



Universidad de Valladolid

Facultad de Ciencias

GRADO EN MATEMÁTICAS

TRABAJO FIN DE GRADO

Algoritmo de Dykstra

Autor: Iñigo San Jose Visiers

Tutor: Luis M. Abia Llera

Curso: 2017-2018

Luis M. Abia Llera, Catedrático de Matemática Aplicada de la Universidad de Valladolid,

CERTIFICA: que la presente Memoria *Algoritmo de Dykstra* ha sido realizada por D. Iñigo San Jose Visiers en la Universidad de Valladolid y constituye la memoria de su trabajo de fin de Grado, preceptiva para la finalización de sus estudios del Grado en Matemáticas de dicha universidad.

Valladolid, 24 de enero de 2018

Fdo.:

Luis M. Abia Llera

Índice general

Introducción	vii
1. Conceptos básicos	1
1.1. Producto interno	1
1.2. Mejor aproximación	5
1.3. Convergencia débil	11
2. Método de proyecciones alternadas	15
2.1. Teoría de operadores	15
2.2. Teorema de Von Neumann	17
3. Algoritmo de Dykstra	23
3.1. Motivación	23
3.2. Algoritmo de Dykstra	24
3.3. Caso de espacios afines	32
4. Aplicación del algoritmo de Dykstra a un problema de mínimos cuadrados con matrices	35
4.1. Explicación del problema	35
4.2. Solución al problema	37
4.3. Experimentos	39
5. ANEXO	43
5.1. Programas auxiliares	43
5.2. Implementación del algoritmo de Dykstra	46

Introducción

Consideremos dos rectas r_1 y r_2 del plano que se cortan en el punto P . Si partiendo de un punto P_0 formamos la sucesión de puntos que se obtiene proyectando ortogonalmente de forma alternada sobre las rectas r_1 y r_2 los puntos que se van obteniendo se recae en una sucesión que converge al punto P . El algoritmo de Dykstra es la generalización de este resultado cuando r_1 y r_2 se reemplazan por dos conjuntos convexos y cerrados K_1 y K_2 de un espacio de Hilbert, con intersección no vacía K , y a partir de un punto P_0 se construye la aproximación óptima a P_0 en $K = K_1 \cap K_2$ resolviendo sucesivamente y de forma alternada problemas de aproximación óptima en K_1 y K_2 . En muchas situaciones prácticas la computación de estas aproximaciones óptimas son relativamente fáciles de obtener: por ejemplo, cuando los K_i son semiespacios, hiperplanos, subespacios de dimensión finita (algoritmo de Von Neumann) o algunas clases de conos.

Los algoritmos de Von Neumann y de Dykstra son ejemplos de métodos de proyecciones alternadas. Con este nombre se denota una clase general de métodos que reducen el problema general de aproximar un elemento de un espacio de Hilbert en una intersección de conjuntos mediante la solución alternada de problemas de aproximación más sencillos sobre cada uno de los conjuntos que forman dicha intersección.

Las técnicas de proyecciones alternadas aparecen frecuentemente en el tratamiento de problemas de aproximación tales como:

1. Solución de sistemas lineales (método de Kaczmarz).
2. Problemas de mínimos cuadrados con restricciones para matrices.
3. Aplicaciones en probabilidad y estadística (aplicación al problema de estimación de matrices de covarianza con un patrón dado).
4. Métodos secante para la aproximación de ceros de sistemas no lineales.
5. Problemas de factibilidad convexa.
6. Problemas de programación convexa.

En cuanto a las aplicaciones de los algoritmos basados en las proyecciones alternadas, éstas son muy diversas: restauración de imágenes, recuperación de

señales, redes neuronales y reconocimiento de patrones, teoría de juegos, diseño de antenas, ciencia de materiales y compresión de datos, entre otras.

El trabajo tiene como objetivo presentar el análisis de los algoritmos de Von Neumann y Dykstra de forma autocontenida desarrollando todos los elementos técnicos necesarios que permiten la prueba y no forman parte del curriculum del plan de estudios del Grado en Matemáticas de la Universidad de Valladolid.

En el capítulo 1 se reúnen además de resultados fundamentales de la teoría de espacios de Hilbert estudiados en la asignatura de *Introducción a los espacios de funciones*, otros resultados como la caracterización de Kolmogorov de la mejor aproximación en un convexo, cerrado, no vacío de un espacio de Hilbert; el concepto de convergencia débil; la compacidad débil de los conjuntos acotados de un espacio de Hilbert y el teorema fuerte de separación para convexos cerrados de un espacio de Hilbert.

El capítulo aborda el teorema de Von Neumann de aproximación óptima a un elemento de un espacio de Hilbert en la intersección de los subespacios cerrados M_1, M_2 de H . La demostración de Von Neumann utiliza las propiedades de los operadores de proyección para establecer que la sucesión de iterantes del producto de dos proyecciones converge puntualmente al operador proyección sobre la intersección de los subespacios.

En el capítulo 3 consideraremos la situación general de la mejor aproximación en un conjunto que es intersección no vacía de un número finito de conjuntos convexos cerrados. Ocurre que la sucesión de las proyecciones alternadas sobre cada uno de los convexos cerrados que forman la intersección no converge hacia la mejor aproximación en la intersección. Dykstra introduce una modificación del algoritmo de Von Neumann que permite recuperar la convergencia y la optimalidad de la aproximación. El teorema de convergencia en este contexto está asociado a los nombres de Dykstra y Boyle. Aunque históricamente la generalización del algoritmo de Von Neumann a más de dos subespacios es anterior al algoritmo de Dykstra, en este capítulo obtenemos dicha generalización como corolario del teorema de Dykstra-Boyle (teorema de Halpenin).

El último capítulo ilustra con experimentos numéricos la efectividad del algoritmo de Dykstra en un problema de aproximación mínimos cuadrados en la norma de Frobenius de una matriz general por una matriz en un conjunto convexo caracterizado por varias restricciones de distinta naturaleza. Los programas Matlab utilizados en estos experimentos se incluyen como anexo a la memoria.

Para terminar con la introducción me gustaría agradecer a mi tutor Luis M. Abia por su dedicación y ayuda, a mi familia y a todos los profesores y amigos de Valladolid, Coimbra y Granada que me han acompañado durante estos *muchos* años y han hecho este viaje más liviano.

Capítulo 1

Conceptos básicos

1.1. Producto interno

Daremos una nociones básicas y propiedades del producto interno que facilitarán la comprensión del trabajo.

Definición 1.1 Sea X un espacio vectorial real. Llamamos **espacio con producto interno** a X si para cada par de elementos x, y de X esta definido un escalar real $\langle x, y \rangle$ con las siguientes propiedades ($\forall x, y, z \in X$ y $\alpha \in \mathbb{R}$)

1. $\langle x, x \rangle \geq 0$.
2. $\langle x, x \rangle = 0$ si y solo si $x = 0$.
3. $\langle x, y \rangle = \langle y, x \rangle$.
4. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$.
5. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.

La función $\langle \cdot, \cdot \rangle$ de $X \times X$ es llamada **producto interno**. De las propiedades (1) y (2) obtenemos que es definida positiva y de la (3) que es simétrica. Usando las propiedades (4) y (5) podemos ver que

$$\left\langle \sum_1^n \alpha_i x_i, y \right\rangle = \sum_1^n \alpha_i \langle x_i, y \rangle$$

para cualesquiera $\alpha_i, i = 1, 2, \dots, n$ reales y $x_i \in X$ De la propiedad anterior y de la (3) concluimos

$$\left\langle x, \sum_1^n \alpha_i y_i \right\rangle = \sum_1^n \alpha_i \langle x, y_i \rangle.$$

De aquí en adelante asumiremos que, salvo especificar lo contrario, X es un **espacio con producto interno real**.

Definición 1.2 Para todo $x \in X$ definimos la **norma** como

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Teorema 1.3 Desigualdad de Schwarz. Para cada par $x, y \in X$,

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

La desigualdad estricta se da si y solo si x e y son linealmente dependientes

Demostración: Consideremos x e y linealmente dependientes, es decir $x = \alpha y$ para algún α escalar no nulo, entonces ambos lados de la igualdad son $|\alpha| \|y\|^2$. En caso de que fueran linealmente independientes, $\forall \lambda$ escalar tenemos que $x - \lambda y \neq 0$, entonces

$$0 < \langle x - \lambda y, x - \lambda y \rangle = \|x\|^2 - 2\lambda \langle x, y \rangle + \lambda^2 \|y\|^2.$$

Haciendo $\lambda = \langle x, y \rangle \|y\|^{-2}$ concluimos que

$$0 < \|x\| - |\langle x, y \rangle|^2 \|y\|^{-2},$$

de donde tenemos la desigualdad estricta. □

Propiedades 1.4 La norma tiene estas propiedades

1. $\|x\| \geq 0$.
2. $\|x\| = 0$ si y solo si $x = 0$.
3. $\|\alpha x\| = |\alpha| \|x\|$.
4. $\|x + y\| \leq \|x\| + \|y\|$.
5. $|\|x\| - \|y\|| \leq \|x - y\|$.

Demostración: Las propiedades (1), (2) y (3) vienen dadas por las propiedades del producto interno. Para demostrar la (4) recurriremos a la desigualdad de Schwarz 1.3

$$\|x+y\|^2 = \langle x+y, x+y \rangle = \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2$$

que implica (4) (Llamada desigualdad triangular).

Usaremos (4) para demostrar (5)

$$\|x\| = \|(x - y) + y\| \leq \|x - y\| + \|y\|$$

tenemos ahora

$$\|x\| - \|y\| \leq \|x - y\|.$$

Intercambiando los papeles de la x e y llegamos a

$$\|y\| - \|x\| \leq \|y - x\|$$

que, combinándolos, llegaríamos a (5). □

Proposición 1.5 Ley del paralelogramo Para todo x, y en un espacio con producto interno

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

Demostración:

$$\|x \pm y\|^2 = \langle x \pm y, x \pm y \rangle = \|x\|^2 \pm 2\langle x, y \rangle + \|y\|^2.$$

Sumando ambos casos llegamos al resultado. □

Vamos a dar unos ejemplos de espacios con **producto interno**.

Ejemplo 1.6 En el espacio \mathbb{R}^n , el producto

$$\langle x, y \rangle = x_1y_1 + \dots + x_ny_n = \sum_i^n x_iy_i$$

define un producto interno, cuya norma es

$$\|x\| = \left(\sum_i^n x_i^2 \right)^{\frac{1}{2}}.$$

Ejemplo 1.7 En el espacio $C^0[a, b]$ de las funciones continuas en el intervalo $[a, b]$, el producto

$$\langle x, y \rangle = \int_a^b x(t)y(t)dt$$

define un producto interno, y su norma viene dada por

$$\|x\| = \left[\int_a^b x^2(t)dt \right]^{\frac{1}{2}}.$$

Ejemplo 1.8 En el espacio $L^2[a, b]$ de las funciones medibles de Lebesgue que son de cuadrado integrable en el intervalo $[a, b]$, el producto

$$\langle x, y \rangle = \int_a^b x(t)y(t)dt$$

define un producto interno, y su norma viene dada por

$$\|x\| = \left[\int_a^b x^2(t)dt \right]^{\frac{1}{2}}.$$

Ejemplo 1.9 En el espacio $\mathbb{R}^{n \times n}$, de las matrices reales $n \times n$,

$$\langle A, B \rangle = \text{traza}(A^T B)$$

define un producto interno y su norma asociada, es la norma de Frobenius, definida

$$\|A_F\| = \left(\text{traza}(A^T A) \right)^{\frac{1}{2}} = \left(\sum_{i,j=1}^n a_{ij}^2 \right)^{\frac{1}{2}}.$$

Definición 1.10 Llamamos a dos vectores x e y **ortogonales**, escrito $x \perp y$, si $\langle x, y \rangle = 0$. Decimos que un vector x es **ortogonal a un conjunto** Y , y escribiremos $x \perp Y$, si para todo $y \in Y$, $x \perp y$. Un conjunto A es llamado **conjunto ortogonal** si para cualquier par de vectores $x, y \in A$, se tiene $x \perp y$. Un **conjunto ortonormal** es un conjunto ortogonal donde cada vector tiene norma 1, es decir

$$\langle x_i, x_j \rangle = \delta_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

Como en cualquier espacio Euclideo de dimension 2, existe una versión análoga del teorema de Pitágoras para cualquier espacio con producto interno.

Teorema 1.11 Teorema de Pitágoras Sea x_1, x_2, \dots, x_n un conjunto ortogonal. Entonces,

$$\left\| \sum_1^n x_i \right\|^2 = \sum_1^n \|x_i\|^2.$$

Demostración:

$$\left\| \sum_1^n x_i \right\|^2 = \left\langle \sum_1^n x_i, \sum_1^n x_j \right\rangle = \sum_i \sum_j \langle x_i, x_j \rangle = \sum_1^n \|x_i\|^2.$$

□

A continuación daremos algunas definiciones y conceptos relacionados con la topología necesarios para el grueso de este trabajo.

Definición 1.12 Una sucesión de vectores x_n en X se dice que **converge** a $x \in X$ y x es el **límite** de x_n , denotado por $x_n \rightarrow x$ o $x = \lim x_n$, cuando $\lim_{n \rightarrow \infty} \|x_n - x\| = 0$. Es decir, para cada $\epsilon > 0$ existe un N entero tal que

$$\|x_n - x\| < \epsilon \text{ para todo } n \geq N.$$

Teorema 1.13 Si $x_n \rightarrow x, y_n \rightarrow y$ para un escalar $\alpha_n \rightarrow \alpha$ entonces

1. $\|x_n\| \rightarrow \|x\|$.
2. $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$.
3. $x_n + y_n \rightarrow x + y$.
4. $\alpha_n x_n \rightarrow \alpha x$.

Demostración:

1. Trivial usando la propiedad (5) de las propiedades de la norma 1.4.
2. Usando la Desigualdad de Scharwz 1.3 tenemos

$$\begin{aligned} |\langle x_n, y_n \rangle - \langle x, y \rangle| &= |\langle x_n - x, y_n \rangle + \langle x, y_n - y \rangle| \\ &\leq |\langle x_n - x, y_n \rangle| + |\langle x, y_n - y \rangle| \leq \|x_n - x\| \|y_n\| + \|x\| \|y_n - y\| \rightarrow 0. \end{aligned}$$

- 3.

$$\|x_n + y_n - (x + y)\| \leq \|x_n - x\| + \|y_n - y\| \rightarrow 0.$$

- 4.

$$\begin{aligned} \|\alpha_n x_n - \alpha x\| &= \|\alpha_n(x_n - x) + (\alpha_n - \alpha)x\| \\ &\leq |\alpha_n| \|x_n - x\| + |\alpha_n - \alpha| \|x\| \rightarrow 0. \end{aligned}$$

□

Definición 1.14 Una sucesión $\{x_n\}$ se dice que es una **sucesión de Cauchy** si para cada $\epsilon > 0$, existe un N entero tal que

$$\|x_n - x_m\| < \epsilon \text{ para todo } n, m \geq N.$$

Definición 1.15 Un conjunto *no vacío* B se dice **acotado** si $\sup\{\|x\| \mid x \in B\} < \infty$.

