OPTIMAL DESIGN IN MEDICAL INVERSION

Lior Horesh, Eldad Haber, Luis Tenorio

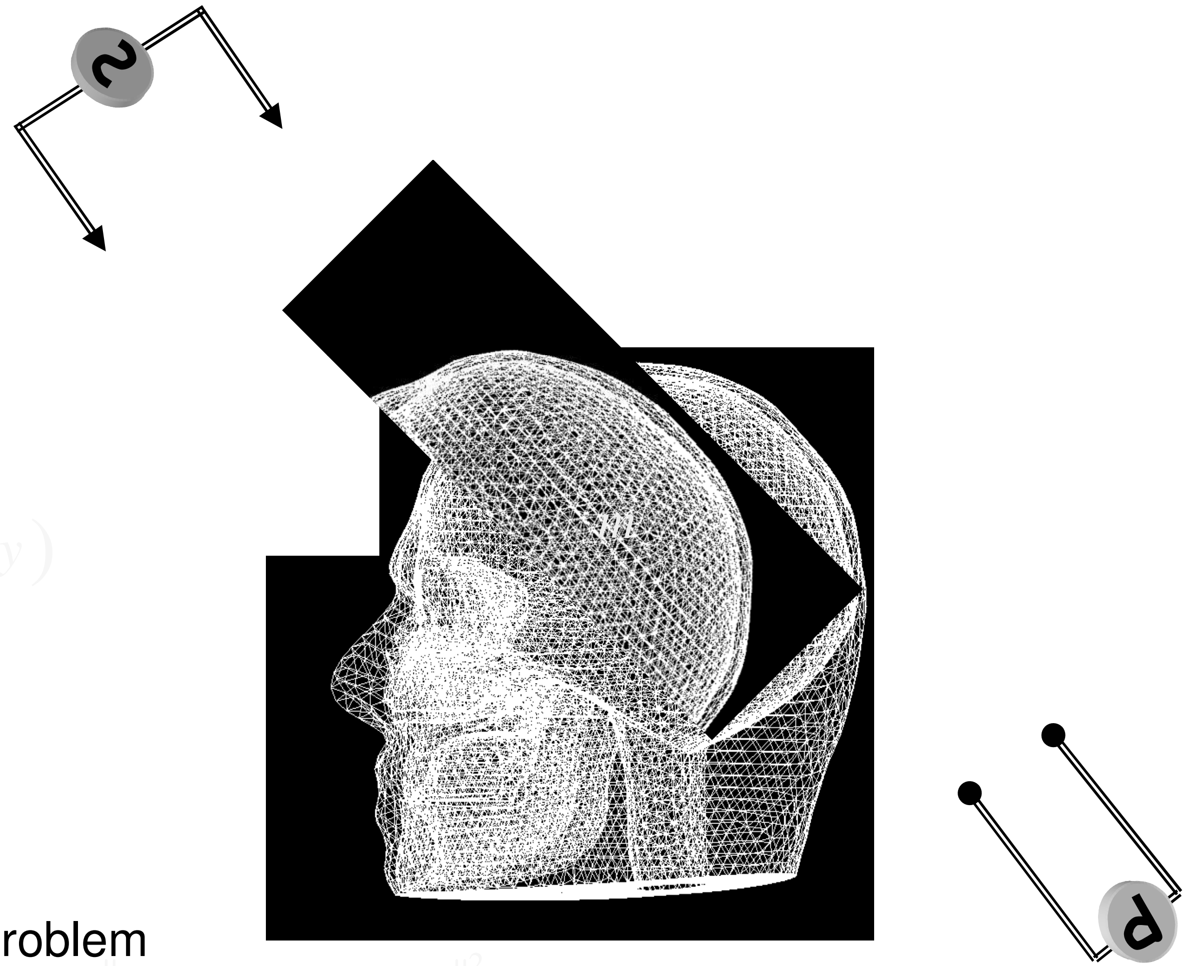
FIELDS

IBM TJ Watson Research Center - Toronto, Canada - June 2011

IBM

INTRODUCTION

EXPOSITION - INVERSE PROBLEMS



- Aim: infer model
- Given
  - Design parameters
  - Measurements
  - Observation model

$$F(m;y) + \eta = d(m;y)$$

- Naïve inversion ... Fails...

$$\|m - \hat{m}\|^2 = O(1)$$

- Cast as an optimization problem

$$\hat{m} = \arg\min \underbrace{\|F(m;y) - d(y)\|^2}_{data\ fit} + \underbrace{S(m)}_{regularization}$$

HOW TO IMPROVE MODEL RECOVERY ?

$$\hat{m} = \arg\min \underbrace{\left\| F(m; y) - d(y) \right\|^2}_{data\ fit} + \underbrace{S(m)}_{regularization}$$

- How can we ...
  - Improve observation model ?
  - Extract more information in the measurement procedure ?
  - Use more meaningful a-priori information ?
  - Provide more efficient optimization schemes ?

PART I

REGULARIZATION DESIGN

REGULARIZATION DESIGN - BACKGROUND

REGULARIZATION APPROACHES

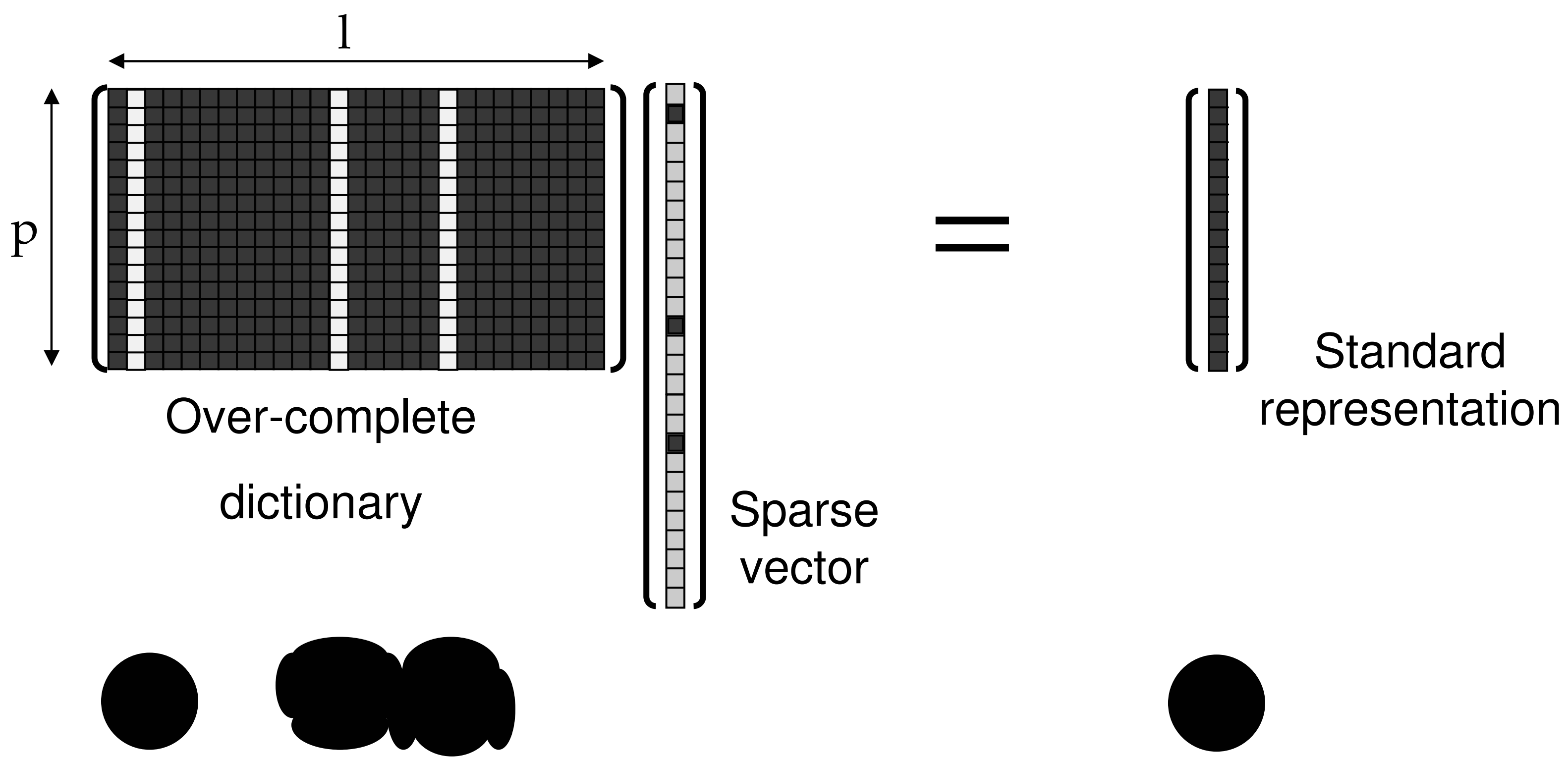
- Why regularization is necessary ?
  - Imposes a-priori information
  - Stabilizes the inversion process
  - Provides a unique solution

$$\hat{m} = \arg\min \underbrace{\left\| F(m;y) - d(y) \right\|^2}_{data\ fit} + \underbrace{S(m)}_{regularization}$$

- Two approaches
  - Explicit
  - Sparse representation

HOW TO REPRESENT SPARSELY ?

