UNIVERSITE CLERMONT AUVERGNE

ECOLE DOCTORALE
SCIENCES POUR L'INGENIEUR DE CLERMONT-FERRAND

T h è s e

Présentée par

ORÉ ALBORNOZ Ernesto

pour obtenir le grade de

DOCTEUR D'UNIVERSITÉ

SPECIALITE : INFORMATIQUE

Sur les Méthodes de Décomposition Proximale Appliquées à la Planification de Production Electrique en Environnement Incertain

Soutenue publiquement le 18 décembre 2018 devant le jury :

M. Michel DE LARA	Président
M. Samir ADLY	Rapporteur et examinateur
M. Jean-Pierre DUSSAULT	Rapporteur et examinateur
M. Eladio OCANA ANAYA	Directeur de thèse
M. Philippe Mahey	Directeur de thèse

Dedicatoria

A toda mi familia,
en especial a mi madre.

AGRADECIMIENTOS

Agradesco a las instituciones peruanas IMCA y CONCYTEC, quienes por medio del Contrato 217-2014 – FONDECYT, me permitieron desarrollar y culminar el trabajo de tesis.

Agradesco a mis asesores, Profesor Eladio Ocaña y Profesor Philippe Mahey, por su paciencia, consejos y asesoramiento durante la elaboración de este trabajo.

Asimismo agradesco al Instituto frances ISIMA, por la acogida, el apoyo científico, económico e inmobiliario brindado.

RÉSUMÉ

Dans cette thèse, nous nous concentrons sur les méthodes d'éclatements d'opérateurs monotones (en anglais "splitting") qui sont appliquées à des problèmes d'optimisation convexe ou d'inclusions monotones présentant des structures décomposables. Une des contributions majeures de cette thèse est de considérer des reformulations de type point-selle (Lagrangiennes) de ces problèmes d'inclusion menant vers des algorithmes de type primal-dual.

Dans un contexte général pour résoudre un problème d'inclusion monotone et obtenir des algorithmes de splitting, nous développons une méthode de point proximal généralisée et construisons une application associée avec des propriétés de contraction similaire à celle de l'opérateur résolvante classique. Notre configuration générale inclut des algorithmes de splitting connus dans l'abondante littérature qui les applique à des modèles d'apprentissage et de régression parcimonieuse très en vogue ces dix dernières années dans les domaines de l'inférence statistique et du traitement de signal et d'images.

Nous présentons également des techniques de décomposition afin de résoudre la version multi-blocs de certains problèmes structurés en exploitant des reformulations appropriées du problème original et puis en appliquant une version particulière de l'algorithme de notre schéma général.

Finalement, nous appliquons la méthode de splitting généralisée à un problème de planification de la production d'énergie à grande échelle avec une décomposition spatiale.

ABSTRACT

We focus in this thesis, on splitting methods which be applied to special optimization or inclusion problems considering its related inclusion problems with an appropriated Lagrangian map.

In a general setting for solving a monotone inclusion problem and obtain splitting algorithms, we develop a generalized proximal point and construct a related map with similar contraction properties as the resolvent map. Our general setting includes popular splitting algorithms.

Also, we show decomposition techniques in order to solve the multi-block version of our model problems, finding adequate formulations of the original problem and then apply a particular algorithm version of our general scheme.

Finally, we apply the splitting method to a large-scale energy production planning problem.

Contents

Introduction **1**

1 Notations, preliminaries and basic results on convex optimization problems **11**

1.1 Notations and basic definitions 11

1.2 The duality scheme 14

1.2.1 The composite model 16

1.2.2 The separable case 17

1.2.3 The optimization problem with linear subspace constraints . . 18

1.3 The gradient and proximal point methods 19

1.3.1 Application to the dual projection problem 20

1.3.2 The resolvent map corresponding to the Saddle Point Problem 22

1.4 The $\alpha-$average maps 23

1.4.1 Douglas-Rachford map 24

1.4.2 Davis-Yin map 25

1.4.3 Convergence Study 25

2 A unified splitting algorithm for composite monotone inclusions **27**

2.1 Introduction 27

2.2 A generalized proximal point method 31

2.3 Generalized splitting algorithms 33

2.3.1 The separable structure on the main step 34

2.4 The co-coercive map associated with GPPM 40

2.4.1 Examples of co-coercive operators G_S^L 43

2.5 Rate of Convergence 47

2.5.1 Bounding the fixed-point residual 49

2.5.2 Bounding the saddle-point gap 50

2.5.3 Bounding the constraint violation 51

2.6 Application to multi-block optimization problems 52

3 Decomposition techniques **59**
3.1 The separable model with coupling variable (SMCV) 60
3.2 Separable model with coupling constraints (SMCC) 64
3.3 Proximal separation into two sub-blocks 67
3.3.1 The separable model with coupling variables 68
3.3.2 The separable model with coupling constraints 70
3.4 Multi-block optimization problem 71
3.4.1 Aplication to a stochastic problem 74

4 A new splitting algorithm for inclusion problems mixing a composite monotone plus a co-coercive operator **77**
4.1 Matrix A injective 80
4.1.1 The average map $G_{\widehat{S}}^{\widehat{L}}$: an appropriate regularization map . . . 81
4.1.2 Constructing the splitting algorithm 83
4.1.3 Switching the proximal step 84
4.1.4 Rate of Convergence 86
4.2 The general case on matrix A 90
4.2.1 The main algorithm for non injective operators 91
4.2.2 Switching the proximal step 94
4.3 A variant of primal–dual Condat's algorithms 95
4.3.1 Relationship with the Condat's method 97
4.3.2 Relationship with the Davis-Yin's method 100
4.3.3 Rate of convergence 101
4.4 General separable optimization problem 104
4.5 Application to multi-block optimization problems 107
4.6 Numerical Example 110

5 Application to stochastic problems **113**
5.1 The stochastic optimization model 113
5.2 Solution of a deterministic formulation 115
5.2.1 ADMM applied to the dynamic model 118
5.2.2 Chambolle-Pock applied to the dynamic model 119
5.2.3 PDA applied to the dynamic model 122
5.3 Uncertainty Environment 125

Conclusion **130**

Introduction

Decomposition techniques have been widely used in Mathematical Programming and Variational Analysis to cope mainly with the large-scale models issued from Decision Systems with many variables and constraints, but also to manage heterogeneous models including discrete choices, uncertainties or even mix of conflictual criteria. The recent explosion in size and complexity of datasets and the increased availability of computational resources has led to what is sometimes called the big data era. The large dimension of these data sets and the often parallel, distributed, or decentralized computational structures used for storing and handling the data, set new requirements on the optimization algorithms that solve these problems. Much effort has gone into developing algorithms that scale favorably with problem dimension and that can exploit structure in the problem as well as the computational environment.

Splitting methods for convex optimization or monotone variational analysis are commonly referred to address the construction of decomposition techniques based on regularization and duality. Indeed, many hard problems can be expressed under the form of a minimization of a sum of terms where each term is given by the composition of a convex function with a linear operator. The main advantage of splitting methods results thus from the fact that they can yield very efficient optimization schemes according to which a solution of the original problem is iteratively computed through solving a sequence of easier subproblems, each one involving only one of the terms constituting the objective function. These algorithms can also handle both differentiable and non smooth terms, the former by use of gradient operators (yielding explicit forward steps) and the latter by use of proximal operators (yielding implicit backward steps), thus giving rise to efficient first-order algorithms.

Since the pioneer works of Martinet [35], Glowinski-Marocco [24], Gabay [22] and Rockafellar [45], many algorithms have been studied for different models. A cornerstone was Lions and Mercier's paper in 1979 [31] about the Douglas-Rachford's

family of splitting methods applied to the following inclusion :

$$\text{Find } x \in X \text{ such that } 0 \in S(x) + T(x) \qquad (V_s)$$

where S and T are maximal monotone operators (typically subdifferential operators of convex functions) on a Hilbert space X.

Back to the motone inclusion problem, i.e. to find $x \in X$ such that $0 \in \mathcal{T}(x)$ where $\mathcal{T}$ is maximal monotone, the proximal point method (PPM) constructs the mapping $J_\lambda^{\mathcal{T}} = (I + \lambda\mathcal{T})^{-1}$ for $\lambda > 0$, the resolvent of $\mathcal{T}$ with known contractive properties, and transforms the above inclusion into an equivalent fixed-point equation, i.e. $x = J_\lambda^{\mathcal{T}}(x)$. (PPM) is thus defined by the following fixed-point iteration :

$$x^{t+1} = J_\lambda^{\mathcal{T}}(x^t)$$

which corresponds, when $\mathcal{T} = \partial f$, the subdifferential of a convex function f, to the following subproblem :

$$x^{t+1} = \text{argmin}_x \; f(x) + \frac{1}{2\lambda}\|x - x^t\|^2$$

That so-called implicit backward step leads to the celebrated Augmented Lagrangian algorithm when f is the dual function associated to the Lagrangian Relaxation of a constrained concave maximization problem with many potential applications (see for instance [28]).

In general the maximality assumption on $\mathcal{T}$ is restrictive, consider for instance the inclusion problem corresponding to the sum of two operators, i.e. inclusion (V). The sum $S + T$ is not necessarily maximal monotone and also its resolvent doesn't necessarily maintain its separability structure. Fortunately Lions and Mercier [31], solve that disadvantage considering an appropriate operator called after "Douglas-Rachford" operator (cf. [17], defined by

$$G_\lambda = I - J_\lambda^T + J_\lambda^S[2J_\lambda^T - I]$$

having splitting properties and whose fixed points are closely related with the solution points of problem (V). Moreover, the maximality assumptions on T and S, ensure that G_λ maintains the contractive properties of the single resolvent and the fullness of its domain.

When $T = \partial f$ and $S = \partial g$, unlike the resolvent map, G_λ is not in general the subdifferential of a function [19], so we need to continue working with monotone operators even in optimization problems in order to obtain splitting methods for optimization problems.

The Douglas–Rachford method leads us to consider in Chapter 2 and 4 the following methodology: given an inclusion problem with separable structure, we construct an appropriated average map with full domain having splitting properties and then apply the fixed point method to this new map.

Gabay [22] noticed that the Douglas-Rachford operator is behind the popular Alternate Direction Method of Multipliers (ADMM), considering the dual variational problem associated with the **composite model**

$$\text{Minimize} \quad f(x) + g(Ax)$$

where $f : I\!R^n \mapsto I\!R$ and $g : I\!R^m \mapsto I\!R$ are proper lower semi-continuous (lsc, for short) convex functions and A a given $(m \times n)$ matrix. The ADMM is an iterative method that consider two sub problems associated with f and g separately at each iteration.

Algorithm (**ADMM**)

$$\left\{ \begin{array}{rcl} x^{k+1} & = & \text{argmin}_x \ f(x) + \frac{\sigma}{2}\|Ax - z^k + \sigma^{-1}y^k\|^2 \\ z^{k+1} & = & \text{argmin}_z \ g(z) + \frac{\sigma}{2}\|Ax^k - z + \sigma^{-1}y^k\|^2 \\ y^{k+1} & = & y^k + \sigma(Ax^{k+1} - z^{k+1}) \end{array} \right.$$

where σ is a positive real parameter.

Recently Shefi and Teboulle [49] have presented a unified scheme algorithm for solving the last composite model based on the introduction of additional proximal terms like in Rockafellar's Proximal Method of Multipliers [44], this algorithm includes a version of a Proximal ADMM and other known algorithms like Chambolle and Pock [10].

In Chapter 2, we consider an extended model problem coming from an energy production planning problem

$$\min_{(x,y)} \quad [f(x) + g(y) : Ax + By = 0] \tag{P}$$

where f and g are again proper lsc convex functions, and A and B are matrices of order $m \times n$ and $m \times p$, respectively. Its saddle-point formulation in the variational setting is

$$\text{Find } (\bar{x}, \bar{z}, \bar{y}) \in I\!R^n \times I\!R^p \times I\!R^m \text{ such that } \ 0 \in L(\bar{x}, \bar{z}, \bar{y}) \tag{V_L}$$

where L is the maximal monotone map defined on $I\!R^n \times I\!R^p \times I\!R^m$ as

$$L(x, z, y) := \begin{pmatrix} \partial f(x) \\ \partial g(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}.$$

For solving (V_L), a generalized resolvent of L, defined as $J_P^L = (L+P)^{-1}P$, is introduced, where P is a symmetric positive semidefinite matrix with special block structure in order to split J_P^L into a separable structure leaving f and g separated. Then we consider the relaxed fixed-point method applying to J_P^L, where after changing the variables we obtain the following generalized algorithm

Generalized Splitting Scheme (GSS)

$$\tilde{z}^{k+1} \in \operatorname{argmin}\left\{g(z) + \frac{1}{2}\|Bz + M^{-1}u^k + Ax^k\|_M^2 + \frac{1}{2}\|z - z^k\|_{V_2}^2\right\} \quad (1)$$

$$v^{k+\frac{1}{2}} = \gamma Ax^k - (\gamma - 1)Bz^k + M^{-1}u^k \quad (2)$$

$$\tilde{x}^{k+1} \in \operatorname{argmin}\left\{f(x) + \frac{1}{2}\|Ax + v^{k+\frac{1}{2}} + 2\gamma B\tilde{z}^{k+1}\|_M^2 + \frac{1}{2}\|x - x^k\|_{V_1}^2\right\} \quad (3)$$

$$\tilde{u}^{k+1} = u^k + M(\gamma Ax^k + (1-\gamma)A\tilde{x}^{k+1} + B\tilde{z}^{k+1}) \quad (4)$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1-\rho)(x^k, z^k, u^k). \quad (5)$$

where the special cases $\gamma = 0$ and $\gamma = 1$, include the Shefi-Teboulle algorithm.

To guarantee the convergence of GSS and find an average map behind it, we developed a general setting in order to solve the inclusion problem: finding $x \in I\!R^n$ such that $0 \in \mathcal{T}(x)$, studying the generalized proximal point associated to the generalized resolvent $J_P^{\mathcal{T}} = (\mathcal{T}+P)^{-1}P$, where P is a symmetric positive semidefinite matrix, and then defining the map $G_S^{\mathcal{T}}$ as

$$G_S^{\mathcal{T}} = S(\mathcal{T} + S^tS)^{-1}S^t \quad (6)$$

where S is a matrix satisfying $S^tS = P$. By the monotonicity of $\mathcal{T}$, $G_S^{\mathcal{T}}$ is $\frac{1}{2}-$average.

In the special case of $\mathcal{T} = L$, we find conditions on the matrix S in order to G_S^L has full domain, which is used in the proof of the convergence of GSS. Also, considering $S = S_3$ (defined in Remark 2.4.3), we recover the Douglas-Rachford operator coinciding with $\lambda^{\frac{1}{2}}G_{S_3}^L\lambda^{-\frac{1}{2}}$, showing that its fundamental properties of splitability and $\frac{1}{2}-$average can be deduced from our generalized setting corresponding the Lagrangian map. Also, the general setting help us to find the ergodic and nonergodic rate of convergence of GSS.

Finally, in the last section of Chapter 2, for $f_i : I\!R^{n_i} \to \overline{I\!R}$ $(i = 1, \cdots, q)$ and $g : I\!R^m \to \overline{I\!R}$ proper lsc convex functions, A_i and B matrices of order $p \times n_i$ and

$p \times m$, respectively, the following S-Model problem is presented:

$$\begin{aligned} \inf_{(x_1,\cdots,x_q,z)} \quad & \sum_{i=1}^{q} f_i(x_i) + g(z) \\ s.t \quad & \sum_{i=1}^{q} A_i x_i + Bz = 0. \end{aligned}$$

This model is a multiblock generalization of problem (P). In [18] the authors propose the so called Separable Augmented Lagrangian Algorithm (SALA) in order to solve the S-Model problem by considering $g = 0$ and $B = 0$. At each iteration k, the algorithm considers a set of subproblems (one subproblem fo each $i = 1, \cdots, q$) defined by

$$\text{argmin}_x \; f_i(x) + \frac{1}{2}\|A_i x - z_i^k + M^{-1} y^k\|_M^2 \tag{7}$$

where M is a parameter matrix considered symmetric positive definite, solved in parallel processing.

As a generalization, they also consider different parameter matrices depending on each iteration.

Reformulating the S-Model problem in order to apply GSS, we get a new algorithm called **"Proximal Multi-block Algorithm (PMA)"** where like SALA it considers a set of problems at each iteration k:

$$\text{argmin}_x \left\{ f_i(x) + \frac{1}{2}\|A_i x - \tilde{z}_i^{k+1} + M_i^{-1}\tilde{y}_c^{k+1}\|_{M_i}^2 + \frac{1}{2}\|x - x_i^k\|_{Q_i}^2 \right\}, \; i = 1, \cdots, q,$$

where M_i is symmetric positive definite and Q_i symmetric positive semi-definite. Matrix Q_i allow us to deal with the not injective case on A_i and choosing it appropriately, each sub problem becomes on proximal step of f_i.

In Chapter 3, we show some decomposition techniques, which consist in finding an adequate formulation of the original problem in order to apply GSS with particular parameter matrices V_1, V_2 and M, and consequently find new splitting algorithms.

For every $i \in \{1, \cdots, q\}$, let $f_i : I\!R^n \to \overline{I\!R}$ be proper lsc convex. We consider the **separable model with coupling variable** (SMCV) defined as

$$\min_x \quad \sum_{i=1}^{q} f_i(x) \tag{S}$$

For this problem we recover algorithm PDA and the algorithm given in [34] by considering a particular parameter. We also show the relationship between these two

algorithms.

Adding linear constraints in the model, we obtain the **separable model with coupling constraint** (SMCC) where $A_i, i = 1, \ldots, q$ are $(p \times n_i)$ matrices

$$\begin{aligned} \min \quad & \sum_{i=1}^{q} f_i(x_i) \\ s.t \quad & \sum_{i=1}^{q} A_i x_i = 0 \end{aligned}$$

Notice that this problem can be seen as a S-Model problem and, since SMCC can be formulated as SMCV, we apply the results obtained for SMCV to SMCC, getting directly two algorithms which can also be recovered from PMA considering special formulation for SMCC.

The precedent algorithms found in this chapter, consider, for each iteration, the proximal step of all family $\{f_i\}_{i=1,\cdots,q}$ or separate the family into two sub-family, one consisting of $\{f_i\}_{i=1,\cdots,q-1}$ and the other consisting of f_q. Then, the proximal step of all $\{f_i\}_{i=1,\cdots,q-1}$ are found in parallel processing and then, after linear combination of all these values, the proximal step of f_q is found.

We show that after special reformulations, we get two splitting algorithms, one for SMCV and the other for SMCC. Each algorithm separates the problem into two sub-block problems, considering the proximal step to one sub-block and then (at a linear combination of the preceding values) the proximal step is found for the other sub-block, both in parallel processing.

In the last part of this chapter, we consider the following multi-block optimization problem

$$\min_{x=(x_1,\cdots,x_q)} \sum_{i=1}^{q} f_i(x_i) + g(\sum_{i=1}^{q} A_i x_i) + s(x) \qquad (P_{sc})$$

where for $i \in \{1, ..., q\}$, $f_i : I\!R^{n_i} \to \overline{I\!R}$, $g : I\!R^p \to \overline{I\!R}$ and $s : I\!R^n \mapsto \overline{I\!R}$ $(n = \sum_{i=1}^{q} n_i)$ are proper lsc convex functions, and A_i are matrices of order $p \times n_i$.

This special structure is deduced from the formulation over the Euclidean space of the following stochastic optimization model problem with finite scenarios Ξ

$$\min_{X \in \mathcal{L}} \left[E_\xi \sum_{t=1}^{T} g_t(X_t(\xi), \xi) : s.t\ X \in \mathcal{N} \text{ and } \sum_{t=1}^{T} B_t^\xi X_t(\xi) = 0,\ \forall \xi \in \Xi \right] \qquad (SP)$$

where $\mathcal{L}$ is the linear space of all mapping X from Ξ to $I\!R^n := I\!R^{n_1} \times \cdots \times I\!R^{n_T}$, and $\mathcal{N}$ the nonanticipativity subspace of $\mathcal{L}$, and for $\xi \in \Xi$ and $t = 1, \cdots, T$, $g_t(\cdot, \xi)$

is a proper lsc convex function and B_t^{ξ} an $m_\xi \times n_t$ matrix.

As a consequence of obtaining a splitting algorithm for (P_{sc}), we get a splitting algorithm for problem (SP).

Coming back to problem (V_s), if one these maps (say T) is co-coercive, then we can apply the Forward-Backward method or Backward-Forward method, which combine the Backward step of S with the Forward step of T. Notice that the Forward step only needs the value of single value map T unlike the Backward step needs the value of the resolvent map of S which in general is not easy.

D. Davis and W. Yin [16] generalized problem (V_s) considering the sum of three maps:

$$\text{Find} \quad x \quad \text{such that} \quad 0 \in S(x) + T(x) + C(x)$$

where S and T are again maximal monotone (with T not necessarily single value) and C a co-coercive operator with full domain. They combine separately the Backward steps on S and T, with the Forward step on C, defining the following map

$$\mathcal{G} := I - J_\lambda^T + J_\lambda^S(2J_\lambda^T - I - \lambda C(J_\lambda^T))$$

which clearly extends the Douglas-Rachford operator G_λ and the operators corresponding to the Forward-Backward and Backward-Forward methods.

In Chapter 4, we consider the more general composite monotone inclusion:

$$\text{Find } x \quad \text{such that} \quad 0 \in S(x) + A^t T(Ax) + C(x) \qquad (Var)$$

where $S : I\!R^n \rightrightarrows I\!R^n$ and $T : I\!R^m \rightrightarrows I\!R^m$ are maximal monotone maps, $C : I\!R^n \to I\!R^n$ a $\beta-$co-coercive with full domain and A an $m \times n$ matrix.

For the particular case $C = 0$, the map $\mathcal{G}$ coincides with the Douglas-Rachford operator which in turn is equal to $\lambda^{\frac{1}{2}} G_{S_3}^L \lambda^{-\frac{1}{2}}$ as mentioned just after (6).

For the general cases (C not necessarily equal to zero) and assuming A injective, we get from (6) considering $\mathcal{T} = \widehat{\mathcal{L}}$ an alternative Lagrangian map and $S = \widehat{D}$ a special matrix, a map $G_{\widehat{D}}^{\widehat{L}}$ which extends the Davis-Yin operator $\mathcal{G}$ (coinciding it with $\lambda^{\frac{1}{2}} G_{\widehat{D}}^{\widehat{L}} \lambda^{-\frac{1}{2}}$ when $A = I$) and maintaining similar properties as $\mathcal{G}$, for instance under mild assumptions, $G_{\widehat{D}}^{\widehat{L}}$ is an average map with full domain.

Then using the generalized resolvent $J_{\widehat{D}}^{\widehat{L}}$, we get a new splitting algorithm which converges to a saddle-point of Lagrangian map associated with primal problem (Var) because, as mentioned previously, $G_{\widehat{D}}^{\widehat{L}}$ is an average map with full domain.

For the general case, where A is not necessarily injective, problem (Var) is reformulated as

$$0 \in S(x) + \begin{pmatrix} A^t M^{\frac{1}{2}} & V^{\frac{1}{2}} \end{pmatrix} \begin{bmatrix} M^{-\frac{1}{2}} T M^{-\frac{1}{2}} \\ 0 \end{bmatrix} \begin{pmatrix} M^{\frac{1}{2}} A x \\ V^{\frac{1}{2}} x \end{pmatrix} + C(x) \qquad (Var_1)$$

where M and V are two symmetric matrices of order $m \times m$ and $n \times n$, respectively, with V positive semi-definite and M positive definite.

In this reformulation the involved matrix $\begin{pmatrix} M^{\frac{1}{2}} A \\ V^{\frac{1}{2}} \end{pmatrix}$ is injective if and only if $A^t M A + V$ is invertible. So, assuming that condition and applying the former splitting algorithm (for the injective case on A) to (Var_1), we get a splitting algorithm for problem (Var) in the general setting, which is termed **"Generalized splitting algorithm for three operators (GSA3O)"**:

(GSA3O)

$$\begin{aligned}
\tilde{z}^{k+1} &= (T+M)^{-1}(y^k + MAx^k) && (8)\\
\tilde{y}^{k+1} &= y^k + MAx^k - M\tilde{z}^{k+1} && (9)\\
r^{k+1} &= C((V + A^t MA)^{-1}(Vx^k + A^t M\tilde{z}^{k+1})) && (10)\\
\tilde{x}^{k+1} &= (S + V + A^t MA)^{-1}\left(Vx^k + A^t M\tilde{z}^{k+1} - A^t\tilde{y}^{k+1} - r^{k+1}\right) && (11)\\
(x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k). && (12)
\end{aligned}$$

The structure of problem (Var) is related to the variational formulation of the minimization of separable convex functions:

$$\text{Minimize} \quad f(x) + g(Ax) + h(x) \qquad (13)$$

where $f : I\!R^n \mapsto \bar{I\!R}$ and $g : I\!R^m \mapsto \bar{I\!R}$ are proper lower semi-continuous convex functions, $h : I\!R^n \mapsto I\!R$ is convex and $(\frac{1}{\beta})$-Lipschitz-differentiable, and A an $m \times n$ matrix.

Condat [13] presents two types of algorithms for solving (13) that we call $CA1$ and $CA2$, for simplicity, we have considered with fixed relaxation parameter $\rho > 0$ and without error term.

Algorithm (**CA1**)

$$\begin{cases}
\tilde{x}^{k+1} &= (\tau\partial f + I_{n\times n})^{-1}(x^k - \tau\nabla h(x^k) - \tau A^t y^k)\\
\tilde{y}^{k+1} &= (\sigma\partial g^* + I_{m\times m})^{-1}(y^k + \sigma A(2\tilde{x}^{k+1} - x^k))\\
(x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k)
\end{cases}$$

and the other one switch the roles of primal and dual variables:

Algorithm (**CA2**)

$$\begin{cases} \tilde{y}^{k+1} & = & (\sigma\partial g^* + I_{m\times m})^{-1}(y^k + \sigma A x^k) \\ \tilde{x}^{k+1} & = & (\tau\partial f + I_{n\times n})^{-1}(x^k - \tau\nabla h(x^k) - \tau A^t(2\tilde{y}^{k+1} - y^k)) \\ (x^{k+1}, y^{k+1}) & = & \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k) \end{cases}$$

The main difference between these algorithms is the order of action of the proximal steps. In the same manner, we present two versions of each algorithm proposed in this chapter.

From our general setting, choosing special parameter matrices and different Lagrangian maps, we get different variants of algorithms $CA1$ and $CA2$. From the variant of $CA1$ we recover in particular the recently algorithm proposed by Yang.

Using the same techniques given in Chapter 2, we show the ergodic and nonergodic rates of convergence of the algorithms found in this chapter.

The last model considered in this chapter is concerned with the more general $S-$Model problem defined as:

$$\begin{aligned} \inf_{x=(x_1,\cdots,x_q),\, z} \quad & \sum_{i=1}^{q} f_i(x_i) + h(x) + g(z) \\ s.t \quad & \sum_{i=1}^{q} A_i x_i - Bz = 0 \end{aligned}$$

where for $i = 1, \cdots, q$, $f_i : I\!R^{n_i} \to \overline{I\!R}$ and $g : I\!R^m \to \overline{I\!R}$ are proper lsc convex functions, $h : I\!R^n \mapsto I\!R$ is convex and $(\frac{1}{\beta})$-Lipschitz-differentiable ($n = \sum_{i=1}^{q} n_i$), and A_i and B are matrices of order $p \times n_i$ and $p \times m$, respectively.

Rewriting this last problem as a variational inclusion problem having similar structure as problem (Var), we apply GSA3O getting a new algorithm called **"Separable Primal -Dual Variant (SPDV)"** which is a generalization of PMA appearing in Chapter 2.

Finally, in Chapter 5, we apply the splitting algorithm developed in the previous chapters in order to solve a model of long-term energy pricing problem. Rewriting this model as problem (P), we apply algorithm GSS, getting three types of algorithms. Finally, we give some ideas on how to deal with the considered model in the stochastic case where many 'scenarios' with given probabilities need to be included, thus minimizing the total expected cost on the horizon.

Chapter 1

Notations, preliminaries and basic results on convex optimization problems

1.1 Notations and basic definitions

Throughout this thesis, we will use the following notations on convex optimization and variational inequality problems . Most of the theoretical material can be found in [43, 4].

We will denote by $\overline{\mathbb{R}}$ the set of extended real numbers $\mathbb{R}\cup\{\pm\infty\} = [-\infty, +\infty]$. For a given subset $C \subset \mathbb{R}^n$ we will denote by cl (C), int (C) and ri (C), the closure, the interior and the relative interior of C, respectively.

For a given set $C \subset \mathbb{R}^n$, the **orthogonal subspace** to C, denoted by $C^\perp$, is the linear subspace

$$C^\perp = \{x^* \in \mathbb{R}^n : \langle x^*, x\rangle = 0 \ \text{ for all } x \in C\}.$$

For a closed convex set $C \subset \mathbb{R}^n$, we denote by $\mathrm{Proj}_C(x)$ the projection of $x \in \mathbb{R}^n$ onto C which consists of all $\bar{y} \in C$ satisfying

$$\|x - \bar{y}\| \leq \|x - y\| \ \text{ for all } \ y \in C$$

where $\|\cdot\|$ denotes a norm of $\mathbb{R}^n$. Of course, if $C \neq \emptyset$, then $\bar{y}$ satisfying this inequality is unique if the considered norm is Euclidean (unless otherwise stated, we will use in all the Thesis this type of norm). For $F \subset \mathbb{R}^n$, the set $\mathrm{Proj}_C(F)$ denotes the collection of all $\mathrm{Proj}_C(x)$ for $x \in F$.

For a closed convex set $C \subset \mathbb{R}^n$, the normal cone of C at a given point $x \in C$ is the set denoted by $N_C(x)$ and defined as

$$\mathcal{N}_C(x) = \{x^* \in \mathbb{R}^n : \langle x^*, y - x\rangle \leq 0 \text{ for all } y \in C\},$$

assuming $N_C(x) = \emptyset$ if $x \notin C$.

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be **convex**, if its epigraph

$$\text{epi}\,(f) := \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \lambda\}$$

is convex; and **concave**, if $-f$ is convex. The function f is said to be lower semi-continuous (lsc, for short) at a given point $\bar{x}$ if for every $\lambda \in \mathbb{R}$ verifying $\lambda < f(\bar{x})$ there exists an open set V containing $\bar{x}$ such that $\lambda < f(x)$ for all $x \in V$. This function is lsc if it is lsc at every point of $\mathbb{R}^n$. Of course, f is lsc if $\text{epi}\,(f)$ is closed in $\mathbb{R}^n \times \mathbb{R}$.

Assuming f convex, it is said to be proper if $f(x) > -\infty$ for all $x \in \mathbb{R}^n$ and its domain defined as

$$\text{dom}\,(f) := \{x \in \mathbb{R}^n : f(x) < +\infty\} = \text{Proj}_{\mathbb{R}^n}(\text{epi}\,(f))$$

is nonempty. Of course, $\text{dom}\,(f)$ is convex if f is convex.

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be **strongly convex** (with modulus $\alpha > 0$) or $\alpha-$strongly convex if for all $x, y \in \mathbb{R}^n$ and all $t \in [0, 1]$, one has

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) - \frac{\alpha}{2}t(1-t)\|x - y\|^2.$$

The function f is said to be $\beta-$**Lipschitz differentiable function** (with $\beta > 0$) if it is differentiable whose gradient is $\beta-$Lipschitz (ie, Lipschitz continuous with constant β). In that case, f is of real value on the whole $\mathbb{R}^n$.

Associated to a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, its **Fenchel-conjugate** is the function f^* defined on $\mathbb{R}^n$ as

$$f^*(x^*) = \sup_{x \in \mathbb{R}^n} [\langle x^*, x\rangle - f(x)]$$

and its **biconjugate** is the function f^{**} which is the conjugate of f^*, ie

$$f^{**}(x) = \sup_{x^* \in \mathbb{R}^n} [\langle x, x^*\rangle - f^*(x^*)].$$

It is clear that f^* and f^{**} are convex and lsc, and f^{**} is the greatest lsc convex function bounded from above by f. Moreover, f^{**} and f coincide if f is proper lsc convex.

The **subdifferential** of f at a point $x \in \mathbb{R}^n$ is the set

$$\partial f(x) := \{x^* \in \mathbb{R}^n : f(x) + \langle x^*, y - x\rangle \leq f(y) \text{ for all } y \in \mathbb{R}^n\}$$

or equivalently

$$\partial f(x) = \{x^* \in \mathbb{R}^n : f(x) + f^*(x^*) \leq \langle x^*, x\rangle\}.$$

Clearly $\partial f(x) = \emptyset$ if $x \notin \text{dom}\,(f)$ or if f is not lsc at x. In general, $\partial f(x)$ is convex and closed, maybe empty. It is nonempty and bounded on $\text{int}\,(\text{dom}\,(f))$.

Another very important property of subdifferential is its **monotonicity**. For all $(x, x^*), (y, y^*)$ in the graph of ∂f one has

$$\langle x^* - y^*, x - y\rangle \geq 0.$$

In general, a multivalued map (or simply, map) $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be monotone if for all $(x, x^*), (y, y^*)$ in the graph of Γ [is the set consisting of all pair $(z, z^*) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $z^* \in \Gamma(z)$] one has

$$\langle x^* - y^*, x - y\rangle \geq 0.$$

It is clear that if Γ is monotone, then its inverse map Γ^{-1} defined by $\Gamma^{-1}(x^*) = \{x : x^* \in \Gamma(x)\}$, is monotone. So, the monotonicity property can be seen as a property on the graph instead on the map itself.

The map Γ is said to be **maximal monotone** if for any monotone map $\Sigma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfying $\Gamma(x) \subset \Sigma(x)$ for all $x \in \mathbb{R}^n$, one has $\Gamma = \Sigma$. It also follows that Γ is maximal monotone if and only if Γ^{-1} is maximal monotone.

A very important characterization of the maximality in the monotone sense is given by Minty's theorem [36]. It say that a monotone map $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone if and only if the inverse map $(I + \Gamma)^{-1}$, which is single-valued and with full domain. Here I denotes the identity map from $\mathbb{R}^n$ into itself.

Analogous to the strongly convexity, a map $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be **strongly monotone** (with modulus $\alpha > 0$) or $\alpha-$strongly monotone if $\Gamma - \rho I$ is monotone, i.e. for all $(x, x^*), (y, y^*)$ in the graph of Γ, it holds

$$\langle x^* - y^*, x - y\rangle \geq \alpha\|x - y\|^2.$$

One deduces that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is strongly convex if and only if its subdifferential ∂f is strongly monotone.

The inverse of a strongly monotone map (with modulus α) is clearly single value and $\alpha^{-1}-$Lipschitz.

A map Γ is said to be **co-coercive** with constant β (or shortly $\beta-$co-coercive) if its inverse Γ^{-1} is $\beta-$strongly monotone. That is, for all $(x, x^*), (y, y^*)$ in the graph of Γ, it holds

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

One deduce that if Γ is **co-coercive** with constant β then Γ is $\beta^{-1}-$Lipschitz. When $\beta \geq 1$, the map Γ is **nonexpansive**. In general, a map Γ is said to be **nonexpansive** if it is Lipschitz with constant ≤ 1, i.e. if there exists $\gamma \leq 1$ such that for all $(x, x^*), (y, y^*)$ in the graph of Γ, it holds that

$$\|x^* - y^*\| \leq \gamma\|x - y\|.$$

Another important property used in many parts of the thesis is the $\alpha-$average of a map. A map Γ is said to be $\alpha-$**average** if

$$\Gamma = (1-\alpha)I + \alpha R$$

where R is a nonexpansive map. A $2^{-1}-$average map is also called **firmly nonexpansive**. For example, the resolvent of a maximal monotone map Γ, $J^{\Gamma} := (I+\Gamma)^{-1}$ is firmly nonexpansive (and defined on the whole space).

We finish this section by introducing the following notations that we will use for instance in Chapters 2, 3 and 4. For arbitrary maps T_1 and T_2 and vectors x and y of appropriated dimensions, we denote

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} T_1(x) \\ T_2(y) \end{pmatrix}.$$

Analogously, for two given functions g_1 and g_2, we denote

$$(g_1, g_2)(z_1, z_2) = g_1(z_1) + g_2(z_2)$$

for all z_1 and z_2 of appropriated dimensions.

1.2 The duality scheme

An optimization problem in the mathematical context can be set as

$$\alpha := \inf [f(x) : x \in I\!R^n] \tag{P}$$

where $f : I\!R^n \to \overline{I\!R}$ is a given function. Problem (P) is commonly called **primal problem**.

In order to develop the duality scheme following Rockafellar's scheme [43], we introduce a duality space $I\!R^p$ and a *perturbation function* $\varphi : I\!R^n \times I\!R^p \to \overline{I\!R}$ satisfying $\varphi(x,0) = f(x)$ for all $x \in I\!R^n$. Then the associated *perturbed primal problems* is defined as

$$h(u) := \inf [\varphi(x,u) : x \in I\!R^n]. \tag{P_u}$$

If φ is convex on $I\!R^n \times I\!R^p$ then h is convex on $I\!R^p$; but if φ is lsc it does not necessarily imply that h is lsc.

It is clear that $h(0) = \alpha$, then the duality arises from the idea to find in other way $h(0)$, for this we use the Fenchel-conjugate function h^* of h,

$$h^*(u^*) := \sup_u [\langle u^*, u\rangle - h(u)] = \varphi^*(0, u^*).$$

The biconjugate of h is

$$h^{**}(u) = \sup_{u^*}[\langle u, u^*\rangle - \varphi^*(0, u^*)].$$

The biconjugate, under some conditions (see [43]), allows us to recover the initial function. In general,

$$\beta := h^{**}(0) \leq h(0) = \alpha.$$

Then the **dual problem** is defined as

$$\beta = h^{**}(0) = \sup\,[-\varphi^*(0, u^*) : u^* \in I\!R^p]. \qquad (D)$$

The primal and dual problem are also related through the Lagrangian function $l : I\!R^n \times I\!R^p \to \overline{I\!R}$ defined as:

$$l(x, u^*) := \inf\,[\varphi(x, u) - \langle u^*, u\rangle : u \in I\!R^p].$$

So, if φ is convex proper lsc, the primal and dual problem are respectively:

$$\inf_x \sup_{u^*} l(x, u^*) \quad \text{and} \quad \sup_{u^*} \inf_x l(x, u^*).$$

In order to obtain optimal solution of primal and dual problems without duality gap ($\alpha = \beta$), the **Saddle Point problem** arises which consist in finding $(\bar{x}, \bar{y}) \in I\!R^n \times I\!R^p$ such that

$$\inf_x l(x, \bar{y}) = l(\bar{x}, \bar{y}) = \sup_y l(\bar{x}, y).$$

Under some regularity condition on φ, the primal and dual problems can be respectively formulated as inclusion problems called **optimality condition**

$$\text{Find } x \in I\!R^n \text{ such that } 0 \in \pi_1^t \partial\varphi(\pi_1 x) \qquad (P_{opc})$$

and

$$\text{Find } y \in I\!R^p \text{ such that } 0 \in \pi_2^t \partial\varphi^*(\pi_2 y) \qquad (D_{opc})$$

where

$$\pi_1 = \begin{pmatrix} I_{n\times n} \\ 0_{p\times p} \end{pmatrix} \text{ and } \pi_2 = \begin{pmatrix} 0_{n\times n} \\ I_{p\times p} \end{pmatrix}.$$

The Saddle Point problem can also be formulated as

$$\text{Find } (x, y) \in I\!R^n \times I\!R^p \text{ such that } (0, 0) \in (\partial_x l) \times (\partial_y[-l])(x, y) \qquad (L_{opc})$$

which, in terms of $\partial\varphi$ and assuming φ proper lsc convex, it holds that

$$(z, y) \in \partial\varphi(x, u) \text{ if only if } z \in \partial_x l(x, y),\ u \in \partial_y[-l](x, y).$$

Since $\partial\varphi^* = (\partial\varphi)^{-1}$, then problems (P_V), (D_V) and (L_V) are equivalent to each other in the sense that the mapping intervening in each inclusion problem is the composite of $\partial\varphi$ or $\partial\varphi^{-1}$ with π_1 or π_2 (and its respective transpose matrix).

Following this construction, [42] (see also [39]) has extended this duality scheme to general variational or inclusion problems which can be set as

$$\text{Find } x \in I\!R^n \text{ such that } 0 \in S(x). \tag{V}$$

A perturbation map associated to problem (V) is a map F satisfying $\pi_1^t F \pi_1 = S$. Then the corresponding dual and lagrangian problems are respectively formulated as

$$\text{Find } y \in I\!R^p \text{ such that } 0 \in \pi_2^t F^{-1}(\pi_2 y) \tag{D_V}$$

and

$$\text{Find } (x,y) \in I\!R^n \times I\!R^p \text{ such that } (0,0) \in L(x,y) \tag{L_V}$$

where

$$(z,y) \in F(x,u) \text{ if and only if } (z,u) \in L(x,y).$$

Coming back to the duality scheme for an optimization problem, we reformulate the duality and its respective Lagrangian problem through their optimality conditions for some particular classes of optimization problems.

1.2.1 The composite model

A composite model is an optimization problem that can be set as

$$\text{Minimize } f(x) + g(Ax) \tag{P_c}$$

where $f : I\!R^n \to \overline{I\!R}$ and $g : I\!R^p \to \overline{I\!R}$ are proper lsc convex functions, and A a $p \times n$ matrix. We consider the following perturbation function $\varphi : I\!R^n \times I\!R^p \to \overline{I\!R}$ defined by

$$\varphi(z,u) = f(x) + g(Ax + u)$$

and its corresponding dual problem

$$\text{Minimize } f^*(-A^t y) + g^*(y) \tag{D_c}$$

and its Lagrangian function $l : I\!R^n \times I\!R^p \to \overline{I\!R}$ defined by

$$l(x,y) := f(x) - g^*(y) + \langle y, Ax \rangle.$$

Under some regularity conditions, the optimal conditions of (P_c) and (D_c) are respectively

$$\text{Find } x \in I\!R^n \text{ such that } 0 \in \partial f(x) + A^t \partial g(Ax) \tag{P_{cv}}$$

and

$$\text{Find } x \in I\!R^n \text{ such that } 0 \in -A\partial f^*(-A^t y) + \partial g^*(y). \qquad (D_{cv})$$

The corresponding Saddle point problem can also be formulated as

$$\text{Find } (x,y) \in I\!R^n \times I\!R^p \text{ such that } 0 \in \begin{pmatrix} \partial f(x) \\ \partial g^*(y) \end{pmatrix} + \begin{pmatrix} 0 & A^t \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (L_{cv})$$

1.2.2 The separable case

We now consider a more general optimization problem regarding to the previous one

$$\min_{(x,z)} \quad [f(x) + g(z) : Ax + Bz = 0] \qquad (P_{sc})$$

where $f : I\!R^n \to \overline{I\!R}$ and $g : I\!R^p \to \overline{I\!R}$ are again proper lsc convex functions, and A and B are matrices of order $m \times n$ and $m \times p$, respectively.

It is clear that problem (P_{sc}) includes the composite model (P_c) by considering $B = -I_{p\times p}$. Conversely, problem (P_{sc}) can be written as the following composite model:

$$\min_{(x,z)} \quad f(x) + g(z) + \delta_{\{0\}} \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix}.$$

So, the dual problem is

$$\min_{y} \quad f^*(-A^t y) + g^*(-B^t y) \qquad (D_{sc})$$

and the lagrangian function $l_{sc} : I\!R^n \times I\!R^p \times I\!R^m \to \overline{I\!R}$ is

$$l_{sc}(x,z,y) = f(x) + g(z) + \langle y, Ax + Bz\rangle.$$

The optimality conditions of (P_{sc}) and (D_{sc}) and the corresponding saddle point problem are respectively:

$$\text{Find } (x,z) \in I\!R^n \times I\!R^p \text{ st } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial f(x) \\ \partial g(z) \end{pmatrix} + \begin{pmatrix} A^t \\ B^t \end{pmatrix} \mathcal{N}_{\{0\}}(Ax + Bz) \quad (P_{scv})$$

and

$$\text{Find } x \in I\!R^n \text{ such that } 0 \in -A\partial f^*(-A^t y) - B\partial g^*(-B^t y) \qquad (D_{scv})$$

and

$$\text{Find } (\bar{x},\bar{z},\bar{y}) \in I\!R^n \times I\!R^p \times I\!R^m \text{ such that } 0 \in L(\bar{x},\bar{z},\bar{y}) \qquad (L_{scv})$$

where L is the maximal monotone map defined on $I\!R^n \times I\!R^p \times I\!R^m$ as

$$L(x,z,y) = (\partial_{x,z} l_{sc}) \times (\partial_y[-l_{sc}]) = \begin{pmatrix} \partial f(x) \\ \partial g(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}. \quad (1.1)$$

1.2.3 The optimization problem with linear subspace constraints

We note that problem (P_{sc}) (in particular problem (P_c)) can be formulated as

$$\min_{z \in V} \psi(z) \tag{P_{sbp}}$$

where $\psi : I\!R^r \to \overline{I\!R}$ is a proper lsc convex function and V a linear subspace of $I\!R^r$.

We present now two equivalent formulations of (P_{sbp}) like the primal and dual problems by expressing V as Image space and as well as kernel space of suitable matrices.

It is clear that (P_{sbp}) can be formulated as the composite model:

$$\min \quad \psi(z) + \delta_V(z) \tag{$\widehat{P}_{sbp}$}$$

whose associated dual problem is

$$\min \quad \psi^*(z^*) + \delta_{V^\perp}(z^*) \tag{$\widehat{D}_{sbp}$}$$

which can be formulated as the following optimization problem with linear subspace constraints

$$\min_{z^* \in V^\perp} \psi^*(z^*). \tag{D_{sbp}}$$

Let K and W be matrices of order $p \times r$ and $r \times n$, respectively, such that $V = \ker(K) = \text{range}\,(W)$. Then (P_{sbp}) and (D_{sbp}) can also be formulated respectively as the following optimization problems termed **primal projection problem**

$$\min_{x \in I\!R^n} \psi \circ W(x) \tag{P^V_{sbp}}$$

and **dual projection problem**

$$\min_{u^* \in I\!R^p} \psi^* \circ K^t(u^*). \tag{D^V_{sbp}}$$

We observe that problems (P_{sbp}) and (P^V_{sbp}) are defined on different linear spaces and, if W is injective, then problem (P^V_{sbp}) is defined on a linear space of dimension $\dim(V)$ which is less or equal than r, that is, the dimension of the linear space where problem (P_{sbp}) is defined. Similarly, problems (D_{sbp}) and (D^V_{sbp}) are also defined on different linear spaces.

