Práctico de Optimización

Modelado de Redes de Telecomunicaciones

1. **Repaso Minimización sin restricciones de una función cuadrática.**

   Encontrar el gradiente, el Hessiano, los puntos estacionarios e indicar cuales son mínimos de

   $$f(x,y) = x^2 + y^2 + \beta xy + x + 2y$$

   Discutir según el valor de β.

2. **Convergencia del algoritmo del gradiente con paso constante.**

   Se desea aplicar el algoritmo de minimización del gradiente (descenso más rápido) con paso constante a la función $f(x) = \|x\|^{\beta+2}$, $\beta \geq 0$, $x \in \mathbf{R^n}$.

   a)calcular el gradiente de f(x)

   b)Calcular la ecuación de iteración es decir $x_{k+1} = g(x_k)$ para el método del gradiente.

   c) Mostrar que si $\|x_1\| < \|x_0\|$ entonces $\|x_{k+1}\| < \|x_k\|$ y que si $\|x_1\| \geq \|x_0\|$ entonces $\|x_{k+1}\| \geq \|x_k\|$

   d) Encontrar la relación que deben cumplir x_0, β y el paso s para que el algoritmo converja.

   e) Implementar el algoritmo en matlab (implementando la iteración de la parte b)) y verificar que la condición de la parte anterior se cumple tomando un caso en que converge y un caso en que no.

   f) ¿La función $\nabla f(x)$ es Lipschitz? es decir que $\forall x, y$ existe una constante $L > 0$ tal que $\|\nabla f(y) - \nabla f(x)\| \leq L \|y - x\|$

   g) Comparar el resultado de convergencia obtenido en este ejercicio con el resultado de convergencia del método del gradiente para paso constante visto en el teórico.

3. **Convergencia del algoritmo del gradiente utilizando el paso mínimo en cada iteración.**

   Sea $f(x,y) = \frac{1}{2}(x^2 + \gamma y^2)$, $\gamma > 0$. Se desea utilizar el algoritmo del gradiente, usando para calcular el paso en cada iteración el método de minimización de $f(x_k - s_k \nabla(f(x_k)))$.

   a) Aplicar el algoritmo anterior a partir del punto $(\gamma, 1)$ y mostrar que

   $x_k = \gamma \left(\frac{\gamma-1}{\gamma+1}\right)^k$

   $y_k = \left(-\frac{\gamma-1}{\gamma+1}\right)^k$

   $f(x_k, y_k) = \left(\frac{\gamma-1}{\gamma+1}\right)^{2k} f(x_0, y_0)$.

   b) Calcular el Hessiano de la función. ¿Cuál es el número de condicionamiento de esta función? Como se comparan los resultados de la parte a) con las cotas para la velocidad de convergencia vistas en el teórico.

4. **Comparación del algoritmo del gradiente y de Newton de acuerdo al número de condicionamiento.**

   Se considera la siguiente función a minimizar $f(x_1, x_2) = e^{x_1+\gamma x_2-0{,}1} + e^{x_1-\gamma x_2-0{,}1} + e^{-x_1-0{,}1}$.

   a) Programar en Matlab el algoritmo del descenso por el gradiente para la función anterior. Para calcular el paso utilizar el método de backtracking con $\alpha = 0{,}1$ y $\beta = 0{,}7$.

   b) Programar el algoritmo de Newton para la función anterior usando el método de backtracking para el paso con $\alpha = 0{,}1$ y $\beta = 0{,}7$.

   c) Correr los algoritmos anteriores con condición inicial $x_1 = 0$, $x_2 = 0{,}1$ y para , $\gamma = 5$, $\gamma = 10$. Graficar para cada caso las curvas de nivel y la trayectoria de cada algoritmo. ¿Como se comparan las velocidades del algoritmo de Newton y del gradiente para los diferentes valores de γ? ¿como es cualitativamente en cada caso el número de condicionamiento?

5. **Mínimos cuadrados. Regresión**.

