Motivation for Gradient Methods

• We have the closed-form information of $f(x)$
  – How to utilize this information

• What information can we get?
  – Why is gradient so important?
  – Easy to find the descending direction locally
  – Hard globally
The Rastrigin Function Terrain

Global minimum at [0 0]
Idea of Gradient Methods

• Form an approximation to the function
  – Minimize the approximated function
• Two ingredients:
  – Direction
  – Stepsize
If $\nabla f(x) \neq 0$, there is an interval $(0, \delta)$ of stepsizes such that

$$f\left(x - \alpha \nabla f(x)\right) < f(x)$$

for all $\alpha \in (0, \delta)$.

If $d$ makes an angle with $\nabla f(x)$ that is greater than 90 degrees,

$$\nabla f(x)'d < 0,$$

there is an interval $(0, \delta)$ of stepsizes such that $f(x + \alpha d) < f(x)$ for all $\alpha \in (0, \delta)$. 
Steepest Descent Example

• Minimize

\[ X^2 + 100 \, y^2 \]

Where \(x, y\) are real numbers
PRINCIPAL GRADIENT METHODS

\[ x^{k+1} = x^k + \alpha^k d^k, \quad k = 0, 1, \ldots \]

where, if \( \nabla f(x^k) \neq 0 \), the direction \( d^k \) satisfies

\[ \nabla f(x^k)' d^k < 0, \]

and \( \alpha^k \) is a positive stepsize. Principal example:

\[ x^{k+1} = x^k - \alpha^k D^k \nabla f(x^k), \]

where \( D^k \) is a positive definite symmetric matrix

- Simplest method: Steepest descent

\[ x^{k+1} = x^k - \alpha^k \nabla f(x^k), \quad k = 0, 1, \ldots \]

- Most sophisticated method: Newton’s method

\[ x^{k+1} = x^k - \alpha^k \left( \nabla^2 f(x^k) \right)^{-1} \nabla f(x^k), \quad k = 0, 1, \ldots \]
Newton’s Method Example

• Minimize

\[ X^2 + 100 \, y^2 \]

Where \( x, y \) are real numbers
SLOW CONVERGENCE OF STEEPEST DESCENT

Fast convergence of Newton’s method w/ $\alpha^k = 1$.

Given $x^k$, the method obtains $x^{k+1}$ as the minimum of a quadratic approximation of $f$ based on a second order Taylor expansion around $x^k$. 

STEEPEST DESCENT AND NEWTON’S METHOD
CHOICES OF STEPSIZE I

- Minimization Rule: $\alpha^k$ is such that

$$f(x^k + \alpha^k d^k) = \min_{\alpha \geq 0} f(x^k + \alpha d^k).$$

- Limited Minimization Rule: Min over $\alpha \in [0, s]$

- Armijo rule:

Start with $s$ and continue with $\beta s, \beta^2 s, \ldots$, until $\beta^m s$ falls within the set of $\alpha$ with

$$f(x^k) - f(x^k + \alpha d^k) \geq -\sigma \alpha \nabla f(x^k)' d^k.$$
CHOICES OF STEPSIZE II

• Constant stepsize: $\alpha^k$ is such that

$$\alpha^k = s : \text{a constant}$$

• Diminishing stepsize:

$$\alpha^k \to 0$$

but satisfies the infinite travel condition

$$\sum_{k=0}^{\infty} \alpha^k = \infty$$