Introducción a Optimización

Semestre 2011-2

Definición de Optimización

- Según Bonnans et. al.:
  - Dado un conjunto X de variables númericas reales y una función f : X, encontrar una serie de valores $x^* \in X$ que, para cualquier otro x: $f(x) \geq f(x^*)$
- ¿Qué significa esto?

En español...

- Encontrar la solución que minimiza una *función objetivo*.

Pero...

- Según Bonnans et. al.:
  - Dado un conjunto X de variables númericas **reales** y una función f : X, encontrar una serie de valores $x^* \in X$ que, para cualquier otro x: $\mathbf{f(x) \geq f(x^*)}$

...

- ¿Valores enteros, complejos, discretos?
- ¿Maximizar?

Terminología

- Función objetivo:
  - La función a optimizar.
  - Uno o más variables.
    - Usualmente restringidas a rangos de valores.
    - Pueden ser discretas (enteras) o contínuas.
  - Define el Espacio de Solución.
  - Ejemplos.
    - $f(x) = x^2$
    - $f(x,y,z) = ?$
      - También conocidas como **Cajas Negras**

Parecido al problema de SAT

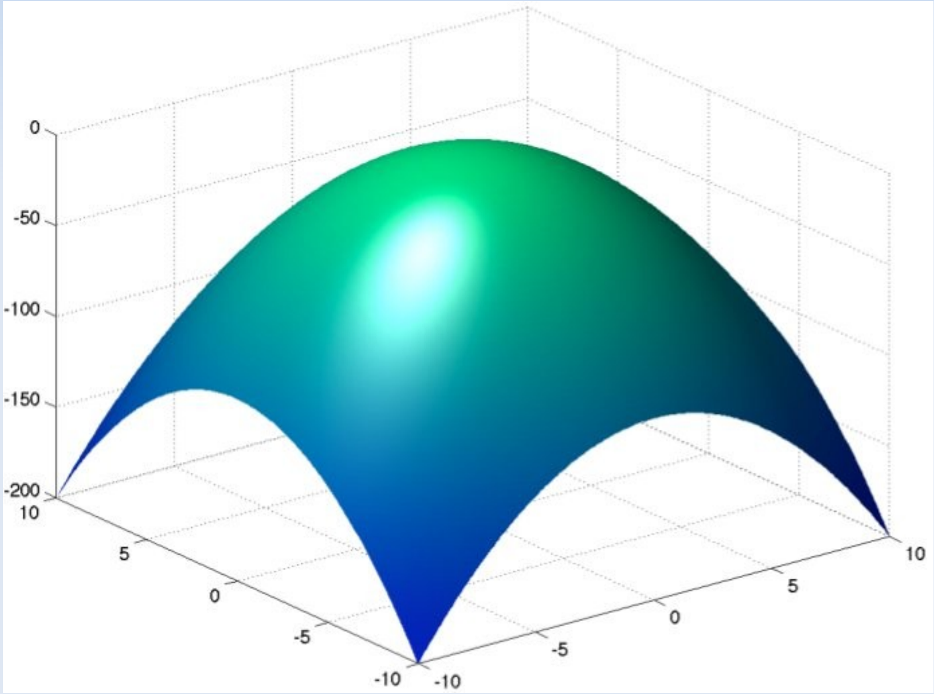
- La función objetivo es análogo al enunciado a satisfacer.
- Se han aplicado algoritmos de optimización como resolvedores SAT.
  - Algoritmos genéticos (ya están en forma binaria).
- Pero, hay una diferencia importante entre satisfacer un enunciado y optimizar una función.

Terminología

- Espacio de Solución:
  - Conjunto de series de valores que proveen una *solución posible* (como los modelos en el problema SAT) a la función objetivo.

Espacio de Solución Convexo

- De Jong (1975)
  - $f(X) = - \sum x^2$
  - Modificado para que sea máximo.
- Todo modelo se puede 'conectar' a otro por una línea recta.
- La función más fácil de optimizar.



¿Por qué es la más fácil?

