Introduction to Unconstrained Optimization: Part 1

James Allison

ME 555
January 29, 2007
Monotonicity Analysis

What can MA be used for?

- Problem size reduction (elimination of variables and constraints)
- Identification of problems such as unboundedness
- Solution of optimization problem in some cases

What can be done if MA does not lead to a solution?

- Application of optimality conditions
  - Use optimality conditions to derive analytical solution
  - Use numerical algorithms based on optimality conditions
Monotonicity Analysis

What can MA be used for?

- Problem size reduction (elimination of variables and constraints)
- Identification of problems such as unboundedness
- Solution of optimization problem in some cases
Monotonicity Analysis

What can MA be used for?

- Problem size reduction (elimination of variables and constraints)
- Identification of problems such as unboundedness
- Solution of optimization problem in some cases

What can be done if MA does not lead to a solution?
Monotonicity Analysis

What can MA be used for?

- Problem size reduction (elimination of variables and constraints)
- Identification of problems such as unboundedness
- Solution of optimization problem in some cases

What can be done if MA does not lead to a solution?

Application of optimality conditions

- Use optimality conditions to derive analytical solution
- Use numerical algorithms based on optimality conditions
### Ch. 4: Unconstrained Optimization

- Concerned only with objective function
- Constrained optimization covered in Ch. 5

$$\min_{x} f(x)$$
**Scope**

**Ch. 4: Unconstrained Optimization**
- Concerned only with objective function
- Constrained optimization covered in Ch. 5

\[
\min_x f(x)
\]

**Assumptions**
- Functions and variables are continuous
- Functions are $C^2$ smooth
Lecture Outline

- Derivation of optimality conditions
- Analytical solutions
- Function approximations
- Numerical methods
  - First order methods (gradient descent)
  - Second order methods (Newton's Method)
- Problem scaling
Optimality Conditions

**Necessary Conditions**

\[ B \Rightarrow A \]

\( B \) is true only if \( A \) is true (\( A \) is necessary for \( B \))

**Sufficient Conditions**

\[ A \Rightarrow B \]

\( B \) is true if \( A \) is true (\( A \) is sufficient for \( B \))

**Necessary and Sufficient Conditions**

\[ B \iff A \]

\( B \) is true if and only if \( A \) is true (\( A \) is necessary and sufficient for \( B \))
Math Review

**Gradient: \( \nabla f(\mathbf{x}) \)**

- Multidimensional derivative
- Vector valued (column in my documents, row in POD2)
- \( \nabla f(\mathbf{x}) = [\partial f / \partial x_1, \partial f / \partial x_2, \ldots, \partial f / \partial x_n]^T \)
- Points in direction of steepest ascent

Example:

\[
f(x, y) = x^2 + 2y^2 - xy
\]
Math Review

Gradient: $\nabla f(x)$

- Multidimensional derivative
- Vector valued (column in my documents, row in POD2)
- $\nabla f(x) = [\partial f / \partial x_1, \partial f / \partial x_2, \ldots, \partial f / \partial x_n]^T$
- Points in direction of steepest ascent

Example:

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

**Hessian: $H$, sometimes written $\nabla^2 f(x)$**

- Multidimensional second derivative
- Matrix valued (symmetric)
- Provides function shape information

$$
H = \begin{bmatrix}
\frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \\
\frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2}
\end{bmatrix}
$$
Hessian: $\mathbf{H}$, sometimes written $\nabla^2 f(x)$

- Multidimensional second derivative
- Matrix valued (symmetric)
- Provides function shape information

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

Example:

$$f(x, y) = x^2 + 2y^2 - xy$$
Math Review: Taylor’s series expansion

Function of one variable:

\[ f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n \]

Function of multiple variables:

\[ f(x) = f(x_0) + \sum_{i=1}^{n} \frac{\partial f(x_0)}{\partial x_i} (x_i - x_{i0}) \]
\[ + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(x_0)}{\partial x_i \partial x_j} (x_i - x_{i0})(x_j - x_{j0}) + o \left( \|x - x_0\|^2 \right) \]
Math Review: Taylor’s series expansion

Function of one variable:

\[ f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \]

Function of multiple variables: (matrix form)

\[ \partial f \triangleq f(x) - f(x_0) = \nabla f(x_0)\partial x + \partial x^T H \partial x + o(\|x - x_0\|^2) \]

where \( \partial x \triangleq x - x_0 \) is the perturbation vector
First order necessity

Suppose that $x_*$ is a local minimum of $f(x)$.
- no perturbations about $x_*$ will result in a function decrease
- first order function approximation can be used to derive necessary conditions (i.e., if $x_*$ is a minimum, then what must be true?)
Derivation of Optimality Conditions

First order necessity

Suppose that \( x_* \) is a local minimum of \( f(x) \).

- no perturbations about \( x_* \) will result in a function decrease
- first order function approximation can be used to derive necessary conditions (i.e., if \( x_* \) is a minimum, then what must be true?)

