WARWICK

Guided sequential ABC schemes for intractable Bayesian models

Massimiliano Tamborrino, Dept.of Statistics, Warwick, warwick.ac.uk/tamborrino

Joint work with Umberto Picchini (Chalmers University of Technology and the University of Gothenburg, Sweden)

U. Picchini, M. Tamborrino. Guided sequential ABC schemes for intractable Bayesian models. arXiv:2206.12235, 2022.

We want to construct ways to improve sequential ABC schemes.

SMC-ABC can be "slow" in the sense that the cloud of parameters "particles" can be very diffuse, especially in the initial iterations.

This is typical as the initial proposal sampler is the (possibly vague) prior $\rightarrow$ very different from the posterior $\rightarrow$ low acceptance rate.

We explore how to make parameter proposals *guided* by the observed data to reduce rejections and hence the computational effort in SIS-ABC and SMC-ABC while keeping their accuracy.

We have:

- unknown parameters θ
- observations y but we consider information encoded into summary statistics $s_y = S(y)$;
- a generative model implicitly simulating from the likelihood $p(y|\theta)$
- prior $\pi(\theta)$
- a proposal $\theta^* \sim q(\cdot|\theta)$

We wish to sample from an approximation $\pi_\delta(\theta|s_y)$ of the posterior $\pi(\theta|s_y)$, where

$$\pi_\delta(\theta|s_y) \propto \pi(\theta) \int \mathbb{I}(d(s^*, s_y) < \delta) p(s^*|\theta) ds^*$$

with $s^* \sim p(s^*|\theta)$ simulated summaries from the generative model.

In the following, $d(s^*, s_y) = ||s^* - s_y||$.

How to simulate the proposal θ^*?

Several possibilities[1] to simulate the proposal θ^*. The most common are:

- (inefficient but parallelisable) acceptance-rejection ABC, proposes from $\theta^* \sim \pi(\theta)$;
- MCMC-ABC (serial, difficult to tune the δ), high-autocorrelations, proposes from kernel $\theta^* \sim q(\theta^*|\theta)$;
- SIS-ABC and SMC-ABC (parallelisable, more plug-and play than MCMC-ABC).

Today we exclusively discuss sequential ABC (SMC-ABC and SIS-ABC)

[1] Sisson, Fan, Beaumont. *Handbook of Approximate Bayesian Computation*, 2018.

SMC-ABC introduces T approximations to the desired posterior and moves a "cloud of particles" between the T approximations as:

- $(t = 1)$ run acceptance-rejection ABC, getting particles $\theta_1^{(1)}, ..., \theta_1^{(N)}$. Assign weights to each particle $w_1^{(i)} = 1/N$.
- **for** t=2:T
  for i=1:N
    repeat until acceptance:
      sample a θ_k^* from $\theta_{t-1}^{(1)}, ..., \theta_{t-1}^{(N)}$ using weights $w_{t-1}^{(1)}, ..., w_{t-1}^{(N)}$.
      (perturb) $\theta_k^{**} \sim q_t(\cdot|\theta_k^*)$
      simulate $z \sim p(z|\theta_k^{**})$, get summaries $s^* = S(z)$
      accept and store $\theta_t^{(i)} := \theta_k^{**}$ if $||s^* - s_y|| < \delta$
      if accept, set $w_t^{(i)} = \pi(\theta_t^{(i)}) / \sum_{j=1}^{N} w_{t-1}^{(j)} q_t(\theta_t^{(i)}|\theta_{t-1}^{(j)})$ and normalise it
    end
  end
end

Output: weighted particles $(\theta_T^{(1)}, \ldots, \theta_T^{(N)}) \sim \pi_{\delta_T}(\theta|s_y)$.

The proposal sampler does not have to be Gaussian, it can be an arbitrary $q_t(\cdot|\theta)$.

