Quick Introduction to Optimization

Haibin Ling

http://www.ist.temple.edu/~hbling/Teaching/13F_5543/index.html

Optimization Problems

Optimization: determining an argument for which a given function has an extreme value on a given domain.

\[ \min_x f(x) \quad f: \mathbb{R}^n \rightarrow \mathbb{R} \]

subject to \[ g(x) = 0 \quad \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \]

\[ h(x) \leq 0 \quad \begin{bmatrix} h_1(x) \\ h_2(x) \\ \vdots \\ h_p(x) \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \]

Example

- Minimum surface areas of a cylinder given its volume.

\[ \min f(x_1, x_2) = 2\pi x_1 (x_1 + x_2) \]

subject to \[ g(x_1, x_2) = \pi x_1^2 x_2 - V = 0 \]

Examples in Practice

- Support vector machine
  - Maximize classifier margin
  - Subject to constraints from training data

- Point cloud matching
  - Minimize total matching cost
  - Subject to neighborhood constraint

- Surface warping
  - Minimize bending energy
  - Subject to smooth constraint (regularization)

- Bayesian inference
  - Maximizing posterior probability
  - Energy minimization (MRF, ...)

Unconstrained Optimality Conditions

- Unconstrained problem

\[ \min f(x) \]

- Derivatives

\[
\begin{array}{c|cc|cc|cc}
\hline
\phi(x) & \frac{\partial \phi}{\partial x_1} & \frac{\partial \phi}{\partial x_2} & \frac{\partial \phi}{\partial x_3} \\
\frac{\partial \phi}{\partial x_1} & \frac{\partial^2 \phi}{\partial x_1^2} & \frac{\partial^2 \phi}{\partial x_1 \partial x_2} & \frac{\partial^2 \phi}{\partial x_1 \partial x_3} \\
\frac{\partial \phi}{\partial x_2} & \frac{\partial^2 \phi}{\partial x_2 \partial x_1} & \frac{\partial^2 \phi}{\partial x_2^2} & \frac{\partial^2 \phi}{\partial x_2 \partial x_3} \\
\frac{\partial \phi}{\partial x_3} & \frac{\partial^2 \phi}{\partial x_3 \partial x_1} & \frac{\partial^2 \phi}{\partial x_3 \partial x_2} & \frac{\partial^2 \phi}{\partial x_3^2} \\
\hline
\end{array}
\]

\[
H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \\
\end{bmatrix}
\]
Unconstrained Optimality Conditions

- **Taylor expansion (1st order)**
  \[ f(x+s) = f(x) + \nabla f(x)^T s \]
- **First-order necessary condition**
  \[ f \text{ reaches local minimum at } x^* \Rightarrow \nabla f(x^*) = 0 \]
- **Taylor expansion (2nd order)**
  \[ f(x+s) = f(x) + \nabla f(x)^T s + \frac{1}{2} s^T \nabla^2 f(x) s + o(s^2) \]
- **Second-order sufficient condition**
  \[ s^T \nabla^2 f(x) s > 0 \quad \forall s \neq 0 \]
  A.k.a. \( Hf \) positive definite

Constrained Optimality Conditions

- **Equality constraints**
  \[ \min f(x) \quad \text{subject to} \quad g(x) = 0 \]
  From the first-order necessary condition, we have for minimum \( x^* \)
  \[ \nabla f(x^*)^T J(x^*) = 0 \]
  Constrain on \( s \) feasible direction
  \[ -\nabla f(x^*) = J(x^*) \lambda \]
  \[ \lambda \in \mathbb{R}^m \]

- **Inequality constraints**
  \[ \min f(x) \quad \text{subject to} \quad g(x) \leq 0 \]
  \[ L(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x) \]
  Karush-Kuhn-Tucker (KKT) conditions
  \[ \nabla L(x^*, \lambda^*, \mu^*) = 0 \]
  \[ g(x^*) = 0 \]
  \[ h(x^*) \leq 0 \]
  \[ \mu^* \geq 0 \]
  \[ \lambda(x^*) = 0 \]

Solutions for Nonconstrained Problems

- **Summary, for** \( \min f(x) \)
  Given a critical point \( x^* \), i.e., \( \nabla f(x^*) = 0 \)
  By studying
  \[ f(x+s) = f(x) + \frac{1}{2} s^T \nabla^2 f(x) s \]
  We have
  - \( Hf \) positive definite \( \Rightarrow x^* \) is a minimum of \( f \)
  - \( Hf \) negative definite \( \Rightarrow x^* \) is a maximum of \( f \)
  - \( Hf \) indefinite \( \Rightarrow x^* \) is a saddle of \( f \)
  - \( Hf \) singular \( \Rightarrow x^* \) is pathological
No Constraints – 1D

- Some straightforward methods
  - Golden Section Search
  - Successive Parabolic Interpolation

