

# **OPTIMIZACIÓN LIBRE DE DERIVADAS**

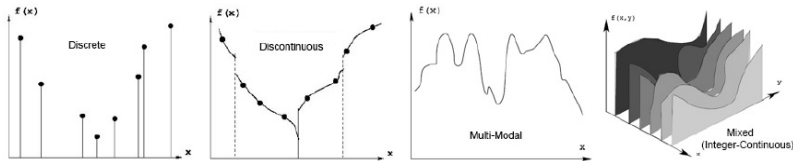
ALAN REYES-FIGUEROA  
INTELIGENCIA ARTIFICIAL

(AULA 12) 24.FEBRERO.2022

# Otros Esquemas de Optimización

Al utilizar optimización en la solución de aplicaciones prácticas, a menudo encontramos uno o más de los siguientes desafíos:

- funciones  $y$  / o restricciones no diferenciables
- espacio factible desconectado y / o no convexo
- espacio discreto factible
- variables mixtas (discretas, continuas, permutación)
- gran dimensionalidad
- múltiples mínimos locales (multimodal)
- múltiples objetivos



Algunas funciones problemáticas: (a) discreta, (b) discontinua, (c) multimodal, (d) mixta.

# Otros Esquemas de Optimización

Los optimizadores basados en gradientes son eficientes para encontrar mínimos locales para problemas convexos de alta dimensión, no linealmente restringidos sin embargo, la mayoría de los optimizadores basados en gradientes tienen problemas para manejar funciones ruidosas y discontinuas, y no están diseñadas para manejar problemas multimodales o variables de diseño discretas y mixtas discretas-continuas.

Para muchos uno de estos desafíos, la optimización ha encontrado respuesta.

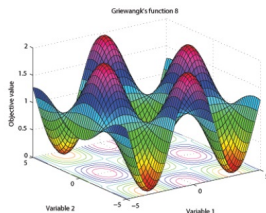
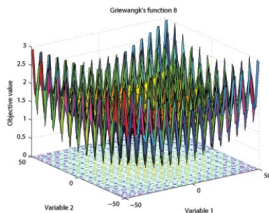
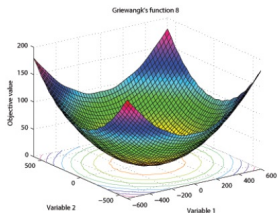
- funciones no diferenciables: métodos de búsqueda.
- espacio factible desconexo/no convexo: variables de estado
- espacio discreto factible: optimización combinatoria
- variables mixtas: programación entera, programación mixta
- múltiples mínimos locales: optimización estocástica.
- múltiples objetivos: optimización multi-objetivo.

# Otros Esquemas de Optimización

Por ejemplo, la función GRIEWANK:

$$f(\mathbf{x}) = \sum_{i=1}^n \left( \frac{x_i^2}{4000} \right) - \prod_{i=1}^n \cos \left( \frac{x_i}{\sqrt{i}} \right) + 1, \quad -600 \leq x_i \leq 600.$$

Esta función se ve inocente: parece engañosamente suave cuando se traza en un dominio grande (gran escala), pero cuando se acerca, puede ver que el espacio de diseño tiene múltiples mínimos locales, aunque la función sigue siendo suave.



Función de GRIEWANK 8.

# Otros Esquemas de Optimización

**Ejemplo:** Dada  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  no diferenciable. ¿Cómo podemos resolver el problema de minimización?

- aproximar gradientes y derivadas con diferencias finitas,
- usar algoritmos de búsqueda (talvez no encuentre el mínimo global, pero puede brindar buenas soluciones sub-óptimas),
- usar métodos que no requieran derivadas (por lo general estos métodos basan su decisión sólo conociendo valores de  $f$ ).

**Ejemplo:** Dada  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  multimodal. ¿Cómo podemos resolver el problema de minimización?

- iniciar los optimizadores tipo gradiente desde diferentes puntos iniciales.
- usar algoritmos de búsqueda estocástica.

# Otros Esquemas de Optimización

Existen varias familia amplias de métodos de optimización que no requieren derivadas.

