Tarea 7

Optimizacion

César Amilkar Rivera Covarrubias

3 de mayo de 2025

1. Dada $\mu \geq 0$, definimos X_μ como el conjunto de funciones $f \in C^2(\mathbb{R}^n)$ que satisfacen

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2}\|y-x\|^2.$$

   a) Muestra que las funciones en X_0 son convexas.
   Solución: Sea $f \in X_0$, dados $x, y \in \mathbb{R}^n$ por hipótesis se cumple que

$$f(y) \geq f(x) + \nabla f(x)^T (y-x). \tag{0.1}$$ `{eq: def_X}`

   Sabemos que una función g es convexa si para todo $\lambda \in [0,1]$ se cumple que

$$g(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y).$$

   Dada $x, y \in \mathbb{R}^n$ fijos, parametricemos el segmento de linea que une ambos puntos como $\gamma(t) = xt + (1-t)y$ para $t \in [0,1]$, notemos que

$$x - \gamma(t) = (1-t)(x-y) \qquad \text{y} \qquad y - \gamma(t) = t(y-x),$$

   de (0.1) se sigue que

$$f(x) \geq f(\gamma(t)) + \nabla f(\gamma(t))^T (x - \gamma(t)),$$

   además

$$f(y) \geq f(\gamma(t)) + \nabla f(\gamma(t))^T (y - \gamma(t)).$$

   Multipliquemos la primer desigualdad por t, obteniendo que

$$f(x)t \geq f(\gamma(t))t + \nabla f(\gamma(t))^T (x - \gamma(t))t$$

   a la segunda por $(1-t)$

$$f(y)(1-t) \geq f(\gamma(t))(1-t) + \nabla f(\gamma(t))^T (y - \gamma(t))(1-t),$$

   tengamos en cuenta que

$$(1-t)(y-\gamma(t)) + (1-t)(x - \gamma(t)) = t(1-t)(y-x) + t(1-t)(x-y) = 0,$$

   por otro lado,

$$\begin{aligned}
&f(\gamma(t))(1-t) + \nabla f(\gamma(t))^T (y-\gamma(t))(1-t) + f(\gamma(t))t + \nabla f(\gamma(t))^T (x-\gamma(t))t \\
&= f(\gamma(t)) + \nabla f(\gamma(t))^T (t(x-\gamma(t)) + (1-t)(y - \gamma(t))),
\end{aligned}$$

   por lo anterior, al sumar las desigualdad anteriores obtenemos que

$$f(x)t + f(y)(1-t) \geq f(\gamma(t)) = f(xt + (1-t)y),$$

   Concluyendo así que la función f es convexa.

b) Muestra que $f \in X_\mu$ si y solo si la función g definida por

$$g(x) = f(x) - \frac{\mu}{2}\|x\|^2$$

es convexa.

Solución: Supongamos que $f \in X_\mu$, definamos

$$g(x) = f(x) - \frac{\mu}{2}\|x\|^2,$$

por la definición de X_μ, para $y \in \mathbb{R}^n$ tenemos que

$$\begin{aligned} g(y) &\geq f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|^2 - \frac{\mu}{2}\|y\|^2 \\ &\geq f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}(-2y^tx + \|x\|^2) \\ &\geq \left(f(x) - \frac{\mu}{2}\|x\|^2\right) + \nabla f(x)^t(y-x) - \mu y^tx + \mu x^tx, \end{aligned}$$

la desigualdad anterior es valida ya que

$$\|y-x\|^2 - \|y\|^2 = \|y\|^2 - 2y^tx + \|x\|^2 - \|y\|^2 = -2y^tx + \|x\|^2.$$

Tengamos en cuenta que

$$\nabla g(x) = \nabla f(x) - \mu x,$$

luego, de la desigualdad de anterior se sigue que

$$g(y) \geq g(x) + \nabla g(x)^t(y-x).$$

Concluyendo que g es convexa.

