



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico II

Métodos Numéricos
Segundo Cuatrimestre de 2015

Integrante	LU	Correo electrónico
Iván Arcuschin	678/13	iarcuschin@gmail.com
Martín Jedwabny	885/13	martiniedva@gmail.com
José Massigoge	954/12	jmmassigoge@gmail.com
Iván Pondal	078/14	ivan.pondal@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

<http://www.fcen.uba.ar>

Índice

1. Introducción	3
2. Modelo	4
2.1. Rankings de Páginas Web	4
2.1.1. PageRank	4
2.1.2. PageRank con matriz Esparsa	5
2.2. Rankings en competencias deportivas	5
2.2.1. Generalized Markov chains Method (GeM)	5
3. Demostración: Eliminación Gaussiana sin pivoteo	7
4. Implementación	8
5. Experimentación	9
6. Conclusión	10

1. Introducción

El objetivo de este Trabajo Práctico es implementar diferentes algoritmos de resolución de sistemas de ecuaciones lineales y experimentar con dichas implementaciones en el contexto de un problema de la vida real.

El problema a resolver es hallar la isoterma $500C^{\circ}$ en la pared de un Alto Horno. Para tal fin, deberemos particionar la pared del horno en puntos finitos, y luego resolver un sistema de ecuaciones lineales, en el cual cada punto de la pared interior y exterior del Horno es un dato, y las ecuaciones para los puntos internos satisfacen la ecuación del calor.

Los experimentos realizados se dividen en dos partes: Comportamiento del sistema y Evaluación de los métodos. En la primera parte, analizaremos con distintas instancias de prueba y se estudiará la proximidad de la isoterma buscada respecto de la pared exterior del horno. En la segunda parte, analizaremos el tiempo de computo requerido para la resolución del sistema en función de la granularidad de la discretización y analizaremos el escenario en el cual las temperaturas de los bordes varían a lo largo del tiempo.

2. Modelo

2.1. Rankings de Páginas Web

Hoy en día la cantidad de páginas web en todo el mundo asciende a una cantidad de 4.79 billones (sólo las indexadas). Es por esta razón que los buscadores (Search Engines) cumplen un rol tan importante en el uso diario de internet desde hace muchos años.

Dichos buscadores nos permiten realizar búsquedas de páginas web mediante distintos criterios, facilitando el acceso a la información.

Un posible criterio (bastante utilizado) es considerar que las páginas web populares son las más buscadas. De esta forma, el buscador puede ofrecernos en orden descendiente de popularidad los resultados obtenidos, esperando que encontremos más rápido lo que buscamos.

Siguiendo esta intuición, se han elaborado diferentes algoritmos para “rankear” las páginas web. A continuación presentaremos el famoso método PageRank, utilizado por Google en sus comienzos.

2.1.1. PageRank

El algoritmo de PageRank¹ se define para un conjunto de páginas Web = $\{1, \dots, n\}$ de forma tal de asignar a cada una de ellas un puntaje que determine la importancia relativa de la página respecto de las demás.

Llamemos x_j al puntaje asignado a la página $j \in \text{Web}$, que es lo que buscamos calcular.

Ahora, un link saliente de la página j a la página i puede significar que i es una página importante. Pero bien podría ser que j sea una página muy poco importante, por lo que deberíamos ponderar sus links salientes para decidir la importancia de las páginas a las que apunta.

Luego, vamos a considerar que la importancia de la página i obtenida mediante el link de j es proporcional a la importancia de j e inversamente proporcional al grado de j . Entonces, si $L_k \in \text{Web}$ es el conjunto de páginas web que apuntan a la página k :

$$x_k = \sum_{j \in L_k} \frac{x_j}{n_j}, \quad k = 1, \dots, n \quad (1)$$

En este algoritmo, hallaremos los x_k modelando el problema como una cadena de Markov, a la cual llamaremos Matriz de Transición, y que construiremos de la siguiente forma:

