Disjoint sparsity for signal separation and applications to hybrid imaging inverse problems

Giovanni S Alberti

Seminar for Applied Mathematics, ETH Zurich

University College London, October 9, 2015

ETH zürich

Outline

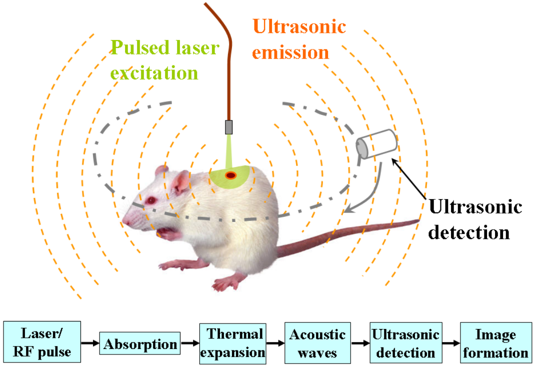
1. Introduction to hybrid imaging inverse problems
2. Disjoint sparsity for signal separation
3. Applications to quantitative photoacoustic tomography

Giovanni S. Alberti and Habib Ammari.
Disjoint sparsity for signal separation and applications to hybrid inverse problems in medical imaging.
Applied and Computational Harmonic Analysis, 2015.
doi:10.1016/j.acha.2015.08.013.

Outline of the talk

1. Introduction to hybrid imaging inverse problems
2. Disjoint sparsity for signal separation
3. Applications to quantitative photoacoustic tomography

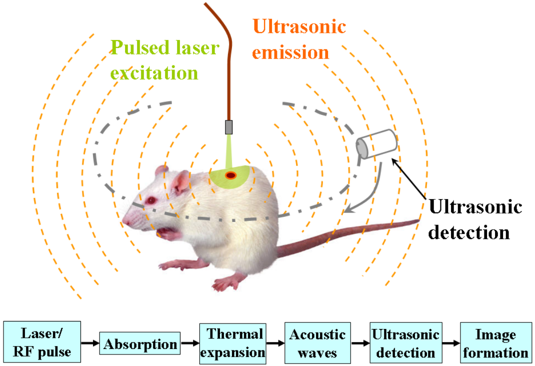
Photoacoustic tomography



(From Wikipedia, http://en.wikipedia.org/wiki/Photoacoustic_imaging_in_biomedicine)

1. The image is $H(x) = \Gamma(x)\mu(x)u(x)$, where
   - μ is the light absorption,
   - Γ is the Grüneisen parameter,
   - and u is the light intensity.
2. How to extract the unknowns from H?

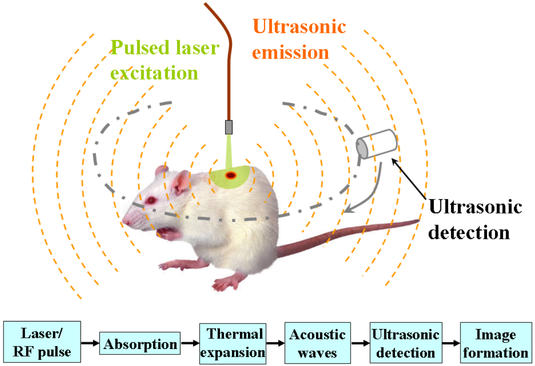
Photoacoustic tomography



(From Wikipedia, http://en.wikipedia.org/wiki/Photoacoustic_imaging_in_biomedicine)

1. The image is $H(x) = \Gamma(x)\mu(x)u(x)$, where
   - μ is the light absorption,
   - Γ is the Grüneisen parameter,
   - and u is the light intensity.
2. How to extract the unknowns from H?

Photoacoustic tomography



(From Wikipedia, http://en.wikipedia.org/wiki/Photoacoustic_imaging_in_biomedicine)

1. The image is $H(x) = \Gamma(x)\mu(x)u(x)$, where
   - μ is the light absorption,
   - Γ is the Grüneisen parameter,
   - and u is the light intensity.
2. How to extract the unknowns from H?

PDE-based methods

- Alessandrini, Arridge, Bal, Beard, Beretta, Cox, Di Cristo, Francini, Gao, Jollivet, Jugnon, Kaipio, Köstli, Laufer, Muszkieta, Naetar, Pulkkinen, Ren, Scherzer, Tarvainen, Uhlmann, Vessella, Zhao, ...
- A possible way to obtain Γ and μ from

$$H = \Gamma \mu u$$

is based on the PDE satisfied by u. In the diffusive regime for light

$$-\mathrm{div}(D\nabla u) + \mu u = 0 \quad \text{in } \Omega.$$

- This approach is sometimes very successful. Possible drawbacks:
  - PDE model non accurate (e.g. transport regime for light), or required boundary conditions not known.
  - Too many unknowns (e.g. if $\Gamma \neq 1$ above)

The focus of this talk is a new approach to this issue based on the separation of the unknowns from the fields via sparsity conditions:

$$h = \log H = \log \Gamma \mu + \log u = f + g.$$

- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u tends to be much smoother.

PDE-based methods

- Alessandrini, Arridge, Bal, Beard, Beretta, Cox, Di Cristo, Francini, Gao, Jollivet, Jugnon, Kaipio, Köstli, Laufer, Muszkieta, Naetar, Pulkkinen, Ren, Scherzer, Tarvainen, Uhlmann, Vessella, Zhao, ...
- A possible way to obtain Γ and μ from

$$H = \Gamma \mu u$$

is based on the PDE satisfied by u. In the diffusive regime for light

$$-\mathrm{div}(D\nabla u) + \mu u = 0 \quad \text{in } \Omega.$$

- This approach is sometimes very successful. Possible drawbacks:
  - PDE model non accurate (e.g. transport regime for light), or required boundary conditions not known.
  - Too many unknowns (e.g. if $\Gamma \neq 1$ above)

The focus of this talk is a new approach to this issue based on the separation of the unknowns from the fields via sparsity conditions:

$$h = \log H = \log \Gamma\mu + \log u = f + g.$$

- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u tends to be much smoother.

PDE-based methods

- Alessandrini, Arridge, Bal, Beard, Beretta, Cox, Di Cristo, Francini, Gao, Jollivet, Jugnon, Kaipio, Köstli, Laufer, Muszkieta, Naetar, Pulkkinen, Ren, Scherzer, Tarvainen, Uhlmann, Vessella, Zhao, ...
- A possible way to obtain Γ and μ from

$$H = \Gamma\mu u$$

is based on the PDE satisfied by u. In the diffusive regime for light

$$-\mathrm{div}(D\nabla u) + \mu u = 0 \quad \text{in } \Omega.$$

- This approach is sometimes very successful. Possible drawbacks:
  - PDE model non accurate (e.g. transport regime for light), or required boundary conditions not known.
  - Too many unknowns (e.g. if $\Gamma \neq 1$ above)

The focus of this talk is a new approach to this issue based on the separation of the unknowns from the fields via sparsity conditions:

$$h = \log H = \log \Gamma\mu + \log u = f + g.$$

- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u tends to be much smoother.

PDE-based methods

- Alessandrini, Arridge, Bal, Beard, Beretta, Cox, Di Cristo, Francini, Gao, Jollivet, Jugnon, Kaipio, Köstli, Laufer, Muszkieta, Naetar, Pulkkinen, Ren, Scherzer, Tarvainen, Uhlmann, Vessella, Zhao, ...
- A possible way to obtain Γ and μ from

$$H = \Gamma\mu u$$

is based on the PDE satisfied by u. In the diffusive regime for light

$$-\mathrm{div}(D\nabla u) + \mu u = 0 \quad \text{in } \Omega.$$

- This approach is sometimes very successful. Possible drawbacks:
  - PDE model non accurate (e.g. transport regime for light), or required boundary conditions not known.
  - Too many unknowns (e.g. if $\Gamma \neq 1$ above)

The focus of this talk is a new approach to this issue based on the separation of the unknowns from the fields via sparsity conditions:

$$h = \log H = \log \Gamma\mu + \log u = f + g.$$

- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u tends to be much smoother.

PDE-based methods

- Alessandrini, Arridge, Bal, Beard, Beretta, Cox, Di Cristo, Francini, Gao, Jollivet, Jugnon, Kaipio, Köstli, Laufer, Muszkieta, Naetar, Pulkkinen, Ren, Scherzer, Tarvainen, Uhlmann, Vessella, Zhao, ...
- A possible way to obtain Γ and μ from

$$H = \Gamma \mu u$$

  is based on the PDE satisfied by u. In the diffusive regime for light

$$-\mathrm{div}(D\nabla u) + \mu u = 0 \quad \text{in } \Omega.$$

- This approach is sometimes very successful. Possible drawbacks:
  - PDE model non accurate (e.g. transport regime for light), or required boundary conditions not known.
  - Too many unknowns (e.g. if $\Gamma \neq 1$ above)

The focus of this talk is a new approach to this issue based on the separation of the unknowns from the fields via sparsity conditions:

$$h = \log H = \log \Gamma\mu + \log u = f + g.$$

- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u tends to be much smoother.

