Laplace deconvolution and its application to the analysis of dynamic contrast enhanced medical imaging data

Marianna Pensky, University of Central Florida, USA

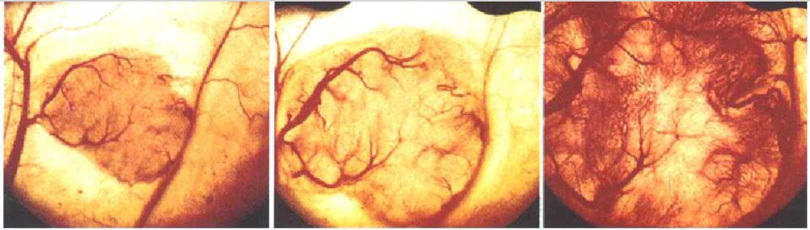
joint work with
Felix Abramovich, Tel-Aviv University, Israel
Fabienne Comte, Charles-Andre Cuenod
and Yves Rozenholc
University of Paris V, France

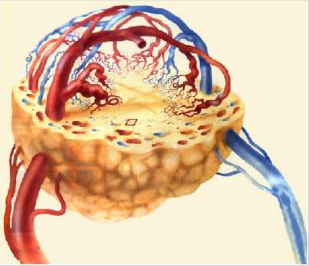
"Inverse Problems", February 2016
Lumini, CIRM, France

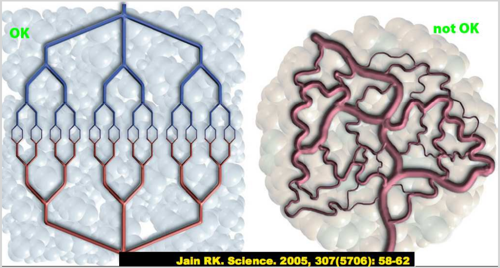
1. Medical framework
2. Formulation of the problem
3. Laplace deconvolution - challenges
4. Laplace deconvolution with a fully known kernel: resolvent solution
5. Laplace deconvolution with a data-driven kernel
6. Application of overcomplete dictionaries

Angiogenesis in cancer

Angiogenesis : growth of new blood vessels from pre-existing vessels.







Angiogenesis in cancer

Facts on tumors

- Population of rapidly dividing and growing cells
- Cannot grow beyond a small size if lack of oxygen and nutrients
- Induce angiogenesis by secreting growth factors
- Develop fast, anarchic and inefficient blood pathways
- Spread in metastases through blood vessels

From the relations: **An early 90's point of view**

$$\textbf{tumor} \rightleftharpoons \textbf{growth} \rightleftharpoons \textbf{energy} \rightleftharpoons \textbf{glucose} \rightleftharpoons \textbf{vascularization}$$

anti-angiogenesis treatments are introduced with expectation that:

"Reducing angiogenesis will asphyxiate the tumor".

Angiogenesis in cancer

An interesting but unfortunately wrong idea

Penalizing angiogenesis induces:

- a regularization of the tumor blood pathways,
- an improvement of the tumor vessel efficiency.

Current point of view

If anti-angiogenesis treatments improve the tumor vessel efficiency, then they can help to bring chemical weapons inside the tumor.

"Anti-angiogenesis treatments help to fight tumors from inside".

Angiogenesis changes blood flow

Blood flow characterization

Blood flow evaluation is useful for:

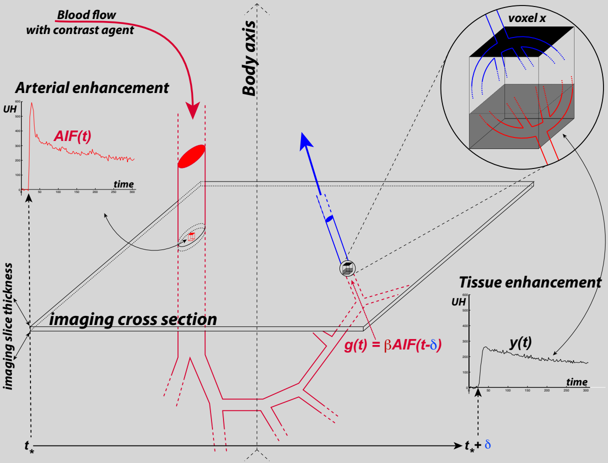
- *Early detection of metastases,*
- *Diagnostic after a stroke,*
- *Prognostic in cancer or after a stroke,*
- *Monitoring the efficiency of treatments, etc.*

