Take Home 2

Álvaro José Álvarez Arranz

May 8, 2024

1. We want to solve the following constrained restriction problem:

$$\begin{aligned} \min\ & x^2 + 2xy + 2y^2 - 3x + y \\ \text{s.t. } & x + y = 1 \\ & x \geq 0, y \geq 0 \end{aligned}$$

   (a) Write its Lagrangian with α, β the multipliers of the inequality constrains.

   (b) Write the KKT conditions.

   (c) Use them to solve the problem.

Solución:

- **Lagrangiana** de f en términos de α, β, λ:

$$\mathcal{L}\left(x, y, \alpha, \beta, \lambda\right) = x^2 + 2xy + 2y^2 - 3x + y - \alpha x - \beta y + \lambda\left(x + y - 1\right)$$

- Ahora, escribimos las condiciones KKT. Las condiciones KKT son:

$$\begin{aligned} \nabla f\left(x^*\right) + \sum_i^m \lambda_i \nabla g_i\left(x^*\right) = 0, \\ \lambda_i g_i\left(x^*\right) = 0 \end{aligned}$$

Lo aplicamos al problema , con lo que **nuestras condiciones KKT** son:

$$\begin{aligned} 2x + 2y - 3 - \alpha + \lambda = 0 \\ 2x + 4y + 1 - \beta + \lambda = 0 \\ \alpha x = 0 \\ \beta y = 0 \end{aligned}$$

- Ahora, resolvemos el sistema de ecuaciones para intentar encontrar un mímimo. Vamos a mirar varios casos:

  (a) Caso $\alpha = \beta = 0$. Nuestro sistema queda de la siguiente manera:

  $$\begin{aligned} 2x + 2y - 3 + \lambda &= 0 \\ 2x + 4y + 1 + \lambda &= 0 \end{aligned}$$

  Sustituimos $2x$ de la primera ecuación en la segunda y obtenemos:

  $$-2y + 3 - \lambda + 4y + 1 + \lambda = 0 \implies 2y = -4 \implies y = -2$$

  Dado que la condición inicial indica $y \geq 0$, no se obtiene ningún punto KKT.

  (b) Caso $\alpha, \beta > 0$. En este caso, de las condiciones KKT se obtiene $x = y = 0$, que no cuadra con la condición inicial $x + y = 1$ y tampoco se obtiene ningún punto KKT.

  (c) Caso $\alpha > 0, \beta = 0$. Mirando las condiciones KKT, con $\alpha > 0 \implies x = 0 \implies y = 1$, con lo que tenemos el punto $(0, 1)$. Comprobamos si el punto obtenido concuerda con el resto de condiciones KKT. Primero, usamos el punto en la segunda condición

  $$0 + 4 + 1 - 0 + \lambda = 0 \implies \lambda = -5$$

  Sustituimos λ en la primera condición KKT y obtenemos

  $$0 + 2 - 3 - \alpha - 5 = 0 \implies \alpha = -6$$

  Como no concuerda con la condición impuesta en este caso $\alpha > 0$, el punto obtenido no es un punto KKT.

  (d) Caso $\alpha = 0, \beta > 0$. Usando el mismo razonamiento que en el caso anterior, obtenemos $(1, 0)$ como posible punto KKT. Usamos el punto en la primera ecuación KKT:

  $$2 + 0 - 3 - 0 + \lambda = 0 \implies \lambda = 1$$

  Hacemos la sustitución en la segunda ecuación KKT y obtenemos:

  $$2 + 0 + 1 - \beta + 1 = 0 \implies \beta = 4$$

  Lo que concuerda con la condición impuesta del caso, por lo que hemos obtenido un **punto KKT**.

Dado que solamente tenemos un único candidato, $(1,0)$, es el valor **mínimo** con un valor de $f(1,0) = -2$.

2. i. Prove that if f is convex and g strictly convex, then f + g is strictly convex.
   ii. Prove that if f is convex and for $x \in \mathbb{R}^d$ and direction $d \neq 0$, we define $g(t) = f(x + td)$, then g is convex.
   iii. Conservely, if for any x and d, the function $g(t) = f(x + td)$ is convex, prove that f is also convex.

