Particle Swarm Optimzation Assisted Markov Chain Monte Carlo

Matthew Simpson*, Scott Holan and Christopher Wikle
*Postdoctoral Fellow, Department of Statistics, University of Missouri,
Supported by NSF and Census under NSF grant SES-1132031, funded through NCRN.

May 20, 2016
Spatial and Spatio-Temporal Design and Analysis for Official Statistics workshop
University of Missouri

Outline

- Heuristic optimization methods such as particle swarm optimization (PSO) allow for numerical optimization in higher dimensional spaces.
- Goal 1: develop better PSO algorithms using a tuning approach often used in MCMC algorithms.
- Goal 2: Use PSO to find posterior modes and use the Laplace approximation as a proposal for independent Metropolis-Hastings (IMH) and IMH within Gibbs (IMHwG) algorithms.

Particle swarm optimization (PSO) [2, 1]

- Goal: maximize $Q(\theta)$, $\theta \in \Theta \subseteq \Re^p$.
- Define particles $\theta_i \in \Theta$ with velocities $v_i \in \Theta$, $i = 1, 2, \dots, n$.
- Define a neighborhood $\mathcal{N}_i$ of "nearby" particles for each particle.
- Evolve the position of a particle over time towards 1) its personal best ($p_i \in \Theta$) and its neighborhood best ($g_i \in \Theta$).

▶ Standard PSO evolution equations:

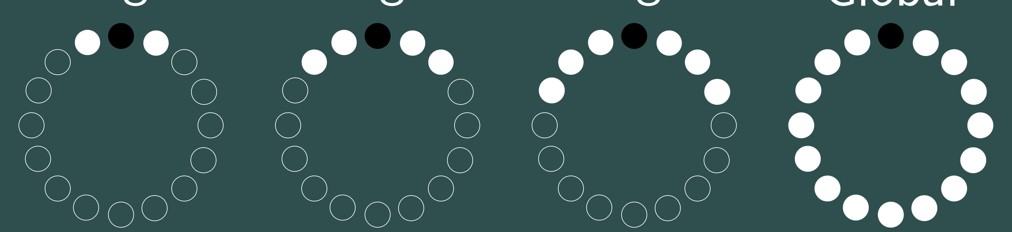
$$\begin{aligned}
\theta_i(t+1) &= \theta_i(t) + v_i(t)\\
v_i(t+1) &= \text{ inertia } + \text{ cognitive } + \text{ social}\\
&= \omega v_i(t) + \phi_1 r_{1i}(t) \circ [p_i(t) - \theta_i(t)]\\
&\quad + \phi_2 r_{2i}(t) \circ [g_i(t) - \theta_i(t)]\\
p_i(t+1) &= \begin{cases} p_i(t) & \text{if } Q(p_i(t)) \geq Q(\theta_i(t+1))\\ \theta_i(t+1) & \text{otherwise,}\end{cases}\\
g_i(t+1) &= \arg\max_{\{p_j(t+1) \mid j \in \mathcal{N}_i\}} Q(p_j(t+1))
\end{aligned}$$

- Parameters: scalars ω, ϕ_1, and ϕ_2 (good defaults known).
- Stochastic: $r_{1i}(t)$ & r_{2i} vectors of iid $U(0,1)$ r.v.'s.

Common neighborhoods

- Global: each particle is a neighbor of each other particle.
- Ring-k: arrange particles in a ring; each particle has k neighbors to the left and k to the right.

Ring-1 Ring-2 Ring-3 Global



Filled in white particles are neighbors of the filled in black particle.

Bare Bones PSO (BBPSO) [5]

▶ Simplify by removing the velocity term:

$$\theta_{ij}(t+1) \sim N\left(\frac{p_{ij}(t) + g_{ij}(t)}{2}, |p_{ij}(t) - g_{ij}(t)|\right) \tag{1}$$

for $j = 1, 2, \dots, p$. Updates for p_i and g_i as in PSO.

- BBPSO variants:
  - BBPSOxp: every iteration $\theta_{ij}(t+1)$ has 50% chance of moving according to (1) and a 50% chance of moving to $g_{ij}(t)$ [5].
  - BBPSO-MC: same as (1) except any particle currently at its neighborhood best moves according to
  $$\theta_{ij}(t+1) = p_{i_1 j}(t) + 0.5(p_{i_2 j}(t) - p_{i_3 j}(t))$$
  where i_1, i_2, and i_3 are distinct, randomly selected particles [7].
  - BBPSOxp-MC: combine both.

Adaptively tuned BBPSO (AT-BBPSO)

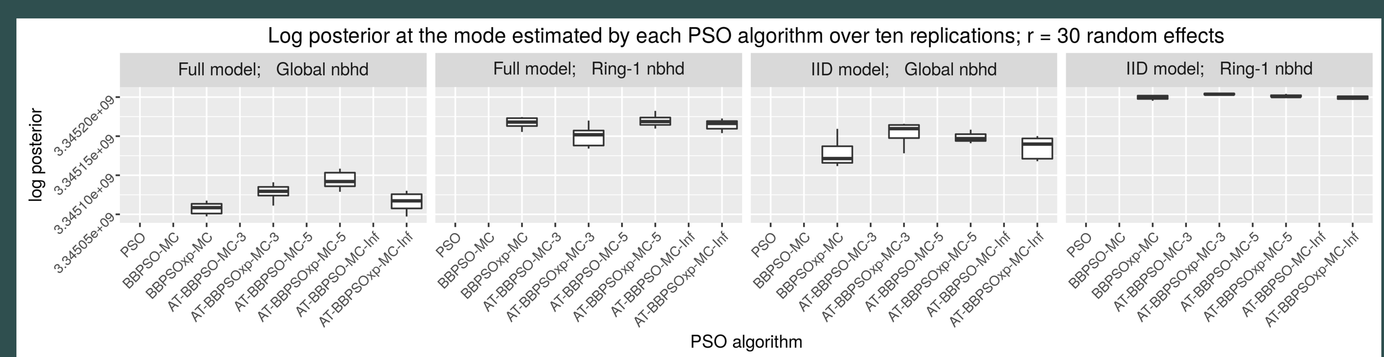
▶ Add scale parameter to BBPSO and tune it (tuned version of [3]):

