The Legacy of Rudolph Kalman

Blending Data and Mathematical Models

Andrew Stuart

Computing and Mathematical Sciences
California Institute of Technology

AFOSR, DARPA, EPSRC, ONR
Potsdam

August 24th 2018

Overview

Kalman State Estimation

Kalman Filter

State Space Model

$$
\begin{aligned}
\text{Dynamics Model:} \quad & v_{n+1} = Mv_n + \xi_n, \quad n \in \mathbb{Z}^+ \\
\text{Data Model:} \quad & y_{n+1} = Hv_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\
\text{Probabilistic Structure:} \quad & v_0 \sim N(m_0, C_0), \quad \xi_n \sim N(0, \Sigma), \quad \eta_n \sim N(0, \Gamma) \\
\text{Probabilistic Structure:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \text{ independent}
\end{aligned}
$$

- Born: Budapest, May 19, 1930.
- Died: Florida, July 2, 2016.
- BS and MS from MIT, 1953, 1954.
- Positions at Stanford, ETH, U of Florida.
- US National Academy of Engineering 1991.
- US National Academy of Sciences 2004.
- US National Medal of Science 2008.
- Draper Prize, Kyoto Prize, Steele Prize . . .

Kalman Filter

State Space Model

$$
\begin{aligned}
\text{Dynamics Model:} \quad & v_{n+1} = M v_n + \xi_n, \quad n \in \mathbb{Z}^+ \\
\text{Data Model:} \quad & y_{n+1} = H v_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\
\text{Probabilistic Structure:} \quad & v_0 \sim N(m_0, C_0), \quad \xi_n \sim N(0, \Sigma), \quad \eta_n \sim N(0, \Gamma) \\
\text{Probabilistic Structure:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \text{ independent}
\end{aligned}
$$

- J. Basic Engineering **82**(1960); see [1].
- 27,307 Google Scholar citations.
- Navigational and guidance systems.
- Apollo 11.
- $Y_n = \{y_\ell\}_{\ell=1}^n$.
- $v_n | Y_n \sim N(m_n, C_n)$.
- $(m_n, C_n) \mapsto (m_{n+1}, C_{n+1})$.

Kalman Filter

Sequential Optimization Perspective

$$\text{Predict:} \quad \widehat{m}_{n+1} = Mm_n, \quad n \in \mathbb{Z}^+$$
$$\text{Model/Data Compromise:} \quad J_n(m) = \frac{1}{2}|m - \widehat{m}_{n+1}|^2_{\widehat{C}_{n+1}} + \frac{1}{2}|y_{n+1} - Hm|^2_\Gamma$$
$$\text{Optimize:} \quad m_{n+1} = \text{argmin}_m J_n(m).$$

- $|\cdot|_A = |A^{-\frac{1}{2}} \cdot|$ for $A > 0$.
- Updating $\widehat{C}_{n+1}$ is expensive: $\mathcal{O}(d^2)$ storage.
- d the state space dimension.

3DVAR Filter

State Space Model

$$\begin{aligned}
\text{Dynamics Model:} \quad & v_{n+1} = \Psi(v_n) + \xi_n, \quad n \in \mathbb{Z}^+ \\
\text{Data Model:} \quad & y_{n+1} = Hv_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\
\text{Probabilistic Structure:} \quad & v_0 \sim N(m_0, C_0), \quad \xi_n \sim N(0, \Sigma), \quad \eta_j \sim N(0, \Gamma) \\
\text{Probabilistic Structure:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \text{ independent}
\end{aligned}$$

- Introduced in UK Met Office.
- Primary proponent: Andrew Lorenc.
- Quart J. Roy. Met. Soc. **112**(1986).
- J. Met. Soc. Japan **99**(1997).
- $\{v_n\} \mapsto \{v_{n+1}\}$.

3DVAR

Sequential Optimization Perspective

$$\text{Predict:} \quad \widehat{v}_{n+1} = \Psi(v_n), \quad n \in \mathbb{Z}^+$$

$$\text{Model/Data Compromise:} \quad J_n(v) = \frac{1}{2}|v - \widehat{v}_{n+1}|^2_{\widehat{C}} + \frac{1}{2}|y_{n+1} - Hv|^2_{\Gamma}$$

$$\text{Optimize:} \quad v_{n+1} = \operatorname{argmin}_v J_n(v).$$

- $\widehat{C}$ is a fixed model covariance (not updated sequentially).
- $\widehat{C}$ chosen to have simple, computable, structure (Fourier).