- Principle of parsimony True model can be represented by a small number of parameters
- Each column is a prototype model atom
- Sparse representation vector



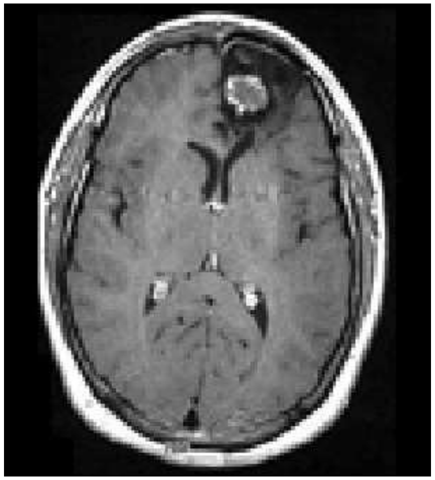
SPARSE REPRESENTATION

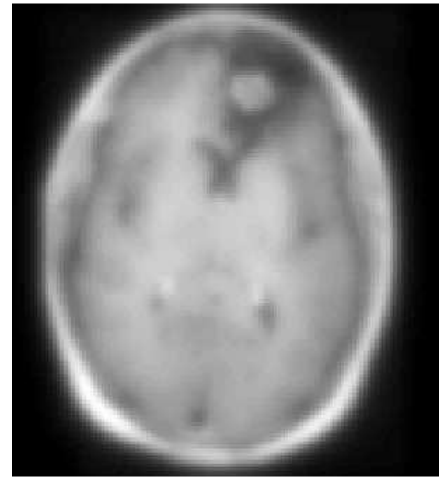
- Ideally sparsest solution achieved by -’norm’ penalty
- Non-convex NP-hard combinatorial problem
- Instead employ -norm (*Donoho 2006*)

$$\hat{m} = \arg\min_{\hat{m}} \underbrace{\left\| F(m;y) - d(y) \right\|^2}_{data\ misfit} + \underbrace{S(m)}_{regularization}$$

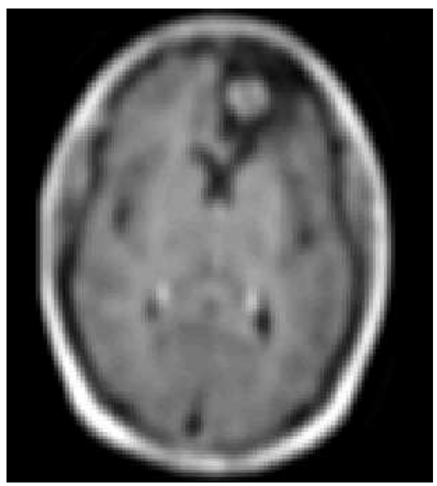
$$\hat{u} = \arg\min_{\hat{u}} \left\| F(Du;y) - d(y) \right\|_2^2 + a \left\| u \right\|_1$$

SPARSE REPRESENTATION PERFORMANCE

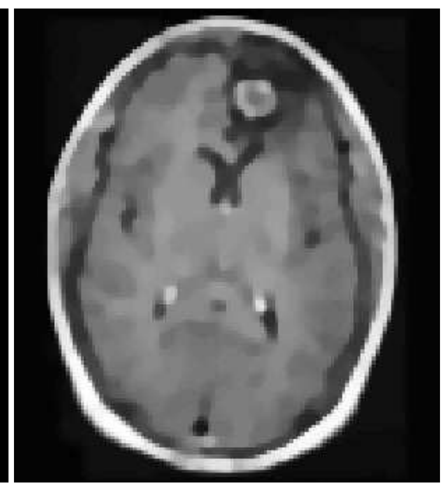




Energy



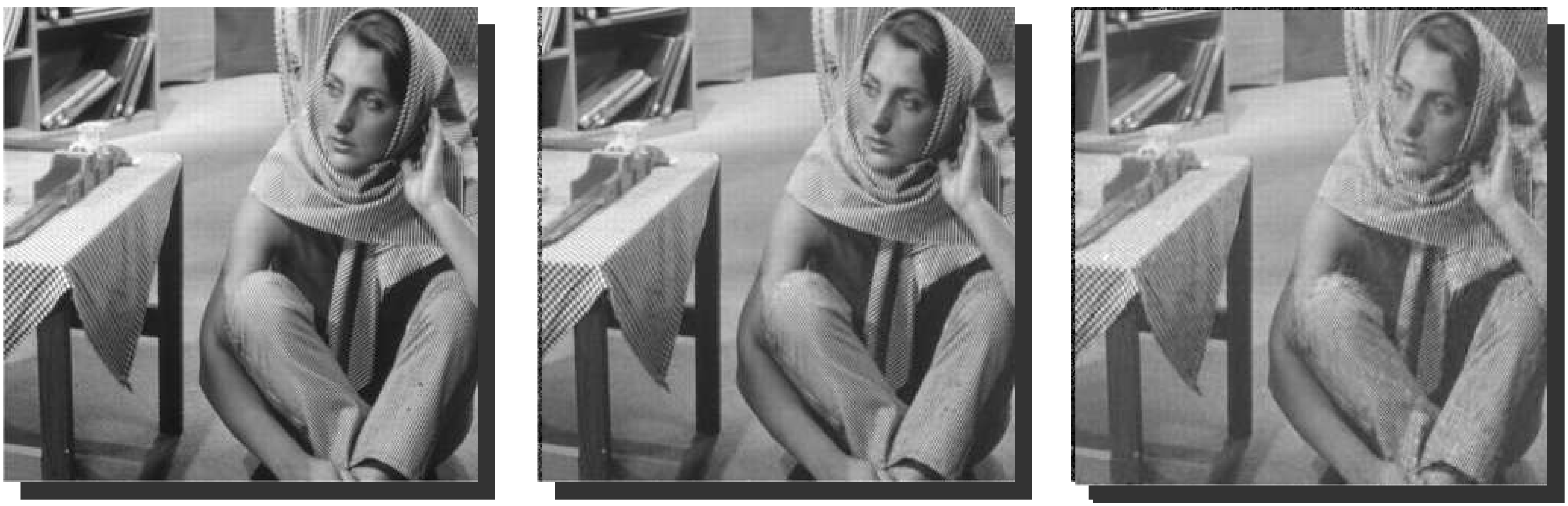
Smoothness



Total Variation

Sparse Representatior

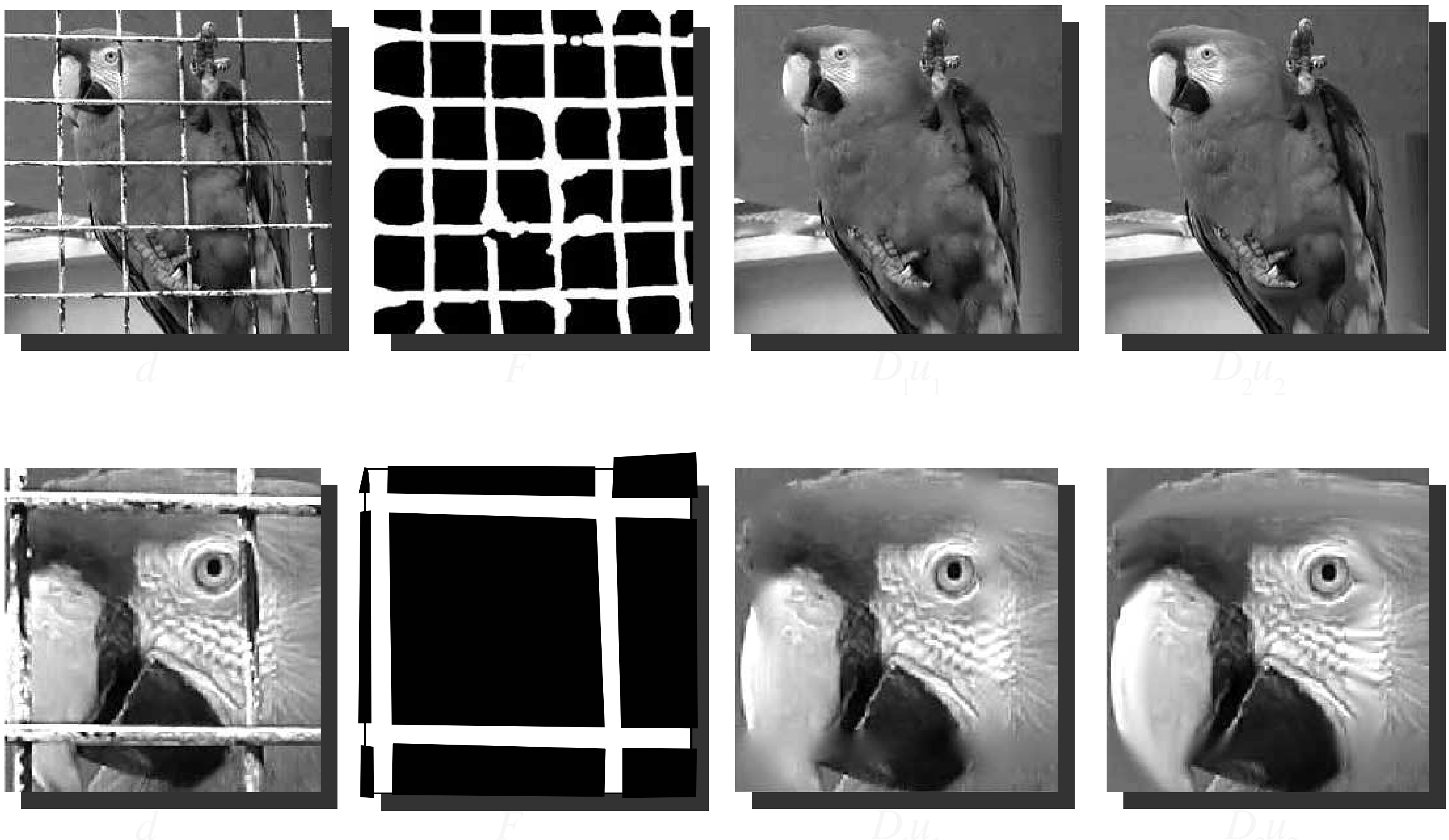
SPARSE REPRESENTATION PERFORMANCE - DIFFERENT OPERATORS



