma} &\ge \frac{1}{\tau}\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) + \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)\|x^k-p^{k+1}\|^2 + \frac{\|u^{k+1}-\hat{u}\|^2}{\gamma}\\ &\quad + \frac{\|u^{k+1}-u^k\|^2}{\gamma} + 2\langle L(p^{k+1}-x^k) \mid u^{k+1}-\hat{u}\rangle\\ &\quad - 2\langle L(p^k-x^{k-1}) \mid u^k-\hat{u}\rangle - 2\|L\|\cdot\|p^k-x^{k-1}\|\cdot\|u^{k+1}-u^k\|\\ &\quad + \sum_{i\in I_0}\frac{\pi^i_{k+1}}{\tau}\|P_{C_i}p^{k+1}-p^{k+1}\|^2 \qquad \text{P-a.s.}\\ &\ge \frac{1}{\tau}\mathbb{E}(\|x^{k+1}-\hat{x}\|^2 \mid \mathcal{X}_k) + \left(\frac{1}{\tau}-\frac{1}{2\mu}\right)\|x^k-p^{k+1}\|^2 + \frac{\|u^{k+1}-\hat{u}\|^2}{\gamma}\\ &\quad + 2\langle L(p^{k+1}-x^k) \mid u^{k+1}-\hat{u}\rangle - 2\langle L(p^k-x^{k-1}) \mid u^k-\hat{u}\rangle\\ &\quad + \left(\frac{1}{\gamma}-\frac{1}{\nu}\right)\|u^{k+1}-u^k\|^2 - \nu\|L\|^2\|p^k-x^{k-1}\|^2\\ &\quad + \sum_{i\in I_0}\frac{\pi^i_{k+1}}{\tau}\|P_{C_i}p^{k+1}-p^{k+1}\|^2 \qquad \text{P-a.s.}, \qquad (3.5)\end{aligned}$$

where $\nu > 0$, in particular if we set $\nu = 2\left(\frac{1}{\gamma} + \frac{2\mu\tau\|L\|^2}{2\mu-\tau}\right)^{-1}$, and $\rho = \frac{1}{2}(\frac{1}{\gamma} - \frac{2\mu\tau\|L\|^2}{2\mu-\tau}) > 0$, we have $\nu\|L\|^2 = \left(\frac{1}{\tau} - \frac{1}{2\mu}\right)(1-\nu\rho)$ and by (3.5) we obtain, P-a.s.

$$
\begin{aligned}
\frac{\|x^k - \hat{x}\|^2}{\tau} + \frac{\|u^k - \hat{u}\|^2}{\gamma} \geq & \left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|x^k - p^{k+1}\|^2 + \frac{\|u^{k+1} - \hat{u}\|^2}{\gamma} + \frac{1}{\tau}\mathbb{E}(\|x^{k+1} - \hat{x}\|^2 \mid \mathcal{X}_k) \\
& + 2\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\rangle - 2\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle + \rho\|u^{k+1} - u^k\|^2 \\
& + \rho\|u^{k+1} - u^k\|^2 - \left(\frac{1}{\tau} - \frac{1}{2\mu}\right)(1-\nu\rho)\|p^k - x^{k-1}\|^2 \\
& + \sum_{i \in I_0} \frac{\pi^i_{k+1}}{\tau}\|P_{C_i}p^{k+1} - p^{k+1}\|^2. \qquad (3.6)
\end{aligned}
$$

Moreover, since $\mathcal{X}_k = \sigma\left(x^0, \ldots, x^k\right)$, we have, P-a.s.,

$$
\begin{aligned}
& \frac{1}{\tau}\mathbb{E}\bigg(\|x^{k+1} - \hat{x}\|^2 \bigg| \mathcal{X}_k\bigg) + \left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|p^{k+1} - x^k\|^2 + 2\langle L(p^{k+1} - x^k) \mid u^{k+1} - \hat{u}\rangle + \frac{\|u^{k+1} - \hat{u}\|^2}{\gamma} \\
& \quad = \mathbb{E}\bigg(\frac{\|x^{k+1} - \hat{x}\|^2}{\tau} + \left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|p^{k+1} - x^k\|^2 + 2\langle L(p^{k+1} - x^k)|u^{k+1} - \hat{u}\rangle + \frac{\|u^{k+1} - \hat{u}\|^2}{\gamma}\bigg| \mathcal{X}_k\bigg)
\end{aligned}
\qquad (3.7)
$$

and, defining

$$
a_k = \left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|p^k - x^{k-1}\|^2 + 2\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle + \frac{\|u^k - \hat{u}\|^2}{\gamma}, \qquad (3.8)
$$

we obtain, P-a.s.

$$
\begin{aligned}
a_k + \frac{\|x^k - \hat{x}\|^2}{\tau} \geq & \mathbb{E}\left(a_{k+1} + \frac{\|x^{k+1} - \hat{x}\|^2}{\tau} \mid \mathcal{X}_k\right) + \rho\|u^{k+1} - u^k\|^2 + \nu\rho\left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|p^k - x^{k-1}\|^2 \\
& + \sum_{i \in I_0} \frac{\pi^i_{k+1}}{\tau}\|P_{C_i}p^{k+1} - p^{k+1}\|^2. \qquad (3.9)
\end{aligned}
$$

Note that from (i) we have, for every $k \in \mathbb{N}$,

$$
\begin{aligned}
a_k &\geq \gamma\|L\|^2\|p^k - x^{k-1}\|^2 + 2\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle + \frac{\|u^k - \hat{u}\|^2}{\gamma} \\
&\geq \frac{1}{\gamma}(\|\gamma L(p^k - x^{k-1})\|^2 + 2\gamma\langle L(p^k - x^{k-1}) \mid u^k - \hat{u}\rangle + \|u^k - \hat{u}\|^2) \\
&\geq \frac{1}{\gamma}(\|\gamma L(p^k - x^{k-1})\|^2 - 2\gamma\|L(p^k - x^{k-1})\|\|u^k - \hat{u}\| + \|u^k - \hat{u}\|^2) \\
&= \frac{1}{\gamma}\left(\|u^k - \hat{u}\| - \|\gamma L(p^k - x^{k-1})\|\right)^2 \geq 0. \qquad (3.10)
\end{aligned}
$$

Thus, for every $k \in \mathbb{N}$, $a_k + \frac{\|x^k - \hat{x}\|^2}{\tau}$ is a $[0, \infty[$-valued $\mathcal{X}_k$-measurable random variable. On the other hand, by definition of $\ell_+(\mathscr{X})$, we have

$$
b_k := \rho\|u^{k+1} - u^k\|^2 + \nu\rho\left(\frac{1}{\tau} - \frac{1}{2\mu}\right)\|p^k - x^{k-1}\|^2 + \sum_{i \in I_0} \frac{\pi^i_{k+1}}{\tau}\|P_{C_i}p^{k+1} - p^{k+1}\|^2 \in \ell_+(\mathscr{X}) \qquad (3.11)
$$

and Lemma 1.4 imply that $(a_k + \frac{\|x^k - \hat{x}\|^2}{\tau})_{k\in\mathbb{N}}$ converges $P - a.s.$ to a $[0, \infty[$-valued random variable and $(b_k)_{k\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$. Therefore, since $0 < \min_{i\in I_0} \inf_{k\in\mathbb{N}} \pi_k^i$, we have that, $P - a.s.$

$$\sum_{k\in\mathbb{N}} \|u^{k+1} - u^k\|^2 < +\infty, \quad \sum_{k\in\mathbb{N}} \|p^k - x^{k-1}\|^2 < +\infty, \text{ and } (\forall i \in I) \sum_{k\in\mathbb{N}} \|P_{C_i} p^{k+1} - p^{k+1}\|^2 < +\infty. \tag{3.12}$$

It follows from (3.10), (3.12), and the convergence of $(a_k + \frac{\|x^k - \hat{x}\|^2}{\tau})_{k\in\mathbb{N}}$ that $(u^k - \hat{u})_{k\in\mathbb{N}}$ is bounded and from (3.12) there exists a set $\Omega_1 \in \mathcal{F}$ such that $\mathbb{P}(\Omega_1) = 1$ and, for every $w \in \Omega_1$, we have that

$$\lim_{k\to\infty} \frac{\|x^k(w) - \hat{x}(w)\|^2}{\tau} + \frac{\|u^k(w) - \hat{u}(w)\|^2}{\gamma} = \lim_{k\to\infty} a_k(w) + \frac{\|x^k(w) - \hat{x}(w)\|^2}{\tau}. \tag{3.13}$$

Then $\frac{\|x^k - \hat{x}\|^2}{\tau} + \frac{\|u^k - \hat{u}\|^2}{\gamma}$ converges P-a.s. to a $[0, \infty[$-valued random variable. It follows from Proposition 1.3 (ii) with $G_{k+1}(p^{k+1}, \epsilon_{k+1}) = P_{C_{\epsilon_{k+1}}} p^{k+1}$ that there exists $\Omega_Z$ such that $\mathbb{P}(\Omega_Z) = 1$ and, for every $\omega \in \Omega_Z$ and for every $(\hat{x}, \hat{u}) \in Z$, $\left(\frac{\|x^k(w) - \hat{x}\|^2}{\tau} + \frac{\|u^k(w) - \hat{u}\|^2}{\gamma}\right)_{k\in\mathbb{N}}$ converges.

