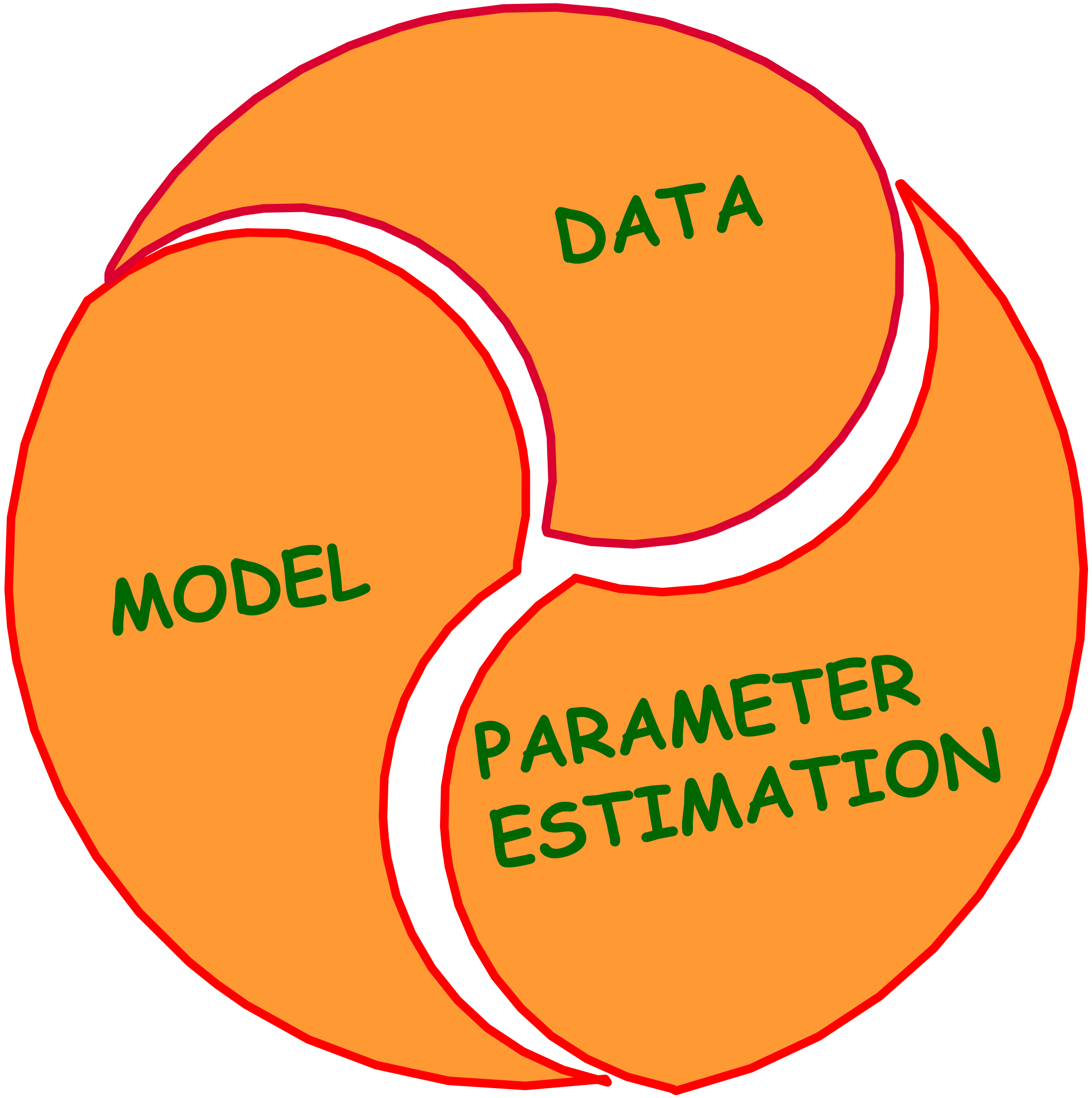
Optimization 101

Mathematical Preliminaries

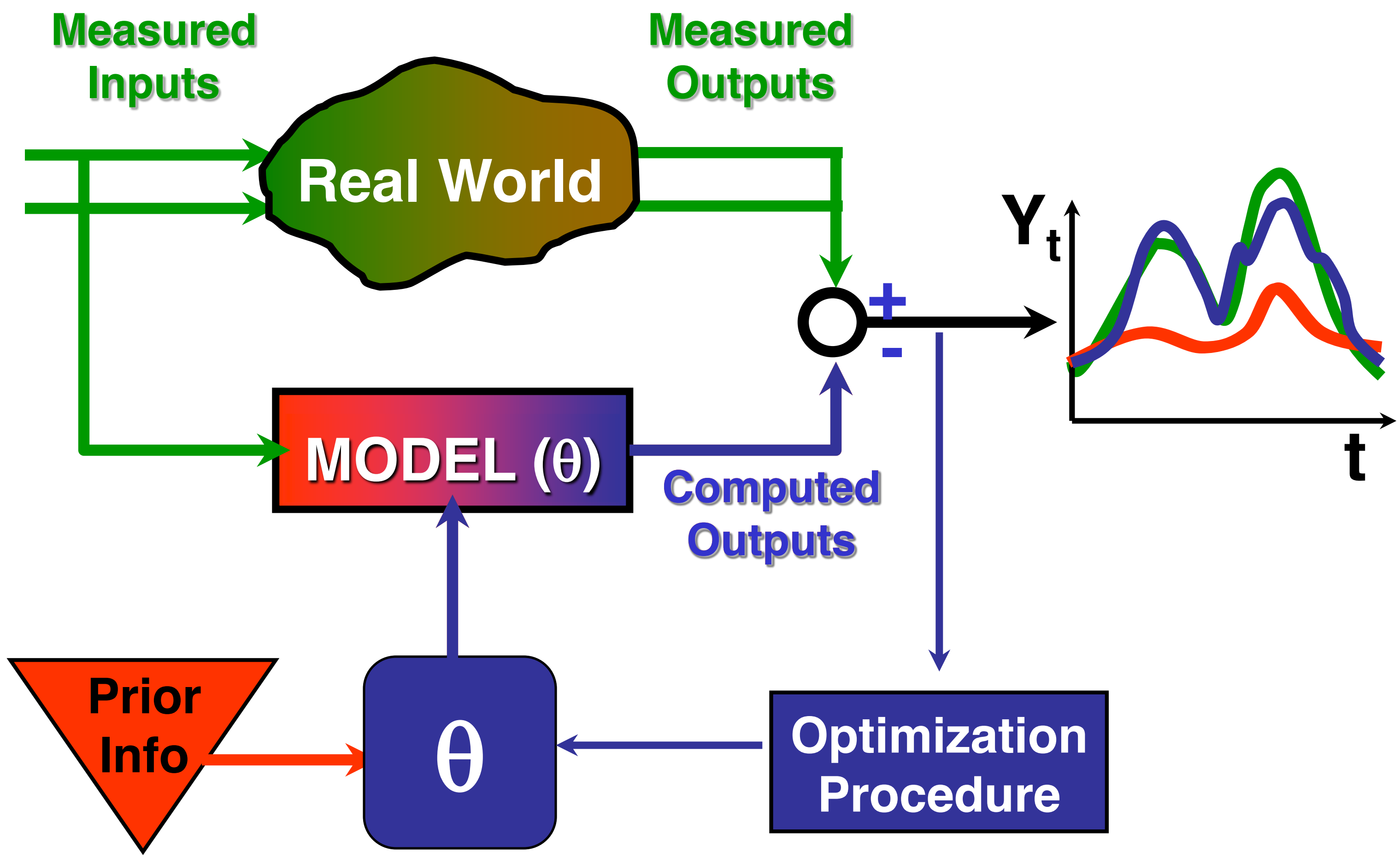
Dr. Nagiza F. Samatova
Department of Computer Science
North Carolina State University
and
Computer Science and Mathematics Division
Oak Ridge National Laboratory

Model Construction



Src: Soroosh Sorooshian, UC Irvine

The Concept of Model Calibration



"Calibration: constraining the model to be consistent with observations"