1. Sea G el grafo de la Web, dónde cada vértice es una página web y un eje de v a u significa que la página v tiene link saliente hacia u .
2. Luego, sea $W \in \{0, 1\}^{n \times n}$ la *matriz de conectividad* de G , tal que la celda $\{i, j\}$ tiene un 1 si hay un link saliente de la j -ésima página a la i -ésima página (los *autolinks* son ignorados, o lo que es lo mismo, $\forall 1 \leq i \leq n \ W_{i,i} = 0$).
3. Si definimos $n_j = \sum_{i=1}^n W_{i,j}$ como el grado de j (la cantidad de links salientes), entonces podemos definir la matriz $P \in \mathbb{R}^{n \times n}$, tal que la $P_{i,j} = 1/n_j * W_{i,j}$, y P es estocástica por columnas.
Además, notese que resolver el sistema dado por 1 es equivalente a encontrar un $x \in \mathbb{R}^n$ tal que $Px = x$. Es decir, encontrar el autovector asociado al autovalor 1 de P tal que $x_i > 0$ y $\sum_{i=1}^n x_i = 1$.
4. Ahora, puede pasar que para algún j , $n_j = 0$ lo que indicaría que la página j no tiene ningún link saliente. Para remediar estos casos, vamos a modificar P utilizando la idea del *navegante aleatorio*, de forma tal que para un j sin links salientes, la probabilidad de que el navegante salte a cualquier otra página i es $1/n$.

Entonces, $P_1 = P + D$, dónde $D = vd^t$, $d \in \{0, 1\}^n$ tal que $d_j = 1$ si $n_j = 0$, y $d_j = 0$ en caso contrario, y $v \in \mathbb{R}^n$ tal que $v_j = 1/n$.

¹Kurt Bryan and Tanya Leise. The linear algebra behind google. SIAM Review, 48(3):569–581, 2006.

5. Entonces, P_1 es estocástica por columnas, pero puede que no sea regular. Para que sí lo sea, extendemos el concepto anterior a todas las páginas (fenómeno de *teletransportación*).

Luego, $P_2 = c * P_1 + (1 - c) * E$, donde $\forall 1 \leq i, j \leq n$, $E_{i,j} = 1/n$, y $c \in (0, 1)$. Llamamos a c coeficiente de teletransportación.

Lo que nos queda es una matriz P_2 estocástica por columnas y $\forall 1 \leq i, j \leq n$, $(P_2)_{i,j} > 0$.

Una vez que tenemos la Matriz de Transición, generaremos el puntaje para cada página buscando el autovector w del autovalor 1 de P_2 , tal que $P_2 w = w$ y w sea un vector de probabilidades (normalizado con norma 1).

Es decir, generar el ranking de páginas equivale a aplicar el método de la potencia a la matriz P_2 y una vez hallado el w mencionado, ordenar los puntajes de mayor a menor:

$$ranking = \{p_1, \dots, p_n\}, \text{ donde } \forall i = 1, \dots, n-1, w_{p_i} \geq w_{p_{i+1}} \quad (2)$$

2.1.2. PageRank con matriz Esparsa

2.2. Rankings en competencias deportivas

Elegir un sistema de puntos que sea justo para todos los participantes en un deporte no es una tarea sencilla. Existen muchos factores que afectan el resultado de una competencia como lo pueden ser el orden en el que deben competir entre si los equipos, generando desbalances respecto las capacidades de cada uno. Es por esto que a continuación presentaremos el modelo GeM² que busca modelar los resultados de forma tal que estos factores impacten lo menos posible en el posicionamiento final de la tabla de puntajes.

Con el fin de experimentar con distintos modelos, a lo largo del informe trabajaremos sobre los resultados del Torneo de Primera División 2015³, donde utilizaremos el sistema de ranking estándar de la AFA como punto de comparación.

