Raytheon

Concentration of measure in nonlinear filters

Fred Daum

22 February 2007

Raytheon

Nonlinear Filtering Problem

$$dx = f(x, t)dt + G(x,t)\,dw$$

x = d-dimensional state vector

t = time

w(t) = process noise vector

$$z(t_k) = h(x(t_k), t_k, v_k)$$

$z(t_k)$ = m-dimensional measurement vector

t_k = time of k^{th} measurement

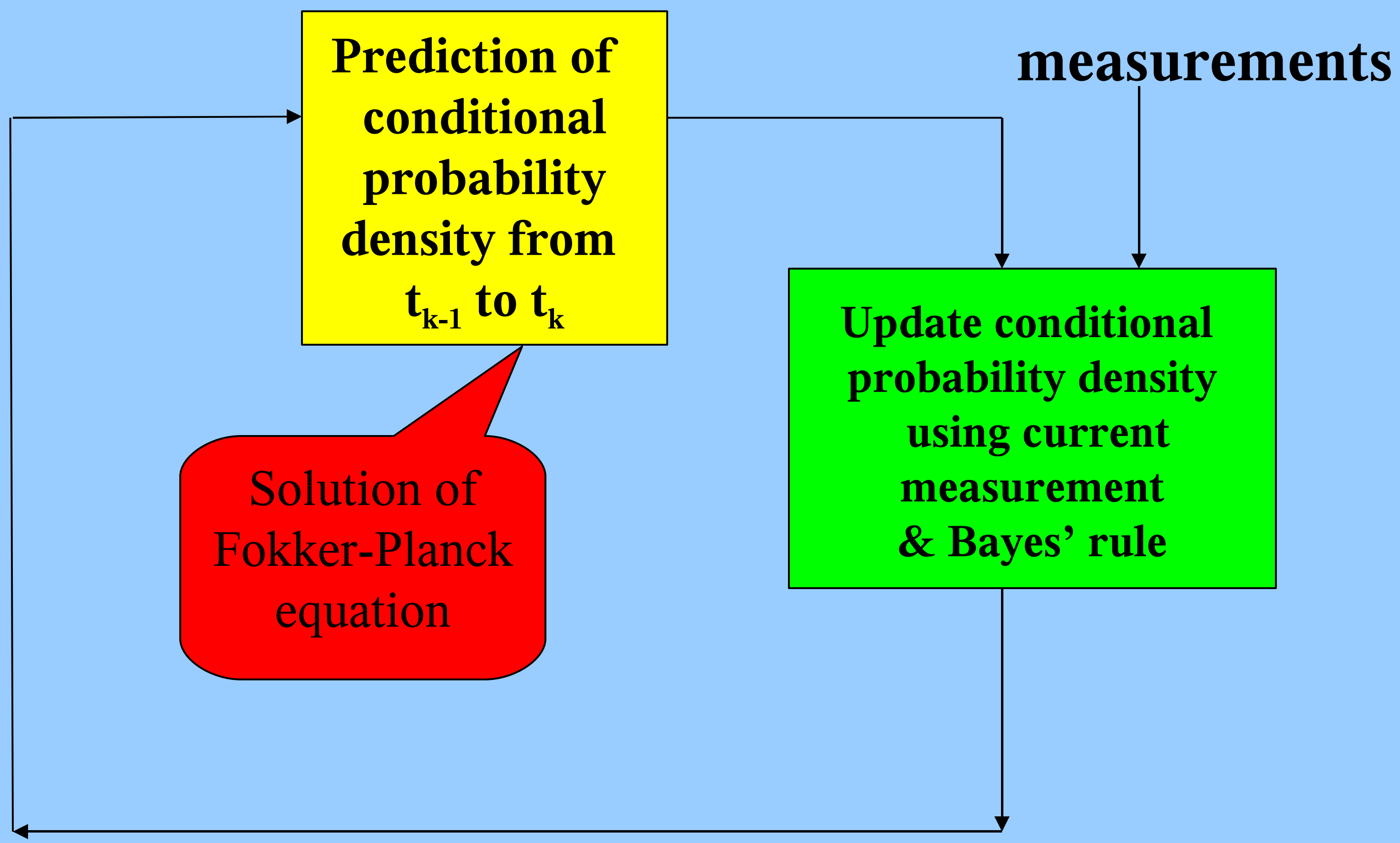
v_k = measurement noise vector

p(x, t | Zk) = probability density of x at time t given Zk

Zk = set of all measurements up to & including time tk

Raytheon

Nonlinear Filter



Raytheon

Fokker-Planck equation

