Optimal Sensor Placement for Data Assimilations

Wei Kang[1] and Liang Xu[2]

[1]Naval Postgraduate School, Monterey, CA, USA
[2]Naval Research Laboratory, Monterey, CA, USA

The 9th Workshop on Adjoint Model Applications in Dynamic Meteorology

Cefalu, Sicily, Italy 10-14 October 2011

Outline

- Background & motivation
- Definition of observability
- Two potential applications of observability
- An example using Burgers' equation
- Concluding remarks

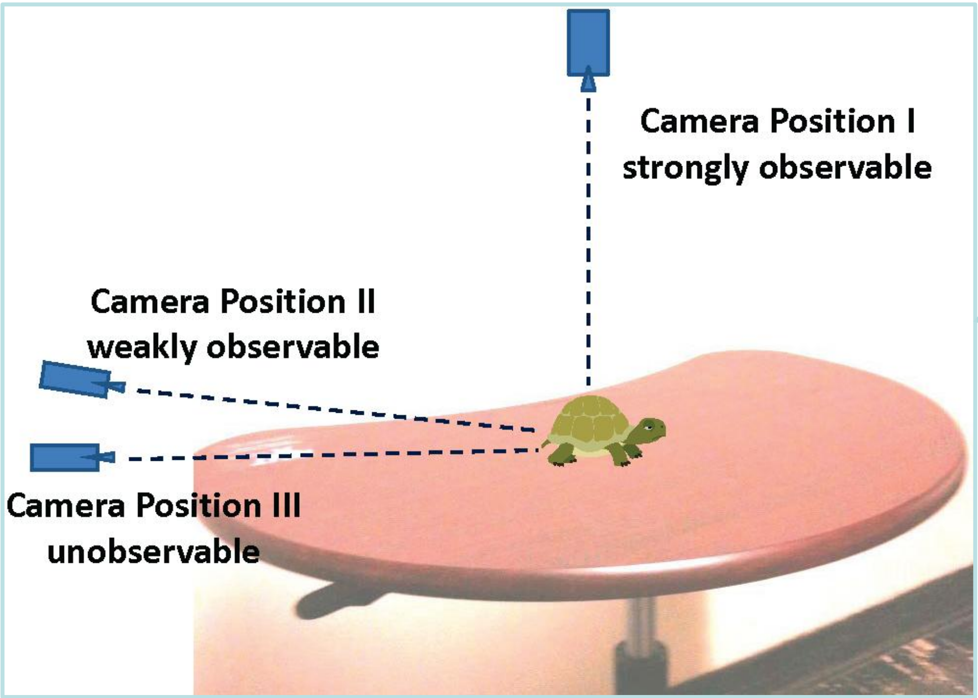
Background
&
Motivation

Motivation

- **Targeting observation: where to put additional observations during a field experiment?**
  - Find the sensitivity of forecast errors with respect to the initial condition using adjoint or ensemble methods
  - Make additional observations at the sensitivity regions
  - Find the observation impact (**after the fact**) on the reduction of forecast errors using adjoint or ensemble methods
- **Assessing the impact of current and/or future sensors: what is the benefit to assimilate the current and/or future sensors?**
  - Conduct OSE (Observing System Experiment)
  - Conduct OSSE (Observing System Simulation Experiment)
- **Propose yet another concept to address similar issues**
  - Define observability
  - Find the optimal observation configuration that provides the maximum observability for a given dynamic system

A turtle on a table

Observability: practical well-posedness of inverse problems.



A turtle on a table

At camera position I, the turtle is strongly observable.

At camera position II, the turtle is weakly observable.

At camera position III, the turtle is unobservable.

Sensor configuration may have significant impact on the effectiveness and efficiency of the observations

Definition
of
Observability

Observability from Wikipedia

Observability, in control theory, is a measure for how well internal states of a system can be inferred by knowledge of its external outputs. The observability and controllability of a system are mathematical duals. The concept of observability was introduced by American-Hungarian scientist Rudolf E. Kalman for linear dynamic systems.

