

Sparsity in Biomedical Imaging and Beyond

Christoph Brune

joint work with [Andrea Bertozzi](#), [Stanley Osher](#)
[Jeff Brantingham](#)¹ and [Guillermo Sapiro](#)²

Department of Mathematics, UCLA

¹Department of Anthropology, UCLA

²University of Minnesota

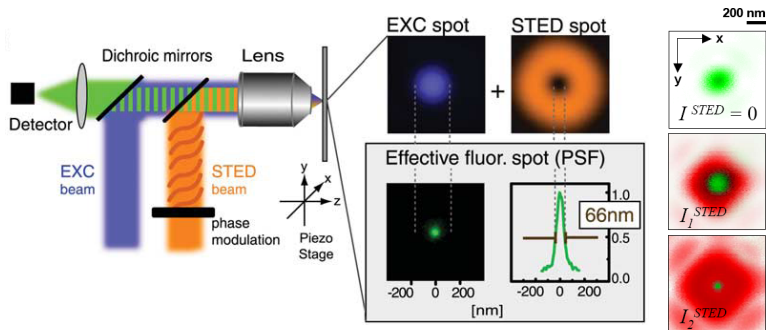
*Compressive Sensing Workshop: Leveraging
Sparsity at UCLA & Beyond, March 6-8, 2012*

The UCLA logo, consisting of the letters "UCLA" in white, bold, sans-serif font, centered within a dark blue rectangular box.

- 1 Sparsity in Static Imaging (3D)
- 2 Sparsity in Dynamic Imaging (4D)
- 3 Sparse Control of Crime

→ Shelley's talk

■ Stimulated Emission Depletion (Hell Group, MPI Göttingen)



■ Fluorescence by focused laser beam; Stimulation + Depletion

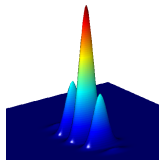
Deconvolution problems with Poisson noise

$$Ku = f \quad \text{with} \quad (Ku)(x) = \int_{\Omega} k(x-y)u(y)dy$$

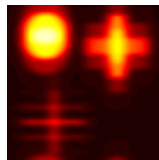
- Nanoscopy provides **blurred** images
- Statistical noise model, described by likelihood $p(f|u)$
→ f_i realization of random var. F_i
- **Poisson**-noise resulting from laser sampling (photon counts)



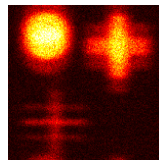
exact data u



kernel k (psf)



convolved data



convolved &
noisy data f

- Maximum a-posteriori probability (MAP) estimation

$$\hat{u} = \arg \min_{u \geq 0} (-p(u|f))$$

- Bayesian model: $p(u|f) = \frac{p(f|u)p(u)}{p(f)}$
- Prior densities: $p(u) \sim e^{-\alpha J(u)}$ (Gibbs measures) $\leftarrow$ sparsity priors
- Different noise models imply different data fidelities, e.g.

$$\text{Gaussian} \quad \leadsto \quad \min_{u \geq 0} \left\{ \frac{1}{2} \|Ku - f\|_{L^2(\Sigma)}^2 + \alpha J(u) \right\}$$

$$\text{Poisson} \quad \leadsto \quad \min_{u \geq 0} \left\{ \int_{\Sigma} (Ku - f \log Ku) d\mu + \alpha J(u) \right\}$$

Kullback-Leibler functional, strong nonlinearity

3D Kullback-Leibler-TV Reconstruction

$$\min_{u \geq 0} \int_{\Sigma} (Ku - f \log Ku) d\mu + \alpha |u|_{BV}$$

- TV-Regularization to preserve edges (**sparsity in the gradient**)

$$|u|_{BV} = \sup_{\substack{g \in C_0^\infty(\Omega; \mathbb{R}^d) \\ \|g\|_\infty \leq 1}} \int_{\Omega} u \operatorname{div} g \quad (= \int_{\Omega} |\nabla u|)$$

EM-TV, modified forward backward splitting

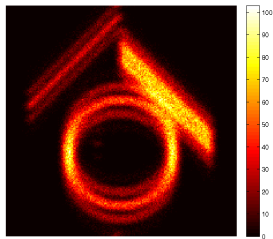
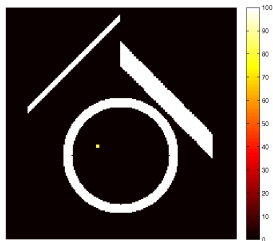
$$u_{k+\frac{1}{2}} = u_k K^* \left(\frac{f}{Ku_k} \right) \quad \text{EM step}$$

$$u_{k+1} = u_{k+\frac{1}{2}} - \alpha u_k p_{k+1} \quad \text{TV step, } p_{k+1} \in \partial |u_{k+1}|_{BV}$$