In practice, the most popular proposal sampler is a multivariate Gaussian centred at a randomly picked particle chosen from the previous iteration

$$\theta_t^{**} \sim q_t(\cdot|\theta^*) = N(\theta_{t-1}^*, 2\Sigma_{t-1})$$

- Beaumont, Cornuet, Marin, Robert. Biometrika, 96(4):983–990, 2009.
- Toni, Welch, Strelkowa, Ipsen, and Stumpf. Journal of the Royal Society Interface, 6(31):187–202, 2008.
- Filippi, Barnes, Cornebise, and Stumpf. Statistical applications in genetics and molecular biology, 12(1):87–107, 2013.

What we do here is to construct ways to incorporate data information s_y in the proposal sampler, so to propose from some $q_t(\cdot|s_y, ...)$.

Therefore we construct **guided sequential proposal samplers**, in the sense that they are guided by data.

Disclaimer: The idea of using information from data summaries s_y to adjust the output of an ABC procedure is not new (e.g., Beaumont et al. [2002], Blum and Francois [2010] and Li et al. [2017]).

Key difference:
- those approaches do not use s_y to improve the proposal sampler while the ABC procedure is running, but only adjust the final output, acting on the already accepted parameters.

- we aim to make use of s_y to guide the parameter proposals *while ABC is running.*

Guided SIS-ABC

1: **for** $i = 1, \ldots, N$ **do**
2: **repeat**
3: $\theta^* \sim q_t(\theta_{t-1}) = \mathcal{N}(\theta_{t-1}, 2\Sigma_{t-1})$.
4: *...etc*
5: **until** $||s^* - s_y|| < \delta$
6: **end for**

Goal : From : $\theta^* \sim q_t(\theta_{t-1})$ to $q_t(\theta_{t-1}|s_y)$

We build upon Picchini, Simola and Corander[2] who noted the following when producing a synthetic likelihoods MCMC proposal sampler:

- denote by $(\theta^{(i)}, s^{(i)})$ a (parameter,summary)-particle accepted at the previous iteration;
- assume that $(\theta^{(i)}, s^{(i)}) \sim \mathcal{N}(m, S)$, with

$$m \equiv (m_\theta, m_s), \qquad S_{t-1} \equiv \begin{bmatrix} S_\theta & S_{\theta s} \\ S_{s\theta} & S_s \end{bmatrix},$$

- Then,

$$\theta^{(i)}|s^{(i)} \sim \mathcal{N}(m_{\theta^{(i)}|s^{(i)}}, S_{\theta^{(i)}|s^{(i)}})$$

[2] Picchini, Simola, Corander. Adaptive MCMC for synthetic likelihoods and correlated synthetic likelihoods, Bayes. Analy. 2022, *Tuesday, 1.30pm*.

The guided sampler we have defined is useful for a sequential importance sampling ABC (SIS-ABC).

We use it for SIS-ABC and not for SMC-ABC because the latter typically considers *local* features, by perturbing a resampled particle.

Instead the sampler we defined has *global features*, since both the mean and the covariance are not particle-specific.

$$q_t(\theta|s_y) \equiv \mathcal{N}(\hat{m}_{\theta|s_y,t-1}, \hat{S}_{\theta|s_y,t-1})$$

$$\hat{m}_{\theta|s_y,t} = \hat{m}_\theta + \hat{S}_{\theta s}(\hat{S}_s)^{-1}(s_y - \hat{m}_s) \quad (1)$$

$$\hat{S}_{\theta|s_y,t} = \hat{S}_\theta - \hat{S}_{\theta s}(\hat{S}_s)^{-1}\hat{S}_{s\theta}. \quad (2)$$

We call this sampler **blocked** because it proposes in block all coordinates of θ.

Guided SMC-ABC

We have built also other proposals. One is named **fullconditional** and is specific for SMC-ABC since this time we do condition on a sampled particle AND s_y.

