UNIVERSITY OF SURREY

A Hamiltonian approach to the adjoint technique

(with application to a Variational Gaussian Process Approximation)

Renata Retkute and Ian Roulstone

Department of Mathematics

Faculty of Engineering and Physical Sciences

University of Surrey

Variational Data Assimilation

$$J(x)=\frac{1}{2}(x_0-x_b)^T B^{-1}(x_0-x_b)+\frac{1}{2}\sum_{i=1}^{N}(y_i-H(x_i))^T R_n^{-1}(y_i-H(x_i))$$

Find

$$\min_{x_0} J(x)$$

Subject to the strong constraint that the model states are a solution to the numerical model and that the tangent linear hypothesis holds.

Adjoint variable λ:

$$\frac{\partial J}{\partial x_0}\Leftarrow -\lambda_0$$

Hamiltonian formulation of 4D-Var

$$L = \int_0^T \left[J + \lambda(t)\left(\dot{x} - f(x)\right)\right] dt = \int_0^T \tilde{L}\, dt$$

$$p \equiv \frac{\partial L}{\partial \dot{x}}$$

$$H(x, \lambda) = p^T \dot{x} - \tilde{L} = -J + f(x)^T \lambda$$

$$\dot{x} = \frac{\partial H}{\partial \lambda} = f(x), \qquad \text{(model)}$$

$$\dot{\lambda} = -\frac{\partial H}{\partial x} = \nabla_x J - \left(\nabla f(x)\right)^T \lambda. \qquad \text{(adjoint)}$$

$$\frac{\partial H}{\partial x_0} = B^{-1}(x_0 - x_b) - \lambda_0 = 0,$$

$$\frac{\partial H}{\partial x(T)} = \lambda(T) = 0.$$

Non-autonomous system

Consider extended system*:*

$$\left(x_1,\ldots,x_d,\lambda_1,\ldots,\lambda_d;x_{d+1},\lambda_{d+1}\right)$$

Define a new Hamiltonian*:*

$$\hat{H}=H\left(x,\lambda,x_{d+1}\right)+\lambda_{d+1}$$

$$\dot{x}_i=\frac{\partial\hat{H}}{\partial\lambda_i}=\frac{\partial H}{\partial\lambda_i},\qquad \dot{x}_{d+1}=\frac{\partial\hat{H}}{\partial\lambda_{d+1}}=1,$$

$$\dot{\lambda}_i=-\frac{\partial\hat{H}}{\partial x_i}=-\frac{\partial H}{\partial x_i},\qquad \dot{\lambda}_{d+1}=-\frac{\partial\hat{H}}{\partial x_{d+1}}=-\frac{\partial H}{\partial t}$$

$$i=1:d$$

Numerical scheme

Euler-B scheme for extended Hamiltonian:

$$x_i^{k+1} = x_i^k + h\frac{\partial \hat{H}}{\partial \lambda_i}\left(x_i^k, \lambda_i^{k+1}, x_{d+1}^k, \lambda_{d+1}^{k+1}\right)$$

$$\lambda_i^{k+1} = \lambda_i^k - h\frac{\partial \hat{H}}{\partial x_i}\left(x_i^k, \lambda_i^{k+1}, x_{d+1}^k, \lambda_{d+1}^{k+1}\right)$$

$$x_{d+1}^{k+1} = x_{d+1}^k + h\frac{\partial \hat{H}}{\partial p_{d+1}}\left(x_i^k, \lambda_i^{k+1}, x_{d+1}^k, \lambda_{d+1}^{k+1}\right)$$

$$\lambda_{d+1}^{k+1} = \lambda_{d+1}^k - h\frac{\partial \hat{H}}{\partial \lambda_{d+1}}\left(x_i^k, \lambda_i^{k+1}, x_{d+1}^k, \lambda_{d+1}^{k+1}\right)$$

Euler-B is a symplectic integrator

Symplectic maps

$$z = \left(x, \lambda, x_{d+1}, \lambda_{d+1}\right)$$
$$\dot{z} = J\nabla_z H\left(z\right)$$
$$J^T = -J$$

$$z\left(t^0\right) = z_0$$
$$\phi_t\left(z\right) = z\left(t; z^0, t^0\right)$$

$$\left(D_z\phi_t\left(z\right)\right)^T J^{-1}\left(D_z\phi_t\left(z\right)\right) = J^{-1}$$

Theorem. If ϕ is a symplectic map of the extended system, then corresponding non-autonomous system is also a symplectic map.

