Local Regression Estimators: A Forgotten Heritage

Gary Dilts
Hydrodynamic Methods Group
Los Alamos National Laboratory

Aamer Haque
George Mason University

John Wallin
George Mason University

LAUR 00-2552
Background

- Compressible flow method for coupled solid-fluid systems
  - Large energy transfer (shocks)
  - Large deformation
  - Strong shear
  - Void opening and closure
  - Complex material histories
- Meshes (Lagrangian or Eulerian) are problematic
- SPH appealing candidate, but numerical difficulties
- Inspired by EFG, tried to fix with Galerkin + MLS
- Much progress, but more difficulties
  - Artificial viscosity: usual tricks don’t work
  - Anti-symmetric complete quadrature schemes needed
    - Can’t seem to do without inverting global matrix
- MLS particles must be met on their own terms: mesh ideas don’t seem to pan out. What to do?
Like Balboa, we sighted a new ocean.

Like the Pacific, it was not \textit{entirely} unknown.
Our Statistical Heritage

- Smoothing kernels originally conceived by Rosenblatt (1956) as a means of generalizing the histogram
- Lucy’s SPH (1977)
  - Motivated by Monte Carlo theory
- Gingold + Monaghan’s SPH (1977)
  - SPH originally conceived as probabilistic method
  - Position of fluid points randomly distributed by mass density
  - This density is given by the statistical smoothing kernel method
  - Monte Carlo efficiency suggested SPH would be efficient
- Stone (1977) conceived local linear regression as a means to generalize kernel density estimators to achieve “consistent” nonparametric regression estimators
- Local linear regression has been extended to the local polynomial regression estimator. Vast literature. 23 years of catch-up.
Recent Books


- Background in theoretical statistics useful.
- Much work is very mathematical and rigorous.
- We can learn a lot!
From Histograms to Smoothing Kernels

\[ P(X \leq x) = F(x), \quad P(a < X < b) = \int_{a}^{b} f(x) \, dx \]

Finite Sample: \( \{x_i\}_{i=1}^{n} \)

\[ f(x) = \frac{d}{dx} F(x) = \lim_{h \to 0} \frac{F(x + h) - F(x - h)}{2h} \]

**Histogram:**

- Bin endpoints: \( \{b_i\}_{i=0}^{n} \)
- \( h = b_i - b_{i-1} \)
- \( \hat{F}_i = \frac{\# \{x_j \leq b_i\}}{n} \)
- \( \hat{f}_i = \frac{\# \{x_j \in (b_{i-1}, b_i]\}}{nh} \)

Agree: \( x = \frac{1}{2}(b_i + b_{i+1}) \)

**Kernel Estimate:**

\[ \hat{F}(x) = \frac{\# \{x_i \leq x\}}{n} \]

\[ \hat{f}(x) = \frac{\# \{x_i \in (x - h, x + h]\}}{2hn} = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \]

\[ K(u) = \begin{cases} 
\frac{1}{2}, & -1 < u \leq 1 \\
0, & \text{otherwise}
\end{cases} \]

\( K \geq 0, \quad \int K = 1, \quad \int u K(u) \, du = 0 \)
Analysis of CD Rates for Long Island

From Simonoff, chs 2 and 3

K=Square

K=Gaussian

Histogram
Parametric Regression Fits to Motorcycle Crash Head Acceleration Data

From Fan & Gijbels

\[ Y_i = a_0 + a_1 X_i + a_2 X_i^2 + \ldots + \text{error} \]
Nonparametric Regression Fits

\[ Y_i = a_0(x) + a_1(x)X_i + \text{error}, \quad X_i \in (x - h, x + h) \]

From Fan & Gijbels
Connecting to the Statistical Literature

\[ Y = u(X) + \vartheta(X)\varepsilon \]

\( X \in \mathbb{R}^d \): spatial sample point

\( f(x) \): distribution of \( X \)

\( Y \in \mathbb{R}^m \): code output

\( u \in \mathbb{R}^m \): true solution (mean)

\( \varepsilon \): random error source

\( X \) and \( \varepsilon \) independent

\( E(\varepsilon) = 0, \ Var(\varepsilon) = 1 \)

\( \vartheta(x_0) \): error scale,

conditional variance of

\( Y \) given \( x_0 \)

✦ Statistical viewpoint of the results of a numerical simulation.

