

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE CIENCIAS



TESIS

**“Una versión general de los métodos Punto
Proximal y Punto Fijo aplicados a una clase de
algoritmos de separación”**

PARA OBTENER EL TÍTULO PROFESIONAL EN
MATEMÁTICA

ELABORADA POR:
ERNESTO ORÉ ALBORNOZ

ASESOR:
ELADIO OCAÑA ANAYA

LIMA-PERÚ

2020

Dedicatoria

**A mi madre Aurelia,
y toda mi familia.**

AGRADECIMIENTOS

Agradezco a mi asesor, Profesor Dr. Eladio Ocaña Anaya, por su exigencia, paciencia, consejos y asesoramiento durante la elaboración de este y anteriores trabajos.

Agradezco al Instituto de Matemática y Ciencias Afines (IMCA) por la acogida, el apoyo científico e inmobiliario brindado.

Asimismo agradezco a mi Familia por el sustento y apoyo en todo momento.

RESUMEN

En este trabajo desarrollamos una variante Relative-error e Inertial-Relaxed de los Métodos Punto Proximal y Punto Fijo, para luego, aprovechando la relación que poseen ciertos algoritmos de separación con los Métodos antes mencionados, aplicar estos métodos más generales para la obtención de las variantes del mismo tipo en esa clase de algoritmos de separación, los cuales permiten su aceleración y/o fácil implementación.

Índice general

Introduction	1
1. Notaciones y resultados preliminares	3
1.1. Notaciones y definiciones básicas	3
1.2. Método del Punto Proximal y sus Variantes	5
1.3. Composición monótona	9
1.4. Composición monótona de dos bloques	10
1.5. Composición monótona más un operador co-coercivo	13
1.5.1. Algoritmo GSA3O, Form II	15
1.5.2. Algoritmo GSA3O, Form I	17
2. Algoritmos de Separación	19
2.1. Composición monotona generalizada de dos bloques	19
2.1.1. Versión Gauss-Seidel	22
2.1.2. Versión Jacobi	27
3. Variante del Método de Punto fijo y Punto Proximal	33
3.1. Variante del Método de Punto fijo	33
3.2. Variante del Método de Punto Proximal	37
4. Variantes Inertial-Relaxed	41
4.1. Composición monótona de dos bloques	41
4.2. Composición monótona generalizada de dos bloques	43
4.2.1. Variantes de los algoritmos de Condat-Vũ y PD3O	47
4.3. Resultados Numéricos	50
5. Variantes Relative-error Inertial-Relaxed	55
5.1. Composición monótona	55
5.2. Composición monótona más un operador co-coercivo	60

6. Debilitando las hipótesis matriciales de convergencia	65
6.1. Composición monótona de dos bloques	65
6.1.1. Versión Gauss-Seidel	66
6.1.2. Versión Jacobi	69
Conclusiones	74

Introduction

Los algoritmos de separación recientemente han adquirido mayor atención, esto debido a la creciente necesidad de solucionar problemas de grandes dimensiones, ya que los algoritmos de separación poseen la propiedad de descomponer el problema original (de estructura separable) en subproblemas más tratables numéricamente e iterativamente aproximándose a la solución del problema original.

En este trabajo consideramos el problema general de estructura separable, cuyo patron surge del modelamiento de problemas de planeación de energía [15] y tratamiento de imagenes [4] :

$$\min_{(x,z)} [f(x) + h_1(x) + g(z) + h_2(z) : Ax + Bz = 0] \quad (\mathcal{P})$$

donde $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ y $g : \mathbb{R}^p \mapsto \bar{\mathbb{R}}$ son funciones propias sci convexas, $h_1 : \mathbb{R}^n \mapsto \mathbb{R}$ y $h_2 : \mathbb{R}^p \mapsto \mathbb{R}$ son convexas, Lipschitz-diferenciables con coeficientes $(\frac{1}{\beta_1})$ y $(\frac{1}{\beta_2})$, respectivamente y, A y B son matrices de órdenes adecuados.

Considerando la formulación variacional y satisfaciendo ciertas condiciones de regularidad, los casos $h_1 = 0$, $h_2 = 0$ (problema de composición de dos bloques) y $h_2 = 0$, $B = -I_{p \times p}$ (problema de composición más una funcion diferenciable) asociados al problemas anterior son formulados como problemas de inclusión monótona bien estructuradas, de donde, a partir de estas reformulaciones, en [19] se logró establecer un marco unificador de una clase de algoritmos de separación clásicos y recientes, esto gracias a la aplicación y desarrollo de una variante del método proximal y del método de punto fijo aplicados a operadores especiales, respectivamente. Estos resultados serán resumidos en el primer capítulo de este trabajo.

Siguiendo el esquema de [19], se definiran dos tipos especiales de operadores cuyos puntos fijos estan asociados a las soluciones de (\mathcal{P}) , mostraremos que estos

operadores poseen propiedades de separación y son average con dominio total, estas propiedades nos permitirán obtener dos tipos de algoritmos de separación (distintos en el manejo de la solución de sus subproblemas siendo del tipo secuencial o paralelo) y su respectiva convergencia.

Por otro lado, en el reciente artículo de Alves-Eckstein [2], se obtuvo una variante Relative-error Inertial-Relaxed del método de punto proximal y, al ser el algoritmo ADMM (Alternating Direction Method of Multipliers, enunciado más adelante) una aplicación del método punto proximal, ver [9], esto les permitió a su vez utilizarlo para obtener una versión general de algoritmo ADMM, el cual admite la inclusión de los parámetros Inertial-Relaxed y el cálculo inexacto de uno de los subproblemas.

Por lo tanto, siguiendo la idea de Alves-Eckstein y la relación que poseen los métodos de separación con los métodos proximal y punto fijo, desarrollamos, en el Capítulo 3, una versión aún más general del método de punto fijo y proximal, para luego en los capítulos posteriores, aplicar estas nuevas variantes y así obtener versiones más generales de los algoritmos de separación.

Dado la importancia de los parámetros inercial y relajado en la aceleración de los algoritmos que los poseen, [3], consideraremos en conjunto, en el Capítulo 4, la inclusión de los parámetros inercial-relajado dentro de los algoritmos generales propuestos en el en el Capítulo 2 y en [19], que lidean con los problemas (\mathcal{P}) y el caso $h_1 = 0$, $h_2 = 0$, respectivamente, obteniendo como casos particulares las variantes inercial-relajado de una clase de algoritmos de separación conocidos en la literatura.

En el capítulo 5, obtendremos una generalización de la variante del *ADMM* propuesto por Alves-Eckstein [2], que considera el problema de minimizar la suma de dos funciones o equivalentemente el problema (\mathcal{P}) en el caso $h_1 = 0$, $h_2 = 0$, $A = I_{n \times n}$ y $B = -I_{n \times n}$. Alternativamente obtenemos un algoritmo general Relative-error Inertial-Relaxed, aplicado al problema (\mathcal{P}) en el caso $h_2 = 0$, $B = -I_{p \times p}$, considerando separadamente los casos $h_1 = 0$ y $h_1 \neq 0$.

Finalmente en el último capítulo mostraremos en algunos casos que es posible debilitar las hipótesis matriciales de convergencia de los algoritmos propuestos, aún cuando estas hipótesis provengan de la definición de los operadores que garantizan la convergencia, usaremos unas técnicas distintas para poder ampliar el dominio de convergencia.

Capítulo 1

Notaciones y resultados preliminares

1.1. Notaciones y definiciones básicas

En este trabajo utilizamos las siguientes notaciones y definiciones necesarias dentro del campo de la Optimización Convexa y del Análisis Variacional, todos ellos dentro del espacio finito dimensional \mathbb{R}^n .

Dada una matriz simétrica definida positiva U de orden $n \times n$, denotaremos la norma inducida por U como :

$$\|x\|_U^2 := \langle x, Ux \rangle \quad \text{para todo } x \in \mathbb{R}^n.$$

Dado un conjunto $C \subset \mathbb{R}^n$, denotaremos como $\text{cl}(C)$, $\text{int}(C)$ y $\text{ri}(C)$, la clausura, el interior y el interior relativo de C , respectivamente.

Dado un conjunto convexo $C \subset \mathbb{R}^n$, el cono normal de C en un punto $x \in C$, es el conjunto denotado por $\mathcal{N}_C(x)$ y definido como

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

asumiendo $\mathcal{N}_C(x) = \emptyset$ si $x \notin C$.

Denotaremos por $\overline{\mathbb{R}}$ al conjunto $[-\infty, +\infty]$. Una función $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ se dirá **convexa**, si su epigrafo,

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

es un conjunto convexo. La función f se dirá **semicontinua inferior**(sci) si su epigrafo es un conjunto cerrado.

Una función f se dirá **propia** si $f(x) > -\infty$ para todo $x \in \mathbb{R}^n$ y su dominio, definido por

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

es no vacío, o equivalentemente si su epigrafo es no vacío y no incluye una entera línea vertical.

La hipótesis de ser f convexa sci y propia es estándar en la literatura, lo cual permite obtener propiedades fundamentales como la monotonía maximal de su subgradiente asociado, estos conceptos son expuestos a continuación.

El operador subdiferencial que generaliza el concepto de la gradiente de una función (en el caso convexo), nos permite obtener caracterizaciones de los puntos optimales asociados a un problema de optimización aún cuando la función no sea diferenciable. El **subdiferencial** de f en un punto $x \in \mathbb{R}^n$ es el conjunto

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

Claramente $\partial f(x) = \emptyset$ si $x \notin \text{dom}(f)$ ó si f no es sci en x . En general, $\partial f(x)$ es convexo y cerrado, siendo tal vez vacío. Es sabido que es no vacío y acotado sobre $\text{int}(\text{dom}(f))$.

Una propiedad algebraica fundamental del subdiferencial es la monotonidad, esto es, para todo $x \in \partial f(x^*)$, $y \in \partial f(y^*)$ se cumple

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

En general, un operador $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ se dirá **monótono** si para todo $(x, x^*), (y, y^*)$ en el grafo de Γ [esto es el conjunto consistente de los pares $(z, z^*) \in \mathbb{R}^n \times \mathbb{R}^n$ tales que $z^* \in \Gamma(z)$] se cumple

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

Claramente si Γ es monótono, entonces su inversa Γ^{-1} definido como $\Gamma^{-1}(x^*) = \{x : x^* \in \Gamma(x)\}$ también lo es.

Otra propiedad analítica del subdiferencial (asociado a una función convexa sci y propia) es la maximalidad, lo cual se define como: un operador Γ se dirá **monótono maximal** si este es monótono y para cualquier operador monótono $\Sigma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfaciendo $\Gamma(x) \subset \Sigma(x)$ para todo $x \in \mathbb{R}^n$, se cumple que $\Gamma = \Sigma$. Se deduce que Γ es maximal monótono si y solo si Γ^{-1} es maximal monótono.

Una caracterización de la monotonia maximal es dada por el Teorema de Minty [16], el cual afirma que un operador monótono $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ es monótono maximal si y solo si el operador $(I + \Gamma)^{-1}$ es univaluado y tiene dominio total (ie, todo \mathbb{R}^n).

La propiedad de monotonia induce en el operador $(I + \Gamma)^{-1}$ una propiedad algebraica más fina que la monotonia y la lipschitzianidad, lo cual conduce a considerar la siguiente definición: un operador Γ se dirá **co-coercivo** con constante β (ó β -co-coercivo) si para todo $(x, x^*), (y, y^*)$ en el grafo de Γ , se cumple

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

Se deduce que si Γ es β -co-coercivo entonces Γ es univaluado y además β^{-1} -Lipschitz y monótono. En general se cumple que Γ es monótono si y solo si $(I + \Gamma)^{-1}$ es 1-co-coercivo. Más aún, del Teorema de Minty se cumple que Γ es monótono maximal si y solo si $(I + \Gamma)^{-1}$ es 1-co-coercivo (1-Lipschitz) con dominio total, mostrandonos así la importancia de $(I + \Gamma)^{-1}$.

Por otro lado, si F es 1-co-coercivo entonces es 1-Lipschitz (o noexpansiva). Más aún, se cumple que F es 1-co-coercivo si y solo si existe una función noexpansiva R tal que $F = \frac{1}{2}(I + R)$. Obteniendose una equivalencia entre los operadores monótonos y las funciones noexpansivas. Otro tipo de funciones que consideraremos son: Dado $\alpha \in]0, 1[$, una función F se dirá α -**average** si

$$F = (1 - \alpha)I + \alpha R,$$

donde R es una función noexpansiva.

1.2. Método del Punto Proximal y sus Variantes

Hallar los puntos optimales asociados a la minimización de una función es equivalente a resolver el problema de inclusión del operador subdiferencial asociado. El problema de inclusión abarca muchos problemas desde ecuaciones lineales hasta problemas variacionales [10], en particular consideraremos el problema de inclusión monótona

$$\text{Encontrar } x \in \mathbb{R}^r \text{ tal que } 0 \in \Gamma(x), \quad (V_\Gamma)$$

donde $\Gamma : \mathbb{R}^r \rightrightarrows \mathbb{R}^r$ es un operador monótono maximal. Denotemos por $\text{sol}(V_\Gamma)$ al conjunto solución del problema (V_Γ) (llamado también zeros de Γ).

El método proximal es un método clásico que lidia con el problema (V_Γ) , aún en su generalidad, sin considerar la buena estructura que pueda tener Γ , ver [23], y

es por ello que su aplicación directa produce algoritmos de difícil implementación, en contraste si es aplicado a las reformulaciones equivalentes de (V_Γ) (que saquen ventaja de su estructura), produce resultados interesantes. Por ejemplo, Rockafellar [22] aplica este método a la formulación variacional dual y Lagrangiano asociado a un problema de optimización, por otro lado, Eckstein [9], muestra que el método Douglas-Rachford (importante en los metodos de separación), resulta de la aplicación del método proximal a un operador especial.

El método proximal propone utilizar alternativamente el operador $J^\Gamma := (\Gamma + I_{r \times r})^{-1}$, llamado operador **resolvente** de Γ , cuyo conjunto de puntos fijos es igual al conjunto de zeros de Γ , luego al tener J^Γ dominio total, se le aplica el método de punto fijo, lo cual genera una secuencia $x^{k+1} = J^\Gamma(x^k)$ que converge por su propiedad adicional de co-coercividad, siempre que $\text{sol}(V_\Gamma)$ es no vacío. Cada iteración genera el siguiente subproblema: dado $x^k \in \mathbb{R}^r$

$$\text{Encontrar } x^{k+1} \in \mathbb{R}^r \text{ tal que } x^k - x^{k+1} \in \Gamma(x^{k+1}). \quad (V)$$

Este subproblema es la principal desventaja de este método, siendo en algunos casos tan complejo como el problema original. Solodov [20] reemplaza este subproblema por uno más tratable computacionalmente, dando origen al método híbrido proximal, en el cual nosotros consideraremos los parámetros fijos y sin el término enlargement de Γ , por simplicidad.

Relaxed Hybrid Proximal Projection (HPP)

Inicialización: Escoger $z^0 \in \mathbb{R}^r$, M una $r \times r$ matriz simétrica definida positiva, $c > 0$, $\theta \in [0, 1[$ y $\rho \in]0, 2[$.

Para $k = 0, 1, 2, \dots$ **hacer**

- Subproblema inexacto:

Encontrar $(\tilde{z}^k, v^k) \in \mathbb{R}^r \times \mathbb{R}^r$ tales que

$$v^k \in \Gamma(\tilde{z}^k), \quad \|cMv^k + \tilde{z}^k - z^k\|_{M^{-1}}^2 \leq \theta^2 \left(\|\tilde{z}^k - z^k\|_{M^{-1}}^2 + \|cMv^k\|_{M^{-1}}^2 \right). \quad (1.1)$$

- Si $v^k = 0$, **PARAR**. En caso contrario, escoger $\rho \in]0, \bar{\rho}]$ y actualizar

$$z^{k+1} = z^k - \rho \frac{\langle z^k - \tilde{z}^k, v^k \rangle}{\|Mv^k\|_{M^{-1}}^2} Mv^k.$$

Terminar el bucle.

Se observa que si $\theta = 0$, el subproblema (1.1) resulta: hallar $\tilde{z}^k = J^{cM\Gamma}(z^k)$, lo cual es el método proximal con parámetro multi-scaling cM y un parámetro de relajación ρ .

Otro tipo de variación del método proximal clásico fue la introducción del término inercial [1], lo cual produce mejoras numéricas. Alves-Eckstein [2] desarrollaron una extensión del método híbrido proximal de Solodov, permitiendo adicionarle el término inercial, logrando así un algoritmo que combina el término inercial y relajado a su vez que sólo exige resolver un subproblema inexacto. Debido a estas características, este algoritmo será usado de base para el desarrollo de este trabajo, donde consideramos los parámetros fijos por simplicidad.

Relative-error Inertial-Relaxed HPP (RIRHPP)

Inicialización: Escoger $z^0 = z^{-1} \in \mathbb{R}^r$, $c > 0$, $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ y $\theta \in [0, 1[$

Para $k = 0, 1, 2, \dots$ **hacer**

- Escoger $\lambda \in [0, \bar{\lambda}[$ y definir (término inercial)

$$w^k = z^k + \lambda(z^k - z^{k-1})$$

- Subproblema inexacto:
Encontrar $(\tilde{z}^k, v^k) \in \mathbb{R}^r \times \mathbb{R}^r$ tales que

$$v^k \in \Gamma(\tilde{z}^k), \quad \|cv^k + \tilde{z}^k - w^k\|^2 \leq \theta^2 \left(\|\tilde{z}^k - w^k\|^2 + \|cv^k\|^2 \right) \quad (1.2)$$

- Si $v^k = 0$, **PARAR**. En caso contrario, escoger $\rho \in]0, \bar{\rho}]$ y actualizar

$$z^{k+1} = w^k - \rho \frac{\langle w^k - \tilde{z}^k, v^k \rangle}{\|v^k\|^2} v^k.$$

Terminar el bucle.

Para obtener la convergencia se considera la siguiente relación entre $\bar{\lambda}$ y $\bar{\rho}$, la cual será una de las hipótesis de las proposiciones que desarrollaremos más adelante :

(H1) Se cumple que $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ y satisfacen la siguiente relación

$$\bar{\rho} = \frac{2(\bar{\lambda} - 1)^2}{2(\bar{\lambda} - 1)^2 + 3\bar{\lambda} - 1} \quad (1.3)$$

Obsérvese que estas relaciones son independientes de los parámetros θ y c asociados al subproblema (1.2).

El siguiente Teorema es una versión simplificada del Teorema 2.1 [2], que muestra la convergencia del algoritmo

Teorema 1.2.1 (Teorema 2.1 [2]) Dado $\Gamma : \mathbb{R}^r \rightrightarrows \mathbb{R}^r$ un operador monótono maximal tal que posee al menos un zero ($0 \in \Gamma(z^*)$). Sean z^k, \tilde{z}^k y v^k generados por el algoritmo RIRHPP. Si $\bar{\lambda}$ y $\bar{\rho}$ satisfacen (H1), entonces z^k y \tilde{z}^k convergen a un mismo zero de Γ y v^k converge a cero.

En [19] mostramos otro tipo de variación del método proximal, en el que extendemos la definición de operador resolvente por

$$J_P^\Gamma := (\Gamma + P)^{-1}P,$$

donde P es una $r \times r$ matriz simétrica semidefinida positiva. Esta ligera generalización es debido a la necesidad de considerar una matriz distinta de la identidad en el cálculo del resolvente, esto se observa en la obtención de algoritmos de separación, ver Capítulo 2 [19], por medio del uso de especiales matrices P . Además se considera un parámetro de relajación ρ . El algoritmo desarrollado lo renombramos como

Multi-scaling Proximal Point (MPP)

Inicialización: Escoger $z^0 \in \mathbb{R}^r$, P una $r \times r$ matriz simétrica semidefinida positiva y $\rho \in]0, 2[$

Para $k = 0, 1, 2, \dots$ **hacer**

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

Terminar el bucle.

Proposición 1.2.1 (Proposición 2.2.2 [19]) Sean $\Gamma : \mathbb{R}^r \rightrightarrows \mathbb{R}^r$ monótono maximal y P una $r \times r$ matriz simétrica semidefinida positiva. Si J_P^Γ es univariado (por lo tanto continua) con dominio cerrado y $\text{sol}(V_\Gamma)$ no vacío, entonces para $\rho \in (0, 2)$, la sucesión $\{x^k\}$ generada por el algoritmo MPP, converge a una solución de (V_Γ) .

El operador J_P^Γ no conserva en general la co-coercividad, pero al algoritmo anterior aún podemos asociarle el método de punto fijo respecto de un operador 1-co-coercivo. Considerando una matriz D de orden $q \times r$ satisfaciendo

$$P = D^t D, \tag{1.4}$$

la cual existe pues P es una matriz simétrica semidefinida positiva, consideramos el operador $G_D^\Gamma : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ definido como

$$G_D^\Gamma := D(\Gamma + D^t D)^{-1} D^t. \tag{1.5}$$

De la monotonía de Γ se sigue que G_D^Γ es 1-co-coercivo, además, como $DJ_P^\Gamma = G_D^\Gamma D$, la secuencia generada por el algoritmo anterior cumple que

$$Dw^{k+1} = \rho G_D^\Gamma(Dw^k) + (1 - \rho)Dw^k, \quad (1.6)$$

es decir, Dw^k es generado por el método de punto fijo con parámetro relajado ρ , aplicado al operador G_D^Γ que es 1-co-coercivo.

En lo que sigue mostraremos algunos importantes algoritmos de separación aplicados a importantes modelos de optimización, luego en los capítulos siguientes se obtendrán las variantes de estos algoritmos.

1.3. Composición monótona

El modelo de composición o **composite model**, es un problema de optimización definido como

$$\min_x f(x) + g(Ax). \quad (Model)$$

donde $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ y $g : \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$ son funciones convexas sci y propias, y A es una $m \times n$ matriz. Esta especial estructura que posee el modelo, se presenta en diversos problemas aplicativos tales como la eliminación de ruido en el tratamiento de imágenes, entre otros, es por ello su alto interés de estudio en los últimos años [4], [7].

El algoritmo Alternating direction method of multipliers (ADMM) propuesto por Glowinski [12], fue uno de los primeros algoritmos con propiedad de separación que lidia con $(Model)$. Este algoritmo surge de considerar una descomposición del tipo Gauss-Seidel al subproblema originado por el algoritmo lagrangiano aumentado:

$$\begin{array}{c} \text{ADMM} \\ \left\{ \begin{array}{l} x^{k+1} \in \operatorname{argmin} \left\{ f(x) + \frac{c}{2} \|Ax + Bz^k + c^{-1}y^k\|^2 \right\} \\ z^{k+1} \in \operatorname{argmin} \left\{ g(z) + \frac{c}{2} \|Ax^{k+1} + Bz + c^{-1}y^k\|^2 \right\} \\ y^{k+1} = y^k + c(Ax^{k+1} + Bz^{k+1}) \end{array} \right. \end{array}$$

donde $B = -I_{m \times m}$. El caso general cuando B es cualquier matriz, está asociado al modelo general (P) desarrollado en la siguiente subsección. En el mismo año, Gabay [11] mostró la equivalencia del ADMM con el método Douglas-Rachford (que es un método de separación para el problema de inclusión de la suma de dos operadores monótonos maximales). La desventaja del ADMM es que el primer subproblema es aún complejo en el caso $A^t A$ no sea múltiplo de la identidad. Un algoritmo alternativo

sin esta desventaja es presentado en los primeros trabajos de Chambolle-Pock [4]:

Primal-Dual algorithm (PD)

$$\begin{cases} y^{k+1} &= (I_{m \times m} + \sigma \partial g^*)^{-1}(y^k + \sigma A \eta^k) \\ x^{k+1} &= (I_{n \times n} + \tau \partial f)^{-1}(x^k - \tau A^t y^{k+1}) \\ \eta^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k) \end{cases}$$

donde cada subproblema sólo consiste en hallar el término proximal de las funciones f y g por separado, obteniendo una fácil implementación numérica, por ejemplo en problemas de regularización l_1 . Más tarde Condat [6] mostró que este algoritmo ($\theta = 1$), se puede obtener aplicando una variante del método punto proximal a la formulación variacional del lagrangiano, esto a su vez muestra que para obtener la convergencia, sólo es necesario usar la propiedad de monotonía y maximalidad de la subdiferencial de las funciones, permitiendo aplicar estos resultados al problema de inclusión de la **Composición monótona**, que es una generalización de la formulación variacional de problema (*Model*), definido como

$$\text{Encontrar } x \in \mathbb{R}^n \text{ tal que } 0 \in S(x) + A^t T(Ax), \quad (\text{CopMon})$$

donde $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ y $T : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ son monótonos maximales y A es una matriz de orden $m \times n$. Este problema de inclusión análogamente a la formulación variacional del problema de optimización (*Model*), posee un marco de dualidad [17],[21], obteniéndose el siguiente problema equivalente de inclusión Lagrangiana

$$\text{Encontrar } (\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ tal que } 0 \in \tilde{L}(\bar{x}, \bar{y}), \quad (V_{\tilde{L}})$$

donde \tilde{L} es el operador monótono maximal definido sobre $\mathbb{R}^n \times \mathbb{R}^m$ como

$$\tilde{L}(x, y) := \begin{pmatrix} S(x) \\ T^{-1}(y) \end{pmatrix} + \begin{pmatrix} 0 & A^t \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (1.7)$$

Para poder hallar otros algoritmos de separación, consideramos otro Lagrangiano con una variable adicional, el cual está asociado a un modelo más general como se muestra en la siguiente subsección.

1.4. Composición monótona de dos bloques

A continuación, en esta sección, se hará un resumen de los resultados necesarios para la elaboración de este trabajo, siendo estos recopilados del Capítulo 2 de la tesis [19]. Consideramos el problema de inclusión de la **Composición monótona**

de dos bloques, definido como

$$\text{Encontrar } (x, z) \in \mathbb{R}^n \times \mathbb{R}^p \text{ tal que } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} S(x) \\ T(z) \end{pmatrix} + \begin{pmatrix} A^t \\ B^t \end{pmatrix} \mathcal{N}_{\{0\}}(Ax + Bz) \quad (P_V)$$

donde $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ y $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ son monotonos maximales y, A y B matrices de orden $m \times n$ y $m \times p$, respectivamente. Es claro que este problema incluye al problema (*CopMon*) al considerar $B = -I$.

