Computación Científica II
Laboratorio 1

Paola Arce - Raquel Pezoa
Profesoras

Hernán Sarmiento - Camilo Zambrano
Ayudantes de Laboratorio

Universidad Técnica Federico Santa María
Departamento de Informática

Santiago, 4 de septiembre de 2012

1. Reglas del Juego

- Los laboratorios se desarrollarán individualmente.
- El laboratorio está conformado por dos entregables: Los códigos y un informe.

1.1. Informe

El informe debe contener la siguiente estructura:

- Introducción
- Objetivos
- Desarrollo (respuesta a las preguntas planteadas)
- Conclusiones
- Anexos

Además, deben tener presentes las siguientes consideraciones:

- Serán evaluadas la ortografía y redacción del informe.
- En los anexos deben ir los casos de prueba que se utilizaron, especificando claramente input y resultados (en caso de que el ejercicio lo requiera).
- El informe debe estar elaborado en LaTeX, de forma **obligatoria**. Asimismo, se tendrá en consideración el correcto uso del formato (en especial ecuaciones y otras fórmulas matemáticas).

1.2. Código

- El código implementado puede ser realizado en MATLAB, Octave o Python.
- Se evaluará el orden (indentación y claridad) y la documentación del código.
- No se permite el uso parcial o total de códigos encontrados en internet o en libros.
- Debe respetarse el input solicitado en el ejercicio, en caso contrario el ejercicio no se revisará y será evaluado con nota cero (0).

2. Preguntas

2.1. Superficies Cuadráticas mediante Valores Propios

Una superficie cuadrática corresponde a una ecuación con tres variables cuya expresión es:

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz + 2a_{14}x + 2a_{24}y + 2a_{34}z + 2a_{44} = 0$$

donde el coeficiente de al menos un término de grado dos es distinto de cero.

La ecuación anterior se puede expresar matricialmente de la siguiente forma:

$$\begin{pmatrix} x & y & z \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 2a_{14} & 2a_{24} & 2a_{34} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} + a_{44} = 0$$

O bien:

$$\begin{pmatrix} x & y & z \end{pmatrix} \cdot A \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} + B \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} + C = 0 \tag{1}$$

Donde A corresponde a una matriz cuadrada y diagonalizable. De lo anterior se puede concebir la existencia de una matriz D diagonal semejante con A, tal que $D = Q^{-1} \cdot A \cdot Q$ donde Q es la matriz ortogonal de cambio de bases entre dos bases ortonormales S (además canónica) y S', por lo que se puede escribir:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}_S = Q_{S' \to S} \cdot \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}_{S'}$$

Reemplazando en (1):

$$\begin{pmatrix} x' & y' & z' \end{pmatrix} \cdot Q^t \cdot A \cdot Q \cdot \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + B \cdot Q \cdot \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + C = 0$$

siendo $D = Q^t \cdot A \cdot Q$

De esta forma:

$$\begin{pmatrix} x' & y' & z' \end{pmatrix} \cdot D \cdot \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + B \cdot Q \cdot \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} + C = 0 \tag{2}$$

Esta expresión permite eliminar los términos rectangulares xy, xz y yz. Además la matriz diagonal D está formada por lo valores propios asociados a la matriz A

Si se efectúa una traslación de ejes paralelos conveniente, se obtiene la ecuación canónica de la expresión dada con la cual es posible elaborar la gráfica.

De acuerdo a lo explicado anteriormente, dada una ecuación cuadrática representada por las matrices A, B y C, desarrolle un programa que permita graficar la superficie correspondiente mediante las transformaciones y uso de valores propios explicados en el contexto mencionado, utilizando como base el siguiente **input**:

```
>>eigen_surface(A,B,C)
```

Recuerde utilizar en todo momento el trasfondo matemático explicado anteriormente, es decir, la utilización de valores propios y la respectiva matriz diagonal D

Ejemplo

La ecuación $5x^2 + 6y^2 + 7z^2 + 4xy + 4yz - 1 = 0$ puede ser representada de forma matricial como sigue:

$$A = \begin{pmatrix} 5 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 7 \end{pmatrix} ; B = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix} ; C = -1$$