   Se desea minimizar la función $f(x) = \|g(x)\|^2 = \frac{1}{2}\sum_{i=1}^{m} \|g_i(x)\|^2$ con $x \in \mathcal{R}^n$, $g_i : \mathcal{R}^n \to \mathcal{R}, g = (g_1(x), .., g_m(x))$.

   a) Calcular el Hessiano $\nabla^2(f(x^*))$ en función de $\nabla^2(g(x^*))$ siendo x^* un punto óptimo (es decir $(g(x^*) = 0)$). Observar que ∇g, como g es un vector de m funciones es una matriz nxm

   ¿Cuando el Hessiano es singular y cuando no lo es? Discutir según m y n. Observar que en una matriz A mxn con $m < n$, su rango es a los sumo m y existen por tanto vectores v no nulos tales que $Av = 0$.

   b) Un problema típico de la formulación anterior es el de regresión lineal donde se toman muestras de la entrada y salida de un sistema, y se quiere encontrar los coeficientes x que minimizen $\|g(x)\|^2$, $g(x) = z - Ax$, con A matriz mxn. Probar que si m¡n existen infinitos mínimos. Se sugiere calcular $f(x^* + \lambda v)$ siendo v un vector que cumple que $Av = 0$, x^* un mínimo y λ un escalar cualquiera. Encontrar una solución si A tiene m filas linealmente independientes.

6. **Balanceo de carga óptimo en una red**.

   Se considera una red donde para cada par de nodos ingreso/egreso $w = 1...W$ tiene una tasa de tráfico que desea ir del nodo de ingreso al de egreso: r_w. Esta tasa se reparte entre un conjunto n_w de posibles caminos entre ese ingreso y egreso y cada camino se nota $C_{w,i}$. Por el camino $C_{w,i}$ se envía una parte del tráfico r_w igual a $x_{w,i}$. Por cada enlace l de la red, pasa un tráfico total $x_l = \sum_{w=1}^{W} \sum_{i:C_{w,i}\in l} x_{w,i}$ donde $C_{w,i} \in l$ significa que el camino i del par ingreso/egreso w utiliza el enlace l. Se desea como objetivo minimizar el retardo total de la red. El retardo total es la suma del retardo en cada enlace. El retardo en cada enlace es una función $f_l(x_l)$ que se supone conocida y que tiene derivadas parciales continuas.

   a) Formular el problema de optimización anterior. Observar que es un problema de minimización de una función sobre un conjunto convexo X.

   b) Utilizando la condición necesaria de optimalidad vista en el teórico $(\nabla(f(x^*))'(x - x^*) \geq 0$, $\forall x \in X)$, probar que es necesario para tener un reparto óptimo que para todos los caminos por donde se envíe tráfico $(x_{w,i} > 0)$ se tenga la misma suma de la derivada de los retardos en el camino y que este valor es menor que la suma de la derivada del retardo en los restantes caminos por los que no se envía tráfico $(x_{w,i} = 0)$.

   c)¿Cuando la condición anterior es además suficiente y permite decir que el reparto obtenido es óptimo global?

   d)Aplicar los resultados anteriores para una red con un solo par ingreso/egreso y dos caminos, cada uno de ellos con un solo enlace. Uno de los enlaces tiene capacidad C_1 y el otro C_2 con $C_1 < C_2$. La función del retardo en cada enlace se asumirá que corresponde al de un enlace M/M/1. Graficar los rates óptimos por cada camino en función del valor del tráfico ingreso/egreso.

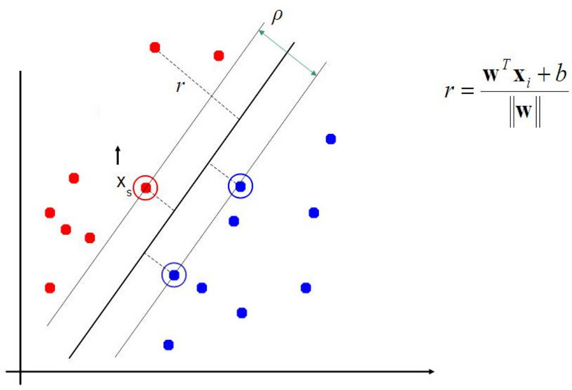

Figura 1: svm

7. **Balanceo de carga óptimo. Métodos de proyección del gradiente y direcciones factibles.Problema dual**.