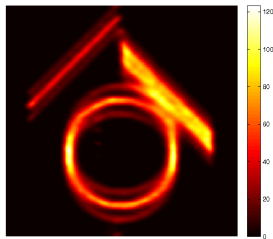
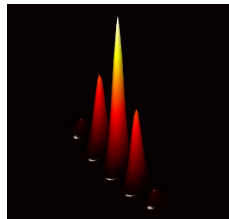
→ **Luminita's talk, Ming Yan**

[B. , Sawatzky, Burger - Bregman-EM-TV Methods with Application to Optical Nanoscopy, 2009]

Example: Optical Nanoscopy

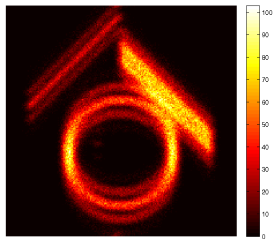
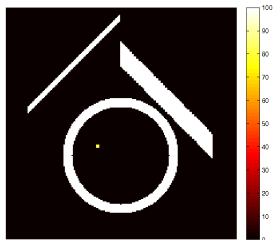


Convolved + Poisson noise

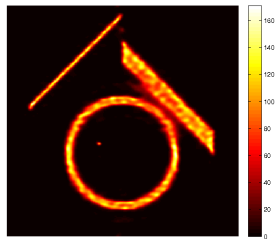
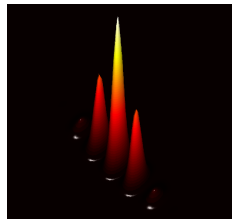


EM Reconstruction, 5 its

Example: Optical Nanoscopy

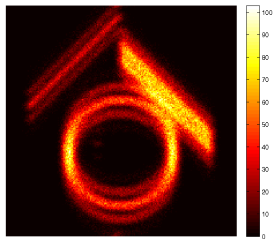
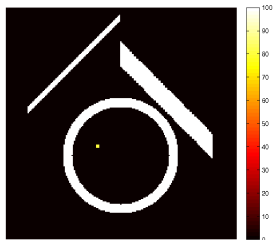


Convolved + Poisson noise

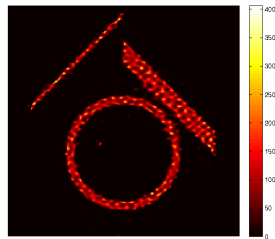
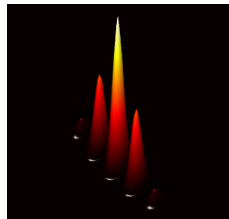


EM Reconstruction, 50 its

Example: Optical Nanoscopy

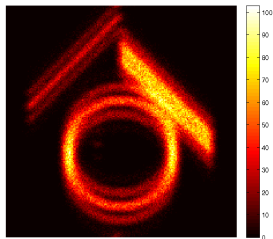
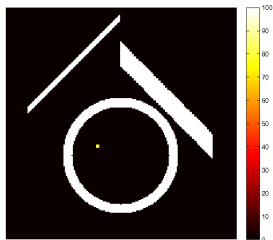


Convolved + Poisson noise

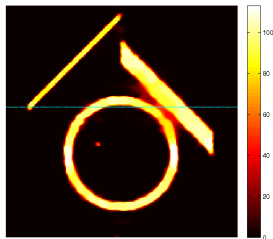
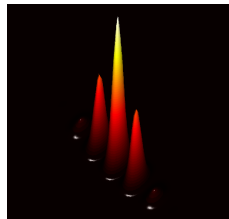


EM Reconstruction, 1000 its

Example: Optical Nanoscopy



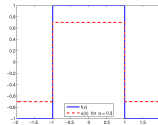
Convolved + Poisson noise



EM-TV Reconstruction

- Systematic error of TV - **contrast reduction**

$$|u|_{BV} = \sup_{\|g\|_{\infty} \leq 1} \int_{\Omega} u \operatorname{div} g \quad (= \int_{\Omega} |\nabla u|)$$

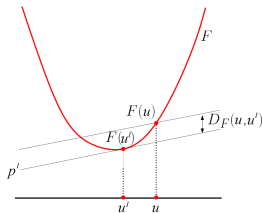


Iterative Bregman-TV Regularization

$$p^l \in \partial |u^l|_{BV}$$

$$u^{l+1} = \arg \min_{u \in BV(\Omega)} \{ H_f(Ku - f) + \alpha \underbrace{(|u|_{BV} - |u^l|_{BV} - \langle p^l, u - u^l \rangle)}_{D_{|\cdot|_{BV}}^{p^l}(u, u^l)} \}$$

$D_{|\cdot|_{BV}}^{p^l}(u, u^l)$ Bregman distance



- $D_{|\cdot|_{BV}}^{p^l}(u, u^l) \geq 0$
 $= 0$ if $p^l \in \partial |u|_{BV}$

particularly if $u \equiv c u^l \quad \forall c \in \mathbb{R}_0^+$

- keeps edges, **enhanced contrast**

[Meyer 01], [Osher-Burger-Goldfarb-Xu-Yin 05]