Los espacios con producto interno en los que cada sucesión de Cauchy converge tienen muy buenas propiedades. Por ello vamos a llamarlos de una forma específica.

Definición 1.16 Decimos que un espacio con producto interno X es **completo** o un **espacio de Hilbert** si cada sucesión de Cauchy en X converge a un punto $x \in X$.

1.2. Mejor aproximación

En esta sección nos dedicaremos a describir el problema general de la mejor aproximación en un espacio de producto interno y demostraremos la unicidad para conjuntos convexos.

Definición 1.17 Sea K un conjunto no vacío de un espacio con producto interno X y sea $x \in X$. Decimos que el elemento $y_o \in K$ es la **mejor aproximación** a x desde K si

$$\|x - y_o\| = d(x, K)$$

donde $d(x, K) = \inf_{y \in K} \|x - y\|$, llamada **distancia de x a K** . Al conjunto de todas las mejores aproximaciones de x en K (puede ser vacío) lo denotamos $P_K(x)$ y esta definido

$$P_K(x) = \{y \in K \mid \|x - y\| = d(x, K)\}.$$

Esto define una aplicación $P_K(x)$ de X en subconjuntos de K llamada **proyección métrica sobre K** . Si para cada $x \in X$ existe por lo menos una mejor aproximación en K , decimos que K es **conjunto proximal**. En caso de que exista exactamente una única mejor aproximación llamaremos a K **conjunto de Chebyshev**. Cuando K es un conjunto de Chebyshev, $P_K(x)$ es una aplicación univalorada de $X \rightarrow X$ para todo $x \in X$.

Definición 1.18 Decimos que un subconjunto K de X es **convexo** si para todo $x, y \in K$ y $\lambda \in [0, 1]$ tenemos que $\lambda x + (1 - \lambda)y \in K$.

Definición 1.19 Un subconjunto C de X es un **cono convexo** si $\alpha x + \beta y \in C$ para todo $x, y \in C$ y $\alpha, \beta \geq 0$

Proposición 1.20 Sea K un subconjunto convexo de X , entonces cada $x \in X$ tiene como máximo una mejor aproximación en K . En particular, todo conjunto convexo proximal es Chebyshev.

Demostración: Sea $x \in X$ y supongamos que tanto y_1 como y_2 están en $P_K(x)$. Entonces $(y_1 + y_2)/2 \in K$ por convexidad y tenemos

$$d(x, K) \leq \|x - \frac{1}{2}(y_1 + y_2)\| = \|\frac{1}{2}(x - y_1) + \frac{1}{2}(x - y_2)\| \leq \frac{1}{2}\|x - y_1\| + \frac{1}{2}\|x - y_2\| = d(x, K).$$

La igualdad tiene que mantenerse en las desigualdades. Por la condición de igualdad de la desigualdad triangular, $x - y_1 = \lambda(x - y_2)$ para algun $\lambda \geq 0$. Pero como $\|x - y_1\| = \|x - y_2\| = d(x, k)$ obliga a que $\lambda = 1$ y, por tanto, $y_1 = y_2$. Con lo que es única. \square

Definición 1.21 Dado $K \subset X$ no vacío e $y \in X$, denotamos

$$y + K = \{y + x \mid x \in K\}.$$

Propiedades 1.22 Sea K un subconjunto de X no vacío. Se tiene

1. $d(x + y, K + y) = d(x, K)$ para todo $x, y \in X$.
2. $P_{K+y}(x + y) = P_K(x) + y$ para todo $x, y \in X$.
3. $d(\alpha x, \alpha K) = |\alpha|d(x, K)$ para todo $x \in X, \alpha \in \mathbb{R}$.

4. $P_{\alpha K}(\alpha x) = \alpha P_K(x)$ para todo $x \in X$, $\alpha \in \mathbb{R}$.
5. K es proximal (respectivamente Chebyshev) si y solo si $K + y$ es proximal (respectivamente Chebyshev) para cualquier $y \in X$ dado.
6. K es proximal (respectivamente Chebyshev) si y solo si αK es proximal (respectivamente Chebyshev) para cualquier $\alpha \in \mathbb{R}$ dado.

Demostración:

1. Para cualquier $x, y \in X$ se tiene

$$d(x + y, K + y) = \inf_{z \in K} \|x + y - (z + y)\| = \inf_{z \in K} \|x - z\| = d(x, K).$$

2. Usando (1) $y_0 \in P_{K+y}(x + y)$ si y solo si $y_0 \in K + y$ y $\|x + y - y_0\| = d(x + y, K + y)$; si y solo si $y_0 - y \in K$ y $\|x - (y_0 - y)\| = d(x, K)$; si y solo si $y_0 - y \in P_K(x)$, esto es $y_0 = P_K(x) + y$.

3. Tenemos que

$$d(\alpha x, \alpha K) = \inf_{y \in K} \|\alpha x - \alpha y\| = |\alpha| \inf_{y \in K} \|x - y\| = |\alpha| d(x, K).$$

4. Si $\alpha = 0$ es trivial. En caso contrario, usando (3) vemos que $y_0 \in P_{\alpha K}(\alpha x)$; si y solo si $y_0 \in \alpha K$ y $\|\alpha x - y_0\| = d(\alpha x, \alpha K)$ si y solo si $\frac{1}{\alpha} y_0 \in K$ y $|\alpha| \|x - \frac{1}{\alpha} y_0\| = |\alpha| d(x, K)$. Sin embargo, esto es equivalente a $\frac{1}{\alpha} y_0 \in P_K(x)$, es decir, $y_0 \in \alpha P_K(x)$.
5. Trivial usando (2).
6. Trivial usando (4).

□

Proposición 1.23 Los conjuntos proximales son cerrados. Sea K un conjunto proximal de X , entonces K es cerrado.

Demostración: Supongamos que K no fuese cerrado. Existiría una sucesión $\{x_n\}$ en K tal que $x_n \rightarrow x$ y $x \notin K$. Entonces $d(x, K) \leq \|x - x_n\| \rightarrow 0$, con lo que $d(x, K) = 0$. Pero como $\|x - y\| > 0$ para cada $y \in K$, contradice que $P_K(x) \neq \emptyset$. □

Definición 1.24 Sea K un subconjunto no vacío de X . Lo llamamos

1. **completo** si cada sucesión de Cauchy en K converge a un punto en K .
2. **aproximadamente compacto** si para cada $x \in X$ dado, toda sucesión $\{y_n\}$ en K con $\|x - y_n\| \rightarrow d(x, K)$ tiene una subsucesión que converge a un punto en K .

Con estas nociones ya podemos demostrar la existencia y unicidad de la mejor aproximación en espacios convexos.

Proposición 1.25 Existencia de la mejor aproximación

1. Todo conjunto compacto es proximal.
2. Todo conjunto cerrado, completo y convexo es un conjunto de Chebyshev.

Demostración:

1. Sea K aproximadamente compacto y $x \in X$. Suponemos sin pérdida de generalidad que $x \notin K$ y escogemos una sucesión minimizante $\{y_n\}$ en K para x . Sea $\{y_{n_j}\}$ una subsucesión que converge a algún $y \in K$; aplicando el teorema 1.13

$$\|x - y\| = \lim \|x - y_{n_j}\| = d(x, K).$$

Con lo que y es mejor aproximación a x desde K , y por ende, K es proximal.

2. Sea K un conjunto completo y convexo, fijamos un $x \in X$. Supongamos que $\{y_n\}$ es minimizante para x : $\|x - y_n\| \rightarrow d(x, K)$. Entonces aplicando el teorema del paralelogramo

$$\begin{aligned} \|y_n - y_m\|^2 &= \|(x - y_m) - (x - y_n)\|^2 \\ &= 2(\|x - y_m\|^2 + \|x - y_n\|^2) - \|2x - (y_m + y_n)\|^2 \\ &= 2(\|x - y_m\|^2 + \|x - y_n\|^2) - 4\|x - \frac{1}{2}(y_m + y_n)\|^2. \end{aligned}$$

Al ser K convexo, $\frac{1}{2}(y_n + y_m) \in K$ y por tanto

$$\|y_n - y_m\|^2 \leq 2(\|x - y_m\|^2 + \|x - y_n\|^2) - 4d(x, K)^2.$$

Como $\{y_n\}$ es una sucesión minimizante para x , la parte de la derecha tiende a cero conforme n y m tienden a infinito. Con lo que $\{y_n\}$ es una sucesión de Cauchy y, al ser K completo, $\{y_n\}$ converge a algún punto $y \in K$, por lo que K es aproximadamente compacto. Por la primera parte, K es proximal y, con la proposición 1.20 y el hecho de que K es convexo, terminamos.

□

Proposición 1.26 Unicidad de la mejor aproximación. Cada subconjunto cerrado y convexo no vacío de un espacio de Hilbert es un conjunto (aproximadamente compacto) de Chebyshev.

Demostración: Sea K un subconjunto cerrado y convexo de un espacio de Hilbert X . Por la proposición 1.25(2) será suficiente verificar que K es completo. Sea ahora $\{y_n\}$ una sucesión de Cauchy en K . Al ser X completo, existe un $y \in X$ tal que $y_n \rightarrow y$. Como K es cerrado, $y \in K$ y por tanto K es completo. \square

Vistas la existencia y unicidad de la mejor aproximación, veamos ahora dos caracterizaciones para casos particulares: para conjuntos convexos y para subespacios convexos.

Proposición 1.27 Caracterización de la mejor aproximación para conjuntos convexos (Criterio de Kolmogorov). Sea K un subconjunto convexo, cerrado, no vacío, de un espacio H de Hilbert y sea $x \in X$. Entonces existe un único $y_0 \in K$ tal que $\|x - y_0\| \leq \|x - y\|$ para todo $y \in K$. Además, una condición necesaria y suficiente para que y_0 sea el único vector que minimiza $\|x - y\|$ es que satisfaga

$$\langle x - y_0, y - y_0 \rangle \leq 0 \text{ para todo } y \in K.$$

Demostración: La existencia y unicidad de la mejor aproximación es consecuencia de las proposiciones 1.25 y 1.26. Veamos la caracterización.

Si se tiene la desigualdad e $y \in K$, entonces

$$\begin{aligned} \|x - y_0\|^2 &= \langle x - y_0, x - y_0 \rangle = \langle x - y_0, x - y \rangle + \langle x - y_0, y - y_0 \rangle \\ &\leq \langle x - y_0, x - y \rangle \leq \|x - y_0\| \|x - y\| \end{aligned}$$

por la desigualdad de Schwarz 1.3. Por tanto $\|x - y_0\| \leq \|x - y\|$ e $y_0 \in K$. Supongamos ahora que no se tiene la desigualdad, entonces $\langle x - y_0, y - y_0 \rangle > 0$ para algún $y \in K$. Por convexidad tenemos que para cualquier $\lambda \in (0, 1)$, el elemento $y_\lambda = \lambda y + (1 - \lambda)y_0 \in K$ y

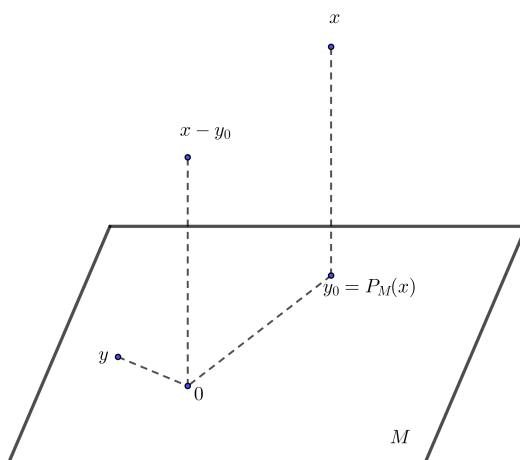
$$\begin{aligned} \|x - y_\lambda\|^2 &= \langle x - y_\lambda, x - y_\lambda \rangle = \langle x - y_0 - \lambda(y - y_0), x - y_0 - \lambda(y - y_0) \rangle \\ &= \|x - y_0\|^2 - 2\lambda \langle x - y_0, y - y_0 \rangle + \lambda^2 \|y - y_0\|^2 \\ &= \|x - y_0\|^2 - \lambda [2\langle x - y_0, y - y_0 \rangle - \lambda \|y - y_0\|^2]. \end{aligned}$$

El término entre corchetes es positivo para cierto $\lambda > 0$ suficientemente pequeño, y por tanto $\|x - y_\lambda\| < \|x - y_0\|$. Entonces $y_0 \neq P_K(x)$. \square

Proposición 1.28 Caracterización de la mejor aproximación para subespacios. Sea M un subespacio en X , y sean $x \in X$ e $y_0 \in M$. Entonces $y_0 = P_M(x)$ si y solo si $x - y_0 \in M^\perp$, es decir

$$\langle x - y_0, y \rangle = 0 \text{ para todo } y \in M.$$

La idea geométrica sería: y_0 es la mejor aproximación a x si y solo si $x - y_0$ es ortogonal a M .