Blood flow quantifications are needed to:

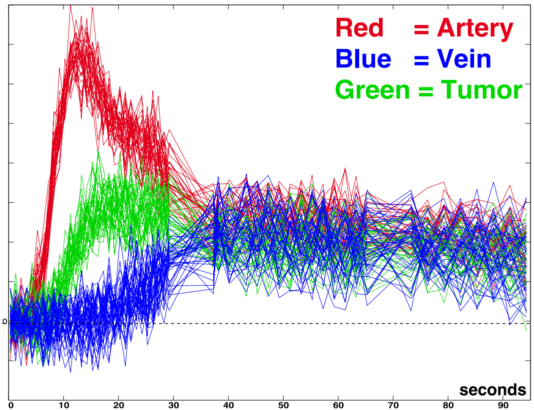
- *clarify medical decisions,*
- *help comparisons in longitudinal or multi-centric studies.*

Dynamical Contrast Enhanced Computed Tomography (DCE-CT) experiment

DCE-CT follow-up of contrast agent injection - about 30-40 images in 100 seconds



DCE-CT pre-processing: Rozenholc & Reiss (2011)



Typical enhancements

aorta (red), veinous (blue), tumor (green)

REMISCAN study

Medical and economic impact of the prediction of response to anti-angiogenic treatment in metastatic renal cell carcinoma by functional CT and functional MRI

Main goal: Define imaging biomarkers using functional CT and MRI to
- Optimize the selection of patients likely to benefit from anti-angiogenic drugs
- Adjust real-time processing to improve efficiency while limiting its side effects

Establishments: 16 French hospital centers
Disease: Metastatic Kidney Cancer
Patients: 100 patients with metastatic renal cell carcinoma under anti-angiogenic therapy
Coordinators: Profs. Cuenod (radiology) and Oudard (Oncology)
Funding: 1 M Euros

A nonparametric model for tissue microvascular circulation

- $AIF(t)$ number of particles inside the aorta at time t, $AIF(0) = 0$
- $\beta\ AIF(t-\delta)$ number of arrivals into voxel x at time t
- $q(t)$ number of particles in the voxel x at time t
- S_i *i.i.d.* sojourn times in the voxel x with *c.d.f.* F

Link between arrivals and sojourn times:

$$\begin{aligned}\mathbb{E}y(t) &= \underbrace{\int_0^{t-\delta} \beta\ AIF(t-\delta-z)\,dz}_{\text{arrived before time } t} - \underbrace{\int_0^{t-\delta} \beta\ AIF(t-\delta-z)\,P(S \le z)\,dz}_{\text{left before time } t} \\ &= \int_0^{t-\delta} \beta\ AIF(t-\delta-z)(1-F(z))dz\end{aligned}$$

The model for time transit distribution

This leads to the following model

$$q(t_i) = \int_0^{t_i-\delta} AIF(t_i - \delta - \tau)\beta\,(1 - F(\tau))d\tau, \quad i = 1, ..., n.$$

The value of delay δ can be measured with the small error using the decay between the jumps after the injection of the contrast agent inside the aorta and the tissue.

$f = \beta(1 - F)$ **is an unknown function to estimate**
$g = AIF(t)$ **can be estimated with relatively small error**
q **is measured with error** $\Longrightarrow$ **ill-posed problem**
Parameter β, **the blood flow**, is of **great interest to physicians**

Laplace deconvolution in the presence of noise

Discrete noisy version of Laplace convolution equation of the first kind :

$$q(t) = \int_0^t g(t-\tau) f(\tau) d\tau, \quad t \geq 0.$$

Observations:

$$y(t_i) = \int_0^{t_i} g(t_i - \tau) f(\tau) d\tau + \sigma \epsilon_i, \quad i = 1, ..., n,$$

where $0 \leq t_1 \leq ... \leq t_n \leq T_n$, ϵ_i are i.i.d. $N(0,1)$ and T_n may grow with the number of observations n.