Este problema está relacionado con el siguiente problema de optimización

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

donde f y g son funciones convexas, sci y propias. Bajo condiciones de regularidad el problema de optimización (P) admite una formulación variacional del tipo (P_V), considerando $S = \partial f$ y $T = \partial g$. Un ejemplo de tal condición de regularidad es

$$\text{Existe } x \in \text{ri}(\text{dom } f) \text{ y } z \in \text{ri}(\text{dom } g) \text{ tales que } Ax + Bz = 0. \quad (H)$$

La aplicación directa del método proximal al problema (P_V) es inviable, pero considerando el esquema de dualidad en operadores [17],[21], se obtienen problemas equivalentes más tratables, en especial se obtiene la formulación Lagrangiana variacional

$$\text{Encontrar } (\bar{x}, \bar{z}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \text{ tal que } 0 \in L(\bar{x}, \bar{z}, \bar{y}), \quad (V_L)$$

donde L es el operador maximal definido sobre $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ como

$$L(x, z, y) := \begin{pmatrix} S(x) \\ T(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}. \quad (1.8)$$

Este operador es la suma de un operador monótono maximal (con estructura separable) y una matriz antisimetrica. Aprovechando esta estructura se considera una matriz P especial para el cálculo de la resolvente J_P^L y así considerando MPP , obtener el siguiente algoritmo separable (en el sentido que poseen subproblemas que involucran a S y T , separadamente).

Generalized Splitting Scheme (GSS)

$$z^{k+\frac{1}{2}} = V_2 z^k - B^t u^k \quad (1.9)$$

$$\tilde{z}^{k+1} = (T + V_2 + B^t M B)^{-1} [z^{k+\frac{1}{2}} - B^t M A x^k] \quad (1.10)$$

$$x^{k+\frac{1}{2}} = V_1 x^k - \gamma A^t M A x^k + (\gamma - 1) A^t M B z^k - A^t u^k \quad (1.11)$$

$$\tilde{x}^{k+1} = (S + V_1 + A^t M A)^{-1} [x^{k+\frac{1}{2}} - 2\gamma A^t M B \tilde{z}^{k+1}] \quad (1.12)$$

$$\tilde{u}^{k+1} = u^k + \gamma M A x^k + (1 - \gamma) M A \tilde{x}^{k+1} + M B \tilde{z}^{k+1} \quad (1.13)$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1 - \rho)(x^k, z^k, u^k) \quad (1.14)$$

Variando los parámetros matriciales y el scalar γ , recuperamos importantes algoritmos desarrollados con anterioridad en la literatura, en especial considerando $\rho = 1$, los parámetros $\gamma = 1$ (considerando un cambio de variable) y $\gamma = 0$ aplicados al problema (P), permiten obtener el siguiente algoritmo, que posee una interpretación como la descomposición del tipo Gauss-Seidel o Jacobi respectivamente del subproblema originado por el algoritmo lagrangiano aumentado proximal, permitiéndonos calcular la solución de los subproblemas de manera secuencial o paralela.

Proximal primal-dual Splitting, PPDS

$$\begin{cases} x^{k+1} \in \operatorname{argmin} \left\{ f(x) + \frac{1}{2} \|Ax + B\bar{z}^k + M^{-1}y^k\|_M^2 + \frac{1}{2} \|x - x^k\|_{V_1}^2 \right\} \\ z^{k+1} \in \operatorname{argmin} \left\{ g(z) + \frac{1}{2} \|A\eta^k + Bz + M^{-1}y^k\|_M^2 + \frac{1}{2} \|z - \bar{z}^k\|_{V_2}^2 \right\} \\ y^{k+1} = y^k + M(Ax^{k+1} + Bz^{k+1}) \end{cases}$$

según como escojamos η^k obtenemos las dos distintas versiones del algoritmo

$$\eta^k := \begin{cases} x^k & \text{para la version Jacobi} \\ x^{k+1} & \text{para la version Gauss-Seidel.} \end{cases}$$

considerando $B = -I$, $M = cI$ corresponden a los algoritmos presentados por Shefi-Teboulle [24], que a su vez contienen los algoritmos *ADMM* y *PD*, obteniendo así un enfoque unificador a los algoritmos de separación o **Splitting algorithms** (en inglés).

Con la finalidad de mostrar la construcción y convergencia de GSS, consideraremos que la matriz simétrica P , involucrada en J_P^L , de tal modo que su cálculo implique la resolución de subproblemas asociado a S y T separadamente, posee la siguiente estructura : Dados las matrices simétricas V_1 y V_2 , los reales α y β tales que $\gamma = \alpha - \beta$,

$$P := \begin{pmatrix} V_1 + (\alpha + 1)^2 A^t M A & \alpha(2 + \beta) A^t M B & (1 + \alpha) A^t \\ \alpha(2 + \beta) B^t M A & V_2 + (\beta + 1)^2 A^t M B & (1 + \beta) B^t \\ (1 + \alpha) A & (1 + \beta) B & M^{-1} \end{pmatrix} \quad (1.15)$$

Cualquiera de las siguientes condiciones matriciales garantizan que la matriz P sea simétrica semidefinida positiva.

A1' $V_1 - (\gamma - 1)^2 A^t M A$ y $N V_2 - B^t M B$ son semidefinidas positivas.

A2' $V_1 - A^t M A$ y $V_2 - (\gamma - 1)^2 B^t M B$ son semidefinidas positivas.

A3' $\gamma \geq 1$ y las matrices $V_1 - (\gamma - 1)A^tMA$ y $V_2 - (\gamma - 1)B^tMB$ son semidefinidas positivas.

A4' $\gamma = 1$ y las matrices V_1 y V_2 son semidefinidas positivas.

Además consideraremos que las matrices $V_1 + A^tMA$ y $V_2 + B^tMB$ son definidas positivas, para garantizar que J_P^L sea univaluado. Luego a partir de la Proposición 1.2.1, se obtiene la siguiente proposición de convergencia asociado a *GSS*.

Proposición 1.4.1 (Proposición 2.3.3 [19]) Sean $\rho \in (0, 2)$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ todas simétricas, con M definida positiva tal que $V_1 + A^tMA$ y $V_2 + B^tMB$ son definidas positivas, y $\gamma \in \mathbb{R}$ tales que una de las condiciones (A1') – (A4') es satisfecha. Si $\text{sol}(V_L)$ es no vacío, entonces para un $(x^0, z^0, u^0) \in \mathbb{R}^{n+p+m}$ arbitrario, la secuencia (x^k, z^k, u^k) en (1.9)-(1.14) converge a una solución de (V_L) .

La relación exacta de este algoritmo *GSS* con el método *MPP* es la siguiente: Considerando y^k e \tilde{y}^{k+1} definidos por $y^k = u^k + (\alpha - \gamma + 1)MAx^k + (1 + \beta)MBz^k$ e $\tilde{y}^{k+1} = \tilde{u}^{k+1} + (\alpha - \gamma + 1)MA\tilde{x}^{k+1} + (1 + \beta)MB\tilde{z}^{k+1}$, se cumple que la secuencia $w^k = (x^k, z^k, y^k)$ proviene del método *MPP* :

$$J_P^L w^k = (\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) \quad \text{y} \quad w^{k+1} = \rho J_P^L w^k + (1 - \rho)w^k. \quad (1.16)$$

En particular, evaluando la última secuencia por una matriz D tal que $D^tD = P$, se obtiene que los términos de la secuencia $\{w^k\}$ están relacionados a través de la aplicación del método de punto fijo con parámetro de relajación ρ ,

$$Dw^{k+1} = \rho G_D^L(Dw^k) + (1 - \rho)Dw^k, \quad (1.17)$$

donde el operador G_D^L (ver definición (1.5)) es 1-co-coercivo con dominio total, que se deduce de la monotonía y maximalidad de L .

Por lo tanto el algoritmo *GSS* está asociado al algoritmo de punto fijo aplicado al operador G_D^L que es $\frac{1}{2}$ -average con dominio total, en similitud al algoritmo *ADMM* que está relacionado al operador *Douglas-Rachford*.

1.5. Composición monótona más un operador co-coercivo

Consideramos otro tipo de extensión del problema (*CopMon*), esto es, el problema **Composición monótona más un operador co-coercivo**, definido como

$$0 \in S(x) + A^tT(Ax) + C(x), \quad (Var)$$

donde $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ y $T : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ son monotonos maximales, $C : \mathbb{R}^n \rightarrow \mathbb{R}^n$ un operador β -co-coercivo con dominio total y A una $m \times n$ matriz. Cuando $C = 0$, el problema (*Var*) se convierte en el problema de inclusión composición monótona (*CopMon*). El caso general (*Var*) también puede ser visto como un problema de composición monótona, considerando la suma de S y C como un único operador, obteniéndose así su formulación lagrangiana variacional,

$$\text{Encontrar } (\bar{x}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ tal que } 0 \in \widehat{L}(\bar{x}, \bar{y}), \quad (V_{\widehat{L}})$$

donde \widehat{L} es el operador monótono maximal definido sobre $\mathbb{R}^n \times \mathbb{R}^m$ como

$$\widehat{L}(x, y) := \begin{pmatrix} S(x) + C(x) \\ T^{-1}(y) \end{pmatrix} + \begin{pmatrix} 0 & A^t \\ -A & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \quad (1.18)$$

Contrariamente a lo obtenido en el problema (V_L), la aplicación de *MPP* a este operador no produce la total separación deseada, principalmente entre C y S .

La estructura del problema (*Var*) está relacionada a la formulación variacional del siguiente problema de optimización

$$\min_x f(x) + g(Ax) + h(x), \quad (1.19)$$

donde $f : \mathbb{R}^n \mapsto \bar{\mathbb{R}}$ y $g : \mathbb{R}^m \mapsto \bar{\mathbb{R}}$ son funciones propias sci convexas, $h : \mathbb{R}^n \mapsto \mathbb{R}$ convexo y $(\frac{1}{\beta})$ -Lipschitz-diferenciable, y A una $m \times n$ matriz. Condat [6] desarrolló dos formas de algoritmos considerando dos diferentes matrices multi-scaling en el método Forward-backward aplicados a su problema de inclusión Lagrangiano ($V_{\widehat{L}}$), obteniendo algoritmos de separación que consdieran pasos proximales para f y g y, su correspondiente evaluación de la gradiente para h .

Condat–Vũ Algorithm, Form I

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

y su versión **switching**, donde se intercambia el orden de los subproblemas

Condat–Vũ Algorithm, Form II

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

Más tarde Chambolle-Pock [5], mostró la rapidez de convergencia de orden $O(1/k)$ y desarrolló una variación inercial del algoritmo Condat-Vũ, forma I (considerando $\rho = 1$)

Inertial Chambolle-Pock Algorithm

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

Otra metodología para lidiar con el problema (1.19), fue presentada recientemente por Yan [25], donde considera un operador average construido a partir del operador Davis-Yin, de donde obtiene el algoritmo *PD3O*, el cual posee restricciones sobre los parámetros, más débiles que los exigidos por los algoritmos de Condat-Vũ y muestra mejores resultados numéricos. Una relación más estricta del método de Davis-Yin con el algoritmo *PD3O* es estudiado por O' Connor [18] donde mostró que después de un cambio de variable, el algoritmo *PD3O* se convierte en la aplicación directa del método Davis-Yin [8] asociado a la suma de tres operadores especialmente contruidos siendo estos monótonos maximales donde uno de ellos es co-coercivo. A esta ligera variación lo seguiremos llamando *PD3O*.

Primal-Dual Three-Operator splitting (PD3O)

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

El problema de encontrar un zero de la suma de tres operadores monotonos maximales donde uno de ellos es co-coercivo, está incluido en el problema (*Var*), al considerar A igual a la matriz identidad. A este problema, Davis-Yin [8], le asocia equivalentemente el problema de encontrar un punto fijo del operador Davis-Yin definido como

$$\mathcal{G} := I - J^{\lambda T} + J^{\lambda S}(2J^{\lambda T} - I - \lambda C(J^{\lambda T})), \quad (1.20)$$

el cual es $\frac{2\beta}{4\beta-\lambda}$ -average (siempre que $\lambda < 2\beta$) con dominio total, propiedad que permite la convergencia del algoritmo de punto fijo aplicado a este. El operador \mathcal{G} resulta ser una extensión del operador Douglas-Rachford (caso $C = 0$) y de los operadores correspondientes a los métodos Forward-Backward y Backward-Forward. En la siguiente sección se muestra la extensión del operador \mathcal{G} y su relación con la construcción de algoritmos de separación para el problema general (*Var*).

1.5.1. Algoritmo GSA3O, Form II

A continuación se hará un resumen de los resultados de [19], Capítulo 4 :

Iniciando con el caso donde la matriz A es inyectiva, se contruye un lagrangiano especial que permite la separación deseada al aplicarle el método MPP , luego reescribiendo el problema original (Var) manteniendo su estructura original pero siendo su matriz asociada inyectiva, y aplicando el resultado previo, se obtienen dos operadores average asociados a dos tipos de algoritmos splitting siendo uno la version switching del otro en similitud con los algoritmos Condat-Vũ.

El primer tipo de algoritmo es como sigue, al cual le adicionamos el término form II, en concordancia con los algoritmos Condat-Vũ.

**Generalized splitting algorithm for three operators
(GSA3O, Form II)**

$$\tilde{z}^k = (T + M)^{-1}(y^k + MAx^k) \quad (1.21)$$

$$\tilde{y}^k = y^k + MAx^k - M\tilde{z}^k \quad (1.22)$$

$$r^k = C((V + A^tMA)^{-1}(Vx^k + A^tM\tilde{z}^k)) \quad (1.23)$$

$$\tilde{x}^k = (S + V + A^tMA)^{-1}(Vx^k + A^tM\tilde{z}^k - A^t\tilde{y}^k - r^k) \quad (1.24)$$

$$(x^{k+1}, y^{k+1}) = \rho(\tilde{x}^k, \tilde{y}^k) + (1 - \rho)(x^k, y^k). \quad (1.25)$$

Se obtiene la siguiente proposición de convergencia.

Proposición 1.5.1 (Proposición 4.2.1 [19]) Sean $V \in \mathbb{R}^{n \times n}$ y $M \in \mathbb{R}^{m \times m}$ simétricas, con V semidefinida positiva y M definida positiva, tales que $V + A^tMA$ es definida positiva satisfaciendo $\|(V + A^tMA)^{-1}\| \in]0, 2\beta[$. Sea $\rho \in]0, \alpha^{-1}[$ donde

$$\alpha := \frac{2\beta}{4\beta - \|(V + A^tMA)^{-1}\|}.$$

Si $\text{sol}(Var)$ es no vacío, entonces dado un punto arbitrario $(x^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m$, la secuencia (x^k, y^k) en (1.21)-(1.25) converge a una solución de $(V_{\hat{L}})$.

Análogamente a (1.17), este algoritmo está relacionado con el método de punto fijo aplicado a un operador, en este caso, α -average. Esto es, considerando $\zeta^k := (x^k, y^k)$ generado por $GSA3O$, tenemos

$$\widehat{Q}\zeta^{k+1} = \rho G_1(\widehat{Q}\zeta^k) + (1 - \rho)\widehat{Q}\zeta^k, \quad (1.26)$$

donde la matriz \widehat{Q} está definida por

$$\widehat{Q} = \begin{pmatrix} M^{\frac{1}{2}}A & M^{-\frac{1}{2}} \\ V^{\frac{1}{2}} & 0 \end{pmatrix}, \quad (1.27)$$

y el operador G_1 definido de $\mathbb{R}^m \times \mathbb{R}^n$ en si mismo es, en (u, x) ,

$$\begin{pmatrix} u - \tilde{J}^T u + M^{\frac{1}{2}} A \tilde{J}^S \left[V^{\frac{1}{2}} x + A^t M^{\frac{1}{2}} (2\tilde{J}^T u - u) - \tilde{C} (V^{\frac{1}{2}} x + A^t M^{\frac{1}{2}} \tilde{J}^T u) \right] \\ V^{\frac{1}{2}} \tilde{J}^S \left[V^{\frac{1}{2}} x + A^t M^{\frac{1}{2}} (2\tilde{J}^T u - u) - \tilde{C} (V^{\frac{1}{2}} x + A^t M^{\frac{1}{2}} \tilde{J}^T u) \right] \end{pmatrix} \quad (1.28)$$

donde

$$\tilde{J}^T = M^{\frac{1}{2}} (T + M)^{-1} M^{\frac{1}{2}}, \quad \tilde{J}^S = (S + V + A^t M A)^{-1} \quad \text{y} \quad \tilde{C} = C (A^t M A + V)^{-1}.$$

Este operador no es 1-co-coercivo pero es α -average con dominio total (donde α es definido como en la proposición anterior), siempre que, $\|(V + A^t M A)^{-1}\| < 2\beta$. Además, su correspondiente conjunto de puntos fijos está relacionado con $\text{sol}(V_{\hat{L}})$, más precisamente, es igual al conjunto

$$\widehat{Q}(\text{sol}(V_{\hat{L}})) := \{(M^{\frac{1}{2}} A \bar{x} + M^{-\frac{1}{2}} \bar{y}, V^{\frac{1}{2}} \bar{x}) : -A^t \bar{y} \in S(\bar{x}) + C(\bar{x}), \bar{y} \in T(A \bar{x})\}. \quad (1.29)$$

1.5.2. Algoritmo GSA3O, Form I

De forma similar al algoritmo Condat-Vũ que posee una versión switching, se obtiene también la versión switching del algoritmo *GSA3O*

GSA3O, Form I

$$\tilde{x}^{k+1} = (S + A^t M A + V)^{-1} (V x^k + A^t M z^k - A^t y^k - V r^k) \quad (1.30)$$

$$\tilde{y}^{k+1} = y^k + M A \tilde{x}^{k+1} - M z^k \quad (1.31)$$

$$\tilde{r}^{k+1} = (A^t M A + V)^{-1} C(\tilde{x}^{k+1}) \quad (1.32)$$

$$\tilde{z}^{k+1} = (T + M)^{-1} (M A \tilde{x}^{k+1} + \tilde{y}^{k+1} - M A \tilde{r}^{k+1}) \quad (1.33)$$

$$(x^{k+1}, z^{k+1}, y^{k+1}, r^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}, \tilde{r}^{k+1}) + (1 - \rho)(x^k, z^k, y^k, r^k) \quad (1.34)$$

Se obtienen similares resultados a *GSA3O*, relacionándolo con el método de punto fijo aplicado a una función α -average : Considerando $\{\xi^k = (x^k - r^k, z^k, y^k)\}$, se cumple que

$$\bar{Q} \xi^{k+1} = \rho G_2(\bar{Q} \xi^k) + (1 - \rho) \bar{Q} \xi^k, \quad (1.35)$$

donde la matriz \bar{Q} es igual a

$$\bar{Q} = \begin{pmatrix} 0 & M^{\frac{1}{2}} & -M^{-\frac{1}{2}} \\ V^{\frac{1}{2}} & 0 & 0 \end{pmatrix},$$

y G_2 es un operador de $\mathbb{R}^m \times \mathbb{R}^n$ en si mismo y cuyo valor en (u, x) es

$$\begin{pmatrix} u - M^{\frac{1}{2}} A \widehat{J}^S(x, u) + \tilde{J}^T \left[2M^{\frac{1}{2}} A \widehat{J}^S(x, u) - u - M^{\frac{1}{2}} A \widehat{C}(\widehat{J}^S(x, u)) \right] \\ V^{\frac{1}{2}} \widehat{J}^S(x, u) - V^{\frac{1}{2}} \widehat{C}(\widehat{J}^S(x, u)) \end{pmatrix}, \quad (1.36)$$

donde

$$\tilde{J}^T = M^{\frac{1}{2}}(T + M)^{-1}M^{\frac{1}{2}}, \quad \hat{J}^S = (S + V + A^tMA)^{-1} \circ \begin{pmatrix} V^{\frac{1}{2}} & A^tM^{\frac{1}{2}} \end{pmatrix}$$

y

$$\hat{C} = (A^tMA + V)^{-1}C.$$

El operador G_2 definido en (1.36) es también α -average con dominio total, desde que $\|(V + A^tMA)^{-1}\| < 2\beta$, siendo su conjunto de puntos fijos igual a

$$\{(M^{\frac{1}{2}}A[\bar{x} - WC(\bar{x})] - M^{-\frac{1}{2}}\bar{y}, V^{\frac{1}{2}}[\bar{x} - WC(\bar{x})]) : -A^t\bar{y} \in S(\bar{x}) + C(\bar{x}), \bar{y} \in T(A\bar{x})\}, \quad (1.37)$$

donde $W = (A^tMA + V)^{-1}$.

Observación 1.5.1 *En el caso que A sea una matriz inyectiva, podemos considerar la matriz parámetro $V = 0$, lo cual permite la restricción de los operadores G_1 y G_2 al espacio \mathbb{R}^m , luego en tal caso, al considerar $A = I$ y $M = \lambda^{-1}I$, reobtenemos en ambas restricciones respectivamente las dos versiones del operador de Davis-Yin, intercambiando el orden de S y T (ver (1.20)).*

Recíprocamente, podemos obtener los operadores G_1 y G_2 a partir del operador de Davis-Yin, aplicándolo a reformulaciones especiales del problema (Var).

Capítulo 2

Algoritmos de Separación

Este Capítulo estudia un modelo más general que incluye a los problemas (P_V) y (Var) (presentados en el capítulo anterior), obteniendo operadores average cuyos puntos fijos están relacionados a la solución de este modelo, luego a partir de estos operadores y su aplicación del método de punto fijo, obtenemos dos tipos de algoritmos de separación, extendiendo los dos tipos de algoritmos considerados en PPDS. El modelo extendido planteado, permite considerar un solo operador average que incluye a ambos operadores average G_1 y G_2 (definidos en 1.28 y 1.36) que a su vez induce un algoritmo secuencial extendiendo los algoritmos GSA3O y la versión Gauss-Seidel del algoritmo PPDS. Además obtenemos otro tipo de operador average asociado a un algoritmo paralelo que extiende la versión Jacobi del algoritmo PPDS. En suma, definiremos dos tipos de operadores average y sus algoritmos de separación asociados: tipo Gauss-Seidel y tipo Jacobi, respectivamente.

2.1. Composición monótona generalizada de dos bloques

Consideramos el problema **Composición monótona generalizada de dos bloques** definido como

$$\text{Encontrar } (x, z) \in \mathbb{R}^n \times \mathbb{R}^p \text{ tal que } \begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} S(x) + C_1(x) \\ T(z) + C_2(z) \end{pmatrix} + \begin{pmatrix} A^t \\ B^t \end{pmatrix} \mathcal{N}_{\{0\}}(Ax + Bz) \quad (\mathcal{V})$$

donde usamos las mismas condiciones sobre los operadores del problema (Var) , adicionando la condición sobre los operadores C_1 y C_2 que son operadores de dominio total y co-coercivos de coeficientes β_1 y β_2 respectivamente. Recíprocamente considerando la suma de S y C_1 , y la suma de T y C_2 como únicos operadores, obtenemos

su formulación lagrangiana

$$\text{Encontrar } (\bar{x}, \bar{z}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \text{ tal que } 0 \in \mathcal{L}(\bar{x}, \bar{z}, \bar{y}), \quad (V_{\mathcal{L}})$$

donde \mathcal{L} es el operador monótono maximal definido sobre $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ como

$$\mathcal{L}(x, z, y) := \begin{pmatrix} S(x) + C_1(x) \\ T(z) + C_2(z) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & A^t \\ 0 & 0 & B^t \\ -A & -B & 0 \end{pmatrix} \begin{pmatrix} x \\ z \\ y \end{pmatrix}. \quad (2.1)$$

La aplicación directa de MPP a este operador no produce la separación total de los operadores, i.e. dado $P = D^t D$ una matriz simétrica semidefinida positiva, si bien el operador $D(\mathcal{L} + D^t D)^{-1} D^t$ es $\frac{1}{2}$ -average, el operador resolvente asociado $(\mathcal{L} + P)^{-1} P$ no implica la resolución separada de términos proximales respecto de S y T , a su vez que no considera la evaluación separada de los operadores C_1 y C_2 . Por ello consideraremos otros operadores con propiedades similares y principalmente con propiedades separables.

Bajo condiciones de regularidad, el problema (\mathcal{V}) está asociado a la formulación variacional del siguiente problema de optimización

$$\min_{(x,z)} [f(x) + h_1(x) + g(z) + h_2(z) : Ax + Bz = 0], \quad (\mathcal{P})$$

donde h_1 y h_2 son convexas $(\frac{1}{\beta_1})$ -Lipschitz-diferenciable y $(\frac{1}{\beta_2})$ -Lipschitz-diferenciable, respectivamente. Debemos notar que este modelo resulta de considerar que las funciones dadas en el problema de composición de 2 bloques (P) , son la suma de dos funciones convexas siendo una de ellas Lipschitz-diferenciable.

En [14], estudian el modelo (\mathcal{P}) , desarrollando el algoritmo de separación *Majorized iPADMM*, alternativamente construimos diferentes algoritmos, tratando de seguir la metodología empleada en [19], obteniendo a nuestro parecer una extensión más natural del algoritmo *GSS* asociado ahora al problema (\mathcal{P}) .

Las siguientes proposiciones son resultados técnicos que nos ayudan a mostrar la propiedad average de los operadores que definiremos en las siguientes secciones

Proposición 2.1.1 Sean $\beta > 0$ y C un operador β -co-coercivo, entonces

$$\langle z_1 - z_2, C(w_2) - C(w_1) \rangle \leq \frac{1}{4\beta} \|z_1 - z_2 + w_2 - w_1\|^2$$

Prueba. Se cumple que $\langle z_1 - z_2, C(w_2) - C(w_1) \rangle$ es igual a

$$\langle w_1 - w_2, C(w_2) - C(w_1) \rangle + \langle z_1 - z_2 + w_2 - w_1, C(w_2) - C(w_1) \rangle$$

luego al ser C un operador β -co-coercivo y usando la desigualdad de Cauchy en el segundo termino, se cumple que

$$\begin{aligned} \langle z_1 - z_2, C(w_2) - C(w_1) \rangle &\leq -\beta \|C(w_2) - C(w_1)\|^2 + \frac{1}{4\beta} \|z_1 - z_2 + w_2 - w_1\|^2 \\ &\quad + \beta \|C(w_2) - C(w_1)\|^2 \end{aligned}$$

obteniéndose la desigualdad deseada. \blacksquare

Proposición 2.1.2 Sean $G : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ y $\mathcal{F} : \mathbb{R}^r \rightrightarrows \mathbb{R}^r$ dos operadores y D una matriz de orden $q \times r$ tales que

$$G = D\mathcal{F}D^t.$$

Si para $\alpha \in [\frac{1}{2}, 1[$ fijo, se cumple

$$\|GDu_1 - GDu_2\|^2 \leq \|Du_1 - Du_2\|^2 - \frac{1-\alpha}{\alpha} \|Du_1 - GDu_1 - Du_2 + GDu_2\|^2,$$

para todo $u_1, u_2 \in \mathbb{R}^r$, entonces G es α -average.