Recall in SMC-ABC we have a resampling step (snippet below):

1: **for** $i = 1, \ldots, N$ **do**
2: **repeat**
3: pick (with replacement) θ^* from the weighted set $\{\theta_{t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^{N}$.
4: $\theta^{**} \sim \mathcal{N}(\theta^*, \Sigma_{t-1})$.
5: *...etc*
6: **until** $||s^* - s_y|| < \delta$
7: **end for**

We are going to sample θ^{**} conditionally on components of θ^* and s_y.

Guided SMC-ABC

As usual we iteratively resample a θ^* from the previous iteration.

How can we condition on both s_y and θ^*?

We cannot condition simultaneously on all components of θ^*: imagine instead decomposing $\theta^* = (\theta^*_k, \theta^*_{-k})$ and write

$$\begin{bmatrix} \theta^*_k \\ \theta^*_{-k} \\ s_y \end{bmatrix} \tag{3}$$

We now place a multivariate Gaussian assumption on (3) so we can produce a "perturbation kernel"

$$q_t(\theta^*_k | \theta^*_{-k}, s_y)$$

that is proposing for the k-th component of θ, conditionally on the remaining coordinates set to the coordinates of the picked θ^*.

Notice k can be a set of indices instead of a single index.

Guided SMC-ABC

It is easy to show that we can write the following guided sampler (this is for the case where k is a single index, but can be generalized)

$$\theta_k^{**} \sim \mathcal{N}(\hat{m}^*_{k|s_y,t-1}, \hat{\sigma}^2_{k|s_y,t-1})$$

$$\hat{m}^*_{k|s_y,t-1} = \hat{m}_k + \hat{S}_{k,-k}(\hat{S}_{-k,-k})^{-1}\left(\begin{bmatrix}\theta^*_{-k} \\ s_y\end{bmatrix} - \begin{bmatrix}\hat{m}_{-k} \\ \hat{m}_s\end{bmatrix}\right),$$
$$\hat{\sigma}^2_{k|s_y,t-1} = \hat{\sigma}^2_k - \hat{S}_{k,-k}(\hat{S}_{-k,-k})^{-1}\hat{S}_{-k,k},$$

where all quantities are computed using accepted particles from the previous iteration.

Unlike the previous **blocked** SIS-ABC, here we conditioning on a randomly sampled particle from the previous population θ^*_{-k}.

We call this **fullcond** guided SMC-ABC sampler ("fully conditional").

A few important remarks

1. **blocked** and **fullcond** may be mode seeking $\Rightarrow$ We developed two variants where the covariance matrix is suitably optimised $\Rightarrow$ **blockedopt** and **fullcondopt**

2. Experiments suggest that the Gaussianity assumption on the joint (θ, s_y) works very well even with highly non Gaussian posterior targets.

To allow for more flexibility, we considered **(Gaussian and t) copulas with different marginals**.

i. What is a copula and how do we use it for **blocked**?

$$\begin{aligned} q_t(\theta|s_y) &\equiv \mathcal{N}(\hat{m}_{\theta|s_y,t-1}, \hat{\Sigma}_{\theta|s_y,t-1}), \\ \hat{m}_{\theta|s_y,t-1} &= \hat{m}_\theta + \hat{S}_{\theta s}(\hat{S}_s)^{-1}(s_y - \hat{m}_s) \\ \hat{S}_{\theta|s_y,t-1} &= \hat{S}_\theta - \hat{S}_{\theta s}(\hat{S}_s)^{-1}\hat{S}_{s\theta}. \end{aligned}$$

ii. Same as Y. Chen, M. Gutmann, *Adaptive Gaussian Copula ABC*, AISTATS, 2019?

We go through a couple of examples now.

