Semi-definite Programming and Nuclear Magnetic Resonance

Christopher Kumar Anand
Anuroop Sharma

http://sqrl.mcmaster.ca/~anand/papers/preprints.html

Advanced Optimization Laboratory

The McMaster School of Computational Engineering and Science

McMaster University
Inspiring Innovation and Discovery

SDP + NMR

- Applications
  - Imaging
  - NMR
- Problem: Maximize Signal
- Trust-Region Method + CSDP
- Results

Applications: Imaging

- Fourier Transform
  - MRI
- Radon Transform
  - CT, PET, SPECT, EPR

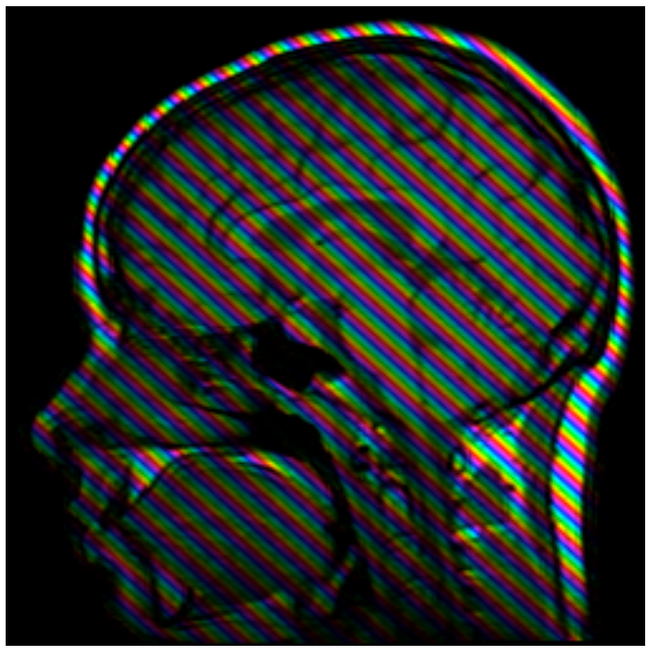
MRI

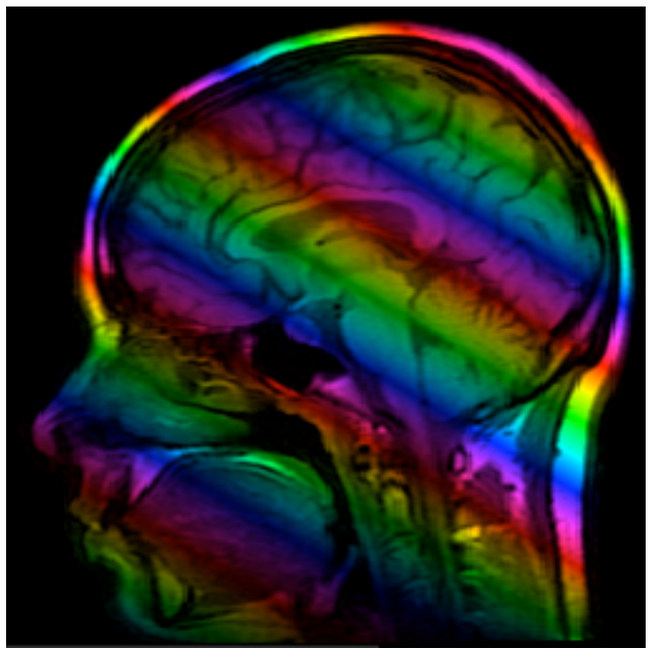
sampled Fourier Transform

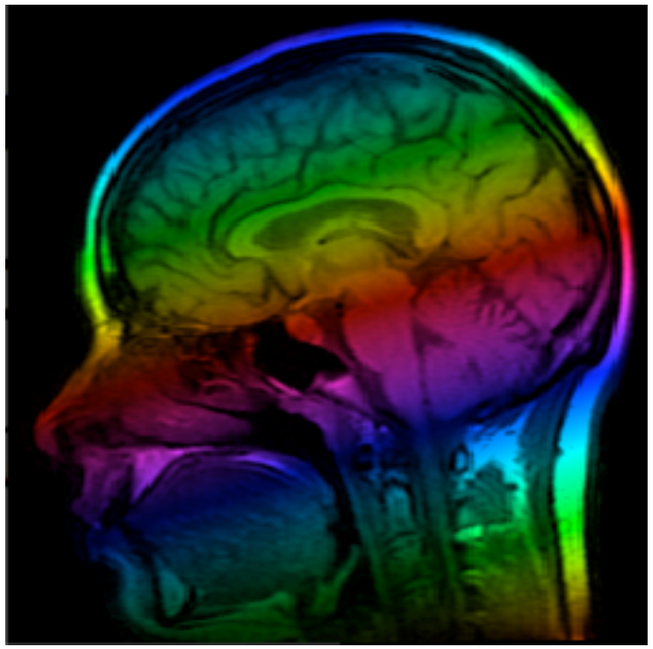
linear phase variation

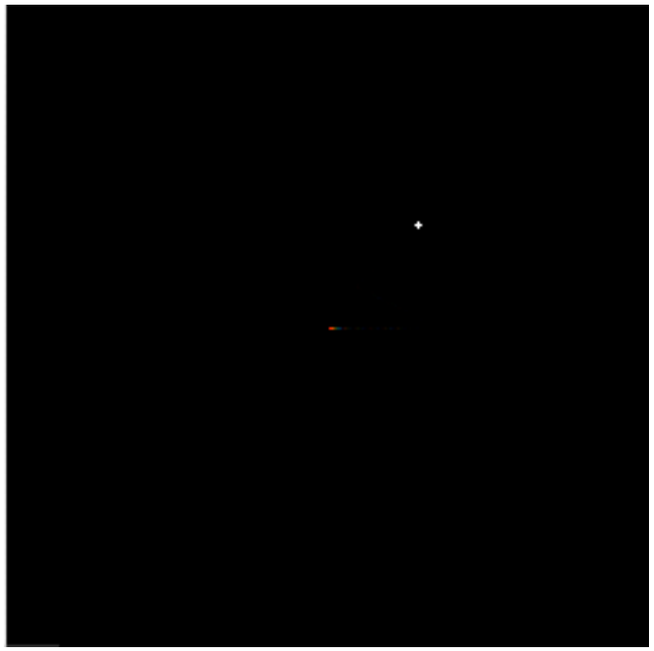
$$s(t) = \int_{\mathbb{R}^3} e^{i\langle x, k(t)\rangle} \rho(x) dx,$$

