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

Tutorial Session, European Molecular Imaging Meeting (EMIM’17), 5-7 April 2017, Köln, Germany

OUTLINE

1. Imaging as an inverse problem
   - Basic imaging operators
   - Comparison of modalities
   - Discretization of the inverse problem

2. Classical reconstruction algorithms
   - Backprojection
   - Tikhonov regularization
   - Wiener / LMSE solution

3. Modern methods: the sparsity (re)evolution
   - Specific examples: Magnetic resonance imaging
     Computed tomography
     Differential phase-contrast tomography

4. What’s next: the learning revolution?
Inverse problems in bio-imaging

- Linear forward model
  \[ y = Hs + n \]

Problem: recover \( s \) from noisy measurements \( y \)

Basic limitations

1) Inherent noise amplification
2) Difficulty to invert \( H \) (too large or non-square)
3) All interesting inverse problems are **ill-posed**

Part 1:

Setting up the problem
Forward imaging model (noise-free)

Unknown molecular/anatomical map: \( s(r), r = (x, y, z, t) \in \mathbb{R}^d \)

*defined over a continuum in space-time*
\[ s \in L_2(\mathbb{R}^d) \quad \text{(space of finite-energy functions)} \]

Imaging operator \( H : s \mapsto y = (y_1, \ldots, y_M) = H\{s\} \)

*from continuum to discrete (finite dimensional)*
\[ H : L_2(\mathbb{R}^d) \to \mathbb{R}^M \]

Linearity assumption: for all \( s_1, s_2 \in L_2(\mathbb{R}^d), \alpha_1, \alpha_2 \in \mathbb{R} \)
\[ H\{\alpha_1 s_1 + \alpha_2 s_2\} = \alpha_1 H\{s_1\} + \alpha_2 H\{s_2\} \]

\[ \Rightarrow \quad [y]_m = y_m = \langle \eta_m, s \rangle = \int_{\mathbb{R}^d} \eta_m(r)s(r)dr \]
(by the Riesz representation theorem)

Images are obviously made of sine waves ...
Basic operator: Fourier transform

\[ \mathcal{F} : L^2(\mathbb{R}^d) \rightarrow L^2(\mathbb{R}^d) \]

\[ \hat{f}(\omega) = \mathcal{F}\{f\}(\omega) = \int_{\mathbb{R}^d} f(x) e^{-j\langle \omega, x \rangle} \, dx \]

Reconstruction formula (inverse Fourier transform)

\[ f(x) = \mathcal{F}^{-1}\{f\}(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\omega) e^{j\langle \omega, x \rangle} \, d\omega \quad \text{(a.e.)} \]

Equivalent analysis functions: \( \eta_m(x) = e^{j\langle \omega_m, x \rangle} \) (complex sinusoids)

2D Fourier reconstruction

Original image: \( f(x) \)

Reconstruction using \( N \) largest coefficients:

\[ \tilde{f}(x) = \frac{1}{(2\pi)^2} \sum_{\text{subset}} \hat{f}(\omega) e^{j\langle x, \omega \rangle} \]
Magnetic resonance imaging

- Magnetic resonance: \( \omega_0 = \gamma B_0 \)

Frequency encode:

\[
\omega_0 = \omega_0(x)
\]

- Linear forward model for MRI

\[
\hat{s}(\omega_m) = \int_{\mathbb{R}^3} s(r) e^{-j(\omega_m \cdot r)} \, dr
\]

(sampling of Fourier transform)

- Extended forward model with coil sensitivity

\[
\hat{s}_w(\omega_m) = \int_{\mathbb{R}^3} w(r) s(r) e^{-j(\omega_m \cdot r)} \, dr
\]

Basic operator: Windowing

\[
W : L_2(\mathbb{R}^d) \rightarrow L_2(\mathbb{R}^d)
\]

\[
W\{f\}(x) = w(x) f(x)
\]

Positive window function (continuous and bounded):

\( w \in C_b(\mathbb{R}^d), w(x) \geq 0 \)

- Special case: modulation

\[
w(r) = e^{j(\omega_0 \cdot r)}
\]

\[
e^{j(\omega_0 \cdot r)} f(r) \quad \xrightarrow{\mathcal{F}} \quad \hat{f}(\omega - \omega_0)
\]

Application: Structured illumination microscopy (SIM)
Basic operator: Convolution

\[ H : L_2(\mathbb{R}^d) \rightarrow L_2(\mathbb{R}^d) \]

\[ H\{f\}(x) = (h \ast f)(x) = \int_{\mathbb{R}^d} h(x - y)f(y)dy \]

Impulse response: \( h(x) = H\{\delta\} \)

Equivalent analysis functions: \( \eta_m(x) = h(x_m \cdot) \)

Frequency response: \( \hat{h}(\omega) = \mathcal{F}\{h\}(\omega) \)

- Convolution as a frequency-domain product

\[ (h \ast f)(x) \overset{\mathcal{F}}{\leftrightarrow} \hat{h}(\omega) \hat{f}(\omega) \]

Modeling of optical systems

\[ f(x, y) \rightarrow g(x, y) = (h \ast f)(x, y) \]

Diffraction-limited optics = LSI system

- Aberation-free point spread function (in focal plane)

\[ h(x, y) = h(r) = C \cdot \left[ \frac{2J_1(\pi r)}{\pi r} \right]^2 \]

where \( r = \sqrt{x^2 + y^2} \) (radial distance)

- Effect of misfocus

![Point source output](in focus) (defocus)
Basic operator: X-ray transform

Projection geometry: \( x = t\theta + r\theta^\perp \) with \( \theta = (\cos \theta, \sin \theta) \)

- Radon transform (line integrals)

\[
R_\theta \{ s(x) \}(t) = \int_{\mathbb{R}} s(t\theta + r\theta^\perp) dr \\
= \int_{\mathbb{R}^2} s(x) \delta(t - \langle x, \theta \rangle) dx
\]

Equivalent analysis functions: \( \eta_m(x) = \delta(t_m - \langle x, \theta_m \rangle) \)