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} f - Tr\left(\frac{\partial f}{\partial x}\right) p + \frac{1}{2} Tr\left(\frac{\partial^2 p}{\partial x^2} Q\right)$$

$$p = p(x, t)$$

p = probability density of x at time t

Bayes' rule

$$p\left(x,t_k/Z_k\right)=p\left(x,t_k/Z_{k-1}\right)p\left(z_k/x,t_k\right)$$

Unnormalized probability density of x at time t given the set of measurements up to & including time t

Gauss

Raytheon

Computational Complexity

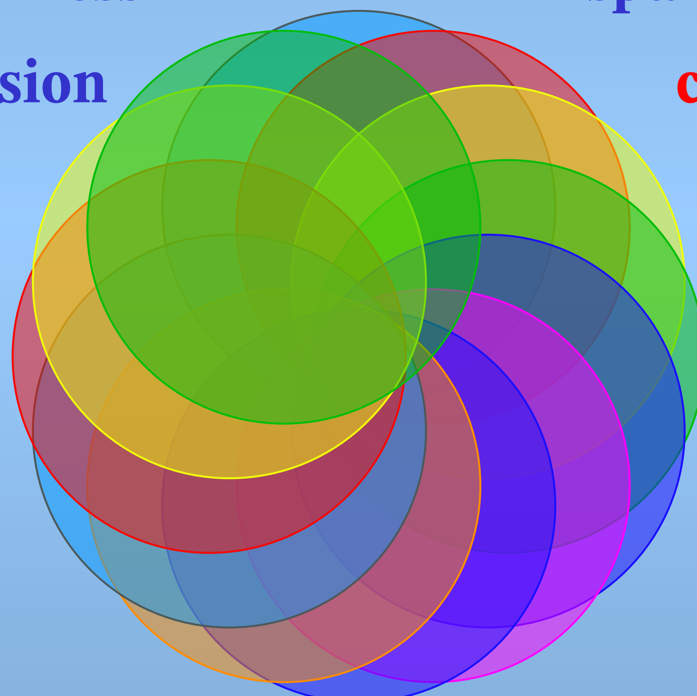
dimension

smoothness

sparseness

effective dimension

concentration of measure



convexity & unimodality, etc.