It remains to prove that, for every $w \in \widetilde{\Omega} := \Omega_1 \cap \Omega_Z$, we have that $\mathfrak{W}(x^k(w), u^k(w))_{n\in\mathbb{N}} \subset Z$. Let $(x(w), u(w)) \in \mathfrak{W}(x^k(w), u^k(w))_{n\in\mathbb{N}}$, say $(x^{k_n}(w), u^{k_n}(w)) \rightharpoonup (x(w), u(w))$. Note that, for every $w \in \Omega_1$ and $(x^{k_n})_{n\in\mathbb{N}}$, $\sum_{n=1}^{\infty} \|x^{k_n+1} - x^{k_n}\|^2 < +\infty \quad P - a.s.$ Indeed, it follows from (3.12) that

$$\begin{aligned}
\sum_{n=1}^{\infty} \|x^{k_n+1}(w) - x^{k_n}(w)\|^2 &= \sum_{n=1}^{\infty} \left(\|P_{C_{\epsilon_{k_n+1}(w)}} p^{k_n+1}(w) - x^{k_n}(w)\|^2\right) \\
&\leq 2\sum_{n=1}^{\infty} \left(\|p^{k_n+1}(w) - x^{k_n}(w)\|^2 + \|P_{C_{\epsilon_{k_n+1}(w)}} p^{k_n+1}(w) - p^{k_n+1}(w)\|^2\right) \\
&\leq 2\sum_{n=1}^{\infty} \left(\|p^{k_n+1}(w) - x^{k_n}(w)\|^2 + \sum_{i\in I} \left(\|P_{C_{i+1}} p^{k_n+1}(w) - p^{k_n+1}(w)\|^2\right)\right) \\
&< +\infty
\end{aligned} \tag{3.14}$$

and the result follows.

On the other hand, since $P_{C_i}$ is nonexpansive, $Id - P_{C_i}$ is maximally monotone operator [3, Example 20.29] and, therefore, it has weak-strong closed graph [3, Proposition 20.38]. Hence, it follows from (3.12), (3.14), and $(x^{k_n}(w), u^{k_n}(w)) \rightharpoonup (x(w), u(w))$ that, for every $i$ in $I$, $(Id - P_{C_i})p^{k_n+1}(w) \to 0$ and $p^{k_n+1}(w) \rightharpoonup x(w)$ and, hence, $x(w) \in Fix(P_{C_i}) = C_i$, for every $i \in I$, which yields $x(w) \in \bigcap_{i=1}^m C_i = C$. Moreover

$$(y^k, v^k) \in (M + Q)(p^{k+1}, u^{k+1}), \tag{3.15}$$

where $M$, $Q$, and $(y^k, v^k)$ are defined in the same way as in (2.17) and (2.18), respectively. Since (3.12), (3.14), and $(x^{k_n}(w), u^{k_n}(w)) \rightharpoonup (x(w), u(w))$ yields $(y^{k_n}(w), v^{k_n}(w)) \to 0$ and $(p^{k_n+1}(w), u^{k_n+1}(w)) \rightharpoonup (x(w), u(w))$, we deduce from the weak-strong closedness of the graph of $M+Q$ that $(x(w),u(w)) \in Z$. The weak convergence and the measurability results follows from [3, Lemma 2.47] and [21, Corollary 1.13], respectively. □

**Remark 3.2** From theorem 3.1, we recover Randomized Kaczmarz and [7], as we detail below:

1. **Randomized Kaczmarz**: Let $R$ be a full rank $m \times n$ matrix such that $m \leq n$, and $b \in \mathbb{R}^m$. We denote the rows of $R$ by $r_1, ..., r_m$ and let $b = (b_1, ..., b_m)^\top$. Consider the case when $\mathcal{H} = \mathbb{R}^n$, $f = g = h = L = 0 \in \Gamma_0(\mathcal{H})$, and $C = \{x \in \mathcal{H} |\ Rx = b\} = \bigcap_{i=1}^m C_i \neq \emptyset$, where, for every $i \in I := \{1, \ldots, m\}$, $C_i = \{x \in \mathcal{H} \mid \langle r_i, x\rangle = b_i\}$. It is clear that $(C_i)_{i\in I}$ are nonempty closed convex sets. Let $(\epsilon_k)_{k\in\mathbb{N}}$ be a sequence of independent $I$-valued random variables, with $\pi_k^0 = 0$, and for every $i \in I$, $\pi_K^i$ is proportional to $\|r_i\|^2$ for all $k \in \mathbb{N}$ .

   Let $x^0 = \overline{x}^0 = u^0 \in \mathcal{H}$ and set $\gamma = \tau = 1$. It follows from $\text{prox}_f(x) = x$, $\text{prox}_{g^*}(x) = x - \text{prox}_g(x) = 0$ [3, Proposition 24.8(ix)], and (3.1) that

$$(\forall k \in \mathbb{N}) \ \ x^{k+1} = P_{C_{\epsilon_{k+1}}} x^k = x^k + \frac{b_{\epsilon_{k+1}} - \langle r_{\epsilon_{k+1}}, x^k\rangle}{\|r_{\epsilon_{k+1}}\|^2} r_{\epsilon_{k+1}}. \tag{3.16}$$

   which is the method proposed in [23], whose convergence is deduced by Theorem 3.1. Note that in [23] there is a convergence in expectation with expected exponential rate, instead, with the algorithm presented in this section, convergence is obtained P-a.s.

2. **Constrained primal-dual method**: In the case $m = 1$, Algorithms (3.1) and (2.1) are equivalent. Additionally by Remark 2.2.3 if $\epsilon_k^{-1}(\{1\}) = \Omega$, for every $k \in \mathbb{N}$, we recover the Algorithm proposed in [7].

# Chapter 4

# Application to the arc capacity expansion problem of a directed graph

In this chapter we apply the stochastic algorithms developed in chapter 2 and 3 to the arc capacity expansion problem of a directed graph. The goal is to obtain the Wardrop equilibrium, along with a expansion vector in each arc. The latter is a positive vector representing the expansion capacity at each arc.

## 4.1 Arc capacity expansion problem

Consider a directed graphs, where $\mathcal{A}$ is the set of arcs of the graph, $\mathcal{O}$ is the set of origin nodes, $\mathcal{D}$ is the set of destination nodes, $R_{o,d}$ is the set of routes from $o \in \mathcal{O}$ to $d \in \mathcal{D}$, $R := \bigcup_{(o,d)\in\mathcal{O}\times\mathcal{D}} R_{o,d}$ is the set of all routes. Consider the incidence matrix $N \in \mathbb{R}^{|\mathcal{A}|\times|R|}$ defined by

$$(\forall a \in \mathcal{A})(\forall r \in \mathcal{R}) \quad N_{a,r} := \begin{cases} 1 & \text{if } a \text{ belongs to } r \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

and we denote the rows of $N$ by $N_a$, for every $a \in \mathcal{A}$.

We will consider a stochastic approach introducing a finite set of scenarios $\Xi$. We denote by $x_\xi := (x_{a,\xi})_{a\in\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|}$ and $f_\xi := (f_{r,\xi})_{r\in R} \in \mathbb{R}^{|R|}$ to the vectors that represents the expandability of each arc and the flow of each route respectively, in every scenario $\xi \in \Xi$. In a similar way we define $x := (x_\xi)_{\xi\in\Xi} \in \mathbb{R}^{|\mathcal{A}||\Xi|}$ and $f := (f_\xi)_{\xi\in\Xi} \in \mathbb{R}^{|R||\Xi|}$.

The demand for every pair $(o,d) \in \mathcal{O}\times\mathcal{D}$ and $\xi \in \Xi$ is denoted by $h_{od,\xi} \in \mathbb{R}$. The capacity for every arc $a \in \mathcal{A}$ and $\xi \in \Xi$ is denoted by $c_{a,\xi} \in \mathbb{R}_+$. We denote by $u_\xi := Nf_\xi = (u_{a,\xi})_{a\in\mathcal{A}}$ the vector of flows in arcs in the scenario $\xi \in \Xi$, where $u_{a,\xi} := \sum_{r\in R} N_{a,r} f_{r,\xi}$.

