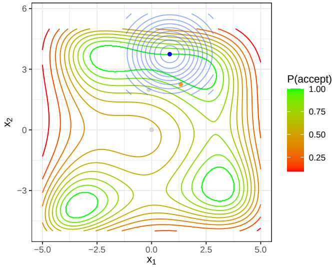
# Optimization in Machine Learning

# Simulated Annealing




## Learning goals

- Motivation
- Metropolis algorithm
- Simulated Annealing

# MOTIVATION

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

Heuristics:

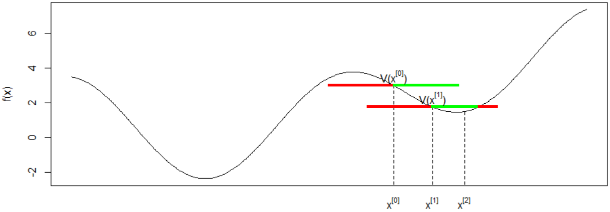
- Procedure for finding good solutions to complex problems.
- Does not guarantee optimal/best result (global optimum), but usually good solutions.
- Goal for complex optimization problems: avoid “getting stuck” in local optima.
- Is often used for difficult discrete problems as well.

# MOTIVATION

- **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.
- Mathematically it is a local search strategy, with a random option to accept even worse values (sometimes).

# STOCHASTIC LOCAL SEARCH STRATEGY

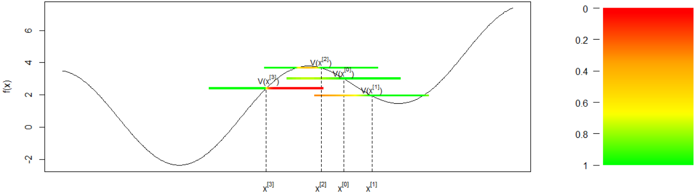
- Given is a multivariate objective function $f(\boldsymbol{x})$
- Define a local neighborhood area $V(\boldsymbol{x})$ for a given $\boldsymbol{x}$
- Stochastic local search produces a new solution $\boldsymbol{x}^{[t+1]}$ from neighborhood $V(\boldsymbol{x}^{[t]})$ of the solution $\boldsymbol{x}^{[t]}$ by sampling of a uniform distribution.
- Calculate $f(\boldsymbol{x}^{[t+1]})$
- If $\Delta f = f(\boldsymbol{x}^{[t+1]}) - f(\boldsymbol{x}^{[t]}) < 0$, $\boldsymbol{x}^{[t+1]}$ is accepted as new solution, otherwise a new candidate solution from neighborhood is selected.




Stochastic local search; green: acceptance range, red: rejection range

# METROPOLIS ALGORITHM

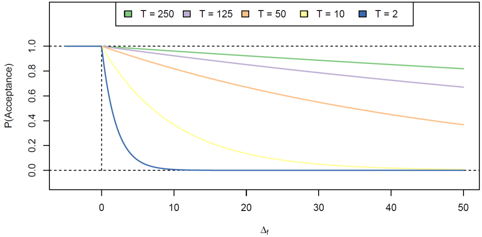
- Stochastic local search strongly depends on the initialization of $x^{[0]}$ and the neighborhood.
- Danger of ending up in local minima.
- Sensible: temporarily allow worse candidate combinations.
- Metropolis: accept candidate solutions from previous rejection range ($\Delta f > 0$) with probability $\mathbb{P}(\text{acceptance}|\Delta f) = exp(-\frac{\Delta f}{T})$.
- $T$ denotes the **temperature**




Simulated annealing schematic, colors: $\mathbb{P}$(acceptance)

# METROPOLIS ALGORITHM

- New parameter $T$ describes temperature/progress of the system.
- The higher $T$, the higher the probability to accept worse $\boldsymbol{x}$.
- Atomical view: individual atoms (solution points) of the system can move more freely
- Local minima can be escaped again, but no convergence can be achieved at constant temperature
- We come across an important principle of optimization: **exploration (high T) vs. exploitation (low T)**



# SIMULATED ANNEALING

- Approach now: start with high temperature to search the whole system (**exploration**)
- Slowly lower temperature to reach a minimum
  $\Rightarrow$ sequence of temperatures $T^{[t]}, t \in \mathbb{N}$
- If temperature depends on simulation time, the procedure is called **simulated annealing**.
- Temperature is often kept constant several iterations at a time to search the space of candidate solutions, then multiplied by coefficient $0 < c < 1$:

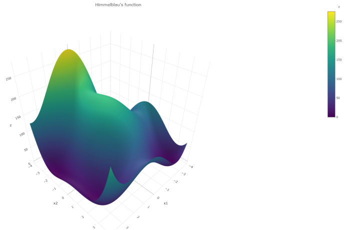
$$T^{[t+1]} = c \cdot T^{[t]}$$

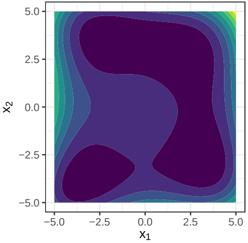
# SIMULATED ANNEALING

(Choice of) optimization parameters

- Temperature $T$: for any optimization problem, the initial temperature can be the average of a number of random function values.
- Temperature coefficient $c$: typically between 0.6 and 0.9 ($c < 1$)
- Iterations at the same temperature: typically between 50-100
- Range $\gamma$: defines area around $\boldsymbol{x}^{[t]}$ in which next candidate solution set $\boldsymbol{x}^{[t+1]}$ is searched (depends strongly on objective function)

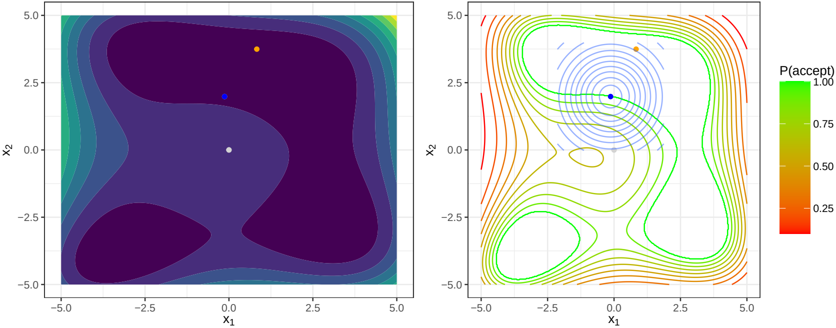
# EXAMPLE: SIMULATED ANNEALING




- Himmelblau's function has several local optima.
- We perform 100 iterations of simulated annealing with the following settings:
  - Proposal points are sampled from a normal distribution ($\sigma = 1.5$) around the current point
  - Initial temperature of $T^{[0]} = 200$
  - Constant temperature for the first 50 iterations
  - Afterwards, temperature drops by a multiplicative factor of $c = 0.8$ in every iteration

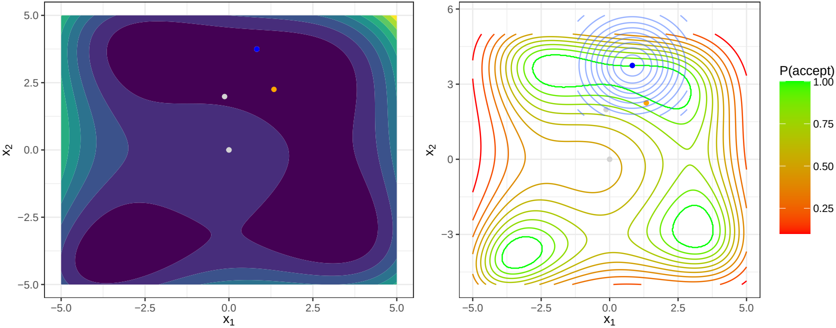
# EXAMPLE: SIMULATED ANNEALING




Left: Optimization surface of the Himmelblau function. Right: Acceptance probability $\mathbb{P}$(acceptance).

- The blue dot is the starting point $(0, 0)$ of optimization.
- For the first 50 iterations, the temperature is set to $T = 200$.
- A point is proposed (orange)
- In the beginning, almost every point is accepted (exploration)

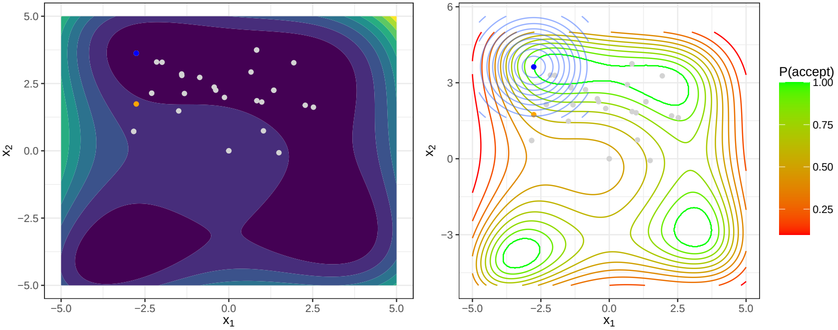
# EXAMPLE: SIMULATED ANNEALING




Left: Optimization surface of the Himmelblau function. Right: Acceptance probability $\mathbb{P}(\text{acceptance})$.

- The blue dot is the starting point $(0, 0)$ of optimization.
- For the first 50 iterations, the temperature is set to $T = 200$.
- A point is proposed (orange).
- In the beginning, almost every point is accepted (exploration).

# EXAMPLE: SIMULATED ANNEALING




Left: Optimization surface of the Himmelblau function. Right: Acceptance probability $\mathbb{P}$(acceptance).

- The blue dot is the starting point $(0, 0)$ of optimization.
- For the first 50 iterations, the temperature is set to $T = 200$.
- A point is proposed (orange)
- In the beginning, almost every point is accepted (exploration).
- Later, more points are rejected.