ICSLP 98

Global Optimisation of Neural Network Models Via Sequential Sampling-Importance Resampling

Sue Johnson

Nando de Freitas

Mahesan Niranjan

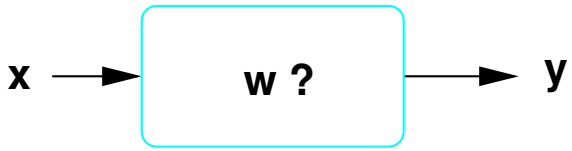
Andrew Gee

http://www-svr.eng.cam.ac.uk/~jfgf

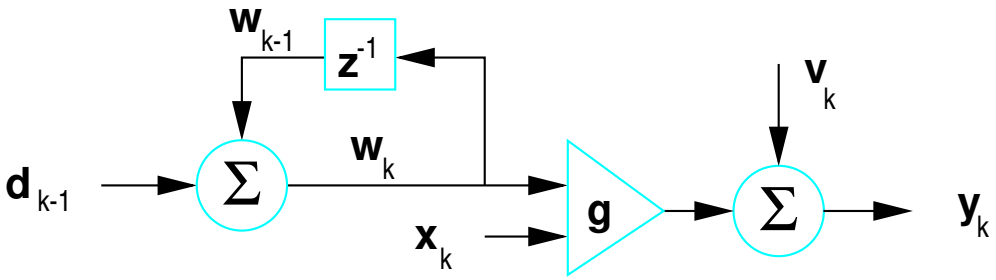
Email: jfgf@eng.cam.ac.uk

Cambridge University Engineering Department

Dynamical Model

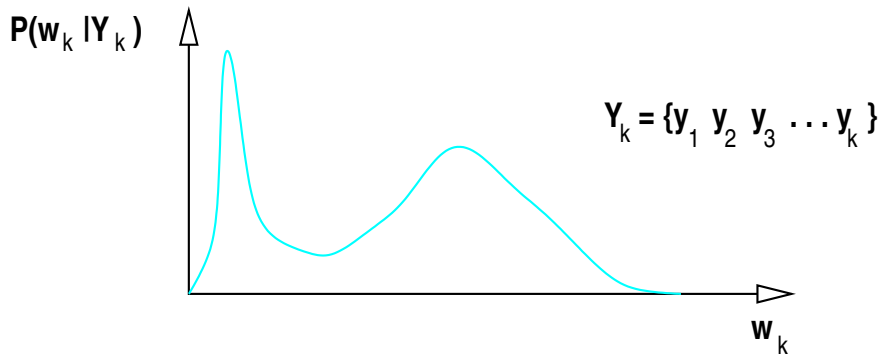


$$w_k = w_{k-1} + d_{k-1}$$
$$y_k = g(w_k, x_k) + v_k$$



{x,y}	Input and output measurements.
w	Unknown neural network weights.
g(.)	Neural network mapping.
v	Measurement noise.
d	Process noise.

Bayesian Solution



$P(w_{k-1}|Y_{k-1})$

d_{k-1} → $p(w_k|Y_{k-1}) = \int p(w_k|w_{k-1})\ P(w_{k-1}|Y_{k-1})\ dw_{k-1}$ **Prediction**

v_k, y_k → $p(y_k|Y_{k-1}) = \int p(y_k|w_k)\ p(w_k|Y_{k-1})\ dw_k$ **Measurement**

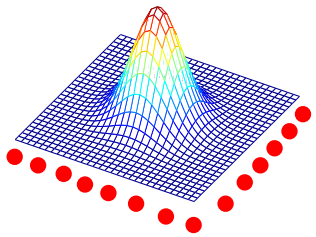
$P(w_k|Y_k) = \dfrac{p(y_k|w_k)\ p(w_k|Y_{k-1})}{p(y_k|Y_{k-1})}$ **Bayes rule**

$P(w_k|Y_k)$

Must specify: $\begin{cases} p(d_{k-1}) = p(w_k|w_{k-1}) \\ p(v_k) = p(y_k|w_k) \end{cases}$

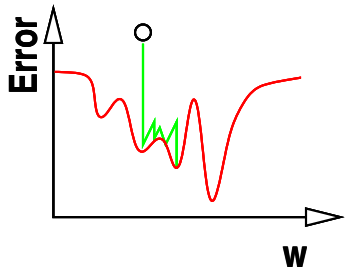
Practical Solutions

Numerical Integration



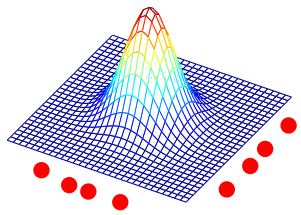
- Too cumbersome.