Prueba. Dados $w_1, w_2 \in \mathbb{R}^q$, existen $u_1, u_2 \in \mathbb{R}^r$ y $v_1, v_2 \in \text{Ker } D^t$ tales que :

$$w_i = Du_i + v_i. \quad \forall i \in \{1, 2\} \quad (2.2)$$

Por hipótesis, para u_1, u_2 , tenemos

$$\|GDu_1 - GDu_2\|^2 \leq \|Du_1 - Du_2\|^2 - \frac{1-\alpha}{\alpha} \|Du_1 - GDu_1 - Du_2 + GDu_2\|^2,$$

de donde, descomponiendo el último termino,

$$\frac{1}{\alpha} \|GDu_1 - GDu_2\|^2 \leq (2 - \frac{1}{\alpha}) \|Du_1 - Du_2\|^2 - 2\frac{1-\alpha}{\alpha} \langle Du_1 - Du_2, GDu_2 - GDu_1 \rangle.$$

Usando la relación $G = D\mathcal{F}D^t$ y (2.2), obtenemos

$$GDu_i = D\mathcal{F}D^t Du_i = D\mathcal{F}D^t w_i = Gw_i, \quad \forall i \in \{1, 2\} \quad (2.3)$$

$$\langle Du_1 - Du_2, GDu_2 - GDu_1 \rangle = \langle w_1 - w_2, Gw_2 - Gw_1 \rangle \quad (2.4)$$

Por (2.2) se tiene que $w_1 - w_2 = D(u_1 - u_2) + v_1 - v_2$, de donde

$$\|D(u_1 - u_2)\| \leq \|w_1 - w_2\|. \quad (2.5)$$

Finalmente, al ser $(2 - \frac{1}{\alpha})$ no negativo, combinando las relaciones (2.3), (2.4) y (2.5), se deduce la afirmación. \blacksquare

2.1.1. Versión Gauss-Seidel

Consideramos las siguientes matrices parámetros : Sea M una $m \times m$ matriz simétrica definida positiva, V_1 y V_2 matrices simétricas semidefinidas positivas de orden $n \times n$ y $p \times p$, respectivamente, tales que $V_1 + A^tMA$ y $V_2 + B^tMB$ son invertibles.

Asociado a estas matrices parámetros, el operador G_1 (asociado al problema (\mathcal{V}) , considerando $B = -I$ y $C_2 = 0$) junto con el operador G_2 (asociado al problema (\mathcal{V}) , considerando $A = -I$ y $C_1 = 0$) nos conduce a agruparlos en un solo operador G_s , que aplica $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ en si mismo, definido como

$$G_s(\hat{x}, \hat{z}, \hat{y}) = \begin{pmatrix} V_1^{\frac{1}{2}}x \\ V_2^{\frac{1}{2}}(z - \tilde{v}) \\ M^{\frac{1}{2}}Ax + M^{\frac{1}{2}}Bz + \hat{y} \end{pmatrix},$$

donde

$$z = (T + V_2 + B^tMB)^{-1} \left(V_2^{\frac{1}{2}}\hat{z} - B^tM^{\frac{1}{2}}\hat{y} \right).$$

$$\tilde{r} = C_1 \left((V_1 + A^tMA)^{-1} (V_1^{\frac{1}{2}}\hat{x} - A^tMBz) \right)$$

$$\tilde{v} = (V_2 + B^tMB)^{-1}C_2(z)$$

$$x = (S + V_1 + A^tMA)^{-1} \left(V_1^{\frac{1}{2}}\hat{x} - A^tM^{\frac{1}{2}}(\hat{y} + 2M^{\frac{1}{2}}Bz - M^{\frac{1}{2}}B\tilde{v}) - \tilde{r} \right)$$

Se deduce que G_s posee dominio total a partir de la monotonía maximal de S y T , y la hipótesis que C_1 y C_2 poseen dominio total. Los puntos fijos de G_s están relacionados con $(V_{\mathcal{L}})$ y es igual a

$$\left\{ \hat{Q} \begin{pmatrix} x^* \\ z^* - (V_2 + B^tMB)^{-1}C_2(z^*) \\ y^* + MB(V_2 + B^tMB)^{-1}C_2(z^*) \end{pmatrix} : (x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}}) \right\}, \quad (2.6)$$

donde \hat{Q} es una matriz definida como

$$\hat{Q} = \begin{pmatrix} V_1^{\frac{1}{2}} & 0 & 0 \\ 0 & V_2^{\frac{1}{2}} & 0 \\ M^{\frac{1}{2}}A & 0 & M^{-\frac{1}{2}} \end{pmatrix}. \quad (2.7)$$

En el caso $C_1 = 0$ y $C_2 = 0$, se cumple que $G_s = \hat{Q}(L + \hat{Q}^t\hat{Q})^{-1}\hat{Q}^t$; en el caso general, G_s se convierte en una buena alternativa del operador $\hat{Q}(\mathcal{L} + \hat{Q}^t\hat{Q})^{-1}\hat{Q}^t$, principalmente por su propiedad separable, conservando la propiedad average aún en el caso general, lo cual mostraremos en lo que sigue.

Para ello consideramos el operador \mathcal{F}_s que aplica $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ en si mismo, definido como

$$\mathcal{F}_s(\tilde{x}, \tilde{z}, \tilde{y}) := \begin{pmatrix} x \\ z - \tilde{\nu} \\ MBz + M\tilde{y} \end{pmatrix},$$

donde

$$\begin{aligned} z &= (T + V_2 + B^tMB)^{-1}(\tilde{z} - B^tM\tilde{y}) \\ \tilde{r} &= C_1((V_1 + A^tMA)^{-1}(\tilde{x} - A^tM\tilde{y} - A^tMBz)) \\ \tilde{\nu} &= (V_2 + B^tMB)^{-1}C_2(z) \\ x &= (S + V_1 + A^tMA)^{-1}(\tilde{x} - 2A^tMBz - 2A^tM\tilde{y} - \tilde{r} + A^tMB\tilde{\nu}). \end{aligned}$$

Comenzamos mostrando la relación de G_s con el método MPP aplicado al lagrangiano L definido en (1.8).

Proposición 2.1.3 *Bajo las hipótesis matriciales consideradas en la definición de G_s , se cumple que*

$$G_s = \widehat{Q}\mathcal{F}_s\widehat{Q}^t. \quad (2.8)$$

Además, dado arbitrariamente $u := (\hat{x}, \hat{z}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, se cumple que

$$L\left(\mathcal{F}_s\widehat{Q}^t\widehat{Q}u + e\right) + \widehat{Q}^t\widehat{Q}\left(\mathcal{F}_s\widehat{Q}^t\widehat{Q}u - u\right) \in \begin{pmatrix} -r + A^tMB\nu \\ -V_2\nu \\ 0 \end{pmatrix}, \quad (2.9)$$

donde

$$\begin{aligned} e &= (0, \nu, 0)^t \\ \nu &= (V_2 + B^tMB)^{-1}C_2(z) \\ r &= C_1\left((V_1 + A^tMA)^{-1}(V_1\hat{x} - A^tMBz)\right) \\ z &= (T + V_2 + B^tMB)^{-1}\left(V_2\hat{z} - B^tMA\hat{x} - B^t\hat{y}\right). \end{aligned}$$

El siguiente Teorema muestra las condiciones necesarias que aseguran la propiedad average de G_s .

Teorema 2.1.1 *Sean $S : \mathbb{R}^n \xrightarrow{\rightarrow} \mathbb{R}^n$ y $T : \mathbb{R}^p \xrightarrow{\rightarrow} \mathbb{R}^p$ monótonos maximales, $C_1 : \mathbb{R}^n \xrightarrow{\rightarrow} \mathbb{R}^n$ β_1 -co-coercivo y $C_2 : \mathbb{R}^p \xrightarrow{\rightarrow} \mathbb{R}^p$ β_2 -co-coercivo, ambos de dominio total con β_1, β_2 ambos positivos y, A y B matrices de orden $m \times n$ y $m \times p$, respectivamente. Asumiendo que $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ son simétricas, con V_1 y V_2 semidefinida positivas y M definida positiva tales que $V_1 + A^tMA$ y $V_2 + B^tMB$ son definidas positivas. Si α satisface*

$$2 - \alpha^{-1} := \frac{\|(V_1 + A^tMA)^{-1}\|}{2\beta_1} + \frac{\|(V_2 + B^tMB)^{-1}\|}{2\beta_2} < 1,$$

entonces, G_s es α -average con dominio total, siendo $\alpha \in]\frac{1}{2}, 1[$.

Prueba. Dados $u_1 = (\hat{x}_1, \hat{z}_1, \hat{y}_1)$ y $u_2 = (\hat{x}_2, \hat{z}_2, \hat{y}_2)$, mostraremos la desigualdad average de G_s . Se cumple que la evaluación de $\mathcal{F}_s \hat{S}^t \hat{S}$ en u_1 y u_2 son, para $i = 1, 2$,

$$\mathcal{F}_s \hat{Q}^t \hat{Q} u_i = \begin{pmatrix} x_i \\ z_i - \nu_i \\ MA\hat{x}_i + MBz_i + \hat{y}_i \end{pmatrix}, \quad (2.10)$$

donde

$$z_i := (T + V_2 + B^t MB)^{-1} (V_2 \hat{z}_i - B^t MA\hat{x}_i - B^t \hat{y}_i)$$

$$r_i = C_1 ((V_1 + A^t MA)^{-1} (V_1 \hat{x}_i - A^t MBz_i))$$

$$\nu_i = (V_2 + B^t MB)^{-1} C_2(z_i)$$

$$x_i := (S + V_1 + A^t MA)^{-1} (V_1 \hat{x}_i - A^t MA\hat{x}_i - A^t \hat{y}_i - 2A^t MBz_i - r_i + A^t MB\nu_i).$$

A partir de la relación (2.9) de la proposición anterior, usando la monotonía de L asociado a los puntos u_1 y u_2 , y la equivalencia (2.8), se obtiene que la siguiente suma

$$\begin{aligned} & \langle G_1 \hat{Q} u_1 - G_1 \hat{Q} u_2, (\hat{Q} u_1 - G_1 \hat{Q} u_1) - (\hat{Q} u_2 - G_1 \hat{Q} u_2) \rangle + \\ & \left\langle \mathcal{F}_s \hat{Q}^t \hat{Q} u_1 - \mathcal{F}_s \hat{Q}^t \hat{Q} u_2, \begin{pmatrix} -r_1 + r_2 + A^t MB(\nu_1 - \nu_2) \\ V_2(-\nu_1 + \nu_2) \\ 0 \end{pmatrix} \right\rangle + \\ & \langle \nu_1 - \nu_2, V_2(\hat{z}_1 - z_1 + \nu_1 - \hat{z}_2 + z_2 - \nu_2) - V_2\nu_1 + V_2\nu_2 \rangle \end{aligned}$$

es mayor o igual que cero. Descomponiendo el segundo y tercer término obtenemos

$$\begin{aligned} & \langle G_1 \hat{Q} u_1 - G_1 \hat{Q} u_2, (\hat{Q} u_1 - G_1 \hat{Q} u_1) - (\hat{Q} u_2 - G_1 \hat{Q} u_2) \rangle + \langle x_1 - x_2, -r_1^1 + r_2^1 \rangle + \\ & \langle x_1 - x_2, A^t MB(\nu_1 - \nu_2) \rangle + \langle z_1 - \nu_1 - z_2 + \nu_2, V_2(-\nu_1 + \nu_2) \rangle + \\ & + \langle \nu_1 - \nu_2, V_2(\hat{z}_1 - z_1 - \hat{z}_2 + z_2) \rangle \geq 0. \end{aligned}$$

Agrupandolos y considerando $\varsigma_i = -B^t MAx_i + V_2 z_i - V_2(\hat{z}_i - z_i + \nu_i)$, obtenemos

$$\begin{aligned} & \langle G_1 \hat{Q} u_1 - G_1 \hat{Q} u_2, (\hat{Q} u_1 - G_1 \hat{Q} u_1) - (\hat{Q} u_2 - G_1 \hat{Q} u_2) \rangle + \\ & \langle x_1 - x_2, -r_1 + r_2 \rangle + \langle \nu_1 - \nu_2, \varsigma_2 - \varsigma_1 \rangle \geq 0 \quad (2.11) \end{aligned}$$

Ahora buscaremos apropiadas cotas superiores de los últimos dos términos considerando $\left\| \hat{S} u_1 - G \hat{S} u_1 - \hat{S} u_2 + G \hat{S} u_2 \right\|$. Al ser C_1 y C_2 co-coercivos con coeficientes de coercividad positivos y usando la proposición 2.1.1, obtenemos

$$\langle x_1 - x_2, -r_1 + r_2 \rangle \leq \frac{1}{4\beta_1} \|x_1 - x_2 + \eta_2 - \eta_1\|^2 \quad (2.12)$$

$$\langle \nu_1 - \nu_2, \varsigma_2 - \varsigma_1 \rangle \leq \frac{1}{4\beta_2} \|z_1 - z_2 + \xi_2 - \xi_1\|^2 \quad (2.13)$$

donde $\eta_i = (V_1 + A^t M A)^{-1}(V_1 \hat{x}_i - A^t M B z_i)$ y $\xi_i = (V_2 + B^t M B)^{-1} \varsigma_i$.

Por otro lado se cumple que $\left\| \widehat{Q}u_1 - G_1 \widehat{Q}u_1 - \widehat{Q}u_2 + G_1 \widehat{Q}u_2 \right\|^2$ es igual a

$$\left\| \widehat{Q} \left[u_1 - \mathcal{F}_s \widehat{Q}^t \widehat{Q}u_1 - u_2 + \mathcal{F}_s \widehat{Q}^t \widehat{Q}u_2 \right] \right\|^2 = \|(p_1, p_2, p_3)\|^2, \quad (2.14)$$

donde

$$\begin{aligned} p_1 &= V_1^{\frac{1}{2}}(\hat{x}_1 - x_1 - \hat{x}_2 + x_2) \\ p_2 &= V_2^{\frac{1}{2}}(\hat{z}_1 - z_1 + \nu_1 - \hat{z}_2 + z_2 - \nu_2) \\ p_3 &= M^{\frac{1}{2}}A(-x_1 + x_2) + M^{\frac{1}{2}}B(-z_1 + z_2). \end{aligned}$$

Denotando las matrices $K_1 = \begin{pmatrix} V_1^{\frac{1}{2}} & A^t M^{\frac{1}{2}} \end{pmatrix}$ y $K_2 = \begin{pmatrix} V_2^{\frac{1}{2}} & -B^t M^{\frac{1}{2}} \end{pmatrix}$, se cumple

$$\begin{aligned} K_1(p_1, p_3)^t &= K_1 K_1^t (\eta_1 - \eta_2 - x_1 + x_2) \\ K_2(p_2, p_3)^t &= K_2 K_2^t (z_1 - z_2 + \xi_2 + \xi_1) \end{aligned}$$

luego combinando estos resultados con (2.33), obtenemos

$$\left\| \widehat{Q}u_1 - G_1 \widehat{Q}u_1 - \widehat{Q}u_2 + G_1 \widehat{Q}u_2 \right\|^2 \geq \frac{1}{\|(K_1 K_1^t)^{-1}\|} \|x_1 - x_2 + \eta_2 - \eta_1\|^2 \quad (2.15)$$

$$\left\| \widehat{Q}u_1 - G_1 \widehat{Q}u_1 - \widehat{Q}u_2 + G_1 \widehat{Q}u_2 \right\|^2 \geq \frac{1}{\|(K_2 K_2^t)^{-1}\|} \|z_1 - z_2 + \xi_2 - \xi_1\|^2 \quad (2.16)$$

luego usando estas desigualdades en (2.12) y (2.13), obtenemos las cotas deseadas de $\langle x_1 - x_2, -r_1 + r_2 \rangle$ y $\langle \nu_1 - \nu_2, \varsigma_2 - \varsigma_1 \rangle$, esto es

$$\begin{aligned} &\langle x_1 - x_2, -r_1 + r_2 \rangle + \langle \nu_1 - \nu_2, \varsigma_2 - \varsigma_1 \rangle \\ &\leq \left(\frac{\|(V_1 + A^t M A)^{-1}\|}{2\beta_1} + \frac{\|(V_2 + B^t M B)^{-1}\|}{2\beta_2} \right) \left\| \widehat{Q}u_1 - G_1 \widehat{Q}u_1 - \widehat{Q}u_2 + G_1 \widehat{Q}u_2 \right\|^2. \end{aligned}$$

Finalmente, usando esta cota en la desigualdad (2.11), obtenemos la desigualdad

$$\left\| G_s \widehat{Q}u_1 - G_s \widehat{Q}u_2 \right\|^2 \leq \left\| \widehat{Q}u_1 - \widehat{Q}u_2 \right\|^2 - \frac{1-\alpha}{\alpha} \left\| \widehat{Q}u_1 - G_s \widehat{Q}u_1 - \widehat{Q}u_2 + G_s \widehat{Q}u_2 \right\|^2$$

que, por la proposición 2.1.2, deducimos el resultado. \blacksquare

Por las propiedades de separación y average de G_s , se obtiene un algoritmo convergente de separación a partir del método de punto fijo aplicado a $\mathcal{F}_s \widehat{Q}^t \widehat{Q}$ (en

lugar del operador $[\mathcal{L} + \widehat{Q}^t \widehat{Q}]^{-1} \widehat{Q}^t \widehat{Q}$: dado un punto arbitrario $(x^0, \bar{z}^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, se genera la siguiente sequencia

$$(x^{k+1}, \bar{z}^{k+1}, y^{k+1}) = \mathcal{F}_s \widehat{Q}^t \widehat{Q}(x^k, \bar{z}^k, y^k). \quad (2.17)$$

Teniendo en cuenta la evaluación de $\mathcal{F}_s \widehat{Q}^t \widehat{Q}$ dado en (2.10), introducimos, por comodidad, las variables z^k y ν^k definidas a partir de z^0 y ν^0 arbitrarios tal que $\bar{z}^0 = z^0 - \nu^0$ y, para $k \geq 0$,

$$z^{k+1} := (T + V_2 + B^t M B)^{-1} (V_2(z^k - \nu^k) - B^t M A x^k - B^t y^k)$$

$$\nu^{k+1} := (V_2 + B^t M B)^{-1} C_2(z^{k+1}).$$

La arbitrariedad de \bar{z}^0 es reflejada en la arbitrariedad de z^0 y ν^0 y se cumple que $\bar{z}^k = z^k - \nu^k$ para todo $k \geq 0$.

Por tanto se obtiene el siguiente algoritmo

**Primal-dual Two-block compositive splitting, Gauus-Seidel version
(PD2BCS, Gauus-Seidel version)**

Escoger $(x^0, z^0, y^0, \nu^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ and $M \in \mathbb{R}^{m \times m}$ tales que $V_1 + A^t M A$ y $V_2 + B^t M B$ sean matrices invertibles

$$z^{k+1} = (T + V_2 + B^t M B)^{-1} (V_2(z^k - \nu^k) - B^t M A x^k - B^t y^k) \quad (2.18)$$

$$y^{k+1} = y^k + M A x^k + M B z^{k+1} \quad (2.19)$$

$$r^{k+1} = C_1 ((V_1 + A^t M A)^{-1} (V_1 x^k - A^t M B z^{k+1})) \quad (2.20)$$

$$\nu^{k+1} = (V_2 + B^t M B)^{-1} C_2(z^{k+1}) \quad (2.21)$$

$$\zeta^{k+1} = V_1 x^k - A^t M B(z^{k+1} - \nu^{k+1}) - A^t y^{k+1} - r^{k+1} \quad (2.22)$$

$$x^{k+1} = (S + V_1 + A^t M A)^{-1} (\zeta^{k+1}) \quad (2.23)$$

A partir del cambio de variable de \bar{z}^k en (2.17), se obtiene que la sequencia generada por el algoritmo anterior $\zeta^k := (x^k, z^k - \nu^k, y^k)$ cumple que :

$$\zeta^{k+1} = \mathcal{F}_s \widehat{Q}^t \widehat{Q}(\zeta^k)$$

la cual se relaciona con el operador G_s al evaluarse la matriz \widehat{Q} en la relación anterior y utilizando la igualdad (2.8) en la proposicion 2.1.3, obteniendose

$$\widehat{Q} \zeta^{k+1} = G_s \widehat{Q}(\zeta^k), \quad (2.24)$$

y así su convergencia a partir de la propiedad avarage de G_s .

Se observa que este algoritmo utiliza de manera secuencial el cálculo de los subproblemas proximales de T y S , es por ello que le adicionamos el término *Gauss-Seidel version*, tal característica también es observado en la definición de G_s . En la siguiente sección alternativamente obtendremos una versión paralela.

En el caso del problema de optimización (\mathcal{P}), y considerando el cambio de variable $(\bar{x}^k, \bar{z}^k, \bar{y}^k, \bar{\nu}^k) = (x^k, z^{k+1}, y^{k+1}, \nu^{k+1})$ se obtiene el siguiente algoritmo :

$$\begin{aligned} \bar{r}^k &= (V_1 + A^t M A)^{-1} \nabla h_1 \left((V_1 + A^t M A)^{-1} (V_1 \bar{x}^k - A^t M B \bar{z}^k) \right) \\ \bar{x}^{k+1} &= \operatorname{argmin} \left\{ f(x) + \frac{1}{2} \|A(x + \bar{r}^k) + B(\bar{z}^k - \nu^k) + M^{-1} \bar{y}^k\|_M^2 + \frac{1}{2} \|x - \bar{x}^k + \bar{r}^k\|_{V_1}^2 \right\} \\ \bar{\nu}^k &= (V_2 + B^t M B)^{-1} \nabla h_2(\bar{z}^k) \\ \bar{z}^{k+1} &= \operatorname{argmin} \left\{ g(z) + \frac{1}{2} \|A \bar{x}^{k+1} + Bz + M^{-1} \bar{y}^k\|_M^2 + \frac{1}{2} \|z - \bar{z}^k + \bar{\nu}^k\|_{V_2}^2 \right\} \\ \bar{y}^{k+1} &= \bar{y}^k + M(A \bar{x}^{k+1} + B \bar{z}^{k+1}) \end{aligned}$$

La estructura del algoritmo anterior nos permite una mejor comparación de las diferencias con el algoritmo *Majorized iPADMM* [14], obteniendo así una extensión distinta del *ADMM*.

2.1.2. Versión Jacobi

Consideramos las siguientes matrices parámetros : Sea M una $m \times m$ matriz simétrica definida positiva, R_1 y R_2 matrices simétricas semidefinidas positivas de orden $n \times n$ y $p \times p$, respectivamente, tales que $R_1 + 2A^t M A$ y $R_2 + 2B^t M B$ sean invertibles.

Tomando en cuenta el operador G_D^L (ver 1.5), considerando la matriz D asociada al caso $\gamma = 0$, el cual corresponde al algoritmo *PPDS version Jacobi*, extendemos este operador considerando el problema general (\mathcal{V}), manteniendo el cálculo paralelo de los subproblemas proximales de T y S , obteniendo así G_p que aplica $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^m$ en si mismo definido como:

$$G_p(\hat{x}, \hat{z}, \hat{w}, \hat{y}) = \begin{pmatrix} R_1^{\frac{1}{2}} x \\ R_2^{\frac{1}{2}} z \\ M^{\frac{1}{2}} A x - M^{\frac{1}{2}} B z \\ M^{\frac{1}{2}} A x + M^{\frac{1}{2}} B z + \hat{y} \end{pmatrix},$$

donde

$$x = (S + R_1 + 2A^t M A)^{-1} \left(R_1^{\frac{1}{2}} \hat{x} + A^t M^{\frac{1}{2}} (\hat{w} - \hat{y}) - \hat{r}_1 \right)$$

$$\begin{aligned}\hat{r}_1 &= C_1 \left((R_1 + 2A^t M A)^{-1} (R_1^{\frac{1}{2}} \hat{x} + A^t M^{\frac{1}{2}} \hat{w}) \right) \\ z &= (T + R_2 + 2B^t M B)^{-1} \left(R_2^{\frac{1}{2}} \hat{z} + B^t M^{\frac{1}{2}} (-\hat{w} - \hat{y}) - \hat{r}_2 \right). \\ \hat{r}_2 &= C_2 \left((R_2 + 2B^t M B)^{-1} (R_2^{\frac{1}{2}} \hat{z} - B^t M^{\frac{1}{2}} \hat{w}) \right)\end{aligned}$$

Análogamente al operador G_s , mostraremos que G_p es average con dominio total, mostraremos también su algoritmo de separación asociado.

Se observa que bajo las hipótesis iniciales G_p posee dominio total, cuyos puntos fijos son

$$\widehat{D}(\text{sol}(V_{\mathcal{L}})) = \left\{ (R_1^{\frac{1}{2}} x^*, R_2^{\frac{1}{2}} z^*, M^{\frac{1}{2}} A x^* - M^{\frac{1}{2}} B z^*, M^{-\frac{1}{2}} y^*) : (x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}}) \right\} \quad (2.25)$$

donde \widehat{S} es una matriz definida como

$$\widehat{D} = \begin{pmatrix} R_1^{\frac{1}{2}} & 0 & 0 \\ 0 & R_2^{\frac{1}{2}} & 0 \\ M^{\frac{1}{2}} A & -M^{\frac{1}{2}} B & 0 \\ 0 & 0 & M^{-\frac{1}{2}} \end{pmatrix}. \quad (2.26)$$

En el caso $C_1 = 0$ y $C_2 = 0$ se cumple que $G_p = G_D^L = \widehat{D}(L + \widehat{D}^t \widehat{D})^{-1} \widehat{D}^t$, en el caso general mostraremos que G_p mantiene la propiedad average, para ello consideramos el operador \mathcal{F}_p que aplica $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ hacia si mismo, definido como

$$\mathcal{F}_p(\tilde{x}, \tilde{z}, \tilde{y}) := \begin{pmatrix} x \\ z \\ M A x + M B z + M \tilde{y} \end{pmatrix}$$

donde

$$\begin{aligned}x &= (S + R_1 + 2A^t M A)^{-1} (\tilde{x} - A^t M \tilde{y} - \tilde{r}_1) \\ \tilde{r}_1 &= C_1 \left((R_1 + 2A^t M A)^{-1} \tilde{x} \right) \\ z &= (T + R_2 + 2B^t M B)^{-1} (\tilde{z} - B^t M \tilde{y} - \tilde{r}_2) \\ \tilde{r}_2 &= C_2 \left((R_2 + 2B^t M B)^{-1} \tilde{z} \right).\end{aligned}$$

La siguiente proposición muestra la relación de G_p con MPP aplicado a L .

Proposición 2.1.4 *Bajo las hipótesis matriciales considerados en la definición de G_p , se cumple que*

$$G_p = \widehat{D} \mathcal{F}_p \widehat{D}^t. \quad (2.27)$$

Además, dado arbitrariamente $u := (\hat{x}, \hat{z}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, se cumple que

$$L\left(\mathcal{F}_p \hat{D}^t \hat{D}u\right) + \hat{D}^t \hat{D}\left(\mathcal{F}_p \hat{D}^t \hat{D}u - u\right) \in \begin{pmatrix} -C_1(\eta_1) \\ -C_2(\eta_2) \\ 0 \end{pmatrix}, \quad (2.28)$$

donde

$$\begin{aligned} \eta_1 &= (R_1 + 2A^t M A)^{-1} \left((R_1 + A^t M A) \hat{x} - A^t M B \hat{z} \right) \\ \eta_2 &= (R_2 + 2B^t M B)^{-1} \left((R_2 + B^t M B) \hat{z} - B^t M A \hat{x} \right). \end{aligned}$$

El siguiente Teorema muestra las condiciones necesarias para obtener la propiedad average de G_p .

