Supporting Material for 'Diffusion Tempering Improves Parameter Estimation with Probabilistic Integrators for Ordinary Differential Equations'

1 Different schedules

In Fig. 1 we illustrate the comparative performance of different schedules.

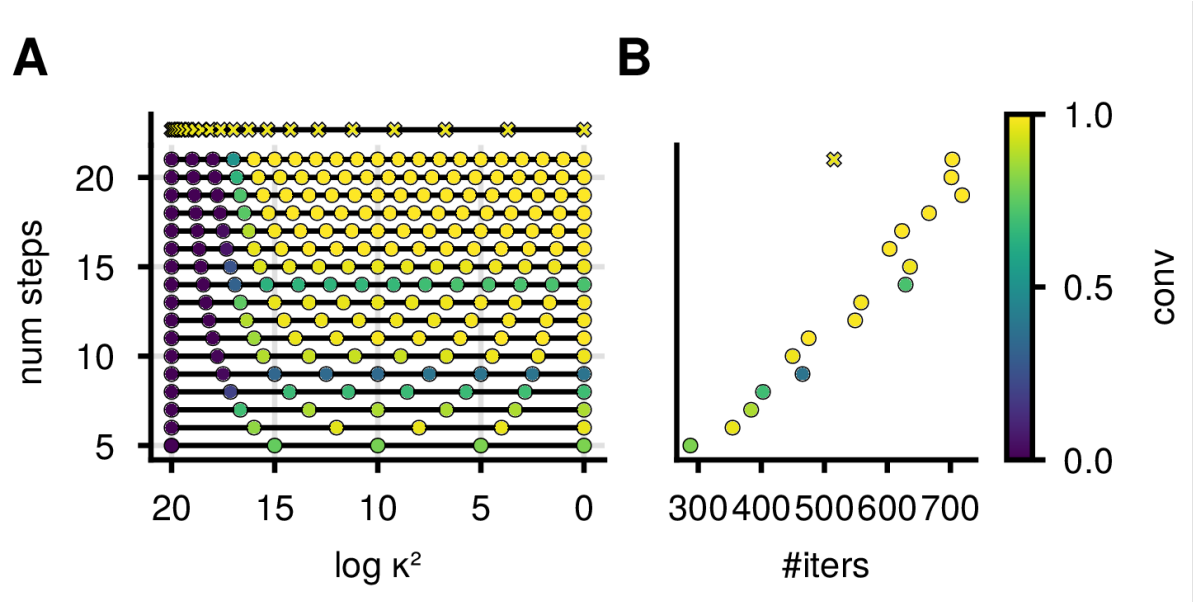

Figure 1: Comparison of different schedules. **A** Percentage of correctly converged runs at every step of the tempering schedule for different schedules. Circles, $\log \kappa^2$ was reduced linearly. Crosses, $\log \kappa^2$ was reduced with an exponential decay. **B** Cost of each schedule and final convergence.

2 Effect of early stopping

Tab.1 shows the effect of early stopping for the pendulum (PD), Lotka Volterra (LV) model and Hodgkin-Huxley (HH) model. The number of parameters is denoted in the model's index.

model	**conv**	**iter**
PD_1	1.00	303.89 ± 12.43
LV_2	0.77	411.18 ± 151.25
LV_4	0.43	727.15 ± 482.93
HH_2	**1.00**	696.22 ± 78.10

model	**conv**	**iter**
PD_1	1.00	$\mathbf{92.21 \pm 0.95}$
LV_2	**0.79**	$\mathbf{132.43 \pm 31.01}$
LV_4	**0.44**	$\mathbf{292.05 \pm 139.98}$
HH_2	0.96	$\mathbf{197.81 \pm 40.38}$

Table 1: Effect of early stopping on the cost of compute. **left** Without early stopping. **right** With early stopping.

3 Updated Table 1

Tab.2 shows the updated Table 1.