Numerical schemes

Euler-B method:

$$x^{k+1} = x^k + h\frac{\partial H}{\partial \lambda}\left(x^k, \lambda^{k+1}\right) = x^k + hf\left(x^k\right)$$

$$\lambda^{k+1} = \lambda^k - h\frac{\partial H}{\partial x}\left(x^k, \lambda^{k+1}\right) = \lambda^k + h\left(\nabla J_k - \nabla f\left(x^k\right)^T \lambda^{k+1}\right)$$

Euler method:

$$x^{k+1} = x^k + hf\left(x^k\right)$$

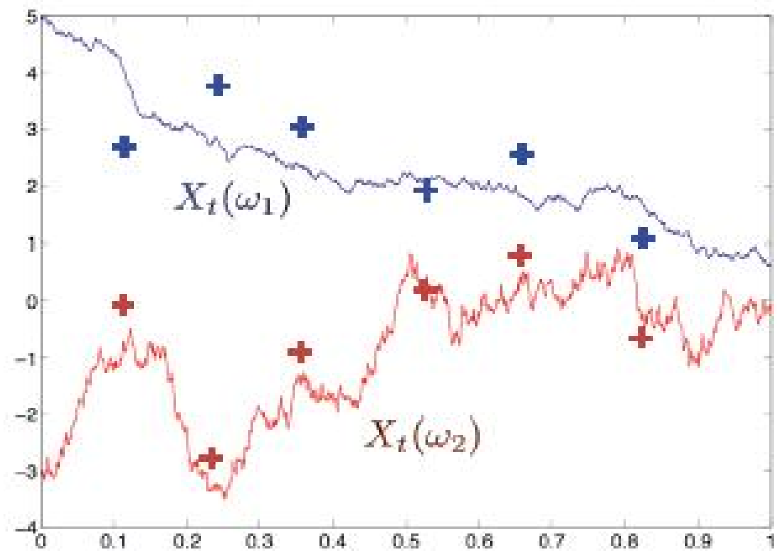
$$\lambda^{k+1} = \lambda^k + h\left(\nabla J_k - \nabla f\left(x^k\right)^T \lambda^k\right)$$

Variational Gaussian Process Approximation

Project **'Variational Inference in Stochastic Dynamic (Environmental) models** (VISDEM)':

Aston University
TU Berlin
UCL
University of Surrey

Variational Gaussian Process Approximation



Stochastic dynamics:

$$dx_t = f(x_t, t)dt + \sigma\, dW_t$$

Noisy observations:

$$y_n = x_{t=t_n} + \varepsilon_n, \quad \text{where}\ \varepsilon_n \sim N(0, R)$$

Variational Gaussian Process Approximation

Non-Gaussian processes are approximated by a Gaussian processes *Q* :

$$dx_t = f_L(x_t, t)dt + \sigma \, dW_t$$
$$f_L(x_t, t) = -\alpha x_t + \beta$$

Kullback-Leibler divergence:

$$KL[Q \| P] = \left\langle \ln \frac{Q}{P} \right\rangle_Q$$

Approximate process:

UNIVERSITY OF SURREY

Linear SDE:

$$dx = \left(-\alpha x + \beta\right)dt + \sigma dW$$

PDF:

$$Q(x) \sim N\left(x \middle| m, S\right) = \frac{1}{\sqrt{2\pi S}} e^{-(x-m)^T S^{-1}(x-m)}$$

Mean and covariance:

$$\dot{m}(t) = -\alpha(t)m(t) + \beta(t)$$

$$\dot{S}(t) = -\alpha(t)S(t) - S(t)\alpha(t)^T + \sigma^2$$

Kullback -Leibler divergence

UNIVERSITY OF SURREY

KL divergence (Archambeau et al 2008) :

$$KL\left[Q\|P\right]=\int_0^T\left(E_{sde}(t)+E_{obs}(t)\right)dt$$

$$E_{sde}(t)=\frac{1}{2}\left\langle\left(f-f_L\right)^T\sigma^{-2}\left(f-f_L\right)\right\rangle_Q$$

$$E_{obs}(t)=\frac{1}{2}\sum_{n=1}^{N}\left\langle\left(y_n-x_n\right)^T R_n^{-1}\left(y_n-x_n\right)\right\rangle_Q$$