If \( x_* \) minimizes \( f(x) \), then \( \nabla f(x_*) = 0 \)
First order necessity

Suppose that $x_*$ is a local minimum of $f(x)$.

- no perturbations about $x_*$ will result in a function decrease
- first order function approximation can be used to derive necessary conditions (i.e., if $x_*$ is a minimum, then what must be true?)

If $x_*$ minimizes $f(x)$, then $\nabla f(x_*) = 0$

If $\nabla f(x^\dagger) = 0$, then $x^\dagger$ is a stationary point, but not necessarily a minimizer
Derivation of Optimality Conditions

Second order sufficiency

Suppose that \( \mathbf{x}^\dagger \) is a stationary point of \( f(\mathbf{x}) \).

- no perturbations about \( \mathbf{x}^* \) will result in a function decrease
- second order function approximation can be used to derive necessary conditions (i.e., what must be true for \( \mathbf{x}^\dagger \) to be a minimum?)
Second order sufficiency

Suppose that $x_\dagger$ is a stationary point of $f(x)$.

- no perturbations about $x_\star$ will result in a function decrease
- second order function approximation can be used to derive necessary conditions (i.e., what must be true for $x_\dagger$ to be a minimum?)

If $H \succ 0$ at $x_\dagger$, then $x_\dagger$ minimizes $f(x)$.
Analytical Solution Example

\[ \min_{x} f(x) = x_{1}^{2} + 2x_{2}^{2} - x_{1}x_{2} + x_{1} \]

1) Identify stationary point

2) Test stationary point
Analytical Solution Example

\[ \min_{x} f(x) = x_1^2 + 2x_2^2 - x_1x_2 + x_1 \]

1) Identify stationary point

\[ x^\dagger = \begin{bmatrix} -\frac{4}{7} \\ -\frac{1}{7} \end{bmatrix} \]

2) Test stationary point
Analytical Solution Example

\[
\min_x f(x) = x_1^2 + 2x_2^2 - x_1x_2 + x_1
\]

1) Identify stationary point

\[x^* = \begin{bmatrix} -\frac{4}{7} \\ -\frac{1}{7} \end{bmatrix}\]

2) Test stationary point

\[
H = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix} \succ 0
\]
Analytical solution to optimization problem frequently impossible (why?)

Numerical algorithms based on FONC and function models can be used
First Order Model: Gradient Descent

\[ f(x) \approx f(x^k) + \nabla f(x^k) \partial x \]

1. Build a linear model about the current point \( x^k \)
2. Move in the direction of steepest descent \( -\nabla f(x^k) \) until \( f(x) \) stops improving
3. Update the linear model and repeat until no descent direction exists \( (\nabla f(x^k) = 0) \)

Iterative formula:

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

\[ f(\mathbf{x}) = 7x_1^2 + 2.4x_1x_2 + x_2^2, \quad \mathbf{x}^k = [10 \ -5]^T \]

\[ \min_{\alpha} f(\alpha) = \mathbf{x}_0 - \alpha \nabla f(\mathbf{x}_0) \]
Discussion on Gradient Descent Method

- Stability/optimality
- Descent
- Speed of convergence
Second Order Model: Newton’s Method

\[ f(x) \approx f(x^k) + \nabla f(x^k) \partial x + \partial x^T H \partial x \]

1. Build a quadratic model about the current point \( x^k \)
2. Go to the quadratic approximation for the stationary point
3. Update the quadratic model and repeat until the current point is a stationary point (\( \nabla f(x^k) = 0 \))

Iterative formula:

\[ x^{k+1} = x^k - H^{-1} \nabla f(x^k) \]
Second Perspective on Newton’s Method

Newton’s method for root finding (solve $f(x) = 0$):

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$$

Multidimensional system of equations (solve $f(x) = 0$):

$$x^{k+1} = x^k - J^{-1}f(x^k)$$
Second Perspective on Newton’s Method

Newton’s method for root finding (solve $f(x) = 0$):

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$$

Multidimensional system of equations (solve $f(x) = 0$):

$$x^{k+1} = x^k - J^{-1}f(x^k)$$

What system of equations do we need to solve in unconstrained optimization?
Second Perspective on Newton’s Method

Newton’s method for root finding (solve $f(x) = 0$):

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$$

Multidimensional system of equations (solve $f(x) = 0$):

$$x^{k+1} = x^k - J^{-1}f(x^k)$$

What system of equations do we need to solve in unconstrained optimization?

$$\nabla f(x) = 0$$
Discussion on Newton’s Method

- Stability/optimality
- Descent
- Speed of convergence
Quadratic Forms

**Quadratic Form**: function is a linear combination of $x_i x_j$ terms

**Matrix representation**:

$$f(x) = x' Ax$$

Example: $f(x) = 2x_1^2 + x_1 x_2 + x_2^2 + x_2 x_3 + x_3^2$
Quadratic Forms: function is a linear combination of $x_i x_j$ terms