good guess

exploit structure

robustness & stability

exact solutions

accuracy

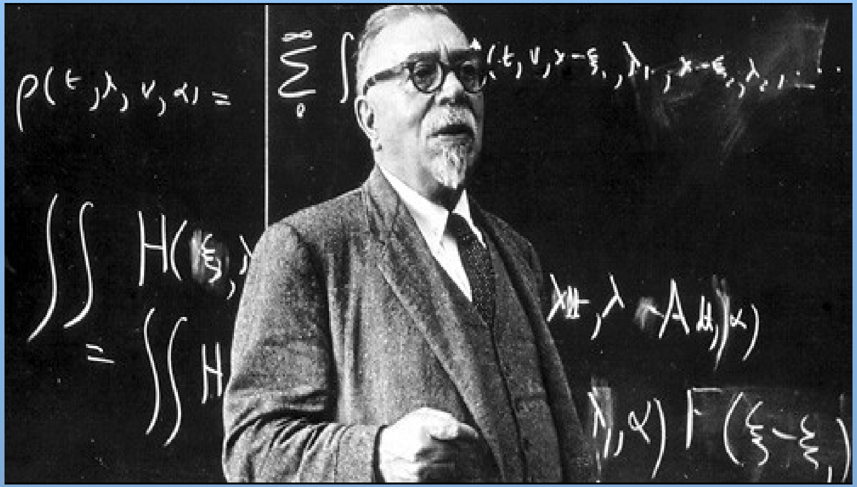
ill-conditioning

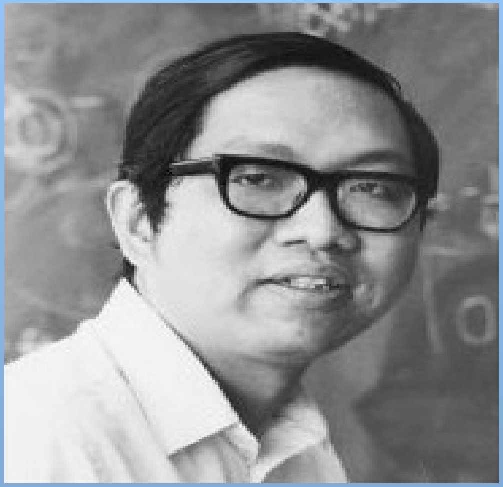
Raytheon

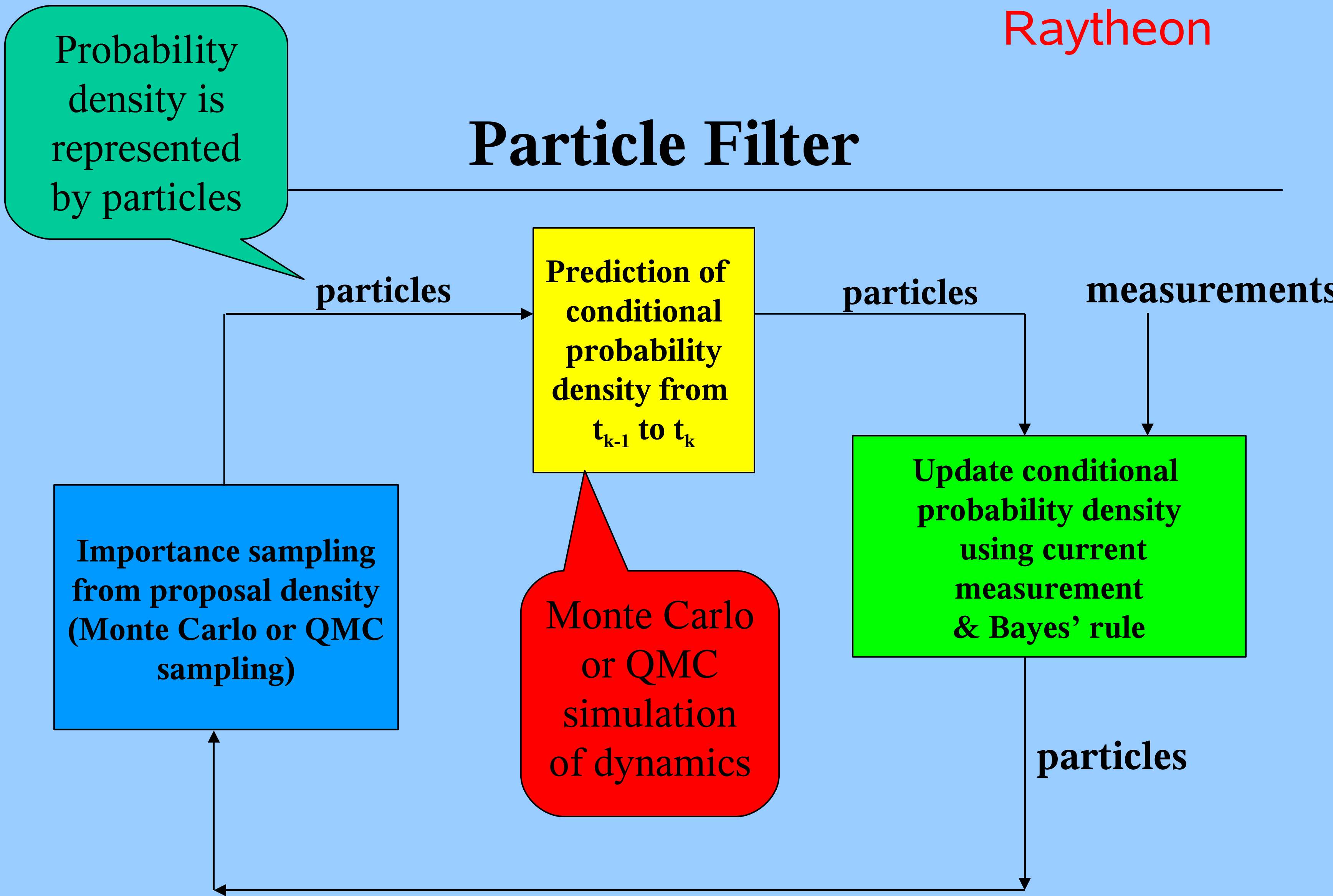
Type of nonlinear filter	Statistics computed	Computational Complexity	Estimation accuracy	Representation of probability density
Extended Kalman filters	Mean vector & covariance matrix	d^3	Sometimes good but often highly suboptimal	Mean vector & covariance matrix
Unscented Kalman filters	Mean vector & covariance matrix	d^3	Sometimes better than EKF	Mean vector & covariance matrix
Batch least squares	Mean vector & covariance matrix	d^3	Sometimes better than EKF	Mean vector & covariance matrix
Numerical solution of Fokker-Planck PDE	Full conditional probability density of state	Curse of dimensionality*	Optimal*	points in state space
Particle filters	Full conditional probability density of state	Curse of dimensionality**	Optimal**	particles
Exact recursive filters (Benes, Daum, Yau, Wonham)	Full conditional probability density of state	Polynomial in d	Optimal for special problems	Sufficient statistics