Extended Kalman Filtering

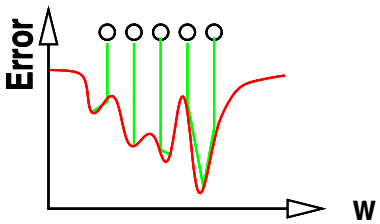


+ Computationally efficient.
- Local optimisation scheme.
  Can get trapped in local minima.
- It assumes Gaussian distributions.
- It requires a Taylor expansion
  of the measurements equation.

Sequential Monte Carlo



+ Global optimisation scheme.
+ No need for approximations.
+ No need for Gaussian assumption.
- Computationally demanding.
+ Parallel implementation.



Sequential Monte Carlo

Use the approximation:

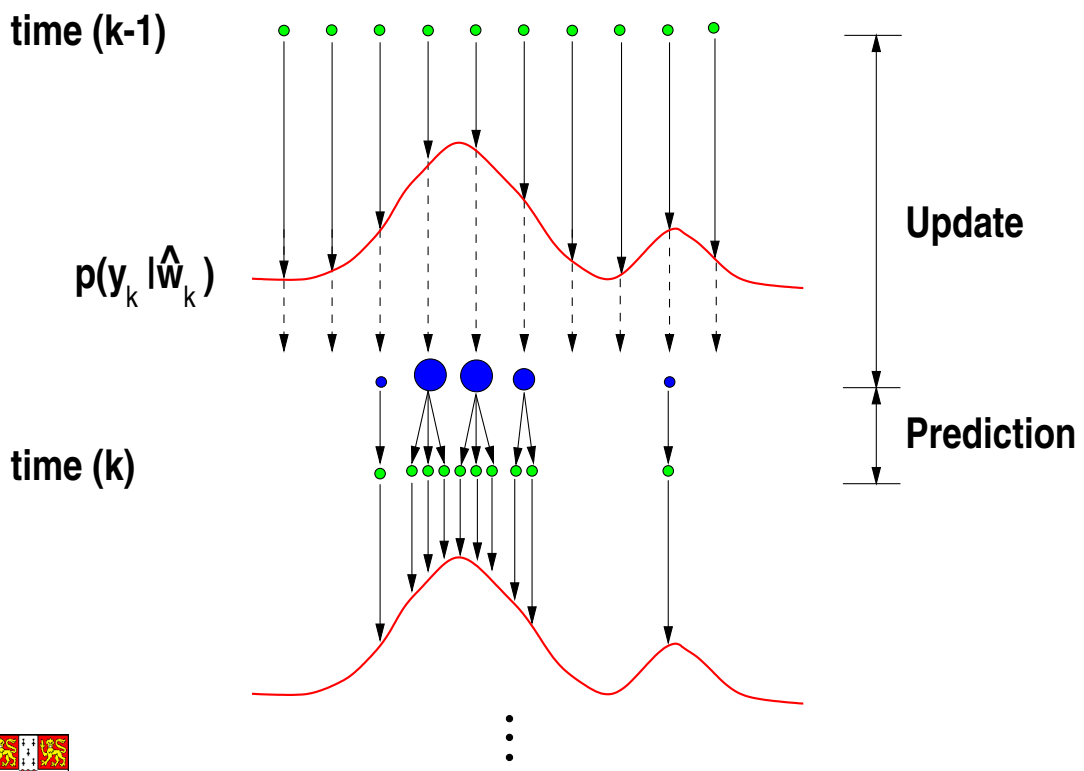
$$p(w_k|Y_k) \sim \frac{1}{S}\sum_{i=1}^{S}\delta(w_k - w_k^{(i)})$$

to map integrals:

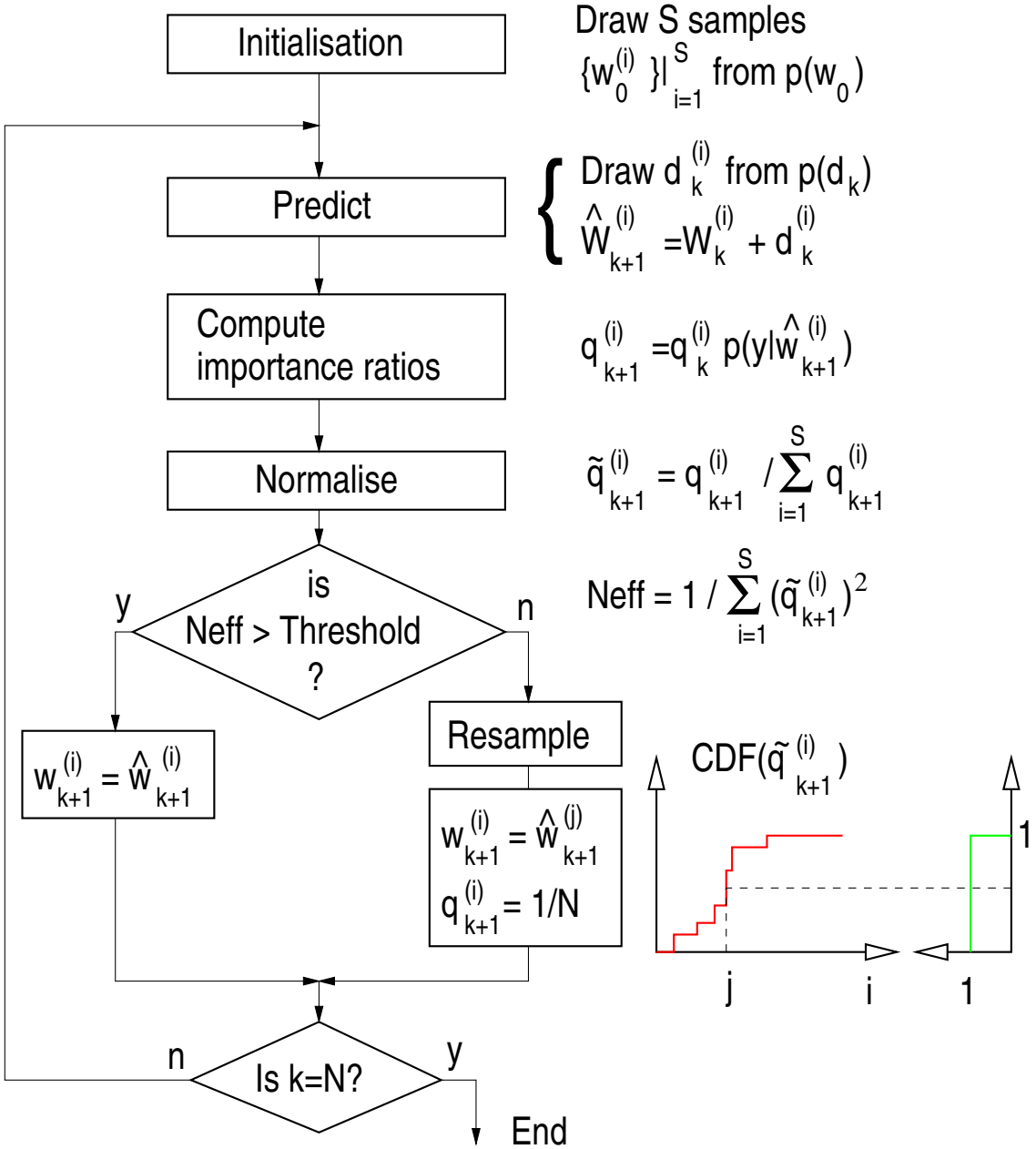
$$E\{f_k(w_k)\} = \int f_k(w_k)p(w_k|Y_k)\, dw_k$$

to weighted sums:

$$E\{f_k(w_k)\} \sim \frac{1}{S}\sum_{i=1}^{S} f_k(w_k^{(i)})$$



SIR Algorithm

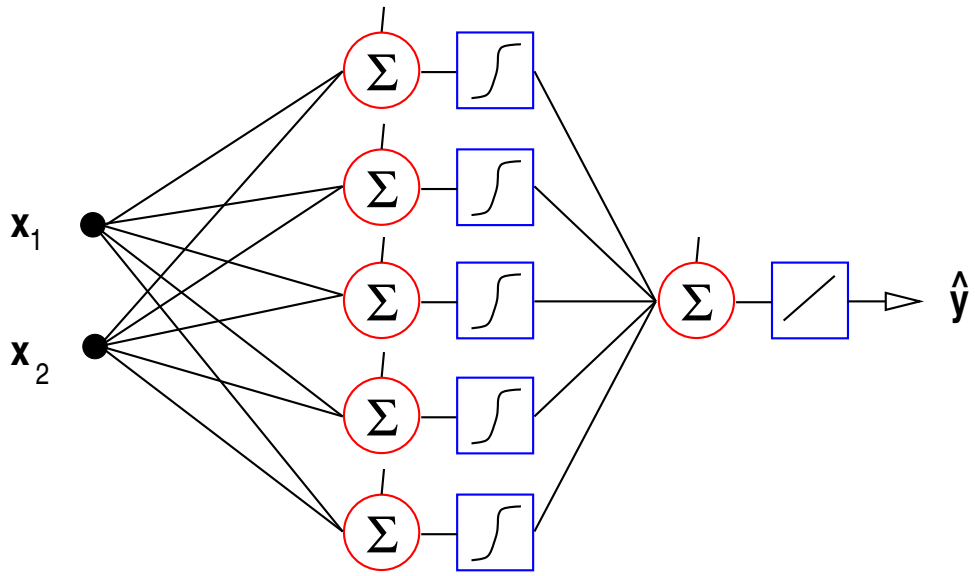


Noisy Function Approximation

We approximated the following noisy function:

$$y(x_1, x_2) = 4 \sin(x_1 - 2) + 2 x_2^2 + 5 + \text{noise}$$

With a two-layer neural network

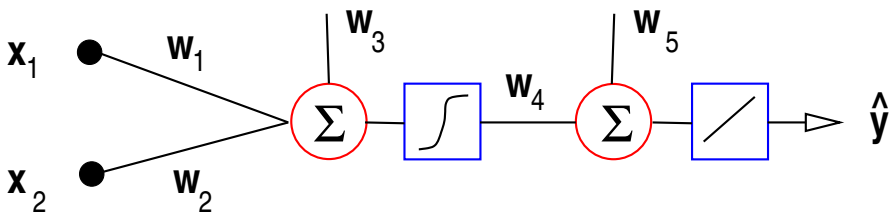


	RMS Error	Time (sec)
EKF	6.04	4
SIPR	4.81	98
SIR	2.83	109

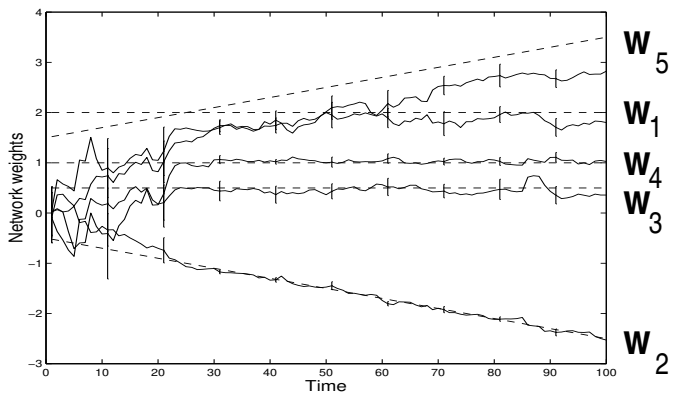
SIR: Resampling at each time step.
SIPR: 50% resampling.

Latent States Tracking

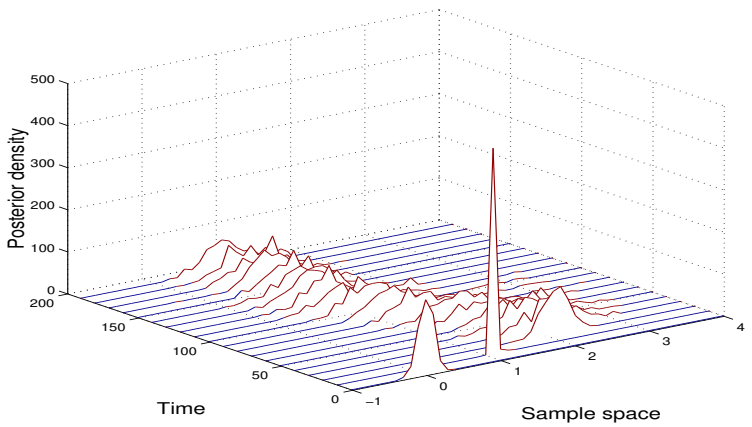
Logistic map (two-neuron MLP)



Tracking weights



Typical pdf for a network weight



Conclusions

+ Global optimisation scheme for NNs.
+ No need for approximations.
+ No need for Gaussian assumption.
- Computationally demanding.
+ Parallel implementation.

For improved algorithms and software visit:

http://svr-www.eng.cam.ac.uk/~jfgf/software.html

http://svr-www.eng.cam.ac.uk/~jfgf/publications.html