Central slice theorem

- Measurements of line integrals (Radon transform)

\( p_\theta(t) = R_\theta \{ f \}(t, \theta) \)

- 1D and 2D Fourier transforms

\[
\hat{p}_\theta(\omega) = \mathcal{F}_{1D} \{ p_\theta \}(\omega) \\
\hat{f}(\omega) = \mathcal{F}_{2D} \{ f \}(\omega) = \hat{f}_{pol}(\omega, \theta)
\]

- Central-slice theorem

\[
\hat{p}_\theta(\omega) = \hat{f}(\omega \cos \theta, \omega \sin \theta) = \hat{f}_{pol}(\omega, \theta)
\]

Proof: for \( \theta = 0 \)

\[
\hat{f}(\omega, 0) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j\omega x} dx dy = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f(x, y) dy \right) e^{-j\omega x} dx = \hat{p}_0(\omega)
\]

then use rotation property of Fourier transform...
<table>
<thead>
<tr>
<th>Modality</th>
<th>Radiation</th>
<th>Forward model</th>
<th>Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D or 3D tomography</td>
<td>coherent x-ray</td>
<td>$y_i = R\theta, x$</td>
<td>parallel, cone beam, spiral sampling</td>
</tr>
<tr>
<td>3D deconvolution microscopy</td>
<td>fluorescence</td>
<td>$y = Hx$</td>
<td>brightfield, confocal, light sheet</td>
</tr>
<tr>
<td>structured illumination microscopy (SIM)</td>
<td>fluorescence</td>
<td>$y_i = \text{H}W_i x$</td>
<td>full 3D reconstruction, non-sinusoidal patterns</td>
</tr>
<tr>
<td>Positron Emission Tomography (PET)</td>
<td>gamma rays</td>
<td>$y_i = H\theta, x$</td>
<td>list mode with time-of-flight</td>
</tr>
<tr>
<td>Magnetic resonance imaging (MRI)</td>
<td>radio frequency</td>
<td>$y = Fx$</td>
<td>uniform or non-uniform sampling in k space</td>
</tr>
<tr>
<td>Cardiac MRI (parallel, non-uniform)</td>
<td>radio frequency</td>
<td>$y_{t,i} = F_iW_i x$</td>
<td>gated or not, retrospective registration</td>
</tr>
<tr>
<td>Optical diffraction tomography</td>
<td>coherent light</td>
<td>$y_i = W_iF_i x$</td>
<td>with holography or grating interferometry</td>
</tr>
</tbody>
</table>

**Discretization: Finite dimensional formalism**

$$s(r) = \sum_{k \in \Omega} s[k] \beta_k(r)$$

Signal vector: $s = \{s[k]\}_{k \in \Omega}$ of dimension $K$

- **Measurement model (image formation)**

  $$y_m = \int_{\mathbb{R}^d} s(r)\eta_m(r)dr + n[m] = \langle s, \eta_m \rangle + n[m], \quad (m = 1, \ldots, M)$$

  $\eta_m$: sampling/imaging function ($m$th detector)

  $n[\cdot]$: additive noise

  $$y = y_0 + n = Hs + n$$

  $(M \times K)$ system matrix: $[H]_{m,k} = \langle \eta_m, \beta_k \rangle = \int_{\mathbb{R}^d} \eta_m(r)\beta_k(r)dr$
Example of basis functions

Shift-invariant representation: $\beta_k(x) = \beta(x - k)$

Separable generator: $\beta(x) = \prod_{n=1}^{d} \beta(x_n)$

- Pixelated model
  $\beta(x) = \text{rect}(x)$

- Bilinear model
  $\beta(x) = (\text{rect} * \text{rect})(x) = \text{tri}(x)$

- Bandlimited representation
  $\beta(x) = \text{sinc}(x)$

Part 2:
Classical image reconstruction

Discretized forward model: $y = Hs + n$

Inverse problem: How to efficiently recover $s$ from $y$?
Vector calculus

- Scalar cost function $J(v) : \mathbb{R}^N \rightarrow \mathbb{R}$

- Vector differentiation: 
  \[
  \frac{\partial J(v)}{\partial v} = \begin{bmatrix}
  \frac{\partial J}{\partial v_1} \\
  \vdots \\
  \frac{\partial J}{\partial v_N}
  \end{bmatrix} = \nabla J(v) \quad \text{(gradient)}
  \]

- Necessary condition for an unconstrained optimum (minimum or maximum)
  \[
  \frac{\partial J(v)}{\partial v} = 0 \quad \text{(also sufficient if $J(v)$ is convex in $v$)}
  \]

- Useful identities
  \[
  \frac{\partial}{\partial v} (a^T v) = \frac{\partial}{\partial v} (v^T a) = a
  \]
  \[
  \frac{\partial}{\partial v} (v^T A v) = (A + A^T) \cdot v
  \]
  \[
  \frac{\partial}{\partial v} (v^T A v) = 2A \cdot v \quad \text{if $A$ is symmetric}
  \]

Basic reconstruction: least-squares solution

- Least-squares fitting criterion:
  \[
  J_{LS}(\tilde{s}, y) = \|y - H\tilde{s}\|^2
  \]
  \[
  \min_{\tilde{s}} \|y - \tilde{y}\|^2 = \min_{s} J_{LS}(s, y) \quad \text{(maximum consistency with the data)}
  \]

- Formal least-squares solution
  \[
  J_{LS}(s, y) = \|y - Hs\|^2 = \|y\|^2 + s^T H^T H s - 2y^T H s
  \]
  \[
  \frac{\partial J_{LS}(s, y)}{\partial s} = 2H^T H s - 2H^T y
  \]

- Basic limitations
  1) Inherent noise amplification
  2) Difficulty to invert $H$ (too large or non-square)
  3) All interesting inverse problems are ill-posed

- Backprojection (poor man's solution): $s \mapsto H^T y$
  - OK if $H$ is unitary

Basic limitations

1) Inherent noise amplification
2) Difficulty to invert $H$ (too large or non-square)
3) All interesting inverse problems are ill-posed
Linear inverse problems (20th century theory)