m_T $\quad$ $F_1(mD) + u_1 h = d_1$ $\quad$ $F_2(mD) + u_2 h = d_2$

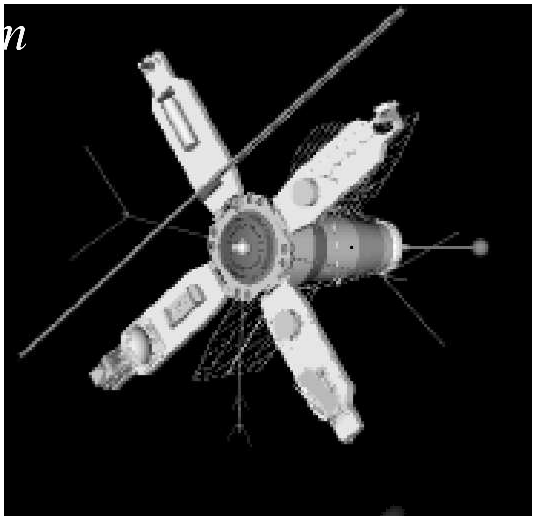
Fadili et al 2007

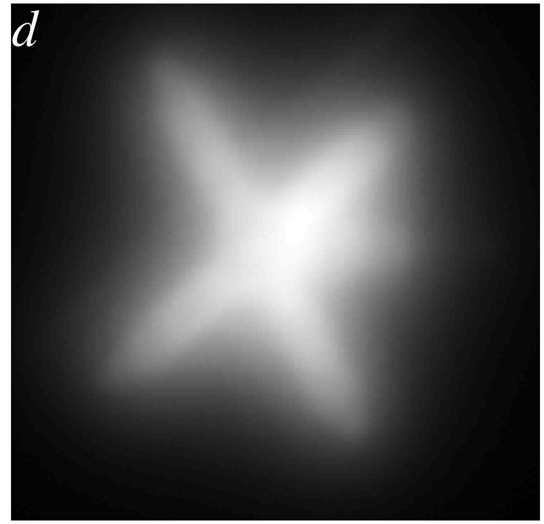
SPARSE REPRESENTATION PERFORMANCE - DIFFERENT DICTIONARIES



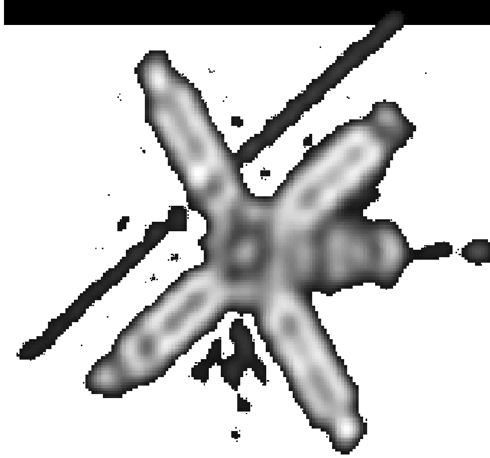
Fadili et al 2007

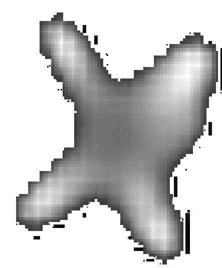
SPARSE REPRESENTATION PERFORMANCE





Singular vectors





Wavelets

Lanczos Hybrid Bidiagonalization Regularization (HyBR)

Gradient Projection Sparse Representation (GPSR)

Chung, Nagy, O'Leary 2006

Figueiredo, Nowak, Wright 2007

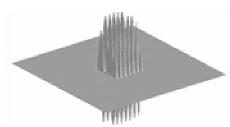
IMPLICIT REGULARIZATION - RATIONALE

- Sometimes sparse representation performs well, sometimes not...

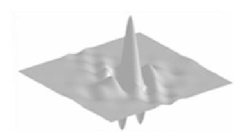
Why?

- Model and operator dependent
- Some dictionaries perform better than others for specific problems

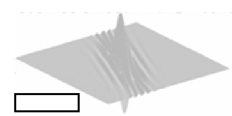
- should be chosen such that it sparsifies the representations



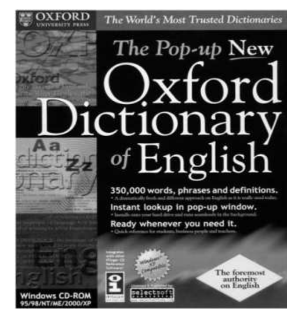
Local DCT
locally oscillatory, stationary texture



Wavelets
piecewise smooth, isotropic structures



Curvelets
piecewise smooth with C^2 contours



- One approach: choose from a known set of transforms (Steerable wavelet, Curvelet, Contourlets, Bandlets, Singular vectors…)

IMPLICIT REGULARIZATION BY DICTIONARY DESIGN

- **Objective** vs. **subjective** function
  - Heuristic choice of regularization functional based on ad-hoc assumptions
  - Solutions are intrinsically subjective to the regularization functional choice

How to construct more objective regularization functionals ?

Design a dictionary by learning from authentic examples $\{m_1, \ldots, m_s\}$

- Adaptability - account for the problem's statistics (model, operator and noise)
- Efficiency and precision - use the right jargon to express a message/model

DICTIONARY DESIGN - PREVIOUS WORK

- Approximated Maximum Likelihood (*Olshusen & Field 1996, 1997*)
- Overcomplete ICA (*Lewicki 2000*)
- Method of Optimal Directions (*Engan et al 2001, 2005*)
- Sparse Bayesian Learning (*Girolami 2001, Wipf 2005*)
- FOCUSS (*Delgado et al 2003*) - Bayesian MAP & relative complexity
- K-SVD (*Aharon & Elad 2006*)
- FOCUSS+ (*Murray & Delgado 2007*)

But

- All addressed sparse coding observation operator was identity F

REGULARIZATION LEARNING – STATISTICAL MERIT

DICTIONARY LEARNING - OPTIMALITY CRITERION

- Loss

$$\mathbf{L}(m, D) := \left\| \hat{m}\left(d(h), D, u\right) - m \right\|_2^2$$

ß Depends on the noise h

ß Depends on an unknown model m

- Mean Square Error

$$MSE\,(m, D) := \mathbf{E}_h \left\| \hat{m}\left(d(h), D, u\right) - m \right\|_2^2$$

ß Depends on an unknown model m

DICTIONARY LEARNING - OPTIMALITY CRITERION

- Bayes risk

$$R_{true}(M, D) := \mathbf{E}_{em} \left\| \hat{m}(D, u) - m \right\|_2^2$$

ß Computationally infeasible

- Bayes empirical risk
  - Assume a set of feasible authentic model examples is available

$$R_{empirical}(m, D) = \mathbf{E}_{h} \sum_{i=1}^{s} \left\| \hat{m}_i(D, u_i) - m_i \right\|_2^2$$

REGULARIZATION LEARNING – OPTIMIZATION FRAMEWORK

OVER-COMPLETE DICTIONARY DESIGN - FORMULATION

- Bi-level optimization problem

$$\hat{D} = \arg\min_{\hat{D}} \frac{1}{s}\mathbf{E}_h \sum_{i=1}^{s} \left\| \hat{m}_i(D,u_i) - m_i \right\|_2^2$$

$$\text{s.t. } u_i = \arg\min_{u_i} \left\| F\left(Du_i; y\right) - d_i\left(y\right) \right\|_2^2 + a\left\|u_i\right\|_1$$

- Non-smooth - norm is replaced by a smooth optimization problem with inequality constraints

- Sensitivity by differentiating the necessary conditions of the decomposition

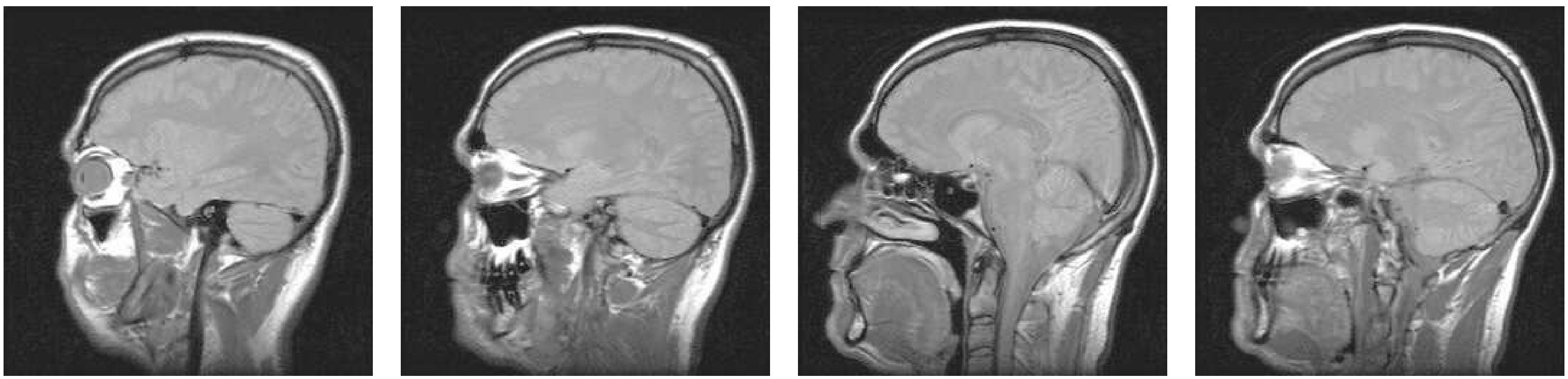
$$\frac{\partial p_l}{\partial D_k} = f\left(D,p,F\right) \qquad \frac{\partial q_l}{\partial D_k} = g\left(D,p,F\right)$$