Corolario 1.29 Las ecuaciones normales. Sean $\{x_1, x_2, \dots, x_n\}$ una base de un subespacio M n -dimensional de X . Entonces M es Chebyshev y para cada $x \in X$ se tiene

$$P_M(x) = \sum_1^n \alpha_i x_i$$

donde los escalares α_i son la única solución a las **ecuaciones normales**

$$\sum_{i=1}^n \alpha_i \langle x_i, x_j \rangle = \langle x, x_j \rangle \quad (j = 1, 2, \dots, n).$$

En particular, si $\{x_1, x_2, \dots, x_n\}$ es una base ortonormal de M , entonces

$$P_M(x) = \sum_{i=1}^n \langle x_i, x \rangle x_i \text{ para todo } x \in X.$$

Demostración: Fijamos $x \in X$ e $y_0 \in M$. Entonces, para algunos α_i tenemos $y_0 = \sum_1^n \alpha_i x_i$. Aplicando la caracterización de mejor aproximación para subespacios 1.28 tenemos que $y_0 = P_M(x)$ y $\langle x - y_0, y \rangle = 0$ para todo $y \in M$ son equivalentes. Al ser también equivalentes a $\langle x - y_0, x_j \rangle = 0$ para cada $j = 1, 2, \dots, n$, tenemos la primera parte. Si $\{x_1, x_2, \dots, x_n\}$ es ortonormal, entonces $\langle x_i, x_j \rangle = \delta_{ij}$ por tanto $\alpha_i = \langle x, x_i \rangle$ para cada i y se cumple la segunda parte. \square

Teorema 1.30 Sea M un subespacio de Chebyshev de un espacio de Hilbert H . Entonces

1. M^\perp es un subespacio de Chebyshev.
2. $x = P_M(x) + P_{M^\perp}(x)$ para todo $x \in H$. Más concretamente, $I = P_M + P_{M^\perp}$, donde I es la aplicación identidad $I(x) = x$ para todo x .

Demostración: Para probar (1) y (2), sea $x \in H$ e $y_0 = x - P_M(x)$. Por la caracterización de mejor aproximación para subespacios 1.28, $y_0 \in M^\perp$ y $y_0 \perp (x - y_0)$. Para todo $y \in M^\perp$,

$$\langle x - y_0, y \rangle = \langle P_M(x), y \rangle = 0.$$

Por tanto $x - y_0 \in (M^\perp)^\perp$ e $y_0 = P_{M^\perp}(x)$. Esto prueba que M^\perp es un subespacio de Chebyshev y que $x = P_M(x) + P_{M^\perp}(x)$. \square

1.3. Convergencia débil

Hablaremos ahora sobre la convergencia débil y algunos resultados que derivan de ella, que serán necesarios para el Capítulo 4.

Definición 1.31 Una sucesión $\{x_n\}$ en un espacio con producto interno X se dice que **converge débilmente** a x , escrito $x_n \xrightarrow{w} x$ si

$$\langle x_n, y \rangle \rightarrow \langle x, y \rangle \text{ para todo } y \in X.$$

Lema 1.32 1. Si $x_n \rightarrow x$ entonces $x_n \xrightarrow{w} x$ (no es recíproco en general).

2. $x_n \rightarrow x$ si y solo si $x_n \xrightarrow{w} x$ y $\|x_n\| \rightarrow \|x\|$.

Demostración:

1. Para todo $y \in X$ se tiene

$$|\langle x_n, y \rangle - \langle x, y \rangle| = |\langle x_n - x, y \rangle| \leq \|x_n - x\| \|y\| \rightarrow 0,$$

entonces $\langle x_n, y \rangle \rightarrow \langle x, y \rangle$, que es la definición de convergencia débil.

2. La primera implicación esta demostrada en (1) y en el teorema 1.13. Veamos la segunda. Si x_n converge débilmente a x

$$\|x_n - x\|^2 = \|x_n\|^2 - 2\langle x_n, x \rangle + \|x\|^2 \rightarrow \|x\|^2 - 2\langle x, x \rangle + \|x\|^2 = 0,$$

es decir, $x_n \rightarrow x$.

□

Teorema 1.33 Teorema de la convergencia débil. Toda sucesión acotada en un espacios de Hilbert tiene una subsucesión que converge débilmente.

Demostración: Sea $\{x_n\}$ una sucesión en un espacio de Hilbert H , con $\|x_n\| \leq c$ para todo n . Consideremos $\{\langle x_1, x_n \rangle\}$, una sucesión de números reales acotada, ya que $|\langle x_1, x_n \rangle| \leq \|x_1\| \|x_n\| \leq c^2$. Al ser acotada existe una subsucesión que converge: $\langle x_1, x_{1(n)} \rangle \rightarrow \delta_1 \in \mathbb{R}$ cuando $n \rightarrow \infty$. De la misma manera, $\{\langle x_2, x_{1(n)} \rangle\}$ está acotada y tiene una subsucesión convergente: $\langle x_2, x_{2(n)} \rangle \rightarrow \delta_2 \in \mathbb{R}$ cuando $n \rightarrow \infty$. Continuando de esta manera vemos que para todo $k \in \mathbb{N}$ hay una subsucesión $\{x_{k(n)}\}$ de $\{x_{(k-1)(n)}\}$ tal que $\langle x_k, x_{k(n)} \rangle \rightarrow \delta_k \in \mathbb{R}$ cuando $n \rightarrow \infty$. Entonces la sucesión "diagonal" $\{x_{n(n)}\}$ es una subsucesión de $\{x_n\}$ tal que, para cada $k \in \mathbb{N}$,

$$\langle x_k, x_{n(n)} \rangle \rightarrow \delta_k \text{ cuando } n \rightarrow \infty \quad (1.1)$$

Sea

$$M = \{x \in H \mid \lim_n \langle x, x_{n(n)} \rangle \text{ existe}\}$$

y definimos f en M como

$$f(x) = \lim_n \langle x, x_{n(n)} \rangle \quad x \in M. \quad (1.2)$$

Por 1.1, $x_k \in M$ ($k = 1, 2, \dots$). También, si $x, y \in M$ y $\alpha\beta \in \mathbb{R}$, se tiene

$$\lim_n \langle \alpha x + \beta y, x_{n(n)} \rangle = \lim_n [\alpha \langle x, x_{n(n)} \rangle + \beta \langle y, x_{n(n)} \rangle] = \alpha f(x) + \beta f(y).$$

Por lo que $\alpha x + \beta y \in M$. Esto prueba que M es un subespacio lineal y que f es lineal en M . Además, para todo $x \in M$,

$$|f(x)| = \lim_n |\langle x, x_{n(n)} \rangle| \leq \lim_n \sup \|x\| \|x_{n(n)}\| \leq c \|x\|.$$

Ergo f está acotada en M y $\|f\| \leq c$. Ahora tenemos que ver que M es cerrado. Fijamos un $y \in \overline{M}$, entonces existe $y_m \in M$ tal que $y_m \rightarrow y$. Como

$$|f(y_n) - f(y_m)| \leq \|f\| \|y_n - y_m\| \leq c \|y_n - y_m\|$$

y $\{y_n\}$ es de Cauchy, se sigue que $\{f(y_n)\}$ es tambien de Cauchy en \mathbb{R} , con lo que $\lim_n f(y_n)$ existe. Si tambien $y'_n \in M$ e $y'_n \rightarrow y$, entonces $y_n - y'_n \rightarrow 0$ y

$$|f(y_n) - f(y'_n)| \leq \|f\| \|y_n - y'_n\| \rightarrow 0.$$

Esto prueba que $F(y) := \lim_n f(y_n)$ existe y es independiente de la sucesión $\{y_n\}$ en M que converge a y . Dado un $\epsilon > 0$, elegimos un entero N_1 tal que

$$c \|y_{N_1}\| < \epsilon/3 \quad y \quad |f(y_{N_1}) - F(y)| < \epsilon/3. \quad (1.3)$$

Entonces elegimos un entero N tal que, para todo $n \geq N$,

$$|\langle y_{N_1}, x_{n(n)} \rangle - f(y_{N_1})| < \epsilon/3. \quad (1.4)$$

Si en (1.3) y en (1.4) tomamos $n \geq N$, tenemos que

$$\begin{aligned} |\langle y, x_{n(n)} - F(y) \rangle| &\leq |\langle y, x_{n(n)} \rangle - \langle y_{N_1}, x_{n(n)} \rangle| + |\langle y_{N_1}, x_{n(n)} \rangle - f(y_{N_1})| + |f(y_{N_1}) - F(y)| \\ &< \|y - y_{N_1}\| \|x_{n(n)}\| + 2\epsilon/3 \leq c\|y - y_{N_1}\| + 2\epsilon/3 < \epsilon. \end{aligned}$$

Con esto vemos que $F(y) = \lim_n \langle y, x_{n(n)} \rangle$ existe y así $y \in M$, y, por tanto, M es cerrado. Al ser M un subespacio cerrado de un espacio de Hilbert H , M es también Hilbert y un subespacio de Chebyshev de H . Por el teorema de representación de Riesz tenemos que f tiene un representante $x_0 \in M$, que es

$$f(x) = \langle x, x_0 \rangle, \quad x \in M \quad (1.5)$$

Además, por 1.30 tenemos que todo $x \in H$ puede ser escrito como $x = P_M(x) + P_{M^\perp}(x)$. Como tanto x_0 como $x_{n(n)}$ están en M , obtenemos de (1.2) y (1.5) que

$$\langle x, x_0 \rangle = \langle P_M(x), x_0 \rangle = f(P_M(x)) = \lim_n \langle P_M(x), x_{n(n)} \rangle = \lim_n \langle x, x_{n(n)} \rangle$$

que muestra que $x_{n(n)} \xrightarrow{w} x_0$ y completa la prueba. □

Definición 1.34 Sea X un espacio con producto interno y sea $K \subset X$. Decimos que K es **débilmente cerrado** si para todo $x \in X$ tal que existe $\{x_n\} \subset K$ que converge débilmente a x , se tiene que $x \in K$.

Veamos ahora el principio de separación fuerte para convexos, que necesitaremos para demostrar el siguiente teorema.

Teorema 1.35 Principio de separación fuerte para convexos. Sea K un conjunto cerrado, no vacío, convexo en el espacio de Hilbert H y sea $x \in H \setminus K$. Entonces existe una única forma lineal continua $g : H \rightarrow \mathbb{R}$ tal que $\|g\| = 1$ y

$$d(x, K) = g(x) - \sup_{x \in K} g(x).$$

En particular

$$\sup_{x \in K} g(x) < g(x).$$

Demostración: Si $x \notin K$, la distancia $d = d(x, K)$ es positiva. Sea $u = P_K(x)$ la mejor aproximación en K a x y pongamos $z = (x - u)/d$. Por la caracterización de la mejor aproximación para convexos 1.27

$$\sup_{y \in K} \langle x - u, y - u \rangle \leq 0,$$

que implica

$$\sup_{y \in K} \langle z, y - u \rangle \leq 0.$$

Se tiene $\|z\| = 1$ y

$$\langle z, x \rangle - \sup_{y \in K} \langle z, y \rangle \leq \langle z, x \rangle - \langle z, u \rangle = \langle z, x - u \rangle = d$$

de forma que $g(x) = \langle z, x \rangle$. □

Teorema 1.36 1. Todo débilmente cerrado es cerrado.

2. Un conjunto convexo es cerrado si y solo si es débilmente cerrado.

Demostración:

1. Sea K un conjunto débilmente cerrado, $x_n \in K$ y $x_n \rightarrow x$. Por el teorema 1.32(1), $x_n \xrightarrow{w} x$. Como K es débilmente cerrado, $x \in K$.
2. Admitiendo (1), nos basta probar que si K es cerrado y convexo, entonces K es débilmente cerrado. Sea $x_n \in K$ y $x_n \xrightarrow{w} x$. Si $x \notin K$, por el teorema 1.35, existe una forma lineal continua $g : H \rightarrow \mathbb{R}$ con $\|g\| = 1$ y $g(x) > \sup_{x \in K} g(x)$. Se sigue que

$$g(x) > \sup_n g(x_n) \geq \lim_n g(x_n) = g(x),$$

que es un absurdo, por tanto $x \in K$ y K es débilmente cerrado. □

Capítulo 2

Método de proyecciones alternadas

En este capítulo describiremos el método de proyecciones alternadas para calcular la mejor aproximación desde un conjunto cerrado y convexo K , que es intersección de un número finito de conjuntos K_i cerrados y convexos, es decir $K = \bigcap_1^r K_i$. Este método consiste en reducir el problema de mejor aproximación a cada uno de esos K_i donde es más fácil ver esa mejor aproximación.

2.1. Teoría de operadores

Veamos primero unas definiciones básicas de teoría de operadores

Definición 2.1 Sean X e Y dos espacios normados. Entonces

1. Una aplicación $T : X \rightarrow Y$ se llama **operador o transformación** y denotamos Tx al valor de $T(x)$.
2. El operador T es un **operador lineal** si $T(x + y) = Tx + Ty$ y $T(\alpha z) = \alpha Tz$ para todo $\alpha \in \mathbb{R}$.
3. Decimos que un operador T está **acotado** si existe un k real positivo tal que $\|Tx\| \leq k\|x\|$ para todo $x \in X$.
4. Un operador T es **uniformemente continuo** si, para $\epsilon > 0$, existe un δ independiente de x_0 tal que $\|Tx - Tx_0\| < \epsilon$ para cualquier x, x_0 con $\|x - x_0\| < \delta$, se tiene $\|Tx - Tx_0\| < \epsilon$.
5. Llamamos **norma inducida de un operador lineal acotado T** a $\|T\| = \sup_{x \neq 0} \left\{ \frac{\|Tx\|}{\|x\|} \right\}$.
6. Si $Y = \mathbb{R}$ entonces T se llama **funcional**.

7. Para un operador T , los conjuntos $R = \{Tx \in Y \mid x \in X\}$ y $N = \{x \in X \mid Tx = 0\}$ se llaman respectivamente **rango** y **núcleo** de T .

Definición 2.2 Definimos $\mathcal{B}(X, Y)$ como al conjunto de todos los operadores lineales acotados de X a Y . La norma de un operador $A \in \mathcal{B}(X, Y)$ viene dada por

$$\|A\| = \sup_{\|x\|=1} \|Ax\|.$$

Ademas

$$\|Ax\| \leq \|A\|\|x\| \text{ para todo } x \in X.$$

Lema 2.3 Si $A \in \mathcal{B}(X, Y)$ y $B \in \mathcal{B}(X, Y)$ entonces $B \circ A = BA \in \mathcal{B}(X, Y)$ y $\|BA\| \leq \|B\|\|A\|$.

Demostración: Sea $x \in X$, por la linealidad de BA

$$\|BAx\| = \|B(Ax)\| \leq \|B\|\|Ax\| \leq \|B\|\|A\|\|x\|,$$

que prueba que BA es acotado y que $\|BA\| \leq \|B\|\|A\|$. □

Definición 2.4 Sean X, Y espacios con producto interno y $A \in \mathcal{B}(X, Y)$. Llamamos **operador adjunto** a la única aplicación $B : Y \rightarrow X$ que satisface

$$\langle Ax, y \rangle = \langle x, B(y) \rangle \text{ para todo } x \in X, y \in Y$$

siendo X, Y espacios con producto interno. Lo denotamos por A^* .