- Dealing with **ill-posed problems**: Tikhonov regularization

\[ R(s) = \|Ls\|_2^2: \text{ regularization (or smoothness) functional} \]

\[ L: \text{ regularization operator (i.e., Gradient)} \]

\[ \min_s R(s) \quad \text{subject to} \quad \|y - Hs\|_2^2 \leq \sigma^2 \]

- Equivalent variational problem

\[ s^* = \arg \min_s \|y - Hs\|_2^2 + \lambda \|Ls\|_2^2 \]

Formal linear solution: \( s = (H^T H + \lambda L^T L)^{-1} H^T y = R_\lambda \cdot y \)

Interpretation: “filtered” backprojection

Statistical formulation (20th century)

- Linear measurement model: \( y = Hs + n \)

\( n: \text{ additive white Gaussian noise (i.i.d.)} \)

\( s: \text{ realization of Gaussian process with zero-mean and covariance matrix } \mathbb{E}\{s \cdot s^T\} = C_s \)

- Wiener (LMMSE) solution = Gauss MMSE = Gauss MAP

\[ s_{\text{MAP}} = \arg \min_s \frac{1}{\sigma^2} \|y - Hs\|_2^2 + \|C_s^{-1/2} s\|_2^2 \]

\[ \uparrow \quad L = C_s^{-1/2}: \text{ Whitening filter} \]

- Quadratic regularization (Tikhonov)

\[ s_{\text{Tik}} = \arg \min_s \left( \|y - Hs\|_2^2 + \lambda R(s) \right) \quad \text{with} \quad R(s) = \|Ls\|_2^2 \]

**Linear solution**: \( s = (H^T H + \lambda L^T L)^{-1} H^T y = R_\lambda \cdot y \)
Iterative reconstruction algorithm

- Generic minimization problem: \( s_{\text{opt}} = \arg\min_s J(s, y) \)

- Steepest-descent solution
  \[
  s^{(k+1)} = s^{(k)} - \gamma \nabla J(s^{(k)}, y)
  \]

- Iterative constrained least-squares reconstruction
  \[
  J_{T,k}(s, y) = \frac{1}{2} \|y - Hs\|^2 + \frac{1}{2} \|Ls\|^2
  \]
  Gradient: \( \frac{\partial J_{T,k}(s, y)}{\partial s} = -s_0 + (H^T H + \lambda L^T L)s \)
  with \( s_0 = H^T y \)
  Steepest-descent algorithm
  \[
  s^{(k+1)} = s^{(k)} + \gamma (s_0 - (H^T H + \lambda L^T L)\tilde{s}^{(k)})
  \]
  Positivity constraint (IC):
  \[
  [\tilde{s}^{(k+1)}]_i = \begin{cases} 
  0, & [s^{(k+1)}]_i < 0 \\
  [s^{(k+1)}]_i, & \text{otherwise}.
  \end{cases}
  \]
  (projection on convex set)

Iterative deconvolution: unregularized case

Degraded image: Gaussian blur + additive noise

van Cittert animation

Ground truth
Effect of regularization parameter

Degraded image: Gaussian blur + additive noise

not enough: $\lambda=0.02$

not enough: $\lambda=0.2$

Optimal regularization: $\lambda=2$

too much: $\lambda=20$

too much: $\lambda=200$

Selecting the regularization operator

- Translation, rotation and scale-invariant operators
  - Laplacian: $\Delta s = (\nabla^T \nabla) s \iff -||\omega||^2 \hat{s}(\omega)$
  - Modulus of gradient: $|\nabla s|$
  - Fractional Laplacian: $(-\Delta)^{\gamma/2} \iff ||\omega||^\gamma \hat{s}(\omega)$

- TRS-invariant regularization functional
  $$||\nabla s||_{L^2(\mathbb{R}^d)}^2 = ||(-\Delta)^{1/2} s||_{L^2(\mathbb{R}^d)}^2 \Rightarrow \mathbf{L}: \text{discrete version of gradient}$$

- Fractional Brownian motion field
  - Statistical decoupling/whitening: $(-\Delta)^{\gamma/2} s = w \iff \frac{1}{|\omega|^\gamma}$ spectral decay
Relevance of self-similarity for bio-imaging

- Fractals and physiology

Designing fast reconstruction algorithms

Normal matrix: \( A = H^T H \)  (symmetric)

Formal linear solution: \( s = (A + \lambda L^T L)^{-1} H^T y = R \lambda \cdot y \)

Generic form of the iterator: \( s^{(k+1)} = s^{(k)} + \gamma (s_0 - (A + \lambda L^T L)s^{(k)}) \)

- Recognizing structured matrices
  - \( L \): convolution matrix \( \Rightarrow L^T L \): symmetric convolution matrix
  - \( L, A \): convolution matrices \( \Rightarrow (A + \lambda L^T L) \): symmetric convolution matrix

- Fast implementation
  - Diagonalization of convolution matrices \( \Rightarrow \) FFT-based implementation

- Applicable to:
  - deconvolution microscopy (Wiener filter)
  - parallel rays computer tomography (FBP)
  - MRI, including non-uniform sampling of k-space
Part 3:
Modern image reconstruction

**Linear inverse problems: The sparsity (r)evolution**

(20th Century) \( p = 2 \to 1 \) (21st Century)

\[
\mathbf{s}_{\text{rec}} = \arg\min_{\mathbf{s}} (\|\mathbf{y} - \mathbf{Hs}\|_2^2 + \lambda \mathcal{R}(\mathbf{s}))
\]

- Non-quadratic regularization
  \[
  \mathcal{R}(\mathbf{s}) = \|\mathbf{Ls}\|_2^2 \to \|\mathbf{Ls}\|_p^p \to \|\mathbf{Ls}\|_{\ell_1}
  \]
- Total variation \((\text{Rudin-Osher, 1992})\)
  \[
  \mathcal{R}(\mathbf{s}) = \|\mathbf{Ls}\|_{\ell_1} \text{ with } \mathbf{L}: \text{gradient}
  \]
- Wavelet-domain regularization \((\text{Figuereido et al., Daubechies et al. 2004})\)
  \[
  \mathbf{v} = \mathbf{W}^{-1}\mathbf{s}: \text{wavelet expansion of } \mathbf{s} \text{ (typically, sparse)}
  \]
  \[
  \mathcal{R}(\mathbf{s}) = \|\mathbf{v}\|_{\ell_1}
  \]
- Compressed sensing/sampling \((\text{Candes-Romberg-Tao; Donoho, 2006})\)
Sparsifying transforms