Outline of the talk

Introduction to sparse representations

- Let $h \in \mathbb{R}^n$ be a column vector ($n = d \times d$ is the resolution of the image).
- Let $A \in \mathbb{R}^{n \times m}$ be a dictionary of m atoms, which are used as building blocks:

$$h = Ay, \tag{1}$$

  for some coefficient vector $y \in \mathbb{R}^m$ (weights).
- If $m > n$ then (1) is in general underdetermined, and has many solutions y.
- Select the *sparsest* one, i.e. with fewest non-zero entries:

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } h = Ay,$$

  where $\|y\|_0 := \#\mathrm{supp}\, y = \#\{\alpha \in \{1, \dots, m\} : y(\alpha) \neq 0\}$.
- If the dictionary A is well chosen, it is possible to represent an n-dimensional vector with much fewer coefficients.
- In practice, we minimise

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } \|h - Ay\|_2 \leq \varepsilon$$

  for some $\varepsilon > 0$ (hard problem, but many available algorithms).

Introduction to sparse representations

- Let $h \in \mathbb{R}^n$ be a column vector ($n = d \times d$ is the resolution of the image).
- Let $A \in \mathbb{R}^{n\times m}$ be a dictionary of m atoms, which are used as building blocks:

  (1) $$h = Ay,$$

  for some coefficient vector $y \in \mathbb{R}^m$ (weights).
- If $m > n$ then (1) is in general underdetermined, and has many solutions y.
- Select the *sparsest* one, i.e. with fewest non-zero entries:

  $$\min_{y\in\mathbb{R}^m} \|y\|_0 \quad \text{subject to } h = Ay,$$

  where $\|y\|_0 := \#\operatorname{supp} y = \#\{\alpha \in \{1,\dots,m\} : y(\alpha) \neq 0\}$.
- If the dictionary A is well chosen, it is possible to represent an n-dimensional vector with much fewer coefficients.
- In practice, we minimise

  $$\min_{y\in\mathbb{R}^m} \|y\|_0 \quad \text{subject to } \|h - Ay\|_2 \leq \varepsilon$$

  for some $\varepsilon > 0$ (hard problem, but many available algorithms).

Introduction to sparse representations

- Let $h \in \mathbb{R}^n$ be a column vector ($n = d \times d$ is the resolution of the image).
- Let $A \in \mathbb{R}^{n \times m}$ be a dictionary of m atoms, which are used as building blocks:

$$h = Ay, \tag{1}$$

  for some coefficient vector $y \in \mathbb{R}^m$ (weights).
- If $m > n$ then (1) is in general underdetermined, and has many solutions y.
- Select the *sparsest* one, i.e. with fewest non-zero entries:

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } h = Ay,$$

  where $\|y\|_0 := \#\operatorname{supp} y = \#\{\alpha \in \{1, \ldots, m\} : y(\alpha) \neq 0\}$.
- If the dictionary A is well chosen, it is possible to represent an n-dimensional vector with much fewer coefficients.
- In practice, we minimise

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } \|h - Ay\|_2 \leq \varepsilon$$

  for some $\varepsilon > 0$ (hard problem, but many available algorithms).

Introduction to sparse representations

- Let $h \in \mathbb{R}^n$ be a column vector ($n = d \times d$ is the resolution of the image).
- Let $A \in \mathbb{R}^{n \times m}$ be a dictionary of m atoms, which are used as building blocks:

$$h = Ay, \tag{1}$$

  for some coefficient vector $y \in \mathbb{R}^m$ (weights).
- If $m > n$ then (1) is in general underdetermined, and has many solutions y.
- Select the *sparsest* one, i.e. with fewest non-zero entries:

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } h = Ay,$$

  where $\|y\|_0 := \#\operatorname{supp} y = \#\{\alpha \in \{1, \dots, m\} : y(\alpha) \neq 0\}$.
- If the dictionary A is well chosen, it is possible to represent an n-dimensional vector with much fewer coefficients.
- In practice, we minimise

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } \|h - Ay\|_2 \leq \varepsilon$$

  for some $\varepsilon > 0$ (hard problem, but many available algorithms).

Introduction to sparse representations

- Let $h \in \mathbb{R}^n$ be a column vector ($n = d \times d$ is the resolution of the image).
- Let $A \in \mathbb{R}^{n \times m}$ be a dictionary of m atoms, which are used as building blocks:

$$h = Ay, \tag{1}$$

  for some coefficient vector $y \in \mathbb{R}^m$ (weights).
- If $m > n$ then (1) is in general underdetermined, and has many solutions y.
- Select the *sparsest* one, i.e. with fewest non-zero entries:

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } h = Ay,$$

  where $\|y\|_0 := \#\operatorname{supp} y = \#\{\alpha \in \{1, \dots, m\} : y(\alpha) \neq 0\}$.
- If the dictionary A is well chosen, it is possible to represent an n-dimensional vector with much fewer coefficients.
- In practice, we minimise

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } \|h - Ay\|_2 \leq \varepsilon$$

  for some $\varepsilon > 0$ (hard problem, but many available algorithms).

Introduction to sparse representations

- Let $h \in \mathbb{R}^n$ be a column vector ($n = d \times d$ is the resolution of the image).
- Let $A \in \mathbb{R}^{n \times m}$ be a dictionary of m atoms, which are used as building blocks:

$$h = Ay, \tag{1}$$

  for some coefficient vector $y \in \mathbb{R}^m$ (weights).
- If $m > n$ then (1) is in general underdetermined, and has many solutions y.
- Select the *sparsest* one, i.e. with fewest non-zero entries:

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } h = Ay,$$

  where $\|y\|_0 := \#\operatorname{supp} y = \#\{\alpha \in \{1, \dots, m\} : y(\alpha) \neq 0\}$.
- If the dictionary A is well chosen, it is possible to represent an n-dimensional vector with much fewer coefficients.
- In practice, we minimise

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } \|h - Ay\|_2 \leq \varepsilon$$

  for some $\varepsilon > 0$ (hard problem, but many available algorithms).

Introduction to sparse representations

- Let $h \in \mathbb{R}^n$ be a column vector ($n = d \times d$ is the resolution of the image).
- Let $A \in \mathbb{R}^{n \times m}$ be a dictionary of m atoms, which are used as building blocks:

$$h = Ay, \tag{1}$$

  for some coefficient vector $y \in \mathbb{R}^m$ (weights).
- If $m > n$ then (1) is in general underdetermined, and has many solutions y.
- Select the *sparsest* one, i.e. with fewest non-zero entries:

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } h = Ay,$$

  where $\|y\|_0 := \#\operatorname{supp} y = \#\{\alpha \in \{1, \ldots, m\} : y(\alpha) \neq 0\}$.
- If the dictionary A is well chosen, it is possible to represent an n-dimensional vector with much fewer coefficients.
- In practice, we minimise

$$\min_{y \in \mathbb{R}^m} \|y\|_0 \quad \text{subject to } \|h - Ay\|_2 \leq \varepsilon$$

  for some $\varepsilon > 0$ (hard problem, but many available algorithms).

Morphological component analysis (Starck, Elad, Donoho, 2004,...)

Back to the signal separation problem.

- Let $h = f + g \in \mathbb{R}^n$ be the sum of two components.
- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries such that:
  - f can be sparsely represented w.r.t. A_f but not w.r.t. A_g;
  - g can be sparsely represented w.r.t. A_g but not w.r.t. A_f;
- Decompose h w.r.t. the concatenated dictionary $A = [A_f, A_g]$:

$$\min_{y \in \mathbb{R}^{m_f + m_g}} \|y\|_0 \quad \text{subject to } [A_f, A_g] \begin{bmatrix} y_f \\ y_g \end{bmatrix} = h.$$

- Recover $f \approx A_f y_f$, $g \approx A_g y_g$.

In other words, the original components can be recovered provided that they have different features. This is expressed in terms of the sparsity of their decompositions with different dictionaries.

Morphological component analysis (Starck, Elad, Donoho, 2004,...)

Back to the signal separation problem.

- Let $h = f + g \in \mathbb{R}^n$ be the sum of two components.
- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries such that:
  - f can be sparsely represented w.r.t. A_f but not w.r.t. A_g;
  - g can be sparsely represented w.r.t. A_g but not w.r.t. A_f;
- Decompose h w.r.t. the concatenated dictionary $A = [A_f, A_g]$:

$$\min_{y \in \mathbb{R}^{m_f + m_g}} \|y\|_0 \quad \text{subject to } [A_f, A_g] \begin{bmatrix} y_f \\ y_g \end{bmatrix} = h.$$

- Recover $f \approx A_f y_f$, $g \approx A_g y_g$.