Algoritmo de Optimización 1: Steepest Descent

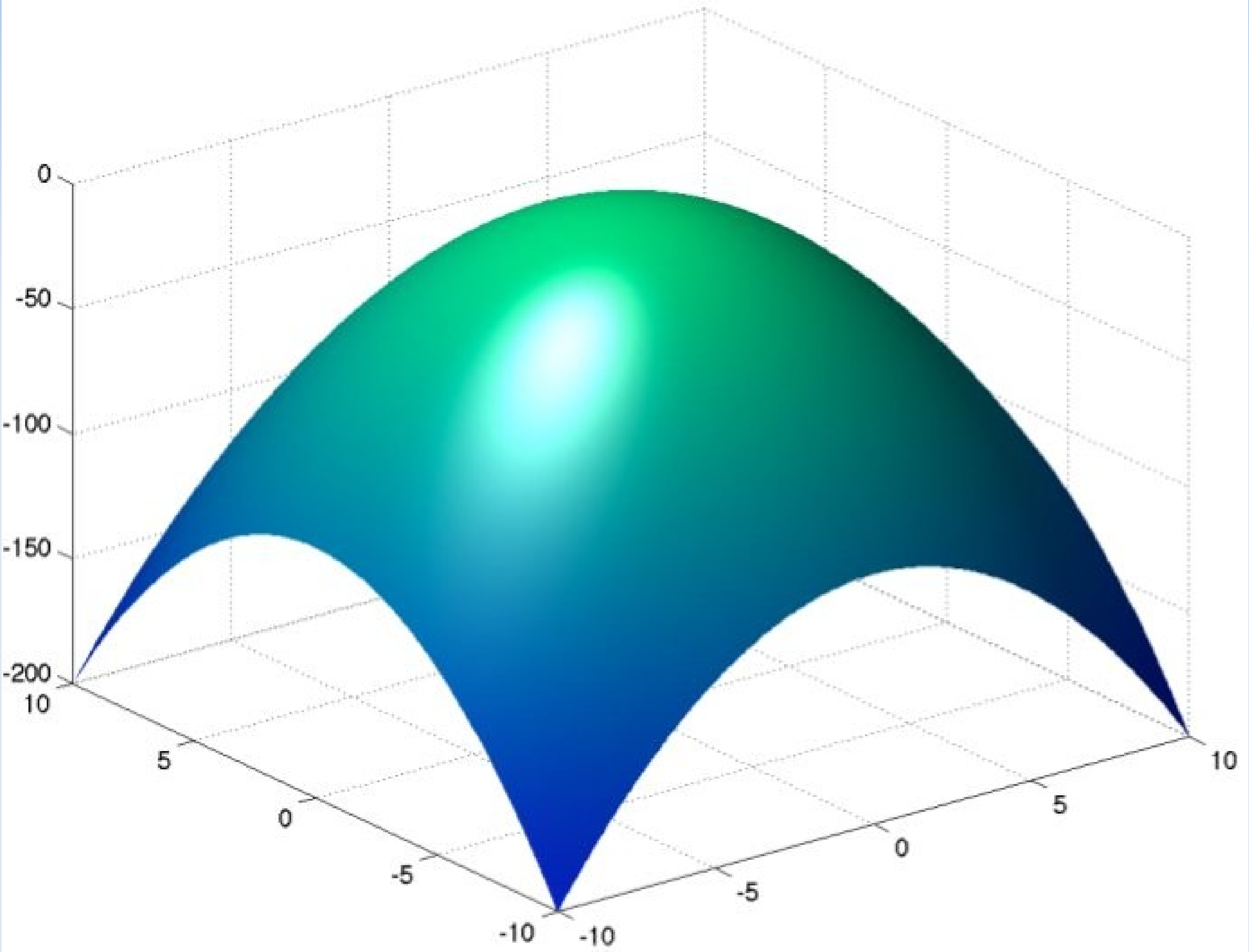
- A.k.a. Descenso por **Gradiente** (Gauss)
  - En caso de maximizar, es Ascenso
- Algoritmo:
  - Se propone una solución X_i inicial al azar.
  - Se actualiza a X_i con:
    - $X_{i+1} \leftarrow X_i - a\nabla f(X_i)$
      - a es un número real pequeño (a << 0).
        - En caso de ascenso, a es negativo.
      - $\nabla f(X_i)$ es la gradiente de f aplicando la solución X_i
  - Hasta X_i **converja**.

Gradiente

- Indica la dirección del mayor incremento de cambio en una función.
- ∇f es un vector de derivadas parciales:
  - $\partial f/\partial x_1, \partial f/\partial x_2, \partial f/\partial x_3, \ldots \partial f/\partial x_n$
- Por lo tanto, $\nabla f(X_i)$:
  - $\partial f(x_{i,1})/\partial x_1, \partial f(x_{i,2})/\partial x_2, \partial f(x_{i,3})/\partial x_3, \ldots \partial f(x_{i,n})/\partial x_n$
- $|\nabla f(X_i)|$ es la magnitud del vector.