Observability related study in atmospheric sciences:
Cohn and Dee (1988), Menard (1994), Daley (1995)

Observability

$$x(n) = M\big(n-1, x(n-1)\big)$$

$$y(n) = H(x(n), \lambda)$$

$$J(x_0, \delta x_0, \lambda) = \delta x_0^{\mathrm{T}} \mathrm{W} \delta x_0$$
$$+ \sum_{n=1}^{N} \|y(n, \lambda; x_0 + \delta x_0) - y(n, \lambda; x_0)\|_{\mathrm{Y}}$$

where x is a state vector (internal state) in model space, y (external output) is observation vector in observation space, λ is the sensor configuration (e.g. the sensor locations), W is a weight matrix, $\|\cdot\|_{\mathrm{Y}}$ is a norm for the observation operator. Cost function J is the square of the distance between two initial states, x_0 and $x_0 + \delta x_0$, and the two sets of observations associated with the corresponding initial states.

Observability …

Definition. Let $\rho > 0$ be a positive number. Then the number ϵ is defined as follows

$$\epsilon^2 = \min_{\delta x 0} J(x_0, \delta x_0, \lambda) \qquad (1)$$

subject to $||\delta x_0|| = \rho, \delta x_0 \in S$

where S is a reduced space for estimation

The scalar $\boldsymbol{\epsilon}$ represents the smallest variation (or distance) of y corresponding to the variation δx_0 in x_0. A small $\boldsymbol{\epsilon}$ implies that x_0 is less observable.

The ratio $\boldsymbol{\rho/\epsilon}$ is a measure of observability. It is called an **unobservability index**. A small value of $\boldsymbol{\rho/\epsilon}$ implies strong observability.

Applications of Observability

Observability of A Sensor

- **Assessing sensor impact**
  - We can in principle indirectly assess the impact of a given sensor, current or future one through the calculation of the **unobservability index**, $\boldsymbol{\rho/\epsilon}$, with respect to the variations of the initial condition.
- **Required major components**
  - Dynamic model - M (e.g. NWP model)
  - Observation operator - H
  - Formulation of the cost function
  - A minimization algorithm (not easy)

Optimal Sensor Placement

The concept of observability provides a quantitative measure of the quality of sensor information.

The best sensor configuration (such as sensor locations λ) are those that maximize the value of ϵ, as defined in (1), following performance measure, i.e.

$$\max_{\delta x0} \epsilon(\lambda) \qquad (2)$$

subject to $\lambda_{min} \leq \lambda \leq \lambda_{max}$

Eq (1) represents a minimization problem and eq (2) represents a maximization problem. While both problems are numerically challenging to solve, it is especially true for the problem represented by eq (2).

An Example
using
Burgers' Equation

Objectives of the Example

- **Illustrate all components/procedures needed**
  - Dynamic model, observation operator, cost function, minimization and maximization algorithms
  - Optimal sensor locations
- **Demonstrate its usefulness in data assimilation**
  - 4D-Var data assimilation experiments
  - Results obtained from both equally spaced sensors and optimal sensor placement using Monte Carlo experiments
  - Robustness analysis

Burgers' Equations

Consider a system

$$\frac{\partial U(x,t)}{\partial t} + U(x,t)\frac{\partial U(x,t)}{\partial x} - \kappa\frac{\partial^2 U(x,t)}{\partial x^2} = 0, \qquad \begin{array}{l} U(0,t) = 0 \\ U(2\pi,t) = 0 \end{array}$$