We compare against several SMC-ABC and SIS-ABC:

- **standard**: Gaussian sampler $\mathcal{N}(\theta^*, 2 \cdot \mathrm{Cov}_{t-1})$ (as in Filippi et al. 2013[3], by generalizing Beaumont et al);
- **olcm**: Gaussian sampler with the optimal local covariance matrix of Filippi et al. 2013

- **blocked**: our guided SIS-ABC approach;
- **blockedopt**: same mean as guided SIS-ABC **blocked** but optimal local covariance;
- **hybrid**: **blocked** at step $t = 1$ and then **blockedopt**.

- **cop-blocked**: our copula-based guided SIS-ABC approach;
- **cop-blockedopt**: same mean as copula-based guided SIS-ABC **cop-blocked** but optimal local covariance;
- **cop-hybrid**: **cop-blocked** at step $t = 1$ and then **cop-blockedopt**.

- **fullcond**: our guided SMC-ABC approach;
- **fullcondopt**: same mean as guided SMC-ABC **fullcond** but optimal local covariance.

---

[3] Filippi, Barnes, Cornebise, and Stumpf. Statistical Applications in Genetics and Molecular Biology, 12(1):87–107, 2013.

Example 1: Twisted prior with highly correlated posterior

* Observed data $y = (y_1, ..., y_{d_\theta}) \sim \mathcal{N}(\theta, \Psi)$, with $\theta = (\theta_1, ..., \theta_{d_\theta}), \Psi = \mathrm{diag}(\sigma_0, ..., \sigma_0))$.

* The prior is the *twisted-normal* prior with density function prop. to

$$\pi(\theta) \propto \exp\left\{ -\frac{\theta_1^2}{200} - \frac{(\theta_2 - b\theta_1^2 + 100b)^2}{2} - \sum_{j=3}^{d_\theta} \theta_j^2 \right\}.$$

* Here: $\sigma_0 = 1, b = 0.1, d_\theta = 5, y = (10, 0, 0, 0, 0)$.

As before, $S(y) = y$.

* Automatically decreasing ABC-tolerance δ:
  - $\delta_1 = 50$.
  - δ decreases across iterations by taking first percentile of the distances.
  - the algorithm stops as $\delta < 0.25$.

How are our approaches coping with highly correlated posteriors?