Teorema 2.1.2 Sean $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ y $T : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ monótonos maximales, $C_1 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ β_1 -co-coercivo y $C_2 : \mathbb{R}^p \rightrightarrows \mathbb{R}^p$ β_2 -co-coercivo (β_1 y β_2 , positivos), ambos con dominio total y, A y B matrices de orden $m \times n$ y $m \times p$, respectivamente. Asumiendo que $R_1 \in \mathbb{R}^{n \times n}$, $R_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ son simétricas, con R_1 y R_2 semidefinida positivas y M definida positiva tales que $R_1 + 2A^t M A$ y $R_2 + 2B^t M B$ son definidas positivas, cumpliéndose que α satisface

$$2 - \alpha^{-1} := \frac{\|(R_1 + 2A^t M A)^{-1}\|}{2\beta_1} + \frac{\|(R_2 + 2B^t M B)^{-1}\|}{2\beta_2} < 1,$$

entonces, G_p es α -average con dominio total, siendo $\alpha \in]\frac{1}{2}, 1[$.

Prueba. Dados $u_1 = (\hat{x}_1, \hat{z}_1, \hat{y}_1)$ y $u_2 = (\hat{x}_2, \hat{z}_2, \hat{y}_2)$, mostraremos la desigual average de G_p . Se cumple que la evaluación de $\mathcal{F}_p \hat{D}^t \hat{D}$ en u_1 y u_2 son: para $i = 1, 2$

$$\mathcal{F}_p \hat{D}^t \hat{D}u_i = \begin{pmatrix} x_i \\ z_i \\ M A x_i + M B z_i + \hat{y}_i \end{pmatrix} \quad (2.29)$$

donde

$$\begin{aligned} x_i &:= (S + R_1 + 2A^t M A)^{-1} \left((R_1 + A^t M A) \hat{x}_i - A^t M B \hat{z}_i - A^t \hat{y}_i - C_1(\eta_i^1) \right) \\ \eta_i^1 &:= (R_1 + 2A^t M A)^{-1} \left((R_1 + A^t M A) \hat{x}_i - A^t M B \hat{z}_i \right) \\ y_i &:= (T + R_2 + 2B^t M B)^{-1} \left((R_2 + B^t M B) \hat{z}_i - B^t M A \hat{x}_i - B^t \hat{y}_i - C_2(\eta_i^2) \right). \\ \eta_i^2 &:= (R_2 + 2B^t M B)^{-1} \left((R_2 + B^t M B) \hat{z}_i - B^t M A \hat{x}_i \right) \end{aligned}$$

Luego de (2.28) considerando u_1 y u_2 , usando la monotonia de L , y la relación (2.27), se tiene que

$$\begin{aligned} &\langle G_p \hat{D}u_1 - G_p \hat{D}u_2, (\hat{D}u_1 - G_p \hat{D}u_1) - (\hat{D}u_2 - G_p \hat{D}u_2) \rangle \\ &\quad + \langle x_1 - x_2, C_1(\eta_2^1) - C_1(\eta_1^1) \rangle + \langle z_1 - z_2, C_2(\eta_2^2) - C_2(\eta_1^2) \rangle \geq 0 \quad (2.30) \end{aligned}$$

Ahora encontraremos una cota superior de los últimos dos terminos de (2.30), considerando $\left\| \widehat{D}u_1 - G_p \widehat{D}u_1 - \widehat{D}u_2 + G_p \widehat{D}u_2 \right\|$. Por la proposición 2.1.1, se sigue que

$$\langle x_1 - x_2, C_1(\eta_2^1) - C_1(\eta_1^1) \rangle \leq \frac{1}{4\beta_1} \|x_1 - x_2 + \eta_2^1 - \eta_1^1\|^2 \quad (2.31)$$

$$\langle z_1 - z_2, C_2(\eta_2^2) - C_2(\eta_1^2) \rangle \leq \frac{1}{4\beta_2} \|z_1 - z_2 + \eta_2^2 - \eta_1^2\|^2 \quad (2.32)$$

De otro lado, se tiene que $\left\| \widehat{D}u_1 - G_p \widehat{D}u_1 - \widehat{D}u_2 + G_p \widehat{D}u_2 \right\|^2$ es igual a

$$\left\| \widehat{D} \left[u_1 - \mathcal{F}_p \widehat{D}^t \widehat{D}u_1 - u_2 + \mathcal{F}_p \widehat{D}^t \widehat{D}u_2 \right] \right\|^2 = \|(p_1, p_2, p_3, p_4)\|^2 \quad (2.33)$$

donde

$$p_1 = R_1^{\frac{1}{2}}(\widehat{x}_1 - x_1 - \widehat{x}_2 + x_2)$$

$$p_2 = R_2^{\frac{1}{2}}(\widehat{z}_1 - z_1 - \widehat{z}_2 + z_2)$$

$$p_3 = M^{\frac{1}{2}}A(\widehat{x}_1 - x_1 - \widehat{x}_2 + x_2) - M^{\frac{1}{2}}B(\widehat{z}_1 - z_1 - \widehat{z}_2 + z_2)$$

$$p_4 = M^{-\frac{1}{2}}(-MAx_1 - MBz_1 + MAx_2 + MBz_2)$$

notando las matrices $K_1 = \begin{pmatrix} R_1^{\frac{1}{2}} & A^t M^{\frac{1}{2}} & A^t M^{\frac{1}{2}} \end{pmatrix}$ y $K_2 = \begin{pmatrix} R_2^{\frac{1}{2}} & -B^t M^{\frac{1}{2}} & B^t M^{\frac{1}{2}} \end{pmatrix}$, se satisface que

$$K_1(p_1, p_3, p_4)^t = K_1 K_1^t (\eta_1^1 - \eta_2^1 - x_1 + x_2)$$

$$K_2(p_2, p_3, p_4)^t = K_2 K_2^t (\eta_1^2 - \eta_2^2 - z_1 + z_2)$$

luego usando estas relaciones y (2.33), se obtiene que

$$\left\| \widehat{D}u_1 - G_p \widehat{D}u_1 - \widehat{D}u_2 + G_p \widehat{D}u_2 \right\|^2 \geq \frac{1}{\|(K_1 K_1^t)^{-1}\|} \|x_1 - x_2 + \eta_2^1 - \eta_1^1\|^2$$

$$\left\| \widehat{D}u_1 - G_p \widehat{D}u_1 - \widehat{D}u_2 + G_p \widehat{D}u_2 \right\|^2 \geq \frac{1}{\|(K_2 K_2^t)^{-1}\|} \|z_1 - z_2 + \eta_2^2 - \eta_1^2\|^2$$

por lo tanto de (2.31) y (2.32), obtenemos la cota deseada

$$\begin{aligned} & \langle x_1 - x_2, C_1(\eta_2^1) - C_1(\eta_1^1) \rangle + \langle z_1 - z_2, C_2(\eta_2^2) - C_2(\eta_1^2) \rangle \\ & \leq \left(\frac{\|(R_1 + 2A^t M A)^{-1}\|}{4\beta_1} + \frac{\|(R_2 + 2B^t M B)^{-1}\|}{4\beta_2} \right) \left\| \widehat{D}u_1 - G_p \widehat{D}u_1 - \widehat{D}u_2 + G_p \widehat{D}u_2 \right\|^2 \end{aligned}$$

Finalmente usando esta cota en (2.30) se obtiene que

$$\left\| G_p \widehat{D}u_1 - G_p \widehat{D}u_2 \right\|^2 \leq \left\| \widehat{D}u_1 - \widehat{D}u_2 \right\|^2 - \frac{1-\alpha}{\alpha} \left\| \widehat{D}u_1 - G_p \widehat{D}u_1 - \widehat{D}u_2 + G_p \widehat{D}u_2 \right\|^2$$

donde

$$\alpha := \left(2 - \frac{\|(R_1 + 2A^tMA)^{-1}\|}{2\beta_1} - \frac{\|(R_2 + 2B^tMB)^{-1}\|}{2\beta_2} \right)^{-1}$$

que por la proposición 2.1.2 , muestra que G_p es α -average. \blacksquare

El algoritmo de separación asociado a G_p , proviene del método de punto fijo aplicado a $\mathcal{F}_p \widehat{D}^t \widehat{D}$, esto es : dado $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^p$ arbitrario generar la sequencia

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \mathcal{F}_p \widehat{D}^t \widehat{D}(x^k, z^k, y^k) \quad (2.34)$$

luego de (2.29), se obtiene el siguiente algoritmo :

**Primal-dual Two-block compositive splitting, Jacobi version
(PD2BCS, Jacobi version)**

Escoger $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, $R_1 \in \mathbb{R}^{n \times n}$, $R_2 \in \mathbb{R}^{p \times p}$ and $M \in \mathbb{R}^{m \times m}$ tales que $R_1 + 2A^tMA$ y $R_2 + 2B^tMB$ son matrices invertibles

$$r_1^k = C_1 ((R_1 + 2A^tMA)^{-1}((R_1 + A^tMA)x^k - A^tMBz^k)) \quad (2.35)$$

$$x^{k+1} = (S + R_1 + 2A^tMA)^{-1} ((R_1 + A^tMA)x^k - A^tMBz^k - A^ty^k - r_1^k) \quad (2.36)$$

$$r_2^k = C_2 ((R_2 + 2B^tMB)^{-1}((R_2 + B^tMB)z^k - B^tMAx^k)) \quad (2.37)$$

$$z^{k+1} = (T + R_2 + 2B^tMB)^{-1} ((R_2 + B^tMB)z^k - B^tMAx^k - B^ty^k - r_2^k) \quad (2.38)$$

$$y^{k+1} = y^k + MAx^{k+1} + MBz^{k+1} \quad (2.39)$$

Dado la sequencia generada por el algoritmo, considerando $\zeta^k := (x^k, z^k, y^k)$ y evaluando la matriz \widehat{D} en (2.34), se obtiene la relación con el algoritmo de punto fijo aplicado a G_p

$$\widehat{D}\zeta^{k+1} = G_p(\widehat{D}\zeta^k) \quad (2.40)$$

y al ser G_p average, se deduce la convergencia del algoritmo anterior.

En el caso del problema de minimización asociado (\mathcal{P}), se obtiene el algoritmo :

$$\begin{aligned} \bar{r}_1^k &= (\bar{V}_1 + A^tMA)^{-1} \nabla h_1 ((\bar{V}_1 + A^tMA)^{-1}(\bar{V}_1 x^k - A^tMBz^k)) \\ x^{k+1} &= \operatorname{argmin} \left\{ f(x) + \frac{1}{2} \|A(x + \bar{r}_1^k) + Bz^k + M^{-1}y^k\|_M^2 + \frac{1}{2} \|x - x^k + \bar{r}_1^k\|_{\bar{V}_1}^2 \right\} \\ \bar{r}_2^k &= (\bar{V}_2 + B^tMB)^{-1} \nabla h_2 ((\bar{V}_2 + B^tMB)^{-1}(\bar{V}_2 z^k - B^tMAx^k)) \\ z^{k+1} &= \operatorname{argmin} \left\{ g(z) + \frac{1}{2} \|Ax^k + B(z + \bar{r}_2^k) + M^{-1}\bar{y}^k\|_M^2 + \frac{1}{2} \|z - z^k + \bar{r}_2^k\|_{\bar{V}_2}^2 \right\} \\ y^{k+1} &= y^k + M(Ax^{k+1} + Bz^{k+1}) \end{aligned}$$

donde consideramos las matrices $\bar{V}_1 := R_1 + A^tMA$ y $\bar{V}_2 := R_2 + B^tMB$.

Observación 2.1.1 *En el caso que uno de los operadores C_1 o C_2 sea nulo, sin pérdida de generalidad sea $C_2 = 0$, entonces la hipótesis en el Teorema 2.1.1 y 2.1.2 se reduce a que se cumpla*

$$2 - \alpha^{-1} := \frac{\|(V_1 + A^t M A)^{-1}\|}{2\beta_1} < 1 \quad y \quad 2 - \alpha^{-1} := \frac{\|(R_1 + 2A^t M A)^{-1}\|}{2\beta_1} < 1$$

respectivamente. Esto se deduce observando la prueba de los teoremas o directamente de ellos, pues en este caso se obtiene que los operadores involucrados respectivamente son $\hat{\alpha}$ -average, para todo $\hat{\alpha} \in]\alpha, 1[$, esto debido a que C_2 β_2 -co-coercivo, para todo $\beta_2 > 0$.

Capítulo 3

Variante del Método de Punto fijo y Punto Proximal

A partir del algoritmo “A relative-error inertial-relaxed HPP (RIRHPP)” propuesto por Alves-Eckstein [2], en este capítulo mostraremos las variantes del tipo Relative-error e Inertial-Relaxed, aplicados en conjunto al método de Punto fijo, para poder hallar el punto fijo de un operador α -average. Luego este resultado permitirá obtener una variante Relative-error e Inertial del algoritmo *MPP* (ver Subsección 1.2), obteniendo así una variante general del método proximal. Estos resultados serán usados luego en los capítulos siguientes para obtener mejoras en los algoritmos de separación, a partir de su relación con el método de punto fijo y punto proximal, respectivamente.

La variante Relative-error es similar a lo desarrollado por Solodov, permitiendo debilitar el cálculo exacto del subproblema en uno donde sólo implica satisfacer una desigualdad (similar al subproblema (1.1)).

La variante Inertial-Relaxed, permite insertar los parametros inercial y relajado en conjunto. Resultados numéricos muestran que ambos términos independientemente producen mejoras en los algoritmos que lo posean [3], por ello es interesante considerar algoritmos que posean estos parámetros en conjunto.

3.1. Variante del Método de Punto fijo

Como se sabe, para resolver el problema correspondiente a la suma de tres operadores donde uno de ellos es co-coercivo, D. Daves [8] desarrolló un algoritmo separable, considerando el método de punto fijo aplicado a un operador especial α -average. Aún para el problema general “inclusión de composición monótona más un ope-

rador co-coercivo” (ver Subsección 1.5), se obtienen dos operadores α –average G_1 y G_2 (definidos en (1.28) y (1.36)), que permiten obtener un algoritmo separable al relacionarlo con el método punto fijo (usando (1.26) y (1.35)). Es por ello que desarrollaremos una variante del método de punto fijo para luego aplicarlo a los operadores G_1 y G_2 respectivamente.

Observando el algoritmo *RIRHPP*, al considerar $J^{c\Gamma} = (I_{r \times r} + c\Gamma)^{-1}$, el subproblema inexacto (1.2) es rescrito como

Encontrar $(\tilde{z}^k, \zeta^k) \in \mathbb{R}^r \times \mathbb{R}^r$ tales que

$$\tilde{z}^k = J^{c\Gamma}(w^k + \zeta^k) \quad \|\zeta^k\|^2 \leq \theta^2 \left(\|\tilde{z}^k - w^k\|^2 + \|w^k + \zeta^k - \tilde{z}^k\|^2 \right). \quad (3.1)$$

Por lo tanto este algoritmo es descrito como una variación del método de punto fijo aplicado a $J^{c\Gamma}$. Esto permite desarrollar una variante del método de punto fijo aplicado a G , un operador 1–co-coercivo con dominio total, considerando el subproblema (3.1) reemplazando G por J^{cT} , pues resultaría en la aplicación del algoritmo *RIRHPP* al operador $G^{-1} - I_{r \times r}$ que por el Teorema de Minty, es un operador monótono maximal.

Esta variante del método de punto fijo es útil también para desarrollar una variante general de este método para operadores más generales como los α –average (recordar que un operador 1–co-coercivo es $\frac{1}{2}$ –average). La siguiente Proposición relaciona los operadores α –average y 1–co-coercivo, mostrándonos que encontrar un punto fijo de un operador α –average es equivalente a encontrar el punto fijo de un operador 1–cocoercivo no más complejo como el inicial, pero con el inconveniente de conocer explícitamente el parámetro α .

Proposición 3.1.1 *Dado $\alpha \in]0, 1[$. Se cumple que F es α –average si y solo si $(1 - \frac{1}{2\alpha})I + \frac{1}{2\alpha}F$ es 1–cocoercivo. Además, F y $(1 - \frac{1}{2\alpha})I + \frac{1}{2\alpha}F$ tienen el mismo conjunto de puntos fijos.*

Prueba. Dado N no expansivo se cumple:

$$F = (1 - \alpha)I + \alpha N \text{ si y solo si } (1 - \frac{1}{2\alpha})I + \frac{1}{2\alpha}F = \frac{1}{2}(I + N)$$

Además $x^* = F(x^*)$ si y solo si $x^* = (1 - \frac{1}{2\alpha})x^* + \frac{1}{2\alpha}F(x^*)$. ■

Por lo tanto, dado $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ un operador α –average con dominio total, para poder hallar uno de sus puntos fijos, consideramos alternativamente el operador $(1 - \frac{1}{2\alpha})I + \frac{1}{2\alpha}F$, que por la proposición anterior es 1–cocoercivo con dominio total, luego remplazándolo por J^{cT} en (3.1), obtenemos el siguiente algoritmo :

Relative-error Inertial-Relaxed Fixed Point (RIRFP)

Inicialización: Escoger $\eta^0 = \eta^{-1} \in \mathbb{R}^n$, $\theta \in [0, 1[$, además $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ y considerar $\hat{\rho} := \frac{\bar{\rho}}{2\alpha}$ donde α es el coeficiente average de F .

Para $k = 0, 1, 2, \dots$ **hacer**

- Escoger $\lambda \in [0, \bar{\lambda}[$ y definir (término inercial)

$$w^k = \eta^k + \lambda(\eta^k - \eta^{k-1})$$

- Subproblema inexacto:
Encontrar $(\tilde{d}^k, \varsigma^k) \in \mathbb{R}^n \times \mathbb{R}^n$ tales que

$$\tilde{d}^k = F(w^k + \varsigma^k) \text{ y } \|\varsigma^k\|^2 \leq \left(\frac{\theta}{2\alpha}\right)^2 \left(\|(2\alpha - 1)\varsigma^k + \tilde{d}^k - w^k\|^2 + \|\vartheta^k\|^2 \right)$$

Considerando $\vartheta^k := w^k + \varsigma^k - \tilde{d}^k$,

- Si $\vartheta^k = 0$, **PARAR**. Caso contrario, escoger $\rho \in]0, \hat{\rho}]$ y actualizar

$$\eta^{k+1} = w^k - \rho \frac{\langle (1 - 2\alpha)\varsigma^k + w^k - \tilde{d}^k, \vartheta^k \rangle}{\|\vartheta^k\|^2} \vartheta^k.$$

Terminar el bucle.

Es claro que si el algoritmo se detiene, entonces \tilde{d}^k es un punto fijo de F . Caso contrario, la siguiente proposición muestra la convergencia del algoritmo, la cual es una aplicación del Teorema 1.2.1.

Proposición 3.1.2 *Dados $\bar{\lambda}$ y $\bar{\rho}$ satisfaciendo (H1). Sea F un operador α -average con dominio total que posee al menos un punto fijo. Si el algoritmo no se detiene, entonces η^k y \tilde{d}^k , ambos, convergen a un mismo punto fijo de F y, ς^k conjuntamente con ϑ^k convergen a cero.*

Prueba. Usando la Proposición 3.1.1, $(1 - \frac{1}{2\alpha})I + \frac{1}{2\alpha}F$ es 1-cocoercivo con dominio total y posee igual conjunto de puntos fijos que F . Luego considerando \tilde{z}^k y v^k definidos como

$$\tilde{z}^k := \left[\left(1 - \frac{1}{2\alpha}\right)I + \frac{1}{2\alpha}F \right] (w^k + \varsigma^k) \text{ y } v^k := w^k + \varsigma^k - \tilde{z}^k$$

se cumple que

$$\tilde{z}^k - w^k = \frac{1}{2\alpha} \left((2\alpha - 1)\varsigma^k + \tilde{d}^k - w^k \right) \text{ y } v^k = \frac{1}{2\alpha} \vartheta^k.$$

por tanto estas secuencias son generadas por el algoritmo RIRHPP aplicado al operador monótono maximal $\Gamma := [(1 - \frac{1}{2\alpha})I + \frac{1}{2\alpha}F]^{-1} - I$, considerando $c = 1$ y el parámetro de relajación $2\alpha\rho$. Luego el resultado se sigue del Teorema 1.2.1. ■

Se observa que los parámetros de acotación $\bar{\lambda}$ y $\hat{\rho}$ son independientes del parámetro θ , y están relacionados por la hipótesis (H1) como

$$\hat{\rho} = \frac{(\bar{\lambda} - 1)^2}{\alpha[2(\bar{\lambda} - 1)^2 + 3\bar{\lambda} - 1]}.$$

luego considerando λ^* como

$$\lambda^* := \frac{\sqrt{8\alpha - 7\alpha^2} - \alpha}{\sqrt{8\alpha - 7\alpha^2} + 3\alpha},$$

se cumple que si $\hat{\rho} = 1$ (donde podemos considerar el parámetro relajado $\rho = 1$) entonces el parámetro inercial λ debe pertenecer a $[0, \lambda^*[$. También se deduce que si no consideramos término inercial ($\lambda = 0$), entonces ρ debe pertenecer a $]0, \alpha^{-1}[$.

En el caso en que el coeficiente average α es $\frac{1}{2}$, el algoritmo es una rescritura de RIRHPP en términos de la resolvente de un operador monótono maximal. En capítulos posteriores este algoritmo será aplicado a operadores α -average (G_s y G_p) con $\alpha > \frac{1}{2}$, en este caso la gráfica 3.1 muestra el comportamiento de las restricciones de los parámetros de acotación $\bar{\lambda}$ y $\hat{\rho}$, cumpliéndose que $\lambda^* < 1/3$.

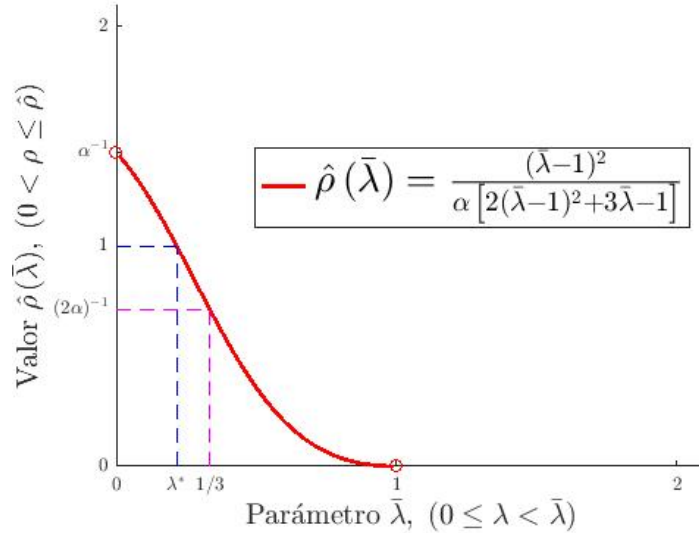


Figura 3.1: Restricciones de los parámetros $\bar{\lambda}$ y $\hat{\rho}$

El considerar el parámetro inercial y relajado juntos ya es interesante de por sí, es por ello que consideramos el siguiente algoritmo, subproducto del anterior, considerando el subproblema como exacto (i.e. escogiendo $\theta = 0$).

Inertial-Relaxed Fixed Point (IRFP)

Inicialización: Escoger $z^0 = z^{-1} \in \mathbb{R}^n$, también $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ y considerar $\hat{\rho} := \frac{\bar{\rho}}{2\alpha}$ donde α es el coeficiente average de F .

Para $k = 0, 1, 2, \dots$ **hacer**

- Escoger $\lambda \in [0, \bar{\lambda}[$ y definir (término inercial)

$$w^k = z^k + \lambda(z^k - z^{k-1}).$$

- Escoger $\rho \in]0, \hat{\rho}]$ y actualizar (término relajado del algoritmo punto fijo)

$$z^{k+1} = (1 - \rho)w^k + \rho F(w^k).$$

Terminar el bucle.

La convergencia es obtenida de la proposición anterior al satisfacer $\bar{\lambda}$ y $\bar{\rho}$ las hipótesis (H_1) , considerando F un operador α -average con dominio total que posee al menos un punto fijo.

3.2. Variante del Método de Punto Proximal

El método proximal es un método clásico e importante que permite el desarrollo de algoritmos más complejos (como los algoritmos de separación), a partir de su relación con él, como se resume en el primer Capítulo. Por lo tanto, variaciones en el método proximal permiten obtener variantes en los métodos relacionados a este.

Dado el problema de inclusión monótona (V_Γ) , en la Subsección 1.2, se presento el algoritmo *MPP* que es una variante del método de punto proximal útil para la obtención de algoritmos de separación, con la ayuda del algoritmo anterior *RIRFP*, desarrollaremos una variante Relative-error Inertial-Relaxed del algoritmo *MPP*, esto por medio de la relación que tiene con el método de punto fijo aplicado a G_D^Γ , un operador $\frac{1}{2}$ -average, obteniendo el siguiente algoritmo.

Relative-error Inertial-Relaxed Multiscaling Proximal Projection (RIRMPP)

Inicialización: Escoger $z^0 = z^{-1} \in \mathbb{R}^r$, P una $r \times r$ matriz simétrica semidefinida positiva, $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ y $\theta \in [0, 1[$.

Para $k = 0, 1, 2, \dots$ **hacer**

- Escoger $\lambda \in [0, \bar{\lambda}[$ y definir

$$z_w^k = z^k + \lambda(z^k - z^{k-1})$$

- Subproblema inexacto:

Encontrar $(\tilde{z}^k, \varepsilon^k) \in \mathbb{R}^r \times \mathbb{R}^r$ tales que

$$\tilde{z}^k = J_P^\Gamma(z_w^k + \varepsilon^k) \quad \text{y} \quad \|\varepsilon^k\|_P^2 \leq \theta^2 \left(\|\tilde{z}^k - z_w^k\|_P^2 + \|\mu^k\|_P^2 \right) \quad (3.2)$$

considerando $\mu^k := z_w^k + \varepsilon^k - \tilde{z}^k$.

- Si $P\mu^k = 0$, **PARAR**. Caso contrario escoger $\rho \in]0, \bar{\rho}]$ y actualizar

$$z^{k+1} = z_w^k - \rho \frac{\langle z_w^k - \tilde{z}^k, P\mu^k \rangle}{\|\mu^k\|_P^2} \mu^k.$$

Terminar el bucle.

La siguiente proposición muestra su convergencia.

Proposición 3.2.1 Dado $\Gamma : \mathbb{R}^r \rightrightarrows \mathbb{R}^r$ un operador monótono maximal y P una $r \times r$ matriz simétrica semidefinida positiva. Asumiendo J_P^Γ univaluado (lo cual implica que es continua) con dominio total y $\text{sol}(V_\Gamma)$ no vacío. Si $(\bar{\lambda}, \bar{\rho})$ satisfacen (H1), entonces

- Si el algoritmo no se detiene, entonces $\{\tilde{z}^k\}$ converge a una solución de (V_Γ) .
- Si el algoritmo se detiene con $P\mu^k = 0$, entonces \tilde{z}^k es una solución de (V_Γ) .