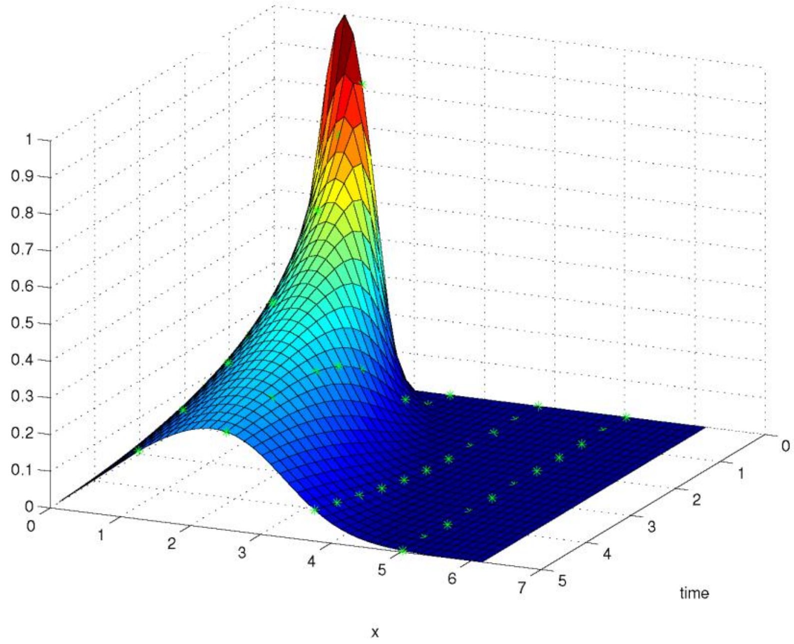
$$y_i(t_j) = U(\lambda_i, t_j), \qquad \begin{array}{l} i = 1, 2, \cdots, N_s \\ j = 0, 1, 2, \cdots, N_t \end{array}$$

λ_i – sensor location , N_s – number of sensors ,

time interval/N_t – sensor sampling rate

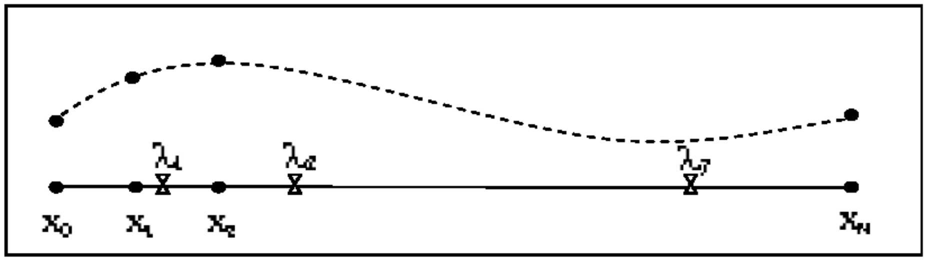
Problem:

- where to place 7 “weather stations” (in x-direction) that is measured at each model time step?
- Dimension of the model is 50 (in x-direction).



Numerical Solution

Discretized Model at uniformly spaced nodes: $u_i(t) = U(x_i, t)$



$$\begin{aligned}
\dot{u}_1(t) &= -u_1(t)\frac{u_2(t) - f_1(t)}{2\Delta x} + \kappa\frac{u_2(t) + f_1(t) - 2u_1(t)}{\Delta x^2} \\
\dot{u}_2(t) &= -u_2(t)\frac{u_3(t) - u_1(t)}{2\Delta x} + \kappa\frac{u_3(t) + u_1(t) - 2u_2(t)}{\Delta x^2} \\
&\vdots \\
\dot{u}_{N-1}(t) &= -u_{N-1}\frac{f_2(t) - N_{N-2}(t)}{2\Delta x} + \kappa\frac{f_2(t) + u_{N-2}(t) - 2u_{N-1}(t)}{\Delta x^2}
\end{aligned}$$

Output and its metric

$$||\hat{y}(t) - y(t)|| = \sum_{k=0}^{N_t} ||\mathrm{interp}(u(t_k; u_0), \lambda) - \mathrm{interp}(u(t_k; u_0 + \delta u_0), \lambda)||^2$$