As a special case, problems (P_c) and (D_c) can also be formulated respectively as the primal projection and dual projection problems regarding the following optimization problem

$$\min_{(x,y) \in \overline{V}} \psi(x, y) \tag{Op}$$

where $\psi(x,y) := f(x) + g(y)$ and

$$\overline{V} = \text{range} \begin{pmatrix} I_{n\times n} \\ A \end{pmatrix} = \ker \begin{pmatrix} -A & I_{p\times p} \end{pmatrix}.$$

In the separable case, problem (D_{sc}) can also be formulated as the dual projection problem related to (Op) by considering $\overline{V} = \ker \begin{pmatrix} -A & -B \end{pmatrix}$. Then, by setting R and D matrices of order $n \times q$ and $p \times q$, respectively, such that

$$\overline{V} = \text{range} \begin{pmatrix} R \\ D \end{pmatrix}$$

we obtain two related problems, the primal projection problem

$$\min_{z\in I\!R^q} \quad f(Rz) + g(Dz)$$

and the optimization problem with linear subspace constrains

$$\min_{(x^*,y^*)} \quad [f^*(x^*) + g^*(z^*) : R^t x^* + D^t z^* = 0].$$

Remark 1.2.1 *Considering $I\!R^r = I\!R^n \times I\!R^p$, the linear subspace V coincides with $I\!R^n \times \{0_p\}$ and the matrices W and K are exactly the projection matrices π_1 and π_2^t, respectively, defined in problems (P_{opc}) and (D_{opc}). So,*

$$(\psi \circ W)(x) = \psi(x,0) \quad \textit{and} \quad (\psi^* \circ K^t)(u^*) = \psi^*(0,u^*) \quad \textit{for all } x \in I\!R^n \textit{ and } u^* \in I\!R^p$$

and hence problems (P^V_{sbp}) and (P^V_{sbp}) are respectively the primal and dual optimization problems associated to perturbation function $\varphi = \psi$ described in the duality scheme.

1.3 The gradient and proximal point methods

The gradient and proximal point methods are apparently the most popular and basic methods to solve an optimization problem or (more generally) an inclusion problem.

Let $h : I\!R^r \mapsto I\!R$ be an $(\frac{1}{\beta})$-Lipschitz-differentiable convex function. In order to find a minimizer of h, the **gradient method** generates, from a given initial point $x^0 \in I\!R^r$, the iterative points defined by:

$$x^{k+1} = x^k - \alpha \nabla h(x^k).$$

The corresponding sequence converges to a minimizer of h if $\alpha \in]0, \frac{2}{\beta}[$.

Regarding the proximal point algorithm in order to find a minimizer of a proper lsc convex function $f : I\!R^r \mapsto \overline{I\!R}$, this algorithm constructs the so called **Moreau**

envelope function, an alternative Lipschitz-differentiable convex function having the same minimizers as function f. Then the proximal point algorithm can be recovered from the gradient algorithm applied to the Moreau envelope function.

Given $\lambda > 0$, the **Moreau envelope function** of f is the function f_λ defined as

$$f_\lambda(z) := \min_x \left[f(x) + \frac{1}{2\lambda}\|x - z\|^2 \right]. \qquad (*)$$

One deduce immediately that

- $\inf f = \inf f_\lambda$ and $\operatorname{argmin} f = \operatorname{argmin} f_\lambda$; and
- f_λ is differentiable on $\mathbb{R}^r$ and its gradient is

$$\nabla f_\lambda(z) = \frac{1}{\lambda}(I - (\lambda \partial f + I)^{-1})(z)$$

  which is $(\frac{1}{\lambda})$-Lipschitz. The set value $(\lambda\partial f + I)^{-1}(z)$ is singleton whose unique element is the minimizer of problem $(*)$.

So, given an initial point $z^0 \in \mathbb{R}^r$, the proximal point method generates a sequence defined by

$$z^{k+1} = z^k - \rho\lambda\nabla f_\lambda(z^k) = (1-\rho)z^k + \rho(\lambda\partial f + I)^{-1}(z^k).$$

This sequence converges to a minimizer of f if $\rho \in]0, 2[$.

1.3.1 Application to the dual projection problem

The dual projection problem (D^V_{sbp}),

$$\min_{u^* \in \mathbb{R}^p} \psi^* \circ K^t(u^*)$$

represents, as we saw, the general formulation of all dual problems described in Subsection 1.2.3. In order to apply the gradient method to this problem, we check under what condition the dual objective function $\psi^* \circ K^t$ is Lipschitz-differentiable on the whole $\mathbb{R}^p$. This is the object of the next proposition.

Proposition 1.3.1 *With the same notations as before, suppose that ψ is proper lsc α-strongly convex and K a nonzero matrix, then $\psi^* \circ K^t$ is differentiable with gradient $K\nabla\psi^* K^t$ which is $\frac{\|K\|^2}{\alpha}-$Lipschitz with full domain.*

Proof. From the assumptions, $\partial\psi^*$ is univalued with full domain. Then the subdiferential of $\psi^* \circ K^t$ is $K\nabla\phi^* K^t$ having full domain. On other hand, for arbitrary points $x, y \in \mathbb{R}^p$, we have that

$$\langle K\nabla\psi^* K^t x - K\nabla\psi^* K^t y, x - y\rangle = \langle \nabla\psi^* K^t x - \nabla\psi^* K^t y, K^t x - K^t y\rangle$$

and since $\partial\psi = (\nabla\psi^*)^{-1}$ is α-strongly monotone, we get

$$\langle K\nabla\psi^*K^tx - K\nabla\psi^*K^ty, x-y\rangle \geq \alpha\|\nabla\psi^*K^tx - \nabla\psi^*K^ty\|^2$$

which implies, if K is nonzero, that

$$\langle K\nabla\psi^*K^tx - K\nabla\psi^*K^ty, x-y\rangle \geq \frac{\alpha}{\|K\|^2}\|K\nabla\psi^*K^tx - K\nabla\psi^*K^ty\|^2.$$

The Lipschitz constant of $K\nabla\psi^*K^t$ is deduced applying the Cauchy-Schwarz's inequality. ■

Regarding the proximal point algorithm applied to before dual problem, for a given $r \times r$ symmetric positive definite matrix Q, we consider a little more general Moreau envelope function of f denoted by f_Q, defined as

$$f_Q(z) := \min_x \left[f(x) + \frac{1}{2}\|x-z\|_Q^2 \right]. \qquad (env)$$

Similarly to the classical Moreau envelope function f_λ, it holds that

- $\inf f = \inf f_Q$ and $\operatorname{argmin} f = \operatorname{argmin} f_Q$;
- f_Q is differentiable on $\mathbb{R}^r$ and its gradient is

  $$\nabla f_Q(z) = Q[I - (\partial f + Q)^{-1}Q](z).$$

  The set $[(\partial f + Q)^{-1}Q](z)$ is singleton whose element is the optimal solution of the minimization problem (env);
- $Q^{-1/2}\nabla f_Q Q^{-1/2}$ is $1-$Lipschitz on $\mathbb{R}^r$.

The next proposition shows another way to express the Moreau envelope function for the objective dual function.

Proposition 1.3.2 *Let $\psi : \mathbb{R}^r \to \overline{\mathbb{R}}$ be a proper lsc convex function, K a $p \times r$ matrix satisfying $\operatorname{Im}(K^t) \cap \operatorname{ri}\operatorname{dom}(\psi^*) \neq \emptyset$. For a $p \times p$ positive definite matrix M one has*

$$(\psi^* \circ K^t)_{M^{-1}}(u^*) = -\inf_x \left[\psi(x) + \frac{1}{2}\|Kx\|_M^2 - \langle u, Kx\rangle \right] \quad \text{for all } u \in \mathbb{R}^p. \qquad (**)$$

Furthermore, denoting $z_u := (\partial(\psi^* \circ K^t) + M^{-1})^{-1}M^{-1}u$ the minimizer of problem (env) with $f = (\psi^* \circ K^t)$ and $Q = M^{-1}$, and $x_u = (\partial\psi + K^tMK)^{-1}K^tu$ a minimizer of problem $(**)$, then

$$z_u = u - MKx_u.$$

Proof. From assumptions, $\psi^* \circ K^t$ is proper lsc convex and then

$$(\psi^* \circ K^t)_{M^{-1}} = [(\psi^* \circ K^t)_{M^{-1}}]^{**} = [(\psi^* \circ K^t)^* + \frac{1}{2}\|\cdot\|_M^2]^*.$$

Also, $(\psi^* \circ K^t)^*(v) = \inf_x [\psi(x) : Kx = v]$ and hence

$$(\psi^* \circ K^t)_{M^{-1}}(u) = -\inf_v \inf_x \quad [\psi(x) + \frac{1}{2}\|v\|_M^2 - \langle v, u\rangle : Kx = v]$$

which implies that

$$(\psi^* \circ K^t)_{M^{-1}}(u) = -\inf_x \left[\psi(x) + \frac{1}{2}\|Kx\|_M^2 - \langle u, Kx\rangle\right].$$

The relationship between the optimal solutions follows from this expression. ■

On the other hand, since both optimization problems

$$\min_x f(x) \quad \text{and} \quad \min_w f_Q(w)$$

have same optimal values and same minimizers, then under the regularization condition given in Proposition 1.3.2, both optimization problems

$$\min_{y \in I\!R^p} \psi^* \circ K^t(y) \quad \text{and} \quad \min_{v \in I\!R^p} \left(-\min_x \left[\psi(x) + \frac{1}{2}\|Kx\|_M^2 - \langle v, Kx\rangle\right]\right)$$

have also same optimal values and same minimizers.

The problem on the right is termed **Augmented Dual Problem** [47] and its objective function is nothing else than $\left(\psi + \frac{1}{2}\|K(\cdot)\|_M^2\right)^* \circ K^t$ and hence the augmented dual problem is the dual projection problem corresponding to problem (P_{sbp}) with objective function $\psi + \frac{1}{2}\|K(\cdot)\|_M^2$.

In particular, for a given perturbation function φ of f in the duality scheme, the objective function of the dual problem corresponding to $\varphi(x,u) + \frac{1}{2}\|u\|^2$ (another perturbation function of f), results to be Lipschitz-differentiable if φ and π_2 satisfy the conditions given in Proposition 1.3.2 considering $\psi = \varphi$ and $K = \pi_2$.

1.3.2 The resolvent map corresponding to the Saddle Point Problem

Rockafellar [44], considered the proximal point method to a saddle point problem corresponding to a convex optimization problem with inequality constraints, getting the so called "proximal multiplier algorithm".

Corresponding to the Saddle Point Problem (L_{scv}) defined in Section 1.2.2 and associated to the diagonal block symmetric positive definite matrix

$$P := \begin{bmatrix} W_1 & 0 & 0 \\ 0 & W_2 & 0 \\ 0 & 0 & M \end{bmatrix}$$

where W_1, W_2 and M are symmetric positive definite matrices of order $n \times n$, $p \times p$ and $m \times m$, respectively, we introduce the following resolvent map defined by

$$(L + P)^{-1}P.$$

Since L is maximal monotone, Minty's theorem guarantees that $(L+P)^{-1}P$ has full domain and its value at each point $(x, z, y) \in I\!R^n \times I\!R^p \times I\!R^m$ is a singleton whose value is

$$(L + P)^{-1}P(x, z, y) = \left\{ (\eta, \nu, M^{-1}(My + A\eta + B\nu) : \begin{array}{l} 0 \in H^1_{(x,z)}(\eta, \nu) \\ 0 \in H^2_{(x,z)}(\eta, \nu) \end{array} \right\}$$

where

$$H^1_{(x,z)}(\eta, \nu) := \partial f(\eta) + W_1(\eta - x) + A^t y + A^t M^{-1}(A\eta + B\nu)$$

and

$$H^2_{(x,z)}(\eta, \nu) := \partial g(\nu) + W_2(\nu - z) + B^t y + B^t M^{-1}(A\eta + B\nu).$$

It is noteworthy that the involved subproblems cannot directly be splitted because of the coupling on their variables η and ν is present and hence the solvability of such subproblems becomes very difficult in practice. In Chapter 2 we present another matrix P avoiding the aforementioned coupling.

1.4 The $\alpha-$average maps

The gradient and proximal point method have common structure in the sense that both methods can be formulated as a relaxed fixed point method [13] for a suitable mapping having the following property defined now.

Definition 1 *An operator T is $\alpha-$average if $\alpha \in]0, 1[$ and there exists a nonexpansive map N such that $T = (1 - \alpha)I + \alpha N$.*

For instance, the $\alpha-$average maps involved in the gradient and proximal point methods are respectively $\beta\nabla h$ and $I-(\lambda\partial f+I)^{-1}$. Baillon-Haddad's theorem shows that if h is convex and $(\frac{1}{\beta})$-Lipschitz-differentiable, then $\beta\nabla h$ is $\frac{1}{2}-$average having full domain. On the other hand, Rockafellar [43] (Proposition 12.11) shows that if f is proper lsc convex, then $(\lambda\partial f + I)^{-1}$ and hence $I - (\lambda\partial f + I)^{-1}$ are $\frac{1}{2}-$average having full domain.

Another important example of $\alpha-$average map is given through a maximal monotone map $T : I\!R^n \longrightarrow I\!R^n$ and a $n \times n$ positive definite matrix M. It is not difficult to show that the resolvent map

$$M^{\frac{1}{2}}(T + M)^{-1}M^{\frac{1}{2}}$$

which is related with the multidimensional scaling proximal point method [29] is $\frac{1}{2}-$average having (due Minty) full domain.

In general, an $\alpha-$average map $\mathcal{T}$ satisfies the following important inequality

$$\|\mathcal{T}x - \mathcal{T}y\|^2 \leq \|x-y\|^2 - \frac{1-\alpha}{\alpha}\|(I-\mathcal{T})x - (I-\mathcal{T})y\|^2$$

which immediately implies the convergence of the relaxed fixed point algorithm as mentioned in next proposition.

Proposition 1.4.1 *Let $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ be an $\alpha-$average map with full domain and $\rho \in]0, \alpha^{-1}[$. Assume that the fixed point set of $\mathcal{T}$ is nonempty. For a given initial point $x^0 \in \mathbb{R}^n$ consider the following iteration points*

$$x^{n+1} = \rho\mathcal{T}(x^n) + (1-\rho)x^n.$$

Then the corresponding sequence $\{x^n\}$ converges to a fixed point of $\mathcal{T}$.

This important convergence result makes it possible to deal with general monotone inclusion problems by transforming them into fixed point equations corresponding to $\alpha-$ average maps. Moreover, for practical treatments is also important that the corresponding $\alpha-$ average map possesses splitting property.

We give two examples of $\alpha-$average maps corresponding to inclusion problems for the sum of two and three monotones maps. The first one is due the Douglas & Rachford scheme discussed in [31] and the second one is due to Davis & Yin [16].

1.4.1 Douglas-Rachford map

Consider the following inclusion problem for the sum of two maps

$$0 \in S(x) + T(x)$$

where S and T are two maximal monotone maps from $\mathbb{R}^n$ into itself. It is well known that $S+T$ is not necessarily maximal monotone which is a condition to apply the proximal point method. Another disadvantage of this method is the absence of splitting structure of its corresponding resolvent map.

Lions and Mercier [31], reformulate the above inclusion into a fixed-point equation with respect to an appropriated operator called after "Douglas-Rachford" operator, defined by

$$G_\lambda = I - J_\lambda^T + J_\lambda^S[2J_\lambda^T - I], \tag{1.2}$$

where $J_\lambda^T = (\lambda T + I)^{-1}$ and $J_\lambda^S = (\lambda S + I)^{-1}$ are the resolvent maps of T and S, respectively. This map is $\frac{1}{2}-$average and, unlike the resolvent map of $S+T$, it has splitting property and having full domain.

This map is behind the popular ADMM algorithm as noticed Eckstein [21]. On the other hand, D. O'Connor and L. Vandenberghe [40] have recently shown that Chambolle-Pock algorithm [10] is also constructed using this map by considering S and T with special structures in the inclusion problem.

1.4.2 Davis-Yin map

We consider the sum of three maps

$$0 \in S(x) + T(x) + C(x) \tag{V_2}$$

where S and T are two maximal monotone maps and C a $\beta-$co-coercive function with full domain, all from $\mathbb{R}^n$ into itself.

It is possible to apply the Douglas-Rachford method considering the sum of S (or T) and C as a unique map, but in general this procedure does not split S and C (or T and C).

Davis and Yin [16] have recently considered the following map

$$\mathcal{G} := I - J_\lambda^T + J_\lambda^S[2J_\lambda^T - I - \lambda C(J_\lambda^T)] \tag{1.3}$$

having splitting property and defined everywhere of $\mathbb{R}^n$. It is also $\alpha-$average for $\alpha = \frac{2\beta}{4\beta-\lambda}$, if $\lambda \in]0, 2\beta[$.

1.4.3 Convergence Study

We recall that $x_n = O(y_n)$ means that there exists a positive C such that for all n sufficiently large

$$\|x_n\| \leq C\|y_n\|.$$

And $x_n = o(y_n)$ means that $\frac{\|x_n\|}{\|y_n\|}$ converges to 0.
Also we say that x_n converge linearly to x^* if there exists a positive $C < 1$ such that for all n sufficiently large

$$\|x_{n+1} - x^*\| \leq C\|x_n - x^*\|.$$

For example, H. Brezis et P.L. Lions [**?**] showed that given a monotone map $\mathcal{T}$ with at least one zero then the fixed point residual (FPR) $\|J_\lambda^{\mathcal{T}}(x_n) - (x_n)\|$ is $O(\frac{1}{\sqrt{k+1}})$. D. Davis and W. Yin [15] improve this result for any average map with has at least a fixed point getting that its FPR is $o(\frac{1}{\sqrt{k+1}})$ and the ergodic FPR is $O(\frac{1}{k})$.

Rockafellar [45] showed that if we consider $\mathcal{T}$ a strongly monotone map (or more generally $\mathcal{T}^{-1}$ Lipschitz continuous at 0) and $\mathcal{T}$ has at least one zero then the proximal point applied to $\mathcal{T}$ generates a linearly convergent sequence.

Lions and Mercier [31] showed that if T in 4.2 is strongly monotone and Lipschitz then the Douglas-Rachford method generates a linearly convergent sequence. Recently Giselsson [23] gave a best upper bound rate as Lions-Mercier and showed linear convergence under other regularity conditions over S or T, proving that in these cases the map G_λ is contractive. In the convex case D. Davis and W. Yin [15] showed the ergodic and nonergodic convergence rate of the feasibility and objective function error related to the relaxed Douglas-Rachford method.

D. Davis and W. Yin in [16] showed that under regularity assumptions the map $\mathcal{G}$ is a contractive map from which the linear convergence is deduced.

In this thesis, we will not focus on the special cases when linear convergence is attained, rather keeping the analysis on global or point wise convergence in the ergodic or non ergodic sense.

Chapter 2

A unified splitting algorithm for composite monotone inclusions

Operator splitting methods have been recently concerned with inclusions problems based on composite operators made of the sum of two monotone operators, one of them associated with a linear transformation. We analyze here a general and new splitting method which indeed splits both operator proximal steps, and avoiding costly numerical algebra on the linear operator. The family of algorithms induced by our generalized setting includes known methods like Chambolle-Pock primal-dual algorithm and Shefi-Teboulle Proximal Alternate Direction method of multipliers. The study of the ergodic and non ergodic convergence rates show similar rates with the classical Douglas-Rachford splitting scheme. We end with an application to a multi-block convex optimization model which leads to a generalized Separable Augmented Lagrangian algorithm[1].

2.1 Introduction

Composite models involving sums and compositions of linear and monotone operators are very common and still challenging problems like in constrained separable convex optimization or composite variational inequalities. We will consider here composite monotone inclusions of the form (X and Y are Hilbert spaces) :

$$0 \in S(x) + A^*T(Ax) \tag{2.1}$$

where $S : X \mapsto X$ and $T : Y \mapsto Y$ are maximal monotone operators and $A : X \mapsto Y$ is a linear transformation (associated with its adjoint operator A^*, which will be denoted by A^t when dealing with finite-dimensional spaces).

[1]This chapter corresponds to the paper [41] submitted to Journal of Convex Analysis

Most existing monotone operator splitting methods can deal with composite models, for example the Douglas-Rachford family (see [31]) and its special decomposition versions, the Alternate Direction Method of Multipliers (ADMM) (see [22, 21]) and the Partial Inverse or Proximal Decomposition Algorithm (PDA) (see [51, 34, 42]).

Lions and Mercier [31] analyzed the Douglas-Rachford's method (including the limiting case of Peaceman-Rachford splitting, PRS) for the case of the sum of two maximal monotone operators $(S+T)$, alternating between proximal steps applied to each operator separately. Gabay [22] analyzed the case $S + A^*TA$ where A is an injective linear transformation (and A^* its adjoint), yielding the celebrated Alternative Direction Method of Multipliers (ADMM). Spingarn [50] studied the case when the operator is the sum of the normal cone of a closed subspace M and a maximal monotone operator T. Later, Pennanen [42] showed how to reformulate that model as a monotone inclusion

The first study which explicitly considered an algorithm to solve the composite inclusion which avoids the use of projection (or proximal) steps on the range of A was proposed in [9] (an extension of Spingarn's Partial Inverse to composite models was proposed too in [1]). The corresponding algorithms solve the dual problem at the same time, which is defined by :

$$0 \in -AS^{-1}(-A^*y) + T^{-1}(y)$$

Many applications surge in the minimization of separable convex functions like :

$$\text{Minimize} \quad f(x) + g(Ax) \tag{2.2}$$

where $f : I\!R^n \to I\!R$ and $g : I\!R^m \to I\!R$ are proper lower semi-continuous convex functions and A is a given $(m \times n)$ matrix. The Dual problem in the sense of Rockafellar-Fenchel theory is :

$$\text{Minimize} \quad f^*(-A^ty) + g^*(y)$$

where f^* is the Fenchel–conjugate of f.

Recently Chambolle-Pock [10] studied model (2.2) and introduced new splitting schemes applied to a Lagrangian formulation of the primal minimization problem. They applied a primal-dual version of (ADMM) to the following saddle-point formulation :

$$\min_x \max_y \quad f(x) - g^*(y) + \langle Ax, y\rangle$$

Observe that we could as well define a Lagrangian operator associated with the composite inclusion (2.1) :

$$\bar{L}(x,y) = [S(x) + A^ty] \times [T^{-1}(y) - Ax] \tag{2.3}$$

Chambolle and Pock's algorithm relies on two Proximal steps on f and g with an additional extrapolation step (in a similar fashion of Varga's iterative principle [53]) as summarized below :

$$\left\{\begin{array}{lcl} x^{k+1} & = & (I+\tau\partial f)^{-1}(x^k-\tau A^t\bar{y}^k) \\ y^{k+1} & = & (I+\sigma\partial g^*)^{-1}(y^k+\sigma Ax^{k+1}) \\ \bar{y}^{k+1} & = & y^{k+1}+\theta(y^{k+1}-y^k) \end{array}\right.$$

where $(I+\tau\partial f)^{-1}$ is the resolvent operator of the subdifferential operator $S = \partial f$ which is known to be defined on the whole space and supposed to be easily computable in a so-called 'backward' proximal step as detailed below.

The difference and presumed advantage of that formulation is the symmetry (considering that x and y can be updated in reverse order) and a potentially decomposable algorithm which depends on three parameters. Their convergence result states that we should choose their values such that $\sigma\tau\|A\|^2 < 1$.

Observe now that (CPA) can be rewritten using Augmented Lagrangian-like functions by using the Moreau identity (see [37]) :

$$(I+\sigma\partial g^*)^{-1}(y)+\sigma(I+\sigma^{-1}\partial g)^{-1}(\sigma^{-1}y) = y$$

Resuming the transformed steps into the following iteration:

Algorithm (**CPA**)

$$\left\{\begin{array}{lcl} x^{k+1} & = & \text{argmin}_x\ f(x)+\frac{1}{2\tau}\|x-x^k+\tau A^t\bar{y}^k\|^2 \\ z^{k+1} & = & \text{argmin}_z\ g(z)+\frac{\sigma}{2}\|z-Ax^{k+1}-\sigma^{-1}y^k\|^2 \\ y^{k+1} & = & y^k+\sigma(A\bar{x}^{k+1}-z^{k+1}) \\ \bar{y}^{k+1} & = & y^{k+1}+\theta(y^{k+1}-y^k) \end{array}\right.$$

Chambolle and Pock confirmed the expected rate of convergence in $O(1/k)$ and even obtain the accelerated rate of $O(1/k^2)$ following the FISTA scheme of Beck and Teboulle [5] (thus reaching Nesterov's optimal rates in convex programming [38]).

In a recent survey, Shefi and Teboulle [49] have presented a unified scheme algorithm for solving model (2.2) based on the introduction of additional proximal terms like in Rockafellar's Proximal Method of Multipliers [44]. The resulting schemes include a version of a Proximal (ADMM) and other known algorithms like Chambolle-Pock's method (CPA). Indeed, a generic sequential algorithm proposed by Shefi and Teboulle is the following three steps scheme :

Algorithm (**STA**)

$$\left\{\begin{array}{lcl} x^{k+1} & = & \text{argmin}_x\ f(x)+\frac{\sigma}{2}\|Ax-z^k+\sigma^{-1}y^k\|^2+\frac{1}{2}\|x-x^k\|^2_{M_1} \\ z^{k+1} & = & \text{argmin}_z\ g(z)+\frac{\sigma}{2}\|Ax^{k+1}-z+\sigma^{-1}y^k\|^2+\frac{1}{2}\|z-z^k\|^2_{M_2} \\ y^{k+1} & = & y^k+\sigma(Ax^{k+1}-z^{k+1}) \end{array}\right.$$

where $\|.\|_M$ is the norm induced by a symmetric positive definite matrix M, i.e. $\|x\|_M^2 = x^t Mx$. Algorithm (STA) makes use of alternate minimization steps on the Augmented Lagrangian function associated with the coupling subspace $Ax - z = 0$. It is noted in [49] that (CPA) with the choice $\theta = 1$ corresponds exactly to (STA) with $M_1 = \tau^{-1} I - \sigma A^t A$ and $M_2 = 0$ (which implies again that $\sigma\tau\|A\|^2 < 1$).

Later, Condat [13] extended the model (2.2) and algorithm (CPA) to the case $f = F + h$ where $F : I\!R^n \to I\!R$ is convex and smooth. He relaxed the restriction on the parameters allowing $\sigma\tau\|A\|^2 = 1$ and also includes the Douglas-Rachford family in the case of $A = I$ (therefore we can say that Chambolle-Pock's method generalized Douglas-Rachford's splitting scheme). Condat showed too that Chambolle-Pock's method is the proximal point method applied to the Lagrangian operator associated with the primal and dual pair of inclusions.

In this chapter we will further extend the algorithms surveyed by Shefi and Teboulle, in order to solve the following convex optimization problem

$$\min_{(x,z)} \quad [f(x) + g(z) : Ax + Bz = 0]. \tag{P}$$

where f and g are again convex lsc functions and, A and B are two matrices of order $m \times n$ and $m \times p$, respectively. It is clear that this problem includes problem (2.2) by considering $B = -I_{p\times p}$.

The primal variational formulation of (P) is the following

$$\text{Find } (x,z) \in I\!R^n \times I\!R^p \text{ such that } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial f(x) \\ \partial g(z) \end{pmatrix} + \begin{pmatrix} A^t \\ B^t \end{pmatrix} \mathcal{N}_{\{0\}}(Ax + Bz) \tag{2.4}$$

where $\mathcal{N}_C(a)$ is the normal cone to set C at point a.

The dual variational formulation of (P) is the following

$$\text{Find } \; y \in I\!R^m \; \text{ such that } \; 0 \in \; - A(\partial f)^{-1}(-A^t y) - B(\partial g)^{-1}(-B^t y) \tag{2.5}$$

In Section 2.2, we propose a generalized proximal point method (GPPM) which was developed implicitly by Condat [13], where we consider specific assumptions to relax the condition of symmetric positive definiteness of the matrix associated with the resolvent, to authorize matrices which are only symmetric positive semidefinite, maintaining the properties of convergence of the proximal method.

In Section 2.3, we apply GPPM in order to find a zero of the Lagrangian map associated with problem (P), selecting an appropriate symmetric positive semi definite matrix in order to obtain a Generalized Splitting Scheme (GSS), which includes

various known algorithms, for instance both types of algorithms studied by Shefi and Teboulle [49] correspond indeed to particular choices of the parameters in GSS.

In Section 2.4, we define a $1-$co-coercive operator G_P^T related to GPPM, which set of fixed points is related to the zeroes of T. When T is the Lagrangian operator and the matrix P has a special structure as considered in Section 2.3, we show examples where we can get that operator explicitly, in particular we can recover the Douglas-Rachford operator.

In Section 2.5, we investigate the rate of converge of the GSS scheme, in the ergodic and non ergodic sense, analyzing the convergence of the sequences of the optimal values and the constraints violations associated with problem (P).

Finally, section 2.6 applies the GSS scheme to some general multi-block convex optimization problem with a composite structure. We show the relationship with a separable Augmented Lagrangian algorithm (SALA) introduced in [32].

2.2 A generalized proximal point method

The classical *Proximal Point method* is used to solve a monotone inclusion

$$\text{Find} \quad x \in I\!R^r \quad \text{such that} \quad 0 \in T(x) \tag{V}$$

where $T : I\!R^r \longrightarrow I\!R^r$ is a maximal monotone operator. We denote by $\text{sol}\,(V)$ the solution set of problem (V). It is closed, convex and may be empty. The iteration exploits the contractive properties of the resolvent operator $J_\tau^T = (I + \tau T)^{-1}$ to define a sequence given by $x^{k+1} = J_\tau^T(x^k)$ which converges weakly to a solution of (V) if it is nonempty.

Following former ideas developed by Condat [13] in the proof of the convergence of a specialized splitting method closely related (CPA), we define the generalized Proximal Point iteration by substituting the classical resolvent by

$$J_P^T := (T + P)^{-1} P \tag{2.6}$$

where P is an $r \times r$ symmetric positive semidefinite matrix.

Since T is monotone, then for any (x, x^*), $(\bar{x}, \bar{x}^*) \in \text{graph}\,(J_P^T)$, one has

$$\langle x^* - \bar{x}^*, Px - P\bar{x} \rangle \geq \langle Px^* - P\bar{x}^*, x^* - \bar{x}^* \rangle \geq 0. \tag{2.7}$$

We deduce immediately the following properties:

- $T + P$ and thereby its inverse $(T + P)^{-1}$ are monotone.

- $R := P + I_{r\times r} - Q$ is a symmetric positive definite matrix, whenever Q is the orthogonal projection onto the image of P, which implies in particular that Q satisfies $QP = PQ = P$ and $Q^2 = Q$.

- $J_P^T = J_P^T Q$, where Q is as above.

As R is symmetric positive definite, it induces an inner product on $\mathbb{R}^r$, $\langle u, v\rangle_R := \langle Ru, v\rangle$ for all $u, v \in \mathbb{R}^r$ with its corresponding norm $\|u\|_R := \sqrt{\langle u, u\rangle_R}$ for all $u \in \mathbb{R}^r$.

Hence, from (2.7), for all $x, \bar{x} \in \mathrm{dom}\,(QJ_P^T) = \mathrm{dom}\,(J_P^T)$,

$$\langle QJ_P^T(x) - QJ_P^T(\bar{x}), x - \bar{x}\rangle_R \geq \|QJ_P^T(x) - QJ_P^T(\bar{x})\|_R^2,$$

which implies that QJ_P^T is $1-$**co-coercive wrt** R on domain of J_P^T.

We deduce immediately the following relationship between the solution set of problem (V) and the fixed points of J_P^T and QJ_P^T.

Proposition 2.2.1 *With the same notations as before, we have*

- $x \in \mathrm{sol}\,(V)$ *if and only if* x *is a fixed point of* J_P^T.

- v *is a fixed point of* QJ_P^T *if and only if* $v = Qx$ *for some* $x \in \mathrm{sol}\,(V) \cap J_P^T(v)$.

Proof. The first property is directly by definition. The second one follows from the fact that $v \in QJ_P^T v$ if and only if there exists x such that $x \in J_P^T(v)$ satisfying $v = Qx$. It follows that $x \in J_P^T(v) = J_P^T(Qx) = J_P^T(x)$. Using the first equivalence we deduce that x belongs to $\mathrm{sol}\,(V)$. ■

Concerning the *regularity* of J_P^T, we have

- If P is positive definite, then $Q = I_{r\times r}$ and $R = P$. We deduce that $J_P^T = QJ_P^T$ and then J_P^T is $1-$co-coercive wrt P on the whole of its domain.

- If P is not positive definite, then J_P^T may not be single valued. But if it is single valued, then it is continuous on the whole of its domain.

We consider now a relaxed version of the generalized proximal iteration. In connection with the resolvent operator J_P^T and a real positive parameter ρ, we consider for an arbitrary point $x^0 \in \mathrm{dom}\, J_P^T$, the sequence $\{x^k\}$ defined by

$$x^{k+1} \in \rho J_P^T(x^k) + (1-\rho)x^k. \tag{2.8}$$

Notice that this sequence is well defined whenever

$$\mathrm{range}\,(\rho J_P^T + (1-\rho)I) \subseteq \mathrm{dom}\,(J_P^T).$$

Concerning the convergence of $\{x^k\}$, we distinguish the following situations:

- If P is positive definite, then J_P^T is $1-$co-coercive wrt P (hence single valued) with full domain which implies that $\{x^k\}$ converges, for $\rho \in (0,2)$, assuming $\mathrm{sol}\,(V)$ nonempty. In fact, given $x^* \in \mathrm{sol}\,(V)$, the convergence follows from the inequality

$$\|x^k - x^*\|_P^2 \geq \frac{2-\rho}{\rho}\|x^{k+1} - x^k\|_P^2 + \|x^{k+1} - x^*\|_P^2.$$

- In general, since QJ_P^T is $1-$co-coercive wrt R, then for $\rho \in (0,2)$ and assuming that QJ_P^T has closed domain and nonempty fixed point set (which is equivalently to $\mathrm{sol}\,(V)$ being nonempty), the sequence $\{Qx^k\}$ is convergent. The convergence of $\{x^k\}$ needs additional assumptions as we show in the following proposition.

Proposition 2.2.2 *Let $T : \mathbb{R}^r \longrightarrow \mathbb{R}^r$ be maximal monotone and P be an $r \times r$ positive semidefinite matrix. Assuming J_P^T single valued (which implies that it is continuous) with closed domain and $\mathrm{sol}\,(V)$ not empty. Then, for $\rho \in (0,2)$, the sequence $\{x^k\}$ converges to some point belonging to $\mathrm{sol}\,(V)$.*

Proof. Since QJ_P^T is $1-$co-coercive wrt R, it is single valued on its domain; and since $J_P^T = J_P^T Q$, then from (2.8) we obtain that

$$Qx^{k+1} = \rho QJ_P^T(Qx^k) + (1-\rho)Qx^k. \tag{2.9}$$

Using again the fact that QJ_P^T is $1-$co-coercive wrt R and, by assumptions with closed domain, $\rho \in (0,2)$ and $\mathrm{sol}\,(V)$ nonempty, then $\{Qx^k\}$ converges to some point a, which is a fixed point of QJ_P^T. From Proposition 2.2.1 and the single valuedness assumption, $J_P^T(a) \in \mathrm{sol}\,(V)$.

On the other hand, using the triangular inequality in (2.8) we have

$$\|x^{k+1} - J_P^T(a)\| \leq \rho\|J_P^T(Qx^k) - J_P^T(a)\| + |1-\rho|\|x^k - J_P^T(a)\|.$$

Since J_P^T is continuous, the sequence $\|J_P^T(Qx^k) - J_P^T(a)\|$ converges to 0. We deduce that $\{x^k\}$ converges to $J_P^T(a)$. ∎

Some examples of specially tailored co-coercive operators will be discussed in Section 2.4.

2.3 Generalized splitting algorithms

With the convex minimization problem (P) defined in Section 2.1, we associate its Lagrangian function defined as

$$l(x,z,y) = f(x) + g(z) + \langle y, Ax + Bz\rangle \tag{2.10}$$

and then its saddle-point problem in the variational setting

$$\text{Find } (\bar{x},\bar{z},\bar{y}) \in I\!R^n \times I\!R^p \times I\!R^m \text{ such that } \ 0 \in L(\bar{x},\bar{z},\bar{y}) \qquad (V_L)$$

where L is the maximal monotone map defined on $I\!R^n \times I\!R^p \times I\!R^m$ as

$$L(x,z,y) := (\partial_{x,z} l) \times (\partial_y[-l]) = \begin{pmatrix} \partial f(x) \\ \partial g(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}. \quad (2.11)$$

The map L, as the sum of maximal monotone operators and a skew-symmetric linear operator, satisfies similar inequalities as the subdifferential of a convex-concave bifunction. These inequalities will be used in order to obtain the rate of convergence studied in Section 2.5.

Proposition 2.3.1 *For any* (d,d^*), $(\bar{d},\bar{d}^*) \in \mathrm{graph}\,(L)$, *considering* $d = (x,z,y)$ *and* $\bar{d} = (\bar{x},\bar{z},\bar{y})$, *it holds*

$$\langle d - \bar{d}, d^* \rangle \geq l(x,z,\bar{y}) - l(\bar{x},\bar{z},y) \geq \langle d - \bar{d}, \bar{d}^* \rangle.$$

These inequalities are still verified if we consider $(d,d^*) \in \mathrm{graph}\,(L)$ *and* $\bar{d} \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g) \times I\!R^m$, *for the first inequality; and* $(\bar{d},\bar{d}^*) \in \mathrm{graph}\,(L)$ *and* $d \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g) \times I\!R^m$, *for the second inequality.*

It is well known that, under some regularity conditions, problem (V_L) admits a saddle-point if and only if problem (P) admits an optimal solution. One instance of such regularity condition is :

$$\text{There exist } x \in \mathrm{ri}\,(\mathrm{dom}\, f) \text{ and } z \in \mathrm{ri}\,(\mathrm{dom}\, g) \text{ such that } Ax + Bz = 0. \qquad (H)$$

We now apply to problem (V_L) the relaxed proximal method described in the previous section for a specially tailored matrix P in order to provide a separable structure to the algorithm.

2.3.1 The separable structure on the main step

In this part we describe the main iteration step of the relaxed proximal method given in (2.8) providing a decomposable structure.

We will choose an appropriate symmetric matrix P in order to split $(L+P)^{-1}$ or equivalently $J_P^L = (L+P)^{-1}P$, into a separable structure leaving f and g separated.

To that end, given $(\tilde{x},\tilde{z},\tilde{y}) \in I\!R^n \times I\!R^p \times I\!R^m$, we analyze the solution of the following inclusion system: Find (x,z,y) such that

$$\begin{pmatrix} \partial f(x) \\ \partial g(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix} + \begin{pmatrix} P_{11} & P_{21}^t & P_{31}^t \\ P_{21} & P_{22} & P_{32}^t \\ P_{31} & P_{32} & P_{33} \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix} \ni \begin{pmatrix} \tilde{x} \\ \tilde{z} \\ \tilde{y} \end{pmatrix}.$$

We introduce now two parameters $\alpha, \beta \in \mathbb{R}$, and a positive definite matrix M to simplify the third row-block of P into $P_3 = [(1+\alpha)A \quad (1+\beta)B \quad M^{-1}]$. So, the last inclusion can be expressed as

$$y = M\tilde{y} - \alpha MAx - \beta MBz \tag{2.12}$$

and hence, replacing it in the second block-system, this results in

$$\partial g(z) + (2+\beta)B^t(M\tilde{y} - \alpha MAx - \beta MBz) + P_{21}x + P_{22}z \ni \tilde{z}.$$

So, in order to express this last system eliminating primal variable x, we need to consider $P_{21} = \alpha(2+\beta)B^tMA$, obtaining

$$z \in (\partial g + P_{22} - \beta(2+\beta)B^tMB)^{-1}(\tilde{z} - (2+\beta)B^tM\tilde{y}). \tag{2.13}$$

Using again (2.12), now in the first block system, we get

$$\partial f(x) + (2+\alpha)A^t(M\tilde{y} - \alpha MAx - \beta MBz) + P_{11}x + \alpha(2+\beta)B^tMAz \ni \tilde{x}$$

which is equivalent to

$$x \in (\partial f + P_{11} - \alpha(2+\alpha)A^tMA)^{-1}(\tilde{x} - (2+\alpha)A^tM\tilde{y} - 2(\alpha-\beta)A^tMBz). \tag{2.14}$$

Summarizing the previous sequence in order to get a separable structure, we must first solve system (2.13), then system (2.14) and finally system (2.12). The corresponding matrix P, of order $(r \times r)$ with $r = n + p + m$, is then of the form

$$P := \begin{pmatrix} C_1 & \alpha(2+\beta)A^tMB & (1+\alpha)A^t \\ \alpha(2+\beta)B^tMA & C_2 & (1+\beta)B^t \\ (1+\alpha)A & (1+\beta)B & M^{-1} \end{pmatrix} \tag{2.15}$$

where $C_1(n \times n), C_2(p \times p)$ are arbitrary symmetric matrices,

From the maximality of ∂f and ∂g, the inclusions in (2.13) and (2.14) are indeed equalities if the matrices defined as

$$W_1 := C_1 - \alpha(2+\alpha)A^tMA \quad \text{and} \quad W_2 := C_2 - \beta(2+\beta)B^tMB,$$

are positive definite. In that case $(L+P)^{-1}$ is single-valued with full domain and therefore J_P^L is continuous with full domain.

It is clear that P is symmetric. It is positive semidefinite (resp. positive definite) if and only if the matrix

$$U := \begin{pmatrix} C_1 - (1+\alpha)^2A^tMA & (\alpha-\beta-1)A^tMB \\ (\alpha-\beta-1)B^tMA & C_2 - (1+\beta)^2B^tMB \end{pmatrix} \tag{2.16}$$

is positive semidefinite (resp. positive definite).

We now list some conditions in order to get a positive semidefinite matrix U :

A1 If $C_1-[(1+\alpha)^2+(\alpha-\beta-1)^2]A^tMA$ and $C_2-[(1+\beta)^2+1]B^tMB$ are positive semidefinite then U is positive semidefinite.

A2 If $C_1-[(1+\alpha)^2+1]A^tMA$ and $C_2-[(1+\beta)^2+(\alpha-\beta-1)^2]B^tMB$ are positive semidefinite then U is positive semidefinite.

A3 If $\beta \leq \alpha-1$, and $C_1-[(1+\alpha)^2+(\alpha-\beta-1)]A^tMA$ and $C_2-[(1+\beta)^2+(\alpha-\beta-1)]B^tMB$ are positive semidefinite then U is positive semidefinite.

A4 If $\beta=\alpha-1$. Then $C_1-(1+\alpha)^2A^tMA$ and $C_2-\alpha^2B^tMB$ are positive semidefinite if only if U is positive semidefinite.

In order to calculate the sequence in (2.8), we first calculate $(\tilde{x}^{k+1},\tilde{z}^{k+1},\tilde{y}^{k+1})=J_P^L(x^k,z^k,y^k)$, which is equal to

$$(\tilde{x}^{k+1},\tilde{z}^{k+1},\tilde{y}^{k+1})=(L+P)^{-1}\begin{pmatrix} C_1x^k+\alpha(2+\beta)A^tMBz^k+(1+\alpha)A^ty^k \\ \alpha(2+\beta)B^tMAx^k+C_2z^k+(1+\beta)B^ty^k \\ (1+\alpha)Ax^k+(1+\beta)Bz^k+M^{-1}y^k \end{pmatrix}$$

Then from (2.13), we have that

$$\tilde{z}^{k+1}=\bar{J}^g_{W_2}\left(\tilde{z}-(\beta+2)B^tMAx^k\right) \tag{2.17}$$

where $\tilde{z}=C_2z^k-(2+\beta)(1+\beta)B^tMBz^k-B^ty^k$ and $\bar{J}^g_{W_2}=(\partial g+W_2)^{-1}$ is the generalized resolvent operator associated with the convex function g.

From (2.14), we have that

$$\tilde{x}^{k+1}=\bar{J}^f_{W_1}\left(\tilde{x}-2(\alpha-\beta)A^tMB\tilde{z}^{k+1}\right) \tag{2.18}$$

where $\tilde{x}=C_1x^k-(2+\alpha)(1+\alpha)A^tMAx^k+(\alpha-2\beta-2)A^tMBz^k-A^ty^k$ and $\bar{J}^f_{W_1}=(\partial f+W_1)^{-1}$ is the generalized resolvent operator associated with the convex function f; and from (2.12), we have that

$$\tilde{y}^{k+1}=y^k+(1+\alpha)MAx^k+(1+\beta)MBz^k-\alpha MA\tilde{x}^{k+1}-\beta MB\tilde{z}^{k+1}. \tag{2.19}$$

The sequence in (2.8) is completed with an extrapolation step for a given $\rho\in(0,2)$:

$$(x^{k+1},z^{k+1},y^{k+1})=\rho(\tilde{x}^{k+1},\tilde{z}^{k+1},\tilde{y}^{k+1})+(1-\rho)(x^k,z^k,y^k). \tag{2.20}$$

We obtain the following proposition directly from Proposition 2.2.2.