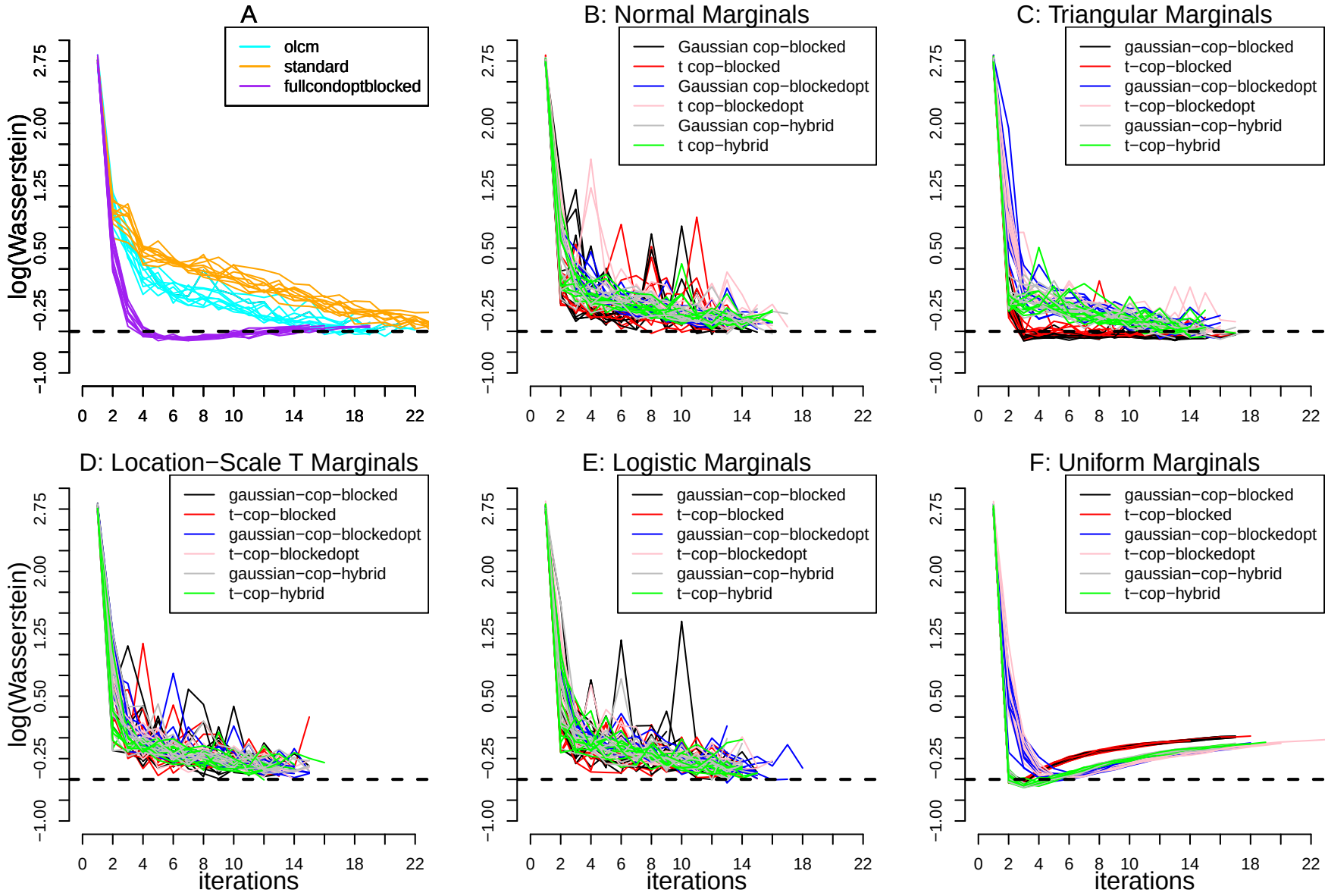


Figure 1: Log-Wasserstein distances of all ten independent estimations at each iteration ($N = 1{,}000$)