Simple standard numerical techniques are used here

Parameters used

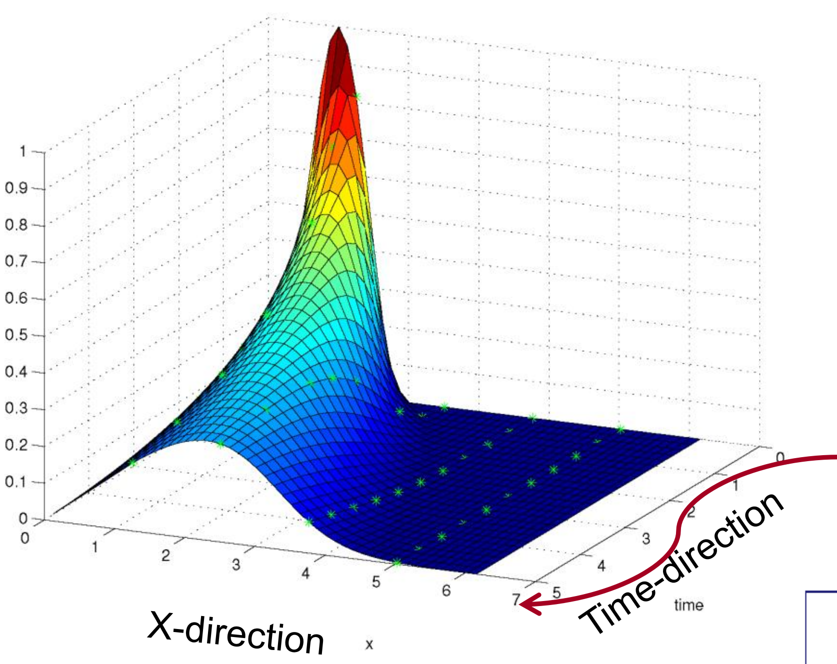
$W = \{\alpha_0 + \sum_{k=1}^{6} \alpha_k \cos(kx) + \beta_k \sin(kx)\}$	(Space for estimation)
$\kappa = 0.14$	
$L = 2\pi$	(length of x-interval)
$N = 50$	(dimension of the model)
$T = 5$	(final time)
$N_t = 20, \Delta t = T/N_t$	(time step size of sensors)
$U_0(x) = \begin{cases} x^3(2-x)^3, & x \leq 2 \\ 0, & x > 2 \end{cases}$	(nominal initial condition)
$f_1(t) = 0, f_2(t) = 0$	(boundary condition)
$\rho = 0.01$	(radius of the variation of u_0)

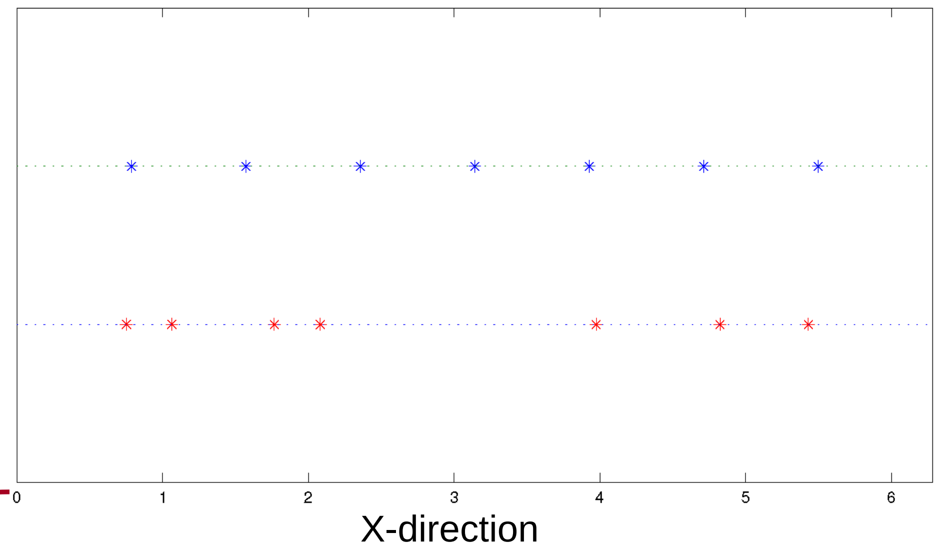
Sensor Locations

Sensor locations:

equally spaced: $\begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_7 \end{bmatrix} = \begin{bmatrix} \frac{2\pi}{8} & 2\frac{2\pi}{8} & \cdots & 7\frac{2\pi}{8} \end{bmatrix}$

optimal locations: $\begin{bmatrix} 0.75 & 1.06 & 1.77 & 2.08 & 3.98 & 4.83 & 5.43 \end{bmatrix}$





	Equally spaced	Optimal location
ρ/ϵ	12.92	1.75

Optimal sensor locations provide strong observability (small ρ/ϵ)!