Lema 2.5 Sean X, Y, Z espacios con producto interno, y $A \in \mathcal{B}(X, Y)$ y $B \in \mathcal{B}(Y, Z)$. Si A^* y B^* existe, también existe $(AB)^*$ y

$$(AB)^* = B^*A^*.$$

Demostración: Como $AB \in \mathcal{B}(X, Z)$, para todo $x \in X$ y $z \in Z$ se tiene

$$\langle ABx, z \rangle = \langle A(Bx), z \rangle = \langle Bx, A^*z \rangle = \langle x, B^*(A^*z) \rangle = \langle x, B^*A^*z \rangle$$

y por tanto $(AB)^* = B^*A^*$. □

2.2. Teorema de Von Neumann

Continuaremos con un tipo especial de operadores, las proyecciones. Si S es un subespacio cerrado en H , al vector $x_0 \in S$ tal que $x - x_0 \in S^\perp$ se le llama proyección ortogonal de x sobre S , y denotamos dicha operación como $P_S(x) = S_0$. Estos operadores nos permitirán el cálculo de la mejor aproximación para un caso especial, usando el teorema de Von Neumann.

Proposición 2.6 Linealidad de las proyecciones métricas. Sea M un subespacio de Chebyshev en un espacio con producto interno X

1. P_M es un operador lineal acotado y $\|P_M\| = 1$ (salvo cuando $M = \{0\}$, en cuyo caso $\|P_M\| = 0$).
2. P_M es idempotente, es decir $P_M^2 = P_M$.
3. P_M es autoadjunto, es decir,

$$\langle P_M(x), y \rangle = \langle x, P_M(y) \rangle \text{ para todo } x, y \in X.$$

4. Para todo $x \in X$ se tiene

$$\langle P_M(x), x \rangle = \|P_M(x)\|^2.$$

5. P_M es no negativo, es decir,

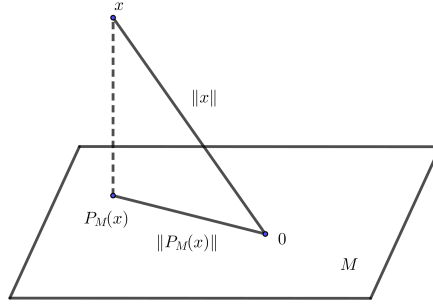
$$\langle P_M(x), x \rangle \geq 0 \text{ para todo } x.$$

Demostración:

1. Sean x, y en X y sean $\alpha, \beta \in \mathbb{R}$. Por la caracterización para subespacios 1.28, $x - P_M(x)$ y $y - P_M(y)$ están en M^\perp , y como es un subespacio,

$$\alpha x + \beta y - [\alpha P_M(x) + \beta P_M(y)] = \alpha(x - P_M(x)) + \beta(y - P_M(y)) \in M^\perp.$$

Sabemos que $\alpha P_M(x) + \beta P_M(y) = P_M(\alpha x + \beta y)$ (por la caracterización para subespacios, ya que $\alpha P_M(x) + \beta P_M(y) \in M$) con lo que es lineal. Nos falta ver que es acotado, para ello, veamos que $\|P_M(x)\| \leq \|x\|$



Vemos que por el teorema de Pitágoras se da la desigualdad, salvo que $x \in M$, en cuyo caso (de hecho, $P_M(x) = x$). Por tanto P_M está acotado y $\|P_M(x)\| \leq 1$. Como $P_M(y) = y$ para cualquier $y \in M$, $\|y\| = \|P_M(y)\| \leq \|P_M\| \|y\|$ implica que $\|P_M\| \geq 1$ y por tanto $\|P_M\| = 1$.

2. Sabemos que $P_M(x) \in X$, y que $P_M(y) = y$ si y solo si $y \in M$. Sea $x_0 = P_M(x)$ y por tanto $x_0 \in M$, entonces

$$P_M(P_M(x)) = P_M(x_0) = x_0 = P_M(x).$$

3. Para todo $x, y \in X$ por la caracterización de subespacios 1.28 sabemos que $\langle P_M(x), y - P_M(y) \rangle = 0$ y por tanto

$$\langle P_M(x), y \rangle = \langle P_M(x), P_M(y) \rangle.$$

Cambiando x por y ,

$$\langle x, P_M(y) \rangle = \langle P_M(y), x \rangle = \langle P_M(y), P_M(x) \rangle = \langle P_M(x), P_M(y) \rangle = \langle P_M(x), y \rangle.$$

4. Aplicamos (3) a $y = x$.
5. Trivial usando (4).

□

Lema 2.7 Sea X un espacio con producto interno y $A \in B(X, X)$, siendo $B(X, X)$ el espacio de aplicaciones lineales continuas de X a X . Entonces A es idempotente y autoadjunto si y solo si $A = P_M$ para algún subespacio de Chebyshev M .

Demostración: Si $A = P_M$, por el lema anterior, $A^2 = A$ (idempotente) y $A^* = A$ (autoadjunto). Veamos ahora la otra implicación, supongamos $A^2 = A$ y $A^* = A$. Sea $M = \mathcal{R}(A)$, el rango de A . Cojamos ahora un $y \in M$ tal que $y = Ax$

para algún x . Como es idempotente, $y = Ax = A^2x = A(Ax) = Ay$ con lo que $y = Ay$ para todo $y \in M$. Para cualquier $x \in X$ e $y \in M$

$$\langle x - Ax, y \rangle = \langle x, y \rangle - \langle Ax, y \rangle = \langle x, y \rangle - \langle x, Ay \rangle = \langle x, y \rangle - \langle x, y \rangle = 0.$$

Por tanto, $x - Ax \in M^\perp$. Por el teorema de caracterización de mejor aproximación para subespacios 1.28, $Ax = P_M(x)$. Puesto que x era arbitrario, tenemos $A = P_M$
□

Lema 2.8 Sea M y N subespacios de Chebyshev del espacio con producto interno X , los siguientes resultados son equivalentes

1. P_M y P_N conmutan, es decir, $P_M P_N = P_N P_M$.
2. $P_N(M) \subset M$.
3. $P_M(N) \subset N$.
4. $P_M P_N = P_{M \cap N}$.
5. $P_M P_N$ es la proyección ortogonal sobre un subespacio Chebyshev.

En particular, si $M \subset N$ o viceversa, entonces P_M y P_N conmutan.

Demostración:

- (1) \Rightarrow (2). Supongamos P_M y P_N conmutan. Entonces para todo $x \in M$, tenemos que

$$P_N(x) = P_N P_M(x) = P_M P_N(x) \in M$$

Por tanto tenemos (2)

- (2) \Rightarrow (3). Supongamos (2), $x \in N$ e $y \in N^\perp$. Entonces

$$\begin{aligned} \langle P_M(x), y \rangle &= \langle P_M P_N(x), y \rangle = \langle P_N(x), P_M(y) \rangle \\ &= \langle x, P_N P_M(y) \rangle = \langle x, P_M P_N P_M(y) \rangle \text{ (como } P_N P_M(y) \in M \text{ por (2))} \\ &= \langle P_M(x), P_N P_M(y) \rangle = \langle P_N P_M(x), P_M(y) \rangle \\ &= \langle P_M P_N P_M(x), y \rangle = \langle P_N P_M(x), y \rangle \text{ (por (2))} \\ &= 0 \text{ (porque } y \in N^\perp). \end{aligned}$$

Tenemos ahora que $P_M(x) \in N^{\perp\perp}$. Pero como $N^{\perp\perp} = N$, concluimos que para todo $x \in N$, $P_M(x) \in N$ y se tiene (3).

- (3) \Rightarrow (4). Supongamos que $P_M(N) \subset N$. Sea $x \in X$ y definamos $x_0 = P_M P_N(x)$. Para mostrar que $x_0 = P_{M \cap N}(x)$, por la caracterización de mejor aproximación para subespacios 1.28, basta demostrar que $x_0 \in M \cap N$ y $x - x_0 \in (M \cap N)^\perp$. Para todo $y \in M \cap N$,

$$\begin{aligned} \langle x - x_0, y \rangle &= \langle x, y \rangle - \langle P_M P_N(x), y \rangle = \langle x, y \rangle - \langle P_N(x), P_M(y) \rangle \\ &= \langle x, y \rangle - \langle P_N(x), y \rangle = \langle x, y \rangle - \langle x, P_N(y) \rangle = \langle x, y \rangle - \langle x, y \rangle = 0. \end{aligned}$$

Así que $x - x_0 \in (M \cap N)^\perp$. Claramente, $x_0 = P_M P_N(x) \in M$. Además, para cada $z \in N^\perp$,

$$\begin{aligned} \langle x_0, z \rangle &= \langle P_M P_N(x), z \rangle = \langle P_N P_M P_N(x), z \rangle \\ &= \langle P_M P_N(x), P_N(z) \rangle = \langle P_M P_N(x), 0 \rangle = 0. \end{aligned}$$

Por lo que $x_0 \in N^{\perp\perp} = N$. Por tanto, $x_0 \in M \cap N$ y esto prueba que $x_0 = P_{M \cap N}(x)$.

- (4) \Rightarrow (5) es trivial.
- (5) \Rightarrow (1) Sea $P = P_M P_N$ la proyección ortogonal sobre un subespacio Chebyshev. Por el Lema 2.7 sabemos que $P_M P_N$ es idempotente y auto-adjunto. Usando el Lema 2.5 sabemos que $P_M P_N = (P_M P_N)^* = P_N^* P_M^* = P_N P_M$, llegando a (1).

□

Teorema 2.9 Teorema de Von Neumann. Sea M_1 y M_2 subespacios cerrados en un espacio de Hilbert X , entonces, para cada $x \in X$,

$$\lim_{n \rightarrow \infty} (P_{M_1} P_{M_2})^n(x) = P_{M_1 \cap M_2}(x).$$

Demostración: Llamaremos P_i a P_{M_i} para suavizar la notación. Fijamos un $x \in X$ y hacemos que

$$\begin{aligned} x_0 &= x \\ x_{2n-1} &= P_1(x_{2n-2}) (= P_1(P_2 P_1)^{n-1} x) \\ x_{2n} &= P_2(x_{2n-1}) (= (P_2 P_1)^n x), \quad \text{para } n = 1, 2, \dots \end{aligned}$$

Basta ahora con probar que $x_{2n} \rightarrow P_{M_1 \cap M_2}(x)$. Probaremos que x_n converge a $P_{M_1 \cap M_2}(x)$, donde, obviamente, se concluye que x_{2n} también converge.

Como $\|P_i\| \leq 1$,

$$\|x_{2n}\| = \|P_2 x_{2n-1}\| \leq \|x_{2n-1}\| = \|P_1 x_{2n-2}\| \leq \|x_{2n-2}\|.$$

Con esto se tiene que la sucesión de normas $\{\|x_n\|\}$ es decreciente y por tanto existe

$$\lambda := \lim_{n \rightarrow \infty} \|x_n\|.$$

Al ser P_i autoadjunto, $x_{2n} \in M_2$ y $x_{2n-1} \in M_1$ deducimos que

$$\begin{aligned}\langle x_{2n}, x_{2m} \rangle &= \langle P_2 x_{2n-1}, x_{2m} \rangle = \langle x_{2n-1}, P_2 x_{2m} \rangle = \langle x_{2n-1}, x_{2m} \rangle \\ &= \langle P_1 x_{2n-1}, x_{2m} \rangle = \langle x_{2n-1}, P_1 x_{2m} \rangle = \langle x_{2n-1}, x_{2m+1} \rangle.\end{aligned}$$

Sumando 1 a los índices y manteniendo la igualdad,

$$\langle x_{2n+1}, x_{2m+1} \rangle = \langle x_{2n}, x_{2m+2} \rangle \quad n, m \in \mathbb{N}.$$

Juntando ambas igualdades llegamos a que, si $m \leq n$ se tiene

$$\langle x_{2n}, x_{2m} \rangle = \langle x_{2n-1}, x_{2m+1} \rangle = \langle x_{2n-2}, x_{2m+2} \rangle = \dots = \langle x_{2n-k}, x_{2m+k} \rangle, \quad 0 \leq k \leq 2n$$

Si tomamos $k = n - m$

$$\langle x_{2n}, x_{2m} \rangle = \langle x_{n+m}, x_{n+m} \rangle = \|x_{n+m}\|^2.$$

Si $m \leq n$,

$$\begin{aligned}\|x_{2n} - x_{2m}\|^2 &= \|x_{2n}\|^2 - 2\langle x_{2n}, x_{2m} \rangle + \|x_{2m}\|^2 \\ &= \|x_{2n}\|^2 - 2\|x_{n+m}\|^2 + \|x_{2m}\|^2.\end{aligned}$$

Conforme $m \rightarrow \infty$ se tiene que $\lambda^2 - 2\lambda^2 + \lambda^2 = 0$, con lo que $\{x_{2n}\}$ es una sucesión de Cauchy en M_2 . Al ser X completo y M_2 cerrado, existe $y \in M_2$ tal que $x_{2n} \rightarrow y$. Además, como

$$\langle x_{2n}, x_{2n-1} \rangle = \langle P_2 x_{2n}, x_{2n-1} \rangle = \langle x_{2n}, P_2 x_{2n-1} \rangle = \langle x_{2n}, x_{2n} \rangle = \|x_{2n}\|^2$$

que implica

$$\begin{aligned}\|x_{2n} - x_{2n-1}\|^2 &= \|x_{2n}\|^2 - 2\langle x_{2n}, x_{2n-1} \rangle + \|x_{2n-1}\|^2 \\ &= -\|x_{2n}\|^2 + \|x_{2n-1}\|^2 \xrightarrow{n \rightarrow \infty} -\lambda^2 + \lambda^2 = 0\end{aligned}$$

de donde podemos ver que

$$\|x_{2n-1} - y\| \leq \|x_{2n-1} - x_{2n}\| + \|x_{2n} - y\| \rightarrow 0.$$

Como $x_{2n-1} \in M_1$ para todo n y M_1 es cerrado, $y \in M_1$. Y queda demostrado que $\lim x_n = y \in M_1 \cap M_2$. Usando la caracterización para subespacios 1.28, basta verificar que $y = P_{M_1 \cap M_2}(x)$, es decir que $x - y \perp M_1 \cap M_2$. Para ello, sea $z \in M_1 \cap M_2$, usando otra vez la caracterización para subespacios

$$\langle x_{2n}, z \rangle = \langle x_{2n} - x_{2n-1} + x_{2n-1}, z \rangle = \langle P_1 x_{2n-1} - x_{2n-1} + x_{2n-1}, z \rangle = \langle x_{2n-1}, z \rangle.$$