Raytheon

Mathematicians & filtering






Raytheon
Particle Filter
Probability density is represented by particles
particles
Prediction of conditional probability density from t_{k-1} to t_k
particles
measurements
Importance sampling from proposal density (Monte Carlo or QMC sampling)
Monte Carlo or QMC simulation of dynamics
Update conditional probability density using current measurement & Bayes' rule
particles

Assertion About Particle Filters

"Particle filtering methods beat the curse of dimensionality as the rate of convergence is independent of the state dimension."

Dan Crisan & Arnaud Doucet

(IEEE Trans. Signal Processing March 2002)

Definition of Dimension Free Error

$$r = E\left[\left(x - \hat{x}\right)^T C^{-1}\left(x - \hat{x}\right)\right] / d$$

in which

x = state vector to be estimated

$\hat{x}$ **= estimate of x**

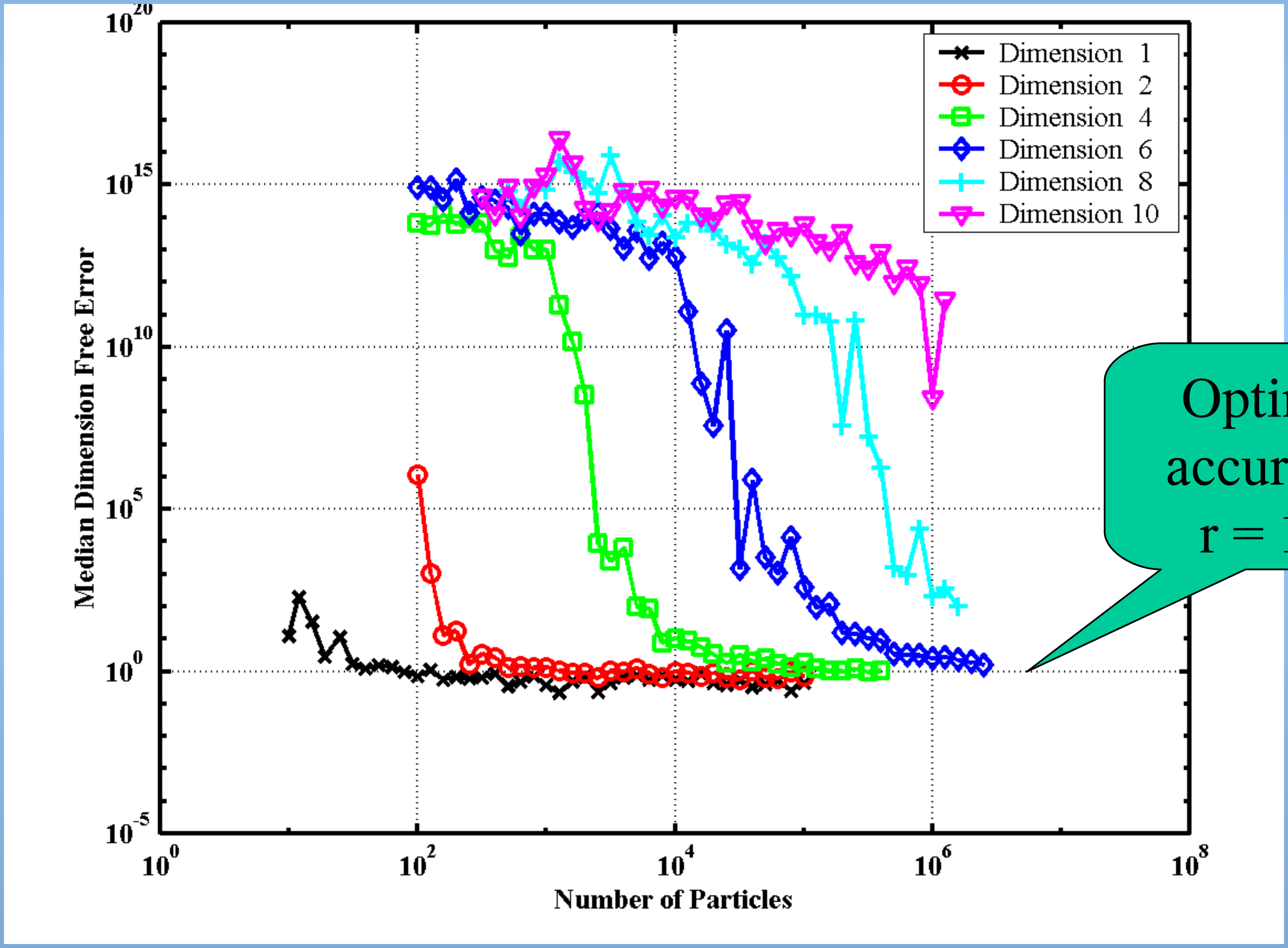
d = dimension of x

C = covariance matrix for the optimal filter

E(.) = expected value of (.)

Raytheon

Curse of dimensionality for classic particle filter



Optimal accuracy: r = 1.0

Quote from Prof. Arnaud Doucet's Website (FAQ)

QUESTION: When you say "particle filters beat the curse of dimensionality," what do you mean?

ANSWER: A rather unfortunate expression we've used ourselves to look good.

Oh's Formula for Monte Carlo errors

$$\sigma^2 \approx \left\{ \left[\frac{1+k}{\sqrt{1+2k}} \right] \exp\left[\frac{\varepsilon^2}{1+2k} \right] \right\}^d / N$$

Assumptions:

(3) Gaussian density (zero mean & unit covariance matrix)

(4) d-dimensional random variable

(5) Proposal density is also Gaussian with mean ε and covariance matrix kI, but it is not exact for k ≠ 1 or ε ≠ 0

(6) N = number of Monte Carlo trials

Raytheon

Variance too small is a disaster!

$$\sigma^2 \approx \left\{ \left(\frac{k^2}{2\mathrm{k}-1} \right)^{d/2} - 1 \right\} / N$$