ODE	D_θ	ALG	ITER	pRMSE	CONV	tRMSE	N_θ
PD	1	Fenrir	74.12 ± 11.55	0.01 ± 0.08	0.99	0.04 ± 0.26	0.99 ± 0.10
PD	1	RK	$\mathbf{34.11 \pm 18.45}$	1.42 ± 1.09	0.32	0.98 ± 0.72	0.32 ± 0.47
PD	1	ours	303.89 ± 12.43	$\mathbf{0.00 \pm 0.00}$	**1.00**	$\mathbf{0.02 \pm 0.00}$	$\mathbf{1.00 \pm 0.00}$
LV	2	Fenrir	$\mathbf{60.68 \pm 36.73}$	1.41 ± 1.01	0.20	2.84 ± 2.71	0.40 ± 0.80
LV	2	RK	97.15 ± 146.95	1.35 ± 1.11	0.24	2.35 ± 1.32	0.48 ± 0.86
LV	2	ours	411.18 ± 151.25	$\mathbf{0.41 \pm 0.82}$	**0.77**	$^*\mathbf{0.54 \pm 1.14}$	$\mathbf{1.54 \pm 0.85}$
LV	4	Fenrir	$\mathbf{112.45 \pm 47.24}$	1.31 ± 0.94	0.23	$\mathbf{4.18 \pm 13.65}$	1.07 ± 1.70
LV	4	RK	271.49 ± 151.05	1.09 ± 0.76	0.24	6.57 ± 34.45	0.97 ± 1.71
LV	4	ours	727.15 ± 482.93	0.81 ± 0.88	**0.43**	$^*\mathbf{10.41 \pm 69.37}$	$\mathbf{1.76 \pm 1.98}$
HH_1	1	Fenrir	46.96 ± 19.08	0.38 ± 0.67	0.68	7.85 ± 10.70	0.68 ± 0.47
HH_1	1	RK	$\mathbf{43.30 \pm 43.45}$	0.42 ± 0.48	0.57	7.54 ± 8.26	0.57 ± 0.50
HH_1	1	ours	382.08 ± 32.19	$\mathbf{0.00 \pm 0.00}$	**1.00**	$\mathbf{0.43 \pm 0.02}$	$\mathbf{1.00 \pm 0.00}$
HH_1	2	Fenrir	110.04 ± 61.70	0.20 ± 0.37	0.75	5.89 ± 10.15	1.53 ± 0.83
HH_1	2	RK	$\mathbf{54.02 \pm 62.60}$	0.28 ± 0.45	0.72	4.88 ± 7.58	1.44 ± 0.90
HH_1	2	ours	696.22 ± 78.10	$\mathbf{0.00 \pm 0.00}$	**1.00**	$\mathbf{0.42 \pm 0.04}$	$\mathbf{2.00 \pm 0.00}$
HH_1	3	Fenrir	$\mathbf{122.15 \pm 49.74}$	0.59 ± 0.65	0.51	9.31 ± 9.63	1.53 ± 1.51
HH_1	3	RK	223.55 ± 117.31	0.90 ± 0.25	0.03	14.33 ± 4.01	0.13 ± 0.56
HH_1	3	ours	676.33 ± 150.72	$\mathbf{0.01 \pm 0.10}$	**0.99**	$\mathbf{0.60 \pm 1.51}$	$\mathbf{2.97 \pm 0.30}$
HH_1	6	Fenrir	$\mathbf{108.06 \pm 108.49}$	13.36 ± 6.97	0.00	26.21 ± 7.38	1.05 ± 0.22
HH_1	6	RK	210.26 ± 120.86	12.27 ± 6.87	0.00	16.80 ± 3.81	1.18 ± 0.39
HH_1	6	ours	2159.60 ± 532.55	$\mathbf{10.36 \pm 7.72}$	0.00	$\mathbf{15.20 \pm 5.41}$	$\mathbf{1.21 \pm 0.46}$
HH_2	4	Fenrir	286.54 ± 205.37	0.28 ± 0.44	0.68	12.00 ± 17.37	2.80 ± 1.78
HH_2	4	RK	$\mathbf{136.50 \pm 200.20}$	0.43 ± 0.56	0.50	7.98 ± 9.86	2.08 ± 1.81
HH_2	4	ours	1492.03 ± 335.17	$\mathbf{0.00 \pm 0.00}$	**1.00**	$\mathbf{0.60 \pm 0.01}$	$\mathbf{4.00 \pm 0.00}$
HH_2	6	Fenrir	$\mathbf{221.28 \pm 144.56}$	0.62 ± 0.72	0.50	13.01 ± 13.10	3.06 ± 2.96
HH_2	6	RK	390.34 ± 195.85	0.88 ± 0.22	0.00	19.36 ± 5.54	0.31 ± 0.75
HH_2	6	ours	1525.57 ± 448.56	$\mathbf{0.12 \pm 0.32}$	**0.88**	$^*\mathbf{3.01 \pm 6.70}$	$\mathbf{5.28 \pm 1.96}$