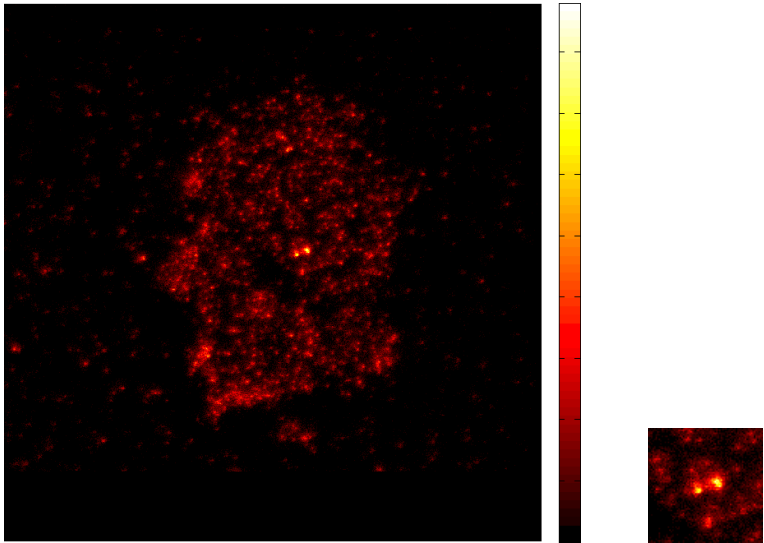


Figure: Protein Syntaxin, 53nm, MPI Göttingen (Stefan Hell)

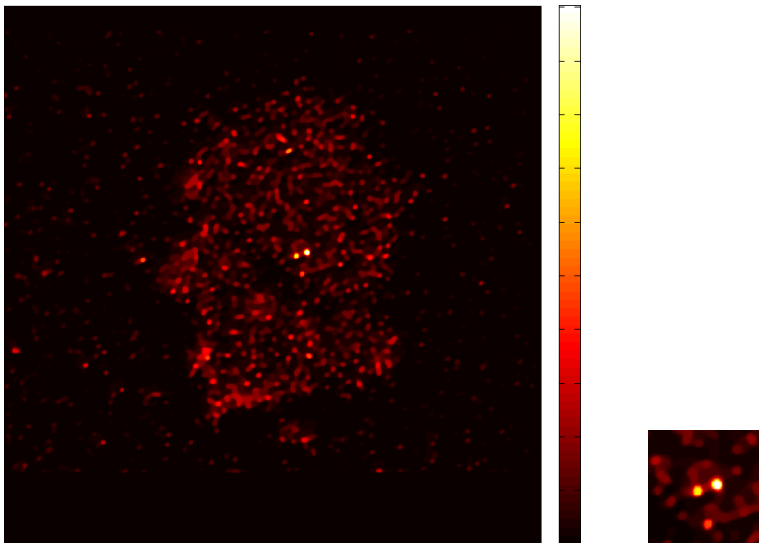


Figure: EM-TV Reconstruction

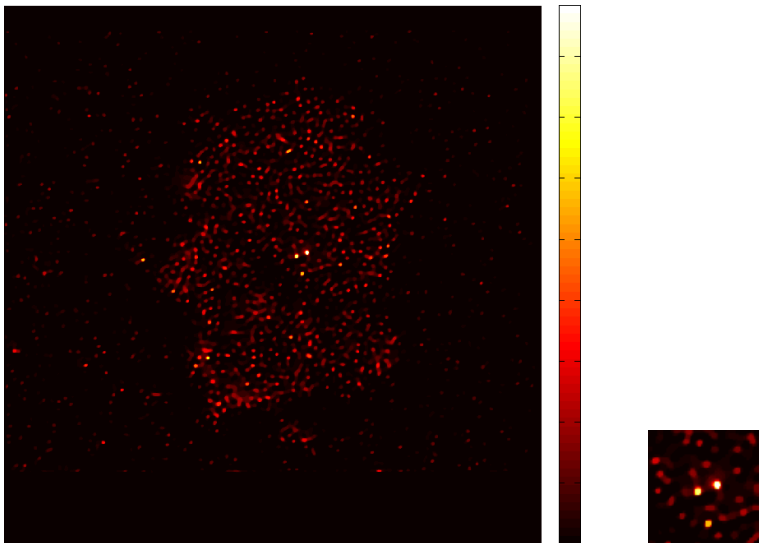
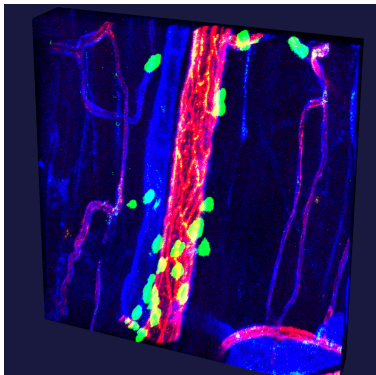
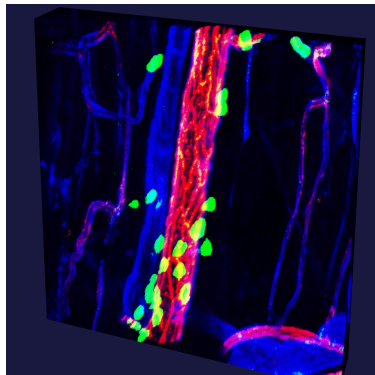


