Lausanne, August 19, 2004

Dear Dr. Liebling,

I am pleased to inform you that you were selected to receive the 2004 Research Award of the Swiss Society of Biomedical Engineering for your thesis work "On Fresnelets, interference fringes, and digital holography". The award will be presented during the general assembly of the SSBE, September 3, Zurich, Switzerland.

Please let us know if
1) you will be present to receive the award,
2) you would be willing to give a 10 minutes presentation of the work during the general assembly.

The award comes with a cash prize of 1000.- CHF. Would you please send your banking information to the treasurer of the SSBE, Uli Diermann (Email: uli.diermann@bfh.ch), so that he can transfer the cash prize to your account?

I congratulate you on your achievement.

With best regards,

Michael Unser, Professor
Chairman of the SSBE Award Committee

cc: Ralph Mueller, president of the SSBE; Uli Diermann, treasurer

Dr. Michael Liebling
Biological Imaging Center
California Inst. of Technology
Mail Code 139-74
Pasadena, CA 91125, USA

BIOMEDICAL IMAGING GROUP (BIG)
LABORATOIRE D’IMAGERIE BIOMEDICALE
EPFL LIB
Bât. BM 4.127
CH 1015 Lausanne
Switzerland

Deep splines

Michael Unser
Biomedical Imaging Group
EPFL, Lausanne
Switzerland

BIG’s 20th Birthday, March 23, 2018, EPFL, Switzerland

Basic problem: how to interpolate data

\[(x_m, y_m), \quad m = 1, \ldots, M\]

B-spline expansion: \[f(x) = \sum_{m=1}^{M} y_m \varphi_m(x)\]
20 years of splines and biomedical imaging

<table>
<thead>
<tr>
<th>Image processing task</th>
<th>Specific operation</th>
<th>Imaging modality</th>
</tr>
</thead>
</table>
| Tomographic reconstruction | • Filtered backprojection  
• Fourier reconstruction  
• Iterative techniques  
• 3D + time | Commercial CT (X-rays)  
PET, SPECT  
Dynamic CT, SPECT, PET |
| Sampling grid conversion | • Polar-to-cartesian coordinates  
• Spiral sampling  
• k-space sampling  
• Scan conversion | Ultrasound (endovascular)  
Spiral CT, MRI  
MRI |
| Visualization | 2D operations  
• Zooming, panning, rotation  
• Re-sizing, scaling | All |
| | • Stereo imaging  
• Range, Topography | Fundus camera  
OCT |
| | 3D operations  
• Re-slicing  
• Max. intensity projection  
• Simulated X-ray projection | CT, MRI, MRA |
| | Surface/volume rendering  
• Iso-surface ray tracing  
• Gradient-based shading  
• Stereogram | CT  
MRI |
| Geometrical correction | • Wide-angle lenses  
• Projective mapping  
• Aspect ratio, tilt  
• Magnetic field distortions | Endoscopy  
C-Arm fluoroscopy  
Dental X-rays  
MRI |
| Registration | • Motion compensation  
• Image subtraction  
• Mosaicking  
• Correlation-averaging  
• Patient positioning  
• Retrospective comparisons  
• Multi-modality imaging  
• Stereotactic normalization  
• Brain warping | IMRI, fundus camera  
DSA  
Endoscopy, fundus camera, EM-microscopy  
Surgery, radiotherapy  
CT/PET/MRI |
| Feature detection | • Contours  
• Rods  
• Differential geometry | All |
| | Contour extraction  
• Snakes and active contours | MRI, Microscopy (cytology) |
Splines as a unifying mathematical concept

**Functional analysis**

- Approximation theory
- Wavelet theory
- Partial differential equations
- Regularization theory
- Machine learning
- Stochastic processes
- Signal processing
- Numerical analysis

**Equivalent variational formulation**

\[ f_{\text{spline}} = \arg \min_{f \in \mathcal{H}^1(\mathbb{R})} \| Df \|_{L_2}^2 \quad \text{s.t.} \quad f(x_m) = y_m, (m = 1, \ldots, M) \]
Splines, operators and (sparse) innovations

$L\{\cdot\}$: (quasi)-invertible differential operator (translation-invariant)

$\delta$: Dirac distribution

**Definition**

The function $s(x)$, $x \in \mathbb{R}^d$ (possibly of slow growth) is a **nonuniform L-spline** with knots $\{x_k\}_{k \in S}$

\[
Ls = \sum_{k \in S} a_k \delta(\cdot - x_k) = w
\]

Spline theory: (Schultz-Varga, 1967; Jerome-Schumaker 1969; Micchelli, 1976)

- Splines are inherently sparse (with a finite rate of innovation)
  - Location of singularities (knots): $\{x_k\}$
  - Strength of singularities (linear weights): $\{a_k\}$

**Formal spline synthesis**

$L$: spline admissible operator (LSI)

- Finite-dimensional null space: $\mathcal{N}_L = \text{span}\{p_n\}_{n=1}^{N_0}$
- Green’s function $\rho_L : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $L\{\rho_L\} = \delta$