Biomedical images are well described by few basis coefficients

![Graph showing Normalised MSE vs Percentage of coefficients kept for different transforms.](image)

Prior = sparse representation

\[ \mathcal{R}(s) = \lambda \| W^T s \|_1 \]

Advantages:
- convex
- favors sparse solutions
- Fast: WFISTA

(Guerquin-Kern *IEEE TMI* 2011)

Theory of compressive sensing

- Generalized sampling setting (after discretization)
  - Linear inverse problem: \( y = Hs + n \)
  - Sparse representation of signal: \( s = Wx \) with \( \|x\|_0 = K \ll N_x \)
  - \( N_y \times N_x \) system matrix: \( A = HW \)

- Formulation of ill-posed recovery problem when \( 2K < N_y \ll N_x \)
  \[
  (P0) \quad \min_x \| y - Ax \|_2^2 \quad \text{subject to} \quad \|x\|_0 \leq K
  \]

- Theoretical result
  Under suitable conditions on \( A \) (e.g., restricted isometry), the solution is unique and the recovery problem \((P0)\) is equivalent to:
  \[
  (P1) \quad \min_x \| y - Ax \|_2^2 \quad \text{subject to} \quad \|x\|_1 \leq C_1
  \]

[Donoho et al., 2005, Candès-Tao, 2006, ...]
Compressive sensing (CS) and $l_1$ minimization

Sparse representation of signal: $s = Wx$ with $\|x\|_0 = K \ll N_x$

Equivalent $N_y \times N_x$ sensing matrix: $A = HW$

Constrained (synthesis) formulation of recovery problem

$$\min_x \|x\|_1 \quad \text{subject to} \quad \|y - Ax\|_2^2 \leq \sigma^2$$

Classical regularized least-squares estimator

Linear measurement model:

$y_m = (h_m, x) + n[m], \quad m = 1, \ldots, M$

System matrix: $H = [h_1 \cdots h_M]^T \in \mathbb{R}^{N \times N}$

$$x_{LS} = \arg \min_{x \in \mathbb{R}^N} \|y - Hx\|_2^2 + \lambda \|x\|_2^2$$

$$\Rightarrow \quad x_{LS} = (H^T H + \lambda I_N)^{-1} H^T y$$

$$= H^T a = \sum_{m=1}^{M} a_m h_m \quad \text{where} \quad a = (H H^T + \lambda I_M)^{-1} y$$

Interpretation: $x_{LS} \in \text{span}\{h_m\}_{m=1}^{M}$

Lemma

$$(H^T H + \lambda I_N)^{-1} H^T = H^T (H H^T + \lambda I_M)^{-1}$$
Generalization: constrained $l_2$ minimization

- Discrete signal to reconstruct: $x = (x[n])_{n \in \mathbb{Z}}$
- Sensing operator $H : \ell_2(\mathbb{Z}) \to \mathbb{R}^M$
  $x \mapsto z = H\{x\} = (\langle x, h_1 \rangle, \ldots, \langle x, h_M \rangle)$ with $h_m \in \ell_2(\mathbb{Z})$
- Closed convex set in measurement space: $C \subset \mathbb{R}^M$

Example: $C_y = \{z \in \mathbb{R}^M : \|z - y\|^2 \leq \sigma^2\}$

Representer theorem for constrained $l_2$ minimization

(P2) $\min_{x \in \ell_2(\mathbb{Z})} \|x\|_{l_2}^2$ s.t. $H\{x\} \in C$

The problem (P2) has a unique solution of the form

$$x_{\text{LS}} = \sum_{m=1}^{M} a_m h_m = H^*\{a\}$$

with expansion coefficients $a = (a_1, \ldots, a_M) \in \mathbb{R}^M$.


Constrained $l_1$ minimization $\Rightarrow$ sparsifying effect

- Discrete signal to reconstruct: $x = (x[n])_{n \in \mathbb{Z}}$
- Sensing operator $H : \ell_1(\mathbb{Z}) \to \mathbb{R}^M$
  $x \mapsto z = H\{x\} = (\langle x, h_1 \rangle, \ldots, \langle x, h_M \rangle)$ with $h_m \in \ell_{\infty}(\mathbb{Z})$
- Closed convex set in measurement space: $C \subset \mathbb{R}^M$

Representer theorem for constrained $l_1$ minimization

(P1) $V = \arg \min_{x \in \ell_1(\mathbb{Z})} \|x\|_{\ell_1}$ s.t. $H\{x\} \in C$

is convex, weak*-compact with extreme points of the form

$$x_{\text{sparse}}[\cdot] = \sum_{k=1}^{K} a_k \delta[\cdot - n_k] \quad \text{with} \quad K = \|x_{\text{sparse}}\|_0 \leq M.$$
Controlling sparsity

Measurement model: $y_m = \langle h_m, x \rangle + n[m], \ m = 1, \ldots, M$

$$x_{\text{sparse}} = \arg \min_{x \in \ell_1(Z)} \left( \sum_{m=1}^{M} \left| y_m - \langle h_m, x \rangle \right|^2 + \lambda \| x \|_{\ell_1} \right)$$

Geometry of $l_2$ vs. $l_1$ minimization

- Prototypical inverse problem
  $\min_x \left\{ \| y - Hx \|_{\ell_2}^2 + \lambda \| x \|_{\ell_2}^2 \right\} \Leftrightarrow \min_x \| x \|_{\ell_2} \text{ subject to } \| y - Hx \|_{\ell_2} \leq \sigma^2$