Laplace deconvolution and Fourier deconvolution

Technically, Laplace deconvolution can be viewed as a particular case of Fourier deconvolution

$$y(t_i) = \int_{-\infty}^{\infty} g(t_i - \tau) f(\tau) d\tau + \sigma \epsilon_i, \quad i = 1, ..., n,$$

which has been extensively studied in the last thirty years

In reality, Laplace deconvolution is very different from Fourier deconvolution

- The problem is **harder than it seems**
- Fourier deconvolution approaches **do not work for Laplace deconvolution**.

Laplace deconvolution - challenges

Can Laplace convolution equation be solved by Fourier transform on an interval (FFT)?

$$y(t_i) = \int_0^{t_i} g(t_i - \tau) f(\tau) d\tau + \sigma \epsilon_i, \quad 0 \le t_1 \le \ldots \le t_n \le T_n$$

This is impossible:

- application of FFT **assumes periodicity** of f and g on $[0, T_n]$ which is **not true**
- application of FFT to Laplace convolution **does not result in the product of Fourier transforms** of f and g on $[0, T]$

Laplace deconvolution - challenges

Can Laplace convolution equation be solved by Fourier transform on a real line?

$$y(t_i) = \int_0^{t_i} g(t_i - \tau) f(\tau) d\tau + \sigma \epsilon_i, \quad 0 \le t_1 \le \ldots \le t_n \le T_n$$

This is problematic!

- there is **no data available for $t > T_n$**: if T_n **is not very large**, estimator

$$\hat{y}(\omega) = \sum_{j=1}^{n} e^{it_j \omega} y(t_j)$$

  of the Fourier transform of q may have **poor precision**
- This is **an artificial problem** since Laplace deconvolution has **causality** property:
- **$f(t)$ for $t < T$ depends only on $y(t)$ for $t < T$**

Laplace deconvolution - challenges

Can Laplace convolution equation can be solved using Laplace transform?

This is difficult:

- Inverse Laplace transform is usually founds using partial fractions, series expansions or Tables of inverse Laplace transforms, and is **not applicable with observational data**
- Finding inverse Laplace transform numerically is **often a more difficult task than solving a Laplace convolution equation**
- There is **no data available for** $t > T_n$.

Resolvent solution: noiseless case

Idea of the approach (no noise)

- Let $r \geq 1$ be such that

$$g^{(j)}(0) = \begin{cases} 0, & \text{if } j = 0, \ldots, r-2, \\ B_r \neq 0, & \text{if } j = r-1. \end{cases}$$

- Taking derivatives of both sides of the equation

$$q(t) = \int_0^t g(t-\tau) f(\tau) d\tau, \quad t \geq 0.$$

one obtains

$$\begin{aligned} q^{(j)}(t) &= \int_0^t g^{(j)}(t-\tau) f(\tau) d\tau, \quad j = 1, \ldots, r-1; \\ &\;\; \vdots \\ q^{(r)}(t) &= B_r f(t) + \int_0^t g^{(r)}(t-\tau) f(\tau) d\tau \end{aligned}$$

Laplace deconvolution with a simple kernel: resolvent solution

- There exists an **unique solution ϕ** of the equation

$$g^{(r)}(t) = B_r \phi(t) + \int_0^t g^{(r)}(t-\tau)\phi(\tau)d\tau$$

called a **resolvent of $g^{(r)}$** (Gripenberg, Londen & Staffans, 1990)

- Resolvent ϕ **can be recovered exactly as an inverse Laplace transform** of Φ, where

$$\Phi(s) = \frac{s^r G(s) - B_r}{s^r G(s)},$$

G is the Laplace transform of g, **known exactly**

- **There exists a unique solution of Laplace convolution equation**

$$f(t) = B_r^{-1} q^{(r)}(t) - B_r^{-1} \int_0^t q^{(r)}(t-\tau)\phi(\tau)d\tau.$$

Laplace deconvolution with a simple kernel: explicit solution

Let s_l be distinct zeros of $G(s)$ of orders α_l, respectively, $l = 1, ..., M$, $M < \infty$. Set $s_0 = 0$ and $\alpha_0 = r$. Then, f is of the form

$$f(t) = B_r^{-1} \left(q^{(r)}(t) - \sum_{j=0}^{r-1} b_j q^{(r-1-j)}(t) - \int_0^t q(t-x)\phi_1^{(r)}(x)dx \right),$$

where

$$\begin{aligned} \phi_1(x) &= \sum_{l=1}^{M} \sum_{j=0}^{\alpha_l - 1} \frac{a_{lj} x^j e^{s_l x}}{j!}, \quad b_j = a_{0j} + \sum_{l=1}^{M} \sum_{i=0}^{\min(j, \alpha_l - 1)} \binom{j}{i} a_{li} s_l^{j-i} \\ a_{lj} &= \frac{1}{(\alpha_l - 1 - j)!} \frac{d^{\alpha_l - j - 1}}{ds^{\alpha_l - j - 1}} \left[(s - s_l)^{\alpha_l} \Phi(s) \right] \Big|_{s = s_l}, \end{aligned}$$