Let $M := \bigtimes_{a\in\mathcal{A}}[0, M_a] \subset \mathbb{R}^{|\mathcal{A}|}$ be the set of constraints for the expansion capacities, where $M_a \in \mathbb{R}$ represents the expansion capacity limit of the arc $a \in \mathcal{A}$. The constraint set on the demands for each scenario is the following affine subspace

$$(\forall \xi \in \Xi) \quad V_\xi^0 := \left\{ f \in \mathbb{R}^{|R|} \;\middle|\; \forall (o,d) \in \mathcal{O} \times \mathcal{D} \quad \sum_{r \in R_{o,d}} f_r = h_{od,\xi} \right\}. \tag{4.2}$$

We denote by $V_\xi^+ := V_\xi^0 \cap \mathbb{R}_+^{|R|}$ the set that restricts flows only to positive values. The set of constraints on the flows of each arc in every scenario is

$$(\forall \xi \in \Xi) \quad H_\xi = \left\{ (x,u) \in \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|} \;\middle|\; (\forall a \in \mathcal{A}) \;\; u_a - x_a \le c_{a,\xi} \right\} \tag{4.3}$$

or equivalently, considering the flows on routes

$$(\forall \xi \in \Xi) \quad H_\xi = \bigcap_{a \in \mathcal{A}} P_\xi \left( C_{a,\xi} \right), \tag{4.4}$$

where

$$(\forall \xi \in \Xi)(\forall a \in \mathcal{A}) \quad C_{a,\xi} := \left\{ (x,f) \in \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|} \mid N_a f_\xi - x_{a,\xi} \le c_{a,\xi} \right\}, \tag{4.5}$$

and $P_\xi : \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|} \mapsto \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|R|}$ is the orthogonal projection $P_\xi : (x,f) \mapsto (x_\xi, f_\xi)$.

Suppose now that the expandability vector does not depends on the scenario $\xi \in \Xi$. That is, we consider the following Nonanticipativity condition

$$(\forall \xi, \xi' \in \Xi) \;\; x_\xi = x_{\xi'}, \tag{4.6}$$

we denote by $W$ the set of $x \in \mathbb{R}^{|\mathcal{A}||\Xi|}$ such that (4.6) is satisfied.

## 4.2 Formulation

Let $C := \bigcap_{i=1}^m C_i \neq \emptyset$, where, for every $i$ in $\{1, ..., m\}$, $C_i$ is a nonempty closed convex subset of $\mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|}$, and define the function

$$\begin{aligned} \varphi_\xi : \mathbb{R}^{|\mathcal{A}|} &\mapsto \mathbb{R} && (4.7) \\ u &\mapsto \sum_{a \in \mathcal{A}} \int_0^{u_{a,\xi}} t_{a,\xi}(z)dz, && (4.8) \end{aligned}$$

where $t_{a,\xi} : \mathbb{R} \mapsto \mathbb{R}_+$ is an increasing $\beta_{a,\xi}$-Lipschitz function representing the travel time in the arc $a \in \mathcal{A}$ for the scenario $\xi \in \Xi$. Consider the following optimization problem

$$\begin{aligned} \min_{\left((M^{|\Xi|} \cap W) \times \mathbb{R}_+^{|R||\Xi|}\right) \cap C} \; & \sum_{\xi \in \Xi} p_\xi \left[ \frac{1}{2} x_\xi^\top Q x_\xi + \varphi_\xi(N f_\xi) \right] \\ s.t. \quad & \\ & \sum_{r \in R_{o,d}} f_{r,\xi} = h_{od,\xi} && (\forall \xi \in \Xi) \;\; (\forall (o,d) \in \mathcal{O} \times \mathcal{D}) \\ & \sum_{r \in R} N_{a,r} f_{r,\xi} - x_{a,\xi} \le c_{a,\xi} && (\forall \xi \in \Xi) \;\; (\forall a \in \mathcal{A}) \qquad (4.9) \end{aligned}$$

where, for every $\xi \in \Xi$, $p_\xi > 0$, $\mathbb{P}$ is the probability of occurrence of $\Xi$, and $Q \in \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|}$ is a positive definite symmetric matrix which represents the cost of expansion. In the deterministic case when $\Xi$ is a singleton and $Q = 0$ any vector flow solution $f = (f_r^*)_{r\in\mathcal{R}}$ to (4.9) is a Wardrop equilibrium satisfying

$$(\forall (o,d) \in \mathcal{O} \times \mathcal{D})\ (\forall r \in R_{o,d}) \qquad f_r^* > 0 \Rightarrow c_r = \min_{r' \in R_{o,d}} c_{r'}, \tag{4.10}$$

where $c_r = \sum_{a\in r} t_a(u_a^*)$ is the cost of the route $r \in R_{o,d}$ and, for every $a \in \mathcal{A}$, $u_a^* = \sum_{r\in R} N_{a,r} f_r^*$. Therefore in a Wardrop equilibrium, for every origin-destination pair, all used paths have minimal cost [4]. Therefore, the optimization problem in (4.9) aims to finding a Wardrop equilibrium with minimum expansion cost in arcs. Note that we can write (4.9) equivalently as

$$\min_{(x,f)\in C} \delta_{(M^{|\Xi|}\cap W)\times \bigtimes V_\xi^+}(x,f) + \delta_{\bigtimes H_\xi}(x_\xi, Nf_\xi) + \sum_{\xi\in\Xi} p_\xi \left[\frac{1}{2} x_\xi^\top Q x_\xi + \varphi_\xi(Nf_\xi)\right]. \tag{4.11}$$

Under the assumption that is not empty we denote by $Z_1$ the set of solutions of (4.11).

**Proposition 4.1** *Consider the following operators*

$$\begin{aligned}
F :\ & \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|} \mapsto \mathbb{R}\\
& (x,f) \mapsto \delta_{(M^{|\Xi|}\cap W)\times \bigtimes_{\xi\in\Xi} V_\xi^+}(x,f)\\
G :\ & \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|\mathcal{A}||\Xi|} \mapsto \mathbb{R}\\
& (x,u) \mapsto \delta_{\bigtimes_{\xi\in\Xi} H_\xi}(x_\xi, u_\xi)\\
H :\ & \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|} \mapsto \mathbb{R}\\
& (x,f) \mapsto \sum_{\xi\in\Xi} p_\xi \left[\frac{1}{2} x_\xi^\top Q x_\xi + \varphi_\xi(Nf_\xi)\right]\\
L :\ & \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|} \mapsto \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|\mathcal{A}||\Xi|}\\
& (x,f) \mapsto (x_\xi, Nf_\xi)_{\xi\in\Xi}.
\end{aligned} \tag{4.12}$$

*Set* $\mu = \max_{\xi\in\Xi} p_\xi^2 \left(\max\left\{\|N\|^4 \max_{a\in\mathcal{A}} \beta_{a,\xi}^2, \|Q\|^2\right\}\right)^{-\frac{1}{2}}$, *set* $\mathcal{G} = \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|\mathcal{A}||\Xi|}$, *and set* $\mathcal{H} = \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|R||\Xi|}$. *Then the following hold:*

(i) $F \in \Gamma_0(\mathcal{H})$ *and* $G \in \Gamma_0(\mathcal{G})$.

(ii) $L$ *is a nonzero bounded linear operator and* $\|L\|^2 \le \max\{1, \|N\|^2\}$.

(iii) $H$ *is a convex and differentiable function with* $\mu^{-1}$*- Lipschitzian gradient.*

*Moreover,* (4.11) *can be written as*

$$\min_{(x,f)\in C} F(x,f) + G(L(x,f)) + H(x,f), \tag{$\mathcal{P}_1$}$$

*which is a particular instance of Problem 1.1.*

*Proof.*

(i) $F \in \Gamma_0(\mathcal{H})$: The convexity of $W$, $M$ and $V_\xi^+$ $(\forall \xi \in \Xi)$ follows directly from the definition of every set. Define the function $\phi_1 : \mathbb{R}^{|\mathcal{A}||\Xi|} \mapsto \mathbb{R} :\ x \mapsto \sum_{(\xi_1,\xi_2)\in\Xi^2} \|x_{\xi_1} - x_{\xi_2}\|^2$, we have that $W = \phi_1^{-1}(0)$ is closed. Similarly, if we define

$$\phi_{2\xi} : \mathbb{R}^{|R||\Xi|} \mapsto \mathbb{R}^{|\mathcal{O}||\mathcal{D}|} :\ f \mapsto \left( \sum_{r\in R_{(o,d)}} f_{r,\xi} \right)_{(o,d)\in\mathcal{O}\times\mathcal{D}}, \tag{4.13}$$

then, for every $\xi \in \Xi$ we have that $V_\xi^+ = \phi_{2\xi}^{-1}((h_{od,\xi})_{(o,d)\in\mathcal{O}\times\mathcal{D}})$ is closed.

Finally, $\left(M^{|\Xi|} \cap W\right) \times \bigtimes_{\xi\in\Xi} V_\xi^+$ is a nonempty closed convex set and the result follows.