  $\min_x \left\{ \| y - Hx \|_{\ell_2}^2 + \lambda \| x \|_{\ell_1} \right\} \Leftrightarrow \min_x \| x \|_{\ell_1} \text{ subject to } \| y - Hx \|_{\ell_2} \leq \sigma^2$

$\ell_2$-ball: $|x_1|^2 + |x_2|^2 = C_2$

$\ell_1$-ball: $|x_1| + |x_2| = C_1$
Geometry of $l_2$ vs. $l_1$ minimization

- Prototypical inverse problem

\[
\begin{align*}
\min_x \left\{ \|y - Hx\|_2^2 + \lambda \|x\|_2^2 \right\} & \iff \min_x \|x\|_2 \quad \text{subject to} \quad \|y - Hx\|_2^2 \leq \sigma^2 \\
\min_x \left\{ \|y - Hx\|_2^2 + \lambda \|x\|_1 \right\} & \iff \min_x \|x\|_1 \quad \text{subject to} \quad \|y - Hx\|_2^2 \leq \sigma^2
\end{align*}
\]

Configuration for **non-unique** $\ell_1$ solution

Variational-MAP formulation of inverse problem

- Linear forward model

\[
y = Hs + n
\]

- Reconstruction as an optimization problem

\[
s_{\text{rec}} = \arg \min \left[ \|y - Hs\|_2^2 + \lambda \|Ls\|_p^p \right], \quad p = 1, 2
\]

\[- \log \text{Prob}(s) : \text{prior likelihood}\]
Discretization of reconstruction problem

Spline-like reconstruction model: \( s(r) = \sum_{k \in \Omega} s[k] \beta_k(r) \quad \leftrightarrow \quad s = (s[k])_{k \in \Omega} \)

- Statistical innovation model

\[
\begin{align*}
L_s w &= s \\
L^{-1} w &= s
\end{align*}
\]

Discretization \( u = Ls \) (matrix notation)

- Physical model: image formation and acquisition

\[
y_m = \int_{\mathbb{R}^d} s(x) \eta_m(x) dx + n[m] = \langle s, \eta_m \rangle + n[m], \quad (m = 1, \ldots, M)
\]

\[
y = y_0 + n = Hs + n
\]

\( n: \text{i.i.d. noise with pdf } p_N \)

Posterior probability distribution

\[
p_{S|Y}(s|y) = \frac{p_Y(s|y)p_S(s)}{p_Y(y)} = \frac{p_N(y - Hs)p_S(s)}{p_Y(y)} \\
= \frac{1}{Z} p_N(y - Hs)p_S(s)
\]

(Mayes’ rule)

Statistical decoupling

\[
u = Ls \quad \Rightarrow \quad p_S(s) \propto p_U(Ls) \approx \prod_{k \in \Omega} p_U([Ls][k])
\]

- Additive white Gaussian noise scenario (AWGN)

\[
p_{S|Y}(s|y) \propto \exp \left( -\frac{||y - Hs||^2}{2\sigma^2} \right) \prod_{k \in \Omega} p_U([Ls][k])
\]

... and then take the log and maximize ...
General form of MAP estimator

\[ s_{MAP} = \arg \min \left( \frac{1}{2} \| y - Hs \|_2^2 + \sigma^2 \sum_n \Phi_U([Ls]_n) \right) \]

- **Gaussian:** \( p_U(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-x^2/(2\sigma_0^2)} \) \( \Rightarrow \Phi_U(x) = \frac{1}{2\sigma_0} x^2 + C_1 \)
- **Laplace:** \( p_U(x) = \frac{\lambda}{2} e^{-\lambda|x|} \) \( \Rightarrow \Phi_U(x) = \lambda|x| + C_2 \)
- **Student:** \( p_U(x) = \frac{1}{B(\frac{r}{2}, \frac{1}{2})} \left( \frac{1}{x^2 + 1} \right)^{r + \frac{1}{2}} \) \( \Rightarrow \Phi_U(x) = (r + \frac{1}{2}) \log(1 + x^2) + C_3 \)

Potential: \( \Phi_U(x) = -\log p_U(x) \)

Proximal operator: pointwise denoiser

\[ \text{prox}_{\Phi_U}(y; \sigma^2) = \arg \min_{u \in \mathbb{R}} \frac{1}{2} |y - u|^2 + \sigma^2 \Phi_U(u) \]

\[ \tilde{u} = \text{prox}_{\Phi_U}(y; 1) \]

\[ \sigma^2 \Phi_U(u) \]

- **linear attenuation**
- **soft-threshold**
- **shrinkage function**

\( \ell_2 \) minimization
\( \ell_1 \) minimization
\( \approx \ell_p \) relaxation for \( p \to 0 \)
Maximum a posteriori (MAP) estimation

- Constrained optimization formulation

  Auxiliary innovation variable: \( u = Ls \)

  \[
  s_{MAP} = \arg \min_{s \in \mathbb{R}^k} \left( \frac{1}{2} \| y - Hs \|_2^2 + \sigma^2 \sum_n \Phi_U([u]_n) \right) \quad \text{subject to} \quad u = Ls
  \]

- Augmented Lagrangian method

  Quadratic penalty term: \( \frac{\mu}{2} \| Ls - u \|_2^2 \)

  Lagrange multiplier vector: \( \alpha \)

  \[
  \mathcal{L}_A(s, u, \alpha) = \frac{1}{2} \| y - Hs \|_2^2 + \sigma^2 \sum_n \Phi_U([u]_n) + \alpha^T(Ls - u) + \frac{\mu}{2} \| Ls - u \|_2^2
  \]

  (Bostan et al. *IEEE TIP* 2013)

Alternating direction method of multipliers (ADMM)

\[
\mathcal{L}_A(s, u, \alpha) = \frac{1}{2} \| y - Hs \|_2^2 + \sigma^2 \sum_n \Phi_U([u]_n) + \alpha^T(Ls - u) + \frac{\mu}{2} \| Ls - u \|_2^2
\]