Adaptive estimation

- **In a general case**, estimate f by

$$\hat{f}(t) = B_r^{-1}\widehat{q^{(r)}}(t) - B_r^{-1}\int_0^t \widehat{q^{(r)}}(t-\tau)\phi(\tau)d\tau$$

  or, **if the number of zeros of $G(s)$ is finite**,

$$\hat{f}(t) = B_r^{-1}\left(\widehat{q^{(r)}}(t) - \sum_{j=0}^{r-1} b_j \widehat{q^{(r-1-j)}}(t) - \int_0^t \hat{q}(t-x)\phi_1^{(r)}(x)dx\right),$$

  where $\widehat{q^{(l)}}(t)$ are estimators of derivatives $q^{(l)}(t)$, $l = 0, \cdots, r$.
- **Abramovich, Pensky& Rozenholc (2013):** kernel estimators with adaptive choice of bandwidth using Lepskii method (Lepski (1991), Lepski, Mammen and Spokoiny (1997))
- The estimator is **minimax optimal** under additional assumptions on the model

Merits and defects of the approach

Merits of the method

- Reduces the problem to a **well-known problem**
- Allows to write an estimator in an **explicit form**
- Allows to apply a **variety of techniques** for estimating regression function and its derivatives
- Produces **asymptotically optimal estimators**

Defects of the method

- **Requires knowledge of kernel** g in general and **parameters** r **and** B_r in particular (hard to use in applications)
- Produces **boundary effects**: does not allow to estimate β, the blood flow

Laplace deconvolution with the data-driven kernel: Laguerre functions approach

Consider **a system of Laguerre functions** defined as

$$\phi_k(t) = \sqrt{2a} e^{-at} L_k(2at), \quad k = 0, 1, \cdots,$$

where $L_k(t)$ are Laguerre polynomials

$$L_k(t) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} \frac{t^j}{j!}, \quad t \geq 0.$$

- Functions $\phi_k(t)$, $k = 0, 1, \cdots$, form **an orthonormal basis** of the $L^2(0, \infty)$ space
- Let $f^{(k)}$, $g^{(k)}$, $q^{(k)}$ and $y^{(k)}$, $k = 0, \cdots, \infty$, be Laguerre coefficients of functions $f(x)$, $g(x)$, $q(t)$ and $y(t)$, respectively.

Relations between Laguerre coefficients

By plugging expansions of $f(x)$, $g(x)$ and $q(t)$ into the Laplace convolution equation

$$q(t) = \int_0^t g(t-\tau) f(\tau) d\tau$$

obtain

$$\sum_{k=0}^{\infty} q^{(k)} \phi_k(t) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} f^{(k)} g^{(j)} \int_0^t \phi^{(k)}(x) \phi^{(j)}(t-x) dx.$$

Using identity

$$\int_0^t \phi_k(x) \phi_j(t-x) dx = (2a)^{-1/2} \left[\phi_{k+j}(t) - \phi_{k+j+1}(t) \right],$$

equation above can be re-written as

$$\sum_{k=0}^{\infty} q^{(k)} \phi_k(t) = (2a)^{-1/2} \sum_{k=0}^{\infty} \phi_k(t) [f^{(k)} g^{(0)} + \sum_{l=0}^{k-1} (g^{(k-1)} - g^{(k-l-1)}) f^{(l)}].$$

System of linear equations

- Equating coefficients for each basis function, obtain **an infinite triangular system of linear equations**.
- Define $f_m(x) = \sum_{k=0}^{m-1} f^{(k)} \phi_k(x) \implies$ Need m equations to recover coefficients $f^{(k)}$, $k = 0, \cdots, m-1$.

