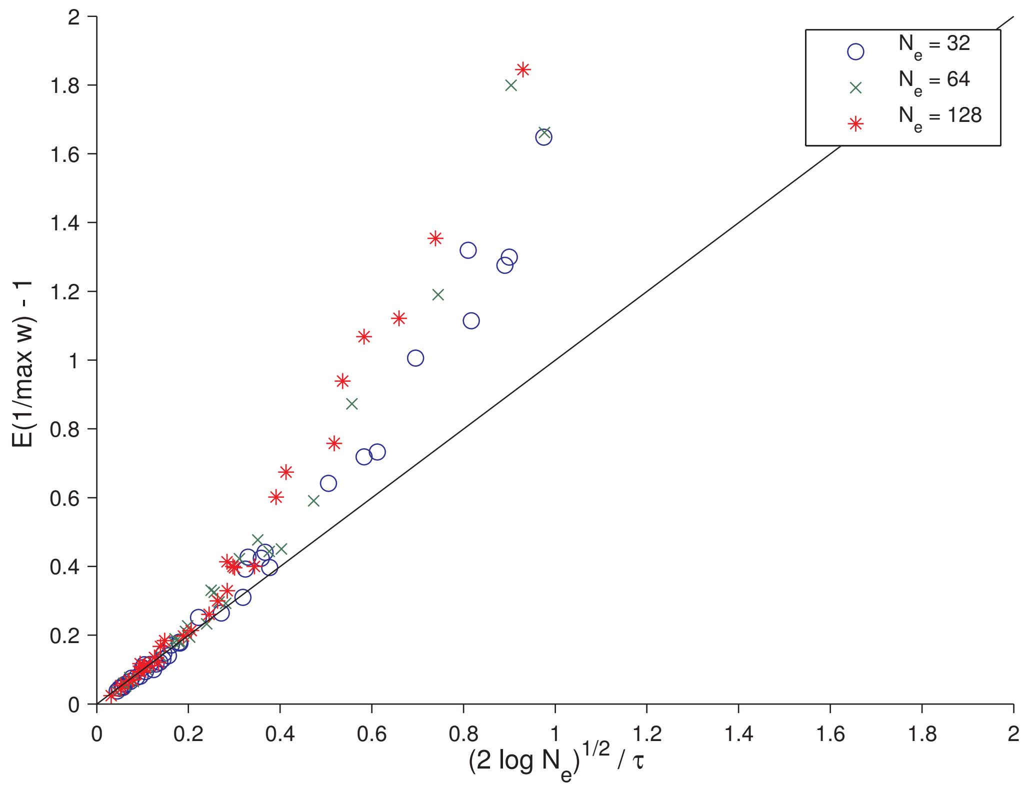
Performance Bounds for Particle Filters using the "Optimal" Proposal Density



▷ Chris Snyder
National Center for Atmospheric Research, Boulder Colorado, USA

Overview

- Simplest particle filter requires very large ensemble size, growing exponentially with the problem size.
- Can the use of the optimal proposal density fix this?

Preliminaries

Notation

- follow Ide et al. (1997) generally, except:
  ... $\dim(\mathbf{x}) = N_x$, $\dim(\mathbf{y}) = N_y$, ensemble size $= N_e$
  ... superscripts index ensemble members
- $\sim$ means "distributed as," e.g. $x \sim N(0,1)$. Also used for "asymptotic to"
- state evolution: $\mathbf{x}_k = M(\mathbf{x}_{k-1}) + \eta_k$
- observations: $\mathbf{y}_k = H(\mathbf{x}_k) + \epsilon_k$

Interchangeable terms

- particles $\equiv$ ensemble members
- sample $\equiv$ ensemble

Preliminaries: The Simplest Particle Filter

- begin with members $\mathbf{x}_{k-1}^i$ and weights w_{k-1}^i that "represent" $p(\mathbf{x}_{k-1}|\mathbf{y}_{k-1}^o)$
- compute $\mathbf{x}_k^i$ by evolving each member to t_k under the system dynamics
- re-weight, given new obs $\mathbf{y}_k^o$: $w_k^i \propto w_{k-1}^i p(\mathbf{y}_k^o|\mathbf{x}_k^i)$
- (resample)

Preliminaries: Sequential Importance Sampling

Basic idea

- suppose $p(\mathbf{x})$ is hard to sample from, but $\pi(\mathbf{x})$ is not.
- draw $\{\mathbf{x}^i\}$ from $\pi(\mathbf{x})$ and approximate

$$p(\mathbf{x}) \approx \sum_{i=1}^{N_e} w^i \delta(\mathbf{x} - \mathbf{x}^i), \quad \text{where } w^i \propto p(\mathbf{x}^i)/\pi(\mathbf{x}^i)$$

- $\pi(\mathbf{x})$ is the *proposal density*

Preliminaries: SIS (cont.)