(ii) $G \in \Gamma_0(\mathcal{G})$: Define the function $\phi_{3\xi} : \mathbb{R}^{|\mathcal{A}||\Xi|} \times \mathbb{R}^{|\mathcal{A}||\Xi|} \mapsto \mathbb{R}^{|\mathcal{A}||\Xi|} :\ (x,u) \mapsto u - x$ then $H_\xi = \phi_{3\xi}^{-1}(\bigtimes_{a\in\mathcal{A}}]-\infty, c_{a,\xi}])$ is closed for every $\xi \in \Xi$. Let $(x_1,u_1),(x_2,u_2) \in H_\xi$ and $\lambda \in [0,1]$ we have that

$$(\forall \xi \in \Xi) \quad \lambda(u_1 - x_1) + (1-\lambda)(u_2 - x_2) \le \lambda c_\xi + (1-\lambda)c_\xi = c_\xi, \tag{4.14}$$

where $c_\xi = (c_{a,\xi})_{a\in\mathcal{A}}$. Since $\bigtimes_{\xi\in\Xi} H_\xi$ is the product of nonempty closed convex sets, we have that $G \in \Gamma_0(\mathcal{G})$.

(iii) $L$ is a nonzero bounded linear operator: Clearly $L$ is a nonzero linear operator. It follows from (4.12) that

$$\begin{aligned}(\forall (x,f) \in \mathcal{G}) \quad \|L(x,f)\|^2 &= \|x\|^2 + \sum_{\xi\in\Xi} \|Nf_\xi\|^2 \\ &\le (\|x\|^2 + \|N\|^2 \sum_{\xi\in\Xi} \|f_\xi\|^2) \\ &\le \max\{1, \|N\|^2\}(\|x\|^2 + \|f\|^2) \\ &= \max\{1, \|N\|^2\}\|(x,f)\|^2 \end{aligned} \tag{4.15}$$

and the result follows. Moreover, the conditions (iii) of Theorem (2.1) is satisfied.

(iv) $H$ is a convex and differentiable with $\mu^{-1}$- Lipschitzian gradient function: Using The Fundamental Theorem of Calculus and the Chain Rule respectively, we obtain

$$\begin{aligned}(\forall (x,u) \in \mathcal{H}) \quad \nabla H(x,f) &= \Big(p_\xi Q x_\xi, p_\xi N^\top \nabla \varphi_\xi (N f_\xi)\Big)_{\xi\in\Xi} \\ &= \Big(p_\xi Q x_\xi, p_\xi N^\top (t_{a,\xi}(N_a f_\xi))_{a\in\mathcal{A}}\Big)_{\xi\in\Xi}. \end{aligned} \tag{4.16}$$

Using the last results, we have that for every $(x^1,f^1),(x^2,f^2)\in\mathcal{H}$

$$\begin{aligned}
\|\nabla H(x^1,f^1)-\nabla H(x^2,f^2)\|^2 &= \sum_{\xi\in\Xi}\left(\|p_\xi N^t((t_{a,\xi}(N_a f^1_\xi)-t_{a,\xi}(N_a f^2_\xi))_{a\in\mathcal{A}})\|^2+\|p_\xi(Qx^1_\xi-Qx^2_\xi)\|^2\right)\\
&\le \sum_{\xi\in\Xi}p_\xi^2\left(\|N\|^2\|((t_{a,\xi}(N_a f^1_\xi)-t_{a,\xi}(N_a f^2_\xi))_{a\in\mathcal{A}})\|^2+\|Q\|^2\|x^1_\xi-x^2_\xi\|^2\right)\\
&\le \sum_{\xi\in\Xi}p_\xi^2\left(\|N\|^2\sum_{a\in\mathcal{A}}\left(\beta^2_{a,\xi}\|N_a\|^2\|f^1_\xi-f^2_\xi\|^2\right)+\|Q\|^2\|x^1_\xi-x^2_\xi\|^2\right)\\
&\le \sum_{\xi\in\Xi}p_\xi^2\left(\|N\|^4\max_{a\in\mathcal{A}}\beta^2_{a,\xi}\|f^1_\xi-f^2_\xi\|^2+\|Q\|^2\|x^1_\xi-x^2_\xi\|^2\right)\\
&\le \sum_{\xi\in\Xi}p_\xi^2\left(\max\left\{\|N\|^4\max_{a\in\mathcal{A}}\beta^2_{a,\xi},\|Q\|^2\right\}\left(\|f^1_\xi-f^2_\xi\|^2+\|x^1_\xi-x^2_\xi\|^2\right)\right)\\
&\le \max_{\xi\in\Xi}p_\xi^2\left(\max\left\{\|N\|^4\max_{a\in\mathcal{A}}\beta^2_{a,\xi},\|Q\|^2\right\}\right)\left(\|f^1-f^2\|^2+\|x^1-x^2\|^2\right)\\
&\le \mu^{-2}\left(\|(x^1,f^1)-(x^2,f^2)\|^2\right) \qquad (4.17)
\end{aligned}$$

Since each $t_{a,\xi}$ is an increasing function we have that $\nabla^2\phi_\xi(u_\xi)=diag((t'(u_{a,\xi}))_{a\in\mathcal{A}})$ is positive semi-definite matrix, therefore, $\phi_\xi\circ N$ is convex, for every $\xi\in\Xi$. Then, since $H$ is the sum of convex functions we obtain that $H$ is convex.

Finally, it follows from (4.12) that (4.11) can be written equivalently as $(\mathcal{P}_1)$. □

In what follows we assume that the Slater condition holds, i.e., there exists $(x,f)\in\left(M^{|\Xi|}\cap W\right)\times\mathbb{R}_+^{|R||\Xi|}$, satisfying the constraints in (4.9) such that

$$(\forall\xi\in\Xi)\,(\forall a\in\mathcal{A})\qquad \sum_{r\in R}N_{a,r}f_{r,\xi}-x_{a,\xi}<c_{a,\xi}. \qquad (4.18)$$

Then by [3, Proposition 27.21] the qualification condition (1.10) is satisfied.

## 4.3 Alternating and Random binary projections into the arc capacity expansion problem of a directed graph.

In this section, we use the Theorem 2.1 to solve the problem $(\mathcal{P}_1)$.

**Corollary 4.2** *Consider the setting of Problem $(\mathcal{P}_1)$. Let $\gamma > 0$, let $\tau \in ]0, 2\mu[$, where $\mu = \max_{\xi\in\Xi} p_\xi^2 \left(\max\left\{\|N\|^4 \max_{a\in\mathcal{A}} \beta_{a,\xi}^2, \|Q\|^2\right\}\right)^{-\frac{1}{2}}$, let $(u^0, v^0) \in \mathcal{G}$, let $(x^0, f^0)$, $(\overline{x}^0, \overline{f}^0) \in \mathcal{H}^2$ be such that $(x^0, f^0) = (\overline{x}^0, \overline{f}^0)$ and, let $(\epsilon_k)_{k\in\mathbb{N}}$ be a sequence of independent $\{0,1\}-$Bernoulli random variables such that $\mathbb{P}(\epsilon_k^{-1}(\{1\})) = \pi_k$, for every $k \in \mathbb{N}$. Consider the following routine*

$$
(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l}
\left(\widetilde{u}^{k+1}, \widetilde{v}^{k+1}\right) = \left(u_\xi^k + \gamma \bar{x}_\xi^k,\ v_\xi^k + \gamma N \bar{f}_\xi^k\right)_{\xi\in\Xi} \\
\left(u^{k+1}, v^{k+1}\right) = \left(\left(\widetilde{u}_\xi^{k+1},\ \widetilde{v}_\xi^{k+1}\right) - \gamma P_{H_\xi}\left(\gamma^{-1}\left(\widetilde{u}_\xi^{k+1}, \widetilde{v}_\xi^{k+1}\right)\right)\right)_{\xi\in\Xi} \\
(\widetilde{p}^{k+1}, \widetilde{g}^{k+1}) = \left(x_\xi^k - \tau u_\xi^{k+1} - \tau p_\xi Q x_\xi^k,\ f_\xi^k - \tau N^\top v_\xi^{k+1} - \tau p_\xi \psi_\xi(f_\xi^k)\right)_{\xi\in\Xi} \\
(p^{k+1}, g^{k+1}) = \left(P_\xi\left(P_{M^{|\Xi|}\cap W}\left(\widetilde{p}^{k+1}\right)\right),\ P_{V_\xi^+}\left(\widetilde{g}_\xi^{k+1}\right)\right)_{\xi\in\Xi} \\
(x^{k+1}, f^{k+1}) = \left(p^{k+1}, g^{k+1}\right) + \epsilon_{k+1}\left(P_{C_{i(k+1)}}\left(p^{k+1}, g^{k+1}\right) - \left(p^{k+1}, g^{k+1}\right)\right) \\
(\bar{x}^{k+1}, \bar{f}^{k+1}) = \left(x^{k+1}, f^{k+1}\right) + \left(p^{k+1}, g^{k+1}\right) - \left(x^k, f^k\right),
\end{array}\right. \tag{4.19}
$$

*where $(\forall \xi \in \Xi)\ \ \psi_\xi : \mathbb{R}^{|R|} \mapsto \mathbb{R}^{|R|} : f \mapsto N^\top(t_{a,\xi}(N_a f))_{a\in\mathcal{A}}$ and $i(k) := (k \bmod |\mathcal{A}||\Xi|) + 1$. Assume that the following hold:*

(i) $0 < \inf_{k\in\mathbb{N}} \pi_k$ *and* $\max\{1, \|N\|^2\} < \frac{1}{\gamma}\left(\frac{1}{\tau} - \frac{1}{2\mu}\right)$.