Lemma. Let $\mathbf{f}_m$, $\mathbf{g}_m$ and $\mathbf{q}_m$ be m-dimensional vectors with elements $f^{(k)}$, $g^{(k)}$ and $q^{(k)}$, $k = 0, 1, \cdots, m-1$, respectively. Then, for any m, one has

$$\mathbf{q}_m = \mathbf{G}_m \mathbf{f}_m$$

where $\mathbf{G}_m$ is **the lower triangular Toeplitz matrix** with elements

$$G^{(ji)} = \begin{cases} (2a)^{-1/2}\, g^{(0)}, & \text{if } \ i = j, \\ (2a)^{-1/2}\, (g^{(j-i)} - g^{(j-i-1)}), & \text{if } \ i < j, \\ 0, & \text{if } \ i > j, \end{cases}$$

Evaluating Laguerre coefficients from discrete noisy data

Recall: one observes $y(t_i) = \int_0^{t_i} g(t_i - \tau) f(\tau) d\tau + \sigma \epsilon_i, \quad i = 1, \cdots, n$, with measurements $y(t_i)$ taken at points $0 \le t_1 \le \ldots \le t_n \le T < \infty$, where both n and T are large.

- **Choose large M**, so **the bias** in representation of f by f_M **is very small**.
- Form an $(n \times M)$ matrix $\mathbf{\Phi}_M$ with elements $\Phi^{(i,k)} = \phi_k(t_i)$, $i = 1, \cdots, n$, $k = 0, \cdots, M-1$.
- Let $\vec{y}$ be n-dimensional vector with components $y(t_i)$ $i = 1, \cdots, n$.
- Let $\mathbf{J}_{m,M} = (\mathbf{I}_m \quad \mathbf{0}_{m,M-m})$ be the $m \times M$ matrix which has the $m \times m$ identity matrix $\mathbf{I}_m$ as its first m columns and the rest of the columns are equal to zero

Non-adaptive estimation

Construction:

Denote the initial estimator of vector $\mathbf{q}_M$ by

$$\widehat{\mathbf{q}_M} := (\Phi_M^T \Phi_M)^{-1} \Phi_M^T \vec{y}$$

Consider a collection $\{1, \ldots, M\}$ of indices and define **a collection of estimators** for all m, $1 \le m \le M$, given by

$$\widehat{\mathbf{f}_m} := \mathbf{G}_m^{-1} \mathbf{J}_{m,M} \widehat{\mathbf{q}_M} = \mathbf{G}_m^{-1} \mathbf{J}_{m,M} (\Phi_M^T \Phi_M)^{-1} \Phi_M^T \vec{y}$$

Motivation:

- Use **regression approach**, not numerical integration for estimating the Laguerre coefficients (the bias depends on M, not T)
- Do not need to **re-fit** estimator $\widehat{\mathbf{q}_M}$ for every model size m

The oracle risk

Denote

$$\mathbf{Q}_m = \frac{n}{T}[(\Phi_M^T \Phi_M)^{-1}]_m([\mathbf{G}_M \mathbf{G}_M^T]_m)^{-1}$$

If T and n are large, T/n is small, then $\mathbf{\Phi}_M^T \mathbf{\Phi}_M \approx n T^{-1} \mathbf{I}$.

Define the set of indices $\mathcal{M}_n = \mathcal{M}_n(\aleph) = \{1, \ldots, M_0\}$ such that condition $T\, n^{-1} \mathrm{Tr}(\mathbf{Q}_m) \le \aleph$, holds for all $m \le M_0$.
The smallest possible risk, **the oracle risk**, is given by

$$R_{oracle} = \min_{m \in \mathcal{M}_n(\aleph)} \left[\|f_m - f\|_2^2 + \sigma^2 T n^{-1}\ \mathrm{Tr}(\mathbf{Q}_m) \right].$$

Model selection using complexity penalty

Denote

$$\mathbf{A}_m = \sqrt{\frac{n}{T}}\, \mathbf{G}_m^{-1} \mathbf{J}_{m,M} (\Phi_M^T\, \Phi_M)^{-1} \Phi_M^T$$

$$v_m^2 = \|\mathbf{A}_m\|_2^2 = \mathrm{Tr}(\mathbf{Q}_m), \quad \rho_m^2 = \rho^2(\mathbf{A}_m).$$

Recall that

$$\widehat{\mathbf{f}_m} = \sqrt{T/n}\ \mathbf{A}_m \vec{y}.$$

For any constant $B > 0$ introduce a **penalty**

$$\mathrm{pen}(m) = 4\sigma^2 T n^{-1} \left[(1+B) v_m^2 + (1+B^{-1}) \rho_m^2 \log(m^2 \rho_m^2)\right].$$

Choose **model size**

$$\widehat{m} = \arg\min \left\{ m \in \mathcal{M}_n : \ -\|\widehat{\mathbf{f}_m}\|^2 + \mathrm{pen}(m) \right\}$$

Risk of penalized estimator

If $M = n^{2/3}$ and $q(t) = \int_0^t g(t-\tau) f(\tau) d\tau$ is sufficiently smooth, then

$$\mathbb{E}(\|\hat{f}_{\widehat{m}} - f\|_2^2) \leq C \left[\log(m_0^2 \rho_{m_0}^2) R_{oracle} + \sigma^2 \frac{T}{m_0 n} + \frac{a}{n} \right],$$

where a is the scaling parameter of the Laguerre functions and $m_0 = m_0(n, T)$ is the value of m which minimizes the right-hand side of the oracle inequality.

The penalized estimator $\hat{f}_{\widehat{m}}$ has the risk within $\log(m_0^2 \rho_{m_0}^2)$ of the oracle risk **under almost no assumptions on the model**

With some **extra assumptions**, one can show that the estimator is **minimax optimal within a logarithmic factor** of n

Merits and defects of the approach

Advantages of the method

- **Does not require exact knowledge** of r and **exact representation of the kernel** g
- **Computationally very easy and fast:** solution of a relatively small system of equations. The system is **triangular with a Toeplitz matrix**.
- **No boundary effects** due to extension at zero and cut-off at T.
- Presence of an extra parameter a **allows to control the error of approximation** of both q and f.

Shortcomings of the method

- **Does not work well** for functions which **do not have simple representation via Laguerre functions basis**
- **Hard to enforce** positivity or shape restrictions

General approach

Consider a **general ill-posed linear inverse problem**

$$y_i = q(t_i) + \xi_i \quad \text{with} \quad q(t) = (Qf)(t) = \int_a^b u(t,\tau) f(\tau) d\tau, \;\; t \in [c,d]$$

Q **does not have a bounded inverse**
ξ_i are **i.i.d.centered sub-gaussian random variables**
$\mathbb{E}\xi_i = 0, \quad \mathbb{P}\left(|\xi_i| > t\right) \le \exp(-t^2/2\sigma^2)$
Our particular case is **a discrete noisy version of Laplace convolution equation** of the first kind where $f \in L^2[0,T]$, $q \in L^2[0,T]$

$$u(t,\tau) = g(t-\tau); \quad g(z) = AIF(z) \;\text{ if }\; z \ge 0; \quad g(z) = 0 \;\text{ otherwise}$$

$0 \le t_1 \le \ldots \le t_n \le T$ are observation times
$n = 90$ for **DCE-MRI** data and $n = 23$ for **DCE-CT** data
$f(\tau) = \beta(1 - F(\tau)$ is a **nonnegative decreasing function**

Application of overcomplete dictionaries

Consider a **dictionary** $\{\varphi_j, j = 1, \cdots, p\}$ such that functions φ_j are **linearly independent** and $\|\varphi_j\|_2 = 1$. For any $\mathbf{t} \in \mathbb{R}^p$, denote

$$f_{\mathbf{t}}(z) = \sum_{j=1}^{p} t_j \varphi_j(z).$$

Estimate f by $f_{\widehat{\boldsymbol{\theta}}}$

If function f **were known**, we would search for the vector of coefficients $\boldsymbol{\theta}$ as a solution of $\boldsymbol{\theta} = \arg\min_{\mathbf{t}} \|f - f_{\mathbf{t}}\|^2$ where the contrast is

$$\|f - f_{\mathbf{t}}\|_2^2 = \|f\|^2 + \|f_{\mathbf{t}}\|^2 - 2\sum_{j=1}^{p} \langle f, \varphi_j \rangle t_j.$$

