

## Tarea 1.

**Instrucciones:** Esta tarea debe realizarse en los grupos constituidos en el curso según las instrucciones dadas. La tarea debe ser respondida en forma ordenada y escrita en computador, usando L<sup>A</sup>T<sub>E</sub>X u otro procesador de texto. **No se aceptarán tareas manuscritas. De ocurrir, no serán corregidas. Debe ser subida a la plataforma de tareas a más tardar el viernes 4 de septiembre de 2020 a las 23:59 horas, como un archivo PDF.** Si usan LaTeX, les daremos 2 décimas de bono en la tarea. Durante su trabajo en la tarea deben respetar las normas éticas correspondientes, cualquier colaboración indebida entre los grupos será sancionada según establecen los reglamentos de la Escuela y Universidad, pudiendo incluso implicar la reprobación del curso. Todos los temas de esta tarea pueden ser evaluados en las interrogaciones.

El siguiente enunciado se aplica a las siguientes dos preguntas y requiere un desarrollo computacional usando los códigos Python que se le entregaron.

Un proceso industrial utiliza cinco materias primas que son mezcladas para obtener un producto. El proceso es contaminante y la empresa tiene interés en estudiar el impacto de las materias primas que utiliza. Se ha postulado que si de una cierta materia prima se usa una cantidad  $x$  (en toneladas), entonces se generarán  $\beta x^\alpha$  kilos de  $CO_2$  como parte de todo el proceso. No se han podido estudiar las materias primas por separado y, además, la generación de contaminantes se produce en la interacción con todo el proceso y las otras materias primas. La empresa, eso sí, ha medido su generación de  $CO_2$  en la producción de diferentes lotes de producto. La planilla Excel que se entrega junto a esta tarea contiene los datos medidos por la empresa. Ahí se muestra, para un total de 55 lotes de producción: las emisiones de  $CO_2$  medidas, y las cantidades de cada una de las 5 materias primas utilizadas.

De esta forma, el  $CO_2$  generado es el resultado de la suma de lo atribuible a cada materia prima, el que se podría asumir que sigue el modelo indicado arriba. En otras palabras, si llamamos  $y$  la contaminación y  $x_k$  es la cantidad usada de la materia  $k$ , se puede postular que

$$y = \sum_{k=1}^5 \beta_k x_k^{\alpha_k}$$

donde los  $\alpha_k$  y  $\beta_k$  son los parámetros que tendrán que ser estimados para cada materia prima.

Para hacer lo anterior, la empresa los ha recibido a ustedes, en un trabajo de práctica, para que encuentren una estimación de los coeficientes. También les han sugerido que la estimación se podría hacer mediante “mínimos cuadrados no lineales”.

### Pregunta 1:

Investigue en qué consiste hacer mínimos cuadrados no lineales (“nonlinear least square”, en Inglés). Escriba una breve explicación de qué se trata esa técnica (no olvide especificar las referencias usadas) y escriba el problema de mínimos cuadrados no lineales que tendrá que resolver para esta aplicación.

### Pregunta 2:

Continuando con el mismo problema del enunciado común y la Pregunta 1, modifique los códigos Python que se entregaron para los métodos de Gradiente, Newton y BFGS, para adaptarlos al problema formulado en la Pregunta 1. Lo que tendrá que, fundamentalmente modificar, es la definición de la función objetivo (la función “subrutina” dentro del código). Acá tendrá que usar los datos de la planilla y esos los tendrá que codificar explícitamente en Python o, si prefiere, use paquetes que permiten interactuar con Excel, como por ejemplo `openpyxl`. También necesitará el gradiente y Hessiano de la función objetivo. Cualquier método que

use para obtener esos gradientes y Hessianos o, incluso, calcularlos en forma numérica, es válido en la medida que expliquen con claridad lo que hacen y, si es necesario, den las referencias correspondientes. Muestre en la solución de esta tarea el código modificado y explique en detalle lo que hace.

### Pregunta 3:

Continuando con el problema del enunciado común y con el código modificado, corra sus programas para encontrar la mejor solución que pueda. Compare el funcionamiento del método del Gradiente, del Método de Newton y de la rutina BFGS, en términos de iteraciones, evolución del error con las iteraciones, tiempo de ejecución, etc. El desarrollo de su respuesta debe incluir, al menos: explicación detallada de su modelación e implementación (no olvide citar cualquier referencia que haya usado), el código modificado y gráficos de evolución del error con las iteraciones.