Ejemplo

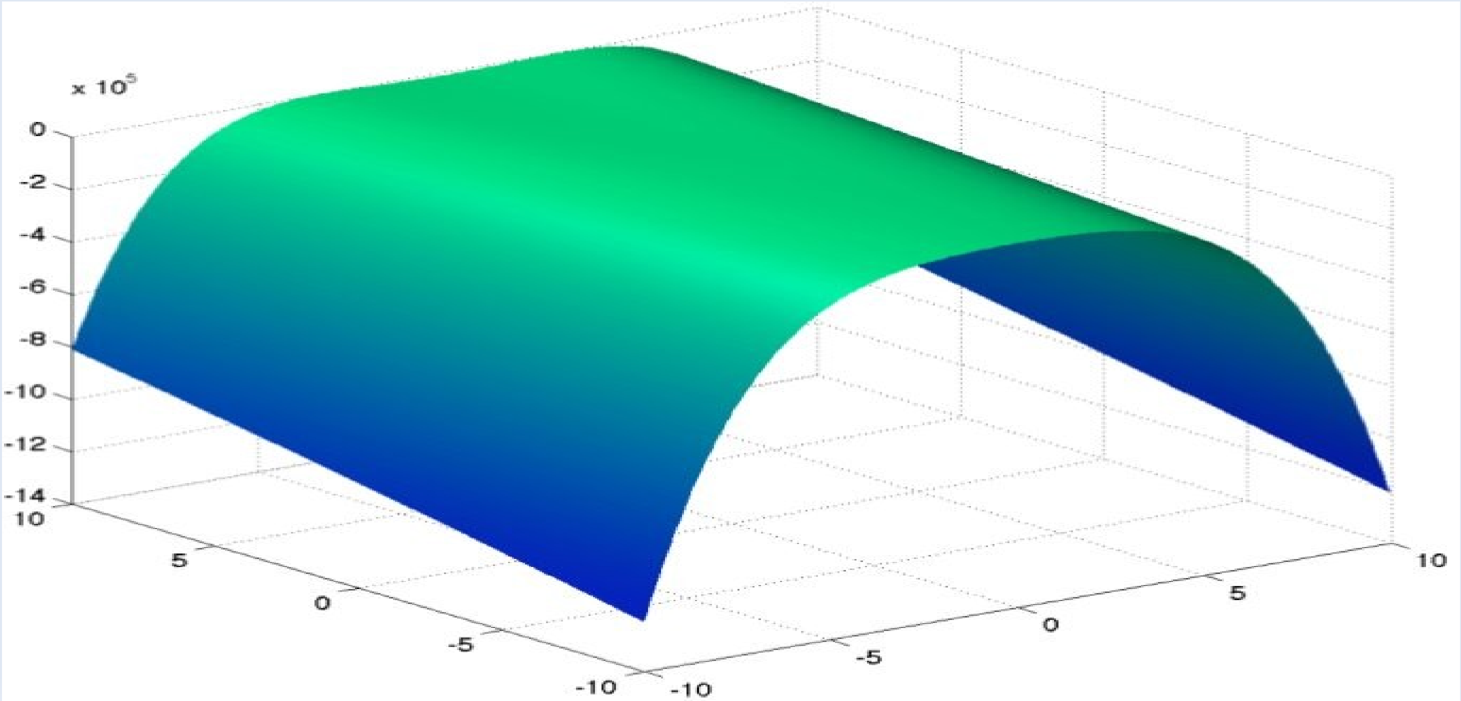
$f(x,y) = -x^2-y^2$



Adelantémonos un poco...