*Then $((x^k, f^k))_{k\in\mathbb{N}}$ converges $P$-a.s. to a $Z_1$-valued random variable $(x, f)$.*

*Proof.*

It follows from Proposition 4.1, [3, Proposition 24.11 & Proposition 24.8(ix)], and (4.19) that

$$
\left(\widetilde{u}^{k+1}, \widetilde{v}^{k+1}\right) = \left(u^k,\ v^k\right) + \gamma L\left(\bar{x}^k,\ \bar{f}^k\right) \tag{4.20}
$$

$$
\begin{aligned}
\left(u^{k+1}, v^{k+1}\right) &= \left(\widetilde{u}^{k+1},\ \widetilde{v}^{k+1}\right) - \gamma P_{\bigtimes_{\xi\in\Xi} H_\xi}\left(\gamma^{-1}\left(\widetilde{u}^{k+1}, \widetilde{v}^{k+1}\right)\right) \\
&= \left(\widetilde{u}^{k+1},\ \widetilde{v}^{k+1}\right) - \gamma\, \mathrm{prox}_{\gamma^{-1}G}\left(\gamma^{-1}\left(\widetilde{u}^{k+1}, \widetilde{v}^{k+1}\right)\right) \\
&= \mathrm{prox}_{\gamma G^*}\left(\left(\widetilde{u}^{k+1}, \widetilde{v}^{k+1}\right)\right) \\
&= \mathrm{prox}_{\gamma G^*}\left(\left(u^k,\ v_\xi^k\right) + \gamma L\left(\bar{x}^k,\ \bar{f}^k\right)\right)
\end{aligned} \tag{4.21}
$$

$$(\widetilde{p}^{k+1}, \widetilde{g}^{k+1}) = \left(x^k,\ f^k\right) - \tau L^* \left(u^{k+1}, v^{k+1}\right) - \tau \nabla H \left(x^k, f^k\right) \tag{4.22}$$

$$\begin{aligned}(p^{k+1}, g^{k+1}) &= P_{\left(M^{|\Xi|} \cap W\right) \times \bigtimes_{\xi \in \Xi} V_\xi^+} \left(\widetilde{p}^{k+1}, \widetilde{g}^{k+1}\right) \\ &= \mathrm{prox}_{\tau F} \left(\widetilde{p}^{k+1}, \widetilde{g}^{k+1}\right) \\ &= \mathrm{prox}_{\tau F} \left(\left(x^k,\ f^k\right) - \tau L^* \left(u^{k+1}, v^{k+1}\right) - \tau \nabla H \left(x^k, f^k\right)\right).\end{aligned} \tag{4.23}$$

Note that (4.19) can be written equivalently

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \left(u^{k+1}, v^{k+1}\right) = \mathrm{prox}_{\gamma G^*} \left(\left(u^k,\ v^k\right) + \gamma L \left(\bar{x}^k,\ \bar{f}^k\right)\right) \\ \left(p^{k+1}, g^{k+1}\right) = \mathrm{prox}_{\tau F} \left(\left(x^k,\ f^k\right) - \tau L^* \left(u^{k+1}, v^{k+1}\right) - \tau \nabla H \left(x^k, f^k\right)\right) \\ \left(x^{k+1}, f^{k+1}\right) = \left(p^{k+1}, g^{k+1}\right) + \epsilon_{k+1} \left(P_{C_{i(k+1)}} \left(p^{k+1}, g^{k+1}\right) - \left(p^{k+1}, g^{k+1}\right)\right) \\ \left(\bar{x}^{k+1}, \bar{f}^{k+1}\right) = \left(x^{k+1}, f^{k+1}\right) + \left(p^{k+1}, g^{k+1}\right) - \left(x^k, f^k\right). \end{array} \right.$$

Finally using Proposition 4.1, Remark 2.2.3, and (i) the conditions (i)-(iii) of Theorem (2.1) holds and we conclude that $((x^k, f^k))_{k \in \mathbb{N}}$ converges P-a.s. to a $Z_1$-valued random variable $(x, f)$. □

**Remark 4.3** In order to implement the algorithm (4.2) we need to calculate the following projections:

1. It follows from [3, Proposition 24.11] that, for every $(x, u) \in \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|}$ and $\xi \in \Xi$

$$P_{H_\xi}(x, u) = \left(P_{H_{a,\xi}}(x_{a,\xi}, u_{a,\xi})\right)_{a \in \mathcal{A}} \tag{4.24}$$

   and applying [3, Example 29.20] we have that, for every $a \in \mathcal{A}$ and $\xi \in \Xi$,

$$P_{H_{a,\xi}} : \mathbb{R}^2 \mapsto \mathbb{R}^2 : (x, u) \mapsto \begin{cases} \left(\frac{x+u-c_{a,\xi}}{2}, \frac{x+u+c_{a,\xi}}{2}\right) & \text{if } x - u + c_{a,\xi} < 0 \\ (x, u) & \text{if } x - u + c_{a,\xi} \geq 0 \end{cases} \tag{4.25}$$

2. It follows [3, Theorem 3.16] that

$$P_{M^{|\Xi|} \cap W} : \mathbb{R}^{|\mathcal{A}||\Xi|} \mapsto \mathbb{R}^{|\mathcal{A}||\Xi|} : x \mapsto \left(P\left(\frac{\sum_{\xi \in \Xi} x_\xi}{|\Xi|}\right)\right)_{\xi \in \Xi}, \tag{4.26}$$

   where

$$P : \mathbb{R}^{|\mathcal{A}|} \mapsto \mathbb{R}^{|\mathcal{A}|} : x \mapsto \left(\min\left(M_a, \max\left(0, x_a\right)\right)\right)_{a \in \mathcal{A}}. \tag{4.27}$$

3. Note that

$$(\forall \xi \in \Xi) \quad V_\xi^+ = \bigtimes_{(o,d) \in \mathcal{O} \times \mathcal{D}} \underbrace{\left\{ f \in \mathbb{R}_+^{|R_{o,d}|} \;\middle|\; \sum_{r \in R_{(o,d)}} f_r = h_{od,\xi} \right\}}_{V_{od,\xi}^+} \tag{4.28}$$

and it follows from [3, Proposition 24.11] that

$$P_{V_\xi^+} : \mathbb{R}^{|R|} \mapsto \mathbb{R}^{|R|} : f \mapsto (P_{V^+_{od,\xi}}(f_{R_{o,d}}))_{(o,d)\in\mathcal{O}\times\mathcal{D}}. \tag{4.29}$$

What is left is to calculate $P_{V^+_{od,\xi}}$ for every $(o,d) \in \mathcal{O}\times\mathcal{D}$, to do this, we use the algorithm proposed in [15], which allows us to calculate the projection on sets of the form $\{x \in \mathbb{R}^n \mid b^t x = r \ \wedge\ l \le x \le u\}$, Formulacion in our case $n = |R_{o,d}|$, $r = h_{od,\xi}$, $b = 1_{|R_{o,d}|}$, $l = 0_{|R_{o,d}|}$ and $u = +\infty$.

On the other hand, let $l \in \{1,\dots,|\Xi|\}$ and consider the following sets

$$(\forall j \in \{1,\dots,|\mathcal{A}||\Xi|\}) \ \ \Delta_j^l := \bigcap_{i=j}^{j+l-1} C_{a(j),\xi(i,j)}, \tag{4.30}$$

where $a(j) := ((j-1) \bmod |\mathcal{A}|)+1$ and $\xi(i,j) = \left(\left(\frac{(j-1)-((j-1) \bmod |\mathcal{A}|)}{|\mathcal{A}|} + i - j\right) \bmod |\Xi|\right) + 1$. The next example illustrates the indexation of the sets $\Delta_j^l$ for $l = 2$ and $j = 118, 119, 120$.

**Example 4.4** Let $\mathcal{A} = \{1,\dots,7\}$, let $\Xi = \{1,\dots,18\}$ and set $l = 2$. Then, for $j = 118, 119, 120$ we have the following 3 sets

$$\Delta^2_{118} = C_{6,17}\cap C_{6,18} \quad \Delta^2_{119} = C_{7,17}\cap C_{7,18} \quad \Delta^2_{120} = C_{1,18}\cap C_{1,1}.$$

Consider the setting of the problem $(\mathcal{P}_1)$ and set $\left(x^0, f^0\right) = \left(\bar{x}^0, \bar{f}^0\right) = 0_{\mathcal{H}}$, $\left(u^0, v^0\right) = 0_{\mathcal{G}}$. Consider the following 10 instances of the algorithm (4.19).