De forma similar,

$$\langle x_{2n-1}, z \rangle = \langle x_{2n-2}, z \rangle.$$

Por inducción, tenemos que $\langle x_{2n}, z \rangle = \langle x_0, z \rangle = \langle x, z \rangle$ para todo n . Concluimos con

$$\langle x - y, z \rangle = \langle x, z \rangle - \langle y, z \rangle = \langle x, z \rangle - \lim \langle x_{2n}, z \rangle = \langle x, z \rangle - \langle x, z \rangle = 0.$$

□

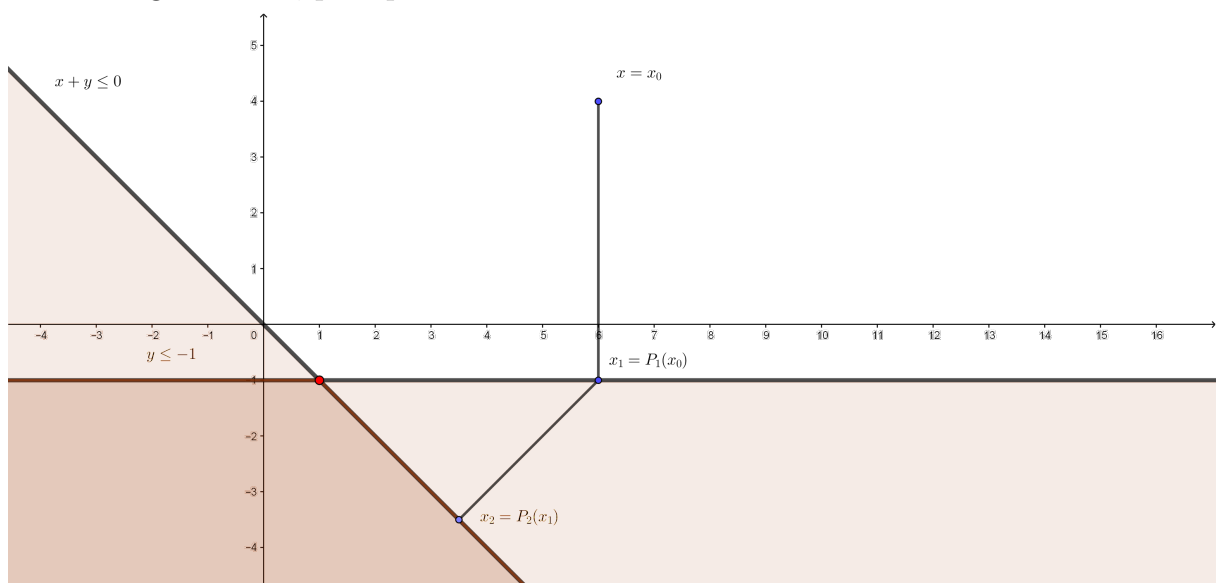
Capítulo 3

Algoritmo de Dykstra

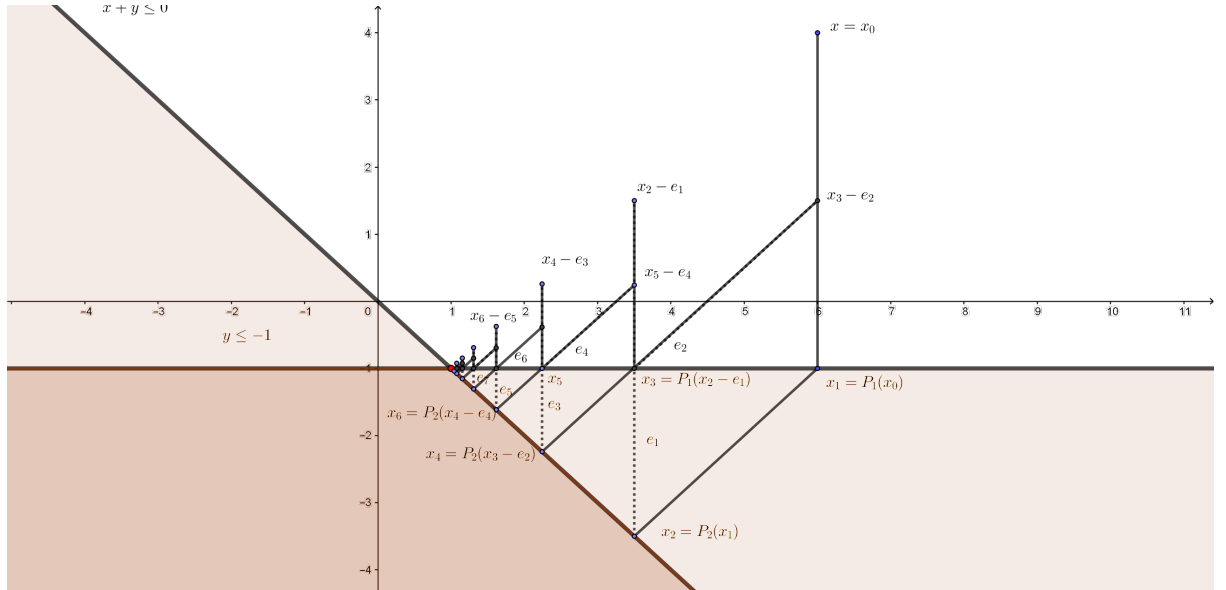
3.1. Motivación

La motivación del estudio de este algoritmo recae en la no convergencia hacia la mejor aproximación del algoritmo de Von Neumann 2.9 en algunos casos (cuando no se aplica a subespacios), mientras que con el algoritmo de Dykstra sí converge. Veamos a continuación un ejemplo que ilustra este fenómeno.

Ejemplo 3.1 Sea $M_1 = \{x, y \in \mathbb{R} \mid y \leq -1\}$ y sea $M_2 = \{x, y \in \mathbb{R} \mid x + y \leq 0\}$. Expondremos la aplicación de los algoritmos mencionados anteriormente de manera geométrica, pues podremos sacar conclusiones de forma más intuitiva.



Podemos ver que con dos iteraciones el algoritmo para sin llegar a la mejor aproximación (el punto rojo). El algoritmo de Dykstra (que definiremos más adelante) incluye unas correcciones, las e_i , que hacen que converja.



Vemos que, añadiendo dichas correcciones, el algoritmo converge a la mejor aproximación. Al final de la siguiente sección veremos un ejemplo gráfico del algoritmo más claro que este ejemplo .

A continuación explicaremos el algoritmo de Dykstra y demostraremos su convergencia, incluyendo los lemas necesarios para su demostración.

3.2. Algoritmo de Dykstra

El algoritmo de Dykstra es un algoritmo iterativo que converge hacia la mejor aproximación a un espacio convexo cerrado K no vacío, siendo K la intersección de r conjuntos convexos cerrados, es decir $K = \bigcap_1^r K_i$. El algoritmo consiste en ir obteniendo la mejor aproximación a cada uno de esos K_i de forma alternada.

Sean K_i ($i = 1, 2, \dots, r$) subconjuntos cerrados y convexos de un espacio X de Hilbert y $K = \bigcap_1^r K_i$ no vacío. Dado un $x \in X$, se calcula $P_K(x)$.

Llamamos $[n]$ a $n \bmod r$, siendo $n \in \mathbb{N}$, y definimos

$$\begin{aligned}
 x_0 &= x & e_{-(r-1)} &= \dots = e_{-1} = e_0 = 0, \\
 x_n &= P_{K_{[n]}}(x_{n-1} + e_{n-r}), \\
 e_n &= x_{n-1} + e_{n-r} - x_n \\
 &= x_{n-1} + e_{n-r} - P_{K_{[n]}}(x_{n-1} + e_{n-r}) \quad (n = 1, 2, \dots).
 \end{aligned}$$

Para ver que este algoritmo converge tendríamos que establecer que

$$\lim_{n \rightarrow \infty} \|x_n - P_K(x)\| = 0.$$

Esto lo haremos con el teorema de Boyle-Dykstra, pero previamente necesitaremos ciertos lemas.

Lema 3.2 Para todo n ,

$$\langle x_n - y, e_n \rangle \geq 0 \text{ para todo } y \in K_{[n]}.$$

Demostraci3n: Usando la caracterizaci3n de mejor aproximaci3n para convexos 1.27 con

$$x = x_{n-1} + e_{n-r}, y_0 = P_{K_{[n]}}(x_{n-1} + e_{n-r}) \text{ y } K = K_{[n]}, \text{ para todo } y \in K_{[n]},$$

$$\langle x_n - y, e_n \rangle = \langle P_{K_{[n]}}(x_{n-1} + e_{n-r}) - y, x_{n-1} + e_{n-r} - P_{K_{[n]}}(x_{n-1} + e_{n-r}) \rangle \geq 0.$$

□

Lema 3.3 Para todo $n \geq 0$,

$$x - x_n = e_{n-(r-1)} + e_{n-(r-2)} + \dots + e_{n-1} + e_n.$$

Demostraci3n: Se demuestra por inducci3n sobre n : Para $n = 0$ tenemos $x - x_0 = x - x = 0$ y $e_{-(r-1)} + e_{-(r-2)} + \dots + e_{-1} + e_0 = 0$ por definici3n de e_n . Asumimos que el resultado es v3lido para alg3n $n \geq 0$, comprobemos que tambi3n es cierto para $n + 1$,

$$\begin{aligned} x - x_{n+1} &= (x - x_n) + (x_n - x_{n+1}) \\ &= (e_{n-(r-1)} + e_{n-(r-2)} + \dots + e_{n-1} + e_n) + (e_{n+1} - e_{n+1-r}) \\ &= e_{n-(r-2)} + e_{n-(r-3)} + \dots + e_{n-1} + e_n + e_{n+1} \\ &= e_{n+1-(r-1)} + e_{n+1-(r-2)} \dots + e_{n-1} + e_n + e_{n+1}. \end{aligned}$$

Con lo que queda demostrado. □

Lema 3.4 Para todo $n \in \mathbb{N}$, $0 \leq m \leq n$, e $y \in K$,

$$\begin{aligned} \|x_m - y\|^2 &= \|x_n - y\|^2 + \sum_{k=m+1}^n \|x_k - x_{k-1}\|^2 + 2 \sum_{k=m+1}^n \langle e_{k-r}, x_{k-r} - x_k \rangle \\ &\quad + 2 \sum_{k=n-(r-1)}^n \langle e_k, x_k - y \rangle - 2 \sum_{k=m-(r-1)}^m \langle e_k, x_k - y \rangle. \end{aligned}$$

Demostraci3n: Para un conjunto de vectores $\{y_m, y_{m+1}, \dots, y_{n+1}\}$ en X se da la identidad

$$\begin{aligned} \|y_m - y_{n+1}\|^2 &= \|(y_m - y_{m+1}) + (y_{m+1} - y_{m+2}) + \dots + (y_n - y_{n+1})\|^2 \\ &= \sum_{k=m+1}^{n+1} \|y_{k-1} - y_k\|^2 + 2 \sum_{m+1 \leq i < j \leq n+1} \langle y_{i-1} - y_i, y_{j-1} - y_j \rangle \end{aligned}$$

$$= \|y_n - y_{n+1}\|^2 + \sum_{k=m+1}^n \|y_{k-1} - y_k\|^2 + 2 \sum_{i=m+1}^n \left(\sum_{j=i+1}^{n+1} \langle y_{i-1} - y_i, y_{j-1} - y_j \rangle \right). \quad (3.1)$$

Este último sumando es igual a

$$\sum_{i=m+1}^n \left\langle y_{i-1} - y_i, \sum_{j=i+1}^{n+1} (y_{j-1} - y_j) \right\rangle = \sum_{i=m+1}^n \langle y_{i-1} - y_i, y_i - y_{n+1} \rangle.$$

Hacemos ahora $y_i = x_i$ para todo $i \leq n$ e $y_{n+1} = y$ y llegamos a

$$\begin{aligned} & \sum_{i=m+1}^n \langle y_{i-1} - y_i, y_i - y_{n+1} \rangle = \sum_{i=m+1}^n \langle x_{i-1} - x_i, x_i - y \rangle \\ &= \sum_{i=m+1}^n \langle e_i - e_{i-r}, x_i - y \rangle = \sum_{i=m+1}^n \langle e_i, x_i - y \rangle - \sum_{i=m+1}^n \langle e_{i-r}, x_i - y \rangle \\ &= \sum_{i=m+1}^n \langle e_i, x_i - y \rangle - \sum_{i=m+1}^n [\langle e_{i-r}, x_i - x_{i-r} \rangle + \langle e_{i-r}, x_{i-r} - y \rangle] \\ &= \sum_{i=m+1}^n \langle e_i, x_i - y \rangle - \sum_{i=m+1}^n \langle e_{i-r}, x_{i-r} - y \rangle + \sum_{i=m+1}^n \langle e_{i-r}, x_{i-r} - x_i \rangle \\ &= \sum_{i=m+1}^n \langle e_i, x_i - y \rangle - \sum_{i=m+1-r}^{n-r} \langle e_i, x_i - y \rangle + \sum_{i=m+1}^n \langle e_{i-r}, x_{i-r} - x_i \rangle \\ &= \sum_{i=n-r+1}^n \langle e_i, x_i - y \rangle - \sum_{i=m-r+1}^m \langle e_i, x_i - y \rangle + \sum_{i=m+1}^n \langle e_{i-r}, x_{i-r} - x_i \rangle. \end{aligned}$$

Sustituyendo los mismos valores en la igualdad (3.1)

$$\begin{aligned} \|x_m - y\|^2 &= \|x_n - y\|^2 + \sum_{k=m+1}^n \|x_{k-1} - x_k\|^2 + 2 \sum_{i=n-r+1}^n \langle e_i, x_i - y \rangle \\ &\quad + 2 \sum_{i=m+1}^n \langle e_{i-r}, x_{i-r} - x_i \rangle - 2 \sum_{i=m-r+1}^m \langle e_i, x_i - y \rangle, \end{aligned}$$

como se quería demostrar. \square

Lema 3.5 $\{x_n\}$ es una sucesión acotada, y además

$$\sum_1^{\infty} \|x_{k-1} - x_k\|^2 < \infty.$$

En particular,

$$\|x_n - x_{n-1}\| \rightarrow 0 \text{ cuando } n \rightarrow \infty$$

Demostración: Usando el lema 3.4, fijamos $m = 0$ y llegamos a que, para todo n

$$\begin{aligned} \|x_0 - y\|^2 &= \|x_n - y\|^2 + \sum_{k=1}^n \|x_k - x_{k-1}\|^2 + 2 \sum_{k=1}^n \langle e_{k-r}, x_{k-r} - x_k \rangle \\ &\quad + 2 \sum_{k=n-(r-1)}^n \langle e_k, x_k - y \rangle - 2 \sum_{k=-(r-1)}^0 \langle e_k, x_k - y \rangle. \end{aligned}$$