Proposition 2.3.2 *Let $\rho\in(0,2)$. Assume that $C_1\in\mathbb{R}^{n\times n}$, $C_2\in\mathbb{R}^{p\times p}$ and $M\in\mathbb{R}^{m\times m}$ are symmetric, with M positive definite; and α, $\beta\in\mathbb{R}$, such that W_1 and W_2 are positive definite and satisfying one of conditions (A1)-(A4). If $\mathrm{sol}\,(V_L)$ is nonempty, then for an arbitrary $(x^0,z^0,y^0)\in\mathbb{R}^{n+p+m}$, the sequence (x^k,z^k,y^k) defined by the sequential update formulas (2.17$\rightarrow$ 2.18$\rightarrow$ 2.19$\rightarrow$ 2.20) converges to some element of $\mathrm{sol}\,(V_L)$.*

We will now further reformulate the iteration to show the alternating steps on separable Augmented Lagrangian functions. We introduce the parameter $\gamma = \alpha - \beta$ and the matrices defined as

$$V_1 := W_1 - A^t M A \quad \text{and} \quad V_2 := W_2 - B^t M B. \tag{2.21}$$

The conditions $(A1) - (A4)$ become:

A1' If $V_1 - (\gamma - 1)^2 A^t M A$ and $V_2 - B^t M B$ are positive semidefinite then U is positive semidefinite.

A2' If $V_1 - A^t M A$ and $V_2 - (\gamma - 1)^2 B^t M B$ are positive semidefinite then U is positive semidefinite.

A3' If $\gamma \geq 1$. Then $V_1 - (\gamma - 1) A^t M A$ and $V_2 - (\gamma - 1) B^t M B$ are positive semidefinite then U is positive semidefinite.

A4' If $\gamma = 1$. Then V_1 and V_2 are positive semidefinite if only if U is positive semidefinite.

We introduce a new primal-dual auxiliary variable $u^k := y^k + (\alpha - \gamma + 1) M A x^k + (1 + \beta) M B z^k$, to obtain the following updates :

$$z^{k+\frac{1}{2}} = V_2 z^k - B^t u^k \tag{2.22}$$

$$\tilde{z}^{k+1} = J^g_{W_2}[z^{k+\frac{1}{2}} - B^t M A x^k] \tag{2.23}$$

$$x^{k+\frac{1}{2}} = V_1 x^k - \gamma A^t M A x^k + (\gamma - 1) A^t M B z^k - A^t u^k \tag{2.24}$$

$$\tilde{x}^{k+1} = J^f_{W_1}[x^{k+\frac{1}{2}} - 2\gamma A^t M B \tilde{z}^{k+1}] \tag{2.25}$$

$$\tilde{u}^{k+1} = u^k + \gamma M A x^k + (1 - \gamma) M A \tilde{x}^{k+1} + M B \tilde{z}^{k+1} \tag{2.26}$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1 - \rho)(x^k, z^k, u^k) \tag{2.27}$$

which is equivalent to the following sequential minimization subproblems :

Generalized Splitting Scheme (GSS)

$$\tilde{z}^{k+1} \in \operatorname{argmin} \left\{ g(z) + \frac{1}{2} \| B z + M^{-1} u^k + A x^k \|^2_M + \frac{1}{2} \| z - z^k \|^2_{V_2} \right\} \tag{2.28}$$

$$v^{k+\frac{1}{2}} = \gamma A x^k - (\gamma - 1) B z^k + M^{-1} u^k \tag{2.29}$$

$$\tilde{x}^{k+1} \in \operatorname{argmin} \left\{ f(x) + \frac{1}{2} \| A x + v^{k+\frac{1}{2}} + 2\gamma B \tilde{z}^{k+1} \|^2_M + \frac{1}{2} \| x - x^k \|^2_{V_1} \right\} \tag{2.30}$$

$$\tilde{u}^{k+1} = u^k + M(\gamma A x^k + (1 - \gamma) A \tilde{x}^{k+1} + B \tilde{z}^{k+1}) \tag{2.31}$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1 - \rho)(x^k, z^k, u^k). \tag{2.32}$$

From Proposition 2.3.2, we obtain the proposition of convergence of (GSS)

Proposition 2.3.3 *Let $\rho \in (0,2)$. Assume that $V_1 \in I\!R^{n\times n}$, $V_2 \in I\!R^{p\times p}$ and $M \in I\!R^{m\times m}$ are symmetric, with M positive definite such that $V_1 + A^tMA$ and $V_2 + B^tMB$ are positive definite. Let $\gamma \in I\!R$ such that one of conditions $(A1')-(A4')$ is satisfied. If $\mathrm{sol}\,(V_L)$ is nonempty, then for an arbitrary $(x^0, z^0, u^0) \in I\!R^{n+p+m}$, the sequence (x^k, z^k, u^k) in (2.28)-(2.32) converges to some element of $\mathrm{sol}\,(V_L)$.*

We analyze now the special cases when $\gamma = 0$ and $\gamma = 1$, which correspond to the two types of algorithms proposed by Shefi and Teboulle [49].

Case $\gamma = 0$

From $(A1')$, if both matrices $V_1 - A^tMA$ and $V_2 - B^tMB$ are positive semi-definite then P is a positive semi-definite matrix.

Switching the order (2.28) for (2.30), we get the following algorithm where the primal updates are performed in parallel:

$$\tilde{x}^{k+1} \in \mathrm{argmin}\left\{ f(x) + \frac{1}{2}\|Ax + Bz^k + M^{-1}u^k\|_M^2 + \frac{1}{2}\|x - x^k\|_{V_1}^2 \right\} \quad (2.33)$$

$$\tilde{z}^{k+1} \in \mathrm{argmin}\left\{ g(z) + \frac{1}{2}\|Ax^k + Bz + M^{-1}u^k\|_M^2 + \frac{1}{2}\|z - z^k\|_{V_2}^2 \right\} \quad (2.34)$$

$$\tilde{u}^{k+1} = u^k + M(A\tilde{x}^{k+1} + B\tilde{z}^{k+1}) \quad (2.35)$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1-\rho)(x^k, z^k, u^k) \quad (2.36)$$

If $B = -I_{p\times p}$, $M = cI_{p\times p}$ and $\rho = 1$, we obtain the algorithm STA type I proposed by Shefi and Teboulle [49].

Summarizing, from Proposition 2.3.3, we obtain the following proposition of convergence of the sequence defined by (2.33)-(2.36).

Proposition 2.3.4 *Let $\rho \in (0,2)$. Assume that $V_1 \in I\!R^{n\times n}$, $V_2 \in I\!R^{p\times p}$ and $M \in I\!R^{m\times m}$ are symmetric, with M positive definite, such that $V_1 + A^tMA$ and $V_2 + B^tMB$ are positive definite and $V_1 - A^tMA$ and $V_2 - B^tMB$ are positive semi-definite. If $\mathrm{sol}\,(V_L)$ is nonempty, then for an arbitrary $(x^0, z^0, u^0) \in I\!R^{n+p+m}$, the sequence (x^k, z^k, u^k) in (2.33)-(2.36) converges to some element of $\mathrm{sol}\,(V_L)$.*

Case $\gamma = 1$

From $(A4')$, it holds that V_1 and V_2 are positive semi-definite if only if P is a positive semi-definite matrix. In this case GSS becomes :

$$\tilde{z}^{k+1} \in \operatorname{argmin}\left\{g(z) + \frac{1}{2}\|Ax^k + Bz + M^{-1}u^k\|_M^2 + \frac{1}{2}\|z - z^k\|_{V_2}^2\right\} \quad (2.37)$$

$$\tilde{u}^{k+1} = u^k + M(Ax^k + B\tilde{z}^{k+1}) \quad (2.38)$$

$$\tilde{x}^{k+1} \in \operatorname{argmin}\left\{f(x) + \frac{1}{2}\|Ax + B\tilde{z}^{k+1} + M^{-1}\tilde{u}^{k+1}\|_M^2 + \frac{1}{2}\|x - x^k\|_{V_1}^2\right\} \quad (2.39)$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1-\rho)(x^k, z^k, u^k) \quad (2.40)$$

If $B = -I_{p\times p}$, $M = \tau I_{p\times p}$, $V_2 = 0$ and $V_1 = \sigma^{-1}I_{n\times n} - \tau A^tTA$ such that $1 \geq \sigma\tau\|A\|^2$, then we obtain the over relaxed algorithm proposed by Chambolle-Pock [10].

Considering $\rho = 1$ and defining, $\bar{x}^k := x^k$, $\bar{z}^k := z^{k+1}$ and $\bar{u}^k := u^{k+1}$, then substituting in (2.37)-(2.39) and switching the order, we get the following algorithm

$$\bar{x}^{k+1} \in \operatorname{argmin}\left\{f(x) + \frac{1}{2}\|Ax + B\bar{z}^k + M^{-1}\bar{u}^k\|_M^2 + \frac{1}{2}\|x - \bar{x}^k\|_{V_1}^2\right\} \quad (2.41)$$

$$\bar{z}^{k+1} \in \operatorname{argmin}\left\{g(z) + \frac{1}{2}\|A\bar{x}^{k+1} + Bz + M^{-1}\bar{u}^k\|_M^2 + \frac{1}{2}\|z - \bar{z}^k\|_{V_2}^2\right\} \quad (2.42)$$

$$\bar{u}^{k+1} = \bar{u}^k + M(A\bar{x}^{k+1} + B\bar{z}^{k+1}) \quad (2.43)$$

If $B = -I_{p\times p}$ and $M = cI_{p\times p}$, we obtain the algorithm STA type II proposed by Shefi and Teboulle [49], which is called the Proximal Alternating Direction Method (PADM).

Further transformations applied to (2.37)-(2.40) lead us to consider two interesting algorithms. The first of them is obtained by considering $V_2 = 0$, and considering the auxiliary variables $\widehat{x}^{k+1}, \widehat{z}^k$, $\widehat{u}^k$, $\widehat{s}^k$ to update the relaxed sequences $\widehat{x}^{k+1} := \frac{1}{\rho}x^{k+1} + (1-\frac{1}{\rho})x^k = \tilde{x}^{k+1}$, $\widehat{z}^k := \frac{1}{\rho}z^{k+1} + (1-\frac{1}{\rho})z^k = \tilde{z}^{k+1}$, $\widehat{u}^k := \tilde{u}^{k+1}$ and $\widehat{s}^k := x^k$, getting

$$\widehat{x}^{k+1} \in \operatorname{argmin}\left\{f(x)+\frac{1}{2}\|Ax+B\widehat{z}^k+M^{-1}\widehat{u}^k\|_M^2+\frac{1}{2}\|x-\widehat{s}^k\|_{V_1}^2\right\} \quad (2.44)$$

$$\widehat{z}^{k+1} \in \operatorname{argmin}\left\{g(z)+\frac{1}{2}\|\rho A\widehat{x}^{k+1}+Bz+M^{-1}\widehat{u}^k+(\rho-1)B\widehat{z}^k\|_M^2\right\} \quad (2.45)$$

$$\widehat{u}^{k+1} = \widehat{u}^k+\rho MA\widehat{x}^{k+1}+(\rho-1)MB\widehat{z}^k+MB\widehat{z}^{k+1} \quad (2.46)$$

$$\widehat{s}^{k+1} = \rho\widehat{x}^{k+1}+(1-\rho)\widehat{s}^k \quad (2.47)$$

The second interesting algorithm is obtained by considering the auxiliary variables $\check{x}^k$, $\check{z}^k$, $\check{u}^k$, $\check{s}^k$ to update the relaxed sequences $\check{x}^k := \frac{1}{\rho}x^{k+1}+(1-\frac{1}{\rho})x^k = \tilde{x}^{k+1}$, $\check{z}^k := \frac{1}{\rho}z^{k+1}+(1-\frac{1}{\rho})z^k = \tilde{z}^{k+1}$, $\check{u}^k := \tilde{u}^{k+1}$ and $\check{s}^k := x^k$, getting

$$\check{z}^{k+1} \in \operatorname{argmin}\left\{g(z)+\frac{1}{2}\|\rho A\check{x}^k+Bz+M^{-1}\check{u}^k+(\rho-1)B\check{z}^k\|_M^2\right\} \quad (2.48)$$

$$\check{u}^{k+1} = \check{u}^k+\rho MA\check{x}^k+(\rho-1)MB\check{z}^k+MB\check{z}^{k+1} \quad (2.49)$$

$$\check{s}^{k+1} = \rho\check{x}^k+(1-\rho)\check{s}^k \quad (2.50)$$

$$\check{x}^{k+1} \in \operatorname{argmin}\left\{f(x)+\frac{1}{2}\|Ax+B\check{z}^{k+1}+M^{-1}\check{u}^{k+1}\|_M^2+\frac{1}{2}\|x-\check{s}^{k+1}\|_{V_1}^2\right\} \quad (2.51)$$

So, by considering in these two last algorithms $B=-I_{p\times p}$, $M=cI_{p\times p}$ and $V_1=0$, the sequences $\widehat{s}^k$ and $\check{s}^k$ becomes unnecessary. Moreover, (2.44)-(2.47) become the generalized ADMM proposed by Eckstein [21], and (2.48)-(2.51) become the algorithm 2 considered in [15].

From Proposition 2.3.3, we obtain the convergence of the sequence (2.37)-(2.40)

Proposition 2.3.5 *Let $\rho\in(0,2)$. Assume that $V_1\in\mathbb{R}^{n\times n}$, $V_2\in\mathbb{R}^{p\times p}$ and $M\in\mathbb{R}^{m\times m}$ are symmetric, with V_1 and V_2 positive semi-definite and M positive definite such that V_1+A^tMA and V_2+B^tMB are positive definite. If $\operatorname{sol}(V_L)$ is nonempty, then for an arbitrary $(x^0,z^0,u^0)\in\mathbb{R}^{n+p+m}$, the sequence (x^k,z^k,u^k) defined in (2.37)-(2.40) converges to some element of $\operatorname{sol}(V_L)$.*

2.4 The co-coercive map associated with GPPM

Lions and Mercier [31] have transformed an inclusion problem of the sum of two maximal monotone operators $(S+T)$ into a fixed-point equation with respect to an appropriated operator, the Douglas-Rachford operator, which is $1-$co-coercive map and, in order to compute its value at each point of its domain, only local calculations

of proximal terms of S and T separately are needed. Eckstein [21] later showed the relationship between the splitting algorithm (ADMM) and the fixed-point method applied to a Douglas-Rachford operator, after a suitable linear transformation.

In our general setting, we show in this section that the sequence generated by the generalized proximal point method (GPPM) corresponding to map J_P^T for arbitrary maximal monote operator T and arbitrary symmetric positive semidefinite matrix P is nothing else but the sequence generated by the fixed point method corresponding to map G_S^T defined in (2.53), after a linear transformation S (satisfying $P = S^tS$). It leads thus in some sense to a generalization of a Douglas-Rachford operator, keeping the property of $1-$co-coercivity.

As pointed out in Subsection 2.3.1, the sequence generated by GPPM for $T = L$ defined in (2.11) and P defined in (2.15) corresponds to the sequence generated by the generalized splitting scheme (GSS) defined in (2.17)–(2.20).

In Section 2.2, we have shown that the sequence generated by GPPM is nothing else but, under the linear transformation Q, the sequence generated by the fixed point method corresponding to the $1-$co-coercive wrt R map QJ_P^T (see (2.9)). But for arbitrary symmetric positive semidefinite matrix P, matrices Q and R are difficult to calculate; when P is symmetric positive definite, then $Q = I$ and $R = P$. Alternately by considering S such that $P = S^tS$, we define G_S^T an operator easier to implement than QJ_P^T and having similar properties, for example, it is $1-$co-coercive property wrt the usual norm. In particular, using G_S^T instead QJ_P^T, we give an alternative proof of Proposition 2.2.2.

Finally, by considering $S = S_3$ defined in Remark 2.4.3, one gets $G_{S_3}^L = S_3^t(L + S_3^tS_3)^{-1}S_3$ which corresponds, under a reparametrization, to the classical Douglas-Rachford operator defined by $M^{-\frac{1}{2}}S_3^t(L+S_3^tS_3)^{-1}S_3M^{\frac{1}{2}}$. In other words, the Douglas-Rachford operator and its fundamental properties of of co-coercivity and splittability will be shown to be a special case of our generalized setting based on the Lagrangian monotone inclusion.

Associated with the $r \times r$ symmetric positive semidefinite matrix P introduced in the former section, let consider a $q \times r$ matrix S satisfying

$$P = S^tS \tag{2.52}$$

and then the map $G_S^T : I\!R^q \longrightarrow I\!R^q$ defined as

$$G_S^T := S(T + S^tS)^{-1}S^t. \tag{2.53}$$

It follows that

$$SJ_P^T = G_S^TS \tag{2.54}$$

and hence, from (2.7), we get for all $w, w' \in I\!R^r$:

$$\langle G_S^T(Sw) - G_S^T(Sw'), Sw - Sw' \rangle \geq \|G_S^T(Sw) - G_S^T(Sw')\|^2.$$

Since for any $s, s' \in I\!R^q$ there exist $w, w' \in I\!R^r$ such that $S^tSw = S^ts$ and $S^tSw' = S^ts'$, we get

$$\langle G_S^T(s) - G_S^T(s'), s - s' \rangle \geq \|G_S^T(s) - G_S^T(s')\|^2$$

which means that G_S^T is $1-$co-coercive with respect to the usual norm.

The following proposition shows in particular that G_S^T is the Moreau-Yosida regularization of $ST^{-1}S^t$. This will be used in the examples considered in this Section and in Section 2.6 (Proposition 2.6.1).

Proposition 2.4.1 *Let $T : I\!R^r \longrightarrow I\!R^r$ be an arbitrary map, S and M two matrices of order $q \times r$ and $q \times q$, respectively, with M invertible. For $z \in I\!R^q$ the value $(ST^{-1}S^t + M)^{-1}Mz$ is nonempty if and only if $(T + S^tM^{-1}S)^{-1}S^tz$ is nonempty. Furthermore, it holds that*

$$(ST^{-1}S^t + M)^{-1}Mz = z - M^{-1}S(T + S^tM^{-1}S)^{-1}S^tz.$$

Proof. The proof follows from the two properties:

- $x \in (ST^{-1}S^t + M)^{-1}Mz$ if and only if there exists $y \in I\!R^m$ such that $S^tx \in T(y)$ and $z - M^{-1}Sy = x$.
- $y^* \in (T + S^tM^{-1}S)^{-1}S^tz$ if and only if exists $x^* \in I\!R^r$ such that $S^tx^* \in T(y^*)$ and $z - M^{-1}Sy^* = x^*$. ■

Similar to Proposition 2.2.1, we get the relationship between the solution set of problem (V) and the fixed points of G_S^T.

Proposition 2.4.2 *With the same notations as before, we have*

- *If $z \in \mathrm{sol}\,(V)$, then Sz is a fixed point of G_S^T.*
- *If w is a fixed point of G_S^T, then $w = Sq$ for some $q \in \mathrm{sol}\,(V) \cap (T+P)^{-1}S^tw$.*

We deduce that the set of fixed point of G_S^T is exactly

$$S(\mathrm{sol}\,(V)) = \{Sw : w \in \mathrm{sol}\,(V)\}.$$

Applying S to the sequence $\{w^k\}$ defined in (2.8) and considering the permutation property (2.54), we get:

$$Sw^{k+1} = \rho G_S^T(Sw^k) + (1 - \rho)Sw^k. \tag{2.55}$$

This equation gives us another alternative proof of convergence of the sequence $\{w^k\}$ under the same conditions of Proposition 2.2.2. In fact, since G_S^T is $1-$co-coercive and from (2.55), we have that, given $w^* \in \operatorname{sol}(V)$

$$\|Sw^k - Sw^*\|^2 - \frac{2-\rho}{\rho}\|Sw^{k+1} - Sw^k\|^2 - \|Sw^{k+1} - Sw^*\|^2 \geq 0. \tag{2.56}$$

Since $\operatorname{rank} S^tS = \operatorname{rank} S^t$, the domain of G_S^T is equal to the domain of J_P^T which is closed, using this fact and from (2.56) we deduce that Sw^k converges to some point b, which is a fixed point of G_S^T. On the other hand, using the triangular inequality and considering $\tilde{w} := (T+P)^{-1}S^tb$, we get

$$\|w^{k+1} - \tilde{w}\| \leq \rho\|(T+P)^{-1}S^t(Sw^k) - \tilde{w}\| + |1-\rho|\|w^k - \tilde{w}\|.$$

From the continuity of J_P^T, we deduce the continuity of $(T+P)^{-1}S^t = J_P^TS^+$, where S^+ denotes the Moore–Penrose pseudo-inverse matrix of S. Therefore we deduce that $\{w^k\}$ converges to $\tilde{w}$.

We now give some explicit expressions of G_S^L for the Lagrangian operator L and matrix S such that $P = S^tS$, considered in Section 2.3.

2.4.1 Examples of co-coercive operators G_S^L

Example 2.4.1 *Let $\gamma = 1$ ($\beta = \alpha - 1$), We consider in (2.15)*

$$C_1 = V_1 + (1+\alpha)^2A^tMA \quad \text{and} \quad C_2 = V_2 + \alpha^2B^tMB,$$

where V_1 and V_2 are as (2.21) assumed positive semidefinite matrices. In (2.37)-(2.40) matrices V_1 and V_2 are associated with the additional proximal term that will be used in ADMM, which, as we have shown in Subsection 2.3.1 (Case $\gamma = 1$), is related to Shefi-Teboulle algorithm type II [49]. We get :

$$P = \begin{pmatrix} V_1 + (1+\alpha)^2A^tMA & (1+\alpha)\alpha A^tMB & (1+\alpha)A^t \\ (1+\alpha)\alpha B^tMA & V_2 + \alpha^2B^tMB & \alpha B^t \\ (1+\alpha)A & \alpha B & M^{-1} \end{pmatrix}.$$

The matrix

$$S_1 = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ 0 & V_2^{\frac{1}{2}} & 0 \\ (1+\alpha)M^{\frac{1}{2}}A & \alpha M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}$$

satisfies (2.52) and the corresponding map $G^L_{S_1}$, that applies $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ into itself, is defined as

$$G^L_{S_1}(\widehat{x},\widehat{z},\widehat{y}) = \begin{pmatrix} V_1^{\frac{1}{2}}x \\ V_2^{\frac{1}{2}}z \\ M^{\frac{1}{2}}Ax + M^{\frac{1}{2}}Bz + \widehat{y} \end{pmatrix}$$

where

$$x = (\partial f + V_1 + A^tMA)^{-1}(V_1^{\frac{1}{2}}\widehat{x} - A^tM^{\frac{1}{2}}(\widehat{y} + 2M^{\frac{1}{2}}Bz))$$
$$z = (\partial g + V_2 + B^tMB)^{-1}(V_2^{\frac{1}{2}}\widehat{z} - B^tM^{\frac{1}{2}}\widehat{y}).$$

Note that $G^L_{S_1}$ has full domain if $V_1 + A^tMA$ and $V_2 + B^tMB$ are assumed positive definite matrices.

Remark 2.4.1 *The map $G^L_{S_1}$ is the Douglas–Rachford operator [31], applied to the two maps*

$$-\begin{pmatrix} V_1^{\frac{1}{2}} & 0 \\ 0 & I_{p\times p} \\ M^{\frac{1}{2}}A & 0 \end{pmatrix}\begin{bmatrix} \partial f \\ 0 \end{bmatrix}^{-1}\left(-\begin{pmatrix} V_1^{\frac{1}{2}} & 0 & A^tM^{\frac{1}{2}} \\ 0 & I_{p\times p} & 0 \end{pmatrix}\right)$$

and

$$-\begin{pmatrix} -I_{n\times n} & 0 \\ 0 & -V_2^{\frac{1}{2}} \\ 0 & M^{\frac{1}{2}}B \end{pmatrix}\begin{bmatrix} 0 \\ \partial g \end{bmatrix}^{-1}\left(-\begin{pmatrix} -I_{n\times n} & 0 & 0 \\ 0 & -V_2^{\frac{1}{2}} & B^tM^{\frac{1}{2}} \end{pmatrix}\right)$$

The corresponding sum of these two maps is exactly the **dual variational map** *(2.5) associated with the following optimization problem*

$$\min_{(x_1,x_2,z_1,z_2)\in\mathcal{F}}(f,0)(x_1,x_2) + (0,g)(z_1,z_2)$$

where $\mathcal{F}$ is the set of all (x_1,x_2,z_1,z_2) satisfying

$$\begin{pmatrix} V_1^{\frac{1}{2}} & 0 \\ 0 & I_{p\times p} \\ M^{\frac{1}{2}}A & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -I_{n\times n} & 0 \\ 0 & -V_2^{\frac{1}{2}} \\ 0 & M^{\frac{1}{2}}B \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 0.$$

Remark 2.4.2 *In the case that $V_2 = 0$, which corresponds to Chambolle-Pock algorithm as we showed in Subsection 2.3.1, we can restrict the map $G^L_{S_1}$, and obtain the map D_1 that applies $\mathbb{R}^n \times \mathbb{R}^m$ into itself, where $D_1(x,u)$ is*

$$\begin{pmatrix} V_1^{\frac{1}{2}}(\partial f + V_1 + A^tMA)^{-1}[V_1^{\frac{1}{2}}x - A^tM^{\frac{1}{2}}(u+2z)] \\ M^{\frac{1}{2}}A(\partial f + V_1 + A^tMA)^{-1}[V_1^{\frac{1}{2}}x - A^tM^{\frac{1}{2}}(u+2z)] + z + u \end{pmatrix}$$

where

$$z = M^{\frac{1}{2}} B(\partial g + B^t M B)^{-1} B^t M^{\frac{1}{2}}(-u).$$

Note if B is injective, then D_1 has full domain.

The map D_1 can be obtained in the form (2.53), considering that when $V_2 = 0$, the matrix

$$S_2 = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ (1+\alpha)M^{\frac{1}{2}}A & \alpha M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}$$

satisfies (2.52), and we obtain that $D_1 = G^L_{S_2}$.

The map D_1 can also be obtained as the Douglas–Rachford operator, applied to the two maps

$$-\begin{pmatrix} V_1^{\frac{1}{2}} \\ M^{\frac{1}{2}}A \end{pmatrix} (\partial f)^{-1} \left(-\begin{pmatrix} V_1^{\frac{1}{2}} & A^t M^{\frac{1}{2}} \end{pmatrix} \right)$$

and

$$-\begin{pmatrix} -I_{n\times n} & 0 \\ 0 & M^{\frac{1}{2}}B \end{pmatrix} \begin{bmatrix} 0 \\ \partial g \end{bmatrix}^{-1} \left(-\begin{pmatrix} -I_{n\times n} & 0 \\ 0 & B^t M^{\frac{1}{2}} \end{pmatrix} \right)$$

The corresponding sum of these two maps is exactly the dual variational map associated with the following optimization problem

$$\min_{(x,z_1,z_2)\in\mathcal{F}} f(x) + (0,g)(z_1,z_2)$$

where $\mathcal{F}$ is the set of all triples (x, z_1, z_2) satisfying

$$\begin{pmatrix} V_1^{\frac{1}{2}} \\ M^{\frac{1}{2}}A \end{pmatrix} x + \begin{pmatrix} -I_{n\times n} & 0 \\ 0 & M^{\frac{1}{2}}B \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 0.$$

Remark 2.4.3 *In the case $V_1 = 0$ and $V_2 = 0$, we can restrict the map $G^L_{S_1}$, and obtain the map D_2 that applies $\mathbb{R}^m$ into itself, where $D_2(x, u)$ is*

$$M^{\frac{1}{2}}A(\partial f + A^t M A)^{-1} A^t M^{\frac{1}{2}}[-u - 2z] + z + u$$

where

$$z = M^{\frac{1}{2}} B(\partial g + B^t M B)^{-1} B^t M^{\frac{1}{2}}(-u).$$

Note that if A and B are injective, then D_2 has full domain.

The map D_2 can be obtained in the form (2.53), considering that when $V_1 = V_2 = 0$, the matrix

$$S_3 = \begin{pmatrix} (1+\alpha)M^{\frac{1}{2}}A & \alpha M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}$$

verifies (2.52), and we obtain that $D_2 = G^L_{S_3}$.

The map D_2 *can also be obtained as the Douglas–Rachford operator [31], applied to the two maps*

$$-M^{\frac{1}{2}}A(\partial f)^{-1}(-A^tM^{\frac{1}{2}}) \quad \textit{and} \quad -M^{\frac{1}{2}}B(\partial g)^{-1}(-B^tM^{\frac{1}{2}}).$$

The corresponding sum of these two maps is exactly the **dual variational map** *(2.5) associated with the following optimization problem*

$$\min_{(x,y)} \quad [f(x)+g(z) : M^{\frac{1}{2}}Ax + M^{\frac{1}{2}}Bz = 0].$$

Alternatively we can consider, instead D_2, *the map* $\tilde{D}_2 := M^{-\frac{1}{2}}D_2M^{\frac{1}{2}}$, *i.e*

$$\tilde{D}_2(\bar{u}) = A(\partial f + A^tMA)^{-1}A^tM[-\bar{u}-2z] + z + \bar{u}$$

where

$$z = B(\partial g + B^tMB)^{-1}B^tM(-\bar{u}),$$

which is co-coercive w.r.t. the metric induced by M.

Example 2.4.2 *Let* $\gamma = 0$ $(\alpha = \beta)$. *We consider in (2.15)*

$$C_1 = (1+(\alpha+1)^2)A^tMA + R \quad \textit{and} \quad C_2 = (1+(\alpha+1)^2)B^tMB,$$

where R *is a positive semidefinite matrix. Then* V_1 *and* V_2 *in (2.21) are equal to*

$$V_1 = A^tMA + R \textit{ and } V_2 = B^tMB.$$

These matrices are associated with the additional proximal term considered in (2.33)-(2.36), which, as we have shown in Subsection 2.3.1 (Case $\gamma = 0$*), is related to Shefi-Teboulle algorithm type I [49]. We get :*

$$P = \begin{pmatrix} (1+(\alpha+1)^2)A^tMA + R & \alpha(2+\alpha)A^tMB & (1+\alpha)A^t \\ \alpha(2+\alpha)B^tMA & (1+(\alpha+1)^2)B^tMB & (1+\alpha)B^t \\ (1+\alpha)A & (1+\alpha)B & M^{-1} \end{pmatrix}.$$

The matrix

$$S_4 = \begin{pmatrix} R^{\frac{1}{2}} & 0 & 0 \\ M^{\frac{1}{2}}A & -M^{\frac{1}{2}}B & 0 \\ (1+\alpha)M^{\frac{1}{2}}A & (1+\alpha)M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}$$

satisfies (2.52) and hence the value $G^L_{S_4}(\widehat{x},\widehat{z},\widehat{y})$ *of the corresponding map* $G^L_{S_4}$, *that applies* $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ *into itself, is*

$$\begin{pmatrix} R^{\frac{1}{2}}x \\ M^{\frac{1}{2}}Ax - M^{\frac{1}{2}}Bz \\ M^{\frac{1}{2}}Ax + M^{\frac{1}{2}}Bz + \widehat{y} \end{pmatrix}$$

where

$$x = (\partial f + 2A^tMA + R)^{-1}(R^{\frac{1}{2}}\widehat{x} + A^tM^{\frac{1}{2}}(\widehat{z} - \widehat{y}))$$

$$z = (\partial g + 2B^tMB)^{-1}B^tM^{\frac{1}{2}}(-\widehat{z} - \widehat{y}).$$

Note that $G^L_{S_4}$ has full domain if $2A^tMA + R$ and $2B^tMB$ are assumed positive definite matrices.

Remark 2.4.4 *In the case that $R = 0$, we can restrict the map $G^L_{S_4}$, and obtain the map D_3 that applies $\mathbb{R}^m \times \mathbb{R}^m$ into itself, where $D_3(\widehat{z}, \widehat{y})$ is*

$$\begin{pmatrix} M^{\frac{1}{2}}A(\partial f + 2A^tMA)^{-1}A^tM^{\frac{1}{2}}(\widehat{z} - \widehat{y}) + M^{\frac{1}{2}}(\partial g + 2B^tMB)^{-1}B^tM^{\frac{1}{2}}(-\widehat{z} - \widehat{y}) \\ M^{\frac{1}{2}}A(\partial f + 2A^tMA)^{-1}A^tM^{\frac{1}{2}}(\widehat{z} - \widehat{y}) - M^{\frac{1}{2}}(\partial g + 2B^tMB)^{-1}B^tM^{\frac{1}{2}}(-\widehat{z} - \widehat{y}) + \widehat{y} \end{pmatrix}$$

The map D_3 can be obtained as the form (2.53), considering that when $V_1 = A^tMA$ and $V_2 = B^tMB$, the matrix

$$S_5 = \begin{pmatrix} M^{\frac{1}{2}}A & -M^{\frac{1}{2}}B & 0 \\ (1+\alpha)M^{\frac{1}{2}}A & (1+\alpha)M^{\frac{1}{2}}B & M^{-\frac{1}{2}} \end{pmatrix}.$$

satisfies (2.52), then we obtain that $D_3 = G^L_{S_5}$.

2.5 Rate of Convergence

The global rate of convergence of ADMM and other monotone operator splitting algorithms has motivated many research contributions that we cannot survey here (see [15] for example). We will recover these results for the generalized splitting scheme GSS with no further refinements (like uniform or strong convexity) and will remain in the framework of finite-dimensional spaces (see [2] for similar results in Hilbert spaces).

D. Davis and W. Yin [15] have show the ergodic and nonergodic convergence rate of the feasibility and objective function error related to the relaxed PRS and relaxed ADMM, which is a particular case of our general scheme as remarked in Subsection 2.3.1. Similarly, in this Section, without regularity assumption, we show the ergodic and nonergodic convergence rate of the constraint violations (feasibility) and objective function error related to the chain of steps (2.17) $\rightarrow$ (2.18) $\rightarrow$ (2.19) $\rightarrow$ (2.20), defined in Subsection 2.3.1, which is our main sequence associated with primal problem (P) defined in the first section.

With the same expressions of matrices P and U defined in (2.15) and (2.16), respectively, we get the following identity by using S satisfying $P = S^tS$ and explicit expressions of P and U,

$$\|(x,z,y)\|_P^2 = \|S(x,z,y)\|^2 = \|(x,z)\|_U^2 + \|M^{\frac{1}{2}}((1+\alpha)Ax + (1+\beta)Bz) + M^{-\frac{1}{2}}y\|^2. \tag{2.57}$$

Notice that for $\gamma = 0$ ($\beta = \alpha$),

$$\|(x,z)\|_U^2 = \|x\|_{V_1 - A^tMA}^2 + \|z\|_{V_2 - B^tMB}^2 + \|Ax - Bz\|_M^2$$

and for $\gamma = 1$ ($\beta = \alpha - 1$),

$$\|(x,z)\|_U^2 = \|x\|_{V_1}^2 + \|z\|_{V_2}^2. \tag{2.58}$$

Back to the sequence (2.17)–(2.20) and considering $w^k = (x^k, z^k, y^k)$, it holds from definition that

$$J_P^L w^k = (\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) \text{ and } w^{k+1} = \rho J_P^L w^k + (1-\rho)w^k. \tag{2.59}$$

The following proposition will be used later in Subsection 2.5.2 in order to estimate an upper bound of the optimal value of problem (P).

Proposition 2.5.1 *With the same notations as before and considering $w = (x,z,y) \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g) \times \mathbb{R}^m$, the following inequality holds:*

$$\|w^k - w\|_P^2 - \frac{2-\rho}{\rho}\|w^{k+1} - w^k\|_P^2 - \|w^{k+1} - w\|_P^2 \geq 2\rho\left[l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l(x,z,\tilde{y}^{k+1})\right]$$

Proof. Let $w = (x,z,y) \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g) \times \mathbb{R}^m$. Since $P(w^k - J_P^L w^k) \in L(J_P^L w^k)$, then using Proposition 2.3.1, it holds that

$$\left\langle J_P^L w^k - w, P(w^k - J_P^L w^k)\right\rangle \geq l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l(x,z,\tilde{y}^{k+1}). \tag{2.60}$$

On the other hand, from the symmetry of P, it holds

$$2\rho\left\langle J_P^L w^k - w, P(w^k - J_P^L w^k)\right\rangle = \|w^k - w\|_P^2 - \frac{2-\rho}{\rho}\|w^{k+1} - w^k\|_P^2 - \|w^{k+1} - w\|_P^2$$

So, replacing this last expression in (2.60), we get the desired inequality. ∎

In particular, from the inequality of last proposition, we get

$$\|w^k - w\|_P^2 - \|w^{k+1} - w\|_P^2 \geq 2\rho\left[l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l(x,z,\tilde{y}^{k+1})\right]. \tag{2.61}$$

This inequality will be used in Theorem 2.5.1 for approximating the optimal value of problem (P).

We note that Proposition 2.5.1 is a general version of the inequality given in Proposition 2 of [15] which is obtained by considering $A = I = -B$, $M = \gamma^{-1}I$ and P as in Remark 2.4.3, $w = (x, x, 0)$ (which implies $M^{-\frac{1}{2}}S_3 w = x$), $z = M^{-\frac{1}{2}}S_3 z^k$, and

$$z^+ = (T_{PRS})_\lambda(z) = (M^{-\frac{1}{2}}G^L_{S_3}M^{\frac{1}{2}})_{2\lambda}(M^{-\frac{1}{2}}S_3 z^k) = M^{-\frac{1}{2}}S_3 w^{k+1}.$$

Similarly, Proposition 2.5.1 is also a general version of the one given in Proposition 11 of [15] by considering $M = \gamma I$ and P as in Remark 2.4.3; $(\bar{x}^*, \bar{z}^*, \bar{y}^*)$ and z^* fixed points of $G^L_{S_3}$ and $(T_{PRS})_\lambda = (M^{\frac{1}{2}}G^L_{S_3}M^{-\frac{1}{2}})_{2\lambda}$, respectively; w^k satisfying $M^{\frac{1}{2}}S_3 w^k = z^k$ and $w = (\bar{x}^*, \bar{z}^*, 0)$ such that

$$M^{\frac{1}{2}}S_3 w = M^{\frac{1}{2}}S_3(\bar{x}^*, \bar{z}^*, \bar{y}^*) - \bar{y}^* = z^* - w^*$$

where $w^* = J_{\gamma(-B)(\partial g)^{-1}(-B^t)}(z^*)$.

2.5.1 Bounding the fixed-point residual

The fixed-point residual of operator $\rho G_S^T + (1-\rho)I_{q\times q}$ is the sequence with general term

$$\|(\rho G_S^T + (1-\rho)I_{q\times q})Sw^k - Sw^k\|^2$$

which, from (2.55), is equal to

$$\|Sw^{k+1} - Sw^k\|^2.$$

Since $\rho \in (0, 2)$, then $\rho G_S^T + (1-\rho)I_{q\times q}$ is non expansive and hence $\{\|Sw^{k+1} - Sw^k\|\}$ is non increasing. Summing over $k = 0, \cdots, N-1$ in (2.56), we get

$$\|Sw^k - Sw^{k-1}\|^2 \leq \frac{\rho}{(2-\rho)k}\|Sw^0 - Sw^*\|^2. \tag{2.62}$$

On the other hand, using the Jensen's inequality, we get

$$\|Sw^k - Sw^0\|^2 \leq 2\|Sw^k - Sw^*\|^2 + 2\|Sw^0 - Sw^*\|^2 \leq 4\|Sw^0 - Sw^*\|^2$$

and hence

$$\left\|\frac{1}{N}\sum_{k=1}^{N}(Sw^k - Sw^{k-1})\right\|^2 = \frac{1}{N^2}\|Sw^N - Sw^0\|^2 \leq \frac{4}{N^2}\|Sw^0 - Sw^*\|^2. \tag{2.63}$$

Notice that upper bounds (2.62) and (2.63) can also be deduced respectively from Theorem 1 "Notes on Theorem 1" and Theorem 2 developed in D. Davis and W. Yin [15].

2.5.2 Bounding the saddle-point gap

We consider the following **ergodic sequences** defined as: for $N \geq 1$,

$$\overline{x}_N := \frac{1}{N}\sum_{k=1}^{N} \tilde{x}^k, \;\; \overline{z}_N := \frac{1}{N}\sum_{k=1}^{N} \tilde{z}^k \quad \text{and} \quad \overline{y}_N := \frac{1}{N}\sum_{k=1}^{N} \tilde{y}^k.$$

Theorem 2.5.1 *With the same notations as before, we get the following rate of convergence:*

- **Ergodic Convergence:** *for any* $w = (x, z, y) \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g) \times I\!R^m$

$$l(\overline{x}_k, \overline{z}_k, y) - l(x, z, \overline{y}_k) \leq \frac{1}{2\rho k}\|Sw^0 - Sw\|^2. \tag{2.64}$$

- **Nonergodic Convergence:** *for any* $w^* = (x^*, z^*, y^*) \in \mathrm{sol}\,(V_L)$

$$l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y^*) - l(x^*, z^*, \tilde{y}^{k+1}) \leq \frac{1 + |1-\rho|}{\rho\sqrt{\rho(2-\rho)(k+1)}}\|Sw^0 - Sw^*\|^2. \tag{2.65}$$

Proof. Summing (2.61) over $k = 0, \cdots, N-1$, and applying the Jensen's inequality to the convex functions $l(\cdot, \cdot, y) - l(x, z, \cdot)$ for arbitrary fixed elements $x \in \mathrm{dom}\,(f)$, $z \in \mathrm{dom}\,(g)$ and $y \in I\!R^m$, where l is the lagrangian function defined in (2.10) of Section 2.3, we deduce the desired ergodic convergence.

Given $w^* \in \mathrm{sol}\,(V_L)$ and considering $w = w^*$ in (2.60), we get

$$\langle G_S^L Sw^k - Sw^*, Sw^k - G_S^L Sw^k\rangle \geq l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y^*) - l(x^*, z^*, \tilde{y}^{k+1}) \geq 0$$

and hence, from the Cauchy-Schwarz inequality and (2.55), we obtain

$$\frac{1}{\rho}\|G_S^L Sw^k - Sw^*\|\|Sw^{k+1} - Sw^k\| \geq l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y^*) - l(x^*, z^*, \tilde{y}^{k+1}). \tag{2.66}$$

On other hand, from (2.55) and since $\{\|Sw^{k+1} - Sw^*\|\}$ is non increasing, we get

$$\|G_S^L Sw^k - Sw^*\| = \|\frac{1}{\rho}(Sw^{k+1} - Sw^*) + (1 - \frac{1}{\rho})(Sw^k - Sw^*)\| \leq \frac{1 + |1-\rho|}{\rho}\|Sw^0 - Sw^*\|.$$

So, replacing this last expression and inequality (2.62) in expression (2.66), we deduce the desired nonergodic convergence. ∎

2.5.3 Bounding the constraint violation

We consider, for $N \geq 1$,

$$\widehat{x}_N := \frac{1}{N}\sum_{k=1}^{N} x^{k-1} \quad \text{and} \quad \widehat{z}_N := \frac{1}{N}\sum_{k=1}^{N} z^{k-1}.$$

We get the following result

Theorem 2.5.2 *With the same notations as before, for any $w^* \in \mathrm{sol}\,(V_L)$, we get the following rate of convergence:*

- **Ergodic Convergence:**

$$\|(\overline{x}_k - \widehat{x}_k, \overline{z}_k - \widehat{z}_k)\|_U^2 + \|A\overline{x}_k + B\overline{z}_k\|_M^2 \leq \frac{4}{\rho^2 k^2}\|Sw^0 - Sw^*\|^2.$$

- **Nonergodic Convergence:**

$$\|(\tilde{x}^k - x^{k-1}, \tilde{z}^k - z^{k-1})\|_U^2 + \|A\tilde{x}^k + B\tilde{z}^k\|_M^2 \leq \frac{1}{(2-\rho)\rho k}\|Sw^0 - Sw^*\|^2.$$

Proof. From (2.59) we have $w^k - w^{k-1} = \rho(\tilde{x}^k - x^{k-1},\ \tilde{z}^k - z^{k-1}, \tilde{y}^k - y^{k-1})$ and hence, from (2.19), we get

$$w^k - w^{k-1} = \rho(\tilde{x}^k - x^{k-1},\ \tilde{z}^k - z^{k-1},\ M[(1+\alpha)Ax^{k-1} + (1+\beta)Bz^{k-1} - \alpha A\tilde{x}^k - \beta B\tilde{z}^k]). \tag{2.67}$$

Summing over $k = 1, \cdots, N$, we obtain

$$\frac{1}{N}\sum_{k=1}^{N}(w^k - w^{k-1}) = \rho(\overline{x}_N - \widehat{x}_N,\ \overline{z}_N - \widehat{z}_N,\ M[(1+\alpha)A\widehat{x}_N + (1+\beta)B\widehat{z}_N - \alpha A\overline{x}_N - \beta B\overline{z}_N]).$$

Then from (2.57), we get

$$\frac{1}{\rho^2}\left\|\frac{1}{N}\sum_{k=1}^{N}(w^k - w^{k-1})\right\|_P^2 = \|(\overline{x}_N - \widehat{x}_N, \overline{z}_N - \widehat{z}_N)\|_U^2 + \|A\overline{x}_N + B\overline{z}_N\|_M^2$$

and hence, given $w^* \in \mathrm{sol}\,(V_L)$, we deduce from (2.63) the ergodic rate of convergence for constraint violations.

Using (2.67), from (2.57), we get

$$\frac{1}{\rho^2}\|w^k - w^{k-1}\|_P^2 = \|(\tilde{x}^k - x^{k-1}, \tilde{z}^k - z^{k-1})\|_U^2 + \|A\tilde{x}^k + B\tilde{z}^k\|_M^2$$

and hence, from (2.62), we deduce the nonergodic rate of convergence for constraint violations. ∎

We note that the particular case $\gamma = 1$, $V_1 = 0$ and $V_2 = 0$, which implies that $U = 0$, the two terms $\|(\overline{x}_k - \widehat{x}_k, \overline{z}_k - \widehat{z}_k)\|_U^2$ and $\|(\tilde{x}^k - x^{k-1}, \tilde{z}^k - z^{k-1})\|_U^2$ of inequalities in Theorem 2.5.2 are null and hence we recover the Theorem 15 of [15].

2.6 Application to the decomposition of multi-block optimization problems

To conclude our study, we consider the application of the generalized scheme GSS to the decomposition of some block structured convex optimization problems.