Table 2: Comparison of different methods, models and model sizes for parameter estimation. LV: Lottka-Volterra, PD: Pendulum, HH_x Hodgkin–Huxley with x compartments, D_θ: Number of parameters, N_θ: The number of correctly identified parameters, CONV: correctly converged. *Runs with a final loss of NaN where excluded from computation if marked with* *.

4 Noisy gradients

In Tab.3 we show the effect of adding Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma_0^2 e^{-t/\tau})$ to the gradients of the RK baseline, with initial variance σ_0^2 and decay constant τ.

σ_0	τ	**conv**	**pRMSE**	**iter**
0.00	-	0.32	1.42 ± 1.09	33.47 ± 18.11
0.01	-	0.35	1.40 ± 1.11	33.79 ± 12.84
0.05	-	0.34	1.41 ± 1.10	29.89 ± 10.77
0.10	-	0.37	1.40 ± 1.13	27.30 ± 11.84
0.50	-	0.49	1.14 ± 1.13	15.19 ± 10.72
1.00	-	0.63	0.77 ± 1.05	15.52 ± 14.12
5.00	-	**0.64**	$\mathbf{0.51 \pm 0.88}$	13.01 ± 8.87
10.00	-	0.39	0.64 ± 0.93	$\mathbf{10.43 \pm 7.19}$
50.00	-	0.45	0.49 ± 0.83	11.66 ± 8.26
100.00	-	0.28	0.44 ± 0.74	10.26 ± 7.42
0.00	5.00	0.32	1.42 ± 1.09	33.47 ± 18.11
0.01	5.00	0.31	1.45 ± 1.08	35.94 ± 18.58
0.05	5.00	0.33	1.45 ± 1.10	32.58 ± 13.39
0.10	5.00	0.35	1.40 ± 1.11	33.65 ± 12.85
0.50	5.00	0.43	1.24 ± 1.13	22.23 ± 14.93
1.00	5.00	0.53	0.98 ± 1.08	22.68 ± 18.51
5.00	5.00	**0.59**	0.67 ± 0.99	15.23 ± 13.33
10.00	5.00	0.45	0.51 ± 0.84	13.11 ± 13.49
50.00	5.00	0.40	$\mathbf{0.51 \pm 0.79}$	$\mathbf{11.36 \pm 11.60}$
100.00	5.00	0.41	0.58 ± 0.86	12.54 ± 14.20
0.00	10.00	0.32	1.42 ± 1.09	33.47 ± 18.11
0.01	10.00	0.33	1.43 ± 1.10	33.42 ± 12.90
0.05	10.00	0.32	1.44 ± 1.09	34.21 ± 11.00
0.10	10.00	0.34	1.39 ± 1.11	33.40 ± 12.70
0.50	10.00	0.46	1.22 ± 1.14	18.89 ± 13.20
1.00	10.00	0.50	1.08 ± 1.10	18.58 ± 14.95
5.00	10.00	**0.61**	0.58 ± 0.94	14.28 ± 12.63
10.00	10.00	0.58	0.55 ± 0.90	12.95 ± 10.90
50.00	10.00	0.32	0.41 ± 0.71	$\mathbf{9.96 \pm 8.40}$
100.00	10.00	0.37	$\mathbf{0.40 \pm 0.72}$	10.92 ± 8.47
Fenrir		0.75	0.20 ± 0.37	110.04 ± 61.70
ours		**1.00**	$\mathbf{0.00 \pm 0.00}$	696.22 ± 78.10

Table 3: Effect of adding noise to the gradients of the RK baseline. The noise was added with different scales / variance and exponential decays. Constant variance is indicated by $\tau = -$.