   Para el problema de balanceo de carga óptimo visto antes encontrar:

   a)Un algoritmo iterativo para resolverlo y que utilice el método del gradiente condicional.

   b)Un algoritmo iterativo para resolverlo y que utilice el método de proyección del gradiente.

   c)Formular el problema Dual y encontrar la relación que cumple el multiplicador de Lagrange de cada restricción y cual es la interpretación de dicho multiplicador.

8. **Support Vector Machines para clasificación**.

   Se supone que se tiene una muestra datos ($x_i \in R^n, i = 1, .., n$) (por ejemplo n características del tráfico de un flujo en la red: puerto origen y destino, direcciones, tamaños de los k primeros paquetes, etc.). Cada uno de estos datos se puede clasificar en dos clases ($y_i = 1$ o -1) (si el flujo es un p2p o no por ej.).

   Se supone que los datos son totalmente separables por un hiperplano (ver figura). Se puede generalizar al caso no totalmente separable. En SVM se busca el hiperplano ($w'x + b = 0$) tal que la separación ρ indicada en la figura sea máxima. Observar que todos los puntos (x_i, y_i) deben verificar la siguiente condición: $\frac{y_i(x_i w + b)}{\|w\|} \geq \frac{\rho}{2}$.

   a)Formular el problema de maximización anterior. Transformarlo en un problema de minimización cuadrática ($\min \|w\|^2$) observando que $\rho = \frac{2}{\|w\|}$.

   b)Formular el problema dual. Resolviendo el problema dual, encontrar la ecuación del hiperplano óptimo (ecuación de w y b) en función de los multiplicadores de Lagrange y verificar que solo depende de los puntos x_i ubicados sobre los hiperplanos paralelos al hiperplano óptimo a una distancia $\frac{\rho}{2}$. Definir la función que permite clasificar un nuevo punto x. Escribir el Lagrangiano en función de los multiplicadores de Lagrange y de los datos de entrenamiento. ¿cómo calcularía los multiplicadores de Lagrange?.

9. **Sensibilidad**.

   El teorema de sensibilidad brinda una interpretación de los multiplicadores de Lagrange. Estos se pueden interpretar como la variación del costo óptimo con respecto a una variación de la restricción asociada a dicho multiplicador. En este ejercicio se busca aplicar dicho teorema a dos ejemplos simples. Se tiene un enlace de capacidad C y un usuario con una función utilidad $U(x)$

donde x es el bitrate que envía el usuario a la red. El problema es maximizar $U(x)$ sujeto a que $x \leq C$ para los siguientes casos:

a) $U(x) = 1 - e^{-x}$, $x \geq 0$

b) $U(x) = -x^2 + kx$, $x \geq 0$

Para ambos casos calcular el óptimo, calcular el valor del multiplicador de Lagrange e interpretar del valor del multiplicador comparando ambos casos según el teorema de sensibilidad. En el segundo caso discutir según k.

10. **Planificación de la capacidad de una red**.

    Un operador de una red va desplegar un enlace de interconexión internacional. Este enlace es vendido a N usuarios y cada uno tiene una función utilidad $U_i(x_i)$ siendo x_i la tasa de bits y U_i cuanto está dispuesto a pagar el usuario i por esa tasa de bits en dicho enlace, que asumimos es una función cóncava. El operador tiene un costo asociado a instalar una cierta capacidad z que es una función $V(z)$ que asumimos es lineal con z o más en general convexa con z. Un regulador fija los precios de este mercado a los efectos de que se maximice el bienestar social es decir: $\sum_{i=1}^{N} U_i(x_i) - V(z)$ sujeto a que $z = \sum_{i=1}^{N} x_i$.

    a) Formular el lagrangiano del problema anterior. Observar que cuando el regulador fija un precio p el problema de maximización del Lagrangiano es equivalente a que cada usuario optimice por separado $U_i(x_i) - px_i$, es decir su ganancia y el operador maximice su propia ganancia es decir $V(z) - pz$.

    b) Encontrar la relación que deben cumplir el precio óptimo la función Utilidad y la función de costo.

    c)Proponga un algoritmo iterativo en el cual el regulador modificando los precios iterativamente llegue al precio óptimo.