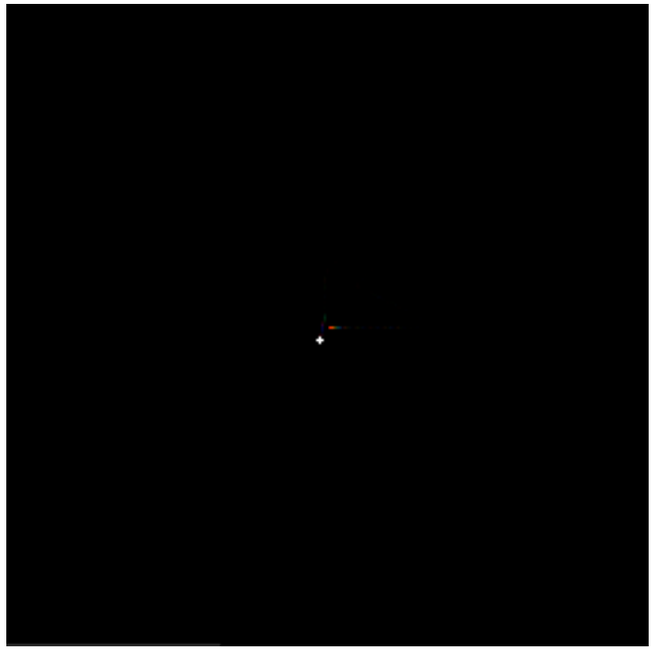
model

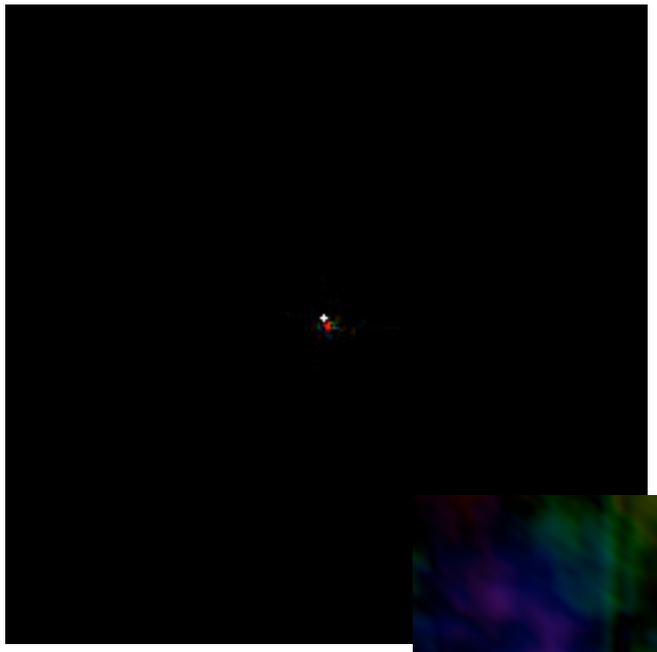


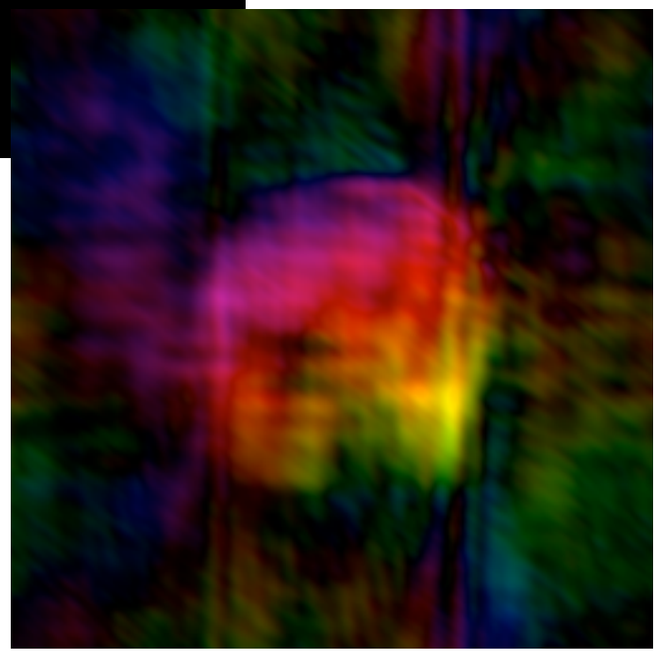




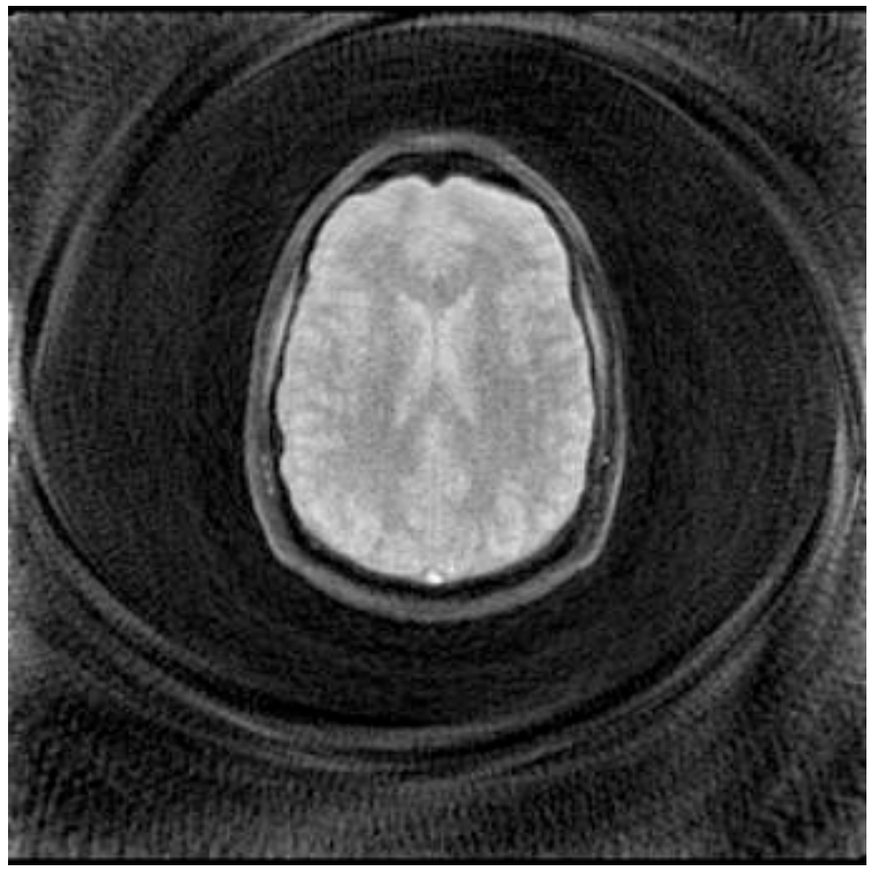






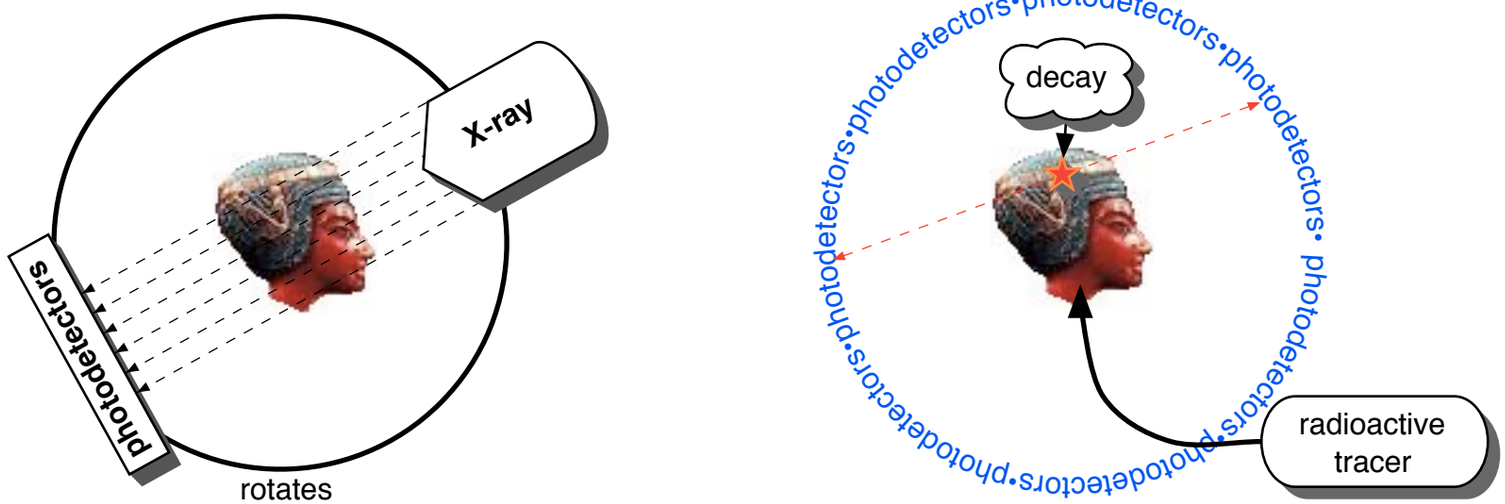