Figure: Bregman-EM-TV Reconstruction

Sharp Microscopy in Cell Transmigration



Data: MPI for Molecular Biomed., Germany



3D EM-TV Reconstruction

[Benning, B., Burger, Müller - Higher-Order TV Methods - Enhancement via Bregman Iteration, UCLA CAM (12-04)]

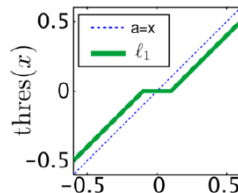
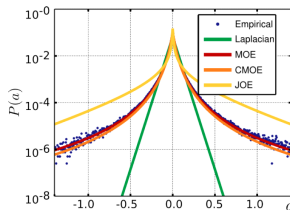
with Guillermo Sapiro and Stanley Osher:

Gaussian Mixture Models (GMM) + Laplacian Mixture Models (LMM)

Total Variation prior:

$$p(u) = \theta_0 e^{-\theta_0 |\nabla u|}$$

shape and θ_0 fixed



Can we control shape and scale of Laplacians based on the data?

Yes! We need hyperpriors and EM.

Conjugate prior of Laplacian distr. is Gamma distr.

$$\Rightarrow Q_{MOE}(\nabla u | \kappa, \beta) = \frac{1}{2} \kappa \beta^\kappa (|\nabla u| + \beta)^{-(\kappa+1)}$$

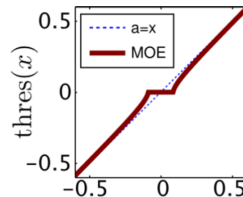
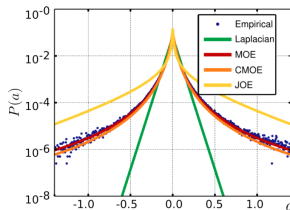
with Guillermo Sapiro and Stanley Osher:

Gaussian Mixture Models (GMM) + Laplacian Mixture Models (LMM)

Total Variation prior:

$$p(u) = \theta_0 e^{-\theta_0 |\nabla u|}$$

shape and θ_0 fixed



Can we control shape and scale of Laplacians based on the data?

Yes! We need hyperpriors and EM.

Conjugate prior of Laplacian distr. is Gamma distr.

$$\Rightarrow Q_{MOE}(\nabla u \mid \kappa, \beta) = \frac{1}{2} \kappa \beta^\kappa (|\nabla u| + \beta)^{-(\kappa+1)}$$

$$f = Ku + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2), \quad u \sim \mathcal{N}(\eta_k, \Sigma_k), \quad Q_{MOE}(\nabla u \mid \kappa, \beta)$$

Reconstruct and learn in alternation:

1.) E-step (signal estimation and model selection (clustering)):

$$\begin{aligned} (u, k) &= \arg \max_{u, k} \log p(u \mid f, (\Sigma_k, \mu_k, \kappa, \beta)) \quad \rightsquigarrow \text{Bayes} \rightsquigarrow \\ &= \arg \min_{u, k} \frac{\|Ku - f\|_2^2}{2\sigma^2} + (u - \mu_k)^T \Sigma_k (u - \mu_k) + (2\kappa + 1) \int_{\Omega} \log(|\nabla u| + \beta) \end{aligned}$$

2.) M-step (model selection, estimate params):

$$\kappa = \frac{2(\nu_2 - \nu_1^2)}{\nu_2 - 2\nu_1^2} \text{ and } \beta = \frac{\kappa - 1}{\nu_1}, \quad \text{with } \nu_1 = \int |\nabla u| \quad \text{and} \quad \nu_2 = \int |\nabla u|^2$$

first and second non-central moments

coming soon:

[B., Sapiro, Osher - Gaussian Mixture Models meet Mixture Total Variation, 2012]

[Yu, Sapiro, Mallat - Solving Inverse Problems with Piecewise Linear Estimators From Gaussian Mixture Models to Structured Sparsity - 2011]