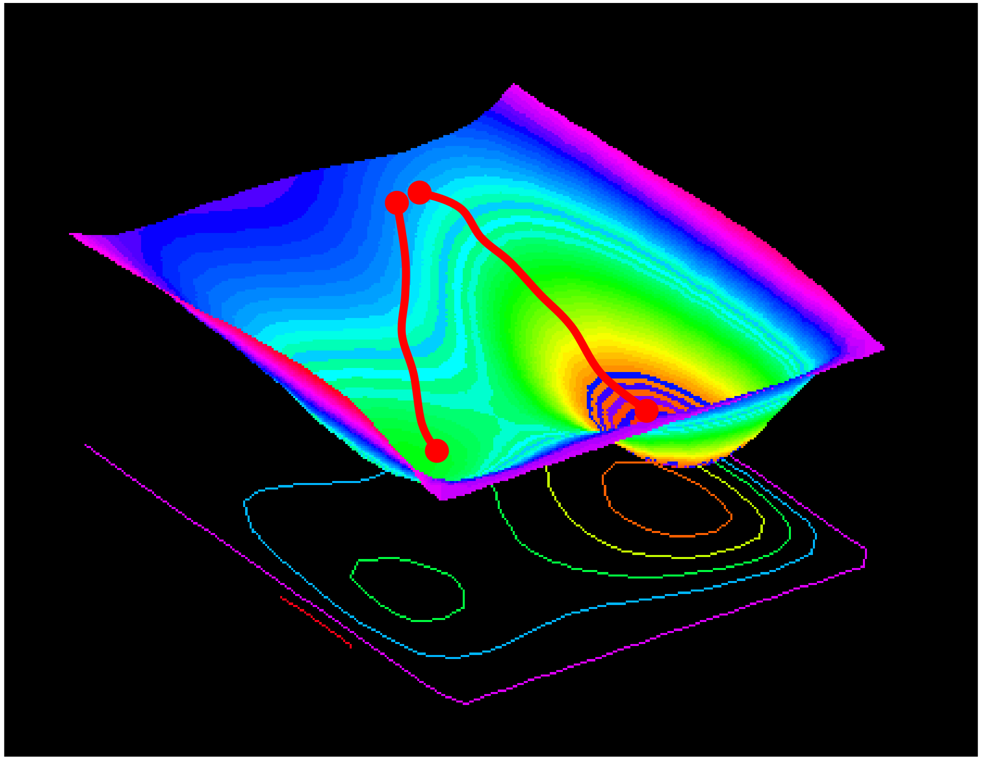
Src: Soroosh Sorooshian, UC Irvine

Calibration Components

Objective Function

Search Algorithm

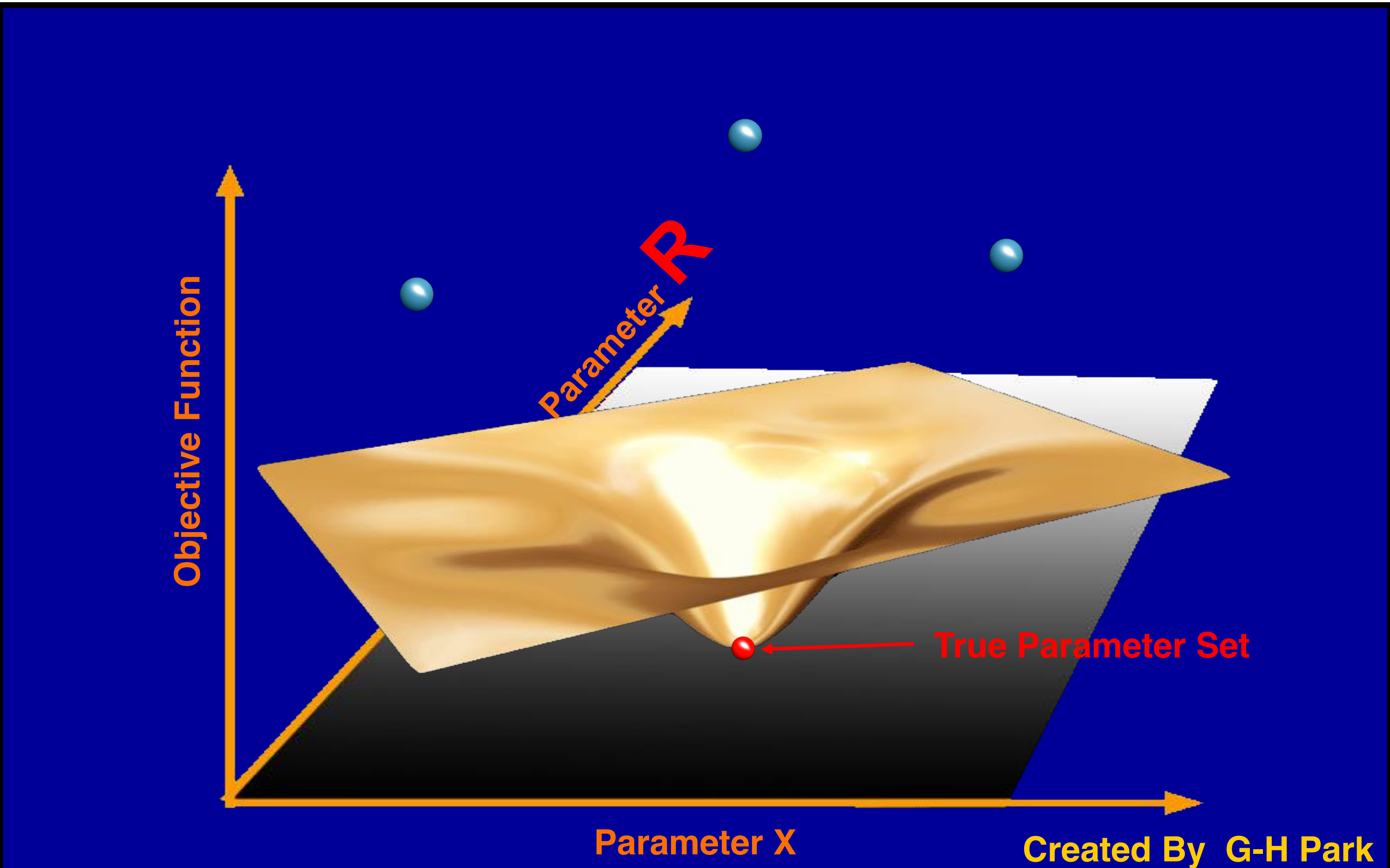
Sensitivity Analysis



Problems with identifiability

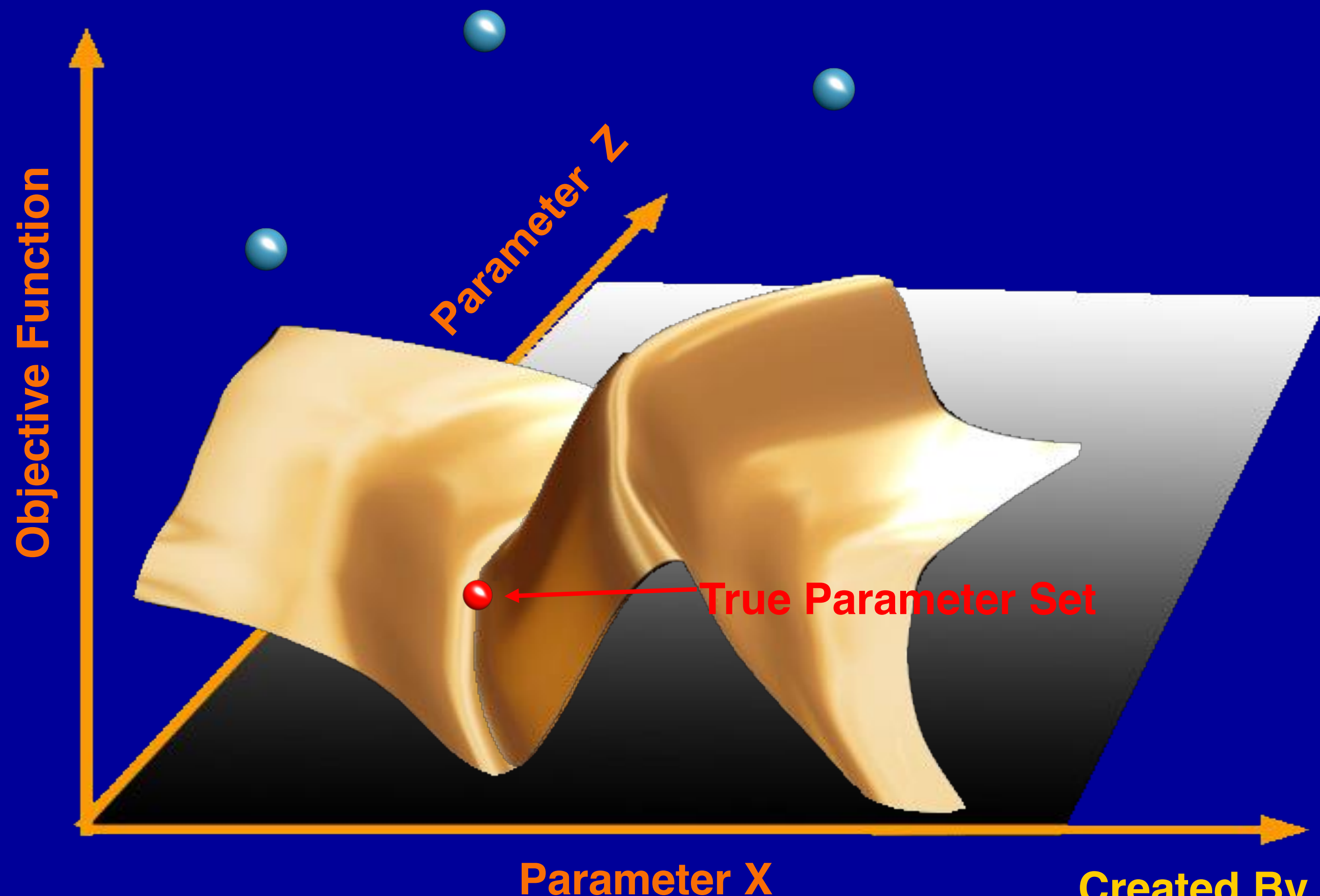
Src: Soroosh Sorooshian, UC Irvine

The Ideal case: Convex Optimization



Src: Soroosh Sorooshian, UC Irvine