For $i \in \{1, ..., q\}$, let $f_i : I\!R^{n_i} \to \overline{I\!R}$ and $g : I\!R^m \to \overline{I\!R}$ are proper lsc convex functions, A_i and B matrices of order $p \times n_i$ and $p \times m$, respectively. We consider the following S-Model problem:

$$\begin{aligned} \inf_{(x_1,\cdots,x_q,z)} \quad & \sum_{i=1}^{q} f_i(x_i) + g(z) \\ s.t \quad & \sum_{i=1}^{q} A_i x_i + Bz = 0. \end{aligned}$$

This problem has been analyzed by many authors (see [29] for instance). We rewrite it into two different forms, (B_1) and (B_2), but with the same structure as (BP) defined below, then we rewrite (BP) as problem $(\overline{P})$ also defined below. Finally, we apply the algorithm (2.37)-(2.40) to this last problem.

The S- Model problem is equivalent to

$$\inf_{(x_1,\cdots,x_q,z)} \sum_{i=1}^{q} f_i(x_i) + g(z) + \delta_{\{0\}}(\sum_{i=1}^{q} A_i x_i + Bz). \qquad (B_1)$$

In this formulation the function g can be viewed as a function f_i. The associated dual problem of (B_1) is

$$\inf_{y^*} \sum_{i=1}^{q} (f_i^* \circ A_i^t) y^* + (g^* \circ B^t) y^*. \qquad (Ds)$$

Now, by considering $n = \sum_{i=1}^{q} n_i$ and $f : I\!R^n \to \overline{I\!R}$ defined as $f(x) := \sum_{i=1}^{q} f_i(x_i)$, the problem (Ds) can be written as

$$\inf_{y^*} \; f^* \circ \begin{bmatrix} A_1^t \\ \vdots \\ A_q^t \end{bmatrix} y^* + (g^* \circ B^t) y^*.$$

This is a composite problem whose associated dual problem is

$$\inf_{(x_1,\cdots,x_q)} \sum_{i=1}^{q} f_i(x_i) + (g^* \circ B^t)^* \circ (-\sum_{i=1}^{q} A_i x_i). \qquad (B_2)$$

We observe that in this last problem we reduce the number of variables considered in the S-Model problem and the function g acts now as regularization function.

Using the same notations as before, we define a problem having the same structures as problems (B_1) and (B_2):

$$V_P = \inf_{(x_1,\cdots,x_q)} \sum_{i=1}^{q} f_i(x_i) + (g^* \circ B^t)^* \circ (\sum_{i=1}^{q} A_i x_i). \qquad (BP)$$

In order to apply the splitting algorithm to problem (BP), we reformulate it to an appropriate optimization problem. To do it, consider
$K := \begin{pmatrix} I_{p\times p} & \cdots & I_{p\times p} \end{pmatrix} \in I\!R^{p\times pq}$ and

$$A := \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_q \end{pmatrix} \in I\!R^{pq\times n}.$$

So, problem (BP) can be formulated as

$$\inf_{x\in I\!R^n,\, z\in I\!R^{pq}} \quad \left[f(x) + (g^* \circ B^t)^* \circ Kz : Ax - z = 0\right]. \qquad (\overline{P})$$

Notice that this last formulation problem has a good separable structure.

We apply to problem $(\overline{P})$ the algorithm (2.37)-(2.40) developed in Subsection 2.3.1 (Case $\gamma = 1$). We assume that g verifies the following identity

$$\partial[(g^* \circ B^t)^* \circ K] = K^t(B(\partial g)^{-1}B^t)^{-1}K.$$

The saddle-point problem of $(\overline{P})$ is

$$\text{Find } (\bar{x},\bar{z},\bar{y}) \in I\!R^n \times I\!R^{pq} \times I\!R^{pq} \text{ such that } \ 0 \in \bar{L}(\bar{x},\bar{z},\bar{y}) \qquad (V_{\bar{L}})$$

where $\bar{L}$ is the maximal monotone map defined on $I\!R^n \times I\!R^{pq} \times I\!R^{pq}$ as

$$\bar{L}(x,z,y) := \begin{pmatrix} \partial f(x) \\ K^t(B(\partial g)^{-1}B^t)^{-1}Kz \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & I \\ -A & I & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}.$$

For $i \in \{1,...,q\}$, let M_i be an $p \times p$ symmetric positive definite matrix and Q_i be an $n_i \times n_i$ symmetric positive semi-definite matrix.

In order to take advantage of the separability of f, we take $V_1 = diag([Q_1,...,Q_q])$ and $M = diag([M_1,...,M_q])$, and we consider $V_2 = 0$ in order to calculate z^{k+1} using

alone the resolvent of ∂g. So, the related algorithm (2.37)-(2.40) takes the following structure:

$$\tilde{z}^{k+1} = (K^t(B(\partial g)^{-1}B^t)^{-1}K + M)^{-1}(MAx^k + y^k) \tag{2.68}$$

$$\tilde{y}^{k+1} = y^k + M(Ax^k - \tilde{z}^{k+1}) \tag{2.69}$$

$$\tilde{x}^{k+1} = (\partial f + A^tMA + V_1)^{-1}(V_1x^k + A^tM\tilde{z}^{k+1} - A^t\tilde{y}^{k+1}) \tag{2.70}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, z^k, y^k) \tag{2.71}$$

Because of the diagonal structure of expression (2.70) the calculation of $\tilde{x}^{k+1}$ is realized in parallel: for $i \in \{1, \cdots, q\}$,

$$\tilde{x}_i^{k+1} = (\partial f_i + A_i^tM_iA_i + Q_i)^{-1}(Q_ix_i^k + A_i^tM_i\tilde{z}_i^{k+1} - A_i^t\tilde{y}_i^{k+1}).$$

Now, in order to calculate $\tilde{z}^{k+1}$, the following identity is relevant:

Proposition 2.6.1 *With the same notations as before, the following identity holds*

$$(K^t(B(\partial g)^{-1}B^t)^{-1}K + M)^{-1}M = I - M^{-1}K^t\Sigma(I - B(\partial g + B^t\Sigma B)^{-1}B^t\Sigma)K$$

where Σ is a $p \times p$ matrix defined by

$$\Sigma := (KM^{-1}K^t)^{-1} = \left(\sum_{i=1}^{q} M_i^{-1}\right)^{-1}.$$

Proof. From Proposition 2.4.1, we have

$$(K^t(B(\partial g)^{-1}B^t)^{-1}K + M)^{-1}M = I - M^{-1}K^t(B(\partial g)^{-1}B^t + KM^{-1}K^t)^{-1}K$$

and hence by combining it with the following identity

$$(B(\partial g)^{-1}B^t + \Sigma^{-1})^{-1}\Sigma^{-1} = I - \Sigma B(\partial g + B^t\Sigma B)^{-1}B^t$$

obtained also from Proposition 2.4.1, we get the desired identity. ∎

So, using the identity of this last proposition, we can obtain an equivalent expression of $\tilde{y}^{k+1}$ in (2.69) but with a more tractable expression for computational purpose :

$$\tilde{y}^{k+1} = K^t\Sigma(I - B(\partial g + B^t\Sigma B)^{-1}B^t\Sigma)K(Ax^k + M^{-1}y^k). \tag{2.72}$$

It follow in particular that $\tilde{y}^{k+1} \in \operatorname{range} K^t$ and, by considering $y^k \in \operatorname{range} K^t$ in (2.71), we have that $y^{k+1} \in \operatorname{range} K^t$ and hence all the block components of $\tilde{y}^{k+1}$

(similarly of y^{k+1}) are equal. We denote by $\tilde{y}_c^{k+1}$ (resp y_c^{k+1}) such a block component of $\tilde{y}^{k+1}$ (resp y^{k+1}). Then,

$$\tilde{y}_c^{k+1} = \Sigma(I - B(\partial g + B^t\Sigma B)^{-1}B^t\Sigma)K(Ax^k + M^{-1}K^t y_c^k) \tag{2.73}$$

By denoting

$$\zeta^{k+1} := (\partial g + B^t\Sigma B)^{-1}B^t(\Sigma\sum_{j=1}^{q}(A_j x_j^k) + y_c^k)$$

we obtain, from (2.73),

$$\tilde{y}_c^{k+1} = y_c^k + \Sigma(\sum_{j=1}^{q}(A_j x_j^k) - B\zeta^{k+1}). \tag{2.74}$$

On the other hand, from (2.69), we get

$$\tilde{z}^{k+1} = Ax^k + M^{-1}K^t(y_c^k - \tilde{y}_c^{k+1})$$

which combining with (2.74), we deduce that for $i \in \{1, \cdots, q\}$,

$$\tilde{z}_i^{k+1} = A_i x_i^k - M_i^{-1}\Sigma\left(\sum_{j=1}^{q}(A_j x_j^k) - B\zeta^{k+1}\right).$$

Therefore we obtain the following algorithm, called **"Proximal Multi-block Algorithm"**.

Proximal Multi-block Algorithm
(PMA)

For $i \in \{1, \cdots, q\}$ set $Q_i \in I\!R^{n_i \times n_i}$ symmetric positive semi-definite, $M_i \in I\!R^{p\times p}$ symmetric positive definite. Set $\Sigma = \left(\sum_{i=1}^{q} M_i^{-1}\right)^{-1}$. Then for an arbitrary $(x^0, z^0, y_c^0) \in I\!R^n \times I\!R^{pq} \times I\!R^p$

Step 1. Find ζ^{k+1} such that

$$\zeta^{k+1} = \text{argmin}\left\{g(w) + \frac{1}{2}\|Bw - \sum_{j=1}^{q}(A_j x_j^k) - \Sigma^{-1}y_c^k\|_\Sigma^2\right\}$$

Step 2. Find $\tilde{z}^{k+1}$

For all $i \in \{1, ..., q\}$ **do**

Find $\tilde{z}_i^{k+1}$ such that

$$\tilde{z}_i^{k+1} = A_i x_i^k - M_i^{-1}\Sigma\left(\sum_{j=1}^{q}(A_j x_j^k) - B\zeta^{k+1}\right).$$

Step 3. Find $\tilde{y}_c^{k+1}$ such that

$$\tilde{y}_c^{k+1} = y_c^k + \Sigma\left(\sum_{j=1}^{q}(A_j x_j^k) - B\zeta^{k+1}\right).$$

end for

Step 4. Find $\tilde{x}^{k+1}$

For all $i \in \{1, ..., q\}$ **do**

Find $\tilde{x}_i^{k+1}$ such that

$$\tilde{x}_i^{k+1} = \operatorname{argmin}\left\{f_i(x_i) + \frac{1}{2}\|A_i x_i - \tilde{z}_i^{k+1} + M_i^{-1}\tilde{y}_c^{k+1}\|_{M_i}^2 + \frac{1}{2}\|x_i - x_i^k\|_{Q_i}^2\right\}$$

end for

Step 5. Find $(x^{k+1}, z^{k+1}, y_c^{k+1})$

$$(x^{k+1}, z^{k+1}, y_c^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}_c^{k+1}) + (1-\rho)(x^k, z^k, y_c^k).$$

The next proposition gives conditions in order to guarantee the convergence of PMA. The proof is a direct consequence of Proposition 2.3.5.

Proposition 2.6.2 *Let $\rho \in (0, 2)$. For $i \in \{1, ..., q\}$, assume that $Q_i \in \mathbb{R}^{n_i \times n_i}$ and $M_i \in \mathbb{R}^{p \times p}$ are symmetric, with Q_i positive semi-definite and M_i positive definite such that $Q_i + A_i^t M_i A_i$ is positive definite. If $\operatorname{sol}(V_{\bar{L}})$ is nonempty, then for an arbitrary $(x^0, z^0, y_c^0) \in \mathbb{R}^n \times \mathbb{R}^{pq} \times \mathbb{R}^p$, the sequence $(x^k, z^k, K^t y_c^k)$ generated by (PMA) converges to some element of $\operatorname{sol}(V_{\bar{L}})$.*

The **Separable Augmented Lagrangian Algorithm (SALA)** with multidimensional scaling has been proposed in [18] to solve a special case of the S-Model where $g = 0$ and $B = 0$. This algorithm can be recovered if instead of applying the algorithm (2.37)-(2.40) to problem $(\overline{P})$, we consider the algorithm (2.41)-(2.43) with $V_1 = V_2 = 0$. Therefore SALA is a particular version of (PMA).
The advantages of (PMA) are twofold: 1) the inclusion of the relaxing term $\rho \in (0, 2)$, which enables the accelaration of the algorithm, and 2) the additional proximal term $\|x_i - x_i^k\|_{Q_i}^2$ considered in the subproblems of Step 4, which improves the strong convexity of the proximal subproblem when we choose an adequate matrix Q_i. More specifically, considering σ_i and τ_i positive numbers holding $\sigma_i \tau_i \|A_i\|^2 \leq 1$ and choosing M_i and Q_i matrices defined as

$$M_i = \sigma_i I_{p \times p} \quad \text{and} \quad Q_i = \tau_i^{-1} I_{n_i \times n_i} - \sigma_i A_i^t A_i,$$

the conditions about matrices Q_i, M_i and $Q_i + A_i^t M_i A_i$ in Proposition 2.6.2 are verified and hence the subproblem in Step 4 of the Algorithm (PMA) becomes

$$\tilde{x}_i^{k+1} = \text{argmin} \left\{ f_i(x_i) + \frac{1}{2\tau_i} \|x_i - x_i^k - \tau_i[\sigma_i A_i^t \tilde{z}_i^{k+1} - \sigma_i A_i^t A_i x_i^k - A_i^t \bar{y}^{k+1}]\|^2 \right\}$$

which has an explicit solution in some particular cases, for instance $f_i(x_i) = \|x_i\|_1$.

Chapter 3

Decomposition techniques

In the last section of previous chapter we have developed a general algorithm termed Proximal Multi-block Algorithm (PMA) in order to solve a S-Model problem as in Section 2.6. This algorithm uses the proximal step of all the family $\{f_i\}_{i=1,\cdots,q}$ through a parallel processing for each iteration.

The special case considering $g \equiv 0$ and $B \equiv 0$ in the previous S-model, which will be termed **separable model with coupling constraint** (SMCC) is considered in Section 3.2. We propose an alternative way to use the proximal step of each f_i, separating the problem into two sub-block problems and considering the proximal step to one sub-block and then (at a linear combination of the preceding values) the proximal step is found for the other sub-block, both in parallel processing.

For this purpose, we first study in Section 3.1 the **separable model with coupling variable** (SMCV) using the duality scheme developed in Chapter 1, finding an adequate formulation for that problem, allowing recovery two know algorithms with multi–scaling parameters, and their relationship. We also show the numerical behavior of these two algorithms. Since SMCC can be formulated as a SMCV, we apply the results obtained for SMCV to SMCC, getting another way to recover SALA and DSALA, the last also contained in PMA.

In Section 3.3, we get two splitting algorithms, one for SMCV and the other for SMCC. Each algorithm separates the problem into two arbitrary sub-block problems considering the proximal step of all functions corresponding to each sub-block in parallel processing similar to the one obtained in the precedent section.

All found algorithms in this chapter are consequences of finding adequate formulations of the original problem and then applying popular algorithms. In this vein, in the last section we study an especial block optimization problem, that will be used for solving a stochastic optimization model problem, reformulating it and applying the multi–scaling ADMM (considering $V_1 = 0$ and $V_2 = 0$ in (2.41)–(2.43)).

In [11] the authors show that **"The direct extension of ADMM is not**

necessarily convergent". The authors in [54] showed an alternative way to deal with a multi-block optimization problem, reformulating it into two-block problem in order to apply the ADMM.

The main difference of the corresponding algorithm found by these authors regarding the one proposed in Section 3.2 of this chapter is the step order of subproblems intervening in the algorithm.

On the other hand, in [20] the author gives another way to solve problem SMCC using projective splitting methods.

3.1 The separable model with coupling variable (SMCV)

In this section we present the sum problem (or consensus problem as termed in [7]). In this book the authors give important references on recent advances and applications regarding the SMCV.

For every $i \in \{1, \cdots, q\}$, let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper lsc convex. We consider the following SMCV

$$\inf_{x \in \mathbb{R}^n} \sum_{i=1}^{q} f_i(x) \tag{S}$$

which by defining $f : \mathbb{R}^{nq} \to \bar{\mathbb{R}}$ as $f(z) := \sum_{i=1}^{q} f_i(z_i)$ for $z = (z_1, \cdots, z_q)$ with $z_i \in \mathbb{R}^n$, and W the $nq \times n$ matrix defined by $W = (I_{n\times n} \cdots I_{n\times n})^t$, that problem can be set as

$$\inf_{x \in \mathbb{R}^n} f \circ W(x) \tag{P_s^V}$$

which is termed **Primal projected problem** regarding to the following optimization problem with linear subspace constraint

$$\inf_{z} \quad \sum_{i=1}^{q} f_i(z_i) \tag{3.1}$$

$$s.t \quad z \in V \tag{3.2}$$

where $V := \text{range}\,(W) = \{z = (z_1, \cdots, z_q) \in \mathbb{R}^{nq} : z_1 = \cdots = z_q\}$. Clearly, this problem is also equivalent to

$$\inf_{z \in \mathbb{R}^{nq}} f(z) + \delta_V(z). \tag{$\widehat{P}_s$}$$

Notice that if we apply the Douglas-Rachford's method (DRM) to problem $(\widehat{P}_s)$, we obtain the so called [29] **"Proximal Decomposition algorithm (PDA)"**.

On the other hand, by making $V = \ker(K)$ where K is the $n(q-1) \times nq$ matrix defined by

$$K = \left[\begin{array}{cc} I_{n(q-1)\times n(q-1)} & -D \end{array} \right]$$

with D being the $n(q-1) \times n$ matrix defined by $D = (I_{n\times n} \cdots I_{n\times n})^t$, we obtain the following problem so called **Dual projected problem**:

$$\inf_{u^* \in \mathbb{R}^{n(q-1)}} f^* \circ K^t u^*. \qquad (D_s^V)$$

Now, by considering $h : \mathbb{R}^{n(q-1)} \to \overline{\mathbb{R}}$ defined as $h(y) := \sum_{i=1}^{q-1} f_i(y_i)$ for $y = (y_1, \cdots, y_q)$ with $y_i \in \mathbb{R}^n$, problem (D_s^V) can be formulated as

$$\inf_{u^* \in \mathbb{R}^{n(q-1)}} h^*(u^*) + f_q^* \circ -D^t u^*. \qquad (\bar{D}_s^V)$$

Since $D^t D = (q-1)I$, applying the DRM to problem $(\bar{D}_s^V)$, we obtain a suitable decomposition algorithm called **"Dual Proximal Decomposition algorithm (DPDA)"**.

By considering $V = \ker(K)$ in problem (3.1)-(3.2), it is transformed in

$$\begin{array}{ll} \inf\limits_{(u,s)} & \sum\limits_{i=1}^{q-1} f_i(u_i) + f_q(s) \\ s.t & u_i - s = 0,\ i = 1, \cdots, q-1 \end{array}$$

which is exactly the dual problem of (D_s^V).

In [7] the authors termed this problem **"Global variable consensus with regularization"**, where f_q represents the regularization function. In order to solve such a problem, they apply ADMM, which is equivalent to apply the Douglas-Rachford's method to problem $(\bar{D}_s^V)$.

Summarizing the previous discussions, PDA is DRM applied to problem (P_s), and DPDA is DRM applied to problem (D_s).

We now describe these two algorithms:

Proximal Decomposition algorithm (PDA)

For every $i \in \{1, \cdots, q-1\}$ set $M_i \in \mathbb{R}^{n\times n}$ symmetric positive definite. Then for an arbitrary $(x^0, y^0, u^0) \in \mathbb{R}^{nq} \times \mathbb{R}^n \times \mathbb{R}^{nq}$,

Step 1. Find x^{k+1}

For all $i \in \{1, ..., q\}$ **do**

find x_i^{k+1} the optimal value of problem

$$\inf_x \left[f_i(x) + \frac{1}{2}\|x - y^k + M_i^{-1} u_i^k\|_{M_i}^2 \right]$$

end for

Step 2. Find y^{k+1} such that

$$y^{k+1} = \left(\sum_{i=1}^{q} M_i\right)^{-1} \sum_{j=1}^{q} (M_i x_i^{k+1} + u_i^k)$$

Step 3. Find u^{k+1}

For all $i \in \{1, ..., q\}$ **do**

find u_i^{k+1} such that

$$u_i^{k+1} = u_i^k + M_i(x_i^{k+1} - y^{k+1})$$

end for

Dual Proximal Decomposition algorithm (DPDA)

For every $i \in \{1, \cdots, q-1\}$ set $M_i \in \mathbb{R}^{n \times n}$ symmetric positive definite matrix. Then for an arbitrary $(x^0, y^0, u^0) \in \mathbb{R}^{n(q-1)} \times \mathbb{R}^n \times \mathbb{R}^{n(q-1)}$

Step 1. Find x^{k+1}

For all $i \in \{1, ..., q-1\}$ **do**

find x_i^{k+1} the optimal value of the problem

$$\inf_x \left[f_i(x) + \frac{1}{2} \|x - y^k + M_i^{-1} u_i^k\|_{M_i}^2 \right]$$

end for

Step 2. Find y^{k+1} the optimal value of the problem

$$\inf_x \left[f_q(x) + \frac{1}{2} \left\| x - \left(\sum_{i=1}^{q-1} M_i\right)^{-1} \sum_{j=1}^{q-1} (M_i x_i^{k+1} + u_i^k) \right\|_{\sum_{i=1}^{q-1} M_i}^2 \right]$$

Step 3. Find u^{k+1}

For all $i \in \{1, ..., q-1\}$ **do**

find u_i^{k+1} such that

$$u_i^{k+1} = u_i^k + M_i(x_i^{k+1} - y^{k+1})$$

end for

Remark 3.1.1 *If $f_q = 0$, then DPDA becomes PDA applied to problem*

$$\inf_{x\in\mathbb{R}^n} \sum_{i=1}^{q-1} f_i(x).$$

So, apply PDA to problem (S) is equivalent to apply DPDA to the following problem

$$\inf_{x\in\mathbb{R}^n} \sum_{i=1}^{q} f_i(x) + 0(x).$$

Example 3.1.1 (Numerical illustration for the sum of three operators) *We apply the methods PDA and DPDA in order to solve the following problem*

$$\min_{x=(x_1,x_2)\in I\!R^2} \left\langle \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} x,\, x \right\rangle + \left\langle \begin{bmatrix} 8 & 0 \\ 0 & 1 \end{bmatrix} x,\, x \right\rangle + \left\langle \begin{bmatrix} 5 & 0 \\ 0 & 7 \end{bmatrix} x,\, x \right\rangle$$

which evidently $x^* = 0$ is the unique solution.

We consider, for $i \in \{1, 2, 3\}$, $M_i = \lambda I$ in the algorithm of (PDA); and for $j \in \{1, \cdots, q-1\}$, we set $M_j = \lambda I$ in the algoritm of $(DPDA)$.

The graph in (fig 3.1) describes the relationship of parameter λ for PDA and DPDA versus the necessary number of iterations in order to get an approximation of the optimal value with an error $(\|x_1^k - x^*\|)$ less than 10^{-8}, 10^{-25} and 10^{-60} respectively.

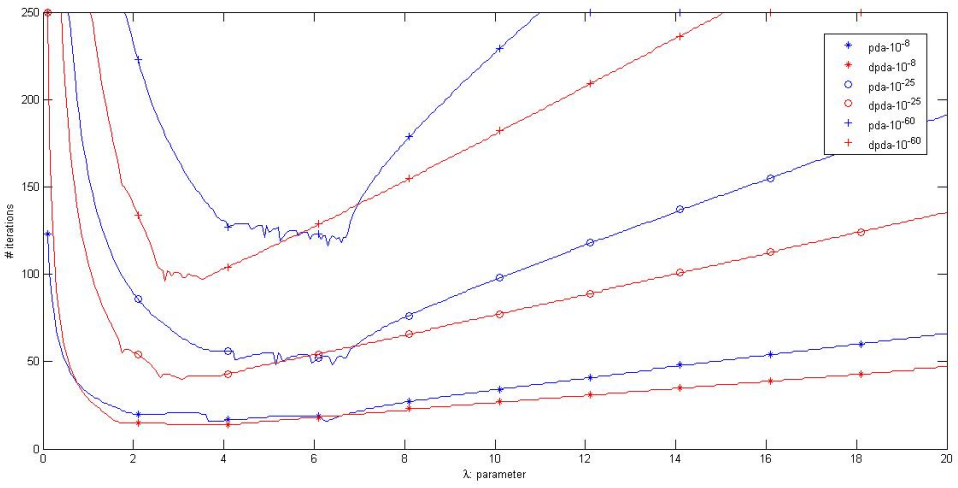


Figure 3.1: iteration vs parameter

3.2 Separable model with coupling constraints (SMCC)

For $i \in \{1, \cdots, q\}$, let $f_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ be proper lsc convex function and A_i be a $p \times n_i$ matrix. We consider the SMCC which can be expressed as

$$\begin{aligned} \inf_x \quad & \sum_{i=1}^{q} f_i(x_i) \\ s.t \quad & \sum_{i=1}^{q} A_i x_i = 0. \end{aligned}$$

This is a minimization problem over a linear subspace whose **dual projection problem** is

$$\inf_{u^* \in \mathbb{R}^p} \sum_{i=1}^{q} (f_i^* \circ A_i^t) u^* \qquad (D_{sm}^V)$$

which is a SMCV as described in the previous section. So, if we apply the algorithms PDA and DPDA to this problem, we obtain respectively the **Separable Augmenting Lagrangian Algorithm (SALA)** and the **Dual Separable Augmenting Lagrangian Algorithm (DSALA)**. In order to apply PDA or DPDA we will use the equivalent expression of the Moreau envelope function to each composite function $f_i \circ A_i^t$ as showed in Proposition 2.4.1.

We now describe the algorithms SALA and DSALA.

Separable Augmenting Lagrangian Algorithm (SALA)

For every $i \in \{1, \cdots, q\}$, set $M_i \in \mathbb{R}^{p \times p}$ symmetric positive definite. Set $\Sigma = (\sum_{i=1}^{q} M_i^{-1})^{-1}$ (similarly to the given in Proposition 2.6.1). Then for an arbitrary $(x^0, y^0, u^0) \in \mathbb{R}^{\sum_{i=1}^{q} n_i} \times \mathbb{R}^p \times \mathbb{R}^{pq}$,

Step 1. Find x^{k+1}

For all $i \in \{1, ..., q\}$ **do**

find x_i^{k+1} the optimal value of the problem

$$\inf_x \left[f_i(x) + \frac{1}{2} \| A_i x - M_i^{-1} y^k + u_i^k \|_{M_i}^2 \right]$$

end for

Step 2. Find y^{k+1} such that

$$y^{k+1} = y^k - \Sigma(\sum_{j=1}^{q} (A_i x_i^{k+1}))$$

Step 3. Find u^{k+1}

For all $i \in \{1, ..., q\}$ **do**

find u_i^{k+1} such that

$$u_i^{k+1} = -A_i x_i^{k+1} + M_i^{-1}\Sigma(\sum_{i=1}^{q}(A_i x_i^{k+1}))$$

end for

Dual Separable Augmenting Lagrangian Algorithm (DSALA)

For $i \in \{1, \cdots, q-1\}$, let $M_i \in \mathbb{R}^{p\times p}$ symmetric positive definite. Set $\Sigma = (\sum_{i=1}^{q-1} M_i^{-1})^{-1}$. Then for an arbitrary $(x^0, y^0, u^0) \in \mathbb{R}^{\sum_{i=1}^{q-1} n_i} \times \mathbb{R}^p \times \mathbb{R}^{p(q-1)}$

Step 1. Find x^{k+1}

For all $i \in \{1, ..., q-1\}$ **do**

find x_i^{k+1} the optimal value of the problem

$$\inf_x \left[f_i(x) + \frac{1}{2}\|A_i x - M_i^{-1} y^k + u_i^k\|_{M_i}^2 \right]$$

end for

Step 2. Calculate w^{k+1} the optimal value of the problem

$$\inf_x \left[f_q(x) + \frac{1}{2}\|A_q x - \Sigma^{-1} y^k + \sum_{j=1}^{q-1}(A_i x_i^{k+1})\|_{\Sigma}^2 \right]$$

Step 3. Find y^{k+1} such that

$$y^{k+1} = y^k - \Sigma(\sum_{j=1}^{q-1}(A_i x_i^{k+1}) + A_q w^{k+1})$$

Step 4. Find u^{k+1}

For all $i \in \{1, ..., q-1\}$ **do**

find u_i^{k+1} such that

$$u_i^{k+1} = -A_i x_i^{k+1} + M_i^{-1}\Sigma(\sum_{i=1}^{q-1}(A_i x_i^{k+1}) + A_q w^{k+1})$$

end for

Remark 3.2.1

- *If $A_q = 0$, then DSALA becomes SALA applied to problem*

$$\begin{aligned} \inf_x \quad & \sum_{i=1}^{q-1} f_i(x_i) \\ s.t \quad & \sum_{i=1}^{q-1} A_i x_i = 0. \end{aligned}$$

- *For $i \in \{1, ..., q\}$, let $f_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^p \to \overline{\mathbb{R}}$ proper lsc convex functions, and A_i be $p \times n_i$ matrix. We consider the Multi-block problem*

$$\min_x \sum_{i=1}^{q} f_i(x_i) + g(\sum_{i=1}^{q} A_i x_i)$$

also termed in [7] as **Sharing problem**, *where f_i is called local cost function and g the shared objective.*

It is clear that Multi-block problem includes SMCC problem by considering $g = \delta_{\{0\}}$. Conversely, we can formulate the Multi-block problem as a SMCC by making $x_{q+1} = \sum_{i=1}^{q} A_i x_i$:

$$\begin{aligned} \inf_x \quad & \sum_{i=1}^{q} f_i(x_i) + g(x_{q+1}) \\ s.t \quad & \sum_{i=1}^{q} A_i x_i - x_{q+1} = 0. \end{aligned}$$

So, by applying the SALA or DSALA algorithms to this last problem, we get a splitting algorithm for Multi-block problem.

Example 3.2.1 (Numerical illustration for the sum of operators) *We consider the algorithms SALA and DSALA in order to solve the following constrained problem*

$$\min_{x_1,x_2,x_3} \quad \left\langle \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} x_1,\, x_1 \right\rangle + \left\langle \begin{bmatrix} 8 & 0 \\ 0 & 1 \end{bmatrix} x_2,\, x_2 \right\rangle + \left\langle \begin{bmatrix} 5 & 0 \\ 0 & 7 \end{bmatrix} x_3,\, x_3 \right\rangle$$

subject to

$$\begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix} x_1 + \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix} x_2 + \begin{bmatrix} 1 & 3 \\ -1 & 5 \end{bmatrix} x_3 = 0$$

whose solution is $x_1^* = x_2^* = x_3^* = 0$.

We consider, for $i \in \{1, \cdots, q\}$, $M_i = \lambda I$ *in SALA; and for* $j \in \{1, \cdots, q-1$, $M_j = \lambda I$ *in DSALA.*

The graph in (fig 3.2) describes the relationship of parameter λ *versus the necessary number of iterations in order to get an approximation of the optimal value with an error* ($\|x_1^k - x_1^*\|$) *less than* 10^{-6}, 10^{-25} *and* 10^{-50} *respectively.*

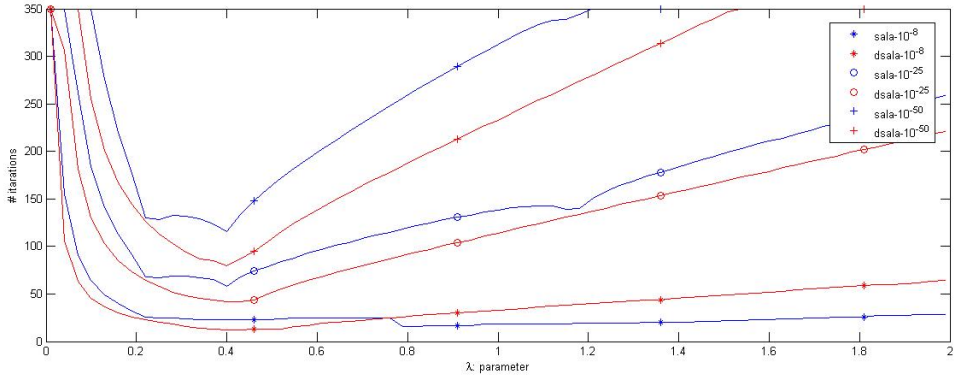


Figure 3.2: iteration vs parameter

3.3 Proximal separation into two sub-blocks

The algorithm PDA described in the first section of this chapter considers the proximal step of all family $\{f_i\}_{i=1,\cdots,q}$ in parallel processing. Unlike this algorithm, DPDA separates the family into two sub-families or sub-blocks: one consisting of $q-1$ functions, $f_1, \cdots, f_{q-1}$, and the other consisting of f_q. Then, the proximal step of each function of sub-block $\{f_i\}_{i=1,\cdots,q-1}$ is found in parallel processing and then, at a linear combination of all these values, the proximal step of f_q is found.

B. He and X.Yuan [25] considered a linear programming model and a corresponding algorithm separating the problem into two adequate sub-blocks. They then determinate the proximal step of all functions corresponding to one of these sub-block in parallel processing and, using these values, determinate the proximal step for all functions corresponding to the second sub-block in parallel processing too.

We show in this section that this procedure can also be applied for general setting by separating a given problem into two arbitrary sub-blocks. In practice, is more adequate to separate into two sub-blocks taking into account the difficulty to determinate their proximal steps.

3.3.1 The separable model with coupling variables

Let r and s be two positive numbers such that $r+s=q$ and $r \geq s$. We separate the block of functions into two sub-blocks: the first one consisting of r functions $\{f_i\}_{i=1,\cdots,r}$ and the second one consisting of s functions $\{f_{i+r}\}_{i=1,\cdots,s}$, and the matrices coupling these two sub-blocks: B_1 and B_2 of order $n(q-1)\times nr$ and $n(q-1)\times ns$, respectively, defined as

$$B_1 = \begin{pmatrix} I_{n\times n} & & & & \\ & C & & & \\ & & \ddots & & \\ & & & C & \\ & & & & I_{n(r-s)\times n(r-s)} \end{pmatrix} \text{ and } B_2 = \begin{pmatrix} -C & & & & \\ & \ddots & & & \\ & & -C & & \\ & & & -I_{n\times n} & \\ & & & -U & \end{pmatrix}$$

with $C = \begin{pmatrix} I_{n\times n} & I_{n\times n} \end{pmatrix}^t$ and U a $n(r-s)\times n$ matrix such that $U^t = \begin{pmatrix} I_{n\times n} & \cdots & I_{n\times n} \end{pmatrix}$.

Since $V = \ker \begin{pmatrix} B_1 & B_2 \end{pmatrix}$, then problem (3.1)-(3.2) can be equivalently formulated as

$$\inf_{(x,z)} \quad \sum_{i=1}^{r} f_i(x_i) + \sum_{j=1}^{s} f_{j+r}(z_j) \tag{3.3}$$

$$s.t \quad B_1 x + B_2 z = 0. \tag{3.4}$$

On the other hand, the diagonal structure of $B_1^t B_1$ and $B_2^t B_2$ allows to solve in parallel processing the proximal step of each sub-block when the algorithm (2.41)-(2.43) with $V_1 = 0$ and $V_2 = 0$, is applies to problem (3.3)-(3.4). Explicitly $B_1^t B_1$ and $B_2^t B_2$ are respectively equal to:

$$\begin{pmatrix} I_{n\times n} & & & & \\ & 2I_{n\times n} & & & \\ & & \ddots & & \\ & & & 2I_{n\times n} & \\ & & & & I_{n(r-s)\times n(r-s)} \end{pmatrix} \text{ and } \begin{pmatrix} 2I_{n\times n} & & & \\ & \ddots & & \\ & & 2I_{n\times n} & \\ & & & \eta I_{n\times n} \end{pmatrix}$$

where $\eta = r - s + 1$.

So, applying the algorithm (2.41)-(2.43) with $V_1 = 0$, $V_2 = 0$ and $M = \lambda$ (which is a slight variant of ADMM) to problem (3.3)-(3.4) by considering $f(x) = \sum_{i=1}^{r} f_i(x_i)$ and $g(z) = \sum_{j=1}^{s} f_{j+r}(z_j)$, we get the following algorithm:

$$\begin{aligned} x^{k+1} &= (\partial f + \lambda B_1^t B_1)^{-1}(\lambda B_1^t B_2 z^k + B_1^t u^k) \\ z^{k+1} &= (\partial g + \lambda B_2^t B_2)^{-1}(\lambda B_2^t B_1 x^{k+1} + B_2^t u^k) \\ u^{k+1} &= u^k + \lambda(B_1 x^{k+1} + B_2 z^{k+1}) \end{aligned}$$

Since $B_1^t B_1$ is diagonal, the calculation of $\tilde{x}^{k+1}$ can be realized in parallel processing: for $i \in \{1, \cdots, r\}$,

$$x_i^{k+1} = (\partial f_i + \lambda \alpha_i I_{n\times n})^{-1}(\lambda (B_1^t B_2 z^k)_i + (B_1^t u^k)_i).$$

Similarly, since $B_2^t B_2$ is diagonal, the calculation of $\tilde{x}^{k+1}$ is also realized in parallel processing: for $j \in \{1, \cdots, s\}$,

$$z_j^{k+1} = (\partial f_{j+r} + \lambda \alpha_{j+r} I_{n\times n})^{-1}(\lambda (B_2^t B_1 x^{k+1})_j + (B_2^t u^k)_j)$$

where $\{\alpha_i\}_{i=1,\cdots,q}$ is defined as

$$\alpha_i = \begin{cases} 1 & \text{if} \quad i \in \{1, s+1, \cdots, r\}, \\ 2 & \text{if} \quad i \in \{2, \cdots, s, r+1, \cdots, q-1\}, \\ r-s+1 & \text{if} \quad i = q. \end{cases} \tag{3.5}$$

Then we get the following algorithm.

Two Sub-blocks Proximal Decomposition algorithm (2sb-PDA)

Set the finite sequence $\{\alpha_i\}_{i=1,\cdots,q}$ previously defined and λ a positive number. Then for an arbitrary $(x^0, z^0, u^0) \in I\!R^{rn} \times I\!R^{sn} \times I\!R^{n(q-1)}$,

Step 1. Find x^{k+1}

For all $i \in \{1, ..., r\}$ **do**
if $i = 1$ **then** $\xi_i = -z_i^k$, $\mu_i = u_i^k$, **end if**
if $2 \le i \le s$ **then** $\xi_i = -z_{i-1}^k - z_i^k$, $\mu_i = u_{2(i-1)}^k + u_{2i-1}^k$, **end if**
if $i \ge s+1$ **then** $\xi_i = -z_s^k$, $\mu_i = u_{i+s-1}^k$, **end if**
find x_i^{k+1} the optimal value of the problem

$$\inf_x \left[f_i(x) + \frac{\lambda \alpha_i}{2} \| x + \alpha_i^{-1} \xi_i + (\lambda \alpha_i)^{-1} \mu_i \|^2 \right]$$

end for

Step 2. Find z^{k+1}

For all $j \in \{1, ..., s\}$ **do**
if $j \le s-1$ **then** $\zeta_j = -x_i^{k+1} - x_{j+1}^{k+1}$, $\nu_j = -u_{2j-1}^k - u_{2j}^k$, **end if**
if $j = s$ **then** $\zeta_i = -\sum_{i=s}^r x_i^{k+1}$, $\nu_i = -\sum_{i=s}^r u_{i+s-1}^k$, **end if**
find z_i^{k+1} the optimal value of the problem

$$\inf_z \left[f_{r+j}(z) + \frac{\lambda \alpha_{r+j}}{2} \| z + \alpha_{j+r}^{-1} \zeta_j + (\lambda \alpha_{r+j})^{-1} \nu_j \|^2 \right]$$

end for

Step 3. Find u^{k+1}

For all $i \in \{1, ..., q-1\}$ **do**

if $1 \leq i \leq 2s-2$ **then**

$a_i = x^{k+1}_{(i-(i \mod 2))/2+1}$, $b_i = -z^{k+1}_{(i+(i \mod 2))/2}$ end if

If $i \geq 2s-1$ **then**

$a_i = x^{k+1}_{i-s+1}$, $b_i = -z^{k+1}_{s}$ end if

Calculated u_i^{k+1} such that

$$u_i^{k+1} = u_i^k + \lambda(a_i^{k+1} + b_i^{k+1})$$

end for

3.3.2 The separable model with coupling constraints

Since SMCC can be formulated as a SMCV (D^V_{sm}), we can get from the previous algorithms given for the last model an splitting algorithm for the SMCC which separates it into two sub-block problems: the first one corresponding to $\{f_i\}_{i=1,\cdots,r}$ and the second one corresponding to $\{f_{r+j}\}_{j=1,\cdots,s}$. Then, the proximal step of all functions corresponding the first sub-block is obtained by parallel processing, and then, using these values, the proximal step of the other functions is also obtained in parallel processing.

So, the SMCC can also be formulated as

$$\inf_{(x,z)} \quad \sum_{i=1}^{r}(f_i^* \circ A_i^t)(x_i) + \sum_{j=1}^{s}(f_{r+j}^* \circ A_{r+j}^t)(z_j) \tag{3.6}$$

$$s.t \quad B_1x + B_2z = 0 \tag{3.7}$$

where B_1 and B_2 are the matrices corresponding to expression (3.4).

Similarly to algorithm (2sb-PDA), applying the algorithm (2.41)-(2.43) with $V_1 = 0$, $V_2 = 0$ and $M = \lambda$ to this new formulation problem we get the following algorithm called (2sb-SALA).

Two Sub-blocks Separable Augmenting Lagrangian Algorithm (2sb-SALA)

Set the finite sequence $\{\alpha_i\}_{i=1,\cdots,q}$ defined in (3.5) and λ a positive real number. Then for an arbitrary $(x^0, z^0, u^0) \in I\!R^{rn} \times I\!R^{sn} \times I\!R^{n(q-1)}$

Step 1. Find x^{k+1}

For all $i \in \{1, ..., r\}$ **do**

if $i = 1$ **then** $\xi_i = -z_i^k,\ \ \mu_i = u_i^k$, **end if**
if $2 \leq i \leq s$ **then** $\xi_i = -z_{i-1}^k - z_i^k,\ \ \mu_i = u_{2(i-1)}^k + u_{2i-1}^k$, **end if**
if $i \geq s+1$ **then** $\xi_i = -z_s^k,\ \ \mu_i = u_{i+s-1}^k$, **end if**
find $\tilde{x}_i^{k+1}$ the optimal value of the problem

$$\inf_x \left[f_i(x) + \frac{1}{2\lambda\alpha_i} \|A_i x + \lambda\xi_i + \mu_i\|^2 \right]$$

calculated $x^{k+1} = -\alpha_i^{-1}\xi_i - (\lambda\alpha_i)^{-1}\mu_i - (\lambda\alpha_i)^{-1}A_i\tilde{x}_i^{k+1}$
end for

Step 2. Find z^{k+1}
For all $j \in \{1, ..., s\}$ **do**
if $j \leq s-1$ **then** $\zeta_j = -x_i^{k+1} - x_{j+1}^{k+1},\ \ \nu_j = -u_{2j-1}^k - u_{2j}^k$, **end if**
if $j = s$ **then** $\zeta_i = -\sum_{i=s}^r x_i^{k+1},\ \ \nu_i = -\sum_{i=s}^r u_{i+s-1}^k$, **end if**
find $\tilde{z}_i^{k+1}$ the optimal value of the problem

$$\inf_z \left[f_{r+j}(z) + \frac{1}{2\lambda\alpha_{r+j}} \|A_{r+j} z + \lambda\zeta_j + \nu_j\|^2 \right]$$

calculated $x^{k+1} = -\alpha_{j+r}^{-1}\zeta_j - (\lambda\alpha_{r+j})^{-1}\nu_j - (\lambda\alpha_{r+j})^{-1}A_{r+j}\tilde{z}_j^{k+1}$
end for

Step 3. Find u^{k+1}
For all $i \in \{1, ..., q-1\}$ **do**
if $1 \leq i \leq 2s-2$ **then**
$a_i = x_{(i-(i \mod 2))/2+1}^{k+1},\ \ b_i = -z_{(i+(i \mod 2))/2}^{k+1}$ **end if**
If $i \geq 2s-1$ **then**
$a_i = x_{i-s+1}^{k+1},\ \ b_i = -z_s^{k+1}$ **end if**
Calculated u_i^{k+1} such that

$$u_i^{k+1} = u_i^k + \lambda(a_i^{k+1} + b_i^{k+1})$$

end for

3.4 Multi-block optimization problem

We consider the following problem

$$\min_{x=(x_1,\cdots,x_q)} \sum_{i=1}^q f_i(x_i) + g(\sum_{i=1}^q A_i x_i) + s(x) \qquad (P_{sc})$$

where for $i \in \{1, ..., q\}$, $f_i : I\!R^{n_i} \to \overline{I\!R}$, $g : I\!R^p \to \overline{I\!R}$ and $s : I\!R^n \mapsto \overline{I\!R}$ $(n = \sum_{i=1}^q n_i)$ are proper lsc convex functions, and A_i are matrices of order $p \times n_i$.

By considering $f : I\!R^n \to \overline{I\!R}$ defined as $f(x) = \sum_{i=1}^{q} f_i(x_i)$ and matrices K and A of order $p \times pq$ and $pq \times n$, respectively, defined by

$$K := \begin{pmatrix} I_{p\times p} & \cdots & I_{p\times p} \end{pmatrix} \quad \text{and} \quad A := \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_q \end{pmatrix},$$

then (P_{sc}) can be set as

$$\min_x \quad f(x) + (g \circ K, s) \begin{pmatrix} A \\ I_{n\times n} \end{pmatrix} x.$$

For every $i \in \{1, ..., q\}$, set M_i an $p \times p$ symmetric positive definite matrix and Q_i an $n_i \times n_i$ symmetric positive definite matrix. We define the blocks diagonal matrices $\widehat{M} = \text{diag}\,(M_1, \cdots, M_q)$ and $Q = \text{diag}\,(Q_1, \cdots, Q_q)$.