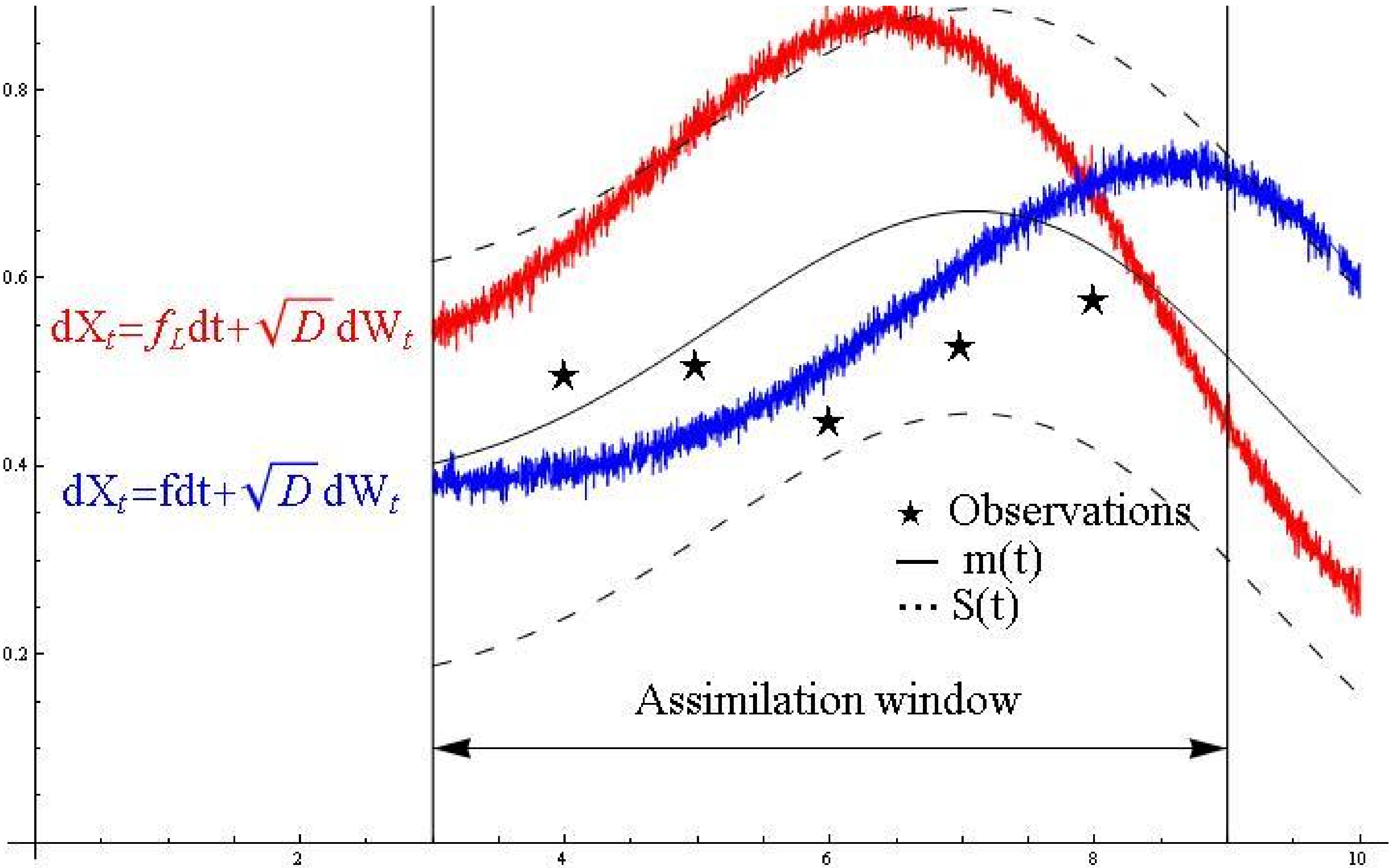


Figure. Schematic view of VGPA framework.

Hamiltonian formulation of VGPA

$$H = -E_{sde} - E_{obs} + \lambda\left(-\alpha m + \beta\right) + \Psi\left(-\alpha S - S^T\alpha + \sigma\right)$$

$$\dot{m}(t) = \frac{\partial H}{\partial \lambda} = -\alpha(t)m(t) + \beta(t)$$

$$\dot{S}(t) = \frac{\partial H}{\partial \Psi} = -\alpha(t)S(t) + S(t)\alpha(t)^T + \sigma^2$$

$$\dot{\lambda}(t) = -\frac{\partial H}{\partial m} = \nabla_m E_{sde}(t) + \nabla_m E_{obs}(t) + \lambda(t)\alpha(t)$$

$$\dot{\Psi}(t) = -\frac{\partial H}{\partial S} = \nabla_S E_{sde}(t) + \nabla_S E_{obs}(t) + 2\Psi(t)\alpha(t)$$