Mencionamos algunos

- método de Nelder-Mead (también llamado Simplex).
- método de división por rectángulos.
- recocido simulado (*simulated annealing*)
- búsqueda tabú
- enjambres de partículas (*particle swarm optimization* PSO),
- evolución diferencial (DE)
- algoritmos genéticos (GA)
- EDAs (algoritmos de estimación de distribución)

# Método de Nelder-Mead

El método de NELDER-MEAD, es un algoritmo de optimización que no requiere derivadas. Realiza una búsqueda en un espacio  $n$ -dimensional usando ideas heurísticas.

También se conoce como *simplex no lineal* y no debe confundirse con el método simplex lineal (usado en programación lineal) con el que no tiene nada en común.

## **Ventajas:**

- no requiere el cálculo de derivadas,
- y que no requiere la función objetivo sea suave.

## **Desventajas:**

- no es un algoritmo muy eficiente (lento),
- en la práctica se limita a problemas con aproximadamente 10 o menos variables de diseño (baja dimensionalidad); por encima de este número de variables la convergencia se vuelve cada vez más difícil.

# Método de Nelder-Mead

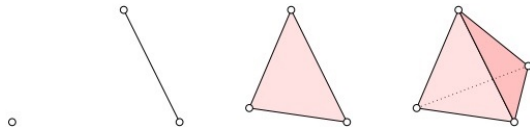
Un **simplejo** o **simplexo** (*simplex* en inglés) de dimensión  $n$ , es una estructura geométrica  $n$ -dimensional formado por  $n + 1$  puntos, de forma que no están en el mismo subespacio.

Esto se matematiza de la siguiente forma

## Definición

Sean  $\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_n \in \mathbb{R}^d$ ,  $n + 1$  puntos en un espacio  $d$ -dimensional,  $d \geq n$ . Decimos que la estructura  $\Delta_n = [\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_n]$  es un **simplejo**  $n$ -dimensional si los vectores

$\mathbf{p}_1 - \mathbf{p}_0, \mathbf{p}_2 - \mathbf{p}_0, \dots, \mathbf{p}_n - \mathbf{p}_0 \in \mathbb{R}^d$ , son l.i..



Simplejos de dimensiones  $n = 0, 1, 2, 3$ .



# Método de Nelder-Mead

En realidad,  $\Delta_n = [\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_n]$  es el conjunto de todas las combinaciones convexas

$$\Delta_n = \left\{ \mathbf{x} = \sum_{j=0}^n t_j \mathbf{p}_j : 0 \leq t_j \leq 1, \sum_{j=0}^n t_j = 1 \right\}.$$

El propósito de construir el simplejo es que es la unidad mínima de volumen  $n$ -dimensional (o más bien, es el volumen  $n$ -dimensional con menor número de vértices posible). La idea del método es que vamos a construir una secuencia de simplejos  $\{\Delta_k\}_k$ , que sirven como espacio de búsqueda, y vamos a modificar estos en función del valor de  $f$  en los vértices de  $\Delta_k$ .

El algoritmo de Nelder Mead comienza con un simplejo  $n$ -dimensional ( $n + 1$  puntos o vértices) y luego modifica este simplejo en cada iteración usando cinco operaciones básicas:

- reflexión,
- expansión,
- contracción interior,
- contracción exterior,
- *shrinking*.

# Método de Nelder-Mead

La secuencia de operaciones a realizar se elige en función de los valores relativos de la función objetivo  $f$  en cada uno de los vértices  $\mathbf{p}_j$  de simplejo.

El primer paso del algoritmo es encontrar los  $n + 1$  puntos del simplejo dado un valor inicial  $\mathbf{x}_0$ . Esto se puede hacer fácilmente simplemente agregando un paso  $\mathbf{d}_j$  a cada componente de  $\mathbf{x}_0$  para generar  $n$  puntos  $\mathbf{p}_j = \mathbf{x}_0 + \mathbf{d}_j$ . nuevos puntos.