Prueba. a) Sea D una matriz de orden $q \times r$ tal que $P = D^t D$. Si se genera una secuencia, entonces evaluándola con D se tiene $Dz_w^k = Dz^k + \lambda(Dz^k - Dz^{k-1})$, $D\mu^k := Dz_w^k + D\varepsilon^k - D\tilde{z}^k$ y finalmente $Dz^{k+1} = Dz_w^k - \rho \frac{\langle Dz_w^k - D\tilde{z}^k, D\mu^k \rangle}{\|D\mu^k\|^2} D\mu^k$. La desigualdad en (3.2) se reescribe como

$$\|D\varepsilon^k\|^2 \leq \theta^2 \left(\|D\tilde{z}^k - Dz_w^k\|^2 + \|D\mu^k\|^2 \right).$$

Además, como $DJ_P^\Gamma = G_D^\Gamma D$, se tiene $D\tilde{z}^k := G_D^\Gamma(Dz_w^k + D\varepsilon^k)$, por lo tanto estas sucesiones satisfacen el algoritmo RIRFP aplicado al operador G_D^Γ que es $\frac{1}{2}$ -average, luego de la Proposición 3.1.2 se deduce que $D\varepsilon^k$ converge a cero y Dz_w^k converge a un punto fijo de G_D^Γ . Luego como $(\Gamma + P)^{-1} S^t = J_P^\Gamma D^+$, donde D^+ denota la matriz

pseudo inversa (en el sentido de Moore–Penrose) de D , se deduce su continuidad, de donde $\tilde{z}^k = J_P^\Gamma(z_w^k + \varepsilon^k) = (\Gamma + P)^{-1}D^t(Dz_w^k + D\varepsilon^k)$ converge a un punto perteneciente a $\text{sol}(V_\Gamma)$.

b) Si se detiene el algoritmo con $P\mu^k = 0$, entonces $\Gamma(\tilde{z}^k) = P(z_w^k + \varepsilon^k - \tilde{z}^k) = P\mu^k = 0$, del cual se sigue la afirmación. ■

Observación 3.2.1 *El subproblema inexacto (3.2) se puede describir como:*

Encontrar $(\tilde{z}^k, \mu^k) \in \mathbb{R}^r \times \mathbb{R}^r$ tal que

$$P\mu^k \in T(\tilde{z}^k), \quad \|\mu^k + \tilde{z}^k - w^k\|_P^2 \leq \theta^2 \left(\|\tilde{z}^k - w^k\|_P^2 + \|\mu^k\|_P^2 \right). \quad (3.3)$$

Considerando $P = \frac{1}{c_k}I$ y $v^k = \frac{1}{c_k}\mu^k$ reobtenemos el algoritmo RIRHPP. En el caso que P sea invertible, al considerar el algoritmo anterior sin término inercial ($\lambda = 0$), reobtenemos el algoritmo HPP, considerando la variable $\tilde{v}^k = P\mu^k$ y los parámetros $cM = P^{-1}$ y $\tau = \rho$. Por lo tanto, este algoritmo se puede considerar como una extensión de los algoritmos MPP, RIRHPP y HPP.

Análogamente al algoritmo IRFP, consideramos el caso donde el subproblema es exacto $\varepsilon^k = 0$ (notar que $\theta = 0$ no implica $\varepsilon^k = 0$), obteniéndose el siguiente algoritmo

Inertial-Relaxed Multiscaling Proximal Point (IRMPP)

Inicialización: Escoger $z^0 = z^{-1} \in \mathbb{R}^n$, P una $r \times r$ matriz simétrica semi-definida positiva, $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$.

Para $k = 0, 1, 2, \dots$ **hacer**

- Escoger $\lambda \in [0, \bar{\lambda}[$ y definir

$$w^k = z^k + \lambda(z^k - z^{k-1})$$

- Escoger $\rho \in]0, \bar{\rho}]$ y actualizar

$$z^{k+1} = (1 - \rho)w^k + \rho J_P^T(w^k)$$

Terminar el bucle.

La convergencia se sigue de la proposición anterior.

Capítulo 4

Variantes Inertial-Relaxed

En este capítulo usaremos los algoritmos *IRMPP* y *IRFP* desarrollados en el capítulo anterior, para incluir en conjunto los parámetros inercial y relajado, dentro de los algoritmos generales *GSS* (presentado en la sección 1.4) y *PD2BCS* (desarrollado en el Capítulo 2), los cuales lidean con los problemas de inclusión monótono (P_V) y (\mathcal{P}) , respectivamente, logrando así algoritmos incluso más generales, lo cual en particular permiten incluir los parámetros Inertial-Relaxed a los algoritmos conocidos que los contenga.

4.1. Composición monótona de dos bloques

Como se sabe, el algoritmo *GSS* brinda un marco unificador de importantes algoritmos de separación que lidean con el problema composición monótona de dos bloques (P_V) . En este esquema general mostraremos que aún podemos obtener una variante Inertial-Relaxed.

La expresión (1.16) muestra la relación de *GSS* con el algoritmo *MPP* aplicado al operador lagrangeano L (definido en (1.8)) considerando una matriz P especial, luego al ser *IRMPP* una versión Inertial-Relaxed del algoritmo *MPP*, considerando el algoritmo *IRMPP* aplicado al operador L con la misma matriz P , obtenemos la siguiente variante de *GSS* :

Inertial-Relaxed GSS

Escoger $(x^0, z^0, u^0) = (x^{-1}, z^{-1}, u^{-1}) \in \mathbb{R}^{n+p+m}$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ y los reales $\bar{\lambda}$ y $\bar{\rho}$ tales que $\lambda \in [0, \bar{\lambda}[$ y $\rho \in]0, \bar{\rho}]$.

$$(x_w^k, z_w^k, u_w^k) = (x^k, z^k, u^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, u^k - u^{k-1}). \quad (4.1)$$

$$z^{k+\frac{1}{2}} = V_2 z_w^k - B^t u_w^k \quad (4.2)$$

$$\tilde{z}^{k+1} = (T + V_2 + B^t M B)^{-1} [z^{k+\frac{1}{2}} - B^t M A x_w^k] \quad (4.3)$$

$$x^{k+\frac{1}{2}} = V_1 x_w^k - \gamma A^t M A x_w^k + (\gamma - 1) A^t M B z_w^k - A^t u_w^k \quad (4.4)$$

$$\tilde{x}^{k+1} = (S + V_1 + A^t M A)^{-1} [x^{k+\frac{1}{2}} - 2\gamma A^t M B \tilde{z}^{k+1}] \quad (4.5)$$

$$\tilde{u}^{k+1} = u_w^k + \gamma M A x_w^k + (1 - \gamma) M A \tilde{x}^{k+1} + M B \tilde{z}^{k+1} \quad (4.6)$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, u_w^k). \quad (4.7)$$

Se obtiene la siguiente proposición de convergencia

Proposición 4.1.1 Sean $\bar{\lambda}$ y $\bar{\rho}$ satisfaciendo la condición (H1). Asumamos que $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ son simétricas, con M definida positiva tal que $V_1 + A^t M A$ y $V_2 + B^t M B$ son definidas positivas. Sea $\gamma \in \mathbb{R}$ tal que una de las condiciones (A1') – (A4') se cumple. Si $\text{sol}(V_L)$ es no vacío, entonces la secuencia $(\tilde{x}^k, \tilde{z}^k, \tilde{u}^k)$ en (4.1)-(4.7) converge a algún elemento de $\text{sol}(V_L)$.

Prueba. Análogamente a la relación (1.16), considerando $y^k = u^k + (\alpha - \gamma + 1)M A x^k + (1 + \beta)M B z^k$, $y_w^k = y^k + \lambda(y^k - y^{k-1})$ e $\tilde{y}^{k+1} = \tilde{u}^{k+1} + (\alpha - \gamma + 1)M A \tilde{x}^{k+1} + (1 + \beta)M B \tilde{z}^{k+1}$, se cumple que las secuencias $\zeta^k = (x^k, z^k, y^k)$ y $\zeta_w^k = (x_w^k, z_w^k, y_w^k)$ provienen del método Inertial-relaxed MPP, esto es

$$\zeta_w^k = \zeta^k + \lambda(\zeta^k - \zeta^{k-1}), \quad J_P^L(\zeta_w^k) = (\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) \quad \text{y} \quad \zeta^{k+1} = \rho J_P^L \zeta_w^k + (1 - \rho)\zeta^k.$$

Luego por la Proposición 3.2.1 se deduce el resultado. ■

Se observa que este algoritmo hereda la propiedad de *RIRHPP*, donde la hipótesis de convergencia requiere que las cotas de los parámetros inercial y relajado $(\bar{\lambda}, \bar{\rho})$ sólo satisfagan la relación (H1), independiente de los otros parámetros.

El algoritmo anterior expresando en términos de las funciones asociadas al problema de Optimización (P) relacionado con (P_V), se rescribe como

Inertial-Relaxed Optimization GSS

Escoger $(x^0, z^0, u^0) = (x^{-1}, z^{-1}, u^{-1}) \in \mathbb{R}^{n+p+m}$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ y los reales $\bar{\lambda}$ y $\bar{\rho}$ tales que $\lambda \in [0, \bar{\lambda}[$ y $\rho \in]0, \bar{\rho}]$.

$$(x_w^k, z_w^k, u_w^k) = (x^k, z^k, u^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, u^k - u^{k-1}). \quad (4.8)$$

$$\tilde{z}^{k+1} \in \operatorname{argmin} \left\{ g(z) + \frac{1}{2} \|Bz + M^{-1}u_w^k + Ax_w^k\|_M^2 + \frac{1}{2} \|z - z_w^k\|_{V_2}^2 \right\} \quad (4.9)$$

$$v^{k+\frac{1}{2}} = \gamma Ax_w^k - (\gamma - 1)Bz_w^k + M^{-1}u_w^k \quad (4.10)$$

$$\tilde{x}^{k+1} \in \operatorname{argmin} \left\{ f(x) + \frac{1}{2} \|Ax + v^{k+\frac{1}{2}} + 2\gamma B\tilde{z}^{k+1}\|_M^2 + \frac{1}{2} \|x - x_w^k\|_{V_1}^2 \right\} \quad (4.11)$$

$$\tilde{u}^{k+1} = u_w^k + M(\gamma Ax_w^k + (1 - \gamma)A\tilde{x}^{k+1} + B\tilde{z}^{k+1}) \quad (4.12)$$

$$(x^{k+1}, z^{k+1}, u^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{u}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, u_w^k). \quad (4.13)$$

La convergencia se sigue, al satisfacer las hipótesis de la proposición anterior, considerando que las funciones asociadas al problema (P) satisfacen la condición de regularidad (H) .

Como subproducto, al considerar $\gamma = 1$ y $\gamma = 0$, se obtienen las variantes Inertial-Relaxed de $PPDS$ (mostrado en la sección 1.4) y en consecuencia la variante Inertial-Relaxed de los dos tipos de algoritmos (en el sentido Gauss-Seidel y Jacobi, respectivamente) presentados por Shefi-Teboulle [24].

4.2. Composición monótona generalizada de dos bloques

Respecto al problema (\mathcal{V}) , considerando C_1 y C_2 ambos con constante de coercividad positivos, en el Capítulo 2 se desarrolló dos tipos de algoritmos $PD2BCS$ versión Gauss Seidel y Jacobi, siendo estos asociados al método de punto fijo aplicado a G_s y G_p respectivamente, los cuales son operadores average con dominio total (ver Teorema 2.1.1 y 2.1.2). Por tanto considerando $IRFP$, una variante del método de punto fijo, aplicados a los operadores G_s y G_p , obtenemos las variantes inercial-relajado de $PD2BCS$ versión Gauss Seidel y Jacobi.

Considerando la relación (2.24), y el algoritmo $IRFP$ aplicado a G_s , construimos las secuencias $\zeta^k := (x^k, z^k - v^k, y^k)$, $\zeta_w^k := (x_w^k, z_w^k - v_w^k, y_w^k)$ y $\tilde{\zeta}^k := (\tilde{x}^k, \tilde{z}^k - \tilde{v}^k, \tilde{y}^k)$ de manera iterativa satisfaciendo

$$\zeta_w^k = \zeta^k + \lambda(\zeta^k - \zeta^{k-1}), \quad \tilde{\zeta}^{k+1} = \mathcal{F}_s \widehat{Q}^t \widehat{Q}(\zeta_w^k), \quad \zeta^{k+1} = (1 - \rho)\zeta_w^k + \rho\tilde{\zeta}^{k+1} \quad (4.14)$$

tomando en cuenta la constante average de G_s obtenido en el Teorema 2.1.1 para estimar cotas de los parametros inercial y relajado (λ, ρ) , se obtiene la siguiente variante inercial relajado de $PD2BCS$ Gauss Seidel version

**Inertial-Relaxed Primal-dual Two-block compositive splitting,
Gauss-Seidel version (IRPD2BCS, Gauss-Seidel version)**

Escoger $(x^0, z^0, y^0, \nu^0) = (x^{-1}, z^{-1}, y^{-1}, \nu^{-1}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p$, $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ tales que $V_1 + A^t M A$ y $V_2 + B^t M B$ son matrices invertibles y, parámetros $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ tales que $\lambda \in [0, \bar{\lambda}[$ y $\rho \in]0, \bar{\rho}\gamma]$, donde

$$\gamma := 1 - \frac{\|(V_1 + A^t M A)^{-1}\|}{4\beta_1} - \frac{\|(V_2 + B^t M B)^{-1}\|}{4\beta_2}.$$

Generar la secuencia de manera iterativa

$$(x_w^k, z_w^k, y_w^k, \nu_w^k) = (x^k, z^k, y^k, \nu^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}, \nu^k - \nu^{k-1}) \quad (4.15)$$

$$\tilde{z}^{k+1} = (T + V_2 + B^t M B)^{-1}(V_2(z_w^k - \nu_w^k) - B^t M A x_w^k - B^t y_w^k) \quad (4.16)$$

$$\tilde{y}^{k+1} = y_w^k + M A x_w^k + M B \tilde{z}^{k+1} \quad (4.17)$$

$$r^{k+1} = C_1((V_1 + A^t M A)^{-1}(V_1 x_w^k - A^t M B \tilde{z}^{k+1})) \quad (4.18)$$

$$\tilde{\nu}^{k+1} = (V_2 + B^t M B)^{-1} C_2(\tilde{z}^{k+1}) \quad (4.19)$$

$$\zeta^{k+1} = V_1 x_w^k - A^t M B(\tilde{z}^{k+1} - \tilde{\nu}^{k+1}) - A^t \tilde{y}^{k+1} - r^{k+1} \quad (4.20)$$

$$\tilde{x}^{k+1} = (S + V_1 + A^t M A)^{-1}(\zeta^{k+1}) \quad (4.21)$$

$$(x^{k+1}, z^{k+1}, y^{k+1}, \nu^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}, \tilde{\nu}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, y_w^k, \nu_w^k). \quad (4.22)$$

Se obtiene la siguiente proposición de convergencia

Proposición 4.2.1 Sean $\bar{\lambda}$ y $\bar{\rho}$ satisfaciendo la condición (H1). Asumiendo que $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ son simétricas, con V_1 y V_2 semidefinidas positivas y M definida positiva tales que $V_1 + A^t M A$ y $V_2 + B^t M B$ son definidas positivas, satisfaciendo

$$\frac{\|(V_1 + A^t M A)^{-1}\|}{2\beta_1} + \frac{\|(V_2 + B^t M B)^{-1}\|}{2\beta_2} < 1$$

Si $\text{sol}(V_{\mathcal{L}})$ es no vacío, entonces la secuencia $(\tilde{x}^k, \tilde{z}^k, \tilde{y}^k - M B \tilde{\nu}^k)$ generada por (4.15)-(4.22) converge a una solución de $(V_{\mathcal{L}})$.

Prueba. A partir de (4.14), se cumple que al evaluarlo con la matriz \widehat{Q} , se obtiene el algoritmo *IRFP* aplicado a G_s que por el Teorema 2.1.1 es α -average,

luego por la convergencia de *IRFP* (Proposición 3.1.2) y la estructura de los puntos fijos de G_s dado en (2.6), se deduce que $\widehat{Q}\zeta_w^k$ igual a

$$\begin{pmatrix} V_1^{\frac{1}{2}}x_w^k \\ V_2^{\frac{1}{2}}(z_w^k - \nu_w^k) \\ M^{\frac{1}{2}}Ax_w^k + M^{-\frac{1}{2}}y_w^k \end{pmatrix} \text{ converge a } \begin{pmatrix} V_1^{\frac{1}{2}}x^* \\ V_2^{\frac{1}{2}}(z^* - WC_2(z^*)) \\ M^{\frac{1}{2}}Ax^* + M^{-\frac{1}{2}}(y^* + MBWC_2(z^*)) \end{pmatrix},$$

donde $(x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}})$ y $W := (V_2 + B^tMB)^{-1}$. Luego al ser $(T + V_2 + B^tMB)^{-1}$ univaluado y continuo, de (4.16) se obtiene que \tilde{z}^k converge a $(T + V_2 + B^tMB)^{-1}((V_2 + B^tMB)z^* - C_2(z^*) - B^ty^*) = z^*$. De esto, por (4.17), (4.18) y (4.19), se tiene que \tilde{y}^k , r^k y $\tilde{\nu}^k$ convergen a $y^* + MBWC_2(z^*)$, $C_1(x^*)$ y $WC_2(z^*)$, respectivamente. Finalmente por la unicidad y continuidad de $(S + V_1 + A^tMA)^{-1}$, a partir de (4.21) se deduce que \tilde{x}^k converge a x^* . ■

De manera análoga se obtiene la variante inercial-relajado de IRPD2BCS, versión Jacobi. Teniendo en cuenta la relación (2.40), y el algoritmo *IRFP* aplicado a G_p , se genera las secuencias $\varsigma^k := (x^k, z^k, y^k)$, $\varsigma_w^k := (x_w^k, z_w^k, y_w^k)$ y $\tilde{\varsigma}^k := (\tilde{x}^k, \tilde{z}^k, \tilde{y}^k)$ de manera iterativa satisfaciendo

$$\varsigma_w^k = \varsigma^k + \lambda(\varsigma^k - \varsigma^{k-1}), \quad \varsigma^{k+1} = \mathcal{F}_p \widehat{D}^t \widehat{D}(\varsigma_w^k), \quad \varsigma^{k+1} = (1 - \rho)\varsigma_w^k + \rho\varsigma^{k+1} \quad (4.23)$$

donde las cotas superiores de (λ, ρ) considerar la constante average de G_p (dado en el Teorema 2.1.2), obteniendo el siguiente algoritmo paralelo

Inertial-Relaxed Primal-dual Two-block compositive splitting, Jacobi version (IRPD2BCS, Jacobi version)

Escoger $(x^0, z^0, y^0) = (x^{-1}, z^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, $R_1 \in \mathbb{R}^{n \times n}$, $R_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ tales que $R_1 + 2A^tMA$ y $R_2 + 2B^tMB$ son matrices invertibles y, parámetros $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ tales que $\lambda \in [0, \bar{\lambda}[$ y $\rho \in]0, \bar{\rho}\gamma]$, donde

$$\gamma := 1 - \frac{\|(R_1 + 2A^tMA)^{-1}\|}{4\beta_1} - \frac{\|(R_2 + 2B^tMB)^{-1}\|}{4\beta_2}.$$

Generar la secuencia de manera iterativa

$$(x_w^k, z_w^k, y_w^k) = (x^k, z^k, y^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}) \quad (4.24)$$

$$r_1^k = C_1((R_1 + 2A^tMA)^{-1}((R_1 + A^tMA)x_w^k - A^tMBz_w^k)) \quad (4.25)$$

$$\tilde{x}^{k+1} = (S + R_1 + 2A^tMA)^{-1}((R_1 + A^tMA)x_w^k - A^tMBz_w^k - A^ty_w^k - r_1^k) \quad (4.26)$$

$$r_2^k = C_2((R_2 + 2B^tMB)^{-1}((R_2 + B^tMB)z_w^k - B^tMAx_w^k)) \quad (4.27)$$

$$\tilde{z}^{k+1} = (T + R_2 + 2B^tMB)^{-1}((R_2 + B^tMB)z_w^k - B^tMAx_w^k - B^ty_w^k - r_2^k) \quad (4.28)$$

$$\tilde{y}^{k+1} = y_w^k + MA\tilde{x}^{k+1} + MB\tilde{z}^{k+1} \quad (4.29)$$

$$(x^{k+1}, z^{k+1}, y^{k+1}) = \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, y_w^k) \quad (4.30)$$

Se cumple la siguiente proposición de convergencia.

Proposición 4.2.2 Sean $\bar{\lambda}$ y $\bar{\rho}$ satisfaciendo la condición (H1). Asumiendo que $R_1 \in \mathbb{R}^{n \times n}$, $R_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ son simétricas, con R_1 y R_2 semidefinidas positivas y M definida positiva tales que $R_1 + 2A^tMA$ y $R_2 + 2B^tMB$ son definidas positivas, satisfaciendo

$$\frac{\|(R_1 + 2A^tMA)^{-1}\|}{2\beta_1} + \frac{\|(R_2 + 2B^tMB)^{-1}\|}{2\beta_2} < 1$$

Si $\text{sol}(V_{\mathcal{L}})$ es no vacío, entonces la secuencia $(\tilde{x}^k, \tilde{z}^k, \tilde{y}^k)$ generada por (4.24)-(4.29) converge a una solución de $(V_{\mathcal{L}})$.

Prueba. Al evaluar en \widehat{D} la relación (4.23), se obtiene el algoritmo *IRFP* aplicado a G_p de donde se deduce que

$$\widehat{D} \begin{pmatrix} x_w^k \\ z_w^k \\ y_w^k \end{pmatrix} = \begin{pmatrix} R_1^{\frac{1}{2}} x_w^k \\ R_2^{\frac{1}{2}} z_w^k \\ M^{\frac{1}{2}} A x_w^k - M^{\frac{1}{2}} B z_w^k \\ M^{-\frac{1}{2}} y_w^k \end{pmatrix} \text{ converge a } \begin{pmatrix} R_1^{\frac{1}{2}} x^* \\ R_2^{\frac{1}{2}} z^* \\ M^{\frac{1}{2}} A x^* - M^{\frac{1}{2}} B z^* \\ M^{-\frac{1}{2}} y^* \end{pmatrix},$$

donde $(x^*, z^*, y^*) \in \text{sol}(V_{\mathcal{L}})$. Por la continuidad de C_2 y C_1 , de (4.25) y (4.27) se obtiene que $\{r_1^k\}$ y $\{r_2^k\}$ convergen a $C_1(x^*)$ y $C_2(z^*)$ respectivamente. Luego al ser $(S + R_1 + 2A^tMA)^{-1}$ y $(T + R_2 + 2B^tMB)^{-1}$ univaluados y continuos, de (4.26) y (4.28) se obtiene que $\{\tilde{x}^k\}$ y $\{\tilde{z}^k\}$ converge a x^* y z^* respectivamente. Finalmente de (4.29) se cumple que $\{y_w^k\}$ converge a y^* . ■

La paralelización del algoritmo anterior, no produce una versión switching distinta. Contrariamente usando la simetría del problema (\mathcal{V}) , se obtiene una versión switching distinta del algoritmo *IRPD2BCS*, *Gauss-Seidel version* al intercambiar el rol de operadores y las matrices.

En particular de la versión *Gauss-Seidel* obtenemos la variante inercial-relajada multiscaling del método Forward-Backward y de la versión switching la variante inercial-relajada multiscaling del método Backward-Forward. Es interesante observar que la variante inercial-relajada Forward-Backward también se obtuvo recientemente en [3], el cual es deducido alternativamente a partir de una variante inercial del método proximal dado por Solodov [20], considerando el término Elergement.

Observación 4.2.1 *Al considerar la aplicación de los algoritmos desarrollados al problema (Var) (que se reescribe en el marco del problema (V) considerando $B = -I$, $C_1 = C$ y $C_2 = 0$), se debe tener en cuenta la observación 2.1.1, a partir de cual se deduce que se puede prescindir del coeficiente de coercividad β_2 , relajando las hipótesis de las proposiciones 4.2.1 y 4.2.2 como : las matrices parámetro deben satisfacer*

$$\frac{\|(V_1 + A^t M A)^{-1}\|}{2\beta} < 1 \quad y \quad \frac{\|(R_1 + 2A^t M A)^{-1}\|}{2\beta} < 1$$

respectivamente. El valor de β_2 también es prescindido de la definición de γ en los algoritmos, redefinido como

$$\gamma := 1 - \frac{\|(V_1 + A^t M A)^{-1}\|}{4\beta} \quad y \quad \gamma := 1 - \frac{\|(R_1 + 2A^t M A)^{-1}\|}{4\beta}$$

en la versión Gauss-Seidel y Jacobi respectivamente.

En particular se reobtiene el algoritmo GSA3O y su versión switching, a partir de IRPD2BCS, Gauss-Seidel versión sin considerar el término inercial ($\alpha = 0$); para ello se escoge $V_2 = 0$ reobteniendo GSA3O, Form II, e intercambiando S y T , A y B , C_1 y C_2 , escogiendo $V_1 = 0$ reobtenemos GSA3O, Form I.

4.2.1. Variantes de los algoritmos de Condat-Vũ y PD3O

Independientemente de las estructuras matriciales de A y B , una práctica forma de escoger las matrices parámetros en los algoritmos PD2BCS, de tal modo que satisfagan las hipótesis de convergencia y que los subproblemas asociados a S y T , sean subproblemas proximales clásicos, son escogerlos como :

Dados σ, τ_1, τ_2 positivos tales que $\sigma\tau_1 \|A\|^2 \leq 1$, $\sigma\tau_2 \|B\|^2 \leq 1$ y $\frac{\tau_1}{2\beta_1} + \frac{\tau_2}{2\beta_2} < 1$, considerar

$$M = \sigma I_{m \times m}, \quad V_1 = \tau_1^{-1} I_{n \times n} - \sigma A^t A \quad y \quad V_2 = \tau_2^{-1} I_{p \times p} - \sigma B^t B \quad (4.31)$$

para el algoritmo PD2BCS Gauss-Seidel versión.

Dados σ, τ_1, τ_2 positivos tales que $2\sigma\tau_1 \|A\|^2 \leq 1$, $2\sigma\tau_2 \|B\|^2 \leq 1$ y $\frac{\tau_1}{2\beta_1} + \frac{\tau_2}{2\beta_2} < 1$, considerar

$$M = \sigma I_{m \times m}, \quad R_1 = \tau_1^{-1} I_{n \times n} - 2\sigma A^t A \quad y \quad R_2 = \tau_2^{-1} I_{p \times p} - 2\sigma B^t B \quad (4.32)$$

para el algoritmo PD2BCS Jacobi versión.

Para mostrar claramente las variantes obtenidas en los algoritmos de Condat-Vũ y *PD3O*, a partir de los algoritmos *PD2BCS*, ecogemos las matrices parámetros como en (4.31) o (4.32) y lo aplicamos al problema de optimización (1.19), (considerando $S = \partial f$, $T = \partial g$, $B = -I$, $C_1 = \nabla h$ y $C_2 = 0$).