$$\nabla_\alpha H = 0$$

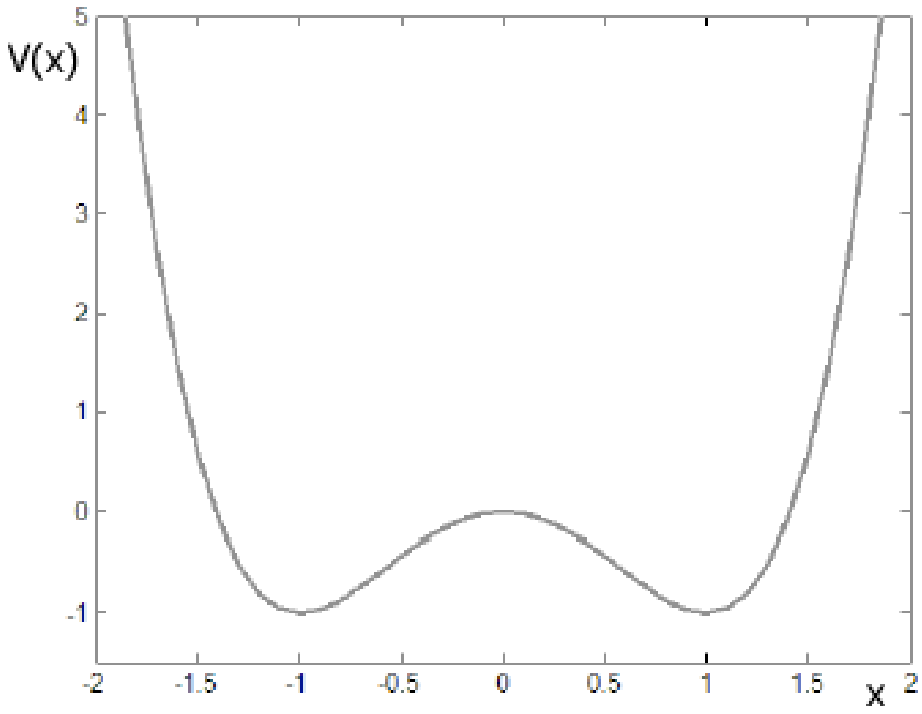
$$\nabla_\beta H = 0$$

Algorithm

1. Initialize α_0, β_0, m_0 and S_0
2. Run model forward for m and S.
3. Calculate E_{sde} and E_{obs}.
4. Run adjoint system backwards for λ and Ψ.
5. Using conjugate gradient, update α, β.
6. Calculate $KL[Q||P]$.
7. Repeat steps 2-6 until required accuracy is reached.

Double-well model with noise

$$dx = \left(4x - 4x^3\right)dt + \sigma dW$$



Double-well model with noise

$$H = -E_{sde} - E_{obs} - \lambda\left(-\alpha m + \beta\right) - \Psi\left(-2\alpha S + \sigma^2\right)$$

$$\dot{m} = \frac{\partial H}{\partial \alpha},\ \dot{S} = \frac{\partial H}{\partial \Psi} \quad \text{(model)}$$

$$\dot{\alpha} = -\frac{\partial H}{\partial m},\ \dot{\Psi} = -\frac{\partial H}{\partial S} \quad \text{(adjoint)}$$

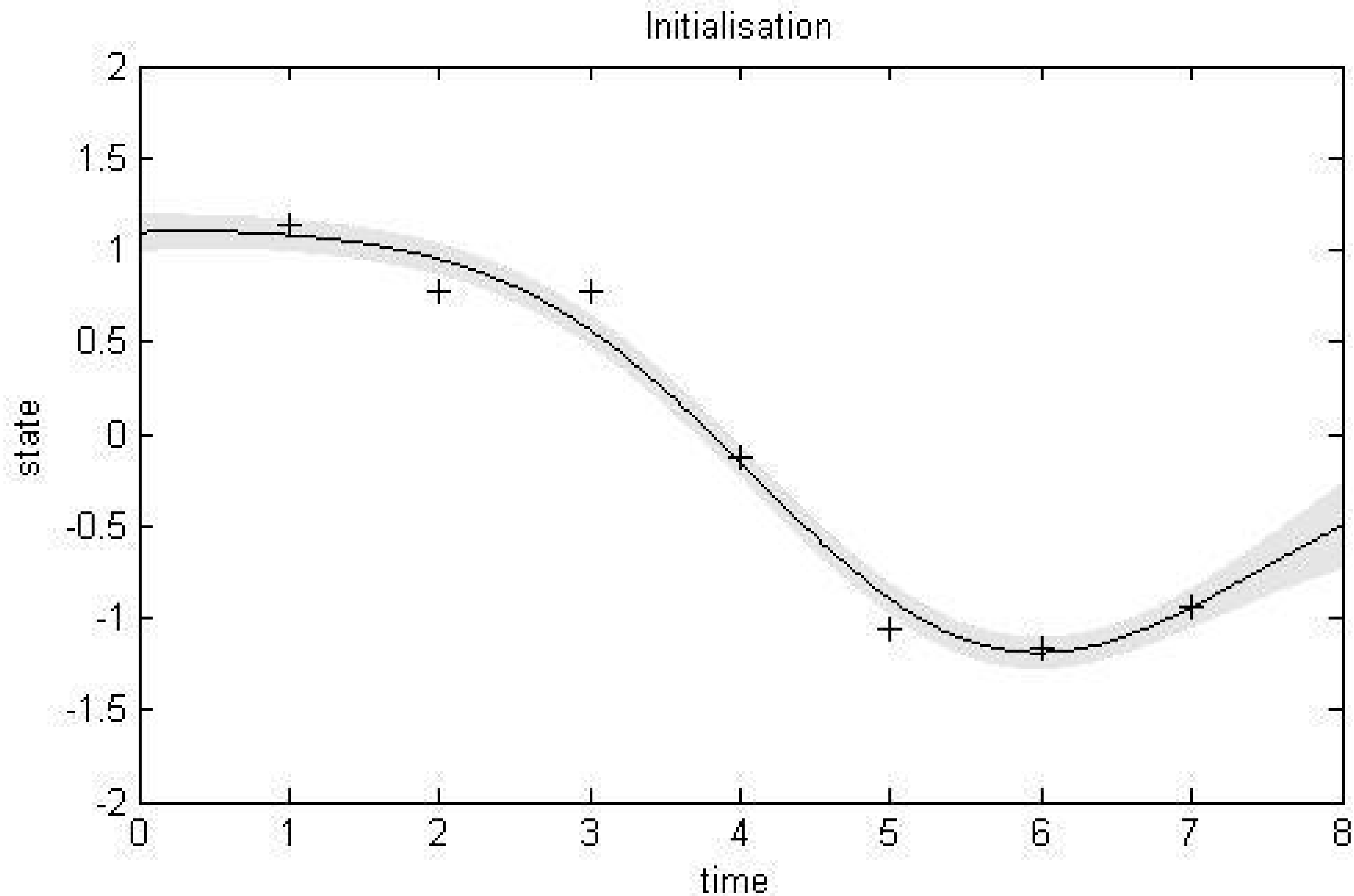
$$E_{sde} = \left(2\sigma^2\right)^{-1}(16(m^6 + 15m^4 S + 45m^2 S^2 + 15S^3)$$