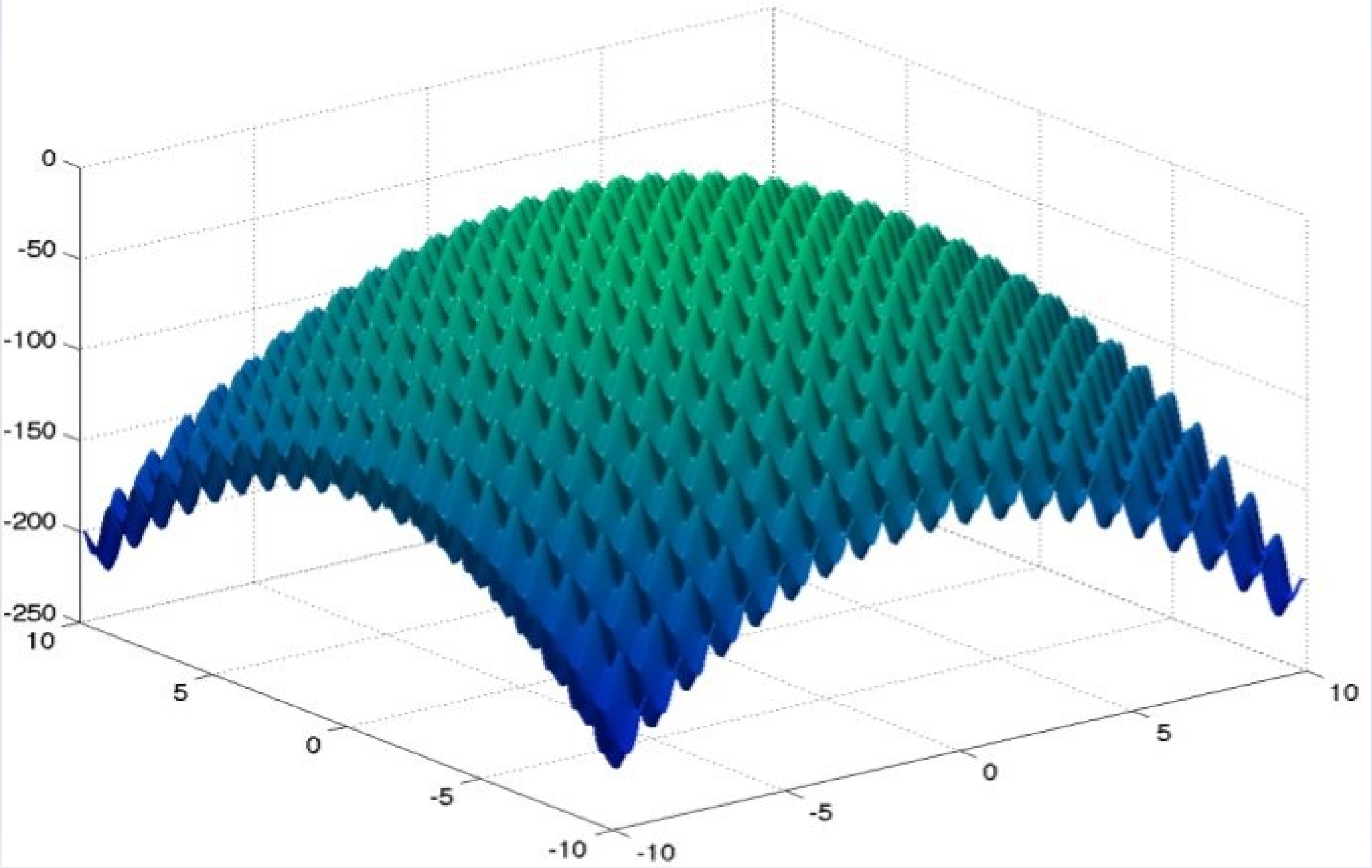
Otro Espacio de Solución: Rosenbrock

- $f(X) = -\sum 100(x_{d-1} - x_d^{\ 2})^2 + (1-x_d)^2$
  - d es el índice de variable dentro de X
- Diseñado para 'engañar' al algoritmo a entrar a un área de *óptimos locales* con el *óptimo global* escondido.