Ensemble Kalman Filter

State Space Model

$$
\begin{aligned}
\text{Dynamics Model:} \quad & v_{n+1} = \Psi(v_n) + \xi_n, \quad n \in \mathbb{Z}^+ \\
\text{Data Model:} \quad & y_{n+1} = Hv_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+ \\
\text{Probabilistic Structure:} \quad & v_0 \sim N(m_0, C_0), \quad \xi_n \sim N(0, \Sigma), \quad \eta_j \sim N(0, \Gamma) \\
\text{Probabilistic Structure:} \quad & v_0 \perp \{\xi_n\} \perp \{\eta_n\} \text{ independent}
\end{aligned}
$$

- Introduced by Geir Evensen.
- J. Geophysical Research **99**(1994).
- Motivated by extended Kalman filter; see [2].
- Jazwinski (1970) [3], Ghil et al (1981) [4].
- Original paper in ocean dynamics.
- Used in weather forecasting centres worldwide.
- $\{v_n^{(k)}\}_{k=1}^K \mapsto \{v_{n+1}^{(k)}\}_{k=1}^K$.

Ensemble Kalman Filter

Sequential Optimization Perspective

$$
\begin{aligned}
\text{Predict:} \quad & \widehat{v}_{n+1}^{(k)} = \Psi(v_n^{(k)}) + \xi_n^{(k)}, \quad n \in \mathbb{Z}^+ \\
\text{Model/Data Compromise:} \quad & J_n^{(k)}(v) = \frac{1}{2}|v - \widehat{v}_{n+1}^{(k)}|^2_{\widehat{C}_{n+1}} + \frac{1}{2}|y_{n+1}^{(k)} - Hv|^2_{\Gamma} \\
\text{Optimize:} \quad & v_{n+1}^{(k)} = \mathrm{argmin}_v J_n^{(k)}(v).
\end{aligned}
$$

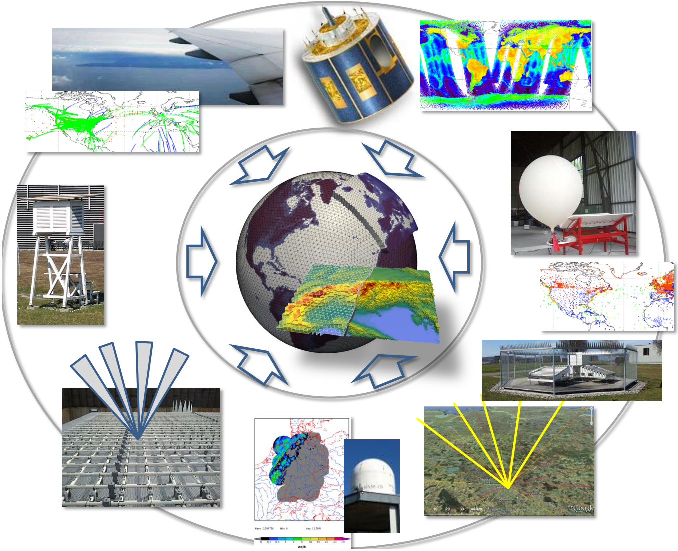
- $\widehat{C}_{n+1}$ is empirical covariance of the $\{\widehat{v}_{n+1}^{(k)}\}$.
- Updating $\widehat{C}_n$ requires only $\mathcal{O}(Kd)$ storage.