4D-Var Data Assimilation

Dynamical System

$$x_{n+1} = \mathcal{M}_n(x_n)$$
$$y_n = Hx_n$$

Estimation

$$x_n^a = x_n^b + g_n, \quad 0 \le n \le N$$
$$g_n = P_n^b H^T z_n$$
$$(HP_n^b H^T + R)z_n = (y_n - Hx_n^b)$$

Computation

$$f_n = M_n^T f_{n+1} + H^T z_n, \quad f_{N+1} = 0$$
$$g_{n+1} = M_n g_n, \quad g_0 = P_0^b f_0$$

$M_n g_n, \;\; M_n^T f_n$ are computed using linear tangent model and conjugate model

- To examine the usefulness of the proposed observability in data assimilation, two sets of 4D-Var data assimilation experiments are carried out.
- The only differences between the two sets of 4D-Var experiments are the sensor configurations.
- One configuration is equally spaces sensors, while the other is optimally placed sensors
- Typical 4D-Var data assimilation setup for a simple problem.

Monte Carlo Experiments

Background

$\{u_k^b(t)|k = 1, 2, \cdots, 200\}$

Sensor data

$$\begin{bmatrix} y_1(t_0) & y_1(t_1) & \cdots & y_1(t_{N_t}) \\ y_2(t_0) & y_2(t_1) & \cdots & y_2(t_{N_t}) \\ \cdots & \cdots & \cdots & \cdots \\ y_{N_y}(t_0) & y_2(t_1) & \cdots & y_{N_y}(t_{N_t}) \end{bmatrix}$$

$y(t_i) = y^{true}(t_i) + Rv_i$

$v_i \in I\!R^{N_y}$ are standard independent white Gaussian noise

$R = 0.001$

Root-mean-square Error

The error of $u^b(\cdot) - u^{truth}(\cdot)$: 0.3638 or 11.14%

The error of $u^b(0) - u^{truth}(0)$: 0.3524 or 15.13%

- Two hundred sets of background are generated.
- They are used to perform two hundred corresponding 4D-Var data assimilation for both sensor configurations, respectively.