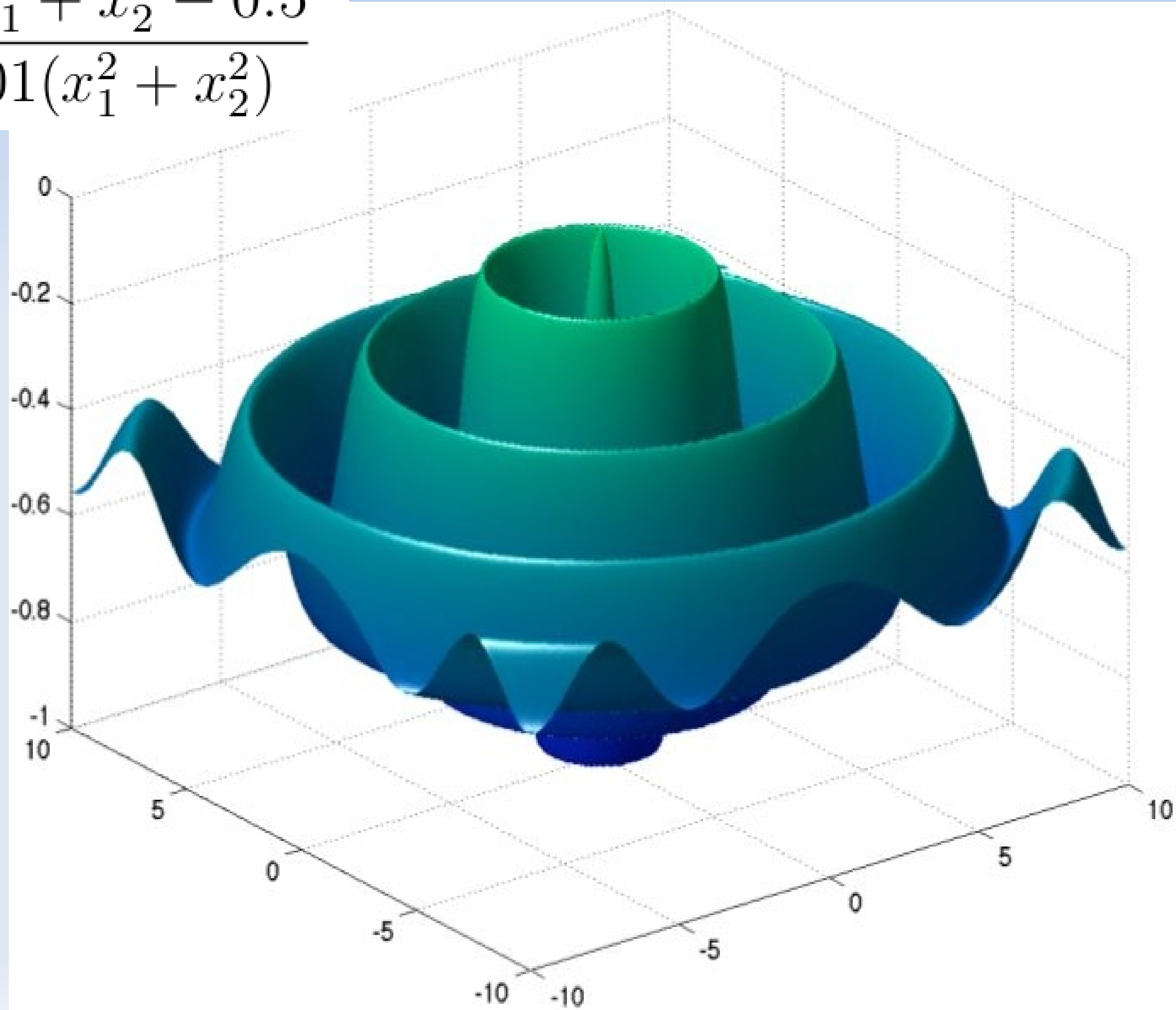
Otro Espacio de Solución: Rastrigin

- $f(X) = -10D - \sum x_d^2 - 10\cos(2\pi x_d)$
  - D es el número de variables
- Como De Jong, pero con más 'topes'.