Comenzamos con la aplicación de *PD2BCS*, *Gauss-Seidel version* al modelo (1.19), escogiendo las matrices parámetros como en (4.31) y $V_2 = 0$, usando $(\sigma T^{-1} + I_{m \times m})^{-1} = I - \sigma(T + \sigma I_{m \times m})^{-1}$, obtenemos el siguiente algoritmo

Inertial-Relaxed Primal-Dual Three Operator, Form II (IRPD3O, Form II)

Escoger $(x^0, y^0) = (x^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$, σ y τ reales positivos, los reales $\bar{\lambda}$ y $\bar{\rho}$ tales que $\lambda \in [0, \bar{\lambda}[$ y $\rho \in]0, \bar{\rho}\gamma]$, donde $\gamma := 1 - \frac{\tau}{4\beta}$

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

Se observa que se reduce el número de variables del algoritmo de donde procede. En comparación con *Condat-Vũ algorithm, Form II*, la diferencia radica en la evaluación de la gradiente de h . En el caso $h = 0$, este algoritmo resulta en la variante inercial de *Condat-Vũ algorithm, Form II*.

Análogamente obtenemos la versión switching del algoritmo anterior al intercambiar los operadores y matrices ($S = \partial g$, $T = \partial f$, $C_1 = 0$, $C_2 = \nabla h$, $A = -I$, $B = A$) en la aplicación de *PD2BCS*, *Gauss-Seidel version*, con los mismos parámetros matriciales intercambiados y considerando además los cambios de variables : $\eta^k = \sigma A x^k + y^k - \sigma z^k - \sigma A r^k$, $\eta_w^k = \sigma A x_w^k + y_w^k - \sigma z_w^k - \sigma A r_w^k$ y $\tilde{\eta}^k = \sigma A \tilde{x}^k + \tilde{y}^k - \sigma \tilde{z}^k - \sigma A \tilde{r}^k$, obtenemos el siguiente algoritmo

Inertial-Relaxed Primal-Dual Three Operator Splitting, Form I (IRPD3O, Form I)

Escoger $(x^0, \eta^0, r^0) = (x^{-1}, \eta^{-1}, r^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$, σ y τ reales positivos, los reales $\bar{\lambda}$

y $\bar{\rho}$ tales que $\lambda \in [0, \bar{\lambda}[$ y $\rho \in]0, \bar{\rho}\gamma]$, donde $\gamma := 1 - \frac{\tau}{4\beta}$.

$$\begin{aligned}
(x_w^k, \eta_w^k, r_w^k) &= (x^k, \eta^k, r^k) + \lambda(x^k - x^{k-1}, \eta^k - \eta^{k-1}, r^k - r^{k-1}) \\
\tilde{x}^{k+1} &= (\tau\partial f + I_{n \times n})^{-1}(x_w^k - \tau A^t \eta_w^k - r_w^k) \\
\tilde{r}^{k+1} &= \tau \nabla h(\tilde{x}^{k+1}) \\
\tilde{\eta}^{k+1} &= (\sigma \partial g^* + I_{m \times m})^{-1}(\eta_w^k + \sigma A(2\tilde{x}^{k+1} - x_w^k) + \sigma A r_w^k - \sigma A \tilde{r}^{k+1}) \\
(x^{k+1}, \eta^{k+1}, r^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{\eta}^{k+1}, \tilde{r}^{k+1}) + (1 - \rho)(x_w^k, \eta_w^k, r_w^k).
\end{aligned}$$

Se observa que este algoritmo es una variante inercial-relajada de *PD3O*, pero considera una variable adicional en comparación con su versión switching. En el caso $h = 0$, el algoritmo resulta en la variante inercial de *Condat-Vũ algorithm, Form I* y la variante relajada de *Inertial Chambolle-Pock Algorithm*.

La convergencia de los algoritmos anteriores se sigue de la proposición 4.2.1, considerando que los parámetros de acotación $\bar{\lambda}$ y $\bar{\rho}$ satisfacen la condición (H1), los reales λ , ρ y los reales positivos σ y τ satisfacen las siguientes relaciones

$$0 \leq \lambda < \bar{\lambda}, \quad 0 < \rho \leq \frac{\bar{\rho}(4\beta - \tau)}{4\beta}, \quad \sigma\tau \|A\|^2 \leq 1 \quad \text{y} \quad \tau < 2\beta.$$

Estas restricciones de los parámetros son menos estrictas que las propuestas por Chambolle-Pock [5] para el algoritmo Inertial Chambolle-Pock y las restricciones para los algoritmos Condat-Vũ. En contraste obtenemos algoritmos que incluyen en conjunto los parámetros Inertial-Relaxed que poseen la misma complejidad de implementación, pero muestran mejores comportamientos numéricos.

Por otro lado al aplicar *PD2BCS, Jacobi version* al problema (1.19), escogiendo las matrices parámetros como en (4.32) y $R_2 = 0$, se obtiene el siguiente algoritmo cuyos subproblemas resultan en el cálculo de las resolventes clásicas hallados en forma paralela.

Inertial-Relaxed Parallel Primal-Dual Three Operator (IRPPD3O)

Escoger $(x^0, \eta^0, r^0) = (x^{-1}, \eta^{-1}, r^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$, σ y τ reales positivos, los reales $\bar{\lambda}$

y $\bar{\rho}$ tales que $\lambda \in [0, \bar{\lambda}[$ y $\rho \in]0, \bar{\rho}\gamma]$, donde $\gamma := 1 - \frac{\tau}{4\beta}$.

$$\begin{aligned}
(x_w^k, z_w^k, y_w^k) &= (x^k, z^k, y^k) + \lambda(x^k - x^{k-1}, z^k - z^{k-1}, y^k - y^{k-1}) \\
r^k &= \nabla h(x_w^k - \tau\sigma A^t A x_w^k + \sigma\tau A^t z_w^k) \\
\tilde{x}^{k+1} &= (\tau\partial f + I_{n \times n})^{-1}(x_w^k - \tau A^t(\sigma A x_w^k - \sigma z_w^k + y_w^k) - \tau r^k) \\
\tilde{z}^{k+1} &= (\partial g + 2\sigma I_{m \times m})^{-1}(\sigma z_w^k + \sigma A x_w^k + y_w^k) \\
\tilde{y}^{k+1} &= y_w^k + \sigma A \tilde{x}^{k+1} - \sigma \tilde{z}^{k+1} \\
(x^{k+1}, z^{k+1}, y^{k+1}) &= \rho(\tilde{x}^{k+1}, \tilde{z}^{k+1}, \tilde{y}^{k+1}) + (1 - \rho)(x_w^k, z_w^k, y_w^k).
\end{aligned}$$

La convergencia se sigue de la proposición 4.2.2, considerando que los parámetros de acotación $\bar{\lambda}$ y $\bar{\rho}$ satisfacen la condición (H1), los reales λ , ρ y los reales positivos σ y τ satisfacen las siguientes relaciones

$$0 \leq \lambda < \bar{\lambda}, \quad 0 < \rho \leq \frac{\bar{\rho}(4\beta - \tau)}{4\beta}, \quad 2\sigma\tau \|A\|^2 \leq 1 \quad \text{y} \quad \tau < 2\beta.$$

4.3. Resultados Numéricos

Consideramos el problema denominado *fused lasso*

$$\min_x \mu_1 \|x\|_1 + \frac{1}{2} \|Qx - b\|_2^2 + \mu_2 \|Ax\|_1, \quad (P_{fl})$$

Escogemos los mismos valores como en [25] : $\mu_1 = 20$, $\mu_2 = 200$, Q una matrix de orden $q \times n$ y b un vector en \mathbb{R}^q generados aleatoriamente, y la matrix A definida por

$$A := \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \dots & \dots & & \\ & & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1) \times n}.$$

Nos limitamos en considerar las dimensiones no tan altas, $n = 400$ y $q = 40$.

Al poseer el problema (P_{fl}) la estructura del problema (1.19), aplicaremos los algoritmos recién propuestos IRPD3O, form I y IRPD3O, form II y IRPPD3O y lo compararemos con el algoritmo Condat-Vũ y Chambolle-Pock (presentados en el primer capítulo). Para ello mediremos el error de aproximación de las iteraciones con la solución de (P_{fl}), esto es, consideraremos $\|x^k - x^*\|$, donde la solución x^* es aproximado como el valor primal de la 18000-iteración del algoritmo $PD3O$.

Comenzamos mostrando la gráfica del número de iteraciones necesarias para obtener una aproximación a una solución de (P_{fl}) con un error menor de 10^{-6} , donde variamos los parámetros τ y σ sin considerar los términos inercial y relajado ($\alpha = 0$ y $\rho = 1$). Además trazamos una curva roja que nos muestra los límites teóricos de convergencia. Se considera como máximo 2000 iteraciones para cada par (τ, σ) de parámetros fijados.

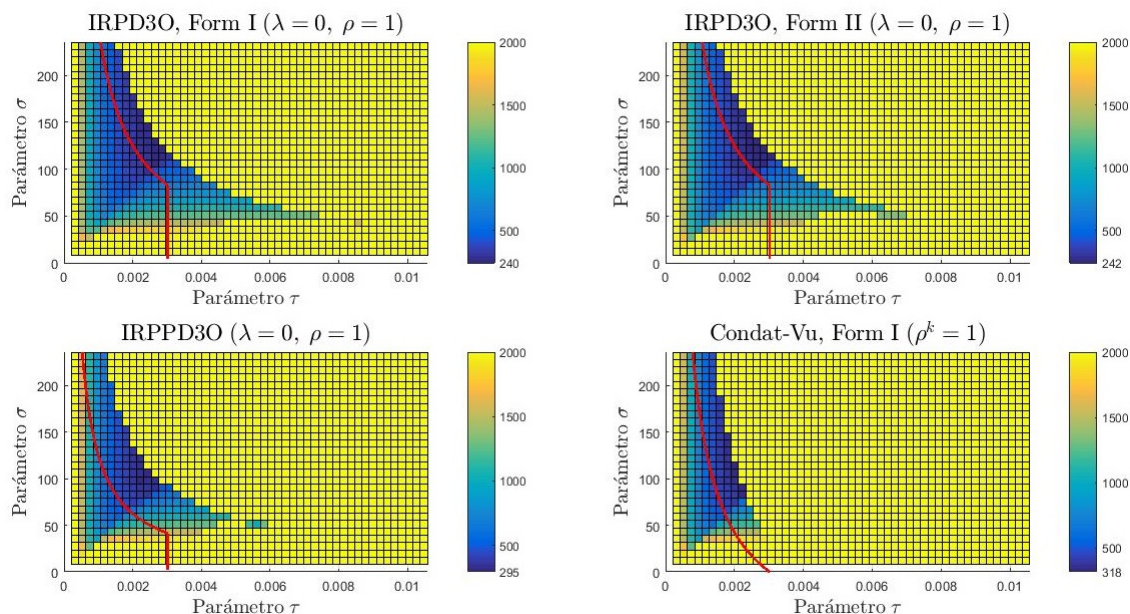


Figura 4.1: Iteraciones necesarias para obtener un error de 10^{-6}

La figura nos muestra que hay una región mayor de convergencia tanto teórica (debajo de la curva roja) como experimental en los algoritmos *IRPD3O* comparado con Condat-Vũ, además de tener zonas dentro de la convergencia teórica que son más rápidas, en especial en la frontera curvada de la línea roja. Por otro lado se observa que las formas I y II del algoritmo *IRPD3O* poseen similar comportamiento (similar a lo observado en las formas I y II de Condat-Vũ)

Las siguientes figuras muestran el número de iteraciones necesarias para obtener una aproximación a una solución de (P_{fl}) con un error menor de 10^{-6} con un máximo de 2000 iteraciones, para cada par $(\sigma(\tau), \tau)$ perteneciente a sus fronteras curvadas respectivamente. En la figura 4.2 variamos los parámetros τ y relajado ρ sin considerar el parámetro inercial ($\lambda = 1$). En la figura 4.3 variamos los parámetros τ e inercial λ , sin considerar el parámetro relajado ($\rho = 1$).

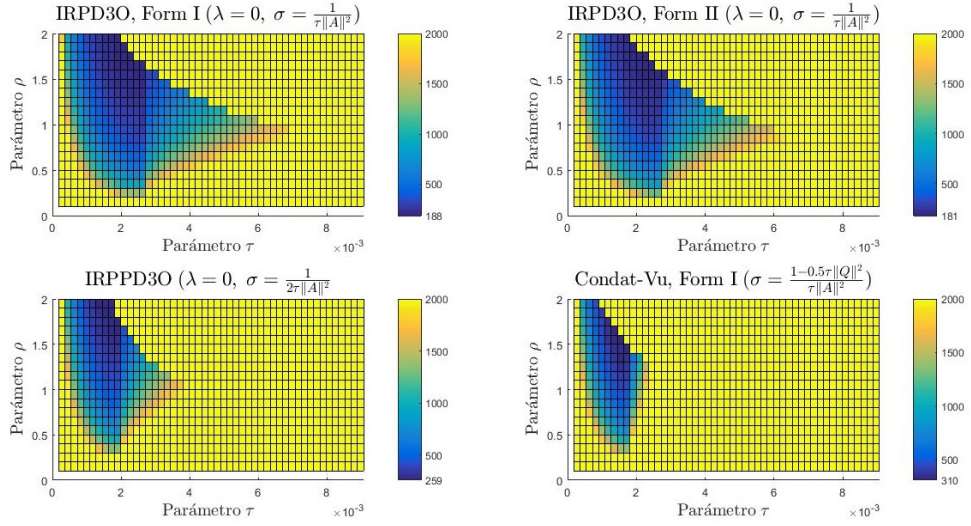


Figura 4.2: Variando el parámetro de relajación para obtener un error de 10^{-6}

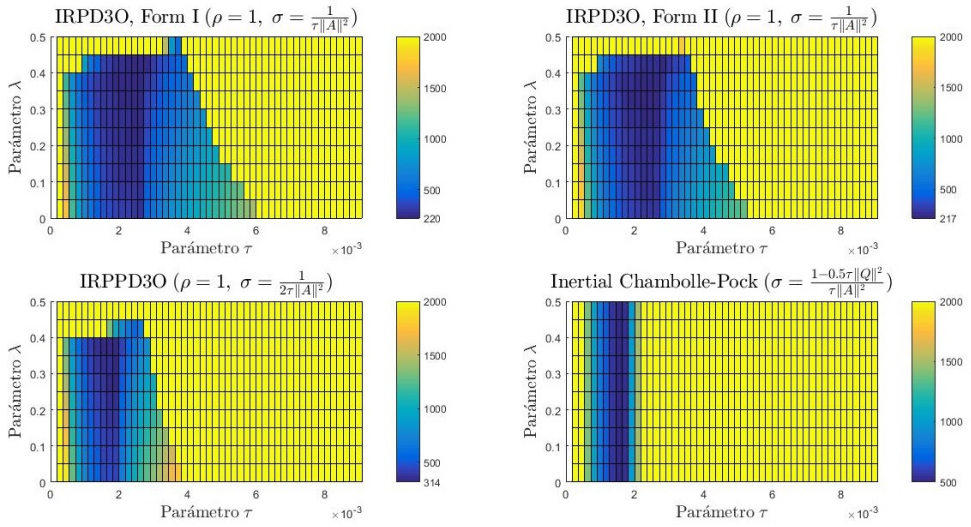


Figura 4.3: Variando el parámetro inercial para obtener un error de 10^{-6}

Se observa que el incremento del parámetro de relajación produce aceleración en todos los algoritmos, siendo este incremento mucho más sensible que lo obtenido al variar el parámetro inercial.

Finalmente variamos el comportamiento de los parámetros inercial y relajado,

para ello consideramos fijos los siguientes puntos incluidos en la frontera roja :

$$\tau = \frac{\gamma_i}{\|Q\|^2} \quad \text{y} \quad \sigma = \frac{1}{\tau \|A\|^2}$$

escogiendo $\gamma_1 = 1$, $\gamma_2 = 1,5$ y $\gamma_3 = 1,99$. Considerando el número de iteraciones necesarias al aplicar el algoritmo IRPD3O, form I para obtener un error menor de 10^{-6} , además graficamos la curva roja de restricciones teóricas de estos parámetros

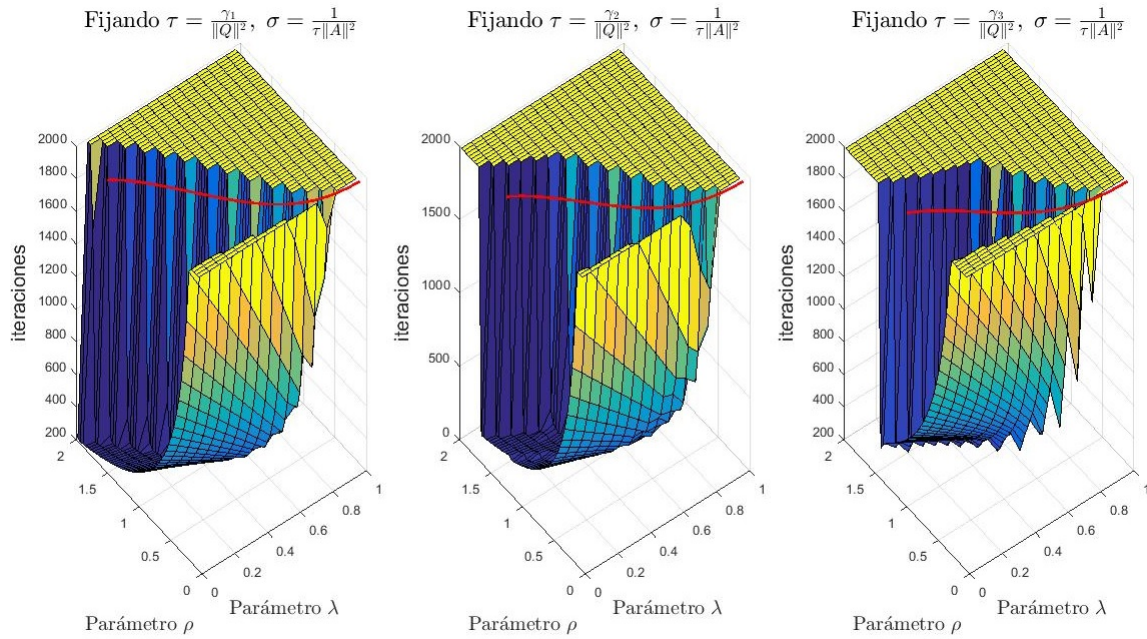


Figura 4.4: Variando λ y ρ para obtener un error de 10^{-6}

Las gráficas anteriores muestran que las restricciones teóricas pueden debilitarse, aún mas, existen zonas fuera de los límites teóricos que muestran un mejor comportamiento. En el Capítulo 6 extenderemos el dominio de convergencia al aplicarlo al problema monótono de dos bloques.

Capítulo 5

Variantes Relative-error Inertial-Relaxed

En el capítulo anterior se logró insertar el parámetro inercial y relajado en los algoritmos separables y generales asociados a dos tipos de problemas. Es de esperar poder también considerar el cálculo inexacto de sus subproblemas asociados, deseando construir un algoritmo análogo a *RIRHPP*. Esto lo podremos realizar de forma parcial, es decir, se trabajará solo con algunos casos particulares de los algoritmos generales, y se podrá considerar el cálculo inexacto de sólo uno de los subproblemas, considerando para ello incluir una subrutina semejante a lo desarrollado por Alves-Eckstein [2], cuando desarrollaron la variante del ADMM aplicado al problema de optimización de la suma de dos funciones.

En estas variantes, consideraremos que existe una subiteración que converge a la solución de uno de los subproblemas (asociado a un mismo operador) originado en cada iteración por el algoritmo de separación, luego en vez de considerarla aislada del algoritmo de separación central, se usa esta subiteración conjuntamente con el otro subproblema para la búsqueda de una solución a una inecuación, lo cual garantizará la convergencia del método.

5.1. Composición monótona

Considerando el problema de composición monótona con la matriz A igual a la identidad, Alves-Eckstein [2], desarrolló una nueva variante del algoritmo Douglas-Rachford admitiendo los parámetros Inertial-Relaxed y el cálculo inexacto de uno de los subproblemas. Esto lo logró apartir del algoritmo *RIRHPP* aplicado al operador monótono maximal asociado al operador de Douglas-Rachford.

De forma análoga, desarrollaremos un algoritmo con las mismas características para el problema composición monótona (siendo A cualquier matriz), para ello consideraremos el algoritmo IRGSS con $\gamma = 1$ (versión Gauus-Sidel) y $V_2 = 0$ o equivalentemente el algoritmo IRGSA3O, Form I (con $C = 0$), el cual posee los parámetros Inertial-Relaxed, esto es:

$$(x_w^k, y_w^k) = (x^k, y^k) + \lambda(x^k - x^{k-1}, y^k - y^{k-1}). \quad (5.1)$$

$$\tilde{z}^k = (T + M)^{-1}(y_w^k + MAx_w^k) \quad (5.2)$$

$$\tilde{y}^k = y_w^k + MAx_w^k - M\tilde{z}^k \quad (5.3)$$

$$\tilde{x}^k = (S + V + A^tMA)^{-1}(Vx_w^k + A^tM\tilde{z}^k - A^t\tilde{y}^k) \quad (5.4)$$

$$(x^{k+1}, y^{k+1}) = (x_w^k, y_w^k) + \rho(\tilde{x}^k - x_w^k, \tilde{y}^k - y_w^k). \quad (5.5)$$

Se observa que la variable \tilde{z}^k es una variable auxiliar, luego este algoritmo en esencia trabaja con menos variables, aún más, podemos reobtenerlo considerando el algoritmo *IRMPP* aplicado a \bar{L} (el operador lagrangiano de dos variables definido en (1.7)) y considerando la matriz $P = \widehat{Q}^t\widehat{Q}$, donde \widehat{Q} es definido en (1.27). Esto equivale a que considerando $\zeta^k := (x^k, y^k)$, $\zeta_w^k := (x_w^k, y_w^k)$, se cumple que

$$\zeta_w^k = \zeta^k + \lambda(\zeta^k - \zeta^{k-1}), \quad J_{\bar{P}}^{\tilde{L}}(\zeta_w^k) = (\tilde{x}^k, \tilde{y}^k) \quad \text{y} \quad \zeta^{k+1} = \rho J_{\bar{P}}^{\tilde{L}}(\zeta_w^k) + (1 - \rho)\zeta_w^k.$$

Ahora en lugar de *IRMPP*, consideraremos la aplicación del algoritmo general *RIRMPP* asociado al mismo operador \bar{L} y la matriz $P = \widehat{Q}^t\widehat{Q}$, obteniéndose una nueva versión donde admita el cálculo inexacto de uno de los subproblemas.

A continuación en el algoritmo (5.1)-(5.5), debilitaremos el cálculo exacto del subproblema

$$\tilde{z}^k = (T + M)^{-1}(y_w^k + MAx_w^k),$$

reemplazándolo por uno inexacto, considerando para ello que alternativamente existe un procedimiento iterativo para resolverlo, siendo este procedimiento iterativo capaz de resolver problemas del tipo: dado $y \in \mathbb{R}^m$ y $x \in \mathbb{R}^n$ arbitrarios,

$$\text{Encontrar } z \in \mathbb{R}^m \text{ tal que } y + MAx \in T(z) + Mz, \quad (5.6)$$

como se define en [2], esto es:

Definición 1 (\mathcal{B} -procedimiento para resolver (5.6)) Un \mathcal{B} -procedimiento para resolver aproximadamente (5.6), es un operador $\mathcal{B}: \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}_{++}^{m \times m} \times \mathbb{N} \rightarrow \mathbb{R}^m \times \mathbb{R}^m$ tales que dado $(y, x, M) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}_{++}^{m \times m}$ arbitrario, se genera la secuencia $(s^l, q^l) = \mathcal{B}(y, x, M, l)$ para todo $l \in \mathbb{N}$, tal que se cumple que $q^l \in T(s^l)$ para todo $l \in \mathbb{N}$ y la secuencia $\{q^l + Ms^l\}_l$ converge a $y + MAx$.

Este \mathcal{B} -procedimiento, incluirá dentro del algoritmo, una subrutina que no se detendrá hasta que se satisfaga una inecuación proveniente de la inecuación (3.2) de *RIRMPP*, reemplazando así la secuencia (5.2)-(5.4) por una subrutina. Luego se debe considerar una actualización del tipo proyección, reemplazando (5.5). En suma, se obtendrá el siguiente algoritmo :

Partially inexact Inertial-Relaxed Primal-Dual (PIRPD)

Inicialización: Escoger $(x^0, y^0) = (x^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$, matrices $V \in \mathbb{R}^{n \times n}$ y $M \in \mathbb{R}^{m \times m}$, reales $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ y $\theta \in [0, 1[$.

Para $k = 0, 1, 2, \dots$ **hacer**

- Escoger $\lambda \in [0, \bar{\lambda}[$ y definir (término inercial)

$$(x_w^k, y_w^k) = (x^k, y^k) + \lambda(x^k - x^{k-1}, y^k - y^{k-1})$$

- **Repetir** { **Para** $l = 1, 2, \dots$ }

Mejorar la solución $(\tilde{z}^k \approx (T + M)^{-1}(y_w^k + MAx_w^k))$

$$\tilde{y}^k \in T(\tilde{z}^k), \quad \tilde{y}^k + M\tilde{z}^k \approx y_w^k + MAx_w^k,$$

considerando

$$(\tilde{z}^{k,l}, \tilde{y}^{k,l}) = \mathcal{B}(y_w^k, x_w^k, M, l). \quad (5.7)$$

Calcular de manera exacta

$$\tilde{x}^{k,l} = (S + V + A^t M A)^{-1} (V x_w^k + A^t M \tilde{z}^{k,l} - A^t \tilde{y}^{k,l}), \quad (5.8)$$

Hasta se satisfaga la siguiente desigualdad:

$$\left\| \widehat{Q} \begin{pmatrix} 0 \\ e^{k,l} \end{pmatrix} \right\|^2 \leq \theta^2 \left[\left\| \widehat{Q} \begin{pmatrix} \tilde{x}^{k,l} - x_w^k \\ \tilde{y}^{k,l} - y_w^k \end{pmatrix} \right\|^2 + \left\| \widehat{Q} \begin{pmatrix} x_w^k - \tilde{x}^{k,l} \\ M \tilde{z}^{k,l} - M A x_w^k \end{pmatrix} \right\|^2 \right],$$

considerando $e^{k,l} := \tilde{y}^{k,l} + M \tilde{z}^{k,l} - y_w^k - M A x_w^k$.

- **Si** $\begin{pmatrix} \tilde{z}^{k,l} - A \tilde{x}^{k,l} \\ V x_w^k - V \tilde{x}^{k,l} \end{pmatrix} = 0$, entonces **PARAR**.