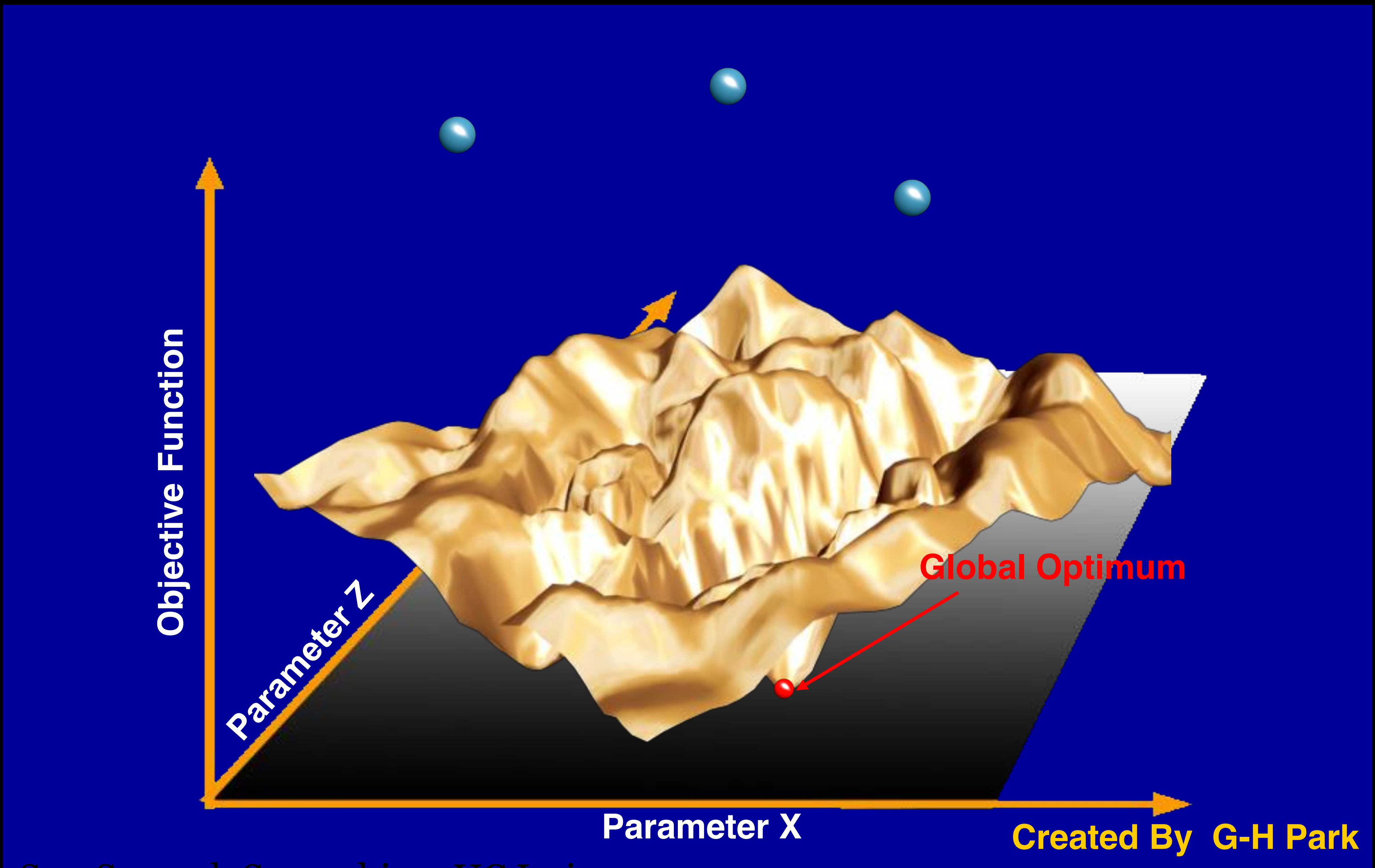
Difficulties in Global Optimization



Created By G-H Park

Src: Soroosh Sorooshian, UC Irvine

Parameter Estimation
(non-convex, multi-optima)

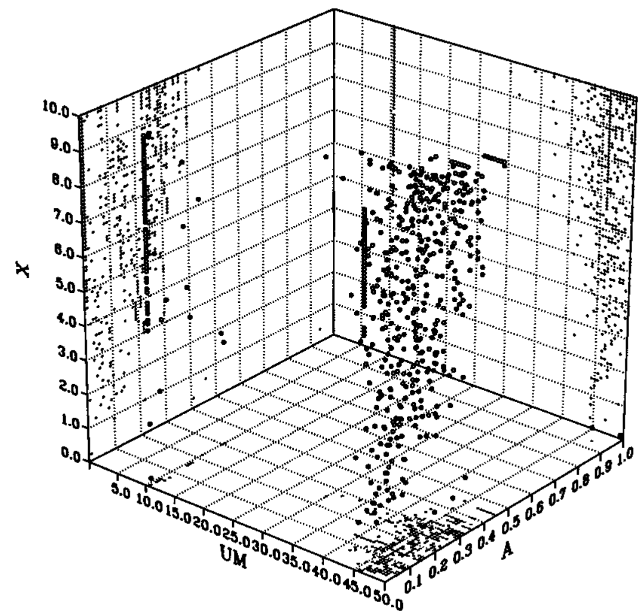


Src: Soroosh Sorooshian, UC Irvine

Difficulties in Optimization

1.- Regions of Attraction — *More than one main convergence region*