- Non-smooth optimization framework Modified L-BFGS (*Overton 2003*)

Horesh & Haber 2009

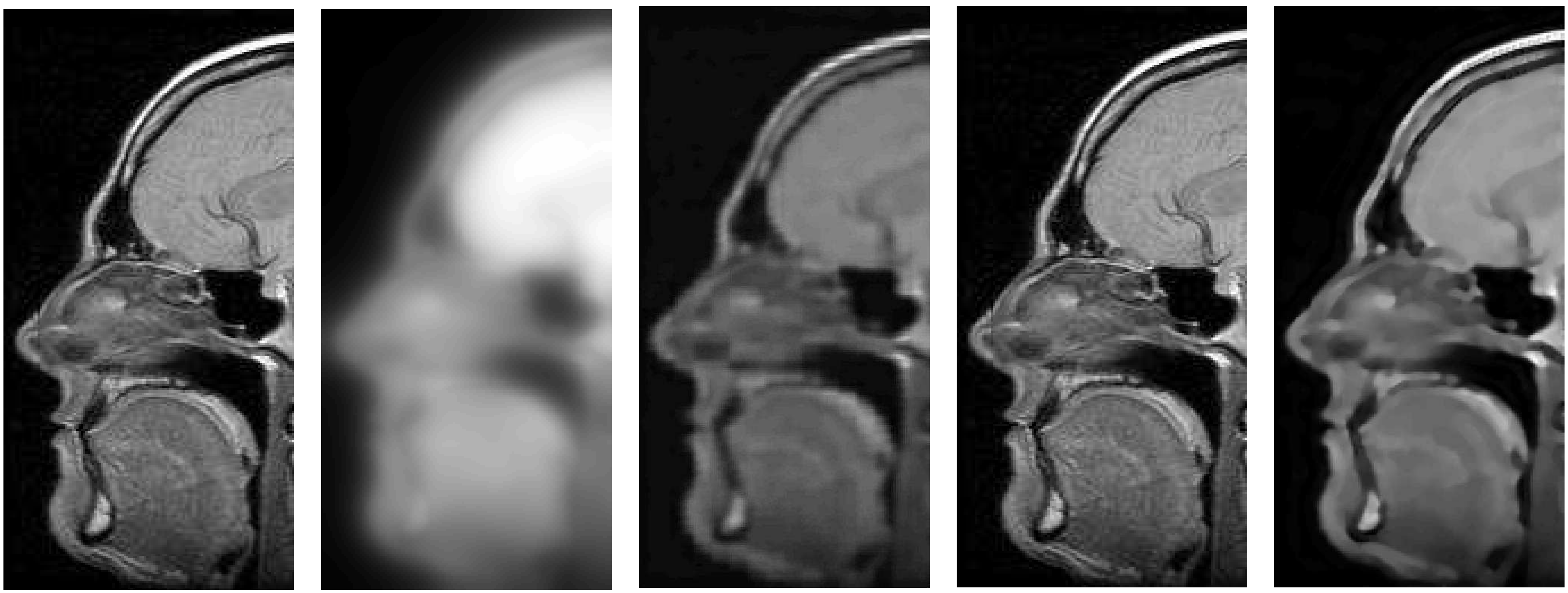
REGULARIZATION DESIGN – NUMERICAL RESULTS