Here f is independent of $\mathbf{t}$ and $f_{\mathbf{t}}$ is completely known

Need to estimate the last term only

Lasso solution

Let Q^* be **the conjugate operator** for Q:
$(Q^* u)(z) = \int_z^T g(x-z)u(x)dx, \quad 0 \le z \le T.$
$\langle Qf, u \rangle = \langle f, Q^* u \rangle$ for any $f, u \in L^2[0,T]$
Assumption A0 There exist ψ_j such that

$$(Q^* \psi_j)(z) = \int_z^T g(x-z)\psi_j(x)dx = \varphi_j(z) \quad \text{and} \quad \|\psi_j\|_2 < \infty$$

$\|\psi_j\|_2$ is the **"price"** of using dictionary element φ_j

Replace $\beta_j = \langle f, \varphi_j \rangle = \langle q, \psi_j \rangle$ by its estimator $\widehat{\beta}_j$ where

$$\widehat{\beta}_j = \frac{1}{n} \sum_{i=1}^{n} y_i \psi_j(x_i) \Delta x_i, \quad \nu_j^2 = \frac{T^2}{n} \sum_{i=1}^{n} \psi_j^2(x_i).$$

Lasso solution

The estimation procedure

Let $\boldsymbol{\Phi}$ be the matrix with elements $\boldsymbol{\Phi}_{kj} = \langle \varphi_k, \varphi_j \rangle$ and $\mathbf{W}^T \mathbf{W} = \boldsymbol{\Phi}$. Estimate f by $f_{\widehat{\boldsymbol{\theta}}}$, where $\widehat{\boldsymbol{\theta}}$ is the solution of

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\mathbf{t}} \left\{ \mathbf{t}^T \boldsymbol{\Phi} \mathbf{t} - 2 \sum_{j=1}^{p} \widehat{\beta}_j t_j + \alpha \sum_{j=1}^{p} \nu_j |t_j| \right\}$$

or

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\mathbf{t}} \left\{ \|\mathbf{W}\mathbf{t} - \gamma\|_2^2 + \alpha \|\boldsymbol{\Upsilon}\mathbf{t}\|_1 \right\}$$

Here, $\gamma = (\mathbf{W}\mathbf{W}^T)^{+} \mathbf{W} \widehat{\beta}$ and $\boldsymbol{\Upsilon}$ is the diagonal matrix $\boldsymbol{\Upsilon} = \operatorname{diag}(\nu_1, \cdots, \nu_p)$.

Advantage: Lasso is used in a **prediction set up** where it requires much milder conditions

Lasso: slow error rates

Let $\mathcal{P} = \{1, \cdots, p\}$ and $\alpha \geq \sqrt{2\, n^{-1} \sigma\, (\tau + 1) \log p}$.
Then, under **no restriction on the dictionary**, with probability at least $1 - 2p^{-\tau}$

$$\|f_{\widehat{\boldsymbol{\theta}}} - f\|_2^2 \leq \inf_{\mathbf{t}} \left[\|\mathbf{W}\mathbf{t} - f\|_2^2 + 4\alpha \sum_{j=1}^{p} \nu_j |t_j| \right]$$

These are the **"slow" Lasso error rates**

Under a compatibility condition for matrix $\mathbf{W}$ obtain **"fast" Lasso error rates**

Advantage: Matrix $\mathbf{W}$ is **the dictionary-generated matrix** and can be chosen to **satisfy compatibility conditions** (e.g., restricted isometry condition)

Sharp oracle inequality: fast rates

Let $\boldsymbol{\Phi}_J$ be a reduction of Gram matrix $\boldsymbol{\Phi}$ to a set of functions $\{\varphi_j, j \in J\}$, and $\mathcal{L}_J = \mathrm{Span}\,\{\varphi_j, j \in J\}$ and compatibility condition hold
Then, with high probability,

$$\|f_{\widehat{\boldsymbol{\theta}}} - f\|_2^2 \le \min_{J \subseteq \mathcal{P}} \left\{ \|f - f_{\mathcal{L}_J}\|_2^2 + C\, \frac{\sigma^2 \log p}{n\lambda_{\min}(\boldsymbol{\Phi}_J)} \sum_{j \in J} \nu_j^2 \right\},$$

If f were known, one would choose $J \subseteq \mathcal{P}$ and recover f as its estimated projection $\hat{f}$ on $\mathcal{L}_J$ with the error

$$\mathbb{E}\|\hat{f} - f\|_2^2 \ge \min_{J \subset \mathcal{P}} \left\{ \|f - f_{\mathcal{L}_J}\|_2^2 + \frac{\sigma^2}{n\lambda_{\min}(\boldsymbol{\Phi}_J)} \sum_{j \in J} \nu_j^2 \right\}$$

