Methods for finding optimal configurations

Milos Hauskrecht
milos@pitt.edu
5329 Sennott Square

Parametric optimization

Optimal configuration search:
• Configurations are often described in terms of variables and their values
• Each configuration has a quality measure
• **Goal:** find the configuration with the best value

When the state space we search is finite, the search problem is called a **combinatorial optimization problem**
When parameters we want to find are real-valued
  – **parametric optimization problem**
Parametric optimization

Parametric optimization:

• Configurations are described by a vector of parameters (variables) $\mathbf{w}$ with real-valued values

• **Goal:** find the set of parameters $\mathbf{w}$ that optimize (maximize or minimize) the given objective function $f(\mathbf{w})$

Parametric optimization techniques

• **Special cases (with efficient solutions):**
  – Linear programming
  – Quadratic programming
  – Convex optimization

• **General optimization methods (also for nonlinear problems)**
  – First-order methods:
    • Gradient-ascent (descent)
    • Conjugate gradient
  – Second-order methods:
    • Newton-Rhapson method
    • Broyden–Fletcher–Goldfarb–Shanno (BFGS) method
      Levenberg-Marquardt

• **Constrained optimization:**
  – Lagrange multipliers
Linear programming

- A special case when:
  - The objective function $f$ is a linear combination of variable values $w$
  - Values variables $w$ can take are constrained by a set of linear constraints
- Assume variables: $w_1, w_2, ... w_k$

Minimize $f(w_1, w_2, ... w_k)= a_1w_1 + a_2w_2 + ... + a_kw_k$

Subject to constraints:

- $b_{1,1}w_1 + b_{1,2}w_2 + ... + b_{1,k}w_k + b_{1,0} \leq 0$
- $b_{2,1}w_1 + b_{2,2}w_2 + ... + b_{2,k}w_k + b_{2,0} \leq 0$
- ...
- $b_{m,1}w_1 + b_{m,2}w_2 + ... + b_{m,k}w_k + b_{m,0} \leq 0$

Gradient ascent method

- A method for finding parameters $w_1, w_2, ... w_k$ optimizing an arbitrary differentiable function $f(w_1, w_2, ... w_k)$
- Example:

\[
\nabla f(w) = \left[ \frac{\partial f(w)}{\partial w_1}, \frac{\partial f(w)}{\partial w_2}, ..., \frac{\partial f(w)}{\partial w_k} \right]
\]
Gradient ascent method

- **Gradient ascent**: the same as hill-climbing, but in the continuous parametric space $w$

\[ \frac{\partial}{\partial w} f(w) \bigg|_{w^*} \]

- What is the derivative of an increasing function?
  - positive

M. Hauskrecht
**Gradient ascent method**

- **Gradient ascent**: the same as hill-climbing, but in the continuous parametric space $w$

![Graph of gradient ascent]

- Change the parameter value of $w$ according to the gradient

\[ w^{*} \leftarrow w^{*} + \alpha \frac{\partial}{\partial w} f(w) \mid_{w^{*}} \]

$\alpha > 0$ - a learning rate (scales the gradient changes)
Gradient ascent method

- To get to the function minimum repeat (iterate) the gradient based update few times

- **Problems:**
  - local optima, saddle points, slow convergence
  - More complex optimization techniques use additional information (e.g. second derivatives)