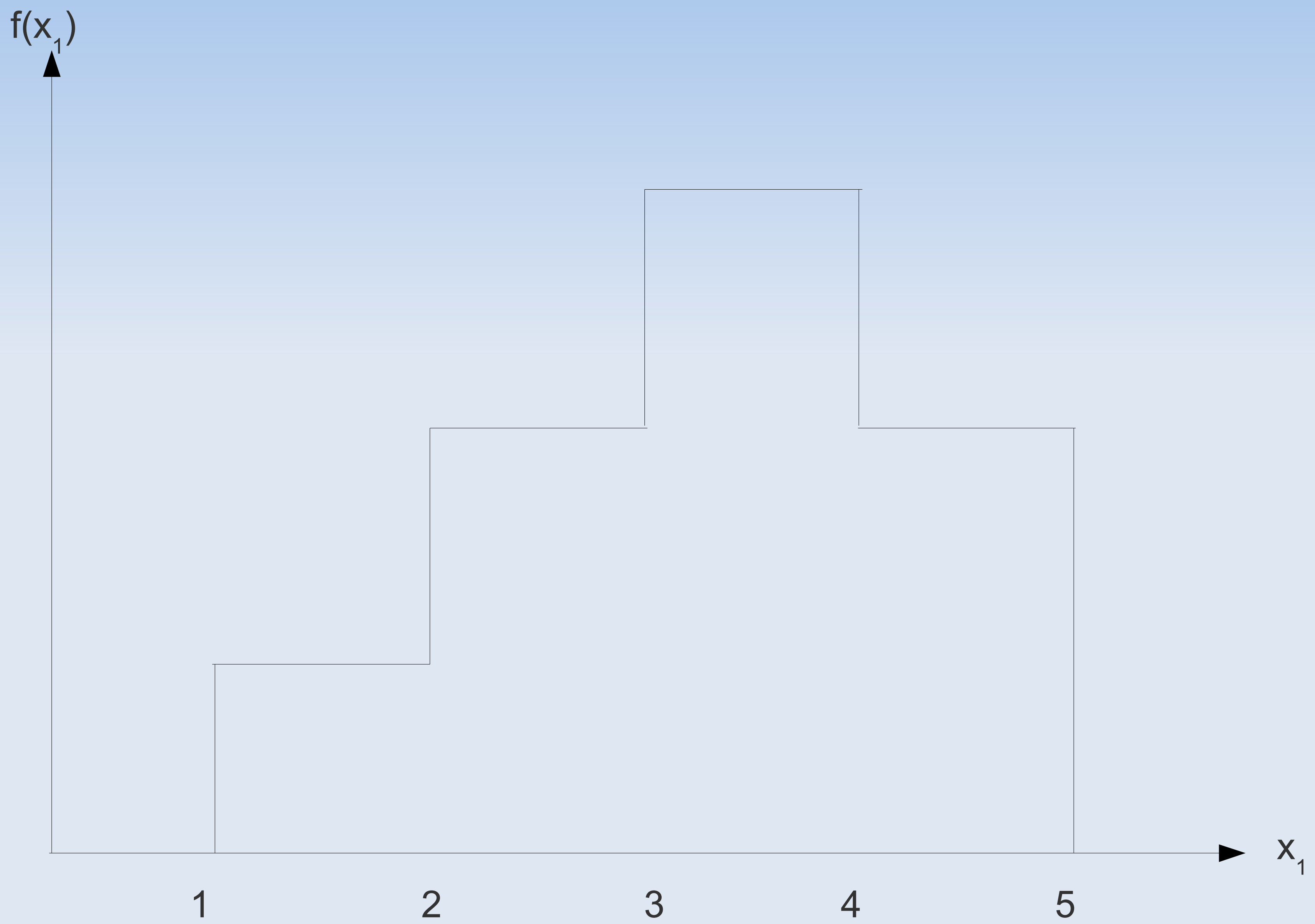
Otro Espacio de Solución: Schaffer F6

$$Z = -0.5 - \frac{\sin^2\sqrt{x_1^2 + x_2^2} - 0.5}{1 + 0.01(x_1^2 + x_2^2)}$$



¿Problemas con Descenso Gradiente?

¿Que tal con éste?



Entonces...

- No funciona en funciones no derivables (no contínuas, aka discretas).
- Se atasca en óptimos locales fácilmente.
- Se **alenta** al estar cerca de algún óptimo.
- El tiempo computacional para el cálculo de la gradiente es exponencial al número de variables.
-
- Pero... **garantiza** resultado correcto en un espacio de solución convexo.

1er Problema: Funciones Discretas

- Hay tal cosa como Gradiente Discreta:
  - Dado un estado x_i, busca el estado *vecino* que da el mayor valor a la función, y es mayor a x_i.
- Un *vecino* de x_i es el estado en el que todas las variables de x_i sólo cambiaron de índice uno o menos.
  - Si x_i = {1,2,3}, x_{i+1} puede ser:
    - {0,2,3} vecino
    - {2,1,2} vecino
    - {4,2,3} no es vecino

Algoritmo de Optimización 2: Hill Climbing

- Es Ascenso por Gradiente, pero en forma discreta.
- La búsqueda del vecino de mayor valor también es exponencial por número de variables.
  - Todas las posibles *combinaciones* de valores en diferencia a 1 de los valores actuales.
- Esta búsqueda de combinaciones es la razón por la que a los problemas de optimización con variables discretas se les conoce como **Optimización Combinatoria**.

2do Problema:
Óptimos Locales

- Cada vez que se crea que se ha llegado a un óptimo, se 'desatasca' al algoritmo cambiando el estado utilizando alguna *regla de cambio*.

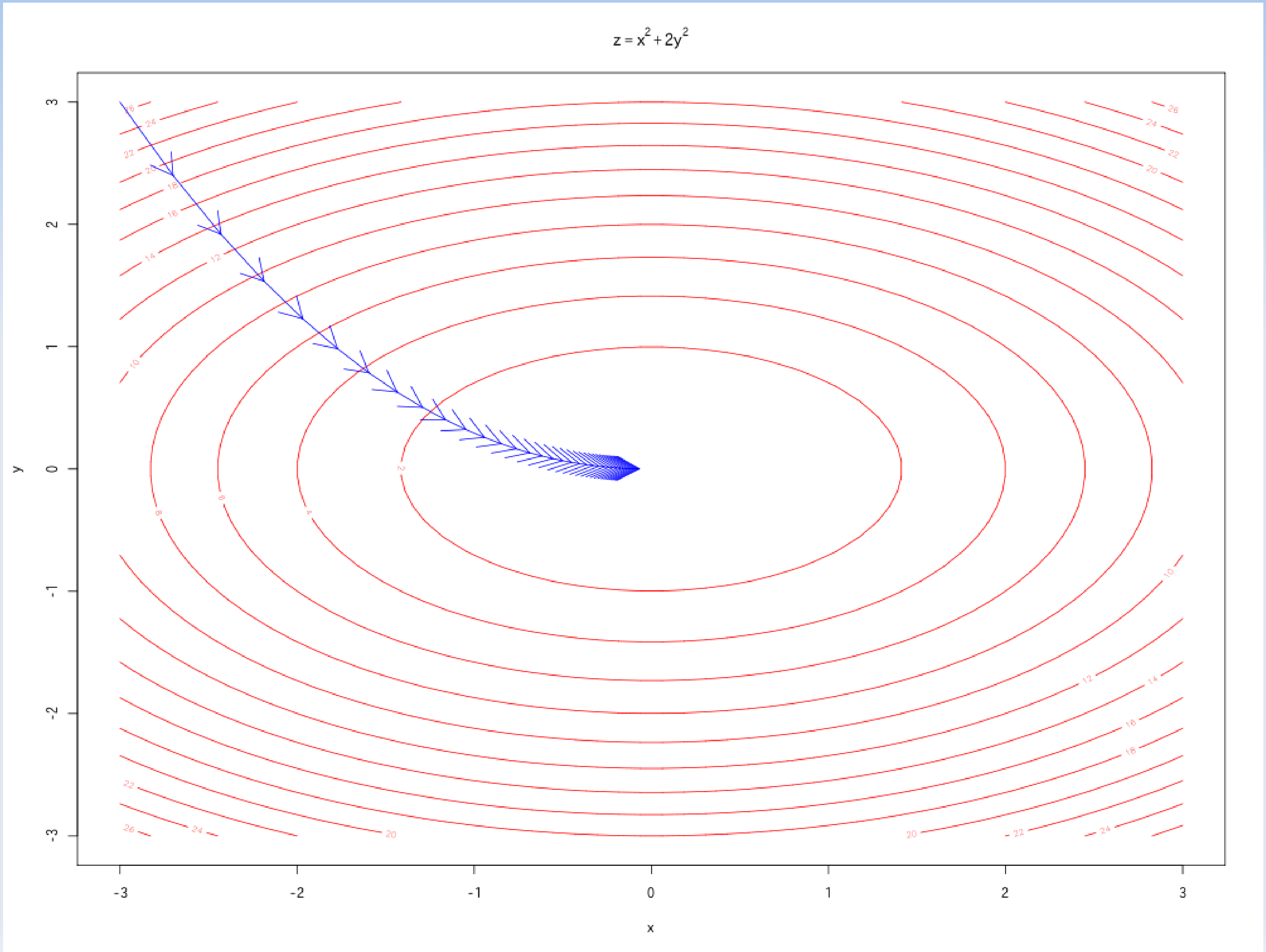
Algoritmo de Optimización 3: Templado Simulado