Por otro lado, supongamos que g, definida como antes, es convexa, entonces, para todo $x, y \in \mathbb{R}^n$ tenemos que

$$g(y) \geq g(x) + \nabla g(x)^t(y-x),$$

al sustituir g y el gradiente de g tenemos que

$$\begin{aligned} f(y) - \frac{\mu}{2}\|y\|^2 &\geq f(x) - \frac{\mu}{2}\|x\|^2 + (\nabla f(x) - \mu x)(y-x) \\ f(y) &\geq f(x) + \nabla f(x)^t(y-x) + \frac{\mu}{2}\left(\|y\|^2 - \|x\|^2 - 2x^t(y-x)\right) \\ f(y) &\geq f(x) + \nabla f(x)^t(y-x) + \frac{\mu}{2}\|y-x\|^2. \end{aligned}$$

Lo anterior se sigue de que

$$\|y\|^2 - \|x\|^2 - 2x^t(y-x) = \|y\|^2 - \|x\|^2 - 2x^ty + 2\|x\|^2 = \|y-x\|$$

Satisfaciendo la desigualdad de la definición de X_μ, concluyendo que $f \in X_\mu$.

c) De 2, deduce la propiedad de los eigenvalores de la Hessiana de las funciones en X_μ.

Solución: Anteriormente calculamos el gradiente de g, que está dado por

$$\nabla g(x) = \nabla f(x) - \frac{\mu}{2}x,$$

la Hessiana de está función es

$$\nabla^2 g(x) = \nabla^2 f(x) - \frac{\mu}{2}I.$$

Ya que g es convexa, entonces se cumple que su Hessiana es semi positiva definida. Sea $x \in \mathbb{R}^n$, y λ un eigenvalor de la Hessiana de f, con eigenvector v de norma unitaria, de la definición de eigenvalor, tenemos que

$$v^t \nabla^2 g(x) v = v^t(\nabla^2 f(x) - \frac{\mu}{2})v = \|v\|^2(\lambda - \mu),$$

de la definición de semipositiva definida obtenemos que la expresión anterior es mayor o igual a 0, ya que $\|v\|^2 = 1$, podemos concluir que

$$\lambda - \mu \geq 0 \Rightarrow \lambda \geq \mu.$$

Obteniendo así que los eigenvalorea de la Hessiana de f son mayores o iguales a μ.

d) Da un ejemplo de una función estrictamente convexa que no pertenezca a ningun conjunto X_μ, para $\mu > 0$.

Solución: Consideremos la función $f(x) = e^x$, tengamos en cuenta que su primer y segunda derivada coinciden con la función. Ya que la segunda derivada es no negativa para todo $x \in \mathbb{R}^n$ podemos concluir que la función es estrictamente convexa.
Para que $f(x)$ perteneza a algún X_μ, debe existir μ tal que para todo $x, y \in \mathbb{R}^n$,

$$e^y \geq e^x + e^x(y - x) + \frac{\mu}{2}(y - x)^2,$$

Consideremos $x = 0$, evaluando en la expresión anterior tenemos que

$$e^y \geq 1 + y + \frac{\mu}{2}y^2,$$

Notemos que si tomamos y suficientemente pequeño, entonces e^y se aproxima a 0, mientras que $1 + y + \frac{\mu}{2}y^2$ se aproxima a infinito, lo cual nos lleva a una contradicción, de que exista un $\mu > 0$ que cumpla esta condición.
Concluyendo así que $f(x) = e^x$ no pertenece a ningun X_μ para $\mu > 0$.

e) Muestra que cualquier función convexa $h \in C^1(\mathbb{R}^n)$ satisface que

$$(\nabla h(x) - \nabla h(y))^T(x - y) \geq 0, \quad \forall x, y \in \mathbb{R}^n.$$

Deduce que $f \in X_\mu$ satisface que

$$(\nabla f(x) - \nabla f(y))^T(x - y) \geq \mu\|x - y\|^2 \quad \forall x, y \in \mathbb{R}^n.$$

