## **Optimization in Machine Learning**

# **Simulated Annealing**





#### Learning goals

- Motivation
- Metropolis algorithm
- Simulated Annealing

### INTRODUCTION

**Heuristics** for the optimization of complex (multivariate, non-linear, non-convex) objective functions

- Procedure for finding good solutions to complex problems.
- Does not guarantee optimal/best result (global optimum), but usually good solutions.
- Goal for complex optimization problems: avoid "getting stuck" in local optima.
- Is often used for difficult discrete problems as well.
- Local search strategy with random option to accept worse values.

× 0 0 × × ×

#### SIMPLE STOCHASTIC LOCAL SEARCH

- Given is a multivariate objective function  $f(\mathbf{x})$
- Define a local neighborhood area  $V(\mathbf{x})$  for a given  $\mathbf{x}$
- Sample proposal  $\mathbf{x}^{[t+1]}$  uniformly at random from neighborhood  $V(\mathbf{x}^{[t]})$
- Calculate  $f(\mathbf{x}^{[t+1]})$
- If Δ*f* = *f*(**x**<sup>[*t*+1]</sup>) − *f*(**x**<sup>[*t*]</sup>) < 0, **x**<sup>[*t*+1]</sup> is accepted as new solution, otherwise a new proposal from neighborhood is sampled.



Simple stochastic local search: Acceptance (green) and rejection range (red)



#### **METROPOLIS ALGORITHM**

- Simple stochastic local search strongly depends on **x**<sup>[0]</sup> and the neighborhood.
  - $\Rightarrow$  Danger of ending up in local minima
- Idea: allow worse candidates with some probability
- Metropolis: accept candidates from previous rejection range  $(\Delta f > 0)$  with probability  $\mathbb{P}(\operatorname{accept} | \Delta f) = \exp(-\Delta f/T)$
- T denotes "temperature"



Simulated annealing: Colors correspond to  $\mathbb{P}(\text{accept})$ 

× 0 0 × × ×

#### METROPOLIS ALGORITHM / 2

- Parameter T describes temperature/progress of the system
- High temperatures correspond to high probability of accepting worse x
- Local minima can be escaped, but no convergence can be achieved at *constant* temperature
- We come across an important principle of optimization:

exploration (high T) vs. exploitation (low T)



× × 0 × × ×

#### SIMULATED ANNEALING

- Start with high temperature to explore whole space
- Slowly reduce temperature to converge
  ⇒ Sequence of descending temperatures T<sup>[t]</sup>, t ∈ ℕ
- Procedure is called simulated annealing
- Temperature is often kept constant several iterations in a row to explore the space, then multiplied by coefficient 0 < *c* < 1:

$$T^{[t+1]} = c \cdot T^{[t]}$$

• Other strategies possible, for example:

$$T^{[t]} = T^{[0]} \left( 1 - \frac{t}{t_{\max}} \right)$$

Choosing neighborhood:

• Many different strategies. Strongly depends on objective function.



### ANALOGY TO METALLURGY

- Simulated annealing draws analogy between a cooling process (e.g. a metal or liquid) and an optimization problem.
- If cooling of a liquid material (amount of atoms) is too fast, it solidifies in suboptimal configuration, slow cooling produces crystals with optimal structure (minimum energy stage).
- Consider atoms of the liquid as a system with many degrees of freedom, analogy to optimization problem of a multivariate function
- Minimum energy stage corresponds to optimum of objective function.

× < 0 × × ×