- A.k.a. Simulated Annealing.
- *Regla de Cambio*:
  - Se comienza con un monto de 'energía' inicial.
  - Al momento de cambio, se propone un nuevo estado *al azar,* dentro de un rango de valores definido por el monto de 'energía' actual.
  - En cada cambio, la 'energía' se disminuye.
  - La búsqueda termina cuando ya no haya 'energía'.
- Inspirado por el proceso de endurecimiento de acero.

Templado Simulado

- El monto de 'energía' tiene que ser grande para garantizar convergencia en un óptimo global.
  - Pero también aumenta el tiempo de convergencia.

3er Problema: Lento al Llegar a Óptimo



Entonces...

- Hacer más grande al parámetro a de:
  - $X_{i+1} \leftarrow X_i - a\nabla f(X_i)$
  - No tan grande para que no *sobredispare* el óptimo.
- O hacerlo dinámico:
  - Incrementarlo cuando $|\nabla f(X_i)|$ se hace pequeño.
- O reemplazarlo completamente...

Método Newton-Raphson

- Asume que se ha llegado al óptimo global cuando: $|\nabla f(X_i)| = 0$.
  - El problema es encontrar las raíces de $\nabla f(X_i)$.
- Encuentra la segunda derivada:
  - $H_f = \partial^2 f/\partial x_1 \partial x_2 \ldots \partial x_D$
- Así la actualización es:
  - $X_{i+1} \leftarrow X_i - H_f^{-1}(X_i)\nabla f(X_i)$
- Y **converge** cuando $|\nabla f(X_i)| = 0$.

Método Newton-Raphson

- Es lento... muy lento.
  - Sigue siendo exponencial, pero a mayor grado:
    - Cálculo de segundas derivadas, además de primeras derivadas.
- De nuevo, la función tiene que ser derivable.
  - Tiene versión discreta, pero requiere de aún más tiempo que la versión contínua.

Otras Opciones para el Cálculo del 'Siguiente Paso'

- Es en sí un problema de optimización.
  - Encontrar el 'paso óptimo' dado x_i y f(X) que asegure que nos lleve más cerca a la convergencia, en los menos pasos posibles.
- Una opción popular es el uso de las condiciones Armijo-Wolfe:
  - El nuevo paso asegura una disminución 'suficiente' de f(X) en una dirección en el que el gradiente se disminuye 'suficientemente'.

¿Los tres problemas fueron resueltos?

¿Se requieren resolver?