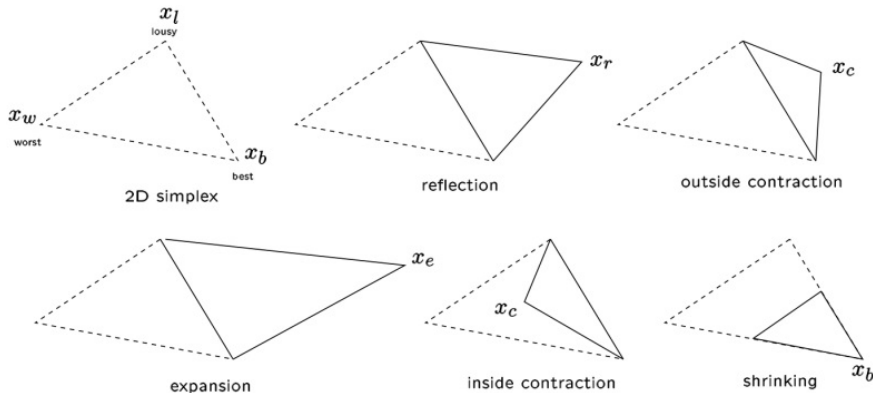
En la práctica, es preferible generar un simplejo con aristas de igual longitud. Si  $c$  es la longitud de aristas,  $\mathbf{x}_0$  es el punto inicial, los  $n + 1$  puntos restantes pueden calcularse sumando como  $\mathbf{p}_j = \mathbf{x}_0 + \mathbf{v}_j$ , sumando un vector  $\mathbf{v}_j$  a  $\mathbf{x}_0$  cuyos componentes son todos  $b$ , excepto por el  $j$ -ésimo componente, que vale  $a$ , donde

$$b = \frac{c}{n\sqrt{2}}(\sqrt{n+1} - 1), \quad a = b + \frac{c}{\sqrt{2}}, \quad \text{para } j = 1, \dots, n+1.$$

Después de generar el simplejo inicial, evaluamos la función objetivo en cada vértice  $\mathbf{p}_j$  con el fin de identificar tres puntos clave: los puntos asociados con el más alto ("peor") el segundo más alto ("pésimo") y el más bajo ("mejor") valores de la  $f$  objetivo.

# Método de Nelder-Mead

El algoritmo de Nelder-Mead realiza cinco operaciones principales para el simplejo: reflexión, expansión, contracción exterior, contracción interior y encogimiento.



Operaciones de transformación en el método de NELDER-MEAD,  $n = 2$ .

# Método de Nelder-Mead

Cada una de estas operaciones genera un nuevo punto  $\mathbf{x}_{k+1}$ , y la secuencia de operaciones realizadas en una iteración depende del valor del objetivo  $f(\mathbf{x}_{k+1})$ , en relación con los otros puntos clave.

Denotamos por  $\mathbf{x}_w$ ,  $\mathbf{x}_\ell$  y  $\mathbf{x}_b$  denoten el peor, el segundo peor y el mejor vértice, respectivamente, de entre los  $n + 1$  puntos del simplejo. Esto es

$$f(\mathbf{x}_b) \leq f(\mathbf{p}_j) \leq f(\mathbf{x}_\ell) \leq f(\mathbf{x}_w), \quad j = 1, 2, \dots, n + 1.$$

Primero calculamos el promedio de los  $n$  vértices, excluyendo el peor  $\mathbf{x}_w$ ,

$$\mathbf{x}_a = \frac{1}{n} \sum_{j=1, j \neq w}^{n+1} \mathbf{p}_j.$$

Luego, la recta de  $\mathbf{x}_w$  a  $\mathbf{x}_a$  es una dirección de descenso. Hallamos un nuevo punto sobre esta línea por reflexión, de la foama

$$\mathbf{x}_r = \mathbf{x}_a + \alpha(\mathbf{x}_a - \mathbf{x}_w).$$

con  $\alpha = 1$ . Si el valor de la función en este punto reflejado  $\mathbf{x}_r$  es mejor que en el mejor

# Método de Nelder-Mead

punto  $\mathbf{x}_b$ , esto es  $f(\mathbf{x}_r) < f(\mathbf{x}_b)$ , entonces la reflexión ha tenido éxito especial y damos un paso más en la misma dirección al realizar una expansión,

$$\mathbf{x}_e = \mathbf{x}_r + \gamma(\mathbf{x}_r - \mathbf{x}_a),$$

donde el parámetro de expansión generalmente se establece en  $\gamma = 1$ .