In other words, the original components can be recovered provided that they have different features. This is expressed in terms of the sparsity of their decompositions with different dictionaries.

Morphological component analysis (Starck, Elad, Donoho, 2004,...)

Back to the signal separation problem.

- Let $h = f + g \in \mathbb{R}^n$ be the sum of two components.
- Let $A_f \in \mathbb{R}^{n\times m_f}$ and $A_g \in \mathbb{R}^{n\times m_g}$ be two dictionaries such that:
  - f can be sparsely represented w.r.t. A_f but not w.r.t. A_g;
  - g can be sparsely represented w.r.t. A_g but not w.r.t. A_f;
- Decompose h w.r.t. the concatenated dictionary $A = [A_f, A_g]$:

$$\min_{y\in\mathbb{R}^{m_f+m_g}} \|y\|_0 \quad \text{subject to } [A_f, A_g]\begin{bmatrix} y_f \\ y_g \end{bmatrix} = h.$$

- Recover $f \approx A_f y_f$, $g \approx A_g y_g$.

In other words, the original components can be recovered provided that they have different features. This is expressed in terms of the sparsity of their decompositions with different dictionaries.

Morphological component analysis (Starck, Elad, Donoho, 2004,...)

Back to the signal separation problem.

- Let $h = f + g \in \mathbb{R}^n$ be the sum of two components.
- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries such that:
  - f can be sparsely represented w.r.t. A_f but not w.r.t. A_g;
  - g can be sparsely represented w.r.t. A_g but not w.r.t. A_f;
- Decompose h w.r.t. the concatenated dictionary $A = [A_f, A_g]$:

$$\min_{y \in \mathbb{R}^{m_f + m_g}} \|y\|_0 \quad \text{subject to } [A_f, A_g] \begin{bmatrix} y_f \\ y_g \end{bmatrix} = h.$$

- Recover $f \approx A_f y_f$, $g \approx A_g y_g$.

In other words, the original components can be recovered provided that they have different features. This is expressed in terms of the sparsity of their decompositions with different dictionaries.

Morphological component analysis (Starck, Elad, Donoho, 2004,...)

Back to the signal separation problem.

- Let $h = f + g \in \mathbb{R}^n$ be the sum of two components.
- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries such that:
  - f can be sparsely represented w.r.t. A_f but not w.r.t. A_g;
  - g can be sparsely represented w.r.t. A_g but not w.r.t. A_f;
- Decompose h w.r.t. the concatenated dictionary $A = [A_f, A_g]$:

$$\min_{y \in \mathbb{R}^{m_f + m_g}} \|y\|_0 \quad \text{subject to } [A_f, A_g] \begin{bmatrix} y_f \\ y_g \end{bmatrix} = h.$$

- Recover $f \approx A_f y_f$, $g \approx A_g y_g$.

In other words, the original components can be recovered provided that they have different features. This is expressed in terms of the sparsity of their decompositions with different dictionaries.

Morphological component analysis (Starck, Elad, Donoho, 2004,...)

Back to the signal separation problem.

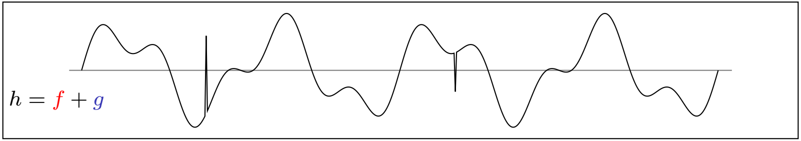
- Let $h = f + g \in \mathbb{R}^n$ be the sum of two components.
- Let $A_f \in \mathbb{R}^{n\times m_f}$ and $A_g \in \mathbb{R}^{n\times m_g}$ be two dictionaries such that:
  - f can be sparsely represented w.r.t. A_f but not w.r.t. A_g;
  - g can be sparsely represented w.r.t. A_g but not w.r.t. A_f;
- Decompose h w.r.t. the concatenated dictionary $A = [A_f, A_g]$:

$$\min_{y\in\mathbb{R}^{m_f+m_g}} \|y\|_0 \quad \text{subject to } [A_f, A_g]\begin{bmatrix} y_f \\ y_g \end{bmatrix} = h.$$

- Recover $f \approx A_f y_f$, $g \approx A_g y_g$.

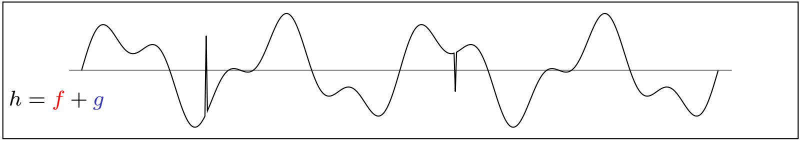
In other words, the original components can be recovered provided that they have different features. This is expressed in terms of the sparsity of their decompositions with different dictionaries.

Example: spikes and sinusoids (Donoho, Huo, 2001,...)

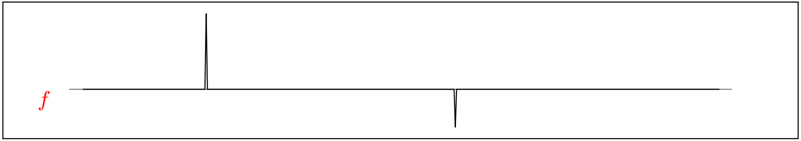


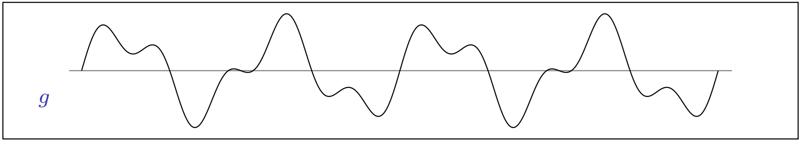
Choose $A_f = A_\delta$ and $A_g = A_F$ and let $y = \begin{bmatrix} y_f \\ y_g \end{bmatrix}$ give the sparsest representation of h w.r.t. $A = [A_f, A_g]$. This clearly provides the right reconstruction: only 4 atoms are used.

Example: spikes and sinusoids (Donoho, Huo, 2001,...)



Choose $A_f = A_\delta$ and $A_g = A_F$ and let $y = \begin{bmatrix} y_f \\ y_g \end{bmatrix}$ give the sparsest representation of h w.r.t. $A = [A_f, A_g]$. This clearly provides the right reconstruction: only 4 atoms are used.





Theoretical justification (Elad & Bruckstein, 2002)

Why does writing $f = A_f y_f$ and $g = A_g y_g$ give the correct reconstruction?

- (Uncertainty principle) If $h \neq 0$ has the representations $h = Ay_A = By_B$ w. r. t. two orthonormal bases $A = [a_1, \dots, a_n]$ and $B = [b_1, \dots, b_n]$, then

  $$\|y_A\|_0 + \|y_B\|_0 \geq 2/M,$$

  where $M = \max_{i,j} |(a_i, b_j)_2|$ is the *mutual coherence*. ($M = 1/\sqrt{n}$ with spikes and sinusoids.)
- If f and g have representations y_f and y_g satisfying $\|y_f\|_0 + \|y_g\|_0 < 1/M$, then the reconstruction is correct.
- In practice, the assumption $\|y_f\|_0 + \|y_g\|_0 < 1/M$ is almost never satisfied, and so the above argument remains only a theoretical speculation.

Focus of this new approach: extension to multiple measurements (many other generalisations by Georgiev, Theis, Cichocki, Bobin, Moudden, Starck, Donoho, Kutyniok...).

Theoretical justification (Elad & Bruckstein, 2002)

Why does writing $f = A_f y_f$ and $g = A_g y_g$ give the correct reconstruction?

- (Uncertainty principle) If $h \neq 0$ has the representations $h = Ay_A = By_B$ w. r. t. two orthonormal bases $A = [a_1, \dots, a_n]$ and $B = [b_1, \dots, b_n]$, then

$$\|y_A\|_0 + \|y_B\|_0 \geq 2/M,$$

where $M = \max_{i,j} |(a_i, b_j)_2|$ is the *mutual coherence*. ($M = 1/\sqrt{n}$ with spikes and sinusoids.)

- If f and g have representations y_f and y_g satisfying $\|y_f\|_0 + \|y_g\|_0 < 1/M$, then the reconstruction is correct.
- In practice, the assumption $\|y_f\|_0 + \|y_g\|_0 < 1/M$ is almost never satisfied, and so the above argument remains only a theoretical speculation.

Focus of this new approach: extension to multiple measurements (many other generalisations by Georgiev, Theis, Cichocki, Bobin, Moudden, Starck, Donoho, Kutyniok...).

Theoretical justification (Elad & Bruckstein, 2002)

Why does writing $f = A_f y_f$ and $g = A_g y_g$ give the correct reconstruction?