✦ Model the output of the code as a fluctuation about some presumed correct value.

✦ We don’t know what the true value is, but will use statistical methods to make the best determination.

✦ Slight departure from tradition:

\[ Y_i \leftrightarrow u_i, \quad X_i \leftrightarrow x_i \]
Preliminary Notions

(Reproducing basis) \( p(x)^T = [1, x, y, x^2/2!, \ldots] \in \mathbb{R}^n \)

(Taylor monomials) \( p_i(x)^T = [1, x_i - x, y_i - y, (x_i - x)^2/2!, \ldots] \in \mathbb{R}^n \)

\[
J(x) = \begin{bmatrix}
p, & \frac{\partial}{\partial x} p, & \frac{\partial}{\partial y} p, & \frac{\partial^2}{\partial x^2} p, & \ldots
\end{bmatrix} \in \mathbb{R}^n \otimes \mathbb{R}^n
\]

\[
e.g. \quad J(x) = \begin{bmatrix}
1 & 0 & 0 \\
x & 1 & 0 \\
\frac{1}{2} x^2 & x & 1
\end{bmatrix}, \quad J(x)^{-1} = \begin{bmatrix}
1 & 0 & 0 \\
-x & 1 & 0 \\
\frac{1}{2} x^2 & -x & 1
\end{bmatrix}
\]

\[
p_i(x) = J(x)^{-1} p(x_i)
\]

No “shifting” necessary.
A Modified Local Polynomial Regression

Estimate of $u$ and derivatives: $\beta(x) \in \mathbb{R}^m \otimes \mathbb{R}^n$

Taylor expansion of $u$ from $x$ to $x_j$: $\beta(x)p_i(x)$

Weighted Taylor objective: $\mathcal{L}(x) = \sum_j \left(u_j - \beta(x)p_i(x)\right)^2 W_j(x)$

Condition for optimum $\beta$: $\frac{\partial \mathcal{L}(x)}{\partial \beta(x)} = 0$

Solution for estimates: $\beta(x) = \sum_j u_j \psi_j(x)^T$

Vector of shape functions: $\psi_i(x) = B(x)^{-1} p_i(x) W_i(x) \in \mathbb{R}^n$

$B(x) = \sum_j p_j(x)p_j(x)^T W_j(x) \in \mathbb{R}^n \otimes \mathbb{R}^n$
Properties of LPR Estimators

Columns of $\beta$ are derivatives

$e_0^T = [1, 0, 0, \ldots], \quad e_1^T = [0, 1, 0, \ldots]$  
$u \equiv \beta e_0 \equiv \hat{u}, \quad \partial_x u \equiv \beta e_1 \equiv \hat{u}^{(1)}, \quad \ldots$

Partition of Identity Matrix

$\sum_j p_j(x)\psi_j^T = I_n$

Reproducing condition

$u_j = \lambda^T p(x_j) \Rightarrow \beta = J\lambda$  
$\lambda = e_a \Rightarrow \beta = \partial^a x p$

Equivalence to MLS

$p_j(x) = J^{-1} p(x_j), \quad B = J^{-1} A \quad J^{-T}$

$\Rightarrow \psi_j^T e_0 = p(x_j)^T A^{-1} J e_0 W_j(x) \equiv \phi_j$

These properties hold if $p$ is any set of independent functions!
Advantages for Derivatives

- LPR is different from MLS in how derivatives are estimated
  - MLS: real derivative of a function estimate
  - LPR: least squares fit for all derivatives simultaneously
- MLS derivatives tend to be noisy
- LPR derivatives tend to be smoother
LPR Diffusion Method (Haque)

- Linear diffusion equation, collocation solution

\[
\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial x^2}
\]

\[T_{i}^{k+1} = T_{i}^{k} + \kappa \Delta t \left( \sum_{j} T_{j}^{k} \psi_{j}^{2} \right)\]

\[T(0, x) = Q \delta(x)\]

- Initialize numerical problem with analytic solution at t=.01 and \(\kappa = 1\).
Numerical Solutions (Haque)

- 40% randomly perturbed grid
- MLS is less stable than LPR
- LPR compares well with CFD on much finer uniform grid.
LPR and the Galerkin Approach to Conservation Laws

- Regrettably, LPR is not magic.
- Use of LPR directly in place of MLS in the Galerkin approach produces hardly a noticeable difference in shock solutions.
- Some care must be taken to do the Galerkin approximation correctly, as there is no explicit derivative in LPR.
- End result is almost identical to MLS in 1D.

\[
\begin{align*}
\int \tau \rho \dot{U} &= \int \tau \nabla \cdot F \\
\tau &= \sum_i \tau_i \psi_i^0, \quad \nabla \cdot F = \sum_j F_j \cdot \bar{\psi}_j \\
\sum_i \tau_i m_i \dot{U}_i &= \sum_{ij} \tau_i F_j \cdot \int \psi_i^0 \bar{\psi}_j \\
m_i \dot{U}_i &= \sum_j F_j \cdot \int \psi_i^0 \bar{\psi}_j \\
\phi_i &\leftrightarrow \psi_i^0, \quad \nabla \phi_i \leftrightarrow \bar{\psi}_i
\end{align*}
\]
**Asymptotic estimates for LPR**

- **MSE** is the figure of merit for an estimator:
  \[ u(x) = E(Y|X = x), \quad \sigma^2(x) = \text{Var}(Y|X = x) \]

  \[
  \text{Bias}\{\hat{u}^{(v)}(x)|X\} = E\left(\hat{u}^{(v)}(x)|X\right) - u^{(v)}(x)
  \]

  \[
  \text{Var}\{\hat{u}^{(v)}(x)|X\} = E\left(\left[\hat{u}^{(v)}(x) - E\left(\hat{u}^{(v)}(x)|X\right)\right]^2|X\right)
  \]

  \[
  \text{MSE}(x)^2 = E\left(\left[\hat{u}^{(v)}(x) - u^{(v)}(x)\right]^2|X\right)
  \]

  \[
  \text{MSE}(x)^2 = \text{Bias}\{\hat{u}^{(v)}(x_0)|X\}^2 + \text{Var}\{\hat{u}^{(v)}(x_0)|X\}
  \]

- **For LPR these have the asymptotic values**

  \[ h \to 0, \quad N h \to \infty \]

  \[
  \text{Var}\{\hat{u}^{(v)}(x)|X\} \to C_1(K,v,n) \frac{\sigma^2(x)}{f(x)Nh^{1+2v}}
  \]

  \[
  \text{Bias}\{\hat{u}^{(v)}(x)|X\} \to C_2(K,v,n)u^{(v+1)}(x)h^{n+1-v}, \quad n - v \text{ odd}
  \]

  \[
  \text{Bias}\{\hat{u}^{(v)}(x)|X\} \to C_3(K,v,n)\left[ u^{(v+2)}(x) + (n+2)u^{(v+1)}(x)\frac{f'(x)}{f(x)} \right]h^{n+2-v}, \quad n - v \text{ even}
  \]
Asymptotics and Smoothing Length

- **Bias** $\propto h^{n+1-v}$ large bias (error) at large $h$,
- **Var** $\propto 1/h^{1+2v}$ large variance (oscillation) at small $h$
- Constant trade-off between error and oscillation versus $h$.
- An optimal choice of smoothing length is possible by minimizing the MSE:

$$h_{opt}^v(x) = C_4(K, v, n) \left( \frac{\vartheta^2(x)}{\{u^{(v+1)}(x)\}^2 f(x)} \right)^{\frac{1}{2v+3}} N^{\frac{-1}{2v+3}}$$

- Contains unknown quantities. Many techniques for estimating these. Art form.
- For first derivatives, $f = 1/L$, $N = L/\Delta x$, $h_{opt} \propto \Delta x^{1/5}$
- Best to use $n - v$ odd. Same variance and bias as next-lowest even but more parameters to fit with.
Effectiveness of Proper Smoothing Length
Finding the Signal in the Noise with a Certain Bandwidth Selector

From Fan & Gijbels

Los Alamos

Hydrodynamic Methods
Help from Statistics

- Statisticians have well-developed notions that may be of use in the analysis of experimental and computational data for mechanics
- Decreasing sensitivity to outliers with LOWESS
- Maximum likelihood
- Confidence intervals
- Optimality
  - best smoother is LPR
- Asymptotic equivalent kernel
Reducing Influence of Outliers with LOWESS

Weights can be manipulated to reduce the effect of outliers.

\[ u \approx \beta e_0 = \hat{u} \]
\[ \beta(x) = \sum_j u_j \psi_j(x)^T \]
\[ \psi_i(x) = B(x)^{-1} p_i(x) W_i(x) \]
\[ B(x) = \sum_j p_j(x)^T p_j(x) W_j(x) \]

Iteratively de-weight the particles with large residuals

\[ W_i(x) \rightarrow W_i(x) L \left( \frac{r_i^k}{R_k} \right) \]
\[ r_i^k = \hat{u}_{k-1}(x_i) - u_i \]
\[ L(x) = \left( 1 - x^2 \right)^+ \]
\[ R_k = 6 \text{ median} \{ r_i^k \}_{i=1}^N \]

From Fan + Gijbels
Confidence Intervals

- It is possible to construct confidence intervals for LPR estimators.
- After folding through time integrator, is possible to set confidence limits on a hydro calculation?

\[
\hat{u}^{(v)}(x) - \hat{b}_{v,n}(x) \pm z_{1-\alpha/2} \hat{V}_{v,n}(x)^{1/2}
\]

\[\hat{b}_{v,n}(x) \quad \text{Estimate for bias}
\]
\[\hat{V}_{v,n}(x) \quad \text{Estimate for variance}
\]
\[z_{1-\alpha/2} \quad \text{Gaussian quantile}
\]

From Fan + Gijbels
The Best Kernel and Smoother

- The optimal kernel for LPR is the Epanechnikov:

\[ K(x) = c \cdot (1 - x^2)_+ \]

- Theorem: With this kernel, LPR is THE optimal linear smoother.

- Didn’t work to well for hydro
- Not smooth at neighbor loss?
- Continue to use cubic B-spline
Equivalent Kernel

- In continuum limit, one can show that the LPR estimator converges to an integral kernel approximation:

\[ \int_{-1}^{1} y^m K^*_v(y) \, dy = \delta_{m\nu}, \quad m, \nu < n \]

\[ \int_{x_0-h}^{x_0+h} f(x) K^*_v \left( \frac{x-x_0}{h} \right) \, dx = f^{(\nu)}(x_0) + O(h^n) \]

- Order of error is at least \( n \) for all derivatives.
- Of great theoretical utility.

From Fan + Gijbels
LRE for Discontinuities
(cf. Loader)

\[
\varepsilon(u, \hat{u}_c) = \lambda_0 \left| \frac{u - \hat{u}_c}{\hat{u}_c} \right| + \lambda_1 \left| \frac{u' - \hat{u}'_c}{\hat{u}'_c} \right| + \lambda_2 \left| \frac{u'' - \hat{u}''_c}{\hat{u}''_c} \right|
\]

\[
\Lambda^{(n)}(\hat{u}_l, \hat{u}_c, \hat{u}_r) = \begin{cases} 
\hat{u}_l^{(n)}, & \varepsilon(u_l, \hat{u}_c) \leq \varepsilon(u_r, \hat{u}_c) \\
\hat{u}_r^{(n)}, & \varepsilon(u_r, \hat{u}_c) \leq \varepsilon(u_l, \hat{u}_c)
\end{cases}
\]

\[e^{-x^2/2} \sin(2\pi x^2)\]
Differential Equations and LRE

\[ p = [1, t, t^2/2]^T \]

**ODE**

\[
\begin{aligned}
\dot{t} &= F(t)u \\
\ddot{t} &= (F(t)^2 + \ddot{F}(t))u \\
\beta e_1 - F(t) \beta e_0 &= 0 \\
\beta e_2 - (F(t)^2 + \dot{F}(t)) \beta e_0 &= 0
\end{aligned}
\]

**PDE**

\[
\begin{aligned}
\frac{\partial}{\partial t} u + \frac{\partial}{\partial x} F(u) &= G(u) \\
\frac{\partial}{\partial x} \frac{\partial}{\partial t} F(u) &= \frac{\partial}{\partial t} G(u) \\
\frac{\partial}{\partial x} \frac{\partial}{\partial x} F(u) &= \frac{\partial}{\partial x} G(u) \\
\beta e_1 + D_u F(\beta e_0) \beta e_2 &= G(\beta e_0) \\
\beta e_4 + D_u F(\beta e_0) \beta e_5 + D_u^2 F(\beta e_0) \beta e_1 \beta e_2 &= D_u G(\beta e_0) \beta e_1 \\
\beta e_5 + D_u F(\beta e_0) \beta e_6 + D_u^2 F(\beta e_0) \beta e_2 \beta e_2 &= D_u G(\beta e_0) \beta e_1
\end{aligned}
\]

Differential relations are algebraic constraints on LRE solution \( \mathcal{D}(\beta) = 0 \)

\[ u_3 = u_1 u_2 \]

**Leibnitz’ Rule**

\[
\begin{aligned}
\partial_x u_3 &= u_2 \partial_x u_1 + u_1 \partial_x u_2 \\
e_3^T \beta e_1 &= (e_2^T \beta e_0)(e_1^T \beta e_1) + (e_1^T \beta e_0)(e_2^T \beta e_1)
\end{aligned}
\]

**Thermodynamics**

\[ p = [1, x, \ldots]^T \]

\[
\begin{aligned}
T \, ds &= d\eta + pdv \\
T \, \partial_x s &= \partial_x \eta + p \partial_x v \\
(e_T \beta e_0)(e_\eta \beta e_1) &= (e_\eta \beta e_1) + (e_p \beta e_0)(e_v \beta e_1)
\end{aligned}
\]

Los Alamos

10/4/01 - 28

Hydrodynamic Methods
Parameterized Differential Relations

- The differential relations we are interested in can often have unknown parameters associated with them
  - Eigenvalues
  - Material constants
  - Unknown design specs
- In general we write

\[ \mathcal{D}(\alpha; \beta) = 0 \]
Point Constraints and LRE

In general, $\beta$ depends on the $u_i$.

Floating interior data are determined by minimization.

Global implicit system results whose sources are boundary specs.

The “stiffness matrix” is the coefficient of the $u_i$ in the expression for $\beta$.

A point on boundary can be used to specify boundary conditions.

Invert the MLS matrix $\phi_i(x_j)$

$\mathcal{L}(x) = \sum_j \left( u_j - \beta(x)p_j(x) \right)^2 W_j(x)$

$x_i \in \mathcal{E}, \quad \beta(x_i)H_i = h_i$

$x_i \notin \mathcal{E}, \quad p_i(x_i) = [1, 0, \cdots]^T, \quad W_i(x_i) = W_0$

$\frac{\partial \mathcal{L}}{\partial u_i}(x_i) = 0 \quad \Rightarrow \quad u_i = \beta(x_i)e_0 = \hat{u}(x_i)$
Tuned Regression Estimators