2.2.1. Generalized Markov chains Method (GeM)

Definición del método

El método GeM es el resultado de tomar el algoritmo PageRank y mediante pequeñas modificaciones utilizar su potencial para establecer un ranking de equipos. Análogo a PageRank, los equipos pasan a formar parte de un grafo dirigido con pesos, donde cada nodo representa un equipo y los pesos de cada arista reflejan el resultado de los partidos jugados entre los vértices conectados.

Formalmente, el modelado se realiza de la siguiente manera:

1. Representamos el torneo como un grafo con pesos dirigidos de n nodos, donde n es igual a la cantidad de equipos que participan. Cada equipo tiene su respectivo nodo y las aristas contienen como peso la diferencia positiva entre los nodos conectados.
2. Definimos la matriz de adyacencia $A \in \mathbb{R}^{n \times n}$.

$$A_{ij} = \begin{cases} w_{ij} & \text{si el equipo } i \text{ perdió contra } j \\ 0 & \text{caso contrario} \end{cases}$$

Donde w_{ij} es la suma total de diferencia positiva de puntaje sobre todos los partidos en los que i perdió contra j .

²Angela Y. Govan, Carl D. Meyer, and Rusell Albright. Generalizing google's pagerank to rank national football league teams. In Proceedings of SAS Global Forum 2008, 2008.

³Campeonato de Primera División 2015, Julio H. Grondona
<http://www.afa.org.ar/html/9/estadisticas-de-primera-division>

3. Definimos la matriz $H \in \mathbb{R}^{n \times n}$.

$$H_{ij} = \begin{cases} A_{ij} / \sum_{k=1}^n A_{ik} & \text{si hay un link de } i \text{ a } j \\ 0 & \text{caso contrario} \end{cases}$$

4. Definimos la matriz GeM, $G \in \mathbb{R}^{n \times n}$ con $u, v, a, e \in \mathbb{R}^n$ y $c \in \mathbb{R}$.

$$\begin{aligned} G &= c(H + au^t) + (1 - c)ev^t \\ \sum_{k=1}^n v_k &= 1 \quad \sum_{k=1}^n u_k = 1 \quad \forall_{i=1..n} e_i = 1 \\ 0 \leq c \leq 1 \quad a_i &= \begin{cases} 1 & \text{si la fila } i \text{ de } H \text{ es un vector nulo} \\ 0 & \text{caso contrario} \end{cases} \end{aligned}$$

5. Por último tenemos que el ranking de los equipos estará definido por el vector $\pi \in \mathbb{R}^n$ tal que

$$\pi^t = \pi^t G$$

o si tomamos la transpuesta en ambos lados

$$G^t \pi = \pi$$

De esta forma al igual que con PageRank, calculando el autovector π obtenemos nuestro ranking. Este modelo permite cierta flexibilidad a partir del u , v y c que tomemos.

El vector de probabilidad u se aplicará en el caso de que un equipo se encuentre invicto, esto es el equivalente a que en PageRank un sitio no tenga ningún link saliente, por lo tanto su tratamiento es el mismo, se le asigna a la fila correspondiente el vector con las probabilidades de saltar a otro nodo. Por lo tanto, el vector u nos permite definir con qué probabilidades un equipo invicto perdería contra el resto de los participantes. En el caso de PageRank, este es un vector de distribución uniforme, donde es igual la posibilidad de saltar a cualquiera de los otros nodos, una posible alternativa sería definirlo como un vector cuyas probabilidades se basen en algún ranking anterior.

El vector de probabilidad v , nos da otro tipo de personalización que es la del *navegante aleatorio*. Esta es la probabilidad de que un equipo cualquiera independientemente de los resultados registrados, pierda contra el resto de los equipos. En PageRank, esto lo veíamos como la posibilidad de que estando navegando el grafo, uno se *teletransportará* a otro nodo independientemente de las conexiones de los mismos. Este vector por defecto también suele tomar el valor de la distribución uniforme.

Por último tenemos nuestro valor c que actúa como un factor de amortiguación donde lo que se modifica es cuánto afecta el *navegante aleatorio* al resultado final, donde con $c = 0$, únicamente influye el *navegante aleatorio* y con $c = 1$ se elimina el efecto del mismo.