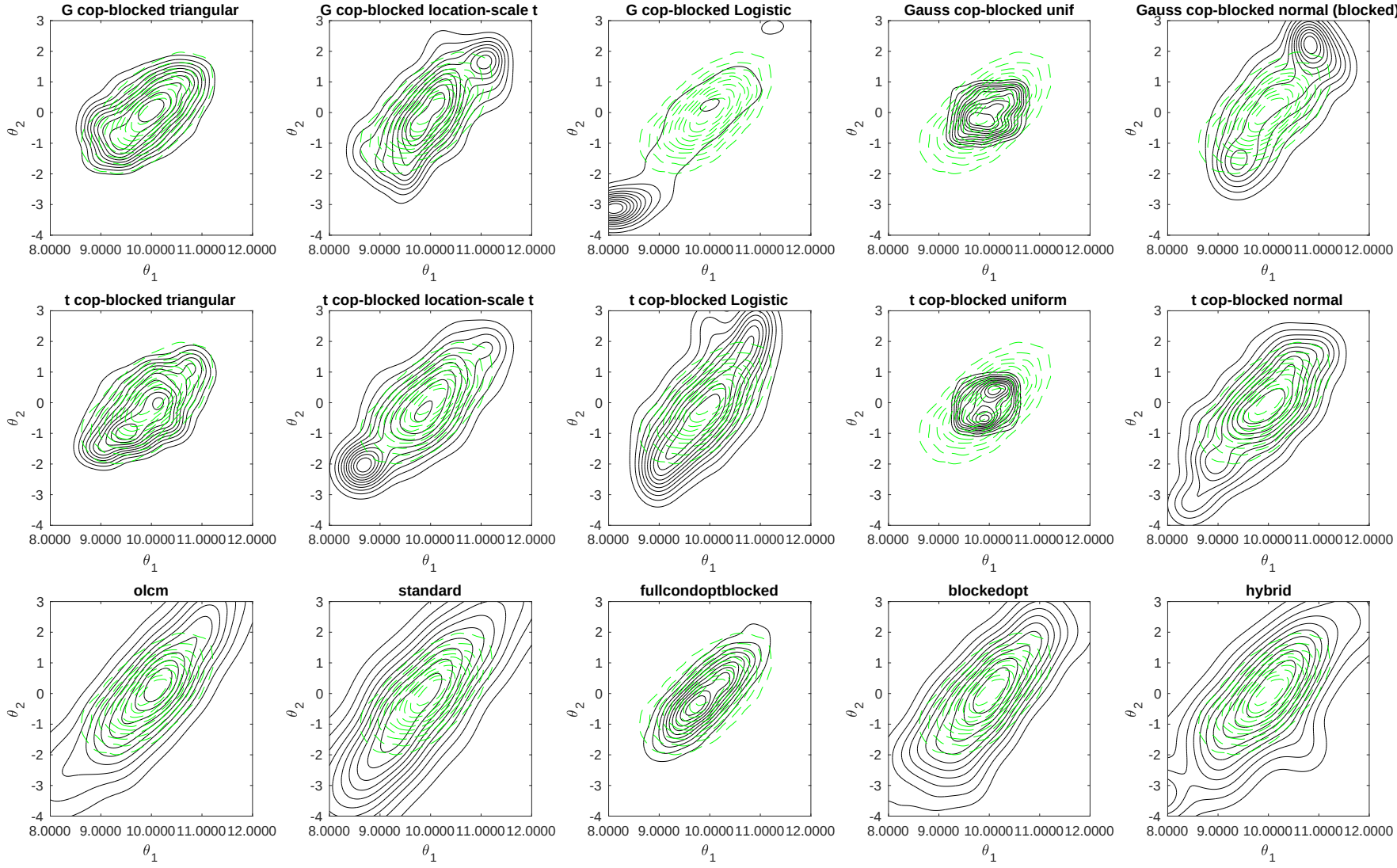
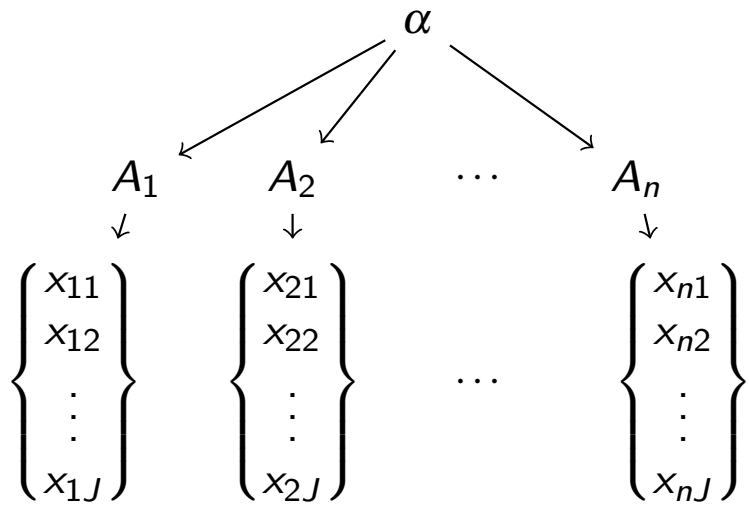


Figure 2: Twisted: Contour plots of the posterior marginal at the 2nd iteration whose acceptance rate is below 1.5% (similar to Del Moral et al., Stat. Comput. 2012).