DICTIONARY DESIGN - TRAINING SET



Horesh & Haber 2009

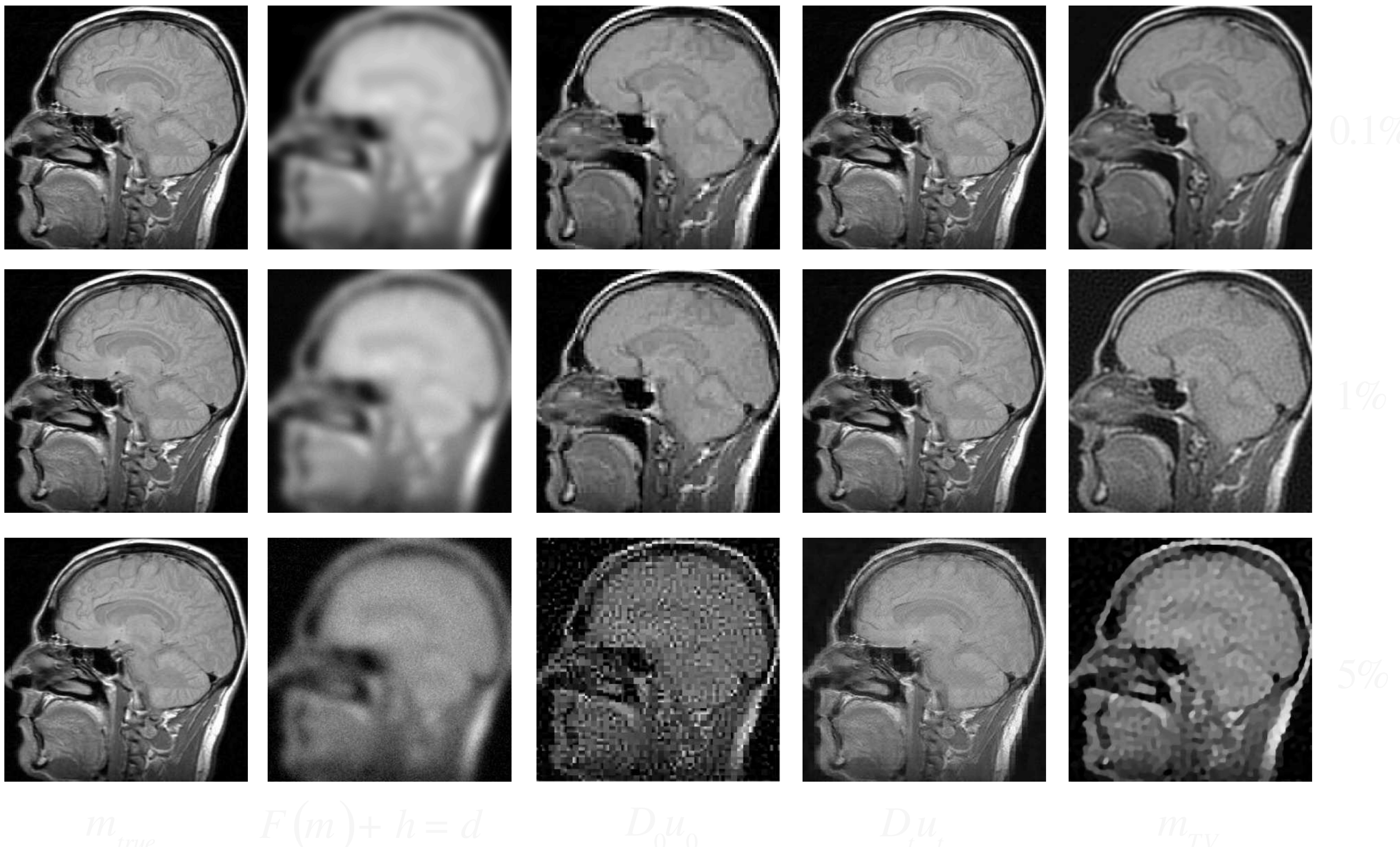
DICTIONARY DESIGN - COMPARISON



m_T $\quad$ $F(m_T) + h = d$ $\quad$ $D_0 u_0$ $\quad$ $D_t u_t$ $\quad$ m_{TV}

Horesh & Haber 2009

DICTIONARY LEARNING – ASSESSMENT WITH NOISE



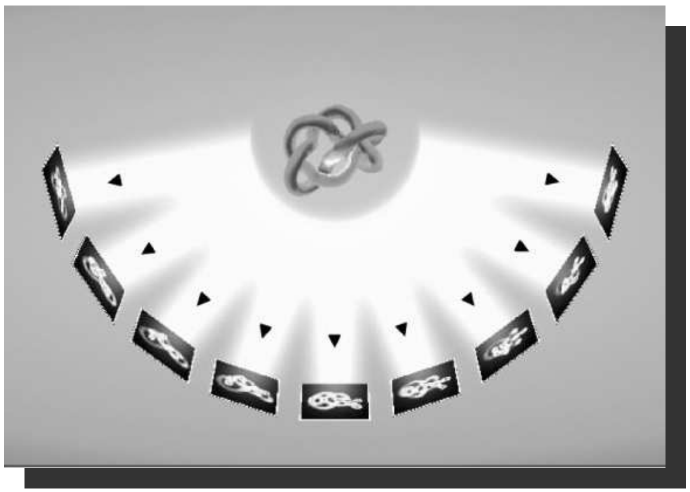
Horesh & Haber 2009

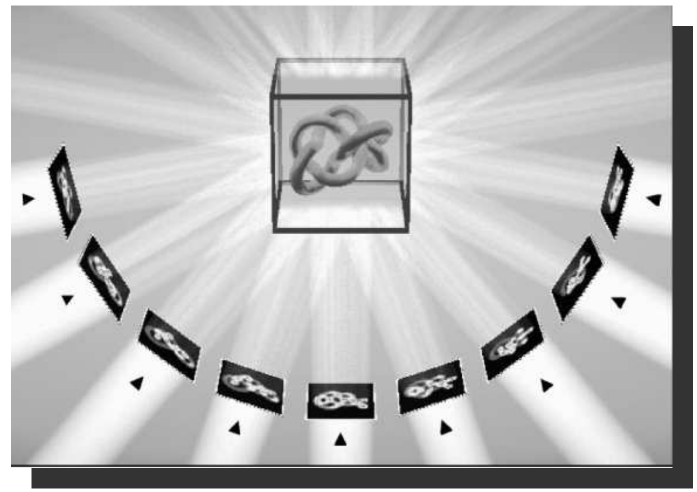
PART II

OPTIMAL EXPERIMENTAL DESIGN

OPTIMAL EXPERIMENTAL DESIGN - MOTIVATION

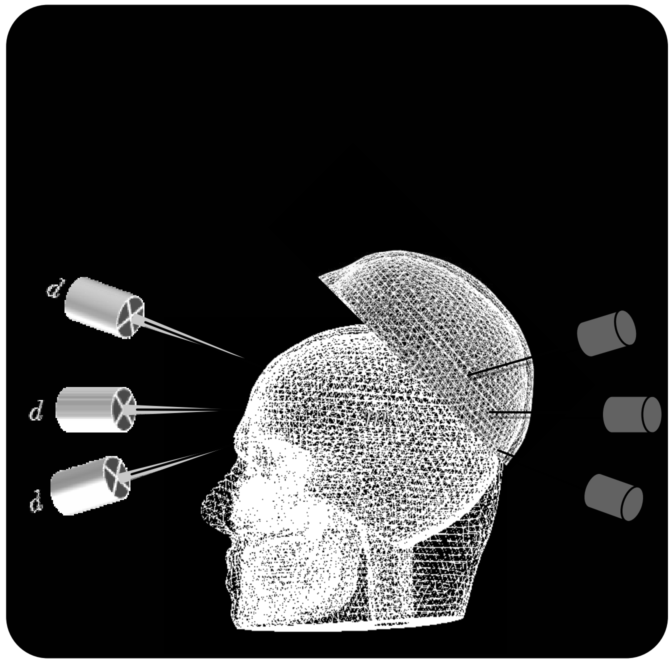
MOTIVATION – LIMITED ANGLE TOMOGRAPHY





Clerbout 2000

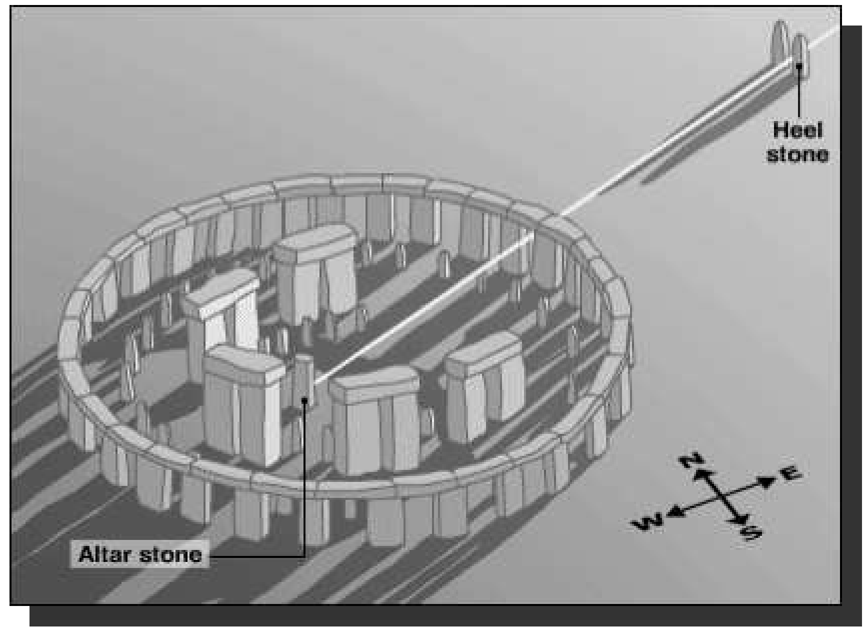
MOTIVATION – DIFFUSE OPTICAL TOMOGRAPHY



MOTIVATION – ULTRASOUND IMAGING

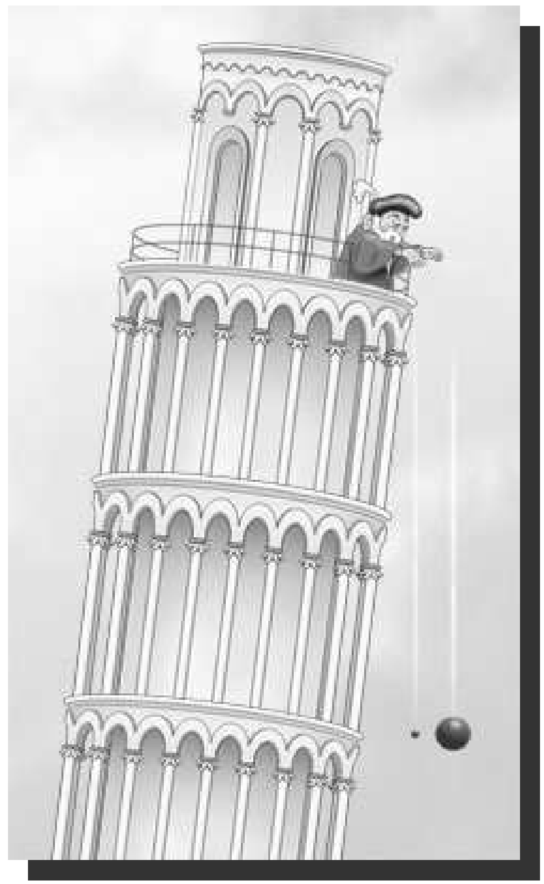
Clerbout 2000

DESIGN EXPERIMENTAL LAYOUT



Stonehenge, 2500 B.C.

DESIGN EXPERIMENTAL PROCESS



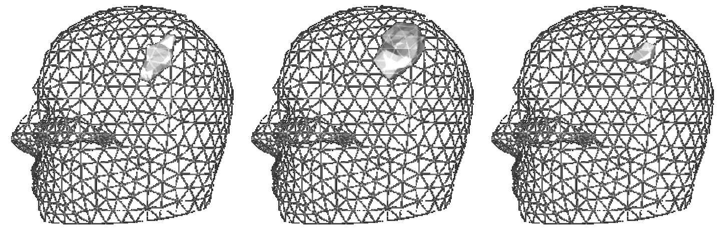
Galileo Galilei, 1564-1642

RESPECT EXPERIMENTAL CONSTRAINTS…

French nuclear test, Mururoa, 1970

OPTIMAL EXPERIMENTAL DESIGN - BACKGROUND

ILL VS. WELL-POSED OPTIMAL EXPERIMENTAL DESIGN



$$\hat{m} = \arg\min \underbrace{\left\| F(m;y) - d(y) \right\|^2}_{data\ fit} + \underbrace{S(m)}_{regularization}$$

- Previous work
  - Well-posed problems - well established (*Fedorov 1997*, *Pukelsheim 2006*)
  - Ill-posed problems - under-researched (*Curtis 1999, Bardow 2008*)

- Many practical problems in engineering and sciences are ill-posed (under-determined)

What makes non-linear ill-posed problems so special ?

OPTIMALITY CRITERIA IN OVER-DETERMINED PROBLEMS

- For linear inversion, employ Tikhonov regularized least squares solution

$$\hat{m} = \underbrace{\left(J^\top J + \alpha L^\top L\right)}_{C(y)}^{-1} J^\top d \qquad J \equiv \frac{\partial F(m,y)}{\partial m}$$

- Bias - variance decomposition

$$\mathbf{E}\left\|\hat{m} - m\right\|_2^2 = \underbrace{\sigma^2 trace\left(J C(y)^{-2} J^\top\right)}_{\text{variance}}$$ [redacted]

- For over-determined problems
- A-optimal design problem

$$\min_{y} \; trace\left(C(y)^{-1}\right)$$

OPTIMALITY CRITERIA IN OVER-DETERMINED PROBLEMS

- Optimality criteria of the information matrix
  - A-optimal design average variance

$$\min_{y} \; trace\left(C(y)^{-1}\right)$$

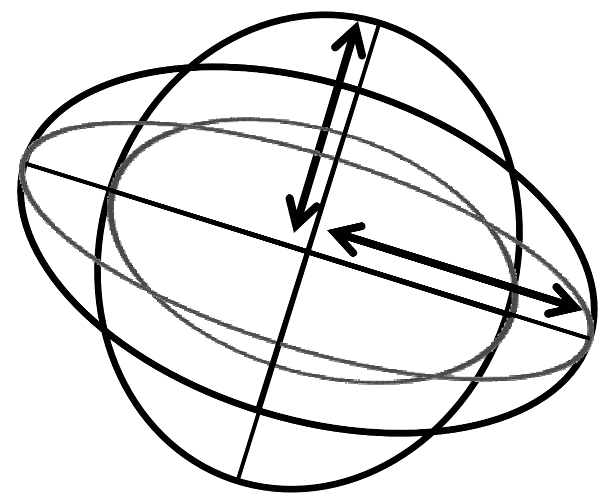
  - D-optimality uncertainty ellipsoid

$$\min_{y} \; \det\left(C(y)^{-1}\right)$$

  - E-optimality minimax

$$\min_{y} \; \max\left(eig\left(C(y)^{-1}\right)\right)$$

  - Almost a complete alphabet…



THE PROBLEM...

- For non-linear ill-posed problems   none of these apply !
  - Non-linearity   bias-variance decomposition is impossible
  - Ill-posedness   controlling variance alone reduces mildly the error

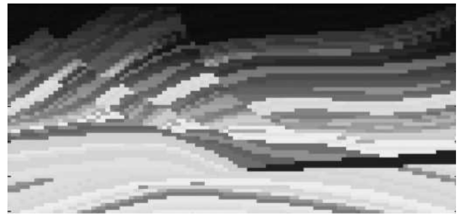
What strategy can be used ?