We apply the algorithm (2.41)-(2.43) with $V_1 = 0$ and $V_2 = 0$, which is equivalent to the scaling ADMM. So, in order to take advantage of the separability of ∂f we consider the matrix of scaling defined as $M = \text{diag}\,(\widehat{M}, Q)$.

$$x^{k+1} = (\partial f + Q + A^t\widehat{M}A)^{-1}(A^t\widehat{M}z_1^k + Qz_2^k - A^tu_1^k - u_2^k) \tag{3.8}$$

$$z_1^{k+1} = (K^t\partial gK + \widehat{M})^{-1}(\widehat{M}Ax^{k+1} + u_1^k) \tag{3.9}$$

$$z_2^{k+1} = (\partial s + Q)^{-1}(Qx^{k+1} + u_2^k) \tag{3.10}$$

$$u_1^{k+1} = u_1^k + \widehat{M}(Ax^{k+1} - z_1^{k+1}) \tag{3.11}$$

$$u_2^{k+1} = u_2^k + Q(x^{k+1} - z_2^{k+1}) \tag{3.12}$$

Then using similar techniques described in Section 6 of Chapter 2, we get the following algorithm called **"Separable Multi-block for sum of three blocks function (SMS3BF)"**

(SMS3BF)

For every $i \in \{1, \cdots, q\}$, set M_i and Q_i symmetric positive definite matrices of order $p \times p$ and $n_i \times n_i$, respectively, and $\Sigma = (\sum_{i=1}^{q} M_i^{-1})^{-1}$. Then for arbitrary $(x^0, y^0, z^0, \bar{v}^0, u^0) \in I\!R^n \times I\!R^{pq} \times I\!R^n \times I\!R^p \times I\!R^n$,

Step 1. Find x^{k+1}

For all $i \in \{1, ..., q\}$ **do**
find x_i^{k+1} such that

$$x_i^{k+1} = (\partial f_i + Q_i + A_i^tM_iA_i)^{-1}(A_i^tM_iy_i^k + Q_iz_i^k - A_i^t\bar{v}^k - u_i^k)$$

end for

Step 2. Calculate w^{k+1}

$$w^{k+1} = (\partial g + \Sigma)^{-1}(\Sigma \sum_{j=1}^{q}(A_j x_j^{k+1}) - \bar{v}^k)$$

Step 3. Find y^{k+1}

For all $i \in \{1, ..., q\}$ **do**

find y_i^{k+1} such that

$$y_i^{k+1} = A_i x_i^{k+1} - M_i^{-1}\Sigma(\sum_{j=1}^{q}(A_j x_j^{k+1}) + w^{k+1}).$$

end for

Step 4. Calculate s^{k+1}

$$z^{k+1} = (\partial s + Q)^{-1}(Qx^{k+1} + u^k)$$

Step 5. Calculate $\bar{v}^{k+1}$

$$\bar{v}^{k+1} = \bar{v}^k + \Sigma(\sum_{j=1}^{q}(A_j x_j^{k+1}) + w^{k+1}).$$

Step 6. find u^{k+1}

For all $i \in \{1, ..., q\}$ **do**

find u_i^{k+1} such that

$$u_i^{k+1} = u_i^k + Q_i(x_i^{k+1} - z_i^{k+1})$$

end for

Remark 3.4.1 *If $s \equiv 0$ in problem (P_{sc}), then this problem becomes a multi-block problem*

$$\min_x \sum_{i=1}^{q} f_i(x_i) + g\left(\sum_{i=1}^{q} A_i x_i\right)$$

which is treated in Section 6 of Chapter 2, considering $B = -I$. In that case, the k-th iteration in Step 4 satisfies

$$Qz^{k+1} = Qx^{k+1} + u^k$$

and hence from Step 6, we get $u_i^{k+1} = 0$ for all $i \in \{1, ..., q\}$. So, the $(k+1)$-th iteration in Step 1, becomes

$$x_i^{k+2} = (T_i + Q_i + A_i^t D_i A_i)^{-1}(A_i^t D_i y_i^{k+1} + Q_i x_i^{k+1} - A_i^t \bar{v}^{k+1}),\ z^{k+2} = x^{k+2} \ \text{ and } \ u^{k+2} = 0$$

implying that the variables z and u in the algorithm turn out obsolete. The resulting algorithm becomes equivalent to algorithm (PMA).

3.4.1 Aplication to a stochastic problem

We consider a stochastic problem with finite scenarios, which can be reformulated as a problem defined over a Euclidean linear space, having the same structure of problem (P_{sc}) considered at the beginning of this section. Then we apply (SMS3BF) in order to obtain a splitting algorithm for such a stochastic problem.

Let consider a finite set Ξ of scenarios and a corresponding positive probability function p. We also consider the linear space $\mathcal{L}$ consisting of all mapping from Ξ to $\mathbb{R}^n := \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_T}$, and the following inner product related

$$\langle X, Y \rangle_{\mathcal{L}} = \sum_{\xi \in \Xi} p(\xi) \langle X(\xi), Y(\xi) \rangle \tag{3.13}$$

Set $E := card(\Xi)$ and $G := \{1, \cdots, T\}$. We consider the linear space $\mathbb{R}^{nE} := \mathbb{R}^{n_1 E} \times \cdots \times \mathbb{R}^{n_T E}$, and the following related inner product

$$\langle ((x_t^\xi)_{\xi \in \Xi})_{t \in G}, ((y_t^\xi)_{\xi \in \Xi})_{t \in G} \rangle_{\mathbb{R}^{nE}} = \sum_{\xi \in \Xi} p(\xi) \langle (x_t^\xi)_{t \in G}, (y_t^\xi)_{t \in G} \rangle.$$

There is a relationship between $\mathbb{R}^{nE}$ and $\mathcal{L}$ through the following isomorphic mapping $W : (\mathbb{R}^{nE}, \langle \cdot, \cdot \rangle_{\mathbb{R}^{nE}}) \to (\mathcal{L}, \langle \cdot, \cdot \rangle_{\mathcal{L}})$ such that for $x := ((x_t^\xi)_{\xi \in \Xi})_{t \in G}$, the value $W(x) \in \mathcal{L}$ satisfies

$$W(x)(\xi) = (x_1^\xi, \cdots, x_T^\xi) \text{ for all } \xi \in \Xi.$$

For each $t \in G$, we consider $\mathcal{A}_t$, a partition of Ξ such that $\mathcal{A}_{t+1}$ is a refinement of $\mathcal{A}_t$. The nonanticipativity subspace of $\mathcal{L}$ is defined as (see (1.3) in [46])

$$\mathcal{N} := \{X \in \mathcal{L} : X_t \text{ is constant on each } A \in \mathcal{A}_t \quad \text{for } t \in G\}.$$

For $t \in G$ and $\xi \in \Xi$, we consider B_t^ξ an $m_\xi \times n_t$ matrix and then $C(\xi) := \ker \begin{pmatrix} B_1^\xi & \cdots & B_T^\xi \end{pmatrix}$.

The following stochastic optimization problem is considered

$$\min_{X \in \mathcal{L}} \left[E_\xi \sum_{t=1}^{T} g_t(X_t(\xi), \xi) : s.t\ X \in \mathcal{N} \text{ and } \sum_{t=1}^{T} B_t^\xi X_t(\xi) = 0,\ \forall \xi \in \Xi \right]. \tag{SP}$$

Reformulating this problem in the Euclidean linear space $\mathbb{R}^{nE}$, we get

$$\min_{x = ((x_t^\xi)_{\xi \in \Xi})_{t \in G}} \sum_{\xi \in \Xi} p(\xi) \sum_{t=1}^{T} g_t(x_t^\xi, \xi) + \sum_{\xi \in \Xi} \delta_{\{0_{m_\xi}\}} (\sum_{t=1}^{T} B_t^\xi x_t^\xi) + \delta_{W^{-1}\mathcal{N}}(x)$$

This last problem can be reformulated having the same structure of (P_{sc}) considered at the beginning of this section:

$$\min_{x=((x_t^\xi)_{\xi\in\Xi})_{t\in G}} \sum_{t=1}^{T}\left[\sum_{\xi\in\Xi} p(\xi)g_t(x_t^\xi,\xi)\right] + \delta_{\bigotimes_{\xi\in\Xi}\{0_{m_\xi}\}}(\sum_{t=1}^{T} B_t(x_t^\xi)_{\xi\in\Xi}) + \delta_{W^{-1}\mathcal{N}}(x) \quad (3.14)$$

where

$$B_t := \begin{pmatrix} B_t^{\xi_1} & & \\ & \ddots & \\ & & B_t^{\xi_E} \end{pmatrix}.$$

Since algorithm SMS3BF applied to last problem uses the proximal step of $\delta_{W^{-1}\mathcal{N}}$ through the isomorphic mapping W, we get an equivalent expression for the general resolvent of $\partial\delta_{W^{-1}\mathcal{N}}$ with respect to matrix $\bar{Q} := \lambda \mathrm{diag}\,(\bar{Q}_1, \cdots, \bar{Q}_T)$, where $\bar{Q}_t$ is defined as

$$\bar{Q}_t := \begin{pmatrix} p(\xi_1)I_{n_t\times n_t} & & \\ & \ddots & \\ & & p(\xi_E)I_{n_t\times n_t} \end{pmatrix}.$$

Proposition 3.4.1 *Given $x = ((x_t^\xi)_{\xi\in\Xi})_{t\in G} \in I\!R^{nE}$ and M defined as before. Then the resolvent value $z := (\partial\delta_{W^{-1}\mathcal{N}} + \bar{Q})^{-1}\bar{Q}(x)$ can be calculated as follows: for every $t \in G$ and every $A \in \mathcal{A}_t$,*

$$z_t^\xi = \frac{1}{\sum_{\xi'\in A} p(\xi')}\left(\sum_{\xi'\in A} p(\xi')x_t^\xi\right) \quad \textit{for all} \;\; \xi \in A.$$

Proof. Set $Y \in \mathcal{L}$ and $PY = \mathrm{Proj}_{\mathcal{N}}Y$. Then

$$\langle Y - PY, \eta\rangle_{\mathcal{L}} = 0 \;\; \text{for all} \;\; \eta \in \mathcal{N}$$

and hence

$$\langle \bar{Q}(W^{-1}Y - W^{-1}PY), W^{-1}\eta\rangle = 0 \;\; \text{for all} \;\; \eta \in \mathcal{N}.$$

Since $W^{-1}PY \in W^{-1}\mathcal{N}$, we get

$$(\partial\delta_{W^{-1}\mathcal{N}} + \bar{Q})^{-1}\bar{Q}(W^{-1}Y) = W^{-1}PY.$$

So, using the equivalent expression of PY given by Rockafellar and Wets [46], we deduce the result. ∎

We now apply algorithm SMS3BF considering for every $t \in G$, the parameter matrices $Q_t = \bar{Q}_t$ and $M_t = \mathrm{diag}\,(p(\xi_1)\bar{M}_t^{\xi_1}, \cdots, p(\xi_E)\bar{M}_t^{\xi_E})$, Then we obtain the following algorithm, called **"Time Scenarios Decomposition (TSD)"**

(TSD)

For every $t \in G$, $\xi \in \Xi$, set $\bar{M}_t^\xi$ a symmetric positive definite matrix. For every $\xi \in \Xi$, set also $\Sigma^\xi = (\sum_{t=1}^T (\bar{M}_t^\xi)^{-1})^{-1}$ and $\bar{p} = \sum_{\xi \in \Xi} m_\xi$. Then for an arbitrary $(x^0, y^0, z^0, v^0, u^0) \in I\!R^{nE} \times I\!R^{\bar{p}T} \times I\!R^{nE} \times I\!R^{\bar{p}} \times I\!R^{nE}$,

Step 1. find x^{k+1}

For all $t \in \{1, ..., T\}$ **and** $\xi \in \Xi$ **do**

find $(x^{k+1})_t^\xi$ such that

$$(x^{k+1})_t^\xi = \left(g_t(\cdot, \xi) + \lambda I_{n_t \times n_t} + (B_t^\xi)^t \bar{M}_t^\xi B_t^\xi \right)^{-1} (\eta)$$

where $\eta = (B_t^\xi)^t \bar{M}_t^\xi B_t^\xi (y^k)_t^\xi + (z^k)_t^\xi - p(\xi)^{-1} B_t^\xi (\bar{v}^k)^\xi - p(\xi)^{-1} (u^k)_t^\xi$

end for

Step 2. Calculate y^{k+1}

For all $t \in \{1, ..., T\}$ **and** $\xi \in \Xi$ **do**

$$(y^{k+1})_t^\xi = B_t^\xi (x^{k+1})_t^\xi - (\bar{M}_t^\xi)^{-1} \Sigma^\xi \left(\sum_{i=1}^T B_i^\xi (x^{k+1})_i^\xi \right).$$

end for

Step 3. Calculate z^{k+1}

For all $t \in \{1, ..., T\}$ **do**

set $A \in \mathcal{A}_t$

$$(z^{k+1})_t^\xi = \left(\sum_{\xi' \in A} p(\xi') [(x^{k+1})_t^{\xi'} + p(\xi)^{-1} (u^k)_t^{\xi'}] \right) / \sum_{\xi' \in A} p(\xi'), \quad \forall \xi \in A.$$

Step 4. Calculate $\bar{v}^{k+1}$

For all $\xi \in \Xi$ **do**

$$(\bar{v}^{k+1})^\xi = (\bar{v}^k)^\xi + p(\xi) \Sigma^\xi \left(\sum_{i=1}^T B_i^\xi (x^{k+1})_i^\xi \right).$$

end for

Step 5. Calculate u^{k+1}

For all $t \in \{1, ..., T\}$ **and** $\xi \in \Xi$ **do**

$$(u^{k+1})_t^\xi = (u^k)_t^\xi + p(\xi)[(x^{k+1})_t^\xi - (z^{k+1})_t^\xi]$$

end for

Chapter 4

A new splitting algorithm for inclusion problems mixing a composite monotone plus a co-coercive operator

In this chapter we consider the following composite monotone inclusion:

$$0 \in S(x) + A^t T(Ax) + C(x) \qquad (Var)$$

where $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $T : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ are maximal monotone maps, $C : \mathbb{R}^n \to \mathbb{R}^n$ is $\beta-$co-coercive with full domain and A an $m \times n$ matrix. Regarding this formulation, D. Davis and W. Yin's [16] analyze the particular case $A = I$ (thus $m = n$). They reformulate (Var) as a fixed-point equation with respect to an appropriate average map with similar properties as the Douglas-Rachford map considered for the sum of two maximal monotone maps ($C = 0$). In fact this new map recovers the Douglas-Rachford map when $C = 0$.

Monotone inclusion problems with three operators have gained a recent increase of interest in the community of splitting methods. It is motivated by many inverse problems in different fields like data analysis, image processing or machine learning when parcimony is a challenge for very large data sets. Primal-dual splitting methods were proposed in the literature as extensions of the classical splitting schemes for two operators, mainly the Douglas-Rachford family [31], which lead to decomposing the corresponding proximal steps for each operator separately (see [9, 12, 52, 13, 26, 16, 27, 6, 55, 30]).

The case $C = 0$, was studied in Chapter 2, where we have shown that the

Douglas-Rachford map is recovered from (2.53) considering the associated lagrangian map and a special matrix (see Remark 2.4.3). On the other hand, the case $C \neq 0$ and under mild assumption, we construct an average map with similar properties as the Davis-Yin map (recovering it when $A = I$) using the same definition (2.53), but considering a variant of the lagrangian map associated to (Var) and a special matrix. Then, we construct a generalized resolvent map defined in (2.6) deducing from it new splitting algorithms in order to solve problem (Var).

The structure of problem (Var) is related to the variational formulation of the minimization of separable convex functions:

$$\text{Minimize} \quad f(x) + g(Ax) + h(x) \tag{4.1}$$

where $f : I\!R^n \mapsto \bar{I\!R}$ and $g : I\!R^m \mapsto \bar{I\!R}$ are proper lower semi-continuous convex functions, $h : I\!R^n \mapsto I\!R$ is convex and $(\frac{1}{\beta})$-Lipschitz-differentiable, and A an $m \times n$ matrix.

Condat [13] considers problem (4.1) where under some regularity conditions it can be equivalently written as the following inclusion problem

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial f(x) + \nabla h(x) \\ \partial g^*(y) \end{pmatrix} + \begin{pmatrix} 0 & A^t \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{V}$$

To solve such problem, the author applies the Forward-Backward method, getting two algorithms described just below where, for simplicity, we fixed the relaxation parameter $\rho > 0$ and without error term, termed for us **CA1** and **CA2**:

Algorithm (**CA1**)

$$\begin{cases} \tilde{x}^{k+1} &= (\tau\partial f + I_{n\times n})^{-1}(x^k - \tau\nabla h(x^k) - \tau A^t y^k) \\ \tilde{y}^{k+1} &= (\sigma\partial g^* + I_{m\times m})^{-1}(y^k + \sigma A(2\tilde{x}^{k+1} - x^k)) \\ (x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k) \end{cases}$$

and

Algorithm (**CA2**)

$$\begin{cases} \tilde{y}^{k+1} &= (\sigma\partial g^* + I_{m\times m})^{-1}(y^k + \sigma A x^k) \\ \tilde{x}^{k+1} &= (\tau\partial f + I_{n\times n})^{-1}(x^k - \tau\nabla h(x^k) - \tau A^t(2\tilde{y}^{k+1} - y^k)) \\ (x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k) \end{cases}$$

The main difference between **CA1** and **CA2** is the parameter matrices chosen in the Forward–Backward method. Choosing special parameter matrices and Lagrangian maps in our general setting, we get different variants of algorithms $CA1$

and $CA2$. For example, a variant of $CA1$ is the recently algorithm **YA** proposed by O'connor [40], a slight variant of algorithm PD3O proposed by Yan [55], in order to solve model (4.1):

Algorithm (**YA**)

$$\begin{cases} x^{k+1} &= (\tau\partial f + I_{n\times n})^{-1}(x^k - \tau\nabla h(x^k) - \tau A^t y^k) \\ y^{k+1} &= (\sigma\partial g^* + I_{m\times m})^{-1}(y^k + \sigma A(2x^{k+1} - x^k + \tau\nabla h(x^k) - \tau\nabla h(x^{k+1}))) \end{cases}$$

Note that **CA1** and **YA** differ on their second update and on the choose of parameter ρ, considering $\rho = 1$ in the second algorithm. Through numerical experiment, Yan [55] noticed that **YA** has more advantages than **CA1** (considering $\rho = 1$).

Similarly, our variant algorithm of **CA2** has the same advantage like **YA** respect to **CA1**, ie. has large range of acceptable parameters ensuring convergence and better numerical result.

In Chapter 2 we have proposed some splitting algorithms for the following separable optimization problem

$$\min_{(x,z)} \quad [f(x) + g(z) : Ax + Bz = 0] \tag{P_0}$$

where f and g are convex and A and B two matrices of appropriated dimensions. In practice (see for instance [55]) f and g have the form $(k + h)$ where h is convex differentiable and k not necessarily differentiable but with a tractable proximal step. So, we need to propose an appropriate algorithm such that instead of finding the value of $(\partial k + \nabla h + Q)^{-1}$ at a given point, where Q is a symmetric positive definite matrix, the algorithm must use the proximal step of ∂k and the evaluation of ∇h, separately, in oder to obtain a splitting structure.

So, we will assume that problem (P_0) has the following form:

$$\min_{(x,z)} \quad [f(x) + h_1(x) + g(z) + h_2(z) : Ax + Bz = 0]. \tag{P}$$

where f and g are convex lsc functions, h_i $(i = 1, 2)$ is convex and $(\frac{1}{\beta_i})$-Lipschitz-differentiable, and A and B two matrices of order $m \times n$ and $m \times p$, respectively. It is clear that this problem includes problem (4.1) by considering $B = -I_{p\times p}$ and $h_2 = 0$.

In Section 4.1, we analyze the case A injective $(m \geq n)$, constructing an average map and a related splitting algorithm for solving (Var). Then, in order to obtain another algorithm switching the proximal steps of S and T with respect to algorithm (4.4)-(4.7) found in this section, we construct an appropriated average map, getting also ergodic and nonergodic convergence.

The general case (A not necessarily injective) is analyzed in Section 4.2 by reformulating (Var) as a problem preserving his original structure but corresponding to an injective matrix. Then, applying the results of Section 4.1, we get two general algorithms in this setting.

So, taking special parameter matrices intervening in the two previous general algorithms, we get in Section 4.3 two new algorithms (**Alg1**) and (**Alg2**) closely related with Condat's splitting algorithms (**CA1**) and (**CA2**). We show that (**Alg1**) and (**Alg2**) can also be obtained from Davis-Yin algorithm [16]. The rate of convergence of the two new algorithms is also analyzed.

In Section 4.4, we study problem (P) by reformulating it in order to apply the same procedure of Section 4.1 corresponding to injective matrix.

In Section 4.5, we apply the algorithm developed in Section 4.2 to a general multi-block convex optimization problem.

Finally, in Section 4.6 a numerical example is given.

4.1 Matrix A injective

Problem (Var) can be set as a composite monotone inclusions by considering the sum of S and C as a unique map, which, similarly to the one given in [41], is equivalent to the primal problem

$$\text{Find } (x,z) \in I\!R^n \times I\!R^m \text{ such that } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} S(x)+C(x) \\ T(z) \end{pmatrix} + \begin{pmatrix} A^t \\ -I \end{pmatrix} \mathcal{N}_{\{0\}}(Ax-z)$$

where, as usual, $\mathcal{N}_K(a)$ is the normal cone to set K at point a.

This problem is in turn equivalent to the following saddle-point inclusion problem

$$\text{Find } (\bar{x},\bar{z},\bar{y}) \in (I\!R^n \times I\!R^m) \times I\!R^m \text{ such that } \ 0 \in L(\bar{x},\bar{z},\bar{y}) \qquad (V_L)$$

where L is the map defined on $(I\!R^n \times I\!R^m) \times I\!R^m$ as

$$L(x,z,y) := \begin{pmatrix} S(x)+C(x) \\ T(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & -I \\ -A & I & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}.$$

From [41] we know that the lagrangian map L allows us to find splitting algorithms alternating between general proximal steps applied to $S+C$ and T, separately. In order to also split the general proximal map of $S+C$ into the general proximal map of S and the evaluation of C, we consider an alternative Lagrangian map which considers S and C defined with different variables.

Take $\overline{M}$ an arbitrary $m \times m$ positive definite matrix. Since A is of full-rank, then $A^t\overline{M}A$ is invertible and hence, from he third row-block of expression L in (V_L), we get $\bar{x} = (A^t\overline{M}A)^{-1}A^t\overline{M}\bar{z}$. One deduces that the solution set of problem (V_L), denoted by $\mathrm{sol}\,(V_L)$, coincides with the solution set of the following inclusion problem

$$\text{Find } (\bar{x},\bar{z},\bar{y}) \in (I\!R^n \times I\!R^m) \times I\!R^m \text{ such that } \ 0 \in \widehat{L}(\bar{x},\bar{z},\bar{y}) \qquad (V_{\widehat{L}})$$

where $\widehat{L}$ is the map defined on $(I\!R^n \times I\!R^m) \times I\!R^m$ as

$$\widehat{L}(x,z,y) := \begin{pmatrix} S(x) + C((A^t\overline{M}A)^{-1}A^t\overline{M}z) \\ T(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & -I \\ -A & I & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}.$$

The maps S and T being defined on different variables, we will get an additional separation of these maps by introducing an appropriate regularization map similarly to the one given in Section 4 of [41].

4.1.1 The average map $G^{\widehat{L}}_{\widehat{S}}$: an appropriate regularization map

First recall that the generalized resolvent operator associated with a maximal monotone map T and a positive semi-definite linear map P is defined by $J^T_P := (T+P)^{-1}P$ which is not necessarily defined on the whole space. Similarly, the corresponding map $G^T_S := S(T+P)^{-1}S^t$ where S satisfies $P = S^tS$, is not necessarily defined on the whole space but has contractive properties under additional assumptions (see Section 2.4 of Chapter 2).

Back to the Lagrangian map $\widehat{L}$ defined above, we take a symmetric positive definite matrix $\overline{M}$ of order $m \times m$ and the matrices

$$\widehat{P} = \begin{pmatrix} A^t\overline{M}A & 0 & A^t \\ 0 & 0 & 0 \\ A & 0 & \overline{M}^{-1} \end{pmatrix} \quad \text{and} \quad \widehat{D} = \begin{pmatrix} \overline{M}^{\frac{1}{2}}A & 0 & \overline{M}^{-\frac{1}{2}} \end{pmatrix},$$

which satisfy $\widehat{P} = \widehat{D}^t\widehat{D}$. We define the generalized resolvent map

$$J^{\widehat{L}}_{\widehat{P}} := (\widehat{L}+\widehat{P})^{-1}\widehat{P}$$

and then the map

$$G^{\widehat{L}}_{\widehat{D}} := \widehat{D}(\widehat{L} + \widehat{D}^t\widehat{D})^{-1}\widehat{D}^t.$$

It is clear that the set of fixed points of $J^{\widehat{L}}_{\widehat{P}}$ and $G^{\widehat{L}}_{\widehat{D}}$ are respectively $\mathrm{sol}\,(V_L)$ and

$$\widehat{D}(\mathrm{sol}\,(V_L)) := \{\overline{M}^{\frac{1}{2}}A\bar{x} + \overline{M}^{-\frac{1}{2}}\bar{y} : -A^t\bar{y} \in S(\bar{x}) + C(\bar{x}),\ \bar{y} \in T(A\bar{x})\}.$$

Also, by simple calculations we get

$$G_{\widehat{D}}^{\widehat{L}} = I - \tilde{J}_{\overline{M}}^{T} + \tilde{J}_{A^t\overline{M}A}^{S}[2\tilde{J}_{\overline{M}}^{T} - I - \widetilde{C} \circ \tilde{J}_{\overline{M}}^{T}] \tag{4.2}$$

where

$$\begin{aligned} \tilde{J}_{\overline{M}}^{T} &= \overline{M}^{\frac{1}{2}}(T+\overline{M})^{-1}\overline{M}^{\frac{1}{2}}, \\ \tilde{J}_{A^t\overline{M}A}^{S} &= \overline{M}^{\frac{1}{2}}A(S+A^t\overline{M}A)^{-1}A^t\overline{M}^{\frac{1}{2}}, \text{ and} \\ \widetilde{C} &= \overline{M}^{\frac{1}{2}}A(A^t\overline{M}A)^{-1}C(A^t\overline{M}A)^{-1}A^t\overline{M}^{\frac{1}{2}}. \end{aligned}$$

Notice that $C = 0$ implies $\widehat{L}$ monotone and thereby the co-coercivity of $G_{\widehat{D}}^{\widehat{L}}$ as shown in expression just before of Proposition 2.4.1. If $C \neq 0$, the map $G_{\widehat{D}}^{\widehat{L}}$ is still co-coercive by making an additional condition on matrix $\overline{M}$ as shows the following proposition :

Proposition 4.1.1 *Let $S : I\!R^n \longrightarrow I\!R^n$ and $T : I\!R^m \longrightarrow I\!R^m$ be maximal monotone maps, $C : I\!R^n \to I\!R^n$ a $\beta-$co-coercive function with full domain, and A an $m \times n$ injective matrix. Assume that $\|(A^t\overline{M}A)^{-1}\| \in]0, 2\beta[$. Then, $G_{\widehat{D}}^{\widehat{L}}$ is $\alpha-$average with full domain, where $\alpha := \frac{2\beta}{4\beta - \|(A^t\overline{M}A)^{-1}\|} < 1$.*

Proof. The fullness of the domain of $G_{\widehat{D}}^{\widehat{L}}$ is deduced from the maximality of S and T and the fullness of the domain of C.

From expression (4.2) the $\alpha-$average of map $G_{\widehat{D}}^{\widehat{L}}$ is deduced from Proposition 2.1 of Davis-Yin [16] taking into account that $\tilde{J}_{\overline{M}}^{T}$ and $\tilde{J}_{A^t\overline{M}A}^{S}$ are both $1-$co-coercive, and $\frac{1}{\|(A^t\overline{M}A)^{-1}\|}\widetilde{C}$ is $\beta-$co-coercive. ■

We note that the particular case $A = I$ and $\overline{M} = \gamma I$ in $\overline{M}^{-\frac{1}{2}} G_{\widehat{D}}^{\widehat{L}} \overline{M}^{\frac{1}{2}}$ correspond to the average map associated to the sum of three maps defined in [16].

Remark 4.1.1 *Using Proposition 2.4.1 the map $G_{\widehat{D}}^{\widehat{L}}$ can also be deduced from the average map associated to the sum of three maps defined in [16] applying to the following equivalent problem of (Var)*

$$0 \in (M^{\frac{1}{2}}AS^{-1}A^tM^{\frac{1}{2}})^{-1}(y) + M^{-\frac{1}{2}}TM^{-\frac{1}{2}}(y) + M^{\frac{1}{2}}A(A^tMA)^{-1}C(A^tMA)^{-1}A^tM^{\frac{1}{2}}(y).$$

This equivalent problem is deducted as follows, first we note that problem (Var) is equal to the composite problem

$$0 \in S(x) + A^tM^{\frac{1}{2}}\left[M^{-\frac{1}{2}}TM^{-\frac{1}{2}} + M^{\frac{1}{2}}A(A^tMA)^{-1}C(A^tMA)^{-1}A^tM^{\frac{1}{2}}\right]M^{\frac{1}{2}}A(x)$$

then the dual problem is equal to

$$0 \in -M^{\frac{1}{2}}AS^{-1}(-A^tM^{\frac{1}{2}}y) + \left[M^{-\frac{1}{2}}TM^{-\frac{1}{2}} + M^{\frac{1}{2}}A(A^tMA)^{-1}C(A^tMA)^{-1}A^tM^{\frac{1}{2}}\right]^{-1}(y)$$

finally taking the dual again considering the last problem as sum of two map problem, we obtained the desirable equivalent problem.

4.1.2 Constructing the splitting algorithm

In connection with the resolvent operator $J_{\widehat{P}}^{\widehat{L}}$ and a real positive parameter ρ, we consider for an arbitrary point $w^0 \in \operatorname{dom} J_{\widehat{P}}^{\widehat{L}}$, the sequence $\{w^k\}$ defined by

$$w^{k+1} \in \rho J_{\widehat{P}}^{\widehat{L}}(w^k) + (1-\rho)w^k. \tag{4.3}$$

Denoting $w^k := (x^k, z^k, y^k)$ and $\tilde{w}^{k+1} := (\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) = J_{\widehat{P}}^{\widehat{L}} w^k$, we get from (4.3) that

$$\begin{aligned}
\tilde{z}^{k+1} &= (T + \overline{M})^{-1}(y^k + \overline{M}Ax^k) && (4.4)\\
\tilde{y}^{k+1} &= y^k + \overline{M}Ax^k - \overline{M}\tilde{z}^{k+1} && (4.5)\\
r^{k+1} &= C((A^t\overline{M}A)^{-1}A^t\overline{M}\tilde{z}^{k+1}) && (4.6)\\
\tilde{x}^{k+1} &= (S + A^t\overline{M}A)^{-1}\left(A^t\overline{M}\tilde{z}^{k+1} - A^t\tilde{y}^{k+1} - r^{k+1}\right) && (4.7)
\end{aligned}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, z^k, y^k) \tag{4.8}$$

and

$$\widehat{D}w^{k+1} = \rho G_{\widehat{D}}^{\widehat{L}}(\widehat{D}w^k) + (1-\rho)\widehat{D}w^k. \tag{4.9}$$

It is worth to mention that $\{\widehat{D}w^k\}$ is the sequence generated by the fixed point method corresponding to operator $\rho G_{\widehat{D}}^{\widehat{L}} + (1-\rho)I$ which is $\rho\alpha-$average if $\rho \in]0, \alpha^{-1}[$ because $G_{\widehat{D}}^{\widehat{L}}$ is $\alpha-$average (see Proposition 4.1.1), and hence the sequence $\{\widehat{D}w^k\}$ converges if $\operatorname{sol}(Var)$ is nonempty.

The next proposition concerns the convergence of w^k; its proof is similar to the given in Section 2.4 of Chapter 2.

Proposition 4.1.2 *With the same hypothesis given in Proposition 4.1.1, let* $\rho \in]0, \alpha^{-1}[$ *and assume that* $\operatorname{sol}(Var)$ *is nonempty. Then for an arbitrary* $(x^0, z^0, y^0) \in (\mathbb{R}^n \times \mathbb{R}^p) \times \mathbb{R}^m$, *the sequence* (x^k, z^k, y^k) *defined by the sequential update formulae (4.4)− (4.8), converges to some element of* $\operatorname{sol}(V_L)$.

Proof. From the comments given just before this proposition, $\{\widehat{D}w^k\}$ converges to some $b \in \operatorname{sol}(V_L)$, which is a fixed point to $G_{\widehat{D}}^{\widehat{L}}$.

Now, from (4.3) and considering $\widetilde{w} := (\widehat{L} + \widehat{P})^{-1}\widehat{D}^t b$, we get by using the triangular inequality

$$\|w^{k+1} - \widetilde{w}\| \leq \rho\|(\widehat{L} + \widehat{P})^{-1}\widehat{D}^t(\widehat{D}w^k) - \widetilde{w}\| + |1-\rho|\|w^k - \widetilde{w}\|.$$

On the other hand, the map $J_{\widehat{P}}^{\widehat{L}}$ being continuous, $(\widehat{L}+\widehat{P})^{-1}\widehat{D}^t = J_P^T\widehat{D}^+$ is also continuous, where $\widehat{D}^+$ denotes the Moore–Penrose pseudo-inverse matrix of $\widehat{D}$, and hence the convergence of $\{w^k\}$ to $\tilde{w}$ is deduced. ■

Remark 4.1.2 *We note that the sequence z^k in (4.4)-(4.8) is only used in the final step of the algorithm, so we can discard it and consider the final step as*

$$(x^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k).$$

Applying (4.4)-(4.8) to the optimization problem (4.1), we get the following sequence:

$$\tilde{z}^{k+1} = \operatorname{argmin}\left\{g(z) + \frac{1}{2}\|z - \overline{M}^{-1}y^k - Ax^k\|_{\overline{M}}^2\right\} \tag{4.10}$$

$$\tilde{y}^{k+1} = y^k + \overline{M}Ax^k - \overline{M}\tilde{z}^{k+1} \tag{4.11}$$

$$\tilde{r}^{k+1} = A(A^t\overline{M}A)^{-1}\nabla h((A^t\overline{M}A)^{-1}A^t\overline{M}\tilde{z}^{k+1}) \tag{4.12}$$

$$\tilde{x}^{k+1} = \operatorname{argmin}\left\{f(x) + \frac{1}{2}\|Ax - \tilde{z}^{k+1} + \overline{M}^{-1}\tilde{y}^{k+1} + \tilde{r}^{k+1}\|_{\overline{M}}^2\right\} \tag{4.13}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, z^k, y^k). \tag{4.14}$$

4.1.3 Switching the proximal step

Applying the forward-backward method to a lagrangian inclusion problem (V), Condat [13] obtains two splitting algorithms $CA1$ and $CA2$ corresponding to two appropriated parameter matrices. The main difference between these algorithms is the order of action of the proximal steps.

In the same manner, we present an algorithm switching the order of action of the proximal steps regarding algorithm (4.4)–(4.8). For this purpose, it is not only necessary to find an appropriate matrix but also consider another alternative Lagrangian map.

For a given $m \times m$ positive definite matrix $\overline{M}$ (similarly to the given at the beginning of this section), we consider the map $\bar{L}$ defined on $(\mathbb{R}^n \times \mathbb{R}^m) \times \mathbb{R}^m$ as

$$\bar{L}(x,z,y) := \begin{pmatrix} S(x) \\ T(z) + \overline{M}A(A^t\overline{M}A)^{-1}C(x) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & -I \\ -A & I & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}$$

whose set of their zeroes is exactly

$$\{(x^*, \overline{M}A(A^t\overline{M}A)^{-1}C(x^*) + y^*, z^*) : (x^*, y^*, z^*) \in \operatorname{sol}(V_L)\}.$$

Note that when $C \equiv 0$, this new Lagrangian coincides with Lagrangian $\widehat{L}$ corresponding to problem $(V_{\widehat{L}})$ defined at the beginning of this section.

Analogously to the one given in Subsection 4.1.1, we define $J_{\overline{P}}^{\overline{L}}$ and $G_{\overline{D}}^{\overline{L}}$ corresponding to matrices

$$\overline{P} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \overline{M} & -I \\ 0 & -I & \overline{M}^{-1} \end{pmatrix} \quad \text{and} \quad \overline{D} = \begin{pmatrix} 0 & \overline{M}^{\frac{1}{2}} & -\overline{M}^{-\frac{1}{2}} \end{pmatrix},$$

which satisfies $\overline{P} = \overline{D}^t\overline{D}$. It follows that

$$G_{\overline{D}}^{\overline{L}} = I - \tilde{J}^S_{A^t\overline{M}A} + \tilde{J}^T_{\overline{M}}[2\tilde{J}^S_{A^t\overline{M}A} - I - \tilde{C} \circ \tilde{J}^S_{A^t\overline{M}A}] \tag{4.15}$$

where $\tilde{J}^T_{\overline{M}}$, $\tilde{J}^S_{A^t\overline{M}A}$ and $\tilde{C}$ are defined just after of expression (4.2).

We note that the main difference between $G_{\overline{D}}^{\overline{L}}$ and $G_{\widehat{D}}^{\widehat{L}}$ (the last one defined in (4.2)) is the switching position of $\tilde{J}^T_{\overline{M}}$ and $\tilde{J}^S_{A^t\overline{M}A}$ in their expressions. So, under the same conditions giving in Proposition 4.1.1, we get that $G_{\overline{D}}^{\overline{L}}$ is also $\alpha-$average.

The fixed point iteration method applied to $J_{\overline{P}}^{\overline{L}}$ generates the following sequences

$$\tilde{x}^{k+1} = (S + A^t\overline{M}A)^{-1}(A^t\overline{M}z^k - A^t y^k) \tag{4.16}$$

$$\tilde{y}^{k+1} = y^k + \overline{M}A\tilde{x}^{k+1} - \overline{M}z^k \tag{4.17}$$

$$r^{k+1} = MA(A^t\overline{M}A)^{-1}C(\tilde{x}^{k+1}) \tag{4.18}$$

$$\tilde{z}^{k+1} = (T + \overline{M})^{-1}\left(\overline{M}A\tilde{x}^{k+1} + \tilde{y}^{k+1} - r^{k+1}\right) \tag{4.19}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, z^k, y^k). \tag{4.20}$$

We note that the main difference between this algorithm and the one of (4.4)-(4.8) is the order position of the (general) proximal step of S and T.

Remark 4.1.3 *Briceño [8] has analyzed the particular case $A = I$ and $T = \mathcal{N}_V$ where V is a linear sub-space of a Hilbert space H. He proposed two alternative methods where the first one was obtained through the composition of two special average maps, and the other one through the forward-backward method applied to the sum problem corresponding to the partial inverse of map S with respect to V and a special co-coercive map. The considered model is*

$$\textit{Find } x \in H \textit{ such that } \; 0 \in \sum_{i=1}^{m} \mathcal{S}_i(x) + C(x) \tag{S_v}$$

where for $i \in \{1, \cdots, m\}$, $\mathcal{S}_i$ is maximal monotone and C co-coercive, all defined on H. The aforementioned algorithms were applied to an appropriate reformulation of this model. Considering $H = \mathbb{R}^n$, an alternative reformulation of problem (S_v) is

$$\text{Find } x \in \mathbb{R}^n \text{ such that } 0 \in \mathcal{S}_1(x) + D^t\mathcal{T}(Dx) + C(x) \qquad (RS_v)$$

where $\mathcal{T} := (\mathcal{S}_2, \cdots, \mathcal{S}_m)$ and D is an $n(m-1) \times n$ matrix defined by $D = (I_{n\times n} \cdots I_{n\times n})^t$. Notice that problem (RS_v) has the same structure as model (Var) considered at the beginning of this chapter, when D is injective. So we can apply algorithm (4.4)-(4.8) or his switched version (4.16)-(4.20), getting in both cases splitting algorithms by considering $\overline{M} = \lambda I$. These splitting algorithms combine proximal steps on each $\mathcal{S}_i$ with the forward step on C, because $D^tD = (m-1)I_{n\times n}$.

4.1.4 Rate of Convergence

This part is dealing with the rate of convergence of algorithm $(4.10) - (4.14)$. In this direction, the next proposition gives an upper bound estimation of the saddle-point gap of optimization problem (4.1) defined in the introduction of this chapter.

Proposition 4.1.3 *With the same notations as before and considering $w = (x, z, y) \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g) \times \mathbb{R}^m$, the following inequality holds:*

$$\|w^k - w\|^2_{\widehat{P}} - \gamma\|w^{k+1} - w^k\|^2_{\widehat{P}} - \|w^{k+1} - w\|^2_{\widehat{P}} \geq 2\rho\left[l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l(x, z, \tilde{y}^{k+1})\right]$$

where $\gamma = \frac{1}{\rho}\left[2 - \rho - \frac{\|(A^t\overline{M}A)^{-1}\|}{\beta}\right]$ and

$$l(x, z, y) = f(x) + h(x) + g(z) + \langle y, Ax + Bz\rangle.$$

Proof. Since $\tilde{w}^{k+1} = J^{\widehat{L}}_{\widehat{P}} w^k$, one has $\widehat{L}(\tilde{w}^{k+1}) \in \widehat{P}(w^k - \tilde{w}^{k+1})$ and hence

$$L'(\tilde{w}^{k+1}) \in \widehat{P}(w^k - \tilde{w}^{k+1}) - \begin{pmatrix} \nabla h(\widehat{z}^{k+1}) \\ 0 \\ 0 \end{pmatrix}$$

where L' is L without the term C, and $\widehat{z}^{k+1} = (A^t\overline{M}A)^{-1}A^t\overline{M}\tilde{z}^{k+1}$. Note that $L' = (\partial_{x,z}l') \times (\partial_y[-l'])$, where

$$l'(x, z, y) = f(x) + g(z) + \langle y, Ax - z\rangle.$$

From Prop. 3 given in [41] and denoting $w := (x, z, y) \in \mathrm{dom}\,(l)$, we get

$$\left\langle \tilde{w}^{k+1} - w, \widehat{P}(w^k - \tilde{w}^{k+1})\right\rangle - \left\langle \tilde{x}^{k+1} - x, \nabla h(\widehat{z}^{k+1})\right\rangle \geq l'(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l'(x, z, \tilde{y}^{k+1}). \tag{4.21}$$

On the other hand, since h is convex and $\beta^{-1}-$Lipschitz differentiable, we have

$$\frac{1}{2\beta}\|\widehat{z}^{k+1}-\tilde{x}^{k+1}\|^2+\langle\nabla h(\widehat{z}^{k+1}),\tilde{x}^{k+1}-\widehat{z}^{k+1}\rangle\geq h(\tilde{x}^{k+1})-h(\widehat{z}^{k+1})$$

and

$$h(x)\geq h(\widehat{z}^{k+1})+\langle\nabla h(\widehat{z}^{k+1}),x-\widehat{z}^{k+1}\rangle.$$

Then, summing the three last inequalities we get

$$\left\langle\tilde{w}^{k+1}-w,\widehat{P}(w^k-\tilde{w}^{k+1})\right\rangle+\frac{1}{2\beta}\|\widehat{z}^{k+1}-\tilde{x}^{k+1}\|^2\geq l(\tilde{x}^{k+1},\tilde{z}^{k+1},y)-l(x,z,\tilde{y}^{k+1}). \tag{4.22}$$

We now find an appropriate upper bound for $\frac{1}{2\beta}\|\widehat{z}^{k+1}-\tilde{x}^{k+1}\|^2$. From the definition of $G_{\widehat{D}}^{\widehat{L}}$ we have

$$\widehat{D}\tilde{w}^{k+1}=G_{\widehat{D}}^{\widehat{L}}(\widehat{D}w^k)=\widehat{D}w^k-\overline{M}^{\frac{1}{2}}\tilde{z}^{k+1}+\overline{M}^{\frac{1}{2}}A\tilde{x}^{k+1}$$

and hence by using $w^{k+1}=\rho\tilde{w}^{k+1}+(1-\rho)w^k$ we get

$$\frac{1}{\rho}\widehat{D}(w^{k+1}-w^k)=\widehat{D}(\tilde{w}^{k+1}-w^k)=\overline{M}^{\frac{1}{2}}A\tilde{x}^{k+1}-\overline{M}^{\frac{1}{2}}\tilde{z}^{k+1}. \tag{4.23}$$

Since

$$\widehat{z}^{k+1}-\tilde{x}^{k+1}=(A^t\overline{M}A)^{-1}A^t\overline{M}^{\frac{1}{2}}[\overline{M}^{\frac{1}{2}}\tilde{z}^{k+1}-\overline{M}^{\frac{1}{2}}A\tilde{x}^{k+1}]$$

we get, from (4.23), the desirable appropriate upper bound

$$\frac{1}{2\beta}\|\widehat{z}^{k+1}-\tilde{x}^{k+1}\|^2\leq\frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta}\|\frac{1}{\rho}\widehat{D}(w^{k+1}-w^k)\|^2=\frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta\rho^2}\|w^{k+1}-w^k\|_{\widehat{P}}^2.$$

On the other hand, from the symmetry of matrix $\widehat{P}$, it holds

$$2\rho\left\langle\tilde{w}^{k+1}-w,\widehat{P}(w^k-\tilde{w}^{k+1})\right\rangle=\|w^k-w\|_{\widehat{P}}^2-\frac{2-\rho}{\rho}\|w^{k+1}-w^k\|_{\widehat{P}}^2-\|w^{k+1}-w\|_{\widehat{P}}^2.$$

So, replacing the two last expressions into (4.49), we get the desired inequality. ■

Upper bound of fixed-point residual

Set $\overline{M}$ and ρ satisfying the hypothesis of Proposition 4.1.2, i.e,

$$\alpha^{-1}=2-\frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta}>\rho>0.$$

With this condition $\rho G_{\widehat{D}}^{\widehat{L}}+(1-\rho)I$ is $\rho\alpha-$average and hence from (4.9), we have

$$\|\widehat{D}w^k-\widehat{D}w^*\|^2-\theta\|\widehat{D}w^{k+1}-\widehat{D}w^k\|^2-\|\widehat{D}w^{k+1}-\widehat{D}w^*\|^2\geq 0 \tag{4.24}$$

where $\theta = \frac{1}{\rho}\left[2 - \rho - \frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta}\right] > 0$ and $w^* \in \mathrm{sol}\,(V_L)$.

