The Legacy of Rudolph Kalman
Blending Data and Mathematical Models

Andrew Stuart

Computing and Mathematical Sciences
California Institute of Technology

AFOSR, DARPA, EPSRC, NSF, ONR
Allen Philanthropies, Mission Control for Earth, Schmidt Futures

⋆ ⋆ ⋆

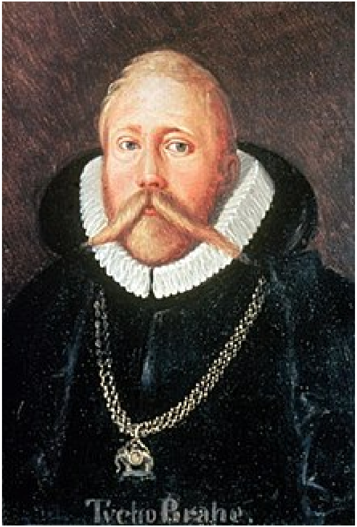
Jacques Morgenstern Colloquium, INRIA

March 28th 2019

Overview

Historical Context: Celestial Mechanics

Brahe

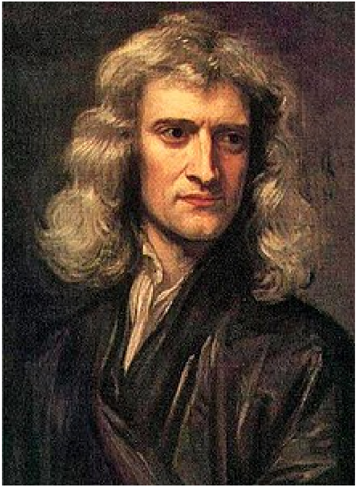


Purely observational data – initially by eye.
"Big data" c. 1600s.

Kepler

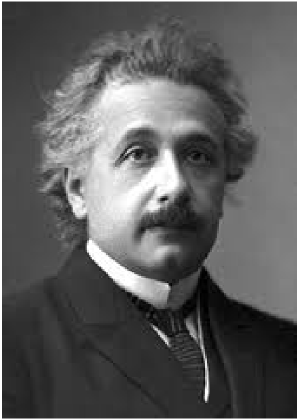
Mathematical formulae which interpolated Brahe's data.
Kepler's Law – a "data-driven model."

Newton



Kepler's Law emerge from Newtonian mechanics.
Led to theory of conservation laws.

Einstein



Discrepancy between data and prections of Newtonian mechanics (Mercury perehilion). Resolved by special, and then general, relativity.

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

Dynamics Model: $v_{n+1} = Mv_n + \xi_n, \quad n \in \mathbb{Z}^+$

Data Model: $y_{n+1} = Hv_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+$

Probabilistic Structure: $v_0 \sim N(m_0, C_0), \quad \xi_n \sim N(0, \Sigma), \quad \eta_n \sim N(0, \Gamma)$

Probabilistic Structure: $v_0 \perp \{\xi_n\} \perp \{\eta_n\}$ independent

- J. Basic Engineering **82**(1960); see [1].
- $29,946$ Google Scholar citations; 27/3/19.
- 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

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

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

3DVAR Filter

State Space Model

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

- 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

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

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

Ensemble Kalman Filter

State Space Model

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

- 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.
- $\{v_n^{(j)}\}_{j=1}^J \mapsto \{v_{n+1}^{(j)}\}_{j=1}^J$.

Ensemble Kalman Filter

Sequential Optimization Perspective

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

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

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

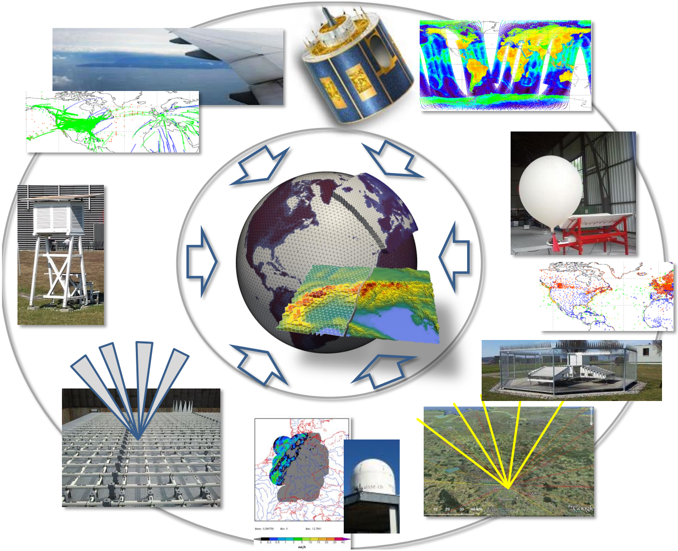
- $\widehat{C}_{n+1}$ is empirical covariance of the $\{\widehat{v}_{n+1}^{(k)}\}$.
- Updating $\widehat{C}_n$ requires only $\mathcal{O}(Jd)$ storage.

Weather Forecasting

w/Law [7]

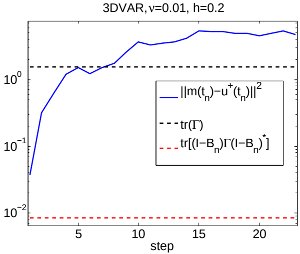
Weather Forecasting: Data

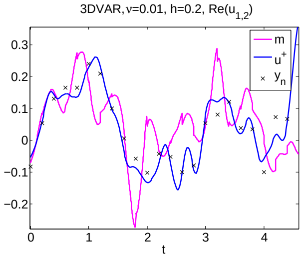
courtesy Roland Potthast(DWD)



Data Fails to Overcome Butterfly Effect

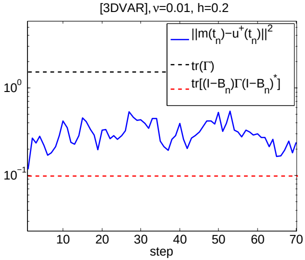
KJH Law and AM Stuart, Monthly Weather Review, 2014.

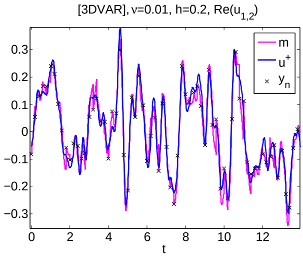




Theory Backed use of Data Overcomes Butterfly Effect

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

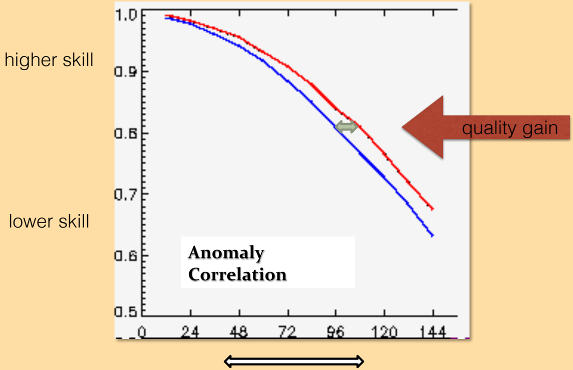




Impact of EnKF over 3DVAR

courtesy Roland Potthast(DWD)

Ensemble Kalman Filter (red) versus 3DVAR (blue)



forecast lead time in hours

Ensemble Kalman Inversion

w/Iglesias and Law [9]

w/Schillings [10]

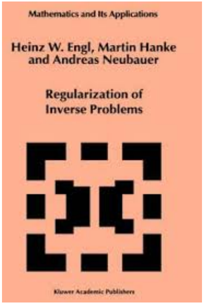
Inverse Problem

Problem Statement

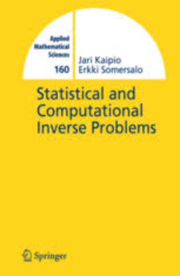
Find u from y where $\mathrm{G}:\mathcal{U}\mapsto\mathcal{Y}$, where $\mathcal{U},\mathcal{Y}$ are Hilbert spaces, η is noise and

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

Optimization $\Phi_R(u)=\frac{1}{2}\|y-G(u)\|^2+R(u);$ *Probability* $e^{-\Phi_R(u)}$.



A.N. Tikhonov (1963)



J. Franklin (1970)

Inverse Problem

Dynamical Formulation

$$\text{Dynamics Model:} \quad u_{n+1} = u_n, \quad n \in \mathbb{Z}^+$$

$$\text{Dynamics Model:} \quad w_{n+1} = \mathrm{G}(u_n), \quad n \in \mathbb{Z}^+$$

$$\text{Data Model:} \quad y_{n+1} = w_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+$$

- $y_{n+1} = y, \quad \eta_{n+1} \sim N(0, h^{-1}\Gamma).$
- 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

Dynamical Formulation

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

State Space Estimation Formulation

$$\text{Reformulate:} \quad v = (u, w), \quad \Psi(v) = (u, \mathsf{G}(u)), \quad H = (0, I)$$
$$\text{Dynamics Model:} \quad v_{n+1} = \Psi(v_n), \quad n \in \mathbb{Z}^+$$
$$\text{Data Model:} \quad y_{n+1} = Hv_{n+1} + \eta_{n+1}, \quad n \in \mathbb{Z}^+$$

Employ Ensemble Kalman Filter with $y_{n+1} \equiv y$.

Theory For EKI

w/Garbuno-Inigo, Hoffmann, Li [8]

The Scheme

EKI: Discrete Time

$$u_{n+1}^{(j)} = u_n^{(j)} + C_n^{uy}(C_n^{yy} + \Gamma)^{-1}\left(y_n^{(j)} - G(u_n^{(j)})\right), \quad u^{(j)}(0) = u_0^{(j)}.$$

$\bar{\cdot}$ denotes ensemble average. $u_n^{(j)} \approx u^{(j)}(nh), \Gamma \mapsto h^{-1}\Gamma$ and $h \to 0$:

EKI: Continuous Time Limit

$$\dot{u}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\left\langle G(u^{(k)}) - \bar{G}, G(u^{(j)}) - y\right\rangle_{\Gamma}\left(u^{(k)} - \bar{u}\right), \quad u^{(j)}(0) = u_0^{(j)}$$