- (Uncertainty principle) If $h \neq 0$ has the representations $h = Ay_A = By_B$ w. r. t. two orthonormal bases $A = [a_1, \ldots, a_n]$ and $B = [b_1, \ldots, b_n]$, then

$$\|y_A\|_0 + \|y_B\|_0 \geq 2/M,$$

  where $M = \max_{i,j} |(a_i, b_j)_2|$ is the *mutual coherence*. ($M = 1/\sqrt{n}$ with spikes and sinusoids.)
- If f and g have representations y_f and y_g satisfying $\|y_f\|_0 + \|y_g\|_0 < 1/M$, then the reconstruction is correct.
- In practice, the assumption $\|y_f\|_0 + \|y_g\|_0 < 1/M$ is almost never satisfied, and so the above argument remains only a theoretical speculation.

Focus of this new approach: extension to multiple measurements (many other generalisations by Georgiev, Theis, Cichocki, Bobin, Moudden, Starck, Donoho, Kutyniok...).

Theoretical justification (Elad & Bruckstein, 2002)

Why does writing $f = A_f y_f$ and $g = A_g y_g$ give the correct reconstruction?

- (Uncertainty principle) If $h \neq 0$ has the representations $h = Ay_A = By_B$ w. r. t. two orthonormal bases $A = [a_1, \ldots, a_n]$ and $B = [b_1, \ldots, b_n]$, then

$$\|y_A\|_0 + \|y_B\|_0 \geq 2/M,$$

  where $M = \max_{i,j} |(a_i, b_j)_2|$ is the *mutual coherence*. ($M = 1/\sqrt{n}$ with spikes and sinusoids.)
- If f and g have representations y_f and y_g satisfying $\|y_f\|_0 + \|y_g\|_0 < 1/M$, then the reconstruction is correct.
- In practice, the assumption $\|y_f\|_0 + \|y_g\|_0 < 1/M$ is almost never satisfied, and so the above argument remains only a theoretical speculation.

Focus of this new approach: extension to multiple measurements (many other generalisations by Georgiev, Theis, Cichocki, Bobin, Moudden, Starck, Donoho, Kutyniok...).

Theoretical justification (Elad & Bruckstein, 2002)

Why does writing $f = A_f y_f$ and $g = A_g y_g$ give the correct reconstruction?

- (Uncertainty principle) If $h \neq 0$ has the representations $h = Ay_A = By_B$ w. r. t. two orthonormal bases $A = [a_1, \ldots, a_n]$ and $B = [b_1, \ldots, b_n]$, then

$$\|y_A\|_0 + \|y_B\|_0 \geq 2/M,$$

  where $M = \max_{i,j} |(a_i, b_j)_2|$ is the *mutual coherence*. ($M = 1/\sqrt{n}$ with spikes and sinusoids.)
- If f and g have representations y_f and y_g satisfying $\|y_f\|_0 + \|y_g\|_0 < 1/M$, then the reconstruction is correct.
- In practice, the assumption $\|y_f\|_0 + \|y_g\|_0 < 1/M$ is almost never satisfied, and so the above argument remains only a theoretical speculation.

Focus of this new approach: extension to multiple measurements (many other generalisations by Georgiev, Theis, Cichocki, Bobin, Moudden, Starck, Donoho, Kutyniok...).

Multi-measurement case

- Let
$$h_i = f + g_i \in \mathbb{R}^n, \qquad i = 1, \dots, N$$
be N measurements. The problem is to recover f and the g_i's.
- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries as before. Assume that A_g is an orthonormal set (and that A_f is an orthonormal basis). In our applications:
  - $A_f =$ Haar wavelets,
  - $A_g =$ low frequency sinusoids.
- The reconstruction method applied here consists in the minimisation of
$$\min_{y \in \mathbb{R}^{m_f + N m_g}} \|y\|_0 \quad \text{subject to } \left\| [A_f, A_g] \begin{bmatrix} y_f \\ y_g^i \end{bmatrix} - h_i \right\|_2 \leq \varepsilon, \ i = 1, \dots, N,$$
where $y = {}^t[{}^t y_f, {}^t y_g^1, \dots, {}^t y_g^N]$.
- The noisy case: $h_i = f + g_i + n_i$, $\|n_i\|_2 \leq \eta$ for some $\eta > 0$.
- Why does this provide a better reconstruction?

Multi-measurement case

- Let
$$h_i = f + g_i \in \mathbb{R}^n, \qquad i = 1, \dots, N$$
be N measurements. The problem is to recover f and the g_i's.
- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries as before. Assume that A_g is an orthonormal set (and that A_f is an orthonormal basis). In our applications:
  - $A_f =$ Haar wavelets,
  - $A_g =$ low frequency sinusoids.
- The reconstruction method applied here consists in the minimisation of
$$\min_{y \in \mathbb{R}^{m_f + N m_g}} \|y\|_0 \quad \text{subject to } \left\| [A_f, A_g] \begin{bmatrix} y_f \\ y_g^i \end{bmatrix} - h_i \right\|_2 \leq \varepsilon, \; i = 1, \dots, N,$$
where $y = {}^t[{}^t y_f, {}^t y_g^1, \dots, {}^t y_g^N]$.
- The noisy case: $h_i = f + g_i + n_i$, $\|n_i\|_2 \leq \eta$ for some $\eta > 0$.
- Why does this provide a better reconstruction?

Multi-measurement case

- Let
$$h_i = f + g_i \in \mathbb{R}^n, \qquad i = 1, \ldots, N$$
be N measurements. The problem is to recover f and the g_i's.
- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries as before. Assume that A_g is an orthonormal set (and that A_f is an orthonormal basis). In our applications:
  - $A_f =$ Haar wavelets,
  - $A_g =$ low frequency sinusoids.
- The reconstruction method applied here consists in the minimisation of
$$\min_{y \in \mathbb{R}^{m_f + N m_g}} \|y\|_0 \quad \text{subject to} \quad \left\| [A_f, A_g] \begin{bmatrix} y_f \\ y_g^i \end{bmatrix} - h_i \right\|_2 \leq \varepsilon,\ i = 1, \ldots, N,$$
where $y = {}^t[{}^t y_f, {}^t y_g^1, \ldots, {}^t y_g^N]$.
- The noisy case: $h_i = f + g_i + n_i$, $\|n_i\|_2 \leq \eta$ for some $\eta > 0$.
- Why does this provide a better reconstruction?

Multi-measurement case

- Let
$$h_i = f + g_i \in \mathbb{R}^n, \qquad i = 1, \dots, N$$
be N measurements. The problem is to recover f and the g_i's.
- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries as before. Assume that A_g is an orthonormal set (and that A_f is an orthonormal basis). In our applications:
  - $A_f =$ Haar wavelets,
  - $A_g =$ low frequency sinusoids.
- The reconstruction method applied here consists in the minimisation of
$$\min_{y \in \mathbb{R}^{m_f + N m_g}} \|y\|_0 \quad \text{subject to } \left\| [A_f, A_g] \begin{bmatrix} y_f \\ y_g^i \end{bmatrix} - h_i \right\|_2 \leq \varepsilon, \ i = 1, \dots, N,$$
where $y = {}^t[{}^t y_f, {}^t y_g^1, \dots, {}^t y_g^N]$.
- The noisy case: $h_i = f + g_i + n_i$, $\|n_i\|_2 \leq \eta$ for some $\eta > 0$.
- Why does this provide a better reconstruction?

Multi-measurement case

- Let

$$h_i = f + g_i \in \mathbb{R}^n, \qquad i = 1, \dots, N$$

be N measurements. The problem is to recover f and the g_i's.

- Let $A_f \in \mathbb{R}^{n \times m_f}$ and $A_g \in \mathbb{R}^{n \times m_g}$ be two dictionaries as before. Assume that A_g is an orthonormal set (and that A_f is an orthonormal basis). In our applications:
  - $A_f =$ Haar wavelets,
  - $A_g =$ low frequency sinusoids.
- The reconstruction method applied here consists in the minimisation of

$$\min_{y \in \mathbb{R}^{m_f + N m_g}} \|y\|_0 \quad \text{subject to} \quad \left\| [A_f, A_g] \begin{bmatrix} y_f \\ y_g^i \end{bmatrix} - h_i \right\|_2 \leq \varepsilon, \; i = 1, \dots, N,$$

where $y = {}^t[{}^t y_f, {}^t y_g^1, \dots, {}^t y_g^N]$.

- The noisy case: $h_i = f + g_i + n_i$, $\|n_i\|_2 \leq \eta$ for some $\eta > 0$.
- Why does this provide a better reconstruction?