2.- Local Optima — *Many small "pits" in each region*

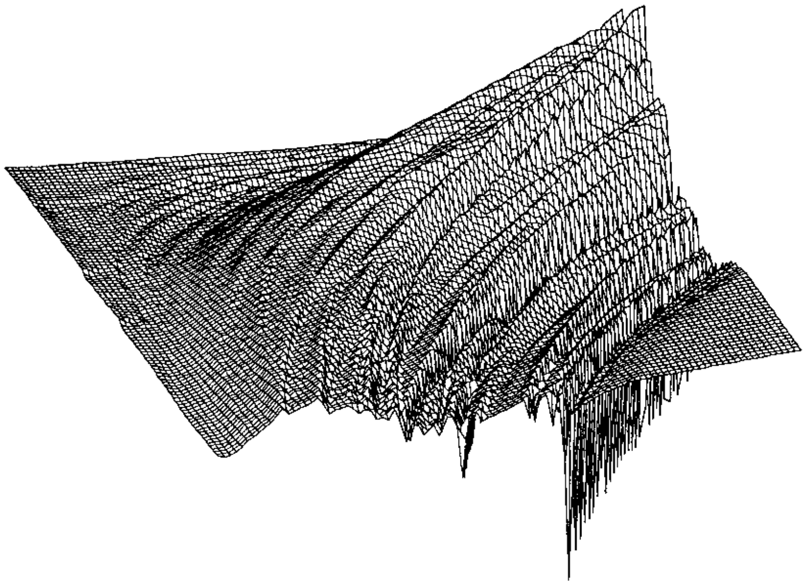


Duan, Gupta, and Sorooshian, 1992, WRR

Src: Soroosh Sorooshian, UC Irvine

Difficulties in Optimization

1.- Regions of Attraction	*More than one main convergence region*
2.- Local Optima	*Many small "pits" in each region*
3.- Roughness	*Rough surface with discontinuous derivatives*