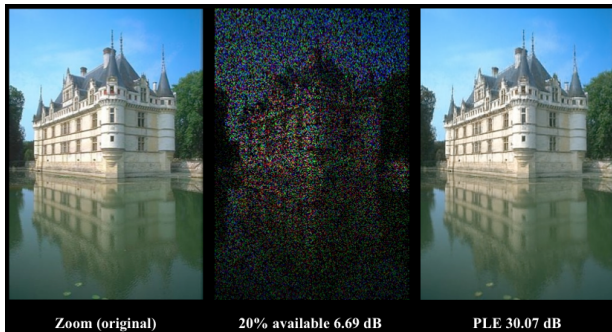
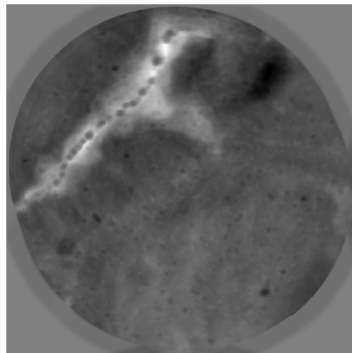
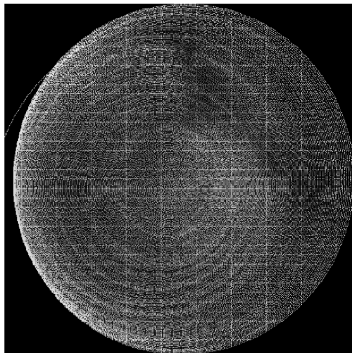


Figure: inpainting and sparsity, Sapiro et al. 2011

Quality very close to BM3D and nonlocal methods.

TV-Inpainting for Realtime Sparse Atomic Force Microscopy (AFM)

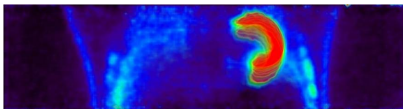
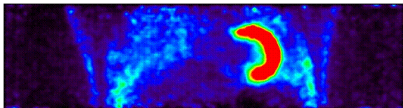


Data: Paul Ashby, Berkeley National Lab

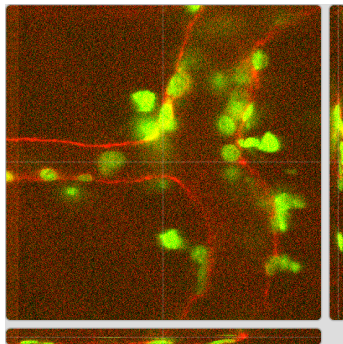
New summer **REU project** on STM with Shelley, Paul and Andrea

Sparsity in Dynamic Imaging (4D)

Motion is Life



4D PET data from EIMI, Germany



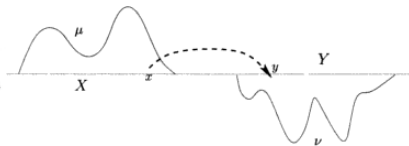
4D live microscopy, MPI Molec.
Biomed., Germany

Challenges in dynamic imaging:

- Patient lung and myocardial motion leads to blurring
- Live cell imaging - cell or protein migration (e.g. dynamic STED)



Earth mover's distance



L^2 -Wasserstein metric: $W_2(\rho_0, \rho_T)^2 := \inf_M \int |M(x) - x|^2 \rho_0(x) dx$

Formulation of Benamou-Brenier, Continuum Mechanics

$$\min_{(\rho, \mathbf{v})} \frac{1}{2} \int_0^T \int_{\Omega} \rho(x, t) |\mathbf{v}(x, t)|^2$$

subject to $\rho_t + \nabla \cdot (\rho \mathbf{v}) = 0$ (mass conservation)

$$\rho(0, \cdot) = \rho_0$$

$$\rho(T, \cdot) = \rho_T$$

4D Reconstruction using Optimal Transport

$$\min_{\rho, \mathbf{v}} \frac{1}{2} \int_0^T H_{f(\cdot, t)}(K\rho(\cdot, t)) dt + \alpha \int_0^T J(\rho(\cdot, t)) dt + \frac{\beta}{2} \int_0^T \int_{\Omega} \rho |\mathbf{v}|^2 dx dt$$

$$\begin{aligned} \text{subject to} \quad & \partial_t \rho + \nabla_{\Omega} \cdot (\rho \mathbf{v}) = 0 \quad \text{in } Q \quad (\text{mass conservation}) \\ & \rho \geq 0 \end{aligned}$$

4D Reconstruction using Optimal Transport

$$\min_{\rho, \mathbf{v}} \frac{1}{2} \int_0^T H_{f(\cdot, t)}(K\rho(\cdot, t)) dt + \alpha \int_0^T J(\rho(\cdot, t)) dt + \frac{\beta}{2} \int_0^T \int_{\Omega} \rho |\mathbf{v}|^2 dx dt$$

$$\text{subject to} \quad \partial_t \rho + \nabla_{\Omega} \cdot (\rho \mathbf{v}) = 0 \quad \text{in } Q \quad (\text{mass conservation})$$

$$\rho \geq 0$$

- Different data fidelities (noise models)

$$H_{f(\cdot, t)}(K\rho(\cdot, t)) := \frac{1}{2} \|K\rho(\cdot, t) - f(\cdot, t)\|_{L^2(\Sigma)}^2 \quad (\text{Gaussian noise})$$

$$H_{f(\cdot, t)}(K\rho(\cdot, t)) := \int_{\Sigma} K\rho(\cdot, t) - f(\cdot, t) \log(K\rho(\cdot, t)) \quad (\text{Poisson noise})$$