### Pregunta 4:

Muchos computadores tienen, por ejemplo, un procesador Intel i7, típicamente con frecuencia de 2.4 GHz. Busque en Internet la información que le permita determinar cuantos “flops” por segundo puede realizar ese procesador (o uno similar). Con esa información en mano, construya una tabla que indique cuanto tiempo se requiere para una iteración del Método de Newton, que asumiremos necesita ejecutar  $n^3$  flops, una iteración de un método Quasi-Newton, que asumiremos requiere  $n^2$  flops, y una iteración del método de Gradiente, que asumiremos requiere  $n$  flops. En la tabla ponga el tiempo que se requiere para realizar 1.000 iteraciones de cada uno de esos métodos. Haga esta tabla para los siguientes valores de  $n$ ,  $n = 10, 100, 1.000, 10.000, 20.000.000$ . (Un problema de optimización con 16 millones de variables podría ser uno en que las variables son valores de pixeles en una imagen de 16 megapixeles, bastante común en cualquier cámara digital de un teléfono celular moderno). Desde luego estos cálculos están ignorando otras operaciones que también requieren tiempo y el hecho que el procesador no necesariamente estará exclusivamente dedicado a su trabajo, pero se obtendrá, al menos, una idea aproximada de los tiempos necesarios y su dependencia en  $n$  para cada método.

### Pregunta 5:

En biología se ha estudiado desde hace tiempo la forma en que las proteínas se combinan para dar origen a diversos tejidos en un organismo, entre otras cosas. La forma espacial que adquiere una proteína tiene que ver con la constitución de sus aminoácidos, los cuales se atraen o repelen según sus características. En los dos problemas que vienen a continuación, queremos modelar esta situación y visualizar cómo resolver el problema de optimización resultante.

Para esto, vamos a considerar un modelo conceptual de la proteína, el que está formado por una secuencia de elementos rígidos. Dichos elementos están unidos por sus extremos en donde se pueden articular. Los distintos elementos se atraen o repelen según la acción de fuerzas eléctricas y nucleares, de acuerdo a características conocidas. Para simplificar, asumiremos que las fuerzas se aplican en los puntos centrales de los elementos. La figura 1 muestra un modelo simplificado de una proteína en 2 dimensiones.

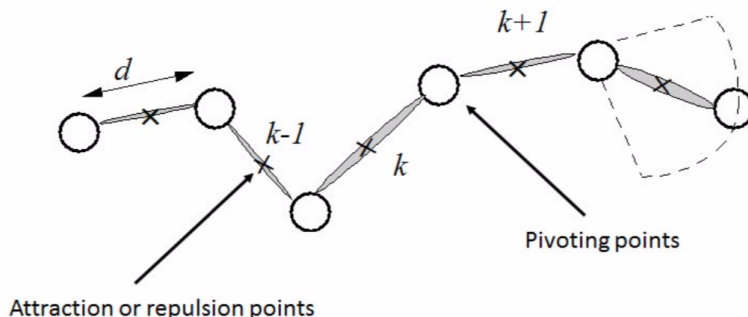


Figura 1: Modelo de una Proteína.

Asumamos que tenemos  $K$  elementos, y que los puntos extremos del elemento  $k$  son  $(x_{k-1}, y_{k-1})$  y  $(x_k, y_k)$ . Asumimos igualmente, también para simplificar, que todos los elementos tienen un largo fijo igual a  $d$ . Luego, una restricción del sistema es que:

$$(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2 = d^2, \quad k = 1, \dots, K.$$

También especificaremos que el punto inicial,  $(x_0, y_0)$  es fijo y conocido.

Para modelar las fuerzas eléctricas y nucleares que producen el doblamiento de la proteína, asumimos que todo el sistema trata de alcanzar un estado de mínima energía. Esta energía está dada por:

$$\sum_{i \neq j} \frac{a_{ij}}{dist_{ij}}$$

donde  $dist_{ij}$  es la distancia entre el elemento  $i$  y el elemento  $j$ , y  $a_{ij}$  es un coeficiente conocido que depende que las características del elemento. Dicha distancia la calculamos entre los puntos centrales de los segmentos. Formule el modelo de optimización que busca minimizar esa energía sujeto a las restricciones del largo de los elementos y la definición de los puntos centrales.

**Pregunta 6:**

Continuando con el problema anterior, el modelo que ustedes escribieron en la Pregunta 5, es un problema de minimización de una función objetivo no lineal, pero sujeto a restricciones de igualdad. Modifique el modelo de modo tal que las restricciones de igualdad puedan ser reemplazadas y de como resultado un modelo de optimización sin restricciones. Una vez hecho eso, analice la función objetivo y discuta las ventajas y/o desventajas de usar el método del gradiente para este problema, o el Método de Newton, un método Quasi-Newton como BFGS o un método de primer orden con paso fijo (el que tiene el paso  $1/L$ , por ejemplo). No necesita desarrollar los métodos, pero queremos que puedan analizar las complejidades de implementación en cada caso.