Universidad de la República - Facultad de Ingeniería

Curso 2023 : Fundamentos de Optimización

Grupo 133

Guido Dinello
5.031.022-5

Mathías Ramilo
5.665.788-5

Tarea Final

Índice

1. Verifique que la solucion no es unica.

Sugerencia: dada una solucion X, busque soluciones de la forma XC, con C matriz de tamaño y características a definir.

1.1. Enfoque

Considerando soluciones de la forma XC siendo X solucion. Para que el producto matricial este bien definido entre las matrices para la expresion X' = XC donde tanto X' como X son solucion al problema y por ende poseen dimensiones nxd entonces C debe tener dimensiones dxd.
Por otro lado, podemos realizar la sustitucion en la expresion original pbteniendo:

$$A - XX^t = A - XC(XC)^t = A - XCC^tX^t$$

Observemos que si $CC^t = Id$ entonces obtenemos la ecuacion inicial, por lo que si X minimza $||A - XX^t||_F^2$ se sigue que XC tambien lo hara.
Por ende, C debe ser una matriz ortogonal.

1.2. Generación de la matriz aleatoria C

Se genera una matriz aleatoria C de tamaño $d \times d$ utilizando la función `rand` del paquete `numpy.random`.

1.3. Descomposición QR

Se realiza la descomposición QR de la matriz C, donde Q es una matriz ortogonal y R es una matriz triangular superior. Esto se logra utilizando la función `np.linalg.qr`.

1.4. Generación de la matriz combinada XC

Obtenemos XC de tamaño $n \times d$ multiplicando la matriz solucion dada X_{svd} y la matriz Q obtenida en la descomposición QR. Es decir $XC = X_{\text{svd}}Q$.

1.5. Cálculo del valor funcional

Se calcula el valor funcional utilizando la función $f(X)$ definida en el problema.

$$f(X) = ||A - XX^t||_F^2$$

Resultados empricos obtenidos.

$$f(XC) \approx 8045{,}3111$$

$$f(X_{\text{svd}}) \approx 8045{,}3111$$

1.5.1. Verificación de la no unicidad

Como podemos observar anteriormente tanto XC como X_{svd} son soluciones validas para el problema planteado para cualquier C matriz ortogonal de tamaño dxd.

2. Compruebe numericamente que la funcion objetivo no es convexa. Sugerencia, tome dos soluciones, basandose en el item anterior.

2.1. Definición de función convexa

Una función $f : \mathbb{R}^n \to \mathbb{R}$ se dice que es convexa si para cualquier par de puntos $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ y cualquier $\alpha \in [0, 1]$, se cumple la siguiente desigualdad:

$$f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1 - \alpha)f(\mathbf{y})$$

En otras palabras, una función es convexa si el segmento de línea entre dos puntos en su gráfica siempre se encuentra por encima o en el mismo nivel que la función en esos puntos.

2.2. Comprobacion numerica

Utilizando las 2 soluciones obtenidas en el inciso anterior y tomando $\alpha = 0{,}5$ verificamos que no se cumple la desigualdad.

$$f(\frac{1}{2}X_{\text{svd}} + (1 - \frac{1}{2})XC) \leq \frac{1}{2}f(X_{\text{svd}}) + (1 - \frac{1}{2})f(XC)$$
$$11612{,}7977 \leq 8045{,}3111$$

Por lo que podemos afirmar que la funcion no es convexa.

3. Calcule el gradiente de la función objetivo, y verifique numéricamente el resultado.

3.1. Calculo del gradiente

Sea $f(X) = ||A - XX^T||^2$
Siguiendo la defincion de la norma Frobenius: $||A||_F = \sqrt{tr(AA^H)}$
Podemos reescribir a $f(X)$ como:

$$((tr([A - XX^T][A - XX^T]^T))^(1/2))^2 =$$
$$tr([A - XX^T][A^T - (XX^T)^T]) =$$
$$tr([A - XX^T][A^T - (XX^T)]) =$$
$$tr(AA^T - AXX^T - XX^TA^T + XX^TXX^T)$$