Proposition 1 - Common practice so far

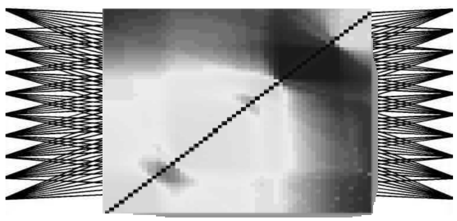
Trial and Error…

EXPERIMENTAL DESIGN BY TRIAL AND ERROR

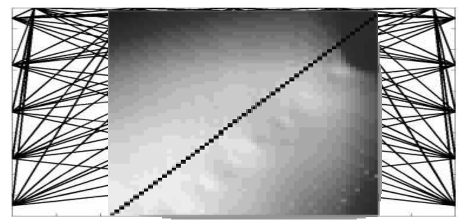
- Pick a model

$m_T =$ 

- Run observation model of different experimental designs, and get data

y_1 

$F(m_T, y_1) + h = d_1$

y_2 

$F(m_T, y_2) + h = d_2$

- Invert and compare recovered models

$\hat{m}_1 =$ 

$\hat{m}_2 =$ 

- Choose the experimental design that provides the best model recovery

THE PROBLEM...

- For non-linear ill-posed problems none of these apply !
  - Non-linearity bias-variance decomposition is impossible
  - Ill-posedness controlling variance alone reduces mildly the error

What **other** strategy can be used ?

Proposition 2 - Minimize bias and variance altogether by some optimality criterion

How to define the optimality criterion ?

Horesh, Haber & Tenorio 2010

OPTIMAL EXPERIMENTAL DESIGN - STATISTICAL MERIT

OPTIMALITY CRITERION

- Loss

$$\mathbf{L}(\mu, y) := \left\| \hat{\mu}(y) - \mu \right\|_2^2$$

  ß Depends on the noise η
  ß Depends on an unknown model μ

- Mean Squared Error

$$MSE(\mu, y) := \mathbf{E}_\eta \left\| \hat{\mu}(y) - \mu \right\|_2^2$$

  ß Depends on an unknown model μ

OPTIMALITY CRITERION

- Bayes risk

$$R_{true\ risk}(M, y) := \mathbf{E}_{em} \left\| \hat{m}(y) - m \right\|_2^2$$

ß Computationally infeasible

- Bayes empirical risk
  - Assume a set of feasible authentic model examples is available

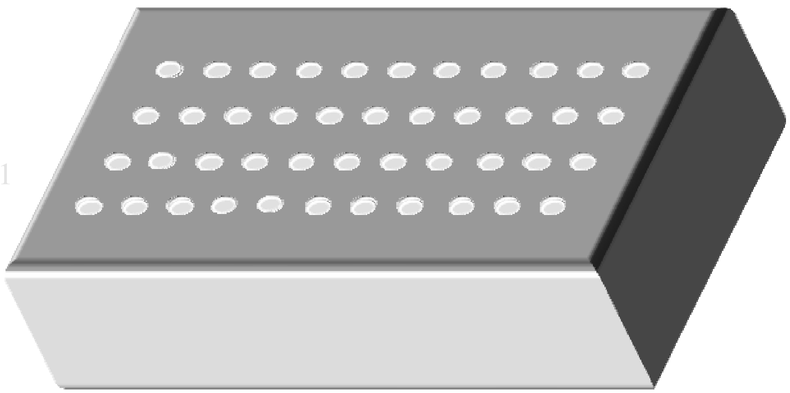
$$R(m, y) := \frac{1}{sk} \sum_{i,j=1}^{k,s} \left\| \hat{m}_{ij}(y) - m_j \right\|_2^2$$

How can be regularized ?

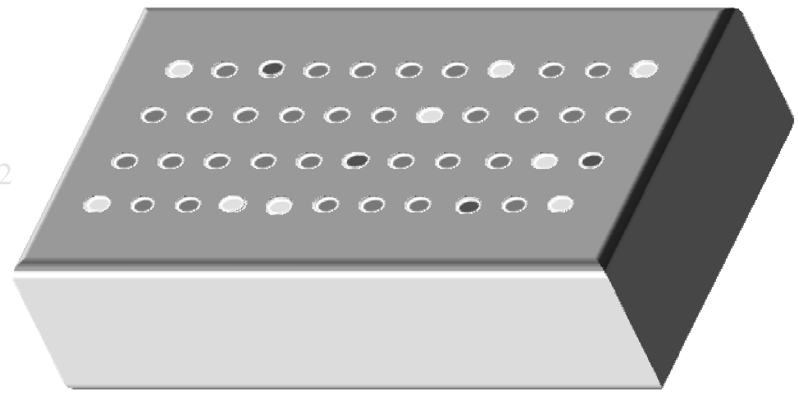
OPTIMALITY CRITERION – SPARSITY CONTROLLED DESIGN

- Regularized empirical risk - Direct density penalty for activation
- Assume: fixed number of experiments

y_1



y_2



- Let y
- The data

$$d(m, y) = F(m, V(y) Q) + \eta = V^{T} A(m)^{-1} Q + \eta$$

- Regularized risk

$$R_{reg}(m, y) := \frac{1}{sk} \sum_{i,j=1}^{k,s} \left\| \hat{m}_{ij}(y) - m_j \right\|_2^2 + \beta \left\| y \right\|_p$$

Horesh, Haber & Tenorio 2011

OPTIMALITY CRITERION – SPARSITY CONTROLLED DESIGN

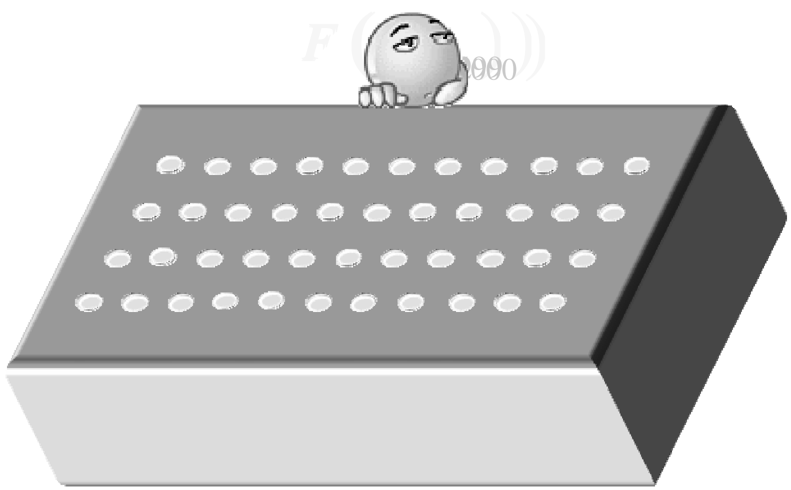
- Regularized empirical risk - Direct approach
  - Total number of experiments may be large

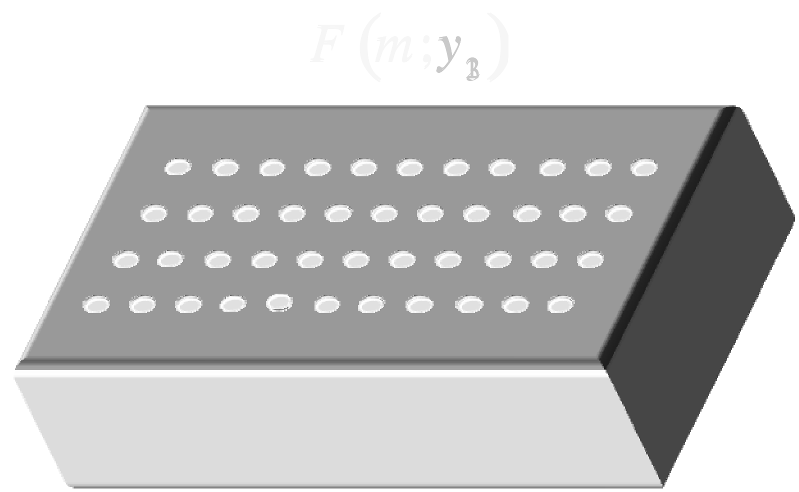
  - Effective when activation of each source and receiver is expensive
  - Derivatives of the forward operator w.r.t. y → Difficult…

Horesh, Haber & Tenorio 2011

OPTIMALITY CRITERION – SPARSITY CONTROLLED DESIGN

- Regularized empirical risk - Weights formulation
- Density penalty over selected experiments from a predefined set





Haber, Horesh & Tenorio 2010

OPTIMALITY CRITERION – SPARSITY CONTROLLED DESIGN

- Regularized empirical risk - Weights formulation
  - Let be discretization of the space
  - Let
  - The observation operator is weighted

$$W^{\frac{1}{2}}\left(F(m)+\eta\right)=W^{\frac{1}{2}}d(m) \qquad W^{\frac{1}{2}}=diag\left(\sqrt{w}\right)$$

  - If experiment is not conducted

$$\mathrm{R}_{reg}(m,w) := \frac{1}{sk}\sum_{i,j=1}^{k,s}\left\|\hat{m}_{ij}(w)-m_j\right\|_2^2+\beta\left\|w\right\|_p, \quad w \ge 0$$

Haber, Horesh & Tenorio 2010

OPTIMALITY CRITERION – SPARSITY CONTROLLED DESIGN

- Regularized empirical risk - Weights formulation
  - Suitable when each experiment conduction is costly
  - Source and receiver activation may be highly populated

  - Less DOF
  - No explicit access to the observation operator needed

Haber, Horesh & Tenorio 2010

OPTIMAL EXPERIMENTAL DESIGN - OPTIMIZATION FRAMEWORK

THE OPTIMIZATION PROBLEMS

- Direct formulation

$$\min_{y} \; \frac{1}{sk}\sum_{i,j=1}^{k,s} \left\|\hat{m}_{ij}(y) - m_j\right\|_2^2 + \beta\left\|y\right\|_p$$

$$\text{s.t.} \quad \hat{m}_{ij} = \arg\min_{\hat{m}_{ij}} \left\|F(m_{ij};y) + \eta_i - F(m_j;y)\right\|_2^2 + S(m_{ij})$$