Si el punto reflejado  $\mathbf{x}_r$  es peor que el peor punto  $\mathbf{x}_w$ , esto es  $f(\mathbf{x}_r) > f(\mathbf{x}_w)$ , asumimos que existe un mejor candidato entre  $\mathbf{x}_w$  y  $\mathbf{x}_a$  y realizamos una contracción interna

$$\mathbf{x}_c = \mathbf{x}_a - \beta(\mathbf{x}_a - \mathbf{x}_w),$$

donde el factor de contracción generalmente se establece en  $\beta = 0.5$ .

Si el punto reflejado  $\mathbf{x}_r$  no es peor que el peor punto  $\mathbf{x}_w$ , pero es peor que el punto pésimo  $\mathbf{x}_\ell$ , ( $f(\mathbf{x}_w) > f(\mathbf{x}_r) > f(\mathbf{x}_\ell)$ ) entonces se realiza una contracción externa

$$\mathbf{x}_o = \mathbf{x}_a + \beta(\mathbf{x}_a - \mathbf{x}_w).$$

Después de la expansión, aceptamos el nuevo punto  $\mathbf{x}_e$  si el valor de la función objetivo satisface  $f(\mathbf{x}_e) < f(\mathbf{x}_b)$ . De lo contrario, aceptamos el punto reflejado anterior  $\mathbf{x}_r$ .

# Método de Nelder-Mead

Si la reflexión, la expansión y la contracción fallan, recurrimos a una operación de encogimiento. Esta operación retiene el mejor punto  $\mathbf{x}_b$  y encoge el simplejo, de modo que todos los puntos del simplejo excepto  $\mathbf{x}_b$ , calculamos

$$\mathbf{x}_i = \mathbf{x}_b + \rho(\mathbf{x}_i - \mathbf{x}_b),$$

donde el parámetro de escala usualmente se establece como  $\rho = 0.5$ .

# Método de Nelder-Mead

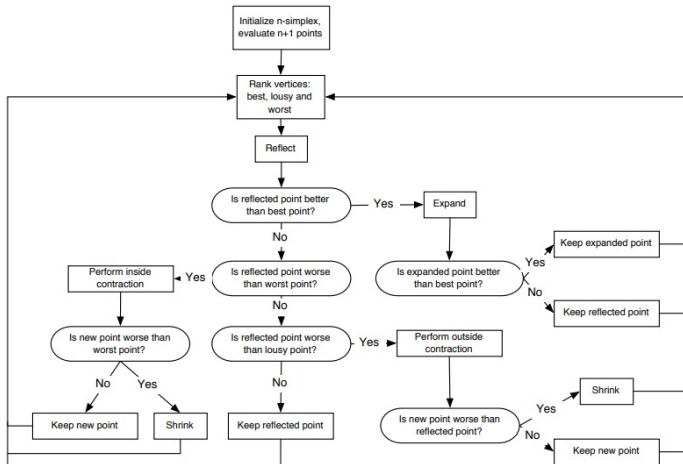


Diagrama de flujo de operaciones en un paso de NELDER-MEAD.

# Método de Nelder-Mead

**Algoritmo:** (NELDER-MEAD).

*Input:* Punto inicial  $\mathbf{x}_0$ , función objetivo  $f$ ,

*Output:* Óptimo local  $\mathbf{x}$  de  $f$ .

Create a simplex with edge length  $c$ ,

Repeat for  $k = 0, 1, 2, \dots$  until stop criterion holds:

Identify  $\mathbf{x}_w, \mathbf{x}_\ell$  and  $\mathbf{x}_b$  with values  $f_w, f_\ell, f_b$  from vertices,

Evaluate  $\mathbf{x}_a$ , the average of vertices excluding  $\mathbf{x}_w$

Perform reflection  $\mathbf{x}_r$ , and evaluate  $f_r$ ,

If  $f_r < f_b$ :

Perform expansion  $\mathbf{x}_e$ , and evaluate  $f_e$ ,

If  $f_e < f_b$ :  $\mathbf{x}_w \leftarrow \mathbf{x}_e, f_w \leftarrow f_e$  (accept expansion),

Else:  $\mathbf{x}_w \leftarrow \mathbf{x}_r, f_w \leftarrow f_r$  (accept reflection).