Weather Forecasting

w/Law

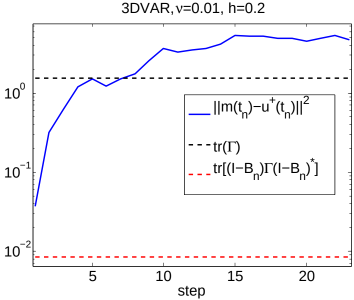
Weather Forecasting: Data

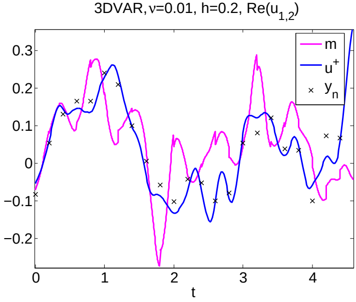
courtesy Roland Potthast(DWD)



Data Fails to Overcome Butterfly Effect

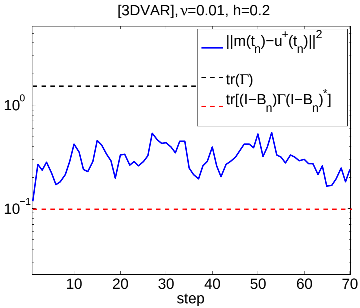
KJH Law and AM Stuart, Monthly Weather Review, 2014.

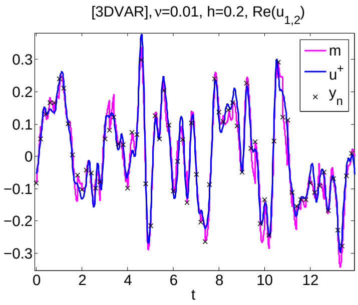




Theory Backed use of Data Overcomes Butterfly Effect

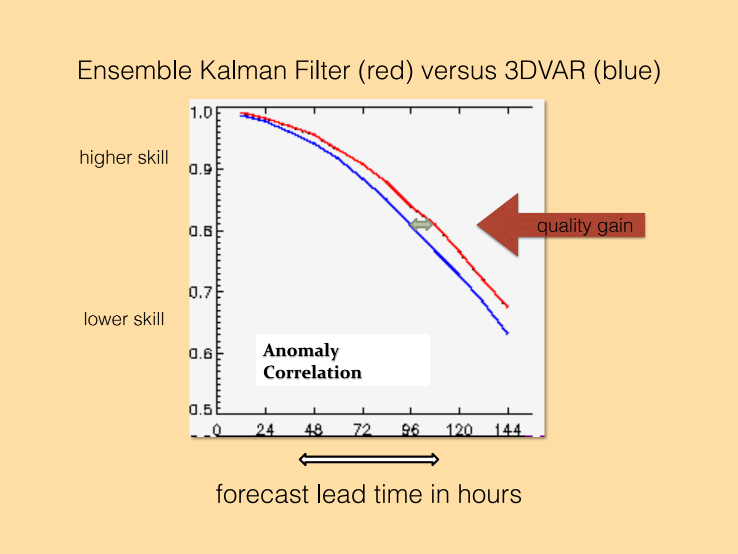
KJH Law and AM Stuart, Monthly Weather Review, 2014.





Impact of EnKF over 3DVAR

courtesy Roland Potthast(DWD)



Ensemble Kalman Inversion

w/Iglesias and Law

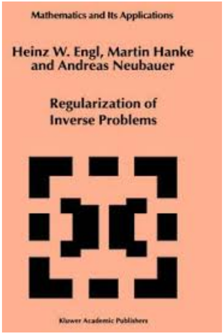
w/Schillings

Inverse Problem

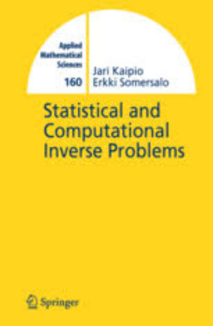
Problem Statement

Find u from y where $G : \mathcal{U} \mapsto \mathcal{Y}$, where $\mathcal{U}, \mathcal{Y}$ are Hilbert spaces, η is noise and

$$y = G(u) + \eta, \quad \eta \sim N(0, \Gamma).$$



A.N. Tikhonov (1963)



J. Franklin (1970)

Inverse Problem

Dynamical Formulation

$$\text{Dynamics Model:} \quad u_{n+1} = u_n, \quad n \in \mathbb{Z}^+$$
$$\text{Dynamics Model:} \quad w_{n+1} = \mathsf{G}(u_n), \quad n \in \mathbb{Z}^+$$
$$\text{Data Model:} \quad y_{n+1} = w_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+$$

- $y_{n+1} = y, \quad \eta_{n+1} \sim N(0, h^{-1}\Gamma)$.
- Evensen moved to Statoil.
- Methodology widely used in oil industry.
- Also in groundwater flow.
- Gier Nævdal 2001, 2002.
- Oliver, Reynolds, Liu (2008) [5].