Then using similar argument given in Section 2.5.1 of Chapter 2, we get that

$$\|\widehat{D}w^k - \widehat{D}w^{k-1}\|^2 \leq \frac{1}{k\theta}\|\widehat{D}w^0 - \widehat{D}w^*\|^2 \tag{4.25}$$

and

$$\left\|\frac{1}{N}\sum_{k=1}^{N}(\widehat{D}w^k - \widehat{D}w^{k-1})\right\|^2 \leq \frac{4}{N^2}\|\widehat{D}w^0 - \widehat{D}w^*\|^2. \tag{4.26}$$

These two relations can also be deduced respectively from Theorem 1 "Notes on Theorem 1" and Theorem 2 given in [15].

Bounding the saddle-point gap

We consider the following **ergodic sequences** defined for $N \geq 1$ as

$$\overline{x}_N := \frac{1}{N}\sum_{k=1}^{N}\tilde{x}^k, \quad \overline{z}_N := \frac{1}{N}\sum_{k=1}^{N}\tilde{z}^k \quad \text{and} \quad \overline{y}_N := \frac{1}{N}\sum_{k=1}^{N}\tilde{y}^k.$$

We get the following result

Theorem 4.1.1 *With the same notations as before. Set $\overline{M}$ and ρ satisfying*

$$2 - \frac{\|(A^t\overline{M}A)^{-1}\|}{\beta} \geq \rho > 0.$$

The following rate of convergence are deduced:

- **Ergodic Convergence:** *for any $w = (x, z, y) \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g) \times I\!R^m$,*

$$l(\overline{x}_k, \overline{z}_k, y) - l(x, z, \overline{y}_k) \leq \frac{1}{2\rho k}\|\widehat{D}w^0 - \widehat{D}w\|^2. \tag{4.27}$$

- **Nonergodic Convergence:** *for any $w^* = (x^*, z^*, y^*) \in \mathrm{sol}\,(V_L)$,*

$$l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y^*) - l(x^*, z^*, \tilde{y}^{k+1}) \leq \left(\frac{\alpha_1}{\sqrt{k+1}} + \frac{\alpha_2}{k+1}\right)\|\widehat{D}w^0 - \widehat{D}w^*\|^2 \tag{4.28}$$

where

$$\alpha_1 = \frac{1 + |1 - \rho|}{\rho^2\sqrt{\theta}} \quad \text{and} \quad \alpha_2 = \frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta\rho^2\theta}.$$

Proof. From Proposition 4.1.3 we get

$$\|w^k - w\|^2_{\widehat{P}} - \|w^{k+1} - w\|^2_{\widehat{P}} \geq 2\rho \left[l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y) - l(x, z, \tilde{y}^{k+1})\right]$$

since $2 - \frac{\|(A^t\overline{M}A)^{-1}\|}{\beta} \geq \rho > 0$. Summing over $k = 0, \cdots, N-1$, and applying the Jensen's inequality to the convex functions $l(\cdot, \cdot, y) - l(x, z, \cdot)$ for arbitrary fixed element $(x, z, y) \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g) \times I\!R^m$, where l is the Lagrangian function defined in Proposition 4.1.3, the desired ergodic convergence is deduced.

Regarding the nonergodic convergence, for $w^* \in \mathrm{sol}\,(V_L)$ and considering $w = w^*$ in (4.3.1), we get

$$\left\langle \tilde{w}^{k+1} - w^*, \widehat{P}(w^k - \tilde{w}^{k+1}) \right\rangle + \frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta\rho^2}\|w^{k+1} - w^k\|^2_{\widehat{P}} \geq l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y^*) - l(x^*, z^*, \tilde{y}^{k+1})$$

and hence, from the Cauchy-Schwarz inequality and (4.9), we obtain

$$\frac{1}{\rho}\|\tilde{w}^{k+1} - w^*\|_{\widehat{P}}\|w^{k+1} - w^k\|_{\widehat{P}} + \frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta\rho^2}\|w^{k+1} - w^k\|^2_{\widehat{P}} \geq l(\tilde{x}^{k+1}, \tilde{z}^{k+1}, y^*) - l(x^*, z^*, \tilde{y}^{k+1}). \tag{4.29}$$

On other hand, from (4.9) and since $\{\|\widehat{D}w^{k+1} - \widehat{D}w^*\|\}$ is non increasing, we get

$$\|G^{\widehat{L}}_{\widehat{D}}\widehat{D}w^k - \widehat{D}w^*\| = \|\frac{1}{\rho}(\widehat{D}w^{k+1} - \widehat{D}w^*) + (1 - \frac{1}{\rho})(\widehat{D}w^k - \widehat{D}w^*)\| \leq \frac{1 + |1 - \rho|}{\rho}\|\widehat{D}w^0 - \widehat{D}w^*\|.$$

So, replacing this last expression and inequality (4.25) in expression (4.29), we deduce the desired nonergodic convergence. ∎

Constraint violations

Following the same arguments given in the proof of Theorem 2.5.2 of Chapter 2, we get the following result.

Theorem 4.1.2 *With the same notations as before. Set $\overline{M}$ and ρ satisfying*

$$2 - \frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta} > \rho > 0.$$

For any $w^* \in \mathrm{sol}\,(V_L)$, the following rate of convergence are obtained:

- **Ergodic Convergence:**

$$\|A\overline{x}_k - \overline{z}_k\|_{\overline{M}} \leq \frac{2}{k\rho}\|\widehat{D}w^0 - \widehat{D}w^*\|.$$

- **Nonergodic Convergence:**

$$\|A\tilde{x}^k - \tilde{z}^k\|_{\overline{M}} \leq \frac{1}{\rho\sqrt{k\theta}}\|\widehat{D}w^0 - \widehat{D}w^*\|.$$

Linear convergence

In the case when $A = I$ (sum of three maps problem), D. Davis and Yin [16] got a linear convergence under additional regularity condition (co-coercive, strong montone or lipschitz properties) over the monotone maps. In the case when matrix A is injective, since the map $G_{\widehat{D}}^{\widehat{L}}$ can be obtained from Davis-Yin map associated to a sum of three maps problem (see remark 4.1.1), and noting that if S $\gamma-$strong monotone then $(M^{\frac{1}{2}}AS^{-1}A^tM^{\frac{1}{2}})^{-1}$ is $\frac{\gamma}{\|A^tMA\|}-$strong monotone and if T is $\theta-$co-coercive then $M^{-\frac{1}{2}}TM^{-\frac{1}{2}}$ is $\frac{\theta}{\|M\|}-$co-coercive. We deduce the linear convergence of algorithm (4.4)-(4.8) and (4.16)-(4.20), if we consider the additional hypothesis: S strong monotone and T is co-coercive.

4.2 The general case on matrix A

We now consider problem (Var) without assuming matrix A injective. In order to cope with this rank deficiency, we reformulate the problem as

$$0 \in S(x) + \begin{pmatrix} A^tM^{\frac{1}{2}} & V^{\frac{1}{2}} \end{pmatrix} \begin{bmatrix} M^{-\frac{1}{2}}TM^{-\frac{1}{2}} \\ 0 \end{bmatrix} \begin{pmatrix} M^{\frac{1}{2}}Ax \\ V^{\frac{1}{2}}x \end{pmatrix} + C(x) \qquad (Var_1)$$

where M and V are two symmetric matrices of order $m \times m$ and $n \times n$, respectively, with V positive semi-definite and M positive definite.

It is clear that matrix $\begin{pmatrix} M^{\frac{1}{2}}A \\ V^{\frac{1}{2}} \end{pmatrix}$ is injective if and only if $A^tMA + V$ is invertible. So, applying the algorithm described in Section 4.1 for matrix $\overline{M} = I$ we get a splitting algorithm for problem (Var) in the general setting.

It is important to note that formulation (Var_1) is motivated by the optimization problem defined in (4.1). Indeed, using the same notations given at the end of the first section of Chapter 1, problem (4.1) can be formulated as

$$\min_{(x,z_1,z_2)\in\mathcal{F}} f(x) + h(x) + (g,0)(z_1,z_2) \qquad (\bullet)$$

where $\mathcal{F}$ is the set of all triples (x, z_1, z_2) satisfying

$$\begin{pmatrix} M^{\frac{1}{2}}A \\ V^{\frac{1}{2}} \end{pmatrix} x + \begin{pmatrix} -M^{\frac{1}{2}} & 0 \\ 0 & -I \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 0.$$

It is clear that the optimal solution set of problem $(\bullet)$ consists of all $(x, Ax, V^{\frac{1}{2}}x)$, where x is an optimal solution of problem (4.1).

Notice that problem $(\bullet)$ has the same structure as problem (P_0). Indeed, by taking $f_1(x) = f(x) + h(x)$ and $f_2 = (g, 0)$ and the matrices

$$B_1 = \begin{pmatrix} M^{\frac{1}{2}}A \\ V^{\frac{1}{2}} \end{pmatrix} \quad \text{and} \quad B_2 = \begin{pmatrix} -M^{\frac{1}{2}} & 0 \\ 0 & -I \end{pmatrix},$$

the dual variational formulation of problem $(\bullet)$ consists in finding a zero of the sum of two composite maps consisting of

$$-\begin{pmatrix} M^{\frac{1}{2}}A \\ V^{\frac{1}{2}} \end{pmatrix} (\partial f + \nabla h)^{-1} \left(-\begin{pmatrix} A^t M^{\frac{1}{2}} & V^{\frac{1}{2}} \end{pmatrix} \right)$$

and

$$K := -\begin{pmatrix} -M^{\frac{1}{2}} & 0 \\ 0 & -I \end{pmatrix} \begin{bmatrix} \partial g \\ 0 \end{bmatrix}^{-1} \left(-\begin{pmatrix} -M^{\frac{1}{2}} & 0 \\ 0 & -I \end{pmatrix} \right).$$

Then the dual problem of this sum problem is

$$0 \in \partial f(x) + \nabla h(x) + \begin{pmatrix} A^t M^{\frac{1}{2}} & V^{\frac{1}{2}} \end{pmatrix} K^{-1} \begin{pmatrix} M^{\frac{1}{2}}Ax \\ V^{\frac{1}{2}}x \end{pmatrix}.$$

Since K^{-1} has the following expression

$$K^{-1} = \begin{pmatrix} M^{-\frac{1}{2}} & 0 \\ 0 & I \end{pmatrix} \begin{bmatrix} \partial g \\ 0 \end{bmatrix} \begin{pmatrix} M^{-\frac{1}{2}} & 0 \\ 0 & I \end{pmatrix} = \begin{bmatrix} M^{-\frac{1}{2}} \partial g M^{-\frac{1}{2}} \\ 0 \end{bmatrix},$$

then the last inclusion problem can be set as

$$0 \in \partial f(x) + \begin{pmatrix} A^t M^{\frac{1}{2}} & V^{\frac{1}{2}} \end{pmatrix} \begin{bmatrix} M^{-\frac{1}{2}} \partial g(x) M^{-\frac{1}{2}} \\ 0 \end{bmatrix} \begin{pmatrix} M^{\frac{1}{2}}Ax \\ V^{\frac{1}{2}}x \end{pmatrix} + \nabla h(x).$$

4.2.1 The main algorithm for non injective operators

Like to map $\widehat{L}$ and matrix $\widehat{D}$ corresponding to problem (Var), we denote respectively by $\widehat{L}'$ and $\widehat{D}'$ the map and matrix corresponding to problem (Var_1). It is clear that

$\widehat{L}'$ is defined on $(I\!R^n \times I\!R^m \times I\!R^n) \times (I\!R^m \times I\!R^n)$ and its value is

$$\widehat{L}'(x, z_1, z_2, y_1, y_2) = \begin{pmatrix} S(x) + C((A^tMA+V)^{-1}(A^tM^{\frac{1}{2}}z_1 + V^{\frac{1}{2}}z_2)) \\ M^{-\frac{1}{2}}TM^{-\frac{1}{2}}(z_1) \\ 0_{z_2} \\ 0_{y_1} \\ 0_{y_2} \end{pmatrix}$$

$$+ \begin{pmatrix} 0 & 0 & 0 & A^tM^{\frac{1}{2}} & V^{\frac{1}{2}} \\ 0 & 0 & 0 & -I & 0 \\ 0 & 0 & 0 & 0 & -I \\ -M^{\frac{1}{2}}A & I & 0 & 0 & 0 \\ -V^{\frac{1}{2}} & 0 & I & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ z_1 \\ z_2 \\ y_1 \\ y_2 \end{pmatrix}.$$

The set of zeroes of $\widehat{L}'$ is

$$\{(\bar{x}, M^{\frac{1}{2}}\bar{z}, V^{\frac{1}{2}}\bar{x}, M^{-\frac{1}{2}}\bar{y}, 0_{y_2}) : (0_x, 0_z, 0_y) \in L(\bar{x}, \bar{z}, \bar{y})\},$$

where 0_w denotes the zero vector of the $w-$space.

On the other hand, the corresponding map defined in (4.2) associated with $\widehat{L}'$ is denoted by $G^{\widehat{L}'}_{\widehat{D}'}$ which applies $I\!R^m \times I\!R^n$ into itself and whose value at (u, x) is

$$\begin{pmatrix} u - \tilde{J}^T u + M^{\frac{1}{2}}A\tilde{J}^S \left[V^{\frac{1}{2}}x + A^tM^{\frac{1}{2}}(2\tilde{J}^T u - u) - \tilde{C}(V^{\frac{1}{2}}x + A^tM^{\frac{1}{2}}\tilde{J}^T u)\right] \\ V^{\frac{1}{2}}\tilde{J}^S \left[V^{\frac{1}{2}}x + A^tM^{\frac{1}{2}}(2\tilde{J}^T u - u) - \tilde{C}(V^{\frac{1}{2}}x + A^tM^{\frac{1}{2}}\tilde{J}^T u)\right] \end{pmatrix}$$

where

$$\tilde{J}^T = M^{\frac{1}{2}}(T+M)^{-1}M^{\frac{1}{2}},\ \tilde{J}^S = (S + V + A^tMA)^{-1} \text{ and } \tilde{C} = C(A^tMA+V)^{-1}.$$

Set $r = \|(V + A^tMA)^{-1}\|$. From Proposition 4.1.1, $G^{\widehat{L}'}_{\widehat{D}'}$ is $\frac{2\beta}{4\beta - r}-$average with full domain, if $r \leq 2\beta$.

Applying the algorithm described in (4.4)-(4.8), we get that

$$\begin{aligned}
\tilde{z}_1^{k+1} &= M^{\frac{1}{2}}(T+M)^{-1}M^{\frac{1}{2}}(y_1^k + M^{\frac{1}{2}}Ax^k) \\
\tilde{z}_2^{k+1} &= y_2^k + V^{\frac{1}{2}}x^k \\
\tilde{y}_1^{k+1} &= y_1^k + M^{\frac{1}{2}}Ax^k - \tilde{z}_1^{k+1} \\
\tilde{y}_2^{k+1} &= y_2^k + V^{\frac{1}{2}}x^k - \tilde{z}_2^{k+1} \\
r^{k+1} &= C((A^tMA+V)^{-1}(A^tM^{\frac{1}{2}}\tilde{z}_1^{k+1} + V^{\frac{1}{2}}\tilde{z}_2^{k+1})) \\
\tilde{x}^{k+1} &= (S + A^tMA + V)^{-1}\left(A^tM^{\frac{1}{2}}\tilde{z}_1^{k+1} + V^{\frac{1}{2}}\tilde{z}_2^{k+1} - A^tM^{\frac{1}{2}}\tilde{y}_1^{k+1} - V^{\frac{1}{2}}\tilde{y}_2^{k+1} - r^{k+1}\right)
\end{aligned}$$

$$(x^{k+1}, y_1^{k+1}, y_2^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{y}_1^{k+1}, \tilde{y}_2^{k+1}) + (1-\rho)(x^k, y_1^k, y_2^k).$$

Notice that $\tilde{y}_2^{k+1} = 0$, and hence, by considering $y_2^0 = 0$, we get $y_2^k = 0$, which implies in particular $\tilde{z}_2^{k+1} = V^{\frac{1}{2}} x^k$. So, by denoting

$$\tilde{z}^{k+1} = M^{-\frac{1}{2}} \tilde{z}_1^{k+1},\ \tilde{y}^{k+1} = M^{\frac{1}{2}} \tilde{y}_1^{k+1} \ \text{ and } \ y^k = M^{\frac{1}{2}} y_1^k,$$

we deduce from the previous sequences our main algorithm termed **"Generalized splitting algorithm for three operators (GSA3O)"**:

(GSA3O)

$$\begin{aligned}
\tilde{z}^{k+1} &= (T+M)^{-1}(y^k + MAx^k) && (4.30)\\
\tilde{y}^{k+1} &= y^k + MAx^k - M\tilde{z}^{k+1} && (4.31)\\
r^{k+1} &= C((V + A^tMA)^{-1}(Vx^k + A^tM\tilde{z}^{k+1})) && (4.32)\\
\tilde{x}^{k+1} &= (S + V + A^tMA)^{-1}\left(Vx^k + A^tM\tilde{z}^{k+1} - A^t\tilde{y}^{k+1} - r^{k+1}\right) && (4.33)\\
(x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k). && (4.34)
\end{aligned}$$

We finally deduce the following proposition directly from Proposition 4.1.2.

Proposition 4.2.1 *Assume that $V \in \mathbb{R}^{n\times n}$ and $M \in \mathbb{R}^{m\times m}$ are symmetric, with V positive semi-definite and M positive definite, such that $V + A^tMA$ is positive definite and satisfying that $\|(V + A^tMA)^{-1}\| \in]0, 2\beta[$. Let $\rho \in]0, \alpha^{-1}[$ be where $\alpha := \frac{2\beta}{4\beta - \|(V+A^tMA)^{-1}\|}$. If $\mathrm{sol}\,(Var)$ is nonempty, then for an arbitrary $(x^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m$, the sequence (x^k, y^k) in (4.30)-(4.34) holds that (x^k, Ax^k, y^k) converges to some element of $\mathrm{sol}\,(V_L)$.*

Similarly to expression (4.9), the sequence $\{\zeta^k := (x^k, y^k)\}$ generated by algorithm (GSA3O), satisfies the following relation

$$\widehat{Q}\zeta^{k+1} = \rho G_{\widehat{D}'}^{\widehat{L}'}(\widehat{Q}\zeta^k) + (1-\rho)\widehat{Q}\zeta^k \tag{4.35}$$

where

$$\widehat{Q} = \begin{pmatrix} M^{\frac{1}{2}}A & M^{-\frac{1}{2}} \\ V^{\frac{1}{2}} & 0 \end{pmatrix}.$$

4.2.2 Switching the proximal step

In a similar way described for the injective context, we get now an algorithm where the order position of the proximal steps corresponding to maps S and T in algorithm GSA3O are switched. In order to do that, like to map $\bar{L}$ and matrix $\overline{D}$ corresponding to problem (Var), we denote respectively by $\bar{L}'$ and $\overline{D}'$ the map and matrix for problem (Var_1). Then, the map $G_{\overline{D}}^{\bar{L}}$ defined in (4.15) is replaced by $G_{\overline{D}'}^{\bar{L}'}$ which applies $\mathbb{R}^m \times \mathbb{R}^n$ into itself whose value at (u, x) is

$$\begin{pmatrix} u - M^{\frac{1}{2}} A \widehat{J}^S(x,u) + \tilde{J}^T \left[2M^{\frac{1}{2}} A \widehat{J}^S(x,u) - u - M^{\frac{1}{2}} A \widehat{C}(\widehat{J}^S(x,u)) \right] \\ V^{\frac{1}{2}} \widehat{J}^S(x,u) - V^{\frac{1}{2}} \widehat{C}(\widehat{J}^S(x,u)) \end{pmatrix}$$

where

$$\tilde{J}^T = M^{\frac{1}{2}}(T+M)^{-1}M^{\frac{1}{2}}, \quad \widehat{J}^S = (S+V+A^tMA)^{-1} \circ \begin{pmatrix} V^{\frac{1}{2}} & A^tM^{\frac{1}{2}} \end{pmatrix}$$

and

$$\widehat{C} = (A^tMA+V)^{-1}C.$$

By setting $r = \|(V + A^tMA)^{-1}\|$, we deduce from Proposition 4.1.1, that $G_{\overline{D}'}^{\bar{L}'}$ is $\frac{2\beta}{4\beta - r}-$average with full domain, if $r \leq 2\beta$.

Now applying algorithm (4.16)-(4.20) for problem (Var_1), we get the following chain of sequences:

$$\begin{aligned}
\tilde{x}^{k+1} &= (S + A^tMA + V)^{-1}(A^tM^{\frac{1}{2}}z_1^k + V^{\frac{1}{2}}z_2^k - A^tM^{\frac{1}{2}}y_1^k - V^{\frac{1}{2}}y_2^k) \\
\tilde{y}_1^{k+1} &= y_1^k + M^{\frac{1}{2}}A\tilde{x}^{k+1} - z_1^k \\
\tilde{y}_2^{k+1} &= y_2^k + V^{\frac{1}{2}}\tilde{x}^{k+1} - z_2^k \\
r_1^{k+1} &= M^{\frac{1}{2}}A(A^tMA+V)^{-1}C(\tilde{x}^{k+1}) \\
r_2^{k+1} &= V^{\frac{1}{2}}(A^tMA+V)^{-1}C(\tilde{x}^{k+1}) \\
\tilde{z}_1^{k+1} &= M^{\frac{1}{2}}(T+M)^{-1}M^{\frac{1}{2}}(M^{\frac{1}{2}}A\tilde{x}^{k+1} + \tilde{y}_1^{k+1} - r_1^{k+1}) \\
\tilde{z}_2^{k+1} &= V^{\frac{1}{2}}\tilde{x}^{k+1} + \tilde{y}_2^{k+1} - r_2^{k+1} \\
(x^{k+1}, z^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, z^k, y^k).
\end{aligned}$$

So, by defining

$$\tilde{r}^{k+1} = (A^tMA+V)^{-1}C(\tilde{x}^{k+1}) \quad \text{and} \quad r^{k+1} = \rho\tilde{r}^{k+1} + (1-\rho)r^k$$

and assuming for arbitrary points x^0, r^0, y_2^0 and z_2^0,

$$V^{\frac{1}{2}}z_2^0 - V^{\frac{1}{2}}y_2^0 = Vx^0 - Vr^0,$$

we get applying the mathematical induction ($k \geq 1$)

$$\begin{aligned} V^{\frac{1}{2}} z_2^k - V^{\frac{1}{2}} y_2^k &= \rho(V\tilde{x}^k - V^{\frac{1}{2}} r_2^k) + (1-\rho)(Vx^{k-1} - Vr^{k-1}) \\ &= Vx^k - Vr^k. \end{aligned}$$

Hence, by denoting

$$z^k := M^{-\frac{1}{2}} z_1^k, \quad \tilde{z}^{k+1} := M^{-\frac{1}{2}} \tilde{z}_1^{k+1}, \quad y^k := M^{\frac{1}{2}} y_1^k \quad \text{and} \quad \tilde{y}^{k+1} := M^{\frac{1}{2}} \tilde{y}_1^{k+1},$$

we deduce from the previous chain of sequences our desired new algorithm:

$$\tilde{x}^{k+1} = (S + A^t M A + V)^{-1}(Vx^k + A^t M z^k - A^t y^k - Vr^k) \tag{4.36}$$

$$\tilde{y}^{k+1} = y^k + MA\tilde{x}^{k+1} - Mz^k \tag{4.37}$$

$$\tilde{r}^{k+1} = (A^t M A + V)^{-1} C(\tilde{x}^{k+1}) \tag{4.38}$$

$$\tilde{z}^{k+1} = (T + M)^{-1}(MA\tilde{x}^{k+1} + \tilde{y}^{k+1} - MA\tilde{r}^{k+1}) \tag{4.39}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}, r^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}, \tilde{r}^{k+1}) + (1-\rho)(x^k, z^k, y^k, r^k) \tag{4.40}$$

Notice that the sequence $\{\xi^k = (x^k - r^k, z^k, y^k)\}$ corresponding to the sequence (4.36)-(4.40), satisfies the following relation

$$\overline{Q}\xi^{k+1} = \rho G_{\overline{D}'}^{\overline{L}'}(\overline{Q}\xi^k) + (1-\rho)\overline{Q}\xi^k \tag{4.41}$$

where

$$\overline{Q} = \begin{pmatrix} 0 & M^{\frac{1}{2}} & -M^{-\frac{1}{2}} \\ V^{\frac{1}{2}} & 0 & 0 \end{pmatrix}.$$

4.3 A variant of primal–dual Condat's algorithms

In this section, considering special parameter matrices M and V corresponding to algorithms (4.36)-(4.40) and (4.30)-(4.34), we deduce two algorithms which can be seen as variants of the primal-dual Condat's algorithms $CA1$ and $CA2$ [13] and their respective construction relationship. We also show the relationship of these new algorithms with Davis-Yin's algorithm, and also their ergodic and nonergodic rate of convergence.

Applying (4.36)-(4.40) considering $M = \sigma I_{m\times m}$ and $V = \tau^{-1} I_{n\times n} - \sigma A^t A$, we get the following sequence:

$$\begin{aligned} \tilde{x}^{k+1} &= (S + \tau^{-1} I_{n\times n})^{-1}((\tau^{-1} I_{n\times n} - \sigma A^t A)(x^k - r^k) + \sigma A^t z^k - A^t y^k) \\ \tilde{y}^{k+1} &= y^k + \sigma A\tilde{x}^{k+1} - \sigma z^k \\ \tilde{r}^{k+1} &= \tau C(\tilde{x}^{k+1}) \\ \tilde{z}^{k+1} &= (T + \sigma I_{m\times m})^{-1}(\sigma A\tilde{x}^{k+1} + \tilde{y}^{k+1} - \sigma A\tilde{r}^{k+1}) \end{aligned}$$

$$(x^{k+1}, z^{k+1}, y^{k+1}, r^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}, \tilde{r}^{k+1}) + (1-\rho)(x^k, z^k, y^k, r^k).$$

Defining the new variables $\eta^k = \sigma Ax^k + y^k - \sigma z^k - \sigma Ar^k$ and $\tilde{\eta}^k = \sigma A\tilde{x}^k + \tilde{y}^k - \sigma\tilde{z}^k - \sigma A\tilde{r}^k$, we obtain the following algorithm

Algorithm (**Alg1**)

$$\begin{cases} \tilde{x}^{k+1} = (\tau S + I_{n\times n})^{-1}(x^k - \tau A^t\eta^k - r^k) \\ \tilde{r}^{k+1} = \tau C(\tilde{x}^{k+1}) \\ \tilde{\eta}^{k+1} = (\sigma T^{-1} + I_{m\times m})^{-1}(\eta^k + \sigma A(2\tilde{x}^{k+1} - x^k) + \sigma Ar^k - \sigma A\tilde{r}^{k+1}) \\ (x^{k+1}, \eta^{k+1}, r^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{\eta}^{k+1}, \tilde{r}^{k+1}) + (1-\rho)(x^k, \eta^k, r^k) \end{cases}$$

This algorithm is indeed a variant of $CA1$ for solving optimization problem (4.1). In particular when $\rho = 1$, we recover **YA algorithm [55]** which is described in the introduction of this chapter.

On the other hand, applying (4.30)-(4.34) with $M = \sigma I_{m\times m}$ and $V = \tau^{-1}I_{n\times n} - \sigma A^tA$, we get the following sequence:

$$\begin{aligned} \tilde{z}^{k+1} &= (T + \sigma I_{m\times m})^{-1}(y^k + \sigma Ax^k) \\ \tilde{y}^{k+1} &= y^k + \sigma Ax^k - \sigma\tilde{z}^{k+1} \\ r^{k+1} &= C((\tau^{-1}I_{n\times n} - \sigma A^tA)\tau x^k + \tau\sigma A^t\tilde{z}^{k+1})) \\ \tilde{x}^{k+1} &= (S + \tau^{-1}I_{n\times n})^{-1}\left((\tau^{-1}I_{n\times n} - \sigma A^tA)x^k + \sigma A^t\tilde{z}^{k+1} - A^t\tilde{y}^{k+1} - r^{k+1}\right) \\ (x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k). \end{aligned}$$

Then using that $(\sigma T^{-1} + I_{m\times m})^{-1} = I - \sigma(T + \sigma I_{m\times m})^{-1}$, and eliminating the term $\tilde{z}^{k+1}$, we obtain the following algorithm:

Algorithm (**Alg2**)

$$\begin{cases} \tilde{y}^{k+1} &= (\sigma T^{-1} + I_{m\times m})^{-1}(y^k + \sigma Ax^k) \\ r^{k+1} &= \tau C(x^k - \tau A^t(\tilde{y}^{k+1} - y^k)) \\ \tilde{x}^{k+1} &= (\tau S + I_{n\times n})^{-1}(x^k - \tau A^t(2\tilde{y}^{k+1} - y^k) - r^{k+1}) \\ (x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k). \end{cases}$$

So, by considering σ, τ and ρ positive parameters such that $\sigma\tau\|A\|^2 \leq 1$, $\tau < 2\beta$ and $\rho < \frac{4\beta-\tau}{2\beta}$, then applying Proposition 4.2.1, we deduce the convergence of sequence (x^k, Ax^k, y^k) to an optimal solution of the lagrangian problem corresponding to problem (4.1).

With respect to the convex problem (4.1), algorithm $Alg2$ is a variant of $CA2$ by changing $\nabla h(x^k)$ by $\nabla h(x^k - \tau A^t(\tilde{y}^{k+1} - y^k))$.

4.3.1 Relationship with the Condat's method

When $C = 0$, algorithms $Alg1$ and $Alg2$ are exactly $CA1$ and $CA2$ respectively. Otherwise, when $C \neq 0$, they are different. We consider the lagrangian function defined as

$$l'(x, y) = f(x) - g^*(y) + \langle Ax, y\rangle \tag{4.42}$$

and its corresponding maximal monotone map L' defined on $\mathbb{R}^n \times \mathbb{R}^m$ as

$$L'(x, y) := (\partial_x l') \times (\partial_y[-l']) = \begin{pmatrix} \partial f(x) \\ \partial g^*(y) \end{pmatrix} + \begin{pmatrix} 0 & A^t \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{4.43}$$

The next inequalities which are immediately deduced by definition will be used later in Proposition 4.3.1.

Lemma 4.3.1 *For any (d, d^*), $(\bar{d}, \bar{d}^*) \in \text{graph}\,(L')$, considering $d = (x, y)$ and $\bar{d} = (\bar{x}, \bar{y})$, it holds*

$$\langle d - \bar{d}, d^*\rangle \geq l'(x, \bar{y}) - l'(\bar{x}, y) \geq \langle d - \bar{d}, \bar{d}^*\rangle.$$

These inequalities are still verified if we consider $(d, d^*) \in \text{graph}\,(L')$ and $\bar{d} \in \text{dom}\,(f) \times \text{dom}\,(g^*)$, for the first inequality; and $(\bar{d}, \bar{d}^*) \in \text{graph}\,(L')$ and $d \in \text{dom}\,(f) \times \text{dom}\,(g^*)$, for the second inequality.

Notice that algorithms $CA1$ and $CA2$ generate the sequences $w_i^k = (x^k, y^k)$ and $\tilde{w}_i^k = (\tilde{x}^k, \tilde{y}^k)$, for $i = 1, 2$, respectively, which satisfy the following inclusion

$$L'(\tilde{w}_i^{k+1}) + P_i(\tilde{w}_i^{k+1} - w_i^k) \ni - \begin{pmatrix} \nabla h(x^k) \\ 0 \end{pmatrix}$$

where P_1 and P_2 are matrices defined as

$$P_1 = \begin{pmatrix} \frac{1}{\tau} I & -A^t \\ -A & \frac{1}{\sigma} I \end{pmatrix} \text{ and } P_2 = \begin{pmatrix} \frac{1}{\tau} I & A^t \\ A & \frac{1}{\sigma} I \end{pmatrix}. \tag{4.44}$$

As showed in the proof of Theorem 2 regarding the relaxed primal-dual algorithm given by Chambolle-Pock [10], it holds that

$$2\rho[\mathcal{L}(\tilde{x}^{k+1}, y) - \mathcal{L}(x, \tilde{y}^{k+1})] \leq \|w^k - w\|_{P_i}^2 - \|\tilde{w}^{k+1} - w\|_{P_i}^2 - \frac{2-\rho}{\rho}\|w^{k+1} - w^k\|_{U_i}^2$$

where U_1 and U_2 are two matrices defined as

$$U_1 = \begin{pmatrix} (\frac{1}{\tau} - \frac{1}{\beta(2-\rho)})I & -A^t \\ -A & \frac{1}{\sigma} I \end{pmatrix} \text{ and } U_2 = \begin{pmatrix} (\frac{1}{\tau} - \frac{1}{\beta(2-\rho)})I & A^t \\ A & \frac{1}{\sigma} I \end{pmatrix},$$

and $\mathcal{L}$ defined as

$$\mathcal{L}(x,y) = f(x) + h(x) - g^*(y) + \langle Ax, y\rangle. \tag{4.45}$$

The last inequality is fundamental to deduce the ergodic convergence of $CA1$ and $CA2$ as showed in the aforementioned Chambolle-Pock's paper.

On the other hand, algorithm $Alg1$ generates the sequences $\tilde{\nu}^k = (\tilde{x}^k, \tilde{\eta}^k)$, $\xi^k = (x^k - r^k, \eta^k)$ and $\tilde{\xi}^k = (\tilde{x}^k - \tilde{r}^k, \tilde{\eta}^k)$ satisfying

$$L'(\tilde{\nu}^{k+1}) + P_1(\tilde{\xi}^{k+1} - \xi^k) \ni -\begin{pmatrix} \nabla h(\tilde{x}^{k+1}) \\ 0 \end{pmatrix}. \tag{4.46}$$

Similarly, algorithm $Alg2$ generates the sequences $w^k = (x^k, y^k)$ and $\tilde{w}^{k+1} = (\tilde{x}^{k+1}, \tilde{y}^{k+1})$ satisfying

$$L'(\tilde{w}^{k+1}) + P_2(\tilde{w}^{k+1} - w^k) \ni -\begin{pmatrix} \nabla h(x^k - \tau A^t(\tilde{y}^{k+1} - y^k)) \\ 0 \end{pmatrix}. \tag{4.47}$$

From (4.47) and (4.46), we deduce the following fundamental upper bound of the saddle-point gap. This will be used later in the analysis of the ergodic and nonergodic rates of convergence.

Proposition 4.3.1 *With the same notations as before; let us assume that $\sigma\tau\|A\|^2 \leq 1$, then for any $w = (x,y) \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g^*)$, the following inequalities hold:*

- *For the sequences generated by algorithm $Alg1$, it holds that*

  $$\|\xi^k - w'\|_{P_1}^2 - \lambda\|\xi^{k+1} - \xi^k\|_{P_1}^2 - \|\xi^{k+1} - w'\|_{P_1}^2 \geq 2\rho[\mathcal{L}(\tilde{x}^{k+1}, y) - \mathcal{L}(x, \tilde{\eta}^{k+1})]$$

  where $w' = w - (\tau\nabla h(x), 0)$

- *For the sequences generated by algorithm $Alg2$, it holds that*

  $$\|w^k - w\|_{P_2}^2 - \lambda\|w^{k+1} - w^k\|_{P_2}^2 - \|w^{k+1} - w\|_{P_2}^2 \geq 2\rho[\mathcal{L}(\tilde{x}^{k+1}, y) - \mathcal{L}(x, \tilde{y}^{k+1})]$$

where $\lambda = \frac{1}{\rho}\left[2 - \rho - \frac{\tau}{\beta}\right]$.

Proof. From (4.47), and applying Proposition 4.3.1, we get

$$\left\langle \tilde{\nu}^{k+1} - w, P_1(\xi^k - \tilde{\xi}^{k+1})\right\rangle - \left\langle \tilde{x}^{k+1} - x, \nabla h(\tilde{x}^{k+1})\right\rangle \geq l'(\tilde{x}^{k+1}, y) - l'(x, \tilde{\eta}^{k+1}).$$

using that $\tilde{\nu}^k = \tilde{\xi}^k + (\tilde{r}^k, 0)$ in the last inequality, we have

$$\left\langle \tilde{\xi}^{k+1} - w, P_1(\xi^k - \tilde{\xi}^{k+1})\right\rangle - \left\langle \widehat{u}^{k+1} - x, \nabla h(\tilde{x}^{k+1})\right\rangle \geq l'(\tilde{x}^{k+1}, y) - l'(x, \tilde{\eta}^{k+1}).$$

where $\widehat{u}^{k+1} = \tilde{r}^{k+1} + x^k - r^k + \tau A^t(\tilde{\eta}^{k+1} - \eta^k)$. Then, since h is convex and β^{-1}–Lipschitz–differentiable, we have

$$\langle \nabla h(\tilde{x}^{k+1}), \tilde{x}^{k+1} - x\rangle - \frac{\beta}{2}\|\nabla h(\tilde{x}^{k+1}) - \nabla h(x)\|^2 \geq h(\tilde{x}^{k+1}) - h(x)$$

and from the properties of norm we have

$$\frac{1}{2\beta}\|\tilde{x}^{k+1} - \widehat{u}^{k+1}\|^2 \geq \langle \tilde{x}^{k+1} - \widehat{u}^{k+1}, \nabla h(\tilde{x}^{k+1}) - \nabla h(x)\rangle - \frac{\beta}{2}\|\nabla h(\tilde{x}^{k+1}) - \nabla h(x)\|^2.$$

Then, summing the three last inequalities we get

$$\left\langle \tilde{\xi}^{k+1} - w', P_1(\xi^k - \tilde{\xi}^{k+1})\right\rangle + \frac{1}{2\beta}\|\tilde{x}^{k+1} - \widehat{u}^{k+1}\|^2 \geq \mathcal{L}(\tilde{x}^{k+1}, y) - \mathcal{L}(x, \tilde{\eta}^{k+1}). \quad (4.48)$$

where $w' = w - (\tau\nabla h(x), 0)$.

Now we find an appropriate upper bound for $\frac{1}{2\beta}\|\tilde{x}^{k+1} - \widehat{u}^{k+1}\|^2$. From the expression of P_1 we have

$$\|(x,y)\|^2_{P_1} = \tau^{-1}\|x - \tau A^t y\|^2 + \|y\|^2_{\sigma^{-1}I - \tau AA^t}$$

Then since $\sigma\tau\|A\|^2 \leq 1$, we have that $\sigma^{-1}I - \tau AA^t$ is positive definite matrix, so we get

$$\begin{aligned}
\|\tilde{x}^{k+1} - \widehat{u}^{k+1}\|^2 &= \|\tilde{x}^{k+1} - \tilde{r}^{k+1} - x^k + r^k - \tau A^t(\tilde{\eta}^{k+1} - \eta^k)\|^2 \\
&\leq \tau\|\tilde{\xi}^{k+1} - \xi^k\|^2_{P_1} \\
&= \frac{\tau}{\rho^2}\|\xi^{k+1} - \xi^k\|^2_{P_1}
\end{aligned}$$

On other hand, from the symmetry of matrix P_1, it holds

$$2\rho\left\langle \tilde{\xi}^{k+1} - w', P_1(\xi^k - \tilde{\xi}^{k+1})\right\rangle = \|\xi^k - w'\|^2_{P_1} - \frac{2-\rho}{\rho}\|\xi^{k+1} - \xi^k\|^2_{P_1} - \|\xi^{k+1} - w'\|^2_{P_1}.$$

So, replacing the two last expressions into (4.48), we get the desired inequality of the first item.

Now we proof of second item, from (4.46), and applying Proposition 4.3.1, we get

$$\left\langle \tilde{w}^{k+1} - w, P_2(w^k - \tilde{w}^{k+1})\right\rangle - \left\langle \tilde{x}^{k+1} - x, \nabla h(\widehat{z}^{k+1})\right\rangle \geq l'(\tilde{x}^{k+1}, y) - l'(x, \tilde{y}^{k+1}).$$

where $\widehat{z}^{k+1} = x^k - \tau A^t(\tilde{y}^{k+1} - y^k)$. Then since h is convex diferentiable with ∇h $\beta^{-1}-$Lipschitz, we have

$$\frac{1}{2\beta}\|\widehat{z}^{k+1} - \tilde{x}^{k+1}\|^2 + \langle \nabla h(\widehat{z}^{k+1}), \tilde{x}^{k+1} - \widehat{z}^{k+1}\rangle \geq h(\tilde{x}^{k+1}) - h(\widehat{z}^{k+1})$$

and

$$h(x) \geq h(\widehat{z}^{k+1}) + \langle \nabla h(\widehat{z}^{k+1}), x - \widehat{z}^{k+1} \rangle.$$

Then, summing the three last inequalities we get

$$\left\langle \tilde{w}^{k+1} - w, P_2(w^k - \tilde{w}^{k+1}) \right\rangle + \frac{1}{2\beta} \|\widehat{z}^{k+1} - \tilde{x}^{k+1}\|^2 \geq \mathcal{L}(\tilde{x}^{k+1}, y) - \mathcal{L}(x, \tilde{y}^{k+1}). \quad (4.49)$$

Then using the same techniques as the first item, we get the desired inequality of the second item. ■

4.3.2 Relationship with the Davis-Yin's method

Recently D. O'Connor and L. Vandenberghe [40], noticed that algorithm **YA** (described early in the introduction of this chapter) can be deduced from the algorithm developed by D. Davis and W. Yin [16] by means a reformulation of the sum of three special operators.

Now we show that algorithm $Alg1$ and $Alg2$ are also obtained from Davis-Yin's algorithm considering these adhoc reformulations.

Fixing $A = I$ in problem (Var) and applying algorithms $Alg1$ and $Alg2$ with $\sigma = \tau^{-1}$, we obtain

$$\begin{aligned}
\tilde{y}^{k+1} &= (\tau^{-1}T^{-1} + I_{m\times m})^{-1}(y^k + \tau^{-1}x^k) && (4.50)\\
r^{k+1} &= \tau C(x^k - \tau(\tilde{y}^{k+1} - y^k)) && (4.51)\\
\tilde{x}^{k+1} &= (\tau S + I_{n\times n})^{-1}(x^k - \tau(2\tilde{y}^{k+1} - y^k) - r^{k+1}) && (4.52)\\
(x^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, y^k). && (4.53)
\end{aligned}$$

and the switching algorithm

$$\begin{aligned}
\tilde{x}^{k+1} &= (\tau S + I_{n\times n})^{-1}(x^k - \tau\eta^k - r^k) && (4.54)\\
\tilde{r}^{k+1} &= \tau C(\tilde{x}^{k+1}) && (4.55)\\
\tilde{\eta}^{k+1} &= (\tau^{-1}T^{-1} + I_{m\times m})^{-1}(\eta^k + \tau^{-1}(2\tilde{x}^{k+1} - x^k) + \tau^{-1}r^k - \tau^{-1}\tilde{r}^{k+1}) && (4.56)\\
(x^{k+1}, \eta^{k+1}, r^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{\eta}^{k+1}, \tilde{r}^{k+1}) + (1-\rho)(x^k, \eta^k, r^k). && (4.57)
\end{aligned}$$

Notice that [40] these two algorithms can also be obtained directly from the Davis-Yin's algorithm.

Now, in order to recover algorithms $Alg1$ and $Alg2$ from algorithms (4.50)–(4.53) and (4.54)–(4.57), respectively, we consider the following inclusion problem corresponding with the sum of three operators, as defined in [40] for optimization problems :

$$0 \in \begin{bmatrix} S \\ \mathcal{N}_{\{0\}} \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} A^t \\ \tilde{V}^{\frac{1}{2}} \end{pmatrix} T \begin{pmatrix} A & \tilde{V}^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{bmatrix} C \\ 0 \end{bmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad (Var_2)$$

where $\tilde{V} = (\tau\sigma)^{-1} I - AA^t$.

Notice that algorithms (4.50)–(4.53) and (4.54)–(4.57) need the resolvent maps of $\begin{bmatrix} S \\ \mathcal{N}_{\{0\}} \end{bmatrix}$ and of the inverse of $\begin{pmatrix} A^t \\ \tilde{V} \end{pmatrix} T \begin{pmatrix} A & \tilde{V} \end{pmatrix}$ which by simple calculations are respectively

$$\begin{bmatrix} (\tau S + I_{n\times n})^{-1} \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{pmatrix} A^t \\ \tilde{V} \end{pmatrix} (T^{-1} + \sigma^{-1} I_{m\times m})^{-1} \tau \begin{pmatrix} A & \tilde{V} \end{pmatrix}.$$

Then the aforementioned two algorithms applied to problem (Var_2) are exactly $Alg1$ and $Alg2$.

The fact that problem (Var_2) can be deduced from problem (Var_1) follows from the following steps: We first apply the dual formulation to (Var), which consists in finding $y \in I\!R^m$ such that

$$0 \in T^{-1}(y) - A(S + C)^{-1}(-A^t y).$$

Then reformulate it as (Var_1) considering $M = I$, resulting

$$0 \in T^{-1}(y) + \begin{pmatrix} -A & V^{\frac{1}{2}} \end{pmatrix} \begin{bmatrix} (S + C)^{-1} \\ 0 \end{bmatrix} \begin{pmatrix} -A^t \\ V^{\frac{1}{2}} \end{pmatrix} y.$$

Finally, the dual formulation of the last inclusion problem considering $\tilde{V} = -V$ is exactly (Var_2). Conversely, using the same previous arguments, we can show that problem (Var_1) is deduced from problem (Var_2).