The **Algorithm 1** is the routine (4.19), in the case when, for every $k \in \mathbb{N}$, $\epsilon_k^{-1}(\{0\}) = \Omega$ and $C = \mathcal{H}$. The **Algorithms 2, 3,** and **4** are the routine (4.19), in the case when, for every $k \in \mathbb{N}$, $\epsilon_k^{-1}(\{1\}) = \Omega$ and $C = \Delta_1^\alpha$, with $\alpha = 1, 9$, and 18 respectively. Note that the algorithm 1 are the algorithms proposed in [16, 24] and the algorithms 2-4 is the algorithm proposed in [7].

The **Algorithms 5, 6,** and **7** are the routine (4.19), in the case when, for every $k \in \mathbb{N}$, $\pi_k = 0.5$ and $C = \bigcap_{i=1}^{|\mathcal{A}||\Xi|}\Delta_i^l$, with $l = 1, 9$, and 18 respectively. The **Algorithms 8, 9,** and **10** are the routine (4.19), in the case when, for every $k \in \mathbb{N}$, $\epsilon_k^{-1}(\{1\}) = \Omega$ and $C = \bigcap_{i=1}^{|\mathcal{A}||\Xi|}\Delta_i^l$, with $l = 1, 9$, and 18 respectively.

**Remark 4.5**

1. Note that

$$(\forall j \in \{1,\dots,|\mathcal{A}||\Xi|\}) \quad \delta_{\Delta_j^l}(x,f) = \bigoplus_{i=j}^{j+l-1} \delta_{\left(C_{a(j),\xi(i,j)}\right)_{\xi(i,j)}}(x_{\xi(i,j)}, f_{\xi(i,j)}) \tag{4.31}$$

and therefore $P_{\Delta_j^l}(x,f) = \left(T_{\xi_1}^{j,l}(x_{\xi_1}, f_{\xi_1})\right)_{\xi_1\in\Xi}$, where

$$(\forall \xi \in \Xi) \ \ T_{\xi_1}^{j,l} := \begin{cases} P_{\left(C_{a(j),\xi}\right)_\xi} & \text{if } \xi_1 \in \bigcup_{i=j}^{j+l-1} \xi(i,j) \\ Id & \text{otherwise.} \end{cases} \tag{4.32}$$

which can be calculated using [3, Proposition 24.11].

## 4.4 Randomized Kaczmarz projections into the arc capacity expansion problem of a directed graph.

In this section, we use the Theorem 3.1 to solve the problem $(\mathcal{P}_1)$.

**Corollary 4.6** *Consider the setting of Problem $(\mathcal{P}_1)$. Let $\gamma > 0$, let $\tau \in ]0, 2\mu[$, where $\mu = \max_{\xi\in\Xi} p_\xi^2 \left(\max\left\{\|N\|^4 \max_{a\in\mathcal{A}} \beta_{a,\xi}^2, \|Q\|^2\right\}\right)^{-\frac{1}{2}}$, let $(u^0, v^0) \in \mathcal{G}$, let $(x^0, f^0)$, $(\overline{x}^0, \overline{f}^0) \in \mathcal{H}^2$ be such that $(x^0, f^0) = (\overline{x}^0, \overline{f}^0)$ and, set $I = \{0, 1, ..., m\}$. Let $(\epsilon_k)_{k\in\mathbb{N}}$ be a sequence of independent $I$-valued random variables. Consider the following routine*

$$
(\forall k \in \mathbb{N}) \quad \left\lfloor
\begin{array}{l}
(\widetilde{u}^{k+1}, \widetilde{v}^{k+1}) = \left(u_\xi^k + \gamma \bar{x}_\xi^k,\ v_\xi^k + \gamma N \bar{f}_\xi^k\right)_{\xi\in\Xi} \\
(u^{k+1}, v^{k+1}) = \left(\left(\widetilde{u}_\xi^{k+1},\ \widetilde{v}_\xi^{k+1}\right) - \gamma P_{H_\xi}\left(\gamma^{-1}\left(\widetilde{u}_\xi^{k+1}, \widetilde{v}_\xi^{k+1}\right)\right)\right)_{\xi\in\Xi} \\
(\widetilde{p}^{k+1}, \widetilde{g}^{k+1}) = \left(x_\xi^k - \tau u_\xi^{k+1} - \tau p_\xi Q x_\xi^k,\ f_\xi^k - \tau N^\top v_\xi^{k+1} - \tau p_\xi \psi_\xi(f_\xi^k)\right)_{\xi\in\Xi} \\
(p^{k+1}, g^{k+1}) = \left(P_\xi\left(P_{M^{|\Xi|}\cap W}\left(\widetilde{p}^{k+1}\right)\right),\ P_{V_\xi^+}\left(\widetilde{g}_\xi^{k+1}\right)\right)_{\xi\in\Xi} \\
(x^{k+1}, f^{k+1}) = P_{C_{\epsilon_{k+1}}}\left(p^{k+1}, g^{k+1}\right) \\
(\bar{x}^{k+1}, \bar{f}^{k+1}) = \left(x^{k+1}, f^{k+1}\right) + \left(p^{k+1}, g^{k+1}\right) - \left(x^k, f^k\right),
\end{array}
\right. \tag{4.33}
$$

*where $(\forall \xi \in \Xi)$ $\psi_\xi : \mathbb{R}^{|R|} \mapsto \mathbb{R}^{|\mathcal{A}|} : f \mapsto N^\top(t_{a,\xi}(N_a f))_{a\in\mathcal{A}}$ and $C_0 = \mathcal{H}$. Assume that the following hold:*

(i) $(\forall i \in I\backslash\{0\})$ $0 < \inf_{k\in\mathbb{N}} \pi_k^i$, *where* $(\forall i \in I)(\forall k \in \mathbb{N})$ $\pi_k^i = \mathbb{P}(\epsilon_k^{-1}(\{i\}))$ *and* $\max\{1, \|N\|^2\} < \frac{1}{\gamma}\left(\frac{1}{\tau} - \frac{1}{2\mu}\right)$.

*Then $((x^k, f^k))_{k\in\mathbb{N}}$ converges $P$-a.s. to a $Z_1$-valued random variable $(x, f)$.*

On the other hand, let $l \in \{1, \ldots, |\Xi|\}$ and let $j$ a bijection from $D_l \times \mathcal{A}^l$ to $I\backslash\{0\}$, where

$$D_l = \{(y_1, \ldots, y_l) \in \Xi^l \mid (\forall i, j \in \{1, \ldots, l\}) \text{ if } i \neq j \Rightarrow y_i \neq y_j\}. \tag{4.34}$$

Let's define the following nonempty closed convex sets

$$(\forall y \in D_l)(\forall x \in \mathcal{A}^l)\ \ K_{j(y,x)}^l = \bigcap_{i=1}^{l} C_{x_i, y_i} \tag{4.35}$$

Consider the setting of the problem $(\mathcal{P}_1)$ and set $\left(x^0, f^0\right) = \left(\bar{x}^0, \bar{f}^0\right) = 0_{\mathcal{H}}$, $\left(u^0, v^0\right) = 0_{\mathcal{G}}$. Consider the following 3 instances of the algorithm (4.33). The **Algorithms 11, 12,** and **13** are the routine (4.33), in the case when $C = \bigcap_{i\in I} K_i^l$, with $l = 1, 9$, and 18 respectively, and for every $k \in \mathbb{N}$, $\epsilon_k$ is $I$-valued random variable such that $\pi_0^k = 0$, and for every $i \in I$, $\pi_i^k = \frac{1}{|I|}$.

**Remark 4.7** In the case when $C = \mathcal{H}$ the routine (4.33) is the **Algorithm 1**.

## 4.5 Numerical experience

Consider the following graph:

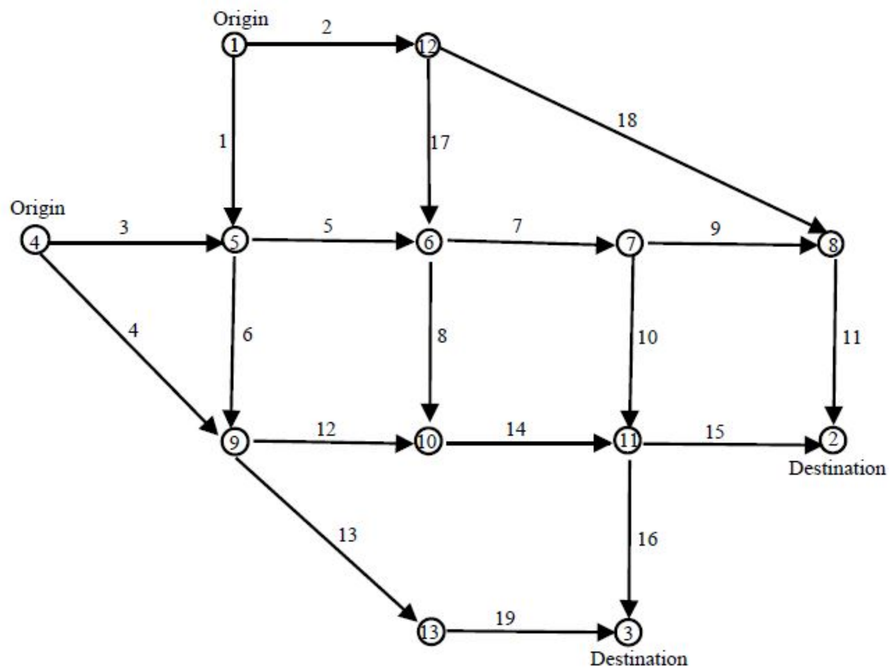


Figure 4.1: Graph of the arc capacity problem

Consider the Algorithms 4.19 and 4.33 in the case when $|\Xi| = 18$, $p_\xi = \frac{1}{|\Xi|}$, $\forall \xi \in \Xi$. According to the Figure 4.1 we have that $|\mathcal{A}| = 19$, $|\mathcal{O}| = |\mathcal{D}| = 2$, $|R_{1,2}| = 8$, $|R_{4,3}| = |R_{1,3}| = 6$, $|R_{4,2}| = 5$ and $|R| = 25$.