$$-8(\alpha+4)(m^4 + 6m^2 S + 3S^2) + (\alpha+4)^2(m^2 + S)$$

$$+8\beta(m^3 + 3mS) - 2\beta(\alpha+4)m + \beta^2)$$

$$E_{obs} = (2R)^{-1}\left(y_n^2 - y_n m + S + m^2\right)$$

UNIVERSITY OF SURREY



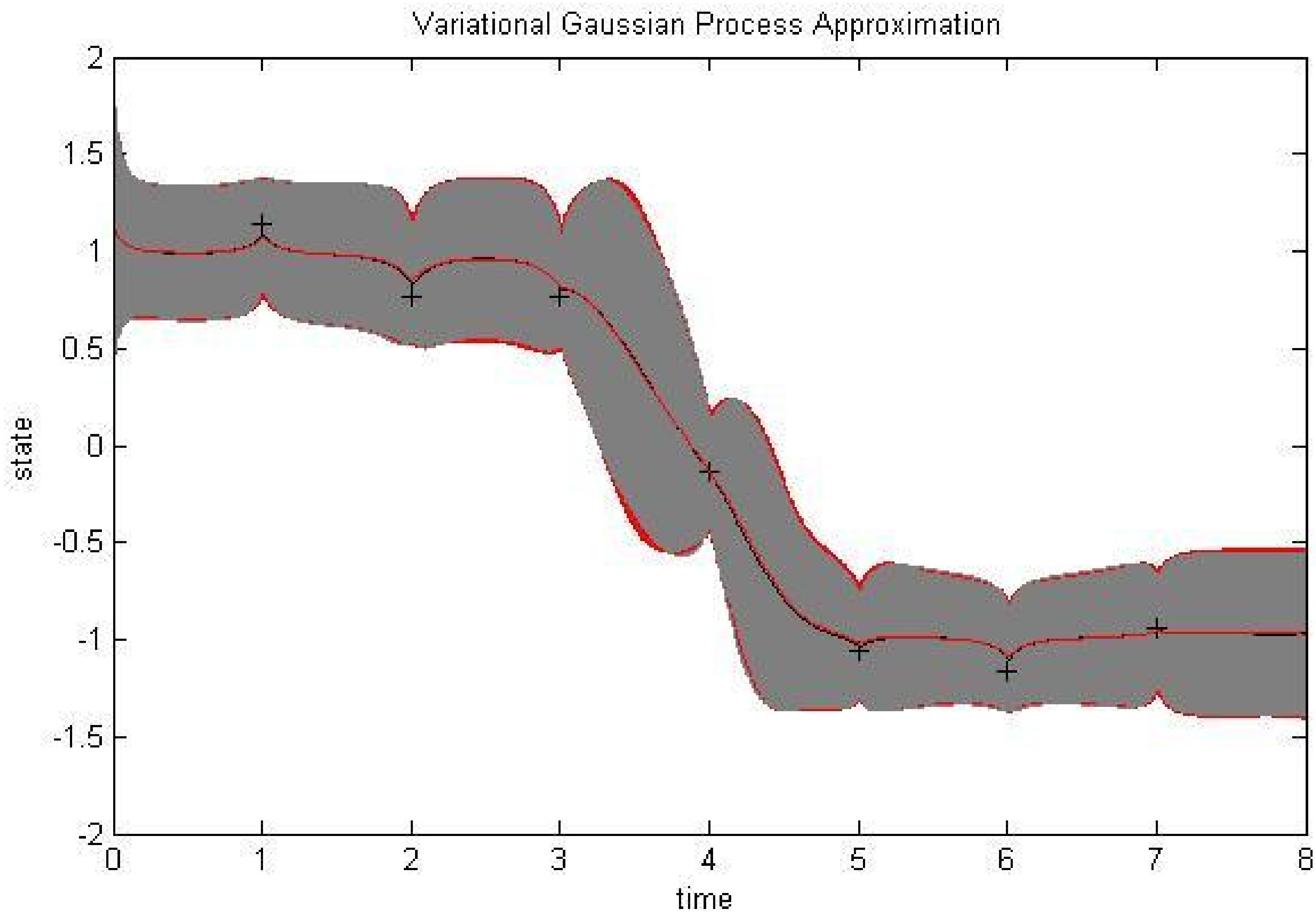


Figure. R = 0.01, σ=0.5.

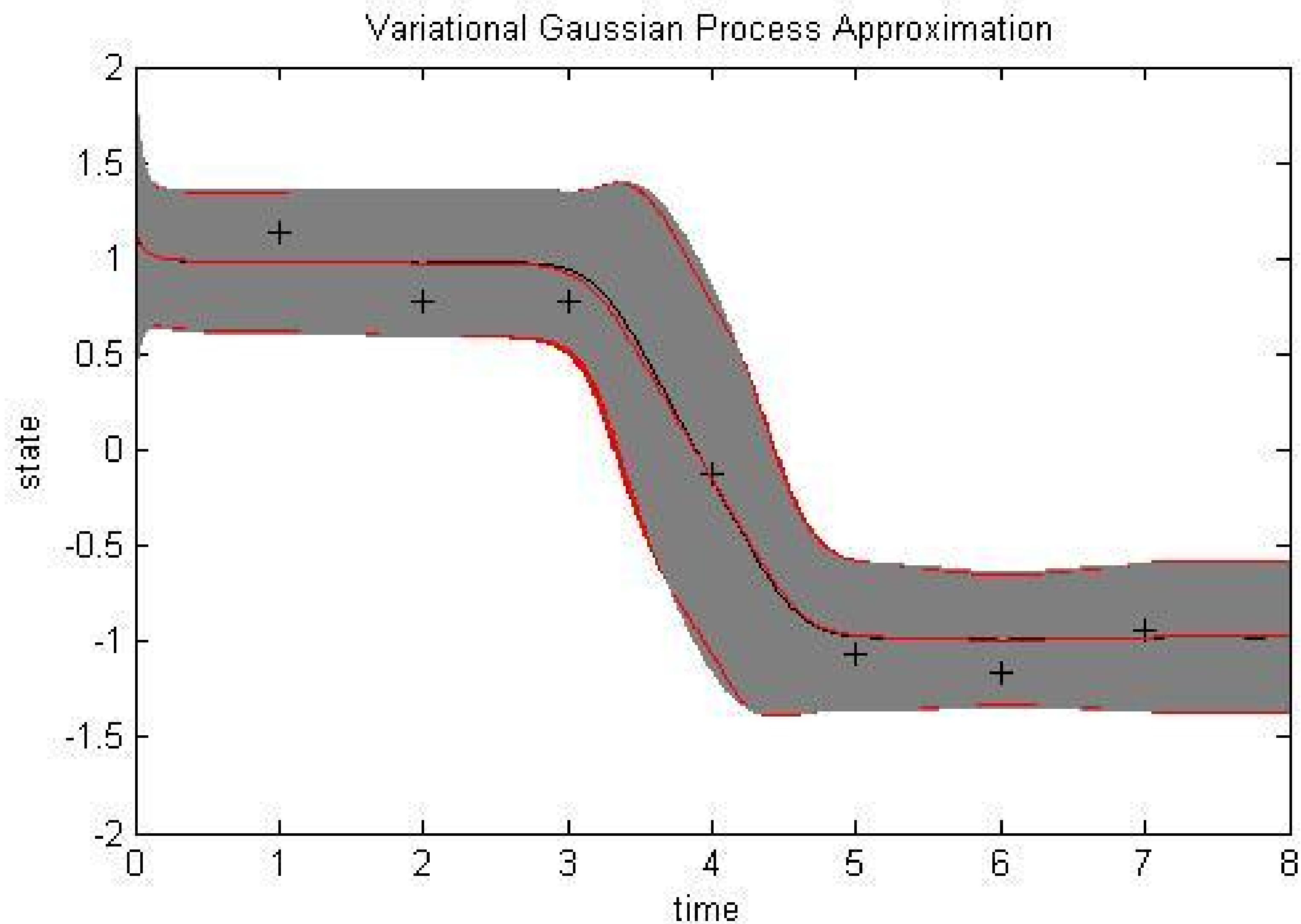


Figure. R = 1, σ=0.5.