El tercer y cuarto término son no negativos (lema 3.2) y el último término vale 0 por como esta definido e_k . Entonces

$$\|x_0 - y\|^2 \geq \|x_n - y\|^2 + \sum_{k=1}^n \|x_k - x_{k-1}\|^2.$$

Con lo que $\|x_0 - y\|$ es una cota superior de $\|x_n - y\|$, y por tanto $\{x_n\}$ es acotado. Además, $\sum_{k=1}^n \|x_k - x_{k-1}\|^2 < \infty$. \square

Lema 3.6 Para todo $n \in \mathbb{N}$,

$$\|e_n\| \leq \sum_{k=1}^n \|x_k - x_{k-1}\|.$$

Demostración: Por inducción sobre n . Para $n = 1$,

$$\|e_1\| = \|x_0 - x_1 + e_{1-r}\| = \|x_0 - x_1\|$$

ya que $e_i = 0$ para todo $i \leq 0$. Asumamos ahora que se cumple para un n fijo, y comprobemos para $n + 1$

$$\begin{aligned} \|e_{n+1}\| &= \|x_n - x_{n+1} + e_{n+1-r}\| \leq \|x_n - x_{n+1}\| + \|e_{n+1-r}\| \\ &\leq \|x_n - x_{n+1}\| + \sum_{k=1}^{n+1-r} \|x_{k-1} - x_k\| \leq \sum_{k=1}^{n+1} \|x_{k-1} - x_k\|. \end{aligned}$$

\square

Lema 3.7

$$\liminf_n \sum_{k=n-(r-1)}^n |\langle x_k - x_n, e_k \rangle| = 0.$$

Demostración: Por la desigualdad de Schwarz 1.3 sabemos que

$$\sum_{k=n-(r-1)}^n |\langle x_k - x_n, e_k \rangle| \leq \sum_{k=n-(r-1)}^n \|e_k\| \|x_k - x_n\|$$

$$\begin{aligned}
&\leq \sum_{k=n-(r-1)}^n \left(\sum_{j=1}^k \|x_{j-1} - x_j\| \right) \left(\sum_{i=k+1}^n \|x_{i-1} - x_i\| \right) \\
&\leq r \left(\sum_{j=1}^n \|x_{j-1} - x_j\| \right) \left(\sum_{k=n-(r-2)}^n \|x_{i-1} - x_i\| \right).
\end{aligned}$$

Llamemos a $a_i = \|x_{i-1} - x_i\|$, bastaría entonces probar

$$\liminf_n \sum_{j=1}^n a_j \left(\sum_{i=n-(r-2)}^n a_i \right) = 0. \quad (3.2)$$

Sabemos ahora que, por el lema 3.5

$$A := \sum_{i=1}^{\infty} a_i^2 = \sum_{i=1}^{\infty} \|x_{i-1} - x_i\|^2 < \infty.$$

Aplicando de nuevo la desigualdad de Schwarz 1.3

$$\sum_{j=1}^n a_j \leq \sqrt{n} \left(\sum_1^n a_j^2 \right)^{1/2} \leq \sqrt{n} A^{1/2} \text{ para todo } n.$$

Bastaría probar que

$$\liminf_n \sqrt{n} \sum_{i=n-(r-2)}^n a_i = 0$$

ya que hemos encontrado una cota para la parte izquierda de (3.2).

Definamos γ como

$$\gamma = \liminf_n \sqrt{n} \sum_{i=n-(r-2)}^n a_i.$$

Supongamos ahora que $\gamma > 0$ (no puede ser menor, por como está definido a_i). Asumamos que $\gamma < \infty$ (en caso de ser infinito la demostración es similar)

$$\sqrt{n} \sum_{i=n-(r-2)}^n a_i > \frac{1}{2}\gamma \quad \text{para un } n \text{ suficientemente grande.}$$

Dividiendo por \sqrt{n} y aplicando la desigualdad de Schwarz 1.3

$$\frac{\gamma^2}{4n} < \left(\sum_{i=n-(r-2)}^n a_i \right)^2 \leq (r-1) \sum_{i=n-(r-2)}^n a_i^2 \quad \text{para un } n \text{ suficientemente grande.}$$

Pongamos que la desigualdad anterior es válida para un $n \geq N$, entonces

$$\begin{aligned} \infty &= \sum_{n=N}^{\infty} \frac{\gamma^2}{4n} \leq (r-1) \sum_{n=N}^{\infty} \sum_{i=n-(r-2)}^n a_i^2 \\ &\leq (r-1) \sum_{n=N}^{\infty} [a_{n-(r-2)}^2 + a_{n-(r-3)}^2 + \dots + a_n^2] \leq (r-1)^2 \sum_1^{\infty} a_i^2 < \infty \end{aligned}$$

que es un absurdo, así que $\gamma = 0$ □

Lema 3.8 Existe una subsucesión $\{x_{n_j}\}$ de $\{x_n\}$ tal que

$$\limsup_j \langle y - \{x_{n_j}\}, x - \{x_{n_j}\} \rangle \leq 0 \text{ para todo } y \in K$$

y

$$\lim_j \sum_{k=n_j-(r-1)}^{n_j} |\langle x_k - x_{n_j}, e_k \rangle| = 0.$$

Demostración: Usando el lema 3.3, para todo $y \in K$, $n \geq 0$

$$\begin{aligned} \langle y - x_n, x - x_n \rangle &= \langle y - x_n, e_{n-(r-1)} + e_{n-(r-2)} + \dots + e_n \rangle \quad (3.3) \\ &= \sum_{k=n-(r-1)}^n \langle y - x_n, e_k \rangle = \sum_{k=n-(r-1)}^n \langle y - x_k, e_k \rangle + \sum_{k=n-(r-1)}^n \langle x_k - x_n, e_k \rangle. \end{aligned}$$

Sabemos que el sumatorio es mayor que 0, ya que es un resultado del lema 3.2, entonces

$$\langle y - x_n, x - x_n \rangle \leq \sum_{k=n-(r-1)}^n \langle x_k - x_n, e_k \rangle.$$

Sabemos que existe una subsucesión $\{n_j\}$ de \mathbb{N} (lema 3.7) tal que

$$\lim_j \sum_{k=n_j-(r-1)}^{n_j} |\langle x_k - x_{n_j}, e_k \rangle| = 0.$$

Juntando la desigualdad anterior con este resultado concluimos la demostración. □

Teorema 3.9 Teorema Boyle-Dykstra Sean K_1, K_2, \dots, K_r , subconjuntos cerrados convexos de un espacio X de Hilbert tal que $K = \bigcap_1^r K_i \neq \emptyset$. Para todo $x \in X$ definimos la sucesión $\{x_n\}$

$$x_0 = x \quad e_{-(r-1)} = \dots = e_{-1} = e_0 = 0,$$

$$\begin{aligned}
x_n &= P_{K_{[n]}}(x_{n-1} + e_{n-r}), \\
e_n &= x_{n-1} + e_{n-r} - x_n \\
&= x_{n-1} + e_{n-r} - P_{K_{[n]}}(x_{n-1} + e_{n-r}) \quad (n = 1, 2, \dots)
\end{aligned}$$

Entonces

$$\lim_{n \rightarrow \infty} \|x_n - P_K(x)\| = 0.$$

Demostración: Sabemos que, por el lema 3.8, existe una subsucesión $\{x_{n_j}\}$ tal que

$$\limsup_j \langle y - x_{n_j}, x - x_{n_j} \rangle \leq 0 \text{ para todo } y \in K.$$

Por el lema 3.5, $\{x_n\}$ es acotado, y por lo tanto lo es la sucesión $\{x_{n_j}\}$. Como estamos en un espacio Hilbert, por el teorema 1.33 existe un $y_0 \in X$ tal que una subsucesión de la $\{x_{n_j}\}$, que seguiremos denotando de la misma forma, es tal que

$$x_{n_j} \xrightarrow{w} y_0 \quad (3.4)$$

y

$$\lim_j \|x_{n_j}\| \text{ existe.} \quad (3.5)$$

Como $x_{n_j} \xrightarrow{w} y_0$,

$$\|y_0\| \leq \liminf_j \|x_{n_j}\| = \lim_j \|x_{n_j}\|.$$

Como hay un número infinito de x_n y solo un número finito de K_i , K_{i_0} debe tener un número infinito de x_{n_j} . Al ser K_{i_0} cerrado y convexo, es débilmente cerrado (ya que todo conjunto convexo es cerrado si y solo si es débilmente cerrado). Por tanto $y_0 \in K_{i_0}$. Como hemos visto en el lema 3.5 $x_n - x_{n-1} \rightarrow 0$. Repitiendo este razonamiento se ve que las sucesiones $\{x_{n_j+1}\}, \{x_{n_j+2}\}, \dots$ convergen débilmente a y_0 , por lo que $y_0 \in K_i$ para todo i , es decir, $y_0 \in K$.

Como $y_0 \in K$ e $y_0 \leq \lim_j \|x_{n_j}\|$, se tiene para todo $y \in K$

$$\begin{aligned}
\langle y - y_0, x - y_0 \rangle &= \langle y, x \rangle - \langle y, y_0 \rangle - \langle y_0, x \rangle + \|y_0\|^2 \\
&\leq \lim_j [\langle y, x \rangle - \langle y, x_{n_j} \rangle + \|x_{n_j}\|^2] \\
&= \lim_j \langle y - x_{n_j}, x - x_{n_j} \rangle \leq 0.
\end{aligned}$$

Por la caracterización de la mejor aproximación para convexos 1.27 sabemos que $y_0 = P_K(x)$. Además, si en la cadena de desigualdades anteriores tomamos $y = y_0$ se da la igualdad y llegamos a

$$\lim_j \|x_{n_j}\|^2 = \|y_0\|^2$$

y

$$\lim_j \langle y_0 - x_{n_j}, x - x_{n_j} \rangle = 0.$$

Por (3.4) y (3.5), por el teorema 1.32(2) se sigue

$$\|x_{n_j} - y_0\| \rightarrow 0. \quad (3.6)$$

y por consiguiente

$$\|x_{n_j} - P_K(x)\| = \|x_{n_j} - y_0\| \rightarrow 0.$$

Nos faltaría ahora ver que la sucesión $\{x_n\}$ converge a y_0 . Poniendo $y = y_0$ y $n = n_j$ en la ecuación (3.3)

$$\langle y_0 - x_{n_j}, x - x_{n_j} \rangle = \sum_{k=n_j-(r-1)}^{n_j} \langle y_0 - x_k, e_k \rangle + \sum_{k=n_j-(r-1)}^{n_j} \langle x_k - x_{n_j}, e_k \rangle.$$

El lado izquierdo de la desigualdad anterior tiende a 0 conforme $j \rightarrow \infty$, mientras el segundo término del lado derecho tiene a 0 por el lema 3.8. Por tanto

$$\lim_j \sum_{k=n_j-(r-1)}^{n_j} \langle y_0 - x_k, e_k \rangle = 0. \quad (3.7)$$

Haciendo $m = n_j$ y $y = y_0$ en los lemas 3.4 y 3.2, vemos que, para todo $n \geq n_j$,

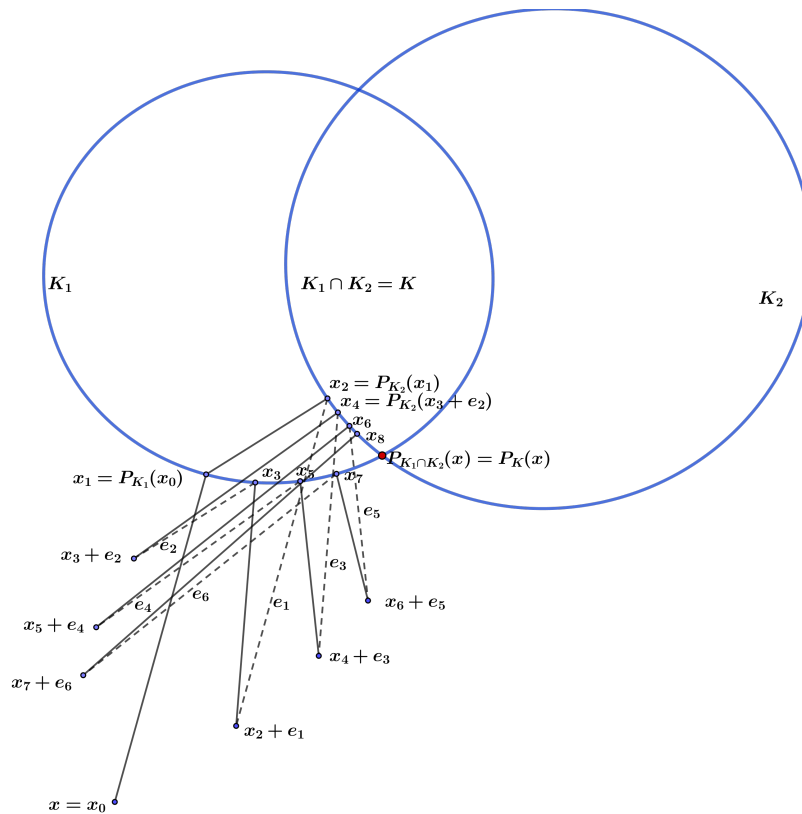
$$\|x_{n_j} - y_0\|^2 \geq \|x_n - y_0\|^2 - 2 \sum_{k=n_j-(r-1)}^{n_j} \langle e_k, x_k - y_0 \rangle,$$

o

$$\|x_n - y_0\|^2 \leq \|x_{n_j} - y_0\|^2 + 2 \sum_{k=n_j-(r-1)}^{n_j} \langle e_k, x_k - y_0 \rangle.$$

Ambos lados de la igualdad tienen a 0 conforme $j \rightarrow \infty$ por (3.6) y (3.7). Concluimos que $\lim_n \|x_n - y_0\| = 0$ y terminamos. \square

Habiendo probado la convergencia del Algoritmo de Dykstra, mostraremos un ejemplo gráfico de él.