11. **Asignación de recursos de una red**.

    En una red con enlaces de capacidad c_l se tiene N usuarios cada uno con una función utilidad $U_i(x_i)$. En un cierto momento cada usuario tiene una ruta que lo conecta con el destino con el que se desea comunicar. La matriz de Ruteo R tiene un 1 en el elemento (i, j) si la ruta del usuario i utiliza el enlace j y 0 en caso contrario. El problema de asignación de recursos en una red en un determinado momento se puede formular como: $\max \sum U_i(x_i)$ sujeto a que $RX \leq C$ siendo C el vector con las capacidades de todos los enlaces, X el vector de las x_i. Escribir un algoritmo distribuido que realice esta optimización. Especificar que información se debe intercambiar entre los usuarios y los enlaces.

12. **Asignación de recursos y justicia** En el teórico se vio que una asignación proporcionalmente justa x_i^* era aquella que cumplía que para toda otra asignación factible x_i se cumple que $\sum_i \frac{x_i - x_i^*}{x_i^*} \leq 0$ es decir, que ninguna otra asignación factible incrementa la suma de las tasas de cambios proporcionales.

    a) Verificar que si en el ejercicio anterior se utiliza $U_i(x_i) = log(x_i)$ la asignación de recursos en la red es proporcionalmente justa. Para esto se sugiere utilizar la condición de optimalidad vista para minimizar funciones sobre conjuntos convexos: $\nabla' f(x^*)(x - x^*) \geq 0$.

    b) Se considera una red con dos enlaces y tres flujos. El primer enlace con capacidad C1=2 y el segundo C2=3. De los tres flujos, el flujo 1 pasa por los dos enlaces, el flujo 2 solo por el primer enlace y el flujo 3 solo por el segundo enlace.

    b.1)Encontrar la asignación que maximiza el bit rate total en la red. ¿a qué función utilidad U_i corresponde?

    b.2) Encontrar la asignación óptima proporcionalmente justa. Comparar ambas asignaciones b.3) Una asignación de justicia maxmin lo que hace es $\max \min(x_i)$ sujeto a las restricciones correspondientes. En el teórico se vio el algoritmo de llenado progresivo para realizar una asignación maxmin justa. Calcular la asignación maxmin para este problema. Comparar las tres asignaciones encontradas.

13. **El throughput de TCP**. Mediante simulaciones en Ns-2 verificar que como se vio en el teórico el throughput de conexiones largas en TCP es inversamente proporcional al RTT y a la raíz cuadrada de la pérdidas.

14. **Asignación de potencia en sistemas OFDM**

    En un sistema con múltiples portadoras como OFDM, sea N el número de sub-canales. En una versión simplificada de dicho sistema se asumirá que cada sub-canal tiene un nivel de potencia de ruido $n_1, .., n_N$, una ganancia de potencia en el sub-canal $h_1, ..h_N$ y sea $p_1, .., p_N$ la potencia asignada a cada sub-canal. Si se asume para cada sub-canal un modelo de canal aditivo con ruido gaussiano (AWGN), la capacidad (bit rate) del subcanal i es igual a :

    $C_i = B \log(1 + \frac{p_i h_i}{n_i})$

    donde B es el ancho de banda del sub-canal. Consideraremos B=1 ya que no cambia el problema de optimización y simplifica la notación.

    a) El objetivo es maximizar la capacidad total del canal sujeto a que la suma de la potencia en todos los sub-canales está acotada por la potencia máxima P de la estación. Formular el problema de optimización y probar usando las condiciones KKT que la asignación óptima verifica que para todo canal con $p_i^* > 0$, se cumple que $\frac{n_i^*}{h_i^*} + p_i^* = \frac{1}{\lambda_0^*}$ siendo λ_0^* el precio (multiplicador de la Lagrange) asociado a la restricción de potencia. Interpretar el resultado.

    b) Ejemplo: si se tienen 4 canales con n=(1,1,1,1), h= (1/4,1/8,3/8,1/2) y la potencia total es 20, todos ellos en unidades adecuadas. Calcular el vector de asignaciones óptimas de potencia.