Perform importance sampling sequentially in time

- Given $\{\mathbf{x}_{k-1}^i\}$ from $\pi(\mathbf{x}_{k-1})$, wish to sample from $p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_k^o)$
- choose proposal of the form

$$\pi(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_k^o) = \pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k^o)\pi(\mathbf{x}_{k-1})$$

- update weights using

$$w_k^i \propto \frac{p(\mathbf{x}_k^i, \mathbf{x}_{k-1}^i|\mathbf{y}_k^o)}{\pi(\mathbf{x}_k^i, \mathbf{x}_{k-1}^i|\mathbf{y}_k^o)} = \frac{p(\mathbf{y}_k^o|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{\pi(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_k^o)} w_{k-1}^i$$

Preliminaries: SIS (cont.)

PF literature shows that choice of proposal is crucial

Standard proposal: transition density from dynamics

- $\pi(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_k^o) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$
- $w_k^i \propto p(\mathbf{y}_k^o|\mathbf{x}_k^i)$
- members at t_k generated by evolution under system dynamics, as in EF

Preliminaries: SIS (cont.)

"Optimal" proposal: Also condition on most recent obs

- $\pi(\mathbf{x}_k, \mathbf{x}_{k-1} | \mathbf{y}_k^o) = p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{y}_k^o)$
- $w_k^i \propto p(\mathbf{y}_k^o | \mathbf{x}_{k-1}^i)$
- members at t_k generated, in a sense, by DA
- optimal in sense that it minimizes variance of weights over $\mathbf{x}_k^i$
- several recent PF studies use proposals that either reduce to or are related to the optimal proposal (van Leeuwen 2010, Morzfeld et al. 2011, Papadakis et al. 2010)

Degeneracy of PF Weights

- degeneracy $\equiv \max_i w_k^i \to 1$
- common problem, well known in PF literature
- for standard proposal, Bengtsson et al. (2008) and Snyder et al. (2008) show N_e must increase exponentially as problem size increases in order to avoid degeneracy
- What happens with optimal proposal?

A Simple Test Problem

Consider the system

$$\mathbf{x}_k = a\mathbf{x}_{k-1} + \eta_{k-1}, \quad \mathbf{y}_k = \mathbf{x}_k + \epsilon_k$$

where $\eta_{k-1} \sim N(0, q^2\mathbf{I})$ and $\epsilon_k \sim N(0, \mathbf{I})$.

For standard proposal:

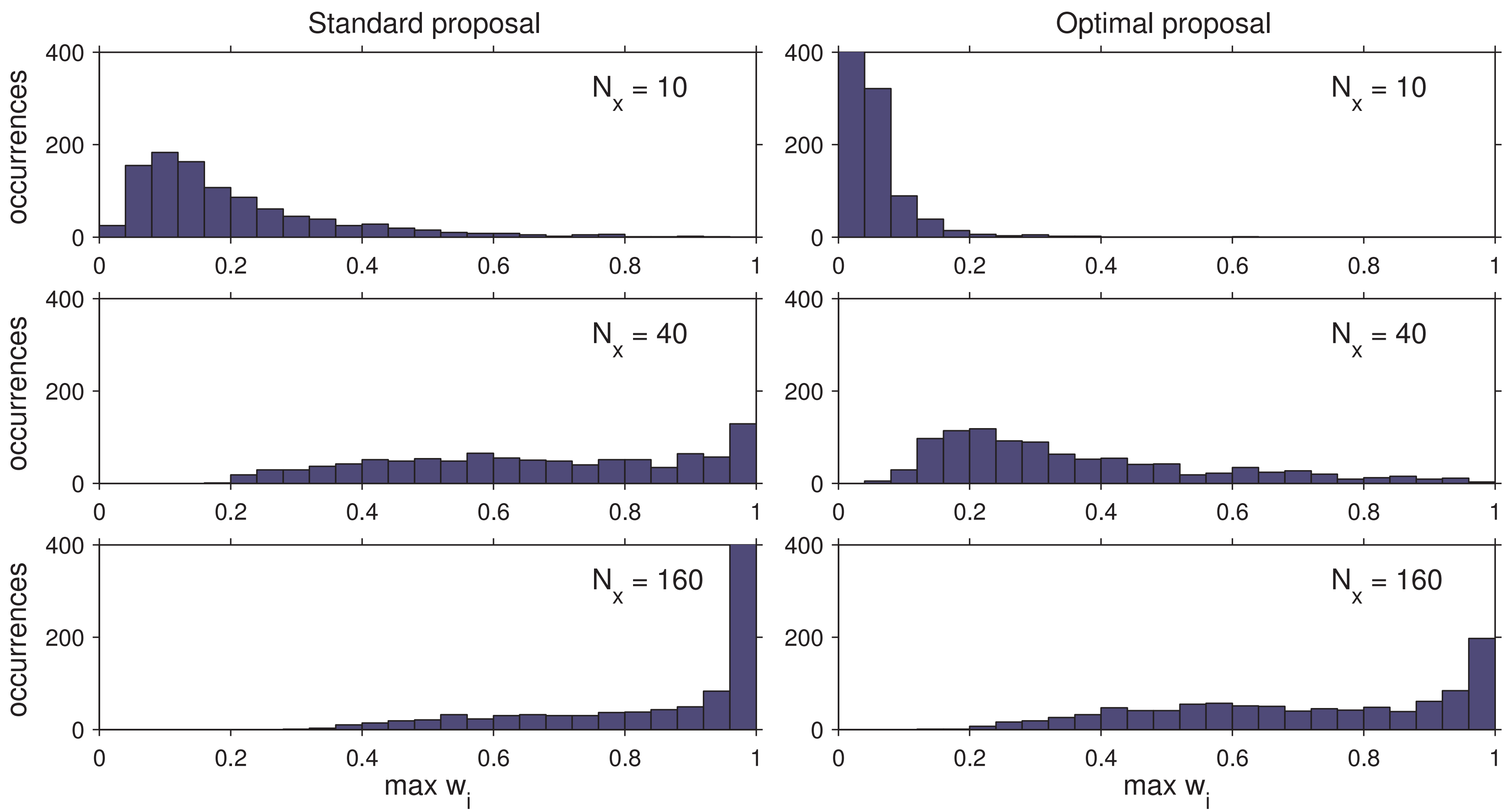
$$\mathbf{y}_k|\mathbf{x}_k \sim N(\mathbf{x}_k, \mathbf{I}), \quad w_k^i \propto \exp\left(-\frac{1}{2}|\mathbf{y}_k - \mathbf{x}_k^i|^2\right)$$

For optimal proposal:

$$\mathbf{y}_k|\mathbf{x}_{k-1} \sim N\left(a\mathbf{x}_{k-1}, \left(1+q^2\right)\mathbf{I}\right), \quad w_k^i \propto \exp\left(-\frac{|\mathbf{y}_k - a\mathbf{x}_{k-1}^i|^2}{2\left(1+q^2\right)}\right)$$

A Simple Test Problem (cont.)

- histograms of $\max_i w_k^i$ for $N_e = 10^3$, $a = q = 1/2$. 10^3 simulations.
- optimal proposal clearly reduces degeneracy



Behavior of Weights

Define

$$V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k) = -\log(w_k/w_{k-1}) = \begin{cases} -\log p(\mathbf{y}_k|\mathbf{x}_k), & \text{std. proposal} \\ -\log p(\mathbf{y}_k|\mathbf{x}_{k-1}), & \text{opt. proposal} \end{cases}$$

Interested in V as random variable with $\mathbf{y}_k$ known and $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$ distributed according to the proposal distribution at t_k and t_{k-1}, respectively.

Suppose each component of obs error is independent.

- $p(\mathbf{y}_k|\mathbf{x}_k)$, $p(\mathbf{y}_k|\mathbf{x}_{k-1})$ can be written as products
- V becomes a sum over log likelihoods for each component
- if terms in sum are nearly independent, $V \to$ Gaussian as $N_y \to \infty$
- infer asymptotic behavior of $\max w_k^i$ from known asymptotics for sample min of Gaussian

Behavior of Weights (cont.)

