# CS 357: Numerical Methods

## Constrained Optimization

Eric Shaffer

# Optimization Problems

- Given function $f : \mathbb{R}^n \to \mathbb{R}$, and set $S \subseteq \mathbb{R}^n$, find $\boldsymbol{x}^* \in S$ such that $f(\boldsymbol{x}^*) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in S$
- $\boldsymbol{x}^*$ is called *minimizer* or *minimum* of $f$
- It suffices to consider only minimization, since maximum of $f$ is minimum of $-f$
- *Objective* function $f$ is usually differentiable, and may be linear or nonlinear
- *Constraint* set $S$ is defined by system of equations and inequalities, which may be linear or nonlinear
- Points $\boldsymbol{x} \in S$ are called *feasible* points
- If $S = \mathbb{R}^n$, problem is *unconstrained*

# Optimization Problems

- General continuous optimization problem:

$$\min f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = \mathbf{0} \quad \text{and} \quad \boldsymbol{h}(\boldsymbol{x}) \leq \mathbf{0}$$

where $f : \mathbb{R}^n \to \mathbb{R}, \quad \boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m, \quad \boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^p$

- *Linear programming*: $f$, $\boldsymbol{g}$, and $\boldsymbol{h}$ are all linear
- *Nonlinear programming*: at least one of $f$, $\boldsymbol{g}$, and $\boldsymbol{h}$ is nonlinear

# Newton's Method (for optimization)

- Another local quadratic approximation is truncated Taylor series

$$f(x+h) \approx f(x) + f'(x)h + \frac{f''(x)}{2}h^2$$

- By differentiation, minimum of this quadratic function of $h$ is given by $h = -f'(x)/f''(x)$
- Suggests iteration scheme

$$x_{k+1} = x_k - f'(x_k)/f''(x_k)$$

  which is *Newton's method* for solving nonlinear equation $f'(x) = 0$
- Newton's method for finding minimum normally has quadratic convergence rate, but must be started close

# Newton's Method: Example

- Use Newton's method to minimize $f(x) = 0.5 - x \exp(-x^2)$
- First and second derivatives of $f$ are given by

$$f'(x) = (2x^2 - 1)\exp(-x^2)$$

and

$$f''(x) = 2x(3 - 2x^2)\exp(-x^2)$$

- Newton iteration for zero of $f'$ is given by

$$x_{k+1} = x_k - (2x_k^2 - 1)/(2x_k(3 - 2x_k^2))$$

- Using starting guess $x_0 = 1$, we obtain

| $x_k$ | $f(x_k)$ |
|---|---|
| 1.000 | 0.132 |
| 0.500 | 0.111 |
| 0.700 | 0.071 |
| 0.707 | 0.071 |

# Newton's Method: Example

- Use Newton's method to minimize $f(x) = 0.5 - x\exp(-x^2)$
- First and second derivatives of $f$ are given by

$$f'(x) = (2x^2 - 1)\exp(-x^2)$$

  and

$$f''(x) = 2x(3 - 2x^2)\exp(-x^2)$$

- Newton iteration for zero of $f'$ is given by

$$x_{k+1} = x_k - (2x_k^2 - 1)/(2x_k(3 - 2x_k^2))$$

- Using starting guess $x_0 = 1$, we obtain

| $x_k$ | $f(x_k)$ |
|---|---|
| 1.000 | 0.132 |
| 0.500 | 0.111 |
| 0.700 | 0.071 |
| 0.707 | 0.071 |

# Steepest Descent

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be real-valued function of $n$ real variables
- At any point $\boldsymbol{x}$ where gradient vector is nonzero, negative gradient, $-\nabla f(\boldsymbol{x})$, points downhill toward lower values of $f$
- In fact, $-\nabla f(\boldsymbol{x})$ is locally direction of steepest descent: $f$ decreases more rapidly along direction of negative gradient than along any other
- *Steepest descent* method: starting from initial guess $\boldsymbol{x}_0$, successive approximate solutions given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \nabla f(\boldsymbol{x}_k)$$

where $\alpha_k$ is *line search* parameter that determines how far to go in given direction

# Steepest Descent

- Given descent direction, such as negative gradient, determining appropriate value for $\alpha_k$ at each iteration is one-dimensional minimization problem

$$\min_{\alpha_k} f(\boldsymbol{x}_k - \alpha_k \nabla f(\boldsymbol{x}_k))$$

  that can be solved by methods already discussed
- Steepest descent method is very reliable: it can always make progress provided gradient is nonzero
- But method is myopic in its view of function's behavior, and resulting iterates can zigzag back and forth, making very slow progress toward solution
- In general, convergence rate of steepest descent is only linear, with constant factor that can be arbitrarily close to $1$