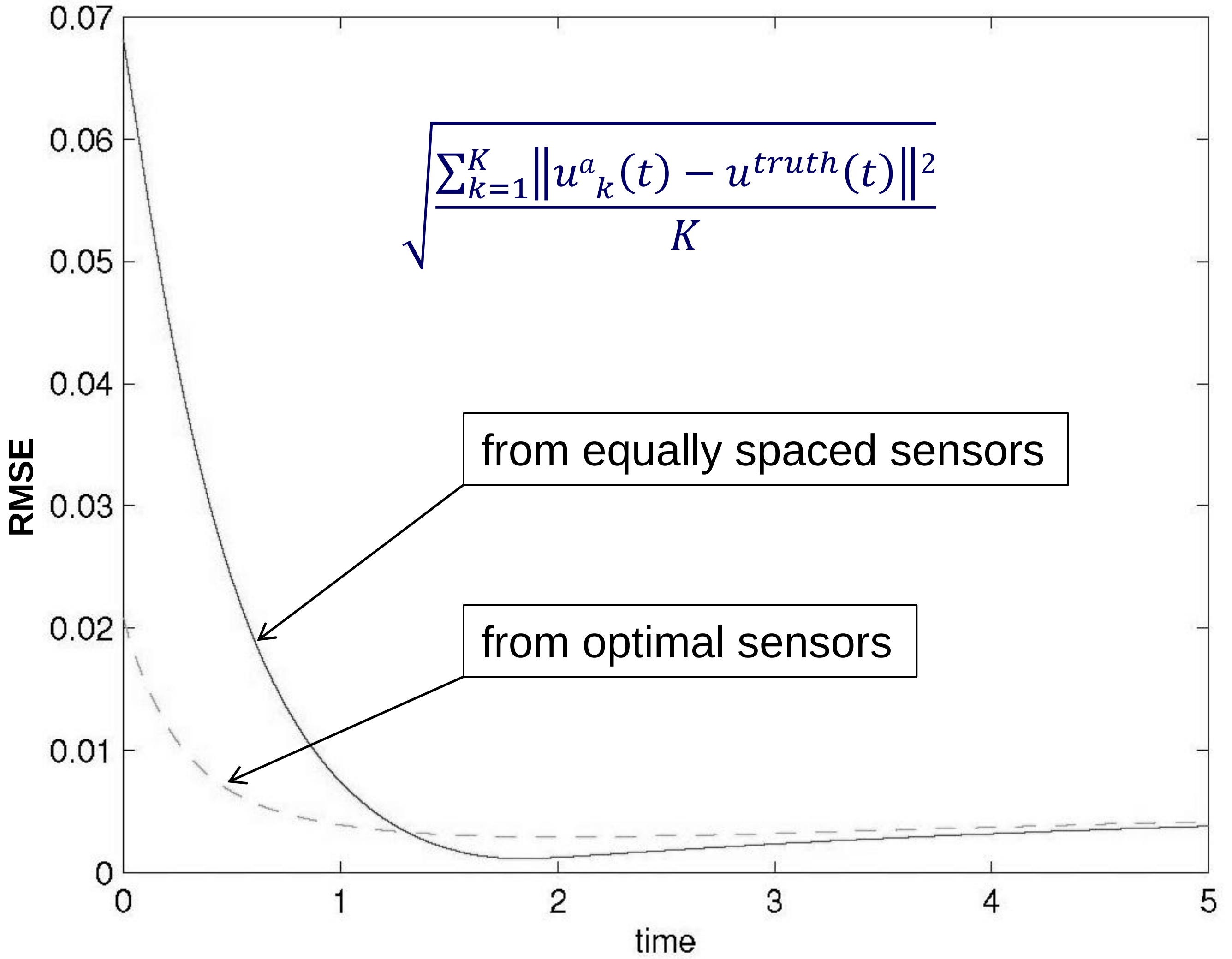
Results

The overall error of the analysis (RMSE) $u^a(t) = \sqrt{\frac{\sum_{k=1}^{K} \| u^a{}_k(t) - u^{truth}(t) \|^2}{K}}$,K=200

	Observability ρ/ϵ	RSME of $u^a(0)$	RSME of $u^a(\cdot)$
Equally spaced sensors	12.92	0.1652	0.0790
Optimal sensor location	1.75	0.0788	0.0326
Improvement	86%	52%	58%

Sensor data with higher observability (smaller ρ/ϵ) results in higher estimation accuracy. It implies that sensor data with higher observability (smaller ρ/ϵ) contains more valuable information than those with low observability.

RMSE of Trajectory



Maximizing observability results in an overall improvement of the estimation accuracy.

Robustness Analysis

Summary of the analysis:

	Variation of P_0^b	Variation of R	RSME of $u^a(\cdot)$	Improvement
Equally spaced	10%	0	0.0938	
Optimal locations	10%	0	0.0351	62%
Equally spaced	50%	0	0.1191	
Optimal locations	50%	0	0.0887	25%
Equally spaced	0	100%	0.0646	
Optimal locations	0	100%	0.0407	37%

In all cases, the optimal sensor locations result in significantly improved estimation accuracy, which ranges from 25% to 62%

Robustness Analysis…

Uniform distribution:

Background error is bounded by $[-0.17, 0.17]$

Sensor error is bounded by $[-0.003, 0.003]$

Summary:

	Observability ρ/ϵ	RSME of $u^a(0)$	RSME of $u^a(\cdot)$
Equally spaced sensors	12.92	0.1376	0.0645
Optimal sensor location	1.75	0.0893	0.0372
Improvement	86%	35%	42%

The results are consistent with the ones with normally distributed errors in both background and sensor noises.