   Hint: Write $f(\lambda y + (1-\lambda)x) = f(x + \lambda(y - x))$, observe that $g(\lambda) = g(\lambda \cdot 1 + (1-\lambda) \cdot 0)$ and use the convexity of g.

Solución:

i. Una función f es convexa en un intervalo x_1, x_2 cuando para todo $\lambda \in [0,1]$ se cumple
$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$$
De manera similar, una función g es estrictamente convexa en el intervalo x_1, x_2 cuando para todo $\lambda \in [0,1]$ se cumple
$$g(\lambda x_1 + (1-\lambda)x_2) < \lambda g(x_1) + (1-\lambda) g(x_2)$$
Vamos a considerar la función $h(x) = f(x) + g(x)$.
Sumando las desigualdades de f y g, obtenemos:
$$\begin{aligned}
&f(\lambda x_1 + (1-\lambda)x_2) + g(\lambda x_1 + (1-\lambda)x_2) \\
&< \lambda f(x_1) + (1-\lambda) f(x_2) + \lambda g(x_1) + (1-\lambda) g(x_2) \\
&= \lambda(f(x_1) + g(x_1)) + (1-\lambda)(f(x_2) + g(x_2))
\end{aligned}$$
Que es precisamente
$$h(\lambda x_1 + (1-\lambda)x_2) < \lambda h(x_1) + (1-\lambda) h(x_2)$$
Por lo que se demuestra que $f + g$ es estrictamente convexa.

ii. Basándonos en las definiciones de convexidad y estrictamente convexo del apartado anterior. Definamos $u = x + t_1 \cdot d$ y $v = x + t_2 \cdot d$. Con ello, podemos reescribir la desigualdad de f del punto anterior como:
$$f(u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v)$$

Ahora vamos a sustituir u y v en la definición de $g(t)$:

$$\begin{aligned} g(\lambda t_1 + (1-\lambda) t_2) &= f(x + (\lambda t_1 + (1-\lambda) t_2) \cdot d) \\ &= f(\lambda (x + t_1 \cdot d) + (1-\lambda)(x + t_2 \cdot d)) \\ &= f(\lambda u + (1-\lambda) v) \end{aligned}$$

De modo que, de acuerdo con la definición de convexidad, tenemos:

$$\begin{aligned} g(\lambda t_1 + (1-\lambda) t_2) &\leq \lambda f(u) + (1-\lambda) f(v) \\ &= \lambda f(x + t_1 \cdot d) + (1-\lambda) f(x + t_2 \cdot d) \\ &= \lambda g(t_1) + (1-\lambda) g(t_2) \end{aligned}$$

Demostrando, por tanto, que $g(t)$ es convexa.

iii. Ahora vamos a hacer el el caso contrario al punto anterior. Dado que $g(t) = f(x + t \cdot d)$ para cualquier x y d, esto implica que se cumple

$$g(\lambda t_1 + (1-\lambda) t_2) \leq \lambda g(t_1) + (1-\lambda) g(t_2)$$

Vamos a tomar $t_1 = 0$ y $t_2 = 1$, que nos queda:

$$\begin{aligned} g(\lambda \cdot 0 + (1-\lambda) \cdot 1) &\leq \lambda g(0) + (1-\lambda) g(1) \\ g((1-\lambda)) &\leq \lambda g(0) (1-\lambda) g(1) \end{aligned}$$

Reemplazando $g(t)$ con $f(x + t \cdot d)$ nos queda:

$$f(x + (1-\lambda) \cdot d) \leq \lambda f(x) + (1-\lambda) f(x + d)$$

Demostrando así que f es convexa.