Hence, **up to a** $\log p$ factor, the estimator $f_{\widehat{\boldsymbol{\theta}}}$ attains **the minimum possible** mean squared error **for a particular function of interest** f
Here, $\log p$ is **the "price" for choosing an appropriate subset of dictionary functions**

Application to DCE imaging data

- We used a dictionary constructed of **Gamma cdfs**

$$\varphi_k(x) = GammaCDF(x; a_k, b_k), \quad k = 1, \cdots, p,$$

  Here (a_k, b_k) take values on the Cartesian product of $(0, 1, \cdots, m_1 - 1)$ and $(\Delta, 2\Delta, \cdots, m_2\Delta)$ with Δ fixed in advance.

- In order to form matrix $\mathbf{W}$ with columns φ_j, $j = 1, \cdots, p$, dictionary functions were evaluated on a fine grid
- We assumed that the errors are **Gaussian**, so that the error vector $\boldsymbol{\xi} \in \mathbb{R}^n$ is Gaussian $\mathcal{N}(0, \sigma^2 \mathbf{I})$
- Matrix $\mathbf{Q}$ was constructed so that it carried out numerical integration for $0 \leq x \leq T$, i.e. $q = Qf$

Estimation on the basis of DCE-CT data

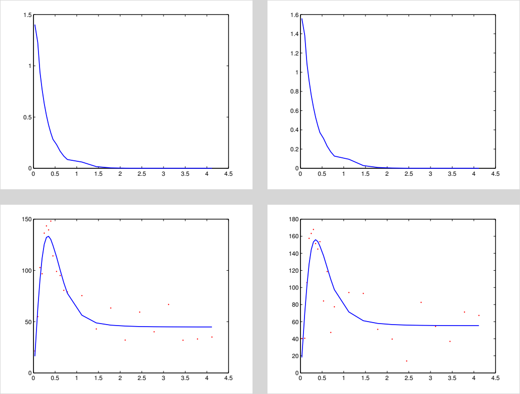


Figure: **Top row:** estimators $\widehat{f}(t)$ of $f(t) = \beta(1 - F(t))$ for two different tissue voxels on the basis of DCE-CT data. **Bottom row:** observations of tissue enhancements (red dots), the recovered values of $\widehat{q}(t) = \int \mathrm{AIF}(t-\tau)\widehat{f}(\tau)d\tau$ (blue line). The number of observations $n = 23$.

Merits and defects of the approach

Advantages of the method

- **Does not require exact knowledge** of r and **exact representation of the kernel** g
- **No boundary effects** due to extension at zero and cut-off at T.
- The dictionary is **extremely flexible**. Can be comprised of kernel functions, splines, frames, etc.
- **Easy to enforce** positivity or shape restrictions
- Works very well when **the noise level is high and the sample size is small**

Shortcomings of the method

- **Computationally is not as easy as Laguerre functions method:** need to evaluate inverse images of the dictionary functions (although this step can be **done in advance**) and solve the optimization problem. **Easier than the resolvent solution**
- The compatibility condition that guarantees fast error rates is **hard to check**

References

- **Abramovich, F., Pensky, M., Rozenholc, Y.** (2013) Laplace deconvolution with noisy observations. *Elect. Journ. Statist.*, **7**, 1094-1128
- **Comte, F., Cuenod, C.-A., Pensky, M., Rozenholc, Y.** (2016) Laplace deconvolution and its application to Dynamic Contrast Enhanced imaging. *Journ. R. Stat. Soc., Ser. B*, accepted.
- **Pensky, M.** (2016) Solution of linear ill-posed problems using overcomplete dictionaries. *Ann. Statist.*, accepted.
- **Rozenholc, Y., and Reiss, M.** (2012) Preserving time structures while denoising a dynamical image, *Mathematical Methods for Signal and Image Analysis and Representation (Chapter 12)*, Florack, L. and Duits, R. and Jongbloed, G. and van Lieshout, M.-C. and Davies, L. Ed., Springer-Verlag, Berlin.
- **Lieury, T., Pouzat, C. and Rozenholc, Y.** (2012) DynClust: non-parametric denoising and clustering method of noisy images both indexed by time and space, *R-package* available at `http://cran.r-project.org/`.