$$\begin{aligned}
\theta_{ij}(t+1) &\sim T_{df}\left(\frac{p_{ij}(t) + g_{ij}(t)}{2}, |p_{ij}(t) - g_{ij}(t)| e^{\lambda(t)}\right)\\
&\text{for } j = 1, 2, \dots, p\\
R(t+1) &= \#\{p_i(t+1) \neq p_i(t) : i = 1, 2, \dots, n\}/n\\
\lambda(t+1) &= \lambda(t) + 0.1 \times sgn(R(t+1) - R^*)
\end{aligned}$$

- The target acceptance rate R^* roughly controls exploitation vs. exploration. $R^* = 0.2$ or 0.3 seems to work well — similar to a random walk Metropolis acceptance rate.
- The degrees of freedom parameter df is harder to interpret, but generally small is good. E.g. $df = 1$ or $df = 3$.
- AT-BBPSOxp-MC w/ Ring-1 & above settings often works well.

Spatially smoothing ACS county population estimates

- Data model: $z_k \sim Pois\left(e^{x_k'\beta + s_k'\delta}\right)$; Fixed effects: $x_k' = 1$.
- Random effects: $s_k' : 1 \times r$ from truncated Moran's I basis [4, 6].
- IID model: $\delta \sim N(0, \sigma^2 \mathrm{I})$; Full model: $\delta \sim N(0, \Sigma)$.



MCMC Algorithms

- PSO-IMH: find posterior mode via PSO and use Laplace approximation as a proposal (a T_ν distribution).
- PSO-IMH within Gibbs (PSO-IMHwG):
  - Conditionally conjugate step for Σ or σ^2.
  - IMH for (β, δ) using conditional distribution implied by T_ν approximation to the full posterior as a proposal (obtained via PSO).
- Block RW within Gibbs (B-RWwG) w/ conjugate draw for Σ or σ^2.

▶ ν: degrees of freedom in T_ν proposal for PSO-IMH and PSO-IMHwG.
  - Choose to optimize the Metropolis acceptance rate.
  - For these models, $\nu = \infty$ — i.e. a Gaussian proposal.

MCMC Results

IID Model

	n_{eff}				time/n_{eff}			
r	IMH	IMHwG	RWwG	B-RWwG	IMH	IMHwG	RWwG	B-RWwG
10	23170	46177	8072	1168	24	23	201	146
20	16958	43005	5739	646	27	26	506	215
30	30237	39739	4440	404	16	24	790	483

Full Model

	n_{eff}				time/n_{eff}			
r	IMH	IMHwG	RWwG	B-RWwG	IMH	IMHwG	RWwG	B-RWwG
5	32	47240	8089	2070	14433	21	131	170
7	37	42459	7811	1743	11188	20	145	167
9	9	717	8298	1417	32197	876	126	153

Effective sample size (n_{eff}) and time in seconds per 10,000 effective samples (time/n_{eff}) for each algorithm in both models with various numbers of random effects (r). IMH algorithms with tiny acceptance rates are indicated by a "—".

- In the IID model the log variance is approximately normal $\Longrightarrow$ PSO-IMH algorithms work well.
- In the Full model we only have one "observation" with covariance Σ $\Longrightarrow$ normal approximation is bad & PSO-IMH works poorly.

References

[1] Blum, C. and Li, X. (2008). Swarm intelligence in optimization. In Blum, C. and Merkle, D., editors, *Swarm Intelligence: Introduction and Applications*. Springer.

[2] Clerc, M. and Kennedy, J. (2002). The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *Evolutionary Computation, IEEE Transactions on*, 6(1):58–73.

[3] Hsieh, H.-I. and Lee, T.-S. (2010). A modified algorithm of bare bones particle swarm optimization. *International Journal of Computer Science Issues*, 7:11.

[4] Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):139–159.

[5] Kennedy, J. (2003). Bare bones particle swarms. In *Swarm Intelligence Symposium, 2003. SIS'03. Proceedings of the 2003 IEEE*, pages 80–87. IEEE.

[6] Porter, A. T., Holan, S. H., and Wikle, C. K. (2015). Bayesian semiparametric hierarchical empirical likelihood spatial models. *Journal of Statistical Planning and Inference*, 165:78–90.

[7] Zhang, H., Kennedy, D. D., Rangaiah, G. P., and Bonilla-Petriciolet, A. (2011). Novel bare-bones particle swarm optimization and its performance for modeling vapor–liquid equilibrium data. *Fluid Phase Equilibria*, 301(1):33–45.

www.thematsimpson.com
thematsimpson@gmail.com