3. Prove **Jensen's inequality**: if f is convex on $\mathbb{R}^d$ and $\sum_1^k \lambda_i = 1$, with $0 \leq \lambda \leq 1$, we have for any $x_1, ..., x_k \in \mathbb{R}^d$

$$f\left(\sum_1^k \lambda_i x_i\right) \leq \sum_1^k \lambda_i f\left(x_i\right)$$

Hint: just write $\sum_1^k \lambda_i x_i = \lambda_1 x_1 + (-\lambda_1)\, v$ for an appropriate v and apply repeatedly the definition of a convex function. Start with $k = 3$ and carry on.

Solución:
Vamos a demostrar la desigualdad por inducción. Empezamos con el **caso base** $k = 3$. Queremos ver que, si $x_1, x_2 x_3 \in \mathbb{R}^d$, entonces

$$f\left(\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3\right) \leq \lambda 1 f\left(x_1\right) + \lambda_2 f\left(x_2\right) + \lambda_3 f\left(x_3\right)$$

Reformulamos el argumento de f en la parte izquierda como:

$$\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 = \lambda_1 x_1 + \left(1 - \lambda_1\right) \underbrace{\left(\frac{\lambda_2}{1 - \lambda_1} x_2 + \frac{\lambda_3}{1 - \lambda_1} x_3\right)}_{\in \mathbb{R}^d}$$

También hay que notar que, dado que $\lambda_1 + \lambda_2 \lambda_3 = 1$, entonces $\lambda_3 = 1 - \lambda_1 - \lambda_2$, y la expresión anterior queda de la siguiente manera:

$$\begin{aligned}
\lambda_1 x_1 + \left(\frac{\lambda_2}{1 - \lambda_1} x_2 + \frac{\lambda_3}{1 - \lambda_1} x_3\right) &= \lambda_1 x_1 + \left(1 - \lambda_1\right) \left(\frac{\lambda_2}{1 - \lambda_1} x_2 + \frac{1 - \lambda_1 - \lambda_2}{1 - \lambda_1} x_3\right) \\
&= \lambda_1 x_1 + \left(1 - \lambda_1\right) \left(\frac{\lambda_2}{1 - \lambda_1} x_2 + \left(1 - \frac{\lambda_2}{1 - \lambda_1}\right) x_3\right)
\end{aligned}$$

Ahora, podemos aplicar la definición de convexidad dos veces:

$$\begin{aligned}
f\left(\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3\right) &= f\left(\lambda_1 x_1 + \left(1 - \lambda_1\right) \left(\frac{\lambda_2}{1 - \lambda_1} x_2 + \left(1 - \frac{\lambda_2}{1 - \lambda_1}\right) x_3\right)\right) \\
&\leq \lambda_1 f\left(x_1\right) + \left(1 - \lambda_1\right) f\left(\frac{\lambda_2}{1 - \lambda_1} x_2 + \left(1 - \frac{\lambda_2}{1 - \lambda_1}\right) x_3\right) \\
&\leq \lambda_1 f\left(x_1\right) + \left(1 - \lambda_1\right) \left(\frac{\lambda_2}{1 - \lambda_1} f\left(x_2\right) + \left(1 - \frac{\lambda_2}{1 - \lambda_1}\right) f\left(x_3\right)\right) \\
&= \lambda_1 f\left(x_1\right) + \lambda_2 f\left(x_2\right) + \left(1 - \lambda_1 - \lambda_2\right) f\left(x_3\right) \\
&= \lambda_1 f\left(x_1\right) + \lambda_2 f\left(x_2\right) + \lambda_3 f\left(x_3\right)
\end{aligned}$$

Finalizando con el **caso base**.