Matrix representation:

$$f(x) = x'Ax$$

Example: $f(x) = 2x_1^2 + x_1 x_2 + x_2^2 + x_2 x_3 + x_3^2$

$$f(x) = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 2 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x'Ax$$
Types of Quadratic Functions

- if $x'Ax > 0 \quad \forall x$, $A$ is positive definite $\Rightarrow$ convex quadratic function
- if $x'Ax < 0 \quad \forall x$, $A$ is negative definite $\Rightarrow$ concave quadratic function
- if $x'Ax \leq 0$, $A$ is indefinite $\Rightarrow$ hyperbolic quadratic function

Physical interpretation?
Quadratic Function Definitions

\[ f_1(x) = x' A_1 x \]
\[ f_2(x) = x' A_2 x \]
\[ f_3(x) = x' A_3 x \]

Where:

\[ A_1 = \begin{bmatrix} 7 & 1.2 \\ 1.2 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -7 & 1.2 \\ 1.2 & -1 \end{bmatrix}, \quad \text{and} \quad A_3 = \begin{bmatrix} -5 & 2.6 \\ 2.6 & 2 \end{bmatrix} \]
Eigenvalues, Eigenvectors, and Function Geometry

\[ Av = \lambda v \]

\( A \) eigenvector \( v \)
\( \lambda \) eigenvalue
Eigenvalues, Eigenvectors, and Function Geometry

\[ \mathbf{A}\mathbf{v} = \lambda \mathbf{v} \]

- Provide insight into function shape
- Facilitate useful coordinate system rotations
- Helpful for understanding problem condition (ellipticity) and scaling
Interpretation of Eigenvalues

\[ f(x) = f_0 + x'b + x'Ax \]

\[ \nabla f(x) = b + 2Ax \]

\[ x^+_t = -\frac{1}{2}A^{-1}b \]

Shift coordinate system:

Rotate coordinate system:
Interpretation of Eigenvalues

\[ f(x) = f_0 + x'b + x'Ax \]
\[ \nabla f(x) = b + 2Ax \]
\[ x^* = -\frac{1}{2}A^{-1}b \]

Shift coordinate system:

\[ f(z) = f^*_t + z'Az \]

Rotate coordinate system:
Interpretation of Eigenvalues

\[ f(x) = f_0 + x'b + x'Ax \]
\[ \nabla f(x) = b + 2Ax \]
\[ x_\dagger = -\frac{1}{2}A^{-1}b \]

Shift coordinate system:

\[ f(z) = f_\dagger + z'Az \]

Rotate coordinate system:

\[ f(p) = f_\dagger + \sum_{i=1}^{n} \lambda_i p_i^2 \]
Numerical Examples

Function 1:

\[
\mathbf{v}_1 = \begin{bmatrix} .189 \\ -.982 \end{bmatrix}, \quad \lambda_1 = .769, \quad \mathbf{v}_2 = \begin{bmatrix} -.982 \\ -.189 \end{bmatrix}, \quad \lambda_2 = 7.23
\]

Function 2:

\[
\mathbf{v}_1 = \begin{bmatrix} -.982 \\ .189 \end{bmatrix}, \quad \lambda_1 = -.723, \quad \mathbf{v}_2 = \begin{bmatrix} -.189 \\ -.982 \end{bmatrix}, \quad \lambda_2 = -.769
\]

Function 3:

\[
\mathbf{v}_1 = \begin{bmatrix} -.949 \\ .314 \end{bmatrix}, \quad \lambda_1 = -5.86, \quad \mathbf{v}_2 = \begin{bmatrix} -.314 \\ -.949 \end{bmatrix}, \quad \lambda_2 = 2.86
\]
Numerical Examples

Convex Quadratic Function

Concave Quadratic Function

Hyperbolic Quadratic Function

\[
\begin{align*}
\lambda_1 &= .769 \\
\lambda_2 &= 7.23 \\
\lambda_1 &= - .723 \\
\lambda_2 &= - .769 \\
\lambda_1 &= -5.86 \\
\lambda_2 &= 2.86
\end{align*}
\]
Connection with Quadratic Form

- \( x'Ax > 0 \quad \forall x \iff \lambda_i > 0 \quad \forall i \)
  \( \Rightarrow A \succ 0 \land \text{convex quadratic function} \)

- \( x'Ax < 0 \quad \forall x \iff \lambda_i < 0 \quad \forall i \)
  \( \Rightarrow A \prec 0 \land \text{concave quadratic function} \)

- if \( x'Ax \leq 0 \iff \lambda_i \leq 0 \)
  \( \Rightarrow A \text{ is indefinite} \land \text{hyperbolic quadratic function} \)
Problem Scaling
Condition Number

\[ C = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}} \]

- \( C \gg 1 \Rightarrow \) numerical difficulties, slow convergence
- \( C \approx 1 \Rightarrow \) faster convergence
Scaling Approaches

- Scale design variables to be the same magnitude: \( y = s'x \)
- Account for \( v \) not aligned with coordinate axes: \( y = S^{-1}x \)
- Implement scaling within algorithm: