LA-UR-16-20916
Approved for public release; distribution is unlimited.

Title: Denoising nonlinear dynamical systems

Author(s): Hengartner, Nicolas W.
Godinez Vazquez, Humberto C.

Intended for: Statistical Inverse Problems, 2016-02-08/2016-02-12 (Marseille, France)

Issued: 2016-02-12

Denoising nonlinear dynamical systems

Nicolas Hengartner[(1)] and Humberto Godinez[(2)],
[(1)] Theoretical Biology and [(2)] Applied Mathematics
Los Alamos National Laboratory

CIRM

February 9, 2016

Model for the observations

- Consider a discrete-time nonlinear dynamical systems generated by the recursion on the state space

$$x_{n+1} = f(x_n), \quad n = 1, 2, 3, \ldots$$

  The function f is known.

Model for the observations

- Consider a discrete-time nonlinear dynamical systems generated by the recursion on the state space

$$x_{n+1} = f(x_n), \quad n = 1, 2, 3, \ldots$$

  The function f is known.

- Observe

$$y_n = g(x_n; \theta) + \varepsilon_n.$$

  The disturbances ε are independent mean zero and finite variance. The parameter θ is unknown.

  Nonparametric estimation of g?

Problem (Extension of Kalman filtering, data assimilation)

Given the observed sequence

$$y_1, y_2, \ldots, y_n$$

without $x_1, \ldots, x_n$, but knowing f,

1. Denoise the sequence to estimate the series to estimate the sequence $g(x_1), g(x_2), \ldots, g(x_n)$
2. Predict future observations, e.g., $g(x_{n+1}), g(x_{n+2}), \ldots$
3. Predict past observations, e.g., $g(x_0), g(x_{-1}), \ldots.$

Easier problem than estimating the dynamical systems series $x_1, x_2, \ldots, x_n$, but still hard.

Simple example: Discrete epidemic evolution model (SIR)

Susceptible-Infected-Recovered epidemic model (S_k, I_k, R_k)

$$
\begin{aligned}
S_{k+1} &= S_k - \beta S_k I_k \\
I_{k+1} &= I_k + \beta S_k I_k - \gamma I_k \\
R_{k+1} &= R_k + \gamma I_k
\end{aligned}
$$

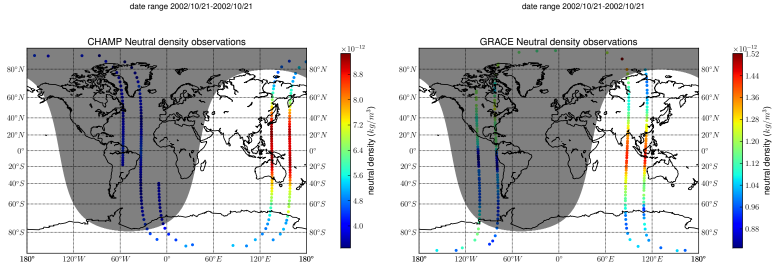
State-space $x_k = (S_k, S_{k-1}, I_k, R_k)$

Observe a "fraction of new cases".

$$Y_k = \text{Poisson}(\alpha(S_{k-1} - S_k))$$

Want to make predictions for the number of new cases.

A real data example



Upper atmospheric densities observational data collected by two satellites, shaded area indicates night side at time of observation collection

Complex global weather evolution model, data are local probing.

Challenging problem...

Because we know f, estimating the unobserved series $x_1, \ldots, x_n$ is equivalent to estimating x_0.

This is an inverse problem.

When g is known, we can estimate the initial state x_0 of the dynamical system by non-linear least squares

$$\widehat{x}_0 = \arg\min_{x_0} \sum_{j=1}^{n} \left(y_j - g(f^{*(j)}(x_0)) \right)^2$$

where $f^{*(j)}(x_0) = f(f^{*(j-1)}(x_0))$ is the j^{th} composition of f.

Potentially ill-posed numerical optimization problem.

Having a good representation is the key to success...

- Our aim is to identify the well posed questions in this ill posed problem.

  A really clever trick! Koopman (PNAS 1931)

Having a good representation is the key to success...

- Our aim is to identify the well posed questions in this ill posed problem.

  A really clever trick! Koopman (PNAS 1931)

- Let $\mathbb{H}$ be a Hilbert space of functions defined on $\mathbb{X}$. For a discrete-time dynamical system $x_{n+1} = f(x_n)$, consider for $x \in \mathbb{X}$ fixed the linear operator

$$
\begin{aligned}
L_x \quad : \quad & \mathbb{H} \longrightarrow \mathbb{R} \\
& h \longrightarrow h(f(x)).
\end{aligned}
$$

Having a good representation is the key to success...

- Our aim is to identify the well posed questions in this ill posed problem.

  A really clever trick! Koopman (PNAS 1931)

- Let $\mathbb{H}$ be a Hilbert space of functions defined on $\mathbb{X}$. For a discrete-time dynamical system $x_{n+1} = f(x_n)$, consider for $x \in \mathbb{X}$ fixed the linear operator

$$\begin{aligned} L_x \quad : \quad & \mathbb{H} \longrightarrow \mathbb{R} \\ & h \longrightarrow h(f(x)). \end{aligned}$$

- For a fixed $x \in \mathbb{X}$, if L_x is bounded, is has a discrete spectrum $\lambda_1(x) \geq \lambda_2(x) \geq \dots$ and right-hand side eigenfunctions

$$L_x \psi_j = \lambda_j(x) \psi_j.$$

Implications

- If ψ_j is a (right-hand) eigenfunction of L_x, then

$$L_x \psi_j \triangleq \psi_j(f(x)) = \lambda_j(x)\psi_j(x).$$

Suppose

$$g(x) = \sum_{j=1}^{p} \alpha_j \psi_j(x).$$

Implications

- If ψ_j is a (right-hand) eigenfunction of L_x, then

$$L_x \psi_j \stackrel{\triangle}{=} \psi_j(f(x)) = \lambda_j(x)\psi_j(x).$$

  Suppose

$$g(x) = \sum_{j=1}^{p} \alpha_j \psi_j(x).$$

- If $\lambda_j(x) \equiv \lambda_j$, then

$$\begin{aligned} g(x_k) &= g(f(x_{k-1})) = \sum_{j=1}^{p} \alpha_j \psi_j(f(x_{k-1})) \\ &= \sum_{j=1}^{p} \alpha_j \lambda_j \psi_j(f(x_{k-1})) = \sum_{j=1}^{p} \alpha_j \psi(x_0) \cdot \lambda_j^k \\ &= \sum_{j=1}^{p} c_j \lambda_j^k. \end{aligned}$$

Estimation of $g(x_0)$ (data assimilation):

If $g \in \text{span}(\psi_1, \dots, \psi_p)$, we estimate $g(x_0)$ by first minimizing

$$(\widehat{c}_1, \dots, \widehat{c}_p) = \arg \min_{c_1,\dots,c_p} \sum_{k=1}^{n} \left(y_k - \sum_{j=1}^{p} c_j \lambda_j^k \right)^2$$

and set

$$\widehat{g}(x_0) = \sum_{j=1}^{p} \widehat{c}_j .$$

Remarks

1. Only knowledge of the spectrum $\lambda_1, \lambda_2, \ldots$ and the assumption that $g \in \text{span}(\psi_1, \psi_2, \ldots, \psi_p)$ is required.
2. Numerical optimization is easy.
3. If $|f(x)| \leq B < \infty$, then $\lambda_1 \leq 1$.
4. Information from later observations exponentially small.

Denoising and prediction:

- Fix k_0. If $g(x) = \sum_{j=1}^{p} \alpha_j \psi_j(x)$, then

$$g(x_k) = \sum_{j=1}^{p} \lambda_j^{k-k_0} \psi_j(x_{k_0}).$$

Denoising and prediction:

- Fix k_0. If $g(x) = \sum_{j=1}^{p} \alpha_j \psi_j(x)$, then

$$g(x_k) = \sum_{j=1}^{p} \lambda_j^{k-k_0} \psi_j(x_{k_0}).$$

- Denoising reduces to solving (via least squares)

$$\min_{c_1,\ldots,c_p} \sum_{k=1}^{n} \left(y_k - \sum_{j=1}^{p} c_j \lambda_j^{k-k_0} \right)^2$$

and set $\widehat{g}(x_k) = \sum_{j=1}^{p} \widehat{c}_j$.

How can we calculate the Koopman basis?
And can we show that indeed, $\lambda_j(x) \equiv \lambda_j$?

A short primer of reproducing kernel Hilbert space (rkhs).

1. A Hilbert space $\mathbb{H}$, endowed with an inner product $< \cdot | \cdot >$ is a rkhs if and only if the valuation function

$$H_x : h \longrightarrow h(x)$$

is a bounded linear functional.

A short primer of reproducing kernel Hilbert space (rkhs).

1. A Hilbert space $\mathbb{H}$, endowed with an inner product $< \cdot | \cdot >$ is a rkhs if and only if the valuation function

$$H_x : h \longrightarrow h(x)$$

is a bounded linear functional.

2. The Riesz representation theorem says that there exists $\ell_x \in \mathbb{H}$, such that for all $h \in \mathbb{H}$, $H_x h =< \ell_x | h >$.

A short primer of reproducing kernel Hilbert space (rkhs).

1. A Hilbert space $\mathbb{H}$, endowed with an inner product $< \cdot | \cdot >$ is a rkhs if and only if the valuation function

$$H_x : h \longrightarrow h(x)$$

   is a bounded linear functional.

2. The Riesz representation theorem says that there exists $\ell_x \in \mathbb{H}$, such that for all $h \in \mathbb{H}$, $H_x h =< \ell_x | h >$.
3. The kernel $K(x, y) =< \ell_x | \ell_y >$ is non-negative definite, symmetric, and satisfies

$$h(x) =< K(x, \cdot) | h > .$$

A short primer of reproducing kernel Hilbert space (rkhs).

1. A Hilbert space $\mathbb{H}$, endowed with an inner product $< \cdot | \cdot >$ is a rkhs if and only if the valuation function
$$H_x : h \longrightarrow h(x)$$
is a bounded linear functional.
2. The Riesz representation theorem says that there exists $\ell_x \in \mathbb{H}$, such that for all $h \in \mathbb{H}$, $H_x h =< \ell_x | h >$.
3. The kernel $K(x, y) =< \ell_x | \ell_y >$ is non-negative definite, symmetric, and satisfies
$$h(x) =< K(x, \cdot) | h > .$$
4. The functional $< K(x, \cdot) | f >$ is the projection of $f \in \mathcal{L}_2$ onto the space $\mathbb{H}$.

A short primer of reproducing kernel Hilbert space (rkhs).

1. A Hilbert space $\mathbb{H}$, endowed with an inner product $< \cdot | \cdot >$ is a rkhs if and only if the valuation function
$$H_x : h \longrightarrow h(x)$$
is a bounded linear functional.
2. The Riesz representation theorem says that there exists $\ell_x \in \mathbb{H}$, such that for all $h \in \mathbb{H}$, $H_x h =< \ell_x | h >$.
3. The kernel $K(x, y) =< \ell_x | \ell_y >$ is non-negative definite, symmetric, and satisfies
$$h(x) =< K(x, \cdot) | h > .$$
4. The functional $< K(x, \cdot) | f >$ is the projection of $f \in \mathcal{L}_2$ onto the space $\mathbb{H}$.
5. Let $\phi_1, \phi_2, \ldots, \phi_p$ be any orthonormal basis for a finite dimensional Hilbert space $\mathbb{H}$, then the reproducing kernel is
$$k(x, y) = \sum_{j=1}^{p} \phi_j(x) \phi_j(y).$$

The Koopman operator is the valuation functional of h evaluated in $f(x)$

- **Proposition 1.** If $\mathbb{H}$ is a reproducing kernel Hilbert space, then the Koopman operator over that space is bounded. It follows that the Koopman operator on that space can be written as

$$L_x h =< K(f(x), \cdot)|h > .$$

The Koopman operator is the valuation functional of h evaluated in $f(x)$

- **Proposition 1.** If $\mathbb{H}$ is a reproducing kernel Hilbert space, then the Koopman operator over that space is bounded. It follows that the Koopman operator on that space can be written as

$$L_x h =< K(f(x), \cdot)|h > .$$

- **Proposition 2.** If $\mathbb{H}$ is finite dimensional, then the Koopman operator can be written as

$$\begin{aligned} L_x h &= \left\langle \sum_{j=1}^{p} \phi_j(f(x))\phi_j(y) \middle| h(y) \right\rangle \\ &= \sum_{j=1}^{p} \phi_j(f(x)) < \phi_j(y)|h(y) > . \end{aligned}$$

Towards calculating the Koopman basis and spectrum

Write, for each j,

$$\phi_j(f(x)) = \sum_{\ell=1}^{p} \beta_{j\ell}\phi_\ell(x) + e_j(x).$$

The terms $e_j(x)$ is nonzero when the Hilbert space $\mathbb{H}$ is not closed under the dynamics of the system.
Write

$$\mathbf{B} = (\beta_{j,\ell})_{j,\ell=1}^{p} \quad \phi(x) = (\phi_1(x), \ldots, \phi_p(x))^t$$

Then

$$K(x, y) = \phi(y)\mathbf{B}\phi(x) + \mathbf{e}^t(x)\phi(y),$$

which implies that for $h(y) = a^t\phi(y)$

$$L_x h = \phi(x)\mathbf{B}^t a + a^t\mathbf{e}(x)$$

Controlling the approximation error

The construction holds for any orthonormal basis $\{\phi_1, \ldots, \phi_p\}$.

Idea: Control the error by selecting a convenient basis.
Given $g_1 \in \mathbb{H}$, define

$$g_k(x) = g_{k-1}(f(x)).$$

Use Gram-Schmidt to build an orthonormal basis $\{\phi_1, \ldots, \phi_p\}$ such that

$$\text{span}(g_1, \ldots, g_k) = \text{span}(\phi_1, \ldots, \phi_k) \quad k = 1, 2, \ldots, p.$$

Controlling the error

Proposition 3 Let $v_k(x) = g_k(x)/\|g_k(x)\|$.

$$e_k(x) = v_k(f(x)) - \text{proj}_{\phi_1,\ldots,\phi_p}(v_k(f(x))).$$

For $k < p$, $e_k(x) = 0$, and for $k = p$

$$|e_p(x)| \longrightarrow 0$$

Expect $e_p \sim \gamma^p$ *for some* $0 < \gamma < 1$.

Justification: power method of linear algebra. If

$$g_1(x) = \sum_{j=1}^{\infty} \alpha_j \psi_j(x)$$

then

$$g_p(x) = \sum_{j=1}^{\infty} \alpha_j \lambda_j^{p-1} \psi_j(x).$$

Approximate Koopman spectrum and basis

If we can neglect the approximation term $a^t\mathbf{e}(x)$, then Koopman basis and eigenvalues are derived from the right-hand eigenvectors of the matrix $\mathbf{B}^t$.

Proposition 4 Let

$$\mathbf{B}^t\xi_j = \lambda_j\xi_j.$$

Then the (approximate) Koopman spectrum is $\lambda_1, \dots, \lambda_p$ and the (approximate) Koopman basis are

$$\psi_j = \xi_j^t\phi(x).$$

Remark: The (approximate) Koopman eigenvalues do not depend on x.

Remark

Matrix **B** is triangular,obviously not symmetric.
Eigenvalues are the roots of the characteristic polynomial

$$det(B - \lambda I).$$

Eigenvectors solve

$$B^t \phi_j = \lambda_j \phi_j.$$

Eigenvectors are linearly independent but are NOT orthogonal.

Implementation

- Recall, all we need are the eigenvalues.
- Select $x_0^{(1)}, \ldots, x_0^{(m)}$ initial conditions, $m \geq n$, possibly as uniformly as possible over the parameter space (i.e., latin hypercube design).
- Compute the generated dynamical system $x_{k+1}^{(\ell)} = f(x_k^{(\ell)})$.
- Compute the $m \times n$ matrix

$$\mathbb{X} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \ldots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \ldots & x_n^{(2)} \\ \vdots & & & \vdots \\ x_1^{(m)} & x_2^{(m)} & \ldots & x_n^{(m)} \end{pmatrix}$$

- Calculate the SVD of $\mathbb{X}$ to get approximate Koopman eigenvectors.

Conclusion

- Cavalier's principle of "finding the well posed questions in ill posed problems" remain true.
- Another example of where embedding a problem into a larger space is helpful
- Opens the door for starting to think how to solve other nonlinear problems using linear methods.

Merci et

Bon Appetit