Vamos ahora a probar el paso de inducción. Considerando que la hipótesis es cierta para $k = n$, se demostrará que es cierto para $k = n + 1$. Procedemos como lo hicimos en el caso anterior:

$$f\left(\sum_{i=1}^{n+1} \lambda_i x_i\right) = f\left(\lambda_1 x_1 + \sum_{i=2}^{n+1} \lambda_i x_i\right)$$

Si consideramos $v = \sum_{i=2}^{n+1} \frac{\lambda_i}{1-\lambda_1},\ x_i \in \mathbb{R}$, podemos escribir la expresión anterior de la siguiente manera:

$$f\left(\lambda_1 x_1 + (1-\lambda_1)\, v\right)$$

Aplicamos la definición de convexidad en f:

$$\begin{aligned} f\left(\lambda_1 x_1 + (1-\lambda_1)\, v\right) &\leq \lambda_1 f(x_1) + (1-\lambda_1)\, f(v) \\ &\overset{(1)}{\leq} \lambda_1 f(x_1) + (1-\lambda_1) \sum_{i=2}^{n+1} \frac{\lambda_i}{1-\lambda_1} f(x_i) \\ &= \sum_{i=1}^{n+1} \lambda_i f(x_i) \end{aligned}$$

En (1) se ha aplicado la hipótesis de la inducción, probando el caso $k = n + 1$, y así terminando la demostración.

4. i. Show that the hinge $f(x) = \max\{0, -x\}$ and the ϵ-insensitive $\max\{0, |x| - \epsilon\}$ loss functions are convex, compute their subgradients and proximals.

   ii. Which are the fixed points of these proximals? Are they related to their minima?

Solución:

i. Se va a demostrar la convexidad de las funciones. Empezamos por la **función bisagra** $f(x) = \max\{0, -x\}$. El método a usar será el cálculo de la segunda derivada.

   La función bisagra se puede escribir de la siguiente manera:

$$f = \begin{cases} 0 & \text{si } x \geq 0 \\ -x & \text{caso contrario} \end{cases}$$

   Ahora, tomamos la segunda derivada en las dos regiones $x < 0, x \geq 0$.

- Para $x < 0$, la función $f(x) = -x$ es una función lineal, por lo tanto su segunda derivada es 0, que es no negativa.
- Para $x \geq 0$, la función $f(x) = 0$ es una constante, por lo que su segunda derivada también es 0, que no es negativa.

Como en ambos casos la segunda derivada no es negativa, implica que $f(x)$ es convexa en todo su dominio.
Para la **función de pérdida ϵ-insensible** $g(x) = \max\{0, |x| - \epsilon\}$ haremos el mismo razonamiento que con la función f en tres regiones: $x < \epsilon, -\epsilon \leq x \leq \epsilon$ y $x > \epsilon$.

- Para $x < -\epsilon$, la función $g(x) = |x| - \epsilon$ es una función lineal, por lo que su segunda derivada es 0, que no es negativa.
- Para $-\epsilon \leq x \leq \epsilon$, la función $g(x) = 0$ es una constante, por lo que su segunda derivada es 0, que no es negativa.
- Para $x > \epsilon$, la función $g(x) = x - \epsilon$ es lineal también.

Como en los tres casos la segunda derivada no es negativa, implica que $g(x)$ también es convexa en todo su dominio.

Ahora vamos a pasar al cálculo de los subgradientes. Primero la función bisagra:

- Para $x \geq 0$, el subgradiente de $f(x)$ es 0 dado que es constante en esta región.
- Para $x < 0$, el subgradiente de $f(x)$ es -1 dado que $f(x) = -x$ en esta región.

Ahora la función ϵ-insensitiva:

- Para $x < -\epsilon$: $g(x) = |x| - \epsilon = -x - \epsilon$, implica que el subgradiente es -1.
- Para $x > \epsilon$: $g(x) = |x| - \epsilon = x - \epsilon$, implica que el subgradiente es 1.
- Para $-\epsilon \leq x \leq \epsilon$ el subgradiente está en el intervalo $[-1, 1]$, dado que $g(x)$ no es diferenciable en $x \pm \epsilon$

Ahora que tenemos los subgradientes, vamos con los **proximales**. Para la función bisagra es:

$$\operatorname{prox}_{\lambda f}(x) = \arg\min_{z} \left\{ f(x) + \frac{1}{2\lambda} ||z - x||_2^2 \right\}$$

$$\operatorname{prox}_{\lambda f}(x) = \begin{cases} 0 & \text{si } x > 0 \\ x & \text{caso contrario} \end{cases}$$