Realizando el computo de acuerdo al marco teórico presentado, el resultado obtenido mediante el **output** es el siguiente:

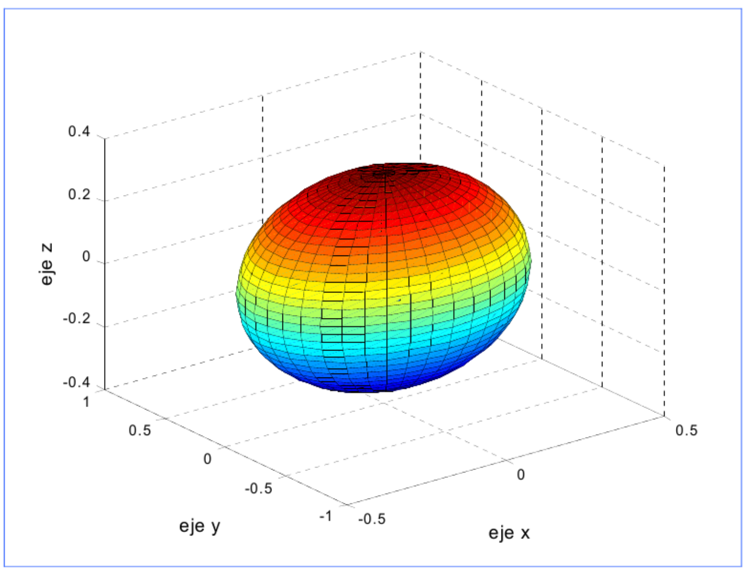

Figura 1: Superficie obtenida mediante valores propios

Para cada ecuación descrita en el correspondiente input y bajo el programa realizado, compare la gráfica obtenida a partir de valores propios versus la gráfica que se genera al confeccionar de manera directa la expresión cuadrática mediante las herramientas o librerías del lenguaje utilizado en su trabajo. Explique y comente sus observaciones.

De igual forma, responda la siguiente pregunta: ¿Qué relación existe entre los valores propios asociados a la matriz A y las superficies generadas a través de las transformaciones?. Explique y comente los resultados obtenidos

Hint: Analice signos y cifras nulas de los respectivos valores propios.

2.2. Ranking de páginas en la Web

Como usted debe saber, la empresa Google se ha hecho famosa en gran medida por que el algoritmo bajo el cual funciona el buscador homónimo, llamado Pagerank, trabaja de una manera distinta a la de los buscadores competidores, diferenciándose principalmente por mostrar en sus resultados "lo mejor" al principio. El buscador Google, ha influenciado fuertemente a las estructuras y al desarrollo de la internet tal como la conocemos, y a que tipo de información y servicios accedemos frecuentemente.
Los motores de búsqueda, como Google, deben básicamente cumplir las siguientes tres tareas:

1. Navegar por la web y ubicar todas aquellas páginas con acceso público.
2. Indexar los datos desde el paso 1, utilizando palabras clave (keywords) o frases relevantes, de modo que se efectúe una búsqueda eficiente.
3. Determinar la importancia de cada página en la base de datos, de modo tal que cuando un usuario busca un sitio, y se encuentra el subconjunto de páginas en la base de datos, las páginas más importantes aparezcan al principio.

En base a lo anterior, ¿cómo es posible determinar la importancia de una página en forma significativa, y cuantificar la importancia de cualquier página dada?

Kurt Bryan y Tanya Leise[1] entregan una idea básica acerca de la forma en que Google determina la importancia de una página usando grafos, matrices, vectores y valores propios. Supongamos estamos interesados en analizar una red de n páginas, cada página indexada bajo el entero k, $1 \leq k \leq n$. Usaremos la notación x_k para representar la importancia de una página k en la red. El valor x_k es no negativo, donde $x_k > x_j$ indica que la importancia de la página k es mayor que la de una página j. Si $x_j = 0$, entonces j tiene el mínimo nivel de importancia posible.