Notas:

1. Por no sobrecargar el dibujo, omitimos la definición de los x_n para algunos n , pero, como se especifica en la descripción del algoritmo, con $r = 2$, se definen $x_n = P_{K_{[n]}}(x_{n-1} + e_{n-2})$.
2. Haría falta seguir haciendo iteraciones para aproximarse a la solución real, no se hacen más para no sobrecargar el dibujo.

3.3. Caso de espacios afines

En el caso de que los conjuntos convexos cerrados sean subespacios vectoriales cerrados de X , el algoritmo de Dykstra se simplifica sustancialmente, pudiendo omitirse el cálculo de las correcciones e_n en todo el algoritmo. En caso de que fueran subespacios (menos general), el algoritmo resultante sería equivalente al de Von Neumann 2.9.

De facto, ésta misma situación resulta cuando los $K_i, i = 1, \dots, r$, son variedades afines de X . La siguiente proposición formaliza esta situación.

Proposición 3.10 Caracterización de la mejor aproximación para espacios afines.

Sea V un espacio afín de un espacio X con producto interno. Entonces, $V = M + v$ donde M es un subespacio y v es un elemento de V . Sea $x \in X$ e $y_0 \in V$, entonces es equivalente

1. $y_0 = P_V(x)$.
2. $x - y_0 \in M^\perp$.
3. $\langle x - y_0, y - v \rangle = 0$ para todo $y \in V$.

y además,

$$P_V(x + e) = P_V(x) \text{ para todo } x \in X, e \in M^\perp. \quad (3.8)$$

Demostración: Por la propiedad 2 del lema 1.22 $P_{K+y}(x + y) = P_K(x) + y$ para cualquier $x, y \in X$.

Por lo anterior, $y_0 = P_V(x)$, si y sólo si, $y_0 = P_{M+v}(x) = P_M(x - v) + v$ y por tanto $y_0 - v = P_M(x - v)$. Usando la caracterización de mejor aproximación para subespacios 1.28 deducimos que esto equivale a que $x - y_0 = (x - v) - (y_0 - v) \in M^\perp$, y tenemos (2). Al ser $M = V - v$ tenemos la implicación (2) \Rightarrow (3).

Para ver (3.8), sea $x \in X$ y $e \in M^\perp$. Usando que $P_{K+y}(x + y) = P_K(x) + y$ junto con la linealidad de P_M y que $P_M(e) = 0$, vemos que

$$\begin{aligned} P_V(x + e) &= P_{M+v}(x + e) = P_M(x + e - v) + v \\ &= P_M(x - v) + v = P_{M+v} = P_{M+v}(x) = P_V(x). \end{aligned}$$

□

Corolario 3.11 Teorema de Halperin. Sean M_1, M_2, \dots, M_r subespacios cerrados en un espacio de Hilbert H y $M = \bigcap_1^r M_i$. Entonces,

$$\lim_{n \rightarrow \infty} (P_{M_r} P_{M_{r-1}} \dots P_{M_1})^n(x) = P_M(x) \text{ para todo } x \in H.$$

Notemos que este teorema es una generalización del teorema de Von Neumann 2.9 para cualquier número finito de subespacios, no solo 2.

Cabe destacar que este teorema fue establecido antes que el teorema de Boyle-Dykstra, que generaliza el teorema de Von Neumann a un número finito de subespacios. El algoritmo de Dykstra y su convergencia se estableció inicialmente para conjuntos K_i , $i = 1, \dots, r$, que son conos convexos de X . Posteriormente fue extendido al caso en que los K_i , $i = 1, \dots, r$ son conjuntos convexos cerrados con intersección no vacía de un espacio de Hilbert H .

Capítulo 4

Aplicación del algoritmo de Dykstra a un problema de mínimos cuadrados con matrices

4.1. Explicación del problema

Aplicaremos el algoritmo de Dykstra de proyecciones alternadas a un problema de mínimos cuadrados con matrices. Buscaremos la matriz simétrica más próxima, usando la norma de Frobenius, a una matriz dada. La matriz tendrá que cumplir ciertas restricciones cuya formulación equivale a la definición de subconjuntos convexos de matrices.

1. Tiene que estar entre dos matrices L y U (Lower-Upper). Decimos que está entre dos matrices si elemento a elemento está en el intervalo, es decir, $L_{ij} \leq X_{ij} \leq U_{ij}$ siendo X la matriz que buscamos.

$$L \leq X \leq U.$$

Llamaremos al conjunto que lo cumple \mathcal{B} y a su proyección $P_{\mathcal{B}}$.

2. La matriz tiene que seguir un patrón \mathcal{P} dado. Dicho patrón está definido por matrices G_1, \dots, G_m , ($1 \leq m \leq n(n+1)/2$), simétricas, cuyos valores son 0 o 1 y tienen la propiedad de que, para cada entrada st , $1 \leq s, t \leq n$, existe un y solo un k tal que $(G_k)_{st} = 1$. Entonces $\mathcal{P} = \sum_i \alpha_i G_i$ con $\alpha_i \in \mathbb{R}$. Llamaremos al conjunto de estas matrices $\mathcal{P} = \{\sum_{i=1}^m \alpha_i G_i \mid \alpha_i \in \mathbb{R}\}$ y a su proyección $P_{\mathcal{P}}$. Para ilustrar esta definición, veamos un ejemplo. Sea \mathcal{P}

$$\mathcal{P} = \begin{pmatrix} \alpha_1 & \alpha_2 & 0 \\ \alpha_2 & \alpha_1 & \alpha_3 \\ 0 & \alpha_3 & \alpha_1 \end{pmatrix}$$

Podemos escribir \mathcal{P} como $\alpha_1 G_1 + \alpha_2 G_2 + \alpha_3 G_3$, donde

$$G_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, G_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, G_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

3. La matriz debe ser simétrica, definida positiva y además, el más pequeño de los autovalores tiene que ser mayor que un $\epsilon > 0$ dado, es decir

$$\lambda_{\min}(X) \geq \epsilon > 0,$$

denotando con $\lambda_{\min}(X)$ al autovalor de X más pequeño. Llamaremos al conjunto que lo cumple ϵpd y a su proyección $P_{\epsilon pd}$.

Vamos a demostrar que podemos aplicar el algoritmo de Dykstra con estas condiciones. Para ello, tenemos que ver que las definiciones anteriores definen conjuntos convexos de matrices.

Demostración:

1. Para ver que \mathcal{B} es convexo tenemos que ver que, para todo par de matrices $A, B \in \mathcal{B}$, $A\lambda + (1 - \lambda)B \in \mathcal{B}$ con $\lambda \in [0, 1]$.

Al ser la comparación con L, U elemento a elemento, tenemos que ver que para todo i, j ,

$$L_{ij} \leq A_{ij}\lambda + (1 - \lambda)B_{ij} \leq U_{ij}.$$

Fijemos un par de índices i, j , con $1 \leq i, j \leq n$, y supongamos, sin pérdida de generalidad que $A_{ij} \leq B_{ij}$ (en caso de ser menor la demostración es similar). Entonces

$$L_{ij} \leq A_{ij} \leq A_{ij}\lambda + (1 - \lambda)B_{ij} \leq B_{ij} \leq U_{ij}$$

ya que $A, B \in \mathcal{B}$. Por lo tanto, $\lambda A + (1 - \lambda)B \in \mathcal{B}$ y concluimos.

2. Para ver que \mathcal{P} es convexo tenemos que ver que, para todo par de matrices $A, B \in \mathcal{P}$, $A\lambda + (1 - \lambda)B \in \mathcal{P}$ con $\lambda \in [0, 1]$. Pongamos

$$A = \sum_i^m \alpha_i G_i, \quad A \in \mathcal{P}$$

$$B = \sum_i^m \beta_i G_i \quad B \in \mathcal{P}$$

entonces,

$$\begin{aligned} \lambda A + (1 - \lambda)B &= \sum_i^m \alpha_i G_i \lambda + (1 - \lambda) \sum_i^m \beta_i G_i \\ &= \sum_i^m (\lambda \alpha_i + (1 - \lambda) \beta_i) G_i \in \mathcal{P} \end{aligned}$$

3. Para ver que ϵpd es convexo tenemos que ver que, para todo par de matrices $A, B \in \epsilon pd$, $A\lambda + (1-\lambda)B \in \epsilon pd$ con $\lambda \in [0, 1]$. La simetría de $\lambda A + (1-\lambda)B$ es obvia.

Sabemos que, como $A, B \in \epsilon pd$, $x^T Ax \geq \epsilon \|x\|^2$ y $x^T Bx \geq \epsilon \|x\|^2$ para todo $x \in \mathbb{R}^n$. Entonces

$$x^T (\lambda A + (1-\lambda)B)x = \lambda (x^T Ax) + (1-\lambda) (x^T Bx) \geq \lambda \epsilon \|x\|^2 + (1-\lambda) \epsilon \|x\|^2 = \epsilon \|x\|^2$$

para todo $x \in \mathbb{R}^n$.

□

En resumen, el problema que tenemos que resolver es

Problema 4.1 Buscamos la matriz X que minimiza

$$\|X - A\|_F^2$$

siendo $\|\cdot\|_F$ la norma de Frobenius, y sujeta a las restricciones siguientes:

$$X^T = X$$

$$L \leq X \leq U$$

$$\lambda_{\min}(X) \geq \epsilon > 0$$

$$X \in \mathcal{P}.$$

4.2. Solución al problema

Procederemos ahora a ver como se calcula la proyección de cada espacio

Proposición 4.2 Sea $A \in \mathbb{R}^{n \times n}$ entonces, la única solución al problema

$$\min_{X \in \mathcal{P}} \|X - A\|_F^2$$

es una matriz $n \times n$ definida

$$(P_{\mathcal{P}}(A))_{ij} = \begin{cases} A_{ij} & \text{si } L_{ij} \leq A_{ij} \leq U_{ij} \\ U_{ij} & \text{si } A_{ij} > U_{ij} \\ L_{ij} & \text{si } A_{ij} < L_{ij} \end{cases}$$

Demostración: Consecuencia directa de la definición de la norma de Frobenius 1.9. □

Proposición 4.3 Sea $A \in \mathbb{R}^{n \times n}$ entonces, la única solución al problema

$$\min_{X \in \mathcal{P}} \|X - A\|_F^2$$

es una matriz $n \times n$ de la forma

$$P_{\mathcal{P}}(A) = \sum_{k=1}^m \alpha_k G_k$$

donde cada α_k esta definida

$$\alpha_k = \frac{\sum_{i,j=1}^n A_{ij}(G_k)_{ij}}{\sum_{i,j=1}^n (G_k)_{ij}} \text{ para todo } 1 \leq k \leq m.$$

Demostración: Vamos a definir $f : \mathbb{R}^m \rightarrow \mathbb{R}$,

$$f(\alpha_1, \dots, \alpha_m) = \left\| \sum_{k=1}^m \alpha_k G_k - A \right\|_F^2.$$

Si calculamos el gradiente y hacemos que sea 0 (para que sea mínimo), vemos que solo es 0 cuando α_k están definidos como previamente se ha visto. □

Proposición 4.4 Sea $A \in \mathbb{R}^{n \times n}$, definimos $B = (A + A^T)/2$ (la parte simétrica) y $C = (A - A^T)/2$ (la parte antisimétrica). La única solución al problema

$$\min_{X \in \text{epd}} \|X - A\|_F^2$$

viene dada por

$$P_{\text{epd}}(A) = Z \text{diag}(d_i) Z^T$$

donde Z es una matriz que cumple $B = Z \Delta Z^T$ es la descomposición espectral de B, es decir, $Z^T Z = I$ y $\Delta = \text{diag}(\lambda_i(B))$, y d_i esta definida

$$d_i = \begin{cases} \lambda_i(B) & \lambda_i(B) \geq \epsilon \\ \epsilon & \lambda_i(B) < \epsilon \end{cases}$$

Además,

$$\min_{X \in \text{epd}} \|X - A\|^2 = \sum_{\lambda_i(B) < 0} (\lambda_i(B) - \epsilon)^2 + \|C\|_F^2.$$

Para demostrar esto necesitamos un teorema previo.

Teorema 4.5 Sea $A \in \mathbb{R}^{n \times n}$, y sean B y C las matrices previamente definidas. Hagamos la descomposición polar sobre B , es decir, $B = UH$ con $U^T U = I$ y $H = H^T \geq 0$. Entonces $X = (B + H)/2$ es el único aproximante positivo de A (usando la norma de Frobenius) y

$$\min_{X \in pd} \|X - A\|_F^2 = \sum_{\lambda_i(B) < 0} \lambda_i(B)^2 + \|C\|_F^2.$$

Demostración: Sea X semidefinida positiva. Sabemos que si $S = S^T$ y $K = -K^T$ se tiene que $\|S + K\|_F^2 = \|S\|_F^2 + \|K\|_F^2$. Entonces tenemos

$$\|A - X\|_F^2 = \|B - X\|_F^2 + \|C\|_F^2$$

y el problema se reduce a encontrar la mejor aproximación para B . Sea $B = Z\Delta Z^T$ la descomposición espectral definida antes, llamemos $Y = Z^T X Z$, entonces

$$\begin{aligned} \|B - X\|_F^2 &= \|\Delta - Y\|_F^2 = \sum_{i \neq j} y_{ij}^2 + \sum_i (\lambda_i - y_{ii})^2 \\ &\geq \sum_{\lambda_i < 0} (\lambda_i - y_{ii})^2 \geq \sum_{\lambda_i < 0} \lambda_i^2 \end{aligned}$$

al ser los $y_{ii} \geq 0$, porque Y es semidefinida positiva. La cota inferior viene dada por la matriz $Y = \text{diag}(d_i)$ donde

$$d_i = \begin{cases} \lambda_i(B) & \lambda_i(B) \geq 0 \\ 0 & \lambda_i(B) < 0 \end{cases}$$

o lo que es lo mismo

$$X = Z \text{diag}(d_i) Z^T.$$

Por tanto, $X = (B + H)/2$, siendo $H = Z \text{diag}(|\lambda_i|) Z^T$. □

Una vez demostrado este teorema, ver el caso de nuestro problema es muy sencillo

Demostración: (de la proposición 4.4) Aplicamos el teorema anterior a $A - \epsilon I$. □

Con esto hemos visto que se puede aplicar el algoritmo de Dykstra y como proyectar en cada conjunto. Siguiendo los pasos que indica el algoritmo Dykstra 3.2 llegaríamos a la solución.

4.3. Experimentos

Haremos dos experimentos para ver si funciona el programa.