# Steepest Descent: Example

- Use steepest descent method to minimize

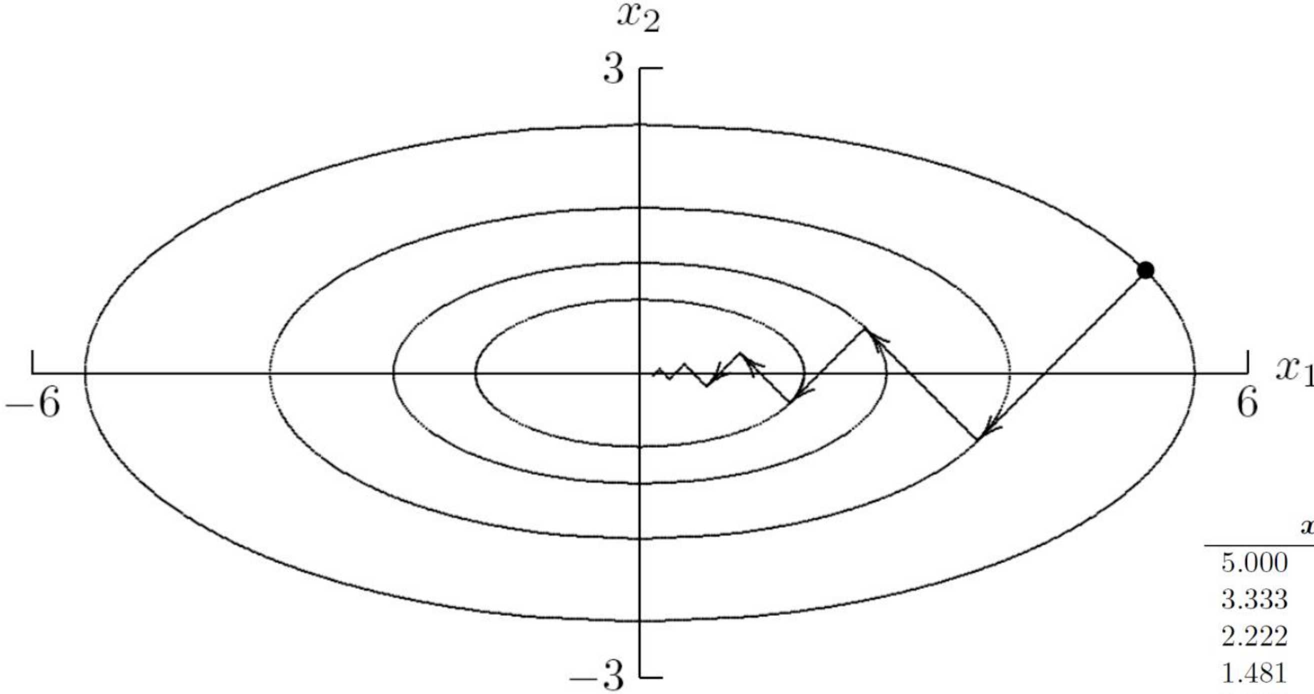
$$f(\boldsymbol{x}) = 0.5x_1^2 + 2.5x_2^2$$

- Gradient is given by $\nabla f(\boldsymbol{x}) = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix}$
- Taking $\boldsymbol{x}_0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, we have $\nabla f(\boldsymbol{x}_0) = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$
- Performing line search along negative gradient direction,

$$\min_{\alpha_0} f(\boldsymbol{x}_0 - \alpha_0 \nabla f(\boldsymbol{x}_0))$$

exact minimum along line is given by $\alpha_0 = 1/3$, so next approximation is $\boldsymbol{x}_1 = \begin{bmatrix} 3.333 \\ -0.667 \end{bmatrix}$

# Steepest Descent: Example




| $\boldsymbol{x}_k$ | | $f(\boldsymbol{x}_k)$ | $\nabla f(\boldsymbol{x}_k)$ | |
|---|---|---|---|---|
| 5.000 | 1.000 | 15.000 | 5.000 | 5.000 |
| 3.333 | −0.667 | 6.667 | 3.333 | −3.333 |
| 2.222 | 0.444 | 2.963 | 2.222 | 2.222 |
| 1.481 | −0.296 | 1.317 | 1.481 | −1.481 |
| 0.988 | 0.198 | 0.585 | 0.988 | 0.988 |
| 0.658 | −0.132 | 0.260 | 0.658 | −0.658 |
| 0.439 | 0.088 | 0.116 | 0.439 | 0.439 |
| 0.293 | −0.059 | 0.051 | 0.293 | −0.293 |
| 0.195 | 0.039 | 0.023 | 0.195 | 0.195 |
| 0.130 | −0.026 | 0.010 | 0.130 | −0.130 |

# Multi-Dimensional Optimization: Newton's Method

- Broader view can be obtained by local quadratic approximation, which is equivalent to Newton's method
- In multidimensional optimization, we seek zero of gradient, so *Newton iteration* has form

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}_f^{-1}(\boldsymbol{x}_k)\nabla f(\boldsymbol{x}_k)$$

where $\boldsymbol{H}_f(\boldsymbol{x})$ is *Hessian* matrix of second partial derivatives of $f$,

$$\{\boldsymbol{H}_f(\boldsymbol{x})\}_{ij} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$$