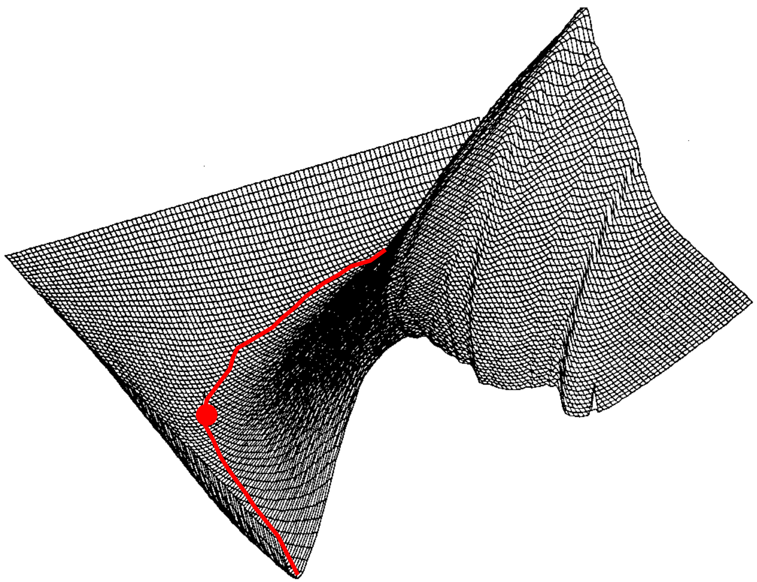


Duan, Gupta, and Sorooshian, 1992, WRR

Src: Soroosh Sorooshian, UC Irvine

Difficulties in Optimization

1.- Regions of Attraction	*More than one main convergence region*
2.- Local Optima	*Many small "pits" in each region*
3.- Roughness	*Rough surface with discontinuous derivatives*
4.- Flatness	*Flat near optimum with significantly different parameter sensitivities*
5.- Shape	*Long and curved ridges*



Duan, Gupta, and Sorooshian, 1992, WRR

Src: Soroosh Sorooshian, UC Irvine

Mathematical Optimization Subfields

- **Linear programming**
- **Integer programming**
- **Quadratic programming**
- **Nonlinear programming**
- **Convex programming**
- **Semidefinite programming**
- **Stochastic programming**
- **Combinatorial optimization**
- **Dynamic programming**
- **....**

Optimization Problem: General Formulation

GIVEN:

Objective function: $f : \mathbb{R}^d \to \mathbb{R}$

cost function, energy function

$$f(x) = f(x_1, x_2, \ldots, x_d)$$

Subject to constraints:

Equality constraints: $g_i(x) = 0, \; i = 1, 2, \ldots, p$

and/or

Inequality constraints: $h_i(x) \le 0, \; i = 1, 2, \ldots, q$

SOLVE:

Minimization problem: $\boldsymbol{f(x_{min}) = \min_{x \in \mathbb{R}^d} f(x_1, x_2, \ldots, x_d)}$

Maximization problem: $\boldsymbol{f(x_{max}) = \max_{x \in \mathbb{R}^d} f(x_1, x_2, \ldots, x_d)}$

Solving Strategies

Optimization

Constrained

Equality

Lagrange Method

Inequality

Lagrange Method
w/ bounded Lagrangian
multipliers

Unconstrained

1. Stationary points, x^*:

$$\left.\frac{df}{dx}\right|_{x=x^*} = 0$$

2. Min. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} < 0$$

3. Max. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} > 0$$

4. Saddle stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} = 0$$

Solving Strategies

Optimization

Constrained

Equality

Lagrange Method

Inequality

Lagrange Method w/ bounded Lagrangian multipliers

Unconstrained

1. Stationary points, x^*:

$$\left.\frac{df}{dx}\right|_{x=x^*} = 0$$

2. Min. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} < 0$$

3. Max. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} > 0$$

4. Saddle stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} = 0$$

Unconstrained Optimization, d=1

$$\boldsymbol{f}\colon \mathbb{R}^{\boldsymbol{d}} \to \mathbb{R}$$
$$\boldsymbol{f}(\boldsymbol{x}) \to \boldsymbol{min} \textbf{ or } \boldsymbol{f}(\boldsymbol{x}) \to \boldsymbol{max}$$

1. **Stationary** points, x^*:

$$\left.\frac{df}{dx}\right|_{x=x^*} = 0$$

2. **Min**. stationary points, x^*:

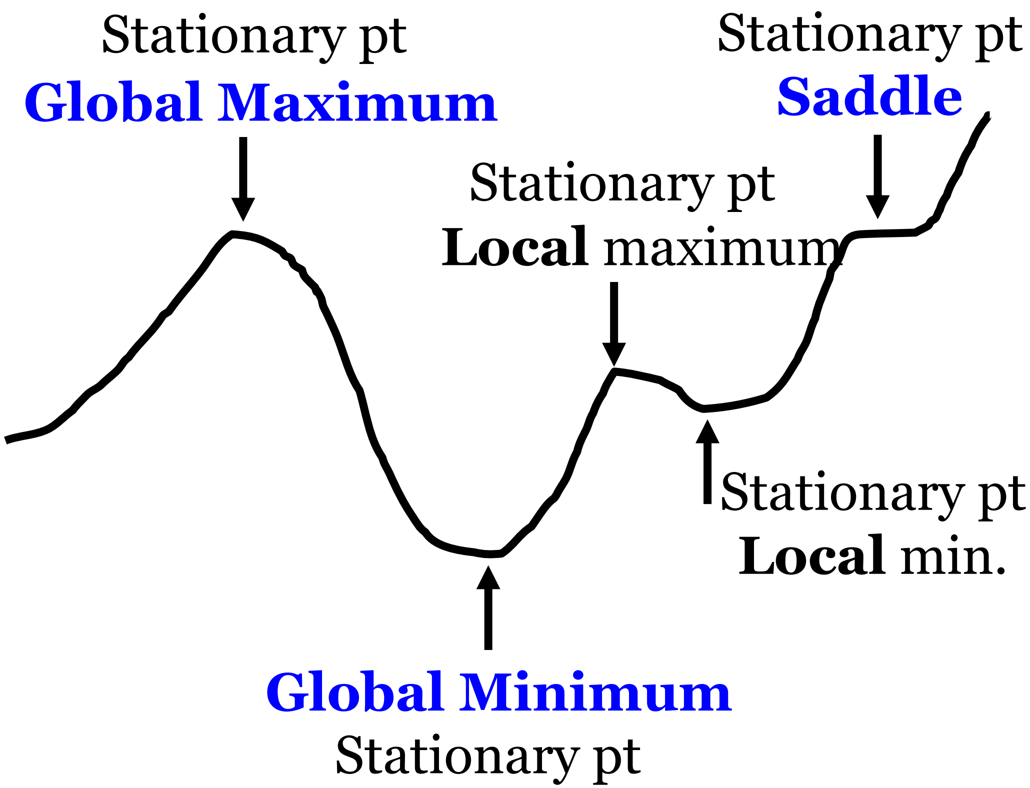
$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} < 0$$