Usando la propiedad de la traza $tr(A + B) = tr(A) + tr(B)$

$$tr(AA^T) - tr(AXX^T) - tr(XX^TA^T) + tr([XX^T]^2)$$

Utilizando la propiedad de la traza $tr(ABC) = tr(BCA) + tr(CAB)y/otr(AB) = tr(BA)$

$$tr(AA^T) - tr(AXX^T) - tr(A^TXX^T) + tr((XX^TX)X^T) =$$
$$tr(AA^T) - tr(AXX^T) - tr(A^TXX^T) + tr(X^TXX^TX)$$

Derivando respecto de X y utilizando la linealidad del operador gradiente en conjunto con las siguientes igualdades[1]:

$$\frac{\partial}{\partial X}[\mathrm{Tr}[X^T X X^T X]] = 4XX^T X \qquad (123\ B = C = Id)$$
$$\frac{\partial}{\partial X}[\mathrm{Tr}(X^T X B)] = XB^T + XB \qquad (113)$$
$$\frac{\partial}{\partial X}[\mathrm{Tr}(B X X^T)] = BX + B^T X \qquad (109)$$

Tenemos que:

$$\frac{\partial}{\partial X}[tr(AA^T)] = 0$$
$$\frac{\partial}{\partial X}[tr(AXX^T)] = AX + A^T X$$
$$\frac{\partial}{\partial X}[tr(A^T XX^T)] = A^T X + A^{TT} X = A^T X + AX$$
$$\frac{\partial}{\partial X}[tr(X^T XX^T X)] = 4XX^T X$$

$$\frac{\partial}{\partial X} f(X) = -AX - ATX - ATX - AX + 4XXTX = \\ -2(AX + ATX) + 4XXTX$$

Finalmente, dado que A es simétrica ($A^T = A$):

$$\frac{\partial}{\partial X} f(X) = -4AX + 4XXTX = \mathbf{4(-A + XXT)X}$$

3.2. Verificación numérica del resultado

Utilizando la implementación 1 para estimar el gradiente numérico obtuvimos un valor para la norma de la diferencia con el gradiente hallado analíticamente de 0,0003884.

Algorithm 1 Numerical approximation of the gradient using incremental ratios

function GRAD_NUMERICO(X)
  eps $\leftarrow 1e - 7$
  grad $\leftarrow$ zeros(X.shape)
  for $row \leftarrow 0$ to $X.rows - 1$ **do**
    for $col \leftarrow 0$ to $X.cols - 1$ **do**
      $h \leftarrow$ zeros(X.shape)
      $h[row, col] \leftarrow$ eps
      grad$[row, col] \leftarrow \frac{f(X+h)-f(X)}{\text{eps}}$
    end for
  end for
  return grad
end function

[1] Extraídas del matrix cookbook https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

4. Implemente un método de descenso por gradiente para encontrar un punto crítico de la función. Compare la solución y su valor funcional con la solución dada por la descomposición espectral de A.

4.1. Implementación

Implementaremos el método de descenso por gradiente acelerado.

Algorithm 2 Nesterov Gradient Descent

function NGD(grad, x_{init}, α, tol $= 1e-5$)
  $x \leftarrow x_{\text{init}}$
  $y \leftarrow x_{\text{init}}$
  iter $\leftarrow 0$
  while $|\text{grad}(y)| >$ tol **do**
    $x_{\text{next}} \leftarrow y - \alpha \cdot \text{grad}(y)$
    $y_{\text{next}} \leftarrow x_{\text{next}} + \frac{\text{iter}}{\text{iter}+3} \cdot (x_{\text{next}} - x)$
    $x, y \leftarrow x_{\text{next}}, y_{\text{next}}$
    iter $\leftarrow$ iter $+ 1$
  end while
  return x
end function

4.2. Comparación de Resultados

El método de Descenso por Gradiente a partir del X inicial proporcionado (elegido aleatoriamente) y utilizando un $\alpha = 1e-3$ convergió en 293 a una "solución"que dista (ver referencia 1). Por otro lado, la diferencia entre los valores funcionales es de (ver referencia 2).

$$||X_svd - X_gd|| = 18{,}3445 \tag{1}$$

$$||f(X_svd) - f(X_gd)|| = 5{,}4570e - 12 \tag{2}$$