Solución: Sea $h \in C^1$, una función convexa, entonces dados $x, y \in \mathbb{R}^n$ se cumple que

$$h(y) \geq h(x) + \nabla h(x)^t(y - x),$$

además

$$h(x) \geq h(y) + \nabla h(y)^t(x - y).$$

Sumando ambas desigualdades obtenemos que

$$\begin{aligned} h(y) + h(x) &\geq h(x) + h(y) + \nabla h(x)^t(y - x) + \nabla h(y)^t(x - y) \\ 0 &\geq (\nabla h(y) - \nabla h(x))^t(x - y) \\ 0 &\leq (\nabla h(x) - \nabla h(y))^t(x - y). \end{aligned}$$

Como queriamos probar.
A continuación. Sea $f \in X_\mu$. Del ejercicio 2 tenemos que $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$ es convexa. Por lo que se cumplen las siguientes desigualdades dadas $x, y \in \mathbb{R}^n$

$$g(x) \geq g(y) + \nabla g(x)^t(x - y),$$

también,

$$g(y) \geq g(x) + \nabla g(y)^t (y - x).$$

Por lo probado anteriormente obtenemos que

$$\begin{aligned}
(\nabla g(x) - \nabla g(y))^t (x - y) &\geq 0 \\
(\nabla f(x) - \mu x - \nabla f(y) + \mu y)^t (x - y) &\geq 0 \\
(\nabla f(x) - \nabla f(y))^t (x - y) - \mu (x - y)^t (x - y) &\geq 0 \\
(\nabla f(x) - \nabla f(y))^t (x - y) &\geq \mu \|x - y\|^2,
\end{aligned}$$

Como queriamos probar.

f) Deduce de 5. la propiedad de los metodos quasi-Newton aplicados a $f \in X_\mu, \mu > 0$.
Solución: Sean f y μ como en el enunciado. La condición de curvatura esta definida como

$$s_k^t y_k > 0,$$

dónde $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ y $s_k = x_{k+1} - x_k$, anteriormente probamos que para $f \in X_\mu$ se cumple que

$$(\nabla f(x) - \nabla f(y))^t (x - y) \geq \mu \|x - y\|^2,$$

para toda x y toda y, así, tomemos $x = x_{k+1}$ y $y = x_k$. Concluyendo que

$$(\nabla f(x_{k+1}) - \nabla f(x_k))^t (x_{k+1} - x_k) \geq \mu \|x_{k+1} - x_k\|^2,$$

por hipótesis $\mu > 0$ y por definición $\|x_{k+1} - x_k\|^2 \geq 0$, por lo tanto obtenemos que

$$0 \leq (\nabla f(x_{k+1}) - \nabla f(x_k))^t (x_{k+1} - x_k) = s_k^t y_k,$$

satisfaciendo la condición de curvatura.

2. Sea $f \in C^2(\mathbb{R}^n)$ y g la función definida por $g(z) = f(Az)$ dónde $A \in \mathbb{R}^{n \times n}$, dónde A es invertible. Consideremos la sucesión $\{x_k\}$ generada por el método de Newton para minimizar f y $\{y_k\}$ la sucesión generada por el metodo de Newton para minimizar g dónde $x_0 = Ay_0$.

a) Establece una relación entre las sucesiones $\{x_k\}$ y $\{y_k\}$.
Solución: Sean f y g como en el enunciado. Notemos que

$$x_0 = Ay_0 \Leftrightarrow A^{-1} x_0 = y_0,$$

luego,

$$g(y_0) = f(Ay_0) = f(AA^{-1}x_0) = f(x_0),$$

con esta definición de y_0 tenemos que g determina el mismo punto inicial que x_0 en f.

De esto, podemos obtener que y_1 debe cumplir que $g(y_1) = f(x_1)$, esto se cumple sí

$$g(y_1) = f(x_1) \Leftrightarrow f(Ay_1) = f(x_1),$$

de lo anterior podemos concluir que $y_1 = A^{-1}x_1$. Esto no depende de la inyectividad de f pues ambas sucesiones comenzaron en el punto $f(x_0)$. Recursivamente podemos realizar esta sustitución y obtener que la relación entre ambas sucesiones está dada por

$$A^{-1} x_k = y_k.$$

b) Concluye.