4.3.3 Rate of convergence

Following the same arguments described in Subsection 4.1.4, we can deduce similar rates of convergence for the sequences generated by $Alg1$ and $Alg2$. For that we need the upper bound of the saddle–point gap given in Proposition 4.3.1 and also an upper bound of the fixed–point residual. The last upper bound can be deduced (see Subsection 2.5.1 in Chapter 2) from the following relations which follow respectively

from (4.41) and (4.35) considering $M = \sigma I_{m\times m}$ and $V = \tau^{-1} I_{n\times n} - \sigma A^t A$ in the definition of matrices $\widehat{Q}$ and $\overline{Q}$ involved in (4.35) and (4.41):

$$D_1 w_1^{k+1} = \rho G_{\overline{D}'}^{\overline{L}'}(D_1 w_1^k) + (1-\rho) D_1 w_1^k$$

and

$$D_2 w_2^{k+1} = \rho G_{\widehat{D}'}^{\widehat{L}'}(D_2 w_2^k) + (1-\rho) D_2 w_2^k$$

where $w_1^k = (x^k - r^k, \eta^k)$, $w_2^k = (x^k, y^k)$ and

$$D_1 = \begin{pmatrix} \sigma^{\frac{1}{2}} A & -\sigma^{-\frac{1}{2}} I \\ V^{\frac{1}{2}} & 0 \end{pmatrix} \quad \text{and} \quad D_2 = \begin{pmatrix} \sigma^{\frac{1}{2}} A & \sigma^{-\frac{1}{2}} I \\ V^{\frac{1}{2}} & 0 \end{pmatrix}$$

which satisfy

$$\overline{Q} = D_1 \begin{pmatrix} I & 0 & 0 \\ \sigma A & -\sigma I & I \end{pmatrix} \quad \text{and} \quad \widehat{Q} = D_2.$$

By Proposition 4.1.1, $G_{\overline{D}'}^{\overline{L}'}$ and $G_{\widehat{D}'}^{\widehat{L}'}$ are $\frac{2\beta}{4\beta-\tau}-$average maps, if $\tau < 2\beta$, and their corresponding fixed point sets are respectively

$$\{(\sigma^{\frac{1}{2}} A(\bar{x} - \tau C(\bar{x})) - \sigma^{-\frac{1}{2}}\bar{y}, V^{\frac{1}{2}}(\bar{x} - \tau C(\bar{x})) : -A^t\bar{y} \in S(\bar{x}) + C(\bar{x}),\ \bar{y} \in T(A\bar{x})\}$$

and

$$\{(\sigma^{\frac{1}{2}} A\bar{x} + \sigma^{-\frac{1}{2}}\bar{y}, V^{\frac{1}{2}}\bar{x}) : -A^t\bar{y} \in S(\bar{x}) + C(\bar{x}),\ \bar{y} \in T(A\bar{x})\}.$$

Now, corresponding to algorithms $Alg1$ and $Alg2$, we consider, for $i = 1, 2$, the following sequences:

$$(\tilde{\zeta}_i^k, \tilde{\nu}_i^k) = (\tilde{x}^k, \tilde{\eta}^k), \quad (\omega_i^k, \nu_i^k) = (x^k - (2-i)r^k, \eta^k), \quad \tilde{\omega}_i^k = \tilde{x}^k - (2-i)\tilde{r}^k, \quad k \geq 1,$$

and the ergodic sequences, for $N \geq 1$,

$$\overline{\zeta}_i^N := \frac{1}{N}\sum_{k=1}^{N} \tilde{\zeta}_i^k, \quad \overline{\nu}_i^N := \frac{1}{N}\sum_{k=1}^{N} \tilde{\nu}_i^k, \quad \widehat{\omega}_i^N := \frac{1}{N}\sum_{k=1}^{N} \omega_i^{k-1} \quad \text{and} \quad \widehat{\nu}_i^N := \frac{1}{N}\sum_{k=1}^{N} \nu_i^{k-1},$$

and, associated to matrix P_i defined in (4.44), we consider the norm

$$\|(x,y)\|_{P_i}^2 = \|x\|_V^2 + \sigma\|Ax + (-1)^i \sigma^{-1} y\|^2,$$

where $i = 1$ corresponds to algorithm $Alg1$, and $i = 2$ to $Alg2$.

Using the upper bound of the saddle-point gap and the fixed–point residual developed in Sections 2.5.2 and 2.5.3 of Chapter 2 the following rate of converge are deduced in the two next results:

Theorem 4.3.1 *With the same notations as before. Set σ, τ and ρ satisfying*

$$2 - \frac{\tau}{\beta} \geq \rho > 0, \quad \textit{and} \quad 1 \geq \sigma\tau\|A\|^2.$$

The following rate of convergence are deduced:

- **Ergodic Convergence:** *for any $(x, y) \in \mathrm{dom}\,(f) \times \mathrm{dom}\,(g^*)$ and $i = 1, 2$,*

$$\mathcal{L}(\overline{\zeta}_i^k, y) - \mathcal{L}(x, \overline{\nu}_i^k) \leq \frac{1}{2\rho k}\|(\omega_i^0, \nu_i^0) - (x - \mu_i, y)\|_{P_i}^2 \tag{4.58}$$

  where $\mu_i = (2 - i)\tau\nabla h(x)$.

- **Nonergodic Convergence:** *for any $(x^*, y^*) \in \mathrm{sol}\,(V)$ and $i = 1, 2$,*

$$\mathcal{L}(\tilde{\zeta}_i^{k+1}, y) - \mathcal{L}(x, \tilde{\nu}_i^{k+1}) \leq \left(\frac{\alpha_1}{\sqrt{k+1}} + \frac{\alpha_2}{k+1}\right)\|(\omega_i^0, \nu_i^0) - (x^* - \mu_i^*, y^*)\|_{P_i}^2 \tag{4.59}$$

  where

$$\mu_i^* = (2 - i)\tau\nabla h(x^*), \quad \alpha_1 = \frac{1 + |1 - \rho|}{\rho^2\sqrt{\theta}} \quad \textit{and} \quad \alpha_2 = \frac{\|(A^t\overline{M}A)^{-1}\|}{2\beta\rho^2\theta}.$$

Respect to the rate of constraint violations we have

Theorem 4.3.2 *With the same notations as before. Set σ, τ and ρ satisfying*

$$2 - \frac{\tau}{2\beta} \geq \rho > 0, \quad \textit{and} \quad 1 \geq \sigma\tau\|A\|^2.$$

For any $(x^*, y^*) \in \mathrm{sol}\,(V)$ and $i = 1, 2$, the following rate of convergence are obtained by setting $u_i^* = (x^* - (2 - i)\tau C(x^*), y^*)$:

- **Ergodic Convergence:**

$$\|\tilde{\omega}_i^k - \widehat{\omega}_i^k\|_V^2 + \sigma\|A\tilde{\omega}_i^k - A\widehat{\omega}_i^k + (-1)^i\sigma^{-1}(\tilde{\nu}_i^k - \widehat{\nu}_i^k)\|^2 \leq \frac{4}{k^2\rho^2}\|(\omega_i^0, \nu_i^0) - u_i^*\|_{P_i}^2.$$

- **Nonergodic Convergence:**

$$\|\tilde{\omega}_i^k - \omega_i^{k-1}\|_V^2 + \sigma\|A\tilde{\omega}_i^k - A\omega_i^{k-1} + (-1)^i\sigma^{-1}(\tilde{\nu}_i^k - \nu_i^{k-1})\|^2 \leq \frac{1}{\alpha k}\|(\omega_i^0, \nu_i^0) - u_i^*\|_{P_i}^2.$$

  where $\alpha = \rho(2 - \rho - \frac{\tau}{2\beta})$ and

Remark 4.3.1 *Considering the sequence $(\mathbf{x^k}, \mathbf{z^k}, \mathbf{s^k})$ generated by PD3O [55], the sequence $(x^k, y^k) = (\mathbf{x^k}, \mathbf{s^{k+1}})$ is generated by* **YA** *(or equivalently by* **Alg1** *with*

$\rho = 1$) using $\tau = \gamma$ and $\sigma = \delta$. Moreover, the sequence r^k in **YA** *verifies $z^{k+1} = x^k - r^k - \tau A^t y^k$. One deduces that*

$$\begin{aligned}\|\xi^k - (x - \tau\nabla h(x), s)\|_{P_1}^2 &= \tau^{-1}\|x - \nabla h(x) - \tau A^t s - z^{k+1}\|^2 + \|s - s^{k+1}\|^2_{\delta^{-1}I - \gamma AA^t} \\ &= \tau^{-1}\|(z, s) - (z^{k+1}, s^{k+1})\|^2_{I,M}\end{aligned}$$

where $\|(a, b)\|^2_{I,M} = \|a\|^2 + \|b\|^2_M$ with $M = \frac{\gamma}{\delta}(I - \gamma\delta AA^t)$ defined in [55]. The last equality relations and the upper bound given in Proposiion 4.3.1, allow us recover the upper bound (36) *given in Theorem* 2 *of [55] related to the aforementioned sequence* $(\mathbf{x^k}, \mathbf{z^k}, \mathbf{s^k})$.

Notice that Theorem 2 *of [55] is exactly the ergodic convergence of Theorem 4.3.1.*

4.4 General separable optimization problem

Following the same scheme described in Section 4.2 regarding the optimization problem (4.1), we reformulate problem (P) keeping the same structure of problem (Var), where the involved matrix is injective and then we apply the algorithm developed in Section 4.1.

Set M, V_1 and V_2 symmetric matrices of order $m\times m$, $n\times n$ and $p\times p$, respectively, with V_1 positive semi-definite and M and V_2 positive definite. Then problem (P) can be formulated as

$$\min_{(x_1,x_2,z_1,z_2)\in\mathcal{F}} (f + h_1, h_2)(x_1, x_2) + (g, 0)(z_1, z_2),$$

where $\mathcal{F}$ denotes the set of all (x_1, x_2, z_1, z_2) satisfying

$$\begin{pmatrix} M^{\frac{1}{2}}A & 0 \\ V_1^{\frac{1}{2}} & 0 \\ 0 & V_2^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} M^{\frac{1}{2}}B & 0 \\ 0 & -I_n \\ -V_2^{\frac{1}{2}} & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 0.$$

The dual problem of its variational formulation consists in finding a zero of the sum of the two composition maps given by

$$-\begin{pmatrix} M^{\frac{1}{2}}A & 0 \\ V_1^{\frac{1}{2}} & 0 \\ 0 & V_2^{\frac{1}{2}} \end{pmatrix} \begin{bmatrix} \partial f + \nabla h_1 \\ \nabla h_2 \end{bmatrix}^{-1} \left(-\begin{pmatrix} A^t M^{\frac{1}{2}} & V_1^{\frac{1}{2}} & 0 \\ 0 & 0 & V_2^{\frac{1}{2}} \end{pmatrix}\right)$$

and

$$-\begin{pmatrix} M^{\frac{1}{2}}B & 0 \\ 0 & -I_n \\ -V_2^{\frac{1}{2}} & 0 \end{pmatrix} \begin{bmatrix} \partial g \\ 0 \end{bmatrix}^{-1} \left(-\begin{pmatrix} B^t M^{\frac{1}{2}} & 0 & -V_2^{\frac{1}{2}} \\ 0 & -I_n & 0 \end{pmatrix}\right).$$

Then the dual formulation of this sum problem is

$$0 \in \begin{bmatrix} \partial f \\ 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{bmatrix} \nabla h_1 \\ \nabla h_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} A^t M^{\frac{1}{2}} & V_1^{\frac{1}{2}} & 0 \\ 0 & 0 & V_2^{\frac{1}{2}} \end{pmatrix} \mathcal{G} \begin{pmatrix} M^{\frac{1}{2}} A & 0 \\ V_1^{\frac{1}{2}} & 0 \\ 0 & V_2^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \tag{4.60}$$

where the map $\mathcal{G}$ is the inverse of the composite map defined by

$$\begin{pmatrix} -M^{\frac{1}{2}} B & 0 \\ 0 & I_n \\ V_2^{\frac{1}{2}} & 0 \end{pmatrix} \begin{bmatrix} \partial g \\ 0 \end{bmatrix}^{-1} \begin{pmatrix} -B^t M^{\frac{1}{2}} & 0 & V_2^{\frac{1}{2}} \\ 0 & I_n & 0 \end{pmatrix}$$

which is clearly monotone.

The next proposition gives an explicit expression of the resolvent of $\mathcal{G}$ and thereby its maximal monotonicity by Minty's theorem.

Proposition 4.4.1 *With the same notations as before, for given $(x, y, z) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p$, it holds that*

$$(\mathcal{G} + I)^{-1}(x, y, z) = (-M^{\frac{1}{2}} Bu, y, V_2^{\frac{1}{2}} u),$$

where I denotes the identity map of order $m + n + p$, and

$$u = (\partial g + V_2 + B^t M B)^{-1}(-B^t M^{\frac{1}{2}} x + V_2^{\frac{1}{2}} z).$$

Proof. Since $(\mathcal{G} + I)^{-1} = I - (\mathcal{G}^{-1} + I)^{-1}$, then using Proposition 2.4.1 of Chapter 2 we obtain

$$(\mathcal{G} + I)^{-1} = \begin{pmatrix} -M^{\frac{1}{2}} B & 0 \\ 0 & I_n \\ V_2^{\frac{1}{2}} & 0 \end{pmatrix} \begin{bmatrix} \partial g + B^t M B + V_2 \\ I_n \end{bmatrix}^{-1} \begin{pmatrix} -B^t M^{\frac{1}{2}} & 0 & V_2^{\frac{1}{2}} \\ 0 & I_n & 0 \end{pmatrix}$$

from which the desired equality is deduced. ∎

Observe that problems (4.60) and (Var) have same structure where the involved matrix in the first one is injective and whose corresponding maps verify the properties required in Proposition 4.1.1. So, the corresponding algorithm described by the sequential update formulae $(4.4) - (4.8)$ converges to a solution of its corresponding saddle-point problem, whose solution set is

$$\{(\bar{x}, \bar{z}, M^{\frac{1}{2}} A\bar{x}, V_1^{\frac{1}{2}} \bar{x}, V_2^{-\frac{1}{2}} \bar{z}, M^{-\frac{1}{2}} \bar{y}, 0_{y_2}, -V_2^{-\frac{1}{2}} \nabla h_2(\bar{z})) : (0_x, 0_z, 0_y) \in \tilde{L}(\bar{x}, \bar{z}, \bar{y})\},$$

where $\tilde{L}$ is the classical Lagrangian map corresponding to problem (P), which is defined as

$$\tilde{L}(x, z, y) := \begin{pmatrix} \partial f(x) + \nabla h_1(x) \\ \partial g(z) + \nabla h_2(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}.$$

So, applying the sequential update formulae (4.4) $\to \cdots \to$ (4.8) for $\overline{M} = I$ and $\rho = 1$, and using Proposition 4.4.1 we get the following update sequences

$$\begin{aligned}
z_1^{k+1} &= -M^{\frac{1}{2}} B(\partial g + V_2 + B^t MB)^{-1}(V_2^{\frac{1}{2}}(y_3^k + V_2^{\frac{1}{2}} x_2^k) - B^t M^{\frac{1}{2}}(y_1^k + M^{\frac{1}{2}} A x_1^k)) \\
z_2^{k+1} &= y_2^k + V_1^{\frac{1}{2}} x_1^k \\
z_3^{k+1} &= V_2^{\frac{1}{2}}(\partial g + V_2 + B^t MB)^{-1}(V_2^{\frac{1}{2}}(y_3^k + V_2^{\frac{1}{2}} x_2^k) - B^t M^{\frac{1}{2}}(y_1^k + M^{\frac{1}{2}} A x_1^k)) \\
y_1^{k+1} &= y_1^k + M^{\frac{1}{2}} A x_1^k - z_1^{k+1} \\
y_2^{k+1} &= y_2^k + V_1^{\frac{1}{2}} x_1^k - z_2^{k+1} \\
y_3^{k+1} &= y_3^k + V_2^{\frac{1}{2}} x_2^k - z_3^{k+1} \\
r_1^{k+1} &= \nabla h_1((V_1 + A^t MA)^{-1}(V_1^{\frac{1}{2}} z_2^{k+1} + A^t M^{\frac{1}{2}} z_1^{k+1})) \\
r_2^{k+1} &= \nabla h_2(V_2^{-\frac{1}{2}} z_3^{k+1}) \\
x_1^{k+1} &= (\partial f + V_1 + A^t MA)^{-1}(V_1^{\frac{1}{2}}(z_2^{k+1} - y_2^{k+1}) + A^t M^{\frac{1}{2}}(z_1^{k+1} - y_1^{k+1}) - r_1^{k+1}) \\
x_2^{k+1} &= V_2^{-1}(V_2^{\frac{1}{2}}(z_3^{k+1} - y_3^{k+1}) - r_2^{k+1}).
\end{aligned}$$

By construction we can reduce some sequences: for $k \geq 1$, one has

- $y_2^k = 0$ and hence $z_2^{k+1} = V_1^{\frac{1}{2}} x_1^k$, and
- on other hand,

$$V_2^{\frac{1}{2}}(y_3^k + V_2^{\frac{1}{2}} x_2^k) = V_2^{\frac{1}{2}} y_3^k + V_2 V_2^{-1}(V_2^{\frac{1}{2}}(z_3^k - y_3^k) - r_2^k) = V_2^{\frac{1}{2}} z_3^k - r_2^k.$$

Hence, by denoting $x^k = x_1^k$, $z^k = V_2^{-\frac{1}{2}} z_3^k$ and $y^k = M^{\frac{1}{2}} y_1^k$, the above chain of sequences is reduced to, for $k \geq 1$,

$$z^{k+1} = (\partial g + V_2 + B^t MB)^{-1}(V_2 z^k - r_2^k - B^t y^k - B^t MAx^k)) \tag{4.61}$$

$$y^{k+1} = y^k + MAx^k + MBz^{k+1} \tag{4.62}$$

$$r_1^{k+1} = \nabla h_1((V_1 + A^t MA)^{-1}(V_1 x^k - A^t MBz^{k+1})) \tag{4.63}$$

$$r_2^{k+1} = \nabla h_2(z^{k+1}) \tag{4.64}$$

$$x^{k+1} = (\partial f + V_1 + A^t MA)^{-1}(V_1 x^k - A^t MBz^{k+1} - A^t y^{k+1} - r_1^{k+1}). \tag{4.65}$$

Notice that the update sequences (4.61) − (4.65) is also satisfied for $k = 0$, if we consider $y_2^0 = 0$, and y_3^0 and x_2^0 satisfying

$$V_2^{\frac{1}{2}}(y_3^0 + V_2^{\frac{1}{2}} x_2^0) = V_2^{\frac{1}{2}} z_3^0 - r_2^0$$

for arbitrary r_2^0 and z_3^0.

From Proposition 4.1.2, we get the following convergence result of the sequences described by (4.61) − (4.65)

Proposition 4.4.2 *With the same notations as before, we assume that* $V_1 \in \mathbb{R}^{n\times n}$, $V_2 \in \mathbb{R}^{p\times p}$ *and* $M \in \mathbb{R}^{m\times m}$ *are symmetric with* V_2 *and* M *positive definite such that* $V_1 + A^tMA$ *is positive definite and satisfying that* $\max(\|V_2^{-1}\|, \|(V_1 + A^tMA)^{-1}\|) \in]0, 2\min(\beta_1, \beta_2)[$. *If* $\mathrm{sol}\,(V_{\tilde{L}})$ *is nonempty, then for arbitrary points* $(x^0, z^0, y^0) \in (\mathbb{R}^n \times \mathbb{R}^p) \times \mathbb{R}^m$ *and* $r_2^0 \in \mathbb{R}^p$, *the sequence* (x^k, z^k, y^k) *generated by* $(4.61) - (4.65)$ *converges to some element of* $\mathrm{sol}\,(V_{\tilde{L}})$.

4.5 Application to the decomposition of multi-block optimization problems

In this section we extend the algorithm (PMA) described in Chapter 2 in order to solve the more general *S-Model* defined below. This extension uses a similar reformulation as described in the mentioned chapter and has similar structure as problem (Var) described in Section 4.2.

Our $S-$Model problem is as follows

$$\begin{aligned} \inf_{x=(x_1,\cdots,x_q),\,z} \quad & \sum_{i=1}^{q} f_i(x_i) + h(x) + g(z) \\ s.t \quad & \sum_{i=1}^{q} A_i x_i - Bz = 0, \end{aligned}$$

where $f_i : \mathbb{R}^{n_i} \to \overline{\mathbb{R}}$ $(i \in \{1, ..., q\})$ and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ are proper lsc convex functions, $h : \mathbb{R}^n \mapsto \mathbb{R}$ is convex and $(\frac{1}{\beta})$-Lipschitz-differentiable $(n = \sum_{i=1}^{q} n_i)$, and A_i and B are matrices of order $p \times n_i$ and $p \times m$, respectively.

It is clear that this problem is equivalent to

$$\inf_{(x=(x_1,\cdots,x_q),\,z)} \sum_{i=1}^{q} f_i(x_i) + g(z) + \delta_{\{0\}}\left(Bz - \sum_{i=1}^{q} A_i x_i\right) + h(x) \qquad (B_1)$$

or again

$$V_P = \inf_{x=(x_1,\cdots,x_q)} \sum_{i=1}^{q} f_i(x_i) + (g^* \circ B^t)^*\left(\sum_{i=1}^{q} A_i x_i\right) + h(x). \qquad (B_2)$$

We note that (B_2) has the same structure as the optimization problem (4.1) given in the introduction of this chapter. If we apply **Alg1** or **Alg2**, we obtain an algorithm with separable structure but unfortunately with two important disadvantages: the necessity to know the norm of A in order to choose the parameters for the convergence result, and also the necessity to know all values of $(\tau\partial f_i + I)^{-1}$ at arbitrary points (parameter τ beings equal for all i).

So, we consider a reformulation of (B_2) allowing us to choose in an independently manner parameters for each block, where we only need to know the norm of each A_i separatly.

By considering $f : I\!R^n \to \overline{I\!R}$ defined as $f(x) = \sum_{i=1}^{q} f_i(x_i)$, and matrices K and A of order $p \times pq$ and $pq \times n$, respectively, defined by

$$K := \begin{pmatrix} I_{p\times p} & \cdots & I_{p\times p} \end{pmatrix} \quad \text{and} \quad A := \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_q \end{pmatrix},$$

problem (B_2) can be set as

$$\inf_{x=(x_1,\cdots,x_q)} \quad f(x) + [(g^* \circ B^t)^* \circ K](Ax) + h(x). \qquad (P_{S-mod})$$

Note that this new problem has also the same structure as problem (4.1) but with a good separable structure since f and A have separable structure for blocks. Then, we apply to this last problem algorithm GSA3O developed in Section 4.2. Alternatively, we can apply algorithm $(4.36) - (4.40)$, but for simplicity we will not do it in this work.

Regarding function g, we assume that

$$\partial[(g^* \circ B^t)^* \circ K] = K^t(B(\partial g)^{-1}B^t)^{-1}K.$$

The saddle-point problem of (P_{S-mod}) is

$$\text{Find } (\bar{x}, \bar{z}, \bar{y}) \in I\!R^n \times I\!R^{pq} \times I\!R^{pq} \text{ such that } \ 0 \in \bar{L}(\bar{x}, \bar{z}, \bar{y}) \qquad (V_{\bar{L}})$$

where $\bar{L}$ is the maximal monotone map defined on $I\!R^n \times I\!R^{pq} \times I\!R^{pq}$ as

$$\bar{L}(x, z, y) := \begin{pmatrix} \partial f(x) + \nabla h(x) \\ K^t(B(\partial g)^{-1}B^t)^{-1}Kz \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & -I \\ -A & I & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}.$$

For $i \in \{1, \cdots, q\}$, let M_i be an $p \times p$ symmetric positive definite matrix and Q_i be an $n_i \times n_i$ symmetric positive semi-definite matrix.

In order to take advantage of the separability of f, we consider the diagonal matrices $V = \text{diag}\,(\mathrm{Q}_1, \cdots, \mathrm{Q}_\mathrm{q})$ and $M = \text{diag}\,(\mathrm{M}_1, \cdots, \mathrm{M}_\mathrm{q})$. So, the related algorithm GSA3O has the following structure:

$$\tilde{z}^{k+1} = (K^t(B(\partial g)^{-1}B^t)^{-1}K + M)^{-1}(MAx^k + y^k) \qquad (4.66)$$

$$\tilde{y}^{k+1} = y^k + M(Ax^k - \tilde{z}^{k+1}) \qquad (4.67)$$

$$r^{k+1} = \nabla h((A^tMA + V)^{-1}(Vx^k + A^tM\tilde{z}^{k+1})) \qquad (4.68)$$

$$\tilde{x}^{k+1} = (\partial f + A^tMA + V)^{-1}(Vx^k + A^tM\tilde{z}^{k+1} - A^t\tilde{y}^{k+1} - r^{k+1}) \qquad (4.69)$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1-\rho)(x^k, z^k, y^k). \tag{4.70}$$

We finally get the following algorithm

Separable Primal-Dual Variant
(SPDV)

For $i \in \{1, \cdots, q\}$ set $Q_i \in I\!R^{n_i \times n_i}$ symmetric positive semi-definite, $M_i \in I\!R^{p \times p}$ symmetric positive definite. Set $\Sigma = \left(\sum_{i=1}^q M_i^{-1}\right)^{-1}$. Then for an arbitrary $(x^0, z^0, y_c^0) \in I\!R^n \times I\!R^{pq} \times I\!R^p$

Step 1. Find ζ^{k+1} such that

$$\zeta^{k+1} = \text{argmin} \left\{ g(w) + \frac{1}{2}\|Bw - \sum_{j=1}^q (A_j x_j^k) - \Sigma^{-1} y_c^k\|_\Sigma^2 \right\}.$$

Step 2. Find $\tilde{z}^{k+1}$

For all $i \in \{1, ..., q\}$ **do**

Find $\tilde{z}_i^{k+1}$ such that

$$\tilde{z}_i^{k+1} = A_i x_i^k - M_i^{-1} \Sigma \left(\sum_{j=1}^q (A_j x_j^k) - B\zeta^{k+1} \right).$$

Step 3. Find $\tilde{y}_c^{k+1}$ such that

$$\tilde{y}_c^{k+1} = y_c^k + \Sigma \left(\sum_{j=1}^q (A_j x_j^k) - B\zeta^{k+1} \right).$$

end for

Step 4. Find $r^{k+1} = (r_1^{k+1}, \cdots, r_q^{k+1})$ such that

$$r^{k+1} = \nabla h((A^t M A + Q)^{-1}(Qx^k + A^t M \tilde{z}^{k+1}))$$

where $A = diag([A_1, ..., A_q])$, $Q = diag([Q_1, ..., Q_q])$ and $M = diag([M_1, ..., M_q])$

Step 5. Find $\tilde{x}^{k+1}$

For all $i \in \{1, ..., q\}$ **do**

Find $\tilde{x}_i^{k+1}$ such that

$$\tilde{x}_i^{k+1} = \text{argmin} \left\{ f_i(x_i) + \frac{1}{2}\|A_i x_i - \tilde{z}_i^{k+1} + M_i^{-1}\tilde{y}_c^{k+1} + A_i \tilde{r}_i^{k+1}\|_{M_i}^2 + \frac{1}{2}\|x_i - x_i^k + \tilde{r}_i^{k+1}\|_{Q_i}^2 \right\}$$

where $\tilde{r}_i^{k+1} = (A_i^t M_i A_i + Q_i)^{-1} r_i^{k+1}$.

end for

Step 6. Find $(x^{k+1}, z^{k+1}, y_c^{k+1})$

$$(x^{k+1}, z^{k+1}, y_c^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}_c^{k+1}) + (1-\rho)(x^k, z^k, y_c^k).$$

The following converge result is deduced.

Proposition 4.5.1 *Assume that Q_i and M_i are symmetric matrices of order $n_i \times n_i$ and $p \times p$, respectively, with Q_i positive semi-definite and M_i positive definite such that $Q_i + A_i^t M_i A_i$ is positive definite and satisfying $\|(Q_i + A_i^t M_i A_i)^{-1}\| \in]0, 2\beta[$. Let $\rho \in]0, \alpha^{-1}[$ where $\alpha := \frac{2\beta}{4\beta - \max_i\{\|(V_i + A_i^t M_i A_i)^{-1}\|\}_{i=1}^q}$. If $\mathrm{sol}\,(V_{\bar{L}})$ is nonempty, then for an arbitrary $(x^0, z^0, y_c^0) \in I\!R^n \times I\!R^{pq} \times I\!R^p$, the sequence $(x^k, z^k, K^t y_c^k)$ generated by (SPDV) converges to some element of $\mathrm{sol}\,(V_{\bar{L}})$.*

Remark 4.5.1 *Choosing matrices*

$$M_i = \sigma_i I_{p\times p} \quad \text{and} \quad Q_i = \tau_i^{-1} I_{n_i \times n_i} - \sigma_i A_i^t A_i,$$

the subproblem in Step 5 of the Algorithm (PMA) becomes

$$\tilde{x}_i^{k+1} = \mathrm{argmin}\left\{ f_i(x_i) + \frac{1}{2\tau_i}\|x_i - x_i^k - \tau_i[\sigma_i A_i^t \tilde{z}_i^{k+1} - \sigma_i A_i^t A_i x_i^k - A_i^t \bar{y}_c^{k+1} - r^{k+1}]\|^2 \right\}.$$

If in addition the positive parameters σ_i and τ_i are chosen satisfying $\sigma_i \tau_i \|A_i\|^2 \leq 1$ and $\tau_i < 2\beta$, then the conditions on matrices Q_i, M_i and $Q_i + A_i^t M_i A_i$ in Proposition 4.5.1 are immediately verified and thereby the sequence $(x^k, z^k, K^t y_c^k)$ generated by (SPDV) converges to some element of $\mathrm{sol}\,(V_{\bar{L}})$ if nonempty.

4.6 Numerical Example

We consider the problem (commonly referred as *fused lasso*) with the least squares loss as in [55]

$$\min_x \ \frac{1}{2}\|Qx - b\|_2^2 + \mu_1\|x\|_1 + \mu_2\|Ax\|_1 \tag{4.71}$$

where $Q \in I\!R^{n\times p}$, $b \in I\!R^n$ and

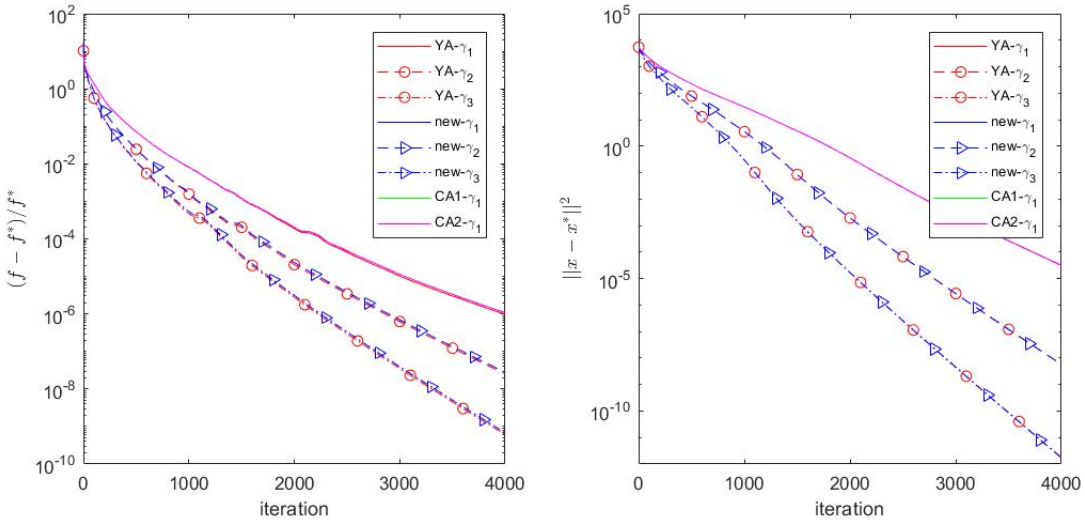
$$A = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \cdots & \cdots & \\ & & & -1 & 1 \end{bmatrix} \in I\!R^{(p-1)\times p}.$$

We consider $n = 200$, $p = 4000$, $\mu_1 = 20$ and $\mu_2 = 200$. Moreover, matrix Q and vector b quoted from [55].

We compare the algorithms $CA1$, $CA2$ with their variants algorithms $Alg1$ and $Alg2$ considering $\rho = 1$ as relaxing parameter. Notice that in this case $Alg1$ coincides with YA. We choose as parameters

$$\tau = \gamma \frac{1}{\|A\|^2} \quad \text{and} \quad \sigma = \frac{1}{8\tau}$$

and we implement the last algorithms considering three variant for γ as $\gamma_1 = 1$, $\gamma_2 = 1.5$ and $\gamma_3 = 1.99$. For $CA1$ and $CA2$ we only use γ_1 because for γ_2 and γ_3 the convergence is not guaranteed.



We observe that all algorithms have the same behavior for γ_1, but for γ_2 and γ_3 algorithms YA and **Alg2** (= New) have more velocity and maintain the same behavior.

Chapter 5

Application to stochastic problems

In this chapter, we will consider a large-scale production planning problem with a multiple separable structure which is favorable to the use of the splitting techniques which constitute the heart of the present thesis. Motivated by a long-term energy production planning problem, we analyze here a stochastic optimal control problem where three levels of coupling structure are present, namely:

- the coupling of the dynamic equations w.r.t. time intervals;
- the coupling of the scenario tree w.r.t. the so-called *non-anticipativity* constraints, i.e. which force the decisions at some period t to be the same for scenarios with identical past history before t;
- the spatial coupling which interconnects the local subsystems

Most approaches in the literature to treat stochastic multistage optimization problems use a scenario tree where the non-anticipativity constraints at each node of the tree are dualized to allow the temporal decomposition as if the model was deterministic (see [48] for instance).

5.1 The stochastic optimization model

We study the model problem presented in [33]. Consider a set of agents Z (geographical zones, markets) with interconnections between them given by a graph ($E \subset Z \times Z$). Given a finite period time $\{0, \cdots, T-1\}$, for each agent ($z \in Z$) there are a production ($p_{z\tau}$), demand ($d_{z\tau}$), storage ($x_{z\tau}$) and interchange ($f_{e\tau}$) of a commodity (electricity, gas). The objective is to minimize the cost associated with the production, interchange, usage of storage of a commodity, in order to satisfy the demand. The **usage of the storage** ($u_{z\tau}$) and the **storage** ($x_{z\tau}$) evolve in time satisfying a dynamic equation. Finally, uncertainty affects the following data:

the local demands ($d_{z\tau}$) are random processes and we consider additional terms $i_{z\tau}$ that are random input of the storage. The distribution of these random processes is supposed to be known and generally approximated by a finite set of historical scenarios.

Since we are working with random variables affecting the dynamic equations defining the state and control decisions, we need to consider the nonanticipativity constraints that rule the sequence of decisions when the successive realizations of the random values are revealed at each stage. More specifically, given the following constraints :

- The demand equation is given by

$$p_{z\tau} + u_{z\tau} + \sum_{e \in z^+} f_{e\tau} - \sum_{e \in z^-} f_{e\tau} = d_{z\tau}$$

  where z^+ (resp. z^-) is the set of outgoing (res. ingoing) arcs incident to zone z.

- The storage $x_{z\tau}$ obeys the dynamics

$$x_{z,\tau+1} = x_{z\tau} - u_{z\tau} + i_{z\tau}$$

  where $i_{z\tau}$ is a random input in the storage.

- The quantity transported through line $f_{e\tau}$ satisfies the capacity constraints as

$$f_{e\tau} \in F_{e\tau}$$

- The variables x_{zt}, u_{zt} and p_{zt} satisfy some constraints

$$x_{z\tau} \in X_{z\tau},\ u_{z\tau} \in U_{z\tau},\ p_{z\tau} \in P_{z\tau}$$

- The control variable $f_{e\tau}$, $u_{z\tau}$ and $p_{z\tau}$ should then satisfy the nonanticipativity equations

$$p_{z\tau}, u_{z\tau}, f_{e\tau} \preceq \mathcal{F}_\tau.$$

  i.e. $p_{z\tau}, u_{z\tau}, f_{e\tau}$ are $\mathcal{F}_\tau-$measurable where $\mathcal{F}_\tau$ is a $\sigma-$field defined as

$$\mathcal{F}_\tau = \sigma(\{(d_{z\tau}, i_{z\tau}) : z \in Z, \tau \in [0,t]\})$$

Summarizing, we consider the following multistage stochastic problem

$$\min_{p,u,f,x\in L^2(\Omega)} I\!E\left[\sum_{\tau=0}^{T-1}\left(\sum_{z\in Z} c_z(p_{z,t}) + \sum_{e\in E} l_{e\tau}(f_{e\tau})\right)\right] \tag{5.1}$$

$$p_{z\tau} + u_{z\tau} + \sum_{e\in z^+} f_{e\tau} - \sum_{e\in z^-} f_{e\tau} = d_{z\tau}, \qquad \forall z \in Z,\ \tau \in [0, T-1] \tag{5.2}$$

$$x_{z,\tau+1} = x_{z\tau} - u_{z\tau} + i_{z\tau}, \qquad \forall z \in Z,\ \tau \in [0, T-1] \tag{5.3}$$

$$x_{z\tau} \in X_{z\tau},\ u_{z\tau} \in U_{z\tau},\ p_{z\tau} \in P_{z\tau} \qquad \forall z \in Z,\ \tau \in [0, T-1] \tag{5.4}$$

$$f_{e\tau} \in F_{e\tau} \qquad \forall e \in E,\ \tau \in [0, T-1] \tag{5.5}$$

$$x_{z,0} = \tilde{x}_{z0}, \qquad \forall z \in Z \tag{5.6}$$

$$p_{z\tau}, u_{z\tau} \preceq \mathcal{F}_\tau \qquad \forall z \in Z,\ \tau \in [0, T-1] \tag{5.7}$$

$$f_{e\tau} \preceq \mathcal{F}_\tau \qquad \forall e \in E,\ \tau \in [0, T-1] \tag{5.8}$$

5.2 Solution of a deterministic formulation

In the first part of this chapter, we study the deterministic case of the last model. The modeling of cost functions c_z and $l_{e\tau}$ is borrowed from [14].

Cost on the final state

The hydroelectric production cost is negligible in the considered model. On the other hand, we add a cost on the final state $x_{zT} \mapsto \Psi(x_{zT})$ to penalize the excess of water reserves, defined as

$$\Psi(x_{zT}) = c_z^{fin} \max\{0, x_{z0} - x_{zT}\}.$$

Thermic production

The thermic production cost is a piecewise affine and convex function of the production levels $p_{z\tau}$. It will be defined with a given number of stages $j = 1, \ldots, Q_{z\tau}$, each one associated with a given slope $c_{z\tau}^j$ valid in the interval $[P_{z\tau}^j, P_{z\tau}^{j+1}]$. We need obviously $0 \le c_{z\tau}^1 \le c_{z\tau}^2 \le \cdots \le c_{z\tau}^{Q_{z\tau}}$ to obtain an increasing convex function of $p_{z\tau}$. The cost function is thus defined by

$$g_{z\tau}(p_{z\tau}) = \begin{cases} c_{z\tau}^1 p_{z\tau} & \text{if } 0 \le p_{z\tau} \le P_{z\tau}^1 \\ c_{z\tau}^1 P_{z\tau}^1 + c_{z\tau}^2 (p_{z\tau} - P_{z\tau}^1) & \text{if } P_{z\tau}^1 \le p_{z\tau} \le P_{z\tau}^2 \\ \vdots & \vdots \\ \sum_{j=1}^{Q_{z\tau}-1} c_{z\tau}^j \widehat{P}_{z\tau}^j + c_{z\tau}^{Q_{z\tau}} (p_{z\tau} - P_{z\tau}^{Q_{z\tau}-1}) & \text{if } P_{z\tau}^{Q_{z\tau}-1} \le p_{z\tau} \le P_{z\tau}^{Q_{z\tau}} \\ +\infty & \text{else} \end{cases}$$

where we defined $\widehat{P}^j_{z\tau} = P^j_{z\tau} - P^{j-1}_{z\tau}$ with $P^{-1}_{z\tau} = 0$.

The cost function $g_{z\tau}$ is reformulated as

$$g_{z\tau}(p_{z\tau}) = \begin{cases} \inf_\theta & \sum_{j=1}^{Q_{z\tau}} c^j_{z\tau}\theta^j_{z\tau} \\ \text{s.t.} & \sum_{j=1}^{Q_{z\tau}} \theta^j_{z\tau} = p_{z\tau} \\ & 0 \le \theta^j_{z\tau} \le \widehat{P}^j_{z\tau} \end{cases}$$

where $\theta^j_{z\tau}$ represents the production of stage j.

Interzonal transfer costs
For an arc $e = (z, z') \in E$ interconnecting two zones z and z', the flow transfer during period τ is the variable $f_{e\tau}$ which is bounded by $0 \le f_{e\tau} \le \kappa_{e\tau}$. The transfer cost is linear and given by $l_{e\tau}(f_{e\tau}) = c^{inter}_e f_{e\tau}$.

Failure cost
The failure quantity corresponds to the part of demand not satisfied during period τ in zone z. It will be denoted by $\eta_{z\tau}$. It is penalized by the cost $c^{fail} >> c^j_{z\tau}$.

Using former formulations of the cost functions, the problem becomes

$$\min_{(\theta,u,f,\eta,v)} \quad \sum_{\tau=0}^{T-1}\left[\sum_{z\in Z}\left[\sum_{j=1}^{Q_{z\tau}} c^j_{z\tau}\theta^j_{z\tau} + c^{fail}\eta_{z\tau}\right] + \sum_{e\in E} l_{e\tau}f_{e\tau}\right] + \sum_{z\in Z} c^{fin}_z v_z \quad (5.9)$$

satisfying

$$u_{z\tau} + \sum_{j=1}^{Q_{z\tau}} \theta^j_{z\tau} - \sum_{e\in z^+} f_{e\tau} + \eta_{z\tau} = d_{z\tau} - \sum_{e\in z^-} f_{e\tau},\ \forall z \in Z,\ \tau \in [0, T-1] \quad (5.10)$$

$$x_{z,\tau+1} = x_{z\tau} - u_{z\tau} + i_{z\tau}, \quad \forall z \in Z,\ \tau \in [0, T-1] \quad (5.11)$$

$$X^{min}_{z\tau} \le x_{z\tau} \le X^{max}_{z\tau}, \quad \forall z \in Z,\ \tau \in [0, T-1] \quad (5.12)$$

$$0 \le u_{z\tau} \le U^{max}_z \delta_h, \quad \forall z \in Z,\ \tau \in [0, T-1] \quad (5.13)$$

$$0 \le \theta^j_{z\tau} \le P^j_{z\tau}\delta_h, \quad \forall z \in Z,\ \tau \in [0, T-1] \quad (5.14)$$

$$0 \le f_{e\tau} \le \kappa_{e\tau}\delta_h, \quad \forall e \in E,\ \tau \in [0, T-1] \quad (5.15)$$

$$v_z \ge 0,\ v_z \ge (x_{z0} - x_{zT}), \quad \forall z \in Z \quad (5.16)$$

We rewrite this problem in the context of problem (P) defined in Chapter 2, in order to apply GSS algorithm also described in that chapter.

Setting $q_{z\tau} = ((\theta^j_{z\tau})_{j\in Q_{z\tau}}, u_{z\tau}, \eta_{z\tau}, x_{z,\tau+1}) \in I\!R^{Q_{z\tau}+3}$ and $f_\tau = (f_{e\tau})_{e\in E}$, relation (5.10) becomes

$$A_{z\tau}q_{z\tau} - B_z f_\tau = d_{z\tau}, \quad \forall z \in Z, \forall \tau \in [0, T-1]$$

where B_z is the row z of the incidence node-arc matrix for graph G, and the matrix $A_{z\tau}$ defined as

$$A_{z\tau} = \begin{pmatrix} 1_{1\times Q_{z\tau}} & 1 & 1 & 0 \end{pmatrix}$$

Considering $\boldsymbol{q}_z = (q_{z\tau})_{\tau\in[0,T-1]}$, the objective cost function (5.9) can be rewritten as

$$\sum_{z\in Z} k_z(\boldsymbol{q}_z, v_z) + \sum_{e\in E}\sum_{\tau=0}^{T-1} l_{e\tau} f_{e\tau},$$

where the zonal cost of production K_z is equal to

$$k_z(\boldsymbol{q}_z, v_z) = \sum_{\tau=0}^{T-1}\left[\sum_{j=1}^{Q_{z\tau}} c^j_{z\tau}\theta^j_{z\tau} + c^{fail}\eta_{z\tau}\right] + c^{fin}_z v_z.$$

We introduce the set of zonal constraints associated to the production $\mathcal{C}_z$ which are the constraints (5.11)-(5.16) except (5.15) and the set of interzonal transfer constraint $\mathcal{C}_{e\tau}$ which is the constraint (5.15). Then the planning problem (5.9)-(5.16) now reads

$$\min_{(\boldsymbol{q}, v, f)} \quad \sum_{z\in Z} k_z(\boldsymbol{q}_z, v_z) + \sum_{e\in E}\sum_{\tau=0}^{T-1} l_{e\tau} f_{e\tau} \tag{5.17}$$

$$A_{z\tau}q_{z\tau} - B_z f_\tau = d_{z\tau}, \quad \forall (z,\tau) \in Z \times [0, T-1] \tag{5.18}$$

$$(\boldsymbol{q}_z, v_z) \in \mathcal{C}_z, \quad \forall z \in Z, \tag{5.19}$$

$$f_{e\tau} \in \mathcal{C}_{e\tau}, \quad \forall (e,\tau) \in E \times [0, T-1]. \tag{5.20}$$

Considering $f = (f_\tau)_{\tau\in[0,T-1]}$ and $d_z = (d_{z\tau})_{\tau\in[0,T-1]}$, relation ((5.18)) becomes

$$A_z\boldsymbol{q}_z - \bar{B}_z f = d_z, \quad \forall z \in Z \tag{5.21}$$

where $A_z = \mathrm{diag}\,(A_{z0}, \cdots, A_{z(T-1)})$ and $\bar{B}_z = \mathrm{diag}\,(B_z, \cdots, B_z)$. Then considering $\boldsymbol{w} = (\boldsymbol{q}_z, v_z)_{z\in Z}$ and $d = (d_z)_{z\in Z}$, we have that (5.21) becomes

$$A\boldsymbol{w} + \bar{B}f = d$$

where $A = \mathrm{diag}\,([A_{z_1}\, 0_{T\times 1}], \cdots, [A_{z_n}\, 0_{T\times 1}])$ and $\bar{B} = \begin{bmatrix} -\bar{B}_{z_1} \\ \vdots \\ -\bar{B}_{z_n} \end{bmatrix}$.