Different Under Sampling



Different Errors

Line-of-Sight



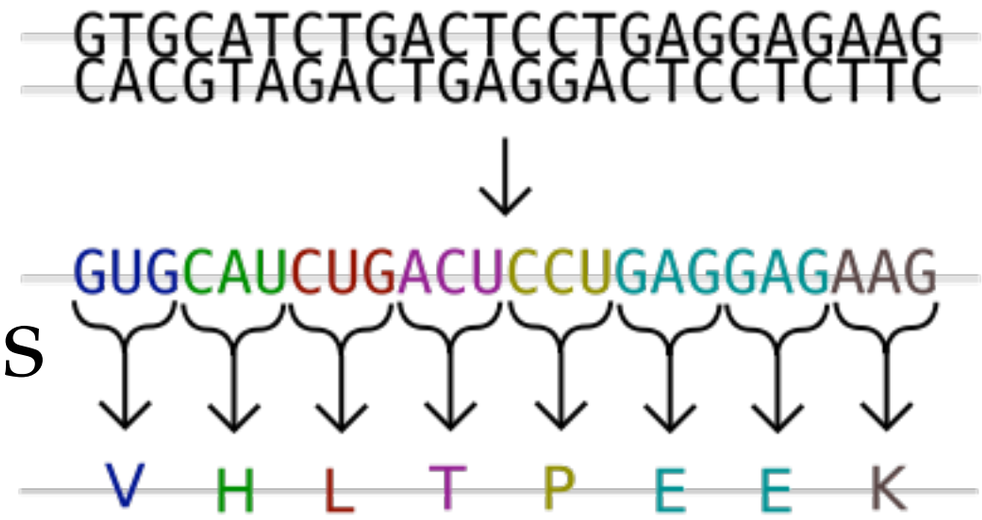
- CT, PET, SPECT, EPR sample projections
- FT(1d Projection) = FT(image) $|_{\text{line}}$
- Constraint on Sampling

