Proximal-based adaptive simulated annealing for global optimization #6676

Émilie Chouzenoux[†], Víctor Elvira[*‡] and Thomas Guilmeau[†]

† Université Paris-Saclay, CentraleSupélec, Inria, CVN, France,

* School of Mathematics, University of Edinburgh, United Kingdom ‡ The Alan Turing Institute, United Kingdom

Global optimization

Given nonconvex $f : \mathbb{R}^d \to \mathbb{R}$ and $\mathcal{X} \subset \mathbb{R}^d$, find the set

$$S_\epsilon := \left\{ x \in \mathcal{X}, \text{ such that } f(x) \leq \min_{x \in \mathcal{X}} f(x) + \epsilon \right\}.$$

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

$$\pi_{T_k}(S_\epsilon) \to 1.$$

Simulated Annealing (SA) [3]

- Metropolis-Hastings kernels: $\pi_{T_k} P_k = \pi_{T_k}$
- Proposals: $\mu_{k+1} = \mu_k P_k$
- Logarithmic cooling schedule: $T_k = \frac{K}{\log(k+1)}$
- Total variation norm convergence [1]

$$||\pi_{T_k} - \mu_k||_{TV} \to 0.$$

SA with parametric proposals [2]

- Parametric proposals: q_{θ_k}, $\theta_k \in \Theta$ (exponential family)
- Parameters are set through

$$\theta_{k+1} = \arg\min_{\theta \in \Theta} KL(\alpha_k \pi_{T_{k+1}} + (1 - \alpha_k) q_{\theta_k}, q_\theta)$$

- This update is then approximated with samples from q_{θ_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

$$\theta_k = \arg\min_{\theta \in \Theta} KL(\pi_{T_k}, q_\theta) + \rho KL(q_{\theta_{k-1}}, q_\theta).$$

- Temperatures are set through

$$T_{k+1} = \arg\min_{T > 0} KL(\pi_T, q_{\theta_k}) + \lambda T^2 + \rho KL(\pi_{T_k}, \pi_T).$$