Solución: Esta relación se cumlpe por el punto incial dado, no directamente por la definición de g, ya que de tener otro punto inicial, no necesariamente se cumple la relación anterior. Pues se tendría la siguiente relación entre los puntos iniciales

$$g(y_0) = f(Ay_0), \qquad f(x_0),$$

si bien, puede existir un k tal que $f(Ay_k) = f(x_k)$, a partir del cual se cumplira la relación descrita anteriormente, también es posible que por el punto inicial, alguna de las sucesiones se cicle o que no se cumpla la relación en ningún punto.

3. Sea $f : \mathbb{R}^n \to \mathbb{R}$ una función continua que satisface la propiedad (P) si

$$\lim_{\|x\| \to +\infty} f(x) = +\infty.$$

Remark: Las funciones que satisfacen (P) son de gran interés en optimización, ya que ellas poseen un mínimo global.

a) ¿Las siguientes funciones satisfacen la propiedad (P)? Justifica tu respuesta

$$f_1(x_1, x_2) = x_1^2 - 2x_1x_2 + x_2^2, \qquad f_2(x_1, x_2) = x_1^4 + x_2^4 - 2x_1x_2.$$

Solución: Consideremos el punto (t, t), notemos que la norma de este punto está dada por

$$\|(t, t)\| = \sqrt{t^2 + t^2} = t\sqrt{2},$$

cuando $t \to \infty$ entonces la norma también tiende a infinito. Sin embargo, el valuar f_1 en este punto obtenemos que

$$f_1(t, t) = t^2 - 2t^2 + t^2 = 0,$$

en este caso, cuando $t \to \infty$ el límite es 0, concluyendo así que f_1 no cumple la propiedad P.

Por otro lado, consideremos el punto (a, b), evaluemos en f_2.

$$f_2(a, b) = a^4 + b^4 - 2ab,$$

La norma del punto elegido es $\sqrt{a^2 + b^2}$, la cual tiende a infinito cuando a ó b tienen a ∞ o $-\infty$, notemos que a y b tienen el mismo comportamiento en la función, es decir, ambos tienen el mismo impacto en valores bastante grandes o pequeños. Por lo cual, sin perdida de generalidad analicemos solo el caso cuando a tiene a los valores de interés.

Observemos que para cualquier valor $a^4 > 0$, más aún, cuando $a \to \infty$ se cumple que $a^4 > -2a$, por lo que $a^4 - 2a$ también tiende a infinito. Consideremos la otra dirección, cuando a tiene a $-\infty$, en ese caso $a^4 - 2a > 0$, nuevamente obtenemos $a^4 - 2a$ también tiene a ∞.

Haciendo un anaálisis análogo para b podemos obtener que f_2 tiende a infinito en todas las direcciones en que la norma crece. Concluyendo que f_2 cumple la propiedad P.

b) Consideremos la función $f : \mathbb{R}^n \to \mathbb{R}$ dada por

$$f(x) = \|Jx - y\|^2 + \alpha\|x\|^2, \quad \alpha \in \mathbb{R},$$

dónde $J \in \mathbb{R}^{n \times n}$ y $y \in \mathbb{R}^n$. ¿Bajo que condiciones en J y α se cumple que f satisface (P)?

Solución: Comencemos expandiendo f de la siguiente manera

$$f(x) = (Jx - y)^t(Jx - y) + \alpha\|x\|^2 = x^t(J^tJ + \alpha I)x - 2y^tJx + \|y\|^2.$$

Es de nuestro interés que $x^t(J^tJ + \alpha I)x$ sea mayor que $-2y^tJx$, lo cual se cumple si la matriz $(J^tJ + \alpha I)$ es definida positiva.

Sea λ_m el minimo eigenvalor de la matriz J^tJ, si $\lambda_m + \alpha > 0$ entonces la matriz de interés es definida positiva.

Notemos que si $\lambda_m > 0$, entonces basta con que $\alpha = 0$, por otro lado, si $\lambda_m = 0$, será necesario que $\alpha > 0$, por último, si $\lambda_m < 0$ entonces estamos interesados en los valores tales que $\alpha > \lambda_m$.