Let $\tau^2 = \mathsf{var}(V)$. Then for large N_e and large τ,

$$E(1/\max w_k^i) \sim 1 + \frac{\sqrt{2\log N_e}}{\tau}$$

(Bengtsson et al. 2008, Snyder et al. 2008)

As τ^2 increases, N_e must increase as $\exp(2\tau^2)$ to keep $E(1/\max w^i)$ fixed.

The Linear, Gaussian Case

Analytic results possible for linear, Gaussian case with general $\mathbf{R} = \mathrm{cov}(\epsilon_k)$, $\mathbf{Q} = \mathrm{cov}(\eta_k)$ and $\mathbf{P}_k = \mathrm{cov}(\mathbf{x}_k)$.

Proceed by transformation in obs space that simultaneously diagonalizes either $\mathbf{R}$ and $\mathbf{H}\mathbf{P}_k\mathbf{H}^T$, or $\mathbf{R} + \mathbf{H}\mathbf{Q}\mathbf{H}^T$ and $\mathbf{H}\mathbf{M}\mathbf{P}_{k-1}(\mathbf{H}\mathbf{M})^T$.

Then

$$\tau^2 = \sum_{j=1}^{N_y} \lambda_j^2 \, (\lambda_j^2/2 + y_{k,j}^2),$$

where λ_j^2 are eigenvalues of

$$\mathbf{A} = \begin{cases} \mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}_k\mathbf{H}^T\mathbf{R}^{-1/2}, & \text{std. proposal} \\ (\mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R})^{-1/2}\mathbf{H}\mathbf{M}\mathbf{P}_{k-1}(\mathbf{H}\mathbf{M})^T(\mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R})^{-1/2}, & \text{opt. proposal.} \end{cases}$$

Simple Test Problem, Revisited

$$V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k) = \begin{cases} -\log p(\mathbf{y}_k|\mathbf{x}_k), & \text{std. proposal} \\ -\log p(\mathbf{y}_k|\mathbf{x}_{k-1}), & \text{opt. proposal} \end{cases}$$

$$2V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k) = \begin{cases} \sum_{j=1}^{N_y}(y_{k,j} - x_{k,j})^2, & \text{std. proposal} \\ \left(1+q^2\right)^{-1} \sum_{j=1}^{N_y}(y_{k,j} - a x_{k-1,j})^2, & \text{opt. proposal} \end{cases}$$

$$\tau^2 = \mathrm{var}(V) = \begin{cases} N_y(a^2+q^2)\left(\frac{3}{2}a^2 + \frac{3}{2}q^2 + 1\right), & \text{std. proposal} \\ N_y a^2 \left(\frac{3}{2}a^2 + q^2 + 1\right)/(q^2+1)^2, & \text{opt. proposal} \end{cases}$$