Concluding Remarks

- Observability is defined as the theoretical foundation and cost function for the optimal design of sensor locations.
- Observability is computed using empirical covariance matrix method.
- For the Burgers equation, optimal sensor locations lead to significantly improved estimation accuracy in 4D-Var data assimilations.
- The optimal sensor locations are robust.
- Many questions are raised for long term research, including optimal trajectory planning subject to complicated constraints; developing computational algorithms for the observability of large scale systems; Optimal sensor configurations in a more general sense.

Thank you

Observability

System:

$$\dot{x} = f(t, x, \mu), \quad \text{—system}$$

$$y = y(t, x, \mu), \quad \text{—system output}$$

Definition

Given a trajectory $(x(t), \mu)$, $t \in [t_0, t_1]$ and $\rho > 0$. The observability of $(x(0), \mu)$ is measured by the ratio ρ/ϵ, where

$$\epsilon = \min_{(\delta x(0), \delta \mu)} ||y(t, \hat{x}(t), \hat{\mu}) - y(t, x(t), \mu)||_Y$$

subject to

$$||(\delta x(0), \delta \mu)|| = \rho$$

$$\dot{\hat{x}} = f(t, \hat{x}, \hat{\mu}), \quad \hat{x}(0) = x(0) + \delta x(0), \hat{\mu} = \mu + \delta \mu$$

$$(\delta x(0), \delta \mu) \in W$$

Observability…

- W: space of estimation in which an estimate of the state with adequate accuracy exists. The estimate is updated at each time step using vectors in W.
- ϵ measures the sensitivity of y relative to the variation of $(x(0), \mu)$. A small value of ρ/ϵ implies strong observability of $(x(0), \mu)$.
- For linear systems, ϵ^2/ρ^2 equals the smallest eigenvalue of observability gramian
- The definition is applicable with general metrics, $||\cdot||_{L^p}$, $||\cdot||_\infty$, ……

Empirical Covariance Matrix Method

Empirical Covariance Matrix Method - A computational algorithm

Suppose the metrics of y and $u_0 = u(x, 0)$ are defined by inner products

$$||y||_Y = \sqrt{< y, y >_Y}, \quad ||u_0||_W = \sqrt{< u_0, u_0 >_W}$$

Let $\{v_1, v_2, \cdots, v_{n_z}\}$ be a basis of W. Define

$$\Delta_i(t) = \frac{1}{2\rho}\left(y(t, u_0 + \rho v_i) - y(t, u_0 - \rho v_i)\right)$$

$$G_Y = (< \Delta_i, \Delta_j >_Y)_{i,j=1}^{n_z}, \quad G_W = (< v_i, v_j >_W)_{i,j=1}^{n_z}$$

Then

$$\rho^2/\epsilon^2 \approx \frac{1}{\lambda_{min}}$$

where λ_{min} is the smallest eigenvalue of G_Y relative to G_W

Projection Gradient Method

Definition

Let λ be the coordinates of the sensors. Then ϵ is a function of λ, $\epsilon(\lambda)$. The optimal sensor locations are defined by