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\|f - A_f \tilde{y}_f\right\|_2 \leq \rho_f \quad \text{and} \quad \left\|g_i - A_g \tilde{y}_g^i\right\|_2 \leq \rho_g.$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \leq \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \neq 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\|p\right\|_2 > D$ and $\left\|{}^t A_g^{\perp} p\right\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \#\left(\{\alpha : |({}^t A_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i\right) > \# \bigcup_{i=1}^{N} \operatorname{supp} \tilde{y}_g^i + \left\|\tilde{y}_f\right\|_0.$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\| f - A_f \tilde{y}_f \right\|_2 \le \rho_f \quad \text{and} \quad \left\| g_i - A_g \tilde{y}_g^i \right\|_2 \le \rho_g.$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \le \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \ne 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\|p\right\|_2 > D$ and $\left\| {}^t A_g^\perp p \right\|_2 \le 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^N \# \left(\{\alpha : |({}^t A_g p)(\alpha)| \ge 1\} \setminus \operatorname{supp} \tilde{y}_g^i \right) > \# \bigcup_{i=1}^N \operatorname{supp} \tilde{y}_g^i + \left\|\tilde{y}_f\right\|_0.$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\| f - A_f \tilde{y}_f \right\|_2 \leq \rho_f \quad \text{and} \quad \left\| g_i - A_g \tilde{y}_g^i \right\|_2 \leq \rho_g.$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \leq \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \neq 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\|p\right\|_2 > D$ and $\left\| {}^t A_g^\perp p \right\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^N \# \left(\{\alpha : |({}^t A_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i \right) > \# \bigcup_{i=1}^N \operatorname{supp} \tilde{y}_g^i + \left\|\tilde{y}_f\right\|_0.$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\| f - A_f \tilde{y}_f \right\|_2 \leq \rho_f \quad \text{and} \quad \left\| g_i - A_g \tilde{y}_g^i \right\|_2 \leq \rho_g.$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \leq \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \neq 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\|p\right\|_2 > D$ and $\left\| {}^tA_g^{\perp} p \right\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^tA_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \#\left(\{\alpha : |({}^tA_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i\right) > \# \bigcup_{i=1}^{N} \operatorname{supp} \tilde{y}_g^i + \|\tilde{y}_f\|_0.$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\| f - A_f \tilde{y}_f \right\|_2 \leq \rho_f \quad \text{and} \quad \left\| g_i - A_g \tilde{y}_g^i \right\|_2 \leq \rho_g .$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \leq \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \neq 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\|p\right\|_2 > D$ and $\left\| {}^t A_g^{\perp} p \right\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \# \left(\{\alpha : |({}^t A_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i \right) > \# \bigcup_{i=1}^{N} \operatorname{supp} \tilde{y}_g^i + \|\tilde{y}_f\|_0 .$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\| f - A_f \tilde{y}_f \right\|_2 \leq \rho_f \quad \text{and} \quad \left\| g_i - A_g \tilde{y}_g^i \right\|_2 \leq \rho_g.$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \leq \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \neq 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\| p \right\|_2 > D$ and $\left\| {}^t A_g^\perp p \right\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \# \left(\{\alpha : |({}^t A_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i \right) > \# \bigcup_{i=1}^{N} \operatorname{supp} \tilde{y}_g^i + \left\| \tilde{y}_f \right\|_0.$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\| f - A_f \tilde{y}_f \right\|_2 \leq \rho_f \quad \text{and} \quad \left\| g_i - A_g \tilde{y}_g^i \right\|_2 \leq \rho_g.$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \leq \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \neq 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\|p\right\|_2 > D$ and $\left\| {}^t A_g^{\perp} p \right\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \# \left(\{\alpha : |({}^t A_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i \right) > \# \bigcup_{i=1}^{N} \operatorname{supp} \tilde{y}_g^i + \left\|\tilde{y}_f\right\|_0.$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\| f - A_f \tilde{y}_f \right\|_2 \leq \rho_f \quad \text{and} \quad \left\| g_i - A_g \tilde{y}_g^i \right\|_2 \leq \rho_g.$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \leq \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \neq 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\|p\right\|_2 > D$ and $\left\| {}^t A_g^\perp p \right\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^N \# \left(\{\alpha : |({}^t A_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i \right) > \# \bigcup_{i=1}^N \operatorname{supp} \tilde{y}_g^i + \left\|\tilde{y}_f\right\|_0.$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Assumptions

- In the applications we have in mind, we shall have (indirect) control on the g_i's.
- We need enough incoherent data: this is measured by their disjoint sparsity.
- Write for some $\rho_f, \rho_g > 0$:

$$\left\| f - A_f \tilde{y}_f \right\|_2 \leq \rho_f \quad \text{and} \quad \left\| g_i - A_g \tilde{y}_g^i \right\|_2 \leq \rho_g.$$

Take $\beta, D > 0$. Assume that:

1. if $|\tilde{y}_g^i(\alpha) - \tilde{y}_g^j(\alpha)| \leq \beta$ and $\tilde{y}_g^i(\alpha)\tilde{y}_g^j(\alpha) \neq 0$ for some α then $i = j$;
2. for every $p \in \mathbb{R}^n$ such that $\left\|p\right\|_2 > D$ and $\left\| {}^t A_g^\perp p \right\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \# \left(\{\alpha : |({}^t A_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i \right) > m_g + \left\|\tilde{y}_f\right\|_0.$$

- 1 is mainly a technical assumption.
- 2 is at the core of the approach: multiple and disjointly sparse measurements $\tilde{y}_g^i$ make it easy to be satisfied.

Uncertainty principles

2. ... for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^t A_g^\perp p\|_2 \le 2/3$ there holds

$$\#(\operatorname{supp} {}^t A_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \# \left(\{\alpha : |({}^t A_g p)(\alpha)| \ge 1\} \setminus \operatorname{supp} \tilde{y}_g^i\right) > m_g + \|\tilde{y}_f\|_0.$$

- "Classical". If M is the mutual coherence of A_f and A_g:

$$\|{}^t A_f p\|_0 + \|{}^t A_g p\|_0 \ge 2/M.$$

- "Normalised". $\exists D > 0$ s.t. for all $p \in \mathbb{R}^n$ with $\|p\|_2 > D$ and $\|{}^t A_g^\perp p\|_2 \le 2/3$

$$\|{}^t A_f p\|_0 + \#\{\alpha : |({}^t A_g p)(\alpha)| \ge 1\} \ge 2/M.$$

Unfortunately, $M \sim 1$ if A_f consists of wavelets and A_g of sinusoids...

- "Haar wavelets". Let A_f be the orthobasis of 2D Haar wavelets in $\mathbb{R}^{2^7 \times 2^7}$ and let A_g be 960 low frequency non-constant sinusoids. There exists $D > 0$ s.t.

$$\|{}^t A_f p\|_0 \ge 1160,$$

for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^t A_g^\perp p\|_2 \le 2/3$.

Uncertainty principles

2. ... for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \le 2/3$ there holds

$$\#(\operatorname{supp} {}^tA_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \#\left(\{\alpha : |({}^tA_g p)(\alpha)| \ge 1\} \setminus \operatorname{supp} \tilde{y}_g^i\right) > m_g + \|\tilde{y}_f\|_0.$$

- "Classical". If M is the mutual coherence of A_f and A_g:

$$\|{}^tA_f p\|_0 + \|{}^tA_g p\|_0 \ge 2/M.$$

- "Normalised". $\exists D > 0$ s.t. for all $p \in \mathbb{R}^n$ with $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \le 2/3$

$$\|{}^tA_f p\|_0 + \#\{\alpha : |({}^tA_g p)(\alpha)| \ge 1\} \ge 2/M.$$

Unfortunately, $M \sim 1$ if A_f consists of wavelets and A_g of sinusoids...

- "Haar wavelets". Let A_f be the orthobasis of 2D Haar wavelets in $\mathbb{R}^{2^7 \times 2^7}$ and let A_g be 960 low frequency non-constant sinusoids. There exists $D > 0$ s.t.

$$\|{}^tA_f p\|_0 \ge 1160,$$

for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \le 2/3$.

Uncertainty principles

2. ... for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \le 2/3$ there holds

$$\#(\operatorname{supp} {}^tA_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \# \left(\{\alpha : |({}^tA_g p)(\alpha)| \ge 1\} \setminus \operatorname{supp} \tilde{y}_g^i\right) > m_g + \|\tilde{y}_f\|_0.$$

- "Classical". If M is the mutual coherence of A_f and A_g:

$$\|{}^tA_f p\|_0 + \|{}^tA_g p\|_0 \ge 2/M.$$

- "Normalised". $\exists D > 0$ s.t. for all $p \in \mathbb{R}^n$ with $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \le 2/3$

$$\|{}^tA_f p\|_0 + \#\{\alpha : |({}^tA_g p)(\alpha)| \ge 1\} \ge 2/M.$$

Unfortunately, $M \sim 1$ if A_f consists of wavelets and A_g of sinusoids...