- Newton’s Method
  - Second-order Taylor
    \[ f(x + s) = f(x) + f'(x)s + \frac{1}{2} f''(x)s^2 \]
  - Minimum in s
    \[ s = -f'(x)/f''(x) \]
  - Iterative algorithm
    \[ x_{k+1} = x_k - f'(x_k)/f''(x_k) \]

No Constraints – High Dimensional

- Direct search
  - Good for nonsmooth objective functions
  - Slow for large dimensional problems

- Steepest Descent
  - Intuition: gradient points to the steepest direction

\[ \text{Solve } H_f(x_k) = -\nabla f(x_k) \text{ for } s_k \]
\[ x_{k+1} = x_k + \alpha_k s_k \]

Solutions for Nonconstrained Problems

- Problem of steepest descent
  - Slow convergence rate

- Newton’s Method
  - Second-order Taylor
    \[ f(x + s) = f(x) + \nabla f(x)^T s + \frac{1}{2} s^T H_f(x)s \]

\[ H_f(x_k) = -\nabla^2 f(x_k) \]
\[ x_{k+1} = x_k + \alpha_k s_k \]

- Variants of Newton’s method
  - Quasi-Newton
  - Scan Updating

Nonlinear Least Squares

Input: \((t_i, y_i), i = 1, \ldots, m\)
Target: fit a function \(f(t, x)\), where \(x\) can be viewed as parameters.

Residual: \(r(x) = y_i - f(t_i, x)\)
Optimization: minimize the total residual
\[
\phi(x) = \frac{1}{2} \sum_i r_i(x)^2 \quad \text{for } i = 1, 2, \ldots
\]

\[ \nabla \phi(x) = -\sum_i J(x_i)r(x_i) \]
\[ H_f(x) = J(x)^T J(x) + \sum_i J(x_i)^T J(x_i) \]

Newton’s method:
\[ J(x_i)J(x_i)^T \]
Hessian: \(H_f\)

Problem: slow due to Hessian

Nonlinear Least Squares

Newton’s method:
\[ \left( J'(x_k)J(x_k) + \sum_i \gamma_{ik} H_f(x_k) \right) h_k = -J'(x_k)r(x_k), \]
Observation: \(r_i\) are usually small for a good fitting

\[ J'(x_k)J(x_k) h_k = -J'(x_k)r(x_k), \]
\[ \Rightarrow \text{ Gauss-Newton: } J'(x_k) h_k = -r(x_k) \]

Drawback: when some \(r_i\) are not that small
Levenberg-Marquardt method:
\[ \left( J'(x_k)J(x_k) + \mu I \right) h_k = -J'(x_k)r(x_k) \]
\[ \mu \text{ regularization} \]
Solutions for Constrained Problems

Sequential Quadratic Programming

Equality-constrained problems: \( \min f(x) \) subject to \( g(x) = 0 \)
Lagrange function: \( L(x, \lambda) = f(x) + \lambda^T g(x) \)

\( \Rightarrow \) Solve the following equations

\[ \frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial \lambda} = 0 \]

Applying Newton’s method and then after some derivation

\[ \min \frac{1}{2} \sum \phi_i x_i^2 + \sum \phi_i \phi_j \sum \lambda_i \]

subject to \( \lambda_i g_i(x) = 0 \)

where \( \lambda_i = H_{ij}(x) \)

This is done iteratively \( \Rightarrow \) Sequential quadratic programming

Penalty Methods

Equality-constrained problems: \( \min f(x) \) subject to \( g(x) = 0 \)

Construct an unconstrained problem:

\[ \min \phi_f(x, \rho) = f(x) + \rho \sum g(x)^2 \]

Penalize for non-zero \( g(x) \) \( \Rightarrow \) Penalty function method.

Drawbacks: becomes unstable when \( \rho \to \infty \).

Barrier Methods

Motivation: keep intermediate points feasible.

Inequality-constrained problems: \( \min f(x) \) subject to \( h(x) \geq 0 \)

Construct barrier functions:

- Inverse barrier: \( \phi_f(x, \mu) = f(x) - \mu \sum \log h(x) \)
- Logarithmic barrier: \( \phi_f(x, \mu) = f(x) - \mu \sum \log h(x) \)

Solves a sequence of unconstrained problems, with decreasing \( \mu \).

Another intuition: from the interior of the feasible set to search for the minimum solution \( \Rightarrow \) interior point methods

Special Optimization Problems

Linear Programming

\[ \min c^T x + d \]
subject to \( Gx \leq h \)
\( Ax = b \)

Simplex method: search in the corners of feasible solution polyhedron.
Practically very efficient.
Theoretically can be exponentially expensive.

Interior method: search start from interior.
Proved to have polynomial efficiency in worst case.
Quadratic Programming

\[ \text{min} \frac{1}{2} x^T P x + q^T x + r \]
subject to
\[ G x \leq h \]
\[ A x = b \]

Objective is quadratic, which is convex.
More general case, convex programming.