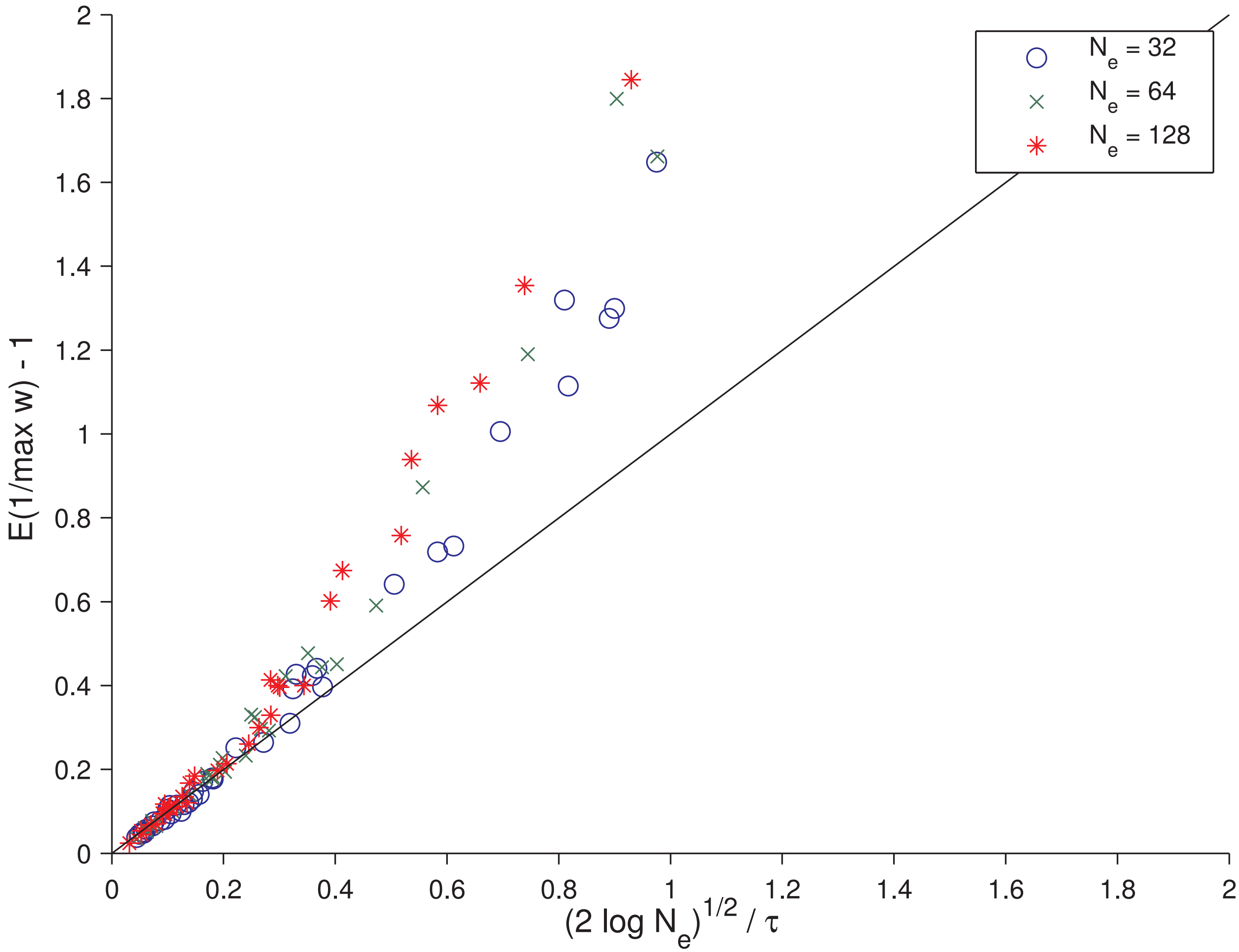
Simple Test Problem, Revisited (cont.)

$$\tau^2 = \mathrm{var}(V) = \begin{cases} N_y(a^2+q^2)\left(\frac{3}{2}a^2+\frac{3}{2}q^2+1\right), & \text{std. proposal} \\ N_y a^2\left(\frac{3}{2}a^2+q^2+1\right)/(q^2+1)^2, & \text{opt. proposal} \end{cases}$$

- ▷ τ^2(opt. proposal) always less than or equal to τ^2(std. proposal)
- ▷ τ^2 from the two proposals is equal only when $q = 0$
- ▷ opt. proposal reduces τ^2 by an $O(1)$ factor for reasonable values of a and q; $a = q = 1/2$ implies a factor of 5 reduction in τ^2.