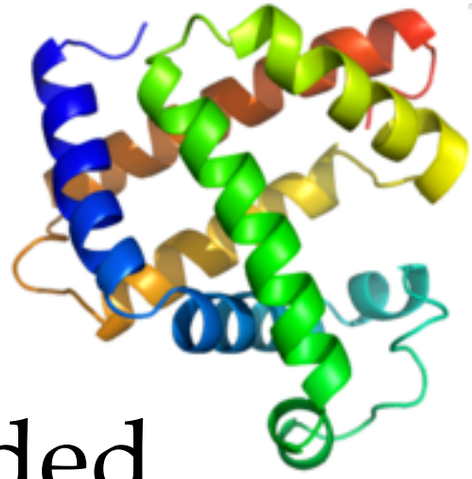
Problem Sneak Peak

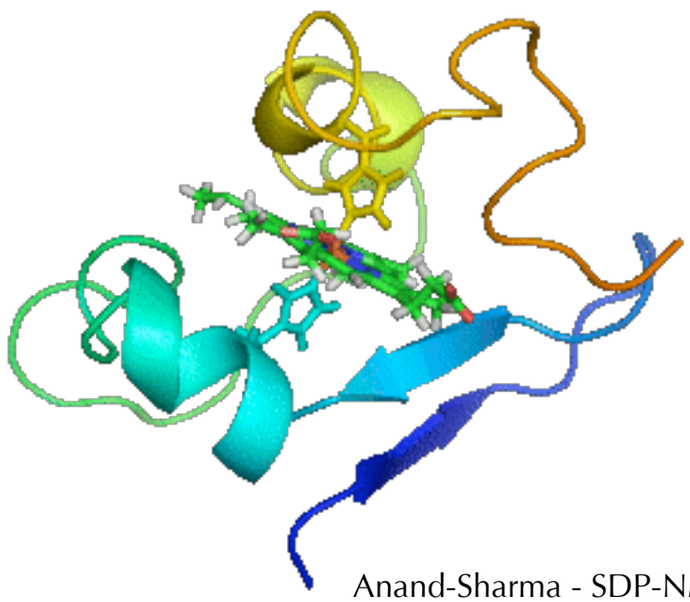
- Inverse Problem with Noisy Data
- Minimize Expected Reconstructed Noise

Application: NMR

- Know DNA Sequences
- Defines Strings of Amino Acids
- Missing Info:
  - Protein Structure
    - only works if folded
  - **Protein Function**
    - **interaction = wiggling**