# Multi-Dimensional Optimization: Newton's Method

- Do not explicitly invert Hessian matrix, but instead solve linear system

$$\boldsymbol{H}_f(\boldsymbol{x}_k)\boldsymbol{s}_k = -\nabla f(\boldsymbol{x}_k)$$

for Newton step $\boldsymbol{s}_k$, then take as next iterate

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{s}_k$$

- Convergence rate of Newton's method for minimization is normally quadratic
- As usual, Newton's method is unreliable unless started close enough to solution to converge

# Multi-Dimensional Optimization: Newton's Method

- Use Newton's method to minimize

$$f(\boldsymbol{x}) = 0.5x_1^2 + 2.5x_2^2$$

- Gradient and Hessian are given by

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{H}_f(\boldsymbol{x}) = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

- Taking $\boldsymbol{x}_0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, we have $\nabla f(\boldsymbol{x}_0) = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$
- Linear system for Newton step is $\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix} \boldsymbol{s}_0 = \begin{bmatrix} -5 \\ -5 \end{bmatrix}$, so

  $\boldsymbol{x}_1 = \boldsymbol{x}_0 + \boldsymbol{s}_0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix} + \begin{bmatrix} -5 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, which is exact solution for this problem, as expected for quadratic function

# Constrained Optimization:Example

# First-Order Optimality Condition

- For function of one variable, one can find extremum by differentiating function and setting derivative to zero
- Generalization to function of $n$ variables is to find *critical point*, i.e., solution of nonlinear system

  $$\nabla f(\boldsymbol{x}) = \boldsymbol{0}$$

  where $\nabla f(\boldsymbol{x})$ is *gradient* vector of $f$, whose $i$th component is $\partial f(\boldsymbol{x})/\partial x_i$
- For continuously differentiable $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$, any interior point $\boldsymbol{x}^*$ of $S$ at which $f$ has local minimum must be critical point of $f$
- But not all critical points are minima: they can also be maxima or saddle points

# Second-Order Optimality Condition

- For twice continuously differentiable $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$, we can distinguish among critical points by considering *Hessian matrix* $\boldsymbol{H}_f(x)$ defined by

$$\{\boldsymbol{H}_f(x)\}_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

  which is symmetric

- At critical point $\boldsymbol{x}^*$, if $\boldsymbol{H}_f(\boldsymbol{x}^*)$ is
  - positive definite, then $\boldsymbol{x}^*$ is minimum of $f$
  - negative definite, then $\boldsymbol{x}^*$ is maximum of $f$
  - indefinite, then $\boldsymbol{x}^*$ is saddle point of $f$
  - singular, then various pathological situations are possible

# Constrained Optimality

- If problem is constrained, only *feasible* directions are relevant
- For equality-constrained problem

$$\min f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{0}$$

  where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{g}\colon \mathbb{R}^n \to \mathbb{R}^m$, with $m \leq n$, necessary condition for feasible point $\boldsymbol{x}^*$ to be solution is that negative gradient of $f$ lie in space spanned by constraint normals,

$$-\nabla f(\boldsymbol{x}^*) = \boldsymbol{J}_g^T(\boldsymbol{x}^*)\boldsymbol{\lambda}$$

  where $\boldsymbol{J}_g$ is Jacobian matrix of $\boldsymbol{g}$, and $\boldsymbol{\lambda}$ is vector of *Lagrange multipliers*
- This condition says we cannot reduce objective function without violating constraints

# Constrained Optimality

- *Lagrangian function* $\mathcal{L}\colon \mathbb{R}^{n+m} \to \mathbb{R}$, is defined by

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{g}(\boldsymbol{x})$$

- Its gradient is given by

$$\nabla \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla f(\boldsymbol{x}) + \boldsymbol{J}_g^T(\boldsymbol{x})\boldsymbol{\lambda} \\ \boldsymbol{g}(\boldsymbol{x}) \end{bmatrix}$$

- Its Hessian is given by

$$\boldsymbol{H}_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} \boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\lambda}) & \boldsymbol{J}_g^T(\boldsymbol{x}) \\ \boldsymbol{J}_g(\boldsymbol{x}) & \boldsymbol{O} \end{bmatrix}$$

where

$$\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{H}_f(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \boldsymbol{H}_{g_i}(\boldsymbol{x})$$

# Constrained Optimality

- Together, necessary condition and feasibility imply critical point of Lagrangian function,

$$\nabla \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla f(\boldsymbol{x}) + \boldsymbol{J}_g^T(\boldsymbol{x})\boldsymbol{\lambda} \\ \boldsymbol{g}(\boldsymbol{x}) \end{bmatrix} = \boldsymbol{0}$$

- Hessian of Lagrangian is symmetric, but not positive definite, so critical point of $\mathcal{L}$ is saddle point rather than minimum or maximum

- Critical point $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ of $\mathcal{L}$ is constrained minimum of $f$ if $\boldsymbol{B}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is positive definite on *null space* of $\boldsymbol{J}_g(\boldsymbol{x}^*)$

- If columns of $\boldsymbol{Z}$ form basis for null space, then test *projected* Hessian $\boldsymbol{Z}^T \boldsymbol{B} \boldsymbol{Z}$ for positive definiteness

# Constrained Optimization: Example

# Constrained Optimization: Example