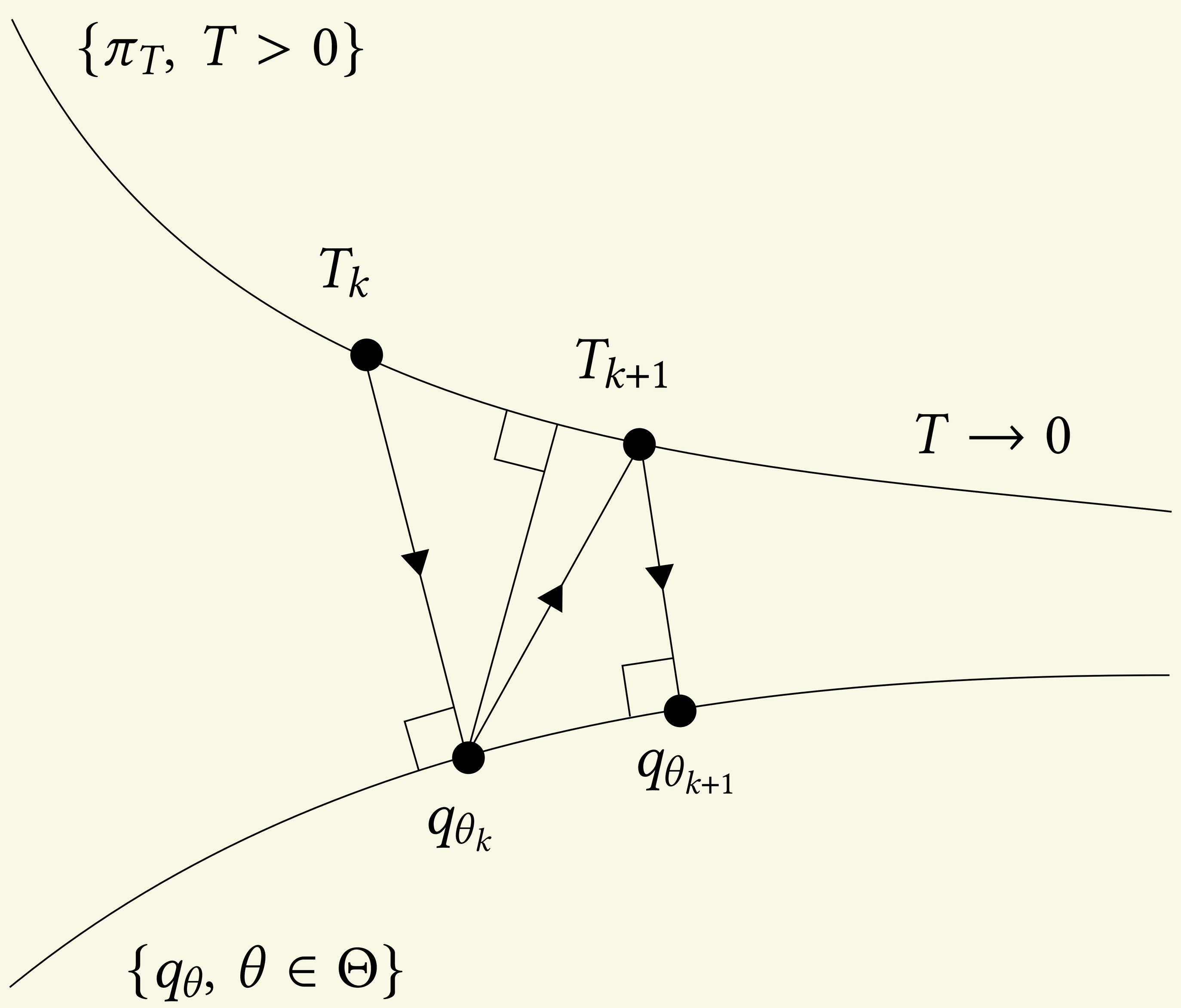

Figure: Geometric interpretation of APSA for $\rho = 0$ and $\lambda > 0$.

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.

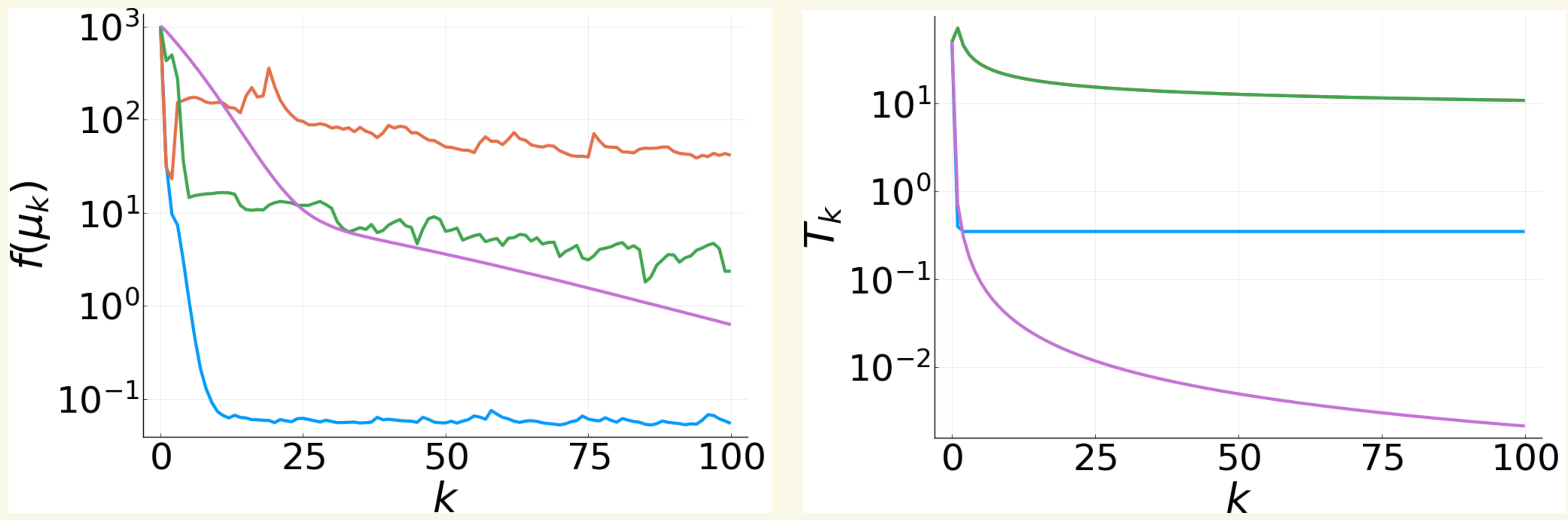

Figure: Comparison of APSA (blue), MARS [2] (red), SMCSA [5] (green) and mFSA [4] (purple), with $\rho = 1$ and $\lambda = 5$. MARS and SMCSA use a logarithmic schedule, while mFSA uses a faster one.

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

References

[1] H. Haario and E. Saksman. "Simulated Annealing Process in General State Space". In: *Advances in Applied Probability* 23.4 (1991), pp. 866–893.

[2] J. Hu and P. Hu. "Annealing adaptive search, cross-entropy, and stochastic approximation in global optimization". In: *Naval Research Logistics* 58.5 (2011).

[3] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. "Optimization by Simulated Annealing". In: *Science* 220.4598 (1983), pp. 671–680.

[4] S. Rubenthaler, T. Rydén, and M. Wiktorsson. "Fast simulated annealing in $\mathbb{R}^d$ with an application to maximum likelihood estimation in state-space models". In: *Stochastic Processes and their Applications* 119.6 (2009), pp. 1912–1931.

[5] E. Zhou and X. Chen. "Sequential Monte Carlo simulated annealing". In: *Journal of Global Optimization* 55 (2013), pp. 101–124.

icassp 2022 Singapore