4D Reconstruction using Optimal Transport

$$\min_{\rho, \mathbf{v}} \frac{1}{2} \int_0^T H_{f(\cdot, t)}(K\rho(\cdot, t)) dt + \alpha \int_0^T J(\rho(\cdot, t)) dt + \frac{\beta}{2} \int_0^T \int_{\Omega} \rho |\mathbf{v}|^2 dx dt$$

subject to $\partial_t \rho + \nabla_{\Omega} \cdot (\rho \mathbf{v}) = 0$ in Q (mass conservation)

$$\rho \geq 0$$

- Different regularization terms (prior assumptions)

$$J(\rho(\cdot, t)) := \frac{1}{2} \|\nabla \rho(\cdot, t)\|_2^2$$

$$J(\rho(\cdot, t)) := \frac{1}{p} (|\rho(\cdot, t)|_{BV})^p \rightsquigarrow \text{choice of } p?$$

4D Reconstruction using Optimal Transport

$$\min_{\rho, \mathbf{v}} \frac{1}{2} \int_0^T H_{f(\cdot, t)}(K\rho(\cdot, t)) dt + \alpha \int_0^T J(\rho(\cdot, t)) dt + \frac{\beta}{2} \int_0^T \int_{\Omega} \rho |\mathbf{v}|^2 dx dt$$

subject to $\partial_t \rho + \nabla_{\Omega} \cdot (\rho \mathbf{v}) = 0$ in Q (mass conservation)

$$\rho \geq 0$$

■ Optimal transport

Joint Reconstruction, Total Variation and Optimal Transport

$$\min_{\rho, \mathbf{v}} \frac{1}{2} \int_0^T \|K\rho - f\|_{L^2(\Sigma)}^2 dt + \frac{\alpha}{p} \int_0^T (|\rho|_{BV(\Omega)})^p dt + \frac{\beta}{2} \int_0^T \int_{\Omega} \rho |\mathbf{v}|^2$$

$$\begin{aligned} \text{subject to} \quad & \partial_t \rho + \nabla_{\Omega} \cdot (\rho \mathbf{v}) = 0 \quad \text{in } Q \quad (\text{mass conservation}) \\ & \rho \geq 0 \end{aligned}$$

Challenges:

- Existence (for $p \in (1, 2]$), uniqueness, convergence
- High dimensional data, computational aspects
- Preconditioning, parallelization on GPUs (CUDA, OpenCL)

[B. - 4D Imaging in Tomography and Optical Nanoscopy, 2010]

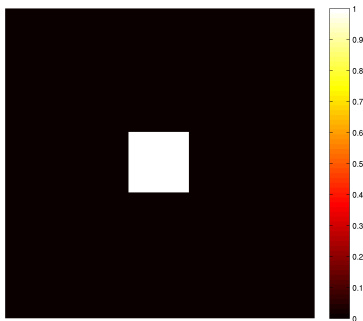
Preconditioning matrix: $P_\delta := \frac{1}{\delta} - \tilde{K}^* \tilde{K}$

$$Q := \Omega \times [0, T]$$

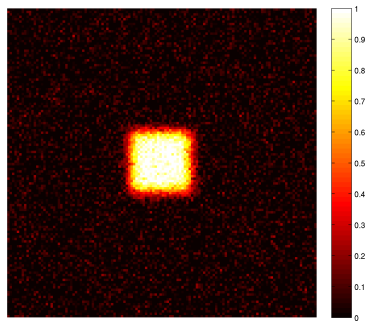
$$\begin{aligned} \rho^{k+1} &= \left(\frac{1}{\delta} + \mu_3 - \mu_1 \partial_t^2 - \mu_2 \Delta_\Omega + P_\delta \right)^{-1} \\ &\quad \left(\tilde{K}^* f + \partial_t(\mu_1 \nabla_\Omega \cdot \mathbf{j}^k + \mathbf{q}^k) + \nabla_\Omega(y^k - \mu_2 z^k) + \mu_3 w^k + r^k + P_\delta \cdot \rho^k \right) \\ \mathbf{j}^{k+1} &= (\mu_1 \nabla_\Omega(\nabla_\Omega \cdot) - \mu_4)^{-1} (-\nabla_\Omega(\mu_1 \partial_t \rho^{k+1} + \mathbf{q}^k) - \mu_4 \nu^k - \eta^k) \\ z^{k+1} &= \textit{Shrink} \left(\nabla_\Omega \rho^{k+1} + \frac{1}{\mu_2} y^k, \frac{\alpha}{\mu_2} \left(\int_\Omega |z^k| dx \right)^{p-1} \right), \quad \forall t \in [0, T] \\ 0 &= (w^{k+1})^3 + \left(\frac{r^k}{\mu_3} - \rho^{k+1} \right) \cdot (w^{k+1})^2 - \frac{\beta}{\mu_3} |\nu^k|^2 \\ \nu^{k+1} &= \frac{\mu_4 \mathbf{j}^{k+1} - \eta^k}{\frac{\beta}{w^{k+1}} + \mu_4} \end{aligned}$$