General Prescription

\[ d \text{ space + time dimensions } \quad x \in \mathbb{R}^d \]

\[ m \text{ unknowns } \quad u: \mathbb{R}^d \rightarrow \mathbb{R}^m \]

\[ q \text{ parameters } \quad \alpha \in \mathbb{R}^q \]

\[ n \text{ derivatives, } r \text{ equations } \quad F: \mathbb{R}^{mn} \otimes \mathbb{R}^q \rightarrow \mathbb{R}^r \]

\[ r \text{ sources } \quad G: \mathbb{R}^d \otimes \mathbb{R}^q \rightarrow \mathbb{R}^r \]

differential relations

\[ \mathcal{D}[\alpha; u(x), D_j u(x), \ldots] = 0 \]

\[
\mathcal{L}(x) = \sum_j \left( u_j - \beta(x) p_i(x) \right)^2 W_j(x)
\]

\[
\mathcal{D}[\alpha; (\beta e_0)^T, (\beta e_1)^T, \ldots] = 0
\]

\[ \beta(x_i) H_i = h_i, \quad i \in \mathcal{B} \]

\[ \frac{\partial \mathcal{L}}{\partial \alpha} = 0 \]

\[ \frac{\partial \mathcal{L}}{\partial \beta} = 0 \]

\[ \left( \frac{\partial \mathcal{L}}{\partial u_i} \right)(x_i) = 0, \quad i \notin \mathcal{B} \]

Weighted error of local series expansion

Match the differential relations (tuning)

Match point constraints

Best fit for parameters

Best fit for derivatives (explicitness)

Best fit for unknown data (implicitness)
Features of Tuned Regression

- Unified formalism for two modes of use:
  - Forward modeling: Simulation + Prediction (physicists and engineers)
  - Inverse modeling: Analysis + Fitting (statisticians)
  - Allows blending both seemlessly
    - E.g. back out material characterizations using finite wave sample
  - Allows for noise, randomness, fuzziness in data
- Nonlinear tuning causes loss of factorization $\beta = \sum_j u_j \psi_j^T$.
- Nonlinear tuning may preclude explicit solution of constraints.
  - Iterative procedure likely required.
- Residual of differential relations is always zero.
  - Price: estimate of derivative $\neq$ derivative of estimate
- Tuning causes elimination of components of $\beta$, giving smaller matrices to invert, giving better answer faster.
Forward Tuned Regression Example

- 201 points
- 40% random x
- $h/dx = 1.0$

- $\frac{\partial \mathcal{L}}{\partial u_i(x_i)} = 0$
- $\beta(\pm 1)e_0 = u(\pm 1)$

- Used tuning
- Inverted matrix to solve

\[ \dot{u} = -4u + e^{-4x} \left( \frac{20\cos[20x]}{x} - \frac{\sin[20x]}{x^2} \right) \]
Inverse Tuned Regression Example

\[ \dot{u} = \begin{bmatrix} -0.2 & \frac{2\pi}{5} \\ \frac{2\pi}{5} & -0.2 \end{bmatrix} u \]

Quadratic LPR

Tuned Quadratic LPR
ODE + ODE'
Inverse Tuned Regression Convergence

-3.28456 - 2.12287 Log h

-3.12521 - 3.00877 Log h

- 6 to 81 points
- 30% random points and h’s sampled from 0 to 10
- Expect cubic convergence from each, as both are quadratic
- Typical methods lose an order of convergence on non-uniform grids.
- LPR is no exception

- Tuning regains the lost order of accuracy.
- Tuning reduces 6 unknowns to 2.
- Invert 2x2 matrix instead of 3x3.
- Better answer at less cost.
- Can use to replace semi-discretization with a space-time particle method for hydro?
Summary

- Much MLS experience: successful mesh-based ideas may lack some essential feature when transferred to meshless world.
- Statistical LPR brings meshless full-circle to its origins.
- Statistical literature is full of rigorous mathematical results which could inspire rigorous developments in meshless mechanics.
- Statistical literature inspires many intriguing naturally meshless analogues to mesh-based ideas.
- 23 years of literature to catch up on. Stone: 277 citations
- Lots of development yet for mechanics applications.
- Take small steps.
  - ODE’S, advection, Poisson, Helmholtz, Euler
  - Deterministic + stochastic
  - Convergence, stability, etc.
- Work enough for an army of students and post-docs.