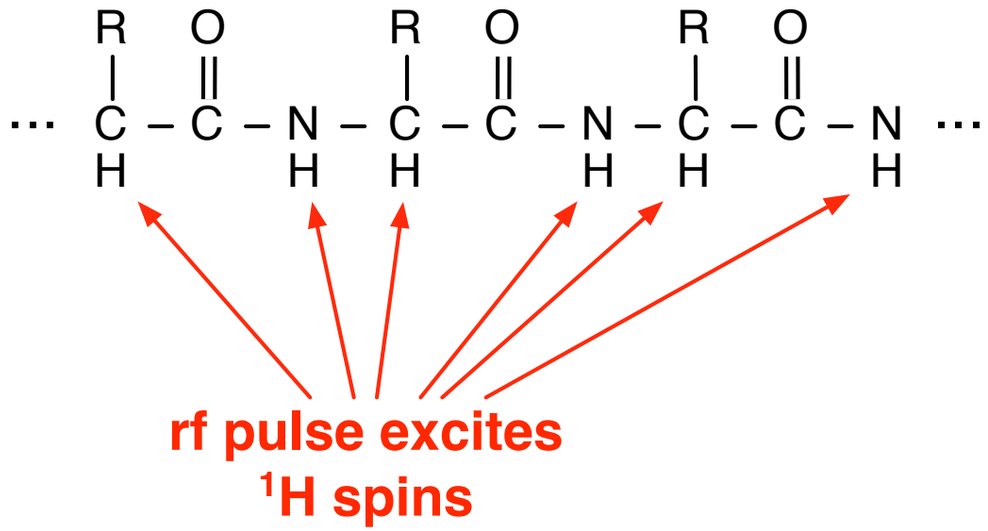


http://en.wikipedia.org/wiki/Protien

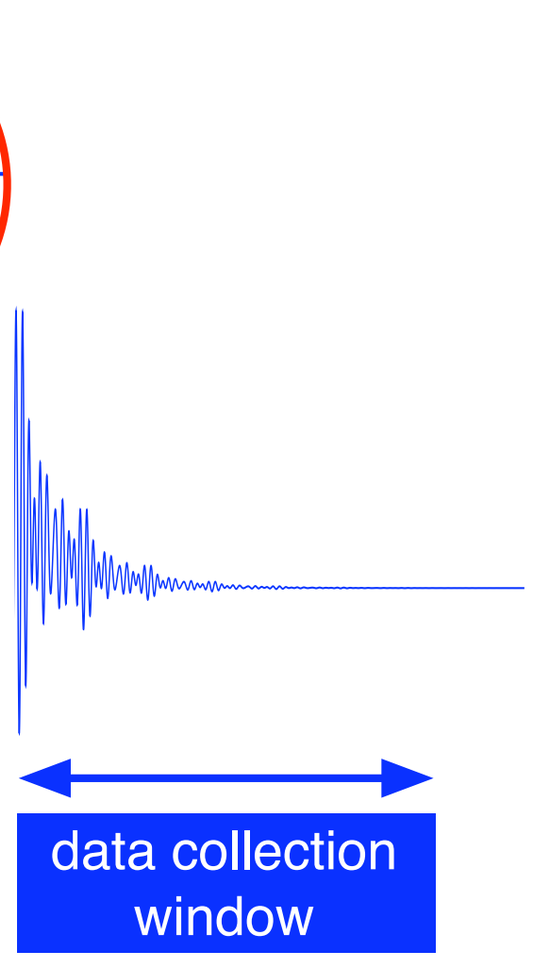
Protein NMR

- Protein Structure
  - 2 methods:
    - heteronuclear, multi-dimensional NMR
    - x-ray Crystallography (faster?)
- Protein Dynamics
  - 1 method
    - repeated n-d NMR

Basic NMR

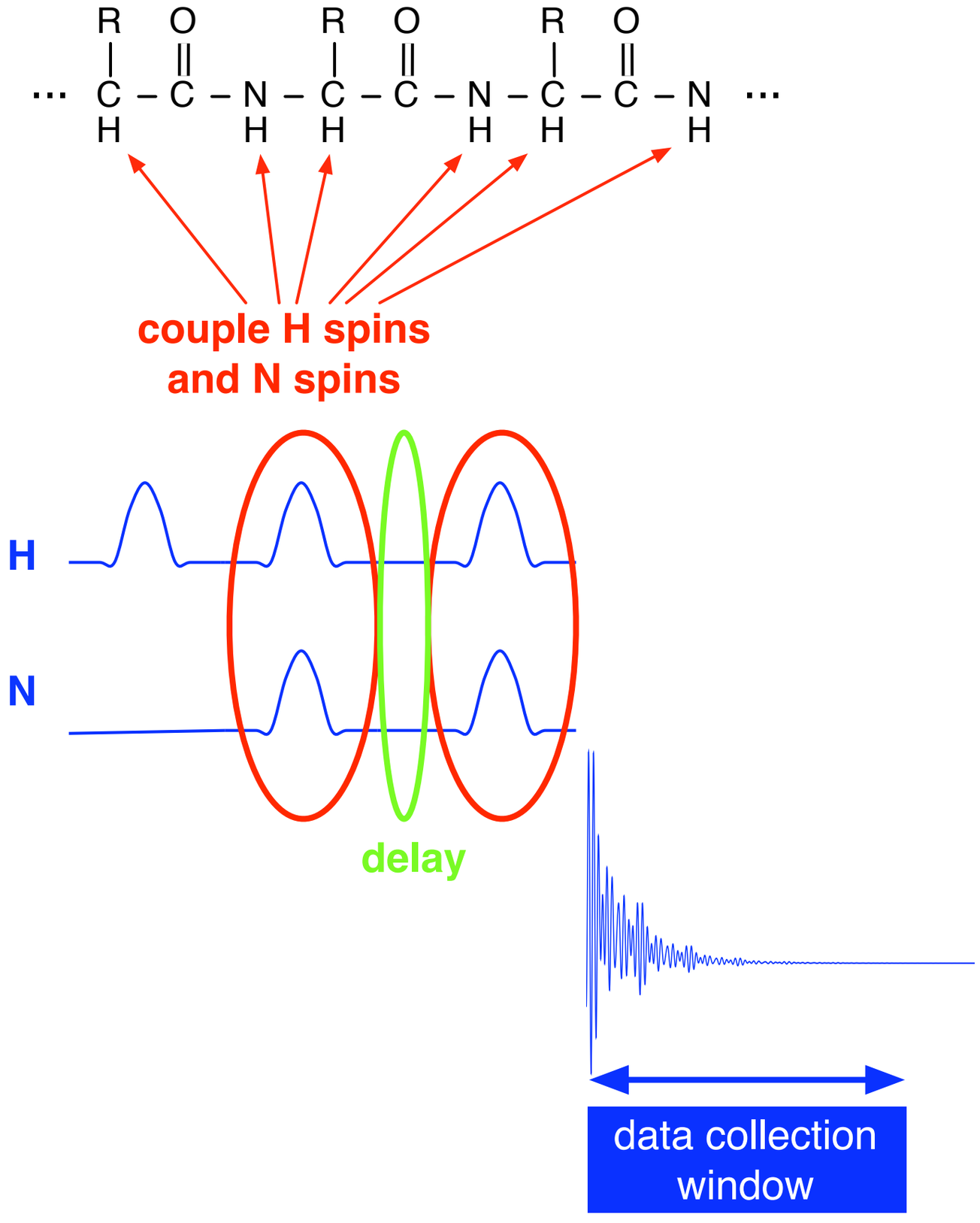


- radio-frequency magnetic field excites spins
- high-energy state decays producing free-induction decay (direct dimension)