Caso contrario, escoger $\rho \in]0, \bar{\rho}[$ y actualizar

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x_w^k \\ y_w^k \end{pmatrix} - \rho \phi_k \begin{pmatrix} x_w^k - \tilde{x}^{k,l} \\ y_w^k + e^{k,l} - \tilde{y}^{k,l} \end{pmatrix}$$

donde ϕ_k es definido como

$$\phi_k = \frac{\left\langle \widehat{Q} \begin{pmatrix} \tilde{x}^{k,l} - x_w^k \\ \tilde{y}^{k,l} - y_w^k \end{pmatrix}, \widehat{Q} \begin{pmatrix} x_w^k - \tilde{x}^{k,l} \\ M\tilde{z}^{k,l} - MAx_w^k \end{pmatrix} \right\rangle}{\left\| \widehat{Q} \begin{pmatrix} x_w^k - \tilde{x}^{k,l} \\ M\tilde{z}^{k,l} - MAx_w^k \end{pmatrix} \right\|^2}.$$

Terminar el bucle

La ventaja de este algoritmo comparado con el algoritmo 2, de Alves-Eckstein [2], es el de poder trabajar con una matriz coupling A cualquiera, además de poder considerar un parámetro multiscaling M y un parámetro adicional V que permite lidiar con más libertad el subproblema (5.8).

Observación 5.1.1 *Para el caso $A = I$ (suma de dos operadores), podemos reobtener el algoritmo 2, propuesto por Alves-Eckstein [2], considerando los parámetros $M = \gamma^{-1}I$ y $V = 0$, con la única diferencia en el último paso de actualización, pero cumpliéndose en ambos algoritmos la relación*

$$\widehat{Q}\zeta^{k+1} = \widehat{Q}\zeta^k + \rho\phi_k\widehat{Q}(\zeta_w^k - \tilde{\zeta}^k + \varepsilon^k),$$

donde $\zeta^k := (x^k, y^k)$, $\tilde{\zeta}^k := (\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)})$, $\zeta_w^k := (x_w^k, y_w^k)$ y $\varepsilon^k := (0, e^{k,l})$.

Observación 5.1.2 *La desigualdad que debe cumplirse en la subrutina se puede reescribir equivalentemente como:*

$$\begin{aligned} \|\tilde{y}^{k,l} + M\tilde{z}^{k,l} - y_w^k - MAx_w^k\|_{M^{-1}}^2 &\leq \theta^2[\|MA\tilde{x}^{k,l} + \tilde{y}^{k,l} - MAx_w^k - y_w^k\|_{M^{-1}}^2 + \\ &\quad + \|M\tilde{z}^{k,l} - MA\tilde{x}^{k,l}\|_{M^{-1}}^2 + 2\|x_w^k - \tilde{x}^{k,l}\|_V^2] \end{aligned}$$

siendo una versión más amplia comparándolo con la inecuación propuesta en el algoritmo 2, dado por Alves-Eckstein [2].

Antes de mostrar la convergencia del algoritmo anterior, provaremos la siguiente proposición técnica.

Proposición 5.1.1 *Sean $V \in \mathbb{R}^{n \times n}$ y $M \in \mathbb{R}^{m \times m}$ simétricas, con V semidefinida positiva y M definida positiva, tales que $V + A^tMA$ es definida positiva. Si se tiene que*

$$y \in T(z) \quad y \quad x = (S + V + A^tMA)^{-1}(V\tilde{x} + A^tMz - A^ty)$$

entonces se cumple que

$$\begin{pmatrix} x \\ y \end{pmatrix} = (\tilde{L} + \tilde{P})^{-1}\tilde{P} \begin{pmatrix} \tilde{x} \\ \tilde{y} + e \end{pmatrix}$$

donde $e := y + Mz - \tilde{y} - MA\tilde{x}$, la matriz $\tilde{P} := \widehat{Q}^t\widehat{Q}$ con \widehat{Q} definido en (1.27) y el operador Lagrangiano \tilde{L} definido en (1.7).

Prueba. Equivalentemente mostraremos que $(\tilde{L} + \tilde{P}) \begin{pmatrix} x \\ y \end{pmatrix} \ni \tilde{P} \begin{pmatrix} \tilde{x} \\ \tilde{y} + e \end{pmatrix}$. Por un lado se cumple que

$$(\tilde{L} + \tilde{P}) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} S(x) + A^t y + (V + A^t M A)x + A^t y \\ T^{-1}(y) + M^{-1}y \end{pmatrix} \quad (5.9)$$

y también

$$\tilde{P} \begin{pmatrix} \tilde{x} \\ \tilde{y} + e \end{pmatrix} = \begin{pmatrix} V\tilde{x} + A^t M z + A^t y \\ M^{-1}y + z \end{pmatrix} \quad (5.10)$$

luego de las hipótesis se cumple que

$$T^{-1}y \ni z \quad \text{y} \quad S(x) + (V + A^t M A)x + A^t y \ni V\tilde{x} + A^t M z$$

por lo tanto de (5.9) y (5.10) se deduce el resultado. \blacksquare

Se obtiene el siguiente teorema de convergencia

Teorema 5.1.1 Sean $\bar{\lambda}$ y $\bar{\rho}$ satisfaciendo la condición (H1). Asumamos que $V \in \mathbb{R}^{n \times n}$ y $M \in \mathbb{R}^{m \times m}$ son simétricas, con V semidefinida positiva y M definida positiva, tales que $V + A^t M A$ es definida positiva. Si $\text{sol}(V_{\bar{L}})$ es no vacío, entonces se cumple

- Si el algoritmo genera una secuencia (sobre k), es decir, no se detiene, donde cada subrutina (sobre l) termina en un número finito de iteraciones $l = l(k)$, entonces $\{(\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)})\}_k$ converge a una solución de $(V_{\bar{L}})$.
- Si el bucle principal se ejecuta sólo un número finito de veces, terminando en $k = \bar{k}$, donde la subrutina genera una secuencia, entonces $\{(\tilde{x}^{\bar{k},l}, \tilde{y}^{\bar{k},l})\}_l$ converge a una solución de $(V_{\bar{L}})$.
- Si el algoritmo se detiene con $\tilde{z}^{k,l} = A\tilde{x}^{k,l}$ y $Vx_w^k = V\tilde{x}^{k,l}$, entonces $(\tilde{x}^{k,l}, \tilde{y}^{k,l})$ es una solución de $(V_{\bar{L}})$.

Prueba. a) Se generan secuencias sujetas al índice k . Considerando $\zeta^k := (x^k, y^k)$, $\zeta_w^k := (x_w^k, y_w^k)$ y $\varepsilon^k := (0, e^{k,l})$, por la proposición 5.1.1, se cumple

$$\zeta_w^k = \zeta^k + \lambda(\zeta^k - \zeta^{k-1}), \quad J_{\tilde{P}}^{\bar{L}}(\zeta_w^k + \varepsilon^k) = (\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)}).$$

Al ser $\tilde{P} = \widehat{Q}^t \widehat{Q}$ y al terminar cada subrutina en $l = l(k)$, entonces se satisface la desigualdad

$$\|\varepsilon^k\|_{\tilde{P}}^2 \leq \theta^2 \left(\|(\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)}) - \zeta_w^k\|_{\tilde{P}}^2 + \|\zeta_w^k + \varepsilon^k - (\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)})\|_{\tilde{P}}^2 \right),$$

y se actualiza, considerando $\mu^k := \zeta_w^k + \varepsilon^k - (\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)})$, como

$$\zeta^{k+1} = \zeta_w^k - \rho \frac{\langle \widehat{Q}(\zeta_w^k - (\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)})), \widehat{Q}\mu^k \rangle}{\|\mu^k\|_P^2} \mu^k.$$

Luego la convergencia se sigue de la Proposición 3.2.1.

b) En este caso la subrutina con índice l , genera la secuencia $(\tilde{z}^{\bar{k},l}, \tilde{y}^{\bar{k},l}) = \mathcal{B}(y_w^{\bar{k}}, x_w^{\bar{k}}, M, l)$, que por definición se tiene $\tilde{y}^{\bar{k},l} \in T(\tilde{z}^{\bar{k},l})$ y la secuencia $\{\tilde{y}^{\bar{k},l} + M\tilde{z}^{\bar{k},l}\}_l$ converge a $y_w^{\bar{k}} + MAx_w^{\bar{k}}$. Luego por la continuidad de $(T+M)^{-1}$, se cumple que $\{\tilde{z}^{\bar{k},l}\}_l$ converge, lo cual implica la convergencia de $\{\tilde{y}^{\bar{k},l}\}_l$, denotando su limite como z^* y y^* , respectivamente. De esto, junto con (5.8), se cumple que $\{\tilde{x}^{\bar{k},l}\}_l$ converge, digamos a x^* .

Por otro lado, como la subrutina no se detiene, se satisface la desigualdad

$$\left\| \widehat{Q} \begin{pmatrix} 0 \\ e^{\bar{k},l} \end{pmatrix} \right\|^2 > \theta^2 \left[\left\| \widehat{Q} \begin{pmatrix} \tilde{x}^{\bar{k},l} - x_w^{\bar{k}} \\ \tilde{y}^{\bar{k},l} - y_w^{\bar{k}} \end{pmatrix} \right\|^2 + \left\| \widehat{Q} \begin{pmatrix} x_w^{\bar{k}} - \tilde{x}^{\bar{k},l} \\ M\tilde{z}^{\bar{k},l} - MAx_w^{\bar{k}} \end{pmatrix} \right\|^2 \right]$$

y al cumplirse que $e^{\bar{k},l} = \tilde{y}^{\bar{k},l} + M\tilde{z}^{\bar{k},l} - y_w^{\bar{k}} - MAx_w^{\bar{k}}$ converge a zero, se cumple que el segundo miembro de la desigualdad converge a zero, luego se deduce que

$$\widehat{Q} \begin{pmatrix} x_w^{\bar{k}} - x^* \\ Mz^* - MAx_w^{\bar{k}} \end{pmatrix} = \begin{pmatrix} z^* - Ax^* \\ Vx_w^{\bar{k}} - Vx^* \end{pmatrix} = 0.$$

Por lo tanto, (x^*, y^*) pertenece a $\text{sol}(V_{\bar{L}})$.

c) De la relación (5.8) se cumple que $-A^t \tilde{y}^{k,l} \in S(\tilde{x}^{k,l})$ y, como $\tilde{y}^{k,l} \in T(\tilde{z}^{k,l})$, implica que $\tilde{y}^{k,l} \in T(A\tilde{x}^{k,l})$, se obtiene lo afirmado. ■

5.2. Composición monótona más un operador co-coercivo

A continuación extendemos los resultados de la sección anterior al problema composición monótona más un operador co-coercivo. Para ello usamos la relación del algoritmo GSA3O, Form I con el método de punto fijo aplicado al operador average G_1 (ver la Subsección 1.5.1), luego usaremos *RIRFP* (una variante de punto fijo) permitiendonos el cálculo inexacto de uno de los subproblemas, por medio de una subrutina similar al algoritmo anterior (usando el \mathcal{B} -procedimiento), que no se detendrá hasta que se satisfaga una inecuacion proveniente de la inecuación de *RIRFP*, obteniendose así el siguiente algoritmo.

Partially inexact Inertial-Relaxed GSA3O (PIRGSA3O)

Inicialización: Escoger $(x^0, y^0) = (x^{-1}, y^{-1}) \in \mathbb{R}^n \times \mathbb{R}^m$, $V \in \mathbb{R}^{n \times n}$ y $M \in \mathbb{R}^{m \times m}$ tales que $V + A^t M A$ sea inversible, reales $(\bar{\lambda}, \bar{\rho}) \in]0, 1[\times]0, 2[$ y $\theta \in [0, 1[$.

Definir

$$\alpha := \frac{2\beta}{4\beta - \|(V + A^t M A)^{-1}\|}.$$

Para $k = 0, 1, 2, \dots$ hacer

- Escoger $\lambda \in [0, \bar{\lambda}[$ y definir (término inercial)

$$(x_w^k, y_w^k) = (x^k, y^k) + \lambda(x^k - x^{k-1}, y^k - y^{k-1})$$

- **Repetir** { **Para** $l = 1, 2, \dots$ }

Mejorar la solución $(\tilde{z}^k \approx (T + M)^{-1}(y_w^k + M A x_w^k))$

$$\tilde{y}^k \in T(\tilde{z}^k), \quad \tilde{y}^k + M \tilde{z}^k \approx y_w^k + M A x_w^k$$

considerando

$$(\tilde{z}^{k,l}, \tilde{y}^{k,l}) = \mathcal{B}(y_w^k, x_w^k, M, l) \quad (5.11)$$

Calcular exactamente

$$r^{k,l} = C((V + A^t M A)^{-1}(V x_w^k + A^t M \tilde{z}^{k,l})) \quad (5.12)$$

$$\tilde{x}^{k,l} = (S + V + A^t M A)^{-1}(V x_w^k + A^t M \tilde{z}^{k,l} - A^t \tilde{y}^{k,l} - r^{k,l}) \quad (5.13)$$

Hasta se satisfaga la siguiente desigualdad:

$$\left\| \hat{Q} \begin{pmatrix} 0 \\ e^{k,l} \end{pmatrix} \right\|^2 \leq \theta^2 \left[\left\| \hat{Q} \begin{pmatrix} \tilde{x}^{k,l} - x_w^k \\ (2\alpha - 1)e^{k,l} + \tilde{y}^{k,l} - y_w^k \end{pmatrix} \right\|^2 + \left\| \hat{Q} \begin{pmatrix} x_w^k - \tilde{x}^{k,l} \\ M \tilde{z}^{k,l} - M A x_w^k \end{pmatrix} \right\|^2 \right]$$

considerando $e^{k,l} := \tilde{y}^{k,l} + M \tilde{z}^{k,l} - y_w^k - M A x_w^k$.

- **Si** $\begin{pmatrix} \tilde{z}^{k,l} - A \tilde{x}^{k,l} \\ V x_w^k - V \tilde{x}^{k,l} \end{pmatrix} = 0$, entonces **PARAR**.

Caso contrario, escoger $\rho \in]0, \frac{\bar{\rho}}{2\alpha}]$ y actualizar

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x_w^k \\ y_w^k \end{pmatrix} - \rho \phi_k \begin{pmatrix} x_w^k - \tilde{x}^{k,l} \\ y_w^k + e^{k,l} - \tilde{y}^{k,l} \end{pmatrix}$$

donde ϕ_k es definido como

$$\phi_k = \frac{\left\langle \hat{Q} \begin{pmatrix} \tilde{x}^{k,l} - x_w^k \\ (2\alpha - 1)e^{k,l} + \tilde{y}^{k,l} - y_w^k \end{pmatrix}, \hat{Q} \begin{pmatrix} x_w^k - \tilde{x}^{k,l} \\ M \tilde{z}^{k,l} - M A x_w^k \end{pmatrix} \right\rangle}{\left\| \hat{Q} \begin{pmatrix} x_w^k - \tilde{x}^{k,l} \\ M \tilde{z}^{k,l} - M A x_w^k \end{pmatrix} \right\|^2}.$$

Terminar el bucle.

Para obtener el teorema de convergencia, primero probaremos la siguiente proposición

Proposición 5.2.1 Sean $V \in \mathbb{R}^{n \times n}$ y $M \in \mathbb{R}^{m \times m}$ simétricas, con V semidefinida positiva y M definida positiva, tales que $V + A^t M A$ es definida positiva. Si se tiene que :

$$\begin{aligned} y &\in T(z) \\ r &= C((V + A^t M A)^{-1}(V\tilde{x} + A^t M z)) \\ x &= (S + V + A^t M A)^{-1}(V\tilde{x} + A^t M z - A^t y - r) \end{aligned}$$

entonces se cumple que

$$\widehat{Q} \begin{pmatrix} x \\ y \end{pmatrix} = G_1 \widehat{Q} \begin{pmatrix} \tilde{x} \\ \tilde{y} + e \end{pmatrix}$$

donde $e := y + Mz - \tilde{y} - MA\tilde{x}$, la matriz \widehat{Q} definido en (1.27) y el operador average G_1 definido en (1.28).

Prueba. Por la definición (1.27) de \widehat{Q} , se cumple que

$$\widehat{Q} \begin{pmatrix} \tilde{x} \\ \tilde{y} + e \end{pmatrix} = \begin{pmatrix} M^{\frac{1}{2}}z + M^{-\frac{1}{2}}y \\ V^{\frac{1}{2}}\tilde{x} \end{pmatrix},$$

por otro lado al ser $y \in T(z)$, se cumple que $z = (T + M)^{-1}(Mz + y)$, luego de la definición de G_1 , se obtiene que $G_1 \widehat{Q} \begin{pmatrix} \tilde{x} \\ \tilde{y} + e \end{pmatrix}$ es igual a

$$\begin{pmatrix} M^{\frac{1}{2}}z + M^{-\frac{1}{2}}y - M^{\frac{1}{2}}z + M^{\frac{1}{2}}A\tilde{J}^S(V\tilde{x} + 2A^t M z - A^t M z + A^t y - \tilde{C}(V\tilde{x} + A^t M z)) \\ V^{\frac{1}{2}}\tilde{J}^S(V\tilde{x} + 2A^t M z - A^t M z + A^t y - \tilde{C}(V\tilde{x} + A^t M z)) \end{pmatrix}$$

finalmente al cumplirse por la hipótesis que $\tilde{x} = J^S(V\tilde{x} + A^t M z - r)$, se obtiene lo pedido. ■

Se obtiene el siguiente teorema de convergencia.

Teorema 5.2.1 Sean $\bar{\lambda}$ y $\bar{\rho}$ satisfaciendo la condición (H1). Asumamos $V \in \mathbb{R}^{n \times n}$ y $M \in \mathbb{R}^{m \times m}$ ambas simétricas, con V semidefinida positiva y M definida positiva, tales que $V + A^t M A$ es definida positiva y satisfaciendo $\|(V + A^t M A)^{-1}\| \in]0, 2\beta[$. Si $\text{sol}(V_{\hat{L}})$ es no vacío, entonces se cumple :

- a) Si el algoritmo genera una secuencia (sobre k), donde cada subrutina (sobre l) termina en un número finito de iteraciones $l = l(k)$, entonces $\{(\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)})\}_k$ converge a una solución de $(V_{\hat{L}})$.
- b) Si el bucle principal se ejecuta sólo un número finito de veces, terminando en $k = \bar{k}$, donde la subrutina genera una secuencia, entonces $\{(\tilde{x}^{\bar{k},l}, \tilde{y}^{\bar{k},l})\}_l$ converge a una solución de $(V_{\hat{L}})$.
- c) Si el algoritmo se detiene con $\tilde{z}^{k,l} = A\tilde{x}^{k,l}$ y $Vx_w^k = V\tilde{x}^{k,l}$, entonces $(\tilde{x}^{k,l}, \tilde{y}^{k,l})$ es una solución de $(V_{\hat{L}})$.

Prueba. a) Considerando $\zeta^k := (x^k, y^k)$, $\tilde{\zeta}^k := (\tilde{x}^{k,l(k)}, \tilde{y}^{k,l(k)})$, $\zeta_w^k := (x_w^k, y_w^k)$ y $\varepsilon^k := (0, e^{k,l})$, por la proposición 5.2.1 se cumple

$$\widehat{Q}\zeta^k = \widehat{Q}\zeta^k + \lambda\widehat{Q}(\zeta^k - \zeta^{k-1}) \quad \text{y} \quad \widehat{Q}\tilde{\zeta}^k = G_1\widehat{Q}(\zeta_w^k + \varepsilon^k),$$

además se satisface la desigualdad

$$\|\widehat{Q}\varepsilon^k\|^2 \leq \sigma^2 \left(\|\widehat{Q}((2\alpha - 1)\varepsilon^k + \tilde{\zeta}^k - \zeta_w^k)\|^2 + \|\widehat{Q}(\zeta_w^k + \varepsilon^k - \tilde{\zeta}^k)\|^2 \right)$$

y la actualización

$$\widehat{Q}\zeta^{k+1} = \widehat{Q}\zeta_w^k - \rho \frac{\langle \widehat{Q}((2\alpha - 1)\varepsilon^k + \tilde{\zeta}^k - \zeta_w^k), \widehat{Q}(\zeta_w^k - \tilde{\zeta}^k + \varepsilon^k) \rangle}{\|\widehat{Q}(\zeta_w^k - \tilde{\zeta}^k + \varepsilon^k)\|^2} \widehat{Q}(\zeta_w^k - \tilde{\zeta}^k + \varepsilon^k).$$

Entonces las sucesiones son generadas por el algoritmo RIRFP, aplicado a G_1 que es un operador α -average, luego por la Proposición 3.1.2 se cumple que $\widehat{Q}\zeta_w^k$ converge a un punto fijo de G_1 y $\widehat{Q}\varepsilon^k$ converge a 0, lo cual conjuntamente con (5.11)-(5.14), se deduce el resultado.

Los resultados b) y c) se obtienen de manera análoga al teorema anterior. ■

Capítulo 6

Debilitando las hipótesis matriciales de convergencia

Como se observa en los experimentos numéricos, las hipótesis sobre las matrices parámetros asociados con *PD2BCS, Gauss-Seidel version* (V_1, V_2 semidefinidas positivas) y los asociados con *PD2BCS, Jacobi version* (R_1, R_2 semidefinidas positivas) podrían debilitarse. En este capítulo obtendremos la convergencia debilitando estas hipótesis matriciales, pero no abarcaremos todos los casos, por ejemplo no consideramos los términos inercial-relajado.

Para la obtención de estos resultados será necesario considerar otra metodología, puesto que al relajar las hipótesis matriciales no podríamos obtener la convergencia a partir de nuestros operadores G_s y G_p , los cuales usan las hipótesis que V_1, V_2 y R_1, R_2 sean semidefinidas positivas respectivamente en sus definiciones.

6.1. Composición monótona de dos bloques

Considerando el problema de composición monótona de dos bloques, recientemente, en [13], se mostró la convergencia de *PPDS Gauss-Seidel version* considerando $V_1 = 0$, $M = \sigma I_{m \times m}$ y debilitando la hipótesis sobre V_2 : dados σ, τ positivos tales que $\tau\sigma \|B^t B\| < 1$ y $\theta \in]3/4, 1[$, considerar

$$V_2 = \theta(\tau^{-1}I_{p \times p} - \sigma B^t B) - (1 - \theta)\sigma B^t B$$

notar que V_2 no es semidefinida positiva pero aún transforma su subproblema asociado en el cálculo de la resolvente clásica.

Teniendo en cuenta las técnicas de [13], extenderemos el resultado anterior manteniendo el uso de los parámetros matriciales de *PD2BCS* tanto en su versión se-

cuencial y paralela (notar que los versiones de *PPDS* y *PD2BCS* estan relacionados respectivamnete por un cambio de variable, ver sección 1.4). Para ello usaremos las siguientes resultados técnicos.

Proposición 6.1.1 Sean U y W matrices simétricas de orden $n \times n$ tales que W y $U + W$ son semidefinidas positivas. Se cumple que

$$\langle a, Ub \rangle \leq \frac{1}{2} \|a\|_{U+2W}^2 + \frac{1}{2} \|b\|_{U+2W}^2$$

para todo $a, b \in \mathbb{R}^n$.

Prueba. Sean las matrices D y S tales que $D^t D = W$ y $S^t S = U + W$, usando la desigualdad de Cauchy-Schwarz se cumple

$$\langle a, Ub \rangle = \langle Sa, Sb \rangle + \langle D(-a), Db \rangle \leq \frac{1}{2} (\|Sa\|^2 + \|Sb\|^2) + \frac{1}{2} (\|Da\|^2 + \|Db\|^2)$$

de donde se obtiene el resultado ■

Proposición 6.1.2 Sea M una matriz simétrica definida positiva de orden $n \times n$. Dado $\theta \in [0, \frac{1}{4}[$, se cumple que

$$\langle a, b \rangle \leq \left(\frac{1}{2} - \theta\right) \|a\|_M^2 + \left(\frac{1}{2} + 2\theta\right) \|b\|_{M^{-1}}^2,$$

para todo $a, b \in \mathbb{R}^n$.

Prueba. Por la desigualdad de Cauchy-Schwarz se cumple

$$\langle a, b \rangle = \langle M^{\frac{1}{2}} a, M^{-\frac{1}{2}} b \rangle \leq \frac{1}{4} \left(\frac{2}{1+4\theta}\right) \|a\|_M^2 + \left(\frac{1}{2} + 2\theta\right) \|b\|_{M^{-1}}^2,$$

luego, al ser $\frac{1}{2} - \theta \geq \frac{1}{2(1+4\theta)}$ desde que $\theta \in [0, \frac{1}{4}[$, se obtiene el resultado ■

6.1.1. Versión Gauss-Seidel

Consideramos *PD2BCS Gauss-seidel version* aplicado al problema composición monótona de dos bloques (o equivalentemente *GSS* con $\gamma = 1$ y $\rho = 1$)

$$z^{k+1} = (T + V_2 + B^t M B)^{-1} (V_2 z^k - B^t M A x^k - B^t y^k) \quad (6.1)$$

$$y^{k+1} = y^k + M A x^k + M B z^{k+1} \quad (6.2)$$

$$x^{k+1} = (S + V_1 + A^t M A)^{-1} (V_1 x^k - A^t M B z^{k+1} - A^t y^{k+1}) \quad (6.3)$$

La siguiente proposición es la clave en la obtencion de la convergencia de la secuencia, en ella observamos una desigualdad alternativa a la desigualdad producida por la propiedad avarage de G_s .

Proposición 6.1.3 Sean $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ simétricas, con M definida positiva. Asumamos que existe $\theta \in [0, \frac{1}{4}[$ tal que

$$V_1 \quad \text{y} \quad V_2 + \theta B^t M B \quad \text{son semidefinidas positivas}$$

y que las matrices $V_1 + A^t M A$ y $V_2 + B^t M B$ son definidas positivas. La secuencia (x^k, z^k, y^k) generada por (6.1)-(6.3), satisface que la siguiente suma es no negativa:

$$\begin{aligned} & \|x^{k-1} - x^*\|_{V_1}^2 + \|z^k - z^*\|_{V_2 + B^t M B}^2 + \|y^k - y^*\|_{M^{-1}}^2 + \frac{1}{2} \|z^k - z^{k-1}\|_{V_2 + 2\theta B^t M B}^2 \\ & - \|x^k - x^*\|_{V_1}^2 - \|z^{k+1} - z^*\|_{V_2 + B^t M B}^2 - \|y^{k+1} - y^*\|_{M^{-1}}^2 - \frac{1}{2} \|z^{k+1} - z^k\|_{V_2 + 2\theta B^t M B}^2 \\ & - \|x^k - x^{k-1}\|_{V_1}^2 - \|z^{k+1} - z^k\|_{V_2 + \theta B^t M B}^2 - 2\left(\frac{1}{4} - \theta\right) \left[\|z^{k+1} - z^k\|_{B^t M B}^2 + \|y^{k+1} - y^k\|_{M^{-1}}^2 \right] \end{aligned} \quad (6.4)$$

donde $(x^*, z^*, y^*) \in \text{sol}(V_L)$.