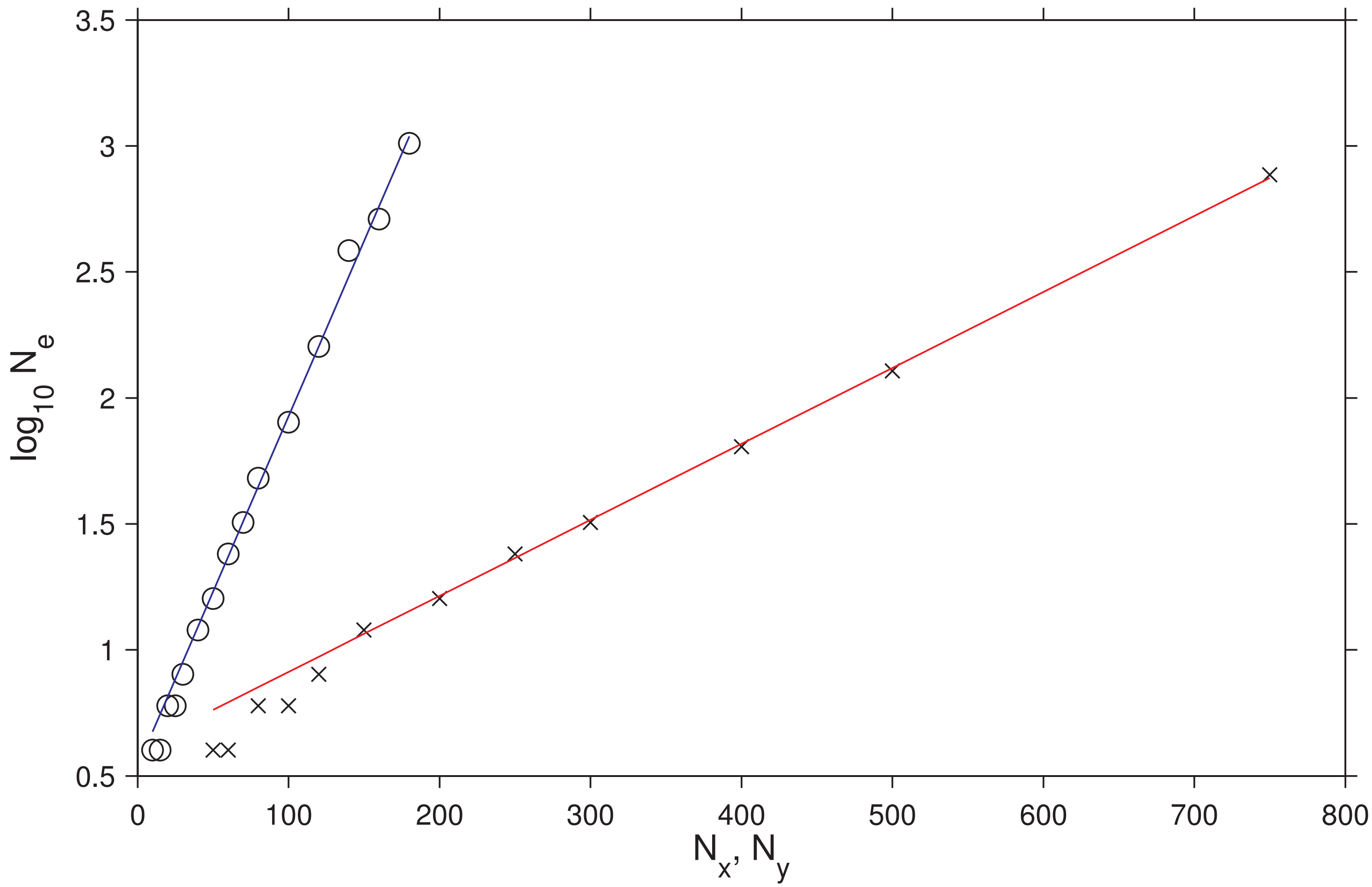
Simple Test Problem, Revisited (cont.)

- Theoretical prediction for $E(1/\max w^i)$ vs. simulations. Expectation is based on 10^3 realizations.



Simple Test Problem, Revisited (cont.)

- minimum N_e such that $E(1/\max w^i) \geq 1/0.8$ for standard proposal (circles) and optimal proposal (crosses) for $a = q = 1/2$.
- ratio of slopes of best-fit lines is 4.6, vs. asymptotic prediction of 5



N_y, N_x and Problem Size

$\tau^2 = \mathrm{var}(\log \text{likelihood})$ measures "problem size" for PF

- as τ^2 increases, N_e must increase as $\exp(2\tau^2)$ if $E(1/\max w^i)$ fixed.

Related to obs-space dimension

- in simple example, $\tau^2 \propto N_y$
- given by sum over e-values of obs-space covariance in general linear, Gaussian case—like an effective dimension

Analogy of τ^2 to dimension is incomplete

- τ^2 depends on obs-error statistics, increasing as $\mathbf{R}$ decreases
- τ^2 depends on proposal

N_y, N_x and Problem Size (cont.)