Experimento 1

Probaremos el algoritmo para matrices 4×4 . La matriz dada es:

$$A = \begin{pmatrix} 1 & 3 & 4 & 2 \\ 0 & 1 & -1 & 6 \\ 7 & -2 & 1 & 2 \\ 2 & 5 & 2 & 0,5 \end{pmatrix}$$

La condición LU viene dada por

$$L = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} \quad U = \begin{pmatrix} 8 & 3 & 0 & 2 \\ 3 & 7 & 4 & 0 \\ 0 & 2 & 6 & 3 \\ 2 & 0 & 3 & 6 \end{pmatrix}$$

Y el patrón a seguir es:

$$\mathcal{P} = \begin{pmatrix} \alpha_1 & \alpha_2 & 0 & \alpha_3 \\ \alpha_2 & \alpha_1 & \alpha_3 & 0 \\ 0 & \alpha_3 & \alpha_1 & \alpha_2 \\ \alpha_3 & 0 & \alpha_2 & \alpha_1 \end{pmatrix}$$

Tomamos el valor para $\epsilon = 0,1$.

La solución exacta es (se muestra la parte triangular superior)

$$X = \begin{pmatrix} 2 & 1,825 & 0 & 0,075 \\ & 2 & 0,075 & 0 \\ & & 2 & 1,825 \\ & & & 2 \end{pmatrix}$$

Y esta es la tabla de resultados computando, donde (de aquí en adelante), *Time* muestra el tiempo CPU en segundos, *It.* el numero de iteraciones para diferentes tolerancias, *Tol* la tolerancia exigida (calculada $\|P_{\mathcal{P}}(A_{i+1}) - P_{\mathcal{P}}(A_i)\|_F$) y el *Error* en la norma de Frobenius entre las aproximaciones obtenidas y la solución exacta X .

Tol	It.	Time	Error
1e-2	19	0.0077	2.1e-2
1e-5	38	0.0119	2.1e-5
1e-10	70	0.017	1.8e-10

NOTA: El Error es calculado haciendo $\|X - S\|_F$, siendo S el resultado obtenido con cada tolerancia.

Experimento 2

Este experimento valdría para cualquier tamaño de matriz, por la forma en la que esta definido: La matriz A dada viene definida por $A_{i,j} = i - j + \frac{i}{i+j-1}$. Definimos las matrices de la condición LU como la matriz de ceros para L y $U_{i,j} = i + j$. Como en el anterior experimento $\epsilon = 0,1$. Y por último, el patrón \mathcal{P} viene dado por la matriz de *Toepliz* simétrica, definida $\mathcal{P}_{i,j} = \mathcal{P}_{i+1,j+1}$ para todo i, j .

Haremos dos computaciones con este experimento, estas son las tablas de los resultados para diferentes tolerancias

1. Matriz tamaño 10×10 .

Tol	It.	Time	Error
1e-2	18	0.072	0.044
1e-5	165	0.723	7.73e-4
1e-7	560	2.48	9.03e-6

NOTAS: El Error es calculado haciendo $\|X - S\|_F$, siendo S el resultado obtenido con cada tolerancia y X el resultado obtenido con tolerancia $1e - 12$. No se exponen los resultados por el tamaño que ocupan.

2. Matriz tamaño 100×100 .

Tol	It.	Time	Error
1e-2	125	573	0.3744
1e-5	874	3889	-

NOTAS: El Error calculado para la tolerancia $1e - 2$ se hace con $\|X - S\|_F$, siendo X el resultado obtenido con tolerancia $1e - 5$. Debido al tamaño de la solución no la escribiremos.

Experimento 3

Con este experimento vamos a ver que la diferencia en los ratios de convergencia viene fuertemente marcada por la matriz inicial.

Sean A_1 y A_2 matrices 5×5 , donde cada entrada i, j de A_1 viene dada por $i + 5(j - i)$, y A_2 es la matriz identidad. El patrón \mathcal{P} es el subconjunto de

matrices simétricas, $\epsilon = 0,1$ y la tolerancia que pediremos será $5e - 6$. Por último, las matrices para la condición LU son

$$L = \begin{pmatrix} 1 & 1 & 0 & 3 & 1 \\ 1 & 2 & 2 & 0 & 3 \\ 0 & 2 & 3 & 2 & 0 \\ 3 & 0 & 2 & 2 & 0 \\ 1 & 3 & 0 & 0 & 1 \end{pmatrix}; U = \begin{pmatrix} 3 & 3 & 2 & 7 & 5 \\ 3 & 5 & 4 & 2 & 6 \\ 6 & 3 & 6 & 10 & 3 \\ 7 & 5 & 8 & 5 & 3 \\ 5 & 9 & 5 & 3 & 3 \end{pmatrix}$$

Veamos la tabla para este experimento

Matrix	It.	Time	Error
A_1	614	0.624	4.37e-4
A_2	66	0.068	2.69e-5

Podemos ver que para el mismo problema y misma tolerancia la diferencia entre el número de iteraciones es más que notable. Podemos ver que A_2 ya cumple algunas de las condiciones que se piden (la condición patrón y la condición ϵpd) con lo que se puede prever que vaya a converger más rápido.

NOTAS: El Error es calculado haciendo $\|X - S\|_F$, siendo S el resultado obtenido con cada tolerancia y X el resultado obtenido con tolerancia $5e - 12$.

La implementación a Matlab está en el ANEXO.

Capítulo 5

ANEXO

Veremos como hemos hecho la **implementación a Matlab** para computar el problema visto en el Capítulo 4.

5.1. Programas auxiliares

Condición Lower-Upper

```
1      %% Iñigo San Jose Visiers  %%
2      %% TFG Algoritmo de Dykstra  %%
3      %% Tutor: Luis Abia  %%
4
5      %M=matriz a aproximar con limites superior e
6      inferior
7      %L y U= limite inferior y superior
8      respectivamente. Comparacion elemento a
9      % elemento
10
11     function [MF]=ConLU(M,L,U)
12     format long
13     n=length(M(1,:));
14     MF=M;
15     for i=1:n
16         for j=1:n
17             if M(i,j)>U(i,j)
18                 MF(i,j)=U(i,j);
19             elseif M(i,j)<L(i,j)
20                 MF(i,j)=L(i,j);
21             end
22         end
23     end
24 end
25 end
```

Nos permite acotar componente a componente entre dos matrices L y U (Lower y Upper), pasándole una matriz M y los límites L y U . Y nos devuelve una matriz dentro de esos límites.

El programa consiste de dos bucles *for* para recorrer la matriz M y una comparación con el condicional *if* para sustituir cada componente ajustándolos a sus límites superior e inferior.

Condición de patrón

```

1 %% Iñigo San Jose Visiers %%
2 %% TFG Algoritmo de Dykstra %%
3 %% Tutor: Luis Abia %%
4
5 %M=matriz a aproximar siguiendo un patron dado
6 %P=matriz de matrices, en la que cada matriz es una
   matriz simetrica
7 % que es representativa del patron
8
9 function [MF]=ConPat(M,P)
10 format long
11
12
13 n=length(M(1,:));
14 p=length(P(1,1,:));
15 alpha=zeros(n,n); %escalar
16 MF=zeros(n,n); %sumatorio matrices
17 for i=1:p %numero de simetricas
18     sumNUM=0;
19     sumDEN=0;
20     for k=1:n
21         for l=1:n
22             sumNUM=sumNUM+M(k,l)*P(k,l,i);
23             sumDEN=sumDEN+P(k,l,i);
24         end
25     end
26     alpha(i)=sumNUM/sumDEN;
27     MF=MF+alpha(i).*P(:, :, i);
28 end
29
30 end

```

Este programa corresponde al subespacio que nos condiciona el patrón que ha de tener la matriz, para llamar a la función se le pasa la matriz a condicionar M , y un grupo de matriz de matrices P donde cada G_i es una matriz simétrica

que forma parte de dicho patrón, tal que $\sum G_i = P$, y nos devuelve la matriz con ese patrón.

Sabemos que la mejor aproximación para patrones es

$$P_{\mathcal{P}}(M) = \sum_{k=1}^N \alpha_k G_k$$

Donde N es el número de matrices simétricas que conforman P y α_k esta definido como indica la proposición 4.3. El programa establece dos sumatorios para cada valor de p (cada matriz G_i) en el primer bucle *for*, el primero hace referencia al numerador y el segundo al denominador. Después recorre la matriz para ir calculando el valor de cada alpha, y suma el resultado de multiplicar αG_i a una matriz sumatorio que nos da el valor final al terminar con todos los G_i

Condición de ϵ -Definida positiva

```

1  %% Iñigo San Jose Visiers  %%
2  %% TFG Algoritmo de Dykstra  %%
3  %% Tutor: Luis Abia  %%
4
5  %M=Matriz a aproximar para que sea definida positiva con
   autovalores mayores que un eps dado
6  %eps= epsilon
7
8  function [MF]=ConDef(M,eps)
9  %format long
10
11 B=(M+M') / 2;
12 C=(M-M') / 2;
13 [Z,AV]=eig(B); %descomposicion espectral B=Z*av(B)*Z'
14 nav=length(AV);
15 diag=zeros(nav,nav);
16 for i=1:nav
17     if AV(i,i)<eps
18         diag(i,i)=eps;
19     else
20         diag(i,i)=AV(i,i);
21     end
22 end
23 MF=Z*diag*Z';

```

Permite calcular la matriz mas próxima que cumple que sea definida positiva y el menor de sus autovalores es mayor o igual a ϵ . Para llamar a la función se le pasa una matriz M a aproximar y el ϵ que se quiera.

El programa calcula los autovalores de B y las Z con la función ya implementada *eig*. Después recorre el vector de autovalores y, si son menores que ϵ los sustituye. Después calcula la matriz más próxima como indica la proposición 4.4

5.2. Implementación del algoritmo de Dykstra

Ya con todas las matrices auxiliares podemos ver el programa que nos resuelve el problema dado.

Algoritmo de Dykstra

```

1 %% Iñigo San Jose Visiers %%
2 %% TFG Algoritmo de Dykstra %%
3 %% Tutor: Luis Abia %%
4
5 %A=matriz a aproximar siguiendo un patron dado, con
   limites sup e inf
6 % exigiendo que sea def positiva con autovalores mayores
   que un epsilon
7 % dado
8 %L y U= limite inferior y superior respectivamente.
   Comparacion elemento a
9 % elemento
10 %P=matriz de matrices, en la que cada matriz es una
   matriz simetrica
11 % que es representativa del patron
12 %eps= epsilon
13 %TOL= tolerancia admitida
14
15 function [X] = DykstraTOL(A,L,U,P,eps,TOL,max)
16 tic
17 format long
18 n=length(A(1,:));
19 %no dependen del E
20 Av(:, :, 1)=A; %A0
21 Av(:, :, 2)=ConLU(A,L,U);
22 Av(:, :, 3)=ConPat(Av(:, :, 2),P);
23 Av(:, :, 4)=ConDef(Av(:, :, 3),eps); %A3
24

```

```

25 zero=zeros(n,n);
26 Ev(:,:,1)=zero; %E -2
27 Ev(:,:,2)=zero;
28 Ev(:,:,3)=zero; %E 0
29 i=1;
30 tol0=999;%garantizar que entra
31
32
33 while i<max && tol0>TOL
34     %ConLu
35     Ev(:,:,4)=Av(:,:,1)+Ev(:,:,1)-Av(:,:,2);
36     Av(:,:,5)=ConLU(Av(:,:,4)+Ev(:,:,4),L,U);
37     %ConPat
38     Ev(:,:,5)=Av(:,:,2)+Ev(:,:,2)-Av(:,:,3);
39     Av(:,:,6)=ConPat(Av(:,:,5)+Ev(:,:,5),P);
40     %ConDef
41     Ev(:,:,6)=Av(:,:,3)+Ev(:,:,3)-Av(:,:,4);
42     Av(:,:,7)=ConDef(Av(:,:,6)+Ev(:,:,6),eps);
43     tol0=norm(Av(:,:,7)-Av(:,:,4),'fro');
44
45     %Ajustamos para la siguiente iteracion
46     Av(:,:,1)=Av(:,:,4);
47     Av(:,:,2)=Av(:,:,5);
48     Av(:,:,3)=Av(:,:,6);
49     Av(:,:,4)=Av(:,:,7);
50     Ev(:,:,1)=Ev(:,:,4);
51     Ev(:,:,2)=Ev(:,:,5);
52     Ev(:,:,3)=Ev(:,:,6);
53     i=i+1;
54
55 end
56 X=Av(:,:,7);
57 i+2 %+2 por los ciclos iniciales
58 tol0
59 toc
60 end

```

El programa resuelve el problema de este Trabajo de Fin de Grado, es decir, el algoritmo de proyecciones alternadas de Dykstra. Para llamar a la función tenemos que pasarle todos los datos de los problemas anteriores y, además, le damos dos salidas, una la tolerancia (con TOL), y un número máximo de iteraciones (con max), como precaución.

El programa sigue lo explicado en el algoritmo 3.2, calcula las primeras aproximaciones (de la $A_0 = A$ a la A_3) ya que esas no dependen de las e por como están definidas. Hace lo mismo con las primeras e (de e_{-2} a e_0) ya que son

nulas, tanto las A_n como e son matrices de matrices. Después hace un ciclo completo de proyecciones (es decir, primero con LU , luego el patrón y por último que sea definida positiva) mientras va calculando las e necesarias para los pasos siguientes, siguiendo las definiciones del algoritmo 3.2. Después, para ahorrar memoria, sustituye los valores para el siguiente ciclo. Se calcula al final de cada ciclo la tolerancia y el programa termina cuando cumple la restricción de la tolerancia o excede el número de iteraciones.

Bibliografía

- [1] Deutsch, F. “Best Approximation in inner Product Spaces” *CMS Books in Mathematics*, 2001.
- [2] Escalante, R., Raydan, M. “Dykstra’s Algorithm for a Constrained Least-squares Matrix Problem” *Numerical Linear Algebra with Applications, Vol.3 (6)*, 459-471, 1996.
- [3] Escalante, R., Raydan, M. “Dykstra’s Algorithm for a Constrained Least-squares Rectangular Matrix Problems” *Computers Math. Applic. Vol. 35, No.6, pp. 73-79*, 1998.
- [4] Higham, N.J. “Computing a nearest symmetric positive semidefinite matrix” *Linear Algebra Appl, 103, 103-118*, 1988.
- [5] Li, J., Hu, X., Lei, Z. “Dykstra’s algorithm for a constrained least-squares doubly symmetric matrix problems” *Theoretical Computer Science 411, 2818-2826*, 2010.