EKI Algorithm

Reformulate In General State Space Notation

$$v_n = (u_n, w_n), \Psi(u, w) = \big(u, \mathrm{G}(u)\big), H(u, w) = w.$$

Continuous Time Limit [10]

Let $NT = h$. Then as $h \to 0$

$$sup_{0 \le n \le N} |u^{(k)}(nh) - u_n^{(k)}| \to 0$$

where

$$\dot{u}^{(k)} = -\sum_{\ell=1}^{K} d_{k,\ell}(u) u^{(\ell)}$$

where

$$d_{k,\ell}(u) = \langle G(u^{(k)}) - y, G(u^{(\ell)}) - \bar{G} \rangle.$$

Properties of EKI

- Linear Case $G(\cdot) = A\cdot$.
- Least Squares Functional

$$\Phi(u) = \frac{1}{2}\|y - Au\|_\Gamma^2.$$

- Gradient Structure

$$\frac{du^{(k)}}{dt} = -C\nabla\Phi(u^{(k)}),$$
$$C = \frac{1}{K}\sum_{\ell=1}^{K}(u^{(\ell)} - \overline{u}) \otimes (u^{(\ell)} - \overline{u}). \quad (1)$$

Theorem (Gradient Structure) [10]

Algorithm minimizes $\Phi(\cdot; y)$ over a finite dimensional subspace defined by the linear span of the initial ensemble.

Inversion Applications

w/Chada, Iglesias, Roininen

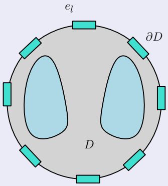
w/Kovachki

Electrical Impedance Tomography (EIT) 1

Forward Problem

Given $(\kappa, I) \in L^{\infty}(D; \mathbb{R}^{+}) \times \mathbb{R}^{m}$ find $(\nu, V) \in H^{1}(D) \times \mathbb{R}^{m}$:

$$
\begin{aligned}
-\nabla \cdot (\kappa \nabla \nu) &= 0 \quad \in \quad D, \\
\nu + z_{\ell} \kappa \nabla \nu \cdot n &= V_{\ell} \quad \in \quad e_{\ell}, \quad \ell = 1, \ldots, m, \\
\nabla \nu \cdot n &= 0 \quad \in \quad \partial D \setminus \cup_{\ell=1}^{m} e_{\ell}, \\
\int \kappa \nabla \nu \cdot n \, ds &= I_{\ell} \quad \in \quad e_{\ell}, \quad \ell = 1, \ldots, m.
\end{aligned}
$$



Ohm's Law: $V = R(\kappa) \times I$.

Inverse Problem

Set $\kappa = \exp(u)$. Given a set of K noisy measurements of voltage $V(k)$ from currents $I(k)$, and $\mathrm{G}_k(u) = R\big(\exp(u)\big) \times I(k)$, find u from y where:

$$
y(k) = \mathrm{G}_k(u) + \eta, \quad \eta \sim \mathrm{N}(0, \gamma^2), \quad k = 1, \ldots, K.
$$

EIT 2

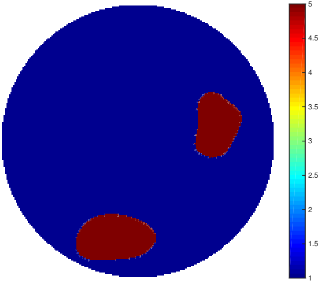


Figure: True Conductivity.

Parameterization

- Continuous level set function.
- Lengthscale of level set function.
- Smoothness of level set function.

EIT 3

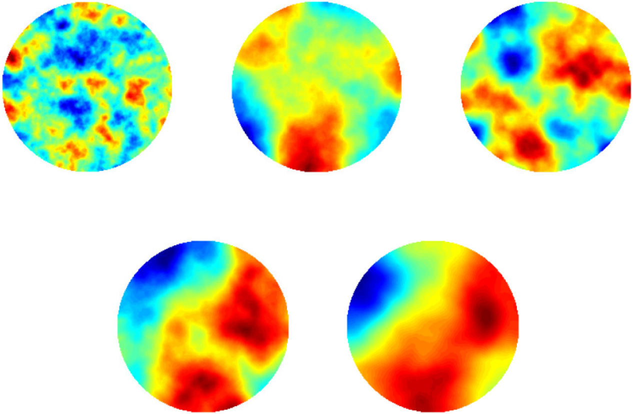


Figure: Five succesive iterations: level set function u.