$\implies$ Efficient subproblems: explicit, thresholding or DCT/FFT (CUDA)

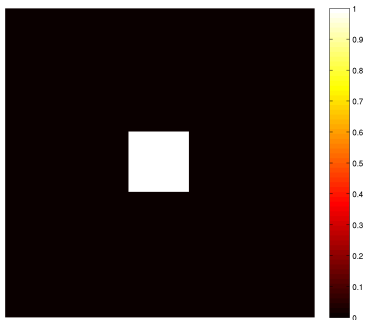
Exact Data



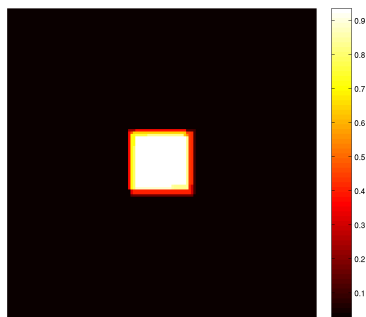
Blurred and noisy data



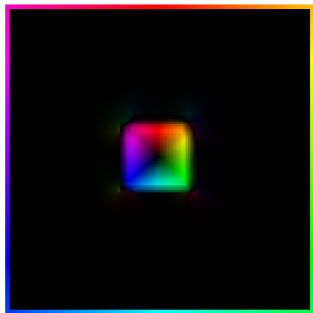
Exact Data



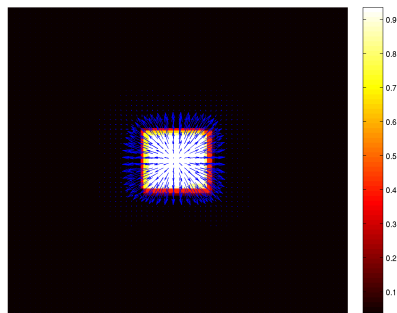
Reconstruction



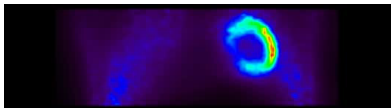
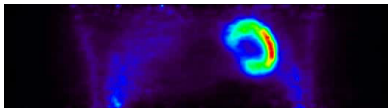
Motion: color visualization



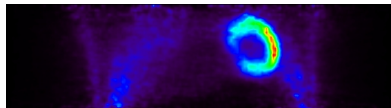
vector plot



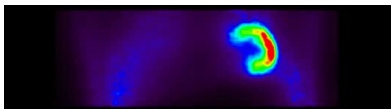
Result Registration PET



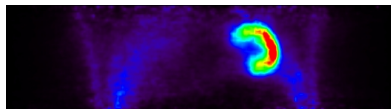
Registration to diastole



Target diastole



Registration to systole



Target systole

- Ongoing project with Hao Gao and Stanley Osher:
4D Imaging via Optimal Transport + Sparsity and Low Rank

$$\begin{aligned} \min_{\rho_1, \rho_2, \mathbf{v}} \quad & \frac{1}{2} \int_0^T \|K\rho - f\|_2^2 dt + \|\rho_1\|_* + \|W\rho_2\|_1 + \int_0^T \int_{\Omega} \rho |\mathbf{v}|^2 \\ \text{s.t.} \quad & \rho = \rho_1 + \rho_2 \\ & \partial_t \rho + \nabla_{\Omega} \cdot (\rho \mathbf{v}) = 0 \quad \text{in } Q, \quad (\text{mass conservation}) \\ & \rho \geq 0 \end{aligned}$$

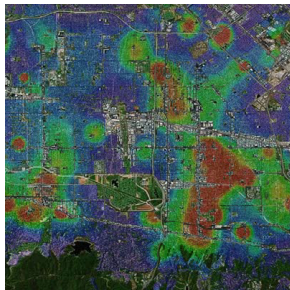
- **Simultaneous** motion estimation and video decomposition
- Addressing components **in time** which are **difficult to detect** by just sparsity and low rank (e.g. relatively high foreground speed (rain), relatively slow objects, occlusions,...)