Example 2: Hierarchical g-and-k model[4]

Observations $x_{ij} = (x_{i1}, \ldots, x_{iJ}), i = 1, \ldots, n$ from hierarchical g-and-k model:

- $x_{i1}, \ldots, x_{iJ}$ iid from from gk model with (A_i, B, g, k, c) parameters with B, g, k, c known constants.
- $A_i \sim N(\alpha, 1), \theta = (\alpha, A_1, \ldots, A_n)$.



- Here: $n = 20, J = 1{,}000$.
- 21-dim. parameter $\theta = (\alpha, A_1, \ldots, A_{20})$.
- $s(x_i) = quant(x_i; l = 8)(l = 0, \ldots, 8)$, where $quant(x, p)$ is the p-th quantile of $x \Rightarrow$ 180 summaries.

[4] Clarté et al. Componentwise approximate Bayesian computation via Gibbs-like steps. Biometrika, 2021.

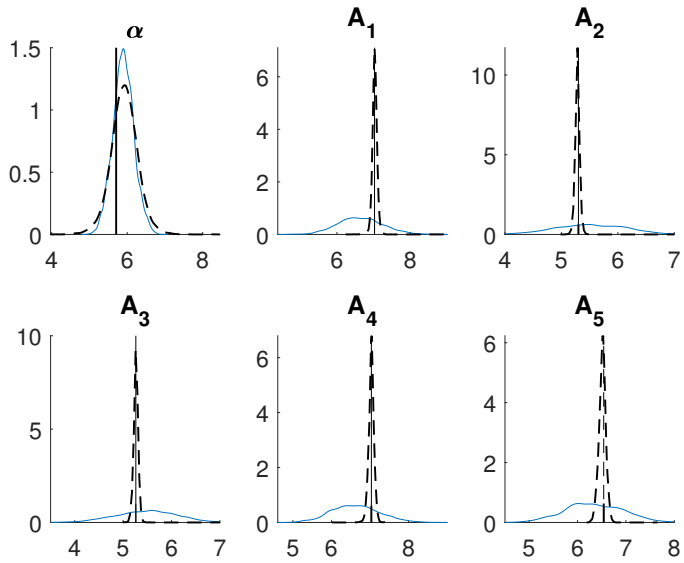
* Automatically decreasing ABC-tolerance δ:
  - $\delta_1 = 50$.
  - δ decreases across iterations by taking 5th percentile of the distances.
  - the algorithm stops as $\delta < 0.62$.

We run `olcm`, `standard`, `hybrid` with $N = 10^4$ particles and compare them with ABC-Gibbs (Clarté et al.,2021).

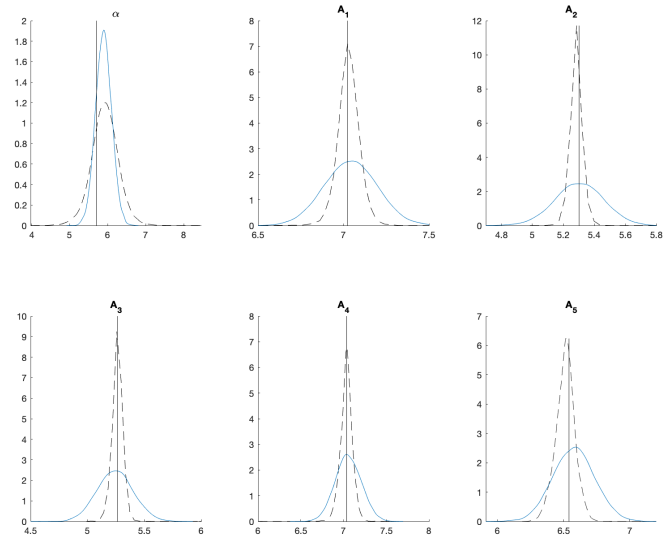
How are our approaches coping with high dimensional parameter and summaries spaces?