EIT 4

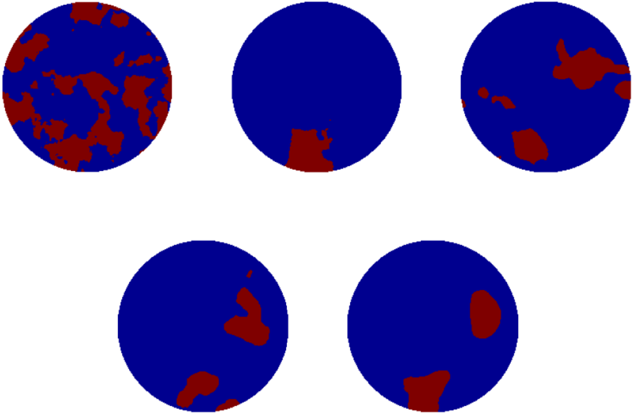


Figure: Five succesive iterations: thresholded level set function $v = S(u)$.

Whittle-Matérn Initial Ensembles

- Create initial ensemble of functions via Gaussian random fields.
- Common choice: Whittle-Matérn family

$$c_{\nu,\tau}(x,x') := \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\tau|x-x'|\right)^{\nu} K_{\nu}\left(\tau|x-x'|\right).$$

  - Smoothness parameter: $\nu \in \mathbb{R}^{+}$.
  - Inverse length-scale parameter: $\tau \in \mathbb{R}^{+}$.

- Corresponding covariance operator

$$\mathcal{C}_{\nu,\tau} \propto \tau^{2\nu}(\tau^{2} I - \triangle)^{-\nu-\frac{d}{2}}.$$

- Hierarchical: invert for as ν, τ as well as field itself.
- F. Lindgren, H. Rue and J. Lindström, (JRSS-B **73**(2011))

Centred vs Non-centred

- Define $\theta = (\alpha, \tau)$.
- Generate samples u by solving the SPDE

$$(\tau^2 I - \triangle)^{\frac{\nu + \frac{1}{2}d}{2}} u = \tau^{\nu} \xi,$$

where $\xi \sim \mathrm{N}(0, I)$ is white noise. $u = T(\xi, \theta)$.
See F. Lindgren, H. Rue and J. Lindström, (JRSS-B **73**(2011))

- Hierarchical: invert for parameters θ as well as field u.
- Centred approach:
  - view (u, θ) as unknowns;
  - initial ensemble samples $\mathbb{P}(u|\theta)\mathbb{P}(\theta)$;
  - $y = \mathcal{G}(u) + \eta$.
- Non-centred approach:
  - view (ξ, θ) as unknowns;
  - initial ensemble samples $\mathbb{P}(\xi)\mathbb{P}(\theta)$;
  - $y = \mathcal{G}\big(T(\xi, \theta)\big) + \eta$.

See O. Papaspiliopoulos, G. O. Roberts, and M. Sköld, (Statistical Science)**22**(2007))

Supervised Learning

Inverse Problem

- **Data**: $\{(x_j, y_j)\}_{j=1}^N$ with $x_j \in \mathcal{X}$, $y_j \in \mathcal{Y}$ and $\mathcal{X}, \mathcal{Y}$ Hilbert spaces.
- **Find**: $\mathcal{G}(u|\cdot) : \mathcal{X} \to \mathcal{Y}$ for parameter $u \in \mathcal{U}$ consistent with the data.
- **Concatenate** x, y and $\mathcal{G}(u|\cdot)$:

$$\mathrm{y} = \mathrm{G}(u|\mathrm{x}) + \eta$$

where $\mathrm{G}(\cdot|\mathrm{x}) : \mathcal{U} \to \mathcal{Y}^N$ and η is model or data error.