Approximate Gradient Structure

Make linear approximation:

$$\boxed{G(u^{(k)}) = G(u^{(j)}) + dG(u^{(j)})(u^{(k)} - u^{(j)}).}$$

EKI Invoking Linear Approximation

$$\dot{u}^{(j)} = -\frac{1}{J}\sum_{k=1}^{J}\left\langle dG(u^{(j)})(u^{(k)} - \bar{u}), G(u^{(j)}) - y\right\rangle_{\Gamma}\left(u^{(k)} - \bar{u}\right), \quad u^{(j)}(0) = u_0^{(j)}$$

- Least Squares Functional

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

- Gradient Structure

$$\dot{u}^{(j)} = -C\nabla\Phi(u^{(j)}),$$

$$C = \frac{1}{K}\sum_{k=1}^{K}(u^{(k)} - \overline{u}) \otimes (u^{(k)} - \overline{u}).$$

Exact Gradient Structure

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

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

- Gradient Structure

$$\dot{u}^{(j)} = -C\nabla\Phi(u^{(j)}),$$

$$C = \frac{1}{K}\sum_{\ell=1}^{K}(u^{(\ell)} - \overline{u}) \otimes (u^{(\ell)} - \overline{u}).$$

Theorem (Gradient Structure) [10]

Flow minimizes $\Phi(\cdot; y)$ over a finite dimensional subspace defined by the linear span of the initial ensemble $\{u_0^{(j)}\}_{j=1}^{J}$.

Noisy EKI

Add Noise

- Gradient flow plus noise:

$$\dot{u}^{(j)} = -C\nabla\Phi(u^{(j)}) + \sqrt{2C}\dot{W}^{(j)},$$

- Approximate gradient flow plus noise:

$$\dot{u}^{(j)} = -C\nabla_{\mathrm{approx}}\Phi(u^{(j)}) + \sqrt{2C}\dot{W}^{(j)},$$

- Empirical Covariance:

$$C = \frac{1}{K}\sum_{\ell=1}^{K}\left(u^{(\ell)} - \overline{u}\right) \otimes \left(u^{(\ell)} - \overline{u}\right).$$

Mean Field Limit

$J \to \infty$.

- Gradient flow plus noise:

$$\dot{u}^{(j)} = -\mathcal{C}\nabla\Phi(u^{(j)}) + \sqrt{2\mathcal{C}}\dot{W}^{(j)},$$

- Empirical Covariance:

$$\mathcal{C} = \int (u-\overline{u})\otimes(u)-\overline{u})\rho(u)du, \quad \overline{u} = \int u\rho(u)du.$$

- Nonlinear Fokker-Planck (NLFP) Equation

$$\partial_t\rho = \nabla\cdot\big(\rho\,\mathcal{C}(\rho)\nabla\Phi_R(u)\big) + \mathcal{C}(\rho) : D^2\rho\,.$$

Theorem (Kalman-Wasserstein Gradient Structure) [8]

The Gibbs measure $\rho_\infty \propto \exp(-\Phi(u))$ is an equilibrium solution of the NLFP equation. If $\mathcal{C}$ is bounded below uniformly in time as an operator then $\|\rho-\rho_\infty\|_{L^1} \to 0$ exponentially fast as $t\to\infty$. In particular this condition is satisfied for the linear imverse problem.

Training Neural Networks

w/Kovachki [11]

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

Architecture, training and generalization.

MNIST Dataset

LeCun and Cortes 1999.

MNIST Supervised

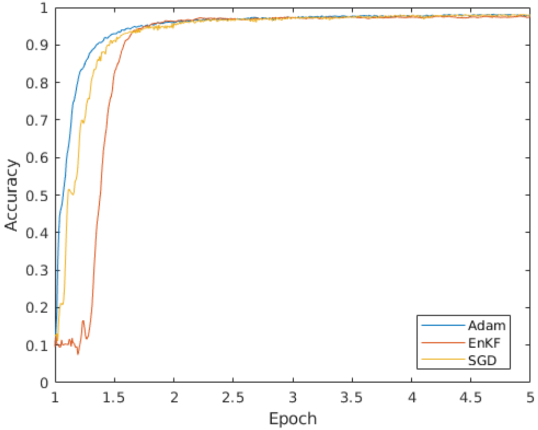


Figure: Test Accuracy of **Net 1** on MNIST (batched).

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.
- Mean Field Limit: novel Kalman-Wasserstien gradient flow.

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**(2010).

[7] K.J.H. Law and A.M. Stuart
Evaluating Data Assimilation Algorithms.
Monthly Weather Review, **18**(2014), 181-217.

[8] A. Garbuno-Inigo, W. Li, F. Hoffmann and A.M. Stuart
Gradient Structure For The Ensemble Kalman Flow With Noise,
arXiv:1903.08866.

[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).

[11] N.B. Kovachki, A.M. Stuart
Ensemble Kalman Inversion: A Derivative-Free Technique For Machine Learning Tasks,
arXiv:1808.03620.