Elif  $f_r < f_\ell$ :  $\mathbf{x}_w \leftarrow \mathbf{x}_r, f_w \leftarrow f_r$  (accept reflected point),

Else:

If  $f_r > f_w$ :

Perform inside contraction  $\mathbf{x}_c$  and evaluate  $f_c$ ,

If  $f_c < f_w$ :  $\mathbf{x}_w \leftarrow \mathbf{x}_c, f_w \leftarrow f_c$  (accept inside contraction),

Else: shrink the simplex.

Else:

Perform outside contraction  $\mathbf{x}_o$  and evaluate  $f_o$ ,

If  $f_o < f_r$ :  $\mathbf{x}_w \leftarrow \mathbf{x}_o, f_w \leftarrow f_o$  (accept outside contraction),

Else: shrink the simplex.



# Método de Nelder-Mead

Como criterios de convergencia se pueden usar los siguientes:

- Tamaño del simplex:

- ◇  $\text{diam } \Delta = \max_{i,j} |\mathbf{p}_i - \mathbf{p}_j|$

- ◇  $\text{size } \Delta = \sum_{i=1}^n |\mathbf{p}_i - \mathbf{p}_{n+1}|$

- Varianza o desviación estándar de la función en los vértices del simplex:

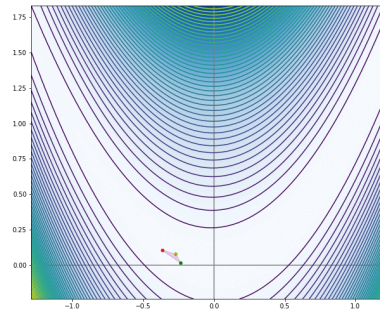
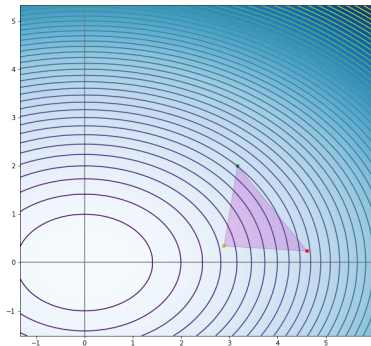
- ◇  $\sigma^2 = \frac{1}{n+1} \sum_{i=1}^{n+1} (f_i - \bar{f})^2$ , donde  $f_i = f(\mathbf{p}_i)$  y  $\bar{f} = \frac{1}{n+1} \sum_{i=1}^{n+1} f_i$

- Distancias o normas entre puntos consecutivos:

Si  $\mathbf{x}_k = \frac{1}{n+1} \sum_{j=1}^{n+1} \mathbf{p}_j$  (promedio de los vértices del simplex en el paso  $k$ )

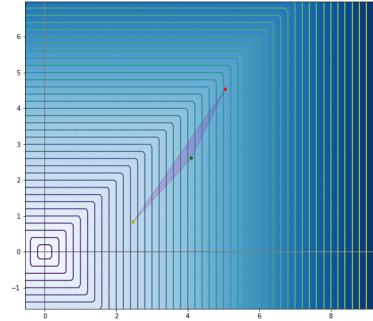
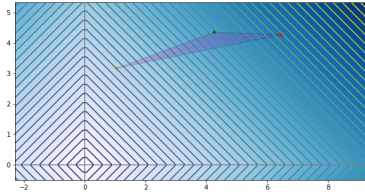
- ◇  $|\mathbf{x}_{k+1} - \mathbf{x}_k|$

# Ejemplos



Ejemplos en  $\mathbb{R}^2$ : (a) Función cuadrática  $f(x, y) = x^2 + 2y^2$ , (b) Función de Rosembrock.

# Ejemplos



Ejemplos en  $\mathbb{R}^2$ : (a) Función  $f(x, y) = |x| + |y|$ , (b) Función  $f(x, y) = \max\{|x|, |y|\}$ .