Sparse Control of Crime

with Joe Zipkin, Martin Short, Andrea Bertozzi and Jeff Brantingham

Optimal Control in Crime Modeling

Prediction and Validation via Inverse Problems



Motivation:

- Instead of **FORWARD** simulation (Aggregation models, Chemotaxis,...)
→ **Optimal Control** / **INVERSE** problem
- **Fitting** to **real crime data**
- Use of known PDE crime models

Question: What is the optimal law enforcement strategy to **explain/predict** a given sequence of crime rates?



Optimal Control Model and Algorithm

$$\min_{\rho_P, \rho_C, A} \frac{1}{2} \int_{\Omega} \sum_{k=1}^N |S_{t_k}(A_0, \rho_{C0}, \rho_P) - Y_k|^2 d\Omega + \frac{\alpha}{2} \|\rho_P\|_{L^2(\Omega \times [0, T])}^2$$

subject to

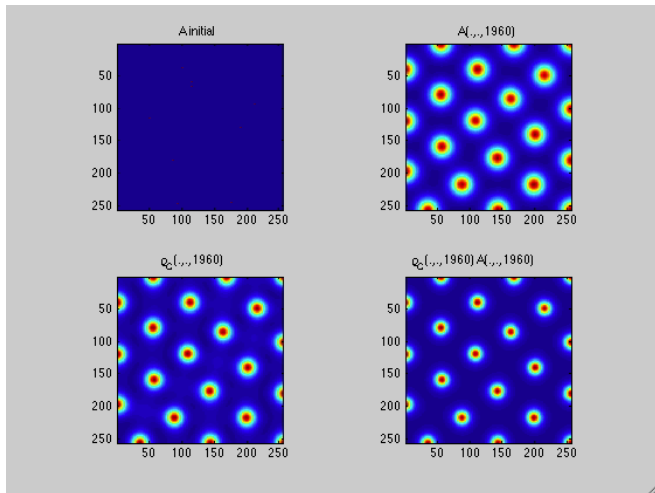
$$\begin{cases} \frac{\partial A}{\partial t} = \Delta A - A + A^0 + \rho_C A + \rho_P(x, t) & \text{in } \Omega \times]0, T] \\ \frac{\partial \rho_C}{\partial t} = \Delta \rho_C - 2\nabla \cdot \left(\frac{\rho_C}{A} \nabla A \right) - \rho_C A + \bar{A} - A^0 & \text{in } \Omega \times]0, T] \\ A(x, 0) = A_0(x), \quad \rho_C(x, 0) = \rho_{C0}(x) . \end{cases}$$

Control-to-state mapping: $S_{t_k} : W \times X \times Y \rightarrow U$

$$(A_0, \rho_{C0}, \rho_P) \mapsto \rho_C(\cdot, t_k) A(\cdot, t_k)$$

Efficient impl. via spectral methods (FFT/DCT), allows for parallelization and GPUs

Inverse Crime Modeling



[Short et al - A Statistical Model Of Criminal Behavior, MMAS 2008]

sparsity in crime control: "police cannot be everywhere" $\rightarrow$ sparse influence and limited resources

- modified L^1 -norm control regularization (**directional sparsity**)

$$\min_{\rho_P, \rho_C, A} \frac{1}{2} \int_{\Omega} |S_T(A_0, \rho_{C_0}, \rho_P) - Y|^2 d\Omega + \frac{\alpha}{2} \|\rho_P\|_{1,2}$$

L^1 resp. L^0 norm: cost of "placing an officer onto the map"

linear control vs. nonlinear control $\rightarrow$ related to cops on the dots???

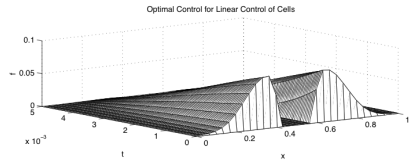
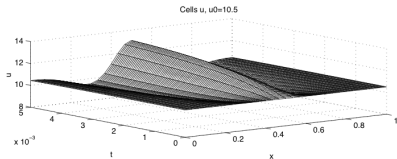
additional control constraints: "limited resources (in space and time)"

$$0 \leq \rho_P(x, t) \leq 1, \quad \text{in } \Omega \times [0, T]$$

[Herzog, Stadler, Wachsmuth - Directional Sparsity in Optimal Control of PDEs, 2011]

[B., Maurer, Wagner - ...Optical Flow via Multidimensional Control, SIAM IS 2009]

"cops on the dots" vs. surrounding or beleaguer



beleaguer?!

[Fistler, McCarthy - Optimal Control of a Chemotaxis System, 2003], ←linear control

[Jones, Brantingham, Chayes - Statistical models of criminal behavior: The effects of law enforcement actions, MMMAS 2010]

→ Talk tomorrow by Jeff Brantingham

Thanks for your attention!

`http://www.math.ucla.edu/~brune`

Department of Mathematics, UCLA