Supervised Learning

Key Issues

- **Approximation:** design of $\mathcal{G}(\cdot|\mathrm{x}_j)$;
- **Optimization**: choosing u to fit data $\{(x_j, y_j)\}_{j=1}^N$;
- **Stability**: ability of $\mathcal{G}(\cdot|\mathrm{x}^\star)$ to predict well for out of sample $x^\star$.

Architecture, training and generalization.

MNIST Dataset

LeCun and Cortes 1999.

MNIST Supervised

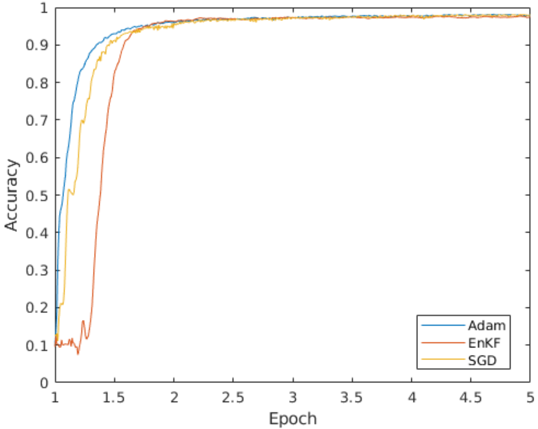


Figure: Test Accuracy of **Net 1** on MNIST (batched).

J	Loss	Momentum	Randomize y	Randomize u
5000	Cross Entropy	✓	✓	χ

Conclusions and References

Conclusions

- Kalman's 1960 paper revoltionized applied mathematics.
- Evensen's 1994 paper introduced a step change in applicability.
- Both state estimation and inverse problems maybe solved.
- Aerospace guidance $\cdots$
- Oceanography, weather forecasting, climate ...
- Geophysical and medical imaging.
- Machine Learning ?

References

[1] R. Kalman and R. Bucy.
A new approach to linear filtering and prediction problems
Journal of Basic Engineering, **82**(1961), 95–108.

[2] G. Evensen.
Data Assimilation: The Ensemble Kalman Filter.
Springer, 2006.

[3] A.H. Jazwinski.
Stochastic Processes and Filtering Theory.
Academic Press, 1970.

[4] M. Ghil, S.E. Cohn, J. Tavantzis, K. Bube, and E. Isaacson.
Application of estimation theory to numerical weather prediction.
Dynamic Meteorology: Data Assimilation Methods, L. Bengtsson, M. Ghil and E. Källén, Eds.
Springer, 139–224, 1981.

[5] D.S Oliver, A.C. Reynolds and N Liu
Inverse theory for petroleum reservoir characterization and history matching
CUP 2008

References

[6] A. M. Stuart
Inverse Problems: a Bayesian perspective
Acta Numerica, 19(May):451-559, 2010.

[7] M. Iglesias, Y. Lu and A.M. Stuart
A Bayesian level set method for geometric inverse problems.
Interfaces and Free Boundaries, **18**(2016), 181-217.

[8] N. K. Chada, M. Iglesias, L. Roininen, A. M. Stuart
Geometric and Hierarchical Ensemble Kalman Inversion,
arXiv:1602.1709.01781

[9] M. A. Iglesias, K. Law, A. M. Stuart
Ensemble Kalman method for inverse problems,
Inverse Problems, 29 (4), 045001, 2013.

[10] C. Schillings, A. M. Stuart
Analysis of the ensemble Kalman filter for inverse problems,
arXiv:1602.02020, SIAM Num. Analysis **55**(2017).

Supervised Learning

$$\Phi(u) = \frac{1}{2}\|\mathrm{y} - \mathrm{G}(u|\mathrm{x})\|^2_{\mathcal{Y}^N} \quad \text{or} \quad -\sum_{j=1}^{N}\langle y_j, \log \mathcal{G}(u|, x_j)\rangle_{\mathcal{Y}}$$

Algorithms

$$\textbf{SGD} : \dot{u} = -\nabla_u \Phi(u); \quad u(0) = u_0, \quad u_* = u(T)$$

$$\textbf{EnKF} : \dot{u}^{(k)} = -\sum_{\ell=1}^{K} d_{k,\ell}(u)u^{(\ell)}; \quad u^{(k)}(0) = u_0^{(k)}, \qquad u_* = \frac{1}{K}\sum_{\ell=1}^{K} u^{(\ell)}(T)$$

Tricks

- Mini-batching.
- Momentum: $\ddot{u}^{(k)} + \frac{3}{t}\dot{u}^{(k)} = -\sum_{\ell=1}^{K} d_{k,\ell}(u)u^{(\ell)}$.
- Randomize y and u.