Assumptions:

(3) Gaussian density (zero mean & unit covariance matrix)

(4) d-dimensional random variable

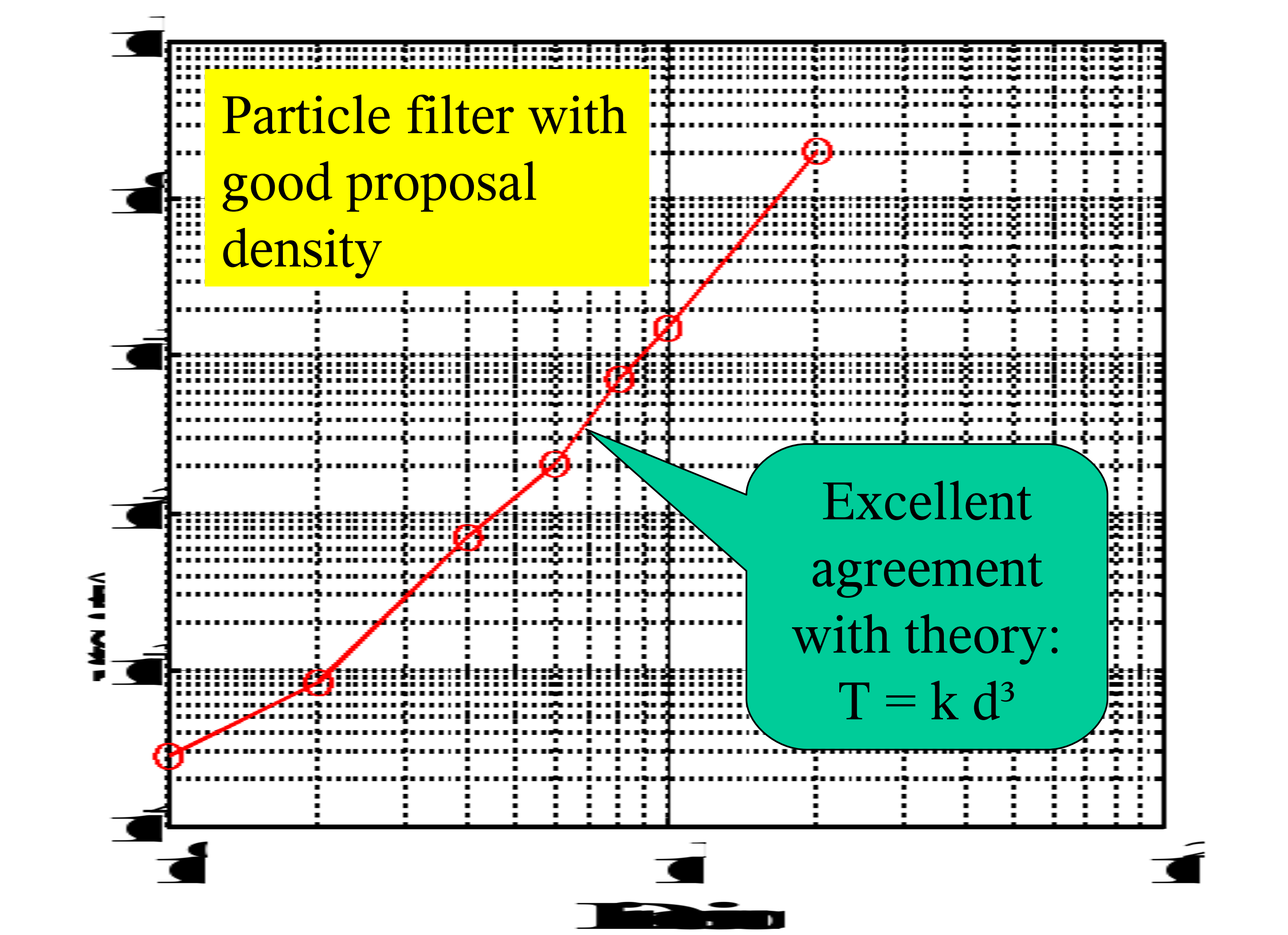
(5) Proposal density is also Gaussian (zero mean & covariance kI), but it is not exact for k ≠ 1