Sequential minimization

\[
s^{k+1} \leftarrow \arg \min_{s \in \mathbb{R}^N} \mathcal{L}_A(s, u^k, \alpha^k)
\]

\[
\alpha^{k+1} = \alpha^k + \mu (Ls^{k+1} - u^k)
\]

\[
u^{k+1} \leftarrow \arg \min_{u \in \mathbb{R}^N} \mathcal{L}_A(s^{k+1}, u, \alpha^{k+1})
\]

Linear inverse problem:

\[
s^{k+1} = (H^T H + \mu L^T L)^{-1} (H^T y + z^{k+1})
\]

with \( z^{k+1} = L^T (\mu u^k - \alpha^k) \)

Nonlinear denoising:

\[
u^{k+1} = \text{prox}_{\Phi_U} \left( Ls^{k+1} + \frac{1}{\mu} \alpha^{k+1}, \frac{\sigma^2}{\mu} \right)
\]

- Proximal operator tailored to stochastic model

  \[
  \text{prox}_{\Phi_U}(y; \lambda) = \arg \min_u \frac{1}{2} \| y - u \|_2^2 + \lambda \Phi_U(u)
  \]

  Cauchy prior with increasing \( \sigma_0 \)
Deconvolution of fluorescence micrographs

- Physical model of a diffraction-limited microscope

\[ g(x, y, z) = (h_{3D} * s)(x, y, z) \]

3-D point spread function (PSF)

\[ h_{3D}(x, y, z) = I_0 \left| p_\lambda \left( \frac{x}{\lambda f_0}, \frac{y}{\lambda f_0}, \frac{z}{\lambda f_0} \right) \right|^2 \]

\[ p_\lambda(x, y, z) = \int_{\mathbb{R}^2} P(\omega_1, \omega_2) \exp \left( j2\pi \left( \frac{\omega_1^2 + \omega_2^2}{2\lambda f_0^2} \right) \right) \exp \left( -j2\pi \frac{x\omega_1 + y\omega_2}{\lambda f_0} \right) \, d\omega_1 \, d\omega_2 \]

Optical parameters

- \( \lambda \): wavelength (emission)
- \( M \): magnification factor
- \( f_0 \): focal length
- \( P(\omega_1, \omega_2) = 1_{||\omega|| < R_0} \): pupil function
- \( NA = n \sin \theta = R_0/f_0 \): numerical aperture

2-D convolution model

- Airy disk: \( h_{2D}(x, y) = I_0 \left| 2J_1(r/r_0)/r/r_0 \right|^2 \)

with \( r = \sqrt{x^2 + y^2} \), \( r_0 = \frac{f_0}{2\pi R_0} \), \( J_1(r) \): first-order Bessel function.

- Modulation transfer function

\[ \hat{h}_{2D}(\omega) = \begin{cases} \frac{2}{\pi} \left( \arccos \left( \frac{||\omega||}{\omega_0} \right) - \frac{||\omega||}{\omega_0} \sqrt{1 - \left( \frac{||\omega||}{\omega_0} \right)^2} \right), & \text{for } 0 \leq ||\omega|| < \omega_0 \\ 0, & \text{otherwise} \end{cases} \]

Cut-off frequency (Rayleigh): \( \omega_0 = \frac{2R_0}{\lambda f_0} = \frac{\pi}{r_0} \approx \frac{2NA}{\lambda} \)
2-D deconvolution: numerical set-up

- Discretization

\[ \omega_0 \leq \pi \] and representation in (separable) sinc basis \{ sinc(x - k) \}_{k \in \mathbb{Z}^2} \\

Analysis functions: \( \eta_m(x, y) = h_{2D}(x - m_1, y - m_2) \)

\[ [H]_{m,k} = \langle \eta_m, \text{sinc}(\cdot - k) \rangle = \langle h_{2D}(\cdot - m), \text{sinc}(\cdot - k) \rangle = (\text{sinc} * h_{2D})(m - k) = h_{2D}(m - k). \]

\( H \) and \( L \): convolution matrices diagonalized by discrete Fourier transform

- Linear step of ADMM algorithm implemented using the FFT

\[
s^{k+1} = \left( H^T H + \mu L^T L \right)^{-1} \left( H^T y + z^{k+1} \right) \quad \text{with} \quad z^{k+1} = L^T \left( \mu u^k - \alpha^k \right)
\]

Deconvolution experiments

![Images](Image)

(Figure 10.3) Images used in deconvolution experiments. (a) Stem cells surrounded by goblet cells. (b) Nerve cells growing around fibers. (c) Artery cells.

Table 10.2 Deconvolution performance of MAP estimators based on different prior distributions.

<table>
<thead>
<tr>
<th></th>
<th>BSNR (dB)</th>
<th>Estimation performance (SNR in dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gaussian</td>
<td>Laplace</td>
</tr>
<tr>
<td>Stem cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>14.43</td>
<td>13.76</td>
</tr>
<tr>
<td>30</td>
<td>15.92</td>
<td>15.77</td>
</tr>
<tr>
<td>40</td>
<td>18.11</td>
<td>16.11</td>
</tr>
<tr>
<td>Nerve cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>13.86</td>
<td>15.31</td>
</tr>
<tr>
<td>30</td>
<td>15.89</td>
<td>16.18</td>
</tr>
<tr>
<td>40</td>
<td>18.58</td>
<td>20.57</td>
</tr>
<tr>
<td>Artery cells</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>14.86</td>
<td>15.23</td>
</tr>
<tr>
<td>30</td>
<td>16.59</td>
<td>17.21</td>
</tr>
<tr>
<td>40</td>
<td>18.68</td>
<td>19.61</td>
</tr>
</tbody>
</table>
2D deconvolution experiment

Disk shaped PSF (7x7)

Deconvolution results in dB

<table>
<thead>
<tr>
<th></th>
<th>Gaussian Estimator</th>
<th>Laplace Estimator</th>
<th>Student's Estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Astrocytes cells</td>
<td>12.18</td>
<td>10.48</td>
<td>10.52</td>
</tr>
<tr>
<td>Pulmonary cells</td>
<td>16.90</td>
<td>19.04</td>
<td>18.34</td>
</tr>
<tr>
<td>Stem cells</td>
<td>15.81</td>
<td>20.19</td>
<td>20.50</td>
</tr>
</tbody>
</table>

L : gradient

Optimized parameters

3D deconvolution of widefield stack

Maximum intensity projections of $384 \times 448 \times 260$ image stacks;
Leica DM 5500 widefield epifluorescence microscope with a $63 \times$ oil-immersion objective;
C. Elegans embryo labeled with Hoechst, Alexa488, Alexa568;
wavelet regularization (Haar), 3 decomposition levels for X-Y, 2 decomposition levels for Z.