- "Haar wavelets". Let A_f be the orthobasis of 2D Haar wavelets in $\mathbb{R}^{2^7 \times 2^7}$ and let A_g be 960 low frequency non-constant sinusoids. There exists $D > 0$ s.t.

$$\|{}^tA_f p\|_0 \ge 1160,$$

for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \le 2/3$.

Uncertainty principles

2. ... for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^{\perp} p\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^tA_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \#\left(\{\alpha : |({}^tA_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i\right) > m_g + \|\tilde{y}_f\|_0.$$

- "Classical". If M is the mutual coherence of A_f and A_g:

$$\|{}^tA_f p\|_0 + \|{}^tA_g p\|_0 \geq 2/M.$$

- "Normalised". $\exists D > 0$ s.t. for all $p \in \mathbb{R}^n$ with $\|p\|_2 > D$ and $\|{}^tA_g^{\perp} p\|_2 \leq 2/3$

$$\|{}^tA_f p\|_0 + \#\{\alpha : |({}^tA_g p)(\alpha)| \geq 1\} \geq 2/M.$$

Unfortunately, $M \sim 1$ if A_f consists of wavelets and A_g of sinusoids...

- "Haar wavelets". Let A_f be the orthobasis of 2D Haar wavelets in $\mathbb{R}^{2^7 \times 2^7}$ and let A_g be 960 low frequency non-constant sinusoids. There exists $D > 0$ s.t.

$$\|{}^tA_f p\|_0 \geq 1160,$$

for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^{\perp} p\|_2 \leq 2/3$.

Uncertainty principles

2. ... for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^{\perp} p\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^tA_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \# \left(\{\alpha : |({}^tA_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i\right) > m_g + \|\tilde{y}_f\|_0.$$

- "Classical". If M is the mutual coherence of A_f and A_g:

$$\|{}^tA_f p\|_0 + \|{}^tA_g p\|_0 \geq 2/M.$$

- "Normalised". $\exists D > 0$ s.t. for all $p \in \mathbb{R}^n$ with $\|p\|_2 > D$ and $\|{}^tA_g^{\perp} p\|_2 \leq 2/3$

$$\|{}^tA_f p\|_0 + \#\{\alpha : |({}^tA_g p)(\alpha)| \geq 1\} \gtrsim 2.$$

Unfortunately, $M \sim 1$ if A_f consists of wavelets and A_g of sinusoids...

- "Haar wavelets". Let A_f be the orthobasis of 2D Haar wavelets in $\mathbb{R}^{2^7 \times 2^7}$ and let A_g be 960 low frequency non-constant sinusoids. There exists $D > 0$ s.t.

$$\|{}^tA_f p\|_0 \geq 1160,$$

for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^{\perp} p\|_2 \leq 2/3$.

Uncertainty principles

2. ... for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \leq 2/3$ there holds

$$\#(\operatorname{supp} {}^tA_f p \setminus \operatorname{supp} \tilde{y}_f) + \sum_{i=1}^{N} \# \left(\{\alpha : |({}^tA_g p)(\alpha)| \geq 1\} \setminus \operatorname{supp} \tilde{y}_g^i\right) > m_g + \|\tilde{y}_f\|_0.$$

- "Classical". If M is the mutual coherence of A_f and A_g:

$$\|{}^tA_f p\|_0 + \|{}^tA_g p\|_0 \geq 2/M.$$

- "Normalised". $\exists D > 0$ s.t. for all $p \in \mathbb{R}^n$ with $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \leq 2/3$

$$\|{}^tA_f p\|_0 + \#\{\alpha : |({}^tA_g p)(\alpha)| \geq 1\} \gtrsim 2.$$

  Unfortunately, $M \sim 1$ if A_f consists of wavelets and A_g of sinusoids...

- "Haar wavelets". Let A_f be the orthobasis of 2D Haar wavelets in $\mathbb{R}^{2^7 \times 2^7}$ and let A_g be 960 low frequency non-constant sinusoids. There exists $D > 0$ s.t.

$$\|{}^tA_f p\|_0 \geq 1160,$$

  for every $p \in \mathbb{R}^n$ such that $\|p\|_2 > D$ and $\|{}^tA_g^\perp p\|_2 \leq 2/3$.

Main result

The following result states that the separation method with multiple measurements gives unique and stable reconstruction.

Theorem

Assume 1 and 2 and that $\varepsilon := \rho_f + \rho_g + \eta \leq \beta/3$. *Assume that* $f, g_i, n_i \in \mathbb{R}^n$ *satisfy* $\|n_i\|_2 \leq \eta$ *and*

$$\left\|A_f \tilde{y}_f - f\right\|_2 \leq \rho_f, \quad \left\|A_g \tilde{y}_g^i - g_i\right\|_2 \leq \rho_g, \qquad i = 1, \ldots, N,$$

and let $y_f \in \mathbb{R}^{m_f}$ *and* $y_g^i \in \mathbb{R}^{m_g}$ *realise the minimum of*

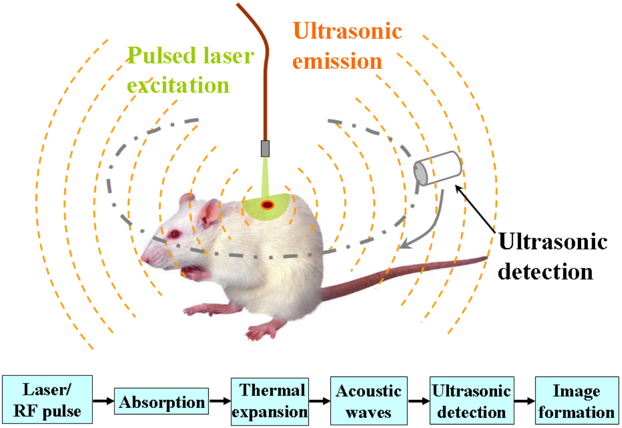
$$\min_{y \in \mathbb{R}^{m_f + N m_g}} \|y\|_0 \quad \text{subject to } \left\|[A_f, A_g]\begin{bmatrix} y_f \\ y_g^i \end{bmatrix} - h_i\right\|_2 \leq \varepsilon,\ i = 1, \ldots, N,$$

where $h_i = f + g_i + n_i$. *Then*

$$\left\|A_f y_f - f\right\|_2 \leq (3D+1)\varepsilon, \quad \left\|A_g y_g^i - g_i\right\|_2 \leq (3D+2)\varepsilon, \qquad i = 1, \ldots, N.$$

Outline of the talk

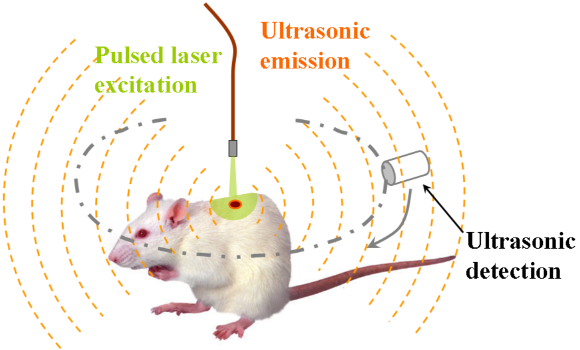
Quantitative photoacoustic tomography, $\Gamma = 1$

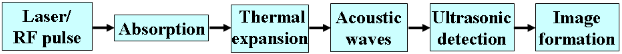


It gives us the image $H(x) = \Gamma(x)\mu(x)u(x) = \mu(x)u(x)$. Need to find μ.

In the general case $\Gamma \neq 1$, apply the same approach and find $\Gamma\mu$ and u. Then by using the PDE approach all the unknowns can be reconstructed.

Quantitative photoacoustic tomography, $\Gamma = 1$

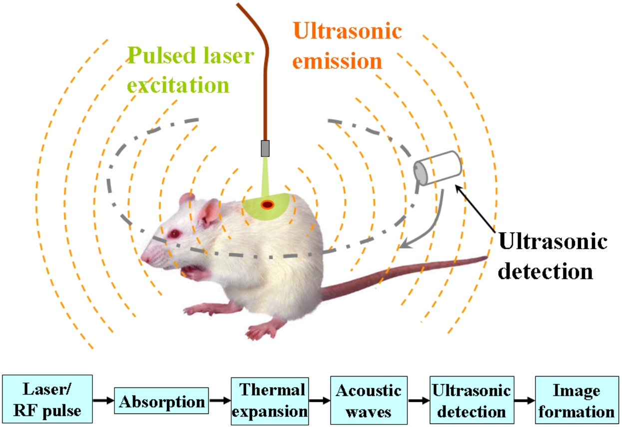




It gives us the image $H(x) = \Gamma(x)\mu(x)u(x) = \mu(x)u(x)$. Need to find μ.

In the general case $\Gamma \neq 1$, apply the same approach and find $\Gamma\mu$ and u. Then by using the PDE approach all the unknowns can be reconstructed.

Quantitative photoacoustic tomography, $\Gamma = 1$



It gives us the image $H(x) = \Gamma(x)\mu(x)u(x) = \mu(x)u(x)$. Need to find μ.