In summary, we rewrite (5.17)-(5.20) as

$$\min_{(\boldsymbol{w},\, f)} \quad \sum_{z\in Z} k_z(\boldsymbol{w}_z) + \sum_{e\in E}\sum_{\tau=0}^{T-1} l_{e\tau} f_{e\tau} \tag{5.22}$$

$$A\boldsymbol{w} + \bar{B} f = d \tag{5.23}$$

$$\boldsymbol{w}_z \in \mathcal{C}_z, \quad \forall z \in Z, \tag{5.24}$$

$$f_{e\tau} \in \mathcal{C}_{e\tau}, \quad \forall e \in E,\ \forall \tau \in [0, T-1]. \tag{5.25}$$

Since this problem has the same structure as problem (P), we apply GSS considering different parameters. In fact, we apply slight modifications of ADMM, Chambolle-Pock, and Spingarn algorithms, which correspond to the GSS algorithms with different parametrized matrices.

5.2.1 ADMM applied to the dynamic model

We apply the algorithm (2.41)-(2.43) with $V_1 = 0$, $V_2 = 0$ and $M = \lambda I$ (which is a slight variant of ADMM) to the problem (5.22)-(5.25) obtaining the algorithm

$$w^{k+1} \in \operatorname{argmin}_{w_z \in \mathcal{C}_z} \left\{ \sum_{z\in Z} k_z(\boldsymbol{w}_z) + \frac{\lambda}{2}\|Aw + \bar{B}f^k - d + \lambda^{-1} y^k\|^2 \right\} \tag{5.26}$$

$$f^{k+1} \in \operatorname{argmin}_{f_{e\tau} \in \mathcal{C}_{e\tau}} \left\{ \sum_{e\in E}\sum_{\tau=0}^{T-1} l_{e\tau} f_{e\tau} + \frac{\lambda}{2}\|Aw^{k+1} + \bar{B}f - d + \lambda^{-1} y^k\|^2 \right\} \tag{5.27}$$

$$y^{k+1} = y^k + \lambda(Aw^{k+1} + \bar{B}f^{k+1} - d) \tag{5.28}$$

Since

$$\|Aw + \bar{B}f^k - d + \lambda^{-1} y^k\|^2 = \sum_{z\in Z}\sum_{\tau=0}^{T-1} \|A_{z\tau} q_{z\tau} - B_z f_\tau - d_{z\tau} - \lambda^{-1} y^k_{z\tau}\|^2$$

the corresponding minimization problem in (5.26) can be solvable in parallel processing with respect to z indices. Similarly, the corresponding minimization problem in (5.27) can also be solvable in parallel processing with respect to τ indices.

Application 1

Step 1. Zonal subproblems

For all $z \in Z$ **do**

Find $q_z^{k+1} = ((\theta^j_{z\tau})^{k+1}_{j\in Q_{z\tau}}, u^{k+1}_{z\tau}, \eta^{k+1}_{z\tau}, x^{k+1}_{z,\tau+1})_{\tau\in[0,T-1]}$ and v_z^{k+1} solution of

$$\min_{(q_z,\, v_z)} \sum_{\tau=0}^{T-1} \left[\sum_{j=1}^{Q_{z\tau}} c^j_{z\tau}\theta^j_{z\tau} + c^{fail}\eta_{z\tau} + \frac{\lambda}{2}\| \sum_{j=1}^{Q_{z\tau}} \theta^j_{z\tau} + u_{z\tau} + \eta_{z\tau} + B_z f^k_\tau - d_{z\tau} + \lambda^{-1} y^k_{z\tau}\|^2 \right] + c^{fin}_z v_z$$

$$
\begin{aligned}
s.t \quad & x_{z,\tau+1} = x_{z\tau} - u_{z\tau} + i_{z\tau}, \quad \forall\ \tau \in [0, T-1] \\
& X_{z\tau}^{min} \leq x_{z\tau} \leq X_{z\tau}^{max}, \quad \forall\ \tau \in [0, T-1] \\
& 0 \leq u_{z\tau} \leq U_z^{max} \delta_h, \quad \forall\ \tau \in [0, T-1] \\
& 0 \leq \theta_{z\tau}^j \leq P_{z\tau}^j \delta_h, \quad \forall \tau \in [0, T-1] \\
& v_z \geq 0, \\
& v_z \geq (x_{z0} - x_{zT}),
\end{aligned}
$$

end for

Step 2. Network subproblem
For all $\tau \in [0, T-1]$ **do**
Find f_τ^{k+1} solution of

$$
\min_{f_\tau} \sum_{e \in E} l_{e\tau} f_{e\tau} + \frac{\lambda}{2} \sum_{z \in Z} \| \sum_{j=1}^{Q_{z\tau}} (\theta_{z\tau}^j)^{k+1} + u_{z\tau}^{k+1} + \eta_{z\tau}^{k+1} + B_z f_\tau - d_{z\tau} + \lambda^{-1} y_{z\tau}^k \|^2
$$

$$
s.t \quad 0 \leq f_{e\tau} \leq \kappa_{e\tau} \delta_h, \quad \forall e \in E.
$$

end for

Step 3. Dual update
For all $(z, \tau) \in Z \times [0, T-1]$ **do**

$$
y_{z\tau}^{k+1} = y_{z\tau}^k + \lambda (\sum_{j=1}^{Q_{z\tau}} (\theta_{z\tau}^j)^{k+1} + u_{z\tau}^{k+1} + \eta_{z\tau}^{k+1} + B_z f_\tau^{k+1} - d_{z\tau})
$$

end for

5.2.2 Chambolle-Pock applied to the dynamic model

Considering in algorithm (2.41)-(2.43) the positive parameters r_1, r_2 such that $1 \geq r_1 \lambda \|B\|^2$ and $1 \geq r_2 \lambda \|\bar{A}\|^2$, and the parameter matrices V_1, V_2 and M defined as

$$
V_1 = r_1 I - \lambda A^t A, \ \ V_2 = r_2 I - \lambda B^t B \quad \text{and} \quad M = \lambda I,
$$

we obtain a variant of Chambolle-Pock algorithm. This variant is applied to problem (5.22)-(5.25) obtaining the following algorithm

$$\begin{aligned} w^{k+1} &\in \operatorname{argmin}_{w_z \in \mathcal{C}_z} \left\{ \sum_{z \in Z} k_z(\boldsymbol{w}_z) + \frac{1}{2r_1} \|w - w^k + r_1 \lambda A^t (A w^k + \bar{B} f^k - d + \lambda^{-1} y^k)\|^2 \right\} \\ f^{k+1} &\in \operatorname{argmin}_{f_{e\tau} \in \mathcal{C}_{e\tau}} \left\{ \sum_{e \in E} \sum_{\tau=0}^{T-1} l_{e\tau} f_{e\tau} + \frac{1}{2r_2} \|f - f^k + r_2 \lambda \bar{B}^t (A w^{k+1} + \bar{B} f^k - d + \lambda^{-1} y^k)\|^2 \right\} \\ y^{k+1} &= y^k + \lambda (A w^{k+1} + \bar{B} f^{k+1} - d). \end{aligned}$$

Notice that each resultant sub-problem is a classical proximal step and can be solvable in parallel procesing with respect to z and $e\tau$ indices respectively.

Now, in order to get a more explicit form the last algorithm, we find explicit expressions of the norms of A^t and $\bar{B}^t$ (for this purpose we assume in the original model (5.9)–(5.16) that the graph is complete), and also explicit expressions of

$$A^t(A w^k + \bar{B} f^k - d + \lambda^{-1} y^k) \quad \text{and} \quad B^t(A w^{k+1} + \bar{B} f^k - d + \lambda^{-1} y^k).$$

For $g = ((g_{z\tau})_{\tau \in [0,T-1]})_{z \in Z}$, the following expressions hold

$$(A^t g)_z = \begin{pmatrix} \tilde{q}_z^t \\ 0 \end{pmatrix} \quad \text{and} \quad (\bar{B}^t g)_\tau = \sum_{z \in Z} g_{z\tau} B_z^t$$

where $\tilde{q}_z = (\tilde{q}_{z\tau})_{\tau \in [0,T-1]}$, with $\tilde{q}_{z\tau} = g_{z\tau} \begin{pmatrix} 1_{Q_{z\tau} \times 1} & 1 & 1 & 0 \end{pmatrix}$.

We now calculate the norms of A^t and $\bar{B}^t$. To calculate the norm of A^t, note that

$$A A^t = \operatorname{diag}(A_{z_1} A_{z_1}^t, \cdots, A_{z_n} A_{z_n}^t)$$

and $A_z A_z^t = \operatorname{diag}(Q_{z_1 0} + 2, \cdots, Q_{z_n (T-1)} + 2)$. Then,

$$\|A\|_2^2 = 2 + \max_{z\tau} \{Q_{z\tau}\}. \tag{5.29}$$

Since that in the original model (5.9)–(5.16) the graph is complete, then given n zones, the $n \times n(n-1)$ matrix B holds that

$$B B^* = 2n I_{n \times n} - 2(1_{n \times n}). \tag{5.30}$$

The following proposition shows some properties of matrix with this structure.

Proposition 5.2.1 *Set x, $y \in \mathbb{R}$. We consider the $n \times n$ matrix*

$$P = x(I_{n \times n}) + y(1_{n \times n}).$$

Then,

- *P has x and $x+ny$ as unique eigenvalues,*
- $\|P\|_2 = \sqrt{\max\{|x|, |x+ny|\}}$
- *if $x \notin \{0, -ny\}$, then P is invertible and*

$$P^{-1} = \frac{1}{x}(I_{n\times n}) - \frac{y}{x(x+ny)}(1_{n\times n}).$$

Since matrix $\bar{B}$ is a row permutation of the $nT\times n(n-1)T$ matrix diag $(B, \cdots, B)$, then $\|\bar{B}\|_2 = \|B\|_2$. Therefore, from (5.30) and the last proposition, we conclude that

$$\|\bar{B}\|_2^2 = 2n. \tag{5.31}$$

So, from (5.29) and (5.31) and choosing the parameters $r_1, r_2 \in I\!R$ satisfying

$$1 \geq r_1\lambda 2n \quad \text{and} \quad 1 \geq r_2\lambda(2 + \max_{z\tau}\{Q_{z\tau}\})$$

we get from the previous algorithm, the desired more explicit algorithm:

Application 2

Step 1. Zonal subproblems
For all $z \in Z$ **do**
calculated
For all $\tau \in \{0, \cdots, T-1\}$ **do**

$$b_\tau := (\sum_{j=1}^{Q_{z\tau}}(\theta_{z\tau}^j)^{k+1} + u_{z\tau}^{k+1} + \eta_{z\tau}^{k+1} + B_z f_\tau^{k+1} - d_{z\tau} + \lambda^{-1}y_{z\tau}^k) \begin{pmatrix} 1_{Q_{z\tau}\times 1} & 1 & 1 & 0 \end{pmatrix}$$

end for
Set $b = (b)_{\tau\in[0,T-1]}$
Find $q_z^{k+1} = ((\theta_{z\tau}^j)_{j\in Q_{z\tau}}^{k+1}, u_{z\tau}^{k+1}, \eta_{z\tau}^{k+1}, x_{z,\tau+1}^{k+1})_{\tau\in[0,T-1]}$ and v_z^{k+1} solution of

$$\min_{(q_z,v_z)} \sum_{\tau=0}^{T-1}\left[\sum_{j=1}^{Q_{z\tau}} c_{z\tau}^j\theta_{z\tau}^j + c^{fail}\eta_{z\tau}\right] + c_z^{fin}v_z + \frac{1}{2r_1}\|q_z - q_z^k + r_1\lambda b\|^2 + \frac{1}{2r_1}\|v_z - v_z^k\|^2$$

$$\begin{aligned} s.t \quad & x_{z,\tau+1} = x_{z\tau} - u_{z\tau} + i_{z\tau}, \quad \forall\ \tau \in [0, T-1] \\ & X_{z\tau}^{min} \leq x_{z\tau} \leq X_{z\tau}^{max}, \quad \forall\ \tau \in [0, T-1] \\ & 0 \leq u_{z\tau} \leq U_z^{max}\delta_h, \quad \forall\ \tau \in [0, T-1] \\ & 0 \leq \theta_{z\tau}^j \leq P_{z\tau}^j\delta_h, \quad \forall\tau \in [0, T-1] \\ & v_z \geq 0, \\ & v_z \geq (x_{z0} - x_{zT}), \end{aligned}$$

end for
Step 2. Network subproblem

For all $\tau \in \{0, \cdots, T-1\}$ **do**
calculated

$$a := \sum_{z \in Z} \left[\sum_{j=1}^{Q_{z\tau}} (\theta_{z\tau}^{j})^{k+1} + u_{z\tau}^{k+1} + \eta_{z\tau}^{k+1} + B_z f_\tau^k - d_{z\tau} + \lambda^{-1} y_{z\tau}^k \right] A_z^*$$

For all $e \in E$ **do**
Find $f_{e\tau}^{k+1}$ solution of

$$\min_{f_{e\tau}} \left[l_{e\tau} f_{e\tau} + \frac{1}{2r_2} \| f_{e\tau} - f_{e\tau}^k + r_2 \lambda a_e \|^2 \right]$$

$$s.t \qquad 0 \leq f_{e\tau} \leq \kappa_{e\tau} \delta_h.$$

end for
end for
Step 3. Dual update
For all $(z, \tau) \in Z \times [0, T-1]$ **do**

$$y_{z\tau}^{k+1} = y_{z\tau}^k + \lambda (\sum_{j=1}^{Q_{z\tau}} (\theta_{z\tau}^{j})^{k+1} + u_{z\tau}^{k+1} + \eta_{z\tau}^{k+1} + B_z f_\tau^{k+1} - d_{z\tau})$$

end for

5.2.3 PDA applied to the dynamic model

We reformulate the problem (5.22)-(5.25) and apply Spingarn method. Set

$$h(\boldsymbol{w}, f) = \sum_z k_z(\boldsymbol{q}_z, v_z) + \sum_{e \in E} \sum_{\tau=0}^{T-1} l_{e\tau} f_{e\tau} \ , \quad V = \{(\boldsymbol{w}, f) : A\boldsymbol{w} + \bar{B} f = 0\}$$

and also a' such that $(A \ \bar{B})a' = d$. Then, problem (5.22)-(5.25) can be set as

$$\begin{aligned} \min_{(\boldsymbol{w}, f)} \quad & h(\boldsymbol{w}, f) \\ & (\boldsymbol{w}, f) \in V + a' \\ & \boldsymbol{w}_z \in \mathcal{C}_z, \quad \forall z \in Z, \\ & f_{e\tau} \in \mathcal{C}_{e\tau}, \quad \forall e \in E, \, \tau \in [0, T-1]. \end{aligned}$$

Using the Spingarn's algorithm to that problem, we get

$$\begin{aligned}
w^{k+1} &\in \operatorname{argmin}_{w_z \in \mathcal{C}_z} \left\{ \sum_z k_z(w_z) + \frac{\lambda}{2} \|w - s^k + \lambda^{-1} d^k\|^2 \right\} \\
f^{k+1} &\in \operatorname{argmin}_{f_{e\tau} \in \mathcal{C}_{e\tau}} \left\{ \sum_{e \in E} \sum_{\tau=0}^{T-1} l_{e\tau} f_{e\tau} + \frac{\lambda}{2} \|f - g^k + \lambda^{-1} r^k\|^2 \right\} \\
(s^{k+1}, g^{k+1}) &= P_{V+a'}((w^{k+1}, f^{k+1}) + \lambda^{-1}(d^k, r^k)) \\
d^{k+1} &= d^k + \lambda(w^{k+1} - s^{k+1}) \\
r^{k+1} &= r^k + \lambda(f^{k+1} - g^{k+1})
\end{aligned}$$

In order to get a more explicit expression of this algorithm, we need to find an adequate manner to express the projection on the affine space $V + a'$. For these we assume that $Q_{z_1\tau} = \cdots = Q_{z_n\tau}$. Since

$$P_{V+a'} = a' - P_V(a') + P_V,$$

it suffices to determinate the projection over V. Set $\widehat{B} = \operatorname{diag}(B, \cdots, B)$ a permutation matrix D such that $D\bar{B} = \widehat{B}$, then

$$V = \{(\boldsymbol{w}, f) : DA\boldsymbol{w} + \widehat{B}f = 0\} \quad \text{and} \quad DAA^tD^t = \operatorname{diag}(\tilde{A}_0\tilde{A}_0^t, \cdots, \tilde{A}_{(T-1)}\tilde{A}_{(T-1)}^t),$$

where, for $\tau = 0, \cdots, T-1$, $\tilde{A}_\tau := \operatorname{diag}(A_{z_1\tau}, \cdots, A_{z_n\tau})$.

On the other hand, since V is the kernel of matrix $\begin{pmatrix} DA & \widehat{B} \end{pmatrix}$, one deduces that the projection $P_V(w, f)$ is equal to

$$(w - A^tD^t(DAA^tD^t + \widehat{B}\widehat{B}^t)^{-1}(DAw + \widehat{B}f),\ f - \widehat{B}^t(DAA^tD^t + \widehat{B}\widehat{B}^t)^{-1}(DAw + \widehat{B}f)).$$

To finish, we calculate the inverse of $DAA^tD^t + \widehat{B}\widehat{B}^t$. Since

$$DAA^tD^t + \widehat{B}\widehat{B}^t = \operatorname{diag}(\tilde{A}_0\tilde{A}_0^* + BB^*, \cdots, (\tilde{A}_{(T-1)}\tilde{A}_{(T-1)}^t, +BB^t),$$

it suffices to determinate the inverse of each diagonal block $\tilde{A}_\tau\tilde{A}_\tau^t + BB^t$ ($\tau = 0, \cdots, T-1$). From definition,

$$\tilde{A}_\tau\tilde{A}_\tau^t + BB^t = \operatorname{diag}(Q_{z_1\tau} + 2, \cdots, Q_{z_n\tau} + 2) + 2nI_{n\times n} - 2(1_{n\times n})$$

and, by considering $Q_\tau = Q_{z_1\tau} = \cdots = Q_{z_n\tau}$, we deduce, from Proposition 5.2.1, that

$$(\tilde{A}_\tau\tilde{A}_\tau^t + BB^t)^{-1} = \frac{1}{Q_\tau + 2n + 2} I_{n\times n} + \frac{2}{(Q_\tau + 2n + 2)(Q_\tau + 2)} 1_{n\times n}$$

and hence the desired explicit expression of the projection is deduced. In particular,

$$a' - P_V(a') = (A^t D^t b, \bar{B}^t D^t b)$$

where $b = (b_\tau)_{\tau\in\{0,\cdots,T-1\}}$ is defined as

$$b_\tau := \frac{1}{Q_\tau + 2n + 2} d_\tau + \frac{2\sum_{z\in Z} d_{z\tau}}{(Q_\tau + 2n + 2)(Q_\tau + 2)} 1_{1\times n}\,.$$

Therefore, the last previous algorithm can be set in the following context

Application 3

Step 1. Zonal subproblems

Find $q_z^{k+1} = (((\theta_{z\tau}^j)_{j\in Q_{z\tau}}^{k+1}, u_{z\tau}^{k+1}, \eta_{z\tau}^{k+1}))_{\tau\in[0,T-1]}$ and $p_z^{k+1} = (v_z^{k+1}, x_z^{k+1})$ solution of

$$\min_{(q_z,p_z)} \sum_{\tau=0}^{T-1} \left[\sum_{j=1}^{Q_{z\tau}} c_{z\tau}^j \theta_{z\tau}^j + c^{fail}\eta_{z\tau}\right] + c_z^{fin} v_z + \frac{\lambda}{2}\|(q_z, p_z) - s_z^k + \lambda^{-1} d_z^k\|^2$$

$$\begin{aligned} s.t \quad & x_{z,\tau+1} = x_{z\tau} - u_{z\tau} + i_{z\tau}, \quad \forall\ \tau \in [0, T-1] \\ & X_{z\tau}^{min} \le x_{z\tau} \le X_{z\tau}^{max}, \quad \forall\ \tau \in [0, T-1] \\ & 0 \le u_{z\tau} \le U_z^{max}\delta_h, \quad \forall\ \tau \in [0, T-1] \\ & 0 \le \theta_{z\tau}^j \le P_{z\tau}^j \delta_h, \quad \forall \tau \in [0, T-1] \\ & v_z \ge 0, \\ & v_z \ge (x_{z0} - x_{zT}), \end{aligned}$$

Step 2. Network subproblem

For all $(e, \tau) \in E \times \{0, \cdots, T-1\}$ **do**
Find $f_{e\tau}^{k+1}$ solution of

$$\min_{f_{e\tau}} \left[l_{e\tau} f_{e\tau} + \frac{\lambda}{2}\|f_{e\tau} - g_{e\tau}^k + \lambda^{-1} r_{e\tau}^k\|^2\right]$$

$$s.t \quad f_{e\tau} \in \mathcal{C}_{e\tau}.$$

end for

Step 3. Projection Step

calculate $g^{k+1} = (g_\tau^{k+1})_\tau$ and $s^{k+1} = (\bar{q}^{k+1}, \bar{p}^{k+1})$ where $\bar{q}_{z\tau}^{k+1} = ((\bar{\theta}_{z\tau}^j)_{j\in Q_{z\tau}}^{k+1}, \bar{u}_{z\tau}^{k+1}, \bar{\eta}_{z\tau}^{k+1})$ and $\bar{p}_z^{k+1} = (\bar{v}_z^{k+1}, \bar{x}_z^{k+1})$.

First calculate

$$c_{z\tau} = \sum_{j=1}^{Q_{z\tau}} (\theta_{z\tau}^j)^{k+1} + u_{z\tau}^{k+1} + \eta_{z\tau}^{k+1} + \lambda^{-1}(\sum_{j=1}^{Q_{z\tau}} (\widehat{\theta}_{z\tau}^j)^k + \widehat{u}_{z\tau}^k + \widehat{\eta}_{z\tau}^k) + B_z(f_\tau^{k+1} + \lambda^{-1} r_\tau^k)$$

then calculate

$$pp_{z\tau} := \frac{1}{Q_\tau + 2n + 2} c_{z\tau} + \frac{2\sum_{z\in Z} c_{z\tau}}{(Q_\tau + 2n + 2)(Q_\tau + 2)}$$

and finally calculate

$$(\bar{\theta}_{z\tau}^j)^{k+1} = (\theta_{z\tau}^j)^{k+1} + \lambda^{-1}(\widehat{\theta}_{z\tau}^j)^k - pp_{z\tau} + b_{z\tau}$$
$$\bar{u}_{z\tau}^{k+1} = u_{z\tau}^{k+1} + \lambda^{-1}\widehat{u}_z^k - pp_{z\tau} + b_{z\tau}$$
$$\bar{\eta}_{z\tau}^{k+1} = \eta_{z\tau}^{k+1} + \lambda^{-1}\widehat{\eta}_z^k - pp_{z\tau} + b_{z\tau}$$
$$\bar{v}_z^{k+1} = v_z^{k+1} + \lambda^{-1}\widehat{v}_z^k$$
$$\bar{x}_z^{k+1} = x_z^{k+1} + \lambda^{-1}\widehat{x}_z^k$$

and

$$g_\tau^{k+1} = f_\tau^{k+1} + \lambda^{-1} r_\tau^k - \sum_{z\in Z} B_z^t(pp_{z\tau} - b_{z\tau})$$

Step 4. Dual update

Calculate $r^{k+1} = ((r^{k+1})_{e\in E})_\tau$ and $d^{k+1} = (\widehat{q}^{k+1}, \widehat{p}^{k+1})$ where $\widehat{q}_{z\tau}^{k+1} := ((\widehat{\theta}_{z\tau}^j)_{j\in Q_{z\tau}}^{k+1}, \widehat{u}_{z\tau}^{k+1}, \widehat{\eta}_{z\tau}^{k+1})$ and $\bar{p}_z^{k+1} = (\widehat{v}_z^{k+1}, \widehat{x}_z^{k+1})$.

$$d^{k+1} = d^k + \lambda((q^{k+1}, p^{k+1}) - s^{k+1})$$

$$r^{k+1} = r^k + \lambda(f^{k+1} - g^{k+1})$$

5.3 Uncertainty Environment

The general case of stochastic production planning models has been studied by many authors and we will not detail the different discussions which are behind these models

when decomposition is the final objective to cope with the curse of dimensionality (see [46, 48, 3]).

Coming back to the stochastic model problem (5.1)-(5.8), we rewrite it in the context of problem (P).

Set $\boldsymbol{w}_{z\tau} = (\boldsymbol{p}_{z\tau}, \boldsymbol{u}_{z\tau}, \boldsymbol{x}_{z\tau})$, $\boldsymbol{w}_z = (\boldsymbol{w}_{z\tau})_{\tau\in[0,T-1]}$, $\boldsymbol{w} = (\boldsymbol{w}_{z\tau})_{z\in Z}$, $d = ((d_{z\tau})_{\tau\in[0,T-1]})_{z\in Z}$ and $\boldsymbol{f} = ((\boldsymbol{f}_{e\tau})_{e\in E})_{\tau\in[0,T-1]}$

$$\min_{(\boldsymbol{w},\, f)} \quad I\!E\left[\sum_{z\in Z}\left(\sum_{\tau=0}^{T-1} c_z(\boldsymbol{p}_{z\tau}) + \psi_z(\boldsymbol{x}_{zT})\right) + \sum_{\tau=0}^{T-1}\sum_{e\in E} l_e(\boldsymbol{f}_{e\tau})\right] \tag{5.32}$$

$$\widehat{A}\boldsymbol{w} + \widehat{B}\boldsymbol{f} = d \tag{5.33}$$

$$\boldsymbol{w}_z \in \mathcal{C}_z, \quad \forall z \in Z, \tag{5.34}$$

$$\boldsymbol{f}_{e\tau} \in \mathcal{C}_{e\tau}, \quad \forall e \in E,\ \forall \tau \in [0, T-1]. \tag{5.35}$$

$$\boldsymbol{w}_{z\tau} \preceq \mathcal{F}_\tau,\ \boldsymbol{f}_{e\tau} \preceq \mathcal{F}_\tau \ \ \forall z \in Z,\ \forall e \in E,\ \forall \tau \in [0, T-1] \tag{5.36}$$

considering matrices $\widehat{A}$ and $\widehat{B}$ defined as

$$\widehat{A} = \operatorname{diag}(A_{z_1}, \cdots, A_{z_n}) \ \text{ and } \ \widehat{B} = \begin{bmatrix} -\widehat{B}_{z_1} \\ \vdots \\ -\widehat{B}_{z_n} \end{bmatrix}$$

where $A_z = \operatorname{diag}([1\,1], \cdots, [1\,1])$ and $\widehat{B}_z = \operatorname{diag}(B_z, \cdots, B_z)$.

Notice that $\widehat{A}\boldsymbol{w}$ and $\widehat{B}\boldsymbol{f}$ in (5.33) are random vectors because w and f are so. They are defined by

$$\widehat{A}\boldsymbol{w}(\xi) = \widehat{A}(\boldsymbol{w}(\xi)) \quad \text{ and } \quad \widehat{B}\boldsymbol{f}(\xi) = \widehat{B}(\boldsymbol{f}(\xi)) \ \text{ for all } \ \xi \in \Xi.$$

We apply algorithm (2.41)-(2.43) with $V_1 = 0$, $V_2 = 0$ and $M = \lambda I$ (which is a slight variant of ADMM) to last problem, assuming the random variable space of finite dimension (finite scenarios) with inner product induced by the expectation function, getting the following algorithm.

Stochastic Application 1(SA1)

Step 1.
For all $z \in Z$ **do**

Find $w_z^{k+1} = (p_z^{k+1}, u_z^{k+1}, x^{k+1})$ a solution of

$$\begin{aligned}
\min_{(p,u)} \quad & I\!E\left[\sum_{\tau=0}^{T-1}\left(c_z(\boldsymbol{p}_{z\tau}) + \frac{\lambda}{2}\|\boldsymbol{p}_{z\tau} + \boldsymbol{u}_{z\tau} + B_z f_\tau^k - d_{z\tau} + \lambda^{-1} y_{z\tau}^k\|^2\right) + \psi_z(\boldsymbol{x}_{zT})\right] \\
& \boldsymbol{x}_{z,\tau+1} = \boldsymbol{x}_{z\tau} - \boldsymbol{u}_{z\tau} + i_{z\tau}, \quad \forall\, \tau \in [0, T-1] \\
& \boldsymbol{x}_{z\tau} \in X_{z\tau},\, \boldsymbol{u}_{z\tau} \in U_{z\tau},\, \boldsymbol{p}_{z\tau} \in P_{z\tau}, \quad \forall\, \tau \in [0, T-1] \\
& \boldsymbol{p}_{z\tau}, u_{z\tau} \preceq \mathcal{F}_\tau, \quad \forall\, \tau \in [0, T-1]
\end{aligned}$$

end for.

Step 2.
Find f^{k+1} a solution of

$$\begin{aligned}
\min_{f} \quad & I\!E\left[\sum_{\tau=0}^{T-1}\left(\sum_{e\in E} l_e(\boldsymbol{f}_{e\tau}) + \frac{\lambda}{2}\sum_{z\in Z}\|p_{z\tau}^{k+1} + u_{z\tau}^{k+1} + B_z \boldsymbol{f}_\tau - d_{z\tau} + \lambda^{-1} y_{z\tau}^k\|^2\right)\right] \\
& \boldsymbol{f}_{e\tau} \in F_{e\tau} \quad \forall e \in E,\ \forall\, \tau \in [0, T-1] \\
& \boldsymbol{f}_{e\tau} \preceq \mathcal{F}_\tau \quad \forall e \in E,\ \forall\, \tau \in [0, T-1]
\end{aligned}$$

Step 3. Dual update

$$y^{k+1} = y^k + \lambda(\widehat{A} w^{k+1} + \widehat{B} f^{k+1} - d).$$

Notice that the sub-problem of Step 2, can be solvable by the *progressing hedging* algorithm proposed by Rockafellar and Wets[46] because the only restriction coupling e and τ is the nonanticipativity constraint.

The sub-problem of Step 1 is a stochastic optimal control (SOC) which has less variables than original problem. But the white noise assumption over the random variable $(d_{z\tau}, i_{z\tau})_{\tau\in[0,T-1]}$ is not enough in order to apply dynamic programming, because we have two another families of random variables $(f_\tau^k)_{\tau\in[0,T-1]}$ and $(y_{z\tau}^k)_{\tau\in[0,T-1]}$ which from Step 3, are not independent over time, since $y_{z\tau}^k$ depends on $(d_{z\tau'})_{\tau'\in[0,\tau]}$. So we cant not solve the sub-problem of Step 1 directly by dynamic programming.

Following the same ideas presented in [33], we now reformulate problem (5.32)-(5.36) considering information relaxation in order to obtain a variant of sub-problem corresponding to Step 1 of previous algorithm (**SA1**) where we can apply DP for solve it.

For each $\tau \in [0, T-1]$ and $z \in Z$, we consider a random variable $U_{z\tau} \preceq \mathcal{F}_\tau$, then we consider the approximate version of our problem (5.32)-(5.36):

$$\min_{(\boldsymbol{w}, f)} \quad I\!E\left[\sum_{z\in Z}\left(\sum_{\tau=0}^{T-1} c_z(\boldsymbol{p}_{z\tau}) + \psi_z(\boldsymbol{x}_{zT})\right) + \sum_{\tau=0}^{T-1}\sum_{e\in E} l_e(\boldsymbol{f}_{e\tau})\right] \tag{5.37}$$

$$I\!E[\boldsymbol{p}_{z\tau} + \boldsymbol{u}_{z\tau} + B_z \boldsymbol{f}_\tau | U_{z\tau}] = I\!E[d_{z\tau}|U_{z\tau}], \quad \forall z \in Z, \forall \tau \in [0, T-1] \tag{5.38}$$

$$\boldsymbol{w}_z \in \mathcal{C}_z, \quad \forall z \in Z, \tag{5.39}$$

$$\boldsymbol{f}_{e\tau} \in \mathcal{C}_{e\tau}, \quad \forall e \in E, \ \forall \tau \in [0, T-1]. \tag{5.40}$$

$$\boldsymbol{w}_{z\tau} \preceq \mathcal{F}_\tau, \ \boldsymbol{f}_{e\tau} \preceq \mathcal{F}_\tau, \quad \forall z \in Z, \forall e \in E, \forall \tau \in [0, T-1]. \tag{5.41}$$

Set matrix $Q_{z\tau}$ ($\tau = 0, \cdots, T-1$ and $z \in Z = \{z_1, \cdots, z_n\}$) satisfying $Q_{z\tau} h = I\!E[h|U_{z\tau}]$, then restriction (5.38) become

$$Q\widehat{A}\boldsymbol{w} + Q\widehat{B}\boldsymbol{f} = Qd \tag{5.42}$$

where $Q = \text{diag}\,(Q_{z_1}, \cdots, Q_{z_n})$, with $Q_{z_i} = \text{diag}\,(Q_{z_i 0}, \cdots .Q_{z_i(T-1)})$.

Therefore, since (5.42) is a coupling linear constraint, similar to our original model (5.32)-(5.36), we get the following algorithm for solving (5.37)-(5.41):

Stochastic Application 2 (SA2)

Step 1.
For all $z \in Z$ **do**
Find $w_z^{k+1} = (p_z^{k+1}, u_z^{k+1}, x^{k+1})$ a solution of

$$\min_{(p,u)} \quad I\!E\left[\sum_{\tau=0}^{T-1} L(\boldsymbol{p}_{z\tau}, \boldsymbol{u}_{z\tau}, Q_{z\tau}B_z f_t^k, Q_{z\tau}y_{z\tau}^k, d_{z\tau}) + \psi_z(\boldsymbol{x}_{zT})\right]$$

$$\boldsymbol{x}_{z,\tau+1} = \boldsymbol{x}_{z\tau} - \boldsymbol{u}_{z\tau} + i_{z\tau}, \quad \forall\ \tau \in [0, T-1]$$

$$\boldsymbol{x}_{z\tau} \in X_{z\tau}, \ \boldsymbol{u}_{z\tau} \in U_{z\tau}, \ \boldsymbol{p}_{z\tau} \in P_{z\tau}, \quad \forall\ \tau \in [0, T-1]$$

$$\boldsymbol{p}_{z\tau}, u_{z\tau} \preceq \mathcal{F}_\tau, \quad \forall\ \tau \in [0, T-1]$$

where the function L is defined as

$$L(p, u, \tilde{f}, \tilde{y}, d) = c_z(p) + \langle Q_{z\tau}(p + u - d) + \tilde{f}), \tilde{y}\rangle + \frac{\lambda}{2}\|Q_{z\tau}(p + u - d) + \tilde{f}\|^2$$

end for.

Step 2.
Find f^{k+1} a solution of

$$\min_{Z\in\mathcal{L}} \quad I\!E\left[\sum_{\tau=0}^{T-1}\left(\sum_{e\in E} l_e(\boldsymbol{f}_{e\tau}) + \frac{\lambda}{2}\sum_{z\in Z}\|Q_{z\tau}(p_{z\tau}^{k+1} + u_{z\tau}^{k+1} + B_z\boldsymbol{f}_\tau - d_{z\tau}) + \lambda^{-1}y_{z\tau}^k\|^2\right)\right]$$

$$\boldsymbol{f}_{e\tau} \in F_{e\tau} \quad \forall e \in E,\ \forall\ \tau \in [0, T-1]$$

$$\boldsymbol{f}_{e\tau} \preceq \mathcal{F}_\tau \quad \forall e \in E,\ \forall\ \tau \in [0, T-1]$$

Step 3. Dual update

$$y^{k+1} = y^k + \lambda(Q\widehat{A}w^{k+1} + Q\widehat{B}f^{k+1} - Qd)$$

Choosing $U_{z\tau}$ equal to $(d_{z\tau}, i_{z\tau})$, we have that $Q_{z\tau}B_z f_t^k$ and $Q_{z\tau}y_{z\tau}^k$ are not noise, on the contrary are function of $(d_{z\tau}, i_{z\tau})$. Therefore we can apply DP for solve the sub-problem of Step 1.

In a future work we will try to apply the algorithm TSD, developed in Subsection 3.4.1 of Chapter 3, to sub-problem of Step 1 of algorithm **SA1**.

Conclusion

The contributions of this thesis are disseminated in the 5 chapters with different relative importance. Chapters 2 and 4 contain the main basic algorithms for two or more operators and they are nearly self-contained. In Chapter 2, the main point is the generalized splitting scheme which includes most of the classical primal-dual splitting methods as particular cases associated with the choices of the blocks of the matrix involved in a generalized proximal point method (a method constructed in this thesis) applied to the saddle-point inclusion problem.

The general setting shows us the relationship between the splitting algorithms and the fixed point algorithms corresponding to special average maps. This general setting also gives us a common point of view of the splitting and convergence properties of the classical primal-dual splitting methods deduced from this general setting, allowing us to improve them by adding for example multi-scaling parameters and a relaxed parameter.

The separable models for multi-block constrained optimization are studied in Chapter 3 and many new decomposition algorithms are derived with block separable augmented Lagrangian subproblems. One of these algorithms (SMS3BF) is applied to a stochastic model defined as (SP), splitting the nonanticipative constrains and the linear temporal coupling constraints.

In Chapter 4, a Lipschitz-differentiable function or its corresponding co-coercive map is added to the models proposed in Chapter 2. So we get extended version of this algorithm, where we add a Forward step correspond to that function in there formulation. Notice that, under mild assumptions, these extended algorithms inherit the properties of the generalized splitting schemes of Chapter 2.

Finally, in Chapter 5, some of the new algorithms considered in the thesis are applied to an applicative model, the stochastic multistage production planning problem with a limited set of numerical experiments based on randomly generated data sets. More is to be done to further validate the proposed algorithms considering the different coupling inherent to the model. An important open question is concerned with the tuning of the numerous parameters which influence these splitting methods. Even if the theoretical convergence rate analysis presented in Chapters 2 and 4 are not surprising, either in the ergodic or non ergodic sense, the numerical behavior of the splitting methods is still very sensitive to the choice of the scaling parameters, as already observed in the literature (see [19, 29]).

Bibliography

[1] M.A. Alghamdi, A. Alotaibi, P.L. Combettes, and N. Shahzad. A primal-dual method of partial inverses for composite inclusions. *Optimization Letters*, 8:2271–2284, 2014.

[2] H. Attouch and M. Soueycatt. Augmented lagrangian and proximal alternating direction method of multipliers in hilbert spaces. applications to games, pde's and control. *Pacific J. Optimization*, 5:17–37, 2008.

[3] K. Barty, P. Carpentier, and P. Girardeau. Decomposition of large scale stochastic optimal control problems. *RAIRO Operations Research*, pages 167–183, 2010.

[4] H.H. Bauschke and P.L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer V., 2011.

[5] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2:183–202, 2009.

[6] R.I. Bot and E.R. Csetnek. Admm for monotone operators : convergence analysis and rates. *Advances in Computational Mathematics*, 2017.

[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning with the alternating direction method of multipliers. In M. Jordan, editor, *Foundations and Trends in Machine Learning*, volume 3, pages 1–122. 2011.

[8] L.M. Briceno-Arias. Forward douglas-rachford splitting and forward partial inverse method for solving monotone inclusions. *Optimization*, 64:1239–1261, 2015.

[9] L.M. Briceno-Arias and P.L. Combettes. A monotone+skew splitting model for composite monotone inclusions in duality. *SIAM J. Optimization*, 21:1230–1250, 2011.

[10] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex programs with applications to imaging. *J. of Math. Imaging and Vision*, 40:1–26, 2011.

[11] Caihua Chen, Bingsheng He, Yinyu Ye, and Xiaoming Yuan. The direct extension of admm for multi block convex minimization problems is not necessarily convergent. *Math. Programming*, pages 1–23, 2014. Electronic access, doi=10.1007/s10107-014-0826-5.

[12] P.L. Combettes and J.C. Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian and paralle-sum type monotone operators. *Set Valued Var. Anal.*, 20:307–330, 2012.

[13] L. Condat. A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *J. Optimization Theory and Applications*, 158:460–479, 2013.

[14] A. Dallagi and A. Lenoir. A stochastic zonal decomposition algorithm for network energy management problems, 2013.

[15] D. Davis and W. Yin. Convergence rate analysis of several splitting schemes. In R. Glowinski, S.J. Osher, and W. Yin, editors, *Splitting Methods in Communication, Imaging, Science and Engineering*, pages 115–163. Springer International, 2016.

[16] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set Valued Var. Anal.*, 25(4):829–858, 2017.

[17] J. Douglas and H. H. Rachford. On the numerical solution of the heat conduction problem in 2 and 3 space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.

[18] J.P. Dussault, O.M. Gueye, and P. Mahey. Separable augmented lagrangian algorithm with multidimensional scaling for monotropic programming. *Journal of Optimization Theory and Application*, 127:329–345, 2005.

[19] J. Eckstein. *Splitting methods for monotone operators with applications to parallel optimization.* PhD thesis, Massachusetts institute of technology, cambridge, June 1989.

[20] J. Eckstein. A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers. *J. Optim. Theory and Appl.*, 173:155–182, 2017.

[21] J. Eckstein and D. P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55:293–318, 1992.

[22] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods : Application to numerical solutions of boundary-value problems*, volume 15 of *Studies in Mathematics and its Applications*, pages 299–331. North-Holland, 1983.

[23] P. Giselsson. Tight global linear convergence rate bounds for douglas rachford splitting. *Journal of Fixed Point Theory and Applications*, 19:2241–2270, 2017.

[24] R. Glowinski and A. Marocco. Sur l'approximation par éléments finis d'ordre 1 et la résolution par pénalisation-dualité d'une classe de problèmes de dirichlet. *RAIRO*, 2:41–76, 1975.

[25] Bing-Sheng He and Xiao-Ming Yuan. Alternating direction method of multipliers for linear programming. *Journal of the Operations Research Society of China*, 4:425–436, 2016.

[26] N. Komodakis and J.C. Pesquet. Playing with duality : an overview of recent primal-dual approaches for solving large-scale optimization problems. *SIAM J. Optim.*, 2014.

[27] P. Latafat and P. Patrinos. Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators. *Comput. Optim. Appl.*, 68:57–93, 2017.

[28] C. Lemaréchal. Lagrangean relaxation. In M. Jünger and D. Naddef, editors, *Computational Combinatorial Optimization*, volume 2241 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2001.

[29] A. Lenoir and P. Mahey. A survey of monotone operator splitting methods for the decomposition of convex programs. *RAIRO Oper. Res.*, 51:17–41, 2017.

[30] J. Liang, J. Fadili, and G. Peyré. Local linear convergence analysis of primal-dual splitting methods. *Optimization*, 67:821–853, 2018.

[31] P.L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16:964–979, 1979.

[32] P. Mahey, J.P. Dussault, A. Benchakroun, and A. Hamdi. Adaptive scaling and convergence rates of a separable augmented lagrangian algorithm. In

V.H. Nguyen, J.J. Strodiot, and P. Tossings, editors, *Optimization*, volume 481 of *Lecture Notes in Economics and Mathematical Systems*, pages 278–287. Springer, 2000.

[33] P. Mahey, J. Koko, and A. Lenoir. Decomposition methods for a spatial model for long-term energy pricing problems. *Math. Methods of Oper. Res*, 85:137–153, 2017.

[34] P. Mahey, S. Oualibouch, and D.T. Pham. Proximal decomposition on the graph of a maximal monotone operator. *SIAM J. Optimization*, 5:454–466, 1995.

[35] B. Martinet. Régularistion d'inéquations variationnelles par approximations successives. *Revue Française d'Informatique et de Recherche Opérationnelle*, pages 154–159, 1970.

[36] G.J. Minty. Monotone operators in hilbert spaces. *Duke Math. Journal*, 29:341–346, 1962.

[37] J.J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique Française*, 93:273–299, 1965.

[38] Y. Nesterov. Smooth minimization of nonsmooth functions. *Math. Programming Series A*, 103:127–152, 2005.

[39] E. Ocaña. *Un schéma de dualité pour les problèmes d'inéquations variationnelles*. PhD thesis, Variables complexes [math.CV]. Université Blaise Pascal - Clermont-Ferrand II, 2005.

[40] D. O'Connor and L. Vandenberghe. On the equivalence of the primal-dual hybrid gradient method and douglas-rachford splitting. *Mathematical Programming*, 2018.

[41] E. Oré, P. Mahey, and E. Ocaña. A unified splitting algorithm for composite monotone inclusions. Technical report, LIMOS, Université Clermont Auvergne, june 2018.

[42] T. Pennanen. A splitting method for composite mappings. *Num. Functional Anal. and Optim.*, 23:875–890, 2002.

[43] Rockafellar. *Convex Analyis*. Princeton University Press, 1970.

[44] R. T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

[45] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J.Control and optimization*, 14:877–898, 1976.

[46] R. T. Rockafellar and R. J-B Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16:119–147, 1991.

[47] R. Tyrell Rockafellar and Roger J-B Wets. *Variational Analyis*. Springer, 1998.

[48] A. Shapiro and A. Ruszczynski. *Stochastic Programming*. Elsevier, 2003.

[49] R. Shefi and M. Teboulle. Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM J. on Optimization*, 24:269–297, 2014.

[50] J.E. Spingarn. Partial inverse of a monotone operator. *Applied Mathematics and Optimization*, 10:247–265, 1983.

[51] J.E. Spingarn. Applications of the method of partial inverses to convex programming:decomposition. *Mathematical Programming*, 32:199–223, 1985.

[52] B.C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, 2013.

[53] R.S. Varga. *Matrix Iterative Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1966.

[54] X. Wang, M. Hong, S. Ma, and Z.Q. Luo. Solving multiple block separable convex minimization problems using two-block alternating direction method of multipliers. *Math. of Computation*, 2013.

[55] M. Yan. A new primal-dual algorithm for minimizing the sum of three functions with a linear operator. *J. Sci. Comput.*, 2018.