(Vonesch-U., IEEE TIP 2009)
Magnetic resonance imaging (MRI)

- Physical image formation model (noise-free)
  \[ \hat{s}(\omega_m) = \int_{\mathbb{R}^2} s(r) e^{-j(\omega_m \cdot r)} \, dr \]  
  (sampling of Fourier transform)

  Equivalent analysis function: \( \eta_m(r) = e^{-j(\omega_m \cdot r)} \)

- Discretization in separable sinc basis

  \[ [H]_{m,n} = \langle \eta_m, \text{sinc}(\cdot - n) \rangle \]
  \[ = \langle e^{-j(\omega_m \cdot \cdot)}, \text{sinc}(\cdot - n) \rangle = e^{-j(\omega_m \cdot n)} \]

  Property: \( H^T H \) is circulant  
  (FFT-based implementation)
The basic problem in MRI is then to reconstruct \( s(\rho) \) based on the partial knowledge of its Fourier coefficients which are also corrupted by noise. While the reconstruction in the case of a dense Cartesian sampling amounts to a simple inverse Fourier transform, it becomes more challenging for other trajectories, especially as the sampling density decreases.

For simplicity, we discretize the forward model by using the same sinc basis functions as for the deconvolution problem of Section 10.3.2. This results in the system matrix

\[
H_{m,n} = h_{\epsilon m, \nu n} \text{ sinc}(\leq \theta n) i = h_{\epsilon m, \nu n} \text{ sinc}(\leq \theta n) i = e^{j \theta m, n},
\]

under the assumption that \( k! m k1 \sum \). The clear advantage of using the sinc basis is that \( H \) reduces to a discrete Fourier-like matrix, with the caveat that the frequency sampling is not necessarily uniform.

A convenient feature of this imaging model is that the matrix \( H^T H \) is circulant so that the linear iteration step of the algorithm can be computed in exact form using the FFT.

**Figure 10.4** Data used in MR reconstruction experiments. (a) Cross section of a wrist. (b) Angiography image. (c) k-space sampling pattern along 40 radial lines.

**Table 10.3** MR reconstruction performance of MAP estimators based on different prior distributions.

<table>
<thead>
<tr>
<th>Radial lines</th>
<th>Estimation performance (SNR in dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gaussian</td>
</tr>
<tr>
<td>Wrist</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>8.82</td>
</tr>
<tr>
<td>40</td>
<td>11.30</td>
</tr>
<tr>
<td>Angiogram</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>4.30</td>
</tr>
<tr>
<td>40</td>
<td>6.31</td>
</tr>
</tbody>
</table>
**ISMRM reconstruction challenge**

$L_2$ regularization (Laplacian) \hspace{1cm} $\ell_1$ wavelet regularization

(Guerquin-Kern *IEEE TMI* 2011)

---

**Differential phase-contrast tomography**

Paul Scherrer Institute (PSI), Villigen

Mathematical model

\[
y(t, \theta) = \frac{\partial}{\partial t} R_\theta \{ s \}(t)
\]

\[
y = H \, s
\]

\[
[H]_{(i,j),k} = \frac{\partial}{\partial t} P_{\theta_j} \beta_k(t_j)
\]

(Pfeiffer, *Nature* 2006)
Properties of Radon transform

- Projected translation invariance
  \[ R_{\theta}\{\varphi(\cdot - x_0)\}(t) = R_{\theta}\{\varphi\}(t - \langle x_0, \theta \rangle) \]

- Pseudo-distributivity with respect to convolution
  \[ R_{\theta}\{\varphi_1 \ast \varphi_2\}(t) = (R_{\theta}\{\varphi_1\} \ast R_{\theta}\{\varphi_2\})(t) \]

- Fourier central-slice theorem
  \[
  \int_{\mathbb{R}} R_{\theta}\{\varphi\}(t)e^{-j\omega t} \, dt = \hat{\varphi}(\omega)|_{\omega = \omega\theta}
  \]

**Proposition**: Consider the separable function \( \varphi(x) = \varphi_1(x)\varphi_2(y) \). Then,
\[
R_{\theta}\{\varphi(\cdot - x_0)\}(t) = \varphi_0(t - t_0)
\]
where \( t_0 = \langle x_0, \theta \rangle \) and
\[
\varphi_0(t) = \left( \frac{1}{|\cos \theta|} \varphi_1\left(\frac{\cdot}{|\cos \theta|}\right) \ast \frac{1}{|\sin \theta|} \varphi_2\left(\frac{\cdot}{|\sin \theta|}\right) \right)(t).
\]

Reducing the numbers of views

Rat brain reconstruction with 181 projections

**ADMM-PCG**
- SSIM = .96
- SSIM = .95
- SSIM = .49
- SSIM = .89

**g-FBP**
- SSIM = .51
- SSIM = .60
- SSIM = .15
- SSIM = .43

Collaboration: Prof. Marco Stampanoni, TOMCAT PSI / ETHZ

(Nichian et al. Optics Express 2013)
Performance evaluation

Goldstandard: high-quality iterative reconstruction with 721 views

\[ J(x, u) = \frac{1}{2} \| y - Hx \|_2^2 + \lambda R(u) + \mu \| Lx - u \|_2^2 \]

Physical model

Statistical model of signal

Schematic structure of reconstruction algorithm:

\[ x^{(n)} = \arg \min_x J(x, u^{(n-1)}) : \text{Linear step (problem specific)} \]
\[ u^{(n)} = \arg \min_u J(x^{(n)}, u) : \text{Statistical or “denoising” step} \]