Consider the following parameters

(i) Capacity of every arc:

$$\begin{aligned} c_\xi :=& 110 \cdot (10\ \ 4.4\ \ 1.4\ \ 10\ \ 3\ \ 4.4\ \ 10\ \ 2\ \ 2\ \ 47\ \ 7\ \ 7\ \ 7\ \ 4\ \ 3.5\ \ 2.2\ \ 4.4\ \ 7) \\ &+ \underbrace{(15\ 6.6\ 2.1\ 15\ 4.5\ 6.6\ 15\ 3\ 3\ 6\ 10.5\ 10.5\ 10.5\ 10.5\ 6\ 5.25\ 3.3\ 6.6\ 10.5)}_{\bar{c}} \cdot X_1 \qquad (\forall \xi \in \Xi). \end{aligned} \tag{4.36}$$

(ii) Origin-Destination Demand:

$$d_\xi := (300\ 700\ 500\ 350) + (120\ 120\ 120\ 120) \cdot X_2 \qquad (\forall \xi \in \Xi), \tag{4.37}$$

where $X_1 \sim \text{Beta}(20, 20)$ and $X_2 \sim \text{Beta}(50, 10)$.

(iii) Arc expansion capacity limit: $M = 200\bar{c}$.

(iv) Cost matrix: $Q = \text{Id}_{|\mathcal{A}|}$.

$\alpha = 0$

| Alg | Time [s] | Iter |
|---|---|---|
| Alg 1 | 110.8630 | 34568 |

$\alpha = 1$

| Alg | Time [s] | Iter |
|---|---|---|
| Alg. 2 | 109.0810 | 34535 |
| Alg. 5 | 107.7653 | 34236 |
| Alg. 8 | 106.6755 | 33773 |
| Alg. 11 | 107.0495 | 33752 |

$\alpha = 9$

| Alg | Time [s] | Iter |
|---|---|---|
| Alg 3 | 109.6725 | 34115 |
| Alg 6 | 99.0522 | 31311 |
| Alg 9 | 92.0858 | 28742 |
| Alg 12 | 94.0510 | 29236 |

$\alpha = 18$

| Alg | Time [s] | Iter |
|---|---|---|
| Alg 4 | 101.7377 | 31234 |
| Alg 7 | 91.7839 | 28618 |
| Alg 10 | 82.2959 | 25080 |
| Alg 13 | 84.4937 | 25703 |

Table 4.1: The average execution time and the average number of iterations of each method. In the case of a graph with 25 routes, 19 arcs and 18 scenarios.

Consider the time travel function:

$$(\forall \xi \in \Xi)(\forall a \in \mathcal{A}) \quad t_{a,\xi}(u) := \eta_a + \tau_a \frac{u}{c_{a,\xi}}, \tag{4.38}$$

where

$$\eta := (7 \ \ 9 \ \ 9 \ \ 12 \ \ 3 \ \ 9 \ \ 5 \ \ 13 \ \ 5 \ \ 9 \ \ 9 \ \ 10 \ \ 9 \ \ 6 \ \ 9 \ \ 8 \ \ 7 \ \ 14 \ \ 11) \tag{4.39}$$

and $\tau := 0.15\eta$. Hence, for every $\xi \in \Xi$ and $\alpha \in \mathcal{A}$, $\beta_{a,\xi} := \frac{\tau_a}{c_{a,\xi}}$ .

We consider stop criteria when the relative error of every iteration is less than a tolerance equal to $10^{-10}$, where the relative error of every iteration is defined by

$$(\forall k \in \mathbb{N}) \ \ e_k = \sqrt{\frac{\|x^{k+1} - x^k\|^2 + \|f^{k+1} - f^k\|^2 + \|u^{k+1} - u^k\|^2 + \|v^{k+1} - v^k\|^2}{\|x^k\|^2 + \|f^k\|^2 + \|u^k\|^2 + \|v^k\|^2}}. \tag{4.40}$$

We test each algorithm and show in the Table (4.1) the average execution time and the number of average iterations of each method, obtained by 20 random realizations of vectors $(c_\xi)_{\xi \in \Xi}$ and $(d_\xi)_{\xi \in \Xi}$. The random generated vectors are obtained via the *random* function of MATLAB, using the same seed.

Note the algorithms that include randomized (Alg. 11, 12, and 13) and alternating projections (Alg. 8, 9, and 10) are similar in terms of the execution time and the number of iterations, with a small advantage for the Alternating algorithms. In addition, both algorithms have significant improvements with respect to Algorithm 1. In Table 4.1, we can see that algorithms that include random alternating projections (Alg. 5, 6, and 7) are more efficient than Algorithm 1 and fixed projections (Alg. 2, 3, and 4). However, they are not faster than randomized and alternating projections. Similarly, fixed projections are more efficient than Algorithm 1.

In Table 4.1, we can see the number of projections on an inequality is directly related to the efficiency of the algorithm with the exception of Algorithm 3. In the case of alternating projections on 18 inequalities, there is an improvement about 35 % in the execution time and 38 % the number of iterations, while the randomized algorithm there is an improvement of approximately 31 % in the time of execution and 35% in the number of iterations.

| Arco | $max_\xi(u_{a,\xi} - c_{a,\xi})$ | $x_a$ | Arco | $max_\xi(u_{a,\xi} - c_{a,\xi})$ | $x_a$ |
|---|---|---|---|---|---|
| 1 | -394.91 | 0.00 | 11 | -218.34 | 0.00 |
| 2 | 18.20 | 18.20 | 12 | -164.31 | 0.00 |
| 3 | -6.95 | 0.00 | 13 | 36.20 | 36.20 |
| 4 | -187.97 | 0.00 | 14 | -164.31 | 0.00 |
| 5 | 24.92 | 24.92 | 15 | 17.67 | 17.67 |
| 6 | 15.62 | 15.62 | 16 | 68.42 | 68.42 |
| 7 | -633.86 | 0.00 | 17 | -126.34 | 0.00 |
| 8 | -221.02 | 0,00 | 18 | -78.68 | 0.00 |
| 9 | -73,31 | 0,00 | 19 | 36.20 | 36.20 |
| 10 | -95,52 | 0,00 | | | |

Table 4.2: The expression $\max_\xi(u_{a,\xi} - c_{a,\xi})$ represents the worst scenario in terms of arc flow. Note that in total 7 arcs are expanded and the expansion capacity coincides with the extra flow needed in the equilibrium for the worst scenario.

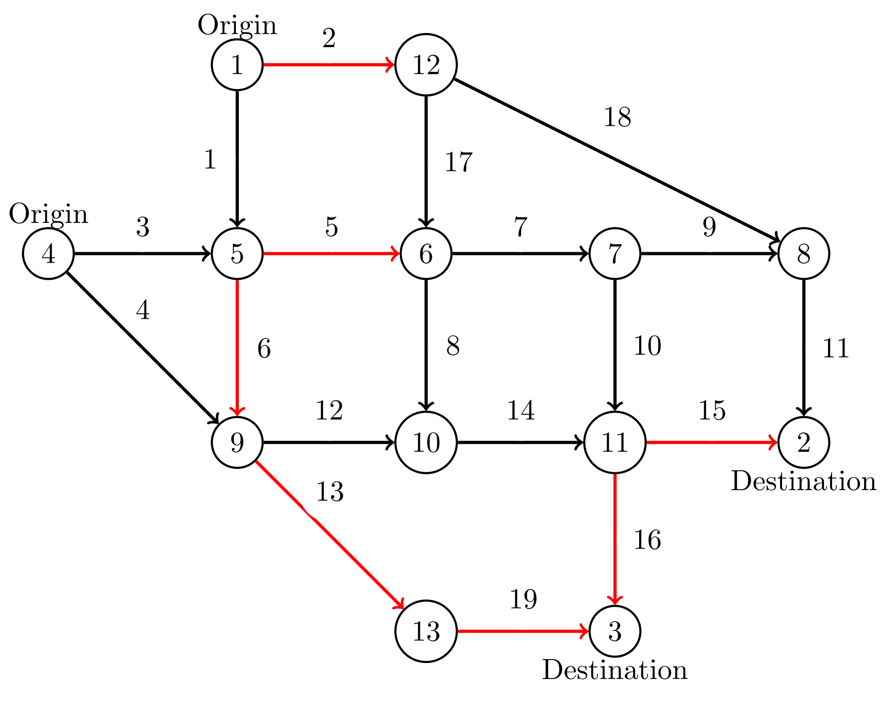


Figure 4.2: Graphical representation of the arcs with more flow according to the Table 4.2. The red arcs represent the arcs that in one or more scenario the maximum arc capacity is reached

Finally, in the Table 4.2 we can see that the expansion of each arc coincides with the worst scenario in terms of flows results in test 20, note that 7 arcs were expanded since there was a scenario such that the flow is greater than the capacity. In the Figure 4.2 the graphical representation of Table 4.2.