(6) N = number of Monte Carlo trials

(7) k > ½ (otherwise the error blows up)

Raytheon

Computation Time vs. Dimension for optimal estimation accuracy (r = 1.0)



Raytheon

Simple Back-of-the-Envelope Formula*

$$T \approx c \frac{d^3}{(\sigma_x^2 / \sigma_0^2)}$$

*assumes concentration of measure similar to Gaussian density and good proposal density; it can be derived from Oh's formula for small ε and k; it can also be derived (more generally & more simply) from the volume of a d-dimensional ball and Chernoff's bound and the usual Monte Carlo error analysis and Stirling's approximation for d!

Multiple Choice Quiz

What is the approximate ratio (r) of the volume of the unit ball to the volume of the smallest cube that contains it for d = 10 dimensions?

(a) r = 0.5
(b) r = 0.1
(c) r = 0.01
(d) r = 0.002

Another Multiple Choice Quiz

What is the approximate ratio (r) of the volume of the unit ball to the volume of the smallest cube that contains it for d = 100 dimensions?

(a) $r = 10^{-4}$

(b) $r = 10^{-6}$

(c) $r = 10^{-10}$

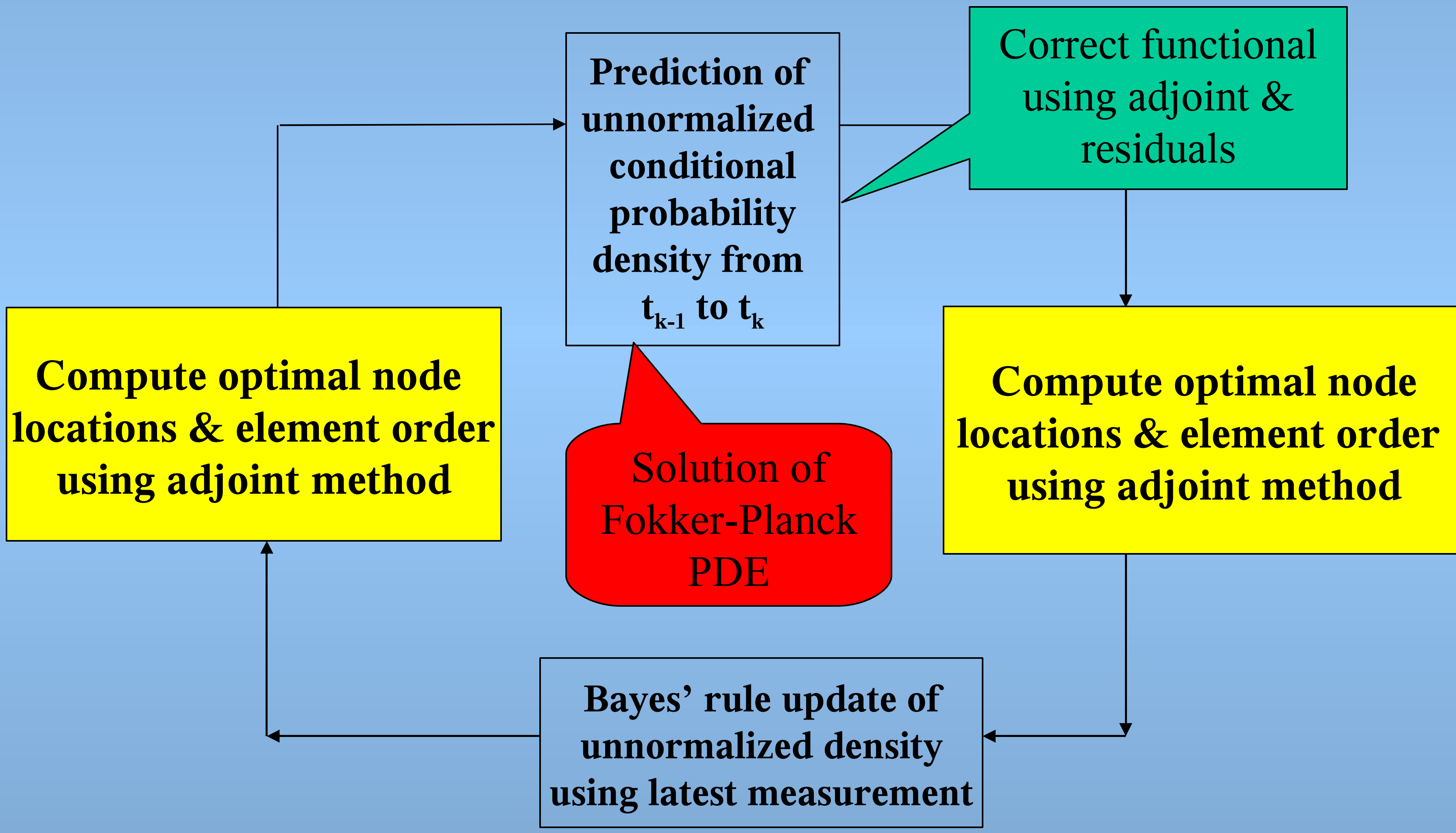
(d) $r = 10^{-69}$

Anti-particle filter vs. particle filter

	Anti-particle filter	Particle filter
1. Prediction of density (drift)	Exact solution of Fokker-Planck PDE	Monte Carlo or QMC
2. Prediction of density (diffusion)	convolution	Monte Carlo or QMC
3. Adaptive method to avoid uniform grid	Adjoint method (MC or QMC sampling)	Proposal density (MC or QMC sampling)
4. Exploits smoothness	Yes	No
5. Feedback	Yes	No
6. Models errors in the density itself	Yes	No

Raytheon

Adjoint Meshfree Nonlinear Filter