3. **Max**. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} > 0$$

4. **Saddle** stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} = 0$$



Unconstrained Multivariate Optimization, d>1

$$\boldsymbol{f}: \mathbb{R}^{d} \to \mathbb{R}, \boldsymbol{d} > 1$$
$$\boldsymbol{f}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d) \to \boldsymbol{min} \text{ or}$$
$$\boldsymbol{f}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d) \to \boldsymbol{max}$$

$$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \frac{\partial^2 f}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_d} \end{bmatrix}$$

Hessian matrix

1. **Stationary** points, x^*:

$$\left.\frac{\partial f}{\partial x_j}\right|_{x_j = x_j^*} = 0, \ \forall j = 1, 2, \ldots, d \longleftarrow \text{partial derivatives}$$

2. **Hessian matrix**, $\boldsymbol{H}(x)$

3. **Min**. stationary points, x^*:

$$\left. x^T H(x) x \right|_{x = x^*} > 0 \longleftarrow \text{positive definite } \boldsymbol{H}$$

4. **Max**. stationary points, x^*:

$$\left. x^T H(x) x \right|_{x = x^*} < 0$$

5. **Saddle** stationary points, x^*:

$\longleftarrow$ indefinite $\boldsymbol{H}$

Ex: Unconstrained Multivariate Case, d=2

GIVEN:

Objective function: $f: \mathbb{R}^2 \to \mathbb{R}, d = 2$

$$\boxed{f(x, y) = 3x^2 + 2y^3 - 2xy}$$

Subject to constraints:

Equality constraints: *None*

and/or

Inequality constraints: *None*

SOLVE:

Minimization problem: $\boldsymbol{f(x_{min}, y_{min}) = \min_{(x,y)\in\mathbb{R}^2} f(x, y)}$

Maximization problem: $\boldsymbol{f(x_{max}, y_{max}) = \max_{(x,y)\in\mathbb{R}^2} f(x, y)}$

Ex: Hessian Matrix

$$f(x,y) = 3x^2 + 2y^3 - 2xy$$

1. **Stationary** points, p^*:

$$\left.\frac{\partial f}{\partial p_j}\right|_{p_j = p_j^*} = 0,\ \forall j = 1,2,\ldots,d \longleftarrow \text{partial derivatives}$$

$$H(x,y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x \partial x} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y \partial y} \end{bmatrix}$$

$$\begin{cases} \frac{\partial f}{\partial x} = 6x - 2y = 0 \\ \frac{\partial f}{\partial y} = 6y^2 - 2x = 0 \end{cases} \Rightarrow \begin{array}{l} p_1^* = (x_1^*, y_1^*) = (0,0) \\ p_2^* = (x_2^*, y_2^*) = (\frac{1}{27}, \frac{1}{9}) \end{array}$$

2. **Hessian matrix**, ***H***(x,y)

$$H(x,y) = \begin{bmatrix} 6 & -2 \\ -2 & 12y \end{bmatrix}$$

3. **Min**. stationary points, p^*:

$$\left. p^T H(p) p \right|_{p=p^*} > 0 \longleftarrow \text{positive definite } \boldsymbol{H}$$

$$\left.(x,y)^T \begin{bmatrix} 6 & -2 \\ -2 & 12\frac{1}{9} \end{bmatrix} (x,y) \right|_{(x,y)=(\frac{1}{27},\frac{1}{9})} = 4x^2 - 2xy + 4y^2/3 = 4(x - \frac{y}{4})^2 + 13\frac{y^2}{4} > 0 \Rightarrow p_{\min}^* = (\frac{1}{27}, \frac{1}{9})$$

Analytical vs. Numerical Solution

- **Analytical solutions**:
  - First and second derivatives exist
- **Numerical methods**:
  - Derivatives can NOT be solved analytically
  - Such cases are abundant

Classes of Numerical Methods

- **Based on smoothness of the objective function:**
  - First order methods
  - Second order methods
  - Combinatorial methods
  - Derivative-free methods
- **Actual methods:**
  - Newton's method
  - Gradient descent method (aka steepest descent/ascent)
  - Conjugate gradient method
  - Quasi-Newton method
  - Simplex method
  - Ellipsoid method
  - …

Newton's Method – Univariate Case

Based on **quadratic** approximation to the function *f*(x).