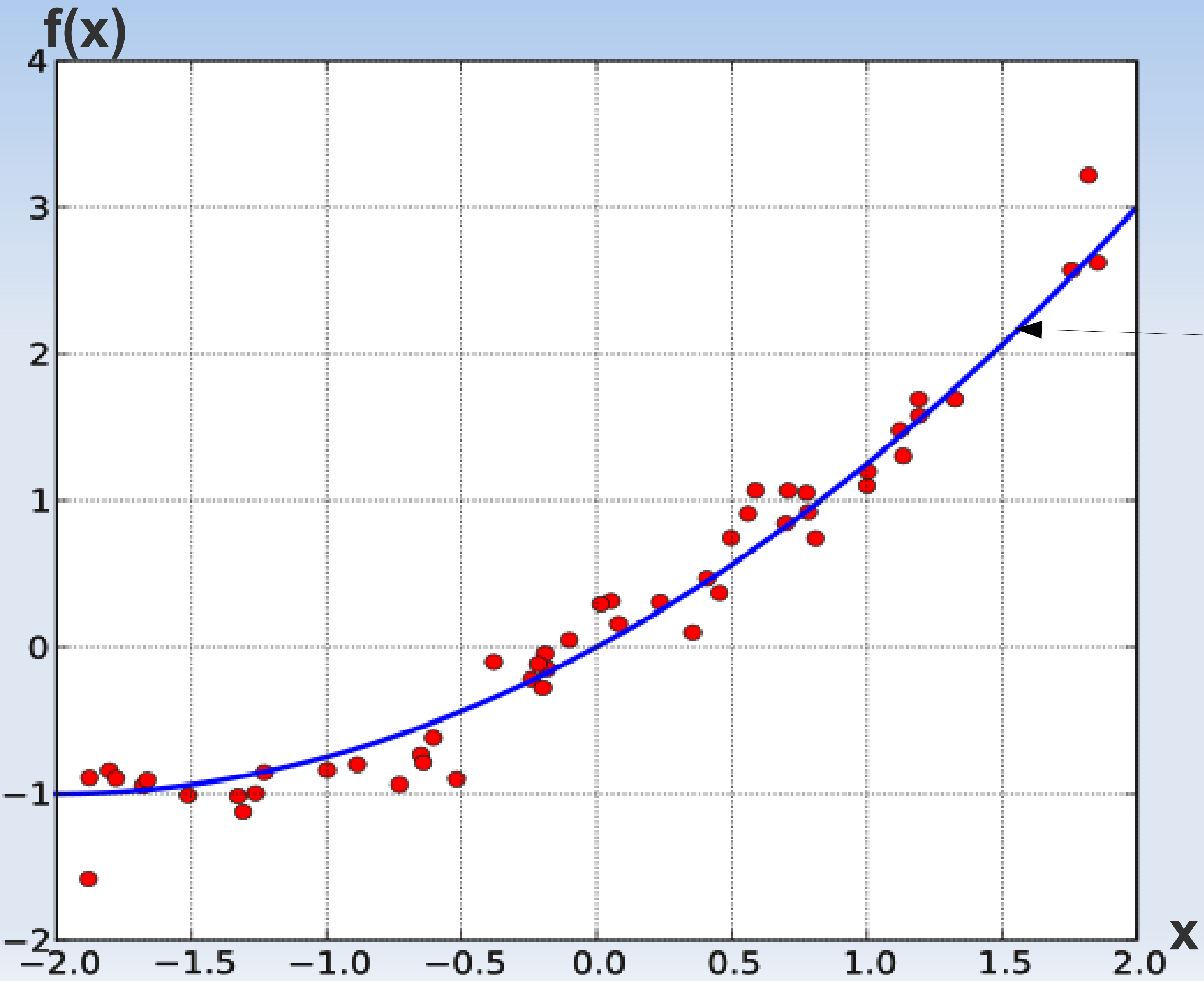
1a Lección de Optimización

- Conoce el espacio de solución.
  - ¿Es convexo?
  - Si no, ¿lo podemos hacer convexo?

Ejemplo: Mínimos Cuadrados

- Se tiene una serie de puntos, a los cuales se les quiere encontrar un modelo que los describa.
- Se propone un orden de modelo, y se requiere que se encuentre los coeficientes del modelo con menor error al describir los puntos.

Mínimos Cuadrados



Se propuso un orden de 2o grado:

$y(x) = Ax^2 + Bx + C$

Funciones Objetivo Posibles

- Donde
  - (x_i,y_i) es un punto
  - $\hat{y}_i[x_i,A,B,C]$ es el estimado de y_i del modelo utilizando los coeficientes A, B, y C
- 1a alternativa: $f(A,B,C) = \sum|y_i-\hat{y}_i[x_i,A,B,C]|$
  - ¿Proporciona un espacio de solución convexo?

Funciones Objetivo Posibles

- 2a alternativa: $f(A,B,C) = \sum(y_i-\hat{y}_i[x_i,A,B,C])^2$
  - ¿Proporciona un espacio de solución convexo?
- De hecho, utilizando esta función objetivo, se ha encontrado que la solución es única:
  - **(X'X)B=X'Y**
  - X es el vector de x_i, Y el de y_i.
  - B es el conjunto de coeficientes óptimos.
  - Cómo se llega a esa ecuación es tema de otro curso, (Wikipedia es su amigo).

Por cierto...

¿Cuál fue la diferencia importante entre optimizadores y resolvedores SAT?

Siguiente Clase

- Hablaremos de:
  - Convergencia
  - Clasificación (informal) de Algoritmos de Optimización