Obteniendo así las condiciones de J y α de interés.

4. *a*) Sea $A \in \mathbb{R}^{n\times n}$ y $U \in \mathbb{R}^{n\times n}$ ortogonal. Muestra que

$$\|A\|_F = \|UA\|_F = \|AU\|_F$$

dónde $\|\cdot\|_F$ es la norma de Frobenius.

Solución: Tengamos en cuenta que la norma de Frobenius se define como

$$\|A\|_F = \sqrt{tra\,(A^TA)}.$$

Dónde $tra(\cdot)$ representa la traza de la matriz. Luego, notemos que

$$\|UA\|_F = \sqrt{tra\,((UA)^TUA)} = \sqrt{tra\,(A^TU^TUA)} = \sqrt{tra\,(A^TA)} = \|A\|_F.$$

Por otro lado

$$\|AU\|_F = \sqrt{tra\,((AU)^TAU)} = \sqrt{tra\,(AU(AU)^T)} = \sqrt{tra\,(AUU^TA^T)} = \sqrt{tra\,(A^TA)} = \|A\|_F.$$

Como queriamos probar.

b) Sea $f \in C^2(\mathbb{R}^n)$ y $\nabla^2 f(x)$ su Hessiana en $x \in \mathbb{R}^n$. Sea Q la matriz de eigenvectores de $\nabla^2 f(x)$, y $\lambda_1, \cdots, \lambda_n$ los eigenvalores de $\nabla^2 f(x)$. Muestra que la matriz ΔA de norma de Frobenius mínima que asegura que

$$\lambda_{min}(\nabla^2 f(x) + \Delta A) \geq \delta$$

está dado por

$$\Delta A = Q diag(\tau_1, \cdots, \tau_n)Q^T, \qquad \text{con } \tau_i = \begin{cases} 0 & \text{si } \lambda_i \geq \delta \\ \delta - \lambda_i & \text{si } \lambda_i < \delta. \end{cases}$$

dónde $\lambda_{\text{mín}}(\nabla^2 f(x) + \Delta A)$ denota el mínimo eigenvalor de la matriz $\nabla^2 f(x) + \Delta A$.

Solución: Tengamos en cuenta que $\nabla^2 f(x)$ es simétrica, por el Teorema de Clairaut. Por lo que se puede descomponer de la siguiente manera

$$\nabla^2 f(x) = Q\Lambda Q^T,$$

dónde Q tiene a los eigenvectores de la Hessiana y Λ es una matriz diagonal que contiene a los eigenvalores. Definamos a

$$\Delta A = Q diag(\tau_1, \ldots, \tau_n)Q^T.$$

Obteniendo así que

$$\nabla^2 f(x) + \Delta A = Q\Lambda Q^T + Q diag(\tau_1, \ldots, \tau_n)Q^T = Q\left(\Lambda + diag(\tau_1, \ldots, \tau_n)\right)Q^T.$$

La matriz resultante tiene como eigenvalores $\lambda_i + \tau_i$. Es de nuestro interés que se cumpla la siguiente desigualdad, para todo i

$$\lambda_i + \tau_i \geq \delta \Rightarrow \tau_i \geq \delta - \lambda_i.$$

Caulquier valor de tau_i que satisfaga la desigualdad anterior garantizará que

$$\lambda_{min}(\nabla^2 f(x) + \Delta A) \geq \delta.$$

Buscamos que ΔA tenga la mínima norma de Frobenius. La cual está dda por

$$\|\Delta A\|_F = \sqrt{\sum_{i=1}^{n} \tau_i^2}.$$

Notemos que si $\lambda_i \geq \delta$ no es necesaria una perturbación a la matriz, pues ya se cumple la desigualdad de interés, en caso contrarío, la pertirbación corresponde a la siguiente diferencia $\delta - \lambda_i$. Motivando así la siguiente definicón para τ

$$\tau_i = \text{máx}(0, \delta - \lambda_i),$$

ya que cualquier valor mayor afectará directamente a la norma de manera negativa para nuestro fin.