Prueba. Por la construcción de la secuencias, se cumple que

$$T(z^{k+1}) \ni V_2(z^k - z^{k+1}) - B^t y^{k+1} \quad (6.5)$$

$$S(x^k) \ni V_1(x^{k-1} - x^k) - A^t M B(z^k - z^{k+1}) - A^t y^{k+1}$$

Considerando $(x^*, z^*, y^*) \in \text{sol}(V_L)$ y por la monotonía de S y T respectivamente se cumple

$$\langle x^k - x^*, V_1(x^{k-1} - x^k) - A^t M B(z^k - z^{k+1}) - A^t y^{k+1} + A^t y^* \rangle \geq 0$$

$$\langle z^{k+1} - z^*, (V_2 + B^t M B)(z^k - z^{k+1}) - B^t M B(z^k - z^{k+1}) - B^t y^{k+1} + B^t y^* \rangle \geq 0$$

Luego sumando las inecuaciones anteriores y (6.2), se tiene

$$\begin{aligned} & \langle x^k - x^*, V_1(x^{k-1} - x^k) \rangle + \langle z^{k+1} - z^*, (V_2 + B^t M B)(z^k - z^{k+1}) \rangle + \\ & \langle M^{-1}(y^{k+1} - y^k), y^* - y^{k+1} \rangle + \langle y^{k+1} - y^k, B(z^{k+1} - z^k) \rangle \geq 0 \end{aligned}$$

Usando que dado una matriz simétrica se cumple $2\langle a - b, U(c - a) \rangle = \|c - b\|_U^2 - \|a - b\|_U^2 - \|a - c\|_U^2$, se obtiene que

$$\begin{aligned} & \|x^{k-1} - x^*\|_{V_1}^2 + \|z^k - z^*\|_{V_2 + B^t M B}^2 + \|y^k - y^*\|_{M^{-1}}^2 \\ & - \|x^k - x^*\|_{V_1}^2 - \|z^{k+1} - z^*\|_{V_2 + B^t M B}^2 - \|y^{k+1} - y^*\|_{M^{-1}}^2 \\ & - \|x^k - x^{k-1}\|_{V_1}^2 - \|z^{k+1} - z^k\|_{V_2 + B^t M B}^2 - \|y^{k+1} - y^k\|_{M^{-1}}^2 \\ & + 2\langle y^{k+1} - y^k, B(z^{k+1} - z^k) \rangle \geq 0 \quad (6.6) \end{aligned}$$

En lo que sigue encontraremos una apropiada cota de $2\langle y^{k+1} - y^k, B(z^{k+1} - z^k) \rangle$. De las hipotesis de V_2 y proposición 6.1.2 se cumple que

$$\left(\frac{1}{2} - \theta\right) \|z^{k+1} - z^k\|_{B^t MB}^2 + \left(\frac{1}{2} + 2\theta\right) \|y^{k+1} - y^k\|_{M^{-1}}^2 \geq \langle B(z^{k+1} - z^k), y^{k+1} - y^k \rangle \quad (6.7)$$

Por otro lado, de (6.5) y por la monotonía de T , se cumple

$$\langle z^{k+1} - z^k, V_2(z^k - z^{k+1}) - B^t y^{k+1} - V_2(z^{k-1} - z^k) + B^t y^k \rangle \geq 0$$

lo cual es rescrito como

$$- \|z^{k+1} - z^k\|_{V_2}^2 + \langle z^{k+1} - z^k, V_2(z^k - z^{k-1}) \rangle \geq \langle y^{k+1} - y^k, B(z^{k+1} - z^k) \rangle$$

usando de las hipótesis de V_2 y la proposición 6.1.1, de la epresion anterior se deduce

$$\frac{1}{2} \|z^{k+1} - z^k\|_{-V_2 + 2\theta B^t MB}^2 + \frac{1}{2} \|z^k - z^{k-1}\|_{V_2 + 2\theta B^t MB}^2 \geq \langle y^{k+1} - y^k, B(z^{k+1} - z^k) \rangle \quad (6.8)$$

Luego sumando (6.8) y (6.7) se obtiene

$$\begin{aligned} & \frac{1}{2} \|z^k - z^{k-1}\|_{V_2 + 2\theta B^t MB}^2 - \frac{1}{2} \|z^{k+1} - z^k\|_{V_2 + 2\theta B^t MB}^2 \\ & \quad - 2\left(\frac{1}{4} - \theta\right) \left[\|z^{k+1} - z^k\|_{B^t MB}^2 + \|y^{k+1} - y^k\|_{M^{-1}}^2 \right] \\ & \quad + \|z^{k+1} - z^k\|_{(1-\theta)B^t MB}^2 + \|y^{k+1} - y^k\|_{M^{-1}}^2 \geq 2\langle B(z^{k+1} - z^k), y^{k+1} - y^k \rangle \end{aligned}$$

A su vez sumando este resultado en (6.6) se demuestra la afirmación. \blacksquare

A partir de la desigualdad de la proposición anterior se obtiene la siguiente convergencia

Teorema 6.1.1 *Asumiendo $V_1 \in \mathbb{R}^{n \times n}$, $V_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ simétricas y cumpliendo : existe $\theta \in [0, \frac{1}{4}[$ satisfaciendo*

$$V_1 \quad y \quad V_2 + \theta B^t MB \quad \text{son semidefinidas positivas ,} \quad (6.9)$$

y M definida positiva tales que $V_1 + A^t MA$ y $V_2 + B^t MB$ son definidas positivas. Si $\text{sol}(V_L)$ es no vacío, entonces la secuencia (x^k, z^k, y^k) generada por (6.1)-(6.3) converge a una solución de (V_L) .

Prueba. La no negatividad de (6.4) de la proposición anterior, se cumple que $\{\|x^{k-1} - x^*\|_{V_1}^2 + \|z^k - z^*\|_{V_2 + B^t MB}^2 + \|y^k - y^*\|_{M^{-1}}^2 + \frac{1}{2} \|z^k - z^{k-1}\|_{V_2 + 2\theta B^t MB}^2\}$ es decreciente del cual se deduce que $\{V_1^{\frac{1}{2}} x^k\}$, $\{z^k\}$ y $\{y^k\}$ son acotadas. También se deduce la convergencia de $\|x^k - x^{k-1}\|_{V_1}^2$, $\|z^{k+1} - z^k\|_{V_2 + \theta B^t MB}^2$, $\|z^{k+1} - z^k\|_{B^t MB}^2$ y

$\|y^{k+1} - y^k\|_{M^{-1}}^2$ todas estas tendiendo a 0, de esto se deduce que $\{V_1^{\frac{1}{2}}(x^k - x^{k-1})\}$, $\{z^{k+1} - z^k\}$ y $\{y^{k+1} - y^k\}$ convergen a 0. Usando lo último en (6.2) se deduce que $\{A(x^k - x^{k-1})\}$ converge a 0 y que $\{Ax^k\}$ es acotado, estos resultados a su vez implican que $\{x^k - x^{k-1}\}$ converge a 0 y que $\{x^k\}$ es acotado.

En suma, se cumple que $\{(x^k, z^k, y^k)\}$ es acotada y que $\{(x^k - x^{k+1}, z^k - z^{k+1}, y^k - y^{k+1})\}$ converge a 0. Luego, dado (x', z', y') un valor de adherencia de (x^k, z^k, y^k) , al cumplirse la relación

$$L(x^{k+1}, z^{k+1}, y^{k+1}) \ni \widehat{Q}^t \widehat{Q}(x^k - x^{k+1}, z^k - z^{k+1}, y^k - y^{k+1}),$$

se deduce de la monotonía maximal de L que $(x', z', y') \in \text{sol}(V_L)$, por lo tanto a partir de (6.4) se obtiene que $\{V_1^{\frac{1}{2}}x^k\}$, $\{z^k\}$ y $\{y^k\}$ converge a $V_1^{\frac{1}{2}}x'$, z' y y' , respectivamente, luego por (6.2) se deduce que $\{Ax^k\}$ converge a Ax' de donde se obtiene finalmente que $\{x^k\}$ converge a x' . ■

6.1.2. Versión Jacobi

Consideramos *PD2BCS Jacobi version* aplicado al problema composición monótona de dos bloques (o equivalentemente *GSS* con $\gamma = 0$ y $\rho = 1$)

$$x^{k+1} = (S + R_1 + 2A^t M A)^{-1} ((R_1 + A^t M A)x^k - A^t M B z^k - A^t y^k) \quad (6.10)$$

$$z^{k+1} = (T + R_2 + 2B^t M B)^{-1} ((R_2 + B^t M B)z^k - B^t M A x^k - B^t y^k) \quad (6.11)$$

$$y^{k+1} = y^k + M A x^{k+1} + M B z^{k+1} \quad (6.12)$$

Gracias a la simetría de la secuencia, podremos mostrar que se pueden debilitar la hipótesis matriciales sobre ambos parámetros R_1 y R_2 , obteniendo su convergencia a partir de la desigualdad que se mostrará en la siguiente proposición, en alternativa de la desigualdad producida por la propiedad average de G_p .

Proposición 6.1.4 *Asumiendo $R_1 \in \mathbb{R}^{n \times n}$, $R_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ simétricas, satisfaciendo la siguiente condición : existe $\theta \in [0, \frac{1}{4}[$ tal que*

$$R_1 + 2\theta A^t M A \quad \text{y} \quad R_2 + 2\theta B^t M B \quad \text{son semidefinidas positivas ,}$$

y M definida positiva tales que $R_1 + 2A^t M A$ y $R_2 + 2B^t M B$ son definidas positivas. Si (x^k, z^k, y^k) es la secuencia generada por (6.1)–(6.3), entonces la siguiente suma

en no negativa:

$$\begin{aligned}
& \|x^k - x^*\|_{R_1+2A^tMA}^2 + \|z^k - z^*\|_{R_2+2B^tMB}^2 + \|y^k - y^*\|_{M^{-1}}^2 + \\
& + \frac{1}{2} \|x^k - x^{k-1}\|_{R_1+4\theta A^tMA}^2 + \frac{1}{2} \|z^k - z^{k-1}\|_{R_2+4\theta B^tMB}^2 + \frac{1}{2} \|A(x^k - x^{k-1}) - B(z^k - z^{k-1})\|_M^2 \\
& \quad - \|x^{k+1} - x^*\|_{R_1+2A^tMA}^2 - \|z^{k+1} - z^*\|_{R_2+2B^tMB}^2 - \|y^{k+1} - y^*\|_{M^{-1}}^2 - \\
& - \frac{1}{2} \|x^{k+1} - x^k\|_{R_1+4\theta A^tMA}^2 - \frac{1}{2} \|z^{k+1} - z^k\|_{R_2+4\theta B^tMB}^2 - \frac{1}{2} \|A(x^{k+1} - x^k) - B(z^{k+1} - z^k)\|_M^2 \\
& \quad - \|x^{k+1} - x^k\|_{R_1+4\theta A^tMA}^2 - \|z^{k+1} - z^k\|_{R_2+4\theta B^tMB}^2 \\
& - (1 - 4\theta) \left[\|x^{k+1} - x^k\|_{A^tMA}^2 + \|z^{k+1} - z^k\|_{B^tMB}^2 + \frac{1}{2} \|y^{k+1} - y^k\|_{M^{-1}}^2 \right], \quad (6.13)
\end{aligned}$$

donde $(x^*, z^*, y^*) \in \text{sol}(V_L)$.

Prueba. La secuencia cumple que

$$S(x^{k+1}) \ni R_1(x^k - x^{k+1}) + A^tM [A(x^k - x^{k+1}) - B(z^k - z^{k+1})] - A^t y^{k+1}$$

y

$$T(z^{k+1}) \ni R_2(z^k - z^{k+1}) + B^tM [B(z^k - z^{k+1}) - A(x^k - x^{k+1})] - B^t y^{k+1}.$$

Dado $(x^*, z^*, y^*) \in \text{sol}(V_L)$, por la monotonía de S y T se obtiene

$$\begin{aligned}
& \langle x^{k+1} - x^*, (R_1 + 2A^tMA)(x^k - x^{k+1}) \rangle + \\
& + \langle MA(x^{k+1} - x^*), -A(x^k - x^{k+1}) - B(z^k - z^{k+1}) - M^{-1}(y^{k+1} - y^*) \rangle \geq 0 \\
& \quad \langle z^{k+1} - z^*, (R_2 + 2B^tMB)(z^k - z^{k+1}) \rangle + \\
& + \langle MA(z^{k+1} - z^*), -B(z^k - z^{k+1}) - A(x^k - x^{k+1}) - M^{-1}(y^{k+1} - y^*) \rangle \geq 0.
\end{aligned}$$

Sumando las dos últimas desigualdades se tiene

$$\begin{aligned}
& \langle x^{k+1} - x^*, (R_1 + 2A^tMA)(x^k - x^{k+1}) \rangle + \langle z^{k+1} - z^*, (R_2 + 2B^tMB)(z^k - z^{k+1}) \rangle + \\
& \langle y^{k+1} - y^k, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle + \langle y^{k+1} - y^k, M^{-1}(y^* - y^{k+1}) \rangle \geq 0.
\end{aligned}$$

Al ser $R_1 + 2A^tMA$, $R_2 + 2B^tMB$ y M^{-1} matrices simétricas, se obtiene

$$\begin{aligned}
& \|x^k - x^*\|_{R_1+2A^tMA}^2 + \|z^k - z^*\|_{R_2+2B^tMB}^2 + \|y^k - y^*\|_{M^{-1}}^2 \\
& \quad - \|x^{k+1} - x^*\|_{R_1+2A^tMA}^2 - \|z^{k+1} - z^*\|_{R_2+2B^tMB}^2 - \|y^{k+1} - y^*\|_{M^{-1}}^2 \\
& \quad - \|x^{k+1} - x^k\|_{R_1+2A^tMA}^2 - \|z^{k+1} - z^k\|_{R_2+2B^tMB}^2 - \|y^{k+1} - y^k\|_{M^{-1}}^2 \\
& \quad \quad 2\langle y^{k+1} - y^k, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle \geq 0. \quad (6.14)
\end{aligned}$$

En lo que sigue obtendremos una cota para $2\langle y^{k+1} - y^k, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle$. Por la proposición 6.1.2 se cumple que

$$\begin{aligned} \left(\frac{1}{2} - \theta\right) \|A(x^{k+1} - x^k) - B(z^{k+1} - z^k)\|_M^2 + \left(\frac{1}{2} + 2\theta\right) \|y^{k+1} - y^k\|_{M^{-1}}^2 \\ \geq \langle y^{k+1} - y^k, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle. \end{aligned}$$

Usando la desigualdad convexa de $\|\cdot\|_M^2$ se obtiene

$$\begin{aligned} (1 - 2\theta) \left[\|x^{k+1} - x^k\|_{A^t M A}^2 + \|z^{k+1} - z^k\|_{B^t M B}^2 \right] + \left(\frac{1}{2} + 2\theta\right) \|y^{k+1} - y^k\|_{M^{-1}}^2 \\ \geq \langle y^{k+1} - y^k, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle \quad (6.15) \end{aligned}$$

Por otro lado, por la monotonía de S y T se deduce que la siguiente suma es no negativa:

$$\begin{aligned} & \langle x^{k+1} - x^k, R_1(x^k - x^{k+1}) - R_1(x^{k-1} - x^k) \rangle + \langle A(x^{k+1} - x^k), y^k - y^{k+1} \rangle + \\ & + \langle M A(x^{k+1} - x^k), A(x^k - x^{k+1}) - B(z^k - z^{k+1}) - [A(x^{k-1} - x^k) - B(z^{k-1} - z^k)] \rangle + \\ & \langle z^{k+1} - z^k, R_2(z^k - z^{k+1}) - R_2(z^{k-1} - z^k) \rangle + \langle B(z^{k+1} - z^k), y^k - y^{k+1} \rangle + \\ & + \langle M B(z^{k+1} - z^k), B(z^k - z^{k+1}) - A(x^k - x^{k+1}) - [B(z^{k-1} - z^k) - A(x^{k-1} - x^k)] \rangle. \end{aligned}$$

Reordenandolo se obtiene

$$\begin{aligned} & - \|x^{k+1} - x^k\|_{R_1}^2 + \langle x^{k+1} - x^k, R_1(x^k - x^{k-1}) \rangle \\ & - \|z^{k+1} - z^k\|_{R_2}^2 + \langle z^{k+1} - z^k, R_2(z^k - z^{k-1}) \rangle - \|A(x^{k+1} - x^k) - B(z^{k+1} - z^k)\|_M^2 \\ & + \langle M [A(x^{k+1} - x^k) - B(z^{k+1} - z^k)], A(x^k - x^{k-1}) - B(z^k - z^{k-1}) \rangle \\ & \geq \langle y^{k+1} - y^k, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle. \end{aligned}$$

Luego por la proposición 6.1.1 y la desigualdad Cauchy-Schwarz se obtiene que

$$\begin{aligned} & \frac{1}{2} \|x^{k+1} - x^k\|_{-R_1 + 4\theta A^t M A}^2 + \frac{1}{2} \|x^k - x^{k-1}\|_{R_1 + 4\theta A^t M A}^2 + \frac{1}{2} \|z^{k+1} - z^k\|_{-R_2 + 4\theta B^t M B}^2 \\ & + \frac{1}{2} \|z^k - z^{k-1}\|_{R_2 + 4\theta B^t M B}^2 - \frac{1}{2} \|A(x^{k+1} - x^k) - B(z^{k+1} - z^k)\|_M^2 + \\ & + \frac{1}{2} \|A(x^k - x^{k-1}) - B(z^k - z^{k-1})\|_M^2 \geq \langle y^{k+1} - y^k, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle. \end{aligned} \quad (6.16)$$

Finalmente, sumando (6.15) y (6.16) se obtiene

$$\begin{aligned}
& \frac{1}{2} \|x^k - x^{k-1}\|_{R_1+4\theta A^t M A}^2 - \frac{1}{2} \|x^{k+1} - x^k\|_{R_1+4\theta A^t M A}^2 \\
& + \frac{1}{2} \|z^k - z^{k-1}\|_{R_2+4\theta B^t M B}^2 - \frac{1}{2} \|z^{k+1} - z^k\|_{R_2+4\theta B^t M B}^2 + \frac{1}{2} \|A(x^k - x^{k-1}) - B(z^k - z^{k-1})\|_M^2 \\
& - \frac{1}{2} \|A(x^{k+1} - x^k) - B(z^{k+1} - z^k)\|_M^2 - 4\left(\frac{1}{4} - \theta\right) \left[\|x^{k+1} - x^k\|_{A^t M A}^2 + \|z^{k+1} - z^k\|_{B^t M B}^2 \right] \\
& - 2\left(\frac{1}{4} - \theta\right) \|y^{k+1} - y^k\|_{M^{-1}}^2 - 4\theta \left[\|x^{k+1} - x^k\|_{A^t M A}^2 + \|z^{k+1} - z^k\|_{B^t M B}^2 \right] + \|x^{k+1} - x^k\|_{2A^t M A}^2 \\
& + \|z^{k+1} - z^k\|_{2B^t M B}^2 + \|y^{k+1} - y^k\|_{M^{-1}}^2 \geq 2\langle y^{k+1} - y^k, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle,
\end{aligned}$$

lo cual, sumándolo con (6.14) se deduce el resultado. \blacksquare

Se obtiene la siguiente convergencia

Teorema 6.1.2 *Asumiendo $R_1 \in \mathbb{R}^{n \times n}$, $R_2 \in \mathbb{R}^{p \times p}$ y $M \in \mathbb{R}^{m \times m}$ simétricas satisfaciendo la siguiente condición: existe $\theta \in [0, \frac{1}{4}[$ satisfaciendo*

$$R_1 + 2\theta A^t M A \quad y \quad R_2 + 2\theta B^t M B \quad \text{son semidefinidas positivas,} \quad (6.17)$$

y M definida positiva tales que $R_1 + 2A^t M A$ y $R_2 + 2B^t M B$ son definidas positivas. Si $\text{sol}(V_L)$ es no vacío, entonces la secuencia (x^k, z^k, y^k) generada por (6.1)-(6.3) converge a una solución de (V_L) .

Prueba. Sea $(x^*, z^*, y^*) \in \text{sol}(V_L)$. Por la desigualdad 6.13 de la proposición anterior se cumple que $\{x^k\}$, $\{z^k\}$ y $\{y^k\}$ son acotadas, además se deduce que $\{\|x^{k+1} - x^k\|_{R_1+4\theta A^t M A}^2\}$, $\{\|x^{k+1} - x^k\|_{A^t M A}^2\}$, $\{\|z^{k+1} - z^k\|_{R_2+4\theta B^t M B}^2\}$, $\{\|z^{k+1} - z^k\|_{B^t M B}^2\}$ y $\|y^{k+1} - y^k\|_{M^{-1}}^2$ convergen hacia 0, de donde se obtiene que $\{x^{k+1} - x^k\}$, $\{z^{k+1} - z^k\}$ y $\{y^{k+1} - y^k\}$ convergen a 0.

Luego, dado (x', z', y') un valor de adherencia de (x^k, z^k, y^k) , al cumplirse la relación

$$L(x^{k+1}, z^{k+1}, y^{k+1}) \ni \widehat{D}^t \widehat{D}(x^k - x^{k+1}, z^k - z^{k+1}, y^k - y^{k+1}),$$

se deduce de la monotonía maximal de L que $(x', z', y') \in \text{sol}(V_L)$. Por lo tanto, a partir de (6.13), se deduce que (x^k, z^k, y^k) converge a (x', z', y') . \blacksquare

Es sabido que escoger las matrices parámetro como en (4.31) y (4.32) para el caso secuencial y paralelo respectivamente, transforman los subproblemas asociados en resolventes clásicos, luego considerando las debilitaciones de las hipótesis de convergencia (6.9) y (6.17) desarrollados en los teoremas anteriores, nos permiten

escogerlos ahora de la siguiente forma :

Dados σ, τ_1, τ_2 positivos y $\theta \in [0, \frac{1}{4}[$, tales que $\sigma\tau_1 \|A\|^2 \leq 1$ y $\sigma(1-\theta)\tau_2 \|B\|^2 \leq 1$, onsiderar

$$M = \sigma I_{m \times m}, \quad V_1 = \tau_1^{-1} I_{n \times n} - \sigma A^t A \quad \text{y} \quad V_2 = \tau_2^{-1} I_{p \times p} - \sigma B^t B \quad (6.18)$$

para el algoritmo *PD2BCS version Gauss-Seidel*.

Dados σ, τ_1, τ_2 positivos y $\theta \in [0, \frac{1}{4}[$, tales que $2\sigma(1-\theta)\tau_1 \|A\|^2 \leq 1$ y $2\sigma(1-\theta)\tau_2 \|B\|^2 \leq 1$, considerar

$$M = \sigma I_{m \times m}, \quad R_1 = \tau_1^{-1} I_{n \times n} - 2\sigma A^t A \quad \text{y} \quad R_2 = \tau_2^{-1} I_{p \times p} - 2\sigma B^t B \quad (6.19)$$

para el algoritmo *PD2BCS version Jacobi*.

Conclusiones

Las importantes contribuciones de este trabajo están distribuidas de la siguiente manera :

En el Capítulo 2 se construye dos tipos de operadores cuyos puntos fijos están relacionadas con la solución de un modelo general, luego gracias a las propiedades de estos operadores se obtienen algoritmos convergentes de separación del tipo secuencial y paralelo.

En el Capítulo 3 se muestra una variante del método de punto fijo y en consecuencia se obtiene la construcción de una versión general del Método de punto proximal, el cual incluye conjuntamente los parámetros inercial y relajado, además considera el cálculo inexacto del subproblema asociado a una resolvente generalizada (reemplazando la matriz identidad por una matriz arbitraria simétrica semidefinida positiva).

En el Capítulo 4 se consideran la inclusión de los parámetros Inercial-Relajado dentro de los dos tipos de algoritmos generales de separación obtenidos en el Capítulo 2, en particular se obtiene las variantes Inercial-Relajado de algunos algoritmos clásicos y los recientemente propuestos.

En el Capítulo 5 desarrollamos una extensión del reciente algoritmo *Relative-Error Inercial-Relajado ADMM*, para así poder aplicarlo a problemas con mayor estructura separable, obteniendo algoritmos separable que no desaprovechan su estructura. Estos algoritmos poseen los parámetros Inercial-Relajado y consideran el cálculo inexacto de uno de sus subproblemas por medio de la inclusión de una subrutina.

Finalmente en el último capítulo mostramos que si debilitamos las hipótesis matriciales, aún se mantiene la convergencia de los algoritmos de separación asociados al problema de composición monótona de dos bloques.

Bibliografía

- [1] F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Analysis*, 9(1-2):3–11, 2001.
- [2] M. M. Alves, J. Eckstein, M. Geremia, and J. G. Melo. Relative-error inertial-relaxed inexact versions of douglas-rachford and admm splitting algorithms. *Computational Optimization and Applications*, pages 1–34, 2019.
- [3] M. M. Alves and R. T. Marcavillaca. On inexact relative-error hybrid proximal extragradient, forward-backward and tseng’s modified forward-backward methods with inertial effects. *Set-Valued and Variational Analysis*, pages 1–25, 2019.
- [4] 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.
- [5] A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal-dual algorithm. *Mathematical Programming*, 159(1-2):253–287, 2016.
- [6] 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.
- [7] 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.
- [8] D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set Valued Var. Anal.*, 25(4):829–858, 2017.
- [9] 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.

- [10] F. Facchinei and J. S. Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer Science Business Media, 2007.
- [11] 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.
- [12] 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.
- [13] Ma F. He, B. and X. Yuan. Optimally linearizing the alternating direction method of multipliers for convex programming. *Computational Optimization and Applications*, 75(2):361–388, 2020.
- [14] Sun D. Li, M. and K.C. Toh. A majorized admm with indefinite proximal terms for linearly constrained convex composite optimization. *SIAM Journal on Optimization*, 26(2):922–950, 2016.
- [15] 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.
- [16] G.J. Minty. Monotone operators in hilbert spaces. *Duke Math. Journal*, 29:341–346, 1962.
- [17] 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.
- [18] D. O’Connor and L. Vandenberghe. On the equivalence of the primal-dual hybrid gradient method and douglas-rachford splitting. *Mathematical Programming*, 2018.
- [19] E. Oré, P. Mahey, and E. Ocaña. A unified splitting algorithm for composite monotone inclusions. Technical report, LIMOS, Université Clermont Auvergne, june 2018.
- [20] L. A. Parente, P. A. Lotito, and M. V. Solodov. A class of inexact variable metric proximal point algorithms. *SIAM Journal on Optimization*, 19(1):240–160, 2008.

- [21] T. Pennanen. Dualization of generalized equations of maximal monotone type. *SIAM Journal on Optimization*, 10(3):809–835, 2000.
- [22] R. T. Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.
- [23] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control and optimization*, 14:877–898, 1976.
- [24] 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.
- [25] M. Yan. A new primal-dual algorithm for minimizing the sum of three functions with a linear operator. *J. Sci. Comput.*, 2018.