Raytheon

Adjoint Method

$$PDE : Lu = f$$

$$functional : J = \langle u , g \rangle$$

$$ADJOINT : L * v = g$$

$$J = \langle u , L * v \rangle = \langle Lu , v \rangle$$

$$\delta J = \langle Lu , v \rangle - \langle L \hat{u} , v \rangle$$

$$\delta J = \langle v - \hat{v} + \hat{v} , f - L \hat{u} \rangle$$

$$\delta J = \langle \hat{v} , f - L \hat{u} \rangle + \langle v - \hat{v} , f - L \hat{u} \rangle$$

Exact solution of Fokker-Planck for Q = 0

$$\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial x} f - Tr\left(\frac{\partial f}{\partial x}\right) p$$

$$\frac{dp}{dt} = \frac{\partial p}{\partial t} + \frac{\partial p}{\partial x} f = -Tr\left(\frac{\partial f}{\partial x}\right) p$$

$$\frac{dp}{p} = -Tr\left(\frac{\partial f}{\partial x}\right) dt$$

$$p(x, t_k) = p(x, t_{k-1}) \exp\left(-\int_{t_{k-1}}^{t_k} Tr\left(\frac{\partial f}{\partial x}\right) dt\right)$$

Variations on this theme

Density representation	nonlinear filter	analogous meshfree PDE method
points	Particles or non-particles	h-version
Exponential family	Smooth functions	p-version
Hybrid of above (e.g., kernels, splines, RBF)	Sums of exponentials	hp-version

Fred Daum, “Nonlinear filters: beyond the Kalman filter,” special tutorial issue of IEEE AES Systems Magazine, August 2005.

Raytheon

Summary

- **Importance of concentration of measure for nonlinear filters**
- **Curse of dimensionality for classical particle filters**
- **Mesh free adjoint method for solving the Fokker-Planck equation**
- **Mesh free adjoint method for Bayes' rule turned into an ODE using homotopy**