Considere la red como un grafo (ver **fig. 2**), y observe que la importancia de la página 1 depende de las importancias de las páginas 3 y 4 bajo la relación $x_1 = x_3 + x_4$. Ahora, tanto las páginas 3 como 4 tienen como *backlink*(un enlace a una página determinada) a la página 1, entonces la referencia es circular, por lo que se produce un problema. Para solucionarlo, se realiza una modificación en la que se considera que si una página j contiene n_j enlaces y uno de ellos va hacia la página k (es decir, es un backlink de la página k),

[1] http://www.rose-hulman.edu/∼bryan/googleFinalVersionFixed.pdf

entonces la importancia de la página k aumenta en $\frac{x_j}{n_j}$, en vez de sólo x_j.

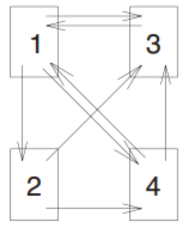

Figura 2: Ejemplo de red compuesta por cuatro páginas enlazadas.

En base a lo anterior, para cuantificar la importancia de una página en base a sus backlinks, considere una red de n páginas y el set $L_k \subset \{1, 2, ...n\}$ de backlinks de una página k. La importancia de una página k está determinada por la siguiente expresión

$$x_k = \sum_{j \in L_k} \frac{x_j}{n_j}$$

donde n_j es el número de links salientes desde la página j. Considere que el link de una página hacia si misma no cuenta como backlink. Considerando el ejemplo de la figura 1, se tendría entonces que $x_1 = x_3 + \frac{x_4}{2}$, $x_2 = \frac{x_1}{3}$, $x_3 = \frac{x_1}{3} + \frac{x_2}{2}$ y $x_4 = \frac{x_1}{3} + \frac{x_2}{2}$. Con ello, estas ecuaciones forman un sistema, que puede ser representado de la forma $\mathbf{Ax} = \mathbf{x}$, donde:

$$x = [x_1, x_2, x_3, x_4]^T, \; A = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix}$$

donde es posible observar que todo se traduce simplemente a un problema de búsqueda de eigenvectores para la matriz cuadrada $\mathbf{A}$ (llamada *matriz de enlaces*), con $\lambda = 1$ y $\mathbf{x} \neq 0$ por definición.
En el caso de ejemplo, se obtiene el eigenvector $[12, 4, 9, 6]^T$ y todos sus múltiplos. Para obtener el ranking, se escala al eigenvector, de modo que la suma de sus componentes den como resultado 1. En este caso se obtiene $x_1 = \frac{12}{31}$, $x_2 = \frac{4}{31}$, $x_3 = \frac{9}{31}$y $x_4 = \frac{6}{31}$.

Finalmente, considere la expresión $V_1(\mathbf{A})$ para denotar el eigenespacio (conjunto de eigenvectores con un eigenvalor en común) para el eigenvalor 1 de una matriz $\mathbf{A}$. Habiendo explicado el contexto, se le solicita lo siguiente:

1. Junto a este laboratorio se ha adjuntado un archivo input llamado `web.txt` que representa a un grafo conexo, el que a su vez representa también a una red de páginas web interconectadas mediante enlaces. El formato del input está creado de modo que cada fila i representa a una página web, y cada valor en la fila representa a un backlink de i. Por ejemplo para el caso de la **figura 1** se tiene el siguiente input:

```
3 4
1
1 2
1 2
```

Donde las páginas 3 y 4 son backlinks para la página 1, la página 1 es un backlink para la página 2 y las páginas 1 y 2 son backlinks para las páginas 3 y 4 respectivamente. Su tarea es construir un programa que genere la matriz de enlaces asociada a la red y a partir de ella obtenga el(los) eigenvectores usando la información entregada a lo largo del enunciado. En base a lo leído y a lo obtenido, calcule el ranking y adjunte el output usando el siguiente formato:

```
Pagina <M>: <Mayor puntaje en el ranking>
...
Pagina <i> : <Puntaje intermedio en orden decreciente>
...
Pagina <m> : <Menor puntaje en el ranking>
```

Donde `<M>`,`<i>`,`<m>` corresponden a los índices de las páginas con mayor puntaje, con un puntaje entre el mayor y el menor, y el menor puntaje, respectivamente. Por ejemplo, para el caso mencionado en el enunciado, el output debería ser el siguiente:

```
Pagina 1: 0.387
Pagina 3: 0.290
Pagina 4: 0.194
Pagina 2: 0.129
```

Hint: Un grafo fuertemente conexo es aquel en el cual se puede llegar desde un nodo a cualquier otro a través de un número finito de pasos.