Double-well model

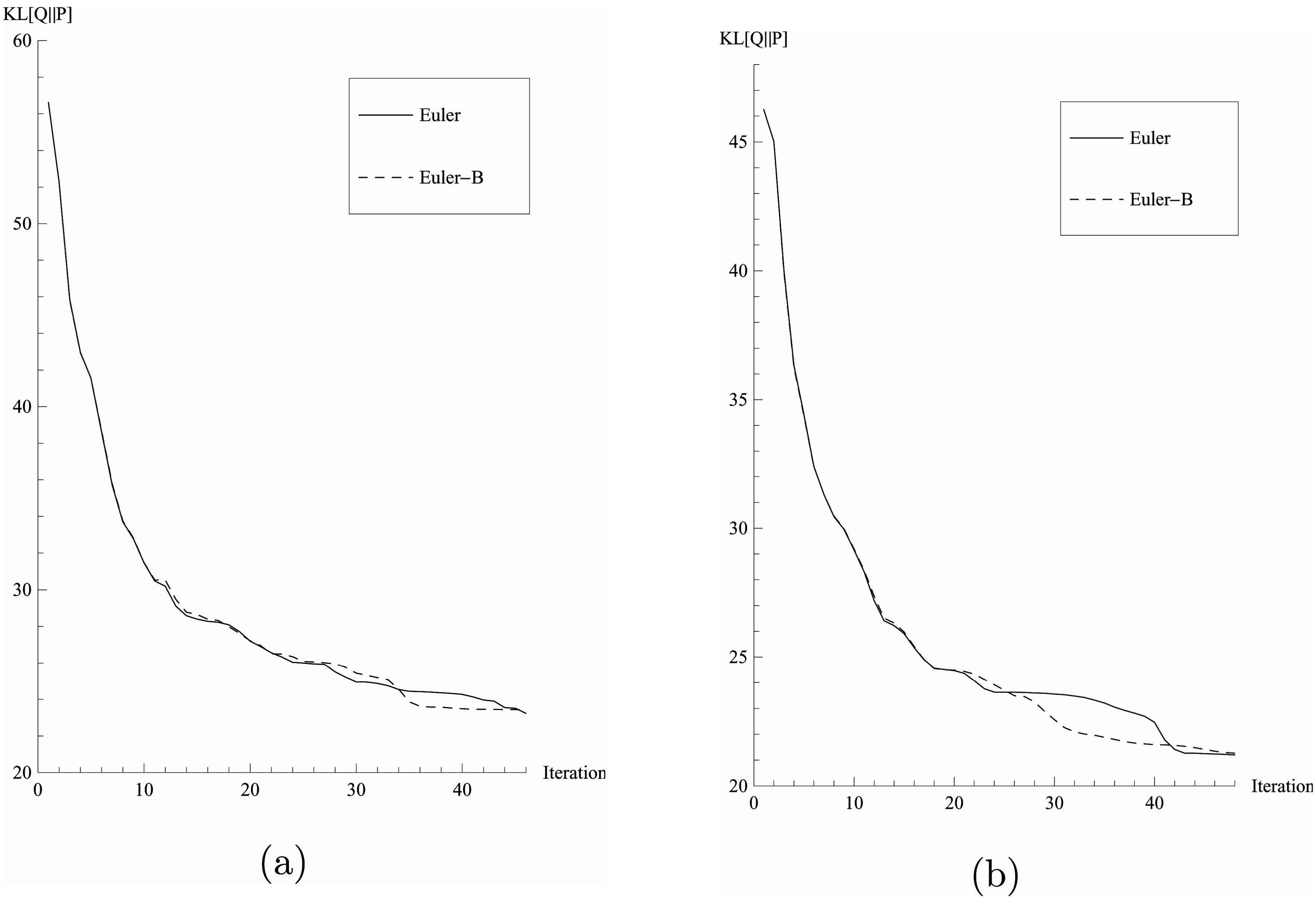
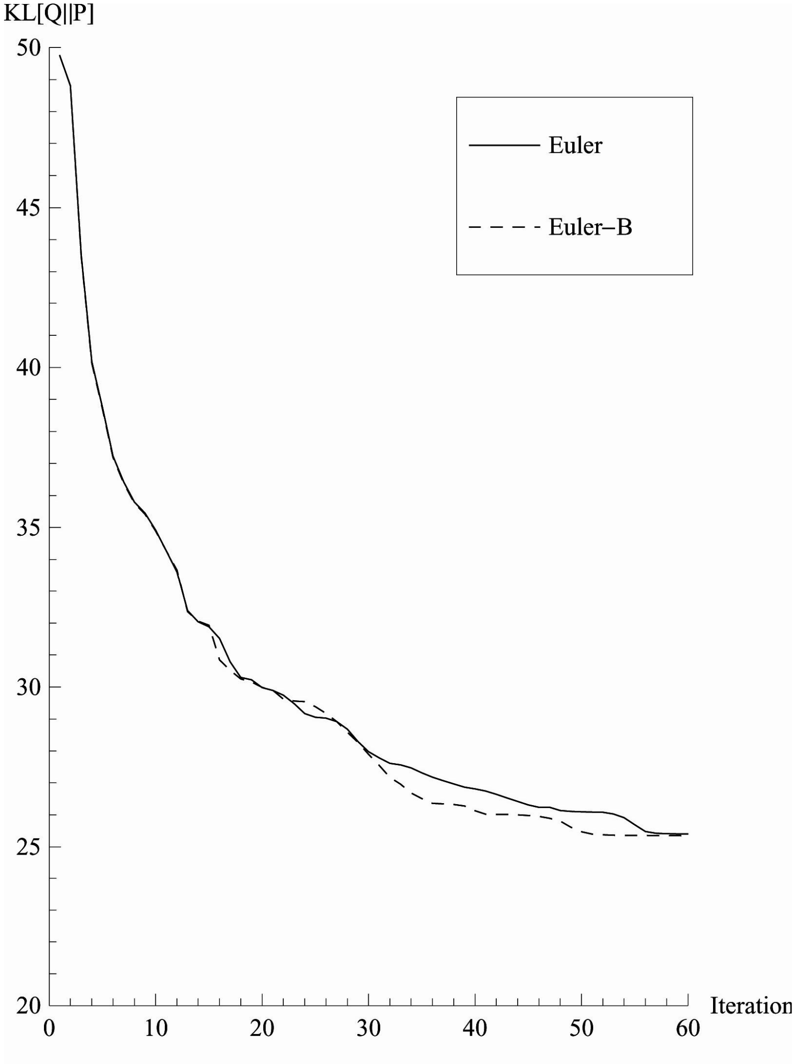
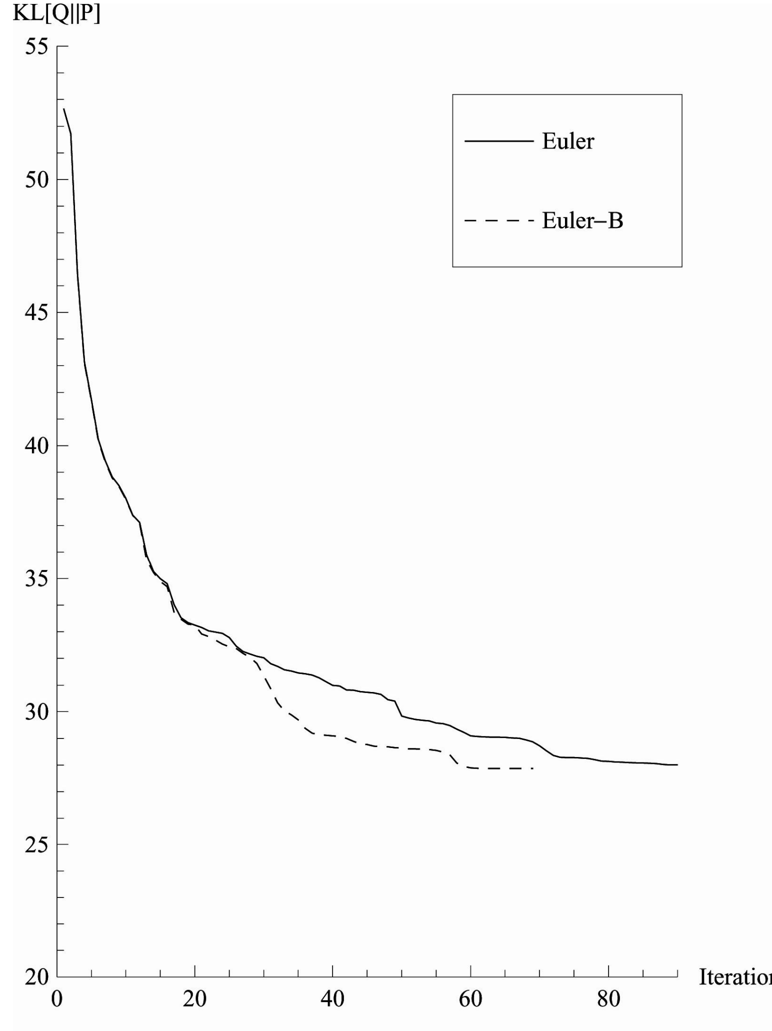


Figure. (a) R = 0.01, (b) R = 0.04.

Double-well model



(c)



(d)

Figure. (c) R = 0.4, (d) R = 1.0.

Conclusions

UNIVERSITY OF SURREY

- We have shown how symplectic methods may be applied to 4DVAR

- We have studied the application of these methods to Variational Gaussian Process Approximation to stochastic dynamical systems

- We have shown that symplectic Euler-B performed better than non-symplectic Euler scheme in tracking the true state of the system in the presence of the measurement noise for stochastically driven double well potential model.