In the general case $\Gamma \neq 1$, apply the same approach and find $\Gamma\mu$ and u. Then by using the PDE approach all the unknowns can be reconstructed.

Quantitative photoacoustic tomography, $\Gamma = 1$

- Multi-measurement data: $H(x) = \mu(x)u_i(x)$, where

$$\begin{cases} -\Delta u_i + \mu u_i = 0 & \text{in } \Omega, \\ u_i = \varphi_i & \text{on } \partial\Omega. \end{cases}$$

- Taking a log: $h_i = \log \mu + \log u_i$.
- The above method may be used if $\log \mu$ and $\log u_i$ can be sparsely represented with respect to two different dictionaries.
- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u_i is a solution of a PDE, and as such enjoys higher regularity properties.
- Thus, the uncertainty principle motivates these for the dictionaries:
  - A_f: Haar wavelets (curvelets, ridgelets, shearlets...)
  - A_g: low frequency trigonometric polynomials (constant included!)
- The disjoint sparsity signal separation method may be used to recover μ.

Quantitative photoacoustic tomography, $\Gamma = 1$

- Multi-measurement data: $H(x) = \mu(x)u_i(x)$, where

$$\begin{cases} -\Delta u_i + \mu u_i = 0 & \text{in } \Omega, \\ u_i = \varphi_i & \text{on } \partial\Omega. \end{cases}$$

- Taking a log: $h_i = \log \mu + \log u_i$.
- The above method may be used if $\log \mu$ and $\log u_i$ can be sparsely represented with respect to two different dictionaries.
- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u_i is a solution of a PDE, and as such enjoys higher regularity properties.
- Thus, the uncertainty principle motivates these for the dictionaries:
  - A_f: Haar wavelets (curvelets, ridgelets, shearlets...)
  - A_g: low frequency trigonometric polynomials (constant included!)
- The disjoint sparsity signal separation method may be used to recover μ.

Quantitative photoacoustic tomography, $\Gamma = 1$

- Multi-measurement data: $H(x) = \mu(x) u_i(x)$, where

$$\begin{cases} -\Delta u_i + \mu u_i = 0 & \text{in } \Omega, \\ u_i = \varphi_i & \text{on } \partial\Omega. \end{cases}$$

- Taking a log: $h_i = \log \mu + \log u_i$.
- The above method may be used if $\log \mu$ and $\log u_i$ can be sparsely represented with respect to two different dictionaries.
- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u_i is a solution of a PDE, and as such enjoys higher regularity properties.
- Thus, the uncertainty principle motivates these for the dictionaries:
  - A_f: Haar wavelets (curvelets, ridgelets, shearlets...)
  - A_g: low frequency trigonometric polynomials (constant included!)
- The disjoint sparsity signal separation method may be used to recover μ.

Quantitative photoacoustic tomography, $\Gamma = 1$

- Multi-measurement data: $H(x) = \mu(x)u_i(x)$, where

$$\begin{cases} -\Delta u_i + \mu u_i = 0 & \text{in } \Omega, \\ u_i = \varphi_i & \text{on } \partial\Omega. \end{cases}$$

- Taking a log: $h_i = \log \mu + \log u_i$.
- The above method may be used if $\log \mu$ and $\log u_i$ can be sparsely represented with respect to two different dictionaries.
- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u_i is a solution of a PDE, and as such enjoys higher regularity properties.
- Thus, the uncertainty principle motivates these for the dictionaries:
  - A_f: Haar wavelets (curvelets, ridgelets, shearlets...)
  - A_g: low frequency trigonometric polynomials (constant included!)
- The disjoint sparsity signal separation method may be used to recover μ.

Quantitative photoacoustic tomography, $\Gamma = 1$

- Multi-measurement data: $H(x) = \mu(x) u_i(x)$, where

$$\begin{cases} -\Delta u_i + \mu u_i = 0 & \text{in } \Omega, \\ u_i = \varphi_i & \text{on } \partial\Omega. \end{cases}$$

- Taking a log: $h_i = \log \mu + \log u_i$.
- The above method may be used if $\log \mu$ and $\log u_i$ can be sparsely represented with respect to two different dictionaries.
- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u_i is a solution of a PDE, and as such enjoys higher regularity properties.
- Thus, the uncertainty principle motivates these for the dictionaries:
  - A_f: Haar wavelets (curvelets, ridgelets, shearlets...)
  - A_g: low frequency trigonometric polynomials (constant included!)
- The disjoint sparsity signal separation method may be used to recover μ.

Quantitative photoacoustic tomography, $\Gamma = 1$

- Multi-measurement data: $H(x) = \mu(x)u_i(x)$, where

$$\begin{cases} -\Delta u_i + \mu u_i = 0 & \text{in } \Omega, \\ u_i = \varphi_i & \text{on } \partial\Omega. \end{cases}$$

- Taking a log: $h_i = \log \mu + \log u_i$.
- The above method may be used if $\log \mu$ and $\log u_i$ can be sparsely represented with respect to two different dictionaries.
- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u_i is a solution of a PDE, and as such enjoys higher regularity properties.
- Thus, the uncertainty principle motivates these for the dictionaries:
  - A_f: Haar wavelets (curvelets, ridgelets, shearlets...)
  - A_g: low frequency trigonometric polynomials (constant included!)
- The disjoint sparsity signal separation method may be used to recover μ.

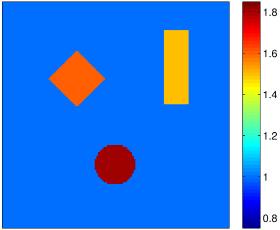
Quantitative photoacoustic tomography, $\Gamma = 1$

- Multi-measurement data: $H(x) = \mu(x) u_i(x)$, where

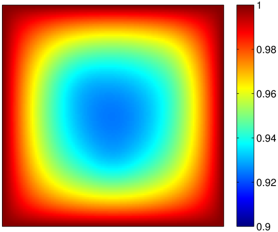
$$\begin{cases} -\Delta u_i + \mu u_i = 0 & \text{in } \Omega, \\ u_i = \varphi_i & \text{on } \partial\Omega. \end{cases}$$

- Taking a log: $h_i = \log \mu + \log u_i$.
- The above method may be used if $\log \mu$ and $\log u_i$ can be sparsely represented with respect to two different dictionaries.
- Following [Rosenthal, Razansky, Ntziachristos, 2009], observe that:
  - the light absorption μ is a constitutive parameter of the tissue, and as such is discontinuous. Its discontinuities are typically the inclusions we are looking for;
  - the light intensity u_i is a solution of a PDE, and as such enjoys higher regularity properties.
- Thus, the uncertainty principle motivates these for the dictionaries:
  - A_f: Haar wavelets (curvelets, ridgelets, shearlets...)
  - A_g: low frequency trigonometric polynomials (constant included!)
- The disjoint sparsity signal separation method may be used to recover μ.

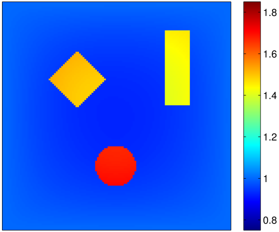
Example 1: noise-free case, $N = 1$



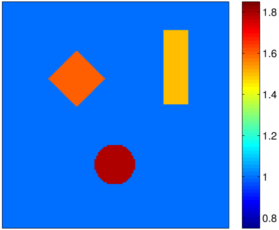
(a) The true absorption $\tilde{\mu}$



(b) The true intensity $\tilde{u}_1$



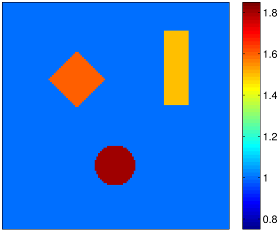
(c) The datum H_1



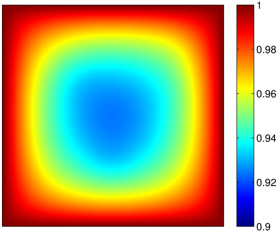
(d) The reconstructed μ

Video: $N = 1$. Video: $N = 2$.

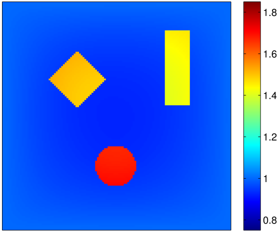
Example 1: noise-free case, $N = 1$



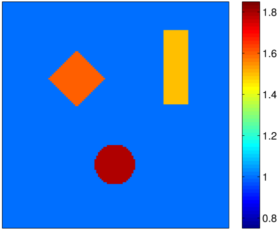
(a) The true absorption $\tilde{\mu}$



(b) The true intensity $\tilde{u}_1$



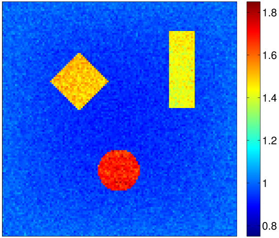
(c) The datum H_1



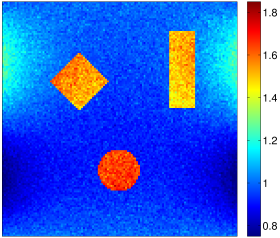
(d) The reconstructed μ

Video: $N = 1$. Video: $N = 2$.