2. Suponga que los dueños de la página 43 se pelean con los dueños de la página 44, y ambos dejan de referenciarse en forma mutua. Realice ese cambio en el input entregado. ¿Qué sucede con la matriz de enlaces?. Obtenga el(los) eigenvectores usando la información entregada a lo largo del enunciado. ¿Se puede calcular el ranking?, si su respuesta es positiva, adjunte el resultado usando el mismo formato utilizado en el **punto 1**. ¿El valor de $V_1(A)$ se mantiene?. ¿Cambia el resultado?. Si es así, ¿Porqué puede ocurrir esto?.

3. Suponga ahora que, luego de la pelea, el dueño de la página 42 se pone de acuerdo con los dueños de la página 44 y ambos se referencian mutuamente, sin embargo, los dueños de la página 43 se dan cuenta de la traición y dejan de referenciar a la página 42, a pesar de que ésta aún sigue siendo un backlink para esa página. Represente este cambio en el input entregado. ¿Qué sucede con la matriz de enlaces?. Obtenga el(los) eigenvectores usando la información entregada a lo largo del enunciado. ¿Se puede calcular el ranking?, si su respuesta es positiva, adjunte el resultado usando el mismo formato utilizado en el **punto 1**. ¿Qué sucede con el valor de $V_1(A)$ con respecto al inicio?. ¿Cambia el resultado?. Si es así, ¿Porqué puede ocurrir esto?.

4. Anote sus conclusiones finales respecto a lo observado.

5. El **teorema de Gershgorin** se aplica tanto a matrices simétricas como no simétricas de tamaño $m \times m$, y dice lo siguiente:

   "Cada eigenvalor de A yace en al menos uno de m discos circulares en el plano complejo con centros a_{ii} y radio $\sum_{j \neq i} |a_{ij}|$. Más aún, si n de estos discos forman un dominio conectado que es disjunto de los otros $m - n$ discos, entonces hay precisamente n eigenvalores de A dentro de este dominio."

Su tarea es generar un programa que obtenga una estimación de los valores propios de la matriz **A** y retorne gráficamente el resultado. Según la estimación realizada, ¿Existen más valores propios aparte de $\lambda = 1$? Adjunte un gráfico que muestre los discos de Gershgorin.

6. Para comprobar si su estimación es correcta, genere un programa que permita obtener los valores propios de la matriz **A** usando el método de Rayleigh, adjunte sus resultados y anote sus conclusiones.

3. Conclusiones

Determinar los efectos del trabajo realizado mediante el estudio de los distintos método utilizados a largo de la experiencia. Es de gran importancia que las hipótesis y el producto de éstas, estén apoyadas de manera concreta a través del análisis de resultados elaborado anteriormente.

4. Sobre la entrega

- El plazo máximo de entrega (del código y del informe impreso) es el **dia 24 de Septiembre del 2012, a las 23:55**. **El informe y el código en versión digital deben ser subidos a la plataforma Moodle**. También se solicita una versión impresa de su informe, la cual debe entregarse en Secretaría de Informática el mismo día de la entrega digital.
- El archivo debe llamarse lab1-nombre-apellido.tar.gz, y en su interior debe contener una carpeta llamada apellido que contenga los archivos .pdf y .tex correspondientes al informe y los archivos correspondientes al código.
- Se sancionará con **15 puntos menos** en la nota del laboratorio por cada día de atraso
- Las copias serán sancionadas con nota cero (0) para todos los grupos involucrados.

5. Evaluación

Item	Puntaje
Superficies Cuadráticas	30 puntos
Ranking de Páginas en la Web	35 puntos
Conclusiones	15 puntos
Redacción y Ortografía	15 puntos
Estructura Informe (Introducción, Objetivos, Conclusiones, Anexos)	5 puntos
Total	100