Repeat \( N_{\text{iter}} \) times

until stop criterion

⇒ Reduction of acquisition time by a factor 10 (or more)?
**Inverse problems in imaging: Current status**

- **Higher reconstruction quality:** Sparsity-promoting schemes almost systematically outperform the classical linear reconstruction methods in MRI, x-ray tomography, deconvolution microscopy, etc...  
  (Lustig et al. 2007)

- **Faster imaging, reduced radiation exposure:** Reconstruction from a lesser number of measurements supported by **compressed sensing.**  
  (Candes-Romberg-Tao; Donoho, 2006)

- **Increased complexity:** Resolution of linear inverse problems using $\ell_1$ regularization requires more sophisticated algorithms (iterative and non-linear); efficient solutions (FISTA, ADMM) have emerged during the past decade.  
  (Chambolle 2004; Figueiredo 2004; Beck-Teboule 2009; Boyd 2011)

**Outstanding research issues**

- Beyond $\ell_1$ and TV: Connection with **statistical modeling & learning**
- Beyond matrix algebra: **Continuous-domain** formulation

---

**Part 4:**

Short guess about the future:  
The (deep) learning revolution (?)

---
Learning within the current paradigm

- Data-driven tuning of parameters: \( \lambda \), calibration of forward model
  Semi-blind methods, sequential optimization

- Improved decoupling/representation of the signal
  Data-driven **dictionary learning**
  (based of sparsity or statistics/ICA)

  \( \Rightarrow \) “optimal” \( L \)

  (Elad 2006, Ravishankar 2011, Mairal 2012)

- Learning of non-linearities / Proximal operators
  CNN-type parametrization, backpropagation

  \( \Rightarrow \) “optimal” potential \( \Phi \)

  (Chen-Pock 2015-2016, Kamilov 2016)

---

Recent appearance of Deep Conv Neural Nets

(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., arXiv:1611.03679)
Dose reduction by 7: 143 views

Reconstructed from 1000 views

(Jin et al., arXiv:1611.03679)

Dose reduction by 20: 50 views

Reconstructed from 1000 views

(Jin et al., arXiv:1611.03679)
Dose reduction by 14: 51 views

μCT data

Ground truth

FBP
SNR 3.265

TV
SNR 7.481

FBPConvNet
SNR 9.003

Reconstructed from from 721 views

COMPARISON OF SNR BETWEEN DIFFERENT RECONSTRUCTION ALGORITHMS FOR EXPERIMENTAL DATASET.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Methods</th>
<th>FBP</th>
<th>TV [13]</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. SNR (dB)</td>
<td>145 views (x5)</td>
<td>5.38</td>
<td>8.25</td>
<td>11.34</td>
</tr>
<tr>
<td></td>
<td>51 views (x14)</td>
<td>3.29</td>
<td>7.25</td>
<td>8.85</td>
</tr>
</tbody>
</table>

Challenges for deep learning methods

- Fundamental change of paradigm
  - Requires availability of **extensive sets of representative training data** together with **gold-standards** = desired high-quality reconstruction
- Research challenges/opportunities
  - How does one assess **reconstruction quality**? Should be “task oriented”!!!
  - Use of CNN to **correct artifacts** of current methods
  - Reconstruction from **fewer measurements** (trained on high-quality full-view data sets).
  - Use of CNN to **emulate/speedup** some well-performing, but “slow”, reference reconstruction methods
  - Development of more **realistic simulators** both “ground truth” images + physical forward model
  - **True 3D** CNN toolbox (still missing)

Can we trust the results?
References

■ Theoretical foundations

■ Algorithms and imaging applications

Acknowledgments

Many thanks to (former) members of EPFL’s Biomedical Imaging Group
  ■ Dr. Pouya Tafti
  ■ Prof. Arash Amini
  ■ Dr. John-Paul Ward
  ■ Julien Fageot
  ■ Dr. Emrah Bostan
  ■ Dr. Masih Nilchian
  ■ Dr. Ulugbek Kamilov
  ■ Dr. Cédric Vonesch
  ■ ....

and collaborators ...
  ■ Prof. Demetri Psaltis
  ■ Prof. Marco Stampanoni
  ■ Prof. Carlos-Oscar Sorzano
  ■ Dr. Arne Seitz
  ■ ....

■ Preprints and demos: [http://bigwww.epfl.ch/](http://bigwww.epfl.ch/)
General convex problems with gTV regularization

\[ \mathcal{M}_L(\mathbb{R}^d) = \{ s : \text{gTV}(s) = \|L\{s\}\|_\mathcal{M} = \sup_{\|\varphi\|_\infty \leq 1} \langle L\{s\}, \varphi \rangle < \infty \} \]

- **Linear** measurement operator \( \mathcal{M}_L(\mathbb{R}^d) \to \mathbb{R}^M : f \mapsto z = H\{f\} \)

- **C**: convex compact subset of \( \mathbb{R}^M \)

- Finite-dimensional null space \( \mathcal{N}_L = \{ q \in \mathcal{M}_L(\mathbb{R}^d) : L\{q\} = 0 \} \) with basis \( \{ p_n \}_{n=1}^{N_0} \)

**Admissibility of regularization**: \( H\{q_1\} = H\{q_2\} \iff q_1 = q_2 \) for all \( q_1, q_2 \in \mathcal{N}_L \)

---

**Representer theorem for gTV regularization**

The extremal points of the constrained minimization problem

\[ \mathcal{V} = \arg \min_{f \in \mathcal{M}_L(\mathbb{R}^d)} \|L\{f\}\|_\mathcal{M} \quad \text{s.t.} \quad H\{f\} \in C \]

are necessarily of the form

\[ f(x) = \sum_{k=1}^{K} a_k p_L(x - x_k) + \sum_{n=1}^{N_0} b_n p_n(x) \]

with \( K \leq M - N_0 \); that is, **non-uniform L-splines** with knots at the \( x_k \) and \( \|L\{f\}\|_\mathcal{M} = \sum_{k=1}^{K} |a_k| \). The full solution set is the **convex hull** of those extremal points.