Modelado del empate

Una particularidad de este sistema que se puede observar en la definición del mismo es que no contempla los partidos donde hubo empate. Un empate equivale a que no exista un perdedor y por ende no se modifica el peso de ningún nodo. Para deportes donde el empate no es algo frecuente esto no sería un problema, pero si tomamos como ejemplo el fútbol, donde los empates son algo mucho más común, el ignorar estos partidos afecta notablemente el ranking.

3. Demostración: Eliminación Gaussiana sin pivoteo

4. Implementación

5. Experimentación

6. Conclusión

En este trabajo pudimos no solo modelar el sistema planteado, sino que apreciar y aprovechar las propiedades del mismo para así resolverlo con los métodos estudiados observando también las características de ellos.

Por un lado mediante la forma en la que construimos nuestro sistema probamos que se podía resolver con Eliminación Gaussiana sin pivoteo. Además produjimos una versión mejorada del algoritmo de eliminación donde aprovechando la propiedad de banda de la matriz del sistema, redujimos drásticamente la cantidad de operaciones necesarias para resolverla.

Así mismo, cabe destacar que al realizar operaciones con aritmética finita, tanto para la solución de los sistemas como para el cálculo de la isoterma donde la reutilización de datos arrastra error, no podemos garantizar que los resultados obtenidos sean exactos, pero dado que realizamos varias instancias de prueba con distintas metodologías y tomando números de condición aceptables, pudimos ver que los valores que obtuvimos eran coherentes a su contexto.

Luego, en lo que respecta el cálculo de la isoterma, al plantear diversas metodologías tuvimos la posibilidad de analizar y discutir los resultados de las mismas, donde en particular pudimos observar cómo al utilizar la búsqueda binaria podíamos llegar al grado de precisión que deseásemos y que para el método por promedio, al aumentar la cantidad de particiones mejoraba la aproximación, mientras que usando la regresión lineal, esta se ajustaba más a una función lineal que no reflejaba el comportamiento de la fórmula de calor, convergiendo así a un valor distinto tanto al del promedio como el de la búsqueda binaria.

Mediante estas aproximaciones, habiendo establecido previamente nuestro criterio para evaluar si una estructura se encontraba en peligro, llegamos a estimar qué sistemas eran seguros dentro de lo estipulado.

Para el análisis del tiempo de ejecución de una así como varias instancias del sistema modelado, vimos cómo se cumplían las complejidades teóricas de la resolución a través de Eliminación Gaussiana y LU. En este análisis corroboramos cómo si se trataba de una sola instancia la Eliminación Gaussiana presentaba una ventaja sobre LU, dado que el último debe calcular su factorización en su primer corrida, mientras que al subir el número de instancias el algoritmo para LU lograba un tiempo sumamente mejor que el de Eliminación Gaussiana, ya que con la factorización LU habiendo pagado un costo cúbico en la primer instancia, luego es del orden cuadrático contra el siempre cúbico de la Eliminación Gaussiana. Además en el análisis para el algoritmo de Eliminación Gaussiana con la optimización de banda llegamos a concluir que su tiempo de ejecución llegaba a reducirse al de orden cuadrático.

Por último, podemos mencionar algunos experimentos que podrían realizarse a futuro, como el aprovechamiento de la matriz banda en lo que es el algoritmo para la factorización LU, ya que esta optimización se realizó sólo para la Eliminación Gaussiana, junto a su correspondiente estudio de tiempo de ejecución. A su vez, quedó pendiente el realizar la mejora no únicamente en lo que son los tiempos de ejecución sino que el espacio que consume nuestro algoritmo dado que en la matriz banda gran parte de la misma permanece inalterada. También se podría haber profundizado en la experimentación del cálculo de la isoterma con sistemas donde la temperatura interna y externa no fueran constantes si no que tuvieran algún tipo de fluctuación donde se pudiera ver con más detalle cómo se comportaba cada método.