$$\max_{\lambda} \epsilon(\lambda)$$

subject to

$$\lambda_{min} \leq \lambda \leq \lambda_{max}$$

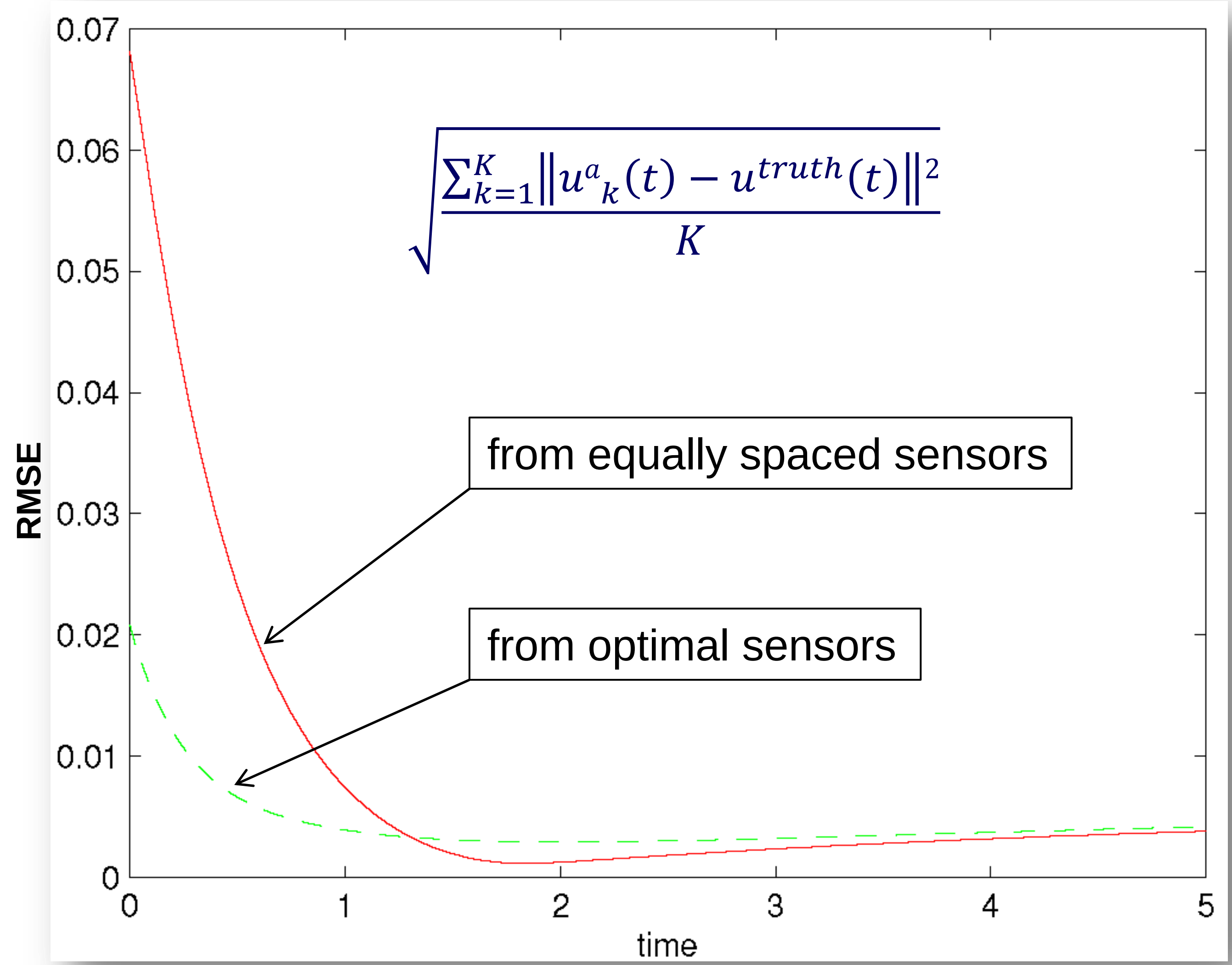
Computation - projection gradient method

Let $\nabla_\lambda = \dfrac{\partial \epsilon}{\partial \lambda}$. The search direction is defined by

$$\bar{\nabla}_{\lambda,i} = \begin{cases} 0, & \text{if } \lambda_i = \lambda_{min,i} \text{ and } \nabla_{\lambda,i} < 0 \\ 0, & \text{if } \lambda_i = \lambda_{max,i} \text{ and } \nabla_{\lambda,i} > 0 \\ \nabla_{\lambda,i}, & \text{otherwise} \end{cases}$$

Armijo algorithm model is applied in the direction $\bar{\nabla}_\lambda$.

RMSE of Trajectory



Maximizing observability results in an overall improvement of the estimation accuracy.