Spline’s innovation:

\[
w = \sum_k a_k \delta(\cdot - x_k)
\]

\[
s(x) = \sum_k a_k \rho_L(x - x_k) + \sum_{n=1}^{N_0} b_n p_n(x)
\]

Requires specification of boundary conditions
\[ \rho_{L^*L}(x) = (L^*L)^{-1}\{\delta\}(x): \text{Green's function of } (L^*L) \]

\textit{(Schoenberg 1964, de Boor-Lynch 1966, Kimeldorf-Wahba 1971)}

\[ \left( P_2 \right) \quad \arg \min_{f \in \mathcal{H}_L} \left( \sum_{m=1}^{M} |y_m - f(x_m)|^2 + \lambda ||Lf||^2_{L^2(\mathbb{R}^d)} \right) \]

\[ f(x) = \sum_{m=1}^{M} a_m \rho_{L^*L}(x - x_m) + \sum_{n=1}^{N_0} b_n p_n(x); \]

\textit{i.e., it is a } (L^*L) \text{-spline with knots at the } \{x_m\}. \]

Example: \( L = D^2 \) with \( \rho_{D^4}(x) \propto |x|^3 \) \( \Rightarrow \) \( f(x) \) is a cubic spline
RKHS representer theorem for machine learning

\[(P2) \quad \arg \min_{f \in \mathcal{H}} \left( \sum_{m=1}^{M} |y_m - f(x_m)|^2 + \lambda \|f\|_H^2 \right) \quad \text{(Poggio-Girosi 1990)}\]

\(r_\mathcal{H} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R} \) is the (unique) reproducing kernel for the Hilbert \(\mathcal{H}\) if

- \(r_\mathcal{H}(x_0, \cdot) \in \mathcal{H}\) for all \(x_0 \in \mathbb{R}^d\)
- \(f(x_0) = \langle r_\mathcal{H}(x_0, \cdot), f \rangle_\mathcal{H}\) for all \(f \in \mathcal{H}\) and \(x_0 \in \mathbb{R}^d\)

Convex loss function: \(F : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}\)

\[(P2') \quad \arg \min_{f \in \mathcal{H}} \left( F(y, f) + \lambda \|f\|_H^2 \right) \quad \text{(Schölkopf-Smola 2001)}\]

Representer theorem for \(L_2\)-regularization

The generic parametric form of the solution of (P2') is

\(f(x) = \sum_{m=1}^{M} a_m r_\mathcal{H}(x, x_m)\)

Supports the theory of SVM, kernel methods, etc.

Convex loss function: \(F : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}\)

Sample values: \(f = (f(x_1), \ldots, f(x_M))\)

\(r_\mathcal{H}(x, x_m) = \rho_{\mathcal{L}^2} (x - x_m)\)

Link with splines:

Machine learning: can we do better?

Emergence of deep learning


Deep ConvNets and biomedical imaging

(Jin et al. 2016; Chen et al. 2017; ...)

- CT reconstruction based on Deep ConvNets
  - Input: Sparse view FBP reconstruction
  - Training: Set of 500 high-quality full-view CT reconstructions
  - Architecture: U-Net with skip connection (Jin et al., IEEE TIP 2017)

Dose reduction by 7: 143 views

CT data

Ground truth

FBP

SNR 24.06

Feedforward deep neural network

- Layers: $\ell = 1, \ldots, L$
- Deep structure descriptor: $(N_0, N_1, \cdots, N_L)$
- Neuron or node index: $(n, \ell), \ n = 1, \cdots, N_\ell$
- Activation function: $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ (ReLU)

- Linear step: $\mathbb{R}^{N_{\ell-1}} \rightarrow \mathbb{R}^{N_\ell}$
  \[ f_\ell : x \mapsto f_\ell(x) = W_\ell x + b_\ell \]

- Nonlinear step: $\mathbb{R}^{N_\ell} \rightarrow \mathbb{R}^{N_\ell}$
  \[ \sigma_\ell : x \mapsto \sigma_\ell(x) = (\sigma(x_1), \ldots, \sigma(x_{N_\ell})) \]

\[ f_{\text{deep}}(x) = (\sigma_L \circ f_L \circ \sigma_{L-1} \circ \cdots \circ \sigma_2 \circ f_2 \circ \sigma_1 \circ f_1)(x) \]

Refinement: free-form activation functions

- Layers: $\ell = 1, \ldots, L$
- Deep structure descriptor: $(N_0, N_1, \cdots, N_L)$
- Neuron or node index: $(n, \ell), \ n = 1, \cdots, N_\ell$
- Free-form activation functions: $\sigma_{n,\ell} : \mathbb{R} \rightarrow \mathbb{R}$

- Linear step: $\mathbb{R}^{N_{\ell-1}} \rightarrow \mathbb{R}^{N_\ell}$
  \[ f_\ell : x \mapsto f_\ell(x) = W_\ell x + b_\ell \]

- Nonlinear step: $\mathbb{R}^{N_\ell} \rightarrow \mathbb{R}^{N_\ell}$
  \[ \sigma_\ell : x \mapsto \sigma_\ell(x) = (\sigma_{n,\ell}(x_1), \ldots, \sigma_{N_\ell,\ell}(x_{N_\ell})) \]

\[ f_{\text{deep}}(x) = (\sigma_L \circ f_L \circ \sigma_{L-1} \circ \cdots \circ \sigma_2 \circ f_2 \circ \sigma_1 \circ f_1)(x) \]
Linear interpolation: can we do better?

Sparsity-promoting regularization: \( TV^{(2)}(f) \overset{\Delta}{=} \|D^2 f\|_{\mathcal{M}} = \sup_{\varphi \in \mathcal{S}(\mathbb{R}) : \|\varphi\|_{\infty} \leq 1} (D^2 f, \varphi) \)

\[
\begin{align*}
  f_{\text{sparse}} &= \arg \min_{f \in BV^{(2)}(\mathbb{R})} TV^{(2)}(f) \quad \text{s.t.} \quad f(x_m) = y_m, (m = 1, \ldots, M)
\end{align*}
\]


Representer theorem for deep neural networks


- neural network \( f : \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_L} \) with deep structure \((N_0, N_1, \ldots, N_L)\)
  \[ x \mapsto f(x) = (\sigma_L \circ \ell_L \circ \sigma_{L-1} \circ \cdots \circ \ell_2 \circ \sigma_1 \circ \ell_1)(x) \]
- normalized linear transformations \( \ell_\ell : \mathbb{R}^{N_{\ell-1}} \rightarrow \mathbb{R}^{N_{\ell}}, x \mapsto U_\ell x \) with weights \( U_\ell = [u_{1,\ell}, \ldots, u_{N_\ell,\ell}]^T \in \mathbb{R}^{N_\ell \times N_{\ell-1}} \) such that \( \|u_{n,\ell}\| = 1 \)
- free-form activations \( \sigma_\ell = (\sigma_{1,\ell}, \ldots, \sigma_{N_\ell,\ell}) : \mathbb{R}^{N_{\ell}} \rightarrow \mathbb{R}^{N_{\ell}} \) with \( \sigma_{1,\ell}, \ldots, \sigma_{N_\ell,\ell} \in BV^{(2)}(\mathbb{R}) \)

Given a series data points \((x_m, y_m)\) \(m = 1, \ldots, M\), we then define the training problem

\[
\arg \min_{(U_{\ell}), (\sigma_{n,\ell} \in BV^{(2)}(\mathbb{R}))} \left( \sum_{m=1}^{M} E(y_m, f(x_m)) + \mu \sum_{\ell=1}^{N} R_\ell(U_\ell) + \lambda \sum_{\ell=1}^{L} \sum_{n=1}^{N_{\ell}} TV^{(2)}(\sigma_{n,\ell}) \right) (1)
\]

- \( E : \mathbb{R}^{N_L} \times \mathbb{R}^{N_L} \rightarrow \mathbb{R}^+ \): arbitrary convex error function
- \( R_\ell : \mathbb{R}^{N_{\ell} \times N_{\ell-1}} \rightarrow \mathbb{R}^+ \): convex cost

If solution of (1) exists, then it is achieved by a deep spline network with activations of the form

\[
\sigma_{n,\ell}(x) = b_{1,\ell,\cdot} + b_{2,\ell,\cdot} x + \sum_{k=1}^{K_{n,\ell}} a_{k,n,\ell} (x - \tau_{k,n,\ell}) + ,
\]

with adaptive parameters \( K_{n,\ell} \leq M - 2 \), \( \tau_{1,n,\ell}, \ldots, \tau_{K_{n,\ell},n,\ell} \in \mathbb{R} \), and \( b_{1,\ell,\cdot}, b_{2,\ell,\cdot}, a_{1,\ell,\cdot}, \ldots, a_{K_{n,\ell},\ell,\cdot} \in \mathbb{R} \).
Deep spline networks: Discussion

- Global optimality achieved with **spline activations**

- State-of-the-art ReLU networks \( (K_{n,\ell} = 1, b_{n,\ell} = 0) \)

  - No need to normalize:
    \[
    (w_{n,\ell}^T x - z_{n,\ell})_+ = (a_{n,\ell} u_{n,\ell}^T x - z_{n,\ell})_+ = a_{n,\ell} (u_{n,\ell}^T x - \tau_{n,\ell})_+ \]

- Backward compatibility

  - Linear regression: \( \lambda \to \infty \Rightarrow K_{n,\ell} = 0 \)

- Key features
  - Direct control of complexity (number of knots): adjustment of \( \lambda \)
  - Ability to suppress unnecessary layers

- Generalizations
  - Broad family of cost functionals
  - Cases where a subset of network components is fixed

CONCLUSION

- **Splines: a unifying mathematical framework**
  - Link between continuous and discrete theories
  - Applicable to many areas of science and engineering

- **A powerful set of tools**
  - B-splines, etc.
  - Best cost/performance tradeoff
  - Optimality and universality

- **An endless source of inspiration**

- **Current frontiers**
  - Non-linear algorithms, optimization in Banach space
  - Sparsity
  - Deep learning