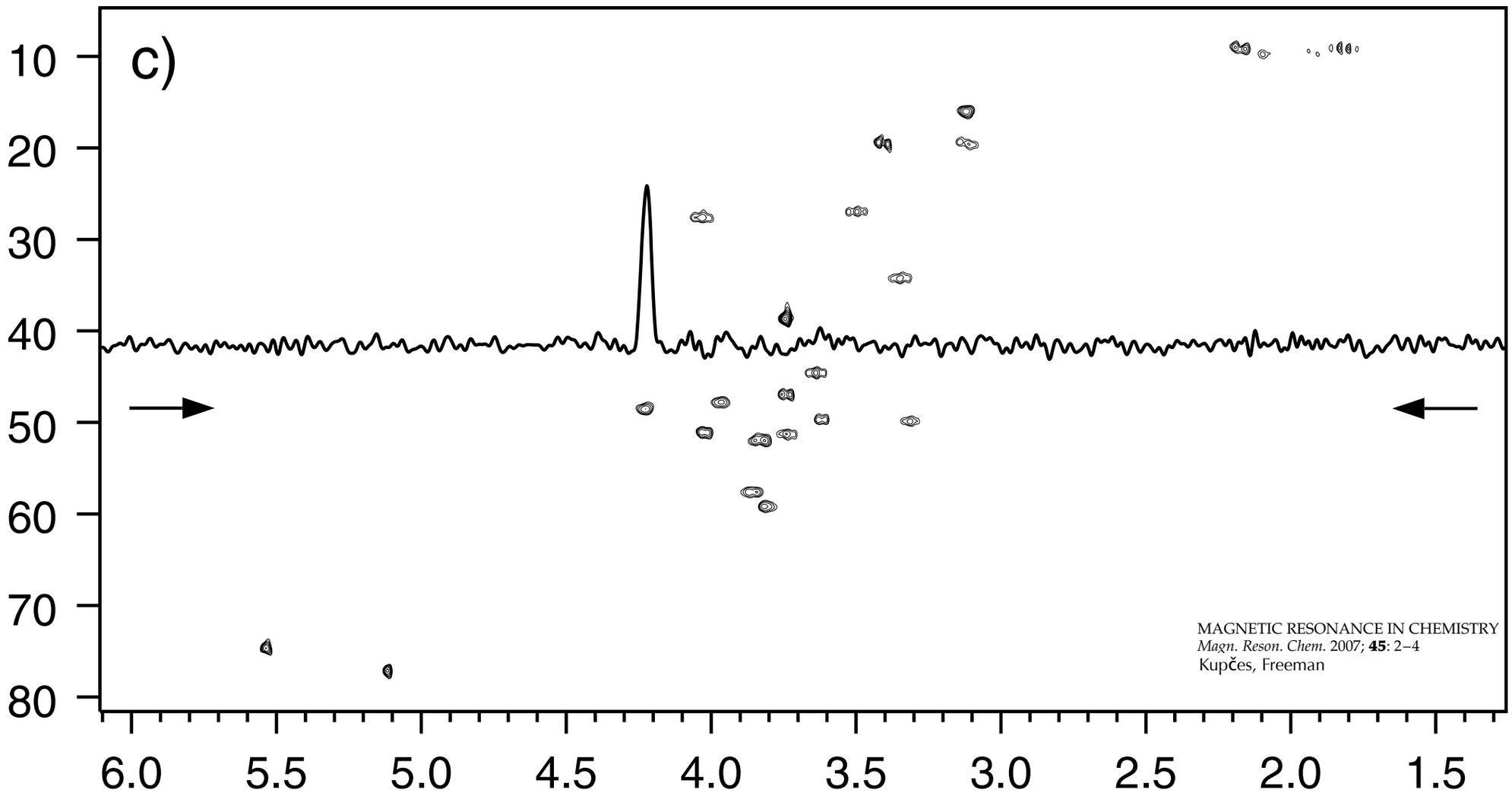


2-d NMR

- pulse @ 2 frequencies
- transfer spin state H-N-H
- phase variation proportional to delay (indirect)



2-d C-H



- regular sampling + FFT
- contour with x-section
- clustered peaks (lorentzian or gaussian)

n-d NMR

- induction on number of colours in diagram
- since delays are positive, can only sample positive values in indirect dimensions
- sampling in indirect dimensions is expensive
  - days, weeks, months, years for full

Protein Dynamics

- in n-d 1 peak per residue
- no overlap for n=4,5,...,10 ?
- measure with additional delay
  - signal of each peak decays exponentially
  - rate linked to mobility

Questions

- How many samples do we need to estimate 200 peak areas?
- Is regular sampling for FFT optimal?

Slow FT Inverse Problem

$$\tilde{f}(k_i) = \sum_{j=1}^{m} f(x_j) e^{\sqrt{-1}\langle k_i, x_j \rangle}$$

- x - known peaks
- k - samples
- general form still a linear system (S)
- Moore-Penrose

$$\begin{pmatrix} \tilde{f}(k_1) \\ \vdots \\ \tilde{f}(k_n) \end{pmatrix} = S \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$S_{i,j} = e^{\sqrt{-1}\langle k_i, x_j \rangle}$$

$$\begin{pmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{pmatrix} = (S^* S)^{-1} S^* \begin{pmatrix} \tilde{f}(k_1) \\ \vdots \\ \tilde{f}(k_n) \end{pmatrix}$$

Noise ~ Conditioning

$$(S^*S)_{i,j} = \sum_{l=1}^{n} e^{\sqrt{-1}\langle k_l, x_j - x_i \rangle}$$

- expected maximum error
  ~ 1 / minimal eigenvalue
- leads to semi-definite constraint

see

Real Nonlinear Problem

$$\begin{aligned} \min_{\{k_i\}} \quad & -\lambda \\ \text{subject to} \quad & A - \lambda I \succeq 0 \end{aligned}$$

$$A_{2i-1,2j-1} = \sum_{l=1}^{n} \cos\langle k_l, x_j - x_i \rangle$$

$$A_{2i,2j} = \sum_{l=1}^{n} \cos\langle k_l, x_j - x_i \rangle$$

$$A_{2i,2j-1} = \sum_{l=1}^{n} \sin\langle k_l, x_j - x_i \rangle$$

$$A_{2i-1,2j} = -\sum_{l=1}^{n} \sin\langle k_l, x_j - x_i \rangle$$

Trust Region

- general non-linear solvers do not use semi-definite cone structure
- use trust region method with linear problem
- shape trust region relative to sensitivity