- Weights formulation

$$\min_{w} \; \frac{1}{sk}\sum_{i,j=1}^{k,s} \left\|\hat{m}_{ij}(w) - m_j\right\|_2^2 + \beta\left\|w\right\|_p$$

$$\text{s.t.} \quad \hat{m}_{ij} = \arg\min_{\hat{m}_{ij}} \left\|W^{\frac{1}{2}}\left(F(m_{ij}) + \eta_i - F(m_j)\right)\right\|_2^2 + S(m_{ij})$$

$$w \geq 0$$

Haber, Horesh & Tenorio 2010 Horesh, Haber & Tenorio 2011

THE OPTIMIZATION PROBLEM

- Bi-level optimization problem

$$\min_{w} \frac{1}{sk}\sum_{i,j=1}^{k,s} \left\| \hat{m}_{ij}(w) - m_j \right\|_2^2 + \beta \left\| w \right\|_p$$

$$\text{s.t.} \quad \hat{m}_{ij} = \arg\min_{\hat{m}_{ij}} \left\| W^{\frac{1}{2}} \left(F(m_{ij}) + \eta_i - F(m_j) \right) \right\|_2^2 + S(m_{ij})$$

$$w \geq 0$$

- Assuming the lower optimization level is:
  - Convex with a well defined minimum
  - With no inequality constraints

$$\min_{y} \frac{1}{sk}\sum_{i,j=1}^{k,s} \left\| m_{ij}(w) - m_j \right\|_2^2 + \beta \left\| w \right\|_p$$

$$\text{s.t.} \quad c_{ij} \equiv c(m_{ij}, m_j, w) = J(m_{ij})^{\top} W \left(F(\hat{m}_j) + \eta_i - F(m_j) \right) + S'(m_{ij}) = 0$$

$$w \geq 0$$

Haber, Horesh & Tenorio 2010

THE OPTIMIZATION PROBLEM

- m is eliminated from the equations and viewed as a function of
- Compute gradient by implicit differentiation

$$\frac{\partial c_{ij}}{\partial m_{ij}} = J(m_{ij})^{T} W J(m_{ij}) + S'(m_{ij}) + K_{ij}$$

$$\frac{\partial c_{ij}}{\partial w} = J(m_{ij})^{T} \operatorname{diag}\left(F(m_{ij}) - d_{ij}\right)$$

- The sensitivity

$$M_{ij} := \frac{\partial m_{ij}}{\partial w} = -\left(\frac{\partial c_{ij}}{\partial m_{ij}}\right)^{-1} \frac{\partial c_{ij}}{\partial w}$$

- The reduced gradient

$$\nabla_w R_b\left(w, m_{ij}(w)\right) = \frac{1}{LK} \sum_{i,j} M_{ij}^{T} \left(m_{ij} - m_i\right) + \beta e$$

Haber, Horesh & Tenorio 2010

OPTIMAL EXPERIMENTAL DESIGN – NUMERICAL STUDIES

IMPEDANCE TOMOGRAPHY – OBSERVATION MODEL

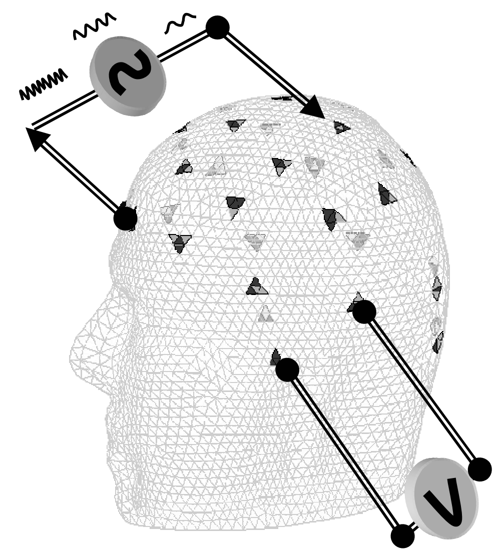
- Governing equations

$$\nabla \cdot (m \nabla u) = 0 \quad \text{in} \quad \Omega$$
$$B.C. \quad \text{on} \quad \partial\Omega$$

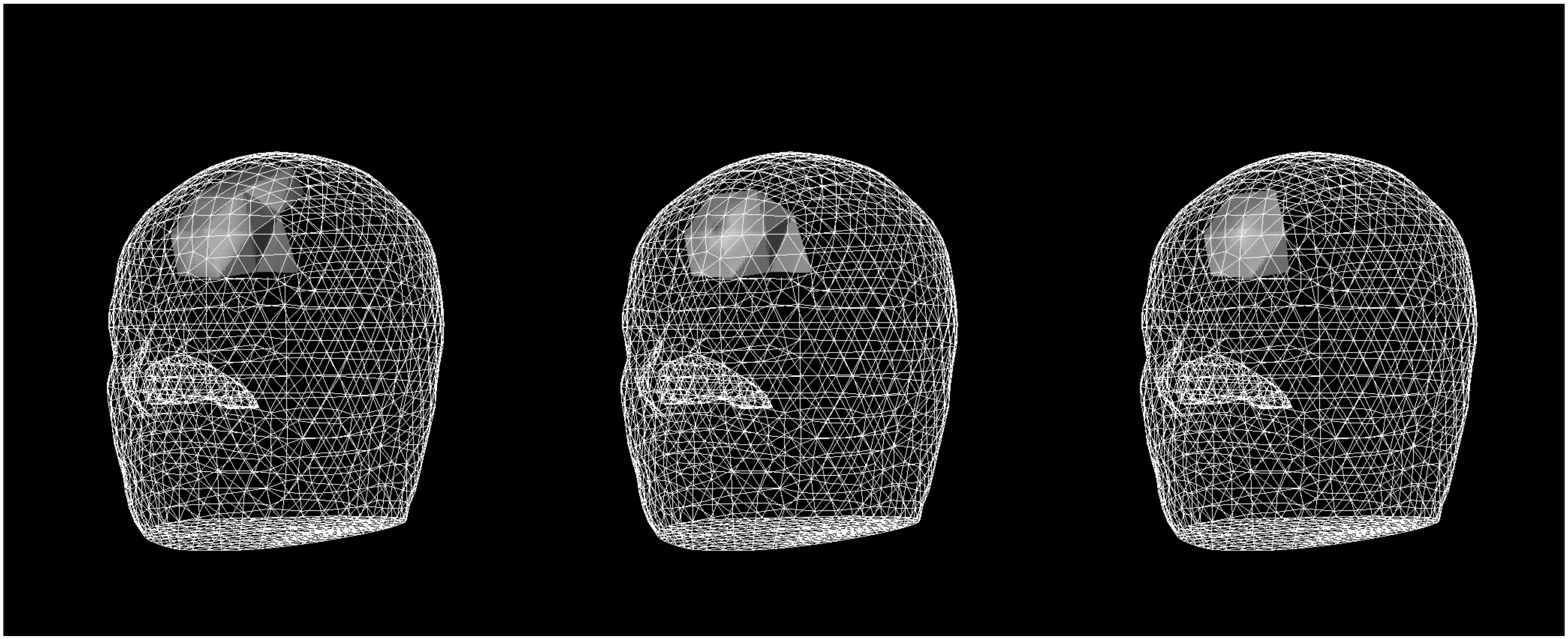
- Following Finite Element discretization
  - Given model and design settings
  - Find data ,

$$A(m)u = Q$$

$$d(m,y) = F(m, \underbrace{V, Q}_{y}) + \eta = V^{\top} A(m)^{-1} Q + \eta$$



IMPEDANCE TOMOGRAPHY – DESIGNS COMPARISON



Naive design True model Optimized design

Horesh, Haber & Tenorio 2011

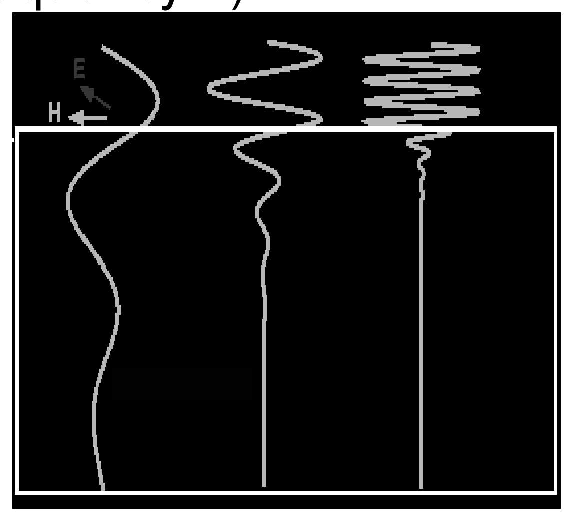
MAGNETO-TULLERICS TOMOGRAPHY – OBSERVATION MODEL

- Governing equations

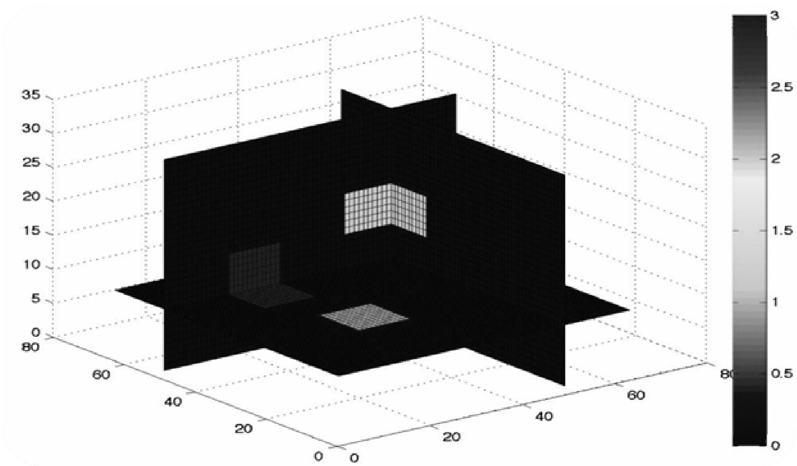
$$\nabla \times \mu_r^{-1} \nabla \times \vec{E} - i\omega\mu\vec{E} = i\omega\vec{s} \qquad \text{in} \quad \Omega$$
$$\nabla \times \vec{E} \times \hat{n} = 0 \qquad \text{on } \partial\Omega$$