1. **Taylor** series expansion of *f()*: $f(x) \approx f(x_0) + (x - x_0) f'(x_0) + \frac{(x - x_0)^2}{2} f''(x_0)$

2. Set the first-derivative of *f*() to 0: $f'(x) = f'(x_0) + (x - x_0) f''(x_0) = 0$

3. Derive the update formula for x: $x = x_0 + \frac{f'(x_0)}{f''(x_0)}$

Newton's method – Algorithm

1: Let x_0 be the initial point

2: **while** $|f'(x_0)| > \varepsilon$ **do**

3: $x = x_0 + \dfrac{f'(x_0)}{f''(x_0)}$

4: $x_0 = x$

5: **end while**

6: return x

Newton's Method – Multivariate Case

Replace:

$$f'(x_1, x_2, \dots, x_d) \rightarrow \nabla f(x_1, x_2, \dots, x_d) = \\ = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_d})$$

gradient operator

$$f''(x_1, x_2, \dots, x_d) \rightarrow H(x_1, x_2, \dots, x_d)$$

Hessian matrix

Update:

$$x = x_0 - H^{-1} \nabla f(x) \big|_{x = x_0}$$

Solving Strategies

Optimization

Constrained | **Unconstrained**

Equality

Lagrange Method

Inequality

Lagrange Method
w/ bounded Lagrangian
multipliers

Unconstrained

1. Stationary points, x^*:

$$\left.\frac{df}{dx}\right|_{x=x^*} = 0$$

2. Min. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} < 0$$

3. Max. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} > 0$$

4. Saddle stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} = 0$$

Constrained Optimization, Equality Constraints

The Lagrange Method

$$f : R^d \to R,\ d \geq 1$$
$$f(x_1, x_2, \ldots, x_d) \to \min$$
$$\text{subject to: } g_i(x_1, x_2, \ldots, x_d) = 0,\ i = 1, 2, .., p$$

1. Define the **Lagrangian**: $L(x, \lambda) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x)$

Lagrangian multipliers

2. Set the first-derivatives to 0:

$$\frac{\partial L}{\partial x_j} = 0,\ \forall j = 1, 2, \ldots, d$$

$$\frac{\partial L}{\partial \lambda_i} = 0,\ \forall j = 1, 2, \ldots, p$$

3. Solve the (d+p) equations in Step 2
   to obtain stationary point x^* and
   the corresponding values for λ_i's

Example

Objective function: $f(x, y) = x + 2y \to \min$

Subject to the constraint: $g(x, y) = x^2 + y^2 - 4 = 0$

$$L(x, \lambda) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x)$$

$$\frac{\partial L}{\partial x_j} = 0, \ \forall j = 1, 2, \ldots, d$$

$$\frac{\partial L}{\partial \lambda_i} = 0, \ \forall j = 1, 2, \ldots, p$$

Solving Strategies

Optimization

Constrained

Equality

Lagrange Method

Inequality

Lagrange Method
w/ bounded Lagrangian
multipliers

Unconstrained

1. Stationary points, x^*:

$$\left.\frac{df}{dx}\right|_{x=x^*} = 0$$

2. Min. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} < 0$$

3. Max. stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} > 0$$

4. Saddle stationary points, x^*:

$$\left.\frac{d^2 f}{dx^2}\right|_{x=x^*} = 0$$

Constrained Optimization, Inequality Constraints

$$f : R^d \to R,\ d \geq 1$$
$$f(x_1, x_2, ..., x_d) \to \min$$
$$\text{subject to: } h_i(x_1, x_2, ..., x_d) \leq 0,\ i = 1, 2, .., q$$

1. Define the **Lagrangian**: $$L(x, \lambda) = f(x) + \sum_{i=1}^{q} \lambda_i h_i(x)$$

2. Set **Karush-Kuhn-Tucker** (KKT) conditions:

$$\frac{\partial L}{\partial x_j} = 0,\ \forall j = 1, 2, ..., d$$
$$h_i(x) \leq 0,\ \forall i = 1, 2, ..., q$$
$$\lambda_i \geq 0,\ \forall i = 1, 2, ..., q$$
$$\lambda_i h_i(x) = 0,\ \forall i = 1, 2, ..., q$$

The Lagrange multipliers are no longer unbounded in the presence of inequality constraints.

3. Solve the system in Step 2
to obtain stationary point x^* and
the corresponding values for λ_i's

Example

$f : R^d \to R,\ d \geq 1$

$f(x_1, x_2, ..., x_d) \to \min$

subject to: $h_i(x_1, x_2, ..., x_d) \leq 0,\ i = 1, 2, .., q$

Objective function: $f(x, y) = (x-1)^2 + (y-3)^2 \to \min$

Subject to the constraints: $h_1(x, y) = x + y - 2 \leq 0$

$h_2(x, y) = x - y \leq 0$

$$L(x, \lambda) = f(x) + \sum_{i=1}^{q} \lambda_i h_i(x)$$

$$\frac{\partial L}{\partial x_j} = 0,\ \forall j = 1, 2, ..., d$$

$$\lambda_i h_i(x) = 0,\ \forall i = 1, 2, ..., q$$

$$h_i(x) \leq 0,\ \forall i = 1, 2, ..., q$$

$$\lambda_i \geq 0,\ \forall i = 1, 2, ..., q$$