A few numbers for one run:

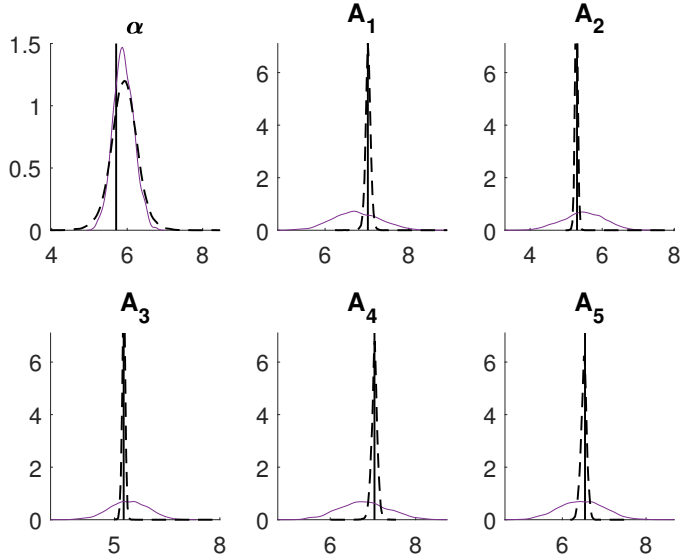
- `standard`: 14.5M simulations to reach $\delta = 2.14$ (t=26) in more than 55h.
- `olcm`: 16M simulations to read $\delta = 0.70$ (t=41) in 55h.
- `hybrid`: 1M simulations to go below $\delta = 0.62$ in 6 minutes!



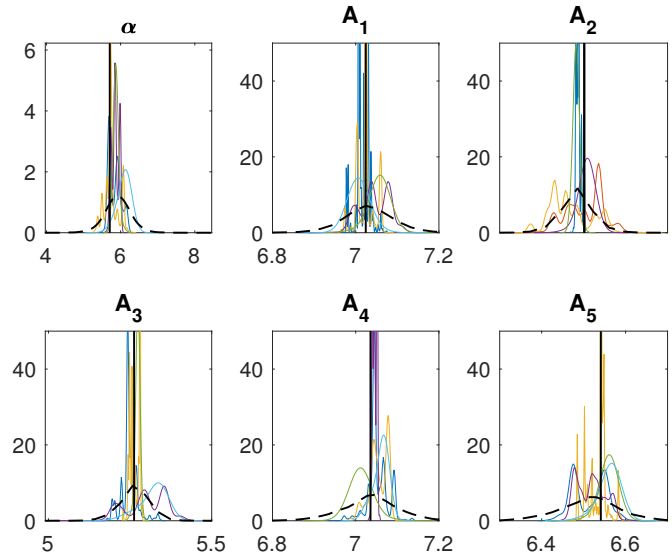
(a) (non guided) `standard` at iteration $t = 26 (\delta = 2.14)$ after 14.5M simulations



(b) (non guided) `olcm` at iteration $t = 41 (\delta = 0.7)$ after 16M simulations.



(c) (non guided) `olcm` after 1M simulations



(d) (guided) `hybrid` (5 runs) after 1M simulations

Take-home message:

- We propose guided samplers making use of s_y while ABC is running.
- We propose different samplers, Gaussian and not, based on local or global features of the particles.
- For copula-based samplers, the mean/covariance/marginals are not learned via DL but chosen a-priori.

A few remarks

- The proposed guided schemes perform at least as well as `standard` and `olcm`, with higher acceptance rate and a (way) lower running time.
- Among the copula-based guided sequential samplers, we recommend
  - `cop-blocked` with triangular marginals.
  - `cop-hybrid` with uniform marginals.
- The EES of the guided samplers are not as high as `olcm,standard`. There's margin for improvement.
- More examples (e.g. one with > 400 summaries) in the ArXived manuscript.
- R and Matlab codes to be uploaded on GitHub soon!

U. Picchini, M. Tamborrino. Guided sequential ABC schemes for intractable Bayesian models. arXiv:2206.12235, 2022.

Thank you for your attention.