- Following Finite Volume discretization
  - Given: model and design settings (frequency)
  - Find: data ,

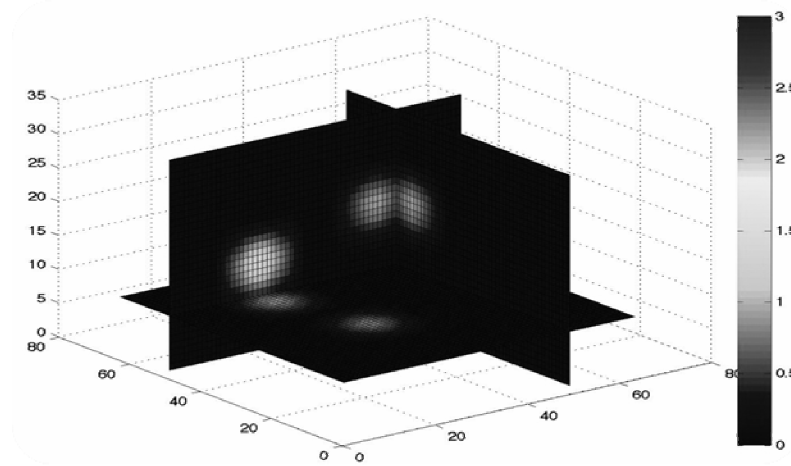
$$d(m;\omega) = V_\omega^{\top} \underbrace{A_\omega(m;\omega)^{-1} i\omega s}_{\psi} + \eta$$



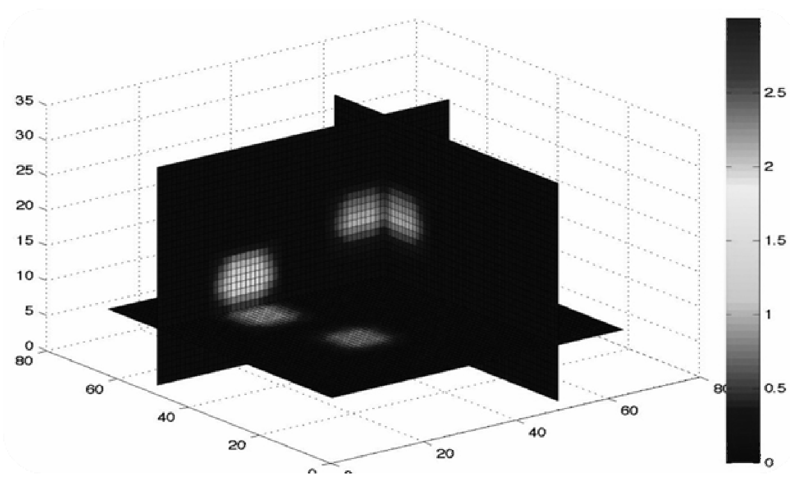
MAGNETOTELLURICS TOMOGRAPHY – DESIGNS COMPARISON



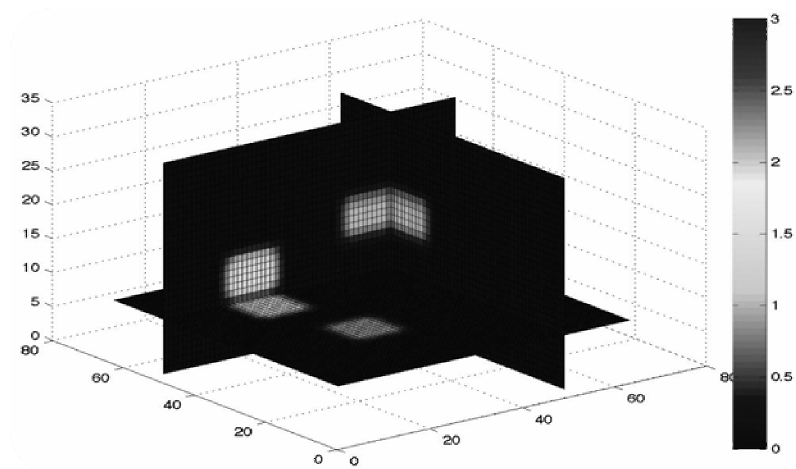
True model



Naive design



Optimal linearized design

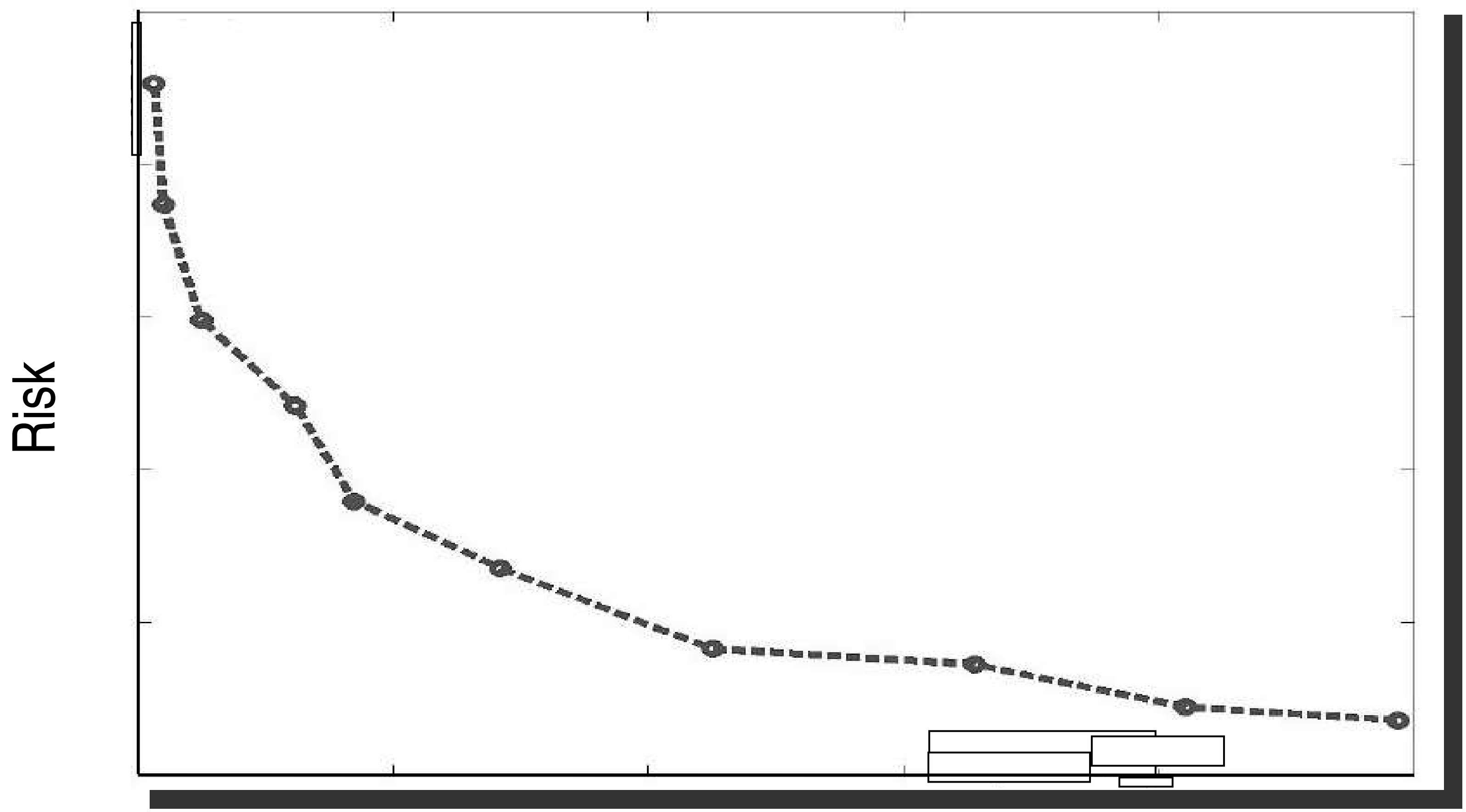


Optimized non-linear design

Haber, Horesh & Tenorio 2008

Haber, Horesh & Tenorio 2010

THE PARETO CURVE – A DECISION MAKING TOOL



Haber, Horesh & Tenorio 2010

SUMMARY

SUMMARY

- Generic approaches for design in ill-posed inverse problems
  - Design of adaptive regularization
  - Optimal experimental design

- Only two (important) elements in the big puzzle...

- New frontiers in inverse problems and optimization

- Vast range of applications in medical imaging, that offers:
  - Faster
  - Safer
  - Higher fidelity image reconstructions

ACKNOWLEDGMENTS

DESIGN IN INVERSION – OPEN COLLABORATIVE RESEARCH

- IBM Research
- MITACS
- University of British Columbia

ACKNOWLEDGMENTS

Michele Benzi

Andy Conn

Eldad Haber

Thank you!

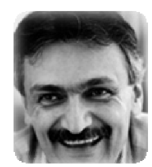
Questions?

Michael Henderson

Raya Horesh

Ulisses Mello

Jim Nagy



David Nahamoo