Global optimization

Given nonconvex \( f : \mathbb{R}^d \to \mathbb{R} \) and \( X \subset \mathbb{R}^d \), find the set \( S_\varepsilon := \left\{ x \in X, \text{ such that } f(x) \leq \min_{x \in X} f(x) + \varepsilon \right\} \).

Boltzmann distributions

- Boltzmann distributions: \( \pi_T(x) \propto \exp \left( -\frac{1}{T} f(x) \right) \)
- Cooling schedule: temperatures \( \{T_k\}_{k \in \mathbb{N}} \), with \( T_k \to 0 \)
- Concentration towards the global minimizer \( \pi_T(S_\varepsilon) \to 1 \).

Simulated Annealing (SA) [3]

- Metropolis-Hastings kernels: \( \pi_k P_k = \pi_k \)
- Proposals: \( \mu_{k+1} = \mu_k P_k \)
- Logarithmic cooling schedule: \( T_k = \frac{K}{\log(k+1)} \)
- Total variation norm convergence [1]
  \[ \|\pi_k - \mu_k\|_{TV} \to 0. \]

SA with parametric proposals [2]

- Parametric proposals: \( q_k, \theta_k \in \Theta \) (exponential family)
- Parameters are set through \( \theta_{k+1} = \arg \min_{\theta \in \Theta} KL(\pi_{T_k} \theta_k, q_\theta) \)
- This update is then approximated with samples from \( q_{\theta_k} \)
- Logarithmic cooling schedule needed for convergence.

The case for adaptive cooling schedule

- The logarithmic cooling schedule is slow.
- Intractable constants are often involved.
- The algorithm can possibly be stopped before \( T = 0 \).

Alternating proximal SA (APSA)

- Parameters are set through
  \[ \theta_k = \arg \min_{\theta \in \Theta} KL(\pi_{T_k} \theta, q_\theta) + \rho KL(q_{\theta_{k-1}}, q_\theta). \]
- Temperatures are set through
  \[ T_{k+1} = \arg \min_{T > 0} KL(\pi_T, q_{\theta_k}) + \lambda T^2 + \rho KL(\pi_{T_k}, \pi_T). \]

\( \{\pi_T, T > 0\} \)

Numerical results

A Rosenbrock-like objective is used:
\[ f(x) = 5(x_2 - x_1^2)^2 + (1 - x_1)^2, \quad \forall (x_1, x_2) \in \mathbb{R}^2. \]
Results are averaged over 1000 iterations, \( N = 500 \) samples per iteration and Gaussian proposals.

Figure: Comparison of APSA (blue), MARS [2] (red), SMCSA [5] (green) and mFSA [4] (purple), with \( \rho = 1 \) and \( \lambda = 5 \). MARS and SMCSA use a logarithmic schedule, while mFSA uses a faster one.

- APSA converges very fast to a low value of \( f \).
- The temperature decreases but stops before \( T = 0 \).

References