Convolutional Models

Net 1 ~ 14k	**Net 2** ~ 30k
Conv12x3x3 MaxPool2x2	Conv12x3x3 Conv12x3x3 MaxPool2x2
Conv24x3x3 MaxPool2x2	Conv24x3x3 Conv24x3x3 MaxPool2x2
Conv32x3x3 MaxPool2x2	Conv32x3x3 Conv32x3x3
FC-100	FC-100
FC-10	FC-10

- ReLU: after each block: $\max(0, x)$;
- Layer normalization: Ba, Kiros and Hinton 2016. (NIPS)

Groundwater Flow 1

Forward Problem

Given $\kappa \in X := L^\infty(D;\mathbb{R}^+)$ find $p \in H_0^1(D;\mathbb{R})$ such that:

$$\begin{aligned} -\nabla \cdot (\kappa \nabla p) &= f, \quad x \in D, \\ p &= 0, \quad x \in \partial D. \end{aligned}$$

Inverse Problem

Set $\kappa = \exp(u)$. Given K linear functionals of the pressure $\mathsf{G}_k(u) = \mathfrak{o}_k(p)$, $\mathfrak{o}_k \in H^{-1}(D;\mathbb{R})$, find u from noisy measurements y where:

$$y = \mathsf{G}(u) + \eta, \quad \eta \sim \mathsf{N}(0, \Gamma).$$

Grdounwater Flow 2

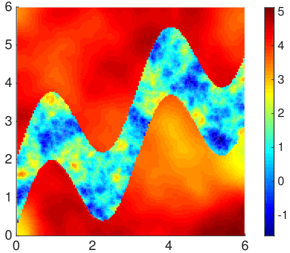


Figure: True log-permeability.

Parameterization

- Five scalars describe channel geometry.
- One random field describes interior of channel.
- One random field describes exterior of channel.
- Lengthscale and smoothness parameters of both random fields.

Groundwater Flow 3

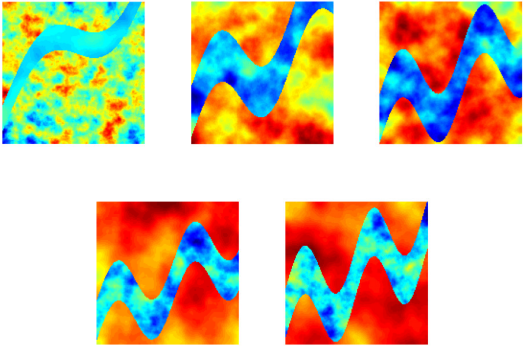


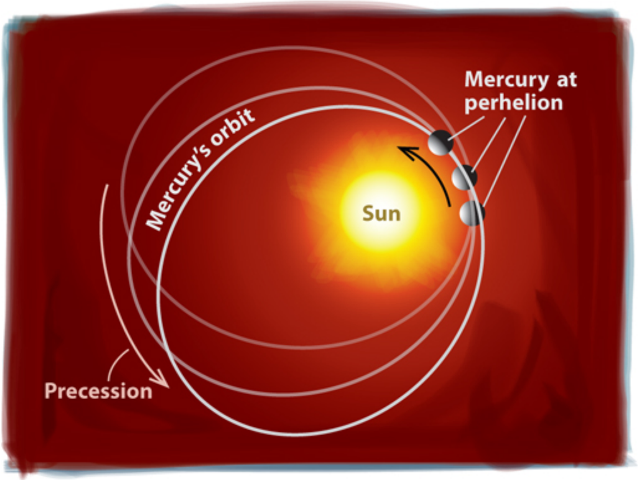
Figure: Five succesive iterations; ensemble mean

Historical Context

Newton

Data: conceptual/sparse $\Longrightarrow$ Models: conservation laws.

Einstein



Data: conceptual/sparse $\Longrightarrow$ Models: special and general relativity.