Para la función de pérdida es:

$$\text{prox}_{\lambda g}(x) = \arg\min_z \left\{ g(x) + \frac{1}{2\lambda}||z-x||_2^2 \right\}$$

$$\text{prox}_{\lambda g}(x) = \begin{cases} 0 & \text{si } |x| \leq \epsilon \\ x - \lambda & \text{si } x > \epsilon \\ x + \lambda & \text{si } x < -\epsilon \end{cases}$$

5. Prove that the following function is convex:

$$f(x) \begin{cases} x^2 - 1, & |x| > 1 \\ 0, & |x| \leq 1 \end{cases}$$

compute its proximal and relate ts fixed points to the minima of f.

Solución:
Para demostrar que $f(x)$ se puede usar el mismo razonamiento que se usó en un ejercicio anterior con la *segunda derivada.* Por partes:

- Para $|x| > 1$, la función $f(x) = x^2 - 1$ su segunda derivada es $f''(x) = 2$, que es no negativa.
- Para $|x| \leq 1$, la función es constante, por lo que su segunda derivada es 0, no negativa.

como en ambos casos la segunda derivada es *no negativa*, esto implica que $f(x)$ es convexa en todo su dominio.

A continuación, calculamos el **proximal**. Sabemos que el proximal de una función se define como:

$$\text{prox}_f(x) = \arg\min_z \left\{ f(z) + \frac{1}{2}||u-x||^2 \right\}$$

En nuestro caso, $u, x \in \mathbb{R}$ y también $|x|^2 = x^2 \ \forall x \in \mathbb{R}$, de modo que el proximal se puede escribir de la siguiente manera:

$$\text{prox}_f(x) = \arg\min_z \left\{ f(z) + \frac{1}{2}(z-x)^2 \right\} = \arg\min_z \phi_x(z)$$

donde

$$\phi_x(z) = \begin{cases} u^2 - 1 + \frac{1}{2}(z-x)^2 & \text{si } |z| > 1 \\ \frac{1}{2}(z-x) & \text{si } |z| \leq 1 \end{cases}$$

Ahora podemos calcular el operador proximal, dependiendo de los diferentes casos.

- En el caso $|z| \leq 1$, el mínimo es $y = x$, lo que significa que $\operatorname{prox}_f(x) = x, \ |x| < 1$.
- En el caso $|z| > 1$ vemos que

$$\left(z^2 - 1 + \frac{1}{2}(z-x)^2\right)' = 3z - x = 0 \iff u = \frac{x}{3}$$

lo que implica que $\operatorname{prox}_f(x) = \frac{x}{3}$ si $x > 3$.

- Por último, tenemo que ver en el resto de valores posibles de x, que se corresponden a los puntos donde $\phi_x(z)$ no es diferenciable, es decir, $\{-1, 1\}$.

$$\phi_x(-1) = \frac{1}{2}(-1+x)^2 \text{ y } \phi_x(1) = \frac{1}{2}(1-x)^2$$

Basicamente, estamosmos calculando la distancia entre x y $-1, 1$ en cada caso.

  - Si $x \in [-3, -1)$, entonces $\phi_x(-1) < \phi_x(1)$ y $\operatorname{prox}_f(x) = -1$.
  - Cuando $x \in (1, 3]$, $\phi_x(1) < \phi_x(-1)$, de modo que $\operatorname{prox}_f(x) = 1$.

En conclusión, el **proximal** de f es:

$$\operatorname{prox}_f(x) = \begin{cases} -1 & \text{si } x \in [-3, 1) \\ x & \text{si } x \in [-1, 1] \\ 1 & \text{si } x \in (1, 3] \\ \frac{x}{3} & \text{resto de casos} \end{cases}$$

Puede verse que, los **puntos fijos** de $\operatorname{prox}_f(x)$ son los que cumplen $\operatorname{prox}_f(x) = x$, que en nuestro caso son los puntos $[-1, 1]$.