Linear Subproblem

$$\min_{k} \quad -\lambda$$

$$\text{subject to} \quad A|_{\tilde{k}} + \sum_{\substack{\alpha=1\ldots n \\ \beta=1\ldots r}} (k_{\alpha,\beta} - \tilde{k}_{\alpha,\beta}) \left.\frac{\partial A}{\partial k_{\alpha,\beta}}\right|_{\tilde{k}} - \lambda I \succeq 0$$

$$\frac{\partial A_{2i-1,2j-1}}{\partial k_{\alpha,\beta}} = -(\sin\langle k_\alpha, x_j - x_i\rangle)(x_{j,\beta} - x_{i,\beta})$$

$$\frac{\partial A_{2i,2j}}{\partial k_{\alpha,\beta}} = -(\sin\langle k_\alpha, x_j - x_i\rangle)(x_{j,\beta} - x_{i,\beta})$$

$$\frac{\partial A_{2i,2j-1}}{\partial k_{\alpha,\beta}} = \quad (\cos\langle k_\alpha, x_j - x_i\rangle)(x_{j,\beta} - x_{i,\beta})$$

$$\frac{\partial A_{2i-1,2j}}{\partial k_{\alpha,\beta}} = -(\cos\langle k_\alpha, x_j - x_i\rangle)(x_{j,\beta} - x_{i,\beta})$$

$$|k_{\alpha,\beta} - \tilde{k}_{\alpha,\beta}| \leq \frac{\pi/2}{\max|x_{j,\beta} - x_{i,\beta}|}$$

Implementation

- C program calling CSDP for subproblem
- use random, and greedy-random seeding in incremental and one-step solvers

Hyperplane Decomposition

- CSDP cannot solve full problem
- use dense H-freq sampling
- reduce dimension of parameters and variables
- only optimize hyperplanes with peaks

Hyperplane Decomposition - II

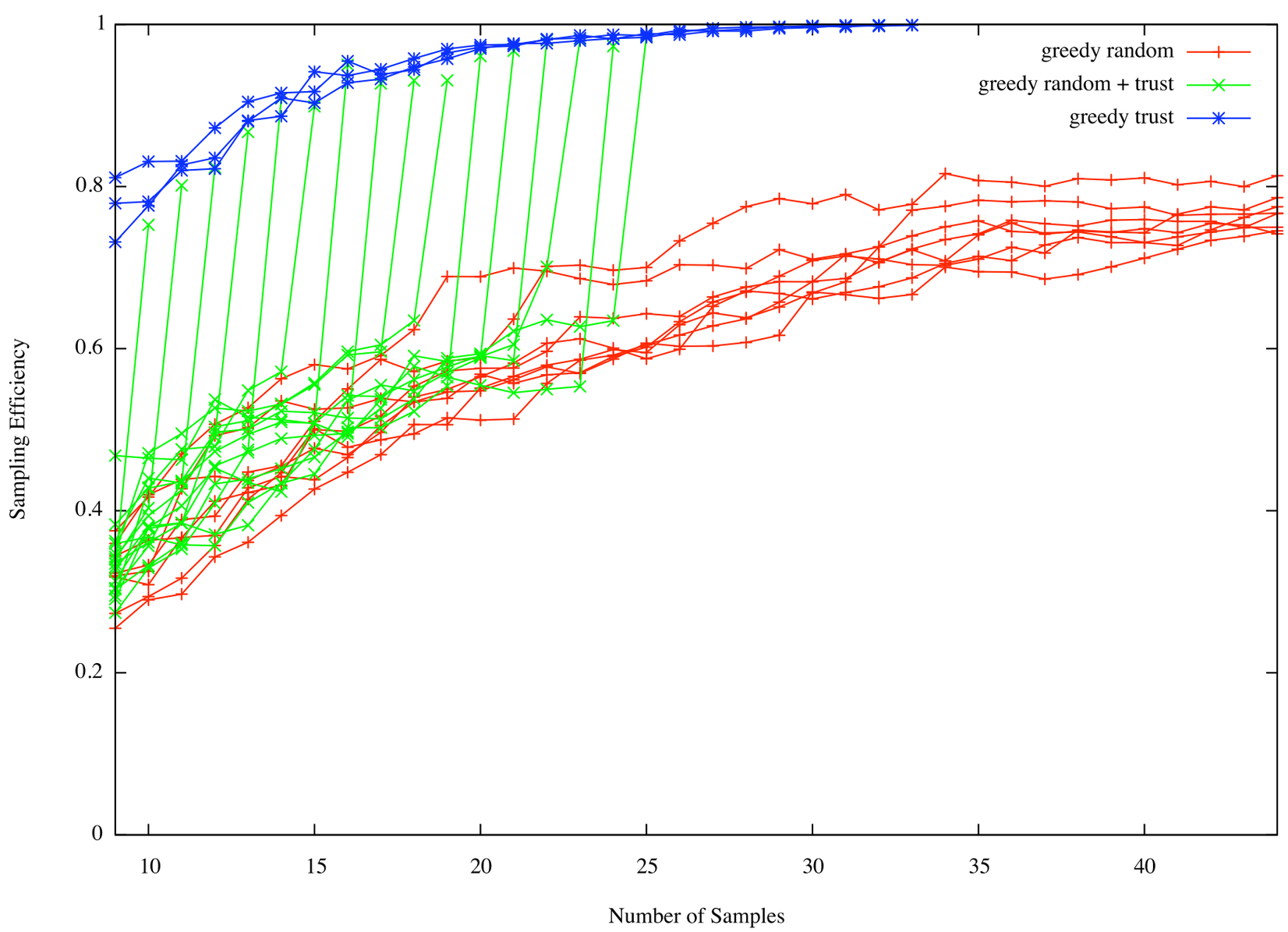
$$(S^*S)_{i,j} = \begin{cases} \sum_{l=0}^{n} e^{\sqrt{-1}\langle k_l, x_j - x_i \rangle} & \text{if } x_j, x_i \text{ belongs to plane } l \\ 0 & \text{otherwise} \end{cases}$$