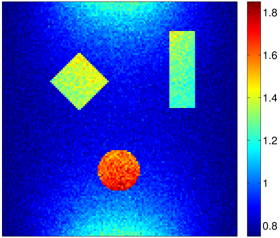
Example 1: noisy case, $N = 1, 3, 5$



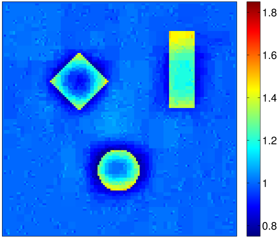
(a) The datum H_1



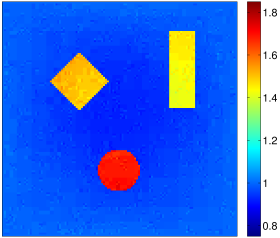
(b) The datum H_3



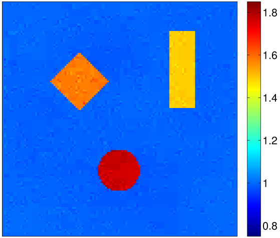
(c) The datum H_5



(d) Reconstructed μ, $N = 1$

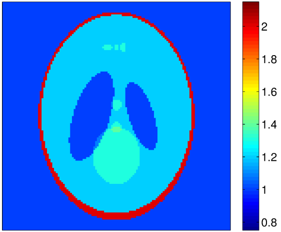


(e) Reconstructed μ, $N = 3$

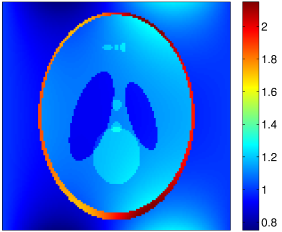


(f) Reconstructed μ, $N = 5$

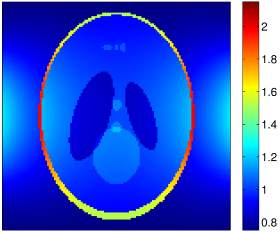
Example 2: the Shepp-Logan Phantom



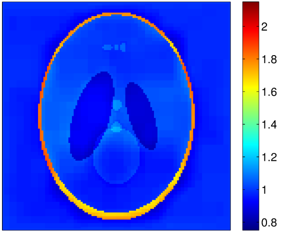
(a) $\tilde{\mu}$



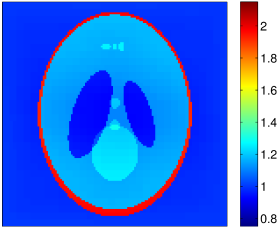
(b) H_2



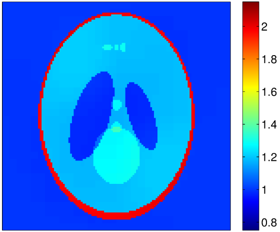
(c) H_4



(d) μ, $N = 1$

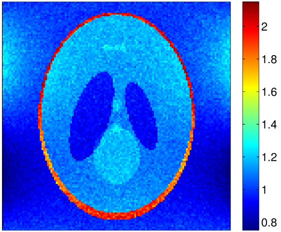


(e) μ, $N = 3$

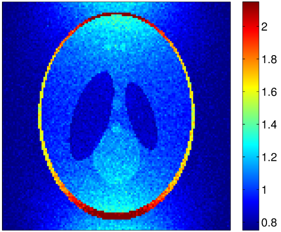


(f) μ, $N = 5$

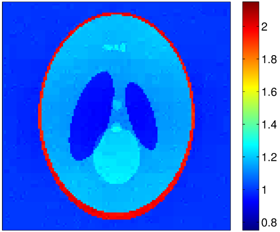
Example 2: the Shepp-Logan Phantom with noise



(a) H_3

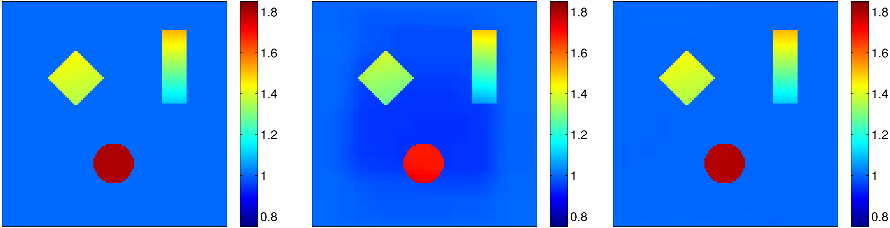


(b) H_5



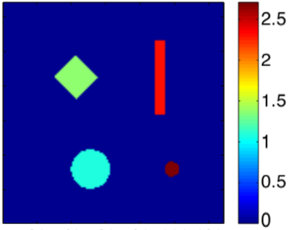
(c) μ, $N = 5$

Example 3: piecewise smooth

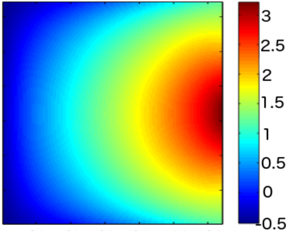


(a) $\tilde{\mu}$ (b) $\mu,\ N = 1$ (c) $\mu,\ N = 2$

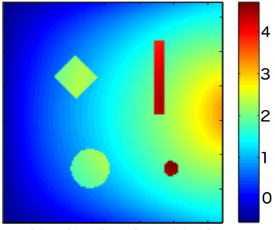
Example 4: lateral illuminations, $N = 4$



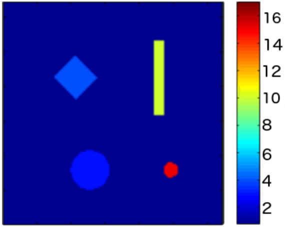
(a) $\log \tilde{\mu}$



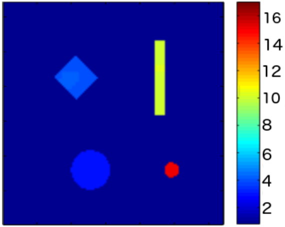
(b) $\log \tilde{u}_1$



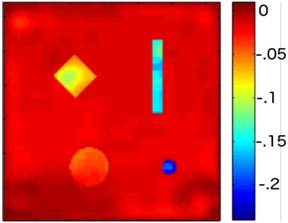
(c) $\log H_1$



(d) $\tilde{\mu}$



(e) μ



(f) $\|\mu - \tilde{\mu}\|_2 \approx 1.5 \cdot 10^{-2}$

Conclusions

Past

- The reconstruction in QPAT (and in many hybrid imaging inverse problems) require the separation of several signals $h_i = f + g_i$.
- PDE techniques are often powerful, but sometimes they are not applicable.

Present

- Multiple measurements and disjoint sparsity can be used to find f and g_i.
- Uniqueness and stability proven.
- Orthogonal matching pursuit performs well in many numerical simulations related to quantitative photoacoustic tomography.

Future

- How can we ensure that the light intensities u_i give the necessary incoherence, measured in terms of their disjoint sparsity? Random illuminations may help.
- Norm $l^0 \rightarrow$ norm l^1.
- Robust uncertainty principles (Candes, Romberg, 2006)

Conclusions

Past

- The reconstruction in QPAT (and in many hybrid imaging inverse problems) require the separation of several signals $h_i = f + g_i$.
- PDE techniques are often powerful, but sometimes they are not applicable.

Present

- Multiple measurements and disjoint sparsity can be used to find f and g_i.
- Uniqueness and stability proven.
- Orthogonal matching pursuit performs well in many numerical simulations related to quantitative photoacoustic tomography.

Future

- How can we ensure that the light intensities u_i give the necessary incoherence, measured in terms of their disjoint sparsity? Random illuminations may help.
- Norm $l^0 \rightarrow$ norm l^1.
- Robust uncertainty principles (Candes, Romberg, 2006)

Conclusions

Past

- The reconstruction in QPAT (and in many hybrid imaging inverse problems) require the separation of several signals $h_i = f + g_i$.
- PDE techniques are often powerful, but sometimes they are not applicable.

Present

- Multiple measurements and disjoint sparsity can be used to find f and g_i.
- Uniqueness and stability proven.
- Orthogonal matching pursuit performs well in many numerical simulations related to quantitative photoacoustic tomography.

Future

- How can we ensure that the light intensities u_i give the necessary incoherence, measured in terms of their disjoint sparsity? Random illuminations may help.
- Norm $l^0 \rightarrow$ norm l^1.
- Robust uncertainty principles (Candes, Romberg, 2006)