# Chapter 5

# Conclusions and future work

In this work, we provide two new primal-dual algorithms for solving constraints optimization problems. The first algorithm includes a random activation step over a cyclic projection scheme, while the second chooses a random element from the set of projection operators. We exploit the properties of Stochastic Quasi-Fejér sequences to prove the almost sure convergence of both algorithms. As special case the algorithms reduces to the method proposed in [7], Kaczmarz [19], randomized Kaczmarz [23], and cyclic projections [3, Theorem 16.47].

The importance of the alternating and randomized projections is illustrated with the arc capacity expansion problem, where the algorithms that include randomized and alternating projections are more efficient than the algorithms proposed in the literature (Alg. 1-4). The randomized Kaczmarz and alternating projections are more efficient than algorithms that include random alternating projections and both algorithms have similar execution time and number of iterations with an small advantage no greater than 4% for the alternating algorithms. The random alternating projections algorithms are more efficient than Alg. 2, 3, and 4. The number of projections on an inequality is directly related to the efficiency of the algorithm with the exception of Algorithm 3. In the case of alternating projections on 18 inequalities, there is an improvement about 35 % in the execution time and 38 % the number of iterations, while the randomized algorithm there is an improvement of approximately 31 % in the time of execution and 35 % in the number of iterations. We think that by including a larger number of scenarios to the problem, the difference in execution time and number of iterations will be greater. In test 20 of the algorithms we can see that 7 scenarios were expanded since there was an scenario such that the flow is greater than the capacity and the expansion is equal to the maximum between 0 and the worst flow minus capacity scenario.

Future work will consist in studying the following problem:

**Problem 5.1** Let $\{T_k\}_{k=1}^m$ be a family of $\alpha_k$-averaged nonexpansive operators. Let $L : \mathcal{H} \to \mathcal{G}$ be a nonzero linear bounded operator. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{G} \rightrightarrows \mathcal{G}$ be two maximally monotone operators, let $C : H \rightrightarrows H$ be an operator $\mu$-cocoercive, and let $D : G \rightrightarrows G$ be an maximally monotone operator and $\delta$ strongly monotone. The problem is to solve the primal-dual inclusions

$$\text{find} \quad x \in C := \bigcap_{i=1}^{m} \operatorname{Fix} T_i \quad \text{such that} \quad 0 \in Ax + L^*(B \,\square\, D)(Lx) + Cx \tag{$\mathcal{P}_2$}$$

Together with the dual inclusion

$$\text{find} \quad u \in G \quad \text{such that} \quad \exists x \in C \begin{cases} 0 \in Ax + L^*u + Cx \\ \\ u \in (B \,\Box\, D)(Lx), \end{cases} \tag{$\mathcal{D}_2$}$$

under the assumption that solutions exist. Where $(B \,\Box\, D) := (B^{-1} + D^{-1})^{-1}$.

In the case when $A = \partial f$, $B = \partial g$, $C = \nabla h$ with $\mu^{-1}$-Lipschitzian gradient, $D = \partial \delta_{\{0\}}$, and for all $i \in \{1, \dots, m\}$ $T_i = P_{C_i}$, the problem 5.1 reduces to the problem 1.2.

The main idea is to extend the algorithms 2.1 and 3.1 to solve the problem 5.1 including a stochastic composition step with $\alpha$-averaged nonexpansive operator, this approach allows us to include new operators, for example the composition of projections and/ or convex combination of projections.

Also, there are still several interesting questions, which need to be addressed in the future: (a) What is the most efficient manner to make projections? (b) is it possible unify both algorithms? (c) how to include the case when $(\epsilon_k)_{k\in\mathbb{N}}$ are continuous random variables? (d) how to include inconsistent linear system?

# Bibliography


[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Alternating proximal algorithms for weakly coupled convex minimization problems · applications to dynamical games and PDE's,* J. Convex Anal. 15, pp. 485· 506 (2008).

[2] H. Attouch, L.M. Briceño-Arias, and P.L. Combettes, *A parallel splitting method for coupled monotone inclusions,* SIAM J. Control Optim. 48, pp. 3246·3270 (2010).

[3] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.*2nd edn. Springer, New York (2017).

[4] M.J. Beckmann, C.B. McGuire, AND C.B. Winsten, *Studies in the economics of transportation.,* Yale University Press, New Haven (1956).

[5] L. M. Briceño-Arias and P. L. Combettes, *A monotone+skew splitting model for composite monotone inclusions in duality,* SIAM J. Optim., vol. 21, pp. 1230–1250, (2011).

[6] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, *Proximal algorithms for multicomponent image recovery problems,* J. Math. Imaging Vision, vol. 41, pp. 3–22, (2011).

[7] L.M. Briceño-Arias and S. López, *A projected primal-dual method for solving constrained monotone inclusions,* J. Optim. Theory Appl.,vol. 180, no. 3, pp. 907-924 (2019).

[8] A. Chambolle and T. Pock, *A first-order primal–dual algorithm for convex problems with applications to imaging,* J. Math. Imaging Vis., 40, pp. 120–145 (2011).

[9] X. Chen, T. K. Pong, and R.J.B. Wets, *Two-stage stochastic variational inequalities: an erm-solution procedure.,* Mathematical Programming, pp. 165(1):71·111 (2017).

[10] P. L. Combettes, Đinh Dũng, and B. C. Vũ, *Dualization of signal recovery problems,* Set-Valued Anal., vol. 18, pp. 373–404 (2010).

[11] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing,* in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering,* (H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Editors), pp. 185-212. Springer, New York (2011).

[12] P. L. Combettes and J.-C. Pesquet, *Stochastic approximations and perturbations in forward-backward splitting for monotone operators,* Pure Appl. Funct. Anal., pp. 13–37 (2016).

[13] P. L. Combettes and J.-C. Pesquet, *Stochastic Quasi-Fejér block-coordinate fixed point iterations with random sweeping,* SIAM J. Optim., vol. 25, pp. 1221–1248 (2015).

[14] P. L. Combettes and V.R. Wajs, *Signal recovery by proximal forward-backward splitting,* Multiscale Modeling and Simulation, vol. 4, no. 4, pp. 1168-1200 (2005).

[15] R. Cominetti, W.F. Mascarenhas, and P.J.S. Silva, *A Newton's method for the continuous quadratic knapsack problem,* Math. Program. Comput., 6(2), pp 151–169 (2014).

[16] L. Condat, *A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,* J. Optim. Theory Appl., 158, pp. 460–479 (2013).

[17] D. Gabay, *Applications of the method of multipliers to variational inequalities,* In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: *Applications to the Numerical Solution of Boundary Value Problems*, pp. 299·331. North-Holland, Amsterdam (1983).

[18] M. Hanke and W. Niethammer, *On the acceleration of Kaczmarz's method for inconsistent linear systems,* Linear Algebra Appl. 130, pp. 83–98 (1990).

[19] S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen.* Bull. Int. Acad. Polon. Sci. Lett. A 335–357 (1937).

[20] C. Molinari, J. Peypouquet, and F. Roldán, *Alternating Forward-Backward Splitting for Linearly Constrained Optimization Problems,* Optimization Letters., pp. 1-18.

[21] B. J. Pettis, *On integration in vector spaces,* Trans. Amer. Math. Soc., 44 (1938), pp. 277–304.

[22] H. Robbins and D. Siegmund, *A convergence theorem for non negative almost supermartingales and some applications, Optimizing Methods in Statistics*, (J. S. Rustagi, Ed.), pp. 233-257. Academic Press, New York (1971).

[23] T. Strohmer and R. Vershynin, *A randomized Kaczmarz algorithm with exponential convergence,* Journal of Fourier Analysis and Applications, 15(2):262–278 (2009).

[24] B.C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators,* Adv. Comput. Math. 38, pp. 667–681 (2013).