- block diagonal structure
- still to large
- separate blocks (hyperplanes) into independent problems

Numerical Tests

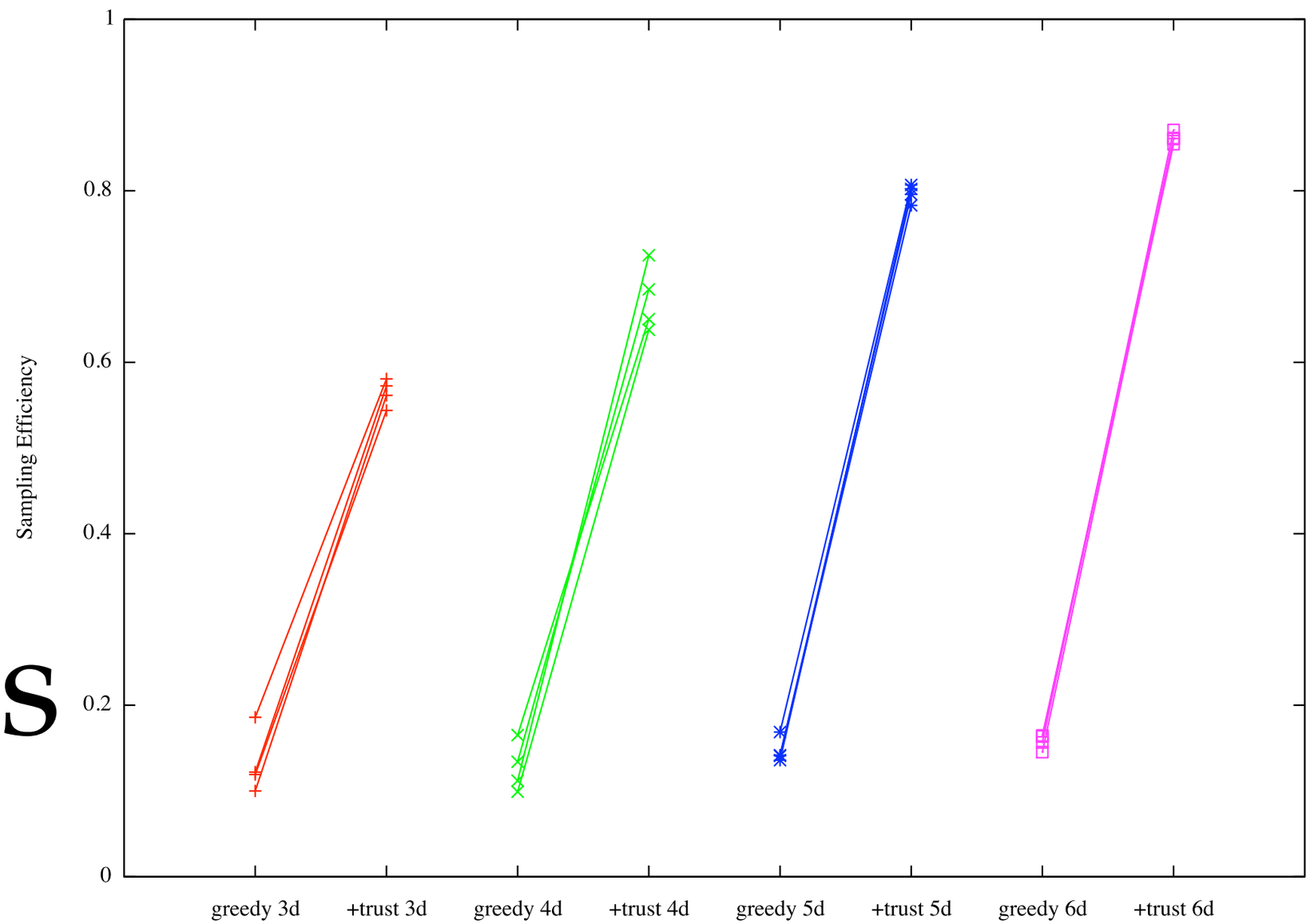
- used peak positions for protein RIa (119-244)
- omitted residues with missing frequencies
- clustered peaks into fat 2D hyperplanes

Robust in 3D



- greedy random optimization limited to 80%
- continuous optimization consistently better
- 2x more samples required with greedy approach

Better in Higher Dimensions



- 17 peaks with full frequency information
- Efficiency increases with dimension (34 samples)
- fewer samples required in higher dimensions (conventionally grows exponentially)

Full Problem

plane	number of peaks	number of samples	efficiency
1	7+2	26	0.99
2	38+2	119	0.72
3	26+2	83	0.82
4	15+2	50	0.87
5	22+2	71	0.80
6	2+2	11	1.00

- overall 88.2% efficient
- 100-fold reduction in sampling for equal noise

Conclusion

- NMR dynamics
  - significant potential cost savings
- NMR structure
  - more complicated prior information
- imaging
  - dimension limited 3d
  - practical problems not yet solvable
- tough dense SDP problems available

Thanks

- Chemistry
  - Alex Bain
  - Giuseppe Melacini
  - Rahul Das
- Advanced Optimization Lab
  - Imre Polik
- CSDP
  - Brian Borchers