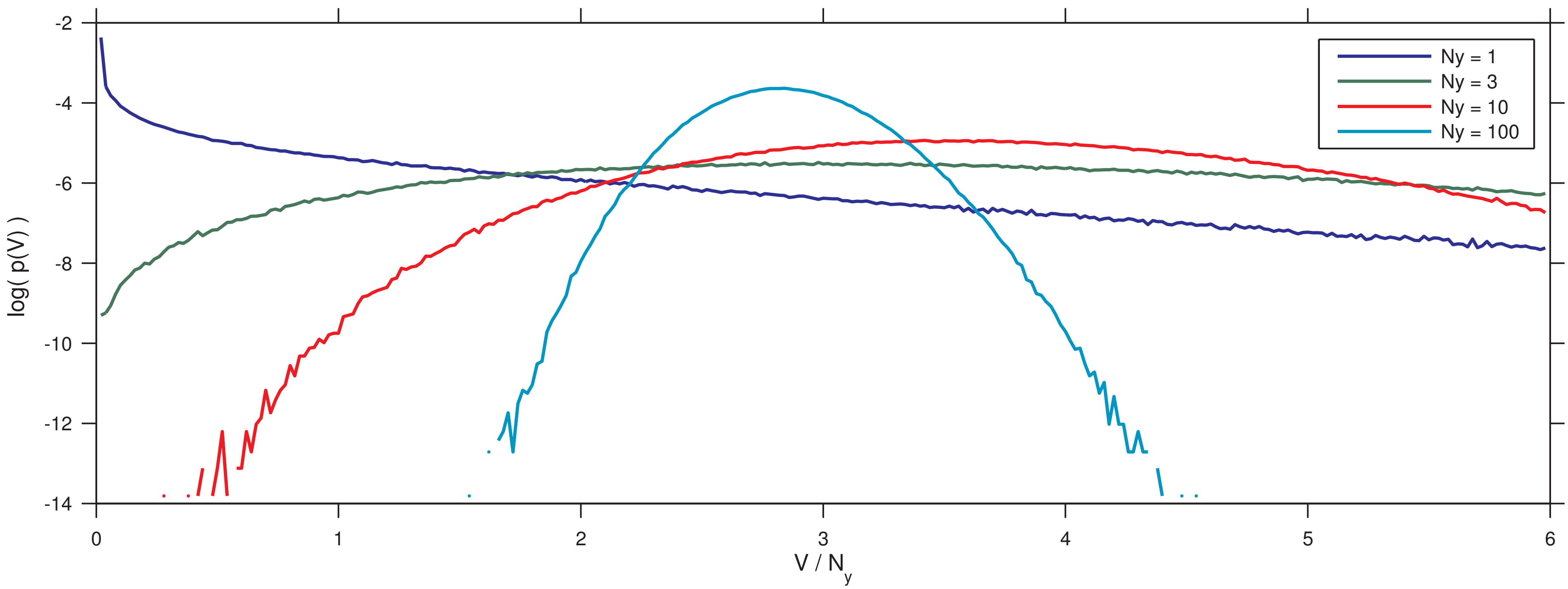
τ^2 depends explicitly *only* on obs-space quantities

How does N_x affect weight degeneracy?

- asymptotic relation of τ^2 and $E(1/\max w^i)$ requires $V(\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{y}_k)$ to be $\sim$ Gaussian over $\mathbf{x}_k$
- $\sim$ Gaussianity of $V(\mathbf{x}_k)$ only if $N_x = \dim(\mathbf{x})$ is large and components of $\mathbf{x}$ are sufficiently independent

N_y, N_x and Problem Size (cont.)

- log of pdf of V: $N_y = 1$, 3, 10, 100; $\mathbf{x} \sim N(0, \mathbf{I})$, $\mathbf{H} = \mathbf{I}$, $\epsilon \sim N(0, \mathbf{I})$
- recall that max weight depends on left-hand tail of $p(V)$, which changes greatly as N_x increases and $V \to$ Gaussian



Summary

- As was the case for the standard proposal, the optimal proposal requires N_e to increase exponentially with the "problem size" to avoid degeneracy.
- Exponential rate of increase is quantitatively smaller for the optimal proposal; necessary ensemble size may therefore be *much* smaller in a given problem.
- Benefits of optimal proposal dependent on magnitude of system noise.
- (In some cases, possible to enhance performance of optimal proposal by artificially increasing system noise in forecast model used in filtering.)

Recommendation

New PF algorithms intended for high-dimensional systems should be evaluated first on the simple test problem given here.

References

Bengtsson, T., P. Bickel and B. Li, 2008: Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *IMS Collections*, **2**, 316–334. doi: 10.1214/193940307000000518.

Morzfeld, M., X. Tu, E. Atkins and A. J. Chorin, 2011: A random map implementation of implicit filters. *J. Comput. Phys.* **231**, 2049–2066.

van Leeuwen, P. J., 2010: Nonlinear data assimilation in geosciences: an extremely efficient particle filter. *Quart. J. Roy. Meteor. Soc.* **136**, 1991–1999.

Papadakis, N., E. Mémin, A. Cuzol and N. Gengembre, 2010: Data assimilation with the weighted ensemble Kalman filter. *Tellus* **62A**, 673–697.

Snyder, C., T. Bengtsson, P. Bickel and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Monthly Wea. Rev.*, **136**, 4629–4640.